

NOVOS ESTIMADORES DE SEMIVARIÂNCIA NO ESTUDO DA ESTRUTURA DE DEPENDÊNCIA ESPACIAL ENTRE AMOSTRAS

ARACÉLI C. MARINS¹, MIGUEL A. U. OPAZO²

¹Mestre em Engenharia Agrícola, Prof^a Assistente, Depto Matemática, Universidade Tecnológica Federal do Paraná, UTFPR, Medianeira – PR, Fone: (45)3264-2762, aracelimarins@yahoo.com.br.

²Estatístico, Prof Doutor Associado, Depto de Engenharia Agrícola, UNIOESTE, Cascavel - PR, mopazo@unioeste.br.

Escrito para apresentação no
XXXV Congresso Brasileiro de Engenharia Agrícola
31 de julho a 4 de agosto de 2006 - João Pessoa – PB

RESUMO: A geoestatística caracteriza-se por estudar a dependência espacial entre as amostras de variáveis regionalizadas. A estrutura de dependência espacial entre amostras de variáveis é verificada pela análise da função semivariância, que pode ser estimada pelos semivariogramas, dados pelo estimador Clássico de Matheron, pelo estimador de Cressie & Hawkins e pelo estimador de Semivariância Relativa Pairwise. Em 1994 LI & LAKE propuseram dois novos estimadores chamados “New1” e “New2”, que segundo os autores são mais eficientes que os demais. Neste estudo foram simulados conjuntos de dados com estrutura de dependência espacial conhecida, pela Simulação de Monte Carlo e os dados simulados foram aplicados aos cinco estimadores da função semivariância e comparados com a semivariância simulada, que é dada pelo modelo teórico. Nesta comparação, os estimadores Matheron, New1 e New2 estimaram valores mais similares aos simulados.

KEYWORDS: Geoestatística, simulação de Monte Carlo, semivariograma.

ABSTRACT: Geostatistic is characterized by the study of the spatial dependence among the samples of regionalized variables. The spatial dependence structure among the samples of variables is verified by the semivariance function analyze, witch may be estimated by the semivariograms, given by the Classical Semivariance Estimator of Matheron, by the Cressie & Hawkins Estimator and by the Pairwise Relative Semivariance Estimator. In 1994 LI & LAKE proposed two new estimators called New1 and New2, which according to the authors are more efficient than the others. In this study were simulated data sets with spatial dependence structure kneed, by the Monte Carlo Simulation and the simulated data were applied in the five estimators of the semivariance function and compared with the simulated semivariances, which is given by the theoretical model. In this comparation, the Matheron, New1 and New2 estimators estimated more similar values to the estimated.

KEYWORDS: Geostatistics, Monte Carlo simulation, semivariogram.

INTRODUÇÃO: No estudo de variáveis regionalizadas, utilizam-se dentre outros métodos, os conhecimentos de geoestatística, definida como um conjunto de procedimentos estatísticos que são aplicados a problemas onde os dados são referenciados espacialmente (RIBEIRO JUNIOR, 1995). Dentre as técnicas geoestatísticas, tem-se o Semivariograma, um gráfico da distância versus a semivariância, que é utilizado para verificar a existência de uma estrutura de dependência espacial nas amostras por meio de um estimador da função semivariância a diferentes distâncias (ISSAKS & SRIVASTAVA, 1989). A semivariância real é desconhecida e pode ser estimada pelos estimadores de semivariância. Os estimadores mais conhecidos da função semivariância são o Estimador Clássico de Matheron (MATHERON, 1962), o Estimador de Cressie & Hawkins (CRESSIE & HAWKINS, 1980) e o Estimador de Semivariância Relativa Pairwise (SRIVASTAVA & PARKER, 1989; ISSAKS & SRIVASTAVA, 1989). O estimador de Matheron é o mais estável sob condição dos dados apresentarem distribuição normal de probabilidade, porém na presença de uma distribuição de caudas pesadas o estimador de

