

# A PLS model based on dominant factor for coal analysis using laser-induced breakdown spectroscopy

Jie Feng · Zhe Wang · Logan West · Zheng Li · Weidou Ni

Received: 10 December 2010 / Revised: 24 February 2011 / Accepted: 1 March 2011 / Published online: 18 March 2011  
© Springer-Verlag 2011

**Abstract** Thirty-three bituminous coal samples were utilized to test the application of laser-induced breakdown spectroscopy technique for coal elemental concentration measurement in the air. The heterogeneity of the samples and the pyrolysis or combustion of coal during the laser-sample interaction processes were analyzed to be the main reason for large fluctuation of detected spectra and low calibration quality. Compared with the generally applied normalization with the whole spectral area, normalization with segmental spectral area was found to largely improve the measurement precision and accuracy. The concentrations of major element C in coal were determined by a novel partial least squares (PLS) model based on dominant factor. Dominant C concentration information was taken from the carbon characteristic line intensity since it contains the most-related information, even if not accurately. This dominant factor model was further improved by inducting non-linear relation by partially modeling the inter-element interference effect. The residuals were further corrected by PLS with the full spectrum information. With the physical-principle-based dominant factor to calculate the main quantitative information and to partially explicitly include

the non-linear relation, the proposed PLS model avoids the overuse of unrelated noise to some extent and becomes more robust over a wider C concentration range. Results show that RMSEP in the proposed PLS model decreased to 4.47% from 5.52% for the conventional PLS with full spectrum input, while  $R^2$  remained as high as 0.999, and RMSEC&P was reduced from 3.60% to 2.92%, showing the overall improvement of the proposed PLS model.

**Keywords** Laser-induced breakdown spectroscopy · Bituminous coal · Dominant factor · Partial least squares

## Introduction

Laser-induced breakdown spectroscopy (LIBS) is an atomic emission spectroscopy technique that employs laser-induced plasma as both the sample volume and the excitation source, dissociating all molecules and fine particulates within the highly energetic micro plasma. The resulting plasma emission can be associated with the elemental concentration in the sample. Because of LIBS' advantages, such as not needing pretreatment of the material, the speed of analysis and the possibility of in situ analysis, LIBS is attractive for industrial applications in a variety of fields [1–10].

In recent years, more and more researchers have focused on LIBS application in field monitoring for coal analysis and related issues in the coal-fired power plants. On-line determination of coal chemical composition prior to combustion is vitally important for a power plant to obtain optimal boiler control. The coal constituents not only impact pollutant emissions, but also the boilers, affecting combustion stability, corrosion, ash deposition and disposal [11]. However, current commercially available on-line

---

Published in the special issue *Laser-Induced Breakdown Spectroscopy* with Guest Editors Jagdish P. Singh, Jose Almirall, Mohamad Sabsabi, and Andrzej Miziolek.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00216-011-4865-y) contains supplementary material, which is available to authorized users.

---

J. Feng · Z. Wang · L. West · Z. Li (✉) · W. Ni  
State Key Lab of Power Systems,  
Department of Thermal Engineering,  
Tsinghua-BP Clean Energy Center, Tsinghua University,  
Beijing 100084, China  
e-mail: lz-dte@mail.tsinghua.edu.cn

techniques such as X-ray fluorescence (XRF) and prompt-gamma neutron activation analysis have strict requirements and difficulties in analyzing low atomic elements such as C and O [12]. Therefore LIBS, which offers fast and simultaneous multi-element analysis, has great potential for on-line coal measurement.

A number of studies have already probed the applicability of LIBS in coal analysis. Yin et al. [12] designed a LIBS system for quality analysis of pulverized coal and the measurement errors were less than 10%. Wallis et al. [13] investigated lignite samples and obtained detection limits of Ca, Al, Na, Fe, Mg and Si. Body et al. [14] developed a LIBS instrument for analysis of pressed coal and the measurement accuracy for inorganic components (e.g. Al, Si, and Mg) was typically within  $\pm 10\%$ . Blevins et al. [15] applied LIBS to detect elements such as Na, K, Ca in the flue gas of a power generation boiler. Ctvrtnickova et al. [16, 17] demonstrated the capability of LIBS for characterization of coal fly ash components and the regression coefficients of calibration curves for elements such as Ba, K, Mg were in the range of 0.86–0.99. Many other applications have also been reported [18–24]. However, as far as we know, the extension of the quantitative method to a wide elemental concentration range still appears to be problematic and few publications have focused on the determination of major elements, such as C, in coal—a task with important implications for coal quality.

The application of LIBS to accurately predict the major elemental constituents in coal, though, faces many obstacles that currently restrict its use in on-line coal analysis. A fundamental difficulty is that coal is a relatively heterogeneous sample containing minerals like carbonates, sulfides and oxides, which cause a strong matrix effect including inter-element interference. Due to the many sources of uncertainty, more accurate coal analysis requires further improvement of the calibration model. Multivariate technique partial least squares (PLS) is such a useful method to utilize the abundant spectral information to compensate for different deviations [25–30]. To our knowledge, there is little work on coal analysis using LIBS with PLS. However, the fundamental flaw in PLS application directly to spectra is that the method neglects any underlying physical principles, focusing purely on mathematical data correlation instead. As a result, the predictions from a PLS model have limited accuracy if the matrix of the measured samples varies away from that of the calibration sample set [25]. Another limitation is that PLS may not satisfactorily model the non-linear relationship between the spectrum and the species concentration, such as the saturation effects of self-absorption [27].

The main philosophy of our approach is to induct non-linear factor to PLS thereby to partially overcome the intrinsic linear correlation deficiency of the PLS approach

in modeling non-linear relationship. Specifically, a PLS model with dominant factor based on the physical principles is proposed for C concentration determination in coal samples. The characteristic line intensities are applied to establish the dominant factor reflecting the elemental concentration in typical LIBS measurements, while PLS is used to minimize the residuals of the calibration model to effectively utilize spectral information to compensate for fluctuations of plasma properties. The model combines advantages of the physical-principle-based univariate model and PLS approach, reducing the possibility of overfitting and thus improving the accuracy.

