GAUSSIAN SPATIAL LINEAR MODEL OF SOYBEAN YIELD USING BOOTSTRAP METHODS

Gustavo H. Dalposso1*, Miguel A. Uribe-Opazo2, Jerry A. Johann2, Manuel Galea3, Fernanda De Bastiani4

1*Corresponding author. Universidade Tecnológica Federal do Paraná/ Toledo - PR, Brasil. E-mail: gustavodalposso@utfpr.edu.br

KEYWORDS
geostatistics, spatial resampling, spatial variability.

ABSTRACT
This study aims to quantify the uncertainties associated to the parameters of a Gaussian spatial linear model (GSLM) and the assumption of normality residuals in the modeling of the spatial dependence of the soybean yield as a function of soil chemical attributes. The spatial bootstrap methods were used to determine the point and interval estimators associated with the model parameters. Hypothesis tests were carried out on the significance of the model parameters and the quantile-quantile probability plot was elaborated to verify the data normality. The uncertainties associated to the parameters of the spatial dependence structure were quantified and the potassium content, phosphorus content and soil pH covariates were significant to explain the soybean yield mean. These covariates were used in the elaboration of a new model, which provided the elaboration of a contour map of soybean yield. Analysis of the quantile-quantile plot indicated that soybean yield data follow a normal probability distribution.

INTRODUCTION
Soybean is the most important legume species in the world (Gallon et al., 2016) and it stands out in the Brazilian agribusiness for its high productive potential in different regions (Cruz et al., 2016). Soybean is widely used for the elaboration of animal feed and oil and as Ávila & Albrecht (2010) explained, it has been emphasized as an alternative in the prevention of diseases and in human food, being able to be transformed into several protein foods.

Among the studies involving soybean yield, we highlight the geostatistics methods used to detect the spatial variability in the crops (Borssoi et al., 2011; Kestring et al., 2015; Guedes et al., 2016). Although geostatistics allows the understanding of spatial variability of soybean yield, the samples used in the analyzes are generally few and sparse (Pardo-Igúzquiza & Olea, 2012), then there are uncertainties associated with the results obtained.

The uncertainties occur because the spatial linear model is estimated by parametric methods, such as those presented by Mardia & Marshall (1984). To quantify the uncertainties associated with the results of a geostatistics analysis, an alternative to traditional inference methods is the use of spatial bootstrap resampling (SB) (Solow, 1985) and parametric spatial bootstrap (PSB) methods (Tang et al., 2006), adaptations of the bootstrap method (Efron, 1979) for spatially dependent data.

The bootstrap method is well known and has been used in studies involving independent samples of soybean yield (Dalposso et al., 2016; Gupta & Manjaya, 2016). The bootstrap methods for spatially dependent data have been highlighted in the literature, due to the importance of uncertainty modeling in the analyzes, as can be observed in the works of Kang et al. (2008), Schelin & Sjöstedt-De Luna (2010), Olea & Pardo-Igúzquiza (2011) and Pardo-Igúzquiza & Olea (2012).

The aim of this study was to use spatial bootstrap (SB) and parametric spatial bootstrap (PSB) methods to quantify the uncertainties associated with spatial dependence structure parameters and the assumption of residuals normality distributed in spatial dependence modeling of soybean yield with chemical attributes of the soil considered as covariates.

2 Universidade Estadual do Oeste do Paraná/ Cascavel - PR, Brasil.
3 Pontifícia Universidad Católica de Chile/ Santiago, Chile.
4 Universidade Federal de Pernambuco/ Recife - PE, Brasil.
Received in: 6-8-2017
Accepted in: 11-3-2017
MATERIAL AND METHODS

Study area and data

The dataset was collected in the 2012/2013 agricultural year and comes from an area (Figure 1) of 167.35 ha located in the western region of Paraná, Brazil, near the municipality of Cascavel, with central coordinates defined by latitude 24°57'25''S and longitude 53°34'29''W, and average altitude of 714 m. According to the Köppen classification, the climate of the agricultural area is the Cfa type (Aparecido et al., 2016) and the soil is classified as Oxisol (Embrapa, 2009).

