CrossMark

ORIGINAL RESEARCH

Hyperspectral analysis of soil organic matter in coal mining regions using wavelets, correlations, and partial least squares regression

Lixin Lin · Yunjia Wang · Jiyao Teng · Xuchen Wang

Abstract Hyperspectral estimation of soil organic matter (SOM) in coal mining regions is an important tool for enhancing fertilization in soil restoration programs. The correlation—partial least squares regression (PLSR) method effectively solves the information loss problem of correlation—multiple linear stepwise regression, but results of the correlation analysis must be optimized to improve precision. This study considers the relationship between spectral reflectance and SOM based on spectral reflectance curves of soil samples collected from coal mining regions. Based on the major absorption troughs in the 400–1006 nm spectral range, PLSR analysis was performed using 289 independent bands of the second derivative (SDR) with three levels and measured SOM values. A wavelet-correlation-PLSR (W-C-PLSR) model was then constructed. By amplifying useful information that was previously obscured by noise, the W-C-PLSR model was optimal for estimating SOM content, with smaller prediction errors in both calibration ($R^2 = 0.970$, root mean square error (RMSEC) = 3.10, and mean relative error (MREC) = 8.75) and validation (RMSEV = 5.85 and MREV = 14.32) analyses, as compared with other models. Results indicate that W-C-PLSR has great potential to estimate SOM in coal mining regions.

L. Lin · Y. Wang · J. Teng
School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China

L. Lin · Y. Wang · J. Teng
Key Laboratory of Land Environment and Disaster Monitoring, National Administration of Surveying, Mapping and Geo-information (NASG), Xuzhou 221116, China

L. Lin · Y. Wang (✉) · J. Teng
Jiangsu Provincial Key Lab of Resources and Environment Information Engineering, Xuzhou 221116, China
e-mail: wycgcumt@163.com

Y. Wang
e-mail: wyj4139@cumt.edu.cn

X. Wang
College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Introduction

Soil organic matter (SOM) is an important indicator of soil quality in coal mining regions. SOM estimation methods for land reclamation and ecological restoration plans have therefore been receiving a great deal of attention (Demirel et al. 2011a, b). A rapid and inexpensive method for determining SOM content is essential for evaluating and managing soil resources (Chang et al. 2001). However, most approaches are based on traditional methods, which tend to be time consuming, expensive, and laborious (Sebag et al. 2006; Seely et al. 2010). Therefore, the development of fast, inexpensive, accurate, and real-time tools for measuring SOM content has become a priority task for scholars.

Hyperspectral remote sensing techniques have the potential for monitoring SOM content in coal mining regions because of an abundance of spectral information (Doetterl et al. 2013; Gomez et al. 2013). Compared with traditional laboratory methods, hyperspectral technique is quicker, cheaper, and can eliminate sample preparation and chemical reagents (Chang and Laird 2002; Demattê et al. 2004). Therefore, there have been many studies of SOM monitoring models (Croft et al. 2012). For example, using a regression tree method and hyperspectral technology, Gmur et al. (2012) quantified several soil properties, including SOM. Steffens et al. (2014) applied Vis-NIR imaging spectroscopy to map SOM quality in visually homogeneous organic surface layers. Yang and Li (2013) quantified SOM content through a combination of soil spectroscopy and multivariate stepwise linear regression. Nevertheless, there has been limited research into hyperspectral monitoring of SOM in coal mining regions, where serious soil degradation is typical.

Among the most efficient methods in constructing reliable models in the hyperspectral remote sensing field, partial least squares regression (PLSR) has been the most frequently used for estimating SOM content. Nocita et al. (2011) quantified that content via a combination of soil spectroscopy and PLSR. Vohland et al. (2011) integrated field visible near infrared spectroscopy and PLSR to predict SOM. Many scholars have indicated that the PLSR method can mitigate effects of the multicollinearity problem and may solve information losses introduced by multiple stepwise regression (which are attributable to characteristic band screening) (Janik et al. 2009; Vohland and Emmerling 2011). However, PLSR analysis and processing is severely affected by having too many variables. Still worse, the accuracy of SOM estimates can be seriously affected by noise (Groenigen et al. 2003). To alleviate these disadvantages of PLSR, some researchers have combined it with canonical correlation analysis (Chen et al. 2013; Kim et al. 2014). However, it is doubtful that this combined method can be applied to coal mining regions, because they are unique geographical and industrial areas with geospatial, social, and environmental factors that are widespread, comprehensive, dynamic, and complicated (Demirel et al. 2011a, b; Erener 2011). Furthermore, some useful information may not be correlated because of noise in the soil spectrum. Methods that reduce noise while retaining as much useful information as possible have become an urgent requirement for macroscopic SOM monitoring in coal mining regions.

In the present study, wavelet and correlation analyses were used to amplify useful information that was previously obscured by noise. Then, a satisfactory model for SOM prediction based on soil samples in coal mining regions was developed by combination with the PLSR method.

Experiments

Study area

The ecological environment of coal mining regions in Datong, China, has been seriously damaged, because it is one of the main coal-producing areas of Shanxi Province, and the damage is ongoing. The Jinghuagong National and Xinzhouyao mines are typical coal mining regions of Datong. They are composed of temperate hill zones, with widespread salinized chestnut soils. Forty-six samples of these soils were randomly selected from those collected from Jinghuagong National Mining Park (40°7′N, 113°7′E) and a coal mining subsidence region of Xinzhouyao mine (40°4′N, 113°5′E), at 0–20 cm depths (Fig. 1). All samples were air dried, crushed to pass through a 2-mm sieve, and then pulverized by grinding. The samples were split into two sets, one for chemical analysis and the other for hyperspectral measurements. SOM contents of the samples were determined using the potassium dichromate, oxidation-ferrous sulfate titrimetric method (Devi et al. 2011; Ramesh et al. 2012).

