Inference and diagnostics in spatial linear models: an application to wheat productivity

Fernanda De Bastiani
Universidade Federal de Pernambuco, Avenida Professor Moraes Rego 1235, Recife 50740-540, Pernambuco, Brazil - debastiani@de.ufpe.br
Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Macul, Santiago 782-0436, Chile - fedebastiani@mat.puc.cl

DOI: 10.1481/icasVII.2016.g40b

Abstract

The wheat production has been grown along the years and is one of the most important food grain source for humans. We analyze the productivity of two varieties of wheat planted in a regular sampling grid in an experimental area in south region of Brazil. We considered as explanatory variables the variety of wheat, spike length, average plant height and average number of tillers in 60 days. To model the mean of the wheat productivity we fitted a Gaussian spatial linear model, with different geostatisticals models for the variance-covariance matrix. To asses the influence of some observations we considered local diagnostics techniques based on Cook's approach. We considered appropriate perturbation scheme in the response variable. Then, we have substantial information to select the final model.

Keywords: Geostatistical; Maximum Likelihood; Spatial Variability; Precision Agriculture.

1 Introduction

Wheat (*t*riticum spp.) originated in southwestern Asia. According to Curis (2002), wheat was one of the first domesticated food crops and for 8000 years has been the basic staple food of the major civilizations of Europe, West Asia and North Africa, whilst is grown on more land area than any other commercial crop and continues to be the most important food grain source for humans. World wheat production increased dramatically during the period 1951-1990, although the expansion of the area sown to wheat has long ceased to be a major source of increased wheat output (CIMMYT, 1996). In Brazil the production is concentrated in south region.

The wheat grain is used to make flour for bread, pasta, pastry, etc. Wheat is also a popular source of animal feed, particularly in years where harvests are adversely affected by rain and significant quantities of the grain are made unsuitable for food use (Curis, 2002). It is considered a good source of protein, minerals, B-group vitamins and dietary fiber (Shewry, 2007). Davy et al (2002) has shown that whole wheat, rather than refined wheat, is a good choice for obese patients. Sidorova et al (2012) performed a geostatistical analysis of the spatial variability of the soil properties, the sowing parameters, and the wheat yield in a field experiment under precision agriculture conditions.

We analyze the wheat productivity data and four explanatory variables from an agricultural area in south Brazil. To consider the dependence between observations, the analysis were conducted using geostatistics techniques, where the data are collected at known sites in space, from a process that has a value at every site in a certain domain. To model the mean of the wheat productivity we fitted a Gaussian spatial linear model (GSLM) by maximum likelihood (ML) method given in Section 2. To asses the influence of some observations we considered local diagnostics techniques based on Cook's approach. We considered appropriate perturbation scheme in the response variable as shown in Section 2. The results are presents in Section 3 and some conclusions are given in Section 4.

2 Methodology

Let $\mathbf{Y} = \mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}_1), \ldots, Y_n(\mathbf{s}_n))^\top$ be an $n \times 1$ random vector of an isotropic and stationary stochastic process, that belong to the family of Gaussian distributions and depend on the sites $\mathbf{s}_j \in S \subset \mathbb{R}^2$, for $j = 1, \ldots, n$, $\mathbf{s} = (\mathbf{s}_1, \ldots, \mathbf{s}_n)^\top$. This stochastic process can be written in matrix form by

$$\mathbf{Y}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}).$$

where, the deterministic term $\boldsymbol{\mu}(\mathbf{s})$ is an $n \times 1$ vector, the means of the process $\mathbf{Y}(\mathbf{s})$, $\boldsymbol{\epsilon}(\mathbf{s})$ is an $n \times 1$ vector of a stationary process with zero mean vector, $\mathrm{E}[\boldsymbol{\epsilon}(\mathbf{s})] = \mathbf{0}$, and $n \times n$ covariance matrix $\boldsymbol{\Sigma} = [C(s_u, s_v)]$. The mean vector $\boldsymbol{\mu}(\mathbf{s})$ can be written as a spatial linear model by $\boldsymbol{\mu}(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta}$, where, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a $p \times 1$ vector of unknown parameters, $\mathbf{X} = \mathbf{X}(\mathbf{s}) = [\mathbf{x}_{j1}(\mathbf{s}) \ldots \mathbf{x}_{jp}(\mathbf{s})]$ is an $n \times p$ matrix of p explanatory variables, for $j = 1, \ldots, n$.
The matrix $\boldsymbol{\Sigma}$ is symmetric and positive defined, where the elements $C(\mathbf{s}_u, \mathbf{s}_v)$ depend on the Euclidean distance $d_{uv} = ||\mathbf{s}_u - \mathbf{s}_v||$ between points $\mathbf{s}_u$ and $\mathbf{s}_v$, sometimes $C(\mathbf{s}_u, \mathbf{s}_v)$ is also denoted by $C(d_{uv})$ or $C(d)$. The covariance matrix structure which depends on parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q)^\top$ as given in Equation (1) (Uribe-Opazo et al, 2012):

