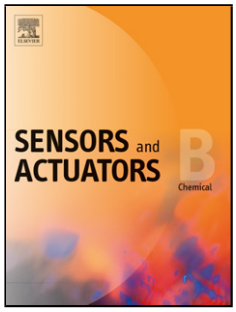


Title: Sliding Window Multi-Curve Resolution: Application to Gas Chromatography - Ion Mobility Spectrometry

Author: S. Oller-Moreno G. Singla-Buxarrais J.M. Jiménez-Soto A. Pardo R. Garrido-Delgado L. Arce S. Marco

PII: S0925-4005(15)00293-2
DOI: http://dx.doi.org/doi:10.1016/j.snb.2015.02.108
Reference: SNB 18172

To appear in: *Sensors and Actuators B*

Received date: 19-7-2014
Revised date: 20-2-2015
Accepted date: 24-2-2015

Please cite this article as: S. Oller-Moreno, G. Singla-Buxarrais, J.M. Jiménez-Soto, A. Pardo, R. Garrido-Delgado, L. Arce, S. Marco, Sliding Window Multi-Curve Resolution: Application to Gas Chromatography - Ion Mobility Spectrometry, *Sensors and Actuators B: Chemical* (2015), http://dx.doi.org/10.1016/j.snb.2015.02.108

Sliding Window Multi-Curve Resolution: Application to Gas Chromatography - Ion Mobility Spectrometry

S. Oller-Moreno[*1,2], G. Singla-Buxarrais[*1], J.M. Jiménez-Soto[1], A. Pardo[2], R. Garrido-Delgado[3], L. Arce[3], S. Marco[1,2]

[1] Signal and Information Processing for Sensing Systems, Institute for Bioengineering of Catalonia, Baldiri i Reixac 4-8, 08028-Barcelona, Spain

[2] Intelligent Signal Processing, Department of Electronics, University of Barcelona, Martí i Franquès 1, 08028-Barcelona, Spain

[2] Department of Analytical Chemistry, University of Córdoba, Campus Rabanales, 14071-Córdoba, Spain.

[*] These authors contributed equally to this work

Abstract

Blind Source Separation (BSS) techniques aim to extract a set of source signals from a measured mixture in an unsupervised manner. In the chemical instrumentation domain source signals typically refer to time-varying analyte concentrations, while the measured mixture is the set of observed spectra. Several techniques exist to perform BSS on Ion Mobility Spectrometry, being Simple-to-use interactive self-modeling mixture analysis (SIMPLISMA) and Multivariate Curve Resolution (MCR) the most commonly used. The addition of a multi-capillary gas chromatography column using the ion mobility spectrometer as detector has been proposed in the past to increase chemical resolution. Short chromatography times lead to high levels of co-elution, and ion mobility spectra are key to resolve them. For the first time, BSS techniques are used to deconvolve samples of the gas chromatography - ion mobility spectrometry tandem. We propose a method to extract spectra and concentration profiles based on the application of MCR in a sliding window. Our results provide clear concentration profiles and pure spectra, resolving peaks that were not detected by the conventional use of MCR. The proposed technique could also be applied to other hyphenated instruments with similar strong co-elutions.

Keywords: Blind Source Separation; Multivariate Curve Resolution; Ion Mobility Spectrometry; Gas Chromatography; Hyphenated instrumentation; SIMPLISMA; co-elution

1. Introduction

Ion Mobility Spectrometry (IMS) is an analytical technique for characterizing chemical substances based on the velocity of gas-phase ions in an electric field [1]. Ion mobility spectrometers are able to detect trace levels of volatile chemicals with high-speed analysis and moderate selectivity. They are also portable enough to be commercialized as handhold instruments, being a recent research interest its miniaturization [2], [3]. IMS was first used in the 1970s for explosives and chemical warfare agents detection in military applications [4]. Over the past twenty years, IMS fields of application have widened and it is currently being used in environmental [5], industrial [6], and biomedical studies [7], drug detection [8], security applications [9], food quality [10] and fraud detection [11].

To improve IMS chemical resolution, it has been proposed to couple a multi-capillary chromatographic column (MCC) to pre-separate the chemical compounds of the sample [12], enabling the analysis of more complex samples. However, as MCCs are characterized by presenting short chromatography times, multiple compounds may still elute from the column at the same time. This effect is known as co-elution. Ion mobility spectra are therefore key to resolve these co-eluted chemicals.

Blind Source Separation (BSS) techniques, also named in chemometrics “resolution techniques”, are commonly applied to hyphenated analytical techniques that provide second order data. In IMS samples, the compounds' original concentration profiles and pure spectra can be deconvolved from the sample using BSS techniques [13]. For the first time, we propose a blind source separation technique in MCC-IMS data. Direct application of MCR techniques to full MCC-IMS data typically fails to resolve co-elution due to the complexity of the data and to the global noise which hinders the detection of weak but significant peaks. The typical approach in this case is the manual selection of the retention time window where the co-elution appears and the application of MCR in this data subset. However, in MCC-IMS chromatography conditions, co-elution is a prevalent phenomenon [14], [15]. Few individual peaks can be isolated in the total chromatogram and mostly very broad peaks are observed. To deal with this complexity, we propose an automatic manner to investigate co-elution across the whole chromatographic axis. The proposed method is able to detect and recover compounds in adverse co-elution conditions and reject spurious spectra with no physical meaning in an unsupervised manner.

The method is applied to real data corresponding to olive oil headspace analysis, with the aim to extract accurate concentration profiles and pure spectra for each sample. The extracted information can be used in a posterior step to discriminate among different regulated olive oil qualities in fraud prevention applications.

1.1 Ion Mobility Spectrometry

An IMS instrument consists of two main parts: an ionization chamber and a drift tube. Gas molecules enter the chamber and are ionized by a source (for instance radioactive, UV-lamp, corona discharge, electro-spray, among others) [16], [17]. The ionization processes take part at atmospheric pressure, being referred to as Atmospheric Pressure Chemical Ionization (APCI) [18]. Ionized molecules are prevented from entering the drift tube using an electrostatic grid, acting as an ion gate [19]. When the gate opens, ions are accelerated by an electric field through the drift tube, colliding with gas molecules and reaching a constant limit velocity. A collector at the end of the drift tube takes the charge from the ions producing a current and leaving neutralized molecules. The current provided by the collector can be interpreted as an ion mobility spectrum that lasts a few milliseconds. Components appear as peaks in the spectrum at different drift times, depending on their mass, shape and size [1].

The sample molecules are indirectly ionized by reactant ions if a radioactive source is used [1]. These ions are formed in the APCI process based on water chemistry reactions. The radioactive source ionizes nitrogen molecules that collide with water and other nitrogen molecules forming hydrated protons (H_3O^+). Depending on the atmosphere conditions, hydrated ammonium (NH_4^+) cations and hydrated superoxide (O_2^-) ions can also appear. The analytes in the sample are then ionized through collisions with the reactant ions. When multiple analytes are present in the ionization region simultaneously, they compete for the charge of the reactant ions. The analytes with largest proton affinity (or electronegativity, depending on the polarity detection mode) acquire most of the available charge and, if their concentration is large enough, they can prevent the ionization and posterior detection of the other analytes.

The remaining reactant ions form a peak -or several peaks- when they reach the detector. These peaks can be used as a reference and are usually known as the Reactant Ion Peaks (RIP).

Multi-Capillary Columns can be used to improve IMS's selectivity, at the expense of some of the IMS's portability and analysis speed [20]. A MCC is similar to a conventional chromatographic column, although it consists of a bundle of parallel capillaries that allow a higher flow rate of the carrier gas. The hyphenated instrument MCC-IMS is similar to conventional Gas Chromatography – Mass Spectrometry (GC-MS), as chromatography is used to separate the compounds, and spectrometry is used to identify them. However, GC-MS instruments offer a much higher resolution and powerful libraries for compound identification than MCC-IMS, but the later offers a portable solution, with lower costs and less maintenance requirements. The pre-separation step provided by the MCC (a) helps discriminating analytes with similar IMS drift times if they present different

elution times and (b) reduces the number of analytes simultaneously present in the ionization region, reducing the charge competition effects described previously.

By continuously analyzing the output of the MCC with the IMS, the analysis of each sample produces a bidimensional matrix formed by consecutive measured spectra. The matrix dimensions are the MCC retention time and the IMS drift time.

1.2 Blind Source Separation techniques

Blind source/signal separation techniques are the collection of algorithms designed to estimate a set of source signals from measured mixtures. As mentioned in [21], techniques are blind because a) the source signals are not observed directly, b) the mixing matrix is unknown and c) no information is available about the composition of the mixture, not even the number of source signals present. These techniques are commonly used in signal processing [22] and are increasingly being used in chemical instrumentation applications [23], such as the analysis of nuclear magnetic resonance data [24], chemical reaction monitoring [25] and Raman spectroscopy [26]. BSS has been recently used to enhance information extraction from temperature-modulated metal oxide gas sensors [27] and to separate interferences from ion activity in ion-sensitive field-effect transistors [28].

As deconvolution problems are under-determined by definition, constraints are required to narrow the space of solutions. For many applications, Independent Component Analysis (ICA) [29] is an appropriate and successful technique if mixing models can be assumed to be linear and source signals to be statistically independent. However, in chemical analysis and specifically in IMS, statistical independence of compounds is not necessarily fulfilled [13]. Therefore other approaches are used to constrain the range of possible solutions [23] being Non-negative Matrix Factorization (NMF) techniques [30] and in particular Multivariate Curve Resolution (MCR) methods [31] common alternatives.

In MCC-IMS applications, BSS techniques are helpful when several analytes elute at the same time from the MCC and are detected by the IMS. If some of the co-eluting analytes present larger proton affinities (or electronegativities) than others, they can mask and hide them due to the charge competition effect described on Section 1.1. In this case, no posterior data analysis techniques (BSS or others that we know of) will be able to detect them.

1.3 Multivariate Curve Resolution Alternating Least Squares

Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) [32] assumes a linear decomposition of the mixing matrix, which can be written as shown in Eq. 1.

$$D = CS^T + E \quad (1)$$

D (*MxN*) is the measured mixing matrix, with *M* spectra of length *N*. *C* (*MxK*) is the abundances or concentrations matrix, that contains the proportions of each unmixed spectrum in the measured matrix and *S* (*NxK*) is the pure (or unmixed) spectra matrix that contains the *K* pure spectra of length *N*. *E* is a matrix of residuals (*MxN*).

Given an initial estimation of K pure spectra, MCR-ALS proceeds as follows:

1. Filter noise from the mixing matrix: First, compute PCA scores and loadings from the D mixing matrix. Then reconstruct a filtered version of D, named D*, using the first K principal components of the computed scores and loadings.
2. Estimate the concentration profiles using least squares:

$$C = \text{argmin}_C \, || D^* - CS^T ||^2 \quad (2)$$

3. Impose constraints on the concentration profiles
4. Estimate the pure spectra using least squares:

$$S = \text{argmin}_S \, || D^* - CS^T ||^2 \quad (3)$$

5. Impose constraints on the pure spectra
6. Iterate steps 2-5 until convergence.

The key to obtaining reliable concentration profiles and pure spectra depends on the estimation of the number of components in the mixture, the initialization of the pure spectra and the imposition of constraints.

The number of components for each window can be estimated with several methods such as [33]–[35]. However, a simpler approach described in [36] is commonly used: the number of components is determined as the number of singular values of the matrix above a given threshold, representative of the noise in the sample.

