

Doi: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v39nep85-95/2019>

Special Issue: Precision Agriculture

SPATIAL DEPENDENCE DEGREE AND SAMPLING NEIGHBORHOOD INFLUENCE ON INTERPOLATION PROCESS FOR FERTILIZER PRESCRIPTION MAPS

Lucas R. do Amaral^{1*}, Diego D. Della Justina²

^{1*}Corresponding author. University of Campinas, School of Agricultural Engineering/ Campinas - SP, Brazil.
E-mail: lucas.amaral@feagri.unicamp.br | ORCID ID: <http://orcid.org/0000-0001-8071-4449>

KEYWORDS

data interpolation,
soil sampling,
geostatistics, site-
specific management.

ABSTRACT

Data interpolation is widely required in precision agriculture. However, its effectiveness is a function of the characteristics of the dataset, being necessary for the evaluation of several parameters. This study aimed to identify how the spatial interpolators, Kriging, and Inverse Distance Weighting, are influenced by the degree of spatial dependence of the variables analyzed and the number of neighbors considered in the interpolation process (sampling neighborhood). Soil samples were collected from three sugarcane fields. By the optimization process, we verified that the sampling neighborhood influences the accuracy of interpolations, but there is not a standard recommendation to follow. Thus, the best sampling neighborhood must ever be optimized for each case when preparing fertilizer prescription maps. Evaluating the performance of interpolations is always important to infer the reliability of the prescription maps, since no index that measures the degree of spatial dependence is effective. Because high prediction errors can occur when spatial dependence is poorly modeled, one cannot expect crop response with the continuous application of fertilizers in variable rates. Thus, work with homogeneous soil zones can be an interesting palliative approach. This study guides precision agriculture practitioners on some points that should be carefully considered in the data interpolation process for generating fertilizer prescription maps.

INTRODUCTION

Precision agriculture is a management approach that aims to address spatial and temporal variabilities of production factors in agricultural fields (Banu, 2015). Among the site-specific interventions enabled by precision agriculture, one of the most widely used is the variable-rate fertilizer application. This technique presupposes that soil fertility varies within fields; therefore, the use of fertilizers can be optimized with its localized application (Mueller et al., 2004).

Spatial variability of soil fertility is routinely surveyed through grid soil sampling (Mallarino & Wittry, 2004). However, the farmer's capacity in employing adequate sampling density to the whole field is limited since soil sampling is laborious and laboratory analyses are expensive. This often leads to inappropriate number and disposition of samples being unsuitable for representing the variability of soil properties intended to be mapped

(Mueller et al., 2004; Nanni et al., 2011; Siqueira et al., 2014). Consequently, the return on soil fertility mapping and variable-rate fertilizer application is impaired.

With the results of soil analyses in hand, these data must be interpolated to obtain continuous maps that can be interpreted and converted into fertilizer rates. The interpolation method that offers the best performance is admittedly the kriging method (Diggle & Ribeiro, 2007). However, for the correct performance of this technique, the analyst must pay attention to several issues of spatial dependence modeling (Oliver & Webster, 2014; Vieira et al., 2010) before effectively obtaining an interpolated map, which often causes this interpolation method to be improperly used or not employed at all. There are several deterministic interpolators alternative to kriging. The most widely used is a weighted average calculation as a function of the distance between samples, the so-called inverse distance weighting (IDW) (Li & Heap, 2011) since it

² University of Campinas, School of Agricultural Engineering/ Campinas - SP, Brazil.
Area Editor: Fabio Henrique Rojo Baio
Received in: 4-9-2019
Accepted in: 6-13-2019



performs well for variables with different behaviors and is easily implemented in software programs and computational routines (Betzek et al., 2019). However, such deterministic interpolators do not take into account the spatial dependence and specific behavior of the data, which may lead to lower efficiency in mapping the spatial distribution of a given variable when compared to kriging (stochastic interpolator) (Betzek et al., 2019). Moreover, Mueller et al. (2004) state that reliable spatial predictions are dependent on the spatial autocorrelation between samples; thus, it is important to understand how the degree of spatial dependence of a given soil property influences the interpolation performance. Nonetheless, there have been many conflicting reports concerning the efficiency of interpolation procedures (Robinson & Metternicht, 2006), mainly because soil fertility variability may present a high level of non-spatial structure (random component) due to anthropic influence (i.e., crop management throughout the seasons). Thus, it is important to determine whether kriging can, in fact, yield higher accuracy in creating prescription maps in comparison to simpler interpolation methods (e.g., IDW).

In addition, the interpolation process is a complex procedure in which the appropriate adjustment of some settings may afford gains in accuracy. In this sense, one of the parameters that influence the quality of the predictions is the local neighborhood surrounding the unsampled

locations, i.e., the number of neighborhood samples taken into account in the interpolation process (here called the sampling neighborhood) as demonstrated by Robinson & Metternicht (2006) and Vieira et al. (2010).

Thus, our main goal was to identify how kriging and IDW are influenced by the number of observations considered in this process (sampling neighborhood) as well as the degree of spatial dependence. Moreover, this study aims to provide important information to practitioners of precision agriculture about the proper procedures to interpolate data and generate fertilizer prescription maps.

MATERIAL AND METHODS

This study comprised three fields cultivated with sugarcane within São Paulo State, Brazil, each with different levels of spatial dependence of potassium availability in the soil ($\text{mmol}_c \text{dm}^{-3}$) (Table 1). Data were interpolated using IDW and ordinary kriging methods (Li & Heap, 2011) to optimize the sampling neighborhood and improve prediction accuracy. Such interpolation efficiency was related to spatial dependence indices proposed in the literature in order to investigate the expected quality of the map prior to interpolation. Two sampling densities were tested to evaluate their impact regarding interpolation procedures and sampling neighborhoods. Fertilizer prescription maps were created to express the results of the above-mentioned variables.

TABLE 1. Characteristics of the study sites, samplings performed and soil analysis.