The detailed model description is given in the next section. To prove the efficacy of the proposed PLS model based on dominant factor, an experiment was conducted to test the proposed model's accuracy against the conventional PLS model. Then, the paper discusses several major factors contributing to the inaccuracy of the raw LIBS data obtained from the experiment. The final sections present and discuss the experiment results, highlighting the gains in prediction accuracy from using the newly proposed model.

## Model descriptions

In LIBS measurement, under the conditions of stoichiometric and local thermodynamic equilibrium, the relationship between the elemental concentration and intensity should be linear when plasma properties are constant, as in the traditional univariate model [31]. However, deviations from different sources would deteriorate the theoretical relationship. Due to the intrinsic plasma characteristics and mechanisms, self-absorption is often unavoidable in LIBS measurement if the concentration of the measured species is high. The emission may be absorbed by cool atoms of the same species in or around the plasma, which leads to a pronounced non-linear relationship between the line intensity and elemental concentration. Furthermore, inter-element interference due to line overlap and matrix effects is also very common, so that the intensity of one special line might not only result from one specific element, but also from other element number densities present in the plasma. Besides, many other factors and processes can alter the measured characteristic line intensity, such as the spatial and temporal heterogeneity of the plasma. These distorting processes shift the line intensity simultaneously, indicating that it is difficult to separate them physically and the only practical way is to apply data processing technology to partially compensate for these deviations.

Due to the above deviations, the detected intensity of the characteristic line may not accurately reflect the elemental concentration, but it still contains the basic information for elemental concentration. Considering this, a potential way

is to extract the major concentration information from these characteristic lines and further correct the model by taking the full-range spectrum into account to compensate for the residual errors. A dominant factor model, in which the characteristic lines of the measured or other elements are explicitly extracted to explain the major part of elemental concentration and PLS is further applied to correct the model, is presented in the paper for carbon concentration measurement in coal. The explicitly extracted expression which takes a dominant portion of the final model results is called “the dominant factor”. The procedure of model establishment can be described as follows.

The first step is to extract the main relationship,  $f(I_i)$  between the characteristic intensity  $I_i$  of the measured element  $i$  and elemental concentration  $C_i$ , which could be non-linear under conditions that self-absorption effect cannot be neglected, as in Eq. 1.

$$C_i = f(I_i) \quad (1)$$

Inter-element interference may be a major source for the difference, called as “residual”, between the real element concentration and the value calculated with Eq. 1. However, the mechanism of inter-element interference is complicated and remains unclear. In the present work, the second step is to model the inter-element interference to further minimize the residuals using best curve-fitting technology with non-linear equation. After this process, the expression of the dominant factor is as in Eq. 2,

$$C'_i = f(I_i) + g(I_j) \quad (2)$$

where  $C'_i$  is the calculated elemental concentration of the dominant factor that considers self-absorption and inter-element interference,  $I_j$  is the characteristic line intensity of the influencing element  $j$  and  $g(I_j)$  is the function to describe inter-element interference. After dominant factor extraction, the remaining difference may come from the imperfectness of the dominant factor, other unknown deviation factors and even signal fluctuations, making it difficult to explicitly model these effects. Considering the entire spectrum contains useful information of the deviation sources, it is logical to utilize the full spectrum to further minimize the deviations. Thus, the powerful multivariate PLS method is applied to compensate for the residual error using the entire spectrum information.

In essence, PLS is a technique for modeling a linear relationship between input variables and output variables. The PLS approach first creates uncorrelated latent variables which are linear combinations of the original input variables. A least squares regression is then performed on the subset of extracted latent variables [32]. Because PLS utilizes all spectral information to implicitly and partially

compensate for the signal deviation due to fluctuations of plasma temperature, electron number density, inter-element interference and other factors using linear correlations, it normally yields better calibration results than conventional univariate model.

The distinction in the present approach is that the general PLS approach uses linear relation to directly model non-linear spectra-to-concentration relationships, whereas the newly proposed model only uses linear PLS to fit the much smaller non-linear spectra-to-residual relationship, theoretically leading to better model results. In addition, the model performance is further improved by explicitly extracting part of the non-linear relation between the spectra and the concentration in the dominant factor since PLS correction only needs to handle less non-linear dependence. Please refer to Wang et al. for details on the approach [33]. Equation 3 shows the final expression of the model.

$$C''_i = f(I_i) + g(I_j) + b_0 + b_1x_1 + \dots + b_nx_n \quad (3)$$

where  $C''_i$  represents the final calculated elemental concentration of PLS model based on the dominant factor,  $x_1, x_2, \dots, x_n$  are the spectral intensities at different wavelengths, and  $b_0, b_1, b_2, \dots, b_n$  are the regression coefficients. Compared with the general PLS model, the PLS model based on dominant factor should be more applicable for a wider matrix range as it is bound to the physical principles and capitalizes on the advantages of the multivariate PLS approach to compensate for residuals.

## Experiment setup

Thirty-three standard powdery bituminous coal samples, which contain C over a wide range of concentration values, were certified by the China Coal Research Institute (CCRI) and used in the experiments. A list of samples is given in Table 1, including the element names and the concentration values (in percentage of weight).

Spectrolaser 4000 from XRF in Australia was used in the experiment. Feng et al. [10] described components of the detection system in details. A Q-switched Nd:TAG laser source with the wavelength of 532 nm was applied and laser repetition rate was set to be 1 pulse/s, while the integration time was fixed at 1 ms. Four Czerny–Turner spectrometers and CCD detectors covered the spectral range from 190 to 940 nm with a nominal resolution of about 0.09 nm. About 3 g of powder of each coal sample were placed into a small aluminum pellet die (30 mm diameter), and then pressed with the pressure of 20 tons. The sample was placed on an auto-controlled  $X$ – $Y$  translation stage and exposed to air. It was found that longer warm-up time greatly reduced signal fluctuations, so the warm-up time was set to be more than

