RESEARCH ARTICLE

Spatial variation and predictions of soil organic matter and total nitrogen based on VNIR reflectance in a basin of Chinese Loess Plateau

Hongfen Zhu^{1,2,3}, Zhanjun Xu^{1,2}, Yaodong Jing^{1,2}, Rutian Bi^{1,2}, Wude Yang^{3*}

¹*College of Resource and Environment, Shanxi Agricultural University, Taigu, Shanxi 030801, China.*

²*National Experimental Teaching Demonstration Center for Agricultural Resources and Environment, Shanxi Agricultural University, Taigu, Shanxi, 030801, China.* ³*College of Agriculture, Shanxi Agricultural University, Taigu, Shanxi 030801, China.* *Corresponding author: sxndywd@163.com

Abstract

Taiyuan basin in Chinese Loess Plateau was characterized as the variety of landform, land use fragmentation, low soil organic matter (SOM) and soil total nitrogen (STN) content; therefore, the predictions of soil nutrients in the area were rather difficult. In this study, three soil sampling transects of cropland soil from northwest to southeast in Taiyuan basin were established and the visible-near infrared reflectance (VNIR) of soil samples were measured. The predicting models for SOM and STN based on VNIR were established, and the predicting accuracies were assessed by traditional evaluating index, wavelet transform, and semivariance structure. The traditional evaluating index showed that the partial least square regression (PLSR) and optimum number of latent variables were suitable for SOM prediction. The accuracies were “good” (RPD ranges from 2.30 to 2.40) for calibration and “moderate” (RPD ranges from 1.80 to 1.95) for validation, whereas the model and parameters of STN were “moderate” (RPD ranges from 1.83 to 1.87) for calibration and “acceptable” (RPD ranges from 1.41 to 1.48) for validation procedure. Based on the wavelet transform, the patterns of global wavelet power spectrum for predicted and measured SOM were closer than that of STN, and their difference in local wavelet spectra could present the predicting errors in the scale and location domain. The nugget effect indicated that the stochastic variability weakened, and the spatial structure of predicted SOM and STN enhanced. The range of predicted SOM and STN were greater than those of measured. Therefore, the predicting models based on independent dataset using PLSR could be used for the prediction of SOM or STN in the un-sampled areas. Wavelet transform and semivariance parameters could be used to guide the utilization of predicted values.

Keywords: Partial least square regression (PLSR), soil organic matter (SOM), soil total nitrogen (STN), spatial variation, scale

1. Introduction

Soil organic matter (SOM) and soil total nitrogen (STN) are important sources and sinks in the global carbon and nitrogen cycles, and are also the key indicators for soil quality assessment (Hu *et al.*, 2014; Rodríguez-Salgado *et al.*, 2017; She *et al.*, 2014). However, due to the spatial variation, the prediction and evaluation of SOM and STN contents are quite challenging, especially at large scales (Hu *et al.*, 2015). Visible-near infrared reflectance (VNIR) is a physically non-destructive, reproducible, rapid and inexpensive method that characterizes soil nutrients according to their reflectance in the wavelength range from 400 to 2500 nm (Gomez *et al.*, 2008), and has been used for SOM and STN predictions (Morellos *et al.*, 2016). Meanwhile, in the new concept of digital soil morphometrics, the use of soil spectroscopy and techniques for analyzing soil properties can enhance the understanding of pedology (Dotto *et al.*, 2017).

Soil spectral curves contain very rich information, but also have redundant data affected by particle size, temperature, soil water, sampling manipulation, and the like. During the process of soil properties' prediction, extracting relative information from the original spectral curve are very important. The partial least square regression (PLSR) is a commonly used method for the correlation between spectral curve and soil properties (Tekin *et al.*, 2012). The accuracy of PLSR models for predicting soil properties is influenced by the optimal number of latent variables (ONLY) that is calculated by the leave-one-out cross validation with calibration samples (Stevens *et al.*, 2008). When the validation and calibration samples are independent, ONLY acquired from the calibration process may not be the best for validation. However, previous studies usually randomly selected validation samples from

total samples with a certain proportion (Reeves *et al.*, 2006; Vuuren *et al.*, 2006), and the validation and calibration samples were dependent. Therefore, it is important to evaluate whether ONLY obtained from the calibration dataset is suitable for predicting independent datasets of SOM or STN.

