A comparison of multivariate curve resolution with endmember extraction methods in hyperspectral Raman imaging

Robert W. Schmidt [a], Freek Ariese [a], Nematollah Omidikia [b,*]

[a] LaserLaB Amsterdam, Department of Physics and Astronomy, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, the Netherlands
[b] NIOZ Royal Netherlands Institute for Sea Research, Department of Marine Microbiology & Biogeochemistry, 't Horntje (Texel), the Netherlands

HIGHLIGHTS

- Endmember extraction and soft modeling in hyperspectral Raman imaging are compared.
- Three simulated cases, a pharmaceutical sample, and plastic foils were studied.
- VCA outputs can be used as initialization for MCR-ALS.
- Pure pixels are required for extracting pure variables by VCA, but not for MCR-ALS.

GRAPHICAL ABSTRACT
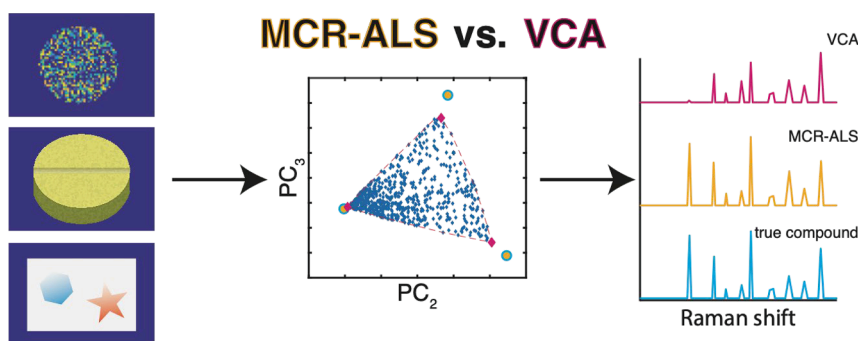
ARTICLE INFO

ABSTRACT

Hyperspectral Raman imaging not only offers spectroscopic fingerprints but also reveals morphological information such as spatial distributions in an analytical sample. However, the spectrum-per-pixel nature of hyperspectral imaging (HSI) results in a vast amount of data. Furthermore, HSI often requires pre- and post-processing steps to extract valuable chemical information. To derive pure spectral signatures and concentration abundance maps of the active spectroscopic compounds, both endmember extraction (**EX**) and Multivariate Curve Resolution (**MCR**) techniques are widely employed. The objective of this study is to carry out a systematic investigation based on Raman mapping datasets to highlight the similarities and differences between these two approaches in retrieving pure variables, and ultimately provide guidelines for pure variable extraction. Numerical simulations and Raman mapping experiments on a mixture of pharmaceutical powders and on a layered plastic foil sample were conducted to underscore the distinctions between MCR and EX algorithms (in particular Vertex Component Analysis, VCA) and their outputs. Both methods were found to perform well if the dataset contains pure pixels for each of the individual components. However, in cases where such pure pixels do not exist, only MCR was found to be capable of extracting the pure component spectra.

* Corresponding author.
E-mail address: nemat.omidikia@nioz.nl (N. Omidikia).

1. Introduction

Raman spectroscopy, as an optical scattering technique, is widely used to investigate the chemical structure of compounds, based on molecular bond vibrations [1]. Each compound exhibits a distinctive combination of bond vibrations, depicted in the measured Raman spectrum, which can be further leveraged for the identification and quantification of chemical compounds. Consequently, Raman spectroscopy has swiftly found applications in various fields: for instance in chemistry for monitoring chemical reactions and studying molecules [2], in biomedical science for disease diagnosis [3,4], and in material sciences for characterizing polymorphism [5] or detecting impurities and defects [6].

In a Raman spectrum, chemical changes are readily distinguishable through the emergence of new peaks, peak shifts, or alterations in relative peak intensities. Analysing a sample through single-point Raman measurements is a subjective and labour-intensive task, suitable for pure samples or for situations where the location of the compound of interest is known. A more comprehensive and objective assessment of a heterogeneous sample can be achieved by integrating micro-Raman spectroscopy with a controllable x-,y-stage, allowing the measurement of the entire area of interest in a mapping (raster scanning) approach [7,8]. Therefore, hyperspectral imaging (HSI) amalgamates the advantages of both spectroscopy and microscopy.

Utilizing the additional spatial information offered by concentration abundance maps, one can derive the location of the target compounds or potential impurities and extract morphological parameters such as shapes, sizes, and patterns. However, Raman mapping experiments yield thousands of spectra per acquisition, making it impractical to manually inspect the spectra for variations within such large datasets.

For exploratory analysis, one simple approach is to generate straightforward distribution maps by selecting the intensity of a peak changing in the spectrum or the ratios between two peaks [9–13]. However, this method is only suitable when the compound of interest is known and exhibits a sufficiently unique Raman peak in the spectrum. It can be a challenging endeavour to create distribution maps for unknown compounds if neither the locations nor the Raman spectra of the compounds are known.

Random [14] or rational selection [15] of a subset of spectra from the pool of the recorded spectra is suggested in the literature against the curse of dimensionality in HSIs. After a representative subset selection, further data processing can be concentrated on the smaller version of the recorded hyperspectral arrays [16]. Unsupervised unmixing algorithms offer a solution for analysing large HSI data by exploring the entire dataset for unique spectra without user subjective constraints. The only requirement is to define the expected number of unique compounds.

Endmember extraction algorithms can be divided into convex geometry and statistical methods [17]. The convex geometry-based methods are favoured in the Raman community, and in this contribution, we will therefore primarily focus on these. The methods can be further divided into simplex volume and orthogonal projection methods, where *N*-FINDR is the most used method in the first case, and vertex component analysis (VCA) in the second case.

Bergner et al. [18] used *N*-FINDR on hyperspectral Raman images to visualize nuclei in brain tissue, whereas Chernenko and colleagues [19] used VCA to map the intracellular distribution of liposomes. Furthermore, Lochocki et al. used VCA to image amyloid deposits in Alzheimer's Disease human brain tissue [3]. Both methods have in common that they assume the presence of pure compound spectra in the dataset and they use principal component analysis (PCA) to reduce the variable dimensionality (wavenumbers) and reveal the variance within the dataset. Depending on the number of components (*N*) to be extracted, each spectrum is projected on an *N*-1 dimensional normalized orthogonalized space surrounded by loadings [17]. Schmidt et al. provided a GUI widget for analyzing hyperspectral Raman images that combines multiple spectral pre-processing methods with various endmember

unmixing algorithms [20]. Endmember extraction operates under the assumption that the dataset primarily comprises mixed spectra of compounds and that the pixels corresponding with the pure compounds exhibit the highest variance within the dataset [17].

On the other hand, Multivariate Curve Resolution [21] is a soft mathematical modeling technique that aims to extract the compound spectra and relative concentrations [22] under chemico-physical constraints [23]. Although in the literature [20] MCR is sometimes also referred to as an endmember technique, it is based on different mathematical principles and in this paper we will regard it as a separate type of approach. MCR is a mixture analysis technique with no assumption about the data except bilinearity and it is widely employed to analyze data from different fields [23].

Whereas MCR and endmember approaches are both extensively utilized in the literature, as far as we know there has been no comprehensive report exploring their similarities and differences based on bilinear data structure. This contribution aims to expand our understanding of when to employ geometrical endmember extraction or soft mathematical modeling techniques for unsupervised unmixing of Raman hyperspectral maps. The study utilizes Raman maps of simulated samples, mixed pharmaceutical powders, and overlapping polymer foils as test cases.

2. Theory

2.1. Bilinear data sets

Typically, the data responses gathered from various spectroscopic systems represent a linear combination of pure variables [23]. Then, the collected responses can be arranged in a data table or data matrix denoted as $D_{I,J}$. Therefore, the problem of mixture analysis can be mathematically modelled in an element-wise manner, expressing it as a linear combination of N sources of data variation. Each source is defined by the sum products of two factors:

$$d_{i,j} = \sum_{n=1}^{N} c_{i,n} s_{n,j} + e_{i,j} \qquad (1)$$

Here, $d_{i,j}$ represents the data entries measured at matrix row i (sample) and matrix column j (variable), while $e_{i,j}$ accounts for the residual data variance not explained by the model, representing the noise contribution. N denotes the number of components contributing to the bilinear data decomposition. The identical model can be expressed in linear algebra matrix form as follows:

$$\mathbf{D}_{I,J} = \mathbf{C}_{I,n} \mathbf{S}_{n,J}^T + \mathbf{E}_{I,J} \qquad (2)$$

The data matrix $\mathbf{D}_{I,J}$ (I is the number of samples and J is the number of columns) can be decomposed into the factor matrices $C_{I,n}$ and $\mathbf{S}_{n,J}^T$. Finally, $\mathbf{E}_{I,J}$ is the residuals matrix that contains the unexplained part of the data by the bilinear model.