**Table 1** Element concentrations of the coal samples from CCRI

No.	C (%)	H (%)
1	81.45	3.52
2	81.54	3.70
3	78.98	4.95
4	79.80	3.80
5	47.12	2.48
6	67.28	3.68
7	58.12	3.40
8	55.14	2.79
9	54.21	2.53
10	72.94	4.46
11	79.22	3.28
12	75.96	4.52
13	60.03	3.73
14	78.78	5.01
15	74.70	4.47
16	76.69	4.42
17	77.28	4.42
18	78.58	4.59
19	52.61	3.45
20	53.42	3.50
21	59.91	3.90
22	67.77	4.25
23	55.67	3.22
24	70.45	3.40
25	77.03	4.70
26	44.60	2.78
27	50.83	3.00
28	59.18	3.07
29	36.90	2.27
30	45.88	3.20
31	39.58	2.05
32	39.53	1.96
33	36.67	2.68

half an hour before analysis. The analysis laser energy and delay time was optimized to be 120 mJ/pulse and 2  $\mu$ s, respectively. Such parameters could produce spectra with negligible Bremsstrahlung radiation and without saturating line intensity to the spectrometer. A laser pulse of 150 mJ was applied to burn off any contaminants before each analysis pulse. For precision calculation, the spectra of 40 replications at different locations on the sample surface were taken to calculate the relative standard error (RSD), while for accuracy modeling, an averaged spectrum of 40 spectra after normalization was obtained for each sample in order to partially average out the experimental parameter fluctuations and the sample heterogeneity.

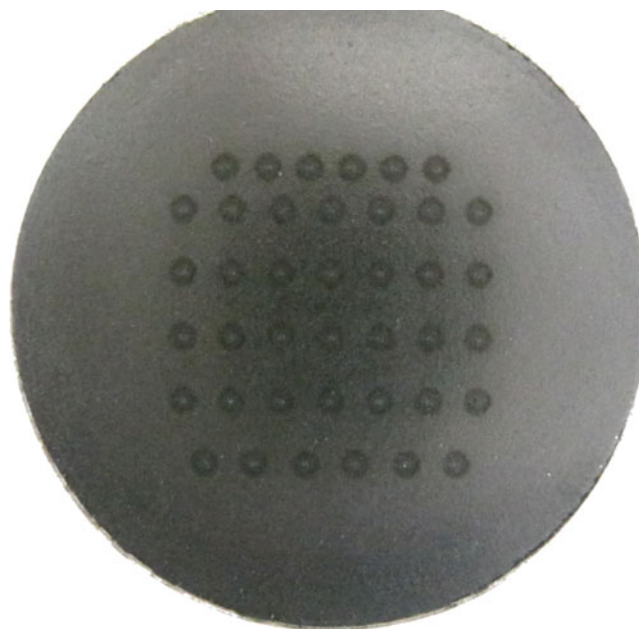
For the purpose of reducing the influence of the background signal, the instrumental and environmental

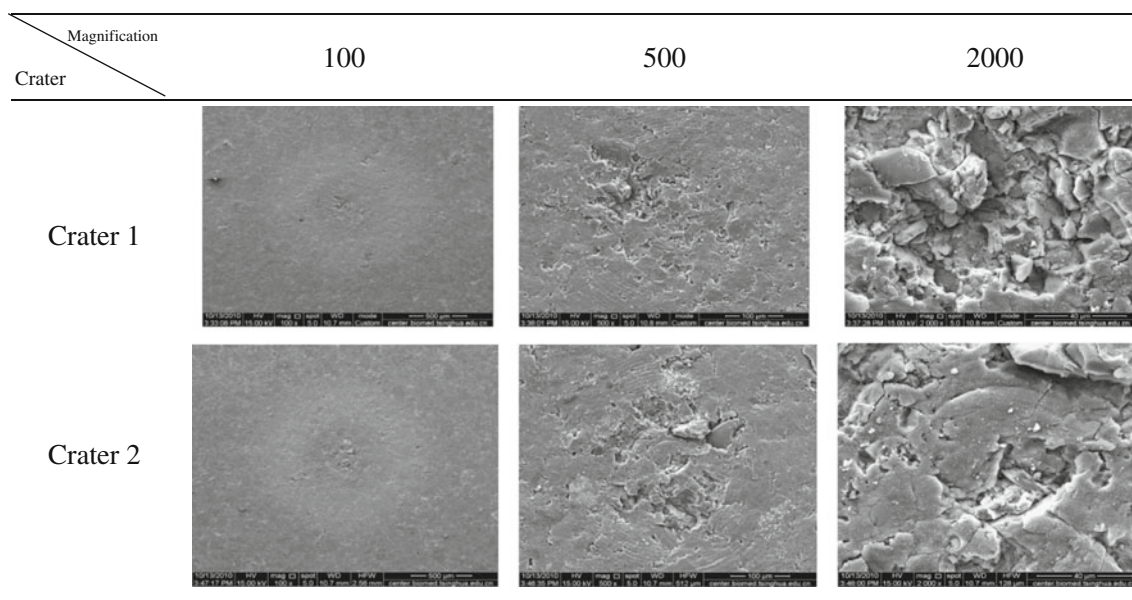
noise was recorded with a long enough delay time and a laser pulse with much lower energy and subtracted from the spectra. Additionally, the spectra were corrected for the efficiency of the detection system to minimize the line intensity distortion from the wavelength dependant efficiency of collecting optics, lenses and fiber optics, the spectrometer gating, the detector sensor and intensifier.

Following the experiment, the data were further adjusted to account for unique phenomena in LIBS coal analysis as discussed below in Special issues in coal analysis with LIBS section. Next, the data were analyzed using both the general PLS method and the new PLS model based on the dominant factor.

### Special issues in coal analysis with LIBS

Figure 1 shows the appearance of the sample surface after analysis. As the image displays, the color of the surface around the laser analysis craters changed, which might result from tar or ash produced by the pyrolysis or combustion of coal. Because these products are chemically distinct from the original coal contents, this might add to the matrix effect in coal analysis with LIBS. Additionally, the laser-sample interaction process will be altered when volatiles in the coal are evaporated after the laser strikes the target. Some of the evaporated volatiles may react with oxygen in the air to produce a flame that can influence the emitted intensity of elemental characteristic lines. This situation should be avoided in future work by optimizing the detection environment, such as using inert gas protection.

