

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Desenvolvimento de uma metodologia
computacional para determinação
automática de petrofácies sedimentares em
poços de exploração de petróleo**

Camila Martins Saporetti

JUIZ DE FORA
JULHO, 2014

Desenvolvimento de uma metodologia computacional para determinação automática de petrofácies sedimentares em poços de exploração de petróleo

CAMILA MARTINS SAPORETTI

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação
Orientador: Leonardo Goliatt da Fonseca

JUIZ DE FORA
JULHO, 2014

DESENVOLVIMENTO DE UMA METODOLOGIA
COMPUTACIONAL PARA DETERMINAÇÃO AUTOMÁTICA DE
PETROFÁCIES SEDIMENTARES EM POÇOS DE EXPLORAÇÃO
DE PETRÓLEO

Camila Martins Saporetti

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Leonardo Goliatt da Fonseca
Doutor em Modelagem Computacional

Custódio Gouvêa Lopes da Motta
Doutor em Engenharia Civil

Leonardo Costa de Oliveira
Mestre em Geologia

JUIZ DE FORA
17 DE JULHO, 2014

*Às minhas meninas, pela força, companhia,
amizade. Por caminharem junto comigo.
Por me proporcionarem momentos de alegria
e por confiarem em mim. Obrigada.*

Resumo

A determinação e o mapeamento de heterogeneidades de reservatórios de hidrocarbonetos é estrategicamente importante para a caracterização destes, uma vez que podem estabelecer uma relação entre a porosidade e permeabilidade, definindo a produtividade e o caráter comercial de um campo de óleo e gás. As petrofácies sedimentares são um conjunto de características petrográficas que individualizam um grupo de rochas, e sua determinação permite a inferência da heterogeneidade do reservatório. Porém este processo usualmente é muito longo e nem toda informação é aproveitada, devido à grande quantidade de dados. Conseqüentemente, torna-se interessante a mudança do uso de métodos manuais para análises automáticas usando ferramentas computacionais. Este trabalho apresenta um método computacional apto a identificar petrofácies e separá-las de acordo com suas características comuns. Dessa forma, o método é capaz de auxiliar o geólogo/petrólogo na tarefa de identificar petrofácies sedimentares.

Palavras-chave: Inteligência Computacional, Mineração de Dados, Métodos de Agrupamentos, Petrografia, Diagênese.

Abstract

To the characterization of hydrocarbon reservoir is strategically important both detection and mapping of its heterogeneity, as it can establish a relationship between porosity and permeability, defining both productivity and commercial character of a field of oil and gas. The sedimentary petrofacies are a set of characteristics that individualize a particular group of rocks, and their determination allows the inference of the heterogeneity of the reservoir's heterogeneity. But this process is usually very long and not all information is harnessed, due to the large amount of data. Consequently, it becomes interesting change of use of manual methods for automatic analysis using computational tools. This work aims to present a computational method able to identify and separate them according to their common features. Thus, the method is able to assist the geologist/petrologist the task of identifying sedimentary petrofacies.

Keywords: Computational Intelligence, Data Mining, Clusters Methods, Petrography, Diagenesis.

Agradecimentos

Agradeço a todos aqueles que de alguma forma contribuíram para que esse trabalho se concretizasse.

À minha mãe e irmã, agradeço pela paciência, pelo apoio e por acreditarem que eu conseguiria. Vocês são fundamentais na minha vida.

À Francislaine, agradeço por ter feito da sua família a minha. Obrigado por sempre estar disponível quando preciso, por me escutar e por ter me dado apoio quando mais precisei.

Aos professores Maicon Ribeiro, Felipe Loureiro, Lorenza Moreno por serem exemplos, pessoas as quais admiro, agradeço pelo incentivo e pelos ensinamentos.

Ao Leonardo Goliatt agradeço a orientação, a confiança, a dedicação e claro a paciência. Obrigada pelo incentivo, por sempre me mostrar o que ainda pode ser feito e as portas que podem se abrir.

Ao Leonardo Oliveira, ao Egberto Pereira e à Juliana Cevolani pela colaboração. Sem dúvidas foram fundamentais para a construção deste trabalho.

Ao professor Custódio Motta, agradeço pelos conceitos de Mineração de Dados ensinados. Foram muito importantes para que alguns conceitos ficassem esclarecidos.

À Ariane e à Jacquelinny agradeço pela companhia, pelos conselhos. Vocês se tornaram minhas irmãzinhas.

Às minhas meninas, Anna Claudia, Bárbara, Isis, Letícia, Liliane e Stephanie, agradeço pela companhia, pela paciência, por sempre estarem perto, por me escutarem. Enfim por tudo, vocês foram mais que fundamentais para que este trabalho se realizasse.

À Alessandra por ser essa pessoa com o coração imenso, obrigada pela amizade e por sempre estar disposta a me ajudar.

Ao Ronaro e ao Wesley agradeço pela amizade, por terem se colocado disponível sempre que precisei. Vocês são especiais.

À UFJF agradeço pelos auxílios disponibilizados, que foram fundamentais para

meu desenvolvimento acadêmico e pessoal.

“Que os vossos esforços desafiem as impossibilidades, lembrai-vos de que as grandes coisas do homem foram conquistadas do que parecia impossível.”

Charles Chaplin

Sumário

Lista de Figuras	9
Lista de Tabelas	10
1 Introdução	11
1.1 Contextualização	11
1.2 Justificativa	12
1.3 Objetivos	14
1.3.1 Objetivos Gerais	14
1.3.2 Objetivos Específicos	14
1.4 Estrutura do Trabalho	15
2 Revisão Bibliográfica	16
3 Fundamentação Teórica	22
3.1 Contexto Geológico	22
3.1.1 Formação de Bacias Sedimentares	22
3.1.2 Ambientes de Sedimentação e Fácies Sedimentares	23
3.1.3 Composição e Gênese dos Hidrocarbonetos	24
3.2 Rocha Reservatório	25
3.3 Formação da Bacia do Paraná	25
3.3.1 Qualidade dos Reservatórios de Exploração e Petrofácies	27
3.4 Análise de Agrupamento	28
3.4.1 Medidas de Dissimilaridade	29
3.4.2 Técnicas de Análise de Agrupamento	31
3.4.3 Critérios de Validação	36
3.4.4 Métricas de Avaliação Externa para Agrupamentos	37
3.5 Análise de Componentes Principais	39
3.5.1 Processo para uma Análise de Componentes Principais	39
4 Metodologia	43
4.1 Caracterização da Base de Dados	43
4.2 Procedimento de Análise Intraçoço	47
4.3 Procedimento de Análise Interçoço	48
5 Experimentos Computacionais	51
5.1 Análise Intraçoço	51
5.1.1 Coeficiente de Silhueta	51
5.1.2 Índice de Calinski-Harabasz	55
5.2 Análise Interçoço	60
6 Conclusão e Trabalhos Futuros	63
Referências Bibliográficas	65
A Base de Dados	70

B	Resultado Análise Intrapoço	76
C	Média dos Constituintes	82

Lista de Figuras

3.1	Eras Geológicas (Milani, 1997).	26
3.2	Métodos de particionamento (Boschi, 2013)	31
3.3	Métodos aglomerativo e divisivo (Kaufman, 1990).	33
3.4	Fluxograma ilustrando o algoritmo DBSCAN	35
4.1	Localização da Bacia do Paraná e a posição dos poços (pontos na cor vermelha) (modificado de (Oliveira, 2009)).	44
5.1	Visualização da Análise Intrapoço para os furos PPG1, PPG2 (Silhueta).	53
5.2	Visualização da Análise Intrapoço para os furos PPG3 e PPG4 (Silhueta).	54
5.3	Visualização da Análise Intrapoço para o furo PPG5 (Silhueta).	55
5.4	Visualização da Análise Intrapoço para os furos PPG1 (Calinski-Harabasz).	56
5.5	Visualização da Análise Intrapoço para os furos PPG2 e PPG3 (Calinski-Harabasz).	57
5.6	Visualização da Análise Intrapoço para o furo PPG4 e PPG5 (Calinski-Harabasz).	58
5.7	Visualização do Primeiro Experimento da Análise Interpoço.	60
5.8	Visualização das petrofácies separadas pelo método convencional.	61
5.9	Visualização da distribuição dos grupos encontrados no Segundo Experimento da Análise Interpoço.	62

Lista de Tabelas

3.1	Dissimilaridades utilizadas	30
4.1	Petrográficos analisados	45
4.2	Amostras retiradas do PPG5	46
4.3	Resultado DBSCAN para o PPG5	49
4.4	Média dos Percentuais dos Constituintes para o PPG5	50
5.1	Parâmetros que maximizaram a silhueta e os valores encontrados para homogeneidade e para integralidade	52
5.2	Parâmetros que maximizaram o índice de Calinski-Harabasz e os valores encontrados para homogeneidade e para integralidade	56
5.3	Relação encontrada para o método DBSCAN	61
A.1	Amostras retiradas do PPG1	71
A.2	Amostras retiradas do PPG2	72
A.3	Amostras retiradas do PPG3	73
A.4	Amostras retiradas do PPG4	74
A.5	Amostras retiradas do PPG5	75
B.1	Resultado DBSCAN para o PPG1	77
B.2	Resultado DBSCAN para o PPG2	78
B.3	Resultado DBSCAN para o PPG3	79
B.4	Resultado DBSCAN para o PPG4	80
B.5	Resultado DBSCAN para o PPG5	81
C.1	Média dos Percentuais dos Constituintes para o PPG1	82
C.2	Média dos Percentuais dos Constituintes para o PPG2	83
C.3	Média dos Percentuais dos Constituintes para o PPG3	84
C.4	Média dos Percentuais dos Constituintes para o PPG4	85
C.5	Média dos Percentuais dos Constituintes para o PPG5	86

1 Introdução

1.1 Contextualização

Nas últimas décadas muitos trabalhos foram desenvolvidos utilizando técnicas de mineração de dados. Isso vem acontecendo pela necessidade de se obter informações relevantes em grandes bases de dados (Liao, 2012). Aplicações podem ser vistas na indústria de óleo e gás, como delimitação de fronteiras de um reservatório de gás natural (Rocha, 2005), análise do potencial de reservatórios de petróleo (Zhang, 1999).

A análise multivariada aparece como uma maneira eficiente de correlacionar amostras, através da análise de múltiplas medidas sobre o objeto estudado. Estudos nas áreas de sociologia (Fonseca, 2008), biologia (Machado *et al*, 2002) e geofísica (Pereira, 2009) estão fazendo uso dos métodos de análise multivariada.

No processo de exploração de petróleo é fundamental ter conhecimento sobre a qualidade e o potencial de um reservatório. Para tal análise, o estudo da diagênese das rochas é de vital importância. O termo diagênese pode ser descrito como um conjunto de fenômenos físicos e químicos que transformam sedimentos desagregados em rochas sedimentares. As propriedades das rochas são investigadas através da análise em campo, quando trata-se de litofácies ou em laboratório, no caso de petrofácies.

As litofácies sedimentares são um aglomerado de características físicas e orgânicas macroscópicas das rochas, já as petrofácies sedimentares são um conjunto de características petrográficas microscópicas, como por exemplo, cor, granulometria, estrutura sedimentar e determinada história diagenética que individualizam um grupo de rochas. Sua determinação permite a inferência da heterogeneidade do reservatório.

Para determinar as fácies (litofácies e petrofácies) é imprescindível realizar um estudo na subsuperfície (testemunhagem) e em muitos casos isso não é viável. Então realiza-se o estudo das logfácies, também conhecidas como eletrofácies, pois nessas situações, tem-se disponível apenas os dados dos perfis de poços. Logfácies é o conjunto de características físicas e químicas das rochas que podem ser identificadas através de perfis

(Rosa *et al*, 2008). A partir do estudo desses dados é feita a identificação das logfácies (Tavakoli, 2006).

Tavakoli (2006) utilizou da análise de agrupamento multivariada para determinar as logfácies. O processo manual de identificação de logfácies acontece através da análise das curvas dos logs, o que pode dar margem para outras interpretações. A partir da interpretação dos perfis, podem ser definidos o topo e a base das eletrofácies por meio dos padrões das curvas e dos valores medidos por cada ferramenta perfilada. Com a necessidade de automatizar esse métodos, técnicas estatísticas e matemáticas foram aplicadas, a fim de obter um resultado mais estável.

Rosa *et al* (2008) apresentaram um método para caracterização automática de eletrofácies, onde utiliza ferramentas de análise multivariada para a convergência entre a classificação de litofácies provenientes de testemunhos e as identificou através dos perfis geofísicos de poços de petróleo, na essência de se caracterizarem eletrofácies. O procedimento convencional é realizado através de softwares que possibilitam a operação com dados de perfis geológicos de poços e os dados associados aos dados de testemunho. O problema está na associação e na representatividade de diferentes tipos de dados.

1.2 Justificativa

A Formação Ponta Grossa é o sistema petrolífero mais importante da Bacia do Paraná (Milani, 1999). Este é subdividido em três membros: Membro Jaguariaíva (inferior), Membro Tibagi (médio) e Membro São Domingos (superior). Segundo (Grahm, 1999), as idades para os três membros são, respectivamente praguiana-emsiana, emsiana e eifeliana-eofrasniana. Essa Formação consiste de folhelhos argilosos, micáceos, finamente lamina-dos, cinzentos, e folhelhos sílticos a arenosos, com siltitos e arenitos muito finos subordinados (Petri, 1983). Os arenitos finos a muito finos, depositados pelo Membro Tibagi no Eifeliano, compreendem possíveis reservatórios para um hipotético sistema petrolífero Ponta Grossa (Oliveira, 2009). A superposição de variados e intensos processos dia-genéticos sobre a deposição ocasionou em uma alta heterogeneidade nesses reservatórios. Isso dificulta a realização de procedimentos de produção e faz com que a recuperação de hidrocarbonetos seja limitada.

A determinação e o mapeamento de heterogeneidades em reservatórios de hidrocarbonetos são estrategicamente importantes para a caracterização destes, uma vez que podem estabelecer uma relação entre a porosidade e permeabilidade, definindo a produtividade e o caráter comercial de um campo de óleo e gás (Cevolani *et al*, 2011). Para uma rocha ser julgada como um bom reservatório, ela deve possuir as seguintes características: uma extensão considerável, boa porosidade, uma apreciável permeabilidade e um eficiente fator de recuperação de hidrocarbonetos. Estas características são denominadas petrofísicas e estão intimamente ligados à história deposicional da bacia, em particular às condições de sedimentação e aos fenômenos de diagênese, sendo de fundamental importância para definição da qualidade do reservatório.

O estudo da diagênese das rochas vem sendo incentivado pelas empresas petrolíferas, com intuito de entender a distribuição da porosidade em arenitos. O interesse vem do fato que estes arenitos podem se constituírem em rochas-reservatórios de hidrocarbonetos. No decorrer do processo de diagênese, minerais podem precipitar-se como cimento nos poros da rocha, o que resulta na diminuição de sua porosidade e da permeabilidade, prejudicando seu potencial como reservatório (Maraschin, 2008).

A partir da análise petrográfica é possível identificar, por meio de um microscópio de luz polarizada, os constituintes de uma rocha. Dessa forma, pode-se realizar uma avaliação das implicações futuras de suas propriedades sobre o comportamento dos produtos formados, como o petróleo (Menezes, 1999). Essa análise ocorre com o uso de microscópio petrográfico, onde o geólogo/petrólogo descreve as lâminas discriminando seus aspectos geológicos. Uma base de dados é criada através das observações realizadas e pode-se, com isso, agrupar os dados em diferentes petrofácies.

Os modelos preditivos bem como o mapeamento da distribuição de heterogeneidades e qualidade em reservatórios são primordiais para exploração e otimização da produção de campos de petróleo. Essa análise pode ser realizada através da contagem em microscópio petrográfico de luz transmitida, a fim de avaliar as heterogeneidades presentes no reservatório estudado. Este processo usualmente é muito longo, pois envolve o processo de amostragem, geração dos dados e posterior interpretação destes, porém nem toda informação é aproveitada, devido à grande quantidade de dados. Conseqüentemente,

torna-se interessante a mudança do uso de métodos manuais para análises automáticas por meio de ferramentas computacionais.

Nesse contexto, métodos de agrupamentos aparecem como um mecanismo útil para auxiliar na identificação de petrofácies. Os métodos de agrupamentos são técnicas de Mineração de Dados cujo propósito consiste em separar objetos de dados em grupos, baseando-se nos atributos que estes objetos possuem. Os objetos que pertencem a um mesmo grupo são similares entre eles, possuindo características semelhantes. Possibilitando que objetos que tenham passado despercebido sejam classificados.

1.3 Objetivos

1.3.1 Objetivos Gerais

O objetivo principal deste projeto é estudar a eficiência do algoritmo de agrupamento DBSCAN (Density Based Spatial Clustering of Applications with Noise) no procedimento de automatizar a identificação e classificação de petrofácies sedimentares. Com o desenvolvimento do método computacional pretende-se auxiliar o geólogo/petrólogo, na análise de dados geográficos.

1.3.2 Objetivos Específicos

Os seguintes objetivos específicos podem ser citados:

1. Estudar e implementar os algoritmos de agrupamentos clássicos. As petrofácies são identificadas a partir dos agrupamentos encontrados;
2. Estudar e implementar critérios para encontrar agrupamentos ótimos, para dessa forma garantir melhores resultados;
3. Realizar uma Análise Intraçoço, para avaliar cada poço de forma individual e uma Análise Interçoço, para estimar como as petrofácies se comportam quando as informações de todos os poços estão reunidas em uma única base de dados;
4. Implementar um modo de visualizar os resultados obtidos de forma que a análise fique mais clara e intuitiva.

1.4 Estrutura do Trabalho

Esse trabalho subdivide-se em seis capítulos. No Capítulo 1 encontra-se uma apresentação dessa monografia que é composta pela contextualização, justificativa e os objetivos.

No Capítulo 2 é exibida a revisão bibliográfica, onde constam trabalhos que utilizaram técnicas de mineração de dados afim de descobrir relações nos dados.

No Capítulo 3 apresenta-se a fundamentação teórica, onde os principais conceitos utilizados para o desenvolvimento desse trabalho estão descritos. Informações sobre diagênese, petrofácies, análise de agrupamentos serão encontradas nesse capítulo.

No Capítulo 4 está descrita a metodologia empregada para o desenvolvimento do projeto em questão, ou seja, na criação de um método computacional capaz de identificar e classificar petrofácies.

O Capítulo 5 é composto pelos resultados encontrados, e uma avaliação da Análise Intrapoço e da Análise Interpoço.

O Capítulo 6 apresenta a conclusão com base nos resultados obtidos e as perspectivas para trabalhos futuros.

2 Revisão Bibliográfica

Devido ao aumento constante na geração de dados, as técnicas tradicionais de exploração de dados não eram mais adequadas para tratar a grande maioria dos repositórios. Com intuito de suprir essa necessidade foi proposta a Mineração de Dados.

A Mineração de Dados pode ser definida como um processo em que há colaboração de homens e máquinas, que tem como objetivo, por meio da exploração de grandes bancos de dados, extrair conhecimento a partir do reconhecimento de padrão e relacionamento entre variáveis. Esse conhecimento podem ser obtidos por técnicas confiáveis, validadas pela sua expressividade estatística (Côrtes, 2002).

Nesse contexto, destacam-se trabalhos relacionados a determinação de propriedades das rochas e identificação de litofácies, eletrofácies e petrofácies. Técnicas para reconhecimento de padrões supervisionados e não-supervisionados estão sendo apresentadas para resolver as questões citadas acima. Dentre os métodos mais aplicados, encontram-se: os algoritmos de agrupamentos, técnicas baseadas em Redes Neurais Artificiais e k-vizinhos mais próximos.

O interesse em métodos multivariados já se espalhou para vários campos de investigação. Podem ser citados como exemplos: educação, química, física, geologia, engenharia, literatura, mineração, biologia (Rencher, 2002). A causa desse interesse é a necessidade em tratar grandes volumes de dados contendo muitas variáveis. Dessa maneira, pretende-se preservar as informações contidas nos bancos de dados.