Traditional statistical indices calculated from comparing measured and predicted SOM or STN are usually used to evaluate the accuracy of VNIR spectral for their predictions (Dotto *et al.*, 2018; Kuang and Mouazen, 2013), but this method ignores the spatial structural of predicted SOM or STN. Actually, spatial variability and spatial structures are very important for soil nutrients' interpolating map based on soil sampling points. Therefore, the objective of this study was (1) to assess the performance of ONLY obtained from the calibration in predicting independent samples of SOM or STN and (2) to analyze the spatial variability and structures of predicted SOM or STN based on VNIR spectra.

2. Materials and Methods

2.1. Study area

The study area is located in Taiyuan basin with an area of 6,159 km² (37°00'–38°20' N, 111°30'–113°00' E) in the Chinese Loess Plateau, and this area is enclosed by hills and mountains (Figure 1a). The basin is a typical semiarid area with a mean annual temperature of 9.5 °C, precipitation of 425–520 mm, and evaporation of 1,780 mm. The landform in Taiyuan basin is characterized by a thickly loess-covered layer due to the dust deposition during the Quaternary (Zhu *et al.*, 2016b),

the thickness of which ranges from 50 to 3,000 m, and the grain size of the loess generally increases from the center to the margin of the basin (Zhu *et al.*, 2016a). Fen River, the second largest tributary of the Yellow River, runs through the basin from

northeast to southwest. The major soil types are Calcaric Fluvisols and Calcaric Cambisols under alkaline conditions according to the FAO-90 soil classification. The dominant crops in the basin are corn and winter wheat.

Figure 1. (a) Geographical location of study area and (b) the distribution and (c) elevation of sampling points at the three transects.

2.2. Experimental design

According to the direction of Fen River, the basin was divided into three sections: upstream, midstream, and downstream. To analyze the soil spectral characteristics at different sections of the basin, three sampling transects along the vertical direction of Fen River were established avoiding large non-arable lands in the different parts of the basin based on the remote sensing images and land use maps. The topography of the sampling transects was bowl-shaped with a depression in the center. The transects were at 330 m intervals and around 42 km long with 121, 128, and 134 sampling points for transect 1, 2, and 3, respectively. For the sampling points located on the non-arable land such as buildings or roads, the nearest points on arable land were used to substitute them. The sampling points were shown in Figure 1b and the elevations were shown in Figure 1c. The transect 1, 2, and 3 are located in the upstream, midstream, and downstream of Taiyuan basin, respectively.

A total of 383 points were sampled during March 12–31, 2016, and the location of soil samples were recorded by GPS. Five soil samples in the surface 0–20 cm were collected and mixed to one sample. The samples were air-dried, gently crushed, and passed through a 2 mm sieve for SOM, STN, and spectral measurement. The SOM content was determined by the Walkley and Black wet oxidation method (Nelson et al., 1996). Sub-samples were finely ground to pass through a 0.15 mm sieve, and the content of STN was determined by the Kjeldahl method (Bremner and Tabatabai, 1972). Soils (approximately 30 g) were placed in a soil holder (1.5 cm in height and 8 cm in diameter) and scanned by VNIR spectroscopy of ASD FieldSpec3 (Analytical Spectral Device Inc.) with a spectral range of 350–2500 nm and 1 nm resampled resolution (1.4 nm in the range of 350–1000 nm and

2 nm in the range of 1001–2500 nm) under laboratory conditions.

2.3. Data analysis

2.3.1. Calibration and validation processes

Based on the previous study (Hu et al., 2015), the pretreatment method of smoothing using cubic smoothing spline and extracting in 10 nm increments from spline fits can produce an accurate prediction for SOM. Therefore, before constructing predicting models for SOM or STN, the pretreatment method for VNIR spectra was used.

The partial least square (PLS) is a dimensional space reduction technique, which can maximize the covariance between the scores of the spectral and the response variables (Adeline et al., 2017), such as SOM and STN. The PLSR is the combination of PLS and classical multivariate linear regression which correlates soil property with VNIR spectral. Based on PLSR, the spectral predicting models of SOM and STN were established. The model was developed with measurements from two transects during the calibration, and the calibrated model was used to predict SOM or STN for the other transect in the validation. Thus, three models, which were model 1 (transect 2 and 3 as the calibration while transect 1 as the validation), model 2 (transect 1 and 3 as the calibration while transect 2 as the validation), and model 3 (transect 1 and 2 as the calibration while transect 3 as the validation), were developed for SOM and STN, respectively.

