

# **ASPECTOS PRÁTICOS RELATIVOS AO PROCEDIMENTO DE GERAÇÃO DE CURVAS SINTÉTICAS DE PERFIS POR MEIO DE REGRESSÃO**

## **PRACTICAL ASPECTS RELATED TO THE PROCEDURE FOR GENERATING SYNTHETIC LOG CURVES BY REGRESSION**

## **ASPECTOS PRÁCTICOS RELATIVOS AL PROCEDIMIENTO DE GENERACIÓN DE CURVAS SINTÉTICAS DE PERFILES POR MEDIO DE REGRESIÓN**

*José Alberto Bucheb<sup>1</sup>  
Fernando da Silva Rodrigues<sup>1</sup>*

### **RESUMO**

O emprego de métodos de regressão para estimar propriedades petrofísicas, utilizando as curvas de perfis como variáveis independentes, é um procedimento de rotina em diversos segmentos da área de E&P da indústria do petróleo. Como estas ferramentas são projetadas para registrar variações na porosidade, tipo de fluido e litologia, pode-se admitir que qualquer curva de perfil pode ser considerada função de qualquer outra, medida para as mesmas amostras. Para garantir a qualidade dos resultados, alguns testes específicos devem ser realizados paralelamente à análise estatística dos resultados, incluindo a verificação do coeficiente de determinação, a significância da correlação e da regressão, bem como a ocorrência de pontos discrepantes. Utilizaram-se perfis de poços da plataforma continental brasileira para ilustrar algumas das possíveis aplicações de curvas sintéticas do perfil sônico, obtidas por meio de regressão, tais como nos cálculos de porosidade e do tempo de trânsito integrado bem como na geração de sismogramas sintéticos. Propõe-se um roteiro para validar o processo de estimativa e verificar quantitativamente as curvas calculadas.

(Originais recebidos em 04.04.96.)

### **ABSTRACT**

*The use of regression methods to estimate petrophysical properties using the log curves as independent variables, is a routine procedure in many E&P sectors of the oil industry. Since the logging tools are designed to record variations in porosity, fluid type and lithology, it can be supposed that any log curve can be considered a function of any other, measured for the same samples. To guarantee result quality, some specific tests must be carried out in parallel with the statistical analysis of the results, including the checking of the goodness of fit coefficient, the significance of the correlation and regression, as well as the occurrence of outliers. Well logs from the Brazilian continental shelf were used to illustrate some of the possible applications of synthetic curves of sonic logs, obtained by regression, such as in porosity and integrated transit time calculations and for the generation of synthetic seismograms. A routine is proposed with the aim of validating the estimation process and quantitatively checking the calculated curves.*

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

<sup>1</sup> PETROBRAS-E&P/GEREX/GESUB - Av. República do Chile, 65, Centro, CEP 20035-900, Rio de Janeiro, RJ, Brasil.  
e-mail: bucheb@ep.petrobras.com.br; ssr@ep.petrobras.com.br

*El empleo de métodos de regresión para estimar propiedades petrofísicas medidas por las herramientas de perfilage en intervalos no perfilados, utilizando las curvas de perfiles como variables independientes, es un procedimiento de rutina en diversos segmentos del área de E&P de la industria del petróleo. Como estas herramientas son proyectadas para registrar variaciones en la porosidad, tipo de fluido y litología, se puede admitir que cualquier curva de perfil puede ser considerada función de cualquier otra, medida para las mismas muestras. Para garantizar la calidad de los resultados, algunas pruebas específicas deben ser realizados paralelamente al análisis estadístico de los resultados, incluyendo la verificación del coeficiente de determinación, la significación de la correlación y de la regresión, así como la ocurrencia de puntos discrepantes. Se utilizaron perfiles de pozos de la plataforma continental brasileña para ilustrar algunas de las posibles aplicaciones de curvas sintéticas del perfil sísmico, obtenidas por medio de regresión, así como en los cálculos de porosidad, tiempo de tránsito integrado y para la generación de sismogramas sintéticos. Se propone un guión para validar el proceso de estimativa y verificar cantitativamente las curvas calculadas.*

## **1. INTRODUÇÃO**

É rotina, em diversos segmentos da área de E&P da indústria do petróleo, empregar métodos de regressão para estimar propriedades petrofísicas medidas pelas ferramentas de perfilagem em intervalos não-perfilados. Para isto, utilizam-se curvas de perfis como variáveis independentes. Como estas ferramentas são projetadas para registrar variações na porosidade, tipo de fluido e litologia, admite-se que qualquer curva de perfil pode ser considerada função de outras variáveis (as outras curvas de perfis), medidas para os mesmos níveis de profundidade.

Do ponto de vista operacional, a rápida difusão destes métodos se deve, dentre outros fatores, às facilidades proporcionadas por novos programas e máquinas. A interface amigável dos pacotes estatísticos, como o SAS® e SPSS®, tornou bastante simples a tarefa de construir modelos para a geração das chamadas “curvas sintéticas” (Souza Jr., 1991; Zapparolli, 1991, e D’Abbadia, 1994). Para garantir a qualidade do produto final e prevenir distorções nos resultados, convém observar certos cuidados. Exemplifica-se, aqui, a aplicação deste procedimento com o estudo de um caso concreto.

## **2. DEFINIÇÃO DOS LIMITES DA REGRESSÃO**

Gerar uma curva sintética do perfil sísmico para poços em que esta curva não estava disponível ao longo de uma seção de interesse foi a tarefa aqui proposta. Utilizaram-se poços de duas áreas de um campo de petróleo. Os limites em área e em profundidade foram definidos demarcando-se os pacotes de rocha onde as variáveis estimada (perfil sísmico) e estimadoras (demais curvas de perfis, neste caso: raios gama, indução profunda, porosidade neutrônica e densidade) apresentam comportamento suficientemente homogêneo, levando em conta as aplicações às quais se destinam as curvas sintéticas. Desta forma, podem ocorrer situações em que seja conveniente subdividir formações ou membros, se ocorrerem variações nos dados que justifiquem tratá-los separadamente.

## **3. TREINAMENTO E VALIDAÇÃO DO MODELO DE REGRESSÃO**

A segunda etapa do processo consiste na separação do conjunto de dados em dois grupos: um para construção do modelo de regressão (treinamento), outro para teste de *performance* (validação). Esta separação pode ser feita por escolha direta, em que o analista seleciona que poços deverão compor cada grupo, ou por amostragem aleatória, de acordo com os passos descritos a seguir:

1. reunião dos poços em conjunto único;
2. divisão aleatória do conjunto acima em um grupo de treinamento e um de validação.

