Spectrochimica Acta Part B 99 (2014) 138–149

Contents lists available at ScienceDirect

Spectrochimica Acta Part B

journal homepage: www.elsevier.com/locate/sab

Automatic correction of continuum background in Laser-induced Breakdown Spectroscopy using a model-free algorithm

Pavel Yaroshchyk *, John E. Eberhardt

CSIRO Mineral Resources, Lucas Heights Science and Technology Centre, Locked Bag 2005, Kirrawee, NSW 2232, Australia

ARTICLE INFO

Article history:
Received 4 March 2014
Accepted 18 June 2014
Available online 5 July 2014

Keywords:
LIBS
Laser-induced Breakdown Spectroscopy
NMR
Background correction

ABSTRACT

We report on the first use of a model-free algorithm (designed to remove baseline in NMR spectra) for background correction of LIBS data. We also present our updated version of the algorithm that is potentially a better candidate for use in automated data processing routines in LIBS. This updated algorithm requires no assumptions to be made on the character of the background or type of spectra, uses no thresholds, and works equally well for high and low signal-to-noise LIBS spectra. The performance of the algorithm was evaluated using a range of randomly generated synthetic spectra and experimental LIBS spectra of varying complexity and noise characteristics. The analytical performance of the algorithm was evaluated by comparing the accuracy of LIBS measurements achieved when using untreated background, when subtracting the background using the original Friedrichs's algorithm and when subtracting the background using our updated algorithm.

1. Introduction

Laser-induced Breakdown Spectroscopy (LIBS) is an analytical technique that uses the atomic emission of laser-generated plasmas for the quantitative analysis of a wide range of materials in various industrial and scientific applications [1–3]. It is common for LIBS spectra to have a noticeable continuum background, which typically occurs due to Bremsstrahlung and recombination radiation [1–3], especially at the early stages of the plasma lifetime (typically up to a microsecond), but could also be contributed by the instrumental response. Since most quantitative LIBS measurements are conducted using time-gated detectors, the data acquisition start is normally delayed to reduce the continuum background intensity. The delay between the laser pulse and the acquisition start is typically set to obtain the maximum signal-to-noise ratio and depends on the sample type, experimental configuration and hardware used for the measurements. Unfortunately, even with delayed acquisition the continuum background intensity is often significant in LIBS spectra. Its accurate detection and removal often remains an issue. In particular, background correction is vital when multivariate algorithms such as Principal Component Analysis (PCA), Principal Component Regression (PCR), and Partial Least Squares Regression (PLS) are used for material identification and analysis. In contrast to standard linear calibration methods that correlate the elemental composition and the intensity of a single peak or a few peaks, multivariate algorithms generate a calibration model that correlates the sample's elemental composition or physical properties with a multitude of peaks spread over tens or hundreds of nanometres of wavelength [4,5].

Recently, two manuscripts have been published that specifically address the problem of continuum background correction in LIBS [6,7]. The method that was proposed by Gornushkin et al. [6] approximates the background by fitting high order polynomial functions through a number of automatically found minima. Their algorithm starts by finding the group of all minima of a spectrum, which is then split into a certain number of sub-groups, N. For each sub-group, a limited set of minima is deemed to contain the background pixels, the authors refer to them as "major minima", and those are the pixels with minimal intensities. For each sub-group, the means and standard deviations of the "major minima" are calculated and used to select another sub-group, "local minima", whose values fall within three standard deviations from the mean. A polynomial is drawn through the "major minima" of all sub-groups using a least-squares fit and a standard deviation, σ_N, is calculated with respect to minor minima. The polynomial power is either predetermined by the operator or is selected automatically, based on the best fit between the polynomial and the background. Once the algorithm iteratively finds the least σ_N that corresponds to the optimal number of groups, the polynomial function drawn through both major and minor minima represents the continuum background.

The approach of Gornushkin et al. [6] was updated by Sun and Yu [7]. Their algorithm also begins with finding all minima of a spectrum, but their use of a threshold based on a signal variance ratio leads to the exclusion of all "unsuitable" minima resulting in less iterations and reduced computational time. The correct determination of the threshold of the variance ratios of the two adjacent minima is crucial in the Sun

* Corresponding author. Tel.: +61 2 9710 6764; fax: +61 2 9710 6789.
E-mail address: pavel.yaroshchyk@csiro.au (P. Yaroshchyk).

http://dx.doi.org/10.1016/j.sab.2014.06.020
0584-8547/

and Yu procedure: if the threshold is overestimated, the resulting background intensity will also be overestimated, and vice versa. A linear interpolation between the widely spaced adjacent points is then made. Finally, one or several polynomial functions of optimal power are used to approximate the continuous background. The optimal number of polynomial functions used depends on the wavelength range of the spectrum and must be chosen to avoid under- or over-estimation of the local baseline.

The problem of spectral background correction is not unique to LIBS and is often a challenge in other analytical techniques including NMR [8,9], Raman[6,10] and ICP–AES[11]. An extensive review of baseline removal techniques as candidates for automated implementation and applicable to various spectroscopic techniques is given by Schulze et al. [12].

