A NOVEL TWO-STEP SPECTRAL RECOVERY FRAMEWORK FOR COAL QUALITY ASSESSMENT BY NEAR-INFRARED SPECTROSCOPY

M. Lei [1,2], X. Yu [1], M. Li [1*], Zh. Rao [1], X. Dai [1]

[1] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China; e-mail: limingsiee@126.com
[2] University of British Columbia, Vancouver, BC, Canada

Near-infrared spectroscopy (NIRS) is an effective and efficient technique for evaluating coal quality. The original spectra might be contaminated by scattering interference and random noise. We propose a novel artifact removal framework to recover the buried information and to overcome limitations of currently available pre-processing techniques, such as the multiplicative scatter correction (MSC), as well as a smoothing process. The two-step framework is mainly constructed by MSC and Savitzky-Golay convolution (S-SGC). Moreover, a particle swarm optimization (PSO) algorithm is used to search the optimal parameters within the framework. In addition, the spectra are collected from coal samples with different particle sizes (i.e., 0.2 and 3 mm), which may carry different characteristics and interfering information. We have analyzed seven kinds of coal properties, such as moisture (%), ash (%), volatile matter (%), and heating value (MJ/kg) via partial least square regression (PLSR) models in order to verify the effectiveness of the proposed method. The results show that the proposed two-step method provides superior performances for zooming in the spectral characteristic peaks and filtering the random noise simultaneously, which mainly benefits from the appropriate combination of MSC and S-SGC.

Keywords: *near infrared spectroscopy, coal quality analysis, multiplicative scatter correction, smoothing processing, fusion mode.*

ДВУХЭТАПНАЯ СИСТЕМА АНАЛИЗА СПЕКТРОВ БЛИЖНЕГО ИНФРАКРАСНОГО ДИАПАЗОНА ДЛЯ ОЦЕНКИ КАЧЕСТВА УГЛЯ

M. Lei [1,2], X. Yu [1], M. Li [1*], Zh. Rao [1], X. Dai [1]

УДК 543.42:622.33

[1] Школа информационной и управляющей инженерии, Китайский горно-технологический университет, Сюйчжоу, Китай; e-mail: limingsiee@126.com
[2] Университет Британской Колумбии, Ванкувер, Британская Колумбия, Канада

(Поступила 5 июня 2018)

Предлагается система устранения артефактов для восстановления скрытой информации и расширения возможностей существующих методов предварительной обработки данных, таких как мультипликативная коррекция рассеяния (MSC) и сглаживание. Система включает в себя два этапа и построена в основном на свертке Савицкого–Голея (S-SGC) и MSC. Для оптимизации параметров в системе использован метод роя частиц (PSO). Спектры получены для образцов угля с различными размерами частиц (0.2 и 3.0 мм) и могут содержать информацию, обусловленную разными характеристиками и помехами. Для проверки эффективности предлагаемого способа с помощью моделей частичной регрессии наименьших квадратов (PLSR) проанализированы свойства угля, в том числе влажность (%), содержание золы (%) и летучих веществ (%), теплотворная способность (МДж/кг). Данный метод обеспечивает превосходные характеристики для масштабирования пиков спек-

тральной характеристики и одновременной фильтрации случайного шума, что достигается в основном соответствующей комбинацией алгоритмов MSC и S-SGC.

Ключевые слова: *спектроскопия ближнего ИК диапазона, анализ качества угля, мультипликативная коррекция рассеяния, сглаживание данных, режим совмещения.*

Introduction. It is highly desired to develop online analyses, which are able to rapidly provide representative information about coal properties, especially in the field of coal production, power generation, and steelmaking [1–3]. Traditional ways, which consist of laboratory evaluation using standards methods, are time consuming and not suitable for rapid multi-parameter analysis of coal properties. Recently, near infrared spectroscopy (NIRS) has become an effective technique, which can provide rapid and online analysis of coal without using any reagent [4, 5]. It is increasingly used to infer several properties of coals, such as moisture, ash, volatile matter, and heating value [6, 7].

