Aplicação de técnicas de estatística multivariada na definição da litologia a partir de perfis geofísicos de poços

Sérgio Sacani Sancevero¹, Armando Zaupa Remacre², Alexandre Campana Vidal³ & Rodrigo de Souza Portugal⁴

Resumo A definição da litologia em poços de petróleo, por meio da análise de múltiplos perfis geofísicos, possui um papel importante dentro do processo de caracterização de reservatórios. A partir dessa é possível gerar modelos litológicos que por sua vez serão preenchidos pelas propriedades petrofísicas e podem então ser utilizados em simuladores de fluxo para entender e estudar o comportamento de um campo de petróleo. Normalmente ao se perfurar um poço, somente em uma pequena parte são extraídos testemunhos, isso ocorre devido ao alto custo na extração deste tipo de dado. Contudo, nos poços são corridos alguns perfis geofísicos em toda a sua extensão. Com esses perfis se tem a indicação indireta das rochas presentes nos poços. Desse modo é importante saber como combinar essas medidas, numa análise conjunta, de modo a definir a litologia nas porções não amostradas dos poços. Com essa motivação, o objetivo deste trabalho é aplicar diferentes técnicas estatísticas multivariadas para definir, a partir de um conjunto de perfis, a litologia presente em um poço específico. Com esse trabalho, pode-se mostrar que além de técnicas tradicionais como a análise de componentes principais, existem outras técnicas como o k-ésimo vizinho mais próximo, o qual pode fornecer resultados satisfatórios e ser aplicado de maneira conclusiva no processo de caracterização de reservatórios.

Palavras-Chave: Caracterização de Reservatórios, Estatística Multivariada, Perfilagem.

Abstract *Application of multivariate statistics techniques for lithology definition from geophysical well logs.* The definition of lithology in petroleum wells from analysis of multiple geophysics logs has an important role in the reservoir characterization process. As a matter of fact, from the definition of lithology is possible to build lithological models that can be filled with their respective petrophysical properties and, ultimately, they can be used in flow simulators in order to study the field's behavior. Usually in the wells, only a little part of it is studied by core samples, mainly because the elevated costs in extract the samples from it. However, in the wells, some geophysical logs are running in all extension of it. With these logs, we have an indirect indication of the rocks that are present in the well. Therefore it is important to know and how to combine that information in a joint analysis to define the lithology in the portions, of the well that have no core samples. Given this motivation, the aim of this work is applying multivariate statistics techniques to define from the geophysical logs available the lithology that are present in the specific well. With this work, it is possible to show that beyond the traditional techniques, such as the principal components analysis, there are other techniques, such as K near neighbor, which can provide satisfactory results and can be applied vastly in the reservoir characterization process.

Keywords: Reservoir Characterization, Multivariate Statistics, Well Log.

INTRODUÇÃO A classificação litológica é uma etapa importante e complexa no processo de caracterização de reservatórios. Normalmente, a litologia é identificada com maior precisão a partir da análise de testemunhos, porém, principalmente em reservatórios de águas profundas a obtenção de testemunhos não é realizada em todos os poços e mesmo nos poços amostrados não é efetuada em toda a extensão do mesmo, sendo assim uma informação não contínua. Além de obter testemunhos para uma análise direta da litologia, ao se perfurar um poço uma série de medidas elétricas e radioativas são também obtidas por meio de perfis geofísicos.

Desse modo, esses perfis podem ser usados para realizar a definição da litologia, uma vez que eles respondem a uma grande quantidade de parâmetros geológicos e são corridos em toda a extensão do poço. Porém, realizar a definição litológica a partir dos perfis de poços pode ser uma tarefa complexa, especialmente quando grande quantidade de informações estão envolvidas. Para tornar essa tarefa eficaz, a estatística multivariada que envolve etapas como a redução da dimensionalidade e o reconhecimento de padrões é uma prática adotada com frequência (Talaat, et al., 1989).

A extração de informações dos resultados de

1 - DGRN/IG/UNICAMP, Cidade Universitária Zeferino Vaz, Campinas (SP), Brasil. E-mail: sacani@gmail.com

2 - DGRN/IG/UNICAMP, Cidade Universitária Zeferino Vaz, Campinas (SP), Brasil. E-mail: armando@ige.unicamp.br

3 - DGRN/IG/UNICAMP, Cidade Universitária Zeferino Vaz, Campinas (SP), Brasil. E-mail: vidal@ige.unicamp.br

