Precision Agric (2017) 18:882–897
DOI 10.1007/s11119-016-9481-6

BRIEF COMMUNICATION

Estimating soil organic matter using interpolation methods with a electromagnetic induction sensor and topographic parameters: a case study in a humid region

Aitor García-Tomillo[1] · José Manuel Mirás-Avalos[1,2] · Jorge Dafonte-Dafonte[3] · Antonio Paz-González[1]

Published online: 14 November 2016

Abstract Soil organic matter (SOM) is a key indicator of soil quality although, usually, detailed data for a given area is difficult to obtain at low cost. This study was conducted to evaluate the usefulness of soil apparent electrical conductivity (EC_a), measured with an electromagnetic induction sensor, to improve the spatial estimation of SOM for site-specific soil management purposes. Apparent electrical conductivity was measured in a 10-ha prairie in NW Spain in November 2011. The EC_a measurements were used to design a sampling scheme of 80 locations, at which soil samples were collected from 0 to 20 cm depth and from 20 cm to the boundary of the A horizon (ranging from 25 to 48 cm). The SOM values determined at the two depths considered were weighted to obtain the results for the entire A Horizon. SOM distribution maps were obtained by inverse distance weighting and geostatistical techniques: ordinary kriging (OK), cokriging (COK), regression kriging either with linear models (LM-RK) or with random forest (RF-RK). SOM ranged from 46.3 to 78.0 g kg^{-1}, whereas EC_a varied from 6.7 to 14.7 mS m^{-1}. These two variables were significantly correlated ($r = -0.6$, $p < 0.05$); hence, EC_a was used as an ancillary variable for interpolating SOM. A strong spatial dependence was found for both SOM and EC_a. The maps obtained exhibited a similar spatial pattern for SOM; COK maps did not show a significant improvement from OK predictions. However, RF-RK maps provided more accurate spatial estimates of SOM (error of predictions was between four and five times less than the other interpolators). This information is helpful for site-specific management purposes at this field.

✉ José Manuel Mirás-Avalos
jmirasa@udc.es; jmmiras@cebas.csic.es

[1] Área de Edafología y Química Agrícola, Facultad de Ciencias, Universidade da Coruña, Campus A Zapateira s/n 15008, A Coruña, Spain

[2] Present Address: Departamento de Riego, Centro de Edafología y Biología Aplicada del Segura (CEBAS-CSIC), Campus Universitario de Espinardo, 30100 Murcia, Espinardo, Spain

[3] Departamento de Ingeniería Agroforestal, Escuela Politécnica Superior de Lugo, Universidade de Santiago de Compostela, Campus de Lugo, 27002 Lugo, Spain

Keywords Cokriging · Geostatistics · Ordinary kriging · Random forest · Regression kriging · Soil quality

Introduction

Soil organic matter (SOM) has great influence on soil physical, chemical, and biological processes. Reduction of SOM levels will result in a decrease of soil fertility, soil nutrient supply, porosity, and consequently, soil productivity (Gray and Morant 2003; Lozano-García et al. 2011). In addition, the rate of SOM loss can vary greatly, depending on cultivation practices, type of crop cover, soil drainage status and weather conditions. Frequently, soil sampling is time-consuming and expensive and, usually, a suitable amount of SOM data is very difficult to obtain (Brevik et al. 2016). This high cost of collecting SOM data through dense sampling across landscapes has created a need for methods of inferring its spatial distribution. Thus, the description of SOM spatial variability through maps obtained by different interpolation methods would be useful for site-specific management (Miller et al. 2015). Soil organic matter maps provide decision makers with a useful tool to identify degraded areas and optimize agro-environmental measures (Piccini et al. 2014). These methods include, for instance, inverse distance weighting (IDW) (e.g. Bregt et al. 1992), several kriging techniques (e.g. Baxter and Oliver 2005; Bishop and Lark 2006; Nerini et al. 2010; Hoffmann et al. 2014; Chen et al. 2015), generalised linear models (e.g. Pachepsky et al. 2001; Dobson and Barnett 2008), and regression trees (e.g. Rudiyanto et al. 2016).

These interpolations are susceptible to improve by using ancillary information. In this sense, electromagnetic induction (EMI) sensors can provide useful data on the spatial variation of certain soil properties and patterns within a field (King et al. 2005) and have been used to support soil surveys and site-specific management (Brevik et al. 2012). Soil apparent electrical conductivity (EC_a) measured using geophysical methods, such as EMI, has important advantages over traditional methods used to collect soil information because of its speed, easy use, relatively low cost, and volume of data collected (Doolittle and Brevik 2014). However, EMI techniques are site-specific, namely the relationships between a given soil property and EC_a have to be determined on a field-by-field basis. This fact diminishes the hope for EMI to be a rapid, inexpensive and widely applied method of soil exploration as thought about 20 years ago (Brevik et al. 2016). Nevertheless, EC_a might improve the accuracy and reliability of maps and provide more detailed information on soil properties when the correlation between this variable and soil attributes has been established (Sudduth et al. 2005; Brevik et al. 2012) and has been widely used as a secondary variable in order to improve spatial estimations (Vitharana et al. 2006).

Geostatistical interpolation techniques are preferred respect to deterministic methods because the latter account neither for estimation errors nor for the spatial autocorrelation of data (Robinson and Metternicht 2006). Geostatistical methods have been extensively applied in agriculture for the study of spatial variability of the main soil attributes (e.g. Goovaerts 1999; Paz-González et al. 2000). Ordinary kriging (OK) is the most common type of geostatistical interpolation; however, the quality of the estimation of the soil properties can be improved and the spatial sampling intensities may be reduced by incorporating ancillary information. In order to incorporate this auxiliary information, a variable more extensively sampled over the studied field should be used through other geostatistical interpolation methods such as cokriging (COK) (Goovaerts 1999), external-drift kriging (Nussbaum et al. 2014) and regression kriging (RK) with a linear regression (LM-RK) model (Zhang et al. 2012) or with a non-linear regression model—random forest (RF-RK) (Guo et al. 2015; Rudiyanto et al. 2016). In these cases, topographical (terrain height, slope, aspect, etc.) and categorical parameters (soil type, texture, etc.) are commonly used as ancillary information.

In this context, the aim of this study was to assess the spatial distribution of SOM in an agricultural field (in NW Spain) through different interpolation techniques (inverse distance weighting, OK, COK using EC_a as an ancillary variable and RK using secondary information from topographic data and EC_a, in order to enhance the estimations. The produced maps could be useful for attaining a site-specific management of agricultural inputs because distinct areas within the field can be observed and farmers can act in consequence when, for instance, applying fertilizers.

