Available online at www.sciencedirect.com
ScienceDirect
International Journal of Mass Spectrometry 260 (2007) 222–236
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

# Extended data analysis strategies for high resolution imaging MS: New methods to deal with extremely large image hyperspectral datasets

Leendert A. Klerk [a], Alexander Broersen [b], Ian W. Fletcher [c], Robert van Liere [b], Ron M.A. Heeren [a,*]

[a] *FOM Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ Amsterdam, The Netherlands*
[b] *Centrum voor Wiskunde en Informatica (CWI), Kruislaan 413, 1098 SJ Amsterdam, The Netherlands*
[c] *ICI Measurement Science Group, Wilton Centre, Wilton, Redcar TS10 4RF, United Kingdom*

Received 7 July 2006; accepted 16 November 2006
Available online 18 December 2006

**Abstract**

The large size of the hyperspectral datasets that are produced with modern mass spectrometric imaging techniques makes it difficult to analyze the results. Unsupervised statistical techniques are needed to extract relevant information from these datasets and reduce the data into a surveyable overview. Multivariate statistics are commonly used for this purpose. Computational power and computer memory limit the resolution at which the datasets can be analyzed with these techniques. We introduce the use of a data format capable of efficiently storing sparse datasets for multivariate analysis. This format is more memory-efficient and therefore it increases the possible resolution together with a decrease of computation time. Three multivariate techniques are compared for both sparse-type data and non-sparse data acquired in two different imaging ToF-SIMS experiments and one LDI-ToF imaging experiment. There is no significant qualitative difference in the use of different data formats for the same multivariate algorithms. All evaluated multivariate techniques could be applied on both SIMS and the LDI imaging datasets. Principal component analysis is shown to be the fastest choice; however a small increase of computation time using a VARIMAX optimization increases the decomposition quality significantly. PARAFAC analysis is shown to be very effective in separating different chemical components but the calculations take a significant amount of time, limiting its use as a routine technique. An effective visualization of the results of the multivariate analysis is as important for the analyst as the computational issues. For this reason, a new technique for visualization is presented, combining both spectral loadings and spatial scores into one three-dimensional view on the complete datacube.


*Keywords:* Multivariate analysis; Principal component analysis; Synthetic polymers; Polymer additives; Imaging mass spectrometry

## 1. Introduction

The development of spectroscopic imaging techniques over the last few decades has created many possibilities for the analysis of complex systems in chemistry, biology, physics, engineering and geology. Applications range from airborne or satellite analysis of features in large land zones for agriculture [1], climate research [2] and military [3], to various microscopic techniques, including imaging Fourier transformed infrared spectroscopy (FTIR) [4–6], Raman spectroscopy imaging [7] and imaging mass spectrometry (MS) [8–10]. These techniques result in hyperspectral imaging datasets that typically contain anywhere between a few tens and a few millions of spectral variables for each image pixel. Recent technological developments allow spectroscopic imaging at higher spatial resolution, shorter acquisition times, larger surfaces and higher spectral resolution. Further sophistication of the measurement techniques and instrumental design has made various imaging techniques more accessible for the routine user.

Because of the large amount of data that is produced with these microspectroscopic techniques, data analysis can get very complicated and automated data mining techniques are required. In many cases, the measurement is not the most time-consuming part, but post-acquisition analysis becomes more elaborate due to the large amount of data obtained. Nevertheless, one would like to be able to efficiently analyze all acquired data and both get a complete overview of the data and find trace features. Moreover, when a comparison between different samples needs to

* Corresponding author. Tel.: +31 20 6081234; fax: +31 20 6684106.
*E-mail address:* heeren@amolf.nl (R.M.A. Heeren).

1387-3806/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2006.11.014

be made, computational routines would speed up the search for corresponding spectral and spatial patterns. In order to match material-specific spectral or spatial patterns between different datasets, they need to be decomposed into profiles that are specific for the various chemical components. Multivariate analysis techniques can be used for this purpose. Once the data is decomposed into components that each represent one specific combination of properties (e.g., a compound-specific spectrum), also database-matching is possible.

Multivariate statistical methods, and especially principal component analysis (PCA), are established ways to efficiently extract information from large multidimensional datasets [11]. Combined with different preprocessing and visualization methods, they form a powerful analytical tool for the analysis of hyperspectral datasets. Using these techniques, chemically relevant spectral features can be extracted from large datasets. Most of the current multivariate analysis methods, however, still require a considerable amount of operator input [12].

One of the fields of science, in which data complexity is becoming an increasing problem, is imaging MS. Imaging MS is becoming an indispensable analytical tool in many different disciplines, such as organic geochemistry, plant sciences, polymer research, biology, biomedical sciences and proteomics. This broad applicability is the driving force behind the numerous instrumental developments that are underlying the current data-explosion. A typical imaging MS measurement results in a three-dimensional datacube, containing a position on the sample and a mass spectrum. Different cross-sections of this cube give either mass-specific images or location-specific mass spectra. To make these cross-sections, either spectral or spatial features need to be known in advance or extracted from the datacube. However, when one relies on pre-existing knowledge, it is very probable that certain features are not remarked. This is what multivariate statistics could be helpful with. Various multivariate statistical methods are currently being used to computationally extract features from imaging time-of-flight secondary ion mass spectrometry (ToF-SIMS) [13–15] and matrix assisted laser desorption/ionization time-of-flight (MALDI-ToF) imaging [16] datasets. ToF-SIMS measurements result in a microscopic image of a sample surface of which every single pixel comprises a full mass spectrum. Thus, an image can be created for each mass-number. The resulting intensity map is usually visualized using a pseudo-color map. The number of measured points in a full datacube makes it impossible to be processed at the highest possible resolution with the computers that are currently readily available. A full data-matrix using 2,000,000 channel numbers (a common measurement unit for flight-time in ToF measurements) and a 256 × 256 pixels image would result in a matrix containing over 131 billion datapoints, mainly containing zeros. Storing these datasets in such a way that zero values are left out is possible when a matrix format is used that only stores the non-zero values. For highly sparse data, this reduces the amount of memory needed when the data is processed and thus yields a much more efficient memory use. The same holds for MALDI-ToF imaging data, which can be processed in a similar way.

Due to the computational constraints that are encountered when imaging mass spectrometric datasets are analyzed, multivariate decomposition methods have only been reported to be executed on either peak-selected datasets (using nominal masses) or on highly binned datasets (typical bin-sizes of not less than 0.5 amu). Only selected parts of the datasets were analyzed using full available resolution [13,17], which did indeed result in increased chemical resolving power.

The increase of computer power, as well as the availability of eigenvalue-analysis methods that are suited for large, but sparse, datasets expands the possibilities to perform multivariate statistics at much higher (spectral) resolution. In this paper, we evaluate different current multivariate analysis methods applied on ToF-SIMS and laser desorption/ionization time-of-flight (LDI-ToF) imaging datasets. Standard PCA as carried out on full matrices, as well as a MatLab™ implementation for PCA on sparse matrixes were evaluated. These two PCA methods were combined with an optimization method using variance maximization (VARIMAX) rotation. A comparison was made with parallel factors analysis (PARAFAC)-analyzed data, which is a completely different multivariate analysis method. Eventually, the usefulness of these different multivariate methods was evaluated by a quantitative comparison of computation time using different datacube-sizes. This comparison will give an indication of the applicability of these multivariate techniques as a routine data-processing method in analytical laboratories. A well-chosen, fast, yet accurate method could eventually find its way as a widely used technique in proteomics, medicine, polymer analysis, various industries and science, opening many possibilities in molecular microscopy.

## 2. Experimental

Evaluation of the different multivariate techniques was performed on two time-of-flight secondary ion mass spectrometry (ToF-SIMS) imaging datasets and one laser desorption/ionization time-of-flight (LDI-ToF) imaging dataset. All samples were analyzed in microprobe mode. The acquired data was subsequently imported and processed using MatLab™-routines.

### 2.1. Sample preparation

Two samples were studied for this paper using ToF-SIMS imaging: a purely synthetic sample containing well-defined chemical components and an embedded hair cross-section. One sample was measured using laser desorption and ionization (LDI)-ToF imaging: a cross-section of paint layers.

A droplet-array of a 1% polyvinylpyrrolidone (PVP-40.000, Sigma–Aldrich) solution in water/methanol (1/1) was spotted on a polyvinylidene difluoride (PVDF) membrane (Bio-Rad Sequi-Blot PVDF Membrane for Protein sequencing, 0.2 μm). The spotted array was created using a CHIP-1000 Chemical Inkjet Printer (Shimadzu Biotech) at 100 pL droplet volume, depositing 20 runs of five droplets at a time, resulting in a total droplet volume of 10 nL solution per spot. The incremental time between the 20 runs was chosen such that the droplets did not completely dry during the process. The pitch between the space-filling array of droplets was set at 250 μm, resulting in a minimum distance

between the centers of the droplets of 176 μm. The choice of these two well-characterized polymers offers a good reference in the comparison between different techniques.

The embedded hair cross-section was created by embedding brown Caucasian human hair in Technovit 2000LC embedding resin and light-cured for 40 min. The hair was embedded in as-received condition. Cross-sections were made using a glass-knife microtome cutting off 10 μm slices until an almost longitudinal cross-section surface was obtained. The bulk block was then gold sputter-coated (1 nm) to enhance the SIMS signal [18,19].

A sample of stacked paint layers was cross-sectioned with a surgical blade and subsequently gold sputter-coated (5 nm) for LDI imaging charge compensation. The sample consisted of alternating layers of two different Liquitex acrylic paints (Lefranc & Bourgeois, Le Mans, France), containing phtalocyanine blue ($C_{32}H_{16}N_8Cu$) and phtalocyanine green ($C_{32}Cl_{16}N_8Cu$) pigments. The alternating layers had an approximate thickness of 100 μm.

### 2.2. Data acquisition

The droplet-array sample was analyzed using an IonTof 'ToFSIMS IV' time-of-flight secondary ion mass spectrometer (IonTof GmbH, Germany) with a Bismuth primary ion gun, using the $Bi_3^{2+}$ clusters. The Bi source was run in "high current bunched" mode. The primary ion energy was set at 20 keV. The primary ion current was 0.095 pA, with a pulse-width of 0.8 ns at 200 ns cycle time. The total ion dose was kept well below the static limit [20] (maximum dose was no more than $10^{12}$ ions/cm$^2$). The beam diameter was between 3 and 4 μm on a sampled area of 500 μm × 500 μm measured at 256 × 256 pixels. The analysis was done in positive ion mode. An electron flood gun was used for charge-compensation.

