HyperPCA: a powerful tool to extract elemental maps from noisy data obtained in LIBS mapping of materials

Riccardo Finotello, Mohamed Tamaazousti, Jean-Baptiste Sirven

HyperPCA: a Powerful Tool to Extract Elemental Maps from Noisy Data Obtained in LIBS Mapping of Materials

—

Supplementary Material

Riccardo Finotello[a,b,*], Mohamed Tamaazousti[b], Jean-Baptiste Sirven[a]

[a] *Université Paris-Saclay, CEA, Service d'Etudes Analytiques et de Réactivité des Surfaces (SEARS), F-91191, Gif sur Yvette, France*
[b] *Université Paris-Saclay, CEA, LIST, F-91120, Palaiseau, France*

---

1. Technical Background

We briefly recall the definitions and properties of the Principal Components Analysis (PCA) and the Discrete Wavelet Transform (DWT) used in the main paper.

1.1. Principal Components Analysis

Let $Y \in \mathbb{R}^{n\times p}$ be a data matrix with zero sample mean whose vector components $y_{(i)} \in \mathbb{R}^p$ for $i = 1, 2, \dots, n$ represent the observed data. The principal components (PCs) are an orthonormal basis of vectors $\left\{w_{(i)}\right\}_{i=1,2,\dots,p}$ representing the directions of the lines which best fit the data. PCA [1] is a statistical and unsupervised machine learning tool to compute the PCs and use them to "rotate" the data in a suitable basis, where the sample variance is the largest.

Mathematically, PCA is a linear map realized as the right multiplication between matrices:

$$\begin{array}{rccc} \text{PCA:} & \mathbb{R}^{n\times p} & \to & \mathbb{R}^{n\times p} \\ & Y & \mapsto & Y' = YW \end{array} \tag{1}$$

where $W \in \mathbb{R}^{p\times p}$ contains the PCs as columns, such that $W^TW = \mathbb{1}$, that is W is an orthogonal transformation $W \in \mathrm{O}(p)$. By its definition, the first component $w_{(1)} \in \mathbb{R}^p$ of W is the direction which maximizes the variance of the transformed data. Let $y'_{(i)\,(1)} = y_{(i)} \cdot w_{(1)}$ for $i = 1, 2, \dots, n$ be the rotated components of the new data vectors, then the PCA computation translates into the maximization problem:

$$\begin{aligned} w_{(1)} &= \arg\max_{||w||=1} \left\{ \frac{1}{n-1} \sum_{i=1}^{n} \left(y'_{(i)\,(1)} \right)^2 \right\} \\ &= \arg\max_{w} \left\{ \frac{w^T C w}{w^T w} \right\}, \end{aligned} \tag{2}$$

where $C = \frac{1}{n-1} Y^T Y \in \mathbb{R}^{p\times p}$ is the sample covariance matrix of the data. The last line is known as *Rayleigh quotient*. The solution to the extremization problem is given if $w_{(1)}$ is the eigenvector of C corresponding to the largest eigenvalue (see the *Rayleigh-Ritz theorem* [2]). Other components $w_{(k)}$ for $k > 1$ can be computed through the iterated application of this procedure after the *deflation* of the original matrix:

$$Y^{(k)} = Y - \sum_{s=1}^{k-1} Y w_{(s)} w_{(s)}^T. \tag{3}$$

*Corresponding author
Email address: riccardo.finotello@cea.fr (Riccardo Finotello)

The k-th PC $w_{(k)} \in \mathbb{R}^p$ is given by the eigenvector associated to the largest eigenvalue of

$$C^{(k)} = \frac{1}{n-1} \left(Y^{(k)}\right)^T Y^{(k)}. \tag{4}$$

In general, the PCs are represented by the orthonormal basis of the eigenvectors of C, ordered by explained variance, that is the value of the corresponding eigenvalues.

In the main paper, we start from individual spectra $x_{(i)} \in \mathbb{R}^p$ for $i = 1, 2 \ldots, n$. The procedure is then as follows:

1. the mean intensity $\overline{x} = \frac{1}{n-1} \sum_{i=1}^{n} x_{(i)} \in \mathbb{R}^p$ is computed across all spectra, i.e. column-wise, and it is subtracted to the original data vectors. Vectors $y_{(i)} = x_{(i)} - \overline{x} \in \mathbb{R}^p$ for $i = 1, 2, \ldots, n$ are defined, and the corresponding centred matrix $Y = \left(y_{(i)}\right)_{i \in [1,n]}$ is constructed;
2. in order to estimate the directions of maximal variance, eigenvalues and eigenvectors of the sample covariance matrix $C = \frac{1}{n-1} Y^T Y \in \mathbb{R}^{p \times p}$ are then computed and collected into the diagonal matrix of the eigenvalues $\Lambda \in \mathbb{R}^{p \times p}$ and the eigenvectors matrix $W \in \mathbb{R}^{p \times p}$. Each column vector $w_{(i)}$ of the orthonormal matrix W is the eigenvector associated to i-th eigenvalue;
3. the data matrix is projected onto its loadings, $L = \Lambda^{\frac{1}{2}} W^T \in \mathbb{R}^{p \times p}$, and scores, $S = Y W \Lambda^{-\frac{1}{2}} \in \mathbb{R}^{n \times p}$, such that $Y = S L$.

The newly constructed representation contains the PCs as columns of the scores, ordered by the fraction of total variance explained, and the projection coefficients as rows of the loadings.

1.2. Wavelet Transforms

In signal processing, a Fourier Transform (FT) is able to extract the frequency content of a signal, but it loses information related to the localization of the frequency in the original domain. A Short-Time FT (STFT) can instead be used to partially reinstate the ability to localize the frequency components by scanning the signal with a window function w:

$$\mathcal{F}_{\text{STFT}}\left[f \,|\, w\right](\omega, t) = \int_{\mathbb{R}} \mathrm{d}\tau \, f(\tau) w(t - \tau) e^{-i\omega\tau}. \tag{5}$$

The window considered is such to always consider intervals of the same length in the original domain: the number of sampled data points used to scan higher and lower frequencies is the same, thus losing the ability of a multi-level resolution.

Wavelets are square-integrable real function of real variables $\psi \in \mathrm{L}^2(\mathbb{R})$ which have been introduced to adapt the width of the slice of the domain to the frequencies which are analysed (see [3, 4] for more details):

$$\mathcal{F}_{\text{WT}}\left[f \,|\, \psi\right](s, t) = \int_{\mathbb{R}} \mathrm{d}\tau \, f(\tau) \psi\left(\frac{t - \tau}{s}\right), \tag{6}$$

where s is the scale factor to adjust the extension of the window, and t represents the localization parameter. Given their ability to adjust the size of the scanned portion of the domain, wavelets can be used to construct multi-level representations of a signal $f \in \mathrm{L}^2(\mathbb{R})$ at various levels of resolution.

A multi-level basis of wavelets is built by considering a sequence of vector spaces:

$$\{0\} \subset \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \subset \mathrm{L}^2(\mathbb{R}), \tag{7}$$

such that $\bigcup_{i \in \mathbb{Z}} V_i$ is dense in $\mathrm{L}^2(\mathbb{R})$, and $\bigcap_{i \in \mathbb{Z}} V_i = \{0\}$. Moreover, $g(x) \in V_i \Leftrightarrow g(2x) \in V_{i+1}$. Thus, $V_m = \overline{\text{span}\{g(2^m x - k), k \in \mathbb{Z}\}}$, where the overline indicates the closure of the set. An orthonormal basis of V_m can be defined as the family of functions

$$\left\{\psi_{m,k}(x) = 2^{\frac{m}{2}} \psi(2^m x - k)\right\}_{k \in \mathbb{Z}}, \tag{8}$$

where m is also referred as the *scale* parameter. At any given value of m, a square integrable function f can be approximated in the space V_m by projection onto the newly defined basis:

$$\mathcal{P}_m\left[f\right](x) = \sum_{k=-\infty}^{\infty} a_{m,k}\,\psi_{m,k}(x), \tag{9}$$

where $\{a_{m,k}\}_{k\in\mathbb{Z}}$ are the Approximation Coefficients (ACs) of f.