Field	Area (ha)	Number of soil samples ¹	Subsamples for each sample point ²	Sampling density ³ (ha sample^{-1})	Sampled depth (cm)	Average potassium soil availability (CV%)
1	115.7	114	6	1.0	0 – 40	1.01 $\text{mmol}_c/\text{dm}^3$ (79.2%)
2	88.4	95	8	1.0	0 – 20	3.61 $\text{mmol}_c/\text{dm}^3$ (35.2%)
3	98.6	194	3	0.5	0 – 20	0.9 $\text{mmol}_c/\text{dm}^3$ (74.7%)

¹ Number of samples after removal of outliers; ² number of soil subsamples collected within 5 m of the center point with an auger; ³ Sampling density originally collected.

The dataset of each field was divided into calibration and validation sets (Table 2). About 90% of the samples were retained for calibration (data to be interpolated), whereas the remaining 10% were used as an external validation set (Fig. 1) to quantify the error of the estimated potassium fertilizer rate. Samples of the validation set were selected to ensure the spatial representativeness of each field. The selection of validation samples was performed based on the spatial distribution of points generated from the function *optimMSSD*, available in the *spsann* package (Samuel-Rosa, 2017). Such functions

aim to minimize the average squared distance between sampling and prediction locations to ensure spatial representativeness of the external validation set. Before dataset division, outliers were removed in order to avoid their influence on interpolation results. Samples were removed according to the criterion of Anselin Local Moran's I (Fu et al., 2016) through the function '*Optimized Outlier Analysis*' available in the Arcmap software (ArcGIS ESRI), which identifies samples with low values within a region where the neighborhood shows high values and vice versa.

TABLE 2. Number of samples and sampling density (ha samples^{-1}) according to the dataset (dense, sparse, and external validation samples) and number of outliers removed in each field.

Field	No. of samples in dense sampling	Final sampling density (ha sample^{-1})	No. of samples in sparse sampling	Final sampling density (ha sample^{-1})	No. of samples in external validation	No. of samples removed as outliers ¹
1	97	1.19	26	4.13	17	4
2	80	1.11	21	4.02	15	1
3	165	0.60	35	2.99	29	3

¹ outliers removed according to the criterion of Anselin Local Moran's I (Fu et al., 2016)

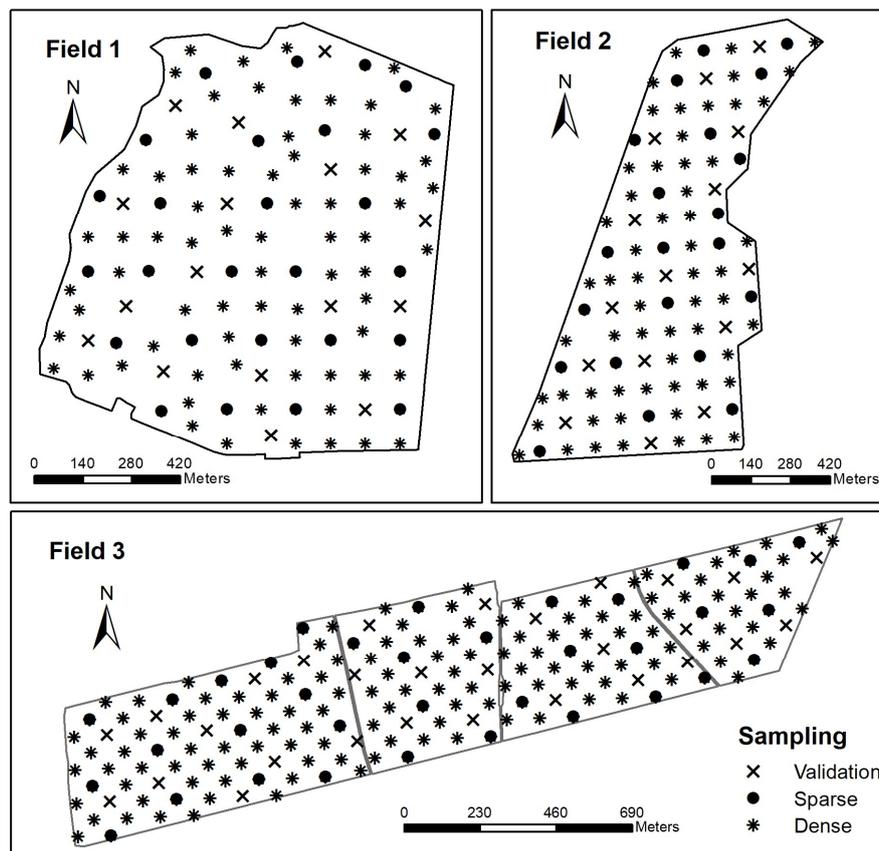


FIGURE 1. Experimental fields, showing the disposition of the samples according to the datasets (dense and sparse sampling and external validation samples).

From the calibration set, two sampling densities were tested (Fig. 1). The first was named “dense sampling” in which all data from the calibration set were maintained (Table 2). In the second set, only a quarter of the samples were kept and named “sparse sampling” (Table 2), which aimed to simulate the most common situation in practice (Siqueira et al., 2014), when the high cost prevents denser sampling.

Potassium prescription maps were interpolated by both ordinary kriging and IDW methods. For kriging, we conducted semivariogram modeling for all scenarios studied (Table 2) using the spatial dependence modeling practices recommended by Oliver & Webster (2014). Thus, we evaluated the presence of spatial trends in the dataset, adopted the lag distance corresponding approximately to the average distance between samples, and tested which model should be used (spherical, exponential or Gaussian) through cross-validation. The Gaussian model did not adjust to any of the datasets.

