# Development of Instrumental and Computational Methods for Accessing Information in Multi-Dimensional Gas Chromatography with Mass Spectrometry

Nathanial E. Watson

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Robert E. Synovec, Chair

Matthew F. Bush

Ashleigh B. Theberge

Program Authorized to Offer Degree:

Chemistry

University of Washington

## Abstract

Development of Instrumental and Computational Methods for Accessing Information in Multi-Dimensional Gas Chromatography with Mass Spectrometry

Nathanial E. Watson

Chair of the Supervisory Committee:
Professor Robert E. Synovec
Department of Chemistry

Three instrumental and computational methods are demonstrated in an endeavor to create novel techniques to extract information from the troves of data generated by multi-dimensional gas chromatography with mass spectrometry. Initially, these methods are considered within the context of targeted and non-targeted experimental design. The tile-based Fisher Ratio with null distribution analysis is first evaluated and validated within the non-targeted realm. The method is shown to be fast and accurate. Forty-six of the fifty-four benchmarked changing metabolites previously discovered were found by the new methodology while consistently excluding all but one of the benchmarked nineteen false positive metabolites previously identified. This was achieved in less than 5% of the time required for the previous method. Later, the three-

dimensional gas chromatograph is improved to include mass spectrometric detection. This instrument provides four dimensions (4D) of chemical selectivity and includes significant improvements to total selectivity (mass spectrometric and chromatographic), peak identification, and operational temperature range relative to previous models of the $GC^3$ reported. Useful approaches to visualize the 4D data are presented. The $GC^3$ - TOFMS instrument experimentally achieved total peak capacity, $n_{c,3D}$, ranging from 5000 to 9600 ($\bar{x}$= 7000, $s$ = 1700) for 10 representative analytes for 50 min separations with component dimensional peak capacities averaging 406, 3.6, and 4.9 for $^1$D, $^2$D, and $^3$D, respectively. Using this instrument and the well understood Parallel Factor Analysis (PARAFAC) model a new option for targeted analysis is presented. Conceptualizing the $GC^3$ - TOFMS as a one-dimensional gas chromatograph with GC × GC-TOFMS detection the instrument was allowed to create the PARAFAC target window natively. Each first dimension modulation thus created a full GC × GC-TOFMS chromatogram totally amenable to PARAFAC. A simple mixture of 115 compounds and a diesel sample were interrogated through this methodology. All test analyte targets were successfully identified in both mixtures. In addition, mass spectral matching of the PARAFAC loadings to library spectra yielded results greater than 900 in 40 of 42 test analyte cases. Twenty-nine of these cases produced match values greater than 950.

# TABLE OF CONTENTS

List of Figures ..... iv

List of Tables ..... vi

Chapter 1. Introduction to Data Analysis in Multi-Dimensional Gas Chromatography and Mass Spectrometry ..... 1

1.1 Scope ..... 1

1.2 Data Management and Analysis of Variance ..... 3

1.3 Deconvolution ..... 4

1.4 Classification and Comparison ..... 8

1.5 Statistical Presentation of Useful Information ..... 9

1.5.1 Targeted Analysis ..... 9

1.5.2 Non-targeted Analysis ..... 10

1.5.3 Hybrid Targeted/Non-targeted Analysis ..... 11

1.6 Hypotheses ..... 12

1.6.1 Chapter 2: Performance Evaluation of Tile-Based Fisher Ratio Analysis using a Benchmark Yeast Metabolome Dataset ..... 12

1.6.2 Chapter 3: Comprehensive Three-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry ..... 13

1.6.3 Chapter 4: Targeted Analyte Deconvolution and Identification by Four-Way Parallel Factor Analysis Using Three-Dimensional Gas Chromatography with Mass Spectrometry 13

1.7 References ..... 13

Chapter 2. Performance Evaluation of Tile-Based Fisher Ratio Analysis using a Benchmark Yeast Metabolome Dataset ..... 18

2.1 Introduction ..... 18

2.2 Experimental ..... 24

2.2.1 Sample Prep and Data Collection ..... 24

2.2.2 Data Analysis ..... 26

2.3 Results and Discussion ..... 28

2.4 Conclusions ..... 42

2.5 References ..... 43

Chapter 3. Comprehensive Three-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry ..... 47

3.1 Introduction ..... 47

3.2 Basic Principles and Instrumental Design ..... 50

3.3 Experimental ..... 52

3.4 Results and discussion ..... 56

3.5 Conclusions ..... 63

3.6 References ..... 63

Chapter 4. Targeted Analyte Deconvolution and Identification by Four-Way Parallel Factor Analysis Using Three-Dimensional Gas Chromatography with Mass Spectrometry Data ..... 67

4.1 Introduction ..... 67

4.2 Experimental ..... 70

4.2.1 Data Collection ..... 70

4.2.2 Data Analysis ..... 72
4.3 Results and Discussion ..... 73
4.4 Conclusion ..... 83
4.5 References ..... 84
Chapter 5. Conclusion ..... 87
Bibliography ..... 90

# LIST OF FIGURES

Figure 1.1. Targeted Analysis ..... 2
Figure 1.2. Non-targeted Analysis ..... 2
Figure 2.1. Representative (A) GC × GC–TOFMS chromatogram of m/z 73 and zoom (B) to region in box. ..... 24
Figure 2.2. Extract of the cystathionine chromatographic elution event at *m/z* 278 represented from various aspect frames. These figures graphically depict the logic behind selecting the 4 modulation by 300 ms tile size. ..... 29
Figure 2.3. F-ratio distribution of the true class comparison for the first replicate data subset. ..... 32
Figure 2.4. Samples composing the true class comparison for the first replicate data subset along with one of the 200 null classes created for determination of an F-ratio threshold.. 33
Figure 2.5. Probability distributions for the first replicate subset..... 34
Figure 2.6. Average DR chromatogram from Fig. 2.1(A) with true hits indicated by circles after removal of redundant hits and matrix artifacts ..... 36
Figure 3.1 Schematic of the major components of the GC3 – TOFMS instrument. ..... 52
Figure 3.2. Chromatogram for adamantane viewed in all dimensions. ..... 55
Figure 3.3. Output from the $GC^3$ – TOFMS for the 115 component test mixture ..... 58
Figure 3.4. Output from $GC^3$ – TOFMS for diesel spiked with a mixture of non-native compounds ..... 60
Figure 4.1. Schematic detail of $GC^3$-TOFMS data structure. ..... 73
Figure 4.2. Schematic detail of the method for targeted analyte finding and deconvolution using GC3-TOFMS and four-way PARAFAC. ..... 74
Figure 4.3. Four-way PARAFAC results for the $^1$D region from Fig. 3 containing 2-nonanone and undecane. ..... 76
Figure 4.4. Method transfer from a standard mix (e.g. Figures 4.1-4.3) to a sample of diesel containing spikes of several non-native compounds..... 78

Figure 4.5. Four-way PARAFAC results on the region from Figure 4.4 containing cyclohexyl benzene (CHB)..... 79
Figure 4.6. Four-way PARAFAC results for ethyl salicylate..... 81
Figure 4.7. Four-way PARAFAC results for 2-decanone..... 82

# LIST OF TABLES

Table 2.1. Sample composition for each replicate data subset. ..... 26
Table 2.2. Results of tile-based F-ratio analysis for metabolites present in both classes. 37
Table 2.3 Results of tile-based F-ratio analysis for metabolites present in only one class.38
Table 2.4. Statistically verified false positives from the previous pixel-based F-ratio analysis [37] ..... 39
Table 3.1. 115 Component Text Mixture Composition ..... 53
Table 3.2. Non-native components spiked into diesel. ..... 54
Table 3.3 Chromatographic peak measurements and figures-of-merit. ..... 57
Table 4.1. Summary of four-way PARAFAC results for six target analytes portrayed in Figures 4.1-4.3. ..... 77
Table 4.2. Summary of four-way PARAFAC results for 18 non-native target analytes spiked in diesel and an additional 18 target analytes native to diesel. ..... 80

# DEDICATION

For the wife who puts up with me.

# Chapter 1. Introduction to Data Analysis in Multi-Dimensional Gas Chromatography and Mass Spectrometry[1]

## 1.1 SCOPE

Data generated by one- and two-dimensional chromatography coupled with mass spectrometry is inherently complex because of the size and multidimensional aspects of the information collected. As technology improves, mass spectrometers can detect at higher mass resolutions with faster scan rates, generating more data per chromatogram than ever before. The complexity of this data, whether it be two-dimensional (2D) data from a one-dimensional separation coupled with mass spectrometry or three-dimensional (3D) data from a comprehensive two-dimensional separation with mass spectrometric detection (eg., GC × GC – MS), requires computer management and interpretation techniques to provide rapid and robust information about the chemical and/or system in question.

Data analysis strategies are best understood by placing them within the larger context of scientific investigations. The broader scientific questions to be addressed by specific research studies dictate the sample preparation methods, the instrumental platform and parameters, and thus the subsequent data analysis techniques employed. While many data analysis methods exist, including both commercially available and academically developed "in-house" software packages, these data analysis methods are not necessarily appropriate for a given scientific study. Experiments generally fall into one of three categories: (1) targeted studies, in which the

[1] Portions of this Chapter have been adapted from B.C. Reaser, N.E. Watson, K.J. Skogerboe, and R.E. Syonvec, "Management and Interpretation of Capillary Chromatography – Mass Spectrometry Data" from Hyphenations of Capillary Chromatography with Mass Spectrometry, Edited by Luigi Mondello and Peter Tranchida (2017), *currently in preparation.*

identities of analytes of interest are known; (2) non-targeted studies, in which the identities of the analytes of interest are not presumed; and (3) a mixture of (1) and (2), in which the identities of some analytes of interest are known, but information about other unknown analytes may be of interest as well.

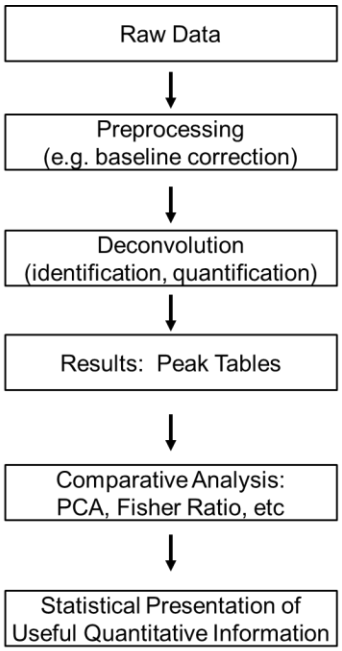


Figure 1.1. Targeted Analysis

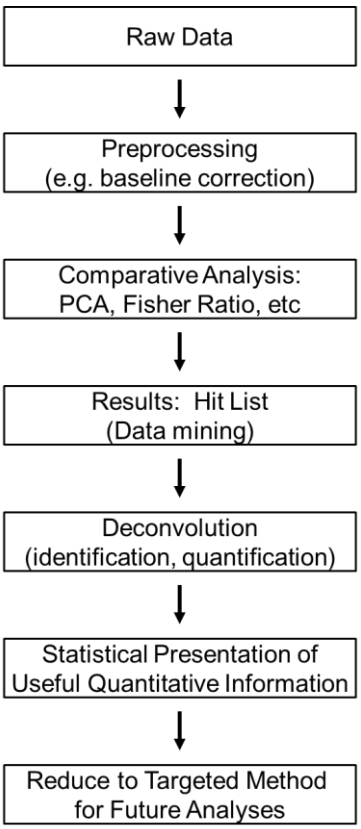


Figure 1.2. Non-targeted Analysis

While the overall data analysis approaches for targeted and non-targeted investigations differ, many of the steps in the analytical workflow are the same. Figure 1.1 and Figure 1.2 show examples of sample workflows for the data in targeted and non-targeted analyses, respectively. The two methods both start with the raw data and pre-processing, but quickly diverge. Both methods involve deconvolution, results determination, comparative analysis and statistical presentation steps, but the order of these steps result in matrices that do not commute. This chapter will discuss some of the many different analytical methods and tools available for data

matrix operations [1] and in conclusion will propose several new ideas which contribute toward the greater body of knowledge in this domain.

## 1.2 Data Management and Analysis of Variance

The term "data management" can have various definitions depending on the context and point of view of the investigator. Here the term is used to describe how one may go about the formatting, storage, access and exchange of the raw data after collection of a chromatogram. Most commercial instrument software includes a file management system for naming, time-and-date stamping, saving, and exporting data either as a proprietary format (e.g., a ".peg" file that can only be analyzed with the LECO ChromaTOF software) or a more universal format (e.g., a ".pdf" or ".csv" file) that can be used with other software programs. Whether a commercially available software incorporates these file management options, it is important for the analyst to plan ahead and to keep detailed records of their data acquisition and storage system. Experiments with multiple samples and injection replicates of comprehensive two-dimensional chromatographic-mass spectrometric data can easily reach several gigabytes (GB) in size, requiring computers with sufficient processors, random-access memory (RAM) and large enough hard drives to be able to acquire, save, and process the data [2–4].

When considering aspects of data management, it is also useful to consider experimental design. Specifically, employing the fewest measured variables to elucidate the various contributors to experimental variance is useful as an initial technique to limit the amount of data produced. A preliminary Analysis of Variance (ANOVA) [5] can be useful to determine all the various contributors to the total experimental variance. Commonly, the analyst will find that one category of variance significantly swamps the other contributors. Once a source of variation has been sufficiently characterized, it may no longer be necessary to always estimate that source of

variation in each set of subsequent analyses. This is typically the case for small, consistent sources of variance that can be disregarded in the shadow of much larger sources of variance.

For example, biological variance may be on the order of 20% relative standard deviation (RSD) or more across multiple samples that are nominally identical, while variance in volume of sample introduced to the instrument is typically no more than 5–10% when using an autosampler [6]. In complex analyses, variances are cumulative, i.e., errors caused at an earlier step are further added upon (and possibly confounded by) variances introduced later in sample handling or analysis. Ultimately, for purposes of data analysis, with the possible exception of preliminary studies to initiate an experimental campaign, only the cumulative variance is of importance. Consistent sources of variance can be minimized using quality control and standard runs in lieu of collecting superfluous replicates of statistically insignificant variance. The size of the total data set can be reduced by 50% or more through replication of only the major sources of variance. Preliminary quantification of all sources of variance excludes the requirement to collect statistically relevant numbers of replicates on insignificant sources of variation.

## 1.3 DECONVOLUTION

Deconvolution is the process of mathematically resolving two or more analytes whose peaks are otherwise chromatographically overlapped. Deconvolution steps are performed during most analyses to identify analyte peaks and produce more accurate and precise peak attributes. This includes such features as peak height, area, and retention time, as well as a more accurate mass spectrum for each of these peaks than can be obtained from, for example, simple integration across a peak. If all the peaks in a chromatogram are sufficiently resolved with a signal-to-noise ratio (*S/N*) greater than ~10, simple methods suffice and deconvolution is not necessary. This is not typically the case, especially for complex samples such as those found

regularly analyzed by one- and two-dimensional chromatography coupled with mass spectrometric detection. Many methods, with varying levels of automation, have been developed to perform deconvolution for specific instrument platforms and/or specific applications [7,8]. While this chapter aims to discuss the most effective and widely-used deconvolution software packages and chemometric methods, it would be impossible to do a comprehensive and thorough review of all possible available deconvolution software packages and methods [9]. Thus, the challenges and opportunities for several different packages are overviewed.

Most analysts employing chromatography coupled with mass spectrometric detection utilize commercial instruments with software programs written and sold by the instrument companies to both operating the instrument and analyzing the data acquired. Most of these commercial software packages include deconvolution algorithms for ease of use. For example, Agilent (Santa Clara, CA, USA) makes commercially available software called ChemStation to go with their GC-MS and LC-MS systems that both utilize deconvolution programs. The GC-MS ChemStation deconvolution program uses the Automatic Mass Spectral Deconvolution Identification System (AMDIS) developed by Stein [10] at the National Institute of Standards and Technology (NIST; Gaithersburg, MD, USA). AMDIS relies on the peak shape modeled by pure selected ion chromatograms to extract individual mass spectral peak profiles with similar shapes to produce a pure mass spectrum and can deconvolute coeluting peaks down to a chromatographic resolution of about 0.3 [10]. Most other commercially available software programs utilize proprietary deconvolution methods. LECO's commercial software, ChromaTOF (LECO Corp, Saint Joseph, MI, USA) utilizes a proprietary deconvolution method called True Signal Deconvolution® for GC-TOFMS and GC x GC – TOFMS data. MARKES International's (Gold River, CA, USA) TargetView for GC-MS deconvolution and TOF-DS for deconvoluting

one- or two-dimensional GC coupled with mass spectrometric detection and Waters' (Milford, MA, USA) ChromaLynx for LC-MS, GC-MS and GC or LC with tandem MS detection also use proprietary methods. Many instrument companies also suggest the use of GC Image (Lincoln, NE, USA), a software development company that has proprietary software deconvolution for gas and liquid chromatography in one and two dimensions coupled with MS detection [2]. Most of these software packages are built to be robust, reproducible and easy to use by analysts with or without formal chemometrics training.

In addition to proprietary methods employed by commercial software packages, there are also several popular chemometric techniques for the deconvolution of chromatographically overlapped peaks. These methods are commercially [11–13] or freely available on the internet [14,15]. Particularly interesting with respect to the studies reported herein is Parallel Factor Analysis (PARAFAC).

Multiway chromatographic instruments with mass spectrometric detection allow access to deconvolution techniques which require three or more dimensions which each respond linearly with concentration, that is, the data are both three-way and trilinear. Trilinear data is natively generated by two serially coupled chromatographic dimensions with mass spectrometric detection. Trilinear data can also be created through replicate stacking of single chromatographic column data with mass spectrometric detection.

PARAFAC can mathematically decompose the trilinear data into individual chromatographic and mass spectral vectors which recreate the original data matrix when multiplied. Mathematically, the PARAFAC model is described as

$$\boldsymbol{R} = \sum_{i=1}^{n} x_i \times y_i \times z_i + \boldsymbol{E} \qquad (1.1)$$

where **R** represents the detected three-way chromatogram. Each of the *n* total components is resolved into **x**, **y**, and **z** vectors which represent the $^{1}$D, $^{2}$D and mass spectral profiles respectively. Any remaining signal is retained in a residuals matrix, **E**. Numerous studies have been published on this technique and the reader is directed to these excellent resources for more details on PARAFAC [16–19] and an enhanced version called PARAFAC2 [20,21]. Employing PARAFAC is generally advantageous in cases where the trilinearity requirement is sufficiently followed [22,23]. As the signal response diverges from the trilinearity criterion, the PARAFAC result deviates, and becomes less accurate. It is also considered best practice for the analyst to create multiple models of varying numbers of factors and select the model which best represents the data.

Generally, deconvolution methods consist of approaches that rely on peak shape or apex identification, approaches that use multivariate resolution techniques, or hybrids of these two approaches. These various strategies each have different advantages and drawbacks, e.g., some cannot easily deal with isotopically labeled analytes that one might use in a metabolomics study, some require selective mass channels, some require multiple samples, some require higher *S/N* than others, some require specific input parameters including *a priori* knowledge of analyte characteristics (peak shape, peak concentration, mass spectrum, etc.), some require bilinearity or trilinearity, and some take more time than others. Ignoring these differences, we can look at how deconvolution fits into the two different approaches to data analysis per Figure 1.1 and Figure 1.2, those applying deconvolution on chromatographic data before performing comparative analysis, or performing comparative analysis prior to deconvolution.

## 1.4 CLASSIFICATION AND COMPARISON

Comparative analysis encompasses all the techniques, including clustering, classification, etc, an analyst may employ to classify or sort data and/or results into useful groupings. In many cases data is collected in a manner where the classes are predefined as part of the experimental design. For example, a biologist subjects a culture of cells to some perturbation while growing an equivalent culture which is not subjected to this perturbation. This kind of experiment naturally sorts into "control" and "experiment" classes. In other cases, these groupings may not be known. In both circumstances (*a priori* or post-experiment categorization) a variety of classification and data reduction techniques are available. Generally, all these techniques center on using sample variance to determine classes or using predefined classes to identify sources of variance between those classes.

The Fisher Ratio (F-ratio) is a simple and useful relation to determine areas of maximum variance between predefined classes. Experimentally these classes are usually defined by conditions; typically "control" and "experimental" classes are convenient and useful groupings for this form of analysis. Mathematically described as

$$F - ratio = \frac{\sigma_{bc}^2}{\sigma_{wc}^2} \tag{1.2}$$

where $\sigma_{bc}^2$ is the variance present between the classes and $\sigma_{wc}^2$ is the average variance within the classes. This returns a value that scales in magnitude from zero to infinity as the variance between classes increases relative to the within class variance. Initially described by Fisher [24], this procedure quantifies regions of data which have particularly large between class variance. Usually these regions will correspond to areas of interest between the two classes as defined.

When the analyst defines the classes as perturbed (or "experimental") and not perturbed (or "control") these regions of high F-ratio will generally correspond to analysts of interest.

The F-ratio procedure can be applied directly to multi-dimensional chromatographic data with mass spectrometric detection [3,25,26]. In these cases, the analyst may choose to apply various binning schemes before application to ensure minimization of the within class variance. The resultant F-ratio chromatogram can be sorted by F-ratio magnitude for detailed analysis of regions of interest. These regions may sometimes be directly identified and quantified. Other cases will apply the F-ratio as a feature selection step to reduce superfluous data before further analysis to ensure only experimentally useful variance (i.e., key chemical features) are considered.

One such novel example of a binning F-ratio method is the tile-based Fisher ratio method for GC × GC – MS [27,28]. This method applies a novel tiling scheme that aims to optimally capture a given analyte peak despite any retention time shift on either the first or second dimension. The method then uses the signal of the tile to compute the F-ratio value of the analyte or chemical feature encompassed by the tile. This method has been successfully applied to the analysis of fuels [29] and metabolomics [30], and provides a robust method for binning and analyzing 2D chromatograms without alignment.

## 1.5 Statistical Presentation of Useful Information

### 1.5.1 *Targeted Analysis*

For targeted analyses, as shown in Figure 1.1, peak defining methods immediately follow the pre-processing steps. Peak defining methods may include peak finding, peak deconvolution, peak identification and peak quantification via signal or peak area. The experimental design of a

targeted investigation generally includes the analysis of standards of the analytes of interest to accurately match the retention times and unique mass spectrum of each targeted analyte in the actual samples. Once peaks of interest are discovered, deconvolution of analytes that coelute with interferents found in the sample matrix may be required. These results are generally viewed first as a peak table. For all subsequent injection replicates or samples that must be analyzed, a new peak table is generated. Each table must subsequently be reviewed to locate each analyte of interest and to adjust the retention time(s) so they are uniform if the alignment was insufficient. After these steps, comparative analyses such as Principal Component Analysis (PCA) or Partial Least Squares (PLS) are often performed on the generated peak information. The reader is directed to some recent research investigations using targeted analysis in this way: Zhu et al. utilized LC-MS/MS and PLS-DA for the analysis of specific metabolites for colorectal cancer screening [31]; Yang et al. utilized both LC – MS/MS and GC × GC – TOFMS in a targeted metabolomics study of *Methylobacterium extorquens* AM1, applying PARAFAC to accurately quantify metabolites of interest [32]; Mi et al. analyzed 26 specific steroids in eggs using LC-MS/MS, first quantifying the steroids and then comparing the various egg types via PCA and HCA [33]; and Sgorbini et al. performed a targeted analysis of biologically active compounds in herbs and spices using GC-MS and PCA [34].

### 1.5.2 *Non-targeted Analysis*

Non-targeted analyses, as shown in Figure 1.2, generally consist of comparative analyses after pre-processing but prior to the peak-defining methods [35]. Rather than a peak table, as for the targeted method, these methods generally result in a "hit list" of chromatographic locations which are indicative of unique or interesting chemical features. In some cases, non-targeted analyses may not include the analysis of standards, so retention times or mass spectra of

standards may not be available to match to. This can make both identification and deconvolution of chemical features more complex. But these methods can also generate new and exciting results about data sets that were previously restricted to targeted analytes. For example, Schmarr and Bernhardt analyzed volatiles from fruits using image analysis and PCA on GC × GC – MS data [36], Cordero et al. analyzed hazelnut volatiles using non-targeted methods and GC × GC – MS [37], Brokl et al. analyzed volatiles in tobacco smoke via GC × GC – MS using ANOVA to optimize their method and both PCA and F-ratio analysis to find statistical meaning in their data set [38]; and Parsons et al. performed a non-targeted, discovery-based analysis of the acid alteration of diesel fuel using GC × GC – MS and the tile-based F-ratio method [29].

