Technical note

A partial least squares based spectrum normalization method for uncertainty reduction for laser-induced breakdown spectroscopy measurements

Xiongwei Li [a], Zhe Wang [a,*], Siu-Lung Lui [a], Yangting Fu [a], Zheng Li [a], Jianming Liu [b], Weidou Ni [a]

[a] State Key Lab of Power Systems, Department of Thermal Engineering, Tsinghua-BP Clean Energy Center, Tsinghua University, Beijing, China
[b] China Guodian Science and Technology Research Institute, Nanjing 100034, China

ARTICLE INFO

ABSTRACT

A bottleneck of the wide commercial application of laser-induced breakdown spectroscopy (LIBS) technology is its relatively high measurement uncertainty. A partial least squares (PLS) based normalization method was proposed to improve pulse-to-pulse measurement precision for LIBS based on our previous spectrum standardization method. The proposed model utilized multi-line spectral information of the measured element and characterized the signal fluctuations due to the variation of plasma characteristic parameters (plasma temperature, electron number density, and total number density) for signal uncertainty reduction. The model was validated by the application of copper concentration prediction in 29 brass alloy samples. The results demonstrated an improvement on both measurement precision and accuracy over the generally applied normalization as well as our previously proposed simplified spectrum standardization method. The average relative standard deviation (RSD), average of the standard error (error bar), the coefficient of determination (R^2), the root-mean-square error of prediction (RMSEP), and average value of the maximum relative error (MRE) were 1.80%, 0.23%, 0.992, 1.30%, and 5.23%, respectively, while those for the generally applied spectral area normalization were 3.72%, 0.71%, 0.973, 1.98%, and 14.92%, respectively.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) is an atomic emission spectroscopic technology that can identify the elemental composition of a gas, liquid, or solid sample [1]. It has become one of the most promising analytical techniques and been applied in archeology, military, industry, and space exploration since the past decade [2,3]. Its advantages over other conventional analytical methods include simple operation, no or minimal sample preparation, rapid and real-time measurement, simultaneous multi-element assay, and sample preservation [3,4]. Yet, the commercialization of LIBS developed slowly, mainly due to its unsatisfying measurement repeatability originating from the variations of experimental parameters (e.g. fluctuations in laser energy, laser-sample interaction, etc.) and the unpredictable chemical and physical matrix effects [5–7]. In addition to LIBS setup modification [8–11], various mathematical and statistical approaches have been applied to process the spectral data [12–15]. A common practice to compensate for the varying ablation energy or ablation mass is to normalize a spectrum with specific emission lines or spectral areas [13–15].

A more sophisticated tactic is to modify the measured signal with the fluctuation of the plasma properties [16–21]. In our previous work, a spectrum standardization method was proposed to improve measurement precision [18]. A "standard plasma state" was defined in the spectrum standardization method by characterizing the plasma with a set of constant parameters (n_{s0}, T_0 and n_{e0}), which were calculated by averaging the corresponding plasma parameters of all the measurements. By assuming the existence of the "standard plasma state", the recorded characteristic line intensity was converted to a standard state intensity with standard plasma temperature (T_0), electron number density (n_{e0}), and total number density of measured species (n_{s0}), and the spectrum standardization method was able to much reduce the measurement uncertainty as well as improve model accuracy. A simplified spectrum standardization method was thereafter introduced [19], and the simplified model further improved measurement precision and accuracy with much less calculation effort than the previous spectrum standardization method.

A drawback of the simplified spectrum standardization method is that it only applied the intensity of a single characteristic line for standard state conversion. Moreover, the signal uncertainty caused by the fluctuation of the plasma temperature is correlated to the ratio of just a single pair of spectral lines, while in real LIBS application, even an isolated line could exhibit unexpected measurement fluctuations, e.g. a noise spike. This can seriously affect the quality of the model if the chosen lines were interfered or affected by others. In practice, a spectrum recorded by a broadband spectrometer usually possesses tens to hundreds of emission lines from different elements. This is generally

---

* Corresponding author. Tel.: +86 10 62795739; fax: +86 10 62795736.
E-mail address: zhewang@tsinghua.edu.cn (Z. Wang).

sufficient to establish multiple pairs of line combinations in order to cross-check any plasma characteristics (e.g. T). In other words, there is potential to further improve measurement accuracy and precision by making use of more spectral line information.

