BARCHAN: Blob Alignment for Robust CHromatographic ANalysis

Camille Couprie, Laurent Duval, Maxime Moreaud,
Sophie Hénon, Mélinda Tebib, Vincent Souchon

February 25, 2017

Abstract

Two dimensional gas chromatography (GC×GC) plays a central role into the elucidation of complex samples. The automation of the identification of peak areas is of prime interest to obtain a fast and repeatable analysis of chromatograms. To determine the concentration of compounds or pseudo-compounds, templates of blobs are defined and superimposed on a reference chromatogram. The templates then need to be modified when different chromatograms are recorded. In this study, we present a chromatogram and template alignment method based on peak registration called BARCHAN. Peaks are identified using a robust mathematical morphology tool. The alignment is performed by a probabilistic estimation of a rigid transformation along the first dimension, and a non-rigid transformation in the second dimension, taking into account noise, outliers and missing peaks in a fully automated way. Resulting aligned chromatograms and masks are presented on two datasets. The proposed algorithm proves to be fast and reliable. It significantly reduces the time to results for GC×GC analysis.

1 Introduction

First introduced in 1991 by Phillips et al. [1], comprehensive two-dimensional gas chromatography (GC×GC) has become in the past decade a highly popular and powerful analytical technique for the characterization of many complex samples such as food derivatives, fragrances, essential oils or petrochemical products [2, 3, 4, 5]. In the field of oil industry, GC×GC gives an unprecedented level of information [6] thanks to the use of two complementary separations combining different selectivities. It is very useful in the understanding of catalytic reactions or in the design of refining process units [7, 8].

From an instrumental point of view, much progress has been made since the early nineties on both hardware and modulation systems [9]. Many modulator configurations are depicted in the

literature or nowadays sold by manufacturers. With the use of leak-free column unions, many of these systems have become robust, easy to use, without cryogenic fluids handling while providing high resolution. Within a series of several consecutive injections, almost no significant shifts in retention times are observed and repeatability of GC×GC experiments is nowadays a minor problem. However, reproducibility of GC×GC results for detailed group-type analysis on complex mixtures is still a great challenge due to column aging, trimming or to slight differences in column features. This results in shifts on retention times that can affect the proper quantification of a single compound, a group of isomers or pseudo-compounds. Experimental retention time locking (RTL) procedures have been proposed to counterbalance shifts on retention times but these procedures must be repeated regularly [10]. On the way to routine analysis for GC×GC, data treatment has therefore become the preferred option to reduce the time to results [11, 12, 13, 14].

A common way of treating GC×GC data is to quantify compounds according to their number of carbon atoms and their chemical families by dividing the 2D chromatographic space into contiguous regions that are associated to a group of isomers. This treatment benefits from the group-type structure of the chromatograms and from the roof-tile effect for a set of positional isomers. For example, for a classical diesel fuel, up to 300 or 400 regions (often referred to as blobs) may be defined. Due to the lack of robustness in retention times, this step often requires human input and is highly time-consuming when moving from an instrument to another or when columns are getting degraded. Several hours may be necessary to correctly recalibrate a template of a few hundreds of blobs on a known sample. This operator-dependent step causes variability in quantitative results which is detrimental to reproducibility. In that goal, 2D-chromatogram alignment methods, consisting in modifying a recently acquired chromatogram to match a reference one, have been a quite active research area.

In this paper, we propose a new algorithm called BARCHAN[1] which aims at aligning GC×GC chromatograms. It relies on a first peak selection step and then considers the alignment of the two point sets as a probability density estimation problem. This algorithm does not require the placement of anchor points by the user.

2   Material and methods

2.1   Datasets and GC×GC methods

The straight-run gas-oil sample named GO-SR which is used in this study was provided by IFP Energies nouvelles and was analyzed on different experimental set-ups. Its boiling point distribution ranges from $180\,°C$ to $400\,°C$.

Dataset 1 was built by considering two GC×GC chromatograms obtained on two different experimental set-ups in the same operating conditions with cryogenic modulation. These GC×GC experiments were carried out with an Agilent 7890A chromatograph (Santa Clara, California, USA) equipped with a split/splitless injector, a LN2 two-stage 4 jets cryogenic modulation system from LECO (Saint-Joseph, Michigan, USA) and an FID. The two evaluated column sets were composed of a first 1D apolar 1D HP-PONA column ($20\,m$, $0.20\,mm$, $0.50\,\mu m$, J&W, Folson, USA) and a mid-polar BPX-50 column ($1\,m$, $0.10\,mm$, $0.10\,\mu m$, SGE, Milton Keynes, United Kingdom) connected together with Siltite microunions from SGE. Experiments were run with a constant flow rate of $1\,mL\,min^{-1}$, a temperature program from $60\,°C$ ($0.50\,min$) to $350\,°C$ at $2\,°C\,min^{-1}$, a $+30\,°C$ offset

---

[1]The name is inspired by wind-produced crescent-shaped sand dunes (<u>barkhan</u> or <u>barchan</u>) reminiscent of 2D chromatogram shapes.

for hot jets and a 8 s modulation period. $0.50\,\mu L$ of neat sample were injected with a 1/100 split ratio.

Dataset 2 includes a reference chromatogram obtained in the previous conditions and a chromatogram of the same sample obtained with a microfluidic modulation system. These GC×GC data were obtained on a Agilent 7890B chromatograph equipped a split/splitless injector, a Griffith [15] type modulation system supplied by the Research Institute for Chromatography (Kortrijk, Belgium) and a FID. The modulation system consists in two Agilent CFT plates (a purged three-way and a two-way splitter) connected to an accumulation capillary. Separation was performed on a DB-1 (20 m, 0.10 mm, $0.40\,\mu m$) 1D column and a DB-17HT (10 m, 0.32 mm, $0.15\,\mu m$, J&W) 2D column. The modulation period was set to 10 s whereas the oven programming and injection conditions were similar to the ones previously described.

2.2   Software

BARCHAN is implemented in C and Matlab. In-house platform INDIGO runs it through a user-friendly interface while the proprietary 2DChrom® software creates template masks (.idn files) and 2D images from GC×GC data.

