



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GEOFÍSICA

TESE DE DOUTORADO

**IDENTIFICAÇÃO LITOLÓGICA COM A HIBRIDIZAÇÃO DOS ALGORITMOS
FIREFLY E AFFINITY PROPAGATION**

NAYARA SAFIRA DA SILVA CALDAS

Belém - Pará

2017

NAYARA SAFIRA DA SILVA CALDAS

**IDENTIFICAÇÃO LITOLÓGICA COM A HIBRIDIZAÇÃO DOS ALGORITMOS
FIREFLY E AFFINITY PROPAGATION**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Geofísica do Instituto de Geociências da Universidade Federal do Pará, em cumprimento às exigências para obtenção do título de doutora em Geofísica.

Área de concentração: Geofísica de Poço

Orientador: André José Neves Andrade

Coorientador: Carlos Eduardo Guerra

Belém - Pará

2017

Dados Internacionais de Catalogação de Publicação (CIP)
Biblioteca do Instituto de Geociências/SIBI/UFPA

Caldas, Nayara Safira da Silva, 1988 -
Identificação litológica com a hibridização dos
algoritmos firefly e affinity propagation / Nayara Safira da
Silva Caldas. – 2017.
54 f. : il. ; 30 cm

Inclui bibliografias

Orientador: André José Neves Andrade

Coorientador: Carlos Eduardo Guerra

Tese (Doutorado) – Universidade Federal do Pará, Instituto
de Geociências, Programa de Pós-Graduação em Geofísica,
Belém, 2017.

1. Perfilagem geofísica de poços. 2. Programação
estruturada. 3. Otimização combinatória. I. Título.

CDD 22. ed.: 622.15

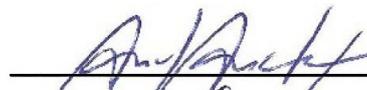
NAYARA SAFIRA DA SILVA CALDAS

**IDENTIFICAÇÃO LITOLÓGICA COM A HIBRIDIZAÇÃO DOS ALGORITMOS
FIREFLY E AFFINITY PROPAGATION**

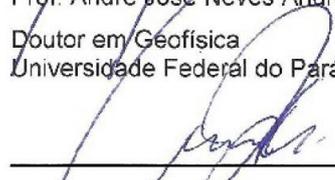
Tese de Doutorado apresentada ao Programa de Pós-Graduação em Geofísica do Instituto de Geociências da Universidade Federal do Pará para obtenção do título de doutora em Geofísica.

Data de Aprovação: 26 de maio de 2017

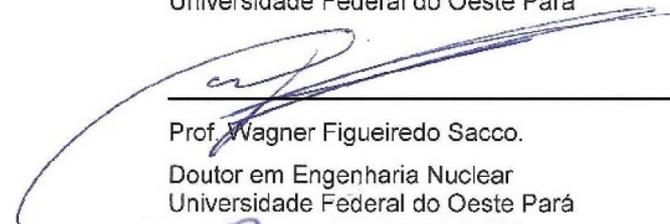
Banca Examinadora:



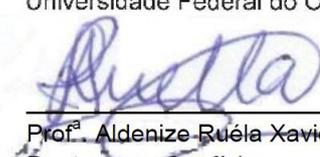
Prof. André José Neves Andrade - Orientador
Doutor em Geofísica
Universidade Federal do Pará



Prof. Carlos Eduardo Guerra
Doutor em Geofísica
Universidade Federal do Oeste Pará



Prof. Wagner Figueiredo Sacco.
Doutor em Engenharia Nuclear
Universidade Federal do Oeste Pará



Prof.ª Aldenize Ruéla Xavier
Doutora em Geofísica
Universidade Federal do Oeste do Pará



Prof.ª Ana Carolina Rios Coelho
Doutora em Engenharia Mecânica
Universidade Federal do Oeste do Pará

AGRADECIMENTOS

Ao Prof.Dr. Carlos Eduardo Guerra pelo apoio e pela orientação na realização deste trabalho.

Ao Prof. Dr. André Andrade pela orientação.

À PETROBRÁS (Rede de Geofísica), pelas oportunidades e suporte ofertado ao curso e a minha pesquisa.

À UFOPA e ao colegiado de Geofísica pela compreensão e flexibilização de horários o que possibilitou o desenvolvimento desta pesquisa.

RESUMO

Este trabalho apresenta uma metodologia para a solução de problemas de extração de informações geológicas tal como a identificação de litologias em profundidade, diretamente dos perfis de poços. Para isto, faz-se o uso do gráfico M-N como técnica para identificação de litologias em formações atravessadas por poços. No entanto, a interpretação visual deste gráfico torna-se limitada quando há um grande espalhamento de pontos no gráfico. A metodologia propõe um algoritmo híbrido para interpretação computacional dos dados do gráfico M-N, por meio de um método de clusterização conhecido como Affinity Propagation. Para otimização da operação de clusterização, adota-se um vetor de preferências com a informação dos pontos M-N que representam os possíveis melhores candidatos para representantes das classes litológicas. Estes possíveis melhores candidatos são obtidos através da otimização de uma função Kernel, construída com base no espaço M-N, com o uso do algoritmo meta-heurístico Firefly. Por fim, a classificação é feita pela associação dos exemplares dos clusters obtidos com os pontos fixos do gráfico M-N pelo critério de minimização de distância entre estes pontos. A metodologia foi aplicada em dados M-N sintético e real, comprovando a sua eficácia, tanto na redução do número de agrupamentos obtidos pelo algoritmo de clusterização, bem como na identificação de classes litológicas em dados M-N com alto grau de espalhamento, comprovado na avaliação do índice kappa onde verificou-se uma boa concordância entre a classificação final com o dado testemunhado.

Palavras-chave: Avaliação de Formação. Perfil de poço. Testemunho. Análise de Agrupamentos. Otimização.

ABSTRACT

This work presents a methodology for solving the problem of extracting geological information, such as the identification of lithologies at depth, directly from wireline logs. For this purpose, the M - N cross-plot is used as the formation evaluation technique for identifying the lithologies in the logged borehole, in terms of the physical properties of the main mineral in each lithology. However, the visual interpretation of this graphic is limited by the large spread of points in the graph. The proposed methodology is a hybrid algorithm for a computational interpretation of the M - N data by means of the affinity propagation clustering technic. For the searching optimization of clusters, it is used a preference vector that flags affinity propagation with M - N points that are the best candidates to be the lithological classes representatives. The possible best candidates to represent the lithological classes are obtained by a firefly metaheuristic optimization of a Kernel function, built over M - N space. Finally, a minimum distance criterion, with respect to M - N fixed points, is applied to the set of exemplary points, acquired from Affinity Propagation, in order to associate the lithologies. The methodology was applied in synthetic and real M - N data, proving its effectiveness, both in reducing the number of groupings obtained by the clustering algorithm as well as in the identification of highly-spread lithological M - N data classes, verified in the evaluation of the kappa index where there was a good agreement between the final classification with the core data.

Key words: Formation evaluation. Log data. Core data. Cluster Analysis Non-Linear optimization.

LISTA DE ILUSTRAÇÕES

2.1 Gráfico M-N. Os quadrados em preto representam os minerais mais comuns com seus respectivos valores de M-N.	9
2.2 Ilustração de um poço sintético sobre uma litologia hipotética (verde-arenito, azul- Calcário, cinza-folhelho).	10
2.3 Gráfico M-N dos perfis sintéticos da Figura 02 e os minerais principais (pontos fixos) expostos na Tabela 1.	11
3.1 Ilustração da troca de mensagens entre os dados segundo AP.	14
6.1 Algoritmo híbrido para mapeamento litológico.	21
7.1 Perfis de dados sintéticos.	23
7.2 Gráfico <i>M-N</i>	24
7.3 (a) Dados <i>M-N</i> (preto), função Pdf (superfície em vermelho). (b) Máximos obtidos com FA (pontos pretos).	24
7.4 (a) Agrupamentos resultantes do processamento AP com vetor de preferência usual. (b) Clusters resultantes do processamento AP com vetor de preferência modificado.	25
7.5 (a) Classificação final por associação de exemplares. (b) Mapeamento de Litologias em profundidade.	26
7.6 Perfis de dado real.	27
7.7 Gráfico M-N com três agrupamentos sugeridos pela interpretação visual.	28
7.8 (a) Resultados do agrupamento com AP com o vetor de preferência usual.(b) Resultados do agrupamento com AP com as modificações do vetor de preferência.	29
7.9 (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade.	29
7.10 Testemunho (à esquerda), perfil de Raio Gama e perfis de porosidade. Intervalos selecionados destacados em cinza e verde.	30
7.11 Gráfico M-N mostrando um espalhamento de pontos e deslocamento na direção dos argilominerías.	31

7.12 (a) Resultado de dez clusters resultantes do AP. (b) Resultados com AP com modificação do vetor P.	32
7.13 (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade.	33
7.14 Testemunho (à esquerda), perfil de Raio Gama e perfis de porosidade. Intervalos selecionados destacados em vermelho.	34
7.15 Gráfico <i>M-N</i> obtidos com os dados do perfil.	34
7.16 (a) Resultado de dez clusters resultantes do AP. (b) Resultados com AP com modificação do vetor P.	35
7.17 (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade.	36
7.18 Testemunho (à esquerda), perfil de Raio Gama e perfis de porosidade referentes ao intervalo de profundidade selecionado.	37
7.19 Gráfico <i>M-N</i> obtidos com os dados do perfil.	38
7.20 (a) Resultado de sete cluster resultantes do AP. (b) Resultados com AP com modificação do vetor P.	39
7.21 (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade. Fonte: Da autora.....	39
7.22 Testemunho (à esquerda), perfil de Raio Gama e perfis de porosidade referentes ao intervalo de profundidade selecionado.	40
7.23 Gráfico <i>M-N</i> obtidos com os dados do perfil.	41
7.24 (a) Resultado resultantes do AP. (b) Resultados com AP com modificação do vetor P.	42
7.25 (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade.	42
7.26 Testemunho (à esquerda), perfil de Raio Gama e perfis de porosidade referentes ao intervalo de profundidade selecionado.	43
7.27 Gráfico <i>M-N</i> obtidos com os dados do perfil.	44

7.28 (a) Resultado resultantes do AP. (b) Resultados com AP com modificação do vetor P.	45
7.29 (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade.	45

LISTA DE TABELA

2.1	Valores de M e N para alguns minerais.	10
7.1	Descrição Geológica das camadas do modelo sintético utilizado.....	22
7.2	Descrição Geologica do dado real 1 utilizado.....	27
7.3	Descrição de testemunhos do poço 2.	30
7.4	Descrição geológica do poço 3.	33
7.5	Descrição geológica do poço 4	37
7.6	Descrição geológica do poço 5.	40
7.7	Descrição geológica do poço 6	43
7.8	Matriz de erro.....	47
7.9	Avaliação do grau de concordância usando K.....	48
7.10	Índice kappa para avaliação das classificações.....	48

SUMÁRIO

1 INTRODUÇÃO	1
2 AVALIAÇÃO DE FORMAÇÃO	3
2.1 PERFIL DE RAIOS GAMA NATURAL	3
2.2 PERFIS DE POROSIDADE.....	4
2.2.1 Perfil de Densidade	4
2.2.2 Perfil de Porosidade Neutrônica (ϕ_N).....	5
2.2.3 Perfil Sônico (Δt).....	6
2.3 GRÁFICO M-N	6
2.3.1 Interpretação do Gráfico M-N	7
3 MÉTODOS DE AGRUPAMENTO	12
3.1 AFFINITY PROPAGATION (AP)	12
4 FUNÇÃO DE DENSIDADE KERNEL	15
5 ALGORITMO FIREFLY (FA)	17
6 METODOLOGIA	19
7 RESULTADOS	22
7.1 DADOS SINTÉTICOS.....	22
7.2 DADOS REAIS	26
7.2.1 Dado Real 1	26
7.2.2 Dado Real 2	30
7.2.3 Dado Real 3	33
7.2.4 Dado Real 4	36
7.2.5 Dado Real 5	40
7.2.6 Dado Real 6	43
7.3 AVALIAÇÃO DA CLASSIFICAÇÃO LITOLÓGICA	46
8 CONCLUSÕES	49
REFERÊNCIAS	52

1 INTRODUÇÃO

A perfilagem de poço é o conjunto de operações e técnicas que envolvem a aquisição, processamento e interpretação das medições de algumas propriedades físicas das rochas atravessadas por um poço. A análise desses registros pode ser realizada de forma qualitativa ou quantitativa. A análise qualitativa é realizada, basicamente, pela interpretação visual das variações das propriedades físicas em profundidade, permitindo eventualmente a distinção das diferentes camadas atravessadas pelo poço. A análise quantitativa, também conhecida como avaliação de formação, é realizada através de cálculos numéricos, para obtenção das propriedades petrofísicas (argilosidade, porosidade e saturação de fluido) das camadas, e o cálculo das reservas de hidrocarbonetos. A avaliação de formação depende fundamentalmente da informação petrofísica e geológica produzida pela análise de testemunhos. A informação geológica obtida em um perfil de poço permite a correlação e atualização estratigráfica do campo de petróleo e tem duas aplicações básicas para a avaliação de formação: a primeira trata do cálculo de porosidade, o que requer informações sobre os parâmetros de porosidade da matriz rochosa (Barros, 2008). A segunda aplicação refere-se ao zoneamento do poço, ou a identificação das profundidades de topo e base de diferentes camadas (Ellis, 2007).

O modo ideal para a identificação e delimitação de camada é a análise geológica de testemunhos retirados ao longo de toda a profundidade do poço. Alguns fatores impedem este procedimento tais como a inviabilidade técnica de realização da aquisição em poços horizontais, o alto custo das operações de testemunhagem e o fator de recuperação, que é a impossibilidade de recuperação de 100% da rocha retida no amostrador em uma operação de testemunhagem.

Ao longo da história da avaliação de formação algumas técnicas para a identificação litológica foram desenvolvidas (Burke, 1969; Busch et al., 1987; Clavier et al., 1976; Savre, 1963). Desta forma, diferentes metodologias têm sido aplicadas para o mapeamento de litologias, através da utilização dos perfis de densidade, porosidade neutrônica e sônico.

Uma das técnicas utilizadas na interpretação da litologia é o Cross-Plot dos perfis de porosidade. No caso de litologias simples, a simples interpretação visual é suficiente para caracterização litológica, tais como perfil sônico versus perfil de densidade para identificação de arenito argiloso e o cross-plot do perfil de densidade e neutrônico para identificação de rochas carbonáticas. Em litologias complexas, o Gráfico M-N, Identificação de Matriz (MID Plot), dentre outros, podem ser utilizados. O Gráfico M-N (Burke, 1969) é

um método amplamente utilizado na avaliação da formação para a identificação de litologias a partir das propriedades físicas do mineral principal. Este método combina os três perfis clássicos de porosidade (Asquith, 1982) e um modelo de rocha (sem argilosidade) para gerar um gráfico que, por construção, atenua o efeito da porosidade sobre os perfis de porosidade. Este tipo de gráfico, em princípio, proporciona uma interpretação visual rápida da litologia de qualquer ponto ao longo da profundidade do poço. No entanto, a presença de ruído nos perfis de porosidade e a simplificação do modelo de rocha, que desconsidera a argilosidade, tornam a interpretação visual do Gráfico M-N complexa e ambígua.

Neste estudo, apresenta-se uma metodologia híbrida para a interpretação computacional do Gráfico M-N com base na técnica de agrupamento Affinity Propagation. Para melhorar a busca de clusters, é definido um vetor de preferência (parâmetro de entrada do Affinity Propagation) com base nos possíveis melhores candidatos a representantes de litologias. Estes possíveis melhores candidatos são obtidos a partir da otimização de uma função kernel, construído sobre o espaço M-N. Finalmente, um critério de distância mínima relaciona os pontos fixos M-N aos pontos exemplares adquiridos pelo Affinity Propagation, para a associação de litologias.

Este trabalho foi dividido em oito seções. Nas cinco primeiras são apresentados o Gráfico M-N, o algoritmo Affinity Propagation, a função densidade de Kernel e o algoritmo meta-heurístico Firefly. Em seguida, seção seis, a metodologia híbrida é introduzida. Nas últimas duas seções apresenta-se e avalia-se esta metodologia com a utilização de dados sintéticos e perfis reais, registrados em poços do Campo de Namorado na Bacia de Campos.

2 AVALIAÇÃO DE FORMAÇÃO

A geofísica de poço é o conjunto de operações e técnicas que envolvem a aquisição, o processamento e a interpretação das medidas de algumas propriedades físicas das rochas atravessadas por um poço. Tais medidas são adquiridas na operação de perfilagem de poço. A perfilagem (well logging), do termo francês *carottage électrique*, surgiu em 1927, como uma técnica de prospecção geofísica (Ellis, 2007). Trata-se do registro de algumas propriedades físicas das rochas atravessadas pelo poço por um dispositivo de medida (ferramenta e sonda de perfilagem), que se desloca de forma contínua, na direção ascendente do poço.

Os perfis de poços são registros pontuais das variações de uma das propriedades medidas pelas ferramentas em função da profundidade. A análise desses perfis pode ser realizada de forma qualitativa ou de forma quantitativa. A primeira é feita basicamente com evidências visuais das variações das propriedades físicas com a profundidade, podendo eventualmente, distinguir as diferentes camadas atravessadas pelo poço. Enquanto que a análise quantitativa, também conhecida como avaliação de formação, é realizada através de cálculos numéricos, nos quais podem-se determinar as propriedades petrofísicas (argilidade, porosidade e a saturação fluida) das diversas camadas, e com isto realizar o cálculo das reservas ou do volume de hidrocarboneto. De um modo geral, a avaliação de formação pode ser conceituada como um conjunto de métodos que visam extrair informações dos perfis geofísicos, com o objetivo de identificar, quantificar e produzir um modelo geológico para uma acumulação de hidrocarbonetos em subsuperfície. Dentre os perfis, os mais utilizados para a identificação de litologias são os perfis de porosidade e o perfil de Raio Gama natural para diferenciar rochas reservatório e rochas selantes.

2.1 PERFIL DE RAIOS GAMA NATURAL

É o registro da variação da radiação gama natural ao longo da profundidade de um poço. A radioatividade natural é devido ao decaimento radioativo de isótopos instáveis com altos valores de meia-vida, tais como Potássio (K^{40}), o Tório (T_h^{232}) e o Urânio (U^{238}) (Ellis, 2007).

O perfil de raios gama natural é utilizado para classificar as rochas perfuradas em rochas reservatório ou rochas selantes devido às diferentes concentrações de isótopos radioativos em cada uma delas. As rochas selantes, como os folhelhos, normalmente emitem

uma maior quantidade de radiação gama que as rochas reservatório, como o arenito, por exemplo, em razão da presença de feldspatos (potássico) e de matéria orgânica.

O perfil de raios gama natural é utilizado como um indicador da argilosidade ou volume relativo de argila nas rochas sedimentares, na forma,

$$V_{sh} = \frac{RG - RG_{min}}{RG_{max} - RG_{min}} \quad (2.1)$$

Na equação 2.1, RG representa os valores registrados no perfil de raios gama natural. RG_{max} e RG_{min} representam, respectivamente, os valores máximos e mínimos da radiação gama, que foram interpretados no intervalo perfilado.

2.2 PERFIS DE POROSIDADE

É uma designação geral para o conjunto formado pelos perfis de densidade, porosidade neutrônica e sônico, que realizam o registro de propriedades físicas das rochas (densidade, porosidade neutrônica e tempo de trânsito) que, segundo o modelo de rocha, podem ser expressos por uma combinação linear da propriedade física mensuradas individualmente em cada material presente na constituição da rocha e os seus volumes relativos (Ellis, 2007).

Considerando um modelo de uma rocha reservatório constituída por matriz (mineral), fluido (água doce) e argila, pode-se escrever a equação geral para um perfil de porosidade qualquer, na forma

$$p = \phi p_w + V_{sh} p_{sh} + (1 - \phi - V_{sh}) p_m. \quad (2.2)$$

Na equação 2.2, p é um perfil de porosidade. p_w é o valor da propriedade medida na água doce, ϕ é a porosidade, p_{sh} é o valor da propriedade medida na argila, V_{sh} é a argilosidade e p_m é o valor da propriedade medida no mineral principal (Barros, 2008).