### 1.5.3 *Hybrid Targeted/Non-targeted Analysis*

Hybrid techniques include aspects of both targeted and non-targeted methods to the benefit of both previously discussed analysis methods [39,40]. Some combination techniques seek to find unidentified and unknown information regarding their sample beyond what is already known or expected [41]. In other cases, a non-targeted method can be used on initial samples to elucidate patterns in the data that point towards specific features or analytes that should be analyzed in a more targeted way with subsequent samples. This helps the analyst focus on the most meaningful information to save time and diminish the number of analytes that need to be deconvoluted and/or quantified. Magagna et al. studied olive oil via GC × GC – MS in this way using the GCImage software and PCA [39]; Gorrochategui et al. provide a workflow for metabolomics studies via LC-MS, which include both classification and deconvolution techniques [40]; Sinanian et al. demonstrate the use of MCR-ALS on high resolution LC-MS data using both targeted amphetamine samples and non-targeted bacterial lipid samples [42].

## 1.6 HYPOTHESES

Three novel concepts are presented which contribute new options in the quest for faster and simpler data management and analysis methods specifically applicable to gas chromatography and mass spectrometry. There are also some implications for the general field of separation science. One evaluates the tile-based F-ratio [27,28] against a well described dataset yielding both a validation of the technique and some options for decreasing the total sample replication necessary for statistical validity. The other two hypotheses focus on new developments in comprehensive three-dimensional gas chromatography [43,44] which aim to both improve the instrument hardware and provide a practical option for implementation of the method for targeted analysis.

### 1.6.1 *Chapter 2: Performance Evaluation of Tile-Based Fisher Ratio Analysis using a Benchmark Yeast Metabolome Dataset*

First, the tile-base F-ratio is evaluated against a well described yeast metabolome dataset [45–47]. The yeast dataset was analyzed in detail and yielded a wealth of information, but at the cost of significant time investment by the analyst. New developments in the tile-base F-ratio combined with null distribution analysis suggest that similar results as previously published could be produced with a significant reduction in human time investment. It is hypothesized that the application of the tile-based F-ratio combined with null distribution analysis will successfully generate the previously determined hit list in less than 10% of the initial time requirement.

### 1.6.2 *Chapter 3: Comprehensive Three-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry*

Secondly, hardware becomes the focus in an effort to provide new and faster tools to extract information from the ever-increasing trove of analytical data. Previously, the now common GC × GC was modified to include a third chromatographic dimension [43]. This provides significant benefits to the total chemical selectivity [44]. However, there remain significant limitations of the presented design using a single channel flame ionization detector. Here the hypothesis is that the addition of a TOFMS will create a novel data structure which solves problems of analyte determination and identification inherent with single channel detectors.

### 1.6.3 *Chapter 4: Targeted Analyte Deconvolution and Identification by Four-Way Parallel Factor Analysis Using Three-Dimensional Gas Chromatography with Mass Spectrometry*

Finally, the modified GC × GC × GC with TOFMS detection is turned toward targeted analyte discovery and identification. Based on previous successes with PARAFAC applied to $GC^3$ [43], the same technique is applied to $GC^3$-TOFMS. This targeted application builds upon previous work with PARAFAC in GC × GC – TOFMS [48–50]. It is hypothesized that the novel four-dimensional data structure generated by the $GC^3$-TOFMS will provide flexibility in defining a target window which will greatly simplify targeted analysis with PARAFAC.

## 1.7 REFERENCES

[1] S. Castillo, I. Mattila, J. Miettinen, M. Orešič, T. Hyötyläinen, Data Analysis Tool for Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry, Anal. Chem. 83 (2011) 3058–3067. doi:10.1021/ac103308x.

[2] S.E. Reichenbach, M. Ni, V. Kottapalli, A. Visvanathan, Information technologies for comprehensive two-dimensional gas chromatography, Chemom. Intell. Lab. Syst. 71 (2004) 107–120. doi:10.1016/j.chemolab.2003.12.009.
[3] T. Gröger, M. Schäffer, M. Pütz, B. Ahrens, K. Drew, M. Eschner, R. Zimmermann, Application of two-dimensional gas chromatography combined with pixel-based chemometric processing for the chemical profiling of illicit drug samples, J. Chromatogr. A. 1200 (2008) 8–16. doi:10.1016/j.chroma.2008.05.028.
[4] K.M. Pierce, J.C. Hoggard, Chromatographic data analysis. Part 3.3.4: handling hyphenated data in chromatography, Anal. Methods. 6 (2014) 645–653. doi:10.1039/C3AY40965A.
[5] D.C. Harris, Quantitative Chemical Analysis, 6th ed., W. H. Freeman and Company, New York, NY, 2003.
[6] S. Yang, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, Gas chromatography and comprehensive two-dimensional gas chromatography hyphenated with mass spectrometry for targeted and nontargeted metabolomics, in: Metabolomics Pract. Success. Strateg. Gener. Anal. Metab. Data, Wiley-VHC, Weinheim, Germany, 2013: pp. 69–92.
[7] K. Hiller, J. Hangebrauk, C. Jäger, J. Spura, K. Schreiber, D. Schomburg, MetaboliteDetector: Comprehensive Analysis Tool for Targeted and Nontargeted GC/MS Based Metabolome Analysis, Anal. Chem. 81 (2009) 3429–3439. doi:10.1021/ac802689c.
[8] W. Niu, E. Knight, Q. Xia, B.D. McGarvey, Comparative evaluation of eight software programs for alignment of gas chromatography–mass spectrometry chromatograms in metabolomics experiments, J. Chromatogr. A. 1374 (2014) 199–206. doi:10.1016/j.chroma.2014.11.005.
[9] K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, Recent advancements in comprehensive two-dimensional separations with chemometrics, J. Chromatogr. A. 1184 (2008) 341–352. doi:10.1016/j.chroma.2007.07.059.
[10] S.E. Stein, An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data, J. Am. Soc. Mass Spectrom. 10 (1999) 770–781. doi:10.1016/S1044-0305(99)00047-1.
[11] Eigenvector Research: Chemometrics Software, Consulting and Training, (n.d.). http://www.eigenvector.com/ (accessed December 9, 2016).
[12] Applied Chemometrics, (n.d.). http://www.chemometrics.com/ (accessed December 9, 2016).
[13] K.R. Murphy, P. Wenig, G. Parcsi, T. Skov, R.M. Stuetz, Characterizing odorous emissions using new software for identifying peaks in chemometric models of gas chromatography–mass spectrometry datasets, Chemom. Intell. Lab. Syst. 118 (2012) 41–50. doi:10.1016/j.chemolab.2012.07.006.
[14] C.A. Andersson, R. Bro, The N-way Toolbox for MATLAB, Chemom. Intell. Lab. Syst. 52 (2000) 1–4. doi:10.1016/S0169-7439(00)00071-X.
[15] S. O'Callaghan, D.P.D. Souza, A. Isaac, Q. Wang, L. Hodkinson, M. Olshansky, T. Erwin, B. Appelbe, D.L. Tull, U. Roessner, A. Bacic, M.J. McConville, V.A. Likić, PyMS: a Python toolkit for processing of gas chromatography-mass spectrometry (GC-MS) data. Application and comparative study of selected tools, BMC Bioinformatics. 13 (2012) 115. doi:10.1186/1471-2105-13-115.
[16] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149–171. doi:10.1016/S0169-7439(97)00032-4.

[17] R.A. Harshman, M.E. Lundy, PARAFAC: Parallel factor analysis, Comput. Stat. Data Anal. 18 (1994) 39–72. doi:10.1016/0167-9473(94)90132-5.
[18] J.C. Hoggard, R.E. Synovec, Parallel Factor Analysis (PARAFAC) of Target Analytes in GC × GC−TOFMS Data: Automated Selection of a Model with an Appropriate Number of Factors, Anal. Chem. 79 (2007) 1611–1619. doi:10.1021/ac061710b.
[19] J.C. Hoggard, W.C. Siegler, R.E. Synovec, Toward automated peak resolution in complete GC × GC–TOFMS chromatograms by PARAFAC, J. Chemom. 23 (2009) 421–431. doi:10.1002/cem.1239.
[20] J.M. Amigo, T. Skov, R. Bro, J. Coello, S. Maspoch, Solving GC-MS problems with PARAFAC2, TrAC Trends Anal. Chem. 27 (2008) 714–725. doi:10.1016/j.trac.2008.05.011.
[21] T. Skov, J.C. Hoggard, R. Bro, R.E. Synovec, Handling within run retention time shifts in two-dimensional chromatography data using shift correction and modeling, J. Chromatogr. A. 1216 (2009) 4020–4029. doi:10.1016/j.chroma.2009.02.049.
[22] D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity deviation ratio: A new metric for chemometric analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data, Anal. Chim. Acta. 871 (2015) 66–76. doi:10.1016/j.aca.2015.02.040.
[23] M. Navarro-Reig, J. Jaumot, T.A. van Beek, G. Vivó-Truyols, R. Tauler, Chemometric analysis of comprehensive LC×LC-MS data: Resolution of triacylglycerol structural isomers in corn oil, Talanta. 160 (2016) 624–635. doi:10.1016/j.talanta.2016.08.005.
[24] R.A. Fisher, Statistical Methods for Research Workers, 14th ed., Oliver and Boyd, 1970.
[25] H.D. Bean, J.E. Hill, J.-M.D. Dimandja, Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography–mass spectrometry data, J. Chromatogr. A. 1394 (2015) 111–117. doi:10.1016/j.chroma.2015.03.001.
[26] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher Ratio Method Applied to Third-Order Separation Data To Identify Significant Chemical Components of Metabolite Extracts, Anal. Chem. 78 (2006) 5068–5075. doi:10.1021/ac0602625.
[27] L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data, Talanta. 115 (2013) 887–895. doi:10.1016/j.talanta.2013.06.038.
[28] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach, Anal. Chem. 87 (2015) 3812–3819. doi:10.1021/ac504472s.
[29] B.A. Parsons, D.K. Pinkerton, B.W. Wright, R.E. Synovec, Chemical characterization of the acid alteration of diesel fuel: Non-targeted analysis by two-dimensional gas chromatography coupled with time-of-flight mass spectrometry with tile-based Fisher ratio and combinatorial threshold determination, J. Chromatogr. A. 1440 (2016) 179–190. doi:10.1016/j.chroma.2016.02.067.
[30] N.E. Watson, B.A. Parsons, R.E. Synovec, Performance evaluation of tile-based Fisher Ratio analysis using a benchmark yeast metabolome dataset, J. Chromatogr. A. 1459 (2016) 101–111. doi:10.1016/j.chroma.2016.06.067.

[31] J. Zhu, D. Djukovic, L. Deng, H. Gu, F. Himmati, E.G. Chiorean, D. Raftery, Colorectal Cancer Detection Using Targeted Serum Metabolic Profiling, J. Proteome Res. 13 (2014) 4120–4130. doi:10.1021/pr500494u.
[32] S. Yang, M. Sadilek, R.E. Synovec, M.E. Lidstrom, Liquid chromatography–tandem quadrupole mass spectrometry and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry measurement of targeted metabolites of Methylobacterium extorquens AM1 grown on two different carbon sources, J. Chromatogr. A. 1216 (2009) 3280–3289. doi:10.1016/j.chroma.2009.02.030.
[33] X. Mi, S. Li, Y. Li, K. Wang, D. Zhu, G. Chen, Quantitative determination of 26 steroids in eggs from various species using liquid chromatography–triple quadrupole-mass spectrometry, J. Chromatogr. A. 1356 (2014) 54–63. doi:10.1016/j.chroma.2014.05.084.
[34] B. Sgorbini, C. Bicchi, C. Cagliero, C. Cordero, E. Liberto, P. Rubiolo, Herbs and spices: Characterization and quantitation of biologically-active markers for routine quality control by multiple headspace solid-phase microextraction combined with separative or non-separative analysis, J. Chromatogr. A. 1376 (2015) 9–17. doi:10.1016/j.chroma.2014.12.007.
[35] S.E. Reichenbach, X. Tian, C. Cordero, Q. Tao, Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography, J. Chromatogr. A. 1226 (2012) 140–148. doi:10.1016/j.chroma.2011.07.046.
[36] H.-G. Schmarr, J. Bernhardt, Profiling analysis of volatile compounds from fruits using comprehensive two-dimensional gas chromatography and image processing techniques, J. Chromatogr. A. 1217 (2010) 565–574. doi:10.1016/j.chroma.2009.11.063.
[37] C. Cordero, E. Liberto, C. Bicchi, P. Rubiolo, P. Schieberle, S.E. Reichenbach, Q. Tao, Profiling food volatiles by comprehensive two-dimensional ga schromatography coupled with mass spectrometry: Advanced fingerprinting approaches for comparative analysis of the volatile fraction of roasted hazelnuts (Corylus avellana L.) from different origins, J. Chromatogr. A. 1217 (2010) 5848–5858. doi:10.1016/j.chroma.2010.07.006.
[38] M. Brokl, L. Bishop, C.G. Wright, C. Liu, K. McAdam, J.-F. Focant, Multivariate analysis of mainstream tobacco smoke particulate phase by headspace solid-phase micro extraction coupled with comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry, J. Chromatogr. A. 1370 (2014) 216–229. doi:10.1016/j.chroma.2014.10.057.
[39] F. Magagna, L. Valverde-Som, C. Ruíz-Samblás, L. Cuadros-Rodríguez, S.E. Reichenbach, C. Bicchi, C. Cordero, Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil, Anal. Chim. Acta. 936 (2016) 245–258. doi:10.1016/j.aca.2016.07.005.
[40] E. Gorrochategui, J. Jaumot, S. Lacorte, R. Tauler, Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow, TrAC Trends Anal. Chem. 82 (2016) 425–442. doi:10.1016/j.trac.2016.07.004.
[41] B.C. Reaser, S. Yang, B.D. Fitz, B.A. Parsons, M.E. Lidstrom, R.E. Synovec, Non-targeted determination of 13C-labeling in the Methylobacterium extorquens AM1 metabolome using the two-dimensional mass cluster method and principal component analysis, J. Chromatogr. A. 1432 (2016) 111–121. doi:10.1016/j.chroma.2015.12.088.
[42] M.M. Sinanian, D.W. Cook, S.C. Rutan, D.S. Wijesinghe, Multivariate Curve Resolution-Alternating Least Squares Analysis of High-Resolution Liquid Chromatography–Mass Spectrometry Data, Anal. Chem. 88 (2016) 11092–11099. doi:10.1021/acs.analchem.6b03116.

[43] N.E. Watson, W.C. Siegler, J.C. Hoggard, R.E. Synovec, Comprehensive Three-Dimensional Gas Chromatography with Parallel Factor Analysis, Anal. Chem. 79 (2007) 8270–8280. doi:10.1021/ac070829x.
[44] W.C. Siegler, J.A. Crank, D.W. Armstrong, R.E. Synovec, Increasing selectivity in comprehensive three-dimensional gas chromatography via an ionic liquid stationary phase column in one dimension, J. Chromatogr. A. 1217 (2010) 3144–3149. doi:10.1016/j.chroma.2010.02.082.
[45] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells, Anal. Chem. 78 (2006) 2700–2709. doi:10.1021/ac052106o.
[46] R.E. Mohler, K.M. Dombek, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, Comprehensive analysis of yeast metabolite GC×GC–TOFMS data: combining discovery-mode and deconvolution chemometric software, Analyst. 132 (2007) 756–767. doi:10.1039/B700061H.
[47] B.P. Tu, R.E. Mohler, J.C. Liu, K.M. Dombek, E.T. Young, R.E. Synovec, S.L. McKnight, Cyclic changes in metabolic state during the life of a yeast cell, Proc. Natl. Acad. Sci. 104 (2007) 16886–16891.
[48] J.C. Hoggard, R.E. Synovec, Parallel Factor Analysis (PARAFAC) of target analytes in GC × GC−TOFMS data: automated selection of a model with an appropriate number of factors, Anal. Chem. 79 (2007) 1611–1619. doi:10.1021/ac061710b.
[49] J.C. Hoggard, R.E. Synovec, Automated Resolution of Nontarget Analyte Signals in GC × GC-TOFMS Data Using Parallel Factor Analysis, Anal. Chem. 80 (2008) 6677–6688. doi:10.1021/ac800624e.
[50] J.C. Hoggard, W.C. Siegler, R.E. Synovec, Toward automated peak resolution in complete GC × GC–TOFMS chromatograms by PARAFAC, J. Chemom. 23 (2009) 421–431. doi:10.1002/cem.1239.

# Chapter 2. Performance Evaluation of Tile-Based Fisher Ratio Analysis using a Benchmark Yeast Metabolome Dataset[2]

## 2.1 INTRODUCTION

Knowledge of a biochemical system is incomplete without the inclusion of the metabolome. The study of metabolomics attempts to identify and quantify the low molecular weight compounds which make up the metabolome [1]. Though not specifically mentioned in the central dogma of molecular biology [2], identified metabolites generally complement and reinforce determinations made of the genome and proteome. In other words, an up/down regulated gene should correspond to a correlated change in the proteome, which generally also induces some measureable change in the metabolome [2]. Conversely, unexpected discoveries in the metabolome could suggest changes in the upper realms of the molecular biology hierarchy.

Analytically, the metabolome can be approached in one of two ways, targeted analysis or non-targeted analysis [1]. Targeted analysis, attempts to verify previous genomic and proteomic results through identification of complementary compounds in the metabolome. Targeted approaches focus on expected changes in the metabolome induced by up and down regulated genes and proteins. The foreknowledge of interesting compounds enables the analyst to define detailed methods and techniques to ensure the targeted metabolites are identified and quantified to the necessary level of fidelity. The alternative is to collect a biological sample of interest and submit it for some form of broad analysis. These non-targeted or discovery-based approaches

[2] This Chapter has been reproduced from N.E. Watson, B.A. Parsons, R.E. Synovec, Performance evaluation of tile-based Fisher Ratio analysis using a benchmark yeast metabolome dataset, J. Chromatogr. A. 1459 (2016) 101–111. doi: 10.1016/j.chroma.2016.06.067.

employ non-specific analytical procedures to identify metabolomic changes induced by experimental or environmental perturbations.

Unlike targeted approaches which are amenable to specific assays or similar techniques, non-targeted metabolic investigation requires instrumentation that probes the sample in a wide-ranging and general way. Instrumental platforms that provide the ability to simultaneously separate and identify the resultant sample components are desirable. Examples include liquid chromatography (LC) [3], gas chromatography (GC) [4], mass spectrometry (MS) [5], nuclear magnetic resonance spectroscopy (NMR) [6,7], or various combinations of these techniques. Comprehensive two-dimensional gas chromatography coupled with time-of-flight-mass spectrometry (GC × GC–TOFMS) [1] is arguably one of the best analytical instrumentation platforms to study metabolites of interest that reside in the 50 to 500 Da mass range. The use of GC-MS for metabolomics studies is widespread, but the benefits of GC × GC–TOFMS applied to metabolomics have begun to emerge in more recent years. While GC × GC–TOFMS is an outstanding instrumental platform for biological and metabolomics studies [8–26], there is an ongoing need to develop software methods to glean useful information from the immensely complex data. For this purpose Fisher Ratio (F-ratio) analysis has been found to be a particularly useful algorithm for the analysis of these complex datasets [27–31].

F-ratio analysis compares the variance between classes relative to the variance within the classes [32–34]. Specifically, the F-ratio compares whether these two variances are different relative to a tabulated F-statistic or other means of threshold determination. When the F-ratio is calculated, a quotient is determined, ranging from zero to infinity; as the value increases the results suggests with greater certainty that the compared variances are different implying that the source data is also different. Points with high F-ratios generally correspond to features which

distinguish between the classes compared. Previously, F-ratio analysis using GC × GC data was applied to several biological and petrochemical models [26–31]. The prior pixel-based F-ratio studies were fruitful, and provide a benchmark for current software development and evolution.

Specifically, Mohler, et. al. [25,26] applied the F-ratio method to intracellular extracts from *Saccharomyces cerevisiae* (yeast) either metabolizing glucose (repressed) or ethanol (derepressed), which define the two classes in the discovery-based analysis to find metabolites that are statistically different in concentration between the two classes. These data were studied in a "pixel-wise" fashion whereby each data point (defined by chromatographic time on column 1 ($^1$D), chromatographic time on column 2 ($^2$D), and the mass-to-charge ratio (*m/z*)) across all the sample replicates was compared as a subset to calculate an F-ratio, which statistically quantifies the variances for the complete dataset at that one data point pixel. The F-ratio calculation was repeated iteratively for every data point collected. Beneficially, the Mohler et. al. dataset was large and well defined biologically. Each culture was thrice replicated providing three samples for each culture class. Also, the samples were extracted in triplicate and each extraction was chromatographically analyzed four times. The copious replication ensured the different sources of variance were sufficiently quantified. The result of this study was a list of metabolites determined to be variant, or up/down regulated and hence changing, between the two sample classes ordered by F-ratio value. The list, referred to as a "hit list", ordered the metabolites from greatest to least F-ratio. This hit list was fully analyzed and quantified through application of parallel factor analysis (PARAFAC) [35]. Usefully, this analysis provided empirical evidence for biochemically suggested metabolic pathways. Regrettably, the identification and quantification of variant metabolites was onerous, since the true positives were found to be intermingled with a significant number of false positive, and data analysis consumed greater than a person-year of

labor to execute. The analyst was required to manually interdict the process at many steps and analyze the hit list down to the lowest level achievable without a useful means of deciding when the hit list was complete. In retrospect, the analysis was specifically confounded by two issues we hope to elucidate and correct in this current study: retention time shifting on $^{1}$D and $^{2}$D, and undesirable false-positive discoveries.

Retention time shifting adds to the difficulty in the analysis of any chromatographic dataset. In the Mohler, et. al. study the retention time shift reduced the sensitivity in which true positives were discovered due to the pixel-based comparison of data. Since the yeast experiment was conducted over the course of months and required many extra chromatographic runs in addition to the extracted samples (e.g. solvent blanks, growth medium blanks, etc.), retention time shifting became especially pronounced. Temperature and pressure fluctuations combined with matrix effects and stationary phase degradation may all lead to retention time variations. Many times alignment [36,37] or warping [38] is chosen as a solution to correct for retention time changes. Alignment, while not particularly time consuming, can impart unfortunate artifacts to the data as the peaks are warped and time-shifted. The tile-based F-ratio software is designed to address this challenge without explicitly requiring alignment [29,30].

The other challenge, how to minimize the false positive discoveries, is tied to determining a useful F-ratio threshold value under which further analysis is deemed fruitless. Statistics has come up with two schools of thought on the topic of threshold determination [39,40]. One has the analyst consult a tabular solution for the applied statistic, in this case the F-test. Based on the degrees of freedom and number of samples, the analyst determines an F-critical value and judges that all features with calculated F-ratios greater than F-critical do significantly differ between/among classes. The other school of thought, less well known outside

statistics circles, suggests that the tabulated values for the textbook distribution may not accurately apply to the experiment at hand. This group is known as the “frequentists.” The frequentist camp approaches the problem of determining the threshold with the expectation that every experiment includes many underlying random and statistical errors which may not be appropriately represented by a Bayesian Distribution.

To address this challenge of determining an appropriate threshold to prudently guide hit list data mining, combinatorial null distribution analysis is coupled with tile-based F-ratio analysis [30,31], leveraging the large volume of related data that may be tested by GC × GC–TOFMS to experimentally determine the distribution of potential false positives. By rearranging the sample classes to nullify the class distinguishing variance, the effects of non-meaningful variation in the dataset may be estimated for the true class comparison. To increase the extensibility of the approach, it is desirable to determine all possible null comparisons in a rigorous and automated fashion. With the recent improvements to the computational performance of the software, it is now possible to glean the benefits of using the complete set of null distributions. By algorithmically determining all possible null combinations it is possible to utilize the resulting Fisher Distributions to determine a rigorous threshold.

Herein tile-based F-ratio analysis is evaluated and validated by application to the already thoroughly analyzed fermenting and respiring yeast benchmark dataset from the Mohler et. al. study [26]. These data were completely identified and quantified previously and provide an insightful opportunity to compare the tile-based F-ratio analysis [29–31], which in turn had evolved from pixel-based F-ratio analysis [26–28]. The novelty of the study herein is based upon the simultaneous demonstration of improved F-ratio analysis software performance for a complex metabolomics dataset, which is manifested as a decisive improvement in the ranking of

true positive hits above false positive hits, in concert with the demonstration that the null distribution threshold accurately identifies this optimized transition from true positive hits to false positive hits in the hit list. In addition, quantification by ChromaTOF (LECO software) is utilized in lieu of PARAFAC to demonstrate the possibilities of rapid peak finding and quantification without the tedium of detailed chemometric deconvolution.

