GEOSTATISTICAL AND DETERMINISTIC INTERPOLATORS IN THE ANALYSIS OF THE SPATIAL DISTRIBUTION OF SOIL pH IN SOROCABA CITY (SÃO PAULO STATE)

Darllan Collins da Cunha e Silva¹, Renan Angrizani de Oliveira², Vanessa Cezar Simonetti², Érico Tadao Teramoto¹, Jomil Costa Abreu Sales²

SAP 26070 Received: 12/10/2020 Accepted: 13/02/2021

ABSTRACT - Soil is a natural resource that allows the maintenance of life on Earth, therefore, it is important to appreciate techniques to assist the conservation of their quality, since the inadequate management of its use causes their degradation. In this sense, the soil pH values of the city of Sorocaba (SP) were analyzed using a pHmeter and a GPS, with 112 points distributed homogeneously in the study area and another 50 random points to verify spatial variability through geostatistical interpolators of ordinary kriging and deterministic of inverse square distance (ISD). The predominant values of pH per soil class were also determined. The pH values predominated between the ranges of 6.0 and 6.5; totaling 74.98% of the study area for the ordinary kriging method and 75.93% for the ISD, observing a low variation among the methods, which result in a root mean square error values of 0.349951 for ordinary kriging and 0.349019 for the ISD. The concordance index for both methods was 0.9944 with mean standard deviation of 0.0374 for ISD and 0.0369 for ordinary kriging. For the different soil classes, pH values ranged from 5.9 to 6.3. Therefore, both the theoretical models applied in the interpolation of the pH data efficiently explain the pH variability of the Sorocaba soil.

Keywords: inverse square distance, ordinary kriging, spatial analysis.

INTRODUCTION

Soil acidity is an important monitoring parameter because it interferes with the absorption of nutrients by plant root systems. In acidic soils, there is greater availability of micronutrients for plants, which can cause damage to production due to interference with their metabolism, and it is necessary to employ acidity correction techniques, such as liming (RAIJ et al., 2001; RAIJ, 2011).

With geoprocessing techniques applied to soil management studies, it is possible to reduce resources and the use of inputs, as well as the more detailed management of agricultural production systems (CARVALHO et al., 2002; MELLO et al., 2006; SALES et al., 2016; SILVA et

¹Universidade Estadual Paulista (Unesp), Campus Experimental, Registro, São Paulo, Brasil. E-mail: darllan.collins@unesp.br *Corresponding author.
²Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia, Sorocaba, São Paulo, Brasil.
However, due to the expensive cost of field collection, it is not always possible to have data from large areas, which the values of the non-sampled points must be estimated for the generation of spatially continuous data. In this case, the spatial interpolation methods make it possible to estimate the unsampled data satisfactorily (SHI et al., 2009; LI; HEAP, 2011).

For LI and Heap (2011), the spatial interpolation methods have several applications, and the ordinary kriging and inverse square distance (ISD) methods are the most used in the environmental and agricultural areas. Silva et al. (2010) used the geostatistical interpolation methods of the ordinary kriging and deterministic of the inverse square distance, comparing them in the representation of the spatial variability of the soil pH in the coffee crop, which found that the ordinary kriging proved to be more adequate in the demonstration of pH changes.

However, the study by Robinson and Metternicht (2006), which compared the kriging, Inverse Distance Weight (IDW) and spline interpolation methods applied to some soil properties, found that all methods had similar values of mean square error. For Oshunsanya et al. (2017), who carried out the mapping of the Ibadan University campus to verify the distribution of soil properties, only the IDW was used to interpolate the data.

Therefore, the estimation of soil pH values using interpolation methods with Geographic Information Systems (GIS) allows the adoption of management techniques that provide better availability of nutrients and plant development. Thus, the present study aims to analyze the spatial variability of soil pH values in the municipality of Sorocaba (SP), and to verify through geostatistical and deterministic interpolators, which theoretical model has a more appropriate application to assess the spatial variability of the collected data.

