

# 5º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS



## TÍTULO DO TRABALHO:

Análise geoestatística de dados geoquímicos de superfície aplicada à exploração de hidrocarbonetos.

## AUTORES:

Michel Ângelo Pérez Bedregal 1. Luiz Landau 2. Felix Thadeu Teixeira Gonçalves 3.

PEC-COPPE/UFRJ <sup>1</sup>. PEC-COPPE/UFRJ <sup>2</sup>. VALE <sup>3</sup>.

## INSTITUIÇÃO:

PEC / COPPE / UFRJ

*Este Trabalho foi preparado para apresentação no 5º Congresso Brasileiro de Pesquisa e Desenvolvimento em Petróleo e Gás- 5º PDPETRO, realizado pela a Associação Brasileira de P&D em Petróleo e Gás-ABPG, no período de 15 a 22 de outubro de 2009, em Fortaleza-CE. Esse Trabalho foi selecionado pelo Comitê Científico do evento para apresentação, seguindo as informações contidas no documento submetido pelo(s) autor(es). O conteúdo do Trabalho, como apresentado, não foi revisado pela ABPG. Os organizadores não irão traduzir ou corrigir os textos recebidos. O material conforme, apresentado, não necessariamente reflete as opiniões da Associação Brasileira de P&D em Petróleo e Gás. O(s) autor(es) tem conhecimento e aprovação de que este Trabalho seja publicado nos Anais do 5ºPDPETRO.*

Análise geoestatística de dados geoquímicos de superfície aplicada à exploração de hidrocarbonetos.

## Abstract

This work examined the distribution of geochemical data from surface (gaseous hydrocarbons), geochemical outliers data, identified by various methods, and compared deterministic interpolation methods (minimum curvature and the inverse square of the distance) and geostatistical (kriging). The measurements obtained in the field were analyzed by classical descriptive statistics and spacial. Statistical analysis showed that the geochemical data of hydrocarbon gases do not follow in general, a lognormal distribution or a normal one. Methods to identify outliers in the geochemical data were reviewed and tested. The rule [mean ( $\bar{x}$ )  $\pm$  2 standard deviations (sdev)] used to estimate the threshold values dividing background of anomalous data, hitherto used since its introduction in the 60s, showed different results of the proposed method in which it should identify approximately 2<sup>1/2</sup>% of data on each end. The median (med)  $\pm$  2 median absolute deviations (mad), the boxplot and 98 th percentile, appeared to be more appropriate in the estimates of threshold and the determination of the extent of background data. However, each can result in different values of threshold. The interpolation of these data in order to generate surface maps through geostatistical method, which is based on the kriging interpolation, takes into account the characteristics of spatial autocorrelation of regional variables, which showed better results than the deterministic methods.

## Introdução

Atividades e estudos que visam determinar o potencial petrolífero de uma determinada região, isto é, a sua capacidade produtiva e a valoração de suas reservas de óleo e gás é um fator de grande interesse da indústria petrolífera. Dentre as técnicas, o método de prospecção geoquímica de superfície tem sido empregado com sucesso como ferramenta de suporte a exploração de petróleo, seja proporcionando uma melhor avaliação do potencial petrolífero de áreas ou bacias sedimentares, ou permitindo uma análise mais criteriosa do risco geológico de leads e prospectos. Esse método de prospecção consiste basicamente na detecção de microexsudações de hidrocarbonetos gasosos em amostras de solo.

A identificação, caracterização e configuração de uma anomalia em superfície de hidrocarboneto gasoso termogênico em um solo de gás inquérito só é possível por meio da análise de um grande número de amostras em uma área específica. As amostras são coletadas e levadas ao laboratório, onde são analisadas para caracterizar a composição e a origem de hidrocarbonetos gasosos. Essas medições são realizadas para determinar a presença ou ausência de reservatórios de hidrocarbonetos.