Materials and methods

Location of the study site

A 10-ha grassland site at Castro Riberas de Lea in NW Spain (43°16′14″N, 7°49′20″W, 403 m above sea level) was selected for this research (Fig. 1). This field is characterized by small variations in topography, with a maximum height difference of about 1.10 m. The soil of the site was developed on Quaternary material rich in gravel over clay Tertiary material and was classified as Umbric Fluvisol according to FAO (IUSS Working Group

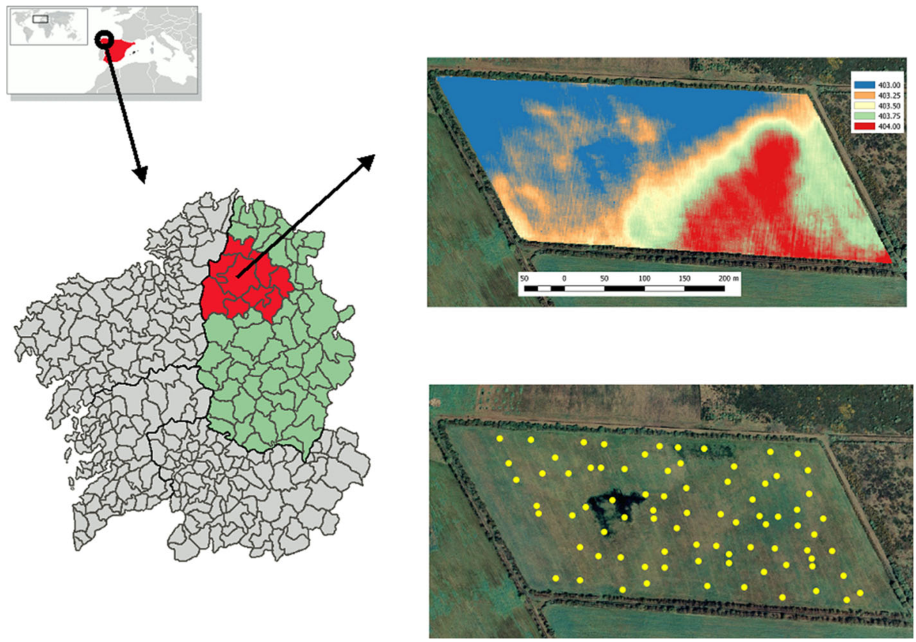


Fig. 1 Location of the experimental site within Galicia, topography of the plot and sampling points for soil organic matter

WRB 2014). Soil is sandy-loam to sandy textured, acidic (pH 4.4), with very low salinity and the field was devoted to grassland. The climate of the study region is humid and temperate with warm summers, Cfb according to the Köppen climate classification (Köppen 1936). According to the data provided by a weather station less than 100 m away from the study site, the mean temperatures (30-year average) during winter and summer are 6.5 and 16.2 °C, respectively. Moreover, mean annual rainfall is about 869 mm (30-year average), with a dry season (June–July–August). During this dry period, rainfall is about 98 mm (30-year average); however, in 2011 it was only 42 mm. In November 2011, when soil samples were collected, mean temperature was 9.7 °C and the monthly rainfall amount was 68 mm.

A digital terrain model (DTM) (1 m cell size) was generated from LIDAR data (0.5 points m^{-2}) obtained from the National Geographical Institute (IGN—Spain) and used for the estimation of topographic features, including terrain height, slope, profile curvature, tangential curvature, aspect and wetness index. These parameters were calculated with QGIS 2.14.3 (QGIS Development Team 2016) and GRASS 7.0.3 (Grass Development Team 2015).

Soil sampling and analysis

Apparent electrical conductivity (EC_a) data were collected on 10 September 2011 using an EM-38DD (Geonics Limited). The EM-38 was seated on a non-metallic trolley that was pulled through the field with a van. Interline spacing was 10 m and the intensity of the collected data was 0.66 measurements m^{-2}. A global positioning system (GPS-RTK) was used to determine the geographical coordinates of the EC_a measurements. In this research, two dipoles (vertical, EC_a-V, and horizontal, EC_a-H) of Geonics EM-38 were used.

The ESAP 2.35 software (Lesch et al. 2000) assesses the spatial dependency of the EC_a data and calculates soil sampling locations which best encompass the variability present in the field, which is crucial information for determining the within-field variability of plant-available nutrients (Mallarino and Wittry 2004). Therefore, 80 sampling points (Fig. 1) were obtained by the ESAP 2.35 software for the studied prairie.

Soil samples were taken at two depths on the 80 spots selected from EC_a measurements: 0–20 and 20 cm to the A horizon lower boundary (highly variable in this field; ranging from 25 to 48 cm, depending on location). Samples were air-dried and sieved to 0.09 mm. Organic matter content was measured after wet digestion following the Walkley and Black (1934) method. SOM was determined for the two depths and then weighted and averaged to obtain the results for the entire A Horizon. Bulk density was determined on the 80 sampling points using the cylindrical core method.

Descriptive statistics

The description of the data set includes examination of the mean, median, mode, standard deviation, coefficient of variation (CV) and extreme minimum and maximum values. The relationship between soil properties (EC_a, bulk density) and topographic features (elevation, slope, profile curvature, tangential curvature and wetness index) and the SOM was assessed through Pearson's correlation coefficient in order to discern if EC_a could be a useful ancillary variable for estimating the spatial distribution of SOM through geostatistical interpolations.

Interpolations

Spatial distribution of SOM over the study plot was assessed through deterministic and geostatistical methods using GSTAT software for R, including IDW, OK and COK (Pebesma 2004) and RK was calculated using GSIF for R (Pebesma and Graeler 2016). The mathematical background of these techniques can be found elsewhere (e.g. Goovaerts 1997; James et al. 2013) and, therefore, they are briefly described.

Inverse distance weighting (IDW)

This technique does not account for the spatial autocorrelation of data and does not provide an estimation of the interpolation errors. However, it offers a rapid mapping of the studied variable, and it can be used as a reference or when no spatial autocorrelation is detected in the dataset. The method of IDW estimates SOM as a linear combination of several surrounding observations, with the weights being inversely proportional to the square distance between observations and the point to be estimated. Observations that are close to each other on the ground tend to be more alike than those further apart, hence observations closer to the record should receive a larger weight (Goovaerts 2000).

Estimation by geostatistical methods

Instead of the Euclidean distance, geostatistics uses the semivariogram as a measure of dissimilarity between observations (Goovaerts 2000). Hence, spatial continuity of SOM was investigated by calculating semivariograms, based on the assumptions of stationarity in accordance with the intrinsic hypothesis (Vieira 2000). Because of the limited number of measured data, only the omnidirectional semivariogram was computed, and hence the spatial variability is assumed to be identical in all directions.

Covariograms between SOM and EC_a were calculated, following the same methodology, in order to use collocated cokriging as interpolation method.

