

Um estimador robusto e o semivariograma cruzado na análise de variabilidade espacial de atributos de solo e planta

Edson Antonio Alves da Silva^{1*}, Miguel Angel Uribe-Opazo¹, Eduardo Godoy Souza¹ e Jansle Vieira Rocha²

¹Centro de Ciências Exatas e Tecnológicas, Unioeste, Rua Maringá, 1719, 85816-280, Cascavel, Paraná, Brasil.

²Departamento de Engenharia Agrícola, Unicamp, Campinas, São Paulo, Brasil. *Autor para correspondência: e-mail: edsonsilva@unioeste.br

RESUMO. O conhecimento da variabilidade espacial dos atributos do solo que influenciam na produtividade agrícola é fundamental para a implantação da agricultura de precisão. Se o comportamento desses atributos apresentar dependência espacial, recomenda-se o emprego da geoestatística na descrição de sua estrutura espacial pela obtenção de estimativas em locais não-amostrados. Nessas estimativas são necessários estimadores não-tendenciosos, precisos e robustos. Muitos trabalhos utilizam o estimador clássico de Matheron; entretanto, para dados com distribuições de caudas longas ou distribuições assimétricas, recomenda-se o emprego de estimador robusto. Os objetivos deste trabalho foram empregar o estimador robusto de Cressie e Hawkins na produção do semivariograma para o fósforo, o potássio e a produtividade de soja em área agrícola experimental; produzir os mapas de contorno correspondentes e correlacioná-los por meio do semivariograma cruzado. O experimento resultou em mapas que ilustram a variabilidade espacial das variáveis e suas correlações, servindo de ferramenta gerencial ou de documentação na prática agrícola moderna.

Palavras-chave: agricultura de precisão, semivariograma, estimador robusto.

ABSTRACT. A robust estimator and the cross semi-variogram in spatial variability analysis of soil and plants attributes. The knowledge of the soil attributes spatial variability that influence crop yield is essential for precision agriculture. If the behavior of these attributes presents spatial dependence, then it is recommended to use geostatistics to describe this spatial structure and to estimate values in non sampled places. In order to have good estimates, unbiased, precise and robust estimators are necessary. Many studies use the classic estimator of Matheron, however, the use of a robust estimator is recommended to data with long tail or asymmetrical distribution. The objectives of this study were to use Cressie and Hawkins' estimator to generate experimental semi-variograms of Phosphorus, Potassium and soybean yield in an experimental agricultural area, to produce their contour maps and to correlate them through crossed semi-variogram. This experiment resulted in maps that illustrate the spatial variability of the variables and their correlations, used as a management tool or documentation in a modern agricultural practice.

Key words: precision agriculture, semi-variogram, robust estimator.

Introdução

Para Molin (1997), a agricultura de precisão é tecnologia de ponta tanto para a otimização de produção quanto para a tomada de decisões no manejo agrícola. A falta de uniformidade espacial nessas áreas de manejo sugere tratamento localizado, visando à economia pela minimização de insumos e/ou aumento de níveis de produtividade. O interesse é obter uma prática agrícola mais eficiente

com respeito à otimização de seus recursos e que cause menor dano ao meio ambiente.

A agricultura de precisão pressupõe a elaboração e a utilização de mapas que descrevam as características e as propriedades de uma dada área. Esses mapas representam espacialmente a variabilidade de propriedades do solo, produtividade e características observadas em campo. Com base nas informações contidas nesses mapas, são tomadas as decisões que envolvem o manejo localizado dessas

áreas. Dessa forma, torna-se essencial que a escolha dos métodos para geração desses mapas seja feita com base em critérios bem estabelecidos. Os mapas de produtividade em agricultura de precisão são gerados a partir de sensores eletrônicos e de sistemas de posicionamento global (GPS) e normalmente são resultantes de denso esquema de amostragem em toda a superfície. No entanto, os mapas que expressam a variabilidade dos atributos do solo e das observações de campo são, em sua maioria, derivados de medições ou de observações realizadas em locais predeterminados, baseados em amostragem.

