



HAL
open science

Guideline for increasing the analysis quality in laser-induced breakdown spectroscopy

D. Syvilay, J. Guezenoc, B. Bousquet

► **To cite this version:**

D. Syvilay, J. Guezenoc, B. Bousquet. Guideline for increasing the analysis quality in laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2019, 161, pp.105696 -. 10.1016/j.sab.2019.105696 . hal-03488783

HAL Id: hal-03488783

<https://hal.science/hal-03488783>

Submitted on 20 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Guideline for increasing the analysis quality in Laser-Induced Breakdown Spectroscopy

D. Syvilay¹, J. Guezenoc¹, and B. Bousquet^{1,*}

¹CELIA, UMR CNRS 5107, Université de Bordeaux, 351 Cours de la Libération, France

Corresponding author*: bruno.bousquet@u-bordeaux.fr

Abstract

Despite well-established advantages and a growing number of applications, laser-induced breakdown spectroscopy (LIBS) still fails to be recognized as a robust analytical technique, especially because many factors varying from an experimental setup to the other are preventing from any inter-comparison. The present paper is a guideline for increasing the quality of the LIBS analyses through a series of good practices, assessments and reporting along five key-steps being: i) monitoring of the LIBS signal, ii) optimization of the measurement conditions, iii) data filtering, iv) sorting, and v) quantification. This guideline offers an opportunity to increase the quality of the LIBS analysis and represents a significant step towards a standardization process. In addition, it allows comparing intra as well as inter-laboratory LIBS results.

Keywords

LIBS; good practices; assessment; quality; quantitative; sorting; guideline

Introduction

Since 2000, Laser-induced breakdown spectroscopy (LIBS) has experienced continuously growing development and has been applied to many different application fields. Several commercial LIBS instruments are now available, either for laboratory use, including imaging, or on-field analyses based on handheld or stand-off equipment. LIBS provides fast multi-elemental analysis based on relatively

simple sample preparation compared to other techniques, and even without any sample preparation in few specific cases, making it a green technique, meaning nonchemical.

However, despite its advantages, LIBS is still judged to be only semi-quantitative since the weakness of its analytical performance is considered as the Achilles' heel of the technique [1] and this point is often mentioned in scientific articles [1-5]. Consequently, LIBS is facing a kind of limit for growing and acceptability of its use in many application fields. This statement is surprising when considering the large number of studies that successfully demonstrated the quantification capabilities of LIBS [4, 6]. In addition, LIBS has also been exploited for sorting and sourcing purposes [1, 6] but, in this case, too, its veritable capability still needs to be demonstrated.

But the real drawback of LIBS is the lack of a standardization process [2, 4]. Indeed, an impressive number of varying experimental parameters related to the laser source (e.g. wavelength, pulse duration, energy [7]), the spectrometer (e.g. spectral range, spectral resolution) and the detector (e.g. sensitivity, gate, delay) [2], as well as a large number of experimental conditions (e.g. number of pre-shots, shots, accumulation, gas flow, focusing conditions) are influencing the LIBS signal. In addition, Motto-Ros *et al.* [2] have demonstrated that data processing was playing a vital role in the performance of the quantification models and that a given series of raw data could drive to significantly different results. Consequently, the variability often observed in raw LIBS signals can be efficiently reduced via a specific work on the control of the experimental parameters and conditions, while the variability on the analytical results of the quantitative models can be efficiently reduced via a specific work on data processing.

Previous works have already introduced the need of adopting good practices in LIBS and selecting adequate figures of merit to assess the performance of the models. Among them, Castle *et al.* [8] proposed to optimize the measurement precision; El Haddad *et al.* [6] emphasized the importance of splitting the dataset into separate training and validation subsets for classification and quantification; and, Motto-Ros *et al.* [2] proposed to apply to several emission lines a series of different sequences of

data processing in order to determine the best string “emission line-processing steps-regression model” *a posteriori* and not *a priori* as usually adopted. The existence of these articles demonstrates that the LIBS scientific community has already started moving towards a quality enhancement of the LIBS technique. However, each previous study only partially addressed the question of quality improvement and, to the best of our knowledge, this is the first time that all the key steps of the LIBS analysis are considered altogether.

In the present paper, we propose a guideline to assist the LIBS analyst from A to Z. The objective is to adopt good practices at each step of the analysis to reach the most valuable and reliable results. Particularly, the added-value comes from the objective comparison between the analytical performance of the different models. This guideline is organized in five points, namely: i) monitoring of the LIBS signal, ii) optimization of the measurement conditions, iii) data filtering, iv) sorting, and v) quantification. Based on this rigorous methodology, we claim that it will become possible to compare the LIBS results obtained by different analysts and then pave the way towards inter-laboratory comparison.

1. *Monitoring the LIBS signal*

Considering spectroscopic techniques in general, the very first step towards quality consists in monitoring the signal with the objective to reduce its fluctuations. In LIBS, if either the laser energy or the lens-to-sample distance is changing, then the measured spectra may exhibit significant variations in intensity. Moreover, the wavelength calibration of the spectrometer can drift and then, LIBS spectra may exhibit a wavelength shift that can thus dramatically affect their interpretation. Nevertheless, the wavelength drifts can be corrected afterwards during data processing and don't need accurate real-time monitoring. The present discussion is thus only focused on the monitoring of intensity fluctuations, which are considered to be quite difficult to correct afterwards. For that, we propose to build and then routinely run a *Control Chart* [9, 10].

Considering a sufficiently large number of LIBS measurements on a reference sample, and corresponding to common and limited experimental conditions, one can reasonably consider that the signal of interest is described by a normal distribution. In this case, 99.7% of the values are expected to be inside the so-called *3-sigma* interval centered on the mean value of the distribution, with *sigma* being the standard deviation of the dataset. The main parameter of the control chart is thus the *3-sigma* limit. Indeed, if the signal related to the control of a LIBS instrument falls outside the *3-sigma* limit, it means the experimental conditions have changed according to unanticipated cause. Two types of alarms can be deduced from the *3-sigma* limit of the control chart: the analyst has to i) investigate the reason that caused the drift, and ii) be aware that a previously built calibration model cannot be applied to interpret any experimental data recorded in this new regime of operation. Finally, it is common to also consider the *2-sigma* limit as an early warning. Thus, any signal falling outside the *2-sigma* limit should drive the analyst to control the experimental conditions at shorter intervals of time since the risk of falling outside the *3-sigma* limit is increased in this case.

1.1. Building a control chart

The number of emission lines present in a LIBS spectrum being sometimes very large, the analyst has to determine first what should be considered as the signal to be monitored. Actually, the analyst may decide to monitor the intensity of one preferred emission line, but also to consider a few lines or even the whole spectrum by computing the sum of all intensity values being related to peaks or not. We recommend this last option to get an overview of the fluctuations since some emission lines can be more sensitive to others to the changes in the experimental conditions. Moreover, considering simultaneously a series of data is expected to be less sensitive to potential drifts in wavelength calibration of the LIBS instrument. Hereafter, the monitored LIBS signal is thus the sum of the raw intensity values extracted from the whole spectrum. It is worth noticing that raw data must be preferred for the monitoring since corrected data may partially hide the effects of the actual experimental variations.

In spectroscopy in general, the construction of a control chart requires the use of a reference sample, which has to be homogeneous and stable in time. A plane metallic sample (e.g. steel), even not certified, is perfect for LIBS since it offers a good repeatability of laser ablation. Then, the experimental protocol, namely the number of laser pre-shots and shots, and sites/craters at the sample surface should be adjusted by the analyst in order to reduce the uncertainties and thus to obtain reliable data.

A control chart dedicated to LIBS should take into account a series of measurements of the reference sample recorded at different times, in order to encompass various experimental conditions, but of course by keeping strictly unchanged all the experimental settings. For instance, if the analyst considers that all the diversity of the normal conditions can be observed during a full week of measurements, at different times every day, then a control chart based on this one-week series of data is well adapted. Then, the *2-sigma* and *3-sigma* limits can be calculated in order to further control the experimental conditions of the LIBS measurements. Figure 1 displays an example of control chart derived from a series of LIBS measurements achieved with a handheld LIBS instrument along 35 days, on a reference steel sample. In this case, the experimental protocol was: 3 pre-shots and 7 shots per location, and 5 locations at the sample surface, giving 35 spectra. Then, the average spectrum over these 35 spectra was considered for the control chart. More precisely, at each point of the control chart the values of the sum of the raw intensity values were reported over the full spectral range. The average value (dotted blue line), as well as the 2-sigma (green) and 3-sigma (red) limits of the control chart are calculated from the 35 points displayed in blue in Figure 1.