Na última década, houve uma grande evolução na sensibilidade dos equipamentos laboratoriais propiciando a detecção de concentrações diminutas de hidrocarbonetos. Esta tecnologia permitiu também o aprimoramento das técnicas de amostragem que tornou mais fácil e simplificada a coleta (de quantidades relativamente pequenas de gases à baixa pressão), o transporte (com menor perigo de

acidentes) e pronta análise dos gases. Contudo, pouco tem sido feito no sentido do desenvolvimento das técnicas interpretativas, tendo se limitado a aplicação de técnicas estatísticas básicas, desconsiderando a dependência espacial das variáveis.

A maioria dos métodos estatísticos é largamente fundamentada na suposição que os dados estudados mostram uma distribuição normal ou lognormal. A fundamentação dessa distribuição iniciou através de Ahrens (1953), que propôs que a distribuição de elementos-traço em materiais naturais tende ser mais próxima de uma distribuição lognormal que normal, isto é, os logaritmos dos dados tendem a ser distribuídos simetricamente sobre a média geométrica, já em uma distribuição simétrica de valores naturais, são distribuídos simetricamente sobre a média aritmética.

Quando uma determinada propriedade varia de um local para outro com algum grau de organização ou continuidade, expresso pela dependência espacial, a estatística clássica deve ser abandonada e dar lugar a uma estatística relativamente nova: a geoestatística. Por estatística clássica entende-se aquela que utiliza parâmetros como média e desvio padrão para representar um fenômeno, e baseia-se na hipótese principal de que as variações de um local para outro são aleatórias. Desse modo, esses dois ramos da estatística têm validade de aplicação em condições perfeitamente distintas. Para se determinar qual das duas deve ser usada em cada caso, utiliza-se o semivariograma que expressa a dependência espacial entre as amostras. Havendo dependência espacial, podem-se estimar valores da propriedade em estudo para os locais não amostrados dentro do campo, sem tendenciosidade e com variância mínima, pelo método denominado krigagem.

### Metodologia

Os dados são provenientes de um conjunto de 220 amostras de medidas de concentrações de gases (ppm) metano (C1), etano (C2), eteno (C2=), propano (C3), propeno (C3=), isobutano (IC4), butano (C4), buteno (C4=), isopentano (IC5), pentano (C5) e hexano+ (C6+) espaçadas regularmente 500 metros umas das outras.

Este trabalho avalia através de diferentes abordagens estatísticas a premissa que os dados geoquímicos seguem uma distribuição lognormal ou normal. Métodos para identificar dados anômalos foram revisados e testados: A regra [média  $\pm$  2 dp] descrita por Hawkes e Webb (1962) e tradicionalmente utilizada até os dias de hoje, a regra [med  $\pm$  2 dma], o boxplot e 98º percentil.

Através de mapas de superfície e QME (quadrado médio do erro) avalia os interpoladores determinísticos (mínima curvatura e inverso do quadrado da distância) e a interpolação geoestatística por krigagem, que leva em consideração as características espaciais de autocorrelação de variáveis regionalizadas. A metodologia geoestatística visa identificar a magnitude da variabilidade espacial dos dados de hidrocarbonetos gasosos, as características qualitativas ligadas à estrutura do fenômeno natural que elas representam, tais como: a localização, a continuidade espacial, a anisotropia e a estrutura de correlação existente entre valores tomados em dois pontos adjacentes no espaço.

## Resultados e Discussão

A maioria dados geoquímicos de hidrocarbonetos gasosos não apresentaram distribuição normal ou lognormal (Tabela 1)

Tabela 1 – Análise estatística descritiva das amostras originais (ppm).

