



Comparação de métodos para ajuste de modelos de semivariograma da precipitação pluvial anual

José R. P. de Carvalho¹, Sidney R. Vieira² & Célia R. Grego³

RESUMO

Comumente dados de precipitação pluvial apresentam variação e a obtenção da estimativa de sua distribuição espacial é primordial no planejamento agrícola e ambiental. O objetivo neste trabalho foi comparar o método de estimação dos mínimos quadrados ponderados para ajuste de modelos ao semivariograma com o método de tentativa e erro, através da técnica de auto-validação "jack-knifing", para dados de precipitação pluvial média anual do Estado de São Paulo. Observações de precipitação correspondentes ao período de 1957 a 1997 foram usadas para trezentos e setenta e nove (379) estações pluviométricas abrangendo todo o Estado de São Paulo, representando uma área de aproximadamente 248.808,8 km². A periodicidade exibida pelos semivariogramas foi ajustada pelo modelo "hole effect", em que os parâmetros foram estimados com maior precisão pelo método de mínimos quadrados ponderados quando comparado com o método de tentativa e erro. O método de auto-validação "jack-knifing" mostrou-se adequado para a definição de métodos e modelos a serem usados para semivariâncias, cujo procedimento permitiu definir dezesseis vizinhos como o número ideal para a estimativa por krigagem de valores de precipitação pluvial para locais não amostrados no Estado de São Paulo.

Palavras-chave: periodicidade, modelo "hole effect", auto-validação, "jack-knifing"

Comparison of methods for adjusting semivariogram model of annual rainfall

ABSTRACT

Rainfall data presents variability and the attainment of the estimate of its space distribution is essential in agricultural and environmental planning. The objective of this work is to compare the weighted least squares estimation method for semivariogram modeling with the trial and error fitting estimation method through the jack-knifing validation technique to annual average rainfall in the São Paulo State. Rainfall observations corresponding to the period from 1957 to 1997, from 379 rain gauge stations covering entire São Paulo State, representing an area of approximately 248,808.8 km² were used in this study. The parameters for the hole effect model, used to describe the periodicity exhibited by the semivariograms, were more precisely estimated by using the weighted least square estimation method when compared with the manual fitting method. The jack-knifing technique showed itself to be an adequate tool for the definition of methods and models to be used for semivariance. It also allowed for the definition of 16 neighbors as the ideal number to estimate by kriging the unsampled locations.

Key words: periodic, hole effect model, jack-knifing, validation method

¹ EMBRAPA/Centro Nacional de Pesquisa Tecnológica em Informática na Agricultura, Avenida Zeferino Vaz 1328, CEP 13083-886, Barão Geraldo, Campinas, SP. Fone (19) 3789 8575, fax: (19) 3789 8575. E-Mail: jruy@cnptia.embrapa.br

² Instituto Agrônomo, C.P. 28, CEP 13001-970, Campinas SP. Fone (19) 3241 5188, ext. 302, fax (19) 3231 4943. E-Mail: sidney@iac.sp.gov.br

³ EMBRAPA/Centro Nacional de Pesquisa de Monitoramento por Satélite, Av. Soldado Passarinho, 303, Jardim Chapadão, CEP 13070-115 Campinas, SP. E-mail: crgrego@cnpm.embrapa.br

INTRODUÇÃO

A experimentação agrônômica tem como uma de suas grandes preocupações, testar diferenças entre tratamentos enquanto a estatística clássica tem sido, normalmente, a ferramenta mais utilizada para este fim; entretanto e dependendo do grau de variabilidade da variável estudada, o coeficiente de variação pode ser tão alto a ponto de mascarar a análise estatística do experimento. Esta variabilidade é devida, muitas vezes, ao solo que, a princípio e por hipótese da estatística clássica é considerado homogêneo. Por outro lado, o coeficiente de variação expressa somente a variabilidade das medidas sem considerar a posição geográfica de cada valor. Esta informação só é possível se uma ferramenta estatística apropriada for usada considerando-se as coordenadas geográficas. Geoestatística é a ferramenta adequada para esta análise, desde que valores separados por distâncias pequenas sejam mais semelhantes uns aos outros do que aqueles separados por distâncias maiores (Druck et al., 2004); tal como as propriedades do solo, dados de precipitação pluviométrica, foco deste trabalho, são fontes de variação e a obtenção da estimativa de sua distribuição espacial é primordial no planejamento agrícola e ambiental (Vieira & Carvalho, 2001; Carvalho & Assad, 2002).

