

# Comparative Study of Four Chemometric Methods for the Quantitative Analysis of the Carbon Content in Coal by Laser-Induced Breakdown Spectroscopy Technology

Qi Ni, Yanqun Zhu,\* Wenyu Zhu, Yong He, and Zhihua Wang



Cite This: *ACS Omega* 2022, 7, 9443–9451

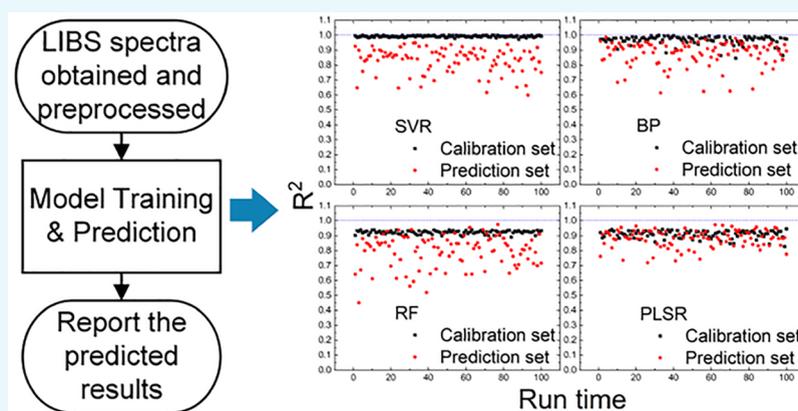


Read Online

ACCESS |

Metrics & More

Article Recommendations



**ABSTRACT:** Coal is a heterogeneous mineral substance mainly composed of carbon, along with various amounts of other elements. The carbon content is an important and pertinent parameter for coal quality. To achieve the rapid and accurate online measurement of the carbon content in coal, four different calibration strategies are applied to coal analysis by laser-induced breakdown spectroscopy (LIBS). Four calibration models based on support vector regression (SVR), back-propagation training (BP), random forest (RF), and partial least-squares regression (PLSR) were proposed, and the prediction accuracy, prediction precision, model stability, and training velocity of the four calibration models were compared for the quantitative analysis of the carbon content. A total of 65 coal samples were ablated, and the plasma spectra were used as the input data. Among the four calibration models, the results indicate that SVR and BP are the most promising calibration models for finding a better optimized model with a better prediction accuracy and prediction precision, and PLSR has a better prediction stability and a faster training velocity; however, RF has a prediction performance worse than those of the other three models.

## 1. INTRODUCTION

Although with the great demand to reduce CO<sub>2</sub> emissions from fossil fuel combustion has become a major task for every government to meet the global warming crisis, switching from the traditional energy system to a renewable one requires time and world-wide efforts. Fossil fuels still play an important role in the current energy system. Coal still acts as the dominant fuel for power generation at a global level, especially for developing countries. The share of coal for power generation in the world was still 35.1% in 2020.<sup>1</sup> According to the BP Statistical Review of World Energy 2021, the world coal reserves in 2020 stood at 1074 billion tones, with the top four coal reserve nations being the USA (23%), Russia (15%), Australia (14%), and China (13%).<sup>1</sup> If the combustion efficiency of the existing power station could be greatly improved, it would make a great contribution to reducing CO<sub>2</sub>

emissions. Therefore, online operation optimization has become more and more important, while the fuel quality online monitoring technique has become the short board. Calorific value is a basic characteristic of fuel, which is very important for combustion adjustment. The calorific value of coal depends mostly on the carbon content, and measuring the carbon content is useful for coal quality analysis. Considering the large amount of coal used in both industry and power generation, the real-time measurement of the carbon content

**Received:** November 30, 2021

**Accepted:** February 23, 2022

**Published:** March 8, 2022



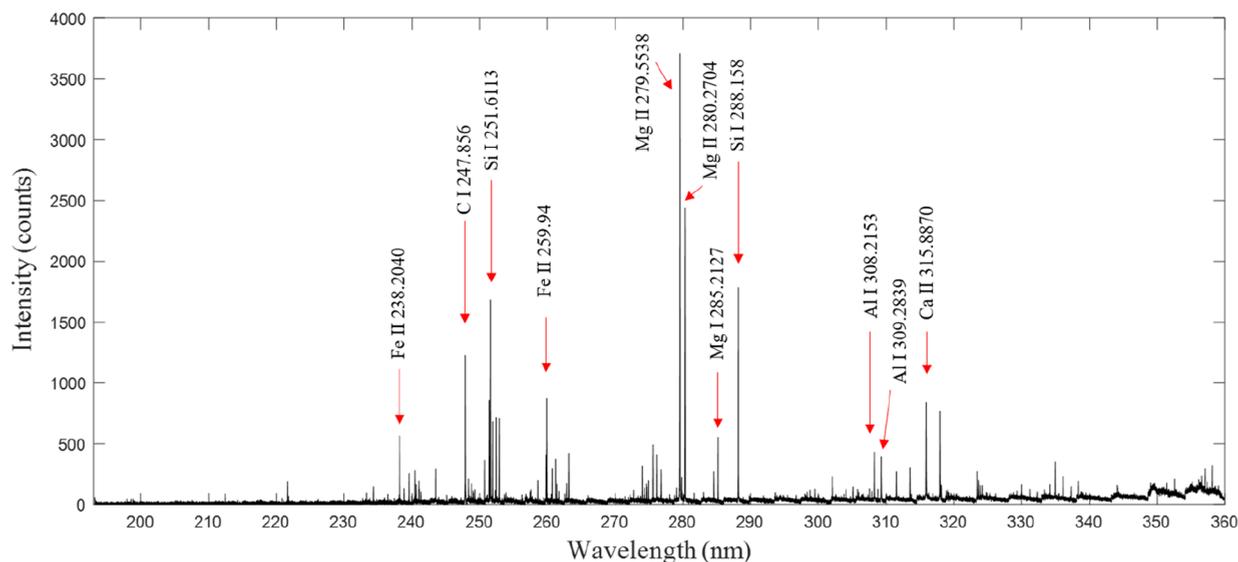


Figure 1. LIBS spectrum of a coal sample.

of coal is of great significance for combustion optimization, fuel quality monitoring, and pollution abatement.

The development of laser-based optical technology, namely laser-induced breakdown spectroscopy (LIBS), as a potential online measurement technology has received a lot of attention in the field of coal quality analysis.<sup>2–4</sup> LIBS as an optical measurement method has several advantages, such as simultaneous and fast in situ multielemental analysis, minimal sample preparation, and better security.<sup>5–7</sup> During the LIBS measurement, a laser beam is focused on the sample surface to generate a plasma spark, and the emission spectra are collected during the plasma cooling process and analyzed.<sup>8</sup> Because of matrix effects and the fluctuation of experimental conditions,<sup>9</sup> there is not a simple linear relationship between the carbon content and the signal intensity of carbon.

