Article

Prediction of Soil Nutrients Based on Topographic Factors and Remote Sensing Index in a Coal Mining Area, China

Hongfen Zhu ^{1,2} , Ruipeng Sun ¹, Zhanjun Xu ^{1,2}, Chunjuan Lv ^{1,2} and Rutian Bi ^{1,2,*}

¹ College of Resource and Environment, Shanxi Agricultural University, Shanxi 030801, China; hongfzh@163.com (H.Z.); sunruipeng1996@163.com (R.S.); zjxu163@126.com (Z.X.); lcjcw@126.com (C.L.)

² National Experimental Teaching Demonstration Center for Agricultural Resources and Environment, Shanxi Agricultural University, Shanxi 030801, China

* Correspondence: sxndbrt@163.com

Received: 15 January 2020; Accepted: 19 February 2020; Published: 21 February 2020

Abstract: (1) Background: Coal mining operations caused severe land subsidence and altered the distributions of soil nutrients that influenced by multiple environmental factors at different scales. However, the prediction performances for soil nutrients based on their scale-specific relationships with influencing factors remains undefined in the coal mining area. The objective of this study was to establish prediction models of soil nutrients based on their scale-specific relationships with influencing factors in a coal mining area. (2) Methods: Soil samples were collected based on a 1×1 km regular grid, and contents of soil organic matter, soil available nitrogen, soil available phosphorus, and soil available potassium were measured. The scale components of soil nutrients and the influencing factors collected from remote sensing and topographic factors were decomposed by two-dimensional empirical mode decomposition (2D-EMD), and the predictions for soil nutrients were established using the methods of multiple linear stepwise regression or partial least squares regression based on original samples (MLSR_{ORI} or PLSR_{ORI}), partial least squares regression based on bi-dimensional intrinsic mode function (PLSR_{BIMF}), and the combined method of 2D-EMD, PLSR, and MLSR (2D-EMD_{PM}). (3) Results: The correlation types and correlation coefficients between soil nutrients and influencing factors were scale-dependent. The variances of soil nutrients at smaller scale were stochastic and non-significantly correlated with influencing factors, while their variances at the larger scales were stable. The prediction performances in the coal mining area were better than those in the non-coal mining area, and 2D-EMD_{PM} had the most stable performance. (4) Conclusions: The scale-dependent predictions can be used for soil nutrients in the coal mining areas.

Keywords: soil nutrients; coal mining; two-dimensional empirical mode decomposition (2D-EMD); bi-dimensional intrinsic mode function; partial linear squares regression

1. Introduction

Soil nutrients, including soil organic matter (SOM), soil available nitrogen (SAN), soil available phosphorus (SAP), and soil available potassium (SAK) are essential for plant growth, and thus, their spatial distribution is important for agricultural management. Traditionally, soil nutrient measurement is based on laboratory analysis of field soil samples, and the discrete sampling points may not provide continuous distributions of soil nutrients.

A number of methods, including geostatistical models [1], inverse distance weighted [2], trend surface analysis [3], and the like, resolve the spatial distribution of soil properties based on geospatial autocorrelation [4,5]. However, these methods are limited by sampling density, and the methods either fail to consider the environmental effects on soil nutrients or consider the

environmental effects only at the sampling scale, causing their heterogeneous distribution at different scales with different intensities [6,7]. The qualitative and quantitative information of vegetation growth or land surface temperatures can be obtained from the electromagnetic radiation of remote sensing (RS), which indirectly demonstrates the distribution of soil nutrients [8–11]. Schillaci et. al. indicated that the application of an RS index could enhance the predicting performance of soil properties [12]. Meanwhile, topographic factors affect the transposition and redistribution of soil nutrient contents [13], and their effects on soil nutrients vary at different scales [14,15]. A number of prediction methods for soil properties based on their relationships with topographic factors are proposed [16], including the ensemble learning model [17], extended models [18], regression model [19], and so on. Nevertheless, these methods do not consider their multiscale relationships between soil nutrients and the covariates. Additionally, scale-specific correlations between soil properties and environmental factors are explored using the geostatistical approach [20], multivariate empirical mode decomposition [21,22], wavelet transform [14], and so on. However, these methods either failed to realize the prediction of soil properties or did not establish the validation model.

For the coal mining area, effects of human activities on the spatial distribution of soil nutrients should not be ignored, and their effects on the relationships between soil nutrients and environmental factors should also be considered [23]. Thus, the traditional methods are not suitable for describing the spatial distribution characteristics of soil nutrients in the coal mining areas. Therefore, the objective of this study was to establish the validation model of soil nutrients based on their multiscale relationships with RS index and topographic factors, and to assess the prediction accuracy of soil nutrients based on different methods in the coal mining and non-coal mining areas.

2. Materials and Methods

2.1. Study Area and Soil Sampling

The study area is the Changhe watershed located in the southeast of the Chinese Loess Plateau. It is bounded by latitude 35°30′14″–35°38′04″ N and longitude 112°27′39″–112°46′13″ E with a total area of about 116 km² (Figure 1). The dominant soil type in the study area is calcaric cambisols [24], and the major crops in the arable land are maize, potato, and wheat. The Changhe river runs through the area from northeast to southwest and divides the study area into the eastern and western parts. In the western part, there are many coal mines containing hard coal. Most lands here were seriously damaged by coal mining activities, and some of them have been reclaimed. The range of elevation is from 718 to 1163 m in the coal mining area, while it is from 711 to 1040 m in the non-coal mining area, and the average elevations are 851 and 857 m for coal mining and non-coal mining areas, respectively.

Figure 1. (a) Geographic location of the Changhe Watershed. (b) Distribution of sampling points.

A total of 120 sampling points was systematically established using the 1×1 km regular grid (Figure 1b), and soil samples were collected during 15–31 July 2015. At each point, three soil samples within a 5 m radius were collected using a soil auger and mixed together for one sample at a top layer of 0–20 cm according to the ploughing depth in the cultivated land of the study area. The samples were air-dried, gently crushed, and passed through a 2 mm sieve for soil nutrient determination. The SOM was measured by the dichromate oxidation method [25], SAN was determined by the diffusion method [26], SAP was obtained by the Olsen extraction method using alkaline sodium bicarbonate as the extractant in a 20:1 ratio [27], and SAK was extracted by a flame photometer [28]. Contents of sand (0.050–2.000 mm), silt (0.002–0.050 mm), and clay (<0.002 mm) were determined by the pipette method [29].

