Quantitative Analysis of Steels using PLS with Three Data Reduction Methods based on LIBS

Haiyang Kong[1,2,3,a] **,** Lanxiang Sun[1,3,b]**,** Jingtao Hu[1,3]**,** Yong Xin[1,3]**,** Zhibo Cong[1,3]

[1]Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;

[2]University of Chinese Academy of Sciences, Beijing 100049, China;

[3]CAS Key Laboratory of Networked Control System, Shenyang 110016, China.

[a]E-mail: konghaiyang@sia.cn, [b]Email: sunlangxiang@sia.cn (corresponding author)

Abstract. Three methods, selecting characteristic lines of elements contained in the samples manually, selecting intensive spectral partitions manually and the whole spectra, were used to reduce dimensions of spectra of 27 steel samples acquired by Laser-Induced Breakdown Spectroscopy. The PLS models were built based on the data after dimension reduction to quantify the Mn concentration of samples. The results show that, PLS model built based on selecting intensive spectral partitions can achieve the best result with the least complexity and the highest generalization ability. Selecting intensive partitions is a promising solution to reduce dimensions for original spectra.

Introduction

It is a matter of great significance to analyze metals quantitatively in the metallurgical industry. However, the widely used methods currently is time-consuming and need tightened pretreatment of samples, so the needs of real-time and *in situ* analytical cannot be met.

Laser-Induced Breakdown Spectroscopy (LIBS) can achieve the purpose of qualitative and quantitative analysis of samples through analyzing the spectra inspired by a high-energy laser beam. It have been paid close attention by an increasing number of researchers in metal analysis because no sample preparation needed and no damage to the samples.[1-5]

As a multivariate analysis method derived from Principal Component Analysis (PCA), Partial Least Squares (PLS) takes into account both the independent and dependent variables when building models, therefore it can achieve better models and results and was used widely by researchers of LIBS recently.[6-11] Stipe *et al.*[8] compared three calibration methods to quantify Cr, Ni and Mn of 23 alloy steel samples. Analysis showed that the model built by PLS can achieve the best result while over-fitting is not occurring comparing to the traditional method using calibration curve. Cong *et al.*[11] quantified the composition of alloy steel samples using the calibration curve and PLS method. The comparative result showed the PLS method can obtain better result than the calibration curve method both for major elements and trace elements. The PLS method can get better results even when the calibration curve method cannot achieve quantitative results.

We built PLS models for spectra by LIBS reduced dimensions by three methods and compared the results. The three methods are selecting characteristic lines of elements, selecting intensive spectral partitions and the whole spectra. Then we analyzed how to reduce dimensions for LIBS spectra when building PLS models.

Experiments and Samples

Experimental Setups. The schematic experimental setups of LIBS are shown in Fig. 1. A Nd-YAG Laser(BigSky, America) was used in the experiment. Its maximum energy and frequency was 200mJ and 10Hz, respectively. The output wavelength of the laser is 1064nm. The energy employed in this experiment was 55mJ and the focus point of lens was 1mm depth under the sample surface. The laser frequency was set to 1 Hz. Spectra emitted from plasma were collected by an Echelle spectrometer

(LLA, Germany) with the gate delay 3µs and the integration time 5µs. The spectral wavelength ranges from 200nm to 780nm and the resolution of spectrum is about 0.02nm (FWHM). Every spectrum contains 57144 pixels denoting intensities. Optimized parameters of the laser and spectrometer in previous studies in our laboratory were directly used here.[1, 2]
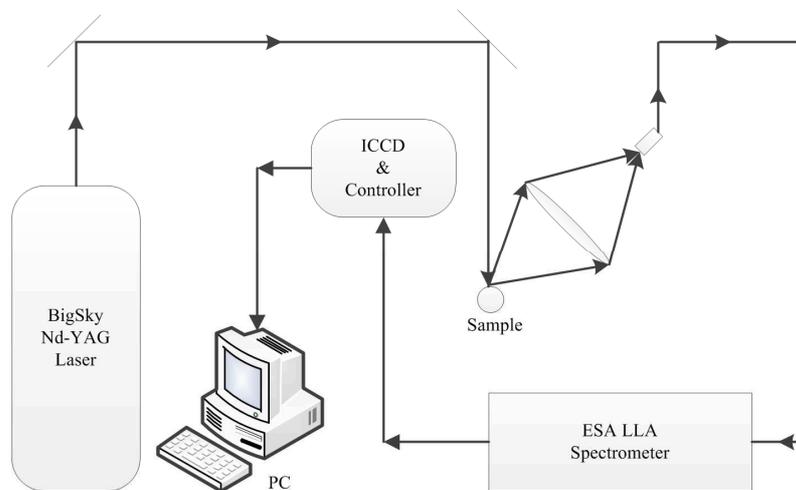

Figure 1. Schematic experimental setups of LIBS.

