



Spatial prediction of soil organic carbon in coal mining subsidence areas based on RBF neural network

Qiangqiang Qi¹ · Xin Yue¹ · Xin Duo¹ · Zhanjun Xu¹ · Zhe Li¹

Received: 4 November 2021 / Revised: 20 July 2022 / Accepted: 25 March 2023
© The Author(s) 2023

Abstract

A quantitative research on the effect of coal mining on the soil organic carbon (SOC) pool at regional scale is beneficial to the scientific management of SOC pools in coal mining areas and the realization of coal low-carbon mining. Moreover, the spatial prediction model of SOC content suitable for coal mining subsidence area is a scientific problem that must be solved. Taking the Changhe River Basin of Jincheng City, Shanxi Province, China, as the study area, this paper proposed a radial basis function neural network model combined with the ordinary kriging method. The model includes topography and vegetation factors, which have large influence on soil properties in mining areas, as input parameters to predict the spatial distribution of SOC in the 0–20 and 2040 cm soil layers of the study area. And comparing the prediction effect with the direct kriging method, the results show that the mean error, the mean absolute error and the root mean square error between the predicted and measured values of SOC content predicted by the radial basis function neural network are lower than those obtained by the direct kriging method. Based on the fitting effect of the predicted and measured values, the R^2 obtained by the radial basis artificial neural network are 0.81, 0.70, respectively, higher than the value of 0.44 and 0.36 obtained by the direct kriging method. Therefore, the model combining the artificial neural network and kriging, and considering environmental factors can improve the prediction accuracy of the SOC content in mining areas.

Keywords Mining area · Soil organic carbon · Radial basis function neural network · Environmental factor · Spatial prediction

1 Introduction

China's coal production was 3.90×10^9 tonnes in 2020, accounting for 50.4% of total world coal production (IEA 2021). Coal mining process have caused serious damage to the land resources, crop production, and the ecological environment (Shrestha and Lal 2011), and induced a very severe disturbance of the soil organic carbon (SOC) pool.

Considering that the SOC pool is the largest potential factor in reducing the carbon emissions of terrestrial ecosystems (Miller et al. 2015; Zhang and Ni 2017), China and other major coal-producing countries in the world must quantitatively study the disturbing influence of coal mining on the SOC pool, so as to improve scientific management

of SOC pool in coal mining areas and realize regional land low-carbon utilization.

At present, due to the frequent human mining activities, which will result in the changes of some ecological environmental factors in coal mining subsidence areas such as land subsidence (Liu et al. 2021), surface destruction (Redondo-Vega et al. 2017), soil erosion (Wang et al. 2020; Su 2021), vegetation destruction (Li et al. 2016), surface runoff and groundwater hydrology (Hu 2021; Song et al. 2021), etc. No matter which factor changes, it will have an impact on the soil carbon pool, so that the SOC content in the mining area usually has strong spatial variability (Cheng et al. 2014; Jun et al. 2015).

In recent years, a plenitude amount of work has been conducted in the impact of coal mining on the SOC pool. For instance, Fu (2017) analyzed the distribution of SOC and the liable organic carbon fraction in the typical subsidence wetland. Furthermore, the main impact factors of SOC formation and distribution have also been studied. In addition, subsidence wetland with different utilization types has been

✉ Zhanjun Xu
zjxu163@126.com

¹ Institute of Land Science, College of Resources and Environment, Shanxi Agricultural University, Taigu 030801, China

chosen to study the human impact on SOC. Huang (2014) found that the carbon sink amount of the vegetation-soil system, affected by coal mining, reduced in the Xinzhouyao coal mine, Datong Mining Area, Shanxi. In order to understand carbon dynamics in mine soil, the spatial variation of SOC contents was investigated in two types of landscapes destroyed by coal mining, i.e., subsidence slope and ground fissure site from Jiaozuo mine area, China (Cheng et al. 2014). Many studies of SOC pool in farmland ecosystems have also been conducted. Tian (2020) took the Changhe Basin mining area as an example and established a method for estimating the carbon sequestration loss of farmland ecosystem caused by coal mining, concluding that the influence of coal mining on carbon sequestration in the farmland ecosystem belongs to a carbon loss effect. The carbon loss effect of coal mining on SOC pool in farmland has also been demonstrated in another study (Xu et al. 2019). These studies have shown that when the soil in mining areas is damaged, the carbon stored in the soil also decreases massively. Therefore, understanding the spatial distribution of SOC in mining areas is of great significance for controlling greenhouse gas emissions and land management in mining areas.

To efficiently and accurately understand the spatial distribution of SOC, various geostatistical methods have been applied to predict SOC. Kriging is one of the most widely used methods among the stochastic techniques and is the best linear unbiased estimator in the sense that it minimizes the variance of the estimation error (Dai et al. 2014; Ren et al. 2021). Therefore, it shows considerable advantages in SOC prediction. However, this method does not consider the relationship between soil properties and environmental factors. Based on the shortcomings of this method, prediction models for SOC, taking into account environmental factors, began to develop. It mainly includes multiple linear regression (MLR) (Zhang et al. 2017), regression kriging (RK) (Zhang et al. 2012), and geographically weighted regression model (GWR) (Wang and Wu 2020). Kriging and regression analysis are both based on the linear relationship between the target and environmental factors, but the relationship between soil and environmental factors is usually a complicated nonlinear relationship. To overcome these problems, machine learning algorithms, driven by big data, have been increasingly applied to spatial prediction of soil organic carbon such as random forest (RF) (Yuan et al. 2021), support vector machine (SVM) (John et al. 2020), artificial neural network (ANN) (Pudelko et al. 2020) and Boosted regression tree (BRT) (Akpa et al. 2016).

Because ANN can automatically learn and analyze the nonlinear relationship between multi-source inputs, researchers have successively applied it to the spatial prediction of SOC, and achieved fairly good prediction performance (Lai et al. 2020). For instance, Morais et al. (2021) combined laboratory NIR spectral data with ANN

to estimate the SOC content of pasture soils in Portugal. Were (2015) compared the performance of SVR, ANN and RF in predicting and mapping SOC stocks in the Eastern Mau Forest Reserve, Kenya. As a traditional ANN, the radial basis neural network (RBFNN) can approximate arbitrary functions with arbitrary accuracy due to its strong nonlinear fitting ability, and is widely used in digital soil mapping. Using RBFNN and high-precision surface model, Luo (2016) achieved high-precision simulation and prediction of the spatial variation of SOC in Purple Soil Hilly area of Mid-sichuan Basin. Lai et al. (2020) used the RBFANN and its model combined with OK (RBFNN-OK) to predict the spatial distribution of SOC content, comparing its performance with MLR, RF, OK.

In summary, many models and methods have been established for the prediction of regional SOC and its spatial distribution, but the model, suitable for prediction of SOC and its spatial distribution in coal mining subsidence areas with intricate terrains, is relatively few and short of relevant case studies. In this paper, the Changhe River Basin was chosen as a study area and the RBFNN was used to predict the spatial distribution of SOC. The prediction precision of this model was compared with the conventional Kriging model to explore a spatial prediction model suitable for soil organic carbon in coal mining subsidence areas.

