



## Modelagem de eletrofácies em minério de ferro por classificador KNN

José Agnelo Soares<sup>1</sup>, Geraldo Girão Nery<sup>2</sup>, Sérgio Brochado Ribeiro de Barros<sup>2</sup>, Fernando Conte Júnior<sup>2</sup>

<sup>1</sup>Universidade Federal de Campina Grande, <sup>2</sup>Hydrolog Serviços de Perfisagens Ltda.

Copyright 2009, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation during the 11<sup>th</sup> International Congress of the Brazilian Geophysical Society held in Salvador, Brazil, August 24-28, 2009.

Contents of this paper were reviewed by the Technical Committee of the 11<sup>th</sup> International Congress of the Brazilian Geophysical Society 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

In this work are showed results for a logfacies procedure applied to a Brazilian iron ore deposit. The procedure has in its central core a KNN classifier, which is a multivariate statistical approach. Some auxiliary tools as step-by-step discriminant analysis, data standardization, and measurement of an n-dimensional Euclidian distance are also used. The well logs used for logfacies modeling are Gamma Ray (GR), Short Normal resistivity (SN), Self Potential (SP) and Sonic (DT). Five facies types are recognized. Due to the high number of available data, it was possible to adopt  $K = 3$ , in a KNN procedure, providing a facies prediction with low error rate and without an excessive smoothed solution. Results show, at the majority of wells, a very good correspondence between lithofacies and logfacies columns, but in two wells the logfacies column show, in their upper part, a higher resolution than that of lithofacies column. The general average index for correct prediction was 91.1%, thus indicating the applied procedure as suitable for iron ore facies recognition, at least for this ore deposit.

### Introdução

A modelagem de eletrofácies consiste na aplicação de métodos numéricos para o reconhecimento das fácies geológicas a partir da análise dos perfis geofísicos de poços. Este tipo de análise é razoavelmente bem difundido na indústria de petróleo, mas ainda pouco utilizado na mineração, provavelmente em consequência do ainda pequeno uso da perfilagem de poços como ferramenta auxiliar exploratória, nesse ramo de atividade.

Existem vários métodos numéricos que tradicionalmente são utilizados para a modelagem de eletrofácies. As principais famílias de métodos são a análise estatística multivariada, as redes neurais e a lógica fuzzy. Cada uma dessas famílias de métodos apresenta características próprias que as individualizam em termos de aplicabilidade para um determinado caso, facilidade de implementação, controle do processo, custo computacional, quantificação da incerteza e facilidade na interpretação dos resultados. Os métodos da análise estatística multivariada são, em geral, os mais trabalhosos, mas que permitem um maior controle de todo o processo de modelagem, o que também resulta

em maior facilidade e segurança na interpretação dos resultados obtidos.

Este trabalho apresenta resultados de um procedimento de modelagem de eletrofácies em treze poços de uma jazida de minério de ferro. Os resultados foram obtidos pela aplicação de um método de classificação baseado em estatística multivariada: o método KNN (*K-nearest neighbors*, ou o método dos K vizinhos mais próximos). Na prática a modelagem de eletrofácies consiste em atribuir uma categoria desconhecida (fácies) a cada profundidade dos poços onde se conhece as propriedades físicas registradas pelos perfis geofísicos.

Uma categoria desconhecida pode ser estimada por análise estatística multivariada em um processo que envolve pelo menos duas fases: treinamento e predição. Na fase de treinamento a regra de classificação é estabelecida e na fase de predição essa regra de classificação é aplicada aos dados para os quais se deseja estimar a categoria desconhecida. Para a fase de treinamento é preciso fornecer as variáveis de análise (no caso os perfis geofísicos) e, na maioria dos métodos, as categorias (fácies) correspondentes, de modo a definir uma relação (regra de classificação) entre as variáveis de entrada (perfis) e as categorias de saída (tipos de fácies). Já na fase de predição basta se conhecer as variáveis de análise (perfis), pois as categorias de saída (fácies) serão estimadas pela aplicação da regra de classificação definida na fase de treinamento.

Quanto ao operador de classificação na fase de treinamento os métodos da estatística multivariada podem ser agrupados em métodos de classificação supervisionada e métodos de classificação não-supervisionada. O primeiro grupo diz respeito aos métodos em que, durante a fase de treinamento, a resposta esperada é conhecida. Neste grupo estão incluídas as regras discriminantes, a correlação canônica e o método KNN. No segundo grupo o treinamento é realizado apenas com os dados de entrada, sem o conhecimento da resposta desejada. Neste caso há apenas uma classificação baseada na separação de grupos de amostras com propriedades similares, como é feito no caso do método da análise de agrupamento.

