

INFLUÊNCIA LOCAL EM UM MODELO ESPACIAL LINEAR DA PRODUTIVIDADE DA SOJA UTILIZANDO DISTRIBUIÇÃO T-STUDENT

NAIMARA V. DO PRADO¹, MIGUEL A. URIBE-OPAZO², MANUEL GALEA³,
ROSANGELA A. B. ASSUMPCÃO⁴

RESUMO: O uso das ferramentas da geoestatística, aliadas à agricultura de precisão permitem o acompanhamento das áreas agrícolas produtoras de soja, estabelecendo as relações de dependência espacial entre os pontos amostrados. A modelagem da estrutura de variabilidade espacial possibilita a construção de mapas temáticos dos atributos estudados, utilizando como método de interpolação a krigagem. Porém, a presença de valores atípicos entre os elementos amostrais pode influenciar na construção e interpretação desses mapas. A distribuição de probabilidades t-Student tem sido utilizada na tentativa de diminuir a influência dos valores atípicos durante a estimativa dos parâmetros de dependência espacial, por ter caudas mais pesadas que a distribuição normal. A detecção dos valores influentes na área em estudo, por meio da análise de diagnósticos de influência local, confere maior confiabilidade na utilização dos mapas gerados, corroborando a aplicação de insumos. Deste modo, o objetivo deste trabalho foi aplicar as técnicas de influência local em dados espacialmente referenciados, com os modelos de perturbação aditiva e utilizando a matriz escala, considerando a distribuição t-Student n-variada. Foi utilizado um modelo espacial linear para o estudo de dados da produtividade da soja em função da altura média de plantas e do número médio de vagens por planta. As técnicas de influência local foram eficientes para detectar pontos que influenciam na escolha do modelo geoestatístico, nas estimativas dos parâmetros e na construção do mapa temático.

PALAVRAS-CHAVE: máxima verossimilhança; variabilidade espacial; diagnósticos.

LOCAL INFLUENCE IN A LINEAR MODEL SPACE USING SOYBEAN PRODUCTIVITY WITH T-STUDENT DISTRIBUTION

ABSTRACT: The use of geostatistical tools combined with precision agriculture, allow the monitoring of agricultural soybean producing areas, establishing relationships of spatial dependence between the sampled points. The modeling of spatial variability structure enables the construction of thematic maps of the attributes studied, using the kriging as the interpolation method. However, the presence of outliers among the elements sampling can influence the construction and interpretation of these maps. The distribution of t-Student probability has been used in attempts to reduce the influence of atypical points in the estimation of parameters of spatial dependence, having heavier tails than the normal distribution of probabilities. The detection of influential points in the study area, through the analysis of local influence diagnostics, provides greater reliability in the use of maps, providing an efficient use of inputs. Then, the objective was to apply the techniques of local influence on spatially referenced data with models of disturbance and using the matrix additive scale, considering the distribution t-Student n-variate. It was used a linear spatial model for the study of soybean yield data as a function of average plant height and number of pods per plant. The local influence techniques were effective to detect points that influence in the geostatistical model selection, estimation of parameters and construction of thematic maps.

KEYWORDS: maximum likelihood; spatial variability; diagnostics.

¹ Engenheira Agrícola, Doutoranda em Estatística e Experimentação Agrônômica, ESALQ/USP, Piracicaba – SP, naimaraprado@gmail.com.

² Estatístico, Dr. em Estatística, Pesquisador de Produtividade do CNPq, Prof. Associado do PGEAGRI, UNIOESTE, Cascavel- PR, Fone: (0xx45) 3220-3228. miguel.opazo@unioeste.br.

³ Estadístico, Dr. em Estadística, Pontificia Universidad Católica de Chile, Departamento de Estadística, Chile, mgalea@mat.puc.cl.

⁴ Matemática, Doutora em Engenharia Agrícola, Prof. Adjunta, UTFPR, Toledo-PR, rosangelaa@utfpr.edu.br.

Recebido pelo Conselho Editorial em: 16-3-2012

Aprovado pelo Conselho Editorial em: 23-4-2013

INTRODUÇÃO

A agricultura de precisão utiliza ferramentas tecnológicas visando a obter maior detalhamento da área cultivada, para aplicação em taxa variada, de acordo com a necessidade de insumos e fertilizantes. A variabilidade espacial dos atributos do solo deve ser considerada quando se deseja o uso racional e equilibrado dos recursos naturais disponíveis. Assim, o gerenciamento da produção agrícola exige tecnologias e procedimentos para que os sistemas de produção sejam otimizados (MOLIN, 2008).

A geração dos mapas temáticos por krigagem não exige que os dados tenham distribuição normal, mas a presença de valores atípicos deve ser levada em consideração. No entanto, é possível encontrar trabalhos que utilizem transformações de variáveis para alcançar a normalidade. A distribuição t-Student faz parte das distribuições simétricas e apresenta como parâmetro adicional o grau de liberdade (ν), acomodando melhor os pontos atípicos. Porém, ainda é possível que a distribuição t-Student sofra o efeito de observações influentes, sendo importante a realização de estudos de sensibilidade sobre ela, por meio de análise de diagnóstico.

Para avaliar a influência dos pontos na estimativa dos parâmetros, a prática de retirada de um ou vários pontos é uma das técnicas mais conhecidas, chamada de Influência Global. Entretanto, um problema que pode ocorrer com a retirada individualizada de observações é a não detecção de pontos conjuntamente influentes. Além disso, quanto maior a quantidade de pontos amostrais, maior é o custo computacional exigido. Para diminuir estes problemas e otimizar o diagnóstico, COOK (1986) apresentou uma proposta, considerada inovadora na área de diagnóstico, na qual propõe avaliar a influência conjunta das observações sob pequenas perturbações no modelo, em vez da avaliação pela retirada individual ou conjunta de pontos, técnica denominada de “influência local”.

A metodologia de influência local (COOK, 1986; OSÓRIO et al., 2007) tem-se tornado uma ferramenta popular de diagnóstico para identificação conjunta de observações influentes em modelos de regressão, com a grande vantagem de aplicação a qualquer modelo paramétrico.

Nos últimos anos, diagnósticos de influência local têm sido apresentados por diversos autores. Por exemplo, GALEA et al. (2003) e GALEA et al. (2005) consideram estudos de influência local em modelos de regressão linear e não-linear para amostras independentes, considerando vários esquemas de perturbação. BORSSOI et al. (2009) usaram técnicas de diagnósticos de influência local em modelos espaciais gaussianos empregados em geoestatística, buscando avaliar a sensibilidade dos estimadores de máxima verossimilhança. ASSUMPÇÃO et al. (2011) estudaram técnicas de influência local para dados com distribuição t-Student, considerando a perturbação aditiva na variável resposta e os graus de liberdade fixos.

BORSSOI et al. (2011b) usaram técnicas de influência local para avaliar pontos influentes em dados agrícolas. URIBE-OPAZO et al. (2012) usam técnicas de diagnósticos de Influência Local para avaliar a sensibilidade das estimativas por máxima verossimilhança e máxima verossimilhança restrita em modelos espaciais lineares gaussianos, quando pequenas perturbações são inseridas no modelo ou nos dados. BORSSOI et al. (2011a) apresentaram técnicas de diagnóstico utilizando o método de influência local quando há perturbação na matriz de variáveis explanatórias do modelo espacial linear com distribuição gaussiana, na análise geoestatística.

