Journal of Chromatography A, 1218 (2011) 2301–2310

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Simultaneous deconvolution and re-construction of primary and secondary overlapping peak clusters in comprehensive two-dimensional gas chromatography

Zhong-Da Zeng[a,b], Sung-Tong Chin[a], Helmut M. Hugel[b], Philip J. Marriott[a,*]

[a] Centre for Green Chemistry, School of Chemistry, Monash University, Wellington Rd, Clayton 3800, Australia
[b] School of Applied Sciences, RMIT University, G.P.O. Box 2476, Melbourne 3001, Australia

ARTICLE INFO

Article history:
Received 6 December 2010
Received in revised form 11 February 2011
Accepted 14 February 2011
Available online 21 February 2011

Keywords:
Comprehensive two-dimensional gas chromatography
GC × GC dataset
Deconvolution
Quantification
Non-linear least square curve fitting
Peak reconstruction

ABSTRACT

In this study, simultaneous deconvolution and reconstruction of peak profiles in the first (^{1}D) and second dimension (^{2}D) of comprehensive two-dimensional (2D) gas chromatography (GC × GC) is achieved on the basis of the property of this new type of instrumental data. First, selective information, where only one component contributes to the peak elution window of a given modulation event, is employed for stepwise stripping of each ^{2}D peak with the help of pure components corresponding to that compound from the neighbouring modulations. Simulation based on an exponentially modified Gaussian (EMG) model aids this process, where the EMG represents the envelope of all ^{2}D peaks for that compound. The peak parameters can be restricted by knowledge of the pure modulated ^{2}D GC peaks derived from the same primary compound, since it is modulated into several fractions during the trapping and re-focusing process of the cryogenic modulation system according to the modulation period. Next, relative areas of all pure ^{2}D components of that compound are considered for reconstruction of the primary peak. This strategy of exploitation of the additional information provided by the second dimension of separation allows effective deconvolution of GC × GC datasets. Non-linear least squares curve fitting (NLLSCF) allows the resolved 2D chromatograms to be recovered. Accurate acquisition of the pure profiles in both ^{1}D and ^{2}D aids quantification of compositions and prediction of 2D retention parameters, which are of interest for qualitative and quantitative analysis. The ratio between the sum of squares of deconvolution residual and original peak response (R_{rr}) is employed as an effective index to evaluate the resolution results. In this work, simulated and experimental examples are used to develop and test the proposed approach. Satisfactory performance for these studies is validated by minimum and maximum R_{rr} values of 1.34e−7% and 1.09e−2%; and 1.0e−3% and 3.0e−1% for deconvolution of ^{1}D and ^{2}D peaks, respectively. Results suggest that the present technique is suitable for GC × GC data processing.

1. Introduction

One-dimensional gas chromatography (1D GC) has been the instrumental pillar for separation of both complicated and simple mixtures of volatile compounds [1,2]. With the increasing demands for the analysis of systems with hundreds or even thousands of chemical components, this technique becomes indispensable in both academic and industrial areas. However, the limits of a 1D GC system are readily exceeded as complexity increases [3]. Fortunately, hyphenated chromatographic instrumentation has provided powerful solutions to problems in recently popularised fields e.g. metabonomics [4–6] allowing qualitative and quantitative information of target molecules in biofluid samples to be further used to interpret the changes in metabolism processes. Thus, in recent years reliable measurement of an increasing number of constituents with a greater total number of identified components, at ever decreasing abundance becomes important. For this task, contaminated peak clusters with a greater number of overlapping analytes require effective resolution [7].

Comprehensive two-dimensional gas chromatography (GC × GC) improves analytical peak capacity and separation effectiveness through coupling of two mutually 'orthogonal' columns [8] and its performance has been confirmed through the analysis of volatile components in complex samples, including traditional Chinese medicine (TCM), wine, coffee, drugs, and others [9–12]. However, complete separation of all detectable components still cannot be attained because of the high complexity of

* Corresponding author. Tel.: +61 3 99059630; fax: +61 3 99058501.
E-mail address: Philip.Marriott@monash.edu (P.J. Marriott).

0021-9673/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2011.02.028

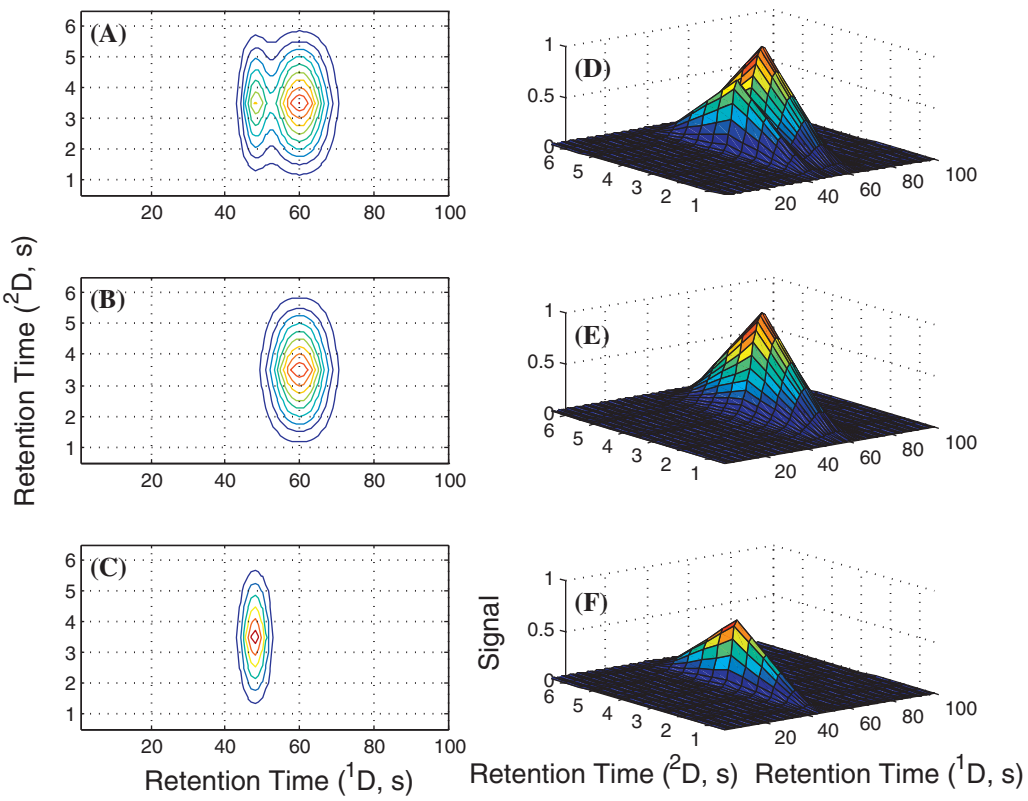

Fig. 1. Illustration of the data structure of comprehensive two-dimensional chromatography, and deconvolution and reconstruction of both primary and secondary chromatographic profiles. (A and D) Contour and three-dimensional graphs of overlapping peak cluster of two simulated components. (B and C) and (E and F) Contour and three-dimensional graphs of the two individual pure simulated components shown in (A) and (D).

real samples, and limitations of the experimental conditions and the instrument [13,14].

Conventional methods for accurate quantification of 1D overlapping peak clusters have been established on the basis of exploitation of different features and models of chromatographic profiles, as well as data transform techniques such as Kalman filter and wavelet analysis [15–19]. For example, Jung et al. extracted four or five data points with the same interval from both the normal and the derivative chromatograms; the four parameters in an exponentially modified Gaussian (EMG) model can then be reconstructed via solving the cubic or quartic equation [20,21]. A study was proposed for automatic deconvolution using a polynomially modified Gaussian (PMG) function, and reasonable selection of the local and global optimization algorithms, including locally optimized genetic algorithm (LOGA), multi-start local search (MSLS), and Powell algorithm; the improved version was found to be effective for the extraction of peak profiles of pure components from overlapping clusters [22]. The result has special relevance to the user for chromatographic data processing, in the absence of strong background response. Generally speaking, most of the reported works on deconvolution of 1D chromatograms seek to recover the pure peaks using full optimization, fitting or searching techniques with different mathematical models. These models include normal and asymmetric Gaussian distribution, generalized exponential function, PMG, EMG, and some modified functions [23].

