Data fusion of LIBS and PIL hyperspectral imaging: Understanding the luminescence phenomenon of a complex mineral sample

Alessandro Nardecchia, Anna de Juan, Vincent Motto-Ros, Michael Gaft, Ludovic Duponchel

Data fusion of LIBS and PIL hyperspectral imaging: understanding the luminescence phenomenon of a complex mineral sample

Alessandro Nardecchia[1], Anna de Juan[2], Vincent Motto-Ros[3], Michael Gaft[4], Ludovic Duponchel[*,1]

1: Univ. Lille, CNRS, UMR 8516 – LASIRE – Laboratoire de Spectroscopie pour Les Interactions, La Réactivité et L'Environnement, Lille, F-59000, France

2: Chemometrics Group. Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona, Diagonal 645, 08028, Barcelona, Spain

3 : Institut Lumière Matière, UMR 5306, Université Lyon 1 – CNRS, Université de Lyon, Villeurbanne, 69622, France

4: Ariel University, Department of Physics, Ariel 40700, Israel

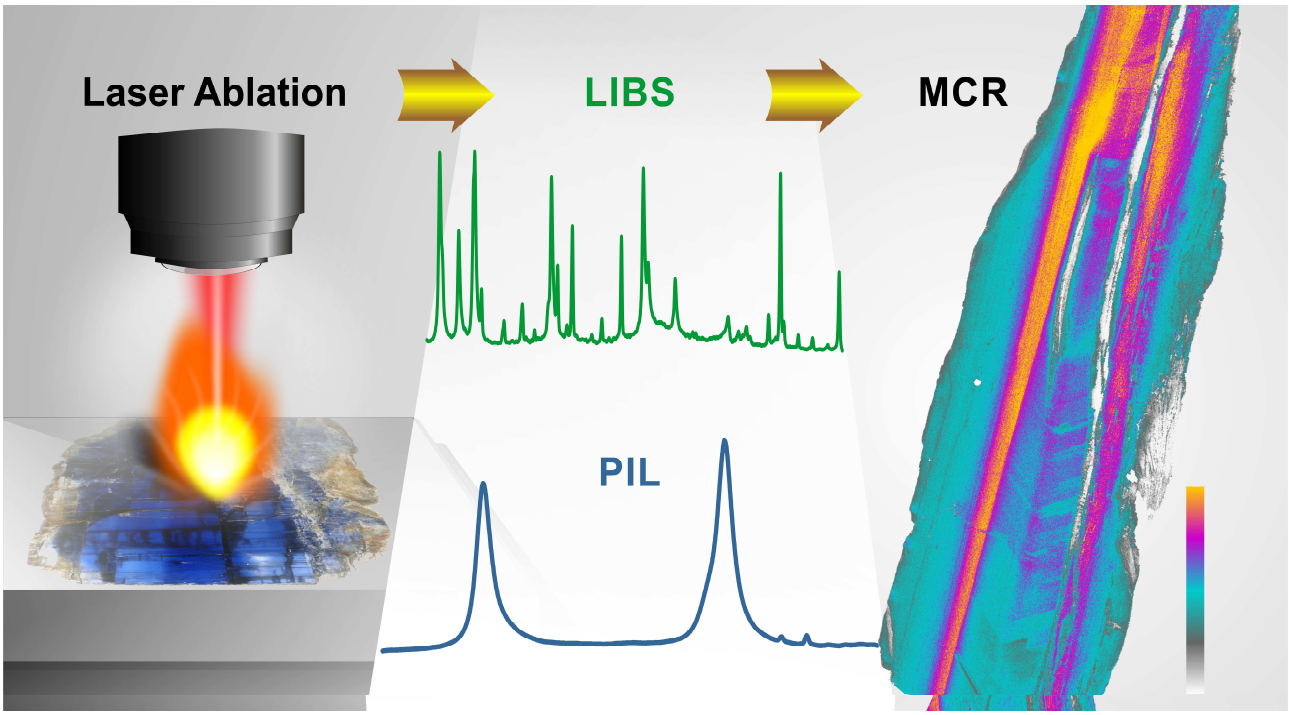
(*) Corresponding Author Email**: ludovic.duponchel@univ-lille.fr**

Keywords: Hyperspectral imaging; LIBS; PIL; Data fusion; chemometrics

ABSTRACT: Laser-induced breakdown spectroscopy (LIBS) imaging is an innovative technique that associates the valuable atomic, ionic and molecular emission signals of the parent spectroscopy with spatial information. LIBS works using a powerful pulse laser as excitation source, to generate a plasma exhibiting emission lines of atoms, ions and molecules present in the ablated matter. The advantages of LIBS imaging are potential high sensitivity (in the order of ppm), easy sample preparation, fast acquisition rate (up to 1 kHz) and µm scale spatial resolution (weight of the ablated material in the order of ng). Despite these positive aspects, LIBS imaging easily provides datasets consisting of several million spectra, each containing several thousand spectral channels. Under these conditions, the current chemometric analyses of the raw data are still possible, but require too high computing resources. Therefore, the aim of this work is to propose a data compression strategy oriented to keep the most relevant spectral channel and pixel information to facilitate, fast and reliable signal unmixing for an exhaustive exploration of complex samples. This strategy will apply not only to the context of LIBS image analysis, but to the fusion of LIBS

with other imaging technologies, a scenario where the data compression step becomes even more mandatory. The data fusion strategy will be applied to the analysis of a heterogeneous kyanite mineral sample containing several trace elements by LIBS imaging associated with plasma induced luminescence (PIL) imaging, these two signals being acquired simultaneously by the same microscope. The association of compression and spectral data fusion will allow extracting the compounds in the mineral sample associated with a fused LIBS/PIL fingerprint. This LIBS/PIL association will be essential to interpret the PIL spectral information, which is nowadays very complex due to the natural overlapped signals provided by this technique.

GRAPHICAL ABSTRACT



INTRODUCTION

Laser-induced breakdown spectroscopy (LIBS) imaging is nowadays a very powerful technique for the elemental analysis of complex samples used in many different scientific fields [1–7]. This technique uses a pulse laser beam focused on the sample surface to generate a plasma that atomizes and excites the ablated matter. As a consequence, the excited atoms, ions and molecules release the excess of energy with electronic relaxations, and a characteristic emission spectrum for each element present in the matrix

can be acquired using an optical microscope coupled with a spectrometer. In LIBS imaging, the sample surface is usually explored in a scanning configuration mode, acquiring one spectrum at a time for each spatial position of a predefined grid. Then, using a classical integration of the acquired signal at a particular wavelength (i.e. an emission line of a given element), it is possible to generate a distribution image of the considered element present in the sample. LIBS technique shows many advantages, such as multi-elemental capabilities including light elements (<Mg), a high acquisition rate (up to 1000 spectra/s), high sensitivity most of the time, high dynamic range (major elements to traces can be observed), and compatibility with optical microscopy. Nevertheless, even if the high acquisition rate of LIBS imaging allows analyzing large sample areas of several cm^2 in a very reasonable time, this advantage becomes a major limitation because a huge amount of data is naturally produced due to both the many spectral channels explored by LIBS and the massive number of sampling points – the pixels – scanned. In fact, it is nowadays common to get images with millions of pixels associated with thousands of spectral channels [8,9]. Another important aspect in the LIBS exploration of a sample is the possibility to obtain an additional plasma induced luminescence (PIL) [10] response using the same instrument. Indeed, the plasma generated by the LIBS laser shot acts as an excitation source and produces the emission of a luminescence response for specific elements present on the sample surface [11]. Nevertheless, despite the relative simplicity of acquiring these additional PIL spectra, the interpretation of such signals remains uncertain [12]. Chemometrics and multivariate data analysis are very suitable approaches for the exploration of this complex kind of imaging datasets. However, the use of these tools for the study of LIBS and/or PIL images is nowadays still limited. Understanding the concept of hyperspectral imaging, finding appropriate tools for data exploration to deal with millions of spectra and able to provide interpretable outputs is still a very complex task, which can be of invaluable help for the LIBS community members.

