

APLICAÇÃO DA ANÁLISE DE COMPONENTES PRINCIPAIS EM DADOS DE PERFIS

SOME APPLICATIONS OF PRINCIPAL COMPONENT ANALYSIS TO WELL-LOG DATA

José Alberto Bucheb¹ e Hilton Bernard Evans²

RESUMO – A análise de componentes principais é uma técnica estatística multivariada utilizada para se estabelecer a quantidade de informação existente num conjunto de dados. O procedimento de conversão dos dados originais em componentes principais compreende sua padronização e o subsequente cálculo da matriz de correlação, bem como de seus autovetores e autovalores utilizados na obtenção dos coeficientes da transformação e da proporção da variância total associada a cada componente principal. Alguns exemplos da utilização dessa metodologia incluem a determinação de eletrofácies (em conjunto com outras técnicas, como a análise discriminante), a estimativa preliminar de porosidade e a correlação de poços.

(Originais recebidos em 16.09.91).

ABSTRACT – *Principal component analysis is a multivariate statistical technique used to ascertain the effective quantity of information existing in a data set. The procedure for converting original data into principal components entails the standardization of these data and subsequent calculation of the correlation matrix, its eigenvectors and eigenvalues (used to obtain the coefficients of transformation), and the proportion of total variance associated with each principal component. Some applications of this methodology include facies determination (in conjunction with other methods, such as discriminant analysis); preliminary porosity estimation; and well-to-well correlation.*

(Expanded abstract available at the end of the paper).

1 – INTRODUÇÃO

O uso de técnicas estatísticas multivariadas tem tido grande aceitação entre os intérpretes de perfil desde o trabalho pioneiro de Serra e Abbott (1980), que definiram o termo eletrofácies como "o conjunto de respostas de perfis que caracteriza uma camada e permite distingui-la das demais".

Outra contribuição importante foi a de Wolff e Pelissier-Combescure (1982), ao utilizarem as técnicas de análise de componentes principais e de distribuição modal para a caracterização de eletrofácies, seguindo-se os trabalhos de Widdicombe *et al.* (1984) e de De la Cruz e Takizawa (1985), que aplicaram a metodologia para identificação de litologia em poços da Bacia de Powder River (EUA) e do Qatar, respectivamente.

Paralelamente, Delfiner *et al.* (1984) mostraram os resultados obtidos com a construção de um banco de dados petrográficos e as correspondentes respostas de perfis, para a identificação de fácies por meio de

métodos estatísticos, e Wilbert (1984) utilizou a técnica da análise discriminante para classificar conglomerados, arenitos e fácies não-reservatório do Campo de Riachuelo (SE).

Na mesma linha, Busch *et al.* (1985) também fizeram uso da análise discriminante, combinando dados de perfis e de testemunhos da Formação Shublik, North Slope, Alaska, e Berteig *et al.* (1985) propuseram a aplicação de métodos de agrupamento automático associados a técnicas de classificação contextual, exemplificando com dados de poços do Mar do Norte.

Posteriormente, Mohn *et al.* (1987) discutiram o emprego de métodos estatísticos multivariados e de modelos estocásticos na determinação de litofácies através de perfis, Descalzi *et al.* (1988) mostraram o uso da análise de agrupamentos em dados de perfis e de testemunhos para a interpretação de reservatórios de litologia complexa, e Sarzenski e Toledo (1988) utilizaram procedimentos estatísticos para a correlação rocha-perfil nos reservatórios heterogêneos

1 - Setor de Tratamento de Dados Geológicos (SETRAG), Divisão de Operações Geológicas (DIGEO), Departamento de Exploração (DEPEX), Av. República do Chile, 65, Centro, CEP 20035, Rio de Janeiro, RJ, Brasil.

2 - Universidade Federal do Pará, Departamento de Geofísica, Caixa Postal 309, Belém, PA, Brasil.

do Campo de Rio Preto (ES).

Seguindo estes desenvolvimentos, Bucheb (1988) comparou os métodos de análise composicional, de agrupamentos e discriminante para a determinação automática de fácies no Campo de Camorim (SE), Aquino e Souto Filho (1988) discutiram a aplicação da técnica de quadratização de perfis na definição de eletrofácies e apresentaram o resultado obtido com sua aplicação no Campo de Upanema (RN), e Sakurai e Melvin (1988) empregaram modelos estatísticos para a discriminação de fácies mediante dados de perfis e de testemunhos, bem como para estimativas de permeabilidade do Campo de Endicott (EUA).

Mais recentemente, Abrahão (1989) mostrou a utilização da análise discriminante e de métodos de regressão múltipla em dados de perfis para a caracterização de rochas geradoras de hidrocarbonetos na Bacia de Campos, Bucheb e Franco (1989) propuseram a utilização do método das cadeias de Markov modificado por Harper (1984), como elemento subordinado à técnica da análise discriminante na caracterização de fácies por meio dos perfis, e Tetzlaff *et al.* (1989) e Hoffman *et al.* (1989) fizeram uso de métodos estatísticos aplicados iterativamente para a identificação de litologia e estimativa de parâmetros petrofísicos como a porosidade e a permeabilidade. Finalmente, Rodrigues e Queiroz Neto (1991) compararam os métodos das redes neurais e discriminante estatístico para a determinação de litologia, Castro (1991) propôs um algoritmo para o reconhecimento semi-automático de eletrofácies fundamentado na análise de agrupamentos, e Souza Jr. (1991) discutiu a aplicação de técnicas de análise de dados multivariados no estudo de reservatórios.

Importantes também são o trabalho de Moura (1985), no qual estão exemplificadas técnicas de tratamento estatístico multivariado para a análise de dados geoquímicos, e de Landim (1976), que estudou a aplicação da análise de agrupamentos em bioestratigrafia.

Há numerosos textos na literatura especializada que apresentam fundamentos e aplicações de métodos estatísticos multivariados. Em particular, devem ser mencionados os trabalhos de Davis (1973), em que são desenvolvidos os aspectos teóricos e mostrados diversos exemplos relacionados às geociências, e de Doveton (1986), que discute extensamente o uso de várias dessas técnicas em dados de perfis. Não há necessidade, portanto, de um tratamento aprofundado das técnicas multivariadas.

Neste estudo, são mostrados, sucintamente, os princípios teóricos da análise de componentes principais, e destacados alguns exemplos de sua utilização em dados de perfis.

2 – ANÁLISE DE COMPONENTES PRINCIPAIS

A técnica da análise de componentes principais é utilizada, fundamentalmente, para estabelecer a

quantidade efetiva de informações existente num grupo de dados. A diversidade nos registros é descrita em termos de um conjunto de eixos mutuamente perpendiculares, de forma que cada um conterá uma parte da variabilidade total e não possuirá correlação com os demais (Moss e Seheult, 1987). Assim, os componentes principais constituem-se numa forma alternativa de se apresentar os dados originais, possibilitando melhor conhecimento de sua estrutura, sem introduzir informação adicional.

Em alguns casos, quase toda a variabilidade pode ser descrita por um número de eixos menor que o de variáveis originais, permitindo, por conseguinte, que se reduza a dimensão do problema. Entretanto, as variações observadas em eixos de ordens superiores podem corresponder a eventos significativos, não devendo, nestas situações, ser desconsideradas (Serra, 1985).

2.1 – Fundamentos Teóricos

Admitindo-se um conjunto de dados formado por m variáveis e n objetos, tem-se a matriz $A = (a_{ij})$, com $i = 1, \dots, n$ e $j = 1, \dots, m$.

Para cada vetor-coluna de A , obtém-se a média \bar{a}_j e a variância s_j^2 , onde:

$$\bar{a}_j = \frac{1}{n} \sum_{i=1}^n a_{ij}$$

e

$$s_j^2 = \frac{1}{n} \sum_{i=1}^n (a_{ij} - \bar{a}_j)^2$$

A covariância s_{jk}^2 entre os vetores-coluna a_j e a_k de A é dada por:

$$s_{jk}^2 = \frac{1}{n} \sum_{i=1}^n (a_{ij} - \bar{a}_j)(a_{ik} - \bar{a}_k),$$

observando-se que $s_{jk}^2 = s_{kj}^2$ e $s_{jj}^2 = s_j^2$.

