

ESTUDO DE PONTOS INFLUENTES NA RESISTÊNCIA DO SOLO A PENETRAÇÃO EM MODELOS ESPACIAIS LINEARES GAUSSIANOS

MIGUEL A. URIBE-OPAZO¹, MANUEL GALEA², FERNANDA DE BASTIANI³, GUSTAVO H. DALPOSSO⁴

¹ Programa de Pós - Graduação em Engenharia Agrícola –PGEAGRI, da Universidade Estadual do Oeste do Paraná – UNIOESTE, Cascavel – PR, Fone: (45) 3220-3228, miguel.opazo@unioeste.br

² Departamento de Estadística, Pontificia Universidad Católica de Chile.

³ Departamento de Estatística, Universidade Federal de Pernambuco.

⁴ COMAT, Universidade Tecnológica Federal do Paraná.

Apresentado no
XLVI Congresso Brasileiro de Engenharia Agrícola - CONBEA 2017
30 de julho a 03 de agosto de 2017 - Maceió - AL, Brasil

RESUMO: Neste trabalho apresentam-se técnicas de diagnóstico de influência local em modelos espaciais lineares Gaussianos, no estudo da dependência espacial de dados da resistência do solo a penetração na camada de 0,0 a 0,10 m de profundidade. A área em estudo é de 57 hectares. Construiu-se o semivariograma experimental e sua respectiva banda de confiança pelo método bootstrap. Os parâmetros foram estimados pelo método de máxima verossimilhança e o modelo final selecionado pelo critério de validação. Mapas de contornos com todas as observações e sem os pontos influentes foram construídos por krigagem. Foi verificada a normalidade dos dados utilizando os resíduos descorrelacionados. Como resultado foi identificado um ponto influente no qual provocou alterações na estimação dos parâmetros, seleção do modelo e nos valores preditos para a construção do mapa de contorno.

PALAVRAS-CHAVE: Dependência espacial, Diagnóstico, Geoestatística.

INFLUENTIAL POINTS STUDY ON GAUSSIAN SPATIAL LINEAR MODELS FOR SOIL RESISTANCE OF PENETRATION

ABSTRACT: In this work we present local influence diagnostics techniques in Gaussian spatial linear models to study the spatial dependence in soil resistance of penetration data in the 0.0-0.10 m layer of depth. The experimental area has 57 hectares. Experimental semivariograms were constructed with the respective confidence band via bootstrap. Maximum likelihood method is considered to obtain the parameters estimates and the cross-validation criterion is used to select the final. Maps with all observations and no influential points were constructed by kriging. The assumption of normality in the data was verified using the uncorrelated residues. One influential poin was identified which caused changes in the parameter estimates, model selection and in the predicted values used to construct the map that shows the spatial variability of the mean of the soil resistance of penetration.

KEYWORDS: Diagnostics, Geostatistics, Spatial dependence.

INTRODUÇÃO: A geoestatística é um método de análise da dependência espacial de variáveis aleatórias espacialmente georreferenciadas. Ela tem como princípio que os elementos amostrais de uma variável se dispõem em forma correlacionada até certa distância, e elas têm influência sobre os pontos próximos não coletados a serem estimados (CRESSIE, 2015). No estudo da dependência espacial os resultados obtidos pelos métodos geoestatísticos podem ser influenciados pela presença de dados discrepantes, podendo-se exercer um peso desproporcional nas estimações dos parâmetros do modelo, na significância dos parâmetros e obter previsões tendenciosas (URIBE-OPAZO et al. 2012). Este trabalho tem como objetivo apresentar os modelos espaciais lineares Gaussianos no estudo da dependência espacial e analisar pontos influentes por meio da análise de diagnósticos de influencia local proposto por COOK (1986) utilizando o esquema de perturbação na variável resposta apropriado.