2.3   Calculations

The quality of the alignments obtained with BARCHAN was evaluated by two different ways. The correlation coefficient CC [16] as well as the Structural Similarity index SSIM [17] between the reference chromatogram and the other one were computed. They directly match global image intensities, without feature analysis. Calculation details for CC and SSIM are provided in the supplementary material. These results were obtained on a restricted area of interest defined by the user. A second indicator to evaluate the quality of the alignment is the match quality between the BARCHAN adjusted template and a fully manually registered template. In practice, this featural similarity index is performed by comparing quantitative results obtained on chemical families with the template mask and with the BARCHAN optimized mask.

3   Theory

3.1   Related works

We may distinguish two classes of alignment methods: the ones that are directly performed on the full chromatographic signal, and the others which require a prior peak selection step. In the first class, the works of [18] and [19] look for shifts minimizing a correlation score between signals. In [20], an affine transformation is assumed between the two chromatograms to register. The recent work of de Boer [16] looks for a warping function parametrized with splines that transforms the chromatogram to be registered into a chromatogram aligned with the reference. Low-degree polynomial alignment is proposed in [21]. Full image registration [22] is however limited for applications in GC×GC because of the variability in chromatograms: positions of peaks in the two chromatograms could be similar, but this is not the case of their intensities. Therefore, the majority of alignment methods choose to first extract peaks in the reference and target chromatograms to only register the informative parts of chromatographic images.

3

Thus, among approaches dedicated to chromatogram alignment, the work of [23] (focused on quantitative analysis) deduces local peaks displacements by correlation computations in slightly shifted blocks surrounding peaks. Variations of peak patterns in different experimental conditions (e.g. temperature) is studied in [24], and exhibits satisfactory results for estimating an affine transformation. Similarly, [25] also models rigid transformations for LC×LC (2D liquid chromatography) template alignment. However, these hypotheses appear to be too restrictive in a general setting. Therefore [26] extended the space of possible deformations by looking for a warping function that transforms signals. Correlation Optimized Warping (COW) is judged effective by [27] that compares three different registration approaches, including target peak alignment (TPA) and semi-parametric time warping (STW) for one specific analysis. However, COW is still not satisfactory when incomplete separation and co-elution problems exist as pointed out by [28]. Instead, the latter uses bilinear peak alignment in addition to COW to correct for progressive within run retention time shifts on the second chromatographic dimension. In [29], the alignment is performed after embedding the chromatograms surfaces into a three-dimensional cylinder, and the parametrization of the transform employs polynomials. The DIstance and Spectrum Correlation Optimization (DISCO) alignment method of [30], extended in [31], uses an elaborate peak selection procedure followed by interpolation to perform the alignment. The approach from [32] also performs peak alignment via correlation score minimization using dynamic programming, comparing favorably to DISCO. Finally, the work of [33] performs an assessment of different GC×GC alignment methods with a new one. Their method requires a manual placement of matching peaks pairs, then the registration is performed differently on each axis: linear deformations along one dimension, and a neighbor based interpolation in a Voronoi diagram defined using the alignment anchor points for the other dimension. The linear constraint is relevant because one dimension displacements are independent of the other dimension elution times. The requirement of user-defined alignment points is robust to large variations in the reference and target chromatogram, at the expense of time-consuming markers placement.

3.2   BARCHAN methodology

A schematic view for the principles of BARCHAN is depicted in the flowchart from Figure 1. First, GC×GC chromatograms of the sample to analyze and the reference 2D image are loaded as images files. Then, the user is provided with a brush to surround, in a user interface, the area of interest on both reference and new 2D chromatograms (see Figure 3). Peaks are extracted in those areas (Section 3.3). Only one centroid per local maximum is retained for the point set registration in order to diminish computation times and to prevent bias for large peaks. Datasets are then assimilated to centroids of a Gaussian Mixture Model (GMM) and a weighted noise model is added. Advantage is taken from recent progresses in point set registration, using a probabilistic and variational approach [34]. This choice is motivated by the fact that a complex transformation must be modeled while remaining robust to noise and outliers. In this context, GMMs (Gaussian Mixture Models) are particularly efficient at reconstructing missing data, which is especially convenient when selected peaks in one point cloud are not included in the other one. Finally, model parameters are optimized to yield registered results. Two types of results are produced:

- if a template mask for the reference chromatogram exists, the transformation of the template points leading to a registered template mask is computed.

- an aligned chromatogram may also be produced by computing the transformation of a grid

Figure 1: Flowchart methodology for BARCHAN.
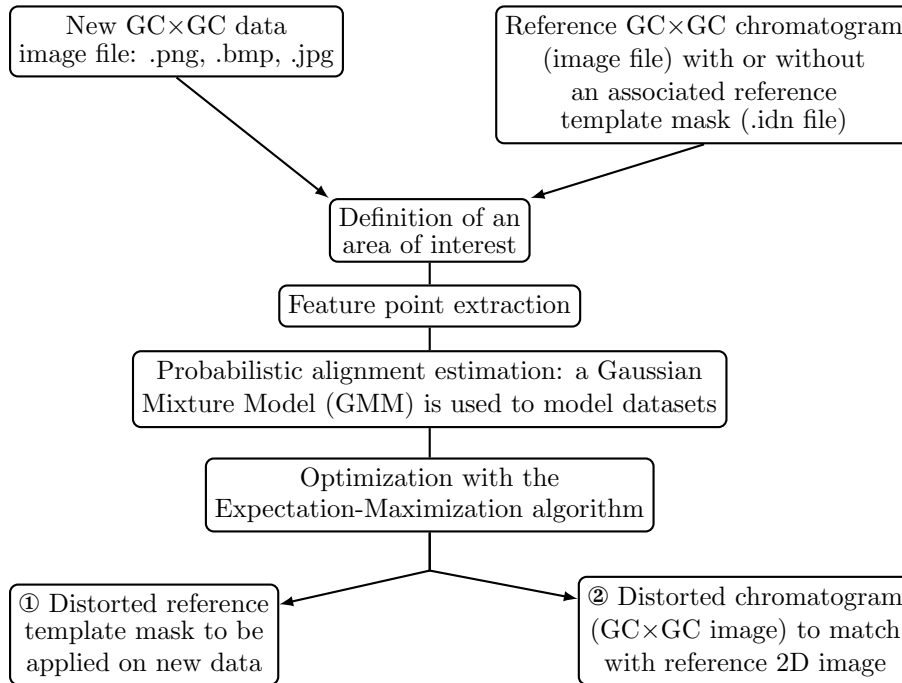
defined as the coordinates of every pixel in an image, and interpolating the target image values at the transformed coordinates.

