Unsupervised fitting of emission lines generated from laser-induced breakdown spectroscopy

Ebo Ewusi-Annan, Noureddine Melikechi [*]

Department of Physics and Applied Physics, Kennedy College of Sciences, University of Massachusetts Lowell, Lowell, MA 01854, USA

ARTICLE INFO

ABSTRACT

This work addresses a bottleneck usually encountered in the processing of large amounts of LIBS data. It shows how the fitting of emission lines in a typical LIBS spectrum can be automated. This is done by exploiting the second derivative of the spectrum to obtain reasonable fitting parameter estimates that, together with a divide and conquer approach, can be used to fit large numbers of lines in a crowded spectrum without user intervention and executed within a reasonable timeframe. As a demonstration, we have successfully fit Voigt profiles to 625 lines in a Martian ChemCam spectrum.

1. Introduction

Laser-induced breakdown spectroscopy's (LIBS) appeal for qualitative and quantitative elemental analysis is due, in large part, to its ability to yield rapid results on the sample of interest. LIBS measurements can produce large amounts of spectra for almost any type of sample. The preprocessing of such data depends on the number and complexity of the spectra acquired. There have been efforts to make the post-experiment analysis of LIBS spectra easier. Examples of such efforts include automatic line identification [1,2], correlation analysis [3,4], peak deconvolution [5], uncertainty estimation [6], and automatic fitting [7,8] of emission lines. The most challenging to accomplish among these is automatic line fitting, especially that for large number of lines in spectra of dissimilar samples. To process such data, it is often necessary to detect, fit, and identify many emission peaks. The fitting process can be time-consuming particularly in situations where the spectra is crowded, and manual fitting is not a viable option.

When the number of spectra under consideration is in the hundreds or even thousands, fitting peaks of LIBS emission lines will be efficient if performed automatically. There are many cases that will benefit from automated peak fitting: analysis of the thousands of LIBS data that have been or will be received from the ChemCam [9–14], SuperCam [15,16] and MaRsCoDe [17] instruments currently on or en route to Mars; mining; environmental; and process monitoring [18,19]. In addition, analysis of a single spectrum can benefit from automatic fitting in cases where there are numerous emission lines present.

However, automating the process for multiple spectra is not straightforward because of numerous complexities that are typically present in the spectra such as line interferences resulting in poorly resolved individual lines, shift of peak positions and the poor signal to noise ratio of peaks in different spectra [20]. The last two issues make automation using a template unrealistic. Thus, any effort aimed at making the process automatic will have to take into consideration such spectral challenges to make the process truly autonomous with ideally no user intervention.

Researchers have reported on efforts to address automatic peak fitting in the literature. Rosas-Roman et al. [7] described a fast approach of fitting peaks in LIBS spectra using Lorentzian profiles. The authors fitted peaks with a polynomial representation of the Lorentzian profile which is of limited use for many spectra. This is because the use of Lorentzian profiles assumes that Stark broadening is the dominant process in the plasma. This makes the applicability of the method to spectra with peaks which have non-negligible Gaussian contributions inappropriate. An example, when the use of a Lorentzian peak is inappropriate is spectra observations or detections in late plasma persistence [21]. Another example is the situation where emission peaks have appreciable Doppler broadening [22], which is Gaussian in nature. It is therefore important to use in general a Voigt profile, which is a convolution of Lorentzian and Gaussian components, in fitting emission peaks.

Yang et al.[8] proposed a continuous-wavelet transform (CWT) method for fitting pseudo-Voigt profiles to emission peaks. They used CWT to detect peaks and obtain initial parameters of the pseudo-Voigt profiles used in the optimization objective function. The peaks were then fit using a trust-region algorithm. A drawback of using CWT as described by

---

* Corresponding author.
E-mail address: Noureddine_Melikechi@uml.edu (N. Melikechi).

the authors is that the scale parameter characterizing the CWT cannot be chosen *a priori* for all cases because of its dependence on the signal-to-noise ratio of the spectrum which can vary from one spectrum to another. In addition, these two approaches are not scalable to spectra containing large numbers of peaks. For this reason, their application to typical LIBS spectra will be, at best, semi-automatic, requiring expert intervention if reasonable execution time of the fitting process is to be achieved.

In this work we show that spectrum derivatives combined with a divide and conquer approach can be used to fit all atomic emission lines present in a typical LIBS spectrum within a reasonable timeframe. Our method does not require the introduction of data-sensitive parameters and scales well with the number of peaks in a spectrum. To illustrate the effectiveness of our method, we apply it to the LIBS spectrum containing over six hundred identified peaks. This work is an extension of the work we recently published [23], and in which we showed that the pre-processing of raw LIBS spectra can be accomplished with machine learning. Together this body of work should make processing of LIBS spectra easier and faster.

2. Methodology

The various steps used for this work are described in this Section. Essentially, our approach is based on, first, smoothening the input spectrum. This is followed by peak detection. Initial estimates of the fitting parameters for each peak detected are then obtained by fitting the second derivative of the spectrum. The initial estimates are then used to fit the peaks identified in a final fitting process employing a divide and conquer approach. In this work, initial fitting refers to fitting peaks using the second derivative of the non-smoothened spectrum to get initial parameter estimates. Final fitting, on the other hand, refers to fitting the non-differentiated, non-smoothened spectrum to obtain optimum Voigt profile parameters describing each identified peak.

2.1. Data smoothening

To aid in the peak detection process, we smoothen the spectrum using a locally weighted linear regression (LLR). This is a nonparametric technique used to approximate nonlinear functions with a piecewise linear model [24] by approximating each point with a linear function based on neighboring points. The linear fit at each point is determined by first multiplying a Gaussian kernel centered at the wavelength of interest, λ_i, with all other points to obtain weights for each point in the spectrum. This process provides a higher weighting to points located closer to the wavelength of interest than others. The weight vector w_i for each data point (λ_i, y_i) considering all other points λ is given as:

$$w_i = e^{-(\lambda-\lambda_i)^2/(2\tau^2)} \tag{1}$$

where τ is a parameter that regulates the importance of neighboring points in the weight calculation. Once weights for all points are determined, coefficients β_i for the model are determined using the normal equation:

$$\beta_i = (\Lambda^T W \Lambda)^{-1} \Lambda^T W y \tag{2}$$

In Eq. (2), Λ is a matrix of wavelength values, W is a matrix of the weight of all points (λ_i, y_i) and y is the vector of spectrum intensities. After the determination of the $\beta_i's$ for all points, a denoised version of the spectrum can be obtained by finding the linear approximation of each point. The level of denoising depends on the value of τ. The larger the τ, the higher the level of denoising. In this work we used $\tau = 0.03$. It should be noted that using locally weighted linear regression is a personal preference of the authors because of the simplicity of the approach. We do expect that using other smoothing techniques such as Savitzky-Golay [25] or wavelets [26] will equally work. The data smoothening was used solely for peak detection. During fitting, the raw

data were used.