A systematic centered sampling with lattice plus close pairs was performed composed of 99 sample elements (Figure 1), georeferenced with a GEOEXPLORE 3 GPS device with a precision of 5 m. At each sampling point, the soybean yield (Prod, t ha$^{-3}$) and the following soil chemical attributes were determined: calcium (Ca, cmolc dm$^{-3}$), magnesium (Mg, cmolc dm$^{-3}$), potassium (K, mg dm$^{-3}$), phosphorus (P, mg dm$^{-3}$), manganese (Mn, mg dm$^{-3}$), and soil pH.

Geostatistical analysis

Consider a Gaussian stochastic process $\{Z(s), s \in S\}$, with $S \subset \mathbb{R}^2$, being $\mathbb{R}^2$ the two-dimensional Euclidean space. Suppose that the data of this process, $Z(s_1), \ldots, Z(s_n)$, are recorded in known spatial locations $s_i$ ($i = 1, \ldots, n$), and generated by the following model:

$$Z(s) = \mu(s) + \varepsilon(s)$$

where,

$Z(s)$: an $n \times 1$ vector of the variable of interest;
$\mu(s)$: an $n \times 1$ vector that represents the mean of the process, and
$\varepsilon(s)$: an $n \times 1$ vector of errors, with zero mean vector $E[\varepsilon(s)] = 0$ and covariance matrix $\Sigma = C[(\varepsilon(s_i), \varepsilon(s_j))]$, where $\sigma_{uv} = \text{cov}(\varepsilon(s_i), \varepsilon(s_j))$, for $i = u = 1, \ldots, n$.

The mean vector $\mu(s)$ can be written as a linear model $\mu(s) = \mathbf{X}\beta$, where $\beta = (\beta_1, \ldots, \beta_p)^T$ is a $p \times 1$ vector of unknown parameters and $\mathbf{X} = \mathbf{X}(s)$ is an $n \times p$ matrix formed with $p$ covariates at site $s \in S$. According to Mardia & Marshall (1984), the covariance matrix can be expressed in the parametric form $\Sigma = \varphi_1I_n + \varphi_2R(\varphi_3)$ where $I_n$ is an $n \times n$ identity matrix, $\varphi_1 \geq 0$ is the parameter called nugget effect, $\varphi_2 \geq 0$ is the parameter called contribution, $\varphi_3 \geq 0$ is the parameter that defines the range ($\rho$) of the model and $R(\varphi_3) = [(r_{ij})]$ is a $n \times n$ symmetric matrix. The elements $r_{ij}$ of the matrix $R$, for $i, j = 1, \ldots, n$, represent the correlation function between the points $s_i$ and $s_j$, with $r_{ij} = 1$ if $i = j$, $r_{ij} = 0$ if $i \neq j$ and $\varphi_2 = 0$ and $r_{ij} = \varphi_2^{-1}\sigma_{ij}$ if $i \neq j$ and $\varphi_2 \neq 0$, with $\sigma_{ij} = C(h_{ij})$ is a function that depends on $h_{ij} = \|s_i - s_j\|$ which is the Euclidean distance between points $s_i$ and $s_j$. Thus, $Z(s)$ follows a normal $n$-ivariate distribution with mean vector $\mu(s)$ and covariance matrix $\Sigma$ (De Bastiani et al., 2015).

In order to identify the spatial dependence structure of soybean yield, the omnidirectional experimental semivariogram $Y(h_{ij})$ was constructed using the Matheron estimator (Cressie, 2015) and to model the structure of spatial dependence were used models of the Matérn family (Matérn, 1986), presented in [eq. (2)]:

$$C(h_{ij}) = \begin{cases} \varphi_1 + \frac{\varphi_2}{2k-1}\Gamma(k) & \text{if } i = j \\ \frac{\varphi_2}{2k-1}\Gamma(k) & \text{if } i \neq j \end{cases}$$

where,

$K_k(\cdot)$: Bessel function of the third order type
$k > 0$,
$\Gamma(\cdot)$: Gamma function, and
$h_{ij}$: Euclidean distance between points $s_i$ and $s_j$.

