

Spatial-temporal Analysis of Soybean Productivity Using Geostatistical Methods

Gustavo Henrique Dalposso

Federal University of Technology – Paraná - (UTFPR), Brazil

Miguel Angel Uribe-Opazo

Western Paraná State University – (UNIOESTE), Brazil

Fernanda De Bastiani

Federal University of Pernambuco - (UFPE), Brazil

Received: April 6, 2021 Accepted: April 28, 2021 Published: May 2, 2021

doi:10.5296/jas.v9i2.18494

URL: <https://doi.org/10.5296/jas.v9i2.18494>

Abstract

To have information about the soybean productivity over the crop years is essential to define strategies to increase profits and reduce costs and most important to reduce environmental impacts. One form of monitoring is the use of Geostatistical methods, which allow us to obtain maps with more accurate predictions. In this paper, an area of 127.16 ha was studied during six crop years between 2012/2013 and 2017/2018. We found that productivity values vary between crop years, mainly due to uncontrollable climatic factors. The removal of influential points caused changes in the predicted values showed in the maps, and the use of scaled semivariograms allowed us to obtain similar maps to those obtained considering the model without influential points, then there was no need to exclude observations. The use of a model with replicates helped to identify regions where productivity was lower. The use of explanatory variables allowed us to elaborate a more accurate thematic map in the 2017/2018 crop year, which was well evidenced by the prediction standard error map.

Keywords: local influence, Prediction standard error map, Scaled semivariogram, Spatial variability

1. Introduction

Soybean (*Glycine max* (L.) Merrill) is a source of food, protein, and oil (Syed et al., 2019), and

one of the most important crops worldwide. Brazil is currently the second-largest soybean producer, standing behind only the USA, and contributes approximately 34.7% of global production (Schwalbert et al., 2020). Given the importance of soy for agribusiness, research is often carried out to develop productivity maps, one of the most widely used precision farming techniques (Vega et al., 2019).

Because the production of soybean has a high cost due to the use of fertilizers, pesticides, machines, and seeds (Costa and Santana, 2018), the use of productivity maps help in identifying regions with greater productive potential. These regions can be managed differently, optimizing inputs, improving profits, and reducing environmental impacts (Vatsanidou et al., 2017). Among the many methodologies used to create thematic maps, geostatistical methods have been widely applied in precision agriculture (Lamour et al., 2020).

Although the result of a geostatistical analysis is, in most cases, a map, it can include a high number of techniques that aim to obtain more accurate maps when compared to maps generated by deterministic methods. The literature presents works that produced maps using, for example, local influence techniques (Grzegozewski et al., 2013 and Uribe-Opazo et al., 2021), scaled semivariograms (Wendpap et al., 2015), spatial linear models with replicates (De Bastiani et al., 2017), and bootstrap methods (Dalposso et al., 2019). Among these works, we also find geostatistical studies effectively related to soybean productivity in a spatial-temporal context. Bakhsh et al. (2000) and Bottega et al. (2017) studied three crop years, and De Bastiani et al. (2017) studied five crop years. In these analyses, obtaining sample data becomes a complicated, time-consuming, and often uncertain task, given that the soybean culture often depends on unpredictable and uncontrollable factors.

In this context, the objective of this paper was to use geostatistical methods to monitor an agricultural area cultivated with soybean for six consecutive crop years. We investigated the impact of removing influential points and the use of scaled semivariograms in the elaboration of maps. We designed a map of soybean productivity ranging from the 2012/2013 to 2016/2017 crop years. Also, we have considered the use of explanatory variables to obtain a more accurate map in 2017/2018, given the missing data this crop year.

2. Material and Methods

2.1 Study Area and Data

The experiment collected the soybean productivity data set ($t\ ha^{-1}$) between the crop years 2012/2013 and 2017/2018. The crop year comprises from October to February of the following year. The agricultural area has 127.16 ha and is in the western region of the state of Paraná, Brazil, close to the municipality of Cascavel, with central coordinates of latitude $24^{\circ}57'25''$ S, longitude $53^{\circ}34'29''$ W, and an average altitude of 714 m (Fig.1).

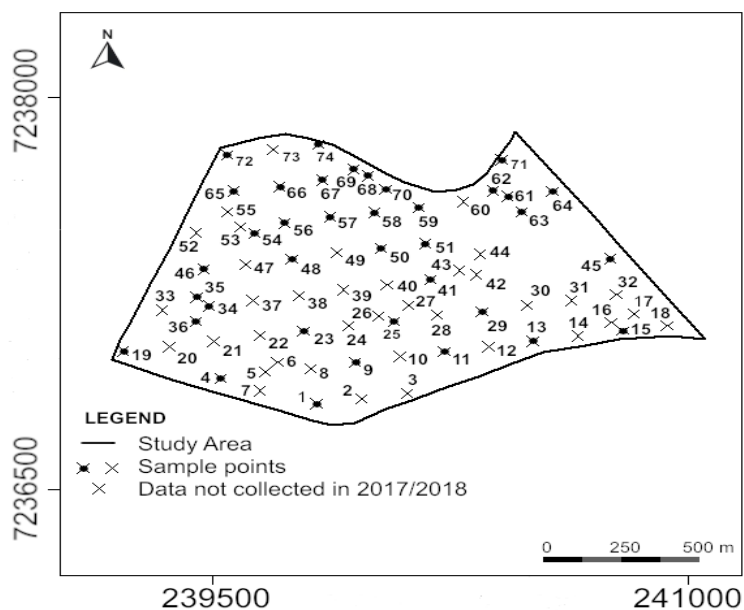


Figure 1. Geographical information of the study area

According to the Köppen classification, the region's climate is type Cfa with an average annual temperature of 21°C. The soil is classified as dystrophic Red Latosol, with a clay texture (EMBRAPA, 2013). A defined sample configuration in this experimental is lattice plus close pairs (Chipeta et al., 2017), with a minimum distance between the points of the regular grid of 141 meters and in some locals, randomly selection, sampling was carried out with minor distances of 75 meters and 50 meters between pairs of points, totalizing 74 sampling points (in each crop year). These samples are located and geo-referenced by means of a signal receiver as a global positioning system (GPS), in a UTM spatial coordinate system. Because of the issues that occurred during the collection in the 2017/2018 crop year, only 37 sampling points were collected this crop year (Fig. 1).

The productivity of soybeans was collected based on the amount of grains harvested from the plants distributed in two rows, along a meter in the length of each point, representing the plot. After sorting, grains are weighed for each plot, checking the water content and correcting for 13% moisture, and then converted in $t\ ha^{-1}$.

2.2 Geostatistical Analysis

To study the spatial dependence was considered $\{Z(s_i), s_i \in S\}$ an isotropic and second-order stationary stochastic process where $S \subset \mathbb{R}^2$, with \mathbb{R}^2 as a bi-dimensional Euclidian space, with an $n \times 1$ random vector $Z = Z(s) = (Z(s_1), \dots, Z(s_n))^T$ of the response variable, which corresponds to the spatial locations known in s_i , with $i = 1, \dots, n$. It can be written matrix form as $Z = X\beta + \varepsilon$, in which X is an $n \times p$ matrix of columns with full rank, with

its lines $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ as a $1 \times p$ vector of explanatory variables in position \mathbf{s}_i , $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is a $p \times 1$ vector of unknown parameters to be estimated, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$, is an $n \times 1$ vector of random errors. It is assumed that the random errors $\varepsilon_i = \varepsilon(\mathbf{s}_i)$ have a zero mean, that is, $E[\varepsilon(\mathbf{s}_i)] = 0$, and the variation between points in space is determined by a covariance function $Cov[\varepsilon(\mathbf{s}_i), \varepsilon(\mathbf{s}_u)] = Cov[Z(\mathbf{s}_i), Z(\mathbf{s}_u)] = C(\mathbf{s}_i, \mathbf{s}_u) = \sigma_{iu}$, for $i, u = 1, \dots, n$.