In contrast to conventional 1D chromatography, two-dimensional (2D) hyphenated chromatographic datasets comprise two separation/identification dimensions, and when combined with spectral information with multichannel detectors offer additional information content. Thus, deconvolution methods, such as iterative optimization, use of selective information, and key variable selection, were developed on the basis of making full use of such information [24–26]. Comprehensive 2D chromatography (C2DC) has special data structure characteristics compared with both conventional 1D and coupled 2D chromatography. Peaks eluting from the first column (^{1}D) are modulated into several fractions and re-injected to the second dimension (^{2}D) [7–12]. It should be interesting and effective if the additional information found from both ^{1}D and ^{2}D can be simultaneously extracted, deconvoluted, and the pure profiles reconstructed for ^{1}D and ^{2}D overlapping peak clusters.

So far, few methods have been specially developed for deconvolution of C2DC datasets. For example, Fraga and Corley combined generalized rank annihilation method (GRAM) and parallel factor analysis (PARAFAC) for chemometric resolution and quantification of overlapped peaks in comprehensive two-dimensional liquid chromatography (LC × LC) [13]. Kong et al. recovered the primary peak profiles using the areas of each fraction acquired from the second dimension of GC × GC [14]. But the strict tri-linear property or complete parameter searching makes these methods difficult for widespread and rapid applications.

The present study attempts to solve the deconvolution issue for both ^{1}D and ^{2}D simultaneously, based on recognition of data structure features. It fully utilizes the information of C2DC. First, the selective information found in ^{1}D is employed to determine the optimization bounds of these same components in ^{2}D for modeling. This has relevance for rapid and accurate determination of the parameters in the EMG model. Then, selective information of the pure component in ^{2}D is employed for deconvolution, using certain boundary information. After extraction of all the peak areas of each pure secondary component, reconstruction and deconvolution of overlapping primary peak profiles in ^{1}D can be achieved using the Levenberg–Marquardt algorithm [27,28]. The main advantage of this work is full utility and exploitation of the new information for simultaneous 2D deconvolution in terms of the data characteristics of C2DC. Such a strategy improves the potential to generate the

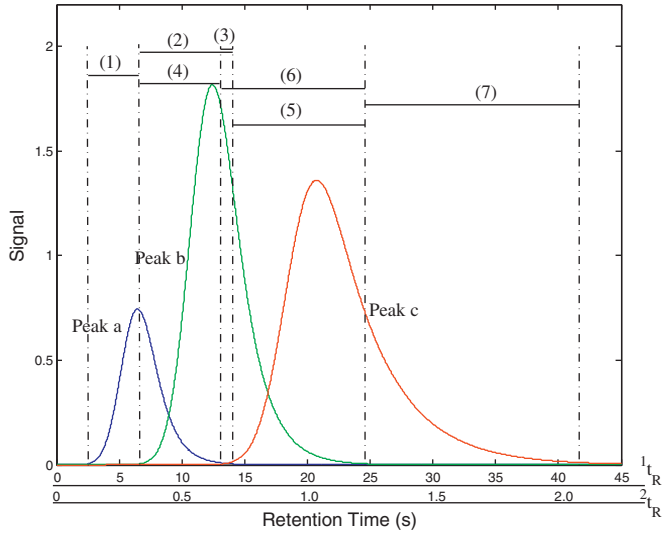

Fig. 2. Principle of deconvolution and reconstruction of primary and secondary peak profiles. Fractions modulated in the two regions (1) and (7) will produce pure ^{2}D peaks (termed selective regions). The six regions (1 and 2), (4 and 5), (6 and 7) represent the left and right elution window of peaks a, b and c. Zone (3) is an overlapping window with the three components. The three peaks here could represent either a 1D chromatogram, or ^{2}D chromatographic data, as shown on the retention scale. The denotations ^1t_R and ^2t_R, respectively, represent the retention times in ^{1}D and ^{2}D with different scales.

requisite results using appropriate fitting techniques for deconvolution of overlapping peaks.

2. Theory

The EMG function is the most successful and widely-used model to simulate real chromatographic peaks [20]. It includes four parameters: the peak magnitude, position, shape and skewness of the target profile, as given in Eq. (1),

$$\mathbf{s}(t) = \frac{A}{2\times\tau}\left(\exp\left(\frac{\delta}{2\times\tau^2} - \frac{t-t_G}{\tau}\right)\left(1 - erf\left(\frac{\delta}{\sqrt{2}\times\tau} - \frac{t-t_G}{\sqrt{2}\times\delta}\right)\right)\right) \quad (1)$$

where $\mathbf{s}(t)$ is a response vector of the chromatographic profile, and denotations A, δ, τ, and t_G, respectively, represents the peak area, the standard deviation of the precursor Gaussian component, time constant of the precursor exponential modifier, and the retention time of the precursor Gaussian peak. These terms may be denoted with superscripts to define the ^{1}D and ^{2}D parameters, respectively. The expression *erf* is an abbreviation of error function. The special case with $\tau = 0$ is the symmetrical Gaussian profile. In this work, EMG model is employed to reconstruct both the ^{1}D and the ^{2}D chromatograms of GC × GC separation.

The importance of deconvolution of overlapping peak clusters and reconstruction of pure chromatographic profiles is represented in Fig. 1. The first three contour plots (Fig. 1(A)–(C)) show overlap profiles of two individual pure components. It may be difficult to accurately determine quantitative information of the two components from Fig. 1(A), and for cases of peak overlap to obtain individual peak parameters such as retention time, peak widths, and other related information for the 1D peak. The visual three-dimensional graphs from (D) to (F) further illustrate the data structure of the mixed and pure peak profiles.

In Fig. 2, a typical overlapping peak cluster with three components is shown to demonstrate the deconvolution strategy. According to the principle of modulation in GC × GC analysis, the primary peak eluting from ^{1}D is collected and focused in time e.g. with a cryogenic modulation system, and then injected to ^{2}D for further separation. Thus, a given mass of each fraction of the ^{1}D chromatographic band will be transferred to the corresponding ^{2}D peak, as Eq. (2) defines [29,30].

$$\sum_{i=1}^{i} A_i = \int_1^{t_0+(i-1)\times P_M} \mathbf{s}(t)dt, \quad i = 1, 2, \ldots n_M, \quad (2)$$

where expressions A_i and $\mathbf{s}(t)$ represent the area of the ith fraction and signal of the secondary profile detected in ^{2}D, respectively. The numbers t_0, P_M and n_M denote the heart-cutting position of the first modulated fraction, the modulation period and the number of modulations detected for a given ^{1}D peak. Using non-linear least squares curve fitting (NLLSCF) techniques, the ^{1}D peak profile may be reconstructed to give effective deconvolution of overlapping peak clusters, provided the profile $\mathbf{s}(t)$ of all secondary components can be extracted in ^{2}D. Otherwise, deconvolution of overlapping ^{2}D peaks need to be accomplished first before reconstruction of the primary profile. In this work, Levenberg–Marquardt algorithm is used for NLLSCF analysis because of its satisfactory performance to treat multi-parameter systems, improving upon algorithms such as those based on immunity and genetics [16]. Additionally, discrimination of all peaks in the second dimension of the corresponding primary components with overlapping peaks may be difficult because of shifts of retention times. Correction of the shift along with recognition of the characteristics of GC × GC datasets improves the accuracy of ^{2}D peak assignment to individual ^{1}D peaks.