In the case of HSIs, the three-dimensional hyperspectral imaging data cube $\underline{\mathbf{D}}(X, Y, J)$, should be unfolded column-wise onto a two-dimensional data matrix of dimensions $\mathbf{D}(X.Y, J)$, where X and Y are the numbers of pixels in the x-and-y directions indicating spatial resolution and J shows the number of measured wavelengths. In the HSI domain, $C_{I,n}$ is the matrix of unfolded abundance maps and $\mathbf{S}_{n,J}^T$ contains pure spectral variables.

2.2. Principal component analysis

Principal components analysis (PCA) is a multivariate statistical technique [24] used for bilinear decomposition, where the factors are constrained to be orthogonal, normalized, and subject to maximum variance constraints. This process yields orthogonal scores ($\mathbf{X}_{I,n}$) and loadings ($\mathbf{Y}_{n,J}^T$)

$$\mathbf{D}_{I,J} = \mathbf{X}_{I,n}\mathbf{Y}_{n,J}^T + \mathbf{E}_{I,J} \tag{3}$$

The truncated PCA decomposition provides a condensed representation of the data variance in a lower-dimensional space. Analyzing the score profiles and their bivariate plots unveils similarities and clusters among the samples (rows of the data matrix) within the new vector space. Simultaneously, the loadings and their plots elucidate the nature of the principal components and their associations with the variables measured in the system. A widely used approach for conducting PCA involves employing fast Singular Value Decomposition (SVD), as described by the equation:

$$\mathbf{D}_{I,J} = \mathbf{U}_{I,N}\Lambda_{N,N}\mathbf{V}_{N,J}^T + \widetilde{\mathbf{E}}_{I,J} \tag{4}$$

Orthonormal $\mathbf{U}_{I,N}$ and $\mathbf{V}_{N,J}^T$ are eigen-vectors and $\Lambda_{N,N}$ is the diagonal matrix of the singular values giving the relative importance of each of the components, while weighted scores $\mathbf{X}_{I,N} = \mathbf{U}_{I,N}\Lambda_{N,N}$ and orthonormal loadings are $\mathbf{Y}_{N,J}^T = \mathbf{V}_{N,J}^T$.

2.3. Endmember extraction

Taking a geometrical approach to the bilinear data set (also bilinear decomposition), measured spectral pixels can be represented as points in the row space of the data set. These points lie in a simplex embedding all the measured data points, with the simplex vertices corresponding to pure/purest spectral pixels [25]. However, all the interior data points that are mixtures are linear combinations of the simplex vertices, they do not carry additional information for linear spectral unmixing [15]. Scheme 1 illustrates a hypothetical example of the data structure for a three-component dataset, where the data cloud (cloud with blue diamonds) is embedded within the pure profiles represented by red dots.

The spectral "point cloud" (e.g. the blue cloud in Scheme 1) is assumed to have a shape similar to a simplex, with the spectra of the purest compounds situated at the vertices of the simplex [26].

In mathematical terminology, a k-simplex is a k-dimensional convex hull with $k + 1$ vertices. For three-component chemical data sets, three pure spectra together with the origin $(0,0,0)$ form the simplex
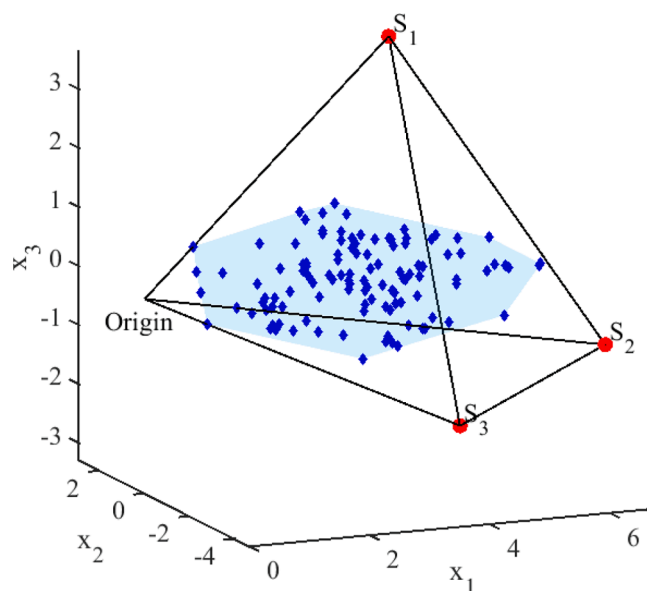
Scheme 1. A graphical illustration of a bilinear data set in the principal component (PC) space. Blue diamonds denote data rows within the data cloud, surrounded by pure spectra (S_1, S_2, and S_3). The linear combination of these pure spectral profiles generates mixtures in a hypothetical dataset, where the coefficients are concentrations. The pure spectra correspond to real-world Raman spectra. x_1, x_2, and x_3 are the first, second, and third scores calculated from Eq. (4).

(tetrahedron) in three-dimensional space (see Scheme 1). To remove intensity ambiguity and fulfil the convex linear combination requirement, a normalization step is conducted [27]. After proper normalization, the data space can be illustrated in a lower-dimensional subspace e. g., three-component cases can be represented in two-dimensional space, and with two-dimensional representation, three pure spectra are enough to create a simplex (triangle).

Both *N*-FINDR and Vertex component analysis (VCA) search for the vertices of the simplex. In the following examples, we unmixed the dataset (mostly) for three spectral profiles, this is done by fitting the PC data cloud in a 2-simplex within 2-dimensional subspace to find the three vertices; to simplify $k + 1$ vertices are plotted in a k-simplex in a k-dimensional space. *N*-FINDR initiates its algorithm with a random starting point. It then iteratively expands a simplex within the Principal Component (PC) plot until the simplex achieves its maximum volume [28]. However, with the iteratively growing simplex in multiple dimensions, the computational complexity of N-FINDR is significantly higher, by up to two orders of magnitude, compared to VCA [29]. Therefore, VCA was used exclusively in this study. VCA determines the vertices of the simplex by iteratively projecting the data points of the simplex onto a direction [29]. The extremes of the projection correspond to the pure/purest spectra, and the next vertices are determined by projecting data onto new directions orthogonal to the subspace of the already determined endmembers until all endmembers are found.

2.4. Multivariate curve resolution

Mathematically, a bilinear map can be decomposed using multivariate curve resolution into two smaller matrices titled component matrices [30]. The objective function of the bilinear model factor decomposition to be minimized is:

$$\min_{\substack{const \\ C,S^T}} \left\| \mathbf{D}_{I,J} - \mathbf{C}_{I,n}\mathbf{S}_{n,J}^T \right\| \tag{5}$$

The objective function is the Frobenius norm of the difference between raw and reconstructed data sets. However, for bilinear model factor decomposition, various optimization schemes under predefined constraints can be implemented [23]. Alternating least squares is one of the flexible and frequently used algorithms. Besides, several constraints are imposed and non-negativity is the minimal constraint in bilinear factor decomposition [31].

MCR-ALS can start a decomposition with the rational or random initialization of one of the factors (either $\mathbf{C}_{I,n}$ or $\mathbf{S}_{J,n}$). Then the unconstrained iterating least squares solutions for $\mathbf{C}_{I,n}$ and $\mathbf{S}_{J,n}$ are:

$$\mathbf{C}_{I,n} = \mathbf{D}_{I,J} \times pinv\left(\mathbf{S}_{n,J}^T\right) \text{ \& } \mathbf{S}_{n,J}^T = pinv(\mathbf{C})^\times \mathbf{D}_{I,J} \tag{6}$$

where *pinv* is the pseudoinverses operation. The ALS algorithm, however, readily allows for incorporating constraints to achieve optimal recovery of physically meaningful profiles in both $\mathbf{C}_{I,n}$ and $\mathbf{S}_{J,n}$ factor matrices. Commonly applied constraints include non-negativity for both concentration and spectral profiles, sparsity for abundance maps.

It should be noted that selectivity plays a dominant role in the endmember extraction. In the following, the concept of selective pixels is emphasized. Each row of the HSI data set can be expressed based on Eq. (1) as:

$$\overrightarrow{r_{i,:}} = c_{i,1} * \overrightarrow{S_1} + c_{i,2} * \overrightarrow{S_2} + c_{i,3} * \overrightarrow{S_3} \tag{7}$$

where $\overrightarrow{r_{i,:}}$ is one of the data rows (e.g., a spectrum recorded at a specific x-y-pixel), and $c_{i,1}$ and $c_{i,2}$ are the concentration of the first and second chemical component for the ith pixel. Besides, $\overrightarrow{S_n}$ ($n = 1, 2, 3$) is the spectral profile of the components. In the event that $c_{i,2}$ and $c_{i,3}$ are zero for a specific pixel, then the data rows contain the pure spectral profile of

$\vec{S_1}$. The same reasoning applies to the pure spectral profiles of the second and third components.