2.2.1 Perfil de Densidade

O perfil de densidade (ρ) é o registro contínuo das variações das densidades das rochas atravessadas por um poço. No caso de rochas porosas, a medição realizada pelo perfil inclui tanto a densidade da matriz da rocha, como a do fluido contido no espaço poroso.

Sabe-se que a atenuação de um feixe de raios gama que atravessa a matéria, pode ser relacionada com a densidade dos elétrons do material, se a interação predominante entre os raios gama e os átomos do material for o efeito Compton (Ellis, 2007). No ambiente do poço, uma medida da atenuação dos raios gama pode ser usada para determinar a densidade da rocha próxima à parede do poço. A fonte de raios gama normalmente utilizada é o Césio (Cs^{137}), o qual emite raios gama com uma energia de 662 KeV, bem abaixo do limite para a ocorrência do efeito de produção de par. Este isótopo tem uma meia vida de cerca de trinta anos, o que dá uma estabilidade na intensidade do feixe energético, durante um período de tempo considerável. O poço é bombardeado por um fluxo de raios gama produzido pela fonte radioativa, que se encontra no interior da ferramenta. Um cintilômetro, montado no corpo ferramental e protegido da radiação direta da fonte, realiza o registro do fluxo de raios gama difundido pela formação. O sinal observado é relacionado com a densidade da formação, uma vez que ele é representativo do número de raios gama que foram desviados de sua trajetória original no interior da formação. Assim, quanto menor for o fluxo de raios gama detectado maior é o número de elétrons (densidade eletrônica) responsáveis pela mudança na trajetória dos raios gama pelo efeito Compton, o que implica em um maior número de átomos por unidade de volume.

Como a ferramenta de densidade não responde diretamente à densidade (massa por unidade de volume), mas sim ao número de elétrons por unidade de volume, ou seja, a densidade eletrônica, a calibração é realizada internamente considerando o calcário ($\rho_{cal} = 2,71 \text{ g/cm}^3$) saturado com água doce como rocha de referência para produzir o registro do valor real da densidade da rocha.

2.2.2 Perfil de Porosidade Neutrônica (ϕ_N)

A ferramenta de porosidade neutrônica, que registra as medidas do perfil de porosidade neutrônica (ϕ_N), foi o primeiro dispositivo nuclear descido em um poço. Baseado no fato que os nêutrons são partículas sem carga elétrica e com massa similar à do núcleo de hidrogênio, elas interagem com as rochas atingindo os núcleos dos elementos dos diversos materiais que as constituem, na forma elástica ou inelástica. Como o hidrogênio é o principal atenuador da energia de um fluxo de nêutrons e se faz presente na constituição da rocha na forma de água ou hidrocarboneto, a medida da atenuação de um fluxo de nêutrons

está diretamente relacionada ao volume de hidrogênio ou à porosidade (Ellis, 2007; Asquith, 1982).

A ferramenta neutrônica é basicamente construída por uma fonte e dois detectores de nêutrons. A fonte emite nêutrons rápidos, os quais penetram nas camadas próximas da parede do poço. Através das múltiplas colisões elásticas, os nêutrons perdem parte da energia com que foram injetados; esta perda de energia está sujeita ao tamanho da massa do núcleo do átomo alvo. A atenuação de energia é máxima quando o núcleo alvo tem massa igual a do nêutron, caso do núcleo de hidrogênio.

O problema da ferramenta de porosidade neutrônica é que esta é sensível a todas as fontes de hidrogênio numa formação, não apenas à contida nos espaços porosos, mas também presente na argila, uma vez que o hidrogênio associado aos minerais de argila é visto pela ferramenta da mesma maneira como o hidrogênio nos fluidos.

2.2.3 Perfil Sônico (Δt)

No início dos anos 50, os experimentos sobre amostra de rocha de Wyllie e outros, na Gulf Oil Co, constataram que a velocidade de propagação de uma onda acústica podia ser representada por uma combinação linear da velocidade de propagação da onda e o volume relativo de cada constituinte da rocha (Crain, 1986). A ferramenta sônica é composta por um ou mais conjuntos formados por um transmissor e dois receptores acústicos. Um receptor está mais próximo e outro mais afastado em relação ao transmissor de modo a atenuar o efeito da lama sobre o tempo de propagação das ondas compressionais (onda P) na rocha. O perfil sônico avalia a velocidade de propagação de uma onda compressional realizando o registro das variações do tempo de trânsito em função da profundidade. O tempo de trânsito (Δt) é conceituado como o tempo de propagação de uma onda refratada na parede do poço na distancia de um pé (30,48 cm), que por construção é a separação entre os dois receptores.

2.3 GRÁFICO M-N

O gráfico M-N (Burke, 1969) resulta de uma particular combinação dos três perfis de porosidade (sônico, densidade e porosidade neutrônica) para produzir a identificação litológica de um trecho de poço perfilado, em termos das propriedades físicas do mineral

principal de cada litologia. O Gráfico M-N é construído com pares ordenados, definidos pelos parâmetros M e N, que são formulados de modo a torná-los relativamente independentes do efeito da porosidade sobre os registros dos perfis de porosidade. O modelo de rocha considerado para a definição dos parâmetros M e N desconsidera a presença da argila ($V_{sh}=0$) na constituição da rocha reservatório. Burke (Burke, 1969) define os parâmetros N e M por:

$$N = \frac{\phi_{Nmf} - \phi_{Nma}}{\rho_{ma} - \rho_{mf}}. \quad (2.3)$$

$$M = \frac{\Delta t_{mf} - \Delta t_{ma}}{\rho_{ma} - \rho_{mf}} \times 0,01. \quad (2.4)$$

Na Eq.(2.4) é introduzido o fator de escala 0,01 para que os parâmetros tenham a mesma magnitude.

2.3.1 Interpretação do Gráfico M-N

O Gráfico M-N é construído por pontos formados por pares ordenados (N,M) calculados para os principais minerais constituintes das rochas sedimentares. Cada mineral produz um único ponto (Tabela 2.1), denominados pontos fixos do Gráfico M-N (**Figura 2.1**). Admite-se que os pontos fixos apresentem valores dos parâmetros M e N suficientemente diferentes, tal que cada ponto fixo ocupe uma posição relativamente distante dos outro, de modo a permitir a classificação das litologias em função das propriedades físicas do seu mineral principal (Barros, 2007).

Os pares (N, M) para os pontos em profundidade de um poço perfilado são calculados a partir da substituição dos parâmetros de matriz ($\rho_{ma}, \phi_{ma}, \Delta t_{ma}$) pelos dados dos perfis nas equações (5) e (6).

$$N = \frac{\phi_{Nmf} - \phi_{Nb}}{\rho_b - \rho_{mf}}. \quad (2.5)$$

$$M = \frac{\Delta t_{mf} - \Delta t_b}{\rho_b - \rho_{mf}} \times 0,01. \quad (2.6)$$

onde $\rho_{mf}, \phi_{Nmf}, \Delta t_{mf}$ e $\rho_b, \phi_{Nb}, \Delta t_b$ são os parâmetros do filtrado de lama e os dados registrados nos perfis, respectivamente. Na **Figura 2.2**, apresenta-se um exemplo de formação com o modelo litológico e os referentes perfis sintéticos onde a profundidade é dada em metro.

Após o cálculo dos parâmetros M-N para cada profundidade dos perfis sintéticos (**Figura 2.2**), estes pontos se distribuem no plano N-M, aproximadamente, nas vizinhanças dos minerais principais de cada litologia (**Figura 2.3**). A identificação da litologia é produzida pela proximidade do ponto do perfil ao ponto fixo do mineral principal da sua litologia. Portanto, os pontos de folhelho, calcário e arenito distribuem-se nas proximidades dos argilominerais, calcita e quartzo, respectivamente.

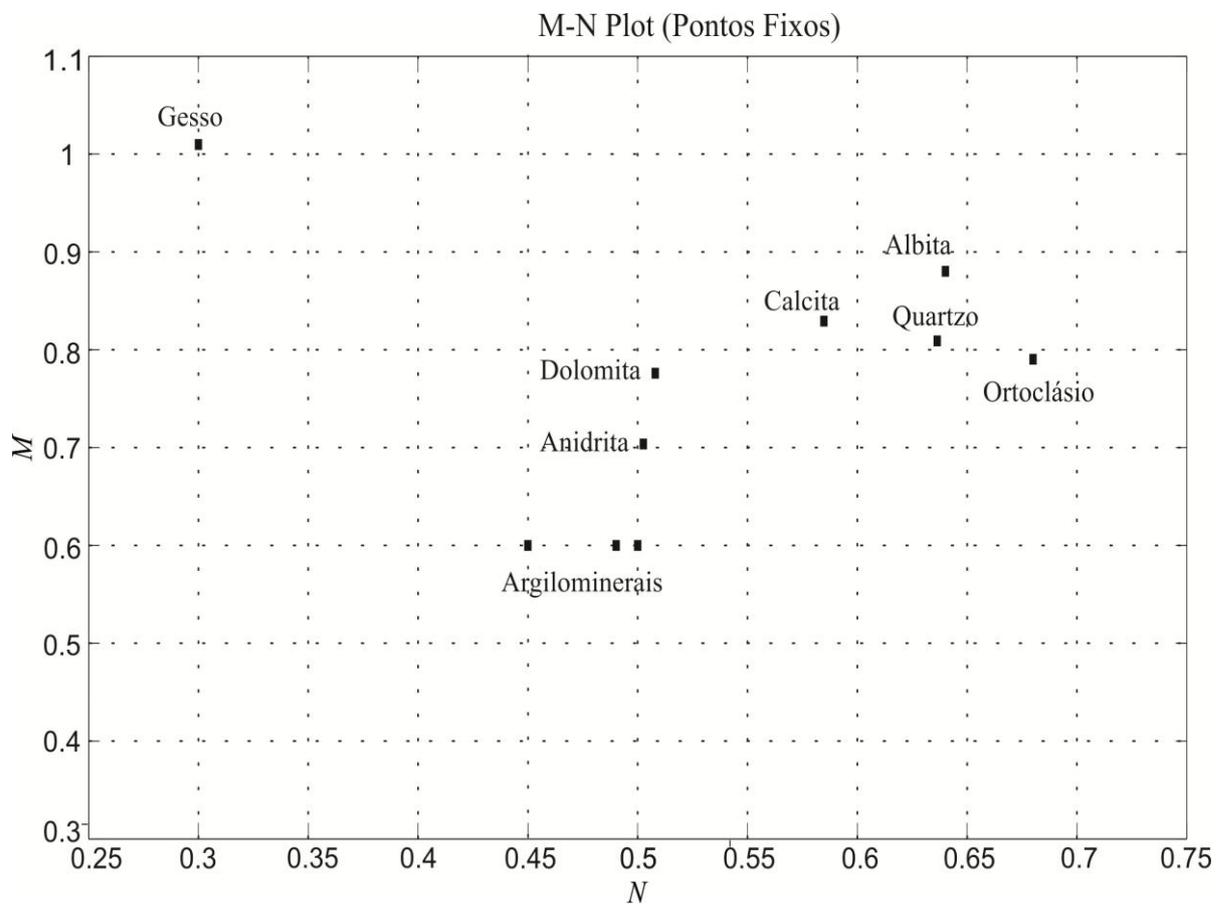


Figura 2.1. Gráfico M-N. Os quadrados em preto representam os minerais mais comuns com seus respectivos valores de M-N. Fonte: Da autora.

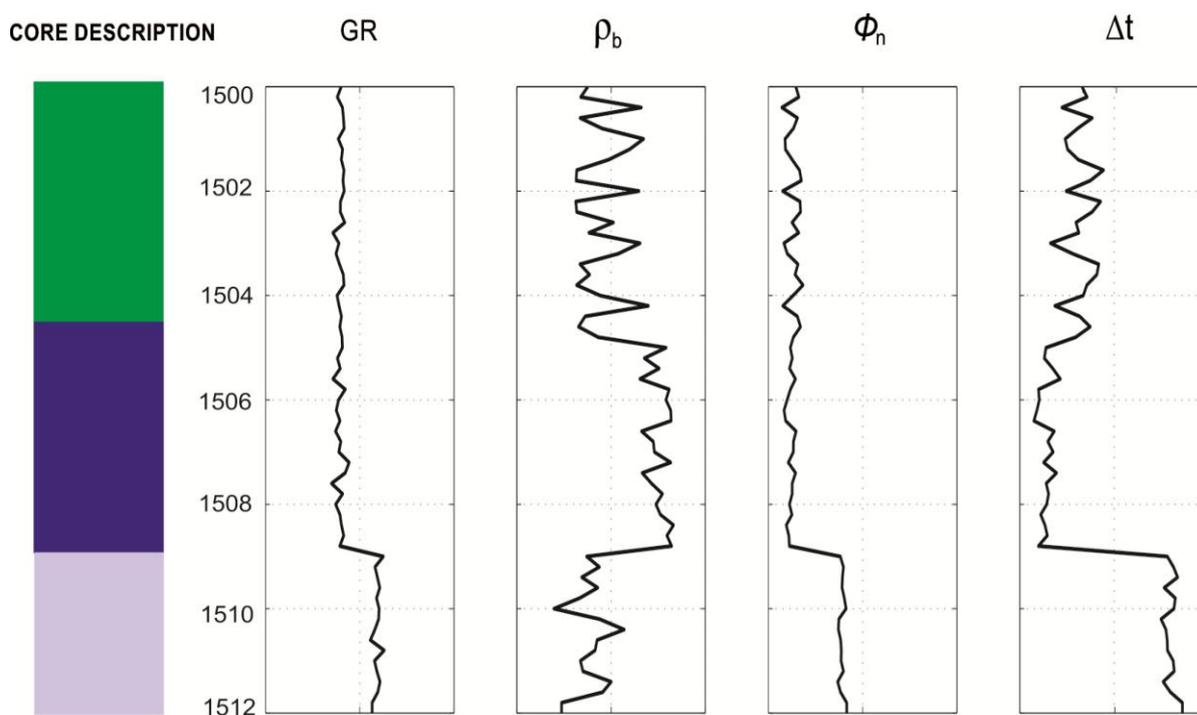


Figura 2.2. Ilustração de um poço sintético sobre uma litologia hipotética (verde-arenito, azul- Calcário, cinza-folhelho).Fonte: Da autora

Tabela 2.1 - Valores de M e N para alguns minerais.

Mineral	ρ	Φ_N	Δt	M	N	Composição
Quartzo	2,65	-0,05	55,5	0,81	0,64	SiO_2
Calcita	2,71	0,00	47,0	0,83	0,59	$CaCO_3$
Dolomita	2,86	0,05	43,6	0,78	0,508	$CaMg(CO_3)_2$
Ortoclásio	2,55	-0,05	66,5	0,79	0,68	$KAlSi_3O_8$
Albita	2,62	-0,04	46,4	0,88	0,64	$NaAlSi_3O_8$
Anidrita	2,96	0,02	51,8	0,70	0,50	$CaSO_4$
Gesso	2,32	0,604	55,7	1,01	0,30	$CaSO_4 \cdot 2H_2O$
Argilomineral						
Caulinita	2,42	0,36	103,8	0,60	0,45	-
Ilita	2,53	0,25	97,2	0,60	0,49	-
Esmectita	2,12	0,44	121,8	0,60	0,50	-

Fonte: Adaptado de Burke, Campbell JR e Smith. (1969).

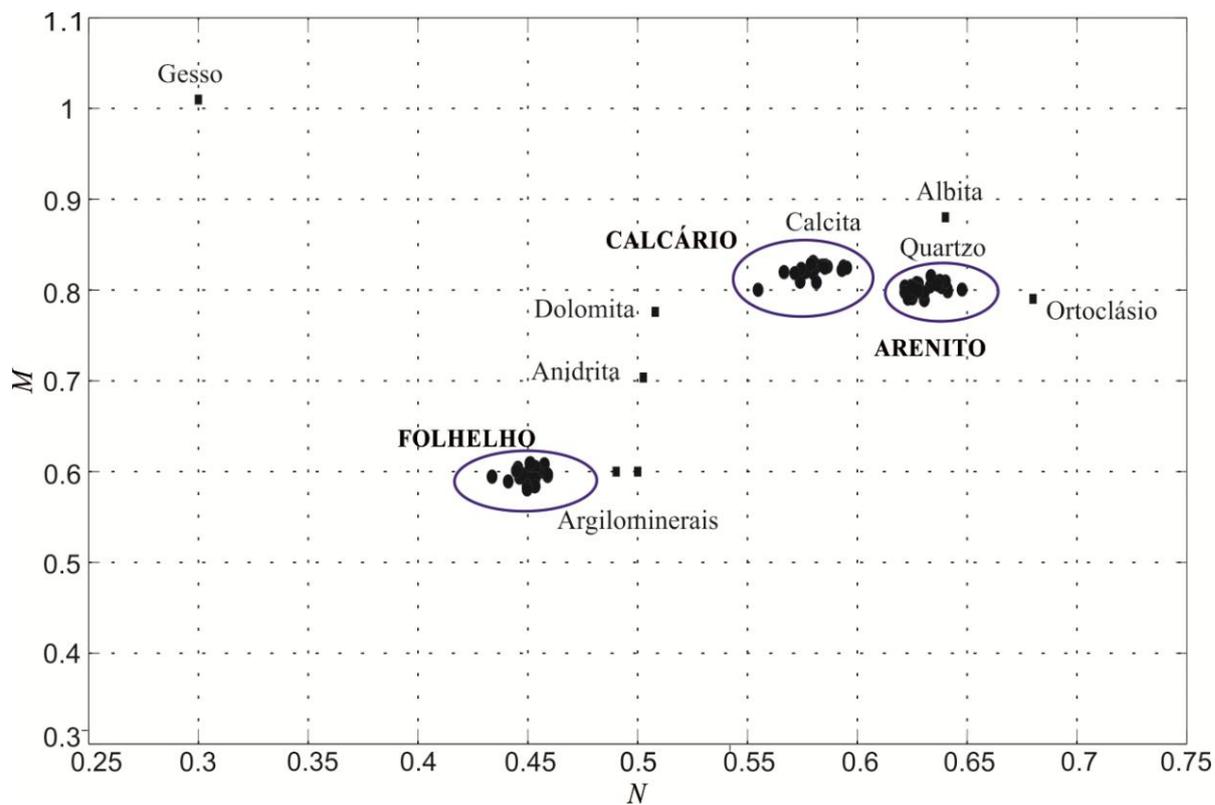


Figura 2.3. Gráfico M-N dos perfis sintéticos da Figura 02 e os minerais principais (pontos fixos) expostos na Tabela 1. Fonte: Da autora

3 MÉTODOS DE AGRUPAMENTO

A análise de agrupamentos é uma ferramenta útil para análise de dados em diferentes situações. Essa técnica pode ser usada para reduzir a dimensão de um conjunto de dados, reduzindo uma ampla gama de objetos à informação do centro do seu conjunto. Técnicas de agrupamentos consistem em reunir objetos com características comuns em grupos (Clusters) ou classes. Um cluster é um conjunto de objetos com características em comum e com características que os diferem de outros clusters. Esta técnica de agrupamento de dados pode ser implementada em sistemas de aprendizado sendo útil para extrair características escondidas dos dados e desenvolver hipóteses a respeito de sua natureza (Linden, 2009).

Existem diversos algoritmos que visam dividir um espaço amostral de dados em grupos cujos elementos possuem algumas características similares. A maioria dos algoritmos usa o critério de dissimilaridade pré-estabelecido, normalmente, a métrica utilizada é a distância euclidiana, mas podem ser usadas outras métricas, de preferência as que retratem melhor alguma característica comum dos dados com os quais se esteja trabalhando.

Aplicações que envolvem análise de agrupamentos são muito utilizadas em distintos campos da ciência e dos negócios como: Bioinformática, análise de dados espaciais, mineração de dados, geração e prova de hipóteses, predição baseada em agrupamento, para posterior classificação (Villanueva, 2008).

3.1 AFFINITY PROPAGATION (AP)

Os principais algoritmos de agrupamento utilizam como parâmetro de entrada um número k pré-determinado de clusters para particionar o espaço amostral de dados. O Affinity Propagation parte do princípio de que todos os pontos ou dados podem ser eleitos com um "exemplar", ou seja, um representante de um dado agrupamento. O conjunto de dados forma a configuração de uma rede onde cada ponto representa um nó e ocorre a transmissão de mensagens entre as arestas da rede que visam otimizar a convergência para a eleição de um conjunto de exemplares que serão considerados representantes do grupo ao qual estão vinculado.