The central point of this work is to provide a data analysis pipeline capable to drastically decrease the amount of imaging data (both the spectral channels and the pixels) used for the investigation of a complex and heterogeneous sample in order to perform a simpler unmixing analysis of the essential information selected for LIBS images or for fused LIBS/PIL data configurations. To do the unmixing task, Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) analysis [13–15] will be applied on the selected small amount of data coming from the previous compression step. We will demonstrate that applying MCR-ALS on such compressed dataset is sufficient to reconstruct high quality full maps and spectral signatures of the compounds in the imaged sample without losing the initial spectral and spatial resolution [16]. The methodology proposed will be tested to study a heterogeneous kyanite mineral sample containing several trace elements analyzed by LIBS and PIL imaging. The results of the analysis of fused

LIBS and PIL datasets will provide the identification and distribution of the different elements present in the sample but, most importantly, will shed light for a better understanding of the luminescence phenomenon in this kind of complex samples. To the best of our knowledge, this is the first time that this data analysis pipeline (data compression and fusion) is used on LIBS/PIL imaging platforms.

MATERIAL AND METHODS

Experimental setup and spectral data acquisition

The LIBS experimental setup has been already described elsewhere [3,11,17,18]. It included a Nd:YAG laser source operating at 100 Hz and emitting at the fundamental wavelength (i.e. 1064 nm) with an 8 ns pulse duration (Centurion, Quantel laser by Lumibird). The laser beam was focused onto the sample using a 15× magnification objective as shown in Figure 1a. All the measurements were conducted in ambient atmosphere with an argon flow of 0.8 l/min acting on the plasma region. A laser line scanning was performed in raster scan mode with the use of a motorized XYZ stage. In this configuration and considering the laser frequency rate, about 360,000 laser shots were produced in 1 hour. The ablation craters, observed afterward with optical microscopy, were less than 8 μm in diameter. Two spectrometers (Shamrock 500 and Shamrock 303, Andor Technology) equipped with intensified charge-coupled device (ICCD) cameras (iStar, Andor Technology) were used to probe simultaneously two spectral ranges in different temporal domains. The Shamrock 500 was used for LIBS experiments and was equipped with a 2400 l/mm grating (Holographic, peak at 220 nm) covering the 425-440 nm spectral range with a resolution of ~0.04 nm. This range was selected to detect primarily iron (Fe), chromium (Cr), vanadium (V) and titanium (Ti), although calcium (Ca) and zirconium (Zr) lines could also be detected (as can be seen on the mean LIBS spectrum in Figure 1b). The Shamrock 303 was used for PIL experiments. It was equipped with a 1200 l/mm grating and setup in the 680-720 nm spectral range, where intense luminescence lines were detected. The mean PIL spectrum is presented in Figure 1c. Both ICCD cameras were synchronized to the Q-switch of the laser. The LIBS acquisition was performed with a delay of 1200 ns and a gate of 4000 ns, while the PIL acquisition was performed with a delay of 100 μs and a gate of 2.2 ms. The light emitted by the plasma was collected by two quartz lenses and focused onto the entrance of round-to-linear fiber bundles connected to each spectrometer. Each fiber bundle was formed by 19 fibers, each with a 200 μm core diameter. Spectra were acquired in full vertical binning mode for the two spectrometers. The laser energy was stabilized throughout the experiment and was fixed to 1 mJ per pulse. Finally, a home-made software developed under the LabVIEW environment was used to control the entire setup, allowing automatic sequences of any selected regions of interest with a preset lateral resolution.

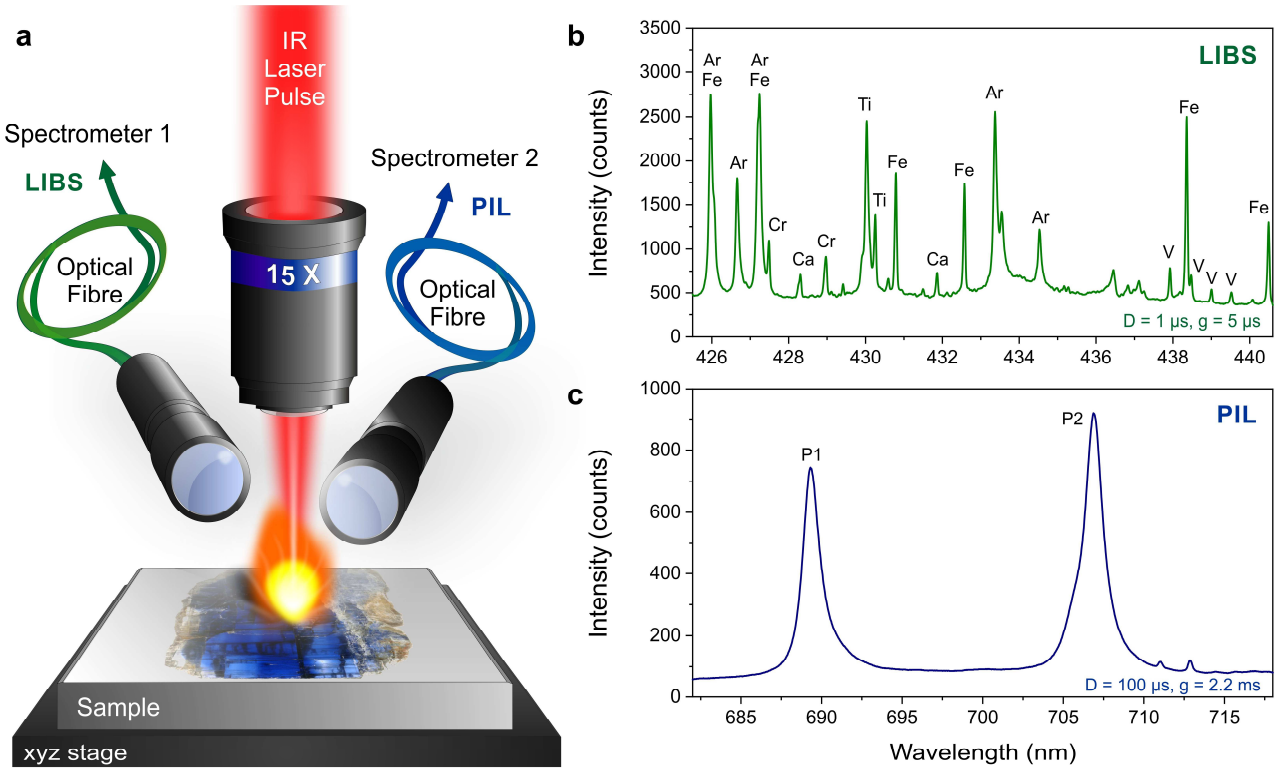


Figure 1. a) Experimental setup. b) Mean LIBS spectrum of the considered sample. c) Mean PIL spectrum of the considered sample.

Sample and dataset description

The sample selected for this study is a section of blue kyanite cristal (also called disten or cyanite) approximately 3 by 1.5 centimeters in size, which isa low temperature - high pressure metamorphic phase mainly formed by Al_2SiO_5 with many heterogeneities and several trace elements (mainly iron, calcium, vanadium, titanium and chromium), collected in Siberia. For LIBS and PIL imaging, the characterized cross-section was embedded in epoxy resin, cut, and finally polished with SiC paper under water to obtain a clean flat surface ready to be scan. The LIBS and PIL images acquired from this sample are sized each 1100 x 2000 pixels (i.e. a total of 2,200,000 spectra) x 2048 spectral channels with a spatial resolution of 20 µm per pixel. The size occupied in terms of storage by the two datasets is equal to 8 gigabytes for the LIBS dataset and more than 6 gigabytes for the PIL one.