O objetivo deste trabalho foi desenvolver a técnica de diagnóstico de influência local, em processos estocásticos, com distribuição t-Student, e parâmetro de grau de liberdade fixo (ν), para um modelo espacial linear, utilizando dados da produtividade da soja em função da altura média de planta e do número médio de vagens por planta, em uma área agrícola com fins comerciais, na região oeste do Paraná. Considerando, para isso, dois esquemas de perturbação, a perturbação aditiva e a perturbação de ZHU et al. (2007) que utilizam a matriz escala, para detectar pontos influentes. A aplicação destas técnicas a dados agrícolas foi realizada com intuito de verificar a existência de pontos influentes que interferem no ajuste dos modelos geoestatísticos, nas

estimativas dos parâmetros e na construção de mapas temáticos.

MATERIAL E MÉTODOS

Para o estudo da variabilidade espacial, considera-se um processo estocástico $\{Y(s), s \in S\}$, com $S \subset R^2$, sendo R^2 espaço euclidiano, bi-dimensional. Suponha-se que os elementos $Y(s_1), \dots, Y(s_n)$, deste processo, sejam registrados em localizações espaciais conhecidas s_i ($i = 1, \dots, n$), e gerados pelo modelo apresentado na Equação (1).

$$Y(s_i) = \mu(s_i) + e(s_i) \quad (1)$$

em que: o termo determinístico $\mu(s_i)$ e o termo estocástico $e(s_i)$ podem depender da localização espacial em que $Y(s_i)$ foi obtida. Assume-se que o erro estocástico e tem $E[e(s_i)] = 0$, e que a variação entre pontos no espaço é determinada por alguma função de covariância $C(s_i, s_u) = Cov[e(s_i), e(s_u)]$ e, em algumas funções conhecidas de s , como $x_1(s), \dots, x_p(s)$, $\mu(s_i)$, é definida como um modelo espacial linear:

$$\mu(s_i) = \sum_{u=1}^p x_u(s_i) \beta_u, \quad (2)$$

sendo: β_1, \dots, β_p parâmetros desconhecidos a serem estimados. De forma matricial, tem-se o modelo espacial linear :

$$Y = X\beta + e, \quad (3)$$

sendo: $E(e) = \mathbf{0}$, vetor nulo, $n \times 1$, e a matriz escala $c = [(\sigma_{iu})]$, $n \times n$, em que $\sigma_{iu} = C(s_i, s_u)$. Assume-se que Σ é não singular, que a matriz X , $n \times p$, na Equação (3) tem colunas com posto completo e que Y segue uma distribuição t-Student n -variada com vetor de médias $X\beta$ e a matriz escala Σ , isto é, $Y \sim t_n(X\beta, \Sigma, \nu)$, em que ν representa os graus de liberdade que, neste caso, são considerados fixos, $\nu > 0$.

A verificação de pontos influentes baseia-se em investigar o comportamento da função de afastamento da verossimilhança em uma vizinhança $LD(\omega)$, na qual o logaritmo da função verossimilhança dos parâmetros estimados $l(\theta)$ e o logaritmo da função verossimilhança perturbada dos parâmetros estimados $l(\theta) = l(\theta | \omega)$ são iguais. O logaritmo da função verossimilhança com a distribuição t-Student n -variada é definido pela Equação 4:

$$l(\theta) = \log \left(\frac{\Gamma\left(\frac{\nu+n}{2}\right)}{(\pi\nu)^{\frac{n}{2}} \Gamma\left(\frac{\nu}{2}\right)} \right) - \frac{1}{2} \log |\Sigma| + \log \left[1 + \frac{1}{\nu} (\mathbf{Y} - \mathbf{X}\beta)^T \Sigma^{-1} (\mathbf{Y} - \mathbf{X}\beta) \right]^{-\left(\frac{\nu+n}{2}\right)} \quad (4)$$

em que: $\theta = (\beta^T, \varphi^T)^T$, com $\beta = (\beta_1, \dots, \beta_p)^T$, $\varphi = (\varphi_1, \varphi_2, \varphi_3)^T$ e $\Sigma = \Sigma(\varphi) = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}$, sendo φ_1, φ_2 e φ_3 são os parâmetros de efeito pepita, contribuição e alcance, respectivamente, que definem a estrutura de dependência espacial. \mathbf{I}_n é a matriz identidade, $n \times n$, e \mathbf{R} representa uma matriz, $n \times n$, que é função de φ_3 e depende do modelo utilizado, isto é, $\mathbf{R} = \mathbf{R}(\varphi_3) = [(r_{iu})]$, matriz simétrica com elementos da diagonal $r_{ii} = 1$, para $i = 1, \dots, n$ e $r_{iu} = \frac{1}{\varphi_2} \sigma_{iu}$, para $i \neq u = 1, \dots, n$.

O logaritmo da função verossimilhança perturbada $l(\theta | \omega)$ é definido pela Equação 5:

$$l(\boldsymbol{\theta} | \boldsymbol{\omega}) = \log \left(\frac{\Gamma\left(\frac{\nu+n}{2}\right)}{(\pi\nu)^n \Gamma\left(\frac{\nu}{2}\right)} \right) - \frac{1}{2} \log |\boldsymbol{\Sigma}| + \log \left[1 + \frac{1}{\nu} (\mathbf{Y}_\omega - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_\omega - \mathbf{X}\boldsymbol{\beta}) \right]^{-\left(\frac{\nu+n}{2}\right)}, \quad (5)$$

em que: \mathbf{Y}_ω é o vetor de perturbação das respostas, em que $\boldsymbol{\omega}$ é um vetor pertencente a um espaço de perturbações Ω . As perturbações utilizadas neste trabalho, para a variável resposta, são: a aditiva $\mathbf{Y}_\omega = \mathbf{Y} + \boldsymbol{\omega}$ com $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ e a perturbação de ZHU et al. (2007), que utilizam a matriz escala, isto é, $\mathbf{Y}_\omega = \mathbf{Y} + \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\omega}$ com $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$.

Assume-se que existe o vetor $\boldsymbol{\omega}_0 \in \Omega$ tal que, nas Equações 4 e 5 $l(\boldsymbol{\theta}) = l(\boldsymbol{\theta} | \boldsymbol{\omega}_0)$, para todo $\boldsymbol{\theta}$ e que $l(\boldsymbol{\theta} | \boldsymbol{\omega}_0)$ seja duas vezes diferenciável em $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\varphi}^T)^T$.

A influência da perturbação $\boldsymbol{\omega}$ no estimador de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\theta}$ pode ser avaliada pelo afastamento da verossimilhança definido por:

$$LD(\boldsymbol{\omega}) = 2(l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_\omega)) \quad (6)$$

em que: $\hat{\boldsymbol{\theta}}$ é estimador de máxima verossimilhança de $\boldsymbol{\theta}$ do modelo postulado e $\hat{\boldsymbol{\theta}}_\omega$ é estimador de máxima verossimilhança de $\boldsymbol{\theta}$ do modelo perturbado.

