Quantitative Analysis of Coal Quality Using Laser-Induced Breakdown Spectroscopy

Rongzhou Zhang , Syed Zaheer Ud Din , Chunling Dang , Xiangming Kong , Rongxin Ma , Jianli Ning , Guangtao Fu , Jiancai Leng , Wenhao Zhang *

Posted Date: 14 October 2025

doi: 10.20944/preprints202510.0947.v1

Keywords: LIBS; coal; Partial Least Squares Regression (PLSR); machine learning

Article

Quantitative Analysis of Coal Quality Using Laser-Induced Breakdown Spectroscopy

Rongzhou Zhang [1], Syed Zaheer Ud Din [1], Chunling Dang [1], Xiangming Kong [2], Rongxin Ma [1], Jianli Ning [1], Guangtao Fu [1], Jiancai Leng [1] and Wenhao Zhang [1,2,*]

[1] Shandong Key Laboratory of Optoelectronic Sensing Technologies, International School for Optoelectronic Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan, Shandong, 250353, China

[2] Shandong Tevinf Intelligent Technology Co., LTD, Jinan, Shandong, 250022, China.

* Correspondence: zhangwenhao@qlu.edu.cn

Abstract

The study presents a novel approach that integrates laser-induced breakdown spectroscopy (LIBS) data with machine learning algorithms for the rapid evaluation of coal quality. The developed framework enables quantitative determination of three critical parameters: Ash Content (Aad), Carbon Content (Cd), Sulfur Content (Stad). The experimental implementation utilized an optimized dataset to construct and evaluate the predictive model. The LIBS prototype system enables spectral data acquisition under controlled experimental conditions. Data preprocessing is carried out by systematically removing background interference, substrate effects, and saturated signals using adaptive filtering techniques. Characteristic emission peaks corresponding to target elements are identified through multivariate analysis, and Partial Least Squares Regression (PLSR) serves as the core algorithm for quantitative analysis. Systematic iterative optimization of multivariate preprocessing parameters and adaptive peak selection strategies yields substantial improvements in both predictive accuracy and computational efficiency, with determination coefficients ($R^2 > 0.90$) demonstrated for all target analytes. This enhanced accuracy validates the viability of LIBS as a robust alternative to conventional analytical methods for coal composition analysis. The LIBS demonstrates substantial advantages in coal quality assessment, thereby enhancing the overall efficiency of both coal extraction and quality evaluation processes.

Keywords: LIBS; coal; Partial Least Squares Regression (PLSR); machine learning

1. Introduction

Since the First Industrial Revolution, coal has served as an indispensable energy source and continues to occupy a crucial position in the global energy structure. Although the development of new energy alternatives has influenced coal consumption, its central role in the energy mix remains irreplaceable. According to the "2025-2030 Analysis Report on the Status and Future Development Trends of the Coal Industry" published by the Zhongyan Puhua Industry Institute, coal consumption in China continues to increase annually. Therefore, rapid and efficient methods for coal quality assessment have become a pressing need. This study proposes a novel approach for coal quality detection that integrates LIBS with machine learning. This approach is operable, fast, safe, environmentally friendly, and highly flexible. It utilizes plasma spectra excited by high-energy lasers to determine the elements present in coal [1,2]. As a non-destructive technique, it ensures repeatability while significantly reducing experimental costs. In recent years, with the advancement of artificial intelligence, the integration of machine learning algorithms with spectral data has substantially enhanced the intelligence and accuracy of material analysis in engineering applications [3–5].

 doi:10.20944/preprints202510.0947.v1

Significant variations in sulfur content (Stad), ash content (Aad), and carbon content (Cd) among different coal types lead to corresponding differences in combustion efficiency and pollutant emissions. High sulfur content increases the risk of low-temperature corrosion and spontaneous combustion of coal dust, while also directly determining the amount of sulfur dioxide (SO2) released during combustion. Accurate sulfur detection, therefore, enables more effective control of the combustion process to mitigate sulfur oxide emissions. Ash content strongly influences coal combustion efficiency. Coals with higher ash content are more difficult to ignite, burn less stably, and experience greater heat loss due to incomplete combustion. In industrial applications, the excessive ash content also impacts the service life of furnace chambers. Conversely, higher carbon content enhances combustibility, reduces combustion losses, and releases greater heat, thereby improving overall combustion efficiency [6].

Sulfur, ash, and carbon contents are key indicators of coal quality. Traditional methods for determining sulfur content include gravimetric techniques (e.g., Eschka method) and coulometric titration. For instance, the gravimetric method involves mixing the coal sample with Eschka reagent and igniting it at 850°C, converting sulfur into sulfate, which precipitates as barium sulfate. The sulfur content is then calculated by weighing the precipitate, a process that typically requires 4-6 hours. Similarly, ash content is traditionally measured by igniting a coal sample in a muffle furnace at 550°C until a constant weight is achieved, with the remaining residue quantified as ash. Carbon content is commonly determined using the two-stage furnace method, in which the furnace temperatures are set to 850°C and 500°C. Following pyrolysis with silver permanganate, interfering substances such as chlorine and nitrogen are removed, and the carbon content is calculated from the amount of carbon dioxide produced. All of these traditional methods for determining sulfur, ash, and carbon contents are based on combustion processes, which are time-consuming, labor-intensive, and destructive, making them unsuitable for the real-time and rapid analysis required by the modern coal industry applications.

In this study, a novel method is developed for rapid and online quantitative analysis of key quality indicators in coal—namely, ash, carbon, and sulfur contents—based on LIBS. Conventional coal quality assessment methods are typically laboratory-dependent, time-consuming, and labor-intensive, making them inadequate for meeting the modern coal industry's demand for real-time quality monitoring and efficient sorting. In contrast, LIBS offers remarkable advantages, including minimal sample preparation, high analysis speed, multi-element detection capability, and strong potential for online applications, rendering it an attractive candidate technique for this purpose. However, the intrinsic complexity and heterogeneity of coal introduce challenges such as spectral interference, matrix effects, and measurement instability, which hinder the analytical performance of LIBS in coal analysis. Therefore, the primary objective of this work is to establish a robust quantitative analytical model through algorithm optimization, thereby significantly enhancing the accuracy and reliability of ash, carbon, and sulfur content determination in coal. The anticipated outcomes are expected to provide powerful technical support for rapid coal quality evaluation, combustion efficiency optimization, and emission control, thus contributing to the advancement of clean and efficient coal utilization.

2. Experimental Setup

LIBS is a technique that focuses high-energy pulsed lasers onto the surface of a sample. The laser ablation generates high-temperature plasma, causing atoms and ions in the sample to be heated and ionized, leading to the formation of highly excited states and the creation of micro-scale ablation pits on the surface. As these excited species relax from higher to lower energy levels, they emit photons. Because each element has a distinct atomic structure, the resulting emission spectra contain distinct spectral lines characteristic of the elements present. By collecting and analyzing the plasma spectra, it is theoretically possible to determine the elemental composition of a sample. LIBS has been successfully applied for detecting trace elements such as chromium, lead, and copper in soil [7–10]. The feasibility of using LIBS to detect sulfur, ash content, and carbon levels in coal has also been successfully validated [11]. The principle of LIBS is illustrated in **Figure 1**.

 doi:10.20944/preprints202510.0947.v1

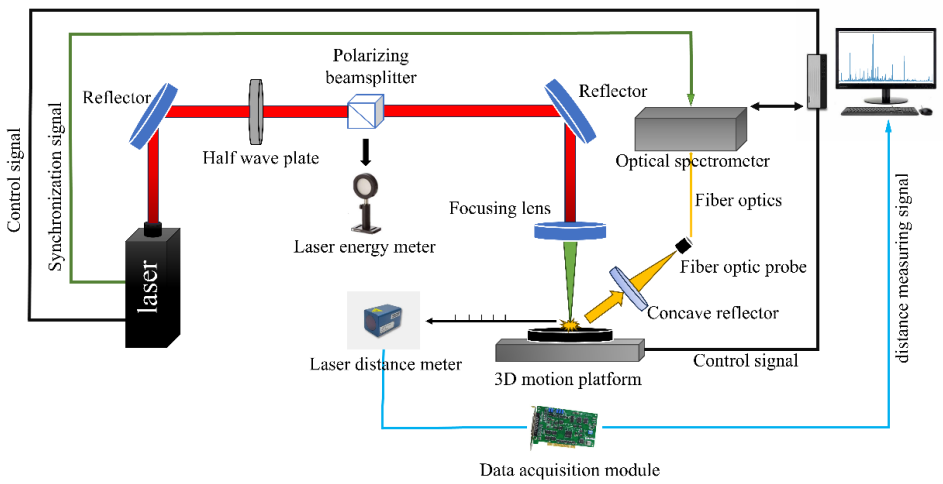

Figure 1. Schematic diagram of a typical LIBS system.