The PLSR calibration model to validation dataset for SOM or STN prediction requires the selection of an optimal number of latent variable. Latent values from 1 to 50 were considered in the validation process to determine whether the optimal number of latent variable calculated in the calibration was the best for the

validation. The ONLY during the calibration and validation processes were calculated by leave-one-out cross validation.

2.3.2. Model accuracy evaluation

The root mean square error (RMSE), ratio of performance to deviation (RPD), and coefficient of determination (R^2) between measured and predicted values are the most widely used parameter in spectroscopy. RMSE was calculate by:

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

where n is the number of samples, y and $\hat{y}$ are the measured and predicted values. RPD is expressed as:

$$RPD = \frac{SD}{RMSE} \quad \text{Equation 2}$$

where SD is the standard deviation. A lower RMSE, larger RPD, and larger R^2 corresponds to better prediction. Meanwhile, $RPD < 1.4$ implies an unreliable prediction of soil properties, $1.4 < RPD < 1.8$ implies an acceptable prediction, $1.8 < RPD < 2.0$ implies a moderate prediction, $2.0 < RPD < 2.5$ implies a very good prediction, while $RPD > 2.5$ implies an excellent prediction (Ma *et al.*, 2012).

2.3.3. Spatial analyses

To unravel the scale-location differentiation between the predicted and measured values, the wavelet spectra of measured and predicted SOM or STN were determined by continuous wavelet transform of Morlet wavelet. A more detailed description of the continuous wavelet spectra can be found in other publication (Farge, 1992). The wavelet spectra of SOM or STN contain both the local and global spectra. The local wavelet spectrum is the square of wavelet coefficient

at each scale of each location while the global wavelet spectrum is calculated based on wavelet power spectra at various scales along the sampling transect. Statistical significance level of 5% was used. The 95% confidence level for the local wavelet spectrum and wavelet coherence was obtained using the Monte Carlo method (Hu and Si, 2016; Torrence and Compo, 1998), and that for the global wavelet spectrum was obtained using the reshuffling method (Si, 2008).

To evaluate the VNIR capacity of capturing the spatial variation and structure for SOM or STN, the spatial variability of measured and predicted SOM or STN were analyzed by the semivariance. Different theoretical models, including the exponential, spherical, and Gaussian models, were used to fit the empirical semivariance calculated from SOM or STN. The exponential model was selected, and the parameters, including the nugget (C_0), structural variance (C), sill (C_0+C), and range (A), were obtained. The ratio of nugget to sill, which >0.75 implies a weak spatial dependence, $0.25-0.75$ a moderate spatial dependence, and <0.25 a strong spatial dependence, was used to characterize the spatial dependence of SOM or STN.

3. Results

3.1. Descriptive statistical analysis for SOM and STN

The descriptive statistical results of the overall variability at SOM or STN of the three transects and all samples are listed in Table 1. The mean values of SOM were 15.99, 17.25, and 20.24 g kg⁻¹ for transect 1, 2, and 3, respectively. The spatial distribution of SOM for transect 1 varied most intensively than those of transect 2 and 3. Meanwhile, the mean value of STN were 1.09, 1.27, and 1.17 g kg⁻¹ for transect 1, 2, and 3, respectively. The variation of STN in transect 1 was identical with that in transect 2, and their variations were notably greater than that

in transect 3. Therefore, the means of SOM or STN content (15.99 g kg⁻¹ and 1.09 g kg⁻¹, respectively) at transect 1 were the smallest, whereas CV values of SOM and STN (48.54 % and 27.17 %, respectively) at transect 1 were the highest in the basin.

According to Jia and Shao (2014), the magnitudes of variations for SOM and STN can be classified as moderate, because their CV values were all between 10 and 100%. Meanwhile, CV values of STN were obviously lower than those of SOM.

Table 1. Descriptive statistics of SOM and STN along the three sampling transects.