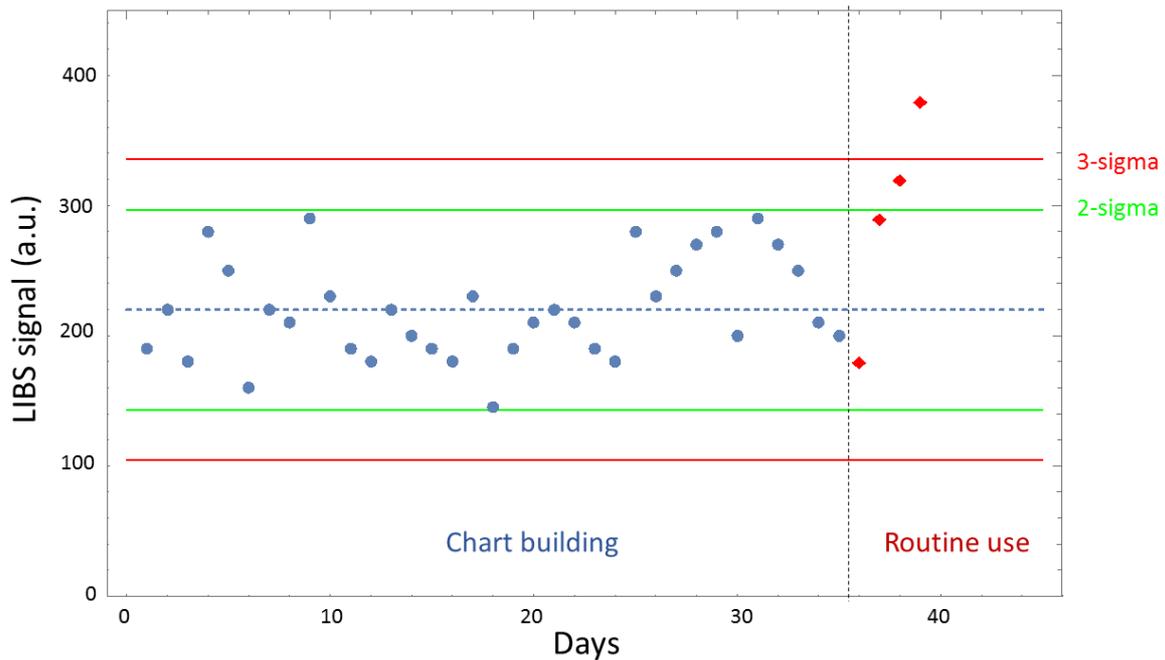


Figure 1: Control chart built from the sum of the intensity values (blue) over the full spectral range, recorded by a handheld LIBS instrument on a reference steel sample. The average value is represented by the dotted blue line, and the 2-sigma and 3-sigma limits are displayed in green and red, respectively. The points displayed in red correspond to routine use of the control chart

1.2. Routine use of a control chart

Once the control chart has been built (from the 35 points displayed in blue in Fig.1), the LIBS analyst can routinely use the chart in order to check the variability of the LIBS signal and by consequence the stability of the experimental conditions. In Figure 1, only the points displayed in blue were used to build the control chart and the other ones, displayed in red, correspond to further routine use of the control chart. The two points (Day 36 and Day 37) located within the 2-sigma limit, correspond to experimental conditions considered as normal. At the opposite, the two last points (Day 38 and Day 39) reveal a significant change in the experimental conditions and the control chart provides an immediate warning about this change to the analyst. More precisely, the point related to Day 38 is located within the 2-sigma and 3-sigma limits, and indicates a detectable change in the experimental conditions that can still be accepted by the analyst, who may continue to run new measurements while

looking for the cause of this change. But the last point (Day 39), outside the 3-sigma limit, illustrates the case of experimental conditions that have dramatically changed. It is worth noticing that it is strictly prohibited to interpret data recorded under the abnormal experimental conditions related to this last point from a calibration model built in the normal experimental conditions related to the blue points. The analyst thus has to identify the origin of the change and act on the experimental setup to move back to the normal experimental conditions. If this return can't be achieved because, for instance, of an irreversible drift of the laser energy linked to ageing, then the analyst must build a new control chart allowing to take into account these new experimental conditions that should be further considered as the normal conditions. And of course, she/he will have to build new models for either quantification or sorting.

In conclusion, the LIBS analysts who are using a control chart are able to early detect an abnormal change in the experimental conditions, or verify, at the opposite, that the setup is operating within the control limits, namely under normal conditions. The consequence is an improvement of the reliability of the technique. Moreover, we want to emphasize that verifying the status of the experimental LIBS conditions via a control chart is mandatory prior to applying any supervised approach for either quantification or sorting purpose. Indeed, LIBS analysts often report on how they build a quantitative or a sorting model, but the routine use of these models, meaning their long-time use, is rarely discussed. Thanks to such kind of control chart, any calibration model is strictly and exclusively applied to data related to normal conditions of analysis.

2. Optimization of the measurement conditions

The quality of the LIBS analyses is first based on the quality of the spectra, and consequently on the choice of the measurement conditions. It is very important to emphasize that the optimization of the measurement conditions has to be made on the basis of LIBS spectra recorded from a sample that is considered as representative of the samples to be further analyzed. For instance, if the LIBS analysis is dedicated to soil, glass or metal, then the reference sample must be also a soil, a glass or a metal,

respectively. Indeed, from one kind of sample to the other, the optimal conditions might be significantly different. The reference sample used for optimizing the experimental conditions must be homogeneous and stable in time, but it doesn't need to be a certified standard since the optimization process is based only on a comparison between a series of LIBS signals recorded under various experimental conditions.

To increase the analytical performance of LIBS, the first objective of the analyst is to maximize the background-corrected signal in order to reduce prediction errors and detection limits. Among the experimental conditions to be optimized, the first ones are linked to instrument parameters such as laser energy, lens-to-sample distance, light collection and injection towards the spectrometer, delay and gate of the detector, etc. It is worth noticing that all the adjustable parameters/conditions are not independent and that optimizing a LIBS measurement is necessarily based either on an iterative process, or on the use of an experimental design [11-13]. In the present section, we will discuss about the optimization related to i) instrument parameters and ii) sampling.

2.1. Instrument parameters

To determine the best instrument parameters for a given type of sample [14], we propose to follow a simple step-by-step approach: i) select the parameter to be optimized; ii) record a series of spectra from a reference sample for different values of the selected parameter (e.g. laser energy, delay); iii) determine the value to be set in order to obtain the best signal, namely the highest background corrected signal for a given spectral line.

To illustrate this point, Figure 2 displays, for different values of time delay varying over a range of 10 microseconds by 1-microsecond steps, the sum of intensity values (after baseline correction) over the full spectral range. It is worth noticing that all the emission lines present in a LIBS spectrum are not varying proportionally when an experimental parameter is changed. Thus, instead of monitoring only one single emission line, we recommend to take into account the full spectrum.

In this example, the points corresponding to the delays 1, 2 and 3 μs clearly exhibit higher signals than the other ones and should thus be preferred. However, it is worth noticing that the plot displayed in Figure 2 corresponds to a given value of laser energy and that the decision about the best delay may change for another value of laser energy, driving to consider optimization through an iterative process.

