

Curve resolution for multivariate images with applications to TOF-SIMS and Raman

Neal B. Gallagher^{a,*}, Jeremy M. Shaver^a, Elaine B. Martin^b, Julian Morris^b,
Barry M. Wise^a, Willem Windig^a

^a *Eigenvector Research, Inc., P.O. Box 561, Manson, WA 98831, USA*

^b *Centre for Process Analytics and Control Technology, Merz Court, School of Chemical Engineering and Advanced Materials, University of Newcastle, Newcastle-Upon-Tyne, NE1 7RU, UK*

Available online 28 July 2004

Abstract

Multivariate curve resolution (MCR) is a powerful technique for extracting chemical information from multivariate images (MI). Two problems with MI are (1) initializing the MCR decomposition and (2) lack of selectivity in the image. Methods derived for initializing MCR with evolving data that are naturally ordered in time are not generally applicable for MI. Purity-based methods show promise and a simple, robust purity-based algorithm has been developed to initialize the MCR decomposition. This method used distance measures to find samples (or variables) on the exterior of a data set. Lack of selectivity, common in MI, generally results in a rotational ambiguity in factors extracted with MCR. Functional constraints were tested as a means to reduce this ambiguity, and the method tested showed that functional constraints could be used to account for offsets and backgrounds in Raman images. Robust initialization and introduction of functional constraints tested here are necessary steps towards the final objective of providing a simple methodology for constraining factors in a general fashion so that knowledge of the physics and chemistry can be easily incorporated in to any MCR decomposition. Additionally, the use of a sequential decomposition method (sequential MCR) is employed to help reduce mixing of recovered components in rotationally ambiguous systems. © 2004 Elsevier B.V. All rights reserved.

Keywords: Multivariate curve resolution (MCR); Multivariate image; Hyperspectral image; Purity; End-member extraction

1. Introduction

Multivariate curve resolution (MCR) has been applied to evolving data [1–6] and is gaining more widespread use for multivariate images (MI) [7–10]. The reason for the growing popularity is that MCR extracts chemically meaningful information that is easy to interpret. For evolving data, MCR is used to identify *what* analytes are present and *when* they occurred, e.g., when an analyte eluted from a GC column. Applying MCR to multivariate images (MI) yields information about *what* analytes are present and *where* in the image they are located. The resulting “concentration images” can be used for process monitoring, quality control, change detection, land mapping in remote sensing, and estimating spatial statistics.

The objective of MCR is to decompose a data set into estimates of pure contributions to each pixel and corresponding spectra. Due to rotational and multiplicative ambiguities in MCR (when not using external information such as standards) there is a lack of an absolute concentration scale for the recovered “concentrations”. Therefore, these are often referred to as scores or contributions images. For similar reasons, the corresponding extracted “spectra” are often referred to as factors. These terms are used interchangeably in this paper with the understanding of the potential ambiguities. In addition, depending on how many channels are included in the spectral axis, multivariate images have been referred to as hyperspectral, omnispectral, ultraspectral, and megaspectral. To avoid confusion, all of these will simply be referred to as multivariate or hyperspectral images.

MCR is known under several names (e.g., self-modeling curve resolution, self-modeling mixture analysis, and end-member extraction), and there are several methods that have a similar objective (e.g., independent components analysis [11], purity-based methods [1,4,9,12,13], positive matrix

* Corresponding author. Tel.: +1-509-687-1039; fax: +1-509-687-2033.

E-mail address: nealg@eigenvector.com (N.B. Gallagher).

factorization [14], direct fitting [15]). The primary difference between these methods is how the decomposition is performed and how constraints are introduced to the problem. In the present work, an alternating and constrained least squares (ALS) approach was used [1,3]. Previous work has shown how to combine “hard” (e.g., using kinetic constraints) and “soft” modeling (MCR) to provide a physically meaningful decomposition [16,17]. The present work shows how this can be generalized using “functional” constraints. These can be applied to each mode, factor-by-factor, or element-by-element and the constraints are highly flexible. This allows the user to incorporate a priori chemical or physical constraints into the decomposition.

A typical MCR ALS decomposition starts with an initial guess for one of the modes and a constrained optimization is used to find a final solution. The initial guess can affect the quality of the final solution. For evolving data, methods have been developed that use the time ordered nature of the data to obtain estimates of the range of existence of different analytes [18–20]. Although versions of these methods can be employed for multivariate images, the quality of the initial guess is typically poor due to lack of chemical selectivity (i.e., each pixel often contains multiple analytes). One alternative is to use pure samples (or variables) as an initial guess [21]. The approach uses the fact that the underlying true analyte spectra must lie on the exterior of the data space. This was also pointed out by Henry and Kim [22] and Neal [23] and is the basis of the NFINDR algorithm [10]. However, in contrast to the volume maximization criteria used in NFINDR, the present approach uses distance measures on the mean-centered one-norm data. The result is a faster algorithm that finds the extreme samples (or variables) of the data space and provides a first estimate of spectra (or contributions) used to initialize the ALS optimization procedure. These guesses, combined with nonnegativity and other constraints, were used to decompose multivariate images. This procedure was used previously for a remote sensing application [24] and other multivariate image work [25].

Pure sample (or variable) initial guesses attempt to provide a good first approximation to the MCR decomposition, and functional constraints assist in minimizing rotational ambiguities. However, some systems have the additional complexity that the components exhibit significantly different magnitude of responses (e.g., a mix of poorly and highly absorbing components in absorption spectroscopy). In the presence of rotational ambiguities, this is manifested as the intermixing of the large response components into the recovered spectra of the smaller response components. The current work demonstrates the use of sequential MCR in which the model is progressively expanded to include additional factors [24]. Sequential MCR works as follows. The model is initially fit with fewer factors than are desired. An additional factor is added to this initial solution and the model is again fit. This process is repeated until the number of desired components has been reached.

Using these three approaches—pure sample/variable initial guesses, functional constraints, and sequential MCR—the challenges to analyzing multivariate images can be lessened.

2. Experimental multivariate images

Three example multivariate images were examined; time-of-flight secondary ion mass spectrometry (TOF-SIMS) surface analysis of a drug bead, a Raman image of an aspirin/polyethylene mixture on a glass slide, and a Raman image of polymer fibers in an epoxy. All analysis was performed using MATLAB 6.5 (The MathWorks) and the PLS_Toolbox 3.0 (Eigenvector Research).

2.1. TOF-SIMS image

The first image was obtained using time-of-flight secondary ion mass spectrometry (TOF-SIMS) of a drug bead. The data is described in detail in Belu et al. [26] (Bead System 1) and the system is described only briefly here. The drug bead contains a core of prednisolone sodium metasulfobenzoate (prednisolone) in microcrystalline cellulose and lactose coated with one part amylose and four parts surelease (ethylcellulose, ammonium hydroxide, coconut oil and oleic acid). Measurements were made using a TRIFT II instrument (Physical Electronics, Eden Prairie, MN). The image of interest is image is $250 \times 250 \mu\text{m}^2$ region of a cross-sectioned bead.

Measured spectra of prednisolone, lactose and amylose were used in Ref. [26] to identify signature peaks of these analytes. Univariate images were then used to characterize the spatial statistics of the drug, core materials, and coating. This was successful because the SIMS has a relatively high selectivity. However, sampling matrix effects may result in measured spectra manifesting differently than the pure material, and measured spectra are what would be used for a process monitoring system. Extracted “pure” analyte spectra are expected to result in better contribution images due to more accurate representation of the species in the given matrix and to signal averaging aspects of the multivariate approach. Ninety-three channels of positive ion spectra were used and spectra of the pure materials were used to guide the choice of channels to use.