2 Materials and methods

2.1 Study area

The study area is the Changhe River Basin (35°30'10" N to 35°38'06" N and, 112°40'37" E to 112°46'04" E), which is located in northwest Jincheng City, Shanxi Province, China, with a total coverage of approximately 113.224 km². There are 48 administrative villages in the region including Chuandi Township, and Dadonggou and Xiacun towns. The location is shown in Fig. 1. The study area has a warm-temperate semi-humid continental monsoon climate. The mean annual air temperature, precipitation and sunshine hours are 10.9 °C, 628.3 mm and 2392.8 h, respectively. The area is located on the southeastern edge of the Loess Plateau and the west terrain is higher than the east, with an elevation between 723 and 1174 m. The topographic relief fluctuates greatly, showing a geographical pattern of two mountains and a river. The east and west are mountains and hills with complex terrain and the Changhe River flows from north to south in the middle, forming river valleys in the central region. The main soil type in the area is cinnamon soil, and there is a small amount of meadow soil. Alkaline soil is the main soil type in the hilly area, which is mainly cultivated. Corn, potatoes and wheat are the main crops in the

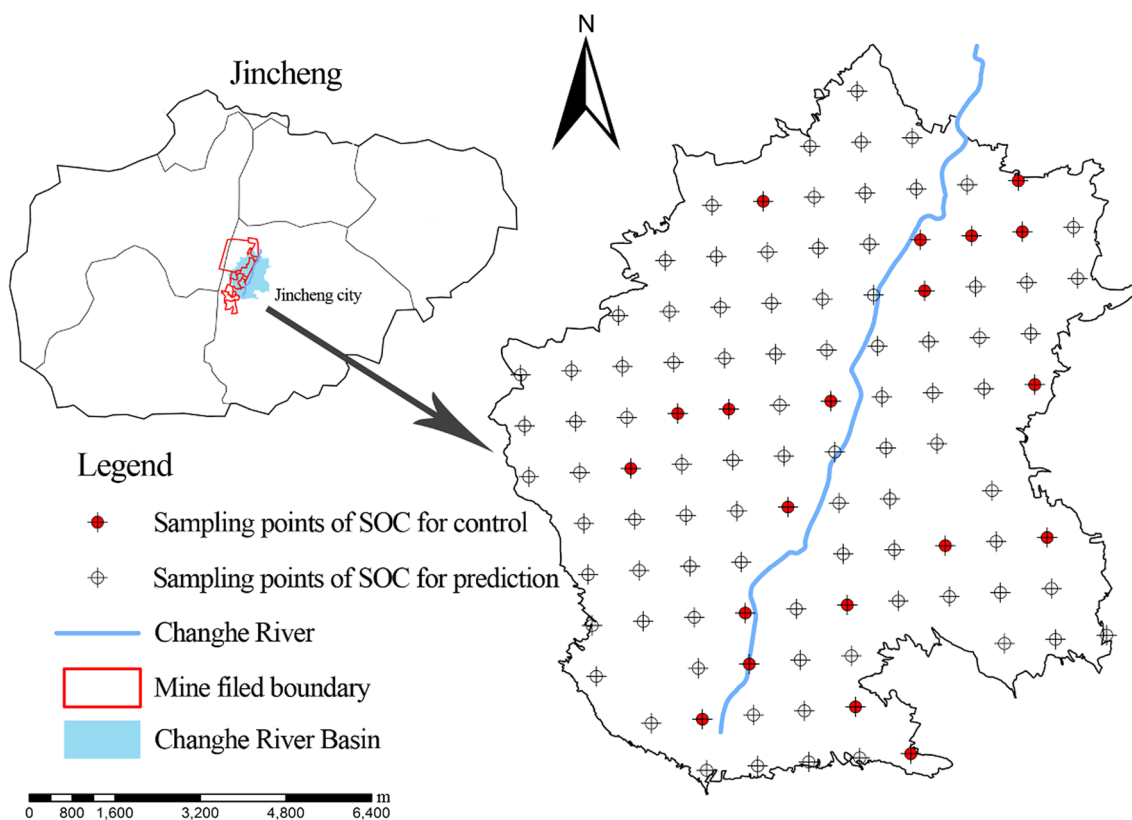


Fig. 1 Location of the study area and distribution of soil sampling sites

agricultural cultivation area, with types of wheat and maize planted according to rotation cropping, producing three crops over two years.

In this area, coal mines are relatively concentrated, with large coal production, abundant coal resources and good coal quality. Currently, coal seams No. 3, 9 and 15 are mainly used. There are several coal mines across the region. The area is therefore affected by high-intensity coal mining and large areas of land have collapsed to different degrees. According to observed data from the mines, after a few decades of mining subsidence, the maximum subsidence is 6500 mm, the maximum incline deformation is 25.7 mm/m, the maximum horizontal movement is 2840 mm. The maximum horizontal deformation is 38.2 mm/m. Therefore, this is a typical study area for coal mining subsidence.

2.2 Soil sampling and analysis

Field sampling was conducted in the Changhe River Basin in July 2015. Based on the location of the study area, the sampling points should be distributed as uniform as possible, and therefore the grid sampling method was used in this study. First, the study area was divided into 1 km \times 1 km grids. Taking the center of the grid as the circle

center and 5 m as the radius, 5 points were set along two diagonal lines in each soil layer. On each grid, five sub-samples of 0–20 cm and 20–40 cm were collected and merged into one composite sample (about 1 kg), respectively. Finally, the soil samples were brought back to the laboratory and their coordinates were recorded using a handheld GPS (Sun et al. 2018). A total of 106 soil samples were collected from each soil layer, and 20 samples were randomly selected as validation samples to validate the accuracy of the SOC prediction model. The remaining 86 samples were used for model prediction. Using the “create subset” function of Geostatistical Analyst in ArcGIS 10.0 to classify these samples. The distribution of the sampling points is shown in Fig. 1.

The soil samples collected outdoors were taken back to the laboratory, air-dried and crushed to pass through a sieve with a 2 mm mesh to remove the animal and plant residues. The soil organic carbon content of the sample was determined using the potassium dichromate ($K_2Cr_2O_7$) oxidation-titration method. During this process, the SOC is oxidized by potassium dichromate and heated to 170–180 °C for approximately 5 min. The excess organic potassium dichromate was then titrated by standard 0.2 mol/L ferrous sulfate ($FeSO_4$) to determine the SOC content (Guo et al. 2019).

2.3 Analytical thinking

2.3.1 Auxiliary variables

Previous studies have shown that factors such as topographic properties (Hao et al. 2002) and vegetation (Lemma et al. 2006; Soleimani et al. 2017), climate change (Coxson and Parkinson 1987) and land use (Brejda et al. 2001) and other factors have a great impact on the spatial distribution of soil properties. For a particular mining subsidence land, climate change is not the dominant factor affecting soil organic carbon change because of the small region, while vegetation and land use are important factors affecting soil organic carbon change.

In the coal mining areas, the goaf areas are formed after the coal mining panels excavated in the study area, which destroyed the original stress equilibrium of the subterranean strata. And the stress transmitted through the stratum induces an inconsistent deformation of the overburden rock above the goaf areas. The stress makes the surface deformation in the horizontal and vertical directions, which changes the topography of the mining subsidence area. The surface deformation caused by coal mining is the root cause of SOC changes in the mining subsidence areas. The conventional indexes describing surface deformation are subsidence, inclination, curvature, horizontal movement, horizontal deformation, distortion and shear deformation. Firstly, the physical changes such as surface subsidence and cracks in the mining area have direct damage to the vegetation (Xu 2012). Secondly, the surface deformation directly affects the erosion intensity of precipitation on surface soil, thereby affecting the loss and accumulation of SOC (Ren et al. 2018).

Coal mining may lead to changes in groundwater systems and surface runoff in subsidence areas, which will change the soil water content, and ultimately change the carbon storage and spatial SOC distribution in mining subsidence areas. The evaporation and infiltration of surface water, water erosion, wind erosion are changed by the cracks and collapses formed by coal mining, which further leads to change of the soil water content (Qie et al. 2015; Wu et al. 2019; Mo et al. 2015). Change in soil water content can further affect crop carbon input (Wang et al. 2017a, b) and characteristics of microorganisms (Chang et al. 2021). In addition, studies have also shown that soil water content will affect soil enzyme activity, and ultimately affect the conversion and circulation of soil nutrients such as carbon, nitrogen and phosphorus (Han et al. 2019).

Land use change, surface subsidence, terrain slope and vegetation coverage induced by coal mining are also important factors affecting SOC pool in the subsidence area. Mining subsidence induces topographic slope in the subsidence

area. Affected by the rainfall, wind and other external factors, soil erosion loss in the areas with lower vegetation fraction occurs, which will change the carbon storage and spatial SOC distribution in mining subsidence areas. Different terrain factors will control the surface water, heat redistribution and vegetation zonality, thereby affecting the accumulation of SOC (Chang et al. 2021; Li et al. 2013; Huang et al. 2018; Meng et al. 2017; Zou et al. 2019). Under different land use patterns and vegetation types, the roots, the quantity and quality of litterfall, and the mineralization rate of SOC are different, resulting in significant differences in SOC content (Chen et al. 2019; Li et al. 2019). In addition, human disturbance, soil structure, physical and chemical properties, soil microbial communities and other differences affect the formation and change of SOC (Huang et al. 2018; Du et al. 2016; Wang et al. 2017a, b).

