Application GIS and remote sensing for soil organic carbon mapping in a farm-scale in the hilly area of central Vietnam

Air, Soil and Water Research
Volume 15: 1–11

DOI: 10.1177/11786221221114777

Chuong Van Huynh[1], Tung Gia Pham[2], Linh Hoang Khanh Nguyen[2], Hai Trung Nguyen[3], Phuong Thuy Nguyen[3], Quy Ngoc Phuong Le[3], Phuong Thị Tran[3], Mai Thi Hong Nguyen[3] and Tuyet Thi Anh Tran[3]

[1]President of Board of Trustees, Hue University, Vietnam. [2]International School – Hue University, Vietnam. [3]University of Agriculture and Forestry, Hue University, Vietnam.

ABSTRACT: Soil Organic Carbon (SOC) influences many soil properties including nutrient and water holding capacity, nutrient cycling and stability, improved water infiltration and aeration. It also is an essential parameter in the assessment of soil quality, especially for agricultural production. However, SOC mapping is a complicated process that is costly and time-consuming due to the physical challenges of the natural conditions that is being surveyed. The best model for SOC mapping is still in debate among many researchers. Recently, the development of machine learning and Geographical Information Systems (GIS) has provided the potential for more accurate spatial prediction of SOC content. This research was conducted in a relatively small-scale capacity in the Central Vietnam region. The aim of this study is to compare the accuracy of Inverse Distance Weighting (IDW), Ordinary Kriging (OK), and Random Forest (RF) methods for SOC interpolation, with a dataset of 47 soil samples for an area of 145 hectares. Three environmental variables including elevation, slope, and the Normalized Difference Vegetation Index (NDVI) were used for the RF model. In the RF model, the values of the number of variables randomly sampled as candidates at each split, (mtry), and the number of bootstrap replicates, (ntree), were determined in terms of 1 and 1,000 respectively The results at our research site showed that using IDW is the most accurate method for SOC mapping, followed by the methods of RF and OK respectively. Concerning SOC mapping based-on auxiliary variables, in areas where there is human activity, the selection of auxiliary variables should be carefully considered because the variation in the SOC may not only be due to environmental variables but also by farming technologies.

KEYWORDS: Digital soil mapping, spatial interpolation, soil organic carbon, Vietnam

RECEIVED: February 26, 2022 . **ACCEPTED:** June 17, 2022.

TYPE: Original Research

CORRESPONDING AUTHOR: Chuong Van Huynh, President, Board of Trustees, Hue University, 03 Le Loi, Hue City, 840234, Vietnam. Email: huynhvanchuong@hueuni.edu.vn

Introduction

Soil Organic Carbon (SOC) is a crucial element in evaluating soil quality, especially in areas of agricultural land use (Tajik et al., 2020). The SOC influences the physical, chemical, and biological aspects of soil (Milne et al., 2015; Zeraatpisheh et al., 2021, 2022). The capacity for water-retention and the infiltration of rainwater that is influenced by SOC creates an increased availability of moisture within the landscape (Hartemink & McSweeney, 2014). In general, SOC is an indispensable consideration in regard to sustainable agriculture (John et al., 2007). Understanding the fundamental aspects and functions of SOC, and its specific measurement, is an important observance in addressing soil degradation (Ramifehiarivo et al., 2017). This knowledge also helps to improve agricultural productivity (Jat et al., 2019). The detailed and accurate spatial distribution of SOC is vital for many processes that are relevant to land use.

The SOC map is a useful tool that has many applications. For instance, to define degraded land areas and thus improve land management (Meersmans et al., 2012), especially in regard to agricultural production (van Den et al., 2017). In 2017, a Global Soil Organic Carbon Map was established by the Food and Agriculture Organization of the United Nations (FAO, 2020), at the resolution of a 1 km grid, and reaching a depth of 0 to 30 cm of the top-soil. This map provides an improved ability to determine the current SOC content stored in a given area and also indicates the potential for carbon sequestration throughout the world (FAO, 2020). However, SOC mapping using the historically standardized method of field surveys conducted physically by researchers is a costly and time-consuming process due to the difficulties of accessing areas where the terrain is treacherous, especially in certain mountainous regions (Vågen & Winowiecki, 2013).

Digital Soil Mapping (DSM) has become popular in recent years. It is favored because of its accuracy and the detailed information that it provides. Also with DSM, the data can be updated easily (G. L. Zhang et al., 2017). DSM also records spatial material information in such a way that it can be integrated into various software programs and serve as a tool that can be applied to other geographical studies such as studies in evaluation and the planning of agricultural land use (Naresh, 2020). The goal of improving the accuracy of the soil mapping process has always been one of the foremost topics of interest among soil scientists. Several methods were applied to map the distribution of soil properties worldwide (Bostani et al., 2017; FAO, 2018; Göl et al., 2017; Pouladi et al., 2019). However, definitive conclusions have not been drawn as to the comparative accuracy of the various interpolated methods of soil

mapping in each research case. The methods have different theoretical bases and thus, in combination, it is possible to study a variety of different conditions (Lai et al., 2021). The different methods have their advantages and disadvantages, for instance, the multiple linear regression (MLR) approach uses auxiliary variables as predictive factors to simulate soil properties, whereas the Ordinary Kriging (OK) interpolation approach emphasizes the spatial heterogeneity of simulated variables (Piccini et al., 2020). Some researchers consider geostatistics to be an efficient method for reducing the variance of assessment errors and also the cost of application (Bhunia et al., 2018; Bostani et al., 2017). Even so, soil science researchers are still seeking a consensus on which one of the spatial interpolation methods is the most accurate. Some researchers found that the OK method seems more accurate than the IDW approach for SOC mapping (Göl et al., 2017; Pouladi et al., 2019). Other researchers (Calvo de Anta et al., 2020; Siyu, 2013) found that the Random Forest (RF) model outperformed the other models. Predictive models are commonly designed based on special hypotheses and may be characterized by some inherent shortcomings (Song et al., 2020). The application of machine learning methods is one of the current trends in the spatial interpolation and distribution of soil features. The vast majority of studies are conducted on a large scale, however on a smaller scale, for example, the area of land occupied by a given farm, have not as yet been addressed by detailed studies of this kind.

Much effort has been made to determine the influences of environmental variables on soil properties (G. L. Zhang et al., 2017). Previous research found that the rock outcrop ratio (ROR), bulk density (BD), altitude, soil depth, slope gradient, and pH level, all have a significant effect on SOC (Hu et al., 2018). Climatic factors also have substantial impact on SOC (Z. Liu et al., 2011). Terrain and environmental variables influence the spatial distribution of small-scale SOC (Piccini et al., 2020). Therefore, taking into account the influence of environmental factors on the distribution of SOC is an important consideration in the SOC mapping processes. The number of auxiliary variables used in the spatial interpolation models of SOC content are diverse, however, it is apparent that the number of regular variables ranges from 5 to 12 (T. Liu et al., 2020; Mishra et al., 2020). Calculating these auxiliary variables requires considerable time and software skills, which causes certain difficulties for less developed areas where human resources are minimal, for example in mountainous regions in developing countries. Conducting research on models that have a small number of auxiliary variables which are easier to calculate for these regions should be developed further.

In Vietnam, the investment into SOC mapping has not been given sufficient attention or become standard practice by the relevant agencies such as agricultural departments and natural resources departments. Most of the current SOC maps were inherited products from more basic research techniques on soil quality that were used many years ago. The main method applied is to take samples from the field for laboratory analysis in combination with observation of the soil color and vegetation cover to create SOC maps. As a result, most of the existing maps created from previous methods are general and outdated. More recently, selected researchers have begun to apply machine learning in soil properties mapping more regularly (Châu, 2020; Gia Pham et al., 2019). However, the influences of environmental variables on soil properties mapping in the mountainous of Central Vietnam are often still lacking in this approach. Therefore, this study aims to compare the popular interpolation methods to find the most suitable approach in SOC mapping in the specific conditions of Central Vietnam, especially considering areas with steep terrain for small-scale areas.