A partir dos dados coletados nos locais de amostra e por meio de interpolação digital dos dados, são gerados os mapas de superfície contínua de variabilidade. Para a produção de mapas de superfície (tridimensionais) ou isolinhas (contornos bidimensionais), têm-se utilizado modelos estatísticos de interpolação matemática e geoestatística na obtenção de estimativas de valores para locais não-amostrados. Os métodos matemáticos e de estatística clássica utilizam amostras independentes e identicamente distribuídas, enquanto os métodos geoestatísticos utilizam amostras espacialmente distribuídas segundo uma grade referenciada geograficamente por um sistema de coordenadas, que possam apresentar estrutura de dependência.

A geoestatística não é um tipo diferente ou especial de estatística, mas uma estatística que leva em conta não só o valor obtido para uma determinada variável resposta Z , mas também a sua posição, expressa por um sistema de coordenadas.

No estudo geoestatístico a interpolação por krigagem apresenta-se como método de predição que permite prever, pontualmente ou por intervalo, valores de um processo estocástico em localizações ou regiões não amostradas previamente. Esse método usa justamente a dependência espacial entre amostras vizinhas, expressa por uma função da variabilidade em termos da distância entre elas, chamada de semivariância. Uma interpolação por esse método permite a construção de mapas de isolinhas tendo estimativas não viciadas de variância mínima, pois a densidade espacial dos dados será maior que outros métodos. Para Vieira (1996), nos problemas em que a estatística clássica tem limitações é que a geoestatística tem aplicações. Assim, esta metodologia vem atender o propósito de lidar com experimentos que apresentem valores amostrais com dependência espacial.

Para Prevedello (1987), devido ao fato de uma variável regionalizada normalmente oferecer

resultados únicos de uma função aleatória, tomados em cada ponto, torna-se impossível o uso da estatística clássica em tais casos. Mesmo a geoestatística precisa atender às hipóteses de estacionariedade de ordem 2, intrínseca e de isotropia, hipóteses essas que devem necessariamente ser verificadas direta ou indiretamente. Tais requisitos são amplamente discutidos por Vieira (1996) e dão foco em sua importância para inferências em geoestatística.

Segundo Isaaks e Srivastava (1989), o semivariograma é definido como o gráfico da função semivariância dada por $\gamma(h) = \frac{1}{2} E[Z(s) - Z(s+h)]^2$ versus a distância h em que $Z(s)$ é uma variável regionalizada para cada posição e s representa o valor obtido da variável $Z(\cdot)$ na coordenada (x_i, y_i) de um espaço euclidiano S . Assim, o semivariograma é ferramenta adequada para medir a dependência entre pontos amostrais separados pela distância h , distribuídos em um sistema espacial de referência. O semivariograma fornece, também, os parâmetros alcance (a), patamar ($C_0 + C_1$) e efeito pepita (C_0), necessários no método de interpolação de valores (krigagem), para a construção de mapas de isolinhas e superfícies.

Estimar a função de semivariância significa determinar convenientemente os parâmetros de um modelo teórico. O estimador mais utilizado para a semivariância $\gamma(h)$ é o estimador de Matheron (estimador dos momentos). Para Braga (1990) se $Z(\cdot)$ é função aleatória estacionária, então esse estimador, sob a hipótese intrínseca é não tendencioso. Ribeiro Júnior (1995) acrescenta, ainda, que esse estimador de Matheron é não viesado para $\gamma(h)$ quando $Z(\cdot)$ representa um processo estocástico intrinsecamente estacionário, mas é pouco resistente e muito afetado por observações atípicas (*outliers*), sobretudo aquelas que produzem distribuições de caudas longas. Nesses casos, surge a necessidade de estimador robusto, que não seja afetado por valores discrepantes que não possam ser eliminados ou substituídos.