A conventional LIBS setup is employed, consisting of a Q-switched Nd:YAG pulsed laser (Laibo). The laser generates a pulsed beam, and its output energy is adjusted using a half-wave plate in combination with a polarizing beam splitter. The beam is then directed by two mirrors and focused through a plano-convex lens onto the surface of the sample, where plasma is generated. The plasma emission is collected by a lens, transmitted through a fiber-optic probe and multimode fiber, and delivered to the spectrometer. The spectrometer converts the optical signals into electrical signals, which are subsequently processed by a computer. An AVANTES spectrometer was used in the experiment, covering a detection wavelength range of 181–769 nm. A laser rangefinder (**Figure 1**), coupled with a data acquisition card, provided real-time focal distance information of the sample to adjust the focusing position of the pulsed laser after the lens. The Z-axis of the three-dimensional translation stage received the feedback signal from the rangefinder to ensure precise focusing. During measurements, coal samples were placed on the translation stage for cyclic scanning. Under computer control, the X- and Y-axes of the stage guided the laser to irradiate the sample in a 5 × 5 matrix. A fixed number of laser pulses was applied at each sampling point, and the averaged spectrum of the coal sample was obtained by accumulating repeated shots.

With advances in artificial intelligence, laser, and optical technologies, laser-induced breakdown spectroscopy (LIBS) has been applied across diverse fields. In the biomedical domain, David Pokrajac et al. employed principal component analysis as a dimensionality reduction method (K-nearest neighbors, classification and regression trees, neural networks, support vector machines, adaptive local hyperplanes, and linear discriminant classification) to achieve high-precision automated classification of complex protein samples from LIBS data [12]. Jonathan Diedrich et al. employed LIBS to identify pathogenic E. coli strains, experimentally demonstrating the technique's potential for distinguishing pathogenic strains from common environmental strains [13]. Additionally, LIBS enables quantitative analysis of trace metal accumulation in teeth [14]. The integration of LIBS with machine learning for diagnosing and staging multiple myeloma offers a rapid, minimally invasive, cost-effective, and robust approach for human malignant tumor diagnosis and staging [15]. In environmental science, Melinda Darby-Dale et al. compared LIBS with PLS regression combined with lasso modeling for analyzing geological samples [16]. Alexander Potnov et al. applied LIBS to characterize emissions from nitroaromatic compounds (NC) and polycyclic aromatic hydrocarbons (PAH) in ambient air samples. Their study demonstrated the potential to distinguish one chemical category from another, and under optimal conditions, even identify specific compounds using LIBS [17]. Bruno Bousquet et al. developed a mobile LIBS-based system specifically for in-situ analysis of contaminated soils, describing the creation of a portable LIBS system dedicated to in-situ analysis of heavy metal-contaminated soils [18]. In deep space exploration, LIBS technology can analyze mineral elements on unknown planets. For instance, LIBS has been employed for the analysis and study of Martian elements and minerals [19–21]. Due to its non-destructive nature, LIBS is also utilized in

 doi:10.20944/preprints202510.0947.v1

archaeology, cultural heritage identification, and conservation [22,23]. LIBS technology offers novel physical approaches for chemical element analysis [24–27].

3. Materials and Methods

3.1. Sample Preparation

The samples used in this study were obtained from the reserve stock of a thermal power plant, with a total of 83 true value samples. The true values of sulfur, ash content, and carbon content from these samples were used as the reference data for modeling. **Table 1** presents the true values of the samples used for modeling.

Table 1. True values of sulfur, ash, and carbon contents in the coal samples.

No.	Aad (kg/mg)	Cd (kg/mg)	Stad (kg/mg)	No.	Aad (kg/mg)	Cd (kg/mg)	Stad (kg/mg)
1	25.62	62.67	0.79	43	25.63	62.47	0.64
2	25.49	62.42	0.70	44	25.20	61.98	0.60
3	25.44	64.30	0.86	45	27.40	60.24	0.63
4	23.94	65.20	0.62	46	29.66	59.78	1.09
5	26.07	63.95	0.70	47	28.66	60.72	1.02
6	27.27	63.71	0.71	48	26.64	61.59	1.06
7	25.92	62.79	0.66	49	24.82	64.94	0.62
8	27.64	62.36	0.72	50	25.92	63.67	0.90
9	30.04	59.27	0.69	51	25.74	63.34	0.86
10	28.71	61.30	0.70	52	29.72	60.36	0.86
11	29.19	60.74	0.53	53	27.02	62.51	0.84
12	32.64	62.92	0.86	54	26.76	62.13	0.94
13	29.83	64.06	0.65	55	29.66	59.09	0.78
14	29.34	59.29	0.70	56	30.41	58.78	0.82
15	33.48	55.37	0.66	57	32.68	56.85	0.98
16	28.40	61.32	0.82	58	33.86	56.60	1.03
17	30.64	57.61	0.60	59	31.51	57.63	0.82
18	27.00	62.75	0.75	60	32.09	56.72	0.90
19	24.82	63.25	0.74	61	29.71	58.60	1.10
20	27.22	62.29	0.92	62	29.72	58.63	1.00
21	28.88	60.92	0.67	63	27.36	61.60	0.74
22	29.17	60.09	0.58	64	24.23	62.53	1.12
23	29.79	62.13	0.66	65	27.77	61.63	0.74
24	27.14	62.58	0.64	66	31.43	58.10	1.10
25	28.88	60.31	0.62	67	29.97	60.00	0.94
26	29.44	60.94	0.62	68	28.85	60.55	0.72
27	29.14	59.37	0.72	69	32.02	58.46	1.12
28	27.86	62.41	0.73	70	30.78	61.56	0.98
29	27.02	63.45	0.59	71	29.86	60.39	0.96
30	25.52	65.03	0.64	72	26.90	62.42	0.84
31	26.20	63.60	0.50	73	27.84	61.29	1.02
32	29.16	60.89	0.70	74	27.60	63.14	0.87
33	24.32	63.87	0.60	75	30.62	58.60	0.98
34	23.10	66.04	0.74	76	28.92	60.20	0.88
35	27.36	63.40	0.52	77	27.36	63.08	0.66
36	22.92	64.37	0.58	78	29.77	58.93	0.88
37	21.26	64.59	0.63	79	33.10	56.42	1.03
38	21.20	65.73	0.54	80	26.10	62.53	0.78
39	23.80	64.41	0.52	81	28.21	61.51	0.86

 doi:10.20944/preprints202510.0947.v1

40	24.29	63.02	0.58	82	25.90	62.74	0.82
41	23.54	63.89	0.58	83	29.50	58.24	0.84
42	23.64	62.17	0.58				

The feasibility of LIBS-based online coal quality analysis has been well established, and equipment for various measurement methods, including coal block analysis, coal powder analysis, and pellet sample preparation, has been developed [28–30]. Currently, pellet sample preparation remains the most suitable approach for LIBS-based online coal quality assessment. In this work, since it is necessary to focus a high-energy pulsed laser onto the sample surface, standard samples with uniform heights were required. A total of 83 true-value samples, each containing the same mass of coal powder, were placed into molds and pressed under identical pressure to form standard pellets. These samples exhibit uniform height, eliminating the need for focus adjustments when changing samples. This uniformity ensures experimental rigor by controlling sample height as a consistent variable. **Figure 2** illustrates the standard sample pressed using the molds.

