JAAS

ROYAL SOCIETY OF CHEMISTRY

PAPER

Cite this: *J. Anal. At. Spectrom.*, 2023, **38**, 2046

Development and application of a rapid coal calorific value analyzer based on NIRS-XRF

Rui Gao,†[ab] Shuqing Wang,†[c] Jiaxuan Li,[ab] Zhihui Tian,[ab] Yan Zhang,[d] Lei Zhang, *[ab] Zefu Ye,[e] Zhujun Zhu,[e] Wangbao Yin*[ab] and Suotang Jia[ab]

Accurately and timely determining the calorific value of coal is essential for optimizing the washing and processing procedures in coal preparation plants and achieving maximum economic benefits. In this study, a rapid coal calorific value analyzer was developed by combining near-infrared spectroscopy (NIRS) and X-ray fluorescence spectroscopy (XRF), and industrial testing and performance evaluation were conducted at a coal preparation plant. This article focuses on the structure, operational process, analysis model, and industrial testing of a rapid NIRS-XRF coal calorific value analyzer. This analyzer consists of six parts: NIRS module, XRF module, sample delivery module, control module, hydrogen production module, and operation software. We proposed a holistic-segmented model quantitative analysis algorithm based on partial least squares regression (PLSR). The analyzer was used to measure the calorific values of four types of product coals, namely, clean coal, middling coal, slime and gangue, from the daily production of the Duanshi coal preparation plant, and compared with the assay results. The test results showed that the root mean square error of prediction (RMSEP), the average absolute error (AAE) and the average relative error (ARE) of the prediction of the coal calorific value by using the holistic-segmented model decreased from 0.65 MJ kg^{-1}, 0.55 MJ kg^{-1} and 4.71% of the traditional holistic model to an average of 0.33 MJ kg^{-1}, 0.28 MJ kg^{-1} and 2.71%, respectively, a decrease of nearly double. The average standard deviation (SD) also decreased from 0.29 MJ kg^{-1} to 0.09 MJ kg^{-1}, which met the Chinese national standard requirement of less than 0.12 MJ kg^{-1}. Our proposed model significantly improved the accuracy and repeatability of the measurements. Furthermore, the measurement results of this analyzer showed good consistency with traditional chemical analysis results, indicating its potential for widespread application in industries such as coal mining, washing, power generation, and coking, among others. Therefore, the rapid NIRS-XRF coal calorific value analyzer provides a new intelligent means for the clean and efficient utilization of coal.

Received 15th June 2023
Accepted 29th August 2023

DOI: 10.1039/d3ja00197k

rsc.li/jaas

1 Introduction

The calorific value[1] is a crucial index that reflects various characteristics of coal. Not only does it determine the degree of coalification but it also serves as an important basis for coal classification and scientific and reasonable pricing.[2] In coal preparation plants, the measurement of the coal calorific value is a key task of coal quality analysis. Based on the obtained calorific value data, guidance and suggestions can be provided for processes such as coal washing and processing, precise coal blending, and sales judgment. To effectively guide the production work of coal preparation plants, fully utilize coal resources, and improve the economic benefits, it is essential to ensure the accuracy and speed of the measurement of the coal calorific value.

At present, Chinese coal preparation plants generally use the chemical analysis method specified in Chinese national standards – the oxygen bomb combustion method[3,4] to measure the coal calorific value. This method requires putting a unit mass of coal sample into the bomb of the calorimeter, burning it in excess high-pressure oxygen and inert gas (usually nitrogen), and calculating the calorific value by measuring the temperature rise of water. Unfortunately, this method is relatively cumbersome, requiring numerous steps and taking a long time from sampling to obtaining measurement results. As a result, coal quality data lags behind the real-time production guidance requirements of coal preparation plants, such as intelligent control of coal-heavy medium sorting and precise coal blending, which causes the process adjustment to lag behind

[a]*State Key Laboratory of Quantum Optics and Quantum Optics Devices, Institute of Laser Spectroscopy, Shanxi University, Taiyuan, China. E-mail: k1226@sxu.edu.cn; ywb65@sxu.edu.cn*

[b]*Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, China*

[c]*SINOPEC Research Institute of Petroleum Processing Co., Ltd, Beijing, China*

[d]*School of Optoelectronic Engineering, Xi'an Technological University, Xian, China*

[e]*Shanxi Gemeng US-China Clean Energy R&D Center Co., Ltd, Taiyuan, China*

† These authors contributed equally to this work.

and seriously restricts the assessment of contract coal in coal preparation plants. With the reform of intelligent construction in coal preparation plants and the continuous improvement of automation levels in coal blending and other technologies, there is an urgent need for efficient and accurate coal quality rapid analysis and measurement technology in order to produce high-standard product coal with good coal quality.

Currently, there is online coal quality measurement equipment based on prompt gamma neutron activation analysis (PGNAA)[5,6] technology produced by Thermofisher Company of the United States and ScanTech Company of Australia. This technology utilizes the thermal neutron capture reaction between neutrons and the nuclei of various elements in the coal sample, causing the excited nuclei to release characteristic γ-rays in the process of deexcitation. Coal quality indicators such as the calorific value can be obtained by modeling and analyzing γ-ray energy spectrum data. However, this method involves radioactive hazards, and the equipment and maintenance costs are expensive.[7] The coal analyzer[8] based on laser-induced breakdown spectroscopy (LIBS)[9–11] uses atomic emission spectroscopy technology. The principle underlying this technology is that when a high-energy pulsed laser hits the sample, the atoms, molecules, and other particles in the focused region absorb the photons producing initial free electrons. As more laser energy is applied, more free electrons are created, causing further ionization of the atoms. This results in a rapid increase in electron density, forming a high-temperature plasma consisting of atoms, ions, and free electrons. By analyzing the characteristic spectral lines emitted as the plasma cools down, information about the types and concentrations of elements in the sample can be obtained.[12,13] However, due to the Rayleigh–Taylor instability of the plasma signal and the influence of self-absorption and matrix effects, the accuracy and repeatability of the measurement by this technology are lower[14,15] and cannot accurately reflect changes in coal quality.

Near-infrared spectroscopy (NIRS)[16–19] technology is one of the widely used and relatively advanced technologies in the field of coal quality analysis and detection in recent years. Near-infrared light refers to electromagnetic waves with wavelengths ranging from 780 to 2526 nm (12 500 to 4000 cm^{-1}) between visible light and mid-infrared light, which can cause vibrational energy level transitions of molecules or atoms. When the light in this wavelength region is irradiated onto a substance containing the hydrogen group X–H (X = C, N, O),[20] the chemical bonds within the substance will absorb the characteristic wave of a certain wavelength, thus generating absorption spectra that carry significant information about the substance's chemical composition and structure, physical state, content, and more. The amount of absorbance is directly related to the content of the specific components in the substance, and in the case of coal combustion, the primary components of interest are those containing hydrogen or carbon. It is this relationship that provides the theoretical basis for the practical application of NIRS technology in the rapid, non-destructive, and accurate measurement of the coal calorific value.[21] In recent years, a large number of studies have been conducted on NIRS coal analysis both domestically and internationally. Kim *et al.*[22] selected the absorbance at specific wavelengths in NIRS spectra to analyze the properties of coal with good accuracy for C and H elements. While the measurement repeatability deviation for the calorific value was higher than that of the ISO/ASTM reference method, it still provided fairly good results at the level of semi-quantitative approximate analysis. Xiao *et al.*[23] proposed a DR_TELM deep learning model, which includes dilated convolution, multi-level residual connection and TELM, with root mean square errors (RMSE) of 1.1%, 1.8% and 0.7 MJ kg^{-1} for coal volatile matter, fixed carbon and higher heating value (HHV), respectively. Le *et al.*[24] proposed a CNN-AELM model based on the artificial bee colony algorithm to optimize the performance of CNN-ELM, and visible-infrared spectroscopy was used to predict the low calorific value of coal samples, which showed excellent accuracy. Begum *et al.*[25] first improved the signal-to-noise ratio of NIRS spectra of coal samples using a variety of pretreatment algorithms, and then the partial least squares regression (PLSR), random forest (RF) and extreme gradient boosting (XGBoost) regression models were used to predict volatile matter, fixed carbon, and the gross calorific value (GCV) with improved accuracy. Liu *et al.*[26] established the PSO-ELM algorithm which combines the particle swarm optimization (PSO) algorithm with an optimized extreme learning machine (ELM), improving the prediction accuracy and computing efficiency, and the normalized RMSE and prediction time for calorific value were 0.026 MJ kg^{-1} and 0.002 s, respectively. Usually, NIRS technology is more suitable for analyzing the content of organic matter in coal, while it has a weak response to inorganic compounds, therefore, it is difficult to analyze the content of K, Ca, Na, Mg, Al, Si and other inorganic compounds in coal which are negatively correlated with the calorific value.[27]