**Fig. 1** Appearance of the sample No. 1 surface after analysis (colored)

**Table 2** Photographs of ESEM for two laser analysis craters of sample No. 1

Further analysis revealed another imperfection in coal sampling that likely leads to greater spectral uncertainty. Local variation in the sample's physical properties and mineralogical characteristics, such as surface texture, caused the morphology of craters formed by the pulse to vary from pulse to pulse. To closely observe the laser impact, several photographs using the environmental scanning electron microscope (ESEM) were taken (Table 2, see also [Electronic Supplementary Material](#)). The morphology of the craters shows significant variation, indicating that the ablation mass and the morphology of the plasma fluctuated greatly between each pulse. Such fluctuation in properties can add to the intensity variations among pulses.

The special conditions for coal analysis by LIBS will lead to large discrepancies in the data. Generally, normalization with whole spectral area was applied to minimize the spectrum uncertainty before establishing the calibration model for LIBS measurement. However, the results showed not much improvement for the spectra obtained from coal samples. As seen in Table 3, the RSD of the line intensities after whole spectral area normalization are still very high. The lack of effects is likely a result of the many imperfections discussed above that prevented accurate representation of fluctuations in plasma properties using the whole spectral area.

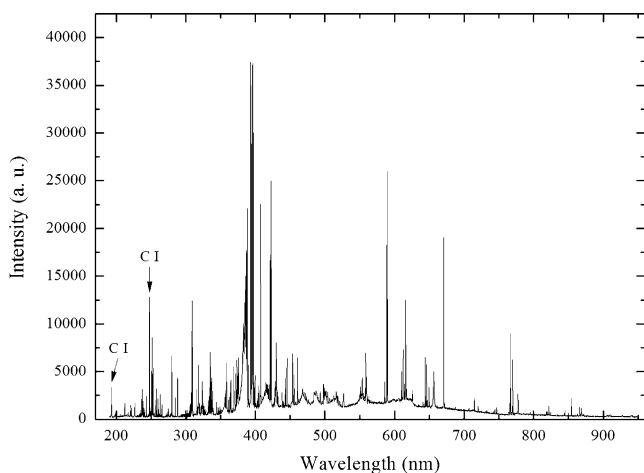
Furthermore, in the typical resulting spectra from coal analysis as shown in Fig. 2, only few atomic emission lines free of overlap for the major element C were observed. This reduces the possibility of selecting an appropriate characteristic line for traditional univariate calibration model. Compared with LIBS analysis for standard metal samples, elemental measurement of coal definitely requires more study for accuracy.

Moreover, the calibration results of the traditional univariate model with the whole spectral area normalization were also not satisfactory. For instance, the data points of intensity  $C(I) 247.856$  nm line versus known C concentration as shown in Fig. 3 are scattered, further indicating the necessity in searching for a suitable data processing method for coal measurement using LIBS.

The whole spectral area represents the total species excitation energy in the laser-induced plasma. Normally, larger whole spectral area usually indicates higher plasma temperature and species concentration inside the plasma caused by higher laser pulse energy and stronger laser-sample interaction processes. Therefore, normalization with the whole spectral area can partly remove the fluctuations of LIBS spectra due to the variations of laser pulse energy, total species, etc. However, as described above, this method does not yield satisfying results for coal analysis, probably

**Table 3** RSD of C(I) line intensity for sample No. 1 after different normalization methods

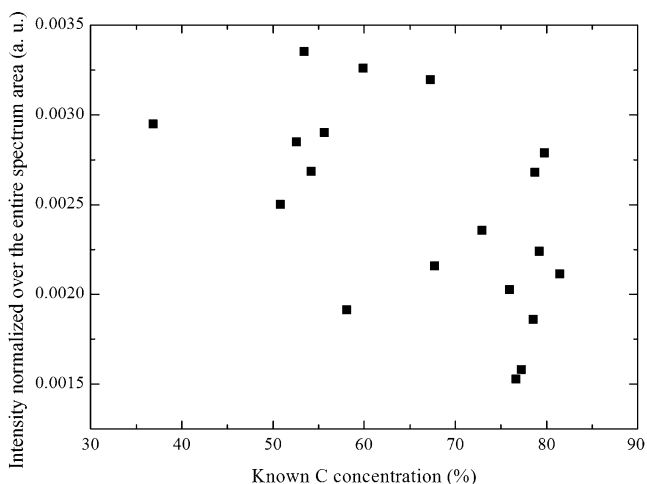
C(I) wavelength (nm)	RSD (whole spectral area normalization, unit: %)	RSD (segmental spectral area normalization, unit: %)
193.092	28.0	12.6
247.856	27.8	11.5



**Fig. 2** A typical LIBS spectrum of coal samples

because the matrix of coal is too complicated. A certain portion of the spectrum is not sensitive to the plasma temperature or total species concentration variation possibly due to the complicated interaction between different elements, weakening the correlation between the whole spectral area and the plasma temperature and total species concentration. That is, the information of spectra fluctuations due to the plasma temperature and total species concentration variation contained in the whole spectral area is partly canceled out due to the non-sensitive segment. Therefore, applying only the area of sensitive segment of the spectra, which should be more correlated with the change of laser power than the whole spectral area, for normalization would possibly yield a good result.

In the present work, as the first choice, the whole spectrum was divided into four segments since naturally our LIBS system used four Czerny–Turner spectrometers to measure the whole spectrum with each spectrometers covering part of the



**Fig. 3** Calibration results of the C(I) 247.856 nm line after normalization with the whole spectral area

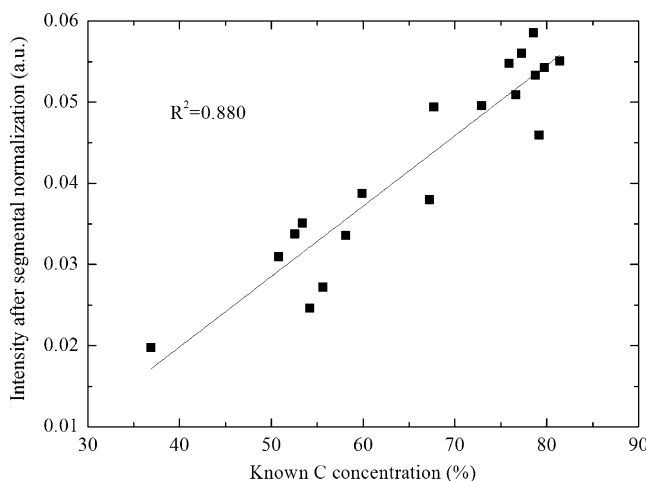
spectral range. The four segments were finally chosen to be 190.048–310.228, 310.228–560.025, 560.025–769.991, and 769.991–940 nm, respectively. The segmental spectral range utilized for the available characteristic lines 193.092 nm and 247.856 nm was set to from 190.048 nm to 310.228 nm. It was found that segmental spectral area normalization was much more effective to reduce the RSD than normalization by the whole spectral area. After segmental spectral area normalization, the RSD values decreased greatly to about 10%, as seen in Table 3.