4 - DGRN/IG/UNICAMP, Cidade Universitária Zeferino Vaz, Campinas (SP), Brasil. E-mail: portugal@ige.unicamp.br

uma campanha de perfilagem de poços, com o objetivo de definir a litologia, envolve a análise de uma grande quantidade de variáveis e de amostras além da aplicação de diferentes métodos para se realizar tal definição. De acordo com Doveton (1994), existem basicamente 3 métodos para se definir a litologia a partir de perfis geofísicos de poços. Num primeiro método, denominado de abordagem teórica, a resposta do perfil para cada fácies é prevista multiplicando a componente da propriedade do perfil pela proporção com a qual ela ocorre em fácies equivalentes, como essa abordagem é baseada num modelo, se esse não é bem definido pode-se gerar fácies adicionais. Numa segunda abordagem, denominada de empírica as fácies observadas em testemunhos são relacionadas a um conjunto de perfis geofísicos, as estatísticas desses perfis como a média, variância e covariância, são descritas como elipsóides multidimensionais para cada uma das fácies, onde existem os perfis e as seções testemunhadas esse método pode ser aplicado de maneira efetiva, porém onde não existe um dos dois fatores os resultados começam a perder a validade. A última abordagem é a interpretativa, onde a definição de fácies se baseia na examinação de gráficos de correlação onde se identificam agrupamentos para cada fácies e que podem ser validados por amostras de testemunhos. Os gráficos construídos para esse tipo de análise podem ser provenientes de uma análise estatística multivariada onde a redundância da informação é minimizada e os agrupamentos se tornam mais claros. Pode-se então dizer que a definição da litologia é um processo de reconhecimento de padrões, baseado na interpretação da característica dos perfis geofísicos e suplementada pelo conhecimento geológico.

Atualmente, pode-se dizer que uma das áreas do conhecimento de maior avanço é a estatística multivariada. Isso ocorre, pelo fato de que a cada dia que se passa o mundo que nos cerca se torna cada vez mais complexo e multivariado. No caso da caracterização de reservatórios e especialmente no caso da perfilagem de poço, essa afirmação é ainda mais verídica. Extrair informações verdadeiramente úteis de um grande número de perfis envolve a análise de um grande número de variáveis. Muitas vezes, ocorre que um pequeno número dessas variáveis contém as informações mais relevantes para o processo de caracterização de reservatórios, enquanto a maioria das variáveis, pouco adiciona à interpretação dos resultados em termos quantitativos. Decidir sobre quais variáveis são relevantes é um processo que exige grande conhecimento geológico sobre a área e tem um importante caráter objetivo, pois definir um modelo estatístico sem esse conhecimento prévio pode agregar grande incerteza ao resultado.

Assim, para se analisar de maneira conjunta as variáveis presentes no conjunto de dados e dar a elas um significado objetivo, foram desenvolvidos os métodos estatísticos multivariados. Num primeiro plano, o objetivo geral de todos os métodos multivariados é realizar de alguma forma a redução na dimensionalidade dos dados. Devido a isso, é possível avaliar as variáveis por simples correlações que podem ser realizadas com

poucas variáveis. Quando se trabalha com um número de variáveis maior que três é necessário realizar correlações entre as variáveis, ou seja, nesse espaço n -dimensional, os métodos conhecidos de análise monovariada, como os histogramas, e os métodos de análise bivariada como os gráficos de correlação, começam a perder a praticidade. Por isso a cada dia tem-se aplicado de maneira mais decisiva a análise estatística multivariada no processamento e interpretação de perfis geofísicos. Esta possui métodos específicos que podem ser aplicados a este problema, primeiro reduzindo o número de variáveis e, na seqüência, classificando as informações para a confecção do perfil litológico. A redução de variáveis é realizada aplicando critérios objetivos, por meio de gráficos bidimensionais que contenham a maior quantidade de informação estatística possível. Isso é feito utilizando análise de componentes principais, neste caso considerada como uma ferramenta de análise exploratória multivariada. Após essa análise, é possível então agrupar as variáveis de acordo com suas similaridades. O agrupamento e a conseqüente classificação de perfis são realizados seguindo duas abordagens, uma supervisionada e a outra não-supervisionada.

Ambas abordagens possuem vantagens e desvantagens. Na abordagem não-supervisionada, todos os perfis de poços são usados e a interferência do geólogo só ocorre no final do processamento, para decidir se os tipos foram corretamente definidos. Em alguns casos, pode ocorrer a definição de tipos inexistentes no reservatório. Por outro lado, na abordagem supervisionada, é necessário conhecer em determinadas posições do testemunho o tipo de rocha analisada, de modo a guiar o processo de classificação. Com isso, não são definidos tipos inexistentes, porém, para se guiar tal classificação, como já foi dito, apenas pequenas porções dos poços são disponíveis (Wong, et al., 2002).

Assim, o objetivo deste trabalho é definir a distribuição litológica de um poço de petróleo por meio da análise estatística multivariada, iniciando com a análise exploratória com a utilização do método de componentes principais, passando para a análise de agrupamentos e classificação, com a aplicação de técnicas não-supervisionadas como o k -médiãs e técnicas supervisionadas como o k -ésimo vizinho mais próximo e a regressão por componentes principais. Com a aplicação dessas técnicas será possível então definir o perfil litológico do poço a partir de um conjunto de perfis geofísicos.

