

Boletim de Pesquisa e Desenvolvimento

11
Julho, 2004

ISSN 1677-9266

Uso da Técnica de Mínimos Quadrados Ponderados para Ajuste de Modelos ao Semivariograma



República Federativa do Brasil

Luiz Inácio Lula da Silva
Presidente

Ministério da Agricultura, Pecuária e Abastecimento

Roberto Rodrigues
Ministro

Empresa Brasileira de Pesquisa Agropecuária - Embrapa

Conselho de Administração

José Amauri Dimárzio
Presidente

Clayton Campanhola
Vice-Presidente

Alexandre Kalil Pires
Dietrich Gerhard Quast
Sérgio Fausto
Urbano Campos Ribeiral
Membros

Diretoria Executiva da Embrapa

Clayton Campanhola
Diretor-Presidente

Gustavo Kauark Chianca
Herbert Cavalcante de Lima
Mariza Marilena T. Luz Barbosa
Diretores-Executivos

Embrapa Informática Agropecuária

José Gilberto Jardine
Chefe-Geral

Tércia Zavaglia Torres
Chefe-Adjunto de Administração

Sônia Ternes Frassetto
Chefe-Adjunto de Pesquisa e Desenvolvimento

Álvaro Seixas Neto
Supervisor da Área de Comunicação e Negócios



*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento*

*ISSN 1677-9266
Julho, 2004*

Boletim de Pesquisa e Desenvolvimento 11

Uso da Técnica de Mínimos Quadrados Ponderados para Ajuste de Modelos ao Semivariograma

José Ruy Porto de Carvalho
Sidney Rosa Vieira
Laurimar Gonçalves Vendrusculo

Campinas, SP
2004

Embrapa Informática Agropecuária
Área de Comunicação e Negócios (ACN)
Av. André Tosello, 209
Cidade Universitária "Zeferino Vaz" Barão Geraldo
Caixa Postal 6041
13083-970 Campinas, SP
Telefone (19) 3789-5743 Fax (19) 3289-9594
URL: <http://www.cnptia.embrapa.br>
e-mail: sac@cnptia.embrapa.br

Comitê de Publicações

Carla Geovana Nascimento Macário
José Ruy Porto de Carvalho
Luciana Alvim Santos Romani (Presidente)
Marcia Izabel Fugisawa Souza
Marcos Lordello Chaim
Suzilei Almeida Carneiro

Suplentes

Carlos Alberto Alves Meira
Eduardo Delgado Assad
Maria Angélica de Andrade Leite
Maria Fernanda Moura
Maria Goretti Gurgel Praxedis

Supervisor editorial: *Ivanilde Dispatto*
Normalização bibliográfica: *Marcia Izabel Fugisawa Souza*
Editoração eletrônica: *Área de Comunicação e Negócios (ACN)*

1ª. edição on-line - 2004

Todos os direitos reservados.

Carvalho, José Ruy Porto de.

Uso da técnica de mínimos quadrados ponderados para ajuste de modelos ao semivariograma / José Ruy Porto de Carvalho, Sidney Rosa Vieira, Laurimar Gonçalves Vendrusculo. — Campinas: Embrapa Informática Agropecuária, 2004.

15 p. : il. — (Boletim de Pesquisa e Desenvolvimento / Embrapa Informática Agropecuária ; 11).