Based on the above analysis on the affecting factors of SOC pool in mining subsidence area, And for the sake of data acquisition convenience by GIS and RS, we elected the following affecting factors to predict the SOC spatial distribution in subsidence area: (1) indicators representing surface deformation in subsidence area: elevation, vertical curvature, horizontal curvature, topographic relief, slope of aspect, slope of slope; (2) Indicators representing runoff change in subsidence area: topographic humidity index; (3) Indicators representing land use, terrain slope and vegetation coverage in subsidence area : land use type, slope, aspect and vegetation coverage index.

These spatial factors are used as auxiliary variables in the spatial prediction of SOC (Mueller et al. 2003; Wu et al. 2009; Mishra et al. 2010; Francaviglia et al. 2012), which will help to improve prediction accuracy. Furthermore, with the increasing development of GIS and remote sensing technology, multi-source remote sensing data has showed great advantages in the spatial prediction of soil properties, which is more practical through GIS and remote sensing data (Summers et al. 2011; Sullivan et al. 2005).

2.3.2 Data acquisition

The digital elevation model (DEM) of the study area with a spatial resolution of 30 m was obtained from the Profession scientific research of public welfare in Ministry of Land and Resources. The calculation of the various environmental factors was based on previous research (Zhang et al. 2010). The Arc GIS spatial analysis tools were used to extract the terrain factors from the DEM data for the study area, including elevation, slope, aspect, vertical curvature, horizontal curvature, the relief degree of land surface, SOS, SOA and the topographic wetness index. The detailed extraction process is shown in Table 1.

Table 1 Terrain factors and the detailed data extraction process

Terrain factors	The detailed acquisition process
Elevation	The difference between the maximum elevation value and the measured elevation in the area, that is $H_{max} - H$
Slope	Formula: $\arctan(\sqrt{f_x^2 + f_y^2})$, where, f_x and f_y are the elevation change rates of north–south, east–west, respectively
Aspect	Formula: $270^\circ + \arctan\left(\frac{f_y}{f_x}\right) - 90^\circ \frac{f_x}{ f_x }$, all parameters are the same as above
Vertical curvature	Formula: $-\frac{p^2r+2pqs+q^2t}{(p^2+q^2)\sqrt{1+p^2+q^2}}$, where, $p = \frac{\delta z}{\delta x}, q = \frac{\delta z}{\delta y}, r = \frac{\delta^2 z}{\delta x^2}, t = \frac{\delta^2 z}{\delta y^2}, s = \frac{\delta^2 z}{\delta x \delta y}$, x, y, z are the distance difference in two horizontal directions and vertical direction, respectively
Horizontal curvature	Formula: $-\frac{p^2r-2pqs+q^2t}{(p^2+q^2)\sqrt{1+p^2+q^2}}$, all parameters are the same as above
The relief degree of land surface	That is, the difference between the highest point and the lowest point in an area
SOA	Slope of the aspect. It is necessary to calculate the slope variability SOA1 of positive topography and SOA2 of anti-topography, and then correct them. Formula: $\frac{(SOA1+SOA2)-(SOA1-SOA2)*\text{abs}}{2}$
SOS	Slope of the slope. It is necessary to calculate the aspect variability SOS1 of positive topography and SOS2 of anti-topography, and then correct them. Formula: $\frac{(SOS1+SOS2)-(SOS1-SOS2)*\text{abs}}{2}$
Topographic wetness index	Formula: $\text{Ln}\left(\frac{\alpha}{\tan(\beta)}\right)$, where, α is the catchment area of a unit contour line, β is the terrain slope

Landsat 8 images were obtained from the International Scientific Data Service Platform, Computer Network Information Center, Chinese Academy of Sciences and the NDVI, and land use type (LUTP) were obtained by raster calculation from ENVI 5.1 in the third and fourth bands. The date of the image is July 2015. Based on Arc GIS, a GIS database of the research area was created, which includes the sample information collected in the research area of the sampling point and the ten environmental pieces of information extracted from remote sensing data. The DEM data and calculated NDVI values are shown in Fig. 2 (Dai et al. 2014).

2.4 Spatial prediction model of regional SOC content using RBF neural network

2.4.1 Prediction model of the RBF neural network

The RBF neural network is a three-layer feedforward neural network model with a single hidden layer. The three-layer data layer includes an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer, with a number of neurons in each. Each input neuron is fully connected to all the hidden neurons, and the hidden neurons and

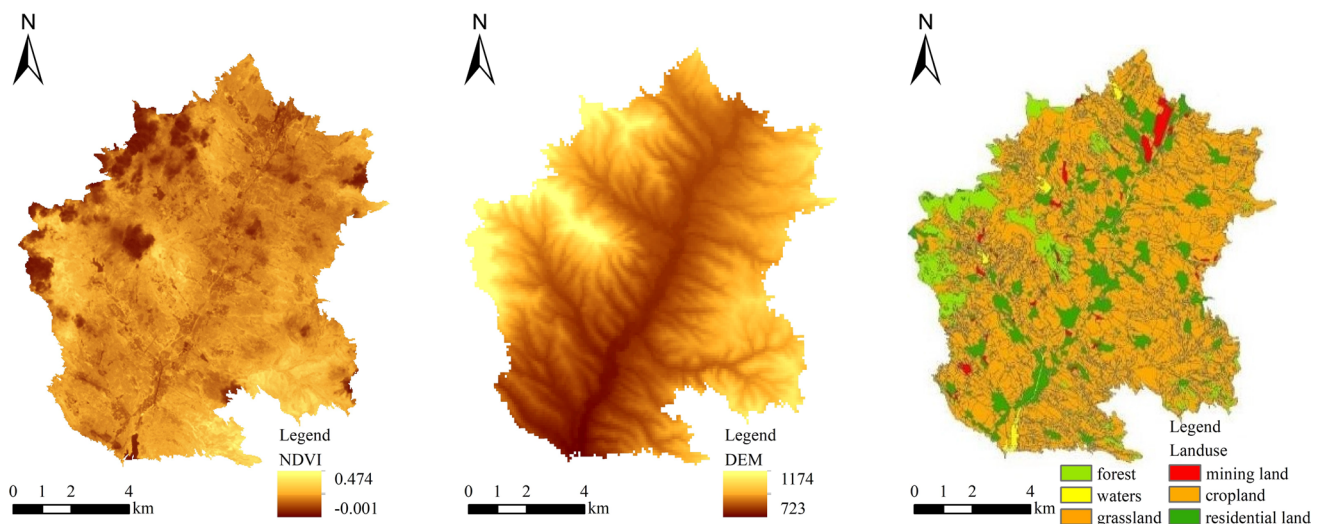


Fig. 2 The main environmental factors in the study area

output neurons are also connected to each other through a set of weights. It has obvious advantages in learning speed and parameter setting, compared with the widely used BP neural network model (Alp et al. 2005). When using the radial basis neural network to predict SOC content, the larger the spread constant, the smoother the function fitting. However, the large spread means that more neurons are needed to adapt to the rapid changes in the function, which places a lot of pressure on the calculation of the function. However, if the spread is set too small, the designed network performance will be poor (Wallisch et al. 2014). Therefore, different spread values need to be tested in the network design to determine an optimal value (Zhou et al. 2014). Similarly, the more hidden neurons, the smaller the prediction error of the model. However, increasing the hidden nodes of the neural network will increase the amount of computation. The longer it takes for neural network training and testing, the lower the learning rate of the neural network, and the lower the real-time performance of the neural network in the application. Conversely, too many hidden nodes may produce over-fitting results (Pan 2017). Therefore, before performing the radial basis neural network simulation, it is first necessary to debug the extended constant spread of the radial basis function and the maximum number of neurons, MN, of the hidden layer, and then select the spread and MN when the error is the smallest.

In this study, the RBF neural network was used as the tool to input 11 quantitative environmental factor variables as the network input, and then the SOC content at the corresponding point was used as the network output to establish an artificial neural network model that could express the quantitative relationship between the environmental factors and SOC content. The environmental factor enters the network through 11 input neurons, including elevation, slope, aspect, vertical curvature, horizontal curvature, the relief degree of land surface, SOS, SOA, topographic wetness index, LUTP and NDVI. Then the information is transmitted to the hidden neurons through $Y = [y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}]^T$. Each hidden neuron then transforms the input neuron using a transfer function \varnothing .