The spatial modeling in $\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ depends on the structure of the covariance matrix $\boldsymbol{\Sigma} = [(\sigma_{iu})]$, in which, $\sigma_{iu} = C(\mathbf{s}_i, \mathbf{s}_u)$ for $i, u = 1, \dots, n$, of the Gaussian stochastic process \mathbf{Z} . The covariance function $C(\mathbf{s}_i, \mathbf{s}_u)$ is used in the spatial dependence study of the stationary process and is specified by a vector $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$, of the form $\boldsymbol{\Sigma} = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}$, (Mardia and Marshall, 1984), in which, \mathbf{I}_n is the $n \times n$ identity matrix, φ_1 is the nugget effect or variance error ($\varphi_1 \geq 0$); φ_2 is the contribution or dispersion variance ($\varphi_2 \geq 0$); $\mathbf{R} = \mathbf{R}(\varphi_3) = [(r_{iu})]$, is a $n \times n$ symmetrical matrix, whose elements are in function of $\varphi_3 > 0$, (Uribe-Opazo et al., 2012; De Bastiani et al., 2015). The parametric form of covariance matrix $\boldsymbol{\Sigma}$ occurs for many stationary and isotropic processes, in which the covariance $C(\mathbf{s}_i, \mathbf{s}_u) = C(h_{iu})$ is defined according to the covariance function $C(h_{iu}) = \varphi_2 r_{iu}$. In the covariance function $C(h_{iu})$, the variance of the Gaussian stochastic process \mathbf{Z} is given by $C(0) = \varphi_1 + \varphi_2$. On the hypothesis that the errors present normal distribution of n-variate probabilities $\boldsymbol{\varepsilon} \sim N_n(0, \boldsymbol{\Sigma})$, we have that the spatial linear model has an n-variate normal distribution, $\mathbf{Z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, in which the vector of unknown parameters of the model $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\varphi}^T)^T$, can be estimated by maximizing the logarithm of the likelihood function, defined in (1):

$$L(\theta) = -\left(\frac{n}{2}\right) \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (\mathbf{Z} - \mathbf{X}\beta)^T \Sigma^{-1} (\mathbf{Z} - \mathbf{X}\beta) \quad (1)$$

The $\theta = (\beta^T, \varphi^T)^T$ parameters of the model were estimated, considering the maximum likelihood method (ML). In modeling considering the scaled semivariogram (see Subsection 2.5) the ordinary least squares method was used. To observe the behavior of the data $\mathbf{Z}(\mathbf{s})$, we performed an exploratory analysis, identifying the discrepant points (outliers). To determine the spatial dependence structure of the data in each agricultural year, it was constructed omnidirectional experimental semivariograms using the Matheron estimator. To model the spatial dependency structures, we used models of the Matérn family with different shape parameters $k > 0$ (Matérn, 1986). The best fits were chosen by cross-validation (Faraco et al., 2008) and the predictions of unsampled data were obtained by ordinary kriging. In the analysis involving explanatory variables (see Subsection 2.7), kriging with external drift (KED) was used.

2.3 Spatial Dependence Index

Seidel and Oliveira (2014, 2016) build the proposed spatial dependence index, called $SDI = FM[\varphi_2 / (\varphi_1 + \varphi_2)] \cdot [a/q.MD]$ where, FM is the fitted model factor ($FM_{\text{Exponential}} = 0.317$, $FM_{\text{Matérn}(k=2)} = 0.424$, $FM_{\text{Gaussian}} = 0.504$); φ_1 is the nugget effect; φ_2 is the contribution; a is the range and $q.MD$ is the value that corresponds to the fraction (q), reached of the maximum distance (MD) between sampled points. Seidel and Oliveira (2016) recommend $q = 0.5$ for considering 50% of the maximum distance to generate the semivariogram as a cutoff.

Seidel and Oliveira (2014) propose the following classification (categorization) for the SDI. For the exponential semivariogram: $SDI_{\text{Exponential}} \leq 6\% \rightarrow$ weak spatial dependence; $6\% < SDI_{\text{Exponential}} \leq 13\% \rightarrow$ moderate spatial dependence; $SDI_{\text{Exponential}} > 13\% \rightarrow$ strong spatial dependence. For the Gaussian semivariogram: $SDI_{\text{Gaussian}} \leq 9\% \rightarrow$ weak spatial dependence; $9\% < SDI_{\text{Gaussian}} \leq 20\% \rightarrow$ moderate spatial dependence; $SDI_{\text{Gaussian}} > 20\% \rightarrow$ strong spatial dependence.

2.4 Local Influence on the Response Variable

In modeling studies, the removal of points (*Jackknife*) is one of the best-known techniques to assess the impact of a possible influential observation on the parameter estimates. However, a problem that can occur with the individualized withdrawal of observations is the effect of failing to detect jointly influential points. There are several robust methods for detecting influential points which low computational cost (Peña and Yohai, 1999). Cook (1986) presents a very innovative proposal in the field of diagnosis, where he proposes to evaluate the joint influence of the observations on small disturbances in the model, instead of the evaluation by

the individual or joint removal of points. This methodology is called local influence and was widely accepted by researchers in the area.

The local influence method proposed by Cook (1986) is the most widely used diagnostic method in the literature. Considered a vector of ω disturbances, its influence can be assessed by measuring the likelihood distance, given by $LD(\omega) = 2\{L(\hat{\theta}) - L(\hat{\theta}_\omega)\}$, in which $\hat{\theta}$ is the ML estimator of θ of the postulated model and $\hat{\theta}_\omega$ is the ML estimator of θ of the disturbed model (Uribe-Opazo et al., 2012).

De Bastiani et al. (2015) considered the perturbation scheme $Z_\omega(s) = Z(s) + \Sigma^{1/2}\omega$, which is based on Zhu et al. (2007), in which $Z(s)$ is the response vector, and Σ is the covariance matrix. To evaluate the existence of influential points, we used the graph B_i versus i (order of data) presented by Poon and Poon (1999) as a diagnostic measure.

2.5 Scaled Semivariogram

The literature (Mercante et al., 2003, Vieira et al., 2010 and Wendpap et al., 2015) brings studies on spatial and temporal processes that have used the technique of scheduling semivariograms as a proposal for modeling dependence in space and time. This technique consists of estimating parameters that define the structure of spatial and temporal dependence by fitting a single model to the semivariograms, called scaled semivariogram.

The experimental semivariance function of soybean productivity ($t\ ha^{-1}$) from crop years 2012/2013 to 2017/2018 was scaled using the equation $\gamma_{esc}(h) = [\gamma_i(h)/Var_i(Z)]$, $i = 1, \dots, r$, (Vieira et al., 2010). In this equation, r is the number of periods (crop years) of the study, h indicates the distance between point pairs, $\gamma_i(h)$ represents the original semivariance function of the i -th period, and $Var_i(Z)$ is the sample variance of the observations of the i -th period.

2.6 Spatial Linear Model With Replicates

De Bastiani et al. (2017) proposed modeling the structure of spatial and temporal dependence using linear spatial models with multiple independent replications and using the method of maximum likelihood for parameter estimation. They consider $Z = vec(Z_1(s), \dots, Z_r(s))$ as

the random vector of stochastic processes of r independent replicates of order $nr \times 1$, belonging to the family of Gaussian distributions dependent on the position $s_j \in S \subset \mathbb{R}^2$, for $j = 1, \dots, n$, in which $s = (s_1, \dots, s_n)^T$. The operator "vec" represents a linear transformation that converts the matrix into a column vector.

This paper, $Z_j(s) = (Z_j(s_1), \dots, Z_j(s_n))^T$ represents soybean productivity (t ha⁻¹) collected from the $n = 74$ sampling points in each crop year, in the interval from 2012/2013 to 2016/2017, thus, $j = 1, \dots, 5$ ($r = 5$). Considering the methodology proposed by De Bastiani et al. (2017) and the samples of soybean productivity obtained from these crop years, we fit the models of the Matérn family with different shape parameters $k > 0$, and the best model is chosen by cross-validation (Faraco et al., 2008).

2.7 Kriging with External Drift

Kriging with external drift (KED) (Wackernagel, 1995) is a methodology that can be used to obtain more accurate thematic maps since, in the model $Z(s) = \beta \mathbf{1} + \varepsilon(s)$, the process mean (β) is described by auxiliary variables, unlike in ordinary kriging, which considers the constant mean. In the case of KED, predictions are made as in ordinary kriging, with the difference that the residual covariance matrix is extended with auxiliary predictors (Webster and Oliver, 2007).

