

Artigo Técnico

Estatística descritiva e estacionaridade em variáveis geoquímicas ambientais

Descriptive statistics and stationarity in environmental geochemical variables

Henrique César Pereira Assumpção¹, Gisele Mara Hadlich²

RESUMO

O objetivo deste trabalho foi analisar os dados de atributos geoquímicos a fim de verificar sua estacionaridade e correlacionar a normalidade estatística com o uso da técnica de krigagem ordinária. A escolha da krigagem ordinária como método geoestatístico aplicado ao trabalho deve-se ao fato de essa ser aconselhada para a realização de estudos em áreas onde existam dados com variáveis que possam apresentar dependência espacial, como é o caso das variáveis geoquímicas, e por ser indicada para dados que apresentam estacionaridade. A metodologia utilizada para a realização desta pesquisa envolveu, além da revisão de literatura, a obtenção de dados dos metais-traço (Cu, Zn, Mn, Fe, Cr e Pb) extraídos parcialmente de amostras superficiais (0 a 10 cm) de solos e sedimentos coletados em campo. Também foram determinados os valores de pH, salinidade, nitrogênio total, fósforo, matéria orgânica e granulometria. Foram conduzidas análises estatísticas, construções de semivariogramas, aplicação da krigagem ordinária e, por fim, validação cruzada para medir a incerteza da medição prévia dos dados. Neste trabalho, por meio dos variogramas, comprovou-se que, apesar de os dados não serem normais, eles apresentaram estacionaridade. Além disso, o parâmetro da estatística descritiva que mais possui correlação direta com a krigagem ordinária é a variância.

Palavras-chave: geoestatística; krigagem ordinária; normalidade estatística; semivariograma.

ABSTRACT

The aim of this work was to analyze geochemical data in order to check their stationarity and to correlate the statistical normality using the ordinary kriging technique. The ordinary kriging technique was chosen as the geostatistical method applied to work because such technique is advised for studies in areas where there are data with variables that might present spatial dependence, like the geochemical variables, and also because it is indicated for data presenting stationarity. The methodology used for this research involved, besides literature review, data collection of trace metals (Cu, Zn, Mn, Fe, Cr and Pb) that were partially extracted from surface samples (0 to 10 cm) of soils and sediments collected in the field. We also determined the values of pH, salinity, total nitrogen, phosphorus, organic matter and particle size. Statistical analyzes, semivariogram development, ordinary kriging use and, lastly, cross validation were performed to measure the uncertainty of the previous measurement of data. It was found, in this work, by means of the variograms that although data were ordinary, they showed stationarity. In addition, the parameter of descriptive statistics that mostly correlates directly with the ordinary kriging is variance.

Keywords: geostatistics; ordinary kriging; statistical normality; semivariogram.

INTRODUÇÃO

A Geoestatística determina o grau de correlação entre pares de amostras com base na direção e distância entre elas. Portanto, extrai-se da aparente desordem e dos dados disponíveis uma imagem da variabilidade dos dados e uma medida da correlação existente entre os valores tomados em pontos no espaço, cuja análise é obtida por meio do semivariograma (ORTIZ, 2002).

Um dos problemas na delimitação do semivariograma é a presença de tendência nos dados avaliados, ou seja, a não estacionaridade da

variável, dificultando o ajuste de modelos adequados ao comportamento espacial da variável (VIEIRA, 1997; GUIMARÃES, 2004). Vieira *et al.* (1983) argumentam que quando a variável apresenta tendência, esta deve ser removida antes do ajuste do semivariograma.

A tendência consiste em uma característica das amostras, a qual pode ser detectada a partir da construção de um semivariograma omnidirecional, com direção angular igual a 0° e tolerância angular de 90° (ISAAKS & SRIVASTAVA, 1989), e por meio de técnicas que testam a normalidade dos dados, como o teste de Pearson e a estatística

¹Geólogo, Mestre em Geoquímica: Petróleo e Meio Ambiente; Professor substituto do Instituto de Geociências da Universidade Federal da Bahia (UFBA) - Salvador (BA), Brasil.