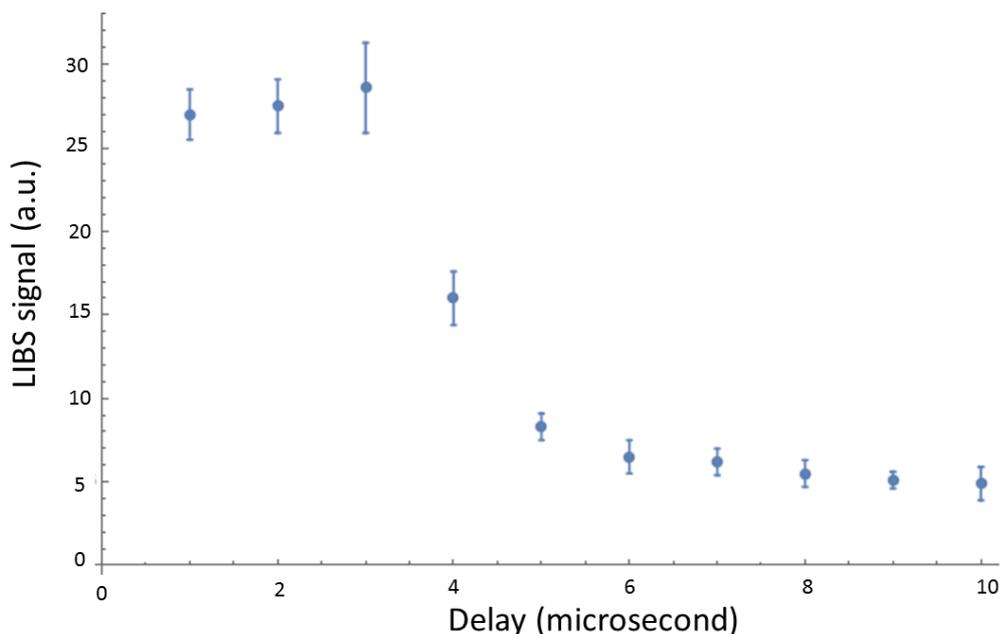


Figure 2: Sum of intensity values (after baseline correction) over the full spectral range of LIBS spectra related to a reference sample as a function of the time delay between the laser pulse and the acquisition start. Error bars correspond to the RSD values calculated over 3 repetitions.

At this stage, it is worth noticing that the optimization of the experimental conditions proposed here might be complicated by the presence of saturated features in the LIBS spectra. To illustrate this point, Figure 3 displays the top view of the raw LIBS spectra (horizontal axis) related to the series of time delay values (vertical axis) examined in Fig. 2. This type of display is similar to the one commonly chosen to show the emission spectra at different excitation wavelengths (excitation-emission matrix).

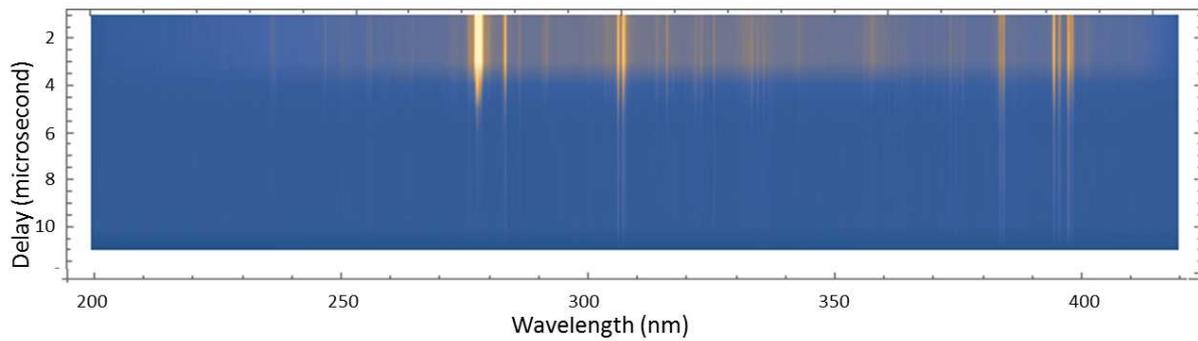


Fig 3: Top view of the average LIBS spectra over 3 repetitions (horizontal axis) related to the 10 values of time delay (vertical axis) presented in Figure 2. Features displayed in white show saturation.

One can observe that, for the delays of 1 to 3 μs , which were considered as the best ones regarding Fig. 2, the LIBS spectra exhibit a saturated feature corresponding to the emission lines of magnesium at 279.553 and 280.270 nm, shown by the white pattern in Fig.3. Based on this new insight, it appears that the choice of the best time delay must include the analysis of saturation, according the following two cases:

- i) If the saturated features are of major interest for interpreting the LIBS spectra, the analyst would prefer to select the value providing the strongest unsaturated signal. From Figure 3, the value of 4 μs should thus be chosen, since no more saturation is observed.
- ii) If the saturated features are without interest for interpreting the LIBS spectra (for instance due to the presence of a well identified major element), they could be excluded from further data processing without loss of information. In this case, the analyst would prefer to select the value providing the strongest but saturated signal. In the example displayed in Figures 2 and 3, one of the values between 1 and 3 μs would have been chosen. This scenario can be particularly advantageous when the LIBS analysis is focused on emission lines characterized by low intensity levels, while they are not affected by the vicinity of saturated features.

To summarize, the LIBS analyst must determine, at the very first step of the optimization process, if the presence of saturated features can be accepted or not. Once this choice is established, the best experimental conditions can be determined by varying the adjustable factors one by one and then iteratively. In the case described in this section, the delay between the laser pulse and the opening time slot of the detector has been optimized in order to increase the background-corrected. But, if most of the analysts choose to maximize the baseline corrected signal, it is well-established in LIBS that the shot-to-shot or site-to-site fluctuations can be significantly reduced when the laser beam is focused not exactly at the sample surface but embedded inside the sample [15-18]. In this case, the signal is not maximized but the standard deviation from a given series of spectra can be drastically reduced. It reveals that the optimization of the lens-to-sample distance should be optimized to reduce the standard deviation, prior to optimizing the sampling conditions.

2.2. Sampling

As it is the case for all spectroscopic techniques, the quality of a LIBS analysis may strongly depend on the sampling. Indeed, the type of sampling may drastically modify the statistical figures related to the fluctuating LIBS signal, namely average signal, noise, and standard deviation. Let's point out that LIBS instruments with a high control level allow to obtain very low levels of fluctuation. However, even for such systems, offering well-controlled measurement conditions, the ablation rate is expected to be different from one sample to another, driving to look for the best number of laser shots to be used in each case. Similarly, the sample surface might be altered by pollutants, dust or oxidation and it might be strategic to exclude the first laser shot(s), referred as pre-shots, from further data processing. Finally, considering that laser ablation and plasma formation might be quite hard to perfectly reproduce, the LIBS analysis should be based on a series of measurements at different locations at the sample surface. Homogeneous samples are expected to generate very small values of standard deviation from point to point, but in the case of heterogeneous samples, the question of sampling is

of major importance in order to get a good representativeness of the bulk sample through the LIBS signal averaged from a series of measurements.

Optimizing the sampling consists in determining the best number of laser shots at the same location/point/crater of the sample surface [15] and the best number of locations at the sample surface [19]. Of course, both the numbers of laser shots and locations have to be minimized in order to save time. Finally, sampling also consists in determining if laser pre-shots are required to remove dust, pollutants or an eventual oxidation layer.

Number of pre-shots

For many applications, the LIBS signal related to the first laser shot(s) is often different to the ones obtained from an ablation induced deeper in the bulk material. Indeed, the sample surface can be affected by dust, pollutants from various origins, and also oxidation. All of this can drive to LIBS signal significantly different for the first laser shot(s) compared with the following ones, driving to consider it/them as pre-shot(s). For quality purpose, instead of taking the decision about the number of pre-shots from a single location at the sample surface, we recommend to run a series of identical measurements at different locations, assuming the sample homogeneous. Thus, for each shot number, both the average value and standard deviation can be calculated for the peak intensity of the selected spectral line, over the different locations at the sample surface. To illustrate this point, in Figure 4, the background corrected signal of the Mg I emission line at 279.581 nm (vertical axis) recorded from a plane metallic sample of aluminum is displayed as a function of the number of laser shots. More precisely, the average value as well as the standard deviation (error bars), both derived from 27 sites at the sample surface, are displayed.

