SIMULAÇÃO GEOESTATÍSTICA COM DIFERENTES GRAUS DE DEPENDÊNCIA ESPACIAL EM UM DELINEAMENTO SISTEMÁTICO

Melissa Lombardi Oda, Depto de Ciências Exatas, ESALQ,Caixa Postal:9, Piracicaba/SP, CEP:13418-900, e-mail: melissa.oda@gmail.com

Paulo Justiniano Ribeiro Jr., Prof. Depto de Estatística, UFPR, e-mail: paulojus@est.ufpr.br

José Luiz Stape, Prof. Depto de Ciências Florestais, ESALQ/USP, e-mail: stape@esalq.usp.br

Décio Barbin, Prof. Depto de Ciências Exatas, ESALQ/USP, e-mail: debarbin@esalq.usp.br

Ana Alice Pilon, Prof. Depto de Estatística, UFPR, e-mail: aapilon@est.ufpr.br

RESUMO: O objetivo da simulação geoestatística é reproduzir a função aleatória, cujas características foram identificadas na análise estrutural. As simulações produzem portanto, a mesma variabilidade dos dados sobre a área de estudo. O objetivo desse trabalho é o de comparar modelos simulados e métodos de estimação de parâmetros que consideram a presença e importância da informação espacial na análise de um delineamento sistemático tipo "leque". As estimativas dos efeitos dos tratamentos serão feitas através dos métodos de estimação de máxima verossimilhança (MV) e do variograma seguido de mínimos quadrados generalizados (VAR-MQG), que serão comparados à ANOVA. A fim de resumir os resultados alcançados nas simulações, a cada conjunto de dados simulados serão construídos os intervalos de cobertura. Podemos concluir que o método MV apresentou o melhor intervalo de cobertura e os resultados obtidos para os dois diferentes níveis da razão entre as variâncias dos erros independentes e dependentes, τ^2/σ^2, foram semelhantes para todos os graus de dependência espacial estabelecidos nesse estudo.

Palavras-chave: simulação, dependência espacial, delineamento sistemático, ANOVA, mínimos quadrados generalizados, máxima verossimilhança.

1. INTRODUÇÃO

Compreender a distribuição espacial de dados oriundos de fenômenos ocorridos no espaço constitui hoje um grande desafio para a elucidação de questões centrais em diversas áreas do conhecimento, seja em saúde, em ambiente, em geologia, em agronomia, entre tantas outras. A ênfase da análise espacial é estudar propriedades e relações, levando em conta a localização espacial do fenômeno em observação, ou seja, a idéia central é incorporar o espaço à análise que se deseja fazer (Câmara et al., 2001).

Assim, na estatística espacial ou geoestatística cada observação é descrita não apenas pelo seu valor, mas também por informação de sua posição, expressa por um sistema de coordenadas (Ribeiro Júnior, 1995). Considerando que observações mais próximas geograficamente tendem a ter valores mais similares e que tal fato pode ser avaliado por medidas de associação, a análise geoestatística determina o grau de associação (correlação) entre amostras, com base na direção e na distância entre elas. A dependência espacial (autocorrelação espacial) faz com que os pontos de amostragem mais próximos entre si sejam mais semelhantes do que aqueles mais distantes (Montagna, 2001).

A utilização da simulação como ferramenta de análise permite realizar diversos cenários num mesmo ponto ou em toda área (Braga, 1990). O objetivo da simulação geoestatística é reproduzir a função aleatória, cujas características foram identificadas na análise estrutural. As simulações produzem portanto, a mesma variabilidade dos dados sobre a área de estudo (Mello, 2004).

O objetivo desse trabalho é comparar os modelos simulados e métodos de estimação de parâmetros que consideram a presença e importância da informação espacial na análise de um delineamento sistemático tipo "leque".

2. METODOLOGIA

Os dados analisados foram simulados a partir de um modelo geoestatístico dado por:

$$Y_i = \mu + a_i + S(x_i) + Z_i, \qquad i = 1, ..., n, \tag{1}$$

em que,

i) Y_i é o valor observado na posição x_i;

ii) μ é média geral, cujo o valor será fixado de acordo com os dados reais;

iii) a_i é o efeito do i-ésimo tratamento, que os valores também serão fixados;

iv) x_i é a posição espacial da i-ésima parcela;

v) $S(x_i)$ o componente espacial, que será gerado aleatoriamente a partir da função grf, do pacote geoR (Ribeiro Jr. & Diggle (2004)), que gera efeitos aleatórios espaciais dados os parâmetros de covariância σ^2 e ϕ, ou seja, $S \sim NMV(0, \sigma^2 R(\phi))$, em que $R(\phi)$ é a matriz de covariância $n \times n$ com (i, j)-ésimo elemento, $\rho(u_{ij})$, sendo $u_{ij} = \|x_i - x_j\|$ a distância euclidiana entre x_i e x_j;

vi) Z_i são variáveis aleatórias, independentes e normalmente distribuídas com média zero e variância τ^2.