O semivariograma é a ferramenta básica na estimação através da geoestatística, haja vista que expressa o grau de dependência espacial entre amostras dentro de um campo experimental e permite a estimativa dos parâmetros com os quais os valores não amostrados são estimados por meio da técnica de interpolação conhecida como krigagem, permitindo a construção de mapas de isolinhas ou tridimensionais para exame e interpretação da variabilidade espacial (Vieira, 2000; Carvalho et al., 2002).

Na prática, a estimação do semivariograma experimental é realizada com mais precisão para distâncias próximas da origem e seu ajuste pode ser feito adotando-se o método por tentativa e erro (Journel & Huijbregts, 1978) usando-se um modelo selecionado pelo usuário, normalmente baseado na sua aparência visual. Segundo McBratney & Webster (1986), em vários trabalhos na área de ciência do solo a função matemática e o procedimento de ajuste dos parâmetros, se mencionados, foram simplesmente declarados de forma subjetiva, em que, praticamente em todos os casos os semivariogramas foram estimados de forma visual. Outros métodos de ajuste sem subjetividade, também foram e estão sendo estudados destacando-se os métodos dos quadrados mínimos ordinários e ponderados; as suas particularidades e pressuposições podem ser encontradas em Cressie (1985; 1991).

O objetivo deste trabalho é aplicar o método de mínimos quadrados ponderados de estimação de parâmetros e compará-lo com o método de tentativa e erro, muito usado na prática, através da técnica de validação “jack-knifing” para dados de precipitação média anual para o Estado de São Paulo, no período de 1957 a 1997.

MATERIAL E MÉTODOS

Usaram-se 379 observações de precipitação pluviométrica média anual, obtidas de estações pluviométricas (Figura 1) abran-

gendo todo o Estado de São Paulo, representando uma área de aproximadamente 248.808,8 km² (2,91% do território nacional), referente ao período de 1957 a 1997 (DAEE, 2005).



Figura 1. Localização das estações pluviométricas no Estado de São Paulo

Os dados foram analisados, usando a estatística descritiva para verificar principalmente a discrepância e a normalidade de distribuição de frequência dos dados.

O semivariograma é uma função matemática definida para representar o nível de dependência entre duas variáveis aleatórias regionalizadas locais. O nível de dependência espacial entre essas duas variáveis regionalizadas, é representado pelo semivariograma estimado pela equação da semivariância (Eq.1) segundo Vieira (2000).

$$\hat{\gamma}(h) = \frac{1}{2 N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (1)$$

em que $N(h)$ é o número de pares de valores medidos $Z(x_i)$, $Z(x_i+h)$, separado pela distância h . O gráfico de $\hat{\gamma}(h)$ versus valores correspondentes de h , é chamado semivariograma. Diversos tipos de modelo estão disponíveis na literatura, em que os mais usados são os modelos linear, esférico, exponencial e gaussiano que, por apresentarem positividade definida condicional (Journel & Huijbregts, 1978; McBratney & Webster, 1986), permitem visualizar a natureza da variação espacial das variáveis estudadas. Para esses modelos, os semivariogramas freqüentemente aumentam com a distância, até atingir um patamar, e são denominados monotônicos crescentes. Entretanto, os semivariogramas não são restritos a uma estrutura monotônica; eles podem apresentar segmentos decrescentes ou cíclicos, os quais são chamados “hole effects”; referidas estruturas não monotônicas podem ou não apresentar patamar; ter as amplitudes de ondas reduzidas e serem isotrópicos ou anisotrópicos (Pyrcz & Deutsch, 2005).