Nutrients	Transacts	N	Mean	Min	Max	SD	CV (%)
SOM (g kg ⁻¹)	Transact 1	121	15.99	3.11	43.27	7.76	48.54
	Transact 2	128	17.25	4.39	36.69	5.74	33.25
	Transact 3	134	20.24	5.88	44.71	8.05	39.74
	All samples	383	17.90	3.11	44.71	7.45	41.67
STN (g kg ⁻¹)	Transact 1	121	1.09	0.42	1.82	0.30	27.17
	Transact 2	128	1.27	0.44	2.43	0.34	27.09
	Transact 3	134	1.17	0.62	1.95	0.26	22.33
	All samples	383	1.18	0.42	2.43	0.31	26.26

Note: N: number; SD: standard derivation; CV: coefficient of variation.

3.2. SOM and STN prediction based on VNIR spectral

The RMSE and RPD versus the number of latent variables for the three models of SOM are shown in Figure 2. With the increase of latent variables, the value of RMSE decreased and then increased, while RPD displayed opposite. Except model 2, the RMSE value in the calibration procedure of SOM prediction was smaller than that in the validation, while RPD value in the calibration was greater

than that in the validation. The ONLY values for SOM calculated from the validation were less than those values calculated from the calibration. In the calibration procedure, they were 14, 17, and 13 for model 1, 2, and 3, respectively (RPD values were 2.40, 2.30, and 2.31 for model 1, 2, and 3, respectively), while in the validation procedure were 9, 11, and 10 for model 1, 2, and 3, respectively (RPD values were 2.10, 1.90, and 1.95, respectively) (Figure 2).

Figure 2. (a) RMSE and (b) RPD versus the number of latent variables in the calibration and validation procedures of SOM

Based on the ONLY values determined in the calibration, the RPD values of SOM in the validation were 1.95, 1.80, and 1.86, respectively (Figure 3). The RPD values were smaller than those based on ONLY values determined in the validation. For the calibration models of SOM, the predicting accuracies were “very good” for RPD values belong to the range of 2.0–2.5. For the validation models of SOM, the predicting ac-

curacies were “moderate” for RPD values belonging to the range of 1.8–2.0. Except the validation of model 1, the predicted values of SOM based on VNIR spectra appeared to be smaller than the measured values at the range of SOM > 20 g kg⁻¹, especially for model 3. For the validation procedure of model 1, the fitting line between measured and predicted SOM values was almost equal to the 1:1 line.

Figure 3. Scatter plots between the measured and predicted SOM (unit: g kg^{-1})

The RMSE and RPD versus the number of latent variables for the three models of STN are shown in Figure 4. The trend of RMSE and RPD for STN were identical with that for SOM, whereas the RPD curve in the validation dropped rapidly after the ONLY values. The ONLY values for STN calculated from the validation were closed to those values calculated

from the calibration. In the calibration procedure, they were 11, 10, and 15 for model 1, 2, and 3, respectively (RPD values were 1.87, 1.84, and 1.83 for model 1, 2, and 3, respectively), while in the validation they were 9, 12, and 11 for model 1, 2, and 3, respectively (RPD values were 1.61, 1.55, and 1.48, respectively) (Figure 4).

Figure 4. (a) RMSE and (b) RPD versus the number of latent variables in the calibration and validation procedures of STN

Based on the ONLY values determined in the calibration, the RPD values of STN in the validation were 1.48, 1.53, and 1.41, respectively. The RPD value for model 2 was close to that value based on ONLY in the validation. For the calibration models of STN, the predicting accuracies were “moderate”. For the validation models of STN, the predicting accuracies were “acceptable”. Except the validation of model 3, the predicted values of STN based on VNIR spectra presented to be greater than the measured values at the range of STN < 1.0 g kg⁻¹, and smaller at the range of STN > 1.0 g kg⁻¹. For the validation procedure of model 3, most of predicted values of STN were greater than its measured values.

3.3. Wavelet spectra for measured and predicted SOM or STN

Global and local wavelet spectrum of the predicted and measured SOM are shown in Figure 5. The trends of global wavelet power spectrum for measured SOM were identical with that for predicted SOM, which meant that the performance of VNIR predicting models for independent SOM was good in the Taiyuan basin. The global wavelet power spectrum of predicted SOM for model 1 and 2 were greater than that of measured SOM at scales > 8 km. The global wavelet power spectrum of predicted SOM for model 3 was less than that of measured SOM at scales around 7 km, which was mainly shown by the disappearance of the significant variance ($P > 5\%$) of local wavelet spectrum at scales 7 km at locations 13.2 km. This might be attributed to the underestimated SOM at high contents (Figure 3).