X-ray fluorescence spectroscopy (XRF)[28–31] technology, which has good repeatability in coal quality analysis, is widely used for measuring inorganic elements. The theoretical basis of XRF analysis is Moseley's law, which states that the frequency of characteristic X-rays of elements is linearly related to the square of the atomic number, and the intensity of characteristic X-rays is positively correlated with the content of elements. When primary X-rays produced by an X-ray tube are incident on the sample, the electrons in the inner shell of each element are excited out of their orbitals and produce holes. This causes the entire atomic system to become unstable and existing electrons at higher energy levels spontaneously transition to the inner layer to fill the holes. This process causes the atom to relax to a lower energy state while radiating characteristic X-rays with a specific energy level difference, known as secondary X-rays or X-ray fluorescence. By analyzing the wavelength (or energy) and intensity of the X-ray fluorescence produced by different elements in the sample, information on the elemental composition and content of the sample can be obtained for qualitative and quantitative analysis.[32] For example, Wen *et al.*[33] used XRF and random forest regression (RFR) to predict coal ash content with RMSE, MAE and R^2 of 1.5423, 1.1136 and 0.9896, respectively. Ward *et al.*[34] used a hand-held energy dispersive X-ray fluorometer (ED-XRF) to measure the content of several main inorganic elements in coal blocks, and the results were basically

consistent with the nominal data. Zhang *et al.*[35] developed a real-time ash analyzer based on pseudo-dual energy X-ray transmission, and the absolute error of measurement was within 1% for clean coal with about 10% ash content, for which the probability of absolute error less than 0.5% was higher than 85%. Guo *et al.*[36] used XRF and PLSR to model and predict the ash content of 45 standard coal samples, and the R^2 of the calibration curve obtained reached 0.946 and the predicted RMSEP was 1.177%. Zhang *et al.*[37] measured heavy metals such as Hg, Pb, Cr and other elements such as Ti, Fe and Ca in pulverized coal using an online XRF analyzer, and the relative standard deviations (RSDs) obtained from the measurements were all less than 7.74%. Although XRF is effective for the measurement of ash-forming elements and ash content in coal, its fluorescence yield for light elements is low, and it is usually not suitable for analyzing elements[38] with atomic numbers less than 11, so it cannot be used to analyze the content of organic elements such as C and H that are positively related to the calorific value[39] of coal.

To sum up, NIRS is well suited for the highly stable measurement of the organic components in coal, while XRF is proficient at measuring the inorganic components in coal with high stability. By combining the two technologies, a fused spectrum containing comprehensive and rich chemical information of coal can be obtained, which can realize the accurate analysis of the coal calorific value. In this paper, we developed a rapid analyzer for quantitative analysis of coal calorific value based on NIRS-XRF fusion spectroscopic technology, and the industrial field testing and performance evaluation were carried out in a coal preparation plant, and the results showed that it can provide fast and accurate measurement of the calorific value of coal. Our work offers a new technology and instrument for the coal machinery industry to measure the coal calorific value quickly and accurately.

2 Experimental

2.1 Instrumentation

Fig. 1(a) and (b) show the physical photo and schematic diagram of this rapid coal calorific value analyzer based on NIRS-XRF, respectively. The external dimension of the instrument is 1200 mm × 1200 mm × 1800 mm, and the weight is 140 kg, which consists of six parts: NIRS module, XRF module, sample delivery module, control module, hydrogen production module and operation software. The interior of the instrument is divided into three layers, with the top layer containing the NIRS module and XRF module encapsulated in a dustproof aluminum shell equipped with vibration-damping rubber gaskets, as well as the sample delivery module; the middle layer with the control module; and the bottom layer with the hydrogen production module. The following is a detailed description of each.

2.1.1 NIRS module. The optical structure of this module is shown in Fig. 2, mainly composed of a near-infrared light source, near-infrared spectrometer, reflection probe, reflection probe bracket, optical fiber, reference tile, *etc.* Among them, the near-infrared light source (AvaLight-HAL-S Mini, Avantes) is a halogen lamp with a wavelength coverage range of 360–2500 nm, and its color temperature is 2700 K when operating in low power mode with a lifetime of over 13 000 hours. The near-infrared spectrometer (Hamamatsu, C15511-01) is a Fourier transform type, which is integrated with a miniature Michelson spectral interferometer and control circuit. The effective operating band is 1100–2500 nm, the signal-to-noise ratio is 10 000 : 1, the spectral resolution is 5.7 nm, and the spectral reproducibility is ±0.5 nm. The reflection probe (FCR-UVIR400-2-ME, Avantes) is a fiber bundle consisting of seven 200 μm core diameter fibers, six for illumination and one for detection, which is fixed to the bottom of the aluminum shell with a bracket. The diffuse reflection reference tile (WS-2, Avantes) is a type of diffuse reflection plastic white adhesive (PTFE) with an ideal Lambert surface that can be used as a standard reference for reflection experiments, and its diffuse reflection spectrum is used to calculate the absorbance spectra of the samples to be measured.

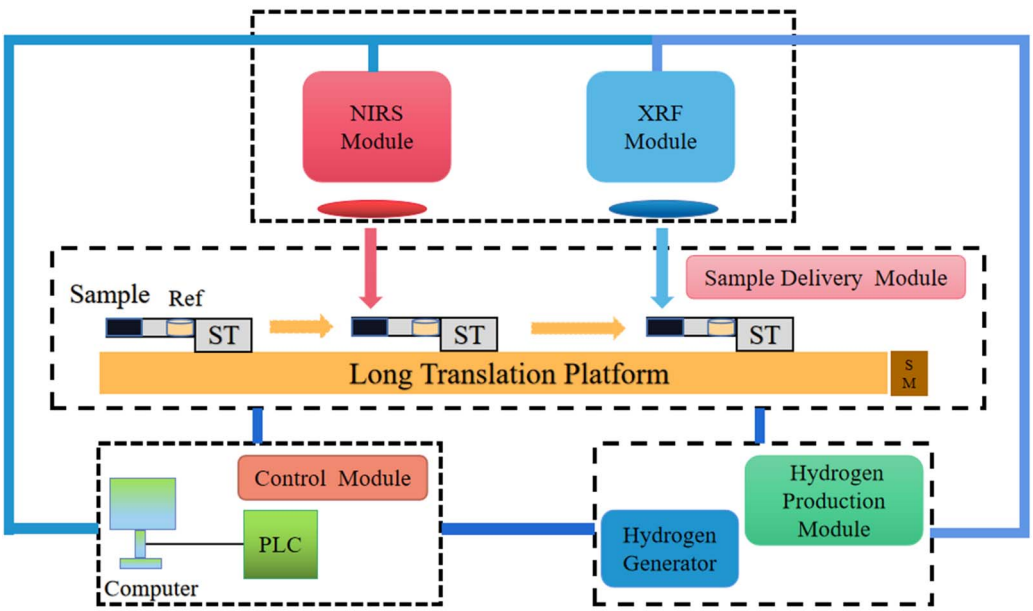
When the module is working, the light emitted by the near-infrared light source is coupled into the six illumination fibers of the fiber bundle and irradiates to the surface of the coal sample by the end of the reflection probe, and the diffuse reflected light is collected by the detection fiber and transmitted to the near-infrared spectrometer for NIRS spectra detection. The obtained spectral signals are then transmitted to the computer *via* the USB data cable. The number of scans of the near-infrared spectrometer in this instrument is set to 4096, with 359 scanning points and a scanning time of 15 seconds.