²Engenheira Agrônoma; Doutora em Geografia; Professora Associada do Instituto de Geociências da UFBA - Salvador (BA), Brasil.

Endereço para correspondência: Henrique César Pereira Assumpção - Rua Caetano Moura, 123 - Bairro Federação - 40201-340 - Salvador (BA), Brasil - E-mail: henri2ue@gmail.com

Recebido: 12/11/14 - **Aceito:** 26/07/16 - **Reg. ABES:** 142739

descritiva dos dados. Segundo Folegatti (1996), para o ajuste dos semivariogramas, a normalidade dos dados não é necessária, porém desejável. Caso a distribuição não seja normal, mas seja razoavelmente simétrica, pode-se admitir as hipóteses necessárias à construção do semivariograma. Em contrapartida, Vieira e Lombardi Neto (1995) afirmam que os cálculos utilizados em Geoestatística não requerem o conhecimento da distribuição de frequências da variável analisada.

Os testes empregados para verificar a distribuição normal dos dados em uma análise geoestatística exploratória têm por objetivo direcionar o pesquisador a saber sobre a tendência dos dados.

Espera-se, com este trabalho, analisar a estacionaridade das variáveis ambientais e verificar se há algum parâmetro da Estatística Descritiva que se correlaciona diretamente com o indicativo de qualidade quanto às estimativas geradas para cada variável. Os estudos foram conduzidos na ilha sede do município de Madre de Deus, localizado na região Norte da Baía de Todos-os-Santos, no estado da Bahia.

METODOLOGIA

Para a realização desta pesquisa, a metodologia usada constou da revisão de literatura, da obtenção de dados de metais-traço nos ambientes estudados e, posteriormente, da aplicação de técnicas estatísticas e geoestatísticas para análise dos dados. Os trabalhos foram realizados em diferentes etapas contemplando campo, laboratório e trabalho em escritório. Na fase de campo, foram coletadas 96 amostras envolvendo os três ambientes estudados em superfície (0 a 10 cm de profundidade), com colheres plásticas previamente lavadas e ambientadas. Antes da coleta, retiraram-se os materiais superficiais que se encontravam soltos, tais como folhas, cascalhos, entre outros. Essas amostras foram encaminhadas para o Laboratório de Estudos Ambientais no Núcleo de Estudos Ambientais do Instituto de Geociências da Universidade Federal da Bahia, no qual foram realizadas as análises químicas.

A partir dos resultados obtidos com as análises químicas, criou-se uma planilha eletrônica com os dados pelo programa de computador Microsoft Excel, a qual continha, para cada ponto amostrado, o dado espacial (coordenadas *Universal Transversal Mercator* – UTM) e as variáveis analisadas.

Para se verificar a normalidade, os dados numéricos coletados inicialmente foram submetidos ao teste D'Agostino-Pearson, que é baseado no fato de que quando os dados são normalmente distribuídos, eles têm distribuição no χ^2 com 2 graus de liberdade, não podendo ser utilizados para os conjuntos de dados com menos de 20 elementos.

A hipótese nula do teste D'Agostino-Pearson é que os dados possuem a distribuição normal; dessa forma, valores p superiores a 0,05 indicam que os dados apresentam a distribuição normal. Sendo assim,

foram considerados valores p inferiores a 0,05 significantes (não apresentando normalidade), e foi realizada uma análise estatística descritiva para verificar o tipo de distribuição. Ambos foram conduzidos e analisados no programa BioEstat, versão 5.0 (AYRES *et al.*, 2007).

O cálculo da semivariância foi realizado seguindo a Equação 1:

$$\gamma(h) = \frac{\sum (x_i - \mu)^2}{2n} \quad (1)$$

Em que:

$\gamma(h)$: semivariância das amostras em função da distância h entre elas;

x : valor das amostras;

μ : média das amostras;

n : número de amostras.