2.2. Raman aspirin/polyethylene mixture

The second example is a Raman image of an aspirin/polyethylene mixture. The sample consists of scrapings from a tablet of aspirin (Bayer) and granulated high-density polyethylene (Dow Chemical) dispersed on a silica-glass slide. The image was collected on a fiber-coupled Raman microscope (Hololab 5000, Kaiser Optical Systems) over a $105 \times 165 \mu\text{m}^2$ region at $5 \mu\text{m}$ steps using a $10 \times$ Olympus objective. The laser wavelength was 785 nm and the Raman

spectra were measured from 0 to 1926 cm^{-1} at 1 cm^{-1} intervals. The Raman data were truncated to the spectral range of greatest interest (600–1660 cm^{-1}) and resampled to 2 cm^{-1} . In this experiment, the objective and glass slide were not constructed of low-background composition glass. This resulted in significant luminescence at the excitation wavelength used.

2.3. Raman image of synthetic polymer fibers

The third example is a Raman image of a sample, obtained from Robert Jan van Wilk of Akzo Nobel through Kaiser Optical Systems, of five synthetic polymer fibers embedded in a low-background medium. Two of the fibers were polyethylene terephthalate (PET), two were Twaron[®], and a third was an unidentified cellulosic material. A $38 \times 92 \mu\text{m}^2$ region of the sample was measured at 2.3- μm steps to form the 16×40 pixel image. The data were collected on equipment similar to that described for the previous image except that low background objectives were used and the data were truncated to the spectral range of 500–1950 cm^{-1} and left at 1 cm^{-1} resolution.

3. MCR objective

3.1. MCR model

A brief discussion of the objective of MCR is provided below. Further reading and details can be found in Refs. [1–10,15,16,20]. The MCR model is based on classical least squares (CLS) as shown in Eq. (1)

$$\mathbf{X} = \mathbf{CS}^T + \mathbf{E} \quad (1)$$

where \mathbf{X} is a $M \times N$ matrix of measured spectra, \mathbf{C} is a $M \times K$ matrix of contributions to be determined, \mathbf{S} is a $N \times K$ matrix of spectra to be determined, \mathbf{E} is a $M \times N$ matrix of residuals, and K is the number of factors in the model. In general, \mathbf{C} is a matrix of coefficients dictating the relative contribution of each of the column vectors in \mathbf{S} to the response matrix \mathbf{X} , and K is the number of independent basis vectors required to span the noise-free data space. Since Eq. (1) is Beer's Law for multicomponent mixtures, \mathbf{C} and \mathbf{S} are often referred to as concentrations and pure analyte spectra matrices. However, without additional information, there is a scaling ambiguity involved with the two matrices; the recovered values for \mathbf{C} and \mathbf{S} are often only proportional to true concentration and spectra respectively. This is discussed in more detail below.

If the pure analyte responses have some selectivity (e.g., there are spectral or contribution features unique to an analyte) [27], a solution to Eq. (1) can be obtained under the constraint that the length of each spectrum is a scalar constant (alternatively this constraint could be applied to the columns of \mathbf{C}). It is typical to constrain the two-norm of the

columns of \mathbf{S} to unit length i.e. $\mathbf{s}_k^T \mathbf{s}_k = 1$ where $k=1, \dots, K$ (if non-zero equality constraints are included for factor k this constraint must be relaxed). Additionally, if \mathbf{C} and \mathbf{S} are constrained to be nonnegative then the resolved contributions and spectra are a close approximation to the true factors. However, there is typically a range of feasible solutions [21,28–31] and the range gets larger as the selectivity decreases. The range of feasible solutions is due to a rotational ambiguity in the solution to Eq. (1). As shown in Eq. (2), multiplying \mathbf{C} and \mathbf{S} by a $K \times K$ invertible rotation matrix \mathbf{A} results in a decomposition with the same fitness (same \mathbf{E}) as the original solution.

$$\mathbf{X} = (\mathbf{CA})(\mathbf{A}^{-1}\mathbf{S}^T) + \mathbf{E} \quad (2)$$

The range of feasible solutions can be restricted by applying constraints (e.g., equality constraints) [30], or functional constraints [16]. The latter can be used to impose smoothness, chemical information (e.g., by fitting the contributions to a kinetic expression or using closure), or physical information (e.g., unimodality is often imposed on contribution profiles obtained from chromatography data). In Example IIb below, we show that other physical constraints can be used. If \mathbf{A} is diagonal, then instead of a rotation it represents a stretching resulting in a multiplicative ambiguity in the solution. This can be resolved, for example, if the concentration is known at some point or if closure is applicable. Another method augments the spectra measured on standards with known concentrations onto \mathbf{X} [28]. Equality constraints can then be imposed on the concentrations for the measurements on the standards.

3.2. Alternating and constrained least squares algorithm

Alternating and constrained least squares (ALS) is often used to get estimates of \mathbf{C} and \mathbf{S} . One reason for its popularity is that it is easy to employ constraints. This algorithm has been described in many places [1,3,7,8,15] and it is described only briefly below. The ALS algorithm with nonnegativity constraints can be described using the following:

- (0) Start with a guess for $\mathbf{C} = \mathbf{C}_0$ and set iteration counter $j=0$
- (1) Estimate \mathbf{S}_{j+1} using least squares from \mathbf{X} and \mathbf{C}_j subject to $\mathbf{S}_{j+1,n,k} \geq 0$
- (2) Estimate \mathbf{C}_{j+1} using least squares from \mathbf{X} and \mathbf{S}_{j+1} subject to $\mathbf{C}_{j+1,m,k} \geq 0$
- (3) Increment j and continue Steps 1 and 2 until $\|\mathbf{E}\|_{j+1}$ is within a set tolerance, $\|\mathbf{E}\|_{j+1} - \|\mathbf{E}\|_j$ is within a set tolerance, or j reaches a preset maximum.

In the above algorithm, $\mathbf{S}_{j+1,n,k}$ corresponds to the n,k element of \mathbf{S}_{j+1} where $n=1, \dots, N$, and $\mathbf{C}_{j+1,m,k}$ corresponds to the m,k element of \mathbf{C}_{j+1} $m=1, \dots, M$. The Frobenius norm, the total sum of squares, is given by $\|\cdot\|$.

In general, nonnegativity was employed in a least squares sense using the FASTNNLS algorithm [31]; however, functional constraints were employed using a resetting approach. The advantage of the ALS algorithm is that constraints can be applied at each iteration by simply “resetting” values. For example, for nonnegativity constraints negative values can be reset to zero before continuing to the next iteration. Resetting makes it very easy to apply functional constraints (e.g., a kinetic equation can be fit to contributions) and impose constraints on a factor-by-factor or element-by-element basis. For this reason, this method was used here. However, ALS and the resetting approach is known to have convergence problems and may not truly solve the problem in a least squares sense [15,31,32]. Plans for future work include incorporating the functional constraints in a weighted penalty function that does lead to a least squares solution.

3.3. Initial guess for images

The algorithm shown above was started with an initial estimate of the contributions C_0 . It is common to start with C_0 , instead of S_0 , because MCR has been applied to evolving data and methods have been developed that provide good initial guesses for the range of existence for contributions. Two complimentary methods are evolving factor analysis (EFA) [18] and evolving window factor analysis (EWFA) [19]. EWFA can also be applied to multivariate images by choosing a window of location in the image rather than a window of time from evolving data. However, the results tend to “smear” the image and it is difficult to translate the EWFA results directly into a good estimate of the range of existence. If a candidate set of pure analyte spectra is available, then window target factor analysis (WTFA) [33] can also be used to estimate which spectra from the candidate set could be used as an initial estimate. WTFA could also be used to provide a first estimate of range of existence for each target analyte. However, the results from WTFA are not always unambiguous and candidate sets are not commonly available for all potential analytes in a system.