É possível simular a escolha aleatória com qualquer pacote estatístico.

Em condições normais, os dois processos são igualmente eficientes para a produção de modelos de regressão confiáveis.

## **4. APLICAÇÃO DO MÉTODO E ANÁLISE DOS RESULTADOS**

### **4.1. Estimativa do Tempo de Trânsito Integrado (TTI)**

Nas figuras 1a, 1b e 1c, estão alguns dos resultados obtidos. Pode-se notar que a curva de valores estimados oscila em torno dos valores medidos (perfil sônico original). Este comportamento é resultante do algoritmo de regressão que procura “acertar na média”. Assim, é importante compreender que a utilização da curva sintética para estimativas localizadas pode estar bastante comprometida, como, por exemplo, no intervalo x139/x154 m do poço A (fig. 1a).

Nas figuras 2a, 2b e 2c, são mostrados os tempos de trânsito integrados original (TTI) e estimado (TTIE), bem como o percentual de erro ao longo do intervalo perfilado:

$$\text{Erro} = (TTI - TTIE) \times 100 / TTI$$

Como é grandeza integrada – obtida pela soma dos registros de um grande número de pontos – as eventuais oscilações em torno dos valores medidos sofrem compensações, fazendo com que o percentual de erro diminua ou se estabilize à medida que o número de termos da soma aumenta. Os erros máximos obtidos – 4,0% no poço A, 1,6% no poço B e 1,4% no poço C – servem de indicadores para decidir se as curvas sintéticas atendem aos níveis de precisão exigidos em cada aplicação específica.

### **4.2. Geração do Sismograma Sintético**

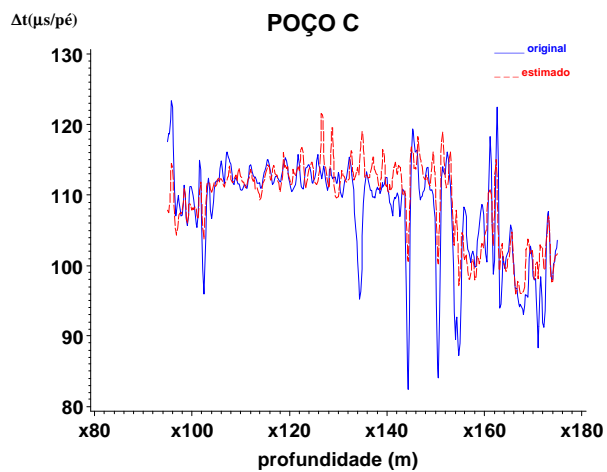
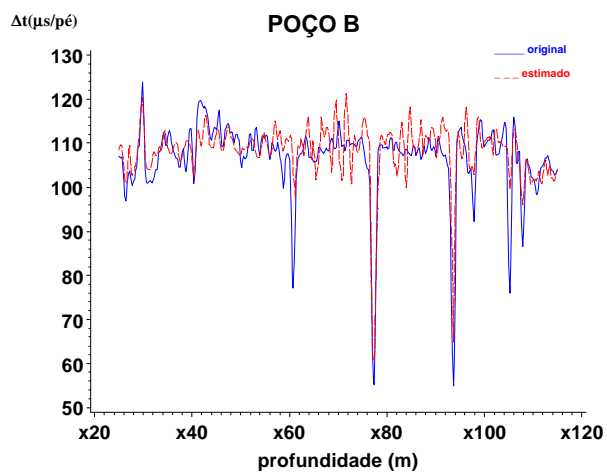
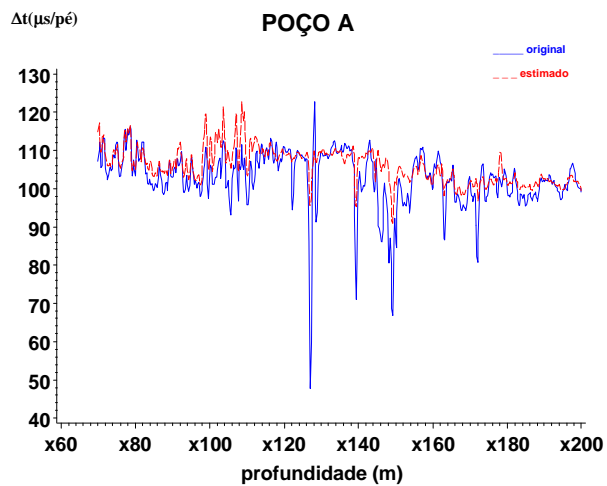
Dentre outras aplicações, o perfil sônico é utilizado em conjunto com o perfil densidade para elaborar sismogramas sintéticos para fazer a calibração entre os registros do levantamento sísmico de superfície e os das ferramentas dos perfis.

Foram construídos os sismogramas sintéticos (8-100 Hz - polaridade normal) para os poços A, B e C (figs. 3a, 3b e 3c). Os traços obtidos com o perfil sônico original (1) são, de maneira geral, bastante similares aos obtidos com o perfil sintético (2), sugerindo que, nos poços B e C, os resultados da regressão podem ser utilizados, com grau de confiabilidade satisfatório, para a geração de sismogramas sintéticos.

