Chilean Journal of Statistics
Vol. 2, No. 2, September 2011, 29–38

SPATIAL STATISTICS AND IMAGE MODELING
SPECIAL ISSUE: "III SEEMI"
RESEARCH PAPER

Local influence of explanatory variables in Gaussian spatial linear models

JOELMIR A. BORSSOI[1], FERNANDA DE BASTIANI[2],
MIGUEL A. URIBE-OPAZO[3]* AND MANUEL GALEA[4]

[1]Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo-SP, Brazil
[2]Departamento de Estatística, Universidade Federal de Pernambuco, Recife-PE, Brazil
[3]Centro de Ciências Exatas e Tecnológicas,
Universidade Estadual do Oeste do Paraná, Cascavel-PR, Brazil
[4]Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile

(Received: 04 March 2011 · Accepted in final form: 20 July 2011)

Abstract

Modeling the structure of spatial dependence by the geostatistical approach is fundamental for the estimation of parameters that define this structure. This is also used in the interpolation of values at unsampled locations by the kriging technique. It is well known the presence of outliers in sampled data can greatly affect the estimation of parameters. The goal of this work is to present diagnostic techniques using the local influence method when there is perturbation in the matrix of explanatory variables of the Gaussian spatial linear model useful for geostatistical analysis. Studies with experimental data have shown that the presence of outliers may cause variations in construction of maps by changing the structure of the spatial dependence. The application of techniques of local influence must be part of all geostatistical analysis to ensure that the information contained in thematic maps are of better quality.

Keywords: Geostatistic · Maximum likelihood · Spatial variability.

Mathematics Subject Classification: Primary 62M30 · Secondary 62J20.

1. INTRODUCTION

In modeling the spatial variability of regionalized variables, the method of maximum likelihood (ML) estimation is used to estimate the parameters that define the structure of spatial dependence. This is also used in the interpolation of values in areas not sampled by the kriging technique (see, e.g., Stein, 1999), generating thematic maps that could be used for a localized treatment in the study area. The quality of these maps depends on the inferences for the models chosen. In order for the kriging technique to give reliable predictions and represent the true local variability, the modeling process should be done with caution, especially in the presence of outliers or influential observations.

*Corresponding author. Miguel A. Uribe-Opazo. Rua Universitária 2069, Sala 62, CEP 85819-110, Cascavel, Paraná, Brazil. Email: miguel.opazo@unioeste.br

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

Cook (1986) developed a diagnostic technique called the local influence method. Cook (1986)'s method has become a popular diagnostic tool for the identification of influential observations in regression. Many papers have been published on this subject. Beckman et al. (1987) presented studies on influence of analysis of variance models with mixed effect. Paula (1993) developed local influence on linear models with restriction in the parameters in the form of linear inequalities. Pan et al. (1997) applied methods of local influence in multivariate analysis. Galea et al. (1997, 2003) and Liu (2000) presented studies of local influence on linear elliptical models. Zhu and Lee (2001) proposed a procedure to assess local influence for incomplete data. Zhu et al. (2007) constructed influence measures by assessing local influence of perturbations to a statistical model. Chen and Zhu (2009) proposed a perturbation selection method for selecting an appropriate perturbation with desirable properties and then developed a second-order local influence measure on the basis of the selected perturbation, in the context of general latent variable models.

The goal of this paper is to apply the local influence method to Gaussian spatial linear models in order to verify if there are observations that cause some kind of influence on likelihood displacement (LD) when there are perturbations on the covariates matrix. This is a continuation of the work by Borssoi et al. (2009), who considered an additive perturbation of the responses vector.

The paper unfolds as follows. Section 2 presents the Gaussian spatial linear model. Section 3 describes the local influence method based on the LD and two perturbation schemes on the explanatory variables. Section 4 carries out an application to illustrate the methodology developed in this paper. The data set is included in the appendix. Finally, Section 5 contains some concluding remarks.

2. Gaussian Spatial Linear Model

Consider a Gaussian stochastic process $\{Z(s), s \in S \subset \Re^d\}$, with $\Re^d$ being a d-dimensional Euclidean space ($d \geq 1$). Suppose that the elements $Z(s_1), \ldots, Z(s_n)$ of this process are recorded in known spatial locations s_i, for $i = 1, \ldots, n$, and generated from the model

$$Z(s_i) = \mu(s_i) + \epsilon(s_i),$$

with the deterministic and stochastic terms, $\mu(s_i)$ and $\epsilon(s_i)$ respectively, may depend on the spatial location at where $Z(s_i)$ is obtained. It is assumed that the stochastic error $\epsilon(\cdot)$ has $\mathrm{E}[\epsilon(s_i)] = 0$. The variation between points in the space is determined by a function of covariance $C(s_i, s_u) = \mathrm{Cov}[\epsilon(s_i), \epsilon(s_u)]$, for $i, u = 1, \ldots, n$. In addition, for some known functions of s, say $x_1(s), \ldots, x_p(s)$, $\mu(s_i)$ is defined by a spatial linear model given by