A partir da criação dos sistemas computacionais, um dos objetivos cruciais das organizações tem sido o armazenamento de dados. Nos últimos anos essa tendência ficou mais notória, pois houve queda nos preços para compra hardware, sendo possível armazenar mais quantidade de dados. O que motivou o desenvolvimento de estruturas de armazenamento mais eficientes, como banco de dados, Web, Data Warehouses (Camilo, 2009).

Lee (1999) propôs uma abordagem para prever a permeabilidade utilizando regressão não paramétrica em conjunto com a análise estatística multivariada. Primeiro foi

realizada a classificação dos dados de perfis de poços em eletrofácies, usando análise discriminante. Em seguida foi aplicada técnicas de regressão não paramétricas para prever a permeabilidade utilizando perfis de poços dentro de cada eletrofácies. Foram examinadas três abordagens não-paramétricas Expectativa Condicional Alternada (ECA), o Modelo Aditivo Generalizado (MAG) e Redes Neurais e as vantagens e desvantagens relativas são exploradas.

Chang (2002) apresentou uma técnica de baixo custo que automatiza o processo de identificação de litofácies. A técnica proposta usa mapas auto-organizáveis (SOM), a fim de identificar de forma sistemática e objetiva litofácies a partir de dados, os quais possuem registros dos poços. SOMs são Redes Neurais Artificiais não-supervisionados que mapeiam o espaço de entrada em agrupamentos em uma forma topológica cuja organização está relacionada com as similaridades dos dados de entrada.

Cunha (2002) propôs um método que usa uma abordagem baseada em Redes Neurais para descobrir conhecimento em uma base de dados de perfis e testemunhos de poços de petróleo e posteriormente identificar as litofácies. A abordagem proposta no trabalho de (Cunha, 2002) é dividida em cinco partes: 1) preparação dos dados para treinamentos, 2) treinamento da rede, 3) predição de litofácies, 4) conversão de TWT para profundidade e 5) impressão de litofácies para a área de estudo. Foi empregado o algoritmo dos k-vizinhos mais próximos com o intuito de obter a melhor combinação entre os dados sísmicos e as correspondentes litofácies, para dessa maneira obter o conjunto de treinamento.

O desenvolvimento de técnicas de perfuração de poços para extração de óleo e gás proporcionou a coleta de grandes quantidades de dados pelas empresas do setor. Como o número de dados aumenta com o passar do tempo, a dificuldade de processar os dados e interpretá-los torna-se maior, tendo em vista à tomada de decisões (Toebe, 2003).

Considere a análise de custo/benefício de um reservatório de óleo e gás natural, visando a sua exploração comercial. Para que a análise seja válida, é fundamental fazer uma descrição precisa da estrutura litológica do reservatório (Toebe, 2003).

Hsieh (2005) desenvolveu um estudo para construir um sistema de litologia fuzzy de perfis de poços para identificar a formação de litologia de um sistema de águas sub-

terrâneas do aquífero para melhor aplicação da interpretação da perfilagem geofísica convencional em estudos hidro-geológicos, pois os registros dos perfis dos aquíferos às vezes são diferentes das de reservatórios convencionais de petróleo e gás.

Lim (2005) sugeriu uma técnica que utiliza a Lógica Fuzzy e Redes Neurais Artificiais para determinar as propriedades dos reservatórios através da análise de perfis de poços. A análise da curva fuzzy, baseada em Lógica Fuzzy, foi utilizada para selecionar os perfis de poços relacionados com a porosidade e a permeabilidade do núcleo. As Redes Neurais foram utilizadas como um método de regressão não-linear para desenvolver uma transformação entre os perfis de poços selecionados e as medições do núcleo.

Braga (2005) desenvolveu uma metodologia para identificação e classificação de litofácies utilizando a teoria bayesiana de reconhecimento de padrões. A metodologia foi separada em cinco partes: 1) identificação das litofácies através do algoritmo de Maximização da Esperança, 2) geração de funções de distribuição de probabilidade, a partir dos dados de treinamento, considerando que os dados de treinamento são os perfis de poços interpretados com as litofácies definidas, 3) pressuposto o treinamento dos dados, geram-se as distribuições de cada classe, 4) emprega-se o algoritmo do k-vizinhos mais próximos com a finalidade de avaliar se uma amostra é determinada litofácies e 5) classifica-se as litofácies encontradas na metodologia realizando uma comparação com as litofácies identificadas pelo método manual, segundo o critério de máxima probabilidade.

Kumar (2006) realizou uma comparação entre a utilização de Redes Neurais Artificiais (RNA) e de Análises de Agrupamentos para identificação das fácies litológicas e deposição de perfis de poços. Nesse caso, foi desenvolvido um programa que realiza a comparação entre essas duas abordagens na tarefa de classificar eletrofácies. Na Análise de Agrupamentos foi primeiramente utilizado no modelo criado, a Análise de Componentes Principais e posteriormente o algoritmo de agrupamento K-Means. Os agrupamentos encontrados estão relacionados com as eletrofácies. Em RNA foi utilizado uma rede supervisionada para classificar as eletrofácies.

De Ros (2007) propôs uma metodologia para identificação de petrofácies. Uma vez que o conceito de petrofácies é útil para a caracterização de reservatórios e para a previsão da sua qualidade durante a exploração, tem-se a necessidade de determiná-

las. A determinação de petrofácies é iniciada a partir do reconhecimento de petrofácies preliminares através de uma descrição sistemática dos atributos listados em amostras recolhidas de rochas, seguido de reconhecimento dos atributos com um impacto maior sobre a porosidade e permeabilidade. As petrofácies preliminares são, então, confrontadas com parâmetros quantitativos petrofísicos e petrográficos, utilizando ferramentas de Redes Neurais. Os valores limites são definidos para os atributos texturais e composicionais que restringem as petrofácies significativas. As petrofácies definidas por esta metodologia são consistentes em termos de porosidade e permeabilidade. Dessa forma, elas podem ser utilizadas para as calibrações e para representações tridimensionais da qualidade dos reservatórios.

Xie (2008) apresentou uma estratégia para caracterizar eletrofácies, que combina a técnica de regressão não paramétrica (ECA) com o procedimento passo a passo. A ideia geral é a substituição da regressão linear pela regressão passo a passo com o algoritmo ECA. Foi utilizado o Critério de Informação de Akaike (CIA) para seleção de modelo e a eliminação quanto para trás quanto para frente no procedimento passo a passo. O processo pode ser descrito nas seguintes fases: 1) os dados de treinamento são colocados com o modelo completo (todos os perfis de poços) usando ECA, 2) o critério CIA é calculado, 3) são feitas as exclusões e adições de termos únicos, 4) é aplicada a regressão não-paramétrica (ACE), com todos os novos modelos, 5) o critério CIA é calculado para os novos modelos. Se houver um valor menor CIA, selecione esse modelo com menor valor CIA e repita o passo 2. Se nenhum modelo tem menor CIA do que o modelo inicial, deve-se parar o procedimento passo a passo e selecionar o modelo inicial como o modelo ideal.

Ponte (2010) utilizou Redes Neurais Artificiais para identificar litofácies, estimando colunas estratigráficas a partir de análises combinadas de perfis geofísicos e testemunhos de poços. As Redes Neurais foram utilizadas em trechos não testemunhados para constituir várias medições geoeletricas e fornecer como resposta qual a litofácies que melhor caracteriza os parâmetros geoeletricos medidos.

Cevolani *et al* (2011) desenvolveu um procedimento para identificação e classificação de petrofácies sedimentares. O processo era dividido em duas partes: 1) ela-

boração de um procedimento computacional para determinação do número de petrofácies e posterior classificação das amostras nas petrofácies determinadas, 2) visualização dos agrupamentos encontrados em duas dimensões. Para a primeira parte foi aplicado o algoritmo de agrupamento de partição em torno de medoides (PAM), a fim de identificar as petrofácies. Para a segunda parte foi utilizado a técnica de Análise de Componentes Principais para reduzir a dimensionalidade dos dados para duas dimensões. Para visualização dos dados foi empregado o algoritmo de Escalonamento Multidimensional (EM).

Raeesi (2012) utilizou Redes Neurais Artificiais (RNA) para interpretar dados sísmicos 3D com intuito de posteriormente identificar litofácies e determinar suas mudanças para exploração de reservatórios de hidrocarbonetos. RNA oferece um manuseio superior quando tem-se não-linearidade nos dados, como é o caso dos dados sísmicos, o que proporciona uma melhor análise em relação a outras técnicas de análise matemática. Foi aplicado uma análise multi-característica com base em métodos RNA em dados de perfis de poços para determinar a alteração de litofácies e a heterogeneidade em um dos campos de petróleo estruturais estratigráficos no Golfo Pérsico.

Sharma (2012) propôs uma abordagem para melhorar a estimativa da permeabilidade de um reservatório de hidrocarbonetos através da caracterização de eletrofácies. A abordagem foi dividida em duas partes. A primeira é classificar os dados de perfis de poços em tipos de eletrofácies. Esta classificação é realizada com base nas medidas das características que individualizam um tipo. Foi aplicado uma combinação dos métodos Análise de Componentes Principais (ACP), Análise de Agrupamentos baseado em Modelo (AAM) e Análise Discriminante, este último usado para identificar e classificar os tipos de eletrofácies. Na segunda parte é utilizado a técnica de regressão não paramétrica, a Expectativa Condicional Alternada (ECA), para prever a permeabilidade usando perfis de poços dentro de cada eletrofácies encontrada.

Chehrazi (2012) propôs uma abordagem sistemática para prever a permeabilidade de um reservatório. A porosidade e o tipo de poros foram identificados como os principais atributos influentes e as petrofácies como a melhor forma de estimar a permeabilidade nos poços não tubulares. O método de agrupamento Fuzzy C-Means foi aplicado para identificar as petrofácies e as relações correspondentes entre porosidade e permea-

bilidade para cada petrofácies. Para classificar as petrofácies encontradas foi utilizado o algoritmo de árvore de decisão, para posteriormente obter a relação permeabilidade-porosidade e assim realizar o cálculo da permeabilidade.

Lacentre (2013) elaborou um método para estimar a permeabilidade de perfis de poços e dados do núcleo. O método faz uso de modernas ferramentas matemática, que se mostraram eficazes em campos da ciência e da engenharia, como Análise de Componentes Principais, Redes Neurais, Análise de Agrupamentos.

3 Fundamentação Teórica

Nesse capítulo encontram-se os principais conceitos teóricos que foram necessários para o desenvolvimento do trabalho. Está dividido em duas seções: Contexto Geológico e Análise de Agrupamento. Na primeira seção estão descritos definições geológicas que são fundamentais para o entendimento de alguns conceitos e da importância desse trabalho. Na segunda seção estão descritos os métodos que foram utilizados no desenvolvimento dessa monografia, o que é importante para compreender a metodologia e os resultados.

3.1 Contexto Geológico

3.1.1 Formação de Bacias Sedimentares

As bacias sedimentares são feições geológicas desenvolvidas em áreas de depressões relativas ou absolutas, que acumulam espessas camadas ou estratificações formadas por rochas sedimentares. As rochas sedimentares são formadas através da desintegração e decomposição de rochas preexistentes (magmáticas, metamórficas ou sedimentares), devido a ação de intemperismo (processos mecânicos, químicos e biológicos que originam na transformação das rochas em sedimentos).

Para a gênese das rochas sedimentares siliciclásticas são necessárias quatro fases (Godoy, 2005): destruição das rochas preexistentes, transporte dos produtos resultantes dessa destruição, deposição desses produtos numa bacia de sedimentação, diagênese. Essas quatro fases são descritas abaixo:

- **Destruição das rochas preexistentes:** será realizada sob ação conjunta do intemperismo e da erosão;
- **Transporte dos produtos originados pelo intemperismo:** os produtos resultantes da fragmentação e decomposição intempérica das rochas, assim como os fragmentos de organismos mortos ou produtos de origem orgânica, serão conduzidos pelos agentes externos como a água, o vento, as geleiras e a gravidade.

- **Sedimentação:** os sedimentos que foram conduzidos mecanicamente passarão pelo processo de decantação e serão acumulados quando a velocidade do agente transportador estiver baixa, originando, de forma inicial, os depósitos sedimentares inconsolidados. Os compostos solúveis, transportados em solução, passarão pelo processo de precipitação química, como consequência direta de mudanças físico-químicas do meio.
- **Diagênese:** conjunto de processos físicos-químicos que atuam após a deposição dos sedimentos, transformando-os em agregados naturais, denominados rochas sedimentares. Os principais processos durante a diagênese são a compactação e a cimentação. A compactação é a união provocada pela pressão das camadas superiores, resultando uma redução dos espaços vazios nos sedimentos. A cimentação é a percolação de soluções nos interstícios dos fragmentos e deposição da substância que estava dissolvida, cimentando os sedimentos. Os principais constituintes cimentantes são: carbonatos, sílica, óxido de ferro e argila.

3.1.2 Ambientes de Sedimentação e Fácies Sedimentares

Ambientes de Sedimentação

Os ambientes de sedimentação podem ser definidos como partes da superfície terrestre com propriedades físicas, químicas e biológicas bem definidas e distintas das apresentadas pelas áreas vizinhas. Estas propriedades compreendem uma gama de variáveis que se interagem, determinando as propriedades dos distintos ambientes de sedimentação.

O estudo das sequências sedimentares é parte de uma pesquisa mais ampla na análise de uma bacia sedimentar. A identificação de ambientes de sedimentação é de grande interesse para pesquisadores e para empresas, uma vez que a exploração de recursos naturais associados às rochas sedimentares, tais como petróleo, carvão, calcário, fosfato, entre outros, ocorrem em ambientes sedimentares específicos. Maiores detalhes do processo de formação desses recursos podem ser encontrados em (Brazil, 2004) e (Petroleum Geoscience Technology, 2011).

Fácies Sedimentares

Gressly (1938) percebeu, enquanto trabalhava na região dos Alpes, que litologias e fósseis diferentes podem ocorrer na mesma época. A partir dessa observação, ele propôs o termo *fácies* para definir unidades de rochas caracterizadas por propriedades litológicas (composição, textura, estruturas sedimentares e cor) e/ou paleontológica (conteúdo e registro fossilífero) semelhantes.

Fácies pode ser definida como uma parte restrita em área de uma determinada unidade estratigráfica, que exibe características diferentes significantes das demais partes da unidade (Fávera, 2001).

3.1.3 Composição e Gênese dos Hidrocarbonetos

Os hidrocarbonetos são misturas orgânicas complexas que são encontradas na natureza, dentro de poros e fraturas de rochas, geralmente sedimentares. Estes se constituem de hidrogênio e carbono, basicamente, com porções que variam de nitrogênio, oxigênio e enxofre (Suguio, 1980).

Os hidrocarbonetos ocorrem nos estados líquido, sólido e gasoso, nas condições naturais. Os líquidos são definidos como petróleo bruto e os sólidos são conhecidos como asfalto, piche e gilsonita. Os gasosos são referidos como gás natural.

A migração de hidrocarbonetos ocorre a partir de uma rocha geradora em destino a rochas porosas e permeáveis. Estas últimas são rochas que possuem chances de armazenar petróleo.

Uma rocha geradora deve possuir matéria orgânica em quantidade e qualidade adequadas e para ser submetida ao estágio de evolução térmica necessário para degradação do querogênio. A espessura e extensão lateral são importantes, pois uma rocha com quantidade e qualidade da matéria orgânica apropriadas pode ser muito delgada para gerar quantidades comerciais de petróleo.

O querogênio é a parte insolúvel da matéria orgânica alterada em consequência das ações geológicas. O querogênio é gerado através dos lipídios, proteínas e carboidratos, dos seres vivos. em condições adequadas pode se transformar em petróleo, gás natural ou grafite (Petroleum Geoscience Technology, 2011).

3.2 Rocha Reservatório

Uma rocha é definida reservatório se a porosidade e permeabilidade são adequadas à acumulação de petróleo. A maioria das reservas descobertas encontra-se em arenitos e rochas carbonáticas, apesar de acumulações de petróleo também ocorrerem em folhelhos, conglomerados, ou mesmo rochas ígneas.

A porosidade é definida como a porcentagem de vazios de uma rocha. Na maior parte dos reservatórios a porosidade varia de 10% a 20%. A quantidade, tamanho, geometria e grau de conectividade dos poros controlam de forma direta a produtividade do reservatório. Calculada diretamente, em amostras de testemunho, ou de forma indireta, por meio de perfis elétricos. A porosidade de uma rocha pode ser classificada como incipiente (0 – 5%), pobre (5 – 10%), regular (10 – 15%), boa (15 – 20%), ou muito boa (> 20%) (Petroleum Geoscience Technology, 2011).

A permeabilidade é a capacidade da rocha de transmitir fluido, sendo expressa em Darcys (D) ou milidarcys (md). Supervisionada pela quantidade, geometria e grau de conectividade dos poros, a permeabilidade de uma rocha é calculada diretamente, em amostras de testemunho, e pode ser classificada como baixa ($< 1md$), regular ($1 - 10md$), boa ($10 - 100md$), muito boa ($100 - 1000md$) e excelente ($> 1000md$). A maior parte dos reservatórios possui permeabilidades de 5 a 500md (Petroleum Geoscience Technology, 2011).

3.3 Formação da Bacia do Paraná

A Bacia do Paraná é uma bacia intratônica de grande extensão situada na porção meridional da América do Sul. Desenvolveu-se sobre a crosta continental, e foi preenchida por rochas sedimentares e vulcânicas, as quais as idades variam entre o Meso-Ordoviciano ao Neo-Cretáceo (Milani *et al*, 2007). O registro estratigráfico da Bacia do Paraná compreende um pacote sedimentar-magmático com uma espessura que pode chegar a depocentros em torno de 8 mil metros.

O conjunto de rochas sedimentares e vulcânicas presentes na Bacia do Paraná demonstra a superposição de sedimentos depositados, de no mínimo três diferentes am-

bientes tectônicos, ocasionados pela evolução do Supercontinente Gondwana ao longo do tempo geológico (Zalán *et al.*, 1990).

Milani (1997) identificou no registro estratigráfico da Bacia do Paraná seis Superseqüências de 2ª ordem, na forma de pacotes rochosos materializando cada um deles intervalos temporais com algumas dezenas de milhões de anos de duração e limitados por superfícies de discordância de caráter regional, quais sejam: Rio Ivaí (Ordoviciano-Siluriano), Paraná (Devoniano), Gondwana I (Carbonífero-Eotriássico), Gondwana II (Meso a Triássico), Gondwana III (Jurássico-Eocretáceo) e Bauru (Cretáceo).

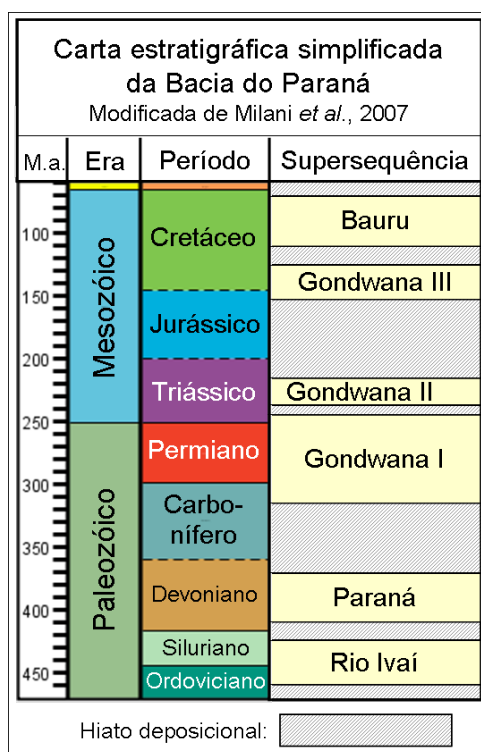


Figura 3.1: Eras Geológicas (Milani, 1997).