Tile-based F-ratio analysis was applied to four randomly selected replicate data subsets composed of six repressed (R) versus six derepressed (DR) chromatograms (originally from a total collection of thirty-five R versus thirty-five DR chromatograms [25,26]). These six versus six subsets were chosen to doubly represent each of the three cultural replicates from each sample class. The biological variance was previously demonstrated to compose the majority of the total variance; therefore, its representation was paramount relative to injection and extraction variance. The six R versus six DR comparison, vice 35 R versus 35 DR in the prior studies [25,26], is necessitated by the combinatorial rearrangement of the chromatogram class assignments. There are 200 possible different null arrangements of a six versus six comparison and this number rapidly increases as the sample class population increases. Each replicate data subset was subjected to F-ratio analysis of both the one true class comparison for metabolite discovery and the additional 200 null classes for threshold determination [30,31]. Four replicate data subsets were analyzed in order to assess whether or not results from any given individual subset adequately represent the whole dataset of thirty-five R versus thirty-five DR chromatograms [25,26]. Generally, the first data subset is utilized for the discussions herein.

## 2.2 EXPERIMENTAL

### 2.2.1 *Sample Prep and Data Collection*

The biological culture , extraction, derivatization and instrument conditions were previously described, with a brief recount included here [25,26]. Three R cultures (AR, BR, CR)

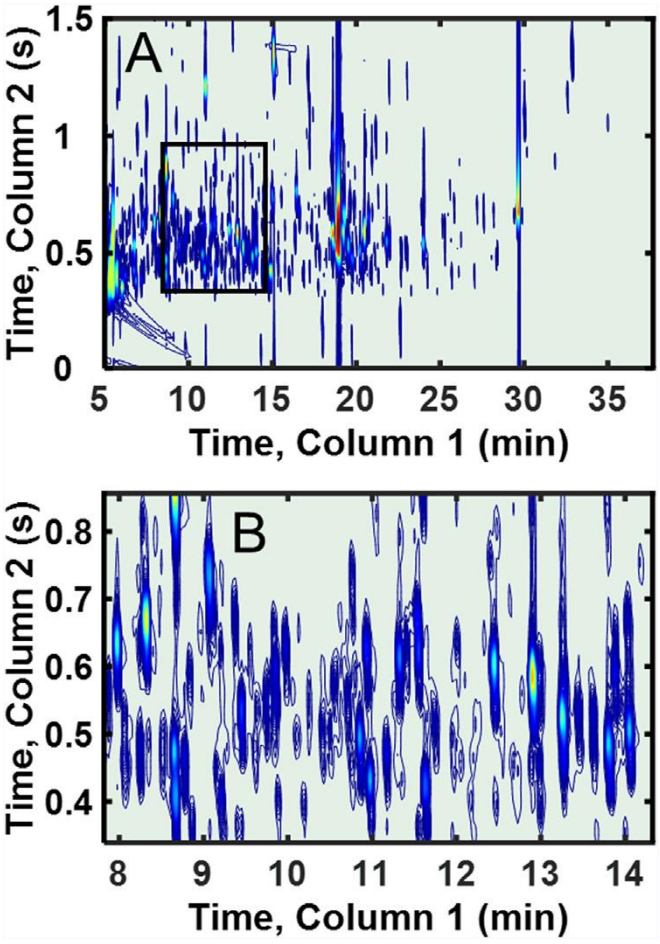


Figure 2.1. Representative (A) GC × GC–TOFMS chromatogram of m/z 73 and zoom (B) to region in box.

and three DR cultures (ADR, BDR, CDR) were grown to obtain a measure of biological variability. At the end of the growth cycle the cells were quenched and metabolites were extracted with ethanol. Each extraction was performed in triplicate, e.g. A1R, A2R, and A3R.

An Agilent 6890N gas chromatograph equipped with an Agilent 7683 auto-injector (Agilent Technologies, Santa Clara, CA, USA) coupled to a LECO Pegasus III TOFMS with a 4D thermal modulator upgrade (LECO, St. Joseph, MI, USA) was used to collect the GC × GC–TOFMS data. The GC inlet and transfer line were held constant at 280 °C. One micro liter split-less injections were made onto the $^{1}$D column (RTX-5MS, 20 m x 250 µm i.d. x 0.5 µm, Restek, Bellefonte, PA, USA). The $^{1}$D column effluent was modulated and transferred to the $^{2}$D column (RTX-200MS, 2 m x 180 µm i.d. x 0.2 µm, Restek, Bellefonte, PA, USA) every 1.5 s. The $^{1}$D column was held at 60 °C for 0.25 min and then increased at 8 °C/min to 280 °C where it was held for 10 min. The $^{2}$D column was initially set at 70 °C and followed the same temperature program as the $^{1}$D column giving a total run time of 37.75 min. The modulator was kept 40 °C higher than the $^{1}$D column. A constant volumetric flow rate of 1 ml/min was held at the head of the $^{1}$D column. The ion source was set to 200°C. Data were collected at a rate of 100 spectra/s after a 5-min solvent peak delay. Each extract had 4 replicate injections except A1R and A1DR, which were injected in triplicate. This resulted in 70 injections (70 GC × GC–TOFMS chromatograms) for the analysis. Samples are labelled from left to right by culture (A, B or C), Extract (1, 2, or 3), biological class (R or DR) and replicate (1, 2, 3, or 4). For example, A1DR:1 indicates Culture A, Extract 1, DR class, and injection replicate 1.

Figure 2.1 depicts a representative chromatogram from the DR class. The complexity of these metabolic extracts is easily discerned from the visually apparent peaks. Chromatographic performance was estimated using the cystathionine signal (Fig. 2): $^{1}W_{b}$ = 4.5 s, $^{2}W_{b}$ = 100 ms, $M_{R}$ = 3, $^{1}n_{c}$ = 13.3 min$^{-1}$, $^{2}n_{c}$ = 15, $n_{c,2D}$ = 200 min$^{-1}$ resulting in a total peak capacity per GC × GC chromatogram of approximately 6400.

### 2.2.2 *Data Analysis*

The GC × GC–TOFMS chromatograms were extracted from ChromaTOF 3.32 (LECO, St. Joseph, MI, USA) to MATLAB R2015b (The Math Works, Natick, MA, USA). Six chromatograms were selected from each biological class, R and DR. Samples were chosen through a managed random selection system by which two replicates were included from each cell culture; the extraction and injection replicate choices were randomized by die rolls. The samples utilized for each replicate data subset are listed in Table 2.1.

Table 2.1. Sample composition for each replicate data subset.

| **Replicate Data Subset 1** | **Replicate Data Subset 2** | **Replicate Data Subset 3** | **Replicate Data Subset 4** |
|---|---|---|---|
| A1DR:2 | A1DR:3 | A2DR:4 | A1DR:3 |
| A2DR:1 | A3DR:2 | A3DR:2 | A2DR:2 |
| B1DR:1 | B1DR:2 | B1DR:3 | B1DR:2 |
| B2DR:1 | B2DR:4 | B3DR:3 | B2DR:1 |
| C1DR:3 | C2DR:1 | C1DR:2 | C2DR:2 |
| C3DR:3 | C3DR:2 | C2DR:1 | C3DR:1 |
| A1R:3 | A1R:1 | A2R:2 | A2R:2 |
| A2R:1 | A2R:2 | A3R:2 | A3R:3 |
| B1R:2 | B2R:4 | B1R:3 | B2R:3 |
| B2R:4 | B3R:3 | B2R:4 | B3R:2 |
| C2R:2 | C1R:2 | C1R:1 | C1R:4 |
| C1R:2 | C2R:3 | C2R:1 | C2R:2 |

The chosen chromatograms were submitted to tile-based F-ratio analysis for metabolite discovery [30]. This metabolite discovery, where the R and DR classes are compared, is referred to as the "true class comparison." Before execution, values were chosen for the required initialization parameters: $^{1}$D dimension tile size of 6 s, $^{2}$D dimension tile size of 300 ms, signal-to-noise (*S/N*) threshold of 3, a minimum F-ratio of zero, at least 5 mass channels required with a measurable F-ratio. A two-dimensional (2D) cluster window of 2 modulations by 200 ms was utilized to minimize redundant hits from the multiple tile schemes used during F-ratio analysis

[30]. The chromatograms were normalized to the sum of the total ion current (TIC). No alignment was performed on any chromatograms in this study.

The reader is directed to the Supporting Information of a previous publication for a detailed description of the tile-based F-ratio software [30]. A brief summary is presented here with respect to how the software reduced the dataset to a manageable list of hit locations for the first replicate subset. Each GC × GC–TOFMS chromatogram is composed of 1.1 x $10^8$ pixel features collected by the TOFMS detector. These pixel features were automatically reduced to 9.1 x $10^5$ tile features by the tile scheme. The *S/N* filter automatically reduced this to 6.1 x $10^4$ features, followed by F-ratio analysis that resulted in 5279 total hits across each of the four tile schemes analyzed. Redundant hits were then automatically removed by a "pin and cluster" algorithm, resulting in the final hit table listing 1696 total hits. Implementation of null distribution analysis, which is explained next, further automatically reduced the number of statistically relevant features that warranted additional analysis.

The true class comparison via the tile-based F-ratio analysis was immediately followed by null class comparison of all two hundred possible null combinations. The number of unique null classes are described by

$$\#\ of\ unique\ null\ arrangements = \frac{1}{2} \times \frac{n_1!}{k_1!(n_1-k_1)!} \times \frac{n_2!}{k_2!(n_2-k_2)!} \qquad (2.1)$$

In Eq. 2.1, *n* represents the number of samples in a given class and *k* represents the number of samples to be utilized in the null class ($k = n / 2$ for a two class system) [30,31]. Null classes are analyzed in the same fashion as the true class. The null class comparisons are binned and plotted as histograms as will be described below. These null distributions were used to determine a F-ratio threshold for the true class comparison and all hits above the threshold were analyzed in greater detail. This left 350 hits from the initial 1696 for the analyst to manually scrutinize.

Visual inspection allowed for removal of a modest number of redundant metabolite hits missed by the algorithm, generally overloaded and tailing peaks along the $^{2}$D dimension. Also removed were hits attributed to reagent artifacts by their presence in corresponding growth medium control chromatograms, which produced many artifact hits that streaked primarily along the $^{1}$D dimension. After removal of reagent artifacts and redundant hits, the 350 hits were reduced to ~ 100 class-distinguishing metabolite features, which were identified by retention time with the results from the benchmark study [26]. The hits were further verified by comparison to a peak table generated by the ChromaTOF software.

## 2.3 Results and Discussion

Discussion of the results of the tile-based F-ratio analysis will begin with a brief tutorial of how the method is executed, as applied specifically to the experimentally obtained benchmark dataset [25,26] (the reader is directed to these previous publications for an overview of the data structure along with the biological implications). Herein, the benchmark dataset is comprehensively interrogated using state-of-the-art tile-based F-ratio analysis [30], and the results are initially detailed by way of a case study on the cystathionine signal event (Fig. 2). Later, the comprehensive results provided by the tile-based approach are presented in various forms, the threshold is determined by null distribution analysis [31], and all of these findings are compared to the previously published results [25,26].

Figure 2.2(A) and (B) present the cystathionine chromatographic signal event for sample A2DR:1 in both unfolded one-dimensional (1D) and 2D format, respectively. The selective ion, m/z 278, is utilized for all the chromatograms in Figure 2.2. The box in Figure 2.2(B) represents the selected tile size used for tile based analysis (4 $^{1}$D modulations by 300 ms on $^{2}$D). The tile

size encompasses the vast majority of the signal generated by each eluted component and is sufficiently wide to compensate for retention shifting on $^{1}$D and $^{2}$D.

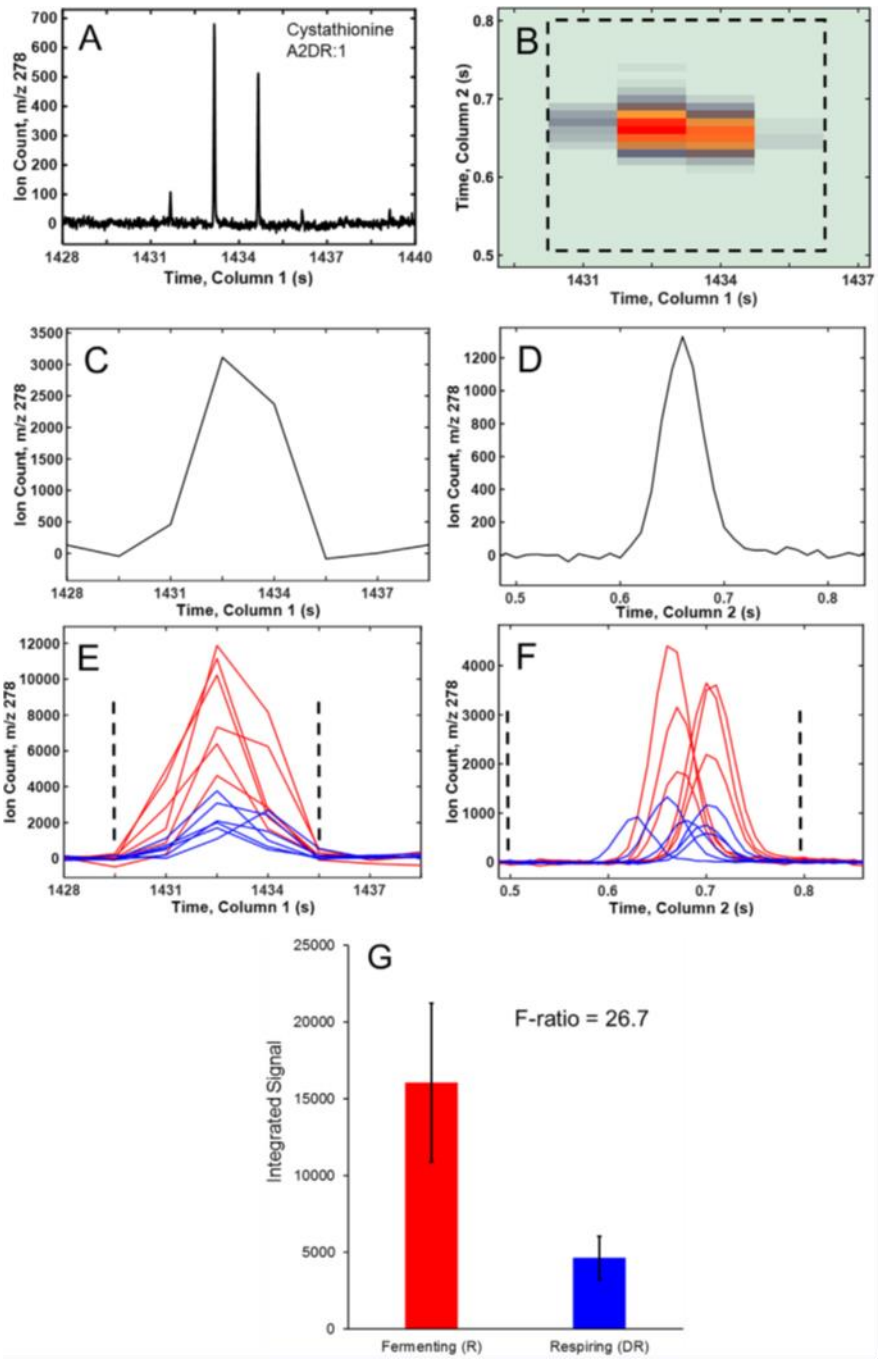


Figure 2.2. Extract of the cystathionine chromatographic elution event at *m/z* 278 represented from various aspect frames. These figures graphically depict the logic behind selecting the 4 modulation by 300 ms tile size.

Figure 2.2(C) and (D) show the result of summing the cystathionine signal event along $^{2}$D and $^{1}$D respectively to construct a 1D peak. Summing a dimension effectively reconstructs, in a very rudimentary way, the peak shape in the other dimension facilitating visualization of the tile-size from multiple aspect frames. Figure 2.2(E) and (F) show the 1D peaks along the $^{1}$D and the $^{2}$D dimensions for all twelve of the samples utilized in the first replicate data subset F-ratio analysis. The vertical dashed lines represent the $^{1}$D and $^{2}$D tile window sizes. The $^{2}$D dimension of the tile size easily encompasses the $^{2}$D retention shifting in this case. One of the significant benefits of the algorithm is the ability to compensate for retention time shifting without the need for chromatographic alignment by warping [38] or peak shifting [36]. It is also observed, as is generally the case, that retention time shifting affects the dataset along the $^{2}$D dimension to a much greater extent than along the $^{1}$D dimension.

The 1.5 s modulation period limited the tile size options along the $^{2}$D dimension to 30, 60, 100, 150, 300 or 500 ms (evenly divisible multiples of 150 spectra per modulation). The best choice was found to be 300 ms, as 500 ms is even larger (introducing too much superfluous signal from nearby eluted compounds) and 150 ms is too small (does not encompass the peak widths coupled with the observed retention shifting). In future applications of tile-based F-ratio analysis, an integer modulation period, eg. 1 or 2 s, is more desirable as there are many evenly divisible multiples that result in an even division of the full modulation period into equally sized tiles along the $^{2}$D dimension. Additionally, numerically resampling the collected data to allow for different divisions of the $^{2}$D separation space can also be performed to address this concern. The tile size is a crucial parameter considered prior to executing the analysis. If properly executed, this data analysis approach presents the opportunity to overcome significant retention time shift in both GC × GC dimensions. The metabolite data utilized here to study and validate

the tile-based F-ratio analysis present some of the most challenging retention time shifts generally expected. Retention time shifting, in particular the shifts on $^{2}$D for this dataset, leads to the use of a $^{2}$D tile dimension that is substantially greater than the peak widths, with the range of shifting ~ 100 ms observed in Figure 2.2(F). One might initially surmise from this that the "practical" peak capacity of the 2D separation has been compromised by the retention time shifting. However, the "pin and cluster" step of the tile-based F-ratio software addresses this issue by focusing the location of each hit back to the original high-fidelity resolution of the pixel-based dataset [30]. Indeed, each hit location is subsequently analyzed by PARAFAC (or other deconvolution software) using a focused tile centered on the hit location, and it is not uncommon to identify and quantify more than one relevant analyte within the focused tile. Accordingly, the tile-based F-ratio software is designed to preserve to a great extent the 2D peak capacity provided by performing high efficiency GC × GC separations.

A simple observation of the difference in the peak heights between these two classes (approximately 3.5 fold higher on average for the R class) indicates cystathionine exhibits a high degree of between class variance. The task of the tile-based F-ratio algorithm is to discover this variance with a high degree of confidence and minimal analyst intervention despite significant background noise and retention time shifting. The algorithm achieves this by following the transformation of Figure 2.2(A) through (F) to the natural conclusion. All the signal within the tile depicted in Figure 2.2(B) to (F) is integrated to single value. These resultant values are then submitted for F-ratio calculation. Class R yields integrated signal of 11761, 22855, 19514, 15831, 8578 and 18530, while Class DR yields integrated signal of 3673, 6548, 2915, 4458, 6189 and 4219, with the average results presented visually in Figure 2.2(G). Determination of the class-to-class variance and within class variance followed by F-ratio calculation yields an F-

ratio value of 26.7 [27]. The algorithm does this across all *m/z* for tiles with a *S/N* exceeding an operator set minimum threshold.

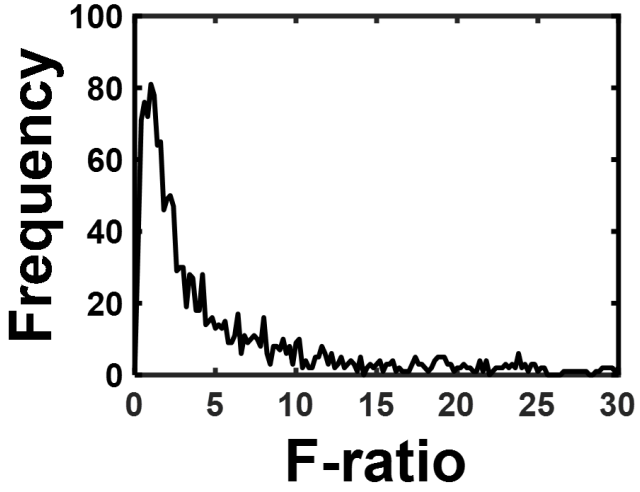


Figure 2.3. F-ratio distribution of the true class comparison for the first replicate data subset.

The result of the true class comparison is a table of hits arranged by descending order of F-ratio, with the top hit representing the chromatographic feature that is most distinctive between the DR and R classes. These results are summarized graphically in Figure 2.3 for the first subset of six randomly selected DR and R samples. Here the tabular results are depicted as histograms binned by F-ratio with a bin window of 0.2. Each curve for the four replicate data subset true class comparisons follows the general form of the F-Distribution [33]. The maximum frequency F-ratio occurs near zero representing non-significant chromatographic features likely resulting from real metabolites that do not change in concentration significantly between the DR and R classes. As the F-ratio increases, the statistical significance of the corresponding feature increases as the value of the class-to-class variance increases relative to the within class variance. Consequently, the frequency of occurrence approaches zero as the F-ratio approaches infinity.

The F-ratio distribution in Figure 2.3 lacks useful features to identify a threshold, below which the hits can *likely* be deemed not statistically different in concentration between the two classes and hence excluded from further analysis. As such, the biological significance of the hits

found above the threshold does not need to be known *a priori*. Bayesian statistics described in a general statistics textbooks can perhaps be used to determine the critical values of an F-Distribution [33]. This is a mathematically sound idea, but it unfortunately only allows for the exclusion of a small portion of the determined hits and leads to an excess of F-ratios deemed as true positive hits in which there is likely no statistical difference in concentration for a given hit between the two classes. The F-critical value for this analysis at the 95% confidence limit is 4.96 leaving approximately 700 hits deemed true positive. Basically, application of Bayesian statistics leads to usage of much too low of a threshold. To address this challenge and shortcoming of applying Bayesian statistics for the threshold determination, combinatorial null distribution analysis was applied to determine an appropriate F-ratio threshold. Implementation of combinatorial null distribution analysis is based on the data collected and therefore allows for the statistical analysis to account for any systematic errors in the data. The threshold determined by combinatorial null distribution analysis follows the empirical or "frequentist" approach.

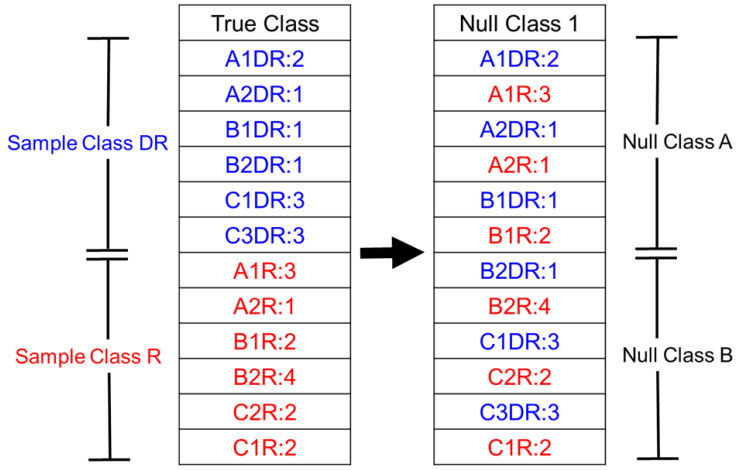


Figure 2.4. Samples composing the true class comparison for the first replicate data subset along with one of the 200 null classes created for determination of an F-ratio threshold.

Combinatorial null distribution analysis was applied and all 200 null distributions for a given subset of six DR and six R samples per Eq. 2.1 were calculated in the same fashion as the

true class comparison in Figure 2.3. An illustration of how the six DR and R samples are scrambled for one of the 200 arrangements is provided in Figure 2.4. Null analysis empirically accounts for the features present in the data that are not attributable to the true class distinguishing features. Figure 2.5(A) shows all 200 of these distributions overlaid on the same axes for the first subset of six DR and six R samples. The maxima of these null distributions are shifted to lower values and the frequency of all the low magnitude F-ratios are increased while the high F-ratios are decreased relative to the true class comparison. These null distributions provide a graphical depiction of the frequency and magnitude of non-class distinguishing features present in the data.