Affinity Propagation tem por entradas uma coleção de valores de similaridades entre dois pontos (dados), onde a similaridade $s(i, k)$ indica o quão adequado o ponto com índice

k seria para ser um exemplar para o ponto de índice i . Quando o objetivo é minimizar o quadrado do erro, cada similaridade é definida como o negativo do quadrado do erro (distância euclidiana): Para os pontos x_i e x_k , $s(i, k) = -\|x_i - x_k\|^2$. Esse método pode ser aplicado em casos mais gerais, onde a similaridade é derivada de pares de imagens, microarrays, sentenças em inglês, e pares de cidades (Frey, 2007).

A matriz de similaridades depende do número de dados em questão, isto é, para n dados, a matriz de similaridades tem ordem n^2 . Aplicando ao nosso problema, n é o número de pontos (N, M). A similaridade esta relacionada a alguma métrica d sobre os dados, conceito esse que incorpora a noção geométrica geral de distância.

Além das similaridades, Affinity Propagation também tem como parâmetro de entrada valores $s(k, k)$ denominados "preferências" para cada ponto k . Sabe-se que, em qualquer métrica, a diagonal principal da matriz de similaridades é nula, o que indica que a maior similaridade ocorre de um ponto a ele mesmo, ao substituírmos este valor por $s(k, k) \neq 0$ indica-se preferências para esses pontos serem exemplares. Esses valores de preferência interferem nos resultados (número de clusters) encontrados pelo AP; altos e baixos valores de $s(k, k)$ resultam em grandes e baixos números de clusters encontrados, respectivamente. Um bom resultado é obtido adotando, para preferências, o valor médio das similaridades.

A troca de mensagens entre os pontos podem ser de dois tipos, transmissão de responsabilidades e disponibilidades. As responsabilidades $r(i, k)$ são mensagens enviadas de um ponto i para um possível exemplar k e indicam o quão adequado seria o ponto k ser exemplar para o ponto i . As disponibilidades $a(i, k)$, enviadas do candidato a exemplar k para o ponto i , indicam evidências de quão adequado seria para o ponto i escolher o ponto k como seu exemplar. A troca de mensagens pode ser visualizada na Figura 3.1.

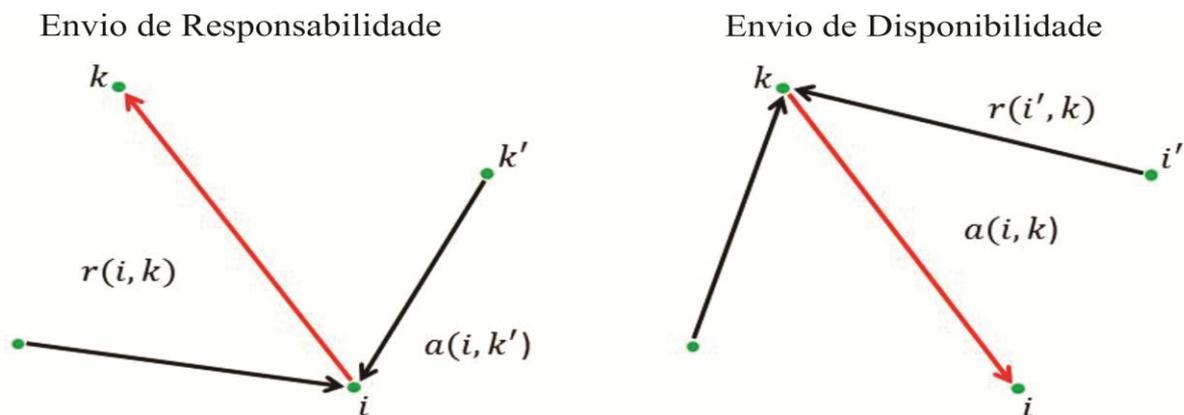


Figura 3.1. Ilustração da troca de mensagens entre os dados segundo AP.

Fonte: Adaptado de Frey e Dueck, 2007.

As responsabilidades são definidas pela seguinte regra:

$$r(i, k) \leftarrow s(i, k) - \max_{k' \neq k} \{a(i, k') + s(i, k')\}, \quad (3.1)$$

onde $a(i, k) = 0$, para a primeira interação. Para $i = k$, a responsabilidade $r(k, k)$ é vista como uma auto-responsabilidade, que reflete a evidencia de que k é um exemplar.

As disponibilidades, que definem se um candidato a exemplar é um bom exemplar, são assim definidas:

$$\begin{cases} a(i, k) \leftarrow \min \left\{ 0, r(k, k) + \sum_{i' \in \{i, k\}} \max \{0, r(i', k)\} \right\}, \\ a(k, k) \leftarrow \sum_{i' \neq k} \max \{0, r(i', k)\}. \end{cases} \quad (3.2)$$

Assim, a definição de um exemplar se dá quando a combinação de responsabilidades e disponibilidades, segundo abaixo, assume valor máximo.

$$AP = \max \{a(i, k) + r(i, k)\}. \quad (3.3)$$

4 FUNÇÃO DE DENSIDADE KERNEL

Como pode ser visto na última seção, a técnica de propagação de afinidade (Affinity Propagation) é apropriada para agrupamento dos dados sem a necessidade de informação prévia do número de classes, uma vez que cada ponto dado pode ser eleito como um exemplar. Sabendo-se que as informações litológicas das camadas são dificilmente obtidas, pode-se utilizar o Affinity Propagation para reconhecimento de padrões e agrupamento dos dados dos perfis relacionados com a presença de minerais. Por outro lado, o vetor de preferência pode ser calibrado com uma informação prévia; isso reduz o número de clusters e sinaliza para a existência de fortes candidatos a exemplares.

A Figura 2.3 mostra que o conjunto de dados M-N é tipicamente multimodal, mesmo para um modelo litológico simples. Esta característica também é frequentemente encontrada em dados reais. Esta tendência para formar agrupamento de pontos mostra a evidência de que, em cada camada, a litologia é expressa predominantemente por um número finito de minerais. Este comportamento pode ser usado para melhorar a performance do Affinity Propagation, pela modificação do vetor de preferências com a inserção da informação dos possíveis representantes para os agrupamentos dos dados. Estes pontos podem ser obtidos diretamente dos cross plots de dados dos perfis com o auxílio de uma estimativa não paramétrica de densidade (Silverman, 1981). Seja N um conjunto de pontos dados com dimensão d , i.e., $x_1, x_2, x_3, \dots, x_N$. A função de densidade de probabilidade para esta amostra de dados é dada por:

$$Pdf(x) = \frac{1}{Nh^d} \sum_{i=1}^N Ker\left(\frac{x-x_i}{h}\right), \quad (4.1)$$

Onde $Ker\left(\frac{x-x_i}{h}\right)$ é a função Kernel e h é o parâmetro de suavização ou largura de banda.

A função Kernel representa a contribuição individual de cada amostra de dados sobre densidade de todos os pontos. Uma vez que a função exata de distribuição para a amostra de dados é desconhecida, é possível utilizar uma variedade de funções Kernel (Silverman,

1981). Para este trabalho, foi escolhida a Gaussiana, pois sua forma assemelha-se a caracterização do problema de determinação de litologias por meio do Gráfico M-N: quanto mais próximo do mineral principal um ponto estiver maior será seu valor na função; quanto maior o distanciamento, menores valores, o que ocorre gradativamente.

A função gaussiana é dada:

$$Ker\left(\frac{x-x_i}{h}\right) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{\|x-x_i\|^2}{2h^2}}. \quad (4.2)$$

Finalmente a função de densidade de probabilidade é dada pela Eq. (4.1) e a Eq. (4.2) e que pode ser reescrita na Eq. (4.3),

$$Pdf(x) = \frac{1}{Nh^d (\sqrt{2\pi})^d} \sum_{i=1}^N e^{-\frac{\|x-x_i\|^2}{2h^2}}. \quad (4.3)$$

Figura 7. mostra um exemplo de uma função de densidade Kernel para os dados *M-N* relativos aos perfis sintéticos da (Figura 2.2). Pode-se notar que a função de Kernel é multimodal e tem três máximos locais. Cada máximo local representa a maior contribuição de um ponto sobre a densidade local; portanto, eles podem ser utilizados como o primeiro representante dos possíveis grupos ou aglomerados. Aqui, a ideia é usar estes máximos locais para modificação do vetor de preferência no AP.

5 ALGORITMO FIREFLY (FA)

Para obter todos os máximos locais deve-se aplicar um algoritmo de otimização eficiente com capacidade de busca multimodal. De forma geral, o problema de otimização tem a seguinte formulação:

$$\begin{cases} \text{minimiza/maximiza } f(x), \\ \text{objeto } x \in \Omega, \end{cases} \quad (5.1)$$

Onde x é uma variável tipo vetor, $\Omega \subset \mathbb{R}^n$ é o domínio e $f(x): \Omega \rightarrow \mathbb{R}$ é uma função real contínua. Para este trabalho o algoritmo de otimização estocástica Firefly (Yang, 2009) foi utilizado devido à sua simplicidade e pela característica multimodal das funções. O algoritmo Firefly é uma Metaheurística com base na inteligência de enxame (Yang et al, 2013) que simula o comportamento dos vaga-lumes, principalmente no que diz respeito à característica de luminosidade. Podemos resumir este comportamento por três regras simples:

1. Todos os vaga-lumes são unissex e atraem uns aos outros.
2. A atratividade é proporcional ao brilho e ambos diminuem à medida que a distância entre quaisquer dois vaga-lume aumente.
3. O brilho ou a intensidade da luz depende da função objetivo a ser otimizada.

A atratividade é uma função $\beta(r)$ determinada pela equação:

$$\beta(r) = \beta_0 e^{-\gamma r^2}, \quad (5.2)$$

onde β_0 é a atratividade em $r=0$, γ é o coeficiente de absorção de luz e r é a distância entre dois pontos. O deslocamento de um vagalume i em uma posição x_i , atraído

por outro vagalume j com maior atratividade (luminosidade) em uma posição x_j é determinado pela equação:

$$x_i(t+1) = x_i(t) + \beta_0 e^{-\gamma r^2} (x_j - x_i) + \alpha \varepsilon_i. \quad (5.3)$$

Na Eq. (5.3), α é um coeficiente randômico e ε_i é um vetor de valores aleatórios baseados em uma distribuição uniforme no instante i . Os parâmetros α , γ e β_0 dependem do domínio de escala da função objetivo; em geral $\alpha \in [0,1]$ e $\gamma, \beta_0 \in [0.9,1]$. É necessário determinar, ao iniciar o algoritmo, o número de vaga-lumes (soluções) e gerações (interações) (Yang, 2010).

6 METODOLOGIA

Nesta seção, mostramos o algoritmo híbrido que consiste em oito etapas principais (**Figura 6.1**). Na primeira etapa, os dados que contêm informações sobre os perfis de porosidade (densidade, Sônico e nêutrons), raio gama e as profundidade associados a cada medida. O segundo passo é calcular os vetores de parâmetros litológicos N e M com os valores de registo dos perfis sobre a zona de estudo. Eq. (2.5) e Eq. (2.6) são usadas para o cálculo dos parâmetros N e M em função de cada profundidade z . Em seguida, é definido o parâmetro h de suavização. A idéia é ajustar a largura de banda do kernel no que diz respeito à distribuição dos ponto $M-N$. Isto significa que, se h aumenta ocorre a suavização da função kernel e aumenta a tendência distribuição normal; Por outro lado, se h diminui uma tendência de distribuição mais sinuosa pode ser observada. Uma vez que os pontos de dados $M-N$ estão relacionadas com os minerais da litologia, isto é, espalhamento de pontos em torno dos minerais principais, pode-se esperar que a estimativa de densidade Kernel apresente uma distribuição multimodal. Portanto, uma estratégia comum para escolha do parâmetro h (Silverman, 1981; Turlach, 1993; Wang, 2004; Zucchini, 2003) é:

$$h = \frac{1.06 \cdot s \cdot esc}{n^{1/5}}, \quad (6.1)$$

Onde n é o número de amostras $M-N$, esc é um fator de escala e s é o desvio padrão da amostra dado por:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (absMN_i - \overline{absMN})^2}. \quad (6.2)$$

Na Eq. (6.2), $absMN_i$ é a amplitude de cada ponto no espaço $M-N$, i.e., $absMN_i = \sqrt{N_i^2 + M_i^2}$ e \overline{absMN} é a média aritmética de todos os pontos $absMN_i$. Na Eq. (6.1), o parâmetro esc comporta-se como um sintonizador de suavidade, ou seja, se esc é igual a 1, temos a clássica aproximação de Silverman's (Silverman, 1981) para um ótimo valor de h . Para outros

valores, tanto h pode sofrer acréscimo (valores de esc maiores que 1 suavizam os valores kernel) ou decréscimo (esc menor que 1 tornam a função kernel bem ondulada).

Em seguida, é definida a função de distribuição Kernel relativa aos dados M-N e o parâmetro de largura de banda h , pelo uso da Eq.(6.1). Para determinação dos máximos da função de densidade kernel usa-se o algoritmo Firefly.

Na terceira etapa, o algoritmo Firefly é inicializado com uma distribuição aleatória (distribuição de Sobol) de 200 vagalumes no espaço $(N \times M)$ para um número máximo de 200 gerações, com coeficiente randômico $\alpha = 0.9$, coeficiente de redução randômico $\varepsilon = 0.9$ e coeficiente de absorção γ dado pela equação:

$$\gamma = \gamma_0 / \sqrt{(Max(N) - Min(N))^2 + (Max(M) - Min(M))^2} \quad (6.3)$$

Onde γ_0 é o coeficiente de absorção inicial usualmente definido como 1 (Yang, 2010).

Em sequencia o vetor de preferências \mathbf{P} é definido da seguinte forma: Primeiramente, calcula-se a distância entre os pontos obtidos pelo Firefly e os pontos (N, M) . Os pontos (N_i, M_i) cuja distância assuma valores mínimos serão relacionados com maiores valores de preferência no vetor \mathbf{P} (posição i no vetor) e, para os outros pontos, os valores de preferência serão mantidos como a média dos valores da matriz de similaridades. Baseado na matriz de similaridades e no vetor de preferências, o AP particiona o conjunto de dados (N, M) em clusters e os respectivos exemplares são armazenados (etapa 3). A classificação (etapa 4) segue pela relação dos exemplares e minerais fixos (**Tabela 1**) que compõem as principais litologias nas rochas sedimentares. A associação ocorre pela mínima distância euclidiana entre os exemplares definidos pelo AP e pontos de mineral fixos. O conjunto de pontos em um cluster será classificado como pertencente à litologia com a qual seu exemplar foi associado. Por fim, após a etapa de classificação, as litologias definidas são apresentadas em

profundidade, o que permite uma posterior comparação das litologias obtidas pela metodologia com os resultados já analisados por meio dos testemunhos.

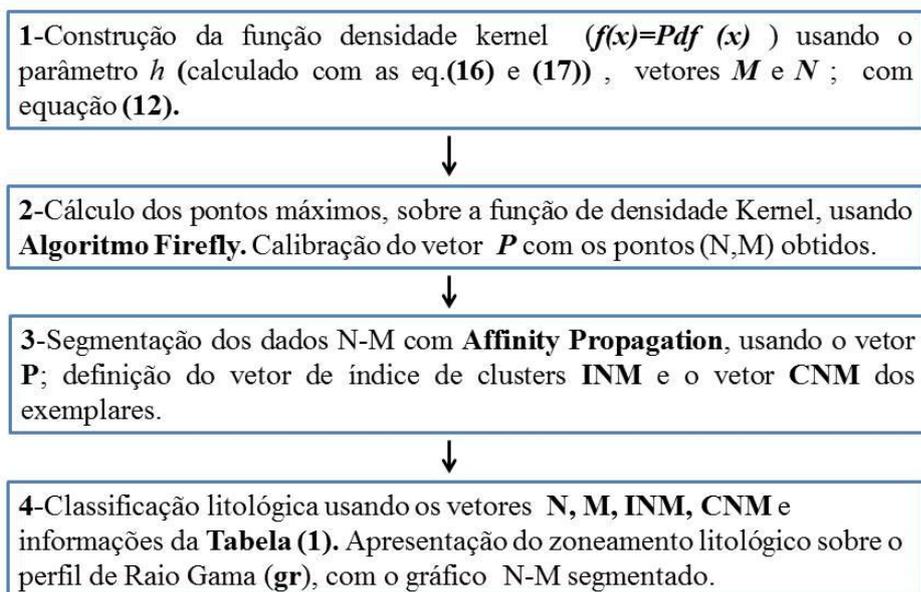


Figura 6.1. Algoritmo híbrido para mapeamento litológico.

7 RESULTADOS

A metodologia foi aplicada em dados sintéticos e reais. O conjunto de dados satisfazem as equações gerais dos perfis de porosidade. Primeiramente, aplica-se a metodologia em dados sintéticos, cujos perfis satisfazem um modelo composto por uma sequência de diferentes litologias em profundidade. Em sequência, a metodologia é aplicada em dados reais para inferir as litologias em profundidade, ainda que na ausência de informação a priori (testemunhos). Os dados reais aqui apresentados são do campo de Namorado, Bacia de Campos, Brasil (Bacoccoli *et al.*, 1980; Rangel *et al.*, 1994).

7.1 DADOS SINTÉTICOS

O uso de dados sintéticos é uma forma de avaliar a eficiência da metodologia devido ao controle sobre o dado processado e o conhecimento do modelo litológico. Pela comparação e interpretação dos dados processados e não processados do Gráfico *M-N* é possível avaliar o quanto de compatibilidade existe entre a interpretação obtida e o modelo litológico conhecido. A **Figura 7.1** apresenta o modelo sintético que consiste de uma sequência de camadas representadas em cores, cujas características geológicas estão descritas na *Tabela 7.1*. Na Figura 6 mostra-se o perfil de raio gama (GR), os perfis de porosidade, i.e, densidade (ρ_b), porosidade neutrônica (ϕ_n) e tempo de trânsito (Δt), respectivamente, referentes ao modelo exibido.

Tabela 7.1: Descrição Geológica das camadas do modelo sintético utilizado

Camada	Litologia	Composição mineral (%)
Azul	Cálcario	Calcita-100%
Vermelho	Dolomita	Dolomita-100%
Verde	Arenito	Quartzo -100%
Cinza	Folhelho	Argilominerais-100%

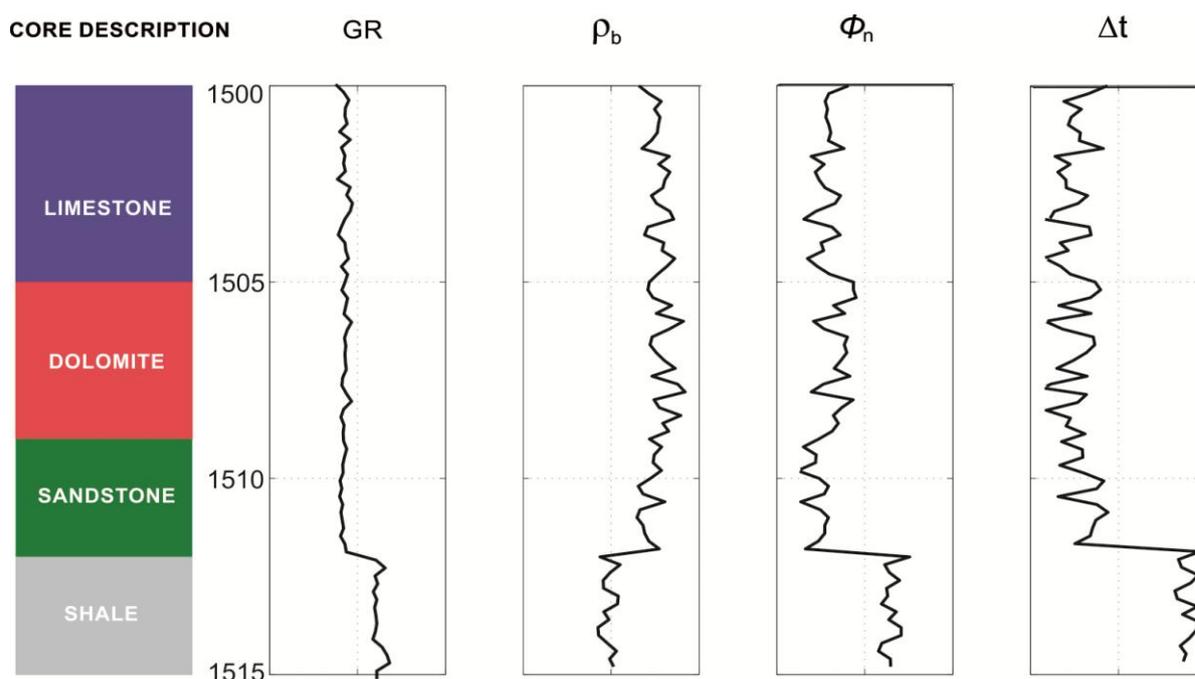


Figura 7.1. Perfis de dados sintéticos. Fonte: Da autora

Na sequência, os parâmetros M e N são definidos com base nos perfis apresentados na **Figura 7.1**, com uso da **Eq. (2.5)** e **Eq. (2.6)**. Os pares (N, M) são exibidos no Gráfico $M-N$, **Figura 7.2**. De acordo com a interpretação visual e quantitativa, baseada na distância euclidiana entre os pontos obtidos pelos perfis e os pontos de mineral fixo, é possível classificar cada ponto, em profundidade, de acordo com a litologia a que está associado. Inicialmente, é definido o vetor de preferências com o auxílio do algoritmo Firefly para busca e definição dos máximos da função de densidade Kernel. O resultado é definido como um conjunto de pontos (N, M) que recebem os mais altos valores de preferência no vetor de Preferências P . **Figura 7.3 (a)** mostra o conjunto de pontos $M-N$ para os dados sintéticos (pontos pretos), a função de densidade de kernel representado pela superfície em vermelho. Na **Figura 7.3 (b)** estão dispostos os valores máximos obtidos pelo algoritmo de FA e as curvas de níveis da superfície da **Figura 7.3 (a)**.