A simulação geoestatística, no presente estudo, gerou 1000 amostras a partir do modelo dado pela equação (1) com três graus de dependência espacial (alto, médio e baixo com 5, 3 e 1 respectivamente) entre as parcelas. Com diferentes níveis da razão entre as variâncias dos erros independentes e dependentes, τ^2/σ^2 (0,05 e 0,3) considerando-se a estrutura de covariância família Matérn com $\kappa = 1,5$. Assim totalizando seis simulações de 1000 amostras cada. Os efeitos de tratamentos e média foram fixados de acordo com os dados reais.

Existem vários métodos para a estimação dos parâmetros do modelo geoestatístico, porém no presente trabalho, a estimativa dos efeitos dos tratamentos serão feitas através dos métodos de estimação de Máxima Verossimilhança e do método do variograma seguido de mínimos quadrados generalizados. Esses métodos também serão comparados À uma análise clássica ANOVA.

O método do variograma seguido de mínimos quadrados generalizados consiste em estimar e modelar a correlação a partir dos resíduos de uma análise clássica através de semi-variogramas e utilizar a teoria dos mínimos quadrados generalizados para estimar os efeitos dos tratamentos. A partir de uma perspectiva da estatística clássica, a função verossimilhança é a fundamentação na qual os métodos de inferências são construídos. Estimações de modelos paramétricos pela maximização da função de verossimilhança sob um modelo assumido, no caso desse trabalho o modelo geoestatístico, fornecem estimadores imparciais e eficientes em amostras de grande tamanho (Pilon, 2004).

A fim de resumir os resultados alcançados nas simulações, a cada conjunto de dados simulados serão construídos os intervalos de cobertura (ICB), que consiste na porcentagem dos intervalos de confiança gerados, conterem o verdadeiro valor dos efeitos de tratamentos.

3. RESULTADOS E DISCUSSÃO

Nas tabelas de resultados apresentadas adota-se a seguinte notação: ANOVA - Análise de variância, VAR-MQG - Variograma seguido de mínimos quadrados generalizados no modelo geoestatístico e MV - Estimação por máxima verossimilhança no modelo geoestatístico.

Nas tabelas 1, 2 e 3 observa-se que a medida que aumenta o grau de dependência espacial o intervalo de cobertura para a ANOVA diminue, comportamento também verificado com o VAR-MQG e MV, porém a diferença foi mais amena, ou seja, os resultados foram relativamente próximos. Comparando-se as tabelas independentemente do nível de dependência espacial, o método de máxima verossimilhança apresentou o melhor intervalo de cobertura. Os resultados acerca dos níveis de τ^2/σ^2 (0,05 e 0,3) foram semelhantes para todos os graus de dependência espacial.

Tabela 1: Intervalos de cobertura de cada efeito de tratamento, resultantes da simulação com baixa dependência espacial com dois níveis de τ^2/σ^2.

$\tau^2/\sigma^2 = 0,05$	$ICB_{(1)}$	$ICB_{(2)}$	$ICB_{(3)}$	$ICB_{(4)}$	$ICB_{(5)}$	$ICB_{(6)}$	$ICB_{(7)}$	$ICB_{(8)}$	$ICB_{(9)}$	$ICB_{(10)}$
ANOVA	0,6870	0,7270	0,7730	0,8310	0,8490	1,0000	0,8960	0,9330	0,9370	0,9420
VAR-MQG	0,8420	0,8610	0,8770	0,9090	0,9180	0,9260	0,9360	0,9640	0,9610	0,9600
MV	0,9440	0,9380	0,9240	0,9420	0,9290	0,9350	0,9210	0,9480	0,9440	0,9420
$\tau^2/\sigma^2 = 0,3$	$ICB_{(1)}$	$ICB_{(2)}$	$ICB_{(3)}$	$ICB_{(4)}$	$ICB_{(5)}$	$ICB_{(6)}$	$ICB_{(7)}$	$ICB_{(8)}$	$ICB_{(9)}$	$ICB_{(10)}$
ANOVA	0,6850	0,7290	0,7730	0,8310	0,8500	1,0000	0,8930	0,9350	0,9370	0,9420
VAR-MQG	0,8340	0,8560	0,8740	0,9020	0,9160	0,9210	0,9320	0,9640	0,9590	0,9590
MV	0,9440	0,9370	0,9260	0,9420	0,9320	0,9350	0,9210	0,9490	0,9420	0,9420

Tabela 2: Intervalos de cobertura de cada efeito de tratamento, resultantes da simulação com média dependência espacial com dois níveis de τ^2/σ^2.