$$\boldsymbol{\Sigma} = \phi_1 \mathbf{I}_n + \phi_2 \mathbf{R}, \tag{1}$$

where, $\phi_1 \geq 0$ is the parameter known as nugget effect; $\phi_2 \geq 0$ is known for sill ; $\mathbf{R} = \mathbf{R}(\phi_3, \phi_4) = [(r_{uv})]$ or $\mathbf{R} = \mathbf{R}(\phi_3) = [(r_{uv})]$ is an $n \times n$ symmetric matrix, which is function of $\phi_3 > 0$, and sometimes also function of $\phi_4 > 0$, with diagonal elements $r_{uu} = 1, (u = 1, \ldots, n)$; $r_{uv} = \phi_2^{-1} C(\mathbf{s}_u, \mathbf{s}_v)$ for $\phi_2 \neq 0$, and $r_{uv} = 0$ for $\phi_2 = 0$, $u \neq v = 1, \ldots, n$, where r_{uv} depends on d_{uv}; ϕ_3 is a function of the model range, ϕ_4 when exists is known as the smoothness parameter, and $\mathbf{I}_n$ is an $n \times n$ identity matrix. The Matérn (Matern, 1960) is a covariance function particularly attractive. Table 1 presents few special cases of the Matérn class of models.

Table 1: Special cases of the Matérn covariance function.

smooth parameter	covariance function	model
$\phi_4 = 1/2$	$C(d_{uv}) = \phi_2 \exp(-d_{uv}/\phi_3)$	exponential
$\phi_4 = 1$	$C(d_{uv}) = \phi_2 (d_{uv}/\phi_3) K_{\phi_4}(d_{uv}/\phi_3)$	Whittle
$\phi_4 \to \infty$	$C(d_{uv}) = \phi_2 \exp(-(d_{uv}/\phi_3)^2)$	Gaussian

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$ be the vector of unknown parameters. The log-likelihood and score functions for the GLSM are given by

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{2}$$

$$\begin{aligned}
\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}, \\
\mathbf{U}(\boldsymbol{\phi}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} &= -\frac{1}{2}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \phi} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}) + \frac{1}{2}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top \boldsymbol{\Sigma}^{-1}),
\end{aligned}$$

where $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$. From the solution of the score function of $\boldsymbol{\beta}$, $\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$, the maximum likelihood estimator $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}\mathbf{y}$. The derivatives of first and second-order of the scale matrix $\boldsymbol{\Sigma}$, with respect to ϕ_1, ϕ_2 and, ϕ_3, for some covariance functions are presented in Uribe-Opazo et al (2012), however the score equation for $\boldsymbol{\phi}$ do not lead to a closed-form solution for $\hat{\boldsymbol{\phi}}$. We consider the parameter ϕ_4 as fixed. The criteria considered to choose the geostatistical model for the covariance matrix were the cross-validation (CV), trace of the asymptotic covariance matrix of an estimated mean (Tr) and the log-likelihood maximum value (LMV) (De Bastiani et al, 2015).
Asymptotic standard errors can be calculated by inverting either observed information matrix, $I(\boldsymbol{\theta})$ or the expected information matrix, $\mathbf{F}(\boldsymbol{\theta})$, where $I(\boldsymbol{\theta})$ is $I(\boldsymbol{\theta}) = -\mathbf{L}(\boldsymbol{\theta})$, evaluated in $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, with $\mathbf{L}(\boldsymbol{\theta}) = \partial^2 \mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ and $\mathbf{F}(\boldsymbol{\theta})$ is given by (see Waller & Gotway, 2004)

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbf{F} = \begin{pmatrix} \mathbf{F}_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{\phi\phi} \end{pmatrix},$$

where $\mathbf{F}_{\beta\beta} = \mathbf{X}^\top \boldsymbol{\Sigma}^{-1}\mathbf{X}$, and $\mathbf{F}_{\phi\phi} = \frac{1}{2}\frac{\partial \operatorname{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}^\top}$.

