Assessment of Essential Information in the Fourier Domain to Accelerate Raman Hyperspectral Microimaging

Laureen Coic, Raffaele Vitale, Myriam Moreau, David Rousseau, José Henrique de M Goulart, Nicolas Dobigeon, Cyril Ruckebusch

analytical chemistry

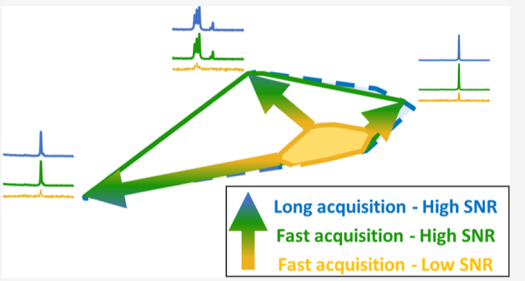
pubs.acs.org/ac
Article

Assessment of Essential Information in the Fourier Domain to Accelerate Raman Hyperspectral Microimaging

Laureen Coic, Raffaele Vitale, Myriam Moreau, David Rousseau, José Henrique de Morais Goulart, Nicolas Dobigeon, and Cyril Ruckebusch*

Cite This: *Anal. Chem.* 2023, 95, 15497−15504
Read Online

ABSTRACT: In the context of multivariate curve resolution (MCR) and spectral unmixing, essential information (EI) corresponds to the most linearly dissimilar rows and/or columns of a two-way data matrix. In recent works, the assessment of EI has been revealed to be a very useful practical tool to select the most relevant spectral information before MCR analysis, key features being speed and compression ability. However, the canonical approach relies on the principal component analysis to evaluate the convex hull that encapsulates the data structure in the normalized score space. This implies that the evaluation of the essentiality of each spectrum can only be achieved after all the spectra have been acquired by the instrument. This paper proposes a new approach to extract EI in the Fourier domain (EIFD). Spectral information is transformed into Fourier coefficients, and EI is assessed from a convex hull analysis of the data point cloud in the 2D phasor plots of a few selected harmonics. Because the coordinate system of a phasor plot does not depend on the data themselves, the evaluation of the essentiality of the information carried by each spectrum can be achieved individually and independently from the others. As a result, time-consuming operations like Raman spectral imaging can be significantly accelerated exploiting a chemometric-driven (i.e., based on the EI content of a spectral pixel) procedure for data acquisition and targeted sampling. The usefulness of EIFD is shown by analyzing Raman hyperspectral microimaging data, demonstrating a potential 50-fold acceleration of Raman acquisition.



INTRODUCTION

Raman microimaging is a well-established technique and has demonstrated great potential for many applications to be found in pharmacology,[1] geology,[2] and life sciences[3] to name a few. However, the mapping process using point-by-point raster scanning of a laser spot through the sample can be terribly slow for spontaneous Raman of weakly scattering systems, with acquisition time of seconds per pixel turning into hours or more to scan a complete sample and get a full image. In contrast, fast spectral mapping may be required in many applications such as quality control,[4] falsified medicine analysis,[5] and biomedical analysis,[6] or to avoid damaging laser exposure time. Different approaches can be used to speed up spontaneous Raman microimaging acquisition. Some are based on instrumental aspects such as line scanning,[7−10] whereas others somehow exploit the properties of the sample itself and notably the presence of spatial features[11−13] to speed up acquisition in the region where little spatial information is observed. One recent illustration of the potential of these approaches was provided by a compressive sensing strategy using a context-aware image prior to improve imaging speed.[14] However, while the combination of chemometrics and Raman spectroscopy is now mainstream in most application fields,[15,16] the use of the chemometric information itself to adjust the acquisition time at each individual pixel based on the observed spectral relevance has not been exploited in the scientific literature to the best of the authors' knowledge.

These observations have motivated us to investigate the possibility of speeding up Raman imaging taking advantage of the so-called essential information (EI), a concept of deep relevance in chemometrics since EI corresponds to the most linearly dissimilar rows and/or columns of a data matrix.[17] Such rows/columns are called essential because they are indispensable to reproduce the full data matrix in a convex linear way. The conceptualization of EI dates back in both the geometric construction of the inner polytope in the multivariate curve resolution (MCR)[18] and in the field of archetypal analysis[19,20] in statistics. On the application side, the selection of EI has proven to be a very useful, significant, and reliable data reduction tool, driven by the properties of linear spectral mixtures, to select the most relevant spectral information before MCR analysis.[21,22] The selection of EI brings advantages in terms of speed and compression and allows to tackle challenging analytical

Received: March 30, 2023
Accepted: July 13, 2023
Published: October 11, 2023

https://doi.org/10.1021/acs.analchem.3c01383
Anal. Chem. 2023, 95, 15497−15504

issues.[17,23−28] In practice, EI is found by computation of the convex hull that encapsulates the data structure in the normalized abstract scores space derived from a truncated singular value decomposition.[29] The score points supporting the vertices of the convex hull correspond to essential rows/columns of the original data matrix. They carry essential samples (ESs) and essential variables, respectively. Note that with spectral imaging data, ESs were originally called essential spectral pixels;[22] however, it should be clear that no information about spatial features or the position in the image is used in their selection. It should be added that the influence of noise on the selection of EI and on the stability of the solutions provided by MCR analysis has been discussed in very recent works.[21,24,25]

The calculation of EI requires dimensionality reduction. By calculating the convex hull of a data matrix in a lower-dimensional principal component analysis (PCA) subspace (normalized scores space), one exploits geometrical properties of convex polytopes.[17] Nevertheless, the application of PCA induces some limitations and prevents a more efficient use of EI for practical purposes of fundamental analytical importance. Most important of these limitations comes from the fact that the PCA coordinate system within which to evaluate EI by convex hull calculations can only be determined after all the spectral data have been acquired by the instrument. This closes the door to a per-spectral pixel evaluation of EI. However, not only factorial representations such as PCA but any linear transformation can be used to evaluate EI.[17] We here propose to utilize a ubiquitous technique in signal processing, namely the discrete Fourier transform (DFT) to achieve this goal.