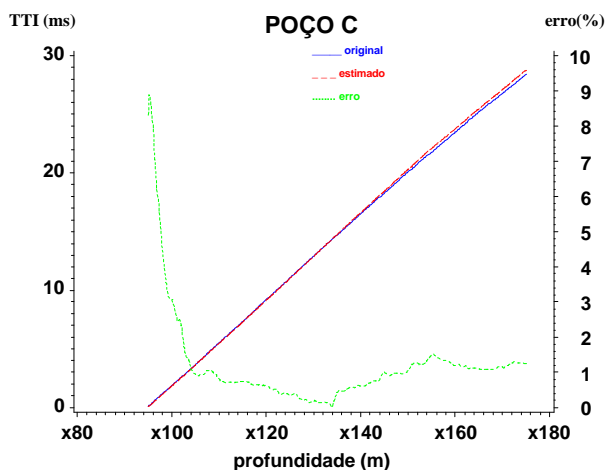
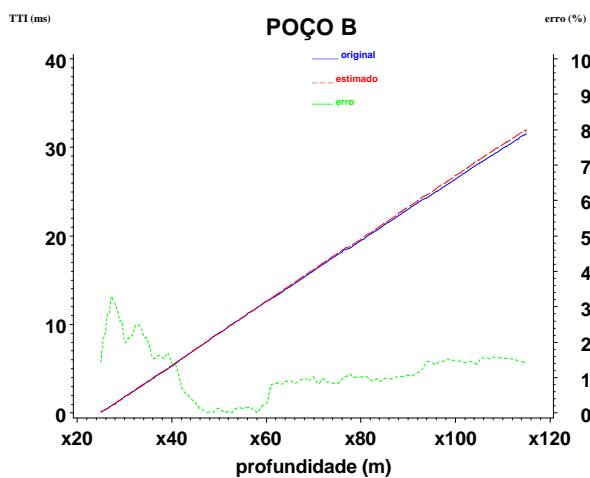
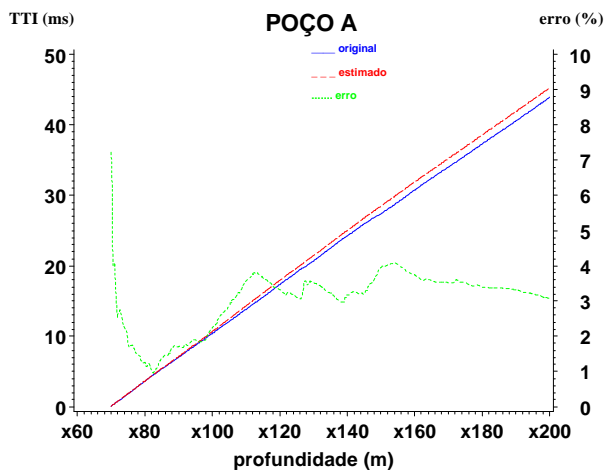
### **4.3. Cálculo da Porosidade**

Nas figuras 4a, 4b e 4c, são mostradas as curvas de porosidade efetiva calculadas com o perfil sônico original e com a curva sintética obtida a partir dos perfis de indução profunda, densidade, porosidade neutrônica e raios gama ( $\Delta t_{si1}$ ). Foi gerada, também, a curva de porosidade efetiva, calculada com o perfil sônico sintético obtido somente com as curvas de indução profunda e raios gama ( $\Delta t_{si2}$ ), para verificar a magnitude dos erros na ausência de perfis de porosidade dentre as variáveis independentes.

De maneira geral, observa-se que, da mesma forma que no cálculo do tempo de trânsito integrado, os erros tendem a se compensar ao longo do intervalo processado. Entretanto, deve-se notar que, localmente, as estimativas de porosidade efetiva estão bastante distorcidas, sugerindo que o uso de curvas sintéticas para estimativas pontuais de porosidade deve ser visto com reservas.



**Fig. 1a, 1b e 1c** - Tempos de trânsito estimado e original *versus* profundidade nos poços A, B e C.  
**Fig. 1a, 1b and 1c** - Estimated and original transit times *versus* depth for wells A, B and C.

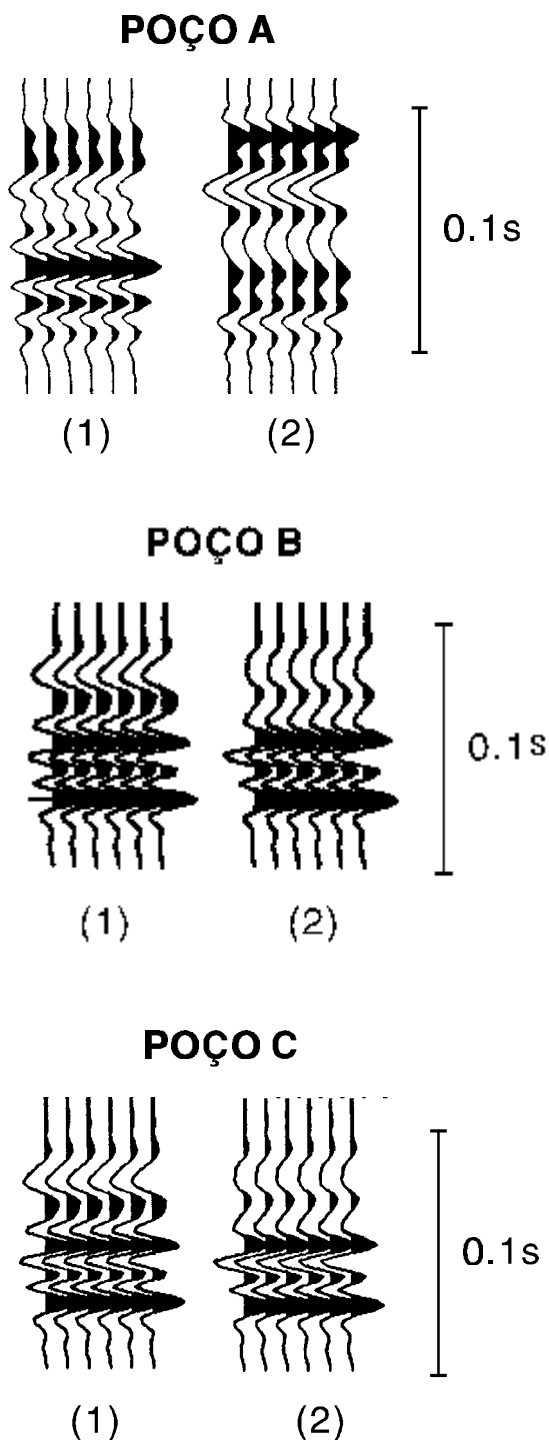


**Fig. 2a, 2b e 2c** - Tempos de trânsito integrado estimado e original *versus* profundidade, e erro percentual *versus* profundidade nos poços A, B e C.

**Fig. 2a, 2b and 2c** - Estimated and original integrated transit times versus depth, and percentage error versus depth for wells A, B and C.

Como se observa na tabela I, os valores médios da porosidade efetiva ao longo da seção de interesse não variam significativamente nos poços B e C.

No poço A, a diferença entre a porosidade efetiva calculada com o perfil sônico original e a calculada com o perfil sônico sintético ( $\Delta t_{si2}$ ) é de 2,35 pontos percentuais, que corresponde a um erro absoluto de 7,14%.