Materials and Methods

Research area

The research site encompasses 145 hectares (ha) and belongs to Nam Dong district, Thua Thien Hue Province, in Central Vietnam. The research area is shown in Figure 1. The climate in this area shows tropical monsoon characteristics with a rainy season from September to December. The average precipitation from 2005 to 2019 is 3,824 mm, and the average temperature is 23°C to 24°C with the highest temperatures reached in June and lowest temperatures in January, 30°C, and 15°C, respectively. Although our research area is not large, the terrain is quite complicated. The elevation fluctuates from 130 to 301 m m.a.s.l and decreases from the southern to the northern areas. The slopes range from more than 0° to 47°. However, 70% of the total area has an average slope of about 10°, so the terrain is generally not very steep (Nam Dong District People's Committee, 2020). This area belongs to Thuong Quang commune's agricultural land, Nam Dong district. Here there are three main types of land-use, including cultivation of acacia, cassava, and rubber trees with mass areas of 80, 11, and 48 ha respectively. The soil types in this area are Ferralic Acrisols, that has a general depth of more than 100 cm. In areas with acacia and rubber plantations, the soil surface is covered with a mat of decaying leaves, while in areas where cassava is grown, the ground is cleared of vegetative cover. Acacia is planted at the highest altitudes, followed by rubber and cassava. Rubber latex extraction is the main source of livelihood for the local people, the plantations of which have been established 15 or 20 years previously, while acacia plantations and logging provide the main immediate cash income due to its rotation cycle of 4 to 5 years.

Methods

Soil sampling and soil organic carbon analysis. There were 65 soil samples collected in March 2020 on a random basis, including 23 samples taken from rubber plantations, five samples from

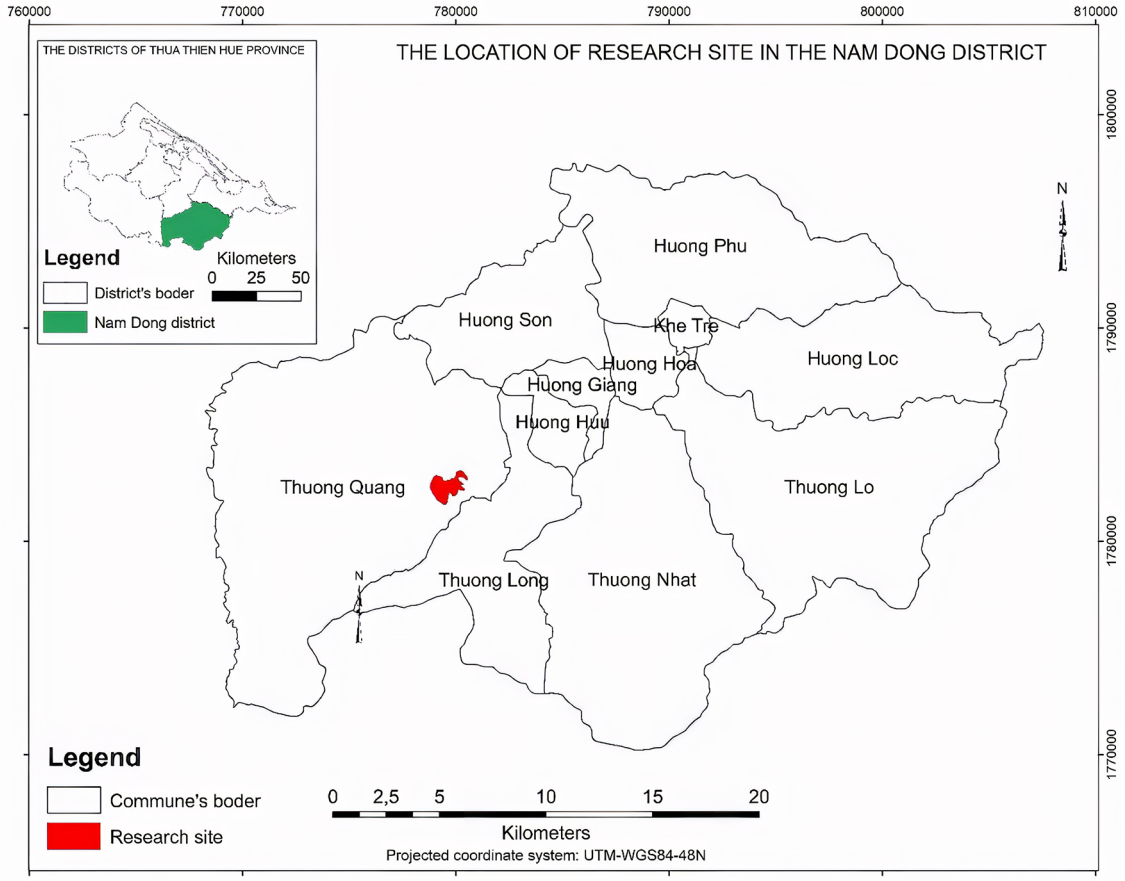

Figure 1. The location of the research site.

cassava plantations, and 37 samples from acacia plantations. The sample locations are presented in Figure 2. The distance between each sampling site ranges from 60 to 200 m. All the samples were selected in the same soil type to ensure that the soil samples were homogenous. The samples have been collected at a depth of 30 cm of topsoil, and from five points around a circle of land area with a radius of 1 m. These were then mixed together to make up one soil sample. All of the samples were dried at room temperature before further processing.

The SOC content was analyzed by the Walkley–Black method (Black, 1965) at the Soil Science and Fertilizer Department laboratory at Hue University of Agriculture and Forestry, Hue city, Vietnam.

Environmental variables. The environmental variables were extracted and calculated based on the remote sensing data. This data is the Landsat 8 image (acquired on March 10th 2020, path 125, row 49) with less than 10% of cloud cover and the Digital Elevation Model (n15, e108, 1 arc global, acquired in 2010). All remote sensing data was downloaded from the United States Geological Survey (USGS) website (https://earthexplorer.usgs.gov/, n.d.). The data of Landsat 8 was atmospherically corrected and converted from digital numbers to reflectance values based on the guidance of USGS (U.S. Geological Survey, 2019). The remote sensing data were then resampled to get a spatial resolution of 30 m. In this research, Band 4 and Band 5 of the Landsat 8 image were used to calculate NDVI as the following equation (U.S. Geological Survey, 2019):

$$NDVI = \frac{\left(Band\,5 - Band\,4\right)}{\left(Band\,5 + Band\,4\right)} \quad (1)$$

where NDVI is Normalized Difference Vegetation Index; Band 5 is the reflectance value of the Near-Infrared band and Band 4 is the reflectance value of the Red band.

The DEM data was used to calculate the slope by ArcGIS software which established the elevation and slope maps. The elevation, slope, and NDVI maps are presented in Figure 3.

Soil organic carbon interpolation techniques

Inverse distance weighting. The IDW technique has been developed based on the relative distance between the estimated points and the known points. The critical levels of the known points were created through the inverse of its distance from the interpolation point. This method has been written as the following equation (Maleika, 2020).

$$Z(x_0) = \sum_{i=1}^{n} w_i Z(x_i); \qquad w_i = \frac{d_i^{-p}}{\sum_{i=1}^{n} d_i^{-p}} \quad (2)$$

