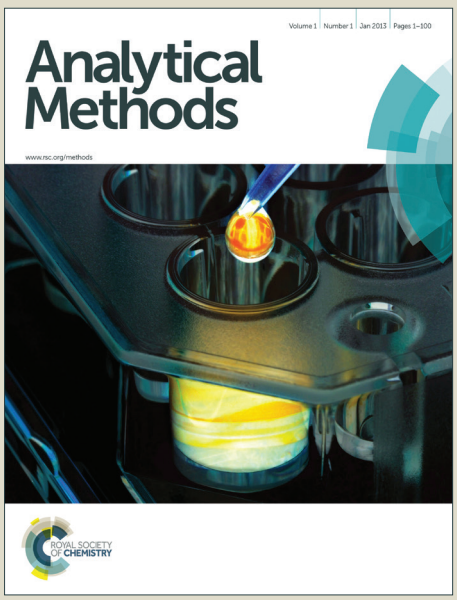
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Quantitative analysis of nonmetal elements in steel using laser-induced breakdown spectroscopy combined with random forest

Shan Wu[a], Tianlong Zhang[a], Hongsheng Tang[a], Kang Wang[b], Xiaofeng Yang[c], Hua Li*[a]

a. Institute of Analytical Science, College of Chemistry & Materials Science, Northwest University, Xi'an, 710069, P. R. China

b. College of Science, Chang'an University, Xi'an, 710064, P. R. China

c. College of Chemistry & Materials Science, Northwest University, Xi'an, 710069, P. R. China

Author to whom correspondence should be addressed:

Dr. Hua Li

Institute of Analytical Science, College of Chemistry & Materials Science, Northwest University

Tel: 86-29-88302635

Fax: 86-29-88303527

E-mail: huali@nwu.edu.cn

Abstract: Laser induced breakdown spectroscopy (LIBS) combined with random forest (RF) was proposed to quantitative analysis sulfur (S) and phosphorus (P) in steel samples. The interference from characteristic spectral line of S and P in steel is difficult to accurately quantitative analysis due to the influence of multi-matrix. RF model was utilized to compensate the negative influence of matrix effect. The influences of laser pulse energy and delay time on the spectral intensity were studied to improve signal-to-noise ratio (SNR) of analytical line for a certain element. Furthermore, the parameters (**ntree** and $\mathbf{m_{try}}$) of RF model were optimized by out-of-bag (OOB) estimation. The final RF calibration model for quantitative analysis of S and P in steel was constructed by means of the spectral range (520-620 nm) as input variable under the optimized experimental conditions. Results showed well prediction of RF calibration model for S (R^2=0.9974) and P (R^2=0.9981) compared with partial least square regression (PLSR) by using the peak signal of S Ⅱ 545.3 nm and P Ⅱ 602.4 nm, respectively, and the averaged relative error (ARE) of S in steel were 2.69% and 3.47% for 8# and 9# sample respectively, and of P were 1.77% and 0.83% for 8# and 9# sample respectively. This confirms that RF model is a promising approach for quantitative detection of the nonmetal elements with LIBS technology in the field of metallurgy.

Keywords: Laser induced breakdown spectroscopy; Random forests; Steel samples; Quantitative analysis; Sulfur and phosphorus

1. Introduction

Various components in steel play an important role on estimating the quality and performance of steel. In the process of metallurgy, a real-time, on-line and effective approach for monitoring the composition of different types of steel is needed to ensure the quality and performance of steel. Strict and accurate control content of nonmetal elements, especially the nonmetal elements of S and P, play an extremely important role on determining some steels of the mechanical and physical properties.[1] The machinability, processability and magnetism of steel can be improved through adding appropriate S. However, the content of S in steel shall not be higher than 0.005% (wt)[2] based on the national standards of China. If not, its plasticity and wear resistance would be significantly decreased and even affect the working life of the steel. Similarly, the content of P should be lower than 0.035% (wt),[3] as it tends to increase the strength and hardness, and reduce

the plasticity and toughness of steel by adding a small amount P. Therefore, a rapid, effective and direct determination method for S and P in steel is needed due to its great significance for steel quality control. The conventional methods for quantitative analysis of S and P in steel[4-8] mainly include chemical analysis, inductively coupled plasma mass spectrometry (ICP-MS), x-ray fluorescence (XRF), atomic emission spectrometry (AES), spark-discharge optical emission spectroscopy (SD-OES) and so on. However, these analytical techniques show some shortcomings such as complex sample preparation or long analysis time, which hinders their application for in-situ, on-line and real-time analysis.

Laser-induced breakdown spectroscopy (LIBS) is an emerging atomic emission-based technique for material analysis. LIBS has rapidly developed into an important analytical technique due to the capability of detecting several elements in the sample simultaneously.[9] In LIBS, the light signal emitted by plasma that vaporizes the samples by low-energy pulsed laser and focusing lens, and then collected and directed to a spectrometer by a fiber optic. The spectrometer disperses the light emitted by excited atoms, ions, and simple molecules in the plasma. The detector records the emission signals to quantitative determine the content of the sample. Compared with the conventional quantitative analysis technologies, the obvious advantages of LIBS generally include the facts that[10-12]: minimal sample requirement, without sample preparation, non-destructive detection, all types of samples (liquid, solid, gas, etc.), and even remote and in-situ analysis.[13,14] LIBS has been widely used in materials,[15] metallurgy,[16] environment,[17] archaeology,[18] space exploration,[19] medical,[20] military,[21-23] and many other fields, particularly in the metallurgy field.