A Sequência de interesse desse trabalho é a Superseqüência Paraná. Os sedimentos Devoniano que ocorrem no Sul do Brasil constituem há muito tempo temas de estudos científicos. Através desse estudos, a Superseqüência do Paraná foi dividida em um pacote arenoso inferior, a Formação Furnas, e um pelítico, sobreposto denominada Formação Ponta Grossa.

A Formação Ponta Grossa foi representada, de forma inicial, no Estado do Paraná, onde é descrita por três membros: Jaguariaíva, Tibagi e São Domingos. O Membro Jaguariaíva, inferior, é apresentado por folhelhos com em torno 100 m de espessura, composto por lentes de arenito fino com estratificações retrabalhadas por ondas. O Membro

Tibagi, areno-siltico, equivale á parte média da Formação Ponta Grossa, revelando um contexto regressivo de progradação de sistemas com constante erosão e deposição provenientes da borda nordeste, onde é bastante expressivo o aporte dos termos arenosos. O Membro São Domingos, predominantemente pelítico, registra nova inundação em ampla escala (Milani *et al*, 2007).

Ao todo, o pacote Ponta Grossa documenta condições de mar alto, sendo as rochas dominantes de composição pelítica pontuada de maneira local por progradações arenosas, corresponde ao Membro Tibagi.

3.3.1 Qualidade dos Reservatórios de Exploração e Petrofácies

A permeabilidade e a porosidade de um reservatório refletem a textura da rocha. De modo geral, essas características são diretamente proporcionais ao grau de seleção e tamanho dos grãos e inversamente proporcional à esfericidade (Petroleum Geoscience Technology, 2011). Do mesmo modo, variações laterais e verticais da permeabilidade e a porosidade são fortemente controladas pelas características do ambiente deposicional.

A diagênese é outro fator que pode alterar as características permoporosas originais de uma rocha reservatório. Em arenitos, os processos diagenéticos mais importantes são a cimentação e a dissolução. A cimentação quando em poucas frações pode ser proveitoso, uma vez que evita a produção de grãos de areia junto com o óleo. Quando em grandes frações, a cimentação pode eliminar a porosidade original, reduzindo a permeabilidade a próximo de zero (Petroleum Geoscience Technology, 2011). A calcita, o quartzo e as argilas autigênicas (caolinita, ilita e montmorilonita) formam os cimentos mais comuns em arenitos. Em rochas carbonáticas as consequências da diagênese são de maior importância, visto que a calcita é menos estável do que o quartzo, uma vez que é composta principalmente por $CaCO_3$ que é muito reativo. Como resultado, a cimentação e a dissolução podem tanto piorar quanto melhorar a qualidade do reservatório (Petroleum Geoscience Technology, 2011).

Afim de avaliar a qualidade de um reservatório surge o conceito de petrofácies. As petrofácies podem ser definidas como uma técnica para o reconhecimento das heterogeneidades de um reservatório auxiliando na análise da evolução diagenética do mesmo.

Segundo De Ros (2007), petrofácies são caracterizadas pela combinação de estruturas específicas de deposição, texturas e composição primária, com processos diagenéticos dominantes. A combinação de aspectos texturais primárias e composicionais com processos e produtos diagenéticos específicos correspondem a gamas de valores definidas de porosidade e permeabilidade, bem como ao registro de atributos e assinaturas sísmicas.

O reconhecimento da petrofácies (De Ros, 2007), inicia com uma petrografia detalhada de amostras representativas da área estudada. Uma análise quantitativa através da contagem de 300 ou mais pontos é importante, mas não é sempre essencial para o reconhecimento petrofácies, pois, em alguns casos, os padrões principais podem ser identificados diretamente a partir de uma descrição qualitativa. As amostras são separadas em grupos, primeiro de acordo com estruturas sedimentares e textura. As amostras devem ser assim agrupadas considerando-se a superposição de atributos de deposição (estrutura, textura e tecido) com as principais categorias de composição primária, e com a distribuição dos processos diagenéticos mais influentes. Os atributos com maior impacto sobre a porosidade e permeabilidade são reconhecidos, e petrofácies preliminares são atribuídas. O agrupamento de amostras nas mesmas petrofácies assume que elas exibem comportamento petrofísico semelhante. As petrofácies preliminarmente definidas são confrontadas com parâmetros quantitativos petrofísicos e petrográficos, utilizando ferramentas de rede estatísticos ou neurais. Os valores limites são então definidos para os atributos, texturais e composicionais, influentes que restringem a significantes petrofácies.

3.4 Análise de Agrupamento

Uma questão básica que muitos pesquisadores de várias áreas enfrentam é como organizar dados observados em estruturas que agrupem subconjuntos semelhantes, isto é, como criar ou desenvolver taxionomias. O conceito de Análise de Agrupamento surgiu como uma forma de automatizar esse processo (Viali, 2000).

As técnicas de análise multivariada tornam possível realizar a avaliação de um conjunto de atributos, considerando as similaridades existentes. A Análise de Agrupamento é uma técnica multivariada que tem como intuito encontrar uma ou várias partições na base de dados, ou seja, grupos, segundo algum critério de classificação, de tal maneira

que exista homogeneidade dentro e heterogeneidade entre os grupos (Albuquerque, 2005).

O procedimento de agrupar pode ser dividido, de um modo geral, em duas fases: a primeira se refere à escolha de uma medida de dissimilaridade entre os objetos e a segunda, à aplicação de uma técnica de obtenção de grupos. Na literatura existem diversas medidas de similaridades/dissimilaridades propostas que têm sido bastante utilizadas em Análise de Agrupamento. A escolha entre essas medidas depende da preferência do pesquisador em relação a abordagem adotada pela medida.

Para adotar um método de agrupamento, o pesquisador deve decidir qual é o mais adequado, entre os vários existentes, de acordo com intuito do trabalho. O uso de diferentes técnicas podem resultar em distintas soluções, daí a importância na escolha do método de agrupamento.

3.4.1 Medidas de Dissimilaridade

Os métodos de agrupamentos, geralmente, admitem que as ligações fundamentais entre os objetos podem ser representados por uma matriz de dissimilaridade ou de similaridade entre os pares de objetos.

Cada elemento da matriz, d_{ij} , é um valor real que mostra o quão próximos os objetos i e j são. Existem métricas que calculam a similaridade, e outras que calculam a dissimilaridade, mas em fundamento elas são iguais.

Os elementos da matriz de dissimilaridade são funções $D : A \times A \Rightarrow \mathcal{R}$, onde A consiste no conjunto de objetos no qual estamos trabalhando. Estas funções possibilitam transformar a matriz de dados A em uma matriz de dissimilaridades D .

$$A = \begin{bmatrix} a_{11} & \dots & a_{1f} & \dots & a_{1p} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{if} & \dots & a_{ip} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \dots & a_{nf} & \dots & a_{np} \end{bmatrix} \quad D = \begin{bmatrix} 0 & & & & \\ d_{21} & 0 & & & \\ d_{31} & d_{32} & 0 & & \\ \vdots & \vdots & \vdots & & \\ d_{n1} & d_{n2} & \dots & \dots & 0 \end{bmatrix}$$

Para uma função ser considerada uma função de dissimilaridade, ela deve satisfazer os seguintes critérios (Kaufman, 1990):

- $d_{ij} \geq 0, \forall i, j \in A$
- $d_{ij} = d_{ji}, \forall i, j \in A$. Este critério afirma que a distância entre dois elementos não muda, não importando ponto o qual começou a ser calculada. Desta forma, a matriz D é mostrada sendo triangular inferior, uma vez que ela é simétrica e os valores não representados estão definidos implicitamente.
- $d_{jk} \leq d_{ij} + d_{ik}, \forall i, j, k \in A$. Este critério é conhecido como Desigualdade Triangular, que diz basicamente que a distância indo diretamente de i para j é menor que fazendo um desvio pelo objeto k .

Neste trabalho usou-se as seguintes dissimilaridades: Manhattan, Braycurtis, Euclideana, Canberra, Chebyshev (Souza *et al*, 2008).

As fórmulas para as distâncias, mostradas na Tabela 3.1, estão representadas para dois pontos $X = (x_1, x_2, \dots, x_n)$ e $Y = (y_1, y_2, \dots, y_n)$.

Tabela 3.1: Dissimilaridades utilizadas

Distância	Fórmula
Manhattan	$D(X, Y) = \sum_{i=1}^n x_i - y_i $
Brayacurtis	$D(X, Y) = \frac{\sum_{k=1}^n x_k - y_k }{\sum_{k=1}^n (x_k + y_k)}$
Euclideana	$D(X, Y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$
Canberra	$D(X, Y) = \sum_{k=1}^n \frac{ x_k - y_k }{ x_k + y_k }$
Chebyshev	$D(X, Y) = \max(x_k - y_k)$

3.4.2 Técnicas de Análise de Agrupamento

Os métodos de agrupamento podem ser divididos em três grandes categorias: particionamento, hierárquico e baseado em densidade.

Métodos de Particionamento

Os métodos de particionamento são métodos baseados na minimização de uma função de custo, onde os objetos são agrupados em um número k de agrupamentos escolhido antecipadamente. Cada objeto é agrupado de forma que a função de custo seja minimizada.

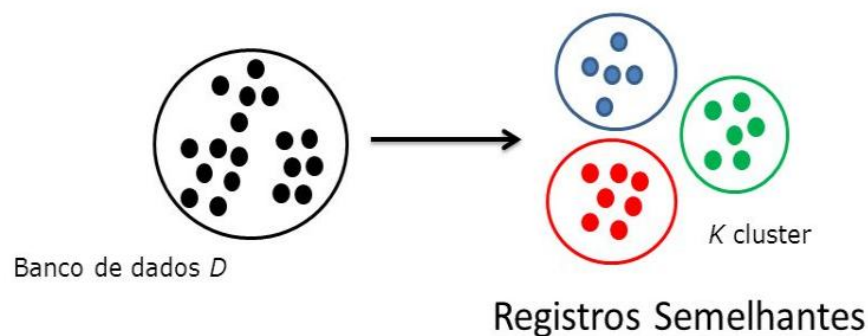


Figura 3.2: Métodos de particionamento (Boschi, 2013)

Um algoritmo muito utilizado é o K-Means, que é um método iterativo onde a soma das distâncias entre cada objeto ao centroide de cada agrupamento é minimizada, sobre todos os agrupamentos. O método desloca os objetos entre os agrupamentos até que a função objetivo fique inalterada ou até um número de iterações pré-estipulados seja alcançado. O resultado é k agrupamentos compactos e bem definidos. O algoritmo pode ser descrito nos seguintes passos

1. Inicializar as médias dos k agrupamentos.
2. Para cada objeto determinar o agrupamento mais próximo.
3. Calcular a média de cada agrupamento.
4. Se houver alteração na média dos agrupamentos, voltar ao passo 2.
5. O resultado obtido são os k agrupamentos.

Métodos Hierárquicos

Os métodos hierárquicos são técnicas onde os dados são particionados de forma sucessiva, produzindo uma representação hierárquica dos agrupamentos. Esse métodos podem apresentar duas abordagens: aglomerativa e divisiva.

Estes métodos não requer que seja definido um número de agrupamentos antecipadamente. A partir da análise do dendograma, que é um diagrama que mostra a hierarquia e a relação dos agrupamentos, pode-se inferir no número de agrupamentos.

Os métodos aglomerativos iniciam-se com cada objeto formando um agrupamento e de forma sucessiva os agrupamentos são unidos até formar um único contendo todos os objetos. No início do procedimento os agrupamentos são pequenos e os objetos de cada grupo possuem uma alta similaridade entre si. No final, obtêm-se poucos agrupamentos, sendo que cada um pode conter muitos objetos, menos similares entre si.

O processo pode ser descrito nos seguintes passos (Vale, 2005)

1. Cada agrupamento tem um único objeto.
2. Calcula-se a matriz de similaridades.
3. Um novo agrupamento é gerado pela junção dos agrupamentos com maior similaridade.
4. Os passos 2 e 3 são processados $(N-1)$ vezes, até que todos os objetos estejam em um único agrupamento.

Os métodos divisivos iniciam-se com um único agrupamento contendo todos os objetos e sucessivamente este vai se dividindo em agrupamentos menores até que os agrupamentos resultantes contenham um único objeto. A ideia é encontrar a partição que minimiza a matriz de similaridades.

O processo pode ser descrito nos seguintes passos (Vale, 2005)

1. Um único agrupamento contendo todos os objetos.
2. Calcula-se a matriz de similaridades entre todos os possíveis pares de objetos.
3. Um novo agrupamento é formado pela divisão dos pares de agrupamentos com menor grau de similaridade.

4. Os passos 2 e 3 são processados até que se tenha um agrupamento por objeto.

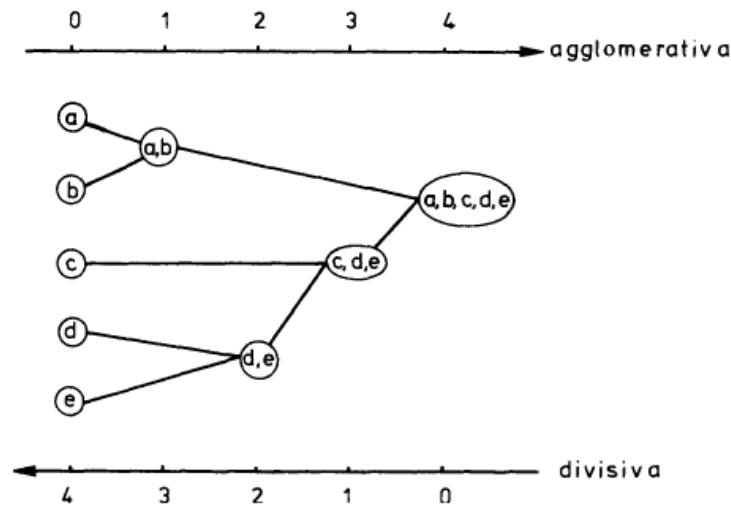


Figura 3.3: Métodos aglomerativo e divisivo (Kaufman, 1990).

Métodos Baseado em Densidade

Em consequência à falta de habilidade que os métodos de particionamento têm em encontrar agrupamentos com formas arbitrárias, ou seja, grupos que apresentam objetos distribuídos em um espaço irregular, foram desenvolvidos métodos baseados em densidade. Estes métodos consideram os agrupamentos como regiões densas de objetos separados no espaço por regiões de baixa densidade que, geralmente, representam algum tipo de ruído (Andrade, 2004). O método empregado neste trabalho, o DBSCAN (*Density-Based Spatial Clustering of Application with Noise*) utiliza essa abordagem e será descrito a seguir.

DBSCAN

O DBSCAN (Ester *et al*, 1996) é um algoritmo baseado em ligações de regiões de alta densidade que aplica a estratégia de assegurar que essas regiões serão os agrupamentos. Este método é capaz de encontrar agrupamentos com formas arbitrárias, inclusive com a presença de ruído nos dados, descrevendo os agrupamentos como a associação máxima de pontos densamente conectados.

Para realizar o agrupamento, o DBSCAN analisa a vizinhança próxima de cada objeto dentro da base de dados. Se a vizinhança de um objeto p possui ao menos o número mínimo de ocorrências $Minpts$, então p será denominado o centro de um novo agrupamento que se inicia em sua volta, chamado de vizinhança próxima. Um ponto p é denominado fronteira se não há pontos alcançáveis pela densidade a partir de p . Quando isso ocorre, visita-se o próximo ponto. O raio de definição da vizinhança - ϵ , bem como $MinPts$, são parâmetros pré-estipulados. A partir desse passo o algoritmo irá agrupar, de maneira iterativa, os objetos alcançados pela densidade em relação ao centro dos agrupamentos, ou seja, são agrupados os objetos mais próximos ao centro de forma que o agrupamento continue com densidade elevada. Este procedimento pode envolver a junção de agrupamentos com densidade baixa e finaliza quando nenhum novo ponto pode ser acrescentado em algum agrupamento. Os objetos não contido em algum agrupamento gerado serão denominados ruído. Abaixo um pseudo-código para ilustrar o algoritmo e a Figura 3.4 apresenta um fluxograma com a ideia do algoritmo.

Algorithm 2: Rotina computacional do DBSCAN

Escolha um ponto p arbitrariamente

while Todos os pontos alcançáveis pela densidade de p , Eps ,

$MinPts$ não forem recuperados **do**

if Se p é um ponto núcleo **then**

forma-se um grupo

end if

if Se p é um ponto fronteira **then**

visitar o próximo ponto

end if

end while

Continue o processo até que todos os pontos tenham sido processados.

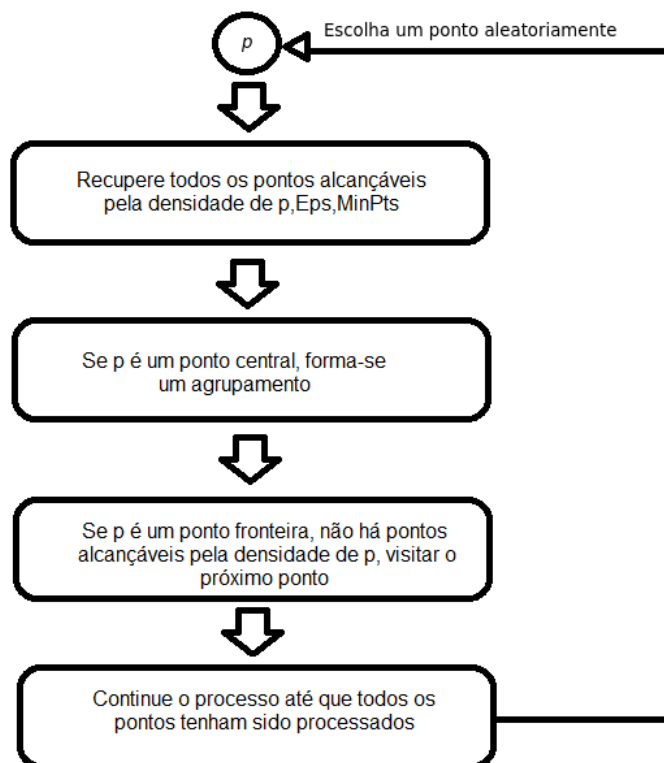


Figura 3.4: Fluxograma ilustrando o algoritmo DBSCAN

3.4.3 Critérios de Validação

Critérios de validação são medidas adotadas para avaliar a qualidade de um agrupamento. A escolha do critério depende do problema a ser tratado. Existem algumas publicações tratando desse assunto e comparando critérios, como em Vendramin (2010). Nesta monografia dois critérios foram testados, os quais serão descritos a seguir.

Silhueta

A análise de silhueta é uma técnica proposta por Rousseeuw (1987). Trata-se de um método geométrico baseado na compactação e separação de agrupamentos com o intuito de analisar a qualidade dos agrupamentos formados. O número ótimo de agrupamentos é definido pelo maior coeficiente de silhueta resultante da análise de silhueta. Para cada amostra i o valor s_i é definido pela seguinte fórmula:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (3.1)$$

Considerando que a amostra i pertença ao agrupamento A, a_i é descrito como a dissimilaridade média da amostra i em relação a todas as outras amostras do agrupamento A. Seja B um agrupamento diferente de A, b_i é a dissimilaridade média mínima da amostra i em relação a todas as outras amostras de B.

O coeficiente de silhueta de um conjunto de dados é dado pela média dos coeficientes individuais das amostras

$$\bar{s} = \frac{\sum_{i=1}^N s_i}{N} \quad (3.2)$$

onde N é o número de amostras do conjunto de dados. A métrica utilizada para o cálculo da dissimilaridade é a distância Euclideana. O valor de s_i varia de -1 a 1. Resultado próximo de 1, indica que os objetos estão bem agrupados.