Os métodos de estatística multivariada podem ainda ser classificados em métodos paramétricos ou métodos não-paramétricos. Os primeiros exigem que todas as variáveis de análise apresentem distribuições estatísticas parametrizadas; por exemplo, que todos os perfis apresentem distribuição gaussiana. Já os métodos não-paramétricos não fazem essa exigência. No primeiro grupo estão as regras discriminantes linear e quadrática e no segundo grupo estão os métodos KNN, regra discriminante passo-a-passo, análise de agrupamentos, e regra discriminante canônica.

### Dados utilizados

Neste trabalho foram utilizados dados de um conjunto de treze poços encravados em uma jazida de minério de ferro localizada no Quadrilátero Ferrífero, estado de Minas Gerais. Esses poços foram perfurados com o objetivo de promover o rebaixamento do nível freático da área de lavra da mina. Posteriormente, esses poços foram perfilados, disponibilizando as curvas de Raios Gama (GR), Resistividade Normal Curta (SN), Potencial Espontâneo (SP), e Tempo de Trânsito (DT).

A descrição de amostras de calha foi realizada em doze dos treze poços e em um dos poços (P10) não foi registrado o perfil SP. A descrição de amostras de calha identificou cinco tipos faciológicos: hematita compacta, hematita branda, itabirito compacto, itabirito brando e filito.

Esses dados foram utilizados em um procedimento de modelagem de eletrofácies com classificador KNN, que gera um modelo para o reconhecimento contínuo das variações faciológicas apresentadas pelo minério de ferro em cada poço. As colunas de eletrofácies foram comparadas com as correspondentes colunas de litofácies, permitindo quantificar o erro de predição em cada poço.

### Procedimento

O procedimento utilizado para a modelagem de eletrofácies consistiu na escolha dos perfis, de acordo com a indicação da regra discriminante passo-a-passo seguido pela classificação KNN. Para o processamento dos dados foi utilizado o pacote estatístico SAS®.

A regra discriminante linear passo-a-passo é um método objetivo para selecionar variáveis que maximizam a previsão com o menor número de variáveis empregadas. Na modelagem de eletrofácies a regra discriminante linear passo-a-passo avalia individualmente cada perfil quanto a sua contribuição na previsão da eletrofácies e acrescenta ou elimina cada perfil ao modelo com base em sua contribuição relativa. Ela examina a contribuição de cada perfil para o modelo de previsão. Todos os perfis são testados no modelo. Cada perfil é considerado para inclusão na etapa inicial de construção iterativa do modelo. O perfil geofísico com a maior contribuição é acrescentado em um primeiro momento. Outros perfis são selecionados para inclusão ou exclusão, com base em sua contribuição incremental sobre os perfis já presentes no modelo. Ao final, a regra discriminante linear passo-a-passo fornece um ranking com os perfis que mais contribuem para a classificação de eletrofácies e excluem aqueles cuja presença no modelo serviria apenas para acrescentar ruído. Esta situação de exclusão é possível quando se tem, por exemplo, dois perfis quase totalmente linearmente dependentes. A mesma informação discriminante está contida nos dois, portanto, acrescentar os dois perfis ao modelo não acrescenta informação, mas prejudica a capacidade preditiva devido ao acréscimo de ruído.

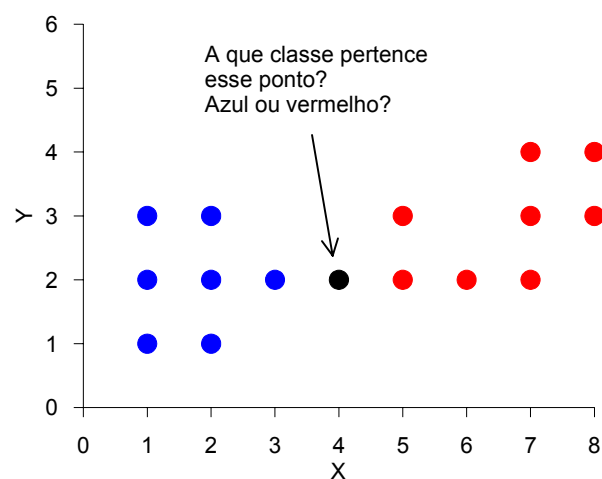
Na modelagem de eletrofácies o método KNN consiste em atribuir a uma dada profundidade o tipo de fácies que

