
# Application of routines for automation of geostatistical analysis procedures and interpolation of data by ordinary kriging.

**[1]Betzek, N. M.; [2]Souza, E. G.; [1]Bazzi, C. L.; [1]Magalhães, P. S. G.; [1]Schenatto, K.; [1]Gavioli A.; [1]Dall´Agnol, R. W.**

[1]Federal University of Technology - Paraná, Brazil; [2]State University of West Paraná, Brazil.

**A paper from the Proceedings of the**
**14th International Conference on Precision Agriculture**
**June 24 – June 27, 2018**
**Montreal, Quebec, Canada**

***Abstract.*** *Ordinary kriging (OK) is one of the most suitable interpolation methods for the construction of thematic maps used in precision agriculture. However, the use of OK is complex. Farmers/agronomists are generally not highly trained to use geostatistical methods to produce soil and plant attribute maps for precision agriculture and thus ensure that best management approaches are used. Therefore, the objective of this work was to develop and apply computational routines using procedures and geostatistical libraries in the R software, to automatically identify the best parameters for OK interpolation. Through the implemented procedures, six different models (spherical, Gaussian, exponential, Matérn 1.0, Matérn 1.5 and Matérn 2.0) and two statistical methods of optimizing the semivariogram (OLS - ordinary least squares and WLS - weighted least squares) were analyzed. We tested 25 values for the semivariogram parameters for each of the 12 models, totalizing 300 different configurations to identify the best parameters used to measure data by OK. To validate the procedure the routines were applied to corn and soybean crop data collected for three years in an agricultural area with approximately 20.9 ha and 73 sampling points. For soybean yield, the best fit model for the experimental semivariogram was Gaussian-WLS and for corn yield was Matérn 1.5-OLS. Other models presented similar adjustments to those chosen as better, emphasizing the quality of the geostatistical procedures implemented that were able to identify, without tendencies, the best adjustments for semivariogram. The obtained parameters were used in the interpolation process by OK onto a 5-m grid (pixels of 25 $m^2$). After the interpolation was processed, thematic maps were generated for each crop. By means of the interpolated maps generated by OK, it was possible to identify regions in which the variations in the average yield occur in both crops, thus being possible to perform differentiated management in these locations. It can be concluded that the computational routines implemented were efficient and able to identify the best fit for the semivariogram to be used when applying OK.*

***Keywords.*** *computational routines, semivariogram, thematic maps.*

## Introduction

Knowledge of the spatial and temporal variability of soil and plant is essential for efficient agricultural management. For this, information of characteristics such as yield and soil chemical and physical attributes are interpolated, usually by Ordinary Kriging (OK) that makes use of geostatistic.

Geostatistic is based on concepts of classical statistics and considers the geographic coordinates of the sample points and the spatial dependence between them to identify the spatial behavior of the variables. The semivariogram is the geostatistic tool that allows to identify the degree of spatial dependence between samples and thus define the model that best describes the behavior of the data in space (Journel and Huijbregts, 1978). The performance of the data interpolation by OK can be influenced by the spatial structure of the data and by the choice of the semivariogram adjustment model, the search radius and the number of nearest neighboring points that are used in the estimates (Guastaferro et al., 2010; Isaaks and Sriivastava, 1989).

R software uses a variety of statistical and graphical techniques extensively, being able to perform geostatistical data analysis and spatial data interpolation by OK. However, although the advantages of R statistical software are recognized, their use is not easy and should be supported by adequate knowledge of the statistical techniques involved.

The objective of this work was to automatically identify, through the procedures implemented in software R, the best parameters for OK interpolation, to present as a thematic map, the spatial distribution of the interpolated variable.

## Methodology

Procedures were implemented in software R using the geoR library to perform geostatistical analysis of the data and to automatically identify the best fit for the experimental semivariogram. We analyzed six different models (spherical, Gaussian, exponential, Matérn 1.0, Matérn 1.5 and Matérn 2.0) and two statistical methods of semivariogram optimization (OLS - minimum squares and WLS), totalizing 12 different models. For each of the models, 25 different sets of parameters (five initial values for the contribution parameter and five for the range) were used, totaling 300 different adjustments analyzed. Cross-validation (Isaaks and Sriivastava, 1989) was used to identify the best fit model for the semivariogram.

The data were interpolated by OK to a 5x5 pixel grid. The classical Matheron estimator (Matheron, 1963) was used to calculate the semivariance, and the range was limited to half the maximum distance between two points (cutoff = 50%). After the interpolation was processed, thematic maps were created for each attribute.

To validate the results, soybean (2013, 2014 and 2015) and maize (2013 and 2014) data were collected based on an irregular sampling grid in an agricultural area, located in the municipality of Serranópolis do Iguaçu-PR, in the western region of the state of Paraná - Brazil. The area has a geographic center of 25º24'17 "S and 54º00'18"O and has a ground elevation of approximately 363 m a.s.l., which approximately 20.9 ha where 73 sampling points were collected.

## Results and discussion

After statistical analysis, the data outside the mean range ± 3 standard deviations were considered outliers and eliminated. Based on the information of each sampling point, it was identified that the minimum distance between sample points was 30 m and the maximum 838 m, with a cutoff of 419 m, characterizing 14 lags with at least 30 pairs of points.

This information was used in the implemented procedures, and after the evaluation of the 300 different adjustment combinations for the semivariogram, it was identified that for the average soybean yield the best model of adjustment of the semivariogram was Gaussian with WLS (Fig 1) and for the average corn yield was Matérn 1.5 with OLS (Fig 2).