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

with $\beta_1, \ldots, \beta_p$ being unknown parameters to be estimated. Equivalently, in matrix notation,

$$Z = \mathbf{X}\boldsymbol{\beta} + \epsilon,$$

where $\mathrm{E}(\epsilon) = \mathbf{0}$ (null vector) and the covariance matrix is $\boldsymbol{\Sigma} = [(\sigma_{iu})]$, with $\sigma_{iu} = C(s_i, s_u)$. Assume $\boldsymbol{\Sigma}$ is an $n \times n$ nonsingular matrix, $\mathbf{X}$ is a full column rank matrix, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is the vector of coefficient of the model. Then, Z has a n-variate normal distribution with mean $\mathbf{X}\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Sigma}$, i.e., $Z \sim \mathrm{N}_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$.

We focus on a particular parametric form for the covariance matrix (see Mardia and Marshall, 1984) given by

$$\mathbf{\Sigma} = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R},$$

where φ_1 is the nugget effect or variance error, φ_2 is the dispersion variance, $\mathbf{R}$ is a matrix that is function of φ_3, and $\mathbf{R} = \mathbf{R}(\varphi_3) = [(r_{iu})]$ is an $n \times n$ symmetric matrix with diagonal elements $r_{ii} = 1$, for $i = 1, \ldots, n$, where φ_3 is function of range of the model and $\mathbf{I}_n$ is the $n \times n$ identity matrix.

The parametric form of the covariance matrix occurs for various isotropic processes, where the covariance $C(s_i, s_u)$ is defined according to the covariance function $C(\delta_{iu}) = \varphi_2 r_{iu}$, where $\delta_{iu} = ||s_i - s_u||$ is the Euclidean distance between points s_i and s_u. In the covariance functions $C(\delta_{iu})$, the variance of the stochastic process Z is $C(0) = \varphi_1 + \varphi_2$ and the variogram can be defined as $\gamma(\delta) = C(0) - C(\delta)$.

3. Local Influence on Likelihood Displacement

The study of outliers and detection of influential observations is an important step in analysis of georeferenced data. The local influence method suggested by Cook (1986) evaluates the simultaneous effect of observations on the ML estimator without removing it from the data set.

Let $\mathcal{L}(\boldsymbol{\theta})$ be the log-likelihood function for the postulated model and $\boldsymbol{\omega}$ a perturbation vector belonging to a perturbation space Ω. It is assumed that exists $\boldsymbol{\omega}_0 \in \Omega$ such that $\mathcal{L}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$. The influence of the perturbation $\boldsymbol{\omega}$ on the ML estimator can be evaluated by the LD given by $\mathrm{LD}(\boldsymbol{\omega}) = 2\{\mathcal{L}(\hat{\boldsymbol{\theta}}) - \mathcal{L}(\hat{\boldsymbol{\theta}}_\omega)\}$, where $\hat{\boldsymbol{\theta}}$ is the ML estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top)^\top$ in the postulated model, with $\boldsymbol{\beta}^\top = (\beta_1, \ldots, \beta_p)$, $\boldsymbol{\varphi}^\top = (\varphi_1, \varphi_2, \varphi_3)$ and $\hat{\boldsymbol{\theta}}_\omega$ is the ML estimator of $\boldsymbol{\theta}$ in the perturbed model by $\boldsymbol{\omega}$. Cook (1986) proposed to study the local behavior of $\mathrm{LD}(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$. He showed that the normal curvature C_l of $\mathrm{LD}(\boldsymbol{\omega})$, at $\boldsymbol{\omega}_0$ in direction of some unit vector l, is given by $C_l = 2|l^\top \mathbf{\Delta}^\top \mathbf{L}^{-1} \mathbf{\Delta} l|$, with $||l|| = 1$, where $\mathbf{L}$ is the observed information matrix, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\mathbf{\Delta}$ is a $(p+3) \times m$ matrix given by $\mathbf{\Delta} = (\mathbf{\Delta}_\beta^\top, \mathbf{\Delta}_\varphi^\top)^\top$, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Here $m = \dim(\Omega)$.

Let $\mathbf{B} = \mathbf{\Delta}^\top \mathbf{L}^{-1} \mathbf{\Delta}$ and $l_{\max}$ be the normalized eigenvector in absolute value associated with the largest eigenvalue of $\mathbf{B}$ matrix. The index plot of the elements of $|l_{\max}|$ may reveal what type of perturbation has higher influence on $\mathrm{LD}(\boldsymbol{\omega})$, around ω_0; see Cook (1986). In addition, given the matrix $\mathbf{B}$, $C_i = 2|b_{ii}|$ can be considered as cut point, where b_{ii} are the elements of the main diagonal of matrix $\mathbf{B}$. We can also use the index plot of C_i to evaluate the existence of influential observations.