A lot of analysis work for coal has been conducted using LIBS, including elemental analysis (C, H, N, and S)<sup>10–13</sup> and proximate analysis (calorific value, ash content, fixed carbon content, and moisture content).<sup>14–17</sup> To improve the accuracy and precision of the LIBS quantitative analysis, researchers have carried out a lot of studies on the experimental setup, operating conditions, and different calibration models to enhance the spectral signal and improve the accuracy of the quantitative analysis.<sup>18,19</sup> The enhanced spectral signal can increase the signal-to-noise ratio of the spectrum and thus improve the accuracy of measurement. Main methods to enhance the spectral signal include ambient conditions,<sup>20,21</sup> microwaves,<sup>22,23</sup> spatial confinement,<sup>24,25</sup> double-pulse,<sup>26,27</sup> etc.

Aside from improving the LIBS signal, the calibration model between the line intensity and the element content of coal is also important. Methods used to improve the accuracy of the calibration model include partial least-squares regression (PLSR), support vector regression (SVR), principle component analysis (PCA), artificial neural network (ANN), etc.<sup>28–30</sup> It is of great significance to evaluate and improve the performance of different chemometric methods. Wang et al.<sup>31</sup> presented a multivariate model based on the dominant factor for LIBS that combined the advantages of both the conventional univariate and PLS models. Zhang et al.<sup>32</sup> compared the modeling efficiencies and prediction accuracies

of four calibration models for the quantitative analysis of ash, volatile matter, and the calorific value of coal based on PLSR, SVR, ANN, and PCR and found ANN could offer the best compromise between modeling efficiency and prediction accuracy. Wei et al.<sup>33</sup> used wavelet neural network (WNN) and ANN for the quantitative analysis of the major components in coal ash and found that the WNN model had a better performance than the ANN model. These studies show that the quantitative analysis of coal quality by LIBS combined with suitable calibration models can be achieved with a high degree of accuracy. However, the performance of the calibration model is related to different input variables and optimization methods. To realize the real-time measurement of the carbon content of coal, it is necessary to carry out comparative studies of different calibration models and select a suitable model to achieve a high measurement accuracy.

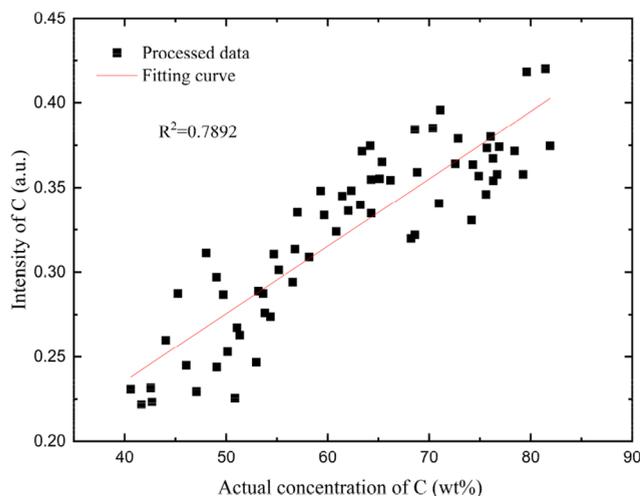
In this study, a LIBS setup was established, and the spectral data of 65 coal samples were acquired for the quantitative analysis of the carbon content of coal. Four calibration models based on SVR, back-propagation neural network (BP), random forest (RF), and PLSR were used to build the relationship between the carbon content and line intensities, and the performances of the four calibration models were evaluated and further discussed.

## 2. RESULTS AND DISCUSSIONS

**2.1. Data Preprocessing.** To measure the carbon content, the carbon atomic line at 247.856 nm (C I) was selected as the characteristic spectral line of carbon based on information from the National Institute of Standards and Technology (NIST). The original LIBS spectrum of a coal sample, which was acquired after 30 laser shots, is shown in Figure 1.

Before training the model, the spectral data need to be preprocessed. To reduce the signal fluctuations due to changes in experimental parameters and the matrix effect of coal, the input spectral lines were preprocessed with local spectral normalization to get the normalized line intensities. In local spectral normalization, the peak area of spectra line was divided by the integral area of the waveband in which the spectral line was located. Every waveband has a width of around 3 nm. For the normalization of the C I 247.856 nm

line, the spectral range of the waveband is between 245.400 and 248.130 nm. As shown in Figure 2, the correlation coefficients ( $R^2$ ) of the fitting curve increased from 0.0167 to 0.7892 after preprocessing.



**Figure 2.** Fitting curve of the carbon content with the intensity of the CI 247.856 nm line.

**2.2. Selection of Characteristic Spectral Lines.** To improve the prediction accuracy, we introduced spectral lines of other elements within the spectral range of 190–350 nm as input data. The introduction lines selected are shown in Table 1. In the spectrum of coal, the main spectral lines in the

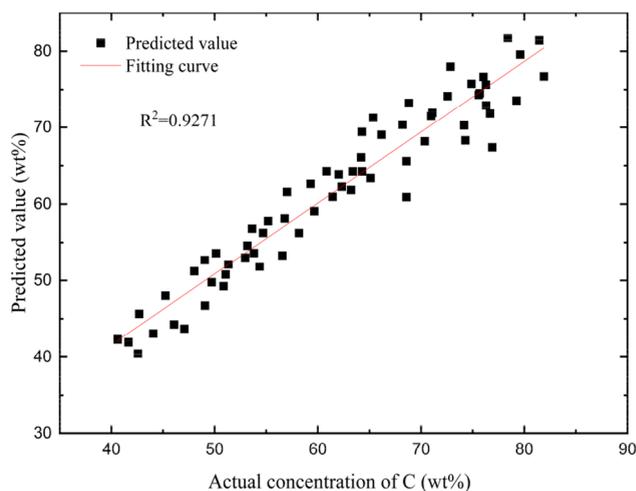
**Table 1. Selected Characteristic Spectral Lines of Different Elements for Model Establishment**

	element spectral emission lines (nm)
C	247.86
Si	221.67, 250.69, 251.43, 251.61, 251.92, 252.41, 252.85, 288.16
Mg	279.55, 280.27, 285.21
Ca	315.89, 317.93
Al	308.22, 309.28
Fe	238.20, 239.56, 240.49, 248.81, 258.59, 259.84, 259.94, 260.71, 261.19, 273.96, 275.57, 323.46, 357.03, 358.12

spectral range of 190–350 nm are carbon lines, Mg lines, Fe lines, and Si lines, followed by some other lines that are independent, distinguishable, and detectable, as shown in Figure 1. PLSR was used to build the linear relationship between the intensity of the spectral lines and the carbon content. After the introduction of multiple spectral lines in the spectral range of 190–350 nm, the  $R^2$  of the calibration curve increased from 0.7892 to 0.9271, as shown in Figure 3.