2.1 Local Influence

One of the purposes of diagnostic techniques is to evaluate the stability of the fitted model in a data set and should be part of all statistical analysis, since influential observations may distort the values of the statistic interest and lead us to misleading results.
In the local influence method, introduced by Cook (1986), a perturbation scheme is introduced into the postulated model through a perturbation vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)^\top$ $(\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^k)$, generating the perturbed model, where $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})$ is the corresponding log-likelihood function. The influence measure is constructed using the basic geometric idea of curvature of the likelihood displacement given by

$$LD(\boldsymbol{\omega}) = 2[\mathcal{L}(\hat{\boldsymbol{\theta}}) - \mathcal{L}(\hat{\boldsymbol{\theta}}_\omega)],$$

where $\hat{\boldsymbol{\theta}}$ is the ML estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$ in the postulated model, with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_q)^\top$ and $\hat{\boldsymbol{\theta}}_\omega$ is the ML estimator of $\boldsymbol{\theta}$ in the perturbed model.
The plot of the elements $|l_{max}|$ versus index (order of data) can reveal what type of perturbation has more influence on $LD(\boldsymbol{\omega})$, in the neighbourhood of $\boldsymbol{\omega}_0$, Cook (1986). Poon & Poon (1999) proposed the conformal normal curvature $B_l = C_l/\operatorname{tr}(2\mathbf{J})$, where $\mathbf{J} = \Delta^\top \mathbf{L}^{-1}\Delta$. The conformal curvature in the unit direction with j-th entry 1 and all other entries 0 is given by $B_i = 2|j_{ii}|/\operatorname{tr}(2\mathbf{J})$. The plot of B_i versus index can reveal potential influential observations.
To verify if a perturbation scheme is appropriate, Zhu et al (2007) proposed to use the Fisher information matrix of $\boldsymbol{\omega}$ in the perturbed model considering the vector $\boldsymbol{\theta}$ as fixed. In the following Section we give the results for the response variable perturbation scheme.

2.1.1 Perturbation on the response variable

Let consider as perturbation scheme the model shift in mean, i.e. $\mathbf{Y} = \boldsymbol{\mu}(\boldsymbol{\omega}) + \epsilon$, with $\boldsymbol{\mu}(\boldsymbol{\omega}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{\omega}$ where $\mathbf{A}$, $n \times n$, is a matrix that does not depend on $\boldsymbol{\beta}$ or on $\boldsymbol{\omega}$. In this case $\boldsymbol{\omega}_0 = \mathbf{0}$. Equivalently we can write $\mathbf{Y}_\omega = \mathbf{X}\boldsymbol{\beta} + \epsilon$, with $\mathbf{Y}_\omega = \mathbf{Y} + (-1)\mathbf{A}\boldsymbol{\omega}$, that corresponds to a perturbation scheme of the response vector.
The perturbed log-likelihood is given by

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}[\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\omega})]^\top \boldsymbol{\Sigma}^{-1}[\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\omega})].$$

To select an adequate matrix $\mathbf{A}$ we can use the methodology proposed by Zhu et al (2007). In effect, the score function for $\boldsymbol{\omega}$ in the perturbed log-likelihood function (2.1.1) is given by

$$\mathbf{U}(\boldsymbol{\omega}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = \mathbf{A}^\top \boldsymbol{\Sigma}^{-1}[\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\omega})].$$