Let now W_m be the orthogonal complement of V_m such that $V_m = V_{m-1} \bigoplus W_{m-1}$, and let

$$\left\{\phi_{m,k}(x) = 2^{\frac{m}{2}}\phi(2^m x - k)\right\}_{k\in\mathbb{Z}} \tag{10}$$

be a basis of W_m. We then have the relation

$$\mathcal{P}_m\left[f\right] = \mathcal{P}_{m-1}\left[f\right] + \mathcal{Q}_{m-1}\left[f\right], \tag{11}$$

where Q_m is the projection operator onto W_m:

$$\mathcal{Q}_m\left[f\right](x) = \sum_{k=-\infty}^{\infty} d_{m,k}\phi_{m,k}(x). \tag{12}$$

Coefficients $\{d_{m,k}\}_{k\in\mathbb{Z}}$ are the Detail Coefficients (DCs) needed in order to move from one level of decomposition to the next. Relation (11) can be chained leading to:

$$\mathcal{P}_m\left[f\right]\Big|_t = \mathcal{P}_{m-t}\left[f\right] + \sum_{n=1}^{k} \mathcal{Q}_{m-t}\left[f\right]. \tag{13}$$

The equation shows the decomposition of the original signal at a given scale m in terms of lower scales, indexed by t: the original decomposition at scale m can be entirely recovered by considering the decomposition at level $m-t$ together with all the DCs from scale $m-t$ to m. In this sense, no information is lost.

A Discrete Wavelet Transform (DWT) is a particular wavelet transform in which the signal S and the basis functions are discretely sampled, and possibly finite. At a given scale m, they effectively act as pass-band filters on the original signal with the DCs, representing the projection (9), being constructed by a high-pass filter $h_{m,k}[n] = h_m[n-k]$ related to $\phi_{m,k}$, and the ACs, representing (12), being the result of the convolution with a low-pass filter $\ell_{m,k}[n] = \ell_m[n-k]$ connected to $\psi_{m,k}$:

$$\begin{aligned} d_{m,k} = d_m[k] &= \sum_{q=-L}^{+L} S[q]\,h_{m,q}[k], \\ a_{m,k} = a_m[k] &= \sum_{q=-L}^{+L} S[q]\,\ell_{m,q}[k], \end{aligned} \tag{14}$$

where L may be theoretically very large. The procedure can then be repeated on the ACs to further increase the frequency resolution as in (13). As for each successive iteration of the filter bank the frequency domain is halved as well as the original domain, the effect of the DWT is to subsample the signal at each scale m. According to the length of the support of S and the wavelet profile considered, the number of decomposition layers which can be computed changes.

2. Determination of the Signal-to-noise Ratio in Synthetic Datasets

As shown in Section 3.1 of the main text, synthetic spectra are generated from the theoretical distribution in local thermodynamical equilibrium, with the addition of stochastic noise. Wavelength channels which do not contain any signal, i.e. pure noise channels, are defined as:

$$x'^{[n]}_{(i)} = |\beta\, N|, \qquad i = 1, 2, \ldots, n, \qquad n = 1, 2, \ldots, n_\lambda, \tag{15}$$

where β is uniformly distributed in the interval $[0, \beta_{\max}]$, and N follows a standard normal distribution. In order to determine the signal-to-noise (SNR), we need to compute the expected value and the population variance of $x'^{[n]}_{(i)}$ in the pure noise channels of the synthetic spectra. This is possible through the computation of the moments of $|\beta|$ and $|N|$:

$$\mathbb{E}[|\beta|] = \mathbb{E}[\beta] = \frac{1}{\beta_{\max}} \int_0^{\beta_{\max}} \mathrm{d}t\, t = \frac{\beta_{\max}}{2}, \tag{16}$$

$$\mathrm{Var}(|\beta|) = \mathrm{Var}(\beta) = \frac{1}{\beta_{\max}} \int_0^{\beta_{\max}} \mathrm{d}t\, (t - \mathbb{E}[\beta])^2 = \frac{\beta_{\max}^2}{12}, \tag{17}$$

and

$$\mathbb{E}[|N|] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{d}t\, |t|\, e^{-\frac{t^2}{2}} = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} \mathrm{d}t\, t\, e^{-\frac{t^2}{2}} = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} \mathrm{d}u\, e^{-u} = \sqrt{\frac{2}{\pi}}, \tag{18}$$

$$\mathrm{Var}(|N|) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{d}t\, (|t| - \mathbb{E}[|N|])^2\, e^{-\frac{t^2}{2}} = \mathrm{Var}(N) - \mathbb{E}[|N|]^2 = 1 - \frac{2}{\pi}. \tag{19}$$

Finally, given the fact that the probability density functions of $|\beta|$ and $|N|$ are respectively independent, we can compute the moments of the distribution of $|\beta\, N|$:

$$\begin{aligned}
\mathbb{E}[|\beta\, N|] &= \mathbb{E}[|\beta|]\, \mathbb{E}[|N|] = \frac{\beta_{\max}}{\sqrt{2\pi}}, \\
\mathrm{Var}(|\beta\, N|) &= \mathbb{E}\Big[\beta^2 N^2 - \mathbb{E}[|\beta\, N|]^2\Big] \\
&= \mathbb{E}\Big[\Big(\beta^2 - \mathbb{E}[\beta]^2\Big) + \mathbb{E}[\beta]^2\Big]\, \mathbb{E}\Big[\Big(N^2 - \mathbb{E}[N]^2\Big) + \mathbb{E}[N]^2\Big] - \mathbb{E}[|\beta\, N|]^2 \\
&= \Big(\mathrm{Var}(\beta) + \mathbb{E}[\beta]^2\Big)\mathrm{Var}(N) - \frac{\beta_{\max}^2}{2\pi} \\
&= \beta_{\max}^2 \left(\frac{1}{3} - \frac{1}{2\pi}\right).
\end{aligned} \tag{20}$$

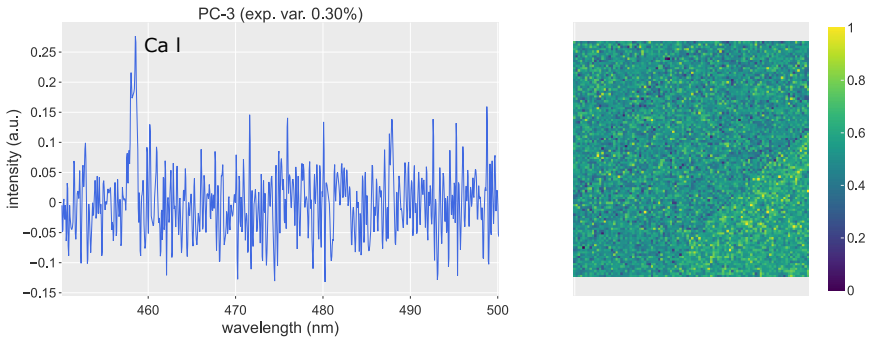
The SNR is defined as $\mathrm{Var}(|\beta\, N|)^{-\frac{1}{2}}$, thus:

$$\mathrm{SNR} = \frac{1}{\beta_{\max}} \sqrt{\frac{6\pi}{2\pi - 3}} \simeq \frac{2.40}{\beta_{\max}}. \tag{21}$$

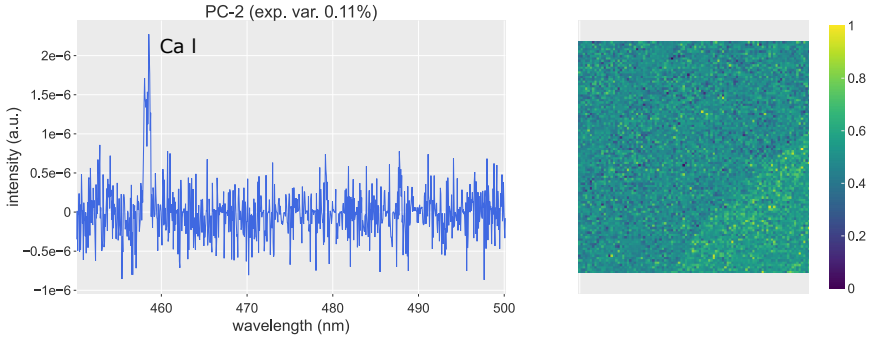
3. Performance Review of the Proposed Algorithm

In the principal text we showed that HyperPCA performs better than standard PCA. The application of simply one of the two techniques, either DWT or Sparse Kernel-PCA (kSPCA), to the data does not achieve the same results. In this additional section, we show that the combined action of the DWT and the application of kSPCA delivers the best outcome: we study separately the contributions of the two techniques we introduce, and we show how they compare to the current methods and the proposed technique. Namely, we show results using:

a. standard PCA applied to the DWT of the original data,
b. kSPCA applied directly to the original data,
c. kSPCA and PCA applied to the Continuous Wavelet Transform (CWT) of the spectra.



(a) PCA on the DWT of the "granite" dataset.



(b) kSPCA on the "granite" dataset (Ca emission line).

Figure 1: Different analyses of the "granite" dataset.

