Spatial variability of soil organic matter content in Eastern Croatia assessed using different interpolation methods

Boris Đurđević1*, Irena Jug1, Danijel Jug1, Igor Bogunović2, Vesna Vukadinović1, Bojan Stipešević1, and Bojana Brozović1

1Faculty of Agriculture in Osijek, Josip Juraj Strossmayer University of Osijek, Vladimira Preloga 1, 31000 Osijek, Croatia
2Faculty of Agriculture, University of Zagreb, Svetošimunska cesta 25, 10000 Zagreb, Croatia

Received March 22, 2018; accepted August 28, 2018

Abstract. Soil organic matter plays a crucial role in soil health and represents one of the key functions for determining soil suitability for crop production. Recently, intensive agricultural production and climatic changes have led to a decline in organic matter level in soils. This paper is to provide the most accurate spatial predictor using different interpolation methods in order to evaluate in detail the status of organic matter in agricultural soils in the Osijek-Baranja County, Croatia. We applied three different interpolation methods, including inverse distance weighting, ordinary kriging and empirical Bayesian kriging. A total number of 9099 soil samples from 0-30 cm layer were compiled and analysed in the laboratory. The average value of soil organic matter in the study area was 2.66% with moderate variability (CV = 30.62%). The best fit variogram model is exponential in the direction of 20 and its spatial variability indicates that soil organic matter varies widely under pedogenetic and soil management practices. Empirical Bayesian kriging method was the most precise (RMSE = 0.457), followed by ordinary kriging (RMSE = 0.466) and inverse distance weighting (RMSE = 0.476). The investigated area shows a heterogeneous spatial pattern of soil organic matter content, with levels below 3% found mostly in western and south-western parts of county.

Keywords: geostatistics, soil organic matter, kriging, interpolation, soil quality

INTRODUCTION

The soil represents a crucial irreplaceable resource for food production (Rojas et al., 2016). Growing population, together with pronounced climate changes put a considerable pressure on agricultural land (FAO and ITYS, 2015). Furthermore, intensive agricultural production (e.g. annual ploughing and use of agrochemicals) leads to soil degradation, decreasing its suitability for crop production (Liu et al., 2010, 2006; FAO and ITYS, 2015). The key function in relation to soil degradation and soil fertility is played by soil organic matter (SOM) content (Gajda et al., 2016; Lehmann and Kleber, 2015; Vázquez et al., 2016). The basic benefits of SOM can be divided into three categories – physical (increases aggregate stability and soil water capacity, and decreases crusting), chemical (increases cation exchange capacity of soil and availability of essential soil nutrients) and biological (provides habitat and food for many living organisms in soil, increases microbiological diversity and expands soil food webs) (Bot and Benites, 2005). Furthermore, SOM represents the most important sink of carbon and plays a crucial role in carbon sequestration, mitigating climate change effects (Milne et al., 2007; Whitmore et al., 2015). The continuing decline of organic matter from soils in agroecosystems due to unappropriated agricultural (incineration and removal of crop residues, overgrazing, inappropriate tillage, etc.) and environmental conditions (rising temperature or heat wave events, frequent floods, erosion, etc.) is the main problem for soil preservation (Jug et al., 2018; Farquharson et al., 2003; Liu et al., 2006; Loveland and Webb, 2003). Therefore, it is very important to exert permanent control over SOM content in soils within agricultural land (Bot and Benites, 2005). Procedures which would facilitate the analysis of the researched area together with an accurate visualization of SOM by geostatistical methods are still unexplored in Croatia. The existence of proper input data enables us to predict the status of unsampled areas and to understand