Details on the calculations for every step are provided in the next paragraphs.

3.3 Feature point extraction

Despite the good behavior of the employed registration algorithm regarding noise and outliers, it is desirable to extract the most resembling point sets. Therefore, preliminary GC×GC enhancement [35, 36] proves useful.

Inherent to the GC×GC experimentation procedure, fragments of the stationary phase are frequently lost by the column resulting into the presence of hyperbolic lines in the chromatogram. Their differentiation from the real peaks is difficult to automate because of possible overlaps with the chromatogram peaks of interest. Therefore, in our treatments, a rough area of interest is delimited by an operator, taking approximately ten seconds.

Rather than using second or third derivatives of the chromatogram [37], which require non-trivial parameters to set, we employ the approach of [7, p. 97–106] and extract the $h-$maxima of the chromatograms. Simply put, all local maxima having a height greater than a scalar h are extracted. Starting from an input signal f from $\mathbb{R}^d$ to $\mathbb{R}$, the positions of the h-maxima may be obtained via a morphological opening by reconstruction, noted $\gamma^{\mathrm{rec}}(f, f - h)$. More specifically, this operation is defined as the supremum of all geodesic dilations of $f - h$ by unit balls in f. More details are provided in [7] and a scheme is displayed in Figure 2.

5

Figure 2: Detection of h-maxima: only one peak is detected in this example (dotted line).
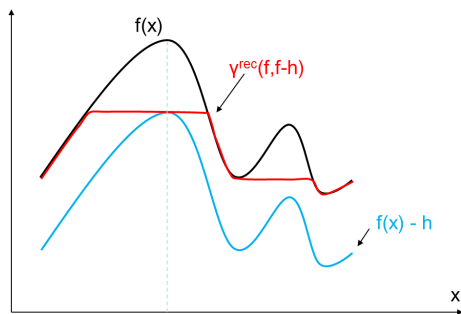
3.4 Data alignment model

To guarantee results where the first point set is similar to the registered point set, while being robust to noise and outliers, we choose to employ a probabilistic approach. Supposing that the first point set X follows a normal distribution, the Coherent Point Drift method [34] seeks to estimate the probability density that explains the data X as a weighted sum of Gaussians initially centered by the second point set Y.

We introduce our notations as follows. The first point set of size $N \times 2$, corresponding to the coordinates of N peaks extracted in the target chromatogram, is denoted $X = \{X_1, \ldots, X_N\}$. The second point set $Y = \{Y_1, \ldots, Y_M\}$ of size $M \times 2$ corresponds to the peak coordinates in the reference chromatogram and is assimilated to centroids of a GMM. Each component X_i is a vector composed of two coordinates denoted $X_i^{(1)}$ and $X_i^{(2)}$. The vector $X^{(i)}$ denotes the ith line of matrix X. Adding a weighted noise model to the GMM probability density function leads to:

$$p(X_n) = \frac{w}{N} + \sum_{m=1}^{M} \frac{1-w}{2M\pi\sigma^2} \exp\left(-\frac{\|X_n - T(Y_m)\|^2}{2\sigma^2}\right) \tag{1}$$

where the first term takes into account uniform noise weighted by the parameter w fixed between 0 and 1, σ is a variance parameter to estimate, and T is the point cloud transform to estimate. In this work, motivated by a failure of global rigid transformation attempts on our data, we modeled two different transforms across the two dimensions. We assume that a rigid displacement occurs along the y-axis second very short column, similarly to [33], and non-rigid transformations are allowed on the x-axis first normal length column. The underlying assumption is a relative anisotropy of the data: two separate pixels in the vertical direction are distant by a much smaller time interval than those aligned horizontally. The x-axis is thus potentially subject to more important nonlinear distortions. Thus, we model the transformation T of point cloud Y as:

$$T(Y^{(1)}) = sY^{(1)} + t, \tag{2}$$

$$T(Y^{(2)}) = Y^{(2)} + GW, \tag{3}$$

where s and t are real numbers, respectively a scale and a translation parameter to estimate, and W is a vector of length M of non-rigid displacements to estimate. The matrix $G \in R^{M \times M}$ is a symmetric matrix defined element-wise by:

6

$$G_{ij} = \exp^{-\frac{\|Y_i - Y_j\|}{2\beta}}, \tag{4}$$

where β is a positive scalar. The minimization of the non-negative likelihood leads to the minimization of:

$$E_1(\sigma, W, s, t) = -\sum_{n=1}^{N} \log p(X_n). \tag{5}$$

A regularization of the weights W, enforcing the motion to be smooth, is necessary for the non-rigid registration, resulting into the following variational problem:

$$\min_{\sigma, W, s, t} E = E_1(\sigma, W, s, t) + \frac{\lambda}{2} \operatorname{Tr}(W^\top GW), \tag{6}$$

where Tr denotes the trace operator of a matrix. The estimation of parameters w, β and λ is discussed in generic terms in [38] and [34]. BARCHAN inherits a similar strategy, within the proposed combined rigid/non-rigid registration procedure. The parameter $w \in [0, 1]$, related to the noise level, is first determined by visual inspection on ten regularly-spaced values. Albeit found to be the most determinant, our chromatograms sharing about the same signal-to-noise ratio, this value is kept constant in all our experiments. For other data types, multiple figures illustrating different registrations with varying amounts of noise and outliers with an appropriate choice of w are presented in [34]. The determination of the other parameters and λ is also discussed in [38]. We have set them to $\beta = 2$ and $\lambda = 2$ as by default in [34]. Slight changes did not affect the registration results sensitively.

3.5 Optimization

We employ the Expectation-Maximization (EM) algorithm [39] that alternates between:

- the E step: we compute the probability P of correspondence for every couple of points.