Quantitative analysis of steel samples with LIBS usually can be accomplished by using a series of standard samples to establish calibration curve between spectral line intensity and corresponding content (wt%). At present, the quantitative methods for S and P in steel mainly include standard curve method and internal standard method, and these methods were univariate. Univariate analysis is the most commonly used calibration curve, namely only using single characteristic spectral line of the measured element as input variable to set up univariate regression model. Unfortunately, the univariate analysis fails to meet the analysis requirements of multi-matrix steel samples[25] which is a sample with multi-component and multi-matrix, and its chemical composition is severely affected by the matrix effect, and there are also complex spectral overlap in iron matrix spectral lines. Therefore, the traditional univariate analysis fails to

compensate the influence of these interference factors. Hence, multivariable analysis methods such as partial least square regression (PLSR),[24] random forests (RF),[26,27] artificial neural network (ANN)[28-31] and support vector machine (SVM)[32-34] are adopted to solve matrix effect and improve the accuracy of quantitative analysis for steel sample. It has been proved that the running speed of RF is very fast and the precision of RF algorithm is incomparable, and even can prevent over-fitting, as well as has a good tolerance for the noise.[35] RF, a new regression algorithm based upon multiple regression trees, was proposed by Leo Breiman in 2001.[36] It is based on an ensemble of regression trees, from which the prediction of a continuous variable is regarded as average of the prediction of all trees. In modeling process, first of all, the bootstrap method[37,38] is used to randomly extract the training set and test set, and then random regression tree forest tree (**ntree**) and the characteristics of random variables ($\mathbf{m_{try}}$) are optimized through out-of-bag (OOB) estimation in order to improve the robustness and accuracy of RF model. Based on the above mentioned advantages, RF has become a new research hotspot in the current field of machine learning.[39,40]

Here we report a method to determine the content of S and P in steel based on the LIBS and RF technology. Normalized LIBS spectra of steel were used as analytical spectrum. The effect of laser pulse energy and the delay time of detector on signal intensity were studied to achieve a high signal-to-noise ratio (SNR). And then the two parameters (**ntree** and $\mathbf{m_{try}}$) of RF model were optimized by OOB error estimation. The performance of calibration model was investigated by different input variables and spectra feature bands under the condition of analysis lines of S and P, respectively. Finally, RF and PLSR were employed to determine the content of S and P in nine kinds of steel samples under the optimized experimental conditions.

2. Methods and materials

2.1 LIBS setup

The schematic diagram of the experimental setup used for LIBS system is described in Fig. 1, it mainly consists of the duration Nd:YAG pulse laser, optical system, Echelle spectrometer and computer system. The laser light is focused by fused-silica planoconvex lens perpendicularly onto the sample surface to form the plasma, which was generated by a Q-switched Nd:YAG laser (10 ns full width at half maximum (FWHM) and 20 Hz repetition rate) with the optional wavelength at 1064 nm. The distance between focusing lens (focal distance: 50 mm, and diameter: 2 mm) and

sample is greater than the focal length of the lens, in order to avoid that the air is staved before laser reach the surface of the sample. For collecting LIBS spectra from different locations of the sample surface, the sample was mounted on X-Y-Z manual micrometric stage. The radiant energy of the pulse must be higher than the breakdown energy for the sample material. Firstly, the atomic and molecular structure of the steel samples will be broken and heated, causing vaporization of a small fraction of the material. Further the incoming energy of the same laser pulse can sustain high temperature plasma in which the vaporized species can be excited and return later to their less energetic levels by emitting electromagnetic radiation. The plasma emission were collected through the optical fiber (with a 1000 nm core diameter and 0.22 numerical aper-ture), and then detected by the Echelle spectrometer (ARYELLE-UV-VIS, LTB150, German) equipped with an Electron-Multiplying CCD camera (UV enhanced, 1004×1002 Pixels, USA). The spectral resolution range of Echelle spectrometer is 200-800 nm, its optical resolution is about 0.1 nm (FWHM). At the same time, the spectrometer provides a constant spectral resolution (CSR) of 6000 over a wavelength range of 220-800 nm displayed in a single spectrum. In this experiment, there are three parameters optimized as follows: the energy of the laser pulse was 60 mJ, pulse frequency of 20 Hz, detector delay time of 1.5 μs, in order to eliminate the influence of background noise produced in the initial formation stage of the plasma bremsstrahlung for spectral signals. To improve the spectral intensity, each recording was obtained by accumulating the signal of 20 ablation events on the same site. The build-in analysis system of Sophi 6.1.5 R2110 was used to acquire the spectral data.

Here is Fig. 1

Fig. 1 Schematic diagram of the experimental set-up for LIBS studies

2.2 Steel samples and LIBS measurements

A total of 9 typical steel samples were kindly provided by the China Xi-ning Special Steel CO., LTD (Xi-ning, Qing-hai, China) for LIBS analysis. The steel samples are all made into Φ20×6 mm cylinder in order to facilitate the experiment conveniently. Before the experiment operation, the surface of samples was swashed by alcohol. In this study, 300 LIBS spectra were collected from different position on each steel sample, and the analytical spectra for each steel sample were the average of fifty spectra. Subsequently, 54 individual records (9 samples × 6 spectra per sample = 54 records) with nine different contents of S and P were input into the RF model. In this case, the

PLSR and RF model were internally calibrated using 42 records (1# - 7# samples × 6 spectra per sample) and tested using 12 records (8# - 9# samples × 6 spectra per sample). The data processing and quantitative analysis for steel samples by chemometrics methods were operated on Matlab (2007a).

2.3 Random Forest

RF is a kind of statistical learning theory that extracts multiple samples from original sample by taking advantages of the bootstrap resampling method for the regression analysis. The RF model was built by integrating the relationship between spectra and content (S and P in steel) of each regression tree of each bootstrap samples, finally the predictions are made by averaging all the regression tree outputs.

In general, the RF algorithm for regression works as follows:

(1) From the training data of **n** steel sequences, draw the ntree bootstrap sample (i.e., randomly sample, with replacement, **n** steel sequences).

(2) For each bootstrap sample, grow a tree with the following modification: at each node, choose the best split among a randomly selected subset of $\mathbf{m_{try}}$ (rather than all) features. The tree is grown to the maximum size (i.e.,until no further splits are possible) and not pruned back.

(3) Predict new data by aggregating the predictions of the **ntree** trees (average for regression).