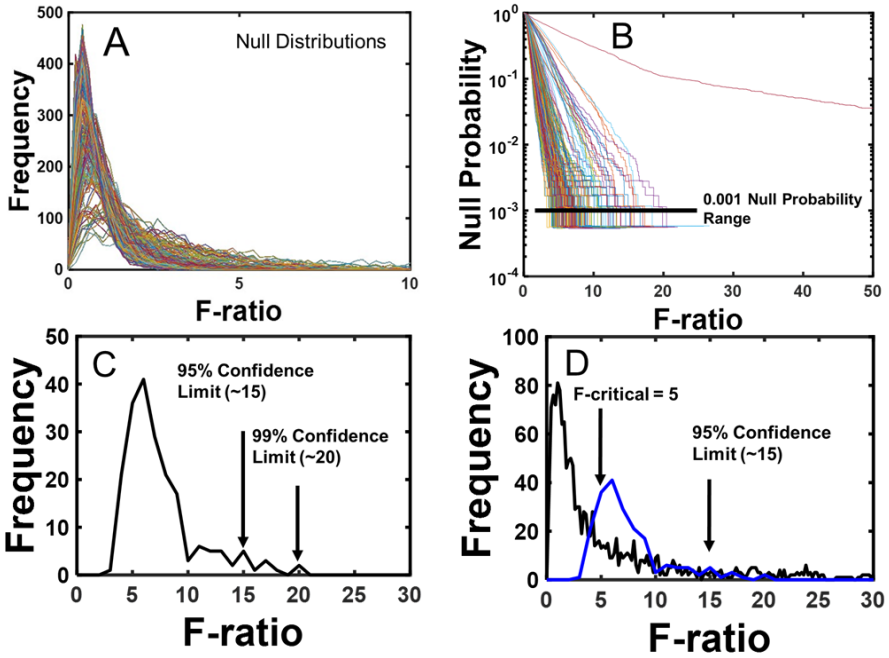


Figure 2.5. Probability distributions for the first replicate subset.

A useful transformation of the null probability distributions shown in Figure 2.5(A) to cumulative probability distributions is depicted in Figure 2.5(B). Here the F-ratios are plotted against their probability of occurrence (plotted with a base ten log scale y-axis for clarity). The null probabilities are calculated by determining the number of features above an F-ratio value divided by the total number of features and multiplying the result by 100%. Serendipitously, this

dataset revealed the power of the combinatorial null distribution analysis. The obvious outlier with significant shifting to higher F-ratio values was attributed to a systematic error previously undiscovered in the dataset. The null class resulting in this outlier is the combination that groups the samples into two classes by date of sample collection. Since these data were collected over the course of two months this likely resulted from a drift in TOFMS detector efficiency. Fortunately, this systematic error was controlled by randomization of the sample analysis sequence. However, its presence still imparts artifacts to the data that would otherwise be ignored by traditional Bayesian Statistics.

Figure 2.5(B) also holds the key to determining a potentially useful threshold for application to the true class comparison hit list. Plotting all the F-ratio values for each of the 200 null distributions corresponding to a 0.001 null probability yields yet another F-Distribution plotted in Figure 2.5(C). Figure 2.5(C) depicts what can be considered a confidence interval for the 0.001 null probability F-ratio value across all 200 possible null combinations. Again, this distribution follows the form of the Fisher Distribution, but the maximum is shifted to the right relative to both the true class comparison (Figure 2.3) and all 200 null class comparisons (Figure 2.5(A)). Integrating the area under this new curve yields a limit which is utilized as a threshold value for the true class comparison. Limits at both 99% and 95% of the integrated area are illustrated. The 99% confidence limit illustrated, at an F-ratio of 20, corresponds to the F-ratio in which 1% (or 2) of the 200 null combinations equal or exceed the selected 0.001 null probability (from Figure 2.5(B)). Similarly, the 95% confidence limit illustrated, at an F-ratio of 15, corresponds to the F-ratio in which 5% (or 10) of the 200 null combinations equal or exceed the selected 0.001 null probability. At a more rudimentary level, the 0.001 null probability is utilized as a surrogate for the false discovery rate (FDR). Each null distribution provides an

approximation for non-class distinguishing variance present in the data that results from both random and systematic errors naturally included within any real instrumental data. The conversion to the F-distribution in Figure 2.5(C) allows the analyst to visualize the range of potential values corresponding to a 1 in 1000 FDR. For this study, the 95% confidence limit with a threshold of 15 was applied to the true class comparison. The choice of a threshold of 15 provides a 95% certainty that we will achieve a 1 in 1000 FDR when the true class comparison is analyzed in detail. The threshold must be determined anew with each replicate data subset. Null distribution analysis already elucidated how the date of sample collection imparted a systematic error to the chromatograms and the effects of this error and others must be redetermined for each replicate data subset.

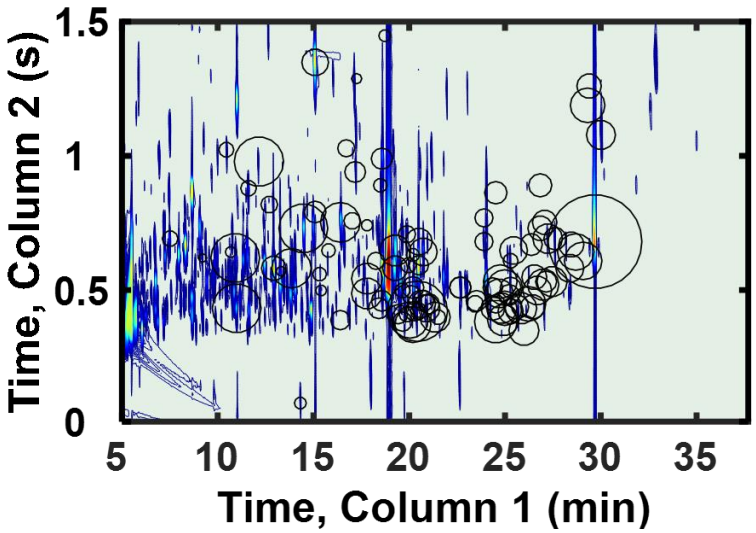


Figure 2.6. Average DR chromatogram from Fig. 2.1(A) with true hits indicated by circles after removal of redundant hits and matrix artifacts

Selection of the threshold completes the preparatory work and leaves the analyst ready for detailed analysis of the hit list generated by the true class comparison. Figure 2.6 shows the average chromatogram for the DR class from the first replicate data subset overlaid with the locations of the 94 hits determined above an F-ratio of 15 (the 95% confidence limit at a 0.001 null probability). Any redundant hits not removed by the otherwise automated software, along

with any reagent artifacts that were identified using culture medium blanks, have been manually removed. The circle diameter is scaled proportionally with F-ratio magnitude. In short, these 94 hits should represent the features that most distinguish sample classes in the 6 DR versus 6 R chromatograms. Though not depicted, a representative R chromatogram looks similar to Figure 2.6 with one visually identifiable feature missing at a retention time of approximately 1800 s. This feature corresponds to the largest F-ratio, most class distinguishing feature, the sugar trehalose.

Table 2.2. Results of tile-based F-ratio analysis for metabolites present in both classes.

| Metabolite | $t_{R1}$, s | $t_{R2}$, s | based F-ratio | [DR]/[R] PARAFAC | $t_{calc}/t_{table}$ | 1 (15) | 2 (22) | 3 (20) | 4 (20) | [DR]/[R] ChromaTOF |
|---|---|---|---|---|---|---|---|---|---|---|
| trehalose | 1782.0 | 0.7 | 1.4350 | 66.8 | 37.97 | 751 | 766 | 701 | 894 | 83.5 |
| glucose | 1225.5 | 0.4 | 0.8998 | 0.02 | 1.17 | 312 | 308 | 664 | 485 | 0.01 |
| glycerol c00116 | 660.0 | 0.4 | 0.3036 | 0.1 | 3.23 | 203 | 172 | 180 | 206 | 0.2 |
| lysine | 1231.5 | 0.6 | 0.1495 | 3.2 | 1.39 | 32 | 63 | 38 | 62 | 3.7 |
| citrate c00158 | 1159.5 | 0.7 | 0.1029 | 13.0 | 1.41 | 83 | 83 | 63 | 87 | 12.4 |
| threonine c00188 | 777.0 | 0.6 | 0.0878 | 3.3 | 2.08 | 51 | 103 | 67 | 90 | 3.3 |
| malate c00149 | 871.5 | 0.7 | 0.0719 | 7.9 | 3.41 | 204 | 239 | 234 | 212 | 8.1 |
| 5-oxoproline c00025b | 906.0 | 1.4 | 0.0639 | 2.4 | 1.82 | 61 | 84 | 81 | 63 | 2.6 |
| glutamic acid c00025 | 987.0 | 0.8 | 0.0491 | 3.1 | 2.55 | 135 | 157 | 119 | 119 | 3.4 |
| tyrosine | 1245.0 | 0.6 | 0.0252 | 3.9 | 4.71 | 66 | 49 | 50 | 56 | 4.1 |
| homoserine | 834.0 | 0.6 | 0.0214 | 5.5 | 2.53 | 133 | 62 | 67 | 117 | 9.0 |
| glutamine c00064 | 1117.5 | 1.0 | 0.0188 | 1.9 | 1.72 | 38 | 60 | 21 | 31 | 2.5 |
| leucine | 657.0 | 0.6 | 0.0174 | 5.0 | 2.44 | 203 | 172 | 180 | 206 | 7.2 |
| succinate c00042 | 697.5 | 0.9 | 0.0148 | 3.0 | 1.82 | 21 | 51 | 52 | 45 | 3.6 |
| Adenosine, 5'-S-methyl-5'-thio# | 1803.0 | 1.1 | 0.0119 | 5.3 | 3.42 | 72 | 165 | 185 | 111 | 10.8 |
| mannitol c00392 | 1254.0 | 0.4 | 0.0082 | 0.4 | 3.32 | 64 | 30 | 52 | 31 | 0.2 |
| fumarate c00122 | 729.0 | 1.0 | 0.0058 | 8.5 | 3.93 | 208 | 219 | 171 | 122 | 14.9 |
| cystathionine | 1434.0 | 0.7 | 0.0054 | 0.3 | 1.20 | 28 | 22 | 42 | 25 | 0.3 |
| glycolic acid | 451.5 | 0.7 | 0.0048 | 2.6 | 6.12 | 21 | 11 | 20 | 6 | 2.9 |
| glycerol 3 phosphate c00093b | 1113.0 | 0.9 | 0.0024 | 2.8 | 2.09 | 16 | 22 | 33 | 25 | 3.3 |
| glucose 1 phosphate c00029b/c00103 | 1117.5 | 0.5 | 0.0020 | 0.6 | 1.79 | 37 | 82 | 51 | 105 | 0.6 |
| ethyl succinate* | 630.0 | 1.0 | 0.0018 | 1.6 | 1.98 | 19 | 21 | 37 | 20 | 1.7 |
| asparginine c00152 | 1033.5 | 0.9 | 0.0016 | 1.9 | 2.40 | 37 | 41 | 23 | 43 | 2.1 |
| octanoic acid | 643.5 | 0.6 | 0.0011 | 0.6 | 2.67 | 10 | 2 | 4 | 4 | 0.4 |
| threonic acid | 927.0 | 0.5 | 0.0006 | 0.3 | 1.84 | 10 | 6 | 15 | 16 | 0.7 |
| glutamine 3TMS c00064b | 861.0 | 0.1 | 0.0005 | 2.8 | 1.77 | 12 | 4 | 20 | 8 | 4.0 |
| glucose 6 phosphate | 1530.0 | 0.7 | 0.0004 | 0.3 | 1.21 | 66 | 23 | 76 | 30 | 0.3 |
| pseudo uridine | 1518.0 | 0.5 | 0.0004 | 1.7 | 3.91 | 37 | 58 | 54 | 55 | 1.9 |

Table 2.2 lists all the named metabolites discovered, quantified, and statistically validated by the Student's t-test [32] from the benchmark study [26]. Each replicate data subset of the tile-based F-ratio analysis is depicted in a column with the threshold listed at the top of the column in parentheses. For example, lysine was determined in the benchmark study to be the fourth most variant metabolite and validated by t-test. The ratio of class DR to R was quantified to be 3.2 by

PARAFAC. Each replicate data subset of the current analysis also discovered lysine at F-ratios greater than the threshold determined for each replicate data subset (e.g. F-ratio of 32 for replicate data subset 1 exceeds the threshold F-ratio value of 15). Discovery of a metabolite above the threshold provides a qualification of statistical significance in a similar fashion to the Student's t-test. If a statistically significant metabolite from the benchmark study was not discovered above the threshold, that cell is shaded grey.

Table 2.3 Results of tile-based F-ratio analysis for metabolites present in only one class.

| Metabolite | $t_{R1}$, s | $t_{R2}$, s | Pixel-based F-ratio | Sample Class | 1 (15) | 2 (22) | 3 (20) | 4 (20) |
|---|---|---|---|---|---|---|---|---|
| glucose | 1239.0 | 0.5 | 0.6443 | R only | 65 | 129 | 147 | 383 |
| glucopyranose | 1281.0 | 0.4 | 0.0825 | R only | 78 | 37 | 69 | 33 |
| glucopyranose | 1218.0 | 0.4 | 0.0546 | R only | 32 | 26 | 44 | 27 |
| butanoic acid 2,4 TMS TMS ester | 798.0 | 0.6 | 0.0146 | DR only | 8 | 12 | 10 | 12 |
| methyl citrate | 1231.5 | 0.7 | 0.0104 | R only | 61 | 27 | 97 | 53 |
| glucose 1 phosphate | 1168.5 | 0.4 | 0.0098 | R only | 32 | 41 | 33 | 62 |
| myo-inositol | 1363.5 | 0.5 | 0.0056 | DR only | 39 | 40 | 44 | 46 |
| adenine | 1192.5 | 0.7 | 0.0037 | DR only | 25 | 44 | 27 | 27 |
| galactofuranose | 1179.0 | 0.4 | 0.0036 | R only | 32 | 41 | 33 | 62 |
| fructose | 1204.5 | 0.4 | 0.0034 | R only | 112 | 34 | 137 | 43 |
| gluconic acid | 1024.5 | 0.8 | 0.0031 | R only | 24 | 33 | 31 | 29 |
| butanoic acid 2 amino | 553.5 | 0.6 | 0.0027 | DR only | 6 | 4 | 4 | 5 |
| glucose | 1207.5 | 0.5 | 0.0020 | R only | 30 | 33 | 133 | 55 |
| orotic acid | 1095.0 | 0.6 | 0.0019 | R only | 25 | 21 | 44 | 16 |
| 2 desoxy-pentose | 1108.5 | 0.4 | 0.0017 | R only | 37 | 82 | 51 | 105 |
| 2-o-glycerol-a-d-galactopyranoside | 1471.5 | 0.4 | 0.0009 | R only | 155 | 30 | 117 | 28 |
| fructose 1,6 bis phosphate | 1765.5 | 1.3 | 0.0008 | R only | 51 | 25 | 61 | 32 |
| fructose 2,6 bis phosphate | 1434.0 | 0.8 | 0.0007 | R only | 28 | 22 | 42 | 25 |
| ribofuranose | 987.0 | 0.4 | 0.0006 | R only | 32 | 29 | 33 | 24 |
| xylonic acid | 1006.5 | 1.0 | 0.0006 | R only | 25 | 29 | 38 | 34 |
| arabinonic acid lactone | 1038.0 | 1.3 | 0.0005 | R only | 9 | 9 | 15 | 27 |
| xylitol | 1069.5 | 0.5 | 0.0005 | R only | 82 | 42 | 61 | 56 |
| fructose 1,6 bis phosphate | 1761.0 | 1.2 | 0.0004 | R only | 102 | 23 | 58 | 32 |
| 6-phospho-d-gluconate c00345 | 1584.0 | 0.7 | 0.0004 | R only | 41 | 20 | 71 | 19 |
| 2-amino-adipinic acid | 1068.0 | 0.7 | 0.0003 | R only | 11 | 16 | 19 | 17 |
| n-acetyl glutamic acid | 1123.5 | 1.5 | 0.0003 | R only | 12 | 11 | 10 | 11 |

Table 2.3 shows similar information for metabolites that were only detected in one of the two classes. In Table 2.3 there is no column for Student's t-test ratio since it would be undefined; in place there is a column stating the biological class in which the metabolite was present. Table 2.4 follows a similar format from Table 2.2, but for the analytes determined to be statistically non-significant for distinguishing the DR and R classes by the Student's t-test. Table 2.4. Statistically verified false positives from the previous pixel-based F-ratio analysis [37] provides

remarkable results suggesting the resounding success of tile-based F-ratio analysis with respect to minimizing the FDR.

Table 2.4. Statistically verified false positives from the previous pixel-based F-ratio analysis

| Metabolite | $t_{R1}$, s | $t_{R2}$, s | Pixel-based F-ratio | [DR]/[R] PARAFAC | $t_{calc}/t_{table}$ | Subset 1 (15) | Subset 2 (22) | Subset 3 (20) | Subset 4 (20) |
|---|---|---|---|---|---|---|---|---|---|
| ornithine | 1155.0 | 0.6 | 0.1082 | 4.2 | 0.90 | 53 | 79 | 64 | 94 |
| isoleucine | 681.0 | 0.6 | 0.0136 | 1.7 | 0.88 | 13 | 18 | 11 | 24 |
| lactate | 436.5 | 0.6 | 0.0130 | 1.7 | 0.93 | 9 | 6 | 14 | 6 |
| serine | 748.5 | 0.6 | 0.0063 | 1.0 | 0.02 | 1 | 3 | 1 | 4 |
| alanine | 480.0 | 0.6 | 0.0043 | 0.9 | 0.28 | 4 | 4 | 4 | 2 |
| stearic acid | 1441.5 | 0.5 | 0.0032 | 1.1 | 0.22 | 2 | 1 | 4 | 17 |
| glucose 6 phosphate | 1518.0 | 0.6 | 0.0026 | 0.4 | 0.88 | 19 | 10 | 41 | 22 |
| 3-OH Propionic | 522.0 | 0.7 | 0.0023 | 2.0 | 0.71 | 1 | 1 | 1 | 1 |
| methionine | 900.0 | 0.7 | 0.0018 | 1.6 | 0.47 | 4 | 7 | 6 | 5 |
| glucose 1 phosphate* | 1168.5 | 0.5 | 0.0013 | 0.6 | 0.36 | Coelutes / indeterminate | | | |
| glycerate 3 phosphate | 1152.0 | 1.0 | 0.0011 | 1.8 | 0.79 | 1 | 5 | 5 | 7 |
| proline | 685.5 | 0.6 | 0.0008 | 1.3 | 0.46 | 13 | 2 | 1 | 5 |
| o-toluic acid | 750.0 | 0.8 | 0.0007 | 1.0 | 0.11 | 2 | 1 | 0 | 2 |
| pyruvate | 424.5 | 0.6 | 0.0005 | 0.4 | 0.74 | 4 | 6 | 6 | 3 |
| asparagine | 883.5 | 1.1 | 0.0005 | 1.5 | 0.98 | 6 | 8 | 6 | 15 |
| CoA phosphoric acid fragment | 663.0 | 1.2 | 0.0004 | 1.0 | 0.09 | 1 | 1 | 2 | 2 |
| arginine | 843.0 | 0.7 | 0.0004 | 1.1 | 0.17 | 3 | 1 | 4 | 3 |
| valine | 600.0 | 0.6 | 0.0004 | 1.0 | 0.05 | 2 | 2 | 1 | 2 |
| phenylalanine | 997.5 | 0.7 | 0.0004 | 1.0 | 0.05 | 4 | 3 | 1 | 4 |

The primary objective of the software is to provide greater segregation of true-positives toward high F-ratio values and false-positives toward low values, concurrent with mitigating the adverse effects of retention shifting. The addition of null distribution analysis provides a means to determine an F-ratio threshold to apply to the hit list, below which a marked increase in the FDR may be expected. The lack of discovery of the metabolites from Table 2.4. Statistically verified false positives from the previous pixel-based F-ratio analysis [37] is again accentuated. Compared to the previously applied pixel-based F-ratio analysis, tile-based F-ratio analysis minimizes the discovery of false positives. The software provides a much better ranking of the features, so the threshold can be applied confidently, saving the user analysis time.

Only one false positive metabolite was consistently discovered across all four replicate data subsets, namely ornithine. Indeed, this metabolite was barely excluded from the true positive list in the benchmark study, because the t-test was only a fraction smaller than the table

value and some minutiae of statistics necessitated the selection of a more stringent number of degrees-of-freedom. In fact, based on the present study, the potential that ornithine should be a true positive is proposed based on the magnitude of the calculated F-ratios and the ratio of the calculated t-value relative to the determined t-critical. The other two false positives, isoleucine and glucose-6-phosphate, were only discovered in one and three of the replicate data subsets respectively. Again, these two analytes' calculated t-test values were only a fraction shy of passing the test of statistical significance, suggesting they also were either potentially inappropriately excluded from the list of true hits or are sitting on the threshold almost exactly.

In the present study, the F-ratio analysis with the accompanying threshold that was applied provided a rapid determination of the major class distinguishing features while excluding virtually all non-class distinguishing features, i.e., minimizing false positives. As a consequence of the relatively stringent F-ratio threshold applied, some true positives from Table 2.2. Results of tile-based F-ratio analysis for metabolites present in both classes. were missed by the tile-based F-ratio analysis. The missed true positives were deemed acceptable in the interest of not exceeding the desired FDR for this demonstration of the combinatorial null distribution analysis. The "undiscovered" metabolites were generally near the bottom of validated hit lists from the previous study, suggesting they were originally more challenging to discover and are less class distinguishing. It should also be noted that many of the missed metabolites eluted in the earlier portion of the chromatogram where many of the reagent contaminants reside, further confounding their discovery. The benchmark study [26] was executed in brute force fashion to the lowest level achievable which brought the benefit of discovering additional true positives, but at a significant penalty to the FDR.

The signal values for each metabolite as quantified using ChromaTOF are also provided in Table 2.2. Results of tile-based F-ratio analysis for metabolites present in both classes.. Beneficially, the between class quantification ratios ([DR]/[R]) are in reasonable agreement relative to the PARAFAC quantification from the original report. The greatest quantification differences (between PARAFAC and ChromaTOF) are approximately a factor of two and likely result from quantification by only a single selective ion (ChromaTOF) relative to the top ten selective ions (PARAFAC). Most importantly, ChromaTOF quantification was completed in about one hour which is several orders of magnitude less than the time required for the PARAFAC analysis from the original report. Overall, the time required to complete the benchmark study by pixel-based F-ratio analysis followed by PARAFAC quantification, while state-of-the-art at the time, required twelve months. One iteration of the current study using tile-based F-ratio analysis followed by ChromaTOF quantification required one week.

The major thrust of this study was to evaluate the performance of tile-based F-ratio analysis in order to validate its effectiveness. This was achieved using a previously analyzed and statistically validated yeast metabolomics dataset. No dataset to this point has provided this level of challenge to the F-ratio software, which was specifically designed to deal with extreme shifting in both chromatographic dimensions. After removal of reagent artifacts and redundant hits resulting from chromatographic tailing of overloaded peaks, 94 class-distinguishing features were at or above the F-ratio threshold of 15. The tile-based F-ratio analysis successfully identified forty-six of the fifty-four changing metabolites from the benchmark study to be class-distinguishing features. The other forty features not covered in detail here were generally unknown, but none the less, very real metabolites. They are addressed in the benchmark study, but were excluded for clarity and brevity from detailed analysis here.

All four replicate data subsets analyzed by tile-based F-ratio analysis provided excellent results. Significantly, all four of the replicate data subsets are in good agreement and only vary at the boundary slightly where the true-positives end and the false-positives begin. Considering the power and speed of this method, the benefits to experimental design cannot be ignored. A true class comparison of these yeast data could have been completed with only six samples from each biological class significantly decreasing the time required for the chromatographic data collection campaign. This is in addition to the benefits gained in terms of the significantly shorter data analysis time requirements.

## 2.4 CONCLUSIONS

Tile-based F-ratio analysis combined with combinatorial null distribution determination of the threshold provides a powerful computational platform for rapid and confident determination of class distinguishing features in complex samples such as those encountered in metabolomics studies. This analysis is particularly useful in cases where chromatographic retention time shifting in both GC × GC dimensions renders other statistical and chemometric techniques untenable and/or substantially reduces their performance. The present study builds upon the three previous reports [29–31], since this study provides the first validation of the full algorithm in concert with validation of the utility of the threshold determined by combinatorial null distribution analysis. The validated benchmark data were rapidly replicated with a significant decrease in the FDR. When a key goal is to rapidly determine features which distinguish sample classes without the burden of overly populated peak tables, the computational platform presented herein is well suited.

## 2.5 REFERENCES


[1] S. Yang, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, Gas chromatography and comprehensive two-dimensional gas chromatography hyphenated with mass spectrometry for targeted and nontargeted metabolomics, in: Metabolomics Pract. Success. Strateg. Gener. Anal. Metab. Data, Wiley-VCH, Weinheim, Germany, 2013: pp. 69–92.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, Molecular biology of the cell, Fifth, Garland Science, New York, NY, 2008.

