Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1056 (2004) 145–154
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

# Multivariate selectivity as a metric for evaluating comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry subjected to chemometric peak deconvolution☆

Amanda E. Sinha[a], Janiece L. Hope[a], Bryan J. Prazen[a], Carlos G. Fraga[b], Erik J. Nilsson[c], Robert E. Synovec[a,*]

[a] *Department of Chemistry, Center for Process Analytical Chemistry, University of Washington, P.O. Box 351700, Seattle, WA 98195-1700, USA*
[b] *Department of Chemistry, 2355 Fairchild Dr. Suite 2N225, United States Air Force Academy, Colorado Springs, CO 80840-6230, USA*
[c] *Insilicos, 4509 Interlake Avenue North, #223, Seattle, WA 98103-6773, USA*

Available online 17 August 2004

**Abstract**

Two-dimensional gas chromatography (GC × GC) coupled to time-of-flight mass spectrometry (TOFMS) [GC × GC–TOFMS)] is a highly selective technique well suited to analyzing complex mixtures. The data generated is information-rich, making it applicable to multivariate quantitative analysis and pattern recognition. One separation on a GC × GC–TOFMS provides retention times on two chromatographic columns and a complete mass spectrum for each component within the mixture. In this report, we demonstrate how GC × GC–TOFMS combined with trilinear chemometric techniques, specifically parallel factor analysis (PARAFAC) initiated by trilinear decomposition (TLD), results in a powerful analytical methodology for multivariate deconvolution. Using PARAFAC, partially resolved components in complex mixtures can be deconvoluted and identified without requiring a standard data set, signal shape assumptions or any fully selective mass signals. A set of four isomers (*iso*-butyl, *sec*-butyl, *tert*-butyl, and *n*-butyl benzenes) is used to investigate the practical limitations of PARAFAC for the deconvolution of isomers at varying degrees of chromatographic resolution and mass spectral selectivity. In this report, multivariate selectivity was tested as a metric for evaluating GC × GC–TOFMS data that is subjected to PARAFAC peak deconvolution. It was found that deconvolution results were best with multivariate selectivities over 0.18. Furthermore, the application of GC × GC–TOFMS followed by TLD/PARAFAC is demonstrated for a plant metabolite sample. A region of GC × GC–TOFMS data from a complex natural sample of a derivatized metabolic plant extract from Huilmo (*Sisyrinchium striatum*) was analyzed using TLD/PARAFAC, demonstrating the utility of this analytical technique on a natural sample containing overlapped analytes without selective ions or peak shape assumptions.
© 2004 Elsevier B.V. All rights reserved.

*Keywords:* Parallel factor analysis; Chemometrics; Deconvolution; Gas chromatography, comprehensive two-dimensional; Net analyte signal; Multivariate selectivity

## 1. Introduction

Two dimensional gas chromatography (GC × GC) time-of-flight mass spectrometry (TOFMS) [GC × GC–TOFMS] is quickly becoming a popular area of research [1–9], fueled by tremendous selectivity and sensitivity for the identification and analysis of components in complex mixtures. Along with providing selectivity and sensitivity, GC × GC–TOFMS is an instrument capable of generating trilinear data, thus broadening the opportunity to utilize state-of-the-art chemometric techniques for signal deconvolution (i.e., mathematical resolution). The trilinear data structure of GC × GC–TOFMS can be a useful attribute for mathematically resolving overlapped signals. Certain conditions need to be met in order for data to be trilinear: "the response in [all] domains of the

☆ This work was presented at the 8th International Symposium on Hyphenated Techniques in Chromatography (HTC-8)

* Corresponding author. Tel.: +1 206 6852328; fax: +1 206 6858665.
*E-mail address:* synovec@chem.washington.edu (R.E. Synovec).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.06.110

instrument arising from a species should be unique, consistent, and independent of the presence of other species" [10]. Trilinear data structure is the key to chemometric deconvolution techniques like the generalized rank annihilation method (GRAM), trilinear decomposition (TLD) and parallel factor analysis (PARAFAC) [11].

Currently available peak deconvolution methods used for GC × GC–TOFMS data analysis are typically extensions of those that are already used for GC–MS and do not utilize the trilinear data structure. These deconvolution methods reduce GC × GC–TOFMS data into a series of column 2 separation dimension GC–MS data. Each column 2 separation is analyzed independently, then recombined along the column 1 separation dimension for the final assessment. These methods essentially do not utilize the column 1 separation dimension during deconvolution of overlapped peaks.

With GC × GC combined with non-spectrometric detection like flame ionization detection, calibration methods such as GRAM are able to deconvolute chromatographic signals using two data sets (standard and sample) where the analytes of interest vary in concentration between the two data sets [12–17]. These techniques stack multiple bilinear data sets to create a trilinear data structure, which can then be deconvoluted. Using trilinear data, such as GC × GC–TOFMS data, it is possible to deconvolute individual components from a group of partially overlapped components using a data set from only one sample. Thus, analytes of interest can be identified in a complex mixture when there is only partial selectivity in both chromatographic dimensions and the mass spectral dimension, unlike currently available methods that rely upon selective ion deconvolution approaches. The ability to deconvolute a single data set from partially resolved signals into the fully resolved signals is known as the third-order advantage [18]. In the case of chromatographic data, third-order data is also advantageous because it relaxes the requirements for sample-to-sample retention time precision, which is a critical issue for other methods based on the trilinear data model such as GRAM [19], thus essentially eliminating the need for retention time alignment prior to analyte deconvolution.

Previously, we made an initial report demonstrating the use of TLD and PARAFAC for GC × GC–TOFMS data on an environmental sample [20]. Here we build upon our previous work, by studying the effects of multivariate selectivity (chromatographic resolution and mass spectral similarity) on GC × GC–TOFMS peak deconvolution methods based on the trilinear data structure. The chemometric methodology will employ TLD to initiate PARAFAC. An investigation is presented with four butyl benzene isomers (*sec*-butyl, *iso*-butyl, *tert*-butyl and *n*-butyl benzenes) that have similar spectra with no selective major ions. In order to investigate TLD-initiated PARAFAC deconvolution of peaks with similar spectra at varying chromatographic resolution, three replicate data sets were collected by GC × GC–TOFMS for each of four isomers (*sec*-butyl, *iso*-butyl, *tert*-butyl and *n*-butyl benzenes) at a concentration of 3% (v/v) in hexane. These individual data sets of the isomers were then added together to create "constructed" samples. These constructed samples have the appearance of a separation of a mixture of the four isomers. Although these isomers can be separated by GC × GC, they are used here as an investigative tool to probe the practical limits of third-order chemometric techniques for peak deconvolution. The major benefits of this approach are that the resolution (i.e., effective chromatographic selectivity) of the isomers in each chromatographic dimension can be readily altered, the true expected deconvolution results are known, and noise in the data is real. All analyses were performed in triplicate and after baseline correction. The chemometric models were calculated with four components specified. The PARAFAC analysis also incorporated non-negative constraints in all dimensions and unimodal constraints for both chromatographic dimensions. In addition, a derivatized metabolic plant extract from Huilmo (*Sisyrinchium striatum*) containing a selected region of overlapped peaks is also analyzed to demonstrate the technique on a novel, complex natural sample. Future studies involving metabolites in complex samples will explore the application of pattern recognition methods that take advantage of the trilinear structure to GC × GC–TOFMS data [21,22].