Therefore, the introduction of multiple spectral lines in spectral range of 190–350 nm is helpful to improve the prediction accuracy of the calibration models.

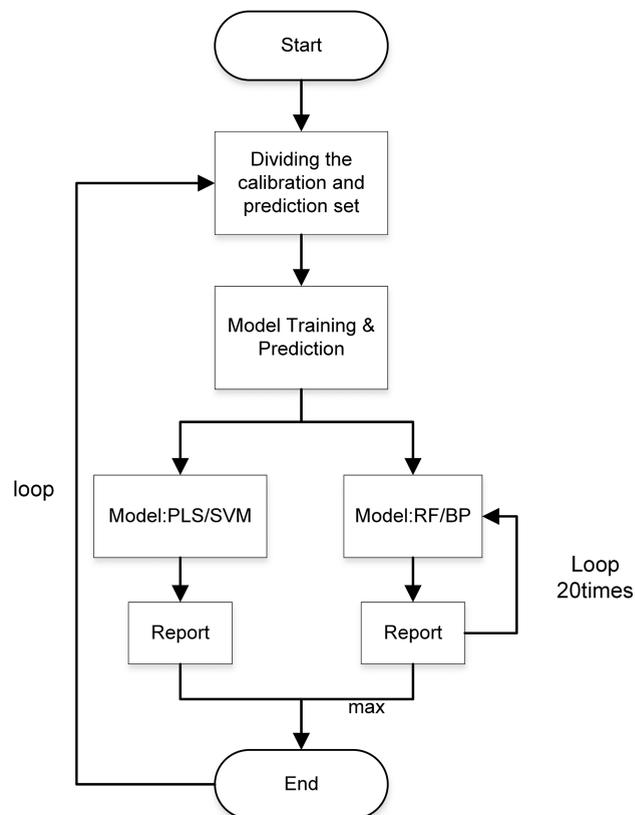
**2.3. Comparison of Different Models.** Comparisons among the prediction accuracy and the model training velocity of the four calibration models (SVR, BP, RF, and PLSR) were performed. When training the four models, we chose 52 samples as the calibration set and 13 samples as the prediction set. The input data for the four models were the intensities of the characteristic spectral lines of C, Si, Mg, Ca, Al, and Fe mentioned in Table 1 and the carbon content of coal samples



**Figure 3.** Predicted results for the carbon content with input lines in the spectral range of 190–350 nm.

measured before the experiment. The output data were the predicted values of the carbon content. All spectral lines were preprocessed, and the key parameters of the four models were optimized. To evaluate the performance of the four calibration models,  $R^2$ , the root-mean-square error of calibration (RMSEC), the root-mean-square error of prediction (RMSEP), and the model training times of the four models were calculated and compared.

Figure 4 shows the flowchart of the training process. For BP and RF, the model with the maximum  $R^2$  for the prediction set among the 20 run times was selected as the final optimized



**Figure 4.** Flowchart of the model training process.

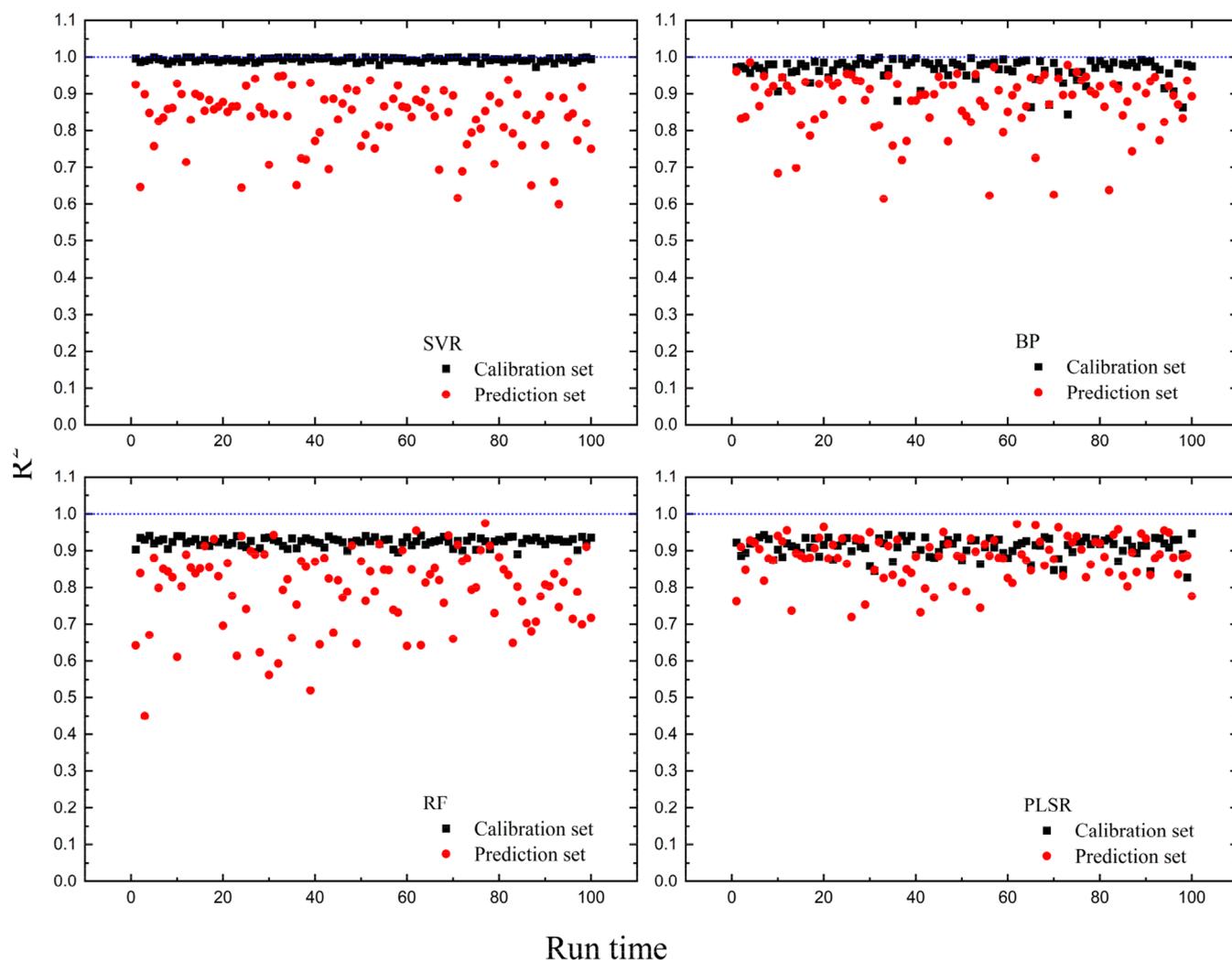


Figure 5. Predicted results from the four different calibration models.