As explained by Hengl et al. (2003), predictions in a new location s_0 are made by

$\hat{z}_{\text{KED}}(s_0) = \delta_0^T z$, where δ_0 is the vector of the KED weights, and z is the vector of the n initial observations.

Considering the soybean productivity data (t ha⁻¹) collected at the same 37 points sampled in 2017/2018 as explanatory variables, the estimation of the model parameters was performed using the maximum likelihood (ML) method and the criteria for selecting the geostatistical model for the covariance matrix was cross-validation (Faraco et al., 2008).

2.8 Prediction Standard Error Maps

The standard error is considered a summary measure of prediction precision since it characterizes how close they are, on average, to the unknown real values. The calculation of prediction variances is necessary to elaborate prediction standard error maps. Kyriakidis and Goodchild (2006) present the theoretical concepts for determining the prediction variance

when interpolation is performed by universal kriging, while Lloyd and Atkinson (2001) described the concept for ordinary kriging. This study presents the prediction variance when interpolation is performed by KED.

2.9 Map Comparison

It was compared the soybean productivity maps ($t\ ha^{-1}$) using the error matrix and the calculate the Kappa index (\hat{K}) (Cohen, 1960) following the levels of agreement proposed by Landis and Kock (1977) which considers the similarity is low if $\hat{K} \leq 0.67$, the similarity is moderate if $0.67 < \hat{K} \leq 0.80$ and the similarity is high if $\hat{K} > 0.80$. It was prepared the soybean productivity ($t\ ha^{-1}$) difference maps in absolute value, highlighting the differences between the map considering all points in 2017/2018 and the map generated using KED.

2.10 Computational resources

The analyses were done using the Microsoft R Open software (Microsoft R Core Team, 2021). The methods were implemented by the authors using resources available in the geoR package (Ribeiro Jr. and Diggle, 2001).

3. Results and Discussion

3.1 Exploratory Analysis

The average soybean productivity obtained in the agricultural year 2012/2013 ($3.26\ t\ ha^{-1}$) was satisfactory, considering that it was close to the average obtained in Paraná ($3.35\ t\ ha^{-1}$), the Brazilian state that received the second-highest average productivity in this crop year (CONAB, 2019). In the crop year 2013/2014, the average soybean productivity ($4.23\ t\ ha^{-1}$) was higher than the average in the state of Roraima ($3.18\ t\ ha^{-1}$), the largest producer in this crop year (CONAB, 2019). The average productivity in the crop year 2014/2015 was low when compared to the national ($2.30\ t\ ha^{-1}$) and state ($3.29\ t\ ha^{-1}$) averages in the same crop year (CONAB, 2019).

The low productivity in the crop year 2014/2015 can be explained by the occurrence of water stress that occurred during crop development. In the crop year 2015/2016, the average productivity ($2.46\ t\ ha^{-1}$) was lower than the state average ($3.09\ t\ ha^{-1}$), which is also a reflection of climatic adversities, which, in turn, limited production and caused losses of 10% of the harvest (SEAB, 2016). The productivities from crop years 2016/2017 and 2017/2018 ($3.06\ t\ ha^{-1}$ and $3.07\ t\ ha^{-1}$, respectively) were low when compared to the state ($3.73\ t\ ha^{-1}$ and $3.5\ t\ ha^{-1}$) and national ($3.36\ t\ ha^{-1}$ and $3.39\ t\ ha^{-1}$) productivities.

It can be observed that in the crop years 2012/2013, 2013/2014 and 2014/2015 the sample points with high productivity 28 ($4.51\ t\ ha^{-1}$), 7 ($5.76\ t\ ha^{-1}$) and 33 ($3.17\ t\ ha^{-1}$), were classified as outliers (Fig. 2).

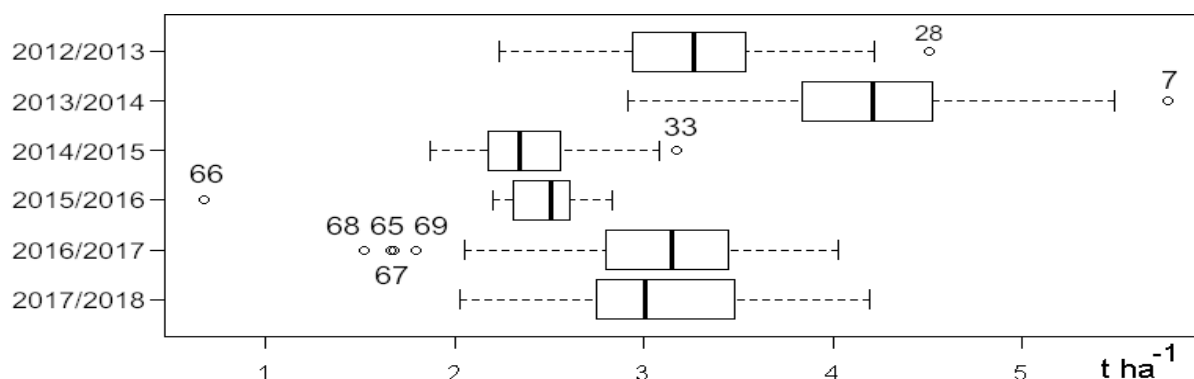


Figure 2. Soybean productivity ($t\ ha^{-1}$) boxplots from 2012/2013 to 2017/2018. The values highlighted in the boxplots indicate the sampling points identified as outliers

In the crop year 2015/2016, sampling point 66 ($0.67\ t\ ha^{-1}$), with low productivity, was classified as an outlier and in the crop year 2016/2017, the sampling points with low productivity 69 ($1.79\ t\ ha^{-1}$), 65 ($1.67\ t\ ha^{-1}$), 67 ($1.66\ t\ ha^{-1}$), and 68 ($1.52\ t\ ha^{-1}$) were classified as outliers (Fig 2.).

3.2 Spatial Analysis

In the crop year 2012/2013, the geostatistical model selected by cross-validation was Matérn model with shape parameter $k=2$ ($M_{2.0}$), presented a range of 101.25 m and spatial dependency index (SDI) of 1.24%, indicating a weak spatial dependence (Table 1). The cross-validation indicated the Gaussian model (M_{inf}) as the best fitted in 66% of the studied crop years, highlighting a weak spatial dependence in crop years 2013/2014, 2014/2015, 2015/2016, and 2016/2017 (Table 1).

In the last crop year studied (2017/2018), cross-validation indicated the exponential model ($M_{0.5}$) as the best fit, which presented a moderate spatial dependence (SDI=6.01%) and the largest spatial dependency radius (326.23 m) among all fitted models.

Table 1. Geostatistical models adjusted to soybean productivity data ($t\ ha^{-1}$) from crop years 2012/2013 to 2017/2018

Crop year	Model	$\hat{\beta}$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	range (m)	SDI
2012/2013	$M_{2.0}$	3.259	0.159	0.052	018.861	101.25	1.24 % (wd)
2013/2014	M_{inf}	4.227	0.000	0.289	037.701	065.25	3.03 % (wd)
2014/2015	M_{inf}	2.374	0.056	0.021	156.422	270.74	4.32 % (wd)
2015/2016	M_{inf}	2.460	0.043	0.034	119.917	207.55	5.37 % (wd)
2016/2017	M_{inf}	3.084	0.066	0.211	107.573	186.19	8.31 % (wd)
2017/2018	$M_{0.5}$	3.015	0.138	0.138	108.899	326.23	6.01 % (md)

M_k : Matérn model with shape parameter $k = \{0.5, 2.0\}$; M_{inf} : Gaussian model; $\hat{\beta}$: estimation of the model mean; $\hat{\varphi}_1$: estimated value of the nugget effect; $\hat{\varphi}_2$: estimated value of the sill; $\hat{\varphi}_3$: estimated value of the parameter that defines the range; wd: weak spatial dependence; md: moderate spatial dependence; SDI: spatial dependence index.