[3] C. Ibáñez, C. Simó, D.K. Barupal, O. Fiehn, M. Kivipelto, A. Cedazo-Mínguez, A. Cifuentes, A new metabolomic workflow for early detection of Alzheimer's disease, J. Chromatogr. A. 1302 (2013) 65–71, doi:10.1016/j.chroma.2013.06.005.

[4] B.C. Reaser, S. Yang, B.D. Fitz, B.A. Parsons, M.E. Lidstrom, R.E. Synovec, Non-targeted determination of 13C-labeling in the *Methylobacterium extorquens* AM1 metabolome using the two-dimensional mass cluster method and principal component analysis, J. Chromatogr. A. 1432 (2016) 111–121, doi:10.1016/j.chroma.2015.12.088.

[5] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics, Mass Spectrom. Rev. 26 (2007) 51–78, doi:10.1002/mas.20108.

[6] S.B. Adams, L.A. Setton, E. Kensicki, M.P. Bolognesi, A.P. Toth, D.L. Nettles, Global metabolic profiling of human osteoarthritic synovium, Osteoarthritis Cartilage. 20 (2012) 64–67, doi:10.1016/j.joca.2011.10.010.

[7] R. La Rosa, V. Behrends, H.D. Williams, J.G. Bundy, F. Rojo, Influence of the Crc regulator on the hierarchical use of carbon sources from a complete medium in *Pseudomonas*, Environ. Microbiol. 18 (2016) 807–818, doi:10.1111/1462-2920.13126.

[8] K.M. Pierce, J.C. Hoggard, Chromatographic data analysis. Part 3.3.4: handling hyphenated data in chromatography, Anal. Methods. 6 (2014) 645–653, doi:10.1039/c3ay40965a.

[9] Humston, Elizabeth M., Synovec, Robert E., Comprehensive chromatography data interpretation technologies, in: Compr. Chromatogr. Comb. Mass Spectrom., John Wiley & Sons, Inc., Hoboken, NJ, 2011: pp. 449–475.

[10] J.V. Seeley, S.K. Seeley, Multidimensional gas chromatography: fundamental advances and new applications, Anal. Chem. 85 (2013) 557–578, doi:10.1021/ac303195u.

[11] J. Dallüge, M. van Rijn, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection applied to the determination of pesticides in food extracts, J. Chromatogr. A. 965 (2002) 207–217, doi:10.1016/S0021-9673(01)01324-3.

[12] W. Welthagen, R.A. Shellie, J. Spranger, M. Ristow, R. Zimmermann, O. Fiehn, Comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry (GC × GC-TOF) for high resolution metabolomics: biomarker discovery on spleen tissue extracts of obese NZO compared to lean C57BL/6 mice, Metabolomics. 1 (2005) 65–73, doi:10.1007/s11306-005-1108-2.

[13] N.P. Vasquez, M. Crosnier de bellaistre-Bonose, N. Lévêque, E. Thioulouse, D. Doummar, T. Billette de Villemeur, D. Rodriguez, R. Couderc, S. Robin, C. Courderot-Masuyer, F. Moussa, Advances in the metabolic profiling of acidic compounds in children's urines achieved by comprehensive two-dimensional gas chromatography, J. Chromatogr. B. 1002 (2015) 130–138, doi:10.1016/j.jchromb.2015.08.006.

[14] E.M. Humston, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Time-Dependent Profiling of Metabolites from Snf1 Mutant and Wild Type Yeast Cells, Anal. Chem. 80 (2008) 8002–8011, doi:10.1021/ac800998j.
[15] D. Ryan, R. Shellie, P. Tranchida, A. Casilli, L. Mondello, P. Marriott, Analysis of roasted coffee bean volatiles by using comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry, J. Chromatogr. A. 1054 (2004) 57–65, doi:10.1016/j.chroma.2004.08.057.
[16] R.A. Shellie, W. Welthagen, J. Zrostliková, J. Spranger, M. Ristow, O. Fiehn, R. Zimmermann, Statistical methods for comparing comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry results: Metabolomic analysis of mouse tissue extracts, J. Chromatogr. A. 1086 (2005) 83–90, doi:10.1016/j.chroma.2005.05.088.
[17] L.C. Marney, S.C. Kolwicz Jr., R. Tian, R.E. Synovec, Sample preparation methodology for mouse heart metabolomics using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry, Talanta. 108 (2013) 123–130, doi:10.1016/j.talanta.2013.03.005.
[18] D. Bressanello, E. Liberto, M. Collino, S.E. Reichenbach, E. Benetti, F. Chiazza, C. Bicchi, C. Cordero, Urinary metabolic fingerprinting of mice with diet-induced metabolic derangements by parallel dual secondary column-dual detection two-dimensional comprehensive gas chromatography, J. Chromatogr. A. 1361 (2014) 265–276, doi:10.1016/j.chroma.2014.08.015.
[19] K.M. Pierce, J.L. Hope, J.C. Hoggard, R.E. Synovec, A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC × GC-TOFMS) separations of metabolites in plant samples, Talanta. 70 (2006) 797–804, doi:10.1016/j.talanta.2006.01.038.
[20] Y.F. Wong, S.-T. Chin, P. Perlmutter, P.J. Marriott, Evaluation of comprehensive two-dimensional gas chromatography with accurate mass time-of-flight mass spectrometry for the metabolic profiling of plant–fungus interaction in *Aquilaria malaccensis*, J. Chromatogr. A. 1387 (2015) 104–115, doi:10.1016/j.chroma.2015.01.096.
[21] P.-H. Stefanuto, K.A. Perrault, S. Stadler, R. Pesesse, H.N. LeBlanc, S.L. Forbes, J.-F. Focant, GC × GC–TOFMS and supervised multivariate approaches to study human cadaveric decomposition olfactive signatures, Anal. Bioanal. Chem. 407 (2015) 4767–4778, doi:10.1007/s00216-015-8683-5.
[22] P.-H. Stefanuto, K.A. Perrault, R.M. Lloyd, B. Stuart, T. Rai, S.L. Forbes, J.-F. Focant, Exploring new dimensions in cadaveric decomposition odour analysis, Anal. Methods. 7 (2015) 2287–2294, doi:10.1039/C5AY00371G.
[23] Z. Alves, A. Melo, A.R. Figueiredo, M.A. Coimbra, A.C. Gomes, S.M. Rocha, Exploring the *Saccharomyces cerevisiae* volatile metabolome: indigenous versus commercial strains, PLoS ONE. 10 (2015) e0143641, doi:10.1371/journal.pone.0143641.
[24] T.U. Greiner, T. Hyötyläinen, M. Knip, F. Bäckhed, M. Orešič, The gut microbiota modulates glycaemic control and serum metabolite profiles in non-obese diabetic mice, PLoS ONE. 9 (2014), doi:10.1371/journal.pone.0110359.
[25] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells, Anal. Chem. 78 (2006) 2700–2709, doi:10.1021/ac052106o.

[26] R.E. Mohler, K.M. Dombek, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, Comprehensive analysis of yeast metabolite GC × GC–TOFMS data: combining discovery-mode and deconvolution chemometric software, Analyst. 132 (2007) 756–767, doi:10.1039/B700061H.
[27] K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: comprehensive GC×GC with ANOVA-based feature selection and principal component analysis, Chemom. Intell. Lab. Syst. 60 (2002) 225–237, doi:10.1016/S0169-7439(01)00198-8.
[28] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher ratio method applied to third-order separation data to identify significant chemical components of metabolite extracts, Anal. Chem. 78 (2006) 5068–5075, doi:10.1021/ac0602625.
[29] L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data, Talanta. 115 (2013) 887–895, doi:10.1016/j.talanta.2013.06.038.
[30] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC × GC–TOFMS) data using a null distribution approach, Anal. Chem. 87 (2015) 3812-3819, doi:10.1021/ac504472s.
[31] B.A. Parsons, D.K. Pinkerton, B.W. Wright, R.E. Synovec, Chemical characterization of the acid alteration of diesel fuel: non-targeted analysis by two-dimensional gas chromatography coupled with time-of-flight mass spectrometry with tile-based Fisher ratio and combinatorial threshold determination, J. Chromatogr. A. 1440 (2016) 179–190, doi:10.1016/j.chroma.2016.02.067.
[32] D.C. Harris, Quantitative chemical analysis, 6th ed., W. H. Freeman and Company, New York, NY, 2003.
[33] R.E. Walpole, R.H. Myers, Probability and statistics for engineers and scientists, 2nd ed., MacMillan Publishing Co., Inc., New York, NY, 1978.
[34] D.L. Massart, B.G.M. VanDeginste, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics: a textbook, Elsevier Academic Press, Amsterdam, 1988.
[35] J.C. Hoggard, R.E. Synovec, Parallel Factor Analysis (PARAFAC) of target analytes in GC × GC−TOFMS data: automated selection of a model with an appropriate number of factors, Anal. Chem. 79 (2007) 1611–1619, doi:10.1021/ac061710b.
[36] K.M. Pierce, L.F. Wood, B.W. Wright, R.E. Synovec, A comprehensive two-dimensional retention time alignment algorithm to enhance chemometric analysis of comprehensive two-dimensional separation data, Anal. Chem. 77 (2005) 7735–7743, doi:10.1021/ac0511142.
[37] J.S. Nadeau, B.W. Wright, R.E. Synovec, Chemometric analysis of gas chromatography–mass spectrometry data using fast retention time alignment via a total ion current shift function, Talanta. 81 (2010) 120–128, doi:10.1016/j.talanta.2009.11.046.
[38] G. Tomasi, F. van den Berg, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, J. Chemom. 18 (2004) 231–241, doi:10.1002/cem.859.
[39] B. Efron, Bayesians, frequentists, and scientists, J. Am. Stat. Assoc. 100 (2005) 1–5, doi:10.1198/016214505000000033.

[40] B. Efron, R. Tibshirani, Empirical bayes methods and false discovery rates for microarrays, Genet. Epidemiol. 23 (2002) 70–86, doi:10.1002/gepi.1124.

# Chapter 3. Comprehensive Three-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry[3]

## 3.1 INTRODUCTION

The promise of comprehensive two-dimensional (2D) gas chromatography (GC × GC) as described by Giddings [1] and realized by Philips and Liu [2] has been largely recognized within the twenty-first century analytical laboratory. The discourse has moved beyond conceptualization, to broad implementation and optimization. Multiple reviews are available on the topic covering many aspects of the field from application and instrumentation [3,4], to data analysis [5]. The most current research focuses on optimizing the total peak capacity of the 2D separation while maintaining the full information imparted by the primary separation dimension [4,6–11]. Enhancement of the chemical selectivity is also provided through the development and use of novel stationary phases and important applications [4,12–14]. Indeed, GC × GC benefits greatly from the large 2D peak capacity that can be regularly provided, in the range of ~ 4,000 to 7,000 [9–11,15], while a relatively long duration, high efficiency one-dimensional (1D) GC separation provides a peak capacity in the range of ~ 600 to 900 [16–19]. In addition, GC × GC provides the added benefit of increased chemical selectivity through the use of the second dimension separation.

In order to further enhance chemical selectivity, it is intriguing to consider higher order instruments, specifically, to provide comprehensive three-dimensional (3D) separations. Although the development of comprehensive 3D separation techniques is still in its infancy,

[3] This Chapter has been reproduced from N.E. Watson, H.D. Bahaghighat, K. Cui, R.E. Synovec, Comprehensive Three-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry, Anal. Chem. 89 (2017) 1793–1800. doi:10.1021/acs.analchem.6b04112.

there are some notable examples based upon various combinations of GC, liquid chromatography (LC), and capillary electrophoresis (CE). In particular LC × LC × CE [20], LC × GC × GC [21], and GC × GC × GC [22,23], have all been reported. All of these "classical" comprehensive 3D separations produce a data cube in which each of the separated sample components literally "hover" in 3D space at a set of three retention time coordinates. Various intriguing options abound, for data reduction from the 3D space to 2D and 1D domains [22] to allow the analyst the opportunity to take advantage of the added chemical selectivity of the 3D separation by focusing on critical regions of interest in the 3D separation. In the strictly GC field, a few related designs have been reported that both used a flow switching device, such as GC x 2GC [24], in which two GC × GC separations are simultaneously produced, and a hybrid GC instrument that coupled GC × GC sequentially with a third GC separation for the purpose of excising the components eluting from a critical region of the GC × GC for more detailed separation on the third GC separation dimension [25].

The previous developments of GC × GC × GC, referred to herein as $GC^3$, were fruitful and encouraging [22,23]. In these initial reports, an FID was implemented for detection, and two diaphragm valves served as modulators, one valve between the primary column $^1$D and secondary column $^2$D separations, and the other valve between the secondary column $^2$D and tertiary column $^3$D separations, respectively. In the first $GC^3$ – FID report, a 3D peak capacity of 3,500 was achieved, and the high quality trilinear data was demonstrated to leverage the benefits of chemometric analysis using PARAFAC [22]. In the subsequent report, the $GC^3$ – FID instrument was demonstrated to provide unique insight into visualizing various chemical compound classes by leveraging the enhanced chemical selectivity of the 3D separation, in particular using an ionic liquid stationary phase column for one of the three separation

dimensions [23]. In this second report the peak capacity production was 4-fold better than the initial report, going from about 45 peaks/min to 180 peaks/min, with a 3D peak capacity of 3,600 achieved in only a 20 min separation of diesel fuel [23]. While use of the FID was satisfactory for the initial development of $GC^3$, limitations of the FID for providing confident analyte identification are obvious. Also, the diaphragm valves had to be face mounted on the wall of the GC oven in order to partially overcome the temperature limitations of these valves [26]. The face mounted diaphragm valves were able to reliably function up to about 250 °C, which was suitable for demonstrating $GC^3$, but not suitable in the long term for wider adoption of this separation technology. In order for $GC^3$ to be more rigorously evaluated as a separation technology platform, and to gain interest in the separations field, it is imperative that $GC^3$ be combined with a more informative method of detection such as a time-of-flight mass spectrometry (TOFMS) [27,28] and to use modulators that overcome the temperature limitations of the face mounted diaphragm valves.

Herein, we describe significant improvements to the previously described $GC^3$ – FID instrument by replacing the FID with TOFMS. The $GC^3$ – TOFMS instrument reported herein includes the major additions of a high-temperature diaphragm valve that has been recently demonstrated to reliably function to 325 °C [29,30], to serve as one of the two modulators, and the stock thermal modulator within the commercial instrument platform with the TOFMS to serve as the other modulator. These improved components bring the benefits of increased operating temperature, decreased sample splitting, and added mass spectral selectivity with peak identification. The $GC^3$ – TOFMS instrument also increases the order of the data, naturally producing fourth order data, which opens up new data analysis options unavailable at lower levels of data dimensionality. This report provides a proof-of-principle for $GC^3$ – TOFMS, and

demonstrates the benefits of the added chemical selectivity afforded by three chemical stationary phases while maintaining if not exceeding the total separation peak capacity of GC × GC – TOFMS. Instrumental design considerations are made, anchored in comprehensive multi-dimensional separation theory, to design the $GC^3$ – TOFMS instrument to maximize the combined peak capacity of the $^2$D and $^3$D separation, which nominally produces a GC × GC – TOFMS separation at every modulation along the $^1$D separation. Two samples were used for demonstration purposes, a test mixture of 115 compounds, and a diesel fuel spiked with non-native compounds which we focus on to highlight the selectivity benefits of $GC^3$ – TOFMS.

## 3.2 Basic Principles and Instrumental Design

For GC × GC, the ideal peak capacity is given by

$$n_{c,2D} = {}^1n_c * {}^2n_c \tag{3.1}$$

where $^1n_c$ and $^2n_c$ are the $^1$D and $^2$D peak capacities, respectively. With the addition of a third separation dimension, the ideal peak capacity for $GC^3$ is given by

$$n_{c,3D} = {}^1n_c * {}^2n_c * {}^3n_c \tag{3.2}$$

where $^3n_c$ is the $^3$D peak capacity. At unit resolution, $R_s = 1$, eq 3.2 can be expressed as,

$$n_{c,3D} = \frac{{}^1t}{{}^1w}\frac{{}^2t}{{}^2w}\frac{{}^3t}{{}^3w} \tag{3.3}$$

where $^2t$ is equivalent to the modulation period, $^1P_m$, for coupling the $^1$D to $^2$D separations, and $^3t$ is equivalent to the modulation period, $^2P_m$, for coupling the $^2$D and $^3$D separations. The nominal peak width-at-base (4σ) for each dimension is given by $^1w$, $^2w$, and $^3w$, respectively.

In terms of modulation periods instead of separation run times of $^2$D and $^3$D, eq 3.3 can be rearranged and expressed as,

$$n_{c,3D} = \frac{{}^{1}P_M}{{}^{1}w}\frac{{}^{2}P_M}{{}^{2}w}\frac{{}^{1}t}{{}^{3}w} \quad (3.4)$$

The arrangement of eq 3.4 facilitates use of the modulation ratio $M_R$, the ratio of the peak width-at-base for a given separation relative to the modulation period coupling the given separation to a subsequent separation [8]. Substitution of ${}^{1}M_R$ and ${}^{2}M_R$, into eq 3.4 results in the following,

$$n_{c,3D} = \frac{1}{{}^{1}M_R}\frac{1}{{}^{2}M_R}\frac{{}^{1}t}{{}^{3}w} \quad (3.5)$$

Peak width minimization on all dimensions per eq 3.4 is critical to optimize $n_{c,3D}$, particularly ${}^{3}w$ for the ${}^{3}$D separation. However, peak width minimization must be balanced with proper selection of ${}^{1}P_m$ and ${}^{2}P_m$ to provide suitable ${}^{1}M_R$ and ${}^{2}M_R$, in order to provide a comprehensive 3D separation to ensure quantitative data [8,31–33]. Here we strive to fully optimize the 3D peak capacity production of the instrument while not degrading the quantitative precision due to valve-based modulator undersampling. A $M_R$ of 2 or greater is needed to ensure the %RSD due to valve-based modulation is between 1% - 2% [8,31–33]. Moreover, the negative implications of statistical overlap and modulator induced band broadening are greatest at low $M_R$ values [7,11,34–36].

Equation 3.5 provides a useful vehicle to theoretically estimate and provide guidance for experimentally maximizing the peak capacity that could be achieved with GC$^3$. Assuming implementation of conditions to achieve a $M_R$ of 2 for both modulation interfaces and an average ${}^{3}w$ of 50 ms, an $n_{c,3D}$ of 15,000 could be achieved for a ${}^{1}t$ of 50 min. An example of conditions to achieve this impressive result could be a ${}^{1}w$ of 6 s, with a ${}^{1}P_m$ of 3 s (separation time on ${}^{2}$D) and ${}^{2}w$ of 500 ms, coupled with ${}^{2}P_m$ of 250 ms (separation time on ${}^{3}$D) and ${}^{3}w$ of 50 ms. The individual peak capacities would then be ${}^{1}n_c$ of 500, ${}^{2}n_c$ of 6, and ${}^{3}n_c$ of 5. It is noteworthy that the product of ${}^{2}n_c$ by ${}^{3}n_c$ is a theoretical peak capacity of 30, which is much higher than typically

obtained for the $^{2}$D dimension of GC × GC [9–11,16,34], with the added benefit of an additional dimension of separation providing more selectivity. Doing the same calculation, with conditions producing a $M_R$ of 2.5 for both modulation interfaces while holding the $^{1}t$ and $^{3}w$ constant, the resulting $n_{c,3D}$ drops from 15,000 to 9,600. Thus, the challenge in instrument design and experimental implementation is to produce low $M_R$, at or approaching a value of 2, for both modulation steps. Experimentally meeting this challenge is investigated in this report.

## 3.3 EXPERIMENTAL

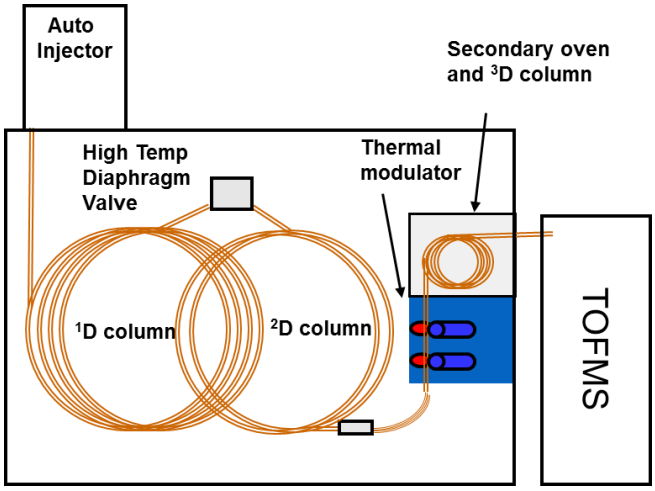


Figure 3.1 Schematic of the major components of the GC3 – TOFMS instrument.

A Pegasus 4D GC × GC – TOFMS (LECO Corporation, St. Joseph, MI) with an integrated Agilent 6890N Gas Chromatograph (Agilent Technologies, Santa Clara, CA, USA) was modified to produce a GC$^{3}$ – TOFMS instrument as shown in Figure 3.1. One high-speed, six port diaphragm valve (Valco Instruments Company Inc, Houston, TX, USA), upgraded by the manufacturer to operate at a maximum temperature of 325 °C [26,29,30], and fitted with a 5 µL sampling loop was installed in the GC oven. The high-temperature valve was utilized as the modulator between $^{1}$D and $^{2}$D separations, and the stock thermal modulator was implemented

between the $^{2}$D and $^{3}$D separations. The $^{3}$D column was contained within the secondary oven in the commercial instrumental platform to provide a small temperature offset.

Three columns with different stationary phases were installed in column 1, column 2 and column 3 positions as depicted in Figure 1 for the $^{1}$D, $^{2}$D, and $^{3}$D separations, respectively. A 30 m, 250 μm inner diameter (id), 0.50 μm film thickness (5% phenyl)-methyl polysiloxane stationary phase column (Rtx-5; Restek, Bellefonte, PA, USA) was installed as column 1. A 3.5 m, 180 μm id, 0.18 μm film thickness polyethylene glycol stationary phase column (Rtx-Wax; Restek, Bellefonte, PA, USA) was installed as column 2. Finally, a 1 m, 100 μm id, 0.1 μm film thickness trifluoropropyl-methyl polysiloxane stationary phase column (Rtx-200; Restek, Bellefonte, PA, USA) was installed as column 3. The orthogonality of these phases was

Table 3.1. 115 Component Text Mixture Composition

| **Alkanes** | **Unsaturated Alkanes** | **Cyclic alkanes** | **Halogenated alkanes** | **Alcohols** | **Ketones & Esters** | **Aromatics** |
|---|---|---|---|---|---|---|
| pentane | 1-hexene | methylcyclopentane | 1-chloroform | 1-propanol | ethyl formate | benzene |
| hexane | 1-heptene | cyclohexane | 1-bromoheptane | 2-butanol | methyl decanoate | toluene |
| heptane | 1-undecene | methylcyclohexane | 1,1,1-trichloroethane | 1-pentanol | methyl caprylate | naphthalene |
| octane | dodecene | cyclooctane | carbon tetrachloride | 1-hexanol | methyl salicylate | mesitylene |
| nonane | 1-hexyne | butylcyclohexane | 1-chlorobutane | 2-heptanol | ethyl salicylate | ethylbenzene |
| decane | 1-heptyne | adamantane | 1-bromooctane | 1-octanol | methyl laurate | butylbenzene |
| undecane | 1-nonyne | bicyclohexane | 1,2-dichloroethane | 1-nonanol | methly caproate | isobutylbenzene |
| dodecane | 5-decyne | cyclopentane | 1,6-dichlorohexane | 1-decanol | diethyl phthalate | tert-butyl benzene |
| tridecane | | cis-1,2-dimethylcyclohexane | 1-chlorohexane | 1-geraniol | 2-butanone | propylbenzene |
| tetradecane | | 2,2,4-trimethylpentane | 1-bromohexane | 1-dodecanol | 2-pentanone | chlorobenzene |
| pentadecane | | 2,3,4-trimethylpentane | | 1-tetradecanol | 3-hexanone | bromobenzene |
| hexadecane | | 2-methylpentane | | 1-hexadecanol | 3-heptanone | 1,3,5-trichlorobenze |
| pristane | | R-(-)-2,6-dimethyloctane | | 2-pentanol | 3-octanone | 1,2,3-trichlorobenze |
| heptadecane | | | | 1-butanol | 2-nonanone | cyclohexylbenzene |
| octadecane | | | | 1-eicosanol | 2-decanone | sec-butyl benzene |
| nonadecane | | | | 1-octadecanol | 2-undecanone | phenanthrene |
| eicosane | | | | cyclohexanol | 2-dodecanone | p-xylene |
| | | | | neopentyl alcohol | 2-pentadecanone | o-xylene |
| | | | | 2-methyl-2-propanol | 2-hexanone | m-xylene |
| | | | | tert-amyl alcohol | 2-heptanone | 1,2,4,5-tetrachlorob |
| | | | | isobutyl alcohol | | 1,2,4-trimethylbenz |
| | | | | isopropyl alcohol | | anisole |
| | | | | benzyl alcohol | | benzophenone |
| | | | | 1,2-propanediol | | |

previously evaluated in detail and the phases were shown to be complementary [22]. Wrap around was also intentionally applied to ensure maximal use of the separation space [22].