Table 2. Results of the Different Calibration Models for the Carbon Content under Different Calibration Sets and Prediction Sets

algorithm	calibration set					prediction set					
	$R^2 > 0.90$ (%)	$R^2 > 0.80$ (%)	$R_{\max}^2$	$R_{\min}^2$	RMSEC <sub>avg</sub>	$R^2 > 0.90$ (%)	$R^2 > 0.80$ (%)	$R^2 > 0.70$ (%)	$R_{\max}^2$	$R_{\min}^2$	RMSEP <sub>avg</sub>
SVR	100	100	0.99	0.97	0.04	16	72	90	0.95	0.60	0.24
BP	95	100	0.99	0.85	0.10	47	85	94	0.99	0.61	0.21
RF	97	100	0.94	0.89	0.17	14	58	80	0.97	0.45	0.27
PLSR	69	100	0.95	0.83	0.17	41	90	100	0.97	0.72	0.21

model. The  $R^2$  and corresponding values of RMSEC and RMSEP for the optimized model were calculated as the final results for BP and RF.

To better evaluate the prediction abilities of the four calibration models, the results with both the selected random prediction set and the calibration set were calculated and compared.

**2.3.1. Randomly Selected Calibration Set and Prediction Set.** To better describe the prediction accuracy and stability of the four calibration models, the calibration set and the prediction set were randomly selected each time. Every model was run 100 times, and the  $R^2$  and RMSE values of each model were calculated. The  $R^2$ , RMSEC, and RMSEP values of the BP and RF models were calculated as mentioned

in the previous section. After changing the calibration set and the prediction set, a different distribution of  $R^2$ , RMSEC, and RMSEP values for the calibration set and the prediction set can be seen when different calibration models are used. Figure 5 provides the distribution of  $R^2$  values for the calibration set and the prediction set of four calibration models in 100 run times. For a more intuitive view, we counted the maximum, minimum, and average value of  $R^2$  for the 100 run times. The statistical results of the four calibration models are listed in Table 2.

It can be clearly seen from Figure 5 that the  $R^2$  values of the calibration set were more stable than those of the prediction set, and the  $R^2$  values for the calibration sets of SVR, BP, and RF were larger than those for the prediction set in most cases.

This may be due to overfitting caused by over training on the calibration set, resulting an overly complex model. Using more coal samples can effectively reduce overfitting.

$R^2$  values for the calibration sets of SVR and BP were larger than those of RF and PLSR in general. This shows that SVR and BP can find a more accurate and complex correlation between the intensities of multiple lines and the carbon content. SVR and BP as nonlinear calibration models are superior to the linear model PLSR. In addition, the average RMSEC and RMSEP values of SVR and BP were smaller than those of RF. This indicates that SVR and BP have better prediction precision.

From the percentage distribution of the  $R^2$  value for the prediction set in Table 2, the  $R^2$  value for the prediction set was larger than 0.80 for 72% of SVR results, 85% of BP results, 58% of RF results, and 90% of PLSR results, showing that BP and PLSR have prediction performances more stable than SVR and RF when changing the calibration set and the prediction set. In addition, 47% of results for the  $R^2$  value of the prediction set in the BP model were larger than 0.90, with a maximum value of 0.99, indicating that BP performed better than other models. The  $R^2$  value of the calibration set in the PLSR model was lower than those in the SVR and BP models. This means that the prediction ability of the PLSR model is limited. In the RF model, only 14% results for the  $R^2$  value of the prediction set were larger than 0.90, and the minimum value was 0.45, which was smaller than those of other models. This indicates that the RF model performed worse relative to other models, with a lower prediction accuracy and a poorer generalization performance.

The PLSR model had the best stability among the four models. This means that PLSR is less dependent on input data because PLSR is a linear calibration model and there is a strong positive correlation between the carbon content and the spectral line intensity of carbon.

In summary, BP and PLSR had better model stabilities, SVR and BP could achieve better training performances, and RF performed worse than other models when the calibration set and the prediction set were changed.

**3.3.2. Selected Calibration Set and Prediction Set.** For a more visual analysis, the total 65 samples were divided into 13 groups according to the carbon content, and one sample from each group was selected as the prediction set. Then, the rest were selected as the calibration set. Each of the four models was trained with the same calibration set and prediction set, and the results were compared.

The final result is shown in Table 3. In general, the results with the selected calibration set and prediction set were in good agreement with the results with a randomly selected calibration set and prediction set. The  $R^2$  values of the prediction set for all models were above 0.91, among which

SVR and BP performed better. The  $R^2$  values of the calibration set for SVR and BP were above 0.97, which demonstrated a high degree of fitness. On the other hand, the RMSEC and RMSEP values of SVR and BP were smaller than those of RF and PLSR. This means that the prediction precision and prediction accuracy of SVR and BP are superior to those of RF and PLSR.

It is worth noting that the  $R^2$  value of calibration set was larger than that of the prediction set for SVR and BP. Additionally, the  $R^2$  value of the calibration set for PLSR was 0.87, which was lower than other calibration models.

To further evaluate the prediction abilities of the different calibration models, the model training time was recorded. The modeling time depends on the algorithm and the optimization method used in the calibration models. When training the four calibration models, the model training time for SVR was 6 s, that for BP was 15 s, that for RF was 14 s, and that for PLSR was 1 s for each calculation. However, it is difficult to achieve a good result in just one run for the BP and RF models, so we ran the BP and RF models 20 times. The model training time for the BP model changed little when the run time increased 20, but the model training time for the RF model increased to about 45 s. It can be seen that PLSR and SVR had faster training velocities, BP ranked second, and RF was the most time consuming.

With this selected calibration set and prediction set, SVR and BP performed better than RF and PLSR considering the prediction precision and prediction accuracy, while SVR and PLSR had faster training velocities. Among the four models, SVR and BP are more likely to find a better relationship between the carbon content and line intensities with a higher prediction accuracy and precision. Considering that BP has a better prediction stability with an acceptable training velocity, BP is a better choice for the quantitative analysis of the carbon content in coal. Meanwhile, when the amount of input data is too small to train the model well, PLSR can be used to establish a linear relationship between the carbon content and line intensities with an acceptable accuracy.

### 3. CONCLUSIONS

In this study, a LIBS-based online experimental setup for the analysis of coal samples was established. Four calibration models (SVR, BP, RF, and PLSR) were employed for the quantitative analysis of the carbon content of coal, and the performances of four calibration models were compared and evaluated. The results show that SVR and BP are more promising calibration models for finding a better regression between input line intensities and the carbon content, PLSR has a better prediction stability and training velocity, and RF has a performance worse than those of the other three models.

In summary, when the amount of data is small it is more suitable to choose the PLSR model considering its better stability, and when the amount of data is large enough it is more suitable to choose BP model considering its better prediction accuracy and prediction precision. In this work, four calibration models were built, developed, and compared, demonstrating that the LIBS technique with appropriate calibration models was a good way to achieve the online analysis of the carbon content.