In relation to the perturbation schemes of the predictors, Thomas and Cook (1990) indicated that: "Local modifications of the explanatory variables may not be meaningful when these variables are discrete". Then, it is supposed that the design matrix $\mathbf{X}$ is composed of discrete and continuous predictors and considered the partition $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, where $\mathbf{X}_1$ is an $n \times p_1$ submatrix of discrete predictors (for example $\mathbf{X}_1$ can be a column vector 1's or a submatrix of predictors that is not relevant to perturb) and $\mathbf{X}_2$ a $n \times p_2$ submatrix of continuous predictors, with $p_1 + p_2 = p$. We consider two perturbation schemes of the form $\mathbf{X}_\omega = [\mathbf{X}_1, \mathbf{X}_{\omega 2}]$; see Cook (1986), Thomas and Cook (1990) and Zhu et al. (2007).

Scheme 1. Consider the perturbation

$$\mathbf{X}_{\omega 2} = \mathbf{X}_2 + \mathbf{W}\mathbf{S},$$

where $\mathbf{W} = [(\omega_{ij})]$, for $i = 1, \ldots, n$ and $j = 1, \ldots, p_2$, is an $n \times p_2$ matrix of perturbations and $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_{p_2})$ is a $p_2 \times p_2$ matrix, with $s_1, \ldots, s_{p_2}$ being respectively the estimates of the standard deviations of the p_2 covariates perturbed. In this case, the perturbation vector is given by the $np_2 \times 1$ vector $\boldsymbol{\omega} = \mathrm{vec}(\mathbf{W})$.

Scheme 2. Another perturbation scheme is

$$\mathbf{X}_{\omega 2} = \mathbf{X}_2 + \mathbf{s}^\top \otimes \boldsymbol{\omega},$$

where $\mathbf{s} = (s_1, \ldots, s_{p_2})^\top$ is a vector $p_2 \times 1$, $\boldsymbol{\omega}$ a $n \times 1$ vector of perturbation and $\otimes$ is the Kronecker product of two matrices.

For Schemes 1 and 2, the perturbed log-likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(Z - \mathbf{X}_\omega\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(Z - \mathbf{X}_\omega\boldsymbol{\beta}).$$

Then, using results of derivatives of matrices, we have, for Scheme 1 (see Nel, 1980),

$$\boldsymbol{\Delta}_{1\beta} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top} = (\mathbf{I}_p \otimes \mathrm{vec}^\top(\mathbf{I}_n))(\mathbf{H} \otimes \mathbf{I}_n),$$

where $\mathbf{H} = \mathbf{A}\{\hat{\boldsymbol{\Sigma}}^{-1}(Z - \mathbf{X}\hat{\boldsymbol{\beta}}) \otimes \mathbf{I}_{p_2}\} - \{\mathrm{vec}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})\beta_2^\top\mathbf{S}\}$ is an $np \times p_2$ matrix, with

$$\mathbf{A} = \begin{pmatrix} \mathbf{O} \\ (\mathbf{S} \otimes \mathbf{I}_n)\mathbf{K}_{p_2 n} \end{pmatrix},$$

$\mathbf{O}$ being an $np_1 \times np_2$ zero matrix and $\mathbf{K}_{p_2 n}$ a $p_2 n \times p_2 n$ commutation matrix, i.e., $\mathbf{A}$ is an $np \times np_2$ matrix. Finally, β_2 is the vector $p_2 \times 1$ obtained from the partition $\boldsymbol{\beta} = (\beta_1^\top, \beta_2^\top)^\top$.

Using a similar procedure, we obtain the matrix of second derivatives

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\omega}^\top},$$

with elements

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\varphi_j\partial\boldsymbol{\omega}^\top} = -\frac{1}{2}\beta_2^\top\mathbf{S}\{(\mathbf{K}_{p_2} + \mathbf{I}_{p_2}) \otimes (Z - \mathbf{X}_\omega\boldsymbol{\beta})^\top(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\varphi_j}\boldsymbol{\Sigma}^{-1})\}, \quad j = 1, 2, 3.$$

Evaluating at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and at $\boldsymbol{\omega} = 0$, we obtain the $\boldsymbol{\Delta}_\varphi$ matrix for Scheme 1, which takes the form $\boldsymbol{\Delta}_{1\varphi} = (\Delta_{1\varphi_1}^\top, \Delta_{1\varphi_2}^\top, \Delta_{1\varphi_3}^\top)^\top$.