2.1.2 XRF module. This module adopts the ED-XRF optical structure as shown in Fig. 3, which is mainly composed of an X-ray tube (VF-50J, VARIAN), silicon drift detector (VIAMP, KETEK), digital pulse processor (DPP), high-voltage power supply (MNX50P50, SPELLMAN), chamber, beryllium window, collimator, *etc.* Among them, the X-ray tube with a collimator and SDD is placed on both sides of the chamber at an angle of 45°. The X-ray tube is powered by a high-voltage power supply, and the output voltage signals from the SDD detector are transmitted to the DPP for digital processing and analysis. The diameter of the beryllium window is 10 mm and the thickness is 20 μm, and it is 2 mm away from the surface of the sample. The inlet of the chamber is connected to the hydrogen production module, and the outlet of the chamber is used to discharge the gas from the chamber.

When the module is working, the filament inside the X-ray tube emits electrons under the action of the high-voltage stable current. These electrons are accelerated to the anode and interact with the atoms of the anode target material to generate primary X-rays. The primary X-rays are transmitted through the collimator and beryllium window and reach the sample. Atoms of each element with low electron binding energy in the sample are excited to generate secondary X-ray fluorescence, which is collected by the SDD detector after passing through the beryllium window. The collected signals are then outputted to the computer through the DPP processor for further analysis.

In this experiment, the ED-XRF spectrometer collected XRF spectra. The module used a 50 W Rh target-anode X-ray tube as the excitation source, and the tube voltage was set to 16 kV, the

(a) Physical photo.



(b) Schematic diagram (Ref: reference tile, ST: sample table, SM: stepper motor, PLC: programmable logic controller).

Fig. 1 Rapid coal calorific value analyzer based on NIRS-XRF.

tube current to 0.6 mA, and the filament voltage and current to 1.5 V and 2.5 A, respectively. The SDD detector peak time was set to 0.1 μs.

2.1.3 Sample delivery module. This module is located below the NIRS and XRF modules, and its 3D models are shown in Fig. 4. Among them, (a) for the overall structure, (b) for the

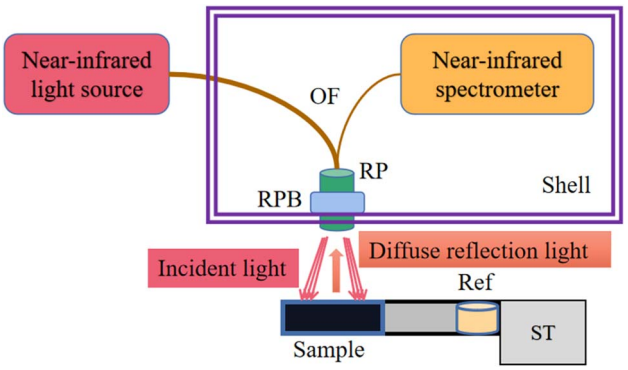


Fig. 2 Schematic diagram of the optical structure of the NIRS module (OF: optical fiber, RP: reflection probe, RPB: reflection probe bracket).

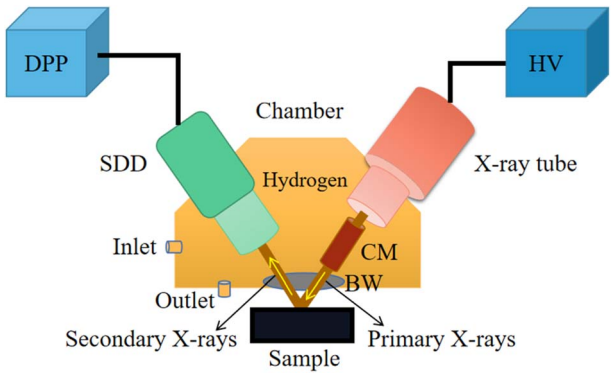


Fig. 3 Schematic diagram of the optical structure of the XRF module (SDD: silicon drift detector, DPP: digital pulse processor, HV: high-voltage power supply, BW: beryllium window, CM: collimator).

electrically controlled long translation platform, (c) for the electrically controlled short translation platform, electrically controlled lifting platform and electrically controlled rotary platform from bottom to top. The long and short translation platforms are both double track ball screw linear guide slide platforms, with effective translation lengths of 600 mm and 400 mm respectively, and a screw rod accuracy of 0.03 mm. The long translation platform is equipped with a sample table for placing the sample cell, which is used to transport the samples to be measured. The short translation platform is fixed with a lifting platform and a rotary platform for regular measurement and calibration of samples. The lifting platform adopts a shear-type lifting support structure, driven by a precision grinding screw rod, with a repeated positioning accuracy of 10 μm, and is used to control the distance between the sample and the beryllium window. The rotary platform is driven by a worm gear, with 6 evenly distributed grooves on top for placing 6 calibration samples. The 3D model of the sample cell is shown in Fig. 5, with an internal groove for placing coal samples, with a groove size of 100 mm × 10 mm × 3 mm, which can hold approximately 2 g of flattened coal samples.

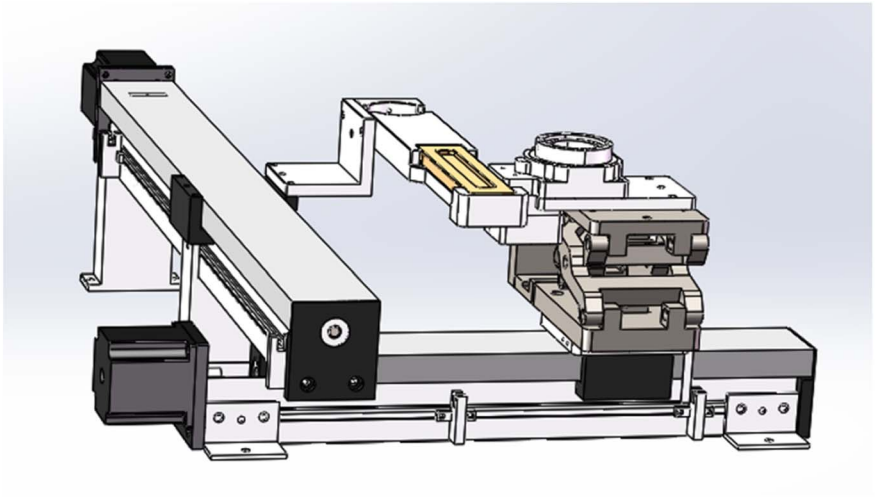
2.1.4 Control module. This module mainly consists of a computer and programmable logic controller (PLC, S7-200CN, SIEMENS), and auxiliary components include solid-state relays, drivers, photoelectric switches, an emergency stop switch, *etc.* Among them, the computer is responsible for communicating with the PLC, processing spectral data, and running the operation software. The PLC is used for sequence control during the operation of the instrument. Solid-state relays are utilized to control the on-off of the near-infrared light source, hydrogen generator, and status indicator light. Drivers, meanwhile, receive pulse signals from the PLC and drive the stepper motor, while photoelectric switches provide position information about the sample to the PLC. The emergency stop switch is used to manually stop the operation of the instrument in an emergency.

2.1.5 Hydrogen production module. This module consists of a hydrogen generator (SFH-300, SFMIT), water tank, float level switch, water pump, *etc.* The hydrogen generator (425 mm × 210 mm × 365 mm) can generate hydrogen gas with a purity of 99.99% at a flow rate of 150 ml min^{-1} and continuously supply the chamber of the XRF module to avoid the absorption of low-energy X-ray fluorescence signals by air. The hydrogen generator integrates liquid storage, electrolytic hydrogen production and oxygen discharge, with low electrolytic cell temperature and long life, and only distilled water needs to be replenished for daily maintenance for continuous use. The float level switch is installed in the hydrogen generator's electrolytic cell. When the water level in the cell falls below the limit, the water pump automatically starts to extract distilled water from the water tank for replenishment.