The prediction performance of the RF regression algorithm is assessed by a type of cross-validation in parallel with the training step by using the so-called OOB samples. In each bootstrap sample, that is a random sample of the original data with replacement and with the same length, some of the data is repeated, and the left out samples are called OOB. The random forest algorithm evaluates the importance of a variable by looking at how much prediction error increases when (OOB) data for that variable is permuted while all others are left unchanged. The number of trees (**ntree**) and the size of the variable subset ($\mathbf{m_{try}}$) in the RF modeling are the two important keys that is reasonably essential to optimize for quantitative analysis of S and P in steel. The best **ntree** and $\mathbf{m_{try}}$ value are determined according to the root mean square error (RMSE) of OOB estimation. The quantitative analysis performance of RF model is estimated by root mean square error of calibration (RMSEC) and the correlation coefficient (R) in this study. The relevant equations are as follows:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n}(x_{i,reference} - y_{i,predictive})^2}{n}}$$

$$R = \frac{\sum_{i=1}^{n}(x_{i,reference} - \bar{x}_{i,reference})(y_{i,predictive} - \bar{y}_{i,predictive})}{\sqrt{\sum_{i=1}^{n}(x_{i,reference} - \bar{x}_{i,reference})^2 \sum_{i=1}^{n}(y_{i,predictive} - \bar{y}_{i,predictive})^2}}$$

where *i*, *n*, represent the index of the sample and the tested number, respectively; $x_{i,reference}$ and $y_{i,predictive}$ are the reference value and predictive value for i-th sample, respectively; $\bar{x}_{i,reference}$ and $\bar{y}_{i,predictive}$ are the average value of reference and predictive value for all samples, respectively.

3. Result and discussion

3.1 Optimize the parameters

3.1.1 The influence of laser pulse energy to the SNR of spectral line for S and P

The laser induced plasma is generated through laser which provides the direct motivation in the process of forming plasma in LIBS system. Laser induced plasma is generated on the condition that the laser pulse energy is more than a certain threshold (i.e. several GW/cm^2). The laser energy directly affects the number of excited atom, electron and molecule, and even the strength of the plasma emission signal. When the pulse energy is too low, the background signal of the spectral line is strong and the intensity of characteristic spectral line is weak, which is unfavorable for quantitative analysis of S and P in steel; In initial stage, spectral line intensity has a linear relation with the increase of laser energy. And then with the increasing of laser energy, it appears nonlinear relation between spectral line intensity and laser energy. Finally, it will lead to the saturation phenomenon (that is to say energy-spilling)[41] if the laser pulse energy exceeds the threshold, which to the disadvantages of accurately quantitative analysis of S and P in steel. In this study, the delay time was set to 1.5 μs, and the laser pulse energy was taken 10 values between 3-100 mJ in which each energy tested on different locations of sample surface. Fig. 2 shows that the changes in the relationship between the laser pulse energy and SNR with analytical lines of S Ⅱ 545.3 nm and P Ⅱ 602.4 nm, respectively. It can be seen the black curve (element S) form Fig. 2, the threshold energy of plasma generated from pulse laser is between 50 - 60 mJ. When the laser pulse energy reaches 65 mJ, it appears the saturation phenomenon. Meanwhile, the element P has the similar change in the relationship between the laser pulse energy and signal-to-noise ratio with the element S. Therefore, in order to get the spectral lines of SNR as high as possible, the

optimized laser pulse energy for the quantitative analysis of S and P in steel was set to 60 mJ.

Here is Fig. 2

Fig. 2 The relationship between the laser pulse energy and SNR with S Ⅱ 545.3 nm and P Ⅱ 602.4 nm, respectively

3.1.2 The influence of delay time to the SNR of spectral line for S and P

Plasma radiation is a dynamic evolution process with the time, the plasma radiation signals at different instants of time, the observing time and the width are all having serious impact on the results of quantitative analysis. At initial phase of the laser plasma formation, the strong continuous background spectra can be produced by violent collisions among a large number of electrons, excited ions and atoms; Then atomic spectra and ionic spectra are superposed on the continuous background spectra; As time goes on, the plasma are inflated and cooled down, accompanied with the background radiation decreased rapidly . However, the attenuation of atomic spectra and ionic spectra is slower than that of the continuous radiation background. Hence, it is crucial for getting a high SNR of the characteristic spectral line through selecting an appropriate delay time and integral gate width back. Fig. 3 presents the influence of different delay time on the spectral signal intensity with analytical line of S Ⅱ at 545.3 nm and P Ⅱ at 602.4 nm, respectively. As shown in Fig. 3, the SNR of S increases continuously along with the delay time between 0-1.5 μs. However, while when the delay time is above 1.5 μs, the SNR decreased accordingly. The P has the same changing trend. Therefore, the delay time was selected at 1.5 μs to obtain the highest SNR.

Here is Fig. 3

Fig. 3 The relationship between the delay time and SNR with S Ⅱ 545.3 nm and P Ⅱ 602.4 nm, respectively

3.2 Selection of analytical lines of S and P in steel

Spectral line intensity can be used to determine the content of elements in steel, while qualitative analysis of the element can be implemented via the characteristic spectral lines. The suitable characteristic spectral lines of S and P in steel are required to improve the predictive accuracy of RF method. The averaged LIBS spectrum of the 1# steel sample recorded at the optimized conditions is displayed in Fig. 4. It is observed that the spectral emission line of iron matrix is very dense in steel spectrum, which will lead to serious interference to characteristic spectral lines of other elements, especially the spectral lines of trace element S and P. The analytical

lines of S and P would be interfered strongly by almost overlapping iron matrix, so a few analytical lines remained were investigated to quantitative analysis of S and P in steel. Therefore, the characteristic LIBS spectral lines of S and P in steel were identified based on the NIST database. As shown as in Fig. 5 (a), the spectrum of S is mostly featureless over the UV range, with the exception of two overlapping S peaks around at 543.2 nm and 545.3 nm. Nevertheless, S Ⅱ 543.2 nm is interfered obviously stronger than that of 545.3 nm that is selected as the primary analytical line of S. LIBS spectra of P with the highest selectivity was obtained at an excitation wavelength corresponding to P Ⅱ 602.4 nm among transitions of atomic phosphorus in 600-601 nm, as shown in Fig. 5 (b).