MATERIAL E MÉTODOS: Seja um processo estocástico $\{Z(\mathbf{s}), \mathbf{s} \in S\}$, com $S \subset R^2$. Os elementos do vetor aleatório $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$ são registrados nos lugares $\mathbf{s}_i, i = 1, \dots, n$ e \mathbf{Z} pode ser escrito por um modelo espacial linear $\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, em que $\boldsymbol{\varepsilon}$ é o vetor de erros aleatórios com $E(\boldsymbol{\varepsilon}) = \mathbf{0}$; \mathbf{X} é uma matriz $n \times p$ de posto completo de p covariáveis e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é o vetor $p \times 1$ de parâmetros desconhecidos. Seja $\boldsymbol{\Sigma} = [(\sigma_{iu})]$, $n \times n$, a matriz covariância não singular com $\sigma_{iu} = Cov[\boldsymbol{\varepsilon}(\mathbf{s}_i), \boldsymbol{\varepsilon}(\mathbf{s}_u)]$, assume-se que $\mathbf{Z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, com $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\varphi}) = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3)$, em que φ_1, φ_2 e φ_3 são os parâmetros: efeito pepita, contribuição e função do alcance, respectivamente; \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 ajustado. A influência local pode ser avaliada pela diferença entre as verossimilhanças com e sem a perturbação, definida como $LD(\boldsymbol{\omega}) = 2 \left(l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}) \right)$, em que, $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ são os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$ do modelo postulado e do modelo perturbado por $\boldsymbol{\omega}$, respectivamente. COOK (1986) propõe estudar o comportamento local de $LD(\boldsymbol{\omega})$ ao redor de $\boldsymbol{\omega}_0 \in \Omega$, tal que $l(\boldsymbol{\theta}) = l(\boldsymbol{\theta} | \boldsymbol{\omega}_0)$, por meio da curvatura normal na direção de algum vetor unitário, definindo: $C_l = 2 | \mathbf{l}^T \boldsymbol{\Delta}^T \mathbf{L}^{-1} \boldsymbol{\Delta} \mathbf{l} |$, em que $\| \mathbf{l} \| = 1$; \mathbf{L} : é a matriz Hessiana avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$; $\boldsymbol{\Delta}$: é uma matriz, $(p + 3) \times n$, dada por $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\boldsymbol{\beta}}^T, \boldsymbol{\Delta}_{\boldsymbol{\varphi}}^T)^T$, avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e em $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Um esquema de perturbação para a variável resposta para modelos espaciais lineares Gaussianos é apresentado por DE BASTIANI et al. (2015), $\mathbf{Z}_{\boldsymbol{\omega}} = \mathbf{Z} + \boldsymbol{\Sigma}^{1/2} \boldsymbol{\omega}$, em que: $\boldsymbol{\Delta}_{\boldsymbol{\beta}} = -\mathbf{X}^T \boldsymbol{\Sigma}^{-1/2}$ e $\boldsymbol{\Delta}_{\boldsymbol{\varphi}}$ tem como elementos $\boldsymbol{\Delta}_{\varphi_j} = \boldsymbol{\varepsilon}^T [\boldsymbol{\Sigma}^{-1} (\partial \boldsymbol{\Sigma} / \partial \varphi_j) - \mathbf{D}_j \mathbf{A}]$, com $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\boldsymbol{\omega}_0) = (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$ e $\mathbf{D}_j = \boldsymbol{\Sigma}^{-1} (\partial \boldsymbol{\Sigma} / \partial \varphi_j) \boldsymbol{\Sigma}^{-1}$, para $j=1,2,3$. Para verificação da suposição de normalidade, ordenou-se os resíduos descorrelacionados (SOLOW, 1985) plotando-os versus os quantis teóricos da distribuição normal padrão, resultando o gráfico *qq-plot*. Os dados em estudo referem-se à resistência do solo a penetração (RSP) [Mpa] na camada de 0,0 a 0,10 m de profundidade do ano agrícola 2004/2005, em uma área com 57 ha. Os elementos amostrais foram georreferenciados com um GPS e a distância entre pontos foi de 75 x 75 m. As análises foram realizadas com o *software* gratuito R.

RESULTADOS E DISCUSSÕES: São apresentadas as semivariâncias obtidas por Matheron (CRESSIE, 2015) e os intervalos de 95% de confiança utilizando o método percentil de Efron (EFRON, 1982) (Tabela 1). A Figura 1 apresenta o gráfico do semivariograma experimental e as bandas de confiança da RSP na camada 0,0 a 0,10 m de profundidade.