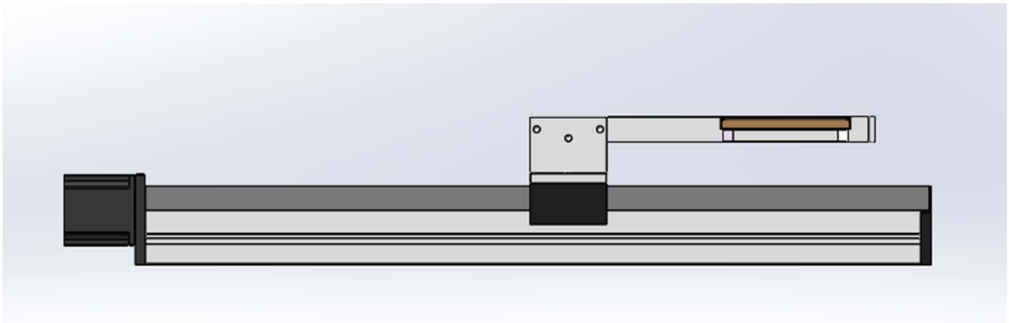
2.1.6 Operation software. The complete operation software for the instrument was written using LabVIEW, and the human–machine interface is shown in Fig. 6. Under the main page column, the user can fill in sample information, select measurement methods, view the measurement process and progress, and display calorific value measurement results. The real-time spectrogram column can display the NIRS and XRF spectrograms of the measured samples in real time. The history column can view past measurement results. The status parameter column displays real-time information on the temperature, voltage, and current of the instrument's internal components.

2.2 Workflow

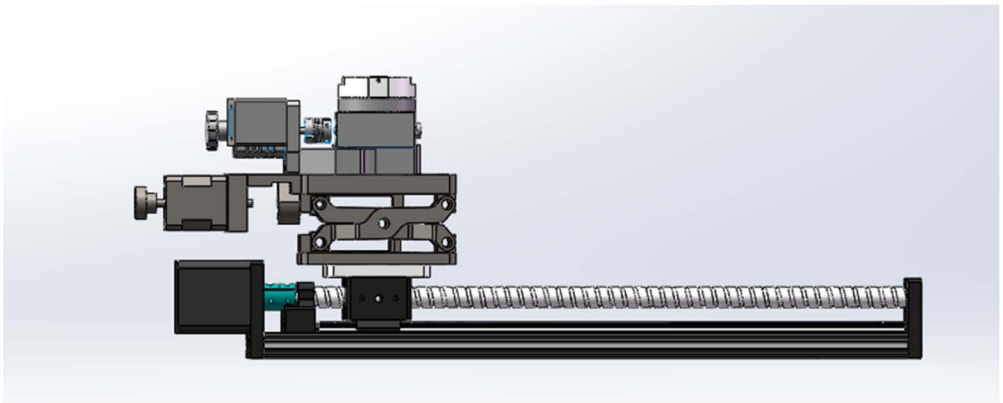
The overall operation process of the rapid NIRS-XRF analyzer can be briefly described as follows. First, after starting the instrument, initialization and preheating are performed, and the sample table is extended from the inside of the instrument to the initial position, while the hydrogen generator and near-infrared light source are turned on, and the chamber is filled with hydrogen after 6 min. In advance, place the coal sample to be measured in the sample cell, scrape it flat, and place it on the sample table. Then, click the measurement button to start the measurement. During the measurement, the sample cell moves below the NIRS module at a speed of 20 mm s^{-1} for measurement, and it takes 15 s to collect the NIRS spectra; the sample cell continues to move below the XRF module at a speed of 20 mm s^{-1} for measurement, and it takes 30 s to acquire XRF spectra. After that, the X-ray tube enters a low-power state, at the



(a) Overall structure.



(b) Electrically controlled long translation platform.



(c) Electrically controlled short translation platform, lifting platform and rotary platform.

Fig. 4 3D models of the sample delivery module.

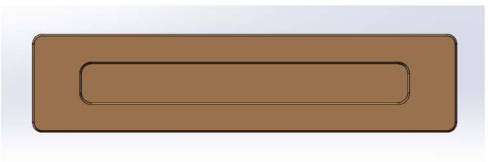


Fig. 5 3D model of the sample cell.

same time, the sample cell returns to its initial position at a speed of 30 mm s^{-1}, while the computer processes the spectral data and displays the analysis result on the software interface. Finally, click the exit button to end the measurement process, the sample table is extended into the instrument, and at the same time, the hydrogen generator, near-infrared light source and other components are turned off. The entire measurement process takes only about 1 min.

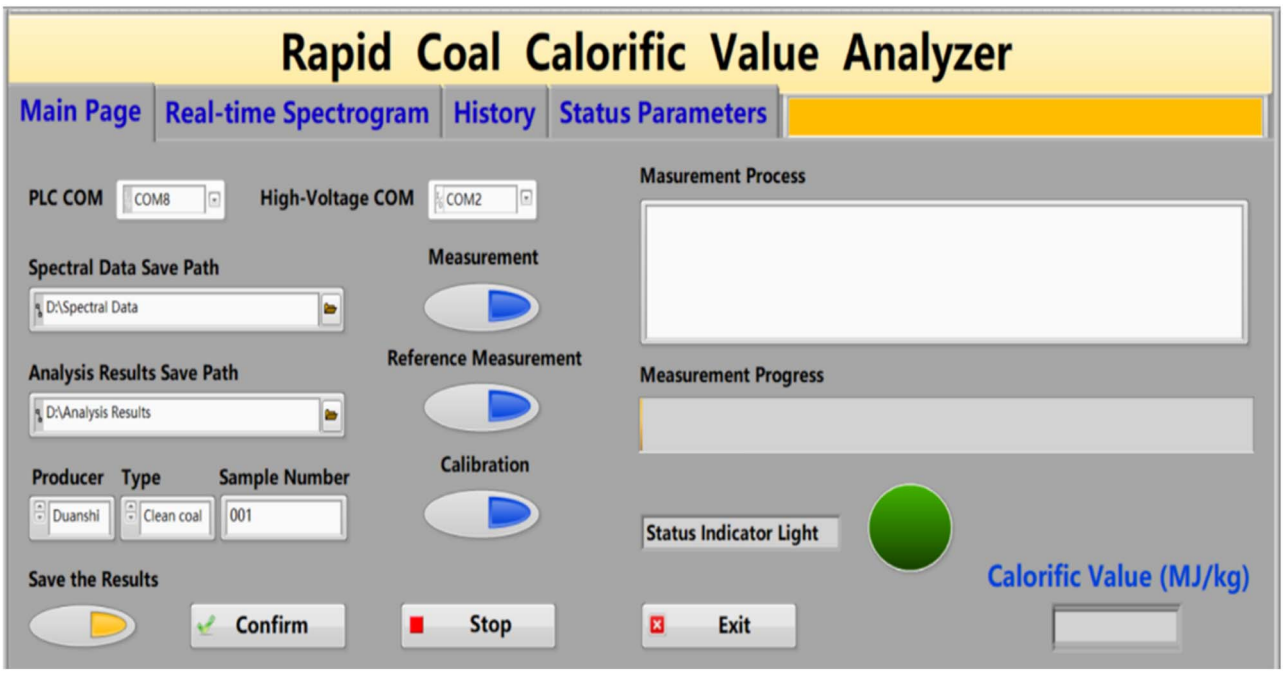


Fig. 6 Human–machine operation interface.