Fourier transform allows breaking a signal (a spectrum) into its frequency components. Computations can be performed very efficiently using the fast Fourier transform algorithm,[30] providing signal representations characterized by sine and cosine functions of varying frequencies. In the Fourier domain, at a given frequency, the Fourier coefficients will carry information about the linear mixture coefficients associated with the spectra of the individual components (unless a spectrum does not contribute at this frequency) and can be represented in a so-called phasor plot.[31−35] By performing convex hull calculations on 2D phasor plots, we provide an approach to estimate EI in the Fourier domain (EIFD) which, we believe, has strong advantages for analytical purposes. Indeed, there is neither the need to choose the dimensionality of the PCA subspace to work with nor the requirement to estimate the number of mixture components. And most importantly, since DFT is performed spectrum-wise, the evaluation of EIFD can be performed independently for each spectral pixel and, thus, potentially as the data acquisition process is ongoing (e.g., for sequential or point-by-point scanning systems). This paves the way to image acquisition processes that would be based on the EI content of a spectral pixel (i.e., chemometrically driven), speeding up the acquisition at those locations where information is not essential and, subsequently, considerably accelerating the imaging process.

In this paper, we aim to propose and illustrate the concept of EIFD. We first work with simulations and then perform EIFD analysis of Raman microimaging data of a ternary mixture tablet. Finally, we emphasize the potential of the approach to drastically reduce the Raman microimaging acquisition time.

MATERIALS AND METHODS

EI Evaluated on Fourier Coefficients. Let $\mathbf{D} = [\mathbf{d}_1, ..., \mathbf{d}_N]^T \in \mathbb{R}_+^{N\times L}$ denote[a] the data matrix where each row $\mathbf{d}_n \in \mathbb{R}^L$ corresponds to the L-dimensional spectrum collected at the n-th spatial position ($n = 1, ..., N$). According to the linear mixture framework, each measured spectrum $\mathbf{d}_n$ is assumed to be approximated by the linear combination of K elementary spectra collected in the matrix $\mathbf{S}^T = [\mathbf{s}_1, ..., \mathbf{s}_K]^T \in \mathbb{R}_+^{K\times L}$, i.e.,

$$\mathbf{d}_n = \sum_{k=1}^{K} c_{nk}\mathbf{s}_k + \mathbf{e}_n \tag{1}$$

where c_{nk} is the mixture coefficient of the k-th mixture component in the n-th measured spectrum and $\mathbf{e}_n$ is an error term accounting for any mismodeling error and measurement noise. In matrix notation, this model is written as

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \tag{2}$$

with $\mathbf{C} = (c_{nk}) \in \mathbb{R}_+^{N\times K}$ and $\mathbf{E} = (e_{nl}) \in \mathbb{R}^{N\times L}$. For the sake of simplicity, we will assume for the moment that, for each n, the nonnegative coefficients c_{nk} satisfy $\sum_{k=1}^{K} c_{nk} = 1$, implying that $\mathbf{d}_n$ is (approximately) a convex combination of the spectra $\mathbf{s}_k$.

Thanks to the linearity property of the DFT, the linear model in 1 easily translates into the Fourier (frequency) domain as

$$\tilde{\mathbf{d}}_n = \sum_{k=1}^{K} c_{nk}\tilde{\mathbf{s}}_k + \tilde{\mathbf{e}}_n \tag{3}$$

where $\tilde{}$ denotes the DFT of the corresponding quantity, i.e.,

$$\tilde{\mathbf{d}}_n = [\tilde{d}_{n1}, ..., \tilde{d}_{nL}]^T = \mathrm{DFT}[\mathbf{d}_n] \tag{4}$$

$$\tilde{\mathbf{s}}_k = [\tilde{s}_{k1}, ..., \tilde{s}_{kL}]^T = \mathrm{DFT}[\mathbf{s}_k] \tag{5}$$

In particular, the coefficients of these DFT are defined as

$$\tilde{d}_{nm} = \sum_{l=1}^{L} d_{nl}\, e^{-j2\pi/L(m-1)(l-1)} \tag{6}$$

$$\tilde{s}_{mk} = \sum_{l=1}^{L} s_{lk}\, e^{-j2\pi/L(m-1)(l-1)} \tag{7}$$

with $m = 1, ..., L$ harmonics and j is the imaginary number for which $j^2 = -1$.

Since such coefficients are complex-valued, by using 3 and 6−7, we can write their real and imaginary parts as

$$G_m(n) := \Re[\tilde{d}_{nm}] = \sum_{k=1}^{K} c_{nk}\Re[\tilde{s}_{mk}] + \Re[\tilde{e}_{nm}] \tag{8}$$

$$S_m(n) := \Im[\tilde{d}_{nm}] = \sum_{k=1}^{K} c_{nk}\Im[\tilde{s}_{mk}] + \Im[\tilde{e}_{nm}] \tag{9}$$

By the inverse DFT, the real numbers $G_m(n)$ and $S_m(n)$ can be seen as coordinates of the m-th phasor involved in the Fourier representation of $\mathbf{d}_n$

