

PAPER

Calibration curve and support vector regression methods applied for quantification of cement raw meal using laser-induced breakdown spectroscopy

To cite this article: Junwei JIA *et al* 2019 *Plasma Sci. Technol.* **21** 034003

View the [article online](#) for updates and enhancements.

Calibration curve and support vector regression methods applied for quantification of cement raw meal using laser-induced breakdown spectroscopy

Junwei JIA (贾军伟)^{1,2}, Hongbo FU (付洪波)¹, Zongyu HOU (侯宗余)³,
Huadong WANG (王华东)^{1,2}, Zhibo NI (倪志波)¹ and Fengzhong DONG (董凤忠)^{1,2}

¹ Anhui Provincial Key Laboratory of Photonic Devices and Materials, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, People's Republic of China

² University of Science and Technology of China, Hefei 230022, People's Republic of China

³ State Key Lab of Power Systems, Department of Thermal Engineering, Tsinghua-BP Clean Energy Center, Tsinghua University, Beijing 100084, People's Republic of China

E-mail: fzdong@aiofm.cas.cn

Received 20 July 2018, revised 22 September 2018

Accepted for publication 24 September 2018

Published 23 November 2018



CrossMark

Abstract

Laser-induced breakdown spectroscopy (LIBS) is a qualitative and quantitative analytical technique with great potential in the cement industrial analysis. Calibration curve (CC) and support vector regression (SVR) methods coupled with LIBS technology were applied for the quantification of three types of cement raw meal samples to compare their analytical concentration range and the ability to reduce matrix effects, respectively. To reduce the effects of fluctuations of the pulse-to-pulse, the unstable ablation and improve the reproducibility, all of the analysis line intensities were normalized on a per-detector basis. The prediction results of the elements of interest in the three types of samples, Ca, Si, Fe, Al, Mg, Na, K and Ti, were compared with the results of the wet chemical analysis. The average relative error (ARE), relative standard deviation (RSD) and root mean squared error of prediction (RMSEP) were employed to investigate and evaluate the prediction accuracy and stability of the two prediction methods. The maximum average ARE of the CC and SVR methods is 34.62% instead of 6.13%, RSD is 40.89% instead of 7.60% and RMSEP is 1.34% instead of 0.43%. The results show that SVR method can accurately analyze samples within a wider concentration range and reduce the matrix effects, and LIBS coupled with it for a rapid, stable and accurate quantification of different types of cement raw meal samples is promising.

Keywords: cement raw meal, calibration curves, SVR, LIBS

(Some figures may appear in colour only in the online journal)

1. Introduction

For a long time, cement as an important cementitious material and the basic ingredient of concrete, is widely used in construction, transportation, water conservancy, national defense and other projects. At the same time, the cement industry is one of the major contributors to energy consuming and greenhouse gas emissions, specifically CO₂

emission, about one-eighth of China's national CO₂ emissions [1]. What's more, due to lack of composition on-line monitoring, especially the raw meal composition monitor, the average qualified rate of cement is actually less than 70%. Therefore, the monitoring of the elemental composition of the cement raw meal has a significant impact on quality and safety of a concrete structure, energy saving and CO₂ emission reduction.

Nowadays, several methods have been applied to the quality control of cement analysis, such as x-ray fluorescence (XRF) method [2–4], wet chemical analysis and spark induced breakdown spectroscopy (SIBS) method [5]. However, XRF requires a vacuum environment, high maintenance costs, stringent sample preparation, and wet chemical analysis methods not only are time consuming, but also require a variety of chemical agents that pollute the environment, and SIBS method is not suitable for measuring minor and non-metallic elements for its poor ablation ability.