A partir dos valores de $\gamma(h)$ e de h constrói-se o semivariograma.

Após a geração do semivariograma, faz-se necessário ajustar a curva da função a um modelo conhecido (Figura 1) a fim de determinar os parâmetros de efeitos pepita, alcance e patamar. Exemplos dos modelos mais utilizados em Geoestatística são: esférico, gaussiano e exponencial (DRUCK, 2002).

Os procedimentos para a análise do variograma foram realizados de acordo com Vieira *et al.* (1983), e as análises foram feitas no programa GS+, versão 7 (GAMMA DESIGN SOFTWARE, 2004). A utilização da krigagem ordinária e da validação cruzada foram realizadas no programa ArcGis 10.

A krigagem é um conjunto de técnicas de regressão linear generalizadas para minimizar uma variância de estimação a partir de um modelo de covariância definido *a priori* (DEUTSCH & JOURNAL, 1992).

Alguns pesquisadores têm trabalhado a aplicação de krigagem indicativa (BASTANTE *et al.*, 2005; SALGUEIRO *et al.*, 2008; LEE *et al.*, 2007), de krigagem fatorial e da cokrigagem. Neste trabalho,

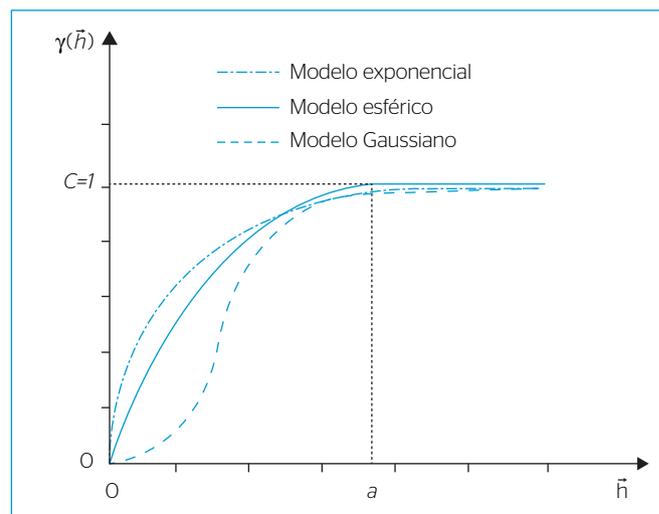


Figura 1 - Representação gráfica dos modelos (Druck *et al.*, 2002).

será utilizada a krigagem ordinária, que, segundo Landim (2006) e conforme constatado em Jung *et al.* (2006) e Zhou *et al.* (2007), é uma das técnicas mais utilizadas dentre os métodos geoestatísticos conhecidos, principalmente quando se tratam de estudos envolvendo as variáveis ambientais.

A estimativa por krigagem ordinária de um atributo Z , na posição i e com coordenadas dadas pelo vetor x , é apresentada pela Equação 2 (ISAAKS & SRIVASTAVA, 1989):

$$Z(x_i) = \sum_{i=1}^n Z(x_i) \cdot \lambda_i \quad (2)$$

Em que:

λ_i : são as ponderações aplicadas aos i -ésimos valores dos atributos que possuam dependência espacial com o valor da variável a ser estimado na posição x_i .

Para a determinação do erro associado ao modelo geoestatístico criado, utilizou-se a validação cruzada para medir a incerteza da medição prévia dos dados (Almeida *et al.*, 2007).

Segundo Isaaks e Srivastava (1989), na etapa de validação cruzada, cada ponto medido é excluído e seu valor é estimado levando-se em conta os dados restantes. Nessa técnica, são estimados os valores dos parâmetros analisados e individualizados dos pontos amostrados, sendo em seguida comparados aos novos totais dos dados medidos.