As grandezas assim calculadas podem ser representadas sob a forma de matriz de variância-covariância $S = (s_{jk}^2)$, de dimensões $m \times m$, que pode ser reescrita na forma adimensional, resultando na matriz de correlação $R = (r_{jk})$, onde:

$$r_{jk} = \frac{s_{jk}^2}{s_j s_k},$$

com $-1 \leq r_{jk} \leq 1$ e $r_{jj} = 1$, sendo que L ($m \times m$) é a matriz de autovetores de R e $\lambda_1, \dots, \lambda_m$ são os autovalores associados.

Analogamente, a versão adimensional de A é dada pela matriz U (n x m), de tal forma que:

$$u_{ij} = \frac{a_{ij} - \bar{a}_j}{s_j}$$

Os dados de A transformados em componentes principais serão representados pela matriz T (n x m), obtida da equação:

$$T = UL$$

Como R é uma matriz quadrada (m x m), sua decomposição é dada na forma:

$$\Lambda = L^T R L$$

sendo que:

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m) \text{ e } R = U^T U$$

ou seja:

$$\Lambda = L^T U^T U L$$

Como T = UL, então:

$$\Lambda = T^T T$$

Assim, Λ representa a matriz de covariância, e λ_j corresponde à variância de j-ésima coluna (j-ésimo componente principal: t_j) de T, ou seja:

$$\lambda_j = \frac{1}{n} \sum_{i=1}^n (t_{ij} - \bar{t}_j)^2$$

onde:

$$\bar{t}_j = \frac{1}{n} \sum_{i=1}^n t_{ij}$$

Como $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$, o primeiro componente principal t_1 possui a maior variância (λ_1), t_2 apresenta a segunda maior variância (λ_2) e não tem correlação com t_1 , e assim por diante até o menor autovalor.

O desenvolvimento mais detalhado, acompanhado de exemplos de utilização da técnica, é dado por Cooley e Lohnes (1971), Davis (1973), Moss e Seheult (1987) e Elek (1988), dentre outros.

A figura 1 ilustra esquematicamente um exemplo de análise de componentes principais em duas dimensões.

Inicialmente, as curvas dos perfis de densidade (ρ_b) e de porosidade neutrônica (ϕ_{nls}) são padronizadas considerando-se suas respectivas médias e desvios-

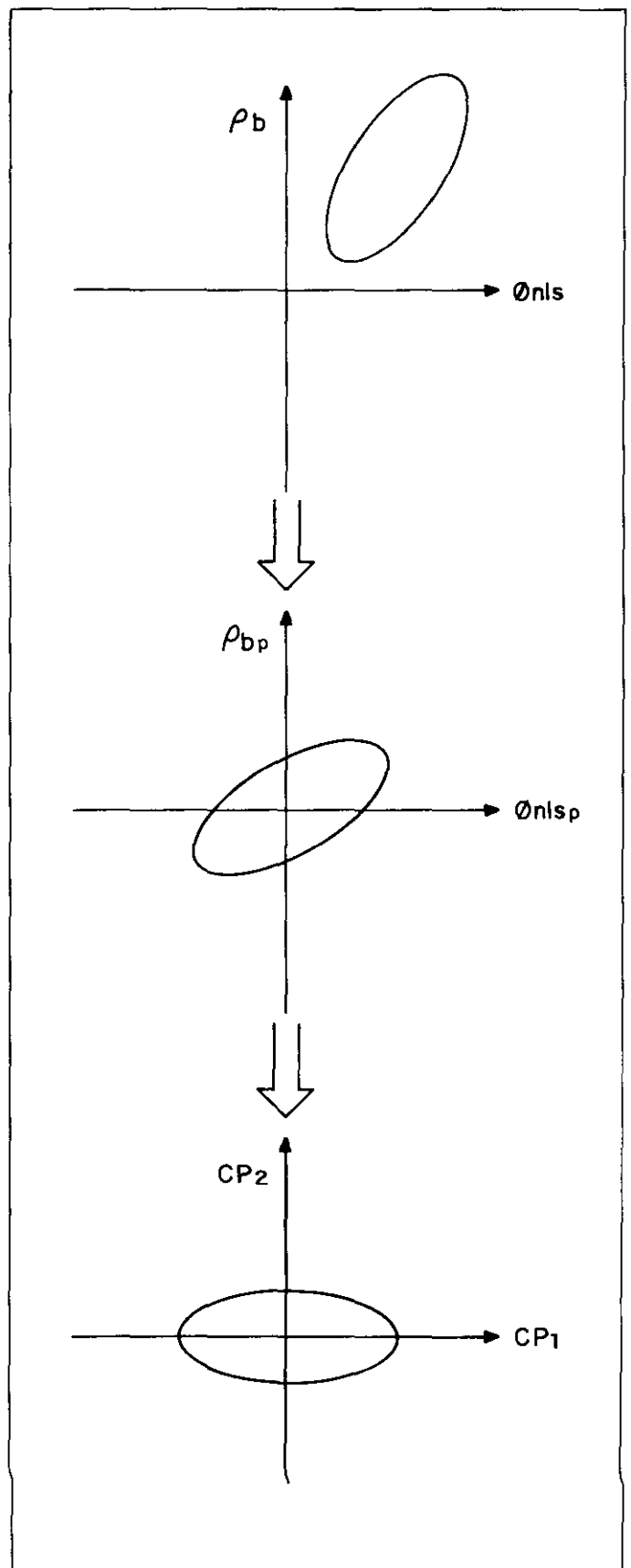


Fig. 1 - Representação esquemática da análise de componentes principais considerando dados de ρ_b e ϕ_{nls} (adaptado de Wolff e Pellissier-Combescur, 1982).

Fig. 1 - Schematic representation of principal component analysis, considering ρ_b and ϕ_{nls} data (adapted from Wolff and Pellissier-Combescur, 1982).

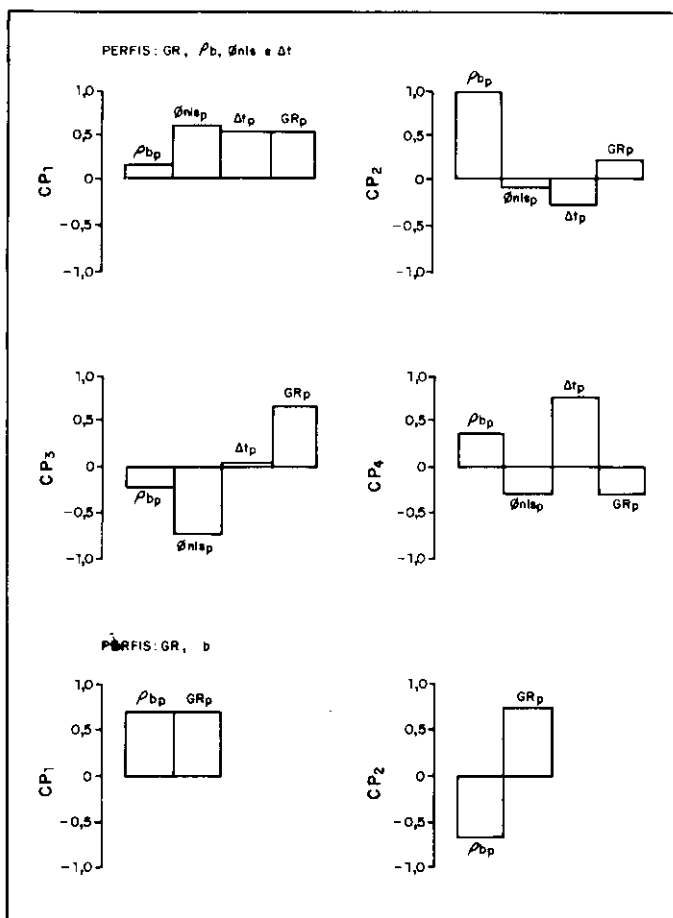


Fig. 2 - Pesos dos componentes principais - Poço CM-A.
Fig. 2 - Principal components, Well CM-A.