There are multiple ways of obtaining an initial estimation of the pure spectra, being SIMPLe-to-use Interactive Self-modeling Mixture Analysis (SIMPLISMA) [37] and Evolving Factor Analysis (EFA) [38] the most common ones. Although EFA works well on samples presenting unimodal concentration profiles, on IMS this condition does not necessarily hold, making SIMPLISMA the most common alternative in this field [13], [39].

Many constraints can be imposed on the concentration profiles and pure spectra depending on the prior knowledge of our particular problem: On IMS spectra, non-negativity can be imposed on both concentration profiles and spectra. Moreover, as the ionization process consists in a charge transfer from the RIP to the compounds, charge conservation can be imposed on the concentration profiles (closure on *C*). Unimodality constraints are not suitable for concentration profiles, but can be imposed to the resolved spectra shapes. Finally, selectivity constraints to the concentration profiles can also be imposed if some components are known to appear at a particular retention time range.

The MCR-ALS algorithm is based on a least squares minimization of the global error of the factorization. As it is shown in our results, local peaks with low intensities appearing in regions with strong co-elution may pass unnoticed by MCR-ALS, as they present a contribution to the error comparable to or smaller than the global noise of the sample. In these cases, increasing the estimated number of components in the mixture leads to extracting spurious compounds with no physical meaning instead of the desired local compounds.

1.4 Proposed technique: Sliding Window Multivariate-Curve Resolution

In order to overcome the limitation in resolving low intensity peaks when the conventional MCR-ALS is applied to the whole MCC-IMS data matrix, we propose to apply MCR in short partially overlapped windows, slicing the matrix in the retention time axis. In addition, window overlap is imposed to avoid splitting peaks on window borders and to avoid detecting spurious compounds inconsistent across windows.

First, the initial estimations of pure spectra and concentration profiles are obtained by applying SIMPLISMA to each window. By using SIMPLISMA in this fashion, we can extract local peaks with low intensities, as they have comparatively higher peak purity within a single window. The number of components for each window is estimated using the threshold on singular values previously described in Section 1.3. To select the threshold, the singular values were plotted in decreasing order (plot not shown), presenting the typical elbow-like shape. The threshold was selected when the singular values begin to stabilize. Given the initial estimations, we use MCR-ALS to extract a set of concentration profiles and pure spectra for each window.

Finally, the results from all the windows are merged into a single set of concentration profiles and spectra representative of the whole sample. To do so, compounds are tracked through consecutive windows based on the similarity of their spectra. The angle between two pure spectra s_i and s_j is computed as shown in Eq. 4.

$$\theta_{i,j} = \arccos\left(\frac{s_i \cdot s_j}{|s_i||s_j|}\right) \quad (4)$$

Fig. 1 shows a diagram with an example of four compounds being tracked along three windows. The link between two spectra of consecutive windows is formed only if their angle is below a given threshold. In this figure, compounds C1 and C2 are being tracked along all the windows N to N+2 while compound C3 disappears on window N+1 because no link can be established on window N+2. Compound C4 does not appear until the N+1 window. The last spectrum in window N+1 does

not establish any link, thus it is considered spurious and is rejected from the final set of tracked compounds.

Windows are highly overlapped to guarantee that if no link can be established for a spectrum, then it can be safely considered as spurious and rejected from the final set. The final estimation of the pure spectra for each compound is computed as the mean of all the tracked spectra. The standard deviation of the mean is used as its error estimation. Averaging and computing the standard deviation are used likewise to obtain the final estimation of the concentration profiles.

2. Materials and Methods

2.1 Description of the samples

The proposed technique was applied to the olive oil dataset described in [10]. Current regulations in the European Union classify olive oils in three different categories according to their quality, namely Extra Virgin Olive Oil (EVOO), Virgin Olive Oil (VOO) and Lampante Olive Oil (LOO), being EVOO the category of highest quality and LOO the lowest one. This classification is based on several chemical parameters (free acidity, peroxide value and Ultra-violet absorbance) and a sensorial analysis. A proper control of olive oil qualities is crucial, not only because of the difference in price but also because LOO is not suitable for human consumption without being previously refined.

Ninety-eight olive oil samples from different qualities (43 samples of EVOO, 28 samples of VOO and 27 samples of LOO) were obtained from the Agrarian Laboratory of Junta de Andalucía and an oil press from Córdoba (Spain) during the 2009-2010 and 2010-2011 harvests. In order to keep the organoleptic features of the samples, they were stored at 4ºC until their analysis.

2.2 Analytical methods

Samples were analyzed with a MCC-IMS instrument (FlavourSpec®) from Gesellschaft für analytische Sensorysteme mbH (G.A.S), Dortmund (Germany). The olive oil headspace was directly sampled with a heated splitless injector, and the instrument was coupled to an automatic sampler unit (CTC-PAL, CTC Analytics AG, Zwingen, Switzerland) to improve reproducibility.

One gram of sample was placed in a 20-mL vial that was closed with magnetic caps. Samples were incubated at 60ºC for 10 minutes and 100 µL of sample headspace was automatically injected into the injector (80ºC) of the MCC-IMS.

The carrier gas going through the injector inserted the sample into the chromatograph, previously heated to 30ºC for pre-separation on a non-polar OV-5 MCC (20 cm long, ~1000 parallel glass

capillaries, filled with 5% diphenyl and 95% dimethylpolysiloxane). The analytes were eluted in an isothermal mode and driven into the IMS.

Inside the IMS, the ionization was produced with a Tritium source (6.5 keV). Ions entered the 6 cm long drift tube operating at a constant electric field of 350 V/cm and at a temperature of 60ºC. Spectra were acquired in the positive ion mode, generating each spectrum with the average of 32 scans, using a grid pulse width of 100 μs. The IMS sampled at 150 kHz and each scan lasted 20 ms. Each spectrum is 3000 points long.

Each sample was analyzed for 15 minutes, obtaining a complete IMS spectrum every 0.7 seconds. Compounds only eluted during the first 4 minutes of the retention time, leading to 340 spectra with information per sample. Each sample can be represented by a 340x3000 matrix.

2.3 Pre-processing

Noise present in each spectrum of the sample was filtered using a second order Savitzky-Golay filter [40] with a window size of 13 data points. The window size was selected assessing that the RIP height distortion caused by the filter was smaller than 1% of its non-filtered maximum value.

Next, a baseline was estimated and subtracted from the spectrum: the estimation of the baseline was computed by fitting a 4^{th} order polynomial to two non-peaked (empty) regions in the spectrum found in the regions 1-5 ms and 14.7-18.7 ms.

Finally, only the drift time region from 4 ms to 14.65 ms (1600 sampled points) contained information, so irrelevant regions were cropped out. Each sample was therefore reduced to a 340x1600 matrix.

2.4 Sliding Window MCR

The proposed Sliding Window MCR (SW-MCR) technique is applied to the sample, using a window length of ten spectra (7 seconds) and a window shift of a single spectrum (0.7 seconds). The window length was selected based on the typical width of a peak in the chromatogram, computed as the median full width at half maximum (FWHM) of ten representative peaks in the sample.

Larger window sizes and smaller window overlaps may be used to reduce the computational cost of the method. A larger window size would imply that more compounds can be found in the same window. If the window is too large we will face the same problem than with conventional MCR-ALS application: we may fail to detect local peaks with low intensities. A larger window overlap would increase our chances of splitting peaks in window borders and would hinder our ability to

distinguish spurious solutions from actual compounds, as actual compounds would not have to necessarily appear among consecutive windows anymore.

After inspecting the distribution of singular values along the windows, we set a threshold to determine the number of components. Data not represented by the selected components was discarded using a PCA filter.

Regarding the MCR-ALS configuration, we initialized the pure spectra and the concentration profiles for each window using SIMPLISMA. We imposed the following constraints: 1) non-negativity to both concentration profiles and pure spectra via fast non-negative least squares, 2) closure to the concentration profiles and 3) unimodality to the resolved spectra. Additionally, we imposed a selectivity constraint to improve the RIP pure spectrum estimation: given that at the end of the sample (high retention times) no compounds elute from the column, the only compound present in the latest spectra is the RIP. From a blind source separation perspective, this information is very valuable, as an accurate estimate of the RIP pure spectrum can be easily obtained.

Finally, in order to track the resolved spectra through the windows, an angle threshold of 15 degrees was used. This angle was chosen after inspecting the angle distribution of the pairwise comparison of the spectra.

3. Results and Discussion

A selected informative region of an olive oil sample can be seen in Fig. 2. The Reactant Ion Peak can be observed close to 6 ms in drift time along all the retention time range. There are multiple peaks in the same retention time range, indicating a strong co-elution of the components. As peaks are created by the transference charge from the reactant ions, the RIP intensity decreases at the retention time when other peaks appear in the ion mobility spectrum. At higher retention times the RIP recovers all the charge returning to a constant intensity as no more compounds elute.

The intensity of the RIP can be used as a non-selective measure of the global elution of compounds. By integrating the RIP (from 6.26 ms to 6.6 ms) and subtracting it from the maximum intensity, we obtain the charge that has been transferred to other compounds throughout the retention time. This figure of merit is called the "Reverse RIP" and it is analogous to a total ion chromatogram in gas chromatography - mass spectrometry samples. Fig. 3 shows the reverse RIP of an olive oil sample. The reverse RIP shows a continuous elution of compounds along approximately the first minute of the sample.

The performance of the proposed SW-MCR method has been assessed by comparing the extracted concentration profiles and pure spectra with the ones resolved using conventional MCR-ALS on the

whole sample, using the same described pre-processing and imposing the same constraints. Regions with strong co-elution are of particular interest, as for those regions conventional MCR-ALS is not able to resolve all compounds, especially the smallest ones. 22 peaks of the first 100 seconds of the sample were randomly selected (covering higher and lower peak intensities) and we checked the retention time range where each peak had been detected by each method. Table 1 shows the actual retention time range of the sample and the one obtained by each method. When the peaks are detected, there is considerable agreement between both methods; however MCR-ALS failed to detect 9/22 of the analyzed peaks.

Fig. 4 shows a sample region with co-elution and peak intensities of different magnitudes: at retention time 40 s two peaks appear: a peak of 2200 intensity units at drift time 10 ms and a less intense peak of 650 intensity units at 7.8 ms. Close to 50 s a third peak of 230 intensity units appears at 8.7 ms. Given the intensity and large tailing shape of the 10 ms peak, it is reasonable to think that the smaller 8.7 ms peak was co-eluting before its detection, but was being masked by the largest peak due to the charge competition effect. Nevertheless, the difference of the peak intensities detected by the IMS is almost of one order of magnitude.

By using MCR-ALS in the whole sample, the resolved pure spectra and concentration profiles on the described region are shown in Fig. 5. The only meaningful compound extracted at that retention time region apart from the RIP is the most intense one, found at 10 ms and marked using a wider line. Other compounds appear, some of them can be interpreted as tails or replicas of the 10 ms peak, but they provide no particular meaning so they must be discarded as spurious compounds. Additionally, the concentration profile for the resolved peak shows non-zero concentration in the 20-40 second retention time region, before the compound has eluted.

When using SW-MCR, the pure spectra and concentration profiles for the three peaks on the described region are extracted (see Fig. 6). The computed error bars of the pure spectra and concentration profiles show a high consistency among different window estimations. As expected, the concentration of the largest peak (at 10 ms) is similar to the concentration resolved using MCR-ALS. The peak with lowest intensity (at 8.7 ms) is well resolved too, with a concentration profile one order of magnitude smaller than the largest peak, as anticipated. The medium intensity peak (at 7.8 ms) is also detected, although its tracking is interrupted in the 47-53 seconds range. This shows a limitation of the proposed technique: Peaks with a constant intensity in the whole window cannot be detected by SIMPLISMA because the standard deviation of the peak maximum along the window is zero leading to zero purity values. However, the peak is tracked again in further windows once the intensity varies again. As expected, neither MCR-ALS nor SW-MCR were able to deconvolve the 8.7 ms peak when it was being completely masked by the 10 ms peak.

