XXXV Congresso Brasileiro de Ciência do Solo
CENTRO DE CONVENÇÕES - NATAL / RN
O SOLO E SUAS MÚLTIPLAS FUNÇÕES
02 a 07 DE AGOSTO DE 2015

# Spatial prediction of Zn, pH, P and clay from statistical and geostatistical methods (1).

**Mariana Ruger Sacco(2); Michele Duarte de Menezes(3); Ana Paula Branco Corguinha(4); Guilherme Amaral de Souza(5); Luiz Roberto Guimarães Guilherme(3).**

(1) Work carried out with funds from International Zinc Association (IZA), CNPq, CAPES, and FAPEMIG.
(2)Undergraduate student; Federal University of Lavras, Lavras, Minas Gerais; mariana.sacco@hotmail.com; (3)Professor; Federal University of Lavras; (4)D.Sc student; Federal University of Lavras; (5) Post D.Sc student; Federal University of Lavras.

**ABSTRACT:** The spatial prediction of Zn, pH, P and clay at non-sampled areas in a continuous way can be source of valuable information for agriculture and land management. Thus, this work aimed to assess the spatial prediction of Zn, pH, P and clay from ordinary kriging (OK), inverse distance weighted (IDW) and multilevel B-spline. A validation data set was used to assess the best prediction method, and two indices were calculated from the observed and predicted values: mean prediction error and root mean square of prediction error. IDW was the best to describe most of the studied soil properties, followed by multilevel B-spline. It was noticed that interpolation of soil properties requires testing different methods once the accuracy can be highly variable depending on the nature of the method and data variability.

**Indexing Terms:** inverse distance weighted, ordinary kriging, multilevel B-splines.

## INTRODUCTION

The spatial prediction of Zn, pH, P and clay at non-sampled areas in a continuous way can be source of valuable information for agriculture and land management. Different interpolation techniques have been used with varying degrees of success, and improved in order to create more accurate maps (Burrough et al., 1994). This is an application of pedometry that refers to a quantitative data-driven generation of soil property map (McBratney et al., 2000).

This work presents a novel initiative to create a harmonized data base related to soil property and its spatial prediction, which the main focus is Zn pH, P, and clay. The soil samples were obtained from georeferenced data base of our partners and Soil Fertility Laboratory of Federal University of Lavras.

All of methods applied in this study provide a smooth and gradual surface, differing on the criteria used in the selection of the weight values in relation to the distance. IDW uses simple distance relation, splines the minimization of curvature and enforcement of smoothness, and ordinary kriging the minimization of variance (Li and Heap, 2008).

Since there is no digital prediction model that fits all regions and purposes, this work aimed to assess the spatial prediction of Zn, pH, P and clay from ordinary kriging, IDW and multilevel B-spline prediction methods.

## MATERIAL AND METHODS

A data set with 38,151 samples of Zn, P and pH and 25,592 samples of clay were used for spatial prediction. The interpolation area was concentrated around the samples, since interpolation methods are distance based and low density points can decrease accuracy. A total of 719 cities and an area of 877.495,8 km² were mapped (**Figure 1**).