Measurement and data processing

An ASD FieldSpec 3 spectroradiometer from Analytical Spectral Devices™ was used to obtain the soil reflectance spectrum over the 350–1000 nm and 1000–2500 nm bands, with increments of 1.4 and 2 nm, respectively. The spectral resolution at 700 nm was 3 nm, and at 1400 and 2100 nm, it was 10 nm. Each soil sample filled a container (10-cm diameter and 2-cm depth) and was illuminated from above with a halogen lamp. After adjusting the zenith angle and distance between the light source and soil surface, 10 scans were made of each sample, and a white reading with a white panel was taken as calibration. All these operations were performed in a dark room to avoid the effect of stray light (Farifteh et al. 2008). Spectral data for each sample were determined using the mean of the 10 scans.
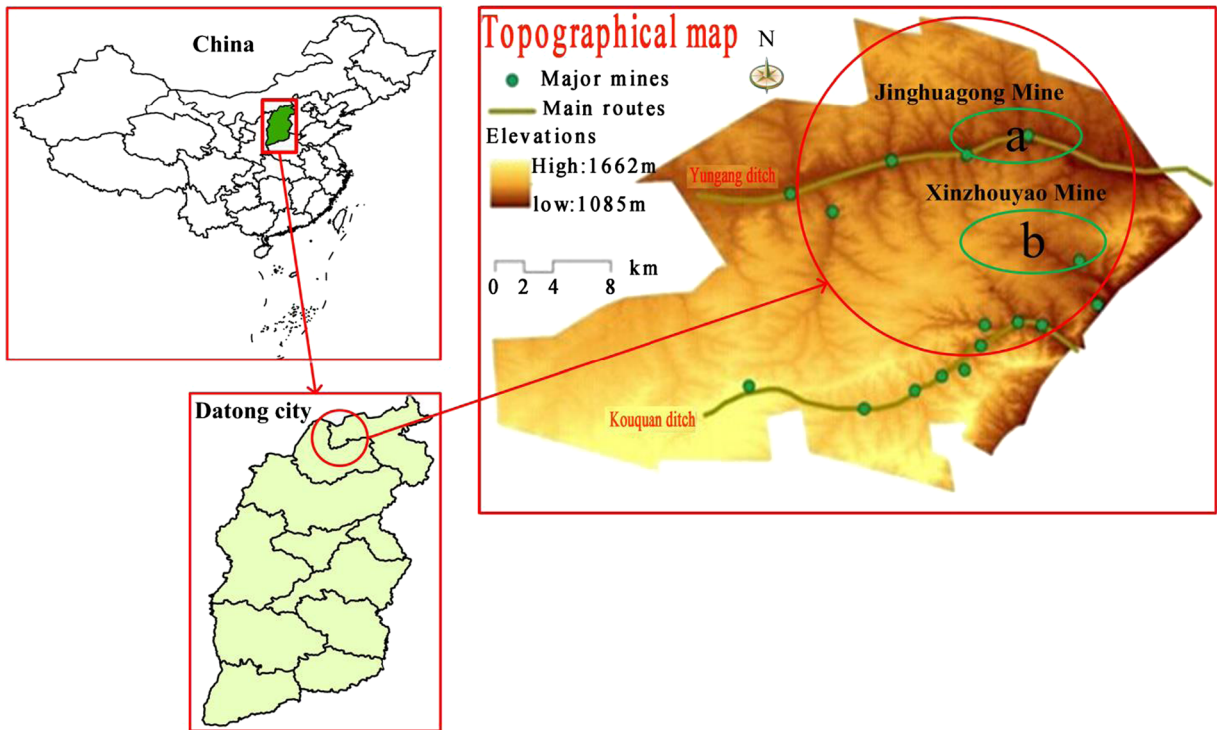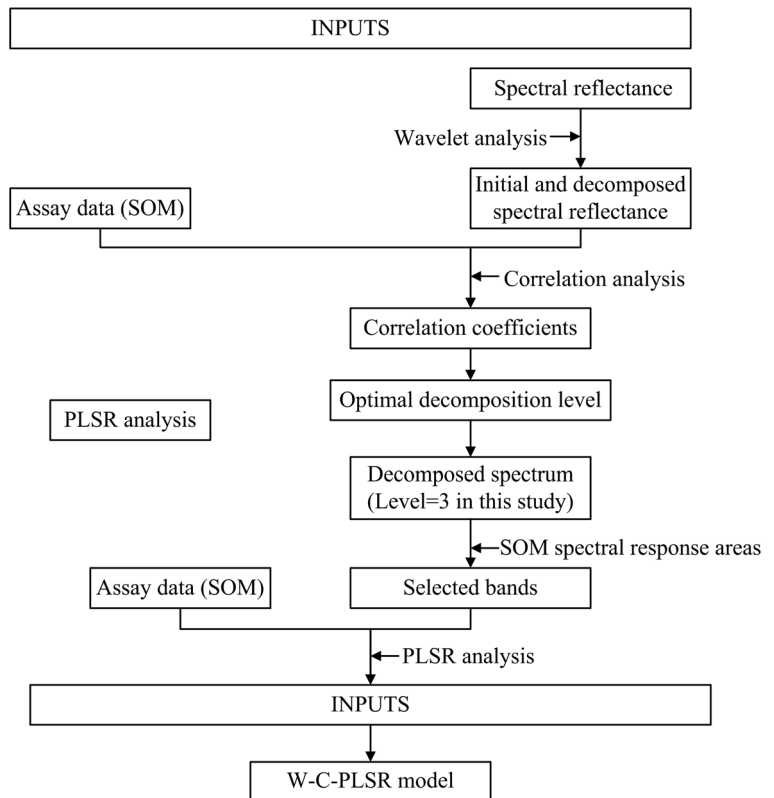
Fig. 1 Soil samples collected from **a** Jinghuagong National Mining Park and **b** coal mining subsidence region of Xinzhouyao mine

Fig. 2 Flowchart for wavelet-correlation-partial least squares regression

Spectral filtering and transformations

Derivate processing reduces the influence of low-frequency noise (Ghiyamat et al. 2013; Liaghat et al. 2014). In the inverse log mode, spectral differences in the visible region can be highlighted, and the influence of illumination variation is minimized (Wang et al. 2009). In the present work, each original spectral reflectance (REF) was transformed into a first derivative (FDR), second derivative (SDR), first derivative of the reciprocal logarithm (log(1/R))', and second derivative of the reciprocal logarithm (log(1/R))".

