Trends in Analytical Chemistry 53 (2014) 150–166

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

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

Review

Interpretation of comprehensive two-dimensional gas chromatography data using advanced chemometrics

Zhongda Zeng [a,b], Jia Li [b], Helmut M. Hugel [c], Guowang Xu [b], Philip J. Marriott [a,*]

[a] Australian Centre for Research on Separation Science, School of Chemistry, Monash University, Wellington Rd, Clayton 3800, Australia
[b] CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China
[c] School of Applied Sciences, RMIT University, GPO Box 2476, Melbourne 3001, Australia

ARTICLE INFO

Keywords:
Chemometrics
Comprehensive two-dimensional gas chromatography (GC × GC)
Data characteristics
Data processing
Deconvolution
Mass spectrometry
Multivariate curve resolution (MCR)
Multivariate data analysis
Orthogonality
Signal processing

ABSTRACT

The power of comprehensive two-dimensional gas chromatography (GC × GC) for the study of complex mixtures has been indisputably proved in the past several decades. This review encompasses the whole of GC × GC-related data processing and summarizes relevant applications. We include theoretical introduction to some specific methods and studies to aid readers' understanding of chemometrics strategies for advanced data interpretation.

Contents

1. Introduction . . . 150
2. GC × GC data deconvolution . . . 152
3. MCR of overlapping peaks in GC × GC-MS . . . 153
   3.1. Single-run-based data resolution . . . 154
      3.1.1. Unique resolution methods . . . 154
      3.1.2. Rational resolution methods . . . 154
   3.2. Multi-run-based data resolution . . . 155
      3.2.1. Trilinear decomposition (TLD) methods . . . 155
      3.2.2. Non-trilinear decomposition (non-TLD) methods. . . . 156
4. New insights from chemometrics for GC × GC structure interpretation with two-dimensional characteristics . . . 157
   4.1. Data preprocessing . . . 157
      4.1.1. Two-dimensional peak detection. . . . 157
      4.1.2. Correction of retention-time shift . . . 158
   4.2. Orthogonality study . . . 158
   4.3. Image processing. . . . 159
5. Summary of other aspects of GC × GC studies with chemometrics . . . 159
   5.1. Experimental optimization . . . 159
   5.2. Pattern recognition . . . 160
   5.3. Model calibration . . . 160
6. Concluding remarks . . . 161
   Acknowledgments . . . 161
   References . . . 161

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

0165-9936/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.trac.2013.08.009

1. Introduction

Comprehensive two-dimensional gas chromatography (GC × GC) is a natural extension in the panoply of development of conventional separations, including single-dimensional chromatography (SDC) and heart-cut (H/C) techniques. H/C is also called multidimensional gas chromatography (MDGC), defined as "the process of selecting a (limited) region or zone of eluted compounds issuing from the end of one GC column, and subsequently subjecting the zone to a further GC displacement" [1].

In the 1990s, a typical example of MDGC separation comprised about 50 individual sampled regions each with around 20 peaks for pattern recognition of pharmaceuticals [2]. In the past decade, the number of "cross-samples" investigated in a specific study and the number of peaks contained in each sample significantly increased (e.g., to 1000 and 500, respectively), because of biology-driven studies, such as proteomics and metabolomics.

Our recent work reported that 2771 compounds were found in an investigation of a Chinese medicinal formulation (CMF) that included nine single herbs, using the platform of GC × GC with time-of-flight mass spectrometry (GC × GC-TOF-MS) [3]. The ability to resolve a sample of such complexity is an evident challenge or even an impossible task for SDC analysis.

However, successful applications exploited the power of GC × GC-related techniques to compositions in mixtures of high complexity, such as herbal medicines and drugs, flavors, foods, petroleum and biological samples [4–11]. The outcomes and the performance of this technique have been introduced and reviewed, with frequent updates [5,12–20]. The basic experiment comprises the connection of two chromatographic columns with complementary polarity that together enhance the separation capacity of the arrangement; the columns are interfaced through a modulator device, which effectively decouples elution on each column [21,22]. The column set pairs two columns that are most often defined as comprising a low polarity (LP)/polar (P) combination, a moderately polar (MP)/polar combination, a P/LP combination, or a P/MP combination. Note, however, that these are relative properties, since a very polar/P combination may perform similarly to a P/LP combination. Such smart configurations help to separate and to re-arrange further the peaks in the first dimension (^{1}D) compared to the second one (^{2}D), with a fixed modulation period (P_M) and the same total analysis time [23,24]. Thus, peak capacity, a theoretical measure of the number of peaks that can be separated in the 2D space, can then ideally attain to the product $^1n_c \times {}^2n_c$, assuming the peak capacities in ^{1}D and ^{2}D separations are ^1n_c and ^2n_c, respectively [25]. This is the essential advantage of GC × GC, enabling the investigation – and separation – of samples with hundreds or even thousands of chemical components in contrast to SDC and MDGC techniques.