COOK (1986) propõe a análise do comportamento local de $LD(\boldsymbol{\omega})$ em torno de $\boldsymbol{\omega}_0$, utilizando a curvatura normal D_l de $LD(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$, na direção de algum vetor unitário \mathbf{l} , definido por:

$$D_l = 2 |\mathbf{l}^T \boldsymbol{\Delta}^T \mathbf{L}^{-1} \boldsymbol{\Delta} \mathbf{l}|, \quad (7)$$

com $\|\mathbf{l}\| = 1$ e, $\boldsymbol{\Delta}$: é uma matriz de ordem $(p+3) \times n$ dada por $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta^T, \boldsymbol{\Delta}_\varphi^T)^T$ avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e em $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, cujos elementos são: $\boldsymbol{\Delta}_\beta = \frac{\partial^2 l(\boldsymbol{\theta} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T}$ e $\boldsymbol{\Delta}_\varphi = \frac{\partial^2 l(\boldsymbol{\theta} | \boldsymbol{\omega})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\omega}^T}$.

Quando se utiliza o método de máxima verossimilhança (MV) para a estimação dos parâmetros, considerando a perturbação aditiva na variável resposta $\mathbf{Y}_\omega = \mathbf{Y} + \boldsymbol{\omega}$, ou a perturbação de ZHU et al. (2007), utilizando a matriz escala, $\mathbf{Y}_\omega = \mathbf{Y} + \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\omega}$, é possível obter a matriz $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta^T, \boldsymbol{\Delta}_\varphi^T)^T$ avaliada em $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ e em $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, para cada uma das perturbações utilizadas.

Os elementos da matriz $\boldsymbol{\Delta}$, para a perturbação aditiva na variável resposta $\mathbf{Y}_\omega = \mathbf{Y} + \boldsymbol{\omega}$, são expressos por ASSUMPCÃO et al. (2011):

$$\boldsymbol{\Delta}_\beta = -2\mathbf{X}^T \boldsymbol{\Sigma}^{-1} [2W_\delta' \boldsymbol{\epsilon}_\omega \boldsymbol{\epsilon}_\omega^T \boldsymbol{\Sigma}^{-1} + W_\delta \mathbf{I}_n], \quad (9)$$

e $\boldsymbol{\Delta}_\varphi$ que tem como elementos $\boldsymbol{\Delta}_{\varphi_j}$, com $j = 1, 2, 3$,

$$\boldsymbol{\Delta}_{\varphi_j} = -2\boldsymbol{\epsilon}_\omega^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} [W_\delta' \boldsymbol{\epsilon}_\omega \boldsymbol{\epsilon}_\omega^T \boldsymbol{\Sigma}^{-1} + W_\delta \mathbf{I}_n]. \quad (10)$$

Para a perturbação proposta por ZHU et al. (2007) na variável resposta, utilizando a matriz escala $\mathbf{Y}_\omega = \mathbf{Y} + \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\omega}$, a matriz $\boldsymbol{\Delta}$ possui como elementos:

$$\boldsymbol{\Delta}_\beta = -2\mathbf{X}^T \boldsymbol{\Sigma}^{-1} [2W_\delta' \boldsymbol{\epsilon}_\omega \boldsymbol{\epsilon}_\omega^T \boldsymbol{\Sigma}^{-1} + W_\delta \mathbf{I}_n] \boldsymbol{\Sigma}^{-1/2}, \quad (11)$$

e ainda, A_φ que tem como elementos A_{φ_j} , com $j = 1, 2, 3$,

$$A_{\varphi_j} = -2\boldsymbol{\varepsilon}_\omega^T \boldsymbol{\Sigma}^{-1} \left[W_\delta' \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_\omega \boldsymbol{\varepsilon}_\omega^T \boldsymbol{\Sigma}^{-1} + W_\delta \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} + W_\delta \boldsymbol{\Sigma}^{-1/2} \frac{\partial \boldsymbol{\Sigma}^{1/2}}{\partial \varphi_j} \right] \boldsymbol{\Sigma}^{-1/2}. \tag{12}$$

\mathbf{L} : é a matriz de informação observada (hessiana), avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$;

A matriz de informação observada \mathbf{L} é definida como:

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_{\beta\beta} & \mathbf{L}_{\beta\varphi} \\ \mathbf{L}_{\varphi\beta} & \mathbf{L}_{\varphi\varphi} \end{pmatrix}, \tag{8}$$

cujos elementos são:

$$\mathbf{L}_{\beta\beta} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}; \mathbf{L}_{\beta\varphi} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \varphi_j} = \frac{\partial U(\varphi)}{\partial \boldsymbol{\beta}^T}; \mathbf{L}_{\varphi\beta} = \mathbf{L}_{\beta\varphi}^T; \mathbf{L}_{\varphi\varphi} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi \partial \varphi^T} = \frac{\partial U(\varphi)}{\partial \varphi^T},$$

sendo: $U(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}$ e $U(\varphi) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \varphi}$ as funções escores.

Considerando:

$\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ e $\delta = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ é a distância de Mahalanobis.

Seja $W_\delta = -\frac{1}{2} \left(\frac{\nu+n}{\nu+\delta} \right)$ e $W_\delta' = \frac{1}{2} \left(\frac{\nu+n}{(\nu+\delta)^2} \right)$ a derivada de W_δ em relação a δ .

Logo, para dados com distribuição t-Student, com ν graus de liberdade fixo, têm-se os elementos da matriz de informação observada:

$$\mathbf{L}_{\beta\beta} = 2W_\delta \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} + 4W_\delta' \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \mathbf{X};$$

$\mathbf{L}_{\beta\varphi}$ tem como elementos:

$$\mathbf{L}_{\beta\varphi_j} = 2W_\delta' \left[\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \left(\boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \right) \right] + W_\delta \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}, \text{ para } j = 1, 2 \text{ e } 3;$$

$\mathbf{L}_{\varphi\varphi}$ tem como elementos:

$$\mathbf{L}_{\varphi\varphi_j} = \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} + \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} \right) \right] + \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left(W_\delta' \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} + W_\delta \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} - W_\delta \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} + W_\delta \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon},$$

$j= 1, 2 \text{ e } 3.$

Observa-se que as matrizes \mathbf{A} e \mathbf{L} são definidas considerando a perturbação utilizada e a forma paramétrica da matriz escala $\boldsymbol{\Sigma}$, respectivamente.

Com as matrizes \mathbf{A} e \mathbf{L} , é possível definir a matriz $\mathbf{B} = \mathbf{A}^T \mathbf{L}^{-1} \mathbf{A}$. Seja $|L_{\max}|$ o autovetor normalizado, associado ao maior autovalor, em módulo, da matriz \mathbf{B} . O gráfico dos elementos $|L_{\max}|$ versus i (ordem dos dados) pode revelar qual ponto tem a maior influência em $LD(\boldsymbol{\omega})$, na

vizinhança de ω_0 (COOK, 1986). O gráfico $|L_{\max_2}|$ versus a ordem dos dados utiliza o autovetor normalizado, associado ao segundo maior autovalor, em módulo, da matriz \mathbf{B} , que também serve para identificar pontos influentes, considerando $C_i = 2|b_{ij}|$, em que b_{ij} são elementos da diagonal principal da matriz \mathbf{B} . Pode-se também utilizar o gráfico C_i versus i (ordem dos dados) como técnica de diagnóstico para identificar a existência de observações influentes.

No ano agrícola de 2006/2007, foram coletados dados da produtividade da soja (Prod) (Mg), altura de plantas (cm) e número de vagens por planta, totalizando 83 observações para cada uma das variáveis.