## 2. Theory

### 2.1. Trilinear data

Mathematically, the trilinear parallel factor analysis (PARAFAC) model is described as:

$$\boldsymbol{R} = \sum_{n=1}^{N} x_n \otimes y_n \otimes z_n + \boldsymbol{E} \tag{1}$$

where $\boldsymbol{R}(I \times J \times K)$ is the instrument response matrix, $x_n$, $y_n$, and $z_n$ are the $n$th columns of the matrices $\boldsymbol{X}(I \times N)$, $\boldsymbol{Y}(J \times N)$, and $\boldsymbol{Z}(K \times N)$ containing the $N$ pure component profiles in each dimension. $\boldsymbol{E}(I \times J \times K)$ is the error matrix, e.g., noise, and $\otimes$ denotes the mathematical function for the outer (or cross) product. For GC × GC–TOFMS data the dimensions are the column 1 separation space ($\boldsymbol{X}$), the column 2 separation space ($\boldsymbol{Y}$), and the mass spectrum ($\boldsymbol{Z}$). Fig. 1 contains a graphical representation of the trilinear model as applied to GC × GC–TOFMS data. Data with the trilinear structure, like GC × GC–TOFMS data, is gainful because signals which are not fully resolved by the instrument can normally be mathematically resolved if there is at least some selectivity in each of the three dimensions. This mathematical resolution, or deconvolution, does not entail peak shape assumptions nor fully selective mass channels.

### 2.2. Multivariate selectivity and net analyte signal (NAS)

When quantifying the selectivity of GC × GC–TOFMS data in real examples, chromatographic resolution alone often does not fully describe the selectivity of GC × GC–TOFMS data. Chromatographic resolution describes the selectivity of

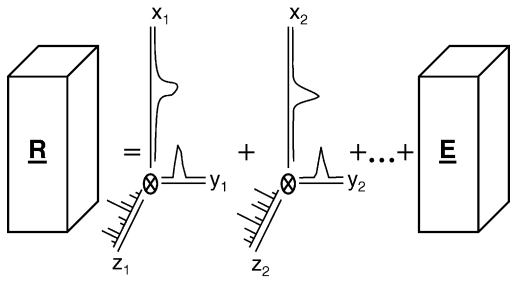


Fig. 1. Illustration representing the trilinear data structure of GC × GC-TOFMS data. For the instrument response $\boldsymbol{R}$ there are unique profiles in both chromatographic dimensions ($x_n$ and $y_n$) and a unique mass spectrum ($z_n$) for each component in a data matrix, which can be described mathematically as $\boldsymbol{R} = \sum_{n=1}^{N} x_n \otimes y_n \otimes z_n + \boldsymbol{E}$ where $\boldsymbol{E}$ is error (e.g., noise).

a peak when it is overlapped with a single interfering peak; however, when there are multiple interfering peaks, resolution cannot fully describe the information. Also, the selectivity contained in the spectra from hyphenated techniques, like GC × GC–TOFMS, can play a critical role in determining the information content of overlapped peaks, but this is not accounted for in the resolution equation. For example, two peaks that have low-resolution on both chromatographic dimensions could have nearly selective mass channels and thus contain a great deal of analytical information. To deal with this issue, chemometricians have introduced a metric known as multivariate selectivity [18,23,24]. Multivariate selectivity is the degree of overlap of an analyte's signal with signals from different sources. Multivariate selectivity ranges between zero (complete overlap) and unity (no overlap). A plot comparing multivariate selectivity to two-dimensional chromatographic resolution is shown in Fig. 2. For low-resolution cases the small changes in resolution change the selectivity drastically. Once two peaks are resolved by about 0.75, further resolution does not substantially increase the multivariate selectivity. These trends are analogous to the results from chromatographic deconvolution. For chromatographic

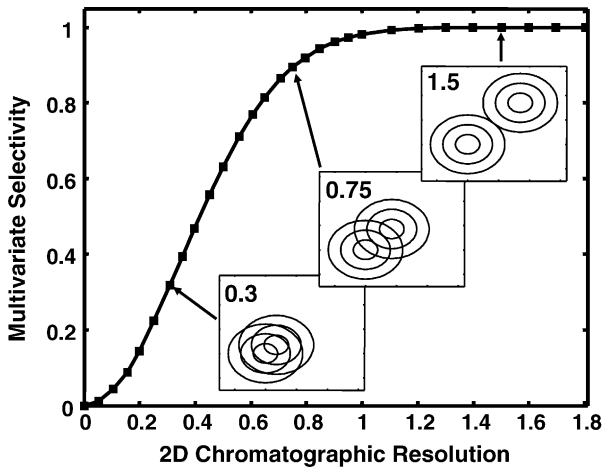


Fig. 2. Plot of data multivariate selectivity as a function of two-dimensional chromatographic resolution.

analysis, multivariate selectivity can describe the information content of peaks overlapped with multiple interference peaks and can be extended to describe data from hyphenated chromatographic separation and spectrometric data. We have found that multivariate selectivity is strongly correlated with the quality of results when techniques that are based on the trilinear data model are used for deconvolution and quantification. Like chromatographic resolution, multivariate selectivity is independent of both concentration and detector sensitivity, and thus can be combined with signal-to-noise or sensitivity for a complete description of data. Multivariate selectivity (SEL) is defined as the ratio of the net analyte signal (nas) to the total analyte signal, as shown in Eq. (2):

$$\text{SEL} = \frac{||\text{nas}(s)||}{||s||} \tag{2}$$

where $||s||$ denotes the norm of the signal ($s$). In geometrical terms, net analyte signal, nas($s$), is defined as the portion of an analyte's signal that is orthogonal to the signals of the other components in the sample matrix. The net analyte signal of a vector of data, such as a chromatogram or spectrum, is described mathematically as:

$$\text{nas}(s) = (\boldsymbol{I} - \boldsymbol{X} \cdot \boldsymbol{X}^{+})s \tag{3}$$

where $\boldsymbol{X}$ is a matrix of chromatograms or spectra of all components in the sample except that of the analyte. $\boldsymbol{X}^{+}$ denotes the pseudoinverse of $\boldsymbol{X}$, and $\boldsymbol{I}$ is the identity matrix with the same dimensions as $\boldsymbol{X}$. For the description of three-dimensional signals like GC × GC–TOFMS the selectivity on each dimension is multiplied to obtain the three-dimensional selectivity. This is the reason that some selectivity is required on each of the chromatographic and spectral dimensions to do mathematical resolution using trilinear techniques. It is important to remember that multivariate selectivity can be misleading in multidimensional data when there is nearly complete selectivity in some of the dimensions but not all of the dimensions. In this case, the acute analyst would only need to use the selective dimensions for quantitation and identification, thus arriving at a multivariate selectivity of nearly one, by discarding the less selective dimensions. In this report, multivariate selectivity is used as a metric to describe the analytical information content of peaks overlapped on both of the chromatographic dimensions and the mass spectral dimension.