MATERIAL AND METHODS
Characterization of the study area

The study area corresponds to the municipality of Sorocaba, São Paulo State, under latitude 23°30'06" South and longitude 47°27'29" West (Figure 1), with a population of 687,357 inhabitants (IBGE, 2021). From 2014, the municipality became part of the Metropolitan Region of Sorocaba, composed of 27 municipalities, having a significant economic expression (OLIVEIRA et al., 2016).

The climate of the municipality, according to Köeppen-Geiger, is dominant of the Cfa type, characterized as hot subtropical, with not very dry winter, with an average annual precipitation of 1,400 mm, also having a more northerly strip area with characteristics of climate Cwa type, characterized as hot and rainy temperate climate, with less than 30 mm pluviometry in the predominantly driest month (SMITH et al., 2014; SIMONETTI et al., 2018).

FIGURE 1 - Location of the study area.
Methodological procedure

The data collection of the potential of hydrogen (pH) of the soil in the municipality of Sorocaba was performed using the pH meter (Sonda Terra, model PH-220S) at 0-20 cm depth. The distribution of the sampling points occurred through a 2,000 m x 2,000 m square mesh with the central points equidistant from each other, totaling 112 sample points (Figure 2A). Within the municipal limit, another 50 points were randomly chosen to verify and compare the values obtained through the interpolators methods (Figure 2B). The sampled values were transcribed in an Excel table and exported to the ArcGIS 10.4 software, while the vectorial limit of the municipality of Sorocaba was extracted from the cartographic database of the Brazilian Institute of Geography and Statistics (IBGE).

For the analysis of the pH distribution of the soil, the interpolators methods of ordinary kriging and deterministic of inverse square distance (ISD) were used, because according to some studies, these methods have satisfactory results to estimate the distribution of physical and chemical attributes of the soils (CORÁ et al., 2004; SILVA et al., 2010; SOUZA et al., 2010). The ISD is a deterministic method, which considers that the nearest points have a greater resemblance when compared to the farthest points. Thus, each point has decreased local influence with distance as presented in Equation 1 (DRUCK et al., 2004; ROBINSON; METTERNICHT, 2006).

\[
\text{ISD} = \frac{\sum_{i=1}^{n} \left( \frac{1}{d_{i}^2} \right) X_i}{\sum_{i=1}^{n} \left( \frac{1}{d_{i}^2} \right)}
\]

(Equation 1)

Where:

- ISD = inverse square distance,
- \(X_i\) = value of the variable of the i-th neighboring locality and
- \(d_i\) = Euclidean distance between the i-th neighboring point and the sampled point.

In the geostatistical method of ordinary kriging, the information of the obtained values allows the calculation of the distances between the observations through the semivariogram (a function of the distance) relating the variance of the obtained points with the distances between them (ROBINSON; METTERNICHT, 2006; SILVA et al., 2010). Therefore, to evaluate the spatial variability of the potential of hydrogen (pH), the spatial dependence of the data was expressed by the semivariance function \(\hat{\gamma}(h)\), estimated by Equation 2:

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

(Equation 2)

Where:

- \(\hat{\gamma}(h)\) = semivariance function,
- \(N(h)\) = sampled values of a given attribute of the studied soil and
- \(Z(x_i)\) e \(Z(x_i+h)\) = separated by \(h\) distance.
Application of geostatistical... 