aparece mais vezes entre as fácies dos seus K vizinhos mais próximos, segundo a métrica adotada, no espaço n-dimensional. A dimensão n é igual ao número de perfis utilizados na entrada do modelo. A métrica é dada por uma medida de similaridade no espaço n-dimensional. No caso deste trabalho a métrica utilizada é a distância euclidiana que, essencialmente, é uma medida do comprimento de um segmento de reta, no espaço n-dimensional, desenhado entre duas amostras (Hair *et al.*, 2005):

$$d(1,2) = [(GR_1 - GR_2)^2 + \dots + (DT_1 - DT_2)^2]^{1/2} \quad (1)$$

O mecanismo do método KNN é o seguinte: ele identifica a classe predominante entre as K amostras mais próximas à amostra desconhecida e atribui a ela essa classe. O classificador KNN possui apenas um parâmetro livre (o número K) que é controlado pelo usuário com o objetivo de obter uma melhor classificação.

A Figura 1 apresenta um exemplo ilustrativo do mecanismo de classificação KNN com apenas duas dimensões (n = 2: X e Y), dois rótulos de saída (as classes azul e vermelho) e uma amostra na qual a classe é desconhecida: o ponto preto. O resultado da classificação depende do valor de K. Se K = 1 não há solução possível, pois o ponto azul mais próximo está à mesma distância do ponto preto que o ponto vermelho mais próximo. Se K = 2 o impasse continua, pois entre os dois pontos mais próximos há um azul e um vermelho. Se K = 3 ao ponto preto é atribuída a classe vermelha, pois entre os 3 pontos mais próximos há dois vermelhos e apenas um azul. Se K = 7 o ponto preto é classificado como sendo azul, pois há quatro pontos azuis e três pontos vermelhos entre os 7 vizinhos mais próximos.



**Figura 1** – Influência do parâmetro K no método KNN (extraído de [www.ppgia.pucpr.br/~soares](http://www.ppgia.pucpr.br/~soares)).

Escolher o valor de K não é um problema trivial. No entanto alguns princípios gerais devem ser observados: a) K deve ser preferencialmente um número ímpar, pois ao adotar um número par para K o procedimento de classificação KNN pode identificar um mesmo número de rótulos entre os K vizinhos mais próximos, o que resulta na não classificação da amostra; b) K deve ser suficientemente grande para garantir um nível aceitável

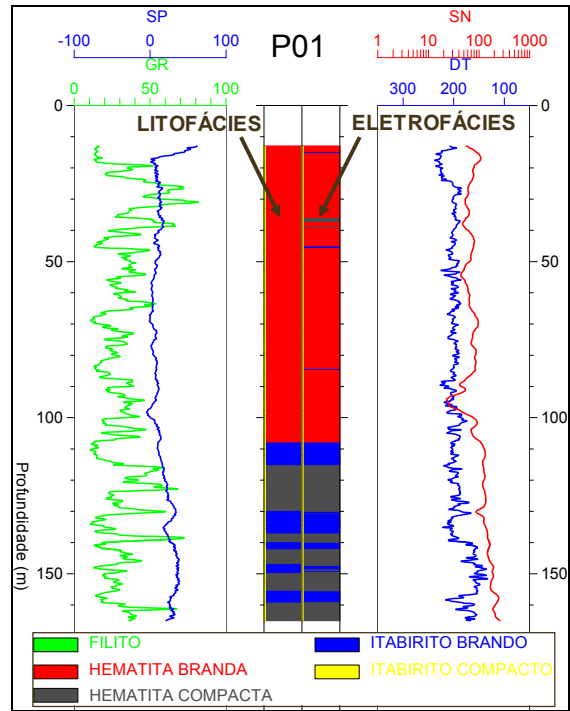
de acertos; e, c) K deve ser suficientemente pequeno para garantir a resolução desejada. Mas quando se dispõe de uma base de dados muito grande - como é caso de uma base de dados de perfis geofísicos de poços - o erro é pequeno mesmo para um K pequeno (K = 1 ou K = 3). Para uma base de dados muito grande (n → ∞) o erro KNN é menor que duas vezes o erro Bayesiano, que é teoricamente o menor erro possível de ser alcançado.