The weak spatial dependence identified by SDI's can be attributed to extrinsic factors and indicates that the soybean productivity values from nearby points do not have a definite relationship. In the crop year 2013/2014, the absence of a nugget effect estimate ($\hat{\phi}_1 = 0$) and the low range of the model (65.25 m) (Table 1) stand out. The low range found (65.25 m) indicates that the spatial continuity quickly disappears (Vieira and Gonzalez, 2003). When analyzing the ranges of geostatistical models (Table 1), the range varies considerably (260.98 m) over the studied agricultural years, indicating a lack of temporal stability in the dependence radius of the spatial structure of the data. This fact can also be observed in the works conducted by Bakhsh et al. (2000) and Bottega et al. (2017), who obtained a variation range of 91 m and 248 m, respectively.

3.3 Local Influence Analysis

Local influence analysis (Fig. 3) indicated as potential influential points the samples 5 and 61 collected in 2012/2013, 15 and 69 collected in 2013/2014, 42 and 55 collected in 2014/2015, 54 and 67 collected in 2015/2016, 34 and 54 collected in 2016/2017 and 11 and 12 collected in 2017/2018.

As stated by Dalposso et al. (2012), influential samples can change a decision in the determination of geostatistical models or the construction of thematic maps. Regarding the sample elements of soybean productivity identified as outliers (Fig. 2), no point was classified as influential.

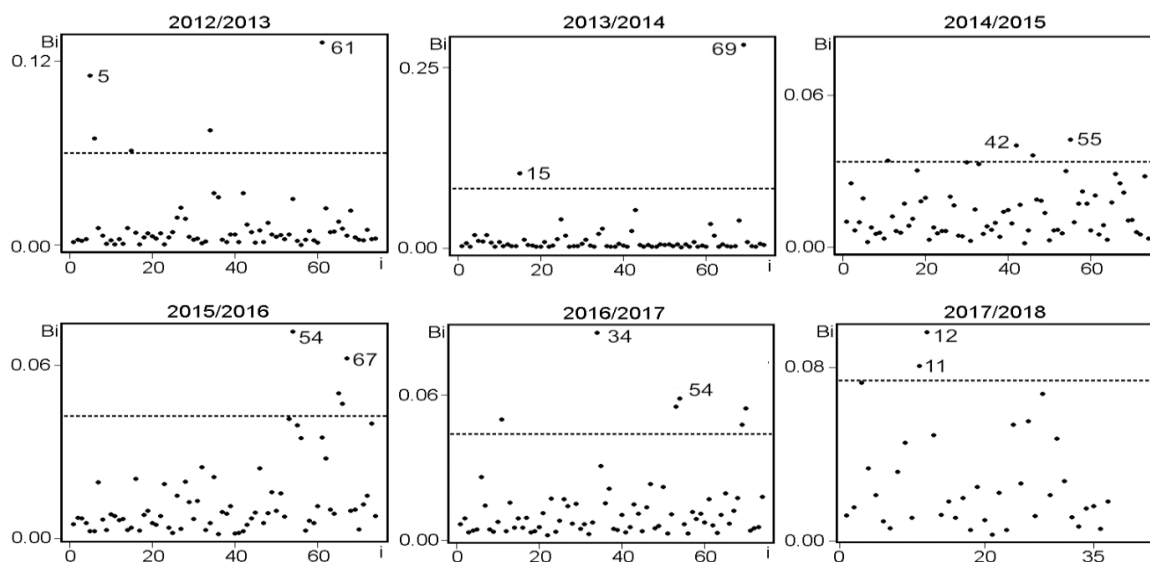


Figure 3. Plots of local influence B_i versus the order of the data. In highlight, the sampling points identified as influential

The wide-spread and non-regularity of these sample points in the initial three crop years stand out when analyzing the positioning of outliers and influential points in Fig. 1. In crop year 2012/2013, sampling point 28 (4.51 t ha^{-1}) was classified as an outlier and is close to the central region of the studied area. Influential points 5 (2.89 t ha^{-1}) and 61 (3.61 t ha^{-1}) are in the southwest and northeast of the studied area, respectively.

In the crop year 2013/2014, sampling point 7 (5.76 t ha⁻¹), classified as an outlier, is in the southwest, while the influential points 15 (3.80 t ha⁻¹) and 69 (3.97 t ha⁻¹) are in the southeast and north of the studied area, respectively. In the crop year 2014/2015, the outlier point 33 (3.17 t ha⁻¹) is in the west, and the influential points 42 (2.24 t ha⁻¹) and 55 (2.12 t ha⁻¹) are in the central and northwest regions, respectively. From crop years 2015/2016 to 2017/2018, there is a tendency for outliers in a stretch that extends from northwest to southwest. In the crop year 2015/2016, sampling point 66 (0.67 t ha⁻¹), classified as an outlier, is in the northwest region, very close to sampling points 54 (2.42 t ha⁻¹) and 67 (2.71 t ha⁻¹), classified as influential.

In the crop year 2016/2017, the sampling points classified as outliers 69 (1.79 t ha⁻¹), 65 (1.67 t ha⁻¹), 67 (1.66 t ha⁻¹), and 68 (1.52 t ha⁻¹) are also in the northwest region, above the points 34 (2.05 t ha⁻¹) and 54 (2.92 t ha⁻¹), classified as influential in the crop year 2016/2017, and points 34 (3.24 t ha⁻¹) and 35 (3.77 t ha⁻¹), classified as influential in 2017/2018. It is essential to highlight the sampling points classified as influential in two consecutive harvests, namely, point 54 in crop years 2015/2016 and 2016/2017 and point 34 in crop years 2016/2017 and 2017/2018.

The fact that outliers are not influential points can occur, as explained by Tóth et al. (2013), considering that the concepts are different. An outlier is an extreme value that does not follow the general tendency. However, this point will not necessarily produce changes in the model, while an influential point is one that, when excluded, causes a significant change in parameter estimates. Therefore, it was decided to conduct new statistical modeling by removing potentially influential sample elements (Table 2).

Table 2. Geostatistical models adjusted to soybean productivity data (t ha⁻¹) from crop years 2012/2013 to 2017/2018, without the influential points

Crop year	Model	$\hat{\beta}$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	range (m)	SDI
2012/2013	M _{0.5}	3.256	0.000	0.214	048.360	144.87	5.34 % (wd)
2013/2014	M _{inf}	4.235	0.000	0.293	044.585	077.16	4.51 % (wd)
2014/2015	M _{inf}	2.376	0.048	0.030	127.372	220.45	11.43 % (md)
2015/2016	M _{inf}	2.457	0.033	0.046	133.491	231.05	6.52 % (wd)
2016/2017	M _{inf}	3.098	0.184	0.248	103.345	178.87	4.96 % (wd)
2017/2018	M _{inf}	2.918	0.000	0.261	130.854	226.48	13.26 % (md)

M_k: Matérn model with shape parameter $k = \{0.5, 2.0\}$; M_{inf}: Gaussian model; $\hat{\beta}$: estimation of the model mean; $\hat{\varphi}_1$: estimated value of the nugget effect; $\hat{\varphi}_2$: estimated value of the sill; $\hat{\varphi}_3$: estimated value of the parameter that defines the range; wd: weak spatial dependence; md: moderate spatial dependence; SDI: spatial dependence index.

In the crop year 2012/2013, after removing potentially influential points, cross-validation indicated the M_{0.5} model as the best fit, with a range of 144.87 m and weak spatial dependence (Table 2). Similar to what occurred in the initial modeling (Table 1), we identified the M_{inf} model as the best fit in most of the studied agricultural years (83%) when the influential points were removed, highlighting a moderate spatial dependence in the crop years 2014/2015 and 2017/2018 (Table 2).

Although the removal of the influential points did not change the model selected by

cross-validation in the crop year 2014/2015, the degree of spatial dependence changed from weak to moderate and the range passing from 270.74 m (Table 1) for 220.45 m (Table 2). In crop year 2015/2016, the removal of the influential points provided an increase in range from 207.55 m (Table 1) to 231.05 m (Table 2), and the degree of spatial dependence remains weak.

Comparing the fitted models without the influencing points (Table 2) with the complete models (Table 1), it is observed that the parameter estimates were affected for the crop years in which cross-validation maintained the Gaussian model as the best fit (namely from crop years 2013/2014 to 2016/2017). For example, in the crop year 2016/2017, the estimated nugget effect of the complete model was 0.066 (Table 1) and 0.184 in the model without the influential points (Table 2). This small difference caused a change in the radius of spatial dependence (from 186.19 m to 178.87 m). However, the interpretation of these models remains similar because both have a weak spatial dependence.