2.5. Rotational ambiguity

Bilinear factor decomposition under the minimal assumption of non-negativity has non-unique solutions in many cases [32]. In the presence of rotational ambiguity, there are several possible solutions that fulfill the imposed constraints during factor decomposition (i.e., non-negativity) and equally fit the data set under study [23]. The extent of the rotational ambiguity can be visualized through a transformation matrix, $T_{N,N}$ which transforms U and V from the PCA loadings to physically interpretable concentrations and spectral profiles based on Eq. (8),

$$S_{J,n} = T_{n,n}V_{n,J}^T$$

$$C_{I,n} = U_{I,n}\Lambda_{n,n}T_{n,n}^{-1} \tag{8}$$

The extent of the rotational ambiguity implies that any invertible matrix $T_{n,n}$ such as $T_{n,n}T_{n,n}^{-1} = I_{n,n}$ transforms an orthonormal loading ($V_{J,n}$) and score ($U_{I,n}$), respectively, to a solution $S_{J,n}$ and $C_{I,n}$.

For several reasons, rotational ambiguity is likely to occur in the field of hyperspectral imaging. A single data matrix is studied, whereas in multi-set analysis there are several data matrices appended into an augmented matrix, and constraints can be applied that connect the profiles in both data dimensions from matrix to matrix. This is not possible in single-matrix hyperspectral imaging. Spectra are overlapped, as may occur with Raman spectra of structurally related chemical components. Also, concentration profiles may be overlapped, as occurring in some of the simulations or experimental data. Finally, MCR-BANDs [33] is used in this contribution to estimate the extent of feasible solutions associated with the components in both simulated and experimental data sets.

3. Materials and methods

3.1. Data sets for numerical simulation

To emphasize the key findings of this research, numerical simulations were conducted on three distinct cases. These simulations mirror real-world hyperspectral Raman imaging scenarios, incorporating abundance maps (depicted in Fig. 1) and spectral profiles (refer to Fig. S1 in the Supporting Information) for mixtures of three components.

In Simulated Case (I), the abundance maps show selective pixels for each species. Each panel in Fig. 1 represents a abundance map for the components, with the last panel in the rows illustrating the mean image of that simulated dataset. Cases (II) and (III) have partially selective and severely overlapped concentration abundance maps, respectively. In case (II), there are selective pixels for the first abundance map (see Fig. 1 second row-C_1) however, C_2 and C_3 show severely overlapped abundance maps with no pure pixels. In case (III), illustrated in the third row of Fig. 1, there is a total overlap between all three abundance maps, and they only differ in relative intensity.

These HSI data sets were simulated to illustrate the difference between endmember extraction and curve resolution for the recovery of pure information. Besides, homoscedastic noise with zero mean and standard deviation of 0.05 maximum value of the data set was added to the simulated cases. The simulated data sets correspond to a three-component multivariate Raman mapping. For each component, the distribution maps correspond to a spot, as shown in Fig. 1. The Raman spectra of components are shown in Fig. S1. The mean image of the data is shown in the last column of Fig. 1. The simulated hyperspectral data set is of dimensions 50×50 pixels by 1101 variables and the unfolded two-way data matrix of 2500 pixels and 1101 pseudo spectral channels.
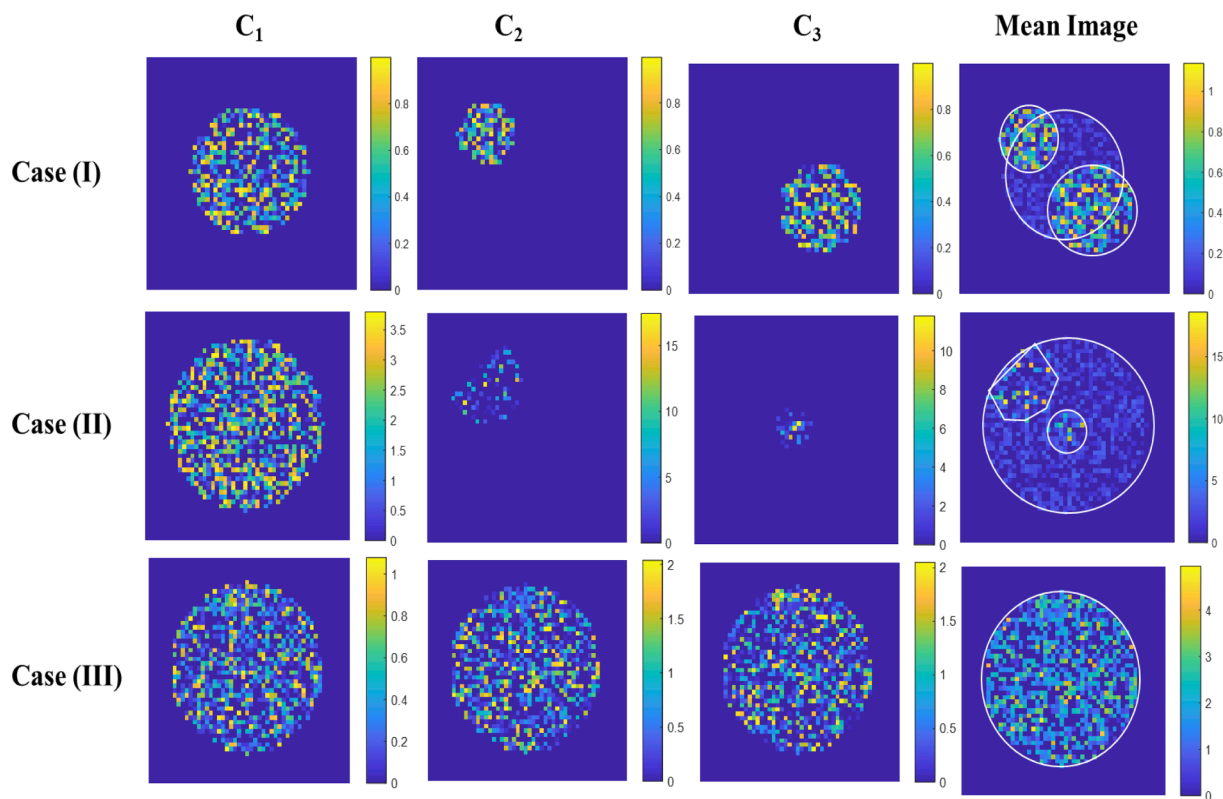

Fig. 1. Concentration abundance maps for the simulated numerical experiments, showing increasing levels of overlap. The first row shows the abundance map of the first, second, and third components in simulated case (I). The last abundance map illustrates the mean image of the simulated data set. The second and third rows highlight the abundance maps of the same components in cases (II) and (III), respectively. For clarity, the boundary of each component is highlighted by a white outline.

Note that deliberate consideration was given to the inclusion or exclusion of pure pixels in these simulated cases.

3.2. Real experiments

3.2.1. APC data set

A pharmaceutical test sample containing three components; aspirin (Merck), paracetamol (Thermo Fischer), and caffeine powder (Sigma-Aldrich) was prepared. Equal amounts (1 g) of each component were weighed, crushed, and mixed in a mortar. The resulting powder was then transferred to a metal microscopy slide and leveled using a spatula to create an even surface. The acquired Raman map contained 17,956 spectra by 1014 variables with an image size of 134 × 134 pixels. Spectra were acquired every 3 μm with an exposure time of 1 s. The pre-processed APC data set (including background removal and outlier detection for removing cosmic rays) is depicted in Fig. 2 (A and B), while the raw unprocessed APC data can be seen in Fig. S2 in the Supplementary Information.

3.2.2. Layered plastic data set

A two-layered plastic foil sample was prepared from polypropylene (PP, transparent, 30 μm thick) and polyethylene (PE, transparent, 200 μm thick) from common packaging material and polyethylene terephthalate (PET, transparent, 13 μm thick) from Goodfellow. The foils were cut into quadrilateral shapes, and pieces of PET and PP were placed on top of a larger piece of PE on a metal microscopy slide (see Fig. 3). For mapping, the Raman microscope was focused at the top of the 2nd (PE) layer; the step size was 30 μm and the exposure time 1 s. Since the foils were not perfectly flat, the relative Raman intensities of the 1st and 2nd layer showed some variation. The acquired Raman map of the full sample (blue rectangle in Fig. 3) contained 35,496 spectra by 1014
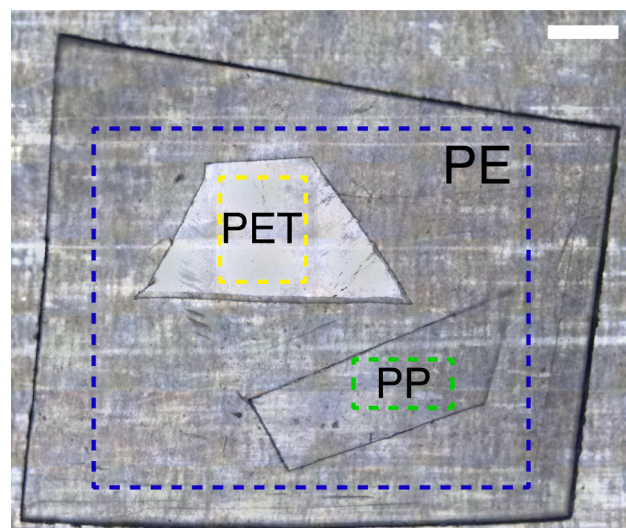

