



## APLICAÇÃO DE INTELIGÊNCIA COMPUTACIONAL NA CARACTERIZAÇÃO GEOQUÍMICA DE ÓLEOS

Marcia Val Springer

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Civil, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Civil.

Orientador(es): Débora de Almeida Azevedo

Luiz Landau

Rio de Janeiro  
Dezembro de 2012

APLICAÇÃO DE INTELIGÊNCIA COMPUTACIONAL NA CARACTERIZAÇÃO  
GEOQUÍMICA DE ÓLEOS

Marcia Val Springer

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ  
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)  
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR  
EM CIÊNCIAS EM ENGENHARIA CIVIL.

Examinada por:

---

Prof<sup>a</sup>. Débora de Almeida Azevedo, D.Sc.

---

Prof. Luiz Landau, D.Sc.

---

Dr. Eugênio Vaz dos Santos Neto

---

Prof<sup>a</sup>. Celeste Yara dos Santos Siqueira, D.Sc.

---

Prof. Nelson Francisco Favilla Ebecken

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2012

Springer, Marcia Val

Aplicação de Inteligência Computacional na  
Caracterização Geoquímica de Óleos/ Marcia Val  
Springer. – Rio de Janeiro: UFRJ/COPPE, 2012.

X, 90 p.: il.; 29,7 cm.

Orientador: Débora de Almeida Azevedo

Luiz Landau

Tese (doutorado) – UFRJ/ COPPE/ Programa de  
Engenharia Civil, 2012.

Referências Bibliográficas: p. 83-90.

1. Caracterização Geoquímica. 2. Agrupamento *fuzzy*.  
3. Biomarcadores e Diamantóides. I. Azevedo, Débora de  
Almeida *et al.* II. Universidade Federal do Rio de Janeiro,  
COPPE, Programa de Engenharia Civil. III. Título.

*Dedico esta tese aos meus pais, ao meu querido irmão, a minha querida Tia Marizete,  
ao meu querido esposo Antonio pela paciência do dia a dia e incansável apoio, e  
todas as pessoas que amo.*

"Não existe sorte. Sorte é quando a preparação encontra oportunidade".

(autor desconhecido)

## AGRADECIMENTOS

*A Deus, graças sempre.*

*Aos meus pais, Luiz e Juçara, por tudo sempre.*

*Ao meu esposo, Antonio, por seu amor, com amor sempre. Ele que foi o mais sacrificado pelas minhas frequentes ausências e indisponibilidades. Pelo apoio incondicional que me deu desde sempre.*

*Ao meu querido irmão, Marcus, um dos grandes incentivadores desta jornada.*

*A minha querida Tia Marizete, por seu apoio e carinho incondicional, estando sempre ao meu lado.*

*A minha querida dinda, Regina, e prima, Juju, pelo incentivo.*

*A minha orientadora Débora pela orientação, dedicação, compreensão e confiança prestadas a mim durante todos esses anos. Todo meu crescimento profissional, Mestrado e agora doutorado, foi regido por ela.*

*Ao coordenador do curso Sistemas Petrolíferos Luiz Landau, pelo empenho de oferecer um curso altamente qualificado e por todo apoio fornecido.*

*A todos os professores que compartilharam os conhecimentos e experiências profissionais.*

*A minha querida amiga Tatiana, pelo incentivo, apoio e animação diária, indispensáveis, e por nunca permitir que eu desanimasse.*

*A minha amiga Kaíza, por todo apoio e carinho.*

*Ao amigo Jean, pelo companheirismo e paciência.*

*Aos funcionários do LAMCE e LAB2M, em especial à Mônica, Sérgio, Marcos e Verônica.*

*Aos funcionários do PEC, em especial à Beth e ao Jairo, por seu excelente trabalho, carinho e atenção.*

*Aos meus amigos do IFRJ, toda a equipe de físico-química, e a amiga Regina pela torcida e incentivo.*

*Aos membros da banca.*

*Gostaria de agradecer ao Programa de Formação de Recursos Humanos da Agência Nacional de Petróleo (PRH-02).*

*A todos vocês que moram no “meu coração” um sincero Obrigado.....*

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

## APLICAÇÃO DE INTELIGÊNCIA COMPUTACIONAL NA CARACTERIZAÇÃO GEOQUÍMICA DE ÓLEOS

Marcia Val Springer

Dezembro/2012

Orientadores: Débora de Almeida Azevedo

Luiz Landau

Programa: Engenharia Civil

Durante os últimos anos, em função da demanda da exploração de petróleo, tem-se verificado um crescimento substancial da quantidade e complexidade de dados armazenados na forma digital na área da geoquímica orgânica. Diante deste cenário, busca-se testar o desempenho do método de descoberta do conhecimento no apoio ao agrupamento de amostras de óleos a partir de parâmetros geoquímicos, para estimar a maturação térmica e identificar a origem. Em vista disso, o presente trabalho teve como objetivo propor um método para extração de informações desconhecidas a partir de uma base de dados, contendo parâmetros geoquímicos envolvendo biomarcadores e diamantóides e tendo como resultados a visualização de agrupamentos *fuzzy* de amostras de óleos. Este método foi desenvolvido sobre dados envolvendo bacias sedimentares brasileiras e a bacia sedimentar colombiana, *Llanos Orientales*. Segundo esse método, as amostras foram agrupadas em cinco grupos diferentes quanto ao parâmetro de maturação térmica. Na caracterização do ambiente deposicional, observou-se que as amostras puderam ser divididas em quatro grupos principais: grupo 1 possui uma contribuição marinha; grupo 2 indicativos de uma mistura (marinho e terrestre); grupo 3 possui características de uma mistura contendo, também, uma contribuição marinha e grupo 4 possui características terrestres.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

APPLICATION OF COMPUTATIONAL INTELLIGENCE IN GEOCHEMICAL  
CHARACTERIZATION OF OILS

Marcia Val Springer

December/2012

Advisors: Débora de Almeida Azevedo

Luiz Landau

Department: Civil Engineering

Organic Geochemistry is a powerful tool to aid oil exploration. During recent years, due to the demand of oil, there has been a substantial increase in the amount and complexity of data stored in digital form. Given this scenario, we seek to test the performance of the method of knowledge discovery in supporting the grouping of oil samples from geochemical parameters to evaluate thermal maturity and identify source. Therefore, the present work aimed to propose a method to extract unknown information from a database, with the results to visualize fuzzy clusters of oil samples from geochemical parameters, involving biomarkers and diamondoid. This method was developed on data using Brazilian sedimentary basins and the Colombian sedimentary basin, named *Llanos Orientales*. Under this method, the samples were grouped into five different groups as the parameter of thermal maturation. In the characterization of the depositional environment, it was observed that the samples could be divided into four main groups: group 1 has a marine contribution, group 2 indicating a mixture (marine and terrestrial), group 3 has characteristics of a mixture containing a marine contribution, and group 4 has terrestrial features.

# Índice

<b>Capítulo1</b> .....	<b>1</b>
1.1) Introdução e Motivação .....	1
1.2) Objetivo.....	6
<b>Capítulo 2: Área de Estudo</b> .....	<b>7</b>
<b>Capítulo3: Parâmetros Geoquímicos</b> .....	<b>14</b>
<b>Capítulo4: Fundamentação Teórica</b> .....	<b>26</b>
4.1) Análise de agrupamento – “ <i>Clustering</i> ” .....	26
4.2) Método de Agrupamento <i>Fuzzy</i> .....	29
4.2.1) Método de Clusterização <i>Fuzzy c-means (FCM)</i> .....	32
4.2.2) Ajuste dos parâmetros .....	36
<b>Capítulo5: Materiais e Métodos</b> .....	<b>40</b>
5.1) Procedimentos Analíticos .....	40
5.1.1) Técnicas Analíticas Utilizadas.....	42
5.2) Avaliação Estatística dos dados .....	44
<b>Capítulo6: Base de Dados - Descrição</b> .....	<b>47</b>
<b>Capítulo7: Resultados e Discussão</b> .....	<b>50</b>
7.1) Parâmetros de Maturação .....	50
7.1.1) Interpretação dos resultados pela caracterização de Biomarcadores e Diamantóides .....	50
7.1.2) Análise de agrupamento FCM – <i>fuzzy c-means</i> .....	60
7.2) Parâmetros de Ambientes Depositionais (Origem).....	66
7.2.1) Interpretação dos resultados pela caracterização de Biomarcadores e Diamantóides .....	66
7.2.2) Análise de agrupamento FCM – <i>fuzzy c-means</i> .....	73
<b>Capítulo 8: Conclusão</b> .....	<b>80</b>
<b>Capítulo 9: Recomendações</b> .....	<b>82</b>
<b>Capítulo 10: Referências Bibliográficas</b> .....	<b>83</b>

# Capítulo 1.

## 1.1) Introdução e Motivação

Atualmente as organizações têm-se mostrado extremamente eficientes em obter, organizar e armazenar grandes quantidades de dados, obtidos de suas operações diárias ou pesquisas científicas, porém a maioria não usa adequadamente essa gigantesca quantidade de dados para transformá-la em conhecimentos e que poderiam ser utilizados em suas próprias atividades, sejam elas comerciais ou científicas.

Para explorar e extrair informações úteis de um banco de dados foi proposto, em 1989, o processo de *KDD (Knowledge Discovery Database)*. Este processo refere-se às etapas que produzem conhecimento a partir de dados e, principalmente, à etapa de mineração de dados, que é a fase que transforma dados em informação (FAYYAD *et al.*, 1996). O conceito de *Data-Mining* (mineração de dados) está se tornando cada vez mais popular como uma ferramenta de descoberta de informações, que podem revelar estruturas de conhecimentos, que possam guiar decisões em condições de certeza limitada. Este processo de obter informações a partir de um banco de dados não é uma tarefa trivial, pois envolve algumas etapas com muitas decisões realizadas pelo analista (Figura 1), como: Criação de um conjunto de dados alvos (SELEÇÃO); Pré-processamento e limpeza dos dados; Redução de dados e projeção (TRANSFORMAÇÃO); Busca por padrões e/ou associações (MINERAÇÃO DE DADOS); Interpretação dos padrões minerados.

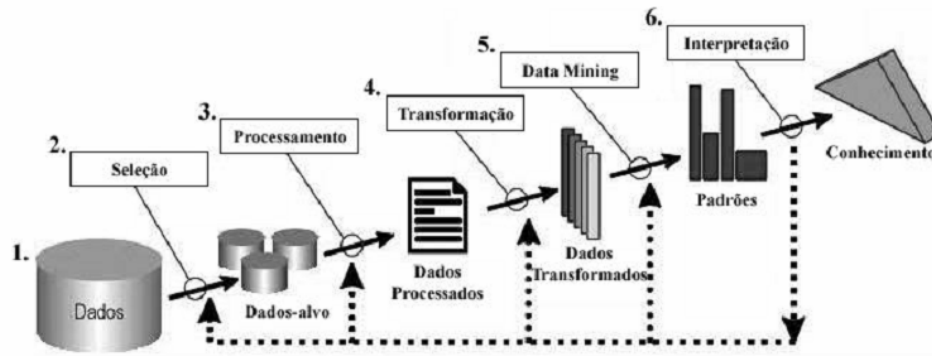


Figura 1: Processo de extração de conhecimento em bases de dados e suas etapas (adaptado FAYAAD *et al.* 1996)

A seleção de dados tem como objetivo identificar as origens internas e externas da informação, extraíndo um subconjunto de dados necessário para a aplicação da mineração de dados, selecionando apenas atributos relevantes aos objetivos do processo de extração de conhecimento (MACEDO *et al.*, 2010).

A etapa de processamento dos dados visa assegurar a qualidade dos dados selecionados. Como o resultado do processo de extração possivelmente será utilizado no processo de tomada de decisão, a qualidade dos dados é um fator extremamente importante. Esta fase inicia-se com uma revisão geral da estrutura dos dados e a definição de medidas de qualidade, utilizando uma combinação de métodos estatísticos e técnicas de visualização de dados. A seguir encontram-se descritos os problemas tratados na etapa de processamento dos dados, segundo BOENTE & GOLDSCHMIDT & ESTRELA (2008):

- Eliminação de dados duplicados - são removidos dados duplicados e/ou corrompidos;
- Tratamento de *outliers*- são valores significativamente fora do esperado para uma variável;
- Valores faltantes - valores que não estão presentes no conjunto selecionado e valores inválidos que foram eliminados durante a detecção de *outliers*.

A etapa de transformação de dados tem como objetivo converter o conjunto bruto de dados em uma forma padrão de uso, tornando os dados úteis para a mineração (BOENTE, OLIVEIRA & ROSA, 2007). Devido às restrições de espaço em memória e tempo de processamento, o número de atributos disponíveis para análise pode inviabilizar a mineração de dados. Com isso, técnicas de redução de dados são aplicadas, sendo feitas de três modos:

- Redução do número de exemplos – deve ser feita mantendo as características do conjunto de dados original, por meio da geração de amostras representativas dos dados;
- Redução do número de atributos – realizada pelo especialista do domínio, seleciona-se um subconjunto dos atributos de forma que não tenha impacto na qualidade do conjunto final;
- Redução do número de valores de um atributo – consiste na redução do número de valores de um atributo, geralmente utilizando-se técnicas de discretização e suavização de valores. A discretização de um atributo consiste na substituição de um atributo contínuo por um atributo discreto, por meio do agrupamento de seus valores. A suavização de valores agrupa determinados atributos em um valor numérico que o represente, podendo ser, por exemplo, a média ou a mediana.

Essa etapa auxilia na redução do tempo de processamento para a técnica de mineração, diminuindo o espaço de busca. As transformações devem, entretanto, preservar ao máximo, nas amostras geradas, as informações presentes nos dados brutos.

A mineração de dados é a etapa mais importante do processo de *KDD* e caracteriza-se pela existência de uma técnica de mineração capaz de extrair conhecimento implícito de um banco de dados em função de um objetivo proposto (JACKSON, 2002).

As técnicas de mineração de dados podem ter gerado uma quantidade enorme de padrões, dos quais podem não ser relevantes ou interessantes ao usuário. Um dos objetivos principais do processo de *KDD* é que o usuário possa compreender e utilizar o conhecimento descoberto (FAYYAD *et al.*, 1996). Mas, podem ocorrer casos em que os modelos são muito complexos ou não fazem sentido para os especialistas. Existem algumas formas de se caracterizar a qualidade dos resultados obtidos, como a “*compreensibilidade*” e a “*interessabilidade*”. A “*compreensibilidade*” de um conjunto de regras relaciona-se com a facilidade de interpretação dessas regras por um ser humano. Assim, quanto menor a quantidade de regras de um dado modelo e menor o número de condições por regra, maior será a *compreensibilidade* das regras descobertas. A “*interessabilidade*” é uma medida de qualidade que tenta estimar o quanto de conhecimento interessante ou inesperado existe, e deve combinar fatores numa medida que reflète como o especialista julga o padrão.

Após a análise do conhecimento, caso esse não seja de interesse do usuário final ou não cumpra os objetivos propostos, o processo de extração pode ser repetido, ajustando-se os parâmetros ou melhorando o processo de escolha dos dados, para a obtenção de melhores resultados em uma próxima iteração.

Diferentes modelos podem ser aplicados aos bancos de dados para dar o suporte necessário aos distintos problemas nas mais diversas áreas. Estes modelos são divididos de acordo com a natureza da atividade, e os principais são: predição, agrupamento (*Clustering*) e classificação. No capítulo 4 serão apresentadas em maiores detalhes as técnicas de análise de *cluster* (agrupamento), que é o foco principal deste trabalho.

Para a obtenção destes modelos, diferentes técnicas computacionais baseadas em paradigmas da inteligência computacional têm sido empregadas na indústria do petróleo. Dentre estas técnicas estão as Redes Neurais, Lógica *Fuzzy*, Regras de Associação, Árvores de decisão, Algoritmos Genéticos, entre outros.

A construção de modelos determinísticos e de interpretação tem sido crescentemente substituída por estes métodos computacionais. A diversidade destas aplicações de mineração de dados usadas nos problemas de campos de óleo e gás e a

aceitação destas metodologias têm se manifestado em grande interesse principalmente por parte de engenheiros e cientistas em todo o mundo.

Durante os últimos anos, em função não apenas da demanda da exploração de petróleo, mas também do desenvolvimento das técnicas analíticas, tem-se verificado um crescimento substancial da quantidade e complexidade de dados armazenados na forma digital na geoquímica orgânica. Este grande volume de dados acaba sendo inviável de ser analisado por especialistas através de métodos convencionais, tais como planilhas de cálculo e relatórios informativos operacionais, onde o especialista testa sua hipótese contra a base de dados. Estatísticas tradicionais são limitadas se levarmos em conta que a análise se torna bastante trabalhosa quando o número de variáveis a serem investigadas cresce seguidamente. Tais métodos possuem condições que limitam o número de casos a utilizar, fazendo com que apenas uma pequena parte do universo esteja disponível para a análise e, quando os relacionamentos dos dados não são lineares, torna-se difícil empregar os métodos estatísticos tradicionais.

Na Geoquímica Orgânica, as pesquisas envolvendo Inteligência Artificial têm sua atuação de forma pontual. De um modo geral, as pesquisas baseadas em grandes volumes de dados na área de Geoquímica Orgânica vêm sendo norteadas a partir de uma perspectiva descritiva e preditiva, onde a estatística é a ferramenta.

Diante deste cenário, busca-se testar o desempenho do método de descoberta do conhecimento no apoio ao agrupamento de amostras de óleos a partir de parâmetros geoquímicos, como maturação e ambiente deposicional (origem).

A análise de agrupamento tem por objetivo agrupar dados semelhantes em conjunto para que possam ser generalizadas como um grupo. Há muitos algoritmos para análise de *cluster*, com base em estatísticas, redes neurais e dos conjuntos *fuzzy*. Em vista disso, o presente trabalho tem como proposta a introdução da técnica de Mineração de Dados baseados em Sistemas *Fuzzy*. O processo de *clusterização* (agrupamento) escolhido para desenvolver o trabalho foi o algoritmo *Fuzzy C-means (FCM)*.

## 1.2) Objetivo

No presente trabalho, foram analisadas amostras de óleos provenientes de bacias sedimentares brasileiras e da bacia sedimentar colombiana *Llanos Orientales*.

O principal objetivo deste estudo é propor um método para extração de informações desconhecidas a partir de uma base de dados, tendo como resultado a visualização de agrupamentos *fuzzy* de amostras de óleo a partir de parâmetros geoquímicos para avaliar a maturação térmica e identificar a origem.

## Capítulo 2.

### Área de estudo

O banco de dados para a realização deste estudo apresenta parâmetros geoquímicos de maturação e ambiente deposicional para amostras provenientes de bacias sedimentares da margem continental brasileira e da bacia sedimentar colombiana *Llanos Orientales*. As amostras de óleo provenientes de bacias sedimentares brasileiras, exploradas neste presente trabalho, foram caracterizadas geoquimicamente nos trabalhos: JESUINO (2005), TAMANQUEIRA (2006), CALDAS (2009) e SILVA (2010). Já as amostras de óleo provenientes da bacia colombiana foram caracterizadas previamente no trabalho desenvolvido por SPRINGER (2007). A seguir, será apresentada uma breve descrição das áreas de estudo envolvidas.

#### **- Bacias sedimentares brasileiras da margem continental**

O Brasil é um país com uma das maiores extensões de margem continental do mundo, englobando diversos segmentos com bacias sedimentares com características geológicas distintas e diferentes graus de conhecimento do potencial exploratório, figura 2 (MELLO, 1988; MELLO *et al.*, 1988a, b). A origem das bacias sedimentares da margem continental brasileira está diretamente relacionada ao processo de ruptura e separação das placas tectônicas africana e sul-americana, figura 3 (BIZZI *et al.*, 2003).

A margem divergente da América do Sul, estendendo-se por mais de 12.000 km desde o Delta do Orinoco no Oriente Venezuelano até a Terra do Fogo, no extremo sul da Argentina, inclui um sistema contínuo de bacias sedimentares originadas pelos mecanismos de distensão litosférica que, a partir do Mesozóico, conduziram à ruptura do paleocontinente Gondwana, e à separação definitiva das placas Africana e Sul-Americana, acompanhando a formação do Oceano Atlântico Sul.

Considerando-se a natureza e a orientação dos campos de tensões regionais durante a fase de “*rifteamento*” e a dinâmica das placas Africana e Sul-Americana

durante a fase de deriva continental, três domínios distintos podem ser reconhecidos ao longo da margem: uma região predominantemente distensiva, entre o sul da Argentina e o extremo nordeste da costa brasileira; um segmento de natureza transformante, correspondente ao Atlântico Equatorial; e a região ao norte da Foz do Amazonas, onde novamente operaram processos de caráter predominantemente distensional (MILANI & THOMAZ FILHO, 2000).

No domínio distensivo meridional, o estilo estrutural durante a fase *rift* foi marcado pela atividade de falhas normais orientadas principalmente na direção paralela à costa. Este estilo pode ser observado em cada uma das bacias individuais ao longo da margem. Falhas de transferência ocorrem a altos ângulos em relação à direção regional de desenvolvimento das falhas normais. O conjunto de falhas normais evoluiu e acabou por definir a orientação regional de abertura da margem passiva, e as zonas de transferência do *rift* nuclearam as grandes falhas transformantes do assoalho oceânico, cuja projeção contra a borda do continente se dá em ângulo reto (MILANI *et al.*, 2000).

O estilo tectônico da margem adjacente ao Atlântico Equatorial foi diferente. Cisalhamento dextrógiro foi o mecanismo responsável pela ruptura crustal, originando um padrão de falhas oblíquas subverticais que controlaram o *rifteamento* e evoluíram para grandes zonas de fratura oceânicas paralelas à costa entre as quais se incluem Fernando de Noronha, Chain, Romanche e São Paulo.

Uma terceira região na margem divergente da América do Sul corresponde ao trecho situado ao norte da Foz do Amazonas. Tal domínio faz parte de outro contexto distensivo, o do Oceano Atlântico Central, mais antigo que o restante da margem uma vez que lá se encontram documentados processos de ruptura ativos já no Triássico.

Os vários setores individuais ou “bacias” ao longo da extensa e contínua margem divergente do continente compartilham algumas características comuns, como por exemplo, os clássicos estágios tectono-sedimentares evolutivos que incluem o *rift*, a fase transicional e a marinha aberta (MILANI *et al.*, 2000).

Como resultado de sua história evolutiva, o registro sedimentar dessas bacias pode ser subdividido em quatro grandes seqüências: *pré-rift*, *rift*, transicional, e *drift*.

A seqüência *pré-rift*, depositada do final do Jurássico ao início do Cretáceo, é caracterizada por sedimentos continentais, principalmente lacustres, fluviais e deltáicos. A seqüência *rift*, de idade Neocomiana a Barremiana, se depositou na fase de extensão crustal e é caracterizada pelo predomínio de sedimentos lacustres de água doce/salobra a salina, e secundariamente, por depósitos flúvio-deltáicos. A seqüência transicional, de idade Aptiana, depositou-se entre o final da fase *rift* e o início do estágio *drift*, sendo caracterizada por depósitos fluviais, deltáicos e lacustres com influência marinha e alta salinidade. Finalmente, a seqüência *drift* depositou-se do Albiano ao presente, em resposta ao processo de separação entre as placas africana e sul-americana e formação do oceano Atlântico, sendo que seus depósitos são essencialmente marinhos (BIZZI *et al.*, 2003).



Figura 2: Mapa de localização das bacias sedimentares da margem continental brasileira (modificado de MELLO *et al.*, 1988b).