**Table 3. Results of the Different Calibration Models for the Carbon Content with the Same Calibration Set and Prediction Set**

algorithm	calibration set		prediction set	
	$R^2$	RMSEC	$R^2$	RMSEP
SVR	0.99	0.04	0.95	0.17
BP	0.97	0.10	0.95	0.16
RF	0.92	0.18	0.91	0.28
PLSR	0.87	0.19	0.92	0.19

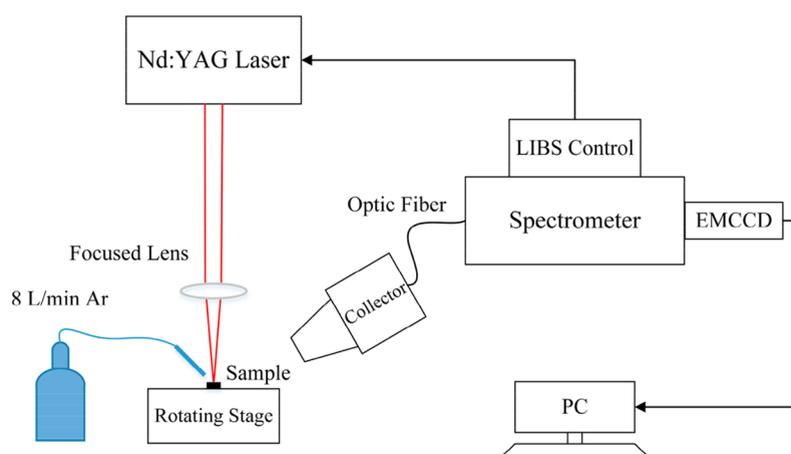


Figure 6. Schematic of the experimental setup.

## 4. EXPERIMENT AND METHODS

**4.1. Experimental Setup.** The experimental system consists of a Q-switched Nd:YAG laser (PRO-250–10H, Spectra Physics) with an output wavelength of 532 nm. The laser pulse energy was set to be 115 mJ, the pulse duration was 9 ns, and the repeat frequency was 10 Hz. The laser beam was focused on the sample surface using a quartz lens with focal length of 150 mm. The focal plane was positioned approximately 2 mm under the sample surface. The plasma emission was collected by a collector and coupled into the spectrometer (Aryelle Butterfly, LTB) equipped with an electron-multiplying CCD (EMCCD) camera (Andor) through a SMA fiber. The spectrometer has a wide wavelength range of 195–950 nm, with a high spectral resolution of 12 pm in the spectral range of 195–350 nm and a spectral resolution of 36 pm in the spectral range of 350–950 nm. To improve the signal-to-noise ratio of the spectra, highly purified argon (99.99%) was used as purge gas with a flow rate of 8 L/min. The coal sample was placed on a two-dimensional rotating stage. The schematic of the setup is shown in Figure 6.

In total, 65 coal samples were utilized in the experiment. The carbon content of these samples is shown in Table 4. The carbon content of the coal samples was analyzed according to the national standard of China (GB/T 476-2008). All coal samples were air-dried, crushed, and sieved to a size of 80–200 mesh. The coal powder was then pressed into pellets by a hydraulic press. The pressure was set to 25 MPa, and the pellets were kept under this pressure for 5 min. In total, 30 shots at different ablation points at the surface of the coal sample were accumulated for one spectrum, five spectra were obtained for each sample, and the average spectrum of the five spectra was used as the final spectrum. In this analysis, 52 coal samples were chosen for calibration and 13 coal samples were chosen for prediction.

**4.2. Methods.** **4.2.1. Support Vector Regression (SVR).** Support vector regression is a regression algorithm that works on the principle of the support vector machine with a few minor differences. The basic principle is to find a fitting curve from which the distance to the data point will be minimized. All SVM models in this study were implemented using the shareware program LibSVM that was developed by Chih-Chung Chang and Chih-Jen Lin.<sup>34</sup> The radial basis function (RBF) was used as the kernel function for nonlinear regression, and the key hyperparameters  $\gamma$  and  $C$  of the RBF kernel were

Table 4. Carbon Content of the Coal Samples

sample number	carbon content (%)	sample number	carbon content (%)
1	70.35	34	51.33
2	63.38	35	56.78
3	68.58	36	64.28
4	59.30	37	57.01
5	65.10	38	63.21
6	64.18	39	50.86
7	58.17	40	62.32
8	76.89	41	62.03
9	76.30	42	40.60
10	72.84	43	60.84
11	71.09	44	61.43
12	66.17	45	68.58
13	64.27	46	59.65
14	76.04	47	53.17
15	75.68	48	41.65
16	81.45	49	42.69
17	79.60	50	56.55
18	74.16	51	55.18
19	68.81	52	53.81
20	74.87	53	54.36
21	75.59	54	47.06
22	81.90	55	51.06
23	79.24	56	49.06
24	78.40	57	46.07
25	76.68	58	49.71
26	76.28	59	42.56
27	68.20	60	53.63
28	50.13	61	52.96
29	74.29	62	49.04
30	72.57	63	48.03
31	70.96	64	44.05
32	65.35	65	45.24
33	54.69		

optimized. Parameters  $\gamma$  and  $C$  took values within a certain range and were optimized by cross-validation. The mean squared error of calibration set with fivefold cross-validation was calculated to optimize the values of  $\gamma$  and  $C$ . Figure 7 provides the results of the mean squared error for different values of  $\gamma$  and  $C$ . When the mean square error was at a

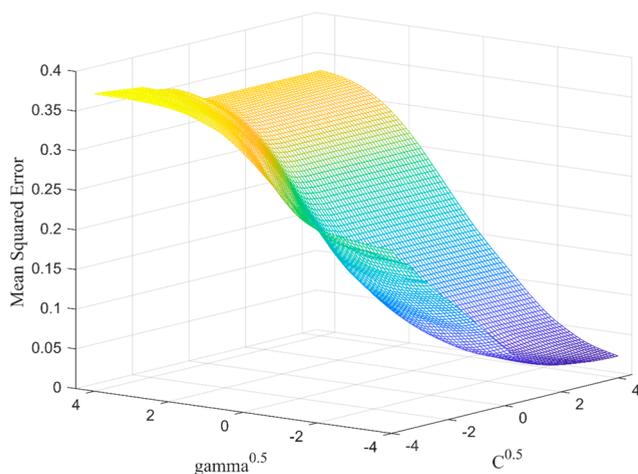


Figure 7. MSE vs gamma and C for SVR.

minimum, the values of  $\gamma$  and  $C$  were determined as the optimal parameters.