The functional relationship between each input neuron and the hidden neuron is (Schmitz et al. 2005):

$$h_t(y_s) = \varnothing\left(-\frac{\|Y - c_t\|}{\sigma}\right) \quad (1)$$

where h_t is the hidden neuron, Y is the output neuron, and $\varnothing()$ is the transfer function, which in this study is the gaussian radial basis function. $\| \cdot \|$ is the Euclidean norm and c_t is the center of the t neuron in the hidden layer, which is the width of the hidden neuron. This can be computed by:

$$\sigma = \frac{d_{\max}}{\sqrt{2T}} \quad (2)$$

where d_{\max} is the maximum distance between the centers of the hidden neurons and T is the number of hidden neurons.

Finally, the output layer responds to the output of the hidden layer through the mapping function, which is a linear function and a linear combination of the output results of the hidden layer through connecting weights. The formula is:

$$\hat{Z}_{\text{ANN}} = \sum_{i=1}^m w_i \varnothing_i(Y) \quad (3)$$

where \hat{Z}_{ANN} is the estimated value of SOC content, w_i is the connecting weight between the hidden neuron and the output neuron, and $\varnothing_i(Y)$ is the response of the i th hidden neuron resulting from all input data.

In MATLAB, the new function is called for the operation of the radial basis function. The call format is:

$$\text{net} = \text{newrb}(\text{P}, \text{T}, \text{goal}, \text{spread}, \text{MN}, \text{df}) \quad (4)$$

where net represents the neural network model that needs to be established; P represents the input matrix, which is the matrix Y that contains all the environmental information; T is the output matrix, and the SOC content is predicted using this function. Goal is a scalar, representing the specified mean square error; Spread refers to the expansion speed of the radial basis function; MN represents the maximum number of hidden neurons; and df represents the number of neurons added between two displays.

2.5 Estimate of residuals by ordinary kriging

The measured value of SOC content was divided into two parts: the sum of the predicted value by the radial basis function and the residual value. The formula is defined as:

$$Z(x_i) = \hat{z}_{\text{ANN}}(x_i) + r(x_i) \quad (5)$$

where $Z(x_i)$ represents the measured value of SOC content at point x_i , $\hat{z}_{\text{ANN}}(x_i)$ represents the predicted value of SOC content at point x_i by an artificial neural network, and $r(x_i)$ represents the residual value.

Using the above formula, the residual value at each sample point was obtained and the residual value was spatially predicted by the ordinary kriging method to calculate the residual value of the whole region. Finally, the predicted values of SOC content and the spatial predicted values of the residual were raster-added in Arc GIS 10.0 to obtain the predicted values of SOC content for the entire region. The formula is:

$$\hat{z}(x_i) = \hat{z}_{ANN}(x_i) + \hat{r}_{ok}(x_i) \tag{6}$$

2.6 Evaluation of the accuracy of the interpolation methods

Based on previous research (Dai et al. 2014; Richard et al. 1991), the three errors of ME, MAE and RMSE were selected for accuracy analysis. The formulas for the three indicators are:

$$ME = \frac{1}{n} \sum_{i=1}^n [\hat{z}(x_i) - z(x_i)] \tag{7}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{z}(x_i) - z(x_i)| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [\hat{z}(x_i) - z(x_i)]^2} \tag{9}$$

where $\hat{z}(x_i)$ represents the predicted value at point x_i , $z(x_i)$ represents the measured value of SOC content at point x_i , and n represents the number of validation sites. The smaller the values of MAE, ME, and RMSE, the smaller the simulation error of the model, and the higher the accuracy.

3 Results and discussion

3.1 Descriptive statistics of the SOC content

The descriptive statistics of the SOC content are shown in Table 2. The SOC content within the 0–20 cm soil layer in the study area ranges from 0.64 to 23.30 g/kg, with an average value of 10.64 g/kg and a coefficient of variation of 0.39, indicating moderate variation. The skewness is 0.13 and kurtosis is 0.14. While the SOC content within the 20–40 cm soil layer in the study area ranges from 0.25 to 19.97 g/kg, with an average value of 9.34 g/kg and a coefficient of variation of 0.43, indicating moderate variation. The skewness is 0.23 and kurtosis is 0.23. Indicating that the data conform to the normal distribution and belong to the positive skewness distribution. Normal distribution is the premise for the kriging interpolation of data (Liu et al. 2015). Therefore, Table 2 further proves that the kriging interpolation of the SOC content and the residual in this study is reasonable and effective.

3.2 Geostatistical analysis on the spatial variability of SOC in the mining subsidence area

According to kriging interpolation theory, C_0 is the nugget variance and the mean random error is the variation jointly caused by experimental error, fertilization, crop variation, management level, and other random factors on a small sampling scale (Sreenivas et al. 2016; Chiles et al. 2009). The large nugget variance indicates that processes on a small scale cannot be ignored. In Changhe River Basin, the nugget values (C_0) for the SOC content were small (Table 3), which indicates that the spatial variations in the SOC caused by experimental error, fertilization, crop variation, management level, and other random factors on a small sampling scale were minimal at a regional scale.

C represents the structural variance and the mean system attribute or maximum spatial variation of a regional variable,

Table 2 Descriptive statistics of the SOC content

Soil properties	Max (g/kg)	Min (g/kg)	Mean(g/kg)	Standard deviation(g/kg)	Variation coefficient	Skewness	Kurtosis
SOC(0–20 cm)	23.30	0.64	10.64	4.22	0.39	0.13	0.14
SOC(20–40 cm)	19.97	0.25	9.34	4.07	0.43	0.23	0.23

Table 3 Semi-variance analysis of the SOC content and residual

Parameter	Soil layer thickness (cm)	Model	Nugget C_0	Sill $C_0 + C$	$C / (C_0 + C)$	Range	Coefficient of determination R^2
SOC	0–20	Exponential	1.60	18.00	0.91	1920	0.74
	20–40	Exponential	8.41	23.40	0.64	5880	0.80
Residual	0–20	Exponential	9.66	19.33	0.50	4680	0.71
	20–40	Exponential	1.67	16.74	0.90	6324	0.54

where this variation is caused by the soil parent material, terrain, climate, and other structural factors (Sreenivas et al. 2016; Chiles et al. 2009). The climate in the subsidence area remained unchanged before and after coal mining, so the spatial variability in the SOC content was basically caused by mining subsidence and other structural factors due to coal mining. $C + C_0$ is the sill variance (the stationary value of the semivariance function after the interval increases progressively to a certain degree) and it represents the total variation in the system. $C/(C_0 + C)$ represents the degree of spatial correlation (the proportion of spatial variation caused by structural factors in the total system variation). The spatial correlation is poor when the specific value is less than 0.25, moderate when the specific value is between 0.25 and 0.75, and good when the specific value is greater than 0.75.

In the study area, the $C/(C_0 + C)$ values of the SOC content (0–20 cm) and 20–40 cm are 0.91 and 0.64, respectively, where $C/(C_0 + C)$ values of the SOC content (0–20 cm) were greater than 0.75 (Table 3), which indicates that the spatial correlation in the SOC (0–20 cm) is mainly caused by structural factors such as mining, surface subsidence and other structural factors due to coal mining at a regional scale. This is mainly due to the subsidence of the surface, the destruction of the original topography and surface vegetation, and soil erosion caused by large-scale coal mining. Changes in the physical, chemical and biological properties of soil in mining areas will result in the destruction of soil aggregates, nutrient loss, reduced microbial activity, and decreased SOC content.

3.3 Spatial distribution of SOC content

The spatial distribution of the SOC content obtained by the two methods is shown in Fig. 3. The spatial distribution of the SOC content obtained by the two methods is in general consistent. On the whole, the SOC content is relatively low in certain areas west of the Changhe River and the highest content is concentrated southeast of the river. In the study region, the mining area is mainly located in the western part of the river basin. Large-scale coal mining activities have caused varying degrees of ground subsidence, soil erosion, and vegetation damage. The land in the mining area has been severely damaged, soil fertility has declined and organic carbon has been destroyed. Therefore, the average SOC content in the western of the study area is lower than the eastern, which is more obvious within the 20–40 cm soil layer. This may be due to the fact that SOC is mainly derived from the biomass that enters the soil, and the activities in the mining area destroy the soil structure, reduce the soil quality, and reduce the productivity, so less biomass was input to soil than in the unmined area. Secondly, the biomass of input soil is mainly concentrated in the topsoil and decreased with the increase of soil depth, resulting the SOC

content of 20–40 cm in the western region is lower than in the eastern region.