In addition, another parameter that deserves to be tested by cross-validation is the sampling neighborhood used for interpolation (i.e., neighborhood size or the number of neighbors) (Robinson & Metternicht, 2006; Vieira et al., 2010). Thus, a grid search was conducted to determine the optimal number of neighbors for each interpolation method (IDW and kriging). We tested combinations with a minimum of four neighbors until the total number of samples (all samples) for each field, minus one; interval of three neighbors was used. The combinations of minimum and maximum numbers of neighbors were evaluated by leave-one-out cross-validation using the calibration dataset.

The combination of neighbors that yielded the highest coefficient of determination (R^2) and the lowest error (root mean square error (RMSE)) was selected.

Moreover, we tested IDW because it is widely used by precision agriculture practitioners. Such deterministic interpolator predictions are proportional to the inverse of the distance raised to a power value p (Li & Heap, 2011). We investigated the effect of the p value by evaluating IDW interpolations using three integer values for p ($p=1$, $p=2$, and $p=3$ for IDW_{p1} , IDW_{p2} , and IDW_{p3} , respectively). We also performed a neighborhood search for IDW by leave-one-out cross-validation to assess the best combination of neighbors for interpolation.

We contrasted the performance of both kriging and IDW, optimized for the sampling neighborhood with IDW interpolation, without performing a neighborhood search (all samples were considered). Predictions weighted by the square of the distance (IDW_{p2}) were considered as the null model (IDW_{null}). This setup aimed to compare what is usually done by practitioners since the null model is the most implemented setup present in interpolation software.

The final evaluation of interpolations was performed using the independent dataset (i.e., external validation samples). This procedure allowed the calculation of the ratio of percent deviation (RPD) - a statistic that measures the suitability of the generated models and the quality of the predictions obtained. According to the classification proposed by Viscarra Rossel et al. (2006), the RPD can be subdivided: $RPD < 1.0$ indicates a very poor model and its use is not recommended; RPD between 1.0 and 1.4 indicates a poor model in which only high and low values are

distinguishable; RPD between 1.4 and 1.8 indicates a regular model that allows its use for interferences and correlations; RPD between 1.8 and 2.0 indicates a good model, which provides quantitative predictions; RPD between 2.0 and 2.5 indicates a very good model for quantifications; and RPD > 2.5 indicates a great model for quantification.

In order to compare the prediction errors in terms of the potassium fertilizer amount (kg ha^{-1}), we defined an equation for fertilizer prescription (Equation 1) from the fertilizer recommendation table for sugarcane planting available in the “Fertilization and Liming Recommendation for the São Paulo State” (van Raij et al., 1996), adopting an expected productivity between 100 and 150 t ha^{-1} .

$$K_{\text{rate}} = -23.781 \times K_{\text{avail.}} + 150.25 \quad (1)$$

Where:

K_{rate} is the potassium rate to be applied in kg ha^{-1} , and

$K_{\text{avail.}}$ is the potassium availability in soil in mmolc dm^{-3} .

Furthermore, we calculated the Moran's index (MI) of the samples in the calibration sets to verify the clustering level of the data (spatial autocorrelation, Driemeier et al., 2016). The MI was computed using the Moran test function available in the ‘*spdep*’ package of the R statistical environment (Bivand & Wong, 2018). The aim here was to relate the interpolation performance with this index, which can be calculated based on both the position and values of the soil samples (i.e., information obtained previous to interpolation) since high spatial autocorrelation leads to high prediction accuracy (Mueller et al., 2004). Thus, practitioners could obtain a prior inference about the interpolation quality to determine whether variable-rate fertilizer applications will bring the expected return on investment.

In addition, we used two other methods related to spatial autocorrelation that can be used to infer the expected interpolation quality prior to actually performing the procedure. One is the degree of spatial dependence (DSD) proposed by Cambardella et al. (1994), which is widely used in literature and classifies the spatial dependence as high, medium, or low depending on the proportion of the nugget to sill effect of the modeled semivariogram. However, this index is dependent of the semivariogram model and does not consider all aspects of the semivariogram geometry (Seidel & Oliveira, 2016), which can compromise comparisons among different datasets. Thus, we also used the index of spatial dependence (ISD) recently proposed by Seidel & Oliveira (2016), which aims to overcome the limitations observed by the DSD of Cambardella et al. (1994), when taking into account more parameters instead of only sill and nugget effect; this index also considers the semivariogram model, the range obtained from it, and the maximum distance between samples from a given dataset.

RESULTS AND DISCUSSION

Both spatial interpolation methods, kriging and IDW, can produce equivalent results in the representation of soil spatial variability within the scope of precision agriculture (Table 3, external validation). This contradicts what several researchers indicated when stating that kriging was the most accurate prediction method (Mueller et al., 2004; Oliver & Webster, 2014). We understand that this occurs in agricultural fields because there is a high amount of anthropic influence on the spatial structure of soil fertility properties; assuming that observations close to each other are more alike than those that are far apart (Goovaerts, 1999) makes the spatial dependence less significant. Thus, interpolation accuracy is impaired, suggesting an equivalence between the methods since the quality of spatial predictions are dependent on the spatial autocorrelation between samples (Mueller et al., 2004). Betzek et al. (2019) recently found similar results working with interpolation optimization of chemical soil attributes and yield data for denser samplings (>3.5 points ha^{-1}). However, in such studies, the authors did not follow the procedures for variogram modeling proposed by Oliver & Webster (2014) since they were developing an automatic procedure for such analysis. This could be the reason why IDW was superior to kriging in some situations in their study. Nevertheless, here we showed that the equivalent accuracy between kriging and IDW interpolation may indeed appear on data that present a low degree of spatial dependence. When such low spatial dependence occurs, caused by anthropic influence on agricultural fields, kriging smooths the data distribution, impairing its prediction efficiency (Yamamoto, 2005). Thus, an IDW interpolator is faithful to the value of the samples, and the predictions can achieve similar prediction accuracy or be more assertive.