Samples. Twenty-seven steel samples belonging to four categories were investigated in the experiment. The first category was composed of 7 carbon steel samples. The second was 9 kinds of low alloy steels. The third was 5 high alloy steels. The last was composed of 6 stainless steel samples containing molybdenum, copper, niobium and nitrogen.

The concentrations of Mn in 27 steel samples analyzed in this paper were listed as Table 1.

Table 1. Concentrations of Mn in 27 steel samples (%)

Cat.	No.	C.	Cat.	No.	C.	Cat.	No.	C.	Cat.	No.	C.
1	1	0.712	2	8	0.15	3	17	0.13	4	22	2.53
	2	0.342		9	0.75		18	0.337		23	0.737
	3	1.02		10	2		19	0.495		24	0.867
	4	0.058		11	1.6		20	0.74		25	1.63
	5	1.27		12	1.29		21	0.983		26	0.127
	6	0.226		13	1.02					27	1.96
	7	0.01		14	0.51						
				15	0.31						
				16	0.1						

Spectra Collection. We excited thirty times for every sample and the first 10 excitations were used to clean the surface of the sample and discarded and then the next 20 excitations were recorded as 20 spectra to analyze the sample. So 20 spectra were collected for every sample and then the 20 spectra were averaged to one spectrum to analyze.

Table 2. Spectra in Calibration and Validation set

Category	Sample number	Calibration set	Validation set
1	1-7	1, 2, 5-7	3, 4
2	8-16	9-11, 13, 15, 16	8, 12, 14
3	17-21	17、19、21	18, 20
4	22-27	22, 24, 26, 27	23, 25

Data Sets: Two data set were constructed to build and validate the PLS model respectively. The first is the calibration set, composed of 2/3 of 27 samples, for building the PLS model and the other is

the validation set, composed of the rest 1/3 samples, for validating the quantitative ability of PLS model. Table 2 lists the sample number belonging to the calibration and validation set respectively.

Methods and Discussions

Criteria of Evaluation. The coefficient of determination (R^2) combined with Root Mean-Squared Errors (RMSE) were used to evaluate the PLS models. RMSE can be calculated as Eq. 1.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - T_i)^2}{n}} .$$ (1)

Where O_i and T_i are the ith predicted and measured value respectively and $\overline{T}$ are the averaged measured value.

Manually Selecting Characteristic Lines. The alloy steel samples mainly contain nine elements, i.e. Mn, Si, Ni, Cr, V, Mo, Ti, Cu and Fe. Eighteen characteristic lines were selected in all with two lines selected for every element. Table 3 listed the characteristic lines selected for all 9 elements.

Table 3. Characteristic lines of nine elements selected manually

Element	Mn	Si	Ni	Cr	V	Mo	Ti	Cu	Fe
Lines(nm)	279.481	288.158	299.259	265.858	275.339	256.705	257.102	279.179	261.382
	403.307	319.950	336.957	266.870	293.081	285.323	315.419	299.735	273.955

The intensities corresponding to these characteristic lines were trimmed from original spectra and a PLS model was built with these intensities. Mean-squared errors (MSE) combined with percentage of variance explained were used to choose the optimal number of principal components (PC) of PLS model. As Fig. 2(a) shows, six PC can explain more than 95% of variance and the MSE is approximately zero simultaneously, so we select the optimal number of PC as 6.

The quantitative result of PLS model was shown in Fig. 2(b) and Table 4. In Fig. 2(b), the blue circles represent the calibration set and the red ones represent the validation set. Obviously, the result is not very good.

Manually Selecting Intensive Spectral Partitions. In this section, an intensive spectral partition ranging from 256nm to 266nm which contains the characteristic lines of Mn was selected to build a PLS model to quantify the steel samples.
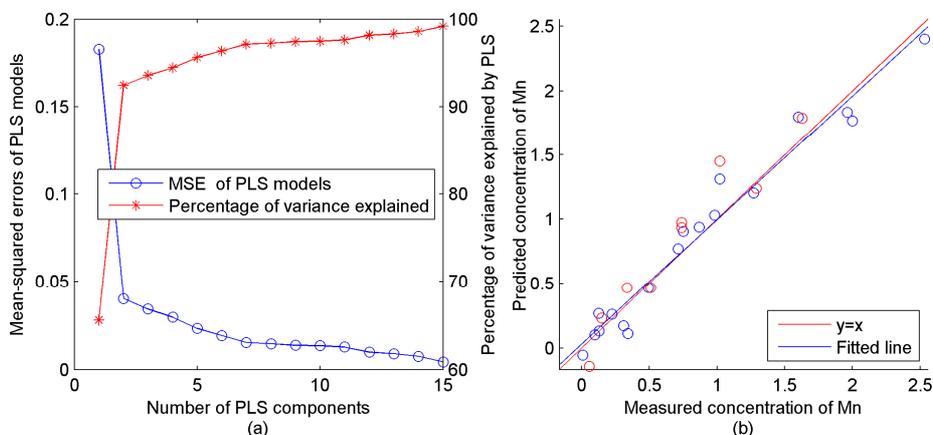

Figure 2. Results of PLS models using characteristic lines manually selected: (a) MSE and percentage of variance explained of PLS models corresponding to different PC; (b) Quantitative results of PLS models: blue circles for calibration set and red circles for validation set, respectively.