The first step in analyzing the spectra is to pre-process the acquired data. The main reason is that the near-infrared (NIR) spectra are often contaminated by systematic noises, such as light scattering and baseline variation [8–10]. These noises are uncorrelated with the coal properties and overlapped with the coal spectra, which may make the spectral characteristic peaks not obvious. In addition, considering the fact that the composition and structure of coal are complex, the analyte of interest may absorb only in small parts of the spectral region. These unwanted noises may disturb the multivariate analysis and cause inaccurate predictions. In order to remove the undesired variations and improve the signal-to-noise ration (SNR), preprocessing of the collected spectral data is the most important step before the subsequent analysis, such as multivariate regression and classification [11–13].

Three types of pre-processing methods are commonly used for NIRS, including data enhancement, scattering correction, and smoothing filters. Popular methods of data enhancement include mean centralization, normalization, and the first/second derivative, which are commonly reported for enhancing the data variation of different spectra [14, 15]. However, these methods may amplify the random noise of spectra. As to the scattering correction, the multiplicative scatter correction (MSC) is often employed to revise the original spectra and eliminate the scatter effect [16, 17]. After being corrected, the spectral characteristic peaks can be enhanced significantly. However, it is ineffective for removing the random noise. Smoothing filters, such as smoothing by moving window average (S-MWA) and smoothing by Savitzky-Golay convolution (S-SGC), are employed to filter the random noise [18, 19]. However, at the same time, this kind of smoothing technique will inevitably smooth the tiny characteristic peaks, which may remove part of the useful information.

Each of the previous methods was specified for one kind of noise. For instance, the multiplicative scatter correction was designed for removing scattering interferences, and the smoothing processing was designed to eliminate random noise. Given these concerns, in this paper we propose a novel framework for removing more artifacts by combining MSC with the smoothing process. However, it is difficult to find the best way to remove the undesired physical phenomena in the spectra for the random setting of the framework parameter. To address this problem, prior to the smoothing processing, we employ the particle swarm optimization algorithm (PSO) to search the optimal parameters. The proposed method is able to enhance the spectral peaks and reduce the random noise at the same time. We predicted seven properties for evaluating coal quality with the partial least square regression (PLSR) model. The predicting accuracy of the analytical model was effectively improved with the processed spectra via the proposed method.

Experimental data and method. *Background.* According to the Kubelka-Munk theory [20], the absorbance of coal spectra mainly depends on the absorption coefficient K_C and scattering coefficient S_C; K_C reflects the composition as well as structure of coal (useful information), and S_C is related to the physical state of coal samples such as particle sizes, distribution, and so on (interfering information) [21, 22]. Even for the same sample, the coal spectra acquired at different particle sizes carry various characteristic information, as well as scatting interference [23]. In view of this, spectral recovery methods are studied to deal with the collected spectra from coal samples of different particle sizes, 0.2 and 3 mm.

Material and instrument. One hundred and seventeen samples of coal were collected from several mining areas in Inner Mongolia Province, Shanxi Province, Guizhou Province, and Northeast regions of China. All of them were sampled and analyzed by the traditional laboratory evaluations. Every coal was pulverized into two kinds of samples of particle size 0.2 and 3 mm. All samples were stored in an air drying chamber where temperature was set to 20°C and humidity was 10%. Each sample was divided into two parts: one for analyzing coal properties and the other one for obtaining the spectra.

Seven coal properties were analyzed in this paper, including total moisture M_t (%), air-dry moisture M_{ad} (%), air-dry ash A_{ad} (%), dry ash A_d (%), air-dry volatile matter V_{ad} (%), dry-ash-free V_{daf} (%), and heating value Q_{grad} (MJ/Kg). Therefore, the dimension of the output vector y is 7 in the analytical model. In this experiment, the properties were analyzed by a FD115 electric thermostatic dry oven, an AAF12/18 ash muffle furnace, a VMF volatile muffle furnace, a C5000 calorimeter, a BS124S electronic analytical balance and so on. All of the experimental operations were executed according to the standards GB/T211-2007, GB/T212-2008, and GB/T 213-2008.