Fig. 3. White light microscopy image of the multilayered plastic foil sample, with pieces of PET and PP on a larger piece of PE, placed on a metal slide. The dashed rectangles show the area of the full Raman map (blue) and the subsets PET@PE (yellow) and PP@PE (green). The scale bar shows 1 mm.

variables with an image size as 204 × 174 pixels. This data set contains spectra of pure PE, but no pixels of pure PP or PET are expected. The raw (unprocessed) layered plastic data set is presented in Fig. S2. On the other hand, the pre-processed data set (background correction with asymmetric least squares [34], and outliers removed) is illustrated in Fig. 2. Panel C shows the recorded spectra per pixel and panel D
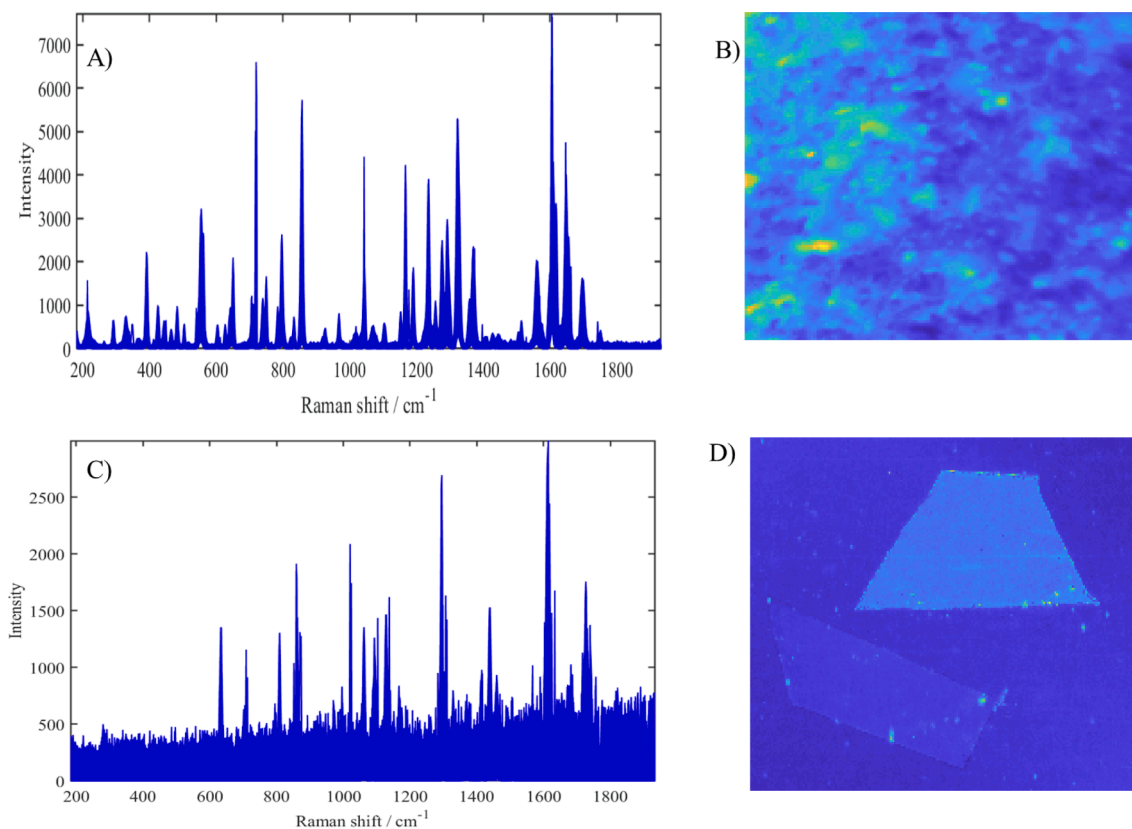

Fig. 2. Summary of the pre-processed real data sets. The asymmetric least squares method [34] was applied for background removal, and outlier detection was conducted to eliminate cosmic rays. Panels (A) and (C) display the pre-processed Raman spectra of the APC and layered plastic data sets, respectively. Panels (B) and (D) present the corresponding mean images of the pre-processed APC and layered plastic samples, respectively.

illustrates the mean image of the data set. The algorithms were also tested on two cropped sub-areas of the Raman map, where no pure pixels are expected (see Fig. 3). The map of the yellow rectangle will show only mixed pixels of PET and PE, whereas the subset of the green rectangle will contain only mixed pixels of PP and PE.

3.3. Raman mapping

Hyperspectral Raman maps of the APC powder mix and the plastic foil sample were acquired with a Renishaw inVia Raman microscope, equipped with a 532 nm excitation laser, and an 1800 L/mm grating. The attached Leica microscope was equipped with a 20x objective (HC PL Fluotar L 20x/0.40 Corr PH1). Numerical simulations, spectra pre-processing and data analysis were carried out in MATLAB. The baseline was removed with asymmetric least squares [34]. For outlier detection, Hotelling's T^2 and Q residuals were calculated for pixels in the PCA score space, and the outlier pixels were replaced with the average of the data set [35]. The data sets were analyzed with homemade PCA, MCR-ALS routines, and a VCA routine from the openly accessible Hyperspectral Toolbox [36]. For reference, Raman spectra were also recorded for the separate pharmaceutical compounds and for the three plastic foils.

4. Results and discussion

To illustrate the differences between endmember extraction methods (e.g., VCA) and a resolution scheme (e.g., MCR-ALS), the simulated cases first underwent PCA decomposition. Subsequently, outputs from VCA and MCR-ALS were calculated and projected into the principal component (PC) space for comparative investigations.

Numerical simulations, spectra pre-processing and data analysis were carried out in MATLAB. It should be highlighted that MCR-ALS was started with random initialization of the concentration abundance maps. Convergence criteria for MCR-ALS decomposition were: maximum iterations of 2000 or when the difference between consecutive convergence steps is below 10^{-6}. To check the fit of MCR-ALS models, the explained variance was calculated as:

$$\boldsymbol{R}^2 = \left(1 - \frac{\sum_{ij}^{IJ}\left(\boldsymbol{d}_{ij} - \widehat{\boldsymbol{d}}_{ij}\right)}{\sum_{ij}^{IJ}\left(\boldsymbol{d}_{ij}\right)^2}\right) \times 100 \qquad (9)$$

where d_{ij} and $\widehat{d}_{ij}$ are the entries of the data and reconstructed matrix based on the model, respectively [23].

4.1. Simulated data

PCA decomposition provides a low-dimensional representation of the original data sets within truncated PC spaces. For the simulated examples with three components, it is possible to illustrate rows of the data sets in a three-dimensional space. Within this presentation, every row can be represented as a point (here, each row is a spectrum related to a pixel). Besides, the space delimited by the data points is the data cloud highlighted in blue in Fig. 4. The pure Raman profiles used for the simulation are depicted as red dots. The pure spectral profiles, together with the origin (0,0,0), form a simplex (here tetrahedron) surrounding the data points [37]. One of the important properties of such a presentation is that all the points located on a straight line starting from the origin have the same spectral shape, but only differ in intensity.

Focusing on the case (I), Fig. 4a shows that the pure simulated spectral profiles (S_1, S_2, and S_3, red dots), VCA extracted endmembers (V_1, V_2, and V_3, magenta diamonds), and MCR-ALS decomposed spectral profiles (M_1, M_2, and M_3, black dots) are all located on the tetrahedron sides starting from the origin. This highlights that the profiles recovered from VCA and MCR-ALS for case (I), are correct, matching the true profiles used for the simulation.

Fig. S3 shows the normalized data space to better visualize the data coordinates and resolved profiles from VCA and MCR-ALS. Finally, the extracted profiles from VCA, MCR-ALS, and true profiles are summarized in Fig. S4 in the Supporting Information. The correlation between the true profiles (S_1, S_2, and S_3), the recovered profiles from VCA (V_1, V_2, and V_3), and the MCR-ALS outputs (M_1, M_2, and M_3) are summarized in Table 1 which shows for case (I) an accurate recovery of the spectral profiles using both VCA and MCR-ALS.

In case (I), there are pure pixels for all components. So, based on equation (7) the spectral profiles of all components can be found among the data rows. In this way, VCA can extract the pure information thoroughly and there are complete matches between the recovered and the true spectral profiles.

