



Um fluxo de trabalho para modelagem de eletrofácies com entrelaçamento de técnicas de classificação supervisionada e não-supervisionada.

José Agnelo Soares (agnelo@geologia.ufrj.br) - Departamento de Geologia da Universidade Federal do Rio de Janeiro.

Copyright 2005, SBGF - Sociedade Brasileira de Geofísica

This paper was prepared for presentation at the 9th International Congress of the Brazilian Geophysical Society held in Salvador, Brazil, 11-14 September 2005.

Contents of this paper were reviewed by the Technical Committee of the 9th International Congress of the Brazilian Geophysical Society. Ideas and concepts of the text are authors' responsibility and do not necessarily represent any position of the SBGF, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of the Brazilian Geophysical Society is prohibited.

Abstract

This paper presents a workflow to build logfacies models which is based on the interlaced application of non-supervised and supervised statistical techniques. The workflow allows answering, in an interactive manner, some basic issues which appear in any logfacies modeling program: How many facies may be recognized? Which is the best log suite to be used in? How the generated logfacies correlate with the original lithofacies? Due to the interactive process between supervised and non-supervised techniques, the proposed workflow is biased to produce a fewer number of logfacies than the number of original lithofacies, but those logfacies tend to correspond to, in a physical properties sense, those most representative lithofacies. A main feature of this workflow is the adoption of the logfacies column provided by a non-supervised classification algorithm as the input column of facies for a supervised classification procedure. This feature warrants that the recognized logfacies suite is the best possible, in terms of log physical properties and lithofacies reproduction. Finally, it is shown a real case where this workflow was applied successfully.

Introdução

A modelagem de eletrofácies pode ser definida como a tentativa de reconhecimento da coluna faciológica de um poço a partir dos perfis geofísicos desse poço. Trata-se de uma importante atividade que permite a construção do modelo geológico da área, mesmo na ausência de testemunhos contínuos na maioria dos poços. Basicamente três famílias de técnicas têm sido utilizadas com o objetivo de reconhecer eletrofácies: as técnicas estatísticas, as redes neurais, a regra *fuzzy*, ou ainda combinações destas.

Este trabalho trata apenas da aplicação de um conjunto de técnicas estatísticas, regularmente usadas para a classificação supervisionada (Tatsuoka, 1971) ou não-supervisionada (Massart & Kaufman, 1983), com o objetivo de reconhecimento das fácies geológicas atravessadas por poços a partir dos perfis geofísicos. Na modelagem de eletrofácies, classificação supervisionada é aquela feita com base na geração de uma regra de reconhecimento de fácies obtida em um intervalo de poço para o qual já se conhece a descrição faciológica de testemunho (litofácies). Por outro lado, a classificação

não-supervisionada é aquela realizada puramente com base na separação dos grupos de amostras com padrões semelhantes em termos das variáveis classificatórias (no caso, os perfis geofísicos dos poços).

Entre os métodos de classificação não-supervisionada estão *Average Linkage*, *Centróide* e *K-means*. Os dois primeiros são baseados na distância entre amostras no espaço n -dimensional, onde n é o número de perfis utilizado. O *Average Linkage* constrói os grupos de acordo com as menores distâncias entre todos os pares possíveis de amostras da população. Esta característica o torna sensível a possíveis valores anômalos. O método *Centróide*, por sua vez, define os grupos com base na distância (no espaço n -dimensional) entre os centros, ou valores médios, de cada nuvem de amostras, o que o torna um método mais robusto à presença de valores anômalos. Já o método *K-means*, que somente pode ser usado quando já se conhece o número K de grupos (ou eletrofácies) que se deseja modelar, age da seguinte maneira: inicialmente, gera K grupos formados aleatoriamente pelas amostras dos perfis, a seguir esse método reorganiza a composição de cada grupo de modo a minimizar a variabilidade interna de cada grupo e maximizar a variabilidade entre grupos.

Pode-se associar, a alguns métodos de classificação não-supervisionada, algumas estatísticas que permitem, com base nos dados disponíveis, estimar o número ideal de fácies a reconhecer. Estas estatísticas são as funções *Pseudo-F*, *Pseudo-t²* e *Cubic Clustering Criterion (CCC)*. Além disso, pode-se avaliar a composição de cada grupo através de um diagrama em árvore.