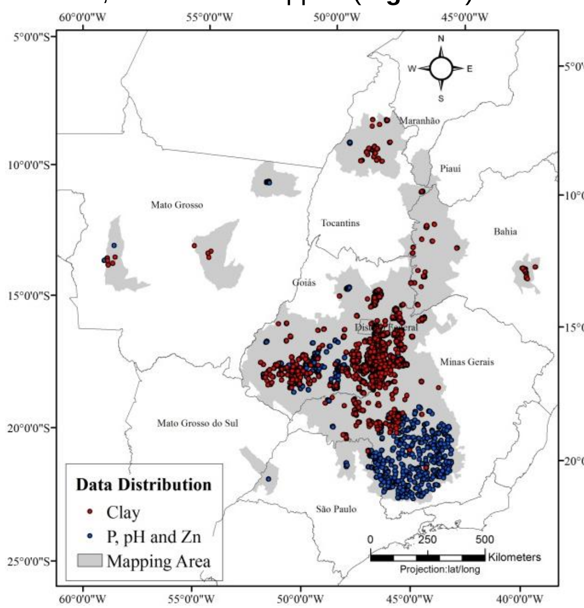



**Figure 1** - Spatial distribution of sampling points and the mapped area.

Samples were collected from 0 to 20 cm. Soil pH was determined in a 1:2.5 soil:water suspension. P and Zn were extracted using 100 mL of the Mehlich$^{-1}$ solution (0.05 mol $L^{-1}$ HCl + 0.0125 mol $L^{-1}$ $H_2SO_4$) reacted with 10 $cm^3$ of soil sample. Zn concentrations in the extracted solutions were determined by flame atomic absorption spectrometer, while P was determined by colorimetry. The clay content was determined with the pipette method (EMBRAPA, 1997).

The values predicted at nonsampled areas from IDW are estimated using linear combination of values at the sampled points, weighted by an inverse function of the distance from the point of interest to the sample points. The weights (*λi*) are expressed as:

$$z_0 = \frac{\sum_{i=1}^{s} zi\left(\frac{1}{d_i^k}\right)}{\sum_{i=1}^{s}\left(\frac{1}{d_i^k}\right)}$$

where $z_0$ is the estimated value at point 0; $z_i$ is the value z value at known point i, $d_i$ the distance between i and 0, k is a power parameter, and s represents the number of sampled points used for the estimation. This interpolation was performed in ArcGIS 10.0 (Esri) with power parameter of 2.

Multilevel B-spline is an algorithm for spatial interpolation of scattered data proposed by Lee et al. (1997). The algorithm makes use of a coarse-to-fine hierarchy of control lattices to generate a sequence of bicubic B-spline functions, whose sum approaches the desired interpolation function. Large performance gains are realized by using B-spline refinement to reduce the sum of these functions into one equivalent B-spline function. It was performed in SAGA GIS version 2.1.4.

For the ordinary kriging interpolation, the first step is to calculate the experimental semivariogram using the following equation:

$$\gamma^*(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[z(x_i) - z(x_i + h)]^2$$

where $\gamma^*(h)$ is the estimated value of the semivariance for lag *h*; *N(h)* is the number of experimental pairs separated by vector *h*; $z(x_i)$ and $z(x_i + h)$ are values of variable *z* at $x_i$ and $x_i$+*h*, respectively; *xi* and $x_i$+*h* are position in two dimensions. The models fitted of semivariograms were gaussian, spherical, and exponential. This interpolation was performed in ArcGIS.

Only a validation data set (20% not applied for prediction) was used to assess the best prediction method. Such data was randomly chosen. Two indices were calculated from the observed and predicted values. The mean prediction of error (MPE) was calculated by comparing estimated values ($\hat{z}(s_j)$) with the validation points ($z^*(s_j)$):

$$MPE = \frac{1}{l}.\sum_{j=1}^{l}[\hat{z}(s_j) - z^*(s_j)]$$

and the root mean square prediction error (RMSPE):

$$RMSPE = \sqrt{\frac{1}{l}\sum_{j=1}^{l}[\hat{z}(s_j) - z^*(s_j)]^2}$$

where *l* is the number of validation points. The $R^2$ from scatterplot graphics 1:1 between real (validation data set) and estimated values from interpolation methods was also calculated.

## RESULTS AND DISCUSSION

The mean, median, QQ-plot graphics are presented in the **Figure 2**. The quantile values of the standard normal distribution are plotted on the x-axis, and the corresponding quantile values of the dataset are plotted on the y-axis. If the data is normally distributed, the points in the QQ plot graphic fall on the 45° reference line. It is possible to notice that most of the properties had its points deviates from the line, except from pH.