The SOC content in surface soil is relatively high along the long river in the middle of the region (Fig. 3a), which is due to better soil moisture conditions along the river. Previous research has shown a positive correlation between soil moisture content and SOC content, the high influence of soil permeability, and soil moisture content in organic carbon mineralization. Therefore, exogenous organic residues in the water under the action of rot have easily degraded into small molecular organic substances and have been preserved in the soil, which helps improve the SOC content. Compared with the soil along the river banks, the surrounding soil has low water content, good soil permeability, high porosity, and easy mineralization and decomposition of organic carbon, which is not conducive to the accumulation of SOC. Therefore, the SOC content is relatively low.

Based on the prediction results, the SOC content ranges in the study area predicted by the RBF neural network are 0.58–23.75 g/kg (within the 0–20 cm soil layer), and 0.55–20.37 g/kg (within the 20–40 cm soil layer), respectively. The SOC content ranges in the study area obtained by the direct kriging method are 1.34–22.13 g/kg (within the 0–20 cm soil layer), and 0.65–19.65 g/kg (within the 20–40 cm soil layer), respectively. It can be concluded that the SOC content within the 0–20 cm soil layer is slightly higher than that within the 20–40 cm soil layer. This is because the surface soil, with its rich hydrothermal resources and animal and plant remains, promotes the decomposition of microorganisms, which is more conducive to the accumulation of organic carbon. Combined with the prediction results, the predicted results of the RBF neural network showed more information than the direct kriging interpolation method, and the changes in the local areas are more obvious. This is because the RBF neural network comprehensively considers different geographical factors, especially the changes in topographic factors caused by coal mining disturbances and the spatial correlation of variables on SOC. In general, the strong spatial dependence of SOC is determined by the intrinsic changes in SOC, while the external variation table controls the variability of the less spatially dependent parameters (Cambardella et al. 1994). Furthermore, because of the influence of coal mining activities in the study area, the terrain change is more severe and had a greater influence on the spatial distribution of SOC content. Therefore, the RBF neural network comprehensively considers the effects of various environmental factors on SOC, taking into account the spatial structure of SOC. At the same time, the calculation of residuals by the ordinary kriging method takes into account the spatial variation of sample point randomness. Compared with the ordinary kriging method, combining kriging with the RBF neural network takes both internal and external factors into consideration to improve the accuracy of SOC spatial distribution prediction.

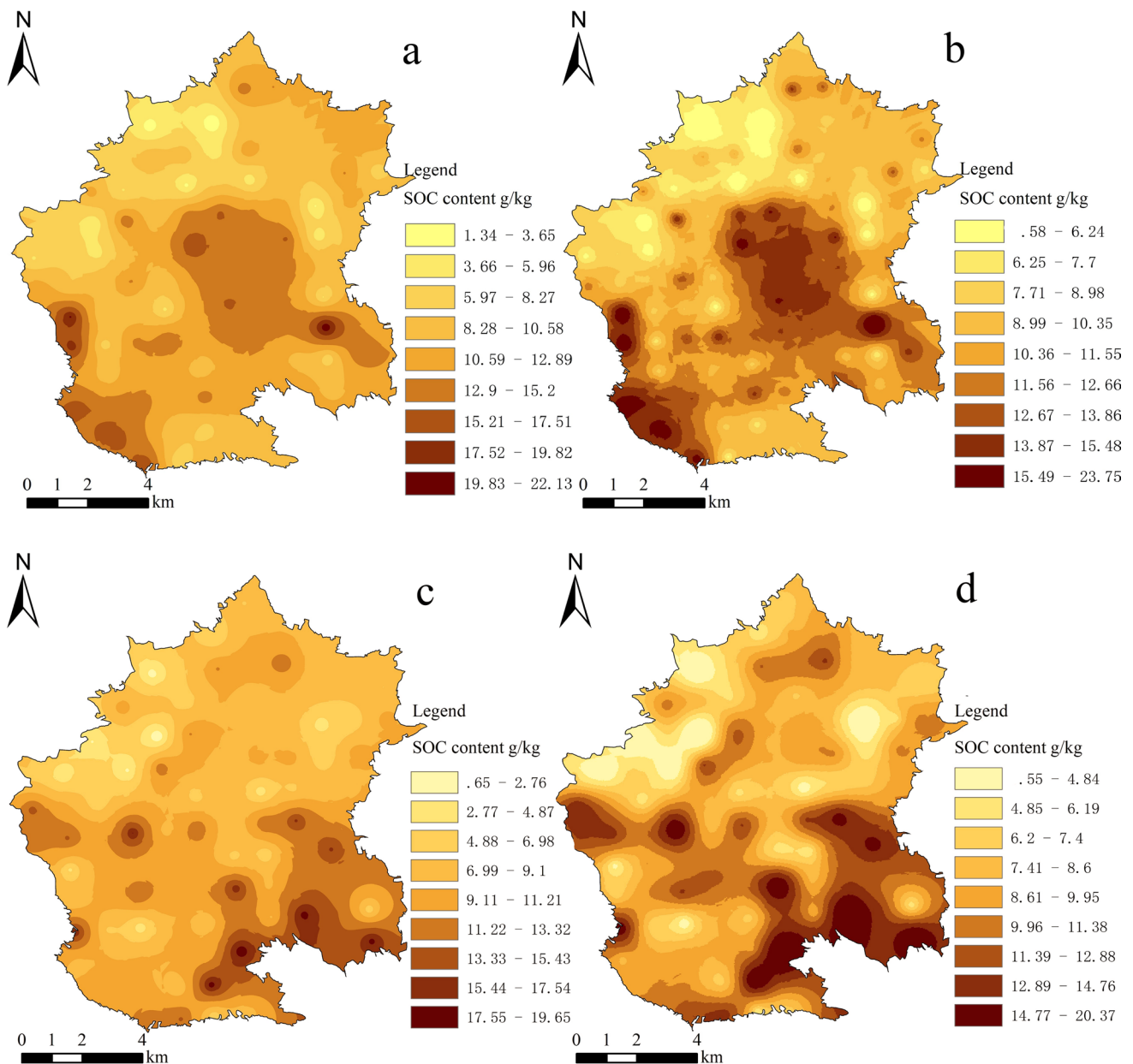


Fig. 3 The spatial distribution of SOC. **a** 0–20 cm, Direct kriging; **b** 0–20 cm, RBF Neural Network; **c** 20–40 cm, Direct kriging; **d** 20–40 cm, RBF Neural Network

Comparing the prediction results of the two methods in Fig. 3 shows that the predicted results of the RBF neural network revealed more detail than the ordinary kriging, which only considers the spatial correlation of SOC, which makes the prediction effect unsatisfactory. The range of variability in the SOC content and the residual is large, indicating that the variables are influenced by other factors within a wide range of regions (Table 3) (Hengl et al. 2004; Takata et al. 2007). Therefore, the spatial distribution of the SOC content using direct kriging is not very accurate, while the RBF neural network that combines

environmental factors is more consistent with the actual state of the study area and is more scientific.