Cressie e Hawkins (1980) apresentam o estimador robusto de semivariância dado por:

$$\hat{\gamma}(h) = \frac{1}{2} \frac{\left[N(h)^{-1} \sum_{N(h)} \sqrt{|(Z(s+h) - Z(s))|} \right]^4}{0.457 + \frac{0.494}{N(h)}}$$

Nesta função, $N(h)$ representa o número de pares separados por uma distância h .

Este estimador, segundo Ribeiro Junior (1995), é adequado naqueles casos nos quais são simuladas distribuições de cauda longas (assimétricas ou

leptocúrticas), pois fornece uma estimativa robusta e estável do semivariograma.

Para Viana Da Mata *et al.* (1997), com o semivariograma cruzado é possível verificar a relação entre duas variáveis espacialmente medidas, mostrando se a variabilidade de uma série é acompanhada pela variabilidade da outra variável. Se duas variáveis regionalizadas se apresentam correlacionadas espacialmente, então o semivariograma cruzado - que representa essa correlação e estabelece o vínculo entre elas - será dado por:

$$\hat{\gamma}(h)_{1,2} = \hat{\gamma}(h)_{2,1} = \frac{1}{2 \cdot N(h)} \left\{ \sum_{i=1}^{N(h)} [z_1(s_i) - z_1(s_i + h)][z_2(s_i) - z_2(s_i + h)] \right\}$$

Nesta função $z_1(\cdot)$ representa o valor de uma variável numa dada localização e $z_2(\cdot)$ o valor da outra variável na mesma posição.

A forma obtida para os gráficos do semivariograma cruzado experimental tem significado diferente do obtido pelo semivariograma simples. Nele, o alcance representa a distância máxima de dependência espacial entre as duas variáveis em estudo; o patamar, se existir, deve aproximar-se do valor da covariância entre as duas variáveis. Quando as duas variáveis apresentarem correlação inversa, a covariância será negativa e o semivariograma cruzado também produzirá resultados negativos (Vieira, 1996). Entretanto, uma padronização da correlação obtida no ajuste de um modelo ao semivariograma cruzado permitirá uma mesma escala de comparação entre os casos avaliados. Uma forma de padronizar tal correlação será ponderar o resultado pelo produto da raiz quadrada dos respectivos patamares das variáveis envolvidas, ou seja:

$$CGP = \frac{(C_0 + C_1)}{\sqrt{C_{Z_1}} \sqrt{C_{Z_2}}}$$

Nesta expressão, CGP é o coeficiente geoestatístico padronizado para a correlação espacial, $(C_0 + C_1)$ é o patamar de ajuste do semivariograma cruzado, C_{Z_1} e C_{Z_2} os patamares das variáveis Z_1 e Z_2 .

Souza *et al.* (1999) analisaram a variabilidade espacial dos atributos fósforo e potássio, matéria orgânica e potencial de hidrogênio (pH) com objetivo de transformar uma área experimental de Latossolo Vermelho Distrófico em Cascavel-PR, em área piloto de agricultura de precisão. Utilizaram o estimador clássico de Matheron para a produção do semivariograma experimental e com os resultados dos parâmetros obtidos com o

ajuste de modelos de variabilidade, interpolaram valores com a técnica de krigagem ordinária para produzir mapas de contorno. Todas as variáveis estudadas apresentaram dependência espacial.

O objetivo deste trabalho foi aplicar o estimador robusto de Cressie e Hawkins para produzir o semivariograma experimental dos atributos fósforo, potássio e produtividade de soja e para descrever suas variabilidades espaciais, por meio de mapas que representem a distribuição da variável na área, utilizando a krigagem ordinária como método de interpolação. Objetivou também determinar a correlação espacial da produtividade observada na área com as variáveis fósforo e potássio por meio do semivariograma cruzado.