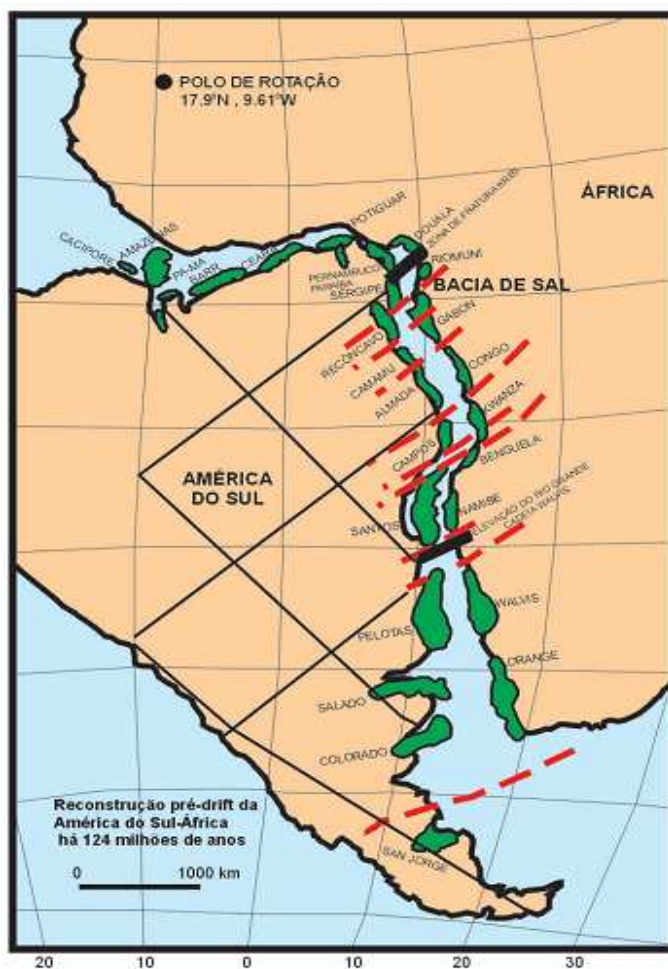


Figura 3: Mapa com a distribuição das bacias marginais dos litorais brasileiro e africano, apresentando reconstrução tectônica das bacias sedimentares na configuração pré-deriva continental, ajustado a reconstrução paleográfica de 124Ma. (BIZZI *et al.*, 2003).

## - Bacia sedimentar colombiana, *Llanos Orientales*

A bacia sedimentar colombiana *Llanos Orientales* está localizada no final norte de uma longa série de bacias subandinas, estendidas desde a Argentina até a Venezuela (Figura 4). A maior parte da bacia está localizada na Colômbia, cobrindo uma área de aproximadamente 200.000 Km<sup>2</sup>, com uma pequena porção no sudeste da Venezuela. A bacia dos *Llanos Orientales* é cercada ao oeste pela Cordilheira Oriental; ao norte pela bacia Apure e pela Cordilheira de Mérida na Venezuela; a leste pelo escudo cratônico Guiana e ao sul pela Bacia Amazônica. A bacia dos Llanos Orientales, situada na Colômbia, tem sua história geológica ligada ao desenvolvimento da parte norte dos Andes. Embora a formação desta bacia e o preenchimento inicial com sedimentos tenham sido provocados por mecanismos de subsidência tectônica e termal associados com eventos de extensão litosférica (*rift* intracratônico e *back-arc*), a geometria atual e os processos de carga do petróleo estão associados a mecanismos compressivos que provocaram inversão tectônica e o soerguimento dos Andes no Neógeno.

Esses últimos eventos criaram novas estruturas ou transformaram drasticamente aquelas mais antigas, modificando os regimes termal e de pressões e tendo importantes implicações no funcionamento dos sistemas petrolíferos, principalmente na distribuição e qualidade das acumulações (BAUTISTA, 2008).

A última seção do Cretáceo do nordeste da América Latina contém excelentes rochas geradoras, que tem originado a maioria dos óleos encontrados na Colômbia, Equador e Venezuela. Os sedimentos ricos em matéria orgânica da formação *La Luna* são descritos como de fontes marinhas, algálicas e carbonáticas (ZUMBERGE, 1983). Entretanto, reconstruções estratigráficas e paleogeográficas mostram que este esquema é um pouco simplista. A complexa evolução tectônica do nordeste da América do Sul resulta em uma grande variedade de ambientes deposicionais e condições paleogeográficas, que controlam o tipo e a quantidade de matéria orgânica preservada. Essas mudanças nas características da rocha geradora resultam em diferentes tipos de óleos (BAUTISTA, 2008).

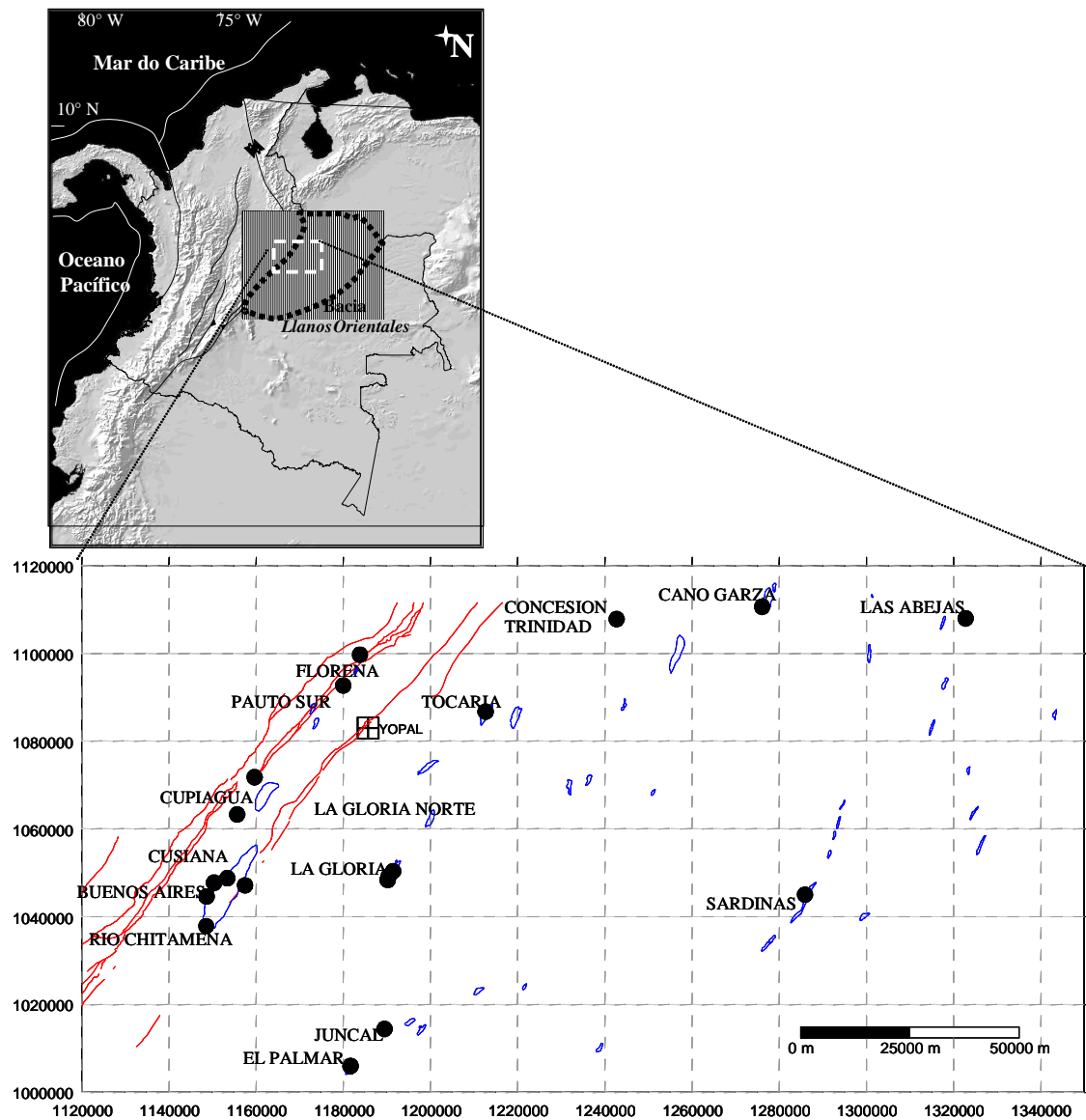


Figura 4: Mapa de localização da bacia sedimentar colombiana *Llanos Orientales*, apresentando os campos explorados (●).

## **Capítulo 3.**

### **Parâmetros Geoquímicos**

O petróleo é uma mistura complexa de produtos metaestáveis, predominantemente de hidrocarbonetos e outras substâncias orgânicas sulfuradas, nitrogenadas e oxigenadas (TISSOT & WELTE, 1984), ou seja, apresenta quatro grupos de compostos: hidrocarbonetos saturados, hidrocarbonetos aromáticos, resinas e asfaltenos. É uma substância oleosa, inflamável, menos densa que a água, com cheiro característico e de cor variável.

A indústria petrolífera foi gradualmente percebendo, ao longo de décadas de exploração, que para se encontrar jazidas de hidrocarbonetos de volume significativo era imperioso que um determinado número de requisitos geológicos ocorresse simultaneamente nas bacias sedimentares. O estudo destas características de maneira integrada e a simulação preliminar das condições ótimas para sua existência concomitante, com o objetivo de permitir a diminuição do risco exploratório envolvido nas perfurações de poços, um item de elevado custo, foram consolidados em um único conceito: sistema petrolífero. O conceito de sistema petrolífero agrupa os diversos elementos que controlam a existência de jazidas de petróleo numa bacia sedimentar. Tal conceito, visualizado numa escala global, parece justificar de maneira adequada as diversas províncias petrolíferas conhecidas.

Um sistema petrolífero ativo compreende a existência e o funcionamento síncronos de quatro elementos (rochas geradoras maduras, rochas-reservatório, rochas selantes e trapas) e dois fenômenos geológicos dependentes do tempo (migração e sincronismo) (MILANI *et al*, 2000).

O petróleo tem origem a partir da matéria orgânica depositada nas bacias sedimentares. Com o progressivo soterramento e aumento da temperatura ao longo do tempo geológico, a matéria orgânica passa por uma série de transformações físico-químicas que podem ser agrupadas de acordo com os seguintes estágios evolutivos: diagênese, catagênese e metagênese (TISSOT & WELTE, 1984; PETERS & MOLDOWAN, 1993; HORSFIELD & RULLKOTER, 1994).

A diagênese é o processo inicial de alteração física, química e biológica da matéria orgânica sob condições de soterramento incipiente e baixas temperaturas. Ao final dessa etapa, a biomassa de origem sedimentar se transforma em querogênio, constituinte orgânico das rochas sedimentares, insolúvel em solventes orgânicos (TISSOT & WELTE, 1984). Com o aumento da pressão e temperatura (soterramento progressivo), o querogênio é degradado para a formação do petróleo, iniciando o estágio denominado catagênese, e a matéria orgânica passa por uma série de transformações químicas, tais como reações de isomerização, aromatização e craqueamento, que resultam na conversão do querogênio em óleo. Sob condições de soterramento ainda maiores, chega-se ao estágio denominado metagênese, onde mudanças mais severas no material orgânico ocasionam a formação de metano e restando apenas resíduos carbonosos.

O grupo dos hidrocarbonetos saturados constitui o grupo mais abundante e é a classe mais investigada, pois é onde se encontram os biomarcadores (terpanos e esteranos) e os diamantóides.

A concentração de hidrocarbonetos saturados no petróleo depende das suas geradoras que por sua vez estão relacionadas ao tipo de matéria orgânica e seu respectivo ambiente deposicional. Óleos mais pesados e óleos originados de matéria orgânica de origem continental possuem maior proporção destes compostos, enquanto que óleos derivados de matéria orgânica de origem marinha, ou uma mistura destes dois tipos de matéria orgânica geram um petróleo mais rico em compostos cíclicos (TISSOT & WELTE, 1984).

A caracterização dos sistemas petrolíferos, através de parâmetros de geoquímica orgânica, é fundamental para o desenvolvimento de modelos preditivos para a exploração de petróleo nas bacias sedimentares. A maturação térmica e o tipo da matéria orgânica constituem as principais variáveis que controlam a geração de hidrocarbonetos nas bacias sedimentares. São vários os parâmetros geoquímicos utilizados para avaliar o nível de evolução térmica e para caracterizar o ambiente deposicional das rochas geradoras (PETERS & MOLDOWAN, 1993); os mais

utilizados se baseiam nas reações de isomerização de biomarcadores, particularmente das famílias dos esteranos e terpanos, e na concentração de diamantóides.

Os biomarcadores são compostos orgânicos complexos que apresentam átomos de carbono e hidrogênio, e são encontrados em rochas e sedimentos. São “fósseis moleculares” derivados de substâncias presentes em organismos vivos, que mostram pequenas mudanças na estrutura herdada das moléculas orgânicas originais (ZUNBERGE, 1987a; PETERS & MOLDOWAN, 1993). Durante a diagênese, as moléculas dos precursores, os triterpenóides e esteróides, que contêm um ou mais átomos de oxigênio e que frequentemente apresentam ligações duplas, passam por reações químicas nas quais perdem os átomos de oxigênio. Suas ligações duplas são reduzidas para produzir o biomarcador saturado mais estável, preservado nas amostras geológicas. Estes compostos representam uma pequena fração do petróleo, porém são de grande importância, pois permitem a correlação entre óleo e a rocha geradora, a avaliação do grau de evolução térmica e do nível de biodegradação que o óleo foi submetido (WAPLES & MACHIHARA, 1991).

Os parâmetros de biomarcadores têm sido amplamente utilizados para a caracterização de ambientes deposicionais das rochas geradoras de petróleo. Vários autores descrevem a distribuição dos biomarcadores como uma ferramenta útil para distinguir rochas geradoras depositadas em diferentes ambientes. As razões de biomarcadores, quando comparadas com outros parâmetros, são úteis na inferência do tipo de rocha geradora mesmo quando apenas amostras de óleos estão disponíveis (PETERS *et al.*, 2005). Os organismos que apresentam os precursores biológicos dos biomarcadores vivem em determinadas condições ambientais. Logo, os biomarcadores podem ser utilizados como indicadores dessas condições de vida (WALPES & MACHIHARA, 1991).

A seguir, estão listados os parâmetros mais utilizados na caracterização de ambientes deposicionais:

- Razão Pristano/Fitano

Os biomarcadores Pristano e Fitano são os compostos mais conhecidos e mais abundantes em sedimentos e óleos. A fonte mais comum desses isoprenóides é a clorofila A, presentes em organismos fototróficos. A cadeia fitil da clorofila é quebrada em condições anóxicas ou redutoras em sedimentos, formando o fitol, que sofre redução à dihidrofitol e chega a fitano. Em condições óxicas o fitol também sofre oxidação à ácido fitênico, que em seguida descarboxila gerando o pristeno e por redução forma o pristano, Figura 5 (BROOKS *et al.*, 1969; PETERS *et al.*, 2005).

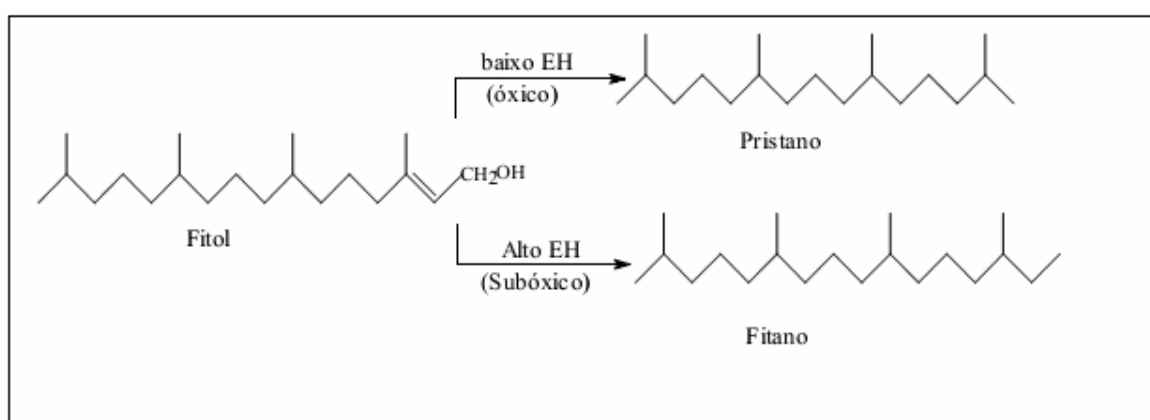


Figura 5: Origem diagenética do pristano e fitano a partir do fitol (PETERS *et al.*, 2005).

DIDYK *et al.* (1978) propuseram uma relação direta entre a razão pristano/fitano (Pr/Fi) e oxigenação do ambiente deposicional. Valores de razão Pr/Fi < 1 indicariam deposição anóxicas, particularmente quando acompanhadas pela alta quantidade de porfirina e de enxofre. Já a razão Pr/Fi > 1 indicaria ambiente deposicional óxico. Em estudos de bacias sedimentares da margem continental brasileira, alguns autores propuseram que em ambientes de água doce e salinos (lacustre de água doce, salino e marinho aberto) ocorre a predominância do pristano, com razão pristano/fitano maior que 1. Já em ambientes marinhos carbonáticos e hipersalinos, ocorre a predominância de fitano, com razão pristano/fitano menor que 1 (MELLO *et al.*, 1988a, 1988b; MELLO & MAXWELL, 1990). Segundo PETERS *et al.* (2005), os valores desta razão normalmente se encontram na faixa entre 0,8 e 3,0, sendo que para ambientes óxicos característicos de matéria orgânica terrestre a razão pristano/fitano é maior que 3,0, enquanto que para ambientes anóxicos, comumente hipersalinos ou carbonáticos, a razão pristano/fitano é menor que 0,8.

- Razão Hopano/Esteranos

Um dos principais parâmetros de biomarcadores para análise de origem é a razão entre os hopanos/esteranos. Esta razão foi obtida abrangendo os esteranos C<sub>27</sub>-C<sub>29</sub> ααα (20S+20R) + esteranos C<sub>27</sub>-C<sub>29</sub> αββ(20S+20R) e os hopanos αββC<sub>29</sub>, C<sub>30</sub>, C<sub>31</sub>-C<sub>33</sub> (22S+22R). Esta razão mostra a maior contribuição de organismos eucarióticos (especialmente algas) ou contribuição de organismos procarióticos (bactérias). Em geral, altas concentrações de esteranos e baixos valores da razão hopanos/esteranos (menor ou igual a 4) indicam deposição de matéria orgânica marinha com maior contribuição de organismos planctônicos e/ou algas. Diferentemente, baixas concentrações de esteranos e altos valores de razão hopano/esterano (maior que 7) indicam deposição de matéria orgânica continental e/ou microbiana retrabalhada (PETERS *et al.*, 2005).

- Razão TPP/DIA

Outro importante parâmetro de origem é a razão dos poliprenóides tetracíclicos (TPPs) / Diasteranos C<sub>27</sub> 13β(H),17α(H) (20S+20R). HOLBA *et al.* (2000, 2003) mostraram que indicadores específicos, os tetracíclicos poliprenóides (TPP), permitem uma melhor diferenciação entre os ambientes lacustres e marinhos. A caracterização desses ambientes é mais complexa porque eles possuem uma enorme variedade de condições deposicionais. Estes autores mostraram que a abundância destes compostos no petróleo reflete a natureza do ambiente. Altas concentrações relativas de TPP's em óleos e rochas geradoras em relação aos diasteranos indicam deposição em ambientes marinhos.

- Razão Diasteranos/Esteranos

A razão diasteranos/esteranos é dada por [(C<sub>27</sub> 13β, 17α(H) 20S + 20R)/(C<sub>27</sub> 5α,14α,17α(H) 20S + 20R)]. Os diasteranos estão presentes em quantidades significativas em óleos que apresentam alta maturação térmica (WAPLES & MACHIHARA, 1991). São formados facilmente em sedimentos clásticos, logo a razão diasteranos/esteranos é frequentemente usada para diferenciar rochas geradoras carbonáticas de clásticas (WALPES & MACHIHARA, 1991; PETERS *et al.*, 2005).

Baixos valores da razão diasteranos/esteranos indicam matéria orgânica anóxica pobre em material argiloso, rocha carbonática ou hipersalina. Já altos valores, são típicos de rochas geradoras ricas em argila como, por exemplo, ambientes lacustres e marinho deltaico (PETERS *et al.*, 2005). Porém, alguns estudos mostraram o oposto a esta afirmação, uma vez que altos valores da razão diasteranos/esteranos foram encontrados em extratos de rochas carbonáticas, MOLDOWAN *et al.* (1991), e em rochas geradoras da Flórida pobres em argila (PALACAS *et al.*, 1984).

MOLDOWAN *et al.* (1986), SEIFERT & MOLDOWAN *et al.* (1978) sugeriram que além do ambiente deposicional, esta razão é influenciada pelo potencial de oxi-redução, pela maturação e/ou biodegradação. Os diasteranos são mais estáveis que os esteranos e se tornam mais abundantes com o aumento da maturação. A intensa biodegradação pode resultar na destruição seletiva dos esteranos e alterar a interpretação do resultado. Assim, esta razão é útil para diferenciar condições deposicionais quando as amostras apresentam níveis de maturação e de biodegradação similares (PETERS *et al.*, 2005).

- Razão Terpanos Tricíclicos/17 $\alpha$ (H), 21 $\beta$ (H)-hopano

Os terpanos tricíclicos foram propostos inicialmente como compostos capazes de avaliar o grau de evolução térmica de rochas e óleos (SEIFERT & MOLDOWAN, 1978). Posteriormente, foram utilizados para caracterizar ambientes deposicionais. MELLO *et al.* (1988 a; 1988 b) sugeriram que a principal importância dos tricíclicos se encontra na sua abundância e não tanto no seu padrão de distribuição.

- Razão terpano tetracíclico/17 $\alpha$ (H), 21 $\beta$ (H)-hopano

O terpano tetracíclico C<sub>24</sub> tem sido encontrado em óleos e extratos de inúmeras rochas evaporíticas e carbonáticas (AQUINO NETO *et al.*, 1983; PALACAS *et al.*, 1984; PETERS *et al.*, 2005). Porém, alguns estudos em óleos australianos de origem terrestre mostram a presença deste composto.

Esta razão é bastante utilizada na caracterização de ambientes deposicionais, e de acordo com alguns estudos, as maiores abundâncias do terpano tetracíclico C<sub>24</sub> foram encontradas em ambientes marinho deltaicos e lacustres salinos (MELLO, 1988). No entanto, este autor afirma que na ausência de outras evidências a alta abundância relativa do terpano tetracíclico C<sub>24</sub> indica abundante deposição de matéria orgânica terrestre.

- Razão Ts/Tm

SEIFERT & MOLDOWAN. (1980), mostraram que esta razão só era afetada pela maturação térmica. Porém, alguns estudos posteriores mostraram a correlação de *Ts* e *Tm* com condições diagenéticas. Apesar da relação entre *Ts/Tm* e a litologia ainda não ser bastante clara, vários estudos já mostraram que os valores de *Ts/Tm* são menores em ambientes hipersalinos e maiores em carbonáticos (RULLKOTTER & MARZI, 1988; WAPLES & MACHIHARA, 1991). ROBINSON, (1987), mostrou também que em óleos de origem terrestre os valores de *Tm* são maiores e nos óleos de origem lacustre são baixos.

O avanço das transformações químicas sofridas pelo querogênio e o petróleo sob condições crescentes de soterramento e temperatura resultam na formação de hidrocarbonetos mais leves que podem ser monitorados por uma série de indicadores geoquímicos, conhecidos como parâmetros de maturação térmica (TISSOT & WELTE, 1984).

Distribuições de biomarcadores são amplamente utilizados na avaliação da maturidade do óleo (PETERS & MOLDOWAN, 1993). Muitos dos parâmetros de maturação, envolvendo biomarcadores, atingem o equilíbrio antes da janela de óleo, e em alguns casos, mostram inversão em níveis elevados de maturidade. Consequentemente, estas proporções não são eficazes como indicadores de maturação para óleos de alta maturidade (FARRIMOND *et al.*, 1998).

Os parâmetros de maturação térmica mais usados são:

- $22S/(22S+22R)$ ; razão de isomerização dos homohopanos

Isomerização no  $C_{22}$  dos  $17\alpha(H)$ -hopanos  $C_{31}$  a  $C_{35}$  ocorre antes que muitas razões de biomarcadores usados para estimar a maturação térmica de óleos e sedimentos. O precursor biológico possui a configuração 22R que é gradualmente convertida em uma mistura de diastereoisômeros 22R e 22S. As proporções de 22R e 22S podem ser calculadas para todos ou apenas um homohopano. Porém, como os homólogos  $C_{33}$ ,  $C_{34}$  e  $C_{35}$  podem apresentar problemas de coeluição de picos, os mais utilizados para calcular esta razão são os homólogos  $C_{31}$  e  $C_{32}$  (PETERS *et al.*, 2005).