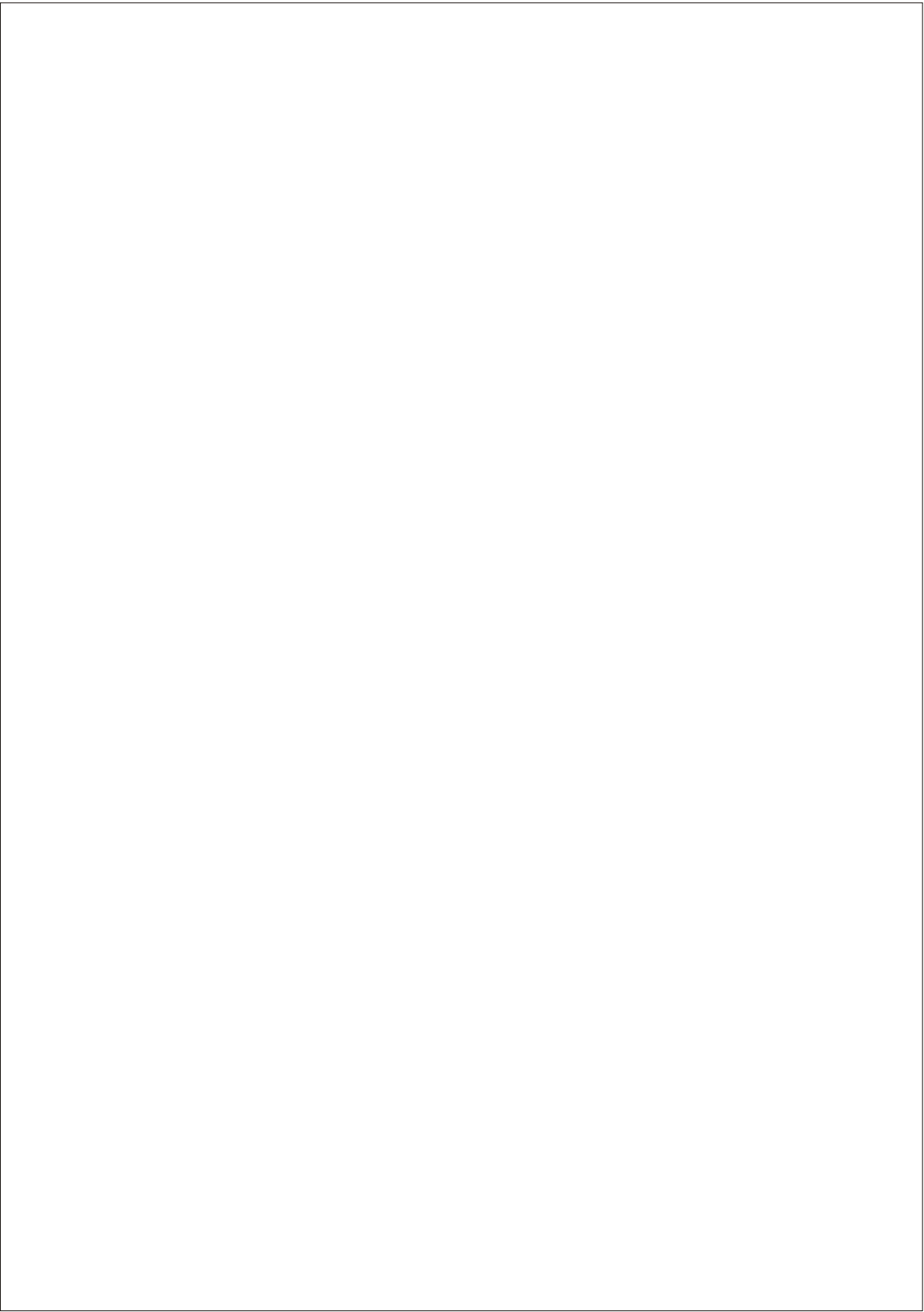
ISSN 1677-9266

1. Métodos de estimação. 2. Geoestatística. 3. Modelagem matemática. I. Vieira, Sidney Rosa. II. Vendrusculo, Laurimar Gonçalves. III. Título. IV. Série.

CDD - 519.544 (21st ed)

Sumário

Resumo.....	5
Abstract.....	6
Introdução.....	7
Material e Métodos	7
Resultados e Discussão.....	10
Conclusões	13
Referências Bibliográficas.....	14



Uso da Técnica de Mínimos Quadrados Ponderados para Ajuste de Modelos ao Semivariograma

José Ruy Porto de Carvalho¹

Sidney Rosa Vieira²

Laurimar Gonçalves Vendrusculo³

Resumo

O objetivo deste trabalho é comparar o método de estimação dos mínimos quadrados ponderados para ajuste de modelos ao semivariograma com o método de tentativa e erro, muito usado na prática, pela técnica de auto-validação "jack-knifing". Foram usadas observações de precipitação pluvial média anual de trezentas e setenta e nove estações pluviométricas abrangendo todo Estado de São Paulo, representando uma área de aproximadamente 248.808,8 km², no período de 1957 a 1997. A periodicidade exibida pelos semivariogramas foi ajustada pelo modelo "hole effect" cujos 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 ser ferramenta adequada para a definição de métodos e modelos a serem usados para semivariâncias. Este procedimento permitiu também, definir dezesseis vizinhos como o número ideal para a estimativa de valores para locais não amostrados por krigagem.

Termos para indexação: periodicidade em semivariâncias, modelo "hole effect", auto-validação, "jack-knifing".