The spectra were acquired by an Antaris Ⅱ Fourier transform NIR spectrometer. The instrument settings were given as follows: scanning number 64; wavelength point range 3799.08–10001.03 cm^{-1}; spectral resolution 4 cm^{-1}; dimension of the spectra 1609. Two kinds of spectral data sets were acquired on the basis of two particle sizes, 0.2 and 3 mm, to study the correlation of between coal properties and the spectra with different noise and scatter interference.

Methods. In this paper, a novel two-step spectral recovery framework was proposed based on the combination of MSC and the smoothing processing method with the optimal key parameters by PSO. The goal of the proposed method includes three parts: (1) enhance the unnoticeable characteristic peaks; (2) remove scattering interference and random noise from coal spectra; (3) obtain a stronger relation between the processed spectra and coal qualities.

The acquired spectrum x_i is processed by the following steps:

Step 1: Estimating the correlation between the original spectrum x_i and the reference spectrum $\bar{x}$,

$$x_i = \bar{x}_i m_i + b_i. \tag{1}$$

Step 2: Describing the scatter-corrected spectrum $\bar{x}_i$,

$$\bar{x}_i = (x_i - b_i)/m_i. \tag{2}$$

Step 3: Selecting a suitable smoothing function.

a) In S-MWA, defining the center point of the spectral data $\hat{x}_{i,j}$ in the window as

$$\hat{x}_{i,j} = \sum_{k=-r}^{r} \left(\tilde{x}_{i,j+k} s_{j,k} \Big/ \sum_{t=-r}^{r} s_{j,t} \right), \quad \tilde{x}_{i,j+k} \in \tilde{x}_i \ ; \tag{3}$$

b) In S-SGC, defining the filtered spectrum $\hat{x}_{i,j+t}$ using a polynomial function as

$$\hat{x}_{i,j+t} = \sum_{k=0}^{n} (\alpha_{i,nk} \phi(\Delta w_{j+t}, k)) = \phi'_{i,j+t} a_{i,j} \, . \tag{4}$$

Step 4: Establishing the quantitative analytical model by the PLSR method to verify the effectiveness of the mentioned methods, where the input is the processed spectrum $\hat{x}_i$ and the output is the predicting value of the coal properties $y'_{i\times 7}$.

Step 5: Optimizing the degree of the polynomial function n and the window width r using PSO.

a) Firstly, setting an average absolute error of the predicting results E_{ave} as a fitness function $f_{fitness}$.

b) Then repeating steps 3–4 until $f_{fitness}$ is reduced to a minimum.

Step 6: Obtaining the final processed spectrum $\hat{x}_i$ with the optimized parameters.

Here the reference spectrum $\bar{x}$ is defined by the average value of the original spectra, i is the samples number of the coal spectra, j is the spectrum dimension, the smoothing operator $s_{j,k} = 1$, m_i and b_i are scalar parameters, $\phi_{i,j+1}$ is the polynomial basis matrix, and $a_{i,j}$ is the coefficient vector calculated by the least square estimation method.

Results and discussion. *Multiplicative scatter correction.* MSC can eliminate the scatter inference caused by the particle size and distribution of coal samples. As shown in Fig. 1, both the characteristic peaks and random noise were enhanced. This means that MSC can eliminate the scatter interference of spectra effectively, but it is not effective for filtering the random noise.

Smoothing processing. S-MWA method filters the spectra by using the weighted arithmetic average several times in the window. It can effectively eliminate the random noise, whose mean approximately equals 0. As shown in Fig. 2, the characteristic peaks are unnoticeable and the spectral curvature changes rapidly in the region of 6000–4000 cm^{-1}. The characteristic peaks are highly similar to the random noise and hard to distinguish. Therefore, the processed spectra are easily over-smoothed by S-MWA. This may lead to the loss of unnoticeable characteristic peaks, which actually contain useful information.