Cressie & Hawkins é mais apropriado (SILVA, 2000). O estimador de semivariância relativa Pairwise caracteriza-se por diminuir o efeito de dados discrepantes, porém, quando a média se aproxima de zero, podem ocorrer semivariogramas com “saltos” nas semivariâncias (ISSAACS & SRIVASTAVA, 1989). LI & LAKE (1994) propuseram dois novos estimadores de semivariância: New1 e New2, a fim de tentar reduzir a imprecisão do estimador quando a distância entre as amostras aumenta. O objetivo deste trabalho é analisar e comparar por meio de simulações estocásticas as funções de semivariância estimada de Matheron, Pairwise, Cressie & Hawkins, New1 e New2, utilizando estruturas de dependência espacial conhecidas, pelo fato da maioria dos estimadores não ser robusta nem resistente.

MATERIAIS E MÉTODOS: Por meio do Processo de Simulação de Monte Carlo com base no método da Decomposição de Sholesky, geram-se dados com uma determinada estrutura de dependência espacial. A Decomposição de Sholesky, descrita por CRESSIE (1993) como segue. Seja $Z = (Z(s_1), Z(s_2), Z(s_3), \dots, Z(s_n))'$ o vetor $n \times 1$ dos dados a serem simulados, representando uma realização de um processo estocástico $\{Z(s) \in S\}$ em n diferentes locais, $s_1, s_2, s_3, \dots, s_n$. Como o processo é estocástico, o vetor das médias μ é dado por: $\mu = (E(Z(s_1)), E(Z(s_2)), E(Z(s_3)), \dots, E(Z(s_n)))'$ e a matriz de Covariância Σ dada por: $\Sigma = [C(Z(s_i), Z(s_j))]$. Nos processos estocásticos que satisfazem as hipóteses de estacionaridade intrínseca e isotropia: $E(Z(s_i)) = \mu$, para $i=1,2,3,\dots,n$, em que μ é uma constante, e $C(Z(s_i), Z(s_j)) = C(h_{ij}) = \gamma(\infty) - \gamma(h)$ onde $h_{ij} = \|s_i - s_j\|$. Assim, cada elemento do vetor μ é igual a um valor constante e cada (i,j) -ésimo elemento da matriz de covariância Σ , quadrada de ordem n , é igual a $C(h_{ij})$. Desta forma, desde que escolhidos o valor μ e a função $C(h)$, o vetor Z é dado pela relação: $Z = \mu + B\varepsilon$ em que a matriz B , triangular inferior, é tal que $BB' = \Sigma$ e $\varepsilon = (\varepsilon(s_1), \varepsilon(s_2), \varepsilon(s_3), \dots, \varepsilon(s_n))'$ é um vetor de variáveis aleatórias não correlacionadas, distribuídas normalmente, com média zero e matriz de covariância I_n (matriz identidade de ordem n). Os modelos escolhidos para a simulação foram o exponencial e o esférico, todos os dados gerados possuíam efeito pepita 0 e patamar 1, e os alcances utilizados com o modelo esférico foram de 4 e 5 unidades, já com o modelo exponencial, os alcances foram de 2, 3, 4 e 5 unidades. Optou-se por uma grade regular