For second-order stationary Gaussian and isotropic processes, the semivariogram function has the following relation: $Y(h_{ij}) = C(0) - C(h_{ij})$ (Uribe-Opazo et al., 2012), for $i, j = 1, \ldots, n$, being $h_{ij} \geq 0$ and $C(0)$ the variance of $Z$. In [eq. (2)], the form parameter $k$ controls the behavior close to the origin and the process analytical smoothing. In this study, we considered the fixed values $k = \{0.5; 1.5; 2.5; 4.5\}$ (Diggle & Ribeiro-Jr., 2007).

The parameters estimation of the models was carried out using the maximum likelihood (ML) method and the criteria for selecting the geostatistics model for the covariance matrix were the cross-validation (CV), the trace of the asymptotic covariance matrix of the estimated mean (Tr) and the maximum value of the loglikelihood function (LML) (De Bastiani et al., 2015). To investigate the significance of covariates, the likelihood ratio test (Rao, 1973) was used, and the kriging with external drift was used to obtain the predicted values (Wackernagel, 1995).

Bootstrap for spatial data

In this study, the spatial bootstrap (SB) methods, proposed by Solow (1985) and presented in Algorithm 1 and parametric spatial bootstrap (PSB), proposed by Tang et al. (2006) and presented in Algorithm 2 were used.

Algorithm 1: Spatial Bootstrap (Solow, 1985).

FIGURE 1 Geographical location map of the study area.
a) Considering the spatial data set \( \{Z(g_1), ..., Z(g_n)\} \), determine the vector of residuals \( \hat{\varepsilon}(s) = Z(s) - \hat{\mu} \) where \( \hat{\mu} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} Z \) is the ML estimator of \( \mu \) and \( \hat{\Sigma} \) the ML estimator of the covariance matrix \( \Sigma \). b) Considering the estimated covariance matrix \( \hat{\Sigma} \), use the Cholesky decomposition method to obtain \( \hat{\Sigma} = \hat{L} \hat{L}^T \), that \( \hat{L} \) is a lower triangular matrix of order \( n \); c) Using the inverse matrix \( \hat{L}^{-1} \), determine \( \hat{\varepsilon}_{\text{unc}} = \hat{L}^{-1} \hat{\varepsilon} \), the vector of uncorrelated residuals; d) Considering the elements of the vector of uncorrelated residuals, carry out \( n \) resamples with replacement to form the vector of bootstrap residuals \( \hat{\varepsilon}_{\text{SB}}^* = (\hat{\varepsilon}_{1}^*, ..., \hat{\varepsilon}_{n}^*)^T \); e) The spatial bootstrap sample is obtained by recorrelating bootstrap residuals \( Z^* = \hat{\mu} + \hat{L} \hat{\varepsilon}_{\text{SB}}^* \).

Algorithm 2: Parametric Spatial Bootstrap (Tang et al., 2006).

As Tang et al. (2006) explained, the BSP method does not uncorrelated the residuals vector \( \hat{\varepsilon} \) as in the BS method. Instead, the residuals are generated independent from a standard normal distribution. The theoretical basis of the BSP method can be seen in the study of Sjöstedt-De Luna & Young (2003).

Quantification of uncertainties in geostatistics analysis

The Algorithms 1 and 2 were used to determine \( B = 1000 \) bootstrap samples from the soybean yield dataset. For each sample, a model was adjusted, which allowed to construct the empirical distribution of the parameters of the model and, consequently, to determine point and interval estimators of the parameters using the percentile bootstrap method (Efron, 1982).

To verify the assumption of normality, the uncorrelated residuals \( \hat{\varepsilon}_{\text{unc}} \) obtained by the spatial bootstrap method (Algorithm 1) were ordered and plotted versus the theoretical quantis of the standard normal distribution, resulting in a quantile-quantile plot (q-q plot).