https://doi.org/10.1021/acs.analchem.3c01383
Anal. Chem. 2023, 95, 15497−15504

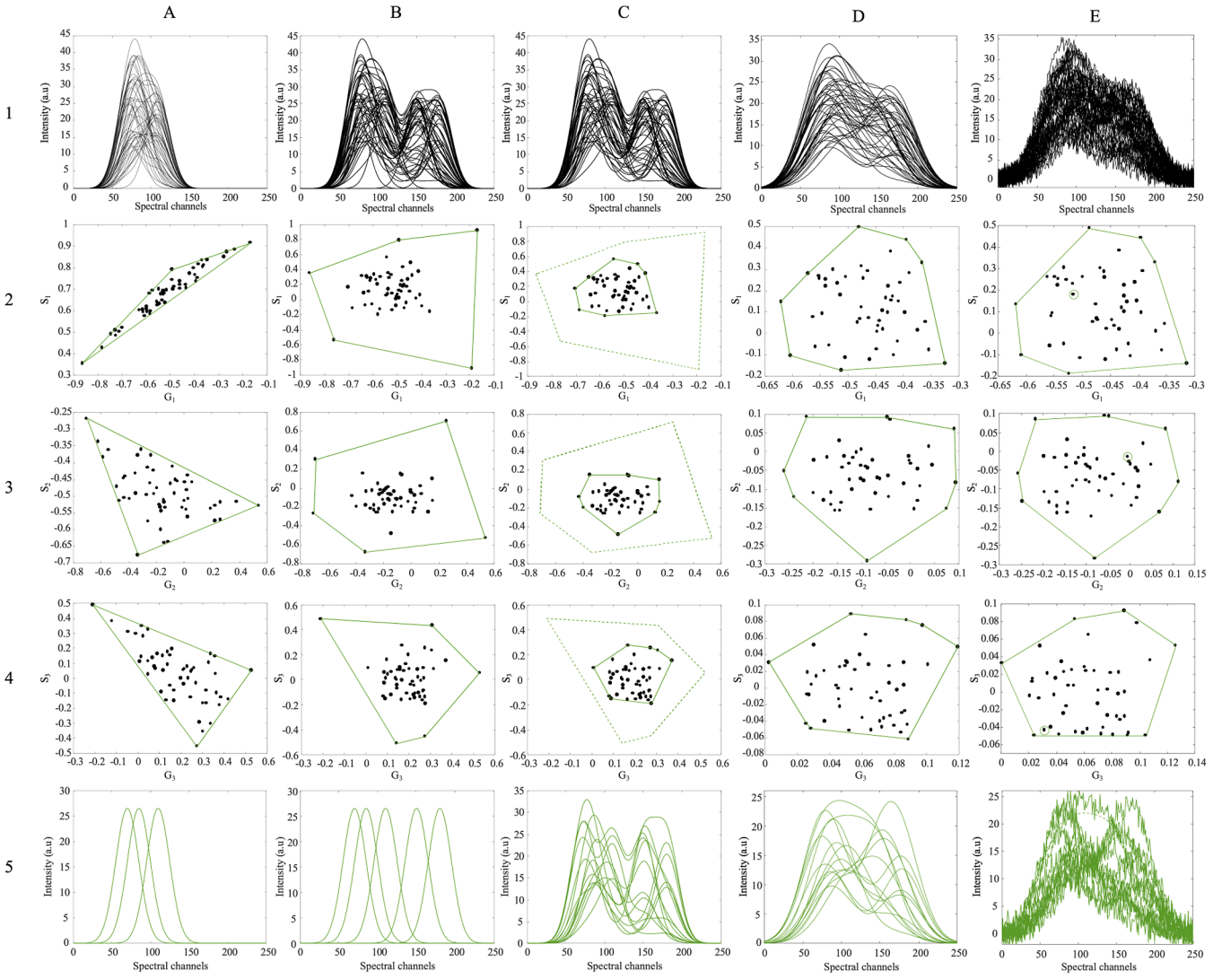


Figure 1. Results obtained applying EIFD to datasets $\mathbf{D}_{1-5}$, distributed over columns A−E, respectively. Simulated spectra are represented along line 1 and phasor plots obtained for $m = 1, 2, 3$ along lines 2−4. Convex hulls are displayed in green; green dashed lines (column C) mark the location of the convex hull obtained from B2. ESs are provided along line 5. The circled points correspond to the ESs not found in $\mathbf{D}_5$.

$$d_{nl} = \sum_{m=1}^{L} \tilde{d}_{nm}\, e^{j2\pi/L(m-1)(l-1)} \quad (10)$$

$$= \sum_{m=1}^{L} (G_m(n) + jS_m(n))\, e^{j2\pi/L(m-1)(l-1)} \quad (11)$$

The crux of our approach relies on the observation that according to 8 and 9, the phasor coordinates $G_m(n)$ and $S_m(n)$ of $\tilde{d}_{nm}$ are (approximately) convex combinations of those of the spectra $\tilde{s}_{km}$, with $m = 1, ..., M$ and $M \leq L$.

Hence, one can retrieve the EI (i.e., the most linearly dissimilar spectra) by computing the convex hull of the data cloud comprising all points $(G_m(n), S_m(n))$, for $n = 1, ..., N$, provided that at least one spectrum $\mathbf{s}_k$ has nonzero (in practice, significant) DFT coefficients. By repeating this reasoning for a few harmonics, a set of candidate ESs can be determined (see, e.g., Figure 1).

In practice, the nonnegative coefficients c_{kn} do not necessarily satisfy $\sum_{k=1}^{K} c_{kn} = 1$ as we have previously assumed. Nevertheless, as detailed in the Supporting Information, this constraint[36,37] can be (approximately) imposed by simply normalizing the data as

$$\bar{\mathbf{d}}_n = \frac{1}{\|\mathbf{d}_n\|_1}\mathbf{d}_n \quad (12)$$

where $\|\cdot\|_1$ denotes the l_1-norm of a vector, that is the sum of the absolue values of its components.

Datasets. *Simulated Datasets.* To illustrate the concept of EIFD and the methodology described in the previous section, 5 datasets were simulated corresponding to different scenarios. The first dataset ($\mathbf{D}_1$) was obtained by considering 3 Gaussian-like pure spectra (FWHM = 35) sampled over $L = 201$ spectral points and their associated mixture coefficients, comprising $N = 53$ spectra. Specifically, 50 linear mixture spectra were generated by taking random coefficients distributed from a uniform distribution in the interval [0, 1], and the 3 individual pure spectra were added to complete the dataset. The second dataset ($\mathbf{D}_2$) was built in the same fashion, but now considering 50 mixtures of 5 Gaussian-like pure spectra to which were added the 5 pure spectra. The third dataset ($\mathbf{D}_3$) consists of 50 mixture spectra, the same as the one of $\mathbf{D}_2$, but without the pure spectra.

https://doi.org/10.1021/acs.analchem.3c01383
Anal. Chem. 2023, 95, 15497−15504

The fourth dataset ($\mathbf{D}_4$) was built by considering the same mixture coefficients as for $\mathbf{D}_3$, but the width of the Gaussians used to build pure spectra was now broadened (FWHM = 58). Finally, white Gaussian noise was added to the spectra of $\mathbf{D}_4$ with a noise power of 1 dBW to obtain the dataset $\mathbf{D}_5$.

Raman Hyperspectral Imaging Data. A three-component system was obtained by mixing powders of calcium carbonate ($CaCO_3$), sodium nitrate ($NaNO_3$), and sodium sulfate (Na_2SO_4) and then pressing a tablet. Image acquisition was performed on a LabRAM HR microspectrometer (Horiba France SAS, Palaiseau, France) using a 50× Olympus objective (0.75 NA). The spectrometer is equipped with a 600 g/mm grating. For excitation, a 632.8 nm HeNe laser was used (15 mW laser power at the sample). In order to speed up the acquisition time, a narrow spectral range was considered, from 497.7 to 901.2 cm^{-1}, which corresponds to the spectral zone where bending and symmetric vibrations of SO_4^{2-}, CO_3^{-}, and NO_3^{-} groups occur. A 101 × 101 pixel image was mapped using point-by-point raster scanning, with a spatial step of 1 μm. Two different accumulation times were considered, 100 ms per pixel (low-SNR spectra) and 5 s per pixel (high-SNR spectra). This resulted in 2 datasets denoted as $\mathbf{D}_{low}$ and $\mathbf{D}_{high}$ (of dimensions 101 × 101 × 343) for low- and high-SNR data, respectively. For the sake of generality, data were also acquired for the spectral range 71.6−496.5 cm^{-1} which contains information about relative translations between cations and anions of the three species under study.