Let $\mathbf{G}(\boldsymbol{\omega}) = E_\omega[\mathbf{U}(\boldsymbol{\omega})\mathbf{U}^\top(\boldsymbol{\omega})] = \text{diag}[\mathbf{g}_{11}(\boldsymbol{\omega}_1), \ldots, \mathbf{g}_{nn}(\boldsymbol{\omega}_n)]$ be the Fisher information matrix with respect to the perturbation vector $\boldsymbol{\omega}$. A perturbation $\boldsymbol{\omega}$ is appropriate if it satisfies $\mathbf{g}_{jj}(\boldsymbol{\omega}_0) = c\mathbf{I}_n$, where $c > 0$. In our case, we have $\mathbf{g}_{jj}(\boldsymbol{\omega}_0) = c\mathbf{A}^\top \boldsymbol{\Sigma}^{-1}\mathbf{A}$, with $c = 1$. Notice that usually $\mathbf{A}^\top \boldsymbol{\Sigma}^{-1}\mathbf{A} \neq \mathbf{I}_n$. However if $\mathbf{A} = \boldsymbol{\Sigma}^{1/2}$, then $\mathbf{g}_{jj}(\boldsymbol{\omega}_0) = c\mathbf{I}_n$ and so $\boldsymbol{\mu}(\boldsymbol{\omega}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\omega}$ is a perturbation scheme appropriate, as shown in De Bastiani et al (2015).
Considering the appropriated perturbation scheme for the response variable, where Δ_β is an $p \times n$ matrix and Δ_ϕ is an $3 \times n$ matrix given by

$$\Delta_\beta = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} = -\mathbf{X}^\top \hat{\boldsymbol{\Sigma}}^{-1/2} \quad \text{and}$$

$$\Delta_\phi = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\omega}^\top} = -\frac{\partial \,\text{vec}^\top(\boldsymbol{\Sigma})}{\partial \boldsymbol{\phi}} \,\text{vec}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1/2}) \,\text{vec}(\boldsymbol{\epsilon} \otimes \mathbf{1}^\top).$$

evaluated in $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, where $\hat{\epsilon} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ and $\mathbf{1}$ is an $n \times 1$ vector of ones.

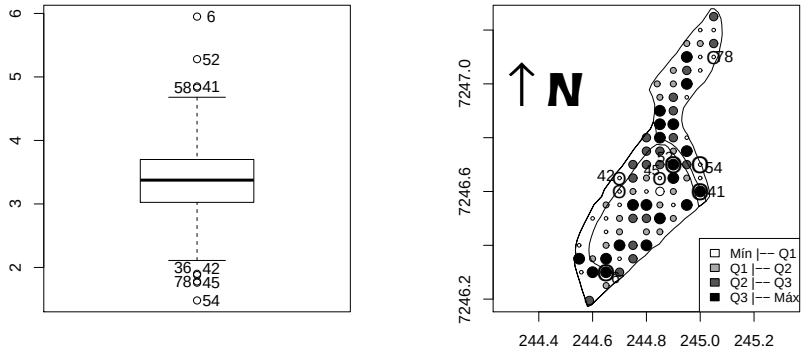
3 Results

The data were collected in 2003 in Cascavel city, south region of Brazil in an area of 22.63 hectares. The climate, according to Kppen is Cfa, temperate mesothermal and super humid. We analyze 84 element samples of two varieties of wheat CD101 and CD103, corresponding to 4.46 ha and 18.17 ha, respectively, collected in a regular grid of $50 \times 50m$. The explanatory variables are: average plant height - `alt60` and average number of tillers - `perfilho60` in 60 days, spike length - `cespigas` and the wheat variety treated as a `dummy` variable (0 or 1). So, $\mathbf{Y}$ represents a vector 84×1.
Table 2 presents a descriptive analysis of the response variable `wheat`, wheat productivity, and the explanatory variables. The wheat productivity mean is $3.372\,t\ ha^{-1}$. The average plant height in 60 days varies from $13.40\,cm$ to $36.60\,cm$. The average number of tillers presents the greatest value for the variance coefficient, however it still can be considered homogeneous. The mean and median of the spike length are the same considering one decimal.

Table 2: Descriptive analysis of response and explanatory variables.