Here is Fig. 4

Fig. 4 The averaged LIBS spectrum of the 1# steel sample recorded at optimized conditions

Here is Fig. 5

Fig. 5 Typical LIBS spectra of 1# sample with analytical line of S Ⅱ 545.3 nm of interest (a) and with analytical line of P Ⅱ 602.4 nm of interest (b)

3.3 Optimization of parameters of RF calibration model

The different spectral range (with the range of 220-800 nm, 500-650 nm, 520-620 nm and 540-610 nm) as input variable for RF calibration model were investigated by means of root mean square error of calibration (RMSEC) and R^2 in order to improve the precision ratio of quantitative analysis result and the prediction ability in this study. Within the scope of each band selected, the smaller the root mean square error, the greater the correlation coefficient, then the better the model predictive ability of RF model. The performance of RF calibration model is not same through RMSEC and R^2 in different spectral range as seen as in Table 1. The R^2 of the RF calibration model with the input variable of 200-800 nm spectral range is the largest than of the rests. However, the smallest RMSEC is 0.0002 in the wavelength range of 540-610 nm. Hence, the 520-620 nm spectral range is selected as input variable for RF calibration model.

The two parameters (**ntree** and $\mathbf{m_{try}}$) of the RF model were optimized by OOB estimation. The stability of RF calibration model can be reflected by the optimized **ntree**, simultaneously the optimized $\mathbf{m_{try}}$ can effectively improve the accuracy and precision of RF calibration model. Even the **ntree** and $\mathbf{m_{try}}$ also have an impact on the running speed and time of the Matlab (2007a). Therefore, it is necessary to optimize these two parameters, and its optimization process are two

steps: (1) Trained in the whole training set, according to OOB error to choose **ntree**, and the choose standard is to make the total error or target category error low and stable enough; (2) The optimal value of $\mathbf{m_{try}}$ parameter is optimized by cross-validation employing training set, which parameter selection criteria can be determined according to the actual problem, minimum or maximum error. In the case of the $\mathbf{m_{try}}$ unchanged, the affect of different **ntree** (**ntree**=100, 200, 300, 400, 500, 600 and 700, respectively) to the predictive ability of RF calibration model is studied in this study. On the contrary, the influence of variational $\mathbf{m_{try}}$ ($\mathbf{m_{try}}$=516, 589, 688, 825, 1031, 1375, 2063 and 4125, respectively) on calibration model predictive ability is investigated by keeping the **ntree** unchanged. Fig.6 shows the relationship of OOB error rate with **ntree** and $\mathbf{m_{try}}$, and the OOB error rate (2.778×10^{-6}) is the smallest when the $\mathbf{m_{try}}$ is 1031, and meanwhile the **ntree** is 200, which has a faster running speed. Consequently, the two optimized parameters of RF are as follows: **ntree** = 200 and $\mathbf{m_{try}}$ = 1031 are based upon the OOB error rate.[42]

Here is Table 1

Table 1 RMSEC and R^2 with different input data of RF calibration

Here is Fig. 6

Fig. 6 Relationship of OOB error rate with **ntree** and $\mathbf{m_{try}}$

3.4 Prediction the content of S and P in steel by RF model

Predictive ability is often a primary goal of data analyses for steel samples. RF model is a particularly appropriate tool and has been broadly used to predict metallurgical outcomes under various high-throughput metallurgical platforms. In this study, the 1# - 7# (7 samples × 6 spectra per sample = 42 records) samples as calibrated data were used to construct the PLSR and RF model, and the rest 8# and 9# (2 samples × 6 spectra per sample = 12 records) as tested data were quantitatively predicted S and P in each steel samples. Here, the reference value and results obtained by RF calibration method were averaged and then plotted. In order to prove the RF calibration method has a better predictive ability, we have compared RF with PLSR to predict the content of S and P in steels. For the calibration model based on PLSR, the best latent variables for S and P optimized by 5-fold cross-validation were 10. The PLSR and RF model performance for S and P in the calibration samples are shown in Fig. 7 and Fig. 8, respectively. RF shows a better correlation relationship (R^2=0.9974 for S, R^2=0.9981 for P) than PLSR (R^2=0.8762 for S, R^2=0.9171 for P) for S and P in steels. Furthermore, the results obtained by PLSR and RF and

Series of Metal CRMs for spectral analysis are also given in Table 2 for comparison, showing that the predictive values of S and P are extremely closer to that of reference values in different steel samples. Therefore, RF model can relatively accurately quantitatively predict S and P in steel. Not only spectral interference from iron matrix and other elements has an effect on the quantitative prediction results, but also the content of S and P in steel. Seen from the Table 2, the higher content of S and P in steel, the smaller averaged relative error, like the averaged relative error of S of 8# (2.69%) in steel was smaller than 9# (3.47%) which the content of S is larger than of 8# no matter what the PLSR or RF. When considered the R^2 and ARE of both S and P, the RF model is more suitable for the determination of P than S in steel.

Here is Fig. 7

Fig. 7 The PLSR and RF model performance for S in the calibration samples using S Ⅱ 545.3 nm as analytical line

Here is Fig. 8

Fig. 8 The PLSR and RF model performance for P in the calibration samples using P Ⅱ 602.4 nm as analytical line

Here is Table 2

Table 2 Comparison of Prediction Results for S and P Obtained by PLSR and RF Calibration Method with Series of Metal CRMs for Spectral Analysis

Conclusion

In this study, S and P in steel can be quantitative detected by LIBS and RF model using analytical lines of S Ⅱ 545.3 nm and P Ⅱ 602.4 nm without spectral interferences from iron matrix and other elements. The experimental condition (delay time-1.5 μs and laser pulse energy-60 mJ) were optimized by SNR of S and P. The RF calibration model was constructed under optimized spectral range (520-620 nm) and parameters (200 **ntree** and 1031 $\mathbf{m_{try}}$). The predictive values obtained by the RF calibration model are close to the reference values. Results also demonstrated RF has a better correlation relationship (R^2=0.9974 for S, R^2=0.9981 for P) than PLSR (R^2=0.8762 for S, R^2=0.9171 for P) for S and P in steels. The above results demonstrated that the RF model is an effective approach for the nonmetal elements quantitative analysis of steel samples. Thus, RF model will be a promising regression method for remote, real-time and in-situ analysis for quality supervision and process control in steel industry, especially for nonmetal elements in steel.