Caso os resultados da validação cruzada apresentem os dados sobre uma reta de regressão próxima da bisetrix e com pequena dispersão dos pontos em torno desta reta, pode-se confiar na estimativa,

sendo que o valor médio das diferenças será tão mais próximo de zero quanto melhor for a estimativa. A validação cruzada pode ser utilizada também para avaliar se o modelo de variograma escolhido está bem ajustado aos dados disponíveis, sendo que um ajuste pobre significa a presença de viés (ANDRIOTTI, 2002).

RESULTADOS E DISCUSSÃO

A partir das variáveis ambientais e da distribuição dos elementos geoquímicos, realizou-se uma análise estatística descritiva com a finalidade de observar a normalidade dos dados (Tabela 1). Sabe-se que o coeficiente de variação alto (maior que 1) e o de assimetria e curtose que se afastam do valor zero podem ser indicativos da não normalidade. No entanto, também foi realizado o teste de normalidade D'Agostino-Pearson, método mais robusto e mais confiável na determinação de normalidade.

Verificou-se, assim, que alguns parâmetros apresentaram assimetria e curtose abaixo de zero e coeficiente de variação acima do valor 1, indicando uma provável normalidade, apenas o pH apresentou normalidade ao ser aplicado o teste de D'Agostino-Pearson ($p > 0,05$).

Com a finalidade de observar os efeitos dos testes descritos em uma análise geoestatística, foram produzidos variogramas para cada variável. O modelo que melhor se ajustou aos dados do atributo fósforo (P) foi o esférico, apresentando o maior coeficiente de determinação (R^2) em relação aos outros.

O melhor modelo para as outras variáveis foi o exponencial, apresentando R^2 superior a 0,1 para todos os atributos e a menor soma de quadrados do resíduo, quando comparado aos demais modelos testados.

Tabela 1 - Estatística descritiva das variáveis geoquímicas estudadas.

Dados estatísticos	pH	Salinidade	Areia (%)	Silte (%)	Argila (%)	N (%)	Matéria orgânica (%)	Valor p (%)	Fe (mg.kg ⁻¹)	Mn (mg.kg ⁻¹)	Cr (mg.kg ⁻¹)	Cu (mg.kg ⁻¹)	Pb (mg.kg ⁻¹)	Zn (mg.kg ⁻¹)
Número de amostras	96	96	96	96	96	96	96	96	96	96	96	96	96	96
Valor mínimo	4,44	0,00	3,50	2,72	0,00	0,04	0,25	0,00	663,2	3,15	0,05	0,05	0,00	4,19
Valor máximo	8,38	100,00	97,28	92,99	3,74	0,64	11,43	1715,69	37446,6	844,52	29,16	39,79	27,34	61,81
Amplitude total	3,94	100,00	93,78	90,27	3,74	0,60	11,18	1715,69	36783,3	841,37	29,11	39,74	27,34	57,62
Mediana	6,82	5,00	79,38	19,99	0,49	0,11	2,00	48,28	2645,2	32,75	3,00	7,01	0,38	15,75
Média aritmética	6,59	30,87	72,32	27,00	0,78	0,14	2,44	130,91	4528,14	95,77	4,57	10,14	2,86	19,75
Desvio padrão	1,12	36,13	23,85	23,16	0,84	0,10	1,88	244,03	5348,60	167,02	5,19	9,37	5,51	12,33
Valor p*	0,07	0,01	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001

*estatística do teste de normalidade D'Agostino-Pearson.

Segundo Robertson (1998), a soma de quadrados do resíduo é um parâmetro mais robusto do que o R^2 e propicia uma medida mais exata do modelo que se ajusta aos dados, sendo os dois parâmetros usados na escolha do melhor entre eles.

Conforme observado na Figura 2, todos os variogramas construídos apresentaram patamar bem definido, demonstrando que o grau de aleatoriedade é baixo e que existe dependência espacial influenciando nos dados. Esse resultado comprova a estacionaridade dos dados e justifica o uso da krigagem para a análise geoestatística das amostras estudadas.

