Multivariate analysis of hyperspectral stimulated Raman scattering microscopy images

Imaiyan Chitra Ragupathy1 | Volker Schweikhard2 | Andreas Zumbusch1

1Department of Chemistry, Universität Konstanz, Konstanz, Germany
2Leica Microsystems CMS GmbH, Mannheim, Germany

Correspondence
Andreas Zumbusch, Department of Chemistry, Universität Konstanz, 78457 Konstanz, Germany.
Email: andreas.zumbusch@uni-konstanz.de

Funding information
European Union’s Horizon 2020, Grant/Award Number: 812922

Abstract
Coherent Raman scattering microscopy is an attractive new technology for label-free imaging. Both coherent anti-Stokes Raman scattering (CARS) and stimulated Raman scattering (SRS) microscopy offer the possibility to record hyperspectral imaging data. While for the analysis of spontaneous Raman microscopy data multivariate methods are nowadays routinely employed, until to date most of the coherent Raman imaging data are interpreted using univariate data analysis. In this work, we report a quantitative comparison of the performance of different multivariate methods used for the analysis of hyperspectral SRS data from different model samples. Our data show for all samples that multivariate methods outperform univariate analysis. Using metrics to quantify method performance, we find that of the methods tested, multivariate curve resolution (MCR) gives the best results. We show that the combination with a selection of essential components based on first-order autocorrelation, gives a simple workflow for the MCR-based analysis of hyperspectral coherent Raman imaging data.

KEYWORDS
multivariate analysis, nonlinear Raman microscopy, SRS microscopy, vibrational imaging

1 | INTRODUCTION

Spontaneous Raman microscopy is an important technique for the investigation of unlabeled samples from material science and biology. Typically, for each pixel in an image, a spectrum over the whole range of molecular vibrations is recorded. This gives a wealth of information that can be exploited to analyze the molecular composition of a sample with high spatial resolution. The simplest approach to visualize the spatial distribution of a specific compound in the sample consists in plotting the Raman scattering intensity integrated over a narrow spectral width as a function of spatial position. This univariate method, however, only has a limited informative value for samples with a complex chemical composition, because the large number of vibrational bands leads to spectral overlap between the various compounds. Over the last two decades, this has motivated the development and application of many multivariate methods for the analysis of hyperspectral data obtained from Raman microscopy.\[1\text{–}4\] Toolboxes offering different multivariate analysis methods are nowadays available from several commercial vendors.

In parallel to the growing use of multivariate analysis in spontaneous Raman microscopy, new nonlinear Raman microscopy techniques such as coherent anti-
Stokes Raman scattering (CARS) and stimulated Raman scattering (SRS) microscopy have been introduced that offer substantial enhancements in signal strength over spontaneous Raman scattering.\(^5\)-\(^7\) Both methods rely on monitoring the nonlinear optical response of the sample system which is excited at two (in the case of CARS sometimes three) frequencies. Vibrational resonance enhancements are observed when the wavenumber difference between the two excitation lasers coincides with a vibrational transition of the sample molecules. Compared with spontaneous Raman microscopy, nonlinear Raman imaging offers insensitivity to background fluorescence and fast image acquisition up to video frame rates.\(^8\) The price to be paid for the fast imaging capability is that commonly only one vibrational wavenumber is monitored at a given time. Therefore, the recording of vibrational spectroscopic imaging data requires the tuning of the excitation frequency between subsequent recordings of individual images at each vibrational wavenumber. These images can then be combined to provide hyperspectral Raman imaging data similar to those obtained in spontaneous Raman scattering microscopy. In practice, nonlinear Raman imaging experiments require the initial definition of a spectral range to be probed. This allows optimizing the relation between data acquisition time and required spectral information content. Many nonlinear Raman imaging experiments have relied on univariate representations of data.\(^9\)-\(^15\) Very early after the introduction of CARS and SRS microscopy, however, also true hyperspectral nonlinear Raman imaging data have been recorded and analyzed.\(^16\)-\(^18\) Especially for biological tissue investigations, often only a small number of bands is investigated in order to allow a histopathological interpretation of the images.\(^19\),\(^20\) By using multivariate analysis techniques for hyperspectral images recorded over extended spectral regions, this approach promises in the long run to provide a spectral histology of tissues.\(^21\)