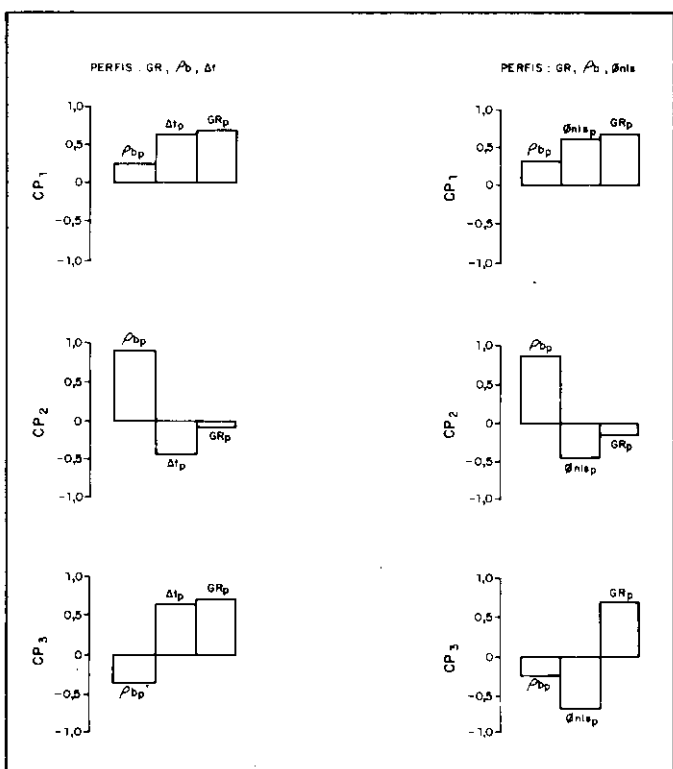


Fig. 3 - Pesos dos componentes principais - Poço CM-A.
Fig. 3 - Principal components, Well CM-A.

padrão. A seguir, no espaço destas curvas padronizadas, dois novos eixos ortogonais que se cruzam no centro de gravidade são definidos, de tal forma que o primeiro (CP1) contenha a maior parte da informação, e o segundo (CP2), a segunda maior quantidade de informação. Dessa forma, as curvas de componentes principais são obtidas por meio da projeção das curvas normalizadas sobre os eixos CP1 e CP2. (Wolff e Pelissier-Combesure, 1982).

2.2 - Descrição da Técnica

A análise de componentes principais é executada em geofísica de poço, considerando-se um conjunto de leituras de perfis formado a partir de m ferramentas com as quais se efetuaram n medições ao longo de uma seção. Esses dados são dispostos numa matriz $A = (a_{ij})$ de dimensões $n \times m$, sendo que a_{ij} é a medida do perfil j tomada na profundidade i , com $i = 1, \dots, n$ e $j = 1, \dots, m$.

Os resultados obtidos para o poço CM-A, usando-se combinações de perfis existentes no Campo de Camorim, são mostrados nas tabelas de I a IV, onde estão representados o desvio-padrão e a média de cada curva (utilizados na padronização dos dados), a matriz de correlação (R) e seus autovalores, a variância total e a variância cumulativa correspondentes a cada componente principal (CP), além da matriz de autovetores de R , que fornece os coeficientes para o cálculo dos componentes principais a partir dos dados padronizados dos perfis (indicados abaixo pelo subscrito p). Assim, para a combinação de curvas $GR/\rho b/\Delta t/\phi_{nls}$, por exemplo, os componentes principais CP_j ($j = 1, \dots, 4$), em cada profundidade ($i = 1, \dots, n$), são dados por:

raios gama (GR)/densidade (ρb)/tempo de trânsito (Δt)/porosidade neutrônica calibrada em matriz (ϕ_{nls}).

$$CP_{1i} = 0,1442\rho_{bpi} + 0,5754\phi_{nlspi} + 0,5696\Delta_{tpi} + 0,5689GR_{pi}$$

$$CP_{2i} = 0,9290\rho_{bpi} - 0,1150\phi_{nlspi} - 0,3009\Delta_{tpi} + 0,1821GR_{pi}$$

$$CP_{3i} = -0,2002\rho_{bpi} - 0,6939\phi_{nlspi} + 0,0638\Delta_{tpi} + 0,6887GR_{pi}$$

$$CP_{4i} = 0,2757\rho_{bpi} - 0,4173\phi_{nlspi} + 0,7622\Delta_{tpi} - 0,4109GR_{pi}$$

A operação em questão equivale a projetar cada conjunto de observações sobre os eixos principais.

Da mesma forma, as figuras 2 e 3 ilustram graficamente a influência de cada curva padronizada nos componentes principais para as combinações de perfis consideradas.

3 - APLICAÇÕES

3.1 - Análise Discriminante

A análise discriminante é uma técnica multivaria-

TABELA I/TABLE I
ANÁLISE DE COMPONENTES PRINCIPAIS
PRINCIPAL COMPONENT ANALYSIS

	P_b	\varnothing_{ns}	Δt	GR
MÉDIA	2,5729	16,4608	73,7334	72,8188
DESVIO - PADRÃO	0,0401	3,5799	4,9985	17,8609
MATRIZ DE CORRELAÇÃO				
	P_{b_p}	\varnothing_{ns_p}	Δt_p	GR _p
P_{b_p}	1,0000	0,1200	-0,0692	0,3530
\varnothing_{ns_p}	0,1200	1,0000	0,8468	0,7384
Δt_p	-0,0692	0,8468	1,0000	0,7612
GR _p	0,3530	0,7384	0,7612	1,0000
	AUTOVALORES	VARIANÇA TOTAL (%)	VARIANÇA TOTAL CUMULATIVA (%)	
CP ₁	2,5984	64,9592	64,9592	
CP ₂	1,0767	26,9180	91,8772	
CP ₃	0,2239	5,5970	97,4742	
CP ₄	0,1010	2,5258	100,0000	
MATRIZ DE AUTOVETORES				
	CP ₁	CP ₂	CP ₃	CP ₄
P_{b_p}	0,1442	0,9290	-0,2002	0,2757
\varnothing_{ns_p}	0,5754	-0,1150	-0,6939	-0,4173
Δt_p	0,5696	-0,3009	0,0638	0,7622
GR _p	0,5689	0,1821	0,6887	-0,4109

TABELA II/TABLE II
ANÁLISE DE COMPONENTES PRINCIPAIS
PRINCIPAL COMPONENT ANALYSIS

	P_b	\varnothing_{ns}	GR
MÉDIA	2,5729	16,4608	72,8188
DESVIO - PADRÃO	0,0401	3,5799	17,8609
MATRIZ DE CORRELAÇÃO			
	P_{b_p}	\varnothing_{ns_p}	GR _p
P_{b_p}	1,0000	0,1200	0,3530
\varnothing_{ns_p}	0,1200	1,0000	0,7384
GR _p	0,3530	0,7384	1,0000
	AUTOVALORES	VARIANÇA TOTAL (%)	VARIANÇA TOTAL CUMULATIVA (%)
CP ₁	1,8696	62,3207	62,3207
CP ₂	0,9074	30,2467	92,5673
CP ₃	0,2230	7,4327	100,0000
MATRIZ DE AUTOVETORES			
	CP ₁	CP ₂	CP ₃
P_{b_p}	0,3649	0,9028	-0,2278
\varnothing_{ns_p}	0,6314	-0,4197	-0,6520
GR _p	0,6843	-0,0941	-0,7231

TABELA III/TABLE III
ANÁLISE DE COMPONENTES PRINCIPAIS
PRINCIPAL COMPONENT ANALYSIS

	P_b	Δt	GR
MÉDIA	2,5729	73,7334	72,8188
DESVIO - PADRÃO	0,0408	4,9985	17,8609
MATRIZ DE CORRELAÇÃO			
	P_{b_p}	Δt_p	GR _p
P_{b_p}	1,0000	-0,0692	0,3530
Δt_p	-0,0692	1,0000	0,7612
GR _p	0,3530	0,7612	1,0000
	AUTOVALORES	VARIANÇA TOTAL (%)	VARIANÇA TOTAL CUMULATIVA (%)
CP ₁	1,8143	60,4780	60,4780
CP ₂	1,0527	35,0885	95,5664
CP ₃	0,1330	4,4335	100,0000
MATRIZ DE AUTOVETORES			
	CP ₁	CP ₂	CP ₃
P_{b_p}	0,2557	0,9073	-0,3338
Δt_p	0,6485	-0,4170	-0,6368
GR _p	0,7170	0,0536	0,6950

TABELA IV/TABLE IV
ANÁLISE DE COMPONENTES PRINCIPAIS
PRINCIPAL COMPONENT ANALYSIS

	P_b	GR	
MÉDIA	2,5729	72,8188	
DESVIO - PADRÃO	0,041	17,8609	
MATRIZ DE CORRELAÇÃO			
	P_{b_p}	GR _p	
P_{b_p}	1,0000	0,3530	
GR _p	0,3530	1,0000	
	AUTO VALORES	VARIANÇA TOTAL (%)	VARIANÇA TOTAL CUMULATIVA (%)
CP ₁	1,3530	67,6492	67,6492
CP ₂	0,6470	32,3508	100,0000
MATRIZ DE AUTOVETORES			
	CP ₁	CP ₂	
P_{b_p}	0,7071	-0,7071	
GR _p	0,7071	0,7071	

da que permite distinguir estatisticamente dois ou mais grupos definidos previamente (Moura, 1985), combinando-se linearmente as variáveis discriminatórias, de modo a maximizar as diferenças entre os conjuntos. Dessa forma, numa seção de poço representativa da área em estudo, as fácies são identificadas, *a priori*, ou nos testemunhos, ou por meio de métodos automáticos de classificação (tais como, a análise de agrupamentos), para a obtenção dos coeficientes a serem aplicados em outros poços não-testemunhados. Se a função discriminante distinguir efetivamente os grupos considerados, será viável o reconhecimento das fácies, com base nas respostas dos perfis, em toda a área investigada.