For Scheme 2, we obtain

$$\boldsymbol{\Delta}_{2\beta} = \mathbf{A}(\hat{\boldsymbol{\Sigma}}^{-1}(Z - \mathbf{X}\hat{\boldsymbol{\beta}}) \otimes \mathbf{I}_n) - (\mathbf{s}^\top\hat{\beta}_2)\mathbf{X}^\top\hat{\boldsymbol{\Sigma}}^{-1},$$

where now

$$\mathbf{A} = \begin{pmatrix} \mathbf{O} \\ \mathbf{s} \otimes \text{vec}^\top(\mathbf{I}_n) \end{pmatrix},$$

with $\mathbf{O}$ being a $p_1 \times n^2$ zero matrix and $\boldsymbol{\Delta}_{2\varphi} = (\Delta_{2\varphi_1}^\top, \Delta_{2\varphi_2}^\top, \Delta_{2\varphi_3}^\top)^\top$, with

$$\Delta_{2\varphi_j} = (\mathbf{s}^\top \hat{\beta}_2)(Z - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \hat{\boldsymbol{\Sigma}}^{-1} \frac{\partial \hat{\boldsymbol{\Sigma}}}{\partial \varphi_j} \hat{\boldsymbol{\Sigma}}^{-1}, \quad j = 1, 2, 3.$$

Note that if $p_2 = 1$, i.e., perturbed a single predictor, both schemes coincide.

The observed information matrix has the form

$$\mathbf{L} = \begin{pmatrix} L_{\beta\beta} & L_{\beta\varphi} \\ L_{\varphi\beta} & L_{\varphi\varphi} \end{pmatrix},$$

where $L_{\beta\beta} = -\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X}$, $L_{\beta\varphi} = \partial^2 \mathcal{L}(\boldsymbol{\theta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\varphi}^\top$, with elements

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \varphi_j} = -\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \epsilon, \quad j = 1, 2, 3$$

and $\epsilon = (Z - \mathbf{X}\boldsymbol{\beta})$; $L_{\varphi\beta} = L_{\beta\varphi}^\top$ and $L_{\varphi\varphi} = \partial^2 \mathcal{L}(\boldsymbol{\theta}) / \partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^\top$, with elements

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \varphi_i \partial \varphi_j} &= \frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} - \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} \right) \right\} \\ &\quad + \frac{1}{2} \epsilon^\top \boldsymbol{\Sigma}^{-1} \left\{ \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} - \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} - \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \right\} \boldsymbol{\Sigma}^{-1} \epsilon. \end{aligned}$$

4. Application

This section contains an application of the local influence method for Gaussian spatial linear model. We analyze a real data set and focus on the $\boldsymbol{\theta}$ parameter. For the geostatistical analysis, we use the R software (`http://www.R-project.org`); see Ribeiro and Diggle (2001) and R Development Core Team (2009).

A sample of 47 points are collected in the crop year 2006/2007 in a 57 ha commercial area, in western region of Paraná State, Brazil. The response variable is the soybean yield (z) [$t\,\text{ha}^{-1}$] and as covariates the soil resistance to penetration through layers 0-0.1m (x_2), 0.1-0.2m (x_3) and 0.2-0.3m (x_4) of depth. The points were georeferenced in a systematic sampling –lattice plus close pairs–, with a maximum distance of 141m between points and in some locations with distances of 75m and 50m between points. These points were located with a GPS receiver signal GeoExplorer 3, in a system of spatial coordinates (UTM). The data collected are displayed in Tables 4 and 3 presented in Appendix.

The mean soybean yield is defined by

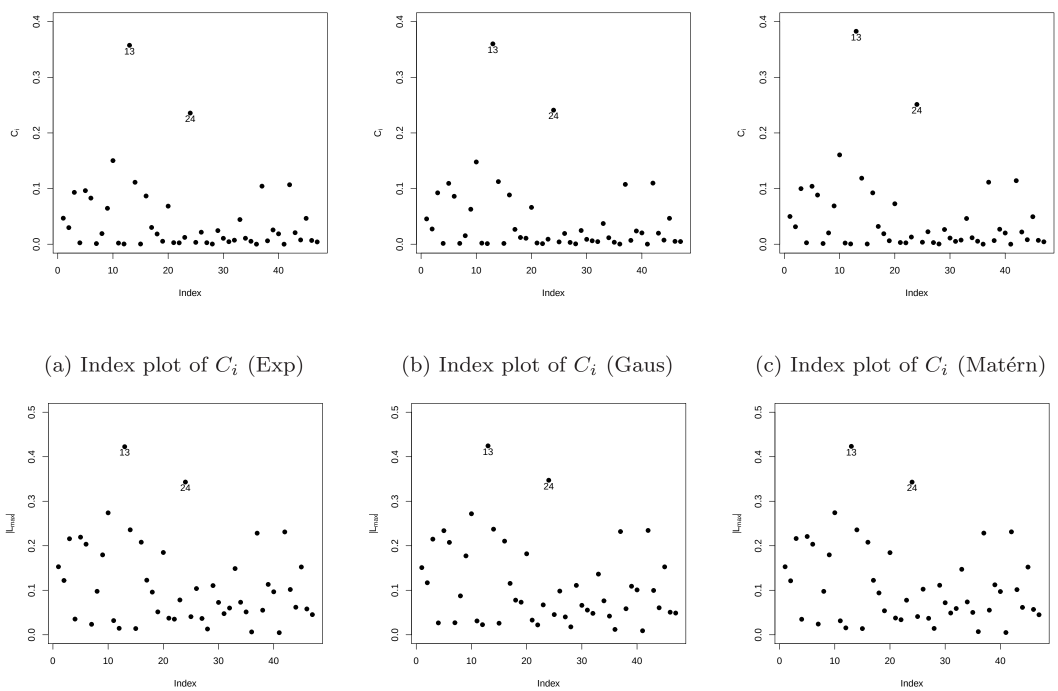
$$\mu(s) = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4,$$