In the presence of spatial dependence between the data, inferences were made to estimate the pH values of the soil in non-sampled locations. For this, was used the ordinary kriging interpolation method (YAMAMOTO; LANDIM, 2013) defined by Equation 3:

\[ Z_{k0}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \quad (Equation 3) \]

Where:
\[ Z^*k0 = \text{estimator for a point } (x0) \text{ in the region,} \]
\[ \lambda_i = \text{weights used in the estimate and} \]
\[ Z(x_i) = \text{set of } n \text{ available data.} \]

To assess the spatial dependence of the phenomenon the Spatial Dependency Index (SDI) proposed by Cambardella et al. (1994) which uses in its calculations the nugget effect (\( C_0 \)) and the contribution (\( C \)), being expressed by Equation 4, with the following classification for its values: if SDI \( \leq 25\% \) the phenomenon has strong spatial dependence; between 25% and 75% have moderate spatial dependence and \( \geq 75\% \) have weak spatial dependence.

\[ SDI = \left( \frac{C_0}{C_0 + C} \right) \times 100 \quad (Equation 4) \]

Where:
\[ SDI = \text{Spatial Dependency Index,} \]
\[ C_0 = \text{nugget effect and} \]
\[ C = \text{contribution.} \]

For the geostatistical analysis, the theoretical models used for adjustment were the spherical, the exponential and the gaussian, being presented in the results the model that better fits in the study. The most appropriate model validation and selection were based on the mean error (ME), the mean standardized prediction error (MSE), the square root of the mean of the square of the differences between the estimated and true values (root mean square - RMS), in the standard error of the prediction error (average standard error - ASE) and in the square root of the mean of the differences between the estimated values and the true values standardized (root mean square standardized - RMSS).

For predictions that represent the real value, the ME and the MSE must have values close to zero, while the RMS and ASE have values close to each other, which reduces the uncertainty associated with the predictions. For analysis of the RMSS, if the value is greater than 1 the variability of the prediction is being underestimated, otherwise it is overestimated, so the ideal value is as close as possible to 1 (TATALOVICH et al., 2006; GEORGAKARAKOS; KITSIOU, 2008; PASINI et al., 2014).

For the analysis of the estimated values, the 50 points randomly collected were used and compared with the estimated pH values. For this comparison, the root mean square error (RMSE) equation was used, being expressed by Equation 5.

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (E_i - O_i)^2} \quad (Equation 5) \]

Where:
\[ RMSE = \text{root mean square error,} \]
\[ E_i = \text{estimated value by the interpolation method,} \]
\[ O_i = \text{value of the real observed variable and} \]
\[ n = \text{number of observations.} \]

In addition to the RMSE, the agreement index (d) was calculated, which expresses a measure of the degree to which the estimated data are or are not free from certain errors (Equation 6). This index ranges from 0 to 1, with 1 indicating a perfect agreement between the observed and estimated values and 0 means total disagreement (WILLMOTT, 1981).

\[ d = 1 - \frac{\sum_{i=1}^{n} (E_i - O_i)^2}{\sum_{i=1}^{n} (|E_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (Equation 6) \]

Where:
\[ d = \text{agreement index,} \]
\[ E_i = \text{estimated value by the interpolation method,} \]
\[ O_i = \text{value of the observed real variable,} \]
\[ n = \text{number of observations and} \]
\[ \bar{O} = \text{arithmetic mean of the observed values.} \]

The average soil pH values were also calculated due to the soil classes present in the municipality of Sorocaba, considering the pedological map of the São Paulo State prepared by the Forestry Institute (ROSSI, 2017) on a scale of 1:100,000 and 1:250,000. For this, the vector file available was cut using the limit of the municipality of Sorocaba as a mask, and the average soil pH value was subsequently calculated for the two spatial distribution methods of soil pH.

RESULTS AND DISCUSSION

The analyzed parameters using the exponential model are shown in Table 1, being possible to observe that the soil pH presented a strong spatial dependence through the calculation of the SDI, which resulted in less than 25% (CAMBARDELLA et al., 1994). Among the theoretical models used, the exponential had best adjusted to the values of ME and MS close to zero and with greater similarity of RMS and ASE, which reduces the uncertainty associated with the predictions. The RMS value was low, indicating low variability between the predicted and true values. The RMSS value found was close to the ideal equal to 1 (TATALOVICH et al., 2006; GEORGAKARAKOS; KITSIOU, 2008; PASINI et al., 2014). Figure 3 shows the experimental semivariogram of soil pH, adjusted to the theoretical model mentioned in Table 1.
TABLE 1 - Parameters of geostatistical analysis of soil pH distribution.