Após aplicar a técnica da krigagem ordinária, utilizando o modelo exponencial e um alcance de 337 m para todos os variogramas, foi utilizada a validação cruzada para avaliar a incerteza associada às estimativas obtidas. A Tabela 2 apresenta a média das diferenças entre os valores medidos e os estimados pela krigagem (erro médio), realizada pela validação cruzada, juntamente com os dados obtidos por meio da estatística descritiva. Observa-se que os maiores valores de erro médio estão associados às maiores variâncias, e o chumbo (Pb), apesar de apresentar um alto valor de coeficiente de variação, curtose e assimetria, teve um baixo erro médio, devido a sua pouca variância.

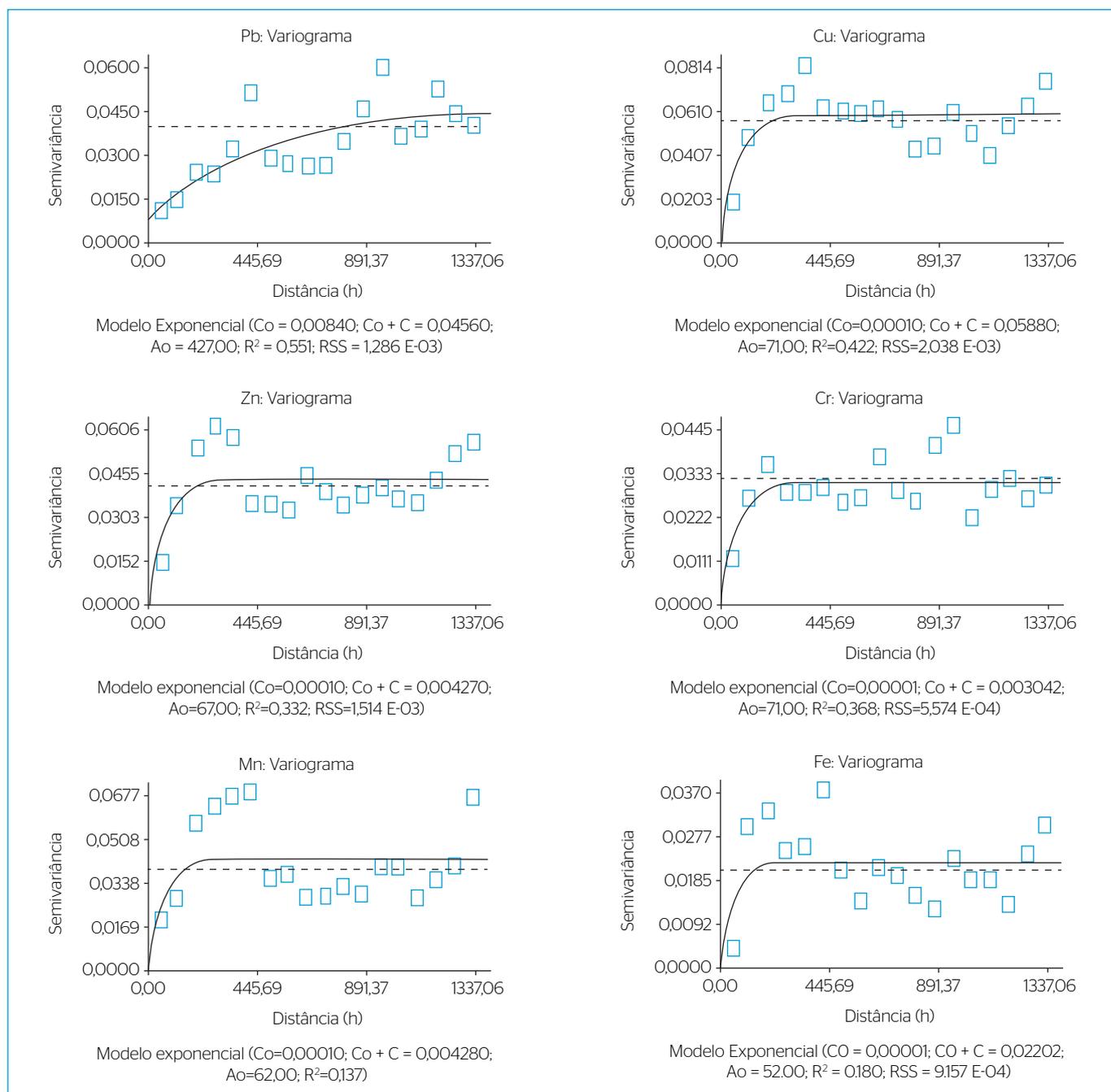


Figura 2 - Semivariogramas omnidirecionais das variáveis geoquímicas analisadas.