For this and similar applications, a broad variety of multivariate analysis techniques has recently been employed to analyze nonlinear Raman imaging data. In all such experiments, the task consists in the spatial segmentation of the data and their spectral unmixing. For segmentation based on the spectral information, different clustering methods such as \(k\)-means\(^18\),\(^22\) and hierarchical cluster analysis\(^23\) have been used. For the spectral unmixing, vertex component analysis (VCA),\(^24\) independent component analysis (ICA),\(^25\) and classical least square analysis (CLS)\(^26\) and recently especially multivariate curve resolution (MCR)\(^18,27\)-\(^29\) have been used to unmix the spectral information from hyperspectral SRS imaging data. In addition, various new methods have been proposed for spectral unmixing and concentration determination, notably blind factorization into susceptibilities and concentrations of chemical components (FSC\(^2\)) for spontaneous Raman, CARS, and SRS data,\(^30\) and spectral phasor analysis (SPA)\(^31\) for SRS data. Of the techniques mentioned, only MCR is frequently used for quantitative concentration measurements.\(^28,29\) When comparing the various methods, it should be remembered that the ultimate aim of the analysis is the quantitative determination of local concentrations. In principle, MCR, FSC, and SPA can be used to this end, but so far, only a few reports on the use of each of these methods have been published. It should be noted that concentration determinations with SPA are not straightforward as they involve the manual selection of regions from the phasor plot. A possible extension of these approaches is N-FINDR, which was recently used for concentration determinations from spontaneous Raman microscopic data.\(^2\)

Given that so far only few reports on the use of multivariate analysis of hyperspectral nonlinear Raman imaging data have been published, in this manuscript, we report a quantitative comparison of the performance of different multivariate analysis techniques on different model-type hyperspectral SRS datasets for spectral unmixing and subsequent quantitative analysis to extract abundance maps reflecting local chemical concentrations. In the spectral region from 2800 to 3100 cm\(^{-1}\) we recorded datasets of PMMA and PS beads as a model for spatially distributed pure compounds, of DMSO in D\(_2\)O as a model for diluted compounds, and of HeLa cells as a typical biological sample. Three different metrics were used to evaluate the quality of the image reconstructions based on the results of the different methods.

## RESULTS AND DISCUSSION

### 2.1 Spectral analysis

From the experiments, three-dimensional datasets consisting of the \(X\) and \(Y\) coordinates and the vibrational frequencies \(\Lambda\) at each spatial point are obtained. For the subsequent analysis with the different methods, the datasets are transferred in two-dimensional sets of the dimension \(XY \times \Lambda\).

### 2.2 Principal component analysis

Principal component analysis (PCA) is an unsupervised method and one of the best-known multivariate-based data analysis techniques. It is mainly used for data reduction and data clustering. The PCA algorithm performs a
linear transformation of raw hyperspectral SRS datasets into a new coordinate system in which the resulting vectors or principal components (PC) are orthogonal to each other and represent the directions of maximal variance in the raw data.\textsuperscript{32,33} Mathematically, the principal component decomposition for a dataset with the dimension \( XY \times \Lambda \) can be expressed as

\[
D = ST^T + E,
\]

where \( T \) represents loading matrix with dimensions \( \Lambda \times P \), \( P \) is the number of PCs requested (see “Determining the number of essential components” below), \( S \) is the score matrix with a dimension of \( XY \times P \), and \( E \) is the error matrix, that is, the residual of hyperspectral data that is not explained by the PCA. \( T^T \) is the transposed loading matrix. For image reconstruction, the \( S \) matrix can be reshaped back to \( X \times Y \times P \). This visualizes the abundance maps of the corresponding PCs. The effective reduction in the dimension performed by the PCA on hyperspectral SRS data explains the ability of the technique to condense correlated information from a large number \( \Lambda \) of spectral bands into a small number \( P \) of significant PCs. However, the retrieved PCs do not necessarily correspond to the spectra of the pure chemical component due to the mathematical orthogonality constraint that does not have a physical correspondence. Thus, a direct correspondence of PCs with chemical components cannot be expected, and it is not possible to generate chemically meaningful abundance maps based on PCA alone.

### 2.3 Vertex component analysis

VCA is an unsupervised multivariate data analysis technique that is mainly used for the analysis of spectral data. VCA is one of the most robust of all component analysis techniques. It works based on the assumption that each spectrum can be viewed as a vector in an \( n \)-dimensional Euclidean space, where the basis vectors are the measured frequencies. All the spectra represented in the \( n \)-dimensional space form an \( n - 1 \) dimensional hull possessing \( n \) vertices. Mathematically, this is known as a simplex. Each spectral vector belongs to a simplex, and vertices of the simplex represents the pure component spectra or endmembers. VCA starts from the original dimension of the hyperspectral SRS data and reduces the dimension step by step by the process of orthogonal subspace projection.\textsuperscript{34} The algorithm terminates when the number of endmembers reaches a predetermined value (see below). The mathematical representation of the analysis is given by

\[
D = AM^T + E,
\]

where \( D \) represents the SRS data with the dimension \( XY \times \Lambda \), \( A \) is the abundance matrix with the dimension \( XY \times P \) for \( P \) endmembers which can be used for generating abundance maps, \( M \) denotes the endmember spectra with the dimension \( \Lambda \times P \) (similar to \( T \) in PCA), and \( E \) that has the dimension \( XY \times \Lambda \) gives the variance not explained by \( A \) and \( M \).

### 2.4 N-FINDR

N-FINDR is yet another unsupervised multivariate data analysis technique based on the identification of endmembers corresponding to the pure component spectra present in the hyperspectral data. Similar to VCA, it is based on convex geometry theory, requiring that in an \( n \)-dimensional space, the volume of the simplex spanned by the pure pixels is larger than the volume enclosed by any other set of pixels.\textsuperscript{35} In order to determine the volume, the dimension of the data should be reduced to \( P - 1 \) where \( P \) is the number of endmembers to be extracted.