As shown in Fig. 2, if the simulated system denotes a ^{1}D overlapping chromatographic cluster, the modulation fractions at the selective region, zone (1) and zone (7), will produce chromatograms of pure fully resolved components in ^{2}D. A selective region of the peak profile means that the elution window has contribution from only one component. Ideally, other ^{2}D peaks of the same chemical component should have the same parameters $^2\delta$, $^2\tau$, and ^2t_G given in Eq. (1), but they will have different peak areas A. Then, deconvolution of the secondary overlapping clusters can be readily attained with the three known parameters. Though variations in experimental parameters may make this difficult to achieve because of the effects of noise, overloading, and disturbance to the distribution arising from temperature and others conditions, this prior information is very helpful to determine the search bounds when NLLSCF is utilized to extract the peak profile of pure ^{2}D components. The search bounds correspond to the range of potential existing values of the target parameters, namely, A, δ, τ, and t_G for which specific values are to be determined for a given peak in this work. The respective values should be included in the upper and lower bounds window. The lower bound value for A will be from the area of a selective region of the compound (which corresponds to a single modulated peak – the area must be larger than this value) and the total area of the whole peak cluster for deconvolution. For the other parameters reasonable values are estimated, and lower/upper bounds (e.g. 0.5–2.0 of this value) are then defined. Theoretically, the optimization procedure has to scan the whole mathematical space if the search bounds cannot be defined before starting the computation. This may make the results go into local optimum. In general, the wider the search bound, the more difficult to get the optimization results, and the narrower bound, the higher precision, as long as it includes the real values inside the boundary limits in terms of the property of Levenberg–Marquardt method. These parameters can be acquired from other zones e.g. zone (3), provided they are completely resolved.

Many model or empirical equations have been developed to extract band broadening parameters of the EMG function [31–33]. In this work, robust formulae are employed for the coefficients of known pure peak profiles, and then used as inputs for modeling [33]. Of course, the pure information of ^{2}D peaks cannot be exactly found in some cases since the selective region of target compo-

nents may be smaller than a modulation period, or may be due to overlapping with the neighbouring components in each modulation event. Then, the empirical formulae to define the relationship between parameters δ and τ can be employed to estimate the search bounds. In Eq. (3), a widely-used expression is given with τ/δ between 0.5 and 6.5 [34,35]. The knowledge and experience about the chromatogram is useful to find the optimum upper and lower thresholds.

$$w_{1/2} = 2.164 \times \delta + 0.718 \times \tau, \quad (3)$$

where denotation $w_{1/2}$ represents the peak width at half height of the target peak, and δ and τ have the same meaning as given in Eq. (1).

If the system in Fig. 2 represents an overlapping peak cluster in ^{2}D, the selective regions zone (1) and zone (7) can be utilized to fit the profiles of pure components a and c with the help of the parameter bounds acquired from the pure component regions in ^{1}D. According to the definition given in Eq. (4), contribution of the resolved profiles can be subtracted from the total 'contaminated' response **x**. Thus, new selective information can then be discovered for deconvolution of other components. As shown in Fig. 2, the left and right selective regions, zone (4) and zone (5), of peak b will appear after extracting the information of peak a or c. It is found that such a strategy can be used to deconvolute complicated profiles with more than 3 peaks as long as sufficient selective regions are available for the target compositions [21]. This stepwise component stripping technique simplifies the system and improves the effectiveness of deconvolution.

$$\mathbf{x} = \sum_{i=1}^{N} \mathbf{c}_i + \mathbf{e}, \quad (4)$$

where denotations **x** and **e** represent the total chromatogram and the corresponding experimental noise, respectively. N is the total number of components with contribution to response **x**.

To the system given in Fig. 2 with three components a, b and c, Eq. (4) can be re-written as Eq. (5).

$$\mathbf{x} = \mathbf{c}_a + \mathbf{c}_b + \mathbf{c}_c + \mathbf{ce}, \quad (5)$$

where $\mathbf{c}_a$, $\mathbf{c}_b$ and $\mathbf{c}_c$ represent the profiles of pure components a, b and c.

The derivatives of the smoothed chromatogram include rich information of the purity of the contaminated profiles with more than one chemical component. Thus, the number of underlying compounds and the corresponding elution ranges of a given overlapping peak cluster can be estimated using first-, second- and third-order derivatives [36]. The selective information also can be assessed in terms of the peak distribution and skewness as recognized by the user with some experience. If unsuccessful, a visual assessment strategy can assist to find acceptable results using the deconvolution residual as an evaluation index. Mathematically, 4 data points of a pure component are sufficient to recover the original chromatographic profile since there are only 4 unknown parameters to be solved, i.e. A, δ, τ, and t_G of the second dimension peak, and including t_0 of the primary peak. However, more selective information is required for model establishment because of interference of experimental noise, possible serious deviation to the EMG function, limitations of the mathematical method, and others. Then, the chromatographic profile of the pure component is fitted using NLLSCF technique with appropriate selective information. Further, prediction of such information derived from both the left and then the right side of the chromatogram is a good strategy to validate the results, since either approach should essentially represent the same pure component. Deconvolution of all the peaks of interest can be achieved based on stepwise stripping analysis. The areas of the resolved pure chromatograms are then utilized as input to obtain the primary peaks.

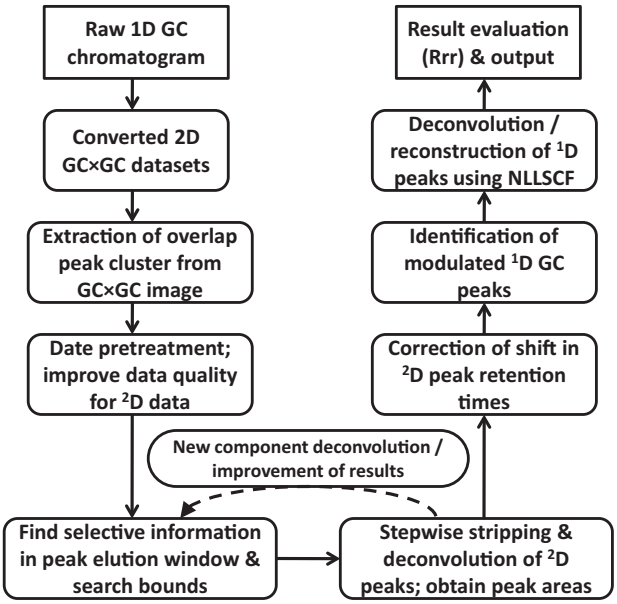

Fig. 3. Illustration of the work flow of the proposed method for simultaneous deconvolution of primary and secondary overlapping peak clusters.

Fortunately, 2D characteristics of GC × GC separations largely reduce the complexity of overlapping peaks, and relatively improve the possibility to find enough selective data points in the component elution window. Deconvolution of the overlapping components can then be attained after extraction of the pure secondary profiles of each of the components. Accurate determination of peak areas of the ^{2}D peaks improves the reconstruction of the primary peak profiles. In ref. 14, a method was introduced to find the linear relationship of parameters δ, τ, t_G and retention time t_R of the components in an experimental run. Such a strategy may be helpful for deconvolution when fewer modulations (n_M) are acquired for components with weak response.

In Fig. 3, a flow diagram is given to illustrate the working procedures of the method proposed in this work. The first four steps include obtainment of two-dimensional GC × GC datasets to be deconvoluted from the raw instrumental chromatogram, data pre-treatment and quality improvement such as background subtraction and noise smoothing. The next two steps are developed to extract the pure chromatographic profiles on the basis of the stepwise component stripping technique and NLLSCF method. These pertain to ^{2}D peaks. The remaining steps focus on the deconvolution and reconstruction of the ^{1}D chromatographic peaks of the GC × GC separation. The areas of the fractions obtained from the previous processes are finally utilized to simulate the primary peak data.