Acknowledgements

This research was financially supported by the National Special Fund for the Development of Major Scientific Instruments and Equipment (No. 2011YQ030113) of China, National Natural Science Foundation of China (No. 21175106 and No. 21375105), Research Fund for the Doctoral Program of Higher Education of China (No.20126101110019), and NWU Graduate Innovation and Creativity Funds (No. YZZ13020).

Notes and references

1. M. A. Khater, *Spectrochimica Acta Part B*, 2013, **81**, 1-10.
2. E. Nagels, J. R. Duflou, J. Van Humbeeck, *Journal of Materials Processing Technology*, 2007, **194**, 159-162.
3. C. Li, Z. Zou, X. Yang, Z. Hao, L. Guo, X. Li, Y. Lu, X. Zeng, *Journal of Analytical Atomic Spectrometry*, 2014, **29**, 1432-1437.
4. A. Varghese, L. George, *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy*, 2012, **95**, 46-52.
5. V. K. Ponnusamy, J. Jen, *Journal of Chromatography A*, 2011, **1218**, 6861-6868.
6. H. Ida, J. Kawai, *Analytical and Bioanalytical chemistry*, 2004, **379**, 735-738.
7. A. Sengupta, B. Rajeswari, R. M. Kadam, R. Acharya, *Atomic Spectroscopy*, 2011, **32**, 200-205.
8. M. Scharun, C. Fricke-Begemann, R. Noll, *Spectrochimica Acta Part B*, 2013, **87**, 198-207.
9. G. P. Gupta, B. M. Suri, A. Verma, M. Sundararaman, V. K. Unnikrishnan, K. Alti, V. B. Kartha, C. Santhosh, *Journal of Alloys and Compounds*, 2011, **509**, 3740-3745.
10. D. A. Cremers and R. C. Chinni, *Applied Spectroscopy Reviews*, 2009, **44**, 457-506.
11. D. W. Hahn and N. Omenetto, *Applied Spectroscopy*, 2010, **64**, 335A-366A.
12. D. A. Rusak, B. C. Castle, B. W. Smith and J. D. Winefordner, *Critical Reviews in Analytical Chemistry*, 1997, **27**, 257-290.
13. R. S. Harmon, F. C. De Lucia, A. W. Miziolek, K. L. McNesby, R. A. Walters and P. D. French, *Geochemistry: Exploration, Environment, Analysis*, 2005, **5**, 21-28.
14. J. E. Carranza, B. T. Fisher, G. D. Yoder and D. W. Hahn, *Spectrochimica Acta Part B-Atom Spectroscopy*, 2001, **56**, 851-864.
15. S. Guirado, F. J. Fortes, V. Lazic, J. J. Laserna, *Spectrochimica Acta Part B-Atomic Spectroscopy*, 2012, **74-75**, 137-143.
16. J. Gruber, J. Heitz, N. Arnold, D. Bauerle, N. Ramaseder, W. Meyer, J. Hochortler, F. Koch,

Applied Spectroscopy, 2004, **58**, 457-62.

17. G. A. Lithgowa, A. L. Robinsonb, S. G. Buckleya, *Atmospheric Environment*, 2004, **38**, 3319-3328.
18. R. Gaudiuso, M. DellAglio, O. De Pascale, S. Loperfido, A. Mangone, A. De Giacomo, *Analytica Chimica Acta*, 2014, **813**, 15-24.
19. B. Salle, J. L. Lacour, P. Mauchien, P. Fichet, S. Maurice and G. Manhes, *Spectrochimica Acta Part B-Atomic Spectroscopy*, 2006, **61**, 301-313.
20. S. Manzoor, S. Moncayo, F. Navarro-Villoslada, J.A. Ayala, R. Izquierdo-Hornillos, F.J. Manuel de Villena, J. O. Caceres, *Talanta*, 2014, **121**, 65-70.
21. B. Salle, P. Mauchien and S. Maurice, *Spectrochimica Acta Part B-Atomic Spectroscopy*, 2007, **62**, 739-768.
22. P. Rohwetter, J. Yu, G. Mejean, K. Stelmaszczyk, E. Salmon, J. Kasparian, J.-P. Wolf, L. Woste, *Journal of Analytical Atomic Spectrometry*, 2004, **19**, 437-444.
23. F. C. De Lucia, J. L. Gottfried, C. A. Munson, and A. W. Miziolek, Applied Optics, 2008, **47**, G112-G121.
24. Z. B. Cong, L. X. Sun, Y. Xin, Y. Li, L. F. Qi, Z. J. Yang, *Spectroscopy and Spectral Analysis*, 2014, **34**, 542-547.
25. L. Xu, V. Bulatov, V. V. Gridin, I. Schechter, *Analytical Chemistry*, 1997, **69**, 2103-2108.
26. L. Zheng, D. G. Watson, B. F. Johnston, R. L. Clark, R. Edrada-Ebel, W. Elseheri, *Analytica Chimica Acta*, 2009, **642**, 257-265.
27. T. Zhang, L. Liang, K. Wang, H. Tang, X. Yang, Y. Duan and H. Li, *Journal of Analytical Atomic Spectrometry*, 2014, DOI: 10.1039/C4JA00217B.
28. J. EI Haddad, M. Villot-Kadri, A. Ismael, G. Gallou, K. Michel, D. Bruyere, V. Laperche, L. Canioni, B. Bousquet, *Spectrochimica Acta Part B-Atomic Spectroscopy*, 2013, **79**, 51-57.
29. B. Sirven, B. Bousquet, L. Canioni, L. Sarger, S. Tellier, M. Potin-Gautier, I. L. Hecho, *Analytical and Bioanalytical Chemistry*, 2006, **385**, 256-262.
30. B. Sirven, B. Bousquet, L. Canioni, L. Sarger, *Analytical Chemistry*, 2006, **78**, 1462-1469.
31. P. Inakollu, T. Philip, A. K. Rai, F. Y. Yueh, J. P. Singh, *Spectrochimica Acta Part B-Atomic Spectroscopy*, 2009, **64**, 99-104.
32. L. Liang, T. Zhang, K. Wang, H. Tang, X. Yang, X. Zhu, Y. Duan, H. Li, *Applied Optics*, 2014, **53**, 544-552.

