Research on coal moisture analysis model based on near infrared spectroscopy

Yiyou Zhou¹, Quanjun Zheng², Xiaoying Hu¹,⁵, Hui Liang³ and Changqing Dong¹,⁴

¹National Engineering Laboratory for Biomass Power Generation Equipment, School of New Energy, North China Electric Power University, Beijing 102206, China
²Beijing Aerospace Petrochemical EC and EP Technology Corporation Limited (BAEEC), Beijing 100176, China
³Beijing Shanren Zhizao Technology Development Co., Ltd., Beijing 100095, China
⁴State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China
⁵E-mail: xiaoying_826@163.com

Abstract. In order to make full use of the advantages of rapid and non-destructive measurement of coal moisture by near-infrared spectroscopy, the 29 original coal samples in the experiment were artificially humidified, and the moisture content of each coal sample was controlled within a different interval of 0-40%; diffuse reflectance near-infrared spectra were collected on the coal samples. The sample was selected according to the random method, Kennard-Stone (KS) method, and Rank-KS method for the selection of calibration set and prediction set. The prediction models of coal moisture content were established by multiple linear regression (MLR) and partial least squares algorithm (PLS) combined with different spectral pretreatment methods, furthermore, water modeling based on BP neural network was established according to optimal sample classification. The results show that when Rank-KS algorithm was used to select correction set and prediction set, the prediction ability of water content prediction model can be significantly improved by using either MLR or PLS, and the root mean square error of prediction (RMSEP) can be minimized; the moisture prediction model of the BP neural network algorithm based on the full-spectrum feature information is the best, the RMSEP value is the smallest among all models, and the model has the best prediction ability.

1. Introduction
In the current coal-fired power plants, due to the large variety of coal quality, if the operation is not adjusted in time during the work, it will seriously affect the economy and safety of the power plant. Moisture is one of the four basic indicators of coal quality economic value, and it is also an important detection parameter for coal quality [1]. Because moisture is closely related to coal calorific value and price, it plays a pivotal role in the utilization of coal at all levels. In the coal-fired system of a power plant, if the moisture content in the coal cannot be accurately detected in time and exceeds the specified range, it may cause drastic changes in the combustion process, which indirectly leads to a reduction in the service life of the boiler, affects the economic operation of the power plant, and even bring safety accident.
The traditional detection of coal moisture is mainly based on the current national standard GB/T 212-2008 [2]. The disadvantage of traditional coal moisture detection is that it takes a long time to determine the moisture content of a batch of coal samples, which cannot meet the actual site requirements. Therefore, it is very important to study how to quickly detect the moisture content of coal.

The near-infrared spectroscopy collects light in the range of 780-2526 nm, and the combined and doubled frequencies of the vibrational frequencies of the C-H, N-H, O-H, C-O, S-H groups in the substance molecules are absorbed [3-4]. The near-infrared analysis of a substance will produce some characteristic wavelengths related to the composition of the substance, and the size of the absorbance is related to the composition of the object to be measured. Therefore, the near-infrared spectroscopy technology is suitable for analyzing and detecting the components related to the above-mentioned groups, the information of the moisture content can be obtained based on the absorption of coal O-H groups by the near-infrared spectroscopy. There are already some researchers have used near-infrared spectroscopy technology to detect coal moisture, in 2009, the coal quality indicators were analyzed and discussed by using near-infrared spectroscopy, and the best modeling band was selected to realize the online detection of moisture [5]; in 2016, the forward interval partial least squares was applied to the near-infrared spectroscopy detection of bituminous coal moisture to reduce the difficulty of modeling and improve the accuracy of prediction [6]. The current near-infrared spectroscopy research on coal moisture is mainly concentrated on a single coal sample, and no in-depth research has been carried out on the simultaneous detection of moisture in multiple coal types, in addition, the classification of sample sets is not provided for optimization rules, leading to a lack of representativeness in the selection of sample sets, directly affect the accuracy of the model.

In this paper, near infrared spectroscopy was used to detect moisture in coal. Firstly, abnormal samples were removed by calculating Mahalanobis distance; then, on the basis of classification of different sample sets, combined with different spectral pretreatment, multiple linear regression (MLR) [7] and partial least square (PLS) [8] were used to establish the moisture models. Based on the above-mentioned optimal sample classification, the non-linear BP neural network [9] was used to continue modeling and forecasting of moisture. The above-mentioned different model evaluation indicators were analyzed to find the establishment of the optimal model.