Continua...

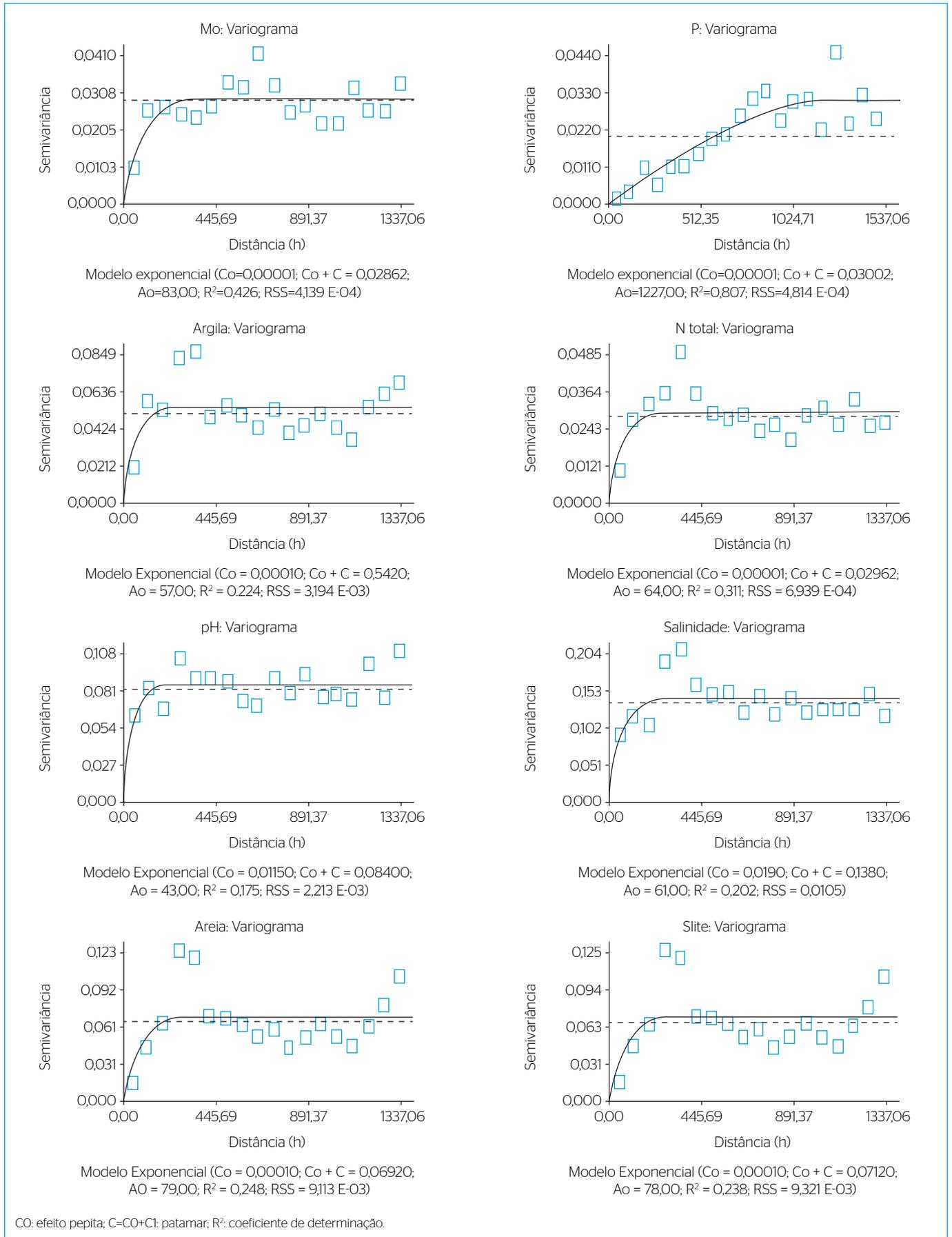


Figura 2 - Continuação.