As the comparison with standard PCA applied to the data was already provided, showing these results enables a better comparison with alternative treatments.

The expected natural outcome of the application of PCA on the DWT of the original data is the retention of a high degree of noise which might prevent the complete separation of the elements in the dataset: the DWT indeed acts as an initial, weak denoising filter, but the lack of the kernel suppressing small entries in the covariance matrix retains most of the background noise. At the top of Figure 1 we show the PC showing Ca in the "granite" dataset of Section 3.1.3 in the main text. Noise is indeed evidently present in the loading vector, and it prevents the complete separation of Ca from the background distribution. This is a feature which is visible also in other datasets, as it is a general trait of the underlying physical nature of the data. The DWT of the original data is not enough to lead to an improvement of the standard PCA.

We then consider kSPCA applied directly to the data, without any kind of sparsity filter. At the bottom of Figure 1, we show the fifth PC computed on the "granite" dataset. By comparison with Figure 14 in the main text, the PC shows the Al contribution correctly, but the quality of the score matrix is worsened by the remaining noise component. The detection of the element is possible, as the application of the kernel lowers the detection threshold of signal. The lack of a sparse basis to appropriately express the signal, does not allow the kSPCA to reach the same quality as HyperPCA.

The comparison between scores and loadings in Figure 1 shows that the action of the DWT is mostly aimed at modifying the line profile and to highlight the presence of physical signal with respect to the noise. The modification of the eigenvalue distribution puts in evidence the physical line by acting directly on the signal-to-noise ratio, which in the case of kSPCA is almost doubled with respect to the the application of DWT only, which nonetheless has a noise filtering effect, too.

The application of PCA or kSPCA over the CWT of the data opens a different kind of discussion. In this section, we show the results obtained using the best fitting wavelet profile, a Mexican hat continuous wavelet, on the "granite" data. As a general trait, we notice two shortcomings to this approach:

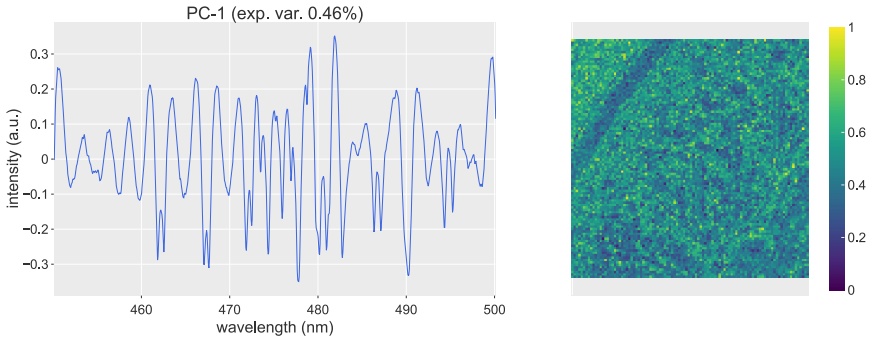


Figure 2: PCA on the CWT of the "granite" data.

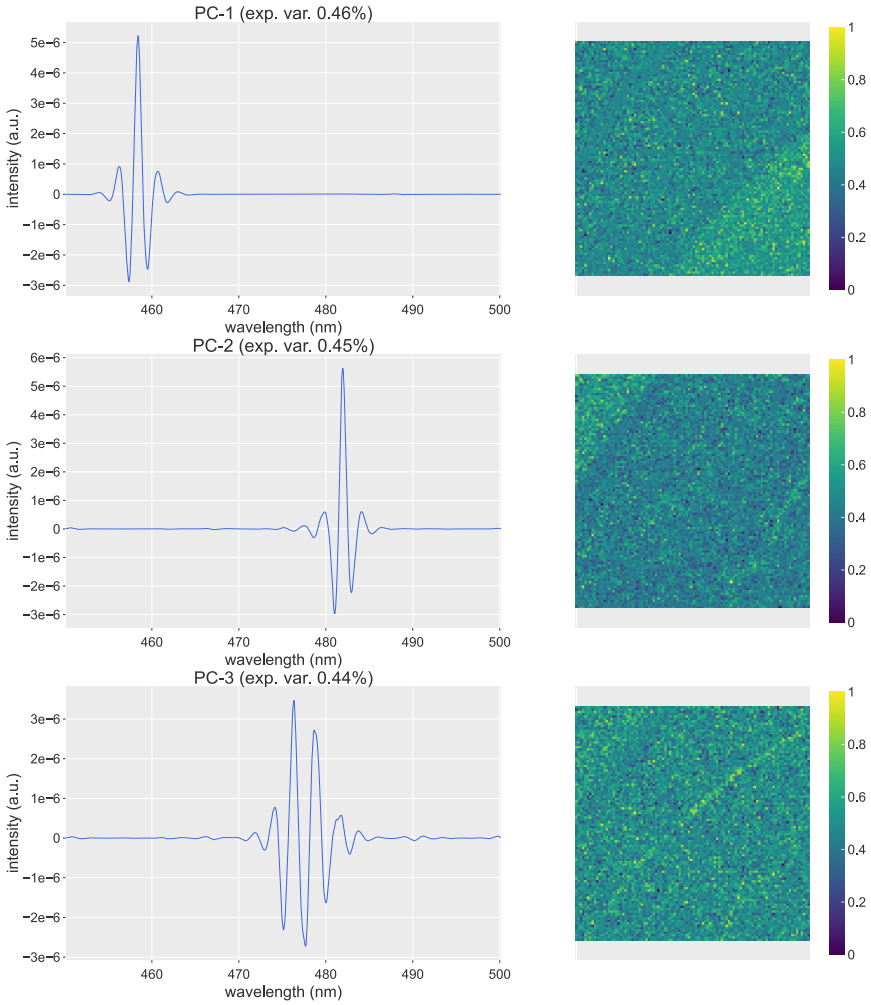
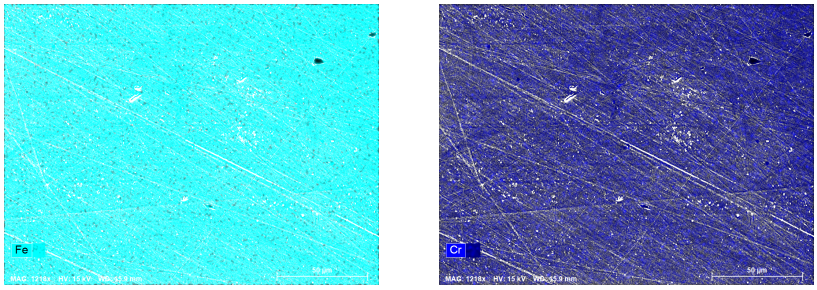


Figure 3: kSPCA on the CWT of the "granite" data.



(a) Fe (b) Cr

Figure 4: Distributions of the main elements in the 316L stainless steel.

	concentration
Fe	0.68
Cr	0.17
Ni	0.10
Mo	0.02
C	0.02
Si	0.01

Table 1: Concentration of elements in the stainless steel.

a. the creation of strong artefacts, not present in the original dataset,
b. the strong modification of the profiles of the emission lines.

In the presence of strong spectral interference, imposing a continuous wavelet profile, negatively impacts neighbouring emission lines, creating strong artefacts, which completely spoil the readability of the loading vectors, as standard PCA on the CWT shows in Figure 2. Imposing the profiles of the continuous wavelet has greatly modified the shape, which becomes rather difficult to interpret as emission spectra. Notice, for instance the characteristic "wells" forming at the base of the lines. Even the coupling of CWT with the kSPCA approach, shown in Figure 3, suffers from the same drawbacks: in this case, though the score maps present good spatial information, the extracted loading vectors are difficult to read and present too many artefacts to be considered physical.

As shown, the various components of HyperPCA play a complementary role in the outcome. The DWTs are able to provide a sparse representation of the input and, at the same time, reduce part of the high frequency noise, as byproduct of the procedure, while retaining the physical information present in the datasets. At the same time, the kernel-based sparse PCA method helps to distinguish noise components from genuine signal, when the latter is sparse in a certain basis. Only the combination of both aspects in HyperPCA has the ability to automatically distinguish the sources of different signals and to disentangle the background noise, thus behaving as an unsupervised Artificial Intelligence, rather than a simple geometrical transformation of the input.

4. Small Elemental Map

In this section we highlight a specific effect of HyperPCA addressing the high dimensionality of LIBS mapping datasets. We show that the use of the kernel function is able to solve the theoretical phase transition issue in cases where the number of variables (wavelength channels) is greater than the number of available samples (spectra). In this case, it can be shown that PCA does not correctly reconstruct the directions of maximal variance [5], which ultimately leads to meaningless PCs.