To obtain the best possible result of interpolations, no matter the method, the analyst has to consider as many factors as possible (Vieira et al., 2010). In this sense, the number of neighbors (sampling neighborhood) is an important parameter to be considered, and its optimization allows significant improvement of interpolation performance. Such importance can be noted when comparing the lower performance of IDW_{null} in predicting K fertilizer rates compared to the other optimized interpolations (kriging and IDW_{p1}) (Table 3). Note that the slope of the relationship between predicted and measured fertilizer rates has improved considerably (it is closer to one) for most models, indicating it to be an important factor to be considered when seeking the correct quantification of an attribute (Vieira et al., 2010). Thus, we suggest that future studies take this factor into consideration, instead of only R^2 and RMSE.

TABLE 3. Performance of interpolation methods in predicting potassium fertilizer rates for external validation samples. We used the best sampling neighborhood indicated by the optimization process (kriging and IDW_{p1}) and the IDW null model (IDW_{null}). The interpolation methods with the best performance for each dataset are indicated in bold and take into account the coefficient of determination, RMSE (kg ha⁻¹) and the slope of the relationship between predicted and measured fertilizer rates.

Interpolation	R ²	RMSE	Slope	R ²	RMSE	Slope
	Field 1 – Dense sampling			Field 1 – Sparse sampling		
Kriging	0.62	11.15	0.70	0.53	12.38	0.54
IDW _{p1}	0.63	11.03	0.73	0.54	12.20	0.46
IDW _{null}	0.60	11.73	0.47	0.46	13.28	0.38
	Field 2 – Dense sampling			Field 2 – Sparse sampling		
Kriging	0.62	17.64	0.82	0.44	19.98	0.49
IDW _{p1}	0.59	18.27	0.71	0.47	18.78	0.57
IDW _{null}	0.61	17.76	0.46	0.41	20.05	0.47
	Field 3 – Dense sampling			Field 2 – Sparse sampling		
Kriging	0.01	16.53	0.04	0.00	19.25	0.04
IDW _{p1}	0.01	16.58	0.05	0.01	17.55	0.04
IDW _{null}	0.00	16.12	0.02	0.00	17.56	-0.01

Despite all variation in the performance of interpolations, we chose to select the spherical model to fit the theoretical semivariogram, followed by kriging interpolation, as well as the use of the IDW_{p1} in the sequence of the analyses in this study. The spherical model presented superior performance in most scenarios and, when it did not, its results were very close to the exponential model (*data not presented*). The same occurred with the IDW_{p1}, which had a scenario with a slightly lower performance than IDW_{p2} which is about 0.01 mmol_c dm⁻³ (RMSE) higher (*data not presented*). Thus, we ensured that there was no other source of variation in our results.

The sparser the soil sampling, the lower the reliability of data interpolation (Table 4). Although Li & Heap (2011) did not find a reduction in interpolation accuracy when working with a lower sampling density, they were dealing with mining and environmental data, which are less susceptible to human interference. Thus, the higher

the anthropic influence, the lower the spatial dependence, which impairs interpolation performance. Also, the harder it was to adjust the semivariogram (Fig. 2), the worse the performance of predictions (Table 4). This difficulty was more pronounced when working with datasets with large distances between samples (i.e., sparse sampling). When this happens, the semivariogram modeling process becomes more dependent on the analyst's experience; this is noted in Figure 2 in which the lags of the experimental semivariograms are larger and exhibit greater variation (inconsistent behavior). This increases the influence on the decision-making of the analyst responsible for the semivariogram analysis and, as a consequence, may jeopardize the final quality of the kriging process. Thus, a fine adjustment in lag size is necessary in the case of poor sampling distribution, which influences the parameters obtained (nugget effect, sill, and range).