2. Model-free algorithm

The algorithm that we report here is a variation of the model-free algorithm reported by Friedrichs for the background correction of NMR data [8]. The basic idea of Friedrichs' method is that the baseline for a spectrum of noise alone can be defined by the criterion that the area above the baseline is equal to the area below it, i.e. the average value of the spectrum. This is not the case in the presence of NMR resonances, since the area under these peaks exceeds the area under noise peaks. In this case the number of local maxima and minima is used as a proxy measure for the area and the baseline is defined by the criterion that the number of local extrema above the line equals the number of local extrema below the line. Using Friedrichs's nomenclature, for each pixel i in the spectrum, the background structure is thus tracked by finding the median value of the extrema in a small window, W, centered about the pixel. An extremum is a point i such that the intensity $I(i)$ is either strictly greater or less than both its neighbouring intensities $I(i-1)$ and $I(i+1)$. The median of extrema, $Med(i)$, is then computed for i. The array of median values Med forms the preliminary distorted baseline. Finally, a Gaussian function G (zero-centred and normalised to 1) is convoluted with Med to smooth any sharp discontinuities and the final estimated baseline at i, $B(i)$, is given by:

$$B(i) = \sum_{j=i-W/2+1}^{j=i+W/2} Med(j) * G(i-j). \quad (1)$$

When applied to LIBS spectra, Friedrichs's algorithm works exceptionally well when those spectra contain continuum background and the spectral lines are not congested with relation to the window size. However, when the method is applied to a very congested data, which is typical of some LIBS spectra (especially those of iron), the computed background can look quite biassed. This bias occurs because many of the extrema in congested regions originate from peaks and not from background. The Friedrichs's algorithm performance can be fine tuned by expanding the size of the window W, but that often results in the baseline that does not plausibly follow the visually "true" LIBS background. Friedrichs describes a method of making the algorithm more robust by using elastic windows that change in width depending on the complexity of the region. In that case, when computing Med, the window size for each pixel is automatically chosen to contain X extrema where X is a maximum number of extrema for a fixed, user-specified window size.

A revision of the Friedrichs's algorithm for baseline recognition that involves the use of a background identification threshold was reported by Golotvin and Williams [9] and was applied by them to FT NMR spectra. Golotvin and Williams argued that since ultimately the baseline contains zero peaks, it is natural to employ for its recognition the same tool used for peak identification. To decide whether the intensity $I(i)$ belongs to a background, the window W centred around i by analogy with the Friedrichs's method. The maximum, I_i^{max}, and minimum, I_i^{min}, values are then computed and their difference is compared to the baseline threshold:

$$I_i^{max} - I_i^{min} = n\sigma_{noise} \quad (2)$$

where n is a definite multiplier typically in the range from 2 to 4, and σ_{noise} is defined by Golotvin and Williams as the minimum standard deviation in one of the 32 sub-regions of original spectrum. Further, to avoid existing spikes in the baseline, there is an extra requirement that the Eq. (2) is satisfied not only for i, but also for at least one of its immediate neighbours: $i-1$ or $i+1$. The smoothed baseline is calculated by averaging the neighbouring data points of the spectrum, an operational equivalent to convolution with a rectangular function:

$$B(i) = \sum_{j=i-\frac{W}{2}+1}^{j=i+\frac{W}{2}} I/(W+1) \quad (3)$$

The final baseline is constructed from fragments of the smoothed baseline (computed using Eq. (3)) interpolated by straight lines. Similarly to Friedrichs's algorithm, when applied to LIBS spectra, the algorithm does a good job in computing the baseline provided that the spectra are not congested and the background is not markedly discontinuous. We found that with increased spectral complexity the Golotvin and Williams method performs worse than the Friedrichs's method with increased spectral complexity. The problem arises from the fact that little to no background points are being selected in greatly congested windows, which results in rather large fragments of our LIBS spectra being omitted from the baseline estimation. In this case the linear interpolation between the existing points often results in the baseline not following the plausible visual background. Increasing W and varying the value of n can potentially improve the performance of the algorithm. However, the existence of a threshold that depends on the complexity of spectra makes the algorithm less suitable to use as a fully automated option in LIBS analysis.

3. The method

Our modification of Friedrichs's method is based on the idea that for a spectrum that has no analytical lines and consists of only a random (white) noise, the baseline computed by tracking median values differs from the baseline computed by tracking minimum values only by the median of the noise magnitude. Below we demonstrate that in the presence of and in relation to the intensity of the spectral lines, and, provided that the background profile is not characterised by the two limiting cases (see 4.1.3), the difference between the two baselines would still be proportional to the median of the noise.

Since most minima in the spectra are not affected by the presence or absence of analytical peaks, the preliminary distorted baseline is computed using a moving minimum for a window size W. Additionally, the smoothing function profile is selected to be the normalised boxcar, which, in our experience, produces a smoother baseline than the Gaussian smoothing function. The resulting baseline is given by the Eq. (4):

$$B(i) = \sum_{j=i-W/2+1}^{j=i+W/2} Min(j) * rect(i-j) \quad (4)$$

where $Min(j)$ is W-points moving minima, and $rect(i-j)$ is a rectangular (boxcar) function which is a constant $1/W$ within the bounds of the W-point domain and zero outside. Fig. 1 demonstrates a fragment of a synthetic spectrum as well as the corresponding moving minima (a minimum value in the $[i - W/2 + 1\ i + W/2]$ interval for each pixel i) and a final background function computed in accordance with (4). It is seen that the computed background tracks the true background albeit in the presence of a relatively small near-constant offset that is approximately equal to the noise magnitude. We note that the existence