**Fig. 3a, 3b e 3c** - Sismogramas sintéticos obtidos com o perfil sônico original (1), e com o perfil sônico estimado (2) dos poços A, B e C.

**Fig. 3a, 3b and 3c** - Synthetic seismographs obtained with the original sonic log (1) with the estimated sonic log (2) for wells A, B and C.

#### 4.4. Outras Aplicações

As curvas sintéticas obtidas por regressão linear não poderão ser utilizadas em conjunto com as originais em aplicações como análise discriminante ou em sistemas de interpretação de perfis, que empreguem métodos lineares de inversão por minimização. A não ser que nestas aplicações ou na regressão as variáveis originais sejam previamente transformadas por operadores não-lineares, como, por exemplo, definindo-se uma nova variável  $ILD_1 = \text{Log}(\text{ILD})$ . Caberá, então, ao intérprete, explicar o modelo físico que sustenta esta transformação. Em aplicações onde o perfil sônico é utilizado qualitativamente, como no procedimento de correlação geológica de perfis, a curva sintética pode ser empregada, desde que verificada sua consistência, de maneira análoga ao discutido acima.

**TABELA I**  
**POROSIDADE MÉDIA (%)**  
*TABLE I*  
*AVERAGE POROSITY (%)*

	POÇO A	POÇO B	POÇO C
$\Delta t_{original}$	32,90	34,03	34,46
$\Delta t_{si1}$	33,92	34,46	35,24
$\Delta t_{si2}$	35,25	34,54	34,89

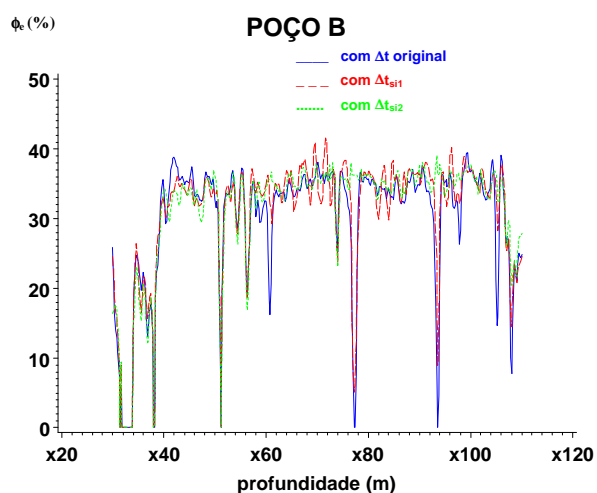
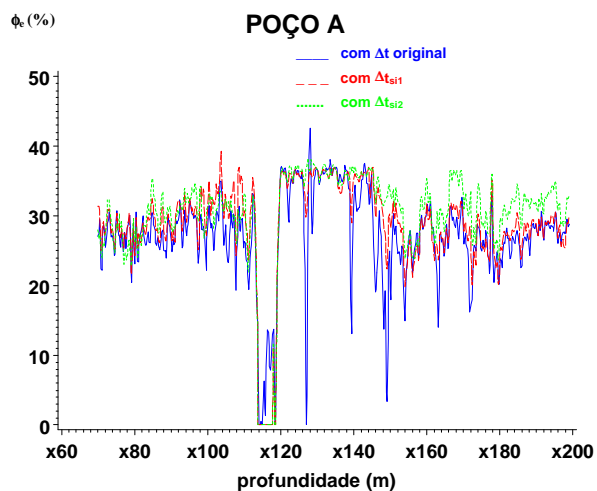
## 5. TESTES ESTATÍSTICOS

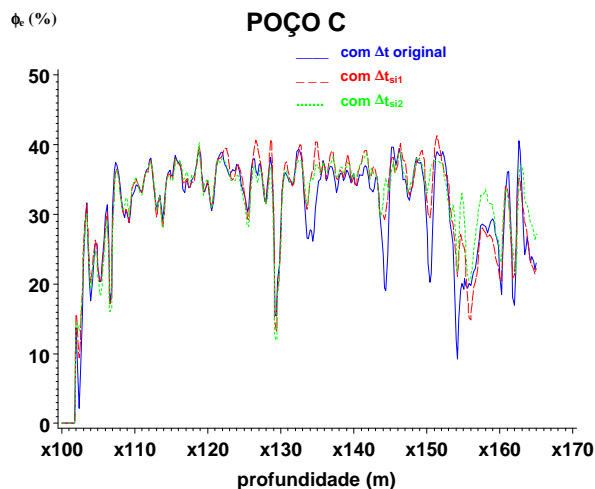
Além de analisar a confiabilidade das curvas sintéticas para cada aplicação específica em que serão usadas, é importante verificar a consistência estatística da regressão. Isto pode ser feito examinando-se o valor absoluto do coeficiente de determinação ( $R^2$ ), a significância da correlação e da regressão, o padrão de distribuição dos resíduos e a ocorrência de pontos discrepantes (*outliers*).

### 5.1. Interpretando o Coeficiente de Determinação ( $R^2$ )

O coeficiente de determinação ( $R^2$ ) expressa a proporção da variância total da variável dependente que pode ser explicada por uma relação linear com as variáveis independentes (Walpole e Myers, 1985; SAS,1990).

Assim, um coeficiente de determinação de 0,64 ( $R=0,80$ ), por exemplo, indica que 64% da variação total nos valores do perfil sônico na amostra analisada pode ser predita pelas variáveis regressoras (as demais curvas de perfis). Valores de  $R^2$  acima de 0,6 são considerados satisfatórios para a geração de curvas sintéticas.





**Fig. 4a, 4b e 4c** - Porosidade efetiva calculada com o perfil sônico original e com o perfil sônico sintético.  $\Delta t_{s11}$  foi gerada a partir dos perfis de indução profunda, densidade, porosidade neutrônica e raios gama.  $\Delta t_{s12}$  a partir dos perfis de indução profunda e de raios gama.