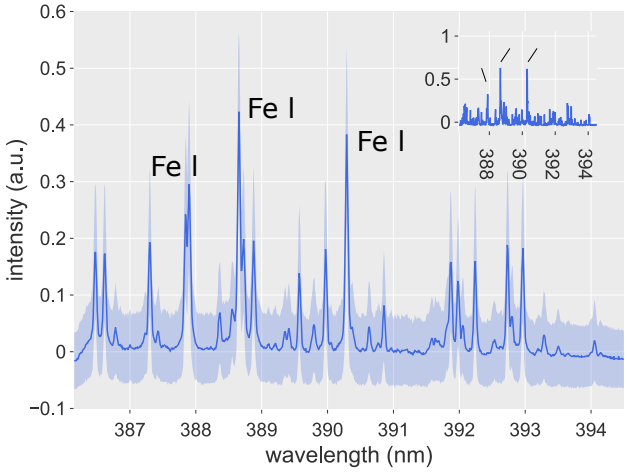


Figure 5: Preprocessed average and 1σ spectra with an example of single-shot spectrum of the 316L stainless steel sample.

The proposed dataset is a small LIBS map taken on a 316L stainless steel sample. In Figure 4 we show the reference elemental distributions as 200 µm × 150 µm images, taken at the SEM in a different location with respect to the LIBS map location, for the main elements. The average spectrum recorded during the LIBS mapping experiment is shown in Figure 5. The maps are mostly homogeneous on very small scales and different positions, and thus it is fair to assume the same composition all over the surface. Table 1 shows the relative concentration of the elements in the sample, estimated by EDS.

As the LIBS analysis of the sample is unable to provide the same resolution to detect possible substructures in the sample, we focus on a different aspect of the data analysis. We also do not provide the line intensities charts as they are not relevant to LIBS mapping in this case. We use a 3 µJ laser pulse on a 25 × 25 matrix on the surface, with a resolution of 10 µm in both directions. Craters are 3 µm both in diameter and depth on average. We use a grating with 2400 grooves/mm, centred at 391.914 nm. The goal of the analysis is to show that HyperPCA can extract meaningful information even below the theoretical threshold for extraction associated to PCA: in this case, the data matrix $X \in \mathbb{R}^{p \times n_\lambda}$ is such that $n_\lambda > p$, that is the number of columns in the dataset exceeds the number of rows. By RMT arguments, it is possible to show that the PCA reconstruction of the PCs is not reliable, and the distance between the true and extracted eigenvectors increases, when the signal-to-noise ratio $\beta^{-1} \lesssim \sqrt{q} = \sqrt{\frac{n_\lambda}{p}}$ [5, 6]. As in this case $q > 1$, we may expect issues in the reconstruction of the loading vectors.

In Figure 6 we show the most readable PCs extracted using the standard PCA. The loadings are all affected by a higher noise rate, even though the average spectrum in Figure 5 shows a strong signal presence: PC-1 is mostly unusable for element recognition tasks, while the line profiles in PC-3 and PC-4 present some positively reconstructed lines together with poorly resolved profiles. Although most of the lines in this wavelength range correspond to either Fe or Cr, some of them are almost undistinguishable and hardly resolved. As anticipated, the score matrices do not provide any information on the map, due to the homogeneity of the sample with LIBS mapping resolution.

In Figure 7 we show PCs extracted with HyperPCA as a proof of concept of the proposed pipeline. We use a kernel parameter $\alpha = 5.0 \times 10^2$ and the first level of the `bior3.5` wavelet filter banks. We impose a HT of 0.99 times the largest value of the power spectrum of the DCs. Even from the first ranked PC, the impact of the noise components on the resolution of the lines is reduced, thus increasing the quality of the loading vector associated to the largest eigenvalue of the covariance matrix: the profiles of the lines are generally better resolved than in standard PCA. Moreover, the absence of noise components in the first PCs is such to enable the resolution of principal and subdominant emission lines which may create spectral interference, such as the Fe lines in PC-3 at 388.63 nm and 390.65 nm, and PC-6 at 387.86 nm and 390.65 nm. The readability of the PCs is ensured also at low rank, such as PC-38 in the figure, where Cr at 391.92 nm

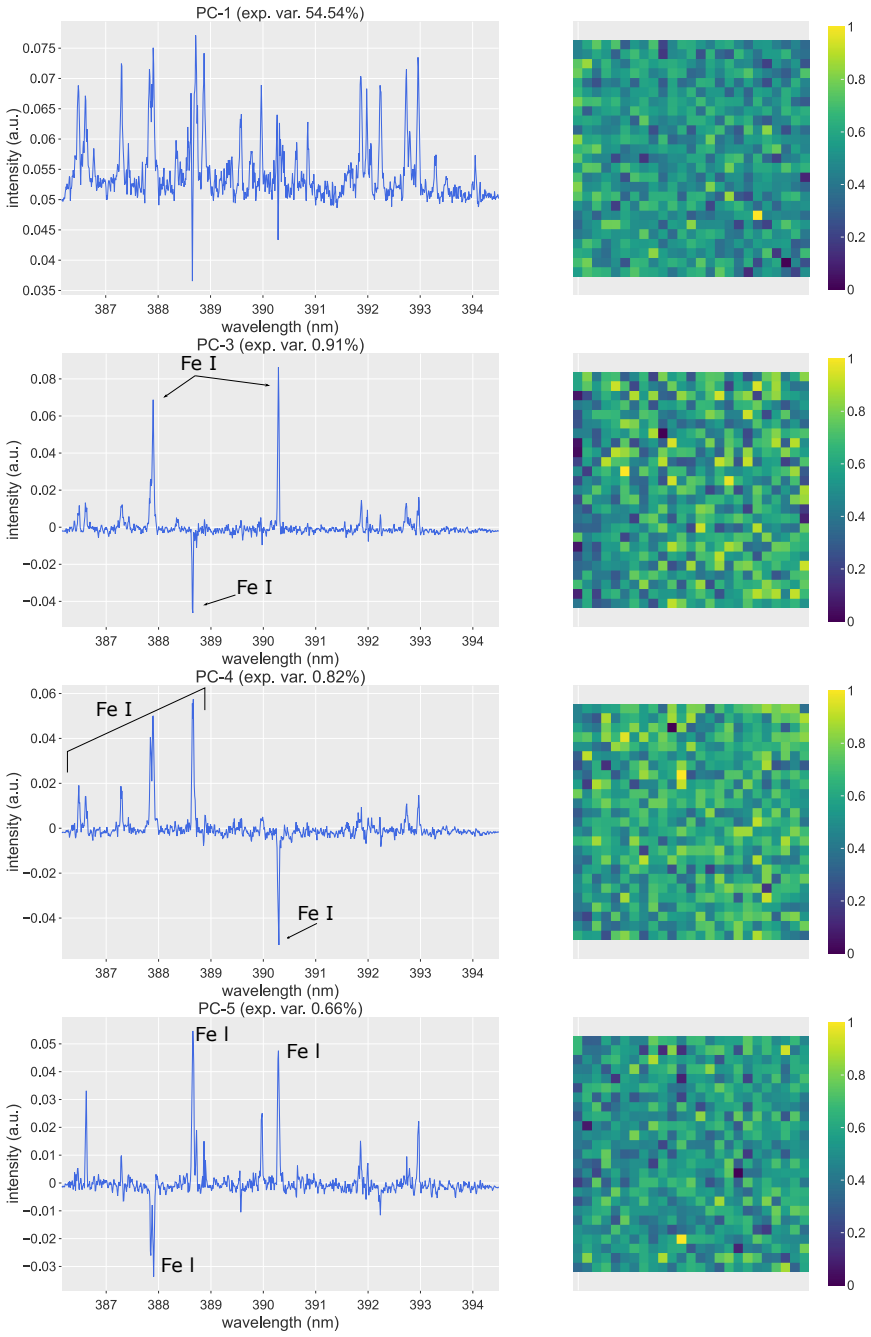


Figure 6: Reconstruction of the 316L stainless steel sample using standard PCA.