**4.2.2. Back Propagation (BP).** The BP neural network is a multilayer feed-forward network trained according to an error back-propagation algorithm.<sup>35</sup> In this study, the obtained optimal architecture of the BP neural network involved an input layer, a single implied layer, and an output layer. The learning rate was set to 1, the training times were set to 1000, the allowable error was 0.001, and the Levenberg–Marquardt algorithm was used for BP neural network training and prediction. In this study, the number of neurons ( $P$ ) in the hidden layer was decided by an experience formula<sup>36</sup>  $P = \sqrt{m + n} + \lambda$ , where  $m$  and  $n$  are the number of neurons in the input and output layers, respectively, and  $\lambda$  is a constant between 1 and 10. In this study, the number of neurons in the input and output layers were set at 30 and 1, respectively; therefore, the value of  $P$  was between 7 and 16. To further optimize the number of neurons in the hidden layer, the average mean squared error of the prediction set was calculated after 15 runs for each number of neurons in the hidden layer. Figure 8 provides the results for the average mean squared error of the prediction set with different numbers of neurons. The number of neurons in the hidden layer was selected when

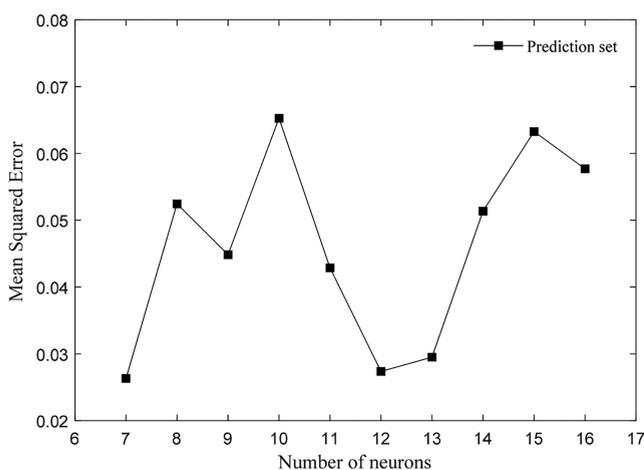


Figure 8. MSE vs the number of neurons in the hidden layer for BP.

the average mean squared error of the prediction set was a minimum in each run.

**4.2.3. Random Forest (RF).** Random forest is an ensemble learning algorithm and is commonly used to solve regression or classification problems. Random forest can handle very high dimensional data and a nonlinear relationship between predictors. In this study, we do regression with the TreeBagger in MATLAB to create bag of decision trees. There are some main parameters that need to be set before training:<sup>37</sup> (1) the number of predictor variables, which were randomly selected from the data (fboot); (2) the number of decision trees (ntree); and (3) the optimum terminal nodes (leaf). In this study, leaf = 5, 10, 15, 20, and 25; ntree = 50, 100, 200, 250, and 400; and fboot = 0.2, 0.4, 0.6, 0.8, and 1 were assumed, and the model was run in MATLAB R2019a software to find a best combination. As shown in Figure 9, the mean squared

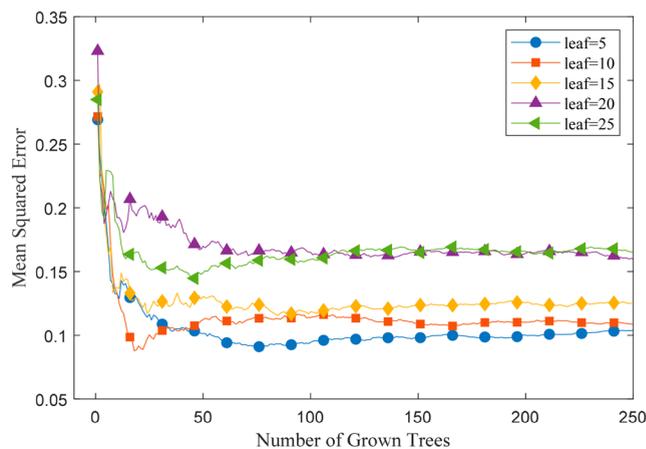


Figure 9. MSE vs the leaf size for RF.

errors obtained by regression for various leaf sizes were compared to verify the optimal leaf size and ntree values with fboot set to 1. After the optimal leaf size and ntree value was determined, the mean squared errors obtained by regression for various fboot values were calculated, as shown in Figure 10. When the MSE was at the minimum, the optimal leaf, ntree, and fboot values were determined and used to construct the

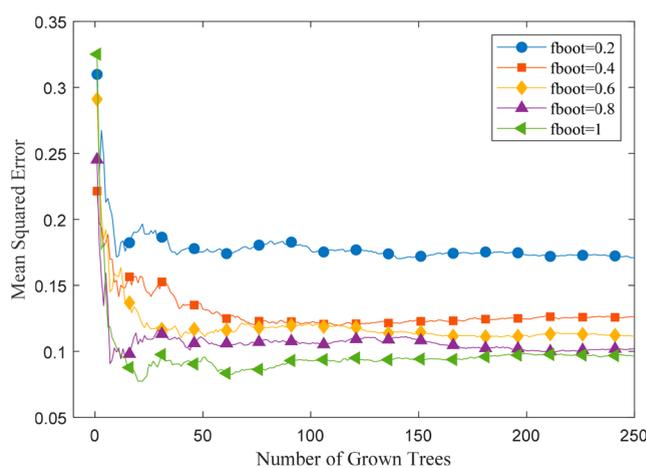


Figure 10. MSE vs the fboot for RF.

RF model. The best combination of parameters to generate the RF model is leaf = 5, ntrees = 250, and fboot = 1.

**4.2.4. Partial Least-Squares Regression (PLSR).** PLSR is a statistical method that integrates advantages of multiple linear regression analysis, principal component analysis, and typical correlation analysis. PLSR finds a linear regression model by projecting the input data into a new space. The number of principal components is a key parameter of PLSR. In this study, the number of principal components was selected by the variance explained in the input data with a fivefold cross-validation, the percentage of variance explained by the model containing the percentage of variance explained in predictor variables by each PLS component, and the percentage of variance explained in response variables. Figure 11 provides the

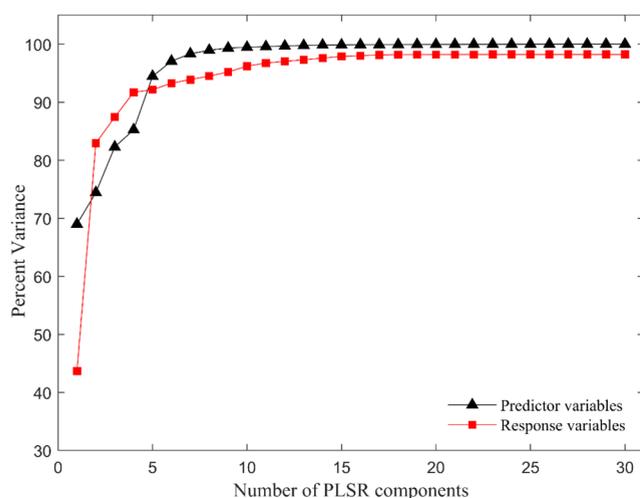


Figure 11. Percentage of variance vs the component for PLSR.

percentage of variance explained in the input data with different PLS components. When determining the optimal number of principle components, both the cumulative percentage of variance explained in predictor variables and that in response variables need to be greater than 0.9.