Índice Calinski-Harabasz

Dado um conjunto $X = x(1), \dots, x(N)$ de N amostras de dados $x(j) \in \mathcal{R}^n$ e uma partição destes dados em k agrupamentos mutuamente disjuntos, o Índice Calinski-Harabasz (CH) avalia a qualidade da partição de dados pela seguinte equação:

$$CH = \frac{\text{traço}(B)}{\text{traço}(W)} \frac{N - k}{k - 1} \quad (3.3)$$

onde W e B são matrizes de dispersão intra-grupo e inter-grupo, respectivamente, sendo definidos com:

$$W = \sum_{i=1}^k \sum_{l=1}^{N_i} (x_i(l) - \bar{x}_i)(x_i(l) - \bar{x}_i)^T \quad (3.4)$$

$$B = \sum_{i=1}^k N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (3.5)$$

onde o N_i é o número de amostras atribuídas ao i -ésimo agrupamento, $x_i(l)$ é a l -ésima amostra atribuída ao agrupamento i , \bar{x}_i é o vetor n -dimensional da média das amostras dentro do agrupamento (centróide do agrupamento) e \bar{x} é o vetor n -dimensional das médias das amostras globais (centróide dos dados). Tal como, a soma das matrizes intra-grupo e inter-grupo resulta na matriz de dispersão da base de dados, isto é, $T = W + B$, onde $T = \sum_{l=1}^N (x(l) - \bar{x})(x(l) - \bar{x})^T$. O traço da matriz W é a soma das variâncias intra-grupo (seus elementos da diagonal). De forma análoga, o traço da matriz B é a soma das variâncias inter-grupos. Como consequência, agrupamentos compactos e separados são esperados para pequenos valores para o $\text{traço}(W)$ e grandes para o $\text{traço}(B)$. O termo de normalização $\frac{N - k}{k - 1}$ impede que essa proporção aumente diretamente com o número de agrupamentos, tornando CH um critério de maximização com relação a k (Vendramin, 2010).

3.4.4 Métricas de Avaliação Externa para Agrupamentos

Homogeneidade

A homogeneidade refere-se a amostras pertencentes a um mesmo agrupamento, que devem ser tão similares quanto possível. Essa medida varia de 0 a 1, sendo que quanto mais

próximo de 1, maior a similaridade existentes nas amostras de um agrupamento.

Para definir a homogeneidade (Rosenberg, 2007), assume-se um conjunto de dados com N objetos e duas partições do mesmo: um conjunto de classes $C = c_i | i = 1, \dots, n$ e um conjunto de agrupamentos $K = k_j | j = 1, \dots, m$. Seja A a tabela de contingência produzida pelo algoritmo de agrupamento, representando sua solução. Tal que $A = a_{ij}$, onde a_{ij} é o número de objetos que são membros da classe c_i e elementos do agrupamento k_j . Para satisfazer a homogeneidade, um agrupamento deve atribuir apenas os objetos que são membros de uma classe única para um único agrupamento ou seja, $H(C|K) = 0$. Afim de normalizar esse valor, faz-se a divisão por $H(C)$. Então $H(C|K) = 1$ e igual a $H(C)$ quando nenhum membros de uma classe c_i é elemento do cluster k_j . A homogeneidade é definida pela fórmula a seguir

$$h = \begin{cases} 1, & \text{se } H(C|K) = 0 \\ 1 - \frac{H(C|K)}{H(C)}, & \text{caso contrário} \end{cases}$$

onde $H(C|K) = - \sum_{k=1}^{|K|} \sum_{c=1}^{|C|} \frac{a_{ck}}{N} \log \frac{a_{ck}}{\sum_{c=1}^{|C|} a_{ck}}$ e

$$H(C) = - \sum_{c=1}^{|C|} \frac{\sum_{k=1}^{|K|} a_{ck}}{n} \log \frac{\sum_{k=1}^{|K|} a_{ck}}{n}.$$

Integralidade

Com intuito de satisfazer os critérios de integralidade (Rosenberg, 2007), um agrupamento deve atribuir todas as amostras pertencentes a uma única classe para um único agrupamento. Para avaliar a integralidade, examina-se a distribuição de atribuições de agrupamentos dentro de cada classe. Em uma solução de agrupamento perfeitamente completo, cada uma destas distribuições será completamente distorcida a um único agrupamento. Pode-se avaliar este grau de inclinação através do cálculo da entropia condicional da distribuição de agrupamento, dada a classe das amostras que o compõem, $H(K|C)$. No caso perfeitamente completo $H(K|C) = 0$. No entanto, na pior das hipóteses, cada classe é representada por cada agrupamento com uma distribuição igual à distribuição dos tamanhos dos agrupamentos, $H(K|C)$ é máximo e igual a $H(K)$. No caso degenerado em que $H(K) = 0$, quando não é um único agrupamento, que define a integralidade ser 1. Portanto, define-se a integralidade como

$$c = \begin{cases} 1, & \text{se } H(K|C) = 0 \\ 1 - \frac{H(K|C)}{H(K)}, & \text{caso contrário} \end{cases}$$

$$\text{onde } H(K|C) = - \sum_{k=1}^{|C|} \sum_{c=1}^{|K|} \frac{a_{ck}}{N} \log \frac{a_{ck}}{\sum_{c=1}^{|K|} a_{ck}} \text{ e}$$

$$H(K) = - \sum_{k=1}^{|K|} \frac{\sum_{c=1}^{|C|} a_{ck}}{n} \log \frac{\sum_{c=1}^{|C|} a_{ck}}{n}.$$

3.5 Análise de Componentes Principais

A Análise de Componentes Principais (ACP) é um método que tem como propósito, a análise dos dados usados objetivando sua redução, eliminação de sobreposições e a escolha dos modos mais representativos de dados através de combinações lineares das variáveis originais. Esta técnica é uma forma de identificar relação entre atributos retirados de dados. É muito útil quando os vetores de atributos possuem muitas dimensões, uma vez que é impossível uma representação gráfica.

3.5.1 Processo para uma Análise de Componentes Principais

O procedimento descrito a seguir foi baseado em Johnson (2007) e em Araujo (2009). A Análise de Componentes Principais tem como objetivo tomar p variáveis X_1, X_2, \dots, X_p e encontrar combinações lineares destas para gerar índices Z_1, Z_2, \dots, Z_p não correlacionados na sua ordem de importância, que represente a variação nos dados. Ser não correlacionados indica que os índices estão medindo dimensões diferentes dos dados. Sendo a ordem de tal forma que $Var(Z_1) \geq Var(Z_2) \geq \dots \geq Var(Z_p)$, onde $Var(Z_i)$ indica a variância de Z_i . Os índices Z são denominados como as componentes principais.

O procedimento se inicia com uma base de dados com p variáveis e n amostras. A combinação linear das variáveis originais X_1, X_2, \dots, X_p denomina a primeira componente principal.

$$Z_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

Dessa forma, a primeira componente principal é escolhida de maneira que $Var(Z_1)$

seja a maior possível sujeito a restrição

$$a_{11}^2 + a_{12}^2 + \dots + a_{1p}^2 = 1$$

A segunda componente principal é escolhida de forma análoga, acrescida da condição que a covariância entre Z_1 e Z_2 seja zero.

$$Z_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

sujeito a

$$\begin{cases} a_{21}^2 + a_{22}^2 + \dots + a_{2p}^2 = 1 \\ Cov(Z_1, Z_2) = 0 \end{cases}$$

As demais componentes principais são definidas do mesmo modo. Para n variáveis originais, podem existir no máximo p componentes principais.

Um dos procedimentos que faz parte é obter autovalores da matriz de covariância amostral. A matriz de covariância tem a seguinte forma

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & & \vdots \\ c_{p1} & c_{p2} & \dots & c_{pp} \end{bmatrix}$$

onde o elemento da diagonal, c_{ii} , é a variância de X_i e c_{ij} , elemento que não faz parte da diagonal, é a covariância entre as variáveis originais $X_i X_j$.

Os autovalores da matriz C são as variâncias das componentes principais. Existindo p autovalores, sendo que estes podem ter o valor zero e não podem ser negativos, uma vez que se trata de matriz de covariância. Parte do pressuposto que os autovalores estão ordenados, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, portanto λ_i equivale a i -ésima componente principal

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$$

Então, $Var(Z_i) = \lambda_i$ e as constantes $a_{i1}, a_{i2}, \dots, a_{ip}$ são elementos do equivalente

autovetor, de maneira que a restrição seja satisfeita

$$a_{i1}^2 + a_{i2}^2 + \dots + a_{ip}^2 = 1$$

Os autovalores possuem uma particularidade importante, o fato que a soma deles é igual ao traço da matriz de covariância C .

$$\lambda_1 + \lambda_2 + \dots + \lambda_p = c_{11} + c_{22} + \dots + c_{pp}$$

Isto deve-se ao fato de c_{ii} ser a variância de X_i e λ_i a variância de Z_i , implicando que a soma das variâncias das componentes principais é igual a soma das variâncias das variáveis originais. Pode-se, então, dizer que as componentes principais contam com toda a variação nos dados originais.

Com intuito de prevenir que variáveis tenham uma influência errônea nas componentes principais, é usual realizar um pré-processamento nos dados, para que as variáveis X_1, X_2, \dots, X_p tenham médias zero e variâncias um no começo da aplicação do procedimento. Isso é realizado afim de garantir que as primeiras componentes principais armazenem maior porcentagem de informação das variáveis originais. A matriz C adota a seguinte forma

$$C = \begin{bmatrix} 1 & c_{12} & \dots & c_{1p} \\ c_{21} & 1 & \dots & c_{2p} \\ \vdots & \vdots & & \vdots \\ c_{p1} & c_{p2} & \dots & 1 \end{bmatrix}$$

onde c_{ij} é a correlação entre as variáveis X_i e X_j . Portanto, a Análise de Componentes Principais é realizada sobre a matriz de correlação. A soma dos autovalores, ou seja, dos elementos da diagonal, é igual a p .

Com base nessa descrição do procedimento, as etapas da Análise de Componentes Principais podem ser relatadas:

1. Realização do Pré-processamento, para que as variáveis X_1, X_2, \dots, X_p tenham médias zeros e variâncias um.

2. Cálculo da matriz de covariâncias C ou matriz de correlação, caso a Etapa 1 tenha sido realizada.
3. Cálculo para encontrar os autovalores $\lambda_1, \lambda_2, \dots, \lambda_p$ e os respectivos autovetores a_1, a_2, \dots, a_p . Os coeficientes da i -ésima componente principal são os termos de a_i , e λ_i sua variância.
4. Exclusão das componentes que expressam uma pequena proporção nos dados.

4 Metodologia

Nesse capítulo encontra-se a descrição da metodologia desenvolvida nessa monografia. Está dividido em três seções: Caracterização da Base de Dados, Procedimento de Análise Intrapoço e Procedimento de Análise Interpoço. A identificação de petrofácies permite a inferência de quão heterogêneo é o reservatório. Determinar e mapear heterogeneidades de reservatórios de hidrocarbonetos é importante para a caracterização destes, uma vez que possibilita obter uma relação entre a permeabilidade e porosidade de uma rocha, o que é essencial para determinar se uma rocha é um bom reservatório.

A metodologia utilizada neste trabalho foi separada em duas partes:

1. Realização de uma Análise Intrapoço, com o intuito de verificar como os métodos procedem diante de cada poço, de forma individual.
2. Desenvolvimento de uma Análise Interpoço, com o objetivo de analisar como ocorre a divisão das lâminas em grupos quando as informações de todos os poços estão armazenadas na mesma base de dados.

A linguagem de programação utilizada nesse trabalho foi Python. Os seguintes fatores contribuíram para a escolha: o fato de conter recursos que possibilitam a aplicação de métodos nas informações e a representação gráfica da base de dados.

Esta linguagem conta com um módulo, *Scikit-learn* (Pedregosa *et al*, 2011) que integra uma gama de algoritmos de aprendizado de máquina para problemas supervisionados e não-supervisionados. A importância é colocada na facilidade de uso, desempenho e documentação.

4.1 Caracterização da Base de Dados

Os dados analisados são lâminas petrográficas obtidas a partir de amostras coletadas em furos de sondagem pertencentes ao Membro Tibagi, Devoniano da Bacia do Paraná

(Projeto Paleosul). A Bacia do Paraná é uma bacia que se localiza em uma região tectonicamente estável, que possui uma área de 1400000 km² com partes localizadas no Brasil, Argentina, Paraguai e no norte do Uruguai, como mostrado na Figura 4.1.

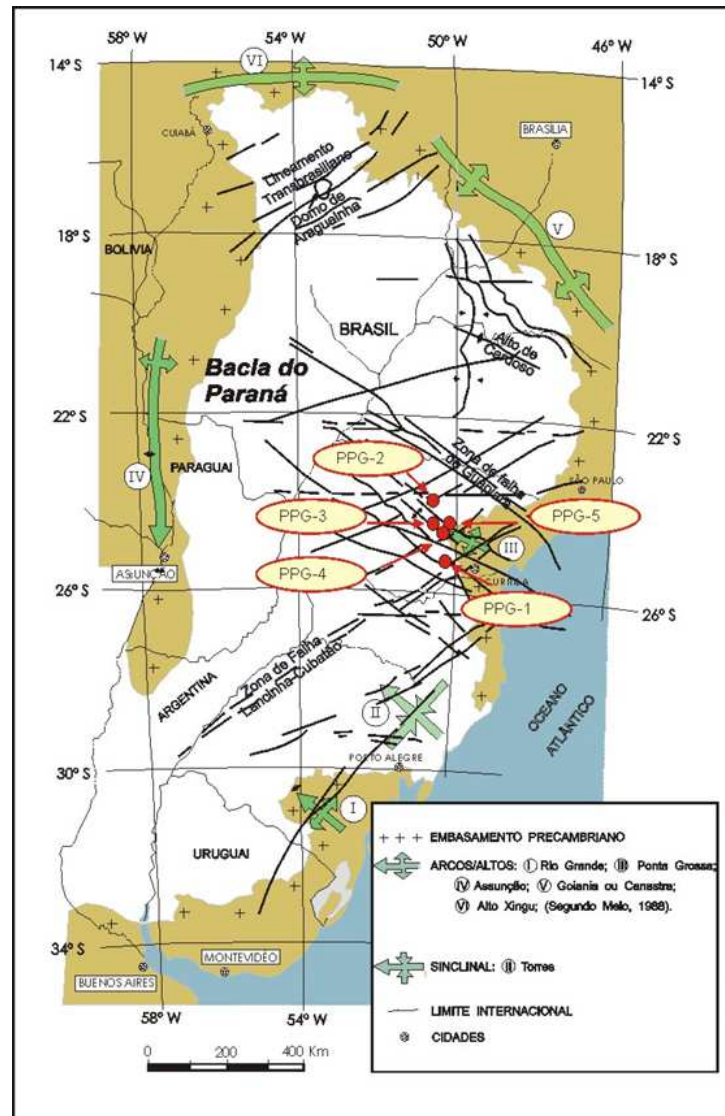


Figura 4.1: Localização da Bacia do Paraná e a posição dos poços (pontos na cor vermelha) (modificado de (Oliveira, 2009)).

A base de dados apresenta informações de 5 poços e 44 lâminas ao todo. A análise quantitativa das lâminas foi feita através da contagem (em microscópio petrográfico de luz transmitida) de 300 pontos em cada lâmina petrográfica com espaçamento de 0.3 mm. Detalhes do processo de obtenção dos dados podem ser encontrados em (Oliveira, 2009). Nas amostras provenientes da Bacia do Paraná, foram contabilizadas, para cada lâmina, as porcentagens de 22 constituintes (Tabela 4.1):

Tabela 4.1: Petrográficos analisados

Bioclasto	Quartzo
Cresc. Sec. Qtz	Feldspato
Caolinita	Muscovita
Ilita/Smectita	Opaco
Pirita	Turmalina
Siderita	Zircão
Cim. Carbonático	Rutilo
Cim. Silicoso	Glauconita
Cim. Ferruginoso	Clorita
Por. Intergranular	Pseudo Matriz
Por. Intragranular	Litoclasto

Neste trabalho, cada lâmina preparada a partir de amostras coletadas no furo de sondagem é uma amostra com 22 características, que correspondem à quantidade de constituintes contabilizados.

De acordo com os resultados obtidos com a análise quantitativa dos dados petrográficos foi possível individualizar os constituintes (pela classificação convencional realizada pelo geólogo/petrólogo) em 6 diferentes petrofácies, de acordo com (Oliveira, 2009): PT-1, PT-2, PT-3, PT-4, I-1 e I-2.

A Tabela 4.2 mostra um exemplo de especificação dos dados. Na base de dados encontra-se na linha 1 a profundidade (metros) de cada lâmina, na linha 2 as lâminas petrográficas, na linha 3 as petrofácies identificadas no procedimento manual, na primeira coluna, após as petrofácies, os constituintes das rochas que foram visualizados nas análises petrográficas, no restante da tabela pode ser visto as porcentagens de cada constituinte em cada lâmina.

Para o Poço 1 (PPG1) foram averiguadas 12 lâminas classificadas de PPG1-1 até PPG1-12. Segundo a análise manual, a lâmina PPG1-1 foi identificada como pertencente à petrofácies denominada I-2; as amostras PPG1-2 a PPG1-8 à petrofácies I-1; as demais foram agrupadas à petrofácies chamada PT-1. No Poço 2 (PPG2) analisou-se 11 lâminas, PPG2-1 a PPG2-11, onde a lâmina PPG2-1 foi classificada na petrofácies PT-2 e o restante na petrofácies PT-1. Para Poço 3 (PPG3) análise foi realizada em 12 lâminas, PPG3-1 a

PPG3-12. A lâmina PPG3-11 foi classificada como pertencente à PT-3 e as 11 amostras restantes foram agrupadas à PT-1. O Poço 4 (PPG4) teve 5 lâminas analisadas, PPG4-1 a PPG4-5, onde todas foram classificadas como referente à PT-4. Foram examinadas 4 lâminas do Poço 5 (PPG5), PPG5-1 a PPG5-4, no qual a lâmina PPG5-3 foi identificada como relativa à PT-3 e o restante à PT-1.

Tabela 4.2: Amostras retiradas do PPG5

Profundidade(m)	21, 5	46, 4	47, 1	55, 3
Lâmina delgada	PPG5-1	PPG5-2	PPG5-3	PPG5-4
Petrofácies	PT-1	PT-1	PT-3	PT-1
Quartzo	40,0	34,3	38,3	41,6
Feldspato	6,0	5,6	1,3	1,7
Muscovita	19,0	23,6	7,3	18,0
Opaco	0,6	1,3	1,0	3,4
Turmalina	0,0	0,0	0,0	0,0
Zircão	0,3	1,0	0,0	0,4
Rutilo	0,0	0,0	0,3	0,0
Glauconita	0,0	0,0	0,0	0,3
Clorita	0,0	0,0	0,0	0,7
Pseudo Matriz	1,3	7,0	2,0	12,7
Litoclasto	0,0	0,0	0,0	0,0
Bioclasto	0,0	0,0	0,0	0,0
Cresc. Sec. Qtz	3,4	3,0	7,6	1,0
Caolinita	10,0	11,3	8,0	1,0
Ilita/Smectita	10,4	12,3	6,3	13,4
Pirita	0,0	0,0	0,0	0,0
Siderita	0,0	0,0	0,0	0,0
Cim. Carbonático	0,0	0,0	27,6	1,4
Cim. Silicoso	0,0	0,0	0,0	0,7
Cim. Ferruginoso	0,0	0,0	0,0	1,4
Por. Intergranular	9,0	0,6	0,3	1,0
Por. Intragranular	0,0	0,0	0,0	0,0

4.2 Procedimento de Análise Intrapoço

A Análise Intrapoço tem o propósito de identificar as petrofácies dentro de um determinado furo de sondagem e a forma como as amostras foram separadas nessas petrofácies (Cevolani *et al*, 2011). Esse procedimento é composto por três técnicas:

- Método de Agrupamento (DBSCAN)
- Análise da Silhueta
- Análise de Componentes Principais.

A escolha do DBSCAN se fez pelo fato dele conseguir agrupar amostras em agrupamentos com formas arbitrárias e não necessitar que o número de agrupamentos seja previamente estabelecido. Outros métodos de agrupamentos foram testados e os resultados podem ser vistos em (Saporetti *et al*, 2014).