N-FINDR starts by assigning a random set of pixels as proposed endmembers and calculates the volume \( V \) of the simplex by replacing each endmember with each pixel. If the volume of the simplex increases, then the current pixel replaces the endmember selected before. These steps are repeated until the volume does not change anymore and the replacement of the endmembers is stopped. Note that because the pure spectra returned by N-FINDR are derived from a single image pixel, these spectra are highly susceptible to noise.

### 2.5 Multivariate curve resolution

Multivariate curve resolution-alternating least-squares (MCR-ALS) is a nonnegative matrix approximation method that decomposes hyperspectral data \( D \) \((XY \times \Lambda)\) into two matrices, pure abundance maps \( C \) \((XY \times P)\) and pure spectra \( S^T \) \((\Lambda \times P)\) as shown in Figure 1. Mathematically, MCR analysis is given by

\[
D = CS^T + E,
\]

where \( E \) is the error matrix containing the residual variation of the data \( D \). It has the same size as the data matrix \( D \). \( C \) and \( S \) are optimized iteratively using
alternating least squares so that the Frobenius norm 
\( \| D - CS \| ^2 \) is minimized under the nonnegativity con-
straints \( C \geq 0 \) and \( W \geq 0 \). Consequently, neither the
derived spectra nor the derived abundance maps can
have negative values. The convergence criterion used is
that the change in percentage of standard deviation of
residuals between two consecutive iterative cycles
reaches the predefined constraint value generally cho-
sen as 0.1%. MCR analysis has an advantage over other
factorization methods, as the results can be interpreted
in a meaningful manner in our case nonnegative abun-
dance maps and spectral intensities. As an initialization
for all the MCR analysis data shown here, we chose
spectra obtained using SIMPLISMA.\(^{36}\) By a compari-
on of this approach to initialization with results from
PCA and a random initialization, we found that the
MCR results are very insensitive to the initialization
(Figure S1).

2.6 | Essential components selection

2.6.1 | Autocorrelation

The first step in multivariate analysis is to select the
number of essential components. There are various
methods mentioned in the literature based on singular
value decomposition (SVD).\(^{37}\) These methods include
cumulative percent variance (CPV), the average eigen-
value approach, first-order autocorrelation, and parallel
analysis. In this work, we calculated the first-order auto-
correlation function of the spectral data to separate noisy
components from signal components using a loading
matrix obtained from SVD.\(^{35}\) The first-order autocorre-
sation function for the loading matrix is given by

\[
\text{AutoK}(j) = \sum_{a=1}^{A-1} m_{aj} m_{a+1,j},
\]

FIGURE 1  Schematic diagram of a spectral unmixing workflow, illustrated with the MCR-ALS method. Hyperspectral SRS data are
unfolded, and the MCR-ALS algorithm is applied to decompose the data into concentration profiles \( C \) and spectral profiles \( S \). The
concentration profile is then refolded back to obtain the abundance maps [Colour figure can be viewed at wileyonlinelibrary.com]
where $\Lambda$ represents the number of spectral bands present in the SRS data, $m$ indicates the current value of the loading matrix, and AutoK is the final autocorrelation result with size $\Lambda \times 1$. The autocorrelation function is not normalized and its values range from $-1$ to $1$. Components closer to 1 correspond to signal, components close to 0 are attributed to noise, and components close to $-1$ would be interpreted as anticorrelated. A threshold of 0.5 was set to select the number of essential components. Although the selection of the threshold is arbitrary, the chosen value worked for various kinds of samples including biological samples. This is an important advantage over using variance explained as a guide for the selection of essential components as is usually done using PCA or SVD. Also in this case, the choice of a threshold value remains arbitrary, but in contrast to the autocorrelation method, new thresholds have to be chosen for each sample investigated (cf. below).

2.6.2 | Metrics for method performance

The hyperspectral SRS data described in this manuscript were analyzed using the multivariate techniques described above. The number of essential components were determined based on the results obtained from the first-order autocorrelation function unless mentioned otherwise. For all the spectral analysis techniques, we used the lack of fit (LOF), the coefficient of determination ($R^2$), and the calculation of the relative spectral error as metrics to compare the performance of the various techniques. The mathematical expression for computing these parameters are as follows:

$$LOF = \sqrt{\sum e_i^2 / \sum d_i^2},$$

$$R^2(\%) = \left(1 - \frac{\sum e_i^2}{\sum d_i^2}\right),$$

where $d_{ij}$ is the element of the original hyperspectral SRS data and $e_{ij}$ is the residual obtained from the difference between the element $d_{ij}$ of the original hyperspectral SRS data and the result obtained from the corresponding multivariate analysis technique. Good performance is indicated by low LOF and high $R^2$ values.