Material e métodos

O conjunto de dados pesquisados refere-se a um experimento conduzido pelo Núcleo de Inovações Tecnológicas - NIT da Unioeste, Estado do Paraná, em uma área de 1,33 ha de Latossolo Vermelho distrófico, no Centro de Pesquisa Eloy Gomes da Cooperativa Central Agropecuária de Desenvolvimento Tecnológico em Cascavel-PR. A amostragem foi conduzida após a colheita de soja da safra de 1998, cultivada por meio da técnica de semeadura direta. O delineamento constou de 256 parcelas de 7,20 x 7,20m, dispostas em uma grade regular. As amostras para determinação dos atributos de solo foram obtidas com trado de 7cm de diâmetro a 15cm de profundidade.

Neste estudo, foram analisados o nutriente fósforo (ppm), por ser importante macronutriente, absorvido em grande quantidade pelas plantas e essencial para o desenvolvimento radicular e para a maturação mais uniforme e o nutriente potássio (cmolc/dm³), também essencial ao correto desenvolvimento das plantas e a variável produtividade (t/ha). O esquema de amostragem utilizado foi o definido por Souza *et al.* (1999). A análise estatística inicialmente avaliou descritivamente a distribuição de frequências, a normalidade, a presença de pontos discrepantes e a estrutura de tendências na área de estudo (gráfico post-plot). A seguir, para uma análise geoestatística, foram produzidos semivariogramas experimentais utilizando-se o estimador robusto de Cressie e Hawkins. A isotropia foi identificada construindo-se semivariogramas nas diferentes direções. Para a obtenção dos parâmetros alcance (a), patamar ($C_0 + C_1$) e efeito pepita (C_0), no ajuste de um modelo teórico matemático aos pontos de cada semivariograma experimental foi utilizado um corte de 50% da distância máxima, desprezando os pares

de longo alcance (Clark, 1979). Empregou-se, também, espaçamento de 8,4m em um total de 10 *lags* de espaçamento máximo. Devido à existência de isotropia dos dados, adotou-se tolerância angular de 90° com direção inicial de 0°.

Foi verificada a adequação do modelo pela validação cruzada. Ela consiste em eliminar observações da amostra e usar os dados restantes para estimá-las e, então, por krigagem ordinária, produziu-se o mapa de contorno para representar a variabilidade espacial de cada atributo na área estudada. Finalmente, foram correlacionados espacialmente esses atributos do solo com a produtividade, empregando o semivariograma cruzado e utilizando o coeficiente de correlação espacial (CCE) como parâmetro de comparação dos casos estudados.

Resultados e discussão

A variável fósforo apresentou média de 4,03mg/dm³, com coeficiente de variação de 31,8% e uma distribuição assimétrica positiva, significando que a variável é razoavelmente homogênea mas tem tendência a assumir valores baixos. A variável potássio apresentou uma média de 0,33 cmolc/dm³, com coeficiente de variação de 26,2% e, também, uma distribuição assimétrica positiva. Com ambas as variáveis rejeitou-se, em nível 5% de probabilidade, a hipótese de normalidade dos dados segundo o teste de Shapiro e Wilk (1965). A variável produtividade apresentou média de 2,75 t/ha, com coeficiente de variação de 17,9% e distribuição simétrica mesocúrtica. Nesse caso, não se rejeitou, em nível de 5% de probabilidade, a hipótese de existência de normalidade dos dados (Tabela 1).

Tabela 1. Estatísticas descritivas das variáveis de estudo

Variável	Unid.	Média	Mediana	D.P.	Mínimo	Máximo	Q1	Q3	CV (%)
Produtividade	ton/ha	2,75	2,73	0,49	1,19	4,14	2,40	3,05	17,9
Fósforo	mg/dm ³	4,03	3,80	1,28	2,00	13,10	3,10	4,50	31,8
Potássio	cmolc/dm ³	0,33	0,32	0,09	0,17	0,74	0,26	0,38	26,2

Nota: D.P. - Desvio Padrão; Mínimo - Valor Mínimo; Máximo - Valor Máximo; Q1 - Primeiro Quartil; Q3 - Terceiro Quartil; CV - Coeficiente de Variação

Os gráficos de média *versus* variância segundo as direções horizontal e vertical, para cada uma das variáveis analisadas, não indicaram estrutura de tendência, fazendo supor satisfeita a hipótese de estacionaridade dos dados.