O modelo matemático seguinte, “hole effect” (Eq.2) pode ser ajustado ao semivariograma:

$$\gamma(h) = C_0 + C_1 \left[1 - \frac{\sin(h/a)}{h/a} \right] \quad h \geq 0 \quad (2)$$

em que C_0 é o efeito pepita, $C_0 + C_1$ é o patamar, h a distância e a é o alcance do semivariograma.

Os métodos analíticos de ajustes de modelos aos semivariogramas podem ser enquadrados em duas categorias (Jian et al., 1996): métodos de máxima verossimilhança e mínimos quadrados. O método de máxima verossimilhança é altamente dependente das pressuposições da distribuição gaussiana obtendo-se, muitas vezes, estimativas viciadas dos parâmetros populacionais (Cressie, 1991).

Para o método de estimação de mínimos quadrados ordinários, os valores desconhecidos do vetor de parâmetros q (C_0 , C_1 e a) da função de semivariograma $\gamma(h, \theta)$, são estimados calculando-se os valores numéricos para os parâmetros que minimizam a soma de quadrados dos desvios entre as respostas estimadas e as observadas do modelo (Draper & Smith, 1981). Matematicamente, a função objetivo pode ser escrita como:

$$Q = \sum_h [\hat{\gamma}(h) - \gamma(h, \theta)]^2 \quad (3)$$

em que $\hat{\gamma}(h)$ denota o semivariograma experimental (Eq. 1), $\gamma(h, \theta)$ o modelo de semivariograma ajustado e q o vetor de parâmetros. As estimativas dos parâmetros são obtidas minimizando-as (Eq. 3), e seguindo a metodologia abaixo:

- Determinar as derivadas parciais de Q em relação a C_0 , C_1 e;
- Igualar a zero as equações de derivadas parciais;
- Resolver o sistema de equações resultantes.

A pressuposição básica para se obter estimativas não viciadas pelo método de estimação de mínimos quadrados ordinários é a de que o desvio padrão do erro seja constante para as variáveis dependentes e independentes do modelo a ser usado. Esta pressuposição, entretanto, nem sempre acontece, pois os valores de semivariâncias para diferentes distâncias de separação não possuem a mesma variância, em virtude de se basearem em diferentes números de pares. Para que se maximize a estimação dos parâmetros (Stasoft, 2004), o método de mínimos quadrados ponderados é normalmente usado, conforme a Eq. 4:

$$Q_1 = \sum_h W_h [\hat{\gamma}(h) - \gamma(h, \theta)]^2 \quad (4)$$

em que W_h é o vetor de peso usado. Jian et al. (1996) sugerem o uso do desvio padrão das estimativas de semivariâncias como peso para determinar quanto cada observação, no conjunto de dados, influencia a estimativa dos parâmetros. Uma grande vantagem deste método é sua capacidade de tratar pequenos conjuntos de dados; além disso, os pesos enfatizam mais as estimativas de semivariâncias para um grande número de pares e também quando os valores do semivariograma estão próximos à origem, que é a parte mais crítica do ajuste.

Para o método de mínimos quadrados ponderados, o modelo foi estimado pelo método iterativo de Gauss-Newton (SAS, 1985), sendo necessárias as especificações dos valores iniciais dos parâmetros para a estimativa do modelo, usando-se uma única variável dependente e se definindo a derivada parcial do modelo em relação aos parâmetros da estimativa.

A qualidade de ajuste dos modelos matemáticos aos semivariogramas usando-se o método dos mínimos quadrados ponderados ou o de tentativa e erro, pode ser verificada através da técnica de “jack-knifing”. Como se tem para cada local, um valor medido e se pode estimar um outro valor através da krigagem ou co-krigagem, então se pode, também, calcular a regressão linear entre esses pares de dados e a interseção (a), o coeficiente angular (b), a correlação entre os pares (r) e o erro reduzido com sua média e variância (Vieira et al., 1983, Vieira, 1997). O melhor ajuste se obtém quando os valores obtidos se aproximam dos seguintes valores ideais: a igual a zero (0), b igual a um (1), r igual a um, média do erro reduzido igual a zero e variância do erro reduzido igual a um. Definido o modelo matemático, os interpoladores de krigagem são usados e os pontos estimados com as propriedades de serem não viciados e com variância mínima (Vieira et al., 1983; Carvalho et al., 2002; Carvalho & Assad, 2003) e ideais para a construção de mapas de isolinhas ou tridimensionais.