The cross-validation technique (Chilés and Delfiner 1999) was used to check the model performance. Two criteria were used to determine the goodness-of-fit of the model and to adjust its parameters (Karnieli 1990): coefficient of correlation (r) and mean squared prediction error (MSPE). For an unbiased prediction, centred on the true values, the MSPE should be close to zero. In this paper, a five-fold cross-validation approach was used; data were randomly split in five parts and, at each time, one of these parts is held out to compare it with the predictions obtained from the remaining four parts.

For spatial interpolation of SOM using geostatistical methods, we used OK, ordinary cokriging and RK.

Ordinary kriging (OK) considers two sources of information regarding the attribute, the variation and the distance between points (Webster and Oliver 2001). It provides each cell with a local, optimal prediction and an estimation of the error that depends on the semivariogram and the spatial configuration of the data (Goovaerts 1997). The OK weights minimize the estimation variance, while ensuring the unbiasedness of the estimator.

When the secondary information is not exhaustively sampled, the estimation can be done using a multivariate extension of the kriging estimator which is referred to as cokriging (COK), when the secondary variable is correlated with the variable of interest (Goovaerts 1997). In our case, EC_a was used as secondary information for estimating the spatial distribution of SOM using ordinary COK.

Regression kriging (RK) is based on the idea that the deterministic component of the main variable is explained by a regression model obtained from secondary information in all locations where we want to estimate the primary variable. Then, the location target residual values are estimated using OK from the observed residuals at all sampled locations (Zhang et al. 2012; Rossiter 2016). In this paper we consider two kinds of regression models: linear (LM-RK) and random forest (RF-RK). For LM-RK, an interactive mode was preferred to an additive one because it allows cross-terms between the different secondary variables (Rossiter 2016). The RF-RK method consists of building a large number of regression trees and average their predictions (Rossiter 2016; James et al. 2013). Random forest classification and regression models were built using the "randomForest" package (Liaw et al. 2016) in the R free statistical software (R Core Team 2016). Random forest regression models were built using 500 trees derived from 500 bootstrapped data sets. The random forest algorithm can rank the relative importance of each predictor variable based on the regression prediction error of the out-of-bag (OOB) portion of data (Breiman 2001; Liaw et al. 2016; Rossiter 2016; Everingham et al. 2016). In the randomForest package, predictor variable importance is reported as mean percent decrease in classification rate for the classification model or mean increase in mean square error for the regression model if that variable was removed from the analysis (Everingham et al. 2016). In addition, the method averages the OOB cross-validations calculated during the construction of the forest (Rossiter 2016).

The secondary topographic features (elevation, slope, profile curvature, tangential curvature, aspect and wetness index) used to make the regression were derived from the DTM, the other secondary parameters used were EC_a-V and EC_a-H estimated in each gridcell using OK.

Results

The SOM content of the topsoil (0 cm to A horizon lower boundary) ranged from 46 to 78 g kg^{-1}, with an average of 64.4 g kg^{-1} and a coefficient of variation (CV) of 10.2%. In addition, EC_a-V varied between 6.7 and 14.7 mS m^{-1}, with an average of 10.7 mS m^{-1} and a CV of 16% (Table 1). These two variables were significantly correlated (r = −0.6;

Table 1 Summary statistics of the soil properties studied

Variable	SOM (g kg^{-1})	SOM	EC_a-V	EC_a-V (mS m^{-1})	EC_a-H (mS m^{-1})	BD (g cm^{-3})
Depth	0–20 cm	20–A horizon	0–A horizon			0–20 cm
Minimum	51.0	40.0	46.3	6.7	5.4	0.92
Maximum	79.0	77.0	78.0	14.7	13.2	1.54
Mean	67.6	60.1	64.4	10.7	7.9	1.24
Median	67.0	60.0	64.7	10.7	7.9	1.24
Variance	38.5	65.3	43.2	3.0	0.4	0.02
Standard deviation	6.2	8.1	6.6	1.7	0.7	0.13
Coefficient of variation	9.2	13.4	10.2	16.0	8.2	0.10
Skewness	−0.40	−0.02	−0.31	0.18	1.72	−0.29
Kurtosis	−0.02	−0.34	0.29	−0.49	6.83	0.13

SOM soil organic matter, *EC_a* apparent electrical conductivity, *BD* bulk density

$p < 0.05$), especially when EC_a was measured using the vertical dipole mode; the correlation coefficient between SOM and elevation was high ($r = 0.57$, $p < 0.05$) (Table 2; Fig. 2). The average EC_a-H value was 7.9 mS m^{-1}, ranging from 5.4 to 13.2 mS m^{-1} and the correlation between SOM with EC_a-H was also significant but lower than with EC_a-V ($r = -0.45$, $p < 0.05$). Bulk density varied between 0.94 and 1.54 g cm^{-3}, with an average of 1.24 g cm^{-3} and a CV of 10% (Table 1), the coefficient of correlation between SOM and bulk density was low ($r = -0.17$). Means and medians were very similar and the kurtosis and skewness coefficients were close to 0 (Table 1); therefore, no transformation of the variables was performed in order to meet the requirements for geostatistical analyses. In addition, the CV values indicate that these soil properties showed spatial variability and suggest the convenience of site-specific management (Moral et al. 2010).

All the variables considered in this study (SOM, EC_a-V and EC_a-H) showed strong spatial dependence (Table 2). Fitted models were isotropic, meaning that the value of the variable varies similarly in all directions and that the semi-variance depends only on the distance between sample points. Semivariogram model parameters differed depending on the soil property considered (Table 2). In the case of the SOM content of the topsoil, nugget was null, sill was 38 and the relation nugget/sill zero. The range was 39 m, indicating no spatial dependence for SOM after 39 m. The semivariograms for EC_a-V and EC_a-H, also presented strong dependence, sill was 3 and 0.4 and range was 97 and 68 m, respectively for EC_a-V and EC_a-H, then OK provided reliable estimations of these variables (Fig. 3).

The variables used for RK were elevation, slope, profile curvature and EC_a-V because they were the most relevant after the OOB analysis and explained 42 % of the variance of the data when used for RF-RK.