Software. All computations were performed using MATLAB 2016a (Mathworks, Natick, MA) and the PLS Toolbox/MIA Toolbox (version 8.6.2, Eigenvector Research, Inc, Wenatchee, USA). The in-house routine built for the computation of EIFD is provided in the Supporting Information code section along with the data used to construct all the figures.

RESULTS AND DISCUSSION

We first evaluated the proposed EIFD approach on simulated data. Figure 1 shows the results obtained for the datasets $\mathbf{D}_{1-5}$. For each dataset, the results obtained by evaluating the convex hull on the data point cloud corresponding to the DFT coefficients for m = 1, 2, 3 are reported, and the corresponding ESs are provided. The results obtained for datasets $\mathbf{D}_{1-2}$ enable highlighting situations for which pure spectra are available. These data correspond to a three- and a five-component spectral mixture, respectively. Focusing first on $\mathbf{D}_1$, the results obtained are provided in the vertical panels of Figure 1A. The simulated data are in Figure 1A(1), whereas the phasor plots of the data cloud are provided in Figure 1A(2−4). The convex hull was calculated for each of the corresponding data and then drawn with a green line. As expected, 3 ESs (actual pure spectra—see Figure 1A(5)) are identified supporting the convex hull. In the absence of noise or sampling artifacts, and unless a pure spectrum has no contribution at a given harmonic, this interpretation does hold for every set of DFT coefficients. Two comments should be made at this step. One is that any data point of coordinates $(G_m(n), S_m(n))$—see 8 and 9—lying inside the convex hull can be obtained by convex linear combinations of the ESs identified by EIFD. Another is that with a three-component system, the data geometry observed for each phasor plot corresponds to a simplex.

Moving to the results obtained for the EIFD analysis of $\mathbf{D}_2$, now a five-component mixture dataset with pure spectra present, it is interesting to note that the discussion above still holds, that is, the data geometry obtained in Figure 1B(2−4) is convex, and ESs correspond to pure spectra, as shown in Figure 1A(5). Whether there are 3, 5, or more components in the spectral mixture, ESs can be identified in the phasor plots of the different DFT coefficients considered. This is a key aspect: identifying EI can be achieved without requiring any preconceived idea about the number of components contributing to the mixture spectral data. One could argue that when considering more than 3 components, the geometry found in the two-dimensional representations of the data cloud is not simplicial (to be so a K component system should be projected in a $K-1$ dimensional subspace), which is true, but this is not a condition required to select EI—only convexity is.

We now investigate situations for which no pure spectrum is available. The results obtained for dataset $\mathbf{D}_3$ are shown in the vertical panels of Figure 1C. It is noticeable that the position of the 50 data points in the phasor plots is the same as the one observed for $\mathbf{D}_2$. This allows pointing at another key aspect of EIFD: the coordinate system of a phasor plot does not depend on the collected data, meaning that, in contrast to PCA, each spectrum of a dataset can be processed individually and independently of the others. Another point to highlight is that, even in the absence of pure information, ESs are still to be found at the vertices of the convex hull of the data cloud. Dataset $\mathbf{D}_4$ was built by considering the same mixture coefficients as for $\mathbf{D}_3$, but now using broader spectra for pure signals. It can be observed in the vertical panels of Figure 1D that the geometry of the data cloud is unchanged and that the exact same 15 ESs are identified. The data structure is indeed driven by the mixture coefficients, unchanged, and the simulated spectra are smooth enough to be well represented by low-frequency DFT coefficients.

A significant amount of noise was then added to $\mathbf{D}_4$ to yield the spectral data in $\mathbf{D}_5$ (see Figure 1E). Very similar outcomes are obtained from the application of EI on the plots (G_1, S_1) and (G_2, S_2)—see Figure 1E(2,3), as expected given the denoising property of FT. Noise will affect the coefficients extracted for higher-frequency Fourier coefficients, as can be seen from the (G_3, S_3) plot (see Figure 1E(4)). A slight change is observed in the positions of the points, which translates into the identification of 14 ESs (spectra provided in Figure 1E(5)—note the point circled in green in Figure 1E(5)). Both data compression and denoising are achieved while maintaining the geometry of the data point cloud, thus providing a very robust approach for analytical studies. For the sake of comparison, we discuss in the Supporting Information the results obtained by calculating the convex hulls in the PCA normalized scores subspaces[22] of datasets $\mathbf{D}_{1-5}$ (see Figure S1). In ideal situations, even if the two methods are underlain by different linear transformations of the original data, both are capable of returning EI. Although PCA yields the best subspace representation of the handled signals, as emphasized in this work EIFD exhibits an additional advantage: it enables the on-the-fly evaluation of EI. We also compare the outcomes resulting from the analysis of a challenging Raman hyperspectral dataset conducted by means of both approaches (see Figures S2 and S3).[23] So far, only a few phasor plots were required for the selection of EI. Nonetheless, more of them might be needed given the specific features of the analyzed signals as illustrated with further simulations (see Figure S4). This does not hamper the applicability and the usefulness of our approach as we will outline below.

In the second step, we applied EIFD to the Raman microimaging data of a three-component synthetic mixture

https://doi.org/10.1021/acs.analchem.3c01383
Anal. Chem. 2023, 95, 15497−15504

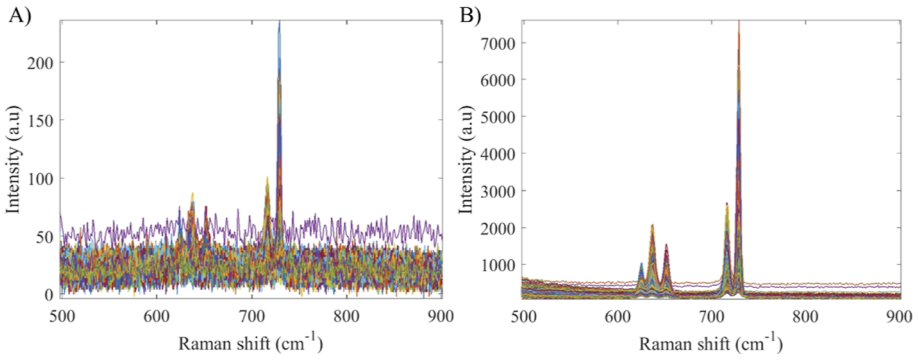


Figure 2. Raman spectra in (A) $\mathbf{D}_{low}$ and (B) $\mathbf{D}_{high}$ (only 1% of the data is shown—the same pixels were selected in both cases).