Unlike the conventional data structures of SDC, MDGC and coupling of chromatographic and spectral instruments, GC × GC data have two special properties:

(1) 2D characteristics with specific retention properties and response in ^{1}D and ^{2}D dimensions; and,
(2) loss of raw chromatographic data in ^{1}D, but continuous modulation of fractions in ^{2}D for each ^{1}D peak.

For GC × GC-MS data with different mass analyzers, such as quadrupole and TOF, ideally a single-component mass spectrum can be detected at each retention-time (t_R) measurement point throughout the 2D GC × GC separation plane. This effectively expands the original data to a three-dimensional (3D) data set, with t_R in both ^{1}D and ^{2}D (^1t_R and ^2t_R, respectively), and spectral intensity at the scanned *m/z*, comprising the x-, y-, and z-axes, respectively. Further, time-dependent and sample-to-sample dynamic variations complicate data processing and information extraction (extended to a four-dimensional arrangement) (e.g., metabolite fingerprinting analysis in metabolomics analysis with evolution of treatment or environmental effects over time). One of the typical examples is correction of t_R shifts among different but related samples of GC × GC-TOF-MS, or different types of detectors on the basis of GC × GC separation [26,27].

The complexity of GC × GC-related data and high-throughput analysis for real mixtures make chemometrics widely applicable to this area [28–31], which has the power to expose buried information in "white, grey and black systems" with different degrees of prior knowledge of multi-components, and draws on multivariate statistics, mathematics and computer science [32], as shown in Fig. 1. Many of the reported reviews of GC × GC incorporate the relevance of chemometrics for the investigation of GC × GC data, and include theoretical development and novel applications [28,29,33]. This work further explores the nexus between GC × GC and chemometrics to mine out hidden information with mathematical interpretation, and aims to provide extra understanding to the researcher without a chemometrics background. Previously reported chemometrics tools for processing of coupled data are introduced to explain GC × GC data, such as multivariate curve resolution (MCR) for bilinear data decomposition based on the principles of the Beer-Lambert Law (BLL). In terms of the 2D, or even 3D, data characteristics introduced above, we review some specific research insights of GC × GC, such as orthogonality and image analysis.

First, chemometrics methods to deconvolute overlapping GC × GC peak clusters in ^{1}D and ^{2}D separations are introduced by using model or fitting techniques. Based on the 2D feature of GC × GC separation, conventional deconvolution methods for 2D or 3D data processing have been applied for GC × GC processing. This further helps to recover lost information of primary peaks. Second, MCR methods based on single or multiple runs are separately summarized, to extract chromatographic data and spectral profiles of pure components from GC × GC-TOF-MS data to support identification and quantification [34,35]. Four important chemometrics methods for 2-way and 3-way data resolution are introduced in theory, with worked examples of processed GC × GC-related data, including heuristic evolving latent projection (HELP) [36,37], parallel factor analysis (PARAFAC) [38,39], MCR-alternating least squares (MCR-ALS) [40,41], and alternative moving window factor analysis (AMWFA) [42,43].

Next, some new research topics applied to GC × GC data are expounded, exploiting the 2D separation characteristics and matrix data structure, such as t_R alignment, orthogonality and image analysis. Last, but not least, some routine considerations of the GC × GC experiment related to data processing aided by chemometrics are reviewed, such as peak detection, experimental design and optimization, signal processing, and component-calibration models. This should familiarize the reader with an appreciation of various chemometrics tools for presentation and interpretation of GC × GC and GC × GC-MS data.

In addition, some commercial and freely downloadable programs or software for GC × GC data analysis are introduced and can be readily used following instructions [44–46]. This includes signal-to-noise filtering, baseline correction, retention-time alignment, normalization, peak picking, deconvolution, integration, and library searching and identification by using retention-index and MS libraries [47,48]. This allows chemometrics strategies to be readily employed by researchers with limited chemometrics experience, such as ChromaToF data-processing software (Leco,