Table 3.2. Non-native components spiked into diesel.

| | |
|---|---|
| 1-chlorohexane | butyrophenone |
| 1,6-dichlorohexane | ethyl salicylate |
| 5-decyne | 2-decanone |
| limonene | dibutylphthalate |
| bromobenzene | 2,5-dimethylthiophene |
| pyridine | diphenyl sulfide |
| 2-heptanol | cyclohexyl isothiocyanate |
| 3-octanone | 1-dodecanethiol |
| methyl caproate | aniline |
| 1-nonanol | 2-mercaptoethanol |

Liquid injections of 1 μL of both a test mixture of 115 compounds and diesel spiked with non-native compounds were made with an Agilent 7683 auto-injector. The composition of the 115 compound test mixture is detailed in Table 3.1. The list of non-native compounds spiked into the diesel are provided in Table 3.2. The inlet was operated at 250 °C in split mode with a split flow of 25 mL/min. Effluent from $^{1}$D was injected onto $^{2}$D with the high-temperature diaphragm valve set to actuate for 400 ms at a modulation period $P_m$ = 3 s. Effluent from $^{2}$D was transferred to $^{3}$D via the thermal modulator with a hot pulse time of 120 ms at a $P_m$ = 250 ms, as previously reported [9]. The $^{1}$D column was operated at a constant volumetric flow of 0.5 mL/min. The $^{2}$D and $^{3}$D columns were controlled via the auxiliary pressure controller under a ramped pressure program held at 25 psi for the first 3 min and ramped to 35 psi at a rate of 0.211 psi/min and held at the final pressure for 1 min. This program resulted in an approximate volumetric flow on $^{2}$D and $^{3}$D columns of 1 mL/min. The $^{1}$D column was operated at less than the optimum flow rate in order to slightly widen the peaks in the $^{1}$D dimension to ensure sufficient peak width and sampling rate at the valve-based modulator. Detection was accomplished with the TOFMS. Effluent was passed through a 0.33 m, 280 °C transfer line into the TOFMS where it was analyzed at 200 Hz between 33 Da to 334 Da. The GC oven was

operated with a temperature program starting at 60 °C and held at that temperature for 3 min. The oven was then ramped to 250 °C at a rate of 4 °C/min. The final temperature was held for 1 min. The secondary oven was held at a constant 10 °C offset above the primary oven temperature. The 115-component test mixture and spiked diesel were both analyzed in triplicate.

All data were collected using ChromaTOF 3.32 and transferred to MATLAB 2016a (The Mathworks, Natick, MA, USA) using in-house software (peg2mat3p8) [37]. All chromatograms were baseline corrected in a 1D fashion and folded into a four-way array (4D data). Compounds were identified through a library search utilizing MS Search 2.0 (NIST, Gaithersburg, MD, USA). All multi-dimensional visualization was achieved using functions and utilities included with MATLAB 2016a. The peak width and retention time, $t_R$, of 10 representative compounds were measured by Gaussian Curve Fitting utilizing the Curve Fitting Toolbox available as an add-on application for MATLAB 2016a.

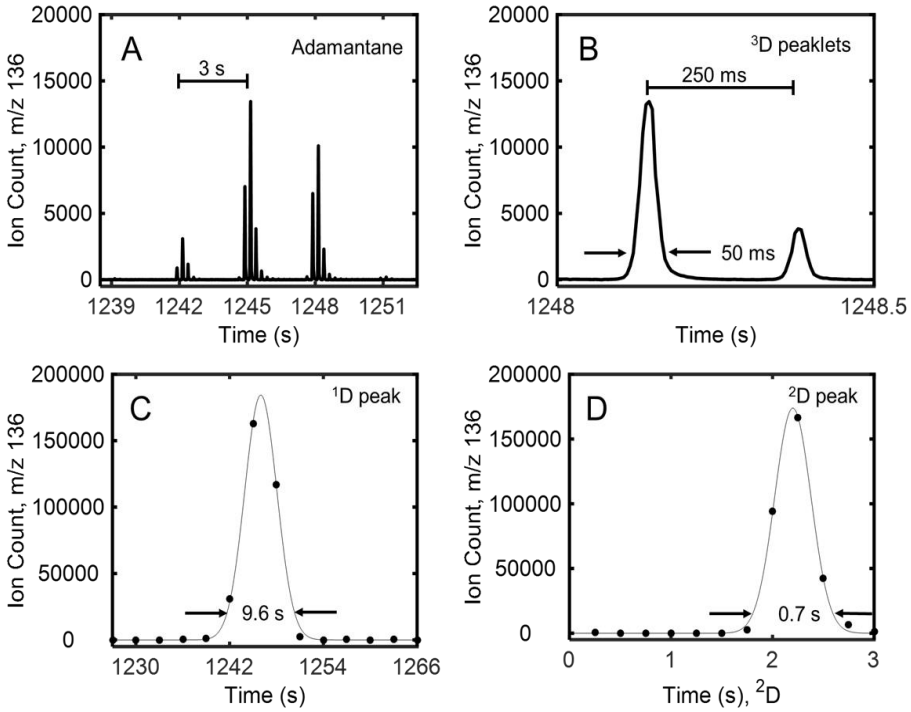


Figure 3.2. Chromatogram for adamantane viewed in all dimensions.

## 3.4 RESULTS AND DISCUSSION

Data collected for adamantane using the $GC^3$ – TOFMS instrument are presented in Figure 3.2, with adamantane serving as a representative analyte in the 115-component test mixture. All subfigures were created using *m/z* 136, a highly selective *m/z* for this compound. Figure 3.2A depicts the raw, baseline corrected data vector at *m/z* 136 showing the expected results for a 3D separation, analogous to a 2D separation. The $^1$D peak is sampled with a $^1P_m$ of 3 s and the resultant $^2$D peaklets are then sampled by $^3$D with a $^2P_m$ of 250 ms. The term "peaklets" refers to a set of peaks in a given separation dimension produced by modulating the analyte peak eluting from the preceding dimension, eg., $^3$D peaklets resulting from modulating a given $^2$D peak.[10,34] Figure 3.2B shows detail of the region in Figure 3.2A with the highest signal intensity, so one can draw their own conclusions regarding the peak widths and peak capacity on the $^3$D dimension. Finally, Figure 3.2C & D provide the peak profiles for adamantane on the $^1$D and $^2$D dimensions, after summing the signal from the remaining two axes. Figure 3.2B-D illustrate how peak widths were measured for subsequent determination of modulation ratios and separation peak capacity on each separation dimension for adamantane and other representative analytes. As indicated in Figure 3.2B, a $^3w_b$ of 50 ms was measured. Using a $^2P_m$ of 250 ms, a $^3n_c$ of 5.0 was determined for adamantane. Figure 3.2C & D were used to measure a $^1w_b$ of 9.6 s and a $^2w_b$ of 0.7 s, but were first fitted to a Gaussian profile to do so. With a $^2P_m$ of 3 s, the $^2n_c$ was 4.3. Finally, with these measured peak widths, a $^1M_R$ of 3.2, and a $^2M_R$ of 2.8 were determined.

Following the same approach illustrated with adamantane, additional measurements and figures-of-merit calculations for five other representative compounds from the separation of the 115-component test mixture and four representative compounds from the spiked diesel fuel separation, *vide infra*, are summarized in Table 3.3. The ten compounds in Table 3.3 were

chosen to represent a wide variety of chemical functional classes including varying degrees of saturated, unsaturated and aromatic hydrocarbons along with alcohol and ketone functionality. Though this is but a small portion of the functionality encountered in a complex chemical analysis, this set of compounds provides a good breadth of chemical functionality and depicts the wide range of chromatographic performance that can be expected across varied compound classes.

Table 3.3 Chromatographic peak measurements and figures-of-merit.

| | 115-component test mixture | | | | | | Spiked Diesel | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | butyl-benzene | 1-octanol | 1-decene | 2-nonanone | undecane | adamantane | pyridine | limonene | 1,6-dichlorohexane | cyclohexyl isothiocyanate |
| $^1t_R$ (min) | 19.5 | 20.0 | 20.4 | 20.7 | 20.7 | 20.9 | 8.2 | 18.4 | 22.6 | 26.7 |
| $^1w$ (s) | 7.2 | 5.9 | 7.8 | 8.1 | 6.3 | 9.6 | 5.4 | 9.3 | 7.3 | 9.8 |
| $^1M_R$ | 2.4 | 2.0 | 2.6 | 2.7 | 2.1 | 3.2 | 1.8 | 3.1 | 2.4 | 3.3 |
| $^1n_C$ | 416 | 506 | 385 | 368 | 477 | 311 | 556 | 323 | 411 | 306 |
| $^2t_R$ (s) | 2.2 | 1.5 | 1.7 | 1.5 | 0.9 | 2.2 | 0.7 | 1.8 | 0.6 | 1.0 |
| $^2w$ (s) | 0.9 | 1.3 | 0.7 | 0.8 | 0.7 | 0.7 | 1.2 | 0.6 | 0.9 | 1.0 |
| $^2M_R$ | 3.5 | 5.3 | 2.7 | 3.3 | 3.0 | 2.8 | 4.8 | 2.5 | 3.7 | 3.9 |
| $^2n_C$ | 3.5 | 2.3 | 4.4 | 3.6 | 4.0 | 4.3 | 2.5 | 4.8 | 3.2 | 3.1 |
| $^3t_R$ (ms) | 165 | 195 | 100 | 85 | 85 | 140 | 155 | 50 | 225 | 215 |
| $^3w$ (ms) | 53 | 58 | 47 | 59 | 50 | 50 | 58 | 46 | 47 | 49 |
| $^3n_C$ | 4.7 | 4.3 | 5.3 | 4.3 | 5.0 | 5.0 | 4.3 | 5.4 | 5.3 | 5.1 |
| $n_{C,3D}$ | **6803** | **4973** | **9106** | **5635** | **9639** | **6821** | **5987** | **8352** | **7010** | **4850** |

The data portions depicted in Figure 3.2 were collected as part of the analysis of the test mixture. Figure 3.3 delves beyond adamantane into the additional benefits gained through the use of the improved GC$^3$ instrument with TOFMS detection. These improvements include mass spectral selectivity and peak identification, and increased maximum temperature while still maintaining the high chromatographic selectivity afforded through the use of three complementary stationary phases. The complete set of images in Figure 3.3 serve as an illustrative depiction of what becomes possible during the analysis of a truly complex sample. Figure 3.3A depicts the full separation of the test mixture, but simplifies the data by summing the $^2$D and $^3$D chromatographic dimensions and mass spectral dimension onto $^1$D (Rtx-5). This provides a figure analogous to a traditional GC-MS total ion current (TIC) chromatogram. Immediately the benefits of mass spectral peak identification become apparent with even the relatively simple test mixture. Figure 3.3B zooms to a region of interest in Figure 3.3A where the

potential for improved selectivity of $GC^3$ – TOFMS can be explored; the compounds butyl-benzene, 1-octanol, 1-decene, 2-nonanone, undecane and adamantane all elute here in the order listed. Even though most of the compounds are baseline resolved on the 1D separation, they were selected to provide clarity in the following illustration.

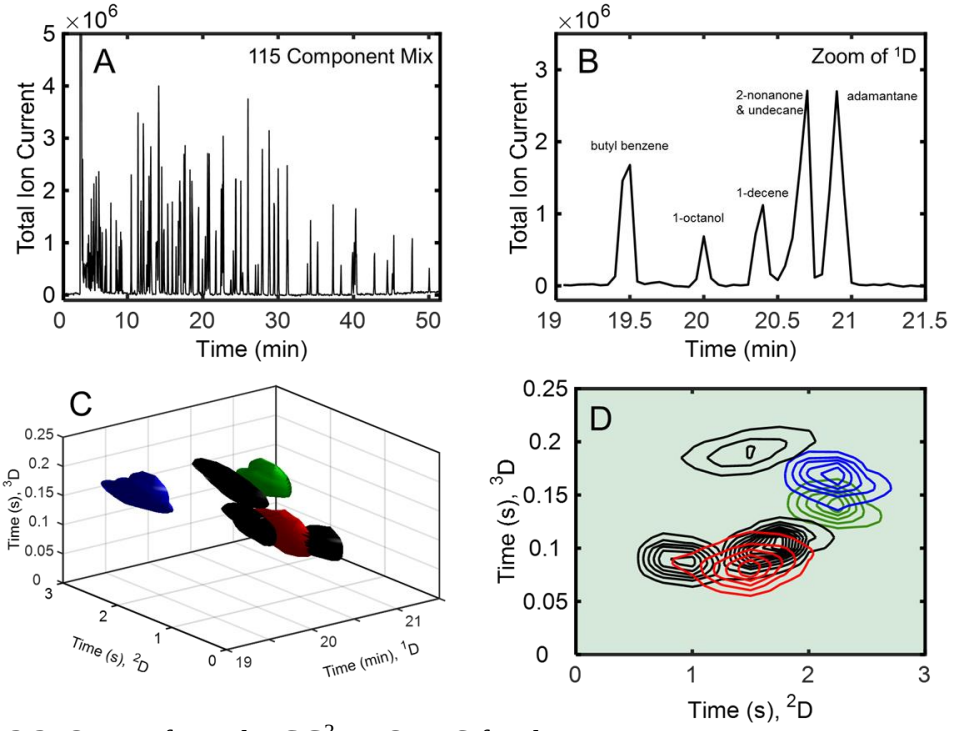


Figure 3.3. Output from the $GC^3$ – TOFMS for the 115 component test mixture

Modern enhancements to 3D visualization allow for significant improvement relative to previous reports on $GC^3$ by way of graphic access to the full 4D selectivity. Figure 3.3C provides a realization of these improvements, a 3D isosurface plot of the same region shown Figure 3.3B. The software connects the dots in 3D space where the signal intensity achieves a user selected value creating a cloud representing the data. The addition of color allows for various *m/z* to be depicted on the same figure. Additional "contours" can be added at different intensities by varying opacity to achieve something akin to a 2D contour plot in 3D space. This feature is not depicted here due to the challenge of presenting such detail on a static 2D surface. Figure 3.3C includes *m/z* 56 (1-octanol, 1-decene and undecane), *m/z* 58 (2-nonanone), *m/z* 91 (butyl-

benzene), and *m/z* 136 (adamantane) at intensities of 500, 1000, 2000, and 800 ion counts respectively. The *m/z* were chosen due to their selectivity for the analytes eluting within the region depicted.

Figure 3.3D shows the chemical selectivity provided by $GC^3$ for the 2D separation on the $^2D$ and $^3D$ dimensions, isolated from the $^1D$ separation for the same analytes per *m/z* shown in Figure 3.3C. In Figure 3D the region of the $^1D$ separation between 19.0 and 21.5 min is summed leaving a 2D contour plot showing the chromatographic profiles for the same compounds as in Figure 3.3B & C on $^2D$ and $^3D$. Note that 2-nonanone and undecane are overlapped in all three chromatographic dimensions, as a result of applying wraparound to maximize peak capacity usage, and only the addition of the TOFMS enabled their resolution. The added selectivity of TOFMS detection will become particularly important as we move on to a truly complex mixture.

The limit-of-detection (LOD) for $GC^3$ – TOFMS was evaluated using adamantane, which was found to be a representative analyte for the other compounds in the test mixture. The injected concentration LOD was ~ 10 ppm at *m/z* 136. A 30:1 split was applied for injection onto the $^1D$ column, and ~ 15% of the material eluting from the $^1D$ column was transferred by the high-temperature valve to the $^2D$ column, for an overall split of ~ 210:1. Thus, the detected mass LOD was ~ 40 pg (*S/N* = 3). This LOD is suitable for this proof-of-principle report, however future designs would benefit by having a thermal modulator for both modulation stages.

The more rigorous evaluation of $GC^3$ – TOFMS is provided by the analysis of diesel fuel, in this study spiked with a mixture of non-native compounds (see Supplemental Material for spiked compound table). Figure 3.4 showcases a portion of the results of this analysis. Figure 4A depicts a reduction of the dimensionality down to a $^1D$ chromatogram akin to Figure 3A, and again provides a familiar benchmark separation. Due to the magnitude of the detector response

generated by the diesel sample, Figure 3.4A is the analytical ion chromatogram (AIC) constructed from summing only *m/z* 41, 43, 53, 55, 74, and 91. Indeed, use of the TIC overwhelms the eluting compounds, reducing the apparent resolution to near zero in a plot of the $^{1}$D separation resulting from the presence of significantly more baseline noise in the entire GC$^{3}$ separation space, which obfuscates the summation of signal of all *m/z* across the $^{2}$D and $^{3}$D separations.

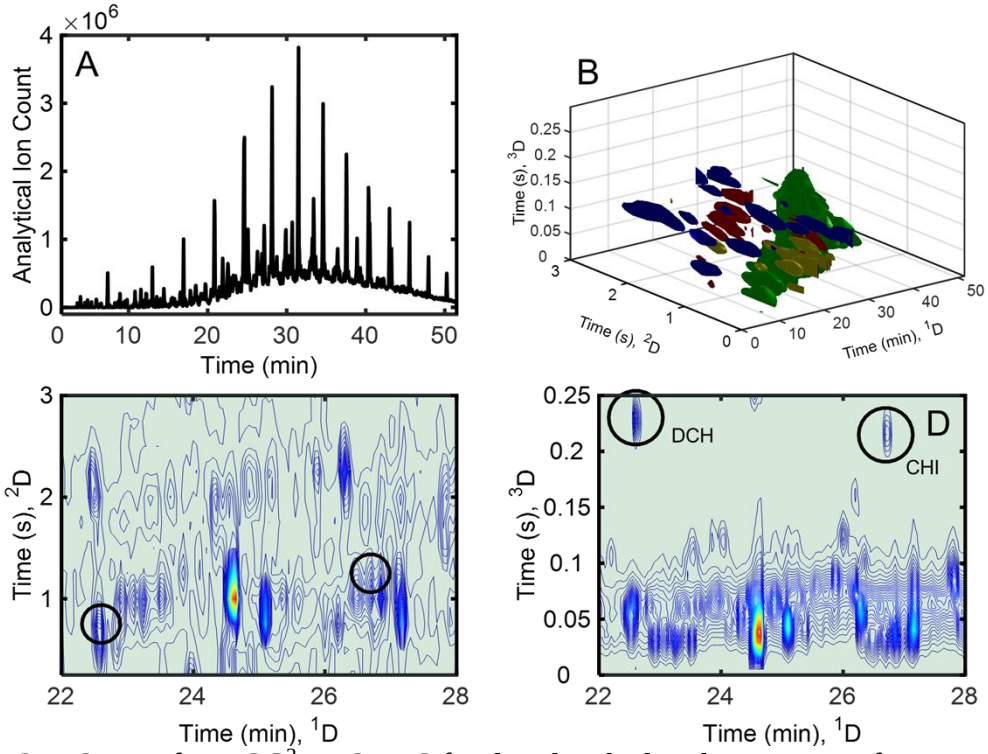


Figure 3.4. Output from GC$^{3}$ – TOFMS for diesel spiked with a mixture of non-native compounds

Figure 3.4B further showcases the true power of GC$^{3}$ – TOFMS. Here, another isosurface plot is shown using the same function described for Figure 3.4C. The complexity of these spiked diesel data allows for a more detailed visualization of the potential of GC$^{3}$ – TOFMS to tackle challenging chemical analysis problems. In this case, the signal at *m/z* 41, 43, 53 and 91 are displayed at intensities of 5000, 5000, 1000, and 2500 ion counts, respectively. The intensities at each contour are varied to scale the size of the clouds so the different *m/z* do not obfuscate each

other, to ensure analyte peaks at lower *S/N* are visible. Considering that the primary composition of diesel fuel is a mixture of hydrocarbons with various degrees of unsaturation and aromaticity we have included four *m/z* that are generally selective for those compound classes. Usefully, the phenomenon whereby homologous series of compounds of the same functional groups elute along a line during 2D GC analysis is shown to exist in three dimensions as well. The green (*m/z* 43) alkane band is most obvious, but the red (*m/z* 41) mono-unsaturated hydrocarbon and yellow (*m/z* 53) di-unsaturated hydrocarbon series are also apparent though both wrap around on $^{2}$D. The varied molecular mass and structures of the aromatics depicted in blue (*m/z* 91) do not create a linear spatial array of peaks in the same fashion as the various alkanes, alkenes, and alkynes; however, the aromatics do separate nicely as a group along the $^{3}$D separation dimension, showcasing the additional chromatographic selectivity by completely isolating the aromatic compounds from the bulk of the hydrocarbons present. This is useful for fingerprinting complex samples which may have a variety of compound classes present.

The benefit of having three dimensions of chromatographic selectivity is illustrated in Figure 3.4C & D. In Figure 3.4C is presented a contour plot of the $^{1}$D separation versus the $^{2}$D separation using the AIC summation of *m/z* 41, 43, 53, 55, 74, and 91, by summing the signal along the $^{3}$D time axis. Significantly, the various alkane, alkene, and alkyne bands are generally recognizable down the diagonal of the plot as demonstrated in Figure 3.2B, but several of the spiked non-native compounds eluting in this region are not well resolved in this 2D view. The mass channels *m/z* 55 and 74 were added to the AIC in Figure 3.4C to better target two of the non-native spiked compounds detailed in Table 3.2: 1,6-dichlorohexane (DCH) and cyclohexyl isothiocyanate (CHI), which have selective *m/z* of 74 and 55, respectively. Neither are well resolved in Figure 3.4C, but are very well separated by the $^{3}$D separation dimension as shown in

Figure 3.4D. Figure 3.4D depicts the $^{1}$D separation versus the $^{3}$D separation for the same portion of the $^{1}$D separation between 22 and 28 min at *m/z* 55 and 74, but in this case summing the signal along the $^{2}$D time axis. Significantly, utilization of the two selective *m/z* coupled with the chromatographic selectivity of the $^{3}$D separation enables the clear separation and detection of these two compounds. For complex mixture analysis such as the spike diesel, GC$^3$ – TOFMS provides the opportunity for novel classification and molecular fingerprinting strategies. The added chemical selectivity presents the opportunity to view different planes within the 4D array that represent different compound classes. As an example, the views in Figure 3.4C & D respectively depict fingerprints of major compound classes and electron rich chemicals in diesel.

Using the GC$^3$ experimental parameters implemented coupled with peak width measurements summarized in Table 3.3, key figures-of-merit were calculated. For the ten representative analytes from the two samples studied, the experimentally achieved total peak capacity $n_{c,3D}$ ranged from 5,000 to 9,600 for GC$^3$ – TOFMS ($\bar{x}$ = 7,000, $s$ = 1,700), which is on par with, if not exceeding, previous GC$^3$ – FID reports [22,23]. This is very competitive with state-of-the-art GC × GC, which provides $n_{c,2D}$ ranging from approximately 4,000 to 7,000 [9–11,15]. The benefits of GC$^3$ – TOFMS stem from the additional chemical selectivity along the $^{3}$D chromatographic dimension (relative to GC × GC – TOFMS) and addition of mass spectral selectivity (relative to GC$^3$ – FID), while maintaining a similar total peak capacity. Particularly, GC$^3$ – TOFMS achieves a combined $^{2}$D × $^{3}$D peak capacity ranging from 10 to 26 (mean = 17.6, 2 = 5.0) for the representative analytes described in Table 1 which is similar to what is achieved by $^{2}$D alone in a GC × GC operating at an equivalent $P_M$. When one considers the benefits of the added chemical selectivity afforded by $^{3}$D with a third stationary phase, the usefulness of GC$^3$ is apparent. Looking into the data in Table 3.3, and upon reflection of the theoretical peak capacity

calculations presented earlier, one can see evidence for where future improvements should be made to increase $n_{c,3D}$. The experimental average of ${}^{1}M_R$ was 2.6 (s=0.5), while the average for ${}^{2}M_R$ was 3.6 (s=0.9). These two $M_R$ are higher than a more desirable $M_R$ of 2. Since the $M_R$ were measured through reconstruction of the modulated peaks, modulator induced band broadening has already been accounted for [34]. Therefore application of eq 3.5 appropriately estimates peak capacity, albeit without the statistical correction [7,35,36]. In terms of peak capacities, the averages were 406, 3.6 and 4.9 for $^{1}$D, $^{2}$D, and $^{3}$D respectively. In particular, additional effort should be placed in reducing the peak widths on the $^{2}$D separation to increase the peak capacity on $^{2}$D. Experimentally obtaining a $n_{c3D}$ in the range of 10,000 to 15,000 certainly is very achievable.