<table>
<thead>
<tr>
<th>Model</th>
<th>$C_0$</th>
<th>C</th>
<th>R (m)</th>
<th>SDI (%)</th>
<th>ME</th>
<th>MSE</th>
<th>RMS</th>
<th>ASE</th>
<th>RMSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>0.00</td>
<td>0.157</td>
<td>4700</td>
<td>0.00%</td>
<td>-0.009</td>
<td>-0.019</td>
<td>0.349</td>
<td>0.350</td>
<td>0.999</td>
</tr>
</tbody>
</table>

$C_0 =$ nugget effect; C = contribution; R (m) = range; SDI (%) = spatial dependency index; ME = mean error; MSE = mean standardized prediction error; RMS = root mean square; ASE = average standard error; RMSS = root mean square standardized.

FIGURE 3 - Semivariogram of the soil pH sample.

Figure 4 presents the map with the spatial distribution of soil pH in classes by the ordinary kriging interpolation method and by the ISD method. It is noticed that for both methods, the spatial distribution followed the same pattern of distribution. The distribution of pH values for the municipality, for both methods applied, is predominantly in the range between 6.0 and 6.5, since the ISD and ordinary kriging presented 6.24 of average value, with deviation standard equal to 0.2307 for the ISD method and 0.2336 for the ordinary kriging method.

FIGURE 4 - Maps showing the pH distribution of soil samples.

Table 2 shows the area and the respective percentage of each pH value classes. As noted, the areas of each interval showed low variation between the methods. The same can be observed in the calculation of the root mean square error for both methods, which presented values of 0.349951 for ordinary kriging and 0.349019 for ISD, representing a percentage variation of only 0.27% in relation to accuracy by the ISD method. For both methods, the agreement index (d) was equal to 0.9944, demonstrating that there are no significant variations between the methods used, and the value is close to 1, that is, there is perfect agreement between the observed values and the estimated ones (WILLMOTT, 1981).

The coefficient of variation obtained for the pH values of the study area, which expresses the relationship between the standard deviation and the mean, was 0.0374 by the ISD method and 0.0369 for ordinary kriging. According to Robinson and Metternicht (2006), the smaller is the coefficient of variation the more homogeneous the
data are, as observed in the pH distribution maps using the ISD and ordinary kriging method presented in Figure 4. The ISD and ordinary kriging methods are widely used in comparative studies that use the interpolation of data from sociodemographic and demographic analyzes, in which kriging presents more satisfactory results (RIGHI; BASSO, 2016). For Robinson and Metternicht (2006), kriging also showed better results for the interpolation of the pH and electrical conductivity data in the soil, while the ISD presented the best results for the pH interpolation in the subsoil. However, the authors conclude that there is not a single method of interpolation that suits the properties of the soil.

Souza et al. (2010) found that for the spatialization of chemical attributes of the soil, the ordinary kriging and the inverse of the distance square are efficient in the inference of the non-sampled values, and do not present great variations in the results, as observed in the present study of the spatial distribution in the municipality of Sorocaba.

**TABLE 2 - Areas of the pH range sections, by ISD and kriging methods.**

<table>
<thead>
<tr>
<th>pH</th>
<th>ISD Area (km²)</th>
<th>ISD Area (%)</th>
<th>Kriging Area (km²)</th>
<th>Kriging Area (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0 – 5.5</td>
<td>3.34 (0.74%)</td>
<td></td>
<td>3.95 (0.88%)</td>
<td></td>
</tr>
<tr>
<td>5.5 – 6.0</td>
<td>101.62 (22.64%)</td>
<td></td>
<td>102.66 (22.86%)</td>
<td></td>
</tr>
<tr>
<td>6.0 – 6.5</td>
<td>340.90 (75.93%)</td>
<td></td>
<td>336.61 (74.98%)</td>
<td></td>
</tr>
<tr>
<td>6.5 – 7.0</td>
<td>3.08 (0.69%)</td>
<td></td>
<td>5.73 (1.28%)</td>
<td></td>
</tr>
</tbody>
</table>