ToF-SIMS analysis of the cross-sectioned hair was done using a Physical Electronics (Eden Prairie, MN) TRIFT-II (triple focusing time-of-flight) ToF-SIMS, using an $^{115}In^+$ primary ion source at 15 keV. The primary ion dose was kept well below the static limit. The sampled area was 150 μm × 150 μm at 256 × 256 pixels. The analysis was done in positive ion mode.

LDI-ToF imaging MS was performed on an extensively modified Physical Electronics (Eden Prairie, MN) TRIFT-II (triple focusing time-of-flight) mass spectrometer equipped with a phosphor screen/CCD camera optical detection combination as described in detail by Luxembourg et al. [21]. This set-up, which was originally designed for MALDI-imaging purposes, offers the possibility of both microscope and microprobe MS imaging. The time-of-flight data is recorded using a digital oscilloscope as described by Luxembourg et al. [21]. LDI microprobe imaging was performed on the paint-layer cross-section using a diode pumped solid-state Nd-YAG laser source, at 355 nm wavelength and 2 ns pulse duration (BrightSolutions, Italy). Seven linescans were made in the direction perpendicular to the layer alternation with an interval of 80 μm/linescan. The scan speed was 50 μm/s at 10 Hz laser frequency. With a laser spot-diameter of approximately 200 μm, this resulted in a microprobe-scanned image of 240 × 7 laser shots, representing an area of approximately 1300 μm × 680 μm. Each microprobe pixel represents an area of the spot-size of the laser, therefore there is an overlap between the data recorded at neighboring sample points.

### 2.3. Data preprocessing

The data was read from .GRD-files (Generic Raw Data) or .RAW-files for IonTof data and TRIFT data, respectively, using MatLab™ (Version 7.0.4, R14, SP2, The MathWorks, Natick, MA). Reading in the full data files ensures inclusion of all information recorded during the analysis. It also reduces operator time as no peak-picking is necessary. From the data read-in, a list is created containing the position as a one-dimensional representation, the channel number, $c$ (which is linearly related to the flight time), and the number of counts ($n$) for that respective occurrence. This dataset, which represents a datacube, can subsequently be converted into an $x \times y$ by $c$ unfolded datacube containing the number of counts for each spectral and spatial combination.

The LDI-ToF spectra were imported into a MatLab™ environment for further analysis. This resulted in an $x \times y$ by $t$ unfolded datacube in which $t$ represents the time-of-flight. After time-of-flight calibration a mass-spectrum is obtained for every shot and every position. No smoothing or background thresholding is applied for further analysis.

Several matching spectral and spatial components can be extracted using the different techniques for multivariate analysis on the same unfolded matrices. Commonly, the resolution that can be used during the analysis is limited by the amount of memory that is available to store a partial solution, for instance the covariance matrix in a PCA. Therefore it is necessary to use a matrix that stores the information as memory-efficient as possible, yet with a resolution high enough to obtain accurate results. Imaging MS data is generally very sparse (it has large spectral areas with zero counts), the use of a sparse matrix format is therefore an obvious choice when using MatLab™. This data-type uses a Harwell–Boeing format which leaves out the storage of zero-counts in the mass spectrometry data without loss of information. Therefore it saves memory space and thus increases the size of the dataset that can be processed.

Although the sparse matrix format allows much larger datasets to be processed, the matrix size has to be reduced due to memory limitations. This data-reduction was done by binning. Binning also reduces computation time.

Binning of the ToF-SIMS data in the ToF dimension was done on channel numbers instead of the mass scale most commonly used. This binning was done by summing the counts of a certain consecutive number of channels. All further analysis was performed on the (binned) channel numbers. Using the 32-bit integer channel numbers instead of mass-numbers, avoids round off-errors that are made when these would be converted into floating-point $m/z$ values. Using the channel integers also ensures that the spectral resolution of the decomposed data is the highest at positions where the original measurement had its highest resolution. Channel-wise binned data therefore gives a higher resolving power at lower masses, which is advantageous for height-mapping purposes [17] and when compounds with

the same nominal mass need to be resolved [13]. The mass resolution at higher masses is lower, but the resolution can still easily be kept high enough around *m/z* 500, so that nominal masses can still be resolved.

Spatial binning is used to enhance the imaging signal-to-noise ratio for lower abundant species and to increase image contrast [22]. This also reduces the amount of memory needed and decreases computation time during analysis. The spatial binning is done with a factor of 2 in each direction, so that neighboring pixels are added and no fitting between pixels is necessary. Although binning results in a decrease in spatial detail, it has turned out to be very effective in increasing image contrast, especially for images with highly sparse features [14].

The LDI-ToF imaging data was spectrally binned with a factor of 50. No spatial binning was necessary as the full dataset only contained 2030 mass-spectra.

Apart from spectral and spatial binning, other preprocessing techniques are not evaluated in this article. The effectiveness of most of the common techniques like mean centering, spectral averaging and various de-noising techniques is either questionable, or not very useful for SIMS imaging at all. Some of them also take too much computational power or operator interaction to be routinely used [14,23–26]. It is beyond the scope of this article to involve in a discussion on these methods. It is very well possible to combine various preprocessing techniques with the techniques presented here. We would only like to mention that mean centering would be not appropriate in our case, as data-processing is performed on full datasets and not only on peaks. The use of spectral mean-centering would then result in negative values for mass-numbers that actually give zero counts and therefore the interpretation of the spectral profiles that result from PCA would be much more complicated. The advantage of the sparse format would be lost as every spectral parameter would give a certain number of mean-centered spectral counts, resulting in almost no zeros in the matrix.

## 3. Methods for multivariate analysis

Data decomposition can be performed, using various different multivariate techniques. Most of these techniques use implicit statistics to compress, de-noise or decompose data by extracting statistical features. A well-known method is PCA which can be applied to compress an image, but more generally to discover patterns in high-dimensional data. PCA is therefore well-suited to be applied on hyperspectral datacubes [27], because they have both a spatial as well as a large spectral dimension. The need for a method that can automatically extract features from spectral data increases with the increasing resolution of the datacubes resulting from ToF-SIMS.

A balance has to be found between the accuracy and the time it takes to produce the results for different methods. We compare some common methods for multivariate analysis with their performance on imaging mass spectrometry data of different resolution and acquired with different acquisition methods. Performance depends on many variables such as available memory, data complexity and the implementation of the algorithm. An indication of this performance is given by relative comparisons of the time it takes to do the calculations. The qualitative performance will be judged based on the contrast in the spectral profiles and the feature-contrast in the image planes.

### 3.1. PCA

One way to find the principal components in an unfolded datacube $X$ is by eigenvector decomposition. Both the spectral and the spatial dimension are decomposed into uncorrelated spectral and spatial components. Eq. (1) describes the PCA decomposition which can be solved by finding the eigenvectors of the covariance matrix of $X$.

$$X = Y \cdot P^{\mathrm{T}} \tag{1}$$

The first dimension of $X$ contains the locations in of the unfolded datacube, and the second dimension describes the channels. The columns of $P$ contain the orthonormal loading vectors and columns in $Y$ the score vectors or spectral profiles in this case. Together, these components describe the original datacube in principal components and can be used to compress the datacube or extract correlated spectral and spatial features.

The resulting component images contain the spatial distribution of the corresponding spectral profiles. The components are sorted according to their variance expressed by the eigenvalues from the eigenvector decomposition. The first components contain the largest contribution to the original datacube and the last components mostly contain the remaining noise. Each spectral and spatial component can contain both positive and negative values which make interpretation not very intuitive. One way to deal with these negative values in $Y$ is by splitting them in a positive and a negative counterpart. Each of these parts will create loading vectors that result in positive-only score images when they are multiplied with the transposed original matrix $X$.

Using the sparse matrix format within MatLab™ saves memory space, computation time and allows larger datasets to be analyzed. MatLab™ also provides a sparse implementation to find the eigenvectors and eigenvalues of a sparse matrix. It uses the FORTRAN library ARPACK [28] which uses an implicitly restarted Arnoldi iteration to solve the eigenvector problem for a sparse matrix [29]. One input parameter is the number of eigenvectors that have to be found. Other parameters control the convergence, number of iterations or model-specific solutions.

Using this function, the two most important limitations in eigenvalue analysis of large datasets can be largely circumvented. Firstly, the amount of memory needed is considerably smaller than the memory needed to store the same data in a full matrix. Secondly, the calculation time can be decreased as the implicitly restarted Arnoldi method only makes an estimation of the first (user defined) number of eigenvectors. The fact that only a few eigenvectors are calculated, implies that the resulting set of eigenvectors does not form a complete orthonormal basis for the original dataset. This means that if this method is used for PCA, the resulting principal components form only the most important part of the original dataset (whereas the PCA results obtained with the use of traditional eigenvalue determination methods give a complete basis and can therefore be

back-transformed into a dataset that is exactly the same as the original data). In practice, this does not give any problems as hyperspectral datasets generally contain only a few predominant principal components. Therefore, within the first 20 (or even less) principal components, almost all original data can be described. From a data-compressing point of view, this is sufficiently accurate as well, as usually only a limited number of principal components are stored, still containing more than 99% of the information from the original dataset.

### 3.2. PCA and VARIMAX

Additional optimizations can be done after a PCA. One method is an additional fitting of the principal components to maximize the variance expressed in each component. There are a number of maximization criteria but the VARIance MAXimization (VARIMAX) from Kaiser [30] is the most common. It can be used as a post-processing step after a PCA. By rotation of the orthogonal axis, more simple structures, and components with a higher contrast are created. The expression in Eq. (2) can explain the relation with PCA.

$$X = Y \cdot R \cdot R^{-1} \cdot P^{\mathrm{T}} \tag{2}$$

$X$ is again the original unfolded datacube, with $Y$ the scores and $P$ the loading vectors. The VARIMAX algorithm tries to find an orthonormal rotation matrix $R$ so that the variance of the squared spectral components is maximized. The value of minimum relative increase of the objective function to keep on iterating is kept on the default of $10^{-5}$. The spectra belonging to the different principal components are then plotted as $R^{-1} \cdot P^{\mathrm{T}}$. Related images are scores of this rotated vector.