Variable	Min.	1st Quartil	Median	Mean	3rd Quartil	Max.	var. coef.
`wheat`	1.480	3.037	3.375	3.372	3.680	5.950	0.23
`alt60`	13.40	20.65	22.50	23.17	24.62	36.60	0.17
`perfilho60`	0.40	1.200	1.700	1.661	2.100	3.40	0.38
`cespigas`	5.00	6.10	6.45	6.47	6.80	7.90	0.09

Figure 1(a) presents the boxplot for wheat productivity where the observations 06, 36, 41, 42, 45, 52, 54, 58 and 78 are outliers with wheat productivity values of 5.95, 1.90, 4.85, 1.88, 1.76, 5.28, 1.48, 4.83 and $1.78\ t\,ha^{-1}$, respectively. The site of these observartions are highlighted in Figure 1(b).



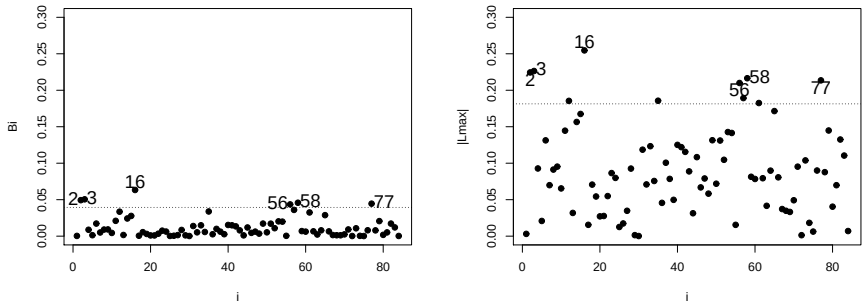
(a) (b)

Figure 1: (a) Boxplot and (b) postplot for wheat productivity samples in a regular grid, $50 \times 50\, m$.

We considered the Matérn family class to model the covariance matrix function. We considered values for ϕ_4 from 0.3 to ∞. According to the criteria LMV, Tr and CV, the choosen covariance matrix function is the value of $\phi_4 \to \infty$, which corresponds to the Gaussian covariance function. According to the likelihood ratio test, all the explanatory variables are significant at a level of 5%. The final choosen model (in parenthesis is given the corresponding asymptotic standard errors) is given by

$$\begin{array}{lccccc} \hat{\boldsymbol{\mu}}(\mathbf{s}_i) = & 0.122+ & 0.354\texttt{dummy}(\mathbf{s}_i)+ & 0.069\texttt{alt60}(\mathbf{s}_i)+ & 0.078\texttt{perfilho60}(\mathbf{s}_i)+ & 0.188\texttt{cespigas}(\mathbf{s}_i) \\ & (1.257) & (0.266) & (0.024) & (0.142) & (0.146), \end{array}$$

with spatial parameters estimates given by $\hat{\phi}_1 = 0.000(0.4058)$, $\hat{\phi}_2 = 0.548(0.4281)$ and, $\hat{\phi}_3 = 0.0349(0.0001)$. Figure 2 presents B_i versus index and $|L_{max}|$ versus index plots where observation #16 is detected as the most potential influent. Non of the observations pointed out where identified in the boxplot given in Figure 1(a).



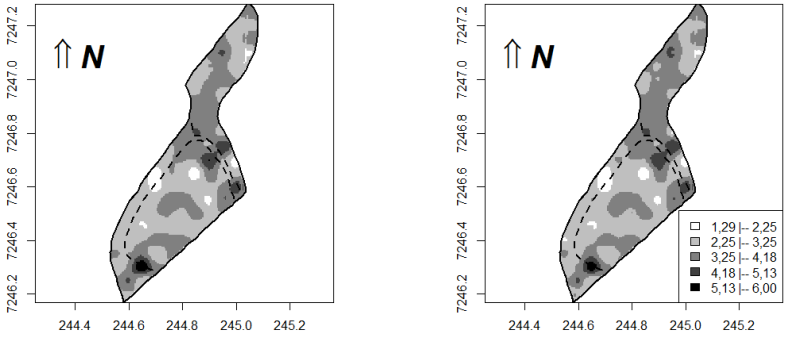
(a) B_i versus index (b) $|L_{max}|$ versus index

Figure 2: Local influence plots, (a) B_i versus index and (b) $|L_{max}|$ versus index considering the appropriate perturbation scheme.