Furthermore, segmental spectral area normalization also largely increased the calibration accuracy. For instance, as shown in Fig. 4,  $R^2$  value of the calibration curve using C(I) line at 247.856 nm after segmental normalization is 0.880, which shows significant improvement in comparison with the data points from normalization with the whole spectral area in Fig. 3. The above results demonstrate that segmental normalization of the spectra is preferable for samples with complicated constituents, such as coal.

As shown by this experiment, a number of factors contribute to the defects of coal analysis using LIBS. While future opportunities exist to refine the experimental design and procedures of analysis, uncertainties in raw LIBS data will remain, requiring that the calibration and prediction model for the composition of coal samples be as accurate as possible. In advancing towards this aim, the section below demonstrates the results from the new dominant factor model with PLS correction.

## Results and discussions

In the following sections, the proposed dominant factor model was evaluated in terms of C concentration determination. The conventional PLS model with the full spectral



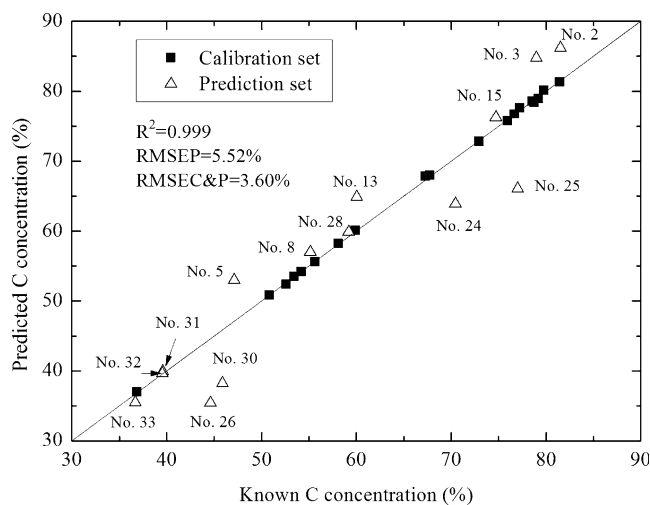
**Fig. 4** Calibration results of the C(I) 247.856 nm line after segmental spectral area normalization

range input was chosen as the baseline against which to clearly judge improvement in the new approach. The software Unscrambler 10.0 (CAMO, Woodbridge, NJ) was used to perform the PLS calculation. The calibration and prediction performance of the model was assessed by the coefficient of determination ( $R^2$ ) and root mean square error of prediction (RMSEP), respectively. For an ideal model,  $R^2$  should be close to 1 while RMSEP is close to 0. Moreover, root mean square error of both calibration and prediction samples (RMSEC&P) was proposed for assessing the overall performance of the model. Smaller RMSEC&P value indicates a better model quality. Nineteen samples were selected to build the calibration model and 14 samples (Nos. 2, 3, 5, 8, 13, 15, 24, 25, 26, 28, 30, 31, 32, and 33) were picked to test the model prediction. Since C concentrations of samples No. 2 and 33 are out of range of the calibration set and the C concentrations of other prediction samples are within the calibration set, the prediction range was chosen purposely to test the robustness of the new PLS model based on dominant factor across a broad range of sample matrixes. In addition, it has been noticed that since the coal matrix is very complicated, C concentration can only partially stand for the matrix of coal. Therefore, the choice of the Nos. 2 and 33 samples as the out-of-calibration samples may not be accurate enough.

### Baseline

PLS is one of the advanced analytical tools in chemometrics and has shown great potential in LIBS quantitative measurements, thus it was chosen as the baseline. In the baseline PLS model, the intensities at different wavelengths of all spectra after segmental spectral area normalization and C concentration of samples were used as input variables and output variables, respectively. The intensity at each wavelength was normalized to the area of the segment which included the intensity wavelength. The number of principal components was chosen to have the smallest RMSEP for the prediction sample set.

As shown in Fig. 5, for the baseline model, the RMSEC&P is 3.60% and  $R^2$  value is 0.999, but the RMSEP is as high as 5.52%. The probable reasons for the large prediction errors are that conventional PLS does not consider the physical principles and the model is not robust over a wide range of C concentrations due to the intrinsic non-linear correlation and overfitting of PLS. That is, conventional PLS application only maintains the accuracy for the range of C concentrations contained in the matrixes of the measured samples within the calibration set. Therefore, when the matrix of the measured sample is out of the calibration sample set, the prediction may not be satisfactory. The absolute relative errors for



**Fig. 5** Baseline model results

sample No. 2, which has the highest C concentration, and sample No. 33, which has the lowest C concentration, are as high as 5.69% and 3.29%, respectively, because they lie outside the range of C concentration contained in the samples used for calibration. Moreover, the widely present noise in the LIBS spectra might make the PLS model less effective because of the excess of unrelated noise input.

From the above results, it was found that for coal samples with complicated constituents, the matrix effect will greatly influence the PLS model accuracy. According to the model description, the major concentration information contained in the characteristic line intensity of the specific element should play a more prominent role and have more weight in calculating model results. Establishing a model with a dominant factor based on the physical principles is a potential way to avoid disadvantages of conventional PLS, which is one of the basic philosophies for the proposed model.

### Dominant factor extraction

The proposed PLS model is mainly determined by the dominant factor based on physical principles between the characteristic line intensity and the specific elemental concentration. The simplest dominant factor was first extracted only applying linear relation (as in the conventional univariate model) as follows.