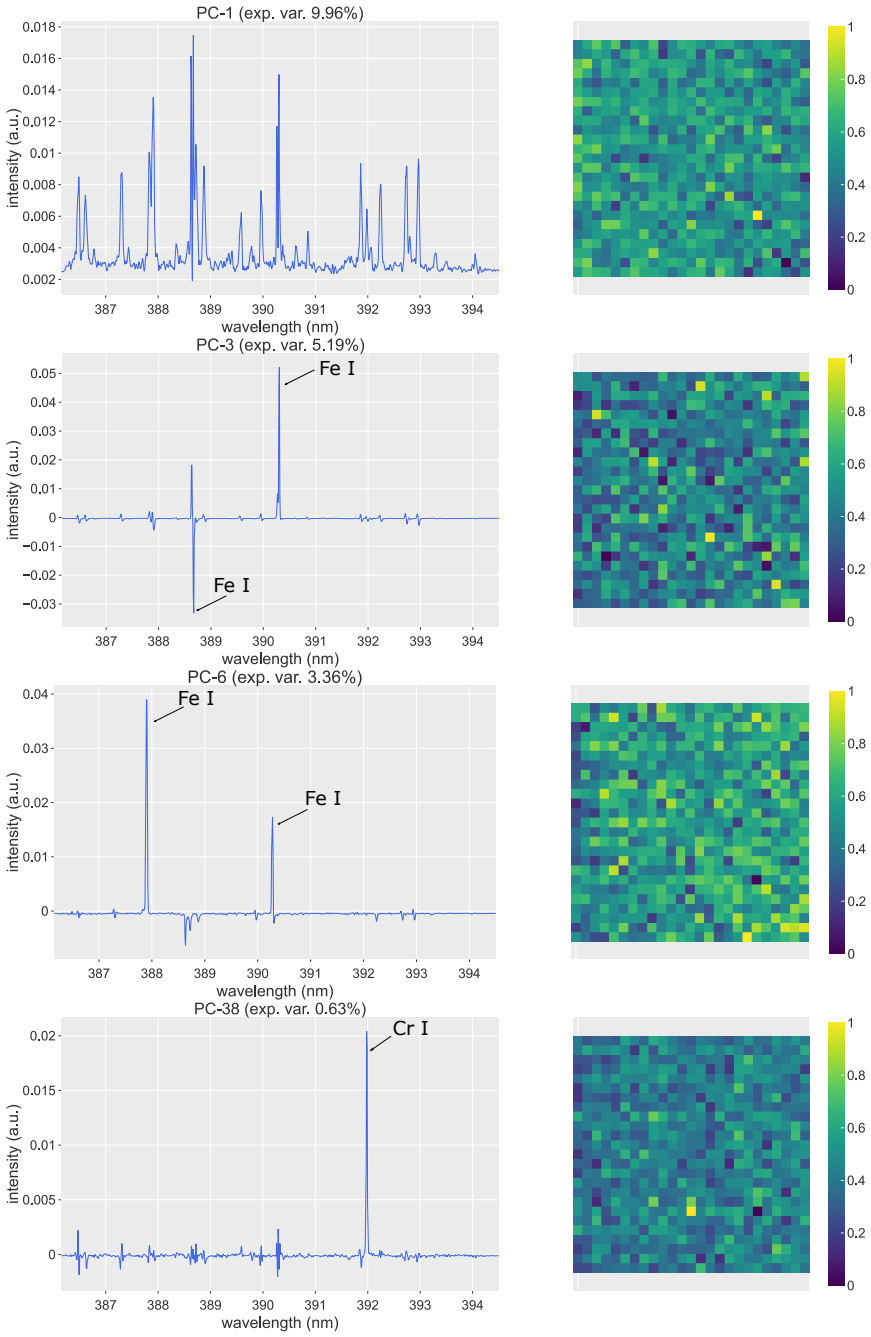


Figure 7: Reconstruction of the 316L stainless steel sample with HyperPCA.

is resolved free of Fe interference, thus also restoring some predictive power to the score map. Even though the quality of the score matrix is reduced by its small size, as the PC isolates Cr, the map can give an indication of the distribution of the element nonetheless. The ratios of the explained variance are also better estimated, such that a better assignment of the reconstruction confidence can be assigned to each PC.

HyperPCA can thus deal also with extreme cases in which the experimental framework may impose strong restrictions, such as small maps, or noisy data, usually defined in RMT as a spiked model where the separation between the largest eigenvalue is not strongly distinct from the bulk. Such situations are usually connected to, but not limited to, cases in which the transition threshold $n_\lambda = p$ is crossed, in which case the noise gains a larger importance with respect to signal. The use of the kernel approach ensures the ability to go beyond the theoretical bounds of standard PCA, thus restoring the readability of the PCs. In addition, beyond the case of LIBS mapping, it is interesting to note that the situation where the number of channels exceeds the number of spectra ($n_\lambda > p$) is very frequent in conventional LIBS, where PCA is often applied to explore small datasets. As shown in this section, standard PCA is therefore limited if the signal-to-noise ratio is low. HyperPCA provides a relevant alternative to extract more meaningful information.

5. Complementary Figures

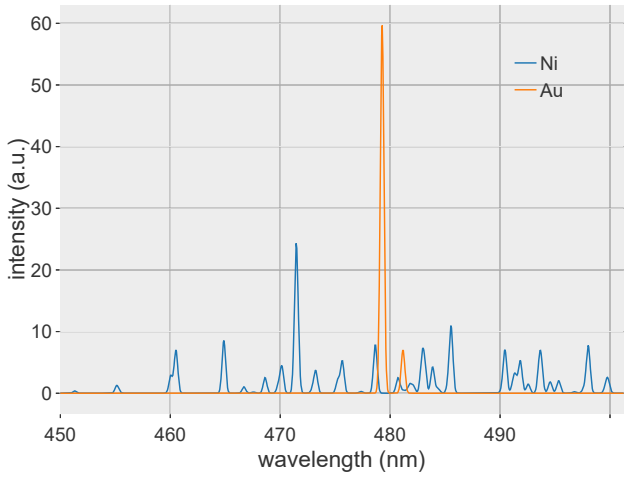


Figure 8: NiAu spectra in LTE.

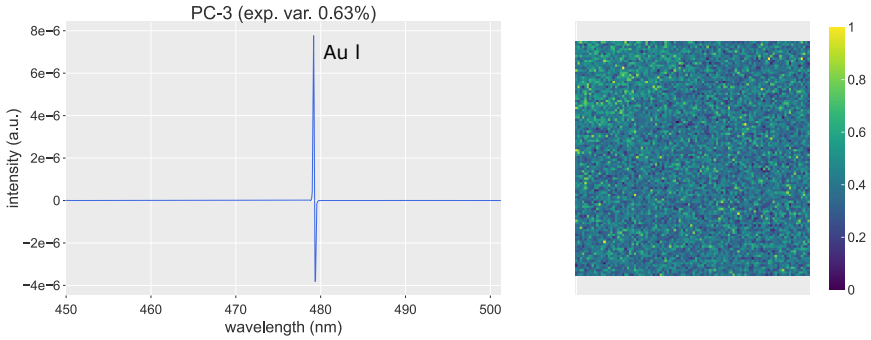


Figure 9: Due to experimental noise, the most intense photon emission can be captured by the CCD sensor in different neighbouring pixels, thus leading to a slightly different shape of the line profile. This is sometimes extracted as a *first derivative* signal by PCA or, in this case, by HyperPCA.

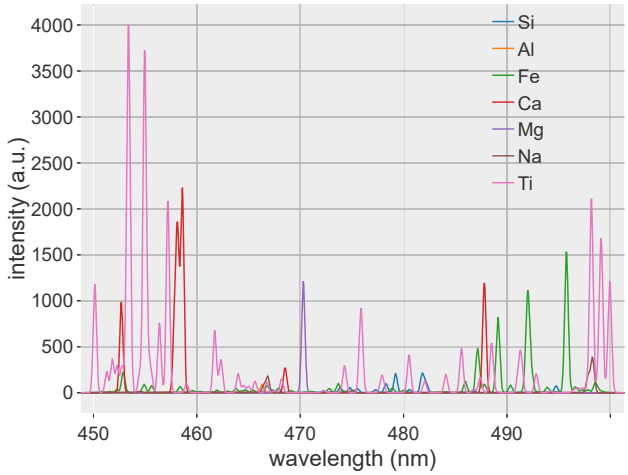


Figure 10: “Basalt” spectra in LTE.