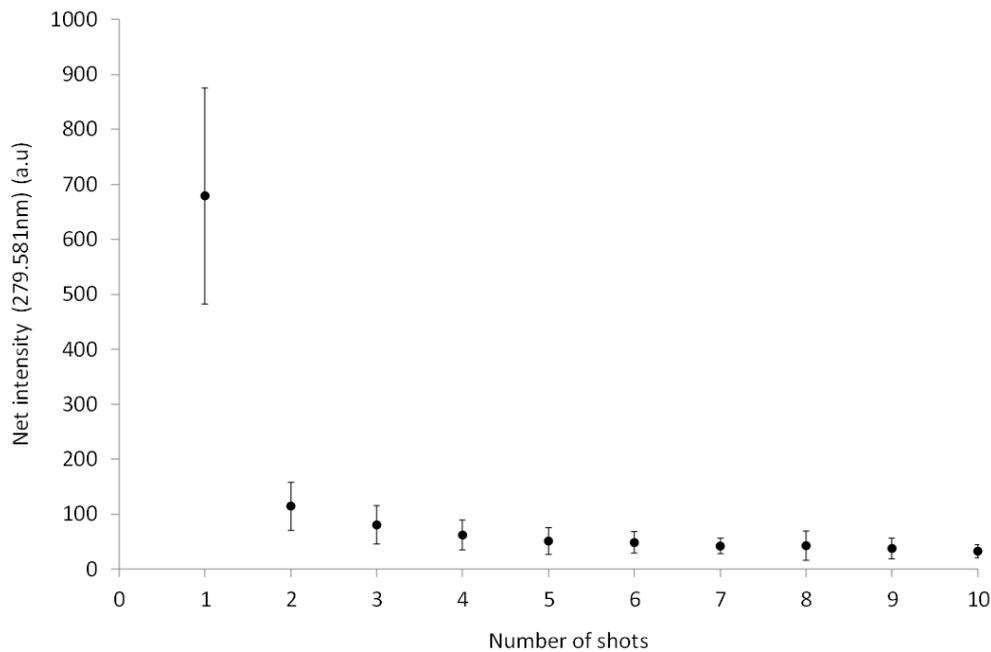


Figure 4: Background corrected intensity of the Mg I emission line at 279.581 nm as a function of the number of shots. Average value and standard deviation (error bars) are calculated from 27 sites.

The analyst can easily conclude from Figure 4 that the signal related to the first shot is significantly different from the next ones. The first shot may thus be considered as a pre-shot and the corresponding signal may be removed from further data analysis. It is worth noticing that this decision of excluding the first signal can be taken after an ANOVA calculation as well as by the naked eye.

Number of shots

Once, the number of pre-shots has been investigated, the LIBS analyst has to determine the best number of laser shots to build the average signal to be included in further data processing. Indeed, this number strongly depends on the materials as well as focusing conditions and irradiance at the sample surface, expected to influence the ablation rate and consequently the LIBS signal at different depths.

- If the ablation rate is high, the crater depth increases fast, the LIBS signal thus decreases fast, and the LIBS measurement should be preferably achieved with a small number of shots to

guarantee a good signal-to-noise ratio of the average signal. ANOVA is expected in this case to allow determining the number of shots after which the signal drastically falls.

- At the opposite, if the ablation rate is low, the crater depth does not change that much from one laser shot to the next, the LIBS signal thus stays quasi unchanged, and the LIBS analyst may prefer to increase the number of shots in order to reduce the standard deviation. Nevertheless, even if quality might be continuously and asymptotically improved when increasing the laser shot number, the best number of laser shots should be minimized in order to save time and money. ANOVA is expected in this case to allow determining if the signal after one single shot is significantly different or not from the average signal from two shots, and so on.

The case displayed in Figure 4 clearly corresponds to the regime of low ablation rate. Thus, applying ANOVA as described above to this dataset allows to demonstrate that, in this case, the signal after one single shot (after pre-shot removal) was not significantly different than the average signal from two or more laser shots, meaning that the number of shots could be reduced, in this particular case, to only one.

To summarize, the LIBS analyst should select, in the case displayed in Figure 4, one pre-shot (to be excluded) and only one shot to be considered for further processing. More generally, based on the results of ANOVA calculations, the analyst should be able to determine the number of pre-shots as well as the optimum number of shots.

Number of sites

The sample analyzed by LIBS might be chemically and/or physically heterogeneous. In this case, averaging the LIBS signal over a series of sites at the sample surface is mandatory in order to be able to compare the LIBS analyses to reference analyses derived from a much larger amount of matter. But, depending on the LIBS instrument and measurement conditions, the LIBS signals of homogeneous

samples might also exhibit fluctuations from one site to another. Thus, once the number of pre-shots and shots have been determined by the analyst, the variations observed from one site to another should also be examined. Considering that the RSD value of the LIBS signal is expected to continuously decrease when increasing the number of sites, the analyst should theoretically prefer considering as many sites as possible for quality purpose. However, in practice, the LIBS analyst has to determine what is the minimum number of sites that provides acceptable statistical figures. Moreover, in several application fields such as cultural heritage, gemology, and forensics, the analyst will also care to reduce the damages at the sample surface.

To illustrate the importance of selecting the right number of sites, Figure 5 displays the background corrected average intensity values at the emission line of Ga I at 403.3 nm (vertical axis), for a sample of Gallium arsenide, as a function of the number of sites (horizontal axis). More precisely, to generate this plot, a series of LIBS measurements has been conducted at 27 locations at the surface of a GaAs wafer. The average value related to these 27 measurements is displayed on the right side of the plot and is considered being the ideal target value. Then, a series of 100 pairs of sites is randomly selected among the 351 possible combinations from the 27 sites, and the 100 corresponding values (average over 2 spectra) are reported in Figure 5 for a number of sites equal to 2. Then, a series of 100 triplets of sites is randomly selected among the 2925 possible combinations from the 27 sites, and the 100 corresponding values (average over 3 spectra) are reported in Figure 5 for a number of sites equal to 3. The same process is repeated for the other numbers of sites. Figure 5 clearly reveals that the average values over 25 points are very close to each other, whatever the locations randomly considered at the surface sample. At the opposite, the average values over 2 points can be drastically different depending on the locations randomly considered. Indeed, the extreme value of 5550 (in arbitrary units, a.u.) has been found, to be compared with the average value calculated from the 27 locations, being 3200 (a.u.). In this case, considering only two locations for the LIBS analysis is clearly insufficient. Moreover, and without surprise, Figure 5 also reveals that the higher the number of sites considered

at the sample surface the lower the risk to report an average value which is very different from the target one calculated from 27 sites.

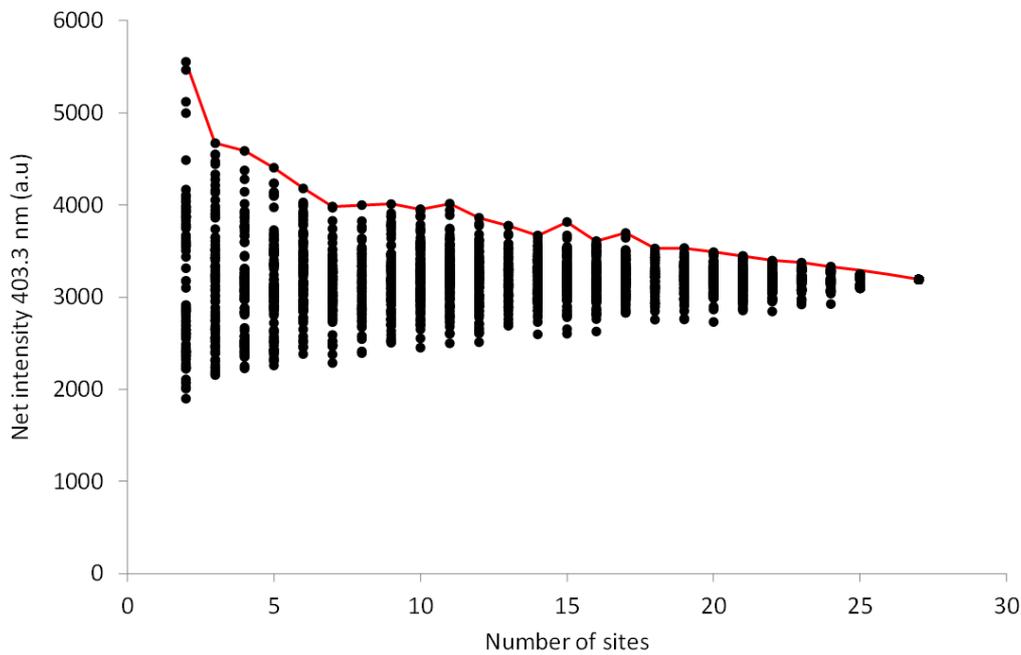


Figure 5: Evolution of the background corrected intensity of the Ga I emission line at 403.3 nm as a function of the number of sites considered at the sample surface. For each number of sites, 100 combinations were randomly chosen among 27 locations.