O principal interesse foi estudar a estrutura de variabilidade espacial da produtividade da soja na área em estudo, considerando como as variáveis explicativas da produtividade a altura média de planta e o número de vagens por planta, coletadas nos mesmos pontos em que foi quantificada a produtividade. Estas grandezas medidas foram utilizadas como covariáveis.

Considerando o uso de covariáveis no estudo, a média da produtividade da soja $\mu(s)$ na posição $s \in S \subset R^2$ é dada por:

$$\mu(s) = \beta_1 + \beta_2 x_2(s) + \beta_3 x_3(s), \tag{13}$$

em que: $\beta_1, \beta_2, \beta_3$ são parâmetros desconhecidos a serem estimados, x_2 e x_3 representam as covariáveis altura média de plantas e número médio de vagens por planta, respectivamente.

Os dados foram obtidos em uma área de 46,6 ha, localizada no município de Cascavel, na região oeste do Paraná, com localização aproximada de 24,83° sul de latitude e 53,60° oeste de longitude, com altitude média de 650 metros. Os dados estão publicados em URIBE-OPAZO et al. (2012). O solo da área é classificado como Latossolo Vermelho distroférico, com textura argilosa (EMBRAPA, 2009), e o clima da região é temperado superúmido, tipo climático Cfa (Köppen). Todas as amostras foram georreferenciadas com auxílio de um receptor de dados GPS (Global Positioning System) GEOEXPLORER 3, da marca Trimble, num sistema espacial de coordenadas (UTM), em um grid regular de 75x75 metros. A Figura 1 apresenta o croqui da área em estudo, no ano agrícola de 2006/2007.

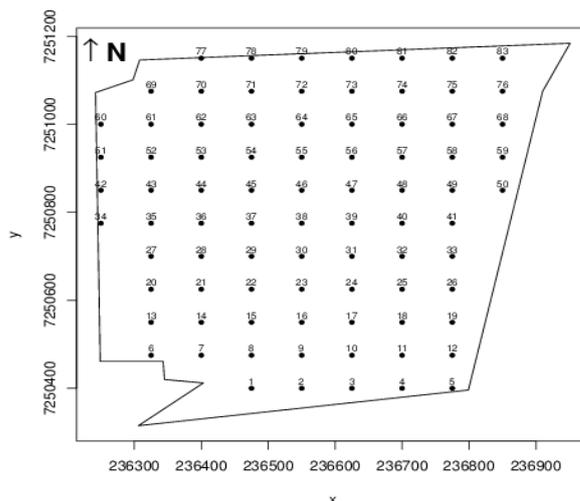


FIGURA 1. Croqui da área no ano agrícola de 2006/2007. **Points localization of the area in the agricultural year 2006/2007.**

A produtividade da soja foi estimada considerando-se a quantidade de grãos de soja colhida de todas as plantas distribuídas em duas fileiras, ao longo de um metro de comprimento,

representando a parcela. Após a trilha, foi feita a pesagem de grãos para cada parcela e verificado o teor de água para posterior correção para 13%. O valor foi convertido em Mg ha⁻¹.

A variável altura média de planta (cm), foi coletada no pico vegetativo da cultura de soja, sendo feita a média de 4 plantas medidas, em um metro linear. E para o número médio de vagens por planta, foram contados o número de vagens na época da colheita, foram escolhidas 4 plantas em cada ponto e obtido o número médio de vagens por planta.

Para a análise geoestatística, foram ajustados os modelos exponencial, gaussiano e da família Matérn com parâmetro de forma $\kappa = 1; 1,5$ e 2 . O melhor modelo ajustado foi selecionado com base nos critérios de validação cruzada (MELLO et al., 2005; FARACO et al., 2008; JOHANN et al., 2010), critério de informação de Akaike (AIC) (AKAIKE, 2002), critério de informação Bayesiano (BIC) (SCHWARZ, 1978) e o máximo valor do logaritmo da função de verossimilhança (LFML) (DE BASTIANI et al., 2012).

Para a análise de influência local, foram utilizados os gráficos C_i , $|L_{\max}|$ e $|L_{\max_2}|$ para identificar os pontos influentes na produtividade da soja. As análises geoestatísticas e de influência local foram realizadas com auxílio do software R (R DEVELOPMENT, 2011) e dos seus módulos: geoR (RIBEIRO JR. & DIGGLE, 2001), mtvnorm (GENZ et al., 2011). A interpolação para locais não amostrados foi realizada por krigagem universal.

RESULTADOS E DISCUSSÃO

Na Tabela 1, são apresentadas as estatísticas descritivas para a variável produtividade (Mg ha⁻¹).

TABELA 1. Estatísticas descritivas dos dados de produtividade da soja (Mg). **Descriptive statistics data of soybean yield (Mg ha⁻¹).**

\bar{x}	S	S^2	CV	Assimetria	Curtose
2,987	0,635	0,403	21,2740%	0,544	2,044
Mínimo	1º Quartil	Mediana	3º Quartil	Máximo	n
1,500	2,600	2,940	3,345	5,530	83

\bar{x} : média; S : desvio-padrão; S^2 : variância amostral; CV: coeficiente de variação, n: número de pontos.

Com os resultados da Tabela 1, verifica-se que a produção média de soja na área foi de 2,98

Mg ha⁻¹. Como CV > 20%, os dados são classificados como heterogêneos (PIMENTEL GOMES, 1970). A área apresentou produtividade mínima de 1,50 Mg ha⁻¹ e máximo de 5,53 Mg ha⁻¹, sendo que 25% da área apresenta produtividade igual ou superior a 3,34 Mg ha⁻¹. Os coeficientes de assimetria e curtose apresentaram valores de 0,54 e 2,04, respectivamente, indicando que os dados da produtividade da soja apresentam alguma simetria e distribuição platicúrtica.

Para a escolha do grau de liberdade (ν) da distribuição t-Student, foi calculado o logaritmo da função de verossimilhança com os parâmetros estimados para os modelos que definem a estrutura da matriz escala Σ , sendo eles, exponencial, gaussiano, família Matérn (ASSUMPTÇÃO et al., 2011), com parâmetro de forma $\kappa = 1; 1,5$ e 2 . Foram testados os graus de liberdade $\nu = 5, 6, \dots, 30$, o grau de liberdade que apresentou o valor máximo da função de verossimilhança, foi o escolhido para a análise. As covariáveis (x_2 e x_3) são conhecidas em todos os pontos de amostragem, sendo chamadas de variáveis controladas. Por simplicidade, quando for citada a variável produtividade da soja, subentende-se que as análises incluem um modelo espacial linear definido na Equação (13), com as covariáveis altura média de planta e número médio de vagens por planta, sendo estas variáveis espacialmente georreferenciadas.

No gráfico *boxplot*, apresentado na Figura 2, observa-se que existe um ponto atípico, sendo um valor de máximo, que corresponde ao ponto 6 (5,53 Mg ha⁻¹), destacado no gráfico *postplot* da área experimental, na Figura 2.