2.2. Peak detection

To locate peak positions in the spectrum, we compare the intensity of each point with neighboring position intensities. If the intensity of the position of interest is higher than its immediate neighbors, it is identified as a peak. To minimize the influence of noise, the LLR smoothened spectrum is used for peak detection. Peak detection was performed using the SciPy python library [27].

2.3. Initial parameter estimates - initial fitting

Optimizing parameters during a fitting process is likely to converge faster to the true parameters if initial parameters used to start the process are close to the true parameters. To ensure efficiency, reasonable estimates of profile parameters that can be used to fit large numbers of peaks in a spectrum is essential. To estimate peak parameters, we exploit the spectrum's second derivative. Second derivative signal processing is widely used in other spectroscopy techniques such as Raman, near-infrared (NIR) and Fourier Transform Infrared (FTIR) for resolution enhancement [28], background correction [29] and peak detection [30]. However, a drawback of using the second derivative for signal processing is the worsening of the signal to noise ratio [31]. Regardless, this is of little concern to this work primarily because we seek to obtain reasonable estimates but not necessarily optimum parameters from its use. As such, possible reduction in accuracy of the parameter estimates due to low SNR will be corrected during the final fitting process. Nevertheless, a noisy spectrum can be denoised for the initial fitting step.

To estimate the Voigt profile parameters of the peaks, we fit each peak in the context of its immediate neighbors on both sides of the peak. The objective function O_n for fitting peak n in the spectrum is given by:

$$O_n = \left\| \frac{d^2 S(\lambda)_{exp}}{d\lambda^2} - \frac{d^2 \left[\sum_{i=1}^{N} A_i \cdot V_i \right]_n}{d\lambda^2} \right\|_2^2 \tag{3}$$

where $S(\lambda)_{exp}$ is the experimental spectrum; λ is the wavelength; A_i is the amplitude of the Voigt profile V_i; N is the total number of peaks in the fit. N is equal to 3 if the peak has two neighbors. Otherwise, $N = 2$ if only a single neighbor exists. In the absence of a neighbor, $N = 1$. The Voigt profile is defined as [32]:

$$V(\lambda; \lambda_{o,i}, \gamma_{g,i}, \gamma_{l,i}) = \frac{1}{\gamma_{g,i}} \sqrt{\frac{4 \ln 2}{\pi}} V_f(v, a) \tag{4}$$

where $\lambda_{o,\,i}$, $\gamma_{g,\,i}$, $\gamma_{l,\,i}$ are the peak wavelength, Gaussian full width at half maximum (FWHM), and Lorentzian FWHM, respectively. $V_f(v, a)$ is the Voigt integral defined as:

$$V_f(v, a) = \frac{a}{\pi} \int_{-\infty}^{+\infty} \frac{e^{-t^2}}{a^2 + (v - t)^2} dt \tag{5}$$

where v and a are defined as:

$$v = \frac{\lambda - \lambda_0}{\gamma_g} 2\sqrt{\ln 2}, \quad a = \frac{\gamma_l}{\gamma_g} \sqrt{\ln 2} \tag{6}$$

v is a Gaussian profile normalized wavelength variable; and a indicates the relative importance of the Gaussian and Lorentzian components. The Voigt integral is evaluated using the real part of the Faddeeva function [27]. We can fit each peak in the second derivative spectrum even in the presence of significant interferences from neighboring peaks in the raw spectrum because of the improved resolution [28] of the peaks in the second derivative space.

Ideally, the parameters obtained by fitting the second-derivative spectrum should suffice in fitting the peaks in the non-differentiated

spectrum. However, this is contingent on the spectrum background being smooth, slow varying and quadratic in nature. For experimental spectra, this is usually not the case: it has been shown that the background of LIBS spectra can be described with polynomials with orders ranging from 1 to 10 [33,34]. This means that for non-quadratic backgrounds, there will be higher order term contributions still present in the second derivative spectrum. The consequence of this on estimating the parameters are fit parameters that underestimate the true peaks in the spectrum. For this reason and the low SNR of the second derivative signal, obtained fit parameters are used as estimates for fitting the peaks in the non-differentiated spectrum. It should be noted that even though the experimental spectrum we use for this work is background corrected, there are spectral regions that show incomplete background removal and as such background contributions to the second derivative fitting is still important.

2.4. Final peak fitting

Once estimates of parameters for all peaks are obtained, each peak n is fitted using the objective function:

$$O_n = \mu \left\| S\left(\lambda\right)_{exp} - \sum_{i=1}^{N} A_i \cdot V \right\|_2^2 + (1-\mu)\left\| \frac{d^2 S\left(\lambda\right)_{exp}}{d\lambda^2} - \frac{d^2 \left[\sum_{i=1}^{N} A_i \cdot V\right]}{d\lambda^2} \right\|_2^2$$
$$+ \sigma \left\| S\left(\lambda_o\right)_{exp} - \sum_{i=1}^{N} A_i \cdot V(\lambda_o) \right\|_2^2 \qquad (7)$$

Eq. (7) has three sum-of-squared-error terms: the first term accounts for the difference between the raw and the fit spectrum; the second term represents the difference between the second derivatives of the raw and fit spectrum; and the last term accounts for differences between the raw and the fit spectrum at the central wavelength λ_o. μ and σ are weighting parameters whose values for this work were determined by hand tuning to a small wavelength range in the UV region. μ is used to regulate the importance of the raw spectrum and its derivative to the fitting process. σ is used to give more importance to peak positions relative to other wavelengths during fitting. μ and σ values used are 0.6 and 10, respectively. Additionally, all points in the first terms whose value exceeded 0.01 were penalized by multiplying the corresponding objective function value by 10. By adding the second derivative to the objective function, the increased resolutions of peaks in the second derivative space enables the optimization to successfully fit peaks which otherwise would have been difficult to tackle due to significant interferences from nearby lines in the raw spectrum.