2.3 Samples

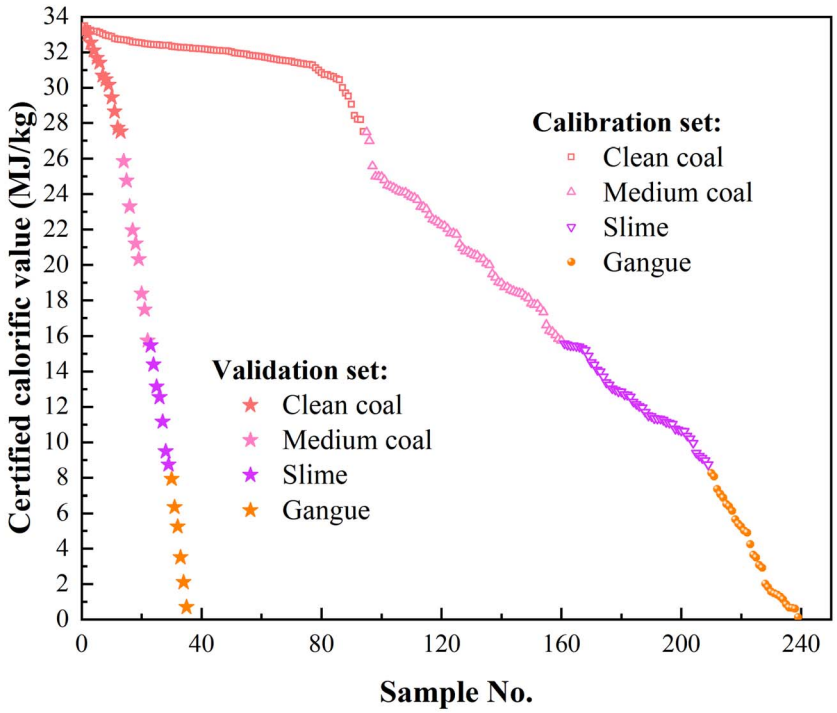
To ensure that the samples collected covered all types of coal produced by the Duanshi coal preparation plant in Qinshui County, Jincheng City, Shanxi Province, a 10-day coal sample collection was conducted on-site. The plant produces four kinds of products: clean coal, medium coal, slime and gangue, with a wide range of calorific value distribution covering 0–34 MJ kg^{-1}. Due to the diversity of coal types and significant differences in coal quality, it is necessary to consider the representativeness of coal sample selection. The coal samples to be selected were divided into calibration set samples and validation set samples. For the calibration set samples, according to the daily coal calorific distribution range produced by the coal preparation plant, a certain number of coal samples were randomly collected from each of the four types of coal, so that the calorific value basically covered the above distribution range and roughly uniform distribution. For the validation set, a certain number of samples were collected from each of the four types of coal, and the calorific value of the collected samples was then compared to the range of calorific value distribution of the calibration set coal to ensure that they covered as much of this range as possible and maintained a uniform distribution. The coal samples collected by the above methods ensure the coverage and uniformity of distribution of the calorific value, so they are widely representative on a large scale. The photo of some coal samples collected is shown in Fig. 7(a).

Due to the large number of samples and the long span of collection time in this experiment, it was necessary to control the conditions of temperature, humidity, and light in the field environment to reduce the influence of the experimental environment on the samples. Therefore, the room temperature was maintained at about 25 °C, the humidity was regulated at about 50%, and the doors and windows were closed and shaded to avoid any external interference. According to the standard procedures for preparing coal samples, all coal samples were crushed, sieved, mixed, downsized, air-dried, and ground to produce air-dried-based pulverized coal samples with an average particle size of 0.2 mm. A total of 274 samples were collected at the end, of which 239 were used as the calibration set for building the prediction model of the calorific value, and the remaining 35 were used as the validation set to evaluate the performance of the prediction model. To verify the measurement repeatability of the established model, 8 coal samples were randomly selected from the validation set and measured 5 times each. The calorific values of all coal samples were measured according to the Chinese national standard GB/T 213-2008, as shown in Fig. 7(b). It can be seen that the calorific values of the calibration set and the validation set of coal samples are basically evenly distributed and representative.

2.4 NIRS-XRF spectra

In the experiment, the near-infrared spectrometer was used to collect the reference spectrum using a reference tile with high reflectivity each time after the instrument was turned on and before the first measurement, and the 4096 spectra of each coal sample were averaged and compared with the reference spectrum as the NIRS spectrum of that coal sample. Fig. 8 shows the typical NIRS spectra of four kinds of coal samples collected, including 10 clean coal, 8 medium coal, 7 slime, and 6 gangue. As shown in the figure, the absorption peaks exist at 2200 nm and 2400 nm for all four types of coal. The absorbance of each type of coal sample shows a decreasing trend with increasing wavelength. In particular, the absorbance of clean coal is the highest, ranging from 1.05 to 1.45, followed by medium coal and

(a) Some of the original samples of clean coal, medium coal, slime and gangue were collected.



(b) Calorific values distribution of coal samples in calibration set and validation set.

Fig. 7 Coal samples and laboratory indicators.

slime, which range from 0.95 to 1.30 and 0.88 to 1.15, respectively, and that of gangue is the lowest, ranging from 0.70 to 1.03.

Fig. 9 shows the typical XRF energy spectra of four kinds of coal samples collected, and the K_α spectral lines of Na, Al, Si, S, K, Ca, Ti, Cr, Mn, Fe, Co and Cu were marked in the energy range from 0 to 9 keV. It can be seen that the intensity of the spectral lines of Si, Fe and Cu in each coal is higher, while that of Na is smaller. This is because Na not only has a low content, but also has a smaller atomic number, and the corresponding fluorescence yield is also lower. In addition, the variability of the intensity of each element in different coal samples is also large. For example, the descending order of spectral line intensities for Al, Si, K and Ti corresponds to the coal types of gangue, slime, medium coal and clean coal, respectively.

2.5 Spectral pretreatment

For NIRS spectra, in order to reduce the interference of invalid spectra and noise, the following pretreatment methods were

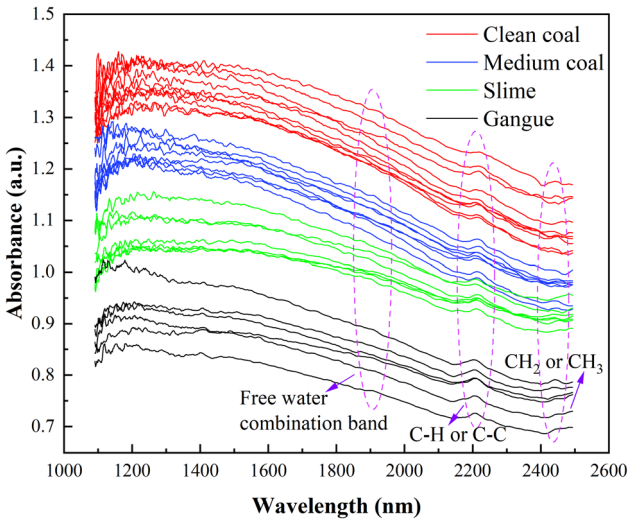


Fig. 8 Comparison of typical NIRS spectra for four types of coal.