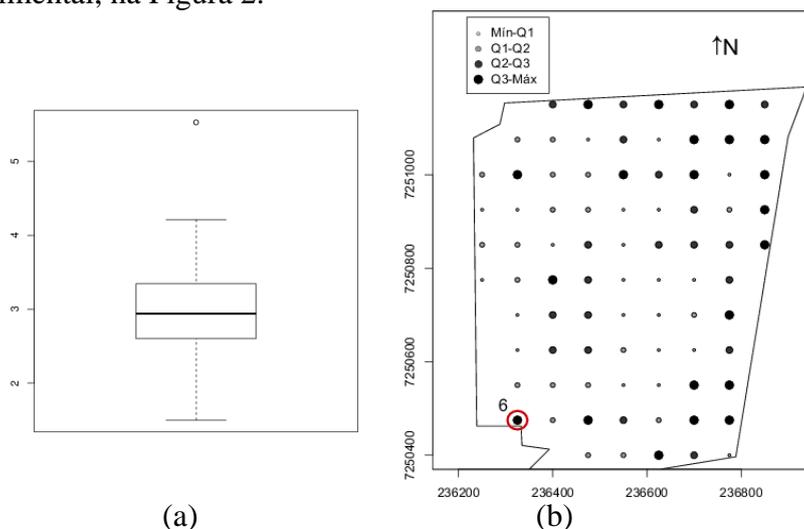


FIGURA 2. Gráficos *boxplot* e *postplot* para os dados de produtividade da soja no ano agrícola 2006/2007: (a) *Boxplot* dos dados de produtividade da soja, (b) *Postplot* dos dados de produtividade da soja. **Boxplot and postplot data for soybean yield in the agricultural year 2006/2007: (a) Boxplot of soybean yield data, (b) Postplot data soybean yield.**

Para estimar o vetor de parâmetros $\theta = (\beta_1, \beta_2, \beta_3, \phi_1, \phi_2, \phi_3)^T$ do modelo espacial linear da Equação (13) e da estrutura de variabilidade espacial com cada modelo, foi utilizado o método de máxima verossimilhança com auxílio do algoritmo BFGS (FRERY & CRIBARI NETO, 2009). Na Tabela 2, apresenta-se o resumo das estimativas obtidas considerando que o grau de liberdade da distribuição t-Student foi definido de acordo com o valor máximo do logaritmo da função de verossimilhança (LFML). O grau de liberdade da distribuição t-Student, que apresentou o maior valor da função de verossimilhança para todos os modelos, foi $\nu = 15$.

TABELA 2. Valor do logaritmo da função de máxima verossimilhança (LFML) e parâmetros estimados para a produtividade da soja, no ano agrícola de 2006/2007, para os modelos da estrutura de variabilidade espacial utilizados, com os respectivos graus de liberdade da distribuição t-Student. **Value function of the logarithm of maximum likelihood (LFML) and estimated parameters for soybean yield in the agricultural year 2006/2007, for models of spatial variability of the structure used, the respective degrees of freedom of t-Student distribution.**

Modelo	ν	LFML	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
Exponencial	15	-69,56	0,9932 (0,1362)	0,0211 (0,0139)	0,0303 (0,0075)	0,2436 (0,1112)	0,1195 (0,0845)	132,7417 (0,9357)
Gaussiano	15	-69,42	0,9546 (0,1439)	0,0219 (0,0321)	0,0308 (0,0041)	0,2840 (0,0830)	0,0819 (0,1155)	229,4728 (0,6428)
Matérn ($\kappa = 1$)	15	-69,45	0,9800 (0,1284)	0,0215 (0,0158)	0,0304 (0,0089)	0,2535 (0,1026)	0,0915 (0,0864)	95,8598 (0,0878)
Matérn ($\kappa = 1,5$)	15	-69,44	0,9722 (0,0150)	0,0217 (0,0652)	0,0304 (0,0069)	0,2715 (0,1031)	0,0836 (0,0917)	83,7966 (0,0638)
Matérn ($\kappa = 2$)	15	-69,44	0,9675 (0,1622)	0,0218 (0,0652)	0,0305 (0,0051)	0,2563 (0,1044)	0,0701 (0,0719)	74,8758 (0,0880)

κ : Parâmetro de forma; ν : grau de liberdade da distribuição t-Student; $\hat{\beta}_1$: estimativa de β_1 ; $\hat{\beta}_2$ estimativa de β_2 ; $\hat{\beta}_3$ estimativa de β_3 ; $\hat{\phi}_1$: efeito pepita; $\hat{\phi}_2$: contribuição; $\hat{\phi}_3$: função do alcance; Os valores entre parenteses indicam o desvio-padrão das estimativas.

Para os modelos utilizados na análise espacial, as estimativas dos parâmetros $\theta = (\beta_0, \beta_1, \beta_2, \varphi_1, \varphi_2, \varphi_3)^T$ apresentaram baixos desvios-padrão. O menor valor de φ_3 foi obtido para o modelo da família Matérn com $\kappa = 2$, cujo valor foi de 74,8758 que corresponde a um alcance prático de aproximadamente 401,33 metros. Para a escolha do melhor modelo ajustado, foram utilizados os critérios de validação cruzada, Critério de Informação de Akaike (AIC) e Critério de Informação Bayesiano (BIC). Na validação cruzada, foi escolhido o modelo que apresentou o erro médio (EM) e erro médio reduzido (ER) mais próximo de zero, o menor valor de S_{EM} e EA e o valor de S_{ER} mais próximo de 1. Os valores de AIC e BIC devem ser os menores possíveis.

Na Tabela 3, são apresentados os valores que atendem aos critérios estabelecidos pela validação cruzada, destacados em negrito; o modelo com maior número de critérios em negrito foi o escolhido.

TABELA 3. Validação cruzada, Critério de Informação de Akaike (AIC) e Critério de Informação Bayesiano (BIC) para os modelos exponencial, gaussiano e família Matérn com parâmetro de forma $\kappa = 1; 1,5$ e 2, para os respectivos graus de liberdade da distribuição t-Student. **Cross-validation, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for models exponential, Gaussian and Maternam family with shape parameter $\kappa = 1, 1.5$ and 2, for the corresponding degrees of freedom t-Student distribution.**

Modelo	ν	EM	ER	S_{EM}	S_{ER}	EA	AIC	BIC
Exponencial	15	0,0019	0,0017	0,6165	1,1125	38,1727	147,1200	161,2142
Gaussiano	15	0,0016	0,0015	0,6134	1,0905	37,3186	146,8400	160,9342
<u>Matérn</u> ($\kappa = 1$)	15	0,0015	0,0014	0,6114	1,1009	37,3308	146,9000	160,9942
Matérn ($\kappa = 1,5$)	15	0,0016	0,0014	0,6117	1,0858	37,3298	146,8800	160,9742
Matérn ($\kappa = 2$)	15	0,0016	0,0014	0,6119	1,1297	37,3174	146,8600	160,9542

κ : Parâmetro de forma; ν : grau de liberdade da distribuição t-Student; EM: erro médio; ER: erro médio reduzido; S_{EM} : Desvio-padrão do erro médio; S_{ER} : desvio-padrão dos erros médios reduzidos; EA: erro absoluto.

Para a produtividade da soja, o modelo que melhor descreve a estrutura de variabilidade espacial foi o da família Matérn, com parâmetro de forma $\kappa = 1$, e $\nu = 15$ graus de liberdade da distribuição t-Student. Neste modelo, os parâmetros estimados foram $\hat{\beta}_1 = 0,9800$, $\hat{\beta}_2 = 0,0215$, $\hat{\beta}_3 = 0,0304$, $\hat{\varphi}_1 = 0,2715$, $\hat{\varphi}_2 = 0,0836$ e $\hat{\varphi}_3 = 95,8598$, com alcance prático de 383,44 metros.