In addition to these two measures for the quality of the performance of each method, we have also calculated the relative spectral error for each method following the procedure outlined in Masia et al. The values calculated in this manner indicate the unexplained spectral intensity in the images; that is, the larger the spectral error value, the larger the unexplained spectral intensity.

2.7 | Qualitative analysis

2.7.1 | PS and PMMA beads

As a first example, we investigated datasets recorded for polystyrene (PS) and poly(methyl methacrylate) (PMMA) beads in water deposited on a glass coverslip. This imaging situation is relevant for samples in which image pixels correspond to pure compounds, as is the case for many samples from material science. The first-order autocorrelation function suggested three essential components present in the data (Figure S2). Therefore, the analysis of the hyperspectral SRS imaging data for the PS-PMMA sample with the different approaches was based on three components. From PCA, we obtained score matrices for PCs 1–3, which were found to represent predominantly PS, PMMA, and water, respectively. The corresponding score and loading matrix are shown in Figure 2a–c, respectively. The three PCs account for 97% of the variance in the data. Because the score and loading matrix are defined purely based on the mathematical algorithm, the spectral information contains both positive and negative values that does not allow its chemical interpretation.

The results from VCA and N-FINDR are shown in Figure S3 and the corresponding endmember spectra are given in Figure S4. Both VCA and N-FINDR performed better than PCA, which is evident from the abundance maps, although the endmember spectrum 3 has less spectral intensity than the endmember spectra 1 and 2. This could be due to the signal intensity in the data being dominated by the endmembers 1 and 2. The abundance map and spectra obtained from MCR are shown in Figure 2d–f, respectively. A quantitative comparison of the techniques’ performance using LOF and $R^2$ values is compiled in Table 1. It shows that VCA and MCR yield slightly higher $R^2$ and lower LOF values than N-FINDR. While the performance of all the spectral unmixing techniques was better than that of PCA, MCR stands out both in the quality of the abundance maps and in the nonnegative spectra obtained from the SRS data.

While the strengths of multivariate analysis techniques are most evident for more complex samples, we were nevertheless interested how MCR would perform even for this simple sample with pure pixel components. Figure 3a,b shows the SRS images taken at Raman shifts of 2960 and 3060 cm$^{-1}$, which are dominated by signal contributions from PS and PMMA, respectively. These
images are a univariate representation of the data. Figure 3d,e, by contrast, represent the abundance maps from the MCR analysis. The overlay of the respective individual images is shown in Figure 3c,f. It is obvious that the results from MCR give crisp spatial distributions of PS and PMMA whereas the individual images taken at single Raman shifts contain contributions from the respective other chemical compound. While it can be argued that for this very simple sample, an image acquisition at the respective spectral peaks is sufficient, even here the superior quality of the MCR based data representation is striking. This finding is confirmed by the calculation of the relative spectral error[30] for images obtained from a mixture of PS and PMMA beads (Figure S5).

2.7.2 | HeLa cells

Biological samples commonly contain a plethora of different compounds and exhibit correspondingly complex spectra. In contrast to the bead sample just described,
pixels only rarely correspond to pure compounds. This is a completely different imaging situation that we modeled by recording hyperspectral SRS data sets from HeLa cells. The autocorrelation analysis again suggests three essential components (Figure S6). Figure 4a–c shows PCA results, where the first PC represents vibrations from lipids whereas the second PC has both lipid and protein (nucleus) contributions. The third PC can be interpreted as representing the PBS buffer, although the spectral window selected is too small to allow the interpretation of the score matrix in this manner. The loading vectors obtained by PCA are shown in Figure 4d. As is generally the case for PCA, these are difficult to match with reference spectra because they are orthogonal to each other, which is not in the case for real chemical compound spectra and distribution maps.

The HeLa SRS data were also analyzed using VCA (Figure 4e–h), N-FINDR (Figure 4i–l), and MCR (Figure 4m–p). It is evident that the three essential components found in VCA, N-FINDR, and MCR show the presence of lipids, protein, and water components, whereas such an assignment is not possible with PCA. Inspection of the $R^2$ and the LOF values shows that both VCA and MCR performed better for the endmember spectra extraction than N-FINDR. The calculation of the relative spectral error shows that MCR has a much lower relative spectral error of 0.10 as compared with N-FINDR with 0.16 and VCA with 0.14 (Figure S5). Most likely this is because N-FINDR is more sensitive to noise in the data. This is largely caused by N-FINDR's search for the location of the pure spectrum in single image pixels of the data and their use as endmembers. This indicates that VCA and MCR are suited to efficiently unmix hyperspectral SRS data also from complex biological samples.

To check whether the determination of the number of essential components by first-order autocorrelation analysis gives reliable results also in the case of this more complex sample, we performed the multivariate analysis also with an additional fourth component. The results of the extraction of abundance maps and the corresponding spectra with four essential components is shown in Figure S7. Apart from the first three components explained by PCA, the fourth PC appeared to mainly represent spectral noise. For N-FINDR, the occurrence of a fourth endmember spectrum similar to the water component with large fluctuations in the spectral signal was
noted. In case of VCA and MCR, the inclusion of an additional spectrum led to the splitting of the protein-rich component into two different spectral components with no meaningful spectra. The presence of signal intensities from protein-rich regions were then seen in the abundance maps 3 and 4, which are displayed in Figure S7b as an overlay image. The corresponding MCR spectra are depicted in Figure S7a. This demonstrates that the selection based on the first-order autocorrelation function indeed is an efficient way for the determination of the number of essential components for data from this more complex sample.