The parameters used in Eq. (7) (μ and σ) were hand tuned by considering a small wavelength range (241–250 nm). The combination of μ and σ that resulted in the least residual values was used for the whole spectrum. For large number of spectra, using a technique such as Bayesian optimization to extract the relevant parameters may be more efficient. However, these parameters do not need to be changed for every spectrum and therefore only one-time tuning is necessary to determine the relative importance of each component in the objective function.

With a handful of peaks present in a spectrum, the simultaneous fitting of all peaks is manageable. However, as the number of peaks increases, simultaneous fitting becomes computationally expensive and impractical especially for routine applications. One is therefore constrained to fitting a few peaks at a time. It thus becomes imperative to select wavelength ranges to fit in succession if the peak count in the spectrum is large. Selection of the wavelength ranges is however not straightforward especially when peak identities and hence peak locations can vary significantly from one spectrum to another making it impossible to use a generic template when fitting a group of spectra. To address this problem, we propose a divide and conquer approach, to efficiently handle this challenge.

2.5. Divide and conquer

The essence of this approach is to divide the task of fitting of many peaks into smaller tasks by fitting a few peaks per task. This is performed by fitting one spectral peak at a time and by considering only the nearest neighbor peaks. Once a peak is fitted, its optimized fitting parameters are fixed when they become neighbors of subsequent peak fits. The advantages of such an approach are a reduction in computation time especially as the number of peaks increase without compromising accuracy, and the lack of user-specified wavelength range for fitting.

In Fig. 1 we show the benefits of this approach when applied to simulated spectra containing 7–106 Voigt profiles. We fit all peaks in each spectrum using two approaches; the divide and conquer (DC) approach, and the simultaneous fitting of all peaks. As can be seen from Fig. 1a, when the number of peaks is less than 40 there is no advantage in using DC. However, with increasing peak numbers, DC is computationally efficient compared to simultaneous fitting. The time taken to fit 100 peaks using DC is 5.9 min compared to 20.8 min using simultaneous fitting. Also, a plot of the root mean square error (rmse) as a function of number peaks in a spectrum, shown in Fig. 1b, is identical for both approaches. This means DC does not sacrifice accuracy for speed of execution. It should be noted that for experimental spectra, the number of peaks beyond which DC will be faster than simultaneous fitting will be less than the 40 observed in the simulation results primarily because of experimental factors such as noise and line asymmetry which were not considered in the simulated spectra. Fig. 1a also illustrates the time complexities characterizing the two approaches: simultaneous fitting scales quadratically with the number of peaks while DC scales linearly. The exponential increase in time for simultaneous fitting arises from the exponential increase in the number of peaks to be optimized. Fitting a single peak with a Voigt profile requires the optimization of four parameters while fitting multiple peaks requires the optimization of parameters 4 times the number of peaks. DC, however, has a limit on the number of peaks (maximum of 13 in this work) and hence a limit on the number of parameters it can fit at a time. Therefore, the time complexity of DC scales as the time required to fit 13 peaks multiplied by the number of 13-peak fits that will be carried out for the whole spectrum. For these reasons, the fitting time of the DC approach scales slower than that of simultaneous fitting.

The application of the DC procedure to the experimental spectrum is performed by looping through the array of identified peaks in a spectrum, fitting only one peak at a time. For initial fitting, we fit each peak in addition to its closest neighbors in the second derivative space. For the final fitting, we also fit in addition to the peak of interest up to six neighboring peaks on either side. This allows us to account for the influence neighboring peaks have on the peak of interest. In total, we fit up to thirteen peaks for each peak of interest. The neighboring peaks parameters are, however, estimates since their actual values are also influenced by their neighbors. Once fitted, peaks of interest have their parameter values fixed when they become neighbors in subsequent peak fits. The parameter optimization steps were done using the trust region reflective algorithm in the SciPy [27] optimization module within LMFIT [35].

In summary, the approach we propose for the unsupervised fitting of emission lines in a spectrum is carried out by:

- Denoising the raw spectrum using local linear regression.
- Detecting all peaks present in the denoised spectrum.
- Fitting each peak detected with Eq. (3) using the second derivative of the raw spectrum and the DC approach considering only nearest neighbor peaks.
- Finally, fitting each peak again in the raw, non-differentiated spectrum using Eq. (7) and the DC approach considering up to 6 neighbors on either side of the peak of interest.
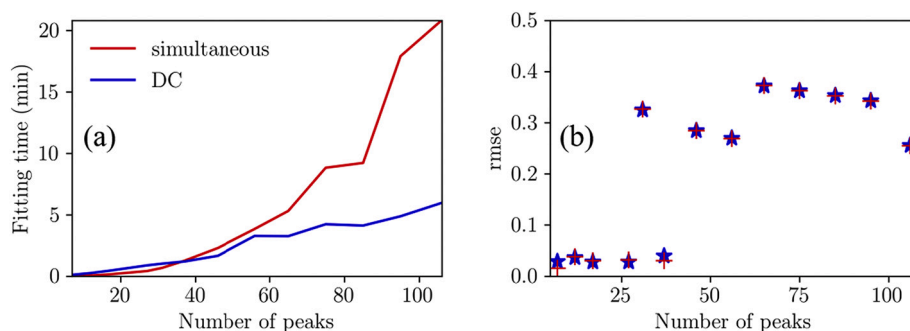
Fig. 1. Comparing the fitting time and root mean square error (rmse) for fitting <u>simulated</u> spectra. (a): Fitting time comparison between divide and conquer (DC) (bottom figure) and simultaneous fitting (top figure). (b): Comparison of the rmse of divide and conquer and simultaneous fitting. The star and plus symbols correspond to simultaneous and DC, respectively.