O uso do DBSCAN permitiu agrupar amostras presentes em um poço, de tal forma que aquelas que fazem parte do mesmo agrupamento possuem características em comum, ou seja, são similares entre si e dissimilares a amostras pertencentes a outros agrupamentos. O quão similares são as amostras é definido a partir de uma medida de dissimilaridade. Os agrupamentos formados são vistos como possíveis petrofácies dos poços da base de dados. Vale ressaltar que nesse procedimento os poços são analisados de forma individual.

Foi realizado uma busca exaustiva em um conjunto de possíveis parâmetros, (*Min_Pts*), Eps e a métrica para calcular a distância, com intuito de maximizar o coeficiente de silhueta (SC) e o índice de Calinski-Harabasz (CH). O número mínimo de pontos (*Min_Pts*) foi variado em um intervalo de 1 a 4, uma vez que um dos poços analisados possuem somente 4 lâminas. O raio máximo (Eps) foi variado de 0.1 a 1 com passos de 0.018. As seguintes distâncias foram testadas Braycurtis, Euclideana, Canberra, Chebyshev, Manhattan (ver descrição na Subseção 3.4.1 deste trabalho).

Os dados foram projetados nas duas primeiras componentes principais que armazenam maior informação dos dados, permitindo que os resultados possam ser visualizados em duas dimensões devido o uso de Análise de Componentes Principais.

4.3 Procedimento de Análise Interpoço

A Análise Intrapoço possui limitações, uma vez que ela permite determinar petrofácies distintas em um único poço. Dessa forma, não é possível verificar se a mesma petrofácies ocorre em poços distintos.

A Análise Interpoço tem o propósito de verificar como as petrofácies se comportam quando as amostras de todos os poços estão reunidas em uma única base de dados. O interesse parte da hipótese que petrofácies similares que se repetem em cada poço apresentam médias dos percentuais dos constituintes semelhantes entre si.

A Tabela 4.3 apresenta um exemplo dos percentuais resultantes dos constituintes depois da aplicação da etapa anterior. No Apêndice B encontram-se as tabelas com os resultados de todos os poços. Nessa análise é realizado o cálculo da média dos constituintes de cada agrupamento encontrado na etapa anterior. Em cada poço, cada agrupamento é representado por seu elemento médio (centróide), e estes são exibidos através do uso da Análise de Componentes Principais.

Tabela 4.3: Resultado DBSCAN para o PPG5

Profundidade(m)	21, 5	46, 4	47, 1	55, 3
Lâmina delgada	PPG5-1	PPG5-2	PPG5-3	PPG5-4
Petrofácies	PT-1	PT-1	PT-3	PT-1
DBSCAN (Agrupamentos)	1	1	2	1
Quartzo	40,0	34,3	38,3	41,6
Feldspato	6,0	5,6	1,3	1,7
Muscovita	19,0	23,6	7,3	18,0
Opaco	0,6	1,3	1,0	3,4
Turmalina	0,0	0,0	0,0	0,0
Zircão	0,3	1,0	0,0	0,4
Rutilo	0,0	0,0	0,3	0,0
Glauconita	0,0	0,0	0,0	0,3
Clorita	0,0	0,0	0,0	0,7
Pseudo Matriz	1,3	7,0	2,0	12,7
Litoclasto	0,0	0,0	0,0	0,0
Bioclasto	0,0	0,0	0,0	0,0
Cresc. Sec. Qtz	3,4	3,0	7,6	1,0
Caolinita	10,0	11,3	8,0	1,0
Ilita/Smectita	10,4	12,3	6,3	13,4
Pirita	0,0	0,0	0,0	0,0
Siderita	0,0	0,0	0,0	0,0
Cim. Carbonático	0,0	0,0	27,6	1,4
Cim. Silicoso	0,0	0,0	0,0	0,7
Cim. Ferruginoso	0,0	0,0	0,0	1,4
Por. Intergranular	9,0	0,6	0,3	1,0
Por. Intragranular	0,0	0,0	0,0	0,0

A Tabela 4.4 apresenta a média dos constituintes para os agrupamentos encontrados no PPG5, que foram apresentados na Tabela 4.3. As Tabelas com a média dos constituintes para os outros poços estão no Apêndice C. Pode-se observar que para os constituintes Muscovita, Pseudo Matriz e Cimento Carbonático houve uma diferença considerável nos valores das médias encontradas para as petrofácies PT-1 e PT-3. Estas diferenças podem levar à diferenciação das petrofácies.

Tabela 4.4: Média dos Percentuais dos Constituintes para o PPG5

Lâminas	PPG5-1,PPG5-2,PPG5-4	PPG5-3
Petrofácies	PT-1	PT-3
DBSCAN (Agrupamentos)	1	2
Quartzo	38.633333	38.3
Feldspato	4.433333	1.3
Muscovita	20.200000	7.3
Opaco	1.766667	1.0
Turmalina	0.000000	0.0
Zircão	0.566667	0.0
Rutilo	0.000000	0.3
Glauconita	0.100000	0.0
Clorita	0.233333	0.0
Pseudo Matriz	7.000000	2.0
Litoclasto	0.000000	0.0
Bioclasto	0.000000	0.0
Cresc. Sec. Qtz	2.466667	7.6
Caolinita	7.433333	8.0
Ilita/Smectita	12.033333	6.3
Pirita	0.000000	0.0
Siderita	0.000000	0.0
Cim. Carbonático	0.466667	27.6
Cim. Silicoso	0.233333	0.0
Cim. Ferruginoso	0.466667	0.0
Por. Intergranular	3.533333	0.3
Por. Intragranular	0.000000	0.0

5 Experimentos Computacionais

Neste capítulo serão expostos os experimentos realizados e os resultados encontrados. Inicialmente foi desenvolvida a Análise Intrapoço a fim de verificar como a metodologia proposta iria se comportar na base de dados em questão. Posteriormente foi elaborada uma Análise Interpoço com intuito de constatar se é possível obter informações relevantes do comportamento das petrofácies quando os dados de todos os poços estão em uma única base. Os procedimentos estão separados em duas seções, Análise Intrapoço e Análise Interpoço. A primeira seção subdivide-se em duas subseções, onde estão divididos os resultados encontrados utilizando o Coeficiente de Silhueta e o Índice de Calinski-Harabasz. Na segunda seção encontram-se os testes realizados e os resultados obtidos com a metodologia proposta para a Análise Interpoço.

5.1 Análise Intrapoço

5.1.1 Coeficiente de Silhueta

A Tabela 5.1 mostra os parâmetros, que dentre os testados, originaram um valor máximo do coeficiente de silhueta. A homogeneidade, que pode ser observada na Tabela 5.1, mostra que as amostras de um mesmo agrupamento são bem similares entre si. Para os poços PPG2 a PPG5, o valor alcançado foi 1, que é o valor máximo da homogeneidade. Já no poço PPG1, a homogeneidade encontrada foi 0.765121, o que mostra que as amostras possuem alguma similaridade entre si. A integralidade, por sua vez, mostrada na Tabela 5.1, define se as amostras pertencente ao mesmo agrupamento foram atribuídas a mesma petrofácies. Para os poços PPG1, PPG2, PPG3 e PPG5 o valor encontrado foi 1. Já no poço PPG4, a integralidade encontrada foi $-1.649634e-16$, que indica que há amostras pertencentes a mesma petrofácies que foram atribuídas a diferentes agrupamentos.

Para realizar a comparação dos resultados obtidos pelo método computacional com a classificação manual foi contabilizado o número de lâminas que foram agrupadas

Tabela 5.1: Parâmetros que maximizaram a silhueta e os valores encontrados para homogeneidade e para integralidade

Poços	Distância	<i>Min_Pts</i>	Eps	SC	Homogeneidade	Integralidade
PPG1	Canberra	3	5.136	0.455913	0.765121	1.000000e+00
PPG2	Chebyshev	1	10.764	0.585717	1.000000	1.000000e+00
PPG3	Canberra	1	6.764	0.656816	1.000000	1.000000e+00
PPG4	Chebyshev	2	9.943	0.426477	1.000000	-1.649634e-16
PPG5	Chebyshev	1	10.895	0.395879	1.000000	1.000000e+00

corretamente e posteriormente dividido pelo número total de lâminas. Esse procedimento foi realizado em todos os poços de forma individual.

As Figuras 5.1, 5.2 e 5.3 mostram as petrofácies identificadas e a forma como as amostras foram divididas nos agrupamentos originados pelo DBSCAN.

A análise para o PPG1 assinala que as duas componentes principais descrevem 76,23% da variabilidade dos dados, sendo 61,91% descrita pela primeira componente e 14,33% pela segunda componente. Para o PPG1, o método identificou dois agrupamentos, coincidindo com o método manual na classificação da petrofácies I-1. Em relação a PT-1 e I-2, o algoritmo agrupou as amostras pertencentes a essas petrofácies em um único agrupamento. Na Tabela 5.1 pode-se observar que nesse poço, as amostras não possuem a homogeneidade alta ao se comparar aos outros poços. Agrupou 92% das lâminas de forma correta.

No poço PPG2, as componentes principais descrevem 88,15% da variabilidade dos dados, 62,34% pela primeira componente e 25,81% pela segunda componente. No poço PPG3, as componentes principais representam 88,15% da variabilidade dos dados, 75,43% pela primeira componente e 13,42% pela segunda componente e PPG5, as componentes principais descrevem 92,66% da variabilidade dos dados, 75,72% pela primeira componente e 16,94% pela segunda componente. Nesses três últimos poços citados, os resultados coincidiram com a classificação manual ao identificar corretamente as petrofácies associadas.

A Figura 5.2 mostra os resultados para o PPG4. As componentes principais reproduzem 88,75% da variabilidade dos dados, 75,55% pela primeira componente e 13,20% pela segunda componente. O método encontrou dois agrupamentos. Pelo método conven-

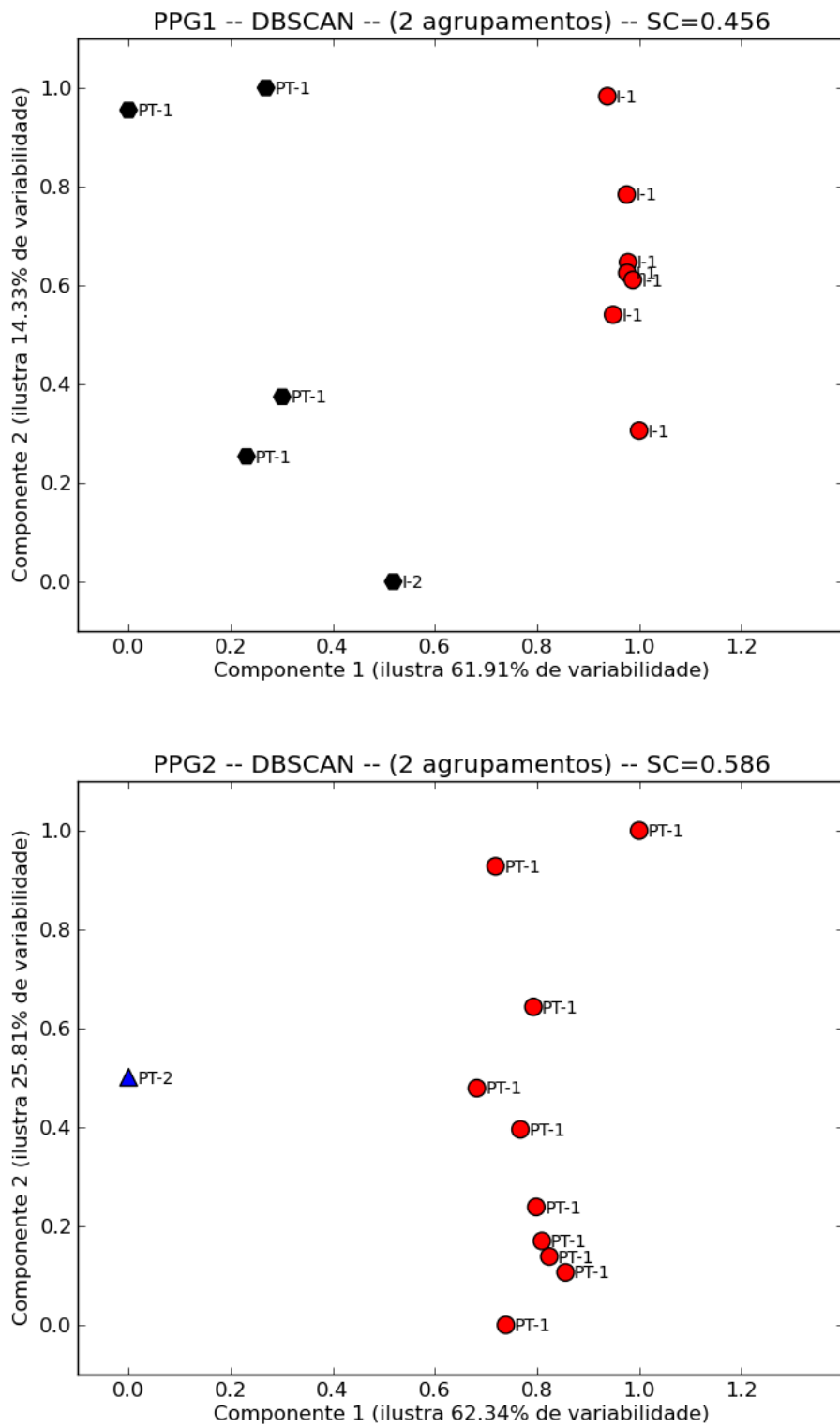


Figura 5.1: Visualização da Análise Intraçoço para os furos PPG1, PPG2 (Silhueta).

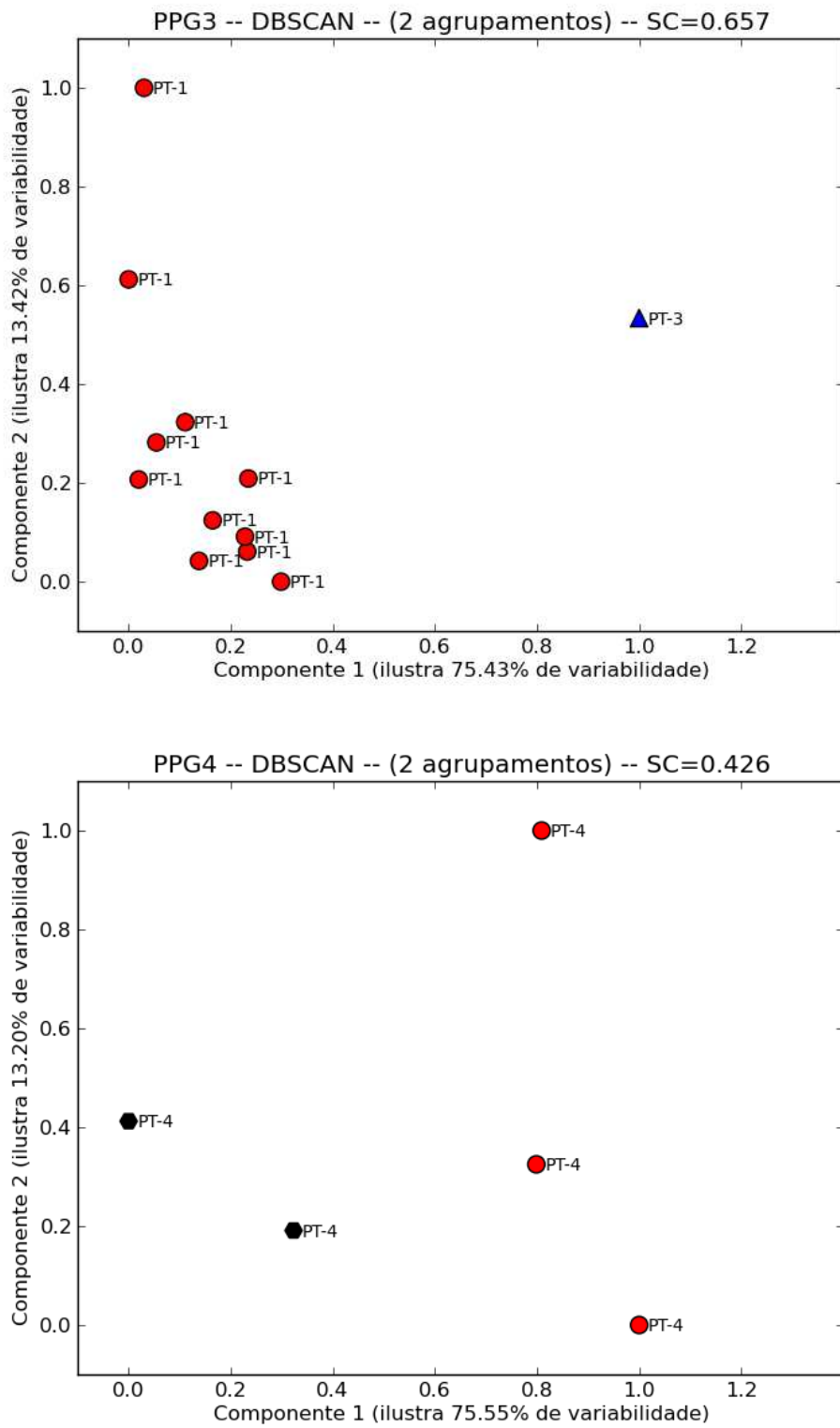


Figura 5.2: Visualização da Análise Intraçoço para os furos PPG3 e PPG4 (Silhueta).

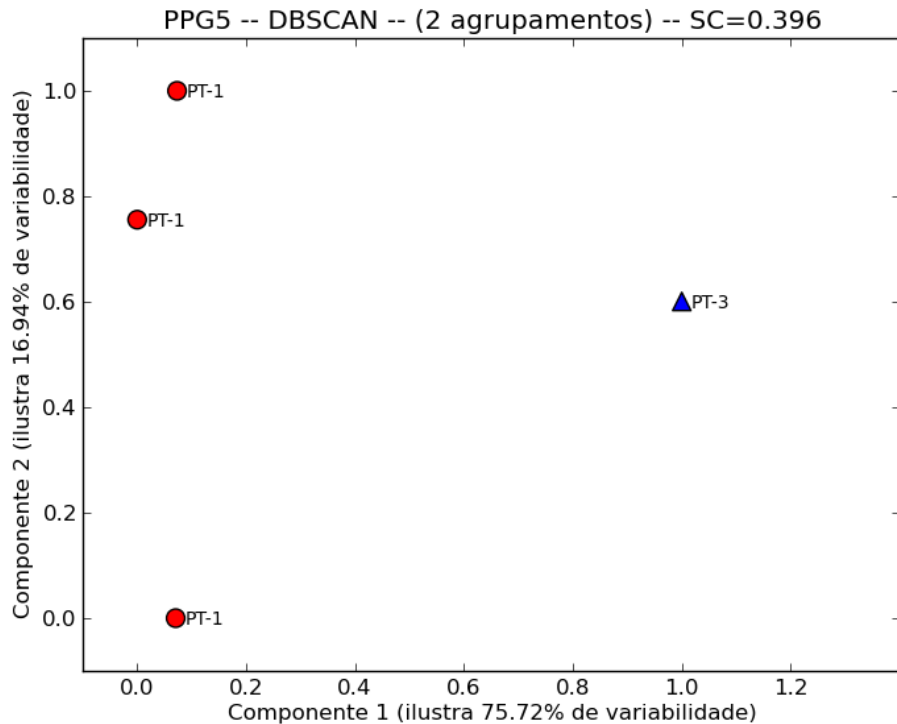


Figura 5.3: Visualização da Análise Intrapoço para o furo PPG5 (Silhueta).

cional foi identificada uma única petrofácies, PT-4. O método não classificou as amostras como pertencentes a mesma petrofácies. O método conseguiu distribuir 60% das lâminas de forma correta.

5.1.2 Índice de Calinski-Harabasz

A Tabela 5.2 mostra os parâmetros, que dentre os testados, originaram um valor máximo do coeficiente de silhueta. A homogeneidade, que pode ser observada na Tabela 5.2, mostra que as amostras de um mesmo agrupamento são bem similares entre si. Para todos os poços, o valor alcançado foi 1, que é o valor máximo da homogeneidade. A integralidade, por sua vez, mostrada na Tabela 5.2, define se as amostras pertencente a mesma petrofácies foram atribuídas ao mesmo agrupamento. Para os poços PPG3 e PPG5 o valor encontrado foi 1. Para os demais poços, os valores encontrados são considerados baixos.