A razão  $22S/(22S+22R)$  varia de 0 a 0,6 (valor de equilíbrio 0,57 a 0,62) durante o processo de maturação térmica. Alguns estudos mostram que essa razão sofre influência de alguns fatores, como por exemplo, a litologia da bacia. Portanto, essa razão deve ser utilizada com cautela (PETERS *et al.*, 2005).

- Terpanos Tricíclicos/ $17\alpha(H)$ -hopanos

Esta razão aumenta com o aumento da maturação térmica. Isto ocorre porque mais tricíclicos são liberados do querogênio em altos níveis de maturação. A origem dos terpanos tricíclicos pode ser o resultado da diagênese de vários precursores biológicos; logo esta razão pode variar consideravelmente entre óleos de diferentes rochas geradoras ou diferentes fácies orgânicas da mesma rocha geradora (PETERS *et al.*, 2005).

- $Ts/(Ts+Tm)$

Durante a catagênese,  $C_{27}17\alpha(H)$ -trisorhopano (Tm ou  $17\alpha(H)$ -22,29,30-trisorhopano) apresenta menor estabilidade do que o  $C_{27}18\alpha(H)$ -trisorhopano II (Ts ou  $18\alpha(H)$ -22,29,30-trisorneohopano).

Esta razão é deveras dependente quanto à maturidade e quanto à origem. Por isso, é um bom indicador de maturação quando utilizada para avaliar óleos de mesma origem com fácies orgânicas compatíveis (PETERS *et al.*, 2005).

- $20S/(20S+20R)$ , razão de isomerização dos esteranos

Devido à facilidade na análise usando os cromatogramas de massa do íon  $m/z$  217, as razões de isomerização dos esteranos são calculadas nos compostos  $C_{29}$  (24-etilcolestanos, estigmastanos ou sistostanos), visto que os compostos  $C_{27}$  e  $C_{28}$  apresentam interferência de coeluição de picos (PETERS *et al.*, 2005).

À medida que aumenta a maturação térmica, a configuração R no  $C_{20}$ , existente nos precursores esteroidais presentes nos organismos vivos, é convertida em uma mistura de configurações R e S. A isomerização no  $C_{20}$  do  $C_{29}$   $5\alpha(H),14\alpha(H),17\alpha(H)$  aumenta a razão de zero para aproximadamente 0,5, sendo o valor de equilíbrio entre 0,52 e 0,55.

Alguns fatores como variações organofácies e a biodegradação podem afetar razões de esteranos. No caso da biodegradação, há a remoção seletiva dos epímeros resultando no aumento da razão para valores acima de 0,55 (PETERS *et al.*, 2005).

- $\beta\beta/(\beta\beta+\alpha\alpha)$ ,  $C_{29}$  esteranos

A isomerização no C-14 e C-17 nos  $C_{29}$  esteranos regulares 20S e 20R gera um aumento na razão  $\beta\beta/(\beta\beta+\alpha\alpha)$ , que varia de valores próximos de zero até 0,7 (valores de equilíbrio 0,67 a 0,71). O gráfico  $\beta\beta/(\beta\beta+\alpha\alpha)$  versus  $20S/(20S+20R)$  para os esteranos  $C_{29}$  são eficazes na caracterização da maturação térmica das rochas geradoras e óleos, sendo muito utilizado no estudo de sistemas petrolíferos (SEIFERT & MOLDOWAN, 1986).

Geralmente, os parâmetros de maturação baseados em biomarcadores saturados não podem ser aplicados em amostras de óleo altamente maturo como dito anteriormente. Sendo assim, utiliza-se uma análise integrada dos parâmetros de biomarcadores e diamantóides, já que estes são mais estáveis à degradação térmica.

Os diamantóides já foram identificados em vários óleos brutos (WEI *et al.*, 2007; AZEVEDO *et al.*, 2008 e referências, SPRINGER *et al.*, 2010; FANG *et al.*, 2012). Eles são rígidos, e mais estáveis do que os hidrocarbonetos: uma vez formados, eles são resistentes à destruição térmica e biológica (WINGERT, 1992). Variações na estabilidade térmica de diamantóides metil-substituídos levaram à utilização de determinadas proporções de isômeros como parâmetros de maturação de óleos brutos e rochas geradoras, especialmente em fases de alta geração de hidrocarboneto (CHEN *et al.*, 1995; 1996).

Usando as relações de diamantóides em combinação com os biomarcadores, podemos efetuar uma avaliação mais refinada de maturidade e de ambiente deposicional (DAHL *et al.*, 1999; AZEVEDO *et al.*, 2008; SPRINGER *et al.*, 2010).

Os parâmetros de diamantóides empregados na avaliação das amostras de óleo do presente estudo foram os índices IMA (índice metil-adamantano) e IMD (índice dimetiladamantano), proposto por CHEN *et al.* (1996), e os índices do gráfico de DAHL *et al.* (1999) para avaliação do grau de maturação térmica, apresentados a seguir.

Os estudos de DAHL *et al.* (1999) propuseram que a estabilidade térmica dos diamantóides conduz à sua concentração progressiva durante o craqueamento do óleo, enquanto a concentração de biomarcadores diminui. Este método é particularmente útil para condensados altamente maduros, onde outros métodos podem não ser confiáveis. Os biomarcadores, por geralmente apresentarem ligações que podem ser rompidas, com o aumento da maturação térmica, sua concentração decresce antes do craqueamento da maioria dos componentes de óleos. Os diamantóides, que apresentam estrutura molecular bastante estável quando comparada a outros componentes do petróleo, tendem a ter concentrações maiores no óleo residual. Desta forma, o craqueamento intenso de um óleo acarreta a destruição da maioria dos componentes e o aumento da concentração de diamantóides. Segundo DAHL *et al.* (1999), diferentes biomarcadores apresentam comportamentos diferentes frente à estabilidade térmica.

Para obtenção da correlação entre diamantóides e biomarcadores, sugeriram preferencialmente a utilização da concentração de um dos biomarcadores menos estáveis, o  $5\alpha(H),14\alpha(H),17\alpha(H)$ -24-etilcolestano 20R (estigmastano), pois sua concentração se aproxima de zero no ponto em que a concentração de diamantóides começa a aumentar (Figura 6). A utilização das concentrações do 3-metildiamantano e 4-metildiamantano foi sugerida pelo fato de que estes compostos não são facilmente perdidos por evaporação.

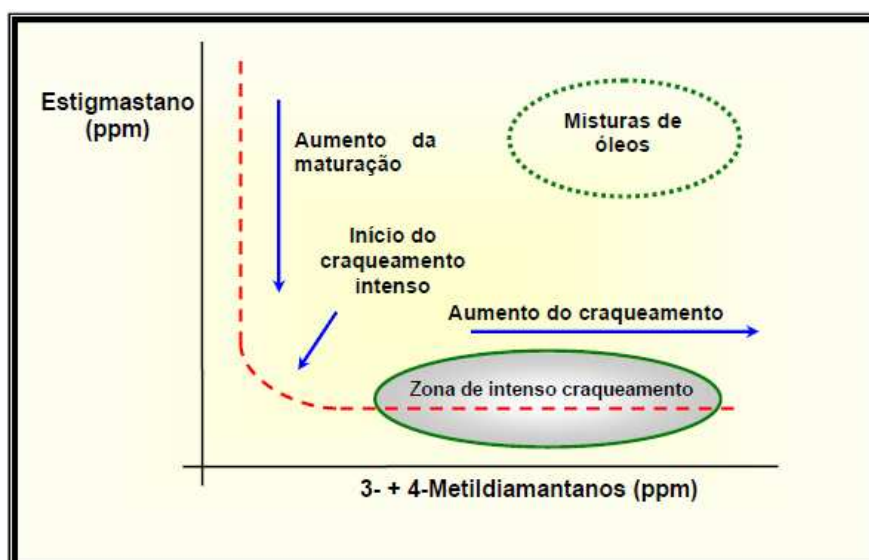


Figura 6: Ilustração da curva assintótica caracterizando a correlação entre as concentrações de diamantóides (metildiamantanos) e biomarcadores (estigmastano) para óleos de diferentes níveis de maturação térmica, craqueados e não craqueados, de um sistema petrolífero específico (adaptado de DAHL *et al.*, 1999).

CHEN *et al.* (1996) identificaram os diamantóides em amostras de óleos e rochas geradoras de algumas bacias sedimentares chinesas. Com a finalidade de avaliar o grau de evolução térmica, foram estabelecidas duas razões de diamantóides: (1) IMA – índice de metiladamantano =  $1MA/(1MA+2MA)$ , (2) IMD – índice de metildiamantano =  $4MD/(1MD + 3 MD + 4 MD)$ . A abundância relativa do 1-metiladamantano e 4-metildiadamantano deve aumentar em função do grau de evolução térmica em razão das maiores estabilidades desses compostos.

Esses parâmetros de maturação foram escolhidos por serem mais resistentes ao estresse térmico, e já terem sido amplamente estudados (PETERS & MOLDOWAN, 1993; AZEVEDO *et al.*, 2008; SPRINGER *et al.*, 2010).

## Capítulo 4.

### Fundamentação Teórica

#### 4.1) Análise de agrupamento – “Clustering”

A análise de agrupamento é uma das etapas na mineração de dados que é rotineiramente utilizada como um primeiro passo para “descobrir” a estrutura oculta do conjunto de dados não ordenados, criando grupos com certa similaridade, ou seja, cria grupos associados a uma função de pertinência (HÜLLERMEIER, 2005). Por esse critério: sendo um conjunto de dados representados pelos registros e valores das variáveis, deseja-se criar uma função que determine qual grupo cada registro pertence. O objetivo desta técnica é dividir o conjunto de dados em subconjuntos homogêneos. A análise de agrupamento possui diversas etapas:

- representação dos padrões (podendo incluir extração ou seleção de características): envolve definição do número, tipo e modo de apresentação dos atributos que descrevem cada padrão;
- seleção de características: processo de identificação do subconjunto mais efetivo dos atributos disponíveis para descrever cada padrão;
- extração de características: uso de uma ou mais transformações junto aos atributos de entrada de modo a salientar uma ou mais característica dentre aquelas que estão presentes nos dados.
- medida de similaridade: fornecida por uma função de distância definida entre pares de dados ou padrões. É possível incluir na medida de distância aspectos conceituais (qualitativos) ou então numéricos (quantitativos).
- agrupamento: os grupos podem ser definidos como conjuntos *crisp* (um padrão pertence ou não-pertence a um dado grupo) ou *fuzzy* (um padrão pode apresentar graus de pertinência aos grupos). O processo de agrupamento pode

ser hierárquico, com um processo recursivo de junções ou separações de grupos, ou não-hierárquico, com o emprego direto de técnicas de discriminação de clusters.

- apresentação: deve permitir que um computador possa utilizar o resultado de forma direta ou então deve ser orientada ao usuário, permitindo a visualização gráfica dos clusters e a compreensão de suas inter relações, através da proposição de protótipos ou outras descrições compactas para os clusters.
- realimentação do resultado do processo de *clusterização* pode levar redefinição dos módulos de “extração ou seleção de características” e “medida de similaridade”.

A Figura 7 apresenta as etapas dessa análise.

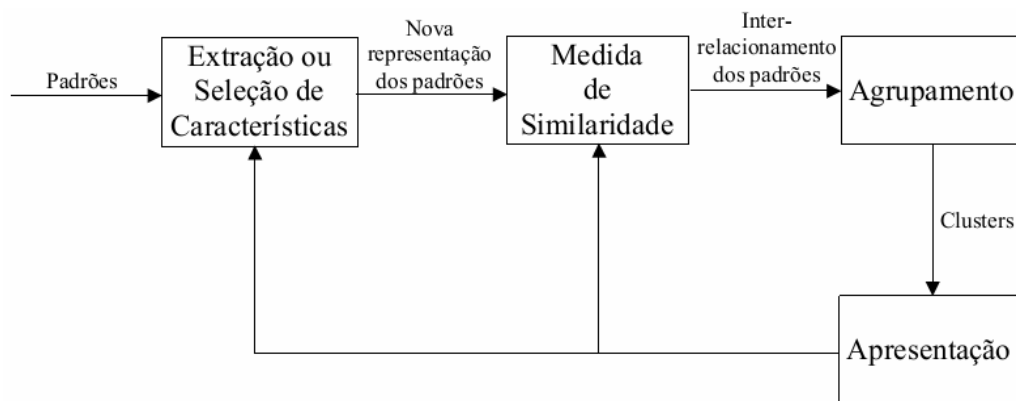


Figura 7: Esquema ilustrando as etapas do processo de agrupamento (adaptado de HÜLLERMEIER, 2005).

Sendo um ramo da estatística, a *clusterização* tem sido vastamente estudada por muitos anos, baseando-se na introdução de uma medida de distância entre os objetos. A partir da utilização dessa medida, determina-se um *critério* que expressa a ideia da *clusterização*, isto é, que particiona o conjunto de objetos em  $n$  *clusters* individuais e homogêneos, nos quais elementos de um *cluster* são tão similares entre si quanto possível, e tão diferentes quanto possível, dos elementos dos outros clusters. Segundo BRAILOVSKY (1991), a minimização desse critério resulta em *clusters* ótimos.

A análise de agrupamentos lida com a descoberta de estruturas e grupos de um conjunto de dados, e para melhor proveito, dados multivariados. Assim sendo, dificilmente alguém poderá afirmar que a descoberta de determinado grupo será perfeita. De fato, ruídos ou imprecisões não podem ser completamente eliminados (HOPPNER *et al.*, 2000).

Na análise de agrupamento clássica, cada objeto é atribuído a um “*cluster*” de uma maneira inequívoca. Consequentemente, os grupos individuais são separados por uma fronteira acentuada, abrupta. Na prática, tais limites não são frequentemente muito naturais. Na verdade, a fronteira entre os “*clusters*” e a transição entre eles são geralmente suave e não abrupta. Isso pode ser feito por meio do uso da Teoria de Conjuntos *Fuzzy* (TCF) para representar valores imprecisos. Nos métodos de agrupamento *fuzzy*, um objeto pode pertencer a diferentes grupos ao mesmo tempo, o grau que cada objeto pertence a um grupo particular é expresso em termos de uma função de pertinência (HÜLLERMEIER, 2005).

No presente trabalho, o algoritmo adotado é o *fuzzy c-means* (FCM), que pertence ao grupo de algoritmos para agrupamentos *fuzzy*, desenvolvido segundo EVSUKOFF *et al.*, 2004. A seguir será apresentado o mecanismo de agrupamento via *fuzzy c-means*.

## 4.2) Método de Agrupamento *Fuzzy*

Como visto anteriormente, na análise de “*cluster*” clássica (*crisp*), essas classes devem formar uma partição de  $X$ ; indivíduos que pertencem ao mesmo bloco (classe) da partição estão totalmente relacionados (são indistinguíveis) e indivíduos que pertencem a blocos distintos não estão relacionados. Em outras palavras, indivíduos estão totalmente relacionados entre si ou não estão relacionados. Em muitas aplicações práticas, no entanto, uma partição *crisp* pode ser muito restritiva e inviável, devido, muitas vezes, à imprecisão ou a não completeza dos dados.

Segundo BEZDEK & PAL (1992), a imprecisão nos dados pode surgir de diversas fontes. Por exemplo, erros em instrumentos ou ruídos no experimento podem levar a valores parcialmente confiáveis de determinados atributos. Em alguns casos, o custo envolvido na extração de valores muito precisos de um atributo pode ser alto. Em outros casos, pode ser difícil decidir quais são os atributos mais relevantes que caracterizam um elemento. Por essas razões, torna-se conveniente o uso de variáveis linguísticas e limitações para descrever os valores de atributos, em vez de tentar fornecer uma representação numérica exata para os dados com valores incertos dos atributos. Isso pode ser feito por meio do uso da Teoria de Conjuntos *Fuzzy* (TCF) para representar valores imprecisos. Conjuntos *Fuzzy* estão inerentemente inclinados a lidar com o conhecimento e domínio linguístico produzindo soluções mais interpretáveis. Essa teoria de conjuntos *fuzzy* foi introduzida primeiramente por Zadeh em meados da década de 60, e construía a partir dos conceitos já estabelecidos da lógica clássica.

Na teoria clássica dos conjuntos, as relações entre um elemento e um conjunto são bem definidas: dado um conjunto  $A$  em um universo  $U$ , os elementos deste universo pertencem ou não pertencem àquele conjunto. Esta relação é expressa por uma função característica  $f_A$ : se  $f_A(x) = 1$ ,  $x$  pertence a  $A$ ; se  $f_A(x) = 0$ ,  $x$  não pertence a  $A$ .

Zadeh propôs uma generalização: a função característica pode assumir um número infinito de valores no intervalo de  $[0,1]$ . Sendo assim, um conjunto *fuzzy*  $A$  em um universo  $U$  é definido por uma função de pertinência  $\mu_A(x): U \rightarrow [0,1]$ . Um determinado elemento do universo pode pertencer simultaneamente a um conjunto e ao seu complementar, sendo representado pela função de pertinência (REZENDE, 2005).

Segundo KLIR & YUAN (1995), a teoria de conjuntos *fuzzy* pode ser utilizada em pelo menos dois níveis no problema de *clusterização*: 1) no nível de atributos, para representar os elementos do conjunto como um vetor de graus de pertinência, sendo que cada um desses graus representa o grau de “posse” do atributo em questão, por parte desses elementos; e 2) no nível de classificação, para representar a pertinência desses elementos às classes, bem como para prover uma estimativa das informações incompletas em termos de valores de pertinência. Em ambos os casos, permitem-se diferentes graus de relacionamento entre elementos do conjunto, sendo que um elemento pode pertencer a mais de uma classe. Com a utilização da abordagem *fuzzy*, o problema passa então a ser caracterizado como um problema de *clusterização fuzzy*, cujo objetivo é a obtenção de uma partição *fuzzy* em um conjunto de dados  $X$ . No agrupamento *fuzzy*, cada “*cluster*” é um conjunto *fuzzy* de todos os padrões, como apresentado na Figura 8, que exemplifica agrupamentos nebulosos e não nebulosos sendo os retângulos H1 e H2 agrupamentos *crisp (hard)* e as elipses F1 e F2 a saída do algoritmo nebuloso.

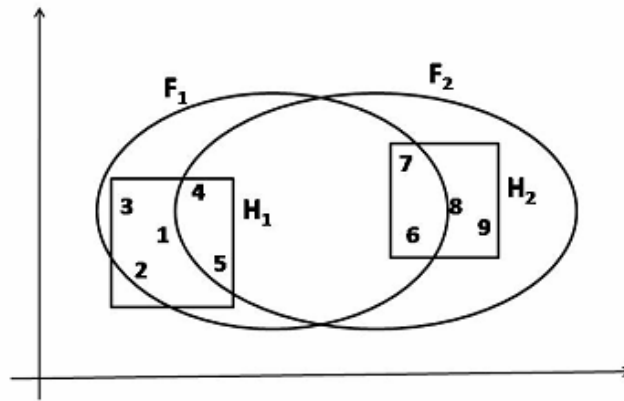


Figura 8: Comparação entre agrupamentos clássicos (conjuntos  $H_1$  e  $H_2$ ) e agrupamentos *fuzzy* (conjuntos  $F_1$  e  $F_2$ ) (modificado de JAIN, MURTY, FLYNN, 1999).

Existem muitos algoritmos de “*clusterização*” disponíveis na literatura, sendo que a sua escolha depende dos tipos de dados disponíveis e da aplicação desejada (GAN *et al.*, 2007). Dois tipos são mais aceitos: algoritmos de métodos hierárquicos e de métodos de partição. O primeiro organiza o conjunto de dados em uma forma hierárquica de acordo com a proximidade entre os indivíduos e os resultados são visualizados na forma de um dendrograma. Já os métodos particionais associam um conjunto de indivíduos a “*k*” grupos sem criar uma estrutura hierárquica.

A maioria dos algoritmos de *clusterização* é de alguma forma dependente de uma distância a qual é utilizada para definir que registros são similares e quais não. Vários métodos têm sido desenvolvidos para obter tanto clusters baseados na teoria clássica quanto *fuzzy* a partir de um determinado conjunto de dados. Para o caso da abordagem *fuzzy*, foco deste trabalho, na seção seguinte será apresentado o algoritmo de relevância e aplicabilidade para este estudo, o algoritmo *Fuzzy c-means*.

### 4.2.1) Método de Agrupamento *Fuzzy C-means* (FCM)

*Fuzzy C-means* é o algoritmo de *clusterização* de método de partição mais popular, e é derivado do algoritmo *K-means*. O método de *clusterização* FCM é feito substituindo a função característica por uma função de pertinência, que gera uma transição suave, ou seja, permite que um dado pertença a dois ou mais agrupamentos simultaneamente (EVSUKOFF *et al.*, 2004; EVSUKOFF *et al.*, 2006; OLIVEIRA & PEDRYCZ, 2007). Pode ser descrito da seguinte maneira:

Supondo um conjunto de indivíduos  $X=\{x_1,\dots,x_n\}$ ,  $x \in \mathbb{R}^p$  (espaço  $p$  –

dimensional), deseja-se organizá-los em clusters,  $C=\{C_1,\dots, C_c\}$ . O algoritmo FCM objetiva fornecer uma partição *fuzzy* de um conjunto de indivíduos em  $c$  clusters. Sendo assim, ele define e minimiza uma função objetivo ( $J_m$ ), a qual mede a adequação entre os indivíduos e os clusters:

$$J_m(U, V) = \sum_{i=1}^c \sum_{j=1}^n (\mu_{ij})^m (d_{ij})^2$$

Onde,  $m>1$  representa o fator de nebulosidade que existe entre uma categoria e outra. O algoritmo possui  $V=\{v_1,\dots,v_c\}$  como um conjunto de vetores que representa os centros das categorias.  $U$  representa a matriz de graus de inclusão nebulosos, e  $\mu_{ij}$  é o grau de inclusão do ponto  $j$  na categoria  $i$ . A matriz  $U$  deve seguir as seguintes condições: a soma dos valores de pertinência de um registro a todas as classes é igual a 1, ou melhor, a soma de todas as colunas de uma linha da matriz  $U$  é igual a 1; cada registro deve pertencer a pelo menos um agrupamento e nenhum deles deve conter todos registros, ou seja, a soma de todas as linhas da matriz  $U$  deve ser maior que 0 e menor que  $n$ .

O método *fuzzy c-means* pode ser equacionado por meio de um algoritmo iterativo, baseado na minimização de um índice de desempenho, que indica a

adequabilidade da pseudopartição gerada. O desempenho do algoritmo é influenciado pela escolha do número de classes  $c$ , dos centros de cluster iniciais, da ordem na qual os vetores são processados, da medida de distância, do critério de parada e pelas propriedades geométricas dos dados. Os conjuntos que apresentam clusters compactos, bem separados e com formas hiperesféricas, são apropriados para este método, mas para encontrar agrupamentos adequados é necessário realizar extensivos testes com vários valores de  $c$ , distâncias, critérios de parada, centros de cluster iniciais e diferentes ordens de amostras.

Assim, o algoritmo assume como entrada os seguintes parâmetros: o número desejado de *clusters*  $c$ ; uma medida de distância  $m \in (1, \infty)$ , que define a distância permitida entre os pontos e os centros de cluster; e um número pequeno  $\varepsilon > 0$ , utilizado como um critério de parada, além da instanciação inicial dos graus de pertinência dos objetos a cada uma das classes e dos centros de cluster dessas classes. A seguir, serão descritos os passos do algoritmo.

**Passo1:** Considere  $t = 0$  representando a iteração 0. Defina  $P^{(0)}$  como uma pseudopartição  $P = \{A_1, A_2, \dots, A_c\}$ , atribuindo os graus de pertinência dos elementos às classes dessa pseudopartição, os quais podem ser informados ou calculados de forma aleatória.