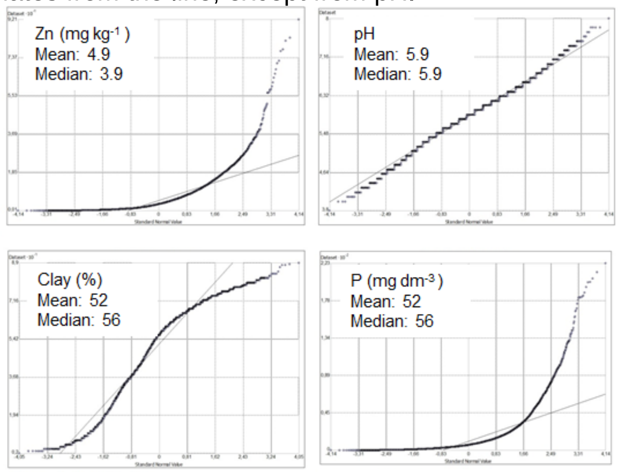
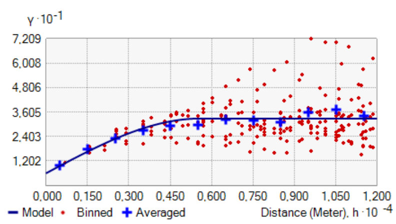



**Figure** 2 - Q-Q plots with mean and median values of Zn, pH, Clay and P.

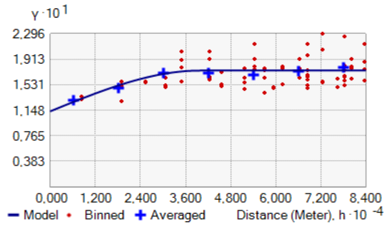
The geostatistical parameters used to fit semivariograms of soil properties are presented in the **Figure 3 (a, b, c and d)**. Spherical model was the main model fitted to experimental semivariograms. It is important to notice the high nugget effect that represents error measurement plus variation that occurs over distances less than the shortest sampling interval. In this case, the latter could be dominant due the sparse location of sampling data, which in turn can generate undesirable large estimates of variances (Webster and Oliver, 2007), providing less reliable ordinary kriging maps.

a) Zn (mg $kg^{-1}$)



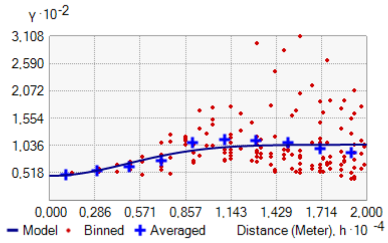

Spherical; range: 5574.64; partial sill:26.79; nugget: 6.04

b) pH



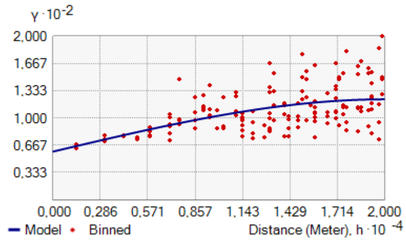

Spherical; range: 39900.00; partial sill:0.06; nugget: 0.11

c) P (mg dm$^{-3}$)




Gaussian; range: 12074.12; partial sill: 58.76; nugget: 46.48

d) Clay (%)




Spherical; range: 20000.00; partial sill: 63.71; nugget: 58.86

**Figure 3** - Geostatistical parameters and semivariograms of soil properties.

From the validation indexes (**Table 1**), the closer to zero MPE and RMSE are, the more accurate the prediction of soil property. The closer the real values are of estimated values, the closer the $R^2$ is of 100. IDW followed by multilevel B-spline presented higher accuracy). Geostatistic is excessively data dependent, requiring a large number of closely spaced data points (Zhu, 1997). Additionally, as long as the nugget effect is high in semivariograms (Figure 6) especially for clay and P, there may be undesirable large estimates of variances, providing less smooth and less reliable ordinary maps, which the low precision of OK according to the calculated statistical indices.