Figure 3(a) demonstrates MSE and percentage of variance explained of PLS and the optimal PC can be chosen based on these, so we construct the PLS model using 5 PC.

The quantitative result of the PLS model was shown in Fig 3(b) and Table 4. The result is very good. The reason is that the noise was eliminated and the effective information was reserved when selecting partitions from spectra, therefore the signal-to-noise ratio (SNR) was improved. So we can achieve an excellent result using these partitions.
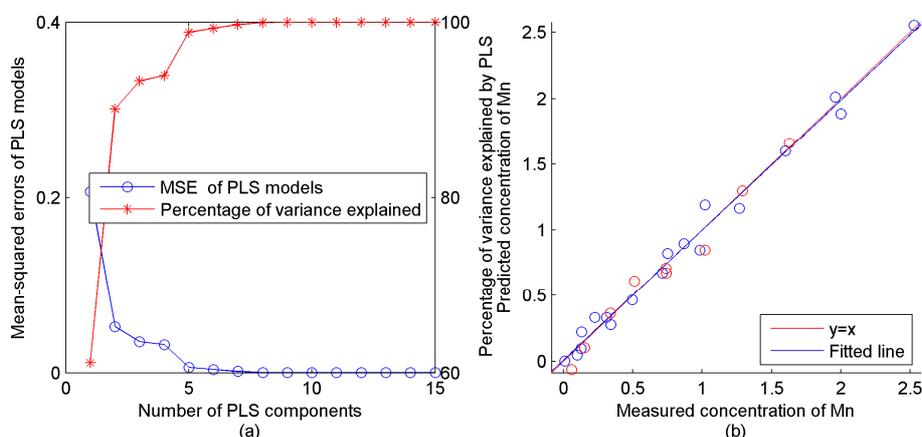
Figure 3. Results of PLS models using spectral partitions manually selected: (a) MSE and percentage of variance explained of PLS models corresponding to different PC; (b) Quantitative results of PLS models: blue circles for calibration set and red circles for validation set, respectively.

PLS of the Whole Spectra. The whole spectra were used to build the PLS model as a comparative solution.
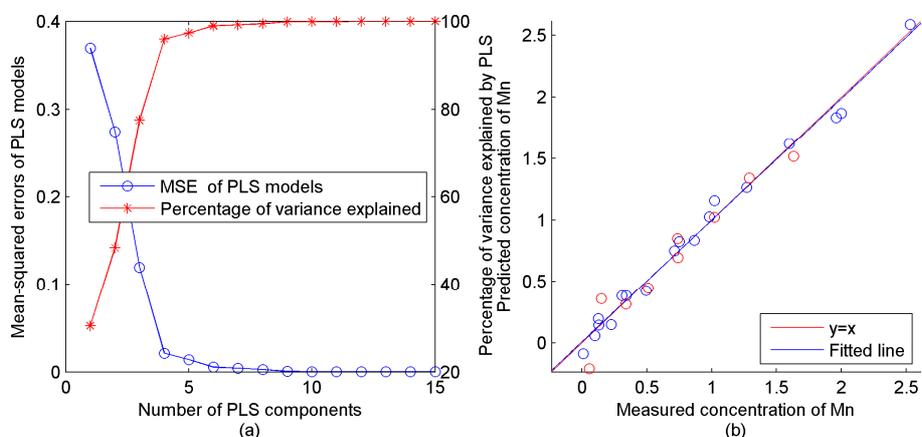
Figure 4. Results of PLS models using the whole spectra: (a) MSE and percentage of variance explained of PLS models corresponding to different PC; (b) Quantitative results of PLS models: blue circles for calibration set and red circles for validation set, respectively.

The MSE and percentage of variance explained of PLS model were shown in Fig. 4(a) and the optimal 6 PCs can be chosen according to those. Therefore a PLS model using 6 PCs was built to quantify the Mn concentration of steel samples.

The quantitative result was shown in Fig. 4(b) and Table 4. As that shows, the result is still very good for calibration set, but the result is a little worse than selecting spectral partition for validation set.

Comparative Analysis. Table 4 lists the quantitative results of PLS models for calibration set and validation set using the three different inputs.