A classificação KNN depende da distância medida, entre amostras, no espaço n-dimensional, definido pelos perfis de entrada do modelo. No entanto, os perfis apresentam intervalos de variação bastante distintos. Por exemplo, a curva GR em geral apresenta uma variação linear entre 0.0 API e 100.0 API, enquanto que a curva SN apresenta uma variação exponencial entre algo como 1.0 ohm.m e 1000 ohm.m. Essa diversidade de escalas dificulta a classificação baseada em uma medida de distância. A fim de contornar esse problema, aplica-se um procedimento de transformação e padronização dos dados, conforme descrito em Soares (2005), do qual resultam curvas com variação linear dentro de um mesmo intervalo de valores. Esta providência melhora substancialmente a medição da distância utilizada no método KNN.

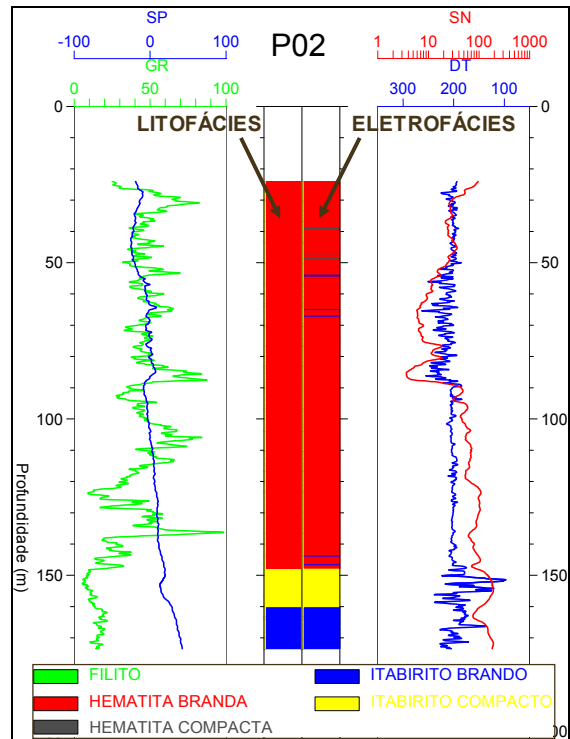
**Aplicação**

O procedimento de modelagem de eletrofácies foi aplicado aos perfis geofísicos de treze poços localizados em uma mina de minério de ferro. Testes preliminares indicaram que a adoção de K = 3 gerava colunas de eletrofácies mais compatíveis com as respectivas colunas de litofácies, sendo este o valor adotado para o parâmetro K. As colunas de litofácies representam a descrição faciológica direta das amostras de calha dos poços. Os poços estudados receberam os seguintes títulos de referência: P01, P02, P03, P04, P05, P06, P07, P08, P09, P10, P11, P12 e P13. O poço P13 não dispõe da descrição de amostras de calha e o poço P10 não tem a curva SP. Os demais poços dispõem da descrição litológica e da suíte completa de perfis.

As Figuras 2 a 14 apresentam os perfis geofísicos dos poços nas colunas à esquerda e à direita de cada figura. A coluna central é a coluna de fácies: na metade esquerda vê-se a seqüência de litofácies, conforme descrição direta de amostras de calha, e na metade direita vê-se a seqüência de eletrofácies, resultante da classificação pelo método KNN. Na coluna central a cor verde indica a ocorrência de filito, que é a rocha encaixante, a cor vermelha indica hematita branda, a cor grafite indica hematita compacta, a cor azul indica a ocorrência de itabirito brando e a cor amarela representa itabirito compacto. Há, na maioria dos poços, uma correspondência muito grande entre as colunas de litofácies e eletrofácies, o que comprova a eficiência do classificador KNN no reconhecimento das variações faciológicas do minério de ferro. Nos poços P09 e P10 a coluna de eletrofácies apresenta, na parte superior da seção, um maior detalhamento em relação à coluna de litofácies. O índice médio de acerto para todas as fácies dos doze poços que possuem descrição de amostras de calha foi de 91,1%.



**Figura 2** – Perfis geofísicos, coluna de litofácies (metade esquerda da coluna central) e coluna de eletrofácies (metade direita da coluna central) para o poço P01.



**Figura 3** – Perfis geofísicos, coluna de litofácies (metade esquerda da coluna central) e coluna de eletrofácies (metade direita da coluna central) para o poço P02.