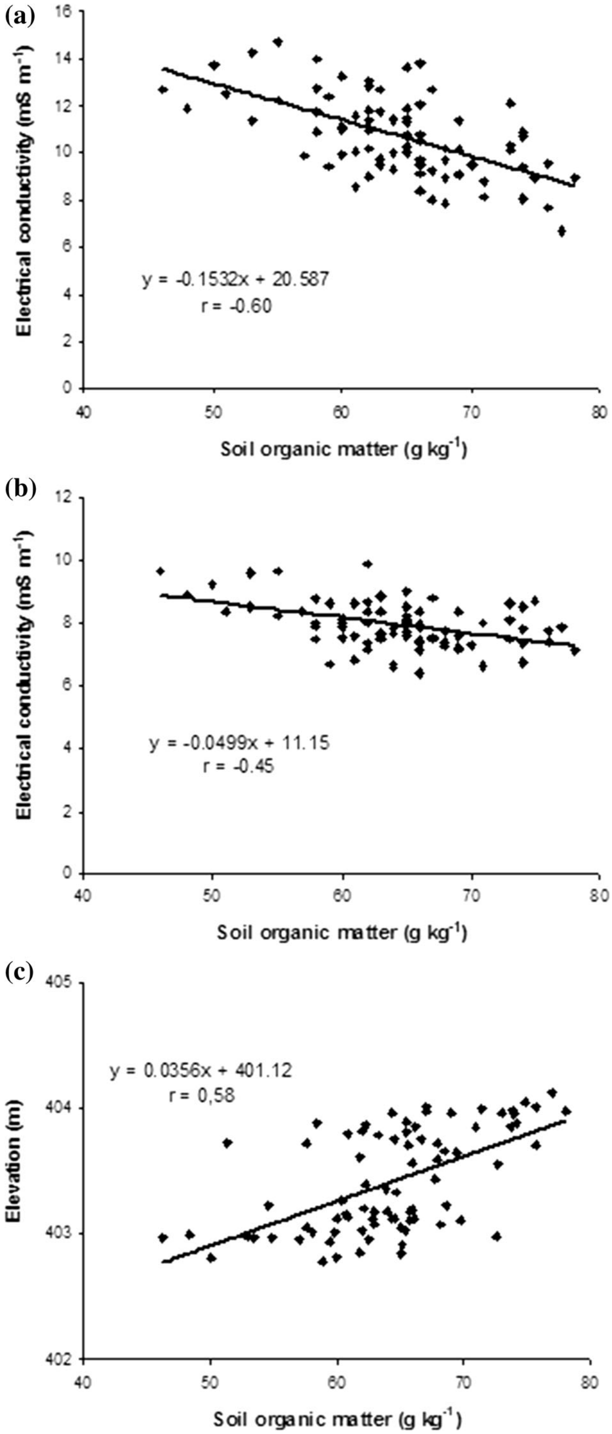
Comparing the SOM maps from the five different estimation methods, similar spatial patterns can be observed (Fig. 4). However, OK maps tended to smooth details and, then, to underestimate the short-distance variability. COK maps did not present a significant improvement from those obtained by OK (Fig. 4), but the SOM map obtained by RF-RK showed a significant improvement, as shown by the cross-validation parameters that are better for RF-RK (Table 3). In fact, the lowest error of prediction (MSPE) was observed for RF-RK, whereas the highest MSPE was detected for IDW, which was five times greater than that observed for RF-RK (Table 3). Moreover, error maps from RF-RK showed lower values than those generated by OK (Fig. 5). It is interesting to note that all methods used predicted the highest SOM concentrations in areas with the lowest EC_a values.

The five interpolation techniques produced rather similar estimated SOM topsoil (0 cm to A horizon lower boundary) contents for the whole field, ranging between 319 and 342 t; being the lowest value predicted by RF-RK and the highest one by COK (Table 4).

Table 2 Theoretical model parameters fitted to experimental semivariograms from the studied datasets

Variable	Model	Nugget	Sill	Range (m)
SOM (A horizon)	Spherical	0.00	38.00	59.0
EC_a-V	Spherical	0.01	3.01	97.0
EC_a-H	Spherical	0.09	0.40	68.0
Cross Variogram SOM to EC_a-V	Spherical	−0.01	−6.20	74.4
Residual SOM LM-RK	Exponential	0	18.47	22.76
Residual SOM RF-RK	Exponential	3.24	4.05	18.60

SOM soil organic matter, *EC_a* apparent electrical conductivity (either vertical, EC_a-V, or horizontal, EC_a-H)


(a)
Electrical conductivity (mS m-1)
y = -0.1532x + 20.587
r = -0.60
Soil organic matter (g kg-1)
(b)
Electrical conductivity (mS m-1)
y = -0.0499x + 11.15
r = -0.45
Soil organic matter (g kg-1)
(c)
Elevation (m)
y = 0.0356x + 401.12
r = 0,58
Soil organic matter (g kg-1)

◀**Fig. 2** Relationships between soil organic matter (SOM) and apparent electrical conductivity, measured either using **a** vertical (EC_a-V) or **b** horizontal (EC_a-H) sensors. The relationship between SOM and elevation is also shown (**c**)

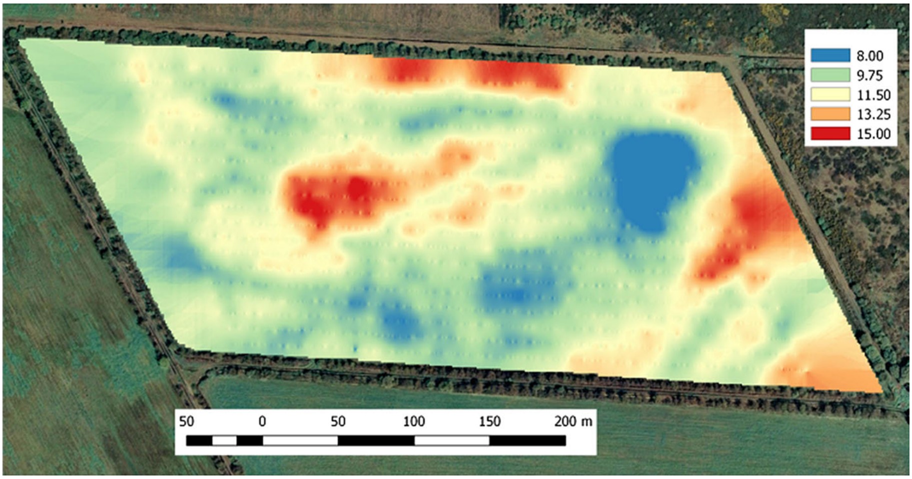


Fig. 3 Estimations map of EC_a-V generated by ordinary kriging (units are mS m^{-1})

However, the relative importance of the different SOM content classes was considerably different from one interpolation technique to the other (Table 4). IDW tended to assign larger areas to the 60–65 and 65–70 g kg^{-1} classes. With OK, the extremes were smoothed and the middle classes of SOM (55–60, 60–65 and 65–70 g kg^{-1}) covered the largest area of the estimated map. Similarly, COK provided larger areas to the 65–70 and 70–75 g kg^{-1} classes (Table 4), LM-RK showed higher values for the extreme classes than the other methods, thus presenting a more heterogeneous picture of the SOM distribution within the studied field than the other four interpolation techniques; whereas RF-RK presented a SOM values distribution similar to OK (Table 4).