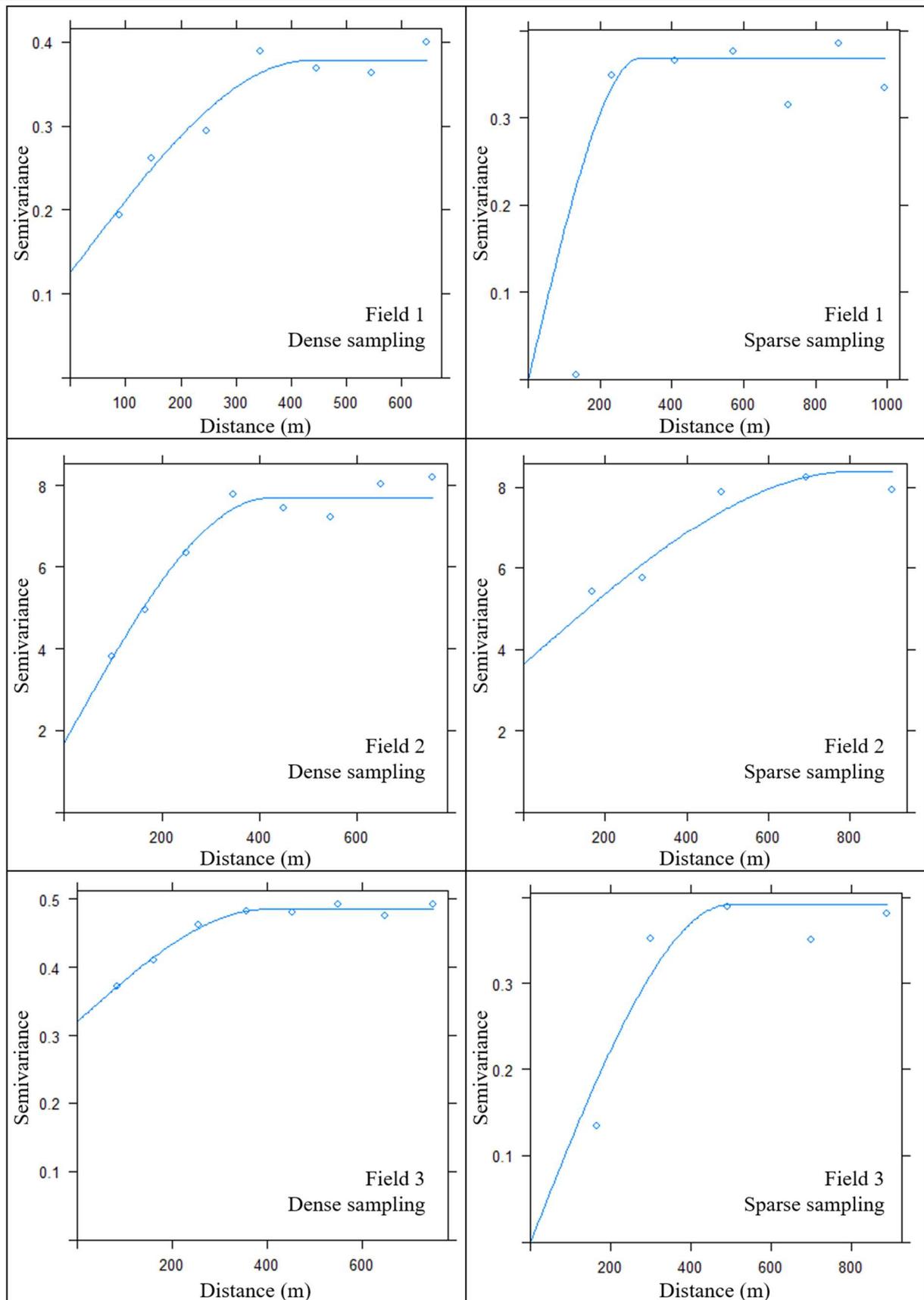


FIGURE 2. Semivariograms obtained with dense (left) and sparse (right) sampling for each field, showing the inconsistency of the lag values and poor model adjustment when the sparse samplings were evaluated.

TABLE 4. The best configuration of sampling neighborhood and interpolation process for each dataset (three fields and two sampling densities) according to cross-validation results. Such configurations were used in all other interpolations carried out in this study (external validation and map preparation).

Interpolation	Sampling	Model ¹	n.min ²	n.max ³	R ²	RMSE ⁴
Field 1						
Kriging	dense	spherical	4	7	0.67	0.46
IDW	dense	power = 1	4	4	0.68	0.46
Kriging	sparse	spherical	4	7	0.23	0.67
IDW	sparse	power = 1	4	4	0.16	0.68
Field 2						
Kriging	dense	spherical	4	64	0.35	2.11
IDW	dense	power = 1	4	7	0.36	2.08
Kriging	sparse	spherical	4	7	0.27	2.43
IDW	sparse	power = 1	4	4	0.20	2.42
Field 3						
Kriging	dense	spherical	4	58	0.16	0.61
IDW	dense	power = 1	4	13	0.15	0.61
Kriging	sparse	spherical	4	10	0.30	0.48
IDW	sparse	power = 1	4	4	0.20	0.50

¹ Model used: kriging = spherical; IDW = inverse distance weighting power equal to 1 (IDW_{p1}). ² and ³ are the minimum and maximum number of neighbors used in interpolation, respectively. ⁴ root mean squared error (mmol. dm.⁻³) between values measured and predicted by interpolation.

An alternative to this difficulty and subjectivity intrinsic to the semivariogram modeling, enforced by the method of moments, would be to fit the semivariogram function using the restricted maximum likelihood method (REML), which allows one to infer the best model by evaluating the data as a whole (Diggle & Ribeiro, 2007); therefore, avoiding the analyst's knowledge dependence. However, more studies with agricultural data must be performed to evaluate the superiority of REML for kriging. In this study, we did not use REML because it could significantly influence the performance of interpolations, which could, in turn, mask the factors that we would like to test (i.e., sampling neighborhood and interpolation methods routinely used in practice).

Our investigation showed that the indices that supposedly quantify the spatial dependence of a given dataset are unsatisfactory in predicting the interpolation accuracy (Fig. 3). From the performance of the interpolators measured by cross-validation in the calibration set (Table 4) and the validation set (Table 3), we noted that the best predictions occurred in Field 1, followed by Field 2, and

then, Field 3, given that the latter had completely unsatisfactory performance (R² between predicted and measured next to zero). Therefore, we can assume that the spatial dependence observed in these fields increases from Field 1 to Field 3. However, none of the spatial dependence indices showed results similar to that inferred by the validation techniques (Fig. 3). When evaluating the RDP, we did not notice similarity among the indices evaluated, showing that they are insufficient to compare the spatial dependence of datasets with different characteristics. It is noteworthy that both ISD and DSD are influenced by the spacing between samples, which may change semivariogram geometry (Fig. 2) and mask the results, especially in sparse datasets. Thus, different sampling densities are not comparable for those indices. As a result, cross-validation (or external validation) of data is the best way to infer the quality of interpolations quantitatively, whereas visual analysis of the semivariogram, despite being subjective, also allows an expectation of performance (qualitative analysis) of the subsequent interpolation when analyzed by experienced analysts.