| ESTATÍSTICAS                              | C <sub>1</sub> | C <sub>2</sub> | C <sub>2</sub> <sup>=</sup> | C <sub>3</sub> | C <sub>3</sub> <sup>=</sup> | i-C <sub>4</sub> | C <sub>4</sub> | C <sub>4</sub> <sup>=</sup> | i-C <sub>5</sub> | C <sub>5</sub> | C <sub>6</sub> <sup>+</sup> |
|-------------------------------------------|----------------|----------------|-----------------------------|----------------|-----------------------------|------------------|----------------|-----------------------------|------------------|----------------|-----------------------------|
| N.º de observações válidas                | 220            | 220            | 220                         | 220            | 220                         | 220              | 220            | 220                         | 220              | 220            | 220                         |
| Média                                     | 20,79          | 2,44           | 8,51                        | 1,21           | 5,37                        | 0,17             | 0,56           | 0,92                        | 0,30             | 0,53           | 1,43                        |
| Mediana                                   | 18,70          | 1,98           | 7,63                        | 1,01           | 2,90                        | 0,15             | 0,44           | 0,76                        | 0,15             | 0,26           | 0,54                        |
| Valor mínimo                              | 4,74           | 0,49           | 0,66                        | 0,35           | 0,46                        | 0,03             | 0,17           | 0,18                        | 0,03             | 0,03           | 0,03                        |
| Valor máximo                              | 51,33          | 8,81           | 31,53                       | 4,08           | 73,19                       | 1,01             | 1,91           | 3,39                        | 4,42             | 6,49           | 38,61                       |
| Amplitude                                 | 46,59          | 8,32           | 30,87                       | 3,73           | 72,73                       | 0,99             | 1,74           | 3,21                        | 4,39             | 6,47           | 38,59                       |
| Quartil Inferior                          | 14,61          | 1,41           | 4,83                        | 0,71           | 1,88                        | 0,12             | 0,32           | 0,56                        | 0,11             | 0,17           | 0,32                        |
| Quartil Superior                          | 25,19          | 3,06           | 10,63                       | 1,49           | 5,24                        | 0,22             | 0,73           | 1,16                        | 0,25             | 0,44           | 0,93                        |
| Amplitude Interquartílica                 | 10,58          | 1,64           | 5,80                        | 0,77           | 3,36                        | 0,10             | 0,41           | 0,60                        | 0,14             | 0,27           | 0,61                        |
| Variância                                 | 79,13          | 2,06           | 26,50                       | 0,48           | 77,92                       | 0,015            | 0,12           | 0,28                        | 0,31             | 0,90           | 19,64                       |
| Desvio Padrão                             | 8,90           | 1,43           | 5,15                        | 0,69           | 8,83                        | 0,12             | 0,34           | 0,53                        | 0,56             | 0,95           | 4,43                        |
| Assimetria                                | 1,10           | 1,64           | 1,24                        | 1,56           | 4,85                        | 2,45             | 1,61           | 1,55                        | 4,80             | 4,22           | 6,92                        |
| Excesso de Curtose                        | 1,06           | 2,95           | 1,96                        | 2,56           | 27,58                       | 11,31            | 2,57           | 2,83                        | 25,46            | 18,50          | 51,66                       |
| Limite Inferior (Q <sub>1</sub> -1,5XAIQ) | -1,26          | -1,05          | -3,87                       | -0,45          | -3,16                       | -0,03            | -0,30          | -0,34                       | -0,10            | -0,24          | -0,60                       |
| Limite Superior (Q <sub>3</sub> +1,5XAIQ) | 41,06          | 5,52           | 19,33                       | 2,65           | 10,28                       | 0,37             | 1,35           | 2,06                        | 0,46             | 0,85           | 1,85                        |
| CV                                        | 0,43           | 0,59           | 0,61                        | 0,57           | 1,64                        | 0,71             | 0,61           | 0,58                        | 1,87             | 1,79           | 3,10                        |
| S-W (*)                                   | 0,000          | 0,000          | 0,000                       | 0,000          | 0,000                       | 0,000            | 0,000          | 0,000                       | 0,000            | 0,000          | 0,000                       |
| K-S (*)                                   | 0,000          | 0,000          | 0,000                       | 0,000          | 0,000                       | 0,000            | 0,000          | 0,000                       | 0,000            | 0,000          | 0,000                       |