Computational resources

The calculations performed in this study were developed in the R software (R Core Team, 2017). The criteria used to choose the best model, the likelihood ratio test, the bootstrap methods and the multivariate plots q-q plots were implemented by the authors.

RESULTS AND DISCUSSION

According to Conab (2016), the average soybean yield in the study area (Table 1) was lower than the average of the State of Paraná (3.38 t ha\(^{-1}\)) and higher than the national average (2.94 t ha\(^{-1}\)).

The calcium (Ca) contents varied from 2.30 cmolc dm\(^{-3}\) to 13.10 cmolc dm\(^{-3}\) and the magnesium (Mg) contents ranged from 0.85 cmolc dm\(^{-3}\) to 5.75 cmolc dm\(^{-3}\) (Table 1), and they were classified as high (Tomé Jr., 1997). Analyzing the classification of the coefficient of variation (CV) (Table 1) proposed by Pimentel Gomes (1985), the high dispersion of potassium (K) and phosphorus (P) contents and homogeneity of pH data are highlighted.

In Table 2, in all models of spatial dependence fitted for covariance matrix \( \Sigma \), the estimated values of the nugget effect parameter \( \Phi_1 \) were equal to zero, the estimated values of the contribution parameter \( \Phi_2 \) were similar and the estimated values of the range parameter \( l(\sigma_2^2) \) indicated a radius of spatial dependence ranging from 109 to 112 m. According to the criteria CrV, LML and Tr (Table 2), the best fit for the covariance matrix \( \Sigma \) was provided by the Matérn model with parameter \( k = 4.5 \).
TABLE 2. Criteria for selecting the soybean yield model elaborated with covariates considering the covariance function Matérn with different form parameters.

<table>
<thead>
<tr>
<th>k</th>
<th>( \hat{\varphi}_1 )</th>
<th>( \hat{\varphi}_2 )</th>
<th>( \delta )</th>
<th>CrV</th>
<th>LML</th>
<th>Tr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.0000</td>
<td>0.1783</td>
<td>37.2544</td>
<td>111.6041</td>
<td>0.2616456</td>
<td>-54.56541</td>
</tr>
<tr>
<td>1.5</td>
<td>0.0000</td>
<td>0.1783</td>
<td>23.7864</td>
<td>112.8393</td>
<td>0.2589309</td>
<td>-54.07624</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0000</td>
<td>0.1785</td>
<td>18.8569</td>
<td>111.6076</td>
<td>0.2576171</td>
<td>-53.89274</td>
</tr>
<tr>
<td>4.5</td>
<td>0.0000</td>
<td>0.1787</td>
<td>14.2486</td>
<td>109.8489</td>
<td>0.2563692</td>
<td>-53.73963</td>
</tr>
</tbody>
</table>

k: form parameter of the Matérn model; CrV: cross validation; \( \hat{\varphi}_1 \): estimator of the nugget effect; \( \hat{\varphi}_2 \): estimator of contribution; \( \hat{\delta} \): practical range; LML: maximum value of the logarithm of the likelihood function; Tr: trace of covariance matrix. The values in bold indicate the best fit for the covariance matrix according to the used criterion.

According to the fitted Gaussian linear spatial model (GSLM), the value associated with the potassium content (K) (Table 3) was the highest and presented a positive sign, indicating that it was the covariate that contributed the most to an increase in the average of the soybean yield (t ha\(^{-1}\)). This result corroborates with the study of Dos Passos et al. (2015), considering that this nutrient is the second most absorbed by the soybean plant, being essential for the growth and development of the plants (Li et al., 2015). The estimates of the parameters associated to the covariates of phosphorus (P) and manganese (Mn) contents indicated that these covariates contributed little to the estimated soybean yield (Table 3).

The fact that the nugget effect is null indicates that, at small distances, spatial dependence is also small (Vieira & Gonzalez, 2003), meaning that the distances considered among the samples were adequate.