2.2. Collection of Environmental Factors

The products of moderate resolution imaging spectroradiometer (MODIS), including gross primary production (GPP, MOD17A2), net primary production (NPP, MOD17A3), land surface temperature for daytime and night (DLST and NLST, respectively, MOD11A2), evapotranspiration (ET, MOD16), and normalized difference vegetation index (NDVI, MOD13Q1) with 1 km resolution (the same with soil sampling scale of 1 km) in 2015 was downloaded (<https://ladsweb.modaps.eosdis.nasa.gov/>). The DLST, NLST, ET, and NDVI were used for extracting RS covariates in ENVI 5.3, including the intra-annual variance of the land surface temperature for daytime (DLSTV), the intra-annual variance of the land surface temperature for night (NLSTV), the annual mean evapotranspiration (ETM), the intra-annual standard deviation of evapotranspiration (ETD), the intra-annual variance of evapotranspiration (ETV), the annual mean NDVI (NDVIM), the inter-annual standard deviation of NDVI (NDVID), and the inter-annual variance of NDVI (NDVIV). The index of GPP, NPP, NDVIM, NDVID, and NDVIV indicated vegetation information and was selected because vegetation was the key source of the SOM. The index of DLSTV, NLSTV, ETM, ETD, and ETV indicated soil temperature and moisture and was selected because they were associated with the decomposition rates of SOM.

A digital elevation model (DEM) with 30 m spatial resolution (<http://gdem.ersdac.jpacesystems.or.jp/>) was downloaded. The DEM was used for extracting topographic factors, including elevation, aspect, slope, convergence index (CI), channel network base level (CNBL), flow accumulation (FA), LS factor (LS), topographic wetness index (TWI), valley depth (VD), and vertical distance to channel network (VDCN) in the software of SAGA ver. 6.4 (Department of Physical Geography, University of Göttingen, Germany) [30] and was resampled into the same resolution with an RS of 1 km in ArcGIS ver. 10.5 (ESRI Inc., Redlands, USA). The factors indicated topographic conditions, which were related to microclimate and soil erosion and thus affected the accumulation, decomposition, and flow of soil nutrients [12].

2.3. Two-Dimensional Empirical Mode Decomposition (2D-EMD)

Two-dimensional empirical mode decomposition has been used to separate the overall variation of any spatial or temporal dataset into different scale components called bi-dimensional intrinsic mode function (BIMFs) [31]. Unlike EMD finding the overall extrema, 2D-EMD determines the local extrema (maxima and minima) of the spatial dataset. The minima and maxima points of each location have to be interpolated, and the mean values of the interpolated minima and maxima points have to be calculated. The 2D-EMD for two-dimensional spatial dataset is defined as Equation (1).

$$z(x, y) = \sum_{i=1}^N d_i(x, y) + r(x, y) \quad (1)$$

where $z(x, y)$ was the original two dimensional spatial dataset, $d_i(x, y)$ is the BIMF, and $r(x, y)$ is the residue. The main process of 2D-EMD can be seen in Figure 2. More detailed description of 2D-EMD can be found in other studies [32,33].

Figure 2. Schematic presentation of the step of two-dimensional empirical mode decomposition (2D-EMD).

2.4. Data Analysis

Soil nutrients and the selected environmental factors were decomposed into three BIMFs and their corresponding residue in the “spemd” package in R ver. 3.6 (The University of Auckland, Oakland, New Zealand) [34] and guaranteed that each scale component contained at least 5% variance of the original data. However, the environmental factors of DLSTV, NLSTV, NDVIM, NDVID, and slope could be decomposed into only one BIMF and its residue, while the environmental factors of ETD and elevation were only decomposed into two BIMFs and its residue.

The percentage variance contributions of each BIMF and residue were obtained through the ratio of each BIMF variance to the variance of original data as Equation (2).

$$\text{Variance (\%)} = \frac{\text{Variance of each BIMF (or residue)}}{\text{variance of original data at the sampling scale}} \times 100 \quad (2)$$

The Pearson's correlation coefficients between soil nutrients and environmental factors at the sampling scale or the scale components (BIMFs and residue) were calculated.

The analysis of variance (ANOVA) was used to analyze soil nutrients between coal mining and non-coal mining areas. When the soil nutrients (SOM and SAN) had significant difference, different prediction models were established for the coal mining area and non-coal mining area, respectively. However, one prediction model was established for the entire area when soil nutrients had no significant difference. Regardless of one model or two different models for the coal mining and non-coal mining areas, 75% of sampling points were selected randomly as calibration samples, and 25% of sampling points were selected as validation samples. There were four different methods of prediction, including prediction from environmental factors at the sampling scale using multiple linear stepwise regression (MLSR_{Ori}), from environmental factors at the sampling scale using partial least squares regression (PLSR_{Ori}), from environmental factors at the scale component using PLSR (PLSR_{BIMF}), and the combined method of 2D-EMD, PLSR, and MLSR (2D-EMD_{PM}). For the method 2D-EMD_{PM}, specifically, soil nutrients and the environmental factors were decomposed into BIMFs and residue using 2D-EMD; soil nutrients at each BIMF (or residue) were predicted from the environmental factors at the corresponding BIMF (or residue) using PLSR; soil nutrients at the sampling scale were estimated from the predicted soil nutrients at the BIMFs and residue using MLSR. The prediction performances were evaluated based on three parameters of coefficient of determination (R²), root mean square error (RMSE), and the ratio of performance to deviation (RPD). The R², RMSE, and RPD are defined in Equations (3)–(5), respectively. The optimal number of latent variables (LV) for PLSR models were determined at the minimum value of root mean squared error of prediction (RMSEP) in the calibration model.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

$$RPD = SD / RMSE = \sqrt{\frac{n \sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - 1)}} \quad (5)$$

where n is the number of samples; $\bar{y}$ is the mean of the measured values; and y_i and $\hat{y}_i$ are the measured and predicted values of soil nutrients in the i th sample, respectively.