Assim, considerando o uso das covariáveis, $x_2(s)$ representa a altura média de plantas e $x_3(s)$ o número médio de vagens por planta na posição s , e a média estimada $\hat{\mu}(s)$ da produtividade da soja pode ser expressa por: $\hat{\mu}(s) = 0,9800 + 0,0215x_2(s) + 0,0304x_3(s)$.

Como o modelo da família Matérn com parâmetro $\kappa = 1$, com $\nu = 15$ graus de liberdade para a distribuição t-Student, foi o que melhor representa a estrutura de dependência espacial da produtividade da soja na área em estudo, foram construídos os gráficos para análise de diagnóstico de influência local, considerando a perturbação aditiva na variável resposta $Y_\omega = Y + \omega$ (Figura 3) e considerando a perturbação de ZHU et al. (2007) na variável resposta, utilizando a matriz escala, $Y_\omega = Y + \Sigma^{-1/2}\omega$ (Figura 4).

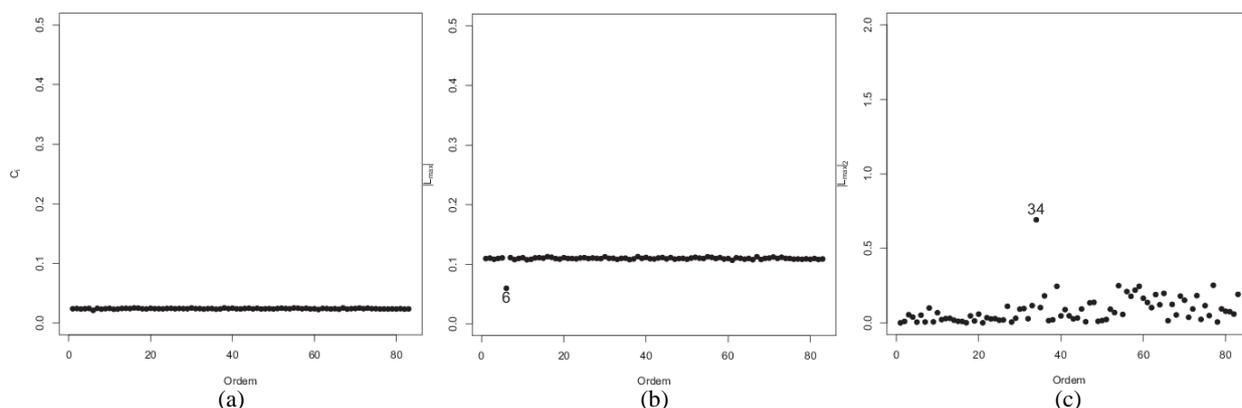


FIGURA 3. Gráficos de diagnóstico de influência local para a produtividade da soja no ano agrícola de 2006/2007, com o modelo da estrutura de variabilidade espacial da família Matérn e parâmetro $\kappa = 1$ e $\nu = 15$ graus de liberdade da distribuição t-Student, considerando a perturbação aditiva na variável resposta ($Y_\omega = Y + \omega$): (a) C_i versus i ; (b) $|L_{max}|$ versus i ; (c) $|L_{max_2}|$ versus i . **Charts for diagnosis of local influence soybean yield in the agricultural year 2006/2007, with the model structure and spatial variability of parameters and Matern family with 15 degrees of freedom of t-Student distribution, considering the additive disturbance in the response variable ($Y_\omega = Y + \omega$): (a) C_i versus i ; (b) $|L_{max}|$ versus i ; (c) $|L_{max_2}|$ versus i .**

As Figuras 3(b) e 3(c) apresentam os gráficos $|L_{max}|$ e $|L_{max_2}|$ versus i , em que i indica a ordem dos dados para a produtividade da soja, considerando a perturbação aditiva na variável resposta, e destacam o ponto 6 ($5,53 \text{ Mg ha}^{-1}$) e o ponto 34 ($2,41 \text{ Mg ha}^{-1}$), respectivamente, como influentes na variável resposta.

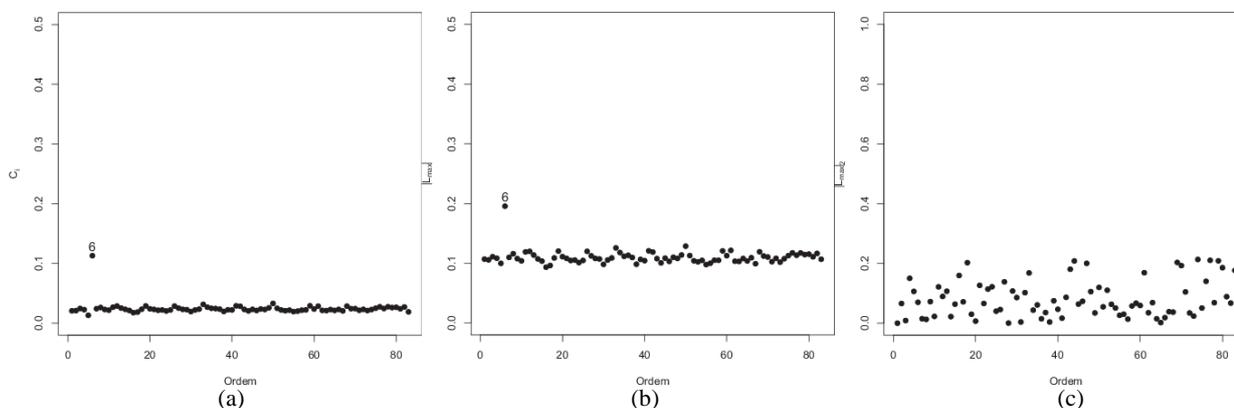


FIGURA 4. Gráficos de diagnóstico de influência local para a produtividade da soja no ano agrícola de 2006/2007 com o modelo da estrutura de variabilidade espacial da família Matérn com $\kappa = 1$ e $\nu = 15$ graus de liberdade para a distribuição t-Student, considerando a perturbação na variável resposta, utilizando a matriz escala ($Y_\omega = Y + \Sigma^{-1/2}\omega$): (a) C_i versus i ; (b) $|L_{max}|$ versus i ; (c) $|L_{max_2}|$ versus i . **Charts for diagnosis of local influence soybean yield in the agricultural year 2006/2007 with the model of the structure of spatial variability of the Matern family and 15 degrees of freedom for the Student-t distribution, considering the disturbance in the response variable using the matrix scale ($Y_\omega = Y + \Sigma^{-1/2}\omega$): (a) C_i versus i ; (b) $|L_{max}|$ versus i ; (c) $|L_{max_2}|$ versus i .**

As Figuras 4(a) e 4(b) apresentam os gráficos C_i , $|L_{max}|$ versus i , em que: i indica a ordem dos dados. Como o ponto 6 aparece distante dos demais pontos no gráfico, este pode ser considerado influente na variável resposta, mediante o esquema de perturbação proposto por ZHU et al. (2007), que utiliza a matriz escala ($Y_\omega = Y + \Sigma^{-1/2}\omega$). O ponto 6 pode ter sido considerado influente, por ser um ponto com alta produtividade cercado por vizinhos com valores baixos, conforme mostra a Figura 1.