($CaCO_3$, $NaNO_3$, and Na_2SO_4), considering two different acquisition times, 100 ms (dataset $\mathbf{D}_{low}$) and 5 s (dataset $\mathbf{D}_{high}$) per pixel. To illustrate how a 50-time faster/slower spectral acquisition time translates into a huge difference in terms of SNR and spectral quality, we provide in Figure 2 a representation of the raw spectra for the two datasets. As the exact same field-of-view was sampled for the collection of both of them, (i) a direct comparison can be carried out between the quality of the spectra in Figure 2A,B and (ii) the same information in terms of mixture coefficients is expected, so as to echo the discussion made for the simulations' results. We provide in Figure 3 the phasor plots for $m = 1, 2, 3$ as well as the results obtained by applying EIFD to $\mathbf{D}_{low}$ and $\mathbf{D}_{high}$. Overall, the observed geometry does not indicate the existence of pure spectral pixels over the scanned surface. A total number of 40 ESs were identified for $\mathbf{D}_{low}$, with the corresponding spectra shown in Figure 3A(4). Since the mixture composition is known and reference spectra are available for the 3 pure chemical compounds concerned, we could identify the ESs most correlated with $CaCO_3$, $NaNO_3$, and Na_2SO_4 (see Figure 3A(5)), pointing out that despite the low spectral quality resulting from a fast acquisition, the selected ESs still embed the most relevant spectral information. We repeated the same analysis for $\mathbf{D}_{high}$. From the phasor plots in Figure 3B(1−3), it is noteworthy that the geometry of the data point cloud is now more structured, way sharper, and somehow closer to the geometry one would expect from the analysis of a three-component mixture. However, a much longer acquisition time is required in order to obtain such information. In this case, 28 ESs were extracted by applying EIFD. The corresponding spectra and, among these, the ones most correlated with the pure spectra are shown in Figure 3B(4,5), respectively.

To make clear how striking the gain of information obtained with $\mathbf{D}_{high}$ in comparison to $\mathbf{D}_{low}$, the convex hull obtained for $\mathbf{D}_{low}$ was superimposed to the plots in Figure 3B(1−3) (see the yellow dashed lined). As discussed previously, because the axis coordinates of the phasors are intrinsic to the method, their direct comparison is possible. This provides a meaningful (and potentially useful, as will be discussed in the next section) representation of how the data points span a much wider area for $\mathbf{D}_{high}$, meaning that more selective information about the mixture coefficients is available in $\mathbf{D}_{high}$. Also interesting is the comparison of the purest spectra extracted for $\mathbf{D}_{low}$ and $\mathbf{D}_{high}$ (see Figure 3A(5),B(5)). Obviously, the former, extracted from fast acquisition data, are not only way noisier than the latter but also provide a clear anticipation of the spectral information that would be obtained at longer acquisition times.

The results provided in Figure 3 have highlighted the link between the spectral information extracted applying EIFD on low-SNR hyperspectral imaging data and the one extracted at high SNR, which exhibit a much higher quality for further processing and unmixing, but required a 50-time longer acquisition. We have also shown how EIFD provides a way to readily compare these two levels of information by drawing convex hulls in phasor plots. We now introduce the concept of an entirely new data acquisition approach to drastically accelerate Raman microimaging.

The proposed approach encompasses the following two steps. First, acquiring the data in $\mathbf{D}_{low}$, very rapidly (here, in 28 min). As explained above, applying EIFD to those data would return 40 ESs corresponding to 40 pixel positions along x and y. We can represent the corresponding convex hull in the phasor plots in Figure 4A−C (see the yellow lines). Second, without moving the sample from the stage, targeting data acquisition at these 40 positions only, now measuring high-SNR Raman spectra with a longer acquisition time (here, 5 s per spectral pixel, for a total time of 3.3 min). Note that we actually have these 40 high-SNR spectra available in $\mathbf{D}_{high}$ (we call this subset of 40 spectra $\mathbf{D}_{EIFD}$). We now apply EIFD to $\mathbf{D}_{EIFD}$, identifying 16 ESs, and report the results in Figure 4A−C (see the convex hull marked with the green lines). As pointed out before, the surface spanned by the convex hull of $\mathbf{D}_{EIFD}$ is much wider and the structure of the data is sharper than that for $\mathbf{D}_{low}$. However, what is really striking is the comparison that can be made with the results obtained for $\mathbf{D}_{high}$, data acquired in 14 h of measurement (convex hull represented as a blue dotted line—dotted lines emphasize the fact that in the proposed rapid measurement scenario $\mathbf{D}_{high}$ would not have been measured). The convex hull for $\mathbf{D}_{EIFD}$ and $\mathbf{D}_{high}$ is strikingly similar, meaning that both spectral datasets carry almost the same information in terms of linear spectral unmixing, but this information can be acquired 50 times faster using the proposed approach than with the state-of-the-art imaging strategy. Of course, there is a compromise to be found, and the results cannot be exactly as good as for full data acquisition, as illustrated in Figure 4D−F, but the price to pay is very small considering the 50-fold improvement of the measurement speed. As already pointed out in the discussion on the simulation results, the choice of how many and which harmonics to be considered might be user-dependent. However, a cautious approach may always be taken: for example, by considering 10 phasor plots (and, therefore, 113 ESs) in the current example, the same results would have been obtained but in 9.4 min. For the sake of comprehensiveness, we also provide in Figure 4G−I the images (virtually indistinguishable) returned when performing the linear unmixing of $\mathbf{D}_{low}$, $\mathbf{D}_{high}$, and $\mathbf{D}_{EIFD}$.