We analized the data set without observation #16, and the chosen model for the covariance function remain the Gaussian one. The model is given by

$$\begin{array}{lccccc} \hat{\boldsymbol{\mu}}(\mathbf{s}_i) = & 0.255+ & 0.307\texttt{dummy}(\mathbf{s}_i)+ & 0.070\texttt{alt60}(\mathbf{s}_i)+ & 0.083\texttt{perfilho60}(\mathbf{s}_i)+ & 0.175\texttt{cespigas}(\mathbf{s}_i) \\ & (1.268) & (0.276) & (0.024) & (0.151) & (0.146), \end{array}$$

with spatial parameters estimates given by $\hat{\phi}_1 = 0.000(0.2323)$, $\hat{\phi}_2 = 0.560(0.278)$ and, $\hat{\phi}_3 = 0.040(0.0002)$. We can note a decrease on the estimate of the asymptotic standard error for $\hat{\phi}_2$.
Figure 3 shows the maps with all observations and without observation #16 where we can note a slightly difference between the maps in the north area. We have information to select the model considering all the observations as the final model.



(a) with all observations (b) without observation #16

Figure 3: Mean wheat productivity maps considering the model (a) with all observations (b) without observation #16.

Figure 3 shows the maps with all observations and the scenarios mentioned above. The maps constructed by kriging with external drift present well defined zones. Note: there is a slight difference between the maps in the northern area. A difference between the varieties CD101 and CD103 was also noted.

4 Conclusions

We proposed a measure based on the likelihood displacement to assess the stability of the likelihood function, using the perturbation on the response variable. We applied the methodology to a data set of wheat productivity collected in south region of Brazil. The spatial linear models enabled us to verify the spatial dependence between the wheat productivity data in the study area, according to the two varieties and plant attributes.
The maps constructed allowed us to estimate the wheat productivity in the study area, allowing us to create management zones with low or high productivity with the purpose of unifying similar areas, apply localized inputs and then maximize the the profit reducing the environment impact. It was observed the deletion of potential influential observations according to Zhu, caused changes in the parameters estimates that define the spatial dependence structure. Then, we have substantial information to select the final model considering all the observations.

5 Acknowledgements

Thank you the partial financial support from Fundação Araucária of Paraná State, FACEPE, Capes, CNPq, Brazil and Pontificia Universidad Catolica de Chile. Thank you the team involved, Professor M. A. Uribe-Opazo, Professor M. Galea and Dr. D. Grzegozewski.

References

Cook, R., 1986. Assessment of local influence. Journal of the Royal Statistical Society, Serie B 48, 133–169.

CIMMYT. 1996. World Wheat Facts and Trends 1995/96: Understanding Global Trends in the Use of Wheat Diversity and International Flows of Wheat Genetic Resources. Mexico, D.F.: CIMMYT.

Curis, B. C. 2002. Wheat in the world. In: Bread wheat: Improvement and Production. FAO Plant Production and Protection Series, 30.

Davy, B. M., Davy, K. P., Ho, R. C. Beske, S. D., Davrath, L. R. & Melby, C. L., 2002. High-fiber oat cereal compared with wheat cereal consumption favorably alters LDL-cholesterol subclass and particle numbers in middle-aged and older men. The American Journal of Clinical Nutrition, 76(2), 351–358.

De Bastiani, F., Mariz deAquino Cysneiros, A. H., Uribe-Opazo, M. A. & Galea, M., 2015. Influence diagnostics in elliptical spatial linear models. TEST 24 (2), 322–340.

Matérn, B., 1960. Spatial variation. Vol. 49. Meddelanden fren Statens Skogséforskningsinstitut.

Poon, W. & Poon, Y. S., 1999. Conformal normal curvature and assessment of local influence. Journal of the Royal Statistical Society, B 61, 51–61.

Shewry, P. R., 2007. Improving the protein content and composition of cereal grain. Journal of Cereal Science, 46, 239–250.

Sidorova, V. A.; Zhukovskiib, E. E.; Lekomtsevb, P. V. & Yakushev, V. V., 2012. Geostatistical Analysis of the Soil and Crop Parameters in a Field Experiment on Precision Agriculture, 45, 8, 783–792.

Uribe-Opazo, M., Borssoi, J. & Galea, M., 2012. Influence diagnostics in gaussian spatial linear models. Journal of Applied Statistics 39 (3), 615–630.

Waller, L. & Gotway, C., 2004. Applied Spatial Statistics for Public Health Data. Wiley-Interscience, Hoboken NJ.

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