Os gráficos Box-plot (Figura 1) confirmam, para as variáveis fósforo e potássio, vários pontos discrepantes acima do terceiro quartil, sugerindo a não-normalidade dos dados. A presença desses valores faz que a distribuição das variáveis seja assimétrica à direita. Esse fato justifica o emprego de

um estimador robusto devido à distribuição em cauda longa, enquanto para a variável produtividade também se justifica o emprego de um estimador robusto por apresentar valores discrepantes que, entretanto, não comprometem a simetria dos dados, indicando normalidade ou ausência de tendência.

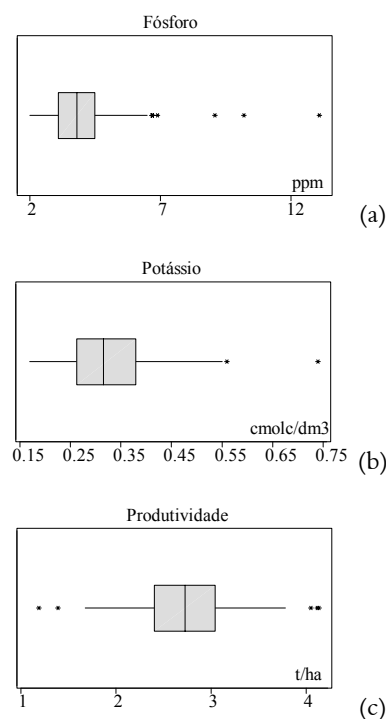


Figura 1. Box-plot das variáveis fósforo (a), potássio (b) e produtividade (c)

Os diagramas de dispersão das três variáveis (fósforo, potássio e produtividade) não indicaram nenhuma tendência para os dados em relação às direções *Norte-Sul*, *Leste-Oeste*, *Noroeste-Sudeste* e *Sudoeste-Nordeste*. No caso do fósforo, a presença de pontos discrepantes (*outliers*) causa uma certa influência no conjunto de dados, mas não pode ser considerada relevante. Nas demais direções não se observou nenhuma correlação.

A Figura 2 mostra o *post-plot* das variáveis estudadas. Esse gráfico é construído pela classificação de cada parcela segundo os quartis. Dados que pertençam ao mesmo intervalo quartílico são classificados e representados como iguais e a concentração de símbolos iguais indica uma sub-região. Se o gráfico apresentar uma variação gradativa numa dada região, isso poderá indicar uma tendência, ou falta de estacionaridade (Isaaks e Srivastava, 1989; Ribeiro Junior, 1995; Gonçalves, 1997). Não foi possível identificar nenhuma região que variasse uniformemente, indicando

estacionaridade para essas variáveis. Assim, todas as análises efetuadas não indicaram problemas que afrontassem as hipóteses necessárias ao emprego do estudo de variabilidade espacial com o emprego da geoestatística.

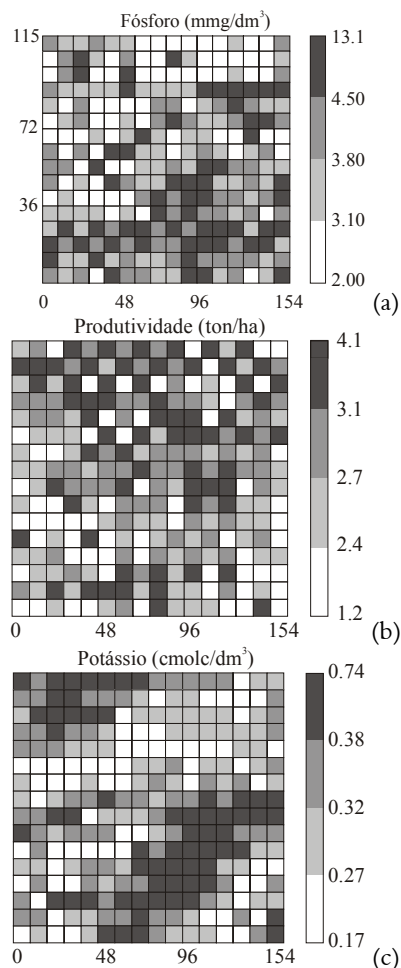


Figura 2. Post-plot das variável fósforo (a), produtividade (b) e potássio (c)