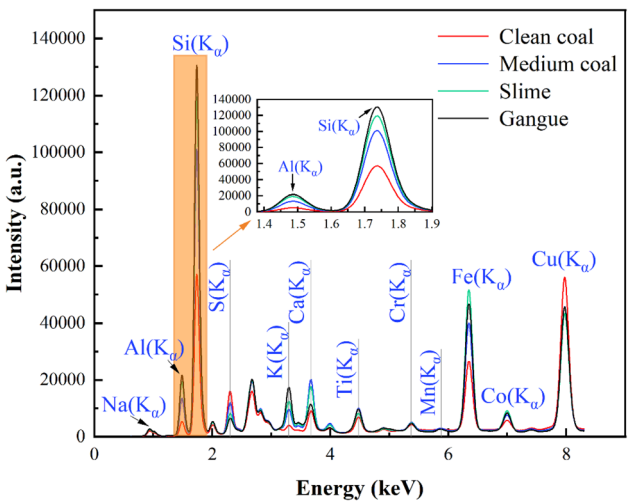


Fig. 9 Comparison of typical XRF energy spectra for four types of coal.

done: the spectral data in the bands of 1000–1089 nm and 2481–2500 nm were removed, leaving the spectral data in the band of 1090–2480 nm (314 variables) to calculate the absorbance from $\log(1/R)$ (R is reflectance); used the Savitzky–Golay (SG) convolutional smoothing algorithm to further reduce noise and eliminate the "spike" phenomenon in the reflection spectrum data of the sample surface, where the window size was set to 5 and the polynomial order was 3; the smoothed spectral data were then standardized using the standard normal variation (SNV) method to eliminate the effects of solid particle size, surface scattering, and optical path changes on NIRS diffuse reflectance spectra. Although the absorbance range of the processed spectral data changed, it did not change the trend of the original spectral data. For the XRF energy spectra, the pretreatment methods were to first reduce the original spectral data range to 0.37–18.52 keV (3425 variables) to eliminate the low energy bands and then to process it using SG smoothing, again using five-point third-order smoothing. The NIRS spectral and XRF energy spectral data were linked end-to-end to form a fused spectrum, and the number of rows of the new matrix was obtained as the number of samples, and the number of columns was the sum of the number of NIRS spectral and XRF energy spectral variables. Due to the large differences in the intensity values between NIRS spectral data and XRF energy spectral data, the fused spectrum needed to be normalized so that the range of NIRS and XRF spectral data was within the range of [−1,1] to enable more accurate comparison and analysis of the spectral characteristics of different samples.

2.6 Predictive model

The fused spectral data and coal calorific value standard values were used to build the calorific value predictive model by PLSR. A holistic-segmented modeling approach was adopted here. The holistic modeling was based on PLSR to establish a model within the entire calorific value range, and the corresponding model was called the holistic model. Considering the large disparity of coal types, a large number of samples, and the obvious differences in the coal matrix, thus showing the potential for classification, segmented modeling was considered to improve the accuracy of model analysis, and the corresponding model was called the segmented model.

2.7 Model evaluation

The basic statistical indicators used to evaluate the measurement accuracy of the built calorific value predictive model include the coefficient of determination (R^2), root mean square error of prediction (RMSEP), average absolute error (AAE) and average relative error (ARE). R^2 reflects the ability of the calibration model to fit sample data, and the closer the value is to 1, the stronger the fitting ability and higher the accuracy of the model. RMSEP reflects the degree of consistency between the predicted and standard values of the validation set, and the closer the value is to 0, the smaller the deviation between the predicted and standard values, and the higher the predictive ability of the built model. AAE is the sum of the absolute values of all errors in the validation set divided by the number of observations, and the closer AAE is to 0, the smaller the error between the predicted and standard values, and the more accurate the predicted values. Similarly, ARE is the sum of the relative values of all errors divided by the number of observations, and the closer ARE is to 0, the smaller the relative error between the predicted and standard values, and the more accurate the predicted values. Using both the AAE and ARE can help in evaluating the accuracy of the predicted values more comprehensively. The definition formulae of the four are as follows:

$$R^2 = 1 - \frac{\sum_{j=1}^{n_c} \left(y_{aj} - y_{bj}\right)^2}{\sum_{j=1}^{n_c} \left(y_{aj} - \bar{y}\right)^2}, \quad (1)$$

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{j=1}^{n_v} \left(y_{cj} - y_{dj}\right)^2}{n_v}}, \quad (2)$$

$$\mathrm{AAE} = \frac{1}{n_v} \sum_{j=1}^{n_v} \left|y_{cj} - y_{dj}\right|, \quad (3)$$

$$\mathrm{ARE} = \frac{1}{n_v} \sum_{j=1}^{n_v} \frac{\left|y_{cj} - y_{dj}\right|}{y_{cj}} \times 100\%, \quad (4)$$

where, y_a and y_b represent the standard and predicted values of the calorific value of the samples in the calibration set, y_c and y_d represent the standard and predicted values of the calorific value of the samples in the validation set, $\bar{y}$ represents the average of the standard values of the calorific value of all samples in the calibration set, j is the sample order, n_c represents the number of samples in the calibration set, and n_v represents the number of samples in the validation set.

Repeatability reflects the degree of consistency between the calorific value results obtained from multiple measurements of the same coal sample under the same conditions. The indicator used here to evaluate the repeatability of the established calorific value model is the standard deviation (SD), and the smaller the SD, the better the repeatability of the model for the measurement of calorific value. The SD is calculated using the following formula:

$$\mathrm{SD} = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}}, \quad (5)$$

where X_i denotes the calorific value obtained from the ith measurement of the same coal sample, $\bar{X}$ denotes the average of all measurements of the same coal sample during repeated measurements, i denotes the order of measurements of the same coal sample, and n denotes the total number of repeated measurements of the same coal sample.

3 Results and discussion

3.1 Holistic-segmented model

The holistic-segmented model based on PLSR is divided into two parts: the holistic model and the segmented model. The holistic model refers to the predictive model established by directly treating the total calorific value of four types of coal as a whole. The comparison between the predicted and standard values of the calorific value of all coal samples using this model is shown in Fig. 10. It can be seen from the figure that R^2, RMSEP, AAE and ARE of the calorific value of the coal sample in the calibration set are 0.998, 0.65 MJ kg^{-1}, 0.55 MJ kg^{-1} and 4.71%, respectively. The accuracy of prediction using the holistic model is not high because the overall number of samples is large, and the properties of clean coal, medium coal, slime and gangue are quite different, and the holistic modeling is easy to cover up the effective spectral information of various coals, ignoring some details, leading to the fuzzy model. In addition, the sample number of slime and gangue is less than

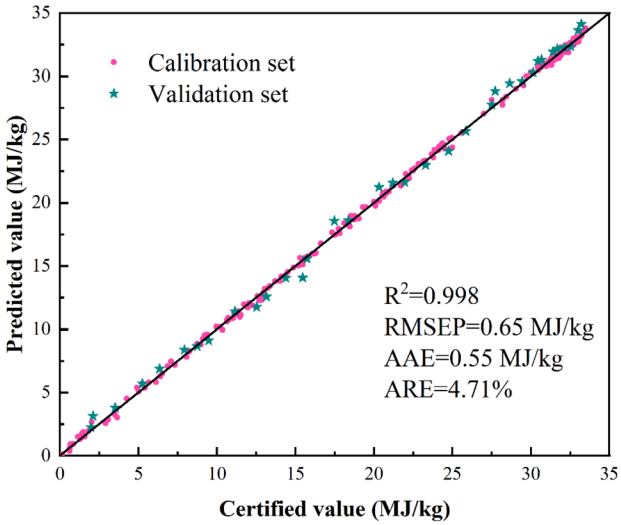


Fig. 10 Comparison between the predicted and standard calorific values of coal samples using the holistic model.

that of clean coal and medium coal, which increases the difference in the calorific value between them and reduces the prediction accuracy of the holistic model. To address this problem, further consideration is given to the use of segmented modeling, based on which more accurate predictions can be made.

Segmented modeling refers to the process of dividing the calorific value range of all coal samples based on the concentrated range of calorific value distribution of gangue, slime, medium coal and clean coal (Fig. 7(b)) into four regions of a (0–8.74 MJ kg^{-1}), b (8.74–15.74 MJ kg^{-1}), c (15.74–27.51 MJ kg^{-1}) and d (27.51–33.50 MJ kg^{-1}), respectively. This classification ensures a higher degree of consistency among the coal types to which each segment belongs, and the corresponding spectral data have a higher homogeneity, which reduces the problem of large differences in the properties of different coal types. As a result, corresponding predictive models can be established for the characteristics of different segments. The holistic-segmented model established is shown in Fig. 11, and the evaluation indicators of the holistic-segmented model and the holistic model are shown in Table 1.