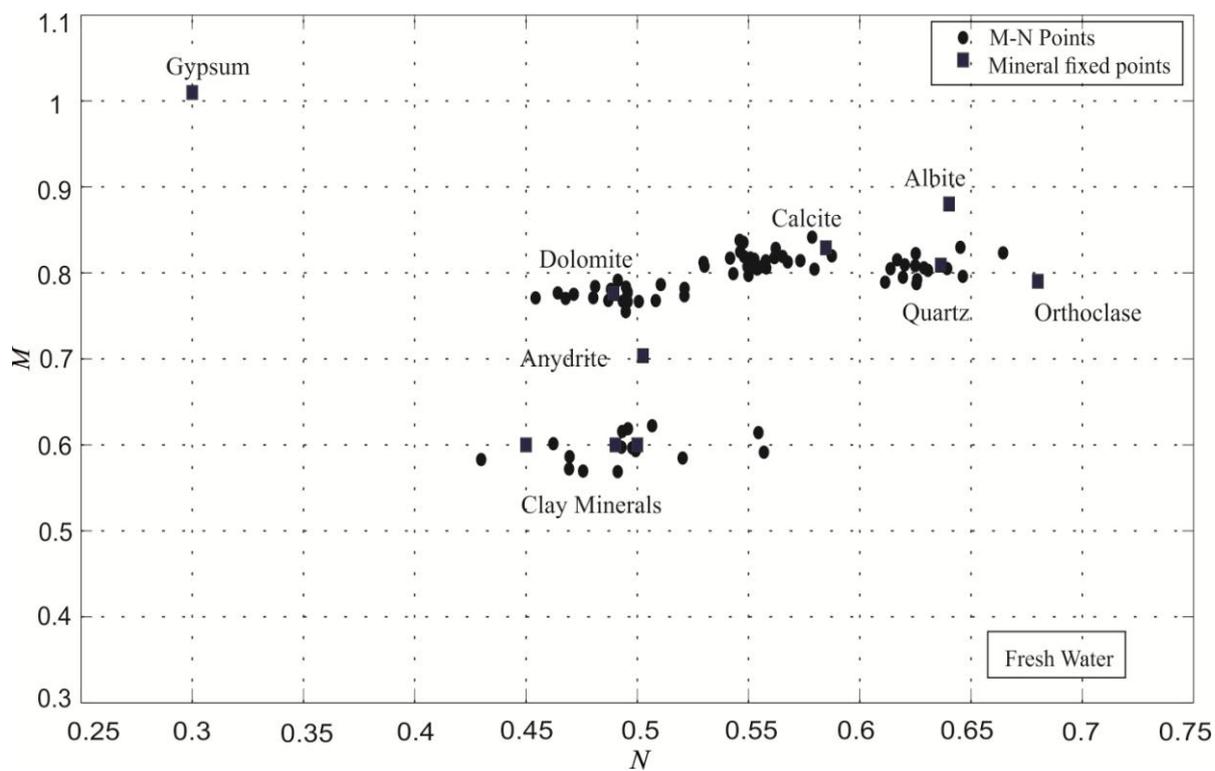


Figura 7.2. Gráfico $M-N$. Fonte: Da autora.

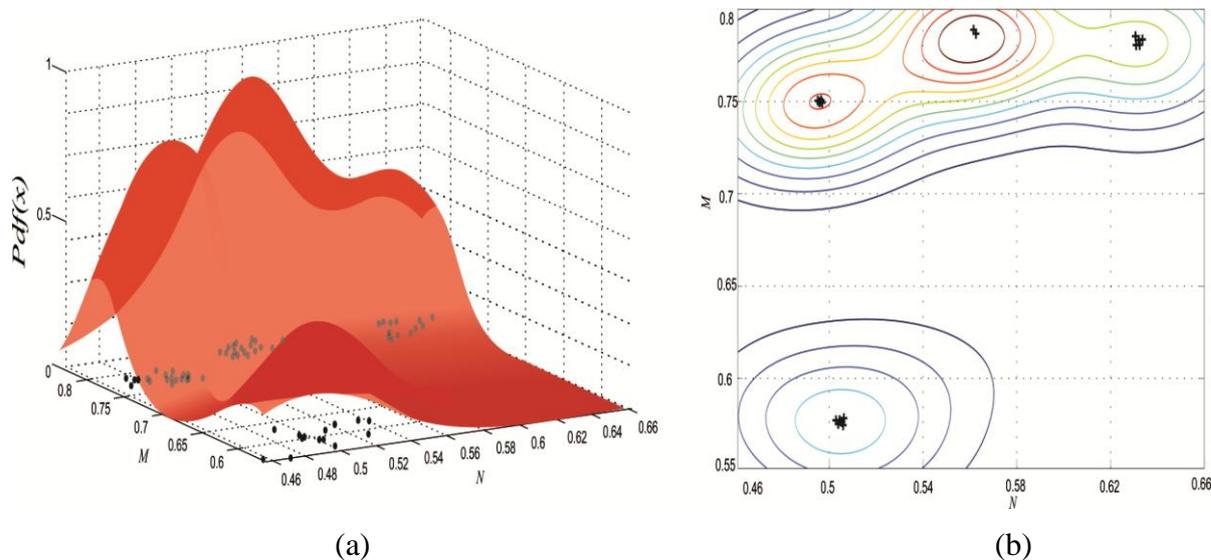


Figura 7.3. (a) Dados $M-N$ (preto), função Pdf (superfície em vermelho). (b) Máximos obtidos com FA (pontos pretos). Fonte: Da autora.

Na **Figura 7.4 (a)** pode-se observar a presença de oito clusters resultantes do processamento AP cujo vetor preferência é definido pela média da matriz de similaridade. Ao aplicar o AP com variações na preferência do vetor (máximos obtidos da FA) pode ser observada uma redução do número de agrupamentos para quatro (**Figura 7.4 (b)**).

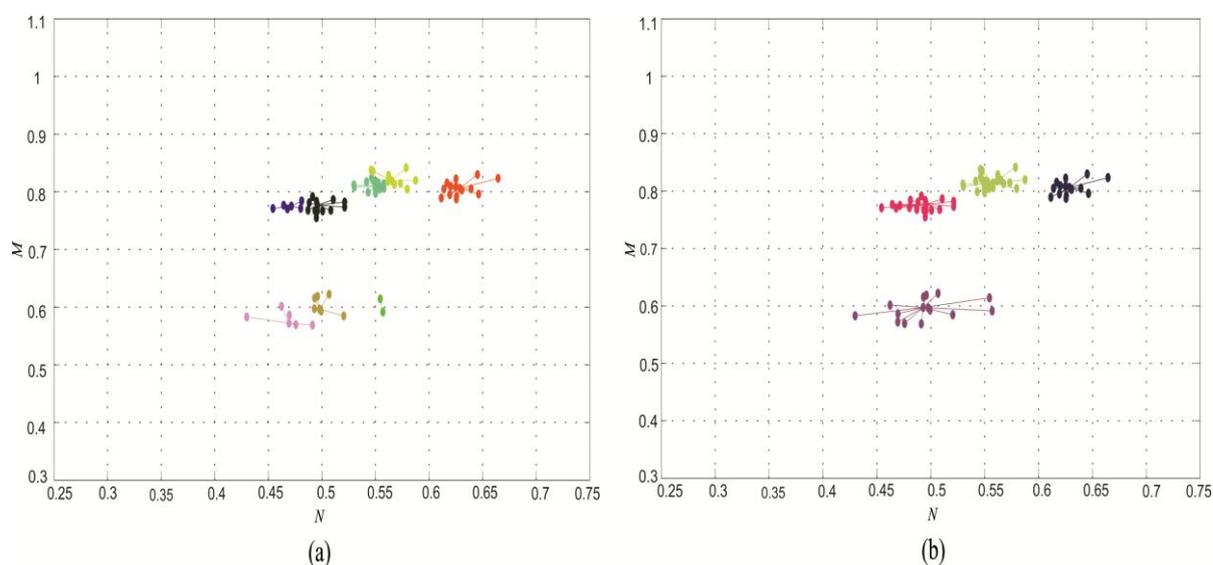


Figura 7.4. (a) Agrupamentos resultantes do processamento AP com vetor de preferência usual. (b) Clusters resultantes do processamento AP com vetor de preferência modificado.
Fonte: Da autora

A classificação da litologia é feita a partir da associação do exemplar de cada cluster com os minerais de pontos fixos do Gráfico $M-N$. Esta associação é dada por uma função mínima distância entre os respectivos exemplares e pontos fixos, ou seja, um exemplar é classificado como pertencente a uma particular litologia se o mineral predominante, compondo a matriz, é o ponto fixo mais próximo. O mapeamento litológico em profundidade e a classificação são apresentados na **Figura 7.5**.

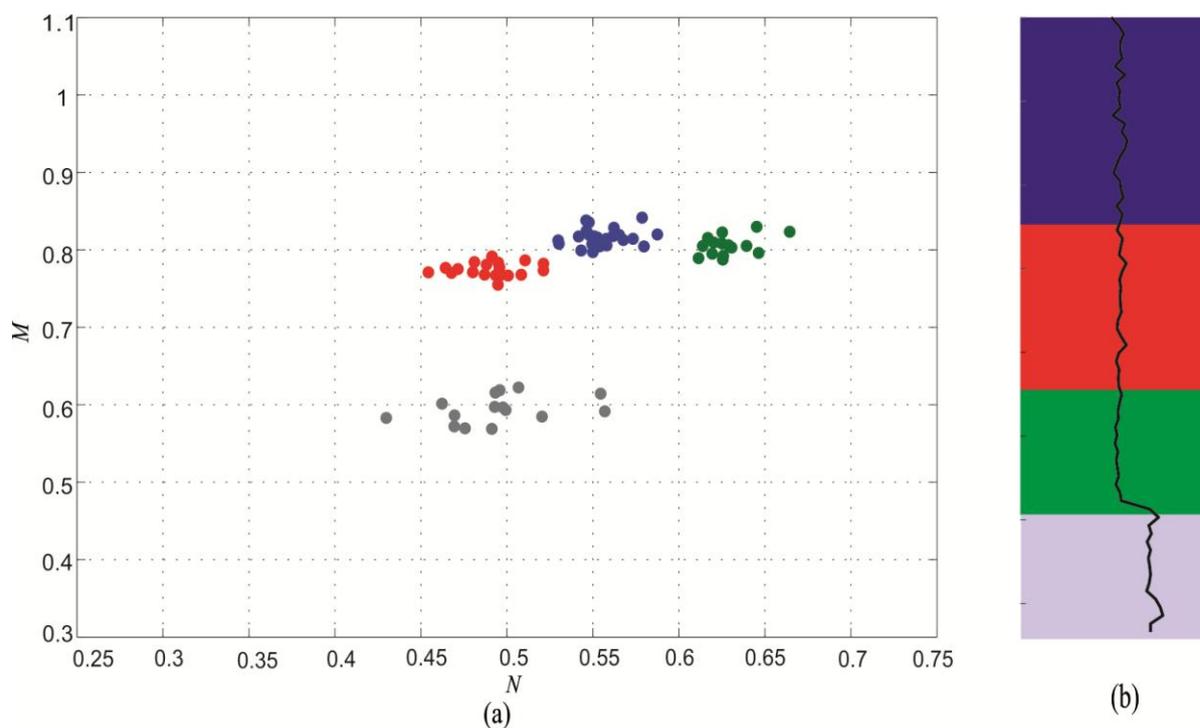


Figura 7.5. (a) Classificação final por associação de exemplares. (b) Mapeamento de Litologias em profundidade. Fonte: Da autora.

7.2 DADOS REAIS

Os dados reais selecionados são referentes aos perfis de seis poços no Campo de Namorado (Bacocoli, 1980) cuja análise de testemunhos foi disponibilizada, permitindo uma posterior comparação litológica com os resultados da metodologia.

7.2.1 Dado Real 1

Figura 7.6 mostra um conjunto de perfis registrados na Bacia de Campos. Os testemunhos estão descritos na *Tabela 7.2* para os intervalos de profundidade selecionados (realçado em verde e cinza nas curvas dos perfis)

Tabela 7.2: Descrição Geológica do dado real 1 utilizado

Camada	Litologia	Composição Mineral
Verde	Arenito	Quartzo
Cinza	Folhelho Radioativo	Argilominerais

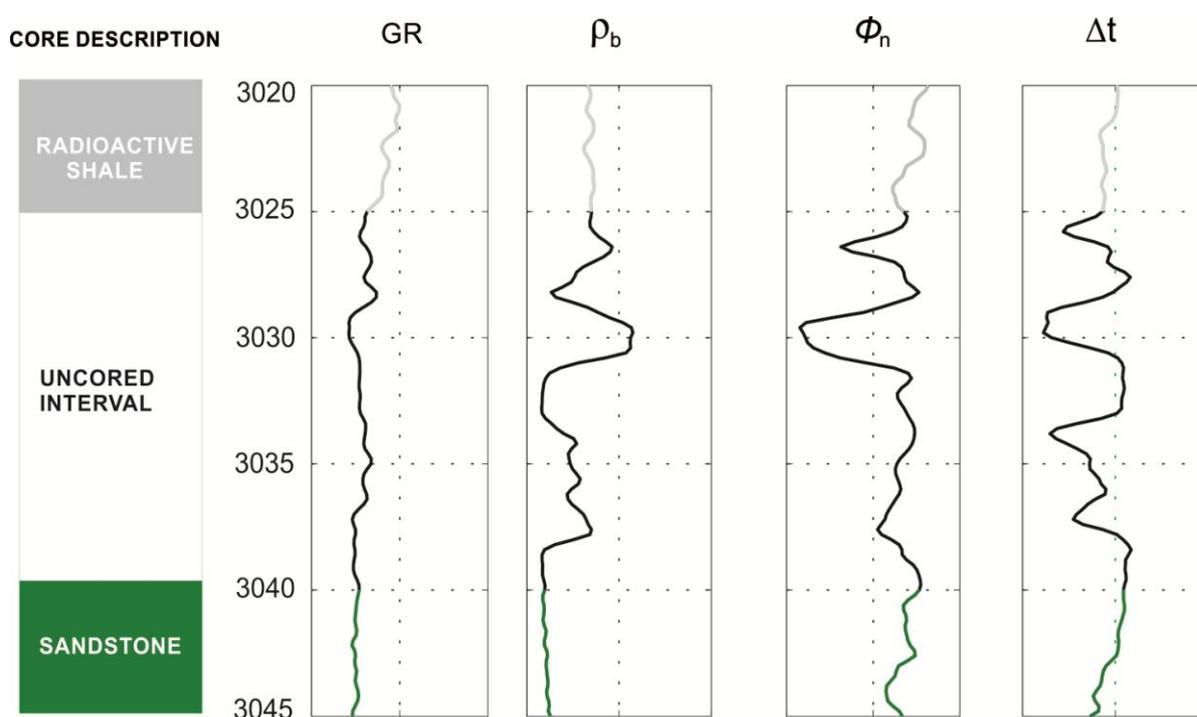


Figura 7.6. Perfis de Dado Real. Fonte: Da autora

Figura 7.7 apresenta o gráfico de M-N baseado nos perfis apresentados na **Figura 7.6**. Os círculos em preto representam os pontos (N, M) e os quadrados pretos os principais minerais de pontos fixos. A interpretação visual do Gráfico M-N sugere a existência de três grupos compactos e bem definidos. No entanto, a interpretação visual não necessariamente reflete o número de diferentes litologias in situ. Um problema comum na avaliação de formação é o deslocamento dos pontos (N, M) devido a argilosidade na direção dos argilominerais, impedindo uma classificação mais fiel. De acordo com o método proposto, define-se uma função de densidade Kernel, com base nos pontos (N, M) para obtenção dos

valores máximos dessa função. Os pontos (N, M) mais próximos dos pontos máximos da função Kernel recebem valores elevados no vetor P de preferências.

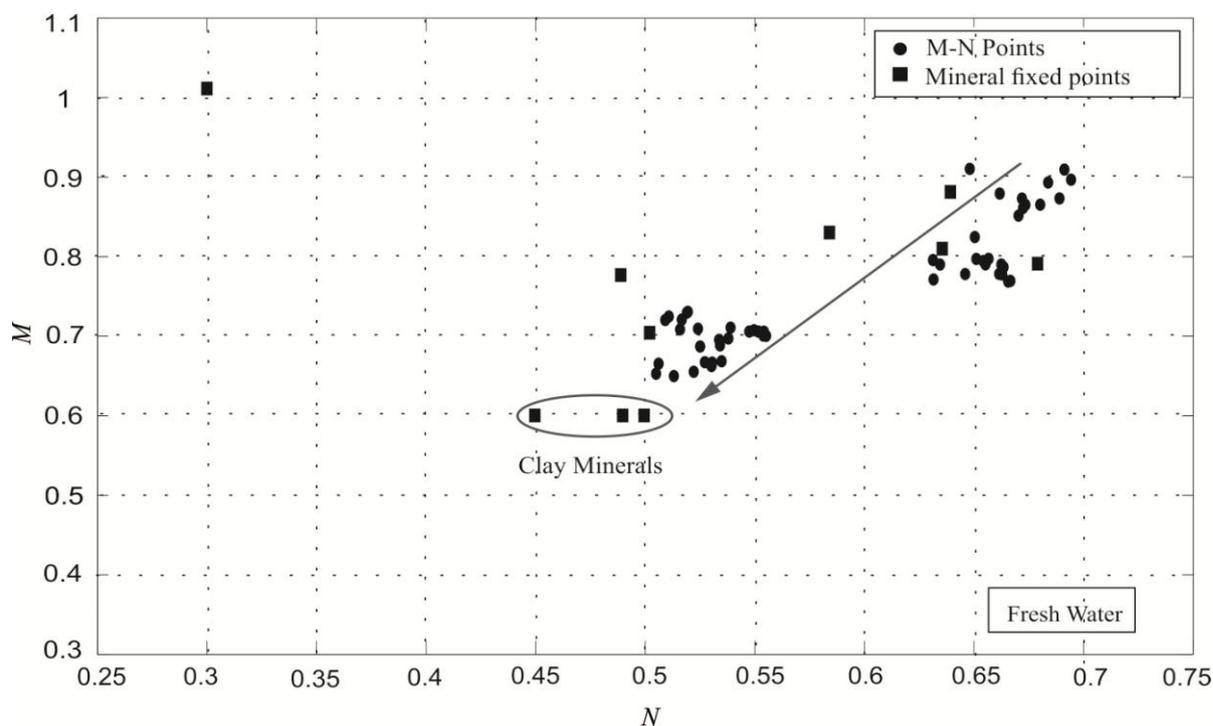


Figura 7.7. Gráfico M-N com três agrupamentos sugeridos pela interpretação visual. Fonte: Da autora

Na **Figura 7.8 (a)**, os resultados do algoritmo AP podem ser visto sobre o Gráfico M-N. Observa-se a presença de cinco clusters resultantes do algoritmo AP sem alteração do vetor de preferência. Ao aplicar o algoritmo de AP com o vetor de preferências modificado, com base em pontos máximos obtidos da FA, o número de grupos é reduzido para três (**Figura 7.8(b)**). A classificação final pode ser visualizada na **Figura 7.9**. O primeiro intervalo, em profundidade, está associado aos minerais componentes do folhelho (argilominerais), o segundo intervalo é associado com o ponto fixo representante do mineral quartzo, principal componente do arenito.