Figure 5. Global (graph) wavelet power spectrum and local (image) wavelet spectrum of measured or predicted SOM. For the global wavelet spectrum, X-axis indicates wavelet power spectrum of each scale, Y-axis indicates scale in km, solid line indicates the variance of SOM content and dashed line indicates the variance in a confidence level of 95%. For the local wavelet spectrum, X-axis indicates distances along the sampling transect in km, Y-axis indicates scales in km, solid black line indicates 95% significance level, and the color bar indicates strength of variance.

Global and local wavelet spectrum of the predicted and measured STN is shown in Figure 6. The difference between the predicted and measured STN was greater than that between the predicted and measured

SOM (Figure 5), which indicated that the performance of SOM predication based on VNIR spectra was better than that of STN prediction. At transect 1, the global wavelet spectrum for predicted STN fluctuate less than

that for measured STN, and the significant global wavelet spectrum for predicted SOM was larger than that of measured, which reflected on the significant wavelet variance at scales >14 km at locations 26.4–40 km in the local wavelet spectrum. At transect 2, the trend between measured and predicted STN was identical, but the shape of significant variance in the local wavelet spectrum were different. At transect 3, the significant variance for predicted STN at scales around

8 km at locations 19.8–33.0 km was greater than that for measured STN, which might be due to the overestimate STN at low contents <1.6 g kg^{-1} .

Compared with the curves between global wavelet power spectrum of SOM and that of STN, the curve shapes were similar at each transect. The significant variance for transect 1 was at scales >14 km, the wave crest for transect 2 was at scales >8 km, and for transect 3 was at scales around 8.

Figure 6. Global (graph) wavelet power spectrum and local (image) wavelet spectrum of measured or predicted STN at three transects.

3.4. Spatial structure between measured and predicted SOM or STN

Based on the geostatistics, different trends were observed at the three transects (Table 2). The sills represented the total variability of predicted

SOM were larger than that of measured SOM at transect 1 and 2, and smaller than that of measured SOM at transect 3. Sills of predicted STN were larger than that of measured STN at transect 1 and 3, and less than that of measured STN at transect 2.

Table 2. Geostatistical summary of the measured and predicted SOM and STN

Nutrients	Models	C ₀	C	C ₀ +C	C ₀ /C ₀ +C	A (m)	R ²
SOM	Measured (model 1)	28.43	27.6	56.03	0.51	4142	0.80
	Predicted (model 1)	28.74	34.31	63.05	0.46	4165	0.87
	Measured (model 2)	19.25	17.36	36.61	0.53	2891	0.44
	Predicted (model 2)	17.14	24.91	42.05	0.41	4611	0.41
	Measured (model 3)	20.48	85.64	106.12	0.19	6273	0.79
	Predicted (model 3)	15.39	49.75	65.14	0.24	10442	0.89
STN	Measured (model 1)	0.067	0.014	0.081	0.83	3698	0.67
	Predicted (model 1)	0.052	0.046	0.098	0.53	13893	0.60
	Measured (model 2)	0.072	0.066	0.138	0.52	3012	0.41
	Predicted (model 2)	0.049	0.033	0.082	0.60	3689	0.46
	Measured (model 3)	0.039	0.051	0.090	0.43	5681	0.73
	Predicted (model 3)	0.046	0.061	0.107	0.43	10966	0.75

Note: C₀: the nugget; C: structural variance; C₀+C: sill; A: range; and R²: coefficient of determination.

The trend of parameter C, which represented the structured variance, was identical with that of sills. The parameter C for predicted SOM at transect 3 was lower than that for measured value. The parameter C for predicted STN at transect 2 was also lower than that for measured STN, but no significant difference at the global and local wavelet spectra was observed. The nugget effect (C₀) for SOM and STN mainly decreased.

The structure of predicted and measured SOM or STN mainly belonged to moderate spatial dependence, for the ratio of nugget to sill was at the level of 0.25-0.75. However, the structure of predicted and measured SOM at transect 3 belonged to strong spatial dependence. The ratio of nugget to sill for predicted SOM at transect 1 and 2, and STN at transect

1 and 3 were less than those of measured. The range of predicted SOM and STN were obviously larger than those of measured.