MATERIAIS E MÉTODOS

Materiais Para a realização deste trabalho, será utilizado um poço pertencente ao Campo de Namorado. Descoberto em 1975, o campo gigante de Namorado está localizado na parte central norte na zona de acumulação de hidrocarbonetos da Bacia de Campos, a 80 km da costa. Em profundidade, o reservatório situa-se entre -2900 e -3400 m. O reservatório principal do campo é constituído por arenitos arcóseos de granulometria fina a grossa, denominados de arenitos turbidíticos de Namorado (Johann, 1997).

O reservatório de Namorado é limitado por fe-

chamentos estratigráficos e por falhas normais. O reservatório possui uma espessura que varia entre 90 m e 180m, cuja porosidade e permeabilidade média são respectivamente 26% e 400mD (Bacoccoli et al., 1980).

Nos últimos anos, foram disponibilizados pela Agência Nacional de Petróleo (ANP) dados referentes ao Campo de Namorado. Essa base de dados possui 56 poços distribuídos na área do reservatório de acordo com o mapa base da figura 1.

Cada poço disponibilizado possui um conjunto de perfis que incluem: sônico (DT), densidade (RHOB), raios gamma (GR), porosidade neutrão (NPHI) e resistividade (ILD). Além disso, para alguns poços foram também disponibilizadas análises sequenciais de testemunhos. Para esse estudo, foi escolhido o poço NA02, mostrado em destaque na figura 1. Esse poço foi escolhido, devido a sua localização e também por possuir todas as informações necessárias para a execução deste trabalho.

Na figura 2, são apresentados os 5 perfis acima descritos para o poço NA02, juntamente com a coluna litológica definida também pela ANP, coluna essa, que foi digitalizada para poder ser utilizada nos algoritmos desenvolvidos.

Pode-se observar na análise de testemunhos a presença de basicamente três tipos de rochas neste poço: (1) Amarelo: arenito; (2) Verde: folhelho; (3) Vermelho: siltito argiloso.

Além disso, em algumas regiões, não existe a informação do tipo de rocha (representada na figura por intervalo não definido). Desse modo com as técnicas multivariadas descritas a seguir, o objetivo é reproduzir

da melhor forma possível os tipos de rochas nas porções testemunhadas e utilizar esse método para preencher as porções não testemunhadas gerando, ao final, o perfil completo da litologia do poço NA02.

Ainda na figura 2 é possível observar que existe uma região por volta da profundidade 3050m, onde não existe amostra de testemunho, porém, pode-se notar pelo padrão principalmente dos perfis de densidade e de porosidade neutrônica, que o comportamento desses nessa região não amostrada é semelhante ao comportamento na porção de arenito, localizada logo abaixo de tal região. Com isso, embora não se tenha o testemunho para poder avaliar a precisão dos métodos, se tem um comportamento de variáveis que sugere ser essa região não amostrada também constituída por arenitos. Essa comparação é apresentada na figura 3.

Métodos A abordagem utilizada neste trabalho consiste basicamente de duas etapas: na primeira etapa realiza-se a análise exploratória multivariada dos dados usando para isso a análise de componentes principais (PCA); na segunda etapa é feita a classificação dos dados utilizando três técnicas distintas: k-médias; regressão por componentes principais (PCR); e k-ésimo vizinho mais próximo (KNN)). Todas essas etapas foram primeiramente realizadas nas regiões testemunhadas e após a avaliação dos resultados, extrapolou-se para todo o perfil, onde se pode assim verificar a consistência do resultado e prever a litologia onde não há testemunho. Uma sequência deste trabalho, não realizada aqui, é a utilização de outros poços, onde após a classificação, os resultados podem ser extrapolados para todo o res-

Figura 1 - Mapa base, mostrando a distribuição dos poços na área de estudo e em destaque o poço utilizado neste estudo.

Figura 2 - Perfis geofísicos referentes ao poço NA02, juntamente com a coluna litológica disponibilizada pela ANP.

ervatório. Na figura 4, é apresentado o fluxograma que descreve as etapas envolvidas na realização deste trabalho.

ANÁLISE DE COMPONENTES PRINCIPAIS (PCA) A análise de componentes principais (PCA) é uma técnica estatística multivariada que seleciona combinações lineares não correlacionadas de variáveis em um espaço

de várias dimensões (Jolliffe, 2002). Por meio do uso da PCA torna-se possível reduzir a dimensionalidade do conjunto de dados. Na análise de PCA, o objetivo é encontrar uma combinação linear $y = Wx$ de modo que a variância seja maximizada. Essa técnica também pode ser vista como uma transformação linear que minimiza o erro de reconstrução. Cada linha da matriz W corresponde ao autovetor ortogonal normalizado da matriz de covariância dos dados. Uma abordagem simples da PCA é usar a decomposição singular de valores (SVD). Se a matriz de covariância dos dados é definida por $R_x(0) = E\{x(t)x^T(t)\}$, então a SVD de $R_x(0)$ é:

$$R_x(0) = UDU^T = [U_s, U_n] \begin{bmatrix} D_s & \\ & D_n \end{bmatrix} [U_s, U_n]^T \quad (1)$$