For case (II), the data space is illustrated in Fig. 4b. Whereas MCR-ALS successfully retrieves the spectral profiles for case (II) (see black dots in Fig. 4b), VCA fails to extract pure signatures for the second and third components. It is important to highlight that there is total overlap in the abundance maps of the second and third simulated components (see mean image of the case (II) in Fig. 1). Since there are no pure spectral profiles of components 2 and 3 inside the data cloud, VCA will retrieve the purest profiles (but not pure ones), which are different from the true profiles.

In case (II), there are selective pixels for component 1 in the abundance maps (where components 2 and 3 are zero). So, based on Eq. (7), the pure spectrum of component 1 is among the data rows. Hence, VCA
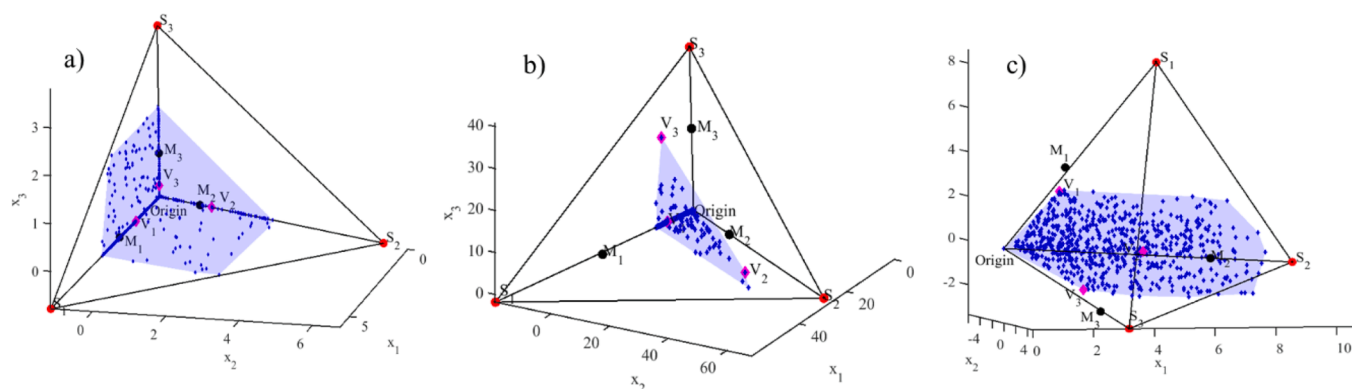

Fig. 4. Data structure of the simulated cases in PC space. Cases (I), (II), and (III) are represented in (a), (b), and (c), respectively. The blue diamonds are data rows (each point is a spectrum), and the red circles marked with S_x are the true profiles used to simulate the cases. The pure spectra together with the origin (0,0,0) create a tetrahedron surrounding the data cloud. VCA and MCR-ALS outputs are depicted as magenta diamonds (V_x) and black dots (M_x), respectively. The pure spectral information is on the vertices/sides, and mixed profiles are inside the tetrahedron.

Table 1

Overview of the correlations obtained between the true profiles and those recovered from MCR-ALS and VCA for the data sets analyzed in this study.

Data	Correlation					
	Corr (VCA outputs, true profiles)*			Corr (MCR-ALS outputs, true)*		
	comp1	comp2	comp3	comp1	comp2	comp3
Simulated Case (I)	0.999	0.999	0.999	0.999	0.999	0.999
Simulated Case (II)	0.999	0.972	0.912	0.999	0.999	0.999
Simulated Case (III)	0.912	0.923	0.901	0.999	0.999	0.999
APC	Aspirin	Paracetamol	Caffeine	Aspirin	Paracetamol	Caffeine
	0.960	0.987	0.977	0.958	0.987	0.979
Layered plastic	PE	PET	PP	PE	PET	PP
	0.991	0.978	0.913	0.998	0.995	0.994
Subset (PET@PE)	PET		PE	PET		PE
	0.928		0.902	0.995		0.991
Subset (PP@PE)	PP		PE	PP		PE
	0.637		0.934	0.999		0.991

* True profiles (S_1, S_2, and S_3), VCA outputs (V_1, V_2, and V_3), MCR-ALS outputs (M_1, M_2, and M_3). The correlation between profiles is calculated using the *corr* function in MATLAB.

can successfully extract S_1. Fig. 4b indicates that V_1, M_1, and S_1 are located on a line starting from the origin which means they are the same. However, V_2 and V_3 are significantly different from S_2 and S_3.

To reiterate, when there is total overlap in the concentration profiles, spectral profiles will be mixed profiles. In this case, VCA and other endmember extraction fail to extract pure profiles. However, the decomposition algorithms such as MCR-ALS can still decompose the data and provide pure information. Fig. S4 (second row) shows the true profiles (red lines), the recovered profiles from VCA (green), and MCR-ALS (blue). The second row of Fig. S4 highlights that the first spectral profile is recovered correctly using both VCA and MCR-ALS. However, VCA fails to extract the pure spectral profiles of the second and third components. The profiles from VCA are different from the true profiles used for simulation, see for instance the large contribution from C_1 in the spectral profile recovered for C_3. In contrast, the spectra recovered using MCR-ALS show an excellent match with the true spectral profiles. The correlation between the true and recovered profiles from VCA and MCR-ALS, summarized in Table 1, quantifies and confirms the previous discussion.

For case (III), concentration abundance maps are even more strongly overlapped and there is no single pure pixel (see Fig. 1, third row). In other words, all rows of the data set are mixtures. Fig. 4c shows the data structure of simulated case (III), where the data cloud is embedded inside the pure spectral profiles. However, rows of the simulated data are inside the tetrahedron and do not touch the sides, indicating that all the pixels are mixed. The recovered profiles from VCA (magenta diamonds) are on the data cloud but they are not pure. On the other hand, the MCR-ALS solutions (depicted as black dots) are situated on the sides of the tetrahedron and align closely with the true solutions (see also correlations in Table 1).

The simulated cases (II) and (III) demonstrate a substantial disparity between endmember extraction and curve resolution techniques. While MCR-ALS aims to decompose the data into meaningful physicochemical solutions, which may not necessarily lie within the data cloud, endmember extraction algorithms search for the purest spectra within the data space. In highly mixed cases, where there is no pure information in the data cloud, VCA (and other endmember extraction methods) will extract the purest information rather than pure spectra.

As mentioned above, MCR-ALS decomposition started with random initialization of the concentration abundance maps for all simulated cases, and the explained variance, calculated based on Eq. (9), was 99.97 %, 99.96 %, and 99.97 % for cases I, II, and III, respectively.

4.2. Aspirin-Paracetamol-Caffeine data set

The APC sample is a well-mixed combination of aspirin, paracetamol, and caffeine with equal contributions (by weight). However, because of inhomogeneity at the micrometer scale, pure pixels are anticipated for all chemicals in the recorded Raman map. The raw APC dataset is illustrated in Fig. S2 and the pre-processed APC spectra are shown in Fig. 2A. Afterward, the pre-processed APC dataset undergoes PCA analysis, with the resulting PC plot presented in Fig. 5. Subsequently, VCA, MCR-ALS, and true profiles (measured from pure materials) are overlaid onto the data cloud. Notably, due to the presence of pure pixels, VCA successfully extracts signatures, matching those obtained by MCR-ALS and the reference spectra from pure aspirin, paracetamol, and caffeine. In essence, the APC data set exhibits similar characteristics as the simulated case (I), where the presence of pure pixels facilitates the extraction of pure variables. Fig. 6 summarizes the
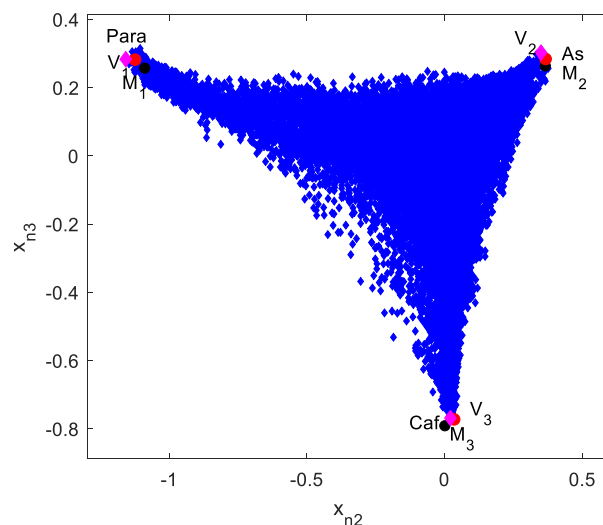