Based on this observation, we propose to calculate the maximum relative bias, namely the percentage of the difference between the maximum average value for a given number of sites and the target value, being the asymptotic value. As an example, in the case of 2 sites, the max average value being 5550 (a.u), the value of the corresponding maximum relative bias is $100 \frac{5550-3200}{3200} \approx 73\%$. Furthermore, when considering a number of sites of 15, the maximum relative bias falls to 19%. Thus, depending on the requirements of the application, the number of sites might be selected by the analyst from the values of the maximum relative bias. However, it is well-known that all the emission lines of a LIBS spectrum don't exhibit the same level of fluctuations. Thus, we recommend to set the number of sites

not only from the data corresponding to a single emission line (Fig.5), but from the values of the maximum relative bias calculated on a large number of emission lines present in the LIBS spectra.

At this step, let's emphasize that the fluctuations of the LIBS signal displayed in Figure 5 are not representative of the analytical performance of LIBS in general. Indeed, there are many cases for which only a few locations at the sample surface are sufficient to get the asymptotic average value. This example has been selected for pedagogical purpose since it allows to clearly understand the strategy we recommend to adopt. Finally, in the case of much lower fluctuations, the methodology discussed here can still be applied, and only the values of deviation and bias are expected to change.

Figure 6 displays the values of the maximum relative bias corresponding to the data displayed in Fig. 5 in a reduced spectral range to facilitate the reading. It is worth noticing that for a given number of sites (vertical axis) the maximum relative bias significantly varies with the wavelength. As an example, the bias can still be higher than 30% for several lines, even when considering 20 points at the sample surface. Moreover, if the requirement of the LIBS analysis is to get a maximum relative bias lower than 10% whatever the spectral line considered, then the number of sites must be at least equal to 25.

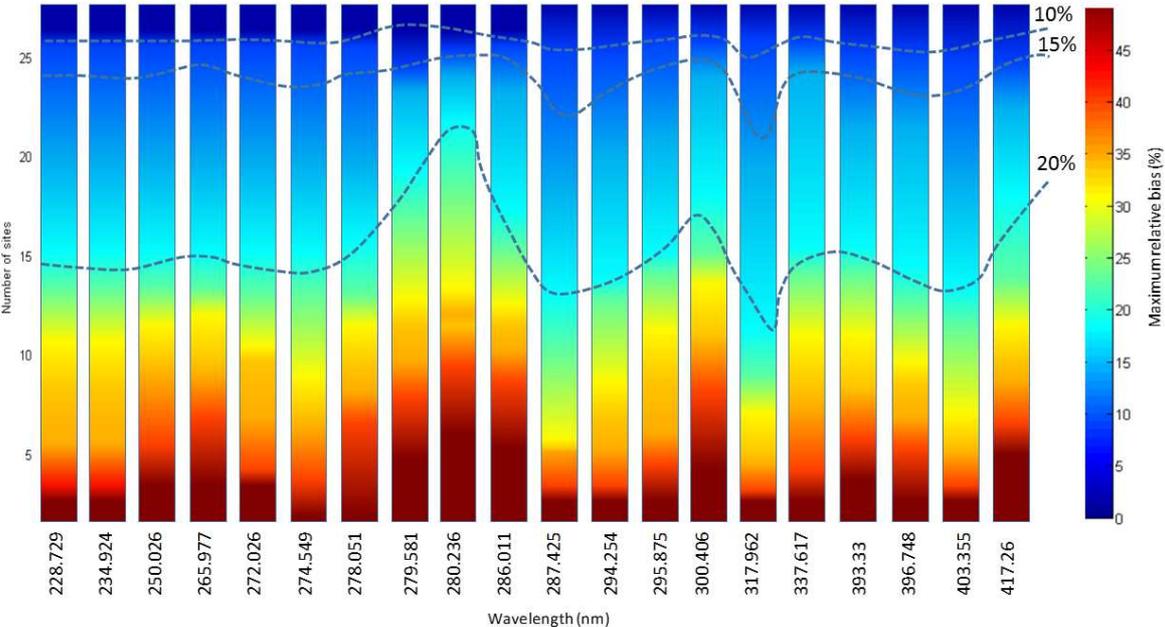


Figure 6: Maximum relative bias (% , in color) displayed for different values of the number of sites (vertical axis) for a selection of emission lines of interest (horizontal axis, wavelength values in nm).

The overview provided by Figure 6 is particularly interesting when further data processing is based on a multivariate approach, and/or normalization. At the opposite, in the case of univariate data processing, the analyst only needs to know what is the maximum relative bias at the wavelength of interest. In this case, the maximum relative bias can be very different from one wavelength to another.

To summarize, we suggest to adopt the strategy described in the present section, and illustrated in Figures 5 and 6 in the frame of a case study in order to determine the number of sites to be considered. Thus, the analyst can make this decision according quality requirements, e.g. maximum relative bias lower than 10%. Of course, the indicator of maximum relative bias might be replaced by any other indicator such as the relative standard deviation, for instance, but the methodology would stay unchanged. Again, we do insist on the fact that the levels of fluctuation reported in Fig. 5 and Fig. 6 are not typical of the LIBS analyses in general. Depending on both the instrument and the sample, the values corresponding to these indicators might be much lower. However, the discussion would be the same.

To conclude, in order to determine from the scratch the optimum sampling conditions the analyst is invited to follow the present guideline:

- i) select a number of laser shots per point (typically 20 to 50), large enough to allow determining the number of pre-shots and shots (cf. Fig. 4),
- ii) select a number of points/locations (typically 20 to 25) at the sample surface in order to allow approaching the average value related to an infinite number of points, with a relative bias compatible with the requirements of the analytical context (cf. Fig. 5 and 6),
- iii) record each single LIBS spectrum in order to be able to process them as described in the present section.

It is worth noticing that the default values given in the present guideline and related to the number of laser shots per point and the number of points should be chosen by the analyst in order to take into account the specificity of the LIBS instrument, such as the ablation rate. For instance, if the LIBS signal falls to zero after 10 shots at a same location, the total number of shots can be limited to 10 for the optimization process. Similarly, depending on the typical degree of heterogeneity of the samples to be analyzed, the value of the number of locations at the sample surface should be adjusted by the analyst.

3. Data filtering

Considering LIBS spectra, one can potentially observe wavelength drifts that may induce issues in element identification, classification and quantification, if they are not identified and corrected. To avoid a mismatch between the experimental and reference wavelengths (according to the NIST atomic database), the experimental spectra potentially need to be wavelength corrected. The risk is probably higher when considering spectra recorded by stand-off or handheld LIBS instruments. And, if the wavelength correction is often proposed and managed by the software communicating with the instrument, it is not taken into account in real-time and thus the recorded spectra are not all corrected. It is worth noticing that the goal of the wavelength correction proposed here is not to assist the wavelength calibration of the spectrometers during the measurement campaigns but to prepare the recorded data to further processing by suppressing potential wavelength drifts.

Technically, the analyst may consider that the first spectrum of a campaign is correctly calibrated in wavelength and that all the next spectra may be compared to it, for instance via a correlation function, and then corrected when needed, in order to finally observe a perfect overlap of selected emission lines in all the spectra. To illustrate the importance of this wavelength correction, we have run two PCA models from the same data set. Fig. 7.a displays the loadings of the first component for the uncorrected data, and Figure 7.b for the corrected ones. It clearly demonstrates that the loadings displayed in Fig. 7.a exhibit abnormal features similar to the derivative function of the atomic emission lines. This result is typical for a wavelength drift within the data set. At the opposite, the features

displayed in Figure 7.b are very different and look like normal features related to the emission lines, indicating that the correction has been efficient.

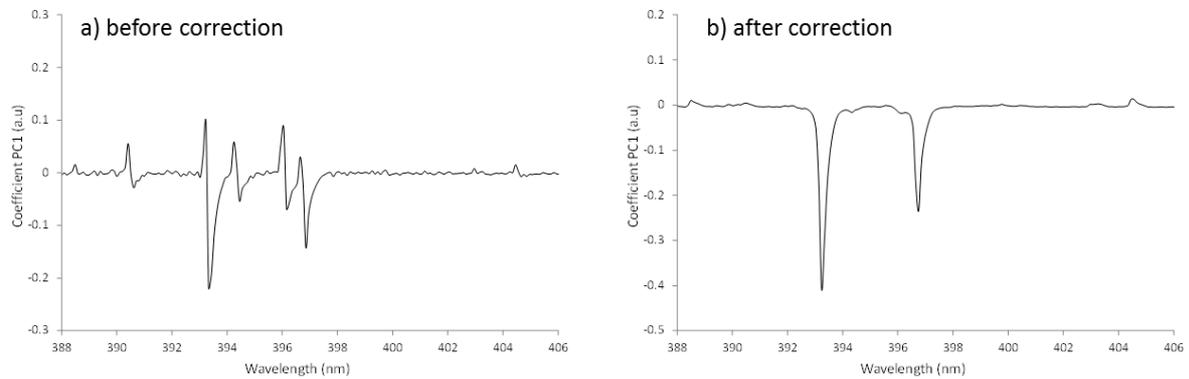


Figure 7: Loadings of the first component of two PCA models applied to LIBS spectra a) before and b) after correction.