where β_1, β_2, β_3 and β_4 are parameters to be estimated by the ML method. Table 1 shows the covariance functions used for modeling the structure of the spatial variability and the ML estimates of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top)^\top$. The respective asymptotic standard errors are in parentheses.

Table 1. ML estimates of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top)^\top$ for the indicated covariance functions.

Covariance	β_1	β_2	β_3	β_4	φ_1	φ_2	φ_3
Exp	3.4444	−0.0001	−0.1444	0.0672	0.0739	0.0565	125.2862
	(0.3549)	(0.0886)	(0.1062)	(0.1232)	(0.0360)	(0.0462)	(1.2927)
Gaus	3.4390	−0.0046	−0.1394	0.0675	0.0911	0.0392	183.7579
	(0.3503)	(0.0872)	(0.1042)	(0.1224)	(0.0242)	(0.0295)	(0.4281)
Matérn	3.4458	−0.0004	−0.1442	0.0665	0.0809	0.0496	112.2012
($\kappa = 0.7$)	(0.3546)	(0.0885)	(0.1060)	(0.1232)	(0.0313)	(0.0406)	(0.3314)

We consider the case of perturbing a single covariate; see Thomas and Cook (1990). Indeed, the covariate (x_4) corresponding to the soil resistance to penetration through layer 0.2-0.3m of depth is perturbed. In this case, $p_2 = 1$ so that, as mentioned, the two perturbation schemes discussed coincide. Figure 1 shows the index plots of C_i and $|l_{\max}|$ for the exponential, Gaussian and Matérn ($\kappa = 0.7$) covariance functions. It may be noted that observations #13 and #24 are potentially influential in all cases.

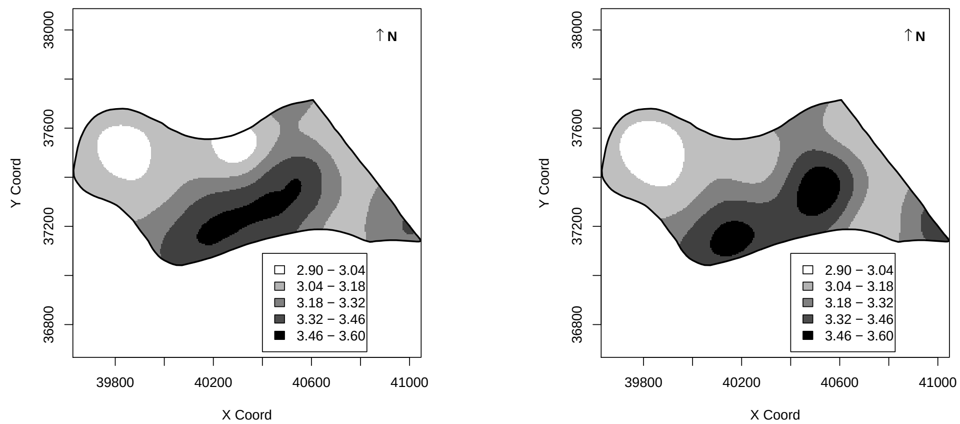


(a) Index plot of C_i (Exp) (b) Index plot of C_i (Gaus) (c) Index plot of C_i (Matérn)

(d) Index plot of $|l_{\max}|$ (Exp) (e) Index plot of $|l_{\max}|$ (Gaus) (f) Index plot of $|l_{\max}|$ (Matérn)

Figure 1. Index plots of C_i and $|l_{\max}|$ for the indicated covariance functions.

Using cross-validation criterion and the maximum value of the logarithm of likelihood function (see Faraco et al., 2008), the Gaussian covariance function is chosen to illustrate the effect of observations identified as potentially influential in thematic maps. We remove observations #13 and #24, which are detected by the plot shown in Figure 1. Consequently, the parameter estimates of the model change to $\beta_1 = 3.4034$, $\beta_2 = -0.0120$, $\beta_3 = -0.0632$, $\beta_4 = 0.0011$, $\varphi_1 = 0.0574$, $\varphi_2 = 0.0408$ and $\varphi_3 = 182.0224$. It can be observed that there is a difference between the maps with all the data and without observations #13 and #24; see Figures 2a) and 2b), respectively.