The present strategy takes advantage of the property of the GC × GC profile in both ^{1}D and ^{2}D. It fully mines out the data characteristics of this new type of dataset. Unlike conventional GC with only 1D chromatographic information, or coupled chromatography with spectral information, modulation of each peak from ^{1}D to ^{2}D in GC × GC provides new features for chromatographic peaks. This characteristic has clear significance required for complete understanding of GC × GC datasets. Simultaneous deconvolution and reconstruction of overlapping peaks from the two dimensions of GC × GC will assist identification and quantification of complex mixtures, since retention parameters and peak areas are fundamental to accurate qualitative and quantitative analysis. Residual response magnitude between raw signal or area and EMG

modeling results of ^{1}D and ^{2}D peak profiles is utilized to evaluate the performance of the present method.

3. Experimental

Simulated and real datasets are employed to validate the performance of the proposed method. The simulated overlapping peak cluster is that given in Fig. 2, and the real mixture is prepared using 61 standard components reported below.

3.1. Materials and preparation of the experimental mixture

The standard mixture contains 4 alkenes, 3 phenols, 16 alcohols, 15 esters, 6 acids, 7 aldehydes, 3 ketones, and 7 other compounds including furan and sulfides. All the components were chromatographically pure standard substances. All components were present in the final solution at 50 mg/L. Details of these components are not given here, since the experimental goal is to generate overlapping peaks, and then utilize the result to deliver examples of the strategy of chemometric interpretation.

3.2. GC × GC system configuration

The two-dimensional GC system comprised two parallel upstream and down-stream CO_2 nozzles for sample trapping and re-focusing. The cryogenic segment was installed at the beginning of the ^{2}D short, narrow diameter column. The cooling gas is alternately delivered from the two nozzles to sub-sample the peaks eluting from the first column into different fractions, and re-injected to the second 'orthogonal' column. The number of fractions depends on the value of modulation ratio (M_R), which simultaneously considers P_M and peak width at base (w_b) of the target peaks [37]. The instrumental system comprised an Agilent 6890 GC (Agilent Technologies, Nunawading, Australia) and a model 7683 Series auto-sampler, with FID detection. The column set consisted of different stationary phases, namely, a 30 m × 250 μm I.D. × 0.25 μm film thickness (d_f) RXi-5Sil MS (5% diphenyl 95% methyl phase, Restek Corp, Bellefonte, PA) ^{1}D column connected to a 1 m × 100 μm I.D. × 0.1 μm d_f Stabilwax (polyethylene glycol, Restek Corp) ^{2}D column. The secondary peaks are directly detected by the FID detector system. The maximum allowable temperature of the column set is 250 °C.

3.3. Experimental conditions for GC × GC analysis

The experimental conditions were applied as follows: the initial oven temperature was set at 40 °C and held for 3 min, programmed to 220 °C at 3 °C/min, and held for 10 min at the final temperature. Hydrogen was used as carrier gas with a flow rate at 1.5 mL/min. The modulation period (P_M) was 4 s; modulation start time was 2 min. The first cryogenic nozzle has 1 s for primary peak heart-cutting, and 3 s for trapping. The injection volume was 1 μL with splitless mode. The data acquisition rate of FID detection was 100 Hz.

3.4. Data transfer and analysis

The original dataset is firstly exported from the Agilent workstation, then transferred to matrix format using 2DGC Converter software (v1.0) coded in-house. All computer programs involved in this study were coded in MATLAB environment (version 7.7.0.471, R2008b). The calculations were performed on a HP compatible personal computer with Intel(R) Core(TM)2 Duo CPU and 1.95 GB RAM memory.

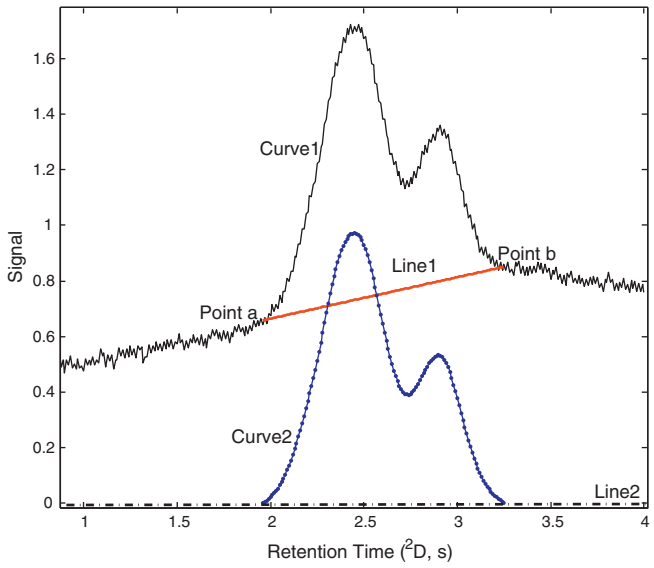

Fig. 4. Illustration to improve data quality for ^{2}D deconvolution and accurate obtainment of peak area. The noisy raw profile Curve1 is improved to Curve2 using smoothing and background subtraction techniques. Data points a and b are selected to construct the background (drift) of the target peak cluster, and Line1 is the mathematical baseline with zero as reference.

4. Results and discussion

4.1. Data pre-treatment and quality improvement

The dataset obtained from the GC × GC instrument had a strong baseline shift over the data range. In order to calculate the peak area directly using the raw data, the first step is to normalize the original data with a data centring operation, where the response of each data point is subtracted by the minimum of the whole dataset. Correct target peak areas of the secondary profiles are required to guarantee good performance in reconstructing the primary chromatograms. Thus, correction of background shift is further conducted in the analyte window using a linear interpolation technique. The peak background $\mathbf{x}_{backg}$ is linearly simulated using the first and last two smoothed data points. It also can be determined manually by the researcher in terms of GC × GC experimental datasets. As given in Eq. (6), the new data $\mathbf{x}_{new}$ without background interference (drift) is employed for analysis.

$$\mathbf{x}_{new} = \mathbf{x} - \mathbf{x}_{backg} \quad (6)$$

A moving window average method is utilized to smooth the noisy data set with a window size of 7 [38], which is important for ^{2}D peaks corresponding to the first or last fraction with small but noisy response of the primary peak.

In Fig. 4, a typical dataset with background shift and noise interference is shown. Each secondary peak acquired from ^{2}D is treated with the same importance, since they equally determine the modulation number (n_M) and time of the ^{1}D peak, and aid reconstruction of the ^{1}D peak profile. However, the first and/or the last modulated peak in ^{2}D have relatively low response because they are remote from the peak centre. Thus, quality improvement of these datasets is necessary to obtain the correct peak areas and modulation information of the ^{1}D peak. Curve 1 in Fig. 4 is the original chromatogram. It is difficult to directly find accurate peak areas of the overlapping peak cluster with the existence of strong background and experimental noise. In this case, data points a and b are selected as the starting and end positions of background of the peak cluster of interest. The background is then linearly interpolated to construct $\mathbf{x}_{backg}$ in Eq. (6), as Line1 shows. Curve 2 is the new

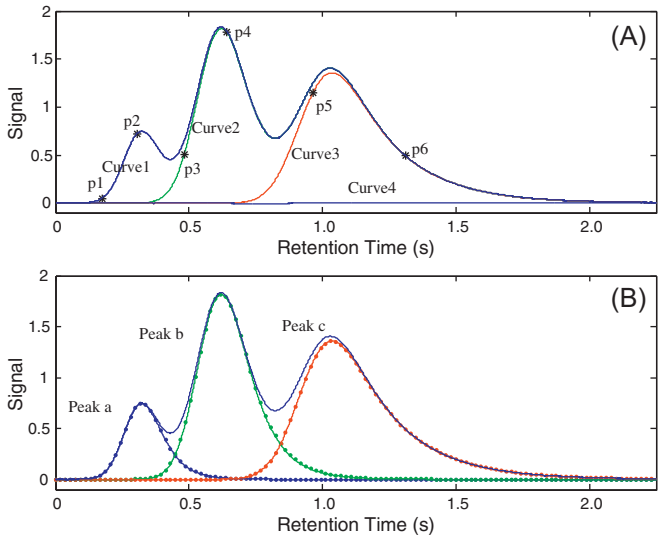

Fig. 5. Deconvolution of simulated overlapping peak cluster to deliver the proposed strategy. Deconvolution of the three components is from the left to right. (A) Curve1 to Curve4 show the changing profiles of the total chromatograms after progressive deconvolution. Curve1 is the original chromatogram; Curve4 is the final residual used as an index to evaluate the results. The asterisked symbols (*) show point-pairs p1 and p2, p3 and p4, p5 and p6 that determine the peak elution windows with contribution of only one target component for deconvolution, respectively. Curve1 can then be stripped from the total response, then Curve2, etc. (B) Comparison of the deconvolution results and the original peak profiles. The solid lines represent the simulated total chromatogram response and the dotted lines represent the deconvolution results.