**4.3. Evaluation Index for the Calibration Models.** To evaluate the performance of the above-mentioned four models, the correlation coefficients ( $R^2$ ), the root-mean-square error of calibration (RMSEC), and the root-mean-square error of prediction (RMSEP) were employed to represent prediction accuracy and precision of the calibration models; the model training time was used to evaluate the model training velocity.  $R^2$  and RMSE are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (1)$$

where  $N$  is the number of samples,  $y_i$  is the carbon content of sample  $i$  measured before the experiments,  $\bar{y}_i$  is the average carbon content of the samples, and  $\hat{y}_i$  is the predicted carbon content.

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^{n_1} (y_i - \hat{y}_i)^2}{n_1}} \quad (2)$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n_2} (y_i - \hat{y}_i)^2}{n_2}} \quad (3)$$

where  $n_1$  is the number of samples for the calibration set and  $n_2$  is the number of samples for the prediction set.

$R^2$  reflects the correlation between the carbon content measured before the experiments and the predicted value of the carbon content. The RMSE reflects the deviation between the predicted value and the measured value for the carbon content. The closer  $R^2$  is to 1, the closer RMSEC and RMSEP are to 0, and the stronger the prediction ability of the model.

## AUTHOR INFORMATION

### Corresponding Author

Yanqun Zhu – State Key Laboratory of Clean Energy Utilization, Zhejiang University, 310027 Hangzhou, China; [orcid.org/0000-0002-0981-2078](https://orcid.org/0000-0002-0981-2078); Email: [yqzhu@zju.edu.cn](mailto:yqzhu@zju.edu.cn)

### Authors

Qi Ni – State Key Laboratory of Clean Energy Utilization, Zhejiang University, 310027 Hangzhou, China

Wenyu Zhu – State Key Laboratory of Clean Energy Utilization, Zhejiang University, 310027 Hangzhou, China

Yong He – State Key Laboratory of Clean Energy Utilization, Zhejiang University, 310027 Hangzhou, China;

[orcid.org/0000-0002-7037-9007](https://orcid.org/0000-0002-7037-9007)

Zhihua Wang – State Key Laboratory of Clean Energy Utilization, Zhejiang University, 310027 Hangzhou, China;

[orcid.org/0000-0002-7521-2900](https://orcid.org/0000-0002-7521-2900)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.1c06752>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This project was supported by Zhejiang Province Public Welfare Technology Application Analysis and Test Project (LGC19E060001) and the Fundamental Research Funds for the Central Universities (2021FZZX001-11).

## REFERENCES

- BP. *Statistical Review of World Energy 2021*. <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html> (accessed 2022-02-05).
- Yuan, T. B.; Wang, Z.; Lui, S. L.; Fu, Y. T.; Li, Z.; Liu, J. M.; Ni, W. D. Coal property analysis using laser-induced breakdown spectroscopy. *J. Anal. At. Spectrom.* **2013**, *28* (7), 1045–1053.
- Li, X.; Yin, H.; Wang, Z.; Fu, Y.; Li, Z.; Ni, W. Quantitative carbon analysis in coal by combining data processing and spatial confinement in laser-induced breakdown spectroscopy. *Spectrochim. Acta, Part B* **2015**, *111*, 102–107.
- Sheta, S.; Afgan, M. S.; Hou, Z.; Yao, S.-C.; Zhang, L.; Li, Z.; Wang, Z. Coal analysis by laser-induced breakdown spectroscopy: a tutorial review. *J. Anal. At. Spectrom.* **2019**, *34* (6), 1047–1082.
- Guo, L. B.; Hao, Z. Q.; Shen, M.; Xiong, W.; He, X. N.; Xie, Z. Q.; Gao, M.; Li, X. Y.; Zeng, X. Y.; Lu, Y. F. Accuracy improvement of quantitative analysis by spatial confinement in laser-induced breakdown spectroscopy. *Opt. Express*. **2013**, *21* (15), 18188–95.
- Gottfried, J. L.; Harmon, R. S.; De Lucia, F. C.; Miziolek, A. W. Multivariate analysis of laser-induced breakdown spectroscopy chemical signatures for geomaterial classification. *Spectrochim. Acta, Part B* **2009**, *64* (10), 1009–1019.