The SW-MCR technique allows extracting detailed information of the co-elution present in the sample. Fig. 7 shows the distribution of the resolved compounds along the retention time. Each row represents a tracked compound, showing the retention time region in which it has been detected and deconvolved. For instance, the first row represents the RIP, which is tracked along all the chromatogram. Fig. 2 and Fig. 3 showed multiple compounds co-eluting from the column on the first seconds of the analysis. Fig. 7 confirms that SW-MCR is able to detect them, resolving more than 6 compounds on a single window. As the retention time increases, Fig. 2 shows less peaks co-eluting, and this is reflected on Fig. 7 as the compound overlap decreases. On the first 100 seconds of the sample, SW-MCR was able to track up to 46 compounds revealing the richness of information present in the MCC-IMS samples.

The estimated concentration profile of the RIP is shown on Fig. 8. Retention time regions with lower RIP concentrations indicate regions with intense peaks, or regions with multiple peaks detected, where (almost) all the charge has already been transferred. The recovered concentration profile for the RIP can be compared with the extracted reverse RIP shown in Fig. 3. Fig. 8 peaks can be matched with Fig. 3 valleys, proving the consistency of our technique.

The angle threshold used for tracking the compounds rejects spurious peaks that appear on a single window. Fig. 9 shows the estimated number of components using singular value decomposition and the actual number of tracked components for a particular window. Regarding the number of tracked components, all windows are able to track the RIP and therefore the number of tracked components is always greater than or equal to one compound. Most of the windows had none or one spurious compound, although some windows had up to five spurious components which were rejected. This rejection allows us to use lower thresholds on the singular value decomposition, as overestimations in the number of components of a window are regulated in the peak tracking step. Lower thresholds on the singular values allow us to detect peaks close to the noise level, as they are consistent across different windows and are not rejected.

Regarding the computational cost of the technique, most of the computing time is spent on the MCR-ALS optimization, as it is an iterative algorithm. For the conventional MCR-ALS, up to 30 iterations are required to reach convergence. For the proposed SW-MCR a maximum of 10 iterations per window were used although for most of the windows 5 iterations were enough for MCR-ALS to converge. In any case, the most expensive part is the (non-negative) least squares estimation required on each concentration profile and pure spectrum estimation. On a 2013 workstation computer with an Intel i7 processor, the conventional MCR-ALS method takes 4.5 minutes to extract the concentration profiles and pure spectra from the sample. This process cannot be parallelized due to its iterative nature. The proposed SW-MCR method requires 1.75 seconds per

window. Given that there is a strong window overlap, the overall cost per sample sums up to 22 minutes. Even though the global time is higher, the SW-MCR method is well suited for parallelization, as each window is independent from the others. For our case, with a window shift of 0.7 seconds, three CPU cores would be needed for a real time application, unfeasible with the conventional algorithm.

4. Conclusions

A novel technique for improved chemometric resolution of gas chromatography ion mobility spectrometry samples has been presented, showing for the first time that blind peak deconvolution techniques can be successfully applied to this analytical hyphenated instrument.

The proposed technique has been tested on olive oil headspace samples. The samples analyzed present a strong co-elution of the individual chemical components, as shown by the wide peaks in the reverse RIP ion chromatogram. Co-eluting peaks are not properly resolved using conventional MCR-ALS methods, as peaks of lower intensities cannot be discriminated from the sample noise. Additionally, spurious compounds appear requiring supervision of the results.

Using the proposed SW-MCR method, we were able to deconvolve the pure spectra and concentration profiles of most of the peaks, even the ones with lower intensities, rejecting spurious solutions automatically.

The computational cost of the proposed technique is higher than the cost of the conventional MCR-ALS method, mainly because of the higher window overlap. However, our method can be easily parallelized, making our method more scalable and even suitable for real time applications.

Further work will be oriented to improving the initial estimation of the concentration profiles and pure spectra, in order to overcome the limitations in SIMPLISMA to resolve peaks of constant intensity along the entire window.

The proposed method has been applied to MCC-IMS samples, although we believe that it can be applied to other hyphenated instruments with similar strong co-elutions, as long as the detector is able to capture information from the analyte.

Acknowledgements

S.M., S.O., J.M.J. and A.P. are part of a consolidated research group recognized by the Generalitat de Catalunya (2014 SGR 1445). S.O. should also acknowledge a FPU doctoral grant to the Spanish Ministerio de Educación (AP2010-0010). This work is also financially supported by the Spanish project SMART-IMS (TEC2011-26143). The authors would also like to thank the anonymous

reviewers for their helpful comments and suggestions that greatly contributed to improving the final version of the paper.

References

[1] G. A. Eiceman, Z. Karpas, H. H. Hill Jr., Ion Mobility Spectrometry, Third Edition. CRC Press, 2013, p. 444.

[2] R. Cumeras, I. Gràcia, E. Figueras, L. Fonseca, J. Santander, M. Salleras, C. Calaza, N. Sabaté, and C. Cané, Finite-element analysis of a miniaturized ion mobility spectrometer for security applications, Sensors Actuators B Chem., vol. 170, pp. 13–20, Jul. 2012.

[3] R. Miller, G. Eiceman, E. Nazarov, and A. King, A novel micromachined high-field asymmetric waveform-ion mobility spectrometer" Sensors Actuators B Chem., vol. 67, no. 3, pp. 300–306, Sep. 2000.

[4] R. Ewing, A critical review of ion mobility spectrometry for the detection of explosives and explosive related compounds, Talanta, vol. 54, no. 3, pp. 515–529, May 2001.

[5] Z. Karpas, Y. Pollevoy, and S. Melloul, Determination of bromine in air by ion mobility spectrometry, Anal. Chim. Acta, vol. 249, no. 2, pp. 503–507, Jan. 1991.

[6] K. J. Budde, Application of Ion Mobility Spectrometry to Semiconductor Technology: Outgassings of Advanced Polymers under Thermal Stress, J. Electrochem. Soc., vol. 142, no. 3, p. 888, Mar. 1995.

[7] S. M., N. J. Vasa, V. Agarwal, and J. Chandapillai, UV photo-ionization based asymmetric field differential ion mobility sensor for trace gas detection, Sensors Actuators B Chem., vol. 195, pp. 44–51, May 2014.

[8] R. L. Eatherton, W. F. Siems, and H. H. Hill, Fourier transform ion mobility spectrometry of barbiturates after capillary gas chromatography, J. High Resolut. Chromatogr., vol. 9, no. 1, pp. 44–48, Jan. 1986.

[9] J. Cline and J. Hobbs, Laboratory Evaluation of Detectors of Explosives' Effluents, Cambridge, MA, 1972.

[10] R. Garrido-Delgado, L. Arce, and M. Valcárcel, Multi-capillary column-ion mobility spectrometry: a potential screening system to differentiate virgin olive oils, Anal. Bioanal. Chem., vol. 402, no. 1, pp. 489–98, Jan. 2012.

[11] R. Alonso, V. Rodríguez-Estévez, A. Domínguez-Vidal, M. J. Ayora-Cañada, L. Arce, and M. Valcárcel, Ion mobility spectrometry of volatile compounds from Iberian pig fat for fast feeding regime authentication, Talanta, vol. 76, no. 3, pp. 591–6, Jul. 2008.

[12] V. Ruzsanyi, J. I. Baumbach, S. Sielemann, P. Litterst, M. Westhoff, and L. Freitag, Detection of human metabolites using multi-capillary columns coupled to ion mobility spectrometers, J. Chromatogr. A, vol. 1084, no. 1–2, pp. 145–151, Aug. 2005.

[13] V. Pomareda, D. Calvo, A. Pardo, and S. Marco, Hard modeling Multivariate Curve Resolution using LASSO: Application to Ion Mobility Spectra, Chemom. Intell. Lab. Syst., vol. 104, no. 2, pp.

318–332, Dec. 2010.

[14] G. A. Eiceman, W. Yuan-Feng, L. Garcia-Gonzalez, C. S. Harden, and D. B. Shoff, Enhanced selectivity in ion mobility spectrometry analysis of complex mixtures by alternate reagent gas chemistry, Anal. Chim. Acta, vol. 306, no. 1, pp. 21–33, Apr. 1995.

[15] J. I. Baumbach, Ion mobility spectrometry coupled with multi-capillary columns for metabolic profiling of human breath, J. Breath Res., vol. 3, no. 3, p. 034001, Sep. 2009.

[16] J. I. Baumbach, S. Sielemann, Z. Xie, and H. Schmidt, Detection of the Gasoline Components Methyl tert -Butyl Ether, Benzene, Toluene, and m -Xylene Using Ion Mobility Spectrometers with a Radioactive and UV Ionization Source, Anal. Chem., vol. 75, no. 6, pp. 1483–1490, Mar. 2003.

[17] Z. Karpas, A. V Guamán, A. Pardo, and S. Marco, Comparison of the performance of three ion mobility spectrometers for measurement of biogenic amines, Anal. Chim. Acta, vol. 758, pp. 122–9, Jan. 2013.

[18] S. E. Bell, R. G. Ewing, G. A. Eiceman, and Z. Karpas, Atmospheric pressure chemical ionization of alkanes, alkenes, and cycloalkanes, J. Am. Soc. Mass Spectrom., vol. 5, no. 3, pp. 177–85, Mar. 1994.

[19] H. R. Shamlouei and M. Tabrizchi, Transmission of different ions through a drift tube, Int. J. Mass Spectrom., vol. 273, no. 1–2, pp. 78–83, Jun. 2008.

[20] A. B. Kanu and H. H. Hill, Ion mobility spectrometry detection for gas chromatography, J. Chromatogr. A, vol. 1177, no. 1, pp. 12–27, Jan. 2008.

[21] J.-F. Cardoso, Blind signal separation: statistical principles, Proc. IEEE, vol. 86, no. 10, pp. 2009–2025, 1998.

[22] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, Vol 1. 2002.

[23] L. T. Duarte, S. Moussaoui, and C. Jutten, Source Separation in Chemical Analysis: Recent achievements and perspectives, IEEE Signal Process. Mag., vol. 31, no. 3, pp. 135–146, May 2014.

[24] D. Nuzillard and J.-M. Nuzillard, Application of blind source separation to 1-D and 2-D nuclear magnetic resonance spectroscopy, IEEE Signal Process. Lett, vol. 5, no. 8, pp. 209–211, Aug. 1998.

[25] C. Carteret, A. Dandeu, S. Moussaoui, H. Muhr, B. Humbert, and E. Plasari, Polymorphism Studied by Lattice Phonon Raman Spectroscopy and Statistical Mixture Analysis Method. Application to Calcium Carbonate Polymorphs during Batch Crystallization, Cryst. Growth Des., vol. 9, no. 2, pp. 807–812, Feb. 2009.

[26] S. Miron, M. Dossot, C. Carteret, S. Margueron, and D. Brie, Joint processing of the parallel and crossed polarized Raman spectra and uniqueness in blind nonnegative source separation, Chemom. Intell. Lab. Syst., vol. 105, no. 1, pp. 7–18, Jan. 2011.

[27] I. Montoliu, R. Tauler, M. Padilla, A. Pardo, and S. Marco, Multivariate curve resolution applied to temperature-modulated metal oxide gas sensors, Sensors Actuators B Chem., vol. 145, no. 1, pp. 464–473, Mar. 2010.

[28] S. Bermejo, C. Jutten, and J. Cabestany, ISFET source separation: Foundations and techniques, Sensors Actuators B Chem., vol. 113, no. 1, pp. 222–233, Jan. 2006.