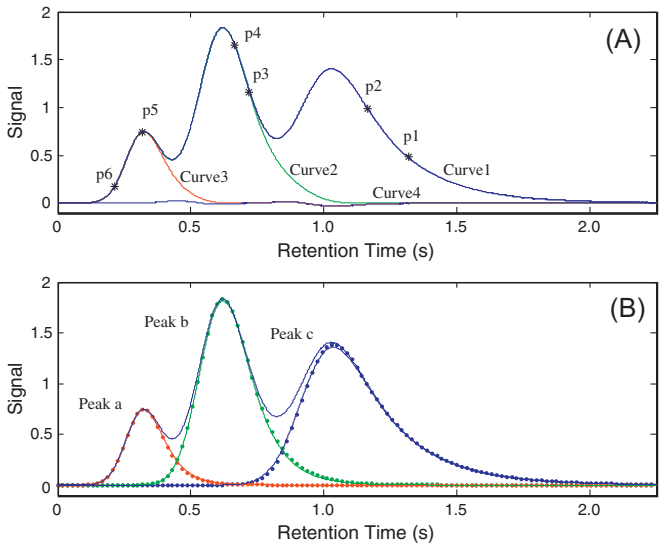

Fig. 6. The deconvolution process of the same data set shown in Fig. 5. Here, deconvolution is from right to left side for the peak cluster. This can be utilized to self-validate the final deconvolution results; (A) and (B) have equivalent meaning to Fig. 5(A) and (B), respectively.

dataset used for deconvolution and area assignment of ^{2}D peaks to recover the primary profile in ^{1}D. The smoothed profile, without background contamination is a satisfactory representation; Line2 is the mathematical zero line.

4.2. Results of simulated dataset with three components

The simulated dataset is utilized to introduce the strategy for deconvolution, and also to identify the advantages and disadvantages of the method, i.e. to assess the effectiveness of the proposed technique. Generally, the complexity of an overlapping peak cluster in GC × GC is significantly reduced compared with 1D GC, both with respect to the number of overlapping components and degree of contamination with neighbouring components. In the theory section, it is mentioned that deconvolution of components can be attained and self-validated through employing information from both the left and the right sides of the peak cluster. It is possible to also commence the deconvolution process using a middle component as long as it is a selected component i.e. a peak with pure regions (no overlap). Thus, the simulated data are analysed in this way to validate the resolution results. Only homoscedastic noise is studied in this work because of the low frequency appearance of heteroscedastic noise in GC × GC datasets. The noise level is simulated at 0.1% of the maximum response of the whole chromatographic profile.

In Fig. 5, results are shown using the proposed method. The curves from 1 to 4 in Fig. 5(A) clearly show the changing process from the original chromatogram (Curve1) to the final residual (Curve4) after deconvolution. The two curves 3 and 4 are the remaining profiles after stripping the left two components, respectively. The asterisked data points from 1 to 6 determine the starting and end positions of selective regions of the three target peaks – i.e. Curve1 from p1 to p2 has no underlying interfering peak, so is a pure peak. Once Curve1 is stripped from the total response, this leaves a result where Curve2 is a pure selective region from p3 to p4. Fig. 5(B) is the comparative result of simulation (solid line) and deconvolution (dotted line) profiles, which match very well. The ratio between the sum of squares of deconvolution residual and original peak response (R_{rr}) is 1.16e−3%, which means most of the signal is expressed using the NLLSCF technique. In Fig. 6, the dataset is resolved commencing from the right side (later component). The symbols have the same meaning as those in Fig. 5. The ratio (R_{rr}) of deconvolution residual and peak response is 1.83e−2% in this case.

Propagation of error has possible effects on deconvolution of complex overlapping systems with more than three components. The resulting deviation from the prior deconvolution process will partly affect subsequent deconvolution of the next component. Careful selection of regions with high response of target components can significantly help to improve the accuracy; this is not always possible with real data.

4.3. Results of three real datasets with different complexities

The first treatment of these experimental data was correction of the baseline using the above approach. After datasets with good quality are obtained, deconvolution of overlapping components in ^{2}D is conducted. Then reconstruction and deconvolution of the ^{1}D peak profile can be achieved using NLLSCF technique. The function for NLLSCF in Matlab is effective to define the search upper and lower bounds, making the method for deconvolution proposed in this work readily performed.

In Fig. 7, an experimental dataset with two partially overlapping peaks in ^{2}D is given as an example. It can be seen that there are two components over this elution window, as shown in Fig. 7(A) and (B). The image shown in Fig. 7(A) is acquired from the peak profiles given in Fig. 7(B), which provides the distribution information of the sub-peaks found in ^{2}D. In most cases, some relative retention shifts will arise for the peaks extracted from ^{2}D. This leads to difficulty in assignment of the overlapping ^{1}D peak clusters to individual components. This has been referred to elsewhere for GC × GC data [39]. Alignment of data for 1D [40,41] and also GC × GC data has been reported [42,43]. Evaluation of the shift in ^{2}D peaks is aided by the stepwise attenuation property of the EMG chromatogram. The partially resolved peak cluster a4 + b1 includes contribution of both components modulated from the two primary peaks for which the ^{2}D column does not provide adequate resolution. Fig. 8 shows

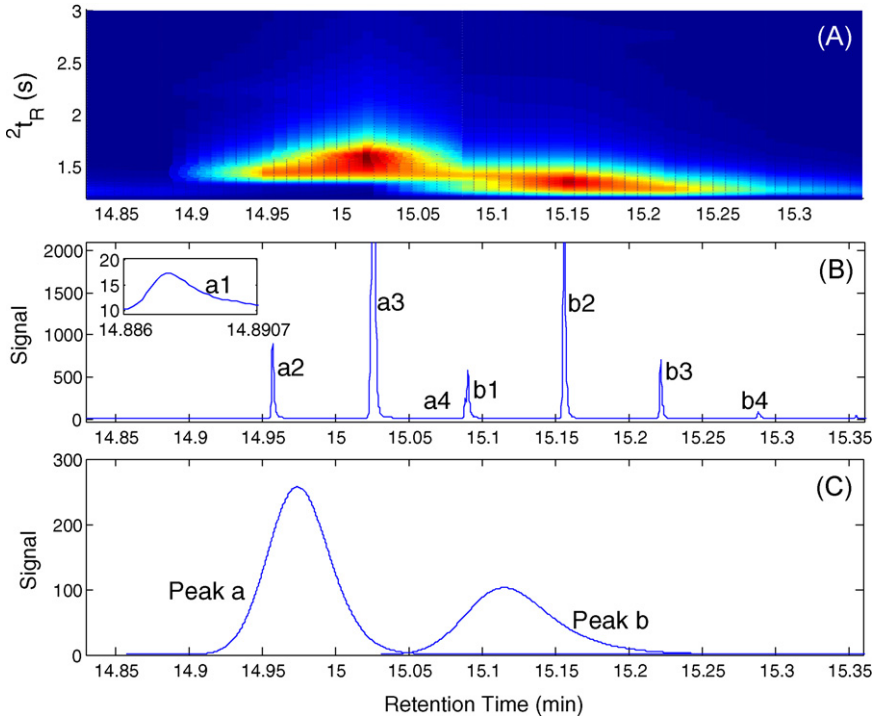

Fig. 7. The image (A), chromatograms of raw peaks eluting from the ^{2}D column (B), and the reconstructed primary overlapping peaks in the first dimension (C), for real example 1. Note that a4 and b1 is a clear overlapping ^{2}D peak where a and b just overlap on the ^{1}D column. Inset shows expanded ^{2}D plot of the a1 result.