Gomez et al. (2008) and Lin et al. (2014) found obvious characteristic absorption troughs using the continuum removal method, so they could more easily distinguish absorption bands and build better PLSR models. The main spectral response areas of SOM content can be evaluated according to previous studies and the results of continuum removal processing. Therefore, we used continuum removal to enhance and standardize spectral absorption features (Kokaly and Clark 1999).

Wavelets-correlations-partial least squares regression method

PLSR is a mainstream, linear multiple regression method that compresses spectral data by reducing measured collinear spectral variables to a few non-correlated latent variables or factors (Geladi and Kowalski 1986; Feret et al. 2011; Singh et al. 2013). The basic aim of PLSR is to build a linear model:

$$Y = Xb + E \qquad (1)$$

where Y is a mean-centered matrix that contains the response variables, X is a mean-centered matrix that contains the predictor variables, b is a matrix that contains the regression coefficients, and E is a matrix of residuals (Cho et al. 2007). Wavelet-correlation-PLSR (W-C-PLSR) is closely related to PLSR, but uses wavelet and correlation analyses instead of using transformed spectra directly. The original soil spectra have obvious burrs, revealing a large number of noisy data in the spectral reflectance curves. This noise is also present in the transformed spectra. Therefore, the transformed data were
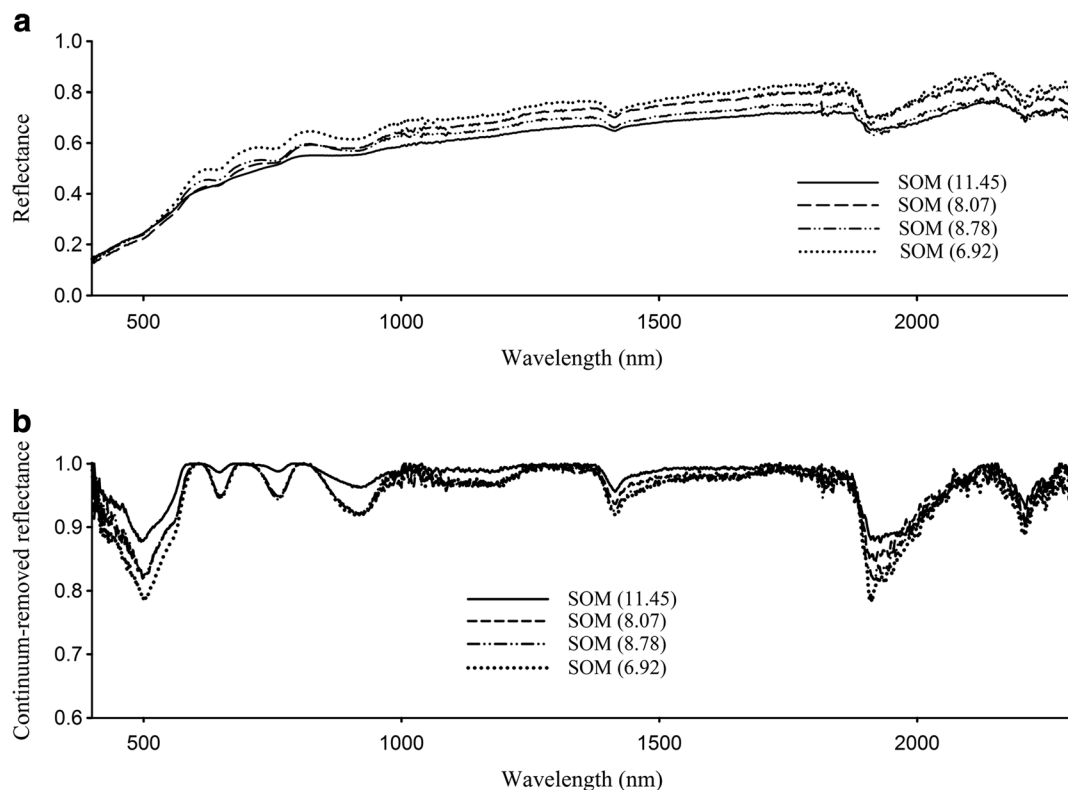

Fig. 3 Original reflectance curves (a) and spectral continuum-removed curves (b) of soil samples with differing SOM contents

decomposed using a wavelet de-noising method based on the Sym8 matrix function (Liu et al. 2011). Liu et al. (2011) noted that a three-level decomposition with the threshold de-noising method based on wavelet analysis provides an appropriate balance between curve smoothing and retention of spectral features. To select the optimal decomposition level, each spectral curve was decomposed into five layers. After the wavelet analysis, correlation between the SOM content and transformed spectra was calculated. The correlation coefficient equations are

$$r = \frac{\sum_{i=1}^{N}\left(R_i(\lambda_j) - \overline{R}(\lambda_j)\right)\left(S_i - \overline{S}\right)}{\left[\sum_{i=1}^{N}\left(R_i(\lambda_j) - \overline{R}(\lambda_j)\right)^2 \sum_{i=1}^{N}\left(S_i - \overline{S}\right)^2\right]^{0.5}} \quad (6)$$

where N is the number of soil samples, $R_i(\lambda_j)$ is transformed spectral reflectance at wavelength j, S_i is the corresponding measured SOM value, $\overline{R}(\lambda_j)$ denotes the sample mean of $\{R_i(\lambda_j)\}_{i=1}^{N}$, and $\overline{S}$ is the sample mean of $\{S_i\}_{i=1}^{N}$.