### 3.3. PARAFAC

A more generalized decomposition method in this study is using the PARAllel FACtors (PARAFAC) model of Harshman [31]. Its exact model was independently proposed by Carroll and Chang [32] as CANonical DECOMPosition (CANDECOMP). This model uses fewer degrees of freedom to fit the data on a simple model for decomposition. It gives a unique solution for the decomposition and makes it possible to put constraints and weights on the resulting components. These constraints can be orthogonality, non-negativity or unimodality with implicit non-negativity. Eq. (3) gives a representation of PARAFAC that can be compared with the model used in PCA.

$$X_k = Y \cdot D_k \cdot P^{\mathrm{T}} \tag{3}$$

$X_k$ is again the unfolded datacube that is decomposed in loading vectors in $P$ and score vectors in $Y$. $D_k$ is a diagonal matrix giving a unique weight to each component. The user has to set the number of components in which $X_k$ has to be decomposed. We used the fast and optimized iterative implementation described by Bro [33] to decompose the datacube with only the non-negativity constraint on matrices $Y$ and $P$. An orthogonality constraint could not be applied together with the non-negativity constraint. The convergence criterion was the default value of a relative change in fit of less than $10^{-6}$. The MatLab™ implementation that was used is able to handle sparse matrices using the same computational algorithm. This enables us to compare the components from a common and sparse PCA with the extracted components from the PARAFAC model.

### 3.4. Three-dimensional visualization of extracted features

Broersen and van Liere [34] describe a technique to visualize correlated spectral and spatial features in a three-dimensional volume using PCA. The main characteristics of the scores and loadings of an extracted principal component are highlighted in the original hyperspectral datacube using opacity maps. This enables a user to select a principal component and interactively view the spectral and spatial contribution within the three-dimensional representation of the datacube. Instead of looking at a solid cube filled with ion-counts, an opacity map gives more 'insight' to the data by hiding specific regions using PCA. The opacity map of each component can be adjusted by changing a threshold which controls the amount of data points that is shown. The resulting principal components after the VARIMAX post-processing and the extracted components of PARAFAC have similar properties as those of the PCA. These components can also be used to create three-dimensional opacity maps to automatically create highlights within a datacube and reveal correlated features.

## 4. Results

PARAFAC and PCA, along with VARIMAX post-processing were applied on the two datasets described earlier. PARAFAC gave the best separation of chemical components. Therefore the assignment of different chemical components in the samples will be done using the PARAFAC results. After that, a comparison is made between PCA with and without post-processing, along with a comparison with PARAFAC. It is very difficult to perform an exact quantification of the quality of a multivariate analysis unless done on a synthetic dataset, which was not done here as it is ambiguous how a synthetic dataset representative for a real sample would look like. We therefore qualitatively assigned the methods and compared the different implementations (for sparse and non-sparse matrix formats) of the various methods within the same dataset. As a quantitative comparison, the computation time of different methods was compared to give an indication on the usefulness in routine analysis.

### 4.1. PARAFAC

A PARAFAC analysis was done on the full acquired dataset of the PVP-droplet sample, with a spectral binning factor of 1024 and a spatial binning as a factor of 2 (resulting in 128 × 128 pixel images). The dataset was cut-off at the 1250th binned channel (which corresponds to $m/z$ 345). This reduction of the dataset was necessary to prevent out-of-memory errors during this computationally demanding method. In this case, the restriction of the dataset to a maximum of $m/z$ 345 does not influence the

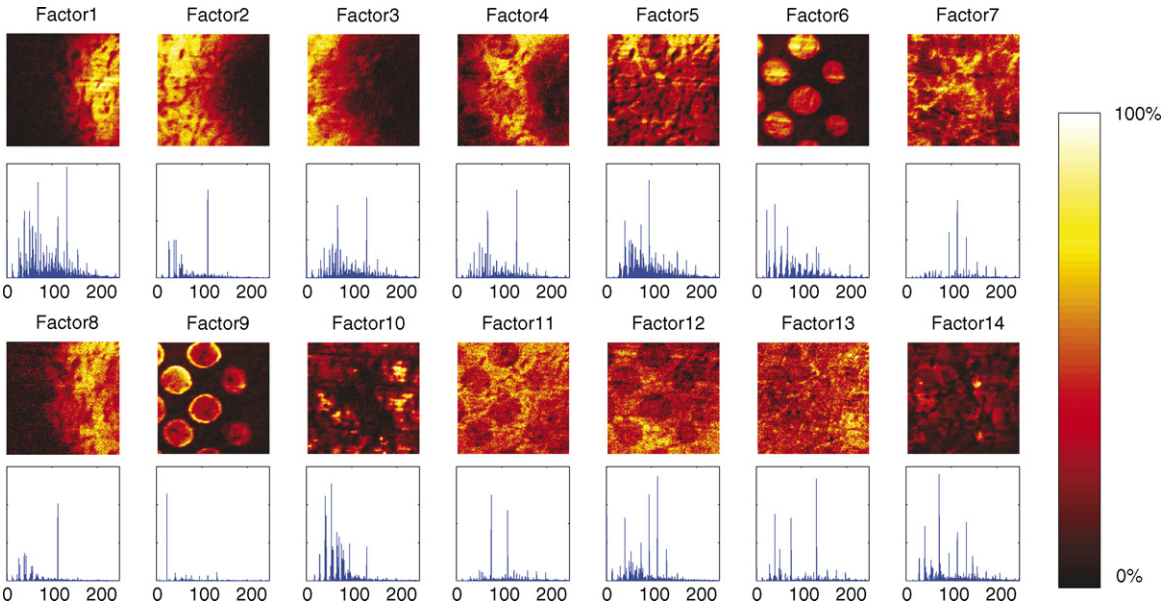


Fig. 1. Score plots and loading vectors for 14-component PARAFAC analysis of +SIMS analyzed PVP spots on PVDF membrane. Only positive scores are possible due to the non-negativity constraint that was used in the analysis. Each image is scaled to its maximum (absolute) intensity.

results of the statistical analysis. In some cases, it may be necessary to either perform an additional analysis within the higher *m/z* region or to increase the binning factor. PARAFAC was then done using 7, 14 and 21 decomposition variables (factors). This number of factors was chosen arbitrarily, based on the convenience of the seven-color plotting scheme in MatLab™ whilst assuring a wide enough range to cover all components. A non-negativity constraint was put on the components, so that only positive scores on the channel numbers were allowed for each variable. The numbered order of the different decomposed components is arbitrary. Therefore, based on the order of the resulting component-spectra, no conclusions can be drawn on the abundance of the corresponding factors. That means that factor 11 could be more abundant than factor 2, which is an important property to recon with when the analysis is done. The random order of the factors is a result of the random initialization that was chosen. Therefore, the order of the factors could vary between different PARAFAC runs on the same dataset.

A few chemical components gave one specific factor in PARAFAC, irrespective of the number of chosen factors to be resolved. These included the "salt rim", which is the result of transportation of salts to the edge of a drying droplet, and the PVP droplet itself (factors 9 and 6, respectively, in the 14-factor PARAFAC analysis, Fig. 1). This can also clearly be seen from the corresponding spectra (Fig. 2): factor 9 gives a very high score at *m/z* 23 ($Na^+$), a low but distinct peak for $Li^+$ (*m/z* 7) and only minor scores for other species. $Na^+$ is highly abundant in factor 6 as well; in any other factors, its presence can be neglected, showing that the $Na^+$ indeed comes from the solution used during the spotting procedure. Factor 6 shows peaks at positions that are specific for PVP [35], along with the peak at *m/z* 133, which seems to be present in almost all factors (closer examination of the unbinned spectrum shows that this nominal mass indeed contains multiple peaks with different exact masses). Assignment of the various peaks in the spectra belonging to factors 6 and 9 is indicated in Fig. 2. Factor 10 and, to a lesser extent factor 14, show distinct structures. These localized components represent a contamination on the PVDF membrane that was introduced when the membrane was attached to the substrate. The inner side of a polyethylene bag was used to tighten the membrane onto the substrate and factor 10 results from erucamide ($CH_3(CH_2)_7CH{=}CH(CH_2)_{11}CONH_2$), which is commonly used as a slip agent for polyethylene. A distinct $[M+H]^+$ peak for erucamide was seen at *m/z* 338. It is striking how well this low-abundant surface component is resolved.

Special attention needs to be paid to the different ways the signal from the PVDF membrane is decomposed into various factors. These factors can be identified by the typical 20 or 38 amu separated peaks due to, respectively, HF and 2F mass difference between the fragments. This PVDF-related chemical component is divided into various factors for PARAFAC using 14 or 21 factors, all containing a different combination of PVDF-specific peaks. However, for the seven-component analysis, the substrate membrane was only divided into a few components (most of which seemingly represented different height zones, as can be concluded from a comparison with the PCA analyses where similar images have height-specific spectra). The division of this single chemical component into multiple factors is a result of the orderless factorization, which seeks for a fixed number of components that have no specific order of importance.