33. J. Cisewski, E. Snyder, J. Hannig, L. Oudejans, *Journal of Chemometrics*, 2012, **26**, 143-149.

34. N. C. Dingari, I. Barman, A. K. Myakalwar, S. P. Tewari, M. K. Gundawar, *Analytical Chemistry*, 2012, **84**, 2686-2694.

35. J. C. W. Chan, D. Paelinckx, *Remote Sensing of Environment*, 2008, **112**, 2999-3011.

36. L. Breiman, *Machine Learning*, 2001, **45**, 5-32.

37. S. Serneels, P. J. Van Espen, *Analytica Chimica Acta*, 2005, **544**, 153-158.

38. N. L. Afanadora, T. N. Tranb, L. M. C. Buydensc, *Analytica Chimica Acta*, 2013, **768**, 49-56.

39. M. Kuhnlein, T. Appelhans, B. Thies, T. Nauss, *Remote Sensing of Environment*, 2014, **141**, 129-143.

40. A. Philibert, C. Loyce, D. Makowski, *Environmental Pollution*, 2013, **177**, 156-163.

41. L. M. Cabalin, J. J. Laserna , *Spectrochimica Acta Part B-Atomic Spectroscopy*, 1998, **53**, 723-730.

42. L. Breiman, *Machine Learning*, 2001, **45**, 5-32.

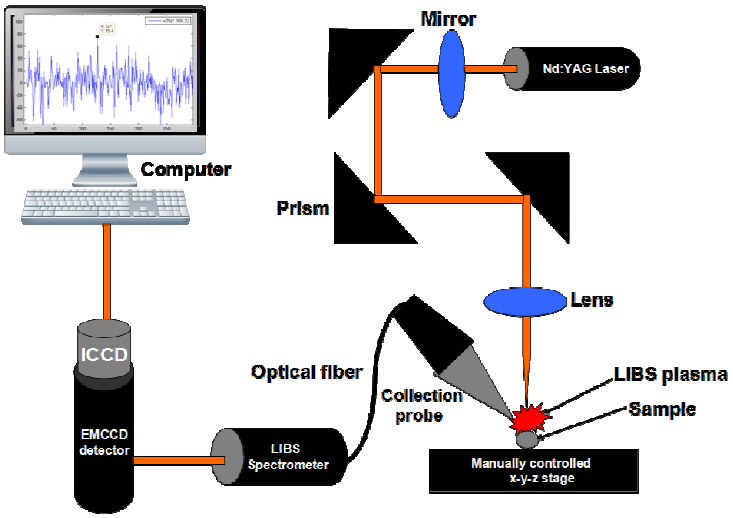


Fig. 1 Schematic diagram of the experimental set-up for LIBS studies

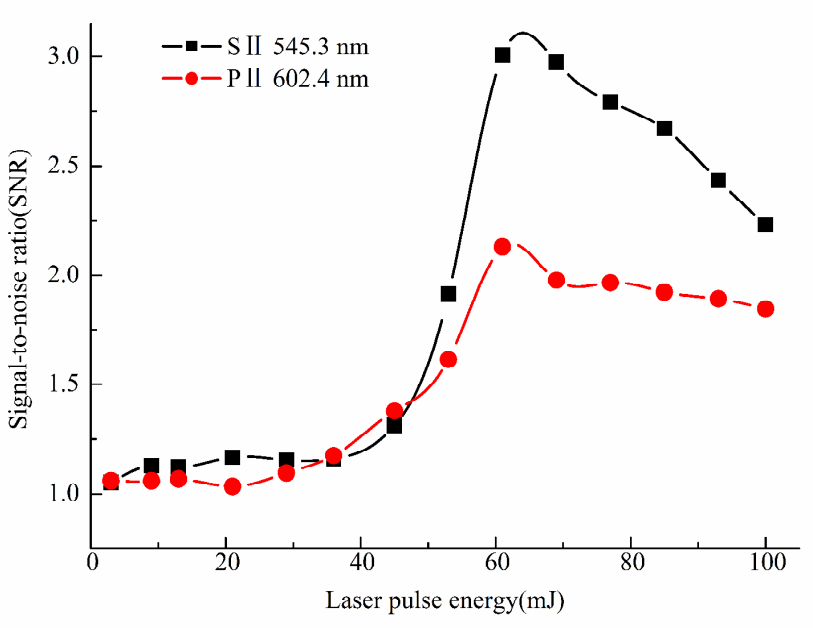


Fig. 2 The relationship between the laser pulse energy and SNR with S II 545.3 nm and P II 602.4 nm, respectively

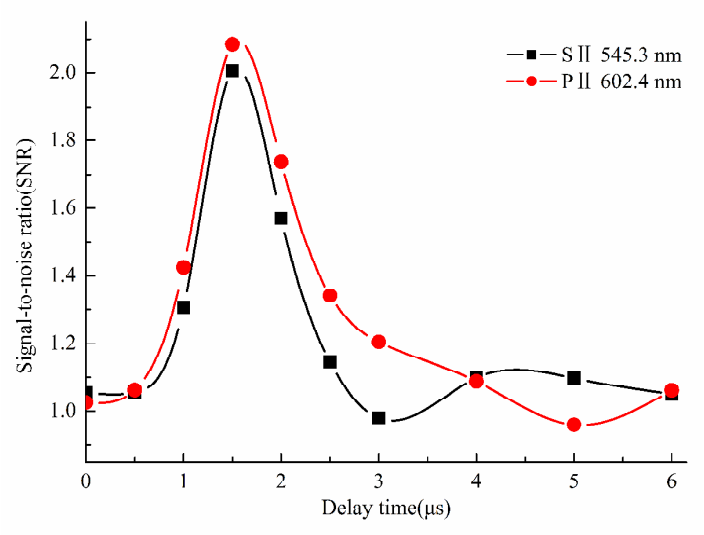


Fig. 3 The relationship between the delay time and SNR with S Ⅱ 545.3 nm and P Ⅱ 602.4 nm, respectively