Onde U é a matriz de autovetores e D é a matriz diagonal sendo os elementos diagonais correspondentes aos autovalores de $R_x(0)$. Então, a transformação de PCA de dados m -dimensional para um subespaço n -dimensional é realizada escolhendo as primeiras n colunas, ou seja, os n componentes principais do vetor y , que são calculados como:

$$y = U_s^T x \quad (2)$$

A técnica de PCA pode ser considerada basicamente como uma técnica de pré-processamento de dados multivariados, porém, sua consequência direta é

Figura 3 - Comparação entre os perfis densidade (RHOB) e porosidade neutrônica (NPHI), juntamente com a litologia, para uma região testemunhada e não-testemunhada com a finalidade de prever a litologia a ser identificada para tal região.

Figura 4 - Fluxograma das atividades realizadas para o processo de definição do perfil litológico no poço NA02.

a aplicação dentro de um processo de classificação e agrupamento.

ANÁLISES DE AGRUPAMENTOS Os métodos de agrupamento fornecem ferramentas poderosas para analisar conjuntos de dados multivariados e têm suas aplicações nas mais distintas áreas. A análise de agrupamento é um termo usado para descrever diversas técnicas numéricas, cujo propósito fundamental é classificar os valores de uma matriz de dados, em grupos discretos. A técnica classificatória de análise de agrupamentos pode ser utilizada quando se deseja explorar as similaridades entre indivíduos ou entre variáveis definindo-se em grupos, considerando simultaneamente no primeiro caso todas as variáveis medidas em cada indivíduo, ou seja, cada atributo em cada amostra, e no segundo todos os indivíduos nos quais foram feitas medições, ou seja, no caso de perfis de poços, as amostras obtidas a cada 20cm. Os métodos de agrupamentos pesquisam na base de dados por grupos homogêneos de itens, representados por pontos num espaço n-dimensional em um número convincente de grupos relacionando-os por meio de coeficientes de similaridade ou de correspondência.

Desse modo, por meio dos métodos de agrupamentos, podem-se separar os dados em grupos numa situação onde nenhuma informação a priori sobre os grupos é disponível (classificação não-supervisionada), ao contrário dos métodos de classificação onde existem informações a priori sobre os grupos e suas características individuais que podem ser utilizadas (classificação supervisionada).

O objetivo de uma análise de agrupamento é que as unidades dentro dos grupos devem ser as mais parecidas possíveis e os grupos formados devem ser o mais distintos possível. Não se deve esperar que a análise de agrupamento sempre divida os dados em grupos completamente distintos, às vezes uma sobreposição ocorre entre os grupos. As principais razões para se fazer uma

análise de agrupamento são: (a) Exploração eficiente da base de dados; (b) Visualização; (c) Redução dos dados; (d) Geração de hipóteses para agrupamentos de novas variáveis.

K-médias A primeira técnica utilizada para agrupar os dados e então se tentar definir a litologia foi o k-médias. O objetivo desta técnica é identificar subconjuntos dentro de um espaço n-dimensional, para tanto se pode utilizar, ou as variáveis originais, ou as variáveis já transformadas por meio da PCA (McQueen, 1967). A técnica do k-médias, não propõe uma redução na dimensão nos dados, mas sim o particionamento dos dados em distintos conjuntos. Esse agrupamento é feito por meio do cálculo de distâncias, para isso pode-se utilizar diversos algoritmos, neste trabalho foi utilizada a distância de Mahalanobis, que é a mais indicada para se trabalhar com dados transformados. Essa técnica de agrupamentos é não-supervisionada, porém é necessário a priori definir o número de classes que se deseja identificar, é uma técnica muito utilizada no agrupamento e na classificação de dados por ser rápida computacionalmente.

Os grupos são definidos com a técnica do k-médias minimizando a soma dos erros ao quadrado de acordo com a seguinte equação:

$$J_K = \sum_{k=1}^K \sum_{i \in G_k} (x_i - m_k)^2. \quad (3)$$

Onde, x_i são os pontos pertencentes a matriz de dados e:

$$m_k = \sum_{i \in G_k} \frac{x_i}{n_k}. \quad (4)$$

É o centróide do grupo c_k e n_k é o número de pontos em c_k .

Talvez, a aplicação mais óbvia do processo do k-médias, seja tratar o problema da similaridade dos grupos, ou o que é conhecido como agrupamento. O objetivo desta aplicação não é encontrar um grupo único e definitivo, mas sim simplificar e auxiliar o intérprete em obter um bom entendimento quantitativo e qualitativo de uma grande quantidade de dados multivariados.