3. Data

To illustrate the approach presented in this work, we have selected a spectrum from the National Aeronautics and Space Administration (NASA) planetary data system [10]. This database contains the Chem-Cam spectra sent to Earth by the Curiosity rover. Details of the Chem-Cam experimental setup and procedure for preprocessing the spectra are described by Wien et al. [9,36]. In brief, the experiment is executed by using a Neodymium doped Potassium-Gadolinium Tungstate (Nd: KGW) laser operating at 1067 nm with a laser energy of 14 mJ/pulse and repetition frequency of 3 Hz to interrogate Martian targets at distances that can be up to 7 m away from the rover using a telescope. Emissions from the plasma formed are propagated back through the telescope and directed simultaneously into three crossed-design Czerny-Turner spectrometers spanning three wavelength ranges: 240.1–342.2 nm (ultraviolet (UV)), 382.1–469.3 nm (violet (VIO)) and 474.0–906.5 nm (visible and near infrared (VNIR)). Attached to each spectrometer is a charged coupled device camera used to record wavelength-dispersed spectra. We use the mean of the preprocessed spectra of 50 single shot interrogations of *Coronation*, the first target analyzed by ChemCam when it landed on Mars.
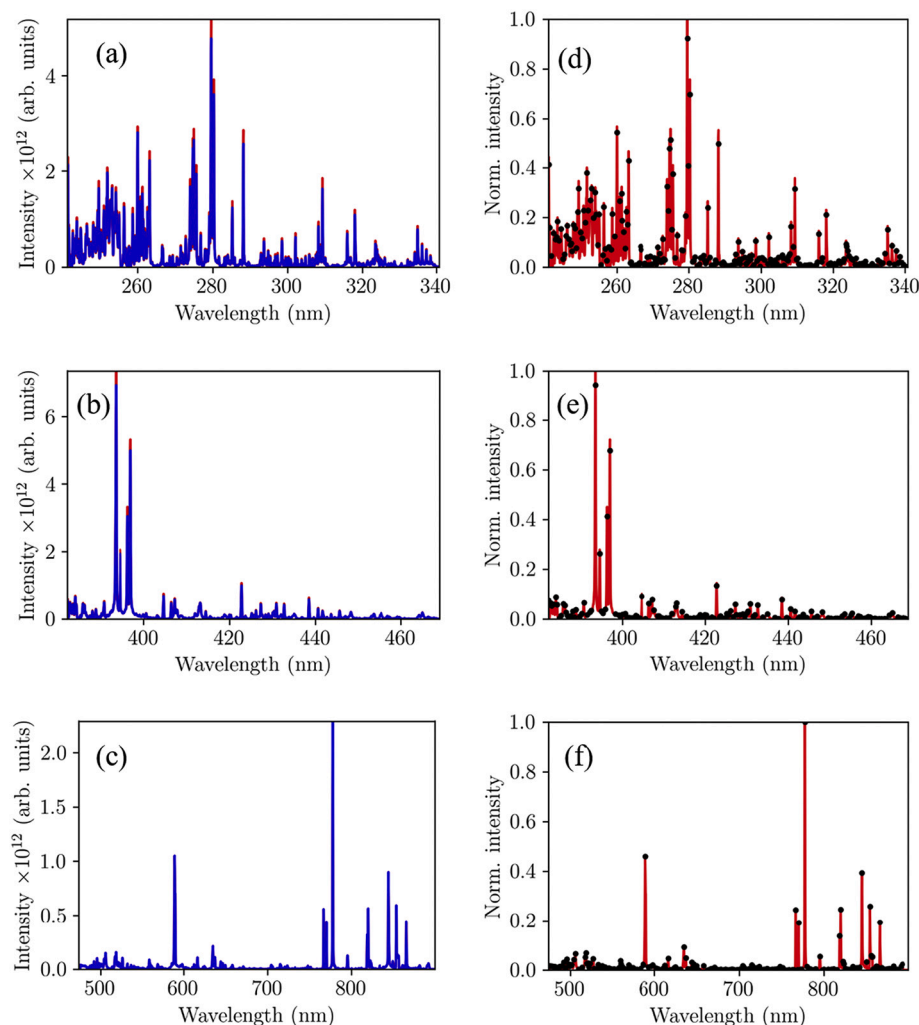

Fig. 2. Spectrum smoothening and peak detection in Coronation. Column 1 (a, b, c): Comparison of the raw spectrum (red line) with smoothened version (blue line) using local linear regression (LLR). Column 2 (d, e, f): Peak detection. The black dots indicate the positions identified as peaks. 247, 146 and 232 peaks were detected for the 241–343 nm (UV), 382–469 nm (VIO), and 474–900 nm (VNIR) regions, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
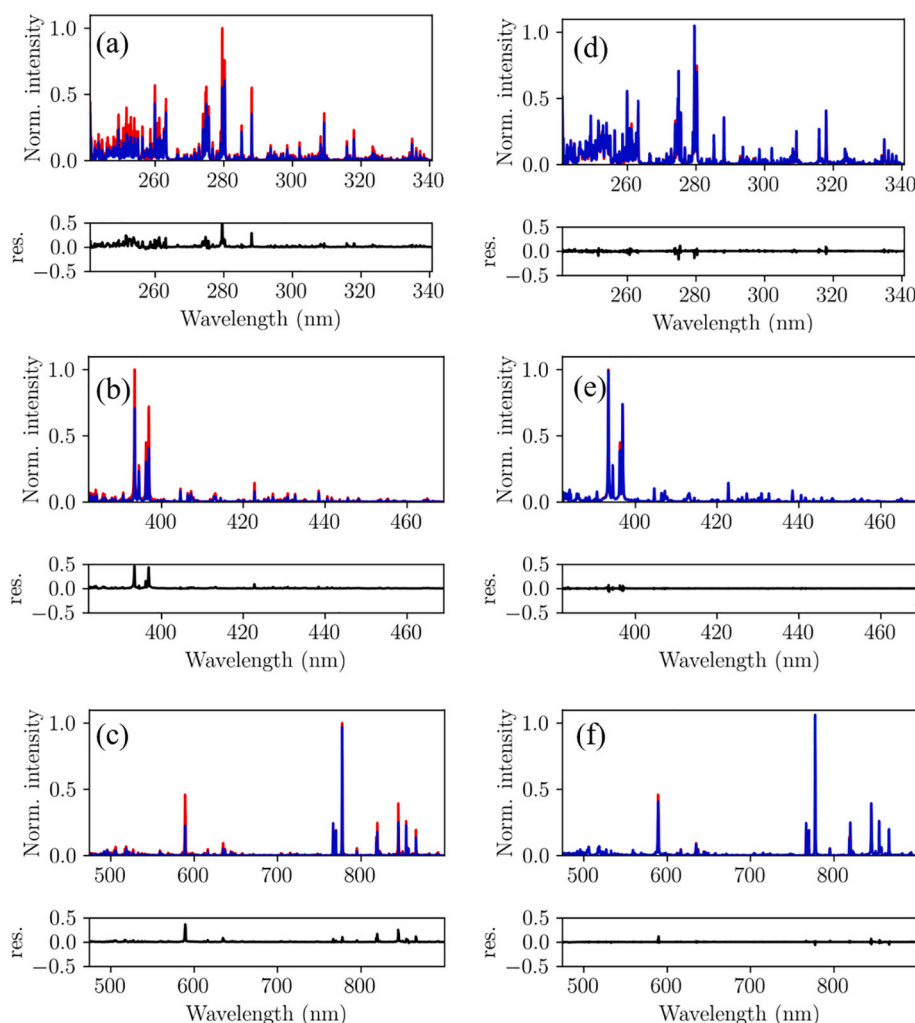
Fig. 3. Fit and raw spectrum comparison where fit spectrum was computed using parameters from second derivative (a, b, c) and final fittings (d, e, f). Column 1 (a, b, c): Comparison of the raw spectrum (red line) and sum of individual peak fits (blue line) obtained using parameters from the second derivative fitting. Column 2 (d, e, f): Comparison of the raw spectrum (red line) and sum of individual peak fits (blue line) obtained using parameters from final fitting. The subplot at the bottom of each figure is the residual (res.) which is the difference between the raw and fit spectrum. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Results and discussion

