Spectrochimica Acta Part B: Atomic Spectroscopy 195 (2022) 106508

Contents lists available at ScienceDirect

Spectrochimica Acta Part B: Atomic Spectroscopy

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

Interesting features finder (IFF): Another way to explore spectroscopic imaging data sets giving minor compounds and traces a chance to express themselves

Qicheng Wu [a], César Marina-Montes [b], Jorge O. Cáceres [c], Jesús Anzano [b], Vincent Motto-Ros [d], Ludovic Duponchel [a,*]

[a] Univ. Lille, CNRS, UMR 8516 – LASIRE – Laboratoire de Spectroscopie pour Les Interactions, La Réactivité et L'Environnement, Lille F-59000, France
[b] Laser Lab, Chemistry & Environment Group, Department of Analytical Chemistry, Faculty of Sciences, University of Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain
[c] Laser Chemistry Research Group, Department of Analytical Chemistry, Faculty of Chemistry, Complutense University of Madrid, Plaza de Ciencias 1, 28040 Madrid, Spain
[d] Institut Lumière Matière, UMR 5306, Université Lyon 1 – CNRS, Université de Lyon, Villeurbanne 69622, France

ARTICLE INFO

Keywords:
LIBS imaging
Chemometrics
Spectral selection
Minor compounds
Traces
Pixel detection

ABSTRACT

Today, we acquire larger and larger spectroscopic imaging data sets on complex samples. Even before proceeding with a spectroscopic analysis, we often have information concerning for example the presence of major compounds. However, we are not in the same position regarding the presence and location of minor compounds and traces in the sample, while this is generally a more interesting piece of information. Modern spectroscopic imaging uses chemometric tools that allow the simultaneous exploration of the entire spectral range available. These well-tested tools, such as principal component analysis, often exploit the variance contained in the data set to extract maximum chemical information. In general, we can say that these approaches are quite efficient but they are not completely adapted to the characteristics of the data acquired during a spectral imaging experiment. Indeed, such data sets have generally a limited signal to noise ratio and minor compounds and traces are often present on a small number of pixels compared to the total number in the considered data set. It is then quite possible that these compounds are present in a sample but not detected by the chosen multivariate tool. The goal of this work is to introduce an approach called *Interesting Features Finder* (IFF) able to retrieve minor compounds and traces independently of the variance they may express in the spectral data set. Like other state of the art methods, it uses the notion of convex hull but with the great advantage of working directly on the raw data, without prior projection to reduce their dimensionality. By means of a synthetic spectral data set and a micro-LIBS (laser-induced breakdown spectroscopy) imaging one acquired on a sample taken in Antarctica, we will show the potential of this approach in terms of sensitivity of detection of compounds but also of robustness against noise.

1. Introduction

Chemometrics began in the 1970s driven by a few far-sighted individuals who perceived the power of applying computer-based algorithms to processing analytical data [1]. This was of course possible as soon as they had enough data on the samples they wanted to explore and this is exactly what spectroscopy would bring. At the same time, it was demonstrated that it is possible to predict the concentrations of products of interest in complex samples by means of their near-infrared spectra introduced in a multivariate regression model, which would completely change the field of analytical chemistry and many other fields [2,3]. It is obvious that the history of vibrational spectroscopy is closely linked to that of chemometrics. Vibrational spectroscopy has first of all shown by its needs the way that chemometrics should take for a better valorization of the acquired spectral data. After several decades, the roles are even reversed in the sense that new concepts of data processing using particular data structures can lead us to think about new instrumental developments able to generate such data.

* Corresponding author.
E-mail address: ludovic.duponchel@univ-lille.fr (L. Duponchel).

https://doi.org/10.1016/j.sab.2022.106508
Received 1 July 2022; Received in revised form 28 July 2022; Accepted 30 July 2022
Available online 3 August 2022

What we have just described for vibrational spectroscopy has been observed and is still observed for all spectroscopies, Laser-Induced Breakdown Spectroscopy (LIBS) being no exception to this evolution with of course different timing. The 1960s saw the birth of LIBS, with instrumental development focusing on microanalysis [4]. For many years, this spectroscopy has been primarily a physicists' technique used for bulk analysis. It is only since the 2000s that researchers have been highlighting the real analytical potential of LIBS microanalysis and imaging [5]. Their works has shown that this imaging technique has many interesting characteristics that few analytical techniques can claim, such as acquisition rates of up to 1000 Hz (i.e. 1000 spectra/s) [6], a spatial resolution at the micrometer scale [7], allowing a spectral and spatial exploration of very large areas of a sample [8]. In the last ten years, we have seen a clear acceleration of LIBS imaging with increasingly fast, accurate and reproducible measurements. The amount of data acquired has also increased significantly since we have gone from about 1000 spectra acquired for a sample in 2010 to more than 10 million spectra today, an increase in the amount of data by a factor of 10 every 2 years [9–11]. It should be noted that LIBS imaging is certainly the only one to have shown such a rapid evolution over such a short period. As a first sight, the low bandwidth of the observed emission lines suggests that a simple integration of the signal at a given wavelength should allow us to generate a map of an element of interest. Specialists in the field know that one must be careful on this last point. Indeed, the richness of the elemental spectra and the presence of many elements in heterogeneous samples does not allow us to unanimously ensure that the selected wavelength is specific to a single element. Therefore, chemical maps generated in this way should be treated with the utmost caution. In order to address this issue but also to propose a more exhaustive exploration of LIBS imaging data sets, chemometric methodologies have been developed in recent years to simultaneously explore the whole spectral range available. Tools as simple as Principal Component Analysis (PCA) [12] but also more advanced methodologies based on classification [13], clustering [14], or even spectral unmixing in the context of data fusion between LIBS and Plasma Induced Luminescence (PIL) [15] or LIBS and Raman spectroscopies [16] have been implemented and much more [17]. However, we had to adapt our data processing methodologies because the application of these classical chemometrics tools is not without difficulty when the spectral data set contains more than one million spectra.

The previous paragraphs tend to show the excellent dynamism of LIBS imaging with more and more reliable and fast instrumentations associated to chemometric analysis for the exploration of always-larger data sets. However, a problem persists in the multivariate processing of spectral data. Indeed, during the exploration of a data set we want to be as exhaustive as possible and therefore detect both major and minor compounds and traces. Unfortunately, minor compounds and traces are often present on a small number of pixels (compared to the total number of pixels in the data set) therefore associated with very low explained variances. This is the problem because the majority of chemometric algorithms using the concept of expressed variance do not (or hardly) allow the detection of these compounds, especially when the signal to noise ratio is limited. The aim of this work is to introduce a new algorithm called *Interesting Features Finder* (IFF), based on the concept of convex hull, capable of detecting the presence of compounds whether they are major, minor or trace. In the following sections, we will first describe the IFF algorithm and what it can bring compared to a state-of-the-art approach, called *Essential Spectral Pixels* (ESP) [18,19]. We will then introduce a synthetic data set as well as a LIBS imaging data set from an Antarctic aerosol measurement campaign. Finally, we will apply our approach to show what it can bring to the exploration of of complex samples.