Figure 2. Standard coal sample prepared for LIBS analysis.

The prepared standard samples were placed into the LIBS prototype, and a 5×5×5 point pattern was set (each sample consisted of five rows and five columns, with each point measured five times). Since the ablation craters are on the micron scale, the depth of the ablation craters is negligible.

3.2. Spectral Preprocessing

The spectra obtained directly from the LIBS prototype are referred to as raw spectra. Due to the presence of noise in these spectra, they cannot be used directly for modeling. Prior to model development, the raw spectra must be preprocessed to minimize interfering factors and ensure the accuracy and reliability of the predictive model.

3.2.1. Background Spectrum Removal

Before standardizing and normalizing the raw spectra, it is necessary to remove background light from data affected by other light sources during collection. The background correction process can be expressed by the following equation:

$$I = I_{raw} - I_{background} \quad (1)$$

 doi:10.20944/preprints202510.0947.v1

Here, I represents the background-corrected spectral intensity, I_{raw} is the measured spectral data affected by background light, and $I_{background}$ is the background spectral data. Since no additional light sources are present in the sample chamber of the LIBS prototype used in this experiment, the background spectral data is set to zero, $I_{background} = 0$. Therefore, the background-corrected data is equivalent to the collected raw spectral data, i.e., $I = I_{raw}$.

3.2.2. Data cleaning

After background correction, the spectral data must be further cleaned. First, data points reaching saturation values (the saturation point of the spectrometer is 65535) are removed to improve data quality. In addition, environmental disturbances (such as aerosols and vibrations) and fluctuations in laser energy can cause inconsistencies in the plasma state excited by each laser pulse. This instability is reflected in the collected spectral data, where single-shot spectra deviate significantly from the true state. To mitigate this effect, anomalous spectra must be excluded. By removing these, the resulting spectral dataset used for model development is more representative and cleaner, enhancing the accuracy of material identification.

The coal spectral data collected in this experiment are stored as one-dimensional matrices of size 1×4094, with 125 spectra acquired for each sample. In the first step, the mean spectrum is calculated by averaging the 125 spectra column-wise at each wavelength, yielding an average spectral intensity vector of size 1×4094. In the second step, the deviation of each individual spectrum from the mean spectrum is computed, followed by the calculation of the sum of squares of these deviations. In the third step, spectra that exceed a threshold are excluded, where the threshold is defined as 1.1 times the sum of squares. The formulas for these steps are expressed as follows:

$$avg_j = \frac{1}{125}\sum_{i=1}^{125} I_{i,j} \quad (2)$$

here, i and j represent the measurement number and the number of specific wavelengths, respectively. The final result is:

$$V = [avg_1, avg_2, avg_3 \cdots avg_{4094}] \quad (3)$$

where V represents the calculated average value vector.

$$A_i = \sum_{j=1}^{4094} (I_{i,j} - avj_j)^2 \quad (4)$$

In the above equation, A_i represents the sum of squares of the difference between the i-th measurement and the average value.

$$C_{threshold} = 1.1 \times (\frac{1}{150}\sum_{i=1}^{150} A_i) \quad (5)$$

here, $C_{threshold}$ represents the threshold value required for the selection process.

If the i-th data point in A_i exceeds $C_{threshold}$, the corresponding i-th row of the spectral matrix $I_{i,j}$ will be excluded.

3.2.3. Normalized De-Basing

Due to significant variations in the characteristic spectral line intensities across different wavelengths corresponding to various elements within the full LIBS spectrum, the disparate scales (dimensionality) of the data adversely impact model performance. To enhance computational efficiency and eliminate factors detrimental to model iterative calculations, data normalization is required. The normalization formula is as follows:

$$I_{normalized} = \frac{I - I_{min}}{I_{max} - I_{man}} \quad (6)$$

where $I_{normalized}$ represents the normalized spectral intensity, I is the spectral intensity value after data cleaning, I_{min} is the minimum spectral intensity within each dataset, and I_{max} is the maximum spectral intensity within each dataset.

Low-frequency signals, such as noise generated by the instrument, environmental light scattering, and sample matrix effects (e.g., fluorescence, scattering), are often present in the baseline

 doi:10.20944/preprints202510.0947.v1

of spectral data. These signals can obscure the characteristic peaks of the target analytes (e.g., elemental spectral lines or molecular absorption peaks) and overwhelm the weak peaks of trace components. In the experiment, baseline removal is employed to improve the signal-to-noise ratio, making the target peaks more prominent and revealing weak peaks that were previously submerged, thus facilitating the identification of key elemental features. Baseline removal in this study is performed using a sliding window method. This approach involves sequentially extracting each specific wavelength within the wavelength range of the spectrometer, from small to large. For each specific wavelength, a left and right window range is set, and the minimum spectral intensity within that range is identified. The spectral intensities corresponding to all individual wavelengths within the window range are then subtracted by this minimum value to achieve baseline removal. The changes in the spectral data before and after baseline removal are shown in **Figure 3**.

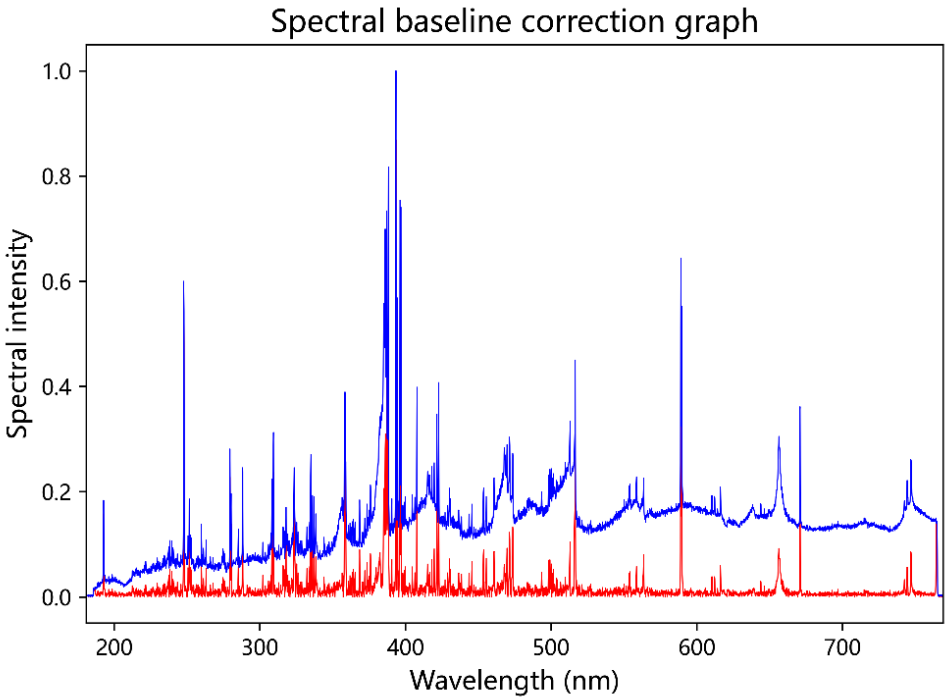

Figure 3. Comparison of before and after basal correction.

3.3. Partial Least Squares Regression (PLSR)

PLSR is a statistical multivariate data analysis method that is particularly effective in situations where multicollinearity exists between the dependent and independent variables. It excels in handling high-dimensional data and small sample problems. The core of PLSR is a black-box algorithm based on latent linear relationships for prediction. It establishes a linear relationship model between the independent and dependent variables through both dimensionality reduction and regression methods. Unlike traditional multiple linear regression, PLSR does not directly fit the original variables but instead extracts a set of new composite variables called "latent variables," which maximize the covariance between the independent and dependent variables, thus avoiding the instability of the model caused by multicollinearity. First introduced in 1983 by S. Wold, C. Albano, and others, PLSR has rapidly developed in theory, methodology, and applications over recent decades [31]. As a multivariate linear regression analysis technique, it has been widely used in fields such as chemistry, environmental science, biomedicine, and finance [32]. Relevant literature indicates that PLSR combined with hyperspectral data is effective for estimating soil organic matter content. This evidence demonstrates that integrating PLSR with spectral data is a viable analytical approach.