3. Results

3.1. Soil Properties at the Sampling Scale

The basic statistics of soil nutrients and soil texture are shown in Table 1. The mean soil nutrients in the coal mining area were 17.42 g/kg, 32.44 mg/kg, 7.95 mg/kg, and 152.50 mg/kg for SOM, SAN, SAP, and SAK, respectively, which were lower than those in the non-coal mining area (19.88 g/kg, 39.40 mg/kg, 9.21 mg/kg, and 163.14 mg/kg for SOM, SAN, SAP, and SAK, respectively). The coefficients of variance (CVs) in the coal mining area were 44.73%, 37.12%, and 26.73% for SOM, SAN, and SAK, respectively, greater than those (28.09%, 30.26%, and 20.74% for SOM, SAN, and SAK, respectively) in the non-coal mining area. The greater CVs in the coal mining area might be attributed to mining practices in the area. The ANOVA analysis indicated that the spatial distribution of SOM and SAN

between coal mining area and non-coal mining area were different. This data indicated that the predicting models for SOM and SAN should be established in the coal mining and non-coal mining areas separately, while the predicting models for SAP and SAK could be constructed in the coal mining and non-coal mining areas together. Obviously, the mean sand in coal mining area (0.23) was higher than that in the non-coal mining area (0.19); the mean silt in the coal mining area (0.48) was lower than that in the non-coal mining area (0.52); the mean clay in the coal mining and non-coal mining areas were similar. The spatial differentiation between coal mining area and non-coal mining area were significant for sand and silt, while it was non-significant for clay.

Table 1. Basic statistics of soil organic matter (SOM), soil available nitrogen (SAN), soil available phosphorus (SAP), and soil available potassium (SAK) and soil texture under coal mining and non-coal mining areas, respectively.

Soil Nutrients	Areas	Min	Mean	Max	Std	CV	ANOVA	
							F	p-value
SOM (g/kg)	Coal mining area	1.10	17.42	36.42	7.79	44.73	a	
	Non-coal mining area	4.72	19.88	32.63	5.58	28.09	b	4.04
SAN (mg/kg)	Coal mining area	7.42	32.44	57.48	12.04	37.12	a	
	Non-coal mining area	5.56	39.40	66.75	11.92	30.26	b	10.12
SAP (mg/kg)	Coal mining area	0.95	7.95	16.91	3.83	48.15	a	
	Non-coal mining area	1.52	9.21	22.30	5.11	55.48	a	2.60
SAK (mg/kg)	Coal mining area	60.30	152.50	311.55	40.75	26.73	a	
	Non-coal mining area	70.35	163.14	244.55	33.83	20.74	a	2.42
Sand (-)	Coal mining area	0.09	0.23	0.64	0.09	40.06	a	
	Non-coal mining area	0.09	0.19	0.67	0.09	47.93	b	4.94
Silt (-)	Coal mining area	0.22	0.48	0.66	0.10	21.37	a	
	Non-coal mining area	0.18	0.52	0.70	0.10	18.56	b	6.13
Clay (-)	Coal mining area	0.14	0.29	0.48	0.08	26.98	a	
	Non-coal mining area	0.13	0.28	0.56	0.08	27.25	a	0.19

CV represents coefficient of variance. ANOVA represents analysis of variance, and different letters between coal mining area and non-coal mining area represent the significant difference of the soil nutrient at $p < 0.05$ level.

The relationships between soil nutrients and soil texture under the coal mining area, non-coal mining area, and the entire watershed are shown in Table 2. Despite the non-significant correlation between SOM and soil texture, SOM was positively related to silt in the coal mining area and positively correlated with clay in the non-coal mining area. Although silt had a non-significant relationship with SAN in the non-coal mining area, silt had a positive effect on SAN in the coal mining area and the entire area. The relationships between clay and SAP in the non-coal mining area and the entire area were significant. Notably, sand had significantly negative effects on SAK in the coal mining area, non-coal mining area, and the entire study area.

Table 2. The correlations between soil nutrients (SOM, SAN, SAP, and SAK) and soil texture under coal mining area, non-coal mining area, and the entire area, respectively.

Soil Texture	Coal Mining Area				Non-Coal Mining Area				The Entire Area			
	SOM	SAN	SAP	SAK	SOM	SAN	SAP	SAK	SOM	SAN	SAP	SAK
Sand	−0.10	−0.07	−0.01	−0.32 *	0.01	0.08	−0.06	−0.44 **	−0.09	−0.05	−0.07	−0.39 **
Silt	0.27 *	0.34 **	−0.12	0.32 *	−0.23	−0.10	−0.15	0.16	0.10	0.18 *	0.10	0.27 **
Clay	−0.22	−0.37 **	0.20	−0.07	0.29 *	0.03	0.26 *	0.32 *	0.02	0.18	0.22 *	0.09

SOM represents soil organic matter; SAN represents soil available nitrogen; SAP represents soil available phosphorus; SAK represents soil available potassium; * significant at $p < 0.05$; ** significant at $p < 0.01$.

3.2. Scale-Specific Relationships of Soil Nutrients with Environmental Factors

The scale components for each soil nutrient were decomposed into three BIMFs and residue by 2D-EMD. The soil nutrients at the sampling scale and at these scale components are shown in Figure 3. The spatial distributions of SOM and SAN at the sampling scale were identical, while the spatial

pattern of SAP and SAK at the residue were similar. Moreover, the distinction between coal mining and non-coal mining areas for SAN were obvious.

Figure 3. Spatial distributions of soil nutrients and their scale components of bi-dimensional intrinsic mode functions (BIMF1, BIMF2, BIMF3) and residue, including (a) soil organic matter (SOM), (b) soil available nitrogen (SAN), (c) soil available phosphorus (SAP), and (d) soil available potassium (SAK).