In this section, we present the results of smoothening, peak detection, initial parameter estimation and final fitting for the *Coronation* spectrum. Fig. 2 (a, b, c), shows a comparison of the LLR smoothened spectrum with its raw counterpart. The intensities of the smooth spectrum (blue line) are slightly lower than that of the raw spectrum (red line) across the UV and VIO wavelength ranges. For the VNIR range, the intensities of lines in the smoothened signal are the same as their raw spectrum analogs. The reason for the discrepancy in the smoothening effect among the three regions is due to differences in spectral resolution in the three regions. The UV and VIO regions both have 2400 groove lines per mm gratings whereas the VNIR region has 600 groove lines per mm [36]. This results in resolutions of 0.2 nm or less for the UV and VIO ranges and 0.6 nm for the VNIR region. The value we used for this work, $\tau = 0.03$, was not only adequate but also for other spectra we considered including those obtained with a different LIBS setup than ChemCam. A more robust way of ensuring that peak intensities are not underestimated in one region compared to another is to use a wavelength variable τ expressed in terms of spectral resolution. This is particularly important when a smooth spectrum is used in the final fitting. However, as we do not use the smoothened spectrum beyond the peak detection step, the slight reductions in peak intensities are inconsequential to the results.

Fig. 2 (d, e, f), shows the peaks identified in the three regions of the spectrum. In these spectral regions, 247, 146 and 232 peaks were identified for the UV, VIO, and VNIR regions, respectively. Some of the peaks have relatively weak intensities compared to others and in typical LIBS analysis these may be ignored. However, as a proof of the robustness of our approach in its ability to handle large numbers of peaks, we retained every peak detected. In applications where such weak intensity peaks are to be ignored adding a threshold based on a metric such as SNR is straightforward. We should also point out an issue that, although not relevant to our spectrum, is commonly seen in LIBS spectra: the presence of peak shoulders due to strong interference between two lines. Our peak detection procedure may not identify such peaks but there exist elaborate peak detection methods such as wavelets [37,38] that can be used to address this. Nevertheless, considering that most peaks in typical LIBS spectra are resolved to an extent that makes them easily detectable with the approach we use, the need for a complex peak detection procedure may rarely be needed.

In Fig. 3, we show the comparison between the fit and raw spectra. Fig. 3(a, b, c) shows a comparison of the raw (red lines) and fit (blue lines) spectra generated using optimized second derivative spectrum fitting parameters. The fit spectra here and in subsequent figures are the sum of individual fitted peaks. Fit lines in all three spectral regions are lower than the raw spectrum with residuals in some cases, such as Na 589 nm doublet, showing an under estimation of more than 50%. The reasons for this are asymmetric lines caused by combination of following unresolved lines, and increased ionic contribution in the Stark broadening [39] of the lines. Also, ineffective continuum removal or experimental noise can affect the derivative of the spectrum. As examples, the peaks around 280 and 589 are asymmetric and hence the inability of the second derivative fit to retrieve the optimum Voigt
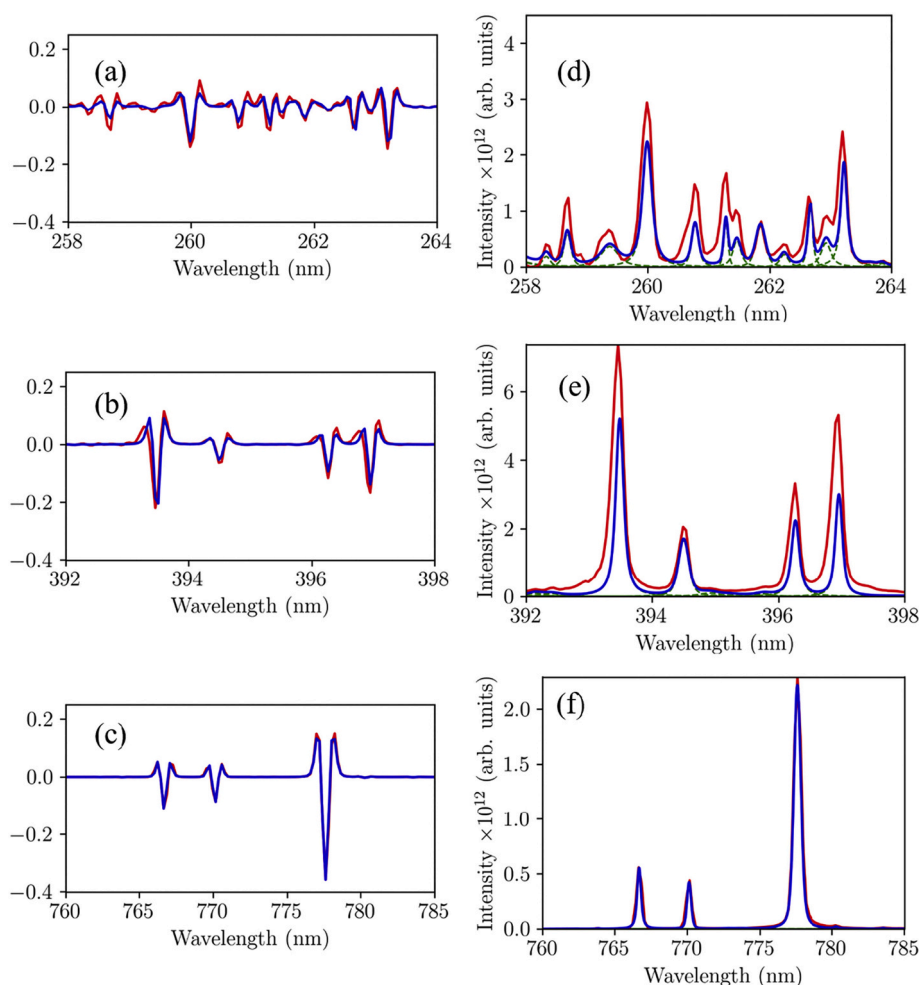
Fig. 4. Selected wavelength regions of Fig. 3 that shows raw (red line) and fit (blue line) second derivative spectra (a, b, c) and the raw spectra obtained using the optimum parameters of the second derivative fit (d, e, f). The dashed green lines show individual peak fits, and the blue lines show the sum of the green dashed lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