RESULTADOS E DISCUSSÃO

Os dados anuais de precipitação correspondem a médias para o período de 1957 a 1997; todas as análises foram realizadas nessas médias sem a preocupação com as flutuações interanuais. As informações da estatística descritiva estão na Tabela 1. O coeficiente de variação apresenta-se baixo e, com os valores de simetria e curtose, distantes de 0 (zero), não ocorre normalidade de distribuição de frequência nos dados, estando de acordo, portanto, com os resultados de Carvalho & Assad (2003) para 1027 postos de medição de precipitação também no Estado de São Paulo.

O modelo “hole effect” foi ajustado aos semivariogramas experimentais para precipitação média anual obtida pelo método de tentativa e erro e mínimos quadrados ponderados, em que os resultados são apresentados nas Figuras 2A e 2B, respectivamente. Por serem periódicos, os semivariogramas obtidos para dados de precipitação pluvial média anual para o Estado de São Paulo, são candidatos naturais para o ajuste deste tipo de estrutura. A periodicidade aqui expressa representa a repetição da variabilidade em pequenos espaços o que segundo Carvalho & Assad (2002) ocorre principalmente na região da Serra do Mar, em virtude das diferenças de altitude. Segundo Bega et al. (2005) outro fator pode estar relacionado a esta periodicidade, como a grande variação da distância entre os pontos de medição: ignorando essas estruturas não

Tabela 1. Estatística descritiva da precipitação pluvial média anual do Estado de São Paulo, referente ao período de 1957 a 1997

Pontos de amostragem	Média	Variância	Desvio padrão	Coefficiente de variação	Valor mínimo	Valor máximo	Simetria	curtose
379	1450,00	54070,00	232,50	16,03	1156	3168	2,86	12,52

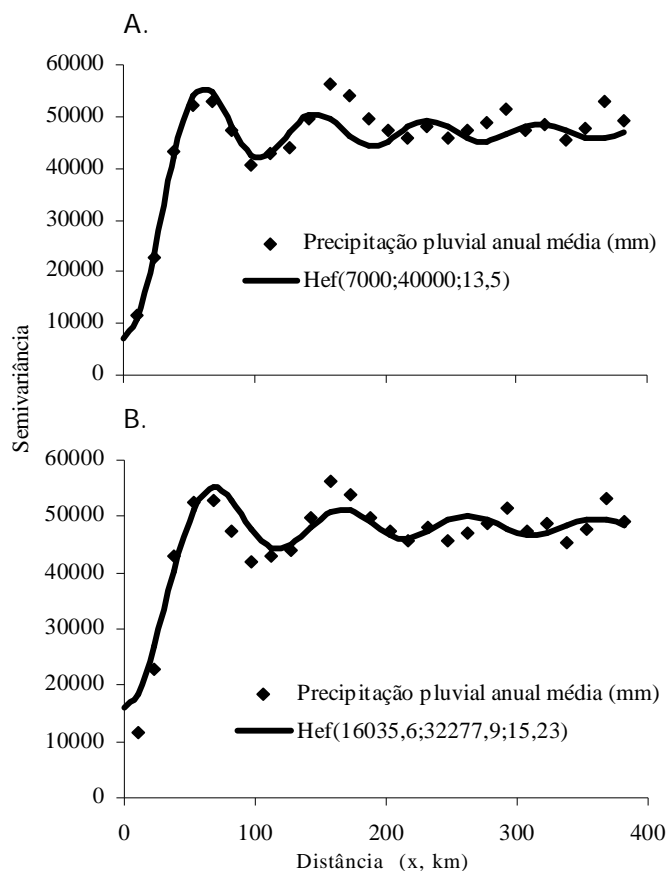


Figura 2. Modelo "hole effect" ajustado aos semivariogramas. (A) Método de tentativa e erro; (B) Método dos mínimos quadrados ponderados

monotônicas, conclui-se que este evento pode resultar em modelos não realísticos os quais, por sua vez, não reproduzem as variabilidades espaciais observadas.