Para cada combinação de perfis corridos nos poços do Campo de Camorim, realizou-se a análise discriminante utilizando-se os dados originais e os transformados em componentes principais do poço CM-A, escolhido como modelo para este estudo (Buchebe, 1991). Objetivou-se, com isso, verificar qual das duas alternativas contribui mais para a eficiência do método.

Os ajustes obtidos no processamento de cinco poços (percentual da espessura total classificada de acordo com a descrição dos testemunhos (Nascimento *et al.* 1981)) são mostrados na tabela V. De maneira geral, a análise discriminante aplicada sobre os dados transformados em componentes principais produz melhores resultados, razão pela qual adotou-se tal procedimento para a determinação das fácies neste estudo.

3.2 – Estimativa Preliminar de Porosidade

A estimativa preliminar de porosidade por meio da análise de componentes principais foi demonstrada por Elek (1990), a partir da constatação de que as elongações do primeiro e do segundo componentes principais se devem às variações na porosidade e na litologia, respectivamente. A metodologia consiste, basicamente, nos seguintes passos:

- seleção dos perfis apropriados;
- transformação dos dados originais em componentes principais;
- exame da relação entre a porosidade efetiva, ϕ_e (calculada segundo os métodos tradicionais, em pelo menos dois pontos) e o primeiro componente principal através de gráficos e análise de regressão; e
- estimativa da porosidade por meio da utilização dos coeficientes obtidos na regressão.

Essa metodologia pode ser empregada quando se deseja estabelecer uma estimativa preliminar de porosidade, como, por exemplo, na análise dos perfis no campo. Assim, calculando o valor de ϕ_e em apenas alguns pontos, o intérprete poderá obter o perfil de porosidade para toda a seção de interesse.

O método foi testado inicialmente no poço CM-A, utilizando-se os perfis neutrônico e densidade. Nas figuras 4 e 5, são mostrados os gráficos $CP_1 \times \phi_e$

TABELA V/TABLE V
AJUSTES OBTIDOS NA ANÁLISE DISCRIMINANTE
ADJUSTMENTS OBTAINED THROUGH
DISCRIMINANT ANALYSIS

POÇO	ORIGINAL	A C P	DIFERENÇA	
GR/P _b /GMS/DT	CM-21 A	69%	69%	0%
	CM-39 B	83%	82%	-1%
	CM-47 C	76%	82%	+6%
GR/P _b /DT	CM-21 A	71%	73%	+2%
	CM-39 B	77%	73%	-4%
	CM-47 C	75%	81%	+6%
GR/P _b /GMS	CM-21 A	71%	69%	-2%
	CM-39 B	77%	80%	+3%
	CM-47 C	81%	84%	+3%
GR/P _b	CM-06 D	68%	80%	+12%
	CM-10 E	61%	65%	+4%
	CM-21 A	67%	72%	+5%
	CM-39 B	78%	74%	-4%
	CM-47 C	75%	77%	+2%

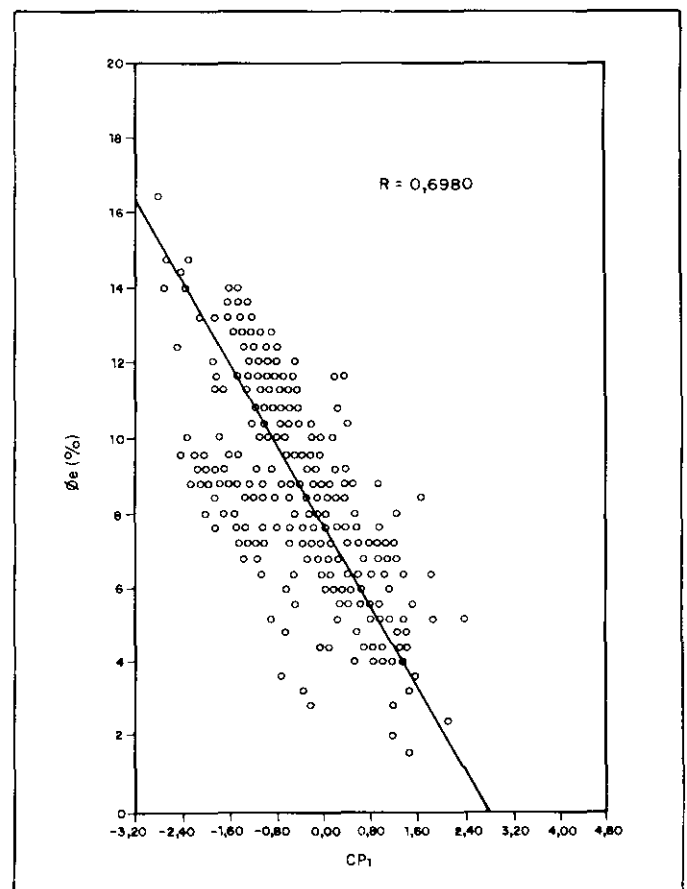


Fig. 4 - Primeiro componente principal versus ϕ_e . Poço CM-A.
Fig. 4 - First principal component versus ϕ_e . Well CM-A.

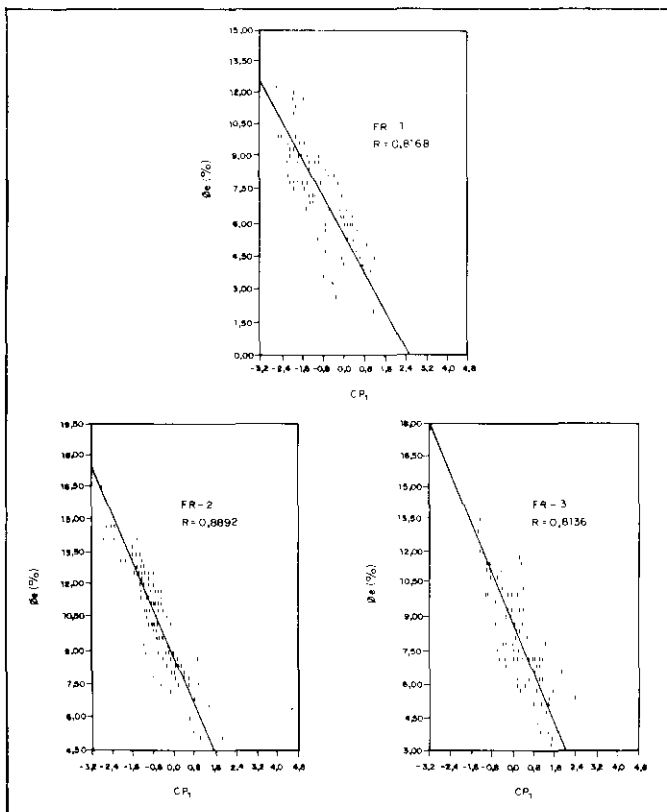


Fig. 5 - Primeiro componente principal versus ϕ_e para as fácies em separado. Poço CM-A.

Fig. 5 - First principal component versus ϕ_e for facies separately, Well CM-A.