Based on the main spectral response areas of SOM content obtained through previous studies and continuum removal, sensitive bands with significant correlation coefficients ($P < 0.01$) were selected for further analysis. Lastly, PLSR analysis was performed using the selected bands and measured SOM values. The W-C-PLSR flowchart is shown in Fig. 2.

By carefully combining wavelet analysis, correlation analysis, and PLSR, the W-C-PLSR method can highlight subtle information that was obscured by noisy data, so as to take full advantage of useful spectral information and enhance accuracy.

Establishment and verification of the model

Thirty soil samples were randomly selected to construct the model. The remaining 16 samples were used for verification. Stabilities and accuracies of all the models were determined by the coefficient of determination (R^2), root mean square error of calibration (RMSEC), and mean relative error of calibration (MREC). Predictive capabilities were evaluated by root mean square error of validation (RMSEV) and mean relative error of validation (MREV). An effective model should
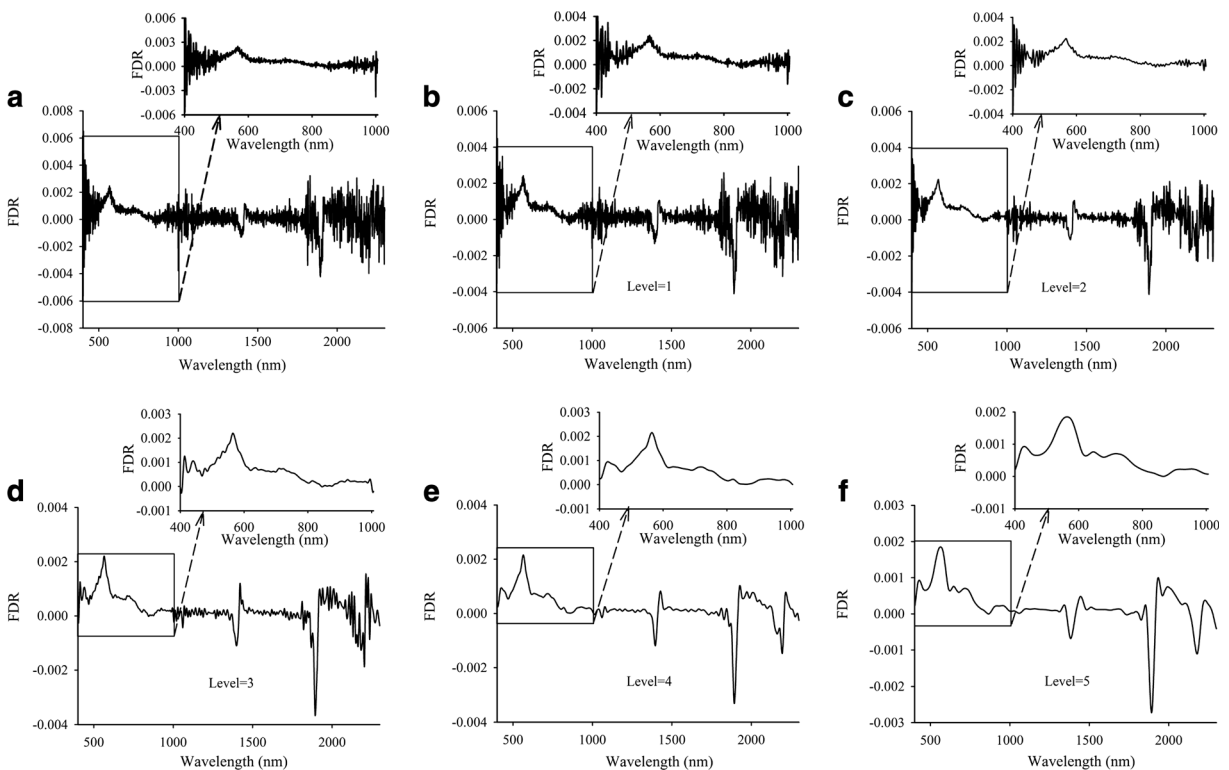

Fig. 4 De-noised first-derivative spectral reflectance curve (FDR) under different wavelet decomposition scales (b–f), compared with the initial FDR (a)

have high R^2, low RMSE (RMSEC and RMSEV), and small MRE (MREC and MREV).

$$RMSE = \sqrt{\frac{1}{N} \sum\nolimits_{i=1}^{N} \left(S_i - S_i \right)^2} \qquad (2)$$

$$MRE = \frac{1}{N} \sum\nolimits_{i=1}^{N} \frac{\left| S_i - S_i \right|}{S_i} \qquad (3)$$

where S_i is the measured value, S_i is the predicted value, and N is the number of validation samples.

Results and discussion

Interpretation of soil spectral reflectance

Differences of spectral reflectance between spectra and samples with varying SOM content (11.45, 8.07, 8.78, and 6.92 mg/kg) are shown in Fig. 3a. This figure indicates that soil reflectance generally decreases with increasing SOM content. The SOM of 11.45 showed lower reflectance values than the others, probably

because of its greater SOM content. Spectral continuum-removed curves (Fig. 3b) show seven major absorption troughs at 400–592, 609–685, 707–813, 826–1006, 1365–1531, 1850–2056, and 2146–2241 nm. Additionally, there were three remarkable water absorption peaks at 1413, 1918, and 2211 nm. Because of apparent differences of spectral characteristics caused by SOM, spectral region fitting for SOM can be predicted by hyperspectral models. Many studies have indicated that the 400–1000 nm spectral range is the main SOM spectral response area. Additionally, some investigators have found that the highest correlation between SOM and reflectance value was ~600 nm (Krishnan et al. 1980; Nocita et al. 2011). According to analysis and results of prior studies and the soil spectral continuum-removed curves in Fig. 3b, 400–1006 nm was determined as the main SOM spectral response area in our study.