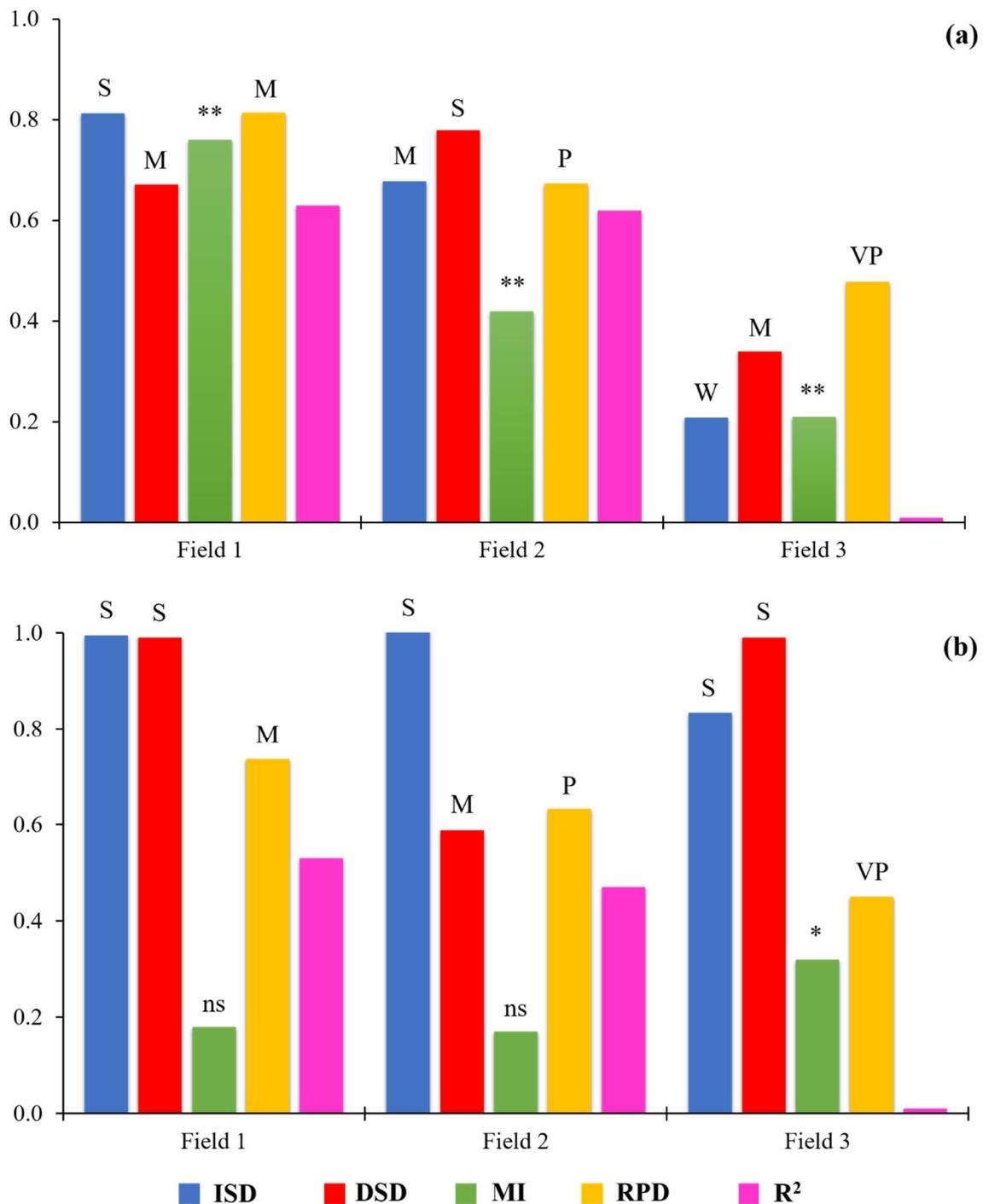


FIGURE 3. Comparison among the indices proposed to evaluate the spatial autocorrelation of data (spatial dependence – ISD, DSD and MI) and the performance of the predictions conducted on the external validation set (RPD and R²). The letters above the bars refer to the classification of spatial dependence proposed by the respective indices, as indicated above. The value of the indices were modified for graphical visualization purposes.

ISD: Index of Spatial Dependence – the values were divided by 20 (ISD/20). The letters indicate: P - weak spatial dependence; M - moderate spatial dependence; S - strong spatial dependence, according to Seidel & Oliveira (2016).

DSD: Spatial Dependence – the values obtained (in decimal) were subtracted from one (1-DSD); thus, higher values indicate greater spatial dependence (Cambardella et al., 1994). The letters indicate: M - moderate spatial dependence; S - strong spatial dependence.

MI: Moran’s Index. The letters/symbols indicate presence (*) or absence (ns) of spatial autocorrelation with 95% confidence.

RPD: Ratio of Percentage Deviation – values divided by two (RPD/2). The letters indicate: VP - very poor model and its use is not recommended; P - poor model, in which only high and low values are distinguishable; M - regular model that allows its use for inferences and correlations, according to Viscarra Rossel et al. (2006).

R²: coefficient of determination.

Due to the reasons mentioned above, analyzing interpolation quality should always be performed since the appearance of the map can deceive the unaware practitioner/analyst. As shown in Figure 4, interpolated maps by the IDW, without the optimization of sampling neighborhood (IDW_{null}), result in a map apparently more “reasonable” since it presents more rounded and smoothed features. This can be intuitively understood as a better representation of a natural variable (e.g., nutrient availability in the soil) than the maps with more defined limits and more acute contours, which were obtained when using fewer neighbors for the interpolation (Fig. 4).

However, the results of the external validation process (Table 3) indicate otherwise. Therefore, the appearance of the maps must be interpreted with caution.

Another question when looking to the IDW_{null} map in Figure 4 is that we can identify some “bull eyes” (concentric circles) which is a characteristic of IDW interpolator. This is reported in the literature as an artificial interruption of the spatial continuity. Thus, the appearance of the map where the IDW parameters were optimized, we note that such spots are no longer concentric. This is the reason for its better performance (IDW_{p1}), as already reported in Table 3.