Multivariate analysis (MVA) is an effective mathematical and statistical approach for LIBS data analysis, since it can utilize much quantitative information from the complex LIBS spectra. Partial least square (PLS) is such a MVA method and has shown great potential for LIBS quantitative measurement in recent years [22–24]. In this paper, a method which exploits the advantages of the spectrum standardization method and the PLS algorithm is proposed.

2. Theory

The proposed PLS based spectrum normalization model is developed from our previous spectrum standardization models [18,19], especially the simplified spectrum standardization model. In the next section, the baseline spectrum standardization model is reviewed, and its implementation with the PLS based spectrum normalization model is introduced as well.

2.1. Brief introduction of the simplified spectrum standardization model

In the simplified spectrum standardization model [19], the deviation of the measured line intensity from the standard state intensity is caused by the fluctuations of plasma parameters (T, n_e, and n_s) from their standard state. Using Taylor expansion, this can be expressed as,

$$I_{ij}(n_s, T, n_e) - I_{ij}(n_{s0}, T_0, n_{e0}) \approx k_1 dn_s + k_2 CdT + k_3 Cdn_e, \tag{1}$$

where n_{s0}, T_0, and n_{e0} are the magnitudes in the standard state; I_{ij} is the raw line intensity; $I_{ij}(n_{s0}, T_0, n_{e0})$ is the standard state line intensity; C is the concentration of the specific element; k_1, k_2, and k_3 are constants.

The deviation of the n_s, T, and n_e may be correlated to the measured spectral information. On the right hand side of Eq. (1), the fluctuation of total number density, dn_s, was assumed to be proportional to the fluctuation of the sum of the multiple emission line intensities of the measured element. According to the local thermodynamic equilibrium (LTE) assumption, the excited states in plasma follow the Boltzmann distribution. The second dT may be associated with the intensity ratio of a pair of lines based on the principle of Boltzmann distribution [25]. As the characteristic spectral line broadening is mainly caused by the Stark broadening for typical LIBS measurements, the full width of half maximum (FWHM) of the spectral line is assumed to be proportional only to electron number density. The third dn_e may be determined from the FWHM of the H_α spectral line through Stark broadening [25]. Assuming there exists a simple linear relationship between the concentration of an element and its standard state line intensity, the concentration may be expressed by Eq. (1) as:

$$C = a_1 I_{ij} + a_2 I_T + a_3 \left(\ln\left(\frac{I_2}{I_1}\right) - \left(\ln\left(\frac{I_2}{I_1}\right)\right)_0 \right) C + a_4 (\Delta\lambda_{stark} - (\Delta\lambda_{stark})_0) C + a_5. \tag{2}$$

In Eq. (2), I_T is the sum of the selected spectral line intensities of the measured element. a_1, a_2, a_3, a_4, and a_5 are constants calculated from regression process. Both $(\ln(I_2/I_1))_0$ and $(\Delta\lambda_{stark})_0$ can be calculated from the average of all measured spectra which can be applied to indicate their standard state values. Rearranging Eq. (2) will give the concentration as:

$$C = \frac{a_1 I_{ij} + a_2 I_T + a_5}{1 - a_3 \left(\ln\left(\frac{I_2}{I_1}\right) - \left(\ln\left(\frac{I_2}{I_1}\right)\right)_0 \right) - a_4 (\Delta\lambda_{stark} - (\Delta\lambda_{stark})_0)}. \tag{3}$$

The line intensity at the standard state is proportional to the measured elemental concentration:

$$I_{ij}(n_{s0}, T_0, n_{e0}) = k_1 C = k_1 \frac{a_1 I_{ij} + a_2 I_T + a_5}{1 - a_3 \left(\ln\left(\frac{I_2}{I_1}\right) - \left(\ln\left(\frac{I_2}{I_1}\right)\right)_0 \right) - a_4 (\Delta\lambda_{stark} - (\Delta\lambda_{stark})_0)}. \tag{4}$$

2.2. PLS based spectrum normalization model

In the simplified spectrum standardization model, only a pair of lines (e.g. I_2/I_1) and only a single line (e.g. I_{ij}) is used for concentration modeling [19]. By using information of multiple lines, the prediction performance may be improved, and the model will become much less sensitive to uncontrollable interference or fluctuations of the several chosen lines.