Como o ponto 6 foi detectado como influente, tanto pela perturbação aditiva $Y_\omega = Y + \omega$ como pela perturbação proposta por ZHU et al. (2007) $Y_\omega = Y + \Sigma^{-1/2}\omega$, o mesmo foi retirado do conjunto de dados da produtividade da soja, sendo o novo conjunto denominado por Prod[6].

URIBE-OPAZO et al. (2012) também detectaram o ponto 6 como influente ao trabalhar com os mesmos dados e a perturbação aditiva em modelos lineares espaciais gaussianos, na área em estudo. Pode-se verificar que, apesar de trabalhar com a distribuição t-Student que engloba pontos atípicos, pois é considerada uma distribuição com cauda mais pesada que o normal, o ponto continua sendo influente.

A estatística descritiva para o conjunto Prod[6] é apresentada na Tabela 4. Com a retirada do ponto 6, houve redução no coeficiente de variação, quando comparados com os valores apresentados na Tabela 1. A produtividade média praticamente não foi alterada com a retirada do ponto de máximo. Os coeficientes de assimetria e curtose apresentaram valores de -0,153 e -0,047, respectivamente, indicando que os dados da produtividade da soja são considerados simétricos e com distribuição platicúrtica.

TABELA 4. Estatísticas descritivas dos dados de produtividade da soja ($Mg\ ha^{-1}$) com a retirada do ponto 6. **Descriptive statistics data of soybean yield ($Mg\ ha^{-1}$) with remove of item 6.**

\bar{x}	S	S^2	CV	Assimetria	Curtose
2,956	0,572	0,328	19,375%	-0,153	-0,047
Mínimo	1º Quartil	Mediana	3º Quartil	Máximo	N
1,500	2,600	2,930	3,308	4,210	82

\bar{x} : média; S : desvio-padrão; S^2 : variância amostral; CV: coeficiente de variação, n: número de pontos.

Para o conjunto Prod[6], foi estimado o vetor de parâmetros $\theta = (\beta_1, \beta_2, \beta_3, \phi_1, \phi_2, \phi_3)^T$ utilizando o método de máxima verossimilhança, com o auxílio do algoritmo BFGS. Foram ajustados os modelos exponencial, gaussiano, família Matérn com $\kappa = 1; 1,5$ e 2, para modelar a estrutura de variabilidade espacial dos dados. O melhor modelo ajustado foi escolhido segundo os critérios de validação cruzada, AIC e BIC. De acordo com o máximo valor da função de verossimilhança, o grau de liberdade escolhido para a distribuição t-Student foi $\nu = 30$; isso indica que, com a retirada do ponto, o conjunto de dados assume a distribuição t-Student com configuração mais próxima da distribuição normal de probabilidades (grau de liberdade $\nu \rightarrow \infty$).

Para os dados de produtividade da soja com a retirada do ponto 6, o melhor modelo ajustado foi o gaussiano, com parâmetros $\hat{\beta}_1 = 1,8349$, $\hat{\beta}_2 = 0,0032$, $\hat{\beta}_3 = 0,0265$, $\hat{\phi}_1 = 0,2316$, $\hat{\phi}_2 = 0,0341$, e $\hat{\phi}_3 = 161,2393$, com alcance prático de 279,08 metros. O máximo valor da função de verossimilhança foi LFML = - 60,85, com AIC = 129,70 e BIC = 143,73.

A Figura 5 apresenta os mapas temáticos para a produtividade com 83 elementos amostrais e o mapa temático com a retirada do ponto 6 Prod[6], confeccionados por meio da interpolação por krigagem universal, visto que, neste estudo, foram introduzidas covariáveis ou variáveis explicativas. Os mapas foram divididos em classes iguais para facilitar a comparação visual dos mesmos.

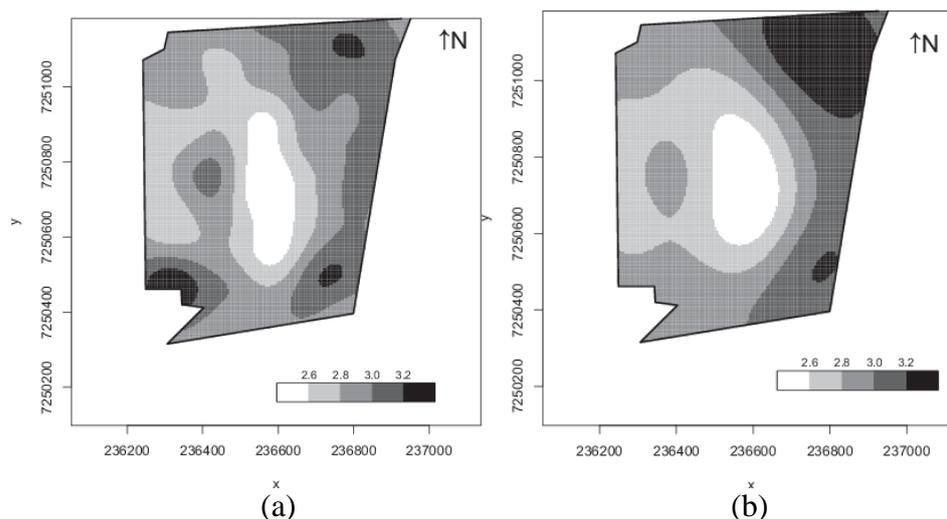


FIGURA 5. Mapas temáticos para a produtividade da soja (Mg ha^{-1}) no ano agrícola de 2006/2007: (a) Produtividade da soja (Mg ha^{-1}) com 83 pontos amostrais, e (b) Produtividade da soja (Mg) com a retirada do ponto 6. **Thematic maps for soybean yield (Mg ha^{-1}) in the agricultural year 2006/2007: (a) soybean yield (Mg ha^{-1}) with 83 sampling points and (b) soybean yield (Mg ha^{-1}) with remove of item 6.**

Pelas Figuras 5(a) e 5(b), pode-se observar que há variação entre as áreas dos mapas. O mapa obtido para a variável produtividade com o conjunto original, com 83 pontos (Figura 5(a)), apresenta pequenas porções isoladas com produtividade acima de $3,2 \text{ Mg ha}^{-1}$. Na Figura 5(b), onde houve a retirada do ponto influente, verifica-se que houve aumento da região superior do mapa com produtividade superior a $3,2 \text{ Mg ha}^{-1}$. Porém, a região central dos dois mapas apresenta baixa produtividade, com valores inferiores a $2,6 \text{ Mg ha}^{-1}$. Estas constatações indicam que a observação 6 exerce influência na interpolação por krigagem com tendência, que foi o método utilizado por inserir covariáveis no estudo. Logo, provoca alterações nos mapas gerados.

A Tabela 5 apresenta a porcentagem de área em cada classe dos mapas apresentados na Figura (5), bem como a porcentagem de diferença em cada classe dos mapas, considerando que a área em estudo possui 47,951 ha. Pode-se observar que a classe 3,0 a $3,2 \text{ Mg ha}^{-1}$ foi a que apresentou maior diferença entre os mapas, seguida da classe de maior produtividade, de 3,2 a $3,4 \text{ Mg ha}^{-1}$. Porém, com a retirada do ponto 6, é que surge uma região de maior produtividade na região norte do mapa apresentado na Figura 6(b).