The percentages of variance explained by each scale component for soil nutrients are shown in Table 3. Obviously, the percentage of variance explained by BIMF1 were greater than half of the total variance, indicating the significantly stochastic effects on soil nutrients, such as coal mining operations. Notably, the percentages of variance explained by BIMF1 for SOM and SAN in the coal mining area were slightly lower than those in the non-coal mining areas, which might be attributed to the less cultivated activities because of the severe subsidence resulting from coal-mining operations. Thus, the effects of cultivated activities on soil nutrients were remarkable in the study area, especially for SAP. Conversely, the structured variance for SOM and SAN, which stand for the sum of BIMF2, BIMF3, and residue, were also greater in the coal mining area than those in the non-coal mining area.

Table 3. Percentage of variance explained by each bi-dimensional intrinsic mode functions (BIMF) or residue for soil nutrients (%).

Soil Nutrients	Area	BIMF1	BIMF2	BIMF3	Residue	SV
SOM	Coal mining area	53.59	24.24	19.47	6.82	50.53
	Non-coal mining area	57.08	19.92	13.88	6.24	40.04
SAN	Coal mining area	55.57	17.99	12.15	7.17	37.31
	Non-coal mining area	58.59	15.11	9.06	10.42	34.59
SAP	The entire area	73.60	19.19	1.05	11.44	31.68
SAK	The entire area	64.48	20.24	4.80	5.65	30.69

SOM: soil organic matter; SAN: soil available nitrogen; SAP: soil available phosphorus; SAK represents soil available potassium; SV: Structured variance, which equals BIMF2 + BIMF3 + Residue.

The correlation coefficients between the scale component of SOM and the corresponding scale component of topographic factors or RS index are shown in Figure 4. In the coal mining area, GPP at BIMF2 and BIMF3 was negatively correlated with SOM at some scales, while GPP at residue was positively correlated with SOM at some scales. NPP at residue negatively correlated with SOM at some scales. The land surface temperatures for daytime and night (DLSTV and NLSTV) had less correlation with SOM at the residue. The factors of evapotranspiration (ETD and ETV) at residue had significantly positive relationship with SOM. The relationships between factors of NDVI and SOM indicated that NDVI had greater effects on SOM, and the greatest coefficient was 0.81 at residue.

Elevation at all scales had a negative relationship with SOM at residue. Aspect at BIMF3 had a positive effect on SOM at almost all scales, while its residue had a negative effect on SOM at the sampling scale, BIMF3, and residue. The effects of slope on SOM distribution at different scales were not greater. CI at the sampling scale, BIMF2, and BIMF3 had positive effects on SOM at the sampling scale, BIMF2, and residue, while CI at the residue had negative effects on SOM at the sampling scale and at some scale components. The significant correlation between FA and SOM were identified at their residues. LS at residue negatively correlated with SOM, while the effects of TWI on SOM at most scales were not significant. VD at residue had a positive relationship with SOM at some scales, while VDCN at residue had a negatively correlation with SOM at some scales.

In the non-coal mining area, the correlation between influencing factors and SOM differed greatly at different scales, and they were slightly less than those in the coal mining area. Generally, the greatest correlation coefficients were identified at the residues of SOM and the influencing factors. The correlation coefficients between the scale component of SAN and the scale component of influencing factors are shown in Figure 5. The correlation coefficients of the influencing factors with SAN were generally greater than those with SOM. The relationships between GPP (NPP, DLSTV, ETD, ETV) at the residue and SAN at almost all scales were significant. The correlation between topographic factors at the residue and SAN at almost all scales were negatively significant except VD. Additionally, the correlations between influencing factors and SAN in the coal mining area were generally greater than those in the non-coal mining area.

Figure 4. Correlation coefficients between soil organic matter (SOM) and topographic index or RS index at the sampling scale and their scale components of bi-dimensional intrinsic mode functions (BIMF1, BIMF2, BIMF3) and residue in (a) the coal mining and (b) the non-coal mining area. SS represents sampling scale, Resi represents residue.

Figure 5. Correlation coefficients between soil available nitrogen (SAN) and topographic index or RS index at the sampling scale and their scale components of bi-dimensional intrinsic mode functions (BIMF1, BIMF2, BIMF3) and residue in (a) the coal mining and (b) the non-coal mining area. SS represents sampling scale, Resi represents residue.

3.3. Soil Nutrient Prediction

The evaluating parameters for soil nutrient prediction are shown in Table 4. There were two methods (MLSR_{Ori} and PLSR_{Ori}) based on the soil nutrients and influencing factors at the sampling scales, and two methods (PLSR_{BIMF} and 2D-EMD_{PM}) based on the scale components of soil nutrients and influencing factors. Although the calibration accuracy of 2D-EMD_{PM} was not the best among the four methods, its validation accuracy was the best, indicating that 2D-EMD_{PM} was more stable than the other three methods.

Table 4. Comparison of prediction parameters of soil nutrients for the methods of MLSR_{Ori}, PLSR_{Ori}, and PLSR_{BIMF} and the combination of 2D-EMD, PLSR, and MLSR (2D-EMD_{PM}).

Soil Nutrients	Methods	Calibration Accuracy			Validation Accuracy		
		R ²	RMSE	RPD	R ²	RMSE	RPD
SOM in coal mining area	MLSR _{Ori}	0.66 **	4.86	1.72	0.08	29.20	0.21
	PLSR _{Ori}	0.31 **	6.85	1.22	0.28 *	5.05	1.19
	PLSR _{BIMF}	0.55 **	5.59	1.50	0.20 *	6.45	0.93
	2D-EMD _{PM}	0.62 **	5.08	1.64	0.64 **	4.23	1.42
SOM in non-coal mining area	MLSR _{Ori}	0.52 **	3.75	1.45	0.02	59.50	0.10
	PLSR _{Ori}	0.24 **	4.68	1.16	0.12	5.82	0.98
	PLSR _{BIMF}	0.18 **	2.92	1.87	0.32 *	5.30	1.08
	2D-EMD _{PM}	0.48 **	3.95	1.38	0.57 **	4.11	1.39
SAN in coal mining area	MLSR _{Ori}	0.62 **	7.42	1.66	0.00	46.72	0.25
	PLSR _{Ori}	0.49 **	8.65	1.42	0.26	9.79	1.20
	PLSR _{BIMF}	0.71 **	6.51	1.89	0.45 **	8.95	1.31
	2D-EMD _{PM}	0.70 **	6.68	1.84	0.61 **	7.30	1.60
SAN in non-coal mining area	MLSR _{Ori}	0.71 **	6.90	1.86	0.21	164.81	0.06
	PLSR _{Ori}	0.32 **	10.45	1.23	0.30 *	8.49	1.07
	PLSR _{BIMF}	0.71 **	6.87	1.87	0.19 *	9.38	0.97
	2D-EMD _{PM}	0.43 **	9.52	1.34	0.40 **	7.27	1.25
SAP in the entire area	MLSR _{Ori}	0.23 **	4.22	1.15	0.01	23.23	0.16
	PLSR _{Ori}	0.10 **	4.58	1.06	0.07	3.56	1.01
	PLSR _{BIMF}	0.31 **	4.00	1.21	0.14 *	4.18	0.86
	2D-EMD _{PM}	0.23 **	4.22	1.15	0.15 *	3.43	1.05
SAK in the entire area	MLSR _{Ori}	0.34 **	31.40	1.23	0.00	137.59	0.25
	PLSR _{Ori}	0.26 **	33.21	1.17	0.01	40.06	0.86
	PLSR _{BIMF}	0.46 **	28.34	1.37	0.06	38.33	0.90
	2D-EMD _{PM}	0.34 **	31.40	1.23	0.20 *	33.02	1.05