Quando analisamos os dados originais e estimamos os limites de variação de *background* usando diferentes procedimentos (Tabela 2), o método [mediana (med) + 2 desvios mediano absoluto (dma)] entrega os mais baixos limiares e identifica, conseqüentemente, os mais altos números de *outliers*, seguidos pelo *boxplot* [*upper whisker*]. A razão pela qual apresentam uma larga proporção de extremos valores identificados é que os estimadores robustos de posição (mediana) e escala (desvio mediano absoluto (med) ou amplitude interquartil (AIQ) são relativamente não influenciados por extremos valores de uma distribuição de dados lognormal ou fortemente assimétrica. Tanto a posição quanto a escala são relativamente baixas em relação à média e desvio padrão, resultando em valores mais baixos de limiares e maiores porcentagens de valores extremos identificados. O limiar definido pelo *boxplot* [*upper whisker*] é em todos os casos mais perto, porém mais baixo que o limite obtido por dados logtransformados usando a regra [média+2 desvios padrão (dp)].

## 5º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS

Tabela 2 – Média, desvio padrão (DP), mediana, desvio mediano absoluto (DMA), para os dados originais e transformados (ln) e resultados da definição do limiar superior via: [média+2 desvios padrão (dp)], [mediana (med) + 2 desvios medianos absolutos (dma)], o *boxplot* [*upper whisker*] e [98º percentil].

|                 | Média |       | DP    |      | Mediana |       | DMA   |      | [Média+2DP] |         | [Med+2DMA] |         | Upper whisker |         | 98º Percentil |         |
|-----------------|-------|-------|-------|------|---------|-------|-------|------|-------------|---------|------------|---------|---------------|---------|---------------|---------|
|                 | Orig. | Ln    | Orig. | Ln   | Orig.   | Ln    | Orig. | Ln   | Orig.       | Anti-ln | Orig.      | Anti-ln | Orig.         | Anti-ln | Orig.         | Anti-ln |
| C1              | 20.79 | 2.95  | 8.90  | 0.41 | 18.70   | 2.93  | 5.11  | 0.27 | 38.59       | 43.38   | 28.92      | 32.14   | 40.75         | 51.41   | 46.05         |         |
| C2              | 2.44  | 0.75  | 1.43  | 0.52 | 1.98    | 0.66  | 0.67  | 0.37 | 5.30        | 5.99    | 3.32       | 4.06    | 5.49          | 8.85    | 6.71          |         |
| C2 <sup>+</sup> | 8.51  | 1.95  | 5.15  | 0.65 | 7.63    | 2.03  | 2.92  | 0.39 | 18.81       | 25.79   | 13.47      | 16.61   | 17.63         | 31.81   | 22.52         |         |
| C3              | 1.21  | 0.05  | 0.69  | 0.51 | 1.01    | 0.01  | 0.33  | 0.36 | 2.59        | 2.92    | 1.67       | 2.08    | 2.61          | 4.09    | 3.19          |         |
| C3 <sup>+</sup> | 5.37  | 1.20  | 8.83  | 0.84 | 2.90    | 1.07  | 1.22  | 0.48 | 23.03       | 17.81   | 5.34       | 7.61    | 9.90          | 23.57   | 38.44         |         |
| IC4             | 0.17  | -2.00 | 0.12  | 0.79 | 0.15    | -1.90 | 0.04  | 0.26 | 0.41        | 0.66    | 0.23       | 0.25    | 0.35          | 0.48    | 0.55          |         |
| C4              | 0.56  | -0.74 | 0.34  | 0.54 | 0.44    | -0.82 | 0.16  | 0.36 | 1.24        | 1.40    | 0.76       | 0.90    | 1.28          | 1.92    | 1.66          |         |
| C4 <sup>+</sup> | 0.92  | -0.22 | 0.53  | 0.53 | 0.76    | -0.27 | 0.24  | 0.34 | 1.98        | 2.32    | 1.24       | 1.51    | 2.04          | 3.39    | 2.47          |         |
| IC5             | 0.30  | -1.83 | 0.56  | 0.98 | 0.15    | -1.90 | 0.05  | 0.39 | 1.42        | 1.14    | 0.25       | 0.33    | 0.44          | 0.68    | 2.48          |         |
| C5              | 0.53  | -1.23 | 0.95  | 0.96 | 0.26    | -1.34 | 0.11  | 0.46 | 2.43        | 1.99    | 0.48       | 0.66    | 0.84          | 1.23    | 4.78          |         |
| C6 <sup>+</sup> | 1.43  | -0.67 | 4.43  | 1.29 | 0.54    | -0.62 | 0.26  | 0.53 | 10.29       | 6.75    | 1.06       | 1.55    | 1.79          | 4.22    | 15.05         |         |