de 49 pontos, separados entre si pela distância mínima de 1 unidade, considerando uma matriz de distâncias quadrada. Em seguida, obteve-se a matriz A , quadrada de ordem 49, das distâncias entre todos os pontos. A partir da matriz A , obteve-se a matriz quadrada Σ , de ordem 49, de Covariância. Determinou-se, pelo Método da Decomposição de Cholesky a matriz B de ordem 49×49 . Em seguida, foram gerados 10 vetores ε de 49×1 , cujos elementos possuem distribuição normal com média 0 e variância 1. Do produto entre as matrizes B e ε_i , $1 \leq i \leq 10$, somados à matriz que represente a média do processo μ de ordem 49×1 (escolhida), resultaram 10 vetores Z de ordem 49×1 , cujos valores representam os dados simulados com estrutura de dependência espacial escolhida, na Tabela 1, para cada um dos 49 pontos. Os dados simulados foram aplicados aos cinco estimadores de semivariância dados pelos estimadores de semivariância, sendo o Estimador Clássico de Matheron da função semivariância dado por $\gamma_M(h) = [2N(h)]^{-1} \{ \Sigma^{N(h)} [Z(s) - Z(s+h)]^2 \}$ em que $N(h)$ é o número de pares de dados utilizados a uma distância h ; Z : é a variável regionalizada ou função aleatória; $s=(x_i, y_i) \in S \subset R^2$ são as posições amostrais. O Estimador proposto por CRESSIE & HAWKINS (1980), é dado por $\gamma_C(h) = [0,914N(h) + 0,988]^{-1} [\Sigma^{N(h)} |Z(s) - Z(s+h)|^{1/2}]^4$. O estimador de semivariância relativa Pairwise (SRIVASTAVA & PARKER, 1989; ISSAACS & SRIVASTAVA, 1989) é dado por $\gamma_P(h) = 2[N(h)]^{-1} \Sigma^{N(h)} \{ [Z(s) + Z(s+h)]^{-1} [Z(s) - Z(s+h)] \}^2$. Os estimadores de Semivariância New1 e New2 são dados respectivamente por: $\gamma_{N1}(h) = n^{-1} \Sigma_{i=1}^n \{ [2m]^{-1} \Sigma_{j \in D_{i,h}} [Z(s_i) - Z(s_j)]^2 \}$ e $\gamma_{N2}(h) = \gamma_{N1}(h) + d^{-1} h \gamma'_{N1}(h)$ em que n é o número de pontos amostrais na área toda; $D_{i,h}$ é um conjunto indexado dos valores da variável regionalizada Z em pontos de uma janela móvel bi-dimensional de área $\Delta_{i,h}$ com tamanho h centrada no bloco i , excluindo o ponto s_i ; m é o número de dados em $D_{i,h}$; $\gamma'_{N1}(h)$ é o valor da derivada da função γ_{N1} na distância h ; d é o momento da semivariância. e o valor da semivariância para cada uma das distâncias calculadas foi comparado com o valor da semivariância simulada, dada pelos modelos teóricos escolhidos, utilizando a soma do quadrado das diferenças entre a semivariância simulada e as estimadas e pela soma do valor absoluto destas diferenças. Foi utilizado um *cutoff* de 50% da distância máxima da área em estudo (CLARK, 1979).