the deconvolution results of the ^{2}D overlapping peaks a4 + b1, as well as the selective regions of each target component. The solid line is the original chromatogram along with the residual after two deconvolution passes, and the dotted line is the profile of individual pure peaks a4 and b1 after deconvolution. The ratio (R_{rr}) of deconvolution residual to peak response is 1.0e−1%. From the image of the target data set given in Fig. 7(A), it is easy to find the possible elution components and peaks in Fig. 7(B). The area of individual pure components a and b can then be obtained by reconstruction of the primary overlapping profiles in ^{1}D. Fig. 7(C) shows the two compositions with clear overlapping regions when the two primary peaks are modulated in ^{2}D. The ratios (R_{rr}) between the fitting residual and original data of the two components are 3.4e−5% and 5.67e−4%, respectively. This should be a satisfactory result for further quantification and determination of other parameters of interest.

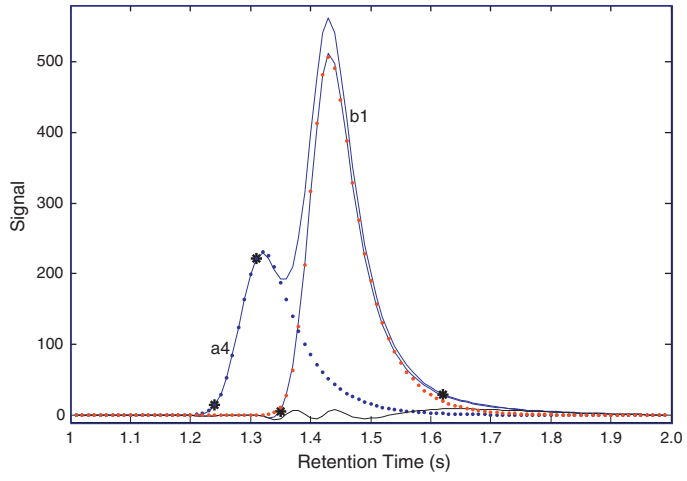

Fig. 8. Deconvolution of the overlapping peak cluster a4/b1 shown in Fig. 7. Solid lines are original chromatogram, and the residual result after stepwise component stripping. The two dotted lines represent the fitting results of peaks a and b using the selective regions of target components. The asterisked symbol pairs (two for a4 and two for b1) have the same meaning as given in Fig. 5.

Fig. 9 shows a further chromatographic result, now with three overlapping components. The image and peak elution pattern is plotted in Fig. 9(A) and (B), respectively. The analytical principle and procedures are very similar to that detailed above. Five potential overlapping clusters a1, a2, a5 and b1, c2 and c3 were tested in ^{2}D (note that in some cases peaks overlap with other matrix components). They should belong to the respective fractions of the three primary peaks a, b and c, as given in Fig. 9(B). The deconvolution results of these components are shown in Fig. 10 with the ratios (R_{rr}) of the sum of square between fitting residual and original data set being 9.33e−2%, 1.0e−3%, 1.34e−1%, 1.20e−2% and 2.4e−2% from Fig. 10(A) to 10(E). In this case, it is important to determine the minor component peaks which interfere with the whole peak profiles. Then, the three components in ^{1}D can be recovered using the acquired areas of secondary peaks with the help of NLLSCF technique, as shown in Fig. 9(C). All the peaks in ^{2}D match very well with the component elution in the first dimension. The fitting ratios (R_{rr}) of residual and original signal of these three components are 1.12e−4%, 6.89e−3%, and 1.34e−7%, respectively.

The case shown in Fig. 11 is more complicated in contrast to the two cases shown in Figs. 7 and 9. Most secondary peaks in ^{2}D have overlapping elution windows with its neighbour, as shown in the expansions in Fig. 12. Fig. 12(G) illustrates a process of relative retention alignment to ensure that the ^{2}D peaks for the same component have the same ^2t_R values. The peaks denoted 'a1' are employed as the reference to correct the ^{2}D retention shift, and the relative shift of peaks b can be clearly found in Fig. 12(G).

All of the components in Fig. 12 need to be firstly resolved to assign the area of each chemical component for subsequent reconstruction of the primary peak profiles in ^{1}D. The deconvolution results are given in Fig. 12(A)–(F); most deconvolution results are acceptable with low residual. For the 5 overlapping systems, the

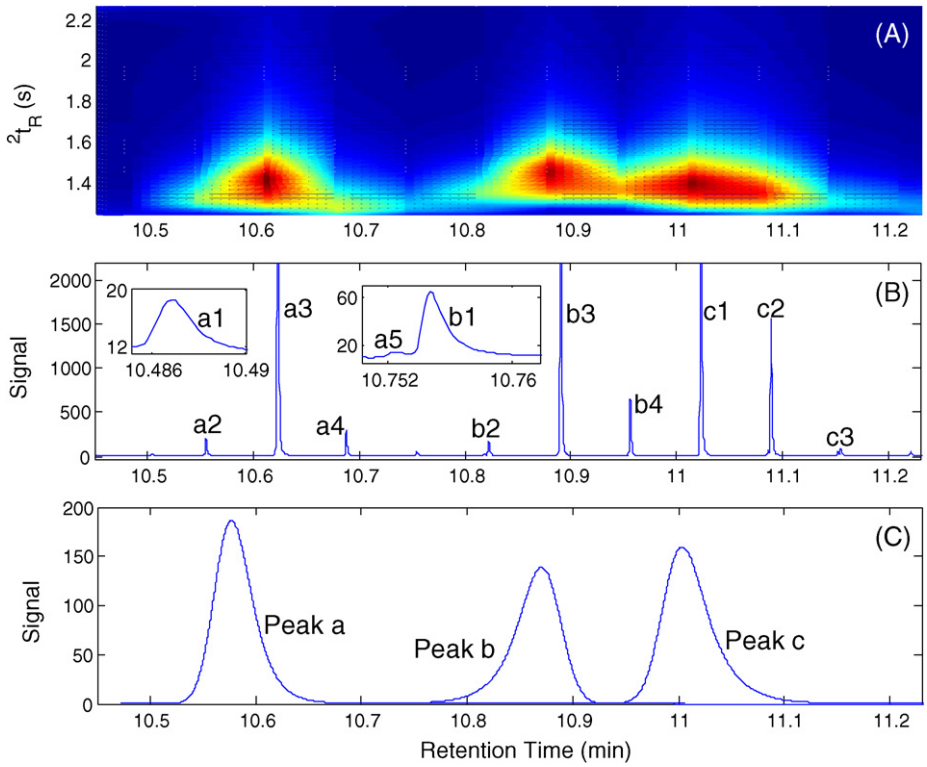

Fig. 9. The two-dimensional image (A), ^{2}D peak elution patterns of raw chromatograms with three components (B), and the recovered results of the ^{1}D overlapping peaks a, b and c (C), for real example 2. The insets are expanded ^{2}D zones. Only slight overlapping exists here.

ratios (R_{rr}) of deconvolution residual and signal are respectively 1.96e−1%, 2.0e−1%, 3.0e−1%, 1.50e−2%, 2.0e−1%, and 1.60e−2%. Using the acquired areas for ^{1}D peak deconvolution, the ratios (R_{rr}) of fitting residual and signal of the two peaks are 1.09e−2% and 3.88e−3%. The results in ^{1}D fully and accurately correspond to the component elution patterns in ^{2}D. It shows the power of the present method for deconvolution of overlapping components and reconstruction of primary profiles in ^{1}D for GC × GC data.