Os efeitos pepita $C_0 = 7000$ e $C_0 = 16035,6$ significam que existe descontinuidade entre valores separados por distância menores que o usado no intervalo de amostragem. A proporção desses valores para o patamar do semivariograma ($C_0 + C_1$), no caso, 14,9 e 33,19%, é um indicativo da quantidade de variação ao acaso, de um ponto para outro, e quanto menor seu valor mais parecidos são os valores vizinhos; outro fato a ser considerado é que, para o método de mínimos quadrados ponderados, o vetor de peso reconhecido foi o desvio padrão. Um outro funcional escolhido poderia evitar a maior ausência de ajuste ocorrida para baixas semivariâncias neste método comparado com o de tentativa e erro.

Pode-se notar fortes flutuações até os 200 km de distância na Figura 2, causadas por variações periódicas da chuva nesta distância. Acredita-se que tal flutuação se deva aos índices pluviométricos obtidos na região litorânea, onde a precipitação pluviométrica apresenta distribuição espacial distinta das demais regiões do Estado, como constatado por Carvalho & Assad (2002). A precipitação aumenta extremamente rápido em uma pequena distância (cerca de 50 km que é a distância média do litoral ao ponto mais alto e de mais alta precipitação da Serra do Mar) e diminui novamente a medida que se afasta da Serra do Mar em direção ao interior.