Fig. 5. Normalized PC space of the APC data set showing the third PC (x_{n3}) vs. the second PC (x_{n2}). Pixels from the data are represented as blue diamonds, showing mostly mixed ingredients. Black dots are MCR-ALS (M_x) outputs under non-negativity constraint. In addition, the magenta diamonds represent VCA (V_x) pure profiles. Lastly, the red dots represent the measured spectra from the standard compounds: 1 = paracetamol, 2 = aspirin, 3 = caffeine.
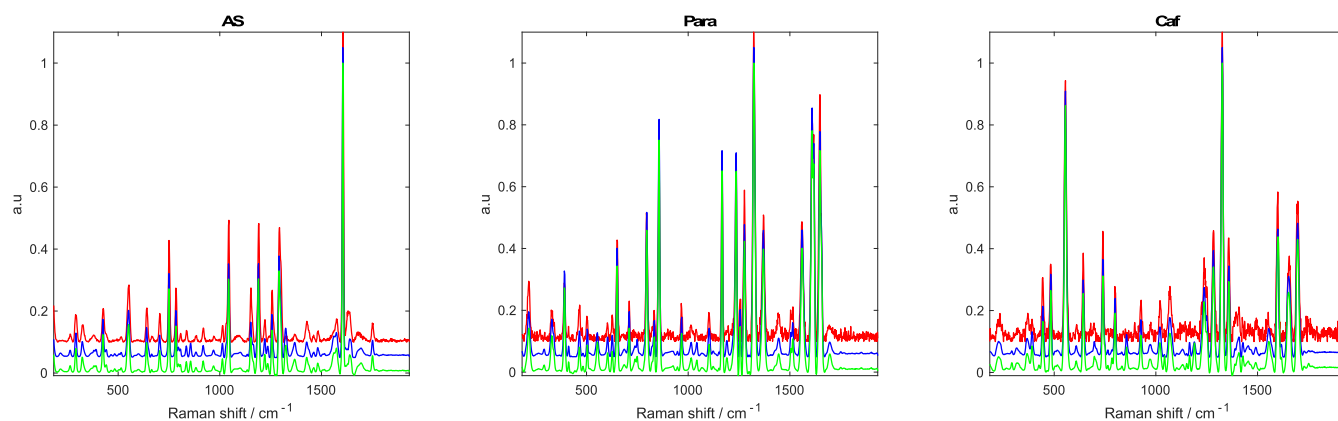
Fig. 6. Recovered spectra using VCA (green lines) and MCR-ALS (blue lines) analysis of the APC data set. True reference spectra are illustrated as red lines. Spectra were normalized, then vertically offset for clarity (0.05 and 0.1 offset for blue and red lines, respectively).

true profiles, VCA outputs, and profiles from MCR-ALS convergence on the APC dataset, highlighting excellent agreement.

The recovered spectral profiles from the APC data set using VCA (green lines), MCR-ALS (blue lines), and true reference profiles (red lines) are summarized in Fig. 6. The correlation between the reference profiles of aspirin, paracetamol, and caffeine and those recovered from VCA and MCR-ALS are {0.960, 0.987, 0.977} and {0.958, 0.987, 0.979}, indicating a high similarity between the reference and recovered profiles with both approaches (see also Table 1). Additionally, corresponding concentration abundance maps of aspirin, paracetamol, and caffeine are provided in Fig. S5, showing the distribution of each chemical within the spatial domain. It should be highlighted that MCR-ALS has an explained variance of 98.9 % on the APC data set, starting from random initialization.

4.3. Plastic foil data set

Finally, the different algorithms were applied to the layered plastic sample shown in Fig. 3. First, the background of the recorded Raman spectral profiles was removed using asymmetric least squares followed by outlier detection. The pre-processed polymer spectral profiles and the mean image of the full dataset are summarized in Fig. 2 C–D. Subsequently, a three-component bilinear decomposition using PCA was performed. Fig. 7 illustrates the PC space of the dataset, with blue diamonds indicating the rows of the data. Red dots represent the co-ordinates of the standard PE (3), PET (2), and PP (1) profiles, forming a triangle that surrounds the data points. Due to the design of the layered plastic sample, there are pure pixels for PE, whereas complete overlap occurs between PET/PE and PP/PE, resulting in no pure pixels for PET and PP in the recorded data set. It's worth noting that the plastic foil sample is similar to simulated case (II), where only one component has pure pixels and the other two are completely overlapped. Since there are pure pixels for PE, the VCA and MCR-ALS extracted profile aligns with the PE standard (see Fig. 7, lower left corner). However, the VCA profiles for PP and PET differ significantly from their true profiles (magenta diamonds are inside the triangle but not on the sides/vertices). On the other hand, MCR-ALS correctly retrieves the pure spectral fingerprints of PET and PP (see black dots in Fig. 7). The three-component MCR-ALS with random initialization showed an explained variance equal to 95.3 % of the total variance of the raw plastic foil data.

Fig. 8 displays the recovered profiles from MCR-ALS as blue lines, VCA as green lines, and the true reference spectral profiles as red lines. Whereas MCR-ALS calculates a meaningful solution for all three plastics with a good match (correlations better than 0.99, see Table 1 for matching indexes), the VCA output for PP and PET, lacking selective pixels, significantly deviates from the true profiles (with a correlation of 0.913 and 0.978 with the true signatures). A closer examination of the
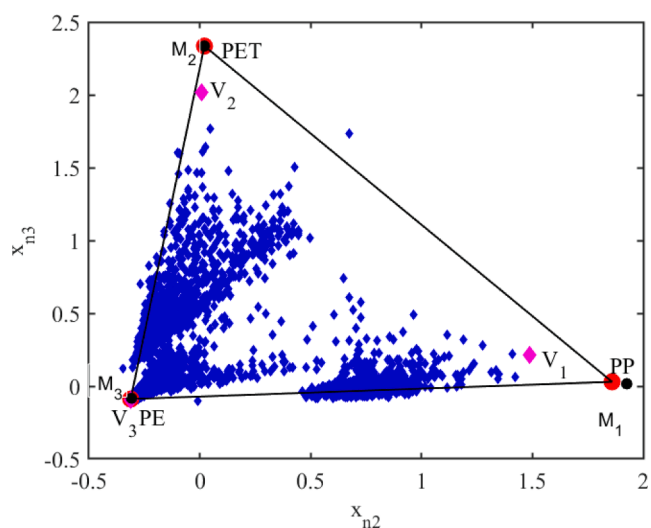

Fig. 7. PC space for the plastic foil data set. The blue diamonds represent rows of the recorded dataset, while the black circles and magenta diamonds denote the profiles from MCR-ALS (M_x) and VCA (V_x), respectively. The red dots indicate the standard plastic spectra measured separately for polyethylene (PE), polyethylene terephthalate (PET), and polypropylene (PP).

VCA-extracted profiles for PET and PP reveals contributions from PE (see Fig. 8 green lines) with some Raman bands related to PE persisting in the VCA-extracted profiles for PET and PP. This is attributed to the fact that all the pixels for PET and PP are mixed with PE, stemming from a lack of selectivity for PET and PP in the designed layered plastic sample. Finally, VCA properly extracts the PE signature from the data because of the selective pixels (with a correlation of 0.991). Additionally, the concentration abundance maps corresponding to the VCA and MCR-ALS solutions for the full data set are presented in Fig. S6. It should be mentioned that here the term "concentration" should not be taken literally, but rather indicates the Raman intensity, which may vary depending on how well the spectrometer's focal point overlaps with the plastic foil pieces (not necessarily perfectly flat).

To further underscore the practical differences between MCR-ALS as a curve resolution algorithm and VCA as an endmember extraction method, we selected two sub-images, cropped from the full Raman map of the plastic foil sample. These chosen sub-images correspond to the layered plastic sections that solely contain PET and PE, or PP and PE, where PE is located underneath the other foils (refer to Fig. 3). Consequently, the sub-images encompass highly mixed spectral profiles without pure pixels for any compound. Selected sub-images have the
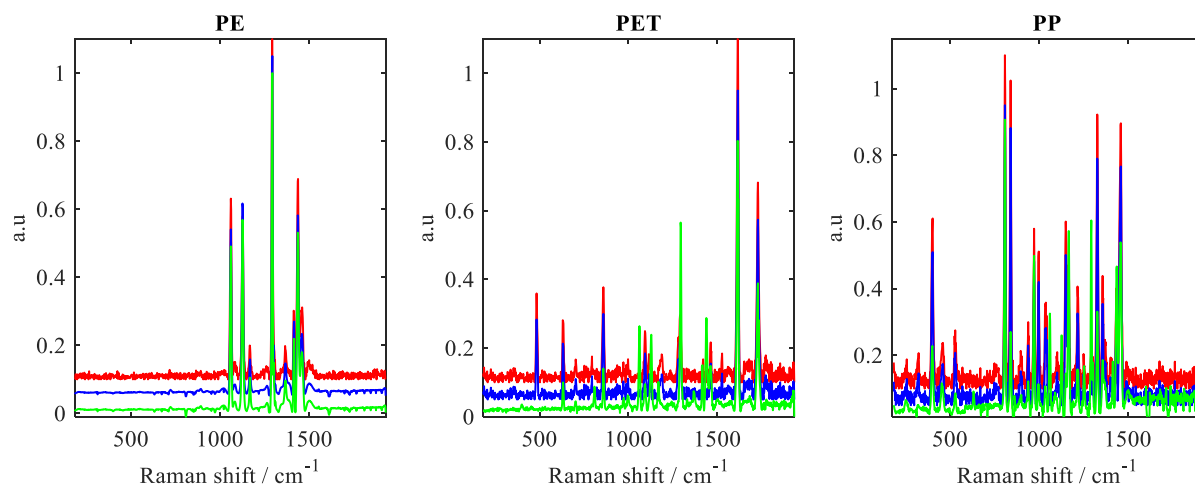
Fig. 8. Spectral signatures of the full plastic foil Raman map, calculated from MCR-ALS analysis (blue lines) and VCA (green lines). True Raman profiles are shown as red lines. From left to right, the panels show the profiles for polyethylene (PE), polyethylene terephthalate (PET), and polypropylene (PP), respectively. Note that the profiles extracted by VCA for PET and PP also contain spectral bands from PE. Spectra were normalized, then vertically offset for clarity (0.05 and 0.1 offset for blue and red lines, respectively).

image size of 50×59 for PET@PE and 47×24 for PP@PE. For these two sub-images, all pixels are mixed; relative intensity variations for the two polymer components in each sub-images are due to differences in focusing for the non-flat foil pieces.