A first idea of the chemical information related to this sample can be obtained observing the mean spectra of the two datasets using LIBS and PIL (Figure 1b and 1c respectively). Thus the observation of the mean PIL spectrum shows a first line near 706 nm that looks not symmetric and evidently has a shoulder at its left side. In fact, under UV excitation, the literature indicates that two close lines with similar intensities can be observed depending on the orientation of the sample at 706.2 and 704.6 nm respectively. This

statement is quite surprising in our case because the line at 704.6 nm is supposed to have a much shorter decay than the one at 706.2 nm (75 µs and 1.2 ms, respectively) and therefore our acquisition parameters should not allow us to see the former. On the other hand, the line near 689 nm seems to be symmetrical and also corresponds to UV excitation. Once again, the literature indicates the presence of two lines this time even closer at 688.9 and 690.1 nm, the second being of very low intensity which finally explains the observation of a single line at first sight. Comparing the two mean spectra, it is clear that LIBS is represented by a larger number of emission signals compared to PIL, which is mainly formed by two broad spectral contributions. Figure 2 shows distribution images of elements present in the sample obtained from the signal integration method classically used on specific wavelengths of the LIBS spectra. Thus, even if this approach takes only into account individual wavelengths and not the full spectra, the spatial distribution of the elements presents some interesting aspects. For instance, while elements Cr and V seem to be strongly correlated and distributed in a very large area, Fe, Ca and Ti are more specific to small zones of the mineral. Note that no distribution image of Al and Si are proposed, although kyanite is an aluminosilicate because this element does not present LIBS emission in the considered spectral range. The two PIL images generated from the two spectral contributions at 689.315 (P1) and 706.865 nm (P2) also show slight spatial locations. However, from a general point of view, it can be seen that it is rather difficult to find spatial correlations between all the LIBS and PIL images by simple visual inspection. The only thing clearly seen is that the luminescence at 689 and 706 nm does not come from the Ti element since we observe some inverse correlation among LIBS and PIL images. Even if the generation of these integration images remains an easy first step to observe the chemical contributions of the sample, it is obvious that the characterization of the correlations between images remains delicate. Unfortunately, this is not the only constraint since the information of elements that can coexist to form mineral phases is lost or, in the best of the cases, incomplete using this univariate approach. Similarly, minor compounds associated with weak signal and small localized areas may be missed. For all these reasons, we propose in this work a multivariate data processing pipeline combining compression, fusion, and signal unmixing for an exhaustive and simultaneous exploration of LIBS and PIL spectroscopies.

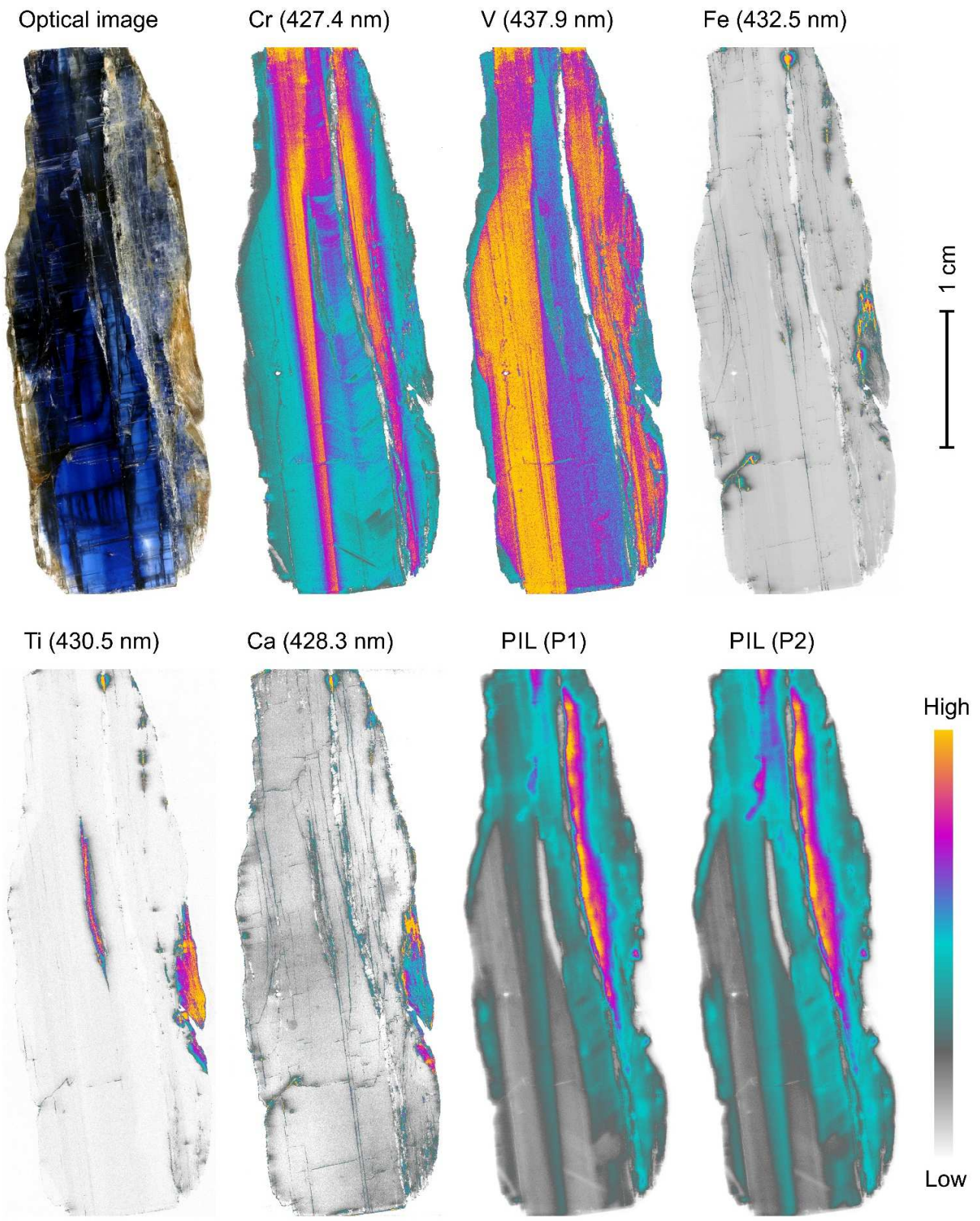


Figure 2. Distribution images of elements generated from the classical signal integration method from LIBS and PIL spectra.

DATA TREATMENT

Data compression and signal unmixing

The massive nature of the datasets to be analyzed demands a mandatory step of data compression to save computational resources and analysis time. The whole process of compression and multivariate resolution (signal unmixing) is described below in several successive steps. For convenience, this section of the text will illustrate the application of the methodology to the LIBS dataset, due to the complexity and rich information provided by this measurement. However, the same procedure was applied to both LIBS and PIL datasets, except for some specificities, that will be described, discussed and shown in Figure 3.

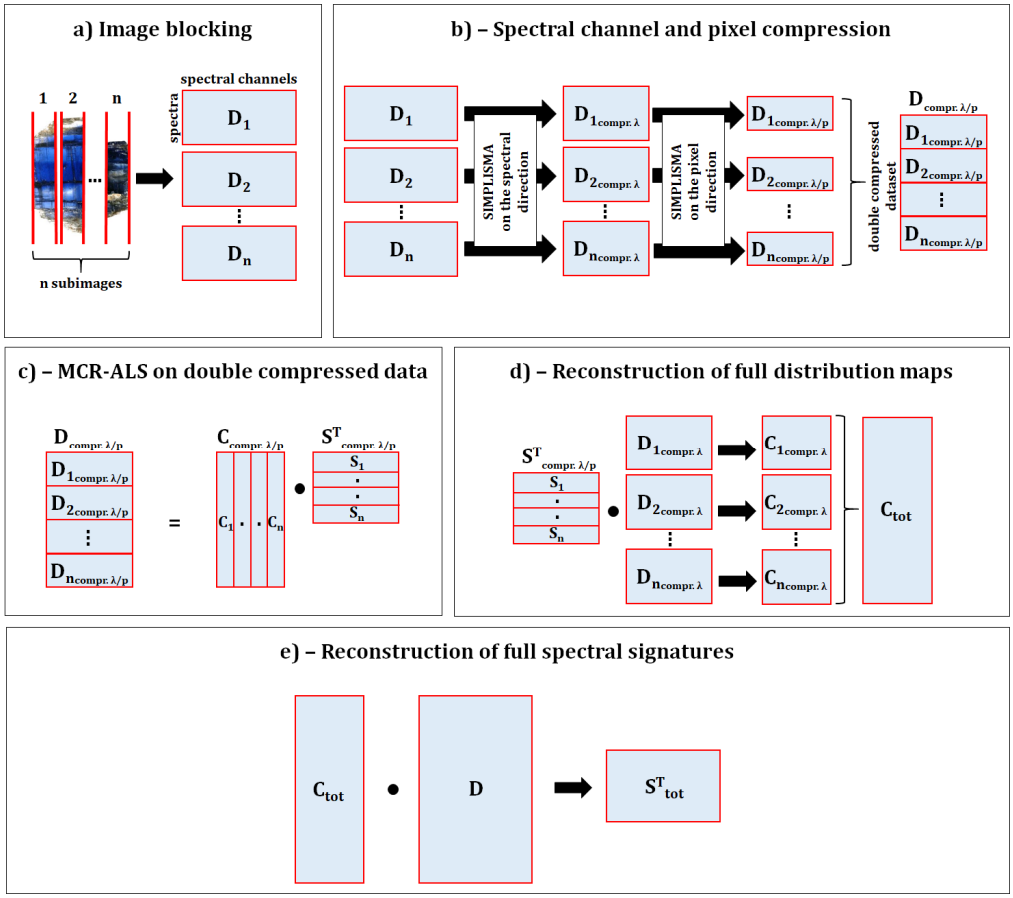


Figure 3. The data fusion and signal unmixing procedure.