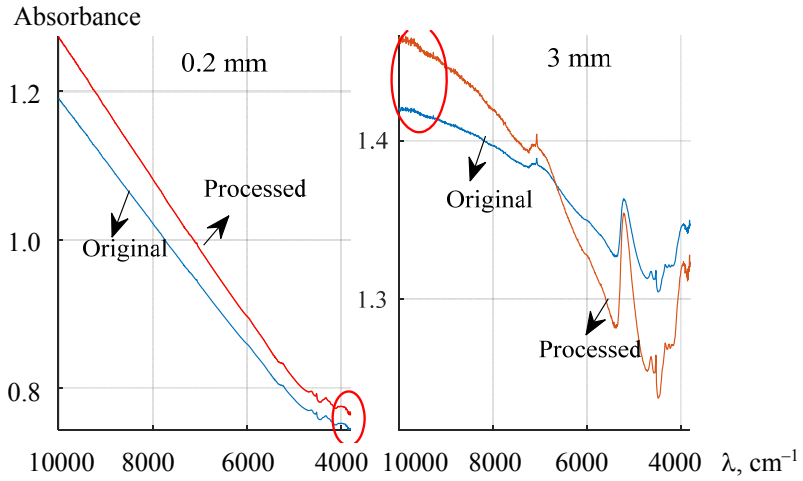

Fig. 1. Correction of coal spectra by MSC.

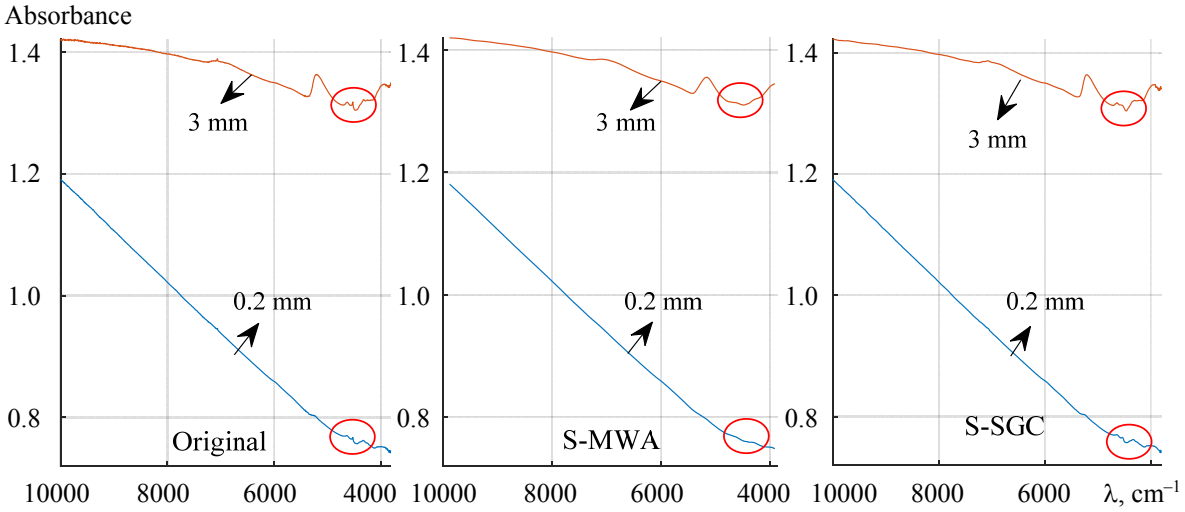

Fig. 2. Smoothing process by S-MWA and S-SGC.

With the PSO algorithm we can get the best possible combinations of the polynomial degree n and window length r to describe the characteristics of the coal spectra and to filter useless information more exactly. The whole searching process is supervised by the PLSR model to find the minimum of E_{ave} as the final objective. Inspired by information from the literature and considering the difference between the targeted samples, we heuristically set the searching range of the polynomial degree n as 1–10 and that of the window width r as 17–100 [24, 25]. In PSO, the main parameters are set as follows: dimension size 2, population size 20, iteration number 100.

For the 0.2 mm spectral data sets, the optimal polynomial degree $n = 2$, window width $r = 17$; for 3 mm spectral data sets, the optimal $n = 6$, $r = 57$. As shown in Fig. 2, the spectra were filtered with a high resolution by S-SGC. The characteristic peaks by S-SGC in the region of 6000–4000 cm^{-1} are reserved much more than by S-MWA. The unnoticeable peaks are still overlooked.