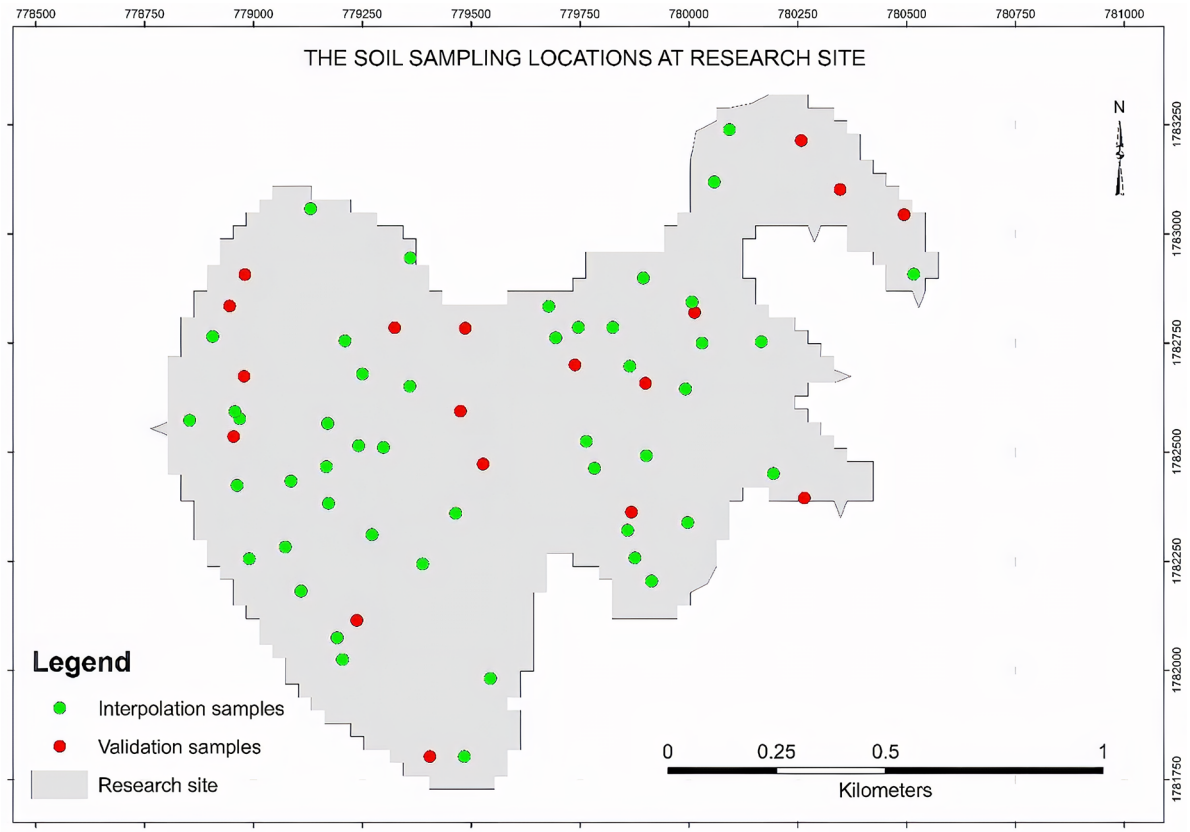

Figure 2. Soil sampling locations at the research site.

where $Z(x_0)$ represents the unknown value point at (x_0); $Z(x_i)$ represents the measured point at (x_i); n are the number of points (in the search radius area); w_i represents the weighting of each soil sample; d_i is the distance from (x_0) to (x_i); and p is the power.

IDW interpolation is a mathematical application which assumes that points that are closer together have a stronger relation than points that are further apart. The weighting of influence is proportional to the inverse of the distance raised to the power value p. A smaller p-value has less effect on the interpolated value. If the p-value is 0, the effectiveness of distance was eliminated. Previous researcher (Tung, 1983) noted that a p-value of 2 is the most suitable in the IDW interpolation. The IDW method is used for spatial interpolation in the environmental research of soil properties, terrain mapping, and air pollution (Salekin et al., 2018; Srivastava et al., 2019; Su et al., 2018). In this research, the IDW method has been implemented by ArcGIS with the value of p being 2, and n being 47.

Ordinary Kriging. Ordinary Kriging (OK) is used in spatial interpolation because the amount of input data that is required is relatively small (Mesić Kiš, 2016). This interpolation method uses only data that is easily accessible, rather than the entire exhaustive dataset, and also excludes the remote points. It calculates the value of unknown points by a linear sum of known points with a weighting coefficient between the known and unknown points. The standard condition when assessing the OK technique is that the sum of all weights is equal to 1. In order to find the significance in the OK method, the Lagrange multipliers is applied. The OK method has been written as the following equation (Cressie, 1993).

$$Z(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{3}$$

where $Z(x_0)$ is the value of the unknown point (x_0); $Z(x_i)$ is the value of unknown at i^{th} point (x_i); λ_i is the weighting coefficient value between (x_0) and (x_i); n is the total number of known points. In this research, the OK has been conducted in R with some previous research frameworks (Hengl, 2009; Omuto & Vargas, 2015).

The semi-variogram depicts the spatial autocorrelation of the measured sample points in the model of the OK method. This semi-variogram has a nugget of 0.02; a sill of 0.13 and a range of 85 m. This was found in the OK method with a spherical model (Figure 4). The nugget/sill ratio of SOC is 0.18, indicating that the sampled spatial dependence is quite important.

Random forest. RF is a machine learning technique which was developed based on the "decision tree" and "bagging" (Prasad et al., 2006). Bagging creates an ensemble algorithm that fits multiple models on different subsets of training datasets, then combines the predictions from all models. Random

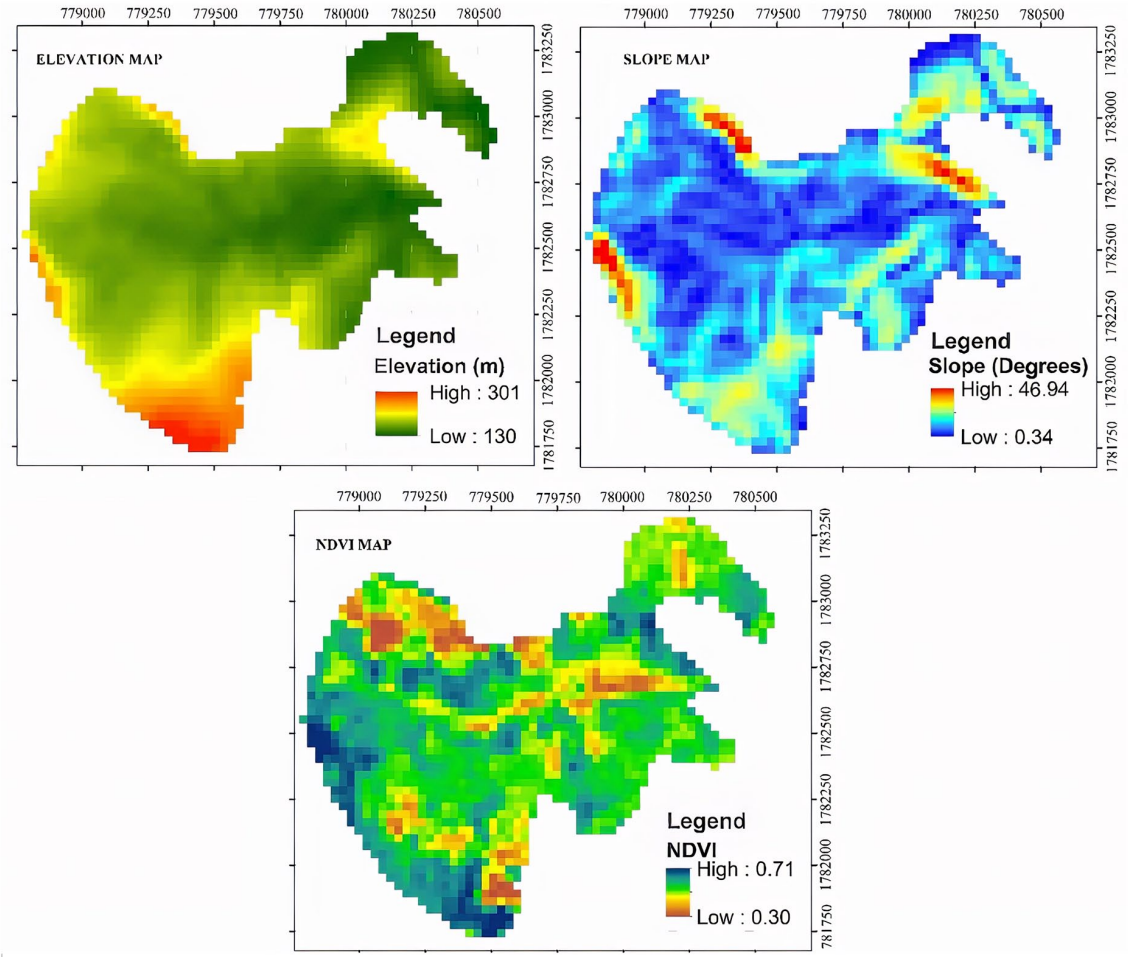

Figure 3. The elevation, slope, and NDVI maps of the entire research area.