Os métodos de classificação supervisionada incluem os chamados métodos paramétricos, que são aqueles para os quais se assume que as distribuições das variáveis obedecem a uma função distribuição conhecida (como funções multinormais, por exemplo), e os métodos não-paramétricos, para os quais essa restrição não se aplica. Os métodos paramétricos incluem a Regra Discriminante Linear (RDL) e a Regra Discriminante Quadrática (RDQ). Já os métodos não-paramétricos incluem, entre outros, os métodos da regra discriminante passo-a-passo, *K-nearest neighbor (KNN)*, e regra discriminante canônica. A regra discriminante linear passo-a-passo permite ordenar os vários perfis disponíveis segundo uma ordem decrescente de poder discriminante. Essa característica torna esse método uma ferramenta muito adequada para utilização na fase inicial do trabalho de modelagem de eletrofácies, quando da escolha dos perfis a serem usados no trabalho. O método KNN se baseia na adoção do rótulo que aparece mais vezes entre os K vizinhos mais próximos, no espaço n -dimensional, segundo a métrica adotada. Já o método da regra canônica consiste num método de redução de dimensão, onde se procura obter algumas variáveis canônicas, pela combinação

linear das variáveis originais, mas não correlacionadas entre si, que juntas, maximizem o poder de discriminação entre fácies.

O fluxo de trabalho

Um fluxograma da estratégia proposta para a modelagem de eletrofácies é apresentado na Figura 1. A primeira operação é constituída pela edição dos dados, que corresponde à preparação do arquivo em colunas, retirada das amostras com dados faltantes, e incorporação da coluna de litofácies, quando disponível.

A seguir os dados passam pela etapa de transformação e padronização das variáveis. Alguns perfis, como é caso dos perfis de resistividade elétrica, variam os valores registrados muito rapidamente, o que normalmente exige uma escala logarítmica para a sua apresentação. Nesses casos recomenda-se aplicar alguma função de transformação para essas variáveis, de modo que as variáveis transformadas tenham um comportamento aproximadamente linear. Para os perfis de resistividade, pode-se aplicar a seguinte transformação:

$$\Omega' = \log_{10}(\Omega + 1) \quad (1)$$

onde Ω é valor original do perfil de resistividade e Ω' é o valor transformado de resistividade.

A padronização é recomendada para todas as variáveis, uma vez que ela deixa todos os perfis com média nula e desvio padrão unitário, fazendo com que todas as variáveis estejam numa mesma escala de valores. O processo de padronização ajuda, inclusive, na interpretação dos coeficientes da regra discriminante, identificando rápida e claramente o poder discriminante relativo de cada perfil. A padronização é dada por

$$x_i' = \frac{X_i - \bar{X}}{\sigma} \quad (2)$$

onde X_i é valor da i -ésima amostra do perfil, \bar{X} é o valor médio do perfil, e σ é o desvio padrão do perfil.

Conforme a Figura 1, os dados devem ser separados em três conjuntos: um conjunto para treinamento da regra, outro para validação, e um terceiro para predição, ou seja, para o reconhecimento propriamente dito das fácies. A proporção de dados em cada um desses conjuntos pode ser bastante variável, pois depende em muito do quão complexa é a geologia local, da suíte disponível de perfis geofísicos, e da quantidade e representatividade das fácies a reconhecer. De modo geral, costuma-se usar cerca de 30% dos dados para treinamento, 10% a 20% para validação e o restante para a fase de predição. É importante ressaltar que, para a estratégia de classificação aqui apresentada, os dados de treinamento e de validação incluem a descrição de litofácies. Esta exigência não se aplica aos dados de predição.

Com o conjunto de dados de treinamento, procede-se à fase de geração da regra discriminante. É nesta fase que se recomenda que sejam usadas, de forma entrelaçada, as técnicas de classificação supervisionada e não-

supervisionada. Essa estratégia permite responder a algumas questões fundamentais como:

1. Para reconhecer as fácies que desejo, dentre os perfis que disponho, quais deles devo usar? Nem sempre usar todos os perfis resulta no melhor reconhecimento de fácies, pois às vezes adicionar mais uma curva não significa agregar informação útil, mas apenas aumentar o ruído do sistema.
2. Qual o número de fácies que meus perfis podem reconhecer? Em geral, os perfis reconhecem um número de fácies muito menor que aquele reconhecido pelo geólogo responsável pela descrição dos testemunhos.
3. E finalmente, qual a relação entre as eletrofácies reconhecidas e as litofácies descritas pelo geólogo?