Laser-induced breakdown spectroscopy (LIBS) is an atomic emission spectroscopy technique that can be used for quantitative and qualitative analysis of solid, liquid, gas, and aerosol samples [6–10], with simple or no sample preparation and rapid non-destructive multi-element analysis [11]. Many scholars have used LIBS technology to quantitatively study cement or cement raw materials. Gondal *et al* used LIBS coupled with Calibration curve (CC) method for determination of elemental composition in three different types of cement samples and the predicted values compared with ICP-ES were in good agreement within the 2% error limit [12]. Mansoori *et al* also used LIBS coupled with CC method for a quantitative analysis of the different elements under different optimum delay time, and many elements got a lower limit of detection [13]. Christoph *et al* compared the LIBS spectral lines in the infrared and ultraviolet-range for chlorine detection in hydrated cement samples using LIBS combined with CC method [14]. Yin *et al* used a new spectrum standardization method to investigate the capability of LIBS for the analysis of three elemental ratios of raw materials [15]. Fan *et al* established a calibration model for on-site quality analysis of cement and the measurement results had a good agreement with the results from the traditional method [16]. Most of the analytes (prediction samples) and the calibration samples used to establish the (CCs) or predicted models are the same series of standard samples, they have a similar matrix, and the concentrations of the analytes are within the calibration samples concentration range. Therefore, quantitative analysis of cement samples using the CC method can give good predictions. However, in the actual cement industry due to the different sources of raw materials, the matrix of the cement raw meal is diversified and the elemental composition of it may also be outside the calibration samples concentration range.

Some scholars have successfully combined LIBS with support vector regression (SVR) method for the quantification of bronze [17], soil [18], slag [19] and rocks [20, 21] samples, these results shown that SVR method has a better performance in dealing with nonlinear problems and reducing matrix effects. However, we have not found any article on applying SVR method to quantify cement or cement materials.

In this study, three types of cement raw meal samples were used as the analytes: (1) samples having a similar matrix to the calibration samples and the concentration within the calibration samples concentration range; (2) samples having a similar matrix to the calibration samples and the concentration

is not within the calibration samples concentration range; (3) samples having a dissimilar matrix to the calibration samples. The analytical concentration range and the ability to reduce matrix effects of the CC and SVR methods were compared. To evaluate the accuracy and stability of the two methods for quantitative analysis, the results were also compared to those obtained by wet chemical analysis method, respectively.

2. Experimental set-up

The principal scheme of the LIBS system is presented in figure 1. A nanosecond Q-switched Nd:YAG laser was employed at 1064 nm with a pulse energy of 32 mJ, a repetition rate of 1 Hz and a pulse duration of 7 ns. The laser pulses focused on the surface of the sample by the beam splitter 1 and a 100 mm focal length quartz lens (Lens1) to generate plasma by the ablation, excitation and ionization processes. After the plasma emission reflected by beam splitter 2 (about 85% reflectivity) and collected by Lens 2, it was coupled into a 1.5 m long optical fiber. A six-channel spectrometer, which has six separate detectors (198–323 nm; 315–424 nm; 413–508 nm; 497–571 nm; 559–785 nm; 779–971 nm), was used to detect the wavelength region from 198 to 971 nm with the delay time of 1.28 μ s and the integration time of 1.05 ms, at an average resolution of 0.1 nm. To prevent the air breakdown in front of the sample, the distance of the sample and the focusing lens was a little shorter than the focal length of the Lens1. The sample mounted on an X–Y–Z motorized stage to provide a fresh surface and to avoid deep crater. Unlike the typical LIBS experimental setup, a fan was used to provide purge gas (air) and prevent the next laser pulse from focusing on the lifted powder, which will affect the stability of the spectral data.

The powder samples were labeled as S1–S11, which were a series of certified reference material (CRM) samples from China Building Material Test & Certification Group Co., Ltd Ts1 and Ts2 were taken from two different cement plants with similar compositions and different matrices. Table 1 shows the certified values (wet chemical method) of the selected thirteen cement raw meal samples. Here, seven CRM samples labeled S2, S3, S4, S7, S8, S9 and S10 were used for preparing CC or training set, and three type of cement raw meal samples were used as unknown samples for prediction set: (1) S1, S11 with a similar matrix to the calibration samples and the concentration within the calibration samples concentration range; (2) S5, S6 with a similar matrix to the calibration samples and the concentration not within the calibration samples concentration range; (3) Ts1, Ts2 with a dissimilar matrix to the calibration samples.