parameters. The asymmetry in the 589 (Na I doublet) arises due to interference between unresolved lines. However, as already stated earlier on, errors in parameters obtained from this fit are not of critical concern because they are intermediate parameters which are further optimized during the final fit.

To account for the ineffective continuum removal, we added an offset parameter to the Voigt profile. However, this did not improve the second derivative fits due to the nature in which the DC approach is executed. Essentially, an offset determined for the first fitted peak automatically becomes the offset of subsequent peaks because parameters of previously fitted peaks are fixed in the DC approach. We also point out that the introduction of a slope parameter does not provide any notable advantage: its improvement of the results pales in comparison to the increase in fitting time and does not compensate for effects resulting from the asymmetric peak shapes.

Fig. 3(d, e, f), shows a comparison between fit and raw spectrum where the final fitting process was done using the optimum parameters from second derivative spectrum fitting as starting parameters of the fit, a DC approach with up to 13 nearest neighbors, and the modified objective function in Eq. (7). Compared to its second derivative analogs, the discrepancies between the raw and fit spectrum have reduced significantly evidenced by the reduction in residuals for all lines. The residuals of lines in the UV region reduced by 0.1 on average for the poorly fit lines in the second derivative fitting. For the 280 nm region, despite the reduction, the discrepancy between fit and raw spectra is still present (residual of 0.2).

The asymmetry here is strong so much so that only the introduction of another peak in the fit or a modification of the Voigt profile to account for asymmetry will reduce the residuals further [40,41]. On the other hand, the residuals of the unresolved Na I doublet reduced significantly by 0.3.

Likewise, the residuals of the Ca II 393 and 397 nm lines reduced from 0.3 and 0.4 to 0.05 and 0.04, respectively. These residual reductions for all lines come from the additional components introduced in the objective function (Eq. (7)). The first component handles the line asymmetries that the second derivative objective function was unable to cope with. The second component ensures that peaks do not wander off from their central wavelength positions; this is important when fitting crowded lines that interfere greatly because, in the absence of the second derivative component, central peak positions are very difficult to optimize. The third component ensures peak amplitudes are comparable to peak intensities at the line center.

In Fig. 4, we show small sections of the three wavelength ranges: 258–264 nm for the UV range containing iron lines; 392–398 nm for the VIO range, which contains aluminum neutral and calcium ionic lines; and 760–785 nm in the VISIR region, which contains neutral potassium and oxygen lines. In addition to the fit and raw spectrum, we show the individual line fits (green dashed lines). Each line was generated using the optimum parameters from the final fit. In Fig. 4 (a, b, c), we show results of the fit to the second derivative spectrum.

The lines in the UV (Fig. 4a) were not fit well as those in the VIS (Fig. 4b) and VISIR (Fig. 4c) regions. This is due to interferences among
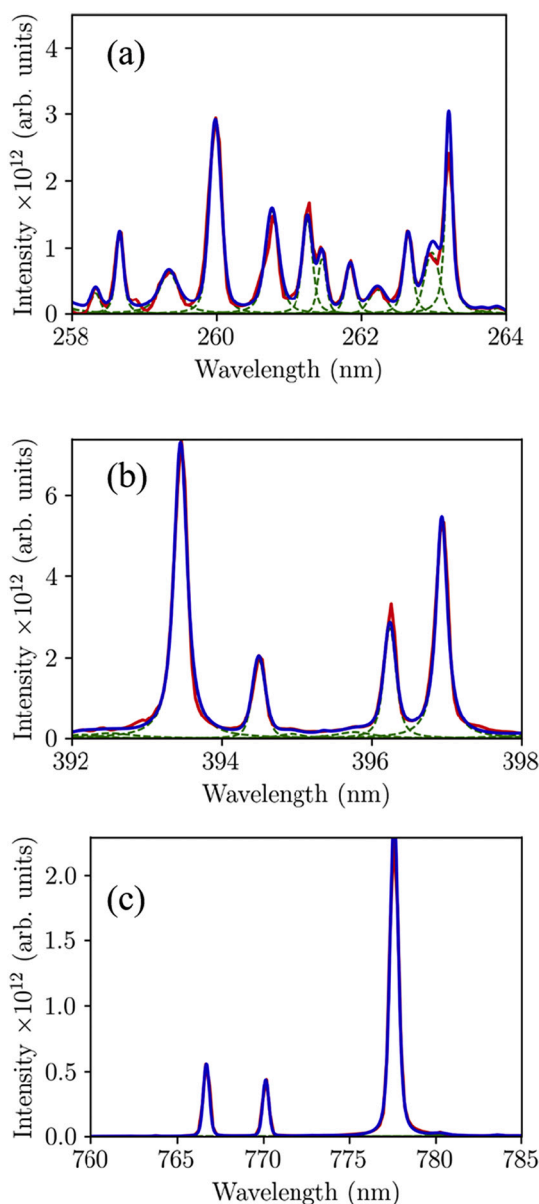
Fig. 5. Selected wavelength regions of Fig. 3 comparing raw and fit spectra from the final fitting of Coronation using Eq. (7), a DC approach with up to 13 nearest neighbor peaks, and the optimum parameters from the second derivative fit. This Figure shows blue, green and red curves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the lines, and lines characterized by asymmetric profiles as can be seen in Fig. 4d. In Fig. 4b and c, there are small differences between the raw and fit derivative spectra. A key feature of the fits to the lines at 260 nm, 393 nm and 397 nm are the presence of asymmetry and our model's inability to fit them; this is indicated by the line wing derivative spectrum, immediate to the central peak position, having different magnitude intensities. The consequence of this is a poor fit of the peaks in the raw spectrum (Fig. 4e and f) because we use a symmetric profile in the model. On the contrary, due to the symmetric nature of the potassium and oxygen lines, the residuals for these lines are close to zero, for both derivative raw spectra, respectively.