**Passo2:** No caso de  $t = 0$ , forneça os  $c$  centros de cluster  $v_1^{(t)}, \dots, v_c^{(t)}$  iniciais. Para  $t > 0$ , calcule os  $c$  centros de cluster  $v_1^{(t)}, \dots, v_c^{(t)}$  pela equação abaixo para  $P^{(t)}$  e o valor de  $m$  fornecido.

$$v_i = \frac{\sum_{k=1}^n [A_i(x_k)]^m x_k}{\sum_{k=1}^n [A_i(x_k)]^m}$$

A variável real  $m > 1$  é chamada de índice de *fuzzificação* e é usada para definir a distância permitida entre os pontos e o centro que está sendo calculado. Quanto maior o valor de  $m$ , mais elementos do conjunto são considerados como pertencentes

a uma pseudopartição. Esse parâmetro é escolhido de acordo com o problema considerado. Não existe nenhuma base teórica para uma escolha ótima do valor de  $m$ . O vetor  $v_i$ , calculado pela equação acima visto como o centro do cluster  $A_i$  é a média ponderada dos dados em  $A_i$ . O peso do dado  $X$  é a  $m$ -ésima potência do seu grau de pertinência ao conjunto *fuzzy*  $A_i$ .

**Passo3:** Atualize  $P^{(t)}$  para  $P^{(t+1)}$  usando o seguinte procedimento: para cada  $x_k \in X$  e para todo  $i \in \{1,2,\dots,c\}$ , se  $\|x_k - v_i^{(t)}\|^2 > 0$ , calcule o grau de pertinência do elemento  $x_k$  à classe  $A_i$ , pela fórmula abaixo, sendo que as classes  $A_i$  podem ser selecionadas em uma ordem pré-definida ou aleatoriamente.

$$A_i^{(t+1)}(x_k) = \left[ \sum_{j=1}^c \left( \frac{\|x_k - v_i^{(t)}\|^2}{\|x_k - v_j^{(t)}\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}$$

Assume-se que  $\|\cdot\|$  é alguma norma indicada para o produto interno no espaço  $\mathbb{R}^p$  e  $\|x_k - v_i^{(t)}\|^2$  representa a distância entre  $x_k$  e  $v_i$ .

Por meio da fórmula acima, a distância de cada elemento  $x_k$  ao centro de cluster  $v_i$  da pseudopartição atual  $A_i$  é comparada à distância do mesmo elemento ao centro de cluster  $v_j$  de todas as demais pseudopartições  $A_j$ , de modo a atribuir um grau de pertinência ao elemento  $x_k$  no cluster  $A_i$  que seja proporcional ao seu grau de pertinência aos demais clusters  $A_j$ .

Quando  $\|x_k - v_i^{(t)}\|^2 = 0$  para algum  $i \in I \subseteq \{1,2,\dots,c\}$ , o que indica que  $x_k$  corresponde ao centro de cluster  $v_i$ , defina  $A_i^{(t+1)}(x_k)$  como um número real não negativo que satisfaz a equação abaixo e defina  $A_i^{(t+1)}(x_k) = 0$  para  $i \in \{1,2,\dots,c\} - I$ .

$$\sum_{i \in I} A_i^{(t+1)}(x_k) = 1$$

Em casos como este, quando o elemento coincide com o centro de *cluster* de uma classe  $A_i$ , o ideal seria que o algoritmo atribuísse a tal elemento o grau de pertinência máximo (1) àquela classe e o grau de pertinência mínimo (0) às demais classes. No entanto, o teste realizado pela fórmula acima se torna necessário, uma vez que esse elemento pode ter sido associado a outras classes com um grau de pertinência maior que zero, se  $A_i$  não for a primeira classe selecionada pelo algoritmo. Nota-se, portanto, a influência da ordem na qual as pseudopartições são escolhidas.

**Passo 4:** Como critério de parada, compare  $P^{(t)}$  e  $P^{(t+1)}$ . Se  $|P^{(t)} - P^{(t+1)}| \leq \epsilon$ , então pare; caso contrário, faça  $t = t + 1$  e retorne ao Passo 2.

Neste teste,  $|P^{(t)} - P^{(t+1)}|$  denota a distância entre  $P^{(t+1)}$  e  $P^{(t)}$ , dada pela fórmula:

$$|P^{(t)} - P^{(t+1)}| = \max_{i \in N_c, k \in N_n} |A_i^{(t+1)}(x_k) - A_i^{(t)}(x_k)|$$

Essa fórmula verifica se a máxima diferença entre o grau de pertinência de qualquer elemento  $x_k$  em uma dada classe  $A_i$  na iteração atual e na iteração anterior é menor que o erro definido  $\epsilon$ , sendo isso feito para todas as classes. Em caso positivo, o algoritmo deve parar e retornar os graus de pertinências atuais de todos os elementos, em todas as classes, como resultado. Note que, quanto menor o valor de  $\epsilon$ , maior o número de passos e, conseqüentemente, mais refinada é a pseudopartição final obtida.

Ao final da última iteração do algoritmo, a pseudopartição obtida deve ser analisada segundo algum critério que expresse a ideia geral de que as associações são fortes dentro do cluster e fracas entre *clusters*. Este critério pode ser definido, por exemplo, em termos de índices de desempenho, tal como o índice  $J_m(P)$ , dado em função dos centros de *cluster* e de  $m$  pela fórmula a seguir.

$$J_m(P) = \sum_{k=1}^n \sum_{i=1}^c [A_i(x_k)]^m \|x_k - v_i\|^2$$

Este índice de desempenho mede, para todos os elementos, a soma das distâncias ponderadas de cada elemento a cada um dos centros de *cluster* da pseudopartição. Quanto menor o valor de  $J_m(P)$ , melhor a pseudopartição *fuzzy*  $P$ . Consequentemente, o objetivo do método de *clusterização fuzzy c-means* é encontrar uma pseudopartição  $P$  que minimize o índice de desempenho  $J_m(P)$ . Assim, o problema de *clusterização* pode ser equacionado como um problema de otimização. Vários testes, utilizando diferentes valores para os parâmetros do algoritmo, podem ser feitos de forma a se obter índices de desempenho menores, ou seja, resultados mais adequados.

#### **4.2.2) Ajuste dos parâmetros**

Diversos são os critérios de validação do resultado do agrupamento propostos na literatura (XIE & BENI, 1991; BEZDEK, 1998, PAKHIRA, 1998).

Após a aplicação desses algoritmos algumas perguntas ficam no ar, como por exemplo: “*Quantos grupos existem no conjunto de dados?*”; “*Os resultados realmente representam o meu conjunto de dados?*”; “*Essa é a melhor maneira de particionar os dados?*”. Para tais questionamentos, existem medidas que visam validar os métodos empregados, medindo a qualidade dos agrupamentos resultantes da aplicação dos algoritmos. A ideia geral presente nesses índices de validação baseia-se em uma visão geométrica da partição, de forma que uma boa partição apresenta clusters compactos e bem separados (EVSUKOFF *et al.*, 2004). Para a análise de *cluster fuzzy*, em que o valor de pertinência é associado a cada ponto, o grau de superposição entre os grupos deve ser levado em conta, sendo a melhor partição aquela que apresentar a menor superposição.

Para produzir uma escolha ótima do número de grupos, alguns índices são utilizados para validar, *a posteriori*, o resultado do agrupamento. No entanto, não existe consenso quanto ao índice ou coeficiente de validade mais eficiente, havendo

ainda pesquisas sendo desenvolvidas a esse respeito, como em KIM *et al.* (2004) e WU & YANG (2005), com uma tentativa de encontrar um índice que se encaixe a qualquer estrutura de dados.

Um dos índices mais recentes de validação, o *PBM*, foi proposto por três pesquisadores: Pakhira, Bandyopadhyay e Maulik. Este índice é o produto de três fatores, cuja maximização garante uma partição formada por um número pequeno de clusters e compactos separados (PAKHIRA *et al.*, 2004). Esse índice pode ser definido por:

$$PBM(c) = \left( \frac{1}{c} \times \frac{E_1}{E_c} \times D_c \right)^2$$

sendo  $c$  é o número de agrupamentos e,

$$E_c = \sum_{k=1}^c E_k$$

tal que,

$$E_k = \sum_{i=1}^N u_{ki} \| \mathbf{x}_i - \mathbf{v}_k \|$$

e,

$$D_c = \max_{k,i=1}^c \| \mathbf{v}_k - \mathbf{v}_i \|$$

$N$  é o número total de pontos do conjunto,  $U(X) = [u_{ki}]_{c \times N}$  é uma matriz de pertinência para os dados e  $\mathbf{v}_i$  é a matriz de centro do  $i$ -ésimo agrupamento. O objetivo é maximizar o índice para obter o número de agrupamentos. Quanto maior o índice PBM, melhor será a *clusterização fuzzy*.

O índice Calinski e Harabasz (CALINSKI & HARABASZ, 1974) escolhe como melhor esquema de agrupamento aquele em que obtém o maior valor do índice, faz testes estatísticos e vale-se apenas da estrutura interna para validar os agrupamentos.

$$CH(K) = \frac{B(K)/(K-1)}{W(K)/(n-K)}$$

em que

$$W(k) = \sum_{j=1}^K \sum_{i=1}^{n_j} \|x_i - c_j\|^2$$

é a dispersão interna dos grupos,

$$B(k) = \sum_{j=1}^K n_j \|c_j - c\|^2$$

é a dispersão entre os grupos e  $c$  o centro do conjunto de dados.

O índice de validação XB (XIE & BENI, 1991) define-se pela seguinte equação:

$$XB(K) = \frac{\sum_{i=1}^K \sum_{j=1}^n u_{ij}^2 \|c_i - x_j\|^2}{n \cdot \min_{i,l} \|c_i - c_l\|^2}$$

em que  $u_{ij}$  são entradas de uma matriz  $U$  que representa um agrupamento de dados difuso. Valores baixos deste índice indicam grupos compactos e bem separados.

Índices de validação como PBM, Calinski-Harabasz, Xie-Beni, de maneira geral, tentam avaliar a capacidade intra-cluster e a separação inter-cluster (PAKHIRA *et al.*, 2004).

Note-se que o uso concomitante de um conjunto de índices diferentes pode representar um subsídio mais robusto na escolha do melhor número de grupos. Isto porque resultados de índices de origens diferentes e com interpretações diversas que apresentam recomendações coincidentes reforçam as convicções para a escolha a ser feita.

O índice de validação PBM obteve bons resultados em vários bancos de dados quando comparado com outros índices (PAKHIRA *et al.*, 2004). Sendo assim, o índice PBM foi aplicado na realização deste trabalho.

## **Capítulo 5.**

### **Materiais e Métodos**

A utilização das técnicas geoquímicas permite correlacionar os óleos entre si e com as rochas geradoras caracterizadas, correlações essas que podem levar a conclusões a respeito dos tipos de petróleo presentes em uma província produtora, e seus graus de evolução térmica.

Para a realização deste estudo foi selecionado um conjunto de amostras de óleos que foram caracterizadas por parâmetros geoquímicos, envolvendo biomarcadores e diamantóides.

A preparação das amostras e os métodos analíticos empregados foram realizadas do laboratório LAGOA - Laboratório de Geoquímica Orgânica Molecular e Ambiental do Instituto de Química da UFRJ.

#### **5.1) Procedimentos Analíticos**

As análises geoquímicas realizadas nas amostras de óleo, descritas a seguir, foram: separação em frações por cromatografia líquida; cromatografia gasosa do óleo total (*whole oil*) e cromatografia gasosa acoplada a espectrometria de massas da fração dos hidrocarbonetos saturados. O procedimento analítico realizado no laboratório está representado na figura 9.

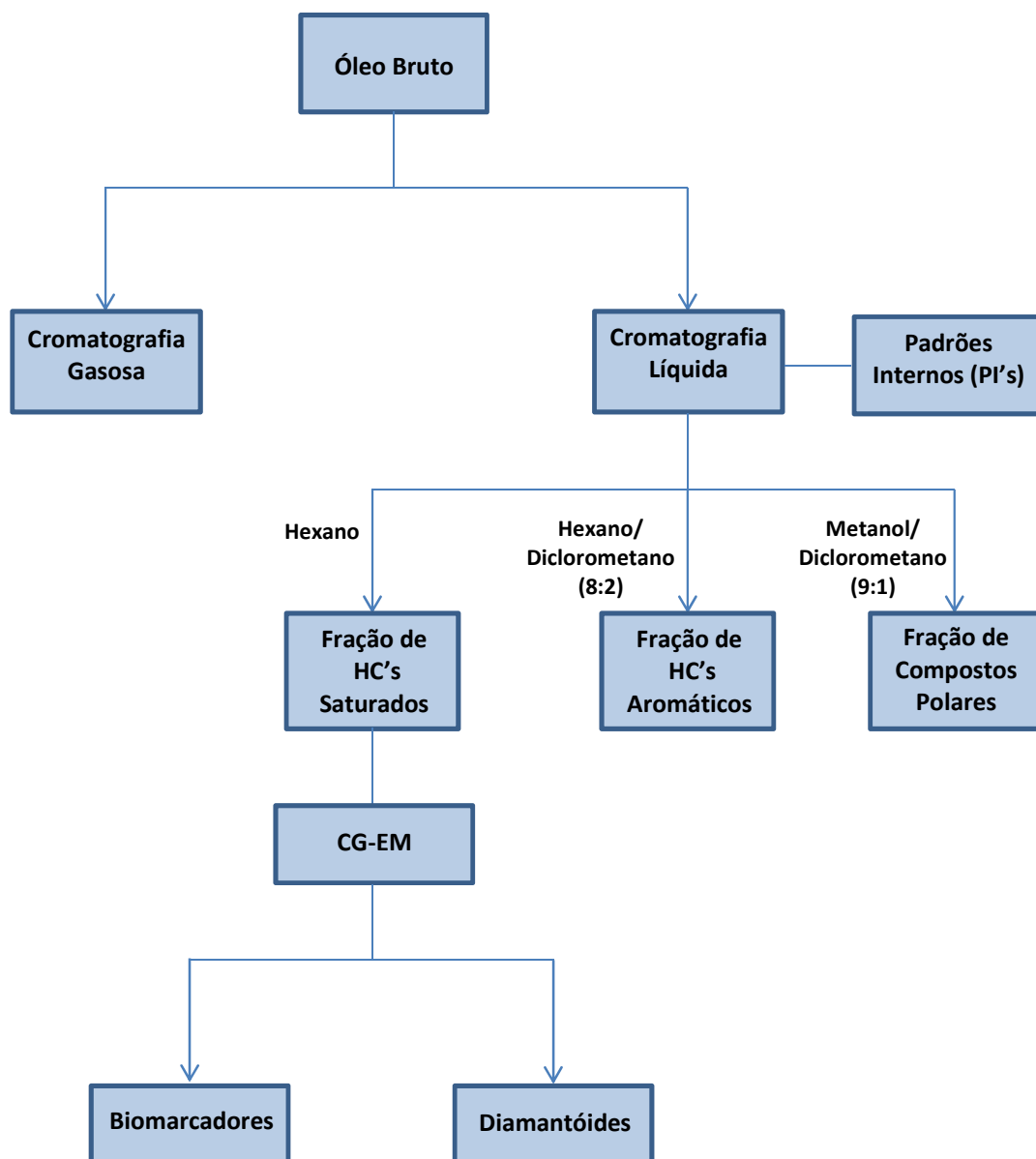


Figura 9: Esquema ilustrativo da sequência de análises para as amostras de óleos selecionados.

### 5.1.1) Técnicas analíticas utilizadas

As amostras de óleos foram analisadas utilizando-se cromatografia líquida, cromatografia gasosa (CG) e cromatografia gasosa acoplada à espectrometria de massas (CG-EM).

#### - Cromatografia líquida

A cromatografia líquida é utilizada para separar os óleos em frações: hidrocarbonetos saturados, hidrocarbonetos aromáticos e compostos polares (denominados NSO), a partir da propriedade da polaridade específica que cada composto possui.

Aproximadamente, 100 mg de óleo bruto foi pesado e dissolvido em 5,0 mL de hexano, e acrescentou-se 0,5 mL da mistura de padrões preparada previamente (*n*-dodecano-d26, *n*-hexadecano-d34 (Chiron S.A., Noruega), e *n*-tetracosano-d50 (Cambridge Isotopes Laboratories, USA)). A solução foi adicionada no topo da coluna cromatográfica contendo 2,5 mg de sílica gel ativada, e foi eluída com hexano. Sendo assim, as frações de hidrocarbonetos saturados (1ª fração), hidrocarbonetos aromáticos (2ª fração) e compostos NSO (3ª fração) foram separadas (GRICE *et al.* 2000; AZEVEDO *et al.*, 2008; SPRINGER *et al.*, 2010).

A fração de hidrocarbonetos saturados (1ª fração) foi eluída com 5,0 mL de *n*-hexano e recolhida em frascos de 2,0 mL, totalizando um volume final de aproximadamente 1,5 mL e logo em seguida analisada por cromatografia gasosa acoplada à espectrometria de massas para a detecção dos diamantóides.

Posteriormente, o volume final foi reduzido a aproximadamente 800  $\mu\text{L}$ , para análise dos biomarcadores.

A fração de hidrocarbonetos aromáticos (2ª fração) foi eluída com 10,0 mL de *n*-hexano/diclorometano (8:2), enquanto a fração dos compostos polares (3ª fração) foi eluída com 10,0 mL de diclorometano/metanol (9:1).

### **- Cromatografia gasosa**

A Cromatografia Gasosa (CG) é uma técnica de separação e análise de misturas de substâncias volatilizáveis. A cromatografia em fase gasosa permite uma separação mais refinada dos compostos orgânicos presentes em óleos ou extrato de rocha do que a cromatografia líquida. Seus resultados são apresentados sob a forma de gráfico (cromatograma) onde é observada a distribuição dos compostos orgânicos presentes.

A análise de óleo total (*Whole oil*) foi realizada em um cromatógrafo gasoso Agilent Technologies modelo HP5890 série II, coluna capilar DB-5 de 30 metros de comprimento, 0,25 mm de diâmetro interno e 0,25  $\mu\text{m}$  de espessura de fase estacionária. As condições cromatográficas utilizadas para as análises foram: temperatura de 40°C (isoterma de 1 minuto) a 310°C (isoterma de 19 minutos), com taxa de aquecimento de 6°C  $\text{min}^{-1}$ . A temperatura do injetor foi de 290°C e do detector foi de 320°C.

### **- Cromatografia Gasosa acoplada à Espectrometria de Massas (CG-EM)**

A cromatografia gasosa não revela a informação estrutural sobre os compostos em análise, por isso é importante o acoplamento de um cromatógrafo a gás com um detector mais específico, como por exemplo, o detector seletivo de massas. Este acoplamento leva à obtenção do espectro de massas de cada analito que é eluído da cromatografia (CG). A interpretação do espectro de massas (EM) permite a identificação dos componentes de interesse.

As frações dos hidrocarbonetos saturados separadas por cromatografia em coluna de sílica foram analisadas por cromatografia gasosa de alta resolução acoplada a espectrometria de massas (CG/EM). Estas foram realizadas em um instrumento Agilent Technologies 6890N acoplado a um detector seletivo de massas Agilent Technologies 5973 utilizando uma coluna capilar de sílica fundida recoberta com DB-5 (J&W; 30 m X 0,25 mm d.i.;  $df = 0,25 \mu\text{m}$ ).

As amostras foram analisadas por CG-EM através do modo de monitoramento seletivo de íons (MSI). Para a caracterização dos biomarcadores foram selecionados os seguintes íons:  $m/z$  217 para os  $\alpha\alpha\alpha$ -esteranos,  $m/z$  218 para os  $\alpha\beta\beta$ -esteranos e  $m/z$  191 para terpanos,  $m/z$  177, 259 e 231, e  $m/z$  66 para os padrões internos ( $n$ -alcanos perdeuterados:  $n$ -dodecano-d26,  $n$ -hexadecano-d34 e  $n$ -tetracosano-d50). As condições cromatográficas utilizadas para as análises foram: temperatura 70°C a 170°C com taxa de aquecimento 20°C  $\text{min}^{-1}$ , de 170°C a 310°C com aquecimento de 2°C  $\text{min}^{-1}$  e isoterma de 310°C durante 10 minutos. Foi realizada injeção automática de 1,0  $\mu\text{L}$  da amostra, com impacto de elétrons a 70 eV, utilizando hélio como gás de carreador, a uma pressão constante de 10 psi.

No caso dos diamantóides foram selecionados os seguintes íons:  $m/z$  135, 136, 149, 163, 177 (adamantanos);  $m/z$  187, 188, 201, 215 (diamantanos) e  $m/z$  66 para os padrões internos ( $n$ -alcanos perdeuterados:  $n$ -dodecano-d26,  $n$ -hexadecano-d34 e  $n$ -tetracosano-d50). As condições cromatográficas utilizadas para as análises foram: temperatura 40°C a 180°C com taxa de aquecimento 3°C  $\text{min}^{-1}$ , de 180°C a 310°C com aquecimento de 6°C  $\text{min}^{-1}$  e isoterma de 310°C durante 10 minutos. Foi realizada injeção automática de 1,0  $\mu\text{L}$  da amostra, com impacto de elétrons a 70 eV, utilizando hélio como gás de carreador, a uma pressão constante de 10 psi.

## 5.2) Avaliação estatística dos dados

Hoje em dia, existem diversas ferramentas encontradas no mercado que seriam passíveis de serem utilizadas na execução deste estudo. Dentre elas destacam-se: o Statistica®, MatLab®, Clementine®, Intelligent Miner® e Weka®. Porém, para a realização deste trabalho optou-se pela utilização do MatLab® versão 6.5, de fácil interface com o usuário e com módulos de visualização já contemplados no pacote.

O MatLab® é um "software" interativo de alta performance voltado para o cálculo numérico, que integra análise numérica, cálculo com matrizes, processamento de sinais e construção de gráficos em ambiente fácil de usar onde problemas e soluções são expressos somente como eles são escritos matematicamente, ao contrário da programação tradicional (SANTOS, 2009).

A modelagem pode ser dividida em quatro blocos principais, conforme apresentado na Figura 10.

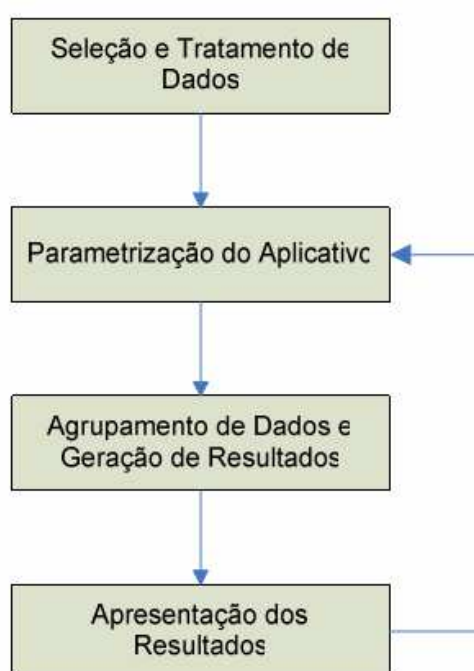


Figura 10: Esquema de visão geral do processo para análise de grupos.

A Seleção e Tratamento de Dados é o processo responsável pelo pré-processamento dos dados. O processo de análise de agrupamento de dados inicia-se com a seleção e preparação do conjunto de dados a ser utilizado na fase de agrupamento. Estas duas etapas envolvem a seleção de um conjunto de dados representativo da base de dados e a realização de processos de tratamento de dados, a fim de garantir a qualidade dos dados que serão utilizados nas etapas seguintes do processo.

A Parametrização do Aplicativo é o processo responsável pela parametrização do processo de agrupamento de dados, bem como os resultados a serem analisados.

Agrupamento de Dados e Geração de Resultados é o processo responsável pelo agrupamento de dados, bem como a geração dos dados a serem utilizados na etapa de apresentação dos resultados.

A fase de Apresentação dos Resultados é a responsável pelo processamento dos dados da etapa anterior para geração dos gráficos e planilhas, disponibilizando as tabelas e gráficos que servem de auxílio à análise dos resultados gerados pelo processo de agrupamento de dados.

## **Capítulo 6.**

### **Base de Dados - Descrição**

A caracterização dos sistemas petrolíferos, através de parâmetros de geoquímica orgânica, é fundamental para o desenvolvimento de modelos preditivos para a exploração de petróleo. A maturação térmica e a caracterização do ambiente deposicional das rochas geradoras constituem os principais parâmetros controladores de geração de hidrocarbonetos nas rochas sedimentares, e muitos autores têm mostrado que as evidências geoquímicas e as distribuições dos biomarcadores e diamantóides podem fornecer critérios para distinguir rochas geradoras depositadas em diferentes ambientes, como já discutido no capítulo 3.