### 2.3. Trilinear decomposition (TLD) and PARAFAC deconvolution

TLD and PARAFAC are multivariate techniques for peak deconvolution and calibration that have been well documented in the literature [11,25–28]. TLD is an eigenvalue-based solution to the trilinear PARAFAC model [25,29]. PARAFAC is the alternating least squares (ALS) based solution to the trilinear PARAFAC model [27,28]. PARAFAC requires a starter solution, in these experiments the TLD results,

to model the data. Other feasible starter solutions include random values, random orthogonalized values and singular values. We have found that TLD initialization for PARAFAC was the fastest and gave the best results for GC × GC–TOFMS data. TLD is advantageous in that it does not require a starter solution and because it is computationally fast, but for the results presented in this report, TLD-initiated PARAFAC provided better results than TLD alone. We attribute the advantage of PARAFAC deconvolution to non-negative and unimodal constraints that are incorporated into the PARAFAC deconvolution algorithm. Constraints cannot be incorporated into the TLD algorithm. Superior deconvolution with PARAFAC agrees with the results of other authors [28]. For brevity, only TLD-initiated PARAFAC results are presented here, and TLD results are omitted. PARAFAC requires some selectivity on each of the chromatographic and spectral dimensions, but avoids the requirement of selective ions.

## 3. Experimental

### 3.1. Butyl benzene isomers

An Agilent 6890 Gas Chromatograph (Agilent Technologies, Palo Alto, CA, USA) was modified to a valve-based GC × GC by mounting the portions of the valve (VICI, Valco Instruments Co. Inc., Houston, TX, USA) which are exposed to sample inside the oven and the remaining portions outside the oven uncovered and exposed to room air [30]. The instrument was equipped with an Agilent 7683 auto-injector (Agilent Technologies, Palo Alto, CA, USA). Column 2 of the GC × GC was connected to a LECO Pegasus III TOFMS (LECO Corporation, St. Joseph, MI, USA) via the heated transfer line. A diagram of the instrument is contained in a recent publication [6]. A set of four butyl benzene isomers with similar spectra was used to investigate the effect of changing selectivity on the TLD and PARAFAC results by changing the chromatographic resolution in the GC × GC separation space.

Four butyl benzene isomers (*sec*-butyl, *iso*-butyl, *tert*-butyl and *n*-butyl benzenes) (99% purity; Sigma-Aldrich Corp., St. Louis, MO, USA) were used to create 3% vol/vol solutions of each individual isomer in hexane (96% *n*-hexane; J.T. Baker, Phillipsburg, NJ, USA). Three replicate data sets were obtained for each solution. Column 1 of the GC × GC–TOFMS for the butyl benzene isomers was a 20 m × 180 μm i.d. capillary column with a 0.18 μm 5% diphenyl/95% dimethyl polysiloxane film (RTX-5; Restek Corp.). Column 2 was a 3 m × 180 μm i.d. capillary column with a 0.05 μm 90% biscyanopropyl/10% phenylcyanopropyl film (RTX-2330; Restek Corp.). Ultra high purity helium was used as the carrier gas. Column 1 was operated with a constant pressure of 20 psi (138 kPa). Column 2 was operated with a constant pressure of 2 psi (13.8 kPa). The injector set point was 275 °C and 0.2 μl injections of 3% (v/v) solutions of each isomer in hexane were split 50:1. The oven was operated isothermally at 100 °C for 4 min. The valve was equipped with a 5 μl sample loop and actuated at a rate of 1 Hz with a 20 ms injection pulse width. A stand-alone pulse generator was used to control the valve actuation [6]. The mass spectrometer had a transfer line temperature of 250 °C and an ion source temperature of 200 °C. The filament bias voltage was −70 V and the detector voltage was −1500 V. All other TOFMS parameters were set from the results of an automated optimization sequence controlled by the LECO software using perfluorotributylamine (PFTBA) as the standard. Data were collected from *m/z* 30 to 150 at a nominal rate of 5 kHz and averaged to 100 full spectra/second by the LECO software. Data were then exported as a comma separated value (.csv) file and loaded into Matlab 6.0 R12 (The Mathworks, Natick, MA, USA) for data processing. The algorithm for TLD was from the PLS_Toolbox (Eigenvector Research, Inc., Manson, WA, USA) and was selected for the advantageous sequencing of the three dimensions of the matrix prior to analysis. The PARAFAC algorithm was from the N-way Toolbox 2.01 [31]. Mass spectral similarity searches were performed with NIST MS Search 2.0 (NIST/EPA/NIH Mass Spectral Library; NIST 98). Baseline correction was done at each mass with a linear correction along the second column time dimension over the region subjected to deconvolution.

### 3.2. Huilmo (Sisyrinchium striatum) metabolite extracts

A natural plant metabolite sample of Huilmo (*Sisyrinchium striatum*) was also analyzed to demonstrate TLD-initiated PARAFAC on a natural sample. Prior to analysis, metabolites were extracted and derivatized via trimethylsilylation [32,33]. The analysis of the extracted sample was performed using a thermally modulated LECO Pegasus 4D GC × GC–TOFMS instrument (LECO Corporation). Column 1 was a 10 m × 180 μm i.d. capillary column with a 0.18 μm 5% diphenyl/95% dimethyl polysiloxane (DB-5; J & W Scientific, Alltech, Deerfield, IL). Column 2 was a 2 m × 100 μm i.d. capillary column with a 0.1 μm film of (50%-phenyl)-methylpolysiloxane (DB-17; Alltech). These columns were joined using a mini union (Scientific Glass Engineering SGE, Austin, TX). Modulation, or delivery of column 1 effluent onto column 2 was performed using cryogenic modulation. Effluent from column 1 was concentrated at the head of column 2 during each column 2 separation. A "hot pulse," occurring when the cryogenic gas was switched off and the heated air jets (40 °C above the oven temperature) were turned on, was used to begin each new column 2 run. The hot pulse was 0.40 s in duration, and the total column 2 run time was 2 s. Ultra high purity helium (0.8 ml/min) was used as the carrier gas. 1 μl of derivatized sample was injected using a 25:1 split and a column 1 oven ramp beginning at 70 °C with a hold time of 5 min then increasing at 5°/min to 250 °C with a hold time at 250 °C of 5 min. Column 2 was held in a separate oven, which was held at a constant 40 °C higher than the column 1 temperature throughout the column 1 oven ramp. The first 5 min of each run was considered a