TABLE 3. Estimated parameters for the linear spatial model by the maximum likelihood method considering the covariance function Matérn with form parameter k = 4.5.

<table>
<thead>
<tr>
<th>( \hat{\beta}_0 )</th>
<th>( \hat{\beta}_{Ca} )</th>
<th>( \hat{\beta}_{Mg} )</th>
<th>( \hat{\beta}_{K} )</th>
<th>( \hat{\beta}_{P} )</th>
<th>( \hat{\beta}_{Mn} )</th>
<th>( \hat{\beta}_{PH} )</th>
<th>( \hat{\varphi}_1 )</th>
<th>( \hat{\varphi}_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.54</td>
<td>0.06</td>
<td>0.06</td>
<td>0.85</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.44</td>
<td>0.00</td>
<td>0.18</td>
</tr>
<tr>
<td>(6.8e-1)</td>
<td>(5e-2)</td>
<td>(6e-2)</td>
<td>(3e-1)</td>
<td>(1e-3)</td>
<td>(1e-3)</td>
<td>(1.8e-1)</td>
<td>(8e-2)</td>
<td>(9e-2)</td>
</tr>
</tbody>
</table>

\( \hat{\beta}_0 \): intercept; \( \hat{\beta}_i \): parameter estimator associated with to variable \( i = \{Ca, Mg, K, P, Mn, PH\} \). Ca: calcium content(cmolc dm\(^{-3}\)); Mg: magnesium (cmolc dm\(^{-3}\)); K: potassium (mg dm\(^{-3}\)); P: phosphorus (mg dm\(^{-3}\)); Mn: manganese (mg dm\(^{-3}\)); pH: soil pH; \( \hat{\varphi}_1 \): estimator of the nugget effect; \( \hat{\varphi}_2 \): estimator of contribution; \( \hat{\delta} \): estimator of the parameter that defines the range. The values in parentheses correspond to the standard deviation of each estimated parameter.

To evaluate the reliability of the parameter estimates, the descriptive statistics of the bootstrap replicates obtained by the BS and BSP methods were calculated and the Efron percentile confidence intervals were determined (Table 4).

From the B = 1000 models adjusted by the BS method, two models (0.2%) were excluded. By the BSP method, only one model (0.1%) was excluded. These exclusions were carried out due to, in these settings; the radius of spatial dependence was greater than the maximum distance between samples (1766 m) in the monitored area. As pointed out by Dalposso et al. (2009), the models that present this behavior should be discarded, considering information that goes beyond the area under study.

For both bootstrap methods, the confidence intervals for the parameters associated with calcium (CA), magnesium (Mg) and manganese (Mn) contents were zero, indicating that these covariates may not be significant (Table 4).

Although the zero value is contained in the confidence intervals of the bootstrap replicates of the contribution (\( \hat{\varphi}_2 \)) (Table 4), there is evidence to assume that the contribution is not null, since in addition to the value obtained in the original sample, 0.18, 94% of the bootstrap replicates obtained by the SB method and 95.1% of the bootstrap replicates obtained by the PSB method presented contribution different than zero. Despite the bootstrap methods used to generate replicates of the soybean yield data set, some models provided high range parameters, emphasizing the positive asymmetry (Table 4).
TABLE 4. Descriptive statistics and Efron percentile 95% confidence intervals of the bootstrap distribution of the model parameters of the spatial dependence structure of soybean yield considering the covariates Ca, Mg, K, P, Mn and pH.