In the initial step, selected sub-images underwent a two-component MCR-ALS decomposition with random initialization of concentration abundance maps. The explained variance was 96.1 % and 97.4 % for the PET@PE and PP@PE data sets, respectively. The recovered profiles from MCR-ALS exhibit an acceptable match with the true signatures from the separate plastic standards (refer to Fig. 9). Additionally, Table 1 summarizes the high degree of similarity between the true and recovered profiles from MCR-ALS. However, owing to a lack of selective pixels, VCA struggles to extract pure profiles (see Table 1). Outputs of VCA extracted from the sub-images (PET@PE and PP@PE) are presented in Fig. 9-last row.

It should be highlighted that VCA outputs are spectra from data. The recorded spectra for the layered plastic sample result from a combination of individual polymer layers, with some variation in relative intensities. As a consequence, the resolved profiles from VCA in Fig. 9-third row are mixed profiles. Applying VCA in such highly overlapped cases is deemed less suitable, necessitating further decomposition. The similarity index between the VCA results on the sub-images and the true profiles is provided in Table 1. In contrast, MCR-ALS or other curve resolution schemes can be employed to retrieve the pure spectral information from such highly mixed cases [38].

5. Conclusion

Hyperspectral Raman imaging offers valuable chemical and morphological details (such as compound distributions) in a sample. However, the analysis of large hyperspectral imaging (HSI) datasets demands innovative, accurate, and fast chemical data analysis approaches. Endmember extraction, such as the Vertex Component Analysis (VCA) solution, involves selecting pure or purest information from the recorded datasets, where the endmember solution lies among data rows or columns. During endmember extraction, the only criterion to be fulfilled is the purity criterion, with no physicochemical constraints imposed on the extraction algorithm. In contrast, resolution techniques like Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS) aim to decompose raw bilinear datasets into meaningful solutions under predefined constraints. Unlike endmember extraction, MCR-ALS solutions may not necessarily reside among data rows and

columns; they can extend beyond the data cloud. It is crucial to emphasize the fundamental difference in perspective between end-member extraction (EX) and curve resolution techniques from a pure information extraction point of view. Whereas EX algorithms search within the data cloud, MCR explores the solution space for bilinear decomposition [39] and is therefore also more computationally intensive. For the data sets analyzed in this study, VCA and MCR-ALS took around 10 and 60 s, respectively. In summary, VCA works very well for exploratory analysis and for datasets that contain pure pixels and it has the advantage of being less computationally intensive.

It is also important to acknowledge that MCR-ALS analysis of highly mixed data sets faces a well-known challenge known as rotational ambiguity [40]. The extent of rotational ambiguity associated with data sets in this work was estimated using the MCR-BANDs algorithm [33] as shown in Table S1 (see Supplementary Information).

Despite the presence of rotational ambiguity in non-selective cases, multivariate curve resolution algorithms, including alternating least squares, are still useful for two reasons. First, MCR-ALS offers flexibility by incorporating constraints to push solutions toward more physico-chemical interpretations [40]. Constraints such as non-negativity on both concentrations and spectra, unimodality on spectra if applicable, and sparsity constraints on concentration abundance maps can be imposed [32]. Second, the MCR-ALS solution can be further enhanced through rational initialization schemes rather than random seeding [41]. Finally, expert judgment and interpretation are essential when working with samples from diverse environments and fields, as the selectivity of pixels is challenging to guarantee.

In conclusion, the recommendation is to utilize curve resolution techniques and reserve endmember extraction methods for exploratory analysis, datasets that contain pure pixels, or as an initialization step for MCR methods, especially in situations where no pure pixels are expected in the data set.

CRediT authorship contribution statement

Robert W. Schmidt: Writing – review & editing, Visualization, Methodology, Investigation. **Freek Ariese:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Nematollah Omidikia:** Writing – original draft, Visualization, Software, Formal analysis, Conceptualization.
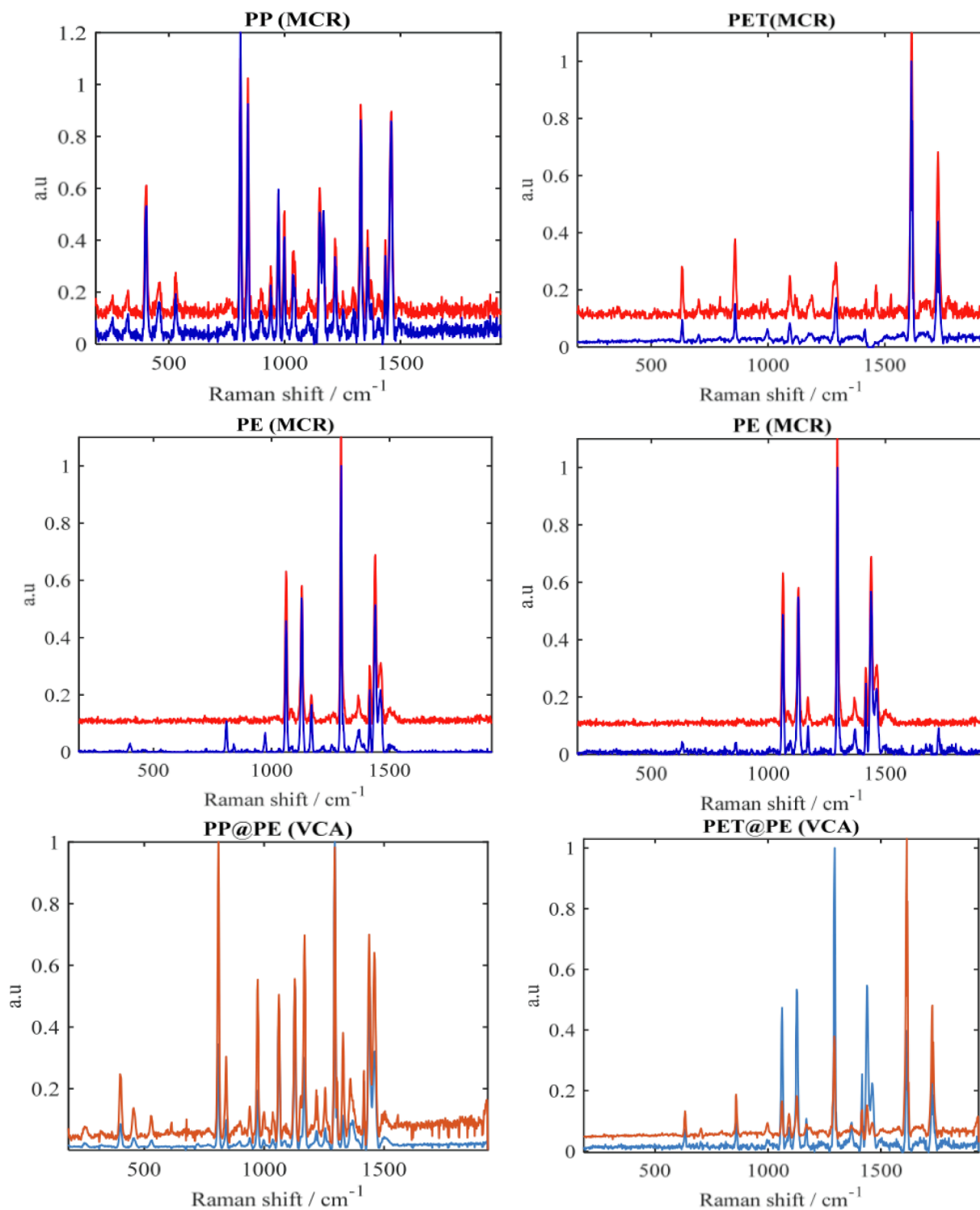
Fig. 9. MCR-ALS and VCA outputs for the cropped data sets PP@PE (left column) and PET@PE (right column). In the first two rows, red lines are reference profiles and blue lines are MCR-ALS solutions. The third row shows the extracted VCA profiles for the cropped images. Blue and red profiles show the first and second endmembers extracted from the PP@PE and PET@PE sub-images (a vertical offset was applied for clarity).