Table 1
Accuracy and precision studies for TLD-initiated PARAFAC results for the high-, intermediate-, and low-resolution constructed data sets comprised of *iso*-butyl, *sec*-butyl, *tert*-butyl, and *n*-butyl benzene isomers for three replicates

| Analyte | Selectivity Eq. (2) | | | Bias (%) | | | R.S.D. (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | H | I | L | H | I | L | H | I | L |
| *Sec*-Butyl benzene | 0.768 ± 0.009 | 0.366 ± 0.004 | 0.05 ± 0.01 | +1.5 | +0.9 | +0.9 | 1.1 | 1.0 | 1.3 |
| *Tert*-Butyl benzene | 0.659 ± 0.003 | 0.29 ± 0.02 | 0.044 ± 0.001 | +0.6 | +1.3 | −8.0 | 0.8 | 0.8 | 2.4 |
| *Iso*-Butyl benzene | 0.1203 ± 0.0007 | 0.058 ± 0.001 | 0.0056 ± 0.0005 | +0.6 | +3.6 | −14 | 0.8 | 1.4 | 6.4 |
| *n*-Butyl benzene | 0.114 ± 0.002 | 0.059 ± 0.001 | 0.0054 ± 0.0005 | −3.0 | −5.8 | +22 | 0.8 | 2.0 | 4.0 |

"H," "I," and "L" refer to the high-resolution, the intermediate-resolution, and the low-resolution cases, respectively. The selectivity corresponds to the overall multivariate selectivity, which is the product of the selectivities in each dimension (Eq. (2)).

solvent delay and no mass spectra were collected during that time. The transfer line was held at 260 °C and the ion source was held at 200 °C. The detector voltage was −1600 V and the filament bias was −70 V. Mass spectra were collected from *m/z* 70–625 at a nominal rate of 5 kHz and averaged to 100 full spectra/second. Data were processed similarly to the butyl benzene isomers.

## 4. Results and discussion

### 4.1. High-resolution deconvolution

We begin with the TLD-initiated PARAFAC investigation of the butyl benzene isomer mixtures. To verify that there was sufficient selectivity among the mass spectra of the four isomers, a benchmark "high-resolution" case was constructed such that the four isomers had at least unit resolution in the column 1 dimension and a resolution of 0.6 in the column 2 dimension (Fig. 3A). It should be noted that since individual data sets were added together, the column 1 and column 2 times in the figure are for gauging relative retention times only and do not indicate real retention times. The column 1 and column 2 chromatographic peak profiles resulting from PARAFAC deconvolution of the high-resolution case yielded excellent quantitative results (Table 1) and excellent peak shapes (not shown for brevity). PARAFAC deconvolution of each dataset took about five minutes on a 1.5 GHz PC. The deconvoluted mass spectra (also not shown for brevity) were all matched to the appropriate library mass spectra with similarity match factors greater than 900 (where 999 is an exact match) and were generally within one standard deviation of the match factors obtained for the signals of the pure compounds prior to constructing the data sets. Standard deviations were determined using the three replicate data sets collected for each analyte. All matches listed in this report were the highest-ranking matches. The reproducibility and bias of the method for this high level of chromatographic resolution was studied by comparing the true (prior to addition) fraction of total peak volume for each peak to the PARAFAC analyzed peak volume fraction for the deconvoluted peak. Normalization of the three deconvoluted replicate peak volume fractions to the mean of the true volume fractions was necessary to eliminate apparent biases due to run-to-run injection variability. The bias is defined

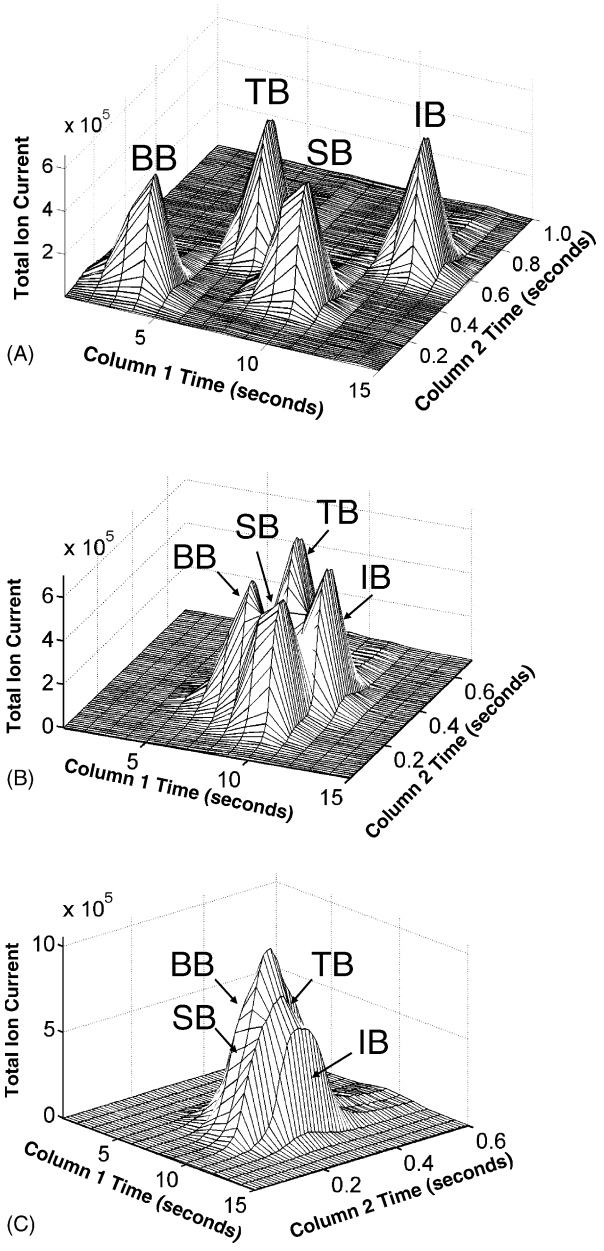


Fig. 3. Three-dimensional total ion current (TIC) image of the (A) high-resolution, (B) intermediate-resolution and (C) low-resolution constructed *iso*-butyl (IB), *sec*-butyl (SB), *tert*-butyl (TB), and *n*-butyl (BB) benzenes data set.

as the difference between a normalized reconstructed peak volume fraction and the mean true peak volume fraction divided by the mean true peak volume fraction. The absolute values of the biases are all less than 3% for the benchmark "high-resolution" case, which suggests an acceptable level of accuracy (Table 1). The R.S.D. (%) calculated and tabulated here is based upon the standard deviation of the normalized peak volume fractions divided by the mean of those fractions. The R.S.D.s (%) for the benchmark case are all less than 1% indicating a high level of precision (Table 1).