MLSR_{Ori}: multiple linear stepwise regression based on original samples; PLSR_{Ori}: partial least squares regression based on original samples; PLSR_{BIMF}: partial least squares regression based on bi-dimensional intrinsic mode function; 2D-EMD_{PM}: the combined method of 2D-EMD, PLSR, and MLSR; SOM: soil organic matter; SAN: soil available nitrogen; SAP: soil available phosphorus; SAK: soil available potassium; * significant at $p < 0.05$; ** significant at $p < 0.01$.

The parameters for the developed 2D-EMD_{PM} models for SOM, SAN, SAP, and SAK are shown in Table 5, and the numbers of latent variables (LVs) are listed. Generally, the LVs gradually increased with the increasing scales. Obviously, the predicting performance was the worst at the scale component of BIMF1, and they were becoming better at the larger scales. Meanwhile, the predicting accuracy for SOM and SAN was greater in the coal mining area than those in the non-coal mining area.

Table 5. The properties of the developed 2D-EMD_{PM} models for soil nutrients.

Soil Nutrients	Procedure	LVs	Calibration Accuracy			Validation Accuracy		
			R ²	RMSE	RPD	R ²	RMSE	RPD
SOM in coal mining area	BIMF1 (PLSR)	3	0.12 *	5.65	1.05	0.13	5.45	1.09
	BIMF2 (PLSR)	22	0.95 **	0.69	4.51	0.64 **	2.02	1.35
	BIMF3 (PLSR)	23	0.99 **	0.21	10.69	0.87 **	0.96	2.71
	Residue (PLSR)	41	1.00 **	0.00	1496.61	1.00 **	0.06	18.04
	SOM (MLSR)	-	0.62 **	5.08	1.64	0.64 **	4.23	1.42
			$-6.31 + 1.64 \text{ IMF2}' + 0.83 \text{ IMF3}' + 1.33 \text{ Residue}'$					
SOM in non-coal mining area	BIMF1 (PLSR)	2	0.11	4.29	0.98	0.16	3.76	0.80
	BIMF2 (PLSR)	15	0.81 **	1.20	2.29	0.39 **	2.45	1.16
	BIMF3 (PLSR)	38	1.00 **	0.05	47.47	0.78 **	1.44	1.70
	Residue (PLSR)	35	1.00 **	0.01	175.19	0.99 **	0.15	9.48
	SOM (MLSR)	-	0.48 **	3.95	1.38	0.57 **	4.11	1.39
			$2.39 + 0.95 \text{ IMF2}' + 0.87 \text{ IMF3}' + 0.87 \text{ Residue}'$					
SAN in coal mining area	BIMF1 (PLSR)	1	0.09 *	8.09	1.06	0.04	7.39	1.04
	BIMF2 (PLSR)	18	0.94 **	4.02	1.10	0.65 **	4.11	1.18
	BIMF3 (PLSR)	40	1.00 **	0.01	657.15	0.98 **	0.54	14.17
	Residue (PLSR)	42	1.00 **	0.00	5327.92	0.99 **	0.20	9.83
	SAN (MLSR)	-	0.70 **	6.68	1.84	0.61 **	7.30	1.60
			$-3.09 + 0.44 \text{ IMF2}' + 1.36 \text{ IMF3}' + 2.69 \text{ Residue}'$					
SAN in non-coal mining area	BIMF1 (PLSR)	1	0.03	9.64	1.02	0.41 **	7.53	0.86
	BIMF2 (PLSR)	20	0.87 **	1.55	2.81	0.35 **	5.03	0.87
	BIMF3 (PLSR)	40	1.00 **	0.00	1024.42	0.92 **	3.70	0.96
	Residue (PLSR)	41	1.00 **	0.00	1742.53	1.00 **	0.27	13.76
	SAN (MLSR)	-	0.43 **	9.52	1.34	0.40 **	7.27	1.25
			$-0.11 + 0.87 \text{ IMF2}' + 1.74 \text{ IMF3}' + 0.92 \text{ Residue}'$					
SAP in the entire area	BIMF1 (PLSR)	1	0.02	4.01	1.02	0.02	6.55	0.30
	BIMF2 (PLSR)	52	0.95 **	0.45	4.40	0.50 **	2.55	1.27
	BIMF3 (PLSR)	59	0.99 **	0.04	11.08	0.66 **	0.28	1.60
	Residue (PLSR)	61	1.00 **	0.01	166.18	0.99 **	0.13	11.65
	SAP (MLSR)	-	0.23 **	4.22	1.15	0.15 *	3.43	1.05
			$1.88 + 0.88 \text{ IMF2}' + 0.96 \text{ IMF3}' + 0.78 \text{ Residue}'$					
SAK in the entire area	BIMF1 (PLSR)	52	0.69 **	17.39	1.80	0.01	103.53	0.26
	BIMF2 (PLSR)	52	0.95 **	3.66	4.61	0.56 **	13.83	1.30
	BIMF3 (PLSR)	54	1.00 **	0.31	25.56	0.98 **	1.32	6.92
	Residue (PLSR)	52	1.00 **	0.12	73.98	0.99 **	0.77	11.75
	SAK (MLSR)	-	0.34 **	31.40	1.23	0.20 *	33.02	1.05
			$-27.27 + 0.96 \text{ IMF2}' + 0.60 \text{ IMF3}' + 1.16 \text{ Residue}'$					