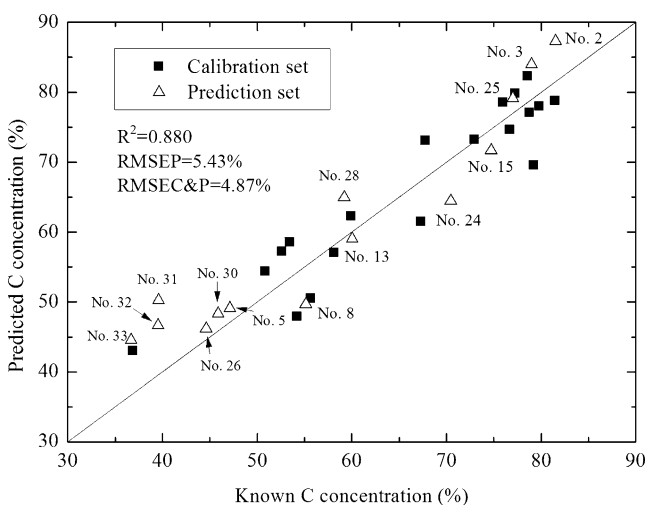
$$C_c = f(I_c) = kI_c + b \quad (4)$$

where  $C_c$  is the C concentration,  $k$  and  $b$  are coefficients and  $I_c$  is the spectra area of C(I) line at 247.856 nm. This line was selected because it provided the best linear calibration results among all applicable C(I) lines after segmental spectral area normalization. In addition, the non-

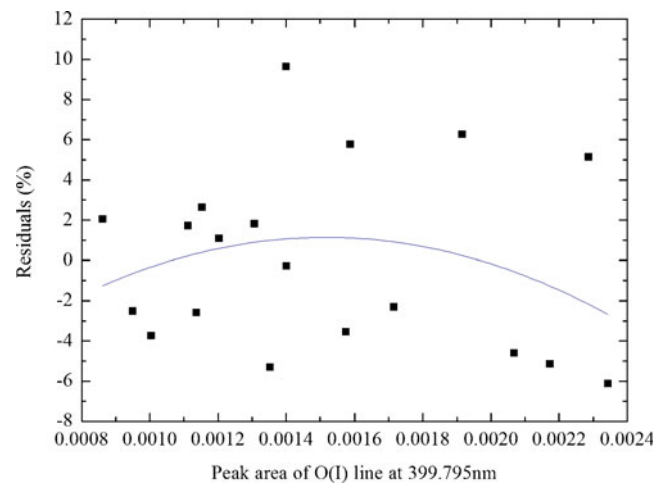
linear empirical expression for self-absorption in the literature [34] was also applied to describe the characteristic line intensity and concentration relation, but it did not improve the dominant factor accuracy compared with the linear relation. The reason might be that the C(I) line at 247.856 nm is not a resonance line of C and thus self-absorption was relatively weak. Therefore, linear relation was preferred in the present work. Moreover, it should be noted that if the self-absorption cannot be neglected, the dominant factor is flexible enough to adopt a suitable expression to describe the non-linear effect.

Figure 6 shows the calibration and prediction results of the linear dominant factor. The RMSEC&P is 4.87% and  $R^2$  is 0.880, which proves that a conventional linear dominant factor less accurately calibrates the true concentration, compared to the baseline model. In contrast, the RMSEP is 5.43%, which is better than the baseline model, showing that physical principles help to make the model more robust than the generally applied PLS method.

Another unavoidable source of uncertainty is inter-element interference, especially for sample like coal with complicated matrix. The inter-element interference effects to the C line intensity (247.856 nm) were modeled by correlating the residual errors of the linear dominant factor with various characteristic line intensities of other elements. The peak area of O(I) line at 399.795 nm was found to have the best correlation, which means that O might be the main species in influencing the C(I) line intensity at 247.856 nm. Therefore, the peak area of O(I) line was taken to further improve the dominant factor. As shown in Fig. 7, there is no clear relation between the residuals of the linear dominant factor and O(I) intensity. The diluted relationship is from the fact that other elements also affected the C(I) intensity and the O(I) intensity was in turn influenced by C and other elements. The model is therefore only intending



**Fig. 6** Calibration and prediction results of the linear dominant factor



**Fig. 7** Correlation between the residuals and O(I) intensity at 399.795 nm

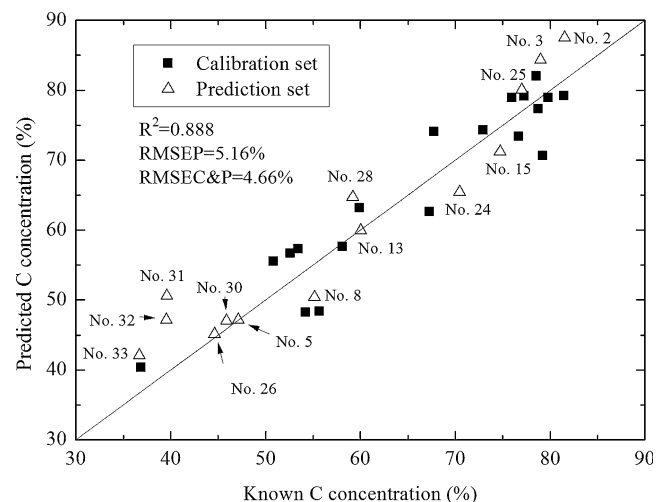
to partially compensate for the effect using curve-fitting technology instead of developing a thorough expression to explicitly describe the interference. A quadratic polynomial equation for  $g(I)$  was obtained using best curve-fitting technology as follows to model the interference effect:

$$e_c = a_0 + a_1 I_o + a_2 I_o^2 \quad (5)$$

where  $e_c$  is the residual error of the linear dominant factor,  $I_o$  is the peak area of O(I) line at 399.795 nm and  $a_0, a_1, a_2$  are constants obtained from the best curve-fitting technology. Now, the dominant factor can be written as:

$$C'_c = kI_c + b + a_0 + a_1 I_o + a_2 I_o^2 \quad (6)$$

The calibration and prediction results of the final dominant factor are shown in Fig. 8. All results are better than the linear dominant factor, which indicates that the



**Fig. 8** Calibration and prediction results of the dominant factor with inter-element interference consideration



**Table 4** List of dominant factor model results

Dominant factor	$R^2$	RMSEP (%)	RMSEC&P (%)
$C'_c = kI_c + b + a_0 + a_1I_o + a_2I_o^2$ <sup>a</sup>	0.888	5.16	4.66

<sup>a</sup> $k=1.014 \times 10^3, b=2.301 \times 10, a_0=-11.695, a_1=1.693 \times 10^4, a_2=-5.583 \times 10^6$

more the physical processes are considered, the more robust the dominant factor model is. For instance, the absolute relative error for sample No. 33 was reduced to 14.8% from 21.5% as in the linear dominant factor alone.