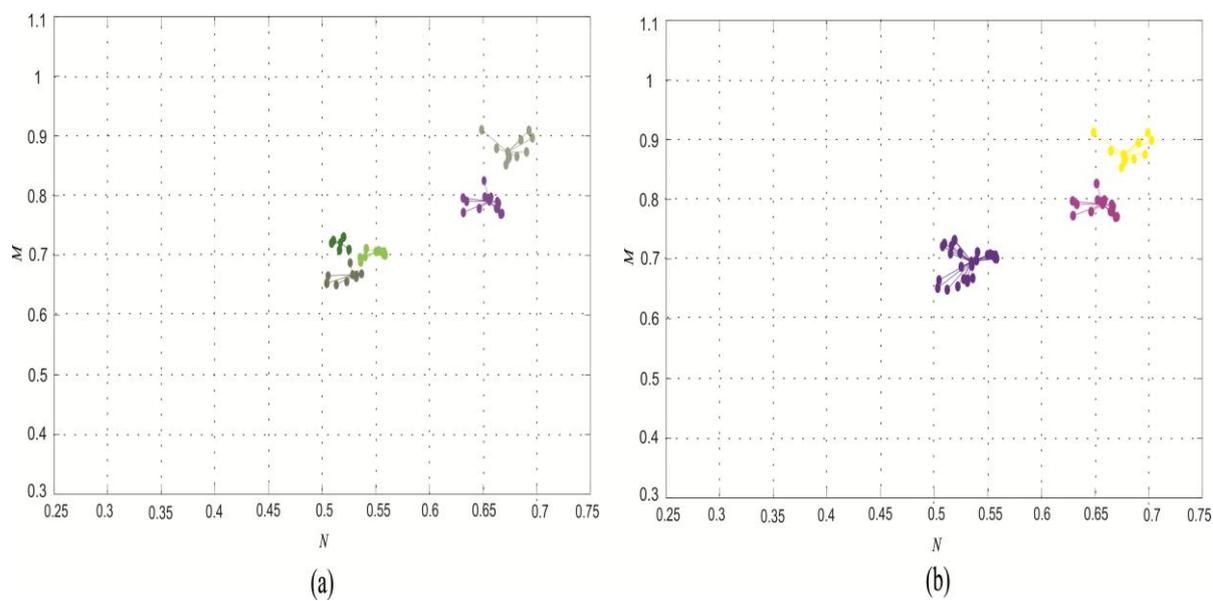


Figura 7.8. (a) Resultados do agrupamento com AP com o vetor de preferência usual.(b) Resultados do agrupamento com AP com as modificações do vetor de preferência.

Fonte: Da autora

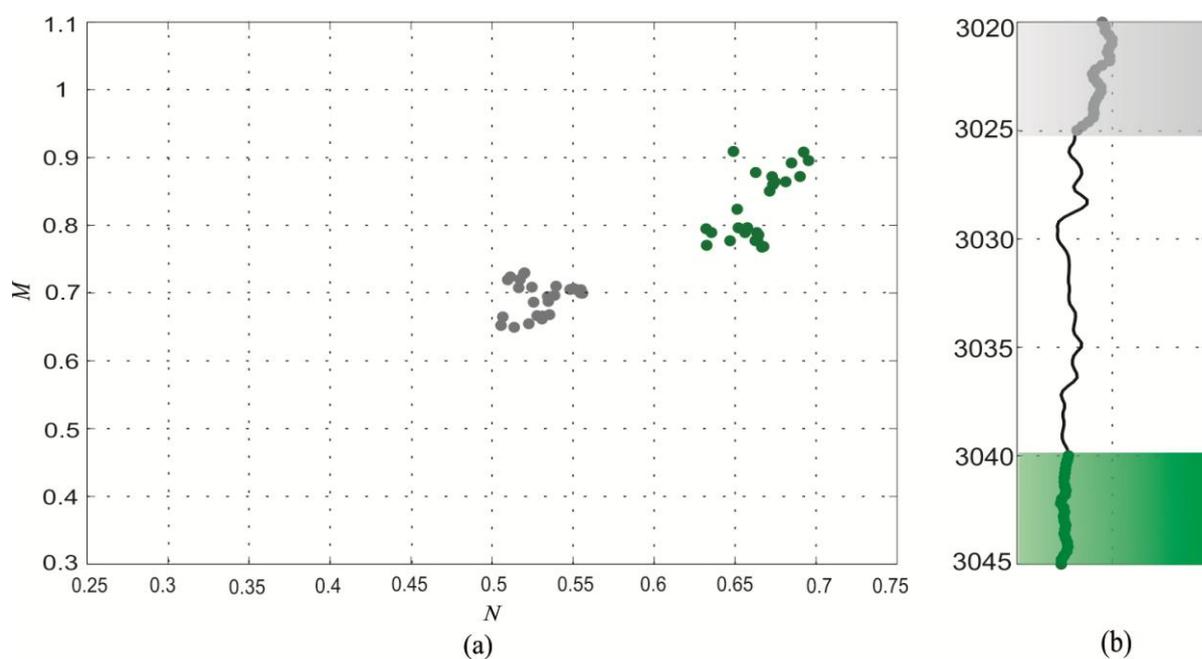


Figura 7.9. (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade. Fonte: Da autora

7.2.2 Dado Real 2

Figura 7.10 apresenta outro dado real de perfis registrados na Bacia de Campos. As amostras de testemunhos estão descritas na *Tabela 7.3* para os intervalos de profundidade selecionados (realçado em cinza e verde no perfil). **Figura 7.11** apresenta os pontos (N, M) no gráfico M-N dos perfis representados na **Figura 7.10**. A interpretação visual do Gráfico M-N torna-se um pouco mais complexa devido à proximidade de pontos, a sua dispersão e deslocamento em direção aos pontos fixos representativos dos argilominerais.

Tabela 7.3: Descrição de testemunhos do poço 2.

Camada	Litologia	Composição da Matriz (%)
Verde	Arenito	Quartzo
Azul	Folhelho	Argilominerías

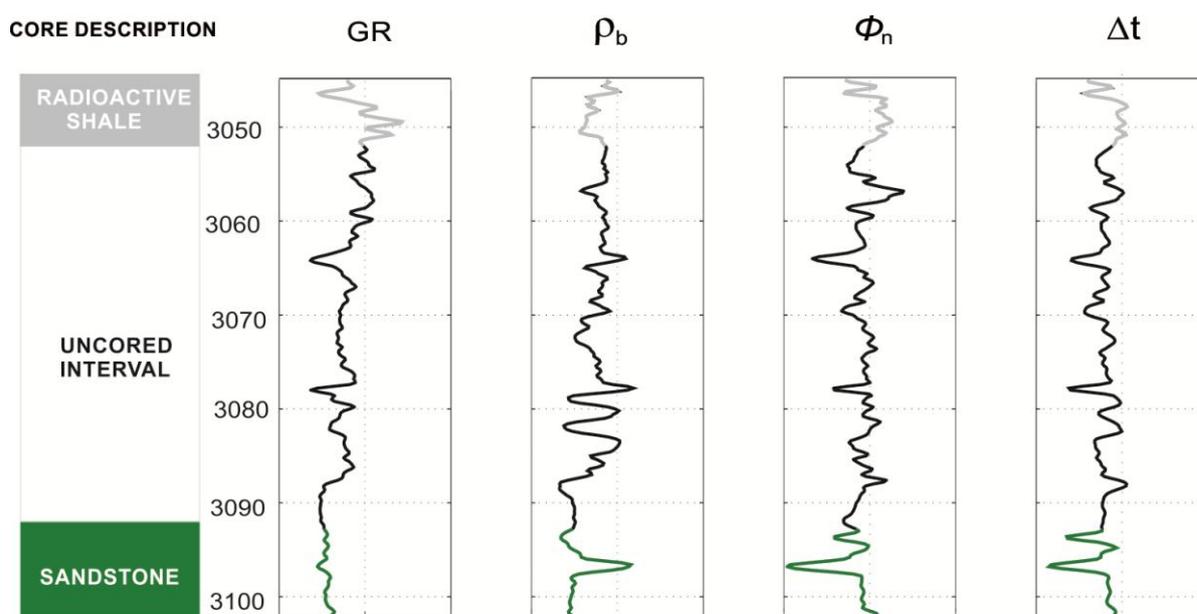


Figura 7.10. Testemunho (à esquerda), perfil de Raio Gama e perfis de porosidade. Intervalos selecionados destacados em cinza e verde. Fonte: Da autora

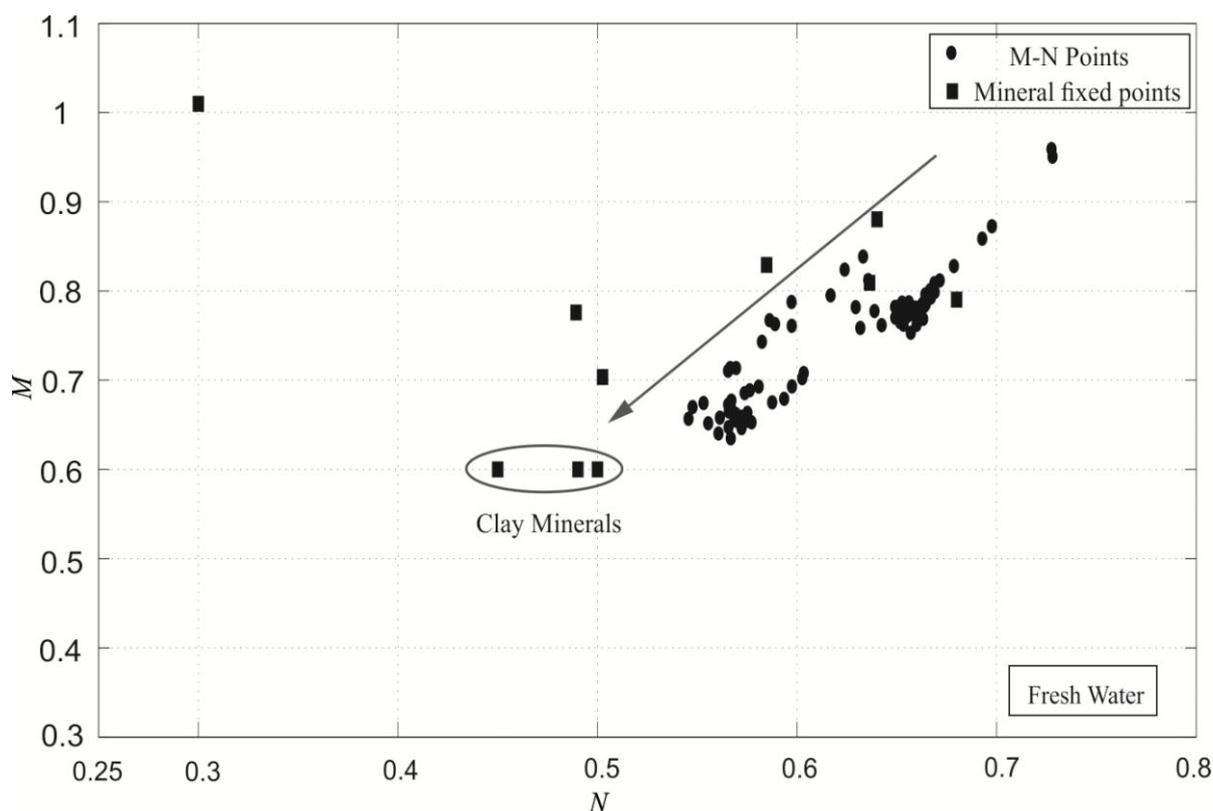


Figura 7.11. Gráfico M-N mostrando um espalhamento de pontos e deslocamento na direção dos argilominerálias. Fonte: Da autora

De acordo com o método, a função de densidade Kernel é definida com base nos pontos (N, M) para obtenção dos máximos pelo FA. Os pontos (N, M) cuja distância é a mínima com os pontos obtidos pelo FA receberam os valores mais elevados no vetor de preferência.

Os resultados da aplicação do algoritmo de agrupamento AP podem ser vistos no Gráfico M-N com os exemplares de cada cluster, **Figura 7.12(a)**. Observa-se a presença de dez clusters resultantes do agrupamento com vetor de preferência definido pela média da matriz de similaridades. Ao aplicar o algoritmo de AP com as mudanças no vetor de preferências, com base nos valores máximos do FA, um resultado diferente pode ser visualizado com a existência apenas três grupos (**Figura 7.12 (b)**).

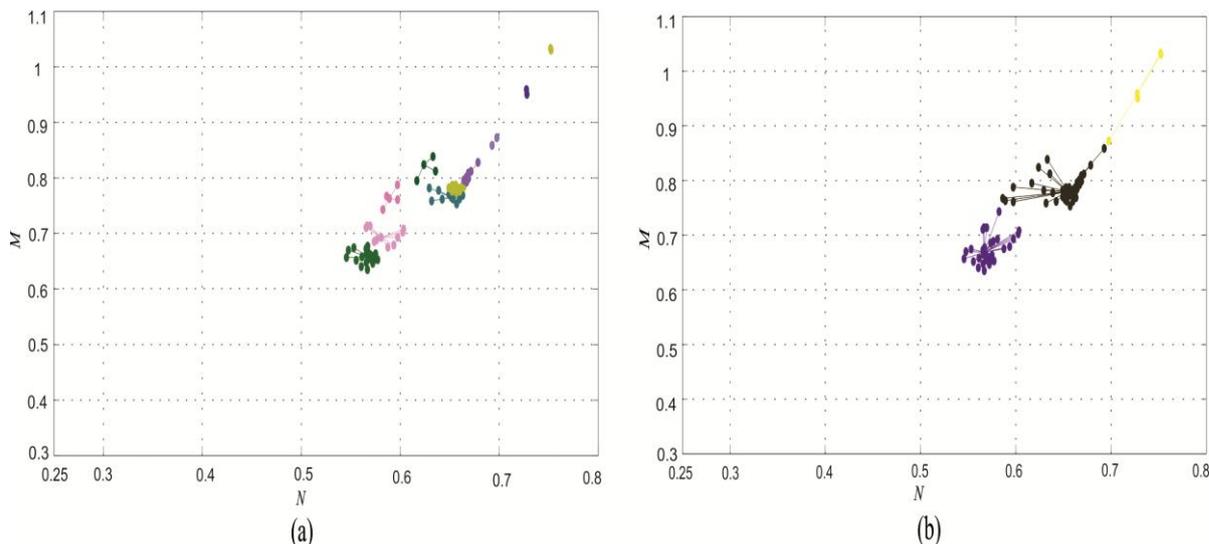


Figura 7.12. (a) Resultado de dez clusters resultantes do AP. (b) Resultados com AP com modificação do vetor P. Fonte: Da autora

A classificação final está exibida na **Figura 7.13**. O primeiro intervalo, em profundidade, está associado aos minerais componentes do folhelho (argila) e o segundo intervalo está associado com o ponto fixo representante do mineral quartzo, o principal mineral do arenito. A classificação final apresenta um erro de quatro pontos que pertencem à litologia folhelho classificados como arenito. Este tipo de erro é típico quando se avalia um conjunto de dados cujos pontos são muito dispersos. À medida que os algoritmos de agrupamento usuais classificam os conjuntos de dados de acordo com as suas semelhanças que, por sua vez, são dadas por alguma métrica (geralmente a distância Euclidiana), a caracterização dos grupos torna-se complicado no caso de dados que estão muito próximos e têm características completamente diferentes. Tal fenômeno é comum quando se trabalha com perfis de poços cuja interpretação é dada pelo Gráfico M-N ou gráficos semelhantes devido ao ruído inerente aos dados e os modelos interpretativos dotados serem de certo modo simplificados. Especificamente para o Gráfico M-N, a presença de argila, a porosidade secundária e o conteúdo de fluidos sob a forma óleo ou de gás causa distanciamento dos pontos fixos dos principais minerais (Burke, 1969). Na **Figura 7.13 (b)**, os pontos isolados cuja classificação varia em comparação com a maioria dos pontos para a mesma faixa de profundidade são, na verdade, pontos de folhelho.

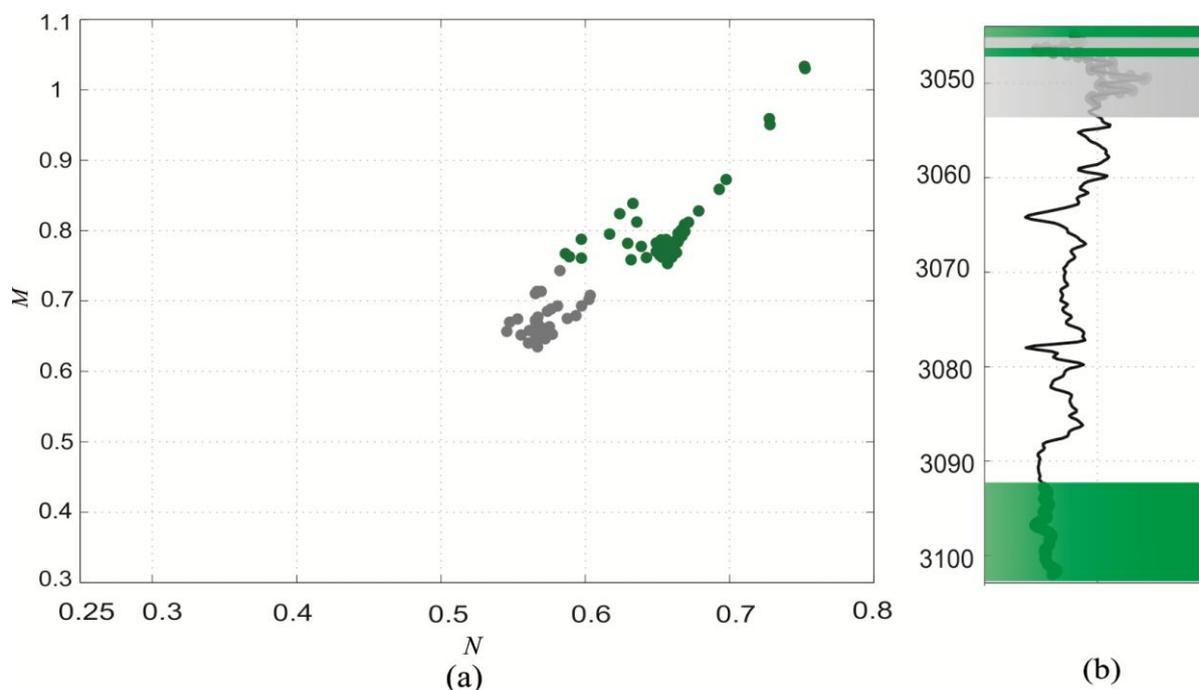


Figura 7.13. (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade. Fonte: Da autora

7.2.3 Dado Real 3

Figura 7.14 apresenta outro dado real de perfis registrados na Bacia de Campos. As amostras de testemunhos estão descritas na *Tabela 7.4* para os intervalos de profundidade selecionados (realçado em verde, cinza e azul no perfil). **Figura 7.15** apresenta os pontos (N, M) no gráfico M-N dos perfis representados na **Figura 7.14**. A interpretação visual do Gráfico M-N torna-se complexa devido ao espalhamento dos pontos.

Tabela 7.4: Descrição geológica do poço 3.