3.4 Accuracy assessment of the prediction methods

The fit of the predicted and measured values (Fig. 4) shows that the determination coefficient R^2 obtained by the RBF neural network are 0.81 and 0.70, respectively, which is much higher than the 0.44 and 0.36 obtained by direct kriging. The determination coefficient R^2 indicates the fitting accuracy of the predicted and measured values. The closer

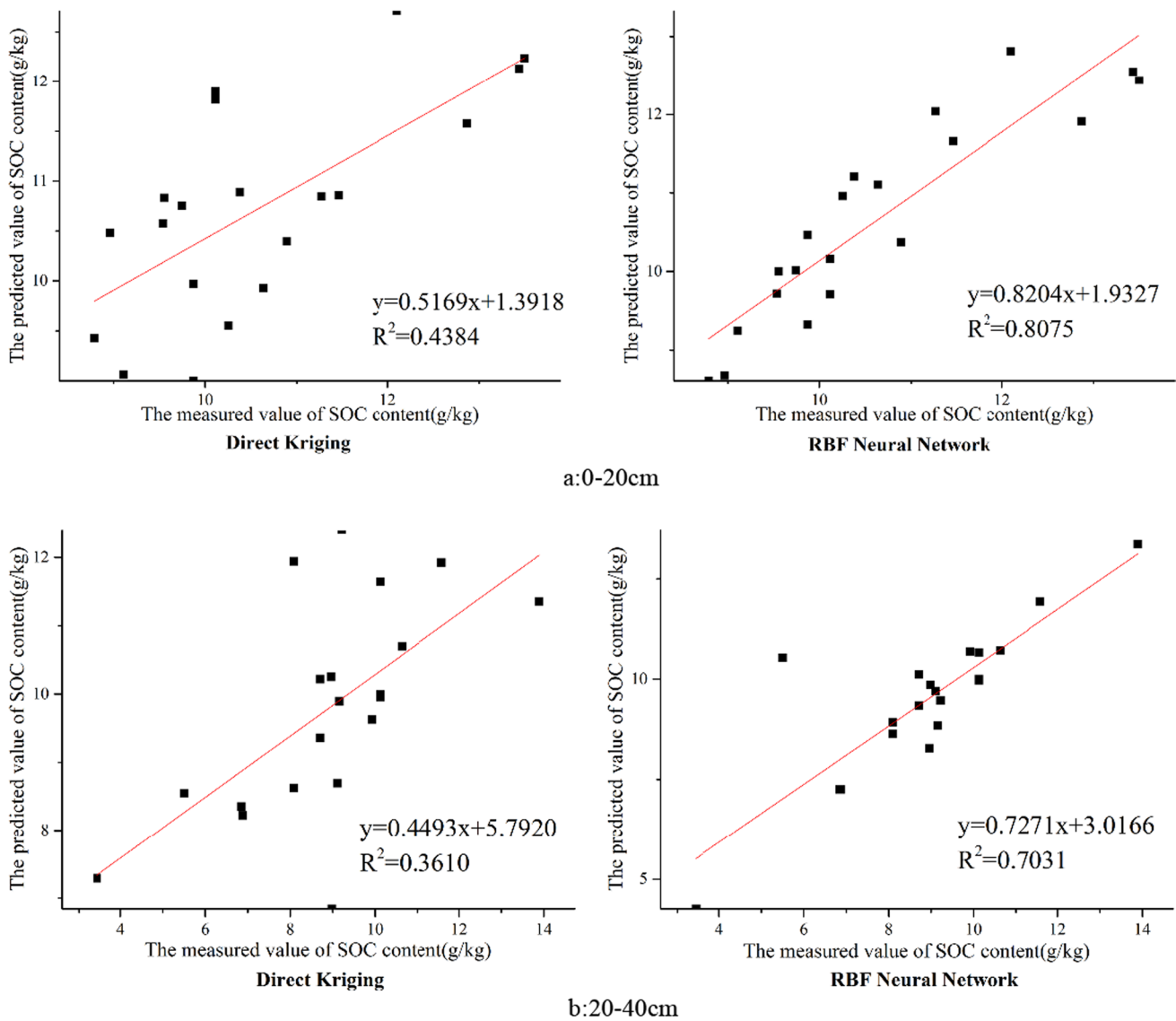


Fig. 4 The scatter plot of predicted and measured values

R^2 is to 1, the higher the fitting accuracy, indicating the better prediction effect of the model (Nakagawa et al. 2013). In conclusion, compared with the ordinary kriging method, the spatial prediction accuracy of the RBF neural network combined with the kriging method for the SOC content in the mining area is higher, and this result has been confirmed by previous studies.

The prediction accuracy indicators of the two methods are shown in Table 4. In terms of prediction accuracy, the ME, MAE and RMSE obtained by direct kriging are 0.12, 0.89, 1.02 (within the 0–20 cm soil layer), and 0.89, 1.45, 1.89 (within the 20–40 cm soil layer), respectively, which are all higher than the 0.03, 0.51, 0.59 (within the 0–20 cm soil layer), and 0.58, 0.76, 1.27 (within the 20–40 cm soil layer) obtained by the RBF neural network (Table 4). Among them,

Table 4 Prediction accuracy indicators for the two methods

Method		ME	MAE	RMSE
Direct Kriging	0–20 cm	0.12	0.89	1.02
	20–40 cm	0.89	1.45	1.89
RBF neural network	0–20 cm	0.03	0.51	0.59
	20–40 cm	0.58	0.76	1.27

Notes: ME Mean error; MAE Mean absolute error; RMSE Root mean square error

ME represents the average deviation of the prediction, indicating that the average level of the RBF neural network prediction is higher. MAE represents the actual prediction error, indicating that the prediction of the RBF neural network is

more consistent with the actual SOC spatial distribution. Furthermore, the RMSE values from both methods are larger than MAE, indicating that the error has strong spatial variability (Dai et al. 2014). The prediction accuracy of the RBF neural network is higher for the spatial distribution of SOC in the study area (Table 4).

4 Conclusions

In this paper, the SOC content of a mining area was analyzed. We also propose a new method to predict the spatial distribution of SOC content in the mining area using the RBF neural network method combined with the ordinary kriging method. First, the RBF neural network is used to construct the nonlinear mapping relationship between the environmental variables of the mining area and the SOC content to calculate the predicted value of SOC content.

Then, the residuals are calculated and spatialized with the ordinary kriging method. Finally, the spatial residuals are added to the results of the RBF neural network, and the predicted value of SOC space in the study area is obtained.

Compared with the ordinary kriging method, this method is scientific and feasible. The conclusions of this study are as follows:

- (1) Coal mining activities have caused great disturbance to the soil in the mining area that has reduced the SOC content in the area, resulting in the loss of soil nutrients. As coal mining will cause land subsidence, soil erosion and vegetation destruction, the main influencing factors of soil property distribution in the mining area are topography and vegetation factors, such as slope, elevation, topographic wetness index and the normalized difference vegetation index. The most important variable for prediction of SOC is slope. Therefore, when estimating the spatial distribution of SOC in a mining area, these topographical and vegetation factors should be taken into account to improve estimation accuracy.
- (2) Compared with direct kriging interpolation, the RBF neural network combined with kriging has a smaller average error, mean absolute error and root mean square error than ordinary kriging in terms of prediction accuracy. Furthermore, the accuracy R^2 of the predicted point and the measured point is higher than that of the ordinary kriging. Therefore, the radial basis neural network combined with environmental factors is more suitable for predicting SOC content in mining areas that have been severely disturbed by human activities. As a result, it can provide a reference for spatial prediction of soil properties in a mining area,

and thus provide a scientific and reasonable basis for land reclamation and land resource management.

Acknowledgements This work was supported by the National Natural Science Foundation of China (51304130), the Natural Science Foundation of Shanxi Province, China (2015021125), Shanxi Provincial People's Government Major Decision Consulting Project (ZB20211703), Program for the Soft Science research of Shanxi (2018041060-2), Program for the Philosophy and Social Sciences Research of Higher Learning Institutions of Shanxi (201803010), Philosophy and Social Sciences Planning Project of Shanxi Province (2020YJ052), Basic Research Program of Shanxi Province (20210302123403).

Author contributions QQ: Data processing, Drafted the manuscript & Revision. ZX: Methodology, Review & Editing. XY: Data processing. XD: Data processing. ZL: Data processing. All authors approved the final manuscript for publication.