Figures 8.a and b display the corresponding scores calculated for the two first components for the uncorrected and corrected spectra, respectively. Interestingly, the two score plots are very different, as a consequence of the wavelength correction applied to the LIBS spectra. And, while Figure 8.b (after wavelength correction) displays the actual variability of three groups of spectra within the dataset, Figure 8.a (uncorrected data), at the opposite, mainly exalts the existence of a spectral drift. This result can be generalized and it clearly illustrates the necessity of correcting spectral drifts prior to any data processing.

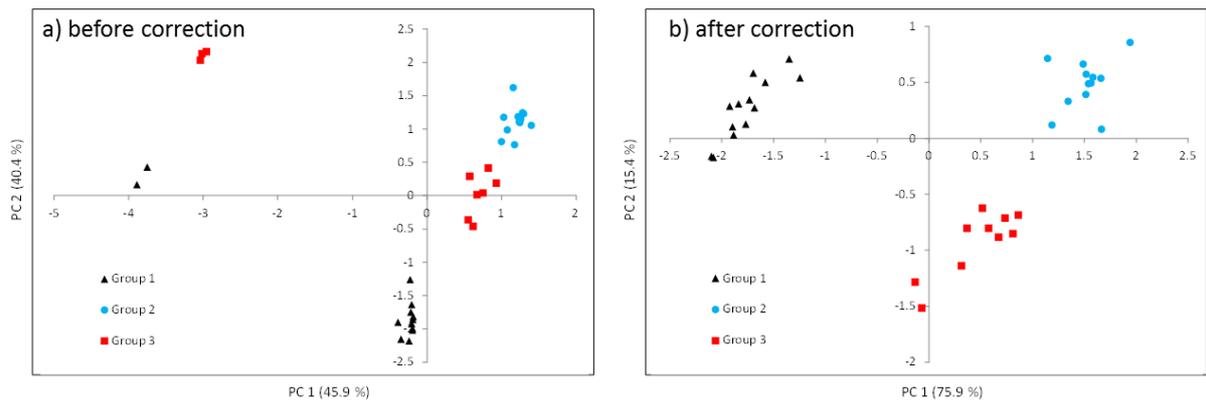


Figure 8: Score plots of two PCA models applied to LIBS spectra a) before and b) after correction.

In addition to the important question of wavelength correction, the second type of data control consists in detecting if any saturated features are present in the LIBS spectra. Indeed, the presence of saturated signals prevents from correctly applying normalization techniques such as SNV (consisting in subtracting the mean value of the spectral intensities to each intensity value and to divide the result by the standard deviation of the initial spectral intensities), or processing the dataset through multivariate approaches. Saturated features are, by definition, not exploitable and must be filtered out. Consequently, if saturated spectral features are detected, the analyst must either exclude them from the spectra or exclude the whole spectra exhibiting those features. In the first type of filtering, the total number of spectra considered in the dataset is kept unchanged but the number of variables is reduced to treat only the unsaturated spectral features. In the second type of filtering, the number of variables considered in the dataset is kept unchanged but the number of spectra is reduced to keep only the ones without saturation. The analyst will have to choose between these two types of filtering depending on the context of the analysis.

In conclusion, the guideline proposed here aims at alerting the analyst to verify the nature of the recorded data, re-calibrate the LIBS spectra in wavelength when needed, and filter out saturated data prior to further data processing. The wavelength calibration can be achieved through a different

method than the one described here and the process of filtering can also be customized by the analyst. However, we claim that these two steps of wavelength recalibration and filtering of saturated data are mandatory to guarantee quality. This basic advice is obviously not limited to LIBS and holds for any spectroscopic technique.

4. Sorting

One of the main objectives of the LIBS analyses is to achieve sorting either in the context of sourcing (archeology, gemology, forensics...) or recycling, based on first the construction, and secondly the routine use of supervised models. In addition, unsupervised models are perfectly adapted to the case of smart sampling, where the LIBS spectra of totally unknown samples are directly compared in order to understand the variability of the dataset, and thus optimize the sampling for further analyses. Efficient sorting can be rarely achieved by considering one single variable and the use of multivariate approaches is very common.

Classification - supervised models

Partial least squares – discriminant analysis (PLS-DA) is the approach the most commonly applied to LIBS spectra for sorting purpose [6]. This supervised multivariate approach allows taking into account the influence of several variables for discriminating the dataset into separate groups. However, it is worth noticing that it is generally quite difficult to strictly compare the sorting ability of different models due to the lack of a common assessment process.

El Haddad *et al.* [6] proposed to adopt good practices by splitting the original dataset into two subsets, namely the training set and the validation set. The training set is used to build the model while the validation set is used to evaluate, *a posteriori*, the sorting ability of the model. They also proposed to calculate for each model the well-known confusion matrix, which allows to estimate the sorting ability of a model, for both the training and the validation set. However, the only information contained in this matrix is *belong/don't belong to a given class* but the errors of the decisions are not given. To

illustrate this point, let's consider a sample assigned to a given class A, among the three classes A, B, and C, according to the confusion matrix. Considering the probability to belong to each class, this sample may be characterized by 40% to belong to class A, 30% to class B, and 30% to class C. This clearly illustrates that the confusion matrix only gives a partial information and doesn't permit to deal with complex cases. Consequently, in order to achieve high quality sorting, we recommend to calculate not only the confusion matrix, but also the probabilities to belong to each class. Syvilay *et al.* [20] applied this approach in the context of geological obsidian sourcing to compare different sorting models, derived from PLS-DA and QDA (quadratic discriminant analysis). Based on this approach, they selected the model presenting the best sorting ability.

Smart sampling - unsupervised models

The analyst may want to check if LIBS spectra belonging to a given dataset are similar or not, without any additional information to supervise the decision. Indeed, after analyzing samples in the field (rocks, soils, etc...), unsupervised sorting allows to avoid collecting very similar samples and to focus further analyses on samples exhibiting different characteristics. In this case, either PCA (principal components analysis) or ICA (independent components analysis) models [21] can be applied as a first step. It is interesting to notice that these multivariate approaches are often interpreted by the naked eye and this is probably why some people consider them as classification techniques, which is not correct, by principle. Thus, instead of sorting the data by the naked eye, we recommend to numerically determine the groups of similarities, or classes, using numerical methods, such as K-means [22] or kernel density [23]. At the end, the analyst needs to assess if the selected approach of unsupervised sorting can be considered as reliable and robust. In the present guideline, we recommend to assess the sorting ability of the studied models through specific figures of merit: i) the probability to belong to each group, and ii) the percentage of overlap between groups. These two indicators drive to establish the number of well-separated groups. In practice, the number of groups doesn't matter and the best model is

expected to exhibit, for a given sample, a high probability to belong to one group and also a percentage of overlap between groups as low as possible.

A case-study has been selected to illustrate the proposed methodology of unsupervised sorting. It concerns a series of 12 cinnabar-based samples prepared according to four well-established techniques, namely casein, glue, fresco, and egg yolk. These samples were prepared by a restorer who mastered the application techniques. Ten single-shot LIBS spectra were recorded for each sample, at ten different locations, and the resulting 120 spectra were examined. Each spectrum consisted in 5949 variables, which were introduced into a PCA model and the corresponding scores plot (PC1-PC2) was displayed in Figure 9.a. In this figure, the four types of preparation are displayed with four markers: blue disc for casein, red triangle for glue, green square for fresco, and black diamond for egg yolk. It's worth noticing that the data exploration performed by the PCA is totally unsupervised, and thus no marker should be used to distinguish the different types of samples. However, we decided to show the colored markers in Fig. 9.a to facilitate the discussion. From this figure, one can distinguish four groups of points with the naked eye. The existence of these four groups was also confirmed by the calculation of the kernel density plot displayed in Fig. 9.a. Thanks to the colored markers, it was easy to verify that, in the present case, each group identified by the kernel density algorithm corresponded to one type of preparation. Furthermore, Figure 9.b displays, for each sample, the probability to belong to a group. In the present case, the probability was equal to 1 (namely 100%) for each sample except for two of them.