As listed in Table 4, though RMSEC&P and  $R^2$  of the dominant factor are worse than that in the baseline model, RMSEP of the dominant factor is better than that in the baseline PLS model, further confirming our conclusion that the physical principle makes the model robust for the situations where the prediction matrix lies outside the calibration set. Still, the  $R^2$  of the dominant factor is not close to 1 because many other deviations are not yet compensated for. PLS is thereafter applied to implicitly minimize the residual errors of the dominant factor using full spectral information.

PLS approach based on the dominant factor

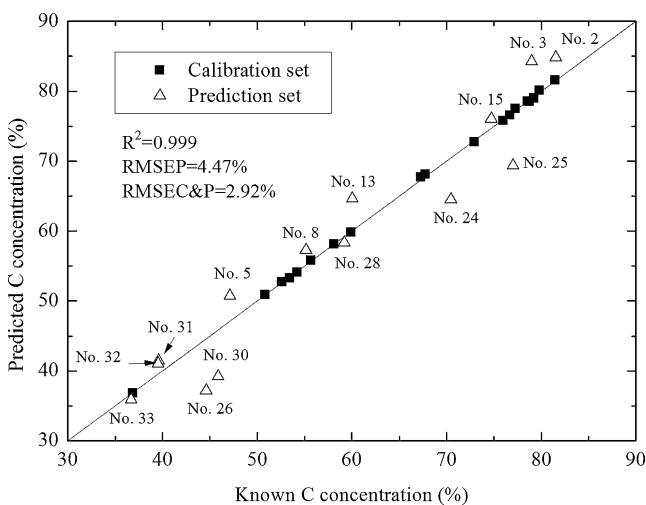
The residual errors of the dominant factor were further corrected with PLS. The number of principle components was also chosen to obtain the smallest RMSEP. After the whole process,  $R^2$  is 0.999 and RMSEP is reduced to 4.47% (Fig. 9). In comparison with the result of the dominant factor alone, the PLS residual correction greatly improves the quality of the calibration curve and the RMSEP. For the overall quality, RMSEC&P is 2.92%, which is also significantly better than the dominant factor

itself and the conventional PLS method, showing the combination of the dominant factor with PLS correction improves the final results.

Table 5 summarizes results of all models presented in the paper. Generally, the dominant-factor-based PLS model has the best calibration and prediction performance. RMSEC&P in the new model is better than the baseline model, being reduced to 2.92% from 3.60% in the baseline model.  $R^2$  of calibration is the same while RMSEP declines significantly. For example, the absolute relative error for sample No. 33 (lowest C concentration) decreased to 2.17% from 3.29% in the baseline model. For sample No. 2 (highest C concentration), the absolute relative error reduced from 5.69% in the baseline model to 3.99% in the dominant factor with PLS correction model. The lower RMSEP indicates that the application of the physical background makes the new model more robust for complicated bituminous samples over a wider C concentration range. Therefore, by combining the advantages of the univariate model and PLS method, the dominant-factor-based PLS model is not only comparable with PLS method for samples having compositions within the range of the calibration set, but is also more robust in measuring unknown samples out of the calibration range compared with the conventional PLS approach.

Conclusions

In the present work, experiments of applying LIBS to measurement of 33 bituminous coal samples with a wide carbon concentration range were carried out. Several special issues for coal analysis with LIBS were discussed. The large RSD of the spectra should result from the heterogeneity of the samples and pyrolysis or combustion of coal during the laser-sample interaction processes. Normalization with segmental spectral area was found to



**Fig. 9** Calibration and prediction results of PLS based on dominant factor

**Table 5** Results of different models for C concentration determination

Models	$R^2$	RMSEP (%)	RMSEC&P (%)
Baseline model	0.999	5.52	3.60
Dominant factor	0.888	5.16	4.66
PLS model based on dominant factor	0.999	4.47	2.92

be more effective to reduce measurement uncertainty and improve the calibration quality than generally applied normalization with the whole spectral area.

Dominant-factor-based PLS method was applied for the C concentration measurement of bituminous coal. Results show great improvement over conventional PLS approach mainly because the new approach induces non-linear correlation to PLS model and partly reduces the noised overfitting by the dominant factor model.

Considering the physics of LIBS, although the intensity of the characteristic lines may not be accurate enough to reflect the measured element concentration, it still contains the most-related information for concentration measurement. Thus, the characteristic line intensity was firstly taken to extract the dominant C concentration. Then, the correlation between the residuals and other elemental emission lines were used to model the non-linear inter-element interference effect to further improve the dominant factor. The residuals of the dominant factor were corrected by PLS using the abundant whole spectral information to compensate for the imperfectness of the dominant factor and other left deviation factors that were not modeled. Compared with the baseline PLS model, RMSEP decreased from 5.52% to 4.47% while  $R^2$  remained as high as 0.999, and RMSEC&P was also reduced from 3.60% to 2.92%, showing that the proposed model is more robust in a wider C concentration range because of its physical principle basis.

**Acknowledgement** The authors acknowledge the financial support from the governmental “863” project (NO. 20091860346) and “973” project (NO. 2010CB227006).