The powder samples were pressed into 20 mm diameter and 3 mm thickness pellets with an electric hydraulic jack under a pressure of 25 MPa in 6 min, 4 min of dwell time, and 2 min of release time. The spectra from the first five shots were discarded to minimize any surface contamination, and each pellet sample was analyzed using 10 locations with 25

3.1. CC method

CC method is a conventional quantitative method, also known as standard curve method for determining the component concentration in an unknown sample by comparing the unknown to a series of standard samples of known concentration.

In order to evaluate the ability of the CC method to analyze cement powder samples, (CCs) of Ca, Si, Fe, Al, Mg, Na, K, and Ti were prepared. The (CCs) for each sample of the elements were shown in figure 2. In figure 2, the error bars corresponded to the standard deviation (SD) of 10 ($n = 10$) independent measurements and were calculated by equation (2)

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2)$$

The x -axis and y -axis refer to the normalized spectral intensity and concentration of the different components in the sample, respectively.

Figure 2 shows that there is a good linear relationship between the elements concentration and the spectral intensity of the calibration samples. Because the calibration samples have similar matrix, the regression coefficients for all (CCs) are between 0.9059 and 0.9956.

Although the (CCs) have high R^2 , there also exist a larger deviation between the predicted value and the certified value and a larger measurement SD, especially for the elements in the samples S1, S11, Ts1, and Ts2, whose concentrations are not within the calibration samples concentration range or have a dissimilar matrix with the calibration samples.

The average relative error (ARE), relative standard deviation (RSD) and root mean squared error of prediction (RMSEP) were employed to evaluate the repeatability and stability of the measurement. Larger SD or RSD indicate instability of multiple measurements, and larger ARE and RMSEP indicate poor measurement accuracy. The definitions of ARE, RSD and RMSEP were described in the following equations

$$ARE = \frac{|\bar{x} - x_{\text{real}}|}{x_{\text{real}}}, \quad (3)$$

$$RSD = \frac{1}{\bar{x}} \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (4)$$

$$RMSEP = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - x_{\text{real}})^2}, \quad (5)$$

where x_i are the values of prediction concentrations, \bar{x} is the average value of the prediction concentrations from different measurements, x_{real} is the certified concentration value of element i , and n is the total number of measurements.

From table 2, it can be seen that the ARE, RSD, and RMSEP values of the Ts1 and Ts2 are significantly larger than those of the S1 and S11 samples, and the values of the S1 and S11 sample are larger than those of the S5 and S6 samples. All of the results for the three types of samples indicate that the ARE, RSD and RMSEP values of the major elements (Ca, Si, Fe, Al) are smaller than those of the minor

elements (Mg, Na, K, Ti). Among them, the prediction result of Ca element is the best and has the largest SNR because the content of Ca is the highest in the sample.

The results illustrate that the CC method is greatly influenced by the matrix effect, and only when the concentration of the sample to be measured is within the calibration sample concentration range can obtain a better predicted result. Therefore, another method called SVR, which can reduce matrix effects, was used to quantify the same samples.

3.2. SVR method

Support vector machine (SVM) is based on statistical theory and has a strict theoretical and mathematical basis. SVM is a method for learning, classifying and predicting (sometimes called regression) small sample data. It can solve the over-learning problem that neural network cannot solve, and it has good generalization ability.

SVR is an extension of SVM and is a kind of multivariate analysis method. The SVR algorithm mainly realizes linear regression by constructing a linear decision function in high-dimensional space after dimensioning. In order to adapt to the nonlinearity of the training sample set, the traditional fitting method usually adds a high order term after the linear equation. The traditional fitting method is effective, but the increased tunable parameters increase the risk of overfitting. The SVR algorithm uses a kernel function to solve this problem. Replacing a linear term in a linear equation with a kernel function can make the original linear algorithm ‘non-linear’, that is, can perform nonlinear regression.