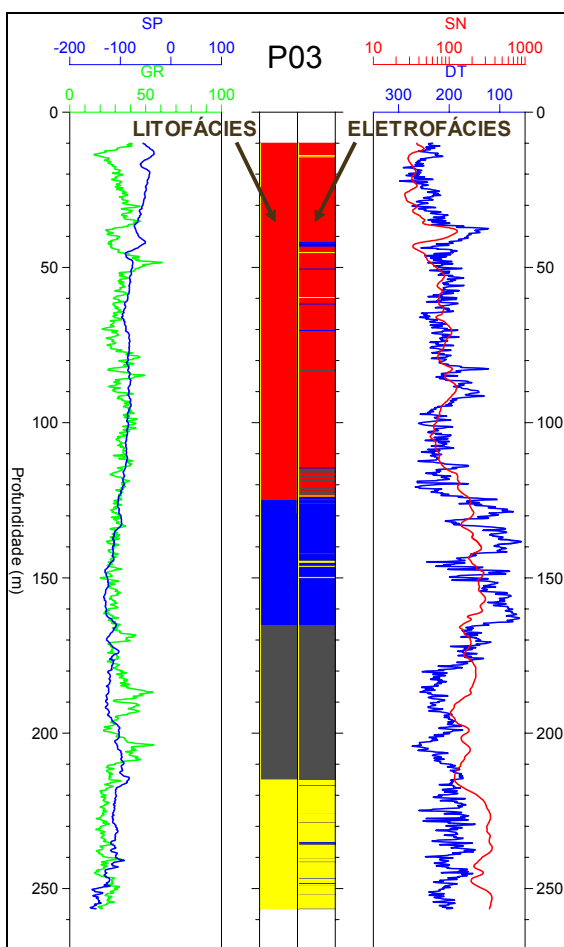


Figura 4 – Perfis geofísicos, coluna de litofácies (metade esquerda da coluna central) e coluna de eletrofácies (metade direita da coluna central) para o poço P03.

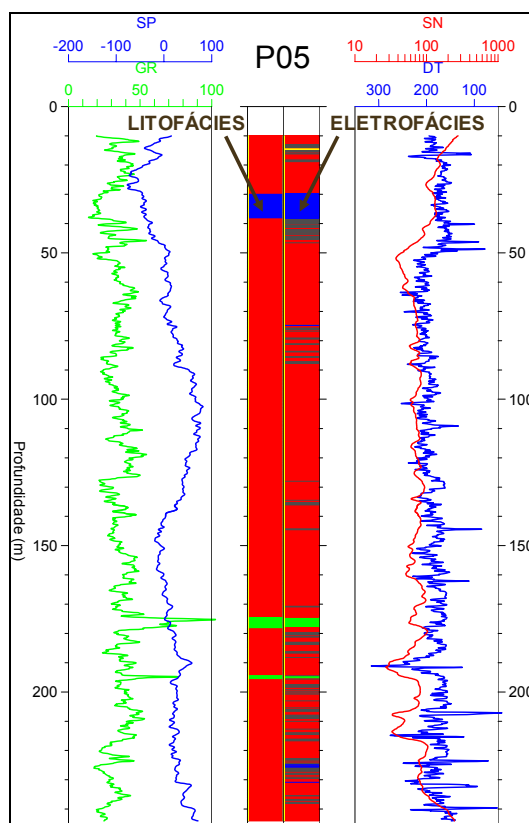


Figura 6 – Perfis geofísicos, coluna de litofácies (metade esquerda da coluna central) e coluna de eletrofácies (metade direita da coluna central) para o poço P05.

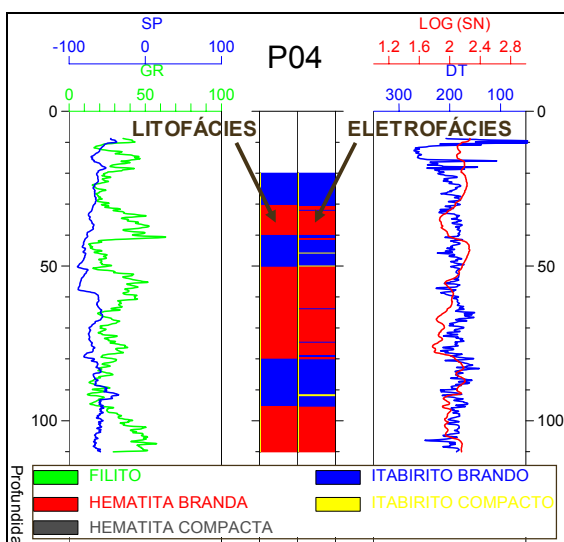


Figura 5 – Perfis geofísicos, coluna de litofácies (metade esquerda da coluna central) e coluna de eletrofácies (metade direita da coluna central) para o poço P04.