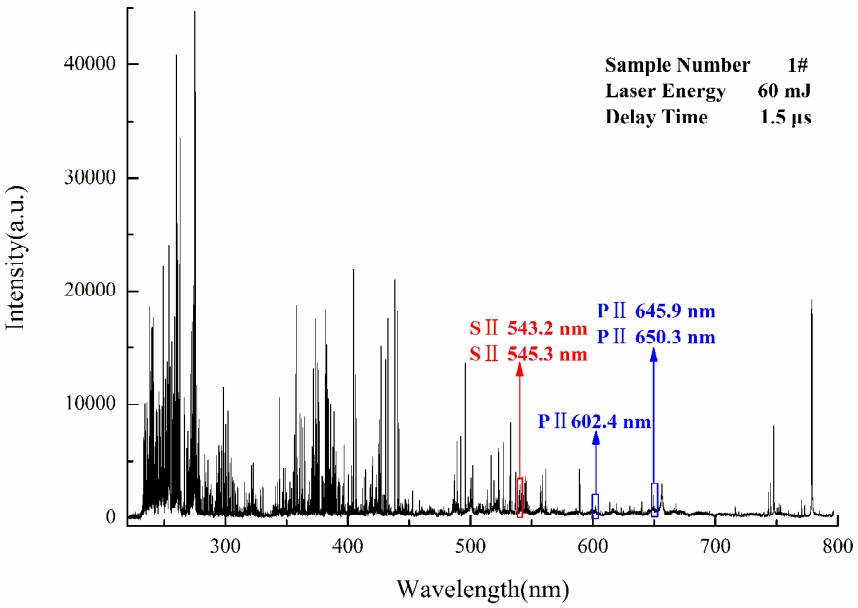


Fig. 4 The averaged LIBS spectrum of the 1# steel sample recorded at optimized conditions

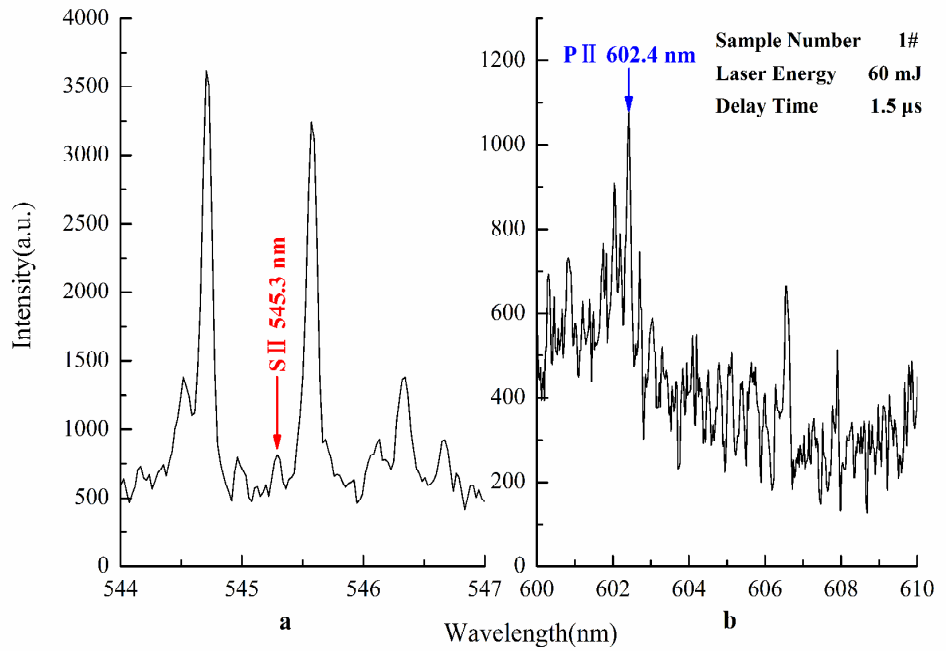


Fig. 5 Typical LIBS spectra of 1# sample with analytical line of S Ⅱ 545.3 nm of interest (a) and with analytical line of P Ⅱ 602.4 nm of interest (b)