Discussion

Soil organic matter is an important indicator of soil quality, and has numerous direct and indirect impacts on it (Lozano-García et al. 2016). The spatial distribution information of SOM in a given region is of paramount importance because it regulates local ecosystems functioning and soil health, hence strongly affecting agricultural productivity and climate change (Wang et al. 2014). Therefore, SOM spatial variation leads to differences in concentration, fertilizer needs, activity of herbicides and crop yield within a field (Mallarino and Wittry 2004). Thus, uniform treatment of the soil will result in zones within a field that are either over- or under-treated (Roy et al. 2006). Lal (2007) indicated that one of the principal challenges for soil scientists regarding SOM was to upscale the sampled point data to landscape, farm, watershed or region. In this context, several studies dealt with finding out the best method for interpolating SOM point data to a farm or region scale (Wu et al. 2009; Mabit and Bernard 2010; Marchetti et al. 2012; Piccini et al. 2014). However, the incorporation of a secondary variable in order to improve estimations is of

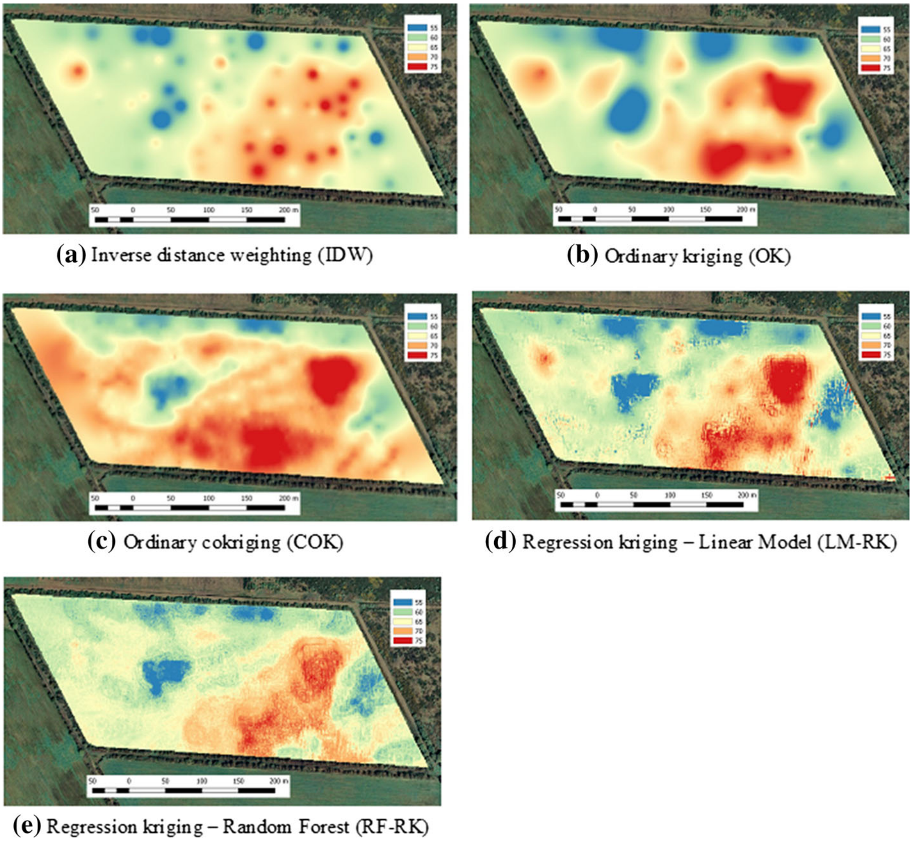


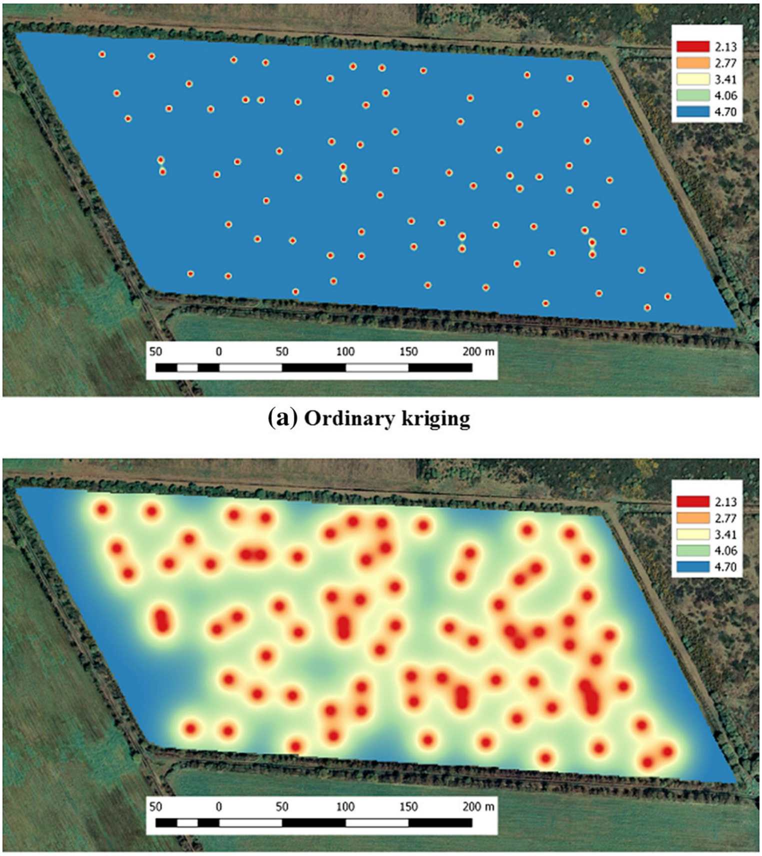
Fig. 4 Estimation maps of SOM generated by inverse distance weighting (IDW), ordinary kriging (OK), ordinary cokriging (COK), regression kriging (LM-RF) and regression kriging random forest (RF-RK) (units are g kg^{-1})

Table 3 Cross-validation parameters for the estimation of SOM using five different interpolation methods

Method	MSPE (g $kg^{-1})^2$	r
IDW	27.17	0.61
OK	25.55	0.64
COK	19.48	0.70
LM-RK	21.26	0.71
RF-RK	4.95	0.94

MSEP mean squared prediction estimation, *r* correlation coefficient between observed and predicted values

particular concern since no clear conclusions have been reached in this regard (Tarr et al. 2005; Wu et al. 2009). In the current study, we used EC_a as a covariate for improving spatial interpolations of SOM over a grassland farm in NW Spain, checking five interpolation methods. Although similar spatial patterns were obtained, a more detailed



(b) Regression kriging Random Forest

Fig. 5 Error maps of SOM generated by ordinary kriging and regression kriging random forest (units are $(\mathrm{g\ kg}^{-1})^2$)

assessment of the SOM variability within the studied field was obtained through RF-RK. In practical applications, the information obtained from these methods can be employed for site-specific management purposes.