https://doi.org/10.1021/acs.analchem.3c01383
Anal. Chem. 2023, 95, 15497−15504

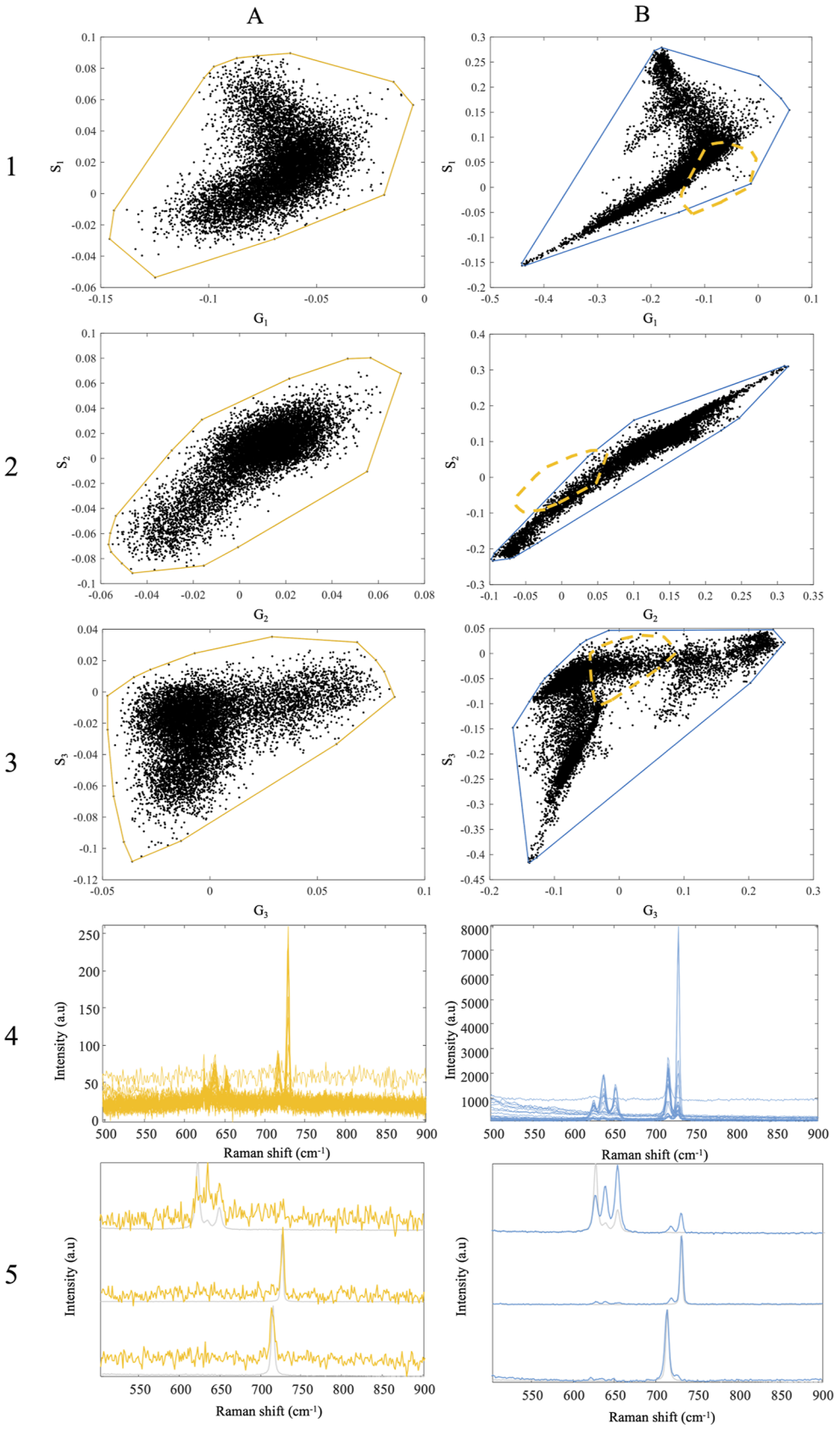


Figure 3. Results obtained applying EIFD to $\mathbf{D}_{low}$ and $\mathbf{D}_{high}$, distributed over columns A and B, respectively. The phasor plots obtained are represented along lines 1−3. The convex hulls estimated by EIFD are displayed in yellow for $\mathbf{D}_{low}$ and blue for $\mathbf{D}_{high}$. Yellow dashed lines (column B) mark the location of the convex hull obtained from $\mathbf{D}_{low}$. ESs are provided along line 4. Along line 5, the spectra exhibiting the highest correlation with the pure spectra of the three compounds under study ($CaCO_3$, $NaNO_3$, and Na_2SO_4) are graphed for $\mathbf{D}_{low}$ and $\mathbf{D}_{high}$, respectively.