2. Material and methods

2.1. Interesting features finder (IFF)

In order to understand the IFF algorithm, we need to represent our spectra in a different but commonly used way in chemometrics. This allows us to make a mental representation quite different from the spectra we observe on our computer. Therefore, a spectrum is considered as a point (or a vector) in a multidimensional space as represented in Fig. 1a. This multidimentional space has as many dimensions as there are wavelengths (p) in a spectrum. As observed, a data set takes the form of a cloud of points with varying densities throughout the volume. The questions that arise now are: what is the convex hull of such a point cloud (i.e. the spectral data set) and why can it be useful for our data exploration problem? A convex hull is a geometric object with applications in many scientific fields as varied as pure mathematics [20], statistics [21], economics [22], ethology [23] or quantum physics [24] to name a few. As represented in Fig. 1b, the convex hull is the region bounded by a balloon that would deflate as it contracts to its minimum, carried by the outer points of the point cloud. The number of points on the convex hull (only 4 spectra for this toy example) is then very limited compared to the total number of spectra in the data set. All the interest of this concept lies in the fact that, mathematically, all the spectra inside the convex hull are linear combinations of the spectra on its surface. Consequently, major / minor compounds and traces can potentially be found on the surface of the convex hull regardless of the variance they may express in the data set. So how to find the convex hull of a data set? The majority of the software proposing the calculation of the convex hull of a data set uses the *Quickhull* algorithm published in 1996 [25]. This is the approach proposed in the well-known *Matlab* computing environment but is also found in many others. However, it is very difficult and even sometimes impossible to calculate the convex hull of a data set when it is huge and more importantly when the number of dimensions of the considered data space exceeds 5–6. Of course, we have always much more than 5–6 wavelengths in our spectral range and our LIBS imaging data sets are getting bigger and bigger (sometimes with several million spectra). Aware of the interest of the convex hull, a team of chemometricians has recently proposed to use PCA beforehand in order to reduce the number of dimensions to investigate in a method they called ESP [18,19]. Very interesting results were published with the method by keeping only the scores of the first 5 or 6 principal components on which the convex hull could be extracted. From our point of view, using a preliminary compression method and fixing a number of components a priori remains rather risky in our case. Indeed, this would mean potentially losing any chance of seeing traces or even minor compounds despite obtaining the convex hull. It is in this sense that we propose in this work a direct calculation of the convex hull on the raw data set (i.e. whatever the number of spectral variables and the total number of pixels) using the *Interesting Features Finder* (IFF) tool we have developed.

The IFF algorithm is finally based on a rather simple idea since we successively project the whole data set on a large number of randomly drawn vectors. Fig. 1c illustrates this principle. Given a first random vector in the data space, we make an orthogonal projection of all spectra of the data set onto it. We then identify the 2 spectra with the most extreme coordinates on this axis. Next, we keep this information in a list and update the selection frequency of a spectrum if it has already been selected for previous random vectors. After the exploitation of a large number of random vectors (10,000−100,000), the final list contains the most interesting spectra of the data set, their selection frequency being used to measure their importance. In order to further reduce the list of selected pixels, we keep only the pixels with a selection frequency higher than 2% compared to the most frequently selected pixel in the data set, which represents 100%. At first sight, the choice of this percentage may seem arbitrary. This criterion can appear quite arbitrary. Nevertheless, we should not forget that this method is unsupervised and thus corresponds to a blind analysis of the dataset. Future users will have to modify

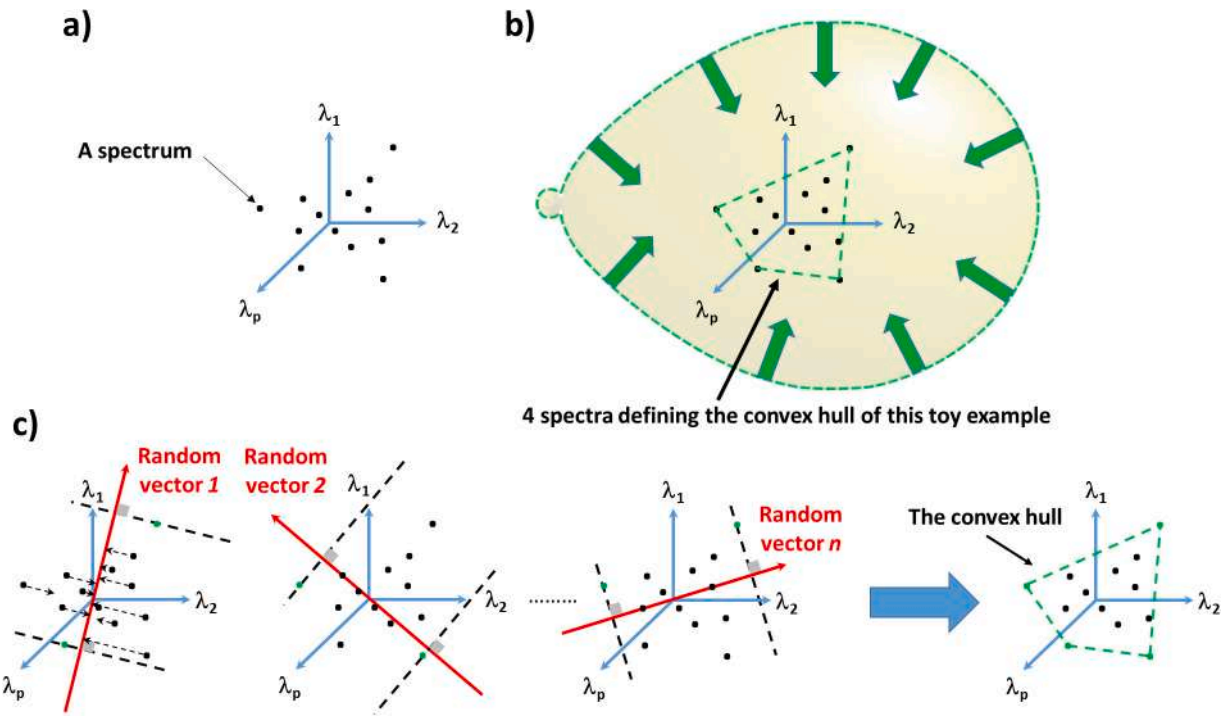

Fig. 1. a) Representation of a spectroscopic data set in a multidimensional space and b) the corresponding convex hull. c) Successive steps of the IFF algorithm in order to obtain the convex hull of the data set from its raw form.

this percentage according to the signal-to-noise ratio of the dataset. However, they will have to take the largest percentage possible in order to keep a limited list of pixels at the risk of losing the interest of the concept with a final list that would be unmanageable.