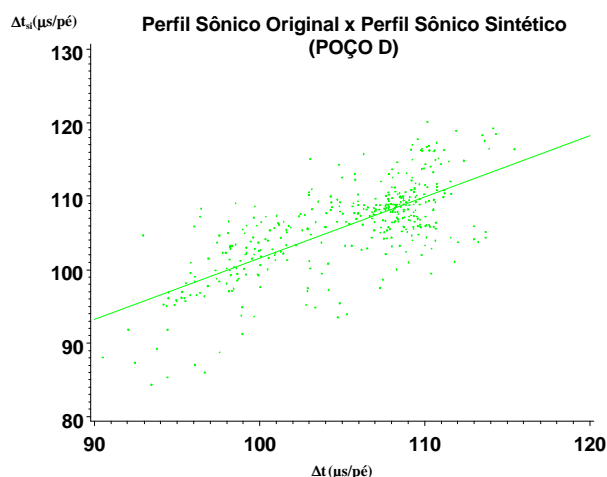
**Fig. 4a, 4b and 4c** - Effective porosity calculated from the original and synthetic sonic logs.  $\Delta t_{s11}$  was generated from deep induction, density, neutron porosity and gamma ray logs.  $\Delta t_{s12}$  from deep induction and gamma ray logs.

## 5.2. Significância da Correlação (R)

O coeficiente de correlação de Pearson ou coeficiente de correlação amostral (R) é uma estimativa do parâmetro  $\rho$  que expressa a relação entre duas variáveis (o perfil original e o perfil sintético) de uma população (o conjunto de todos os possíveis registros de determinado perfil na área em estudo). Se ambas as variáveis foram amostradas aleatoriamente e forem normalmente distribuídas, pode-se testar a significância de R por meio da estatística t de Student dada por:

$$t = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

com  $n - 2$  graus de liberdade, onde  $n$  é o número de observações da amostra, como descrito por Davis (1986). No processo de geração de curvas sintéticas, o número de observações disponíveis é, em geral, suficientemente grande para compensar eventuais valores pequenos do coeficiente de correlação amostral (R), e assim possibilitar a verificação de que R é significativamente diferente de zero, isto é, para se rejeitar a hipótese de que as duas variáveis (perfil sônico original e perfil sônico estimado) são independentes. No exemplo da figura 5:  $R^2 = 0,6466$ ,  $R = 0,8041$ ,  $n = 1378$ ,  $t = 50,17$  e  $t_{\text{crítico}} \cong 2,60$  (para 1376 graus de liberdade e nível de significância ( $\alpha$ ) de 1%).



**Fig. 5** - Exemplo de regressão com  $R^2=0,6466$ ,  $R=0,8041$ ,  $n=1378$  e  $t=50,17$ .

**Fig. 5** - Regression example with  $R^2 = 0.6466$ ,  $R = 0.8041$ ,  $n = 1378$  and  $t = 50.17$ .



O valor  $t_{\text{crítico}}$  corresponde ao menor valor da estatística  $t$  para que, com este teste estatístico, se possa rejeitar a hipótese de que  $\rho = 0$ . Como, neste exemplo,  $t > t_{\text{crítico}}$ , pode-se concluir que  $\rho \neq 0$ . Isto significa que há uma correlação entre a curva sintética e a curva original.

### 5.3. Significância da Regressão

O objetivo do teste de significância da regressão é verificar se a variância sobre a linha de regressão é diferente da variância das observações, o que é feito por meio da estatística  $F$ , como esquematizado na tabela II (Davis, 1986).

**TABELA II**  
**ANÁLISE DE VARIÂNCIA**  
*TABLE II*  
*VARIANCE ANALYSIS*

Fonte de Variação	Graus de liberdade (A)	Soma dos quadrados (B)	(B)/(A)	F
Modelo	m	$SS_r$	$MS_r=SS_r/m$	
Erro	n-m-1	$SS_d$	$MS_d=SS_d/(n-m-1)$	$MS_r/MS_d$
Varição total	n-1	$SS_t$		

onde:

$$SS_t = \sum_{i=1}^m (\Delta t - \overline{\Delta t})^2 = \text{variação total};$$

$$SS_r = \sum_{i=1}^m (\Delta t_{si} - \overline{\Delta t})^2 = \text{variação explicada pela regressão};$$

$$SS_d = SS_t - SS_r = \text{variação não explicada pela regressão};$$

$\Delta t$  = perfil sônico original;

$\Delta t_{si}$  = perfil sônico sintético;

$\overline{\Delta t}$  = média das leituras do perfil sônico original;

$n$  = número total de leituras do perfil sônico original;

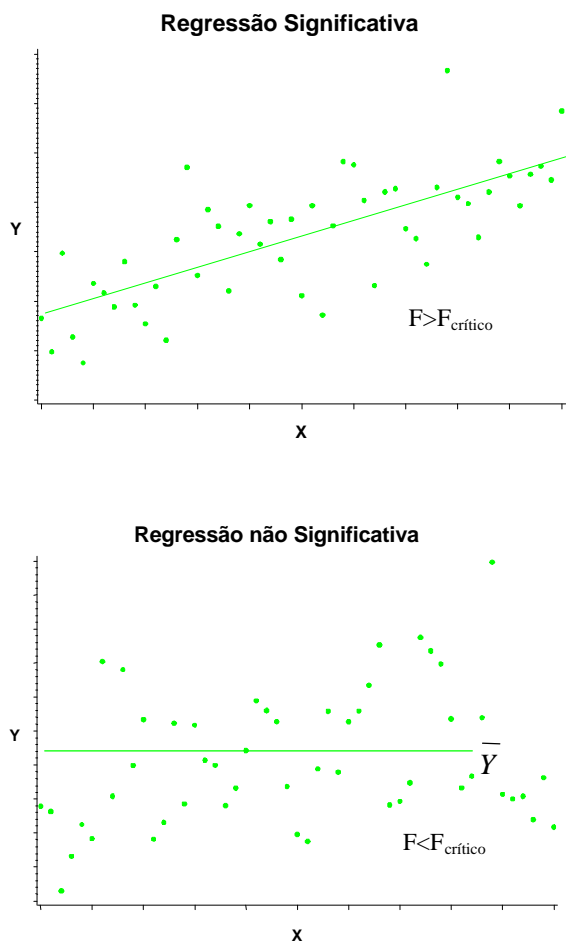
$m$  = número de variáveis independentes utilizadas na regressão;

$$R^2 = \frac{SS_r}{SS_t} = \text{coeficiente de determinação}.$$

Na figura 6, apresentam-se, esquematicamente, exemplos de regressões lineares significativa e não-significativa. O valor  $F_{\text{crítico}}$  é obtido por meio das tabelas da distribuição  $F$ , com os graus de liberdade  $m$  (numerador) e  $n-m-1$  (denominador), de acordo com o nível de significância ( $\alpha$ ) estabelecido anteriormente.