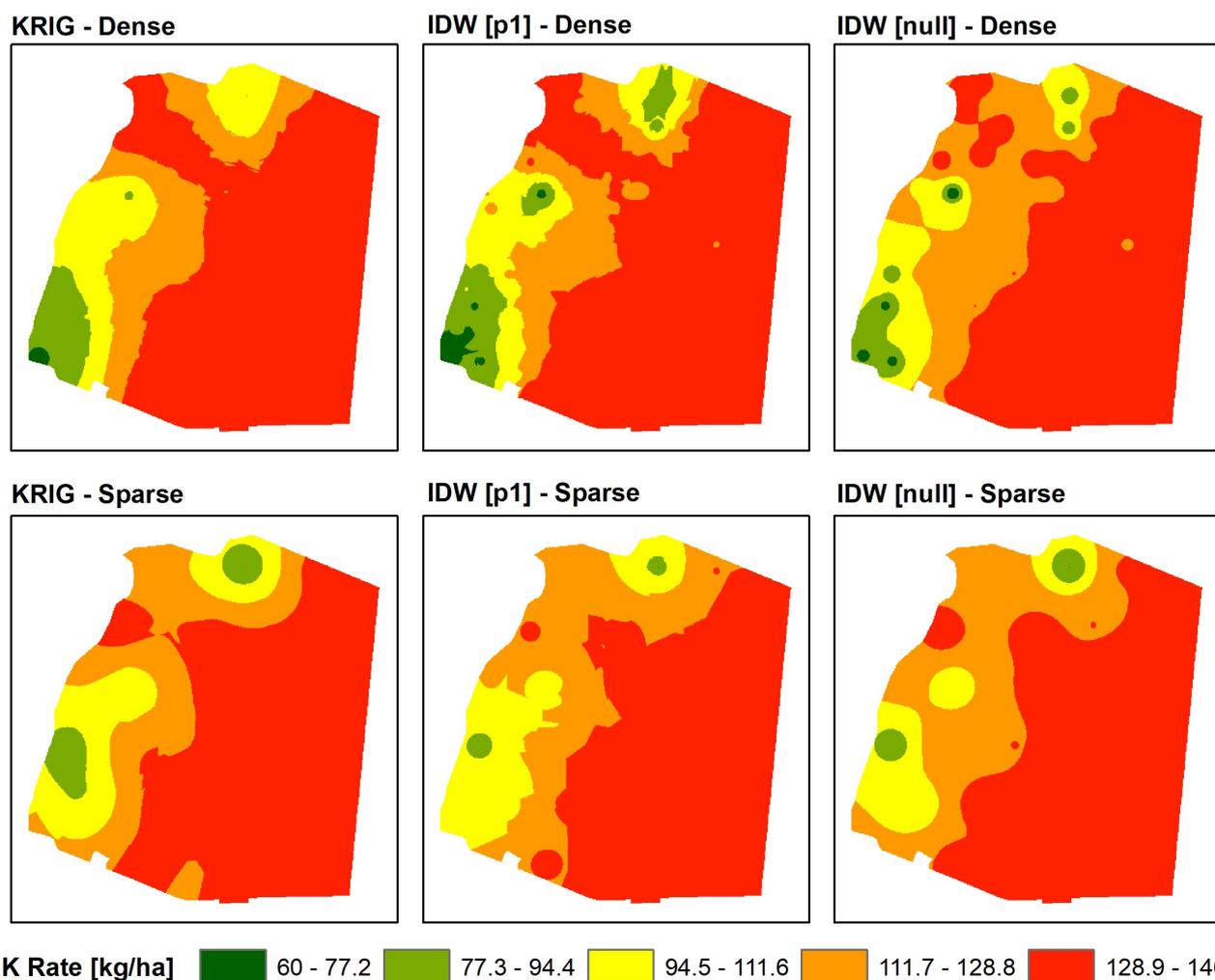


FIGURE 4. Potassium fertilizer prescription maps interpolated using KRIG, IDW_{p1} and IDW_{null} for sparse and dense samplings – example of Field 1.

KRIG: Ordinary Kriging with neighborhood search.

IDW_{p1}: Inverse Distance Weighting with power of 1 ($p=1$ – IDW_[p1]) with neighborhood search.

IDW_{null}: Inverse Distance Weighting with power of 2 ($p=2$ – IDW_[null]) without neighborhood search.

Misinterpretation of the quality of the predictions can also be conducted when evaluating maps from different sampling densities (Fig. 4). When dealing with sparse samples, the maps naturally present larger regions with more rectilinear and smoother contours as noted by others (Nanni et al., 2011; Siqueira et al., 2014). Our results show that, in sugarcane fields, soil samples every four hectares is inefficient in the characterization of the spatial variability of potassium (Table 3). Although we used the potassium variability as an example, this behavior can be generalized to all variables that have high anthropic influence,

especially those of soil chemical fertility. This shows that it is mandatory to evaluate the quality of interpolations before deciding whether to adopt the generated prescription map as this can directly impact the payback of precision agriculture.

In this sense, the error associated with interpolation predictions can be so high that it can jeopardize the entire result of the variable-rate fertilizer application. In this study, we found mean errors of interpolation between 12 and 20 kg ha⁻¹ (Table 3, RMSE), which represents errors in relation to an average fertilizer rate of 9, 28 and 13% for Fields 1, 2

and 3, respectively. In the case of high prediction error, the variable-rate application could be replaced by homogeneous zones of application (Mallarino & Wittry, 2004; Khosla et al., 2008), perhaps following the standard recommendation tables for fertilizers established by traditional experimentation (van Raij et al., 1996), which generally recommend only four or five fertilizer rates based on soil nutrient availability classes. This would facilitate the implementation of site-specific fertilization management and avoid high errors in the prescribed fertilizer rates. Thus, nutrient availability required for proper crop development would be ensured without risk of nutritional deficiency by application of insufficient fertilizer rates. At the same time, the environmental impact would be reduced by avoiding excessive rates.

CONCLUSIONS

The sampling neighborhood influences the accuracy of interpolations and is important to be considered when preparing variable-rate fertilizer prescription maps. However, there is no optimal number of neighbors to be used. Thus, this parameter should always be individually evaluated and optimized for each dataset prior to final interpolation in order to obtain optimal fertilizer prescription maps.

The quality of interpolations depends on the spatial structure of the variable under study, that is, the soil property must present spatial dependence. To access the performance of interpolations by procedures of external or cross-validation, it is mandatory to infer the reliability of the maps obtained. Although we have tested three indices that supposedly quantify spatial dependence without needing the validation process, they all showed limitations and did not present direct similarity with the real accuracy of the interpolation predictions.