Methods designed for evolving data are not as useful for initializing MCR for images because the data are not logically ordered in time. Results from principal components analysis (PCA) have also been used to initialize the ALS algorithm. The advantage of this approach is that it is simple and yields a well-conditioned basis. The disadvantage is that the initial solution does not generally satisfy nonnegativity and the orthogonal basis is usually a poor representation of the true underlying basis S that is generally oblique. PCA followed by a VARIMAX rotation, results from independent components analysis or projection pursuit might be expected to perform better than PCA [34]. D-optimality can also be used [7], but care must be taken to account for magnitude in the sample spectra. High magnitude samples tend to be in the interior of a data cloud and

thus do not generally satisfy nonnegativity. This is handled below by using mean-centered one-norm data. Another approach uses random initialization, but this can result in poor first guesses and (especially when using ALS with resetting of the constraints) can result in a solution that doesn't converge.

Other approaches to initialization are purity-based, and two typical algorithms are SIMPLSMA [4,5] and NFINDR [10]. If the image of interest has sufficient selectivity, these routines can give good estimates for the decomposition. However, it is often the case that images do not have good selectivity and the purity-based approach yields estimates that do not fully satisfy nonnegativity. This may not be a problem for interpretation, but can lead to inaccuracies in quantification [25] and process monitoring. However, purity-based solutions can be used to provide an initial guess that can be refined using ALS, for example [21]. An alternative approach used in this work builds on the purity concept to identify the extreme measured spectra in the image (i.e., the extreme sample points). This approach is related to the purity-based approaches, but uses the extremes of the mean-centered one-norm data as the initial guess. This method can be used to find an initial guess for C_0 or S_0 to initialize the ALS algorithm. It is described below for obtaining S_0 .

The method for obtaining a first estimate S_0 is based on the observation that the true spectra must lie on the exterior of the data space, and that the measured spectra at the extremes of the data space provide a useful first estimate. Fig. 1 shows an example from an MI. The measured responses X at three channels are plotted as points (i.e., three columns of X are plotted against each other). Responses at the extremes of the data cloud are shown with larger points, and the estimated pure spectra S are shown as lines. For the projections of the data points onto the K spectra to be positive, the S must be on the exterior of all the data points. The extreme samples are a good first approximation, and although a majority of the data are interior to these samples, there are a significant number that are exterior. Using the extreme samples as a first guess in the ALS algorithm and employing nonnegativity constraints moves the estimated spectra out to the location indicated by the solid lines. The resulting estimated S approach the true spectra as selectivity increases, e.g., if one or more measured responses in a pixel contain a scalar multiple of only one of the K pure analyte spectra. However, to reiterate, MI often do not have full selectivity and there is most likely a range of feasible solutions.

One method for finding the extreme samples would be to select those at the largest distance from the mean of the data cloud. The problem with this is that the magnitude of the measured responses can contribute to a sample being considered at the extreme. A one-norm of the measured responses (e.g., rows of X) restricts the data to a hyperplane with K vertices as shown in the left graph in Fig. 2 for $K=3$. (One-norm scaled data will be referred to as

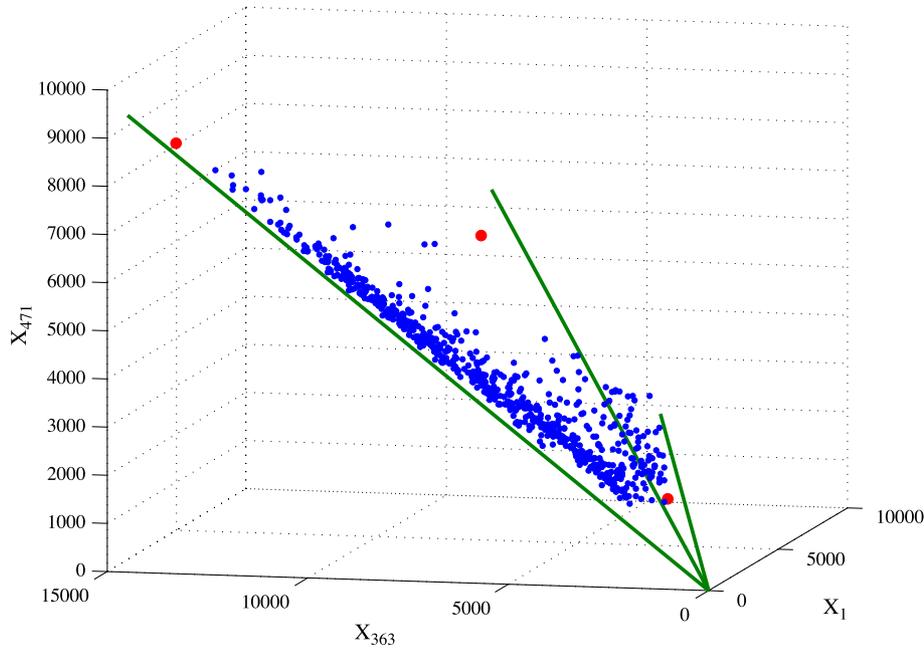


Fig. 1. Plot of typical image pixel values for three channels. The data have not been preprocessed.

\mathbf{X}_{norm}). Mean centering the one-norm data transforms the identification of extreme samples to a problem of finding the samples at the vertices of the data cloud in the hyperplane as shown in the right graph of Fig. 2 for $K=3$. (One-norm scaled data that have been mean-centered will be referred to as $\mathbf{X}_{\text{norm,mean}}$). An interesting property of a matrix with rows scaled to one-norm is that mean centering reduces the rank by one. This property holds when all the data are in a single quadrant of the data space as is the case when nonnegativity is applied to both \mathbf{C} and \mathbf{S} . It also holds when \mathbf{S} all have positive and negative elements, as for second derivative spectra, and \mathbf{C}

is nonnegative. (A good discussion of preprocessing with norms is given in reference [35]). Consequently, scores from a PCA model of mean-centered, one-norm data with $K-1$ factors can be used as a surrogate data set for identifying the K data points closest to the extremes (i.e., closest to the vertices of the data space). The PCA model of interest is given by Eq. (3)

$$\mathbf{X}_{\text{norm,mean}} = \mathbf{TP}^T + \mathbf{E} \quad (3)$$

where $\mathbf{X}_{\text{norm,mean}}$ is $M \times N$, \mathbf{T} is a $M \times (K-1)$ scores matrix, \mathbf{P} is a $N \times (K-1)$ loadings matrix, and \mathbf{E} is a $M \times N$ matrix of residuals [which is not the same as in Eq.

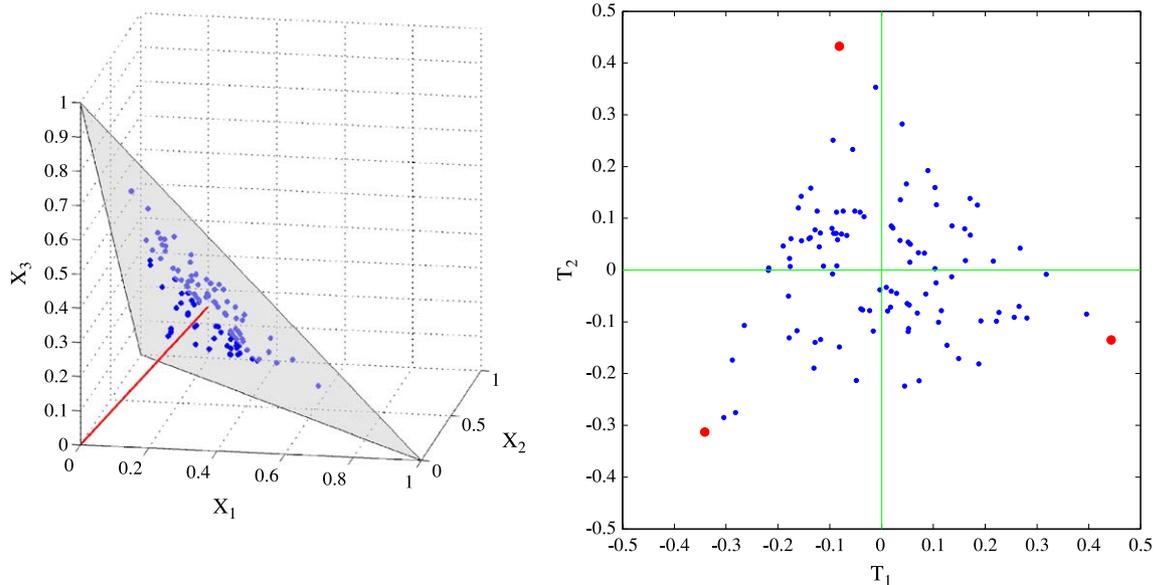


Fig. 2. Example of preprocessing image pixels to one-norm (left), and an example of mean-centered one-norm data (right).