2. Materials and methods

2.1. Experimental materials

The 29 coal samples utilized in the experiment came from the Test Center of China Coal Research Institute and Shandong Metallurgical Research Institute. These samples contain bituminous coal and anthracite, different coal samples in bituminous coal and anthracite have corresponding reference material numbers and batches. This series of standard materials took pre-selected coal samples, after natural drying and crushing, all passed through a 0.2 mm (80 mesh) sieve, with a particle size of less than 0.2 mm. Under dry basis, the distribution of oxygen-carbon ratio and hydrogen-carbon ratio of each coal sample is shown in Figure 1.

The method of artificial humidification was used for each of the original coal samples mentioned above, and the moisture content of each coal sample was roughly divided into 8 intervals: 0-5%, 5-10%, 10-15%, 15-20%, 20-25%, 25-30%, 30-35%, 35-40%.

2.2. Experimental method

2.2.1. Determination of moisture calibration value. According to the three methods provided in GB/T 212-2008 "Proximate analysis of coal", because the coal sample in the experiment contains bituminous coal and anthracite, the method B-air drying method was selected to measure the moisture content in coal samples [2].
2.2.2. Spectral collection. Near-infrared spectroscopy was collected on 231 coal samples with moisture content ranging from 0–40% in different intervals. The detection principle of the near infrared spectrum acquisition instrument is diffuse reflection spectrum sample measurement, the wavelength acquisition range is 900-1700 nm, the data sampling interval is 6.4 nm, the scanning speed is 5 times/second, and each near infrared spectrum data contains 127 data points. In the experiment, each test sample to be tested was measured 5 times continuously, and all were saved as the original spectrum of the corresponding sample, as shown in Figure 2.

Figure 1. Coal sample element ratio distribution.

Figure 2. Original spectrum of coal sample.

Figure 3. Different preprocessing of the original spectrum: (a) Second-order; (b) MSC + second-order; (c) SNV + second-order; (d) Smoothing + second-order.
3. Results and discussion

3.1. Spectral data preprocessing
In the laboratory, when the coal sample to be tested is collected for near-infrared spectroscopy, there are usually some interference factors, so that the measured spectrum not only contains the information of the quality index of the coal sample to be tested, but also contains some abnormal interference signals. The quality and spectral information of these spectra will be affected by interference signals, which will further affect the establishment of the calibration model and the prediction of the composition of the sample to be tested.

Spectral preprocessing methods mainly include spectral movement smoothing, second-order derivation, multiplicative scatter correction (MSC), and standard normal variate (SNV) [10-12].

In order to find the characteristic peaks of the near-infrared spectrum, Mahalanobis distance was calculated on the original near-infrared spectrum data, and after the abnormal data was removed, the results of multiple spectral preprocessing are shown in Figure 3, second-order derivation (second-order) (a), MSC combined with second-order derivation (MSC + Second-order) (b), smoothing processing combined with second-order derivation (smoothing + second-order) (c), SNV combined with second-order derivation (SNV + second-order) (d), at the same time, the correlation coefficient method was also used to find the characteristic peaks of the near-infrared spectrum (Figure 4), besides, the wavelength absorption characteristic information of the full spectrum was utilized.

![Figure 4](image)

**Figure 4.** Correlation between spectrum and coal moisture content.

3.2. Selection of calibration set and prediction set
In the process of spectrum analysis, the selection of sample calibration set and prediction set is necessary before establishing the spectrum multivariate calibration model, this is an important part of spectrum analysis and directly affects the accuracy of the calibration model built. The spectrum correction set is used to establish the spectrum model, and the prediction set is used to verify the prediction performance of the built spectrum.

The commonly used methods for dividing sample set are as follows: random method (RS), Kennard-Stone (KS) method, content gradient method (Rank), Rank-KS method, etc. The RS method is quite contingent, and it is impossible to be sure that the selected sample is representative; the KS method is based on putting the sample data with large spectral differences into the calibration set, and the remaining samples are the prediction sets, When the moisture content difference is small and the spectral changes between the samples are not large, the sample selected by this method is less representative at this time; the Rank method sorts the water content from small to large or large to small, and then samples taken at a certain interval as a calibration set, this method ignores the difference in the spectrum itself and only focuses on the properties of the sample; the Rank-KS method, which combines the Rank and KS method, the first step is the "Rank" part, which refers to the
idea of the content gradient method, ranks the moisture values to be studied, and then divides the entire property interval into n parts, the second step is the "KS" part, in each cell, the KS method is used to select representative samples in the spectral spatial distribution as the correction set, and determine the sample selection of the prediction set [13, 14]. This paper mainly focuses on the random method (RS), the Kennard-Stone (KS) method, and the Rank-KS method to carry out the corresponding coal sample moisture content modeling analysis and comparison.