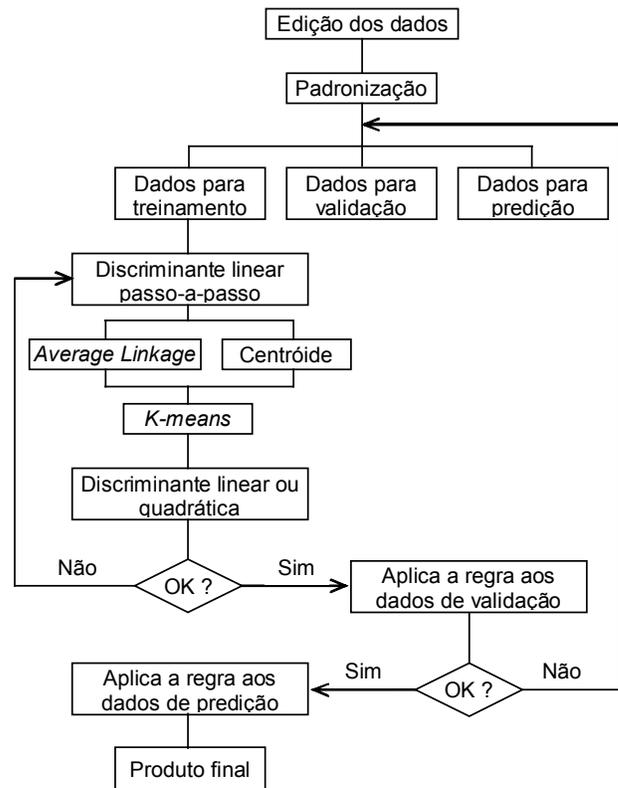


Figura 1 – Fluxograma para modelagem de eletrofácies através do uso entrelaçado de técnicas de classificação supervisionada e não-supervisionada.

Com o objetivo de determinar a melhor suíte de perfis, dentre aqueles disponíveis, aplica-se, inicialmente, a regra discriminante linear passo-a-passo nos dados de treinamento. Essa técnica de classificação supervisionada enumera, em ordem decrescente de prioridade, os perfis com maior poder discriminante e, adicionalmente, exclui do modelo aqueles perfis que contribuíam de forma negativa para o reconhecimento das fácies pretendidas. Nesta etapa as fácies a reconhecer ainda são as litofácies originais, descritas diretamente da observação dos testemunhos.

A seguir, com o objetivo de definir o número ideal de fácies a reconhecer, procede-se à aplicação, em paralelo, dos procedimentos de classificação não-supervisionada *Average Linkage* e Centróide sobre o conjunto de dados de treinamento, desta vez desprezando o conhecimento prévio das litofácies. Com o suporte dos gráficos das estatísticas *Pseudo-F*, *Pseudo-t²* e CCC, define-se o número ideal de fácies a reconhecer. As figuras 2, 3 e 4 apresentam exemplos de saídas gráficas para essas estatísticas.

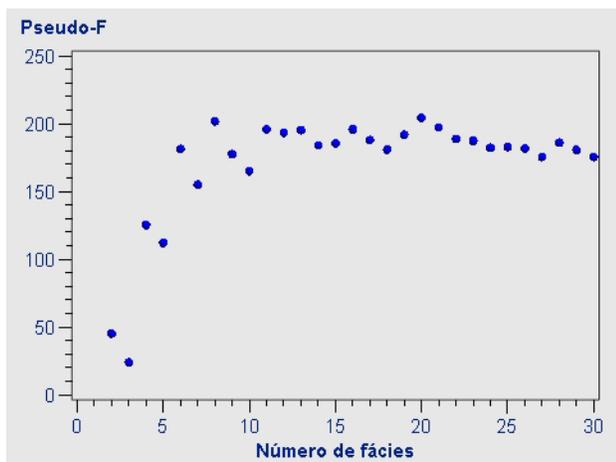


Figura 2 – Estatística *Pseudo-F* contra o número de fácies, segundo a técnica de classificação não-supervisionada *Average Linkage*.