(1)]. Only $K-1$ factors are obtained since that is the estimated rank of $\mathbf{X}_{\text{norm,mean}}$. For the example given in Fig. 2, the right graph shows a plot of the scores on principal component (PC) 2 vs. PC 1; that is, Column 2 of \mathbf{T} is plotted against Column 1. The samples with scores at the vertices correspond to the rows of \mathbf{X} that are at the extreme boundaries of the data.

The first of the K vertices can be found by identifying the sample with the greatest Euclidean distance from the center. The distance for the m th sample d_m is defined by Eq. (4)

$$d_m = \mathbf{t}_m \mathbf{t}_m^T \quad (4)$$

where \mathbf{t}_m is the m th row of \mathbf{T} . Then $d_{i_1} = \max(d_m)$ where i_1 is the index of the sample furthest from the center and the objective is to find a vector $\mathbf{i} = [i_1 \ i_2 \ \dots \ i_K]$ that contains the indices of the samples at all K vertices. The next vertex is expected to be the furthest point from the center in a direction orthogonal to the previous vertex. Thus, the next vertex is the sample furthest from the center in the matrix \mathbf{T}_{j+1} defined as in Eq. (5). Here, $j = 1, \dots, K$ is the step number.

$$\mathbf{T}_{j+1} = \mathbf{T}_j (\mathbf{I} - \mathbf{t}_{i_j} \mathbf{t}_{i_j}^T)^{-1} \mathbf{t}_{i_j} \quad (5)$$

Note that \mathbf{T}_1 is the original score matrix. This procedure can be continued until $K-1$ vertices have been found because the rank of the data decreases by 1 at each step. This procedure was originally outlined as a method for selecting samples to use in instrument standardization [36]. The K th vertex is estimated as the sample furthest from all the selected samples. The distance for the m th sample is given by $d_{iK,m}$ defined in Eq. (6).

$$d_{iK,m} = \sum_{j=1}^{K-1} (\mathbf{t}_m - \mathbf{t}_{i_j})(\mathbf{t}_m - \mathbf{t}_{i_j})^T \quad (6)$$

Thus, $d_{iK} = \max(d_{iK,m})$. This procedure can be seen as finding a set of samples that satisfy the D-optimal criteria for the mean-centered one-norm data and can be used as an initialization procedure for the D-optimal algorithm [37]. The algorithm for selecting samples based on a distance measure is referred to as the DISTSLCT routine.

Even when initial guesses are the optimally selected, the presence of rotational ambiguity makes extracting pure analyte responses difficult. For systems in which one or more components of a multicomponent system contribute a large fraction of the signal, selected samples all have a significant contribution from factors with large signal. This is manifested as widespread mixing of the large-signal component into the recovered smaller signal components. In previous work [25], sequential MCR was used to help remove some of this bias. In sequential MCR, ALS is initially performed with a subset of factors to identify the contribution distribution of the main variance-capturing components. Following that fitting, the next most unique sample (or variable) of the original data is selected, using the DISTSLCT algorithm. In this case, the DISTSLCT algorithm was used to extract the sample (or variable)

furthest from the recovered components. The selected sample (or variable) is added to the previous solution as the initial guess for the next MCR decomposition. In this way, factors are added sequentially. These steps are repeated until the model contains the desired number of components. The result is less biasing of the subsequent factors by the larger variance components. The whole process can be viewed as a sequential addition and refinement of pure-variable initial guesses as compared to a simultaneous fit starting with rough guesses for all components.

3.4. Discussion, comments, and cautions

It should be noted that using PCA scores to identify the vertices means that minor variation captured in the factors $k > K-1$ is not included in distance measures. In addition, if a norm other than the one-norm is used, the result will be a curved surface instead of a hyperplane. The curvature adds variance that influences how the PCs are determined. It is possible that variance due to curvature will be captured in the first $K-1$ PC factors which may result in a poor estimate of the vertices. In some applications, the original data, rather than PCA scores, is used.

The one-norm was used to remove sample magnitude from the data and make it easier to identify extreme samples. A problem with using the one-norm is that samples with little, or no, signal are upweighted and thus have an undue influence on which samples are considered extreme. There are at least two methods for handling this problem. One method, used in SIMPLSMA [4], is to add a small constant offset a to each spectrum prior to obtaining the norm. The effect of this is to rotate each original spectrum towards the spectrum $[1 \ 1 \ \dots \ 1]_1 \times_N$ i.e. towards the spectrum representing a vector that runs perpendicular from the one-norm hyperplane to the origin. Spectra with small norm compared to a will be moved towards the center of the data cloud more than those with large norm. Thus, a spectrum with small norm is effectively removed from the boundary of the data set and they have little impact on the solution \mathbf{S}_0 . The danger then is using an offset, a , that is too large such that small norm spectra, with high signal, are removed from the boundary. The overall result of adding an offset is to squeeze the entire data set inward and this moves the boundaries of the data set inward. The consequence is that the estimated \mathbf{S}_0 will more likely lie interior to the true solution. A second method for dealing with small norm, high noise spectra is to not include them in the distance measures (DISTSLCT). In this case, only spectra with norm greater than some lower limit are used. The danger here, like in the previous method, is that small norm spectra with high signal are removed from the boundary. Although this technique doesn't squeeze the entire data set inwards, it may remove true signal spectra and again move the boundaries of the data cloud inwards. The consequence again is that the estimated \mathbf{S}_0 will likely lie interior to the true solution. In the case where small signal-high noise pixels

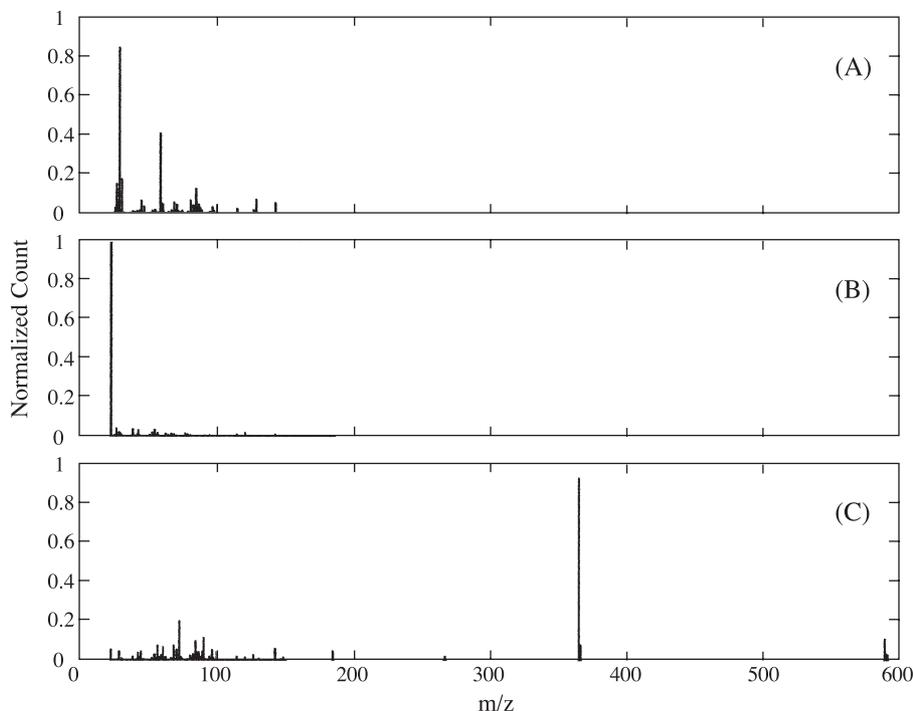


Fig. 3. Normalized count spectra extracted for the TOF-SIMS image. (A) Surelease coating, (B) sodium, and (C) prednisolone/lactose.

exist in the data, they were removed prior to using the DISTSLCT algorithm in this work.