In Fig. 5, we show the results of the final fitting step using the optimum parameters from the second derivative fitting process as initial estimates, a DC approach that considers up to 13 nearest neighbors, and Eq. (7). There is marked improvement in the fits compared to Fig. 4(d, e,

f). The reasons for this improvement are the same as discussed earlier on for Fig. 3.

For the 625 peaks identified, the total computational time used for all processes was 25.4 min on a single processor core of an Intel Xeon desktop with a processor speed of 3.60 GHz and a RAM of 64 GB (~130 MB used during the fitting process). With code refactoring and parallelization, we believe the analysis time can even be reduced further. Also, under certain conditions, the Gaussian FWHM can be fixed in the Voigt profile thereby reducing the convergence time of the optimization algorithm. Similarly, in situations where Lorentzian components dominate over their Gaussian counterparts, the lines can be fitted using a Lorentzian profile.

5. Conclusions

We have presented a general approach that enables the automatic identification and fitting of large numbers of emission lines within laser-induced breakdown spectra using Voigt profiles. We have shown how the Voigt parameters can be initialized by fitting identified peaks using the second derivative of the spectrum. This step provides reasonable estimates for the Voigt profile. Using these initialized parameters as starting parameters, we employ a divide and conquer approach together with an objective function that considers the raw spectrum and its second derivative to successfully obtain optimum parameters of the Voigt profiles that faithfully describe all peaks identified in the spectrum. We have demonstrated the success of this approach on a Martian LIBS spectrum containing 625 identified lines. All peaks in this spectrum were successfully fit in 25.4 min with fit residuals that are globally negligeable. With the necessary code refactoring we believe analysis time can be further reduced. With further investigations, this work, combined with our previous work [22] on automatic preprocessing of LIBS spectra using partial least square regression and feedforward artificial neural network can contribute to significant improvements in the analysis of large numbers and line-dense LIBS spectra that may be acquired in many applications. Finally, we note that in this work we did not observe any self-reversed line or molecular band in the spectra we considered. Fitting of such profiles requires first identifying their presence and then using appropriate fitting models. Our plan for future work does include investigating such situations.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] I. Rosas-Román, M.A. Meneses-Nava, O. Barbosa-García, J.L. Maldonado, Semi-automatic elemental identification of laser-induced breakdown spectra using wavelength similarity coefficient, Appl. Spectrosc. 71 (2017) 627–633.

[2] T.A. Labutin, S.M. Zaytsev, A.M. Popov, Automatic identification of emission lines in laser-induced plasma by correlation of model and experimental spectra, Anal. Chem. 85 (2013) 1985–1990.

[3] I.B. Gornushkin, M. Mueller, U. Panne, J.D. Winefordner, Insights into linear and rank correlation for material identification in laser-induced breakdown spectroscopy and other spectral techniques, Appl. Spectrosc. 62 (2008) 542–553.

[4] A. Ferrero, et al., Libraries for spectrum identification: method of normalized coordinates versus linear correlation, Spectrochim. Acta Part B At. Spectrosc. 63 (2008) 383–388.

[5] K. Liu, et al., Interference correction for laser-induced breakdown spectroscopy using a deconvolution algorithm, J. Anal. At. Spectrom. 35 (2020) 762–766.

[6] D.M. Surmick, H. Boukari, J. Woodward, A.C. Stowe, N. Melikechi, Uncertainty of integrated intensity following line profile fitting of multiline spectra, Appl. Spectrosc. 72 (2018) 787–792.

[7] I. Rosas-Román, M.A. Meneses-Nava, O. Barbosa-García, J.L. Maldonado, Simultaneous height adjust fitting: an alternative automated fitting procedure for

laser-induced plasma spectra composed by multiple Lorentzian profiles, Spectrochim. Acta Part B At. Spectrosc. 134 (2017) 1–5.

[8] W. Yang, B. Li, J. Zhou, Y. Han, Q. Wang, Continuous-wavelet-transform-based automatic curve fitting method for laser-induced breakdown spectroscopy, Appl. Opt. 57 (2018) 7526–7532.

[9] R.C. Wiens, et al., Pre-flight calibration and initial data processing for the ChemCam laser-induced breakdown spectroscopy instrument on the Mars Science Laboratory rover, Spectrochim. Acta Part B At. Spectrosc. 82 (2013) 1–27.

[10] S. Maurice, et al., The ChemCam instrument suite on the Mars Science Laboratory (MSL) rover: science objectives and mast unit description, Space Sci. Rev. 170 (2012) 95–166.

[11] A. Mezzacappa, et al., Application of distance correction to ChemCam laser-induced breakdown spectroscopy measurements, Spectrochim. Acta Part B At. Spectrosc. 120 (2016) 19–29.