Tabela 2 - Parâmetros da análise estatística descritiva e do erro médio associado à modelagem da krigagem ordinária das variáveis geoquímicas estudadas.

Variáveis	Erro médio (validação cruzada)	Variância	Coefficiente de variação	Assimetria (g1)	Curtose (g2)
pH	0,05	1,2539	0,1698	-0,3699	-1,076
Salinidade	1,40	1305,55	1,17	0,82	-0,82
n total	0,00	0,01	0,73	2,37	7,90
Matéria orgânica	0,04	3,52	0,77	1,94	5,30
P	6,40	59550,84	1,86	4,36	23,15
Fe	107,87	28608000,00	1,18	3,60	17,21
Mn	2,3	27894,01	1,74	2,92	8,60
Cr	0,40	26,89	1,13	2,04	5,72
Cu	0,90	87,84	0,92	1,69	2,26
Pb	0,05	30,33	1,92	2,88	8,84
Zn	0,60	152,08	0,62	1,22	1,19
Areia	0,80	568,95	0,33	-1,62	1,92
Silte	0,80	536,47	0,86	1,61	1,87
Argila	0,01	0,71	1,08	1,54	2,14

Tabela 3 - Matriz da correlação de Pearson ($p < 0,05$) dos parâmetros obtidos na análise descritiva e do erro médio associado à krigagem ordinária dos dados das variáveis estudadas.

Parâmetros	Erro médio (validação cruzada)	Variância	Coefficiente de variação	Assimetria (g1)	Curtose (g2)
Erro médio (validação cruzada)	1,00	-	-	-	-
Variância	1,00	1,00	-	-	-
Coefficiente de variação	0,10	0,08	1,00	-	-
Assimetria (g1)	0,36	0,34	0,79	1,00	-
Curtose (g2)	0,50	0,47	0,65	0,81	1,00

Para a comprovação da relação existente entre a variância e o erro médio estimado pela validação cruzada, utilizou-se uma matriz de correlação (Tabela 3). Observou-se, então, que a variância apresenta uma correlação direta com o erro médio, durante a etapa da validação cruzada. No entanto, o coeficiente de variação, a assimetria e a curtose apresentaram baixa correlação com a validação cruzada, comprovando assim que a normalidade dos dados não está diretamente relacionada com uma boa modelagem geoestatística utilizando a técnica da krigagem ordinária.

CONCLUSÕES

Apesar da maioria das variáveis analisadas não apresentarem normalidade, os variogramas gerados demonstraram que os dados possuíam estacionaridade, indicando, portanto, que é preciso um estudo prévio dos dados para se obter um bom resultado na modelagem geoestatística.

A variância apresentou-se como o melhor parâmetro da estatística descritiva para ser investigado quando se utiliza a técnica da krigagem ordinária.