4. Discussion

The traditional statistical analysis indicated that SOM content in different sections of Taiyuan basin ranked as: downstream > midstream > upstream. This might be due to the elevation decreased from north to south, which were 742-894, 723-807, and 707-1002 m for upstream, midstream, and downstream, respectively. However, STN content ranked as midstream > downstream > upstream, which might be attributed to the phenomenon that STN was easier influenced by the land use method than SOM (Li *et al.*, 2011), and the

combined effects of landform and human activities affected the spatial distribution of STN. Additionally, the lowest mean contents of SOM and STN, and their highest CV values at upstream might be attributed to mosaic or patchwork land use type pattern in the area, because this part is close to the capital city of Taiyuan and affected by intensive human activities.

Compared with the statistical values in the entire Chinese Loess Plateau region (620,000 km²), the mean of SOM and STN value in Taiyuan basin were greater (mean SOM value = 17.31 g kg⁻¹, mean STN value = 0.74 g kg⁻¹) and CV value (CV on SOM = 70 %, CV on STN = 59.5%) was lower than values from previous reports (Liu *et al.*, 2012; Liu *et al.*, 2013). This indicated that the statistical values of SOM and STN in Taiyuan basin could not represent those of entire Chinese Loess Plateau.

During the SOM and STN prediction with VNIR spectra, the values of RMSE and RPD versus the number of latent variables presented the trend that the predicting accuracy increased and then decreased along with the latent variables number, which was in accordance with previous study (Villar *et al.*, 2014). This result suggested that increasing the number of latent variables improved the predicting accuracies on SOM and STN, but over fitting could decrease the SOM and STN evaluation. The smaller RMSE of SOM in the validation for model 2 could be attributed to the lower values in the validation dataset of transect 2. Thus, RMSE would be affected by the sampling values, whereas RPD standardized RMSE and not be affected by the sampling values (Chang *et al.*, 2002). The RPD of SOM and STN in the calibration were better than that in the validation, indicating that the ONLY values determined in the calibration were not the best for the SOM prediction in the independent validation dataset, and the independent dataset for validation and calibration procedure could decrease the performance of their predictions.

The SOM predicting accuracies in the calibration procedure were “very good”, while those in the validation procedure belonged to “moderate” level. The STN predicting accuracies in the calibration procedure were “moderate”, while those in the validation procedure belonged to the “acceptable” level. Thus, SOM or STN predicting accuracies in the validation procedure were one level below those in the calibration procedure. Meanwhile, predicting accuracy of STN was one level below that of SOM, which was consistent with previous study (Jiang *et al.*, 2017). This might be attributed to the lower CV values of STN.

The trends of global wavelet power spectrum for predicted and measured SOM were much closer than that for predicted and measured STN, which indicated that the predicting accuracies of SOM were better than that of STN. The lower accuracy of SOM at scales 7 km at locations 13.2 km of the downstream and STN at scales >14 km at locations 26.4~40 km of the upstream, at scales >8 km at all locations of the midstream, and at scales around 8 km at locations 19.8~33.0 km of the downstream indicated that more soil sampling points was required. Moreover, comparing the local wavelet spectrum between the predicted and measured values could show the predicting accuracy in different scales and locations, which could guide the properly use of predicted values at different scales and locations.

The shape of global wavelet spectrum between SOM and STN suggested the similar trend at each transect, and the scales of significant variance ranked as downstream < midstream < upstream, which might be attributed to the gradually decreased elevation from the north to the south. Thus, the scales of significant variance for SOM and STN in Taiyuan basin were > 14 km, > 8 km, and around 8 km, and gradually decreased from the north to the south.

Based on geostatistical parameters, sills of predicted SOM at downstream or predicted STN at midstream

were less than those of measured contents might be due to the greatest variability and mean of measured SOM at downstream and of measured STN at midstream. This result indicated that the variability of predicted area was large, and SOM or STN variability would be underestimated if their variability of calibrated area was less than that of predicted area, and vice versa. The structural variance for SOM at downstream was underestimated, which might be attributed to the significant local wavelet variance disappeared at scales around 7 km and locations around 13.2. Meanwhile, the structural variance for STN at midstream was underestimated, which might be due to the worse fitting lines for the experimental and fitted semivariates (R^2 were between 0.41 and 0.46). The nugget effect among the three transects mainly decreased, indicating that the stochastic variability weakened, and the spatial structure of predicted SOM and STN enhanced.