Also the cross-sectioned hair was analyzed with PARAFAC. Decomposition into seven factors gives one factor that is specific for the embedding medium. All other factors seem to be non-specific, suggesting that there are no other chemical components present in the analyzed dataset. However when 14 factors are allowed, a specific spot is resolved, spectrally corresponding to a peak at *m/z* 39, which results from $K^+$. More close analysis of the factors showed that also the very low-abundant $Na^+$ (*m/z* 23) is specifically localized at this position and only present in this certain factor (Fig. 3). This once more shows the power

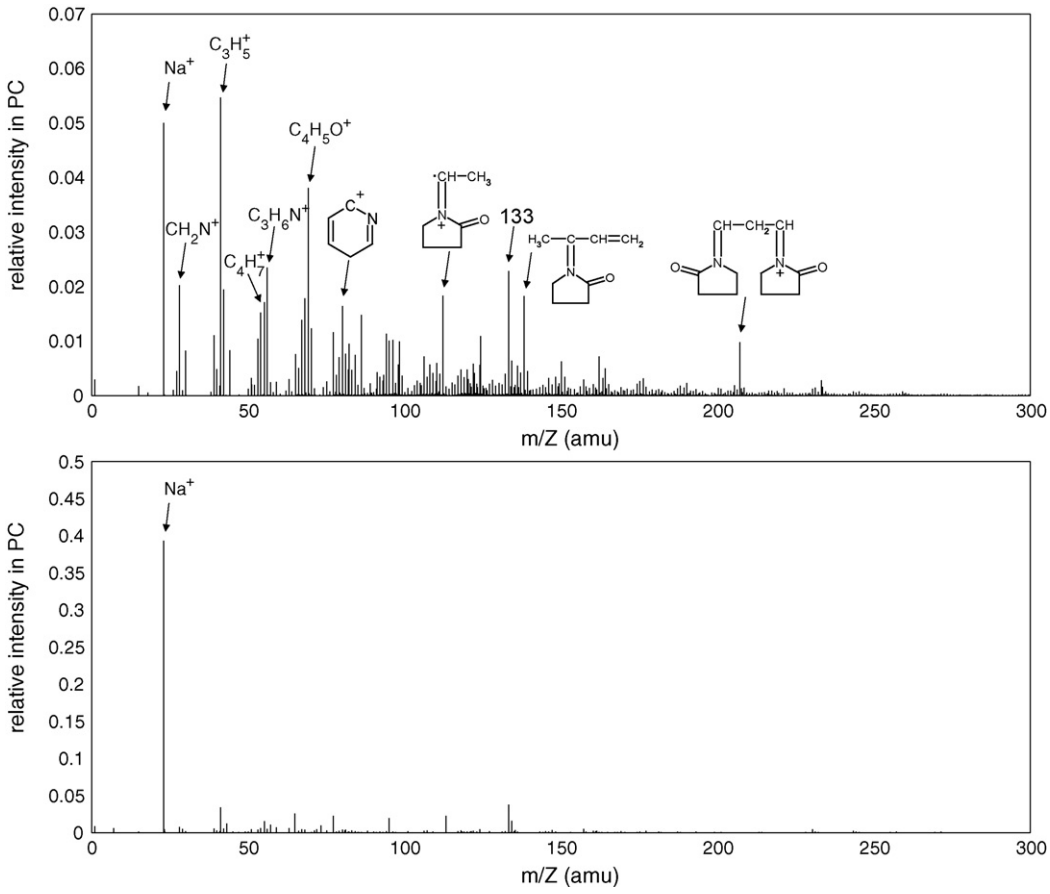


Fig. 2. The loading vectors of PC 6 (top) and PC 9 (bottom) as obtained from 14-component PARAFAC analysis of +SIMS analyzed PVP droplets spotted on PVDF membrane.

of computational analysis as this feature would not be resolved without the use of statistical data analysis.

LDI-ToF microprobe imaging data was analyzed with PARAFAC. The two known components (phtalocyanine blue and phtalocyanine green) were clearly resolved by PARAFAC, irrespectively, the number of chosen factors (Fig. 4). Phtalocyanine blue is seen as $M^{+\bullet}$ ion at *m/z* 575 and as $M_2{}^+$ at *m/z* 1150. Phtalocyanine green shows a 35 *m/z* spaced profile from *m/z* 1127 ($M^{+\bullet}$) down to 915 ($[M-6Cl]^+$). Each spacing of approximately 35 *m/z* represents a chlorine loss. Overestimation of the number of factors results in the splitting of single chemical components into different factors, as can be seen from the 14-factor analysis of the LDI-analyzed paint-sample. Factors 3 and 13 (phtalocyanine blue), as well as 4 and 14 (phtalocyanine green) show very similar localization but are nevertheless represented as different factors. This is not necessarily an artifact from the high number of factors, and could as well result from correlation between various measured ions within one layer.

Conclusively, PARAFAC was able to extract chemical features into single components. However, pre-knowledge is favorable as it will factorize the data into a certain, user defined, number of factors. In essence this number can be made high enough to surely exceed the number of actual chemical components. However, this will lead to the factorization of one actual component into a number of factors, as can be seen from the 14-factor analysis. This over-factorization cannot always be avoided as was shown for the extraction of the salt-crystal in the hair cross-section.

### 4.2. PCA on sparse datasets

PCA was done using an in-house developed toolbox based on standard library functions in MatLab™. PCA was done on the PVP array spotted on PVDF membrane. All decompositions were restricted to the first 20 principal components, unless mentioned otherwise. This number of PCs was chosen after analysis of higher PCs, which only yielded non-specific spectra and features. The preferred spectral and spatial binning factor should be chosen dependant on the character of the dataset (which in turn depends on the measurement circumstances), the type of compounds of interest and the intensity of the signal. A trade-off has to be made between these parameters and the amount of memory that is available for the analysis.

A complicating factor of the comparison with PARAFAC is that no model information can be used during PCA. Therefore a non-negativity constraint cannot be used, which implies that one PC can actually contain two chemical components, if they are anti-correlated. This means that in a two-phase system, all chemical information could be contained in a single PC. In the case of the PVP droplet array on PVDF, this was seen as a combination of a PVP-specific loading vector, together with an anti-correlated PVDF spectrum because the PVP partially

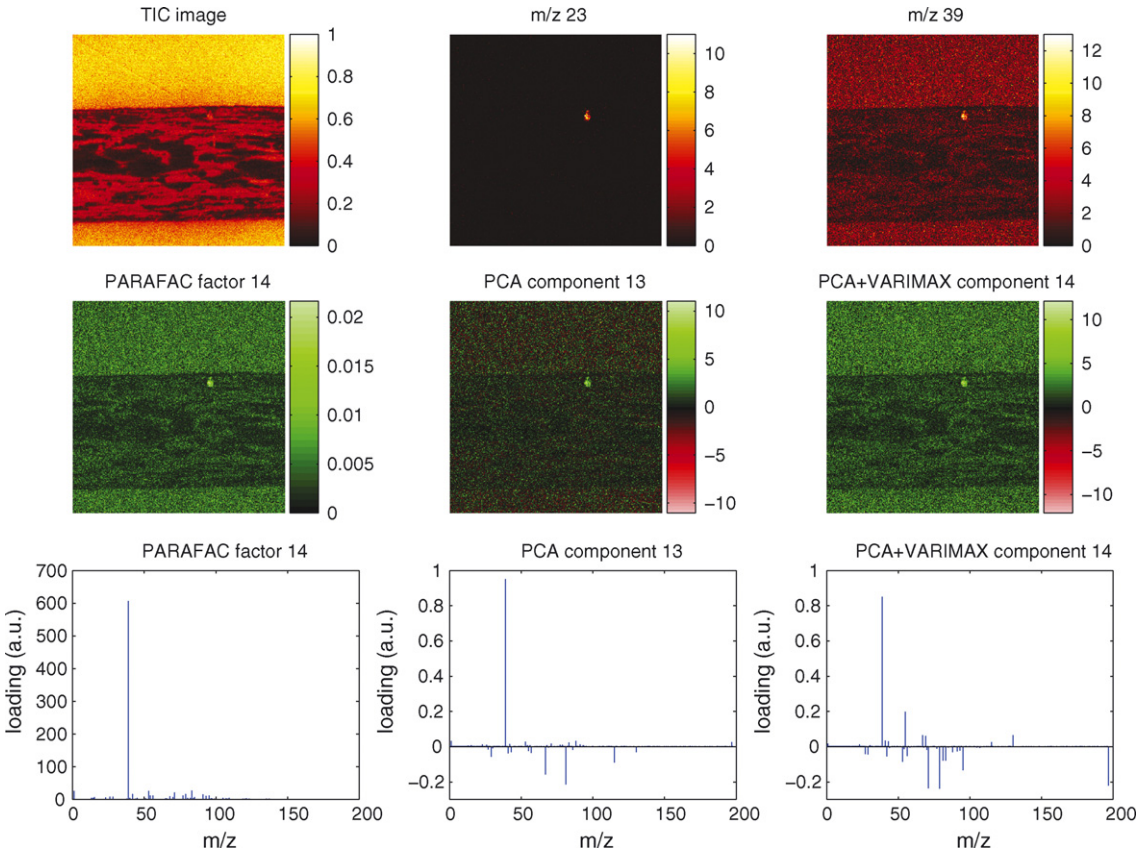


Fig. 3. SIMS images of the cross-section of hair embedded in a methacrylate embedding medium. The total ion image (top left) does not give any indication for the salt crystal as shown with the images at *m/z* 23 and *m/z* 39 (top-middle and -right, respectively). The bottom images show the score images of the major resulting factors from the three investigated methods. The different color maps are chosen for visibility reasons: a "hot" color map as used in the top images would not be appropriate in the PCA score images because of negative values.

covers the PVDF membrane and therefore absence of PVDF comes together with presence of PVP. For the cross-sectioned hair, just like with PARAFAC, the only chemical components that were resolved was the salt crystal (seen in one of the higher PCs) and the embedding medium. This implies that PCA is a suitable technique to resolve small features as long as the chosen number of components is high enough.

The spectral features specific for the Liquitex layered-paint sample that was measured using LDI, were decomposed into more than two principal components by PCA. When 20 PCs were chosen, both the phtalocyanine blue and phtalocyanine green are found in various PCs, often as anti-correlated features. This results in PC score-images that give very little localization (Fig. 8a).

The results obtained with PCA using the implementation for sparse matrixes were compared with PCA as done on full matrixes for the PVP droplet-sample. A comparison was made by calculating the average of $q_{in} = p^{F}_{in}/p^{S}_{in}$ in which $p^{S}_{in}$ is the $n$th sparse-type PC result and $p^{F}_{in}$ the $n$th standard (full-matrix) result in the $i$th spectral dimension. For identical datasets this would give $q_{in} = 1$ for any $i$, $n$. For datasets that are not correlated, the variation in values for $q_{in}$ would be very high. The average values as well as the standard deviation for the first 100 PCs are plotted in Fig. 5, together with the first 10 maximum values of $|q_{in}|$ for each $n$ and their average. This plot shows the high correspondence of the results of the two PCA methods, as can be concluded from the few high values for $q_{in}$. This is confirmed by the average of all $q_{in}$ values, which is close to 1 (only a little bit smaller due to a few $q_{in}$ values that are equal to zero due to a $p^{F}_{in} = 0$; $p^{S}_{in} \neq 0$ for all $i$, $1 \leq n \leq 100$). Larger standard deviations are more common at higher PC numbers. This is explained by the fact that higher PCs represent less chemical information. This can result in a different PC order for different computational methods as well as a less-exact definition of the PCs by themselves (actually noise is compared with noise). Ill-defined PCs result in high values for $q_{in}$ in some cases. Fig. 5 indeed shows that up to PC12 the two methods give identical results, with increasing PC number, this error also increases, as well as the variance in the highest 10 values for $q_{in}$.