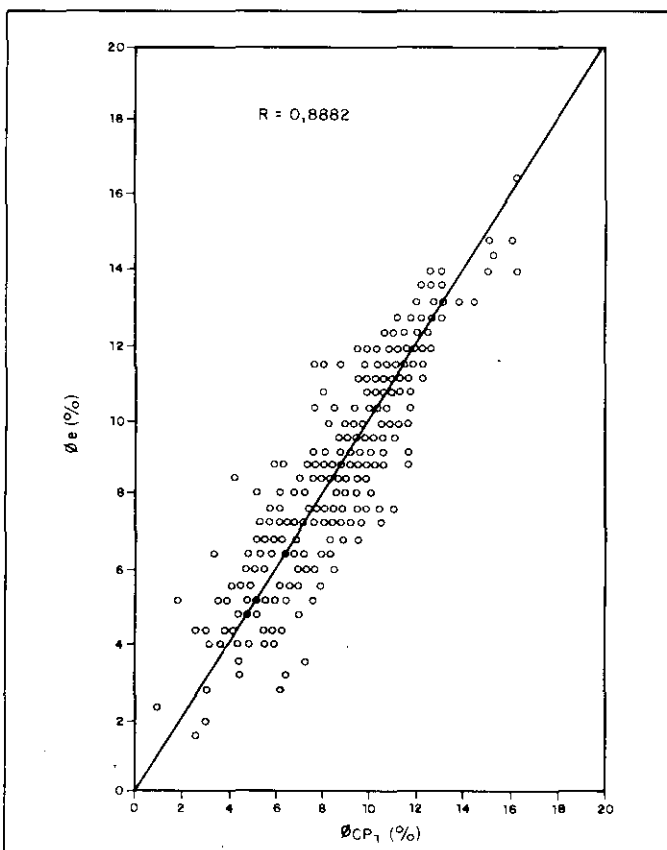


Fig. 6 - ϕ_{CP1} versus ϕ_e . Poço CM-A.

Fig. 6 - ϕ_{CP1} versus ϕ_e . Well CM-A.

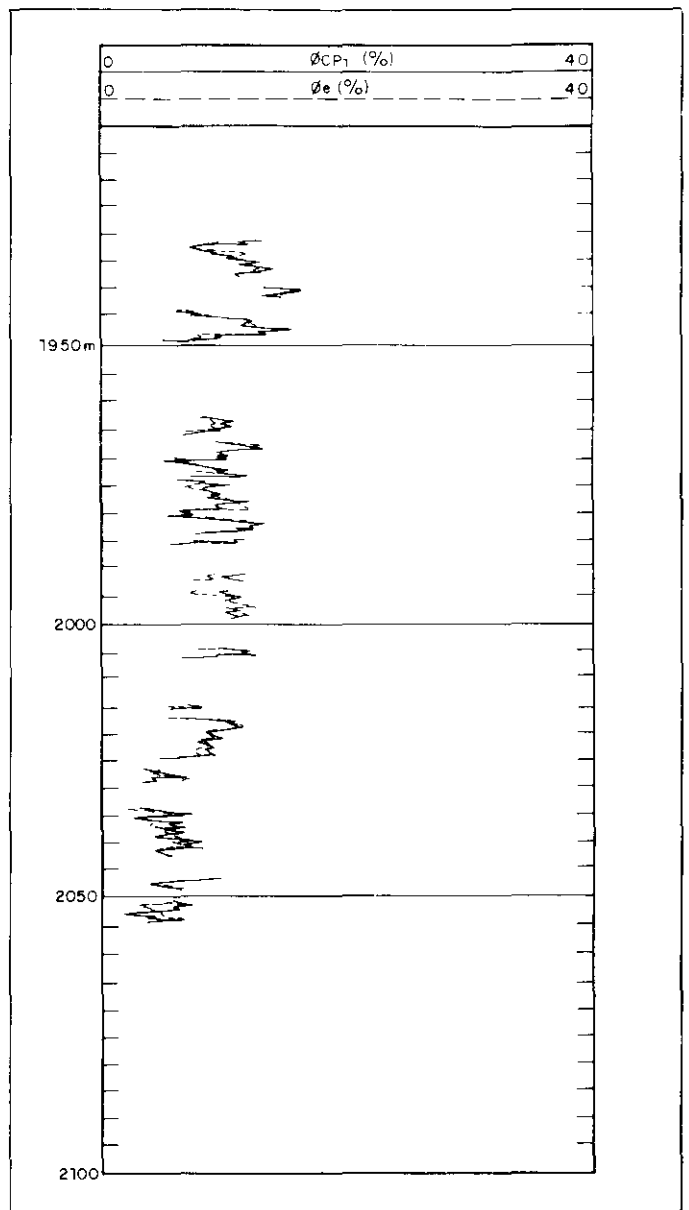


Fig. 7 - Comparação ϕ_e versus ϕ_{CP1} . Poço CM-A.

Fig. 7 - Comparison ϕ_e versus ϕ_{CP1} . Well CM-A.

para toda a seção de interesse e para as fácies-reservatório em separado, onde se observam correlações significativamente melhores.

Assim, por meio de regressão linear, obtém-se:

- a) FR-1 $\phi_e = -2,19CP_1 + 5,68$ e $R = 0,8168$,
- b) FR-2 $\phi_e = -2,68CP_1 + 8,80$ e $R = 0,8892$,
- c) FR-3 $\phi_e = -2,88CP_1 + 8,74$ e $R = 0,8136$.

Estas equações são utilizadas para a estimativa preliminar de porosidade a partir do primeiro componente principal (ϕ_{CP1}). Nas figuras 6 e 7, observa-se o bom ajuste obtido entre ϕ_e e ϕ_{CP1} .

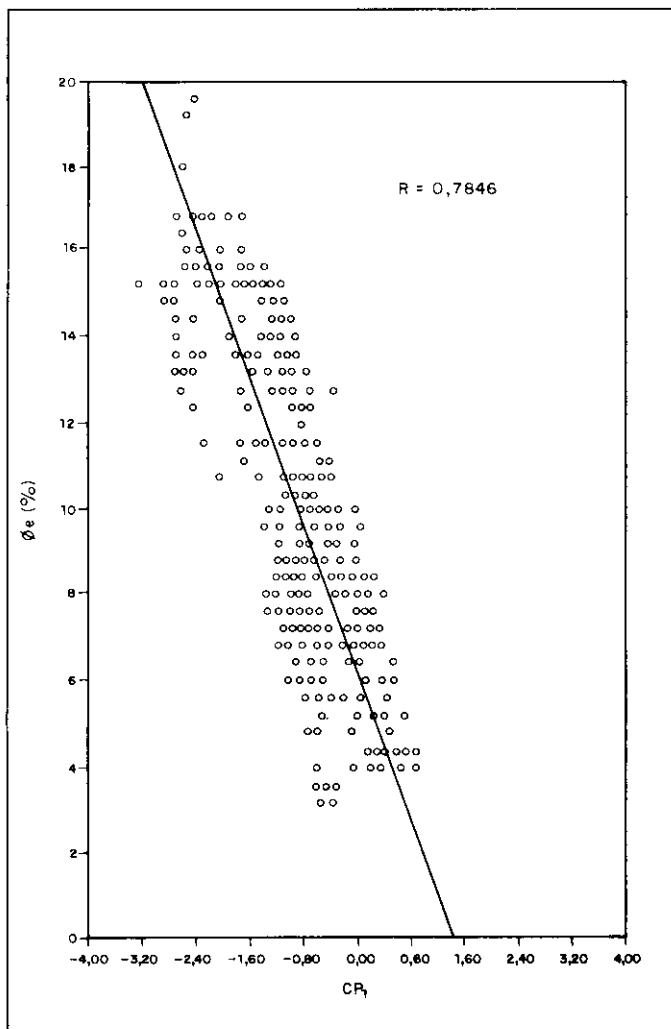


Fig. 8 - Primeiro componente principal versus ϕ_e . Poço CM-B.
Fig. 8 - First principal component versus ϕ_e . Well CM-B.

O mesmo procedimento foi aplicado ao poço CM-B (figs. de 8 a 11), e da mesma forma os resultados foram satisfatórios. Neste caso, as equações resultantes foram:

a) FR-1 $\phi_e = -3,77CP_1 + 4,43$ e $R = 0,9128$,

b) FR-2 $\phi_e = -3,87CP_1 + 7,57$ e $R = 0,8793$,

c) FR-3 $\phi_e = -4,90CP_1 + 7,96$ e $R = 0,9524$.