As Figuras 5.4, 5.5 e 5.6 mostram as petrofácies identificadas e a forma como as amostras foram divididas nos agrupamentos originados pelo DBSCAN.

A análise para o PPG1 assinala que as duas componentes principais descrevem

Tabela 5.2: Parâmetros que maximizaram o índice de Calinski-Harabasz e os valores encontrados para homogeneidade e para integralidade

Poços	Distância	<i>Min_Pts</i>	Eps	CH	Homogeneidade	Integralidade
PPG1	Manhattan	1	4.687	96.087	1.000000	3.746522e-01
PPG2	Euclideana	1	5.159	16.112	1.000000	1.340905e-01
PPG3	Canberra	1	6.764	19.794	1.000000	1.000000e+00
PPG4	Chebyshev	2	9.943	6.324	1.000000	-1.649634e-16
PPG5	Chebyshev	1	10.895	6.122	1.000000	1.000000e+00

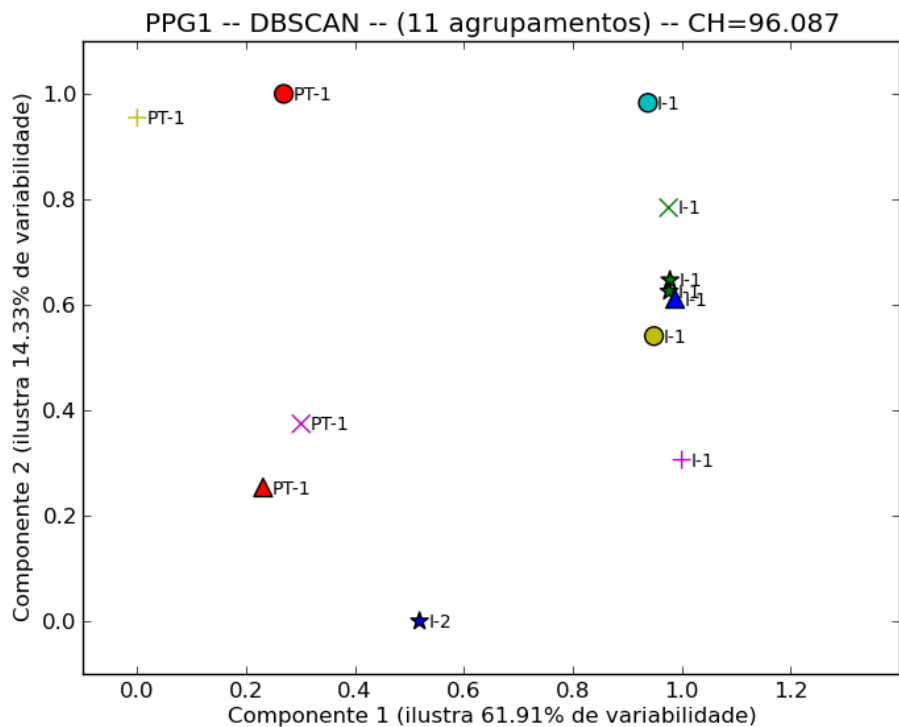


Figura 5.4: Visualização da Análise Intraçoço para os furos PPG1 (Calinski-Harabasz).

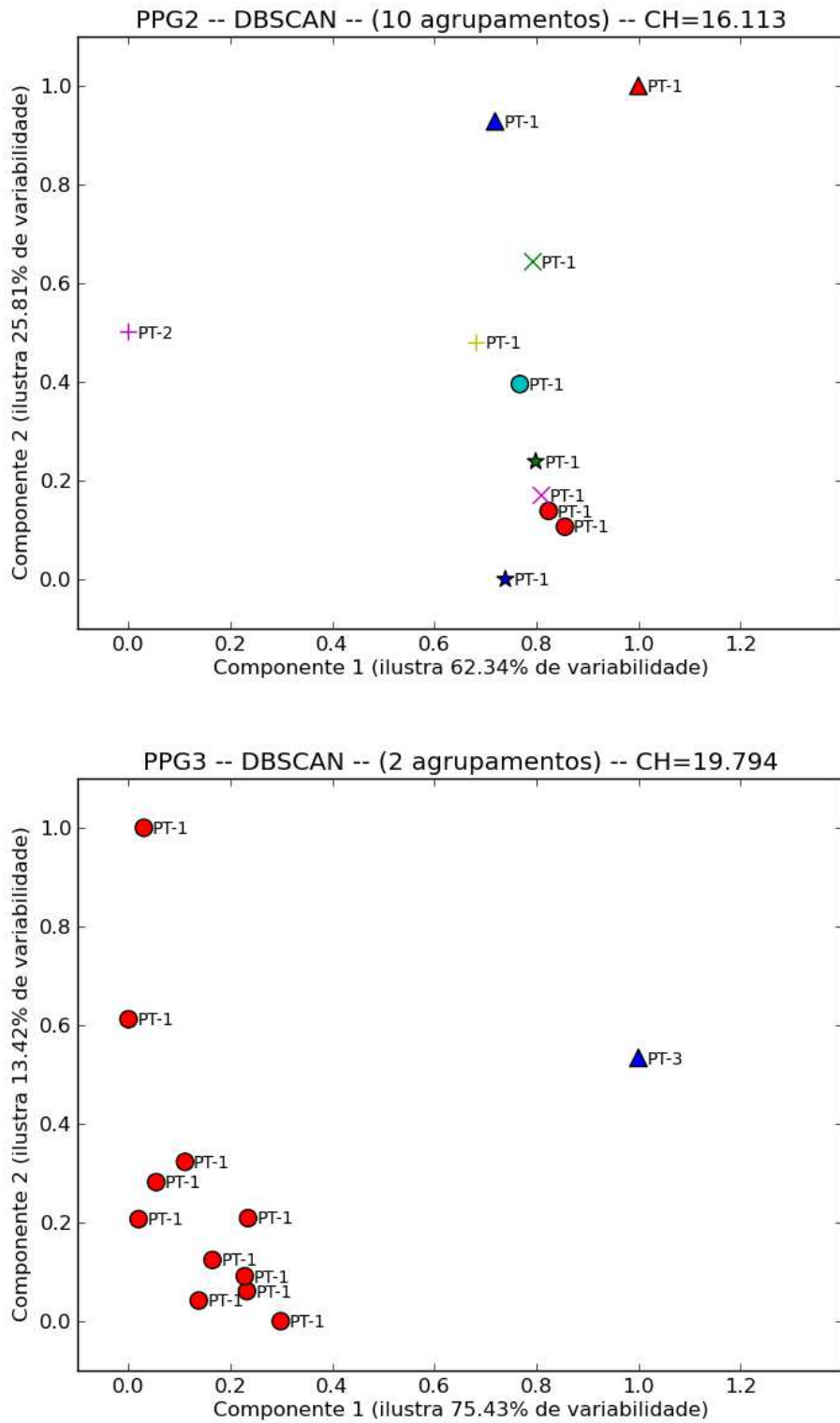


Figura 5.5: Visualização da Análise Intraçoço para os furos PPG2 e PPG3 (Calinski-Harabasz).

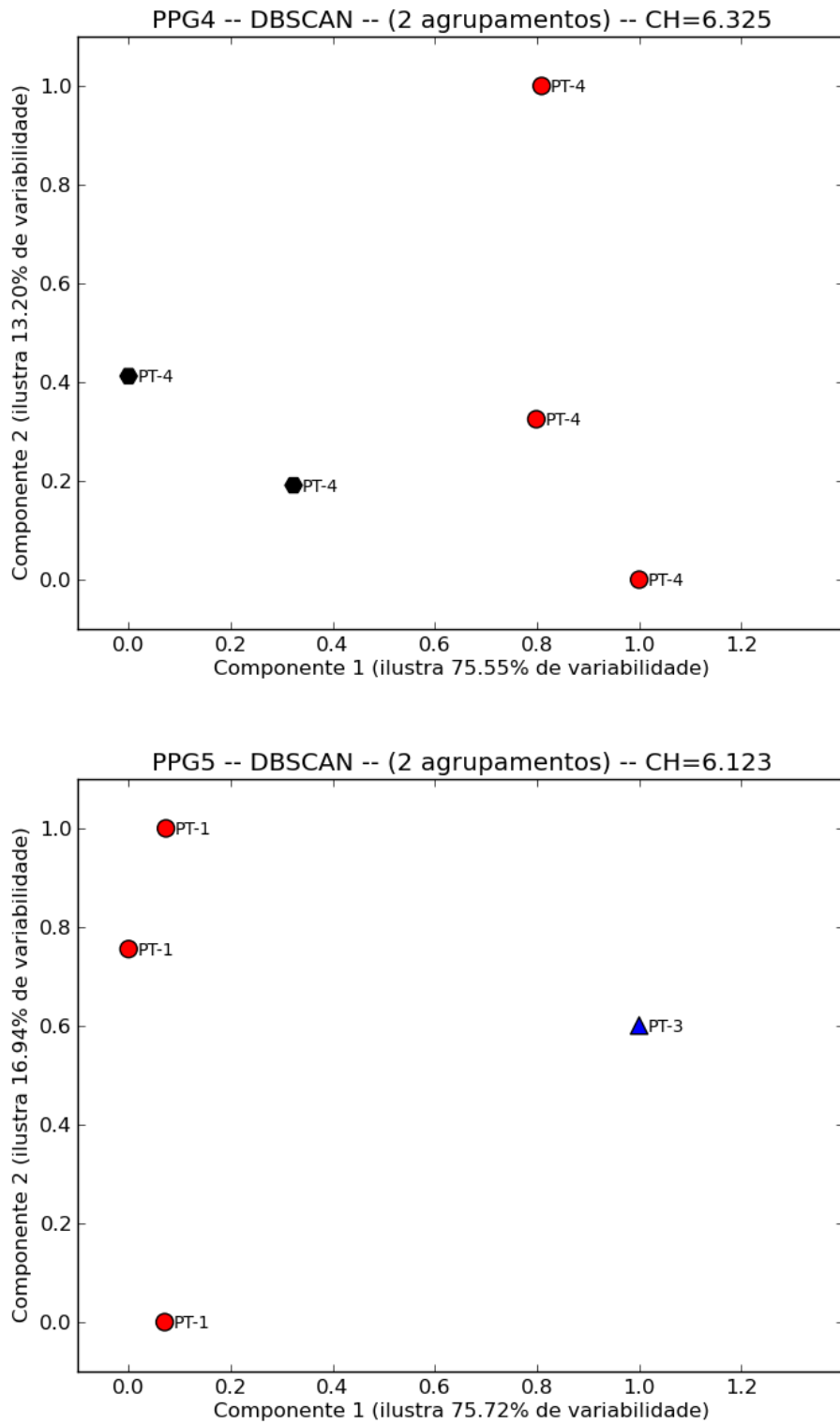


Figura 5.6: Visualização da Análise Intraçoço para o furo PPG4 e PPG5 (Calinski-Harabasz).

76,23% da variabilidade dos dados, sendo 61,91% descrita pela primeira componente e 14,33% pela segunda componente. Para o PPG1, o método identificou onze agrupamentos, separando amostras pertencentes a mesma petrofácies. Agrupou 25% das lâminas de forma correta.

No poço PPG2, as componentes principais descrevem 88,15% da variabilidade dos dados, 62,34% pela primeira componente e 25,81% pela segunda componente. No poço PPG3, as componentes principais representam 88,15% da variabilidade. Para o PPG2, o método identificou dez agrupamentos, separando amostras pertencentes a mesma petrofácies. O método conseguiu distribuir 27% das lâminas de forma correta.

Para o PPG4, as componentes principais reproduzem 88,75% da variabilidade dos dados, 75,55% pela primeira componente e 13,20% pela segunda componente. O método encontrou dois agrupamentos. Pelo método convencional foi identificada uma única petrofácies, PT-4. O método não classificou as amostras PPG4-1 e PPG4-5 como pertencentes a PT-4. O método conseguiu distribuir 60% das lâminas de forma correta.

No poço PPG3, as componentes principais representam 88,15% da variabilidade dos dados, 75,43% pela primeira componente e 13,42% pela segunda componente e PPG5, as componentes principais descrevem 92,66% da variabilidade dos dados, 75,72% pela primeira componente e 16,94% pela segunda componente. Nesses dois últimos poços citados, os resultados coincidiram com a classificação manual ao identificar corretamente as petrofácies associadas.

5.2 Análise Interpoço

O primeiro experimento realizado para a Análise Interpoço foi aplicar o método de agrupamento DBSCAN em uma base de dados contendo as lâminas de todos os poços. O resultado pode ser observado na Figura 5.7.

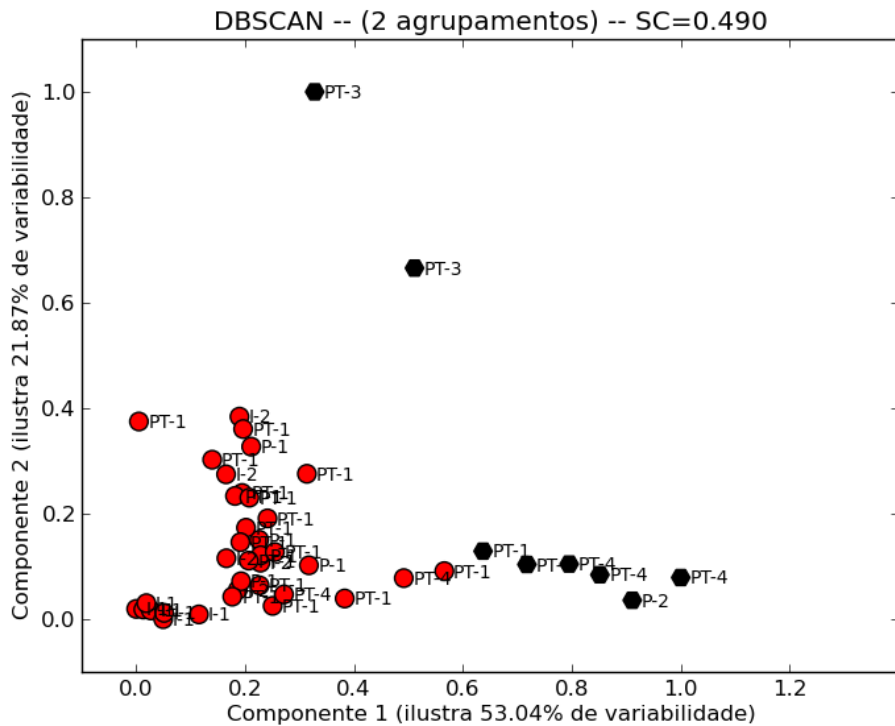


Figura 5.7: Visualização do Primeiro Experimento da Análise Interpoço.

O resultado não foi satisfatório, uma vez que foram encontrados somente dois agrupamentos. Assim, lâminas pertencentes a distintas petrofácies foram agrupadas em um mesmo agrupamento, o que não tem significado geológico. Além disso, lâminas que pertencem a diferentes petrofácies aparecem como dissimilares entre si.

A partir disso, outro procedimento foi realizado, o descrito na Seção 4.3. A Tabela 5.3 mostra as relações das lâminas e petrofácies com os grupos encontrados. Os métodos associaram, para cada grupo, lâminas pertencentes a uma petrofácies, com exceção do grupo GPPG1-1, que possui lâminas de duas petrofácies.

Tabela 5.3: Relação encontrada para o método DBSCAN

Grupos	Poços	Lâminas em cada grupo	Petrofácies
GPPG1-0	PPG1	PPG1-2 a PPG1-8	I-1
GPPG1-1	PPG1	PPG1-1, PPG1-9 a PPG1-12	I-2,PT-1
GPPG2-0	PPG2	PPG2-2 a PPG2-11	PT-1
GPPG2-1	PPG2	PPG2-1	PT-2
GPPG3-0	PPG3	PPG3-1 a PPG3-10, PPG3-12	PT-1
GPPG3-1	PPG3	PPG3-11	PT-3
GPPG4-0	PPG4	PPG4-2 a PPG4-4	PT-4
GPPG4-1	PPG4	PPG4-1, PPG4-5	PT-4
GPPG5-0	PPG5	PPG5-1, PPG5-2 e PPG5-4	PT-3
GPPG5-1	PPG5	PPG5-3	PT-1

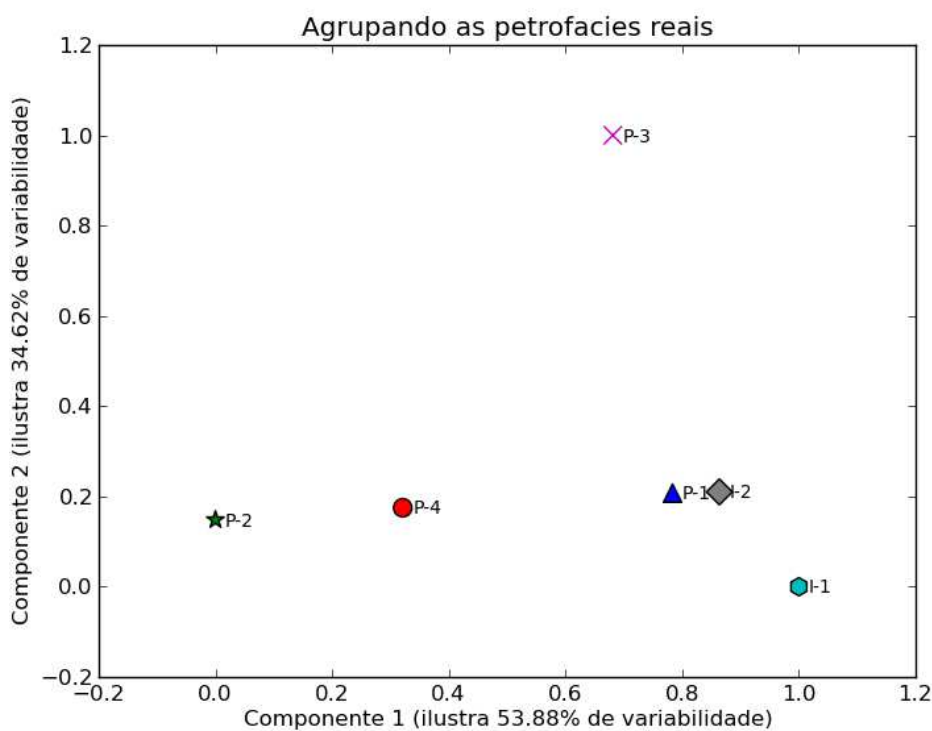


Figura 5.8: Visualização das petrofácies separadas pelo método convencional.

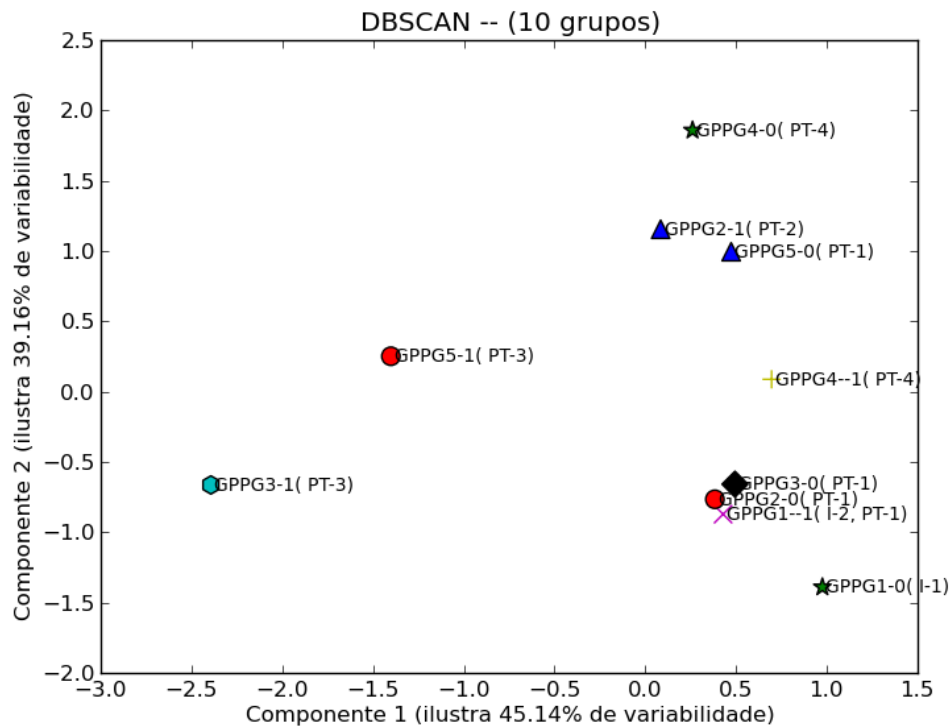


Figura 5.9: Visualização da distribuição dos grupos encontrados no Segundo Experimento da Análise Interpoço.