### 4.3. VARIMAX post-processing

VARIMAX rotation was used to enhance the spectral contrast of the PCs. This axis-rotation results, as expected, in higher contrast not only in the spectra, but also in the images. The resulting PCs do not necessarily correspond with the original PCs. This is shown for the PVP droplet-array, where the loadings vector of PC3 looks completely different after VARIMAX rotation. The rotation did indeed increase both spectral and image contrast, as

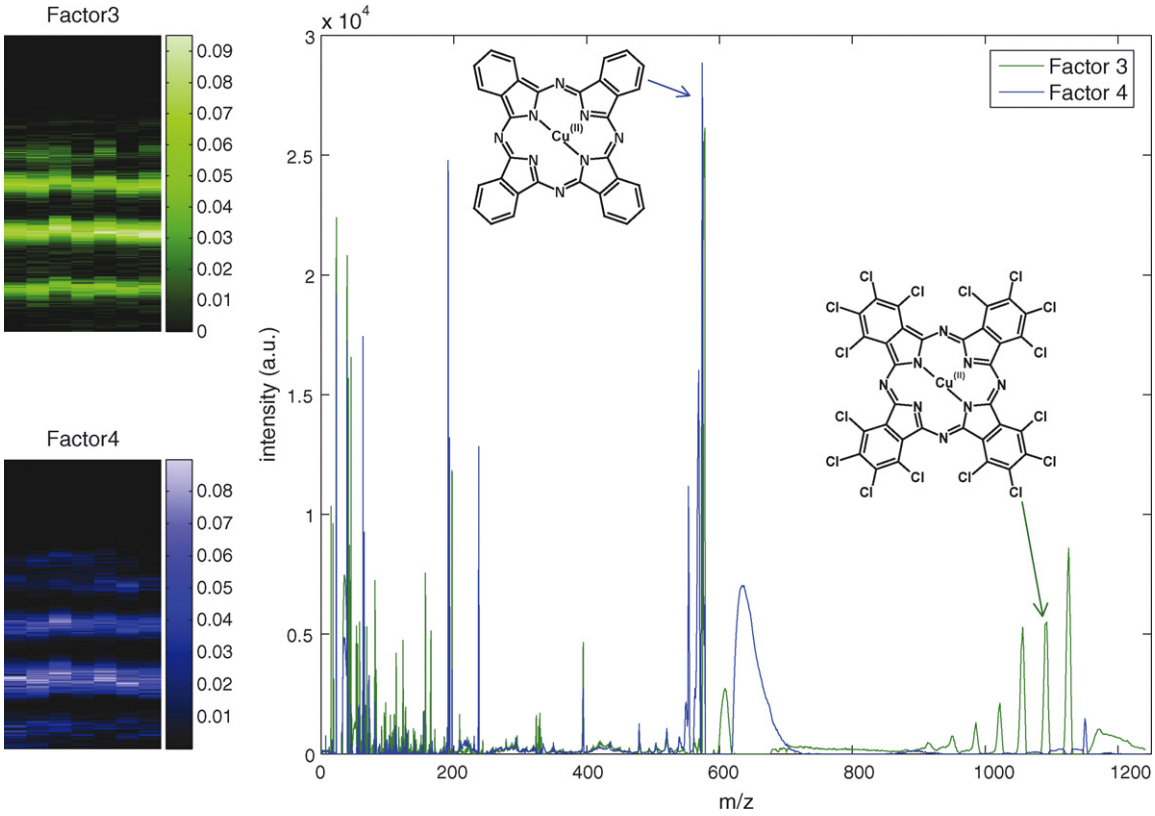


Fig. 4. Score images (left) and loading vectors represented as mass spectrum (right) of the PARAFAC analysis of a LDI-microprobe imaged stack of Liquitex paint layers.

can be seen from Fig. 6 (before optimization) and Fig. 7 (after optimization). Especially, the predominant chemical features that were mentioned earlier in the PARAFAC analysis were represented by single components. After VARIMAX, PC2 contains only the PVP-specific spectral features (as negative peaks) with a few anti-correlated components (as positive peaks) whereas many other peaks were observed before optimization. The optimized PC6 shows hardly any peaks apart from the $Na^+$ signal at *m*/*z* 23 whereas it was hardly resolved before VARIMAX.

The representation of height differences, which is typically observed in the PC-spectrum as the combination of both a positive and a negative peak within the same nominal mass-number

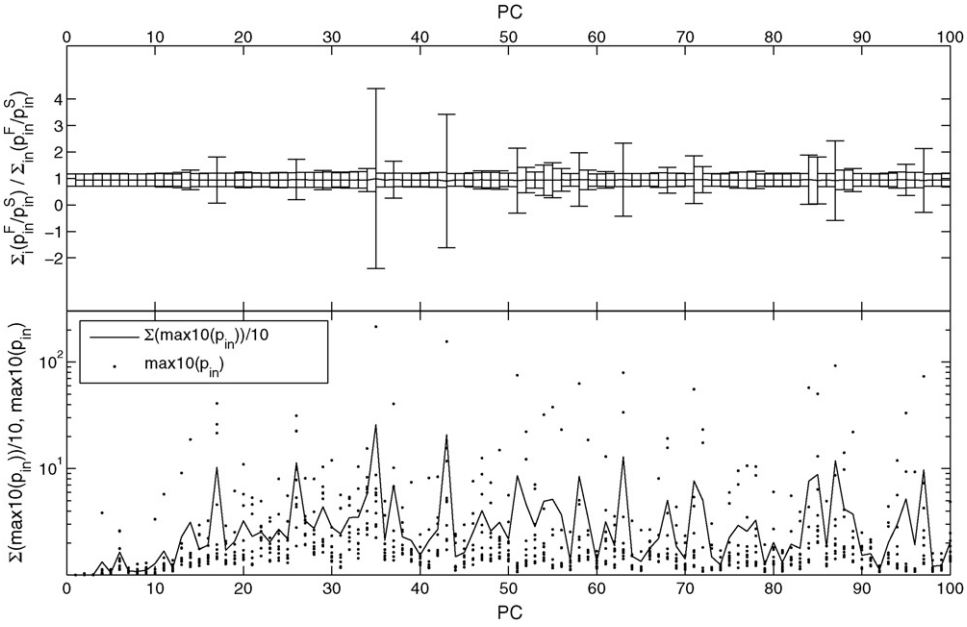


Fig. 5. The average $q_{in}$ values with their corresponding errors (top), and the 10 maximum values for $q_{in}$ as well as their average (bottom).

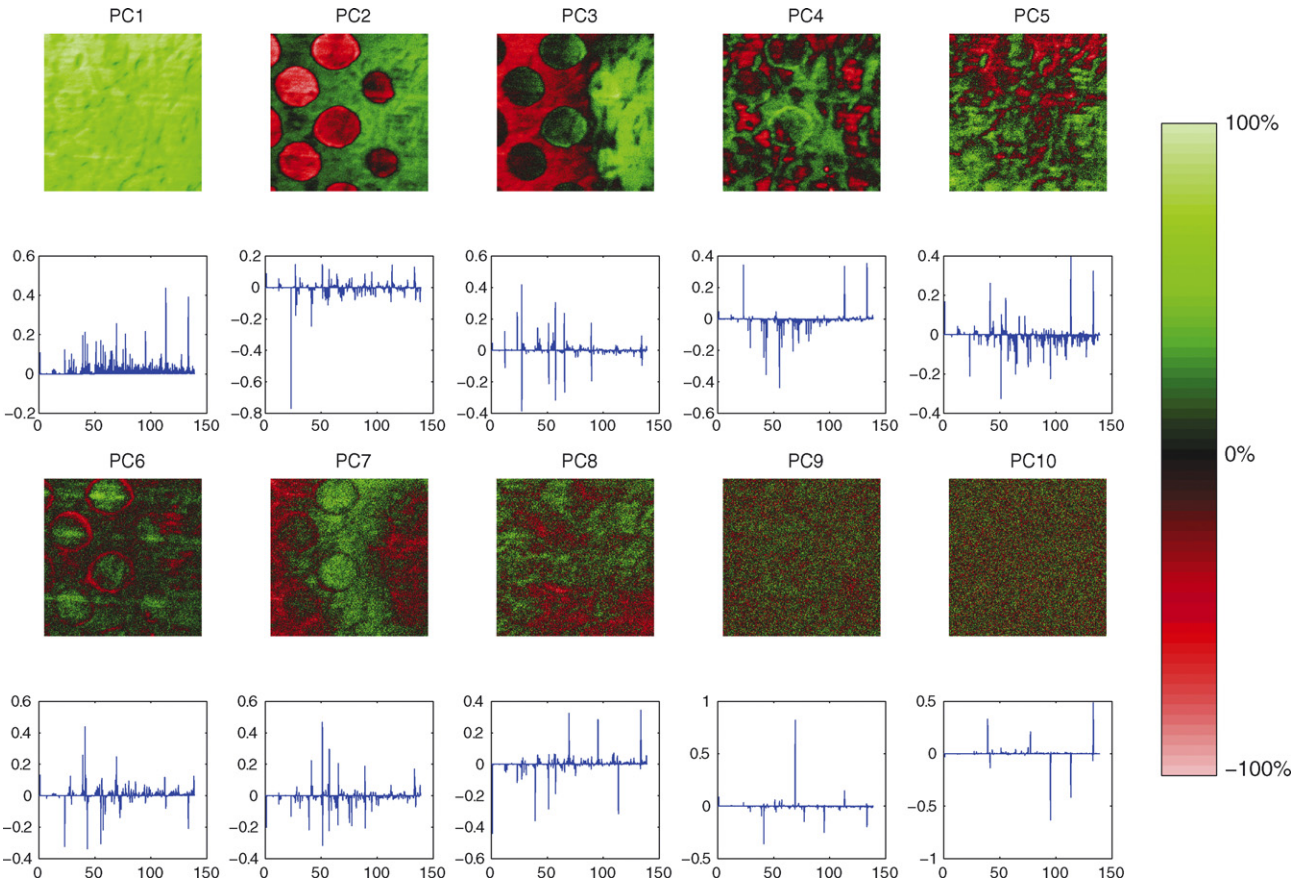


Fig. 6. The first 10 principal components as obtained with sparse-type PCA on a droplet array spotted with PVP on a PVDF membrane. The pseudo-colors represent positive (green) and negative scores (red) on the principal components. For clarity reasons, only the low-mass part of the mass spectrum is shown. Each image is scaled to its maximum (absolute) intensity.

is concentrated into two PCs (1 and 7) that only contain this type of peak. The rough PC result showed these peaks in a mixture with other peaks (PCs 3 and 7).