### 4.2. Intermediate-resolution deconvolution

In this case, the resolution between adjacent isomers on column 1 was nominally 0.25 and 0.2 on column 2 (Fig. 3B). After baseline correction, the three replicate intermediate-resolution constructed data sets were analyzed. Similar to the high-resolution data sets, TLD-initiated PARAFAC of the intermediate-resolution data sets produced good deconvoluted chromatographic peak shapes and mass spectra (also not shown for brevity), suitable for quantification (results in Table 1). The mass spectral match factors for the intermediate-resolution PARAFAC results indicated good qualitative identification of the components of interest with similarity match values greater than 900 achieved. The biases obtained and reported in Table 1 indicate a decreased level of accuracy compared to the high-resolution case with the biases ranging from −5.8 to +3.6%. The R.S.D.s (%) for the intermediate-resolution case are less than or equal to 2% (Table 1). This indicates that although there was some redistribution of signal from the PARAFAC deconvolution, the results are still quite precise even if the accuracy is suffering due to the low multivariate selectivity achieved by some isomers.

### 4.3. Low-resolution deconvolution

An extremely challenging case was then studied with the resolution between adjacent isomers on column 1 nominally 0.25 and a resolution of 0.1 on column 2 (Fig. 3C). Only 1 s on column 1 and only 20 ms on column 2 separated the peak maxima. The deconvoluted chromatographic peak shapes, shown for one data set in Fig. 4A and B, were quite satisfactory, although they exhibited slight deviations from the true peak shapes. The column 1 and 2 peak shapes resulting from the PARAFAC deconvolution of the high- and intermediate-resolution cases resulted in slightly better peak shapes than that shown in Fig. 4A and B. The peak shape deviations resulting from the low-resolution case also resulted in a more significant redistribution of signal, or bias, as summarized in Table 1. The biases indicate a decreased level of accuracy

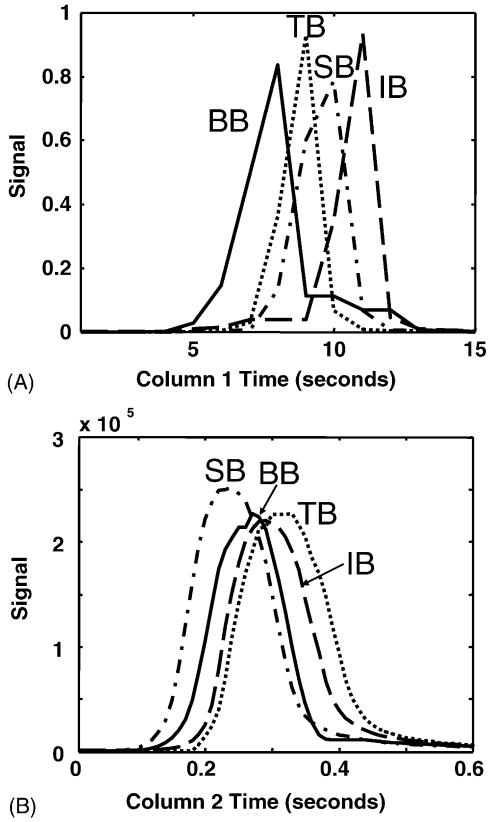


Fig. 4. (A) PARAFAC deconvoluted column 1 pure component profiles of the low-resolution isomer data set. (B) PARAFAC deconvoluted column 2 pure component profiles of the low-resolution isomer data.

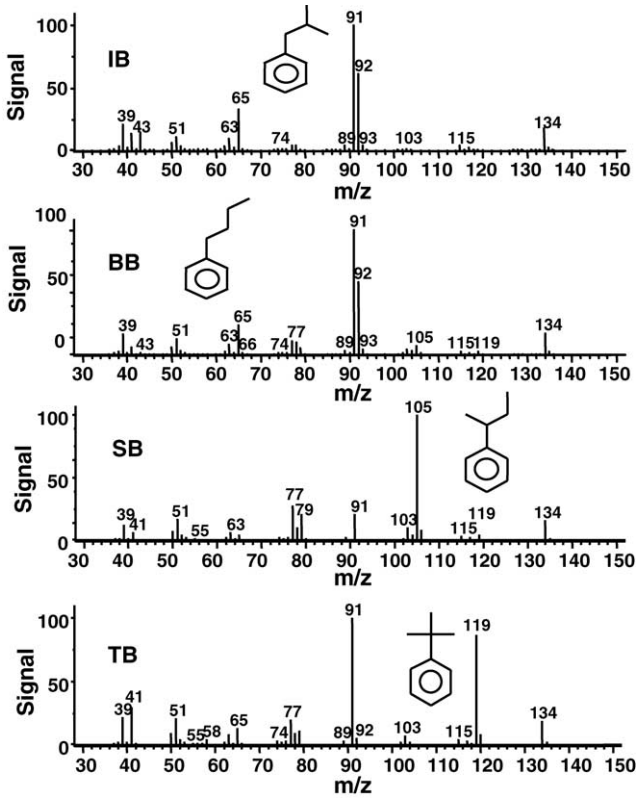


Fig. 5. Mass spectra resulting from PARAFAC deconvolution of the low-resolution constructed *iso*-butyl, *sec*-butyl (SB), *tert*-butyl (TB), and *n*-butyl (BB) benzenes data set, obtained simultaneously with plots in Fig. 4.

compared to the high-resolution case with the biases ranging from −14 to +22%. The R.S.D. (%) for the low-resolution case are less than 6% indicating more error in the modeling compared to the high-resolution case (Table 1). The deconvoluted mass spectra for the four butyl benzene isomers for the low-resolution case are shown in Fig. 5. The spectral match factors for the PARAFAC results indicated good qualitative identification with similarity match factors of 898 or greater even with the very low chromatographic resolution. For the low-resolution case, the comparisons between NIST library spectra and the four deconvoluted spectra display a high degree of similarity (with the NIST spectra not shown for brevity). Additionally, it can be seen by comparing *iso*-butyl and *n*-butyl benzenes (IB and BB of Fig. 5) that the pure component spectra are highly similar and in fact only achieve a multivariate selectivity of ∼0.12 each in the mass spectral dimension.