The spatial dependence of predicted values were mainly stronger than those of measured, which could be attributed to the increased structured variability or decreased nugget effect. Thus, predicted SOM and STN with VNIR spectra had stronger spatial dependence than those of measured values. The greater range of predicted values indicated that predicted SOM and STN based on VNIR hyperspectral spectroscopy tended to be auto-correlated at a longer distance, which were in accordance with previous reported by Hu *et al.* (2015). Therefore, the interpolating maps of predicted values based on kriging generally smoother than that of measured values.

5. Conclusions

The performance of VNIR spectra in predicting SOM and STN for independent samples in Taiyuan basin was evaluated. SOM or STN predicting accuracies in the validation procedure were one

level below those in the calibration procedure, and predicting accuracy of STN was one level below that of SOM. Comparing the significant variances from wavelet transform of measured and predicted could present the predicting errors in the scale and location domains. Meanwhile, there existed similar patterns of high variation for SOM and STN in scale domain, and the scales of significant variance ranked as downstream < midstream < upstream. The wavelet transform between the measured and predicted values could guide the utilization of predicted values in the location and scale domains based on hyperspectral VNIR spectra, whereas the geostatistical analysis could direct interpolating maps using their predicted values.

Acknowledgments

This research was supported financially by the Research Project of Shanxi Scholarship Council of China (2017-075) and Special Fund for Land and Resource Ministry of China in the Public Interest (201411007). We would like to express our appreciations to Zhen Xu, Peng Wang, and Ying Guo from Shanxi Agricultural University for their assistance in soil sampling collection and sample analysis.

References

- Adeline, K.R.M., Gomez, C., Gorretta, N., Roger, J.M. 2017. Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. *Geoderma*. 288, 143-153.
- Bremner, J., Tabatabai, M. 1972. Use of an ammonia electrode for determination of ammonium in Kjeldahl analysis of soils. *Communications in Soil Science & Plant Analysis*. 3, 159-165.

- Chang, ChengWen, Laird, David, A. 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Science*. 167, 110-116.
- Dotto, A.C., Dalmolin, R.S.D., Grunwald, S., Caten, A.T., Filho, W.P. 2017. Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. *Soil & Tillage Research*. 172, 59-68.
- Dotto, A.C., Dalmolin, R.S.D., ten Caten, A., Grunwald, S. 2018. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. *Geoderma*. 314, 262-274.
- Farge, M. 1992. Wavelet transform and their application to turbulence. *Annual Review of Fluid Mechanics*. 24, 395-457.
- Gomez, C., Rossel, R.A.V., McBratney, A.B. 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma*. 146, 403-411.
- Hu, K.L., Li, H., Wang, S.Y., Huang, F., Li, B.G. 2014. Spatial scaling effects on variability of soil organic matter and total nitrogen in suburban Beijing. *Geoderma*. 226-227, 54-63.
- Hu, W., Chau, H.W., Si, B.C. 2015. Vis-Near IR Spectroscopy for Soil Organic Carbon Content Measurement in the Canadian Prairies. *CLEAN-Soil, Air, Water*. 43, 1215-1223.
- Hu, W., Si, B.C. 2016. Technical Note: Multiple wavelet coherence for untangling scale-specific and localized multivariate relationships in geosciences. *Hydrology and Earth System Sciences Discussions*. 1-33.
- Jia, Y.H., Shao, M.A. 2014. Dynamics of deep soil moisture in response to vegetational restoration on the Loess Plateau of China. *Journal of Hydrology*. 519, 523-531.
- Jiang, Q., Li, Q., Wang, X., Wu, Y., Yang, X., Liu, F. 2017. Estimation of soil organic carbon and total nitrogen in different soil layers using VNIR spectroscopy: Effects of spiking on model applicability. *Geoderma*. 293, 54-63.
- Kuang, B., Mouazen, A.M. 2013. Non-biased prediction of soil organic carbon and total nitrogen with vis-NIR spectroscopy, as affected by soil moisture content and texture. *Biosystems Engineering*. 114, 249-258.
- Li, X.F., Chen, Z.B., Chen, H.B., Chen, Z.Q. 2011. Spatial Distribution of Soil Nutrients and Their Response to Land Use in Eroded Area of South China. *Procedia Environmental Sciences*. 10, 14-19.
- Liu, Z.P., Shao, M.A., Wang, Y.Q. 2012. Large-scale spatial variability and distribution of soil organic carbon across the entire Loess Plateau, China. *Soil Research*. 50, 114-124.
- Liu, Z.P., Shao, M.A., Wang, Y.Q. 2013. Spatial patterns of soil total nitrogen and soil total phosphorus across the entire Loess Plateau region of China. *Geoderma*. 197-198, 67-78.
- Ma, Z.Y., Du, C.W., Zhou, J.M. 2012. Characterization of soil calcium carbonate using mid-infrared photoacoustic spectroscopy. *spectroscopy and spectral analysis*. 32, 1255-1258.
- Morellos, A., Pantazi, X.-E., Moshou, D., Alexandridis, T., Whetton, R., Tziotziou, G., Wiebensohn, J., Bill, R., Mouazen, A.M. 2016. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosystems Engineering*. 152, 104-116.
- Nelson, D.W., Sommers, L.E., Sparks, D.L., Page, A.L., Helmke, P.A., Loeppert, R.H., Soltanpour, P.N., Tabatabai, M.A., Johnston, C.T., Sumner, M.E. 1996. Total carbon, organic carbon, and organic matter. *Methods of Soil Analysis Part—chemical Methods*. 961-1010.