As expected, smaller densities of samples may jeopardize data interpolation since this tends to impair the spatial dependence evaluation. Thus, the definition of homogeneous zones for fertilizer application appears to be a palliative approach.

ACKNOWLEDGEMENTS

To the PhD students, Guilherme M. Sanches (Bioenergy graduate program-UNICAMP) and Igor Q. M. Valente (Agricultural Engineering graduate program-UNICAMP), for the field data collection. To the Agricultural Engineer, Thiago Luis Brasco, for support in data organization. To the PhD student (Agricultural Engineering graduate program-UNICAMP), Julyane Vieira Fontenelli, for assistance with programs in R.

REFERENCES

Banu S (2015) Precision agriculture: tomorrow's technology for today's farmer. *Journal of Food Processing and Technology* 6:1-6. DOI: <https://doi.org/10.4172/2157-7110.1000468>

Betzek NM, Souza ED, Bazzi CL, Schenatto K, Gavioli A, Magalhães PSG (2019) Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. *Computers and Electronics in Agriculture* 157:49-62. DOI: <https://doi.org/10.1016/j.compag.2018.12.004>

Bivand RS, Wong DWS (2018) Comparing implementations of global and local indicators of spatial association TEST 27: 716-748. DOI: <https://doi.org/10.1007/s11749-018-0599-x>

Cambardella CA, Moorman TB, Novak JM, Parkin TB, Karlen DL, Turco R, Konopka AE (1994) Field-scale variability of soil properties in Central Iowa soil. *Soil Science Society of America Journal* 58:1501-1511.

Diggle P J, Ribeiro Jr. P J (2007) *Model-based geostatistics*. New York, USA, Springer, 229p.

Driemeier C, Ling LY, Sanches GM, Pontes AO, Magalhães PSG, Ferreira, JE (2016) A computational environment to support research in sugarcane agriculture. *Computers and Electronics in Agriculture* 130:13-19. DOI: <https://doi.org/10.1016/j.compag.2016.10.002>

Fu W, Zhao K, Zhank C, Wu J, Tunney H (2016) Outlier identification of soil phosphorus and its implication for spatial structure modeling. *Precision Agriculture* 17:121-135. DOI: <https://doi.org/10.1007/s11119-015-9411-z>

Goovaerts P (1999) Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89:1-45. DOI: [https://doi.org/10.1016/S0016-7061\(98\)00078-0](https://doi.org/10.1016/S0016-7061(98)00078-0)

Khosla R, Inman D, Westfall DG, Reich RM, Frasier M, Mzuku M, Koch B, Hornung A (2008) A synthesis of multi-disciplinary research in precision agriculture: site-specific management zones in the semi-arid western Great Plains of the USA. *Precision Agriculture* 9:85-100. DOI: <https://doi.org/10.1007/s11119-008-9057-1>

Li J, Heap AD (2011) A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics* 6:228-241. DOI: <http://dx.doi.org/10.1016/j.ecoinf.2010.12.003>

Mallarino AP, Wittry DJ (2004) Efficacy of grid and zone soil sampling approaches for site-specific assessment of phosphorus, potassium, pH, and organic matter. *Precision Agriculture* 5:131-144. DOI: <https://doi.org/10.1023/B:PRAG.0000022358.24102.1b>

Mueller TG, Pusuluri NB, Mathias KK, Cornelius PL, Barnhisel RI (2004) Site-specific soil fertility management: a model for map quality. *Soil Science Society of America Journal* 68:2031-2041. DOI: <https://doi.org/10.2136/sssaj2004.2031>

Nanni MR, Povh FP, Demattê JAM, Oliveira RB, Chicati ML, Cezar E (2011) Optimum size in grid soil sampling for variable rate application in site-specific management. *Scientia Agricola* 68:386-392.

Oliver MA, Webster R (2014) A tutorial guide to geostatistics: computing and modelling variograms and kriging. *Catena* 113:56-69. DOI: <https://doi.org/10.1016/j.catena.2013.09.006>

Robinson TP, Metternicht G (2006) Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture* 50:97-108. DOI: <https://doi.org/10.1016/j.compag.2005.07.003>

Samuel-Rosa A (2017) *spsann*: optimization of sample configurations using spatial simulated annealing. R package version 2.1-0. <https://CRAN.R-project.org/package=spsann>

Seidel EJ, de Oliveira MS (2016) A classification for a geostatistical index of spatial dependence. *Revista Brasileira de Ciência do Solo* 40:1-10. DOI: <https://doi.org/10.1590/18069657rbcS20160007>

Siqueira DS, Marques J, Pereira GT, Barbosa RS, Teixeira DB, Peluco RG (2014) Sampling density and proportion for the characterization of the variability of Oxisol attributes on different materials. *Geoderma* 232-234:172-182. DOI: <https://doi.org/10.1016/j.geoderma.2014.04.037>

Van Raij B, Cantarella H, Quaggio JA, Furlani AMC (1996) Fertilization and liming recommendation for the São Paulo State (In Portuguese). São Paulo, Instituto Agrônômico & Fundação, 2 ed. 285p.

Vieira SR, Carvalho JRP, González AP (2010) Jackknifing for semivariogram validation. *Bragantia* 69:97-105. DOI: <https://doi.org/10.1590/S0006-87052010000500011>

Viscarra Rossel RA, McGlynn RN, McBratney AB (2006) Determining the composition of mineral-organic mixes using UV-Vis-NIR diffuse reflectance spectroscopy. *Geoderma* 137:70-82. DOI: <https://doi.org/10.1016/j.geoderma.2006.07.004>

Yamamoto JK (2005) Correcting the smoothing effect of ordinary kriging estimates. *Mathematical Geology* 37:69-94. DOI: <https://doi.org/10.1007/s11004-005-8748-7>