Also for spectral data acquired from HeLa cells, we compared results from multivariate MCR analysis with a univariate data description. Figure 5 shows SRS images containing both protein- and lipid-rich structures, taken at Raman shifts of 2860 cm\(^{-1}\) (CH\(_2\) predominantly from lipids) and 2930 cm\(^{-1}\) (CH\(_3\) predominantly from proteins), respectively, as well as the first two MCR abundance maps. It is apparent from the overlay of the two single wavenumber images that a univariate analysis does not suffice for a clean separation of protein- and lipid-rich structures. This is due to the nonvanishing spectral intensities of the lipid-rich and protein-rich
structures at both Raman shifts. By contrast, MCR yields abundance maps that clearly separate lipids and proteins. One should note here that there are several ways for the univariate analysis to separate components in similar cases. One of the most commonly used methods is spectral subtraction, in our case, for example, of the lipid SRS image (2860 cm\(^{-1}\)) from the protein SRS image (2930 cm\(^{-1}\)). This approach in fact can be used to isolate the protein-rich regions but also forces some of the intensity values in the resultant image to negative values, which makes a quantitative interpretation difficult.

2.8 | Quantitative analysis

The ultimate goal of Raman microscopy is a faithful determination of abundance maps for the various chemical compounds present in a sample. We therefore were also interested in evaluating how the MCR algorithm performed with respect to the determination of concentrations of chemical compounds. To this end, we recorded hyperspectral data for defined mixtures of DMSO in D\(_2\)O and of HeLa cells treated with different concentrations of an unsaturated fatty acid. Although multivariate techniques such as PCA, VCA, N-FINDR, and MCR are widely used for spectral analysis and to visualize the distribution of chemical compounds in vibrational imaging, we prefer MCR over the other methods. The main reason for not choosing PCA as one of the quantification methods is that the spectra retrieved from PCA cannot be expected to correspond to those of the chemical compounds present in the sample. PCA abundance maps can therefore not be interpreted as concentration maps. The differences between the other approaches tested can be seen in data recorded for differently concentrated DMSO/D\(_2\)O mixtures (Figure S8). In the original implementation of the VCA algorithm, the intensities of the different endmember vectors are scaled to the full intensity range. Therefore, also here the reconstruction of the abundance map does not directly correspond to the concentrations of the respective chemical compounds. N-FINDR, by contrast, could be used for this task,\(^{[2]}\) but as shown above, the noise level present in the endmembers is high because the endmember spectra are derived from single image pixels. Hence, we do not consider N-FINDR for the quantitative analysis. We therefore used MCR for quantification. Here, the concentrations of the respective chemical species are

![FIGURE 5](Image link) Comparison between univariate and multivariate analysis of fixed HeLa cells in the high-wavenumber region. (a) SRS image taken at Raman shift of 2930 cm\(^{-1}\) (showing protein and lipid signals); (b) SRS image taken at Raman shift of 2860 cm\(^{-1}\) (mostly lipid signals); (c) overlay of images (a) and (b); (d) component 1 (protein) from MCR analysis; (e) component 2 (lipid) from MCR analysis; (f) overlay of images (d) and (e). Scale bar: 8 μm [Colour figure can be viewed at wileyonlinelibrary.com]
directly proportional to the intensities of the abundance maps. Note that, in our implementation of MCR, the y-axis of the component spectra represents arbitrary values and cannot be used for quantification.

2.8.1 | Chemical mixtures

The DMSO/D$_2$O mixtures used here can serve as a model for compounds diluted in a solvent. For the analysis of the datasets recorded, all the multivariate techniques discussed were used to unmix the spectral components. The hyperspectral SRS data were recorded in the spectral range from 2800 to 3060 cm$^{-1}$. In the spectral region investigated, only DMSO shows a vibrational resonance that leads to only one essential component found in the autocorrelation analysis for all DMSO concentrations. However, the variance explained by the first PC decreases as the concentration decreases (Figure S9). This is one of the reasons for not determining the number of essential components based on the explained variance from PCA.

A concentration determination comparison was made between univariate and multivariate analysis techniques for DMSO/D$_2$O mixtures. Mean spectral intensities for the entire spectral range were calculated. For each concentration, the mean spectral intensity is plotted, which is shown in Figure 6a. The observed shift in the peak of DMSO toward the higher wavenumbers is caused by the variation of the polarity in the immediate molecular environment of DMSO due to the addition of D$_2$O.$^{[38]}$ Relative DMSO concentrations for the univariate analysis were calculated by considering the value at the spectral peaks for the respective concentrations. For multivariate analysis, the MCR analysis technique was used to determine the relative concentration by finding the mean of the abundance map. Relative concentration curves for both univariate and multivariate analysis are presented in Figure 6b. It is found that for the determination of the relative concentrations, both MCR and univariate analysis shows similar results.