A eficiência da técnica pode ser verificada pela regressão $\phi_{CP_1} \times \phi_e$, ou seja, quanto mais próximos de 1 e de 45° forem o coeficiente de correlação (R) e a inclinação da reta de ajustamento (θ), respectivamente, melhor será a estimativa da porosidade efetiva. Nos dois exemplos mostrados obtiveram-se:

a) CM-A $R = 0,8882$ e $\theta = 45^\circ$,

b) CM-B $R = 0,9377$ e $\theta = 54^\circ$.

3.3 – Correlação de Poços

Normalmente, a correlação de poços se faz por meio de uma ou duas curvas (tipicamente, resistividade e/ou raios gama). No caso mais geral, este procedimento demonstra ser eficiente, atendendo às necessidades da interpretação. Por outro lado, a técnica da análise de componentes principais possibilita a utilização conjunta da informação disponível, mediante a obtenção da curva do primeiro componente principal.

Dessa forma, a tarefa de reconhecimento de padrões, executada no processo de correlação poço a poço, pode ser efetuada com uma única curva que contém a maior parte da variância dos dados originais (Elek, 1988).

Um exemplo do emprego desta técnica é mostrado na figura 12.

3.4 – Outras Aplicações

Além dos exemplos aqui mostrados, esta técnica tem sido utilizada em dados de perfil com outros

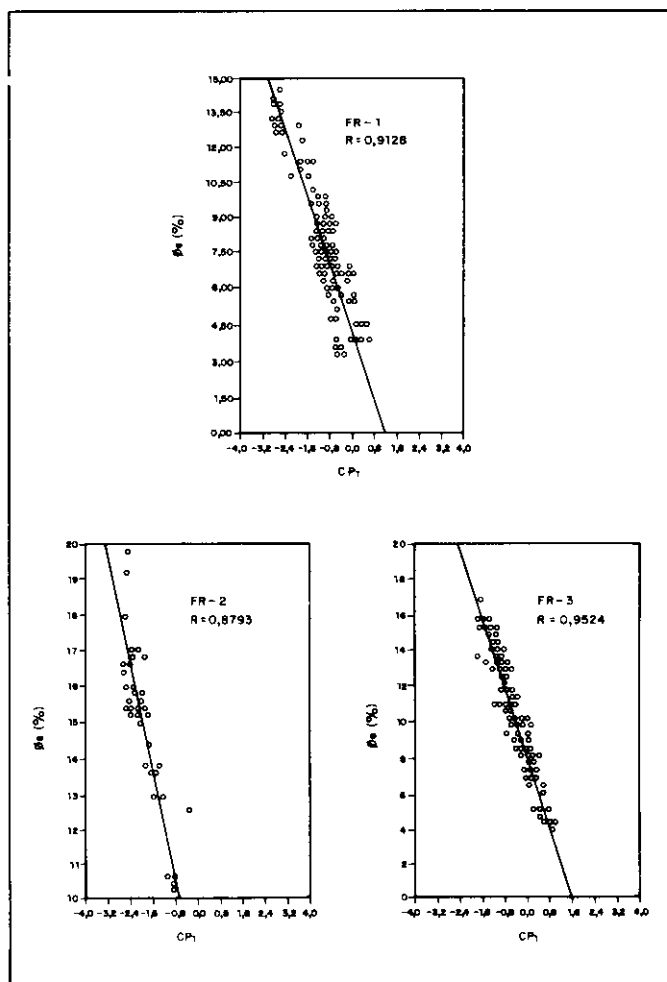


Fig. 9 - Primeiro componente principal versus ϕ_e para as fácies em separado. Poço CM-B.

Fig. 9 - First principal component versus ϕ_e for facies separately. Well CM-B.

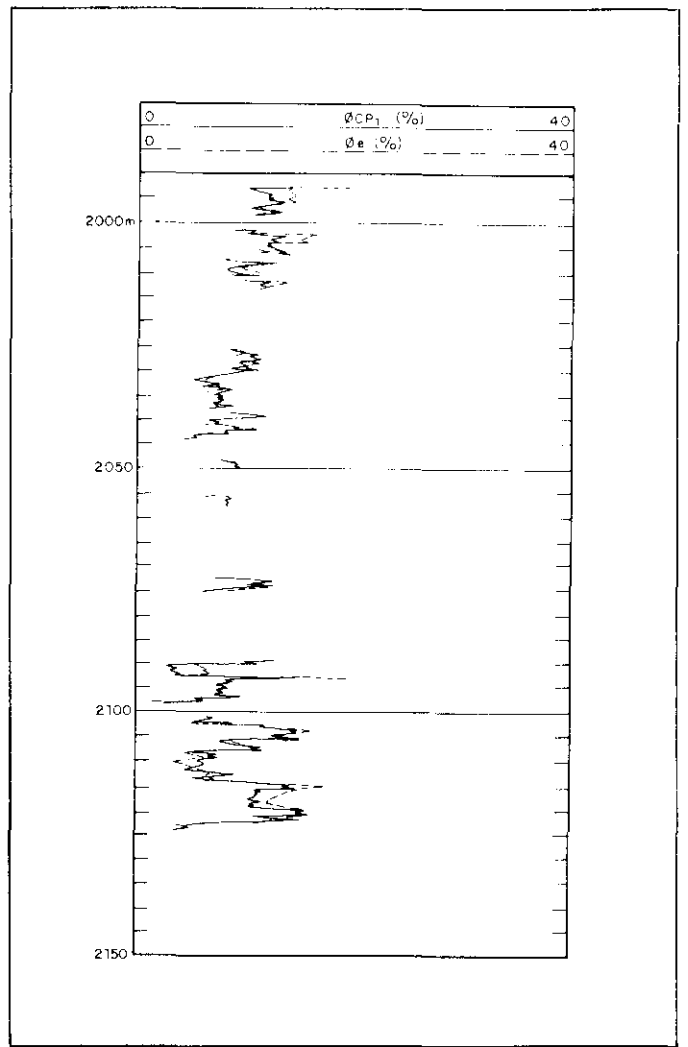
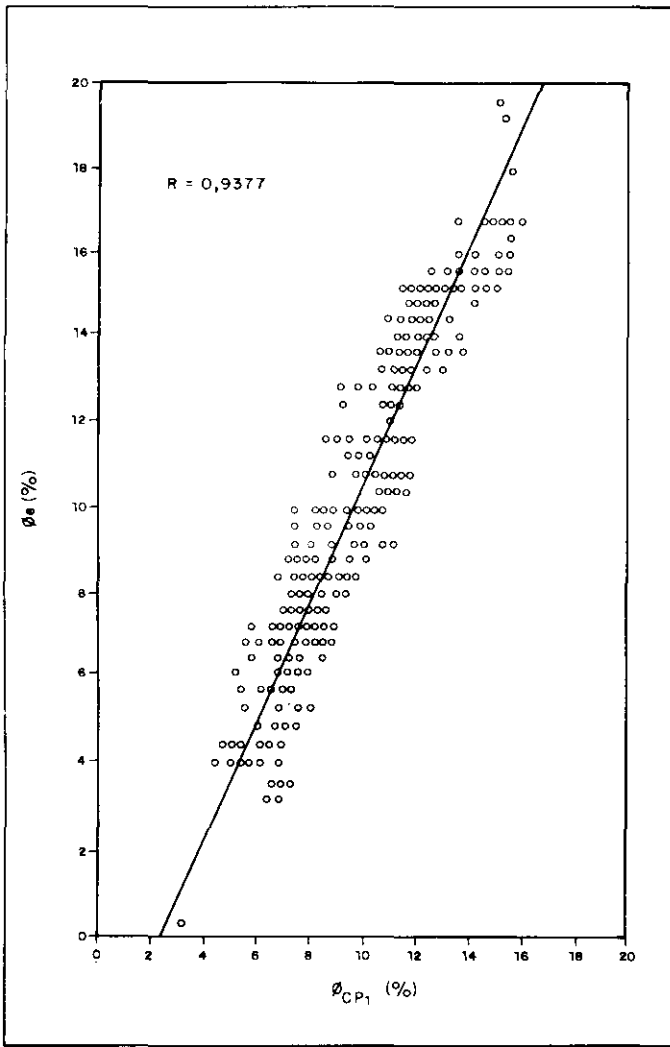


Fig. 10 - ϕ_{CP_1} versus ϕ_e . Poço CM-B.
 Fig. 10 - ϕ_{CP_1} versus ϕ_e . Well CM-B.

Fig. 11 - Comparação ϕ_e versus ϕ_{CP_1} . Poço CM-B.
 Fig. 11 - Comparison ϕ_e versus ϕ_{CP_1} . Well CM-B.