The software of LIBSVM [23] was used to implement SVR. The SVR theory is briefly introduced as follows. For a detailed theoretical background, refer to [24, 25].

$T = \{(x_1, y_1), \dots, (x_n, y_n)\}$ is a given data set, where x_n and y_n are the input and output vectors. The standard form of SVR can be posed as:

$$f(x) = \omega^T \varphi(x) + b, \quad (6)$$

where $\varphi(x)$ maps the input vector x to a vector in the feature space, ω and b denote weight vectors and offsets, respectively. In order to minimize the estimated true risk of equation (6), the minimization objective function is constructed as follows:

$$\begin{aligned} \min_{\omega, b, \xi, \xi^*} & \frac{1}{2} \omega^T \omega + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{Subject to } & \varepsilon + \xi_i^* \leq y_i - (\omega \varphi(x_i) + b) \leq \varepsilon + \xi_i, \\ & \xi_i \geq 0, \xi_i^* \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (7)$$

where $\frac{1}{2} \omega^T \omega$ represents the structural information of the model, and $C \sum_{i=1}^n (\xi_i + \xi_i^*)$ represents the penalty of the error for the regression function. C is the balance coefficient called penalty factor, used for the adjustment between the model complexity and the training error to obtain a better generalization ability. ξ_i and ξ_i^* are the relaxation variables that represent upper and lower training errors, respectively. The minimization objective function can be converted to