Como já citado anteriormente, a base de dados contém registros de parâmetros de maturação térmica e ambiente deposicional de 69 amostras, envolvendo razões de biomarcadores e diamantóides, que foram selecionadas a partir de trabalhos já desenvolvidos anteriormente. Essas amostras de óleos são provenientes de bacias sedimentares brasileiras e da bacia sedimentar *Llanos Orientales*, sendo no total 7 bacias diferentes, como apresentado na Tabela 2. A seguir, será apresentado um resumo destes estudos, mostrando a caracterização geoquímica quanto aos parâmetros de maturação e origem.

Tabela 2: Relação das amostras de óleos usadas neste estudo. As amostras de A a G pertencem às bacias brasileiras (indicadas em verde), e as H, à bacia colombiana (indicadas em laranja).

<b>BA1</b>	<b>BC9</b>	<b>BC26</b>	<b>BH4</b>
<b>BA2</b>	<b>BC10</b>	<b>BE1</b>	<b>BH5</b>
<b>BA3</b>	<b>BC11</b>	<b>BE2</b>	<b>BH6</b>
<b>BA4</b>	<b>BC12</b>	<b>BF1</b>	<b>BH7</b>
<b>BA5</b>	<b>BC13</b>	<b>BF2</b>	<b>BH8</b>
<b>BA6</b>	<b>BC14</b>	<b>BF3</b>	<b>BH9</b>
<b>BB1</b>	<b>BC15</b>	<b>BF4</b>	<b>BH10</b>
<b>BB2</b>	<b>BC16</b>	<b>BF5</b>	<b>BH11</b>
<b>BB3</b>	<b>BC17</b>	<b>BF6</b>	<b>BH12</b>
<b>BC1</b>	<b>BC18</b>	<b>BG1</b>	<b>BH13</b>
<b>BC2</b>	<b>BC19</b>	<b>BG2</b>	<b>BH14</b>
<b>BC3</b>	<b>BC20</b>	<b>BG3</b>	<b>BH15</b>
<b>BC4</b>	<b>BC21</b>	<b>BG4</b>	<b>BH16</b>
<b>BC5</b>	<b>BC22</b>	<b>BG5</b>	<b>BH17</b>
<b>BC6</b>	<b>BC23</b>	<b>BH1</b>	<b>BH18</b>
<b>BC7</b>	<b>BC24</b>	<b>BH2</b>	<b>BH19</b>
<b>BC8</b>	<b>BC25</b>	<b>BH3</b>	<b>BH20</b>

Os estudos envolvendo as amostras de óleo selecionadas indicam que os óleos apresentam graus de maturação diferente, isto é, sofreram aquecimento diferenciado no processo de geração, ao longo do tempo geológico.

Segundo JESUINO (2005), TAMANQUEIRA (2006), CALDAS (2009) e SILVA (2010), cujo objetivo foi caracterizar amostras de bacias sedimentares brasileiras (bacias BA a BG), as amostras estudadas apresentam diferentes graus de evolução térmica, estando alguns óleos na fase de craqueamento. O emprego dos parâmetros de biomarcadores mostrou-se restrito para a avaliação de óleos com alto grau de evolução térmica, além do limite de estabilização da maioria dos parâmetros de esteranos e hopanos (que ocorre antes ou durante a “janela de geração do óleo”).

Portanto, a integração das análises de diamantóides e biomarcadores foi decisiva para a caracterização dos óleos. Os diamantóides permitiram a obtenção de resultados mais consistentes. As análises destes compostos sugerem que os óleos das bacias envolvidas apresentam diferentes estágios de maturação: foram encontradas amostras com alto grau de evolução térmica e amostras pouco evoluídas termicamente.

A origem dos óleos foi diagnosticada como sendo marinha, mas diferentes graus de influência de matéria orgânica depositada em ambiente continental (ou lacustre).

No estudo desenvolvido por SPRINGER (2007), cujo objetivo foi a caracterização geoquímica de óleos provenientes da bacia sedimentar colombiana, *Llanos Orientales*, as amostras de óleo analisadas apresentaram, também, diferentes estágios de maturação. O estudo realizado demonstrou que os óleos estudados foram classificados como biodegradados de baixa maturidade a alta maturidade e, também, como craqueados. Essa caracterização também só foi possível de ser determinada pela integração de parâmetros de biomarcadores e diamantóides, segundo Dahl et al. (1999).

Na caracterização do ambiente deposicional, baseados em parâmetros de biomarcadores, o estudo demonstrou que algumas amostras possuem características de matéria orgânica marinha, outras de matéria orgânica terrestre e, a grande maioria se comporta como mistura.

## **Capítulo 7.**

### **Resultados e Discussão**

Este capítulo apresenta primeiramente a interpretação dos resultados através dos “métodos tradicionais”, ou seja, correlações entre os parâmetros envolvidos, tanto para a caracterização da maturação quanto para a caracterização da origem. Em seguida, apresenta os resultados da execução do algoritmo *Fuzzy c-means*.

#### **7.1) Parâmetros de Maturação**

##### **7.1.1) Interpretação dos resultados pela caracterização de Biomarcadores e Diamantóides**

Para a caracterização geoquímica dos óleos em estudo, alguns parâmetros de biomarcadores e diamantóides propostos na literatura como indicadores do nível de evolução térmica (SEIFERT & MOLDOWAN, 1991; WAPLES & MACHIHARA, 1991; CHEN et al., 1996; WASEDA & NISHITA, 1998; DAHL et al., 1999; PETERS et al., 2005) foram calculados e avaliados, como mostra a Tabela 3.

Tabela 3: Valores dos parâmetros de biomarcadores dependentes de maturação térmica para as amostras de óleo estudadas. As amostras de BA a BG pertencem às bacias brasileiras, e as BH, à bacia colombiana.

AMOSTRAS	IMD	IMA	3-+4- Metildiamantanos (ppm)	Estigmastano (ppm)	Esteranos $\beta\beta/(\beta\beta+\alpha\alpha)$ (C29)	Esteranos S/(S+R) C29
BA1	0,58	0,67	90,89	9,25	0,61	0,55
BA2	0,68	0,69	137,17	21,32	0,61	0,63
BA3	0,57	0,68	91,77	106,02	0,52	0,56
BA4	0,58	0,68	130,39	50,65	0,51	0,54
BA5	0,55	0,69	111,08	8,33	0,41	0,5
BA6	0,53	0,76	29,77	3,78	0,49	0,58
BB1	0,39	0,63	0,37	34,19	0,34	0,49
BB2	0,33	0,57	1,84	83,13	0,43	0,5
BB3	0,40	0,64	2,41	73,63	0,31	0,41
BC1	0,46	0,61	45,35	342,66	0,53	0,57
BC2	0,44	0,68	44,34	721,40	0,35	0,54
BC3	0,47	0,68	21,04	28,64	0,24	0,41
BC4	0,44	0,69	22,87	235,57	0,55	0,69
BC5	0,38	0,60	16,67	766,88	0,3	0,48
BC6	0,44	0,68	25,83	688,90	0,23	0,49
BC7	0,42	0,61	20,03	41,25	0,48	0,51
BC8	0,41	0,65	5,54	29,80	0,6	0,58
BC9	0,40	0,68	7,48	30,82	0,6	0,54
BC10	0,39	0,64	4,38	24,64	0,39	0,59
BC11	0,30	0,62	86,56	20,09	0,5	0,58
BC12	0,31	0,62	30,08	44,66	0,32	0,48
BC13	0,38	0,68	9,21	19,69	0,6	0,7
BC14	0,38	0,62	8,16	294,14	0,43	0,43
BC15	0,27	0,69	45,85	6,16	0,4	0,38
BC16	0,44	0,74	4,00	8,69	0,35	0,52
BC17	0,34	0,62	83,52	25,99	0,45	0,24
BC18	0,44	0,69	4,62	7,61	0,39	0,42
BC19	0,42	0,57	26,82	76,34	0,4	0,52
BC20	0,41	0,64	5,14	11,28	0,42	0,53
BC21	0,43	0,72	32,33	13,46	0,52	0,52
BC22	0,45	0,71	30,00	23,74	0,34	0,52
BC23	0,45	0,73	150,16	0,50	0,33	0,54
BC24	0,44	0,70	19,03	148,70	0,41	0,52
BC25	0,44	0,72	19,96	112,44	0,37	0,4

<b>BC26</b>	0,59	0,71	52,38	19,52	0,31	0,36
<b>BE1</b>	0,50	0,64	2,15	135,74	0,37	0,51
<b>BE2</b>	0,46	0,61	3,39	70,26	0,35	0,48
<b>BF1</b>	0,50	0,76	23,30	56,24	0,50	0,37
<b>BF2</b>	0,52	0,77	41,09	122,80	0,44	0,28
<b>BF3</b>	0,43	0,75	40,63	57,01	0,44	0,38
<b>BF4</b>	0,49	0,64	47,97	29,30	0,57	0,22
<b>BF5</b>	0,49	0,65	59,60	37,77	0,58	0,26
<b>BF6</b>	0,47	0,63	158,36	19,16	0,49	0,28
<b>BG1</b>	0,40	0,85	11,64	36,91	0,47	0,44
<b>BG2</b>	0,61	0,67	302,70	44,13	0,62	0,33
<b>BG3</b>	0,41	0,70	33,41	23,48	0,51	0,38
<b>BG4</b>	0,50	0,67	110,28	0,69	0	0
<b>BG5</b>	0,44	0,75	63,30	22,51	0,54	0,38
<b>BH1</b>	0,42	0,70	224,61	10,16	0,50	0,51
<b>BH2</b>	0,41	0,66	34,36	7,19	0,55	0,36
<b>BH3</b>	0,43	0,69	44,91	7,40	0,63	0,35
<b>BH4</b>	0,43	0,69	49,23	8,67	0,54	0,51
<b>BH5</b>	0,42	0,68	50,27	6,40	0,76	0,33
<b>BH6</b>	0,42	0,67	51,22	12,05	0,60	0,58
<b>BH7</b>	0,42	0,67	37,24	4,85	0,75	0,27
<b>BH8</b>	0,41	0,67	46,71	11,01	0,75	0,34
<b>BH9</b>	0,41	0,68	111,30	17,68	0,68	0,46
<b>BH10</b>	0,41	0,66	121,76	6,72	0,67	0,38
<b>BH11</b>	0,42	0,69	131,14	15,01	0,70	0,33
<b>BH12</b>	0,44	0,69	110,07	31,20	0,68	0,56
<b>BH13</b>	0,43	0,70	125,72	52,18	0,68	0,38
<b>BH14</b>	0,44	0,68	111,47	58,12	0,70	0,33
<b>BH15</b>	0,45	0,68	104,33	17,68	0,61	0,33
<b>BH16</b>	0,43	0,65	87,24	16,59	0,64	0,36
<b>BH17</b>	0,40	0,61	37,85	38,77	0,53	0,50
<b>BH18</b>	0,40	0,66	59,58	25,95	0,67	0,38
<b>BH19</b>	0,43	0,64	52,78	32,92	0,38	0,55
<b>BH20</b>	0,43	0,68	106,41	25,78	0,53	0,56

IMD: Índice de metildiamantano.

IMA: Índice de metiladamantano.

Para a caracterização geoquímica das amostras de óleos em estudo, podemos destacar duas razões de biomarcadores que envolvem a isomerização de esteranos, muito utilizadas na avaliação da maturação de óleos: a razão dos  $C_{29}\alpha\alpha$  20S/ (20S+20R) e a razão dos  $C_{29}\beta\beta / (\beta\beta+\alpha\alpha)$  (FARRIMOND *et al.*, 1998).

A razão 20S/ (20S + 20R) – esterano C29 é usada para medir a evolução térmica da rocha geradora ao longo da zona imatura e parte da zona principal de geração do óleo, chamada “*janela de geração do óleo*” (PETERS & MOLDOWAN, 1993). Apenas a configuração R no C-20 está presente nos precursores esteroidais oriundos de organismos vivos. Esta configuração é convertida ao longo do processo de maturação para uma mistura de esteranos de configurações R e S e atinge seus valores de equilíbrio antes ou durante o início da “*janela de geração do óleo*”. Esses valores variam de *zero* até aproximadamente 0,5, sendo que os valores 0,52 e 0,55 se encontram na faixa de equilíbrio (SEIFERT & MOLDOWAN, 1986; PETERS *et al.*, 2005).

A razão  $\beta\beta / (\alpha\alpha + \beta\beta)$ -esterano C29 é uma razão de isomerização importante nos estudos de maturação. Seus valores de equilíbrio atingem a janela de geração do óleo, variando de *zero* a aproximadamente 0,7, sendo que os valores 0,67 e 0,71 se encontram na fase de equilíbrio. A configuração  $\alpha\alpha$  dos esteranos é a produzida biologicamente. Com o avanço do processo da diagênese, esta forma progressivamente, converte-se em uma mistura de epímeros  $\alpha\alpha$  e  $\beta\beta$ . Esta transformação ainda não é bem conhecida, acredita-se que dois átomos de hidrogênio próximos mudam simultaneamente da posição alfa ( $\alpha$ ) para a posição beta ( $\beta$ ). Já que existem os diastereoisômeros 20S e 20R para cada forma,  $\alpha\alpha$  e  $\beta\beta$  respectivamente, existirão quatro espécies distintas em cada uma das famílias ( $C_{27}$ ,  $C_{28}$  e  $C_{29}$  esteranos). A isomerização no C-14 e C-17 nos  $C_{29}$  esteranos regulares 20S e 20R gera um aumento na razão  $\beta\beta / (\alpha\alpha + \beta\beta)$ , que com o progressivo aumento da maturação, varia de valores próximos a zero até valores próximos a 0,7 (PETERS *et al.*, 2005).

Uma grande ferramenta usada para a avaliação mais completa destas duas razões é o gráfico de correlação entre elas, apresentado na Figura 11.

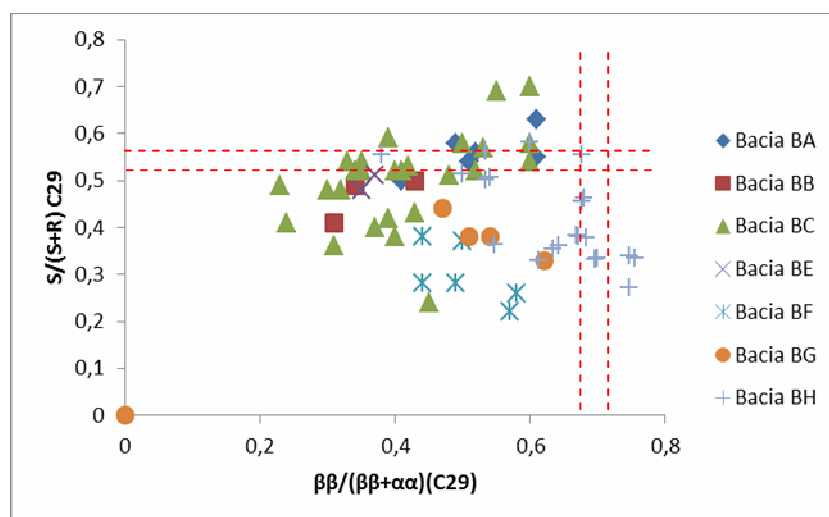


Figura 11: Gráfico de relação entre as razões  $\beta\beta / (\alpha\alpha + \beta\beta)$  versus  $20S / (20S + 20R)$ .

Para a razão  $\beta\beta / (\alpha\alpha + \beta\beta)$ , o gráfico mostrou que apenas as amostras da bacia BH (bacia sedimentar colombiana *Llanos Orientales*) atingiram ou ultrapassaram o limite de estabilização. As demais amostras de óleo, todas provenientes de bacias sedimentares brasileiras, não atingiram o limite de estabilização para esta razão.

Os resultados da razão  $20S / (20S + 20R)$  mostram que amostras provenientes de três bacias – BA, BC e BH – atingiram ou ultrapassaram a zona de equilíbrio.

As amostras de óleo provenientes das bacias BB, BF e BG se encontram abaixo das zonas de estabilização para ambas os parâmetros.

A partir da figura 11, não se observa uma tendência de evolução linear para todo o grupo de óleos analisado, e sim certa dispersão e uma tendência de agrupamento por bacia, o que evidenciaria, aparentemente, um controle da origem dos óleos sobre alguns parâmetros de maturação.

A fim de auxiliar no estudo do grau de evolução térmica dos óleos selecionados, utilizaram-se os diamantóides através dos índices propostos por Chen et al., (1996), detalhados no capítulo 3:  $MAI = 1-MA/(1-MA + 2-MA)$  e  $MDI = 4-MD/(1-MD + 3-MD + 4-MD)$ . A Figura 12 apresenta o gráfico que correlaciona esses índices.

Na avaliação das amostras sob a influência dos parâmetros de CHEN et al. (1996) foi observado que as amostras de óleo estudadas estão distribuídas em regiões de altos valores dos índices IMD e IMA. Os óleos da bacia BA e alguns óleos das bacias BC, BF e BG se destacam (círculo azul) apresentando os maiores valores para ambos os índices, indicando um elevado grau de evolução térmica. Por outro lado, alguns óleos da bacia BC e uma amostra da bacia BB (destacado no círculo vermelho) apresentam os menores valores de MAI e MDI, sendo considerado os óleos menos evoluído termicamente em relação aos demais.

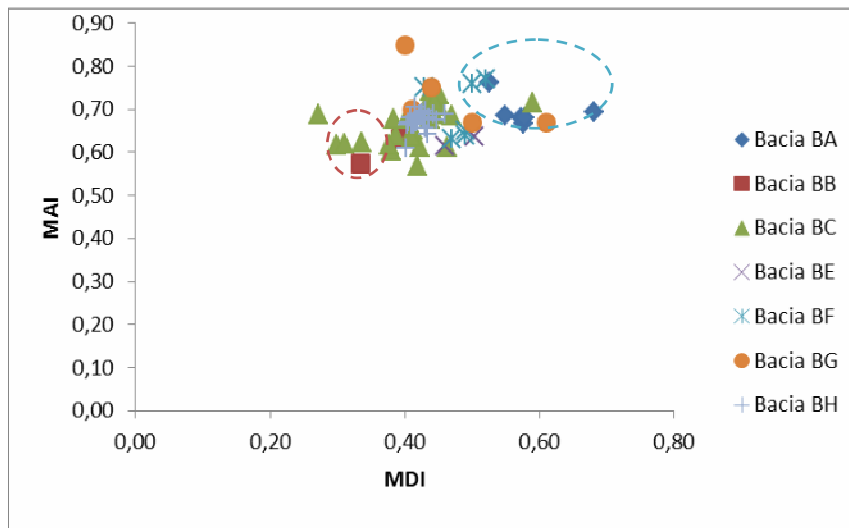


Figura 12: Gráfico entre os índices de metiladamantanos (MDI) e metildiamantanos (MAI) para as amostras de óleo selecionadas.

Outras correlações realizadas foram entre o MDI e as razões de biomarcadores  $\beta\beta / (\alpha\alpha + \beta\beta)$  e  $20S / (20S + 20R)$  – esterano C29, Figuras 13A e 13B, respectivamente. A partir destes gráficos, podemos observar que o parâmetro origem pode ser um grande controlador sob alguns fatores maturação: não existe uma tendência linear para todo o grupo, e sim um visível agrupamento por bacia. Na figura 13B, observa-se uma clara separação entre as amostras da bacia BH (bacia colombiana) e as demais amostras, provenientes de bacias sedimentares brasileiras. No gráfico da figura 13A, observou-se uma dispersão linear para as amostras das bacias brasileiras, destacando as amostras da bacia BA como as mais evoluídas termicamente.

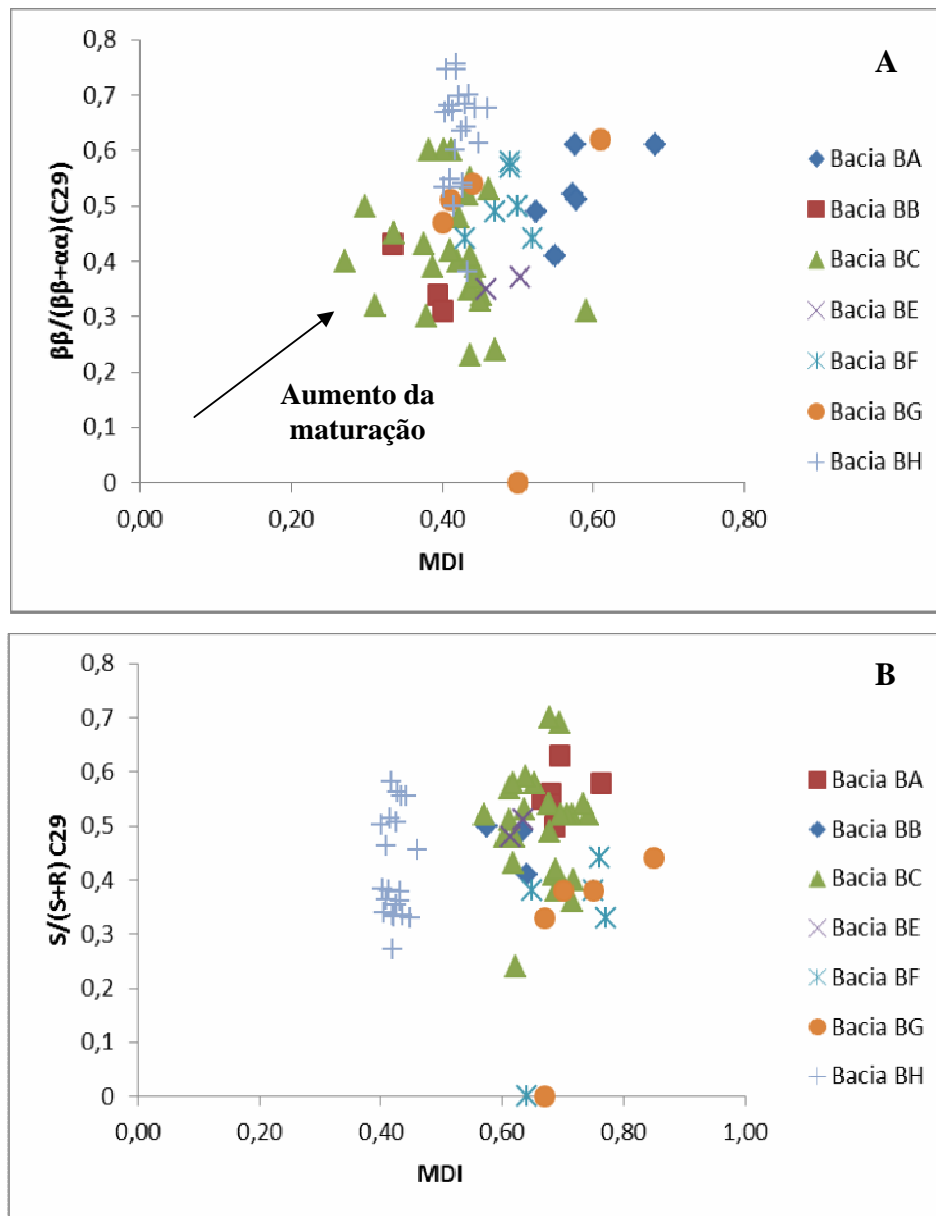


Figura 13: Gráficos de : (A) correlação entre MDI e a razão  $\beta\beta / (\alpha\alpha + \beta\beta)$ ; (B) correlação entre MDI e a razão  $20S / (20S + 20R)$  – estero C29.

Outra correlação muito estudada para analisar a evolução térmica dos óleos foi proposta por DAHL et al. (1999). A resistência à estabilidade térmica dos diamantóides leva a um conseqüente aumento da sua concentração durante a evolução térmica do óleo, enquanto a concentração dos biomarcadores, que são menos resistentes a essas condições, diminui. Com o aumento do estresse térmico os biomarcadores são degradados, este fato pode ser observado pela diminuição da concentração do estigmastano ( $5\alpha,14\alpha,17\alpha(H)$ -24-etil-colestano 20 R). Por outro lado, ocorre um enriquecimento da concentração de diamantóides, neste caso foi usado o somatório do 3-metildiamantano e 4-metildiamantano.