All calculations in this work have been performed under the Matlab 2016b environment. The Matlab code of the *IFF* algorithm is provided in the supplementary material.

2.2. Spectroscopic data sets

The first data set used in this work consist of synthetic LIBS spectra we have generated considering a small spectral range. The advantage of using such simulations lies in the fact that all the parameters potentially influencing a given problem are under control. Then it will be possible to generate spectral interferences between the present elements but also apply different noise levels. Since we will have generated these data, we have access to a ground truth concerning, in our case, the presence of pure pixels in the selection of spectra proposed by the ESP and IFF approaches. Fig. 1S in supplementary material shows the seven pure spectra considered in this simulation containing 500 spectral variables each. We have then generated 230,230 synthetic spectra by linear combinations of these 7 pure spectra considering concentrations varying from 0 to 1 by step of 0.05 imposing the constraint to have the sum of the 7 concentrations equal to 1. A white noise varying from 2 to 8% of the average intensity of the spectrum has also been added to each spectrum, which allow to test the robustness of the approaches to this often very critical parameter in the analysis of spectroscopic data. Fig. 2S in supplementary material shows as an example a spectrum of a mixture extracted from the data set containing approximately the same concentration for the 7 elements as a function of the applied noise level. It is easy to notice that the observation of the 7 contributions in the considered spectrum is already difficult when the noise is absent but also that the four most intense spectral contributions initially observed are at the limit of the detection for the most important noise levels.

The second data set considered in this work is a real one. It was obtained through LIBS imaging measurements made on one of the atmospheric aerosol samples collected during the 2018/2019 Spanish Antarctic campaign in the surroundings of the Spanish Antarctic Research station "Gabriel de Castilla" on Deception Island (62°58′09"S, 60°42′33"W). The objective of the study was to better understand the origin of airborne particles in this specific region of the globe. We only give here the main information concerning the sample preparation and the acquisition parameters for this LIBS imaging experiment, readers interested in the subject being invited to a previous work already published [26]. Regarding the preparation of the sample, air particles were first collected in a circular quartz microfiber filter paper through a high-volume sampler (30.6 m^3 /h) considering a total sampling time of 72 h. We then cut a small part of the filter for analysis by LIBS imaging. The micro LIBS instrumental setup used in this study has also been previously described in detail elsewhere [27,28]. In short, a Q-switched Nd:YAG laser (Centurion GRM, Quantel) working at 1064 nm, with a pulse duration of 8 ns, 8 μm beam diameter and a repetition rate of 100 Hz, was used for plasma creation. The laser pulse was vertically focused onto the sample by a 15× magnification objective (LMM-15XeP01, Thorlabs) creating a typical crater size in the range of 6–7 μm. The instrument was equipped with a motorized translation 3-axis (XYZ) stage in order to generate the mapping. Plasma light was collected by a quartz lens and fiber bundle connected to a Czerny-Turner spectrometer (Shamrock 500, Andor Technology) assembled with an intensified charge-coupled device (ICCD) camera (iStar, Andor Technology). Under these conditions, we were able to acquire a spectral data cube of size 380 pixels × 126 pixels × 1845 wavelengths, i.e. 47,800 spectra acquired in just under 8 min in the 294.45–374.35 nm spectral domain.

3. Results and discussion

The objective of this first part is to evaluate the potential of our IFF algorithm against the state of the art in the field of chemometrics, i.e. the ESP algorithm, by considering an increasing level of complexity through a decrease of the signal to noise ratio on the simulated data set. It seemed interesting to start with the exploration of a trivial case, i.e. the data set of 230,230 spectra without added noise. Thus, the ESP approach was

applied on the data set considering 5 principal components to compress the initial data while 10,000 random vectors were considered for IFF. Not surprisingly, both approaches extracted only the 7 pure spectra. Nevertheless, we notice an interesting property of the ESP algorithm, already underlined in previous works [18,19], which despite a limited number of 5 principal components is able to select the 7 pure spectra. We will now see that things change radically as soon as noise is present. The ESP and IFF methods are then applied to the 2% noise data set under the same conditions as before. The 7 pure spectra are again well selected by both approaches in these new conditions but they are now included in a list of 294 spectra for ESP and 88 for IFF. In order to observe the distribution of the selected spectra within the data set, we represent them all within a PCA space in Figs. 2 and 3 for the ESP and IFF methods respectively. Thus, for each hyperplane defined by a combination of two principal components, the green dots represent the pure spectra selected by the considered algorithm. The red dots represent the other selected spectra also present in the list, while the blue dots represent all the other spectra in the data set. Beyond the number of selected spectra, which is very different between the two approaches, we notice in these two figures that the spectra selected by IFF are localized around the pure spectra while the selection is much more diffuse for ESP. We notice here an interesting property of IFF which, in spite of the absence of filtering (since it works on the raw data), finds the pure spectra present in the data set. If we now look at a noise level of 4%, things become much more complicated. Under these new conditions, the ESP approach selects 327 spectra, a list in which only 5 pure spectra are present (the first and the sixth being absent). On its side, the IFF method selects 186 spectra, the 7 pure spectra being included in this list. There is a clear increase in the number of spectra selected for the IFF approach compared to the 2% noise level. As before, these two selections are represented in a PCA space in Figs. 3S and 4S of supplementary material for the ESP and IFF approaches respectively. The color coding of the spectra is the same as before. Additionally, back dots represent the pure spectra present in the data set but not selected by the approach. We can make the same observation as before on the distribution of the spectra selected by the two approaches. Finally, if we now look at the highest noise level of 8%, the situation is naturally even more extreme. The ESP method selects 232 spectra in which only 4 pure spectra are present. The IFF method selects 788 spectra containing the 7 pure spectra. It is undeniable that the number of spectra selected by IFF increases a lot compared to the previous noise level. Nevertheless, this number must be relativized in the sense that it remains limited compared to the 232,230 spectra present in the data set but also because we have simulated here more than degraded measurement conditions. These two selections are also represented in a PCA space in Figs. 5S and 6S of supplementary material. At first sight, it seems surprising to see the difference between the IFF and ESP approaches in the face of an increasing noise level, even though the latter uses a prior PCA filtering step. We could then naturally say to ourselves that this number of 5 components used in the ESP approach is perhaps too limited. Table 1 shows the evolution of the spectra selection made by the ESP approach considering an increasing number of principal components on the same data set at 8% noise. We observe that the increase in the number of components allows us to go from 4 to 5 pure spectra detected without ever being able to detect all the pure ones in the data set even for a number of 9 principal components. The use of a larger number of components is also accompanied by a clear increase in the number of spectra selected and the calculation time. This first part, based on the exploitation of simulated spectral data, allowed us to show the relevance of the selection of spectra proposed by the IFF algorithm with a certain robustness against the noise very often observed in single shot LIBS analysis.