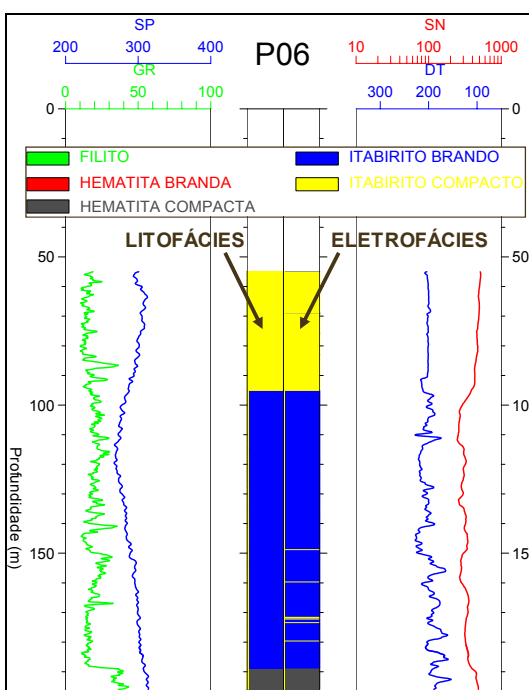
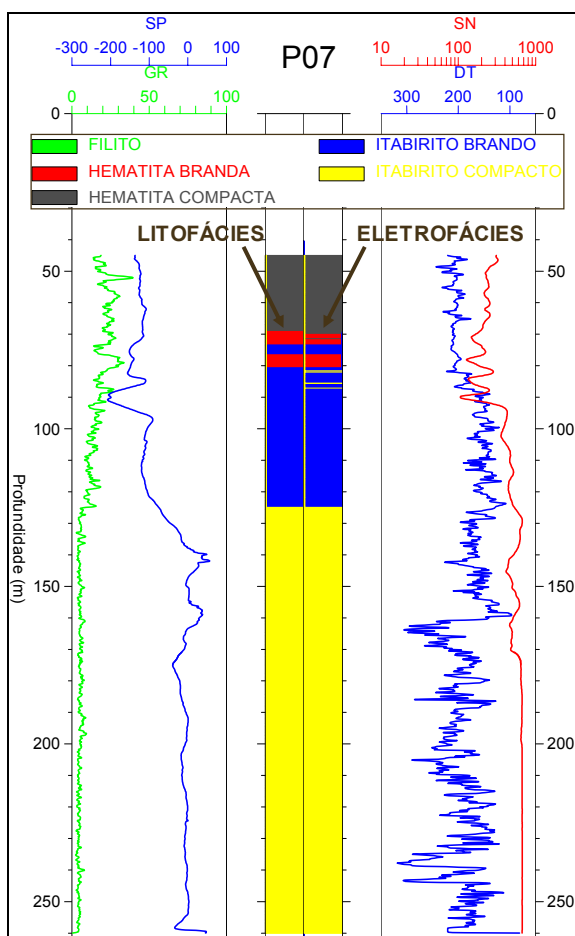
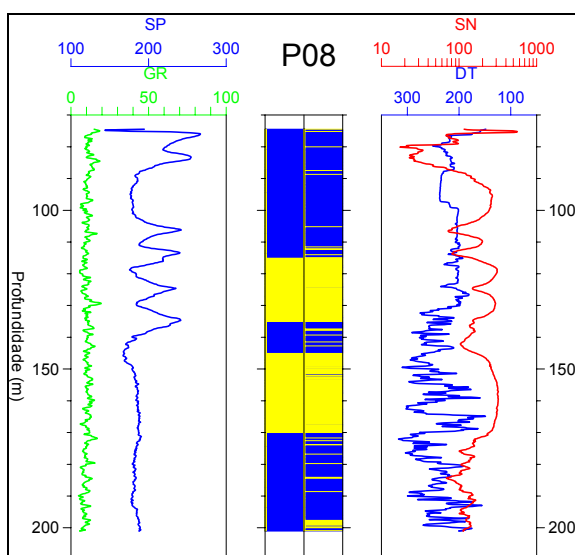