- the M step: we estimate the parameters σ, s, t, and W. To that goal, we compute the partial derivative of E with respect to σ, s, t, and W and set them to zero leading to an estimate of every parameter. Details are provided in the supplementary material, as well as the final algorithm itself.

4 Results and discussion

The areas of interest for both dataset 1 and 2 were defined so that every compound present in the sample is taken into account while limiting the number of peaks due to the bleeding (Figure 3). The detected peaks appear as small blue dots on both chromatograms whereas the selected areas are colored in green and delimited with a purple line. Peaks were extracted with a height parameter h from Section 3.3 equal to 120 and 60 for dataset 1 and 2, respectively.

Three types of transformations were evaluated: rigid transformations on both the x- and the y-axis, non-rigid transformations on both axes and BARCHAN transformation (non-rigid transformation on x-axis, rigid on y-axis). They are compared with the Curfit2D algorithm [16]. Significant changes in scores, especially for the CC index, suggest a better alignment of two chromatograms

Figure 3: Dataset 1: selection of the areas of interest (purple lines), new (left) and reference (right) chromatograms.
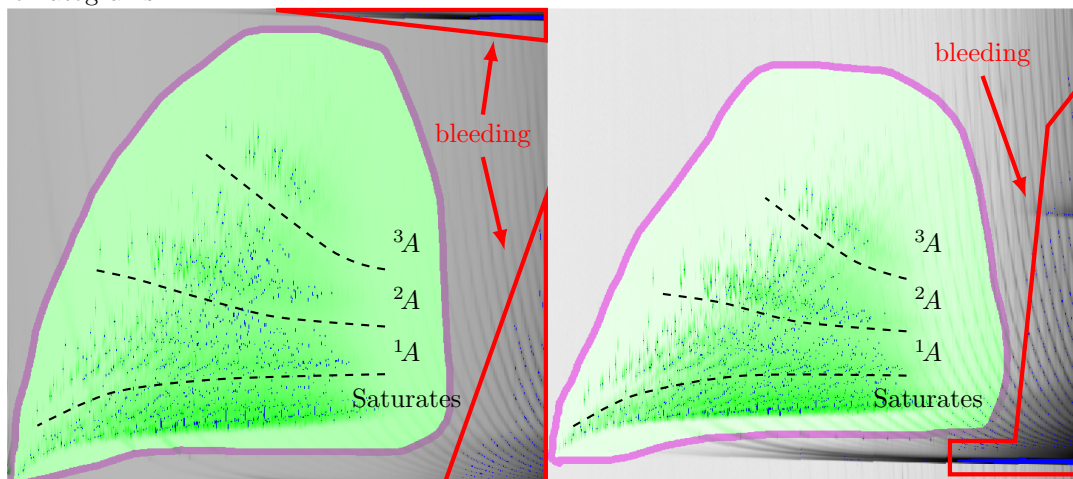

Table 1: Correlation coefficient (CC) and Similarity index (SSIM) on the two datasets.

		No registr.	Rigid [34]	Non-rigid [34]	Curfit2D [16]	BARCHAN
Dataset 1	CC	0.57	0.81	0.94	0.47	0.93
	SSIM	0.86	0.88	0.95	0.87	0.94
Dataset 2	CC	0.30	0.91	0.81	0.73	0.83
	SSIM	0.91	0.93	0.93	0.95	0.94

for dataset 1 with BARCHAN. However, small variations in these global indices demonstrate the need for a closer inspection of the results.

Figure 4 shows the optimization results for the three tested transformations on dataset 1 thanks to scatterplots [40]. Blue circles correspond to extracted peaks from the reference chromatogram whereas red crosses represent the extracted and transformed peaks for the new GC×GC chromatogram. These images show that a fully rigid transformation (Figure 4, top-right) does not allow a good match between the reference chromatogram and the new one. A better agreement is obtained with the BARCHAN algorithm and the fully non-rigid transformation. However, when looking into details in some specific areas of the 2D chromatogram where the number of extracted peaks is highly different between the reference and the new image (see red boxes at the bottom of Figure 4), BARCHAN algorithm outperforms the fully non-rigid approach. The interest of BARCHAN over the fully non-rigid approach is also shown on the transformation of template masks (see supplementary material). Whereas BARCHAN algorithm leads to a coherent transformation of the template mask including for blobs in the upper right part of the chromatogram which are extrapolated, the fully non-rigid deformation is not relevant.

To illustrate the changes modeled by BARCHAN on the chromatograms, the reference and the new chromatograms from dataset 1 are displayed in Figure 5, as well as the resulting aligned chromatogram.

The featural efficiency of the chromatogram alignment was evaluated from a more informative

Figure 4: Point sets before and after three different transformations: rigid, non-rigid and BARCHAN.