Thus the proposed data processing pipeline is divided into the following steps:

1) *Image blocking*: first of all, the whole dataset was divided into 16 subimages. In this way, every single subimage had a reduced size of 780 x 125 pixels and provides an unfolded submatrix $\mathbf{D_i}$ of 97500 spectra. Blocking allows performing data analysis tasks much faster in each $\mathbf{D_i}$ subset of

97500 spectra than if the work was carried out in the initial **D** matrix of 2 million spectra of the full image. Each subsequent step described was applied separately to each of the $\mathbf{D_i}$ submatrices (Figure 3a).

2) *Spectral and spatial preprocessing*: heterogeneous samples analyzed by LIBS imaging often contain saturated spectra in the acquired datasets. Recent works have shown the importance of correcting these signals prior to any data processing by proposing effective but relatively complex correction strategies [19,20]. In this work, we have proposed a simple procedure for managing the saturated spectra. First, the use of a threshold on the spectra (set by the maximum A/D converter dynamic range) was used to identify and locate the saturated pixels in the image spectra. Second, each saturated spectrum was replaced by the average of the non-saturated neighboring pixels in the image. For the LIBS dataset, a baseline correction was afterwards applied using the asymmetric least squares (AsLS) algorithm based on the Whittaker smoother [21], with $\lambda = 104$ and an asymmetry parameter of 0.0003. For the PIL dataset, which had a radically lower signal-to-noise ratio, it was first necessary to apply a Savitzky-Golay smoothing (filter width = 20; polynomial order = 2) [22], and then, again, the AsLS algorithm ($\lambda = 10^7$ and asymmetry parameter of 0.0003) for baseline correction. Finally, the use of a threshold on the global spectral intensity of the pixels helped to generate a mask to separate the spectra from the mineral sample, used for further analysis, from the spectra of the epoxy resin surrounding it, discarded in all the following data treatments steps. This image cropping step reduced the total amount of spectra to be analyzed from more than 2 million to around 1 million. The comparison between the starting raw spectra, and the ones with this first spectral and spatial pretreatment are represented in the supplementary material (Figure S1).

3) *Double data compression*: this is a key central step in all the proposed chemometric strategy. In fact, despite the already performed signal corrections and the massive reduction of spectra, the dimensions of the datasets (both LIBS and PIL) were still very huge. In chemometrics, and particularly in the signal unmixing framework, it is common to use methods oriented to the purest selection of variables (understood as pixels or spectral channels) to generate initial estimates for iterative unmixing methods, such as MCR-ALS. In the present work, a SIMPLe-to-use Interactive Self-modeling Mixture Analysis (SIMPLISMA) [23–26] based-method was first used to select only the purest image information and drastically compress the number of selected pixels and spectral channels prior to the final MCR-ALS analysis. For each $\mathbf{D_i}$ image submatrix in Figure 3b, taking advantage of the fine spectral features of LIBS, the selection of purest information was applied first on the spectral channels. In this way, only

the most important spectral variables (i.e. wavelengths) related to different chemical elements in the mineral were selected, discarding the redundant information. It is also important to underline the fact that at this stage of the procedure, as the final number of purest spectral variables needed is unknown, it was decided to overestimate this value. Thus, each $\mathbf{D_i}$ image submatrix was first divided into blocks of 200 spectra from which the first 20 purest spectral variables were selected. At the end, a small list of spectral channels considering all the selected channels in the 16 $\mathbf{D_i}$ submatrices was used to generate the spectral-compressed $\mathbf{D_{i,compr.\lambda}}$ image submatrices, which had a much lower number of spectral channels than the original $\mathbf{D_i}$ blocks, but had all image pixels. In a next step, the same approach based on SIMPLISMA was used in each of the spectral-compressed $\mathbf{D_{i,compr.\lambda}}$ image submatrices to select the purest pixels. So, each of the $\mathbf{D_{i,compr.\lambda}}$ image submatrices was divided into blocks of 500 spectra, and the first 40 purest pixels of each block were selected, providing a spectral- and pixel-compressed $\mathbf{D_{i,compr.\lambda/p}}$ submatrix with a much lower number of pixels and spectral channels than the initial $\mathbf{D_i}$ related block. The extracted information of all the $\mathbf{D_{i,compr.\lambda/p}}$ submatrices was fused together in order to create a final, compressed, $\mathbf{D_{compr.\lambda/p}}$ dataset with selected information from the full initial image (Figure 3b). The sequential use of a purest variable selection method on small blocks of the initial image not only helps in speeding up the data analysis process but, most importantly, ensures that even minor compounds present in local zones of the raw dataset will be kept in the image compressed version. At this point it is important to note that the selection of information is driven by the difference in spectral and spatial features and not by the percentage of variance expressed by the different compounds in the considered sample. Here, a difference between LIBS and PIL spectra has to be pointed out. While the LIBS spectra show numerous and very fine peaks, only two broad bands are observed with PIL spectroscopy. So, in order to further compress the PIL spectral domain, the baseline part between the two bands was suppressed. In the rest of the dataset, the selection of the purest pixels was applied as in the LIBS dataset.

4) *Unmixing MCR-ALS analysis on the double compressed image data*: once the compression is accomplished, the MCR-ALS process can be started to retrieve the spectral signatures and related distribution maps of the compounds in the image. Using the double-compressed dataset $\mathbf{D_{compr.\lambda/p}}$, the first step was to newly apply the SIMPLISMA-based method to extract the purest spectra to be used as initial estimates for the signal unmixing technique, as normally done in the routine approach, and, most importantly, to assist in the selection of the number of components for the unmixing step. Indeed, selecting the optimal number of components for an MCR model is always a challenge on complex imaging samples, particularly if they contain minor compounds [27], since the presence of these contributions

may not be detected by methods based on analyzing the variance explained to decide the MCR model size, such as Singular Value Decomposition (SVD) does. The innovation in this work is that instead of using SVD to estimate the number of components, the purity of the selected spectra, p_i, as defined by SIMPLISMA, will be the adopted criterion [28]. Hence, a graphical representation plotting purity vs. nr. of components is used to set the threshold for component selection. For the sake of a better interpretability, the y-axis represents p_i/p_1, being p_1 the purity of the first selected spectrum and p_i the purity of the following *i* selected spectra. In this way, the y-axis goes from 0 to 1, i.e., a value of 0.9 will mean that the purity of a certain selected spectrum, meaning a new component in the model, is 90% the value of the first selected spectrum. Therefore, it is possible to observe the difference in purity of any selected spectrum with the first one, related to the difference in spectral shape between them and not to explained variance. Of course, the purity value decreases from component to component and a threshold expressed as, e.g., 1% of purity with respect to the first spectrum selected, can be set. In this manner, it is possible to estimate the right interval of components to consider in the MCR-ALS analysis, avoiding to include either too few or too many possible chemical contributions. An example of this graphical representation is reported in the supplementary material (Figure S2), where a number of components around 8-10 seems a reasonable estimate. Different MCR-ALS resolutions will then be calculated for different values of rank within the range estimated but a single solution will finally be adopted, based on the quality of the extracted pure spectra compared to LIBS database spectra (Figure 3c). The MCR model obtained can be expressed as: $\mathbf{D}_{\mathbf{compr.\,\lambda/p}} = \mathbf{C}_{\mathbf{compr.\,\lambda/p}}\,\mathbf{S^T}_{\mathbf{compr.\,\lambda/p}}$ where $\mathbf{C}_{\mathbf{compr.\,\lambda/p}}$ and $\mathbf{S^T}_{\mathbf{compr.\,\lambda/p}}$ are compressed versions of the information in the distribution maps and spectral fingerprints, respectively. In these analyses, only the constraint of non-negativity was used in the concentration and spectral mode, respectively.