¹ Ph.D. em Estatística, Pesquisador da Embrapa Informática Agropecuária, Caixa postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. Telefone: (19) 3789-5797, e-mail: jrui@cnptia.embrapa.br

² Ph.D. em Solos, Pesquisador do Centro de Solos e Recursos Agroambientais do Instituto Agrônomo, Caixa Postal 28 - 13001-970 - Campinas, SP. Telefone (19) 3241-5188, e-mail: sidney@iac.sp.gov.br

³ M.Sc. em Planejamento e Desenvolvimento Rural Sustentável, Pesquisadora da Embrapa Informática Agropecuária, Caixa postal 6041, Barão Geraldo - 13083-970 - Campinas, SP. Telefone: (19) 3789-5733, e-mail: laurimar@cnptia.embrapa.br

Using Weighted Least Square Technique for Semivariogram Modeling

Abstract

The objective of this work is to compare the weighted least square estimation method for semivariogram modeling with the manual fitting estimation method, very used in practice, through the jack-knifing validation technique. Observations of annual precipitation average of three hundred and seventy nine pluviometer stations were used including all State of São Paulo, representing an area of approximately 248.808,8 km², in the period from 1957 to 1997. The hole effect model whose parameters were estimated with more precision using the weighted least square estimation method when compared with the manual fitting method adjusted the periodicity exhibit by the semivariograms. The jack-knifing technique showed to be an adequate tool for the definition of methods and models to be used for semivariance. It also allowed defining sixteen neighbors as the ideal number for kriging estimations for the unsampled locations.

Index terms: periodic semivariance, hole effect model, jack-knifing, validation method.

Introdução

A experimentação agrônômica tem como uma das suas grandes preocupações testar diferenças entre tratamentos e a estatística clássica tem sido, normalmente, a ferramenta mais utilizada. Entretanto, dependendo da dimensão da variabilidade da variável estudada, o coeficiente de variação pode ser tão alto a ponto de mascarar a análise estatística do experimento. Esta variabilidade muitas vezes é devido ao solo que é a princípio considerado homogêneo. Com o estudo da distribuição espacial das amostras, a geoestatística pode ser aplicada e os parâmetros que caracterizam a estrutura da dependência espacial podem ser conhecidos. Assim como as variáveis de solo, dados de precipitação pluvial, 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, 2003).

O semivariograma é a ferramenta básica na estimação por geoestatística. Ele 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 através 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 feita com mais precisão para distâncias próximas a origem e seu ajuste é feito por tentativa e erro de um respectivo modelo selecionado pelo usuário, normalmente baseado na sua aparência visual.

O objetivo deste trabalho é apresentar 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 "jack-knifing" para dados de precipitação média anual para o Estado de São Paulo no período de 1957 a 1997.

Material e Métodos

Foram usadas trezentos e setenta e nove observações de precipitação pluvial média anual obtidas de estações pluviométricas (Fig. 1) abrangendo todo Estado de São Paulo, representando uma área de aproximadamente 248.808,8km² (2,91% do território nacional), referente ao período de 1957 a 1997.



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

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 seguinte equação:

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

onde $N(h)$ é o número de pares de valores medidos $[Z(x_i), Z(x_i + h)]$, separado pela distância h , se a variável for escalar. O gráfico de $\hat{g}(h)$ versus os valores correspondentes de h , chamado semivariograma, é uma função do vetor h , e, portanto, depende da magnitude e direção de h . Diversos tipos de modelos estão disponíveis na literatura, 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 estes modelos experimentais, os semivariogramas freqüentemente aumentam com a distância até atingir um patamar, sendo denominados monotônico crescente. Entretanto, os semivariogramas não são restritos a uma estrutura monotônica. Ele pode apresentar segmentos decrescentes ou cíclicos os quais são chamados de "hole effects". Estas estruturas não monotônicas podem ou não apresentar patamar; ter as amplitudes de ondas reduzidas; ser isotrópico (variabilidades idênticas em todas as direções) ou anisotrópico (Pyrz & Deutsch, 2004).

Os semivariogramas obtidos para dados de precipitação pluvial média anual para o Estado de São Paulo por serem periódicos, são candidatos naturais para o ajuste deste tipo de estrutura não monotônica. Ignorando estas estruturas, pode resultar

em modelos não realísticos os quais não reproduzem as variabilidades espaciais observadas e influenciam negativamente na estimativa de pontos não amostrados pelos métodos de krigagem.