No transformation (centered cloud points)   Rigid

Non-rigid   BARCHAN
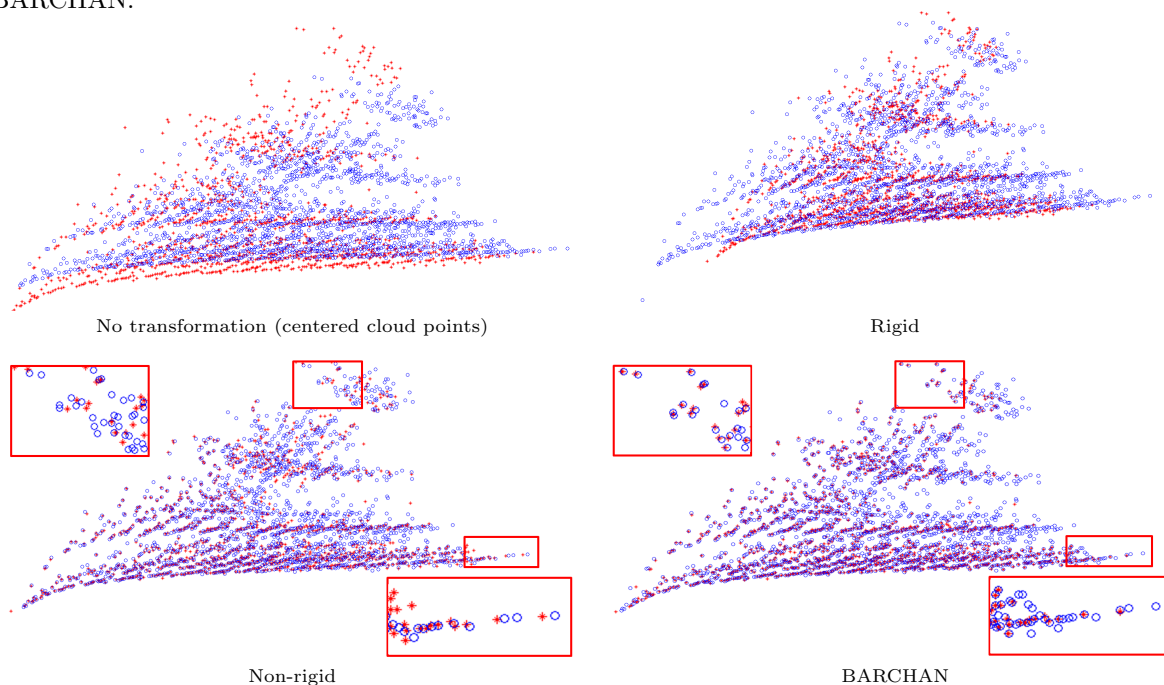
quantitative point of view on dataset 1. Three different ways of integrating the newly acquired chromatogram with 2DChrom® were tested: 1) an hundred-percent manual adjustment (MA) procedure during which the user has moved every point of the template reference mask to make it match with the new data with only simple local or global translation tools; 2) the sole application of BARCHAN alignment algorithm on the raw data and 3) the combination of BARCHAN with light manual editing. The modified mask, after transformation ② from flowchart in Figure 1 with BARCHAN, is displayed on Figure 6 for both datasets 1 and 2, together with the reference template mask on the reference analysis. Concerning dataset 1, it is clearly visible that the new analysis differs from the reference analysis despite the use of the same chromatographic conditions: the new data are slightly shifted to the left and 2D retention times are higher mainly due to lower elution temperatures from the 1D column. Realignment of the template mask however looks satisfactory, with an overall good match between the readjusted mask and the analysis. The same conclusions can be drawn for dataset 2 even if the changes between the reference chromatogram and the new one are huge as these data were not obtained with same type of modulation system. This tends to show that the algorithm is robust and is able to handle large deviations between reference and new data.

Results for the quantification of chemical families are reported in Table 2. These are compared with reference data previously obtained on GO-SR sample during an intra-laboratory reproducibility study on two different chromatographs with two different users so as to take into account both instrumental and user variability.

The BARCHAN transformation leads to coherent quantitative results for every chemical family

9

Figure 5: Calculated deformation of the 2D chromatogram with the BARCHAN algorithm on dataset 1.


| New. | Reference. | BARCHAN transformed. |
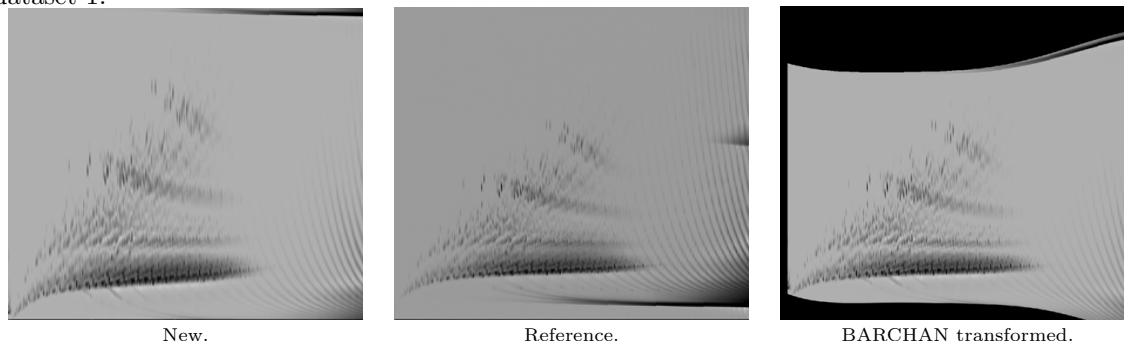
except for normal and iso-paraffins (n-C_nH_{2n+2} and i-C_nH_{2n+2} respectively) and to a lesser extent naphthenes (C_nH_{2n}). The quantification of n-paraffins is underestimated while iso-paraffins are overestimated because of slight misalignments of the template mask as depicted on the bottom of Figure 6 and on Figure 7. Indeed, some blobs identified in the reference mask as n-paraffins or naphthenes are only a few modulation periods wide as they correspond to single compounds. Small deviations in the alignment procedure impact the accurate quantification of these blobs. An additional manual fitting is therefore required to satisfactorily correct the transformed integration mask for these specific compounds. It consists in manually moving the blob points of these small blobs to make them perfectly match with the measured individual peaks. Movements are generally smaller than one or two pixels on the first dimension and minor on the second dimension. This overall procedure is typically applied to 20 to 40 blobs for a classical gas-oil template mask and requires a few minutes.

When looking at the data analysis time required to correctly apply a sophisticated template mask on a new chromatogram, the complexity of the GO-SR sample and of the complex mask with its 280 blobs implies several hours of work for an experienced user with a non-automated procedure. With an anchor point based approach, at the very least 50 similar points in both chromatograms would need to be defined. It would result into a processing time approaching one hour. In contrast, the processing time for dataset 1 was of two minutes, including the peak selection step. Nevertheless, depending of the samples complexity, their range of differences, and the quality of the chromatographic acquisition, the resulting masks may still require light post-processing modifications. In this case, we verified that defining typically five anchor points in an interactive registration post-processing step was enough to get a result as good as a fully manually operated one. Time saving is therefore still significant compared to manual procedures.