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

the top sectors (grant # 741.018.202 "Soft Advanced Materials") and from the Ministry of Economic Affairs in the framework of the *"PPS-Toeslagregeling"*.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2024.124868.

References

[1] P. Larkin, Infrared and Raman Spectroscopy, vol. 44, Elsevier, 2011.

[2] G. Amberchan, A.L. Allen, J.H. Golden, B. Singaram, J.Z. Zhang, Real-time monitoring of aqueous organic reduction reactions using ex situ fiber optic Raman spectroscopy, ACS Sust. Chem. Eng. 9 (2021) 6068–6078.

[3] B. Lochocki, B.D.C. Boon, S.R. Verheul, L. Zada, J.J.M. Hoozemans, F. Ariese, J. F. de Boer, Multimodal, label-free fluorescence and Raman imaging of amyloid deposits in snap-frozen Alzheimer's disease human brain tissue, Commun. Biol. 4 (2021) 1–13.

[4] C. Krafft, J. Popp, P. Bronsert, A. Miernik, Raman spectroscopic imaging of human bladder resectates towards intraoperative cancer assessment, Cancers (Basel) 15 (2023).

[5] M. Inoue, H. Hisada, T. Koide, T. Fukami, A. Roy, J. Carriere, R. Heyler, Transmission low-frequency Raman spectroscopy for quantification of crystalline polymorphs in pharmaceutical tablets, Anal. Chem. 91 (2019) 1997–2003.

[6] D.R. Neuville, D. de Ligny, G.S. Henderson, Advances in Raman spectroscopy applied to earth and material sciences, Rev. Min. Geochem. 78 (2014) 509–541.

[7] S. Schlücker, M.D. Schaeberle, S.W. Huffman, I.W. Levin, Raman microspectroscopy: a comparison of point, line, and wide-field imaging methodologies, Anal. Chem. 75 (2003) 4312–4318.

[8] A. Adibi, A. Toshimitsu Asakura, S.W. Theodor Hänsch, G. Takeshi Kamiya, T. Ferenc Krausz, J. Bo, G.A. Monemar, L. Herbert Venghaus, B. Horst Weber, B. Harald Weinfurter, Raman Imaging, vol. 168, in: A Zoubir (Ed.), 2012, Springer Berlin Heidelberg, Berlin, Heidelberg.

[9] I.A. Larmour, K. Faulds, D. Graham, Rapid Raman mapping for chocolate analysis, Anal. Methods 2 (2010) 1230.

[10] J. Qin, K. Chao, M.S. Kim, Investigation of Raman chemical imaging for detection of lycopene changes in tomatoes during postharvest ripening, J. Food Eng. 107 (2011) 277–288.

[11] S. Šašić, D.A. Clark, J.C. Mitchell, M.J. Snowden, Raman line mapping as a fast method for analyzing pharmaceutical bead formulations, Analyst 130 (2005) 1530–1536.

[12] H. Lin, P. Braeuninger-Weimer, V.S. Kamboj, D.S. Jessop, R Degl'Innocenti, H. E. Beere, D.A. Ritchie, J.A. Zeitler, S. Hofmann, Contactless graphene conductivity mapping on a wide range of substrates with terahertz time-domain reflection spectroscopy, Sci. Rep. 7 (2017) 1–9.

[13] A.P. de Oliveira, W. Chase, M.P. Confer, S. Walker, D. Baghel, A. Ghosh, Colocalization of β-sheets and carotenoids in Aβ plaques revealed with multimodal spatially resolved vibrational spectroscopy, J. Phys. Chem. B 128 (2024) 33–44.

[14] S. Kucheryavskiy, Blessing of randomness against the curse of dimensionality, J. Chemom. 32 (2018).

[15] M. Ghaffari, N. Omidikia, C. Ruckebusch, Essential spectral pixels for multivariate curve resolution of chemical images, Anal. Chem. 91 (2019) 10943–10948.

[16] C. Ruckebusch, R. Vitale, M. Ghaffari, S. Hugelier, N. Omidikia, Perspective on essential information in multivariate curve resolution, TrAC - Trends Anal. Chem. 132 (2020).

[17] V.K. Kale, M. Solankar, B.D. Nalawade, Hyperspectral endmember extraction techniques, Process. Anal. Hyperspectral Data 11 (IntechOpen) (2020) 13.

[18] N. Bergner, A. Medyukhina, K.D. Geiger, M. Kirsch, G. Schackert, C. Krafft, J. Popp, Hyperspectral unmixing of Raman micro-images for assessment of morphological and chemical parameters in non-dried brain tumor specimens, Anal. Bioanal. Chem. 405 (2013) 8719–8728.

[19] T. Chernenko, R.R. Sawant, M. Miljkovic, L. Quintero, M. Diem, V. Torchilin, Raman microscopy for noninvasive imaging of pharmaceutical nanocarriers: intracellular distribution of cationic liposomes of different composition, Mol. Pharm. 9 (2012) 930–936.

[20] R.W. Schmidt, S. Woutersen, F. Ariese, RamanLIGHT—a graphical user-friendly tool for pre-processing and unmixing hyperspectral Raman spectroscopy images, J. Opt. 24 (2022) 064011.

[21] A. De Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Meth. 6 (2014) 4964–4976.

[22] C.H. Camp, PyMCR: a python library for multivariatecurve resolution analysis with alternating regression (MCR-AR), J. Res. Natl. Inst. Stand Technol. 124 (2019).

[23] N. Omidikia, M. Ghaffari, J. Jansen, L. Buydens, R. Tauler, Bilinear model factor decomposition: a general mixture analysis tool, Chemomet. Intell. Lab. Syst. 240 (2023).

[24] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intel. Lab. Syst. 2 (1987) 37–52.

[25] O.S. Borgen, B.R. Kowalski, An extension of the multivariate component-resolution method to three components, Anal. Chim. Acta 174 (1985) 1–26.

[26] B.-V. Grande, R. Manne, Use of Convexity for Finding Pure Variables in Two-Way Data from Mixtures 50 (2000).

[27] R. Rajkó, K. István, Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR) method based on computational geometry, J. Chemom. 19 (2005) 448–463.

[28] E. Winter M 1999 N-FINDR: an algorithm for fast spectral endmember determination in hyperspectral data International Geoscience and Remote Sensing Symposium (IGARSS) 3753, 266–75.

[29] J.M.P. Nascimento, J.M.B. Dias, Vertex component analysis: a fast algorithm to unmix hyperspectral data, IEEE Trans. Geosci. Remote Sens. 43 (2005) 898–910.

[30] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intel. Lab. Syst. 30 (1995) 133–146.

[31] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, J. Chemom. 9 (1995) 31–58.

[32] N. Omidikia, M. Ghaffari, R. Rajkó, Sparse non-negative multivariate curve resolution: L0, L1, or L2 norms? Chemom. Intel. Lab. Syst. 199 (2020).

[33] R. Tauler, Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution, J. Chemom. 15 (2001) 627–646.

[34] J. Peng, S. Peng, A. Jiang, J. Wei, C. Li, J. Tan, Asymmetric least squares for multiple spectra baseline correction, Anal. Chim. Acta 683 (2010) 63–68.

[35] D. Ballabio, A MATLAB toolbox for principal component analysis and unsupervised exploration of data structure, Chemomet. Intell. Lab. Syst. 149 (2015) 1–9.

[36] D. Kun, S. Bakker, Hyper Spectral Toolbox, GitHub, 2018.

[37] R. Rajkó, N. Omidikia, H. Abdollahi, M. Kompany-Zareh, On uniqueness of the non-negative decomposition of two- and three-component three-way data arrays, Chemom. Intel. Lab. Syst. 160 (2017) 91–98.

[38] M. Ghaffari, G.H. Tinnevelt, M.C.P. Van Eijk, S. Podchezertsev, G.J. Postma, J.J. Jansen, Mono/Multi-material Characterization Using Hyperspectral Images and Multi-Block Non-Negative Matrix Factorization.

[39] H. Abdollahi, M. Maeder, R. Tauler, Calculation and meaning of feasible band boundaries in multivariate curve resolution of a two-component system, Anal. Chem. 81 (2009) 2115–2122.

[40] N. Omidikia, S. Beyramysoltan, J. Mohammad Jafari, E. Tavakkoli, M. Akbari Lakeh, M. Alinaghi, M. Ghaffari, S. Khodadadi Karimvand, R. Rajkó, H. Abdollahi, Closure constraint in multivariate curve resolution, J. Chemom. 32 (2018).

[41] A.C. Olivieri, N. Omidikia, Initialization effects in two-component second-order multivariate calibration with the extended bilinear model, Anal. Chim. Acta 1125 (2020) 169–176.