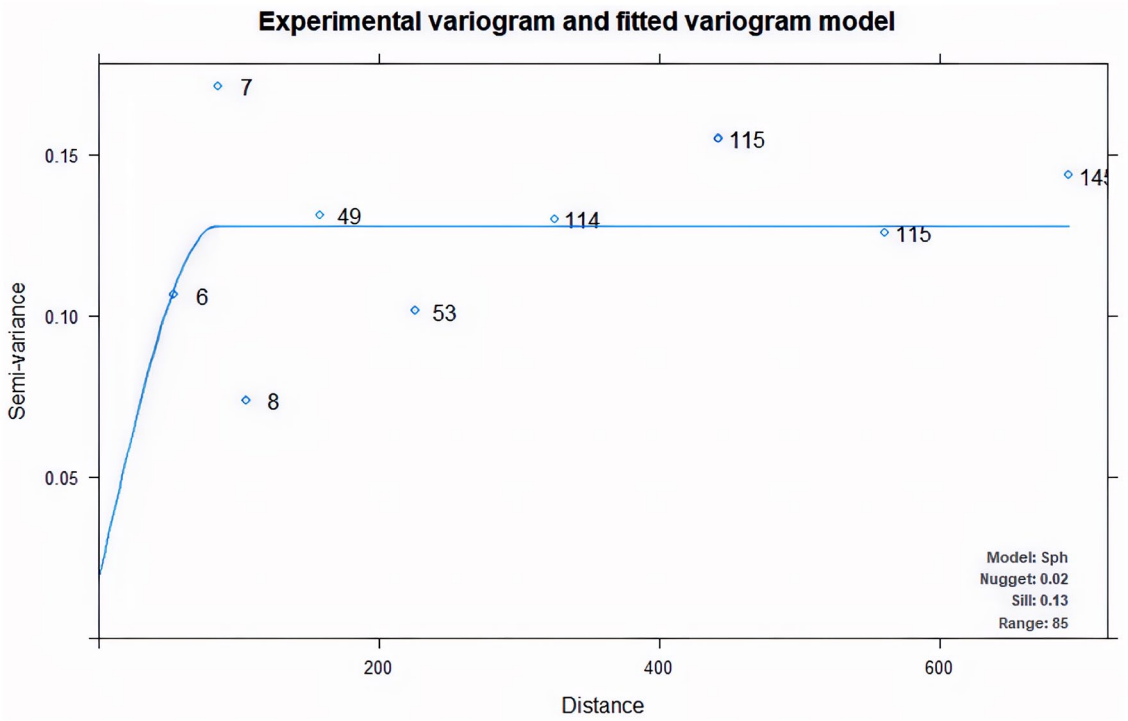

Figure 4. The semi-variogram of SOC in the OK method.

forest is an extension of bagging that is more detailed and also randomly selects subsets of features used in each data sample. It is a versatile method and has been applied in determining many various classifications and regression predictions. For SOC mapping, RF is quite effective when compared to other predictive models (FAO, 2018). In the RF model, the "decision tree" makes a series of conclusions based on a set of features/attributes that are present in the data; on the other hand, bagging is a more general procedure that can be used to reduce the variance for those algorithms that display a high variance (Yohei et al., 2014). RF provides the ability to measure the importance of variables, which in turn quantifies how much each feature influences the accuracy of the RF model (Sekulić et al., 2020). To compare the accuracy of different methods, this study, therefore, uses the same input data for RF as for training data with 47 soil samples used for other methods. The parameters of the number of bootstrap replicates (*ntree*) and the number of variables is randomly sampled at each split (*mtry*) play an essential role in the accuracy of the model. Some studies have stated that satisfactory results can be achieved with the default parameters (Andy & Matthew, 2002; Zhang & Roy, 2017). However, we operated with a multiple *ntree* of 100, 500, 1,000 and a set *mtry* of 1, 2, 3 for further investigation. This study was calculated by using R software with a framework introduced by Tomislav Hengl (Hengl et al., 2018). In general, the RF method can be written as a multi variables regression formula as follows (Sekulić et al., 2020).

$$Z(x_0) = f(x_i(x_0),....x_n(x_0))$$

where $x_i(x_0)$ $(i = 1,2,.....n)$ are covariates at the location (x_0).

Prediction values are made by aggregation of the predictions of the trees. The final prediction will be the most common average value returned by the decision trees that compose the forest (da Silva Júnior et al., 2019).

Assessment accuracy. Eighteen soil samples, (27% of the total soil samples) were used to validate the predicted maps of all the methods. This was evaluated by comparing our observations with predictions at the validated points. In this study, root means square errors (RMSE) were selected as an index of validation. The method which has the lowest RMSE is the most accurate method for SOC mapping in this study. The RMSE was calculated by the following equation (Gia Pham et al., 2019):

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(Z_{oi} - Z_{pi})^2} \quad (4)$$

The prediction accuracy (Acc) was also used to evaluate the accuracy of predicted and measured SOC content (Gao et al., 2021). It was calculated by following equation:

$$Acc = \left[1 - \frac{1}{m}\sum_{i=1}^{m}\left|\frac{Z_{pi} - Z_{oi}}{Z_{oi}}\right|\right]$$

Table 1. Summary of SOC at the sampling points and environmental variables of the entire area of the research site.

INDICATOR	MIN	MAX	MEDIAN	MEAN	SD
SOC (%)	0.26	1.73	0.81	0.86	0.34
NDVI	0.30	0.71	0.55	0.54	0.08
Elevation (m)	130	301	170	175	29.8
Slope (°)	0.34	46.94	9.11	10.57	7.17

where m is the number of validation points, Z_{oi} is the observed value at the i^{th} position, and Z_{pi} is the predicted value at the i^{th} position.

In general, *RMSE* closer to zero and a higher *Acc* give a better prediction.

Results

Description of soil organic carbon and environmental variables

The SOC content of 65 soil samples is presented in Table 1. The data shows that the SOC content of the research site fluctuates widely, with the lowest value being 0.26%, and the highest value being 1.73% of soil weight. The coefficient of variation of SOC content is 0.39 and it is considered to be a low variance. Therefore, the level of dispersion of SOC value around the mean is lower. The soil samples' SOC content on the western side of the research site is higher than in other locations. Although the skewness index, (0.58, less than 1), indicates that the data may be a normal distribution, however, through the variance of samples and the mean of the data set it shows that it does not meet the criterion of normal distribution.

There is some variance of SOC content between different land use types. The highest SOC value was observed in the acacia land use type with the mean value of 0.89%, followed by cassava and rubber, 0.80% and 0.79%, respectively.

The NDVI value ranges from 0.3 to 0.7 and 75% of the total areas have an NDVI value that is higher than 0.5. This means that the land cover surface within the research site is significantly dense. There is no clear difference in NDVI values in relation to land uses. In regard to land use, acacia plantation has an average NDVI value of 0.56; the rubber tree area is 0.54 and cassava area is 0.52. The western part of the study area also has a higher NDVI value than other regions.

Many researchers found that the elevation level significantly impacted the SOC content in agricultural and forest land areas due to the occurrence of soil erosion (Feng-bo et al., 2015; Joel et al., 2016). The sampling data showed that elevation has a correlation not only to SOC content but also to other environmental variables, as shown in Table 2.

Soil organic carbon interpolation

The spatial distribution of the predicted SOC percentage is shown in Figure 5 using three methods. The results showed

that they were different when using different methods, especially between the environmental variable (RF) and other methods (IDW, OK). The maps of both IDW and OK are almost the same. The variance of SOC values of these maps is 1.40% even though the SOC content of the IDW method is slightly higher than OK, by 0.01%.

The accuracy of interpolated methods has been shown via the RMSE and Acc values. The RMSE value within the OK model is the highest being 0.29, followed by RF and IDW with 0.28, 0.25, and 0.24, respectively. The prediction accuracy (Acc) of the IDW, RF and OK method was 0.56, 0.50 and 0.42 respectively. These indicates that the IDW method is the most accurate in comparison to the other methods we used. The procedure of the IDW method is also the simplest, which makes it more advantageous to use than other more complex methods that require much more data input.

The advantage of the RF model is in showing the importance of the effect of the given variables. It explains the influence of the predictor variable upon a given dependent variable. For example, in our model it shows that elevation is most relevant to the increase of Mean Square Error (MSE) of RF model with 8%, followed by slope and NDVI, 2.5% and 0.5% respectively.

Table 2. The correlation between SOC and environmental variables of 47 soil samples.

