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

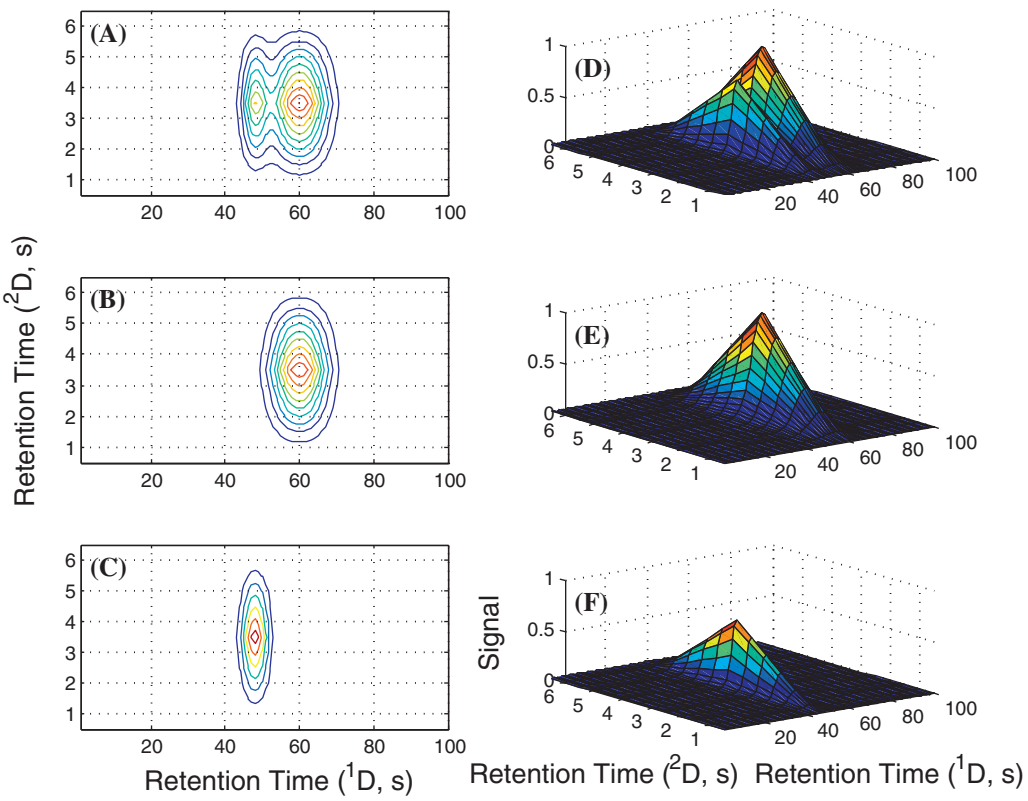

Fig. 1. Illustration of the data structure of comprehensive two-dimensional chromatography, and deconvolution and reconstruction of both primary and secondary chromatographic profiles. (A and D) Contour and three-dimensional graphs of overlapping peak cluster of two simulated components. (B and C) and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