The second part of this study focusses on the exploration of the LIBS imaging data set acquired on the atmospheric aerosol sample. Before going into the actual application of the IFF approach on such a data set, it is interesting to go back to the exploration procedure that we usually use in spectroscopic imaging. In a first step, we usually calculate the mean spectrum from the 47,800 spectra of the data set (Fig. 4a). On this spectrum, we quickly identify the element Si which is mainly contained in the quartz filter but also other elements such as Na, Al, Ca, and Fe. Without a prior knowledge of the analyzed sample, the usual procedure would be to select a given wavelength from the observation of this spectrum in order to generate a distribution map of the supposed element. As a second step, we can also generate a global integration image by summing for each pixel all the intensities over the whole spectral domain (Fig. 4b). The color scale used here is logarithmic because there is a very large dynamic on the integrated values. This representation validates the particulate character of our sample but naturally without giving any chemical information since this information is lost during the integration of the signals. We then have two options for further exploration of the data set depending on whether or not we want to use chemometric tools. If not, we will generate classical distribution maps for the elements mentioned above by selecting specific wavelengths a priori. This task, which seems to be very simple, hides in fact a rather long procedure if one wishes to generate the least biased chemical maps. Indeed, for a given element, we must find a wavelength in the spectral range without interference, which is very likely in LIBS due to the richness of the spectra and the large number of elements present. If we want to have another point of view on these data, it is then advisable to use chemometric tools for a multivariate exploration that is to say exploiting simultaneously all the wavelengths. The first tool used in this case is often PCA, which is a simple, quick and quite effective approach. It is a multivariate factorial method based on the criterion of maximum expressed variance. It compresses the data for an easier observation of the spectral variations present in the data set. The application of this approach to our LIBS data set showed that 7 principal components were needed to describe the spectral variations it contains, expressing 97.93% of the total variance. Higher principal components where not considered since they only exhibit noise structures. This evaluation was done using the scree plot presented in the supplementary material (Fig. 7S). Fig. 5 shows these 7 principal components and the associated scores maps. As can be seen, the multivariate nature of PCA brings to light correlations and/or anti-correlations between elements at the spectral level, but also their relative contributions to the score maps at the spatial level. So for example, the red pixels of the score map 1 have relatively higher concentrations of the elements Si, Na, Fe, Al and Ca (which are correlated) than the blue pixels in this same map. Component 2 shows an anti-correlation between several elements. More precisely, the red pixels of the score map 2 have relatively more Na and Si and less Fe and Al than the blue ones and vice versa. Principal component 3 is interesting since it shows for the first time the presence of Cr correlated to Fe and Ca but also anti-correlated to Na and Al. It was not possible to see Cr in the mean spectrum. The principal component 4 shows above all a Si contribution, this time specific to the quartz filter as one can suppose in front of the absence of particulate structure on the associated score map. While principal component 5 contains elements already mentioned with of course different concentration ratios, components 6 and 7 allow in turn to detect the presence of Ti, not easily observable on the mean spectrum. As we can see, PCA offers an interesting exploration of the data set but the spatially unstructured nature of our sample (with no continuous region often observed on other samples) makes the comparison of score maps and pixel collocations more than delicate. Moreover, it focuses primarily on the relative variations at the spectral level without really highlighting the particular character of certain particles located on specific pixels. We then applied the IFF approach on the same data set considering 10,000 random vectors. As a result, 45 interesting pixels out of the 48,132 present in the data set have been selected. The spatial localization of these spectra and their selection frequencies are shown in Fig. 8S and Table 1S respectively in supplementary material. At the time, this result seemed quite surprising because it was quite unlikely to have 45 types of particles in such a sample. The correlation matrix between all these spectra was therefore calculated (Fig. 9S in supplementary material). Not surprisingly, many correlations greater

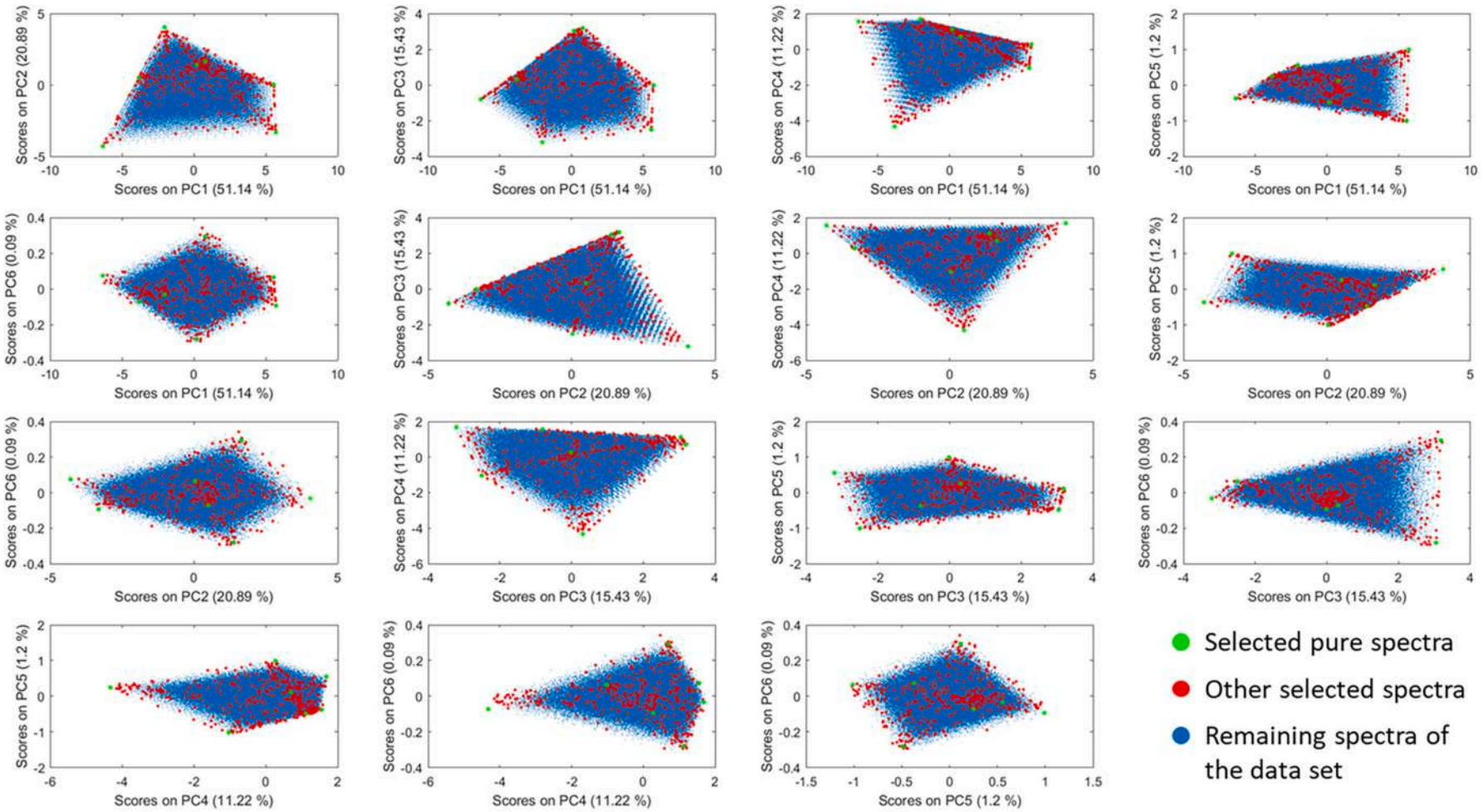

Fig. 2. Representation of the pixels selected by ESP vs the rest of the data set in a PCA space considering a 2% noise level.