In this context, a number of proximal and remote sensing tools are available nowadays for mapping soil properties (Brevik et al. 2016); however, not all of them are equally useful at all scales. Near infrared spectra (NIS) tools have been proven useful at large scales, such as basins (Viscarra Rossel and Chen 2011). Electromagnetic induction has been increasingly used to support soil surveys and site-specific management (Brevik et al. 2012). At field and landscape scales, EC_a maps have the potential to provide higher levels of resolution and greater distinction of soil types than those maps prepared with traditional tools and survey methods provided that there is significant variation in at least one of the factors that affects soil EC_a. For instance, EM-38 has been successfully used for characterising SOM in sandy and non-saline fields of 50–70 ha (Farahani et al. 2005). Other authors (Martínez et al. 2009; Sun et al. 2013; Gozdowski et al. 2015; Stadler et al. 2015) used EM-38 for evaluating soil properties and crop yields in fields ranging from 0.75 to 10 ha. These studies suggest that the EM-38 technique can be useful at field scales (1–35 ha).

Table 4 Areas of soil organic matter (SOM) classes obtained using inverse distance weighting (IDW), ordinary kriging (OK), cokriging (COK), regression kriging with linear model (LM-RK) and regression kriging with random forest (RF-RK); and SOM budgets associated

SOM classes (g kg^{-1})	Area (m^2) in SOM class as evaluated by				
	IDW	OK	COK	LM-RK	RF-RK
40–45	0	0	0	128	0
45–50	142	454	0	477	20
50–55	1 152	4 163	182	3 289	1 150
55–60	7 067	13 215	6 629	16 152	14 156
60–65	52 348	38 258	17 387	39 272	44 414
65–70	26 138	26 176	44 532	22 204	24 566
70–75	5 283	8 780	20 535	8 680	6 009
75–80	137	1 221	3 002	1 341	409
80–85	0	0	0	343	0
SOM average (g kg^{-1})	65.59	65.44	68.94	64.94	64.19
Total SOM content (t)	325.73	324.95	342.37	322.48	318.75

The SOM budget was established for the whole A horizon using an average bulk density of 1.2415 t m^{-3} (n = 80) on the 10-ha field

The number of points needed for reliably estimating semivariograms and the spatial variability of a given soil property depends on the accuracy required and the resources available for the survey (Goovaerts 1999); sampling techniques have evolved to deal with this issue (Ladoni et al. 2010; Roberts et al. 2011). In the current study we took into account the information provided by the EC_a measurements for determining 80 sampling points within the field, thus optimizing both accuracy and available resources. Results indicated that our sampling strategy was appropriate to capture the spatial variability of SOM distribution in the studied field.

As expected, both SOM and EC_a distributions in the A horizon were not homogeneous in the studied field. Greater SOM concentrations coincided with the higher areas within the field, whereas EC_a behaved in the opposite way, being higher in the lower areas of the studied field. This result would help farmers to adjust fertilizer inputs or herbicide applications for site-specific management within this field, thus reducing costs.

The relationship between SOM and EC_a is highly variable and will depend on the characteristics of the studied soil, since this relationship seems to be non-causal. Tarr et al. (2005) found a positive moderate correlation between SOM and soil EC_a (r = 0.24), whereas Kitchen et al. (2003) detected a high correlation (r = 0.81). Similarly, Peralta et al. (2015) found a high correlation between SOM and EC_a in a pasture land (r = 0.84). These findings indicated that correlations between EC_a and soil properties must be established on a field-by-field basis. In the current study, a significant negative relationship between both variables was detected, with a rather strong correlation coefficient (r = −0.6). Previous studies (Siqueira et al. 2014) carried out on a plot 500 m away from the one used in the current study yielded very similar relationships between SOM and EC_a-V and EC_a-H (r = −0.6 and r = −0.45, respectively) to those found in the current research. The soil of that plot, although of different classification, showed similar texture and a slightly greater average SOM (Siqueira 2009). The agreement between our results and those found by Siqueira et al. (2014) encourage the use of EMI techniques for exploring soil properties in similar areas.

High EC_a values are associated with soils with finer texture and high SOM; however, in the acidic soil studied here, SOM is not mineralized and does not contribute to soil structure in order to increase electrical conductivity. This support the idea that the relationship between SOM and EC_a is non-causal. Due to the small proportions of SOM in a field soil, it is difficult that a modification in them would alter the measured EC_a of the soil profile. In fact, there is no physical basis to expect neither a direct positive nor a negative relationship between SOM and EC_a, as reflected by the different relationships determined by other researchers (Tarr et al. 2005; Siqueira et al. 2014; Peralta et al. 2015). Topography and general local soil conditions might modify SOM and EC_a spatial distribution over a field but EC_a would probably depend on other soil properties.

Tarr et al. (2005) used soil EC_a as an ancillary variable for mapping the spatial distribution of several soil properties, including SOM, in Iowa soils. When they used COK, they obtained slight improvements since the correlation between SOM and EC_a was moderate. In our study, local detail improvements of the COK maps were due to the finer sampling grid of the covariate, soil EC_a (McBratney and Webster 1983), hence, COK maps were slightly more detailed than those obtained by OK. However, the improvement in estimation errors was not significant and the effort to model the cross semivariogram for EC_a, SOM did not compensate for the results of cross-validation, which were worse than those obtained by the RF-RK method. Thus, in the studied field and with the ancillary information used, RF-RK outperformed the other methods, this assertion coincides with Rudiyanto et al. (2016) and Guo et al. (2015), who consider the random forest method a good technique to estimate SOM associated to RK.

Both IDW and geostatistical (OK, COK, LM-RK and RF-RK) methods used provided similar total SOM contents for the entire field. However, the relative importance of the different SOM content classes differed from one technique to the other. Furthermore, the spatial distribution patterns obtained by the different approaches used in this study were rather similar. The IDW method tended to under-estimate the extreme values of the distribution and, hence, the accuracy of the map might be affected. In contrast, geostatistical approaches gave softer maps and provided estimations of the interpolation errors.

Conclusions

The SOM content in the study area had a strong spatial dependence and showed a moderate negative correlation with the EC_a ($r = -0.6$) and a moderate positive correlation with the elevation ($r = 0.58$); thus, the predicted SOM map by COK with EC_a as a covariate represented a slight improvement over that by OK. However, other authors have reported very different relationships between these two variables, suggesting that this relationship is non-causal. Nevertheless, electromagnetic data such as EC_a might act as a useful auxiliary variable for improving the accuracy and reliability of SOM spatial predictions in the field studied here. In fact, the results for LM-RK using EC_a, elevation and profile curvature were similar to those obtained by COK. However, RF-RK improved significantly the results for SOM estimation using topographic parameters that are readily available as well as EC_a, providing lower estimation errors than the other interpolation methods.