Regressão por componentes principais (PCR) Além da técnica não-supervisionada do k-médias, duas técnicas supervisionadas também foram utilizadas neste trabalho. A primeira foi a regressão por componentes principais (PCR) (Filzmoser, 2001). A proposta da regressão por componentes principais (PCR) é estimar os valores de uma variável tendo como base uma seleção de componentes realizadas na aplicação da PCA. Existem duas razões fundamentais para se realizar a regressão nas componentes principais ao invés de se utilizar as variáveis originais. Primeiro, as variáveis originais

podem ser altamente correlacionadas o que pode causar uma estimativa não precisa dos coeficientes de regressão. Essa característica pode ser contornada usando as componentes principais (PCs) no lugar das variáveis originais, pois as PCs são não-correlacionadas. Em segundo lugar, a dimensionalidade da regressão é diminuída e mais fácil de ser tratada ao analisar um determinado conjunto de PCs na regressão. Na técnica de PCR, os dados são primeiramente submetidos a uma análise de PCA e os escores são usadas para realizar a predição. A idéia básica é que após a escolha do número de componentes as feições mais importantes da matriz dos dados sejam retidas, assim, é possível realizar uma regressão multilinear de modo a se obter as classes que representam os dados. Essa técnica é robusta, pois se aproxima muito de uma regressão multilinear tradicional. O algoritmo utilizado para a realização da PCR neste trabalho consiste em:

(1) Realizar uma PCA por meio da SVD na matriz dos dados:

$$[X]=[U][S][V]^T \quad (5)$$

(2) Calcular o modelo de calibração utilizando para isso as componentes escolhidas:

$$\begin{aligned} Y &= XB + e \\ Y &= USV^T B + e \end{aligned} \quad (6)$$

(3) Por fim, calcular a matriz de regressão B e então realizar a previsão da propriedade de interesse em novas amostras:

$$B = VS^{-1}U^T \text{ e } Y = XB \quad (7)$$

Onde a matriz X é a matriz que representa os perfis de poços e a matriz Y é a matriz de aprendizagem, ou seja, as porções do poço que foram testemunhadas. Assim, com a técnica de PCR é possível confirmar as litologias analisadas nos testemunhos e prever o tipo de rocha nas porções não testemunhadas.

K-ésimo vizinho mais próximo (KNN) O método não-paramétrico mais comum aplicado a problemas de classificação é o k-ésimo vizinho mais próximo (KNN). O KNN é um método simples, porém muito eficiente de classificação, é muito utilizado em estudos genéticos e em quimiometria, porém sua aplicação dentro do processo de caracterização de reservatórios ainda é nova. Se comparado com outros métodos supervisionados, o KNN pode ser considerado mais eficaz e robusto, o que motiva a sua utilização dentro da geologia de petróleo e principalmente na análise multivariada de perfis geofísicos.

O método KNN é usado para generalizar a informação de um conjunto de dados multivariados. A regra de decisão classifica cada amostra a uma classe pertencente à k-ésima classe mais próxima definida por um conjunto de treino. A proximidade é definida em termos

de uma distância métrica no espaço multi-espectral. O método KNN é altamente apreciado, pois permite prever todas as variáveis medidas ao mesmo tempo.

Supondo que $D = \{p1, \dots, pn\}$ seja um conjunto de treinamento com amostras pi . Para cada amostra um vetor diferente de variáveis m é conhecido. Dado um conjunto de treinamento $D = \{p1, \dots, pn\}$, a distância espectral entre a amostra p a ser classificada e cada amostra pertencente ao conjunto de treinamento é calculada. Para essa amostra desconhecida um vetor de classe m é assinalado; m é obtido, calculando a média ou a moda sobre todos as k-ésimas amostras pi mais próximas e as classes respectivamente.

O método KNN agrupa os k-ésimos vizinhos mais próximos dando a eles votos, a classe que receber mais votos vence e então a amostra é assinalada como sendo dessa classe. Teoricamente, quanto mais vizinhos serem considerados durante a classificação, menor será o erro, porém o tempo computacional também aumenta, por isso deve-se otimizar esse processo de escolha.

RESULTADOS E DISCUSSÕES Ao iniciar um estudo multivariado, o primeiro problema consiste na grande quantidade de variáveis, porém medindo grandezas distintas. No caso da perfilagem de poços, isso não é diferente, uma vez que temos 5 perfis, onde cada um representa uma determinada propriedade, seja vagariedade (DT), densidade (RHOB) ou a porosidade (NPHI). Por meio de uma análise estatística clássica, pode-se observar a dificuldade em se trabalhar com um grande número de variáveis, aqui se entende um grande número de variáveis, qualquer sistema que possua mais de 3 variáveis. Ao se calcular o histograma da distribuição das variáveis originais e construir os gráficos de correlações entre as mesmas, é possível notar que por meio de uma análise estatística convencional, e até mesmo uma análise bivariada é impossível definir grupos que posteriormente serão atribuídos a diferentes tipos de rochas. Na figura 5, observa-se o histograma da distribuição das 5 variáveis estudadas e na figura 6, a matriz de correlação que fornece as relações bivariadas entre as variáveis.