(a) With all the data (b) Without observations #13 and #24

Figure 2. Estimated thematic maps using the Gaussian covariance function.

Table 2 is derived from soybean thematic maps generated with and without the influential points indicated by the study. Changes can be seen in the percentages of classes in areas of productivity due to the influential points.

Table 2. Percentage of productive areas for the cases under study.

Productivity $[t\,\mathrm{ha}^{-1}]$	With all the data (%)	Without observations #13 and #24 (%)
$[2.90 - 3.03]$	6.59	5.67
$[3.04 - 3.17]$	42.00	40.52
$[3.18 - 3.31]$	32.26	32.47
$[3.32 - 3.45]$	14.90	16.47
$[3.46 - 3.60]$	4.25	4.86
Total	100	100

5. Conclusions

In this paper, we have proposed a study of atypical observations by using the technique of local influence considering a perturbation in the matrix of explanatory variables, which is useful to detect changes in the maximum likelihood estimator. It is well known influential observations can change the estimates of the parameters and, consequently, these may change the conclusions of the statistical analysis. Then, diagnostic techniques should be employed to identify departures from the postulated model and influential observations. We have discussed an application using real data set, where we perturb one explanatory variable. The study have shown that the presence of influential observations in the data set can cause distortions in the construction of thematic maps, changing the structure of the spatial dependence. The application of influence diagnostic techniques should be part of all geostatistical analysis, ensuring that the information contained in thematic maps has higher quality and reliability.

Appendix II: Data Set Used in the Application

Table 3. coordinate (X,Y) of the contour.