As Figuras 5.8 e 5.8 mostram a distribuição das petrofácies reais e das petrofácies encontradas pelo DBSCAN. As petrofácies reais estão representadas pela média dos constituintes das suas amostras. Uma vez que a Figura 5.9 pode estar rotacionada, percebe-se que as distribuições são similares.

É possível observar que o grupo GPPG1-0 aparece isolado, pois é o único que contém lâminas da petrofácies I-1. Apesar da classificação manual ter designado a petrofácies PT-2 para o grupo GPPG2-0 e PT-1 para o grupo GPPG5-0, a metodologia empregada sugere que esses dois grupos possuem características similares. De acordo com a Análise Intraçoço a petrofácies PT-4 foi atribuída para os grupos GPPG4-0 e GPPG4-1. Os resultados sugerem que o grupo GPPG1-1 compartilha características com o grupo GPPG2-1 e GPPG3-0, assim como os grupos GPPG3-1 e GPPG5-1, indicando que em cada conjunto as lâminas compartilham as mesmas características.

6 Conclusão e Trabalhos Futuros

A metodologia utilizada para identificação e classificação de petrofácies sedimentares baseou-se na aplicação de método de agrupamento e técnica para visualização dos resultados, as quais possibilitaram que os objetivos inicialmente propostos na Seção 1.3 fossem cumpridos. O estudo e a implementação de diferentes métodos de agrupamentos foram realizados anteriormente, e o resultado pode ser visto em (Saporetti *et al*, 2014). Nessa monografia utilizou-se somente o método que apresentou o melhor resultado, no caso o DBSCAN.

Na Análise Intrapoço, o uso do coeficiente de silhueta se mostrou mais eficaz para o propósito ao ser comparado ao índice de Calinski-Harabasz. Os resultados, com o primeiro critério de validação, foram mais próximos dos encontrados pelo método manual. O uso do DBSCAN se mostrou eficiente para a identificação de similaridades entre as amostras da base de dados estudada, obtendo grandes índices de acerto em comparação com o método manual. Dessa forma otimizou-se esse procedimento que era realizado de forma convencional para identificação de petrofácies.

A Análise Interpoço permite visualizar e identificar similaridades entre lâminas de diferentes grupos (de acordo com a média dos constituintes e segundo a distância Euclideana) e também sugere como a mesma petrofácies (ou conjunto de lâminas que compartilham as mesmas características) se distribui ao longo de diferentes poços. Os resultados foram coerentes com os encontrados na Análise Intrapoço.

De forma geral, a metodologia proposta apresentou ser praticável e válida para o propósito. Proporcionando assim, que conhecimentos sobre a heterogeneidade sejam obtidos, o que implica na determinação dos atributos do reservatório, o que permite que a metodologia possa ser aplicada em outras bases de dados semelhantes as usadas nessa monografia.

O método computacional pode assistir ao geólogo/petrólogo na tarefa de identificação das petrofácies, reduzindo o tempo de análise em comparação com a classificação manual.

Em decorrência desse trabalho, algumas possibilidades para trabalhos futuros são aplicar diferentes técnicas como de classificação, redes neurais, seleção de características, com o propósito de melhorar os resultados. Otimizando o processo de caracterização do reservatório podendo ser uma ferramenta a mais no processo de aumentar o fator de recuperação em reservatórios muito heterogêneos.

Referências Bibliográficas

- de ALBUQUERQUE, M. A. **Estabilidade em análise de agrupamentos**. 2005. Dissertação de Mestrado - Dissertação (Mestrado em Biometria) – Universidade Federal Rural de Pernambuco, Recife.
- de ANDRADE, L. P. **Agrupamentos de dados: Avaliação de métodos e desenvolvimento de aplicativo para análise de grupos**. 2004. Dissertação de Mestrado - Dissertação (Mestrado em Ciências em Engenharia Civil) – Programa de Pós-Graduação de Engenharia da Universidade Federal do Rio de Janeiro.
- de Oliveira ARAUJO, W. **Análise de componentes principais (acp)**. Technical report, Relatório Técnico - Mestrado Sociedade, Tecnologia e Meio Ambiente. Centro Universitário de Anápolis, 2009.
- BOSCHI, R. S. **Técnicas de mineração de dados na agricultura**. In: XV Semana Integrada do CEATEC, 2013.
- de Freitas BRAZIL, F. A. **Estratigrafia de sequências e processo diagenético: Exemplo dos arenitos marinho-rasos da formação ponta grossa, noroeste da bacia do paraná**. 2004. Dissertação de Mestrado - Dissertação de Mestrado – Programa de pós-graduação em Análise de Bacias e Faixas Móveis. Universidade do Estado do Rio de Janeiro.
- BRAGA, I. L. S. **Identificação e classificação de litofácies com o uso da teoria bayesiana de reconhecimento de padrões**. 2005. Dissertação de Mestrado - Dissertação (Mestrado em Engenharia de Reservatório e de Exploração) – Centro de Ciência e Tecnologia da Universidade Estadual do Norte Fluminense.
- CAMILO, C. O.; da SILVA, J. C. **Mineração de dados: Conceitos, tarefas, métodos e ferramentas**. Technical report, Instituto de Informática, Universidade Federal de Goiás, 2009.
- CEVOLANI, J. T.; de OLIVEIRA, L. C.; GOLIATT, L. ; PEREIRA, E. **Visualização e classificação automática de petrofácies sedimentares**. In: 6º Congresso Brasileiro de Pesquisa e Desenvolvimento em Petróleo e Gás, 2011.
- CHANG, H.-C.; KOPASKA-MERKEL, D. C. ; CHEN, H.-C. Identification of lithofacies using kohonen self-organizing maps. **Computers and Geosciences**, v.28, p. 223–229, 2002.
- CHEHRAZI, A.; REZAEI, R. A systematic method for permeability prediction, a petrofacies approach. **Journal of Petroleum Science and Engineering**, v.82, p. 1–16, 2012.
- da CUNHA, E. S. **Identificação de litofácies de poços de petróleo utilizando um método baseado em redes neurais artificiais**. 2002. Dissertação de Mestrado - Dissertação (Mestrado em Informática) – Programa de Pós-Graduação em Informática da Universidade Federal de Campina Grande.

- da Costa CÔRTEZ, S.; PORCARO, R. M. ; LIFSCHITZ, S. Mineração de dados: Funcionalidades, técnicas e abordagens. **PUC**, 2002.
- ROS, L. F. D.; GOLDBERG, K. **Reservoir petrofacies: A tool for quality characterization and prediction**. In: AAPG Annual Convention and Exhibition, 2007.
- ESTER, M.; KRIEGEL, H.; SANDER, J. ; XU, X. **A density-based algorithm for discovering clusters in large spatial databases with noise**. 1996.
- FONSECA, J. R. S. **Os métodos quantitativos na sociologia: Dificuldades de uma metodologia de investigação**. In: VI Congresso Português de Sociologia, Mundos Sociais: Saberes e Práticas, volume 346, 2008.
- FÁVERA, J. C. D. 2001.
- GODOY, L. C. Apostila de geologia: Rochas. **Universidade Estadual de Ponta Grossa**, 2005.
- GRAHN, Y. Recent progress in the silurian and devonian biostratigraphy of the paraná basin in brazil and paraguay. **Universidade do Estado do Rio de Janeiro**, p. 147–163, 1999.
- GRESSLY, A. Observations géologiques sur le jura soleurois. **Neue Denkschr**, v.2, p. 1–112, 1938.
- HSIEH, B.-Z.; LEWIS, C. ; LIN, Z.-S. Lithology identification of aquifers from geophysical well logs and fuzzy logic analysis: Shui-lin area, taiwan. **Computers and Geosciences**, v.31, p. 263–275, 2005.
- JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 6.ed.ED. Person, 2007.
- ans Peter J. ROUSSEEUW, L. K. **Finding Groups in Data: An Introduction to Cluster Analysis**. John Wiley & Sons, 1990.
- KUMAR, B.; KISHORE, M. **Electrofacies classification: A critical approach**. In: 6th International Conference and Exposition on Petroleum Geophysics, p. 822–825, 2006.
- LACENTRE, P. E.; CARRICA, P. M. A method to estimate permeability on uncored wells. **World Oil**, v.225, 2013.
- LEE, S. H.; DATTA-GUPTA, A. **Electrofacies characterization and permeability predictions in carbonate reservoirs: Role of multivariate analysis and nonparametric regression**. In: Society of Petroleum Engineers – Annual Technical Conference and Exhibiton, 1999.
- LIAO, S.-H.; CHU, P.-H. ; HSIAO, P.-Y. Data mining techniques and applications - a decade review from 2000 to 2011. v.39, 2012.
- LIM, J.-S. Reservoir properties determination using fuzzy logic and neural networks from well data in offshore korea. **Journal of Petroleum Science and Engineering**, v.49, p. 182–192, 2005.

- de Fátima MACHADO, C.; de Sousa NUNES, G. H.; FERREIRA, D. F. ; dos SANTOS, J. B. Divergência genética entre genótipos de feijoeiro a partir de técnicas multivariada. v.32, p. 251–258, 2002.
- MARASCHIN, A. J.; MIZUSAKI, A. M. Datação de processos diagenéticos em arenitos-reservatório de hidrocarbonetos: Uma revisão conceitual. v.35(1), p. 27–41, 2008.
- de MENEZES, M. R. F. **Estudos sedimentológicos e o contexto estrutural da formação serra do martins, nos platôs de portaalegre, martins e santana, rn.** 1999. Dissertação de Mestrado - Dissertação (Mestrado em Geodinâmica) – Programa de Pesquisa e Pós-Graduação em Geodinâmica e Geofísica da UFRN, Natal.
- MILANI, E. J. **Evolução tectono-estratigráfica da Bacia do Paraná e seu relacionamento com a geodinâmica fanerozóica do Gondwana Sul-Occidental.** 1997. Tese de Doutorado - Tese Doutorado – Instituto de Geociências – Universidade Federal do Rio Grande do Sul, Porto Alegre.
- MILANI, E. J.; ZALAN, P. V. An outline of the geology and petroleum systems of the paleozoic interior basins of south america. **Episodes**, v.22, p. 199–205, 1999.
- MILANI, E. J.; de MELO, J. H. G.; de SOUZA, P. A.; FERNANDES, L. A. ; FRANÇA, A. B. Bacia do paraná. **Boletim de Geociências da PETROBRAS, Rio de Janeiro**, v.15, p. 265–287, 2007.
- OLIVEIRA, L. C.; PEREIRA, E. **Aplicação de parâmetros diagenéticos para a caracterização do arcabouço estratigráfico do devoniano da bacia do paraná.** In: 5o Congresso Brasileiro de Pesquisa e Desenvolvimento em Petróleo e Gás, 2009.
- OLIVEIRA, L. C. **Estudos das relações entre o arcabouço estratigráfico e as alterações diagenéticas observadas na seção devoniana da bacia do paraná.** 2009. Dissertação de Mestrado - Universidade do Estado do Rio de Janeiro.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M. ; DUCHESNAY, E. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v.12, p. 2825–2830, 2011.
- da Silva PEREIRA, J. E. **Aplicação de técnicas estatísticas à integração de dados geológicos, geofísicos na região do município de Santa Maria, RS, Brasil.** 2009. Tese de Doutorado - Tese (Doutorado em Engenharia) – Escola de Engenharia, PPGEM, Porto Alegre.
- PETRI, S.; FÚLVARO, V. J. **Geologia do Brasil.** Editora Universidade de São Paulo, 1983, 631p.
- Technology, P. G. 2011.
- PONTE, F. B. Fluxo de trabalho para mapeamento de litofácies com aplicação no campo de namorado, bacia de campos, rio de janeiro. **Monografia (Graduação em Geologia) – Universidade Federal Rural do Rio de Janeiro**, 2010.

- RAEESI, M.; MORADZADEH, A.; ARDEJANI, F. D. ; RAHIMI, M. Classification and identification of hydrocarbon reservoir lithofacies and their heterogeneity using seismic attributes, logs data and artificial neural networks. **Journal of Petroleum Science and Engineering**, v.82, p. 151–165, 2012.
- RENCHER, A. C. **Methods of Multivariate Analysis**. Second. ed., Brigham Young University, 2002.
- ROCHA, C. L. **Análise de fronteiras de reservatório de petróleo através de geoquímica de superfície e mineração de dados**. 2005. Dissertação de Mestrado - Dissertação (Mestrado em Ciências em Engenharia Civil) – Programa Pós-Graduação de Engenharia da UFRJ.
- ROSENBERG, A.; HIRSCHBERG, J. **V-measure: A conditional entropy-based external cluster evaluation measure**. In: Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, p. 410–420, 2007.
- ROSA, H. **Estudo de Caracterização de Eletrofácies por meio de Perfis Geofísicos de Poços e de Amostras de Testemunhos Utilizando Estatística Multivariada**. 2008. Tese de Doutorado - Tese (Doutorado em Ciências e Engenharia de Petróleo) – Universidade Estadual de Campinas, Faculdade de Engenharia Mecânica e Instituto de Geociências.
- ROSA, H.; SUSLICK, S. B.; VIDAL, A. C. ; SAKAI, G. K. Caracterização de eletrofácies por meio de ferramentas estatísticas multivariadas. v.61(4), p. 415–422, 2008.
- ROUSSEEUW, P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, 1987.
- SAPORETTI, C. M.; CEVOLANI, J. T.; de OLIVEIRA, L. C.; GOLIATT, L. ; PEREIRA, E. **Comparação de métodos de agrupamento para classificação de dados petrográficos**. In: XI Simpósio de Mecânica Computacional e II Encontro Mineiro de Modelagem Computacional, 2014.
- Pushpa SHARMA, G. M.; BAHUGUNA, V. K. ; LAL, C. **Improved permeability estimates in carbonate reservoirs using electrofacies characterization: A case study of mumbai high south**. In: 2nd South Asian Geoscience Conference, 2012.
- de SOUZA, W. A. **Análise comportamental da optimum-path forest em diferentes funções métricas**. Monografia (Bacharel em Ciência da Computação) - Faculdade de Ciências - UNESP/Bauru, 2008.
- SUGUIO, K. **Rochas Sedimentares: propriedades, gênese, importância econômica**. Edgard Blucher: Ed. da Universidade de São Paulo, 1980.
- TAVAKOLI, V.; AMINI, A. Application of multivariate cluster analysis in logfacies determination and reservoir zonation, case study of marun field, south of iran. v.32(2), p. 69–75, 2006.
- TOEBE, J.; SAMPAIO, M. C. Mineração de dados aplicada à prospecção de petróleo e gás natural. **Revista Eletrônica da UNESC**, v.2, 2003.

- do VALE, M. N. **Agrupamentos de dados: Avaliação de métodos e desenvolvimento de aplicativo para análise de grupos**. 2005. Dissertação de Mestrado - Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro.
- VENDRAMIN, L.; CAMPELLO, R. J. G. B. ; HRUSCHKA, E. R. Relative clustering validity criteria: A comparative overview. **Statistical Analysis and Data Mining**, v.3(4), p. 209–235, 2010.
- VIALI, L. **Análise de Agrupamentos**. Notas de Aula – Universidade de Caxias do SUL, 2000.
- XIE, J. **Improved permeability prediction using multivariate analysis methods**. 2008. Dissertação de Mestrado - Thesis (Master of Science) – Office of Graduate Studies of Texas A and M University.
- ZALÁN, P. V.; WOLFF, S.; de Jesus CONCEIÇÃO, J. C. ; MARQUES, A. Bacia do paran. **Origem e Evoluao de Bacias Sedimentares**. Rio de Janeiro: PETROBRAS, 1990.
- ZHANG, X. **Collaboration of information from different sources for petroleum reservoir prediction**. In: The Second International Conference on Information Fusion. Sunnyvale, USA: International Society of Information Fusion, 1999.

A Base de Dados

As petrofácies mostradas nas tabelas abaixo foram individualizadas pela classificação manual.

Tabela A.1: Amostras retiradas do PPG1

Profundidade (m)	7.7	11	16.25	20.7	23	24	27.25	27.26	27.27	27.28	27.29	27.30
Lâmina delgada	PPG1-1	PPG1-2	PPG1-3	PPG1-4	PPG1-5	PPG1-6	PPG1-7	PPG1-8	PPG1-9	PPG1-10	PPG1-11	PPG1-12
Petrofácies	I-2	I-1	I-1	I-1	I-1	I-1	I-1	I-1	PT-1	PT-1	PT-1	PT-1
Quartzo	61.6	69.2	68.7	68.3	67.4	67.6	68.3	62.4	61	61	64	59.6
Feldspato	1.7	4	4	4.7	2.7	1.7	1.7	2.3	0.4	0	0.6	0.4
Muscovita	6.4	0.6	0.4	0	0.4	0.6	0.3	0.3	7.4	6.7	12.3	9.6
Opaco	4	0.6	1	0	0	1.7	0	0.3	3.6	1.6	4.3	3.6
Turmalina	0.3	0.4	0	0	0	0	0	0	0	1	0	0
Zircão	0.4	0	0.3	0	0.3	0	0	0.3	0.3	0.3	0.3	0.3
Rutilo	0.3	0	0	0	0	0	0	0	0	0	0	0
Glauconita	0	0	0	0	0	0	0	0	0	0	0	0
Clorita	0	0	0	0	0	0	0	0	0	0	0	0
Pseudo Matriz	6	1.7	2.9	4.6	6.4	1	8.3	5.3	1.6	1	0.4	1
Litoclasto	0	0	0.3	1	0	0	3.6	1.3	0	0	0	0
Bioclasto	0	0	0	0	0	0	0	0	0	0	0.4	0.4
Cresc. Sec. Qtz	3.7	3.6	3.4	2.7	2.7	2	0.7	1.4	1.4	0.7	1.7	1.7
Caolinita	9	11.3	11	10.3	16.4	7	8.7	12.7	6	6	3	3.6
Ilita/Smectita	2.6	0	0	0	0	0.3	0.4	1	10.3	4	5	2.4
Pirita	0	0	0	0	0	0	0	0	1.6	0.3	1	0
Siderita	4	0	0	0	0	0	0	0	0	0	3.3	0.4
Cim. Carbonático	0	0	0	0	0	0	0	0	3	11	2	15
Cim. Silicoso	0	0	0	0	0	0	0	0	0	0	0	0
Cim. Ferruginoso	0	0	0	0	0	0	0	0	0	0	0	0
Por. Intergranular	0	8	8	7.7	3.7	13.7	7	12.4	3	5.4	1.7	1
Por. Intragranular	0	0.6	0	0.7	0	0	1	0.3	0.4	1	0	1