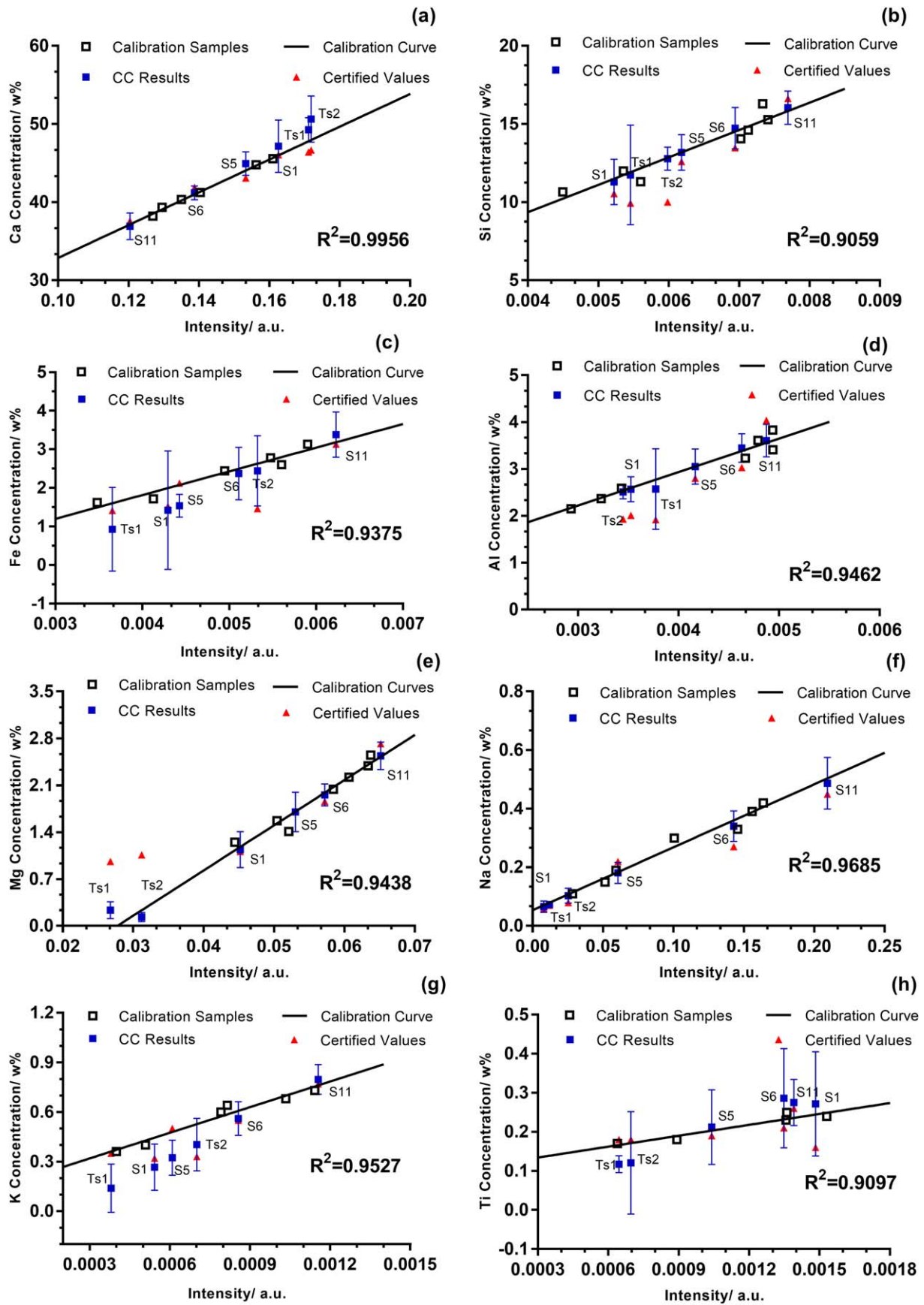


Figure 2. (CCs) and CC predicted values versus certified values of elements: (a) Ca, (b) Si, (c) Fe, (d) Al, (e) Mg, (f) Na, (g) K, (h) Ti.

Table 2. The average ARE, RSD and RMSEP of CC for three types of cement raw meal samples.

Samples	Elements	Ca (%)	Si (%)	Fe (%)	Al (%)	Mg (%)	Na (%)	K (%)	Ti (%)	Average (%)
S1, S11	ARE	2.11	5.36	7.23	19.31	14.65	18.62	30.08	37.93	16.91
	RSD	5.86	9.78	62.62	10.06	15.91	21.35	32.06	35.35	24.12
	RMSEP	2.56	1.37	1.03	0.58	0.26	0.06	0.12	0.11	0.76
S5, S6	ARE	1.97	3.96	5.32	11.44	12.77	21.85	25.50	24.01	13.35
	RSD	2.77	8.85	23.89	10.57	15.86	17.35	43.45	44.80	20.94
	RMSEP	1.72	1.50	0.65	0.47	0.23	0.07	0.15	0.12	0.61
Ts1, Ts2	ARE	7.33	22.97	50.82	32.06	81.43	7.16	41.25	33.94	34.62
	RSD	4.49	16.45	77.35	19.57	52.04	20.76	72.79	63.64	40.89
	RMSEP	4.03	3.19	1.31	0.82	1.06	0.01	0.23	0.11	1.34

equation (8), by introducing the Lagrange coefficients α_i, α_i^*

$$\sum_{i=1}^n (-\alpha_i^* + \alpha_i)K(x_i, x) + b, \tag{8}$$

where K is the kernel function, α_i and α_i^* are the Lagrange coefficients, the nonlinear radial basis function was used in this study

$$K(x_i, x) = \exp\left(-\frac{1}{2\delta^2}|x_i - x|\right), \tag{9}$$

where $\frac{1}{2\delta^2}$ is a kernel parameter, in the LIBSVM methods often using g to represent it. The parameters of optimized significant penalty parameter C and the kernel parameter of g are more sensitive to the SVR model.

The grid search method has advantages for predicting small sample data, and is often used for parameter optimization in SVM and SVR models. It is to try all possible (C, g) value pairs, traverse each pair of parameters in the search grid, and use cross-validation to evaluate each pair of parameters to obtain an evaluation score indicator (mean square error). The evaluation indexes of each pair of parameters are compared to obtain an optimal parameter pair for model training. The optimal (C, g) pair has the highest cross-validation accuracy, and the lowest cross validation mean squared error (MSE). The grid search method combined with the five-fold cross-validation method was used to select the optimal (C, g) pair in this paper. As shown in figure 3, the parameters C and g selected by the grid search method are 256 and 0.329 88, respectively. The cross validation MSE is 0.25%. Based on the optimal parameters C and g , the SVR prediction model was built using the characteristic lines of the calibration samples.