5) *Reconstruction of full distribution maps and spectral signatures*: in this last step, the full spectral signatures and complete distribution maps of the initial image were recovered with two suitable single least-squares steps. Note that the MCR model that corresponds to the analysis of full image would be $\mathbf{D} = \mathbf{C}_{\mathbf{tot}}\,\mathbf{S^T}_{\mathbf{tot}}$, meaning **D** the matrix containing all the spectra of the original image, $\mathbf{S^T}_{\mathbf{tot}}$ the pure complete spectra of the compounds in the image and $\mathbf{C}_{\mathbf{tot}}$ the matrix of related concentration profiles that conveniently refolded provide the complete distribution maps. First, the double compressed spectral signatures ($\mathbf{S^T}_{\mathbf{compr.\,\lambda/p}}$, where λ represents the compressed spectral channels and p the compressed pixels) obtained from the previous MCR results were combined with each of the *i* subimages ($\mathbf{D}_{\mathbf{i,compr.\,\lambda}}$), where the spectral dimension was compressed but the pixel dimension was as in the original image, to rebuild the concentration profiles ($\mathbf{C}_{\mathbf{i\ compr.\,\lambda}}$), as represented in Eq. 1:

$$\mathbf{C}_{\text{i compr. }\lambda} = \mathbf{D}_{\text{i,compr. }\lambda} (\mathbf{S}^{\mathbf{T}}_{\text{compr. }\lambda/\text{p}})^{+} \quad (1)$$

where $(\mathbf{S}^{\mathbf{T}}_{\text{compr. }\lambda/\text{p}})^{+}$ is the pseudoinverse of matrix $\mathbf{S}^{\mathbf{T}}_{\text{compr. }\lambda/\text{p.}}$ All the $(\mathbf{C}_{\text{i compr. }\lambda})$ matrices were appended together to reobtain the matrix of concentration profiles corresponding to the whole initial image, $\mathbf{C}_{\text{tot}}$ (Figure 2d), the profiles of which can be conveniently refolded to give distribution maps. Finally, by combining the obtained concentration profile $\mathbf{C}_{\text{tot}}$ matrix with the original data matrix **D** (represented by the whole spectral domain), the full spectral signatures ($\mathbf{S}^{\mathbf{T}}_{\text{tot}}$) were finally obtained (Figure 2e), as represented in Eq. 2:

$$\mathbf{S}^{\mathbf{T}}_{\text{tot}} = (\mathbf{C}_{\text{tot}})^{+}\,\mathbf{D} \quad (2)$$

Data fusion strategy

As already described above, at first the LIBS and PIL datasets will be separately investigated by the use of the MCR-ALS analysis, in order to obtain a general idea about the information coming from the different techniques. This will form the first part of the results section. Without revealing these results, the observation of the average spectra acquired in LIBS and PIL already allow us to think that the exploitation of the latter will certainly be very limited due to the low number of spectral contributions when it is used alone. The interpretation of PIL spectral data acquired on complex natural samples remains a real challenge today. So, the main idea of this data fusion approach was to increase the possibility to better understand the luminescence phenomenon, investigating potential correlation or anticorrelation between the two spectroscopic techniques. Different fusion strategies have already been proposed in the literature [29–31]. In this work, the low-level data fusion was used for this study, which means concatenating the LIBS and PIL spectrum related to each image pixel to form a multiset configuration. Note that this step is easily done because the instrumental setup of the LIBS/PIL system ensures a perfect spatial congruency between LIBS and PIL pixel spectra. As a first step, all the LIBS and PIL pixel spectra, already compressed in the spectral dimension were concatenated together. It is important to stress here that only the LIBS and PIL matrices from the compression on the spectral variables were used for the fusion, not those from the double compression. The reason for using this strategy is that PIL spectra alone, as already explained, have much poorer information than LIBS. Using all the pixels initially in the fused structure ensures that the subsequent pixel selection will be carried out in such a way that the relevant correlation between the two techniques is appropriately captured. Finally, the extraction of the initial estimates with SIMPLISMA and the MCR-ALS analysis were carried out as previously described, but using the fused LIBS/PIL dataset. Again, the full spectral signatures and complete distribution maps were recovered using

two single least-squares steps combining the MCR results and the information in the extended pixel and spectral dimensions.

RESULTS AND DISCUSSION

LIBS dataset analysis. MCR-ALS results.

Due to its complexity, the LIBS dataset was the first one to be analyzed and investigated by the use of MCR-ALS. With the proposed data analysis pipeline, only 89400 over the more than 2 million of spectra (the 4% of the whole information) and only 489 spectral channels over the initial 2048 (the 24% of the initial value) were used. Considering the double compression, only 1% of the initial image information was used to perform MCR-ALS analysis and finally extract the full size maps and resolved LIBS spectra of each pure chemical contribution. The number of components needed to describe this compressed information, as estimated from the purity-based graphical method then suggested to select 6 significant chemical contributions. Figure 4 shows that the first four contributions from the MCR-ALS model on the compressed data are related to elements already observed in Figure 2. Comparison of the resolved LIBS features with simulated spectra confirmed the identity of these compounds. However, the results offer as additional information the evidence that Cr and V are spatially correlated because they are present in the same MCR pure component. It is then interesting to see that the other 5 pure contributions correspond to single elements which is quite unusual in the context of LIBS signal unmixing. If we look at the distributions of these elements, we could first say that Ca, Fe and Ti are approximately located in the same areas. However, a more detailed analysis shows specific sub-zones of the sample for each of them. Additionally, it should be noted that Ca and Fe are often present along cracks of the mineral. The Ti contribution is also very interesting because it highlights areas of the sample for which no PIL signal was observed in Figure 2. The last two components deserve a separate discussion. The first one seems to be equally distributed in all the mineral (signal contribution #4). The extracted spectrum is undeniably that of argon. However, the reader should not misunderstand the location of this element, which is not in fact part of the sample but comes from the gas flow above it, used to stabilize the plasma. As for the last contribution #6, its distribution image seems at first sight to be mainly related to noise. However, the extracted pure spectrum shows that the LIBS signature can be unambiguously attributed to the emission spectrum of zirconium. This is a very interesting result because this trace element could not be detected by classical single band LIBS integration because of the low signal-to-noise ratio of the related signal and the strong spatial and spectral overlap with other major elements.

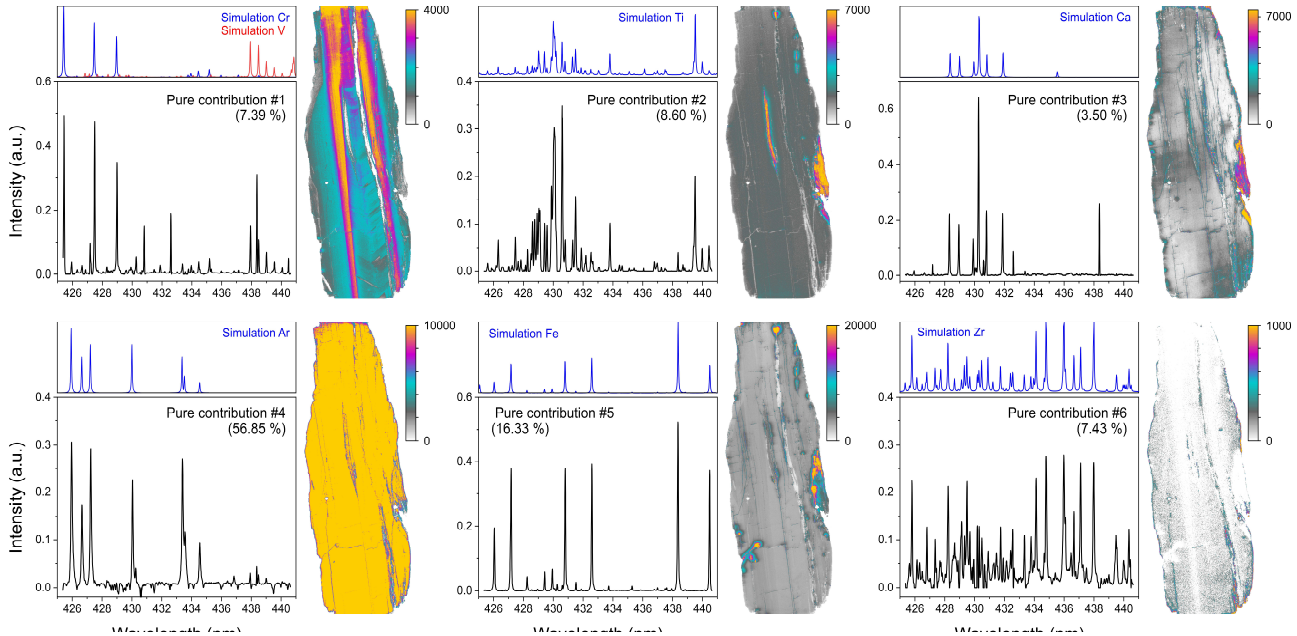


Figure 4. MCR-ALS (Multivariate Curve Resolution - Alternating Least Squares) results on the LIBS dataset.