VARIMAX done on the cross-sectioned hair did not improve the chemical contrast. Although image contrast improved, the specificity for the observed salt crystal decreased (Fig. 9). Chemical features were resolved in a comparable way as those found with PCA. The image contrast was improved and used to identify the sharp boundaries of various spatial features.

The LDI-ToF PCA results optimized with VARIMAX show a tremendous increase in contrast when compared with the PCA results without VARIMAX optimization (Fig. 8b). Although the two different paints are found in various PCs, VARIMAX proves to be a very powerful tool in the optimization of PCA-aided analysis of this microprobe LDI dataset, giving phtalocyanine blue (combined PCs 2, 4, 5 and 19) and phtalocyanine green (PC 7)-specific spectral profiles.

### 4.4. Computation time versus results

As mentioned in Section 1, computational power is one of the main aspects for multivariate data analysis. An estimate of computation time was made for the studied methods. All time-measurements were done on the same computer (single processor 32 bit AMD Athlon, 2.2 GHZ, 1 GB of memory), using MatLab™ 7.1 with the N-way toolbox 2.11 [33] and VARIMAX implementation. Calculations were done in a 32-bit environment. This limits memory allocation (and therefore the maximum size of the analyzed dataset) to 4 GB. The use of a 64-bit environment would circumvent this memory problem and therefore make the use of larger datasets possible. However this would also increase calculation time. The size of the quantitatively analyzed datasets was chosen such that the total calculation could be done without the need of virtual memory. Using virtual memory would dramatically increase the total calculation time because hard disk access is much slower than RAM access. This would not give a representative measure when the algorithms are compared.

Computation time was evaluated for all three datasets mentioned earlier (Table 1). Two different datacubes were used for the ToF-SIMS datasets: one with a large spectral dimension and one with a large spatial dimension (datacubes were unfolded into $x \times y$ by $c$). The number of components was varied from 7 to 14–21. The LDI-ToF imaging datacube was analyzed at full spatial resolution ($7 \times 290$) and with 1850 spectral variables.

The standard PCA method first calculates the full and exact PC decomposition and then restricts the resulting dataset to the requested number of components. PCA performed on sparse

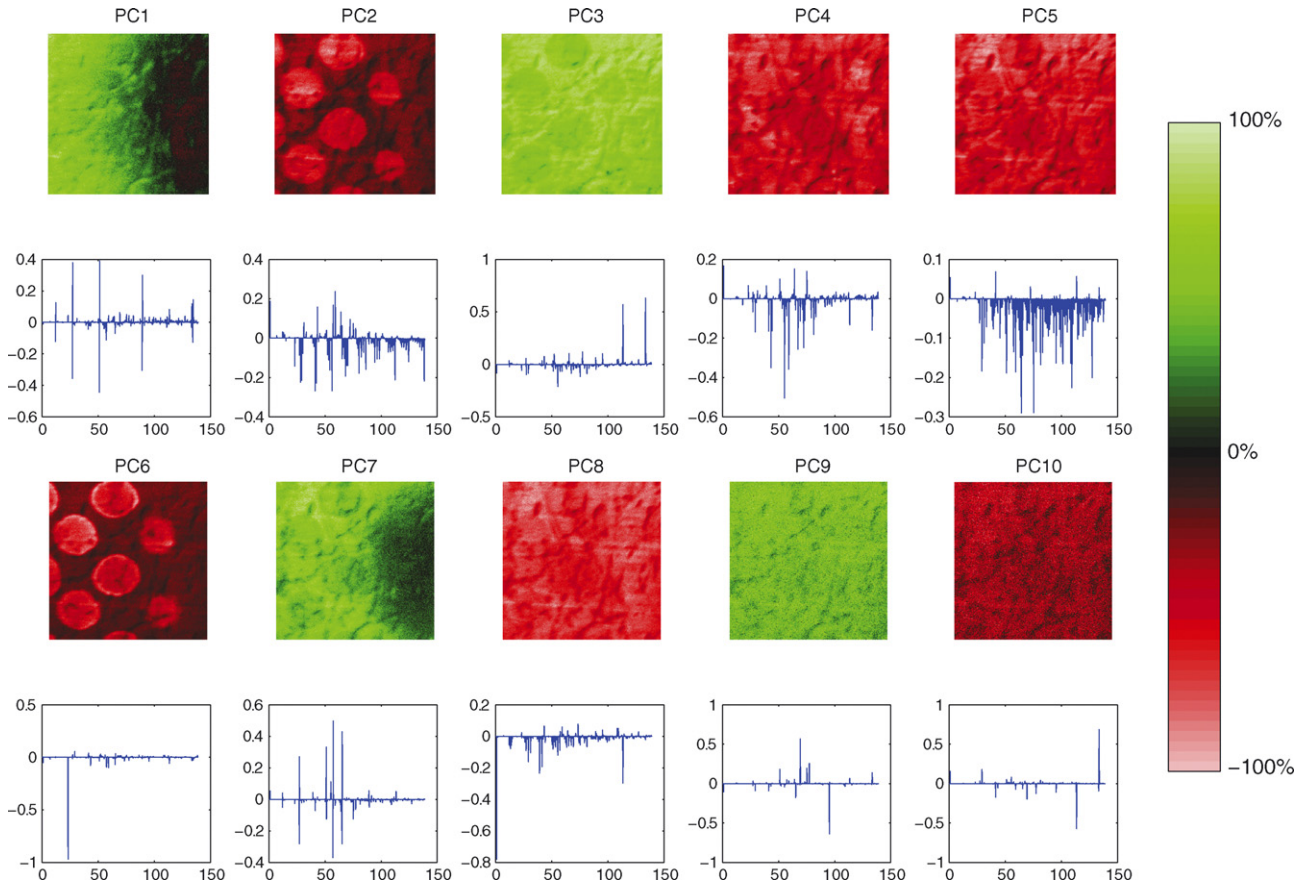


Fig. 7. The first 10 principal components as obtained with sparse-type PCA followed by VARIMAX rotational optimization on a droplet array spotted with PVP on a PVDF membrane. The pseudo-colors represent positive (green) and negative scores (red) on the principal components. For clarity reasons, only the low-mass part of the mass spectrum is shown. Each image is scaled to its maximum (absolute) intensity.

matrixes produces an approximation by itself, not giving a full representation of the original datacube, but only resulting in the requested number of PCs. This difference in methodology contributes to the time-reduction that is involved in the use of sparse matrixes. The continuous nature of the LDI data, with a non-zero entry at almost each sampling point resulted in an increased computation time when the sparse matrix format was used. This can be explained from the fact that the in-memory size is larger for the sparse-type matrix than for the full matrix, which inevitably leads to larger processing times.

VARIMAX as a post-processing optimization step after PCA results in only a small increase in calculation time. This justifies

Table 1
Table with an indication of computation time in s using various methods on various samples

| Set | Components | Dataset size | PCA | PCA (sparse) | PCA + VARIMAX | PARAFAC ($\times 10^3$) | PARAFAC (sparse) ($\times 10^3$) |
|---|---|---|---|---|---|---|---|
| Hair | 7 | $300 \times 256 \times 256$ | 3 | 3 | +0.15 | 3.5 | 2 |
| Droplet | 7 | $300 \times 256 \times 256$ | 3 | 3 | +0.15 | 12 | 6.5 |
| Hair | 14 | $300 \times 256 \times 256$ | 3 | 3 | +0.25 | 6 | 5 |
| Droplet | 14 | $300 \times 256 \times 256$ | 3 | 3 | +0.25 | 40 | 50 |
| Hair | 21 | $300 \times 256 \times 256$ | 3 | 5 | +0.35 | 14 | 13 |
| Droplet | 21 | $300 \times 256 \times 256$ | 3 | 4 | +0.35 | 160 | 85 |
| Hair | 7 | $5053 \times 64 \times 64$ | $5 \times 10^2$ | 25 | +0.2 | 0.9 | 1 |
| Droplet | 7 | $5053 \times 64 \times 64$ | $5 \times 10^2$ | 20 | +0.2 | 0.7 | 0.6 |
| Hair | 14 | $5053 \times 64 \times 64$ | $5 \times 10^2$ | 25 | +0.3 | 3.5 | 3 |
| Droplet | 14 | $5053 \times 64 \times 64$ | $5 \times 10^2$ | 20 | +0.3 | 9 | 8 |
| Hair | 21 | $5053 \times 64 \times 64$ | $5 \times 10^2$ | 30 | +0.4 | 6 | 4 |
| Droplet | 21 | $5053 \times 64 \times 64$ | $5 \times 10^2$ | 20 | +0.4 | 30 | 27 |
| LDI | 7 | $1850 \times 290 \times 7$ | 30 | 35 | +0.15 | 52 | 55 |
| LDI | 14 | $1850 \times 290 \times 7$ | 30 | 35 | +0.30 | | 214 |
| LDI | 21 | $1850 \times 290 \times 7$ | 30 | 35 | +0.40 | | |

The VARIMAX processing time is given as the time added to PCA.