Therefore, the method used here for determining the best soil sampling scheme was successful in capturing the spatial variability of soil properties of interest. The distribution maps obtained in this study allowed us for identifying areas with homogeneous SOM contents, which may help for implementing precision agriculture practices through site-specific fertilization or soil management.

Acknowledgments This work was supported by Spanish Ministry of Economy and Competitiveness (Project CGL2013-47814-C2). The helpful comments from two anonymous reviewers are deeply acknowledged. The authors thank two anonymous reviewers for their helpful insights on previous versions of this manuscript.

References

Baxter, S. J., & Oliver, M. A. (2005). The spatial prediction of soil mineral N and potentially available N using elevation. *Geoderma, 128*, 325–339.

Bishop, T. F. A., & Lark, R. M. (2006). The geostatistical analysis of experiments at the landscape-scale. *Geoderma, 133*, 87–106.

Bregt, A. K., Gesing, H. J., & Alkasuma, M. (1992). Mapping the conditional probability of soil variables. *Geoderma, 53*, 15–29.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Brevik, E. C., Calzolari, C., Miller, B. A., Pereira, P., Kabala, C., Baumgarten, A., et al. (2016). Soil mapping, classification, and pedologic modeling: History and future directions. *Geoderma, 264*, 256–274.

Brevik, E. C., Fenton, T. E., & Jaynes, D. B. (2012). Use of electrical conductivity to investigate soil homogeneity in Story County, Iowa, USA. *Soil Survey Horizon, 53*(5), 50–54.

Chen, C., Hu, K., Li, H., Yun, A., & Li, B. (2015). Three-dimensional mapping of soil organic carbon by combining kriging method with profile depth function. *PLoS ONE, 10*, e012903.

Chilés, J. P., & Delfiner, P. (1999). *Geostatistics. Modeling spatial uncertainty*. New York: Wiley.

Dobson, A. J., & Barnett, A. G. (2008). *An introduction to generalized linear models*. London: Chapman and Hall.

Doolittle, J. A., & Brevik, E. C. (2014). The use of electromagnetic induction techniques in soils studies. *Geoderma, 223–225*, 33–45.

Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development, 36*, 27.

Farahani, H. J., Buchleiter, G. W., & Brodahl, M. K. (2005). Characterization of apparent soil electrical conductivity variability in irrigated sandy and non-saline fields in Colorado. *Transactions of the ASAE, 48*, 155–168.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Applied Geostatistics Series: Oxford University Press.

Goovaerts, P. (1999). Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma, 89*, 1–45.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology, 228*, 113–129.

Gozdowski, D., Stępień, M., Samborski, S., Dobers, E. S., Szatyłowicz, J., & Chormański, J. (2015). Prediction accuracy of selected spatial interpolation methods for soil texture at farm field scale. *Journal of Soil Science and Plant Nutrition, 15*, 639–650.

GRASS Development Team. (2015). Geographic Resources Analysis Support System (GRASS) Software, Version 7.0.3 Open Source Geospatial Foundation. Retrieved June 9, 2016 from http://grass.osgeo.org.

Gray, L. C., & Morant, P. (2003). Reconciling indigenous knowledge with scientific assessment of soil fertility changes in southwestern Burkina Faso. *Geoderma, 111*, 425–437.

Guo, P. T., Li, M. F., Luo, W., Tang, Q. F., Liu, Z. W., & Lin, Z. M. (2015). Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. *Geoderma, 237*, 49–59.

Hoffmann, U., Hoffmann, T., Jurasinski, G., Glatzel, S., & Kuhn, N. J. (2014). Assessing the spatial variability of soil organic carbon stocks in an alpine setting (Grindelwald, Swiss Alps). *Geoderma, 232–234*, 270–283.

IUSS Working Group WRB. (2014). *World reference base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps.* World Soil Resources Reports No. 106. Rome: FAO.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Number 103 in Springer texts in statistics. New York: Springer.

Karnieli, A. (1990). Application of kriging technique to areal precipitation mapping in Arizona. *GeoJournal, 22*, 391–398.

King, J. A., Dampney, P. M. R., Lark, R. M., Wheeler, H. C., Bradley, R. I., & Mayr, T. R. (2005). Mapping potential crop management zones within fields: Use of yield-map series and patterns of soil physical properties identified by electromagnetic induction sensing. *Precision Agriculture, 6*, 167–181.

Kitchen, N. R., Drummond, S. T., Lund, E. D., Sudduth, K. A., & Buchleiter, G. W. (2003). Soil electrical conductivity and topography related to yield for three contrasting soil–crop systems. *Agronomy Journal, 95*, 483–495.

Köppen, W. (1936) Das geograsphica system der Klimate [On a geographic system of climate]. In W. Köppen & G. Geiger (Eds.), *Handbuch der Klimatologie* [*Handbook of Climatology*], 1.C. (pp. 1–44). Gebr, Bontraerger.

Ladoni, M., Bahrami, H. A., Alavipanah, S. K., & Norouzi, A. A. (2010). Estimating soil organic carbon from soil reflectance: A review. *Precision Agriculture, 11*, 82–99.

Lal, R. (2007). Farming carbon. *Soil and Tillage Research, 96*, 1–5.

Lesch, S. M., Rhoades, J. D., & Corwin, D. L. (2000). *ESAP-95 version 2.01R. User manual and tutorial guide*. Research Report N° 146, June 2000. USDA-ARS. George E. Brown, Jr., Salinity Laboratory, Riverside, CA.

Liaw, A., Wiener, M, Breiman, L., & Cutler, A. (2016). Package 'random forest'. Retrieved May 18, 2016 from https://www.stat.berkeley.edu/~breiman/RandomForests/.

Lozano-García, B., Parras-Alcántara, L., & Brevik, E. C. (2016). Impact of topographic aspect and vegetation (native and reforested areas) on soil organic carbon and nitrogen budgets in Mediterranean natural areas. *Science of the Total Environment, 544*, 963–970.

Lozano-García, B., Parras-Alcántara, L., & Del Toro, M. (2011). The effects of agricultural management with oil mill by-products on surface soil properties, runoff and soil losses in southern Spain. *Catena, 85*, 187–193.

Mabit, L., & Bernard, C. (2010). Spatial distribution and content of soil organic matter in an agricultural field in eastern Canada, as estimated from geostatistical tools. *Earth Surface Processes and Landforms, 35*, 278–283.

Mallarino, A. P., & Wittry, D. J. (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.

Marchetti, A., Piccini, C., Francaviglia, R., & Mabit, L. (2012). Spatial distribution of soil organic matter using geostatistics: A key indicator to assess soil degradation status in central Italy. *Pedosphere, 22*(2), 230–242.

Martínez, G., Vanderlinden, K., Ordóñez, R., & Muriel, J. L. (2009). Can apparent electrical conductivity improve the spatial characterization of soil organic carbon? *Vadose Zone Journal, 8*, 586–593.