PIL dataset analysis. MCR-ALS results.

As previously explained, the interpretation of only the PIL image can be very challenging because of the broad spectral features provided by this technique, and this sample is not an exception. Although the luminescence of kyanite has been studied for more than 80 years [12], the interpretation of the emission characteristics is still not clear; hence the interest in proposing an original spectroscopic setup and an associated data processing approach. Researchers agree that the emissions observed in the luminescence spectrum are attributed to different Cr^{3+} centers in the aluminosilicate mineral. They also often associate the differences in luminescence behavior of Cr^{3+} with its substitution in different positions of Al^{3+} inside the kyanite structure. Nevertheless, these positions are so similar that it is challenging to explain significant differences in luminescence properties. It is therefore time to see whether the MCR-ALS approach can help us in this exploration of the PIL dataset exploited alone. The pure spectra and the corresponding distribution maps extracted from this signal unmixing method are shown in Figure 5. First, it is interesting to see that the number of components estimated for the PIL dataset was two.

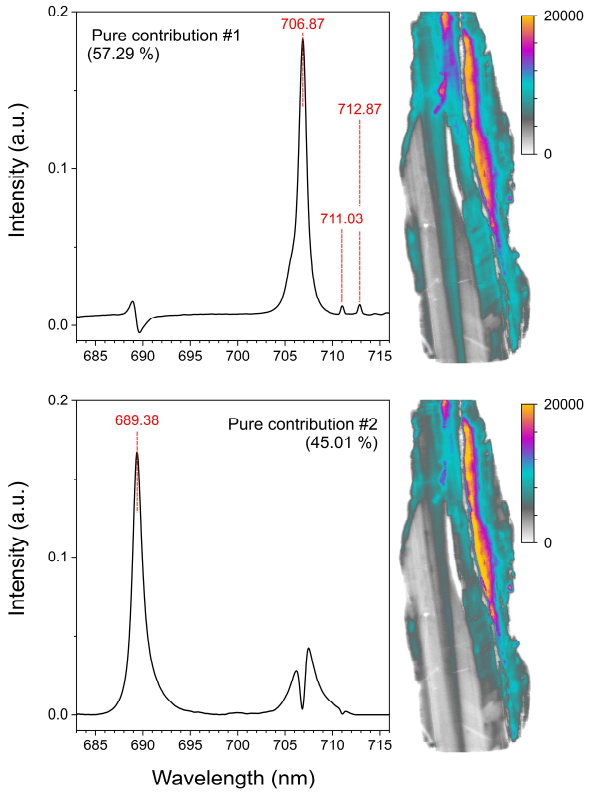


Figure 5. MCR-ALS results on the PIL dataset of the considered kyanite sample.

We were far from observing this with the classical integration method since the two PIL images at 689 and 706 nm presented in Figure 2 were at first sight very similar to each other. Thus, even if the two spatial distributions extracted with MCR-ALS are very close, they still show areas with some variations in intensity. From a spectral point of view, we see that the two most intense emission bands of the dataset are now separated in these two contributions. In the pure contribution #1, an intense luminescence signal at 706.87 nm is accompanied by a doublet of low intensity at 711.03 and 712.87 nm. In addition, a luminescence signal is observed at 688.93 nm. It is in this same spectral zone that we can find the maximum of luminescence in the pure component #2 centered on 689.38 nm. Thus we now understand that the luminescence initially observed around 689 nm from the raw data was in fact at least coming from two distinct signals. Finally, a last observation of the pure component #2 could make us think of the presence of a doublet at 706.19 and 707.48 nm. This is not the case since it is in fact the representation of a broadening of the emission band observed on the component #1 at 706.87 nm. With the results obtained from the MCR-ALS analysis of PIL data, the complexity of the PIL spectra can be confirmed. However, the very strong overlap between the components both in the spectral and spatial directions limits the signal unmixing power of the method and the ambiguity in the solutions obtained hinders the proper understanding of the luminescence phenomenon. To improve this situation, the fusion of the PIL and LIBS datasets

will be mandatory to reliably understand which kind of elements are located in a specific zone of the mineral and associated with the luminescence effect.

LIBS/PIL data-fusion. MCR-ALS results.

After the observation of the individual MCR-ALS results for the LIBS and the PIL images separately, the fusion of the two datasets was carried out. The main idea of this strategy was to deepen the chemical information from both techniques by finding correlations among the different spectroscopic signals and components. In particular, by combining the finer spectroscopic features of LIBS with the broad PIL signals, it now seems possible to give a suitable chemical interpretation to the luminescence signals. Using the proposed compression pipeline, only 89400 pixels and 1165 spectral channels were selected from the fused dataset. More precisely, considering the initial size of the merged dataset (more than 2 million spectra and about 4000 channels for both spectroscopies), the amount of information resulting from the double compression and thus finally used for the signal unmixing constituted only 1% of the original data. The MCR-ALS results on the fused dataset are given in Figure 6 after the estimation of the number of components suggested seven significant chemical contributions. As a reminder, the pure spectra extracted by MCR-ALS on the fused dataset contain simultaneously a LIBS part and a PIL one. Thus, the extended LIBS/PIL fingerprint of the extracted components will directly differentiate the elements that present a LIBS signal associated with a PIL luminescence phenomenon from those that only have a LIBS signal and no luminescence induced. The observation of the first three pure contributions in Figure 6 clearly show that the elements Ca and Fe present a very low luminescence phenomenon (689.5 nm), which is completely absent for Ti. Moreover, there is no significant luminescence for the Ar element either (pure contribution #6), which it is not part of the sample but of the atmosphere above its surface. The strongest luminescence signals are observed for pure contributions #5 and #7 corresponding to a mixture of Cr and V followed by a weaker but nevertheless significant luminescence for Zr in pure contribution #4. This last contribution is very interesting because Zr does not form luminescence center. We can therefore say that the luminescence observed on contribution #4 is certainly due to an indirect correlation of another element with Zr. If we look at these three PIL contributions at the spectral level, they are quite singular. First, the pure contribution #7 has two peaks at 689.34 and 706.89 nm. On the other hand, the pure contribution #5 shows only one peak at 706.89 nm. Finally, pure contribution #4 has a peak at 689.70 nm and an unclear very low signal at 711.03 and 712.87 nm. These last results show the power of the MCR-ALS approach associated with LIBS/PIL data fusion since the luminescence is not due to two intense contributions, but potentially several contributions slightly shifted in wavelength that can be differentiated. In conclusion, we can state that the luminescence of this sample comes mainly from the simultaneous presence of Cr and

V. These results are fully consistent with previous work suggesting that luminescence could potentially originate from the Cr^{3+} and V^{2+} centers [12].