Declarations

Ethics approval and consent to participate All the authors of this manuscript have approved the article's submission for publication. This paper has not been published elsewhere and is not under consideration by another journal. There has been no improper falsification or tampering of data. After accepting the manuscript, we promise not to change the author's identity and author sequence by adding or removing the author.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akpa SIC, Odeh IOA, Bishop TFA et al (2016) Total soil organic carbon and carbon sequestration potential in Nigeria. *Geoderma* 271:202–215. <https://doi.org/10.1016/j.geoderma.2016.02.021>
- Alp M, Cigizoglu HK (2005) Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. *Environ Model Softw* 22(1):2–13. <https://doi.org/10.1016/j.envsoft.2005.09.009>
- Brejda JJ, Mausbach MJ, Goebel JJ et al (2001) Estimating surface soil organic carbon content at a regional scale using the national resource inventory. *Soil Sci Soc Am J* 65:842–849. <https://doi.org/10.2136/sssaj2001.653842x>
- Cambardella CA, Moorman TB, Parkin TB et al (1994) Field-scale variability of soil properties in central Iowa soils. *Soil Sci Soc Am J*. <https://doi.org/10.2136/sssaj1994.03615995005800050033x>
- Chang S, Yu HB, Cao CM, Ma XC, Liu YX, Li X (2021) Distribution characteristics of soil organic carbon in Xilin Gol steppe and its influencing factors. *Arid Zone Res* 38(5):1355–1366. <https://doi.org/10.13866/j.azr.2021.05.17>

- Chen XT, Xu TL, Li XJ, Zhao AH, Feng HY, Chen BD (2019) Soil organic carbon concentrations and the influencing factors in natural ecosystems of northern China. *Chin J Ecol* 38(4):1133–1140. <https://doi.org/10.13292/j.1000-4890.201904.004>
- Cheng JX, Nie XJ, Liu CH (2014) Spatial variation of soil organic carbon in coal-mining subsidence areas. *J China Coal Soc* 39:2495–2500. <https://doi.org/10.13225/j.cnki.jccs.2013.1806>
- Chiles JP, Delfiner P (2009) *Geostatistics: modeling spatial uncertainty*, vol 497. John Wiley & Sons
- Coxson DS, Parkinson D (1987) Winter respiratory activity in aspen woodland forest floor litter and soils. *Soil Biol Biochem* 19:49–59. [https://doi.org/10.1016/0038-0717\(87\)90125-8](https://doi.org/10.1016/0038-0717(87)90125-8)
- Dai F, Zhou Q, Lv Z et al (2014) Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. *Ecol Ind* 45:184–194. <https://doi.org/10.1016/j.ecolind.2014.04.003>
- Du H, Zeng FP, Sun TQ, Wen YG, Li CG, Peng WX, Zhang H, Zeng ZX (2016) Spatial pattern of soil organic carbon of the main forest soils and its influencing factors in Guangxi, China. *Chin J Plant Ecol* 40(4):282–291. <https://doi.org/10.17521/cjpe.2015.0199>
- Francaviglia R, Coleman K, Whitmore AP, Doro L, Urracci G, Rubino M, Ledda L (2012) Changes in soil organic carbon and climate change—application of the RothC model in agro-silvo-pastoral Mediterranean systems. *Agric Syst* 112:48–54. <https://doi.org/10.1016/j.agsy.2012.07.001>
- Fu YH (2017) Soil organic carbon properties of subsidence wetland in the coal mining areas with high ground water table. Dissertation, China University of Mining & Technology, Beijing
- Guo L, Zhang H, Shi T et al (2019) Prediction of soil organic carbon stock by laboratory spectral data and airborne hyperspectral images. *Geoderma* 337:32–41. <https://doi.org/10.1016/j.geoderma.2018.09.003>
- Han Y, Wang Q, Zhao W, Shi NN, Xiao NW, Zhang ZA, Quan ZJ (2019) Effects of opencast coal mining on soil properties and plant communities of grassland. *Chin J Ecol* 38(11):3425–3433. <https://doi.org/10.13292/j.1000-4890.201911.011>
- Hao Y, Lal R, Owens LB et al (2002) Effect of cropland management and slope position on soil organic carbon pool at the North Appalachian Experimental watersheds. *Soil Tillage Res* 68:133–142. [https://doi.org/10.1016/S0167-1987\(02\)00113-7](https://doi.org/10.1016/S0167-1987(02)00113-7)
- Hengl T, Heuvelink GBM, Stein A (2004) A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 120(1–2):75–93. <https://doi.org/10.1016/j.geoderma.2003.08.018>
- Hu JC (2021) Changes and forecasts of water resources and water environment under the condition group mining. *Ground Water* 43:28–32. <https://doi.org/10.19807/j.cnki.DXS.2021-01-009>
- Huang Y, Wang YJ, Tian F, Hou F (2014) Carbon disturbance effects in the vegetation-soil system caused by coal. *Resour Sci* 36:817–823
- Huang XF, Zhou YC, Zhang ZM (2018) Characteristics and affecting factors of soil organic carbon under land uses: a case study in Houzhai river basin. *J Nat Resour* 33(6):1056–1067
- IEA (2021) *Key world energy statistics 2021*. OECD Publishing, Paris. <https://doi.org/10.1787/2ef8cebc-en>
- John K, Abraham Isong I, Michael Kebonye N et al (2020) Using machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil. *Land* 9:487. <https://doi.org/10.3390/land9120487>
- Jun T, Ting-hui D, Jun-chao LI, Jiang XUE (2015) Soil organic carbon profile distribution of rehabilitated grasslands on the opencast coal mine dump of Loess area. *Acta Agres Sin* 23:718. <https://doi.org/10.11733/j.issn.1007-0435.2015.04.008>
- Lai YQ, Sun XL, Wang HL (2020) Mapping of soil organic carbon using neural network and its mixed model with geostatistics in a small area of typical Hilly Region. *Chin J Soil Sci* 51:1313–1322. <https://doi.org/10.19336/j.cnki.trtb.2020.06.08>
- Lemma B, Kleja DB, Nilsson I, Olsson M (2006) Soil carbon sequestration under different exotic tree species in the southwestern highlands of Ethiopia. *Geoderma* 136:886–898. <https://doi.org/10.1016/j.geoderma.2006.06.008>
- Li L, Wu LZ, Yao YF, Qin FC, Guo YF, Jiao SX (2013) Study on spatial variations of soil organic carbon in small watershed-taking huanghuadianzi watershed as an example. *Res Soil Water Conserv* 20(5):18–23
- Li HX, Lei SG, Shen YQ (2016) Impacts of ground subsidence and fissures caused by coal mining on vegetation coverage. *J Ecol Rural Environ* 32:195–199
- Li L, Qin FC, Jiang LN, Yao XL, Wang XJ (2019) Effects of land use type and terrain on soil organic carbon (SOC) content in semi-arid region. *Soils* 51(2):406–412. <https://doi.org/10.13758/j.cnki.tr.2019.02.027>
- Liu YL, Guo L, Jiang QH et al (2015) Comparing geospatial techniques to predict SOC stocks. *Soil Tillage Res* 148:46–58. <https://doi.org/10.1016/j.still.2014.12.002>
- Liu HY, Wu WQ, WANG ZH (2021) Surface settlement monitoring and analysis of coal dam mining area based on SBAS-InSAR Technology. *Geomat Spat Inf Technol* 44:69–72
- Luo YL (2016) Spatial variability change modeling of soil organic carbon at a county scale in hilly area of Mig-Sichuan. Dissertation, Sichuan Agricultural University
- Meng GX, Zha TG, Zhang XX, Zhang ZQ, Zhu LS, ZhouY, Liu YH, Lin Z (2017) Effects of vegetation type and terrain on vertical distribution of soil organic carbon on abandoned farmlands in the Loess Plateau. *Chin J Ecol* 36(9):2447–2454. <https://doi.org/10.13292/j.1000-4890.201709.018>
- Miller BA, Koszinski S, Wehrhan M, Sommer M (2015) Comparison of spatial association approaches for landscape mapping of soil organic carbon stocks. *Soil* 1:217–233. <https://doi.org/10.5194/soil-1-217-2015>
- Mishra U, Lal R, Liu DS, Meirvenne MV (2010) Predicting the spatial variation of the soil organic carbon pool at a regional scale. *Soil Sci Soc Am J* 74(3):906–914. <https://doi.org/10.2136/sssaj2009.0158>
- Mo A, Zhou YZ, Yang JJ, Liu W, Shi QD, Lu HY (2015) Influence of mountain coal mining on physical and chemical properties of soil. *J Soil Water Conserv* 29(1):86–89. <https://doi.org/10.13870/j.cnki.stbxb.2015.01.018>
- Morais TG, Tufik C, Rato AE et al (2021) Estimating soil organic carbon of sown biodiverse permanent pastures in Portugal using near infrared spectral data and artificial neural networks. *Geoderma* 404:115387. <https://doi.org/10.1016/j.geoderma.2021.115387>
- Mueller TG, Pierce FJ (2003) Soil carbon maps: enhancing spatial estimates with simple terrain attributes at multiple scales. *Soil Sci Soc Am J* 67(1):258–267. <https://doi.org/10.2136/sssaj2003.0258>
- Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods Ecol Evol* 4(2):133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Pan Q (2017) Research on establishing an improved method to determine the radial basis function center of radial basis function neural network. Dissertation, Northeast Agricultural University
- Pudełko A, Chodak M, Roemer J, Uhl T (2020) Application of FT-NIR spectroscopy and NIR hyperspectral imaging to predict nitrogen and organic carbon contents in mine soils. *Measurement* 164:108117. <https://doi.org/10.1016/j.measurement.2020.108117>
- Qie CL, Bian ZF, Yang DJ, Lei SG, Mou SG (2015) Effect of high-intensity underground coal mining disturbance on soil physical properties. *J China Coal Soc* 40(6):1448–1456. <https://doi.org/10.13225/j.cnki.jccs.2014.1580>