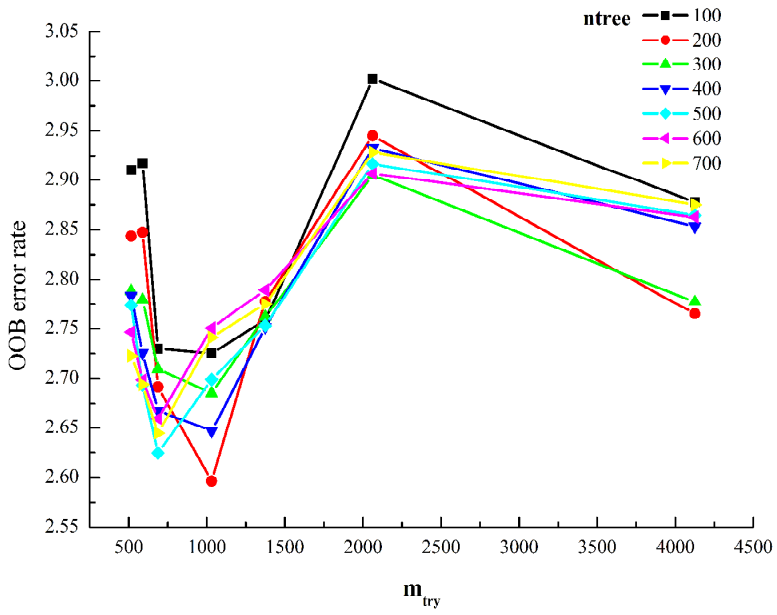


Fig. 6 Relationship of OOB error rate with **ntree** and $\mathbf{m_{try}}$

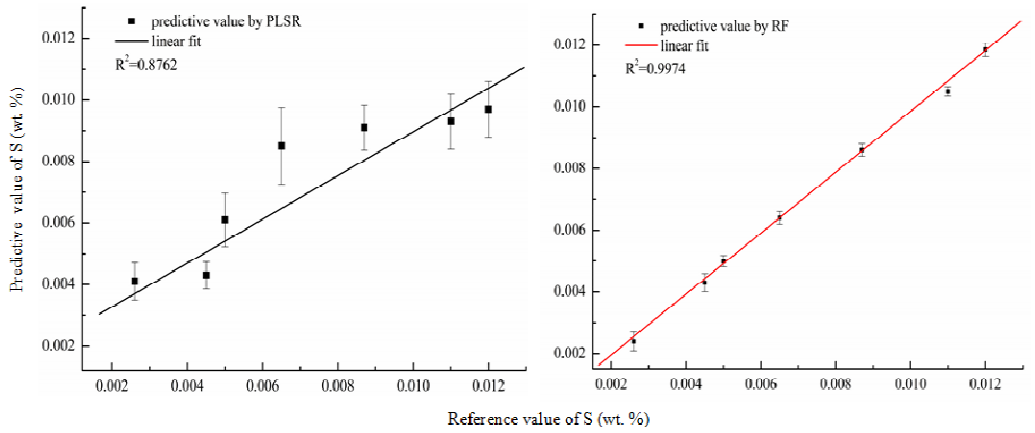


Fig. 7 The PLSR and RF model performance for S in the calibration samples using S Ⅱ 545.3 nm as analytical line

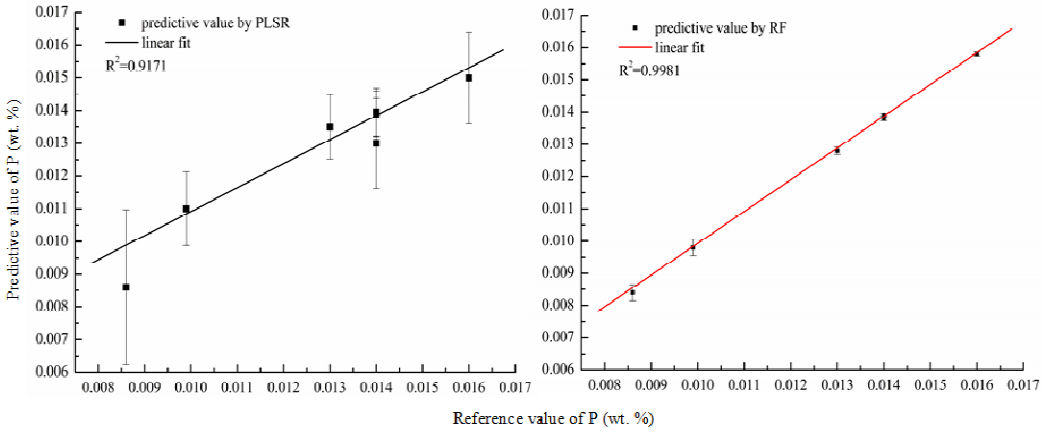


Fig. 8 The PLSR and RF model performance for P in the calibration samples using P Ⅱ 602.4 nm as analytical line

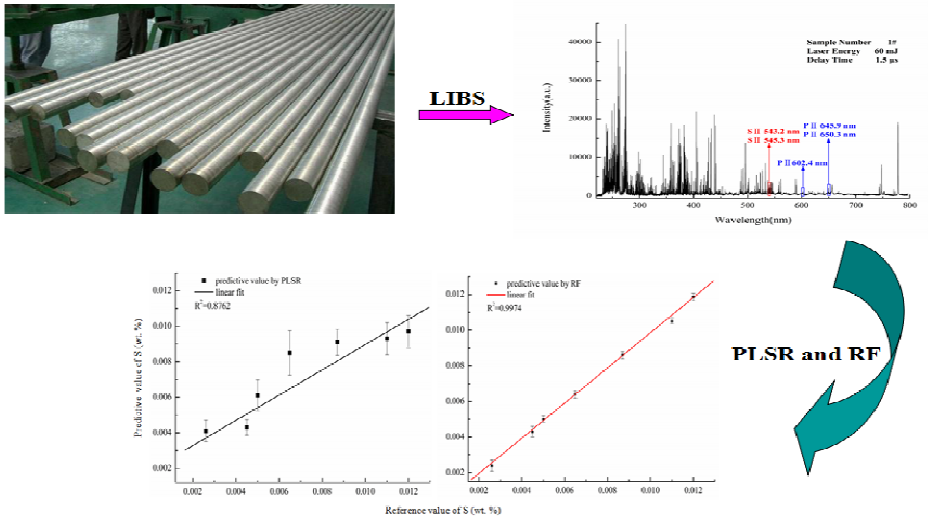
Table 1 RMSEC and R^2 with different input data of RF calibration

Wavelength range (nm)	RMSEC	R^2 of calibration
200-800	0.0007	0.9772
500-650	0.0006	0.9753
520-620	0.0004	0.9769
540-610	0.0002	0.9598

Table 2 Comparison of Prediction Results for S and P Obtained by PLSR and RF Calibration Method with Series of Metal CRMs for Spectral Analysis

No. Sample	Reference value (wt%)		PLSR				RF			
	S	P	S	ARE (%)	P	ARE (%)	S	ARE (%)	P	ARE (%)
8#	0.013	0.014	0.0139	13.97	0.0136	10.11	0.0127	2.69	0.0142	1.77
			0.0141		0.0116		0.0128		0.0143	
			0.0152		0.0113		0.0132		0.0145	
			0.0156		0.0115		0.0132		0.0138	
			0.0154		0.0143		0.0134		0.0137	
			0.0147		0.0138		0.0127		0.014	
9#	0.0072	0.014	0.0089	16.43	0.0143	9.17	0.0074	3.47	0.0137	0.83
			0.0084		0.0132		0.0075		0.0139	
			0.0071		0.0133		0.0071		0.0139	
			0.0086		0.0145		0.0075		0.0141	
			0.0064		0.0145		0.007		0.014	
			0.0062		0.0144		0.0076		0.0141	

Graphical Abstract



Laser induced breakdown spectroscopy (LIBS) combined with PLSR and RF was employed for detection of nonmetal elements in steels.