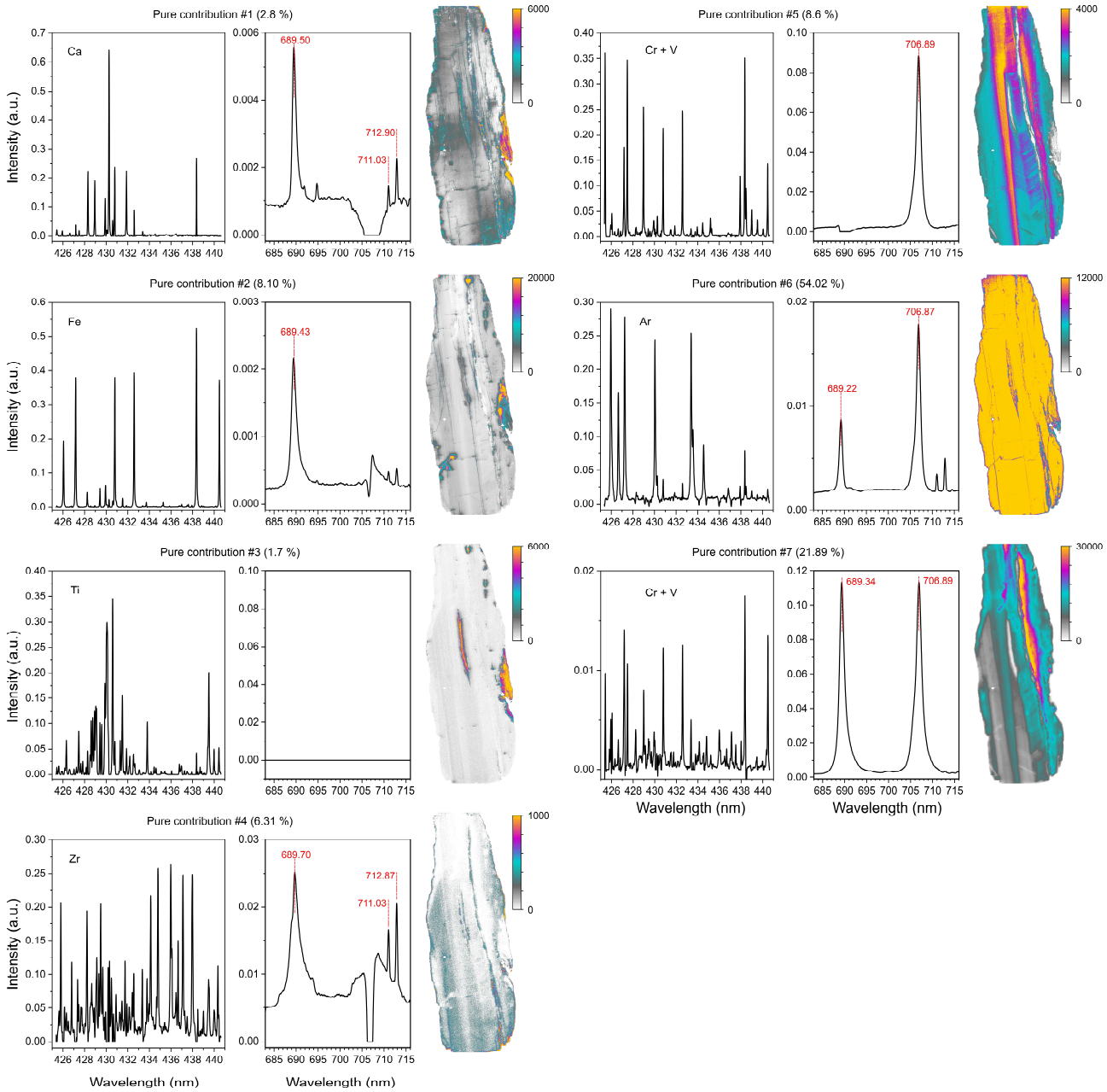


Figure 6. MCR-ALS results on the fused LIBS/PIL dataset of the considered kyanite sample.

However, even if we can state this, there is still the end of the story to write. Indeed, even though the spatial distributions of the pure #5 and #7 contributions are relatively close and partially overlapped, they both potentially show singular chemical information. The same statement can also be made at the spectral level. Thus, the joint exploitation of new LIBS and PIL spectral domains and even an extension of the

fusion concept to other spectroscopic imaging techniques should help further in elucidating the whole nature of the complex luminescence phenomenon.

CONCLUSIONS

LIBS imaging is now clearly a tool of choice for the elemental characterization of complex samples with applications in many fields. Nevertheless, its high acquisition rate, which is an undeniable advantage, is also constrained by the millions of spectra (each containing thousands of wavelengths) acquired from a single sample that require the use of powerful multivariate data analysis tools. This difficult aspect leads to the interest in proposing a data analysis procedure capable of extracting the most distinct information, i.e., purest variables, at the spectral and spatial level, both for major and minor compounds, to facilitate the unmixing analysis without losing quality in the spatial and spectral definition of the imaged components. With this simple procedure, it was possible to reduce the initial amount of data and keep the best and most unmixed 1% of the total information. The size and quality of the selected information allowed not only speeding up the analysis, but obtaining extremely reliable spectral fingerprints and distribution maps for the extracted MCR-ALS components. The data analysis pipeline has been tested on the LIBS/PIL dataset, but can be used in any other kind of large imaging dataset coming from an individual platform or from the fusion of several of them.

The study of the kyanite dataset showed that each resolved component was potentially related to one or two elements present in the mineral. Last but not least, another important aspect of this work was fusing together LIBS and PIL datasets to provide a chemical interpretation for the PIL bands and better understand which elements were related to this luminescence effect. This work represents the first published work on the fusion of LIBS and PIL imaging data and their simultaneous exploitation in a signal unmixing approach such as MCR-ALS. Thus even if we have demonstrated for this particular kyanite sample that the luminescence phenomenon was mainly associated with the Cr and V elements, our next work could be focused on the exploitation of new spectral domains or the addition of another spectroscopy such as Raman and Laser-Induced Time-Resolved Luminescence to the first two in a fusion process.

ASSOCIATED CONTENT

Supporting Information

Figure S1: Comparison of raw spectra and pretreated ones.

Figure S2: Relative purity values graph.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

Author Contributions

The manuscript was written through contributions of all authors. All authors contributed equally and have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

The authors warmly thank Rodrigo Rocha de Oliverira and Adrián Gómez-Sánchez, PhD students in the chemometric group of the department of Chemical Engineering and Analytical Chemistry (Universitat de Barcelona) for fruitful discussions about chemometrics strategies. The authors also thank the University of Lille for providing a mobility grant (called MOB-LIL-EX) to carry out this research at the Universitat de Barcelona. A.J. acknowledges financial support from the Spanish govern-ment through the research project PID 2019-1071586B-IOO.

REFERENCES

[1] Jolivet, L.; Leprince, M.; Moncayo, S.; Sorbier, L.; Lienemann, C.-P.; Motto-Ros, V. Review of the Recent Advances and Applications of LIBS-Based Imaging. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2019**, *151*, 41–53. https://doi.org/10.1016/j.sab.2018.11.008.

[2] Motto-Ros, V.; Moncayo, S.; Trichard, F.; Pelascini, F. Investigation of Signal Extraction in the Frame of Laser Induced Breakdown Spectroscopy Imaging. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2019**, *155*, 127–133. https://doi.org/10.1016/j.sab.2019.04.004.

[3] Labutin, T. A.; Lednev, V. N.; Ilyin, A. A.; Popov, A. M. Femtosecond Laser-Induced Breakdown Spectroscopy. *J. Anal. At. Spectrom.* **2016**, *31* (1), 90–118. https://doi.org/10.1039/C5JA00301F.

[4] Carvalho, R. R. V.; Coelho, J. A. O.; Santos, J. M.; Aquino, F. W. B.; Carneiro, R. L.; Pereira-Filho, E. R. Laser-Induced Breakdown Spectroscopy (LIBS) Combined with Hyperspectral Imaging for the Evaluation of Printed Circuit Board Composition. *Talanta* **2015**, *134*, 278–283. https://doi.org/10.1016/j.talanta.2014.11.019.

[5] Gaudiuso, R.; Melikechi, N.; Abdel-Salam, Z. A.; Harith, M. A.; Palleschi, V.; Motto-Ros, V.; Busser, B. Laser-Induced Breakdown Spectroscopy for Human and Animal Health: A Review. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2019**, *152*, 123–148. https://doi.org/10.1016/j.sab.2018.11.006.