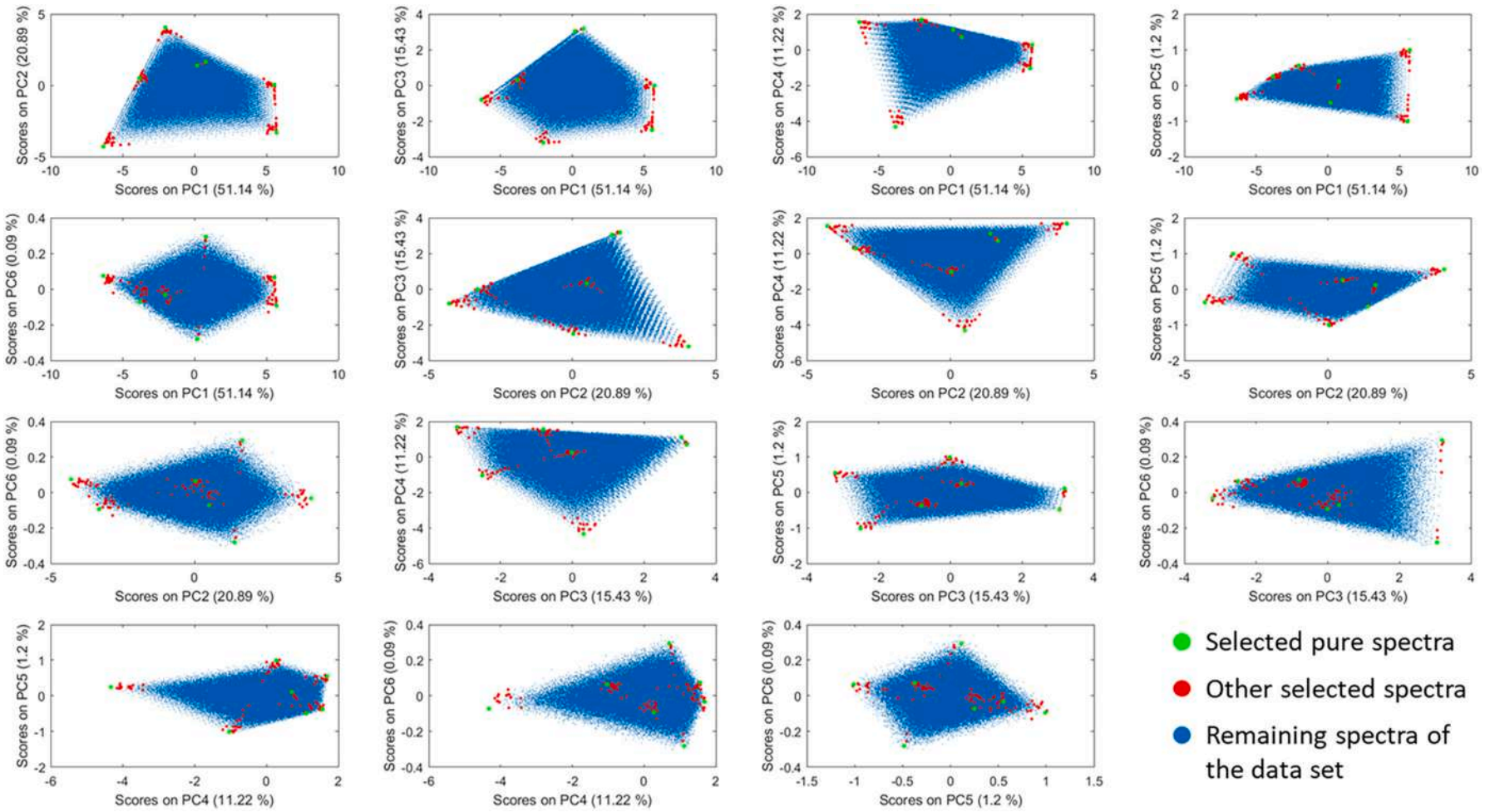

Fig. 3. Representation of the pixels selected by IFF vs the rest of the data set in a PCA space considering a 2% noise level.

Table 1
Evolution of the spectra selection made by the ESP approach considering an increasing number of principal components on the same data set at 8% noise.

Number of PCs used in the ESP approach	Number of selected spectra	Pure spectra included in the selection (/ 7)	Computation time (s)
5	232	4	1
6	707	5	10
7	1468	5	21
8	2764	5	62
9	4708	5	650

than 0.96 are observed. For example, spectrum 1 is very similar to the spectra 4, 29, 39 and 41 with correlations of 0.96, 0.99, 0.99 and 0.96 respectively. This is due to the fact that there are very small variations in elemental concentration between these spectra with, to a lesser extent, the noise that also participates in the phenomenon. As a consequence, a correlation higher than 0.96 allowed us to group the spectra into 16 distinct families obtained using the mean spectrum of each group. We should not misunderstand here the efficiency of the IFF algorithm which accounts for all types of variations from the largest to the most subtle. The 45 spectra are thus well located on the convex hull but from an analytical point of view, there are in fact only 16 typical spectra represented in Fig. 6. First of all, we find in these spectra the contributions of the elements detected during the PCA but we now also have specific elemental concentration ratios for each of them. We recall here that by definition, none of these spectra can be found by linear combinations of the others, which allows us to insist on their representativeness within the data set. On the other hand, all the spectra in the data set are potentially linear combinations of these 16 representative spectra. The most interesting in this selection is certainly the detection of Cu and Ni present in the contributions 5, 14 and 16 respectively. It was not possible to observe these elements either in the mean spectrum or even with PCA despite its recognized sensitivity. For information, the application of the ESP approach considering 6 principal components on this same data set, generates a list of 133 spectra containing all the types of particles presented in Fig. 6 except the three previously mentioned. The detection of Cu and Ni is therefore not possible in this case. Moreover, an increase in the number of components does not change the situation. This work has shown that the IFF algorithm has interesting capabilities to identify compounds present in a complex sample independently of their concentration while being relatively robust to the noise often present on the spectra. It is obvious that such an approach complements the range of chemometric tools such as PCA by offering a completely different view of the spectral data.