### 4.4. Multivariate selectivity

Similar analyses were carried out on six additional cases for a total of nine constructed butyl benzene isomer data sets with variable resolution in both chromatographic dimensions. Approximate resolutions for column 1 were 0.25, 0.6 and 1.0. Column 2 chromatographic resolutions were nominally 0.1, 0.2, and 0.6. At each of the three nominal column 1 resolutions, the four isomers were arranged at each of the three different column 2 resolutions resulting in the nine different cases, three of which were just discussed above in detail. PARAFAC was used to analyze the six additional cases as described above. Since each of the nine cases contained four components, this resulted in 36 data points with overall

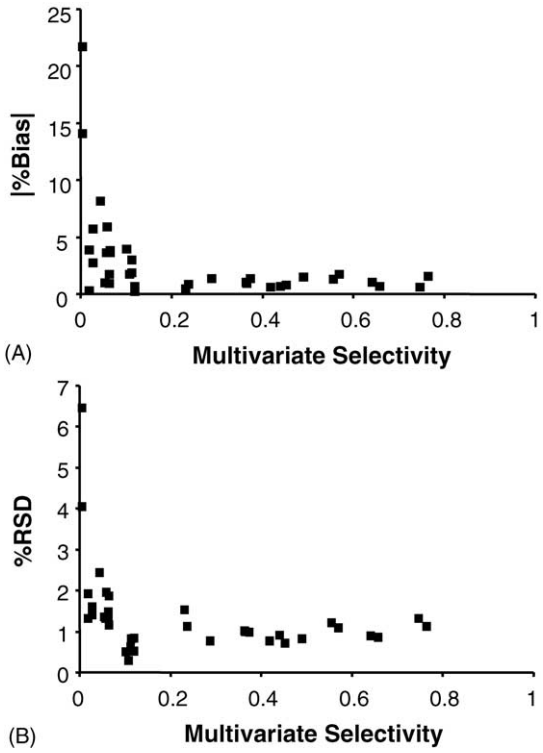


Fig. 6. (A) Plot of the absolute value of the bias (%) for the four compounds (*iso*-butyl, *sec*-butyl, *tert*-butyl, and *n*-butyl benzenes) in nine different chromatographic resolution cases vs. multivariate selectivity. Three replicates were obtained for each compound in each of the nine cases. The average standard deviation of the multivariate selectivities for all compounds and cases was ±0.007. (B) Plot of the R.S.D. (%) for the four compounds in nine different chromatographic resolution cases vs. multivariate selectivity.

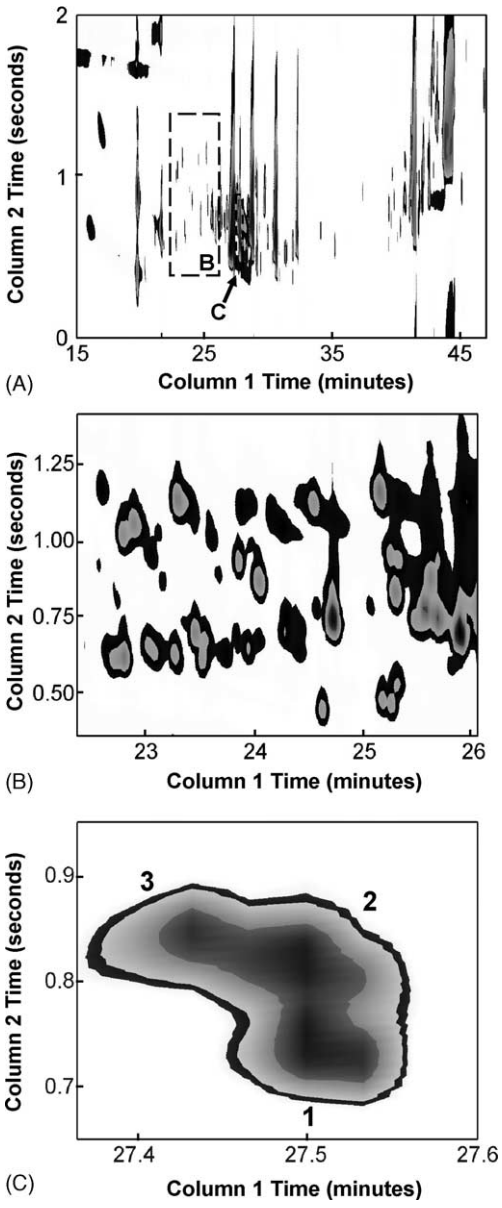


Fig. 7. (A) Chromatogram of *m*/*z* 73 of GC × GC–TOFMS analysis of trimethylsilylation derivative of Huilmo (*Sisyrinchium striatum*) plant extract. Region appearing in (B) is denoted by dashed box with label 'B'. Region used for PARAFAC analysis depicted in (C) is denoted by region labeled with 'C'. (B) Chromatogram of *m*/*z* 73 of a complex sub-region of the Huilmo (*Sisyrinchium striatum*) plant extract chromatogram. (C) Total ion current chromatogram of a region in the GC × GC–TOFMS analysis of trimethylsilylation derivative of Huilmo (*Sisyrinchium striatum*) plant extract that was analyzed with PARAFAC. The three overlapping signals were analyzed using a four component model.

multivariate selectivities ranging from 0.0054 to 0.768. The absolute value of the bias was plotted versus multivariate selectivity (Eq. (2)) for all four components in all nine cases (Fig. 6A). As can be seen from Fig. 6A, as the multivariate selectivity decreases from the maximum of ∼0.8, the bias results are all relatively constant, below a bias of 3%, until the selectivity reaches ∼0.12. At this point the results are more unreliable, although they still follow a general trend resulting in greater bias with less multivariate selectivity. A plot of R.S.D. (%) versus multivariate selectivity shows a similar trend (Fig. 6B); however, the R.S.D. (%) remains below 3% to a much lower level of multivariate selectivity (∼0.005). This indicates that PARAFAC results are consistent for replicate analyses, even if the accuracy is not as optimal. As can be seen from the butyl benzene isomer case studies, TLD-initiated PARAFAC analysis can result in highly accurate and precise results for well-resolved analytes, but it begins to suffer with extreme chromatographic overlap and low mass spectral selectivity. This emphasizes that some selectivity in each dimension needs to be present, and if there is only a very