O gráfico da estatística *Pseudo-F* indica bons números de grupos quando essa estatística apresenta valores elevados. Assim, para o caso da Figura 2, vemos que 4 ou 6 fácies seriam, de acordo com a estatística *Pseudo-F*, bons números de fácies para se tentar identificar. Quantidades muito altas (como 10, por exemplo) ou muito baixas de fácies (como 2, por exemplo) não são desejáveis para a modelagem de eletrofácies.

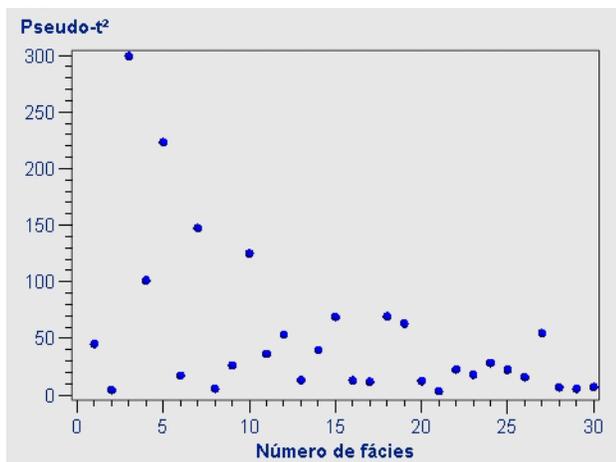


Figura 3 – Estatística *Pseudo-t²* contra o número de fácies, segundo a técnica de classificação não-supervisionada *Average Linkage*.

O gráfico da estatística *Pseudo-t²* apresenta valores altos para os números de fácies imediatamente anteriores aos ideais. Para o caso da Figura 3, cujos dados processados

são os mesmos utilizados na Figura 2, os bons números de fácies seriam, novamente, 4 ou 6.

No caso da estatística CCC, valores positivos maiores que 2 ou 3 são indicativos de bons números de fácies, valores entre 0 e 2 indicam valores potenciais de número de fácies, enquanto que valores negativos elevados são indicativos da presença de valores anômalos. No caso dos dados da Figura 4, que são os mesmos das figuras 2 e 3, um bom número de fácies seria 6.

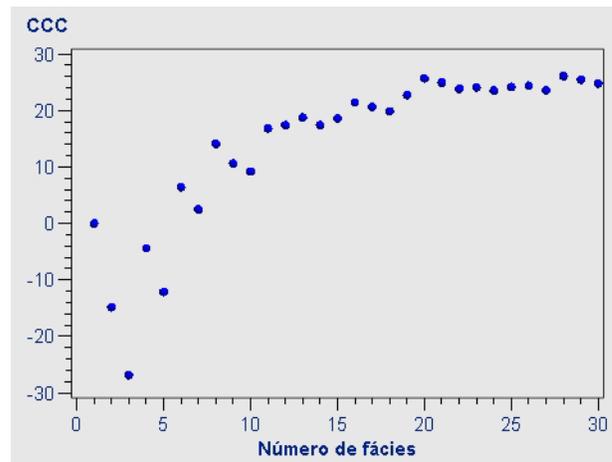


Figura 4 – Estatística CCC contra o número de fácies, segundo a técnica de classificação não-supervisionada *Average Linkage*.

Considerando a intersecção entre os valores indicados pelas três estatísticas, pode-se então considerar que 4 ou 6 seriam os números mais indicados de fácies a reconhecer, nesse caso. O intérprete deve então decidir qual número de fácies considerar, dentre aqueles indicados nesta última etapa. Considerar quatro fácies provavelmente significa reconhecer uma fácies não-reservatório e mais três fácies portadoras de distintos níveis de potencialidade para reservatório. Por outro lado, considerar 6 fácies talvez seja excessivamente detalhado para a modelagem de fluxo. Adotar qualquer outro número de fácies, no entanto, provavelmente resultaria numa precisão menor de predição.

Uma vez definido o número de fácies a reconhecer, deve-se proceder a uma classificação não-supervisionada segundo o método *K-means*. Vale a pena lembrar que as fácies agora reconhecidas não são mais as fácies originais, resultantes da descrição dos testemunhos, mas são fácies compostas, as quais rigorosamente somente podem ser definidas em termos das suas propriedades físicas. Felizmente, o procedimento aqui adotado garante algum nível de correlação entre as principais litofácies e algumas das fácies compostas preditas. Quanto às litofácies de menor representatividade, essas tendem a sofrer um processo de pulverização entre as diversas fácies compostas.