O seguinte modelo matemático "hole effect" foi ajustado ao semivariograma:

$$g(h) = C_0 + C_1 \left[1 - \frac{\sin(h/a)}{h/a} \right] h^3 a \quad (2)$$

onde c_0 é o efeito pepita, $c_0 + c_1$ é o patamar, h a distância e $a > 0$ é o alcance do semivariograma.

Os métodos analíticos de ajustes de modelos aos semivariogramas estão dentro de 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 em 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 $g(h, q)$, são estimados calculando os valores numéricos para os parâmetros que minimizam a soma de quadrados dos desvios entre as respostas estimadas e observadas do modelo (Draper & Smith, 1981). Matematicamente, a função objetivo pode ser escrita como:

$$Q = \sum_h \left[\hat{g}(h) - g(h, q) \right]^2 \quad (3)$$

Onde $\hat{g}(h)$ denota o semivariograma experimental (equação 1), $g(h, q)$ o modelo de semivariograma ajustado (equação 2) e q o vetor de parâmetros. As estimativas dos parâmetros são obtidas de tal forma que a soma dos quadrados das distâncias entre o valor observado e o estimado seja a menor possível (equação 3), segundo a metodologia a seguir:

- determinar as derivadas parciais de Q em relação a c_0 , c_1 e a ;
- igualar a zero as equações de derivadas parciais;
- resolver este sistema de três equações com três incógnitas.

A pressuposição básica para se obter estimativas não-viciadas pelo método de estimação de mínimos quadrados ordinários é 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 não possuem a mesma variância por serem baseados em diferentes números de pares. Para que se maximize a estimação dos parâmetros (Statsoft Inc., 2004), o método de mínimos quadrados ponderados é normalmente usado cuja equação é:

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

Onde 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 eficientemente grandes e 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 a distância nula que é a parte mais crítica do ajuste.

Ajustar modelos matemáticos aos semivariogramas é um procedimento subjetivo. A qualidade de ajuste usando o método de mínimo quadrado ponderado ou de tentativa e erro, pode ser verificada através da técnica de "jack-knifing". Como para cada local têm-se um valor medido e pode-se estimar um outro valor através da krigagem ou co-krigagem, então pode-se calcular a regressão linear entre estes pares de dados e calcular 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 resultantes se aproximam dos seguintes valores ideais: $a = 0$, $b = 1$, $r = 1$, média do erro reduzido = 0 e variância do erro reduzido = 1. Definido o modelo matemático, os interpoladores de krigagem são usados e os pontos são 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 para verificação e interpretação da variabilidade espacial.

Resultados e Discussão

Os dados anuais de precipitação correspondem a médias para o período de 1957 a 1997. Todas as análises foram realizadas nestas médias sem a preocupação com as flutuações interanuais.

O modelo "hole effect" foi ajustado aos semivariogramas experimentais para precipitação média anual obtido pelo método de tentativa e erro e mínimos quadrados ponderados. Os resultados são apresentados nas Fig. 2a e 2b, respectivamente. Para o método de mínimos quadrados ponderados, o modelo foi estimado pelo método iterativo de Gauss-Newton (SAS Institute, 1998) para o qual são necessários a especificação dos valores iniciais dos parâmetros que se quer estimar o modelo usando uma única variável dependente e as derivadas parciais do modelo com respeito aos parâmetros.

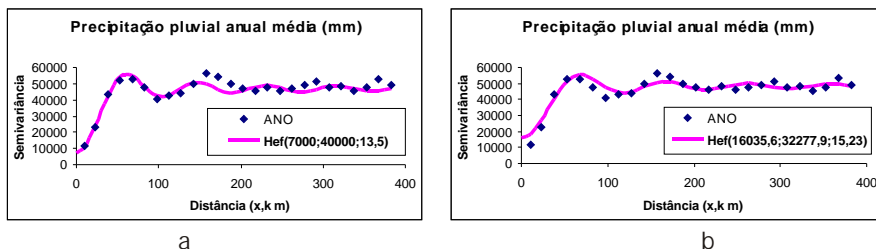


Fig. 2. Modelo "hole effect" ajustado aos semivariogramas. (a) Método de tentativa e erro; (b) Mínimos quadrados ponderados.