Wavelet analysis

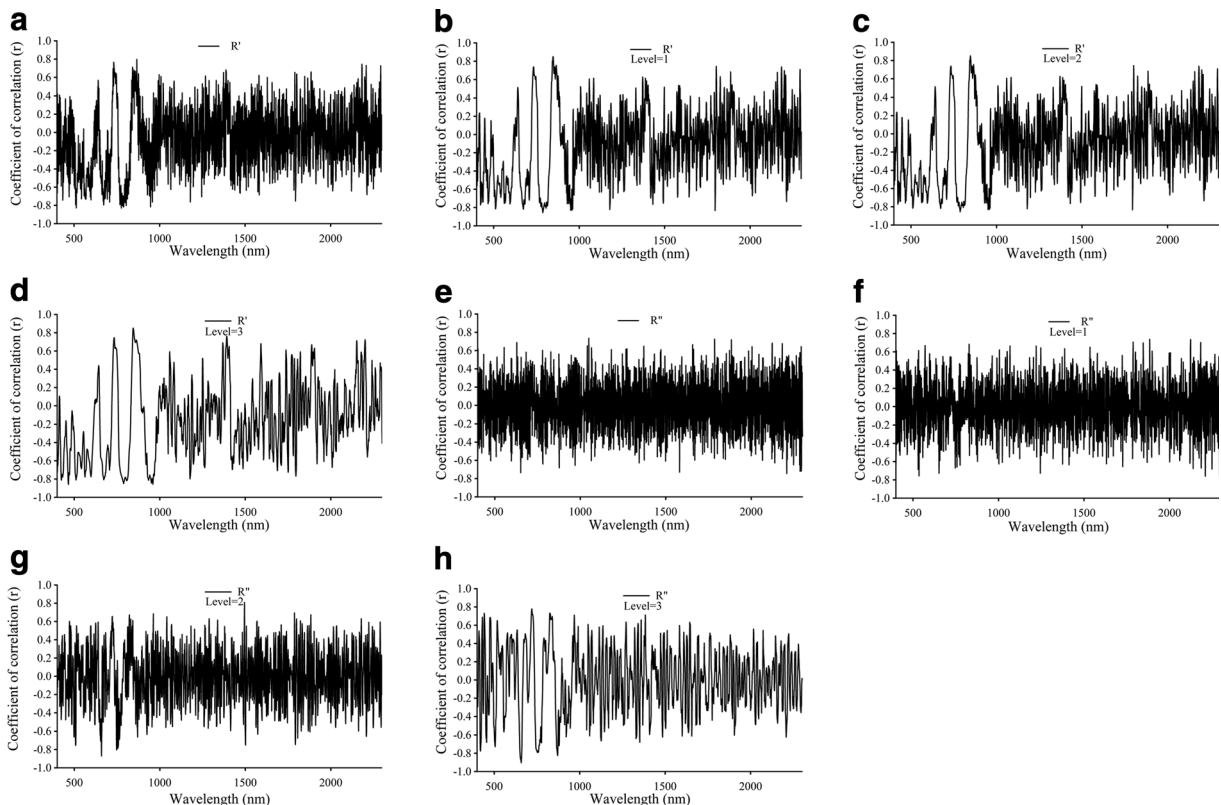Each REF was transformed into FDR, SDR, (log(1/R))', and (log(1/R))". To retain as much useful information as

Fig. 5 Correlation analysis of SOM contents with FDR (**a**–**d**) and SDR (**e**–**h**) (initial and decomposed)

possible, each transformed spectra was decomposed using five levels of approximation. Noise was reduced using the Sym8 matrix function of the wavelet. Figure 4 compares the initial FDR and decomposed results of the wavelet analysis at various levels (data of SDR, log(1/R) and (log(1/R))' not shown). It is clear that the de-noised results were not satisfactory using one level (Fig. 4b) as compared with the initial FDR (Fig. 4a). The smoothness of the spectral curve clearly improved using two levels (Fig. 4c), but there some noise remained. Using three levels (Fig. 4d), the method dramatically reduced the noise while preserving spectral characteristics, especially in the 400–1006 nm spectral range. At the same time, the maximum value was achieved near 600 nm, and the response characteristics were obvious. However, using four and five levels (Fig. 4e–f) produced curves that were too smooth, so that extreme points were not clear and useful information was lost.