IDW is an interpolation method that does not allow values outside the range of the known points, this could be the reason why this method had higher accuracy when compared to the others.

The range of values of Zn varied from 0.1 to 92.1 mg dm$^{-3}$, pH values varied from 3.8 to 8.0, clay from from 2 to 89% and P from 0.05 to 225.2 mg dm$^{-3}$ .

Multilevel B-spline is directed for scattered or non-uniform distribution of data samples. The algorithm is suitable for mapping large areas from a large number of points. In the literature it was not found any study applying this method to interpolate soil properties, therefor there is a need on further studies to comprehend its suitability.

On **Figure 4** it is shown the maps for the chemical properties which had higher accuracy. Among those, pH was best described by multilevel b-spline and it was the map to show higher smoothness. By the maps analyses it was possible to recognize that about half of the area had high concentration of Zn, the adequate range of pH (5.6 – 6.0) also had half of the area, about 60% of the area had high or very high concentrations of clay and P was mainly low in the area.

## CONCLUSIONS

Interpolation of soil properties requires testing different methods once the accuracy can be highly variable depending on the nature of the method and data variability.

## ACKNOWLEDGEMENTS

We would like to thanks our partners: CAMPO Environmental and Agronomical Laboratory; Fundação MT Foundation for Support to Agricultural Research of Mato Grosso State; EMBRAPA Brazilian Corporation of Agricultural Research; SIAP Agribusiness; AMPAR Associação dos Mineradores de Pains, Arcos e Região; COMIGO Agro-industrial Cooperative of Rural Producers of Southwest Goiás; UNEMAT State University of Mato Grosso.

## REFERENCES

BURROUGH, P.A.; BOUMA, J.; YATES, S.R. 1994. The state of the art in pedometrics. Geoderma 62: 311-326.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa Solos. Manual de métodos de análise de solo. 2.ed. Rio de Janeiro, 1997. 212p.

LEE. S.; WOLBERG. G.; SHIN. S.Y. (1997): 'SCATTERED DATA INTERPOLATION WITH MULTILEVEL B-SPLINES'. IEEE TRANSACTIONS ON VISUALISATION AND COMPUTER GRAPHICS. VOL.3. NO.3.

LI, J.; HEAP, A.D. A review of spatial interpolation methods for environmental scientists. Australia: Record, 2003. 137p.

McBRATNEY, A.B., MENDONÇA-SANTOS, M.L.; MINASNY, B. Geoderma: 117: 3-52, 2003.

WEBSTER, R.; OLIVER, M.A. Geostatistics for environmental scientists. 2ed. United Kingdon: John Wiley and Sons. 2007. 333p.

**Table 1**. Comparison of interpolation methods of soil properties.

| | Multilevel B-Splines | | | IDW | | | Ordinary kriging | | |
|---|---|---|---|---|---|---|---|---|---|
| | RMSE | ME | $R^2$ | RMSE | ME | $R^2$ | RMSE | ME | $R^2$ |
| Zn | 10.76 | 0.12 | 0.48 | 3.88 | 0.04 | 0.57 | 25.02 | 0.29 | 0.26 |
| Clay | 53.15 | 0.76 | 0.78 | 3.10 | -0.04 | 0.80 | 489.17 | 6.96 | 0.27 |
| P | 77.05 | 0.89 | 0.44 | 70.51 | 0,81 | 0.45 | 142.28 | 1.64 | 0.29 |
| pH | 1.36 | 0.02 | 0.39 | 1.50 | 0.02 | 0.48 | 2.87 | 0.03 | 0.27 |

IDW – inverse distance weighted; RMSE – root mean squared of error; ME: mean error.