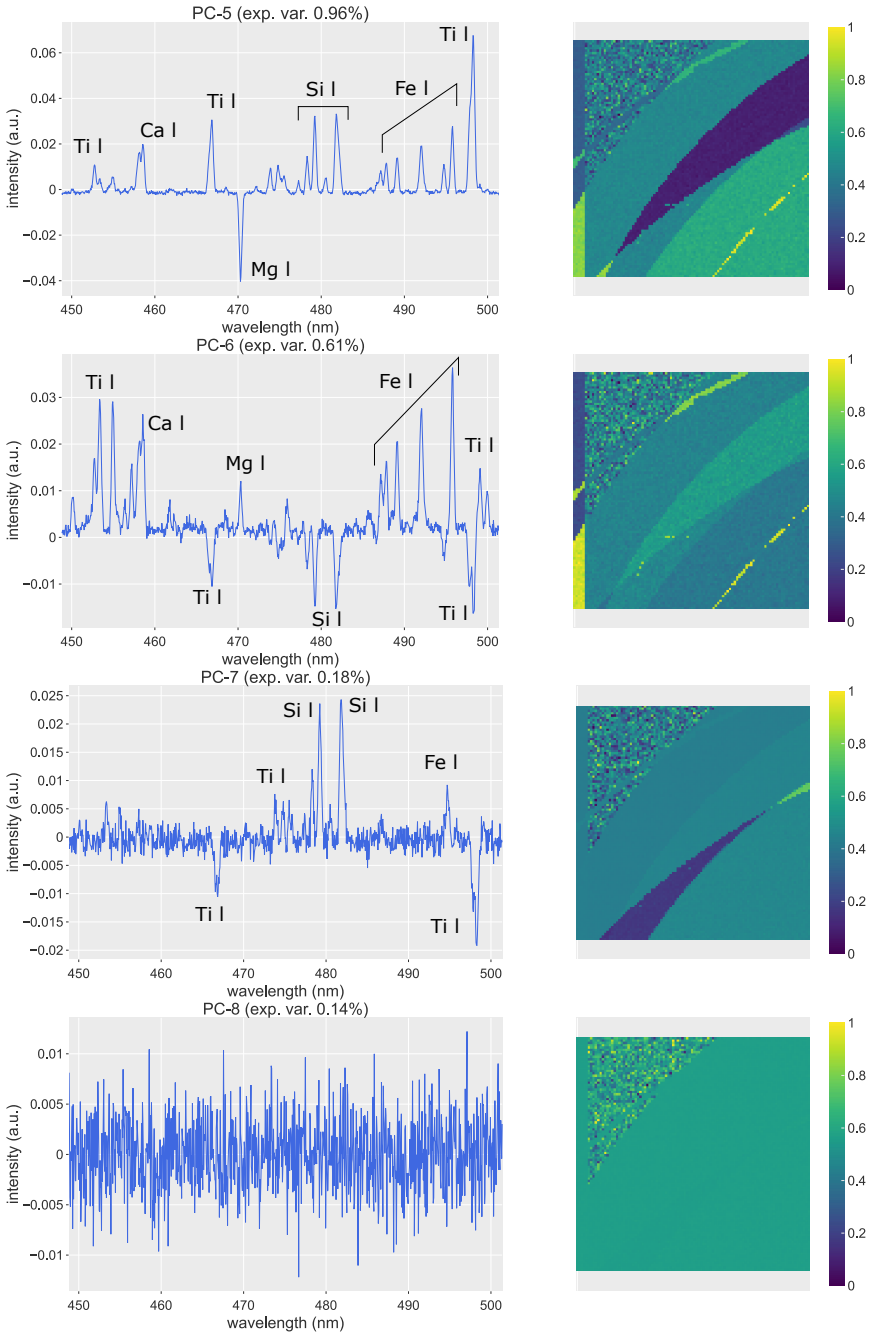


Figure 11: Complementary PCs of the "basalt" specimen using the standard PCA.

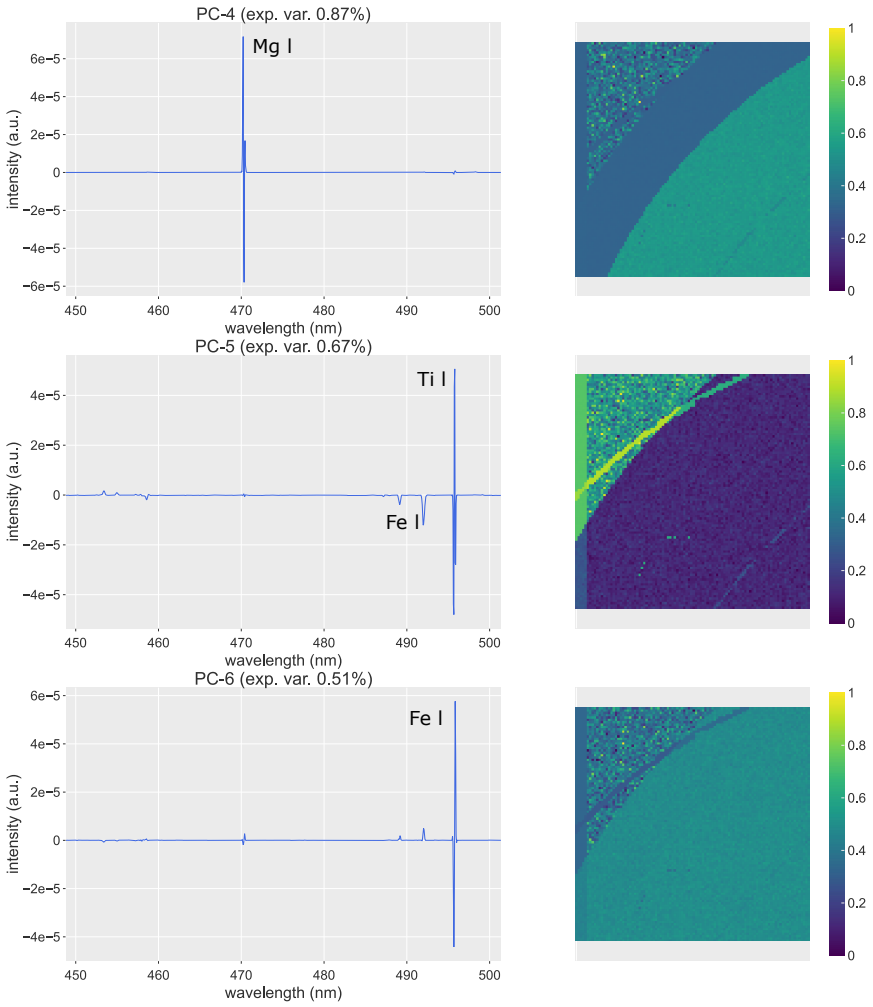


Figure 12: Complementary PCs of the "basalt" specimen using HyperPCA (part 1).

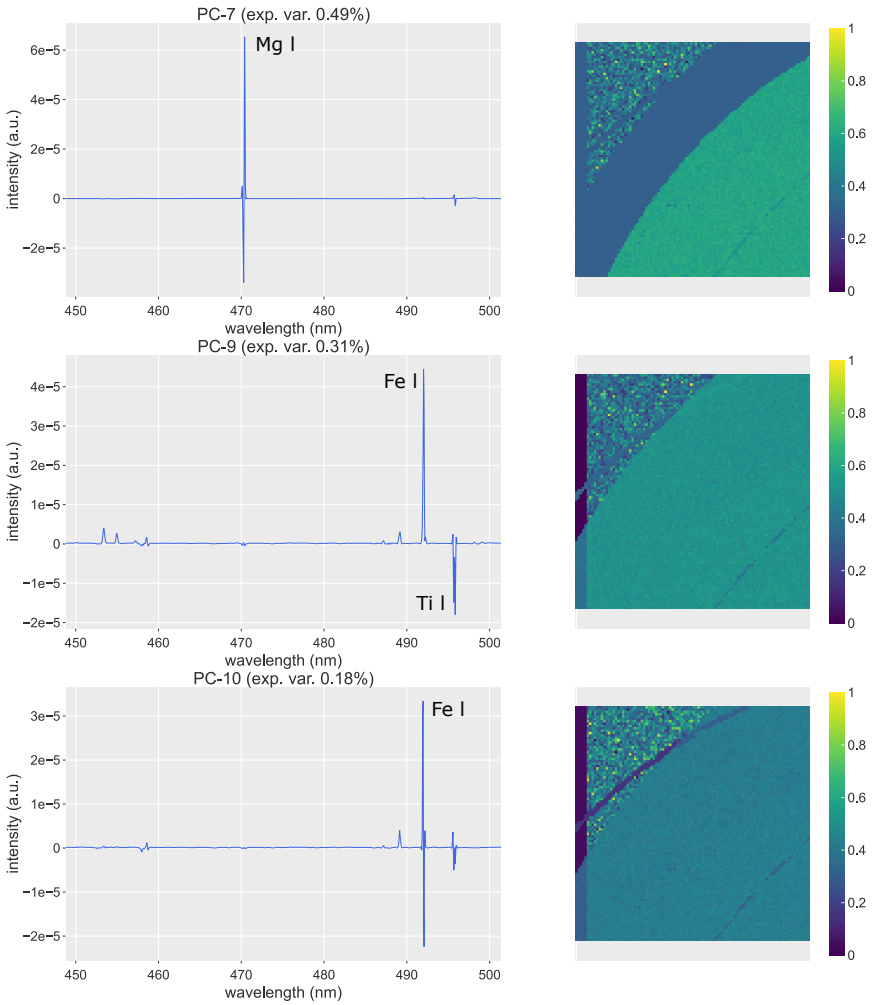


Figure 13: Complementary PCs of the "basalt" specimen using HyperPCA (part 2).