3.4. Evaluation Parameters

 doi:10.20944/preprints202510.0947.v1

The evaluation parameters for PLSR include the coefficient of determination (R^2), root mean square error of the test set (RMSEP), and root mean square error of cross-validation (RMSECV), which are defined as follows:

$$R^2 = 1 - \frac{\sum_i^n (y_i - \hat{y}_i)^2}{\sum_i^n (y_i - \bar{y}_i)^2} \quad (7)$$

$$RMSEP = \sqrt{\frac{\sum_i^m (y_i - \hat{y}_i)^2}{m}} \quad (8)$$

$$RMSECV = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n-1}} \quad (9)$$

where y_i and $\hat{y}_i$ represent the reference values and predicted values, respectively, $\bar{y}_i$ is the mean reference value, m denotes the number of coal samples in the test set, and n represents the number of coal samples in the training set.

4. Results

In this experiment, 67 processed full-spectrum datasets are used as the training set for the PLSR model, and 16 datasets are used as the test set. According to baseline correction theory, an excessively small window value may remove useful spectral information. Based on empirical evaluation, a window value of 16 is selected. The number of latent variables (n_components) is a core parameter of the PLSR model, and the size of latent variables directly affects the complexity, prediction accuracy, and generalization ability of the established model. The core goal of the PLSR experiment is to achieve the maximum prediction accuracy with the fewest components, and cross-validation was preferentially adopted to select the optimal number of latent variables.

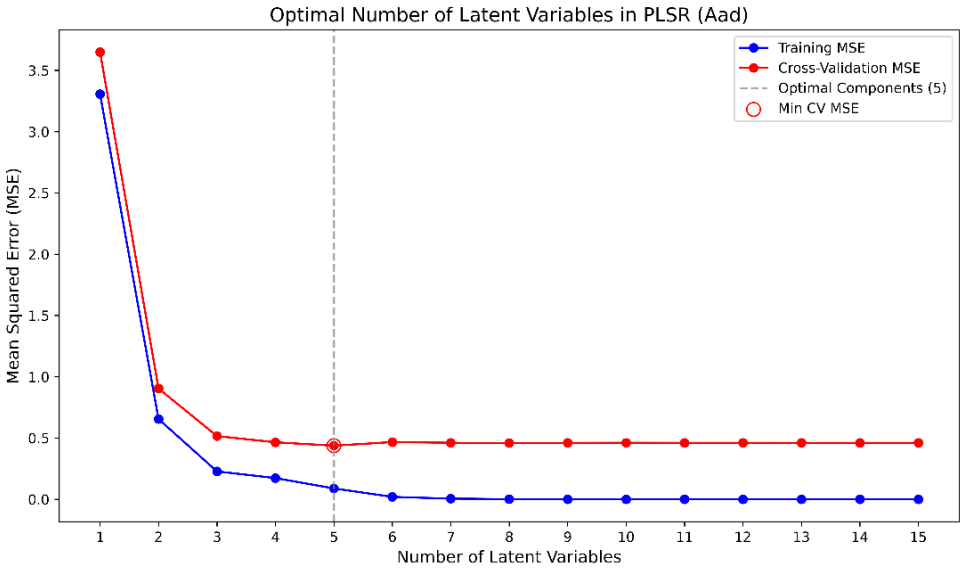

Figure 4. Optimal selection diagram of latent variables for ash content.

 doi:10.20944/preprints202510.0947.v1

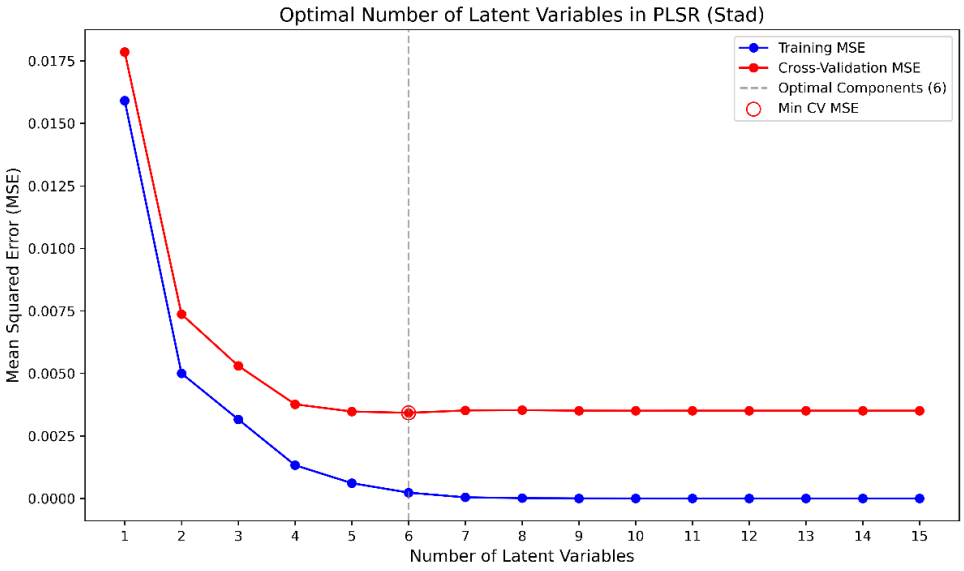

Figure 5. Optimal Selection Diagram of Latent Variables for Sulfur Component.

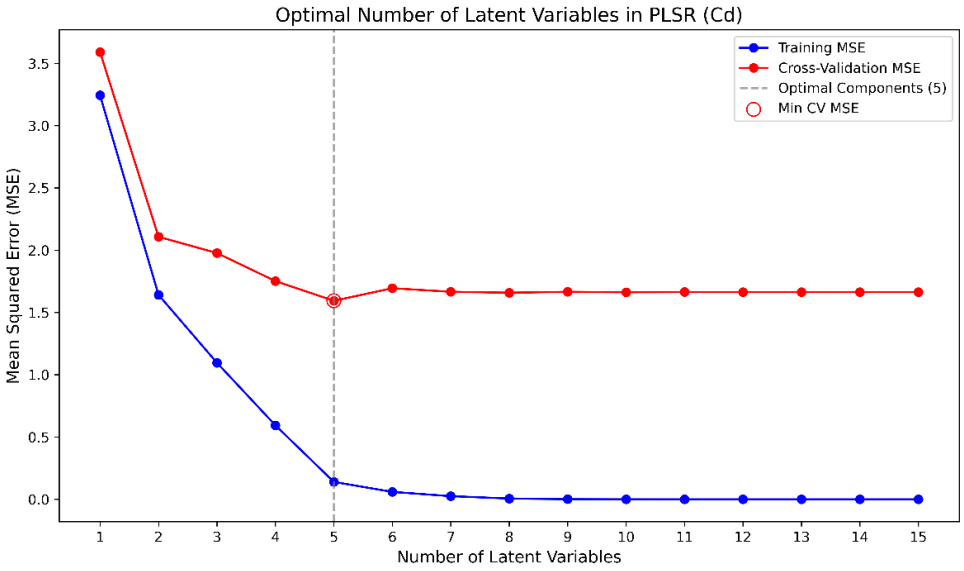

Figure 6. Optimal Selection Diagram of Latent Variables for Carbon Component.

Figure 4 illustrates the determination of the optimal number of latent variables for the PLSR model of ash content in coal samples, obtained by comparing the root mean square errors (RMSE) of cross-validation with those of the training set. The resulting optimal latent variable number for the ash model was 5. Similarly, **Figures 5** and **6** present the selection of latent variables for the PLSR models of sulfur and carbon in coal samples, respectively. The optimal numbers of latent variables determined for sulfur and carbon were 6 and 5. **Figures 4–6** show the determination of the optimal number of latent variables for ash, sulfur, and carbon prediction, with the results summarized in Table 2. Based on a comprehensive evaluation, a latent variable count of five (L=5) is ultimately selected for all models.