A comparação entre a coluna de fácies compostas e a coluna original de litofácies, mostra que a primeira corresponde à segunda com algum grau de suavização. A perda de detalhes não representa, necessariamente, um fator negativo, desde que as feições significativas

sejam preservadas. Qualquer resultado experimental, inclusive a descrição de testemunhos, está sujeita à presença de ruídos indesejáveis. Ao captar prioritariamente as fácies mais marcantes, esse método de geração de fácies compostas exclui, em algum grau, esses ruídos. Além disto, este fenômeno representa uma mudança de escala (*upscaling*) desejável para as modelagens geológica e de fluxo, etapas posteriores à modelagem de eletrofácies.

Assim, uma característica fundamental deste fluxo de trabalho é a adoção da coluna de fácies compostas como sendo representativa da coluna original de litofácies. É essa coluna de fácies compostas que será utilizada, na etapa seguinte, para gerar a regra discriminante a qual será finalmente utilizada para a modelagem definitiva das eletrofácies.

A próxima etapa consiste na geração da regra discriminante que será doravante utilizada em todas as demais etapas do trabalho de modelagem de eletrofácies. Para isto, a coluna de fácies resultante da classificação não-supervisionada pela técnica *K-means* é dada como dado de entrada para a classificação supervisionada pela RDL ou RDQ. Os resultados desta etapa são a regra discriminante propriamente dita e uma nova classificação de eletrofácies, agora baseada na regra gerada. Neste ponto deve-se avaliar a eficiência do processo examinando se a coluna de fácies gerada pela regra respeita as principais feições da coluna litológica original. Em caso negativo, deve-se refazer o processo a partir da aplicação da regra discriminante passo-a-passo, escolhendo melhor os perfis a utilizar, revendo o número ideal de fácies a modelar e reavaliando o uso da RDL ou da RDQ. Em caso positivo, o processo evolui para a aplicação da regra gerada nos dados de validação.

Após submeter os dados de validação à classificação segundo a regra discriminante gerada, deve-se avaliar novamente o processo. Uma vez que a regra discriminante utilizada foi gerada usando apenas os dados de treinamento, aplicá-la aos dados de validação pode resultar numa classificação errônea, mas isto somente acontecerá se houver uma grande variação das características litológicas ou das propriedades físicas dos tipos litológicos entre os dois conjuntos de dados. Obviamente, nesse caso não se pode extrapolar a regra obtida com os dados de treinamento para outros conjuntos dados e a solução seria redistribuir os dados entre os três conjuntos de dados (treinamento, validação e predição) de forma mais representativa. Para evitar esse tipo de problema, recomenda-se que, antes de começar o trabalho de modelagem de eletrofácies, seja feita uma pesquisa exploratória nos dados, observando quais litofácies estão presentes em quais poços, quais perfis estão disponíveis em quais poços, quais poços possuem descrição de litofácies e como se distribuem as variáveis nos diversos poços. Como resultado dessa pesquisa exploratória de dados, o conjunto de dados de treinamento poderá ser escolhido de forma mais representativa, fazendo com que a aplicação da regra aos dados de validação atinja resultados satisfatórios.

A escolha dos dados de validação deve procurar obedecer a um critério de representatividade espacial, ou

seja, os poços de validação devem ser escolhidos de tal forma que os poços de predição estejam inscritos dentro de um polígono definido pelos poços de validação.

Uma vez aceita a classificação dos dados de validação, segundo a regra discriminante gerada com os dados de treinamento, pode-se então proceder à classificação definitiva dos dados do conjunto de predição. Após essa última etapa, todas as classificações, segundo a regra discriminante (RDL ou RDQ) utilizada, compõem o produto final que é a modelagem de eletrofácies para o campo.

Estudo de caso

A fim de exemplificar a modelagem de eletrofácies, conforme o fluxo de trabalho proposto neste trabalho, serão apresentados, a seguir, os resultados obtidos sobre os dados referentes a quatro poços de um determinado campo produtor de petróleo. Para cada um dos quatro poços, foram disponibilizados os perfis geofísicos de raios gama (GR), resistividade profunda por indução (ILD), porosidade de nêutrons (NPHI), densidade (RHOB), além da descrição faciológica dos testemunhos (coluna de litofácies).