Devido à isotropia dos dados, os semivariogramas foram construídos com tolerância angular de 90° e direção angular inicial de 0°. Para as variáveis fósforo, potássio e produtividade, os melhores modelos ajustados foram o Exponencial, o Esférico e o Gaussiano, respectivamente, conforme está pode ser visualizado na Tabela 2 e Figura 3. Utilizando a classificação para o Coeficiente de Efeito Pepita (%CO) dado por Souza *et al.* (1999), verifica-se ainda que as todas as variáveis apresentaram uma variabilidade espacial média.

A validação cruzada, com o emprego de krigagem ordinária, parametrizada pelos valores obtidos no

ajuste do modelo, ofereceu uma correlação de 0,354 para o fósforo, 0,374 para produtividade e 0,663 para potássio. Isto equivale a dizer que o modelo ajustado ao semivariograma experimental, produzido para a variável potássio, oferece uma correlação mais forte, tornando-o o modelo mais bem ajustado.

Tabela 2. Parâmetros de ajuste dos semivariogramas experimentais

Variável	Modelo	C ₀	C ₀ +C ₁	a	%CO
Fósforo (ppm)	Exponencial	0,6000	1,3670	60,0	43,9
Potássio (cmolc/dm ³)	Esférico	0,0025	0,0081	43,1	30,9
Produtividade (t/ha)	Gaussiano	0,1840	0,2800	124,6	53,0

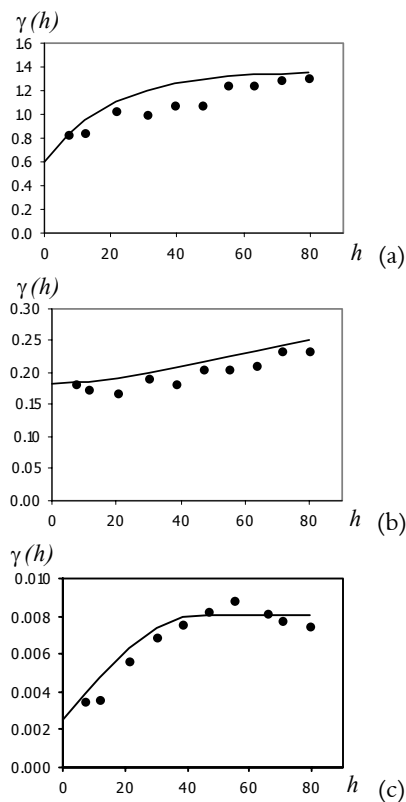


Figura 3. Semivariograma experimental das variáveis fósforo (a), produtividade (b) e potássio (c)

Com os parâmetros obtidos pelo ajuste do modelo ao semivariograma experimental e, ainda, com a krigagem ordinária, foram obtidos os mapas de contorno que descreve a variabilidade na área estudada da variável fósforo (Figura 4a), potássio (Figura 4b) e produtividade (Figura 4c). Essas figuras facilitam a visualização da variabilidade espacial e ilustram o comportamento, as características ou a distribuição potencial da variável na área de interesse.