4. Conclusion

The detection of minor and trace compounds is always a delicate task in analytical chemistry. It is even more challenging today in LIBS imaging as we generate ever larger spectral data sets. As we have seen in this work, classical chemometrics tools are not always optimal to extract such information. It is in this sense that we have introduced the IFF approach in order to extract these compounds of interest independently of the variance they may express in the data set. The originality of the algorithm concerns the direct exploitation of the data (i.e. without prior compression) in order to extract the convex envelope. By means of a simulated dataset and a LIBS imaging one acquired on atmospheric aerosols, we have highlighted the sensitivity of our approach but also its potential robustness against noise, which is particularly important in the context of spectroscopic imaging. Two questions could then legitimately be asked: when and how to use such a tool? Many researchers who are new to chemometrics think that there must be an ideal tool to exhaustively extract all the information contained in a data set. This is not the case and we must instead use different tools, each with its own advantages and disadvantages, in order to obtain the least biased vision of the chemical system we are exploring. In other words, we add the IFF approach to the chemometric toolbox next to PCA or many others. If we are not familiar with the concepts of chemometrics, we can use the IFF approach without much thought. It will then present us with a selection of spectra that we will be able to observe and from which we will most probably be able to observe unsuspected emission lines since they are not visible on the mean spectrum. We can then generate maps by integrating the signal at these wavelengths. If, on the contrary, we want to go further in chemometric exploration, we could use this selection of spectra to generate initial estimates of the pure compounds present in the explored sample in order to do a signal unmixing as for example with MCR-ALS but this is beyond the scope of this work [15].

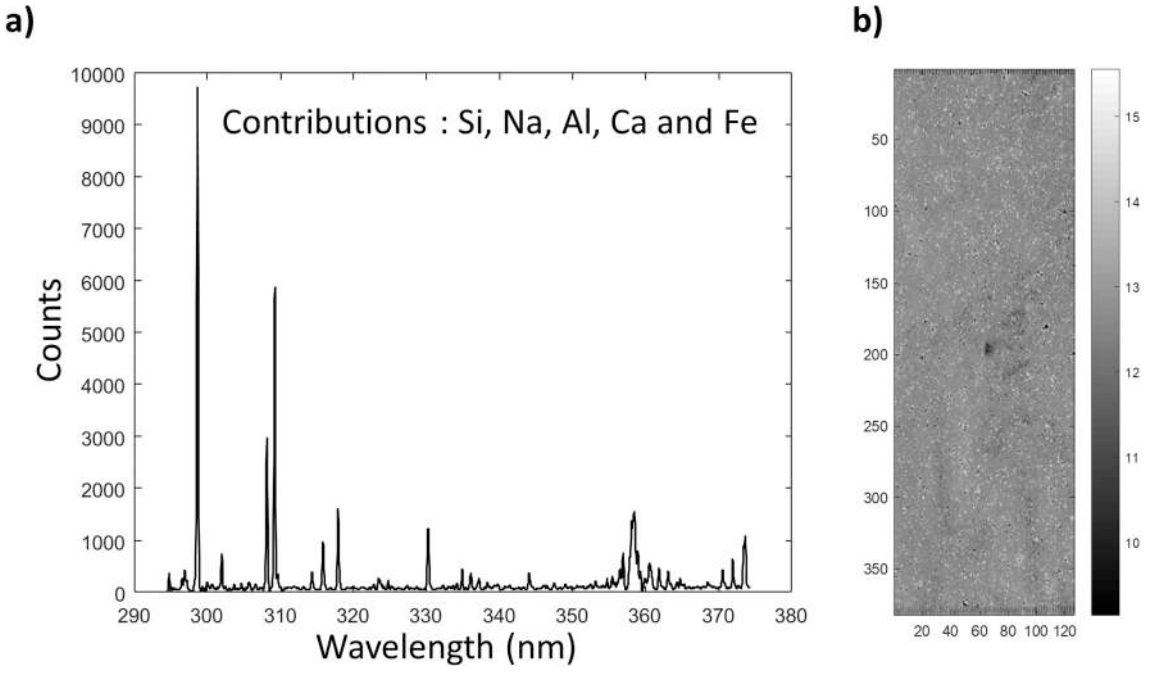

Fig. 4. a) The mean spectrum of the data set. b) The global integration image.