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Average</th>
<th>Q3</th>
<th>Max</th>
<th>SD</th>
<th>As</th>
<th>Lₜ</th>
<th>Uₜ</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB</td>
<td>β₀</td>
<td>2.50</td>
<td>4.03</td>
<td>4.54</td>
<td>4.53</td>
<td>4.99</td>
<td>7.16</td>
<td>0.71</td>
<td>0.15</td>
<td>3.15</td>
<td>5.94</td>
</tr>
<tr>
<td></td>
<td>β₁</td>
<td>-0.12</td>
<td>0.02</td>
<td>0.05</td>
<td>0.05</td>
<td>0.09</td>
<td>0.27</td>
<td>0.06</td>
<td>0.31</td>
<td>-0.04</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>β₂</td>
<td>-0.16</td>
<td>0.02</td>
<td>0.07</td>
<td>0.06</td>
<td>0.11</td>
<td>0.28</td>
<td>0.07</td>
<td>-0.11</td>
<td>-0.07</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>β₃</td>
<td>-0.05</td>
<td>0.64</td>
<td>0.84</td>
<td>0.84</td>
<td>1.05</td>
<td>1.83</td>
<td>0.30</td>
<td>0.01</td>
<td>0.25</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>β₄</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td>β₅</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.02</td>
<td>-0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>β₆</td>
<td>-1.14</td>
<td>-0.56</td>
<td>-0.43</td>
<td>-0.43</td>
<td>-0.31</td>
<td>0.15</td>
<td>0.19</td>
<td>-0.15</td>
<td>-0.80</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>β₇</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
<td>0.24</td>
<td>0.06</td>
<td>0.90</td>
<td>0.00</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>β₈</td>
<td>0.00</td>
<td>0.05</td>
<td>0.14</td>
<td>0.12</td>
<td>0.17</td>
<td>0.30</td>
<td>0.07</td>
<td>-0.47</td>
<td>0.00</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>β₉</td>
<td>0.00</td>
<td>13.00</td>
<td>16.40</td>
<td>22.30</td>
<td>283.00</td>
<td>24.80</td>
<td>4.52</td>
<td>11.40</td>
<td>26.20</td>
<td>27.70</td>
</tr>
<tr>
<td></td>
<td>φ₁</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>φ₂</td>
<td>0.00</td>
<td>0.13</td>
<td>0.20</td>
<td>0.23</td>
<td>0.68</td>
<td>0.09</td>
<td>3.12</td>
<td>5.86</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

SB: spatial bootstrap, PSB: parametric spatial bootstrap, β₀: model intercept, βᵢ: parameter estimate associated with variable i = {Ca, Mg, K, P, Mn, PH}, Ca: calcium content (cmolc dm⁻³), Mg: magnesium (cmolc dm⁻³), K: potassium (mg dm⁻³), P: phosphorus (mg dm⁻³), Mn: manganese (mg dm⁻³), pH: soil pH; φ₁: nugget effect, φ₂: range parameter, Min: minimum, Q: first quartile, Q: third quartile, Max: maximum, SD: standard deviation, As: asymmetry coefficient, Lₜ: lower limit of the confidence interval, Uₜ: upper limit of the confidence interval.

When comparing the standard deviations of the model parameters (Table 3), with the respective standard deviations obtained by the bootstrap (Table 4), except for the standard deviation of the range parameter (φ₂). the others presented similar values. In the analysis of significance of the GSLM parameters (Table 5) the hypotheses \( H₀: \betaᵢ = 0 \), \( i = \{ Ca, Mg, K, P, Mn, PH \} \) and \( H₀: \beta₅ = 0 \) are not rejected at 5% significance, demonstrating that spatial bootstrap methods are an alternative to test the individual effect of covariates.

Due to the parameters associated with calcium (Ca, cmolc dm⁻³), magnesium (Mg, cmolc dm⁻³) and manganese (Mn, mg dm⁻³) contents were not significant (Table 5), a new spatial linear model considering potassium (K, mg dm⁻³), phosphorus (P, mg dm⁻³) and soil pH was adjusted, which presented significant parameters.

TABLE 5. Likelihood ratio test (LR) value and p-value for hypotheses \( H₀: \betaᵢ = 0 \), \( i = \{ Ca, Mg, K, P, Mn, PH \} \) in the linear spatial model.