Tabela 1. Intervalos bootstrap de 95% de confiança para a função semivariância.

h	$\gamma(h)$	LI	LS
0,07	0,3928263	0,2642954	0,5375818
0,14	0,4163526	0,2970848	0,5864695
0,21	0,4032218	0,2996580	0,6182100
0,28	0,3972936	0,2971017	0,6107450
0,35	0,4752909	0,3023787	0,6175271
0,42	0,4611496	0,2982209	0,6167010
0,49	0,4164217	0,2983421	0,6117259
0,56	0,4059023	0,3015051	0,6213486
0,63	0,4461946	0,2973614	0,6296881
0,70	0,4423422	0,2974299	0,6382823

h : distância; $\gamma(h)$: função semivariância; LI: limite inferior; LS: limite superior.

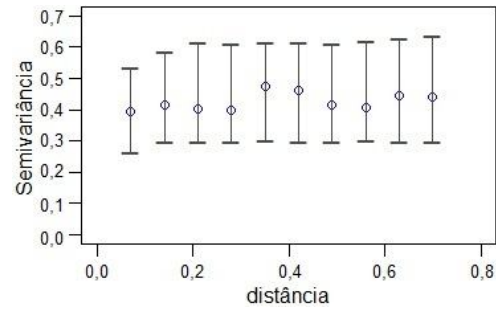


Figura 1. Gráfico do semivariograma experimental com intervalos bootstrap de 95% de confiança para a RSP na camada 0,0 a 0,10 m de profundidade.

No gráfico *box-plot* (Figura 2) observa-se quatro *outliers*: 01, 55, 89 e 90. No gráfico *postplot* (Figura 2) identifica-se os *outliers* e observa-se valores altos na região Leste. O gráfico de coeficientes C_i versus i (Figura 2) considera as observações 54, 61 e 87 como influentes e o gráfico $|L_{max}|$ versus i (Figura 2) considera as observações 07, 61 e 92 como influente.

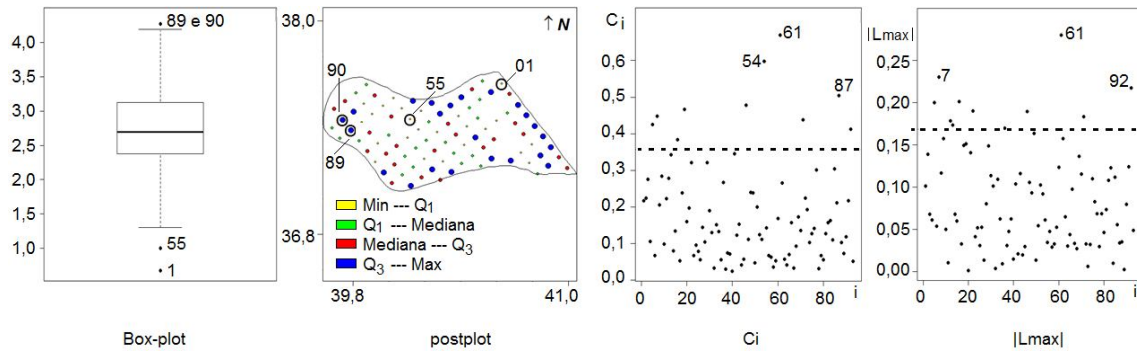


Figura 2. Gráficos *box-plot*, *postplot* e de influência local coeficientes C_i e $|L_{max}|$.

Na Tabela 2 são apresentados os melhores ajustes de modelos espaciais realizados por validação cruzada (MLL) com e sem o ponto influente #61. Observa-se que existem diferenças entre as estimativas de $\hat{\phi}_1$, $\hat{\phi}_2$ e $\hat{\phi}_3$. Os valores estimados para o alcance (\hat{a}) variam de 126,0 a 151,0 m. O coeficiente de efeito pepita relativo E indica forte dependência espacial ($E \leq 25\%$) para o modelo sem o ponto influente #61 e moderada dependência espacial ($25\% < E < 75\%$) para o modelo ajustado com todos os dados.