*Corresponding author e-mail: bjurdevic@pfos.hr

© 2019 Institute of Agrophysics, Polish Academy of Sciences
problems in agricultural production (Krivoruchko, 2012; Lipiec and Usowicz, 2018; Mirzaei and Sakizadeh, 2016; Robinson and Metternicht, 2006; Schueller, 2010). At the regional or national levels, the precise monitoring of SOM could be a labour-intensive, time-consuming and expensive measure, but today’s technologies (e.g. GIS, geostatistics) make it possible to monitor spatio-temporal changes of all chemical properties of soil at various levels. There are many scientific papers that are dealing with spatial prediction of SOM content and predictions in those papers are made at different levels, namely field (Mabit and Bernard, 2010), regional (Dai et al., 2014) and country/continental (Jones et al., 2005). When it comes to the methods used for spatial analyses of SOM, the most frequently used are the deterministic methods, such as inverse distance weighting (IDW) and Splines that do not take into consideration the spatial autocorrelation of data, and they are followed by stochastic geostatistical methods that take into account an estimation of every variable at each point through the selected model, like ordinary kriging (OK) (Mabit and Bernard, 2010; Marchetti et al., 2012; Robinson and Metternicht, 2006). Not long ago, empirical Bayesian kriging (EBK) has become an interesting alternative for mapping soil properties. Unlike classic kriging methods, it automatically estimates and, by using a distribution of semivariogram models instead of just one model, calculates unsampled locations compiled into an image representing a specific area. For example, OK uses only one semivariogram model for calculation and its accuracy significantly depends on the analyst’s skills (Krivoruchko, 2012; Mirzaei and Sakizadeh, 2016). An approach with multiple simulated variograms could increase their reliability, when their parameters are used for kriging (Fabijańczyk et al., 2017). Several other studies showed an increased mapping precision by EBK (e.g. Adhikary et al., 2011; Fabijańczyk et al., 2017). Considering the mentioned importance of SOM content, this study tests a new interpolation method to 1) describe the spatial variability of SOM, 2) compare and choose the most suitable method for SOM spatial prediction, 3) create a map by using geostatistical and GIS tools to detect areas with depleted SOM content status.

MATERIAL AND METHODS

The research was conducted in the Osijek-Baranja County (4152 km²), Eastern Croatia. The Osijek-Baranja County is settled in lowlands and hilly parts of Eastern Croatia (Bašić et al., 2007). The temperatures in the researched area increase from west towards the east, and from northwest to northeast and they are dependent on the influence of the sea. The mean annual temperature is 10°C. Mean monthly temperature ranges from -1 to 21°C, with the coldest period in January when the minimum temperature may reach -25°C or lower. On the other hand, the warmest period is in July and August when maximum temperature exceeds 40°C. Precipitation in the research area increases inversely to air temperature. The mean annual rainfall varies from east (609 mm) to west (792 mm). The main soils types of the Osijek-Baranja County, found in about 85% of the area, are: Gleysols (28.8%), Luvisol (25.9%), Stagnosols (14.9%), Fulvisols (7.4%), Chernozem (6.0%), and Cambisols (1.8%) (Bogunović et al., 1996).

9099 soil samples, with associated latitude and longitude, were compiled for the study area (Fig. 1). Soil samples were collected with an agrochemical and hydraulic automatic probe. One composite sample consisted of 25 cores collected within a 15 m radius of the main sample point with the sample recovery depth of 30 cm. Soil samples were taken after the harvest, before application of any kind of fertilization and prior to soil preparation for new crops (Đurđević, 2014).

Fig. 1. Study area map including the distribution of sampling points in the Osijek-Baranja County.
Soil samples were stored in plastic bags, homogenized, dried, milled and analysed in the laboratory according to the modified Walkley-Black method (Nelson and Sommers, 1982). Calculation of SOM was made under the assumption that the organic matter contained 50% of carbon (Pribyl, 2010).

Descriptive statistics, with arithmetic mean, standard deviation (SD), minimum (min), maximum (max), coefficient of variation (CV), kurtosis (kur) and skewness (skew) were computed for SOM on the assumption that the data were spatially independent. The descriptive parameters were calculated by Microsoft Excel for Windows. Prior to data analysis, SOM dataset was checked for normality using Kolmogorov-Smirnov test. Data normality is desirable for the correct interpretation of spatial interpolation, because it may have some implications for geostatistical analysis (Kerry and Oliver, 2007). If SOM dataset did not follow normal distribution (p < 0.05) it was subjected to logarithm and Box-Cox transformations (Box and Cox, 1964; Osborne, 2010). If the transformed dataset did not achieve normality after performed transformations, the criterion for choosing the best method was normalized skewness in range of values between -0.5 and 0.5 (Bogunovic et al., 2017a).

Spatial analysis of SOM was conducted by spatial tools such as classed post maps, directional variograms and variogram maps (Fotheringham and Rogerson, 2013; Hengl et al., 2004; Webster and Oliver, 2001). After studying the datasets searching for extremes, outliers, local errors and spatial non-stationarity, the following procedure had to be observed: the spatial correlation of SOM with experimental variogram modelling in ArcGIS software (ESRI, 2011). Based on the present sampling geometry and distances between the samples, semivariance was calculated using 12 lags and a maximum lag distance of 200 m related to the average spacing of the data.