As Table 4 shows, selecting characteristic lines manually has little significance and the results of PLS model based on them are not good. This may because the selected characteristic lines cannot include

enough informantion for quantitative analysis. Selecting appropriate lines manually is not an easy work. Selecting intensive spectral partitions can build the best PLS model using the least number of PCs. Although the fitting results for calibration set of this model are a little worse than that derived from the whole spectra, they are still comparative. In addition to that, the results for validation set of this model are the best in all three inputs. So selecting intensive spectral partitions is an outstanding way of dimension reduction for the original spectra with the complexity reduced and the generalization ability enhanced.

Table 4. PLS models based on three different inputs

Input	Number of PC	R^2		RMSE	
		Calibration	Validation	Calibration	Validation
Characteristic lines	6	0.9639	0.8317	0.1380	0.2030
Spectral partitions	5	0.9884	0.9711	0.0785	0.0841
PLS of the whole spectra	6	0.9896	0.9320	0.0742	0.1290

Conclusions

PLS models based on three spectral dimension reduction methods, selecting characteristic lines, selecting intensive spectral partitions and the whole spectra, were built to do the quantitative analysis of 27 steel samples by Laser-Induced Breakdown Spectroscopy. The results show that, the best solution to reduce dimensions is selecting intensive spectral partitions which can achieve the best result with the least complexity.

Acknowledgements

References

[1] Sun L, Yu H, Cong Z and Xin Y: Acta Optica Sinica Vol. 30 (2010), p. 2757.
[2] Sun L, Yu H, Xin Y and Cong Z: Spectroscopy and Spectral Analysis Vol. 30 (2010), p. 3186.
[3] Sun L, Yu H, Cong Z and Xin Y: Chinese Journal of Scientific Instrument Vol. 32 (2011), p. 2602.
[4] Zhang Xu YM-Y, Liu Mu-Hua: Acta Phys Sin Vol. 62 (2013), p. 44211.
[5] Wang Z, Yuan T-B, Hou Z-Y, Zhou W-D, Lu J-D, Ding H-B, et al.: Front Phys Vol. (2013), p. 1.
[6] Labbe N, Swamidoss IM, Andre N, Martin MZ, Young TM and Rials TG: Applied Optics Vol. 47 (2008), p. G158.
[7] Andrade JM, Cristoforetti G, Legnaioli S, Lorenzetti G, Palleschi V and Shaltout AA: Spectrochimica Acta Part B-Atomic Spectroscopy Vol. 65 (2010), p. 658.
[8] Stipe CB, Hensley BD, Boersema JL and Buckley SG: Appl Spectrosc Vol. 64 (2010), p. 154.
[9] Ayyalasomayajula KK, Dikshit V, Yueh FY, Singh JP and Smith LT: Anal Bioanal Chem Vol. 400 (2011), p. 3315.
[10] Dyar MD, Carmosino ML, Breves EA, Ozanne MV, Clegg SM and Wiens RC: Spectrochimica Acta Part B: Atomic Spectroscopy Vol. 70 (2012), p. 51.
[11] Cong ZB, Sun LX, Xin Y, Li Y, Qi LF and Yang ZJ: Spectroscopy and Spectral Analysis Vol. 34 (2014), p. 542.

Frontiers of Chemical Engineering, Metallurgical Engineering and Materials III
10.4028/www.scientific.net/AMR.997

Quantitative Analysis of Steels Using PLS with Three Data Reduction Methods Based on LIBS
10.4028/www.scientific.net/AMR.997.578

DOI References

[1] Sun L, Yu H, Cong Z and Xin Y: Acta Optica Sinica Vol. 30 (2010), p.2757.
http://dx.doi.org/10.3788/AOS20103009.2757

[7] Andrade JM, Cristoforetti G, Legnaioli S, Lorenzetti G, Palleschi V and Shaltout AA: Spectrochimica Acta Part B-Atomic Spectroscopy Vol. 65 (2010), p.658.
http://dx.doi.org/10.1016/j.sab.2010.04.008

[8] Stipe CB, Hensley BD, Boersema JL and Buckley SG: Appl Spectrosc Vol. 64 (2010), p.154.
http://dx.doi.org/10.1366/000370210790619500

[9] Ayyalasomayajula KK, Dikshit V, Yueh FY, Singh JP and Smith LT: Anal Bioanal Chem Vol. 400 (2011), p.3315.
http://dx.doi.org/10.1007/s00216-011-4852-3

[10] Dyar MD, Carmosino ML, Breves EA, Ozanne MV, Clegg SM and Wiens RC: Spectrochimica Acta Part B: Atomic Spectroscopy Vol. 70 (2012), p.51.
http://dx.doi.org/10.1016/j.sab.2012.04.011