2.8.2 | Concentration measurements in HeLa cells

We next attempted to determine the relative concentration of a compound in the more complex environment of biological cells. To this end, we performed an experiment in which we incubated HeLa cells with linoleic acid. It is known that free fatty acids are cytotoxic and that upon addition to the cell culture medium, they are sequestered into lipid droplets in cultured mammalian cells.$^{[39]}$ The addition of linoleic acid thus leads to the emergence of lipid droplets.

Hyperspectral SRS data obtained from HeLa cells treated with linoleic acid at different concentrations were analyzed using MCR. Although three essential components were used for analysis, only the abundance maps of lipid- and protein-rich components and the corresponding spectra are shown in Figures S10 and S11, respectively, for better comparison. For cells treated with high linoleic acid concentration, the lipid component spectrum obtained by MCR closely resembled the Raman spectrum of linoleic acid (Figure S12d), possessing a broad peak in the aliphatic CH stretch range and an additional small peak for the unsaturated CH stretch vibration at 3022 cm$^{-1}$. For treatments with lower linoleic acid concentrations, there is a change in the shape of the lipid MCR component spectra. In particular, the strength of the peak at 3022 cm$^{-1}$ decreases as the concentration of unsaturated linoleic acid added to the cells decreases. No signal of unsaturated CH stretch vibrations was found in lipid droplets of untreated HeLa cells. For univariate analysis, 15 lipid droplets were selected and the corresponding average intensity value at 3022 cm$^{-1}$ was calculated. Note that the average intensity of the entire SRS image at 3022 cm$^{-1}$ was not suitable to determine linoleic acid concentrations, as it overlaps with the tail of the O–H vibration band of water molecules. By comparison, relative concentrations were determined easily by MCR by finding the average intensity from the abundance map of the lipid component and plotted as a

![Figure 6](https://wileyonlinelibrary.com)
concentration curve, which is shown in Figure S13. This shows that it is possible to determine the relative concentration of various compounds present in biological samples. Multivariate analysis techniques determine the abundance map and the corresponding concentration measurements in less time with high accuracy. Though these kinds of quantification are widely used for spontaneous Raman spectral data, very few articles were published on the quantitative analysis on hyperspectral SRS data.[28]

3 | CONCLUSION

In conclusion, in this manuscript, we report a quantitative comparison between different multivariate spectral unmixing methods for the analysis of hyperspectral SRS imaging datasets of various model systems. We show that for the selection of essential components as the first important step in the application of these methods, using first-order autocorrelation yields good results. Of the methods taken into account, PCA turns out to be of limited use for the construction of abundance maps because the derived spectra do not allow a chemical interpretation. This is not the case for VCA, N-FINDR, and MCR, which were also tested. The main difference that we find for these methods is the susceptibility of N-FINDR to noise in the hyperspectral imaging data. This leads to a slightly inferior performance of N-FINDR as compared to the other two methods. Both VCA and MCR performed well. In comparison, however, MCR stands out as it can factorize the input data into abundance maps that are suitable for finding relative concentrations of different chemical compounds. As a side note, it appears worth mentioning that while all the analysis reported in this work was done using a standard laptop, also the computational cost of performing an MCR analysis was lower than that of VCA and N-FINDR. For the simple test sample chosen, the quality of the results from the concentration determinations using MCR were similar to that using univariate analysis. Both matched the actual concentration of the compound investigated. Finally, we have also shown that the hyperspectral SRS data taken at FPR can be used to unmix spectral components. When comparing the multivariate analysis methods to univariate data analysis, we find that already for limited spectral range data recorded from simple model systems as in our case, multivariate analysis delivers results of superior quality, revealing more details in the images.

The hyperspectral data sets used for the study reported here were limited to the CH stretch vibration region. Analysis of fingerprint region hyperspectral SRS imaging data is not yet as common as for spontaneous Raman data. Figures S14–S16 show that an analysis of data recorded for the PS-PMMA samples in the spectral region from 800 to 1750 cm$^{-1}$ gives results comparable with those obtained from the CH stretch region, with VCA and MCR clearly performing best. One should note, however, that this is a very simple model sample. It can be expected that the full potential of multivariate approaches will only be realized in more complex samples in which many different chemical species with overlapping vibrational spectra are present. In this respect, the possibility to record full fingerprint spectra using state-of-the-art SRS microscopy systems is a great promise for the future.[17]