- Reeves, J.B., III., Follett, R.F., McCarty, G.W., Kimble, J.M. 2006. Can Near or Mid-Infrared Diffuse Reflectance Spectroscopy be Used to Determine Soil Carbon Pools? *Communications in Soil Science and Plant Analysis*. 37, 2307-2325.
- Rodríguez-Salgado, I., Pérez-Rodríguez, P., Santás, V., Nóvoa-Muñoz, J.C., Arias-Estévez, M., Díaz-Raviña, M., Fernández-Calviño, D. 2017. Carbon mineralization in acidic soils amended with an organo-mineral bentonite waste. *Journal of soil science and plant nutrition*. 17, 624-634.
- She, D., Xuemei, G., Jingru, S., Timm, L.C., Hu, W. 2014. Soil organic carbon estimation with topographic properties in artificial grassland using a state-space modeling approach. *Canadian Journal of Soil Science*. 94, 503-514.
- Si, B.C. 2008. Spatial scaling analyses of soil physical properties: A review of spectral and wavelet methods. *Vadose Zone Journal*. 7, 547-562.
- Stevens, A., Wesemael, B.V., Bartholomeus, H., Rosillon, D., Tychon, B., Ben-Dor, E. 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma*. 144, 395-404.
- Tekin, Y., Tumsavas, Z., Mouazen, A.M. 2012. Effect of Moisture Content on Prediction of Organic Carbon and pH Using Visible and Near-Infrared Spectroscopy. *Soil science society of America journal*. 76, 188-198.
- Torrence, C., Compo, G.P. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*. 79, 61-78.
- Villar, A., Fernández, S., Gorritxategi, E., Ciria, J.I., Fernández, L.A. 2014. Optimization of the multivariate calibration of a Vis-NIR sensor for the on-line monitoring of marine diesel engine lubricating oil by variable selection methods. *Chemometrics & Intelligent Laboratory Systems*. 130, 68-75.
- Vuuren, J.A.J.V., Meyer, J.H., Claassens, A.S. 2006. Potential Use of Near Infrared Reflectance Monitoring in Precision Agriculture. *Communications in Soil Science and Plant Analysis*. 37, 2171-2184.
- Zhu, H., Hu, W., Bi, R., Peak, D., Si, B. 2016a. Scale- and location-specific relationships between soil available micronutrients and environmental factors in the Fen River basin on the Chinese Loess Plateau. *Catena*. 147, 764-772.
- Zhu, H., Zhao, Y., Nan, F., Duan, Y., Bi, R. 2016b. Relative influence of soil chemistry and topography on soil available micronutrients by structural equation modeling. *Journal of Soil Science and Plant Nutrition*. 16, 1038-1051.