Correlation analysis

Correlation coefficients for the measured SOM contents were calculated and compared with both the initial transformed spectra and decomposed spectral reflectance (one, two, and three levels) in the range 400–2300 nm. The transforms, including FDR, SDR, (log (1/R))', and (log(1/R))", are shown in Figs. 5 and 6. These figures demonstrate that the Sym8 wavelet decomposition remarkably improved correlations between SOM content and spectrum transformations in the range 400–1006 nm, especially for SDR and (log(1/R))". There was stronger correlation between SOM and SDR using three levels, with maximum correlation coefficient −0.9044 (at 658 nm). For 400–1006 nm, maxima of all correlation coefficients and the number of sensitive bands ($P<0.01$) are given in Table 1. Table and Fig. 5 demonstrate that the wavelet analysis amplified useful information that was previously obscured by noise. The decomposed SDR with three levels was the
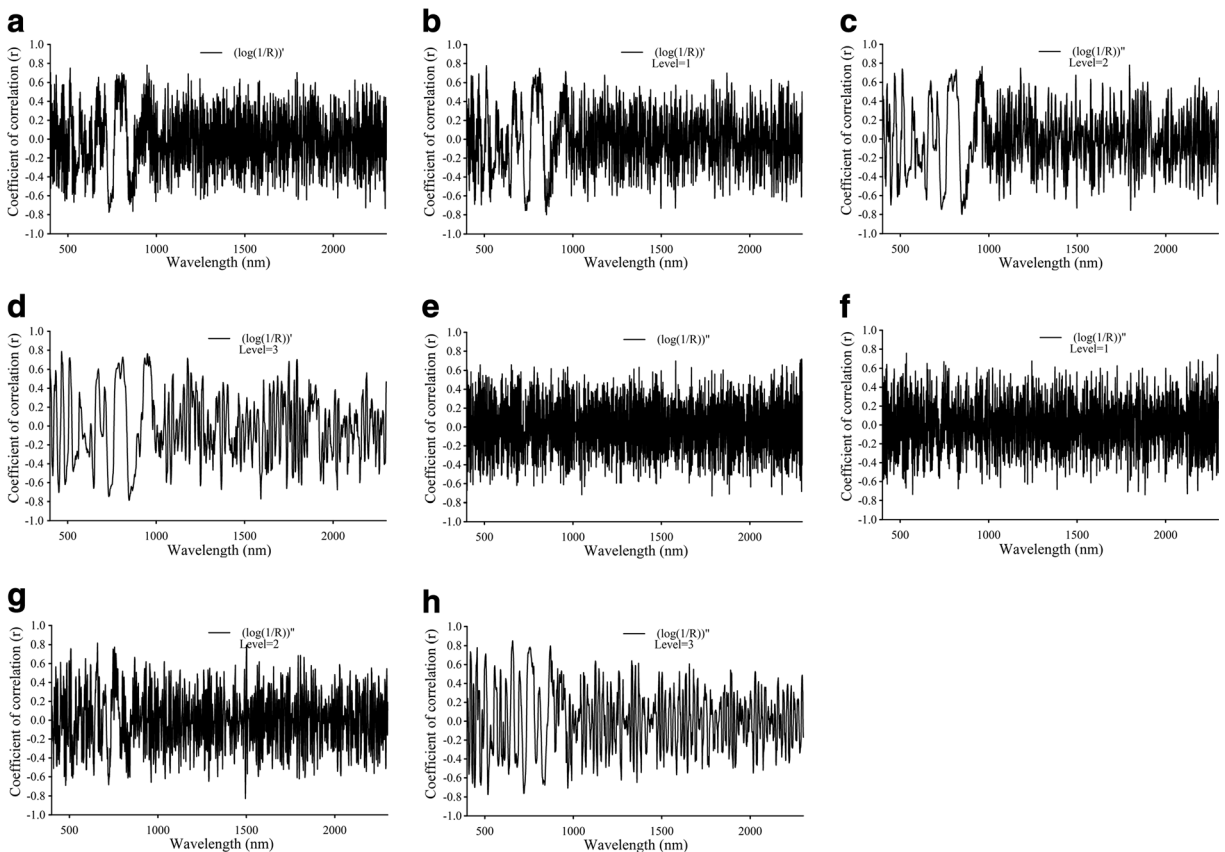

Fig. 6 Correlation analysis of SOM contents with (log(1/R))' (**a–d**) and (log(1/R))" (**e–h**) (initial and decomposed)

Table 1 Correlation analysis between SOM and mathematically transformed spectra (initial and decomposed) in the range 400–1006 nm

TSP	MPCB (nm)	CC	MNCB (nm)	CC	NSB (**)
FDR	867	0.7968	776	−0.8299	338
FDR (L = 1)	851	0.8108	776	−0.8359	356
FDR (L = 2)	847	0.8519	787	−0.8556	439
FDR (L = 3)	845	0.8513	466	−0.8593	533
SDR	652	0.7356	652	−0.7356	131
SDR (L = 1)	828	0.6835	535	−0.7570	127
SDR (L = 2)	962	0.6842	659	−0.8703	190
SDR (L = 3)	720	0.7791	658	−0.9044	289
$(Log(1/R))'$	947	0.7811	732	−0.7754	252
$(Log(1/R))'(L = 1)$	512	0.7767	851	−0.8001	264
$(Log(1/R))'(L = 2)$	962	0.7644	848	−0.7988	301
$(Log(1/R))'(L = 3)$	466	0.7868	847	−0.7881	312
$(Log(1/R))''$	652	0.6594	402	−0.6728	159
$(Log(1/R))''$ (L = 1)	535	0.7569	571	−0.7365	160
$(Log(1/R))''$ (L = 2)	659	0.8137	479	−0.6884	190
$(Log(1/R))''$ (L = 3)	657	0.8488	518	−0.7746	272

TSP types of spectral parameters, *L* level, *MPCB* maximum positive correlation band, *CC* correlation coefficient, *MNCB* minimum negative correlation band, *NSB* number of sensitive bands

**At the 0.01 significance level

most significant. The initial FDR performed better than the other initial spectrum. According to these correlation results, the SDR with three levels after continuum removal was used to build the W-C-PLSR models.

Applicability of W-C-PLSR method

The decomposed SDR with three levels maximized correlations between SOM contents and the major spectral response areas (400–1006 nm). Based on the main spectral response areas of SOM content obtained in other studies and continuum removal, 289 sensitive bands were selected for further analysis, whose correlation coefficients were significant ($P < 0.01$). Then, PLSR analysis was done using 289 independent bands

and measured SOM values. The C-MLSR method based on correlations and multiple linear stepwise regression (MLSR) often guarantees very accurate evaluation results and has great potential (Chang et al. 2001). Thus, for comparison, using wavelet, correlation, and MLSR methods, we constructed the W-C-MLSR model based on the 289 sensitive bands of the SDR, with three levels. There was information loss upon screening the bands during MLSR, and the initial FDR performed better than the other initial spectrum. Accordingly, results of a correlation-partial least squares regression (C-PLSR) model based on the 338 sensitive bands ($P < 0.01$) of the initial FDR were also compared (Pu 2012).

Table 2 shows results of the W-C-PLSR, W-C-MLSR, and C-PLSR models.