Os exemplos mostrados demonstram a grande variedade de limiares que pode ser estimada dependendo do método e da transformação escolhidos. Contudo, fica a questão: Qual desses métodos de estimação de limiar é o melhor? Se é que algum deles serve para esta função. Esta questão não pode ser respondida, porque desconhecemos a verdadeira distribuição estatística e espacial dos dados reais. Não obstante, a questão é bastante significativa, porque deduções por meio de estatística podem ser questionáveis. Entretanto, elas servem para orientar na exploração geoquímica e estabelecer critérios. No entanto, a determinação do limiar entre anomalias e *background* deve ser definida levando em consideração os objetivos da pesquisa, o tamanho e a forma esperados da anomalia, a variação natural prevista nas medidas de superfície utilizando-se de todos os dados disponíveis da região, conhecimentos adquiridos e interdisciplinares.

Na análise de continuidade espacial através da análise semivariográfica, verificamos que todas as variáveis apresentaram dependência espacial. Na geração dos mapas de contorno, verificou-se que a interpolação pelo método de krigagem ordinária resulta numa distribuição espacial muito mais homogênea e com um menor QME em relação aos demais métodos de interpolação utilizados, como podemos verificar para a variável metano (Figura 1). Esse resultado é obtido principalmente pelo fato desse interpolador ser não-viciado, com variância mínima. O desempenho desses interpoladores foi avaliado utilizando o critério do quadrado médio do erro (Tabela 3).

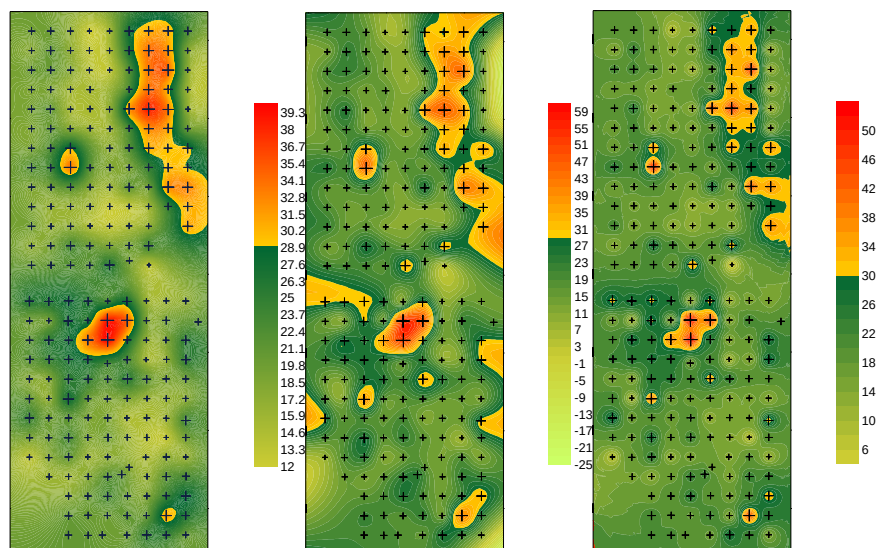


Figura 1 - Mapas de contorno interpolado para o metano (ppm) pelo método de Krigagem ordinária considerando o modelo esférico (a), pelo método de curvatura mínima (b) e inverso do quadrado da distância (c).