In a similar analysis, considering the crop year 2012/2013, the removal of the influential points provided a considerable change. In addition to the difference in the models selected by cross-validation, the spatial dependency radius increased by 43.62 m. These changes corroborate the results found by Uribe-Opazo et al. (2012), in which the points identified as influential affected the parameter estimates.

3.4 Scaled Semivariogram

The purpose of considering a scaled semivariogram to model the spatial variability of soybean productivity in the studied crop years is to obtain more accurate estimates, with a view to better characterizing spatial variability. As explained by Silva et al. (2018), one of the advantages of the scaled semivariogram is that several semivariograms can be drawn on the same graph. Although the sample set is not small (Fig. 1), depending on the pre-defined distances (lags) used to calculate the semivariogram, the amount of semivariances obtained may be low. As in this case, the parameters estimates were obtained by least squares, the results obtained may be inaccurate since the regression models can produce biased estimates (Dalposso et al., 2016).

After adjusting the models to the scaled semivariogram, the cross-validation values indicated the $M_{0.5}$ as the best adjustment, with parameters $\hat{\varphi}_1 = 0.000$, $\hat{\varphi}_2 = 1.027$, $\hat{\varphi}_3 = 82.48$, and a range of 247.09 m. The strong spatial dependence stands out when analyzing the parameters of this model.

3.5 Comparisons Between Maps

The soybean productivity maps of the studied crop years were elaborated using the scaled semivariogram to observe the behavior of soybean productivity in non-sampled locations and quantify the impacts generated by the exclusion of influential points (Fig. 4).

Comparing the 2012/2013 maps (Fig. 4), it was observed that the region of the sampling point identified as an outlier (point 28 in Fig. 1) was more evident in the maps obtained without the influential points and by the scaled semivariogram. Visually, it was verified a higher similarity between the map generated without the influential points and the map created with the scaled semivariogram. This similarity is evidenced by the Kappa index (\bar{K}) since, according to the

classification established by Ladis and Kock (1997), the similarity is regular ($\widehat{K} = 0.24$) when comparing the map obtained with all points with the map obtained without the influential points, and too the similarity is low ($\widehat{K} = 0.13$) when comparing with the map obtained from the scaled semivariogram. The similarity is good ($\widehat{K} = 0.64$) when comparing the map obtained without the influential points with the one obtained from the scaled semivariogram.

A high similarity ($\widehat{K} = 0.90$) was prominent in the crop year 2013/2014 between the map obtained with all points and obtained without the influential points. However, circular regions formed close to the sample points stand out, a phenomenon not observed in the map obtained by the scaled semivariogram. The circular regions close to the sampling points as explained by Menezes et al. (2016), represent the phenomenon known as the "bull eyes effect." This phenomenon was also observed by Dalposso et al. (2018) and occurs when the geostatistical model has a range close to the minimum distance between sample points. Because the models used to create the maps above had a range of 65.25 m (Table 1) and 77.16 m (Table 2), respectively, and the shortest distance between points was 49.1 m, the presence of circular regions is justified.

In the crop year 2014/2015, the highest calculated \widehat{K} was 0.76 when comparing the map obtained with all points and the map obtained without the influential points, indicating a very good similarity. In crop year 2015/2016, all comparisons between maps indicated a good similarity, with emphasis on the correlation between the map obtained without influential points and the map obtained by the scaled semivariogram, which provided the highest Kappa index ($\widehat{K} = 0.79$).

In crop year 2016/2017, there is a good similarity between the map obtained with all points and the map obtained without influential points ($\widehat{K} = 0.65$) and a moderate similarity between the map obtained without influential points and the map obtained by the scaled semivariogram ($\widehat{K} = 0.52$). In the last crop year studied 2017/2018, the removal of the influential points (points 34 and 35 in Fig. 1) provided a considerable change in the thematic map, resulting on the map obtained with all points and the map without the influential points being considered as presenting low similarity ($\widehat{K} = 0.10$).

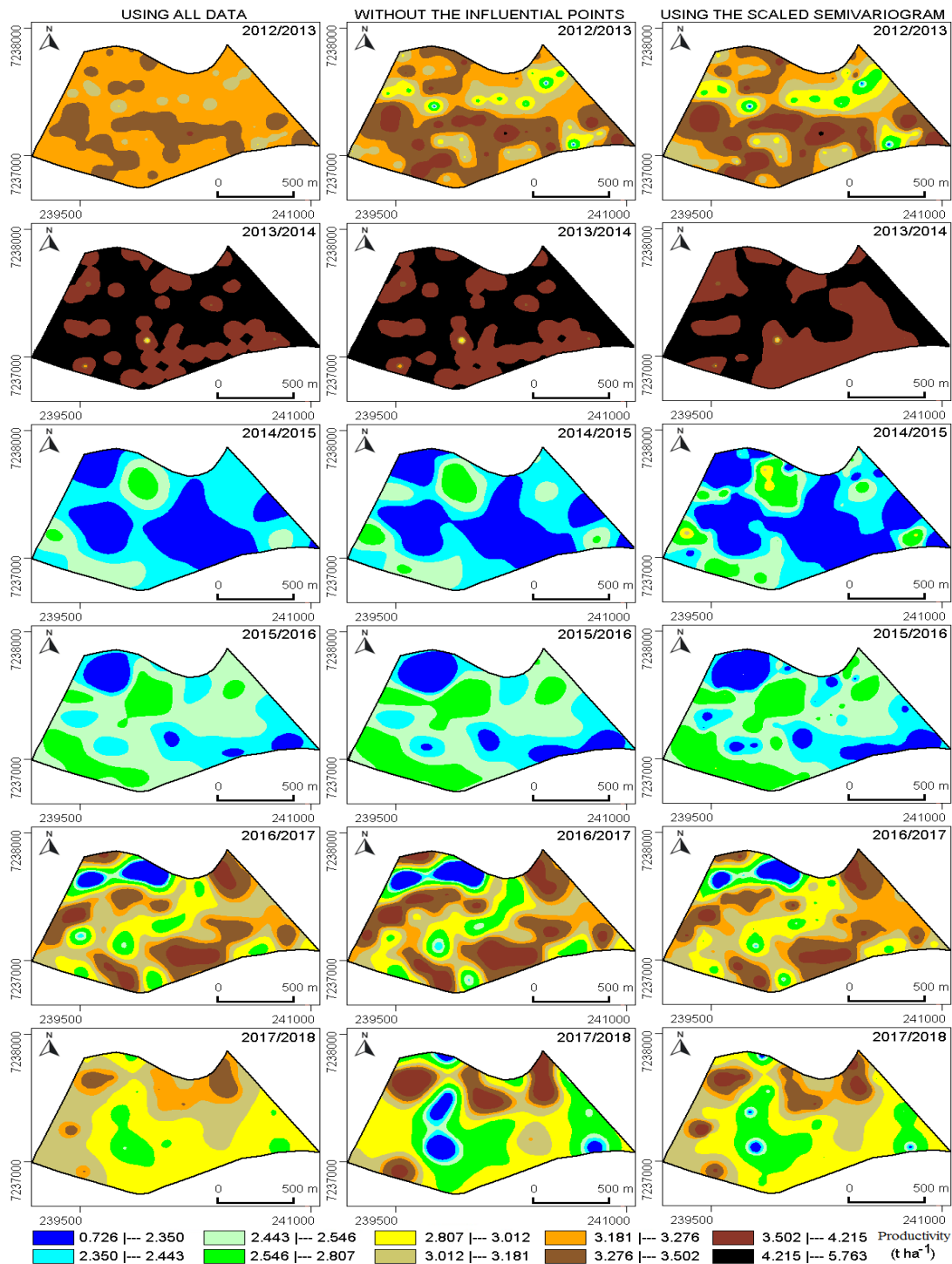


Figure 4. Maps of soybean productivity (t ha⁻¹) in the studied area

The differences found between the maps obtained with all points and maps without the influential points in 2012/2013, 2014/2015 to 2016/2017, and 2017/2018 allow us to state that the exclusion of influential points causes changes in the levels of soybean productivity. This fact corroborates the results obtained by Uribe-Opazo et al. (2012) in the elaboration of maps for soybean productivity. The authors concluded that the use of techniques to evaluate the influence of observations should be part of the geostatistical analysis.