Table 2. Optimal Selection of Latent Variables in PLSR Model.

element	n_components
Aad	5
Cd	6
Stad	5

 doi:10.20944/preprints202510.0947.v1

Following the selection of relevant parameters, models were constructed to predict the ash content, sulfur content, and carbon content in coal samples. To enhance model generalizability, the actual measured values of ash content, sulfur content, and spectral data were sorted in ascending order. A subset of spectral data points was then systematically sampled from this sorted sequence to form the test set. This subset was selected to approximately cover the low, medium, and high ranges of the actual measurement values, ensuring it broadly represents the distribution of the entire dataset. The remaining data constituted the training set. The performance plots generated by the resulting models are presented below (**Figures 7–9**).

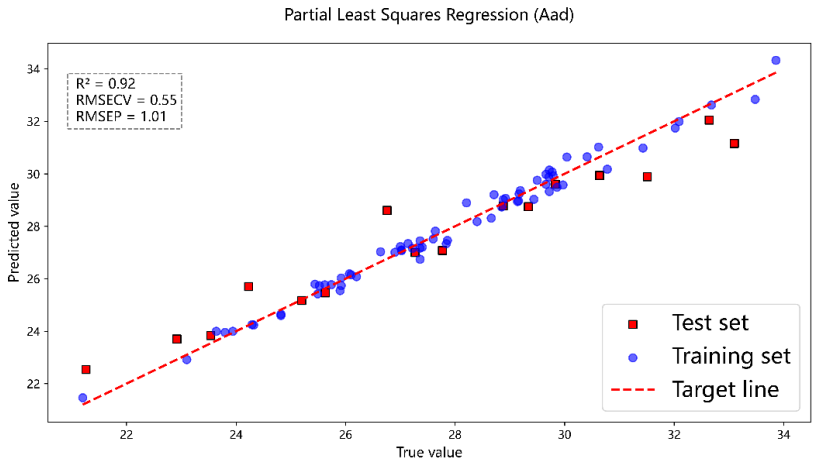

Figure 7. Full-spectrum modeling effect of ash content.

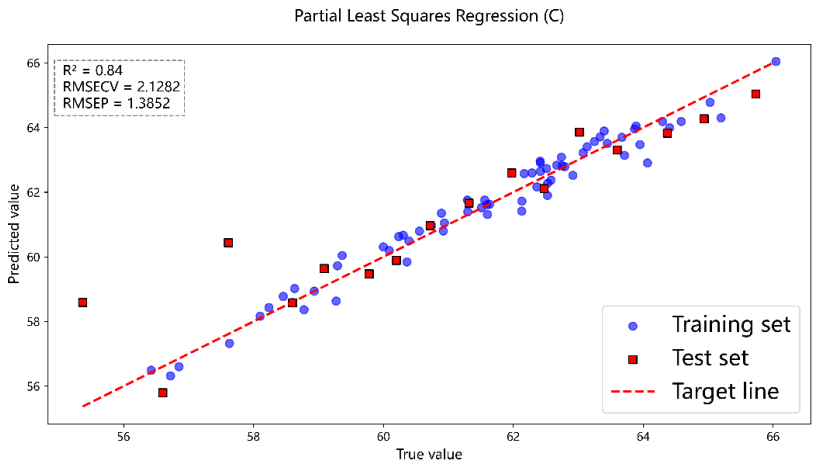

Figure 8. Spectral modeling effect of carbon content composition.

 doi:10.20944/preprints202510.0947.v1

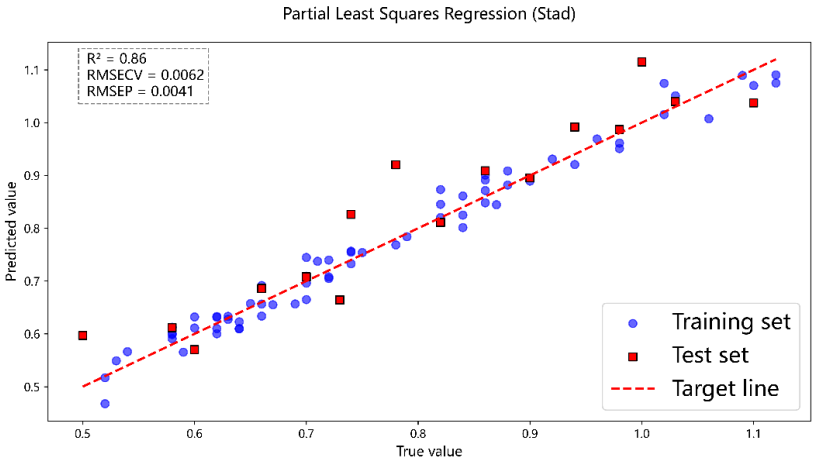

Figure 9. Spectral modeling effect diagram of sulfur content.

Figure 7 presents the prediction results of ash content in coal samples using the full-spectrum data, where the PLSR model achieved $R^2 = 0.92$, $RMSECV = 0.55$ and $RMSEP = 1.01$. **Figures 8** and **9** show the prediction results for carbon and sulfur, respectively. For carbon, the PLSR model yielded $R^2 = 0.84$, $RMSECV = 2.1282$ and $RMSEP = 1.3852$, and for sulfur, the corresponding values were $R^2 = 0.86$, $RMSECV = 0.0062$ and $RMSEP = 0.0041$. In comparison, the prediction accuracy for carbon in coal was relatively lower. This may be attributed to the fact that coal is an extremely complex mixture containing dozens of elements. Its spectral background is complicated and characterized by a low signal-to-noise ratio. The weak carbon signals are easily obscured or interfered with by stronger or denser spectral lines from other elements, making it difficult to accurately identify and quantify the spectral intensity of carbon.

4.1. Selection of Spectral Peaks in LIBS Data

To further enhance model performance, this study builds upon the fundamental principles of LIBS. Within the full-spectrum data, peak identification algorithms are employed to isolate and retain exclusively the peak data, discarding all non-peak information. As distinct elements possess unique atomic structures, their corresponding excited ionic spectra exhibit characteristic differences. These elemental fingerprints are preserved within the retained peak data. However, potential instrumental limitations of the LIBS prototype, including spectrometer wavelength shifts, nonlinear grating dispersion, detector pixel misalignment, and insufficient spectral resolution, may cause the acquired spectral peaks to deviate from the reference wavelengths reported in the NIST database. Consequently, the entire set of peak data is directly utilized for modeling and analysis, and its effectiveness is compared with that of models constructed using the full spectrum.

Following the establishment of underlying principles, the next step involves peak detection in the spectroscopic data. In LIBS data analysis, peak finding constitutes a fundamental step in processing, underpinning the entire workflow from raw spectral interpretation to elemental quantification. This process serves as the critical bridge converting LIBS spectral data into elemental information. Models utilize characteristic peaks for both qualitative identification and quantitative analysis. The positions of identified peaks at specific wavelengths form the basis of qualitative models, while the intensities of these peaks are employed in quantitative models. Furthermore, peak finding facilitates dimensionality reduction by eliminating a portion of non-relevant spectral data. The core objective of the peak finding algorithm is to identify local maxima (peaks) within the data sequence. The selection of peak significance is controlled by implementing a threshold filter; in this experiment, a peak intensity threshold of 0.01 was applied. **Figure 10** shows the peak-finding effect diagram of the full spectral data.