From the linear fitting curves of SVR predicted values and certified values in figure 4, the R^2 of each element in all samples reached to 0.95, indicating a small difference between the predicted values and the certified values. In addition, the lower SD also means that the 10 measurements are closer and the measurements are stable.

According to table 3, the ARE, RSD, and RMSEP values of S1, S11, S5, S6, Ts1 and Ts2 samples indicated that using the SVR method, better prediction results were obtained for both main and minor elements in the three types of samples. Hence, the SVR method can not only reduce the effect of the matrix effects, but also obtain an acceptable prediction result

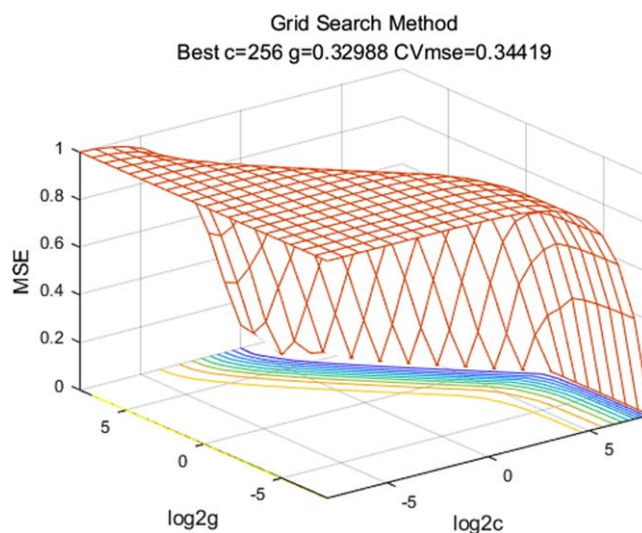


Figure 3. SVR model parameters C and g selection results by the grid search method.

when the sample concentration is not within the calibration sample concentration range.

Obviously, the maximum average ARE of the CC and SVR methods is 34.62% instead of 6.13%, RSD is 40.89% instead of 7.60% and RMSEP is 1.34% instead of 0.43%. Although the CC method predicts that the results of some samples are close to the certified values, the larger SD or RSD illustrates the instability of multiple measurements.

The results show that the SVR method can reduce the matrix effect and has a wider concentration measurement range. Even if the predicted sample and the calibration sample have a dissimilar matrix or the predicted sample concentration is not within the calibration sample concentration range, stable and accurate results can be obtained when the cement raw material sample is quantitatively analyzed using the LIBS combined with the SVR method.

4. Conclusions

In conclusion, LIBS coupled with CC and SVR methods have been successfully applied for quantification of the interest