Tabela 2. Análise da estrutura de dependência espacial para a RSP, utilizando ML

Método	Modelo	$\hat{\beta}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	\hat{a} (m)	E (%)	MLL
ML	Gaussiano	2,739	0,317	0,127	0,087	151,0	71,40	-92,83
	(total)	(0,089)	(0,083)	(0,088)	(0,00039)			
ML	Exponencial	2,739	0,099	0,383	0,042	126,0	15,66	-92,15
	(sem # 61)	(0,093)	(0,236)	(0,259)	(0,00005)			

$\hat{\beta}$: média estimada dos dados; $\hat{\phi}_1$: efeito pepita estimado; $\hat{\phi}_2$: contribuição estimada; $\hat{\phi}_3$: função do alcance estimado; \hat{a} : alcance estimado (m); E : $(\hat{\phi}_1 / (\hat{\phi}_1 + \hat{\phi}_2)) \cdot 100\%$ efeito pepita relativo; ML: máxima verossimilhança; MLL: máximo valor do logaritmo da função de verossimilhança; entre parêntesis são apresentados os desvios padrões.

A Figura 3 apresenta os mapas da RSP com todos os pontos e sem o ponto influente # 61. observa-se a diferença em certos setores devido à eliminação do ponto influente #61.

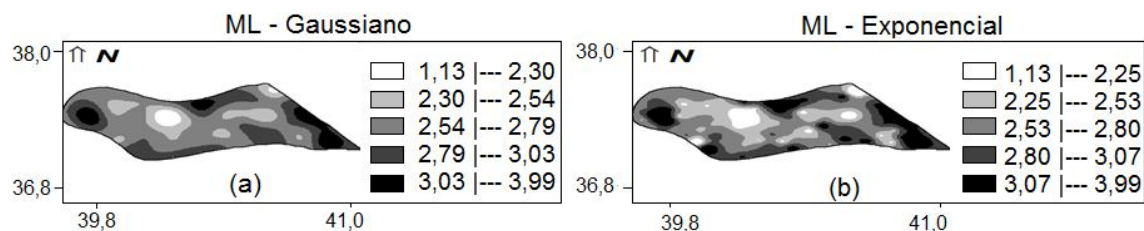


Figura 3. Mapas de contorno da RSP (Mpa) na camada 0,0 a 0,10 m de profundidade construídos com todos os dados (a) e sem o ponto influente (b).

Na Figura 4 verifica-se que os resíduos descorrelacionados encontram-se sobre a linha de referência, ou seja, não existe indício de pontos discrepantes e de anomalias nas caudas das distribuições e, portanto, é coerente assumir que a RSP na camada 0,0 a 0,10 m de profundidade do solo seguem distribuição normal.

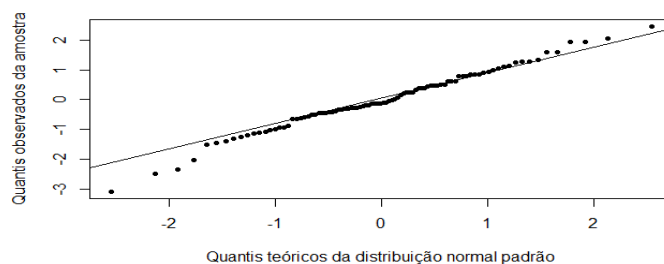


Figura 4. Gráfico *qq-plot* dos resíduos descorrelacionados da RSP (Mpa) na camada 0-0,10 m de profundidade.

CONCLUSÕES: As técnicas de diagnósticos de influência local identificaram um ponto influente, o qual provoca alterações na estimação dos parâmetros, na seleção do modelo e na construção do mapa.

AGRADECIMENTOS

A CNPq, CAPES, FONDECYT e FUNDAÇÃO ARAUCÁRIA pelo apoio financeiro.

REFERÊNCIAS

- CRESSIE,N. Statistics for spatial data. New York: John Wiley & Sons, 2015.
- COOK,R. Assessment of local influence. **Journal of the Royal Statistical Society**, Serie B 48, 133–169, 1986.
- DE BASTIANI,F., CYSNEIROS,A.H.M.A., URIBE-OPAZO,M.A., GALEA. M. Influence diagnostics in elliptical spatial linear models. **Test**, v.24, n.2, p.322-340, 2015.
- EFRON,B. **The jackknife, the bootstrap and other resampling plans**. Philadelphia: SIAM, 1982.
- SOLOW,A. Bootstrapping correlated data. **Mathematical Geology**, v.17, n.7, p.769-775, 1985.
- URIBE-OPAZO, M.A., BORSSOI, J.A., GALEA,M. Influence diagnostics in gaussian spatial linear models. **Journal of Applied Statistics**, v.3, n.39, p.615-630, 2012.