Da mesma forma que no teste  $t$  de Student (usado para testar a significância da correlação), o número de observações utilizadas no processo de geração de curvas sintéticas é, em geral, suficientemente grande para produzir valores de  $F$  superiores ao  $F_{\text{crítico}}$ , permitindo, assim, concluir que a regressão é significativa. No exemplo da figura 5;  $F = 628$  e  $F_{\text{crítico}} \cong 3,4$  (para 4 e 1 373 graus de liberdade e nível de significância ( $\alpha$ ) de 1%). Como  $F > F_{\text{crítico}}$ , pode-se rejeitar a hipótese de que a variância sobre a linha de regressão coincide com a variância das observações.

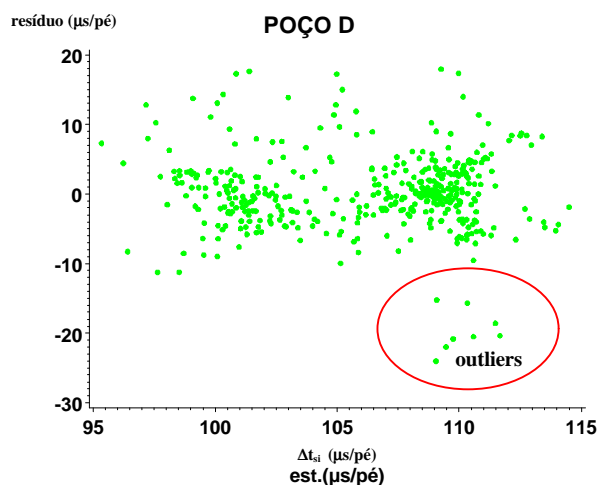
Caso se deseje selecionar o conjunto de variáveis independentes que produza a melhor regressão, podem-se utilizar os algoritmos disponíveis nos pacotes estatísticos (*backward*, *forward* ou *stepwise selection*), que selecionam as variáveis que irão compor o modelo de regressão, passo a passo, levando em conta os valores mais elevados da estatística  $F$  (SAS, 1990).



**Fig. 6** - Exemplos de regressão linear significativa e não-significativa. Adaptado de Draper e Smith (1981).  
**Fig. 6** - Examples of significant and non-significant linear regression. Modified from Draper and Smith (1981).

#### 5.4. Análise de Resíduos

Um importante diagnóstico da qualidade da regressão é fornecido pela análise dos resíduos (diferença entre os valores originais e os valores estimados), como apresentado na figura 7.



**Fig. 7** - Gráfico do resíduo (diferença entre o valor original e o valor estimado) versus valor estimado.  
**Fig. 7** - Residuals (difference between the original and estimated values) versus the estimated value.

Em condições ideais, os resíduos não terão correlação com os valores estimados ( $R^2_{\Delta t_{si}, \text{resíduo}}=0$ ), terão distribuição normal centrada em zero (fig. 8) e a mesma variância ao longo de toda a faixa de valores das variáveis original e estimada. Variações deste padrão podem indicar a necessidade de um termo quadrático ou de transformação das variáveis independentes, ou ainda a ocorrência de tendência temporal (*time trend*) nos dados (SAS, 1987). No exemplo mostrado nas figuras 7 e 8, a correlação (R) entre os valores preditos ( $\Delta t_{si}$ ) é de 0,01 e a média dos resíduos 0,45  $\mu\text{s}/\text{pé}$ . O gráfico em que se relacionam os valores estimados e os resíduos é útil, também, para a identificação de pontos discrepantes (*outliers*), de acordo com a figura 7.

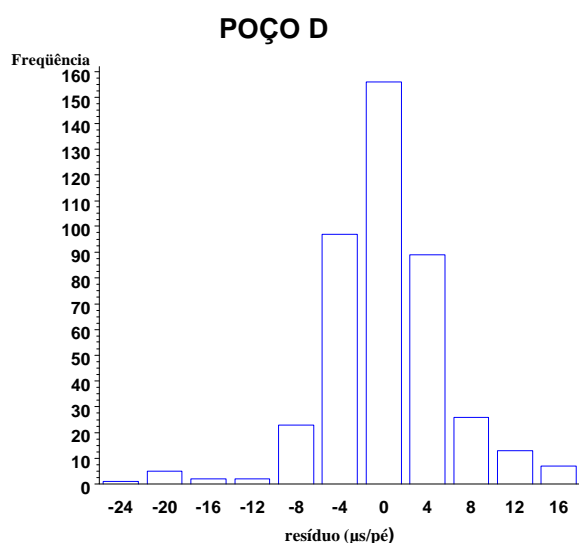
Outros eficientes indicadores de *outliers* são fornecidos pelo cálculo do resíduo padronizado (*studentized residual*), que corresponde à divisão do resíduo pelo seu erro-padrão e da estatística D de Cook, que reflete a influência de cada observação nas estimativas. Uma vez reconhecidos, os *outliers* devem ser examinados cuidadosamente e só poderão ser removidos dos grupos de treinamento e validação dos resultados caso se identifiquem as razões pelas quais essas leituras não são confiáveis, o que pode ocorrer, por exemplo, devido a más condições mecânicas de poço (desmoronamentos e rugosidade, dentre outros).

## 6. MODELO DE REGRESSÃO

Alguns sistemas de tratamento de dados de perfis, como o LOGCALC<sup>®</sup>, têm implementados três métodos de obtenção da reta de ajuste da regressão (figs. 9a, 9b e 9c): mínimos quadrados convencional, mínimos quadrados com critério de ajuste ortogonal e eixo maior reduzido.

Se não se deseja obter a relação de dependência entre as variáveis analisadas, recomenda-se o uso dos métodos eixo maior reduzido ou mínimos quadrados com critério de ajuste ortogonal (Davis, 1986), como, por exemplo, no *Picket plot* (Dewan, 1983), onde as variáveis Log (porosidade) e Log (resistividade) são utilizadas para determinar o parâmetro m, que corresponde à inclinação da reta de  $S_w$  (saturação de água) = 1, traçada pelos pontos de menor resistividade, correspondentes a diferentes valores de porosidade (fig. 10). Neste caso, ambas as variáveis estão sujeitas a erros de medida. Conseqüentemente, não há razão para que os desvios da reta de regressão sejam atribuídos a somente uma delas.