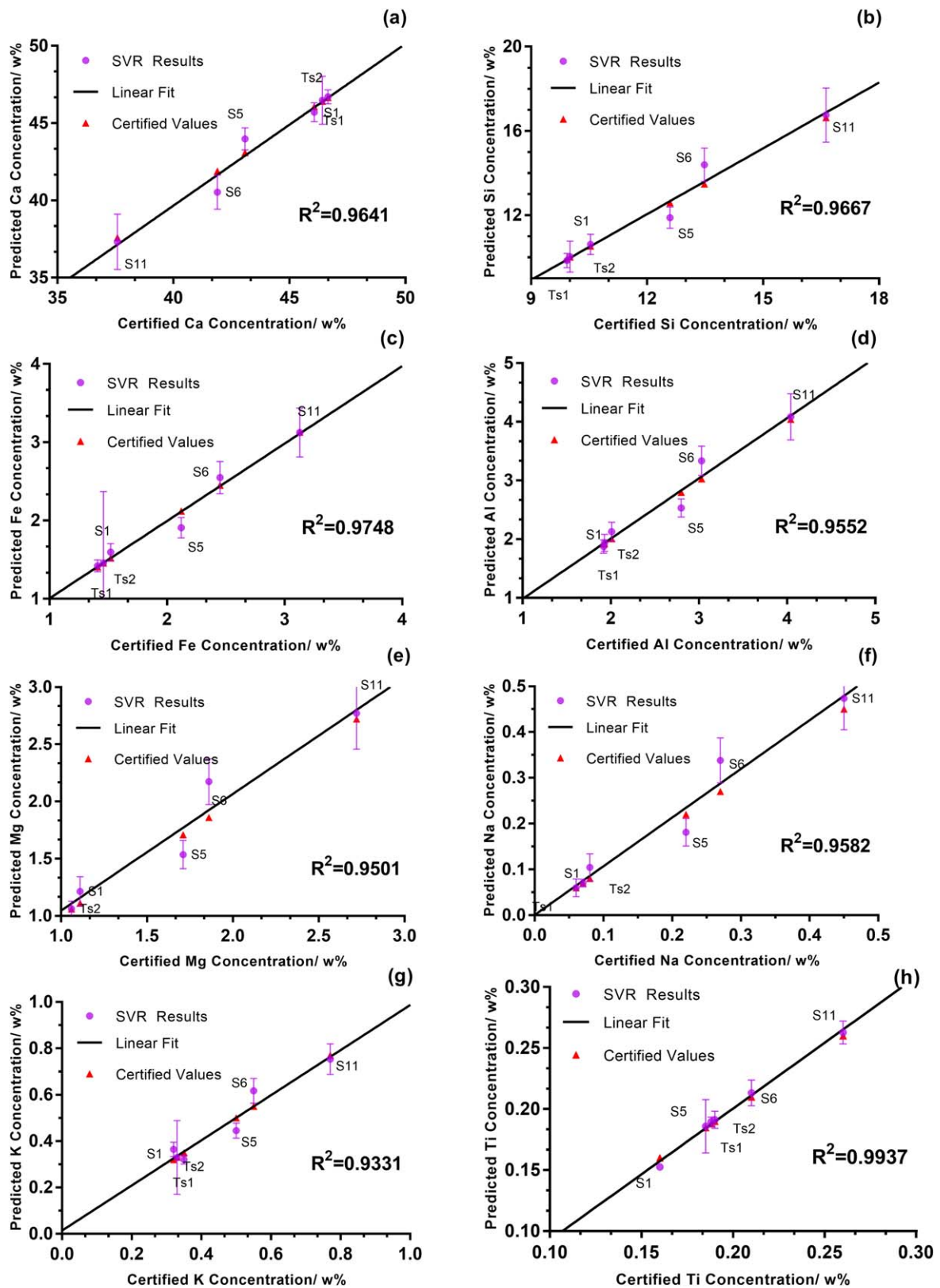


Figure 4. Linear fit of SVR predicted values versus certified values. (a) Ca, (b) Si, (c) Fe, (d) Al, (e) Mg, (f) Na, (g) K, (h) Ti.

elements of Ca, Si, Al, Mg, Fe, Na, K and Ti in three types of cement raw meal samples. In order to reduce the effects of pulse fluctuation fluctuations, unstable ablation and improved repeatability, all analysis line intensities were standardized on a per-

detector basis. The prediction results were compared with the wet chemical analysis results to evaluate their analytical concentration range and the ability to reduce matrix effects. The results proved that SVR method and wet chemical methods have

Table 3. The average ARE, RSD and RMSEP of SVR for three types of cement raw meal samples.