McBratney, A. B., & Webster, R. (1983). Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. *Journal of Soil Science, 34*, 137–162.

Miller, B. A., Koszinski, S., Wehrhan, M., & Sommer, M. (2015). Comparison of spatial association approaches for landscape mapping of soil organic carbon stocks. *Soil, 1*, 217–233. doi:10.5194/soil-1-217-2015.

Moral, F. J., Terrrón, J. M., & Marques da Silva, J. R. (2010). Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil and Tillage Research, 106*, 335–343.

Nerini, D., Momnestiez, P., & Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis, 101*, 409–418.

Nussbaum, M., Papritz, A., Baltensweiler, A., & Walthert, L. (2014). Estimating soil organic carbon stocks of Swiss forest soils by robust external-drift kriging. *Geosciences Model Development, 7*, 1197–1210.

Pachepsky, Y. A., Timlin, D. J., & Rawls, W. J. (2001). Soil water retention as related to topographic variables. *Soil Science Society of America Journal, 65*, 1787–1795.

Paz-Gonzalez, A., Vieira, S. R., & Taboada Castro, M. T. (2000). The effect of cultivation on the spatial variability of selected properties on an umbric horizon. *Geoderma, 97*, 273–292.

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences, 30*, 683–691.

Pebesma, E., & Graeler, B. (2016). Package 'GSIF'gstat'. Retrieved May 18, 2016 from https://cran.r-project.org/web/packages/gstat/gstat.pdf.

Peralta, N. R., Cicore, P. L., Marino, M. A., Marques da Silva, J. R., & Costa, J. L. (2015). Use of geophysical survey as a predictor of the edaphic properties variability in soils used for livestock production. *Spanish Journal of Agricultural Research, 13*(4), e1103. doi:10.5424/sjar/2015134-8032.

Piccini, C., Marchetti, A., & Francaviglia, R. (2014). Estimation of soil organic matter by geostatistical methods: Use of auxiliary information in agricultural and environmental assessment. *Ecological Indicators, 36*, 301–314.
QGIS Development Team. (2016). QGIS geographic information system 2.14.3. Open Source Geospatial Foundation Project. Retrieved May 18, 2016 from http://www.qgis.org/.
R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved February 22, 2016 from http://www.R-project.org/.
Roberts, D. F., Adamchuck, V. I., Shanahan, J. F., Ferguson, R. B., & Schepers, J. S. (2011). Estimation of surface soil organic matter using a ground-based active sensor and aerial imagery. *Precision Agriculture, 12*, 82–102.
Robinson, T. P., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculure, 50*, 97–108.
Rossiter, D. G. (2016). *An introduction to geostatistics with R/gstat*. Cornell University. Retrieved May 18, 2016 from http://www.css.cornell.edu/faculty/dgr2/teach/R/gs_short_ex.pdf.
Roy, S. K., Shibusawa, S., & Okayama, T. (2006). Textural analysis of soil images to quantify and characterize the spatial variation of soil properties using a real-time soil sensor. *Precision Agriculture, 7*, 419–436.
Rudiyanto, Minasny, B., Setiawan, B. I., Arif, C., Saptomo, S. K., & Chadirin, Y. (2016). Digital mapping for cost-effective and accurate prediction of the depth and carbon stocks in Indonesian peatlands. *Geoderma, 272*, 20–31.
Siqueira, G. M. (2009). Medida de la conductividad eléctrica aparente del suelo por inducción electromagnética y variabilidad espacial de propiedades físicas y químicas del suelo [Measurement of soil apparent electrical conductivity by electromagnetic induction and spatial variability of physical and chemical soil properties]. Ph.D. Thesis. Universidade de Santiago de Compostela, Spain.
Siqueira, G. M., Dafonte Dafonte, J., Valcárcel Armesto, M., & França e Silva, E. F. (2014). Using multivariate geostatistics to assess patterns of spatial dependence of apparent soil electrical conductivity and selected soil properties. *The Scientific World Journal,*. doi:10.1155/2014/712403.
Stadler, A., Rudolph, S., Kupischa, M., Langensiepen, M., van der Kruk, J., & Ewert, F. (2015). Quantifying the effects of soil variability on crop growth using apparent soil electrical conductivity measurements. *European Journal of Agronomy, 64*, 8–20.
Sudduth, K. A., Kitchen, N. R., Wiebold, W. J., Batchelor, W. D., Bollero, G. A., Bullock, D. G., et al. (2005). Relating apparent electrical conductivity to soil properties across the north-central USA. *Computers and Electronics in Agriculture, 46*, 263–283.
Sun, Y., Cheng, Q., Lin, J., Schellberg, J., & Lammers, P. S. (2013). Investigating soil physical properties and yield response in a grassland field using a dual-sensor penetrometer and EM38. *Journal of Plant Nutrition and Soil Science, 176*, 209–216.
Tarr, A., Moore, K. J., Burras, C. L., Bullock, D. G., & Dixon, P. M. (2005). Improving map accuracy of soil variables using soil electrical conductivity as a covariate. *Precision Agriculture, 6*, 255–270.
Vieira, S. R. (2000). Uso de geoestatística em estudos de variabilidade espacial de propriedades do solo [Use of geostatistics in studies of spatial variability of soil properties]. In R. F. Novais (Ed.), *Tópicos em Ciência do Solo* [*Topics on soil science*] (pp. 3–87). Viçosa: Sociedade Brasileira de Ciência do Solo.
Viscarra Rossel, R. A., & Chen, C. (2011). Digitally mapping the information content of visible near infrared spectra of superficial Australian soils. *Remote Sensing of Environment, 15*, 1443–1455.
Vitharana, U. W. A., Van Meirvenne, M., Cockx, L., & Bourgeois, J. (2006). Identifying potential management zones in a layered soil using several sources of ancillary information. *Soil Use and Management, 22*, 405–413.
Walkley, A., & Black, I. A. (1934). An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science, 37*(1), 29–38.
Wang, K., Zhang, C. R., Li, W. D., Lin, J., & Zhang, D. X. (2014). Mapping soil organic matter with limited sample data using geographically weighted regression. *Journal of Spatial Science, 59*, 91–106.
Webster, R., & Oliver, M. A. (2001). *Geostatistics for environmental scientists* (p. 149)., Statistics in practice series Chichester: Wiley.
Wu, C., Wu, J., Luo, Y., Zhang, L., & DeGloria, S. D. (2009). Spatial prediction of soil organic matter content using cokriging with remotely sensed data. *Soil Science Society of America Journal, 73*(4), 1202–1208.
Zhang, S. W., Huang, Y. F., Shen, C. Y., Ye, H. C., & Du, Y. C. (2012). Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. *Geoderma, 171*, 35–43.