O resultado do gráfico entre as concentrações de 3-+4-metildiamantanos e Estigmastano está representado nas Figuras 14A e 14B. Para as amostras em estudo, é possível observar que as amostras da bacia BC têm características tanto imaturas (amostras destacadas pelo círculo vermelho) quanto maduras e altamente maduras. As amostras da bacia BB são maduras, mas ainda não se encontram na fase de craqueamento. A maioria das amostras das bacias BA e BH são altamente maduras, sendo que algumas já se encontram na fase de craqueamento, assim como uma amostra da bacia BG. Algumas amostras das bacias BA e BH apresentam altas concentrações de estigmastano e diamantóides, que segundo DAHL e colaboradores (1999), pode ser considerado mistura de óleos em diferentes pulsos de migração com níveis diferenciados de maturação.

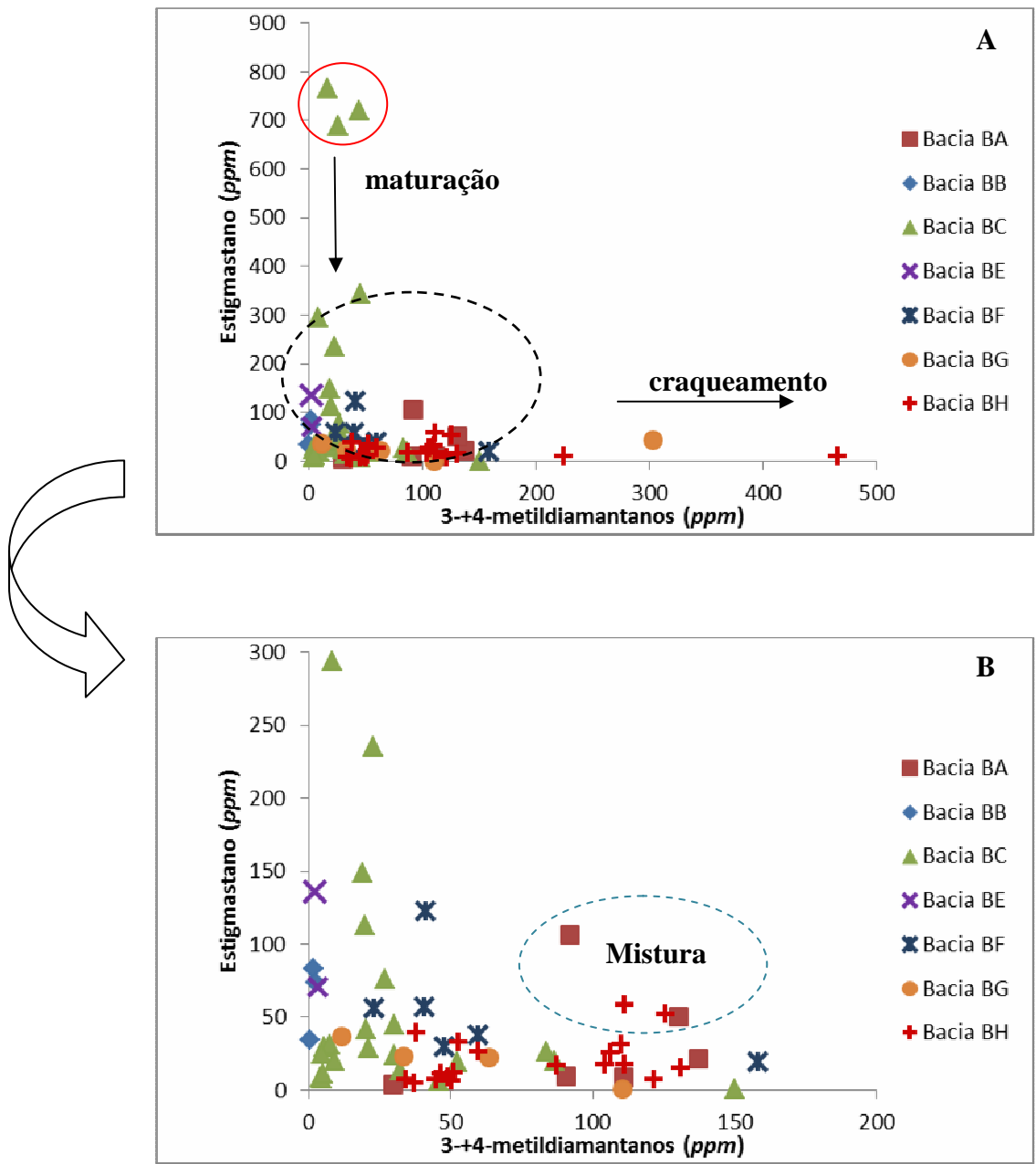


Figura 14: Gráfico de correlação entre 3+4- metildiamantano *versus* Estigmastano, proposto por DAHL et al. (1999).

### 7.1.2) Análise de agrupamento FCM – *Fuzzy c-means*

A análise de agrupamento foi realizada utilizando os parâmetros e valores mostrados na Tabela 2. Pela análise de *cluster* as amostras foram diferenciadas em cinco grupos de diferentes graus de maturação. Os resultados são discutidos a seguir.

Antes de realizar o processo de clusterização foi efetuado o pré-tratamento dos dados, analisando a presença de valores ausentes e correlações entre as variáveis, seguindo o procedimento de Data Mining.

Com a finalidade de determinar o melhor número possível de agrupamentos foi utilizado o índice de validação *PBM*, detalhado no capítulo 4 (PAKHIRA *et al.*, 2004).. Analisando a Tabela 4, que apresenta diferentes valores de *m* para cada agrupamento, o índice de validação indica cinco grupos.

Tabela 4: Determinação do número de grupos.

<b>Grupos</b>	<b>m=1,2</b>	<b>m=1,4</b>	<b>m=1,6</b>	<b>m=1,8</b>	<b>m=2,0</b>
2	0,125	0,081	0,044	0,026	0,011
3	0,185	0,124	0,073	0,050	0,033
4	0,360	0,158	0,081	0,047	0,033
<b>5</b>	<b>1,207</b>	<b>0,417</b>	<b>0,166</b>	<b>0,053</b>	<b>0,028</b>
6	1,116	0,145	0,085	0,050	0,023

Nas Figuras 15, 16 e 17, podemos observar a distribuição e a disposição desses grupos. A partir das figuras 15 e 16, pode-se observar que o grupo 5 apresenta o menor número de amostras envolvidas, e o grupo 2 o maior número.

Na figura 17, observa-se, a seguir, a disposição dos agrupamentos para 5 grupos, obedecendo à função de pertinência *fuzzy* ( $m=1,2$ ).

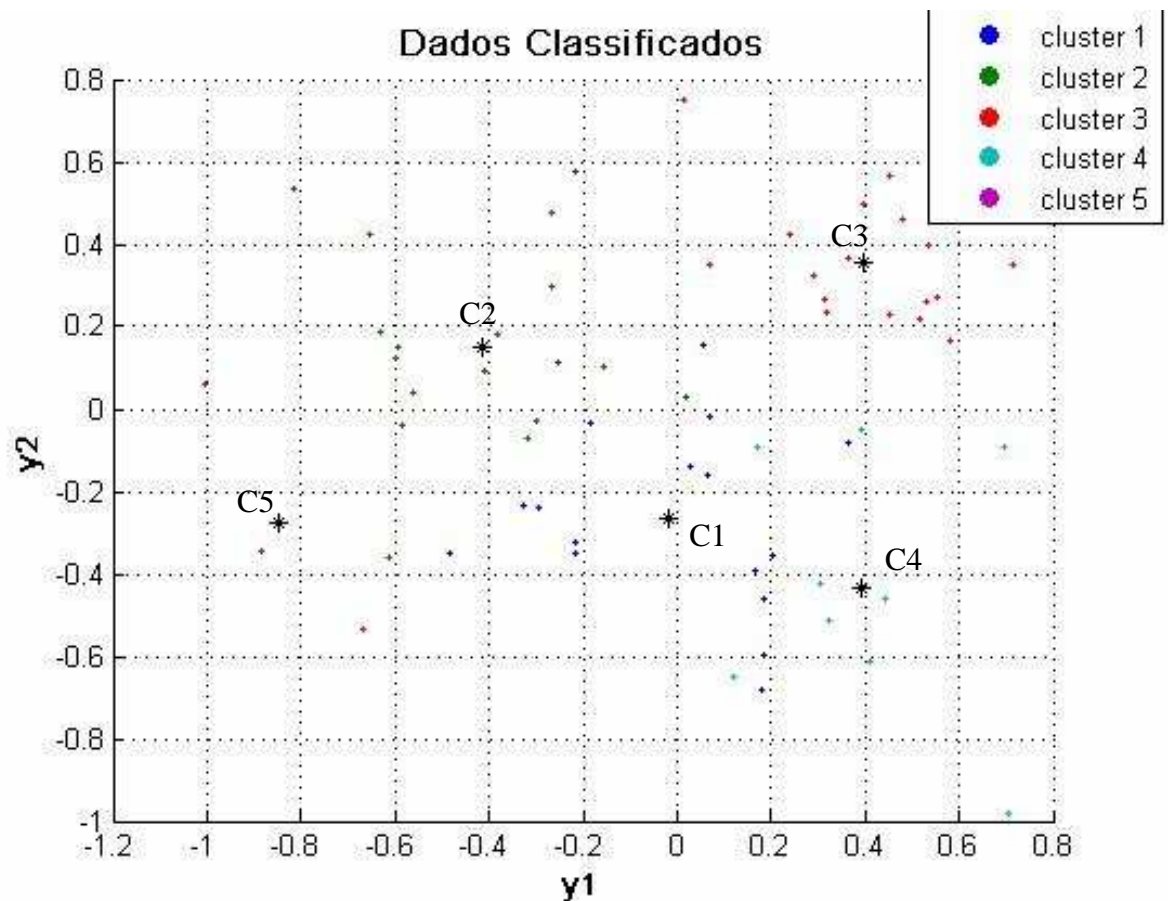


Figura 15: Gráfico de distribuição dos grupos para os parâmetros de maturação selecionados de acordo com os centros dos grupos e as suas cardinalidades Y1 e Y2.



Figura 16: Gráfico de distribuição dos grupos em porcentagem para os parâmetros de maturação selecionados.

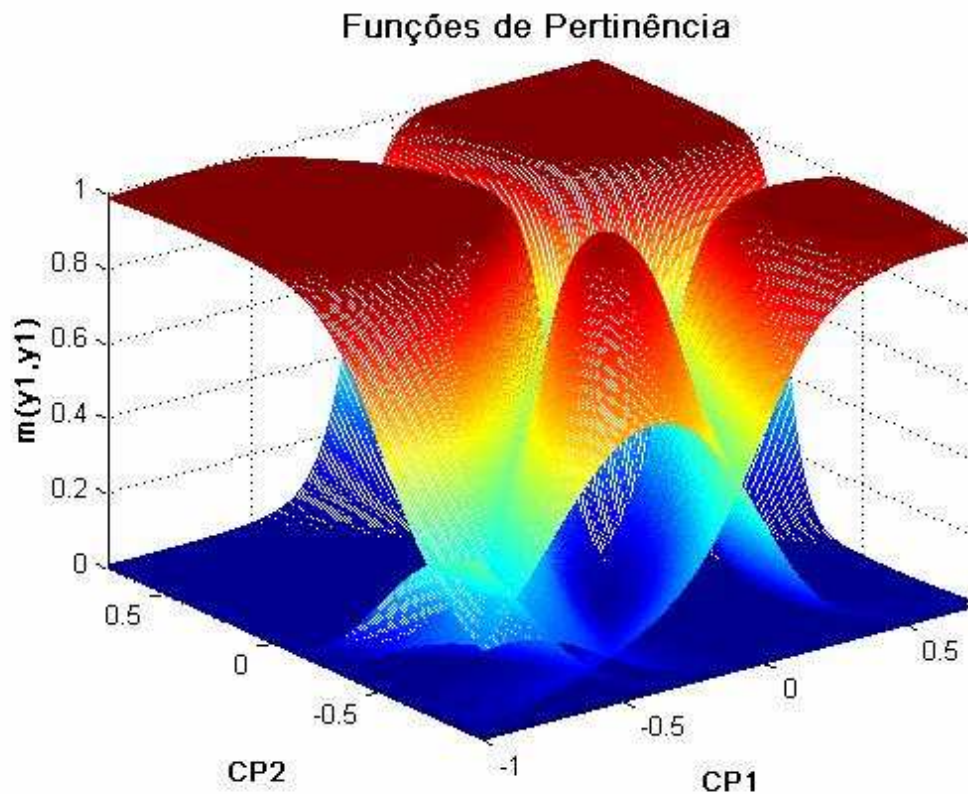


Figura 17: Gráfico de função de pertinência para os parâmetros de maturação, usando *FCM*, com índice de “fuzzificação”  $m=1,2$  e, componentes principais CP1 e CP2.

Como já discutido no capítulo de fundamentação teórica (capítulo 4), e observando a tabela 4, com o aumento do valor do parâmetro de *fuzzificação* ( $m$ ) o valor do índice diminui, pois mais imprecisas são as partições geradas. Nas Figuras 18A e 18B observa-se a disposição das funções de pertinência para diferentes valores de  $m$ ,  $m=1,2$  e  $m=1,8$ , respectivamente.

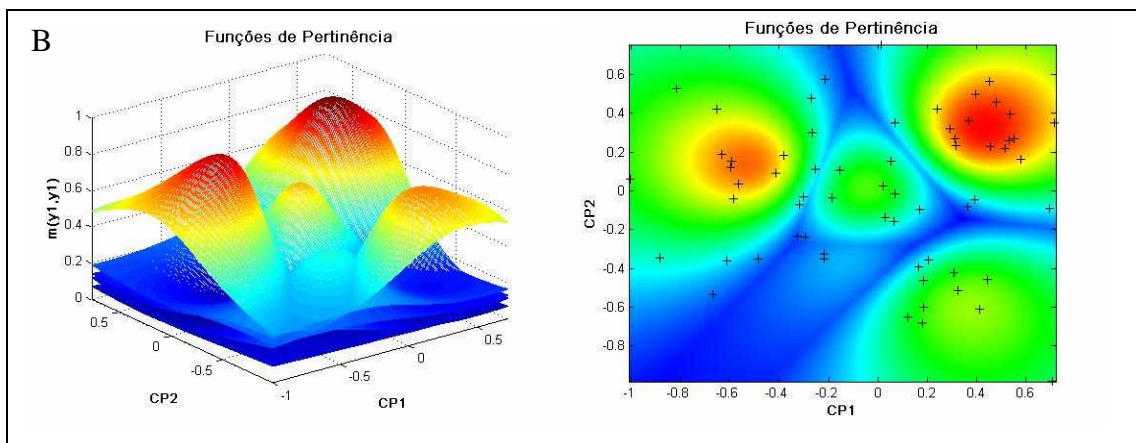
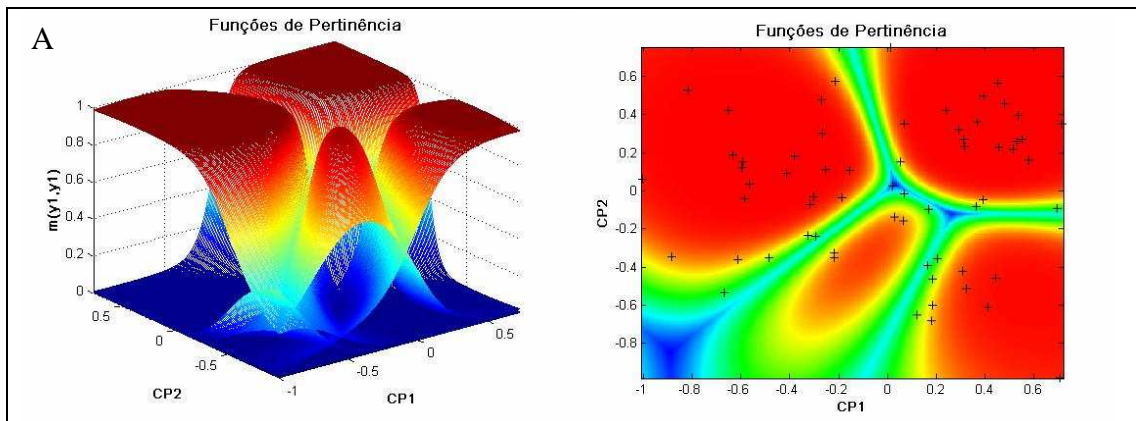


Figura 18: Gráficos de funções de pertinência para os parâmetros de maturação, usando *FCM*: (A)  $m = 1,2$ ; (B)  $m = 1,8$ .

A interpretação de cada *cluster* pode ser realizada pelos centros de suas respectivas coordenadas. Observando a Tabela 5, podem-se fazer algumas considerações relacionadas a seguir.

Tabela 5: Valores dos Parâmetros de Maturação nos Centros de Grupos.

Parâmetros	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
MAI	1,68	1,67	1,69	1,67	1,73
MDI	1,44	1,50	1,50	1,44	1,45
<b>3+4 metildiamantanos</b>	84,64	98,80	119,58	111,37	138,93
<b>Estigmastano</b>	108,96	105,10	99,00	129,02	78,06
$\beta\beta/(\beta\beta+\alpha\alpha)(C29)$	1,52	1,58	1,61	1,66	1,50
$S/(S+R) C29$	1,51	1,48	1,50	1,56	1,53

Os parâmetros **3+4- metildiamantanos** e **Estigmastano** são os que possuem maior ponderação em cada grupo, ou seja, são mais significativos. No grupo 1, a contribuição de Estigmastano supera a contribuição de metildiamantanos, o que pode sugerir um grupo de amostras de óleo pouco maturo, evento que também ocorre no grupo 2. Porém, este apresenta uma maior contribuição dos metildiamantanos em relação aquele. Sendo assim, o grupo 2 pode ser considerado mais evoluído termicamente que o grupo 1.

Já nos grupos 3 e 5 a contribuição dos metildiamantanos é maior que a do biomarcador estigmastano: o grupo 5 apresenta o menor valor de estigmastano e o maior valor de metildiamantano em relação aos demais grupos, mostrando ser assim o mais evoluído termicamente; o grupo 3 possui menor valor de metildiamantano e maior valor de estigmastano que o grupo 5, mostrando ser menos maturo que este, mas com maior maturidade que os grupos 1 e 2.

Analisando os resultados para o grupo 4, observam-se valores elevados de biomarcadores e valores elevados de diamantóides. Conseqüentemente, suspeita-se de mistura de óleos em diferentes pulsos de migração com níveis diferenciados de maturação (DAHL *et al.*, 1999).

Conforme os resultados discutidos anteriormente, os parâmetros 3+4- metildiamantanos e Estigmastano apresentam uma contribuição muito superior aos demais parâmetros. Sendo assim, considerados parâmetros decisivos para o agrupamento das amostras do estudo. Estes resultados, vão de encontro com os

resultados obtidos pela “análise tradicional”, ou seja, mais uma vez a utilização dos parâmetros envolvendo diamantóides foi decisivo para o entendimento do comportamento das amostras em estudo.

## **7.2) Parâmetros de Ambientes Depositionais (Origem)**

### **7.2.1) Interpretação dos resultados pela caracterização de Biomarcadores e Diamantóides**

Como mencionado no capítulo 3, os biomarcadores permitem a caracterização de ambientes deposicionais das rochas geradoras. Sendo assim, razões de biomarcadores consideradas como indicadores de origem são correlacionadas para distinguir tais ambientes, Tabela 6.

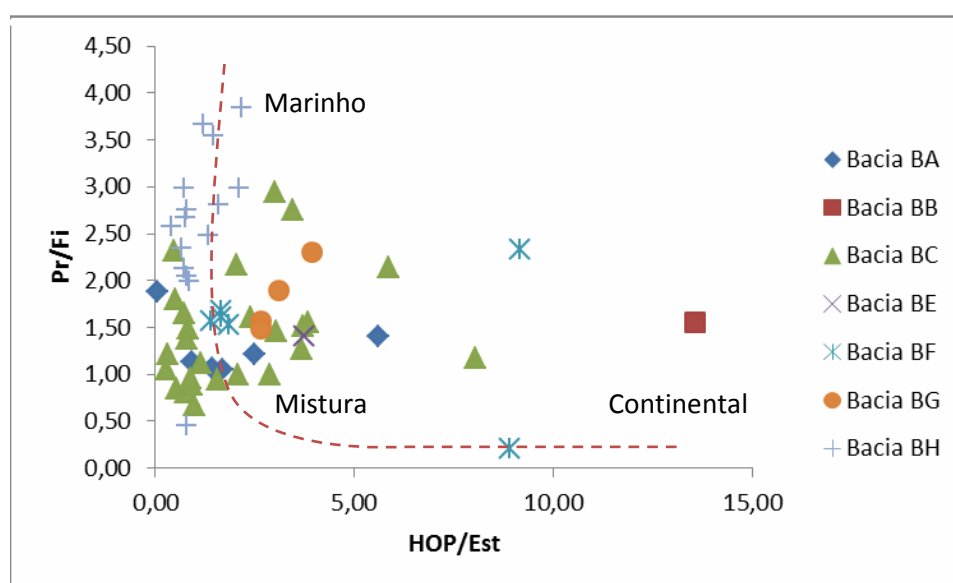
É importante ressaltar que são desconhecidas as bacias sedimentares brasileiras e os poços de origem dos óleos analisados, sabe-se apenas que as amostras da bacia BH pertencem à bacia colombiana *Llanos Orientales*.

Tabela 6: Valores dos parâmetros de biomarcadores dependentes de ambientes deposicionais para as amostras de óleo estudadas, sendo: as amostras em verdes pertencem a Bacias Sedimentares Brasileiras, e as amostras destacadas em laranja, Bacia Sedimentar Colombiana.