Com o objetivo de reproduzir todas as etapas do fluxo de trabalho apresentado na Figura 1, os dados foram divididos da seguinte maneira: os dados do poço 1 foram utilizados como conjunto de treinamento, os dados do poço 2 foram considerados como de validação, e os dados dos poços 3 e 4 foram usados para predição. Dessa forma, apenas para o poço 1 a coluna de litofácies foi utilizada como dado de entrada para o processo de modelagem de eletrofácies. Para os demais poços a coluna de litofácies foi usada apenas como um controle de qualidade para a coluna de eletrofácies gerada.

A Figura 5 apresenta os resultados obtidos para o poço 1, o qual forneceu os dados de treinamento ou, em outras palavras, de geração da regra discriminante. Esta figura contém cinco colunas que, da esquerda para a direita, representam as seguintes variáveis: coluna de litofácies, coluna de indicação da presença de hidrocarbonetos, coluna de indicação de cimentação, coluna de eletrofácies resultante da classificação não-supervisionada pelo algoritmo *K-means*, e coluna de probabilidade de eletrofácies resultante da classificação supervisionada pela regra discriminante gerada com os dados do poço 1.

As eletrofácies das Figuras 5 a 8, identificadas por cores, são definidas em termos das suas propriedades físicas padronizadas médias, de acordo com a Tabela 1.

Tabela 1 – Média e desvio-padrão das propriedades padronizadas de perfil das eletrofácies dos quatro poços.

Fácies	GR _p	ILD _p	NPHI _p	RHOB _p
Amarela	-1,2 ± 0,2	1,9 ± 0,9	-1,7 ± 0,4	1,1 ± 0,7
Laranja	-0,6 ± 0,5	0,2 ± 0,7	-0,5 ± 0,5	0,3 ± 0,8
Vermelha	1,0 ± 0,8	-1,0 ± 0,2	1,6 ± 0,8	-3,4 ± 1,1
Verde	0,9 ± 0,7	-0,6 ± 0,5	0,8 ± 0,4	-0,4 ± 0,7

Das propriedades médias padronizadas descritas na Tabela 1, vê-se que as fácies amarela e laranja são as melhores fácies-reservatório, a fácies verde corresponde a rochas pelíticas (não-reservatório) e a fácies vermelha é, em geral, uma fácies argilosa de qualidade inferior como rocha-reservatório.

poços. Além disso, foram reconhecidas quatro eletrofácies, sendo três de rochas-reservatório e uma de rochas não-reservatório, todas definidas em termos das propriedades físicas médias registradas pelos perfis geofísicos dos poços.

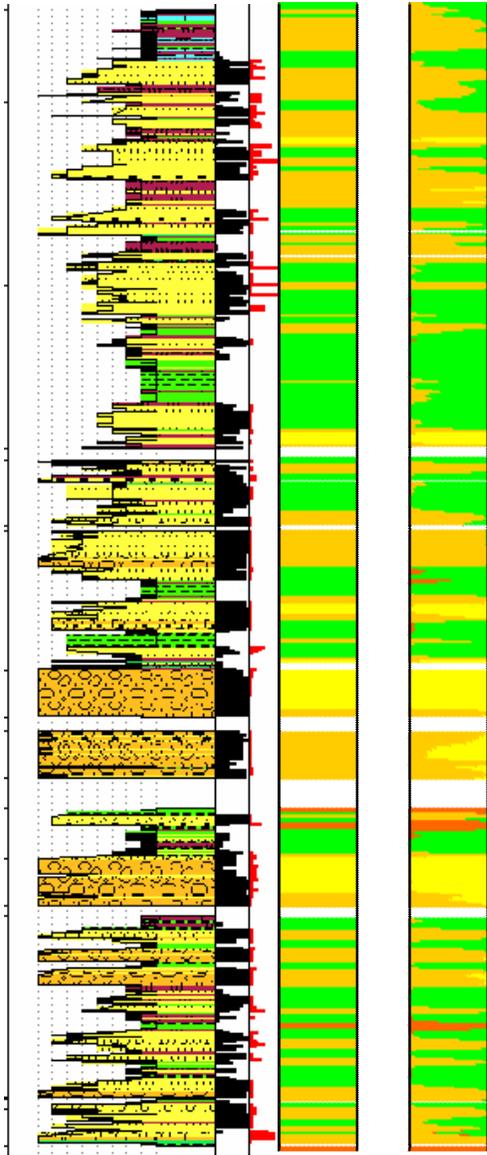


Figura 5 – Colunas de litofácies, hidrocarbonetos, cimentação, eletrofácies por classificação não-supervisionada e eletrofácies por classificação supervisionada para o Poço 1 (dados de treinamento).