 doi:10.20944/preprints202510.0947.v1

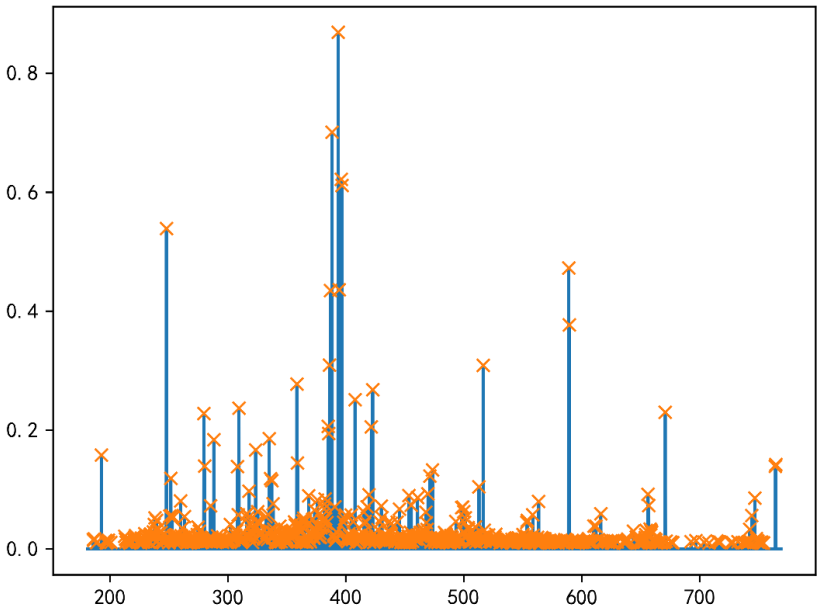

Figure 10. Peak effect diagram. Yellow markers indicate the identified spectral peaks.

4.2. Comparison of Predicted Carbon and Sulfur Contents

Models are trained using the peak-identified data, with non-peak data points set to zero. This procedure ensures that spectral regions lacking elemental fingerprint information are excluded from subsequent algorithmic computations. Applying the processed data to the established model revealed an improvement in the coefficient of determination (R^2) for the predicted sulfur concentration. Specifically, **Figure 11** illustrates the sulfur concentration prediction results obtained using the complete set of peak data as the modeling input, while **Figure 12** presents the corresponding prediction results for carbon concentration.

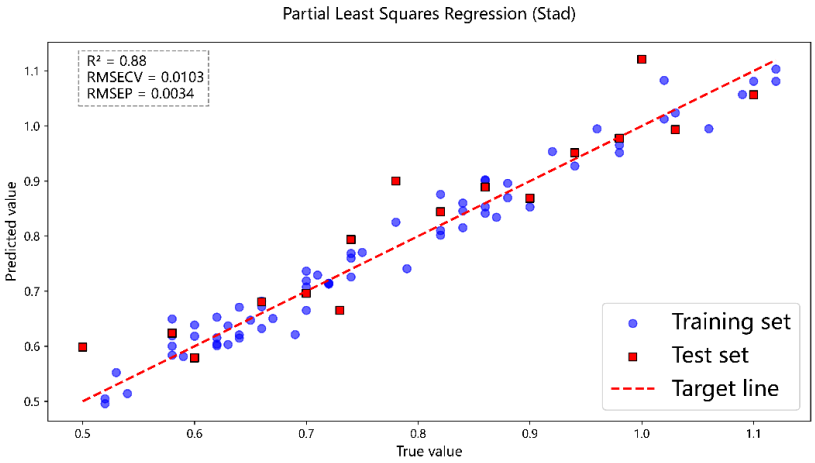

Figure 11. Prediction results of sulfur content using all peak data as the training and test sets in the PLSR model.

 doi:10.20944/preprints202510.0947.v1

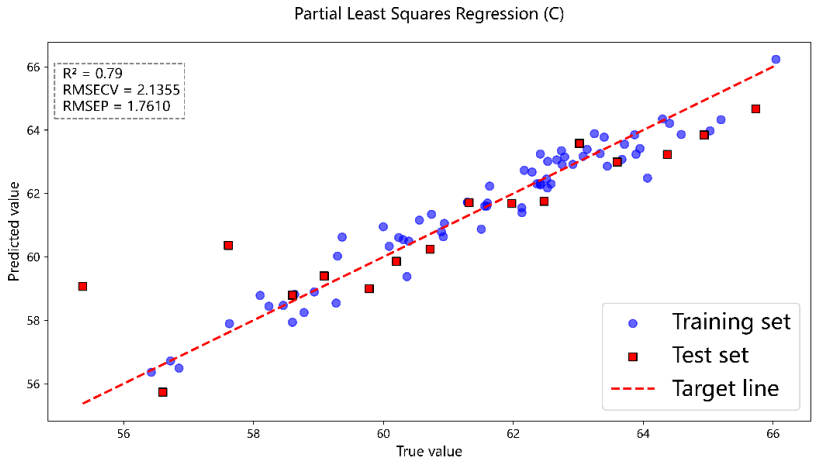

Figure 12. Prediction results of carbon content using all peak data as the training set and test set for the PLSR model.

As shown in **Figures 11** and **12**, after the peak-finding operation, the coefficient of determination (R^2) for sulfur increases from the original value of 0.86 to 0.88, whereas the R^2 for carbon decreases from 0.84 to 0.79. A comparative summary of these effects is presented in **Table 3**. The decline in R^2 for carbon may be attributed to the peak-identified data model no longer being optimally suited to the previously selected number of latent variables. Adjusting the number of latent variables to 4 resulted in an increase in the R^2 for the peak-based model to 0.87. **Figure 13** illustrates the predictive performance of the PLSR model, trained and tested using the peak data with the latent variables set to 4, for carbon concentration.

Table 3. Comparison of modeling effects before and after peak search.

Element Assessment	Stad	Peak spectral Stad	Cd	Peak spectral Cd
R^2	0.86	0.88	0.84	0.79
RMSECV	0.0062	0.0103	2.1282	2.1355
RMSEP	0.0041	0.0034	1.3852	1.7610

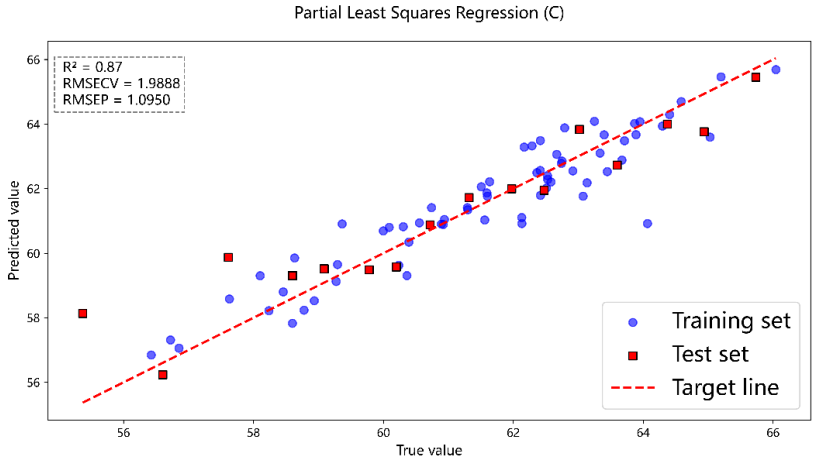

Figure 13. Predictive performance of carbon concentration using peak data for both training and test sets in the PLSR model with four latent variables.

4.3. Handling of Ash Content

 doi:10.20944/preprints202510.0947.v1

Ash is a detrimental component of coal, as it adversely affects both utilization and combustion efficiency [33]. When coal is used as a power fuel, the higher the ash content, the lower the content of combustible substances in coal. Thus, the ash content is related to multiple factors, which can be roughly classified into three categories: (1) primary minerals, i.e., the inorganic matters inherent in the plants forming coal; (2) secondary minerals, i.e., the minerals entering the coal seam during the coal formation process; (3) extraneous minerals, i.e., rock and gangue debris mixed in during coal mining. The components include, but are not limited to, SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO, SO_3, K_2O, Na_2O, etc., and may also contain trace heavy metals and harmful elements (e.g., As, Hg, Pb, etc.). Combustion of these components releases toxic gases (e.g., arsenic trioxide, mercury vapor, etc.), polluting the environment [34,35]. Additionally, the carbon content also affects the ash proportion: the higher the carbon content, the lower the ash content after complete combustion of coal. Given the numerous factors influencing ash content, manual selection of elements for ash prediction is complex and inefficient. Therefore, this experiment adopted variable recombination combined with the variable importance in the projection (VIP) index for modeling ash prediction. The VIP index is a key indicator in the PLSR model to measure the relative contribution of independent variables to explaining dependent variables.