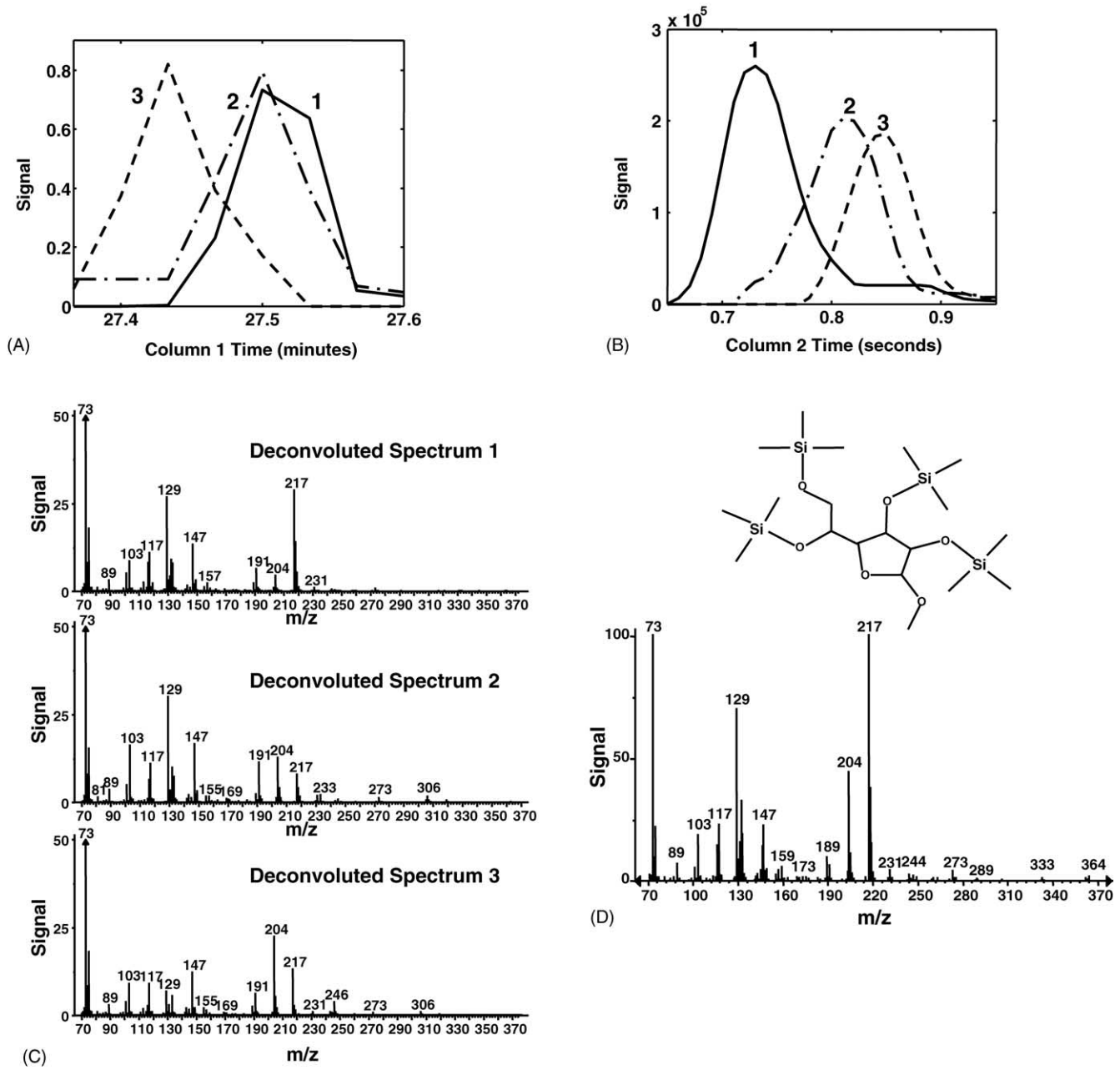


Fig. 8. (A) PARAFAC deconvoluted column 1 pure component profiles (component 4 consisting of baseline offset omitted for clarity), resulting in three deconvoluted species labeled 1, 2, and 3. (B) PARAFAC deconvoluted column 2 pure component profiles (component 4 consisting of baseline offset omitted for clarity). (C) Deconvoluted mass spectral profiles. (D) Best quality NIST library mass spectrum similarity match for deconvoluted species 1: methyl 2,3,5,6-tetrakis-*O*-(trimethylsilyl)-α-D-Glucofuranoside (match: 812, reverse: 819, probability: 23.6).

small amount of selectivity in all dimensions, the results will be less than optimal. As a practical guideline from this study, an overall multivariate selectivity of greater than 0.18 results in acceptably accurate and precise results for the deconvolution of overlapped signals. For most cases, there will be more mass spectral selectivity than that for *iso*-butyl and *n*-butyl benzenes (∼0.12), hence PARAFAC provides a useful tool to identify and quantify constituents in a sample.

### 4.5. Plant metabolite extracts

To demonstrate the deconvolution technique on a real, complex sample, metabolic plant extracts were analyzed. The derivatized metabolic extracts of Huilmo (*Sisyrinchium striatum)* result in a highly complex data set as seen in the chromatogram obtained at *m*/*z* 73, which is a characteristic ion for the trimethylsilyl group $-(CH_3)_3Si$ associated with the derivatization products (Fig. 7A). A relatively small portion of the entire chromatogram for selective mass channel *m*/*z* 73 is depicted in Fig. 7B to highlight the complexity of the sample. There are several instances of signal overlap in the sample where a multivariate deconvolution technique could be useful. A region of the sample that was analyzed with PARAFAC is shown in Fig. 7C. There are three overlapping components posing a challenging case for TLD-initiated PARAFAC, with very little chromatographic resolution between peaks labeled 1 and 2 on column 1 and between peaks 2 and 3 on column 2 (Fig. 7C). A TLD model was built with 4 components to account for baseline and background offsets (results not shown for brevity). Often in complex samples, the best results are obtained by giving the model an extra component to account for baseline and background offset. The TLD results were used to initiate PARAFAC with unimodal and non-negative constraints on column 1 and 2 and only non-negative constraints in the mass spectral dimension. PARAFAC deconvolution provided successful deconvolution of the three components of interest (Fig. 8A and B). The fourth component in the model consisting of baseline offset and noise was omitted for clarity. The deconvoluted mass spectra for the three analytes are depicted in Fig. 8C. Deconvoluted spectrum 1 resulted in a reasonable match with the derivatized monosaccharide methyl 2,3,5,6-tetrakis-*O*-(trimethylsilyl)-α-D-glucofuranoside (match: 812, reverse: 819, probability: 23.6) shown in Fig. 8D as a typical example. This and other sugars are probable components in plant metabolite extracts. The other two deconvoluted spectra did not result in as high a similarity match in the NIST mass spectral database most likely due to the obscurity of the compounds and/or the incompleteness of the database. Standards for these compounds were not currently available for this study. All three deconvoluted spectra were the most similar to isomeric derivatized monosaccharides akin to the one depicted in Fig. 8D. Most important here is that the deconvoluted spectra are very similar (Fig. 8C), do not contain selective ions, and are most likely isomers. Yet, PARAFAC was still able to successfully deconvolute the three compounds. Further analysis of standards and a user-compiled library is necessary for unambiguous identification and quantification of these compounds.

Based on the success of PARAFAC in this and other studies [20], a four-step method to identify and quantify analytes of interest is summarized here. Two data files are collected: (A) "sample" and (a) "sample + standard addition," in which all the analytes of interest are spiked into the standard addition case. The region around each analyte of interest in the sample data set and the standard addition data set are analyzed by TLD-initiated PARAFAC, resulting in individual chromatographic peak profiles and mass spectra for both data sets. The pure chromatographic retention times and mass spectra are selective data for identification of analytes. Comparing the deconvoluted mass spectra from the sample and the standard addition cases identifies the analytes in the sample data set. This could improve the quality of similarity matches because the reference spectrum was obtained on the same instrument as the sample as opposed to the NIST library spectra that are obtained on a number of different instruments resulting in different fragmentation ion ratios. Quantification occurs by reconstructing the analyte of interest in both the sample and the standard addition and applying the usual techniques for quantification via standard addition. The problems associated with small retention time shifts between sample and standard runs are eliminated because the deconvolution and identification are performed separately on both the sample and the standard addition, thus simplifying the method substantially.