Os efeitos pepitas $C_0 = 7000$ e $C_0 = 16035,6$ significam que existem descontinuidade entre valores separados por distância menores do que o usado no intervalo de amostragem. A proporção destes valores para o patamar do semivariograma ($C_0 + C_1$), no caso, 14,9% e 33,19%, são um indicativo da quantidade de variação ao acaso de um ponto para outro, e quanto menor seu valor, mais parecidos são os valores vizinhos. Pode-se notar fortes flutuações até os 200 km de distância na Fig. 2, causada por variações periódicas da chuva nesta distância. Acredita-se que tal flutuação se deva aos índices pluviométricos obtidos na região litorânea, onde a precipitação pluvial apresenta uma distribuição espacial distinta das demais regiões do Estado. Ela 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 à medida que se afasta da Serra do Mar em direção ao interior.

A interpretação visual da Fig. 2, indica que o modelo é melhor ajustado pelo método de tentativa e erro (Fig. 2a) a partir de distâncias inferiores a 150 km. Ou seja, neste caso o efeito pepita é melhor delineado. Para modelos com patamar definido (Esférico) este procedimento é adequado. Entretanto para modelos com patamar definido assintoticamente ("hole effect", exponencial, gaussiano), o método de mínimos quadrados ponderados melhor se ajusta dentro da totalidade do intervalo de distância (Fig. 2b).

O método de tentativa e erro ao mesmo tempo que é mais dependente do usuário, também permite que ele possa dar mais ou menos importância para uma determinada região do semivariograma, cujo resultado pode ser importante como mostrado na Fig. 2a. Nesta Fig., o método representa muito bem a parte periódica do semivariograma. Segundo Vieira et al. (1991), a parte periódica do semivariograma é formada pelos dados de precipitação obtidos pelas estações pluviométricas localizadas na Serra do Mar as quais são a minoria. Os demais dados das estações pluviométricas que estão localizadas por todo o Estado não apresenta periodicidade, forçando o melhor ajuste determinado pelo método de mínimos quadrados ponderados (Fig. 2b).

Os indicadores obtidos por "jack-knifing" (Vieira, 2000; Vieira et al., 2004) são apresentados na Fig. 3.

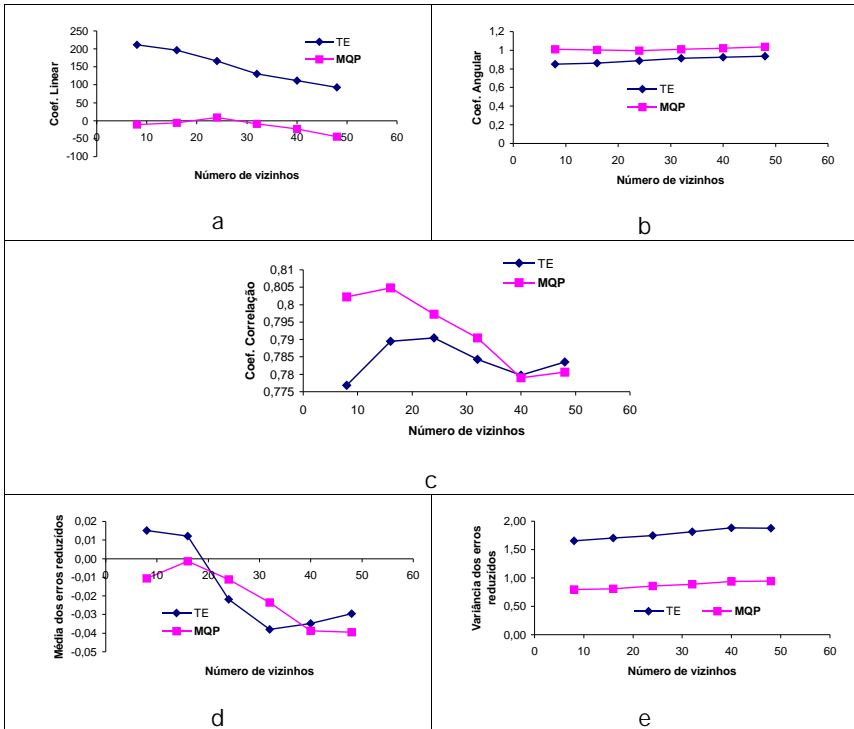


Fig. 3. Resultados de “jack-knifing” para os métodos de tentativa e erro (TE) e mínimos quadrados ponderados (MQP). Gráfico de indicadores (a - coeficiente linear; b - coeficiente angular; c - coeficiente de correlação; d - média dos erros reduzidos e e - variância dos erros reduzidos) versus número de vizinhos.