Pode-se observar na figura anterior que de uma maneira geral as variáveis estudadas apresentam uma distribuição normal, as vezes bimodal como é o caso da densidade (RHOB), dos raios gamma (GR) e da resistividade (ILD), esse comportamento indica a presença clara de tipos diferentes de rocha. Já nas demais variáveis como na porosidade neutra (NPHI) e no perfil sônico (DT) é observado um comportamento normal clássico, ou seja, com essas variáveis não é possível distinguir as diferentes litologias presentes nos poços. Nota-se então a dificuldade de se obter uma análise decisiva utilizando apenas uma descrição estatística monovariada como é o caso dos histogramas, isso ocorre pois, mesmo nas variáveis que apresentam um comportamento bimodal, como essa separação não é claramente observada, ocorre uma ambiguidade na interpretação. Desse modo, por meio de técnicas multivariadas, o objetivo é capturar a correta influência de cada uma das variáveis na defini-

Figura 5 - Histogramas mostrando a distribuição de cada uma das variáveis estudadas neste trabalho.

ção da litologia.

Nos gráficos de correlação apresentados na figura anterior, também pode-se observar que em determinadas correlações como é o caso da resistividade (ILD) com os raios gamma (GR), observa-se comportamento distinto, onde é possível realizar uma separação para diferentes tipos de rochas. Mais uma vez observa-se que não é possível somente com uma análise bivariada estimar de maneira satisfatória a litologia presente no poço em estudo.

Desse modo observa-se que só uma análise multivariada é capaz de fornecer os resultados adequados. Como neste caso se trata de variáveis que medem grandezas distintas, antes de se iniciar o processo de análise estatística multivariada é necessário realizar um pré-processamento nos dados. Esse pré-processamento tem como objetivo normalizar as medidas para comparar e calcular as componentes principais. No caso de se trabalhar com perfis de poços, o pré-processamento mais indicado é o auto-escalamento. Por meio desse processo, todos os perfis são usados e normalizados, subtraindo cada amostra do valor médio e dividindo pela soma-tória da variável. Assim, após o pré-processamento dos dados foi utilizada a técnica de PCA, com o objetivo

de reduzir a dimensionalidade dos dados disponíveis e também preparar a informação que servirá como entrada no processo de agrupamento e classificação. Essa etapa foi executada em toda a extensão do poço, a distinção entre as regiões utilizadas como treinamento só será necessária durante o processo de classificação supervisionada, que será apresentado posteriormente.

A primeira etapa na análise de PCA se refere à extração das componentes. Na figura 7 apresenta-se as 5 componentes extraídas dos perfis de poços.

As componentes apresentadas na figura 7, representam cada uma, a combinação linear de todas as 5 variáveis originais utilizadas, observar somente as componentes não nos leva a grandes conclusões. Assim, para se entender como cada variável impacta em cada uma das componentes, uma estratégia comumente utilizada é o cálculo do círculo de correlação. Esse gráfico mostra quais variáveis contribuem para cada uma das componentes calculadas e qual é o grau dessa contribuição. O círculo de correlação para as componentes principais 1 e 2 e para as variáveis é apresentado na figura 8.

Nota-se no círculo de correlação que na primeira componente existe uma maior contribuição das

Figura 6 - Matriz de correlação com os gráficos de correlação entre as 5 variáveis aqui estudadas.

variáveis DT e NPHI, seguida pela variável RHOB, enquanto na componente 2 a maior contribuição se deve a variável ILD. Essa contribuição da variável resistividade (ILD) fica clara também na figura 7, onde a segunda componente apresentada possui feições características da variável original, como por exemplo, a região localizada em torno da profundidade 3050 metros.

Como o objetivo da aplicação da PCA é reduzir a dimensionalidade do conjunto de dados, isso é feito por meio de uma análise que mostra a porcentagem da variância retida em cada uma das componentes extraídas. Na figura 9 é apresentado o gráfico que relaciona a porcentagem da variância com as componentes principais.

É possível observar na figura 9 que as 3 primeiras componentes representam 87,33% da variância dos dados, ou seja, as cinco variáveis originais podem a partir de agora serem analisadas mediante as 3 primeiras componentes, o que garante uma reprodução da variabilidade igual a aproximadamente 90%. Além disso, como mostrado no círculo de correlação, nas componentes 1 e 2 se tem uma grande contribuição das principais variáveis, já as componentes 4 e 5 retem grande parte dos ruídos presentes nos perfis, não contribuindo

para a análise aqui desejada.

Com a extração das componentes realizada é possível então executar o agrupamento e a classificação dos dados de modo a construir o perfil litológico do poço em estudo. Primeiramente aplicou-se a técnica do k-médias. Para isso estabeleceu-se primeiro o número de classes que se deseja obter dos dados, como existem 3 litologias presentes nesse poço o número de classes definidas foi igual a 3. Na figura 10, é possível observar a separação das 3 classes em um gráfico entre a PC1 e a PC2.