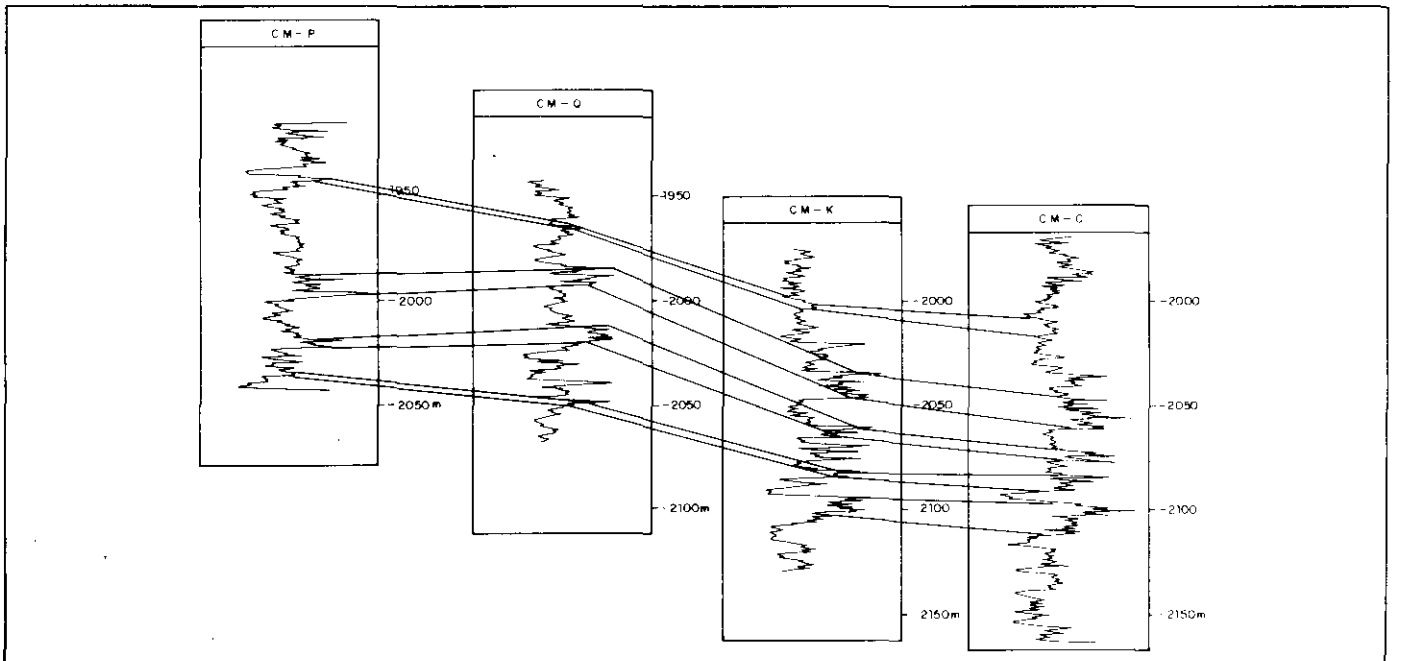


Fig. 12 - Correlação de poços através do primeiro componente principal.
 Fig. 12 - Well-to-well correlation using first principal component.

objetivos, destacando-se os trabalhos de Elek (1988), que desenvolveu um algoritmo de quadratização de perfis, fazendo uso dos componentes principais, visando à definição dos limites e dos valores das respostas dos perfis de uma dada camada, e de Elek (1990), para o cálculo do volume de argila (V_{cl}), executado de maneira análoga à ilustrada neste estudo para a estimativa preliminar de porosidade.

Uma discussão acerca do significado físico e das possibilidades da utilização da análise de componentes principais em dados petrofísicos é efetuada por Moss e Seheult (1988).

4 – CONCLUSÃO

A rápida expansão dos recursos de processamento de dados tem favorecido a utilização de métodos quantitativos, permitindo a otimização do manuseio do expressivo acervo de dados acumulados no estudo das bacias sedimentares brasileiras.

Ao lado das demais técnicas estatísticas multivariadas, a análise de componentes principais constitui uma ferramenta auxiliar que pode ser utilizada como complemento ou alternativa aos métodos tradicionais de avaliação de formações no processamento e na interpretação de dados de perfis.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAHÃO, D. Well-log evaluation of lacustrine source rocks of the Lagoa Feia Formation, Lower Cretaceous, Campos Basin, offshore Brazil. In: ANNUAL LOGGING SYMPOSIUM, 30, 1989. Denver. *Papers*. . . London: Society of Professional Well Log Analysts, 1989. Paper I.
- AQUINO, P. C. A., SOUTO FILHO, J. D. Quadratização de perfis e sua aplicação na definição de eletrofácies no Campo de Upanema-RN. In: SEMINÁRIO TÉCNICO DE OPERAÇÕES GEOLÓGICAS, 2, 1988. Salvador. *Anais*. . . Salvador: PETROBRÁS/DEPEX/DIRGEO, 1988. p. 349-358.
- BERTEIG, V., HELGELAND, J., MOHN, E. Lithofacies prediction from well data. In: ANNUAL LOGGING SYMPOSIUM, 26, 1985. Dallas. *Transactions*. . . London: Society of Professional Well Log Analysts, 1985. Paper TT.
- BUCHÉB, J. A. *Aplicação de tratamento estatístico multivariante em dados de poços da Bacia de Sergipe/Alagoas*. Belém: Universidade Federal do Pará, 1991. 136 p. Dissertação-Mestrado.
- BUCHÉB, J. A. Determinação automática de fácies para o Campo de Camorim. In: SEMINÁRIO TÉCNICO DE OPERAÇÕES GEOLÓGICAS, 2, 1988. Salvador. *Anais*. . . Salvador: PETROBRÁS/DEPEX/DIRGEO, 1988. p. 325-335.
- BUCHÉB, J. A., FRANCO, C. A. Estudo de seqüência de fácies-reservatório no Campo de Camorim. *Testemunho*, Rio de Janeiro, v. 2, n. 6, p. 5-8, set. 1989.
- BUSCH, J. M., FORTNEY, W. G., BERRY, L. N. Determination of lithology from well logs by statistical analysis. In: ANNUAL FALL TECHNICAL CONFERENCE AND EXHIBITION OF THE SOCIETY OF PETROLEUM ENGINEERS OF AIME, 60, 1985. New Orleans. *Papers*. . . Dallas: Society of Petroleum Engineers, 1985. (Special Paper, 14301).
- CASTRO, W. B. M. Um algoritmo para reconhecimento semi-automático de eletrofácies em perfis de poços. In: SEMINÁRIO TÉCNICO DE OPERAÇÕES GEOLÓGICAS, 3, 1991. Cabo Frio. *Anais*. . . Rio de Janeiro: PETROBRÁS/DEPEX/DIRGEO, 1991. v. 1, p. 183-190.
- COOLEY, W. W., LOHNES, P. R. *Multivariate data analysis*. New York: J. Wiley, 1971. 364 p.
- DAVIS, J. C. *Statistics and data analysis in geology*. New York: J. Wiley, 1973. 550 p.
- DE LA CRUZ, M. A., TAKIZAWA, H. Automatic facies analysis in the Arab Formation, El Bunduq field, offshore Abu Dhabi/Qatar. In: ANNUAL FALL TECHNICAL CONFERENCE AND EXHIBITION OF THE SOCIETY OF PETROLEUM ENGINEERS OF AIME, 60, 1985. New Orleans. *Papers*. . . Dallas: Society of Petroleum Engineers, 1985. (Special Paper, 1480).
- DELFINER, P. C., PEYRET, O., SERRA, O. Automatic determination of lithology from well logs. In: ANNUAL FALL TECHNICAL CONFERENCE AND EXHIBITION OF THE SOCIETY OF PETROLEUM ENGINEERS OF AIME, 59, 1984. New Orleans. *Papers*. . . London: Society of Petroleum Engineers, 1984. (Special Paper, 13290).
- DESCALZI, C., ROGNONI, A., CIGNI, M. Sinergetic log and core data treatment through cluster analysis: a methodology to improve reservoir description. In: INTERNATIONAL MEETING ON PETROLEUM ENGINEERING, 1988. Tianjin. *Papers*. . . Richardson: Society of Petroleum Engineers, 1988. (Special Paper, 17637).
- DOVETON, J. H. *Log analysis of subsurface geology*. New York: J. Wiley, 1986. 273 p.
- ELEK, I. Fast porosity estimation by principal component analysis. *Geobyte*, Tulsa, v. 5, n. 3, p. 25-34, 1990.
- ELEK, I. Some application of principal component analysis: well-to-well correlation, zonation. *Geobyte*, Tulsa, v. 3, n. 2, p. 25-31, 1988.
- HAPER, C. W. Improved methods of facies sequence analysis. In: FACIES Models, 2. ed., Ottawa: Geological Society of Canada, 1984. p. 11-13.
- HOFFMAN, L. J. B., HOOPERBRIDGE, P. J., LOMAS, A. T. Logix, a knowledge-based system for petrophysical evaluation. In: LOG ANALYSIS SOFTWARE EVALUATION AND REVIEW SYMPOSIUM, 1989. London. *Anais*. . . London: Society of Professional Well Log Analysts London Chapter, 1989. 13 p.
- LANDIM, P. M. B. Aplicação da análise de agrupamentos (cluster analysis) em Bioestratigrafia. In: CONGRESSO BRASILEIRO DE GEOLOGIA, 29, 1976. Ouro Preto. *Anais*. . . Ouro Preto: Sociedade Brasileira de Geologia, 1976.
- MOHN, E., BERTEIG, V., HELGELAND, J. A review of statistical approaches to lithofacies determination from well data. In: NORTH SEA OIL AND GAS RESERVOIRS SEMINAR, 1987. Trondheim. *Anais*. . . Trondheim: The Norwegian Institute of Technology, 1987. p. 301-309.
- MOSS, B., SEHEULT, A. Does principal components analysis have a role in the interpretation of petrophysical data? In: ANNUAL LOGGING SYMPOSIUM, 28, 1987. London. *Transactions*. . . London: Society of Professional Well Log Analysts, 1989. Paper T.
- MOURA, C. A. V. Aplicação de tratamento estatístico multivariante em dados geoquímicos de solo no mapeamento geológico na província de Carajás (Alvo 2 – Corpo 4). *Revista Brasileira de Geociências*, São Paulo, v. 15, n. 3,