As can be seen from Table 1, the determination coefficients R^2 of both the holistic and holistic-segmented models reach above 0.99; however, the average RMSEP, AAR and ARE of the latter are significantly lower than those of the former, especially for regions c and d. The reason is that these two regions mainly correspond to medium coal and clean coal, which have uniform, stable physicochemical properties, and a larger number of samples, resulting in better predictive performance. The predictive performance of the model for region b is average, but still better than that of the holistic model. The prediction model for region a performs the worst, as gangue contains various impurities and has non-uniform and significantly different components. The evaluation results of the prediction performance show that the average RMSEP, AAE and ARE of the holistic-segmented model decreased from 0.65 MJ kg^{-1}, 0.55 MJ

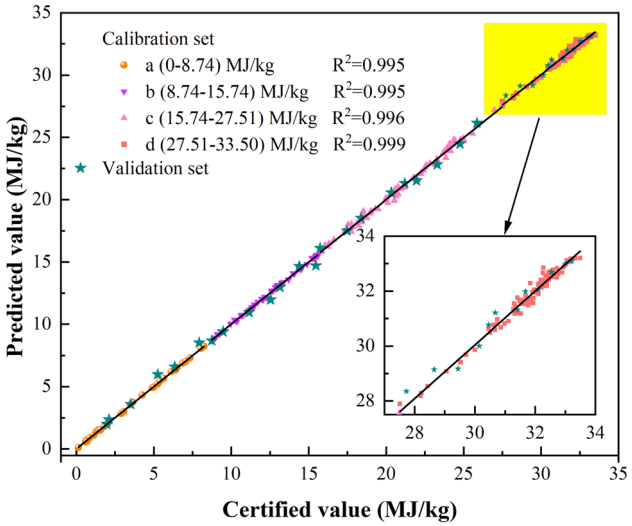


Fig. 11 Comparison between the predicted and standard calorific values of coal samples using the holistic-segmented model.

kg^{-1} and 4.71% of the holistic model to 0.33 MJ kg^{-1}, 0.28 MJ kg^{-1} and 2.71%, respectively, which was almost a times decrease. It can be seen that the segmented modeling method based on PLSR significantly improves the fitting effect of the traditional holistic modeling method and effectively improves the modeling efficiency and measurement accuracy. This AAE has met the requirement of no more than 0.42 MJ kg^{-1} in the coal preparation plant.

When using the holistic-segmented model to predict coal samples, the initial rough value of the calorific value is predicted using the holistic model, then the spectral data are classified into corresponding segmented region based on the discriminant condition, and finally the exact calorific value of the coal sample is calculated using the corresponding segmented model.

For samples with calorific values located in the intersection region, the difference between the predicted calorific values by two adjacent segmented models is around 0.05 MJ kg^{-1}, which can be ignored compared to the average RMSEP of 0.33 MJ kg^{-1} of the holistic-segmented model. For example, there is a coal sample with a standard calorific value of 27.51 MJ kg^{-1} in this validation set, and the segmented models for regions c and d predicted its calorific value as 27.50 MJ kg^{-1} and 27.47 MJ kg^{-1}, respectively, with only a difference of 0.03 MJ kg^{-1} between them.

3.2 Repeatability evaluation

In this paper, we compared the measurement repeatability of the proposed holistic-segmented model and the conventional holistic model. Eight repeatability validation samples were randomly selected from 35 validation set samples, and the corresponding coal types were clean coal, clean coal, medium coal, medium coal, slime, slime, gangue and gangue, numbered from 1 to 8. Using this instrument and these two models, each coal sample was measured 5 times, and their SDs were calculated. The obtained results are shown in Fig. 12.

As shown in Fig. 12, the blue and red horizontal lines represent the average SDs obtained by measuring the calorific values of the 8 samples using the holistic model and the holistic-segmented model, respectively. The purple horizontal line represents the repeatability requirement of 0.12 MJ kg^{-1} for coal calorific value measurement as stipulated by the Chinese national standard. The SDs predicted by the holistic model are all higher than 0.12 MJ kg^{-1}, indicating that the holistic model is unstable and the repeatability is poor. However, the SDs predicted by the holistic-segmented model are significantly improved compared with the holistic model. Furthermore, the repeatability of the model measurement is better in the regions of c and d, while that in the regions of a and b is relatively poor. This may be because slime and gangue belong to non-

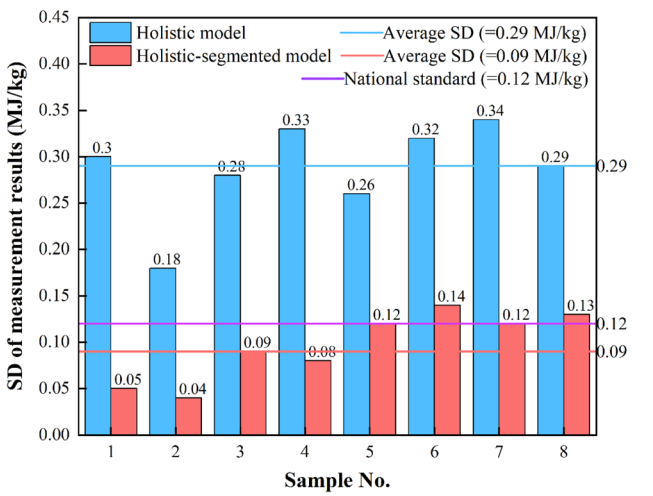


Fig. 12 Comparison of the SDs of the calorific value measured using the holistic model, the holistic-segmented model, and the national standard method.

Table 1 Evaluation of predictive performance for the holistic and holistic-segmented models

Model type	Segmented region	R^2	RMSEP (MJ kg^{-1})	AAE (MJ kg^{-1})	ARE (%)
Holistic model	—	0.998	0.65	0.55	4.71
Holistic-segmented model	a	0.995	0.43	0.32	6.57
	b	0.995	0.35	0.30	2.22
	c	0.996	0.31	0.27	1.27
	d	0.999	0.22	0.24	0.81
	Average	0.996	0.33	0.28	2.71

homogeneous samples, which not only contain more elemental compounds and impurities but also have more fluctuations in chemical components, resulting in poor measurement repeatability. In contrast, medium coal and clean coal are typically purer, with relatively stable chemical component content, resulting in better measurement repeatability. The holistic-segmented model achieved an average SD of 0.09 MJ kg^{-1} for repeated measurements of the 8 samples, which is lower than the 0.12 MJ kg^{-1} specified by the national standard method. This indicates that the holistic-segmented model can greatly improve the stability of measurement and has met the requirements of the national standard.

4 Conclusions

In this work, a rapid coal calorific value analyzer based on NIRS-XRF was developed, and field testing and feasibility verification were conducted in a coal preparation plant. The instrument utilizes NIRS to measure organic groups such as C–, H– and O–, which are positively related to the calorific value of coal. It also utilizes XRF to measure inorganic ash-forming elements such as K, Ca, Na, Mg, Al and Si, which are negatively related to the calorific value of coal. This allows for accurate quantitative analysis of the coal calorific value. The optical modules efficiently capture NIRS spectra and XRF energy spectra, while the control module uses PLC to control the timing of each module. The hydrogen production module is used to maintain a pure hydrogen atmosphere in the XRF measurement chamber. To use this instrument, simply place the pulverized coal in the sample cell, flatten it, and then click the measurement button on the human–machine interface to obtain the calorific value result within 1 minute.

For the analysis of spectral data, we proposed a set of spectral fusion preprocessing algorithms based on smoothing, standard normal and normalization, as well as a quantitative analysis algorithm of PLSR based on the holistic-segmented model. The holistic-segmented model is a new type of prediction model that combines the traditional holistic model with four segmented models. The holistic model roughly determines the region to which the sample belongs, and then the corresponding segmented model to make an accurate prediction of the calorific value. The test results showed that the RMSEP, AAE and ARE of the holistic-segmented model were reduced by a factor of 1 compared to the conventional holistic model, reaching 0.33 MJ kg^{-1}, 0.28 MJ kg^{-1} and 2.71%, respectively. Additionally, the SD of measurement repeatability decreased from 0.29 MJ kg^{-1} to 0.09 MJ kg^{-1}, meeting the requirements of the Chinese national standard.