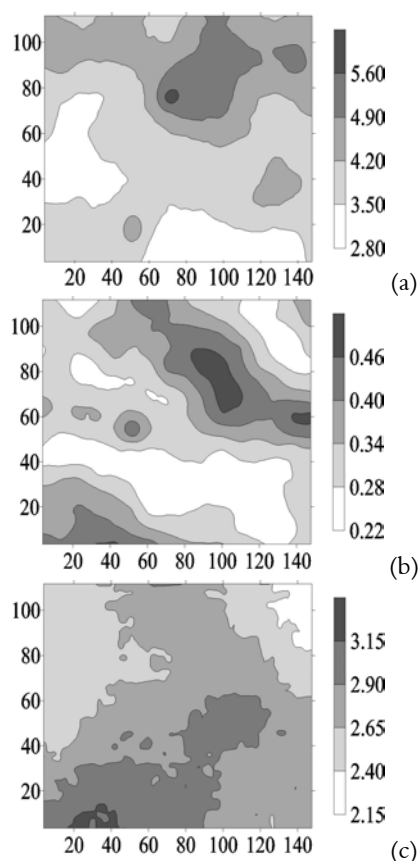


Figura 4. Mapa de contorno das variáveis fósforo (a) em mg/dm^3 , potássio (b) em cmol/dm^3 e produtividade (c) em t/ha . A distância nos eixos das figuras é dada em metros

É possível verificar pelas Figura 5, após reclassificação segundo os critérios do Manual Técnico do Subprograma de Manejo e Conservação do Solo (Paraná, 1989), que, de um modo geral, o solo apresenta nível médio de fósforo e nível alto ou muito alto de potássio. Esses níveis estão adequados para o cultivo de plantas, não caracterizando-se, portanto, como fatores limitantes à produção. Esse fato torna desnecessária adubação diferenciada, característica da agricultura de precisão.

Para verificar como as variáveis fósforo e potássio estavam correlacionadas espacialmente com a produtividade, foi construído o semivariograma cruzado (Figura 6) e foi constatado que, nos dois casos, o melhor modelo teórico ajustado aos pontos obtidos foi o esférico (Tabela 3). De acordo com o coeficiente geoestatístico padronizado, a correlação espacial das variáveis fósforo e potássio com a variável produtividade mostrou-se muito fraca, sendo praticamente nula entre o fósforo e a produtividade e discretamente positiva entre o potássio e a produtividade. O coeficiente de efeito

pepita mostra que a variabilidade de correlação espacial é fraca entre fósforo e produtividade e forte entre potássio e produtividade.

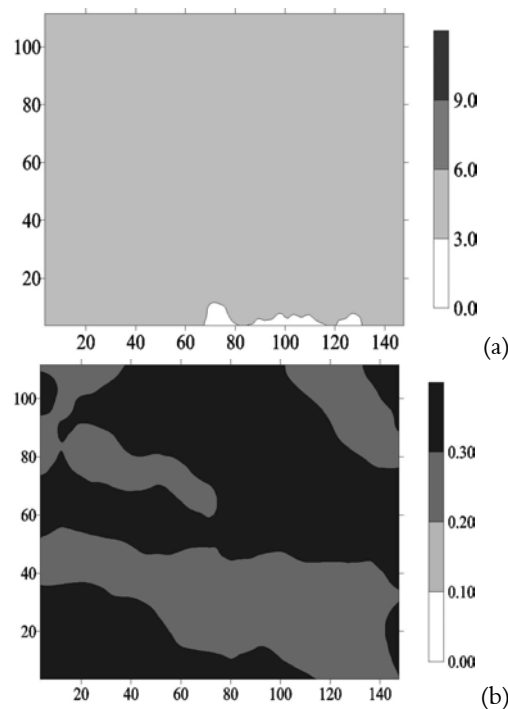


Figura 5. Mapa de contorno das variáveis fósforo (a) em mg/dm^3 e do potássio (b) em cmol/dm^3 , reclassificados segundo Paraná, 1989. A distância nos eixos das figuras é dada em metros