Base on this idea, terms that characterize the fluctuations of the plasma parameters on the right hand side of Eq. (2) are updated. The third term which modeled the temperature variation is now expressed by several pairs of lines as $\sum_{j=1}^{n}\left(c_j \left(\ln\left(\frac{I_{2j}}{I_{1j}}\right) - \ln\left(\frac{I_{2j}}{I_{1j}}\right)_0 \right) \right)$.

As a result, Eq. (2) becomes:

$$C = a_1 I_1 + b_1 I_T + \sum_{j=1}^{n}\left(c_{1j} \left(\ln\left(\frac{I_{2j}}{I_{1j}}\right) - \ln\left(\frac{I_{2j}}{I_{1j}}\right)_0 \right) C \right) + d_1 (\Delta\lambda_{stark} - (\Delta\lambda_{stark})_0) C + e_1, \tag{5.1}$$

where I_1 is a characteristic line intensity of the interested element. a_1, b_1, c_{1j}, d_1 and e_1 are the constants which can be acquired by a regression algorithm. Assuming that the number of the characteristic lines is m ($m \geq 2$), the concentration can also be expressed by the same method with employing other characteristic lines of the measured element:

$$C = a_2 I_2 + b_2 I_T + \sum_{j=1}^{n}\left(c_{2j} \left(\ln\left(\frac{I_{2j}}{I_{1j}}\right) - \ln\left(\frac{I_{2j}}{I_{1j}}\right)_0 \right) C \right) + d_2 (\Delta\lambda_{stark} - (\Delta\lambda_{stark})_0) C + e_2 \tag{5.2}$$

$$C = a_m I_1 + b_m I_T + \sum_{j=1}^{n}\left(c_{mj} \left(\ln\left(\frac{I_{2j}}{I_{1j}}\right) - \ln\left(\frac{I_{2j}}{I_{1j}}\right)_0 \right) C \right) + d_m (\Delta\lambda_{stark} - (\Delta\lambda_{stark})_0) C + e_m. \tag{5.m}$$

From Eqs. (5.1) to (5.m), the concentration can be expressed by a linear combination of the selected spectral line intensities:

$$C = \sum_{i=1}^{m}(a_{mi} I_i) + b I_T + \sum_{j=1}^{n}\left(c_j \left(\ln\left(\frac{I_{2j}}{I_{1j}}\right) - \ln\left(\frac{I_{2j}}{I_{1j}}\right)_0 \right) C \right) + d (\Delta\lambda_{stark} - (\Delta\lambda_{stark})_0) C + e, \tag{6}$$

where $a_{mi} = \frac{a_i}{m}, b = \frac{1}{m}\sum_{i=1}^{m} b_i, c_j = \frac{1}{m}\sum_{i=1}^{m} c_{ij}, d = \frac{1}{m}\sum_{i=1}^{m} d_i, e = \frac{1}{m}\sum_{i=1}^{m} e_i$. a_{mi}, b, c_j, d and e are the constants which can be acquired by the PLS regression algorithm.

Eq. (6) may also be re-arranged as,

$$C = \frac{\sum_{i=1}^{m}(a_{mi} I_i) + b I_T + e}{1 - \sum_{j=1}^{n}\left(c_j \ln\left(\frac{I_{2j}}{I_{1j}}\right) - \ln\left(\frac{I_{2j}}{I_{1j}}\right)_0 \right) - d (\Delta\lambda_{stark} - (\Delta\lambda_{stark})_0)} \tag{7}$$

to determine the exact elemental concentration.

It needs to be pointed out that part of the LIBS facilities would record a space and time integrated spectra such as the setup applied in the present work. That is, the plasma temperature, electron number density, and