The developed rapid analyzer, based on NIRS-XRF fusion, enables accurate, rapid, and highly stable measurement of the calorific value of coal. It also offers user-friendly operation with a one-button fully automatic measurement feature. The instrument provides coal preparation plants with valuable calorific value data for timely adjustment of the washing process and precise control of the coal quality. Future research will be devoted to enhancing the long-term stability of the instrument's performance in a harsh industrial environment, as well as exploring its potential application in other industries.

Author contributions

Rui Gao and Shuqing Wang are the main authors and are responsible for the instrument design, conceptualization, methodology, software, data curation and formal analysis of the experiment, and the writing of the original draft. Jiaxuan Li and Zhihui Tian are responsible for investigation, validation and supervision. Yan Zhang and Lei Zhang are responsible for conceptualization and methodology, while Lei Zhang is also responsible for writing reviews, editing and project management. Zefu Ye and Zhujun Zhu are responsible for providing the necessary resources for the experiment. Wangbao Yin and Suotang Jia are responsible for project management and funding acquisition, while Wangbao Yin is also responsible for writing reviews and editing. All authors reviewed and commented on subsequent versions of the manuscript, and read and approved the final manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

National Key R&D Program of China (2017YFA0304203); National Energy R&D Center of Petroleum Refining Technology (RIPP, SINOPEC); Changjiang Scholars and Innovative Research Team in University of Ministry of Education of China (IRT_17R70); National Natural Science Foundation of China (NSFC) (61975103 and 627010407); 111 Project (D18001); Fund for Shanxi "1331KSC".

References

1 S. U. Patel, B. Jeevan Kumar, Y. P. Badhe, B. K. Sharma, S. Saha, S. Biswas, A. Chaudhury, S. S. Tambe and B. D. Kulkarni, *Fuel*, 2007, **86**, 334–344.
2 S. S. Matin and S. C. Chelgani, *Fuel*, 2016, **177**, 274–278.
3 L. Wang, X. Wu, X. Gao, Y. Wu and K. Cen, *Thermochim. Acta*, 2021, **703**, 179011.
4 N. Bech, P. A. Jensen and K. Dam-Johansen, *Biomass Bioenergy*, 2009, **33**, 534–537.
5 L. Yongsheng, J. Wenbao, H. Daqian, S. Qing, C. Can, Z. Haojia, H. Wenyu, H. Yanquan and C. Da, *Appl. Radiat. Isot.*, 2015, **95**, 233–238.
6 M. Borsaru, M. Biggs, W. Nichols and F. Bos, *Appl. Radiat. Isot.*, 2001, **54**, 335–343.
7 C. S. Lim and D. A. Abernethy, *Appl. Radiat. Isot.*, 2005, **63**, 697–704.
8 L. Zhang, Y. Gong, Y. Li, X. Wang, J. Fan, L. Dong, W. Ma, W. Yin and S. Jia, *Spectrochim. Acta, Part B*, 2015, **113**, 167–173.
9 S. Sheta, M. S. Afgan, Z. Hou, S.-C. Yao, L. Zhang, Z. Li and Z. Wang, *J. Anal. At. Spectrom.*, 2019, **34**, 1047–1082.

10 Z. Lu, J. Mo, S. Yao, J. Zhao and J. Lu, *Energy Fuels*, 2017, **31**, 3849–3855.
11 W. Yin, L. Zhang, L. Dong, W. Ma and S. Jia, *Appl. Spectrosc.*, 2009, **63**, 865–872.
12 L.-B. Guo, D. Zhang, L.-X. Sun, S.-C. Yao, L. Zhang, Z.-Z. Wang, Q.-Q. Wang, H.-B. Ding, Y. Lu, Z.-Y. Hou and Z. Wang, *Front. Phys.*, 2021, **16**, 22500.
13 Z. Wang, T.-B. Yuan, Z.-Y. Hou, W.-D. Zhou, J.-D. Lu, H.-B. Ding and X.-Y. Zeng, *Front. Phys.*, 2013, **9**, 419–438.
14 Y.-T. Fu, W.-L. Gu, Z.-Y. Hou, S. A. Muhammed, T.-Q. Li, Y. Wang and Z. Wang, *Front. Phys.*, 2020, **16**, 22502.
15 C. Yan, T. Zhang, Y. Sun, H. Tang and H. Li, *Spectrochim. Acta, Part B*, 2019, **154**, 75–81.
16 Y. Ozaki, *Anal. Sci.*, 2012, **28**, 545–563.
17 J. M. Andrés and M. T. Bona, *Anal. Chim. Acta*, 2005, **535**, 123–132.
18 J. M. Andres and M. T. Bona, *Talanta*, 2006, **70**, 711–719.
19 M. T. Bona and J. M. Andres, *Talanta*, 2007, **72**, 1423–1431.
20 E. A. Cloutis, *Fuel*, 2003, **82**, 2239–2254.
21 Y. Wang, M. Yang, G. Wei, R. Hu, Z. Luo and G. Li, *Sens. Actuators, B*, 2014, **193**, 723–729.
22 D. W. Kim, J. M. Lee and J. S. Kim, *Korean J. Chem. Eng.*, 2009, **26**, 489–495.
23 D. Xiao, Z. Yan, J. Li, Y. Fu and Z. Li, *Spectrochim. Acta, Part A*, 2023, **287**, 122042.
24 B. T. Le, D. Xiao, Y. Mao and D. He, *Infrared Phys. Technol.*, 2018, **93**, 34–40.
25 N. Begum, A. Maiti, D. Chakravarty and B. S. Das, *Fuel*, 2020, **280**, 118676.
26 W. Liu, B. Peng, X. Liu, F. Ren and L. Zhang, *J. Appl. Spectrosc.*, 2021, **88**, 645–652.
27 A. R. Shirazi, O. Börtin, L. Eklund and O. Lindqvist, *Fuel*, 1995, **74**, 247–251.
28 Y.-L. Liu, Q.-X. Zhang, J. Zhang, H.-T. Bai and L.-Q. Ge, *Nucl. Sci. Tech.*, 2019, **30**, 52.
29 S. Carter, R. Clough, A. Fisher, B. Gibson, B. Russell and J. Waack, *J. Anal. At. Spectrom.*, 2019, **34**, 2159–2216.
30 C. Vanhoof, J. R. Bacon, A. T. Ellis, L. Vincze and P. Wobrauschek, *J. Anal. At. Spectrom.*, 2018, **33**, 1413–1431.
31 C. Vanhoof, J. R. Bacon, U. E. A. Fittschen and L. Vincze, *J. Anal. At. Spectrom.*, 2020, **35**, 1704–1719.
32 T. D. T. Oyedotun, *Geol. ecol. landsc.*, 2018, **2**, 148–154.
33 Z. Wen, H. Liu, M. Zhou, C. Liu and C. Zhou, *Fuel*, 2023, **332**, 125991.
34 C. R. Ward, S. J. Kelloway, J. Vohra, D. French, D. R. Cohen, C. E. Marjo and I. E. Wainwright, *Int. J. Coal Geol.*, 2018, **191**, 172–188.
35 X. Zhang, L. Liang, T. Li, J. Tan, X. Liang and G. Xie, *Minerals*, 2021, **11**, 1433.
36 R. Guo, L. Zhang, Y. Zheng and H. Yao, *Front. Phys.*, 2022, **10**, 1054796.
37 Y. Zhang, X. Zhang, W. Jia, S. Qing, Y. Ling, D. Hei and D. Chen, *Appl. Spectrosc.*, 2016, **70**, 272–278.
38 I. F. Mikhailov, A. A. Baturin, A. I. Mikhailov, S. S. Borisova and L. P. Fomina, *Rev. Sci. Instrum.*, 2018, **89**, 023103.
39 S. C. Chelgani, B. Hart, W. C. Grady and J. C. Hower, *Int. J. Coal Prep. Util.*, 2011, **31**, 9–19.