Table 2 Test results of W-C-PLSR (SDR at level 3), C-PLSR (FDR), and W-C-MLSR (SDR at level 3) models of SOM content

Models	NIB	Calibration ($n = 30$)			Validation ($n = 16$)	
		R^2	RMSEC	MREC (%)	RMSEV	MREV (%)
C-PLSR(FDR)	338	0.945	4.92	18.04	7.74	24.83
W-C-PLSR(SDR(L = 3))	289	0.970	3.10	8.75	5.85	14.32
W-C-MLSR(SDR(L = 3))	289	0.869	7.48	36.41	/	/

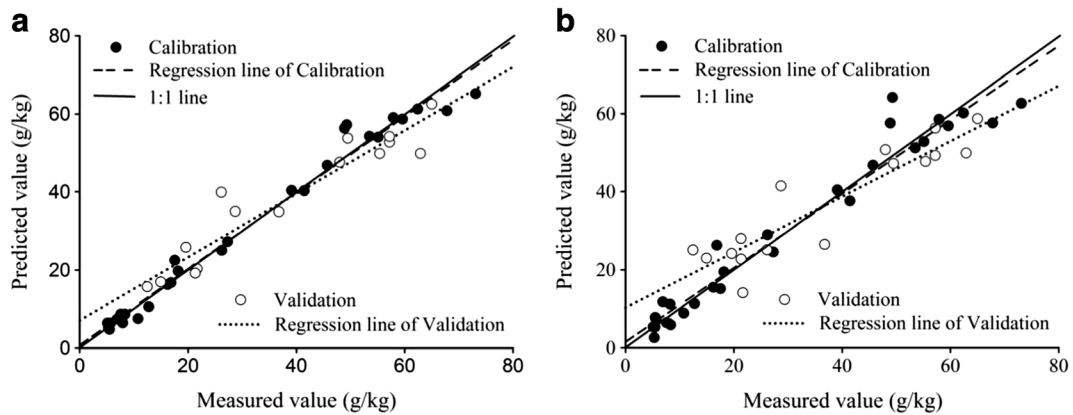*L* level, *NIB* number of independent bands

Fig. 7 Calibration and validation results using **a** W-C-PLSR method based on SDR with three levels and **b** C-PLSR model based on initial FDR

The W-C-MLSR model had small prediction errors in calibration analysis ($R^2 = 0.869$, RMSEC = 7.48, MREV = 36.41 %), indicating that the W-C-MLSR method is unreliable. The reason is clearly because of the complicated spectra of poor-quality soils in coal mining regions and information loss caused by band screening. Fortunately, the C-PLSR model gave satisfactory results in predicting SOM content (RMSEV = 7.74, MREV = 24.83 %; Table 2), but the W-C-PLSR model produced smaller errors for both calibration ($R^2 = 0.970$, RMSEC = 3.10, MREC = 8.75 %) and validation (RMSEV = 5.85, MREV = 14.32 %). As shown in Fig. 7a, W-C-PLSR samples were almost all near the 1:1 line, a much better performance than the C-PLSR model (Fig. 7b). In summary, it was demonstrated that in an environment such as the coal mining region of Datong, W-C-PLSR generated less error in the prediction of SOM content, showing that the W-C-PLSR method can augment useful information that was previously obscured by noise and greatly improve the accuracy of SOM estimates.

was determined as the main SOM spectral response region. Based on the 289 sensitive bands of SDR with three levels, whose correlation coefficients were significant ($P < 0.01$), the W-C-PLSR model was obtained. For comparison, we built the W-C-MLSR model based on 289 sensitive bands of the SDR with three levels and C-PLSR model based on 338 sensitive bands of the initial FDR. The results indicate that the W-C-MLSR method was very unreliable. The C-PLSR model produced favorable predictions of SOM content. However, the W-C-PLSR model performed the best, giving smaller prediction errors than C-PLSR. In conclusion, the W-C-PLSR method has great potential to monitor SOM in coal mining restoration regions.

Conclusions

To establish satisfactory SOM estimation models, the issue of how to reduce noise while retaining as much useful information as possible was investigated herein. By carefully applying wavelet, correlation, and PLSR techniques, the potential of the W-C-PLSR method for rapid SOM quantification was studied. According to the analysis, results of previous studies and the soil spectral continuum-removed curves in Fig. 3b, 400–1006 nm

References

Chang, C. W., & Laird, D. A. (2002). Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Science, 167*, 110–116.

Chang, C. W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R. J. (2001). Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Crop Science Society of America Journal, 65*, 480–490.

Chen, X., Liu, A. P., Wang, Z. J., & Peng, H. (2013). Corticomuscular Activity Modeling by Combining Partial

Least Squares and Canonical Correlation Analysis. *Journal of Applied Mathematics, 2013*, 1–11.

Cho, M. A., Skidmore, A., Corsi, F., van Wieren, S. E., & Sobhan, I. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation, 9*, 414–424.

Croft, H., Kuhn, N. J., & Anderson, K. (2012). On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems. *Catena, 94*, 64–74.

Demattê, J. A. M., Campos, R. C., Alves, M. C., Fiorio, P. R., & Nanni, M. R. (2004). Visible-NIR reflectance: a new approach on soil evaluation. *Geoderma, 121*, 95–112.

Demirel, N., Duzgun, S., & Emil, M. K. (2011a). Landuse change detection in a surface coal mine area using multi-temporal high-resolution satellite images. *International Journal of Mining, Reclamation and Environment, 25*, 342–349.

Demirel, N., Emil, M. K., & Duzgun, H. S. (2011b). Surface coal mine area monitoring using multi-temporal high-resolution satellite imagery. *International Journal of Coal Geology, 86*, 3–11.

Devi, O. Z., Basavaiah, K., Revanasiddappa, H. D., & Vinay, K. B. (2011). Titrimetric and spectrophotometric assay of pantoprazole in pharmaceuticals using cerium (IV) sulphate as oxidimetric agent. *Journal of Analytical Chemistry, 66*, 490–495.

Doetterl, S., Stevens, A., van Oost, K., Quine, T. A., & van Wesemael, B. (2013). Spatially-explicit regional-scale prediction of soil organic carbon stocks in cropland using environmental variables and mixed model approaches. *Geoderma, 204*, 31–42.

Erener, A. (2011). Remote sensing of vegetation health for reclaimed areas of Seyitomer open cast coal mine. *International Journal of Coal Geology, 86*, 20–26.