Observa-se que as classes ficaram bem separadas com o mínimo de sobreposição entre elas. Com a aceitação das três classes projetou-se as mesmas nas respectivas profundidades obtendo-se então o perfil litológico. A comparação entre a análise sequencial de testemunhos e a classificação realizada pelo algoritmo de K-médias é apresentada na figura 11.

Como era esperado, por meio do algoritmo de k-médias ocorre uma classificação deficitária, principalmente na parte inferior do poço, onde uma camada de siltito é colocada no meio de espessa camada de arenito na porção basal do poço. Pode-se notar também na parte superior que não há correlação entre as camadas

Figura 7 - Componentes principais extraídas dos perfis de poços.

Figura 8 - Círculo de correlação para as 5 variáveis estudadas com relação as componentes principais 1 e 2.

de siltito, arenito e folhelho. Em vista disso, optou-se por aplicar as duas técnicas supervisionadas de classificação. A primeira técnica supervisionada aplicada foi a PCR. Para isso, devido as análises já realizadas, escolheu-se 3 componentes principais para a realização da classificação. O resultado da classificação em comparação com a análise sequencial de testemunhos é apresentado na figura 12.

Diferente da classificação não-supervisionada por k-médias pode-se observar que na parte inferior, a fácies arenito foi integralmente classificada com sucesso, sem a presença de outro tipo de rocha. Da mesma maneira na parte superior, o tipo de rocha siltito (vermelho) foi corretamente definido. Isso ocorre, pois os dados de treinamento condicionam o método a acertar onde existem amostras. O que se pode observar, porém com a aplicação do PCR é que na parte heterogênea do perfil, localizada logo abaixo da camada de siltito, não se consegue reproduzir satisfatoriamente essa heterogeneidade. Embora essa região faça parte do conjunto de treinamento e por apresentar um alto grau de heterogeneidade, o algoritmo não consegue resolver os tipos de rochas presentes.

Visando reproduzir da melhor forma possível as heterogeneidades, optou-se então por realizar a classificação via KNN. A classificação por KNN foi feita

Figura 9 - Análise da variância retida nas componentes principais.

Figura 10 - Projeção das 3 classes definidas pelo algoritmo de k-médias nas duas primeiras componentes principais.

Figura 11 - Comparação entre a análise seqüencial de testemunhos (esquerda) e a classificação utilizando o algoritmo de k-médias (direita).

Figura 12 - Comparação entre a análise seqüencial de testemunhos (esquerda) com a classificação feita por meio da PCR com 3 componentes principais (direita).

Figura 13 - Comparação entre a análise seqüencial de testemunhos (esquerda) e a classificação realizada por meio do KNN com 6 vizinhos (direita).

usando os 6 vizinhos mais próximos. O resultado da classificação comparado com a análise seqüencial de testemunho é apresentado na figura 13.

A figura 13 mostra que como o PCR, no KNN,

a parte inferior do poço ficou bem caracterizada, sem apresentar anomalias no resultado. A camada de siltito (vermelho) na parte superior também foi caracterizada corretamente. Pode-se observar que a heterogeneidade abaixo do siltito, com a aplicação do KNN ficou mais bem representada. A intercalação arenito/folhelho é reproduzida de forma satisfatória.

De modo a conferir os resultados, se pode fazer uma comparação entre as porções dos perfis densidade e porosidade neutrônica, na profundidade aproximada de 3050 m, comparando-o com a análise de testemunho realizada e com a classificação realizada. Como dito anteriormente, esperava-se que a região não amostrada nessa profundidade fosse arenito, pois como mostra o comportamento dos perfis ele é o mesmo para a zona testemunhada e não-testemunhada. Assim, na figura 14, pode-se observar como essa previsão se concretizou com exatidão mostrando que a litologia classificada para a região não-testemunhada consiste verdadeiramente de arenitos.

Para resumir os resultados obtidos com esse trabalho, pode-se comparar a análise de testemunhos do poço com os resultados das classificações realizadas com os diferentes métodos aqui utilizados (Fig. 15).

Na figura 15, observa-se que no caso da aplicação do k-médias, ocorre uma incorreta definição litológica tanto na parte superior como na parte inferior do poço, o que invalida a sua utilização na definição da litologia onde não existem dados de testemunhos. Já com a utilização da regressão por componentes principais, embora a parte inferior do poço seja corretamente caracterizada, em regiões heterogêneas o método não fornece resultados confiáveis. Com a aplicação do k-

Figura 14 - Comparação entre as porções dos perfis densidade e porosidade neutrônica, juntamente com a litologia obtida da análise de testemunho e a litologia obtida com a classificação realizada por meio da técnica KNN e com 6 vizinhos.

Figura 15 - Comparação entre a análise de testemunhos (A) com a classificação obtida por meio da aplicação do K-médias (B), da regressão por componentes principais (C) e pelo k-ésimo vizinho mais próximo (D).