4. Results

4.1. TOF-SIMS image of a prednisolone drug bead

Univariate images of key channels in the TOF-SIMS measurements yielded interpretable images in the original work [26]. This suggests some selectivity in the measurements and this resulted in an MCR decomposition that was relatively easy. Only nonnegativity was applied and six factors were extracted from the positive-ion image, and the initial guess was extracted from pixels with one-norm >2 . Results for three of the six factors are shown in Figs. 3 and 4. Contribution images were contrasted to enhance visualization.

The extracted spectra are shown in Fig. 3 and are different than the spectra measured on the corresponding pure materials reported in Ref. [26] and this difference was attributed to sampling artifacts. However, the pure material spectra were used to make peak assignments in the extracted spectra. Fig. 3A shows large peaks at $m/z=29$ (CH_2CH_3^+) and 59 ($\text{CH}_2\text{O CH}_2\text{CH}_3^+$). This factor was associated with the Surelease coating, and the corresponding image is shown in Fig. 4A. Note that there is a contribution of this factor in the upper left-hand corner of the image inside the drug bead. Mixing of the coating inside the drug bead is undesirable, but it may be that this coating was transported when the bead was sectioned. Fig. 3B shows a large single peak at $m/z=23$ (Na^+) and Fig. 4B shows the corresponding contribution image. This is likely due to sodium associated with the prednisolone and disassociated and mixed with the Avicel. Fig. 3C shows a peak at $m/z=365$ (lactose + Na^+) and at

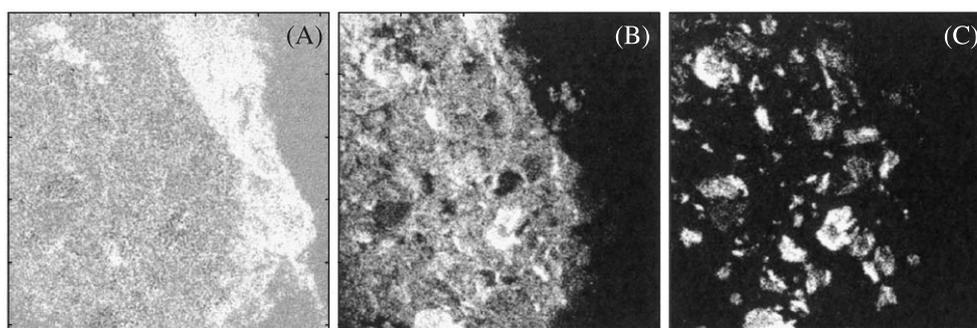


Fig. 4. Concentration images extracted for the TOF-SIMS image. (A) Surelease coating, (B) sodium, and (C) prednisolone/lactose.

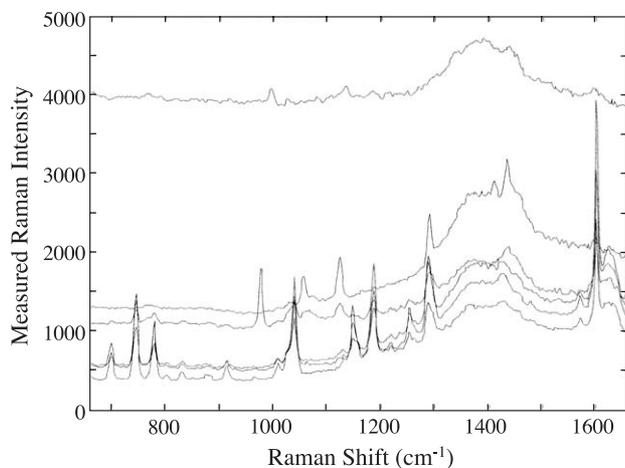


Fig. 5. The six most unique spectra selected from the aspirin and polymer image based on the distinct algorithm. Note the luminescence background present in all spectra.

589 associated with prednisolone + Na⁺. The corresponding drug image is shown in Fig. 4C. The size and distribution of the drug is readily apparent. Note that no drug is found in the bead coating. The remaining three factors (not shown) corresponded to the following. The first was “typical low-mass hydrocarbon” (major peak at $m/z=41$, CH₂CH₂CH₃⁺) found mostly interior to the bead. The second was hydrocarbons (major peaks at $m/z=27$, C₂H₃⁺; 43, C₃H₇⁺; 57, C₄H₉⁺; and 73, C₄H₉O⁺) distributed over the entire bead. The final factor was an unknown

(major peak at $m/z=130$, and minor peak at 23, Na⁺) distributed throughout the bead’s interior.

4.2. Raman aspirin/polyethylene mixture

Every spectrum collected in the MI of the aspirin/polyethylene mixture contains a significant amount of luminescence from the glass in the objective and slide. Luminescence from the glass is the major source of signal in the image and, in fact, the total integrated intensity displays the highest signal in regions where no aspirin or polymer is present to obscure the underlying glass slide. Regions of the image with aspirin and/or polymer particles show lower integrated intensity. In addition, there are several higher intensity pixels that display luminescence greater than the immediate surrounding pixels. The apparent cause was embedded luminescent particles. Finally, some luminescence was expected to be associated with the polymer. Therefore, most of the luminescence (that associated with the glass) could be considered interference, and some could be associated with signal, but the overall selectivity is low requiring external information (constraints) to resolve the image factors.

Eigenanalysis of the image data indicates that six components account for 99.996% of the sum squared signal and that additional components appear to be indistinguishable from noise based on their eigenvalues. Using this as a guideline, the DISTSLCT routine was used to select six

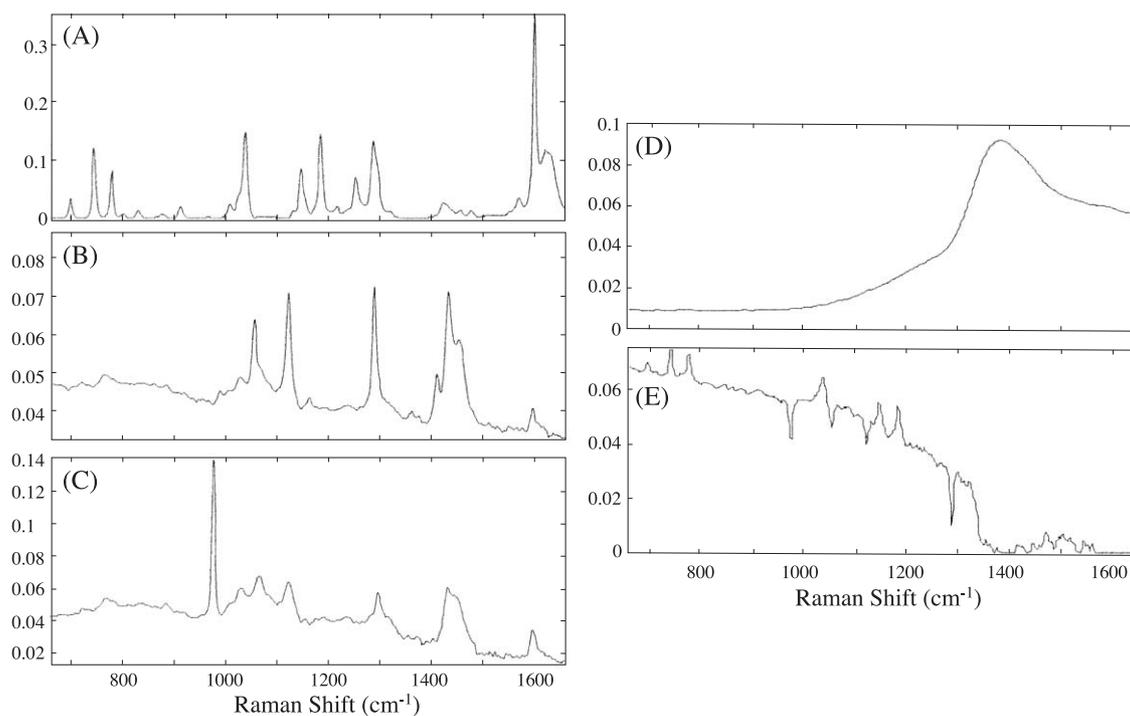


Fig. 6. Pure component spectra recovered from the aspirin and polymer mixture image using sequential ALS and equality constraints. They represent (A) pure aspirin, (B) high-density polyethylene, (C) low-density polyethylene and, (D) and (E) luminescence backgrounds. One additional spectrum was fixed to a constant offset and is not shown.