5   Conclusion

We present in this paper a 2D-chromatogram and template alignment named BARCHAN. It is based on three key ingredients: 1) a peak registration step which is performed on both the reference and the target 2D chromatograms; 2) two different types of transforms: a non-rigid one on the first chromatographic dimension and a rigid one on the second; 3) the use of the probabilistic Coherent Point Drift motion estimation strategy, that is proven to be robust to noise and outliers. It results

Figure 6: Chromatograms with overlaid masks. Top: reference and datasets 1 and 2 transformed with BARCHAN. Middle and bottom: zooms on the top (diaromatics) and bottom (paraffins and naphthenes) red areas.

Reference.   Dataset 1 ($h = 120$).   Dataset 2 ($h = 60$).
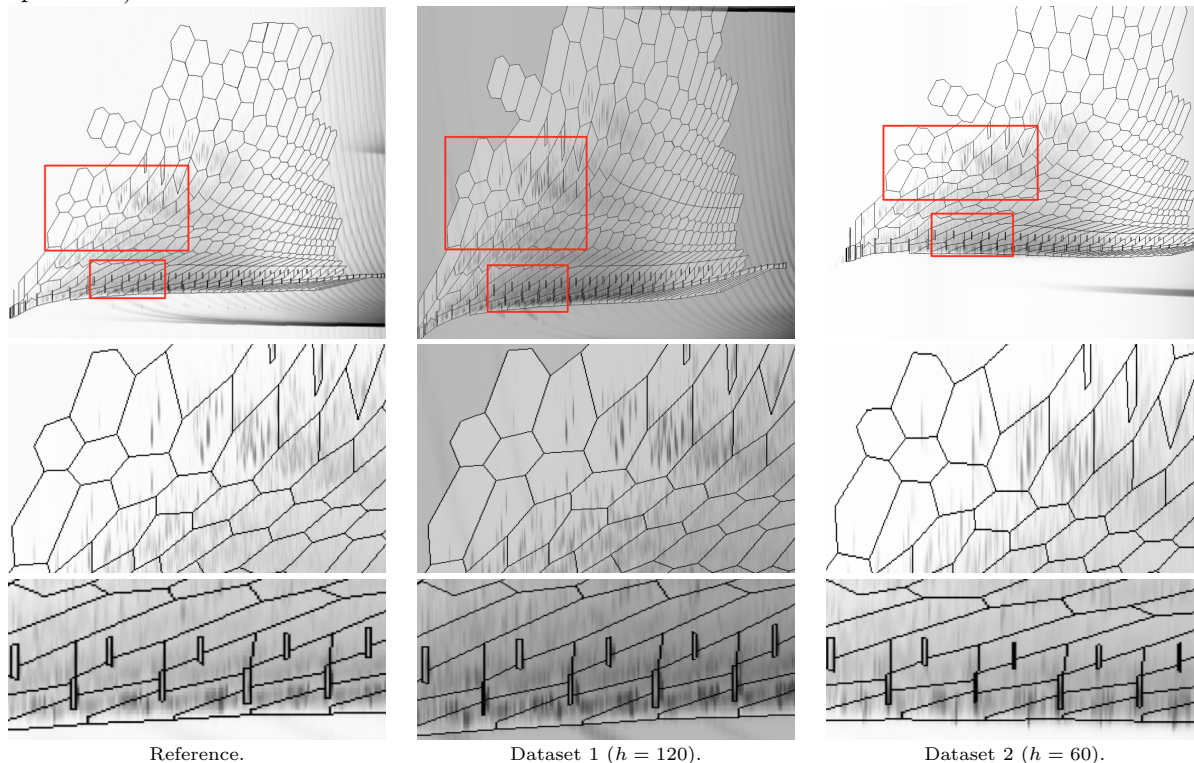
into an overall procedure that is an order of magnitude faster than the competing user-interactive alignment algorithms, with an accuracy as good as manual registration while guarantying a better reproducibility. This fast procedure may have a great interest when changing GC×GC configurations or when translating GC×GC−MS template masks on other GC×GC analysis (GC×GC−FID analysis for example). Finally, feature point selection may benefit from the Bayesian peak tracking recently proposed in [41].

Acknowledgments

References

[1] Z. Liu, J.B. Phillips, Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface, J. Chromatogr. Sci. 29 (6) (1991) 227–231, doi: \bibinfo{doi}{10.1093/chromsci/29.6.227}.

Table 2: Dataset 1: Comparison of chemical family quantification on with respect to manual adjustment (MA).

Chemical family	Weight percent			
	Reference	100% MA	BARCHAN	BARCHAN + MA
n-C_nH_{2n+2}	17.8 ± 1.0	17.2	12.6	17.2
i-C_nH_{2n+2}	25.9 ± 1.2	25.8	31.6	26.0
C_nH_{2n}	14.6 ± 0.9	14.7	13.9	14.9
C_nH_{2n-2}	11.6 ± 0.8	11.8	11.7	11.6
C_nH_{2n-6}	9.5 ± 0.7	9.5	9.4	9.5
C_nH_{2n-8}	5.0 ± 0.5	5.2	4.8	4.9
C_nH_{2n-10}	2.8 ± 0.4	2.8	2.8	2.8
C_nH_{2n-12}	6.6 ± 0.6	6.8	6.6	6.6
C_nH_{2n-14}	2.4 ± 0.4	2.4	2.4	2.4
C_nH_{2n-16}	1.6 ± 0.3	1.5	1.7	1.7
C_nH_{2n-18}	2.0 ± 0.3	2.0	2.1	2.1
C_nH_{2n-20}	0.12 ± 0.08	0.2	0.2	0.2
C_nH_{2n-22}	0.03 ± 0.04	0.0	0.1	0.1
Analysis time		$>4\,\mathrm{h}$	$<2\,\mathrm{min}$	$<30\,\mathrm{min}$

[2] M. Adahchour, J. Beens, U.A.T. Brinkman, Recent developments in the application of comprehensive two-dimensional gas chromatography, J. Chrom. A 1186 (1-2) (2008) 67–108 doi: \bibinfo{doi}{10.1016/j.chroma.2008.01.002}.

[3] C. Meinert, U.J. Meierhenrich, A New Dimension in Separation Science: Comprehensive Two-Dimensional Gas Chromatography, Angew. Chem. Int. Ed. 51 (42) (2012) 10460–10470, doi: \bibinfo{doi}{10.1002/anie.201200842}.