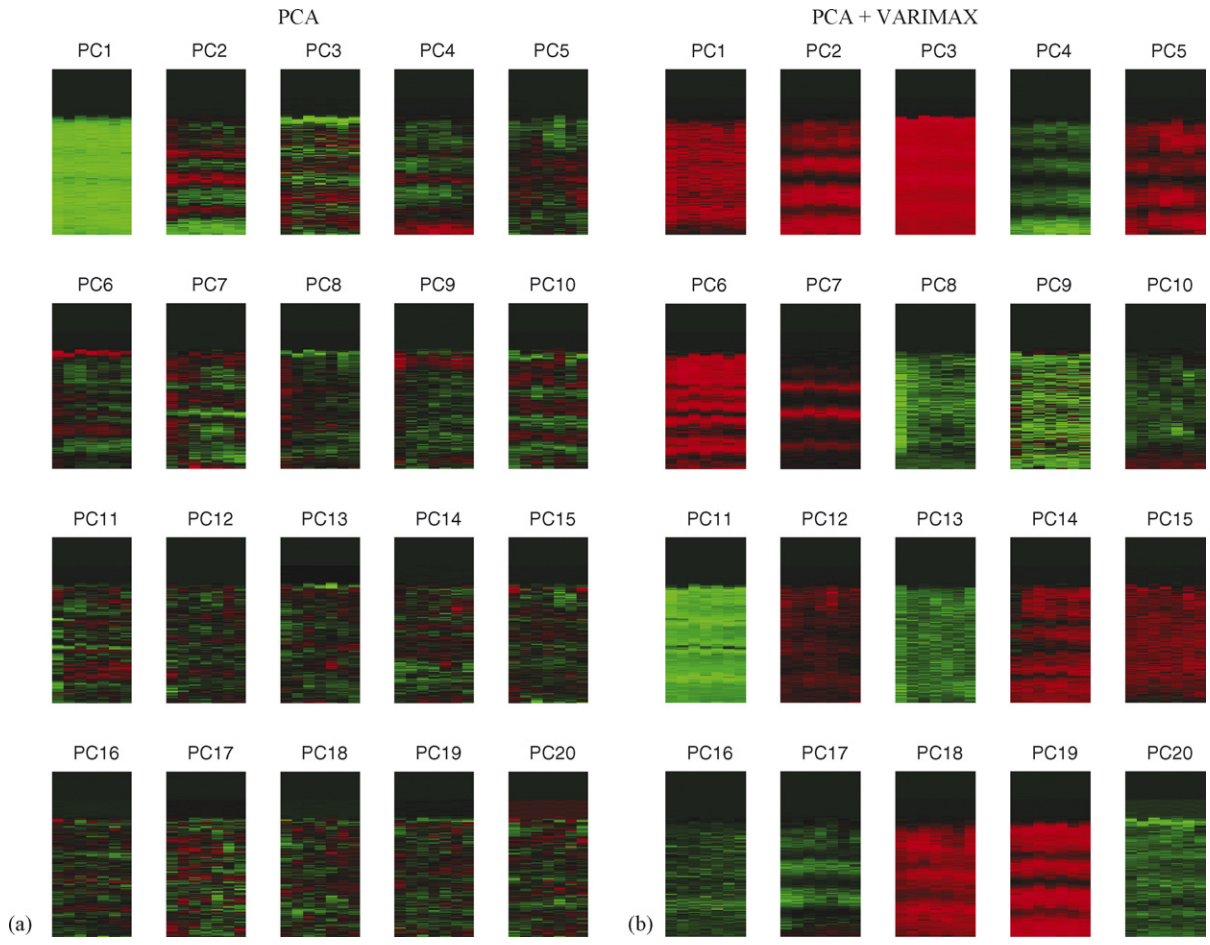


Fig. 8. Score images resulting from 20-component PCA (a) and PCA + VARIMAX (b) analysis of a layered structure of Liquitex paint. The size of the area represented by the images is approximately 1300 μm × 680 μm, the paint layers are overlapping and show up thicker than 100 μm due to the laser spot size of 200 μm.

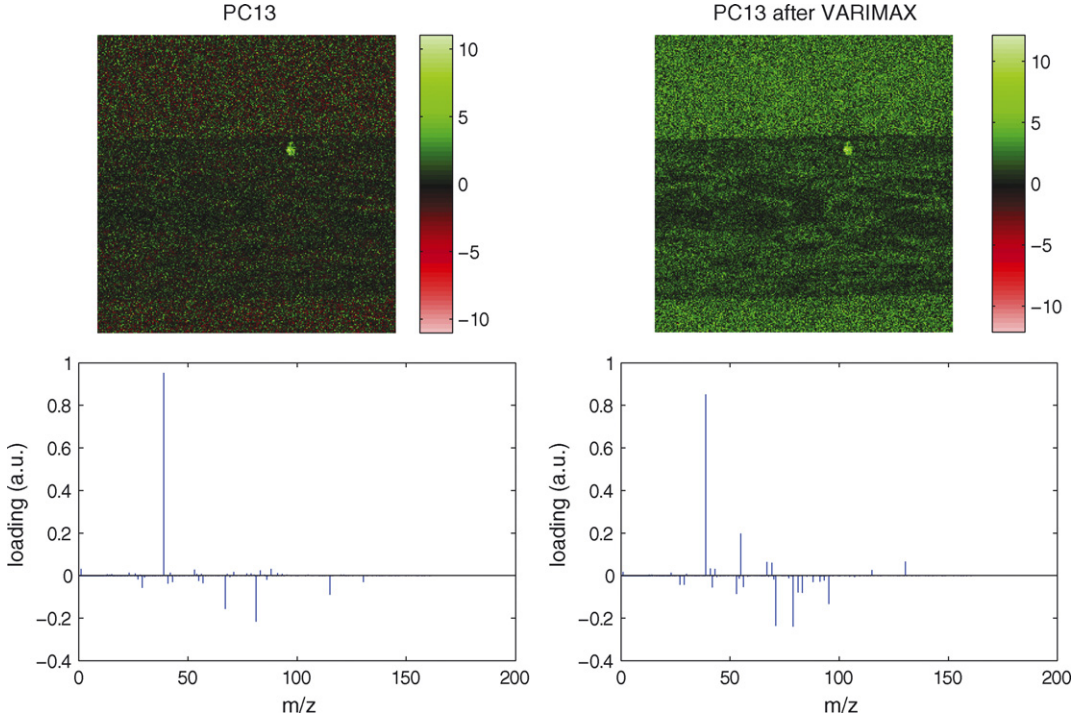


Fig. 9. Score images (top) and loading vectors (bottom) after analysis of the cross-sectioned hair with PCA (left) and PCA + VARIMAX (right). Each image is scaled to its maximum (absolute) intensity.

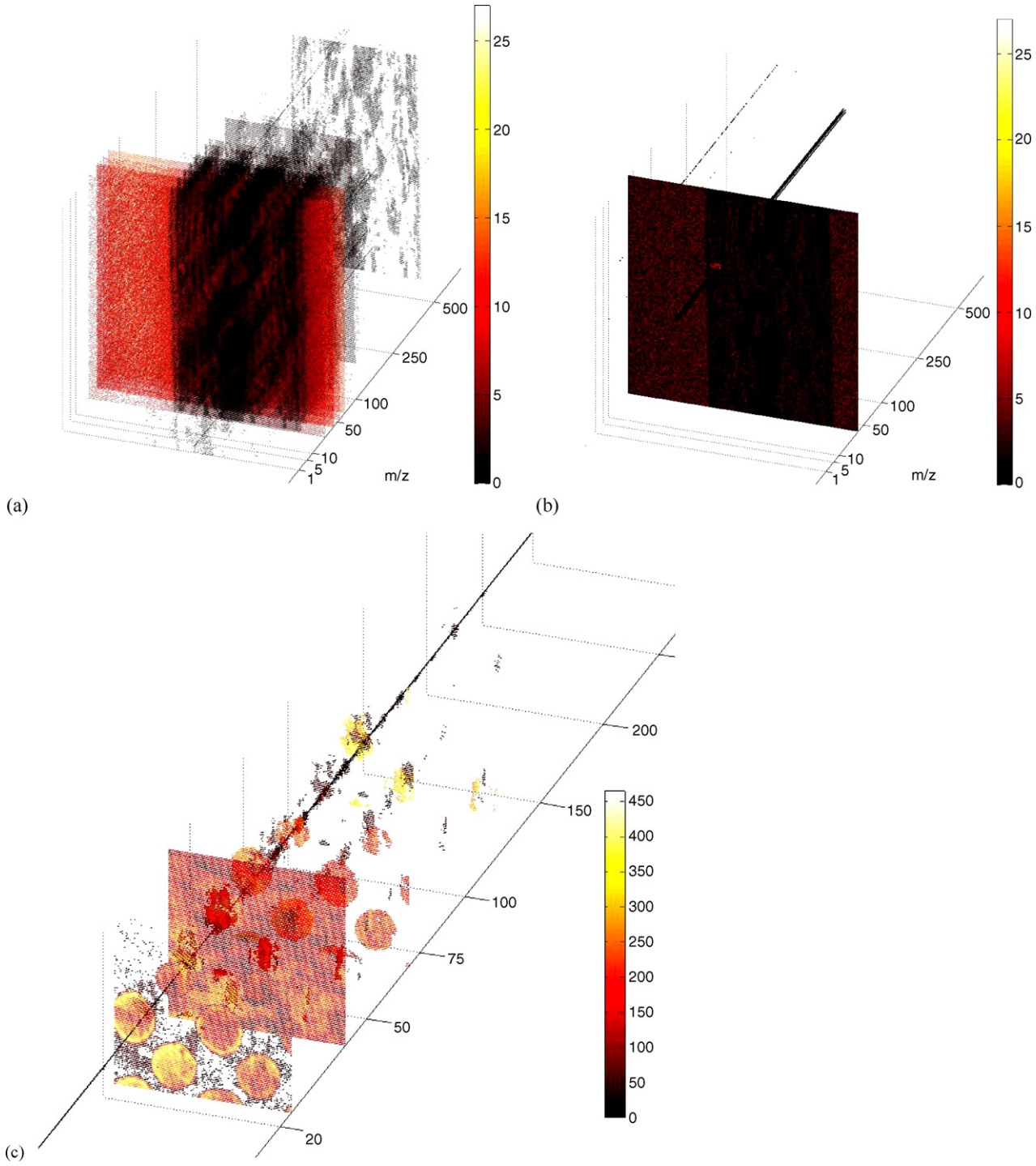


Fig. 10. An alternative representation of the complete hyperspectral datacube of the embedded hair with an overlay using PC1 (a) and PC13 (b) and the PVP droplet series with the 6th PARAFAC factor (c).

the use of VARIMAX after PCA in any case to increase chemical contrast in both PC images and spectra, as shown in previous sections.