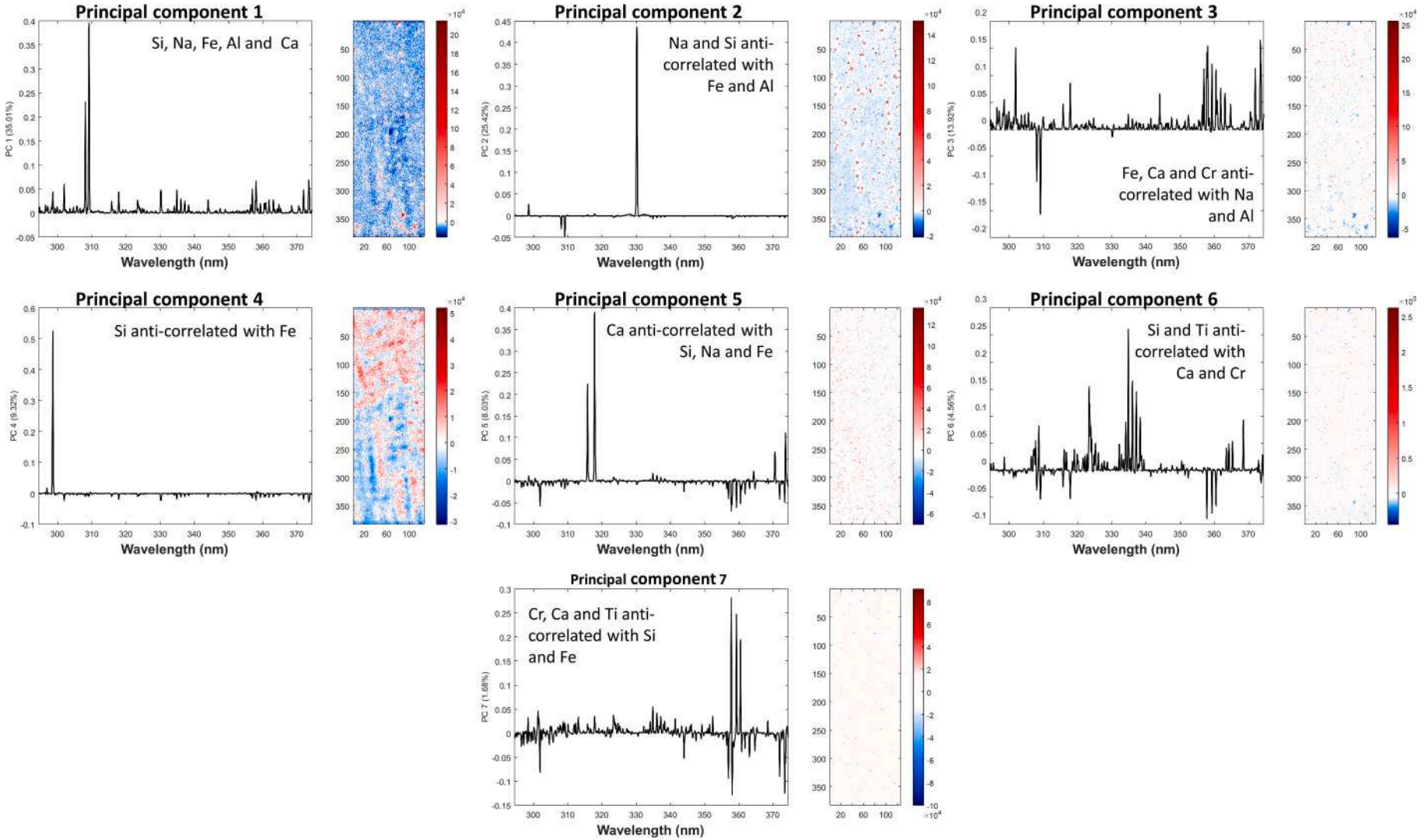

Fig. 5. Principal component analysis (PCA) of the LIBS imaging data set.

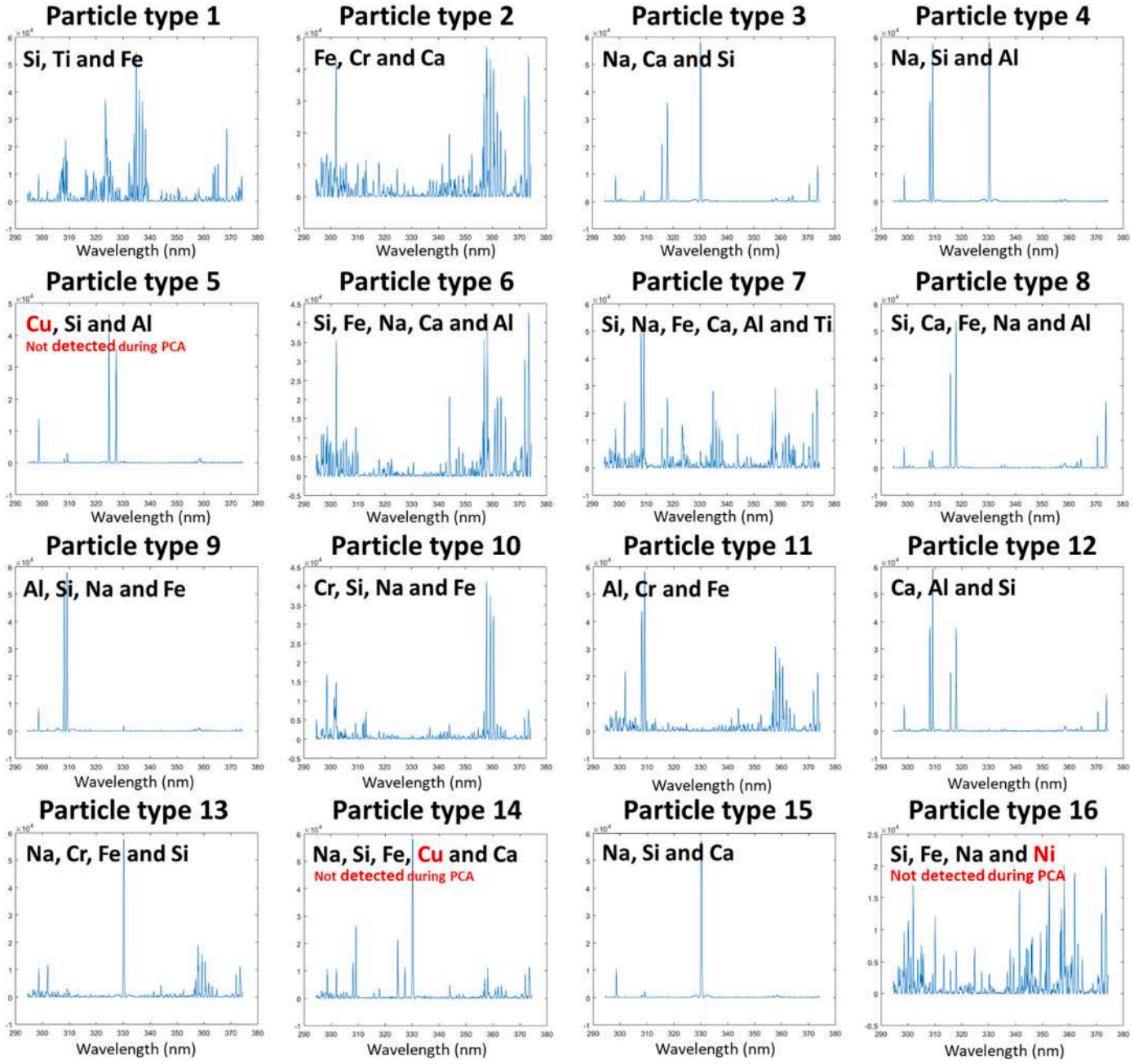

Fig. 6. Representation of the final list of the 16 typical spectra.

Author statement

All authors participated equally in this work.

Aknowledgement

This work was partially supported by the French region Rhônes Alpes Auvergne (Optolyse, CPER2016) and the French National Research Agency (ANR- Diagem project).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sab.2022.106508.

References

[1] D. Duewer, Bruce R. Kowalski, 7 Mar 1942 to 1 Dec 2012, Aust. J. Chem. 28 (2014) (7 Mar 1942) 319–320, https://doi.org/10.1002/cem.2622.
[2] I. Ben-Gera, K.H. Norris, Determination of moisture content in soybeans by direct spectrophotometry, Israel, J. Agric. Res. 18 (1968) 125–132.
[3] I. Ben-Gera, K.H. Norris, Direct spectrophotometric determination of fat and moisture in meat products, J. Food Sci. 33 (1968) 64–67.