[4] J.V. Seeley, Recent advances in flow-controlled multidimensional gas chromatography, J. Chrom. A 1255 (2012) 24–37, doi:\bibinfo{doi}{10.1016/j.chroma.2012.01.027}.

[5] H.J. Cortes, B. Winniford, J. Luong, M. Pursch, Comprehensive two dimensional gas chromatography review, J. Sep. Sci. 32 (5-6) (2009) 883–904, doi:\bibinfo{doi}{10.1002/jssc.200800654}.

[6] C. Vendeuvre, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, D. Thiébaut, M.-C. Hennion, Characterisation of middle-distillates by comprehensive two-dimensional gas chromatography (GC × GC): A powerful alternative for performing various standard analysis of middle-distillates, J. Chrom. A 1086 (1-2) (2005) 21–28, doi:\bibinfo{doi}{DOI:10.1016/j.chroma.2005.05.106}.

[7] F. Bertoncini, M. Courtiade-Tholance, D. Thiébaut (Eds.), Gas chromatography and 2D-gas chromatography for petroleum industry. The race for selectivity, Éditions Technip, 2013.

[8] K.D. Nizio, T.M. McGinitie, J.J. Harynuk, Comprehensive multidimensional separations for the analysis of petroleum, J. Chrom. A 1255 (2012) 12–23, doi:\bibinfo{doi}{10.1016/j.chroma.2012.01.078}.

Figure 7: Misalignment for thin blobs on n-paraffins.
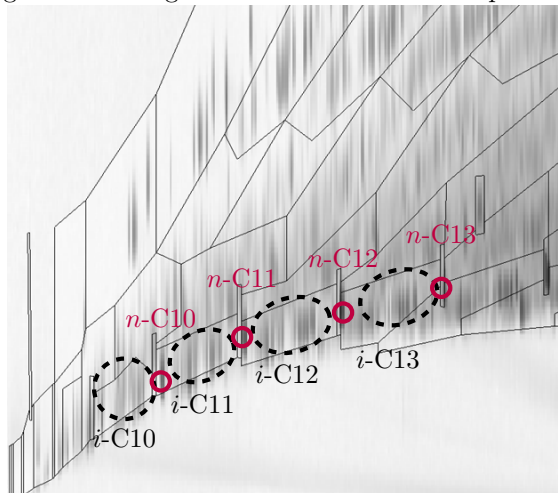
[9] M. Edwards, A. Mostafa, T. Górecki, Modulation in comprehensive two-dimensional gas chromatography: 20 years of innovation, Anal. Bioanal. Chem. 401 (8) (2011) 2335–2349, doi:\bibinfo{doi}{10.1007/s00216-011-5100-6}.

[10] J. Mommers, J. Knooren, Y. Mengerink, A. Wilbers, R. Vreuls, S. van der Wal, Retention time locking procedure for comprehensive two-dimensional gas chromatography, J. Chrom. A 1218 (21) (2011) 3159–3165, doi:\bibinfo{doi}{10.1016/j.chroma.2010.08.065}.

[11] C. Vendeuvre, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, D. Thiébaut, Comprehensive Two-Dimensional Gas Chromatography for Detailed Characterisation of Petroleum Products, Oil Gas Sci. Tech. 62 (1) (2007) 43–55, doi:\bibinfo{doi}{10.2516/ogst:2007004}.

[12] J.A. Murray, Qualitative and quantitative approaches in comprehensive two-dimensional gas chromatography, J. Chrom. A 1261 (2012) 58–68, doi:\bibinfo{doi}{10.1016/j.chroma.2012.05.012}.

[13] S.E. Reichenbach, X. Tian, C. Cordero, Q. Tao, Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography, J. Chrom. A 1226 (2012) 140–148, doi:\bibinfo{doi}{10.1016/j.chroma.2011.07.046}.

[14] Z. Zeng, J. Li, H.M. Hugel, G. Xu, P.J. Marriott, Interpretation of comprehensive two-dimensional gas chromatography data using advanced chemometrics, Trends Anal. Chem. 53 (2014) 150–166, doi:\bibinfo{doi}{10.1016/j.trac.2013.08.009}.

[15] J.F. Griffith, W.L. Winniford, K. Sun, R. Edam, J.C. Luong, A reversed-flow differential flow modulator for comprehensive two-dimensional gas chromatography, J. Chrom. A 1226 (2012) 116–123, doi:\bibinfo{doi}{10.1016/j.chroma.2011.11.036}.

[16] W.P.H. de Boer, J. Lankelma, Two-dimensional semi-parametric alignment of chromatograms, J. Chrom. A 1345 (2014) 193–199, doi:\bibinfo{doi}{10.1016/j.chroma.2014.04.034}.

[17] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image Quality Assessment: From Error Visibility to Structural Similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612, doi: \bibinfo{doi}{10.1109/TIP.2003.819861}.

[18] V.G. van Mispelaar, A.C. Tas, A.K. Smilde, P.J. Schoenmakers, A.C. van Asten, Quantitative analysis of target components by comprehensive two-dimensional gas chromatography, J. Chrom. A 1019 (1-2) (2003) 15–29, doi:\bibinfo{doi}{10.1016/j.chroma.2003.08.101}.

[19] K.M. Pierce, L.F. Wood, B.W. Wright, R.E. Synovec, A Comprehensive Two-Dimensional Retention Time Alignment Algorithm To Enhance Chemometric Analysis of Comprehensive Two-Dimensional Separation Data, Anal. Chem. 77 (23) (2005) 7735–7743, doi:\bibinfo{doi}{10.1021/ac0511142}.

[20] B.V. Hollingsworth, S.E. Reichenbach, Q. Tao, A. Visvanathan, Comparative visualization for comprehensive two-dimensional gas chromatography, J. Chrom. A 1105 (1–2) (2006) 51–58, doi:\bibinfo{doi}{10.1016/j.chroma.2005.11.074}.

[21] S.E. Reichenbach, D.W. Rempe, Q. Tao, D. Bressanello, E. Liberto, C. Bicchi, S. Balducci, C. Cordero, Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors, Anal. Chem. 87 (19) (2015) 10056–10063, doi:\bibinfo{doi}{10.1021/acs.analchem.5b02718}.