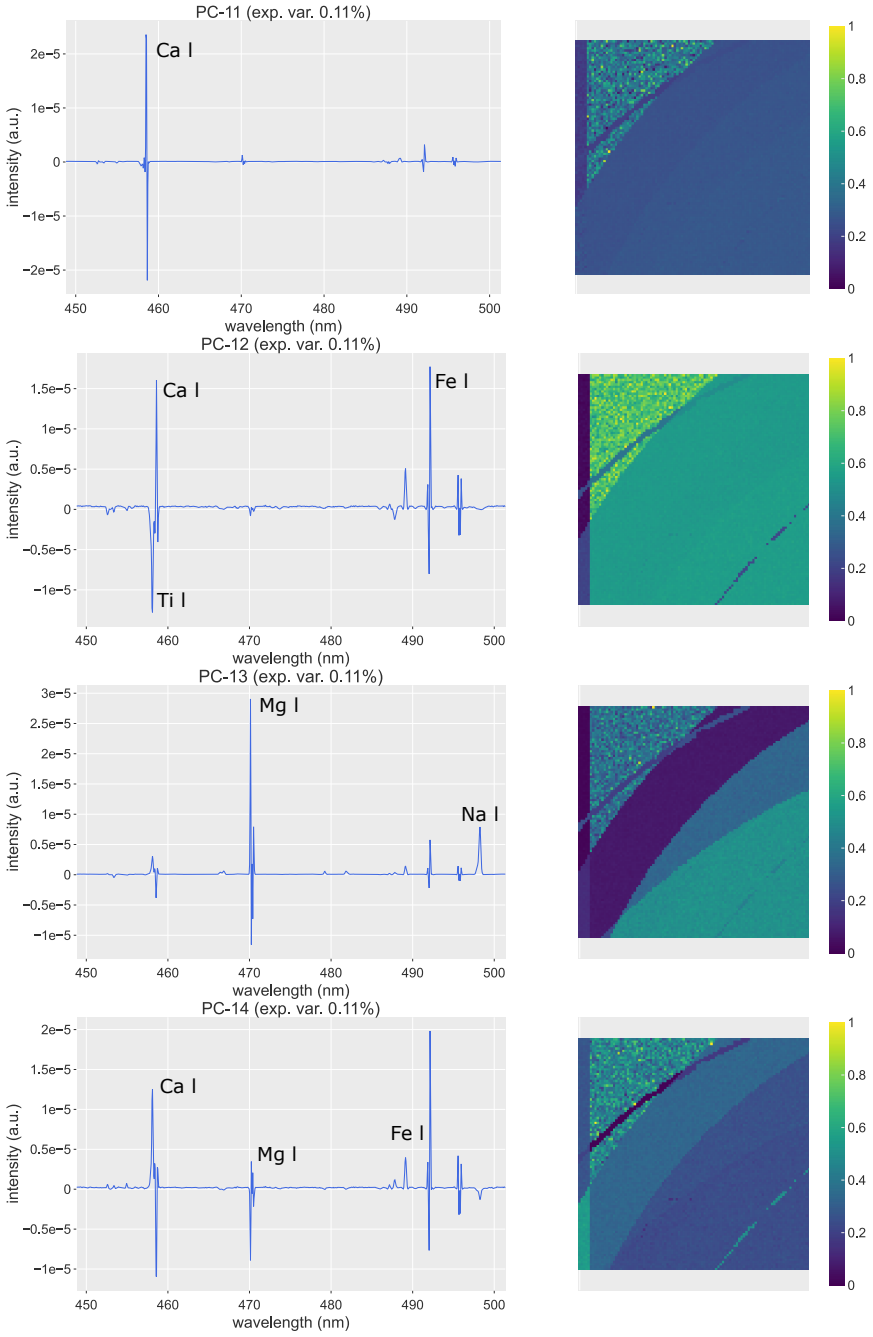


Figure 14: Complementary PCs of the "basalt" specimen using HyperPCA (part 3).

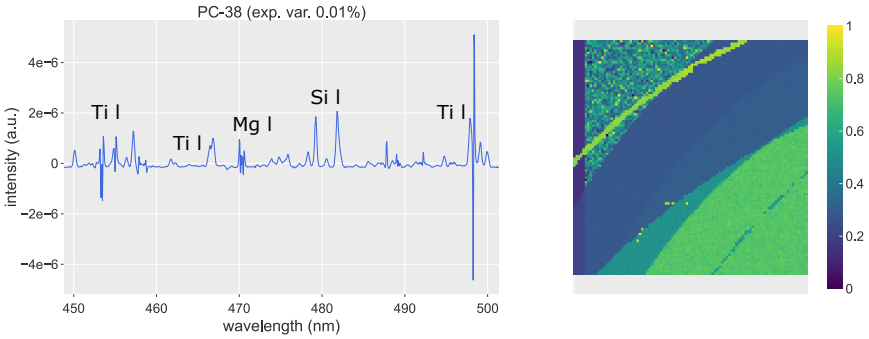


Figure 15: Complementary PCs of the "basalt" specimen using HyperPCA (part 4).

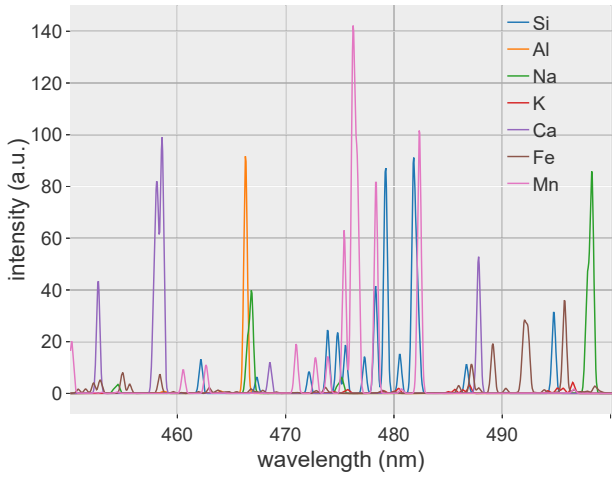


Figure 16: "Granite" spectra in LTE.

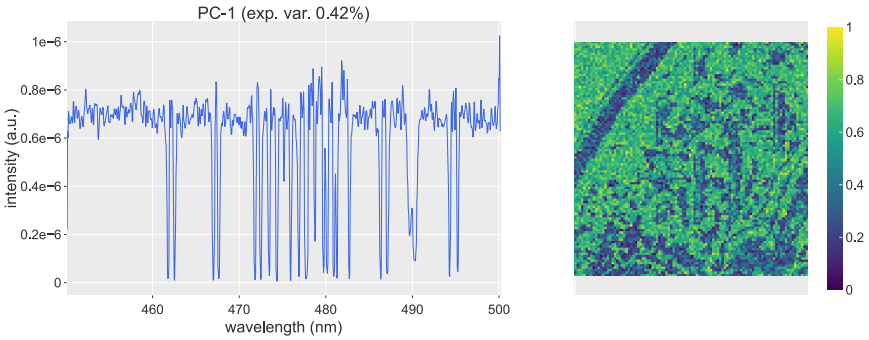


Figure 17: First PC of the "granite" dataset computed using HyperPCA: the noise present in the dataset deeply spoils the first PC of both PCA and HyperPCA rendering it mostly unusable.