[29] A. Hyvärinen, J. Karhunen, and E. Oja, Independent Component Analysis. John Wiley & Sons, 2004, p. 504.

[30] A. Cichocki, R. Zdunek, and S. Amari, New Algorithms for Non-Negative Matrix Factorization in Applications to Blind Source Separation, in 2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings, 2006, vol. 5, pp. V–621–V–624.

[31] W. H. Lawton and E. A. Sylvestre, Self Modeling Curve Resolution, Technometrics, vol. 13, no. 3, Apr. 1971.

[32] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst., vol. 30, no. 1, pp. 133–146, Nov. 1995.

[33] T. L. Buxton and P. de B. Harrington, Rapid multivariate curve resolution applied to identification of explosives by ion mobility spectrometry, Anal. Chim. Acta, vol. 434, no. 2, pp. 269–282, May 2001.

[34] W. Windig, N. B. Gallagher, J. M. Shaver, and B. M. Wise, A new approach for interactive self-modeling mixture analysis, Chemom. Intell. Lab. Syst., vol. 77, no. 1–2, pp. 85–96, May 2005.

[35] S. Gourvénec, D. Massart, and D. Rutledge, Determination of the number of components during mixture analysis using the Durbin–Watson criterion in the Orthogonal Projection Approach and in the SIMPLe-to-use Interactive Self-modelling Mixture Analysis approach, Chemom. Intell. Lab. Syst., vol. 61, no. 1–2, pp. 51–61, Feb. 2002.

[36] J. Diewok, A. de Juan, M. Maeder, R. Tauler, and B. Lendl, Application of a Combination of Hard and Soft Modeling for Equilibrium Systems to the Quantitative Analysis of pH-Modulated Mixture Samples, Anal. Chem., vol. 75, no. 3, pp. 641–647, Feb. 2003.

[37] W. Windig and J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem., vol. 63, no. 14, pp. 1425–1432, Jul. 1991.

[38] M. Maeder, Evolving factor analysis for the resolution of overlapping chromatographic peaks, Anal. Chem., vol. 59, no. 3, pp. 527–530, Feb. 1987.

[39] P. de B. Harrington, E. S. Reese, P. J. Rauch, L. Hu, and D. M. Davis, Interactive Self-Modeling Mixture Analysis of Ion Mobility Spectra, Appl. Spectrosc., vol. 51, no. 6, pp. 808–816, 1997.

[40] A. Savitzky and M. J. E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem., vol. 36, no. 8, pp. 1627–1639, Jul. 1964.

Figures

Figure 1: Diagram of the tracking of spectra through three windows. Links between spectra are established if their angle is lower than a given threshold.

Figure 2: Region of a MCC-IMS olive oil sample. The Reactant Ion Peak (RIP) is observed at 6 ms. Multiple peaks on the same retention time indicate a strong co-elution. Note that both axes are reversed to prevent high intensity peaks from hiding the low intensity ones.

Figure 3: Reverse RIP. Analytes eluting from the column and detected by the IMS will appear as a peak in the reverse RIP. The RIP is computed as the integral (from 6.26 to 6.6 ms) of each IMS spectrum and it represents the charge that has not been transferred to other analytes at a given retention time. The RIP's maximum represents the total charge available. By subtracting the charge that has not transferred to the total charge, the reverse RIP is obtained.

Figure 4: MCC-IMS region (top view). This region shows co-elution of different compounds: a section of the RIP found at 6.4 ms and three other compounds appear 7.8 ms, 8.6 ms (less intensely) and 10 ms.

Figure 5: Pure spectra and concentration profiles resolved by MCR-ALS. Thick line: main peak resolved. Dashed lines: tails and replicas of the resolved peak. Thin lines: spurious compounds.

Figure 6: Pure spectra and concentration profiles resolved by SW-MCR, deconvolving three co-eluting spectra. The small size of the error bars shows a high consistency among the tracked windows.

Figure 7: Tracked compounds along different windows. Each row represents a different pure spectrum. The first one is the RIP, which is being tracked along all the windows. This plot shows how co-elution of three or more components can be resolved at several retention times.

Figure 8: The extracted Reactant Ion Peak using SW-MCR. The reverse shape of the concentration profile is comparable to the reverse RIP shown in Fig. 3.

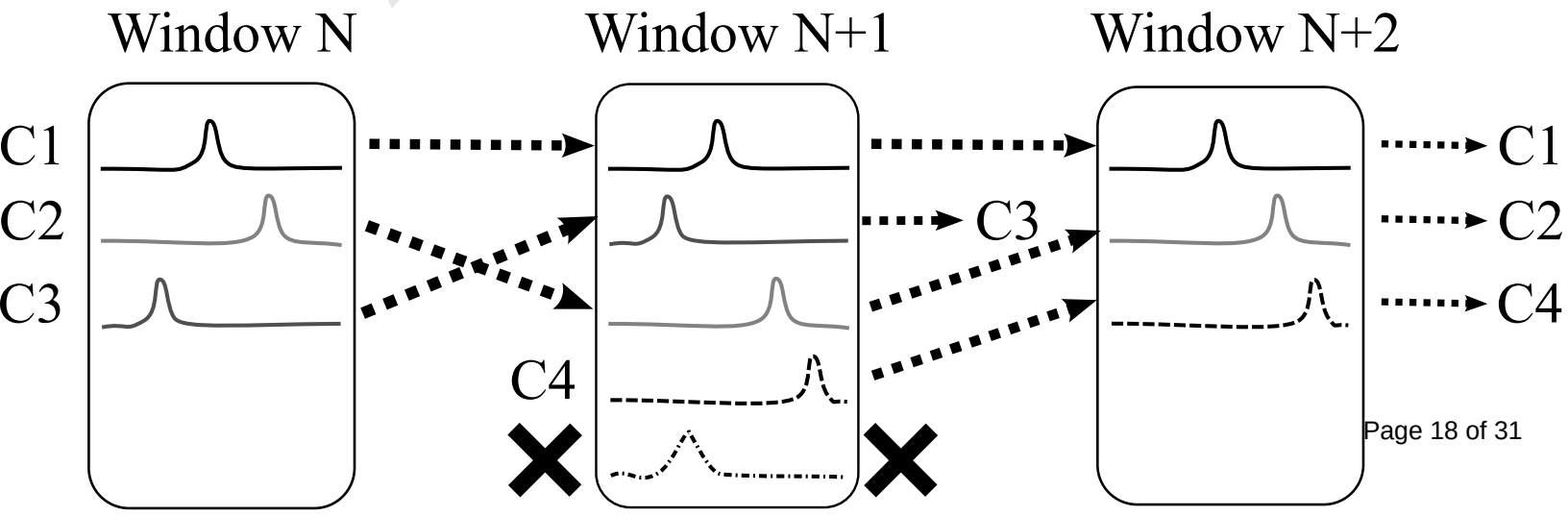
Figure 9: Rejection of spurious compounds. In black: the number of tracked compounds for each window. In white: the initial estimation of the number of components using singular value decomposition. The difference of both black and white values is the number of rejected spurious compounds.