Two-step spectral recovery framework. The parameters are optimized by the PSO algorithm. For the 0.2 mm spectral data sets $n = 3$, $r = 63$; for the 3 mm spectral data sets, $n = 7$, $r = 39$. As shown in Fig. 3, ΔA represents the absorbance error between the final-processed spectrum and the MSC-processed spectrum. Compared to the MSC-processed spectrum, both of the spectra processed by the proposed method contain a less high-frequency random noise in the 0.2 and 3 mm spectral data sets. Meanwhile the unnoticeable characteristic peaks are enhanced and can be clearly distinguished from random noise comparing to Fig. 2 (processed by S-SGC). The SNR is significantly increased. This means that this method is more effective to get the recovered spectra with stronger peaks and lower random noise.

Result of quantitative analysis model. As shown in Fig. 3, the proposed method is able to reduce both random noise and scatter inference. We process two kinds of spectral data sets by three recovery methods, including MSC, S-SGC (PSO optimized), the proposed method, as well as the raw spectra. The average absolute error E_{ave}, correlation coefficient R, and root-mean-square error RMSE obtained by the PLSR model have been considered for comparison of the mentioned methods, where the desired results are $E_{ave}\to 0$, $R\to 1$, and RMSE$\to 0$.

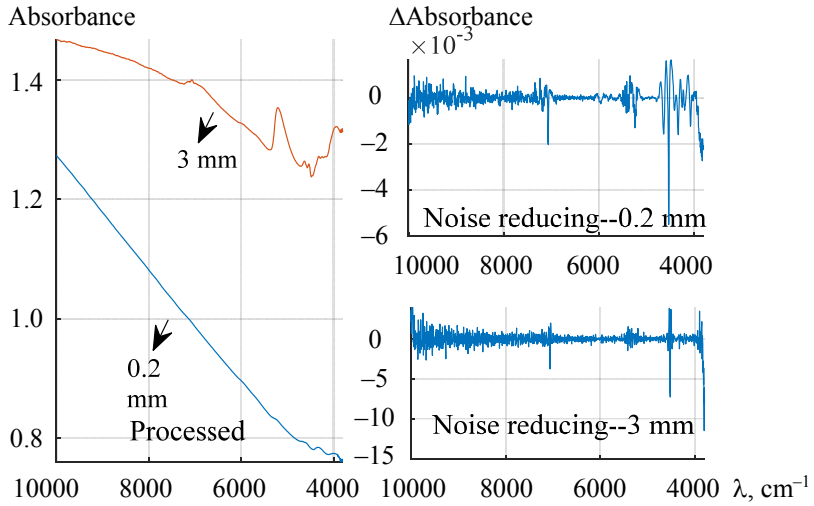

Fig. 3. Spectrum recovered by the proposed method.

The results are shown in Table 1. The spectra processed by the proposed framework give the best performance, with the smallest value of E_{ave} and RMSE as well as the highest R (most approximates to 1). This shows that the correlation between the combined-framework-processed spectra and coal properties is much stronger than for the spectra processed by the other two methods.

TABLE 1. Performance Comparison of Different Pre-processing Methods

Method	0.2 mm			3 mm		
	E_{ave}	R	RMSE	E_{ave}	R	RMSE
Raw	1.1825	0.8924	1.5327	1.9414	0.7650	2.4334
MSC	1.1014	0.8947	1.4042	1.9561	0.7758	2.4341
S-SGC	1.2157	0.8887	1.5764	1.8632	0.7978	2.2004
Proposed	1.0048	0.9272	1.2722	1.5372	0.8161	1.8407

Conclusion. In order to recover spectra with higher SNR, we analyzed the characteristics of interference and noise in the coal spectra with different particle size. We processed the spectra with the common recovery methods, such as MSC, S-MWA, and S-SGC. Considering the complexity of the coal spectra and the shortcoming of current available methods, a novel two-step spectral recovery framework was proposed to enhance the spectral characteristic peaks and to filter the random noise simultaneously. The prediction performance of the PLSR model shows that the proposed method improved the accuracy and reliability of the spectra. It also facilitates the subsequent analysis of near infrared spectra for evaluating coal qualities. In addition, the proposed spectral recovery strategy does not rely on the special characters of the targeted samples and therefore can be applied to a wide class of spectral signals from other samples.