TABELA 5. Porcentagem de área em cada classe dos mapas temáticos da produtividade da soja. **Percentage of area in each class of thematic maps of soybean yield.**

Classes	% da área do mapa construído com 83 pontos amostrais	% da área do mapa construído com 82 pontos amostrais*	% da diferença entre os mapas
2,4 - 2,6	9,54	9,96	-0,42
2,6 - 2,8	26,16	28,01	-1,85
2,8 - 3,0	28,35	30,74	-2,39
3,0 - 3,2	26,46	18,78	7,68
3,2 - 3,4	9,49	12,51	-3,02
Total	100,00	100,00	

*sem considerar o ponto influente.

A aplicação da técnica de diagnóstico de influência local mostrou-se eficiente na identificação dos pontos influentes. A presença de pontos influentes em um conjunto de dados exige maior atenção na construção dos mapas temáticos e, posteriormente, em sua utilização a campo, para que os mesmos representem a realidade da área. Assim, será possível a aplicação correta de insumos à taxa variável e a limitação de zonas de manejo, permitindo que a área seja trabalhada de acordo com suas necessidades específicas.

CONCLUSÃO

Verificou-se que as técnicas de diagnóstico em influência local, quando aplicadas para dados experimentais, foram eficientes para identificar a presença de valores atípicos, que foram considerados influentes na determinação da estrutura de dependência espacial, na escolha do melhor modelo ajustado e na construção dos mapas temáticos.

O uso da distribuição t-Student como alternativa no estudo de dados com valores atípicos pode diminuir a interferência destes pontos. A identificação da influência destes, utilizando as duas perturbações propostas neste trabalho, oferece maior confiabilidade no estudo, pois, mesmo alterando o esquema de perturbação, o ponto continua sendo identificado como influente.

O uso da técnica de diagnóstico de influência local, para avaliar a influência de observações no conjunto de dados e nos resultados das análises, deve fazer parte das análises geoestatísticas, pois, desta forma, garante-se que as informações contidas nos mapas temáticos tenham maior qualidade e possam ser recomendadas ao agricultor.

AGRADECIMENTOS

Os autores agradecem ao CNPq, CAPES e Fundação Araucária, pelo apoio financeiro.

REFERÊNCIAS

- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. *Journal of the Royal Statistical Society*, London, v. 51, p. 469-483, 2002.
- ASSUMPÇÃO, R. A. B.; URIBE-OPAZO, M. A.; GALEA, M. Local Influence for spatial analysis of soil physical properties and soybean yield Student's t-distribution. *Revista Brasileira de Ciência do Solo*, Viçosa-MG, v.35, n.5, p.1.917-1.926, 2011.
- BORSSOI, J. A.; URIBE-OPAZO, M. A.; GALEA, M. Diagnostic techniques applied in geostatistics for agricultural data analysis. *Revista Brasileira de Ciência do Solo*, Viçosa-MG, v. 33, n.6, p.1.561-1.570, 2009.
- BORSSOI, J. A.; DE BASTIANI, F.; URIBE-OPAZO, M. A.; GALEA, M. Local influence of explanatory variables in Gaussian spatial linear models. *Chilean Journal of Statistics*, Santiago, v.2, n.2, p. 29-38. 2011a.
- BORSSOI, J. A.; URIBE-OPAZO, M.; GALEA, M. Técnicas de diagnóstico de influência local na análise espacial da produtividade da soja. *Engenharia Agrícola*, Jaboticabal, v.31, n.2, p.376-387, mar./abr. 2011b.
- COOK, R. D. Assessment of Local Influence (with discussion). *Journal of the Royal Statistical Society*, Series B, London, v.48, n. 2, p. 133-169, 1986.
- DIGGLE, P. J.; RIBEIRO JR., P. J. *Model-based geostatistics*. 2007, 230p. Hardcover.
- DE BASTIANI, F.; URIBE-OPAZO, M. A.; DALPOSSO, G. H. Comparison of maps of spatial variability of soil resistance to penetration constructed with and without covariables using a spatial linear model. *Engenharia Agrícola*, Jaboticabal, v. 32, n.2, p. 394-404, 2012.
- EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. Centro Nacional de Pesquisa de Solos. Sistema brasileiro de classificação de solos. Brasília, 2009. 412p.

- FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A.; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. *Revista Brasileira de Ciência do Solo*, Viçosa-MG, v. 32, n.2, p. 463-476, 2008.
- FRERY, A. C.; CRIBARI-NETO, F. Elementos de estatística computacional usando plataformas de software livre/gratuito: *Publicações Matemáticas*, Rio de Janeiro: IMPA, 2009, 99p.
- GALEA, M.; PAULA, G. A.; URIBE-OPAZO, M. A. On influence diagnostic in univariate elliptical linear regression models. *Statistical Papers*, New York, v. 44, n.1, p.23-45, 2003.
- GALEA, M.; PAULA, G. A.; CYSNEIROS, F. J. A. On diagnostics in symmetrical nonlinear models. *Statistics e Probability Letters*. v. 73, n.4, p. 459-467, 2005.
- GENZ, A. et al. mvtnorm: Multivariate Normal and t Distributions. R package version 0.9-9991. Disponível em: <<http://CRAN.R-project.org/package=mvtnorm>>. Acesso em: 5 abr. 2011.
- JOHANN, J. A.; SILVA, M. C. A.; URIBE-OPAZO, M. A.; DALPOSSO, G. Variabilidade da rentabilidade, perdas na colheita e produtividade do feijoeiro. *Engenharia Agrícola*, Jaboticabal.v. 30, n.4, p. 700-714, jul./ago. 2010.
- MELLO, J. M.; BATISTA, J. L. F.; RIBEIRO JR, P. J.; OLIVEIRA, M. S. Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus grandis*. *Scientia Florestalis*, Piracicaba, v.1, n. 69, p. 25-37, 2005.
- MOLIN, J. P. *Agricultura de precisão: o gerenciamento da variabilidade*. 3. ed. Piracicaba: Escola Superior de Agricultura Luiz de Queiroz, USP, 2008. 83p.
- OSORIO, F.; PAULA, G. A.; GALEA, M. Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics & Data Analysis Journal*, Amsterdam, v. 51, n.9, p. 4.354-4.368, 2007.
- PIMENTEL, G. *Curso de estatística experimental*. Piracicaba: Escola Superior de Agricultura Luiz de Queiroz, USP, 1970. 208p.
- R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna. Disponível em: <<http://www.R-project.org>>. Acesso em: 1 set. 2011.
- RIBEIRO JR., P. J.; DIGGLE, P. J. *geoR: a package for geostatistical analysis*. R-NEWS, New York, v.1, n.2, p.15-18, 2001.
- SCHWARZ, G. Estimating the dimension of a model. *Annals of Statistics*, Hayward, v.6, n.2, p.461-464, 1978.
- URIBE-OPAZO, M.; BORSSOI, J. A.; GALEA, M. Influence diagnostics in gaussian spatial linear models. *Journal of Applied Statistics*, Abingdon, v.39, n.3, p.615-630, 2012.
- ZHU, H.; IBRAHIM, J. G.; LEE, S.; ZHANG, H. Perturbation selection and influence measures in local influence analysis. *Annals of Statistics*, Hayward, v. 35, n. 6, p.2.565-2.588, 2007.