Studies that address the spatial variability of the potential of hydrogen (pH) are considered extremely important, since they quickly provide information of soil characteristics. According to Ribeiro et al. (1999), the pH ranges for agronomic classification purposes have pH values below 4.5 classified as very low; values between 4.5 and 5.4 are considered low; pH between 5.5 and 6.0 is considered good; between 6.1 and 7.0 it is considered high; and a pH greater than 7.0 is considered to be very high.

According to the Soil Fertility Commission of Minas Gerais State, for agronomic classification, pH values ranging from good to high are considered adequate (RIBEIRO et al., 1999), as these ranges require little or no pH correction measures for agriculture (RAIJ et al., 2001). Thus, it can be said that the soil in the municipality of Sorocaba does not need pH correction, since for each adopted method, more than 99% of the soil is in the range of good to high, with no observed values greater than 7.0 (very high).

Figure 5 shows the average values of soil pH depending on the soil classes present in the municipality of Sorocaba. The average values of the pH depending on the soil classes presented a variation of 5.9 to 6.3. According to the classification by Ribeiro et al. (1999), the values found are in the range of good and high, with Haplic Gleysol being the only one classified as good. These soils are hydromorphic and, therefore, are usually in areas subject to periodic flooding because they have low drainage capacity.

**FIGURE 5 - Average values of soil pH by soil classes.**

The highest pH mean value found by soil class is present in the region with the largest forest fragments in medium and advanced successional stages, as Smith et al. (2014), and classified as Haplic Cambisol, and according to Spera et al. (2011), are considered dystrophic, presenting base saturation below 50%, which gives it acidic characteristics.
The Red-Yellow Latosol and Red Latosol presented pH mean values of 6.0. According to Spera et al. (2011), the Latosols are characterized by having in their granulometric fraction about 14% to 24% clay and having good permeability. However, for the Red-Yellow Acrisol, which had the second-highest average pH value (6.1), they are characterized by presenting an average texture ranging from 36% to 50% of clay fractions, and good drainage capacity.

According to Sun et al. (2013), the effect of urbanization is increasing worldwide and, as a result, soils have been undergoing slight changes pH values post-urbanization, a fact corroborated by the present study, since the soil class of the Urban Area presented an average pH value of 6.2, being slightly different from the found in the surroundings, suggesting that the value found may be associated with the accelerated urbanization process that the municipality has been going through, however, it is suggested to use other deterministic and geostatistics methods, to verify the adjustment of spatial pH variability in the municipality.

CONCLUSIONS

The distribution of pH values in the municipality of Sorocaba was predominant in the range between 6.0 and 6.5; both by the ISD method and by the ordinary kriging method, covering, respectively, 75.93% and 74.98% of the municipality. The pH range between 5.5 and 6.0 covered 22.64% by ISD and 22.86% by ordinary kriging.

The pH values estimated by the ordinary kriging and ISD methods came close to the values obtained in loco, as expressed by the values of the Root Mean Square Error (RMSE) of 0.349951 for ordinary kriging, and 0.349019 for the ISD method. The agreement index for both methods was 0.9944. Therefore, both theoretical models applied in the interpolation of the municipality’s pH data efficiently explain the variability of soil pH in the municipality of Sorocaba.

The average pH values for the different classes of soils present in the municipality of Sorocaba presented a variation between 5.9 and 6.3, being inserted in the range of good and high values, therefore, they need little or no application of acidity correction techniques for agricultural production.

REFERENCES


Application of geostatistical...