<b>AMOSTRAS</b>	<b>Pr/Fi</b>	<b>Pr/nC17</b>	<b>Fi/nC18</b>	<b>Hop/Est</b>	<b>TTP/DIA</b>	<b>Tr25/Tr26</b>
<b>BA1</b>	1,88	0,96	1,04	0,05	0,40	1,11
<b>BA2</b>	1,13	0,88	0,92	0,91	0,40	1,15
<b>BA3</b>	1,05	0,70	0,95	1,69	0,30	0,72
<b>BA4</b>	1,06	0,73	0,98	1,43	0,50	0,59
<b>BA5</b>	1,22	0,66	0,86	2,48	0,30	0,62
<b>BA6</b>	1,40	2,02	0,91	5,61	3,80	1,12
<b>BB1</b>	0,00	0,00	0,00	3,30	0,60	0,89
<b>BB2</b>	1,56	0,53	0,36	13,57	0,50	1,38
<b>BB3</b>	0,00	0,00	0,00	5,81	1,96	1,13
<b>BC1</b>	1,22	0,98	1,01	0,32	0,22	0,96
<b>BC2</b>	0,85	0,90	1,51	0,54	0,25	1,12
<b>BC3</b>	1,05	0,80	1,12	0,28	0,28	1,00
<b>BC4</b>	1,80	0,24	0,14	0,49	2,10	0,90
<b>BC5</b>	1,00	0,56	0,74	2,89	0,23	1,00
<b>BC6</b>	2,31	1,75	0,84	0,46	0,12	0,90
<b>BC7</b>	1,61	0,54	0,38	2,40	0,86	0,60
<b>BC8</b>	1,65	0,36	0,24	0,73	1,81	0,30
<b>BC9</b>	1,49	0,34	0,24	0,81	2,57	1,00
<b>BC10</b>	2,14	0,50	0,27	5,86	4,90	0,90
<b>BC11</b>	1,38	0,45	0,35	0,79	0,18	0,80
<b>BC12</b>	2,94	0,57	0,50	3,00	0,61	0,90
<b>BC13</b>	1,27	0,90	0,95	3,68	1,03	0,70
<b>BC14</b>	0,80	0,47	0,91	0,77	0,06	0,60
<b>BC15</b>	0,67	0,71	1,58	0,97	0,07	0,70
<b>BC16</b>	1,46	0,29	0,21	3,02	0,82	0,80
<b>BC17</b>	2,16	0,67	0,35	2,04	0,80	0,80
<b>BC18</b>	0,94	0,93	1,15	1,56	0,15	0,90
<b>BC19</b>	2,75	0,39	0,16	3,45	0,81	1,10
<b>BC20</b>	1,56	0,26	0,17	3,84	1,32	1,40
<b>BC21</b>	1,18	0,23	0,22	8,03	0,19	0,60
<b>BC22</b>	1,00	0,39	0,52	2,09	0,25	0,57
<b>BC23</b>	1,51	0,29	0,22	3,71	0,59	0,64
<b>BC24</b>	1,12	0,36	0,39	1,14	0,32	0,61
<b>BC25</b>	0,89	0,50	0,88	0,91	0,08	0,62
<b>BC26</b>	0,95	0,65	1,16	0,88	0,09	0,84
<b>BE1</b>	0,00	0,00	0,00	3,86	0,00	0,65
<b>BE2</b>	1,40	1,47	0,89	3,74	0,00	0,60
<b>BF1</b>	1,52	0,98	0,76	1,84	0,19	0,74
<b>BF2</b>	1,61	1,23	1,09	1,66	0,19	0,55

<b>BF3</b>	1,68	0,91	0,92	1,66	0,42	0,54
<b>BF4</b>	2,33	0,76	0,32	9,16	25,94	1,34
<b>BF5</b>	0,21	0,07	0,34	8,91	6,60	0,63
<b>BF6</b>	1,57	1,00	1,21	1,40	0,29	0,64
<b>BG1</b>	1,57	0,92	0,62	2,64	0,57	0,61
<b>BG2</b>	1,48	1,25	1,01	2,65	0,19	0,67
<b>BG3</b>	1,90	1,03	0,74	3,09	0,46	0,73
<b>BG4</b>	2,58	1,46	1,16	0,00	0,00	0,80
<b>BG5</b>	2,30	1,22	0,71	3,93	0,53	1,00
<b>BH1</b>	2,99	2,03	5,53	2,10	0,96	0,96
<b>BH2</b>	3,84	1,66	5,97	2,17	0,76	0,86
<b>BH3</b>	3,54	1,96	6,55	1,45	1,77	1,28
<b>BH4</b>	3,67	1,92	6,41	1,20	1,93	1,17
<b>BH5</b>	2,66	1,61	3,34	0,77	2,83	1,28
<b>BH6</b>	2,35	1,66	3,26	0,65	3,25	1,08
<b>BH7</b>	2,81	1,51	3,63	1,58	2,76	1,32
<b>BH8</b>	2,12	1,75	3,08	0,73	3,34	1,27
<b>BH9</b>	2,48	1,60	3,29	1,32	3,69	0,88
<b>BH10</b>	2,10	1,61	2,97	0,32	0,00	1,47
<b>BH11</b>	0,00	2,04	4,12	1,93	1,02	0,94
<b>BH12</b>	2,05	0,92	2,17	0,78	1,02	1,01
<b>BH13</b>	0,45	0,76	0,48	0,41	0,87	1,10
<b>BH14</b>	0,00	0,00	0,00	0,37	0,85	0,98
<b>BH15</b>	0,00	0,00	0,00	0,42	1,11	0,51
<b>BH16</b>	2,75	1,84	4,50	0,86	1,37	0,86
<b>BH17</b>	0,00	0,00	0,00	0,78	0,51	0,90
<b>BH18</b>	2,58	1,44	3,44	0,71	1,18	1,09
<b>BH19</b>	2,00	1,89	3,41	0,83	1,32	1,12
<b>BH20</b>	2,99	1,53	4,38	0,55	0,74	0,93
<b>BH21</b>	0,00	0,00	0,00	1,18	0,00	0,00

Uma maneira de observarmos a origem da matéria orgânica é através da construção do gráfico Hopanos/Esteranos *versus* Pristano/Fitano, Figura 19. Pode-se observar que a maioria das amostras da bacia BH possui características marinhas acentuadas. Amostras das bacias BA, BC, BE, BF, BG apresentam altos valores para ambas as razões, sendo considerado mistura. Algumas amostras provenientes das bacias BB, BC e BF se destacam das demais, apresentando características terrestres mais acentuadas.



Pelos parâmetros avaliados no gráfico da Figura 20 (Fitano/ $nC_{18}$  versus Pristano/ $nC_{17}$ ), pode-se observar que as amostras da bacia BH e algumas amostras da bacia BC apresentam predominância de características marinhas. Uma amostra de óleo da bacia BA apresenta altos valores de ambas razões, sendo caracterizada como mistura. As amostras da bacia BG apresentam características terrestres.

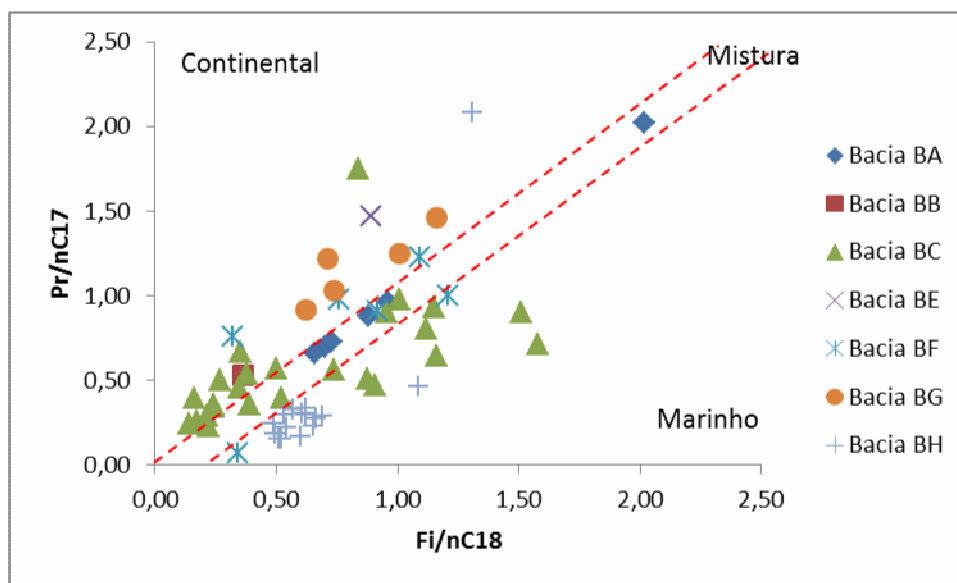


Figura 20: Correlação entre as razões Fitano/ $nC_{18}$  versus Pristano/ $nC_{17}$  para as amostras deste estudo.

Outro importante parâmetro de origem é a razão dos Diasteranos  $C_{27}$   $13\beta(H),17\alpha(H)$  (20S+20R)/Poliprenóides Tetracíclicos (TPP's). Estudos mostram que ambientes com altas concentrações relativas de TPP's em óleos e rochas geradoras indicam deposição em ambientes lacustres de água (doce/salobra), enquanto baixas concentrações sugerem deposição em ambientes marinhos (HOLBA *et al.*, 2003). Analisando o gráfico da Figura 21, pode-se perceber que a maioria das amostras se comporta como mistura, com um caráter marinho acentuado. Algumas amostras das bacias BA, BB, BC e BF possuem uma maior predominância de características terrestres, e outras amostras da bacia BC apresentam características marinhas bem acentuadas.

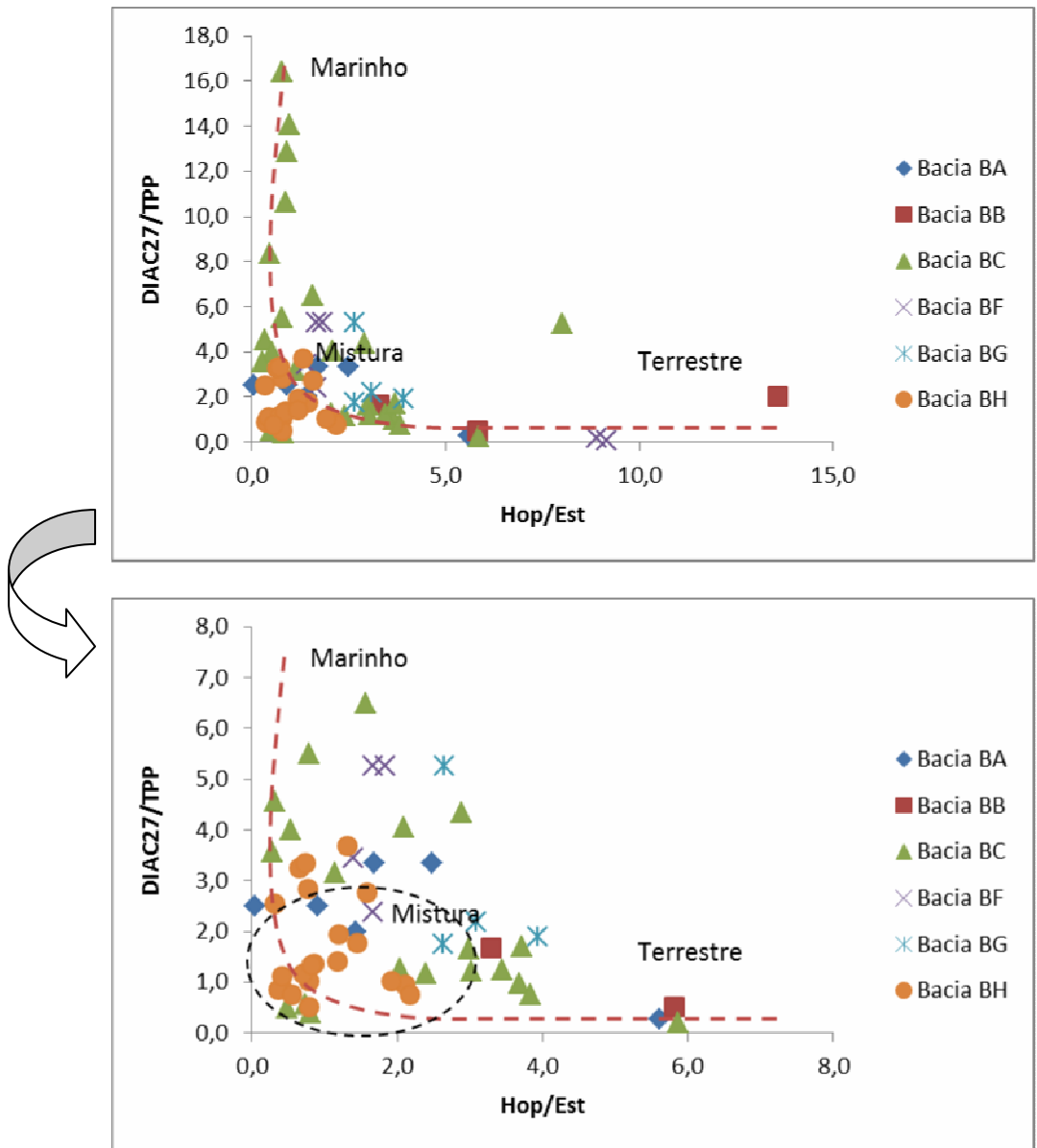


Figura 21: Correlação entre as razões Hopanos/Esteranos e DIA C27/TPP para as amostras deste estudo.

A razão Terpano Tricíclico  $C_{25}/C_{26}$  é outro parâmetro importante que nos mostra a influência dos ambientes deposicionais. No gráfico da Figura 22, podemos observar esta razão contra a razão Hop/Est. A concentração dos compostos tricíclicos  $C_{25}$  é maior que a dos compostos tricíclicos  $C_{26}$  ( $C_{25}/C_{26} > 1$ ) em alguns óleos provenientes das bacias BH, BC e BA, indicando uma contribuição marinha.

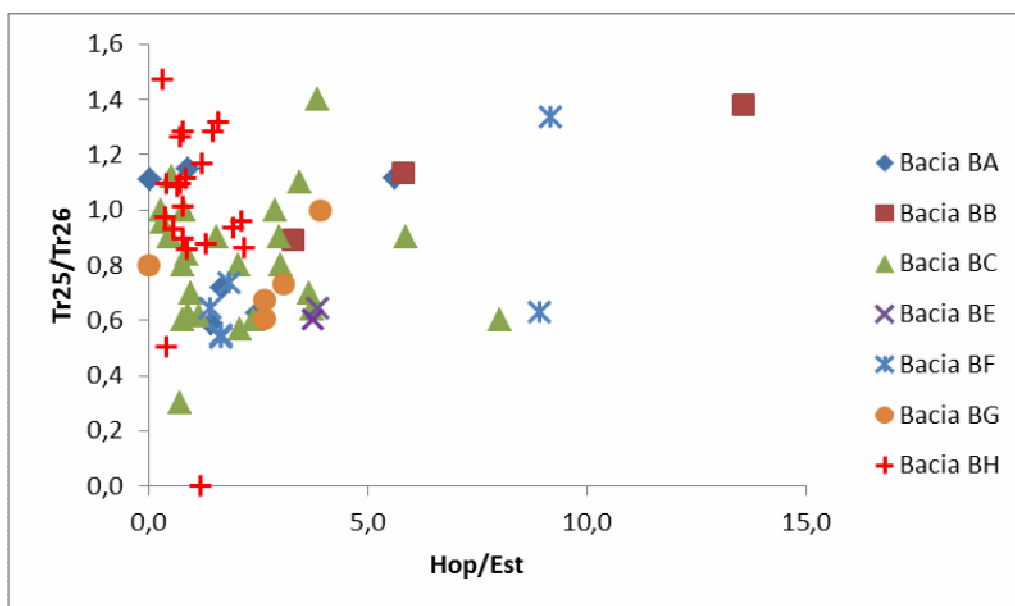


Figura 22: Gráfico de correlação entre as razões Hop/Est versus Tr25/Tr26.

### 7.2.2) Análise de agrupamento FCM – *Fuzzy c-means*

A análise de agrupamento foi realizada utilizando como parâmetros as razões Hop/Est, DiaC27/TPP e TrC25/TrC26. Antes de realizar o processo de clusterização foi efetuado o pré-tratamento, analisando a presença de valores ausentes e correlações entre as variáveis, seguindo o procedimento de *Data Mining*. Pela análise de *cluster* as amostras foram diferenciadas em quatro grupos (Tabela 7), que será discutido adiante.

Com a finalidade de determinar o melhor número possível de agrupamentos foi utilizado o índice de validação *PBM*, detalhado no capítulo 4 (PAKHIRA *et al.*, 2004). Analisando a Tabela 7, que apresenta diferentes valores de *m* para cada agrupamento, o índice de validação indica quatro grupos.

Tabela 7: Determinação do número de grupos.

<b>Grupos</b>	<b>m=1,2</b>	<b>m=1,4</b>	<b>m=1,6</b>	<b>m=1,8</b>	<b>m=2,0</b>
2	0,065	0,051	0,042	0,033	0,026
3	0,165	0,261	0,176	0,090	0,043
<b>4</b>	<b>0,431</b>	<b>0,364</b>	<b>0,235</b>	<b>0,153</b>	<b>0,109</b>
5	0,316	0,325	0,190	0,107	0,104

Nas Figuras 23, 24, 25 e 26, podemos observar a distribuição e a disposição desses grupos. A partir das figuras 24 e 25, pode-se observar que o grupo 4 apresenta o menor número de amostras envolvidas, e o grupo 1 o maior número.

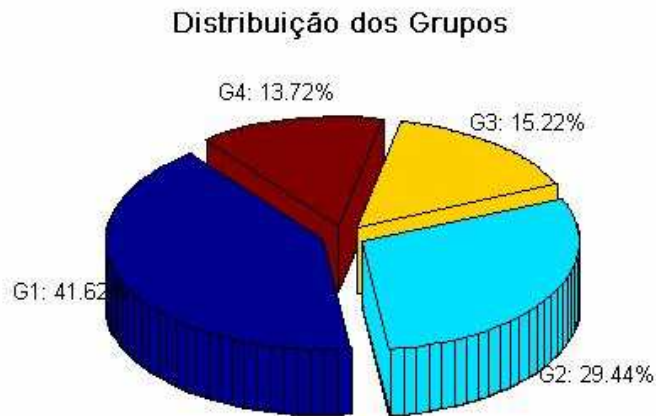


Figura 23: Gráfico de distribuição dos grupos em porcentagem para os parâmetros de origem seleccionados.

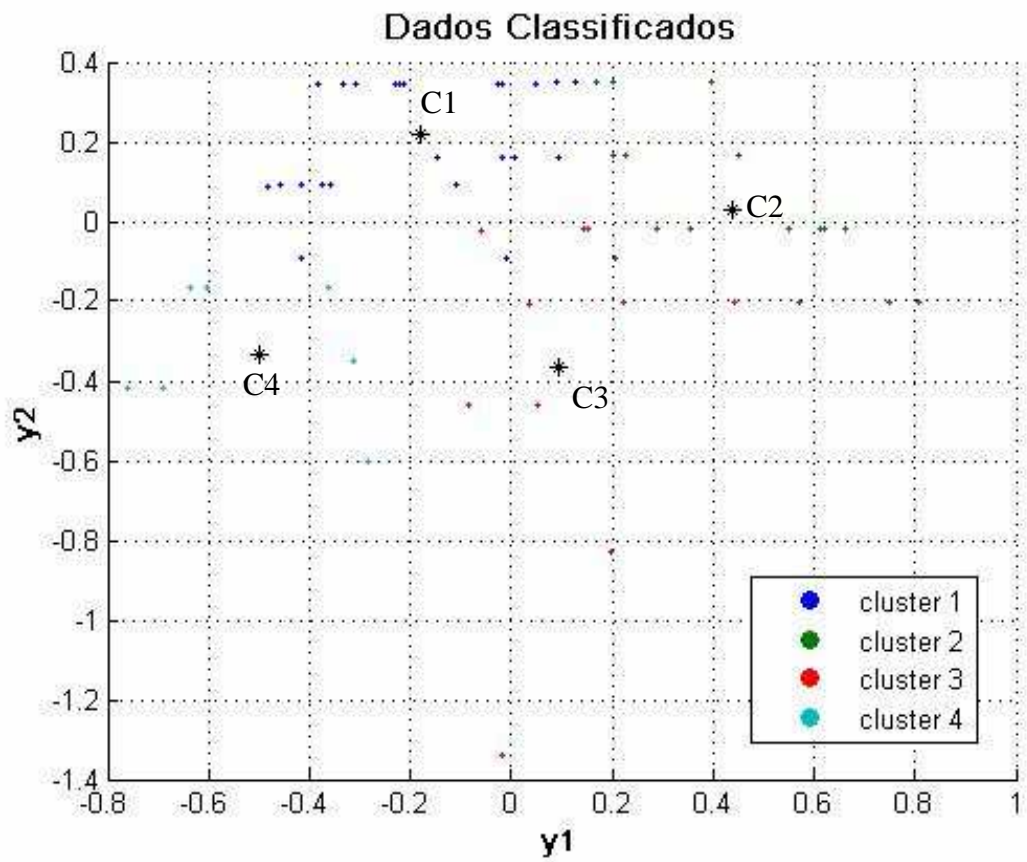


Figura 24: Gráfico de distribuição dos grupos para os parâmetros de origem seleccionados de acordo com os centros dos grupos e as suas cardinalidades  $Y_1$  e  $Y_2$ .

Nas figuras 25 e 26, observa-se a disposição dos agrupamentos para 4 grupos, obedecendo à função de pertinência *fuzzy* ( $m=1,2$ ).

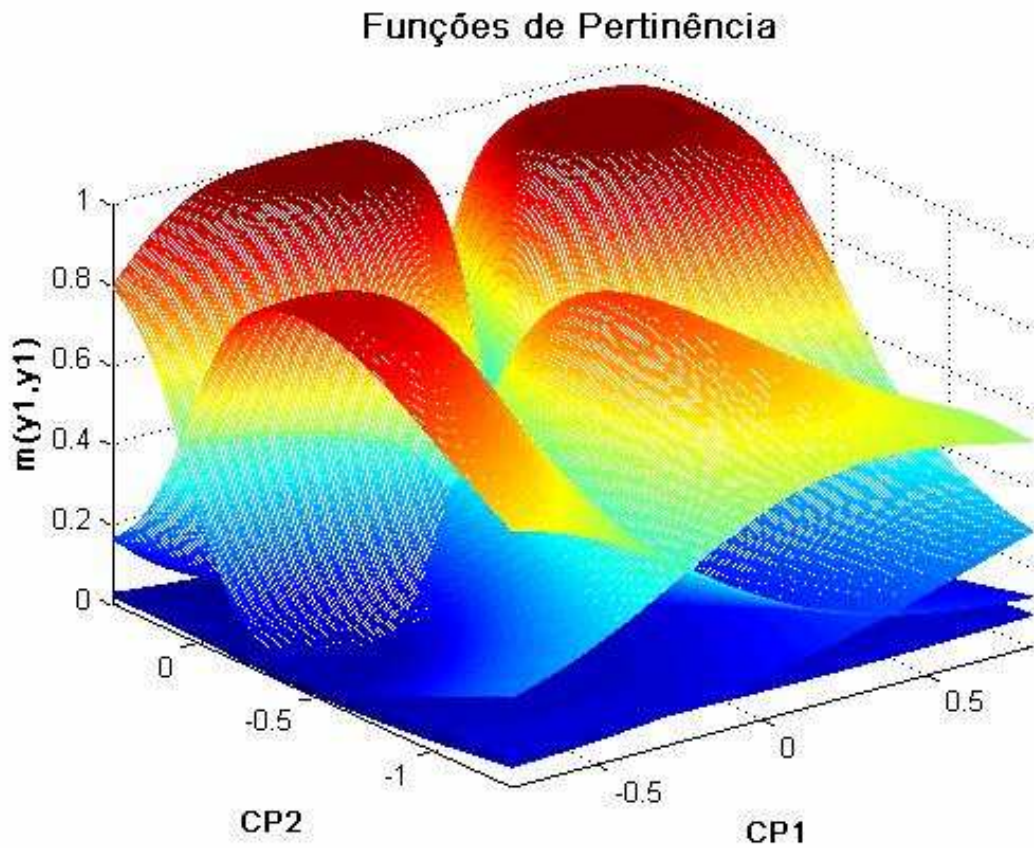


Figura 25: Gráfico de função de pertinência (3D) para os parâmetros de origem, usando *FCM*, com índice de "fuzzificação"  $m=1,2$  e, componentes principais CP1 e CP2.

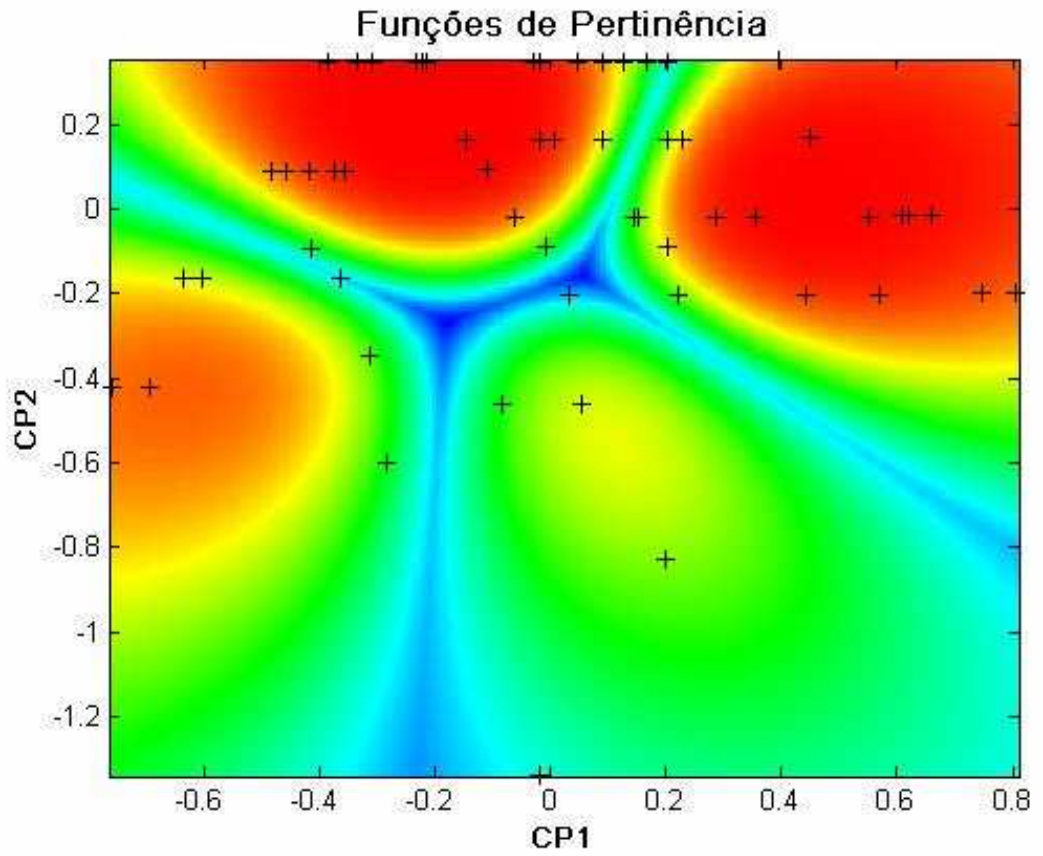


Figura 26: Gráfico de função de pertinência (2D) para os parâmetros de origem, usando *FCM*, com índice de “fuzzificação”  $m=1,2$  e, componentes principais 1 e 2 (CP1 e CP2).