REFERÊNCIAS

- ALMEIDA, R.; BERNARDI, J.V.E.; BASTOS, W.R.; NASCIMENTO, E.L.; OLIVEIRA R.C.; CARVALHO D.P. (2007) Método geoestatístico para modelagem ambiental de poluentes em sistemas lacustres – Amazônia Ocidental. In: 13º Simpósio Brasileiro de Sensoriamento Remoto, *Anais...*, Florianópolis, p. 2247-2253.
- ANDRIOTTI, J.L.S. (2002) Notas de geoestatística. *Acta Geologica Leopoldensia*, v. 25, n. 45, p. 03-14.
- AYRES, M.; AYRES, J.R.M.; AYRES, D.L.; SANTOS, A.S. (2007) *BioEstat 5.0*: aplicações estatísticas nas áreas das ciências biológicas e médicas. Belém: CNPQ, 290p.
- BASTANTE, F.G.; TABOADA, J.; ALEJANO, L.R.; ORDONEZ, C. (2005) Evaluation of the resources of a slate deposit using indicator kriging. *Engineering Geology*, v. 81, p. 407-418.
- DEUTSCH, C.V. & JOURNEL, A.G. (1992) *GSLIB: Geostatistical Software Library and user's guide*. New York: Oxford University Press, 339p.
- DRUCK, S. (2002) A geoestatística e os sistemas de informações geográficas. *Espaço & Geografia*, v. 5, n.1, p. 01-15.
- FOLEGATTI, M.V. (1996) *Estabilidade temporal e variabilidade espacial da umidade e do armazenamento de água em solo siltoso*. Tese (Doutorado) – Universidade de São Paulo, Piracicaba, 84p.
- GAMMA DESIGN SOFTWARE. (2004) *GS+ - Geostatistics for the Environmental Sciences*. Michigan, 1 CD-ROM, versão 9.0.
- GUIMARÃES, E.C. (2004) *Geoestatística básica e aplicada*. 77p. Disponível em: <files.ednaldoufu.webnode.com.br/200000109-9dfa99ef3f/apgeo1.pdf>. Acesso em: 10 dez. 2013.
- ISAAKS, E.H. & SRIVASTAVA, R.M. (1989) *An introduction to applied geostatistics*. New York: Oxford University Press, 560p.
- JUNG, K.; STELZENMÜLLER, V.; ZAUKE, G.P. (2006) Spatial distribution of heavy metal concentrations and biomass indices in *Cerastoderma edule* Linnaeus (1758) from the German Wadden Sea: an integrated biomonitoring approach. *Journal of Experimental Marine Biology and Ecology*, v. 338, p. 81-95.
- LANDIM, P.M.B. (2006) Sobre geoestatística e mapas. *Terra e Didática*, v. 2, n. 1, p. 19-33.
- LEE, J.J.; JANG, C.S.; WANG, S.W.; LIU, C.W. (2007) Evaluation of potential health risk of arsenic-affected groundwater using indicator kriging and dose response model. *Science of the Total Environment*, v. 384, n. 1-3, p. 151-162.
- ORTIZ, G.C. (2002) *Aplicação de métodos Geoestatísticos para identificar a magnitude e a estrutura da variabilidade espacial de variáveis físicas do solo*. Dissertação (Mestrado) – Universidade de São Paulo, Piracicaba, 75p.
- SALGUEIRO, A.R.; ÁVILA, P.F.; PEREIRA, H.G.; OLIVEIRA, J.M.S. (2008) Geostatistical estimation of chemical contamination in stream sediments: the case study of Vale das Gatas mine (northern Portugal). *Journal of Geochemical Exploration*, v. 98, p. 15-21.
- VIEIRA, S.R. (1997) Variabilidade espacial de argila, silte e atributos químicos em parcela experimental de um Latossolo Roxo de Campinas (SP). *Bragantia*, v. 56, n. 1, p. 181-190.
- VIEIRA, S.R.; HATFIELD, J.L.; NIELSEN, D.R.; BIGGAR, J.W. (1983) Geoestatistical theory and application to variability of some agronomical properties. *Hilgardia*, v. 51, p. 01-75.
- VIEIRA, S.R. & LOMBARDI NETO, F. (1995) Variabilidade espacial do potencial de erosão das chuvas do Estado de São Paulo. *Bragantia*, v. 54, n. 2, p. 405-412.
- ZHOU, F.; GUO, H.; HÃO, Z. (2007) Spatial distribution of heavy metals in Hong Kong's marine sediments and their human impacts: a GIS-based chemometric approach. *Marine Pollution Bulletin*, v. 54, p. 1371-1384.