	ELEVATION	SLOPE	NDVI	SOC
Elevation	1.00			
Slope	0.36***	1.00		
NDVI	0.28**	−0.02	1.00	
SOC	0.47***	0.14	0.18	1.00

**Significant level of 0.05; ***Significant level of 0.01.

Discussion

Finding the most suitable interpolation methods for soil properties mapping is still a challenge for researchers worldwide. Previous research (Y. H. Wu & Hung, 2016) has indicated that no specific model is the best under all conditions, and each spatial interpolation algorithm preforms differently. This means that in our particular case, IDW might be the most suitable method, but it may not reach the most effective conclusions for

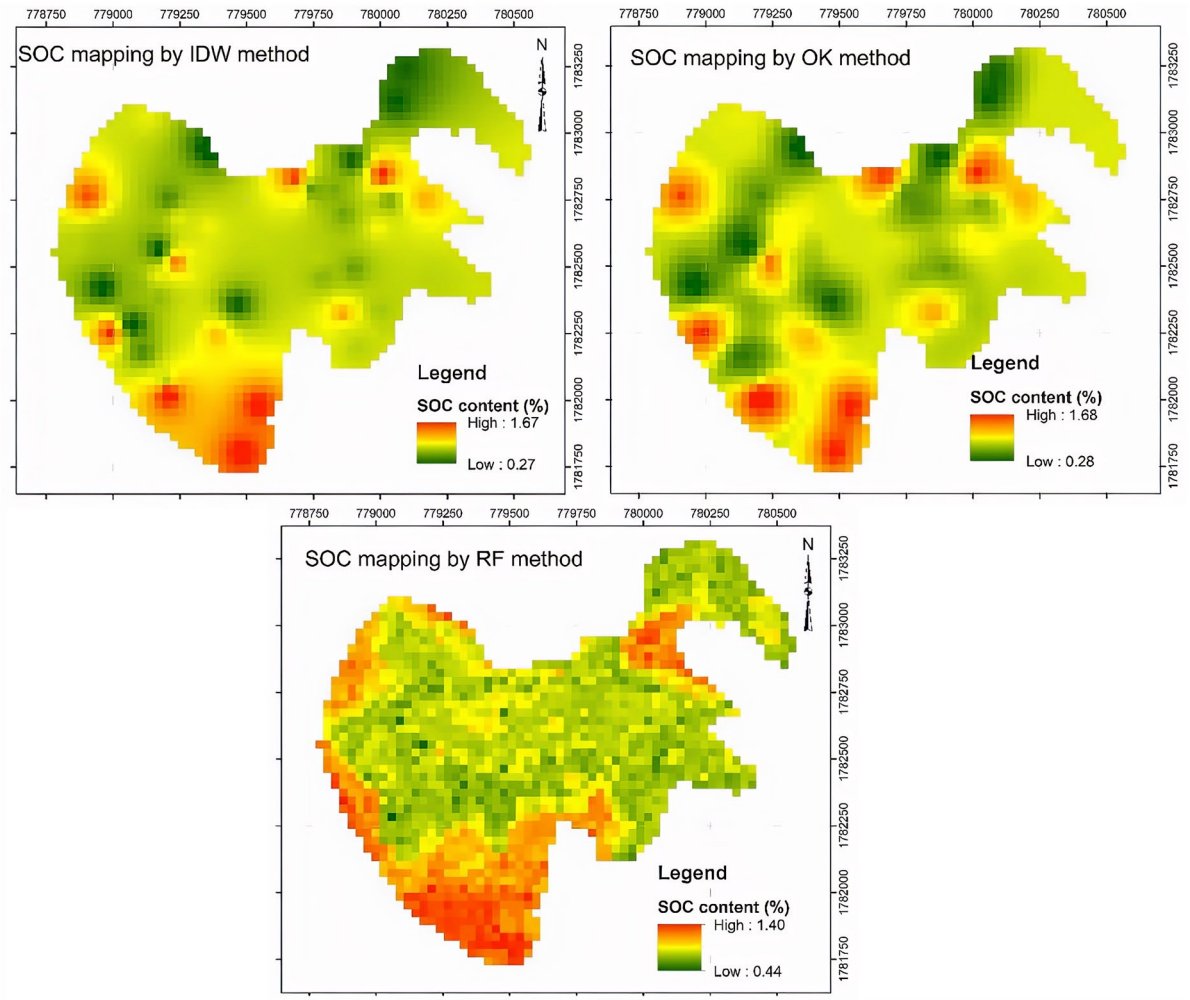

Figure 5. The SOC maps by IDW, OK, and RF methods.

other topographical regions, or when using different sampling methods.

Many researches indicated that the OK model is more accurate than IDW due to the regularity in sample density and high spatial correlation (Zimmerman et al., 1999). However, other researchers (Qiao et al., 2018; Setianto & Triandini, 2013) found that the IDW model is more suitable for ecological interpolation than the OK technique. The IDW method can predict the changing of spatial feature interpolation better than the OK method (Shi et al., 2007; Zhao et al., 2019). Our research finds that the OK method has highest RMSE value and lowest Acc index which means that OK is least method in comparison to other methods used. Contrary to this, the highest in Acc and lowest in $RMSE$ is the IDW method, which subsequently was determined to be the most accurate. There are many reasons that can explain this situation, however, the most important factor is attributed to that of the interpolation data itself. The kriging predictions are dependent on data that satisfies the statistical criteria in an unbiased way and takes into consideration variances (Ikechukwu et al., 2017). A prerequisite of using the kriging method is that the data set must be used with normal distribution (Gorai & Kumar, 2013; Wu et al., 2006). This is an important requirement when using the OK method (Ikechukwu et al., 2017). In addition to this, the data must be distributed consistently with the rule that 68% of all values of x fall between $\mu \pm 1\sigma$; 95% of all values of x will fall between $\mu \pm 2\sigma$; and 99.7% of all values of x will fall between $\mu \pm 3\sigma$, with x as the SOC content, μ represents mean of SOC content, and σ is the variance of the sample. The dataset used in this research of spatial interpolation is 47 samples and only 38%, 45%, and 75% of samples respectively are distributed according to the parameters mentioned above. The lack of a normal distribution within the results of the OK method also shows it to be less effective than the other methods mentioned in this article.

An optimal choice is the semi-variogram model as an important application for spatial interpolation when using the kriging methods. Since the semi-variogram expresses the relationship between measured values, it is obvious that model recognition strongly influences the evaluation process (Mazzella & Mazzella, 2013). As mentioned, our dataset has only 47 sample points, and the range value of the semi-variogram is quite minimal, only 85 m. It proves that the participation of outliers outside the 85 m range for each interpolated point is not large. In this instance, the accuracy will be negatively affected when using the OK method. These findings are common with individual models and small data sets (Biswas & Cheng, 2013). Moreover, the data does not fit to the function in the semi-variogram, which is also a factor in determining the accuracy of the OK method. This again shows it to be inferior worst because it does not determine important criterion in a satisfactory way within the kriging interpolation group (Ikechukwu et al., 2017).

When the RF model is used, the auxiliary variables also impact the accuracy of the RF model. In the RF model, when predicting a value at a given location, the spatial information in the neighboring locations is not taken into consideration (Hengl et al., 2018; Leo, 2001). Mariano and Mónica (2021), shows that the performance of RF is based on the characteristics of the training dataset. Therefore, when compared with the kriging method, the accuracy of RF is better because it is not affected by the sampling location and distance between samples. Our research was conducted in a small area with homogeneous natural conditions; therefore, the density of samples and their distances are similar to previous studies. As was found in previous research (Xie et al., 2020), the IDW approach works well under uniform sample distribution due to the local variance, which is a driver of the estimated surface. RF accuracy for each interpolation is different, (even if the initial input dataset is the same) due to the "random nature" of the method. One consideration when evaluating the effectiveness of the various models of research is that the physical conditions of each individual study are unique. And therefore, each study may require a different specific model to be the most effective within the specific conditions of that study. Recently, a combination of the OK and RF (called RFK) or IDW and RF (called RF-IDW) have performed well in current digital soil mapping research (Szabó et al., 2019; Tan et al., 2021). Which suggests that further investigation into the most effective model, or combination of models, should be pursued.

The influences of auxiliary variables on SOC content have been noted by previous research (Calvo de Anta et al., 2020). Our study confirmed the same conclusions about the influences of environmental variables on SOC maps as had been drawn from those previous studies (Abalori et al., 2022). The map created by RF method is shown clearly. According to the RF method, the variations in the SOC map are similar to those of the elevation map. The RF model indicated that the elevation variable is the most important variable of SOC prediction, followed by slope and NDVI. When generating models of predictor variables by RF, the spatial relationships between the points are not taken into account, this means that the RF model only uses the characteristics of environmental variables. It should be noted that the correlation between the environment variables and the SOC is calculated based on the sampled data set. Therefore, the influence of the variables should be carefully considered on the basis of the size of the sample area and the characteristics of the sampling area together.