https://doi.org/10.1021/acs.analchem.3c01383
Anal. Chem. 2023, 95, 15497−15504

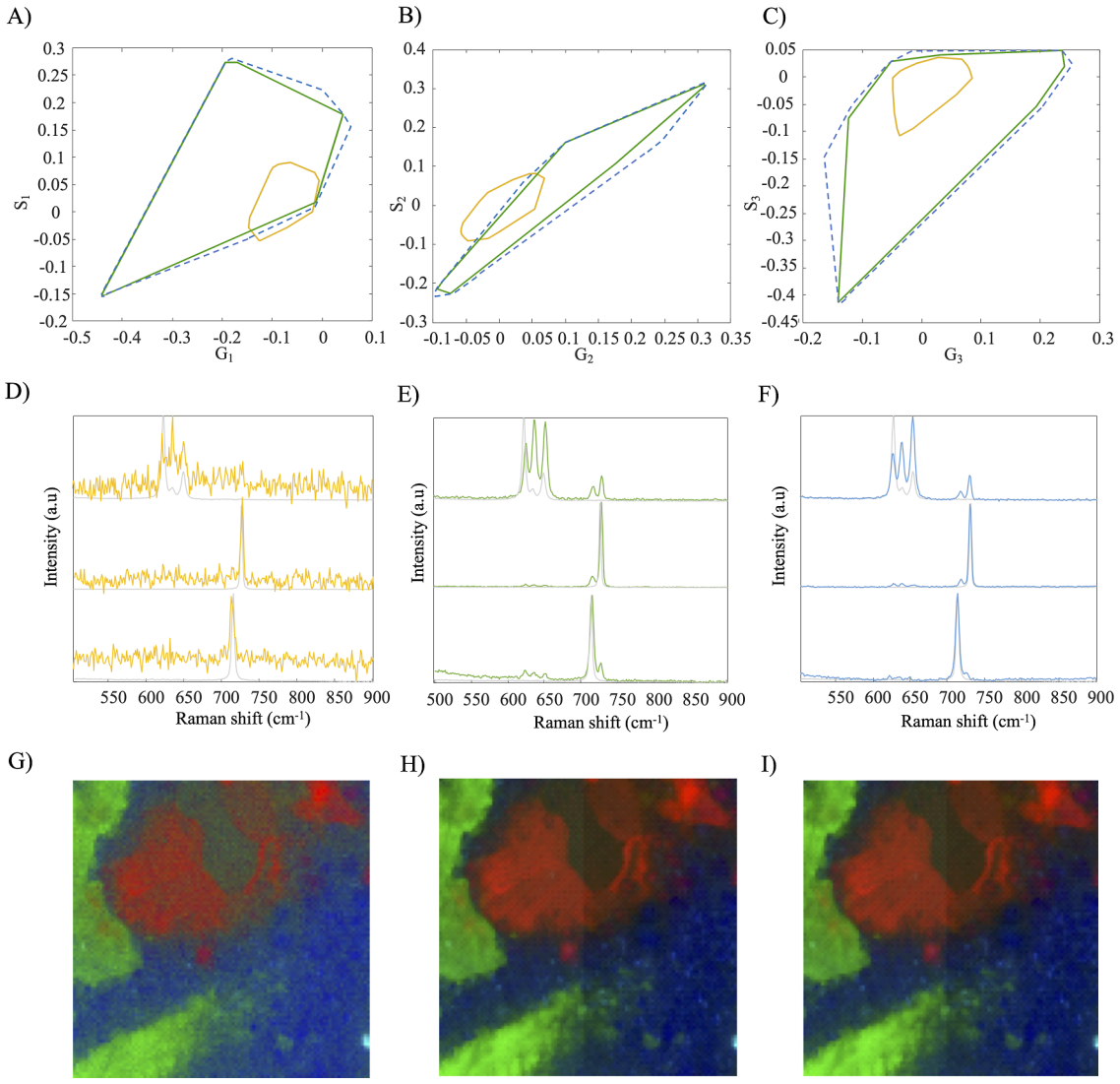


Figure 4. Graphical representation of the geometry of the data in $\mathbf{D}_{low}$ (yellow), $\mathbf{D}_{high}$ (blue), and $\mathbf{D}_{EIFD}$ (green). (A−C) Phasor plots obtained for m = 1, 2, 3, respectively. (D) Spectral profiles in $\mathbf{D}_{low}$ found to exhibit the highest correlation with the reference spectra of the three chemical compounds constituting the analyzed mixture ($CaCO_3$, $NaNO_3$, and Na_2SO_4). (E) Spectral profiles in $\mathbf{D}_{EIFD}$ found to exhibit the highest correlation with the reference spectra of the three chemical compounds constituting the analyzed mixture ($CaCO_3$, $NaNO_3$, and Na_2SO_4). (F) Spectral profiles in $\mathbf{D}_{high}$ found to exhibit the highest correlation with the reference spectra of the three chemical compounds constituting the analyzed mixture ($CaCO_3$, $NaNO_3$, and Na_2SO_4). (G) False RGB image yielded by the superposition of the 3 individual distribution maps resulting from the linear unmixing[26] of $\mathbf{D}_{low}$ (pure spectral estimates in D). (H) False RGB image yielded by the superposition of the 3 individual distribution maps resulting from the linear unmixing of $\mathbf{D}_{high}$ (pure spectral estimates in E). (I) False RGB image yielded by the superposition of the 3 individual distribution maps resulting from the linear unmixing of $\mathbf{D}_{high}$ (pure spectral estimates in F).

Finally, the analysis was also repeated for a data acquisition covering the spectral range from 71.6 to 496.5 cm^{-1}, corroborating the conclusions drawn below (see Supporting Information).

CONCLUSIONS

In this paper, we have demonstrated that EIFD allows extracting ESs from the convex hull analysis of the data clouds in the phasor plots obtained by DFT of the original spectral data. The compression and robustness to noise of EIFD enable a fantastic gain in terms of speed with very low loss of information. The proposed EIFD approach is chemometrics-driven (i.e., based on the relevance of the processed spectra for unmixing) and broadly applicable. We analyzed Raman hyperspectral imaging data and could show the efficiency and robustness of the approach in the context of MCR. We also proposed a new approach for data acquisition with targeted sampling based on EIFD and showed that raster scanning in Raman could be sped-up 50 times.

The method is nature generic, could be applied to different types of data, and could employ other linear transforms than DFT. When dealing with relatively smooth signals such as Raman spectra, we have shown that it is sufficient to consider the DFT coefficients of the first few harmonics in order to fully recover the EIFD. However, we have also pointed out that, when more harmonics are considered, the benefits of our approach are preserved.

We provide the code to implement EIFD which offer the possibility to perform active spectral subsampling identification during the hyperspectral imaging analysis. In essence, the measurement procedure does not need to be performed in two steps. The assessment of EI could be carried out on-the-fly, and

https://doi.org/10.1021/acs.analchem.3c01383
Anal. Chem. 2023, 95, 15497−15504

for each spectral pixel, the decision on whether a longer spectral acquisition is required or not can be taken as long as the imaging experiments are ongoing. This way, EIFD would bring a remarkable acceleration as mentioned above, and it would also allow circumventing sample damaging due to long exposure, opening the door to biological or sensitive sample analysis.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c01383.

MATLAB code; data normalization; and additional results (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Cyril Ruckebusch − *Université Lille, CNRS, LASIRe, F-59000 Lille, France;* orcid.org/0000-0001-8120-4133; Phone: +33 3 20 43 66 61; Email: cyril.ruckebusch@univ-lille.fr

Authors

Laureen Coic − *Université Lille, CNRS, LASIRe, F-59000 Lille, France;* orcid.org/0000-0002-1042-7441

Raffaele Vitale − *Université Lille, CNRS, LASIRe, F-59000 Lille, France;* orcid.org/0000-0002-7497-1673

Myriam Moreau − *Université Lille, CNRS, LASIRe, F-59000 Lille, France*

David Rousseau − *Université d'Angers, LARIS, UMR IRHS INRA, 49000 Angers, France*

José Henrique de Morais Goulart − *Université de Toulouse, IRIT/INP-ENSEEIHT, 31071 Toulouse, France*

Nicolas Dobigeon − *Université de Toulouse, IRIT/INP-ENSEEIHT, 31071 Toulouse, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c01383

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

All authors acknowledge the financial support from the "ANR-21-CE29-0007" project (Agence Nationale de la Recherche) and Eric Ziemons (Université de Liège) for fruitful discussion.

■ ADDITIONAL NOTE

[a]Hereafter, $\mathbb{R}_+$ will denote the set of nonnegative real numbers.