Semivariograms were checked for anisotropy to calculate the directional semivariograms. According to previous works, 150 samples were required for a reliable identification with regard to the presence of anisotropy (Webster and Oliver, 2001). The spatial dependence of variables was assessed using the nugget effect and nugget/sill ratio. A ratio < 25% shows a strong dependence, 25-75% a moderate dependence, and > 75% a weak dependence (Cambardella et al., 1994). The analysis of classed post map and variogram detects the presence of trends. Best fit variogram model was selected on the basis of visual assessment, lowest nugget effect, nugget/sill ratio and longest range of spatial autocorrelation. Any uncertainties between two or more similar models were resolved by choosing the variogram with the lowest root mean square error (RMSE) during cross-validation procedure.

In order to test and analyse spatial variability and provide the most accurate map of SOM, three interpolation methods were used: inverse distance weighting (IDW), ordinary kriging (OK) and empirical Bayesian kriging (EBK). Inverse distance weighting prediction is a deterministic, nonlinear interpolation technique, which on the basis of the average known values of neighbouring points calculates prediction of unsampled location (Xie et al., 2011, Eq. (1)):

$$ Z(x) = \frac{\sum_{i=1}^{n} w_{ij} Z(x_i) d_{ij}^p}{\sum_{i=1}^{n} d_{ij}^p}, $$  \hspace{1cm} (1)

where: $Z(x)$ is an estimated value for a particular point, $Z$ represents known point value, and $r$ is weight, which is associated with the distance between the estimated and the data points $d_{ij}$.

Ordinary kriging (OK) uses a weighted average of known neighbouring values, which depends on the distance between them, their grouping and their values, to estimate the unsampled location. OK is similar to IDW and it aims to estimate the value of random function $z$ for one or more unsampled locations (Eq. (2)):

$$ z^*(x_0) = \sum_{i=1}^{n} w_i Z(x_i), \hspace{1cm} (2) $$

where: $w_i$ are the weights assigned to the known value of $Z(x_i)$, and $z^*(x_0)$ is the estimated value (Xie et al., 2011).

Empirical Bayesian kriging (EBK) assesses the unsampled locations by automatically applying several different semivariogram models, unlike OK which uses only one fitted semivariogram model for the entire area. This process is divided into three main steps: (1) Semivariogram model is estimated on the basis of data. (2) Using the resulting semivariogram model, the new values are then simulated at the locations of the input data. (3) A new semivariogram model is estimated from the simulated data. A weight for this semivariogram is then calculated by using Bayes’ rule, which shows the probability with which the observed data can be generated from the semivariogram. Rules 2 and 3 are repeated, and with each repetition the estimated semivariogram in step 1 is used to simulate a new set of values at input locations. The simulated data are then used to estimate a new semivariogram model and its weight. This process is fully automated and allows the creation of spectrum semivariograms, which can be used to assess a particular area (Krivoruchko, 2012; Mirzaei and Sakizadeh, 2016).

The assessment of the interpolation methods was carried out using the leave-one-out cross-validation method, which estimates the sampling point from the surrounding samples. From the errors produced (observed-predicted) we calculated the mean error (ME) and the root mean square error (RMSE) according to the following formulas:

$$ ME = \frac{1}{N} \sum_{i=1}^{n} (z(x_i) - z^*(x_i)), $$ \hspace{1cm} (3)

$$ RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (z(x_i) - z^*(x_i))^2}, $$ \hspace{1cm} (4)

where: $z^*(x_i)$ is the predicted value, $z(x_i)$ the known value, and $N$ the number of samples (Robinson and Metternicht, 2006). The most accurate method is the one with the lowest RMSE.
RESULTS AND DISCUSSION

The mean value of SOM in the study area was 2.66% (Table 1), while 70.7% of the analysed soil samples had a content of SOM lower than the 3% (Fig. 2) which indicates the presence of agricultural soil degradation (Liu et al., 2006) and can lead to a decline in soil quality (Loveland and Webb, 2003). The value of CV (30.62%) for SOM in the investigated area reveals their moderate variability (according to Nielsen and Bouma, 1985). Other studies also reported similar findings (e.g. Bogunovic et al., 2017a; Jiang et al., 2012). Moderate variability of SOM is the result of the topography, climate and soil type, as well as land use. The investigated area has different geologic, geomorphologic and climatic conditions which impacts on the diversity of soil types, while intensive soil management increases the spatial heterogeneity of SOM. Hilly areas are covered with permanent plantations and intensively tilled which accelerates SOM depletion (Paltineanu et al., 2016) and losses of SOM through erosion (Bogunovic et al., 2018). Lowland areas have unequal water regime. Areas close to rivers where groundwater is shallow record high SOM (Ceddia et al., 2015). Nevertheless, the amelioration of these soils for agricultural purposes has an impact on the duration of the conditions of pedogenesis and increases the heterogeneity of SOM. Frequently flooded soils usually record higher SOM content compared to soils located at higher elevations due to poor decomposition of organic material (Eglin et al., 2008).