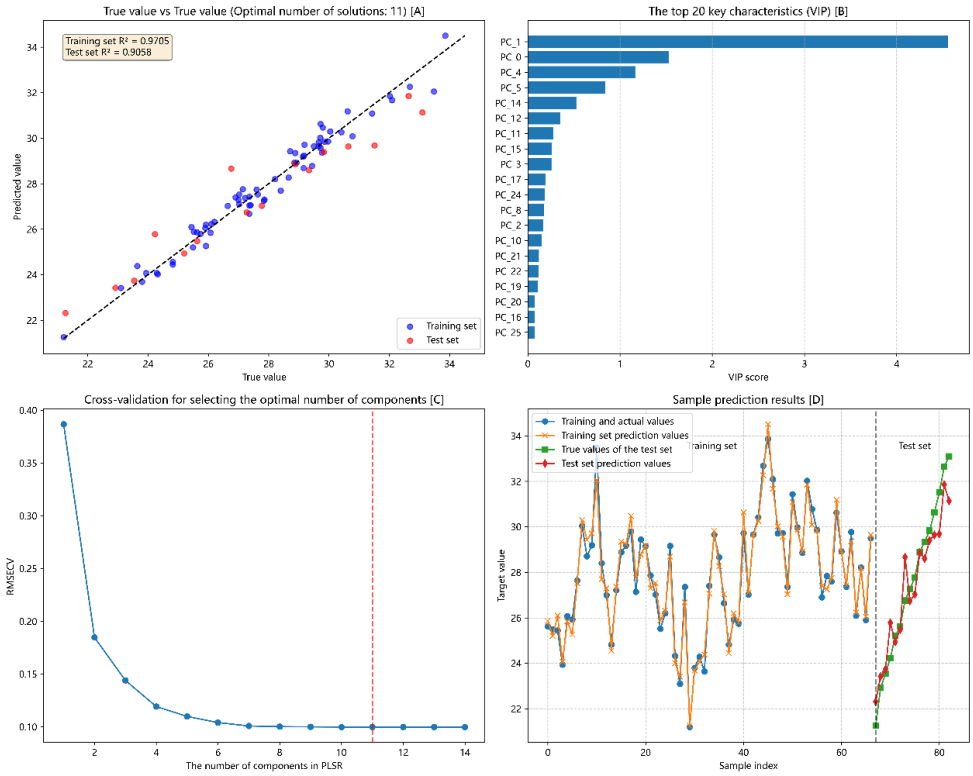

Figure 14. The final effect after reorganizing the gray-scale variables.

Figure 14A shows the ash content prediction results following variable reorganization. The coefficient of determination (R^2) for the test set reached 0.9058 after variable reorganization. Although this is lower than 0.92, it resolves the issue of manually selecting latent variables for each training session. **Figure 14B** illustrates the relative contributions of independent variables selected through the VIP index. **Figure 14C** depicts the optimal latent variables automatically matched by the model. This process automatically selects the best latent variables for different datasets based on changes in the RMSECV during cross-validation prior to model training. The figure indicates that the optimal latent variables are obtained when the RMSECV is 0.0998, and this approach improves the performance of the constructed model. **Figure 14D** demonstrates the predictive performance of the PLSR model on both the test and training datasets, visually comparing the prediction outcomes for these two sets within the model. The above model realizes the optimal selection through automated latent variable selection.

 doi:10.20944/preprints202510.0947.v1

5. Conclusions

This study integrates LIBS spectral data acquired from a prototype system with PLSR modeling to predict coal composition. Significant predictive performance was achieved using the full spectrum, supported by theoretical modeling grounded in LIBS principles. Given the substantial influence of latent variables on PLSR model performance and the inability to determine an optimal number uniquely for each dataset, an automated latent variable selection method is developed, demonstrated through variable reconfiguration for the ash component. Experimental results confirm the feasibility of LIBS–PLSR integration for coal quality analysis, offering superior detection efficiency and enhanced sample reusability compared with traditional analytical methods. Furthermore, the resource consumption associated with model-based prediction of coal constituents is significantly lower than that of conventional techniques. The combination of data elimination and baseline correction further enhances the accuracy and repeatability of ash content detection in pulverized coal particle streams [36]. However, the predictive performance for individual elements such as sulfur and carbon remains lower than that achieved for ash, which reflects a multi-element composition. Subsequent studies should therefore focus on model refinement to optimize predictions for specific elemental constituents.

Author Contributions: Rongzhou Zhang: Conceptualization, Data curation, methodology, Software, Validation, Visualization, Writing – original draft; Syed Zaheer Ud Din: Writing – review & editing, Formal analysis; Chunling Dang: Formal analysis; Xiangming Kong: Formal analysis ,Resources, Supervision; Rongxin Ma: Formal analysis; Jianli Ning: Formal analysis; Guangtao Fu: Formal analysis; Jiancai Leng: Formal analysis; Wenhao Zhang: Formal analysis, Resources, Funding acquisition, Supervision, Project administration, Writing – review & editing.

Funding: The research was funded by the Natural Science Foundation of Shandong Province (ZR2022QF083), the first batch of talent research project of Qilu University of Technology (Shandong Academy of Sciences) (2023RCKY033), the International Science and Technology Cooperation of Shandong Province (2025KJHZ031).

Data Availability Statement: The data involved in this paper cannot be publicly provided as it contains corporate trade secrets. If needed, you may contact the author to obtain the relevant data.

Acknowledgments: Some sentences in this paper have been grammar-checked using Tencent Yuanbao (https://yuanbao.tencent.com).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Qian, Y.; Zhong, S.; He, Y.; Whiddon, R.; Wang, Z.H.; Cen, K.F. Effects of Laser Wavelength on Properties of Coal LIBS Spectrum. *Spectrosc Spect Anal* **2017**, *37*, 1890-1895.
2. Kim, C.K.; In, J.H.; Lee, S.H.; Jeong, S. Independence of elemental intensity ratio on plasma property during laser-induced breakdown spectroscopy. *Opt. Lett* **2013**, *38*, 3032-5.
3. Agrawal, N.; Govil, H. A deep residual convolutional neural network for mineral classification. *Adv. Space Res* **2023**, *71*, 3186-3202.
4. Guan, F.Y.; Liu,Y.C.; Niu, X.C.; Huang, W.H.; Li, W.; Zheng, P.C.; Zhang, D.; Xu, G.; Guo, L.B. AI-enabled universal image-spectrum fusion spectroscopy based on self-supervised plasma modeling. *AP Nexus* **2024**, *3*, 127-139.
5. Azmat, F.; Chen, Y.F.; Stocks, N. Analysis of Spectrum Occupancy Using Machine Learning Algorithms. *IEEE TVT* **2016**, *65*, 6853 – 6860.
6. Haykiri-Açma, H.; Ersoy-Meriçboyu, A.; Küçükbayrak, S. Combustion reactivity of different rank coals. *Energy Convers. Manage* **2002**, *43*, 0196-8904.
7. Ciucci, A.; Palleschi, V.; Rastelli, S.; Barbini, R.; Colao, F.; Fantoni, R.; Palucci, A.; Ribezzo, S.; van der Steen, H.J.L. Trace pollutants analysis in soil by a time-resolved laser-induced breakdown spectroscopy technique. *Appl. Phys. B* **1996**, *63*, 185-190.
8. Bousquet, B.; Sirven, J.B.; Canioni, L. Towards quantitative laser-induced breakdown spectroscopy analysis of soil sample. *SPECTROCHIM ACTA B* **2007**, *62*, 1582-1589.