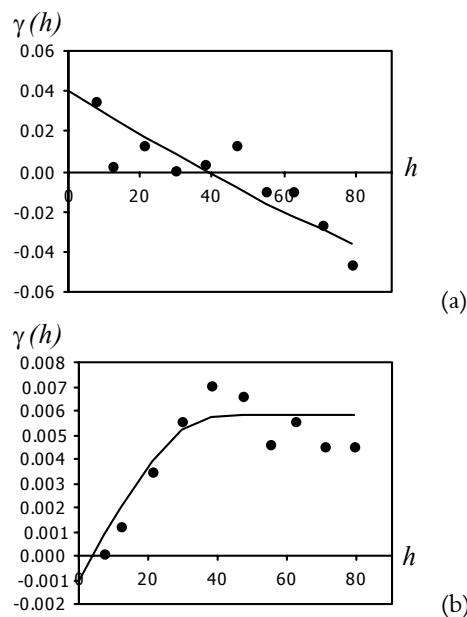


Figura 6. Semivariograma cruzado de produtividade com fósforo (a) e de produtividade com potássio (b)

Tabela 3. Parâmetros de ajuste dos semivariogramas cruzados

Variáveis	Modelo	C_0	C_0+C_1	a	%CO	CGP
Fósforo x Prod.	Esférico	0,040	-0,0500	124,0	80,0	-0,08
Potássio x Prod.	Esférico	-0,001	0,0058	40,0	17,2	0,12

Conclusão

O emprego de estimador robusto na descrição da variabilidade espacial de variáveis com distribuição assimétrica ou com cauda longa, tais como os casos aqui analisados, permitiu modelar adequadamente essas variáveis sem a necessidade de transformações. Além disso, os parâmetros obtidos para modelos ajustados com o emprego de um estimador robusto permitiram elaborar mapas de contorno do comportamento das variáveis de estudo na área pesquisada, utilizando krigagem ordinária. Por fim, a correlação espacial das variáveis fósforo e potássio com a variável produtividade por meio da modelagem do semivariograma cruzado mostrou-se muito fraca.

Referências

BRAGA, L.P.V. Geoestatística e Aplicações. In: SIMPÓSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, 9, São Paulo: IME/USP, 1990, 36p.
 CLARK, I. *Practical geostatistics*. Essex, Applied Science Publ., 1979, 129p.
 CRESSIE, N.; HAWKINS, D. M. Robust estimation of the variogram: *Mathematical Geology*. New York, v.12, n.2, p.115-125, 1980.

GONÇALVES, A.C.A. *Variabilidade espacial de propriedades físicas do solo para fins de manejo da irrigação*. 1997. Tese (Doutorado) - Esalq/Universidade de São Paulo, Piracicaba, 1997.

ISAACS, E.H.; SRIVASTAVA, R.M. *Applied Geostatistics*. New York: Oxford University Press, 1989, 561p.

MOLIN, J. P. *Agricultura de precisão Parte 1: O que é estado da arte em sensoriamento*. *Engenharia Agrícola*, Jaboticabal, v.17, n.2. p.97-107, 1997.

PARANÁ. Secretaria da Agricultura e Abastecimento. *Manual técnico do subprograma de manejo e conservação do solo*. Curitiba, 1989. 306p.

PREVEDELLO, B.M.S. *Variabilidade espacial de parâmetros do solo e planta*. 1987. Tese (Doutorado) - Esalq, Universidade de São Paulo, Piracicaba, 1987.

RIBEIRO JUNIOR, P.J. *Métodos geoestatística no estudo da variabilidade espacial de parâmetros de solo*. 1995. Dissertação (Mestrado) - Esalq, Universidade de São Paulo, Piracicaba, 1995.

SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality (complete samples). *Biometrika*, London, v. 53, n. 3/4, p.591-611, 1965.

SOUZA, E.G *et al.* Variabilidade espacial dos atributos químicos do solo em um Latossolo roxo Distrófico da região de Cascavel - PR. *Engenharia Agrícola*, Jaboticabal, v.18, n.3, p.80-92, 1999.

VIANA DA MATA, J.D. *et al.* Relação entre produtividade e resistência à penetração em área irrigada por pivô central, sob dois sistemas de preparo. *Acta Scientiarum*, Maringá, v. 21, n. 3, p. 519-525, 1997.

VIEIRA, S. R. *Geoestatística: Curso de extensão universitária*. Botucatu: Unesp, 1996.

Received on October 22, 2002.

Accepted on April 23, 2003.