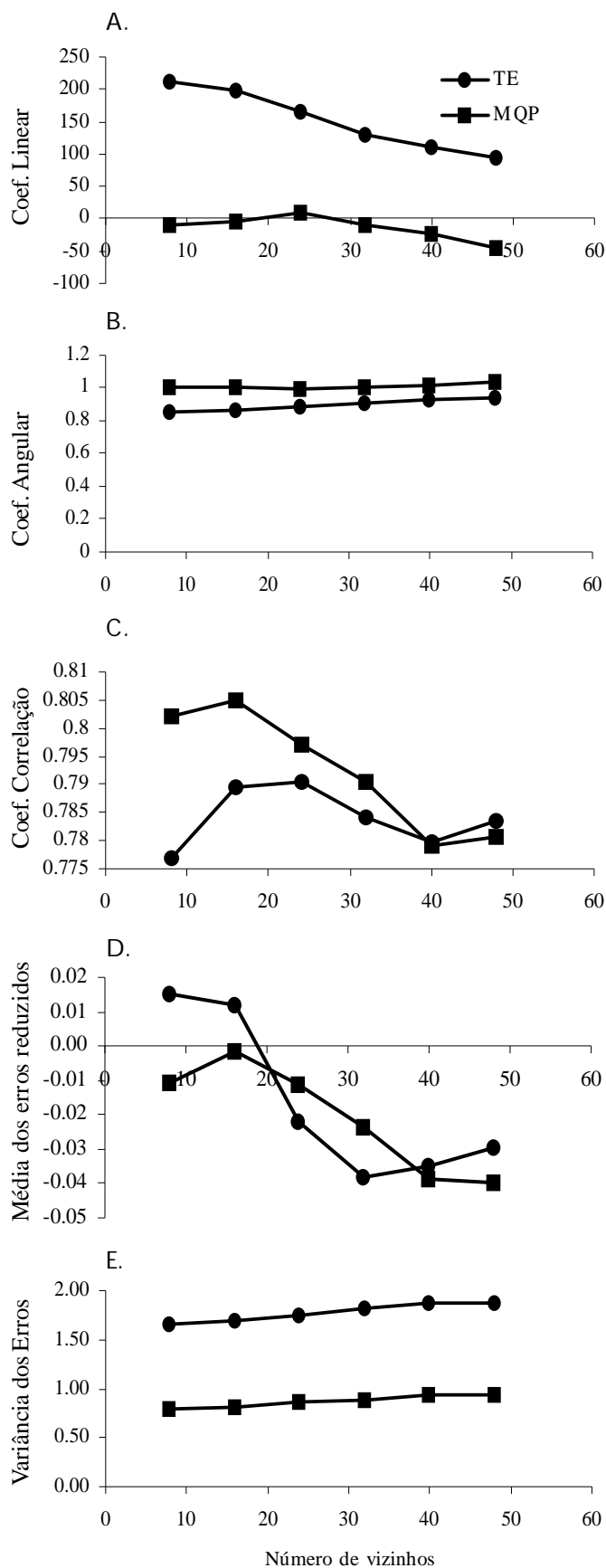


Figura 3. Resultados de jack-knifing para os métodos de tentativa e erro (TE) e mínimos quadrados ponderados (MQP). Gráfico de indicadores (A - coeficiente linear; B - coeficiente angular; C - coeficiente de correlação; D - média dos erros reduzidos e E - variância dos erros reduzidos versus número de vizinhos)

O método de tentativa e erro ao mesmo tempo em que é mais dependente do usuário, também permite que ele possa dar mais ou menos importância para determinada região do semivariograma, representando melhor a parte periódica do semivariograma. A interpretação visual da Figura 2 indica que, para o método de tentativa e erro, o efeito pepita é melhor delineado (Figura 2A) para valores da distância próximo a zero. Para modelos com patamar definido (Esférico) este procedimento é adequado, mas para modelos com patamar definido assintoticamente (“hole effect”, exponencial, gaussiano), o método de mínimos quadrados ponderados se ajusta melhor dentro da totalidade do intervalo de distância (Figura 2B). Segundo Vieira et al. (1991) a maioria dos dados pluviométricos do Estado de São Paulo não é periódica e é bem mais representada pelo método de mínimos quadrados ponderados (Figura 2B).

Os indicadores obtidos por “jack-knifing” (Vieira, 2000; Vieira et al., 2004) apresentados na Figura 3, mostram os resultados do método de validação para os cinco indicadores utilizados (Figura 3). Para todos os gráficos, o modelo ajustado pelo método de mínimos quadrados ponderados se aproxima mais dos valores ideais que o método de tentativa e erro. Na possibilidade de um ajustamento melhor do efeito pepita, o resultado favorável para o método de mínimos quadrados ponderados seria ainda mais evidente. O exame dos indicadores mostra que a melhor combinação está para o uso de 16 vizinhos, em que todos os indicadores se aproximam dos valores ideais, considerando-se a distância máxima de 600 m. A formação de uma vizinhança simétrica definida por grids quadrados em volta da estimativa de valores não observados, é a responsável por esta aparente coincidência (Vieira et al., 2004).

Na obtenção dos parâmetros de ajuste a partir do melhor método, o de mínimos quadrados ponderados e com uso de 16 vizinhos, os valores de precipitação podem ser interpolados sem viés e com variância mínima pela krigagem que, segundo Vieira (2000), permite a elaboração de mapas de isolinhas para a verificação e interpretação da variabilidade espacial.

CONCLUSÕES

1. O modelo matemático “hole effect” mostrou-se apropriado para o ajuste de dados de precipitação pluvial anual média.

2. Apesar de obter o maior valor do efeito pepita em relação ao método de tentativa e erro, o método de mínimos quadrados ponderados permitiu a estimação dos parâmetros do modelo com mais precisão.

3. A técnica de jack-knifing validou a escolha do método apropriado de estimação dos parâmetros do semivariograma e identificou dezesseis vizinhos como número ideal para estimativa de valores não amostrados de precipitação pluvial média, para o estado de São Paulo.