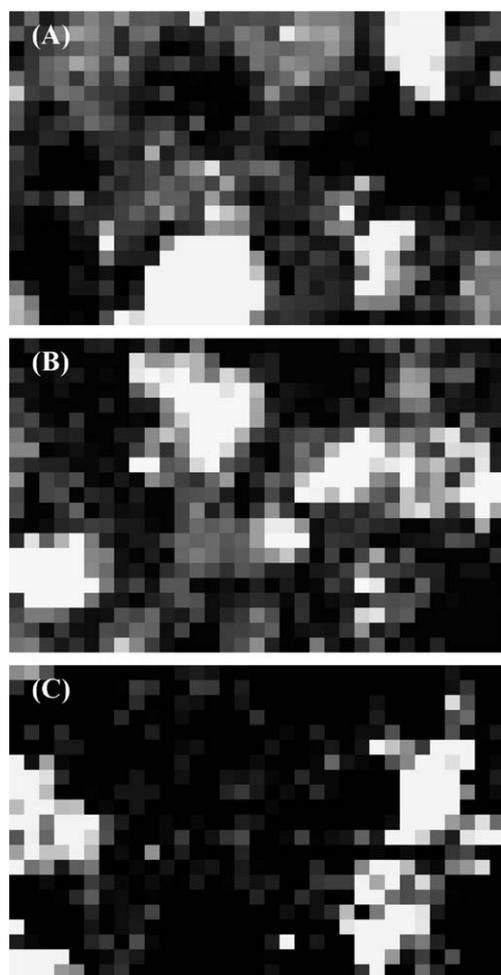


Fig. 7. Concentration images recovered for (A) aspirin, (B) high-density polyethylene and (C) low-density polyethylene.

spectra from the image. These spectra, shown in Fig. 5, demonstrate the ubiquitous nature of the luminescent background. This poses a problem of severe lack of selectivity in both the spectral domain (no unique wavelength exists in which the background is not present) and the spatial domain (no pixel exists without the background). Even with non-negative constraints, the rotational ambiguity in this system leads to ALS results in which the Raman signal and luminescent backgrounds are severely mixed and difficult to interpret.

Two approaches were used to help resolve the luminescence and Raman in this system. First, sequential MCR was used along with simple equality constraints, and second, functional constraints were used to segregate high- and low-frequency spectral shapes into different factors.

4.2.1. Raman aspirin/polyethylene mixture: sequential MCR with equality constraints

To reduce the rotational ambiguity, two equality constraints were used. First, the upper left 2×6 pixel region was identified as being only the underlying glass slide. This region was constrained to include contributions from only the first two factors (each row of C representing these 12 pixels were given the constraint that $c_{3..k} = 0$). Second, the spectral profile of another component was constrained to be constant. This has the effect of providing a fixed offset baseline throughout the ALS. The five recovered spectral profiles are shown in Fig. 6 (fixed offset factor not shown). These profiles are easily identifiable. (D) is the general profile observed for background from silica glass. Profile (A) is an unambiguous aspirin spectrum [38].

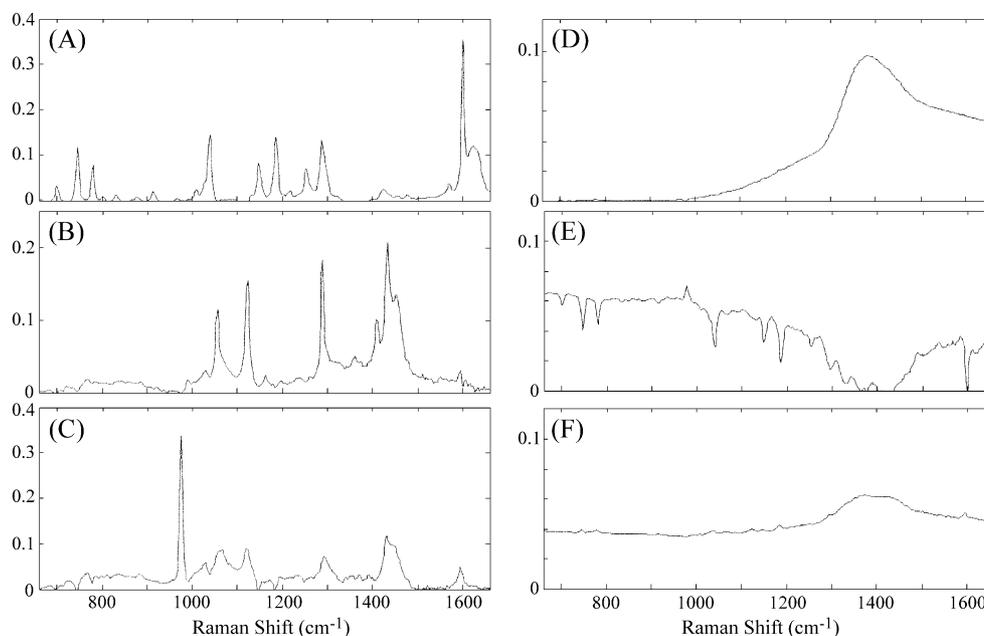


Fig. 8. Pure component spectra recovered from the aspirin and polymer mixture image using ALS with functional constraints. They represent (A) pure aspirin, (B) high-density polyethylene (C) low-density polyethylene and (D)–(F) luminescence backgrounds.

Profiles (B) and (C) are high- and low-density polyethylene spectra, respectively [39,40]. The only unexpected band in these components is the unidentified strong band at 980 cm^{-1} in profile (C). The background observed in the two polyethylene spectra and not in the aspirin spectra may be due to the intrinsic luminescence of the polymer. The final background profile, (E), is a combination of offset and variations in Raman bands. The three contribution images for the aspirin and two polyethylene factors, shown in Fig. 7, indicate the segregation of these domains on the slide.

4.2.2. Raman aspirin/polyethylene mixture: ALS with functional constraints

The previous treatment of the aspirin/polyethylene mixture required the knowledge that no particles existed in the upper left corner of the image. In many cases, such assumptions can not be made. Another solution for the same system can be achieved through the use of factor-specific functional constraints. The objective of the functional constraints on the first three factors was to reduce the low-frequency offsets from three of the factors. This was imposed using a filter on three of the extracted spectra during the ALS iterations. The filter used in this application consisted of three steps. (1) The first step estimates the derivative of an extracted factor using Savitsky–Golay. (2) The second step reintegrates the result using Bode’s rule. Steps 1 and 2 generate a reference spectrum with the offset removed. (3) Finally, the extracted factor is corrected to the reference using multiplicative signal correction. This last step was used to reduce artifacts that can be generated when applying numerical differentiation and reintegration. A smoothing filter was applied to the fourth extracted factor, again using Savitsky–Golay. The filters (functional constraints) were applied at every 10th iteration.