SOM: soil organic matter; SAN: soil available nitrogen; SAP: soil available phosphorus; SAK: soil available potassium; MLSR: multiple linear stepwise regression; PLSR: partial least squares regression; BIMF: bi-dimensional intrinsic mode functions; 2D-EMD_{PM}: the combined method of 2D-EMD, PLSR, and MLSR; LVs: Number of latent variables; IMF2', IMF3', and Residue' were the predicted IMF2, IMF3, and Residue by PLSR; * significant at $p < 0.05$; ** significant at $p < 0.01$.

The spatial distributions of soil nutrient predicted errors based on 2D-EMD_{PM} are shown in Figure 6. Considering the distinction between the coal mining area and non-coal mining area of SOM and SAN, different models were established. The greater predicted errors were distributed in a dispersed pattern and mainly located in the validated samples.

The measured SOM were generally lower than the predicted in the coal mining area, while the measured SOM were generally higher than the predicted in the non-coal mining area. This might be attributed to the mining operations resulting in the SOM leakage in the coal mining area, and the fertilizer application resulted in the higher content of SOM in the non-coal mining area. The predicted SAN were higher than the measured, which were mainly along the Changhe River in the non-coal mining area. The highly predicted SAP and SAK were also found in the non-coal mining areas, which might be attributed to the farming practices. However, the highly predicted SAP was also found in the coal mining area, which might be due to the great transformation easily affected by stochastic factors [35].

Figure 6. The spatial distributions of soil nutrients predicted errors based on 2D-EMD_{PM}, including (a) soil organic matter (SOM) predicted error and (b) soil available nitrogen (SAN) predicted error with different models in the coal mining and non-coal mining areas; (c) soil available phosphorus (SAP) predicted error and (d) soil available potassium (SAK) predicted error with one prediction model in the entire area.

4. Discussion

Coal mining operations reduced the mean content of soil nutrients, which were consistent with previous studies [36,37], showing that land subsidence caused by coal mining could induce losses in surface soil nutrients. This might be due to the subsidence in the coal mining area resulting in the vertical leakage of soil nutrients from the top layer to the deeper layer [38]. Meanwhile, mining activities increased the variances of soil nutrients except SAP. In the Changhe Watershed, spatial distribution of SOM and SAN in the coal mining area and non-coal mining area were different, which were affected by the coal mining activities. However, the distributions of SAP and SAK were not altered by the coal mining activities, which might be attributed to the horizontal running off of soil nutrients decreasing the different patterns between the coal mining area and non-coal mining area.

The correlation coefficient between soil nutrients and the influencing factors varied greatly, and their correlation types were also different at the sampling scale and each scale component, similar to previous studies [39–42]. The correlations between soil nutrients and the influencing factors were generally non-significant at the scale component of BIMF1, which might be attributed to the stochastic effects (i.e., the effect of farming practices) resulting in the random distribution of SOM at the small scale of BIMF1. The spatial distributions of BIMF1 varied greatly (Figure 3), and the worst predicting performance at BIMF1 (Table 5) demonstrated the stochastic variation at the small scale of BIMF1. In addition, a previous study [31] also proved the stochastic variation of soil nutrients at the small scale of BIMF1. Consequently, predicting models that excluded BIMF1 could improve their stability of predicting performance. On the other hand, the total percentage of variance explained by BIMF2, BIMF3, and residue was even less than the percentage of variance explained by BIMF1, which revealed the lesser effect of the structured variance and the greater effect of stochastic variance affected by human activities. Therefore, the predicting performance based on the structured variance of soil nutrients would not be better even if excluded the stochastic variance (BIMF1) in the study area.

SOM predictions using PLSR based on the scale components were better than the other methods in the calibration model except SOM in the coal mining area, indicating that predicting methods based on the scale components could improve the predicting performance, in line with previous reports [14,21]. However, these studies only established the calibration model and did not validate the predicting accuracies. The 2D-EMD_{PM} was chosen as the ideal method based on the validation accuracy, which indicated the stability of 2D-EMD_{PM}. However, the prediction performance in the coal mining area was better than those in the non-coal mining area. This might be attributed to the lesser effect of cultivated activity on SOM and SAN in the coal mining area because of the severe coal mining subsidence. Meanwhile, the large variance in the coal mining area could also result in the better performance of soil nutrients prediction. Therefore, the effects of cultivated activities were stronger than the effects of mining operations on the spatial distribution of soil nutrients. The prediction models

for SAP and SAK were not satisfied in the entire area, which might be due to the farming practices in the cultivated land, such as fertilizer applications.

Because the prediction accuracy was not significant at the stochastic variance of BIMF1, and the prediction accuracy was significant at the larger scales, the predictions at the sampling scale should exclude the stochastic variance of BIMF1. These data further indicated that the predicting method of 2D-EMD_{PM} excluding BIMF1 was stable in the coal mining area and cultivated fields affected by human activities. Meanwhile, topographic factors and RS index were easily acquired from the internet, and soil samplings and measurement were not necessary. If 2D-EMD_{PM} was stable in certain coal mining areas, this would facilitate the prediction of soil nutrients in the coal mining area. However, 2D-EMD_{PM} was only applied for the validated dataset, which was randomly selected only once, and was not applied for all the possible datasets randomly selected. Meanwhile, 2D-EMD_{PM} application in other coal mining areas was not included in this study, and further study is needed in the future for better predictions of soil nutrients.

5. Conclusions

(1) The mean soil nutrients were less, and their CVs were greater in the coal mining area than those in the non-coal mining area. Meanwhile, the correlation coefficients between soil nutrients and influencing factors were generally higher in the coal mining area than those in the non-coal mining area.

(2) The small scale of BIMF1 represented stochastic variance, and their relationships with soil nutrients were almost non-significant, which were influenced by human activities. The predictions for BIMF1 were also non-significant because of the unstable distributions.