RESULTADOS E DISCUSSÕES: O estimador de semivariância relativa Pairwise apresentou saltos nos valores de semivariância em todos os modelos estudados, e conseqüentemente uma grande diferença entre os demais, devido ao fato dos valores de semivariância simulada terem médias próximas de zero. Para o modelo $\text{esf}(0,1,4)$, os estimadores New2 e New1 são mais semelhantes a semivariância simulada, pela soma dos quadrados das diferenças e soma do valor absoluto das diferenças, respectivamente, por apresentarem menores diferenças a semivariância simulada. Quando os dados foram simulados de acordo com o modelo $\text{exp}(0,1,4)$, o estimador que apresentou valores de semivariância mais parecidos com a semivariância simulada foi o estimador de Matheron, por ambos os critérios. O estimador New2 foi o segundo em termos de semelhança com a semivariância simulada, segundo o critério da soma dos quadrados das diferenças e de acordo com o critério do valor absoluto da diferença, o segundo estimador a apresentar valores de semivariância mais semelhantes a semivariância simulada foi o estimador Cressie & Hawkins. Na Figura 1, são apresentados os gráficos do valor médio de dez repetições da semivariância simulada e da estimada pelos cinco estimadores.

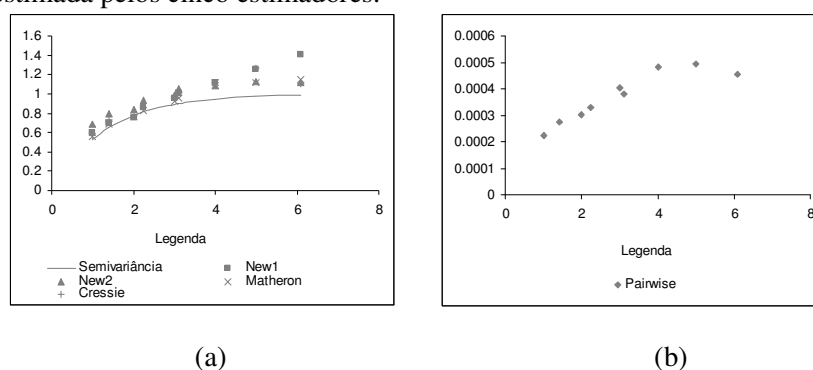


Figura 1 Gráficos da função semivariância simulada com modelo exponencial de parâmetros $C_0=0$, $C_1=1$ e $\alpha=4$ e estimadas pelos estimadores em estudo.

Segundo o modelo $\text{exp}(0,1,5)$, os estimadores que apresentaram valores mais próximos à semivariância simulada foram os estimadores Matheron e New1, nesta ordem, por ambos os critérios. Com estes mesmos parâmetros, mas utilizando o modelo esférico, os resultados mostram que o estimador Matheron apresentou melhores resultados. A Figura 2a apresenta os valores médios de dez repetições da semivariância simulada com modelo $\text{exp}(0,1,5)$ e o valor estimado pelos estimadores Matheron, Cressie & Hawkins, New1 e New2. A Figura 2b apresenta os valores da semivariância simulada pelo estimador Pairwise.

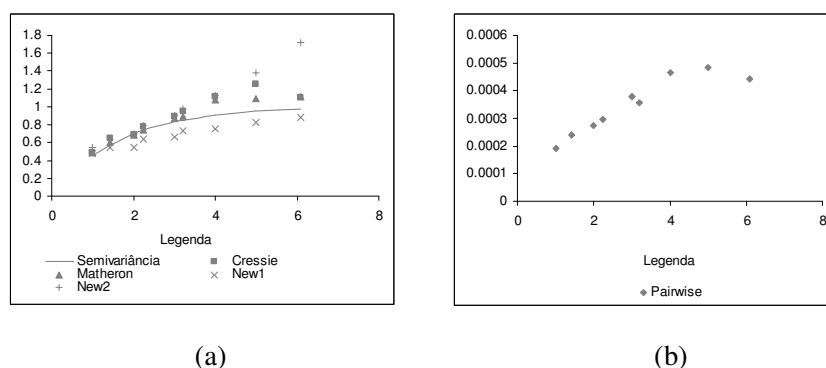


Figura 2 Gráficos da função semivariância simulada com modelo exponencial de parâmetros $C_0=0$, $C_1=1$ e $\alpha=5$ e estimadas pelos estimadores de semivariância.

A Figura 3a apresenta o valor médio de dez repetições da função semivariância simulada com modelo $\text{esf}(0,1,5)$ e seu valor estimado pelos estimadores New1, New2, Cressie & Hawkins e Matheron, e a Figura 3b representa o gráfico da semivariância simulada pelo estimador Pairwise. Observa-se na Figura 3 que os estimadores apresentam padrão de comportamento diferente da semivariância simulada, a medida

que aumenta a distância h entre os dados, e na distância de até 3 unidades os estimadores Matheron, Cressie & Hawkins e New1 apresentam comportamento semelhante.