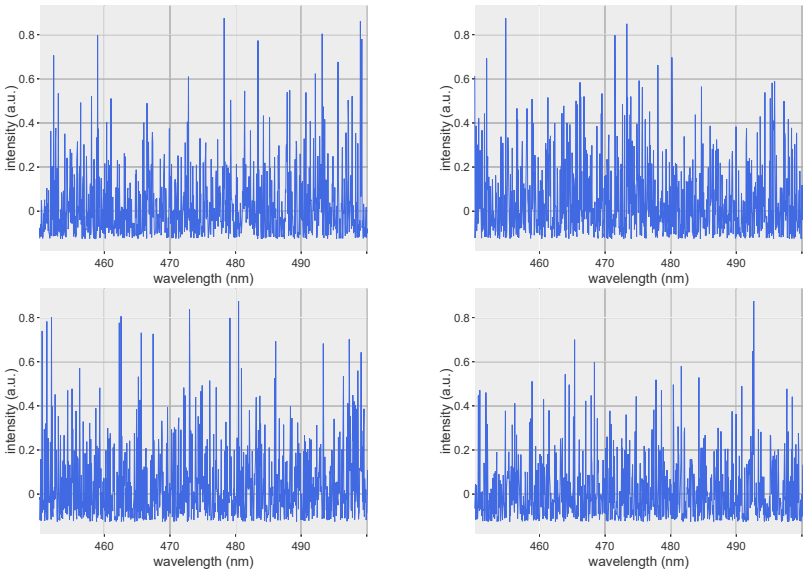
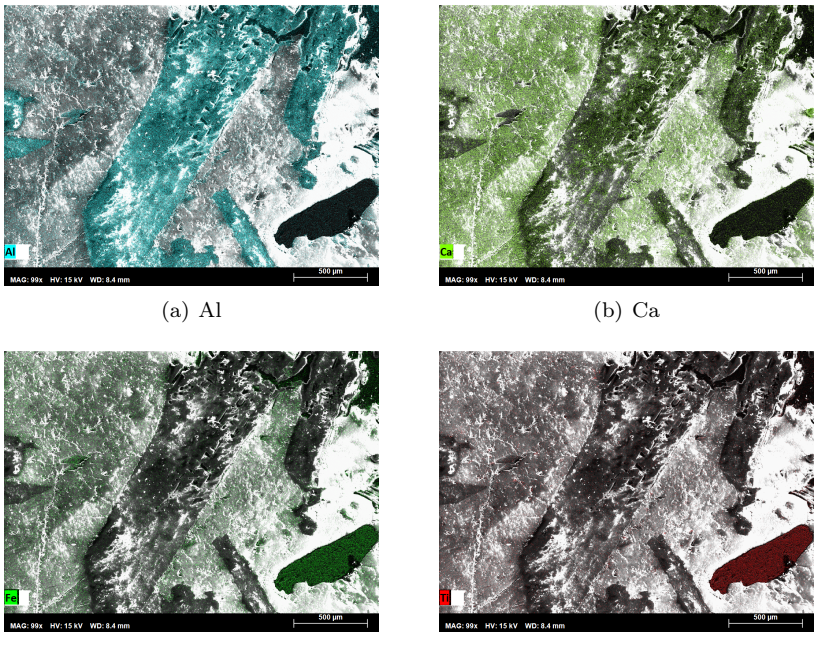


Figure 18: Example spectra in the "granite" dataset which illustrate the difficulty in extracting the signal due to the high noise level.



(a) Al (b) Ca

(c) Fe (d) Ti

Figure 19: Distributions of the main elements in the gabbro specimen via SEM analysis.

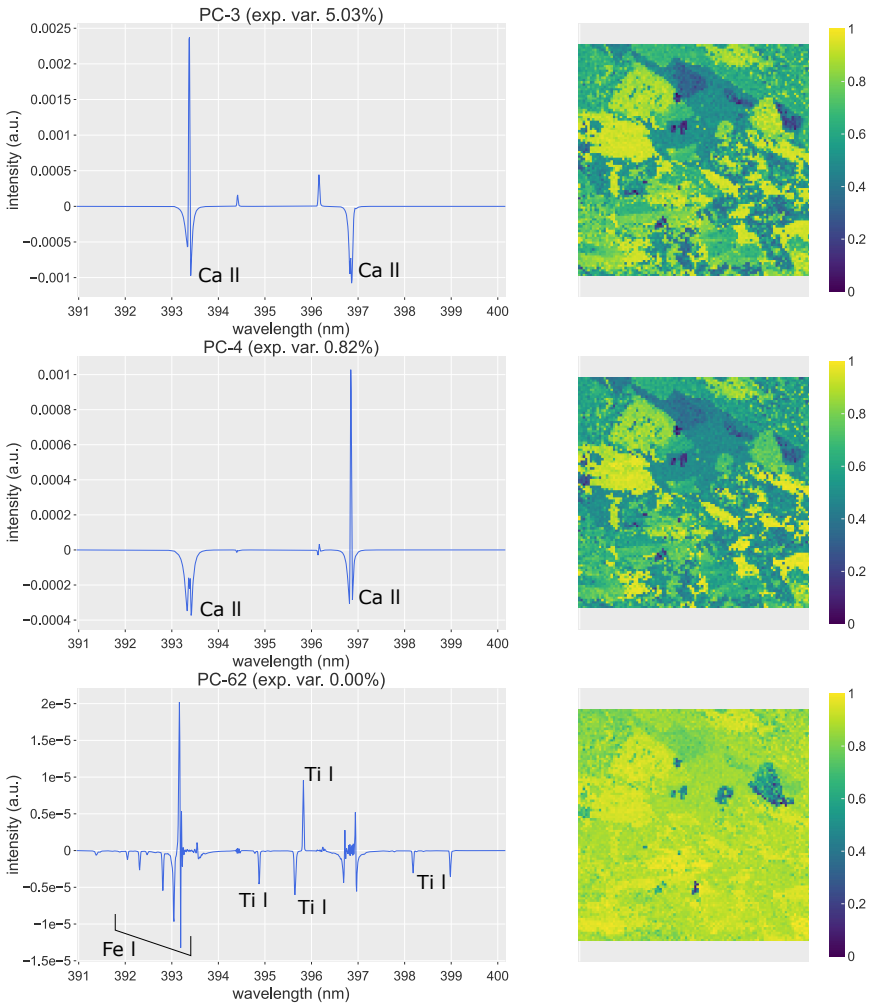


Figure 20: Complementary PCs of the gabbro specimen using HyperPCA.

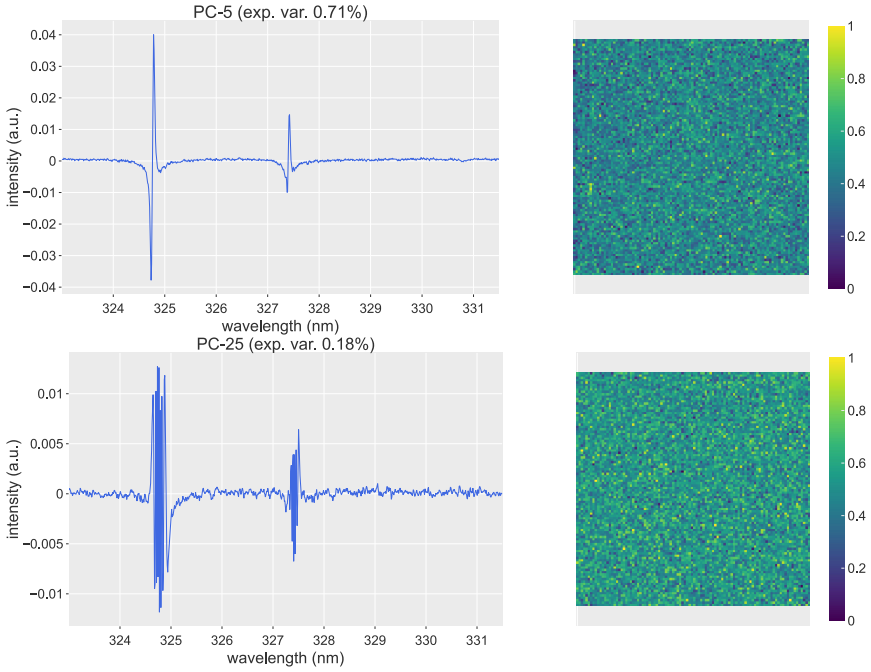


Figure 21: Complementary PCs of the AlCu sample using the standard PCA.

References

[1] K. Pearson, On lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2 (1901) 559–572. doi:10.1080/14786440109462720.

[2] R. A. Horn, C. R. Johnson, Matrix analysis, second edition, corrected reprint ed., Cambridge University Press, New York, NY, 2017.

[3] I. Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, 1992. doi:10.1137/1.9781611970104.

[4] J. R. Williams, K. Amaratunga, Introduction to wavelets in engineering, International Journal for Numerical Methods in Engineering 37 (1994) 2365–2388. doi:10.1002/nme.1620371403.

[5] J. Baik, G. B. Arous, S. Peche, Phase transition of the largest eigenvalue for non-null complex sample covariance matrices, Annals of Probability 33 (2005) 1643–1697. doi:10.1214/009117905000000233, arXiv: math/0403022.

[6] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, Statistica Sinica (2007) 1617–1642. URL: https://www.jstor.org/stable/24307692, publisher: JSTOR.