4 | METHODS AND MATERIALS

4.1 | Data acquisition

Hyperspectral SRS data acquisition was performed on a Leica SP8 CARS laser scanning microscope with SRS option (Leica Microsystems, Mannheim, Germany) equipped with a PicoEmerald S Optical Parametric Oscillator (APE, Berlin, Germany) with a fixed Stokes beam wavelength at 1031.306 nm and a tunable pump beam wavelength in the range of 720–980 nm (4200–500 cm$^{-1}$). The pump beam intensity is modulated at 20 MHz using an electro-optical modulator, and the two beams are spatially and temporally overlapped and tightly focused onto the sample using 25× water objective (HCX IRAPO L 25×/0.95 water, Leica Microsystems, Mannheim, Germany). The signal intensity is collected using an oil condenser (1.4 NA, Leica Microsystems, Mannheim, Germany) in the forward direction and demodulated using a Lock-In Amplifier (Zürich Instruments, Zürich, Switzerland). For the recording of the HeLa cell images, the combined pump and Stokes laser power on the sample was 43 mW with the Stokes power being twice as high as the pump power. Images with 800 × 800 pixels were recorded with a pixel dwell time of 3.84 μs and a lock-in amplifier time constant of 1.9 μs. A frame averaging of 2 was employed.

For qualitative analysis, a mixture of 3-μm PS and 6-μm PMMA beads in water was used to acquire hyperspectral SRS data at 24 spectral points from 2800 to 3100 cm$^{-1}$ with a pixel dwell time of 8 μs. For the same region, hyperspectral SRS data were acquired at 101 spectral points from 800 to 1700 cm$^{-1}$. For quantitative analysis, dimethyl sulfoxide (DMSO) in D$_2$O with different concentrations (100%, 80%, 50%, and 20%) were measured from 2800 to 3050 cm$^{-1}$ with a total of 40 spectral points and a pixel dwell time of 10 μs. Chemical samples were measured by adding a droplet into a well of 9-mm
diameter with a thickness of 0.12-mm adhesive imaging spacer attached to a coverslip and finally covered with another coverslip. The hyperspectral SRS data acquired is of size $X \times Y \times \Lambda$, where $X$ and $Y$ are spatial dimensions and $\Lambda$ is the spectral dimension. All data analysis was performed using the MATLAB platform (MathWorks, USA) with the algorithms written in-house. For a spectral comparison, we also recorded the spontaneous Raman spectra of the compounds used (Figure S12).

### 4.2 Cell culture

HeLa cells were cultured using Gibco Dulbecco’s modified Eagle medium (DMEM), 1% streptomycin and penicillin, and 10% fetal bovine serum (FBS). Cells were subcultured on ibidi μ-Dish 35 mm high (glass bottom), at 37°C and 5% CO₂. Linoleic acid was added to the dishes with four different concentrations (33% V/V [100%], 25% V/V [75%], 18% V/V [55%], and 6% V/V [18%]) into the respective dishes and incubated for 24 h. Before the imaging, cells were washed with 1× phosphate buffer solution (PBS) to remove the traces of media and fixed using 4% paraformaldehyde (PFA) in PBS for 15 min. Finally, PBS was added to the dish and the hyperspectral SRS data were taken at 30 spectral points covering 2800–3100 cm⁻¹ at 12-μs pixel dwell time.

### ACKNOWLEDGMENTS

We thank Anayat Bhat and Franziska Rabold for help with the cell culture, Dr. Mustafa Kocademir for assisting in spontaneous Raman data collection, and Dr. Martin Winterhalder for support with the stimulated Raman scattering measurements. Discussions with Dr. Vishnu Krishnamachari are gratefully acknowledged. This project has received funding from the European Union’s Horizon 2020 research and innovation program under the grant agreement no. 812922.

### ORCID

Andreas Zumbusch 𝖠 [https://orcid.org/0000-0002-3492-0011]

### REFERENCES

[1] M. Hedegaard, C. Matthes, S. Hassing, C. Krafft, M. Diem, J. Popp, *Theor. Chem. Acc.* 2011, 130, 1249.

[2] M. A. B. Hedegaard, M. S. Bergholt, M. M. Stevens, *J. Biophotonics* 2016, 9, 542.

[3] B. Prats-Mateu, M. Felhofer, A. de Juan, N. Gierlinger, *Plant Methods* 2018, 14, 1.

[4] V. Olmos, L. Benitez, M. Marro, P. Loza-Alvarez, B. Pina, R. Tauler, A. de Juan, *Trac-Trends Anal. Chem.* 2017, 94, 130.

[5] A. Zumbusch, G. R. Holtom, X. S. Xie, *Phys. Rev. Lett.* 1999, 82, 4142.

[6] C. W. Freudiger, W. Min, B. G. Saar, S. Lu, G. R. Holtom, C. W. He, J. C. Tsai, J. X. Kang, X. S. Xie, *Science* 2008, 322, 1857.

[7] P. Nandakumar, A. Kovalev, A. Volkmer, *New J. Phys.* 2009, 11, 033026.

[8] B. G. Saar, C. W. Freudiger, J. Reichman, C. M. Stanley, G. R. Holtom, X. S. Xie, *Science* 2010, 330, 1368.

[9] E. O. Potma, X. S. Xie, L. Muntean, J. Preusser, D. Jones, J. Ye, S. R. Leone, W. D. Hinsberg, W. Schade, *J. Phys. Chem. B* 2004, 108, 1296.