Soil organic matter data show a non-normal distribution by an evident positive skewness (Table 1). Performed logarithmic and Box-Cox transformations reduce CV and shape soil parameters, although the transformed datasets did not pass Kolmogorov Smirnov test (p < 0.01). However, for modelling purposes, we used logarithmic transformed SOM data because this was the closest to a normal distribution following procedure as in previous studies (e.g. Bogunovic et al., 2017b; Sharma et al., 2011).

The SOM experimental variogram (Fig. 3) was best-fit with the exponential model, which is in agreement with previous studies (Bogunovic et al., 2017a; Lipiec and Usowicz, 2018; Reza et al., 2016) where the majority of

Table 1. Summary statistics for soil organic matter (SOM)

<table>
<thead>
<tr>
<th>Attributes</th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>CV %</th>
<th>Kurtosis</th>
<th>K-S p</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOM</td>
<td>9099</td>
<td>1.02</td>
<td>6.99</td>
<td>2.66</td>
<td>0.91</td>
<td>1.12</td>
<td>30.62</td>
<td>4.19</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Log_SOM</td>
<td>9099</td>
<td>-0.06</td>
<td>0.82</td>
<td>0.34</td>
<td>0.14</td>
<td>0.42</td>
<td>41.59</td>
<td>-0.37</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>BC_SOM</td>
<td>9099</td>
<td>-1.14</td>
<td>-0.15</td>
<td>-0.48</td>
<td>0.15</td>
<td>-0.26</td>
<td>30.52</td>
<td>-0.38</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

N – number of samples, SD – standard deviation, CV – coefficient of variation, LogSOM – logarithmically transformed SOM, BC_SOM – box-cox transformed SOM.

Fig. 2. Levels of soil organic matter content.

Fig. 3. Directional semivariogram model calculated for soil organic matter with log transformed data. Bold line represents the best fit model.
soil properties were best fitted with the exponential model. Range of spatial correlation for SOM was high (1892 m) in the direction of 20. This range indicates a proper sampling number for mapping of SOM in the investigated area. According to Kerry and Oliver (2004), the sample interval should be less than half of the variogram range. Our data confirmed that the sampling interval was appropriate to measure the spatial variability of SOM. Nugget/sill effect was 34.27%, indicating moderate spatial dependence. According to Cambardella et al. (1994) soil spatial dependence is controlled by intrinsic and extrinsic factors. In the present study, according to variogram modelling, SOM is highly influenced by human activity (Liu et al., 2014), which is also confirmed by quite high nugget effect. The presence of nugget indicates a short-range spatial variability, probably related to the influence of human activities strongly connected with agricultural practices.

The results of the geostatistical techniques tested for SOM are shown in Table 2. Among the tested geostatistical techniques, the most accurate technique for mapping SOM was EBK, while the least accurate was IDW. Considering IDW, the most accurate model parameters for SOM assessment are found to be at weight power of one with 10 neighbouring points included. Modelling with this method reveals that an increase in the weight parameter decreases the mapping accuracy. Kriging estimation of unsampled locations depends highly on variogram modelling performance.

An optimal geostatistical method should have a minimum cross-validation error, which depends mostly on the number and distribution of the samples and their distance from one another (Mirzaei and Sakizadeh, 2016). The lowest RMSE was recorded for EBK. Hence, the mean error (ME) which is very close to the ideal 0, proves that the prediction is relatively unbiased with a small bias. Only the IDW method has a considerably higher ME and consequently stronger bias.