Point	X	Y	Point	X	Y	Point	X	Y
1	240606.55	7237714.28	67	240484.59	7237165.74	133	239670.32	7237585.62
2	240653.33	7237654.36	68	240469.36	7237162.14	134	239678.03	7237600.62
3	240665.83	7237638.98	69	240445.98	7237156.60	135	239687.27	7237614.68
4	240677.56	7237623.01	70	240430.42	7237153.11	136	239703.03	7237633.59
5	240694.14	7237597.85	71	240414.91	7237149.41	137	239714.88	7237645.17
6	240708.25	7237582.47	72	240391.66	7237142.95	138	239728.02	7237654.69
7	240721.23	7237566.07	73	240376.14	7237138.18	139	239749.36	7237665.54
8	240739.09	7237540.02	74	240352.04	7237131.63	140	239763.93	7237671.26
9	240752.09	7237523.54	75	240335.83	7237126.73	141	239778.87	7237675.09
10	240765.76	7237507.46	76	240319.94	7237121.04	142	239801.47	7237677.85
11	240785.25	7237481.68	77	240296.79	7237112.51	143	239816.06	7237678.65
12	240798.25	7237464.28	78	240281.73	7237107.09	144	239830.32	7237679.32
13	240812.43	7237447.54	79	240266.73	7237101.78	145	239851.34	7237677.80
14	240833.76	7237422.21	80	240244.54	7237092.34	146	239865.11	7237674.36
15	240847.30	7237404.84	81	240229.20	7237086.27	147	239885.57	7237667.99
16	240860.56	7237387.25	82	240213.79	7237080.73	148	239899.29	7237663.19
17	240880.59	7237361.26	83	240190.57	7237072.50	149	239911.87	7237657.46
18	240893.62	7237343.72	84	240175.06	7237068.27	150	239930.06	7237648.19
19	240914.19	7237318.17	85	240152.32	7237061.64	151	239941.78	7237642.41
20	240927.65	7237301.19	86	240137.28	7237057.38	152	239953.18	7237637.28
21	240940.28	7237283.98	87	240122.41	7237053.64	153	239968.76	7237630.52
22	240951.56	7237266.08	88	240100.41	7237048.49	154	239977.56	7237626.55
23	240962.65	7237248.44	89	240086.48	7237045.08	155	239985.06	7237623.20
24	240980.93	7237224.80	90	240072.28	7237041.63	156	239991.24	7237620.23
25	240998.54	7237203.44	91	240049.64	7237042.16	157	240016.13	7237601.54
26	241013.15	7237184.32	92	240034.97	7237046.07	158	240027.40	7237595.70
27	241025.45	7237168.54	93	240020.64	7237051.57	159	240038.25	7237590.92
28	241033.65	7237159.59	94	239999.95	7237061.73	160	240050.60	7237585.71
29	241042.31	7237148.40	95	239987.04	7237069.74	161	240070.83	7237575.90
30	241046.11	7237143.63	96	239975.89	7237079.80	162	240085.79	7237569.86
31	241042.72	7237139.17	97	239961.37	7237096.48	163	240101.53	7237565.39
32	241038.67	7237138.07	98	239952.61	7237108.46	164	240124.84	7237559.99
33	241029.88	7237138.05	99	239941.61	7237127.64	165	240140.37	7237557.43
34	241016.46	7237138.98	100	239935.18	7237140.04	166	240163.86	7237555.25
35	240991.80	7237140.67	101	239927.00	7237151.19	167	240179.89	7237555.20
36	240964.35	7237142.46	102	239914.10	7237167.53	168	240195.93	7237555.72
37	240945.99	7237143.62	103	239905.81	7237177.93	169	240219.46	7237557.99
38	240920.76	7237143.70	104	239898.05	7237188.48	170	240234.76	7237560.85
39	240906.25	7237143.12	105	239890.93	7237198.78	171	240249.72	7237563.93
40	240893.82	7237142.23	106	239883.75	7237208.58	172	240272.62	7237568.52
41	240882.64	7237141.08	107	239873.22	7237224.27	173	240287.61	7237573.56
42	240871.66	7237140.37	108	239859.21	7237238.39	174	240309.61	7237582.60
43	240863.09	7237139.92	109	239843.70	7237253.01	175	240324.56	7237589.05
44	240853.34	7237138.54	110	239833.12	7237263.13	176	240347.22	7237599.82
45	240848.69	7237137.49	111	239817.08	7237277.59	177	240361.59	7237607.91
46	240838.31	7237136.75	112	239806.13	7237285.73	178	240381.26	7237622.13
47	240827.42	7237139.84	113	239794.89	7237291.83	179	240393.78	7237631.70
48	240814.09	7237143.67	114	239780.09	7237298.79	180	240413.31	7237643.81
49	240803.00	7237148.11	115	239744.53	7237312.76	181	240425.13	7237651.83
50	240790.39	7237153.76	116	239731.85	7237316.46	182	240437.48	7237660.27
51	240769.01	7237162.72	117	239712.8	7237324.10	183	240457.14	7237671.94
52	240754.20	7237167.99	118	239700.47	7237330.19	184	240470.53	7237679.01
53	240739.14	7237172.44	119	239681.50	7237340.73	185	240484.16	7237685.43
54	240716.69	7237178.63	120	239669.30	7237349.57	186	240505.56	7237693.10
55	240701.62	7237181.79	121	239658.55	7237360.52	187	240519.16	7237697.59
56	240686.51	7237184.78	122	239644.83	7237378.37	188	240539.79	7237702.37
57	240663.75	7237187.35	123	239637.07	7237390.81	189	240552.52	7237704.44
58	240648.60	7237187.64	124	239631.37	7237404.79	190	240565.22	7237706.72
59	240625.99	7237187.72	125	239629.77	7237428.06	191	240583.04	7237710.31
60	240611.13	7237187.49	126	239632.34	7237444.56	192	240593.51	7237712.84
61	240589.03	7237185.63	127	239635.19	7237461.50	193	240601.26	7237714.38
62	240574.50	7237183.29	128	239640.22	7237487.24	194	240606.55	7237714.28
63	240559.84	7237180.90	129	239643.42	7237504.24	195	240606.55	7237714.28
64	240537.62	7237176.57	130	239646.99	7237521.31	196	240606.55	7237714.28
65	240522.88	7237173.53	131	239653.97	7237546.03			
66	240507.91	7237170.42	132	239659.68	7237562.17			

Table 4. X,Y: coordinate (UTM); z: soybean yield ($t\,\text{ha}^{-1}$); x_2: soil resistance to penetration through layer 0-0.1m; x_3: soil resistance to penetration through layer 0.1-0.2m; x_4: soil resistance to penetration through layer 0.2-0.3m.