The corresponding Mahalanobis distance was calculated according to the collected spectral data, and the abnormal points were eliminated by Mahalanobis distance, the number of correction sets was 860 and prediction sets was 291 according to different sample division methods. When the samples were divided by random method, in order to reduce the contingency caused by the random method of sample division, the method of generating random numbers was randomly divided 10 times, and the final evaluation was based on the average value of the data analysis divided by 10 random samples.

3.3. Model establishment and evaluation

The correction set and prediction set are selected by random method (RS), Kennard stone (KS) method and Rank-KS method respectively, then after corresponding spectral preprocessing and feature peak selection, MLR and PLS are used to model and evaluate the sample sets in different sample classification methods. The root mean square error of prediction (RMSEP) is taken as the evaluation index of the mathematical model of moisture content, the smaller the value of RMSEP, the better the prediction effect of the model, which can be used to evaluate the pros and cons of the models established by different sample classification methods and different spectral preprocessing [13-15]. Under the condition of judging the optimal classification method, the non-linear BP neural network is used to continue to model and analyze the moisture content.

If the wavelength related to the moisture content of coal in the near-infrared spectrum can be found for modeling, the prediction accuracy of moisture content modeling can be improved. In this experiment, after the Mahalanobis distance was used to eliminate the abnormal data of near-infrared spectra, firstly, according to the preprocessing method of each spectrum, the band with the characteristic peak of the spectrum was selected for the moisture modeling of MLR, then according to the correlation between each wavelength of the spectrum and moisture content, the band selection was carried out to establish the MLR moisture model, finally, MLR and PLS modeling were performed in the full spectral band. The prediction model of coal moisture content is shown in Table 1.

Table 1. Modeling results of coal moisture content.

| Modeling algorithm | Spectral processing method          | RS classification | K-S classification | Rank-KS classification |
|-------------------|------------------------------------|-------------------|--------------------|-----------------------|
|                   |                                     | RMSEP             | RMSEP              | RMSEP                 |
| MLR               | Second order derivation            | 9.5792            | 9.9948             | 9.4147                |
| MLR               | MSC+ Second-order                  | 9.6616            | 10.0645            | 9.4411                |
| MLR               | smoothing + second-order           | 9.5584            | 10.2036            | 9.4031                |
| MLR               | SNV + second-order                 | 9.4950            | 9.9694             | 9.3778                |
| MLR               | Correlation coefficient method      | 9.9011            | 10.4316            | 9.6360                |
| MLR               | Full spectrum analysis             | 10.2324           | 10.6999            | 10.2228               |
| PLS               | Full spectrum analysis             | 9.4412            | 10.0127            | 9.3213                |

In the RS classification, the PLS moisture prediction model established by full spectrum analysis is better, with an RMSEP value of 9.4412. In the K-S classification, the SNV+ second-order moisture prediction model is better, with an RMSEP value of 9.9694. In the Rank-KS classification, the PLS moisture prediction model based on full spectrum analysis is better, with an RMSEP value of 9.3213.

In each spectrum preprocessing method, compare the value of RMSEP in the three sample classification methods, comparing the data in Table 1, it can be seen that the RMSEP value is the
smallest in the Rank-KS method of sample classification, the modeling effect is the best, and its model prediction ability is significantly improved. In the K-S classification, the value of RMSEP is the largest and the modeling effect is the worst. In addition, in the Rank-KS classification of the samples, the PLS moisture prediction model established by full spectrum analysis is the best, and the RMSEP value is the smallest.

From the above modeling analysis, it can be concluded that each moisture prediction model is better when using Rank-KS classification, which illustrates the feasibility of using rank-KS to classify samples. The non-linear BP neural network was used to further model and analyze the moisture prediction, when the first layer neuron 18 and the second layer neuron 5 of the double hidden layer were set, the node transfer functions were tansig, tansig, purelin respectively, and the training function was trainlm, the RMSEP value is the best.

| Spectral processing method | Double hidden layer | Node transfer function | Training function | Rank-KS classification | RMSEP |
|----------------------------|---------------------|------------------------|-------------------|-----------------------|-------|
| BP neural network          | 18                  | 5                      | tansig            | tansig                | purelin| trainlm | 8.7558 |

According to the comparative analysis in Table 1 and Table 2, for the full-spectrum feature information of the near-infrared spectrum, the coal moisture content prediction model established by the BP neural network algorithm has an RMSEP value of 8.7558, which has the best prediction effect on the coal moisture content. The prediction results of coal moisture content in the correction set and prediction set are shown in Figure 5:

![Figure 5. Optimal treatment results for moisture content.](image)

4. Conclusions
In order to study the moisture analysis model in coal by near infrared spectroscopy, the collected 29 kinds of coal samples with different properties were classified by different sample classification methods, and the near-infrared spectrum data were preprocessed, or the full spectral characteristic information was directly adopted. Then, MLR, PLS and BP neural network algorithms were utilized to model and analyze the near infrared spectrum. The following conclusions are drawn:

(1) This paper provides a Rank-KS method for selecting samples, the method considers the distribution of coal samples in spectral space and property space to achieve the purpose of selecting coal samples, the sample set selected according to this method can significantly improve the ability of the prediction model.
(2) For the samples with different moisture content in this experiment, the sample set established by the Rank-KS algorithm, both spectral preprocessing and full spectral characteristic information, the coal sample moisture prediction models established by MLR and PLS can significantly improve the moisture prediction capabilities of their models.