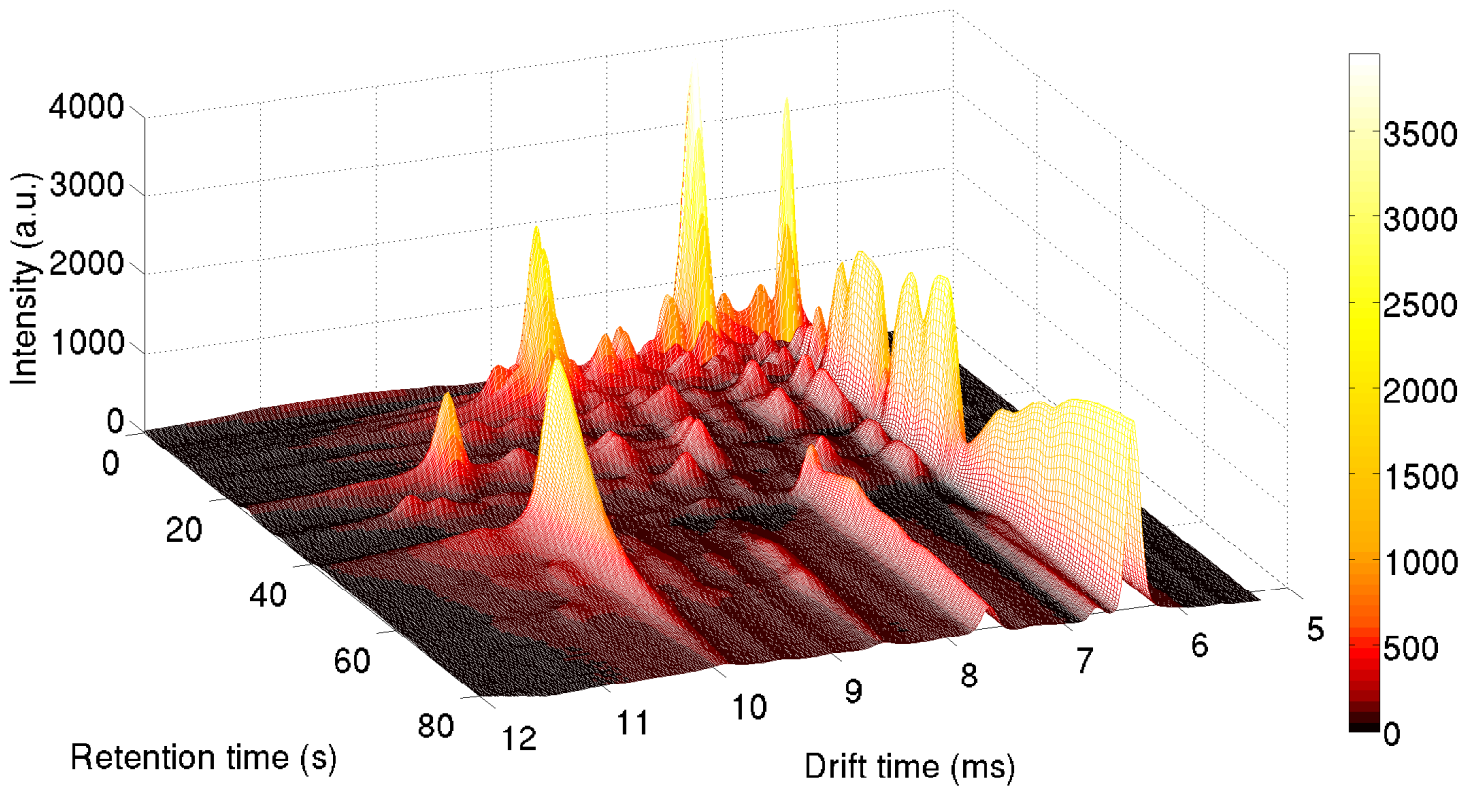
Tables

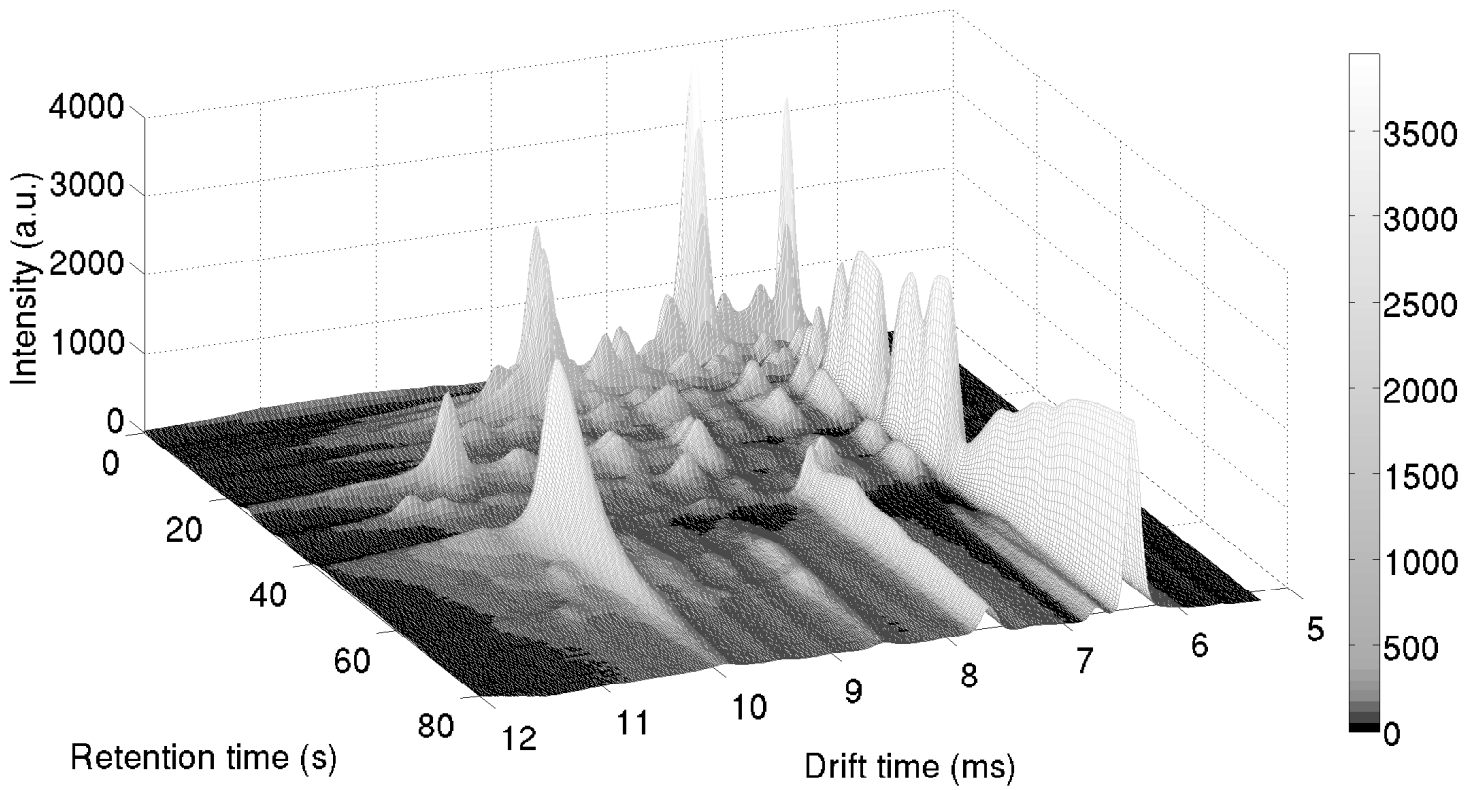
Table 1: Localization of 22 randomly selected peaks from the sample, on MCR-ALS and on SW-MCR deconvolution. The retention time ranges are in agreement on the detected peaks, however MCR-ALS was unable to extract 9/22 peaks.

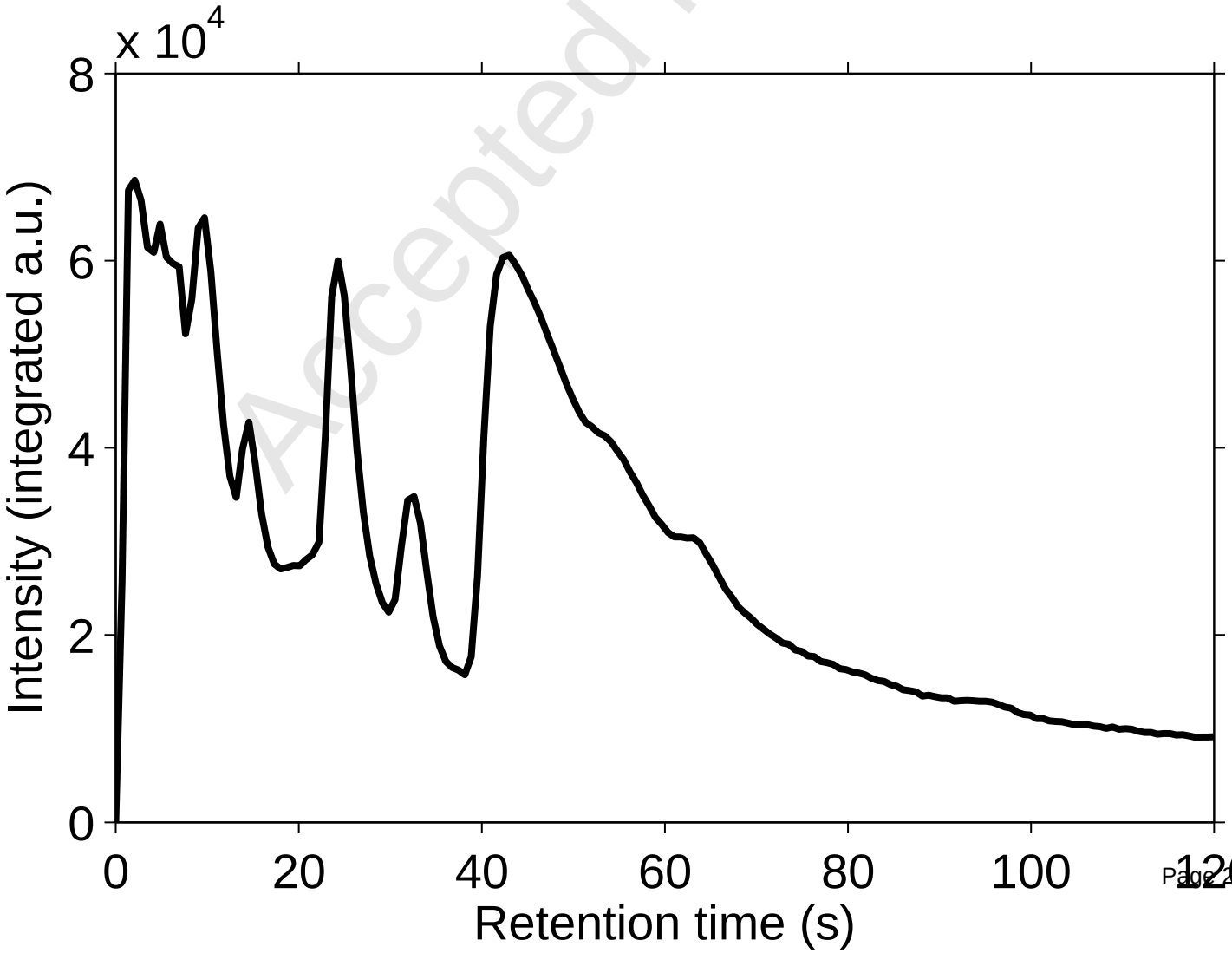
Table 1

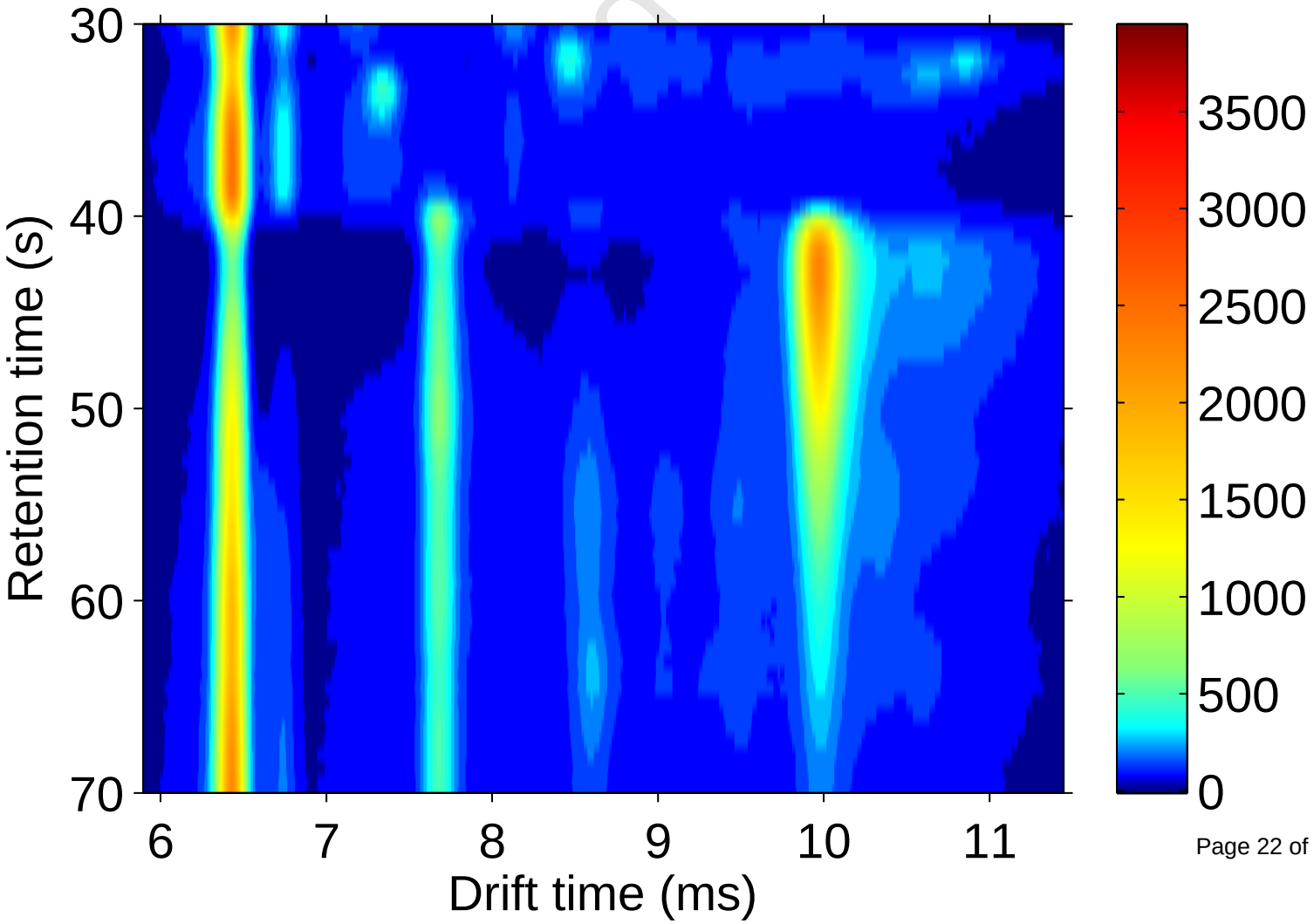
Peak #	Drift time (ms)	Max. intensity (a.u.)	Retention time range (s)		
			Sample	MCR-ALS	SW-MCR
1	6.45 (RIP)	3951	All	All	All
2	7.30	3936	1-4	0-10	0-4
3	7.60	1002	3-7	2-7	3-9
4	7.75	400	4-10	NF	4-9
5	8.30	782	5-9	4-8	4-10
6	9.10	177	5-12	NF	4-10
7	8.10	695	4-9	0-20	4-10
8	8.78	731	5-12	4-12	5-12
9	8.60	623	4-12	4-7	5-13
10	7.15	474	4-13	3-7	5-15
11	8.90	2100	8-13	8-12	8-13
12	8.10	425	11-19	NF	12-20
13	6.75	490	11-23	NF	12-23
14	10.30	1190	20-27	22-27	21-27
15	7.20	188	21-29	NF	21-28
16	8.20	481	21-32	NF	21-31
17	8.50	390	30-38	30-35	30-37
18	7.80	317	28-40	NF	30-40
19	7.30	445	31-40	NF	32-40
20	9.90	2200	40-55	40-60	40-50
21	8.70	220	50-63	NF	50-56
22	7.65	650	45-80	50-80	55-80