Acknowledgment. The authors are grateful for support from the Science and Technology Planning Project of Xuzhou City (KC17075) and the China Postdoctoral Foundation (2014M551695).

REFERENCES

1. J. M. Andrés, M. T. Bona, *Talanta*, **70**, No. 4, 711–719 (2006)
2. Q. Zhu, *IEA Clean Coal Centre*, **4** (2014).
3. Y. Zhao, L. Zhang, S. X. Zhao, Y. F. Li, Y. Gong, L. Dong, W. G. Ma, W. B. Yin, S. C. Yao, J. D. Lu, L. T. Xiao, S. T. Jia, *Front. Phys.-Beijing*, **11**, No. 6, 114211 (2016).
4. Y. S. Wang, M. Yang, G. Wei, R. Hu, Z. Luo, G. Li, *Sensor. Actuat. B: Chem*., **193**, No. 3, 723–729 (2014).
5. M. Lei, M. Li, Y. Shi, *CIESC J*., **63**, No. 12, 3991–3995 (2012).
6. M. T. Bona, J. M. Andr´es, *Talanta*, **72**, 1423–1431 (2007).
7. W. K. Dong, J. M. Lee, J. S. Kim, *Korean. J. Chem. Eng*., **26**, No. 2, 489–495 (2009).
8. J. M. Andrés, M. T. Bona, *Anal. Chim. Acta*, **535**, No. 1-2, 123–132 (2005).
9. H. Jia, Q. Fu, C. J. Han, D. B. Zou, W. L. Chen, *Spectrosc. Spectr. Anal*., **32**, No. 11, 3010 (2012).
10. P. Sirisomboon, J. Nawayon, *J. Near. Infrared Spectr*., **24**, No. 2, 191–198 (2016).
11. R. Tang, K. Chen, C. Jiang, Ch. Li. *J. Appl. Spectr*., **84**, No. 4, 627–632 (2017).
12. H. Sato, M. Kiguchia, F. Kawaguchib, A. Makiac, *Neuroimage*, **21**, No. 4, 1554–1562 (2004).
13. Å. Rinnan, F. Berg, S. Engelsen, *Trac-Trend. Anal. Chem*., **28**, No. 10, 1201–1222 (2009).
14. X. Yang, F. Wang, *Adv. Mater. Res*., **898**, 831–834 (2014).
15. D. B. Kovačević, J. G. Kljusurić, P. Putnik, T. Vukušić, Z. Herceg, V. Dragović-Uzelac, *Food Chem*., **212**, 323–331 (2016).
16. Z. D. Lin, Y. B. Wang, R. J. Wang, L. S. Wang, C. P. Lu, Z. Y. Zhang, L. T. Song, Y. Liu, *J. Appl. Spectr*., **84**, No. 3, 529–534 (2017).
17. P. Geladi, D. Macdougall, H. Martens, *Appl. Spectrosc*., **39**, No. 3, 491–500 (1985).
18. P. A. Gorry. *Anal. Chem*., **62**, No. 6, 570–573 (1990).
19. Y. Shao, R. S. Lunetta, B. Wheeler, J. S. Liames, J. B. Campbell, *Remote Sens. Environ.*, **174**, 258–265 (2016).
20. P. Kubelka, F. Munk, *Zeit Für Tekn*., *Physik*, **12**, 593 (1931).
21. H. Martens, J. P. Nielsen, S. B. Engelsen, *Anal. Chem*., **75**, No. 3, 394–404 (2003).
22. H. Q. Yang, B.Y. Kuang, A. M. Mouazen, *Key Eng. Mater*., **10**, No. 3, 467–469 (2011).
23. M. Lei, M. Li, N. Wu, Y. N. Li, *Spectrosc. Spectr. Anal*., **33**, No. 1, 65–68 (2013).
24. H. Chen, T. Pan, J. Chen, Q. Lu, *Chemometr. Intell. Lab*., **107**, No. 1, 139–146 (2011).
25. H. Chen, Q. Song, G. Tang, Q. Feng, L. Lin, *ISRN Spectrosc*., **2013**, 1–9 (2013).