■ REFERENCES

(1) Bøtker, J.; Wu, J. X.; Rantanen, J. *Data Handling in Science and Technology*; Elsevier Ltd., 2020; Vol. *32*, pp 567−581.

(2) Amigo, J. M., Ed. *Hyperspectral Imaging*, 1st ed.; Elsevier B.V., 2019.

(3) Musto, P.; Calarco, A.; Pannico, M.; La Manna, P.; Margarucci, S.; Tafuri, A.; Peluso, G. *Spectrochim. Acta, Part A* **2017**, *173*, 476−488.

(4) Yaseen, T.; Sun, D. W.; Cheng, J. H. *Trends Food Sci. Technol.* **2017**, *62*, 177−189.

(5) Coic, L.; Sacré, P.-Y.; Dispas, A.; Sakira, A. K.; Fillet, M.; Marini, R. D.; Hubert, P.; Ziemons, E. *Talanta* **2019**, *198*, 457−463.

(6) Duncan, J. S.; Insana, M. F.; Ayache, N.; Fellow, L.; Insana, M. F. *Proc. IEEE* **2020**, *108*, 3−10.

(7) Samuel, A. Z.; Yabumoto, S.; Kawamura, K.; Iwata, K. *Analyst* **2015**, *140*, 1847−1851.

(8) Kong, L.; Navas-Moreno, M.; Chan, J. W. *Anal. Chem.* **2016**, *88*, 1281−1285.

(9) Kumamoto, Y.; Mochizuki, K.; Hashimoto, K.; Harada, Y.; Tanaka, H.; Fujita, K. *J. Phys. Chem. B* **2019**, *123*, 2654−2661.

(10) Reitzig, S.; Hempel, F.; Ratzenberger, J.; Hegarty, P. A.; Amber, Z. H.; Buschbeck, R.; Rüsing, M.; Eng, L. M. *Appl. Phys. Lett.* **2022**, *120*, 162901.

(11) Ahmad, M.; Vitale, R.; Silva, C. S.; Ruckebusch, C.; Cocchi, M. *J. Chemom.* **2020**, *34*, No. e3295.

(12) Zhang, J.; Perrin, M. L.; Barba, L.; Overbeck, J.; Jung, S.; Grassy, B.; Agal, A.; Muff, R.; Brönnimann, R.; Haluska, M.; Roman, C.; Hierold, C.; Jaggi, M.; Calame, M. *Microsyst. Nanoeng.* **2022**, *8*, 19.

(13) Feng, X.; Fox, M. C.; Reichenberg, J. S.; Lopes, F. C.; Sebastian, K. R.; Dunn, A. K.; Markey, M. K.; Tunnell, J. W. *J. Biophotonics* **2020**, *13*, No. e201960109.

(14) Hu, C.; Wang, X.; Liu, L.; Fu, C.; Chu, K.; Smith, Z. J. *Analyst* **2021**, *146*, 2348−2357.

(15) Šašić, S., Ed. *Pharmaceutical Applications of Raman Spectroscopy*, 1st ed.; John Wiley and Sons, 2007.

(16) Tarcea, N.; Popp, J. *Eur. Mineral. Union Notes Mineral.* **2012**, *12*, 193−226.

(17) Ruckebusch, C.; Vitale, R.; Ghaffari, M.; Hugelier, S.; Omidikia, N. *TrAC, Trends Anal. Chem.* **2020**, *132*, 116044.

(18) de Juan, A.; Tauler, R. *Anal. Chim. Acta* **2021**, *1145*, 59−78.

(19) Cutler, A.; Breiman, L. *Technometrics* **1994**, *36*, 338−347.

(20) Mørup, M.; Hansen, L. K. *Neurocomputing* **2012**, *80*, 54−63.

(21) Vitale, R.; Ruckebusch, C. *J. Chemom.* **2023**, *37*, No. e3442.

(22) Ghaffari, M.; Omidikia, N.; Ruckebusch, C. *Anal. Chem.* **2019**, *91*, 10943−10948.

(23) Coic, L.; Sacré, P. Y.; Dispas, A.; De Bleye, C.; Fillet, M.; Ruckebusch, C.; Hubert, P.; Ziemons, É. *Anal. Chim. Acta* **2021**, *1155*, 338361.

(24) Coic, L.; Sacré, P. Y.; Dispas, A.; De Bleye, C.; Fillet, M.; Ruckebusch, C.; Hubert, P.; Ziemons, É. *Anal. Chim. Acta* **2022**, *1198*, 339532.

(25) Sawall, M.; Ruckebusch, C.; Beese, M.; Francke, R.; Prudlik, A.; Neymeyr, K. *Anal. Chim. Acta* **2022**, *1233*, 340448.

(26) Ghaffari, M.; Omidikia, N.; Ruckebusch, C. *Anal. Chim. Acta* **2021**, *1141*, 36−46.

(27) Wu, Q.; Marina-Montes, C.; Cáceres, J. O.; Anzano, J.; Motto-Ros, V.; Duponchel, L. *Spectrochim. Acta, Part B* **2022**, *195*, 106508.

(28) Ahmad, M.; Vitale, R.; Cocchi, M.; Ruckebusch, C. *J. Chemom.* **2023**, *37*, No. e3478.

(29) Zhang, L.; Marron, J. S.; Shen, H.; Zhu, Z. *J. Comput. Graph. Stat.* **2007**, *16*, 833−854.

(30) Cooley, J.; Tukey, J. *Math. Comput.* **1965**, *19*, 297−301.

(31) Chiang, H. J.; Koo, D. E. S.; Kitano, M.; Burkitt, S.; Unruh, J. R.; Zavaleta, C.; Trinh, L. A.; Fraser, S. E.; Cutrale, F. *Nat. Methods* **2023**, *20*, 248−258.

(32) Scipioni, L.; Rossetta, A.; Tedeschi, G.; Gratton, E. *Nat. Methods* **2021**, *18*, 542−550.

(33) Malacrida, L.; Ranjit, S.; Jameson, D. M.; Gratton, E. *Annu. Rev. Biophys.* **2021**, *50*, 575−593.

(34) Digman, M. A.; Caiolfa, V. R.; Zamai, M.; Gratton, E. *Biophys. J.* **2008**, *94*, L14−L16.

(35) Budak, A. *IEEE Trans. Educ.* **1967**, *10*, 48−49.

(36) Grande, B. V.; Manne, R. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 19−33.

(37) Rajkó, R. *J. Chemom.* **2009**, *23*, 265−274.