- (7) Jantzi, S. C.; Motto-Ros, V.; Trichard, F.; Markushin, Y.; Melikechi, N.; De Giacomo, A. Sample treatment and preparation for laser-induced breakdown spectroscopy. *Spectrochim. Acta, Part B* **2016**, *115*, 52–63.
- (8) Musazzi, S.; Perini, U., LIBS Instrumental Techniques. In *Laser-Induced Breakdown Spectroscopy: Theory and Applications*; Musazzi, S., Perini, U., Eds.; Springer-Verlag: Berlin, Germany, 2014; pp 59–86. DOI: 10.1007/978-3-642-45085-3.
- (9) Yu, Z. Y.; Yao, S. C.; Jiang, Y.; Chen, W. Z.; Xu, S. X.; Qin, H. Q.; Lu, Z. M.; Lu, J. D. Comparison of the matrix effect in laser induced breakdown spectroscopy analysis of coal particle flow and coal pellets. *J. Anal. At. Spectrom.* **2021**, *36*, 2473–2479.
- (10) Zhang, W.; Zhou, R.; Liu, K.; Yan, J.; Li, Q.; Tang, Z.; Li, X.; Zeng, Q.; Zeng, X. Sulfur determination in laser-induced breakdown spectroscopy combined with resonance Raman scattering. *Talanta* **2020**, *216*, 120968.
- (11) Zhong, S.; He, Y.; Qiu, K.; Wang, Z.; Li, B.; Liu, J.; Cen, K. Measurement of alkali content in Zhundong coal by LIBS method. *High Power Laser Part. Beams* **2015**, *27* (9), 099002.
- (12) Qian, Y.; Zhong, S.; He, Y.; Whiddon, R.; Wang, Z.; Cen, K. Effects of Laser Wavelength on Properties of Coal LIBS Spectrum. *Spectrosc. Spectral Anal.* **2017**, *37* (6), 1890–1895.
- (13) Li, J.; Lu, J. D.; Lin, Z. X.; Gong, S. S.; Xie, C. L.; Chang, L.; Yang, L. F.; Li, P. Y. Effects of experimental parameters on elemental analysis of coal by laser-induced breakdown spectroscopy. *Opt. Laser Technol.* **2009**, *41* (8), 907–913.
- (14) Li, W.; Dong, M.; Lu, S.; Li, S.; Wei, L.; Huang, J.; Lu, J. Improved measurement of the calorific value of pulverized coal particle flow by laser-induced breakdown spectroscopy (LIBS). *Anal. Methods* **2019**, *11* (35), 4471–4480.
- (15) Dong, M.; Lu, J.; Yao, S.; Li, J.; Li, J.; Zhong, Z.; Lu, W. Application of LIBS for direct determination of volatile matter content in coal. *J. Anal. At. Spectrom.* **2011**, *26* (11), 2183.
- (16) Yao, S.; Lu, J.; Dong, M.; Chen, K.; Li, J.; Li, J. Extracting coal ash content from laser-induced breakdown spectroscopy (LIBS) spectra by multivariate analysis. *Appl. Spectrosc.* **2011**, *65* (10), 1197–201.
- (17) Legnaioli, S.; Campanella, B.; Pagnotta, S.; Poggialini, F.; Palleschi, V. Determination of Ash Content of coal by Laser-Induced Breakdown Spectroscopy. *Spectrochim. Acta, Part B* **2019**, *155*, 123–126.
- (18) Yang, G.; Lin, Q.; Ding, Y.; Tian, D.; Duan, Y. Laser induced breakdown spectroscopy based on single beam splitting and geometric configuration for effective signal enhancement. *Sci. Rep.* **2015**, *5*, 7625.
- (19) Hahn, D. W.; Omenetto, N. Laser-induced breakdown spectroscopy (LIBS), part II: review of instrumental and methodological approaches to material analysis and applications to different fields. *Appl. Spectrosc.* **2012**, *66* (4), 347–419.
- (20) Mateo, M. P.; Pinon, V.; Anglos, D.; Nicolas, G. Effect of ambient conditions on ultraviolet femtosecond pulse laser induced breakdown spectra. *Spectrochim. Acta, Part B* **2012**, *74–75*, 18–23.
- (21) Idris, N.; Pardede, M.; Jobilong, E.; Lie, Z. S.; Hedwig, R.; Suliyanti, M. M.; Kurniawan, D. P.; Kurniawan, K. H.; Kagawa, K.; Tjia, M. O. Enhancement of carbon detection sensitivity in laser induced breakdown spectroscopy with low pressure ambient helium gas. *Spectrochim. Acta, Part B* **2019**, *151*, 26–32.
- (22) Khumaeni, A.; Motonobu, T.; Katsuaki, A.; Masabumi, M.; Ikuo, W. Enhancement of LIBS emission using antenna-coupled microwave. *Opt. Express* **2013**, *21* (24), 29755–68.
- (23) Viljanen, J.; Sun, Z.; Alwahabi, Z. T. Microwave assisted laser-induced breakdown spectroscopy at ambient conditions. *Spectrochim. Acta, Part B* **2016**, *118*, 29–36.
- (24) Li, X. W.; Yin, H. L.; Wang, Z.; Fu, Y. T.; Li, Z.; Ni, W. D. Quantitative carbon analysis in coal by combining data processing and spatial confinement in laser-induced breakdown spectroscopy. *Spectrochim. Acta, Part B* **2015**, *111*, 102–107.
- (25) Zhao, S.; Gao, X.; Chen, A.; Lin, J. Effect of spatial confinement on Pb measurements in soil by femtosecond laser-induced breakdown spectroscopy. *Appl. Phys. B: Lasers Opt.* **2020**, *126* (1), No. 7, DOI: 10.1007/s00340-019-7354-1.
- (26) Babushok, V. I.; DeLucia, F. C.; Gottfried, J. L.; Munson, C. A.; Miziolek, A. W. Double pulse laser ablation and plasma: Laser induced breakdown spectroscopy signal enhancement. *Spectrochim. Acta, Part B* **2006**, *61* (9), 999–1014.
- (27) Gautier, C.; Fichet, P.; Menut, D.; Lacour, J. L.; L'Hermite, D.; Dubessy, J. Study of the double-pulse setup with an orthogonal beam geometry for laser-induced breakdown spectroscopy. *Spectrochim. Acta B* **2004**, *59* (7), 975–986.
- (28) Yao, S. C.; Lu, J. D.; Li, J. Y.; Chen, K.; Li, J.; Dong, M. R. Multi-elemental analysis of fertilizer using laser-induced breakdown spectroscopy coupled with partial least squares regression. *J. Anal. At. Spectrom.* **2010**, *25* (11), 1733–1738.
- (29) Li, L.-N.; Liu, X.-F.; Yang, F.; Xu, W.-M.; Wang, J.-Y.; Shu, R. A review of artificial neural network based chemometrics applied in laser-induced breakdown spectroscopy analysis. *Spectrochim. Acta B* **2021**, *180*, 106183.
- (30) Yang, H. X.; Fu, H. B.; Wang, H. D.; Jia, J. W.; Sigrist, M. W.; Dong, F. Z. Laser-induced breakdown spectroscopy applied to the characterization of rock by support vector machine combined with principal component analysis. *Chin. Phys. B* **2016**, *25* (6), 065201.
- (31) Wang, Z.; Feng, J.; Li, L.; Ni, W.; Li, Z. A multivariate model based on dominant factor for laser-induced breakdown spectroscopy measurements. *J. Anal. At. Spectrom.* **2011**, *26* (11), 2289.
- (32) Zhang, Y.; Xiong, Z.; Ma, Y.; Zhu, C.; Zhou, R.; Li, X.; Li, Q.; Zeng, Q. Quantitative analysis of coal quality by laser-induced breakdown spectroscopy assisted with different chemometric methods. *Anal. Methods* **2020**, *12* (27), 3530–3536.
- (33) Wei, J.; Dong, J.; Zhang, T.; Wang, Z.; Li, H. Quantitative analysis of the major components of coal ash using laser induced breakdown spectroscopy coupled with a wavelet neural network (WNN). *Anal. Methods* **2016**, *8* (7), 1674–1680.
- (34) Chang, C.-C.; Lin, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2* (3), No. 27. Software available at <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- (35) Li, J.; Cheng, J.-H.; Shi, J.-Y.; Huang, F., Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement. In *Advances in Computer Science and Information Engineering*, Vol. 169; Jin, D., Lin, S., Eds.; Springer-Verlag: Berlin, Germany, 2012; pp 553–558. DOI: 10.1007/978-3-642-30223-7.
- (36) Shen, H.; Wang, Z.; Gao, C.; Qin, J.; Yao, F.; Xu, W. Determining the number of BP neural network hidden layer units. *J. Tianjin Univ. Technol.* **2008**, *24* (5), 13–15.
- (37) Neupane, A.; Raj, N.; Deo, R.; Ali, M. Development of data-driven models for wind speed forecasting in Australia. In *Predictive Modelling for Energy Management and Power Systems Engineering*. Deo, R., Samui, P., Roy, S. S., Eds.; Elsevier, 2020; pp 143–190. DOI: 10.1016/B978-0-12-817772-3.00006-9.