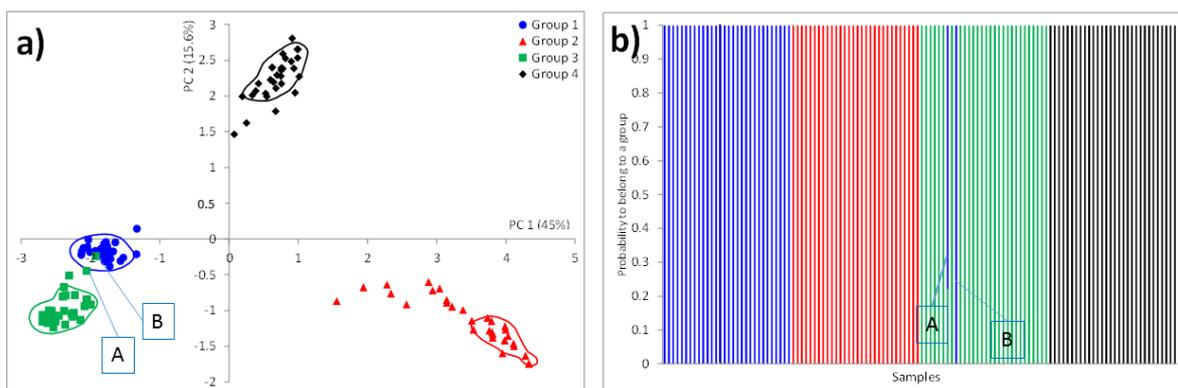


Figure 9: a) Score plots of a PCA model applied to the LIBS spectra of cinnabar-based samples prepared with different techniques (casein – group 1, fresco – group 2, glue – group 3, and egg yolk – group 4); the kernel density plot (closed curves) reveals the presence of four groups; and b) for each sample, the probability to belong to a group. The points labelled “A” and “B” are not assigned to the right group with a probability equal to 1.

More precisely, Figure 9.b reveals that two samples corresponding to the fresco preparation were not assigned to group 3 with a probability equal to 1, as expected. The sample related to the point labelled “A” was characterized by a probability only equal to 0.2 to belong to group 3, and a probability of 0.8 to belong to group 1. The sample related to the point labelled “B” was wrongly assigned to group 1 with a probability equal to 1. From Fig. 9.a, these probabilities were quite easy to interpret since the point “B” was located inside the density kernel plot (blue curve) of group 1, in perfect agreement with the probability equal to 1 to belong to this group. In addition, the point “A” is much closer to the kernel density plot related to group 1 than those related to group 3, in good agreement with the probabilities of 0.8 and 0.2.

Finally, in the present case, the probability of overlapping between the groups was equal to zero, as one can easily verify from the kernel density plots displayed in Fig. 9.a. In complement to the probability to belong to a group for a given sample, this second indicator is quite useful since it confirms that the groups 1 and 3 are significantly separated, indicating that the unsupervised sorting model investigated here is robust. At the opposite, a model exhibiting either a high probability of overlapping between the groups, or many values of probability to belong to a group different to 1 would indicate that the corresponding unsupervised sorting model is less robust than the one presented here.

The advantages of following the present guideline are twofold; the analyst can assess if the sorting model can be considered as reliable, and if several models are tested, they can also be compared between them. The best model can be identified in order to make the decision as reliable as possible. In conclusion, unsupervised sorting is particularly useful in the frame of smart sampling. Indeed, the

similarities/differences between the data are the key-information to optimize the selection of the samples to be further analyzed by reference techniques that might be time-consuming and expensive. We consider that unsupervised sorting of LIBS data recorded on-site is very interesting, but in this case, like in the other ones, the reliability should be maximized and the present guideline is expected to contribute to it.

5. Quantitative analysis

Most of the LIBS applications require quantitative analyses. However, Motto Ros *et al.* [2] have demonstrated that the steps involved in data processing are critical and can drive to very different results not only regarding the predicted values of concentration, and consequently the RMSE (root-mean square error) values, but also regarding the uncertainties as well as the limits of detection and quantification. As a first recommendation about quantitative LIBS, we would like to point out that the univariate approach, i.e. the classical calibration curve, should be preferred whenever it allows answering the requirements of the analysis. Good practices and figures of merit corresponding to this case have been given by Motto-Ros *et al.* [2]. They allow comparing different univariate models between them and to select the best one, without *a priori* decision. Nevertheless, it has been demonstrated that univariate models were not able to deal with matrix effects and this is the reason why multivariate approaches are often applied to analyze the LIBS spectra of complex samples. Among them, the partial least-squares regression (PLS) is the most commonly used. PLS can be run on the entire dataset as well as on a selection of variables. There are well-known figures of merit [6] that enable the analyst to choose the most relevant variables and to assess the performance of the PLS models.

In the present discussion, we would like to point out that it is very important to be able to compare multivariate and univariate models in order to objectively demonstrate the relevance of adopting multivariate approaches, which is very rarely discussed in the scientific literature. And, since multivariate approaches provide as an output of the models the relationship between the predicted

and reference concentration values, we recommend to display the graph “predicted vs. reference concentration values” for both univariate and multivariate approaches, prior to calculating any figure of merit.

A univariate model consists in building the regression function between the LIBS signal and the reference concentration values. This regression function is commonly used to predict the concentration value related to a measured signal. It means that, for a given LIBS signal obtained from a sample belonging to the calibration set, we have simultaneously the reference concentration value and the predicted concentration value from the regression model. It thus becomes possible to plot the “predicted vs. reference concentration values” even for a univariate model. It is not common to display this graph in the case of univariate models, and it is worth noticing that a second regression model is now calculated between the predicted and the reference concentration values. Consequently, the figures of merit proposed by Motto-Ros *et al.* [2] have to be totally revisited in order to be adapted to this new representation. In the present guideline, we recommend to systematically calculate three figures of merit, namely the coefficient of determination R^2 , the limit of detection LOD, and the root-mean square error (RMSE) since they allow to compare the different quantitative models to each other. Let’s notice that LOD or LOQ can be indifferently chosen since they are closely related ($LOQ = \frac{10}{3}LOD$).

To illustrate this discussion, we have introduced the dataset already exploited by Motto-Ros *et al.* [2] into a PLS model dedicated to the quantification of aluminum, and compared the resulting figures of merit to those related to two selected univariate models, from the plot “predicted vs. reference concentration values”. The first univariate model corresponds to a linear regression between the LIBS signal and the reference concentration values, while the second one corresponds to a quadratic regression. Motto-Ros *et al.* [2] have demonstrated that the quadratic regression provided the best model. In Table 1, are reported the values of R^2 , LOD and RMSECV for the two selected univariate

models, derived from the plot “predicted vs. reference concentration values”, and one PLS model built from 3 components and a pre-selection of 5500 variables.

Table 1: RMSECV, LOD and R^2 values for two selected univariate and multivariate models (see in the text for details) for the quantification of aluminum applied to the dataset previously exploited by Motto-Ros *et al.* [2].

| Model | RMSECV (ppm) | LOD (ppm) | R^2 |
|---|--------------|-----------|-------|
| Univariate (linear) 309.27 nm Peak intensity | 80 | 140 | 0.979 |
| Univariate (quadratic) 309.27 nm Area of the Voigt profile | 40 | 70 | 0.995 |
| Multivariate (PLS, 5500 variables, 3 components) | 360 | 230 | 0.942 |

From the values reported in Table 1, the analyst can conclude that, in the present case, the univariate model originating from a quadratic relationship between the LIBS signals and the reference values of concentration provides the best quantitative performance. However, it should be noticed that the PLS model given here was not particularly optimized. Moreover, Table 1 just aims to illustrate that univariate and multivariate models can be rigorously compared.