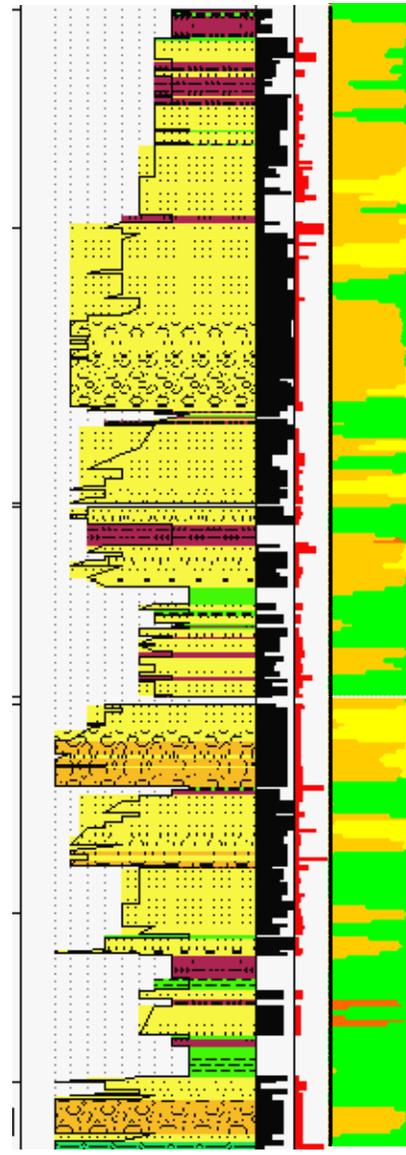


Figura 6 - Colunas de litofácies, hidrocarbonetos, cimentação, e eletrofácies para o Poço 2. A coluna de eletrofácies foi obtida aplicando-se a regra discriminante gerada apenas com os dados do Poço 1.

Discussão

A comparação entre a coluna de probabilidade de classificação de fácies, de cada um dos poços, com a correspondente coluna de litofácies, mostra que as eletrofácies reconhecidas nos Poços 1, 2, 3 e 4, apesar de geradas através de uma regra discriminante definida apenas com dados do Poço 1, reproduzem, na grande maioria das vezes, as litofácies de cada um desses

Deve-se abrir uma discussão a respeito da conveniência de se fazer a padronização dos perfis de modo separado para cada poço, como foi feito neste exemplo, ou se a padronização deve ser feita de uma só vez para todos os poços. Quando uma mesma fácies, em poços diferentes, possui propriedades distintas, é preferível fazer a padronização de forma separada para cada poço. Já quando há falta de alguma fácies, em algum dos poços, deve-se padronizar todos os poços de uma só vez.