The impact of ecosystem variables such as elevation on SOC is usually the result of the long-term interactions between climate, vegetation, and soil type (Garten, 2004). Our results were also consistent with previous studies in finding a positive correlation between elevation and SOC content (Grömping, 2009). The temperature is the primary environmental element that governs soil C dynamics through the effects on soil organic matter decomposition. Cooler temperatures at high elevations

Table 3. Comparison of MSE values of different *ntree* and *mtry*.

ntree / *mtry*	100	500	1,000
1	0.17	0.16	0.16
2	0.19	0.17	0.17
3	0.19	0.17	0.17

limit the decomposition of organic matter. In our research, the elevation variance was minimal. In these instances, the correlation of SOC and elevation reached the medium level of correlation. In high-altitude areas, farmers often grow perennial crops such as acacia hybrid and rubber. As a result, the better ground cover shown in the NDVI index is because these areas are often significantly higher than the other study areas. Our findings here are again consistent with previous studies (Yang et al., 2020; Zhang et al., 2019) who found a significant positive correlation between the SOC content and NDVI. However, we also found that the correlation between NDVI and SOC is very low when compared with some other previous studies (Kumar et al., 2016; Rajeev et al., 2015). This difference may be due to the depth of sampling, studies that sampled close to the surface have a higher correlation between NDVI and SOC. The influence of slope on SOC content is not significant, so our results are similar to the recent study mentioned (Jakšić et al., 2021). This may be due to the influence of farming practices. In areas with high slopes, local farmers often plant forests and perennial crops, while in flat areas, cassava is dominant. Afforestation areas will have more accumulated SOC, (due to leaf decay), but face SOC leaching due to erosion. In the case of normative farming methods, flat areas often receive SOC washed away from areas with steep slopes, but SOC is lost due to those same farming practices (Karchegani et al., 2012). Therefore, the correlation between SOC and slope is complex and unpredictable.

Concerning the impact of *mtry* and *ntree* within the RF model, our results indicated that for the RF interpolation method for multi variables, the most suitable is 1 and 1000 respectively, seen in Table 3. (Sekulić et al., 2020) introduced that $mtry = \frac{m}{3}$ (where m is the number of auxiliary variables). We used three environmental variables which made an *mtry* of 1 consistent with previous researchers. In regard to the *ntree* value, we recognized that there is no significance of MSE when the *ntree* value changes. Our research area was small, and the number of samples was not large. The distances between sampling locations was also short. These factors all contribute to the effectiveness of the *ntree* value. Another previous study also stated that the optimal value of *ntree* depended on the degree of spatial correlation and the sample size (Yohei et al., 2014). The *ntree* value is proportional to the sample size and range of semi-variogram. If the *ntree* increases, it does not always mean the performance of the RF is significantly improved (Oshiro et al., 2012). In general, in a small area with a dense sample area, the $mtry = \frac{m}{3}$ and *ntree* as the default value of the Random Forest package is an appropriate parameter.

Conclusions

The results did not meet our expectations that the environmental auxiliary variable method can improve the accuracy of SOC mapping in small areas. SOC mapping in a small area with a high and random sampling density shows that IDW is the most acceptable interpolation method, followed by RF, and OK techniques. A comparison of the best method for SOC interpolation needs to be conducted with a wide range of auxiliary variables and sampling sizes in the different natural conditions. The evaluation of the distribution of the original data set in the kriging method should be noted because this can be a cause of poor interpolation results. Sampling strategies need to be determined to ensure that the sample size is large enough and also has appropriate sample spacing. This can be a solution to improve the accuracy of the SOC interpolation results.

The selection of auxiliary variables should be considered on the basis of the specific conditions of each study area. In areas where there is human activity such as farming, the conditions of naturally occurring variables can be compromised and therefore these areas may not be suitable for interpolation studies of SOC using the auxiliary variable methods. In these cases, the research should focus on variables related to human activities and land-use types, and also should take into account the use of fertilizers, or other substances used in modern farming methods.

In the RF method, the *mtry* value is a third of the number of variables and the *ntree* value as the default of the RF package. This should be taken as the most important parameter.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors also acknowledge the support of Hue University under the Core Research Program, Grant No. NCM.DHH.2019.06

ORCID iD

Chuong Van Huynh https://orcid.org/0000-0003-1383-6491

REFERENCES

Abalori, T. A., Cao, W., Weobong, C. A.-A., Wang, S., Anning, D. K., Sam, F. E., Liu, W., & Wang, W. (2022). Spatial variability of soil organic carbon fractions and aggregate stability along an elevation gradient in the alpine meadow grasslands of the Qilian Mountains, China. *Chilean Journal of Agricultural Research*, 82(1), 52–64. https://doi.org/10.4067/s0718-58392022000100052

Andy, L., & Matthew, W. (2002). Classification and regression by random Forest. *R Newa*, 2–3, 18–22. https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf

Bhunia, G. S., Shit, P. K., & Maiti, R. (2018). Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *Journal of the Saudi Society of Agricultural Sciences*, 17(2), 114–126. https://doi.org/10.1016/j.jssas.2016.02.001

Biswas, A., & Cheng, B. (2013). Model averaging for semivariogram model parameters. In *Advances in agrophysical research*. InTech. https://doi.org/10.5772/52339.

Black, C. (1965). Methods of soil analysis (A. G. Norman (Ed.)). *American Society of Agronomy, Soil Science Society of America*. https://doi.org/10.2134/agronmonogr9.2

Bostani, A., Salahedin, M., Rahman, M. M., & Khojasteh, D. N. (2017). Spatial mapping of soil properties using geostatistical methods in the Ghazvin Plains of Iran. *Modern Applied Science*, 11(10), 23. https://doi.org/10.5539/mas.v11n10p23

Calvo de Anta, R., Luís, E., Febrero-Bande, M., Galiñanes, J., Macías, F., Ortíz, R., & Casás, F. (2020). Soil organic carbon in peninsular Spain: Influence of environmental factors and spatial distribution. *Geoderma*, 370, 114365. https://doi.org/10.1016/j.geoderma.2020.114365

Châu, T. T. M. (2020). Application of kriging regression to soil organic carbon mapping: A case study in Huong Lam commune, A Luoi district, Thua Thien Hue province. *Hue University Journal of Science: Agriculture and Rural Development*, 129(3A), 1–10. https://doi.org/10.26459/hueuni-jard.v129i3A.5545

Cressie, N. A. C. (1993). *Statistics for spatial data* (Revised ed.). John Wiley & Sons, Inc.

da Silva Júnior, J. C., Medeiros, V., Garrozi, C., Montenegro, A., & Gonçalves, G. E. (2019). Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian's Northeast. *Computers and Electronics in Agriculture*, 166, 105017. https://doi.org/10.1016/j.compag.2019.105017

FAO. (2018). *Soil Organic Carbon Mapping Cookbook* (Y. Yusuf, B. Rainer, & V. Ronald (Eds.), 2nd ed.). Food and Agriculture Organization of the United Nations.

FAO. (2020). Global Soil Organic Carbon Map (GSOCmap) Version 1.5.Author. https://doi.org/10.4060/ca7597en

Feng-bo, L., Guang-de, L., Xi-yue, Z., Hui-xiang, N., Chun-chun, X., Chao, Y., Xiu-mei, Y., Jin-fei, F., & Fu-ping, F. (2015). Elevation and land use types have significant impacts on spatial variability of soil organic matter content in Hani terraced field of Yuanyang County, China. *Rice Science*, 22(1), 27–34. https://doi.org/10.1016/j.rsci.2015.05.005

Gao, L., Huang, M., Zhang, W., Qiao, L., Wang, G., & Zhang, X. (2021). Comparative Study on spatial digital mapping methods of soil nutrients based on different geospatial technologies. *Sustainability*, 13(6), 3270. https://doi.org/10.3390/su13063270

Garten, C. J. (2004). *Soil carbon dynamics along an elevation gradient in the southern Appalachian Mountains*. https://info.ornl.gov/sites/publications/Files/Pub57407.pdf

Gia Pham, T., Kappas, M., Van Huynh, C., & Hoang Khanh Nguyen, L. (2019). Application of ordinary kriging and regression kriging method for soil properties mapping in hilly region of central Vietnam. *ISPRS International Journal of Geo-Information*, 8(3), 147. https://doi.org/10.3390/ijgi8030147

Göl, C., Bulut, S., & Bolat, F. (2017). Comparison of different interpolation methods for spatial distribution of soil organic carbon and some soil properties in the Black Sea backward region of Turkey. *Journal of African Earth Sciences*, 134, 85–91. https://doi.org/10.1016/j.jafrearsci.2017.06.014

Gorai, A., & Kumar, S. (2013). Spatial distribution analysis of groundwater quality index using GIS: A case study ofm Ranchi Municipal Corporation (RMC) area. *Geoinformatics & Geostatistics An Overview*, 1(02), 1–11. https://doi.org/10.4172/2327-4581.1000105

Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *The American Statistician*, 63(4), 308–319. https://doi.org/10.1198/tast.2009.08199

Hartemink, A. E., & McSweeney, K. (2014). Soil carbon. In A. E. Hartemink & K. McSweeney (Eds.), *Soil carbon* (pp. 7–16). Springer International Publishing.