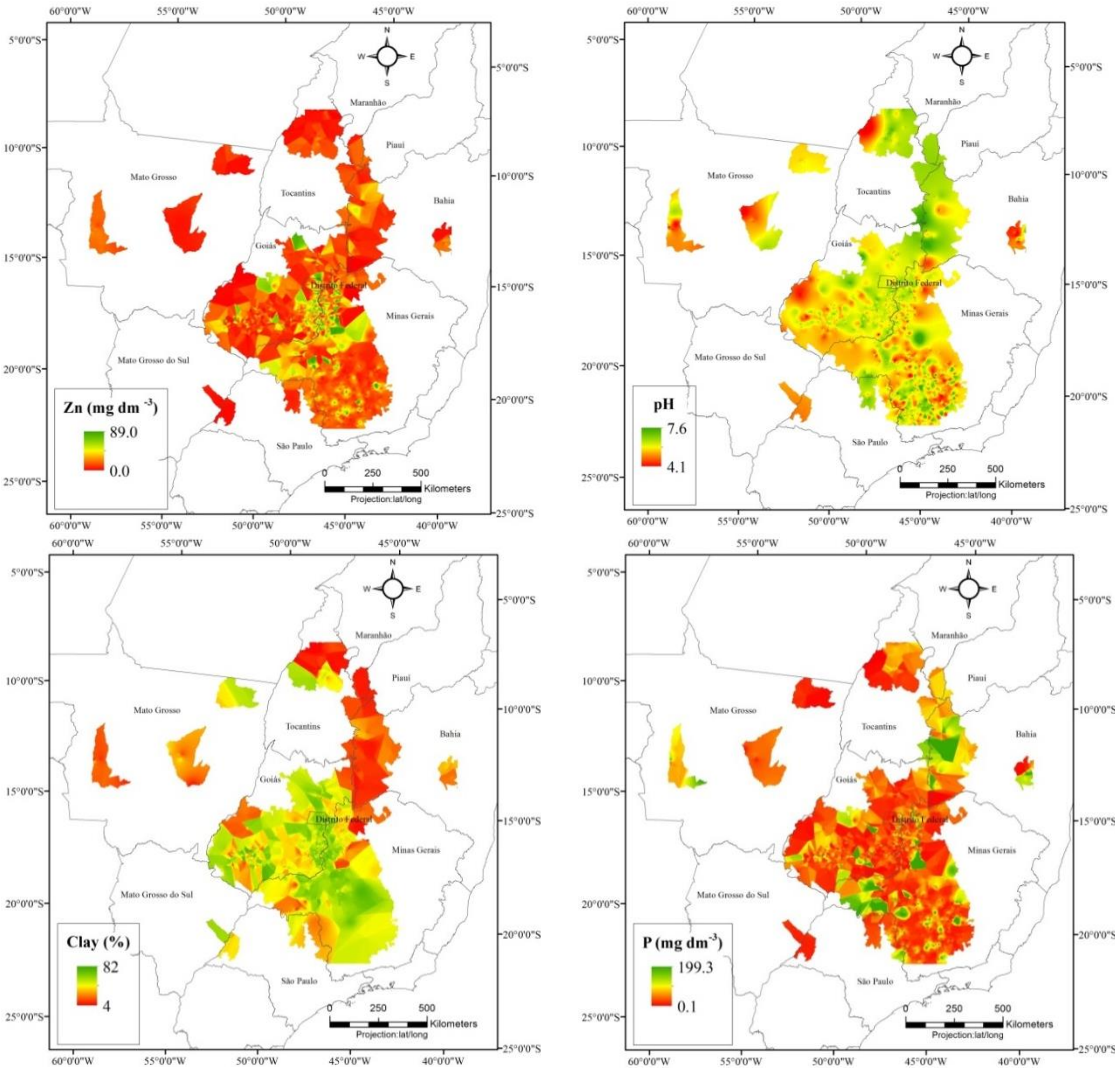



**Figure 4** - Spatial prediction of the highest accuracy methods for Zn (IDW), pH (multilevel B-spline), Clay (IDW) and P (IDW).