ésimo vizinho mais próximo, pode-se caracterizar corretamente todas as porções do poço em estudo, garantindo assim a confiabilidade na aplicação de tal método nas porções não testemunhadas.

Desse modo, é possível mostrar que, embora o KNN não seja uma técnica aplicada vastamente no processo de caracterização de reservatórios, esta consegue a melhor reprodução das litologias existentes. Com isso, pode-se aplicar essa técnica em todos os poços disponíveis no campo de modo a gerar uma base de dados litológicos que posteriormente podem ser utilizados na modelagem de fácies e na modelagem petrofísica. Assim, a definição litológica, por meio de técnicas estatísticas multivariadas, pode ser considerada uma das primeiras etapas a serem realizadas dentro do processo de caracterização de reservatórios.

CONCLUSÕES A definição litológica é com certeza uma das etapas cruciais do processo de caracterização de reservatórios, porém com as pequenas porções que são testemunhadas e com o grande número de variáveis e amostras disponíveis, somente com o auxílio da estatística multivariada é possível definir de forma consistente os tipos de rochas presentes nos poços.

Para realizar a definição da litologia é necessária, anteriormente a aplicação das técnicas de agrupamento e classificação, uma análise estatística multivariada realizada por meio da análise de componentes principais. A análise de componentes principais tem o objetivo reduzir a dimensionalidade dos dados retendo as componentes que melhor representam a variabilidade presente no conjunto de dados. Mostrou-se nesse trabalho que, para o poço em estudo 3 componentes representam aproximadamente 90% da variância dos dados originais. Essa escolha de 3 componentes é posteriormente recuperada na aplicação de determinados algoritmos de classificação.

Para se realizar a classificação litológica é possível utilizar tanto métodos não-supervisionados como métodos supervisionados. Neste trabalho o método não-supervisionado aplicado foi o k-médias. Pôde-se observar que, como nesse método não se tem um grupo de treinamento, o algoritmo define tipos litológicos onde os mesmos não existem como foi o caso da parte inferior do poço, com isso, não há confiabilidade nas regiões não testemunhadas. Uma maneira de contornar tal desvantagem é aplicar técnicas supervisionadas, onde as porções testemunhadas funcionam como dados

condicionantes ao processo. Entre as técnicas supervisionadas aqui aplicadas, o KNN se mostrou mais adequado que o PCR, pois conseguiu na região de maior heterogeneidade, resolver de forma satisfatória em relação à definição dos tipos de rochas presentes, com isso se tem uma boa confiabilidade nas regiões não testemunhadas, podendo-se adotar esse método nos demais poços do campo.

Com isso, pôde-se mostrar que as técnicas estatísticas multivariadas ao serem aplicadas de forma robusta, podem auxiliar na definição dos tipos de rochas

presentes nos poços. Essa definição pode ser considerada como o ponto de partida para o processo de caracterização de reservatórios, pois, a partir dessa definição, poderá se construir um modelo litológico, como também o preenchimento pelas respectivas propriedades petrofísicas.

Agradecimentos Os autores agradecem a ANP, pela disponibilização dos dados do Campo de Namorado, o que proporciona a realização de trabalhos científicos com dados reais.

Referências

- Bacoccoli G., Morales R.G., Campos O.J. 1980. The Namorado Oil Field: A Major Discovery in the Campos Basin, Brazil. *AAPG Bulletin*, **63**(3):329-338.
- Doveton J.H. 1994. *Geologic log analysis using computer methods*. AAPG computer applications in geology. Tulsa, American Association of Petroleum Geologists, No. 2, 169 p.
- Johann P. 1997. *Inversion Sismostratigraphique et Simulations Stochastiques en 3D: Réservoir Turbiditique, Offshore du Brésil*. Intégration Géologique, Géophysique et Géostatistique. Docteur Thèse. Université Pierre et Marie Curie, Paris, 352 p.
- Jolliffe I.T. 2002. *Principal Components Analysis*. 2nd ed., New York, Springer Series in Statistics, 487 p.
- McQueen J. 1967. Some Methods for Classification and analysis of Multivariate Observations. In: University of Berkeley, Berkeley Symposium on Mathematical Statistics and Probability, 5th, Berkeley, p. 325-350.
- Talaat S. El-Sheikh & Mostafa M.S. 1989. Na Efficient Technique for Lithology Classification. *IEEE Transactions on Geoscience and Remote Sensing*, **27**(5):465-480.
- Webb A. 2003. *Statistical Pattern Recognition*. Second Edition. London, John Wiley and Sons Ltd., 450 p.
- Wong P.M., Choi S., Niu Y. 2002. A Comparison of PCA/ICA for Data Preprocessing in a Geoscience Application. *IEEE Transactions on Geoscience and Remote Sensing*, **40**(5):410-420.

Manuscrito BR 03

Submetido em 19 de dezembro de 2007

Aceito em 20 de abril de 2008