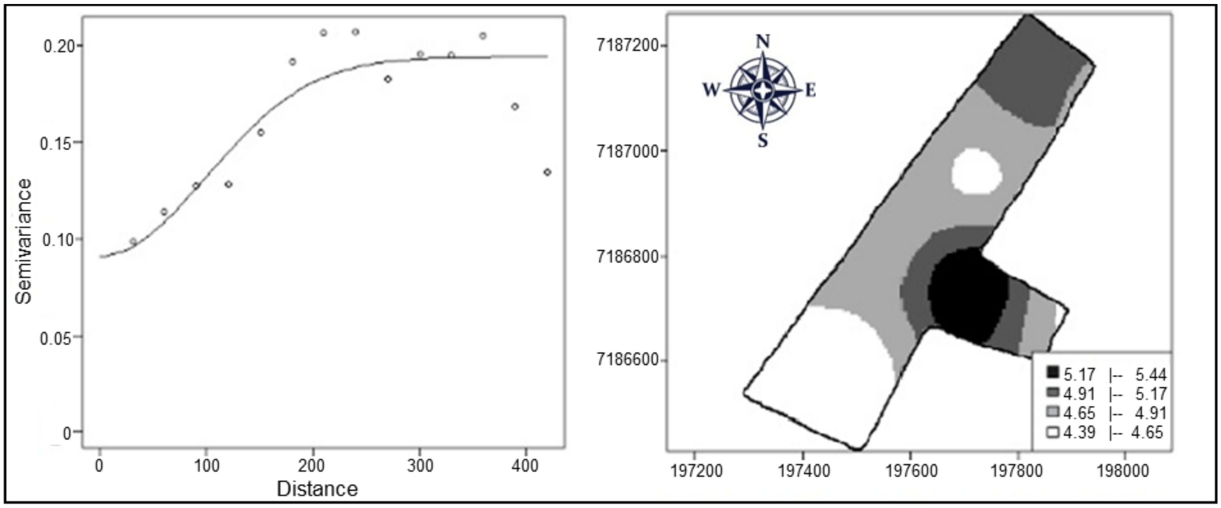


**Fig 1. Semivariogram (best model: Gaussian: method: WLS) and thematic map generated with average yield data of soybean by OK**

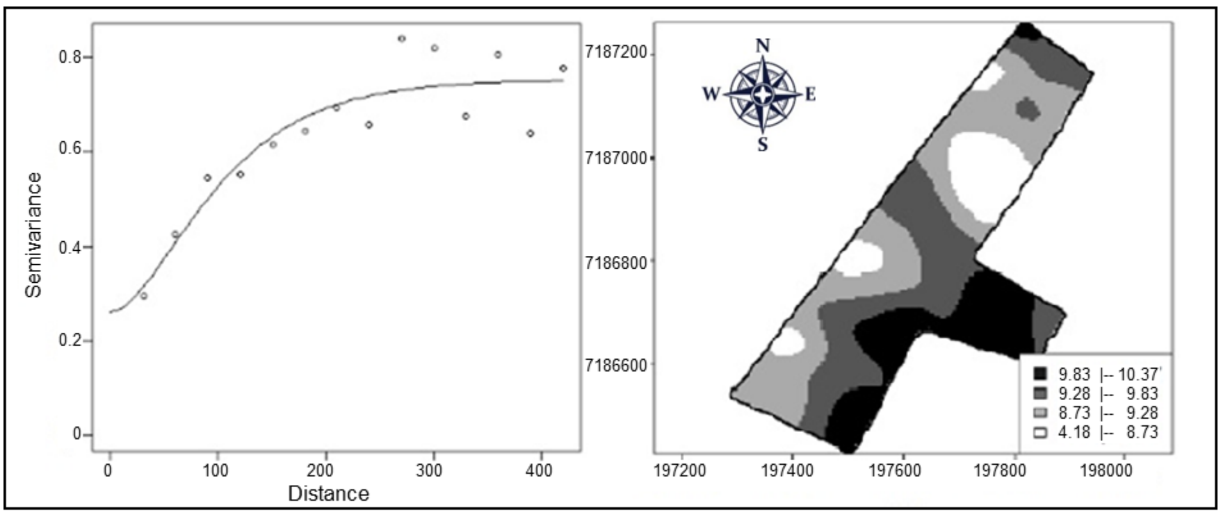


**Fig 2. Semivariogram (best model: Matérn 1.5; method: OLS) and thematic map generated with average corn yield data by OK**

Other models presented similar adjustments to those chosen as better, this emphasizes the quality of the implemented geostatistic procedures that were able to identify, without tendency, the best adjustments for semivariogram. By means of the interpolated maps generated by OK, it was possible to identify regions in which variations in the average yield occur in both crops, being thus possible to carry out differentiated management in these places.

# Conclusion

The computational functions implemented in the R software were efficient and able to identify the best fit for semivariogram parameters automatically. The thematic maps generated based on the parameters found allowed to identify the spatial variability of soybean and corn yield in the studied area. The implemented procedures can also be used with chemical and physical soil attributes, facilitating the decision making regarding the application of agricultural fertilizers.


## Acknowledgements

The authors would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES), and the National Council for Scientific and Technological Development (CNPq) for funding.

## References


Guastaferro, F., Castrignanò, A., de Benedetto, D., Sollitto, D., Troccoli, A., Cafarelli, B., 2010. A comparison of different algorithms for the delineation of management zones. Precis. Agric. 11, 600–620. doi:10.1007/s11119-010-9183-4

Isaaks, E.H., Sriivastava, R.M., 1989. Applied Geostatistics, first. ed. Oxford University Press, New York.

Journel, A., Huijbregts, C., 1978. Mining geostatistics, first. ed. Academic press., New York.

Matheron, G., 1963. Principles of Geostatistics. Econ. Geol. 58, 1246–1266.