The high similarity observed between the maps obtained without the influential points and the maps obtained by the scaled semivariogram indicates that using the scaled semivariogram is an alternative to obtain representative maps. A similar result was found by Wendpap et al. (2015). They obtained maps using scaled semivariograms classified by the Kappa index as presenting high similarity when compared to maps obtained by linear spatial models with replicates.

3.6 Spatial Linear Models with Replicates

A temporal comparison of the maps (Fig. 4) shows that some regions have similarities. For example, the northwestern region can be identified as with lower productivity, which is more evident from 2014/2015 to 2016/2017. The same occurs with the southeastern region, evidenced mainly in the maps obtained with the scaled semivariogram from 2013/2014 to 2015/2016. The similarities identified in the maps (Fig. 4) can be highlighted by fitting a linear spatial model considering the soybean productivity values from 2012/2013 to 2016/2017 as replicates. The cross-validation indicated the adjusted model as the best fit considering the exponential covariance function (M_{0.5}) with parameters $\hat{\beta} = 3.0762$, $\hat{\varphi}_1 = 0.0833$,

$\hat{\varphi}_2 = 0.4096$, $\hat{\varphi}_3 = 54.5580$ and a range of 163.44 m. This model has a strong spatial dependence, and the thematic map of soybean productivity associated with it (Fig. 5a) shows the northwest and southeastern regions as presenting the lowest productivity values. The areas between the west and southwestern regions also stand out for presenting the highest productivity values. However, in most of the areas (71.43%), the soybean productivity (t ha⁻¹) ranged between 3.012 t ha⁻¹ and 3.121 t ha⁻¹.

Because the thematic map of soybean productivity from 2012/2013 to 2016/2017 (Fig. 5a) presents the average behavior of productivity over a long period, we expect the possibility of defining zones of more reliable management. This is possible because, as explained by Fleming et al. (2000), considering the production history allows the identification of different management zones in the agricultural area. In practice, the northwest and southeastern regions had the lowest productivity values. Therefore, a different treatment in these regions tends to increase productivity in future harvests. It is worth mentioning that small regions with an indicator of lower productivity, scattered in the studied area (Fig. 5a), can hinder a localized application. As highlighted by Al-Kaisi et al. (2016), this is one of the problems related to maps that have small regions, resulting in difficulty in defining management zones for a localized application of fertilizers.

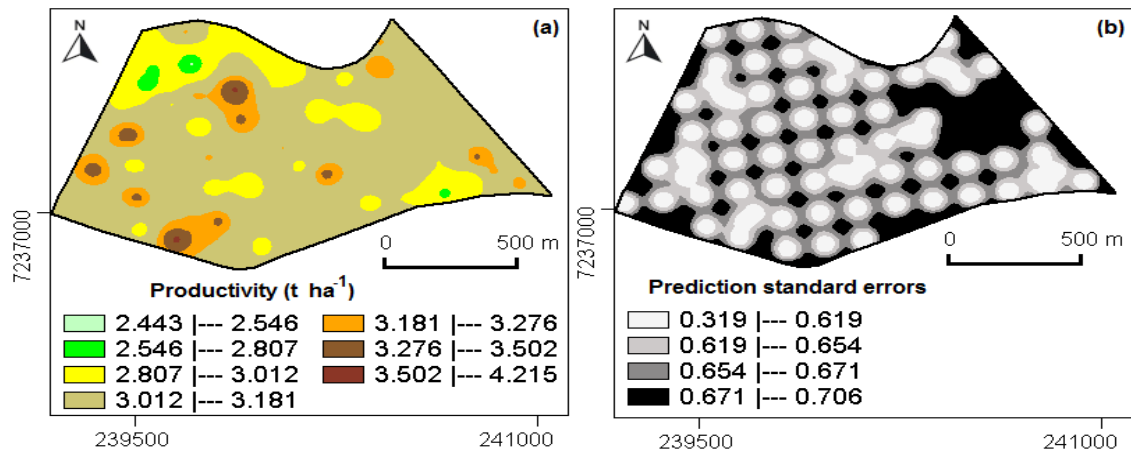


Figure 5. (a) Map of the average soybean productivity ($t\ ha^{-1}$) in the period between 2012/2013 and 2016/2017 and (b) its prediction standard error map

Fig. 5b shows the prediction standard error map, which quantifies the uncertainty related to the values predicted by kriging. The error values are lower close to the sampling points and higher in regions with no sampling points, such as the area highlighted in the east. Considering that the estimates are less accurate in locations distant from the sampling points, the soybean productivity map from 2017/2018 (Fig. 4) is less reliable than the thematic productivity maps of the other agricultural years since it is made with a lower number of sample points.

Finding that the prediction standard error is lower near the points of interpolation and that it gradually increases with the distance of the points (Fig 5b) is a characteristic of this type of map. Similar results were reported by Meusburguer et al. (2012) and Olea (2018).

3.6 Prediction Model

An alternative to obtaining more accurate productivity estimates in 2017/2018, in a S_i position within the studied area, is to develop a Gaussian spatial linear model considering soybean productivity information from previous agricultural years (2012/2013 to 2016/2017) as explanatory variables. After fitting the models, the cross-validation values indicated the spatial linear model as the best fit (Equation 2):

$$\text{Prod}_{17/18}(s) = 4.0740 + 0.1205\text{Prod}_{12/13}(s) - 0.1881\text{Prod}_{13/14}(s) + 0.2872\text{Prod}_{14/15}(s) - 0.3265\text{Prod}_{15/16}(s) - 0.1865\text{Prod}_{16/17}(s) \quad (2)$$

It was adjusted this model considering the Matérn covariance function with parameters $k = 2.0$, $\hat{\varphi}_1 = 0,1412$, $\hat{\varphi}_2 = 0,1426$, $\hat{\varphi}_3 = 113,7036$, and a range of 610.40 m. The structure has a moderate spatial dependence (SDI = 15.14%), and its spatial dependence radius is 87% greater than that of the model generated only with the sampling points collected in 2017/2018 (326.23 m). The soybean productivity map ($t\ ha^{-1}$) for this model (Fig. 6a) shows apparent differences when compared to the map generated only with the sampling points collected in 2017/2018 (Fig. 4).

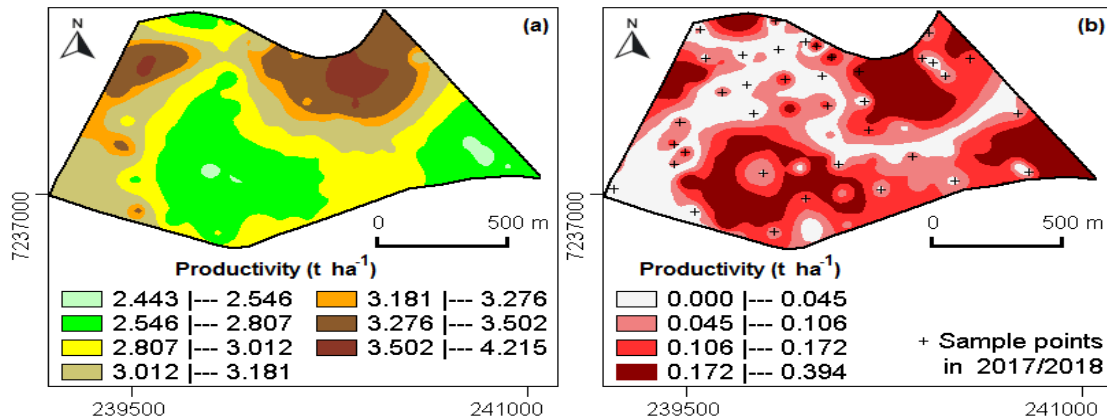


Figure 6. (a) Map of soybean productivity ($t\ ha^{-1}$) in 2017/2018 generated using KED and (b) Productivity differences ($t\ ha^{-1}$) in absolute values between the map considering all points in 2017/2018 (Fig. 4) and the map generated using KED (Fig. 6a)

Numerically, these differences are quantified by the Kappa index ($\bar{K} = 0.36$), which indicates a low similarity between the maps. Visually, the differences between both maps are highlighted in the map of the productivity differences ($t\ ha^{-1}$) in absolute values (Fig. 6b). The differences between both maps are lower than $0.106\ t\ ha^{-1}$ in most places where there were sample points in 2017/2018 (Fig. 6b). However, the differences are bigger (Fig. 6b), where data were not collected in 2017/2018 (highlighted in Fig. 1).