[6] Trichard, F.; Gaulier, F.; Barbier, J.; Espinat, D.; Guichard, B.; Lienemann, C.-P.; Sorbier, L.; Levitz, P.; Motto-Ros, V. Imaging of Alumina Supports by Laser-Induced Breakdown Spectroscopy: A New Tool to Understand the Diffusion of Trace Metal Impurities. *Journal of Catalysis* **2018**, *363*, 183–190. https://doi.org/10.1016/j.jcat.2018.04.013.

[7] Nardecchia, A.; Fabre, C.; Cauzid, J.; Pelascini, F.; Motto-Ros, V.; Duponchel, L. Detection of Minor Compounds in Complex Mineral Samples from Millions of Spectra: A New Data Analysis Strategy in LIBS Imaging. *Analytica Chimica Acta* **2020**, *1114*, 66–73. https://doi.org/10.1016/j.aca.2020.04.005.

[8] Cáceres, J. O.; Pelascini, F.; Motto-Ros, V.; Moncayo, S.; Trichard, F.; Panczer, G.; Marín-Roldán, A.; Cruz, J. A.; Coronado, I.; Martín-Chivelet, J. Megapixel Multi-Elemental Imaging by Laser-Induced Breakdown Spectroscopy, a Technology with Considerable Potential for Paleoclimate Studies. *Sci Rep* **2017**, *7* (1), 1–11. https://doi.org/10.1038/s41598-017-05437-3.

[9] Fabre, C.; Devismes, D.; Moncayo, S.; Pelascini, F.; Trichard, F.; Lecomte, A.; Bousquet, B.; Cauzid, J.; Motto-Ros, V. Elemental Imaging by Laser-Induced Breakdown Spectroscopy for the Geological Characterization of Minerals. *J. Anal. At. Spectrom.* **2018**, *33* (8), 1345–1353. https://doi.org/10.1039/C8JA00048D.

[10] Gaft, M.; Nagli, L.; Groisman, Y. Plasma Induced Luminescence (PIL). *Optical Materials* **2011**, *34* (2), 368–375. https://doi.org/10.1016/j.optmat.2011.05.024.

[11] Gaft, M.; Raichlin, Y.; Pelascini, F.; Panzer, G.; Motto Ros, V. Imaging Rare-Earth Elements in Minerals by Laser-Induced Plasma Spectroscopy: Molecular Emission and Plasma-Induced Luminescence. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2019**, *151*, 12–19. https://doi.org/10.1016/j.sab.2018.11.003.

[12] Gaft, M.; Reisfeld, R.; Panczer, G. *Modern Luminescence Spectroscopy of Minerals and Materials*; Springer Mineralogy; Springer International Publishing: Cham, 2015. https://doi.org/10.1007/978-3-319-24765-6.

[13] Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. A Graphical User-Friendly Interface for MCR-ALS: A New Tool for Multivariate Curve Resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems* **2005**, *76* (1), 101–110. https://doi.org/10.1016/j.chemolab.2004.12.007.

[14] de Juan, A.; Tauler, R. Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications. *Critical Reviews in Analytical Chemistry* **2006**, *36* (3–4), 163–176. https://doi.org/10.1080/10408340600970005.

[15] de Juan, A.; Tauler, R. Multivariate Curve Resolution: 50 Years Addressing the Mixture Analysis Problem – A Review. *Analytica Chimica Acta* **2021**, *1145*, 59–78. https://doi.org/10.1016/j.aca.2020.10.051.

[16] Ghaffari, M.; Omidikia, N.; Ruckebusch, C. Joint Selection of Essential Pixels and Essential Variables across Hyperspectral Images. *Analytica Chimica Acta* **2021**, *1141*, 36–46. https://doi.org/10.1016/j.aca.2020.10.040.

[17] Bassel, L.; Motto-Ros, V.; Trichard, F.; Pelascini, F.; Ammari, F.; Chapoulie, R.; Ferrier, C.; Lacanette, D.; Bousquet, B. Laser-Induced Breakdown Spectroscopy for Elemental Characterization of Calcitic Alterations on Cave Walls. *Environ Sci Pollut Res* **2017**, *24* (3), 2197–2204. https://doi.org/10.1007/s11356-016-7468-5.

[18] Motto-Ros, V.; Negre, E.; Pelascini, F.; Panczer, G.; Yu, J. Precise Alignment of the Collection Fiber Assisted by Real-Time Plasma Imaging in Laser-Induced Breakdown Spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2014**, *92*, 60–69. https://doi.org/10.1016/j.sab.2013.12.008.

[19] Nardecchia, A.; Motto-Ros, V.; Duponchel, L. Saturated Signals in Spectroscopic Imaging: Why and How Should We Deal with This Regularly Observed Phenomenon? *Analytica Chimica Acta* **2021**, *1157*, 338389. https://doi.org/10.1016/j.aca.2021.338389.

[20] D'Angelo, C.; Diaz Pace, D.; Bertuccelli, D.; Bertuccelli, G. Spectroscopic Analysis of Signals from LIBS Experiments; Marcano O., A., Paz, J. L., Eds.; 2004; pp 1037–1042. https://doi.org/10.1117/12.591209.

[21] Whittaker, E. T. On a New Method of Graduation. *Proceedings of the Edinburgh Mathematical Society* **1922**, *41*, 63–75. https://doi.org/10.1017/S0013091500077853.

[22] Savitzky, Abraham.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36* (8), 1627–1639. https://doi.org/10.1021/ac60214a047.

[23] Windig, W.; Guilment, J. Interactive Self-Modeling Mixture Analysis. *Analytical Chemistry* **1991**, *63* (14), 1425–1432. https://doi.org/10.1021/ac00014a016.

[24] Windig, W.; Lippert, J. L.; Robbins, M. J.; Kresinske, K. R.; Twist, J. P.; Snyder, A. P. Interactive Self-Modeling Multivariate Analysis. *Chemometrics and Intelligent Laboratory Systems* **1990**, *9* (1), 7–30. https://doi.org/10.1016/0169-7439(90)80050-G.

[25] Gourvénec, S.; Massart, D. L.; Rutledge, D. N. Determination of the Number of Components during Mixture Analysis Using the Durbin–Watson Criterion in the Orthogonal Projection Approach and in the SIMPLe-to-Use Interactive Self-Modelling Mixture Analysis Approach. *Chemometrics and*

Intelligent Laboratory Systems **2002**, *61* (1–2), 51–61. https://doi.org/10.1016/S0169-7439(01)00172-1.

[26] Nardecchia, A.; Duponchel, L. Randomised SIMPLISMA: Using a Dictionary of Initial Estimates for Spectral Unmixing in the Framework of Chemical Imaging. *Talanta* **2020**, *217*, 121024. https://doi.org/10.1016/j.talanta.2020.121024.

[27] Boiret, M.; de Juan, A.; Gorretta, N.; Ginot, Y.-M.; Roger, J.-M. Distribution of a Low Dose Compound within Pharmaceutical Tablet by Using Multivariate Curve Resolution on Raman Hyperspectral Images. *Journal of Pharmaceutical and Biomedical Analysis* **2015**, *103*, 35–43. https://doi.org/10.1016/j.jpba.2014.10.024.

[28] Eckart, C.; Young, G. The Approximation of One Matrix by Another of Lower Rank. *Psychometrika* **1936**, *1* (3), 211–218. https://doi.org/10.1007/BF02288367.

[29] de Juan, A.; Gowen, A.; Duponchel, L.; Ruckebusch, C. Image Fusion. In *Data Handling in Science and Technology*; Elsevier, 2019; Vol. 31, pp 311–344. https://doi.org/10.1016/B978-0-444-63984-4.00011-9.

[30] Piqueras, S.; Bedia, C.; Beleites, C.; Krafft, C.; Popp, J.; Maeder, M.; Tauler, R.; de Juan, A. Handling Different Spatial Resolutions in Image Fusion by Multivariate Curve Resolution-Alternating Least Squares for Incomplete Image Multisets. *Anal. Chem.* **2018**, *90* (11), 6757–6765. https://doi.org/10.1021/acs.analchem.8b00630.

[31] Gómez-Sánchez, A.; Marro, M.; Marsal, M.; Loza-Alvarez, P.; de Juan, A. 3D and 4D Image Fusion: Coping with Differences in Spectroscopic Modes among Hyperspectral Images. *Anal. Chem.* **2020**, *92* (14), 9591–9602. https://doi.org/10.1021/acs.analchem.0c00780.