Samples	Elements	Ca (%)	Si (%)	Fe (%)	Al (%)	Mg (%)	Na (%)	K (%)	Ti (%)	Average (%)
S1, S11	ARE	0.76	0.73	2.56	3.48	5.60	6.88	8.00	3.24	3.91
	RSD	3.07	6.07	8.47	8.62	10.97	12.36	8.59	2.67	7.60
	RMSEP	1.21	0.84	0.21	0.28	0.23	0.05	0.06	0.03	0.36
S5, S6	ARE	0.11	1.36	0.42	1.36	2.92	0.17	4.31	4.32	1.87
	RSD	0.62	2.17	3.52	3.66	5.37	9.45	3.67	1.18	3.71
	RMSEP	0.28	0.21	0.05	0.21	0.06	0.01	0.02	0.01	0.11
Ts1, Ts2	ARE	2.66	0.17	7.00	9.81	3.54	2.47	11.53	5.89	6.13
	RSD	2.17	4.89	7.46	6.85	8.62	5.62	7.98	4.09	5.96
	RMSEP	1.42	1.02	0.23	0.35	0.29	0.07	0.07	0.02	0.43

good consistency than CC, especially when the analyte has a dissimilar matrix with the calibration samples or the concentration is outside of the calibration samples concentration range. The ARE, RSD, and RMSEP values of the two methods also show that the repeatability and stability of SVR method are better than those of CC method. The LIBS technology combined with the SVR method to achieve stable and accurate quantitative analysis of different types of raw cement samples is promising.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Grant Nos. 61505223, 41775128), the Knowledge Innovation Program of the Chinese Academy of Sciences (Grant No. Y03RC21124), the External Cooperation Program of Chinese Academy of Sciences (Grant No. GJHZ1726) and the project of China State Key Lab. of Power System (Grant Nos. SKLD18KM11, SKLD18M12).

References

- [1] Lei Y et al 2011 *Atmos. Environ.* **45** 147
- [2] Lemberge P, Van Espen P J and Vrebos B A R 2000 *X-Ray Spectrom.* **29** 297
- [3] Polat R et al 2004 *J. Quant. Spectrosc. Radiat. Transfer* **83** 377
- [4] Chang M T et al 2017 *Int. J. Adv. Eng. Sci. Appl. Math.* **9** 136
- [5] Taefi N, Khalaji M and Tavassoli S H 2010 *Cem. Concr. Res.* **40** 1114
- [6] Wang J G et al 2015 *Plasma Sci. Technol.* **17** 649
- [7] Zhang L et al 2012 *Front. Phys.* **7** 690
- [8] Hu L et al 2015 *Plasma Sci. Technol.* **17** 699
- [9] Zhang S et al 2018 *Front. Phys.* **13** 135201
- [10] Wang Z et al 2012 *Front. Phys.* **7** 708
- [11] Cremers D A and Radziemski L J 2006 *Handbook of Laser-Induced Breakdown Spectroscopy* (New Jersey: Wiley) (<https://doi.org/10.1002/0470093013>)
- [12] Gondal M A et al 2009 *Spectrosc. Lett.* **42** 171
- [13] Mansoori A et al 2011 *Opt. Lasers Eng.* **49** 318
- [14] Gehlen C D et al 2009 *Spectrochim. Acta B* **64** 1135
- [15] Yin H L et al 2016 *J. Anal. At. Spectrom.* **31** 2384
- [16] Li Y F et al 2016 *Spectrosc. Spectral Anal.* **36** 1494
- [17] Owolabi T O and Gondal M 2017 *J. Anal. At. Spectrom.* (<https://doi.org/10.1039/C7JA00229G>)
- [18] Gu Y H et al 2016 *Chin. Phys. Lett.* **33** 085201
- [19] Zhang T L et al 2015 *J. Anal. At. Spectrom.* **30** 368
- [20] Shi Q et al 2015 *J. Anal. At. Spectrom.* **30** 2384
- [21] Boucher T F et al 2015 *Spectrochim. Acta B* **107** 1
- [22] Barnett W B, Fassel V A and Kniseley R N 1968 *Spectrochim. Acta B* **23** 643
- [23] Fan R E et al 2005 *J. Mach. Learn. Res.* **6** 1889 (<http://jmlr.org/papers/v6/fan05a.html>)
- [24] Awad M and Khanna R 2015 *Support Vector Regression Efficient Learning Machines* (Berkeley, CA: Apress) (https://doi.org/10.1007/978-1-4302-5990-9_4)
- [25] Smola A J and Schölkopf B 2004 *Stat. Comput.* **14** 199