In general, the IDW method is simpler and there are not so many different possibilities regarding the selection of the model parameters, unlike OK that needs more skill or perhaps more artistic approach (Gong et al., 2014), because semivariance function fitting process is very subjective (Xie et al., 2011). Of course, OK has already proven to be one of the best methods for estimation, especially when it comes to the prediction of soil chemical properties (Bogunovic et al., 2014). The main advantage of EBK is that the process of creating a valid kriging model is automated so manual parameter adjustment is eliminated (Krivoruchko, 2012). In addition, EBK takes into account uncertainties related to variogram plotting, simulating variograms for different parts of the analysed area, and creating separate subpopulations of data. This plays an important role when parts of the analysed area are affected by different factors (different soil management, fertilizer rates, tillage systems, etc.) (Samsonova et al., 2017). Finally, the advantage of EBK in comparison to other models has already been reported during mapping soil contamination (Adhikary et al., 2011), water contamination (Fabijańczyk et al., 2017), and mapping of forest soil properties (Beguin et al., 2017).

Final maps of SOM are presented in Fig. 4. The areas of low SOM values lower than 3.0% can be identified mainly in western and south-western parts of the Osijek-Baranja County, which mostly has very acid soil reaction (predominantly Stagnosols) with only several neutral zones (Luvisol) (Đurđević et al., 2011) (Fig. 4c). Furthermore, a vast majority of these soils (90%) are generally tilled conventionally (deep mouldboard ploughing) (Jug et al., 2007), with insufficient care about the incorporation of crop residues (Jug et al., 2011). In addition, a large part of crop residues is used for the production of biofuels and other biomass energy (Bilandzija et al., 2018), which is a good precondition for losing large amounts of SOM. The described open cycle system can have a significant impact on SOM concentration that results in a decrease in SOM. This has been confirmed by many studies for various agroecosystems (Mann et al., 2002; Kravchenko et al., 2016; Busari et al., 2015; Durán Zuazo et al., 2014; Bot and Benites, 2005). The only exceptions are the north and south-eastern areas which mostly have neutral soil reaction and they belong to the county areas which, according to Đurđević et al. (2011) and Vukadinović et al., (2014), are more suitable for agricultural production (predominantly Gleysols, Chernozem and Luvisol) (Fig. 4c). But agricultural practice in the research area is mostly intensive with predominant use of mineral fertilizers, while farmyard manure is mostly spread in close vicinity of a few larger farms. Consequently, one could easily speculate that in the near future a similar situation with a decrease in SOM concentrations will also affect areas with more than 3% of SOM. In this context, it is necessary to highlight the great importance of preserving and raising SOM content through good agricultural practices, use of conservation soil management with well-planned crop rotation and optimal fertilization recommendations (mineral and organic) based on expert knowledge (Bot and Benites, 2005; Shrestha et al., 2015).

### Table 2. Efficiencies and errors of the interpolation methods (summary)

<table>
<thead>
<tr>
<th>Method</th>
<th>Neighbours</th>
<th>ME</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBK</td>
<td>20</td>
<td>0.003</td>
<td>0.457</td>
</tr>
<tr>
<td>OK</td>
<td>5</td>
<td>0.001</td>
<td>0.466</td>
</tr>
<tr>
<td>IDW</td>
<td>10</td>
<td>0.011</td>
<td>0.476</td>
</tr>
</tbody>
</table>

IDW – inverse distance weighting, OK – ordinary kriging, EBK – empirical Bayesian kriging, ME – mean error, RMSE – the root mean squared error.
Fig. 4. Spatial distribution maps of soil organic matter (SOM %) content calculated with a) IDW, b) OK, and c) EBK.
CONCLUSIONS

1. The investigated soils had low soil organic matter concentration indicating possible future treats in the form of land degradation.

2. The best fit model for soil organic matter is the exponential one. The spatial dependence was moderate and the range was high, revealing that soil organic matter varied widely under the pedogenetic and soil management practices, indicating a strong human influence.

3. The most accurate interpolation method was empirical Bayesian kriging, while inverse distance weighting was the least accurate. The map of SOM indicates low levels (< 3%) in the western and south-western parts of the investigated area. Heterogeneous spatial patterns of soil organic matter suggest that site-specific soil management strategies should be implemented.

4. This kind of approach to problem solving in agriculture can be applied for various agroecological conditions and can significantly facilitate or accelerate the decision-making process, and thus directly affect the profitability and sustainability of agricultural production.

Conflict of interest: The Authors do not declare any conflict of interest.

REFERENCES


Bogunović M., Vidaček Č., Racz Z., Husnjak S., and Sraka M., 1996. Soil map of Croatia at the scale 1: 300 000. Department of Soil Science, Faculty of Agriculture, University of Zagreb, Croatia.