Let’s also notice that El Haddad *et al.* [6] emphasized the importance of evaluating the risk of overfitting by splitting the original dataset into two parts, a calibration set to build the calibration model and a validation set to assess its prediction ability. Indeed, they mentioned that overfitting is observed when the root-mean square error is much lower for the calibration set than for the validation set. They also advised to verify if the performance of a quantitative model were obtained by chance by applying the so-called Y-randomization method, consisting in randomly changing the order of the concentration values in order to verify that, after randomization, the figures of merit significantly change and that the prediction ability of the new calibration model becomes very poor. These good practices, which have been detailed by El Haddad *et al.* should obviously be applied to any LIBS data,

when possible, and thus be taken into account to reinforce the present discussion. However, considering the case presented here, the number of data was too small to split the dataset into calibration/validation subsets and this is why the question of overfitting has not been addressed here.

To summarize, based on the methodology proposed here, the LIBS analyst is able to systematically compare univariate to multivariate models between each other. This approach opens the way for any LIBS analyst to objectively find the best quantitative model for a given dataset, and thus to efficiently participate to future interlaboratory comparison.

Conclusion

We have proposed a guideline to assist the LIBS analyst from A to Z, and to assess the relevance of each step of data processing thanks to relevant figures of merit and we have illustrated, through selected examples, the risks of making non-robust or even wrong conclusions when the basic advices proposed in the present guideline are not taken into account. This guideline describes how to verify that quality is at its highest level, at each step of the measurement campaign and data processing. The LIBS analyst now possesses all the tools necessary to check the status of the instrument and optimize the parameters and sampling prior to any measurement campaign. Then, he/she can check the quality of the dataset in order to correct some spectra in wavelength when needed and to filter out spectra exhibiting saturated features. The guideline also indicates how to compare between different sorting models, being supervised or not, as well as quantitative models, being either univariate or multivariate. We claim that this guideline should allow to drive future high-quality LIBS measurements as well as highly reliable analytical results.

Acknowledgement

The authors acknowledge the regional council of *Region Nouvelle Aquitaine* for the funding of J. Guezenoc's doctoral position in the frame of the project AGROLIBS.

References

- [1] DW. Hahn, N. Omenetto, Laser-induced breakdown spectroscopy (LIBS), part II: review of instrumental and methodological approaches to material analysis and applications to different fields, *Applied Spectroscopy*,
- [2] V. Motto-Ros , D. Syvilay, L. Bassel, E. Negre, F. Trichard, F. Pelascini, J. El Haddad, A. Harhira, S. Moncayo, J. Picard, D. Devismes, B. Bousquet, Critical aspects of data analysis for quantification in laser-induced breakdown spectroscopy, *Spectrochimica Acta Part B* 140 (2018) 54–64
- [3] E. Tognoni, V. Palleschi, M. Corsi, G. Cristoforetti, Quantitative micro-analysis by laser-induced breakdown spectroscopy: a review of the experimental approaches, *Spectrochimica Acta Part B* 57 (2002) 1115–1130
- [4] F. Fortes, J. Laserna, The development of fieldable laser-induced breakdown spectrometer: No limits on the horizon, *Spectrochimica Acta Part B: Atomic Spectroscopy*, 65 (2010) 975-990.
- [5] E. Tognoni, M. Corsi, G. Cristoforetti, V. Palleschi, N.Omenetto, I. Gornushkin, B.W. Smith, J.D. Winefordner, From sample to signal: a complex route to quantitative analysis, in: A. Miziolek, V. Palleschi, I. Schechter (Eds.), *LIBS: Fundamentals and Applications*, Cambridge University Press, in press.
- [6] J. El Haddad, L. Canioni, B. Bousquet, Good practices in LIBS analysis: Review and advices, *Spectrochimica Acta Part B* 101 (2014) 171–182
- [7] A. M. Popov, F. Colao, R. Fantoni, Spatial confinement of laser-induced plasma to enhance LIBS sensitivity for trace elements determination in soils, *Journal of Analytical Atomic Spectrometry*, Volume 25, Issue 6 (2010), 837–848
- [8] B. C. Castle, K. Talabardon, B. W. Smith, and J. D. Winefordner, Variables Influencing the Precision of Laser-Induced Breakdown Spectroscopy Measurements, *Applied Spectroscopy* Volume 52, Number 5 (1998) 649-656
- [9] L. S. Nelson, The Shewhart Control Chart—Tests for Special Causes, *Journal of Quality Technology*, Volume 16, Issue 4, (1984) 237-239

- [10] S. W. Roberts, A Comparison of Some Control Chart Procedures, *Technometrics*, Volume 8, Issue 3 (1966) 411-430
- [11] J. Castro, E. R. Pereira-Filho, Twelve different types of data normalization for the proposition of classification, univariate and multivariate regression models for the direct analyses of alloys by Laser-induced breakdown spectroscopy (LIBS), *Journal of Analytical Atomic Spectroscopy*, Issue 10, (2016)
- [12] J. S. Cowpe, J. S. Astin, R. D. Pilkington, A. E. Hill, Application of response surface methodology to laser-induced breakdown spectroscopy: Influences of hardware configuration, *Spectrochimica Acta Part B*, Volume 62, Issue 12 (2007) 1335-1342.
- [13] L. E. García-Ayuso, J. Amador-Hernández, J. M. Fernández-Romero, M.D. Luque de Castro, Characterization of jewellery products by laser-induced breakdown spectroscopy, *Analytica Chimica Acta*, Volume 457, Issue 2 (2002) 247-256.
- [14] M. A. Khater, J. T. Costello, E. T. Kennedy, Optimization of the Emission Characteristics of Laser-Produced Steel Plasmas in the Vacuum Ultraviolet: Significant Improvements in Carbon Detection Limits, *Applied Spectroscopy*, Volume 56, Issue 8 (2002) 970-983
- [15] C. Aragon, J. A. Aguilera, and F. Penalba, Improvements in Quantitative Analysis of Steel Composition by Laser-Induced Breakdown Spectroscopy at Atmospheric Pressure Using an Infrared Nd:YAG Laser, *Applied Spectroscopy*, Volume 53, Issue 10 (1999) 1259-1267
- [16] J-B. Sirven, P. Mauchien, B. Sallé, Analytical optimization of some parameters of a Laser-Induced Breakdown Spectroscopy experiment, *Spectrochimica Acta Part B*, Volume 63 (2008) 1077–1084
- [17] B. Lal, H. Zheng, F. Y. Yueh, J. P. Singh, Parametric study of pellets for elemental analysis with laser-induced breakdown spectroscopy, *Applied. Optics*, volume 43(2004) 2792–2797

[18] J. Li, J.Lu, Z. Lin, S. Gong, C. Xie, L. Chang, L. Yang, P. Li, Effects of experimental parameters on elemental analysis of coal by laser-induced breakdown spectroscopy, *Optics and Laser Technology*, Volume 41 (2009) 907–913

[19] T. Trejos, J. R. Almirall, Sampling strategies for the analysis of glass fragments by LA-ICP-MS : Part I. Micro-homogeneity study of glass and its application to the interpretation of forensic evidence, *Talanta*, Volume 67 (2005) 388–395

[20] D. Syvilay, B. Bousquet, R. Chapoulie, M. Orange, F.-X. Le Bourdonnec, Advanced statistical analysis of LIBS spectra for the sourcing of obsidian samples, *Journal of Analytical Atomic Spectrometry*, (2019).

[21] O. Forni, S. Maurice, O.Gasnault, R. C. Wiens, A. Cousin, S. M. Clegg, J-B. Sirven, J. Lasue, Independent component analysis classification of laser induced breakdown spectroscopy spectra, *Spectrochimica Acta Part B*, Volume 86 (2013) 31-41

[22] N. Dupuy, O. Galtier, Y. Le Dréau, C. Pinatel, J. Kister, J. Artaud, Chemometric analysis of combined NIR and MIR spectra to characterize French olives, *European Journal of Lipid Science and Technology*, Volume 112, Issue4 (2010) 463-475

[23] W-J. Wang, Y-X. Tan, J-H. Jiang, J-Z Lu, G-L. Shen, R-Q. Yu, Clustering based on kernel density estimation: nearest local maximum searching algorithm, *Chemometrics and Intelligent Laboratory Systems*, Volume 72, Issue 1 (2004) 1-8

High quality data

Assessment

Comparison

High performance
Reliability

Are relevant figures of merit exploited ?

Are **good practices** applied during data processing ?

Does a part of the **data** need to be filtered out ?

Do the **measurement conditions** need to be optimized ?

Does the **instrument** works well ?

Increasing analysis quality in LIBS