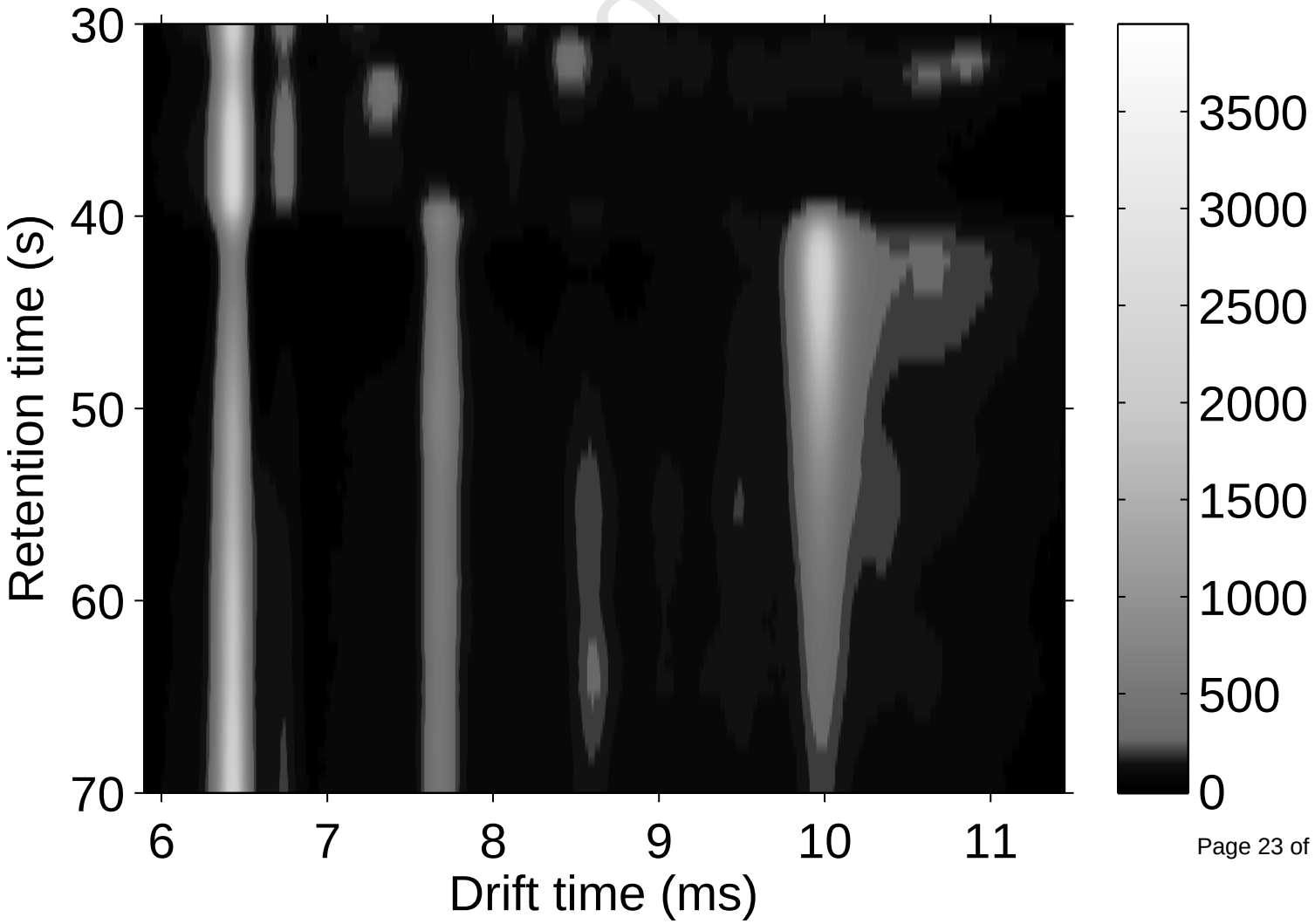


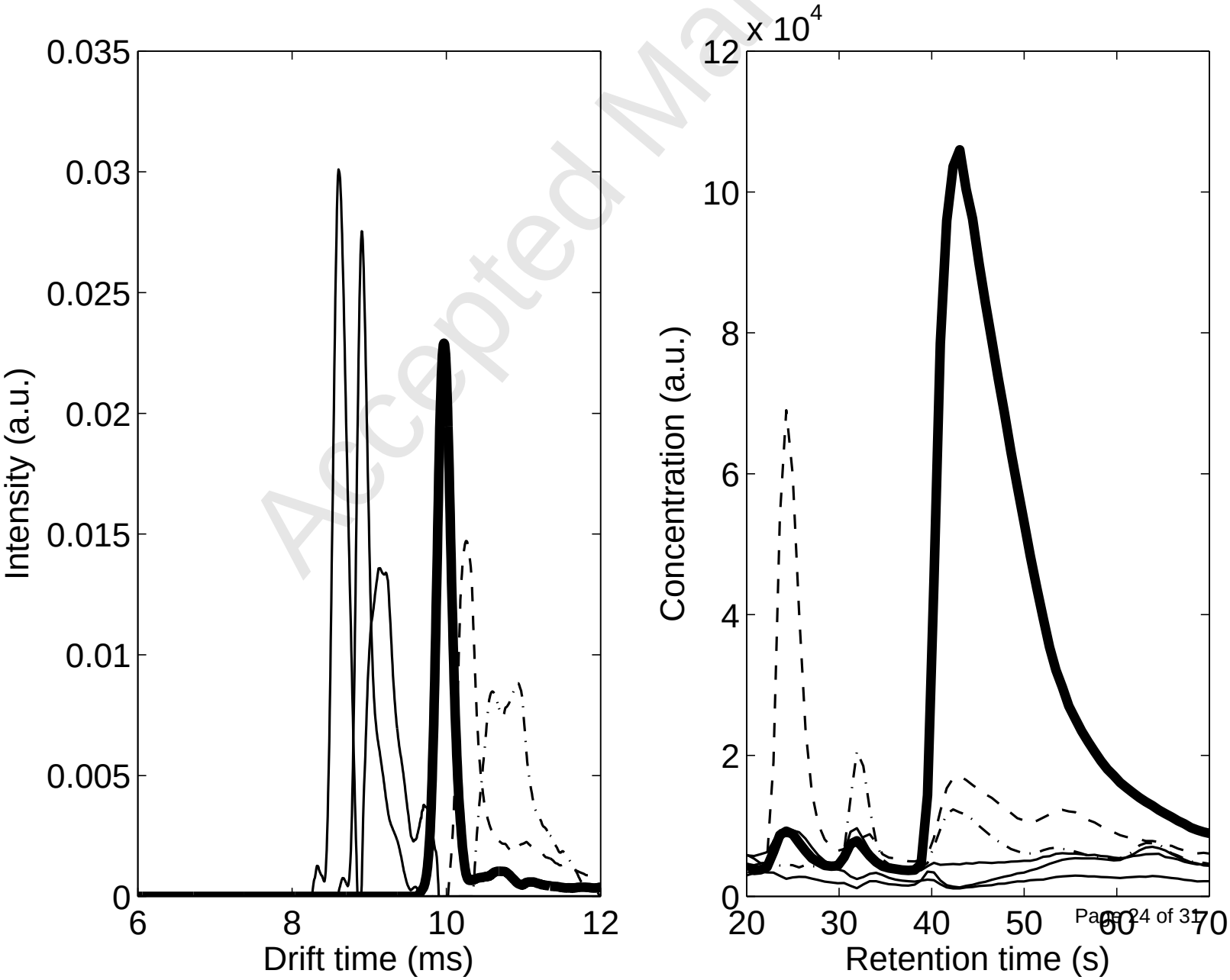


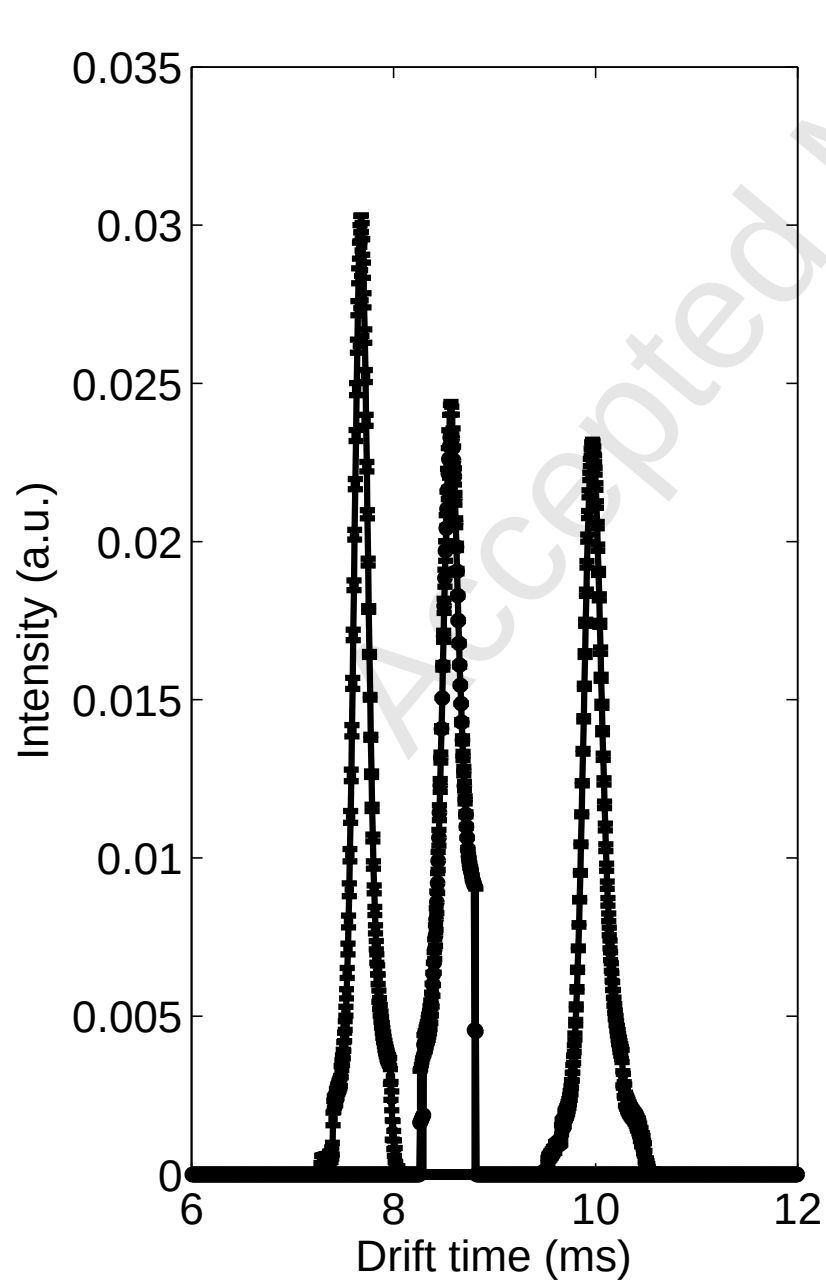


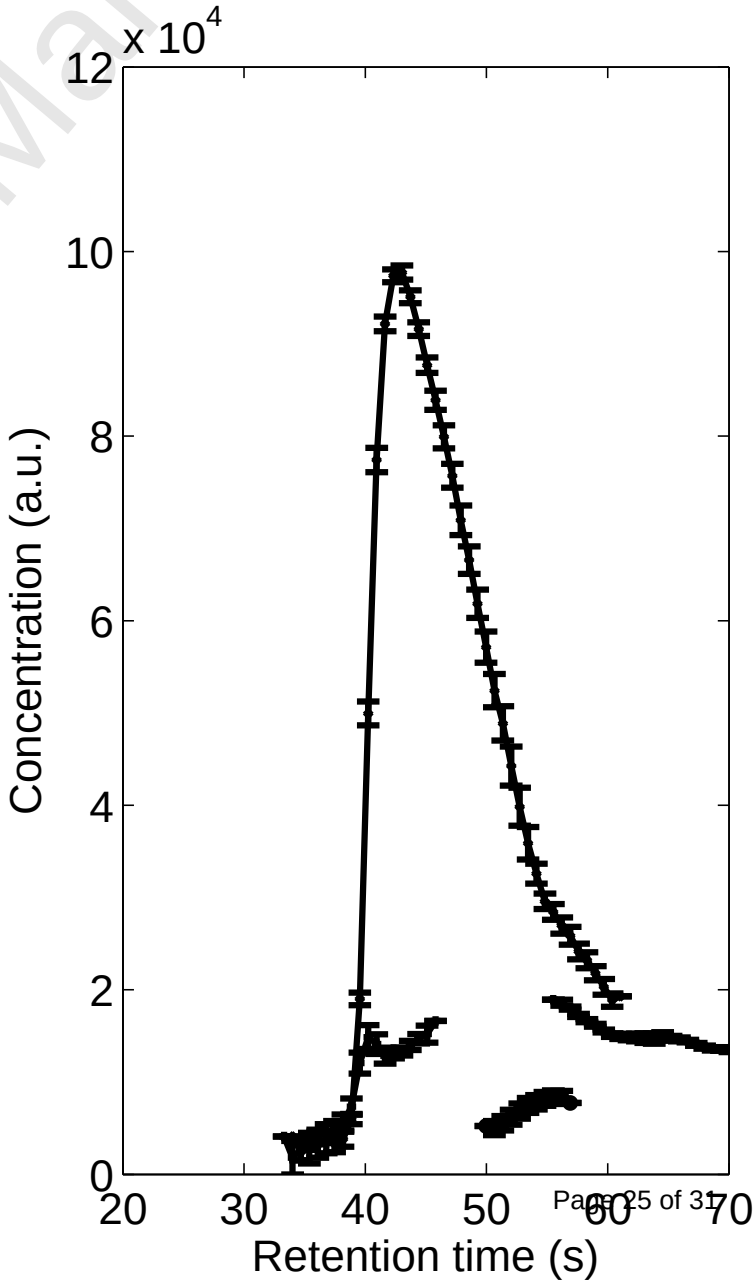


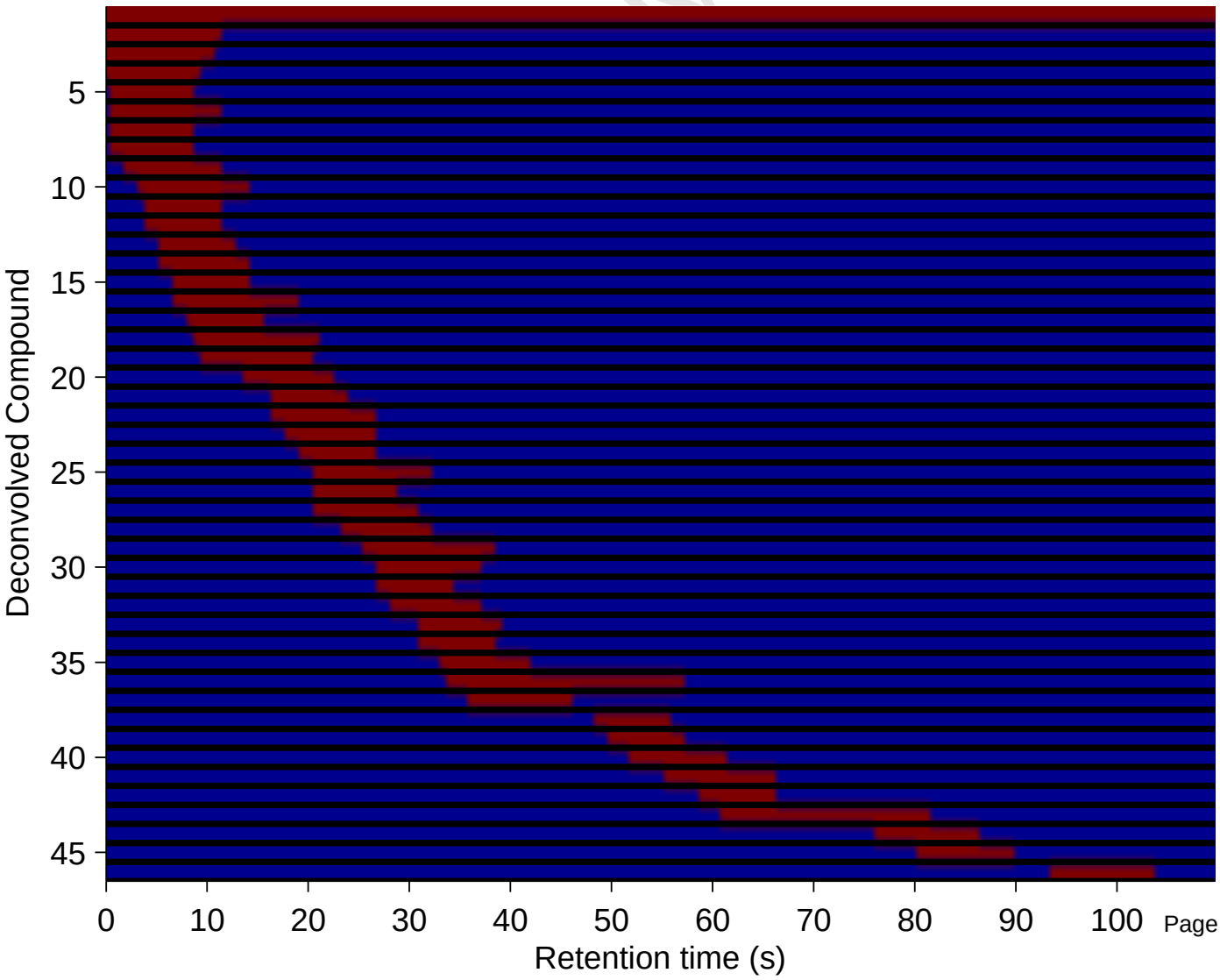


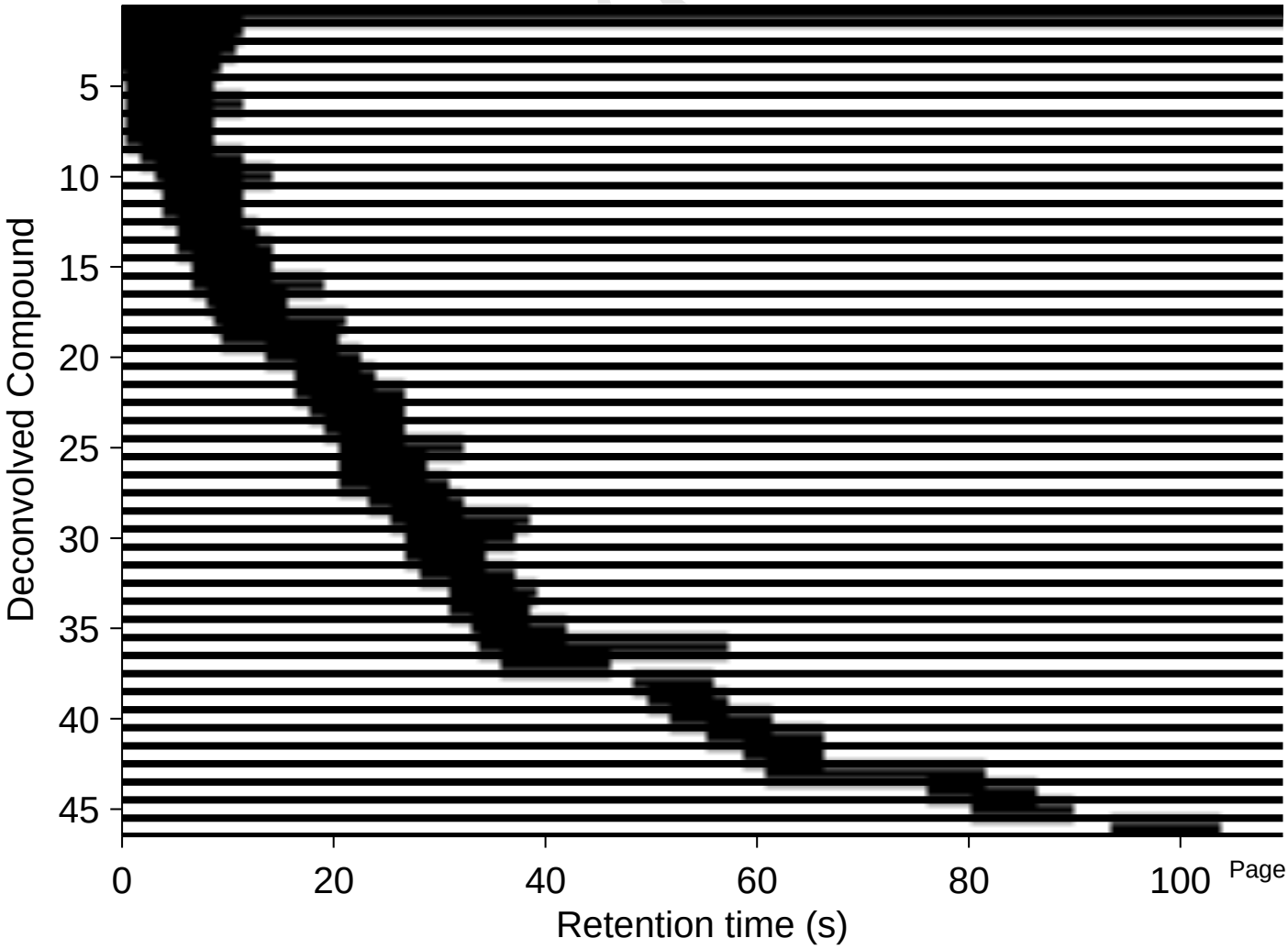


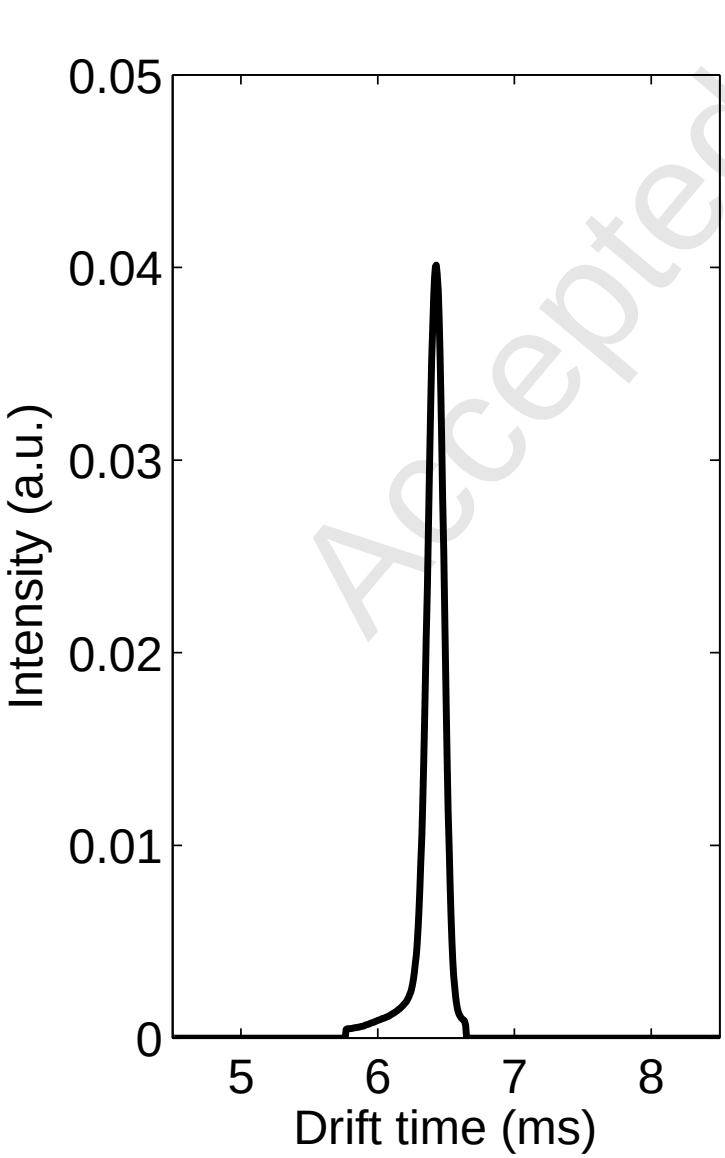


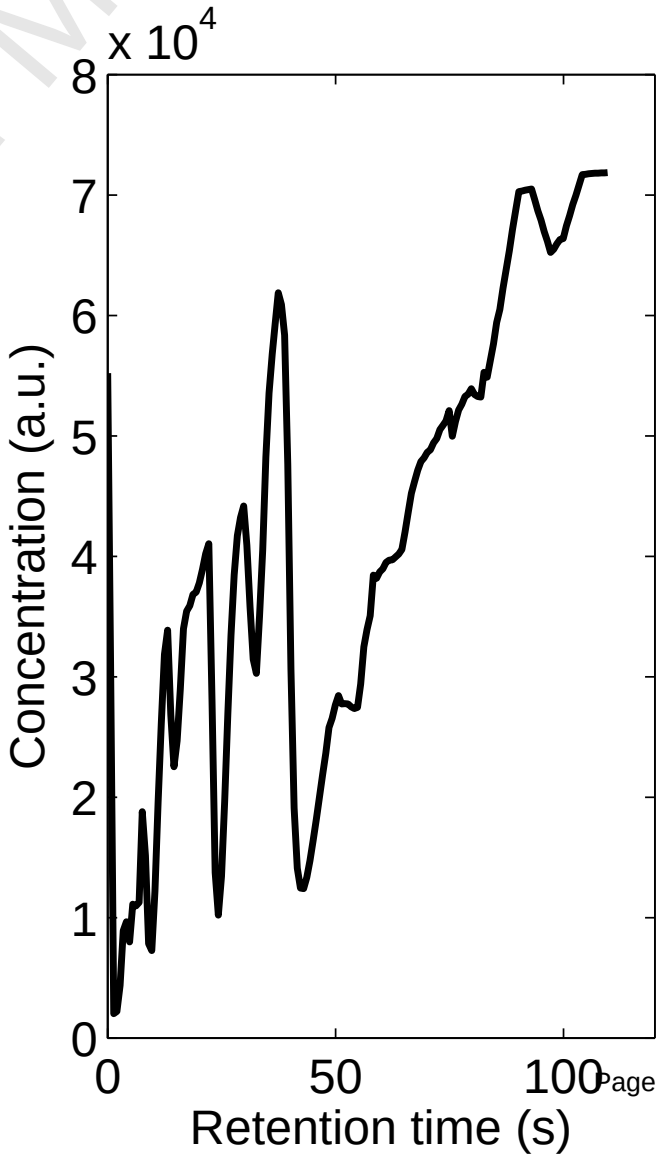


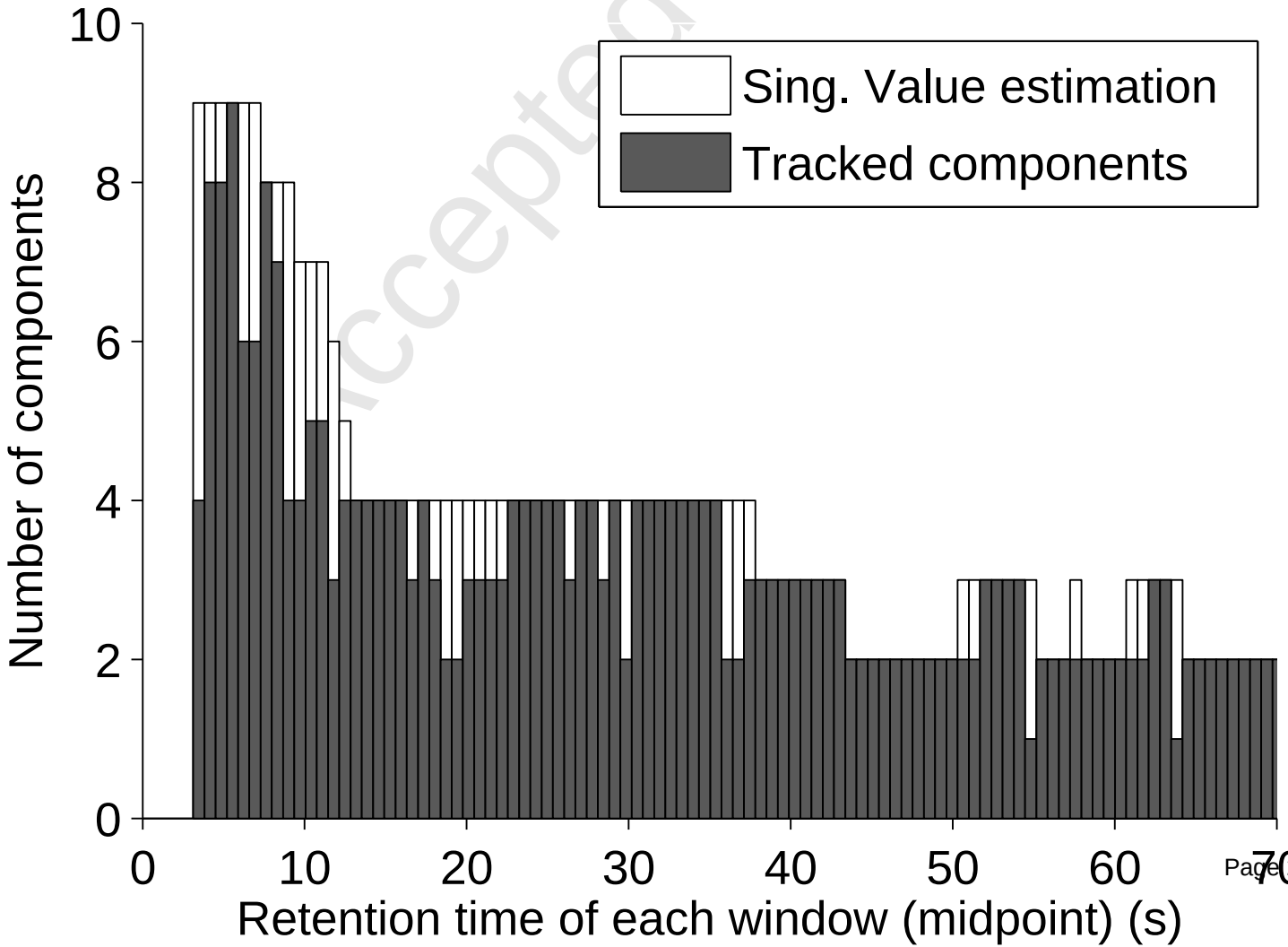


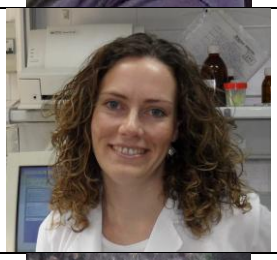










	Sergio Oller received the B.Sc degree in Physics from University of Barcelona and the M.Sc. degree in Computational and Applied Physics from BarcelonaTECH - Polytechnic University of Catalonia in 2011. He is a PhD. student at the Institute for Bioengineering of Catalonia and his current research interests are signal and data processing for analytical instrumentation with -omics applications. http://orcid.org/0000-0002-8994-1549
	Guillem Singla-Buxarrais received his B.Sc. degree in Biomedical Engineering from the University of Barcelona in June 2014. He is now pursuing the MSc in Biomedical Engineering with Neurotechnology at the Department of Bioengineering at Imperial College London. He has conducted research in the fields of lipid metabolism, signal and information processing for sensing systems, integrated cell and tissue dynamics, and he is currently undertaking a research project on human and robotic haptic exploration at the Brain and Behaviour Lab at Imperial College.