## 5. Conclusions

It was demonstrated that the GC × GC–TOFMS trilinear data structure is compatible with third-order chemometric analysis techniques such as TLD and PARAFAC. This was true for both valve modulated and thermally modulated GC × GC instruments. The effects of altering chromatographic resolution on the results of a PARAFAC analysis was investigated with constructed data sets of four isomers that exhibit similar mass spectra. Two of the compounds had extremely similar mass spectra with mass spectral multivariate selectivities of only ∼0.12. It was shown that PARAFAC leads to successful qualitative identification of isomeric analytes with a wide range of selectivities (0.0054–0.768), but quantitative results were better for cases where the overall multivariate selectivity for a given analyte was greater than 0.18. The ability of PARAFAC to successfully deconvolute isomers was demonstrated on three overlapping species of possibly isomeric monosaccharide derivatives from a complex plant metabolite sample of Huilmo (*Sisyrinchium striatum*). Quantification of analytes of interest, including isomers, can therefore be identified and quantified using PARAFAC without selective ions, peak shape predictions or retention time alignment between the sample and the standard prior to analysis. PARAFAC is easily automated and could potentially be applied to complete GC × GC–TOFMS chromatograms in the way NIST's automated mass spectral de-

convolution and identification system (AMDIS) [34,35] is applied to GC–MS chromatograms. AMDIS takes a GC–MS chromatogram, identifies the location of all peaks, deconvolutes the unresolved peaks and then searches a library for matching compounds. A version of AMDIS extended to GC × GC–TOFMS and PARAFAC could work as complementary peak deconvolution techniques. While AMDIS relies on selective ions, PARAFAC relies on some selectivity in each dimension.

## Acknowledgments

This work was supported, in part, by the Center for Process Analytical Chemistry (CPAC), a National Science Foundation, University/Industry Cooperative Center at the University of Washington. The metabolites analysis portion of this work was funded in part by a grant from the National Science Foundation (DMI-0320427). The authors would also like to thank Neil Gallagher (Eigenvector Research) for insights into third-order chemometrics.

## References

[1] M. van Deursen, J. Beens, J. Reijenga, P. Lipman, C. Cramers, J. Blomberg, HRC-J. High Res. Chrom. 23 (2000) 507.
[2] R. Shellie, P.J. Marriott, P. Morrison, Anal. Chem. 73 (2001) 1336.
[3] J. Dallüge, M. van Rijn, J. Beens, R.J.J. Vreuls, U.A.Th. Brinkman, J. Chromatogr. A 965 (2002) 207.
[4] J. Dallüge, R.J.J. Vreuls, J. Beens, U.A.Th. Brinkman, J. Sep. Sci. 25 (2002) 201.
[5] J. Dallüge, L.L.P. van Stee, X. Xu, J. Williams, J. Beens, R.J.J. Vreuls, U.A.Th. Brinkman, J. Chromatogr. A 974 (2002) 169.
[6] A.E. Sinha, B.J. Prazen, C.G. Fraga, R.E. Synovec, J. Chromatogr. A 1019 (2003) 79.
[7] J.-M.D. Dimandja, Am. Lab. 35 (2003) 42.
[8] P. Korytar, L.L.P. van Stee, P.E.G. Leonards, J. de Boer, U.A.Th. Brinkman, J. Chromatogr. A 994 (2003) 179.
[9] X. Lu, J. Cai, H. Kong, M. Wu, R. Hua, M. Zhao, J. Liu, G. Xu, Anal. Chem. 75 (2003) 4441.
[10] Z. Lin, K.S. Booksh, L. Burgess, B.R. Kowalski, Anal. Chem. 66 (1994) 2552.
[11] K.S. Booksh, Z.H. Lin, Z.Y. Wang, B.R. Kowalski, Anal. Chem. 66 (1994) 2561.
[12] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Anal. Chem. 70 (1998) 2796.
[13] B.J. Prazen, C.A. Bruckner, R.E. Synovec, B.R. Kowalski, Anal. Chem. 71 (1999) 1093.
[14] C.G. Fraga, B.J. Prazen, R.E. Synovec, HRC-J. High Res. Chrom. 3 (2000) 215.
[15] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 72 (2000) 4154.
[16] C.G. Fraga, C.A. Bruckner, R.E. Synovec, Anal. Chem. 73 (2001) 675.
[17] R.E. Synovec, B.J. Prazen, K.J. Johnson, C.G. Fraga, C.A. Bruckner, Advances in Chromatography, Marcel Dekker, Inc., New York, 2003.
[18] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782A.
[19] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 73 (2001) 5833.
[20] A.E. Sinha, C.G. Fraga, B.J. Prazen, R.E. Synovec, J. Chromatogr. A 1027 (2004) 269.
[21] K.J. Johnson, R.E. Synovec, Chemom. Intell. Lab. Syst. 60 (2002) 225.
[22] B.J. Prazen, K.J. Johnson, A. Weber, R.E. Synovec, Anal. Chem. 73 (2001) 5677.
[23] K. Faber, A. Lober, B.R. Kowalski, J. Chemom. 11 (1997) 419.
[24] A. Lorber, K. Faber, B.R. Kowalski, J. Chemom. 10 (1996) 215.
[25] E. Sánchez, B.R. Kowalski, J. Chemom. 4 (1990) 29.
[26] S. Li, P.J. Gemperline, J. Chemom. 7 (1993) 77.
[27] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149.
[28] N.K.M. Faber, R. Bro, P.K. Hopke, Chemom. Intell. Lab. Syst. 65 (2003) 119.
[29] N.M. Faber, Anal. Bioanal. Chem. 372 (2002) 683.
[30] A.E. Sinha, K.J. Johnson, B.J. Prazen, S.V. Lucas, C.G. Fraga, R.E. Synovec, J. Chromatogr. A 983 (2003) 195.
[31] C.A. Andersson, R. Bro, Chemom. Intell. Lab. Syst. 52 (2000) 1.
[32] M.A. Adams, Z. Chen, P. Landman, T.D. Colmer, Anal. Biochem. 266 (1999) 77.
[33] O. Fiehn, J. Kopka, R.N. Trethewey, L. Willmitzer, Anal. Chem. 72 (2000) 3573.
[34] J.M. Halket, A. Przyborowski, S.E. Stein, W.G. Mallard, S. Down, R.A. Chalmers, Rapid Commun. Mass Spectrom. 13 (1999) 279.
[35] S.E. Stein, J. Am. Soc. Mass Spectrom. 10 (1999) 770.