## 3.5 CONCLUSIONS

The GC$^3$ – TOFMS instrument may arguably achieve the maximum selectivity available for a gas chromatographic instrument. Herein, we presented data which support this assertion. The coupling of TOFMS with GC$^3$ brings all the benefits inherent with mass spectrometry, namely mass spectral peak identification and added chemical selectivity. The high-temperature diaphragm valve completes the ensemble and brings the GC$^3$ – TOFMS instrument into the same temperature regime as other GC-based instruments. The new design achieves a $n_{c,3D}$ approaching 10,000 for select compounds and on average maintains a $n_{c,3D}$ of 7,000. Future studies will be aimed at additional instrumental improvements and leveraging the 4D data structure to tackle challenging chemical analysis problems using chemometric analysis tools.

## 3.6 REFERENCES

[1] J.C. Giddings, Unified Separation Science, John Wiley & Sons, Inc., New York, NY, 1991.

[2] Z. Liu, J.B. Phillips, Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface, J. Chromatogr. Sci. 29 (1991) 227–231. doi:10.1093/chromsci/29.6.227.
[3] Mondello, L, Comprehensive Chromatography in Combination with Mass Spectrometry, John Wiley & Sons, Inc., Hoboken, NJ, 2011.
[4] J.V. Seeley, S.K. Seeley, Multidimensional gas chromatography: fundamental advances and new applications, Anal. Chem. 85 (2013) 557–578. doi:10.1021/ac303195u.
[5] K.M. Pierce, B. Kehimkar, L.C. Marney, J.C. Hoggard, R.E. Synovec, Review of chemometric analysis techniques for comprehensive two dimensional separations data, J. Chromatogr. A. 1255 (2012) 3–11. doi:10.1016/j.chroma.2012.05.050.
[6] L.M. Blumberg, F. David, M.S. Klee, P. Sandra, Comparison of one-dimensional and comprehensive two-dimensional separations by gas chromatography, J. Chromatogr. A. 1188 (2008) 2–16. doi:10.1016/j.chroma.2008.02.044.
[7] J.M. Davis, D.R. Stoll, P.W. Carr, Effect of First-Dimension Undersampling on Effective Peak Capacity in Comprehensive Two-Dimensional Separations, Anal. Chem. 80 (2008) 461–473. doi:10.1021/ac071504j.
[8] W. Khummueng, J. Harynuk, P.J. Marriott, Modulation Ratio in Comprehensive Two-dimensional Gas Chromatography, Anal. Chem. 78 (2006) 4578–4587. doi:10.1021/ac052270b.
[9] B.D. Fitz, R.B. Wilson, B.A. Parsons, J.C. Hoggard, R.E. Synovec, Fast, high peak capacity separations in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A. 1266 (2012) 116–123. doi:10.1016/j.chroma.2012.09.096.
[10] D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity deviation ratio: A new metric for chemometric analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data, Anal. Chim. Acta. 871 (2015) 66–76. doi:10.1016/j.aca.2015.02.040.
[11] M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, J. Chromatogr. A. 1383 (2015) 151–159. doi:10.1016/j.chroma.2015.01.031.
[12] J.L. Anderson, D.W. Armstrong, Immobilized Ionic Liquids as High-Selectivity/High-Temperature/High-Stability Gas Chromatography Stationary Phases, Anal. Chem. 77 (2005) 6453–6462. doi:10.1021/ac051006f.
[13] T. Payagala, Y. Zhang, E. Wanigasekara, K. Huang, Z.S. Breitbach, P.S. Sharma, L.M. Sidisky, D.W. Armstrong, Trigonal Tricationic Ionic Liquids: A Generation of Gas Chromatographic Stationary Phases, Anal. Chem. 81 (2009) 160–173. doi:10.1021/ac8016949.
[14] T.D. Ho, C. Zhang, L.W. Hantao, J.L. Anderson, Ionic Liquids in Analytical Chemistry: Fundamentals, Advances, and Perspectives, Anal. Chem. 86 (2014) 262–285. doi:10.1021/ac4035554.
[15] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells, Anal. Chem. 78 (2006) 2700–2709. doi:10.1021/ac052106o.

[16] R.B. Wilson, W.C. Siegler, J.C. Hoggard, B.D. Fitz, J.S. Nadeau, R.E. Synovec, Achieving high peak capacity production for gas chromatography and comprehensive two-dimensional gas chromatography by minimizing off-column peak broadening, J. Chromatogr. A. 1218 (2011) 3130–3139. doi:10.1016/j.chroma.2010.12.108.
[17] R.B. Wilson, J.C. Hoggard, R.E. Synovec, Fast, High Peak Capacity Separations in Gas Chromatography–Time-of-Flight Mass Spectrometry, Anal. Chem. 84 (2012) 4167–4173. doi:10.1021/ac300481k.
[18] H. Snijders, H.-G. Janssen, C. Cramers, Optimization of temperature-programmed gas chromatographic separations I. Prediction of retention times and peak widths from retention indices, J. Chromatogr. A. 718 (1995) 339–355. doi:10.1016/0021-9673(95)00692-3.
[19] H. Snijders, H.-G. Janssen, C. Cramers, Optimization of temperature-programmed gas chromatographic separations II. Off-line Simplex optimization and column selection, J. Chromatogr. A. 756 (1996) 175–183. doi:10.1016/S0021-9673(96)00626-7.
[20] A.W. Moore Jr, J.W. Jorgenson, Comprehensive three-dimensional separation of peptides using size exclusion chromatography/reversed phase liquid chromatography/optically gated capillary zone electrophoresis, Anal. Chem. 67 (1995) 3456–3463.
[21] R. Edam, J. Blomberg, H.-G. Janssen, P.J. Schoenmakers, Comprehensive multi-dimensional chromatographic studies on the separation of saturated hydrocarbon ring structures in petrochemical samples, J. Chromatogr. A. 1086 (2005) 12–20. doi:10.1016/j.chroma.2005.02.048.
[22] N.E. Watson, W.C. Siegler, J.C. Hoggard, R.E. Synovec, Comprehensive Three-Dimensional Gas Chromatography with Parallel Factor Analysis, Anal. Chem. 79 (2007) 8270–8280. doi:10.1021/ac070829x.
[23] W.C. Siegler, J.A. Crank, D.W. Armstrong, R.E. Synovec, Increasing selectivity in comprehensive three-dimensional gas chromatography via an ionic liquid stationary phase column in one dimension, J. Chromatogr. A. 1217 (2010) 3144–3149. doi:10.1016/j.chroma.2010.02.082.
[24] P.A. Bueno Jr., J.V. Seeley, Flow-switching device for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1027 (2004) 3–10. doi:10.1016/j.chroma.2003.10.033.
[25] B. Mitrevski, P.J. Marriott, Novel Hybrid Comprehensive 2D – Multidimensional Gas Chromatography for Precise, High-Resolution Characterization of Multicomponent Samples, Anal. Chem. 84 (2012) 4837–4843. doi:10.1021/ac300429y.
[26] A.E. Sinha, K.J. Johnson, B.J. Prazen, S.V. Lucas, C.G. Fraga, R.E. Synovec, Comprehensive two-dimensional gas chromatography of volatile and semi-volatile components using a diaphragm valve-based instrument, J. Chromatogr. A. 983 (2003) 195–204. doi:10.1016/S0021-9673(02)01651-5.
[27] P.-H. Stefanuto, K.A. Perrault, S. Stadler, R. Pesesse, H.N. LeBlanc, S.L. Forbes, J.-F. Focant, GC × GC–TOFMS and supervised multivariate approaches to study human cadaveric decomposition olfactive signatures, Anal. Bioanal. Chem. 407 (2015) 4767–4778. doi:10.1007/s00216-015-8683-5.
[28] W. Welthagen, R.A. Shellie, J. Spranger, M. Ristow, R. Zimmermann, O. Fiehn, Comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry (GC × GC-TOF) for high resolution metabolomics: biomarker discovery on spleen tissue extracts of obese NZO compared to lean C57BL/6 mice, Metabolomics. 1 (2005) 65–73. doi:10.1007/s11306-005-1108-2.

[29] C.E. Freye, L. Mu, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1424 (2015) 127–133. doi:10.1016/j.chroma.2015.10.098.
[30] C.E. Freye, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta. 161 (2016) 675–680. doi:10.1016/j.talanta.2016.09.002.
[31] W.C. Siegler, B.D. Fitz, J.C. Hoggard, R.E. Synovec, Experimental Study of the Quantitative Precision for Valve-Based Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 83 (2011) 5190–5196. doi:10.1021/ac200302b.
[32] J.V. Seeley, Theoretical study of incomplete sampling of the first dimension in comprehensive two-dimensional chromatography, J. Chromatogr. A. 962 (2002) 21–27. doi:10.1016/S0021-9673(02)00461-2.
[33] J.V. Seeley, N.J. Micyus, S.V. Bandurski, S.K. Seeley, J.D. McCurry, Microfluidic Deans Switch for Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 79 (2007) 1840–1847. doi:10.1021/ac061881g.
[34] D.K. Pinkerton, B.A. Parsons, R.E. Synovec, Method to determine the true modulation ratio for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1476 (2016) 114–123. doi:10.1016/j.chroma.2016.11.015.
[35] J.M. Davis, Statistical theory of spot overlap for n-dimensional separations, Anal. Chem. 65 (1993) 2014–2023.
[36] J.M. Davis, D.R. Stoll, P.W. Carr, Dependence of Effective Peak Capacity in Comprehensive Two-Dimensional Separations on the Distribution of Peak Capacity between the Two Dimensions, Anal. Chem. 80 (2008) 8122–8134. doi:10.1021/ac800933z.
[37] K.M. Pierce, J.C. Hoggard, Chromatographic data analysis. Part 3.3.4: handling hyphenated data in chromatography, Anal. Methods. 6 (2014) 645–653. doi:10.1039/c3ay40965a.

# Chapter 4. Targeted Analyte Deconvolution and Identification by Four-Way Parallel Factor Analysis Using Three-Dimensional Gas Chromatography with Mass Spectrometry Data[4]

## 4.1 INTRODUCTION

The body of knowledge surrounding comprehensive two-dimensional (2D) gas chromatography (GC × GC) is seemingly expanding at an ever-growing rate [1]. Moving from a theoretical possibility pioneered during the latter half of the twentieth century [2] to a novelty in the early 1990s [3], separation of complex mixtures is now almost unimaginable without two serially coupled GC columns [1]. Creative options for column selection and coupling are still being developed. In addition, GC × GC has been joined to most GC detectors, be they single- or multi-channel. Indeed, the combination of GC × GC with time-of-flight mass spectrometry (TOFMS) has proven particularly interesting (GC × GC-TOFMS) [4–7].

When GC × GC-TOFMS is implemented, the resultant three-dimensional (3D) data enable application of various chemometric deconvolution techniques providing mathematical resolution [8,9]. Among these methods, the Parallel Factor Analysis (PARAFAC) algorithm has proven particularly useful [10–16]. PARAFAC models the data array as the linear combination sum of signals for each eluting analyte described by the outer product of chromatographic and mass spectrometric vectors in each of the acquired dimensions for an analyst determined number of components. PARAFAC requires, at a minimum, a 3D array which implicates GC × GC-TOFMS as one of the few instruments which create data of this structure naturally.

[4] This Chapter has been submitted for publication as N.E. Watson, R.E. Synovec, Targeted Analyte Deconvolution and Identification by Four-Way Parallel Factor Analysis Using Three-Dimensional Gas Chromatography with Mass Spectrometry Data, Anal. Chim. Acta. (2016).

Previously, comprehensive 3D gas chromatography (GC × GC × GC or $GC^3$) was described as an alternate gas chromatographic configuration which enables implementation of PARAFAC on the native data [17,18]. $GC^3$ with PARAFAC deconvolution was shown to provide a benefit to both the detection limit and provide the ability to resolve convoluted elution signals. Recently the $GC^3$ instrumental design was improved to include a TOFMS as the detector creating a 3D gas chromatograph with mass spectrometric detection ($GC^3$-TOFMS) [19]. The principal benefits being both the added mass spectrometric selectively and the ability to identify analyte peaks based on the acquired mass spectrum. This advanced instrumental platform natively creates four-dimensional data (4D) which permits new options for visualizing and chemometrically interrogating the data array provided by complex samples.

Four-way PARAFAC, in itself, is not novel [9,20]. Four-way data can be simply created by stacking replicates of 3-way data with varying concentration profiles. The results produce a reconstruction of the three instrumental dimensions and the fourth dimension contains the concentration information in each of the replicates. This procedure of replicate stacking is even more common in 2D methods (e.g. one chromatographic dimension with multichannel detection) where the replicate stacking enables access to PARAFAC which has a minimum dimensionality of three. Instruments, like the $GC^3$-TOFMS, which produce native four-way data are significantly less typical, but the development of such has the potential to address emerging challenges in chemical analysis.

Herein we present a targeted method for analysis of four-way data, specifically data generated by $GC^3$-TOFMS. Embracing the native dimensionality of the data produced by this instrument, we explore a technique which overcomes one of the inherent challenges of applying chemometrics to more traditional GC × GC-TOFMS data: assigning a target window for

application of PARAFAC [14–16]. For instance, if one were inclined to apply PARAFAC to a whole dataset, the computational intensity of the technique would quickly overcome the available computing capacity requiring that subordinate windows be selected within the dataset. In GC × GC-TOFMS this window determination step is one of the more time consuming and subjective steps of the analysis. $GC^3$-TOFMS enables a new paradigm by which each modulation of the first "primary" chromatographic dimension ($^1$D) produces a complete GC × GC-TOFMS chromatogram which is suitable, as a whole, for PARAFAC analysis. More simply, we suggest this method could be envisioned as a primary GC separation coupled with GC × GC-TOFMS detection. By implementing a non-polar stationary phase for the $^1$D separation, a complex sample is subjected to an initial separation roughly equivalent to a fractional distillation whereby each fraction is analyzed by GC × GC-TOFMS. The "fractions" are the modulations from the $^1$D separation onto the $^2$D separation, and so on. This paradigm is demonstrated with PARAFAC, whereby data portions amenable to third order chemometrics are produced in a natural fashion and should be applicable to other chemometric methods [21].

In this study, a stack of seven consecutive GC × GC-TOFMS $^1$D modulation "slabs" produce a window across $^1$D without any need to window the second GC separation ($^2$D), third GC separation ($^3$D) and mass spectrometric dimensions. The GC × GC separations in each slab, i.e., $^2$D by $^3$D separations, are designed to provide a two-dimensional chromatographic peak capacity $n_{c,2D}$ of ~ 15 to 20 that is ideally suited to both provide ample chemical selectivity concurrent with keeping the rank suitably low for successful implementation of PARAFAC. The temporal width of the seven $^1$D modulation window is defined to fully encompass the elution event for a target analyte, allowing for some ambiguity in $^1$D retention time selection. Standards or retention indices [22] provide an excellent method for approximate $^1$D target analyte window

location assignment prior to PARAFAC analysis. Herein, two samples are submitted to study this approach for targeted analyte discovery and deconvolution: a standard mixture of 115 components and a diesel fuel spiked with several non-native compounds [19]. Analytical success is determined by both visual verification of the chromatographic loadings and mass spectral matching to a library mass spectrum [14,23].

## 4.2 EXPERIMENTAL

### 4.2.1 *Data Collection*

The $GC^3$-TOFMS was operated as previously described [19] and briefly recalled here. A Pegasus 4D GC × GC-TOFMS (LECO Corporation, St. Joseph, MI) with an integrated Agilent 6890N Gas Chromatograph (Agilent Technologies, Santa Clara, CA, USA) was modified to produce a $GC^3$-TOFMS instrument as shown in Figure 3.1. One high-speed, six port diaphragm valve (Valco Instruments Company Inc, Houston, TX, USA), upgraded by the manufacturer to operate at a maximum temperature of 325 °C, and fitted with a 5 μL sampling loop was installed in the GC oven. The high-temperature valve was utilized as the modulator between the $^1$D and $^2$D separations, and the stock thermal modulator was implemented between the $^2$D and $^3$D separations. The $^3$D column was contained within the secondary oven in the commercial instrumental platform to provide a small temperature offset.

Three columns with different stationary phases were installed in column 1, column 2 and column 3 positions as depicted in Figure 1 for the $^1$D, $^2$D, and $^3$D separations, respectively. A 30 m, 250 μm inner diameter (id), 0.50 μm film thickness (5% phenyl)-methyl polysiloxane stationary phase column (Rtx-5; Restek, Bellefonte, PA, USA) was installed as column 1. A 3.5 m, 180 μm id, 0.18 μm film thickness polyethylene glycol stationary phase column (Rtx-Wax;

Restek, Bellefonte, PA, USA) was installed as column 2. Finally, a 1 m, 100 μm id, 0.1 μm film thickness trifluoropropyl-methyl polysiloxane stationary phase column (Rtx-200; Restek, Bellefonte, PA, USA) was installed as column 3.

Liquid injections of 1 μL of both a test mixture of 115 compounds [19] and diesel spiked with non-native compounds [19] were made with an Agilent 7683 auto-injector. The inlet was operated at 250 °C in split mode with a split flow of 25 mL/min. Effluent from $^{1}$D was injected onto $^{2}$D with the high-temperature diaphragm valve set to actuate for 400 ms at a modulation period $^{1}P_{m}$ = 3 s. Effluent from $^{2}$D was transferred to $^{3}$D via the thermal modulator with a hot pulse time of 120 ms at a $^{2}P_{m}$ = 250 ms. The $^{1}$D column was operated at a constant volumetric flow of 0.5 mL/min. The $^{2}$D and $^{3}$D columns were controlled via the auxiliary pressure controller under a ramped pressure program held at 25 psi for the first 3 min and ramped to 35 psi at a rate of 0.211 psi/min and held at the final pressure for 1 min. This program resulted in an approximate volumetric flow on $^{2}$D and $^{3}$D columns of 1 mL/min. The $^{1}$D column was operated at less than the optimum flow rate in order to slightly widen the analyte peaks in the $^{1}$D dimension to ensure sufficient peak width and sampling rate at the valve-based modulator to properly sample the $^{1}$D peaks. Detection was accomplished with the TOFMS. Effluent was passed through a 0.33 m, 280 °C transfer line into the TOFMS where it was analyzed at 200 Hz between 33 Da to 334 Da. The GC oven was operated with a temperature program starting at 60 °C and held at that temperature for 3 min. The oven was then ramped to 250 °C at a rate of 4 °C/min. The final temperature was held for 1 min. The secondary oven was held at a constant 10 °C offset above the primary oven temperature. The 115-component test mixture and spiked diesel were both analyzed in triplicate.

### 4.2.2 *Data Analysis*

Analysis was conducted on a personal computer employing an Intel i5 4430 processor with 32 GB of RAM running Windows 10. All data were collected using ChromaTOF 3.32 and transferred to MATLAB 2016b (The Mathworks, Natick, MA, USA) using in-house software (peg2mat3p8) [24]. These data were exported from ChromaTOF as a 2D array of *m/z* versus time. Chromatograms were reorganized to four-way arrays of *m/z* by $^{3}$D by $^{2}$D by $^{1}$D. Zero filling at the end of the chromatograms was conducted when necessary to ensure a rectangular array. Retention times of the target analyte compounds were inferred from retention indices and verified by the ChromaTOF peak finder when possible. Utilizing these target $^{1}$D retention times, seven modulation $^{1}$D windows centered around each target analyte $^{1}$D retention time were submitted to PARAFAC. The PARAFAC algorithm included as part of PLS Toolbox 8.2.1 (Eigenvector Research Inc., Wenatchee, WA) was implemented. All target windows were submitted as four-way arrays and subjected to the non-negativity constraint in all dimensions. Models containing up to 12 factors were constructed in all cases and models up to 18 factors were built for three of the more challenging cases (1-nonanol, 2-decanone, and ethyl salicylate).

Appropriate factor selection was determined based on the splitting criteria [14–16]. Match values versus the target spectra were calculated for all PARAFAC loadings and the appropriate loading from the model prior to splitting was selected. All loadings were then visually verified and retention times across all three dimensions were determined. Target spectra were previously collected utilizing the TOFMS employed for this study. NIST target spectra were utilized when an in-house standard was not available. Generally, match values to the NIST library are lower due to instrument specific detector and ion source variations which result in varied mass spectral peak abundances [23,25,26].

## 4.3 Results and Discussion

Management and visualization of data beyond three dimensions is inherently challenging and the four-way data produced by the GC$^3$-TOFMS is no exception in this regard. Figure 4.1 attempts to overcome this challenge through various simplifications which build upon each other to convey the full composition of the data and sample at hand. Here the separation of the 115-component mixture is simplified to an elongated 3D rectangular cuboid where each dimension represents one of the chromatographic separations. The length being $^1$D, depth being $^2$D and the height being $^3$D. These dimensions are broken down into a virtual grid where the subordinate section's sizes are dictated by the two modulation periods $^1P_m$, and $^2P_m$ and the detection frequency, respectively. A complete mass spectrum resides at each grid location creating the full structure of the four-way array.

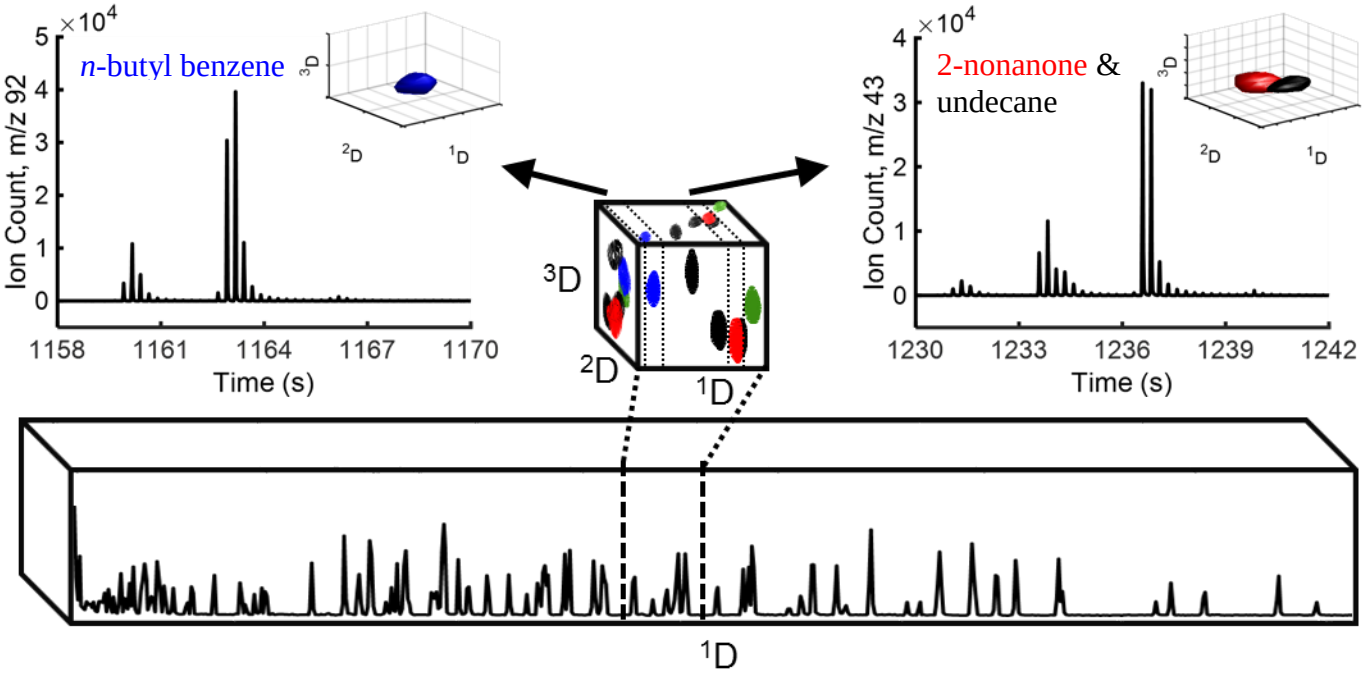


Figure 4.1. Schematic detail of GC$^3$-TOFMS data structure.