Farifteh, J., van der Meer, F., van der Meijde, M., & Atzberger, C. (2008). Spectral characteristics of salt-affected soils: a laboratory experiment. *Geoderma, 145*, 196–206.

Feret, J. B., Francois, C., Gitelson, A., Asner, G. P., Barry, K. M., Panigada, C., Richardson, A. D., & Jacquemoud, S. (2011). Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modelling. *Remote Sensing of Environment, 115*, 2742–2750.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. Anal. *Analytica Chemica Acta, 185*, 1–17.

Ghiyamat, A., Shafri, H. Z. M., Mandiraji, G. A., Shariff, A. R. M., & Mansor, S. (2013). Hyperspectral discrimination of tree species with different classifications using single- and multiple-endmember. *International Journal of Applied Earth Observation and Geoinformation, 23*, 177–191.

Gmur, S., Zabowski, D., & Moskal, L. M. (2012). Hyperspectral analysis of soil nitrogen, carbon, carbonate, and organic matter using regression trees. *Sensors, 12*, 10639–10658.

Gomez, C., Lagacherie, P., & Coulouma, G. (2008). Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma, 148*, 141–148.

Gomez, C., Le Bissonnais, Y., Annabi, M., Bahri, H., & Raclot, D. (2013). Laboratory Vis-NIR spectroscopy as an alternative method for estimating the soil aggregate stability indexes of Mediterranean soils. *Geoderma, 209*, 86–97.

Groenigen, J. W. V., Mutters, C. S., Horwath, W. R., & Kessel, C. V. (2003). NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field. *Plant and Soil, 250*, 155–165.

Janik, L. J., Forrester, S. T., & Rawson, A. (2009). The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. *Chemometrics and Intelligent Laboratory Systems, 97*, 179–188.

Kim, J. K., Kim, E. H., Park, I., Yu, B. R., Lim, J. D., Lee, Y. S., Lee, J. H., Kim, S. H., & Chung, I. M. (2014). Isoflavones profiling of soybean [Glycine max (L.) Merrill] germplasms and their correlations with metabolic pathways. *Food Chemistry, 153*, 258–264.

Kokaly, R. F., & Clark, R. N. (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment, 67*, 267–287.

Krishnan, P., Alexander, J. D., Butler, B. J., & Hummel, J. W. (1980). Reflectance technique for predicting soil organic-matter. *Soil Science Society of America Journal, 44*, 1282–1285.

Liaghat, S., Ehsani, R., Mansor, S., Shafri, H. Z. M., Meon, S., Sankaran, S., & Azam, S. H. M. N. (2014). Early detection of basal stem rot disease (Ganoderma) in oil palms based on hyperspectral reflectance data using pattern recognition algorithms. *International Journal of Remote Sensing, 35*, 3427–3439.

Lin, L. X., Wang, Y. J., & Xiong, J. B. (2014). Hyperspectral extraction of soil available nitrogen in nan mountain coal waste scenic spot of Jinhuagong mine based on Enter-PLSR. *Spectroscopy and Spectral Analysis, 34*, 1656–1659.

Liu, W., Chang, Q. R., Guo, M., Xing, D. X., & Yuan, Y. S. (2011). Extraction of first derivative spectrum features of soil organic matter via wavelet de-noising. *Spectroscopy and Spectral Analysis, 31*, 100–104.

Nocita, M., Kooistra, L., Bachmann, M., Müller, A., Powell, M., & Weel, S. (2011). Predictions of soil surface and topsoil organic carbon content through the use of laboratory and field spectroscopy in the Albany Thicket Biome of Eastern CapeProvince of South Africa. *Geoderma, 167–68*, 295–302.

Pu, R. L. (2012). Comparing canonical correlation analysis with partial squares regression in estimating forest leaf area index with multitemporal landsat TM imagery. *GIScience & Remote Sensing, 49*, 92–116.

Ramesh, P. J., Basavaiah, K., Devi, O. Z., Rajendraprasad, N., & Vinay, K. B. (2012). Titrimetric assay of ofloxacin in pharmaceuticals using cerium(IV) sulphate as an oxidimetric reagent. *Journal of Analytical Chemistry, 67*, 595–599.

Sebag, D., Disnar, J. R., Guillet, B., Di Giovanni, C., Verrecchia, E. P., & Durand, A. (2006). Monitoring organic matter dynamics in soil profiles by 'Rock-Eval pyrolysis': bulk characterization and quantification of degradation. *European Journal of Soil Science, 57*, 344–355.

Seely, B., Welham, C., & Blanco, J. A. (2010). Towards the application of soil organic matter as an indicator of forest ecosystem productivity: deriving thresholds, developing monitoring systems, and evaluating practices. *Ecological Indicators, 10*, 999–1008.

Singh, A., Jakubowski, A. R., Chidister, I., & Townsend, P. A. (2013). A MODIS approach to predicting stream water quality in Wisconsin. *Remote Sensing of Environment, 128*, 74–86.

Steffens, M., Kohlpaintner, M., & Buddenbaum, H. (2014). Fine spatial resolution mapping of soil organic matter quality in a Histosol profile. *European Journal of Soil Science, 65*, 827–839.

Vohland, M., & Emmerling, C. (2011). Determination of total soil organic C and hot water-extractable C from VIS-NIR soil reflectance with partial least squares regression and spectral feature selection techniques. *European Journal of Soil Science, 62*, 598–606.

Vohland, M., Besold, J., Hill, J., & Fründ, H. C. (2011). Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma, 166*, 168–205.

Wang, Y., Wang, F. M., Huang, J. F., Wang, X. Z., & Liu, Z. Y. (2009). Validation of artificial neural network techniques in the estimation of nitrogen concentration in rape using canopy hyperspectral reflectance data. *International Journal of Remote Sensing, 30*, 4493–4505.

Yang, H. F., & Li, J. L. (2013). Predictions of soil organic carbon using laboratory-based hyperspectral data in the northern Tianshan mountains, China. *Environmental Monitoring and Assessment, 185*, 3897–3908.