PARAFAC is clearly a much more demanding technique. Although it turned out to be better at resolving certain features, it is not suitable for routine use with the current standings of on-desk computer facilities. It could be very helpful in very complex systems or in systems where trace amounts of a certain chemical components are expected. Prior knowledge, which is favorable to make a sensible choice for the number of components to be looked for, could be obtained using PCA. Like PCA, PARAFAC turned out to be faster on sparse matrixes. It should be mentioned that the random initialization as used in our PARAFAC calculations, results in a large variation in calculation time and the order of the factors. PARAFAC is a computationally much more demanding technique because it seeks an exact fit of the data using optional constraints, spread over the defined number of factors.

### 4.5. Three-dimensional visualizations

The components from the different methods for multivariate analysis yield a specific extracted spectral profile with a corresponding spatial view. It is hard to give an interpretation using only the individual spectral or spatial component. PCA was able to extract the location of a feature in the hair, but the spectral view revealed that it was caused by a salt-crystal. A combined view would directly reveal the connection between both views. Each pair of extracted scores and loadings can be combined in one three-dimensional overview to gain more insight in the correlations between spectral profile and the location. Each value in the cube is the intensity on a certain position in a spectral plane and is given a color using the '*hot*'-color map from MatLab™. Because most values are zero within a MS dataset, the complete cube would result in an image of a black box. Large parts of this box can be discarded as they do not contain any interesting properties. An opacity map is introduced to hide uninteresting features within the datacube which, in this case, is created with the extracted components from the multivariate analysis. Instead of a continuous switch between spectral and spatial view, a complete view of the cube can directly reveal this connection. A user is able to interactively rotate the cube and instantly get an overview of all the data in three dimensions.

The complete hyperspectral datacube of the hair is shown in Fig. 10a and b. Only the high values in the spectral profile and image component of PC1 are made opaque by the opacity map. In this way PC1 is highlighted in the original datacube which contains mostly the areas and peaks from the hair itself. The component with the extracted features from the crystal is shown in Fig. 10b. It clearly shows the relation between the highlighted image plane on $m/z$ 39 and the small group of pixels on the location of the crystal, while other areas of the cube remain hidden. The significant peak on 39 $m/z$ in the spectral component highlights the complete image plane at this spectral position. Similarly, the high intensity of the pixels in the spatial component results in the appearance of a 'rod', spanning the whole spectral dimension of the cube. The number of points of this feature that are shown can be adjusted by changing the threshold in the opacity map.

This representation provides better overall insight in the data by visualizing the direct correlation between spectral peaks and spatial occurrences. Fig. 10c shows several isolated drops in the spectral datacube using the sixth PARAFAC factor. The different components or factors can be highlighted together or separately in the same cube by combining their opacity maps. The resulting three-dimensional view becomes more accurate and discriminating when the resulting components from the multivariate analysis contain more contrast. This advantage makes it easier to compare the quality of results from the different multivariate analyses.

## 5. Conclusions

We made a comparison between various multivariate statistical methods for the analysis of hyperspectral datasets as acquired with ToF-SIMS and LDI-ToF imaging mass spectrometry. Obviously, the same methods used for LDI-imaging can be used in MALDI-imaging experiments.

The use of the sparse matrix format allows larger datasets to be handled and drastically decreases computation time. Memory problems are circumvented because zero values are disregarded which is a more efficient way of data storage when most values in the datacube are zero. The sparse matrix format makes the analysis of larger datasets possible and allows them to be analyzed at higher resolution. No significant difference was found between the resulting extracted information of the different implementations for normal and sparse matrixes in specific multivariate analytical techniques.

Of the methods compared in this report, PCA turns out to offer the best trade-off between results and computation time. Although PARAFAC gave a better overall performance, the high amount of computational power needed, restricts this technique to the use in specific cases. A sensible choice of the number of components to be calculated is needed in PARAFAC, as an excess number of components dramatically increases computation time. To make an estimation of the number of components to be calculated in PARAFAC, pre-knowledge is needed. This makes the technique less suitable for routine analysis. The application of VARIMAX rotation as a post-processing technique increases both chemical and imaging contrast when used after PCA. The almost negligible amount of computation time needed for this, suggest that is should be used in any case when PCA is used. However, the original results from the PCA should still be considered in some cases, especially for small features, the chemical specificity may decrease when VARIMAX is used. In most cases however, a pseudo-color plot together with manual analysis of the spectra is sufficient to resolve the different chemical components. A three-dimensional presentation of the complete datacube or selected components, was shown to be a useful tool for quick insight into a hyperspectral datacube. Although a scientific expert is still needed to analyze the resulting components, these multivariate statistical methods are an indispensable tool in the analysis of complex imaging MS datasets.

## Acknowledgements

This work is part of the research program of the "Stichting voor Fundamenteel Onderzoek der Materie (FOM)", which is financially supported by the "Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO)". Part of this work was carried out in the context of the Virtual Laboratory for e-Science project (www.vl-e.nl). This project is supported by a BSIK grant from the Dutch Ministry of Education, Culture and Science (OC&W) and is part of the ICT innovation program of the Ministry of Economic Affairs (EZ). We acknowledge funding from the Strategic Research Fund (SRF) of ICI plc.

## References

[1] K.R. Thorp, L.F. Tian, Precis. Agric. 5 (2004) 477.
[2] J.M. Piwowar, E.F. LeDrew, Prog. Phys. Geogr. 19 (1995) 216.
[3] B.H.P. Maathuis, J.L. van Genderen, Int. J. Remote Sens. 25 (2004) 5201.

[4] I.W. Levin, R. Bhargava, Annu. Rev. Phys. Chem. 56 (2005) 429.
[5] R. Bhargava, S. Wang, J.L. Keonig, Adv. Polym. Sci. 163 (2003) 137–191.
[6] N.J. Everall, J.M. Chalmers, L.H. Kidder, E.N. Lewis, M. Schaeberle, I. Levin, Abstr. Pap. Am. Chem. Soc. 219 (2000) U502.
[7] N.J. Everall, Jct Coat. 2 (2005) 38.
[8] P.J. Todd, T.G. Schaaff, P. Chaurand, R.M. Caprioli, J. Mass Spectrom. 36 (2001) 355.
[9] A.M. Belu, D.J. Graham, D.G. Castner, Biomaterials 24 (2003) 3635.
[10] L.A. McDonnell, S.R. Piersma, A.F.M. Altelaar, T.H. Mize, S.L. Luxembourg, P.D.E.M. Verhaert, J. van Minnen, R.M.A. Heeren, J. Mass Spectrom. 40 (2005) 160.
[11] R.R. Meglen, Mar. Chem. 39 (1992) 217.
[12] M.S. Wagner, D.J. Graham, B.D. Ratner, D.G. Castner, Surf. Sci. 570 (2004) 78.
[13] V.S. Smentkowski, J.A. Ohlhausen, P.G. Kotula, M.R. Keenan, Appl. Surf. Sci. 231/232 (2004) 245.
[14] B. Tyler, Appl. Surf. Sci. 203 (2003) 825.
[15] M.C. Biesinger, P.Y. Paepegaey, N.S. McIntyre, R.R. Harbottle, N.O. Petersent, Anal. Chem. 74 (2002) 5711.
[16] G. McCombie, D. Staab, M. Stoeckli, R. Knochenmuss, Anal. Chem. 77 (2005) 6118.
[17] L.A. McDonnell, T.H. Mize, S.L. Luxembourg, S. Koster, G.B. Eijkel, E. Verpoorte, N.F. de Rooij, R.M.A. Heeren, Anal. Chem. 75 (2003) 4373.
[18] A. Delcorte, N. Medard, P. Bertrand, Anal. Chem. 74 (2002) 4955.
[19] A. Delcorte, P. Bertrand, Appl. Surf. Sci. 231/232 (2004) 250.
[20] J.C.B. Vickerman, David, ToF-SIMS: Surface Analysis by Mass Spectrometry, IM Publications and SurfaceSpectra Ltd., 2001.
[21] S.L. Luxembourg, T.H. Mize, L.A. McDonnell, R.M.A. Heeren, Anal. Chem. 76 (2004) 5339.
[22] B.T. Wickes, Y. Kim, D.G. Castner, Surf. Interf. Anal. 35 (2003) 640.
[23] M.R. Keenan, P.G. Kotula, Appl. Surf. Sci. 231/232 (2004) 240.
[24] M.R. Keenan, P.G. Kotula, Surf. Interf. Anal. 36 (2004) 203.
[25] M.R. Keenan, J. Vacuum Sci. Technol. A 23 (2005) 746.
[26] D.J. Graham, B.D. Ratner, Langmuir 18 (2002) 5861.
[27] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, 2002.
[28] R.B. Lehoucq, D.C. Sorensen, C. Yang, ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods, SIAM Publications, Philadelphia, 1998.
[29] D.C. Sorensen, Siam. J. Matrix Anal. A 13 (1992) 357.
[30] H.F. Kaiser, Proc. Psychomet. (1958) 187.
[31] R.A. Harshman, Proc. UCLA Work. Pap. Phonet. (1970) 1.
[32] J.D. Carroll, J.J. Chang, Proc. Psychomet. (1970) 283.
[33] R. Bro, Proc. Chemom. Intell. Lab. Syst. (1997) 141.
[34] A. Broersen, R. van Liere, Proceedings of the Eurographics/IEEE VGTC Symposium on Visualization, June, 2005, p. 117.
[35] J.G. Newman, B.A. Carlson, R.S. Michael, J.F. Moulder, T.A. Hohlt, Static SIMS Handbook of Polymer Analysis, Perkin-Elmer Corporation, Physical Electronics division, Eden Prairie, MN, 1991.