Camada	Litologia	Composição Mineral (%)
Azul	Calcário	Calcita(CaCO ₃)- 35/50%
Verde	Arenito	Quartzo -100%
Cinza	Folhelho	Siltito/Marga interlaminar

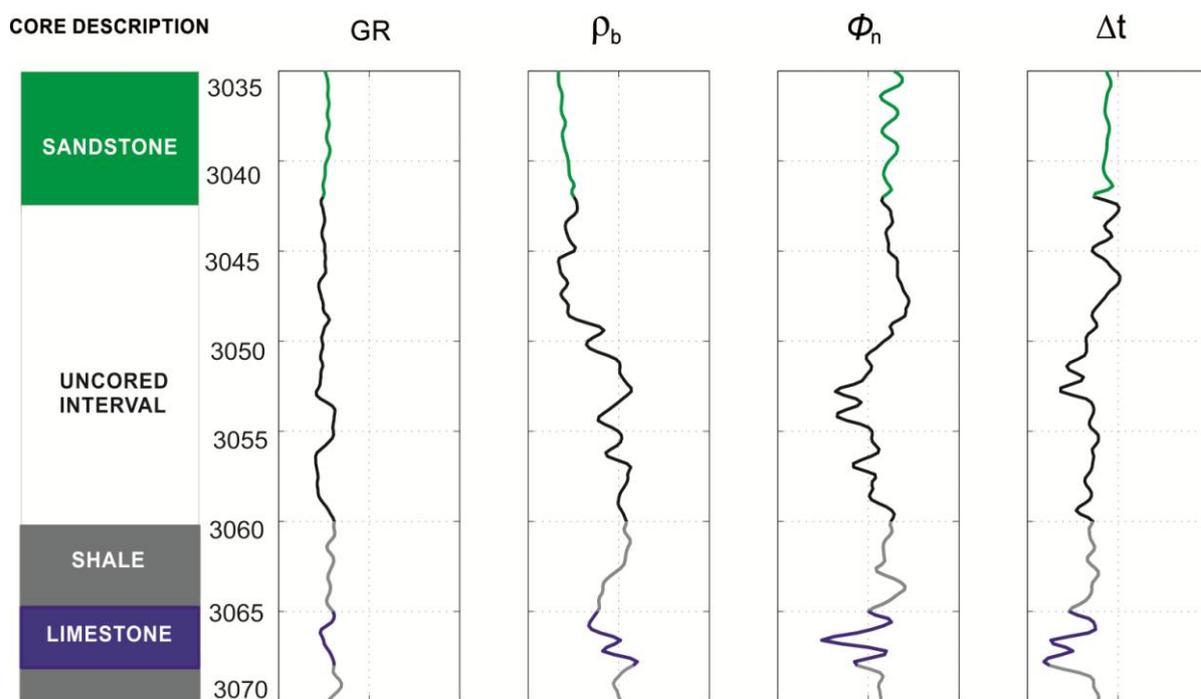


Figura 7.14. Testemunho (à esquerda), perfil de Raio Gama e perfis de porosidade. Intervalos selecionados destacados em vermelho. Fonte: Da autora.

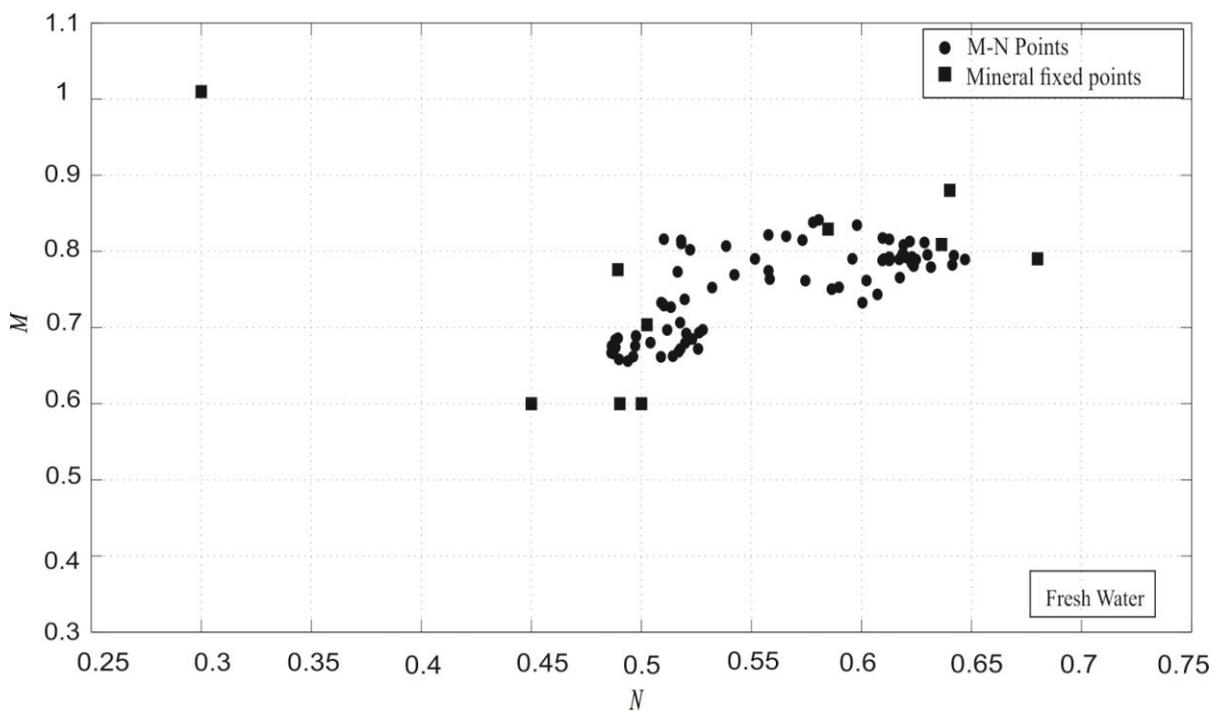


Figura 7.15. Gráfico $M-N$ obtidos com os dados do perfil. Fonte: Da autora

Na **Figura 7.16 (a)** estão os resultados obtidos pelo algoritmo AP. Verifica-se a presença de oito agrupamentos para o vetor de preferências definido como a média da matriz de similaridades. Na **Figura 7.16(b)** podemos visualizar três agrupamentos resultantes da mudança no vetor de preferência.

A classificação final está disposta na **Figura 7.17**. O primeiro intervalo de profundidade está associado ao mineral principal componente do arenito (quartzo) e o segundo intervalo associado aos argilominerías (folhelho) e a calcita, principal componente do calcário.

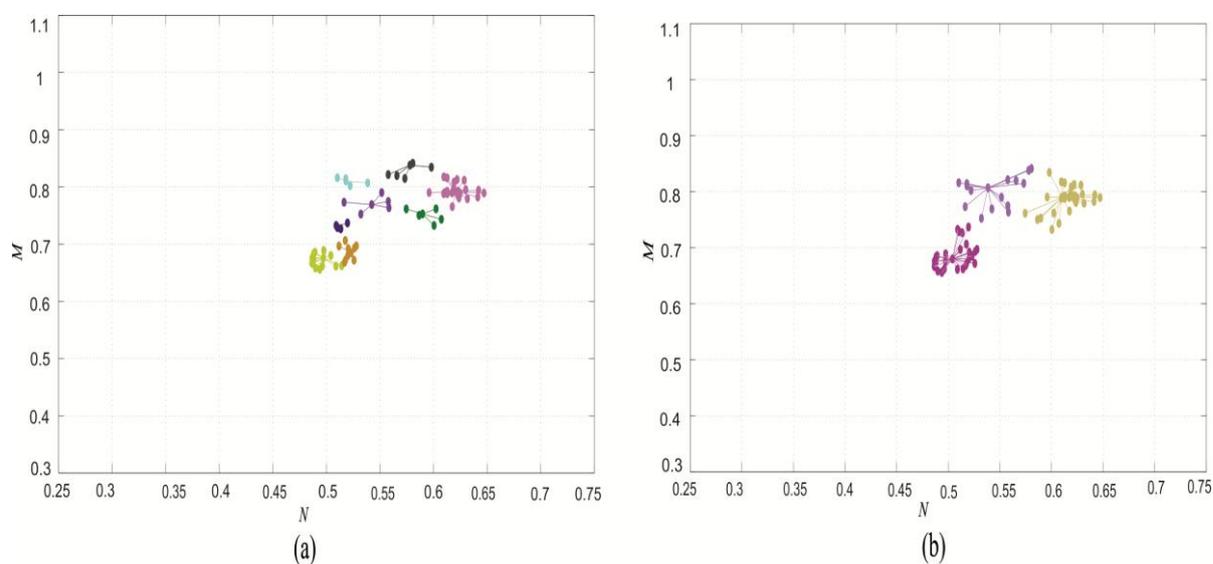


Figura 7.16. (a) Resultado de dez clusters resultantes do AP. (b) Resultados com AP com modificação do vetor P. Fonte: Da autora.

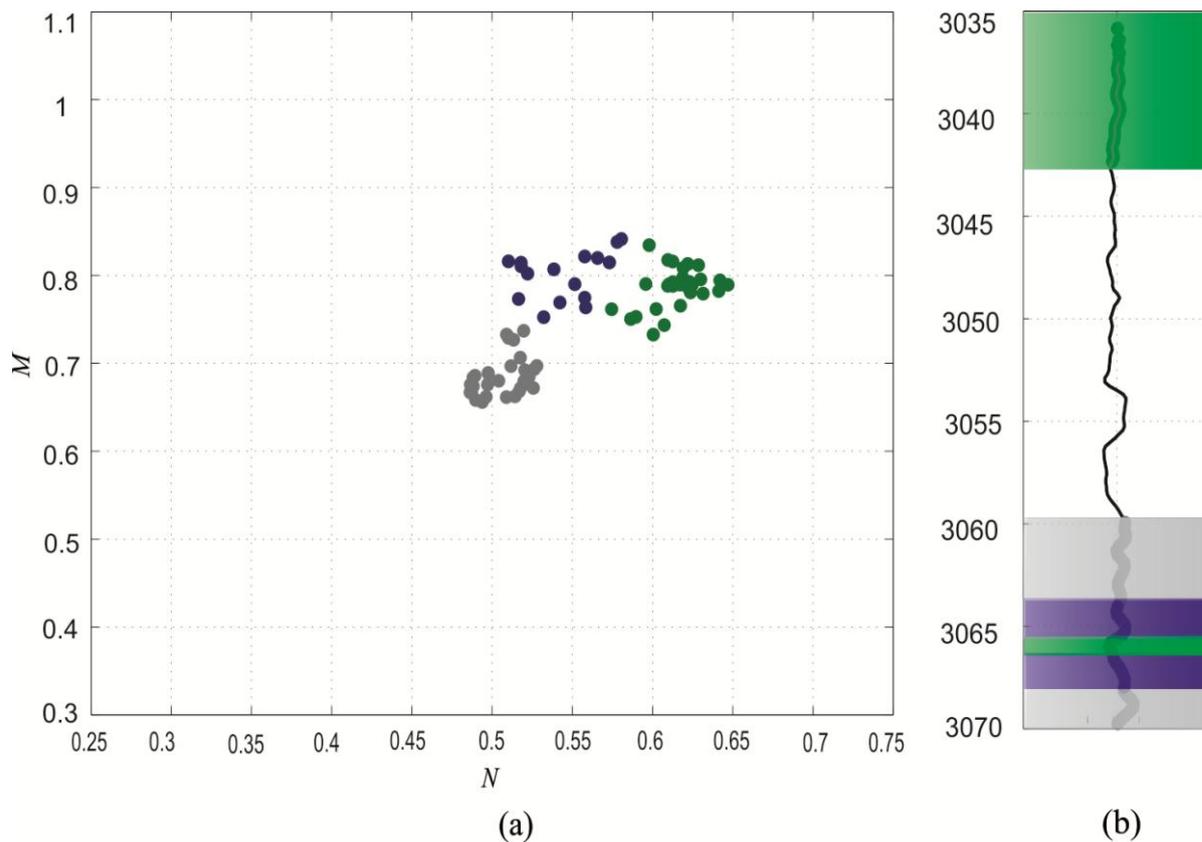


Figura 7.17. (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade. Fonte: Da autora

7.2.4 Dado Real 4

Figura 7.18 apresenta um dado real de perfis registrados na Bacia de Campos. As amostras de testemunhos estão descritas na *Tabela 7.5* para os intervalos de profundidade selecionados para processamento (curvas em preto). **Figura 7.19** apresenta os pontos (N, M) no gráfico M-N dos perfis representados na **Figura 7.18**. A interpretação visual do Gráfico M-N sugere a presença de camadas composta por arenito, pela distribuição de pontos ao redor do mineral quartzo. Observa-se também um espalhamento de pontos próximos ao mineral calcita e em direção aos argilominerais.

Tabela 7.5: Descrição geológica do poço 4

Camada	Litologia	Composição (%)	Mineral
Verde	Arenito	Quartzo -100%	
Verde claro	Conglomerados Residuais	Quartzo Feldspático	
Lilás	Diamectito Arenoso	Siliclástica Lamosa	Areno
Roxo	Interlaminado Lamoso Deformado	Intercalações de calcilutita, marga e folhelho	

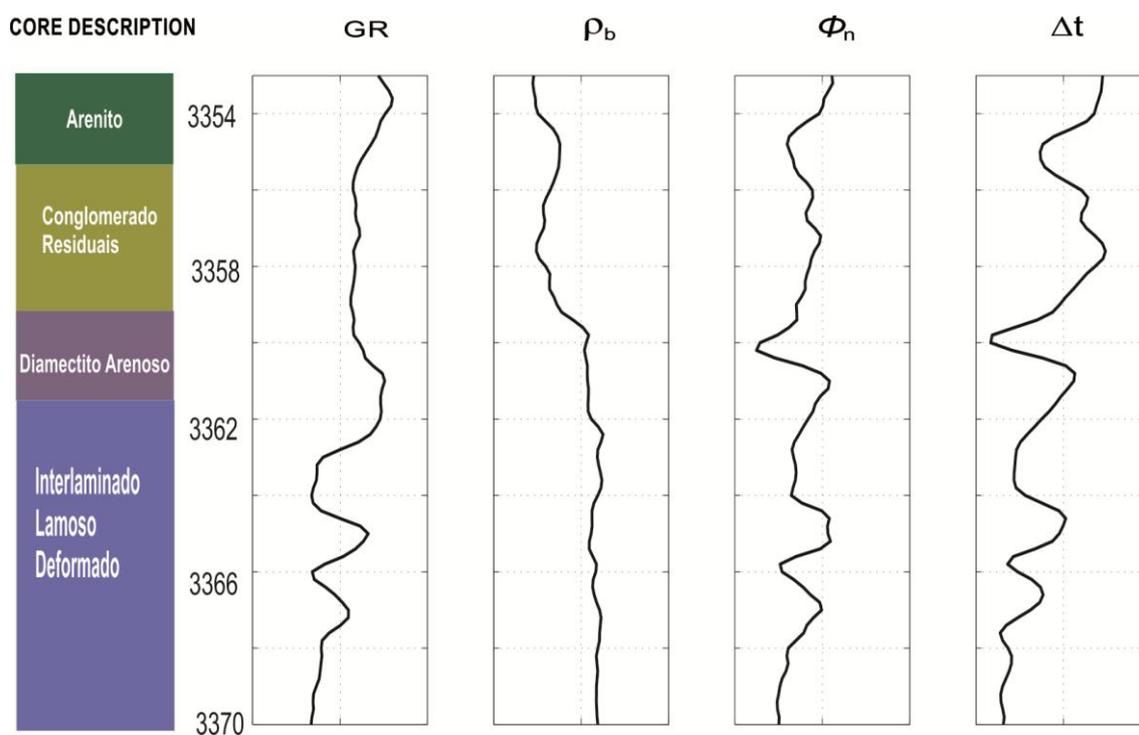


Figura 7.18. Testemunho (à esquerda), perfil de Raio Gama e perfis de porosidade referentes ao intervalo de profundidade selecionado. Fonte: Da autora

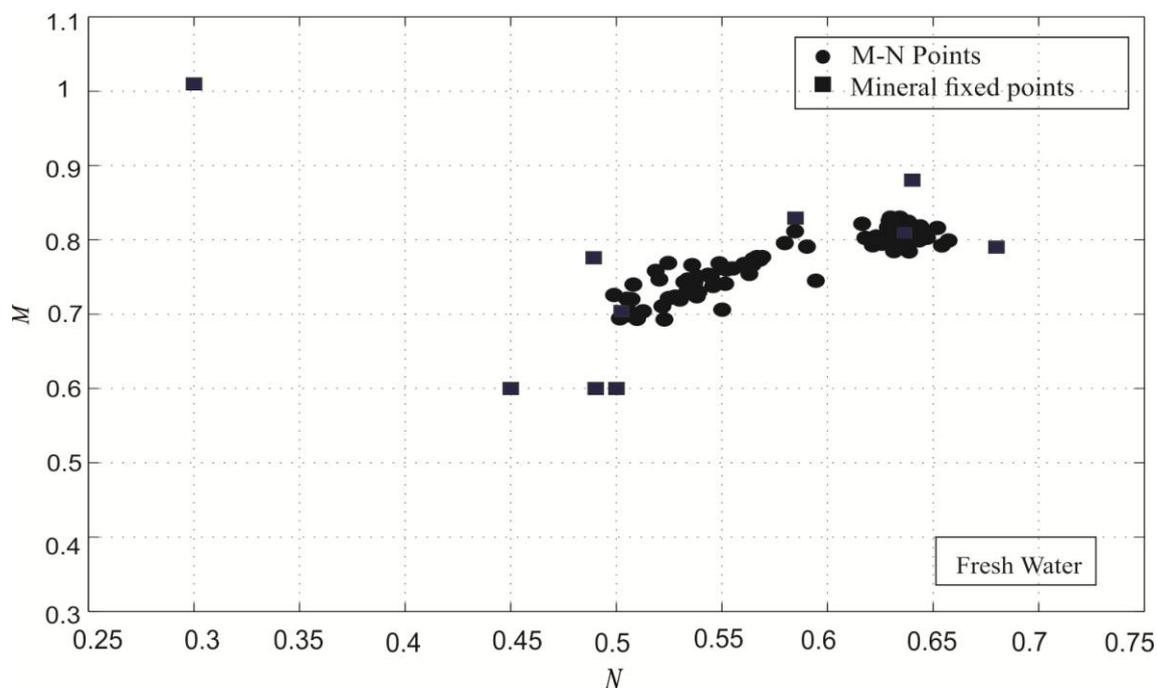


Figura 7.19. Gráfico M-N obtidos com os dados do perfil. Fonte: Da autora

Na **Figura 7.20 (a)** estão os resultados obtidos pelo algoritmo AP sem a atualização do vetor de preferências, definido usualmente como média da matriz de similaridades, o que resulta em um expressivo número de agrupamentos. Na **Figura 7.20(b)** podemos visualizar os agrupamentos resultantes da mudança no vetor de preferência.

A classificação final está disposta na **Figura 7.21**. Sendo a metodologia baseada na classificação por meio de composição mineral, o resultado da classificação está relacionado aos minerais que compõem, predominantemente, as camadas descritas nos testemunhos. O primeiro intervalo de profundidade, referente a camadas de arenito, conglomerados residuais estão associados ao mineral quartzo. Seguindo a sequência em profundidade está a camada referente ao diamectito arenoso composto por matriz siliclástica areno lamosa (Quartzo, calcilutita e folhelho), o que explica o zoneamento em profundidade relacionados à calcita e argilominerais. A camada referente às últimas profundidades é um interlaminado areno lamoso cuja matriz é composta de intercalações entre calcilutita, marga e folhelho, razão

pela qual a classificação para esse intervalo são intercalações entre profundidades cujos minerais que estão compondo são argilominerais e calcita.