Obs	X	Y	z	x_2	x_3	x_4
1	240674	7237587	2.84	2.697	2.359	1.750
2	240811	7237409	2.86	3.692	3.709	2.123
3	240948	7237231	3.64	3.124	2.053	2.172
4	240569	7237601	3.15	3.511	2.776	1.969
5	240463	7237615	3.59	2.920	2.564	2.524
6	240509	7237555	2.83	2.680	2.327	1.911
7	240600	7237437	3.54	1.398	1.954	2.738
8	240783	7237199	3.00	2.016	1.356	2.547
9	240678	7237213	2.79	3.455	3.207	2.109
10	240541	7237391	4.00	2.367	2.405	2.296
11	240449	7237510	3.26	2.560	2.178	2.102
12	240404	7237569	3.14	3.259	3.682	2.536
13	240344	7237524	2.09	3.651	2.452	2.197
14	240435	7237405	3.90	2.170	2.782	2.096
15	240481	7237346	3.33	3.672	2.980	2.049
16	240572	7237227	3.63	3.423	2.477	2.099
17	240513	7237182	3.19	2.416	2.824	2.406
18	240467	7237241	3.63	2.519	2.755	2.211
19	240421	7237300	3.44	1.808	2.310	2.463
20	240376	7237360	3.02	2.700	2.366	1.954
21	240284	7237478	2.90	3.508	3.274	2.403
22	240179	7237492	3.11	2.102	2.421	1.630
23	240225	7237433	3.02	2.389	2.331	1.986
24	240316	7237314	4.09	2.556	2.735	2.224
25	240362	7237255	3.46	2.469	3.193	2.736
26	240348	7237150	3.47	3.590	2.212	1.478
27	240211	7237328	3.30	2.392	2.360	1.735
28	240074	7237506	3.19	2.237	1.915	1.826
29	239877	7237639	3.29	2.640	2.196	1.574
30	239923	7237579	2.87	2.750	1.919	4.958
31	240014	7237461	3.15	2.992	2.068	1.728
32	240151	7237283	3.49	2.418	2.262	1.897
33	240242	7237164	3.65	3.160	2.048	2.011
34	240288	7237105	3.28	2.813	2.707	1.580
35	240137	7237178	3.55	2.388	1.973	1.677
36	239909	7237475	3.07	2.369	1.988	1.879
37	239863	7237534	2.64	1.929	1.705	1.481
38	239817	7237593	2.94	2.537	2.286	1.762
39	239757	7237548	2.80	3.004	2.593	2.141
40	239803	7237489	3.10	3.330	2.152	1.452
41	239895	7237370	3.08	2.763	2.471	2.289
42	239940	7237311	2.73	2.329	2.132	1.683
43	240077	7237132	3.65	2.544	2.897	2.095
44	240123	7237073	3.49	3.245	3.085	2.663
45	239881	7237265	3.38	2.913	1.861	1.337
46	239835	7237324	3.02	2.541	2.140	1.639
47	239729	7237338	3.04	2.627	2.202	1.637

Acknowledgements

The authors would like to thank the Executive and Guest Editors and two referees for their very helpful comments that led to a much improved presentation. The authors would also like to acknowledge the partial financial support to Projects Fondecyt 1110318, Chile, and CNPq, Capes and Fundação Araucária of Paraná State, Brazil.

References

Beckman, R.J., Nachtsheim, C.J., Cook, R.D., 1987. Diagnostics for mixed-model analysis of variance. Technometrics, 29, 413–426.

Borssoi, J.A., Uribe-Opazo, M.A., Galea, M., 2009. Diagnostics tecniques applied in geostatistics for agricultural data analysis. Revista Brasileira de Ciência do Solo, 6, 1561–1570.

Chen, F., Zhu, H.T., 2009. Perturbation selection and local influence analysis for nonlinear structural equation model. Psychometrika, 74, 493–516.

Cook, R.D., 1986. Assessment of Local Influence. Journal of The Royal Statistical Society Series B - Statistical Methodology, 48, 133–169.

Faraco, M.A., Uribe-Opazo, M.A., Silva, E.A., Johann, J.A., Borssoi, J.A., 2008. 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, 32, 463–476.

Galea, M., Paula, G.A., Bolfarine, H., 1997. Local influence in elliptical linear regression models. The Statistician, 46, 71–79.

Galea, M., Paula, G.A., Uribe-Opazo, M.A., 2003. On influence diagnostic in univariate elliptical linear regression models. Statistical Papers, 44, 23–45.

Liu, S.Z., 2000. On local influence for elliptical linear models. Statistical Papers, 41, 211–224.

Mardia, K.V., Marshall, R.J., 1984. Maximum likelihood models for residual covariance in special regression. Biometrika, 71, 319–332.

Nel, D.G., 1980. On matrix differentiation in statistics. South African Statistical Journal, 14, 137–193.

Pan, J.X., Fang, K.T., von Rosen, 1997. Local influence assessment in the growth curve model with unstructured covariance. Journal of Statistical Planning and Inference, 62, 263–278.

Paula, G.A., 1993. Assessing local influence in restricted regression models. Computational Statistics and Data Analysis, 16, 63–79.

R Development Core Team, 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ribeiro Jr., P.J, Diggle, P.J., 2001. GeoR: A package for geoestatistical analysis. R Journal, 01. `http://cran.r-project.org/doc/Rnews.2001`

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Thomas, W., Cook, R.D., 1990. Assessing influence on predictions from generalized linear models. Technometrics, 32, 59–65.

Zhu, H.T., Lee, S.Y., 2001. Local influence for incomplete-data models. Journal of The Royal Statistical Society Series B - Statistical Methodology, 63, 111-126.

Zhu, H., Ibrahim, J.G., Lee, S., Zhang, H., 2007. Perturbation selection and influence measures in local influence analysis. The Annals of Statistics, 35, 2565–2588.