(3) Although the predicting method of ML_{SR}_{ORI} was sometimes best in the calibration models, its performance was particularly unstable in the validation model and should not be used in the areas affected significantly by coal mining operations or farming practices. The proposed method was 2D-EMD_{PM}, which decomposed the soil nutrients and their covariates into four scale components, then predicted soil nutrients at each scale component from the corresponding scale of the covariates, and finally predicted soil nutrients at the sampling scale from the predicted soil nutrients at each scale except BIMF1. The method was stable in the calibration and validation models and should be validated in other coal mining areas in the world.

Author Contributions: H.Z. and R.B. conceived and designed the research. R.S. analyzed the data. Z.X. and C.L. discussed the results. H.Z. wrote the paper, while the other authors supported writing the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science foundation of Shanxi Province (201801D221103), the Research Project of Shanxi Scholarship Council of China (2017–075), and the Innovation Grant of Shanxi Agricultural University (2017ZZ07).

Acknowledgments: The authors would like to acknowledge to all reviewers who have given us valuable information.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McBratney, A.B.; Mendonça Santos, L.M.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]
2. Gong, G.; Matveeva, S.; O'Bryant, S.E. Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. *Environ. Res.* **2014**, *130*, 59–69. [[CrossRef](#)] [[PubMed](#)]
3. Tang, X.; Xia, M.; Pérez-Cruzado, C.; Guan, F.; Fan, S. Spatial distribution of soil organic carbon stock in Moso bamboo forests in subtropical China. *Sci. Rep.* **2017**, *7*, 42640. [[CrossRef](#)] [[PubMed](#)]
4. Zhang, Y.; Guo, L.; Chen, Y.; Shi, T.; Luo, M.; Ju, Q.; Zhang, H.; Wang, S. Prediction of Soil Organic Carbon based on Landsat 8 Monthly NDVI Data for the Jiangnan Plain in Hubei Province, China. *Remote Sens.* **2019**, *11*, 1683. [[CrossRef](#)]

5. Kumar, S.; Singh, R.P. Spatial distribution of soil nutrients in a watershed of Himalayan landscape using terrain attributes and geostatistical methods. *Environ. Earth Sci.* **2016**, *75*, 473. [\[CrossRef\]](#)
6. Liu, Y.; Guo, L.; Jiang, Q.; Zhang, H.; Chen, Y. Comparing geospatial techniques to predict SOC stocks. *Soil Tillage Res.* **2015**, *148*, 46–58. [\[CrossRef\]](#)
7. Malone, B.P.; Jha, S.K.; Minasny, B.; McBratney, A.B. Comparing regression-based digital soil mapping and multiple-point geostatistics for the spatial extrapolation of soil data. *Geoderma* **2016**, *262*, 243–253. [\[CrossRef\]](#)
8. Angelopoulou, T.; Tziolas, N.; Balafoutis, A.; Zalidis, G.; Bochtis, D. Remote Sensing Techniques for Soil Organic Carbon Estimation: a Review. *Remote Sens.* **2019**, *11*, 676. [\[CrossRef\]](#)
9. Nocita, M.; Stevens, A.; Wesemael, B.V.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Dor, E.B.; Brown, D.J.; Clairotte, M.; Csorba, A. *Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring*; Advances in Agronomy; Elsevier: Cambridge, MA, USA, 2015; Volume 132, pp. 139–159.
10. Zhu, H.; Xu, Z.; Jing, Y.; Bi, R.; Yang, W. Spatial variation and predictions of soil organic matter and total nitrogen based on VNIR reflectance in a basin of Chinese Loess Plateau. *J. Soil Sci. Plant Nutr.* **2018**, *18*, 1126–1141. [\[CrossRef\]](#)
11. Hu, W.; Chau, H.W.; Si, B.C. Vis-Near IR Spectroscopy for Soil Organic Carbon Content Measurement in the Canadian Prairies. *CLEAN Soil Air Water* **2015**, *43*, 1215–1223. [\[CrossRef\]](#)
12. Schillaci, C.; Acutis, M.; Lombardo, L.; Lipani, A.; Fantappiè, M.; Märker, M.; Saia, S. Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. *Sci. Total Environ.* **2017**, *601*, 821–832. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Zhu, H.; Zhao, Y.; Nan, F.; Duan, Y.; Bi, R. Relative influence of soil chemistry and topography on soil available micronutrients by structural equation modeling. *J. Soil Sci. Plant Nutr.* **2016**, *16*, 1038–1051. [\[CrossRef\]](#)
14. Zhu, H.; Wei, H.; Yaodong, J.; Yi, C.; Rutian, B.; Wude, Y. Soil organic carbon prediction based on scale-specific relationships with environmental factors by discrete wavelet transform. *Geoderma* **2018**, *330*, 9–18. [\[CrossRef\]](#)
15. Hu, W.; Si, B.C. Soil water prediction based on its scale-specific control using multivariate empirical mode decomposition. *Geoderma* **2013**, *193*, 180–188. [\[CrossRef\]](#)
16. Liu, E.; Liu, J.; Yu, K.; He, P.; Zhao, Z. Spatial Prediction of Forest Soil Organic Matter Based on Environmental Factors and R-STPS Interpolation Methods. *Nongye Jixie Xuebao* **2015**, *46*, 133–137. [\[CrossRef\]](#)
17. Tajik, S.; Ayoubi, S.; Zeraatpisheh, M. Digital mapping of soil organic carbon using ensemble learning model in Mollisols of Hyrcanian forests, northern Iran. *Geoderma Reg.* **2020**, *20*, e00256–e00263. [\[CrossRef\]](#)
18. Zhao, Z.; Yang, Q.; Sun, D.; Ding, X.; Meng, F.-R. Extended model prediction of high-resolution soil organic matter over a large area using limited number of field samples. *Comput. Electron. Agric.* **2020**, *169*, 105172. [\[CrossRef\]](#)
19. Zhang, C.; Yang, Y. Can the spatial prediction of soil organic matter be improved by incorporating multiple regression confidence intervals as soft data into BME method? *CATENA* **2019**, *178*, 322–334. [\[CrossRef\]](#)
20. Bourennane, H.; Salvador-Blanes, S.; Couturier, A.; Chartin, C.; Pasquier, C.; Hinschberger, F.; Macaire, J.-J.; Daroussin, J. Geostatistical approach for identifying scale-specific correlations between soil thickness and topographic attributes. *Geomorphology* **2014**, *220*, 58–67. [\[CrossRef\]](#)
21. She, D.; Qian, C.; Timm, L.C.; Beskow, S.; Hu, W.; Leitzke Caldeira, T.; Montebello de Oliveira, L. Multi-scale correlations between soil hydraulic properties and associated factors along a Brazilian watershed transect. *Geoderma* **2017**, *286*, 15–24. [\[CrossRef\]](#)
22. Zhu, H.; Cao, Y.; Jing, Y.; Liu, G.; Bi, R.; Yang, W. Multi-scale spatial relationships between soil total nitrogen and influencing factors in a basin landscape based on multivariate empirical mode decomposition. *J. Arid Land* **2019**, *11*, 385–399. [\[CrossRef\]](#)
23. Xu, Z.; Zhang, Y.; Yang, J.; Liu, F.; Bi, R.; Zhu, H.; Lv, C.; Yu, J. Effect of Underground Coal Mining on the Regional Soil Organic Carbon Pool in Farmland in a Mining Subsidence Area. *Sustainability* **2019**, *11*, 4961. [\[CrossRef\]](#)
24. Nachtergaele, F.; Van Velthuizen, H.; Verelst, L.; Batjes, N.; Dijkshoorn, K.; Van Engelen, V.; Fischer, G.; Jones, A.; Montanarella, L.; Petri, M. *Harmonized World Soil Database*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2008.
25. Nelson, D.W. *Total Carbon, Organic Carbon and Organic Matter*; ASA-SSSA: Madison, WI, USA, 1982; Volume 9, pp. 961–1010.
26. Jackson, M.L. *Soil Chemical Analysis: Advanced Course*; UW-Madison Libraries Parallel Press: Madison, WI, USA, 2005.
27. Olsen, S.R.; Cole, C.V.; Watanabe, F.S. *Estimation of Available Phosphorus in Soils by Extraction with Sodium Bicarbonate*; USDA: Washington, WA, USA, 1954.