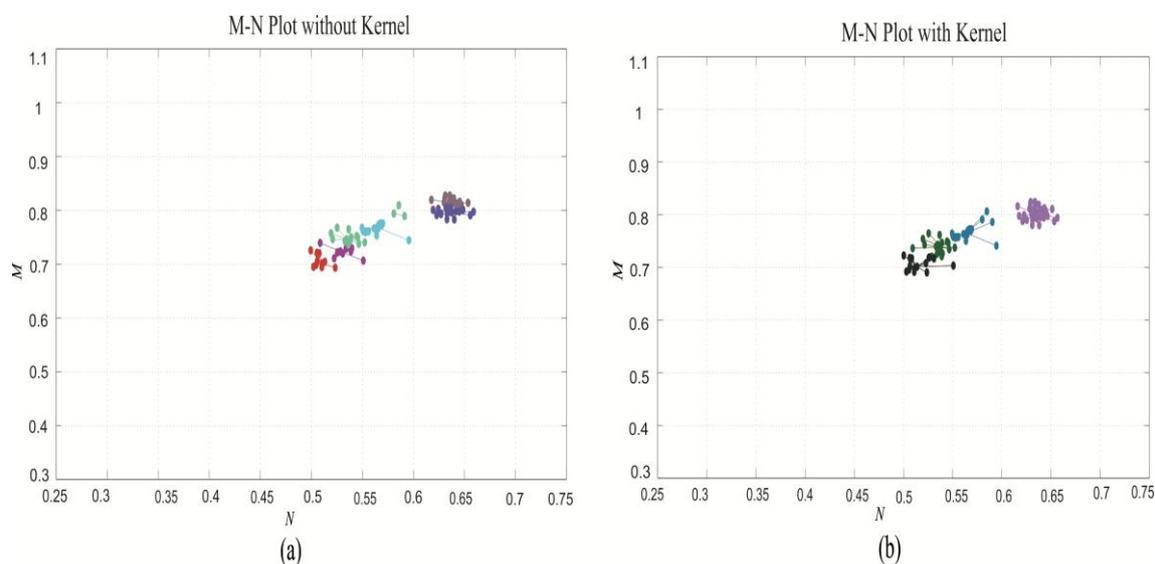


Figura 7.20. (a) Resultado de sete cluster resultantes do AP. (b) Resultados com AP com modificação do vetor P. Fonte: Da autora

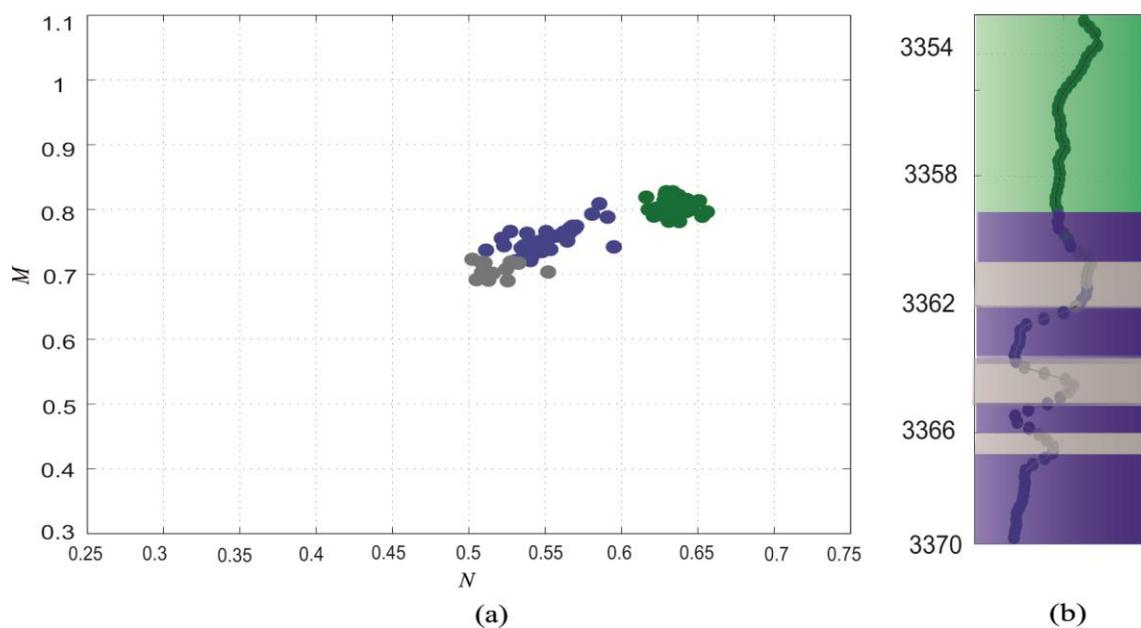


Figura 7.21. (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade. Fonte: Da autora

7.2.5 Dado Real 5

Figura 7.22 apresenta dados de perfis registrados na Bacia de Campos. As amostras de testemunhos estão descritas na *Tabela 7.6* para os intervalos de profundidade selecionados para processamento (curvas em preto). **Figura 7.23** apresenta os pontos (N, M) no gráfico M-N dos perfis representados na **Figura 7.22**. A interpretação visual do Gráfico M-N sugere a presença de camadas composta por quartzo, calcita e argilominerais, pela distribuição de pontos ao redor nas vizinhanças destes minerais.

Tabela 7.6 Descrição geológica do poço 5.

Camada	Litologia	Composição (%)	Mineral
Azul	Calcilutita/Marga	CaCo ₃ /Argilominerais	
Verde	Arenito	Quartzo	
Lilás	Diamectito Arenoso	Siliclástica Lamosa	Areno

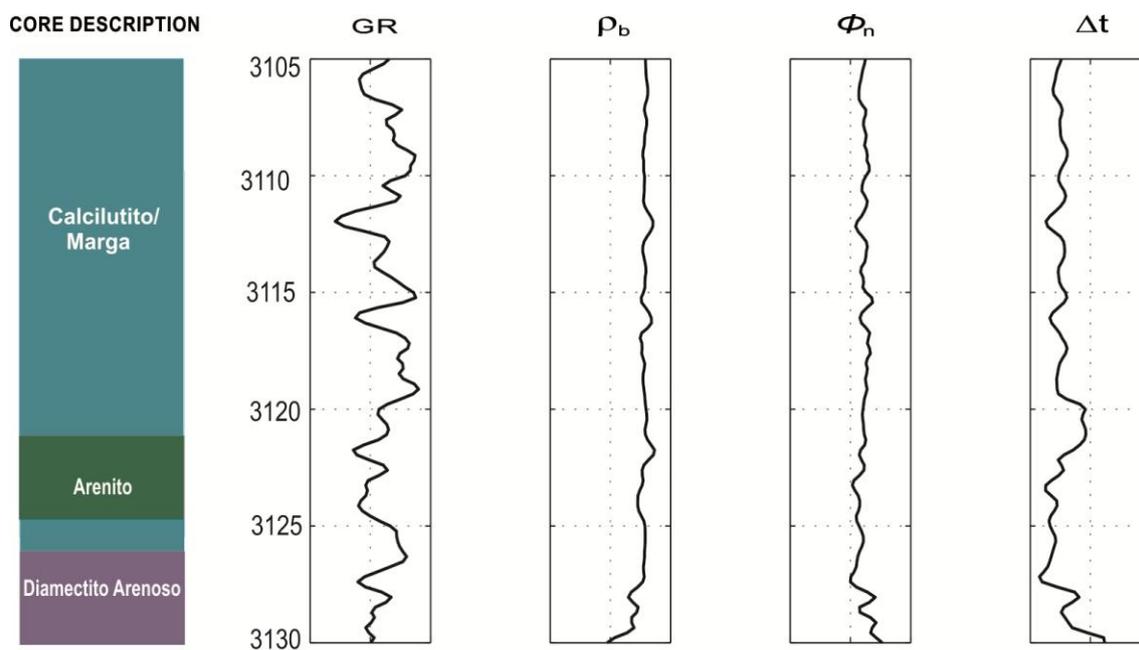


Figura 7.22. Testemunho (à esquerda), perfil de Raio Gama e perfis de porosidade referentes ao intervalo de profundidade selecionado. Fonte: Da autora

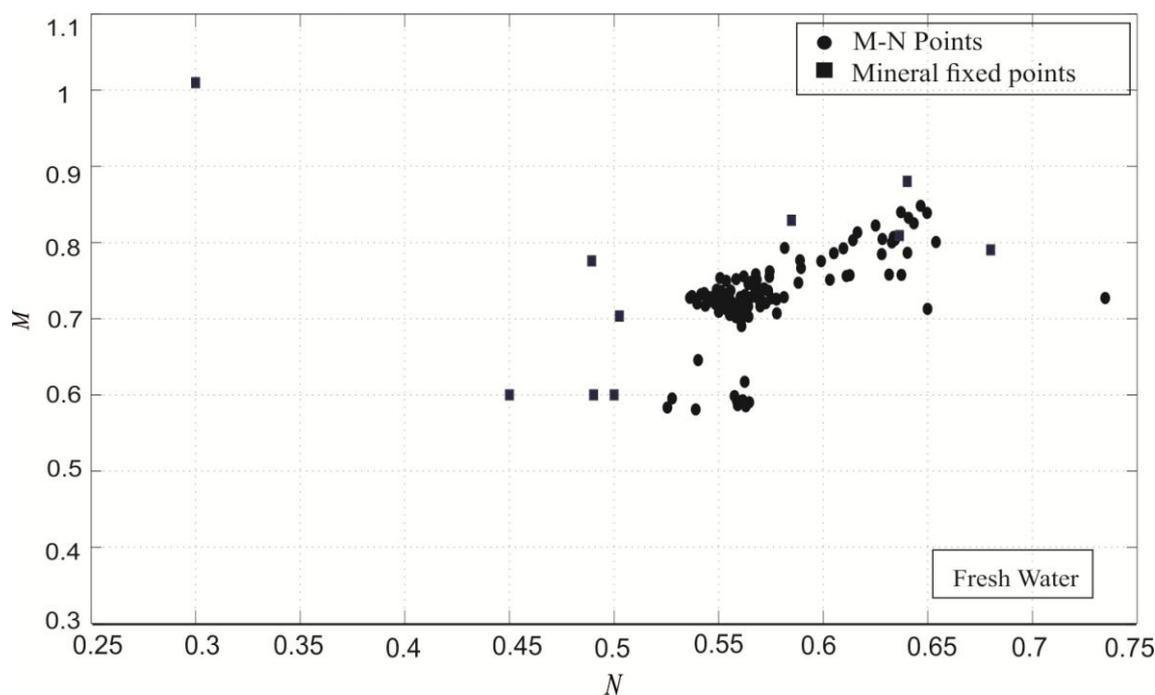


Figura 7.23. Gráfico M-N obtidos com os dados do perfil. Fonte: Da autora

Na **Figura 7.24 (a)** estão os resultados obtidos pelo algoritmo AP sem a atualização do vetor de preferências, definido usualmente como a média da matriz de similaridades, o que resulta em um número de 10 agrupamentos. Na **Figura 7.24 (b)** podemos visualizar os agrupamentos resultantes da mudança no vetor de preferência.

A classificação final está disposta na **Figura 7.25(a)**. Essa classificação é feita pela relação dos exemplares obtidos e representados na **Figura 7.25(b)** com os minerais principais (pontos fixos) do Gráfico M-N. A **Figura 7.25(b)** apresenta a disposição em profundidade dos pontos classificados e representados na **Figura 7.25 (a)**. O primeiro intervalo de profundidade, referente a camadas de calcilutita, marga e folheho, cuja composição da matriz é calcita e argilominerias (como observa-se no zoneamento). Na sequência observa-se o Arenito cuja composição principal foi bem relacionada ao quartzo. A camada de base, referente ao diamectito arenoso composto por matriz siliclástica areno

lamosa (Quartzo, calcilutita e folhelho), e cujos pontos foram relacionados ao mineral Quartzo.

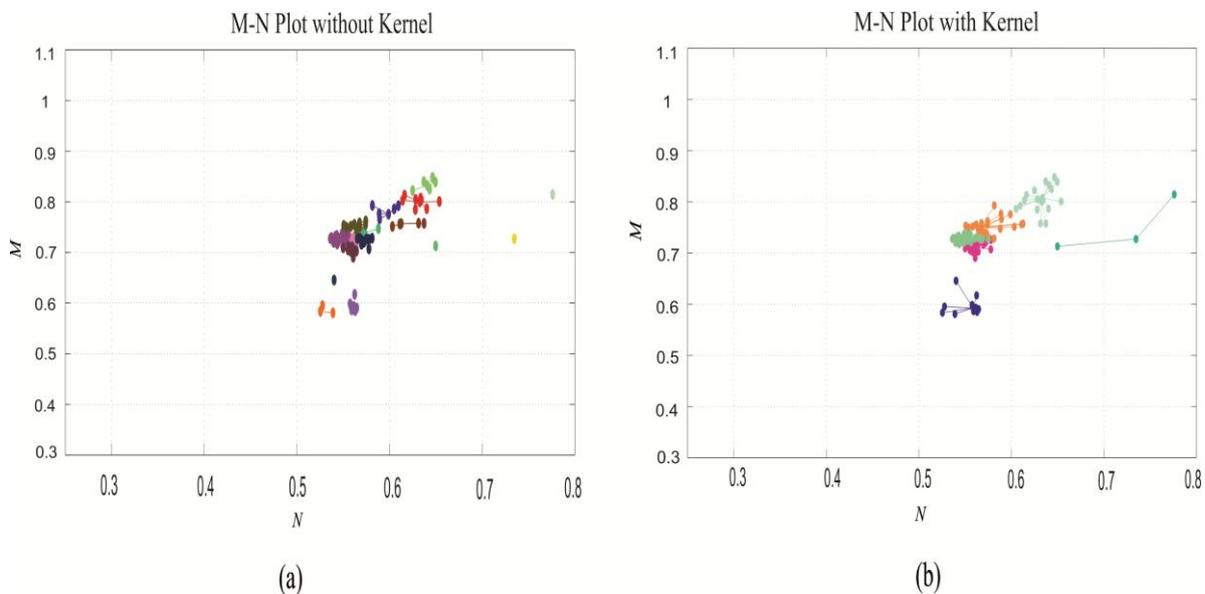


Figura 7.24. (a) Resultado resultantes do AP. (b) Resultados com AP com modificação do vetor P. Fonte: Da autora

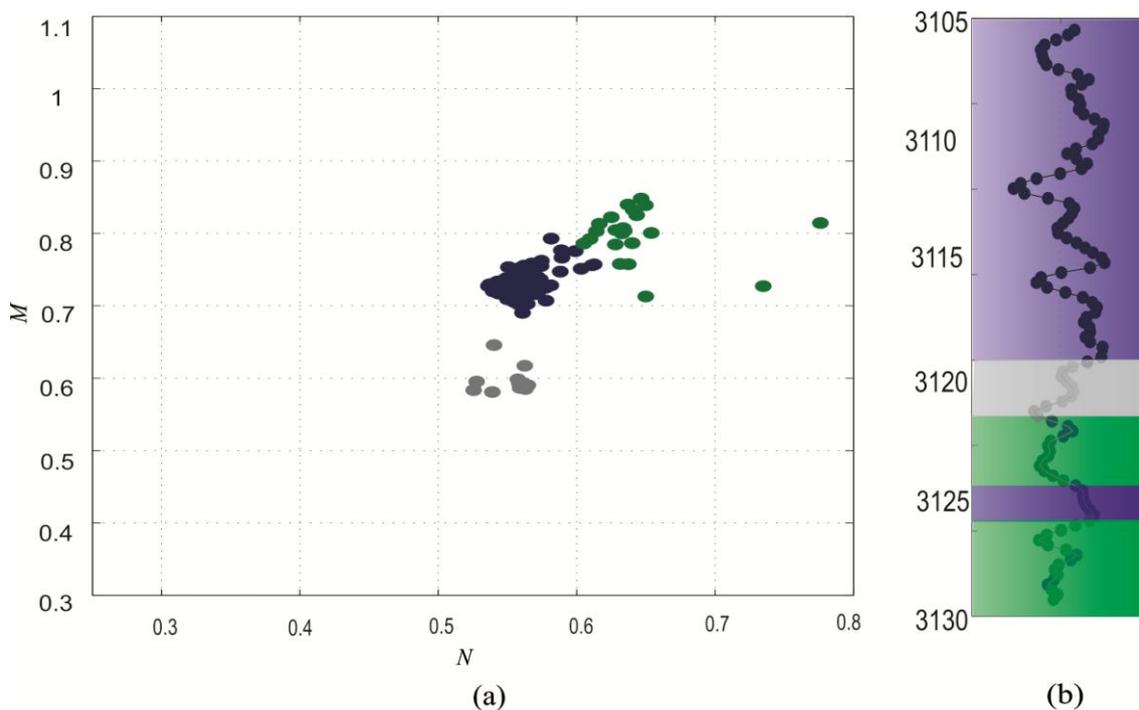


Figura 7.25. (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade. Fonte: Da autora

7.2.6 Dado Real 6

Figura 7.26 apresenta dados de perfis registrados na Bacia de Campos. As amostras de testemunhos estão descritas na *Tabela 7.7* para os intervalos de profundidade selecionados para processamento (curvas em azul). **Figura 7.27** apresenta os pontos (N, M) no gráfico M-N dos perfis representados na **Figura 7.26**. A interpretação visual do Gráfico M-N sugere apenas um espalhamento de pontos na direção de argilominerais e nas proximidades do mineral calcário.

Tabela 7.7: Descrição geológica do poço 6

Camada	Litologia	Composição Mineral
Azul	Arenito/Marga	Quartzo/Argilominerais
Cinza	Folhelho	Argilominerais
Lilás	Diamectito Arenoso	Siliclástica Areno Lamosa
Verde	Siltito estratificado	Argilominerais/quartzo

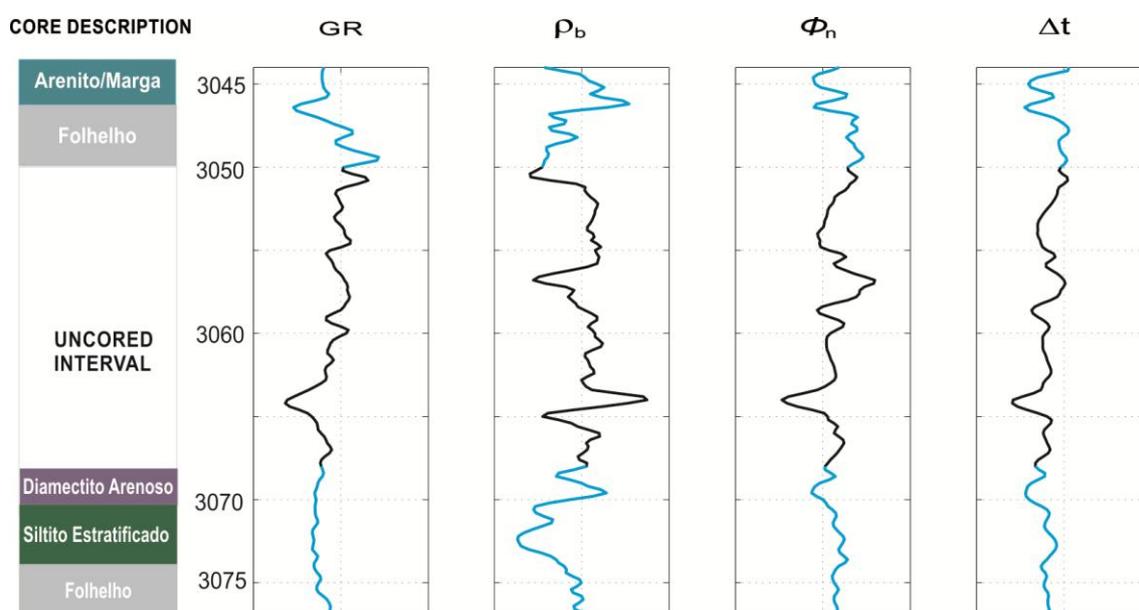


Figura 7.26. Testemunho (à esquerda), perfil de Raio Gama e perfis de porosidade referentes ao intervalo de profundidade selecionado. Fonte: Da autora

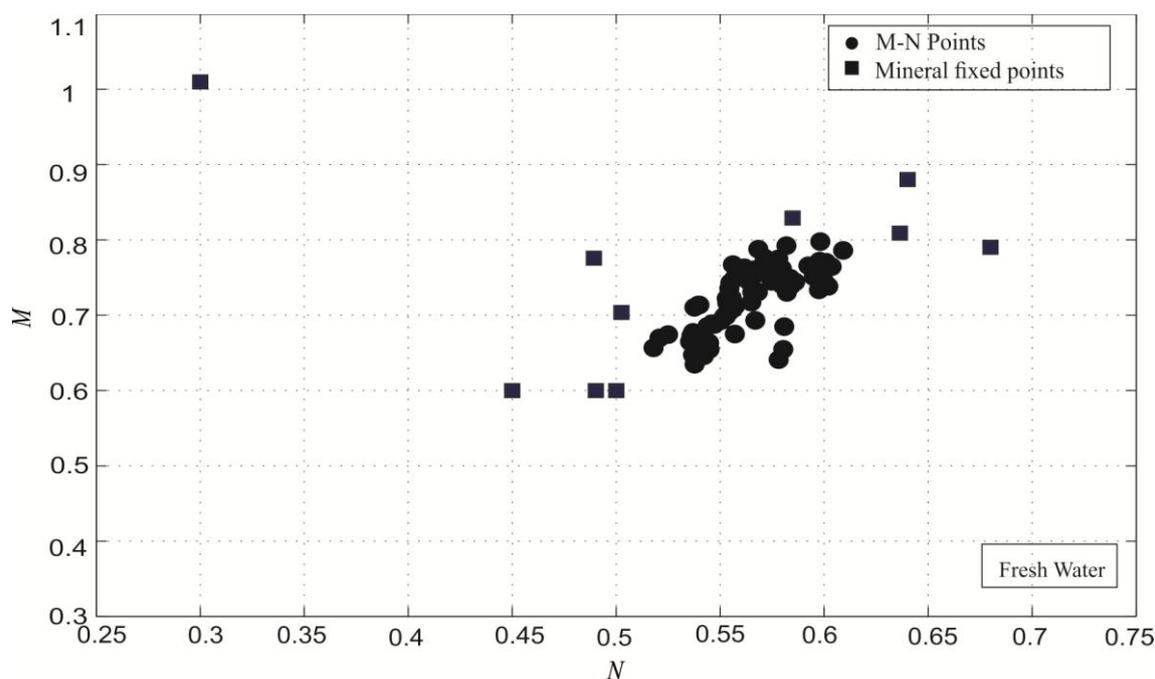


Figura 7.27. Gráfico $M-N$ obtidos com os dados do perfil. Fonte: Da autora

Na **Figura 7.28 (a)** estão os resultados obtidos pelo algoritmo AP sem a atualização do vetor de preferências, definido usualmente como média da matriz de similaridades, o que resulta em um número de 10 agrupamentos. Na **Figura 7.28(b)** podemos visualizar os agrupamentos resultantes da mudança no vetor de preferência.