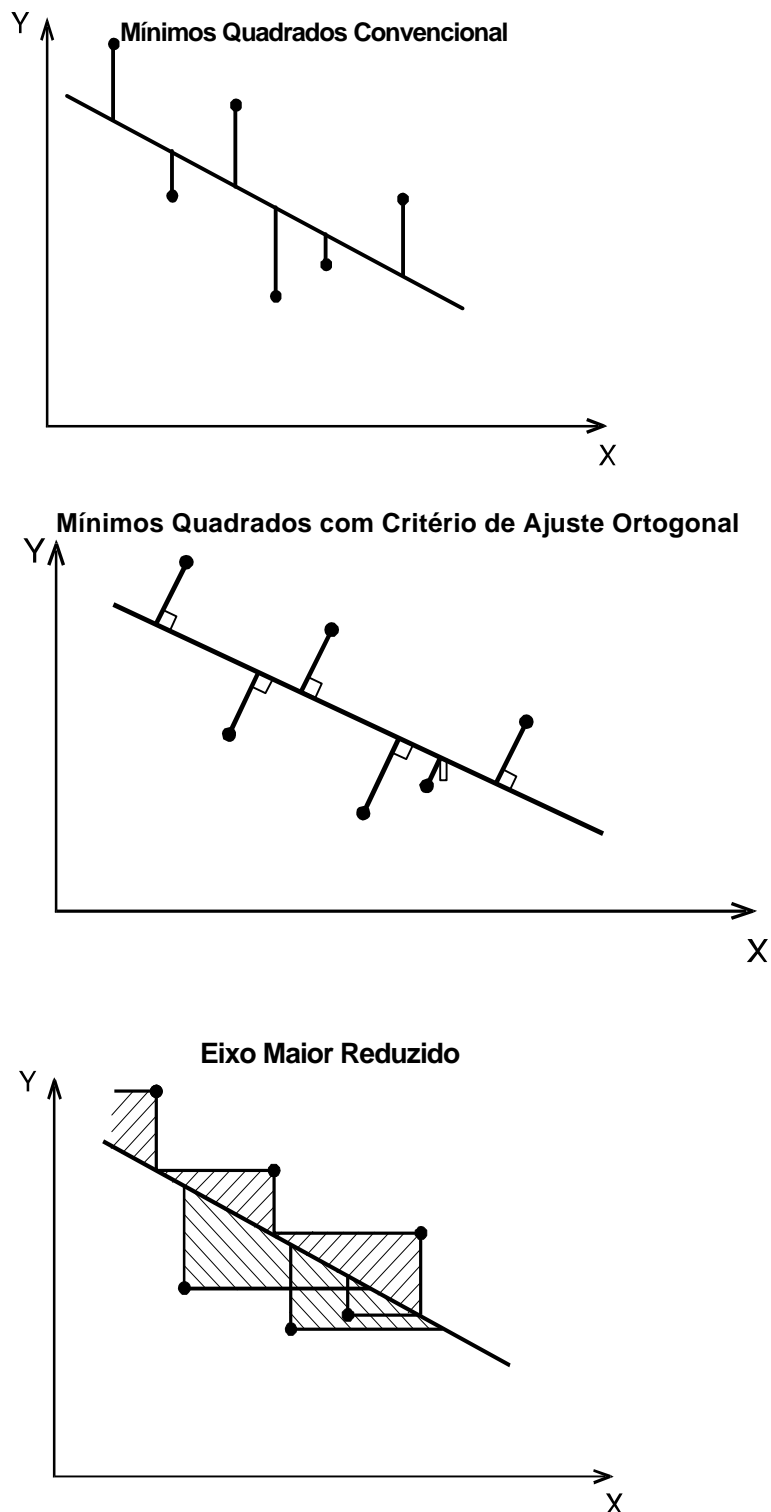
Entretanto, quando se pretende estimar uma variável em função de outra(s), o método que produz as melhores estimativas da variável dependente é o dos mínimos quadrados, onde se consideram somente os erros de medida da variável dependente (SSI, 1987).



**Fig. 8** - Histograma dos resíduos.

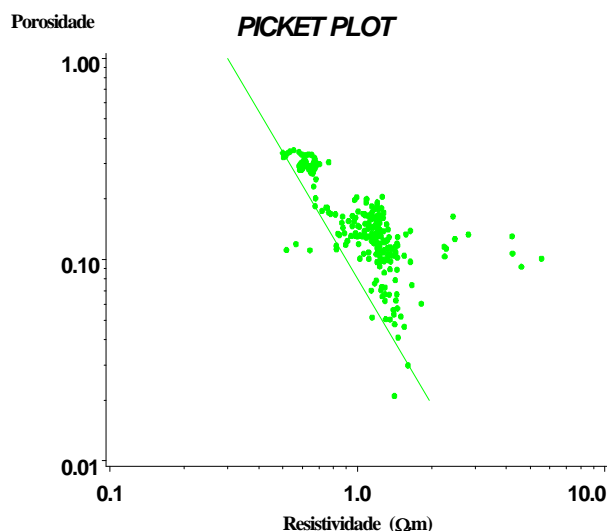
*Fig. 8 - Residuals block diagram.*

Na geração do perfil sônico sintético, o que se busca estimar é o tempo de trânsito da onda compressional. O perfil sônico é o melhor estimador desta grandeza. Dessa forma, a melhor reta de ajuste será a que minimizar a soma dos quadrados das diferenças entre os valores estimados e observados do perfil sônico, o que é obtido por meio do método dos mínimos quadrados convencional.



**Fig. 9a, 9b e 9c** - Métodos de obtenção da reta de ajuste da regressão.  
**Fig. 9a, 9b and 9c** - Methods for obtaining the regression line.

Outro ponto que merece atenção são as curvas utilizadas como variáveis independentes no modelo de regressão. Assim, por exemplo, devem-se evitar registros fora dos limites de acurácia das ferramentas, como no caso das leituras de resistividade profunda do perfil de indução acima de 100  $\Omega\text{m}$ .



**Fig. 10** - Picket plot utilizado para determinar o parâmetro  $m$  da equação de saturação de água. Neste caso, para obter a reta de ajuste da regressão, recomenda-se o uso dos métodos eixo maior reduzido ou mínimos quadrados com critério de ajuste ortogonal.

*Fig. 10* - Picket plot used to determine the parameter  $m$  of the water saturation equation. In this case, to obtain the regression line, the use of the reduced major axis or least square methods, with the major axis criterion is recommended.