28. Isaac, E.; Kerber, J.D. *Atomic Absorption and Flame Photometry: Techniques and Uses in Soil, Plant and Water Analysis*; SSSA: Madison, WI, USA, 1972.
29. Gee, G.W.; Bauder, J.W. Particle-size analysis. In *Methods of Soil Analysis, Part 1. Physical and Mineralogical Methods*, 2nd ed.; Klute, A., Ed.; The Agronomy Series; Elsevier: Madison, WI, USA, 1986; pp. 383–411.
30. Schillaci, C.; Braun, A.; Kropáček, J. *Terrain Analysis and Landform Recognition*; British Society for Geomorphology: London, UK, 2015.
31. Zhu, H.; Sun, R.; Bi, R.; Li, T.; Jing, Y.; Hu, W. Unraveling the local and structured variation of soil nutrients using two-dimensional empirical model decomposition in Fen River Watershed, China. *Arch. Agron. Soil Sci.* **2019**, 1–14. [[CrossRef](#)]
32. Huang, J.; Wu, C.; Minasny, B.; Roudier, P.; McBratney, A.B. Unravelling scale- and location-specific variations in soil properties using the 2-dimensional empirical mode decomposition. *Geoderma* **2017**, 307, 139–149. [[CrossRef](#)]
33. Xu, G.L.; Wang, X.T.; Xu, X.G. Improved bi-dimensional empirical mode decomposition based on 2D-assisted signals: Analysis and application. In *IET Image Processing*; Institution of Engineering and Technology: Stevenage, UK, 2011; Volume 5, pp. 205–221.
34. Roudier, P. a Bi-Dimensional Implementation of the Empirical Mode Decomposition for Spatial Data. *R Package Version 0.1-0*. Available online: <https://github.com/pierrerroudier/spemd> (accessed on 1 July 2018).
35. Takeda, M.; Nakamoto, T.; Miyazawa, K. Phosphorus availability and soil biological activity in an Andosol under compost application and winter cover cropping. *Appl. Soil Ecol.* **2009**, 42, 86–95. [[CrossRef](#)]
36. Ma, K.; Zhang, Y.; Ruan, M.; Guo, J.; Chai, T. Land Subsidence in a Coal Mining Area Reduced Soil Fertility and Led to Soil Degradation in Arid and Semi-Arid Regions. *IJERPH* **2019**, 16, 3929. [[CrossRef](#)]
37. Sun, W.; Li, X.; Niu, B. Prediction of soil organic carbon in a coal mining area by Vis-NIR spectroscopy. *PLoS ONE* **2018**, 13, e0196198. [[CrossRef](#)]
38. Jing, Z.; Wang, J.; Zhu, Y.; Feng, Y. Effects of land subsidence resulted from coal mining on soil nutrient distributions in a loess area of China. *J. Clean. Prod.* **2018**, 177, 350–361. [[CrossRef](#)]
39. Hu, W.; Biswas, A.; Si, B. Application of multivariate empirical mode decomposition for revealing scale-and season-specific time stability of soil water storage. *Catena* **2014**, 113, 377–385. [[CrossRef](#)]
40. Zhou, Y.; Biswas, A.; Ma, Z.; Lu, Y.; Chen, Q.; Shi, Z. Revealing the scale-specific controls of soil organic matter at large scale in Northeast and North China Plain. *Geoderma* **2016**, 271, 71–79. [[CrossRef](#)]
41. Zhao, R.; Biswas, A.; Zhou, Y.; Zhou, Y.; Shi, Z.; Li, H. Identifying localized and scale-specific multivariate controls of soil organic matter variations using multiple wavelet coherence. *Sci. Total Environ.* **2018**, 643, 548–558. [[CrossRef](#)] [[PubMed](#)]
42. Si, B.C. Spatial scaling analyses of soil physical properties: a review of spectral and wavelet methods. *Vadose Zone J.* **2008**, 7, 547–562. [[CrossRef](#)]