A classificação final está disposta na **Figura 7.29(a)**. Essa classificação é feita pela relação dos exemplares obtidos e representados na **Figura 7.29(b)** com os minerais principais (pontos fixos) do Gráfico $M-N$. A **Figura 7.29(b)** apresenta a disposição em profundidade dos pontos classificados e representados na **Figura 7.29(a)**. O primeiro intervalo de profundidade, referente a camadas de Arenito, marga e folheho, cuja composição da matriz é quartzo, calcita e argilominerías (como observa-se no zoneamento). Na sequência observa-se o segundo trecho, referente ao diamectito arenoso composto por matriz siliclástica areno lamosa (Quartzo, calcilutita e folhelho), siltito estratificado e folhelho cujos pontos foram relacionados ao mineral Quartzo e argilominerías.

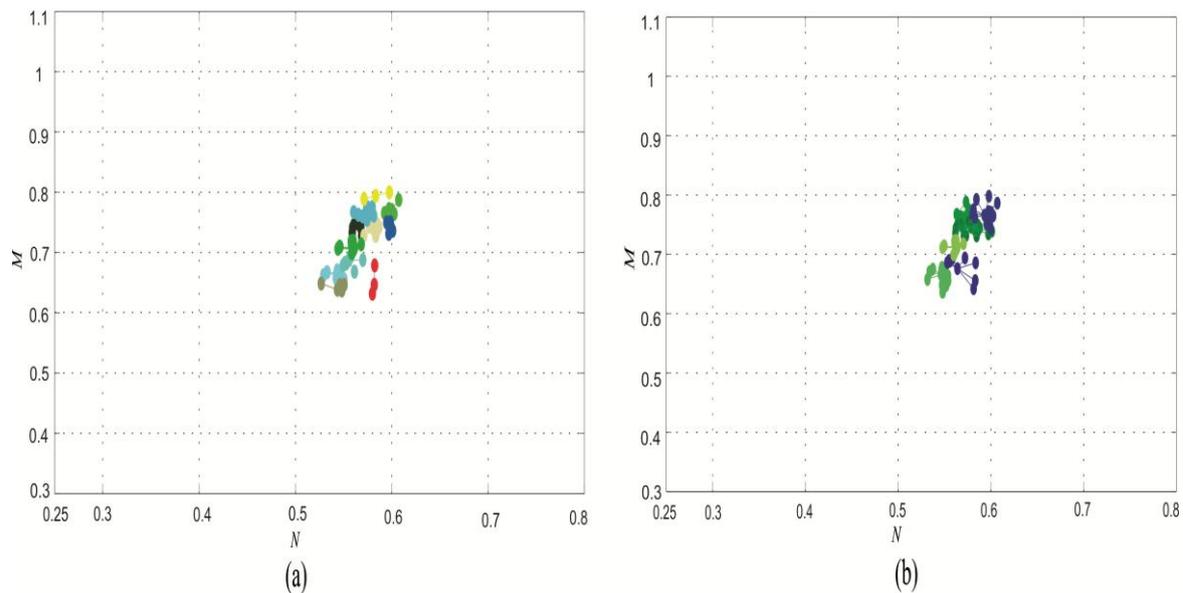


Figura 7.28. (a) Resultado resultantes do AP. (b) Resultados com AP com modificação do vetor P. Fonte: Da autora

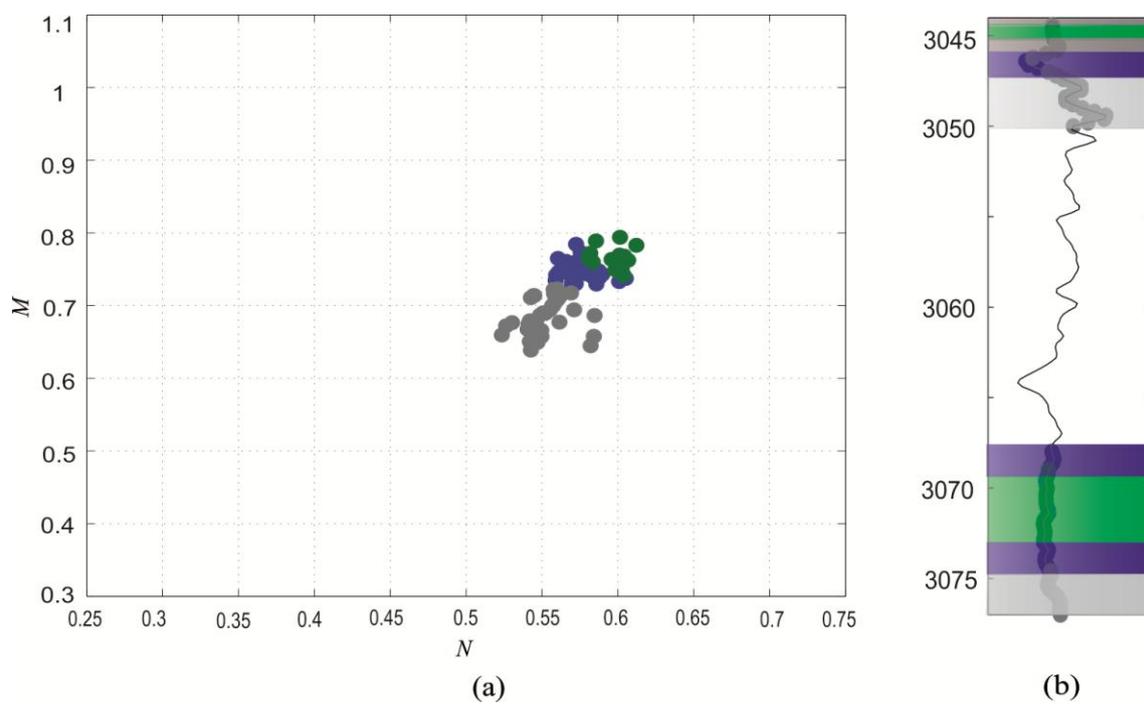


Figura 7.29. (a) Classificação final por associação de exemplar. (b) Mapeamento de Litologias em profundidade. Fonte: Da autora

7.3 AVALIAÇÃO DA CLASSIFICAÇÃO LITOLÓGICA

Os índices de validação têm por objetivo avaliar qualidade de grupos resultantes de um processo de agrupamento. Essa avaliação serve para comparar diferentes algoritmos de agrupamento, partições, e fazer comparação entre dois grupos (clusters).

Os algoritmos de validação podem assumir alguns aspectos, dentre eles: determinar a tendência de agrupamentos de dados (verificar se os dados são ou não aleatórios); Comparar os resultados de uma análise de agrupamento com resultados conhecidos (tendo um conhecimento prévio dos dados, verificar que índices correspondem); Comparar diversos algoritmos de agrupamentos ou determinar o valor mais apropriado do número de grupos (clusters).

O índice Kappa(K) proposto por Cohen (Cohen, 1960) é um método de avaliação entre a concordância entre dois conjuntos de dados. Os valores de K variam em uma escala de 0 – 1 e para sua estimativa é necessário a determinação de uma matriz de erro (matriz de confusão) a qual apresenta os resultados dados classificados pela metodologia em relação a classificação real, obtida pelo testemunho. Como exemplo de matriz de erro temos a matriz A na *Tabela 9*, para um número hipotético de classes n. A diagonal da matriz representa a frequência referente à concordância entre o dado classificado e o dado real, ou número de dados classificados corretamente. A coluna $\sum_{i=1}^n A_{i.}$ representa o número total de elemento para cada classe i. A linha $\sum_{i=1}^n A_{.i}$ representa o número de dados atribuídos a cada classe i.

Tabela 7.8: Matriz de erro.

Core	Classificação				
	1	2	n	$\sum_{i=1}^n A_{i.}$
1	A_{11}				
2		A_{22}			
:					
:					
n				$A_{.n}$	
$\sum_{i=1}^n A_{i.}$					$\sum_{i=1}^n A_{ii}$

O cálculo do índice (k) é feito com base na matriz de erros pela equação (7.1)

$$k = \frac{P_0 - P_e}{1 - P_e} \quad (7.1)$$

Onde

$$\begin{cases} P_0 = \frac{\sum_{i=1}^n A_{ii}}{N} \\ P_e = \frac{\sum_{i=1}^n A_{i.} A_{.i}}{N^2} \end{cases} \quad (7.2)$$

N é o número de amostras, P_0 é a proporção de pontos classificados corretamente pelo total de pontos, enquanto P_e é a proporção de pontos atribuídos a cada classe ao acaso (Cohen, 1960; Antunes, 2008).

Para avaliação do Índice Kappa usou-se a *Tabela 7.9*, adaptada de Galparsoro e Fernández (2001)(Galparsoro et al.,2001; ,Leão et al., 2007).

Tabela 7.9: Avaliação do grau de concordância usando K

Índice (K)	Concordância
<0,2	Pobre
0,21-0,4	Fraca
0,41-0,6	Moderada
0,61-0,8	Boa
0,81-1	Muito Boa

Fonte: Adaptado de Galparsoro e Fernández

A *Tabela 7.10* apresenta os índices Kappa calculados para os dados reais processados. Segundo a avaliação do índice Kappa para os dados reais observa-se que as classificações resultantes da aplicação da metodologia obtiveram bons resultados. Para o dado real 1 o valor do índice K foi 1, ou seja, houve uma total concordância entre a classificação e o dado obtido no testemunho. Para os demais dados, os valores de concordância foram menores devido alguns pontos não classificados corretamente (outliers) e, ainda, algumas camadas cuja constituição é multimineral.

Tabela 7.10: Índice kappa para avaliação das classificações

Dado	Índice (K)
Real 1	1
Real 2	0.8750
Real 3	0.8816
Real 4	0.7642
Real 5	0.8156
Real 6	0.6876

Fonte: Da autora

8 CONCLUSÕES

A identificação de litologias, no sentido estrito da geofísica de poço, é um problema bastante complexo de ser solucionado exclusivamente com os perfis geofísicos, uma vez que não é completamente válida a premissa de correspondência entre variação de propriedades físicas e litologias.

Uma interpretação quantitativa dos perfis é proporcionada pelo Gráfico M-N que, em princípio, poderia proporcionar uma rápida interpretação visual dos pontos ao longo das profundidades do poço. No entanto, a presença de ruído nos perfis geofísicos e o modelo de rocha simplificado que desconsidera a argilosidade, torna a interpretação visual do Gráfico M-N complexa e ambígua. Uma solução para este tipo de problema de classificação foi apresentada aqui através de uma metodologia não supervisionada que, independente da disponibilidade ou não dos testemunhos, é capaz de realizar a identificação e a distinção litológica das rochas atravessadas por um poço.

A representação de um cluster por seu exemplar, produzida pelo algoritmo Affinity Propagation na interpretação do Gráfico M-N resulta em uma redução significativa do conjunto de dados de conteúdo a ser processado, o que atenua a ambiguidade relativa a dispersão do conjunto de dados no Gráfico M-N devido ao ruído nos dados de poço e à argilosidade.

As vantagens obtidas por meio do algoritmo Affinity Propagation é o fato de que não é necessário determinar a priori o número de agrupamentos, como um parâmetro de entrada, um requisito necessário na maioria dos métodos clássicos de agrupamento. Este requisito obriga existência de algum conhecimento sobre dados a serem processados, o que, na maioria dos casos, não é possível. Outra vantagem de utilizar o AP em comparação com outro algoritmo de classificação é a substituição do centróide pelo exemplar, em que os centróides são representantes de um cluster no gráfico, mas não são necessariamente pontos dos dados processados. Os exemplares, no entanto, são representantes dos clusters e ainda um dado relevante.

A utilização da função de densidade Kernel sobre os dados N-M, combinado com o algoritmo de otimização Firefly foi crucial para a detecção de padrões de grupo e determinação de bons candidatos para exemplares. Estes candidatos foram utilizados, pelo AP, como uma informação prévia para modificar o vetor de preferência, melhorando o processo de busca e redução do número de clusters. Uma vez que o exemplar de cada grupo é relacionado a um ponto fixo N-M, pela mínima distância, estas melhorias apresentadas

pela modificação do vetor de preferência, resultou numa melhor caracterização e descrição dos dados, principalmente, com uma classificação litológica mais eficaz, mesmo em dados M-N com alto grau de espalhamento. Em geral, a metodologia híbrida obteve resultados satisfatórios, o que pode ser constatado pela avaliação do índice kappa (Tabela 10), cuja concordância entre a classificação e a descrição do testemunho variou entre Boa e muito Boa, e pode ser utilizada para a interpretação de dados de perfis de poços com o objetivo de definir os tipos de litologias atravessadas por poço.

Em trabalhos posteriores pode-se melhorar esta pesquisa para classificação de litologias em casos de camadas extremamente finas e cuja composição mineral seja mais complexa, ou seja, a composição seja multimineral e ainda ampliar o estudo para a classificação de fácies, termo que abrange diversas características além da composição mineral.

REFERÊNCIAS

- Asquith, G.B., C.R. Gibson, 1982, Basic well log analysis for geologists, American Association of Petroleum Geologists.
- Antunes, A.F.B., Lingnau, C, 2008, Uso de Índices de Acurácia para Avaliação de Mapas Temáticos Obtidos por meio de Classificação Digital. UFPR.
- Bacocoli, G., Morales, R.G., Campos, O.A., 1980, The Namorado oil field: a major oil discovery in the Campos basin, Brazil, AAPG Bulletin, 63 /329-338.
- Barros, C.A., 2008, Determination of water saturation by intelligent algorithm, in SBGf (Ed.) SIMPOSIO BRASILEIRO DE GEOFÍSICA, SBGf, Belém, Para, Brazil.
- Burke, J.A.C.J., Schimidt, A.W., 1969, The Litho-porosity cross plot: a method of determining rock characteristics for computation of log data., in: Society of Petroleum Engineers (Ed.) SPE ILLINOIS BASIN REGIONAL MEETING, Society of Petroleum Engineers, Illinois Basin.Evansville, Indiana, p. 187- 198.
- Busch, J., Fortney, W., Berry, L., 1987, Determination of Lithology From Well Logs by Statistical Analysis, SPE Formation Evaluation.
- Clavier, C., D. Rust, 1976, MID-plot: a new lithology technique, The Log Analyst, 16-24.
- Crain, E.R. , 1986, The Log Analysis Handbook: quantitative log analysis methods, Tulsa: Pennwell Books, 1 v.
- Cohen, J., 1960, A coeficient of Agreement for Nominal Scales. Educational and Measurement. Vol XX, Nº 1, p. 37-46.
- Delfiner, P., O. Peyret, O. Serra, 1987, Automatic determination of lithology from well logs, SPE Formation Evaluation, 2, 303-310.
- Ellis, D.V., Singer, J.M., 2007, Well Logging for Earth Scientists, Springer.

Floudas, C.A., Pardalos, P.M., 2008, Encyclopedia of optimization, Springer.

Frey, B.J., Dueck, D., 2007, Clustering by passing messages between data points, Science, 315 / 972-976.

Frey, B.J., 2007, Affnity Propagation | Frey Lab.

Frey, B.J., Kschischang, F.R., Loeliger, H.A., Wiberg, N., 1997, Factor graphs and algorithms, in Proceedings of the Annual Allerton Conference on Communication Control and Computing, UNIVERSITY OF ILLINOIS, p. 666-680.

Galparsoro, L.U., Fernández, S.P, 2001, Medidas de concordância: el índice Kapp.

Gan, G.; Ma, C. ; Wu, J., 2007, Data clustering: theory, algorithms, and applications, Society for Industrial and Applied Mathematics.

Leão, C., Krug, L.A., Kampel, M., Fonseca, L.M.G., 2007, Avaliação de métodos de classificação em imagens TM/Landsat e CCD/CBERS para o mapeamento do uso e cobertura da terra na região costeira do extremo sul da Bahia. XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, 21-26 abril 2007, INPE, p. 939-946.

Limited, S., 1974, Log interpretation, Schlumberger.

LINDEN, R., 2009, Técnicas de Agrupamento. Revista de Sistema de Informação da FSMA, [s.l.], 3, 18-36.

PUMMA-VILLANUEVA, W.J., ZUBEN, F.J.V, 2008, Índices de validação de Agrupamentos. Unicamp.

Rai, P., Singh, S., 2010, A survey of clustering techniques, International Journal of Computer Applications, 7, 1-5.

Rangel, H., Martins, F., Esteves, F., Feijó, F., 1994, Bacia de campos, Boletim de Geociências da PETROBRAS, 8, 203-217.

- Savre, W.C., Burke, J., 1963, Determination of true porosity and mineral composition in complex lithologies with the use of Sonic, Neutron and Density Survey, *The Log Analyst*, 4, 439-474.
- Savre, W., 1963, Determination of a more accurate porosity and mineral composition in complex lithologies with the use of the sonic, neutron and density surveys, *Journal of Petroleum Technology*, 15, 945-959.
- Silverman, B.W., 1981, Using kernel density estimates to investigate multimodality, *Journal of the Royal Statistical Society. Series B (Methodological)*, 97-99.
- Spath, H., 1980, Cluster analysis algorithms for data reduction and classification of objects, Ellis Horwood, Ltd.
- Steinbach, M., Karypis, G., e Kumar, V., 2000, A comparison of document clustering techniques.
- Wang, W.-J., Tan, Y.-X., Jiang, J.-H., J.-Z. Lu, G.-L. Shen, R.-Q. Yu, 2004, Clustering based on kernel density estimation: nearest local maximum searching algorithm, *Chemometrics and intelligent laboratory systems*, 72, 1-8.
- Yang, X.S., 2010, Firefly algorithm, stochastic test functions and design optimisation, *International Journal of Bio-Inspired Computation*, 2, 78-84.
- Yang, X.S., He, X., 2013, Firefly algorithm: recent advances and applications, *International Journal of Swarm Intelligence*, 1, 36-50.
- Yang, X.S., 2010, Firefly Algorithm, in, Mathworks, MatlabCentral.
- Turlach, B.A., 1993, Bandwidth selection in kernel density estimation: A review, CORE and Institut de Statistique, 19 /1-33.