Though the present least squares fitting method generates satisfactory results for the analysis of datasets acquired from C2DC,

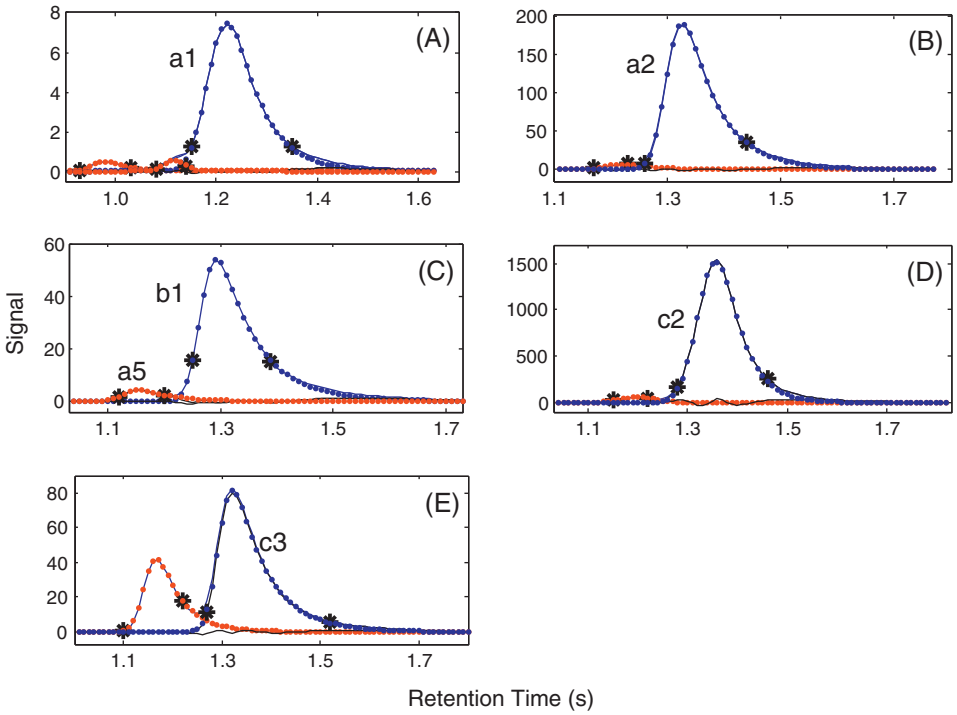

Fig. 10. Deconvolution of five overlapping peak clusters (A–E) for determination of ^{2}D peak areas shown in Fig. 9. Lines have the same meaning as in Fig. 8. Results of all the five cases are acceptable using residual as an evaluation index (see text). The asterisked symbol pairs (two for each peak) have the same meaning as given in Fig. 5.

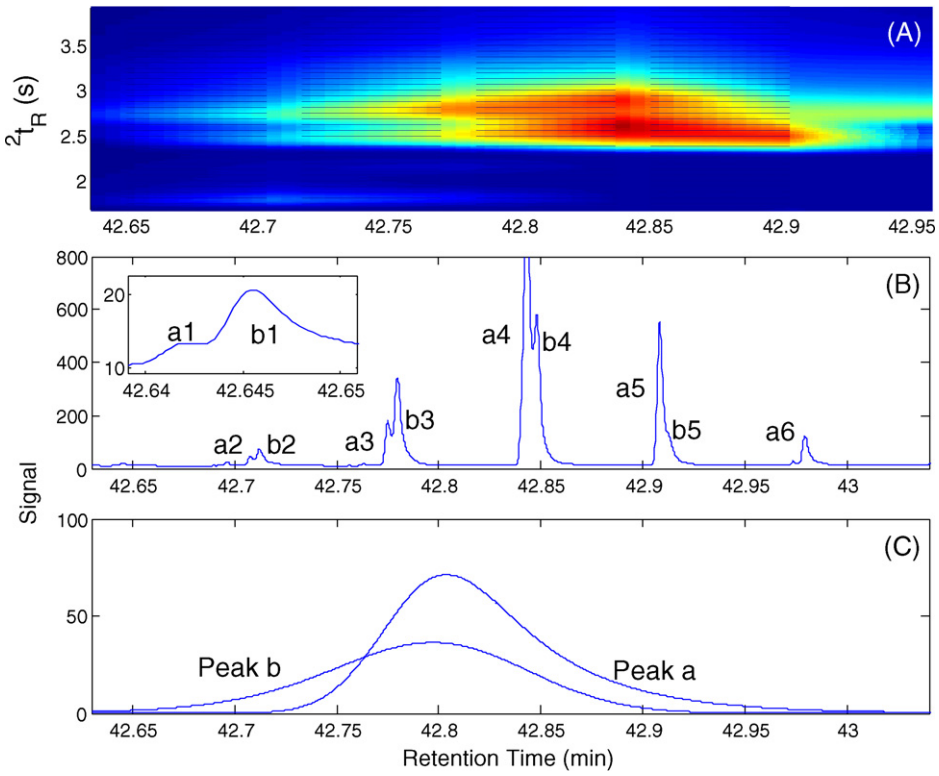

Fig. 11. The 2D image (A), raw chromatograms of two fully overlapping components (B), and the reconstruction results of primary peaks a and b (C), for real example 3. The inset is an expanded ^{2}D zone.

it should clarify the factors with possible effect on the results. Levenberg–Marquardt algorithm is a local optimization method to find acceptable resolution, but it has no uniqueness property to solve the present issue. The results may further be improved if it can be constructed on the basis of a global searching strategy. Some software systems like 1stopt developed by 7D-Soft High Technology Inc. claim that global optimization results can be found using the NLLSCF technique, but with time-consuming computation. This

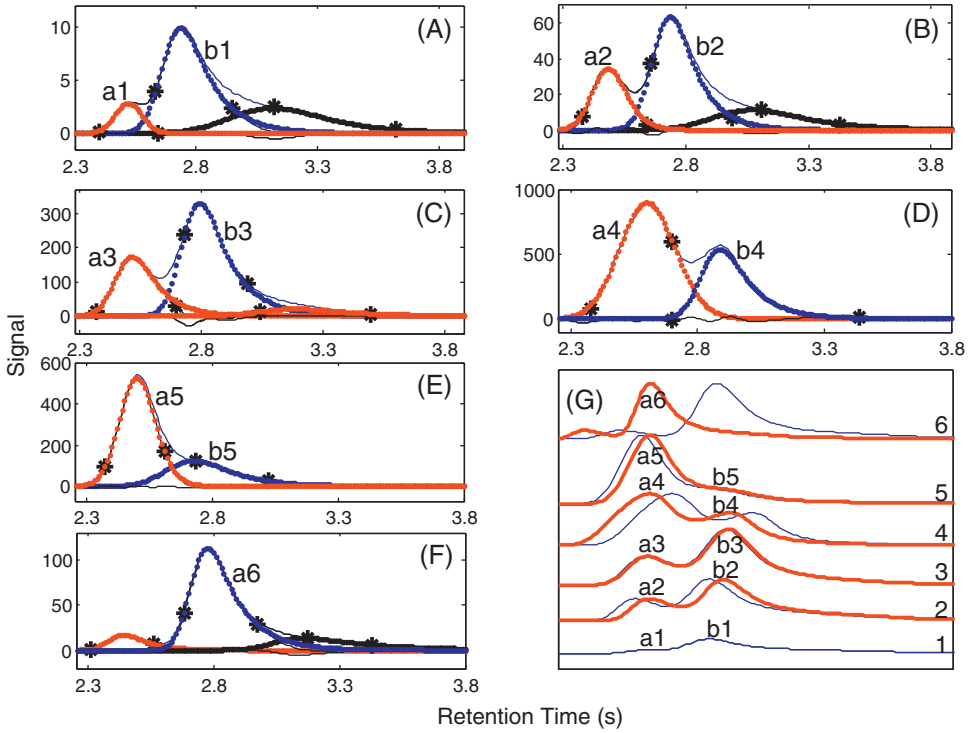

Fig. 12. Deconvolution results of the six secondary overlapping peak clusters (A–F) shown in Fig. 11. The acquired areas of pure peaks in the six figures are utilized to reconstruct the primary profiles a and b shown in Fig. 11(C). (G) Illustrates a suitably retention aligned data set for the above traces (A–F). Asterisked pairs have the same meaning as given in Fig. 5.