 doi:10.20944/preprints202510.0947.v1

9. Martin, M.Z.; Wullschleger, S.D.; Garten, C.T.; Palumbo, A.V. Laser-induced breakdown spectroscopy for the environmental determination of total carbon and nitrogen in soil. *Appl. Opt* **2003**, *42*, 2072-7.
10. Eppler, A.S.; Cremers, D.A.; Hickmott, D.D.; Ferris, M.J.; Koskelo, A.C. Matrix Effects in the Detection of Pb and Ba in Soils Using Laser-Induced Breakdown Spectroscopy. *Appl. Spectrosc* **1996**, *50*, 1175-1181.
11. Legnaioli, S.; Campanella, B.; Pagnotta, S.; Poggialini, F.; Palleschi, V.; Determination of Ash Content of coal by Laser-Induced Breakdown Spectroscopy. *SPECTROCHIM ACTA B* **2019**, *155*, 123-126.
12. Pokrajac, D.; Lazarevic, A.; Kecman, V.; Marcano, A.; Markushin, Y.; Vance,T.; Reljin, N.; McDaniel, S.; Melikechi, N. Automatic Classification of Laser-Induced Breakdown Spectroscopy (LIBS) Data of Protein Biomarker Solutions. *Appl. Spectrosc* **2014**, *68*, 1067-75.
13. Diedrich, J.; Rehse, S.J.; Palchaudhuri, S. Pathogenic Escherichia coli strain discrimination using laser-induced breakdown spectroscopy. *J. Appl. Phys* **2007**, *102*, 6184.
14. Samek, O.; Beddows, D.C.S.; Telle, H.H.; Morris, G.W.; Liska, M.; Kaiser, J. Quantitative analysis of trace metal accumulation in teeth using Laser-Induced Breakdown Spectroscopy. *Appl. Phys. A* **1999**, *69*, S179-S182.
15. Chen, X.; Zhang, Y.; Li, X.; Yang, Z.; Liu, A.; Yu, X. Diagnosis and staging of multiple myeloma using serum-based laser-induced breakdown spectroscopy combined with machine learning methods. *Biomed Opt Express.* **2021**, *12*, 3584-3596.
16. Dyar, M.D.; Carmosino, M.L.; Breves, E.A.; Ozanne, M.V.; Clegg, S.M.; Wiens, R.C. Comparison of partial least squares and lasso regression techniques as applied to laser-induced breakdown spectroscopy of geological samples. *Acta Pt. B-Atom. Spectr* **2012**, *70*, 0584-8547.
17. Portnov, A.; Rosenwaks, S.; Bar, I. Emission following laser-induced breakdown spectroscopy of organic compounds in ambient air. *Appl Opt.* **2003**, *42*, 2835-42.
18. Bousquet, B.; Travaillé, G.; Ismaël, A.; Canioni, L.; Michel-Le Pierrès, K.; Brasseur, E.; Roy, S.; le Hecho, I.; Larregieu, M.; Tellier, S.; Potin-Gautier, M.; Boriachon, T.; Wazen, P.; Diard, A.; Belbèze, S. Development of a mobile system based on laser-induced breakdown spectroscopy and dedicated to in situ analysis of polluted soils. *Acta Pt. B-Atom. Spectr* **2008**, *63*, 1085-1090.
19. Arp, Z.A.; Cremers, D.A.; Wiens, R.C.; Wayne, D.M.; Sallé, B.A.; Maurice, S. Analysis of water ice and water ice/soil mixtures using laser-induced breakdown spectroscopy: Application to Mars polar exploration, *Appl. Spectrosc* **2004**, *58*, 897–909.
20. Lanza, N.L.; Wiens, R.C.; Clegg, S.M.; Ollila, A.M.; Humphries, S.D.; Newsom, H.E.; Barefield, J.E.; Calibrating the ChemCam laser-induced breakdown spectroscopy instrument for carbonate minerals on Mars. *Appl. Opt* **2010**, *49*, C211-C217.
21. Sallé, B.; Lacour, J.L.; Vors, E.; Fichet, P.; Maurice, S.; Cremers, D.A.; Wiens, R.C. Laser-Induced Breakdown Spectroscopy for Mars surface analysis: capabilities at stand-off distances and detection of chlorine and sulfur elements. *Acta Pt. B-Atom. Spectr* **2004**, *59*, 1413-1422.
22. Tzortzakis, S.; Anglos, D.; Gray, D. Ultraviolet laser filaments for remote laser-induced breakdown spectroscopy (LIBS) analysis: applications in cultural heritage monitoring. *Opt. Lett* **2006**, *31*, 1139-1141.
23. Giakoumaki, A.; Melessanaki, K.; Anglos, D. Laser-induced breakdown spectroscopy (LIBS) in archaeological science-applications and prospects. *Anal Bioanal Chem* **2007**, *387*, 749-60.
24. Fortes, F.J.; Ctvrtnícková, T.; Mateo, M.P.; Cabalín, L.M.; Nicolas, G.; Laserna, J.J. Spectrochemical study for the in situ detection of oil spill residues using laser-induced breakdown spectroscopy. *Anal. Chim. Acta* **2010**, *683*, 52-57.
25. Myakalwar, A.K.; Sreedhar, S.; Barman, I.; Dingari, N.C.; Rao, S.V.; Kiran, P.P.; Tewari, S.P.; Kumar, G.M. Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis. *Talanta* **2011**, *87*, 53-9.
26. Zhang, T.L.; Shan, W.U.; Tang, H.S.; Wang, K.; Duan, Y.X.; Li, H. Progress of Chemometrics in Laser-induced Breakdown Spectroscopy Analysis. *CHINESE J ANAL CHEM* **2015**, *43*, 939-948.
27. Parmar, D.; Srivastava, R.; Baruah, P.K. Laser induced breakdown spectroscopy: A robust technique for the detection of trace metals in water. *Mater. Today* **2023**, *77*, 234-239.

 doi:10.20944/preprints202510.0947.v1

28. Yao, S.C.; Xu, J.L.; Dong, X.; Zhang, B.; Zheng, J.P.; Lu, J.D. Optimization of laser-induced breakdown spectroscopy for coal powder analysis with different particle flow diameters. *SPECTROCHIM ACTA B* **2015**, *110*, 146-150.
29. Legnaioli, S.; Campanella, B.; Pagnotta, S.; Poggialini, F.; Palleschi, V. Determination of Ash Content of coal by Laser-Induced Breakdown Spectroscopy. *SPECTROCHIM ACTA B* **2019**, *155*, 123-126.
30. Yan, C.H.; Qi, J.; Ma, J.X.; Tang, H.S.; Zhang, T.L.; Li, H. Determination of carbon and sulfur content in coal by laser induced breakdown spectroscopy combined with kernel-based extreme learning machine. *Chemom. Intell. Lab. Syst.* **2017**, *167*, 226-231.
31. Wold, S.; Ruhe, A.; Wold, H.; Dunn, I.W.J. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J SCI COMPUT.* **1984**, *5*, 735-743.
32. Liu, J.M.; Wu, D.; Fu, C.L.; Hai, R.; Yu, X.; Sun, L.Y.; Ding, H.B. Improvement of quantitative analysis of molybdenum element using PLS-based approaches for laser-induced breakdown spectroscopy in various pressure environments. *Plasma Sci. Technol* **2019**, *21*, 034017.
33. Kurose, R.; Ikeda, M.; Makino, H. Combustion characteristics of high ash coal in a pulverized coal combustion. *Fuel* **2001**, *80*, 1447-1455.
34. Selçuk, N.; Gogebakan, Y.; Gogebakan, Z. Partitioning behavior of trace elements during pilot-scale fluidized bed combustion of high ash content lignite. *J. Hazard. Mater* **2006**, *137*, 1698-1703.
35. Lee, M.G.; Yi, G.; Ahn, B.J.; Roddick, F. Conversion of Coal Fly Ash into Zeolite and Heavy Metal Removal Characteristics of the Products. *Korean J. Chem. Eng* **2000**, *17*, 325-331.
36. Wallis, F.J.; Chadwick, B.L.; Morrison, R.J.S. Analysis of Lignite Using Laser-Induced Breakdown Spectroscopy. *Applied Spectroscopy.* **2000**, *54*, 1231-1235.