[22] B. Zitová, J. Flusser, Image registration methods: a survey, Image Vis. Comput. 21 (11) (2003) 977–1000, doi:\bibinfo{doi}{10.1016/S0262-8856(03)00137-9}.

[23] V.G. van Mispelaar, A.K. Smilde, O.E. de Noord, J. Blomberg, P.J. Schoenmakers, Classification of highly similar crude oils using data sets from comprehensive two-dimensional gas chromatography and multivariate techniques, J. Chrom. A 1096 (1-2) (2005) 156–164, doi:\bibinfo{doi}{10.1016/j.chroma.2005.09.063}.

[24] M. Ni, S.E. Reichenbach, A. Visvanathan, J. TerMaat, E.B. Ledford, Jr., Peak pattern variations related to comprehensive two-dimensional gas chromatography acquisition, J. Chrom. A 1086 (1–2) (2005) 165–170, doi:\bibinfo{doi}{10.1016/j.chroma.2005.06.033}.

[25] S.E. Reichenbach, P.W. Carr, D.R. Stoll, Q. Tao, Smart Templates for peak pattern matching with comprehensive two-dimensional liquid chromatography, J. Chrom. A 1216 (16) (2009) 3458–3466, doi:\bibinfo{doi}{10.1016/j.chroma.2008.09.058}.

[26] D. Zhang, X. Huang, F.E. Regnier, M. Zhang, Two-Dimensional Correlation Optimized Warping Algorithm for Aligning GC×GC–MS Data, Anal. Chem. 80 (8) (2008) 2664–2671, doi:\bibinfo{doi}{10.1021/ac7024317}.

[27] A.M. van Nederkassel, M. Daszykowski, P.H.C. Eilers, Y. Vander Heyden, A comparison of three algorithms for chromatograms alignment, J. Chrom. A 1118 (2) (2006) 199–210, doi:\bibinfo{doi}{DOI:10.1016/j.chroma.2006.03.114}.

[28] H. Parastar, M. Jalali-Heravi, R. Tauler, Comprehensive two-dimensional gas chromatography (GC–GC) retention time shift correction and modeling using bilinear peak alignment, correlation optimized shifting and multivariate curve resolution, Chemometr. Intell. Lab. Syst. 117 (2012) 80–91, doi:\bibinfo{doi}{10.1016/j.chemolab.2012.02.003}.

[29] J.J.A.M. Weusten, E.P.P.A. Derks, J.H.M. Mommers, S. van der Wal, Alignment and clustering strategies for GC×GC–MS features using a cylindrical mapping, Anal. Chim. Acta 726 (0) (2012) 9–21, doi:\bibinfo{doi}{10.1016/j.aca.2012.03.009}.

[30] B. Wang, A. Fang, J. Heim, B. Bogdanov, S. Pugh, M. Libardoni, X. Zhang, DISCO: Distance and Spectrum Correlation Optimization Alignment for Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry-Based Metabolomics, Anal. Chem. 82 (12) (2010) 5069–5081, doi:\bibinfo{doi}{10.1021/ac100064b}.

[31] B. Wang, A. Fang, X. Shi, S.H. Kim, X. Zhang, DISCO2: A Comprehensive Peak Alignment Algorithm for Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry, in: D.-S. Huang, Y. Gan, P. Premaratne, K. Han (Eds.), Bio-Inspired Computing and Applications, vol. 6840 of Lect. Notes Comput. Sci., Springer, 486–491, doi:\bibinfo{doi}{10.1007/978-3-642-24553-4_64}, 2012.

[32] S. Kim, I. Koo, A. Fang, X. Zhang, Smith-Waterman peak alignment for comprehensive two-dimensional gas chromatography-mass spectrometry, BMC Bioinformatics 12 (1) (2011) 235–245, doi:\bibinfo{doi}{10.1186/1471-2105-12-235}.

[33] J. Gros, D. Nabi, P. Dimitriou-Christidis, R. Rutler, J.S. Arey, Robust Algorithm for Aligning Two-Dimensional Chromatograms, Anal. Chem. (2012) 9033–9040, doi:\bibinfo{doi}{10.1021/ac301367s}.

[34] A. Myronenko, X. Song, Point Set Registration: Coherent Point Drift, IEEE Trans. Pattern Anal. Mach. Intell. 32 (12) (2010) 2262–2275, doi:\bibinfo{doi}{10.1109/tpami.2010.46}.

[35] X. Ning, I.W. Selesnick, L. Duval, Chromatogram baseline estimation and denoising using sparsity (BEADS), Chemometr. Intell. Lab. Syst. 139 (2014) 156–167, doi:\bibinfo{doi}{10.1016/j.chemolab.2014.09.014}.

[36] S. Samanipour, P. Dimitriou-Christidis, J. Gros, A. Grange, J.S. Arey, Analyte quantification with comprehensive two-dimensional gas chromatography: Assessment of methods for baseline correction, peak delineation, and matrix effect elimination for real samples, J. Chrom. A 1375 (2015) 123–139, doi:\bibinfo{doi}{10.1016/j.chroma.2014.11.049}.

[37] M.J. Fredriksson, P. Petersson, B.-O. Axelsson, D. Bylund, An automatic peak finding method for LC-MS data using Gaussian second derivative filtering, J. Sep. Sci. 32 (22) (2009) 3906–3918, doi:\bibinfo{doi}{10.1002/jssc.200900395}.

[38] A.L. Yuille, N.M. Grzywacz, A mathematical analysis of the motion coherence theory, Int. J. Comp. Vis. 3 (2) (1989) 155–175, doi:\bibinfo{doi}{10.1007/bf00126430}.

[39] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1) (1977) 1–38.

[40] F. J. Anscombe, Graphs in Statistical Analysis, Am. Stat. 27 (1973) 17–21.

[41] A. Barcaru, E. Derks, G. Vivó-Truyols, Bayesian peak tracking: a novel probabilistic approach to match GCxGC chromatograms, Anal. Chim. Acta 940 (2016) 46–55, doi:\bibinfo{doi}{10.1016/j.aca.2016.09.001}.