The extracted pure component spectra are shown in Fig. 8. The spectra associated with aspirin (A) and polymer, (B) and (C), are qualitatively similar to those recovered using sequential ALS, but they no longer have background (some luminescence is expected for these analytes resulting in a constant offset). In addition, both Figs. 6E and 8E show

“dips” in the extracted “baseline” spectra and Raman bands but are largely due to luminescence signal only. These artifacts are attributed to a lack of selectivity in the image and a rotational ambiguity. Because the aspirin and polymer spectra have been forced not to have a background offset, this background variance is fit by the true background factors. This implies that, to quantitatively describe a theoretically “pure” pixel containing only polymer requires contributions from the extracted polymer Raman spectrum as well as one, or a combination of the background spectra. The result is an incomplete separation of factors into their true pure analyte spectra which is directly attributable to the lack of selectivity in the image, and the type of constraints used.

4.3. Raman image of synthetic polymer fibers

The white-light transmission image of the fibers (Fig. 9) shows two larger fibers with two smaller fibers alternating in a horizontal row and a single fiber just below that row of fibers. Although standard nonnegative least squares gives reasonable results (not shown), rotational ambiguity led to some negative-going peaks which appear as narrow Raman bands subtracted from the expected luminescence offset but which do not drop below zero in each factor. Although the hard-baseline constraints used in the previous example could be invoked to reduce the negative-going peaks, an alternative functional constraint can also be used without the side-effect of artificial separation of Raman and intrinsic (collinear) fluorescence. The alternative requires that all recovered spectra fall above an approximate polynomial baseline for each spectrum. To impose this constraint each cycle of ALS was followed by each spectral factor being automatically fit with an approximate third-order polynomial baseline via a weighted-least squares (WLS) algorithm. The WLS algorithm uses an asymmetric error function in which values below the fit baseline are upweighted by a factor of ten before calculating the overall error in fit. Given sufficient baseline points, this allows quick approximation of the baseline shape.



Fig. 9. White-light transmission image of a cross-section of PET and Twaron fibers embedded in a low-background medium.

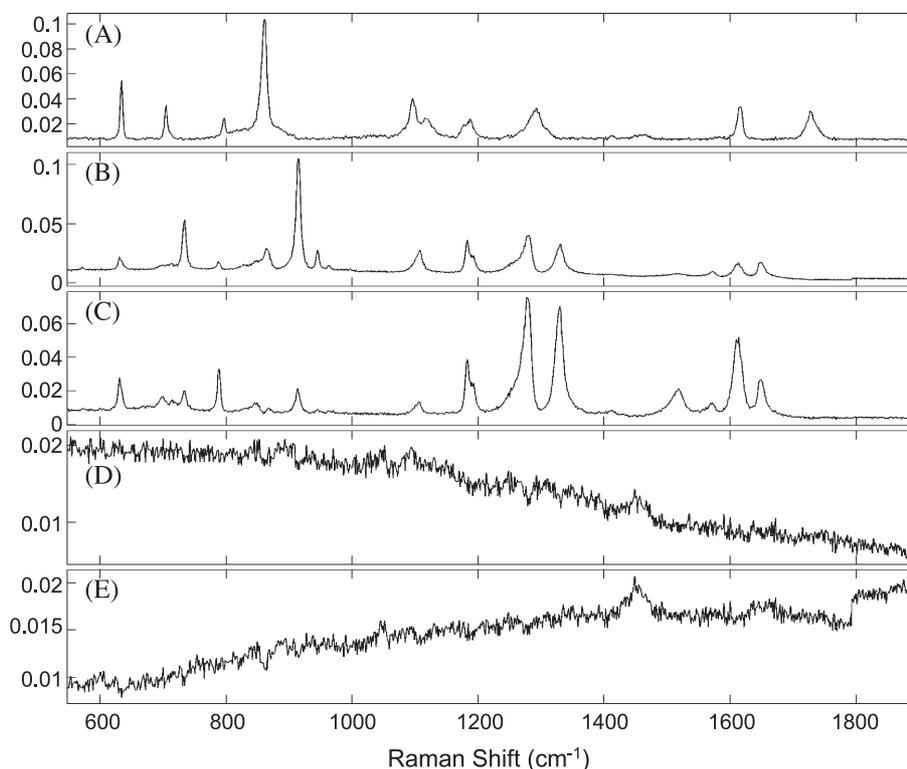


Fig. 10. Extracted spectra from PET and Twaron fibers embedded in a low-background medium: (A) PET, (B) “altered” Twaron, (C) Twaron, (D,E) luminescence.

After calculating the approximate baseline for a given factor, any spectral features below this baseline by 0.1% of the total signal were reset to the baseline. It is important to note that the baseline is not subtracted from the factor, but used only as a reference for resetting “below baseline” values and that the 0.1% of the total signal criterion allows for random noise about the baseline. That is, the offset remains in the factor after the constraint is imposed, but points significantly below this baseline (although potentially still above zero) are not permitted.

Using this method, five components were easily recovered from the synthetic fiber image. Three of the recovered spectra (Fig. 10) are unmistakable Raman spectra. One (Plot A, Fig. 10) compares favorably with a reference PET spectrum [40] and another (Plot C) with a reference Twaron spectrum [41]. The other Raman spectrum (Plot B) contains most of the same vibrational modes as in the Twaron spectrum but with different band ratios and some modes significantly enhanced. By examining the contribution images (Fig. 11), it is observed that the body of the small fibers exhibit the “altered” Twaron spectrum and the edges exhibit the “standard” Twaron spectrum. Such changes in spectra might be observed with changes in crystallinity, cross-linking, or strain which may be the cause for these two forms.[40,42]

The two additional recovered spectra (Plots D and E in Fig. 10) appear to be mostly luminescence, although the

contribution image for component D indicates a localized higher contribution near the single fiber in the lower portion of the imaged region. Previously collected Raman spectra of the cellulose fiber (not shown) indicated that the only observable band for this compound was near 300 cm^{-1} , a spectral region not analyzed in these studies. The step edge at 1800 cm^{-1} in component E is due to a change in gain at this point in the spectrum and is an expected feature.

5. Conclusions

The objective is to simplify MCR for multivariate images by (1) finding the best first estimate of the decomposition, (2) extract all chemically meaningful factors, and (3) provide a simple methodology for constraining factors in a general fashion so that knowledge of the physics and chemistry can be easily incorporated in to any MCR decomposition. Three methodologies were tested here: (1) purity-based initialization, (2) including functional constraints, and (3) sequential MCR.

A purity-based algorithm was developed to initialize an ALS-based MCR decomposition. This method used distance measures to find the extreme samples (or variables) of mean-centered one-norm preprocessed data. The method is simple, robust and easily automated (it can easily be used in an interactive environment). This method was used to