[12] N. Melikechi, et al., Correcting for variable laser-target distances of laser-induced breakdown spectroscopy measurements with ChemCam using emission lines of Martian dust spectra, Spectrochim. Acta Part B At. Spectrosc. 96 (2014) 51–60.

[13] W. Rapin, et al., Quantification of water content by laser induced breakdown spectroscopy on Mars, Spectrochim. Acta Part B At. Spectrosc. 130 (2017) 82–100.

[14] N.L. Lanza, et al., Understanding the signature of rock coatings in laser-induced breakdown spectroscopy data, Icarus 249 (2015) 62–73.

[15] R.C. Wiens, S. Maurice, F.R. Perez, The SuperCam remote sensing instrument suite for the mars 2020 rover: a preview, Spectrosc. (Santa Monica) 32 (50–55) (2017).

[16] T. Nelson, et al., The SuperCam instrument for the Mars 2020 rover, in: 2020 IEEE Aerospace Conference, 2020, pp. 1–12, https://doi.org/10.1109/AERO47225.2020.9172661.

[17] L.-N. Li, X.-F. Liu, W.-M. Xu, J.-Y. Wang, R. Shu, A laser-induced breakdown spectroscopy multi-component quantitative analytical method based on a deep convolutional neural network, Spectrochim. Acta Part B At. Spectrosc. 169 (2020), 105850.

[18] U. Panne, in: P. Hering, J.P. Lay, S. Stry (Eds.), Laser Induced Breakdown Spectroscopy (LIBS) in Environmental and Process Analysis BT - Laser in Environmental and Life Sciences: Modern Analytical Methods, Springer Berlin Heidelberg, 2004, pp. 99–123, https://doi.org/10.1007/978-3-662-08255-3_6.

[19] S.G. Buckley, H.A. Johnsen, K.R. Hencken, D.W. Hahn, Implementation of laser-induced breakdown spectroscopy as a continuous emissions monitor for toxic metals, Waste Manag. 20 (2000) 455–462.

[20] W. Hübert, G. Ankerhold, Elemental misinterpretation in automated analysis of LIBS spectra, Anal. Bioanal. Chem. 400 (2011) 3273–3278.

[21] D.W. Hahn, N. Omenetto, Laser-induced breakdown spectroscopy (LIBS), part I: review of basic diagnostics and plasma–particle interactions: still-challenging issues within the analytical plasma community, Appl. Spectrosc. 64 (2010) 335A–366A.

[22] E. Ewusi-Annan, Laser-induced Breakdown Spectroscopy: Diagnostic and Analytical Approaches Using Atomic and Molecular Emission Spectroscopy to Characterize the Quantitative Aspects of the Technique, University of Florida, 2015.

[23] E. Ewusi-Annan, D.M. Delapp, R.C. Wiens, N. Melikechi, Automatic preprocessing of laser-induced breakdown spectra using partial least squares regression and feed-forward artificial neural network: applications to earth and Mars data, Spectrochim. Acta Part B At. Spectrosc. 171 (2020), 105930.

[24] S. Schaal, C.G. Atkeson, S. Vijayakumar, Scalable techniques from nonparametric statistics for real time robot learning, Appl. Intell. 17 (2002) 49–60.

[25] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem. 36 (1964) 1627–1639.

[26] B. Zhang, L. Sun, H. Yu, Y. Xin, Z. Cong, Wavelet denoising method for laser-induced breakdown spectroscopy, J. Anal. At. Spectrom. 28 (2013) 1884–1893.

[27] P. Virtanen, et al., SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods 17 (2020) 261–272.

[28] G.L. Verdine, K. Nakanishi, Use of differential second-derivative UV and FTIR spectroscopy in structural studies of multichromophoric compounds, J. Am. Chem. Soc. 107 (1985) 6118–6120.

[29] Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC Trends Anal. Chem. 28 (2009) 1201–1222.

[30] B.J.G. de Aragão, Y. Messaddeq, Peak separation by derivative spectroscopy applied to ftir analysis of hydrolized silica, J. Braz. Chem. Soc. 19 (2008) 1582–1594.

[31] T.C. O'Haver, et al., Derivative spectroscopy and its applications in analysis, Anal. Proc. (1982), https://doi.org/10.1039/AP9821900022.

[32] H.-J. Kunze, Introduction to Plamsa Spectroscopy, Springer-Verlag, 2009.

[33] L. Sun, H. Yu, Automatic estimation of varying continuum background emission in laser-induced breakdown spectroscopy, Spectrochim. Acta Part B At. Spectrosc. 64 (2009) 278–287.

[34] I.B. Gornushkin, P.E. Eagan, A.B. Novikov, B.W. Smith, J.D. Winefordner, Automatic correction of continuum background in laser-induced breakdown and Raman spectrometry, Appl. Spectrosc. 57 (2003) 197–207.

[35] M. Newville, et al., lmfit/lmfit-py 1.0.0, 2019, https://doi.org/10.5281/zenodo.3588521.

[36] R.C. Wiens, et al., The ChemCam instrument suite on the Mars Science Laboratory (MSL) rover: body unit and combined system tests, Space Sci. Rev. 170 (2012) 167–227.

[37] J.M. Gregoire, D. Dale, R.B. van Dover, A wavelet transform algorithm for peak detection and application to powder x-ray diffraction data, Rev. Sci. Instrum. 82 (15105) (2011).

[38] G. Yang, J. Dai, X. Liu, M. Chen, X. Wu, Spectral feature extraction based on continuous wavelet transform and image segmentation for peak detection, Anal. Methods 12 (2020) 169–178.

[39] N. Konjević, M. Ivković, S. Jovićević, Spectroscopic diagnostics of laser-induced plasmas, Spectrochim. Acta Part B At. Spectrosc. 65 (2010) 593–602.

[40] M. Schmid, H.-P. Steinrück, J.M. Gottfried, A new asymmetric pseudo-Voigt function for more efficient fitting of XPS lines, Surf. Interface Anal. 46 (2014) 505–511.

[41] A.L. Stancik, E.B. Brauns, A simple asymmetric lineshape for fitting infrared absorption spectra, Vib. Spectrosc. 47 (2008) 66–69.