## References

- Bustamante MF, Rinaldi CA, Ferrero JC (2002) Laser induced breakdown spectroscopy characterization of Ca in a soil depth profile. *Spectrochim Acta B* 57:303–309
- Maravelaki-Kalaitzaki P, Angelos D, Kilikoglou V, Zafropoulos V (2001) Compositional characterization of encrustation on marble with laser induced breakdown spectroscopy. *Spectrochim Acta B* 56:887–903
- Rials TG, Kelley SS, So CL (2002) Use of advanced spectroscopic techniques for predicting the mechanical properties of wood composites. *Wood Fiber Sci* 34:398–407
- Pandhija S, Rai AK (2008) Laser-induced breakdown spectroscopy: a versatile tool for monitoring traces in materials. *Pramana J Phys* 70:553–563
- Gomba JM, D’Angelo C, Bertuccelli D, Bertuccelli G (2001) Spectroscopic characterization of laser-induced breakdown in aluminum-lithium alloy samples for quantitative determination of traces. *Spectrochim Acta B* 56:695–705
- Lee WB, Wu JY, Lee YI, Sneddon J (2004) Recent applications of laser-induced breakdown spectrometry: a review of material approaches. *Appl Spectrosc Rev* 39:27–97
- Li J, Lu J, Lin Z, Gong S, Xie C, Chang L, Yang L, Li P (2009) Effects of experimental parameters on elemental analysis of coal by laser-induced breakdown spectroscopy. *Opt Laser Technol* 41:907–913
- Taschuk MT, Tsui YY, Fedosejevs R (2006) Detection and mapping of latent fingerprints by laser-induced breakdown spectroscopy. *Appl Spectrosc* 60:1322–1327
- Beldjilali S, Borivent D, Mercadier L, Mothe E, Clair G, Hermann J (2010) Evaluation of minor element concentrations in potatoes using laser-induced breakdown spectroscopy. *Spectrochim Acta B* 65:727–733
- Feng J, Wang Z, Li Z, Ni W (2010) Study to reduce laser-induced breakdown spectroscopy measurement uncertainty using plasma characteristic parameters. *Spectrochim Acta B* 65:549–556
- Zhang L, Dong L, Dou H, Yin W, Jia S (2008) Laser-induced breakdown spectroscopy for determination of the organic oxygen content in anthracite coal under atmospheric conditions. *Appl Spectrosc* 62:458–463
- Yin W, Zhang L, Dong L, Ma W, Jia S (2009) Design of a laser-induced breakdown spectroscopy system for on-line quality analysis of pulverized coal in power plants. *Appl Spectrosc* 63:865–872
- Wallis FJ, Chadwick BL, Morrison RJS (2000) Analysis of lignite using laser-induced breakdown spectroscopy. *Appl Spectrosc* 54:1231–1235
- Body D, Chadwick BL (2001) Optimization of the spectral data processing in a LIBS simultaneous elemental analysis system. *Spectrochim Acta B* 56:725–736
- Blevins LG, Shaddix CR, Sickafoose SM, Walsh PM (2003) Laser-induced breakdown spectroscopy at high temperatures in industrial boilers and furnaces. *Appl Opt* 42:6107–6118
- Ctvrtnickova T, Mateo MP, Yanez A, Nicolas G (2010) Laser induced breakdown spectroscopy application for ash characterization for a coal fired power plant. *Spectrochim Acta B* 65:734–737
- Ctvrtnickova T, Mateo MP, Yanez A, Nicolas G (2009) Characterization of coal fly ash components by laser-induced breakdown spectroscopy. *Spectrochim Acta B* 64:1093–1097
- Gaft M, Sapir-Sofer I, Modiano H, Stana R (2007) Laser induced breakdown spectroscopy for bulk minerals online analyses. *Spectrochim Acta B* 62:1496–1503
- Gaft M, Dvir E, Modiano H, Schone U (2008) Laser induced breakdown spectroscopy machine for online ash analyses in coal. *Spectrochim Acta B* 63:1177–1182
- Mateo MP, Nicolas G, Yanez A (2007) Characterization of inorganic species in coal by laser-induced breakdown spectroscopy using UV and IR radiations. *Appl Surf Sci* 254:868–872
- Romero CE, De Saro R, Craparo J, Weisberg A, Moreno R, Yao Z (2010) Laser-induced breakdown spectroscopy for coal characterization and assessing slagging propensity. *Energy Fuels* 24:510–517
- Noda M, Deguchi Y, Iwasaki S, Yoshikawa N (2002) Detection of carbon content in a high-temperature and high-pressure environment using laser-induced breakdown spectroscopy. *Spectrochim Acta B* 57:701–709
- Burakov VS, Tarasenko NV, Nedelko MI, Kononov VA, Vasilev NN, Isakov SN (2009) Analysis of lead and sulfur in environmental samples by double pulse laser induced breakdown spectroscopy. *Spectrochim Acta B* 64:141–146
- Yu L, Lu J, Xie C, Chen W, Wu G, Shen K, Feng W (2005) Analysis of pulverized coal by laser-induced breakdown spectroscopy. *Plasma Sci Technol* 7:3041–3044
- Clegg SM, Sklute E, Dyar MD, Barefield JE, Wiens RC (2009) Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques. *Spectrochim Acta B* 64:79–88
- Fink H, Panne U, Niessner R (2002) Process analysis of recycled thermoplasts from consumer electronics by laser-induced plasma spectroscopy. *Anal Chem* 74:4334–4342
- Sirven JB, Bousquet B, Canioni L, Sarger L (2006) Laser-induced breakdown spectroscopy of composite samples: comparison of advanced chemometrics methods. *Anal Chem* 78:1462–1469

28. Amador-Hernandez J, Garcia-Ayuso LE, Fernandez-Romero JM, Luque De Castro MD (2000) Partial least squares regression for problem solving in precious metal analysis by laser induced breakdown spectrometry. *J Anal At Spectrom* 15:587–593
29. Freedman A, Iannarilli FJ, Wormhoudt JC (2005) Aluminum alloy analysis using microchip-laser induced breakdown spectroscopy. *Spectrochim Acta B* 60:1076–1082
30. Braga JWB, Trevizan LC, Nunes LC, Rufini IA, Santos D, Krug FJ (2010) Comparison of univariate and multivariate calibration for the determination of micronutrients in pellets of plant materials by laser induced breakdown spectrometry. *Spectrochim Acta B* 65:66–74
31. Miziolek AW, Palleschi V, Schechter I (2006) Laser induced breakdown spectroscopy. Cambridge University Press, Cambridge, UK
32. Wold S, Sjostrom M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* 58:109–130
33. Wang Z, Feng J, Li L, Ni W, Li Z (2010) A novel multivariate model based on dominant factor for laser-induced breakdown spectroscopy measurements, arXiv:1012.2735v1 [physics.optics]
34. Aragon C, Aguilera JA, Penalba F (1999) Improvements in quantitative analysis of steel composition by laser-induced breakdown spectroscopy at atmospheric pressure using an infrared Nd: YAG laser. *Appl Spectrosc* 53:1259–1267