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>LR</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>βᵢ = 0</td>
<td></td>
<td>1.18</td>
</tr>
<tr>
<td>βᵢ = 0</td>
<td></td>
<td>0.86</td>
</tr>
<tr>
<td>βᵢ = 0</td>
<td></td>
<td>7.68</td>
</tr>
<tr>
<td>βᵢ = 0</td>
<td></td>
<td>4.40</td>
</tr>
<tr>
<td>βᵢ = 0</td>
<td></td>
<td>2.76</td>
</tr>
<tr>
<td>βᵢ = 0</td>
<td></td>
<td>5.66</td>
</tr>
</tbody>
</table>

LR: likelihood ratio test, p-value: descriptive level of the test at 5% significance.

According to the Crv, LML and Tr criteria, the best fit for the GSLM was provided by the Matern model with parameter of k form = 4.5 (Table 6).

TABLE 6. Parameters estimated by ML for the linear spatial model considering the covariance function Matern with form parameter k = 4.5 and the covariates K, P and pH.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>β₀</td>
<td>3.99</td>
</tr>
<tr>
<td>β₁</td>
<td>0.84</td>
</tr>
<tr>
<td>β₂</td>
<td>-0.01</td>
</tr>
<tr>
<td>β₃</td>
<td>-0.16</td>
</tr>
<tr>
<td>β₄</td>
<td>0.00</td>
</tr>
<tr>
<td>β₅</td>
<td>0.19</td>
</tr>
<tr>
<td>β₆</td>
<td>13.14</td>
</tr>
</tbody>
</table>

ML: maximum likelihood; β₀: intercept; βᵢ: parameter estimator associated with the variable i = {K, P, PH}, K: potassium (mg dm⁻³), P: phosphorus (mg dm⁻³), pH: soil pH; φ₁: estimator of the nugget effect; φ₂: estimator of the parameter that defines the range. The values in parentheses correspond to the standard deviation of each estimated parameter.

Comparing the complete model (Table 3) with the model elaborated with the covariates identified as significant (Table 6), the value associated with potassium content (K) remained high and positive. The value near zero of the parameter associated with the phosphorus content indicates that its influence on the average yield is minimal.

In relation to soil pH, the estimated value of the parameter presented a loss of intensity, going from -0.44 to -0.16. The negative signal of this parameter evidences an inverse relation with the yield, which was already expected, since, in general, the increase of soil pH
decreases the availability of the micronutrients (Sousa et al., 2007), which, consequently, can result in yield losses.

Observing the productivity contour map (Figure 2) elaborated with the parameters of the fitted spatial linear model (Table 6), the majority of the area (84%) presented yield values varying from 3.175 t ha\(^{-1}\) to 3.592 t ha\(^{-1}\).

CONCLUSIONS

The SB and PSB methods allowed to quantify the uncertainties associated to the spatial dependence structure and to evaluate the individual significance of the parameters associated to the average of the spatial linear model, allowing the determination of a model with a smaller number of parameters.

The fact that the linear spatial model formed with the potassium (K, mg dm\(^{-3}\)), phosphorus (P, mg dm\(^{-3}\)) and soil pH covariates had a spatial dependence radius close to the minimum distance between points, provided a map characterized by circular regions and a high amount of near-average values.

The use of the uncorrected residuals in the q-q plot elaboration allowed verifying the normality assumption of the soybean yield data, a presupposition that, in many cases, is simply assumed without verification. This finding is extremely important in the analysis since a violation of the normality assumption can invalidate statistical hypothesis tests.

ACKNOWLEDGEMENTS

The authors would like to thank to UTFPR, UNIOESTE, UFPE, FUNDAÇÃO ARAUCÁRIA and CAPES and to the Fondecyt Project 1150325 - Chile for the support provided to carry out this study.

REFERENCES


Dos Passos AMA, Rezende PM, Carvalho ER, Ávila FW (2015) Biochar, farmyard manure and poultry litter on chemical attributes of a Distrophic Cambissol and soybean crop. Revista Brasileira de Ciências Agrárias 10(3):382-388. DOI: http://dx.doi.org/10.5039/agraria.v10i3a4546


Solow A (1985) Bootstrapping correlated data. Mathematical Geology 17(7):769-775. DOI: http://dx.doi.org/10.1007/BF01031616