(3) For the full-spectrum feature information of the near-infrared spectroscopy data, the coal moisture prediction model established by the BP neural network algorithm has the best effect, and the RMSEP is the smallest.

(4) In order to further improve the effect of measuring moisture in coal based on near-infrared spectroscopy, quantitative analysis models should be established for different coal samples.

Acknowledgement

This work was supported by the Natural Science Foundation of China (Grant Number 51776070).

References

[1] Jia Hao, Fu Qiang, Han Chuanjuan, et al. 2012 Improving Precision in Coal Moisture Detection Using Wavelet Transform[J]. *Spectroscopy and Spectral Analysis* 32(11) 3010-3013

[2] GB/T 212-2008, Proximate analysis of coal[S].

[3] Ma Gongzhe, Yang Wenli, Wang Wenchao, et al. 2015 NIR Detecting of Bituminous Coal Moisture Based on Main Components Analysis and Partial Least Squares Regression Analysis[J]. *Yunnan Chemical Technology* 42(01) 45-47

[4] Zhang Shujuan, Zhang Yugui, Li Donghui, et al. 2021 Near-Infrared Spectral Analysis of Licorice and Its Different Gun Products[J]. *Chemical World* 62(10) 618-622

[5] Dong Won Kim, Lee Jong Min and Kim Jae Sung 2009 Application of near infrared diffuse reflectance spectroscopy for on-line measurement of coal properties[J]. *Korean Journal of Chemical Engineering* 26(2) 489-495

[6] Yang Xiaoli, Ma Gongzhe, Chen Yunxiu, et al. 2016 Determination of Bituminous Coal Moisture Based on FiPLS[J]. *Guangzhou Chemical Industry* 44(04) 26-28

[7] Zheng Yuqiao, Wei Jianfeng, Zhu Kai, et al. 2021 Fault Monitoring Method of Wind Turbine Main Bearing[J]. *Journal of Vibration, Measurement & Diagnosis* 41(02) 341-347+415

[8] Huang Jianjun, Li Xuemei and Teng Hongquan 2021 Evaluation Model of Loess Collapsibility Based on the Partial Least Squares Method[J]. *Journal of Catastrophology* 36(02) 60-64

[9] Bai Jiliang, Li Bin, Zhu Jinqi, et al. 2020 Modeling of carbon content in fly ash of CFB boiler based on BP neural network[J]. *Clean Coal Technology* 26(51) 212-217

[10] Cheng Fang, Fan Yuxia and Liao Yitao 2012 Qualitative and Quantitative Detection of Minced Pork Quality by Near Infrared Reflectance Spectroscopy[J]. *Spectroscopy and Spectral Analysis* 32(02) 354-359

[11] Chen Pengqiang, Lu Huishan and Yan Hongwei 2013 Research of quantitative detection of coal samples based on near infrared spectrum[J]. *Industry and Mine Automation* 39(08) 68-71

[12] Shangguan Haibing 2016 Pretreatment Method of Near-Infrared Spectroscopy Used for Different Coal Partical Size[J]. *Coal Technology* 35(01) 308-309

[13] Chen Yiyun, Zhao Ruoing, Qi Tianci, et al. 2017 Constructing Representative Calibration Dataset Based on Spectral Transformation and Kennard-Stone Algorithm for VNIR Modeling of Soil Total Nitrogen in Paddy Soil[J]. *Spectroscopy and Spectral Analysis* 37(07) 2133-2139

[14] Liu Wei, Zhao Zhong, Yuan Hongfu, et al. 2014 An Optimal Selection Method of Samples of Calibration Set and Validation Set for Spectral Multivariate Analysis[J]. *Spectroscopy and Spectral Analysis* 34(04) 947-951

[15] Li Jie, Li Shangke, Jiang Liwen, et al. 2020 A Nondestructive Method Identifying Varieties of Green Tea Based on Near Infrared Spectroscopy and Chemometrics[J]. *Journal of Instrumental Analysis* 39(11) 1344-1350