- p. 241-248, 1985.
- NASCIMENTO, O. S. (Coord.). *Projeto Camorim*. Rio de Janeiro: PETROBRÁS/DEPEX/DIGED, 1981. 152 p. Relatório Interno.
- RODRIGUES, F. S., QUEIROZ NETO, I. A. Aplicação de inteligência artificial na identificação de eletrofácies, Redes Neurais x Discriminante Estatístico. In: SEMINÁRIO TÉCNICO DE OPERAÇÕES GEOLÓGICAS, 3, 1991. Cabo Frio. *Anais...* Rio de Janeiro: PETROBRÁS/DEPEX/DIRGEO, 1991, v. 1, p. 131-142.
- SAKURAI, S., MELVIN, J. Facies discrimination and permeability estimation from well logs for the Endicote field. In: ANNUAL LOGGING SYMPOSIUM, 29, 1988. London. *Papers...* London: Society of Professional Well Log Analysts, 1988. Paper F.
- SARZENSKI, D. J., TOLEDO, J. B. Correlação rocha-perfil em reservatórios heterogêneos. In: SEMINÁRIO TÉCNICO DE OPERAÇÕES GEOLÓGICAS, 2, 1988. Salvador. *Anais...* Salvador: PETROBRÁS / DEPEX / DIRGEO, 1988. v. 1, p. 59-70.
- SERRA, O. *Sedimentary environments from wireline logs*. [s.l.]: Schlumberger, 1985. 211 p. il.
- SERRA, O., ABBOTT, H. T. The contribution of logging data to sedimentology and stratigraphy. In: ANNUAL FALL TECHNICAL CONFERENCE AND EXHIBITION OF THE SOCIETY OF PETROLEUM ENGINEERS, 55, 1980. Dallas. *Papers...* Dallas: Society of Petroleum Engineers, 1985. (Special Paper, 9270).
- SOUZA JR., O. G. Análise de dados multivariados, uma eficiente ferramenta para descrição e caracterização de reservatórios. In: SEMINÁRIO TÉCNICO DE OPERAÇÕES GEOLÓGICAS, 3, 1991. Cabo Frio. *Anais...* Rio de Janeiro: PETROBRÁS/DEPEX/DIRGEO, 1991, v. 1, p. 121-130.
- TETZLAFF, D. M., RODRIGUEZ, E., ANDERSON, R. L. Estimating facies and parameters from integrated well data. In: LOG ANALYSIS SOFTWARE EVALUATION AND REVIEW SYMPOSIUM, 1989. London. *Anais...* London: Society of Professional Well Log Analysts London Chapter, 1989. 22 p.
- WIDDICOMBE, R., NOON, P., BEST, D. L. Multiwell facies evaluation, Hartzog Draw field, Powder River Basin, Wyoming. In: ANNUAL LOGGING SYMPOSIUM, 25, 1984. New Orleans. *Papers...* London: Society of Professional Well Log Analysts, 1984. Paper. BBB.
- WILBERT, A. *Petrologia, análise eletrofaciológica e características petrofísicas do reservatório Carmópolis no Campo de Riachuelo, Sergipe, Brasil*. Ouro Preto: Universidade Federal de Ouro Preto, 1984. 259 p. Tese-Mestrado.
- WOLFF, M., PELISSIER-COMBESCURE, J. Faciolog-Automatic electrofacies determination. In: ANNUAL LOGGING SYMPOSIUM, 23, 1982. Corpus-Christi. *Papers...* London: Society of Professional Well Log Analysts, 1989. Paper FF.

EXPANDED ABSTRACT

Principal component analysis is basically used to establish the effective quantity of information existing in a data set. The diversity present in the data is described in terms of a set of mutually perpendicular axes, where each contains part of the total variability and each is not correlated with the others (Moss and Seheult, 1987). The principal components thus constitute an alternative form of the original data — one which provides better knowledge of their structure without introducing additional information. In some cases, almost all of the variability can be described by a number of axes fewer than the original variables, thereby permitting a reduction in the dimension of the problem. However, the variation observed in axes of higher orders may correspond to significant events, and in such situations these axes must be considered (Serra, 1985).

To apply principal component analysis in borehole geophysics, consider a set of well-log readings achieved using m tools, with which n measurements are obtained in a borehole interval. These data constitute a matrix A ($n \times m$), where a_{ij} represents the reading of the j -th log taken at depth i , with $i = 1, \dots, n$ and $j = 1, \dots, m$. Some examples of the use of this technique include facies determination procedures (in conjunction with other methods, such as discriminant analysis); fast porosity estimation; and well-to-well correlation, all dealt with in the present article.

For each combination of logs run in the Camorim field (offshore state of Sergipe, Brazil), discriminant analysis was performed using both the original data and the data transformed into principal components. The

objective was to determine which of the two data sets would yield best results using this statistical method. When applied to transformed data, discriminant analysis generally produces better results in the calibration of rock-log data.

Preliminary porosity estimation through principal component analysis was demonstrated by Elek (1990), based on his observation that the elongations of the first and second principal components derive from variations in porosity and lithology, respectively. The methodology consists of the following steps:

- a) selection of the appropriate logs;*
- b) transformation of original data into principal components;*
- c) examination of the relationship between effective porosity ϕ_e (calculated using traditional methods with at least two points) and the first principal component (using graphs and regression analysis);*
- d) examination of porosity using the coefficients obtained in regression analysis.*

This methodology can be employed when one wishes to establish a preliminary porosity examination — for a quick-look examination, for instance. With this approach, the interpreter can define the porosity log for the whole section of interest by calculating the ϕ_e value at only a few points.

Well-to-well correlation is usually carried out using one or two curves (typically, resistivity and/or gamma-ray), a procedure which is generally sufficient for the purposes of the interpretation process. On the other hand, if one uses the first principal component as the correlation element, almost all of the available

information is utilized. The pattern-recognition task undertaken in the well-to-well correlation process may thus be performed using only that curve which contains the most significant proportion of the total variance in the data (Elek, 1988).

In addition to the examples described here, this technique has been employed with other objectives. Elek (1988 and 1989) developed an algorithm to square

log data by relying on principal components to define the limits and log values for a given layer and in the calculation of clay volume, using procedures similar to the fast porosity estimation shown here. Moss and Seheult (1987) discuss the physical meaning and possible application of principal components in petrophysical data handling.