[10] M. Paar, C. Jungst, N. A. Steiner, C. Magnes, F. Sinner, D. Kolb, A. Lass, R. Zimmermann, A. Zumbusch, S. D. Kohlwein, H. Wolinski, *J. Biol. Chem.* 2012, 287, 11164.

[11] W. S. Chiu, N. A. Belsey, N. L. Garrett, J. Moger, M. B. Delgado-Charro, R. H. Guy, *Proc. Natl. Acad. Sci. U. S. A.* 2015, 112, 7725.

[12] D. Lee, J. J. Du, R. N. Yu, Y. P. Su, J. R. Heath, L. Wei, *Anal. Chem.* 2020, 92, 13182.

[13] Z. L. Zhao, Y. H. Shen, F. H. Hu, W. Min, *Analyst* 2017, 142, 4018.

[14] L. Wei, Z. X. Chen, L. X. Shi, R. Long, A. V. Anzalone, L. Y. Zhang, F. H. Hu, R. Yuste, V. W. Cornish, W. Min, *Nature* 2017, 544, 465.

[15] L. Zada, H. A. Leslie, A. D. Vethaak, G. H. Tinnevelt, J. J. Jansen, J. F. de Boer, F. Ariese, *J. Raman Spectrosc.* 2018, 49, 1136.

[16] T. Iino, K. Hashimoto, T. Asai, K. Kuchitsu, Y. Ozeki, *Analyst* 2021, 146, 1234.

[17] V. Schweikhard, A. Baral, V. Krishnamachari, W. C. Hay, M. Fuhrmann, *bioRxiv* 2019, 789248.

[18] P. Wang, B. Liu, D. L. Zhang, M. Y. Belew, H. A. Tissenbaum, J. X. Cheng, *Angew. Chem. Int. Ed.* 2014, 53, 11787.

[19] T. Meyer, N. Bergner, A. Medyukhina, B. Dietzek, C. Krafft, B. F. M. Romeike, R. Reichart, R. Kalff, J. Popp, *J. Biophotonics* 2012, 5, 729.

[20] B. Sarri, R. Canongel, X. Audier, E. Simon, J. Wojak, F. Caillol, C. Cador, D. MargueV, F. Poizat, M. Giovannini, H. Rigneault, *Sci. Rep.* 2019, 9, 10052.

[21] M. T. Cicerone, C. H. Camp, *Analyst* 2018, 143, 33.

[22] J. B. MacQueen, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Los Angeles, 1967, 281.

[23] G. I. Petrov, R. Arora, V. V. Yakovlev, *Analyst* 2021, 146, 1253.

[24] A. Alfonso-García, S. G. Pfisterer, H. Riezman, E. Ikonen, E. O. Potma, *J. Biomed. Opt.* 2016, 21, 061003.

[25] Y. Ozeki, W. Umemura, Y. Otsuka, S. Sato, H. Hashimoto, K. Sumimura, N. Nishizawa, K. Fukui, K. Itoh, *Nat. Photonics* 2012, 6, 844.

[26] Y. J. Lee, D. Moon, K. B. Migler, M. T. Cicerone, *Anal. Chem.* 2011, 83, 2733.

[27] D. L. Zhang, P. Wang, M. N. Slipchenko, D. Ben-Amotz, A. M. Weiner, J. X. Cheng, *Anal. Chem.* 2013, 85, 98.

[28] K. Bae, W. Zheng, K. Lin, S. W. Lim, Y. K. Chong, C. Tang, N. K. King, C. B. T. Ang, Z. W. Huang, *Anal. Chem.* 2018, 90, 10249.

[29] K. Bae, W. Zheng, Y. Ma, Z. W. Huang, *Anal. Chem.* 2020, 92, 740.

[30] F. Masla, A. Glen, P. Stephens, P. Borri, W. Langbein, *Anal. Chem.* 2013, 85, 10820.
[31] D. Fu, X. S. Xie, Anal. Chem. 2014, 86, 4115.
[32] J. W. Einax, Anal. Bioanal. Chem. 2008, 390, 1225.
[33] M. Ringner, Nat. Biotechnol. 2008, 26, 303.
[34] J. M. P. Nascimento, J. M. B. Dias, IEEE Trans. Geosci. Remote Sens. 2005, 43, 898.
[35] M. E. Winter, Imaging Spectrom. V 1999, 3753, 266.
[36] W. Windig, J. Guilmant, Anal. Chem. 1991, 63, 1425.
[37] S. Valle, W. H. Li, S. J. Qin, Ind. Eng. Chem. Res. 1999, 38, 4389.
[38] M. Zimmerley, R. A. McClure, B. Choi, E. O. Potma, Appl. Optics 2009, 48, D79.
[39] M. D. Mantzaris, E. V. Tsianos, D. Galaris, FEBS J. 2011, 278, 519.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: I. Chitra Ragupathy, V. Schweikhard, A. Zumbusch, J Raman Spectrosc 2021, 1. https://doi.org/10.1002/jrs.6184