Tabela 3 – Valores do quadrado médio do erro para interpoladores usados na distribuição espacial da variáveis  $C_1$ ,  $C_2$  e  $C_3$ .

| Variável | Krigagem Ordinária | Inverso do Quadrado da Distância | Curvatura Mínima |
|----------|--------------------|----------------------------------|------------------|
|          | QME                |                                  |                  |
| $C_1$    | 51,29              | 65,5                             | 97,35            |
| $C_2$    | 13,55              | 17,31                            | 22,73            |
| $C_3$    | 20,31              | 21,74                            | 34,6             |

## Conclusões

Os dados geoquímicos de hidrocarbonetos gasosos não seguem de maneira geral uma distribuição lognormal ou normal, sendo uma exceção e não uma regra. Até quando outros diferentes métodos de transformação são usados, os dados não se aproximaram de uma distribuição normal. Portanto, não existe uma boa razão para continuar a usar a regra [média  $\pm$  2 desvios padrão (dp)].

A regra [mediana  $\pm$  2 desvios mediano absoluto], seguido do boxplot parecem ser mais adequados para determinar o valor limiar quando utilizamos dados geoquímicos, que são caracterizados por forte assimetria.

O 98º percentil parece não ser adequado para estimar valores de limiar, pois entregou a mesma quantidade de valores anômalos para qualquer tipo de distribuição.

O interpolador geoestatístico de krigagem ordinária apresenta melhor resultado que os demais interpoladores que ignoram a dependência espacial entre observações. Sendo possível indicar com maior precisão os locais de distribuição das variáveis analisadas, dando subsídios para delimitar áreas com possível potencial petrolífero e novas pesquisas com custo reduzido.

## Agradecimento

PEC – LAMCE – COPPE - UFRJ - ANP

## Referências Bibliográficas

- ABRAMS, M. A., 2005, “Significance of Hydrocarbon Seepage Relative to Petroleum Generation and Entrapment”. *Marine and Petroleum Geology*, 22: 457-477.
- ANDRIOTTI, J. L. S., 2003, *Fundamentos de Estatística e Geoestatística*. São Leopoldo: Unisinos. 165 p.
- BABINSKI, N. A., MELLO, M. R., 1984, *Geoquímica de Superfície*. In: *Geoquímica do Petróleo*. CENPES, PETROBRAS.
- CLARK, I. & HARPER, W. V., 2000, *Practical geostatistics*. Ecosse North America Llc, Colombus, Ohio, USA. 442p.
- CRESSIE, N. A. C., 1993, *Statistics for Spatial Data*. Revised Edition. John Wiley & Sons, Inc. 928p.
- DUTTER R., FILZMOSE P., GATHER U., ROUSSEEUW, P., 2003, *Developments in Robust Statistics. International Conference on Robust Statistics*, 2001. Heidelberg: Physik-Verlag.
- FRIZZO, S. J., LICHT, O. A. B., 2007, “Estatísticas Uni e Bivariadas Aplicadas à Prospecção Geoquímica”. In: LICHT, O. A. B., MELLO, C. S. B., Silva, C. R.(Eds), *Prospecção Geoquímica de Depósitos Minerais Metálicos, Não-Metálicos, Óleo e Gás*. SBGq - Sociedade Brasileira de Geoquímica / CPRM – Serviço Geológico do Brasil, Rio de Janeiro. Cap. 19: 595-618.
- ISAAKS, E. H., SRIVASTAVA, R. M., 1989, *An Introduction to Applied Geoestatistics*. New York: Oxford University Press. 561 p.
- LICHT. O. A . B., 1998, *Prospecção Geoquímica: Princípios, Técnicas e Métodos*. Rio de Janeiro: CPRM. 236 p.
- MIESCH, A.T., 1981, “Estimation of The Geochemical Threshold and its Statistical Significance”. *J. Geochem. Explor.* 16, 49–76.