Hengl, T. (2009). *A practical guide to geostatistical mapping* (2nd ed.). Office for Official Publications of the European Communities. http://spatial-analyst.net/book/system/files/Hengl_2009_GEOSTATe2c1w.pdf

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e5518. https://doi.org/10.7717/peerj.5518

Hu, P.-L., Liu, S.-J., Ye, Y.-Y., Zhang, W., Wang, K.-L., & Su, Y.-R. (2018). Effects of environmental factors on soil organic carbon under natural or managed vegetation restoration. *Land Degradation and Development*, 29(3), 387–397. https://doi.org/10.1002/ldr.2876

https://earthexplorer.usgs.gov/ (n.d.). Retrieved March 15, 2021, from https://earthexplorer.usgs.gov/

Ikechukwu, M. N., Ebinne, E., Idorenyin, U., & Raphael, N. I. (2017). Accuracy assessment and comparative analysis of IDW, spline and kriging in spatial interpolation of landform (topography): An experimental study. *Journal of Geographic Information System*, 9(3), 354–371. https://doi.org/10.4236/jgis.2017.93022

Jakšić, S., Ninkov, J., Milić, S., Vasin, J., Živanov, M., Jakšić, D., & Komlen, V. (2021). Influence of slope gradient and aspect on soil organic carbon content in the region of Niš, Serbia. *Sustainability*, 13(15), 8332. https://doi.org/10.3390/su13158332

Jat, H. S., Datta, A., Choudhary, M., Sharma, P. C., Yadav, A. K., Choudhary, V., Gathala, M. K., Jat, M. L., & McDonald, A. (2019). Climate Smart Agriculture practices improve soil organic carbon pools, biological properties and crop productivity in cereal-based systems of north-West India. *CATENA*, 181, 104059. https://doi.org/10.1016/j.catena.2019.05.005

Joel, L. M., Balthazar, M. M., Didas, N. K., John, B. H., Proches, M., LothS, H., Jozef, M., & Hubert, D. (2016). Variability of soil organic carbon with landforms and land use in the Usambara Mountains of Tanzania. *Journal of Soil Science and Environmental Management*, 7(9), 123–132. https://doi.org/10.5897/JSSEM2016.0557

John, M. K., Charles, W. R., Debbie, R., Mooney, S., Ronald, F. F., & Rattan, L. (2007). *Soil carbon Management* (J. M. Kimble, C. W. Rice, D. Reed, S. Mooney, R. F. Follett, & R. Lal (Eds.)). CRC Press.

Karchegani, P. M., Ayoubi, S., Mosaddeghi, M. R., & Honarjoo, N. (2012). Soil organic carbon pools in particle-size fractions as affected by slope gradient and land use change in hilly regions, western Iran. *Journal of Mountain Science*, 9(1), 87–95. https://doi.org/10.1007/s11629-012-2211-2

Kumar, P., Pandey, P. C., Singh, B. K., Katiyar, S., Mandal, V. P., Rani, M., Tomar, V., & Patairiya, S. (2016). Estimation of accumulated soil organic carbon stock in tropical forest using geospatial strategy. *The Egyptian Journal of Remote Sensing and Space Science*, 19(1), 109–123. https://doi.org/10.1016/j.ejrs.2015.12.003

Lai, Y.-Q., Wang, H.-L., & Sun, X.-L. (2021). A comparison of importance of modelling method and sample size for mapping soil organic matter in Guangdong, China. *Ecological Indicators*, 126, 107618. https://doi.org/10.1016/j.ecolind.2021.107618

Leo, B. (2001). Random forests. *Machine Learning*, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Liu, T., Zhang, H., & Shi, T. (2020). Modeling and predictive mapping of soil organic carbon density in a small-scale area using geographically weighted regression kriging approach. *Sustainability*, 12(22), 9330. https://doi.org/10.3390/su12229330

Liu, Z., Shao, M., & Wang, Y. (2011). Effect of environmental factors on regional soil organic carbon stocks across the Loess Plateau region, China. *Agriculture Ecosystems & Environment*, 142(3–4), 184–194. https://doi.org/10.1016/j.agee.2011.05.002

Maleika, W. (2020). Inverse distance weighting method optimization in the process of digital terrain model creation based on data collected from a multibeam echosounder. *Applied Geomatics*, 12(4), 397–407. https://doi.org/10.1007/s12518-020-00307-6

Mariano, C., & Mónica, B. (2021). A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. *Computers and Electronics in Agriculture*, 184, 106094. https://doi.org/10.1016/j.compag.2021.106094

Mazzella, A., & Mazzella, A. (2013). The importance of the model choice for experimental semivariogram modeling and its consequence in evaluation process. *Engineering Journal*, 2013, 1–10. https://doi.org/10.1155/2013/960105

Meersmans, J., Martin, M. P., Lacarce, E., De Baets, S., Jolivet, C., Boulonne, L., Lehmann, S., Saby, N. P. A., Bispo, A., & Arrouays, D. (2012). A high resolution map of French soil organic carbon. *Agronomy for Sustainable Development*, 32(4), 841–851. https://doi.org/10.1007/s13593-012-0086-9

Mesić Kiš, I. (2016). Comparison of ordinary and Universal Kriging interpolation techniques on a depth variable (a case of linear spatial trend), case study of the Šandrovac Field. *Rudarsko-Geolosko-Naftni Zbornik*, 31(2), 41–58.

Milne, E., Banwart, S. A., Noellemeyer, E., Abson, D. J., Ballabio, C., Bampa, F., Bationo, A., Batjes, N. H., Bernoux, M., Bhattacharyya, T., Black, H., Buschiazzo, D. E., Cai, Z., Cerri, C. E., Cheng, K., Compagnone, C., Conant, R., Coutinho, H. L. C., de Brogniez, D., . . . Zheng, J. (2015). Soil carbon, multiple benefits. *Environmental Development*, 13, 33–38. https://doi.org/10.1016/j.envdev.2014.11.005

Mishra, U., Gautam, S., Riley, W. J., & Hoffman, F. M. (2020). Ensemble machine learning approach improves predicted spatial variation of surface soil organic carbon stocks in data-limited Northern Circumpolar Region. *Frontiers in Big Data*, 3, 528441. https://doi.org/10.3389/fdata.2020.528441

Nam Dong District People's Committee. (2020). *Land use planning of Nam Dong district, Thua Thien Hue province*.

Naresh, D. R. K. (Ed.). (2020). *Advances in Agriculture Sciences*. AkiNik Publications.

Omuto, C. T., & Vargas, R. R. (2015). Re-tooling of regression kriging in R for improved digital mapping of soil properties. *Geosciences Journal*, 19(1), 157–165. https://doi.org/10.1007/s12303-014-0023-9

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? In *Machine learning and data mining in pattern recognition* (pp. 154–168). https://doi.org/10.1007/978-3-642-31537-4_13

Piccini, C., Francaviglia, R., & Marchetti, A. (2020). Predicted maps for soil organic matter evaluation: The case of Abruzzo Region (Italy). *Land*, 9(10), 349. https://doi.org/10.3390/land9100349

Pouladi, N., Møller, A. B., Tabatabai, S., & Greve, M. H. (2019). Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. *Geoderma*, 342, 85–92. https://doi.org/10.1016/j.geoderma.2019.02.019

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9(2), 181–199. https://doi.org/10.1007/s10021-005-0054-1

Qiao, P., Lei, M., Yang, S., Yang, J., Guo, G., & Zhou, X. (2018). Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing. *Environmental Science and Pollution Research*, 25(16), 15597–15608. https://doi.org/10.1007/s11356-018-1552-y

Rajeev, R., Ankita, J., & Ajeet, S. N. (2015). Soil organic carbon estimation using remote sensing in *Tarai* region of Uttarakhand. *Annals of Plant and Soil Research*, 17, 361–364.