Comparing the predicting standard error maps (Fig.7a and Fig.7b), it is evident that the map obtained by the model considering the information from previous crop years (2012/2013 to 2016/2017) is more accurate given the lower standard errors.

The lower precision of the thematic map of soybean productivity ($t\ ha^{-1}$) in 2017/2018 (Fig. 4) is shown in the prediction standard error map (Fig. 7a). The reduction in the number of sample points (50% less) implied a considerable area in which the prediction errors are high (last class of the legend). In this sense, using a linear spatial model considering the productivity information of previous agricultural years as explanatory variables was useful for elaborating a more accurate thematic map. This was verified in the prediction standard error map of the soybean productivity ($t\ ha^{-1}$) in 2017/2018 generated using KED (Fig 7b).

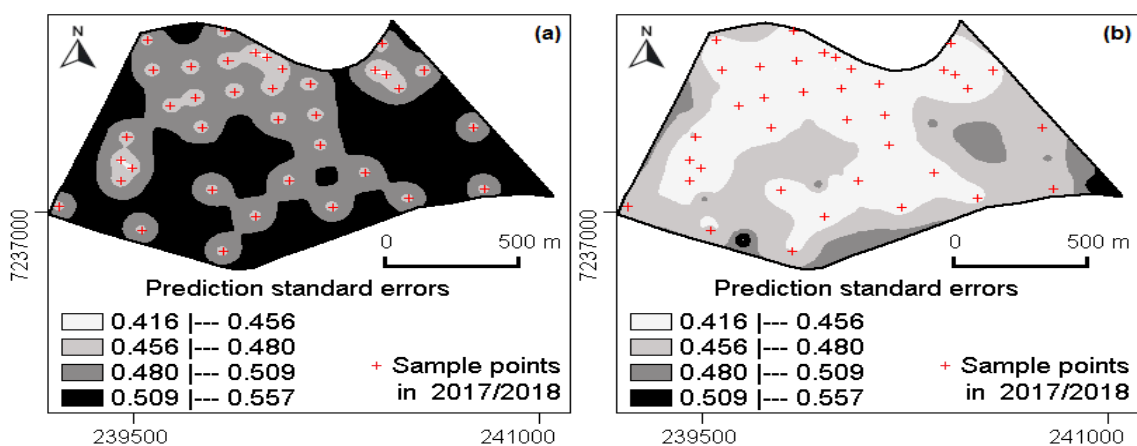


Figure 7. (a) Prediction standard error map of the soybean productivity ($t\ ha^{-1}$) considering all points in 2017/2018 (Fig. 4) and (b) Prediction standard error map of the soybean productivity ($t\ ha^{-1}$) in 2017/2018 generated using KED (Fig 6a)

4. Conclusion

The results showed that in the same agricultural area, soybean productivity could vary considerably, often due to uncontrollable factors such as those related to climatic events. It was found that the applied methodology was efficient in detecting influential points, and the removal of influential points affected the estimates of the parameters that define the spatial dependence structure and influence on the construction of the maps, which can promote changes in a thematic map. However, the use of scaled semivariograms allows us to obtain representative maps without excluding sample points. Although soybean productivity has shown a non-homogeneous behavior between crop years, we identified regions where productivity was lower (northwest and southeast of the studied area), which was shown in the map generated with replicates. The elaboration of the prediction standard error maps was essential to visually highlight that regions without sampling points present less precise values, showing that the use of explanatory provides predictions with a higher level of accuracy.

Acknowledgments

Council for Scientific and Technological Development (CNPq); Coordination for the Improvement of Higher Education Personnel (CAPES); Post-Graduate Program in Agricultural Engineering (PGEAGRI); Federal Technological University of Paraná (UTFPR) and Spatial Statistics Laboratory (LEE/UNIOESTE-LISA).

References

- Al-Kaisi, M. M., Archontoulis, S., & Kwaw-Mensah, D. (2016). Soybean spatiotemporal yield and economic variability as affected by tillage and crop rotation. *Journal of the American Society of Agronomy*, 108(3), 1267-1280. <https://doi.org/10.2134/agronj2015.0363>
- Bakhsh, A., Jaynes, D. B., Colvin, T. S., & Kanwar, R. S. (2000). Spatio-Temporal Analysis of Yield Variability for a Corn-Soybean Field in Iowa. *Transactions of the American Society of Agricultural Engineers*, 43(1), 31-38. <https://doi.org/10.13031/2013.2684>
- Bottega, E. L., Queiroz, D. M., Pinto, F. A. C., Valente, D. S. M., & Souza, C. M. A. (2017). Precision agriculture applied to soybean: Part III - Spatial and temporal variability of yield. *Australian Journal of Crop Science*, 11(07), 799-805. <https://doi.org/10.21475/ajcs.17.11.07.pne383>
- Chipeta, M. G., Terlouw, D. J., Phiri, K. S., & Diggle P. J. (2017). Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure. *Environmetrics*, 28(1), e2425. <https://doi.org/10.1002/env.2425>
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- CONAB, Companhia Nacional de Abastecimento. (2019). *Soybeans - Brazil: Historical series of productivity*. <http://www.conab.gov.br>. Accessed 22 Dezember 2019.
- Cook, R. D. (1986). Assessment of Local Influence. *Journal of the Royal Statistical Society Series B*. 48(2),

133-169. <https://doi.org/10.1111/j.2517-6161.1986.tb01398.x>

Costa, N. L., & Santana, A. (2018). Analysis of the Soybean Market: Conjunctural Aspects of the Formation of the Price Paid to the Brazilian Producer. *Revista Plantio Direto*, 28, 20-39.

Dalposso, G. H., Uribe-Opazo, M. A., & Johann, J. A. (2016). Soybean yield modeling using bootstrap methods for small samples. *Spanish Journal of Agricultural Research*, 14(3), e0207. <https://doi.org/10.5424/sjar/2016143-8635>

Dalposso, G. H., Uribe-Opazo, M. A., Johann, J. A., De Bastiani, F., & Galea M. (2019). Geostatistical modeling of soybean yield and soil chemical attributes using spatial bootstrap. *Engenharia Agrícola*, 39(3), 350-357. <https://doi.org/10.1590/1809-4430-Eng.Agric.v39n3p350-357/2019>

Dalposso, G. H., Uribe-Opazo, M. A., Johann, J. A., Galea, M., & De Bastiani, F. (2018). Gaussian spatial linear model of soybean yield using bootstrap methods. *Engenharia Agrícola*, 38(1), 110-116. <https://doi.org/10.1590/1809-4430-eng.agric.v38n1p110-116/2018>

Dalposso, G. H., Uribe-Opazo, M. A., Mercante, E., Johann, J. A., & Borssoi, J. A. (2012). Comparison measures of maps generated by geostatistical methods. *Engenharia Agrícola*, 32(1), 174-183. <https://doi.org/10.1590/S0100-69162012000100018>

De Bastiani, F., de Aquino Cysneiros, A. H. M., Uribe-Opazo, M. A., & Galea M. (2015). Influence diagnostics in elliptical spatial linear models. *Test*, 24, 322-340. <https://doi.org/10.1007/s11749-014-0409-z>

Bastiani, F., Galea, M., Cysneiros A. H. M. A., & Uribe-Opazo, M.A. (2017). Gaussian spatial linear models with repetitions: An application to soybean productivity. *Spatial Statistics*, 21(A), 319-335. <https://doi.org/10.1016/j.spasta.2017.07.013>

EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária. (2013). *Sistema brasileiro de classificação de solos*. Rio de Janeiro, RJ: Centro Nacional de Pesquisa de Solos.