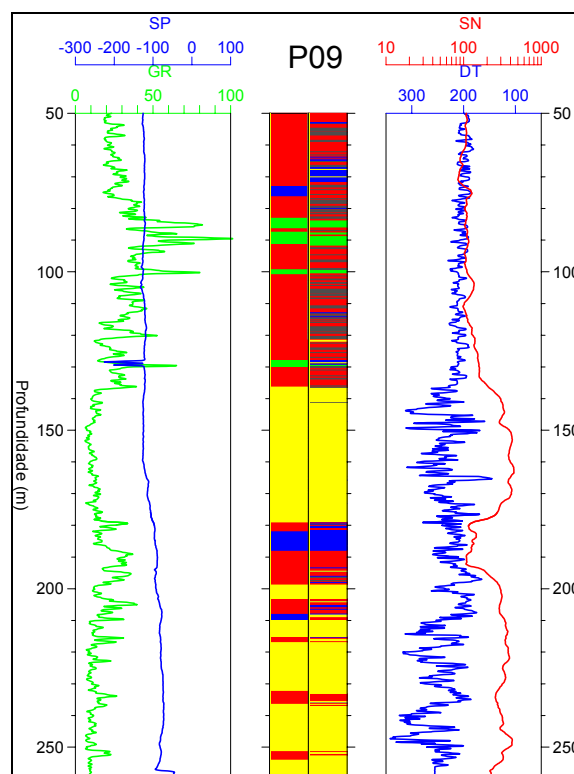
Figura 7 – Perfis geofísicos, coluna de litofácies (metade esquerda da coluna central) e coluna de eletrofácies (metade direita da coluna central) para o poço P06.



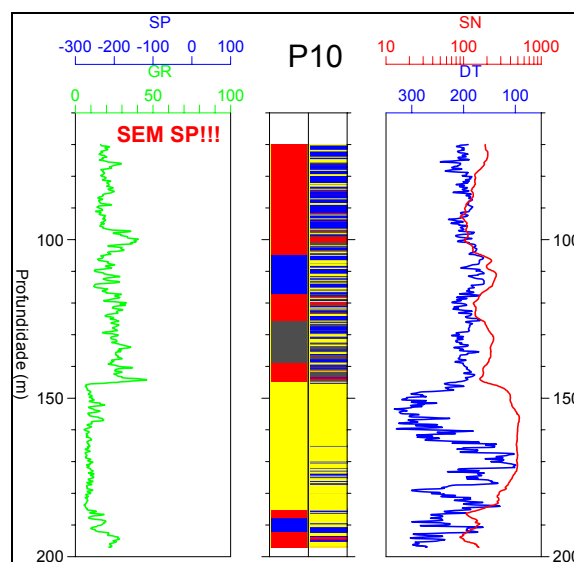
**Figura 8** – Perfis geofísicos, coluna de litofácies (metade esquerda da coluna central) e coluna de eletrofácies (metade direita da coluna central) para o poço P07.



**Figura 9** – Perfis geofísicos, coluna de litofácies (metade esquerda da coluna central) e coluna de eletrofácies (metade direita da coluna central) para o poço P08.



**Figura 10** – Perfis geofísicos, coluna de litofácies (metade esquerda da coluna central) e coluna de eletrofácies (metade direita da coluna central) para o poço P09.



**Figura 11** – Perfis geofísicos, coluna de litofácies (metade esquerda da coluna central) e coluna de eletrofácies (metade direita da coluna central) para o poço P10. Observar que neste poço não foi registrado o perfil SP.

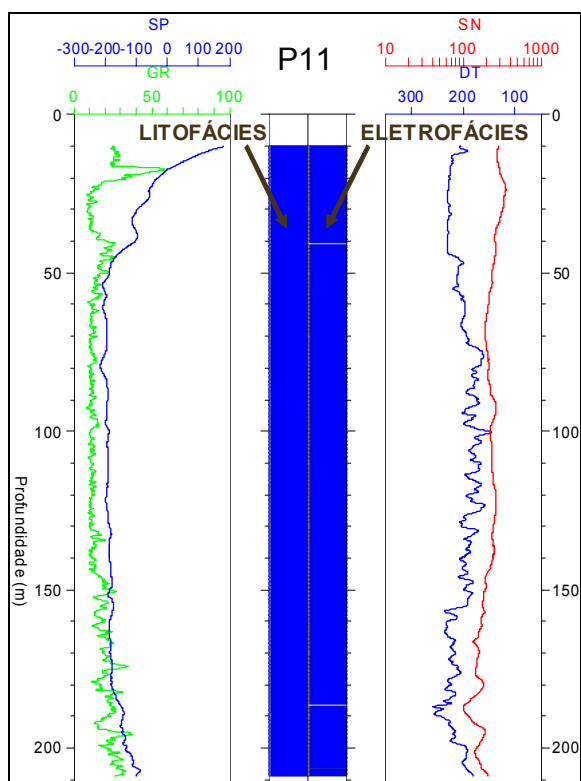


Figura 12 – Perfis geofísicos, coluna de litofácies e coluna de eletrofácies para o poço P11.