$\tau^2/\sigma^2 = 0,05$	$ICB_{(1)}$	$ICB_{(2)}$	$ICB_{(3)}$	$ICB_{(4)}$	$ICB_{(5)}$	$ICB_{(6)}$	$ICB_{(7)}$	$ICB_{(8)}$	$ICB_{(9)}$	$ICB_{(10)}$
ANOVA	0,3970	0,4400	0,4990	0,5540	0,5840	1,0000	0,6530	0,7180	0,7510	0,7710
VAR-MQG	0,8080	0,8120	0,8130	0,8330	0,8410	0,8360	0,8410	0,8570	0,8670	0,8630
MV	0,9370	0,9290	0,9240	0,9300	0,9360	0,9300	0,9120	0,9170	0,9290	0,9150
$\tau^2/\sigma^2 = 0,3$	$ICB_{(1)}$	$ICB_{(2)}$	$ICB_{(3)}$	$ICB_{(4)}$	$ICB_{(5)}$	$ICB_{(6)}$	$ICB_{(7)}$	$ICB_{(8)}$	$ICB_{(9)}$	$ICB_{(10)}$
ANOVA	0,3990	0,4390	0,5000	0,5560	0,5840	1,0000	0,6560	0,7170	0,7530	0,7680
VAR-MQG	0,8060	0,8130	0,8140	0,8360	0,8400	0,8360	0,8410	0,8590	0,8750	0,8630
MV	0,9320	0,9280	0,9240	0,9320	0,9360	0,9290	0,9100	0,9180	0,9290	0,9150

Tabela 3: Intervalos de cobertura de cada efeito de tratamento, resultantes da simulação com alta dependência espacial com dois níveis de τ^2/σ^2.

$\tau^2/\sigma^2 = 0,05$	$ICB_{(1)}$	$ICB_{(2)}$	$ICB_{(3)}$	$ICB_{(4)}$	$ICB_{(5)}$	$ICB_{(6)}$	$ICB_{(7)}$	$ICB_{(8)}$	$ICB_{(9)}$	$ICB_{(10)}$
ANOVA	0,2940	0,3160	0,3470	0,3760	0,4250	1,0000	0,5080	0,5470	0,5830	0,6080
VAR-MQG	0,8240	0,8120	0,8150	0,8040	0,8240	0,8210	0,8170	0,8210	0,8190	0,8110
MV	0,9220	0,9180	0,9220	0,9250	0,9230	0,9200	0,9070	0,8990	0,9030	0,8870
$\tau^2/\sigma^2 = 0,3$	$ICB_{(1)}$	$ICB_{(2)}$	$ICB_{(3)}$	$ICB_{(4)}$	$ICB_{(5)}$	$ICB_{(6)}$	$ICB_{(7)}$	$ICB_{(8)}$	$ICB_{(9)}$	$ICB_{(10)}$
ANOVA	0,2960	0,3160	0,3470	0,3780	0,4250	1,0000	0,5110	0,5460	0,5820	0,6080
VAR-MQG	0,8240	0,8080	0,8140	0,8040	0,8250	0,8190	0,8150	0,8190	0,8200	0,8110
MV	0,9160	0,9150	0,9190	0,9200	0,9190	0,9170	0,9080	0,9000	0,8920	0,8810

4. CONCLUSÕES

Face as metodologias empregadas nesse trabalho e os resultados obtidos, pode-se concluir que: o método de máxima verossimilhança apresentou o melhor intervalo de cobertura e os resultados acerca dos diferentes níveis estabelcidos para τ^2/σ^2 (0,05 e 0,3) foram semelhantes para todos os graus de dependência espacial.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Braga, L.P.V.(1990). *Geoestatística e aplicações.* In: Simpósio Brasileiro de Probabilidade e Estatística, 9., São Paulo, 1990. Resumos, São Paulo: IME-USP, 1990, p.36.

Câmara, G.; Monteiro, A.M.; Fuks, S.D.; Felgueiras, C.(2001). *Análise espacial de dados geográficos.* 2.Ed. São José dos Campos: INPE, 2001.

Mello, J.M. (2004). *Geoestatística aplicada ao inventário florestal.* Piracicaba, 2004. 111p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

Montagna, M.A.(2001). *Distribuição espacial e amostragem seqüencial da mosca branca Bemisia tabaci Raça B (HOMOPTERA: ALEYRODIDAE) no agroecossistema do melão.* Ribeirão Preto, 83p. Dissertação (Mestrado) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo.

Pilon, A. A.(2004). *Métodos para incorporação da dependência espacial na análise de dados experimentais.* Piracicaba, 137p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

Ribeiro Júnior, P.J. (1995). *Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo.* Piracicaba, 1995, 99p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

Ribeiro Júnior, P.J.; Diggle, P.J. (2004). *The geoR package functions for geostatistical data analysis:* , 119p.