Tabela A.3: Amostras retiradas do PPG3

Profundidade (m)	41.55	41.56	41.57	41.58	41.59	41.6	41.61	41.62	41.63	41.64	41.65	41.66
Lâmina delgada	PPG3-1	PPG3-2	PPG3-3	PPG3-4	PPG3-5	PPG3-6	PPG3-7	PPG3-8	PPG3-9	PPG3-10	PPG3-11	PPG3-12
Petrofácies	PT-1	PT-1	PT-1	PT-1	PT-1	PT-1	PT-1	PT-1	PT-1	PT-1	PT-3	PT-1
Quartzo	58.7	57.4	53.4	56	59.7	57	57.7	56	58	61	45.6	57.7
Feldspato	1	1.6	2	2.3	4	2.6	7	3	3.5	2.3	2	3
Muscovita	4.3	6	5	8.4	5	2.6	2.4	4	6.4	6.3	3	5.7
Opaco	2	1	0.7	2.3	1.7	1.7	5	3	2.7	2.4	1.6	6.7
Turmalina	0	0	0.6	0.3	0	0.6	0	0	0	0.4	0	0
Zircão	0.3	0	0.3	0	0.4	1	0.3	0.3	0.4	0	0.3	0.7
Rutilo	0	0	0	0	0.3	0	0	0	0	0	0.3	0
Glauconita	0	0	0	0	0	0	0	0.3	0.4	0.4	0.7	0.6
Clorita	0	0	0	0	0	0	0	0	0	0	0	0
Pseudo Matriz	0.3	1.3	1.3	2	0.6	0.6	0.3	0.6	0	0.3	0	1
Litoclasto	0	0	0	0	0	0	0	0	0	0	0	0
Bioclasto	0.4	0.7	0.3	0	0.3	0	0	0.4	0	0.3	0	0
Cresc. Sec. Qtz	3.3	4.7	4	3	3	5	4.6	3.4	5.4	3	2	2.3
Caolinita	15	13.3	10.4	15	18	10.3	9	11	9	4.6	0.4	4.6
Ilita/Smectita	12	5	18.7	4.4	3.7	2.6	1	1	1	0	0	1.7
Pirita	0	0	0	0	0	0	0	0	0	0	0	0
Siderita	1.7	2	0	0	0.3	0	0	0.4	5.4	1	0	0
Cim. Carbonático	0	1	0.4	4	0.6	9	8.1	4.7	3.4	11	43.7	6.4
Cim. Silicoso	0	0	0	0	0	0	0	0	0	0	0	0
Cim. Ferruginoso	0	0	0	0	0	0	0	0	0	0	0	0
Por. Intergranular	1	5.6	2.6	2.3	2.4	6	4	4	4	3.3	0	0.6
Por. Intragranular	0.0	0.4	0.0	0.0	0.3	1.0	0.0	2.0	0.4	0.3	0.0	0.0

Tabela A.4: Amostras retiradas do PPG4

Profundidade (m)	17.8	18.7	19.4	20.7	24.3
Lâmina delgada	PPG4-1	PPG4-2	PPG4-3	PPG4-4	PPG4-5
Petrofácies	PT-4	PT-4	PT-4	PT-4	PT-4
Quartzo	55.3	20.6	30	33.6	44
Feldspato	1.6	0.7	4	2.6	8.3
Muscovita	4	5.4	9.3	16	9.3
Opaco	3	4	2	3.7	2.4
Turmalina	0	0	0	0	0
Zircão	0.4	0.7	0.6	0.3	1
Rutilo	0	0.3	0	0	0.7
Glauconita	0	0.3	0	0	0
Clorita	0.3	0	0	0	0
Pseudo Matriz	0	0	0	0	0
Litoclasto	0	0	0	0	0
Bioclasto	0	0	0	0	0
Cresc. Sec. Qtz	2.3	1.3	3.4	2.4	3
Caolinita	9	7	6.3	1.7	13
Ilita/Smectita	7	17.3	20	20	10
Pirita	6.7	10.3	7	15	3
Siderita	3.3	4.6	3	3.6	3.6
Cim. Carbonático	0	0	0	0.7	0
Cim. Silicoso	0	0	0	0	0
Cim. Ferruginoso	0.3	0.6	0	0	0
Por. Intergranular	11	2.6	2.4	0.4	0.7
Por. Intragranular	0	0	0	0	0

Tabela A.5: Amostras retiradas do PPG5

Profundidade(m)	21, 5	46, 4	47, 1	55, 3
Lâmina delgada	PPG5-1	PPG5-2	PPG5-3	PPG5-4
Petrofácies	PT-1	PT-1	PT-3	PT-1
Quartzo	40,0	34,3	38,3	41,6
Feldspato	6,0	5,6	1,3	1,7
Muscovita	19,0	23,6	7,3	18,0
Opaco	0,6	1,3	1,0	3,4
Turmalina	0,0	0,0	0,0	0,0
Zircão	0,3	1,0	0,0	0,4
Rutilo	0,0	0,0	0,3	0,0
Glauconita	0,0	0,0	0,0	0,3
Clorita	0,0	0,0	0,0	0,7
Pseudo Matriz	1,3	7,0	2,0	12,7
Litoclasto	0,0	0,0	0,0	0,0
Bioclasto	0,0	0,0	0,0	0,0
Cresc. Sec. Qtz	3,4	3,0	7,6	1,0
Caolinita	10,0	11,3	8,0	1,0
Ilita/Smectita	10,4	12,3	6,3	13,4
Pirita	0,0	0,0	0,0	0,0
Siderita	0,0	0,0	0,0	0,0
Cim. Carbonático	0,0	0,0	27,6	1,4
Cim. Silicoso	0,0	0,0	0,0	0,7
Cim. Ferruginoso	0,0	0,0	0,0	1,4
Por. Intergranular	9,0	0,6	0,3	1,0
Por. Intragranular	0,0	0,0	0,0	0,0

B Resultado Análise Intrapoço

Tabela B.1: Resultado DBSCAN para o PPG1

Profundidade (m)	7.7	11	16.25	20.7	23	24	27.25	27.26	27.27	27.28	27.29	27.30
Lâmina delgada	PPG1-1	PPG1-2	PPG1-3	PPG1-4	PPG1-5	PPG1-6	PPG1-7	PPG1-8	PPG1-9	PPG1-10	PPG1-11	PPG1-12
Petrofácies	I-2	I-1	I-1	I-1	I-1	I-1	I-1	I-1	PT-1	PT-1	PT-1	PT-1
DBSCAN (Agrupamentos)	1	2	2	2	2	2	2	2	1	1	1	1
Quartzo	61.6	69.2	68.7	68.3	67.4	67.6	68.3	62.4	61	61	64	59.6
Feldspato	1.7	4	4	4.7	2.7	1.7	1.7	2.3	0.4	0	0.6	0.4
Muscovita	6.4	0.6	0.4	0	0.4	0.6	0.3	0.3	7.4	6.7	12.3	9.6
Opaco	4	0.6	1	0	0	1.7	0	0.3	3.6	1.6	4.3	3.6
Turmalina	0.3	0.4	0	0	0	0	0	0	0	1	0	0
Zircão	0.4	0	0.3	0	0.3	0	0	0.3	0.3	0.3	0.3	0.3
Rutilo	0.3	0	0	0	0	0	0	0	0	0	0	0
Glauconita	0	0	0	0	0	0	0	0	0	0	0	0
Clorita	0	0	0	0	0	0	0	0	0	0	0	0
Pseudo Matriz	6	1.7	2.9	4.6	6.4	1	8.3	5.3	1.6	1	0.4	1
Litoclasto	0	0	0.3	1	0	0	3.6	1.3	0	0	0	0
Bioclasto	0	0	0	0	0	0	0	0	0	0	0.4	0.4
Cresc. Sec. Qtz	3.7	3.6	3.4	2.7	2.7	2	0.7	1.4	1.4	0.7	1.7	1.7
Caolinita	9	11.3	11	10.3	16.4	7	8.7	12.7	6	6	3	3.6
Ilita/Smectita	2.6	0	0	0	0	0.3	0.4	1	10.3	4	5	2.4
Pirita	0	0	0	0	0	0	0	0	1.6	0.3	1	0
Siderita	4	0	0	0	0	0	0	0	0	0	3.3	0.4
Cim. Carbonático	0	0	0	0	0	0	0	0	3	11	2	15
Cim. Silicoso	0	0	0	0	0	0	0	0	0	0	0	0
Cim. Ferruginoso	0	0	0	0	0	0	0	0	0	0	0	0
Por. Intergranular	0	8	8	7.7	3.7	13.7	7	12.4	3	5.4	1.7	1
Por. Intragranular	0	0.6	0	0.7	0	0	1	0.3	0.4	1	0	1

Tabela B.3: Resultado DBSCAN para o PPG3

Profundidade (m)	41.55	41.56	41.57	41.58	41.59	41.6	41.61	41.62	41.63	41.64	41.65	41.66
Lâmina delgada	PPG3-1	PPG3-2	PPG3-3	PPG3-4	PPG3-5	PPG3-6	PPG3-7	PPG3-8	PPG3-9	PPG3-10	PPG3-11	PPG3-12
Petrofácies	PT-1	PT-1	PT-1	PT-1	PT-1	PT-1	PT-1	PT-1	PT-1	PT-1	PT-3	PT-1
DBSCAN (Agrupamentos)	1	1	1	1	1	1	1	1	1	1	2	1
Quartzo	58.7	57.4	53.4	56	59.7	57	57.7	56	58	61	45.6	57.7
Feldspato	1	1.6	2	2.3	4	2.6	7	3	3.5	2.3	2	3
Muscovita	4.3	6	5	8.4	5	2.6	2.4	4	6.4	6.3	3	5.7
Opaco	2	1	0.7	2.3	1.7	1.7	5	3	2.7	2.4	1.6	6.7
Turmalina	0	0	0.6	0.3	0	0.6	0	0	0	0.4	0	0
Zircão	0.3	0	0.3	0	0.4	1	0.3	0.3	0.4	0	0.3	0.7
Rutilo	0	0	0	0	0.3	0	0	0	0	0	0.3	0
Glauconita	0	0	0	0	0	0	0	0.3	0.4	0.4	0.7	0.6
Clorita	0	0	0	0	0	0	0	0	0	0	0	0
Pseudo Matriz	0.3	1.3	1.3	2	0.6	0.6	0.3	0.6	0	0.3	0	1
Litoclasto	0	0	0	0	0	0	0	0	0	0	0	0
Bioclasto	0.4	0.7	0.3	0	0.3	0	0	0.4	0	0.3	0	0
Cresc. Sec. Qtz	3.3	4.7	4	3	3	5	4.6	3.4	5.4	3	2	2.3
Caolinita	15	13.3	10.4	15	18	10.3	9	11	9	4.6	0.4	4.6
Ilita/Smectita	12	5	18.7	4.4	3.7	2.6	1	1	1	0	0	1.7
Pirita	0	0	0	0	0	0	0	0	0	0	0	0
Siderita	1.7	2	0	0	0.3	0	0	0.4	5.4	1	0	0
Cim. Carbonático	0	1	0.4	4	0.6	9	8.1	4.7	3.4	11	43.7	6.4
Cim. Silicoso	0	0	0	0	0	0	0	0	0	0	0	0
Cim. Ferruginoso	0	0	0	0	0	0	0	0	0	0	0	0
Por. Intergranular	1	5.6	2.6	2.3	2.4	6	4	4	4	3.3	0	0.6
Por. Intragranular	0.0	0.4	0.0	0.0	0.3	1.0	0.0	2.0	0.4	0.3	0.0	0.0

Tabela B.4: Resultado DBSCAN para o PPG4

Profundidade (m)	17.8	18.7	19.4	20.7	24.3
Lâmina delgada	PPG4-1	PPG4-2	PPG4-3	PPG4-4	PPG4-5
Petrofácies	PT-4	PT-4	PT-4	PT-4	PT-4
DBSCAN (Agrupamentos)	1	2	2	2	1
Quartzo	55.3	20.6	30	33.6	44
Feldspato	1.6	0.7	4	2.6	8.3
Muscovita	4	5.4	9.3	16	9.3
Opaco	3	4	2	3.7	2.4
Turmalina	0	0	0	0	0
Zircão	0.4	0.7	0.6	0.3	1
Rutilo	0	0.3	0	0	0.7
Glauconita	0	0.3	0	0	0
Clorita	0.3	0	0	0	0
Pseudo Matriz	0	0	0	0	0
Litoclasto	0	0	0	0	0
Bioclasto	0	0	0	0	0
Cresc. Sec. Qtz	2.3	1.3	3.4	2.4	3
Caolinita	9	7	6.3	1.7	13
Ilita/Smectita	7	17.3	20	20	10
Pirita	6.7	10.3	7	15	3
Siderita	3.3	4.6	3	3.6	3.6
Cim. Carbonático	0	0	0	0.7	0
Cim. Silicoso	0	0	0	0	0
Cim. Ferruginoso	0.3	0.6	0	0	0
Por. Intergranular	11	2.6	2.4	0.4	0.7
Por. Intragranular	0	0	0	0	0

Tabela B.5: Resultado DBSCAN para o PPG5

Profundidade(m)	21, 5	46, 4	47, 1	55, 3
Lâmina delgada	PPG5-1	PPG5-2	PPG5-3	PPG5-4
Petrofácies	PT-1	PT-1	PT-3	PT-1
DBSCAN (Agrupamentos)	1	1	2	1
Quartzo	40,0	34,3	38,3	41,6
Feldspato	6,0	5,6	1,3	1,7
Muscovita	19,0	23,6	7,3	18,0
Opaco	0,6	1,3	1,0	3,4
Turmalina	0,0	0,0	0,0	0,0
Zircão	0,3	1,0	0,0	0,4
Rutilo	0,0	0,0	0,3	0,0
Glauconita	0,0	0,0	0,0	0,3
Clorita	0,0	0,0	0,0	0,7
Pseudo Matriz	1,3	7,0	2,0	12,7
Litoclasto	0,0	0,0	0,0	0,0
Bioclasto	0,0	0,0	0,0	0,0
Cresc. Sec. Qtz	3,4	3,0	7,6	1,0
Caolinita	10,0	11,3	8,0	1,0
Ilita/Smectita	10,4	12,3	6,3	13,4
Pirita	0,0	0,0	0,0	0,0
Siderita	0,0	0,0	0,0	0,0
Cim. Carbonático	0,0	0,0	27,6	1,4
Cim. Silicoso	0,0	0,0	0,0	0,7
Cim. Ferruginoso	0,0	0,0	0,0	1,4
Por. Intergranular	9,0	0,6	0,3	1,0
Por. Intragranular	0,0	0,0	0,0	0,0

C Média dos Constituintes

Tabela C.1: Média dos Percentuais dos Constituintes para o PPG1

Lâminas	PPG1-2 a PPG1-8	PPG1-1, PPG1-9 a PPG1-12
Petrofácies	I-1	I-2, PT-1
DBSCAN (Agrupamentos)	GPPG1-0	GPPG1-1
Quartzo	67.414286	61.44
Feldspato	3.014286	0.62
Muscovita	0.371429	8.48
Opaco	0.514286	3.42
Turmalina	0.057143	0.26
Zircão	0.128571	0.32
Rutilo	0.000000	0.06
Glaucônita	0.000000	0.00
Clorita	0.000000	0.00
Pseudo Matriz	4.314286	2.00
Litoclasto	0.885714	0.00
Bioclasto	0.000000	0.16
Cresc. Sec. Qtz	2.357143	1.84
Caolinita	11.057143	5.52
Ilita/Smectita	0.242857	4.86
Pirita	0.000000	0.58
Siderita	0.000000	1.54
Cim. Carbonático	0.000000	6.20
Cim. Silicoso	0.000000	0.00
Cim. Ferruginoso	0.000000	0.00
Por. Intergranular	8.642857	2.22
Por. Intragranular	0.371429	0.48

Tabela C.2: Média dos Percentuais dos Constituintes para o PPG2

Lâminas	PPG2-2 a PPG2-11	PPG2-1
Petrofácies	PT-1	PT-2
DBSCAN (Agrupamentos)	GPPG2-0	GPPG2-1
Quartzo	61.06	30.0
Feldspato	1.32	0.4
Muscovita	10.46	1.0
Opaco	2.59	1.0
Turmalina	0.23	0.0
Zircão	0.30	0.3
Rutilo	0.03	0.0
Glauconita	0.12	0.0
Clorita	0.00	0.0
Pseudo Matriz	0.79	1.0
Litoclasto	0.00	0.0
Bioclasto	0.11	0.0
Cresc. Sec. Qtz	2.39	0.6
Caolinita	2.59	1.0
Ilita/Smectita	6.69	1.0
Pirita	1.29	0.0
Siderita	2.49	1.0
Cim. Carbonático	6.54	0.0
Cim. Silicoso	0.00	0.0
Cim. Ferruginoso	0.00	0.0
Por. Intergranular	0.86	4.0
Por. Intragranular	0.00	0.0

Tabela C.3: Média dos Percentuais dos Constituintes para o PPG3

Lâminas	PPG3-1 a PPG3-10, PPG3-12	PPG3-11
Petrofácies	PT-1	PT-3
DBSCAN (Agrupamentos)	GPPG3-0	GPPG3-1
Quartzo	57.509091	45.6
Feldspato	2.936364	2.0
Muscovita	5.100000	3.0
Opaco	2.654545	1.6
Turmalina	0.172727	0.0
Zircão	0.336364	0.3
Rutilo	0.027273	0.3
Glauconita	0.154545	0.7
Clorita	0.000000	0.0
Pseudo Matriz	0.754545	0.0
Litoclasto	0.000000	0.0
Bioclasto	0.218182	0.0
Cresc. Sec. Qtz	3.790909	2.0
Caolinita	10.927273	0.4
Ilita/Smectita	4.645455	0.0
Pirita	0.000000	0.0
Siderita	0.981818	0.0
Cim. Carbonático	4.418182	43.7
Cim. Silicoso	0.000000	0.0
Cim. Ferruginoso	0.000000	0.0
Por. Intergranular	3.254545	0.0
Por. Intragranular	0.400000	0.0

Tabela C.4: Média dos Percentuais dos Constituintes para o PPG4

Lâminas	PPG4-2 a PPG4-4	PPG4-1, PPG4-5
Petrofácies	PT-4	PT-4
DBSCAN (Agrupamentos)	GPPG4-0	GPPG4-1
Quartzo	28.066667	49.65
Feldspato	2.433333	4.95
Muscovita	10.233333	6.65
Opaco	3.233333	2.70
Turmalina	0.000000	0.00
Zircão	0.533333	0.70
Rutilo	0.100000	0.35
Glauconita	0.100000	0.00
Clorita	0.000000	0.15
Pseudo Matriz	0.000000	0.00
Litoclasto	0.000000	0.00
Bioclasto	0.000000	0.00
Cresc. Sec. Qtz	2.366667	2.65
Caolinita	5.000000	11.00
Ilita/Smectita	19.100000	8.50
Pirita	10.766667	4.85
Siderita	3.733333	3.45
Cim. Carbonático	0.233333	0.00
Cim. Silicoso	0.000000	0.00
Cim. Ferruginoso	0.200000	0.15
Por. Intergranular	1.800000	5.85
Por. Intragranular	0.000000	0.00

Tabela C.5: Média dos Percentuais dos Constituintes para o PPG5

Lâminas	PPG5-1,PPG5-2,PPG5-4	PPG5-3
Petrofácies	PT-1	PT-3
DBSCAN (Agrupamentos)	GPPG5-0	GPPG5-1
Quartzo	38.633333	38.3
Feldspato	4.433333	1.3
Muscovita	20.200000	7.3
Opaco	1.766667	1.0
Turmalina	0.000000	0.0
Zircão	0.566667	0.0
Rutilo	0.000000	0.3
Glauconita	0.100000	0.0
Clorita	0.233333	0.0
Pseudo Matriz	7.000000	2.0
Litoclasto	0.000000	0.0
Bioclasto	0.000000	0.0
Cresc. Sec. Qtz	2.466667	7.6
Caolinita	7.433333	8.0
Ilita/Smectita	12.033333	6.3
Pirita	0.000000	0.0
Siderita	0.000000	0.0
Cim. Carbonático	0.466667	27.6
Cim. Silicoso	0.233333	0.0
Cim. Ferruginoso	0.466667	0.0
Por. Intergranular	3.533333	0.3
Por. Intragranular	0.000000	0.0