limits the application to deconvolution of GC × GC datasets with a large number of data points. Extended computation times should assist to find better results, as well as using the acquired results of the prior iteration as input for the next step. In this work, 10^3 computation iterations are used to generate the results. Strong overlapping systems with highly asymmetrical peaks may also challenge the present method because of the difficulty to obtain sufficient pure component information. But our global searching goals may help to overcome this limitation in future work.

5. Conclusions

Deconvolution of overlapping peaks is not a new topic in the study of chromatography. However the data structure of C2DC is different in contrast to conventional 1D chromatograms. The present work discovers the data information of this new type of dataset to permit simultaneous deconvolution and re-construction of peak profiles for both the ^{1}D and the ^{2}D separations. This makes accurate quantification and finding of retention parameters of target components practical. Though only simulated and experimental GC × GC data sets are employed to introduce the principles of the proposed technique, it also can be utilized to treat the data acquired from LC × LC system. Complete automatic recovery of peak profiles of all the primary and secondary compositions detected by C2DC, and then derivation of information of specific concern to the analyst should be our next concern.

Acknowledgment

This work is financially supported by the Australian Research Council Discovery Grant No. DP0988656. This work is conducted under our affiliation with the Australian Centre for Research on Separation Science (ACROSS).

References

[1] L. Novakova, H. Vlckova, Anal. Chim. Acta 656 (2009) 8.
[2] E. Sobhanzadeh, N.K. Abu-Bakar, K. Nemati, M. Radzi Abas, World Appl. Sci. J. 7 (2009) 923.
[3] J.M. Davis, J.C. Giddings, Anal. Chem. 55 (1983) 418.
[4] W.S. Law, P.Y. Huang, E.S. Ong, C.N. Ong, S.F.Y. Li, K.K. Pasikanti, E.C.Y. Chan, Rapid Commun. Mass Spectrosc. 22 (2008) 2436.
[5] R.S. Plumb, C.L. Stumpf, M.V. Gorenstein, J.M. Castro-Perez, G.J. Dear, M. Anthony, B.C. Sweatman, S.C. Connor, J.N. Haselden, Rapid Commun. Mass Spectrosc. 16 (2002) 1991.
[6] E.M. Lenz, J. Bright, R. Knight, F.R. Westwood, D. Davies, H. Major, I.D. Wilson, Biomarkers 10 (2005) 173.
[7] R.G. Brereton, Applied Chemometrics for Scientists, John Wiley, Chichester, 2006.
[8] R.C.Y. Ong, P.J. Marriott, J. Chromatogr. Sci. 40 (2002) 276.
[9] X. Di, R.A. Shellie, P.J. Marriott, C.W. Huie, J. Sep. Sci. 27 (2004) 451.
[10] S.M. Song, P. Marriott, A. Kotsos, O.H. Drummer, P. Wynne, Forensic Sci. Int. 143 (2004) 87.
[11] C. Cordero, C.O. Bicchi, P. Rubiolo, J. Agric. Food Chem. 56 (2008) 7655.
[12] Y. Shao, P. Marriott, H. Huegel, Chromatographia 57 (Suppl.) (2003) S349.
[13] C.G. Fraga, C.A. Corley, J. Chromatogr. A 1096 (2005) 40.
[14] H.W. Kong, F. Ye, X. Lu, L. Guo, J. Tian, G.W. Xu, J. Chromatogr. A 1086 (2005) 160.
[15] J.R. Torres-Lapasió, J.J. Baeza-Baeza, M.C. García-Alvarez-Coque, Anal. Chem. 69 (1997) 3832.
[16] A. Felinger, T. Pap, J. Inczedy, Talanta 41 (1994) 1119.
[17] S.N. Chesler, S.P. Cram, Anal. Chem. 45 (1973) 1354.
[18] R.A. Vaidya, R.D. Hester, J. Chromatogr. 287 (1984) 231.
[19] J.T. Lundeen, R.S. Juvet, Anal. Chem. 53 (1981) 1369.
[20] K.H. Jung, S.J. Yun, S.H. Kang, Anal. Chem. 56 (1984) 457.
[21] D.Y. Youn, S.J. Yun, K.H. Jung, J. Chromatogr. 591 (1992) 19.
[22] G. Vivo-Truyols, J.R. Torres-Lapasio, A.M. van Nederkassel, Y. Vander Heyden, D.L. Massart, J. Chromatogr. A 1096 (2005) 146.
[23] P. Nikitas, A. Pappa-Louisi, A. Papageorgiou, J. Chromatogr. A 912 (2001) 13.
[24] Z.D. Zeng, Y.Z. Liang, Y.L. Wang, X.R. Li, L.M. Liang, Q.S. Xu, C.X. Zhao, B.Y. Li, F.T. Chau, J. Chromatogr. A 1107 (2006) 273.
[25] G.X. Chen, P.D. Harrington, Anal. Chim. Acta 484 (2003) 75.
[26] U. Evans, O. Soyemi, M.S. Doescher, U.H.F. Bunz, L. Kloppenburg, M.L. Myrick, Analyst 126 (2001) 508.
[27] K. Levenberg, Quart. Appl. Math. 2 (1944) 164.
[28] W.H. Press, B.P. Flanerry, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes, The Art of Scientific Computing, Cambridge University Press, Cambridge, 1986.
[29] J.L. Adcock, M. Adams, B.S. Mitrevski, P.J. Marriott, Anal. Chem. 81 (2009) 6797.
[30] L.L. Xie, P.J. Marriott, M. Adams, Anal. Chim. Acta 500 (2003) 211.
[31] W.W. Yau, Anal. Chem. 49 (1977) 395.
[32] J.P. Foley, J.G. Dorsey, Anal. Chem. 55 (1983) 730.
[33] D.S. He, S.J. Li, Chin. J. Chromatogr. 9 (1991) 20.
[34] N.S. Wu, A.M. Qiu, M.L. Wu, Chin. J. Chromatogr. 6 (1988) 153.
[35] D.S. He, S.J. Li, Acta Chim. Sinica 48 (1990) 673.
[36] G. Vivo-Truyols, J.R. Torres-Lapasio, A.M. van Nederkassel, Y. Vander Heyden, D.L. Massart, J. Chromatogr. A 1096 (2005) 133.
[37] W. Khummueng, J. Harynuk, P.J. Marriott, Anal. Chem. 78 (2006) 4578.
[38] J.E. Kuo, H. Wang, S. Pickup, Anal. Chem. 63 (1991) 630.
[39] R. Ong, R. Shellie, P. Marriott, J. Sep. Sci. 24 (2001) 367.
[40] N.P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, J. Chromatogr. A 805 (1998) 17.
[41] F. Gong, B.T. Wang, F.T. Chau, Y.Z. Liang, Anal. Lett. 38 (2005) 2475.
[42] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 73 (2001) 5833.
[43] K.M. Pierce, L.F. Wood, B.W. Wright, R.E. Synovec, Anal. Chem. 77 (2005) 7735.