Alguns intérpretes têm incluído a profundidade no conjunto das variáveis independentes, em vista da boa correlação desta variável com o perfil sônico, o que pode ser observado em seções com pacotes espessos de folhelhos, verificando-se a diminuição do tempo de trânsito com o aumento da profundidade. Entretanto, este procedimento deve ser evitado em áreas de pressão anormalmente alta, onde esta tendência pode ser alterada. Além disso, especial cuidado deve ser tomado em relação ao emprego da regressão em faixas de profundidade diferentes daquelas utilizadas no treinamento e na validação do modelo. Assim, uma equação obtida com a profundidade entre as variáveis independentes, situando-se, por exemplo, entre 1 000 m e 2 000 m, não pode ser automaticamente estendida para intervalos fora dessa faixa. A relação linear estimada pode não se manter constante além desses limites (Chatterjee e Price, 1977).

## 7. CONSIDERAÇÕES FINAIS

A técnica de regressão múltipla para a geração de curvas sintéticas tem sido amplamente utilizada, nos últimos anos, em função da facilidade de sua aplicação e dos resultados produzidos. Ainda assim, é importante verificar a consistência dos resultados, tanto em termos estatísticos – por meio do exame do coeficiente de determinação ( $R^2$ ) dos testes de significância da correlação e da regressão, bem como do padrão de distribuição dos resíduos – como do ponto de vista das aplicações em que a curva sintética será utilizada, examinando-se se a mesma atende aos níveis de precisão exigidos em cada caso.

Da mesma forma que as demais curvas processadas de perfis (porosidade efetiva e saturação de água, por exemplo), as curvas sintéticas também são resultado de um processo de interpretação que envolve, dentre outros fatores, a escolha das variáveis independentes e os limites da representatividade da regressão, a critério do analista de perfis. Por esta razão, é recomendável, também que, ao fim do estudo, as curvas sintéticas sejam armazenadas em arquivos distintos dos que contêm as curvas originais, para evitar a utilização não-criteriosa destes dados.

Os procedimentos descritos para a obtenção da curva sintética do perfil sônico podem ser aplicados, com as devidas adaptações, na geração de outras curvas sintéticas.

## **AGRADECIMENTOS**

Ao geólogo Luis Carlos S. de Freitas, pela seleção das áreas de estudo.  
À geofísica Maria Angelina Carvalho e aos técnicos Jânio Carlos de Andrade e Luiz Fernando Muniz da Rocha, pela confecção dos sismogramas sintéticos, bem como aos geólogos Nuno André Pomar Couto e Dirceu Abrahão pela revisão dos originais.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

- CHATERJEE, S., PRICE, B. *Regression analysis by example*. Chichester : J. Wiley, 1977. 228p.
- D'ABBADIA, M. R. I. *Processo de geração de perfil sônico sintético com o suporte de regressões lineares*. Salvador, BA : PETROBRAS. DEXBA. DIRGEO. SERTRAG, 1994. (Relatório interno).
- DAVIS, J. C. *Statistics and data analysis in geology*. 2. ed. Chichester : J. Wiley, 1986. 550p.
- DEWAN, J. T. *Essentials of modern open-hole log interpretation*. Tulsa, Okla : Penn Well, 1983. 361p.
- DRAPER, N. R., SMITH, H. *Applied regression analysis*. 2nd ed. Chichester : J. Wiley, 1981. 709p.
- STATISTICAL ANALYSIS SYSTEM INSTITUTE INCORPORATION. *SAS<sup>®</sup> system for elementary statistical analysis*. Cary, NC : SAS Institute Inc., 1987. 416p.
- STATISTICAL ANALYSIS SYSTEM INSTITUTE INCORPORATION *SAS/STAT<sup>®</sup> user's guide*, volume 2. Cary, NC : SAS Institute Inc., 1990. 1686p.
- SCIENTIFIC SOFTWARE INTERCOMP. *LOGCALC II Newsletter*. Denver, Colo : SSI, 1987. 12 p.
- SOUZA JUNIOR, 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 : PETROBRAS, 1991. v.1, p. 121-130.
- WALPOLE, R. E., MYERS, R. H. *Probability and statistics for engineers and scientists*. London : Macmillan, 1985. 639 p.
- ZAPPAROLLI, L. H. *Confecção da curva sintética do perfil sônico a partir da regressão de dados de perfis*: Informe Técnico. Natal : PETROBRAS. DEBAR. DIRGEO. SERTRAG, 1991.

## **EXPANDED ABSTRACT**

*The use of regression methods to estimate petrophysical properties is a routine procedure in many segments of the E&P area of the oil industry. In order to do this, it is assumed that, since these logging tools are designed to record variations in porosity, fluid type and lithology, it can be assumed that any log curve can be considered a function of other log curves, measured for the same samples.*

*In order to guarantee final product quality and avoid any distortions in the results, certain precautions should be taken. With the aim of discussing some of the possible situations found in practice, the application of this procedure is exemplified using the study of a real situation.*

*It was proposed to generate synthetic of sonic logs in wells where this curve was not available over a required section using data from wells in two separate areas of an oilfield. The area and depth limits, for which each regression could be generalized were chosen by defining the rock packages where the estimated variables (sonic log) and the known parameters (the remaining log curves, in this case: gamma ray, deep induction, neutron porosity and bulk density), show homogeneous behavior, taking into account the applications for which the synthetic curves are intended. Situations may occur where it is useful to subdivide the rock packages, if there are variations which justify treating them separately.*

*The second stage of the process consists of separating the data into two groups: one for constructing the regression model (training) and the other to test the performance of the model (validation). This separation may be carried out by direct choice  $\frac{3}{4}$  where the analyst selects, according to his own criteria, the wells that should compose each group  $\frac{3}{4}$  or by random sampling.*

*Beside the reliability analysis of the synthetic curves for each specific application, it is important to check the statistical consistency of the regression. This can be done by examining the absolute value of the goodness of fit coefficient ( $R^2$ ), the significance of the correlation and regression, the residuals distribution pattern and the occurrence of outliers.*

*As for the other processed log curves (effective porosity, water saturation, etc.), the synthetic curves are also the result of an interpretation process which involves, amongst other factors, the choice of independent variables and limits for the application of the regression, at the log analyst's discretion. For this reason, at the end of the study, it is advisable to store the synthetic curves in separate files from the original curves, to avoid undiscerning use of the data.*