[4] M. Baudelet, B.W. Smith, The first years of laser-induced breakdown spectroscopy, J. Anal. At. Spectrom. 28 (2013) 624–629, https://doi.org/10.1039/C3JA50027F.
[5] P. Fichet, J.-L. Lacour, D. Menut, P. Mauchien, A. Rivoallan, C. Fabre, J. Dubessy, M.-C. Boiron, Micro LIBS technique, in: A.W. Miziolek, I. Schechter, V. Palleschi (Eds.), Laser Induced Breakdown Spectroscopy, Cambridge University Press, Cambridge, 2006, pp. 539–555, https://doi.org/10.1017/CBO9780511541261.017.
[6] H. Bette, R. Noll, High speed laser-induced breakdown spectrometry for scanning microanalysis, J. Phys. D. Appl. Phys. 37 (2004) 1281–1288, https://doi.org/10.1088/0022-3727/37/8/018.
[7] D. Menut, P. Fichet, J.-L. Lacour, A. Rivoallan, P. Mauchien, Micro-laser-induced breakdown spectroscopy technique: a powerful method for performing quantitative surface mapping on conductive and nonconductive samples, Appl. Opt. 42 (2003) 6063, https://doi.org/10.1364/AO.42.006063.
[8] J.M. Vadillo, J.J. Laserna, Laser-induced plasma spectrometry: truly a surface analytical tool, Spectrochim. Acta B At. Spectrosc. 59 (2004) 147–161, https://doi.org/10.1016/j.sab.2003.11.006.
[9] V. Motto-Ros, L. Sancey, X.C. Wang, Q.L. Ma, F. Lux, X.S. Bai, G. Panczer, O. Tillement, J. Yu, Mapping nanoparticles injected into a biological tissue using laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 87 (2013) 168–174, https://doi.org/10.1016/j.sab.2013.05.020.
[10] Y. Gimenez, B. Busser, F. Trichard, A. Kulesza, J.M. Laurent, V. Zaun, F. Lux, J. M. Benoit, G. Panczer, P. Dugourd, O. Tillement, F. Pelascini, L. Sancey, V. Motto-Ros, 3D imaging of nanoparticle distribution in biological tissue by laser-induced breakdown spectroscopy, Sci. Rep. 6 (2016) 29936, https://doi.org/10.1038/srep29936.
[11] X. Le Guével, M. Henry, V. Motto-Ros, E. Longo, M.I. Montañez, F. Pelascini, O. de La Rochefoucauld, P. Zeitoun, J.-L. Coll, V. Josserand, L. Sancey, Elemental and optical imaging evaluation of zwitterionic gold nanoclusters in glioblastoma mouse models, Nanoscale. 10 (2018) 18657–18664, https://doi.org/10.1039/C8NR05299A.
[12] S. Moncayo, L. Duponchel, N. Mousavipak, G. Panczer, F. Trichard, B. Bousquet, F. Pelascini, V. Motto-Ros, Exploration of megapixel hyperspectral LIBS images using principal component analysis, J. Anal. At. Spectrom. 33 (2018) 210–220, https://doi.org/10.1039/C7JA00398F.
[13] J. Vrábel, E. Képeš, L. Duponchel, V. Motto-Ros, C. Fabre, S. Connemann, F. Schreckenberg, P. Prasse, D. Riebe, R. Junjuri, M.K. Gundawar, X. Tan, P. Pořízka, J. Kaiser, Classification of challenging laser-induced breakdown spectroscopy soil sample data - EMSLIBS contest, Spectrochim. Acta B At. Spectrosc. 169 (2020), 105872, https://doi.org/10.1016/j.sab.2020.105872.
[14] A. Nardecchia, C. Fabre, J. Cauzid, F. Pelascini, V. Motto-Ros, L. Duponchel, Detection of minor compounds in complex mineral samples from millions of spectra: a new data analysis strategy in LIBS imaging, Anal. Chim. Acta 1114 (2020) 66–73, https://doi.org/10.1016/j.aca.2020.04.005.
[15] A. Nardecchia, A. de Juan, V. Motto-Ros, M. Gaft, L. Duponchel, Data fusion of LIBS and PIL hyperspectral imaging: understanding the luminescence phenomenon of a complex mineral sample, Anal. Chim. Acta 1192 (2022), 339368, https://doi.org/10.1016/j.aca.2021.339368.
[16] A. Nardecchia, A. de Juan, V. Motto-Ros, C. Fabre, L. Duponchel, LIBS and Raman data fusion: an original approach based on the use of chemometric methodologies, Spectrochim. Acta B At. Spectrosc. (2022). Submitted.
[17] A. Limbeck, L. Brunnbauer, H. Lohninger, P. Pořízka, P. Modlitbová, J. Kaiser, P. Janovszky, A. Kéri, G. Galbács, Methodology and applications of elemental mapping by laser induced breakdown spectroscopy, Anal. Chim. Acta 1147 (2021) 72–98, https://doi.org/10.1016/j.aca.2020.12.054.
[18] M. Ghaffari, N. Omidikia, C. Ruckebusch, Essential spectral pixels for multivariate curve resolution of chemical images, Anal. Chem. 91 (2019) 10943–10948, https://doi.org/10.1021/acs.analchem.9b02890.
[19] L. Coic, P.-Y. Sacré, A. Dispas, C. De Bleye, M. Fillet, C. Ruckebusch, P. Hubert, É. Ziemons, Selection of essential spectra to improve the multivariate curve resolution of minor compounds in complex pharmaceutical formulations, Anal. Chim. Acta 1198 (2022), 339532, https://doi.org/10.1016/j.aca.2022.339532.
[20] V.V. Prasolov, Polynomials, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, https://doi.org/10.1007/978-3-642-03980-5.
[21] P.J. Rousseeuw, I. Ruts, J.W. Tukey, The Bagplot: A Bivariate Boxplot, 2022, p. 7.
[22] P. Nicola, General Competitive Equilibrium, in: Mainstream Mathematical Economics in the 20th Century, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000, pp. 197–215, https://doi.org/10.1007/978-3-662-04238-0_16.
[23] E.B. Nilsen, S. Pedersen, J.D.C. Linnell, Can minimum convex polygon home ranges be used to draw biologically meaningful conclusions? Ecol. Res. 23 (2008) 635–639, https://doi.org/10.1007/s11284-007-0421-9.
[24] E. Rieffel, W. Polak, Quantum Computing: A Gentle Introduction, The MIT Press, Cambridge, Mass, 2011.
[25] C.B. Barber, D.P. Dobkin, H. Huhdanpaa, The quickhull algorithm for convex hulls, ACM Trans. Math. Softw. 22 (1996) 469–483, https://doi.org/10.1145/235815.235821.
[26] C. Marina-Montes, V. Motto-Ros, L.V. Pérez-Arribas, J. Anzano, M. Millán-Martínez, J.O. Cáceres, Aerosol analysis by micro laser-induced breakdown spectroscopy: a new protocol for particulate matter characterization in filters, Anal. Chim. Acta 1181 (2021), 338947, https://doi.org/10.1016/j.aca.2021.338947.
[27] J.O. Cáceres, F. Pelascini, V. Motto-Ros, S. Moncayo, F. Trichard, G. Panczer, A. Marín-Roldán, J.A. Cruz, I. Coronado, J. Martín-Chivelet, Megapixel multi-elemental imaging by laser-induced breakdown spectroscopy, a technology with considerable potential for paleoclimate studies, Sci. Rep. 7 (2017) 1–11, https://doi.org/10.1038/s41598-017-05437-3.
[28] L. Jolivet, V. Motto-Ros, L. Sorbier, T. Sozinho, C.-P. Lienemann, Quantitative imaging of carbon in heterogeneous refining catalysts, J. Anal. At. Spectrom. 35 (2020) 896–903, https://doi.org/10.1039/C9JA00434C.