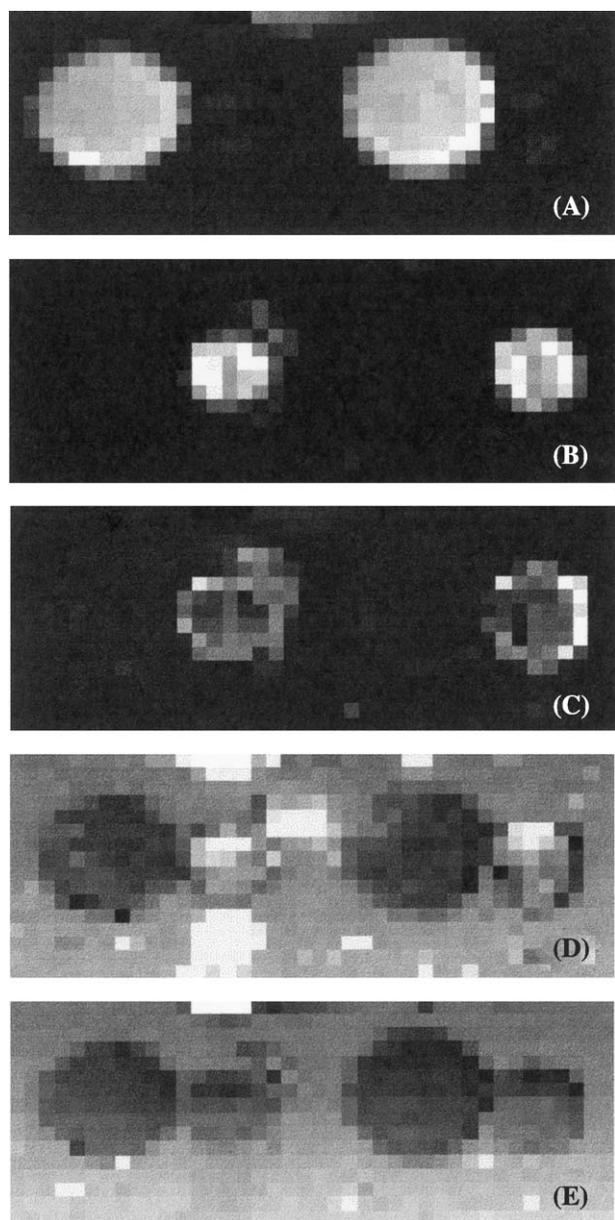


Fig. 11. Extracted contributions from PET and Twaron fibers embedded in low-background medium: (A) PET, (B) “altered” Twaron, (C) Twaron, (D, E) luminescence.

initialize an alternating least squares (ALS) algorithm when extracting all components at once and one-at-a-time in sequential MCR.

Functionality constraints were also tested using ALS. The ALS algorithm was used because it was easy to quickly incorporate functionality constraints. The prototype methodology used here showed that functional constraints could be used to account for offsets and backgrounds, and suggests that this method could be used to account for other sources of interference. The results were encouraging, and future work will incorporate the functional constraints in a least squares sense.

Acknowledgements

A significant portion of funding for this work was from the EPSRC (GR/S19073/01) and the University of Newcastle; Physical Electronics, Eden Prairie, MN, and Anna Belu provided the TOF-SIMS image; Kaiser Optical Systems provided the Aspirin and polymer fiber data; and Anne Leugers at DOW Chemical provided the polyethylene samples.

References

- [1] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics, Part B*, Elsevier, New York, 1998, pp. 243–306.
- [2] J.C. Hamilton, P.J. Gemperline, *J. Chemom.* 4 (1990) 1–13.
- [3] R. Tauler, I. Marqués, E. Casassas, *J. Chemom.* 12 (1998) 55–75.
- [4] W. Windig, B. Antalek, J.L. Lippert, Y. Batonneau, C. Brémard, *Anal. Chem.* 74 (2002) 1371–1379.
- [5] W. Windig, *Chemom. Intell. Lab. Syst.* 36 (1997) 3–16.
- [6] M.E. Kargacin, B.R. Kowalski, *Anal. Chem.* 58 (1986) 2300–2306.
- [7] F.C. Sánchez, B.G.M. Vandeginste, T.M. Hanczewicz, D.L. Massart, *Anal. Chem.* 69 (1997) 1477–1484.
- [8] B.O. Budevska, S.T. Sum, T.J. Jones, *Appl. Spectrosc.* 57 (2003) 124–131.
- [9] J.J. Andrew, T.M. Hanczewicz, *Appl. Spectrosc.* 52 (1998) 797–807.
- [10] M.E. Winter, *Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing, Altatum Geologic Conferences, Vancouver, B.C., Canada, vol. II, 1999*, pp. 337–344.
- [11] L. De Lathauwer, B. de Moor, J. Vandewalle, *J. Chemom.* 14 (2000) 123–149.
- [12] E.R. Malinowski, *Anal. Chim. Acta* 134 (1982) 129–137.
- [13] K.J. Schostack, E.R. Malinowski, *Chemom. Intell. Lab. Syst.* 6 (1989) 21–29.
- [14] P. Paatero, U. Tapper, *Environmetrics* 5 (1994) 111–126.
- [15] P.D. Wentzell, J. Wang, L.F. Loucks, K.M. Miller, *Can. J. Chem.* 76 (1998) 1144–1155.
- [16] A. de Juan, M. Maeder, M. Martínez, R. Tauler, *Anal. Chim. Acta* 442 (2001) 337–350.
- [17] S.P. Gurden, J.A. Westerhuis, S. Bijlsma, A.K. Smilde, *J. Chemom.* 15 (2001) 101–121.
- [18] H.R. Keller, D.L. Massart, *Chemom. Intell. Lab. Syst.* 12 (1992) 209–224.
- [19] H.R. Keller, D.L. Massart, *Anal. Chim. Acta* 246 (1991) 379–390.
- [20] E.R. Malinowski, *Factor Analysis in Chemistry*, Second edition, Wiley, New York, 1991.
- [21] M. Vives, R. Gargallo, R. Tauler, *Anal. Chem.* 71 (1999) 4326–4337.
- [22] R.C. Henry, B.M. Kim, *Chemom. Intell. Lab. Syst.* 8 (1990) 205–216.
- [23] S.L. Neal, *J. Chemom.* 8 (1994) 245–261.
- [24] N.B. Gallagher, D.M. Sheen, J.M. Shaver, B.M. Wise, J.F. Shultz, *SPIE Proceedings*, 5093, 2003, pp. 184–194.
- [25] Y. Batonneau, J. Laureyns, J.C. Merlin, C. Brémard, *Anal. Chim. Acta* 446 (2001) 23–37.
- [26] A.M. Belu, M.C. Davies, J.M. Newton, N. Patel, *Anal. Chem.* 72 (2000) 5625–5638.
- [27] R. Manne, *Chemom. Intell. Lab. Syst.* 27 (1995) 89–94.
- [28] R. Tauler, A. Smilde, B. Kowalski, *J. Chemom.* 9 (1995) 31–58.
- [29] P.J. Gemperline, *Anal. Chem.* 71 (1999) 5398–5404.
- [30] R. Tauler, *J. Chemometrics, J. Chemom.* 15 (2001) 627–646.
- [31] R. Bro, S. De Jong, *J. Chemom.* 11 (1997) 394–401.
- [32] M.H. Van Benthem, M.R. Keenan, D.M. Haaland, *J. Chemom.* 16 (2002) 613–622.

- [33] M.T. Lohnes, R.D. Guy, P.D. Wentzell, *Anal. Chim. Acta* 389 (1999) 95–113.
- [34] B.G.M. Vandeginste, W. Derks, G. Kateman, *Anal. Chim. Acta* 173 (1985) 253–264.
- [35] S.N. Deming, J.A. Palasota, J.M. Nocerino, *J. Chemom.* 7 (1993) 393–425.
- [36] Y. Wang, M.J. Lysaght, B.R. Kowalski, *Anal. Chem.* 64 (1992) 562–564.
- [37] P.F. de Aguiar, B. Bourguignon, M.S. Khots, D.L. Massart, R. Phan-Thau-Luu, *Chemom. Intell. Lab. Syst.* 30 (1995) 199–210.
- [38] W.E. Smith, P.C. White, C. Rodger, G. Dent, in: I.R. Lewis, H.G.M. Edwards (Eds.), *Handbook of Raman Spectroscopy*, Marcel Dekker, New York, 2001, p. 734.
- [39] K.P.J. Williams, N.J. Everall, *J. Raman Spectrosc.* 26 (1995) 427–433.
- [40] I.R. Lewis, in: I.R. Lewis, H.G.M. Edwards (Eds.), *Handbook of Raman Spectroscopy*, Marcel Dekker, New York, 2001, p. 945.
- [41] W.-Y. Yeh, R.J. Young, *Polymer* 40 (1999) 857–870.
- [42] N. Everall, in: M.J. Pelletier (Ed.), *Analytical Applications of Raman Spectroscopy*, Blackwell, Oxford, 1999, p. 127.