Faraco, M. A., Uribe-Opazo, M. A., Silva, E. A. A., Johann, J. A., & Borssoi, J. (2008). Selection criteria of spatial variability models used in thematic maps of soil physical attributes and soybean yield. *Revista Brasileira de Ciência do Solo*, 32(2), 463-476. <https://doi.org/10.1590/S0100-06832008000200001>

Fleming, K. L., Westfall, D. G., Wiens, D. W., & Brodahl, M. C. (2000). Evaluating farmer defined management zone maps for variable rate fertilizer application. *Precision Agriculture*, 2, 201-215. <https://doi.org/10.1023/A:1011481832064>

Grzegozewski, D. M., Uribe-Opazo, M. A., De Bastiani, F., & Galea M. (2013). Local influence when fitting Gaussian spatial linear models: an agriculture application. *Ciencia e Investigacion Agraria*, 40(3), 537-545. <http://dx.doi.org/10.4067/S0718-16202013000300006>

Hengl, T., Geuvelink, G. B. M., & Stein, A. (2003). *Comparison of kriging with external drift and regression-kriging*. ITC. Technical note. https://webapps.itc.utwente.nl/librarywww/papers_2003/misca/hengl_comparison.pdf Accessed 25 June 2020.

Kyriakidis, P. C., & Goodchild, M. F. (2006). On the prediction error variance of three common spatial interpolation schemes. *International Journal of Geographical Information Science*, 20(8), 823-855. <https://doi.org/10.1080/13658810600711279>

Lamour, J., Naud, O., Lechaudel, M., Mogueédec, G. L., Taylor, J., & Tisseyre, B. (2020). Spatial analysis

and mapping of banana crop properties: issues of the asynchronicity of the banana production and proposition of a statistical method to take it into account. *Precision Agriculture*, 21, 897-921. <https://doi.org/10.1007/s11119-019-09700-7>

Landis, J. R., & Kock, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-75. <https://doi.org/10.2307/2529310>

Lloyd, C. D., & Atkinson, P. M. (2001). Assessing uncertainty in estimates with ordinary and indicator kriging. *Computers & Geosciences*, 27(8), 929-937. [https://doi.org/10.1016/S0098-3004\(00\)001321](https://doi.org/10.1016/S0098-3004(00)001321)

Mardia, K. V., & Marshall, R. J. (1984). Maximum likelihood models for residual covariance in special regression. *Biometrika*, 71(1), 319-332. <https://doi.org/10.2307/2336405>

Matérn, B. (1986). *Spatial Variation*. Lecture notes in Statistics. New York, NY: Springer.

Menezes, M. D., Silva, S. H. G., Mello, C. R., Owens, P. R., & Curi, N. (2016). Spatial prediction of soil properties in two contrasting physiographic regions in Brazil. *Scientia Agricola*, 73(3), 274-285. <http://dx.doi.org/10.1590/0103-9016-2015-0071>

Mercante, E., Uribe-Opazo, M. A., & Souza, E. G. (2003). Spatial and temporal variability of the soil mechanical resistance to penetration in areas with and without site-specific chemical application. *Revista Brasileira de Ciência do Solo*, 27(6), 1149-1159. <https://doi.org/10.1590/S0100-06832003000600019>

Meusburger, K., Steel, A., Panagos, P., Montanarella, L., Alewell, C. (2012). Spatial and temporal variability of rainfall erosivity factor for Switzerland. *Hydrology and Earth System Sciences*, 16, 167-177. <https://doi.org/10.5194/hess-16-167-2012>

Microsoft R Core Team, (2021). *Microsoft R Open*. Microsoft, Redmond, WA. <https://mran.microsoft.com>. Accessed 11 August 2021.

Olea, R. A. (2018). *A practical primer on geostatistics* (ver. 1.4, December 2018): U.S. Geological Survey Open-File Report 2009-1103.

Peña, D., & Yohai, V. (1999). A fast procedure for outlier diagnostics in large regression problems. *Journal of the American Statistical Association*, 94(446), 434-445.

Poon, W., & Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society, Series B (Methodology)*, 61(1), 51-61. <https://doi.org/10.1111/1467-9868.00162>

Ribeiro Jr., P. J., & Diggle, P. J. (2001). *geoR: A package for geostatistical analysis*. R-News, 1(2), 15-18.

Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P. V. V., & Ciampitti, I. A. (2020). Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology*, 284, 107886. <https://doi.org/10.1016/j.agrformet.2019.107886>

SEAB, Secretaria da Agricultura e do Abastecimento. (2016). *Soja – Informativo 01 de Agosto – (2016)*. <http://www.agricultura.pr.gov.br>. Accessed 17 January 2019.

Seidel, E. J., & Oliveira, M. S. (2014). New geostatistical index for measuring spatial dependence. *Revista Brasileira de Ciência do Solo*, 38, 699-705. <http://dx.doi.org/10.1590/S0100-06832014000300002>

Seidel, E.J., & Oliveira, M.S. (2016). A Classification for a Geostatistical Index of Spatial Dependence. *Revista Brasileira de Ciência do Solo*, 40, e0160007. <http://dx.doi.org/10.1590/18069657rbc20160007>

Silva, R. A., Siqueira, G. M., Costa, M. K. L., Filho, O. G., & Silva, E. F. (2018). Spatial Variability of Soil Fauna Under Different Land Use and Managements. *Revista Brasileira de Ciência do Solo*, 42, e0170121. <https://doi.org/10.1590/18069657rbcS20170121>

Syed, N., Mahesar, S. A., Sherazi, S. T. H., & Soylak, M. (2019). Quality assessment and safety measurement of different industrial processing stages of soybean oil. *Turkish Journal of Agriculture: Food Science and Technology*, 1(2), 28-33. <http://doi.org/10.14744/turkjfas.2019.006>

Tóth, G., Bodai, Z., & Héberger, K. (2013). Estimation of influential points in any data set from coefficient of determination and its leave-one-out cross-validated counterpart. *Journal of Computer-Aided Molecular Design*, 27(10), 837-844. <https://doi.org/10.1007/s10822-013-9680-4>

Uribe-Opazo, M. A., Borsoi, J. A., & Galea, M. (2012) Influence Diagnostics in Gaussian Spatial Linear Models. *Journal of Applied Statistics*, 39(3), 615-630. <https://doi.org/10.1080/02664763.2011.607802>

Uribe-Opazo, M. A., De Bastiani, F., Galea, M., Schemmer, R. C., & Assumpção, R. A. B. (2021). Influence diagnostics on a reparameterized t-Student spatial linear model. *Spatial Statistics*, 41, 100481. <https://doi.org/10.1016/j.spasta.2020.100481>

Vatsanidou, A., Nanos, G. D., Fontas, S., Baras, J., Castrignano, A., & Gemtos T. A. (2017). Nitrogen replenishment using variable rate application technique in a small hand-harvested pear orchard. *Spanish Journal of Agricultural Research*, 15(4), e0209. <https://doi.org/10.5424/sjar/2017154-10986>

Vega, A., Córdoba, M., Castro-Franco, M., & Balzarini, M. (2019). Protocol for automating error removal from yield maps. *Precision Agriculture*, 20, 1030-1044. <https://doi.org/10.1007/s11119-018-09632-8>

Vieira S. R., & Gonzalez, A. P. (2003). Analysis of the spatial variability of crop yield and soil properties in small agricultural plots. *Bragantia*, 62(1), 127-138. <https://doi.org/10.1590/S0006-87052003000100016>

Vieira, S. R., Garcia, M. A. G., González, A. P., & Siqueira, G. M. (2010). Spatial and temporal variability of soil water content in two land uses. *Bragantia*, 69(1), 181-190. <https://doi.org/10.1590/S0006-87052010000100023>

Wackernagel, H. (1995). *Multivariate geostatistics*. Berlin, GER: Springer-Verlag

Webster, R., Oliver, M.A. (2007). *Geostatistics for Environmental Scientists (Statistics in Practice)*. Chichester, UK: John Wiley and Sons.

Wendpap, B. G., Uribe-Opazo, M. A., Johann, J. A., Vilas Boas, M. A., & Souza, E. G. (2015). Geostatistical models in a study on the spatio-temporal dependence of agricultural data. *International Journal of Agriculture, Environment and Food Sciences*, 13(3&4), 21-28. <https://doi.org/10.1234/4.2015.4034>

Zhu, H., Ibrahim, J. G., Lee, S., & Zhang, H. (2007). Perturbation selection and influence measures in local influence analysis. *Annals of Statistics*, 35(6), 2565-2588. <https://doi.org/10.1214/009053607000000343>

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).