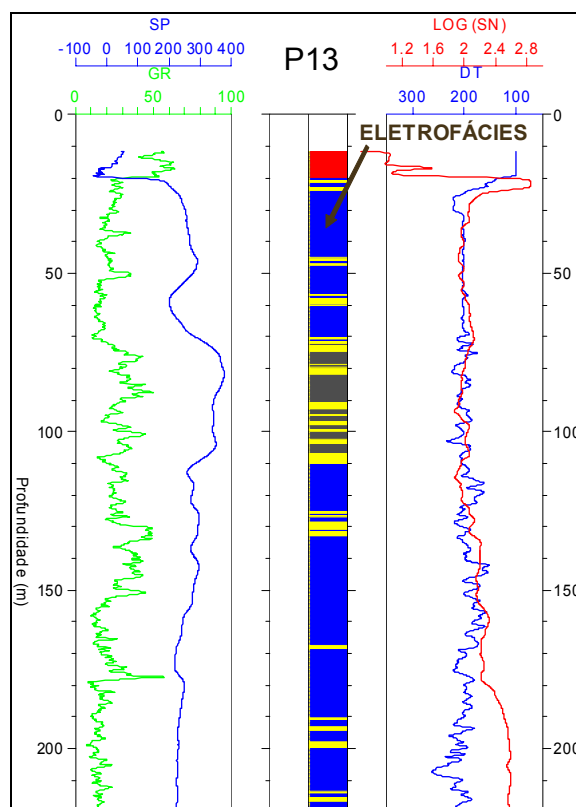


Figura 14 – Perfis geofísicos e coluna de eletrofácies para o poço P13. Este poço não possui registro da coluna de litofácies.

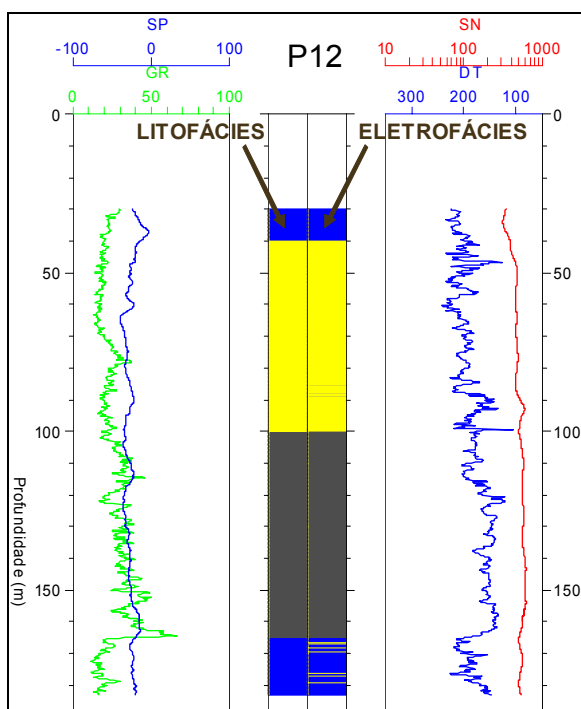


Figura 13 – Perfis geofísicos, coluna de litofácies e coluna de eletrofácies para o poço P12.

### Conclusões

O procedimento utilizado para a modelagem de eletrofácies, com o classificador KNN no seu núcleo central, se mostrou adequado para a análise da jazida de minério de ferro em epígrafe. A análise comparativa entre as colunas de litofácies e de eletrofácies indicou um índice de acerto médio de 91,1%. O uso do parâmetro K igual a três é possível porque na modelagem de eletrofácies normalmente se dispõe de uma base de dados muito grande. Esse pequeno valor de K garante um alto nível de acerto sem suavizar demais a solução. Apesar de o método proposto oferecer uma solução automática para o reconhecimento das fácies, é fundamental reconhecer a importância do conhecimento especialista, especialmente na definição das litofácies, para o sucesso da modelagem de eletrofácies.

### Referências

Hair, F.H.J.; Anderson, R.E.; Tatham, R.L.; Black, W.C. (2005) Análise Multivariada de Dados. 5ª ed., Porto Alegre, Bookman.

Soares, J.A. (2005) Um fluxo de trabalho para modelagem de eletrofácies com entrelaçamento de técnicas de classificação supervisionada e não-supervisionada. 9º Congresso Internacional da SBF. Salvador.