Ramifehiarivo, N., Brossard, M., Grinand, C., Andriamananjara, A., Razafimbelo, T., Rasolohery, A., Razafimahatratra, H., Seyler, F., Ranaivoson, N., Rabenarivo, M., Albrecht, A., Razafindrabe, F., & Razakamanarivo, H. (2017). Mapping soil organic carbon on a national scale: Towards an improved and updated map of Madagascar. *Geoderma Regional*, 9, 29–38. https://doi.org/10.1016/j.geodrs.2016.12.002

Salekin, S., Burgess, J., Morgenroth, J., Mason, E., & Meason, D. (2018). A comparative study of three non-geostatistical methods for optimising digital elevation model interpolation. *ISPRS International Journal of Geo-Information*, 7(8), 300. https://doi.org/10.3390/ijgi7080300

Sekulić, A., Kilibarda, M., Heuvelink, G. B. M., Nikolić, M., & Bajat, B. (2020). Random forest spatial interpolation. *Remote Sensing*, 12(10), 1687. https://doi.org/10.3390/rs12101687

Setianto, A., & Triandini, T. (2013). Comparison of Kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis. *Journal of Southeast Asian Applied Geology*, 5(1), 21–29.

Shi, Y., Li, L., Zhang, L., & Pu, Y. (2007). *Application and comparing of IDW and Kriging interpolation in spatial rainfall information* (J. Chen & Y. Pu (Eds.); p. 67531I). https://doi.org/10.1117/12.761859

Siyu, Z. (2013). *Comparison of statistical methods for digital soil mapping of sub-Saharan Africa*. Wageningen University.

Song, X.-D., Wu, H.-Y., Ju, B., Liu, F., Yang, F., Li, D.-C., Zhao, Y.-G., Yang, J.-L., & Zhang, G.-L. (2020). Pedoclimatic zone-based three-dimensional soil organic carbon mapping in China. *Geoderma*, 363, 114145. https://doi.org/10.1016/j.geoderma.2019.114145

Srivastava, P. K., Pandey, P. C., Petropoulos, G. P., Kourgialas, N. N., Pandey, V., & Singh, U. (2019). GIS and remote sensing aided information for soil moisture estimation: A comparative study of interpolation techniques. *Resources*, 8(2), 70. https://doi.org/10.3390/resources8020070

Su, P., Lin, D., & Qian, C. (2018). Study on air pollution and control investment from the perspective of the environmental theory model: A case study in China, 2005–2014. *Sustainability*, 10(7), 2181. https://doi.org/10.3390/su10072181

Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., & Pásztor, L. (2019). Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. *Hydrology and Earth System Sciences*, 23(6), 2615–2635. https://doi.org/10.5194/hess-23-2615-2019

Tajik, S., Ayoubi, S., & Zeraatpisheh, M. (2020). Digital mapping of soil organic carbon using ensemble learning model in Mollisols of Hyrcanian forests, northern Iran. *Geoderma Regional*, 20, e00256. https://doi.org/10.1016/j.geodrs.2020.e00256

Tan, J., Xie, X., Zuo, J., Xing, X., Liu, B., Xia, Q., & Zhang, Y. (2021). Coupling random forest and inverse distance weighting to generate climate surfaces of precipitation and temperature with multiple-covariates. *Hydrology Journal*, 598, 126270. https://doi.org/10.1016/j.jhydrol.2021.126270

Tung, Y. (1983). Point rainfall estimation for a mountainous region. *Journal of Hydraulic Engineering*, 109(10), 1386–1393. https://doi.org/10.1061/(asce)0733-9429(1983)109:10(1386)

U.S. Geological Survey. (2019). *Landsat 8 (L8) data users handbook* (Version 5.). Author. https://www.usgs.gov/media/files/landsat-8-data-users-handbook

Vågen, T.-G., & Winowiecki, L. A. (2013). Mapping of soil organic carbon stocks for spatially explicit assessments of climate change mitigation potential. *Environmental Research Letters*, 8(1), 015011. https://doi.org/10.1088/1748-9326/8/1/015011

van Den, F., Berg, A., Tiktak, T., Hoogland, A., Poot, J. J. T. I., Boesten, A. M. A., van der Linden, J. W., & Pol. (2017). *An improved Soil Organic Matter map for GeoPEARL_NL, Model description of version 4.4.4 and consequences for the Dutch decision tree on leaching to groundwater*. https://edepot.wur.nl/424920

Wu, J., Norvell, W. A., & Welch, R. M. (2006). Kriging on highly skewed data for DTPA-extractable soil Zn with auxiliary information for pH and organic carbon. *Geoderma*, 134(1–2), 187–199. https://doi.org/10.1016/j.geoderma.2005.11.002

Wu, Y. H., & Hung, M. C. (2016). Applications of spatial statistics. In M. C. Hung (Ed.), *Comparison of spatial interpolation techniques using visualization and Quantitative Assessment* (pp. 2–16). InTech. https://doi.org/10.5772/65996.

Xie, B., Jia, X., Qin, Z., Zhao, C., & Shao, M. (2020). Comparison of interpolation methods for soil moisture prediction on China's Loess Plateau. *Vadose Zone Journal*, 19(1), 2–16. https://doi.org/10.1002/vzj2.20025

Yang, L., He, X., Shen, F., Zhou, C., Zhu, A.-X., Gao, B., Chen, Z., & Li, M. (2020). Improving prediction of soil organic carbon content in croplands using phenological parameters extracted from NDVI time series data. *Soil and Tillage Research*, 196, 104465. https://doi.org/10.1016/j.still.2019.104465

Yohei, M., Masamitsu, T., & Hironobu, F. (2014). *Boosted random forest* [Conference session]. Proceedings of the 9th International Conference on Computer Vision TheoryandApplications,pp.594–598.https://doi.org/10.5220/0004739005940598

Zeraatpisheh, M., Ayoubi, S., Mirbagheri, Z., Mosaddeghi, M. R., & Xu, M. (2021). Spatial prediction of soil aggregate stability and soil organic carbon in aggregate fractions using machine learning algorithms and environmental variables. *Geoderma Regional*, 27, e00440. https://doi.org/10.1016/j.geodrs.2021.e00440

Zeraatpisheh, M., Garosi, Y., Reza Owliaie, H., Ayoubi, S., Taghizadeh-Mehrjardi, R., Scholten, T., & Xu, M. (2022). Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. *CATENA*, 208, 105723. https://doi.org/10.1016/j.catena.2021.105723

Zhang, G. L., Liu, F., & Song, X. D. (2017). Recent progress and future prospect of digital soil mapping: A review. *Journal of Integrative Agriculture*, 16(12), 2871–2885. https://doi.org/10.1016/s2095-3119(17)61762-3

Zhang, H. K., & Roy, D. P. (2017). Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. *Remote Sensing of Environment*, 197, 15–34. https://doi.org/10.1016/j.rse.2017.05.024

Zhang, Y., Guo, L., Chen, Y., Shi, T., Luo, M., Ju, Q., Zhang, H., & Wang, S. (2019). Prediction of soil organic carbon based on Landsat 8 monthly NDVI data for the Jianghan Plain in Hubei Province, China. *Remote Sensing*, 11(14), 1683. https://doi.org/10.3390/rs11141683

Zhao, W., Cao, T., Li, Z., & Sheng, J. (2019). Comparison of IDW, cokriging and ARMA for predicting spatiotemporal variability of soil salinity in a gravel–sand mulched jujube orchard. *Environmental Monitoring and Assessment*, 191(6), 376. https://doi.org/10.1007/s10661-019-7499-8

Zimmerman, D., Pavlik, C., Ruggles, A., & Armstrong, M. P. (1999). An experimental comparison of ordinary and Universal Kriging and inverse distance weighting. *Mathematical Geology*, 31, 375–390.