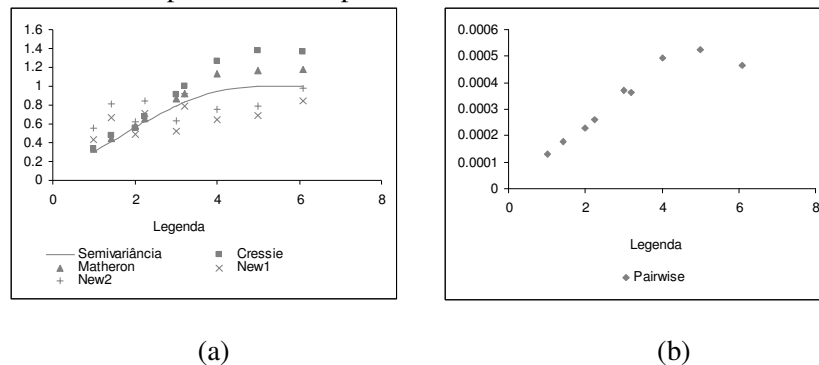


Figura 3 Gráficos da função semivariância simulada e estimada com modelo esférico de parâmetros $C_0=0$, $C_1=1$ e $\alpha=5$ e estimada pelos diferentes estimadores em estudo.

Quando os dados apresentam estrutura de dependência espacial segundo o modelo $\exp(0,1,2)$ o estimador New1 apresenta resultados mais semelhantes de semivariância estimada com a semivariância simulada segundo ambos os critérios. Quando os dados que possuam uma estrutura de dependência espacial do modelo $\exp(0,1,3)$, foram aplicados aos estimadores de semivariância, mostrou-se mais semelhante o estimador Matheron, o segundo estimador mais semelhante foi o estimador de Cressie & Hawkins, segundo ambos critérios de avaliação utilizados. Observa-se que são semelhantes a semivariância, nesta ordem, os estimadores: Matheron, New1, New2, Cressie & Hawkins. Considerando que o estimador Matheron só pode ser usado quando os dados apresentam distribuição normal de probabilidade, aconselha-se a implementação em softwares geoestatísticos dos estimadores New2 e New1, para que desta forma estes estimadores possam ser usados na análise da semivariância de conjuntos de dados georreferenciados. O estimador Pairwise, embora apresente a mesma forma gráfica da semivariância simulada, apresenta os resultados numa escala menor que a simulada nas diferentes distâncias h . Verifica-se também que com pequenos valores de distância h , os estimadores Matheron, Cressie & Hawkins, New1 e New2 têm o mesmo comportamento. LI & LAKE (1994) concluíram que o estimador New2 é o que melhor estima a semivariância mesmo a grandes distâncias, por não distorcer a representação da estrutura espacial, porém, neste estudo o estimador New1 mostrou-se mais eficiente que o New2.

CONCLUSÕES: O estimador Pairwise apresentou valores com a mesma estrutura gráfica da semivariância simulada, porém com parâmetros diferentes e numa escala menor. Segundo o estudo da simulação, os estimadores mais semelhantes a semivariância simulada foram Matheron, New1 e New2.

REFERÊNCIAS BIBLIOGRÁFICAS:

- CRESSIE, N.; HAWKINS, D. M. **Robust Estimation of the variogram**, J. Int. Assoc. Math. Geol., v. 12, p. 115-125, 1980.
- ISAAKS, E.H.; SRIVASTAVA, R.M. **An introduction to applied geoestatísticas**. New York, Oxford University Press, 1989. 560p.
- LI, D.; LAKE, L.W. A moving window semivariance estimator. **Water Resources Research**, vol. 30, n. 5, p. 1479-1489, 1994.
- MATHERON, G. **Traite de geoestatistique appliquée**, vol. I, Mem. BRGM, v. 14, 1962.
- RIBEIRO JUNIOR, P.J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. Piracicaba, SP, 1995, 99p. (dissertação de Mestrado, ESALQ/USP).
- SRIVASTAVA, R.M.; PARKER, H.M. **Robust measures of spatial continuity**, in Geoestatistics, vol. 1, edited by M. Armstrong, p. 295-308, Kluwer Academic, Hingham, Mass., 1989.