LITERATURA CITADA

- Bega, R. M.; Vieira, S. R.; Maria, I. C. de; Dechen, S. C. F.; Castro, O. M. Variabilidade espacial das precipitações pluviométricas diárias em uma estação experimental, em Pindorama, SP. *Bragantia*, v.64, n.1, p.149-156, 2005.
- Carvalho, J. R. P. de; Assad, E. D. Comparação de interpoladores espaciais univariados para precipitação pluvial anual no Estado de São Paulo. Campinas: Embrapa Informática Agropecuária. 2002, 6p. Comunicado Técnico.
- Carvalho, J. R. P. de; Assad, E. D. Análise espacial da precipitação pluviométrica no Estado de São Paulo; comparação de interpoladores. In: Congresso Brasileiro de Engenharia Agrícola, 32, 2003, Goiânia. Anais ... Goiânia: SBEA, 2003. CD-Rom.
- Carvalho, J. R. P. de; Silveira, P. M. da; Vieira, S. R. Geostatística na determinação da variabilidade espacial de características químicas do solo sob diferentes preparos. *Pesquisa Agropecuária Brasileira*, v.37, n.8, p.1151-1159, 2002.
- Cressie, N. A. C. Fitting variogram models using weighted least squares. *Journal of the International Association of Mathematical Geology*, v.17, p.563-586, 1985.
- Cressie, N. A. C. *Statistics for spatial data*. New York: John Wiley & Sons, 1991. 900p.
- DAEE - Departamento de Águas e Energia Elétrica/SP, Recursos Hídricos e Saneamento. Banco de dados hidrometeorológicos. <http://www.daee.sp.gov.br/hidrometeorologia/bancodados.htm>. 13 Mai. 2005.
- Draper, N.; Smith, H. *Applied regression analysis*, 2.ed. New York: John Wiley & Sons, 1981, 709p.
- Druck, S.; Carvalho, M. S.; Câmara, G.; Monteiro, A. M. V. Análise espacial de dados geográficos. Planaltina: Embrapa Cerrados, 2004. 208p.
- Jian, X.; Olea, R. A.; Yu, Y. S. Semivariogram modeling by weighted least squares, *Computer & Geosciences*, v.22, n.4, p.387-397, 1996.
- Journel, A. G.; Huijbregts, C. J. *Mining geostatistics*. London: Academic Press, 1978. 600p
- Mcbratney, A. B.; Webster, R. Choosing functions for semivariograms and fitting them to sampling estimated. *Journal of Soil Science*, v.37, p.617-639, 1986.
- Pyrzc, M. J.; Deutsch, C. V. The whole story on the hole effect. http://www.gaa.org.au/pdf/gaa_pyrzc_deutsch.pdf. 9 Abr. 2005.
- SAS User's guide: Statistics, version 5 ed., Cary: SAS Institute Inc., 1985, 956p.
- Statsoft, Inc. *Electronic statistics textbook*. Tulsa: StatSoft. <http://www.statsoft.com/textbook/stathome.html>. 2004. 10 Abr. 2005
- Vieira, S. R. Variabilidade espacial de argila, silte e atributos químicos em uma parcela experimental de um Latossolo Roxo de Campinas, SP. *Bragantia*, v.56, n.1, p.1-17, 1997
- Vieira, S. R. Geostatística em estudos de variabilidade espacial do solo. In: Novais, R. F. de; Alvarez, V. V. H.; Schaefer, (ed.), C. E. G. R. *Tópicos em Ciência do Solo*, Sociedade Brasileira de Ciência do Solo, v.1, p.1-54, 2000.

- Vieira, S. R.; Carvalho, J. R. P. de. Estudo da periodicidade temporal de chuvas em bacia hidrográfica dos Rios Turvo/Grande – uma proposta. Campinas: Embrapa Informática Agropecuária, 2001. 17p. Documentos, 10
- Vieira, S. R.; Carvalho, J. R. P. de; Moraes, J. F. L. de. Uso de jackknifing para validação geoestatística. In: Congresso Brasileiro de Agricultura de Precisão, 2004, Piracicaba. Anais ... Piracicaba: Sociedade Brasileira de Ciência do Solo, 2004. CD-Rom
- Vieira, S. R.; Hatfield, J. L.; Nielsen, D. R.; Biggar, J. W. Geoestatistical theory and application to variability of some agronomical properties. *Hilgardia*, v.51, n.3, p.1-75, 1983.
- Vieira, S. R.; Lombardi Neto, S.; Burrows, I. T. Mapeamento da chuva diária máxima provável para o Estado de São Paulo. *Revista Brasileira de Ciência do Solo*, v.15, n.1, p.93-98, 1991.