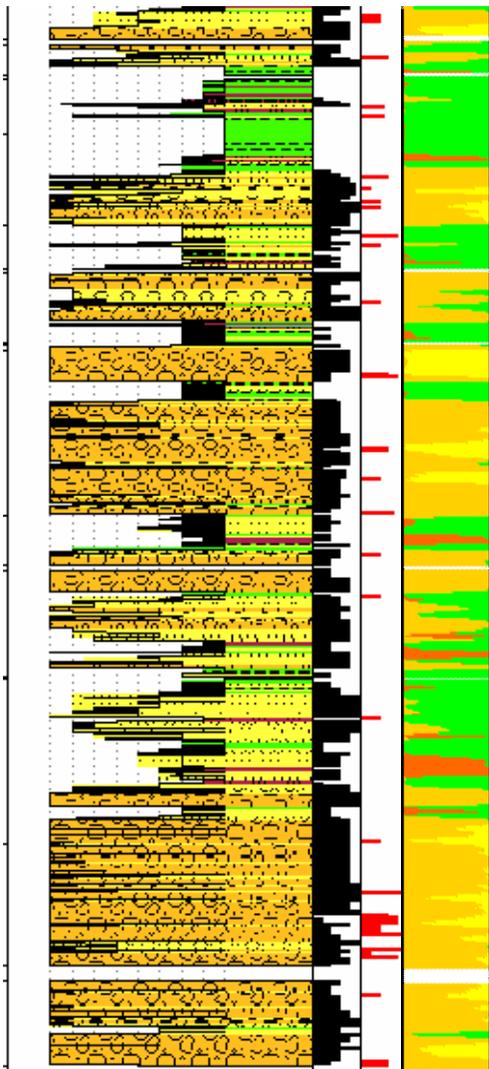


Figura 7 - Colunas de litofácies, hidrocarbonetos, cimentação, e eletrofácies para o Poço 3. A coluna de eletrofácies foi obtida aplicando-se a regra discriminante gerada apenas com os dados do Poço 1.

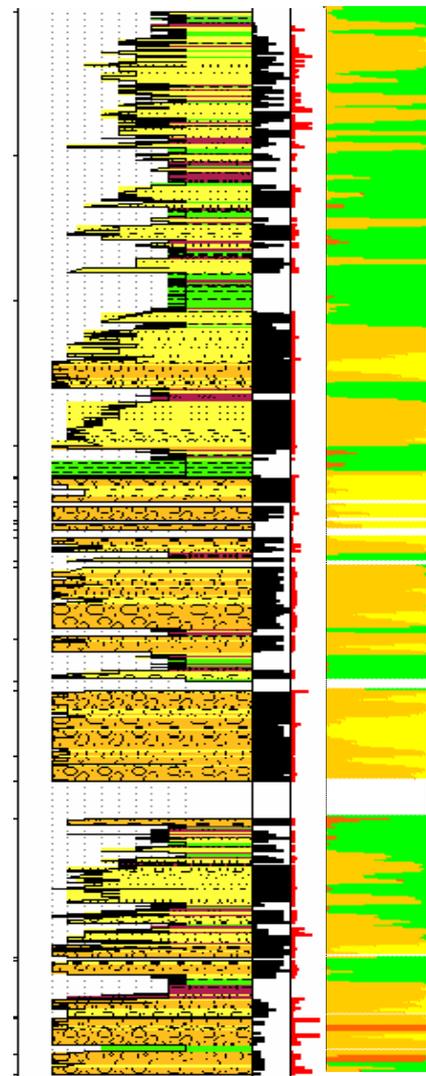


Figura 8 - Colunas de litofácies, hidrocarbonetos, cimentação, e eletrofácies para o Poço 4. A coluna de eletrofácies foi obtida aplicando-se a regra discriminante gerada apenas com os dados do Poço 1.

Conclusões

A modelagem de eletrofácies com entrelaçamento de técnicas de classificação supervisionada e não-supervisionada permite reconhecer de forma adequada as principais fácies, segundo as suas propriedades físicas. A estratégia adotada reduz, em geral, o número de fácies em comparação com o número de fácies originais. Isto funciona como uma espécie de filtro, preservando apenas as fácies mais representativas, o que, do ponto de vista da construção dos modelos geológico e de fluxo de fluidos, é uma característica desejável, pois trabalha favoravelmente ao *upscaling* exigido naquelas etapas posteriores da engenharia de reservatórios. O fluxo de trabalho proposto permite também definir o número ideal de fácies a reconhecer, qual a melhor suíte de perfis a utilizar, dentre todos os perfis disponíveis, e qual a relação entre as eletrofácies encontradas e as litofácies originais.

Agradecimentos

O autor agradece a PETROBRAS e FINEP pelo financiamento do Projeto CTPETRO Perfil, que possibilitou a execução deste trabalho. Agradece ainda às inúmeras colaborações dos professores Cláudio Bettini e João Ismael, da UFRJ, e da geóloga Maria do Socorro de Souza, do Centro de Pesquisas da Petrobras.

Referências

- Massart, D.L. and Kaufman, L. (1983)** *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, New York: John Wiley & Sons, Inc.
- Tatsuoka, M.M. (1971)** *Multivariate Analysis*, New York: John Wiley & Sons, Inc.