A Fig. 3 mostra os resultados do método de validação para os cinco indicadores utilizados. Para todos os gráficos o modelo ajustado pelo método de mínimos quadrados ponderados se aproxima mais dos valores ideais do que o método de tentativa e erro. O exame dos indicadores mostra que se forem usados 16 vizinhos, todos se aproximam dos valores ideais. A formação de uma vizinhança simétrica definida por malhas quadradas em volta da estimativa de valores não-observados é a responsável por esta aparente coincidência (Vieira et al., 2004).

Definidos as estimativas dos parâmetros do modelo e o número de vizinhos, os valores para locais não-amostrados podem ser obtidos pelos métodos de interpolação geostatístico de krigagem. Os valores obtidos através de krigagem são não-viciados, têm variância mínima (Vieira, 2000) e são ideais para a construção de mapas de isolinhas para verificação e interpretação da variabilidade espacial.

Conclusões

- O modelo matemático "hole effect" mostrou-se apropriado para o ajuste de dados de precipitação pluvial anual média.
- Quando os intervalos de distância se reduzem, o método de tentativa e erro apresenta melhor ajuste para o modelo.
- O método de mínimos quadrados ponderados permitiu a estimação dos parâmetros do modelo com mais precisão.
- A técnica de auto-validação "jack-knifing" ajudou significativamente na escolha de qual método de estimação dos parâmetros do semivariograma é mais apropriado e na escolha do número ideal de vizinhos.
- São necessários dezesseis vizinhos na estimativa de valores não-amostrados.

Referências Bibliográficas

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; ASSAD, E. D. Comparação de interpoladores espaciais univariados para precipitação pluvial anual no Estado de São Paulo. Campinas: Embrapa Informática Agropecuária, 2002. 6 p. (Embrapa Informática Agropecuária. Comunicado Técnico, 33). Disponível em: < <http://www.cnptia.embrapa.br/modules/tinycontent3/content/2002/comuntec33.pdf> > . Acesso em: 14 maio 2004.

CARVALHO, J. R. P. de; SILVEIRA, P. M. da; VIEIRA, S. R. Geoestatística na determinação da variabilidade espacial de características químicas do solo sob diferentes preparos. *Pesquisa Agropecuária Brasileira*, Brasília, DF, v. 37, n. 8, p. 1151-1159, ago. 2002.

CRESSIE, N. A. C. *Statistics for spatial data*. New York: John Wiley, 1991. 900 p.

DRAPER, N.; SMITH, H. *Applied regression analysis*. 2nd. ed. New York: John Wiley, 1981, 709 p.

JIAN, X.; OLEA, R. A. A.; YU, Y. S. Semivariogram modeling by weighted least squares. *Computer & Geosciences*, v. 22, n. 4, p. 387-397, 1996.

JOURNAL, A. G.; HUIJBREGTS, C. J. *Mining geostatistics*. London: Academic Press, 1978. 600 p.

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.

PYRCZ, M. J.; DEUTSCH, C. V. The whole story on the hole effect. Disponível em: < http://www.gaa.org.au/pdf/gaa_news_18.pdf > . Acesso em: 14 maio 2004.

SAS INSTITUTE. *SAS/STAT users guide: version 6. 4 .ed*. Cary, 1998. v. 1, 943 p.

STATSOFT INC. Electronic Statistics textbook. Tulsa, OK, 2004. Disponível em: < <http://www.statsoft.com/textbook/stathome.html> > . Acesso em: 14 maio 2004.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F. de; ALVAREZ V., V. H.; SCHAEFER, C. E. G. R. (Ed.). Tópicos em ciência do solo. Viçosa: Sociedade Brasileira de Ciência do Solo, 2000. v. 1, p. 1-54.

VIEIRA, S. R. Variabilidade espacial de argila, silte e atributos químicos em uma parcela experimental de um latossolo roxo de Campinas (SP). *Bragantia*, Campinas, v. 56, n. 1, p. 1-17, 1997.

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. 17 p. (Embrapa Informática Agropecuária. Documentos, 4). Disponível em: < <http://www.cnptia.embrapa.br/modules/tinycontent3/content/2001/DOCUMENTO4int.pdf> > . Acesso em: 14 maio 2004.

VIEIRA, S. R.; CARVALHO, J. R. P. de; MORAES, J. F. L. de. Uso de jack-knifing para validação geoestatística. In: CONGRESSO BRASILEIRO DE AGRICULTURA DE PRECISÃO, 2004, Piracicaba. [Anais ...]. Piracicaba, 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*, Berkeley, 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*, Campinas, v. 15, n. 1, p. 93-98, 1991.



Informática Agropecuária

Ministério da Agricultura, Pecuária e Abastecimento Governo Federal