A interpretação de cada cluster pode ser realizada pelos centros de suas respectivas coordenadas, apresentado na Tabela 8. Observando a Tabela 8 e Figura 27, podem-se fazer algumas considerações relacionadas a seguir.

Tabela 8: Valores dos Parâmetros de Origem nos Centros de Grupos.

Valores dos Parâmetros de Origem nos Centros de Grupos				
	Grupo1	Grupo 2	Grupo 3	Grupo 4
<b>Hop/Est</b>	0,38	4,14	6,86	1,91
<b>TPP/DIA</b>	2,80	5,36	5,59	2,60
<b>Tr25/Tr26</b>	1,95	2,00	2,08	2,04

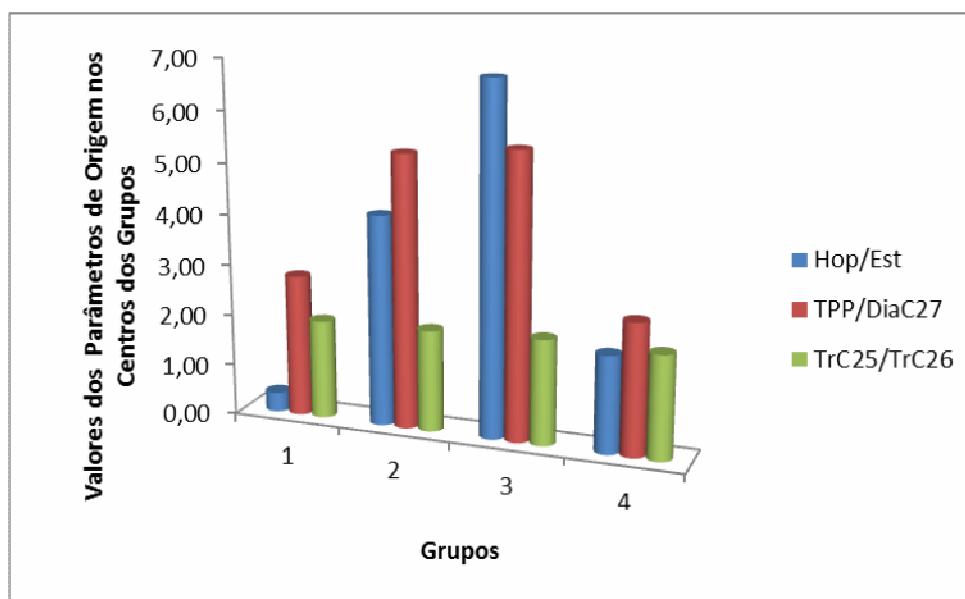


Figura 27: Gráfico em barra relacionando os Valores dos Parâmetros de Origem nos Centros de Grupos e os 4 Grupos estimados.

### **Grupo 1 –**

Analisando os valores dos parâmetros de origem para o grupo 1, observa-se que a contribuição da razão Hop/Est é a menor e, a razão TPP/DIA é a maior. Como discutido anteriormente, ambientes com baixa contribuição Hop/Est indicam deposição de matéria orgânica marinha, a partir dos dados da tabela 8, pode-se sugerir de uma forte contribuição marinha para as amostras pertencentes a este grupo.

### **Grupo 2 –**

Este grupo de amostras apresenta uma forte contribuição das razões Hop/Est, TPP/DIA, sendo os valores da razão TPP/DIA superiores. Logo, a partir desses dados pode-se sugerir que estes valores são indicativos de uma mistura (contendo também caráter continental) para estas amostras.

### **Grupo 3 –**

No grupo 3, a contribuição das razões TPP/DIA e Hop/Est é muito maior em relação aos demais grupos. O valor da razão Hop/Est é superior ao valor da razão TPP/Dia. Estes resultados indicam que os óleos com altas razões TPP/Dia e Hop/Est, estão relacionados a uma mistura contendo, também, uma contribuição marinha.

### **Grupo 4 –**

O grupo 4, assim como o grupo 1, também possui valores mais baixos que os grupos 2 e 3. Porém a contribuição da razão Hop/Est é bem maior em relação ao grupo 1, o que pode caracterizar um caráter continental.

Podemos concluir que a maioria das amostras está localizada nos grupo 1 e 2, ou seja, podem ser caracterizadas por uma contribuição marinha (grupo 1) e uma mistura contendo também caráter continental (grupo 2). Sendo assim, os resultados obtidos a partir da *clusterização fuzzy* se aproximam dos resultados das correlações envolvendo biomarcadores, como discutidos no item 6.2.1.

## Capítulo 8.

### Conclusão

Neste estudo buscou-se a caracterização geoquímica quanto a origem e evolução térmica de 69 amostras de óleo, provenientes de algumas bacias sedimentares brasileiras e da bacia sedimentar colombiana *Llanos Orientales*, a partir de parâmetros indicativos de maturação térmica e de ambiente deposicional das rochas geradoras, envolvendo razões de biomarcadores e diamantóides. Para tanto, foram construídos diversos gráficos de correlação envolvendo os parâmetros geoquímicos - metodologia tradicional – e, buscando uma melhor visualização e interpretação dos resultados, os dados foram avaliados por análise de agrupamento *fuzzy*.

Esta metodologia foi desenvolvida usando dados gerados a partir de trabalhos (teses e dissertações) desenvolvidos por estudantes desta instituição de ensino e pesquisa.

Com relação ao grau de maturação, observou-se uma grande diversidade. Entretanto, de acordo com o agrupamento *fuzzy* foram observados cinco grupos. Para tal, foi fundamental a integração dos parâmetros de biomarcadores e diamantóides para a interpretação do comportamento das amostras de óleo. Os parâmetros sugeridos por DAHL et al. (1999), ou seja, o biomarcador estigmastano e a soma dos diamantóides 3-+4- metildiamantanos, foram os que possuíram maior ponderação em cada grupo, ou seja, são os mais significativos. Para o grupo 1, a contribuição de estigmastano superou a contribuição de metildiamantanos, sugerindo um grupo de amostras de óleo pouco evoluídas termicamente, evento que também ocorreu no grupo 2. Porém, este apresentou uma maior contribuição dos metildiamantanos, sendo considerado mais evoluído termicamente que o grupo 1. Os grupos 3 e 5, apresentaram uma contribuição dos metildiamantanos superior a do biomarcador estigmastano, sendo o grupo 5 caracterizado como o mais evoluído termicamente pois apresentou o menor valor de estigmastano e o maior valor de metildiamantano em relação aos demais grupos; o grupo 3 apresentou menor valor de metildiamantano e maior valor de estigmastano que o grupo 5, mostrando ser menos maturo que este,

mas com maior maturidade que os grupos 1 e 2. O grupo 4 foi caracterizado como mistura de óleos em diferentes pulsos de migração com níveis diferenciados de maturação, pois apresentou elevados valores de estigmastano e diamantóides.

De acordo com o agrupamento *fuzzy* observou-se que as amostras puderam ser divididas em quatro grupos principais quanto à origem: grupo 1 possui uma contribuição marinha; grupo 2 indicativos de uma mistura (contendo também caráter continental); grupo 3 possui características de uma mistura contendo, também, uma contribuição marinha e grupo 4 possui características continentais.

Para tal, os parâmetros TTP/Dia27 e Hop/Est foram os mais significativos. O parâmetro Tr25/Tr26 apresentou contribuições semelhantes para os grupos.

Os resultados obtidos a partir da *clusterização fuzzy* se aproximaram dos resultados das correlações tradicionais envolvendo biomarcadores e diamantóides.

Pode-se concluir que para uma avaliação mais detalhada de um conjunto elevado de parâmetros como é o caso das razões de biomarcadores e diamantóides para óleos, o uso da análise de agrupamento *fuzzy* foi uma ferramenta satisfatória. Esta, embora já venha sendo utilizada em diversas áreas e em alguns centros de pesquisa da indústria de petróleo, ainda não é uma ferramenta de rotina.

A inclusão de parâmetros de diamantóides na avaliação dos óleos foi de real importância no refinamento dos dados e melhor diferenciação dos óleos quanto à maturação.

## Capítulo 9.

### Recomendações

Para uma melhor avaliação do método, recomenda-se aplicar esta técnica em um banco de dados com um número superior de amostras analisadas e, se possível, pertencentes a uma mesma bacia sedimentar, possuindo também outros parâmetros de ambiente deposicional.

Recomenda-se avaliar a técnica aplicada neste trabalho em relação à influência da biodegradação, outro processo geoquímico, mediado por microorganismos que degrada parcialmente os óleos.

Apesar de o número de agrupamentos para os parâmetros geoquímicos envolvidos apresentar bem a variabilidade dos dados em estudo, recomenda-se testar outros índices de validação de grupos, visto que já existem indicadores mais recentes voltados para a lógica *fuzzy* e usando algoritmos genéticos.

## Capítulo 10.

### Referências Bibliográficas

- AZEVEDO, D.A., TAMANQUEIRA, J.B., DIAS, J.C.M., CARMO, A.P.B., LANDAU, L., GONÇALVES, F.T.T., 2008. “*Multivariate statistical analysis of diamondoid and biomarker data in selected oil samples from Brazilian basins*”. Fuel, v.87, p. 2122-2130.
- AQUINO NETO, F.R., TRENDEL, J.M., RESTLE, A., CONNAN, J., ALBRECHT, P.A., 1983. “*Occurrence and formation of tricyclic and tetracyclic terpanes in sediments and petroleum*”, In: Bjoroy, M. et al (eds), *Advances in Organic Geochemistry* 1981, Wiley, p. 659-667.
- BEZDEK, J.C.; PAL, S.K., 1992. “*Fuzzy models for pattern recognition: method that search for structures in data*”. Piscataway: IEEE Press.
- BIZZI, A.L., SCHOBENHAUS, C., VIDOTTI, M.B., GONÇALVES, H.J., 2003. “*Geologia, Tectônica e Recursos minerais do Brasil. Serviço Geológico do Brasil (CPRM)*”. Livro disponível em formato pdf no site: <http://www.cprm.gov.br/>.
- BOBROWSKI, L., BEZDEK, J.C., 1991. “C-Means Clustering with the  $l_1$  and  $l_\infty$  Norms”. In *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, n. 3, Maio/Junho, p. 545-554.
- BOENTE, A.N.P., OLIVEIRA, F.S.G., ROSA, J.L.A., 2007. “Utilização de Ferramenta de KDD para Integração de Aprendizagem e Tecnologia em Busca da Gestão Estratégica do Conhecimento na Empresa”. Anais do Simpósio de Excelência em Gestão e Tecnologia, v.1, p. 123-132.
- BOENTE, A.N.P, GOLDSCHMIDT, ESTRELA, V.V., 2008. “Uma Metodologia de Suporte ao Processo de Descoberta de Conhecimento em Bases de Dados”. In: V

*Simpósio de Excelência em Gestão e Tecnologia*, V SEGeT, , v.1, p. 4-5, Resende – RJ.

CALDAS, B. 2009. “*Diamantóides como ferramenta na avaliação de óleos em bacias sedimentares brasileiras*”. Monografia, IQ/UFRJ, Rio de Janeiro, RJ, Brasil.

CALIŃSKI, T., HARABASZ, J., 1974. “A dendrite method for cluster analysis”, *Communications in Statistics*, v. 3, p. 1–27.

CHEN, J., FU, J., SHENG, G., LIU, D., ZHANG, J., 1996. “Diamondoid hydrocarbon ratios: novel maturity indices for highly mature crude oils”, *Organic Geochemistry*, v.25, p. 179–190.

CHEN, J., DENG, C., LIANG, D., WANG, X., ZHONG, N., SONG, F., SHI, X., JIN, T., XIANG, S., 2003. “Mixed oils derived from multiple source rocks in the Cainan oil field, Junggar basin, northwest china. part I: genetic potential of source rocks, features of biomarkers and oil sources of typical crude oils”, *Organic Geochemistry*, v.34, p. 889–909.

CHEN, J., DENG, C., LIANG, D., WANG, X., ZHONG, N., SONG, F., SHI, X., JIN, T., XIANG, S., 2003. “Mixed oils derived from multiple source rocks in the Cainan oil field, Junggar basin, northwest china. part 2: artificial mixing experiments on typical crude oils and quantitative oil-source correlation”. *Organic Geochemistry* 34, p. 911–930.

DAHL, J., MOLDOWAN, M., PETERS, K., CLAYPOOL, G., ROONEY, M., MICHAEL, G., MELLO, M., KOHNEN, M., 1999. “Diamondoid hydrocarbons as indicators of natural cracking”, *Nature*, v. 399, p.54-56.

DIDYK, B.M., SIMONEIT, B.R.T., BRASSEL, S.C., EGLINTON, G., 1978. “Organic geochemical indicators of paleoenvironmental conditions of sedimentation”, *Nature*, v.272, p. 216-222.

- EVSUKOFF, A. G., GONÇALVES, F. T. T., BEDREGAL, R. P., EBECKEN, N. F. F., 2004. “Fuzzy Classification of Surface Geochemistry Data Applied to Determination of HC Anomalies”, IEE Joint 2<sup>nd</sup> International Conference on Soft Computing and Intelligent Systems.
- EVSUKOFF, A.G., FAIRBAIRN, E.M.R., FARIA, E.F., SILVOSO, M.M., TOLEDO, R.D., 2006. “Modeling adiabatic temperature rise during concrete hydration: A data mining approach”. *Computers and Structures*, v.84, p. 2351–2362.
- FANG, C., XIONG, Y., LIANG Q., LI Y., 2012. “Variation in abundance and distribution of diamondoid during oil cracking”, *Organic Geochemistry* 47, p. 1-8.
- FAYYAD, U.M., PIATETSKY-SHAPIRO, G., SMYTH, P., UTHURUSAMY, R., 1996. *Advances in knowledge discovery and data mining*. Massachusetts, MIT Press.
- FARRIMOND, P., TAYLOR, A., TELNAES, N., 1998. “Biomarker maturity parameters: the role of generation and thermal degradation”, *Organic Geochemistry*, v.29, p. 1181-1197.
- GAN, G., MA, C., WU, J., 2007. *Data Clustering: Theory, Algorithms, and Applications*. First ed. Alexandria, Virginia, ASA-SIAM.
- GRICE, K., ALEXANDER, R., KAGI, R., 2000, “Diamondoid hydrocarbons as indicators of biodegradation in Australian crude oils”, *Organic Geochemistry*, v. 31, p.67-73.
- HOLBA, A.G., TEGELAAR, E., ELLIS, L., SINGLETARY, M.S., 2000. “Tetracyclic polyprenoids: Indicators of freshwater (lacustrine) algal input”, *Geology*, v.28, p. 251- 254.

- HOLBA, A.G., DZOU, L.I., WOOD, G.D., ELLIS, L., ADAM, P., SCHAEFFER, P., ALBRECHT, P., GREENE, T., HUGHES, W.B., 2003. “Application of tetracyclic polyprenoids as indicators of input from fresh-brackish water environments”, *Organic Geochemistry*, v. 34, p. 441-469.
- HOPPNER, F., KLAWONN, F., KRUSE, R., RUNKLER, T., 2000. *FUZZY CLUSTER ANALYSIS: Methods for classification, data analysis and image recognition*. 2 ed. New York, NY 10158-0012, USA, John Wiley & Sons - LTD.
- HÜLLERMEIER, E., 2005. Fuzzy methods in machine learning and data mining: Status and prospects. *Fuzzy Sets Syst.*, v. 156, p. 387–407.
- JACKSON, J. 2002. “Data mining: a conceptual overview”. *Communications of the Association for Information Systems*. v. 8, p. 267-296.
- JAIN, A.K.; MURTY, M.N.; FLYNN, P.J., 1999. “Data Clustering: A Review.” *ACM Computing Surveys (CSUR)*. ACM Press, New York, vol. 31 (3), p. 264-323.
- JESUINO, L.S., 2005, *Aplicação dos Diamantóides na Avaliação Geoquímica de óleos em Bacias Sedimentares Brasileiras*. Dissertação de MSc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- KLIR, G.J., YUAN, B., 1995. *Fuzzy Sets and Fuzzy Logic –Theory and Applications*. New Jersey, Prentice Hall.
- MILANI, E.J., BRANDÃO J.A.S.L., ZALÁN P.V., GAMBOA, L.A.P., 2000. Petróleo na margem continental brasileira: geologia, exploração, resultados e perspectivas. *Revista Brasileira de Geofísica*, v. 18, cap. 3, p. 351-396.
- MILANI, E.J., THOMAZ FILHO, A., 2000. Sedimentary basins of South America. In: Cordani, U.G.; Milani, E.J.; Thomaz Filho, A.; and Campos, D.A. (Eds.),

*Tectonic evolution of South America*. Rio de Janeiro, 31<sup>st</sup> International Geological Congress, p. 389-452.

MACEDO, D.C., MATOS, S.N., 2010. Extração de conhecimento através da Mineração de Dados. *Revista de Engenharia e Tecnologia*, v.2.

MELLO, M.R., 1988. “Geochemical and Molecular Studies of the Depositional Environments of Source Rocks and Their Derived Oils from the Brazilian Marginal Basins”. University of Bristol, Ph. D. Thesis.

MELLO, M. R., GAGLIANONE, P.C., BRASSEL, S.C., MAXWELL, J.R., 1988a. “Geochemical and Biological Marker Assessment of Depositional Environments Using Brazilian Offshore Oils”. *Marine and Petroleum Geology*, v. 5, p. 205-223.

MELLO, M.R., TELNAES, N., GAGLIANONE, P.C., CHICARELLI, M.I., BRASSEL, S.C., MAXWELL, J.R., 1988b. “Organic Geochemical Characterization of Depositional Paleoenvironments of Source Rocks and Oil in Brazilian Marginal Basins”, *Organic Geochemistry*, v.13, p.31-45.

MOLDOWAN, J.M, LEE, C.Y., WATT, D.S., 1991. “Analysis and occurrence of C<sub>26</sub>-steranes in petroleum and source rocks”, *Geochimica et Cosmochimica Acta*, v.55, p. 1065-1081.

OLIVEIRA, J.V.D., PEDRYCZ W., 2007. *Advances in Fuzzy Clustering and its Applications*. Wiley, New York.

PAKHIRA, M. K., BANDYOPADHYAY, S., MAULIK, U., 2004, “Validity index for crisp and fuzzy clusters”, *Pattern Recognition*, v. 37, n. 3, pp. 487–501. ISSN: 0031-3203.

PALACAS, J.G., ANDERS, D.E., KING, J.D., 1984. “*South Florida basin - a prime example of carbonate source rocks in petroleum*”. In: Palacas, J.G. (eds), *Petroleum Geochemistry and Source Rocks Potential of Carbonate Rocks*, American Association of Petroleum Geologists, Tulsa, Oklahoma, p. 71-96.

PETERS, K.E., MOLDOWAN, J.M., 1993. *The Biomarker Guide: Interpreting Molecular Fossils in Petroleum and Ancient Sediments*. New Jersey, Prentice-Hall Inc.

PETERS, K.E., WALTERS, C.C., MOLDOWAN, J.M., 2005. *The Biomarker Guide: Biomarkers and Isotopes in the Environment and Human History*. 2 ed, v. 1 e 2, New York, Cambridge University Press.

REZENDE, S.O., 2003. “*Mineração de dados*”. In: *Sistemas inteligentes: fundamentos e aplicações*. São Paulo, Malone, p. 307- 333.

ROBINSON, K.M., 1987. “*An overview of source rocks and oils in Indonesia*”. In: *Proceedings of the Indonesian Petroleum Association, Sixteenth Annual Convention*, Indonesian Petroleum Association, v.1, Jakarta, p. 97-122.

RULLKOTTER, J., MARZI, R., 1988. “*Natural and artificial maturation of biological markers in a Toarcian shale from northern Germany*”, In: Mattavelli, L. and Novelli, L. (eds), *Advances in Organic Geochemistry* 1987, Pergamon Press, Oxford, p. 639-645.

SANTOS, R.J., 2009. *Introdução ao Matlab*. Departamento de Matemática. Universidade Federal de Minas Gerais - UFMG. <http://www.mat.ufmg.br/~regi>

SCHULZ, L.K., WILHELMS, A., REIN, E., STEEN, R.A., 2001. “*Application of Diamondoids to Distinguish Source Rock Facies*”, *Organic Geochemistry*, v.32, p.365-375.

SEIFERT, W.K., MOLDOWAN, J.M., 1986. “*Use of biological markers in petroleum exploration*.” In: Jonhs, R.B. (eds), *Methods in Geochemistry and Geophysics*, Elsevier, Amsterdam, v.24, p. 261-290.

- SILVA, D.B., 2010. “*Integração de parâmetros geoquímicos para caracterização de óleos leves através de biomarcadores, diamantóides, hidrocarbonetos leves e dibenzotiofenos*”. Dissertação de MSc., IQ/UFRJ, Rio de Janeiro, RJ, Brasil.
- SPRINGER , M.V., 2007, “*Caracterização de Biomarcadores e Diamantóides em Amostras de Óleos da Bacia Llanos Orientales, Colômbia*”. Dissertação de MSc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- SPRINGER, M.V., GARCIA, D.F., GONÇALVES, F.T.T., LANDAU, L., AZEVEDO, D.A., 2010. ”*Diamondoid and Biomarker Characterization in Oils from the Llanos Orientales Basin, Colombia*”. *Organic Geochemistry*, v.41, p.1013-1018.
- TAMANQUEIRA, J. B., 2006, “*Caracterização de Diamantóides em Amostras de Óleo de Bacias Sedimentares Brasileiras*”. Dissertação de MSc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- TEN HAVEN, H.L., DE LEEUW, J.W., RULLKOTTER, J., SINNINGHE DAMSTÉ, J.S., 1987. “*Restricted utility of the pristane/phytane ratio as a palaeoenvironmental indicator*”. *Nature*, v.330, p. 641-643.
- TISSOT, B.P., WELTE, D.H., 1984. “*Petroleum Formation and Occurrence*”. 2<sup>a</sup> ed., Berlin, Springer-Verlag.
- WAPLES, D.W., MACHIHARA, T., 1991. “*Biomarkers for geologists: a practical guide to the application of steranes and triterpanes in petroleum geology*”. *AAPG methods in exploration series*, v. 9, p.1-76.
- WINGERT, W.S., 1992. “*GC-.M.S Analysis of Diamondoid Hydrocarbons in Smackover Petroleum*”. *Fuel*, v.71, p. 37-43.
- XIE, X., BENI, G., 1991.“*Validity Measure for Fuzzy Clustering*”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 13, p. 841–847.

ZUMBERGE, J.E., 1983. "*Tricyclic diterpane distributions in the correlation of Paleozoic crude oils from the Williston basin*". In: M.Bjoroy *et al.* (eds), *Advances in Organic Geochemistry* 1981, Wiley, p. 738-745.

ZUMBERGE, J.E., 1987a. "*Prediction of source rock characteristics based on terpane biomarkers in crude oils: a multivariate statistical approach*". *Geochimica et Cosmochimica Acta*, v.51, p. 1625-1637.

ZUMBERGE, J.E., 1987b. "*Terpenoid biomarker distributions in low maturity crude oils*". *Organic Geochemistry*, v.11, p. 479-496.