- Redondo-Vega JM, Gómez-Villar A, Santos-González J et al (2017) Changes in land use due to mining in the north-western mountains of Spain during the previous 50 years. *Catena* 149:844–856. <https://doi.org/10.1016/j.catena.2016.03.017>
- Ren GQ, Jia XX, Jia YH, Guo CJ (2018) Spatial variation of soil organic carbon content and its driving factors along south-north transect in the loess plateau of China. *Arid Zone Res* 35(3):524–531. <https://doi.org/10.13866/j.azr.2018.03.04>
- Ren BW, Chen HY, Zhang LM, Nie XQ, Xing SH, Fan XY (2021) Comparison of machine learning for predicting and mapping soil organic carbon in cultivated land in a subtropical complex geomorphic region. *Chin J Eco-Agriculture* 29:1042–1050. <https://doi.org/10.13930/j.cnki.cjea.200939>
- Richard A, Bilonick, (1991) An introduction to applied geostatistics. *Technometrics* 33(4):483–485. <https://doi.org/10.1080/00401706.1991.10484886>
- Schmitz GH, Puhlmann H, Dröge W et al (2005) Artificial neural networks for estimating soil hydraulic parameters from dynamic flow experiments. *Eur J Soil Sci* 56(1):19–30. <https://doi.org/10.1111/j.1351-0754.2004.00631.x>
- Shrestha RK, Lal R (2011) Changes in physical and chemical properties of soil after surface mining and reclamation. *Geoderma* 161:168–176. <https://doi.org/10.1016/j.geoderma.2010.12.015>
- Soleimani A, Hosseini SM, Massah Bavani AR et al (2017) Simulating soil organic carbon stock as affected by land cover change and climate change, Hyrcanian forests (northern Iran). *Sci Total Environ* 599–600:1646–1657. <https://doi.org/10.1016/j.scitotenv.2017.05.077>
- Song J, Yang Z, Xia J, Cheng D (2021) The impact of mining-related human activities on runoff in northern Shaanxi, China. *J Hydrol* 598:126235. <https://doi.org/10.1016/j.jhydrol.2021.126235>
- Sreenivas K, Dadhwal VK, Kumar S et al (2016) Digital mapping of soil organic and inorganic carbon status in India. *Geoderma* 269:160–173. <https://doi.org/10.1016/j.geoderma.2016.02.002>
- Su H (2021) Erosion evolution process and hydrodynamic mechanism of the platform-slope system in opencast coal mine dump, the north of Shaanxi province. Dissertation, Northwest A&F University
- Sullivan DG, Shaw JN, Rickman D (2005) IKONOS imagery to estimate surface soil property variability in two Alabama physiographies. *Soil Sci Soc Am J*. <https://doi.org/10.2136/sssaj2005.0071>
- Summers D, Lewis M, Ostendorf B et al (2011) Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. *Ecol Ind* 11(1):123–131. <https://doi.org/10.1016/j.ecolind.2009.05.001>
- Sun W, Li X, Niu B (2018) Prediction of soil organic carbon in a coal mining area by Vis-NIR spectroscopy. *PLoS One* 13:e0196198. <https://doi.org/10.1371/journal.pone.0196198>
- Takata Y, Funakawa S, Akshalov K et al (2007) Spatial prediction of soil organic matter in northern Kazakhstan based on topographic and vegetation information. *Soil Sci plant Nutr* 53(3):289–299. <https://doi.org/10.1111/j.1747-0765.2007.00142.x>
- Tian HW, Zhang XX, Bi RT, Zhu HF, Xi X (2020) An assessment of the carbon sequestration loss of farmland ecosystems caused by coal mining. *J China Coal Soc* 45:1499–1509. <https://doi.org/10.13225/j.cnki.jccs.2019.0511>
- Wallisch P, Lusignan ME, Benayoun MD et al (2014) MATLAB for neuroscientists: an introduction to scientific computing in MATLAB. Academic Press
- Wang D, Wu XD (2020) Modelling soil organic carbon distribution in the seasonally frozen ground area on the Qinghai-Tibet Plateau using the geographically weighted regression. *J Glaciol Geocryol* 42:1036–1045
- Wang LJ, Li R, Sheng MY (2017) Distribution of soil organic carbon related to environmental factors in typical rocky desertification ecosystems. *Acta Ecol Sinica* 37(5):1367–1378
- Wang LJ, Sheng MY, Du JY, Wen PC (2017) Distribution characteristics of soil organic carbon and its influence factors in the karst rocky desertification ecosystem of Southwest China. *Acta Ecol Sin* 37(4):1358–1365. <https://doi.org/10.5846/stxb201607051377>
- Wang X-T, Zhang J-Q, Yang M-Y, Wang Y-J (2020) Effects of “Grain for Green” program and coal mining on sediment production in a typical small watershed of Yushenfu Mining Region, Northwest China. *Ying Yong Sheng Tai Xue Bao* 31:1971–1979. <https://doi.org/10.13287/j.1001-9332.202006.016>
- Were K, Bui DT, Dick ØB, Singh BR (2015) A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape. *Ecol Ind* 52:394–403. <https://doi.org/10.1016/j.ecolind.2014.12.028>
- Wu CF, Wu JP, Luo YM, Zhang LM et al (2009) Spatial prediction of soil organic matter content using cokriging with remotely sensed data. *Soil Sci Soc Am J* 73(4):1202–1208. <https://doi.org/10.2136/sssaj2008.0045>
- Wu ZY, Pen SP, Du WF, Cui F (2019) Effect of coal mining on surface soil physicochemical properties of sandy land in the arid region. *Res Soil Water Conserv* 26(5):75–80. <https://doi.org/10.13869/j.cnki.rswc.2019.05.011>
- Xu ZJ (2012) Carbon effects on the soil carbon and vegetation carbon pools by mining disturbance in mining area with higher groundwater level. Dissertation, China University of Mining and Technology
- Xu Z, Zhang Y, Yang J et al (2019) Effect of underground coal mining on the regional soil organic carbon pool in farmland in a mining subsidence area. *Sustainability* 11:4961. <https://doi.org/10.3390/su11184961>
- Yuan YQ, Chen HY, Zhang LM et al (2021) Prediction of spatial distribution of soil organic carbon in farmland based on multi-variables and random forest algorithm—a case study of a subtropical complex geomorphic region in Fujian as an example. *Acta Pedol Sin* 58:887–899
- Zhang GS, Ni ZW (2017) Winter tillage impacts on soil organic carbon, aggregation and CO₂ emission in a rainfed vegetable cropping system of the mid-Yunnan plateau, China. *Soil Tillage Res* 165:294–301. <https://doi.org/10.1016/j.still.2016.09.008>
- Zhang SM, Wang ZM, Zhang B et al (2010) Prediction of spatial distribution of soil nutrients using terrain attributes and remote sensing data. *Trans Chin Soc Agric Eng* 26(5):188–194. <https://doi.org/10.3969/j.issn.1002-6819.2010.05.033>
- Zhang W, Wang KL, Chen HS, Zhang JG (2012) Use of satellite information and gis to predict distribution of soil organic carbon in depressions amid clusters of karst peaks. *Acta Pedol Sin* 49:601–606
- Zhang H, Wu P, Yin A et al (2017) Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: a comparison of multiple linear regressions and the random forest model. *Sci Total Environ* 592:704–713. <https://doi.org/10.1016/j.scitotenv.2017.02.146>
- Zhou WH (2014) Optimization study of the hidden structure and parameters in the RBF neural networks. Dissertation, East China University of Science and Technology
- Zou RY, Zhou HY, Guo X, Dan CL, Lv TG, Li HY (2019) Spatial distribution and mapping of influencing factors of farmland soil organic carbon in the poyang lake region. *Resour Environ Yangtze Basin* 28(5):1121–1131

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.