	Juan Manuel Jiménez-Soto did his PhD in the Department of Analytical Chemistry at the University of Córdoba from 2008 to 2012. Previously he completed his M.Sc. degree in Advanced Fine Chemistry and a B.Sc. degree in Environmental Sciences at the University of Córdoba. Currently he is developing his research at the Institute for Bioengineering of Catalonia since 2012.
	Dr. Antonio Pardo (Barcelona, 1967) received his diploma in physics 1991 and his PhD in 2000 from the University of Barcelona. During his PhD studies he worked in system identification with applications in gas sensor systems. Since 2005 he is associate professor in the Departament d'Electrònica at Universitat de Barcelona. He has been involved in several scientific and technologic projects as a researcher and as coordinator, in which the chemical instrumentation has a key role. His research interests include signal processing for gas sensors and pattern recognition as well as hardware and software developments for artificial olfaction.
	Rocío Garrido Delgado got her Master's Degree in Chemistry at the University of Córdoba (Spain) in 2007. Since then, she joined at Analytical Chemistry department at the University of Cordoba working with Ion Mobility Spectrometry. She presented her Doctoral Thesis in November 2011. Currently she continues developing her research in the Prof. Miguel Valcárcel group.
	Lourdes Arce obtained her Master's Degree in Chemistry from the University of Córdoba (Spain). She obtained her PhD in Sciences in 1999. Since 2007, she is associate professor in the Analytical Chemistry Department at the University of Córdoba. Her main research project is devoted to the "Development and application of vanguard (Ion Mobility Spectrometry)-rearguard (Capillary Electrophoresis, Gas Chromatography, etc.) new metrology analytical strategies". Current research is directed towards the pursuit of studying the potential of IMS and CE in the agro-food and forensic field.

Dr. Santiago Marco completed his university degree in Applied Physics in 1988 and received a PhD in Microsystem Technology from the University of Barcelona in 1993. He held a European Human Capital Mobility grant for a postdoctoral position at the Department of Electronic Engineering at the University of Rome "Tor Vergata". Since 1995, he is Associate Professor of Electronic Instrumentation at the Department of Electronics at the University of Barcelona. In 2004 he had a sabbatical leave at EADS-Corporate Research, Munich, working on Ion Mobility Spectrometry. In 2008 he was appointed leader of the Signal and Information Processing for Sensing Systems Lab at the Institute of Bioengineering of Catalonia. His research concerns the development of signal/data processing algorithmic solutions for smart chemical sensing based in sensor arrays or microspectrometers integrated typically using Microsystem Technologies (more at http://www.ibecbarcelona.eu/signalinfo).