The method we present here both embraces the inherent complexity of the native four-way data generated by the GC$^3$-TOFMS while allowing the analyst to simplify the initial data

structure back down to the well understood 1D chromatogram displayed on the front face of the rectangular cube in the lower portion of Figure 4.1. Recalling the detailed analysis previously published [19], the reader is invited to consider the cubic section depicted at the center. Six analytes elute in this region: *n*-butyl benzene, 1-octanol, 1-decene, 2-nonanone, undecane, and adamantane. The faces display the resultant 2-way contour plots the observer sees when viewing the respective face directly from above. The colors are varied across the four selective *m/z* used to exhibit the eluting chemicals: *m/z* 92 (blue), *m/z* 43 (black), *m/z* 58 (red) and *m/z* 136 (green).

For added detail, two regions are excised and the raw data vectors are displayed as they arrived at the detector. The upper left shows the elution event for *n*-butyl benzene at *m/z* 92 and the upper right shows the coelution of 2-nonanone and undecane at *m/z* 43. The insets show the results after refolding the depicted vector into three dimensions. The region excised in the upper right encompassing 2-nonanone and undecane present an opportunity to explore the analytical potential of four-way data. Though co-eluting in all three chromatographic dimensions, the TOFMS detection provides the final portion of selectivity necessary to fully resolve these two

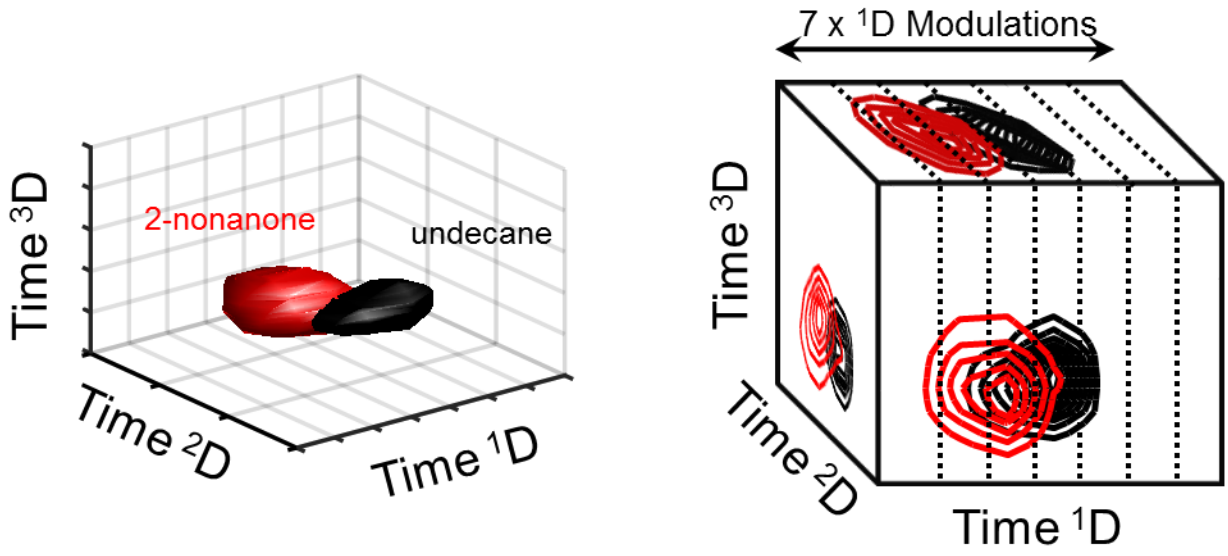


Figure 4.2. Schematic detail of the method for targeted analyte finding and deconvolution using GC3-TOFMS and four-way PARAFAC.

compounds without chemometric deconvolution. However, this simple case of near total co-elution provides an initial test bed which affords ready visualization of the basic elements of the targeted analyte discovery and deconvolution method which is presented next.

Figure 4.2 considers this overlap region of the $GC^3$-TOFMS chromatogram for the co-elution of 2-nonanone and undecane in greater detail. 2-nonanone is shown in red using the totally selective *m/z* of 58 and undecane using *m/z* 56. The isosurface plot on the left of Figure 4.2 is simplified on the right to a cube where each surface again shows what the analyst observes when viewing the cube directly from above the face as was described for Figure 4.1. Taking undecane as the target analyte, this four-way section of data is interrogated chemometrically. As with any target analysis the procedure is initialized with as much data as is known about the target. The NIST Database includes a plethora of information about common compounds including reference library mass spectra and experimental retention indices. In many cases the analyst will also have run a standard thus rendering a mass spectrum specific to the instrument used along with an accurate estimation of retention times for the target, in particular for the $^1D$ separation axis. While incorporation of all available data simplifies the search, only a target mass spectrum and retention index are necessary to employ this technique. The retention index provides a fairly accurate initial estimate for the $^1D$ retention time in most cases and the target $^1D$ section can be adjusted in cases where large variation in the NIST database log of retention indices fails to yield identification of the target on the first attempt.

Visualizing the 3D rectangular array as a stack of $^2D$ by $^3D$ slabs leads to a synonymous way to visualize $GC^3$-TOFMS as GC with GC × GC-TOFMS detection. This paradigm simplifies the transfer to a PARAFAC method whereby the user submits a subarray of GC × GC-TOFMS chromatograms for analysis dictated by approximate $^1D$ retention time of the target.

Assuming the modulation ratio $M_R$~2 to 3 for the $^1$D to $^2$D transfer, a seven-slab window will generally cover the $^1$D elution event for most analytes [27]. Utilizing the *a priori* knowledge of undecane's instrument specific mass spectrum and $^1$D retention time, these data are submitted to PARAFAC. The target window includes the entirety of the seven $^1$D slabs roughly centered on the retention time of undecane. The mass spectrum of undecane is also input as the initialization vector for the mass spectral axis of the four-way array.

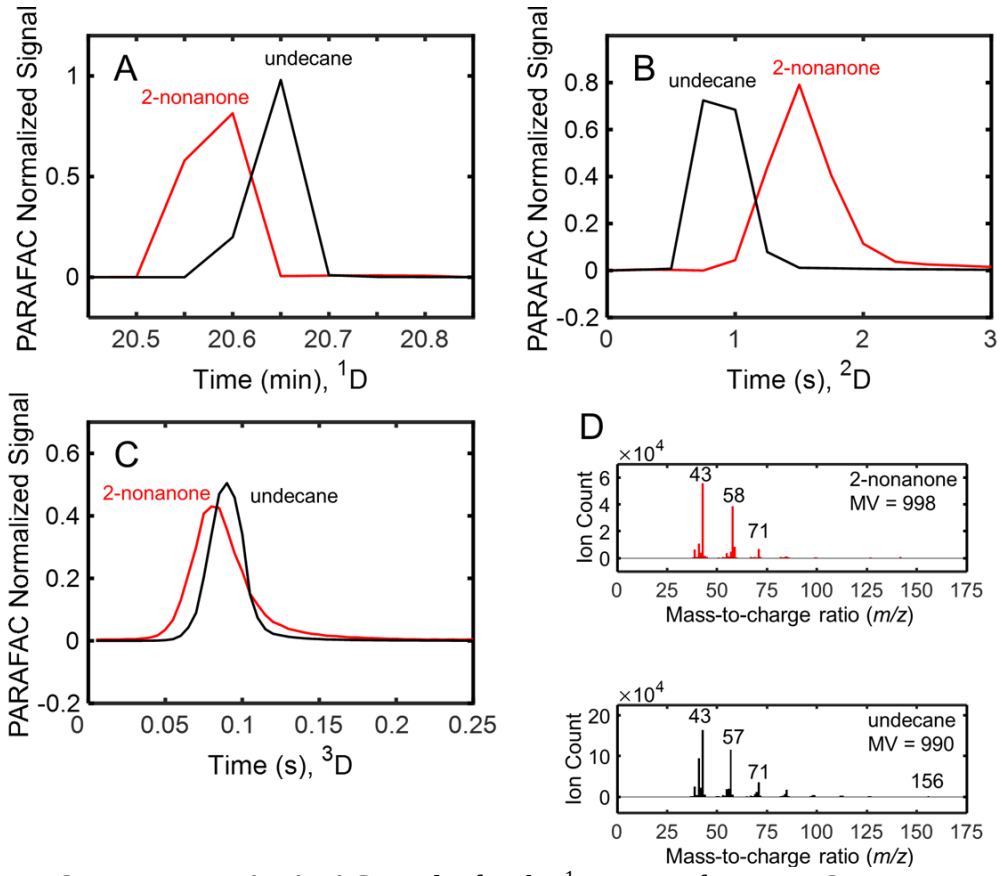


Figure 4.3. Four-way PARAFAC results for the $^1$D region from Fig. 3 containing 2-nonanone and undecane.

The results of this chemometric analysis are depicted graphically in Figure 4.3. The target analyte undecane is indicated in black and the interfering analyte, 2-nonanone is indicated in red. Figure 4.3A-C depict the $^1$D, $^2$D and $^3$D dimensional loadings, respectively. Figure 4.3D depicts the mass spectral loadings for 2-nonanone and undecane. This straightforward example rendered

PARAFAC results which correlate almost exactly with the data determined by more traditional, non-chemometric methods previously reported [19]. Particularly, the exceptionally high mass spectral match values obtained for 2-nonanone and undecance of 998 and 990, respectively, suggest the effectiveness of this method. Table 4.1 reports the PARAFAC results for the targeted analysis of the entire cube depicted at the center of Figure 4.1 with similarly exceptional performance. MmFf indicates a particular factor from a PARAFAC model as the number of factors in the model m, and the index of the factor in that model f (an integer from 1 to m). [15]. However, a truly challenging demonstration and validation of the method with a complex sample matrix is warranted as is considered next.

Table 4.1. Summary of four-way PARAFAC results for six target analytes portrayed in Figures 4.1-4.3.

| **Analyte** | ***n*-butyl benzene** | **1-octanol** | **1-decene** | **2-nonanone** | **undecane** | **adamantane** |
|---|---|---|---|---|---|---|
| **MmFf** | M1F1 | M1F1 | M2F1 | M6F3 | M3F1 | M2F1 |
| **Match** | 996 | 995 | 906* | 998 | 990 | 993 |
| **Time 1 (min)** | 19.40 | 19.95 | 20.35 | 20.60 | 20.65 | 20.80 |
| **Time 2 (s)** | 2.25 | 1.50 | 1.75 | 1.50 | 0.75 | 2.25 |
| **Time 3 (ms)** | 170 | 190 | 100 | 80 | 90 | 140 |

***Match to NIST library**

For the purpose providing a challenging method evaluation, analysis of diesel spiked with several test analytes was undertaken. Numerous non-native spiked and native non-spiked analytes were considered. Cyclohexyl benzene (CHB) provides an appropriate testbed to truly display the power of the method suggest herein. Cyclohexyl benzene is a compound that occurs naturally in most diesel fuels, including the diesel shown in Figure 4.4A. This analytical ion chromatogram provides a reconstruction of the $^{1}$D response for diesel at *m/z* 41, 43, 53, 55, 75, and 91. The typical diesel hump is observed and the experienced chromatographer recognizes that the hump is indicative of a multitude of co-eluting hydrocarbons present when only the $^{1}$D

chromatogram is viewed [28]. Indeed, Figure 4.4B & C provide further evidence of the complexity of a diesel separation, even when subjected to three-dimensional separation. Figure 4.4B depicts the $^{1}$D by $^{2}$D contour plot at the *m/z* listed above for a 6 min time window along $^{1}$D. Likewise, Figure 4.4C depicts the $^{1}$D by $^{3}$D contour plot at the same *m/z* for the same 6 min time window. The circle in both chromatograms indicates the location of the target, cyclohexyl benzene. In both 2D separation visualizations, the target elutes near the center of the alkane band.

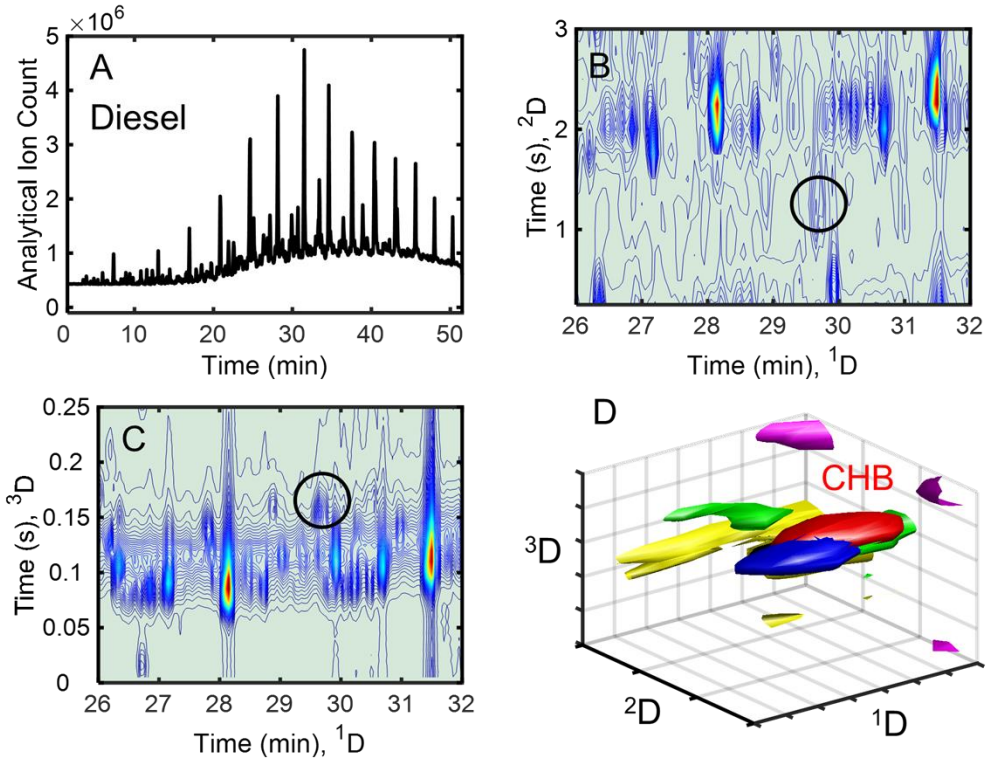


Figure 4.4. Method transfer from a standard mix (e.g. Figures 4.1-4.3) to a sample of diesel containing spikes of several non-native compounds.

Targeted analyte discovery and deconvolution analysis is initialized with all the data available. In this case, cyclohexyl benzene was run as a standard to determine both an instrument specific mass spectrum along with retention times in all three dimensions. The method relies upon sufficient knowledge of the target analyte retention time on the $^{1}$D separation for the

purpose of roughly centering the target window, as demonstrated in Figure 4.2, hence the $^{2}$D and $^{3}$D analyte retention times are not utilized nor needed for successful PARARAC implementation. For cyclohexyl benzene, the GC$^{3}$-TOFMS data in the target window based upon its $^{1}$D retention time of 29.7 min is presented in Figure 4.4D, which is functionally a composite zoom of all three GC separation dimensions. Unlike the case of undecane in the simple mixture before (per Figure 4.1 and Figure 4.2), the GC$^{3}$-TOFMS data in Figure 4.4D do not provide sufficient chromatographic and mass spectral selectivity to render purely resolved peak profiles by

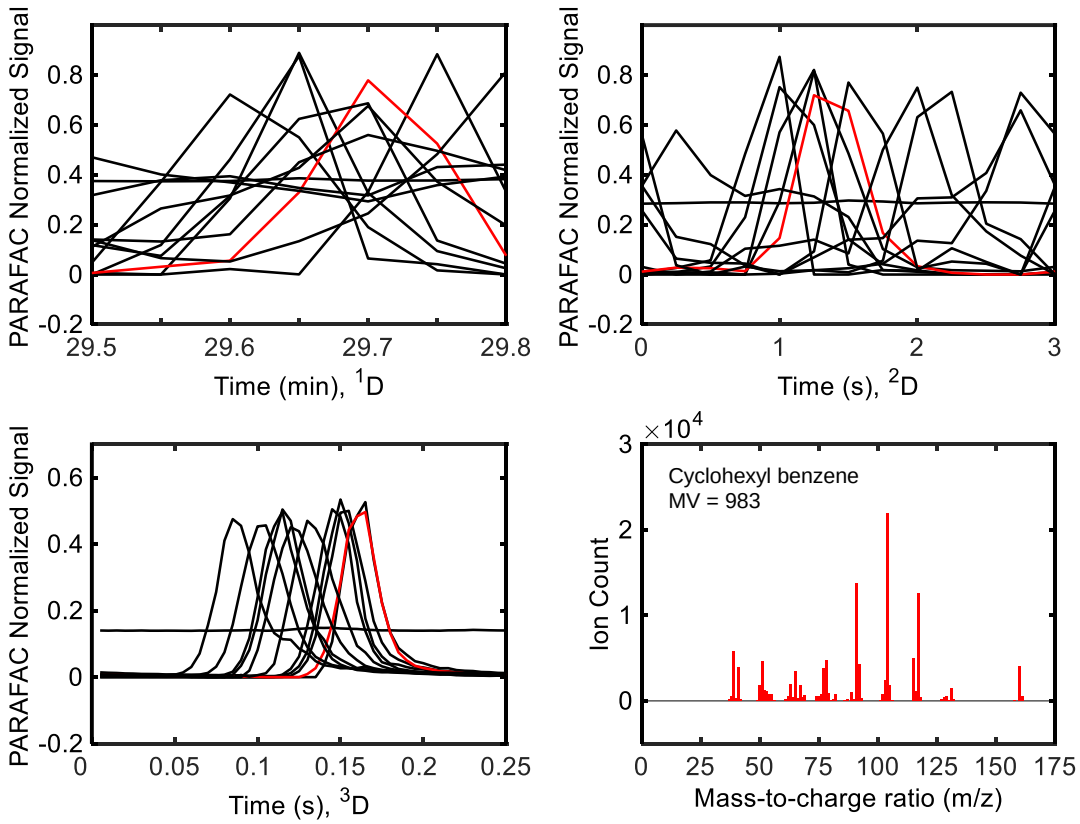


Figure 4.5. Four-way PARAFAC results on the region from Figure 4.4 containing cyclohexyl benzene (CHB).

traditional analytical methods alone. PARAFAC becomes not just beneficial, but necessary to resolve cyclohexyl benzene from the co-eluting peaks in the chemical background.

The results of the PARAFAC analysis, analagous to Figure 4.3, are depicted in Figure 4.5 for cyclohexyl benzene. Figure 4.5A-C depict the $^{1}$D, $^{2}$D and $^{3}$D dimensional loadings

respectively. Figure 4.5D depicts the mass spectral loading for cyclohexyl benzene. The red trace indicates the target and the remaining 11 traces indicate co-eluting signals. PARAFAC required 12 factors to resolve cyclohexyl benzene, and in some cases this caused an interfering compound to split between two factors. Near total co-elution is also present in $^{3}$D between cyclohexyl benzene and one of the interfering compounds. Only the remaining resolution in the other dimensions allowed for the successfully PARAFAC deconvolution of these two analytes.

A previously validated splitting criteria [14] was utilized to determine the appropriate factor model for the target analyte. When comparing the target PARAFAC loading to the standard mass spectrum of cyclohexyl benzene a match value of 983 was obtained. This is an

Table 4.2. Summary of four-way PARAFAC results for 18 non-native target analytes spiked in diesel and an additional 18 target analytes native to diesel.

| Analyte | NIST RI | Calculated RI | MmFf | Match | Time $^{1}$D (min) | Time $^{2}$D (s) | Time $^{3}$D (ms) |
|---|---|---|---|---|---|---|---|
| **pyridine*** | 745 | 725 | M7F5 | 930±1.0 | 8.20 | 1.75 | 200 |
| **1-chlorohexane** | 855 | 829 | M4F2 | 998±0.0 | 11.55 | 2.50 | 145 |
| **2,5-dimethylthiophene*** | 880 | 852 | M7F2 | 984±0.6 | 12.30 | 1.50 | 125 |
| **2-heptanol** | 905 | 883 | M4F3 | 996±0.6 | 13.30 | 2.75 | 165 |
| **methyl caproate** | 925 | 906 | M6F2 | 999±0.0 | 14.05 | 3.00 | 235 |
| **bromobenzene** | 945 | 934 | M2F2 | 999±0.6 | 14.95 | 2.75 | 165 |
| **3-octanone** | 985 | 984 | M9F4 | 998±0.6 | 16.60 | 0.75 | 60 |
| **limonene*** | 1035 | 1043 | M9F3 | 909±12.1 | 18.45 | 2.75 | 100 |
| **5-decyne** | 1030 | 1050 | M11F2 | 980±2.0 | 18.70 | 1.50 | 80 |
| **1,6-dichlorohexane** | 1070 | 1171 | M14F8 | 997±0.6 | 22.60 | 1.75 | 25 |
| **nonanol*** | 1170 | 1210 | M15F14 | 939±4.0 | 23.85 | 2.50 | 170 |
| **2-decanone*** | 1190 | 1230 | M18F17 | 943±11.0 | 24.55 | 2.25 | 55 |
| **cyclohexyl isothiocyanate*** | 1255 | 1298 | M12F9 | 908±2.1 | 26.68 | 2.25 | 15 |
| **butyrophenone*** | 1270 | 1317 | M12F10 | 946±9.8 | 27.35 | 2.25 | 50 |
| **ethyl salicylate** | 1275 | 1338 | M15F15 | 929±12.4 | 28.05 | 1.00 | 30 |
| **dodecanethiol*** | 1520 | 1588 | M11F10 | 912±4.9 | 36.10 | 2.25 | 135 |
| **diphenyl sulfide*** | 1620 | 1668 | M4F3 | 877±2.3 | 38.65 | 1.00 | 180 |
| **dibutylphthalate*** | 1970 | 1956 | M11F7 | 959±4.5 | 47.95 | 1.75 | 50 |
| **Toluene** | 780 | 739 | M4F2 | 995±0.0 | 8.70 | 2.25 | 100 |
| **Octane** | 800 | | M3F2 | 997±0.0 | 9.40 | 1.25 | 95 |
| **ethyl benzene** | 870 | 841 | M4F3 | 993±0.6 | 11.95 | 3.00 | 140 |
| **p-xylene** | 860 | 849 | M4F2 | 995±0.0 | 12.25 | 0.50 | 145 |
| **Nonane** | 900 | | M3F1 | 995±0.0 | 13.00 | 1.50 | 70 |
| **propylbenzene** | 950 | 951 | M5F3 | 995±0.6 | 15.50 | 3.00 | 145 |
| **3-ethyltoluene*** | 960 | 962 | M5F2 | 935±0.0 | 15.85 | 3.00 | 150 |
| **Decane** | 1000 | | M3F1 | 976±0.0 | 16.95 | 1.75 | 75 |
| **butylcylcohexane** | 1030 | 1044 | M7F2 | 970±6.8 | 18.50 | 3.00 | 100 |
| **Undecane** | 1100 | | M3F1 | 991±0.6 | 20.85 | 2.00 | 80 |
| **dodecane*** | 1200 | | M12F8 | 902±6.7 | 24.15 | 2.00 | 90 |
| **Tridecane** | 1300 | | M2F1 | 992±0.0 | 28.15 | 2.25 | 85 |
| **2,7-dimethyltetralin*** | 1340 | 1386 | M5F1 | 954±0.6 | 29.65 | 1.25 | 155 |
| **cyclohexylbenzene** | 1315 | 1389 | M12F6 | 974±10.1 | 29.70 | 1.25 | 165 |
| **1,3-dimethylnaphthalene*** | 1410 | 1487 | M9F9 | 899±4.6 | 32.80 | 2.50 | 235 |
| **Hexadecane** | 1600 | | M2F1 | 978±9.6 | 37.55 | 2.75 | 85 |
| **pristane** | 1710 | 1676 | M3F1 | 982±1.0 | 38.90 | 2.25 | 90 |
| **Heptadecane** | 1700 | | M2F1 | 966±0.0 | 40.35 | 3.00 | 90 |

***Match to NIST library**

exceptional result considering the complexity of the sample, and in particular in this region of the separation in which the rank of the data was exceptionally high. Generally, analysts look for match values of at least 800 when searching for compounds of interest, matches above 950 are the mass spectroscopic equivalent of a bullseye [14,23,26].

Further demonstration and validation of the method was performed. Table 4.2 reports results of cyclohexyl benzene and an additional 35 compounds evenly mixed between diesel natives and spiked non-natives. The results are equally compelling for these additional target analyte compounds and suggest the power of the method described. In some cases the match values are lower, but this generally is the result of utilizing the NIST database in lieu of an in-house standard. Mass spectrometers tend to have individualistic quirks resulting from electron ionization variation that preclude high match values when comparing to spectra obtained from

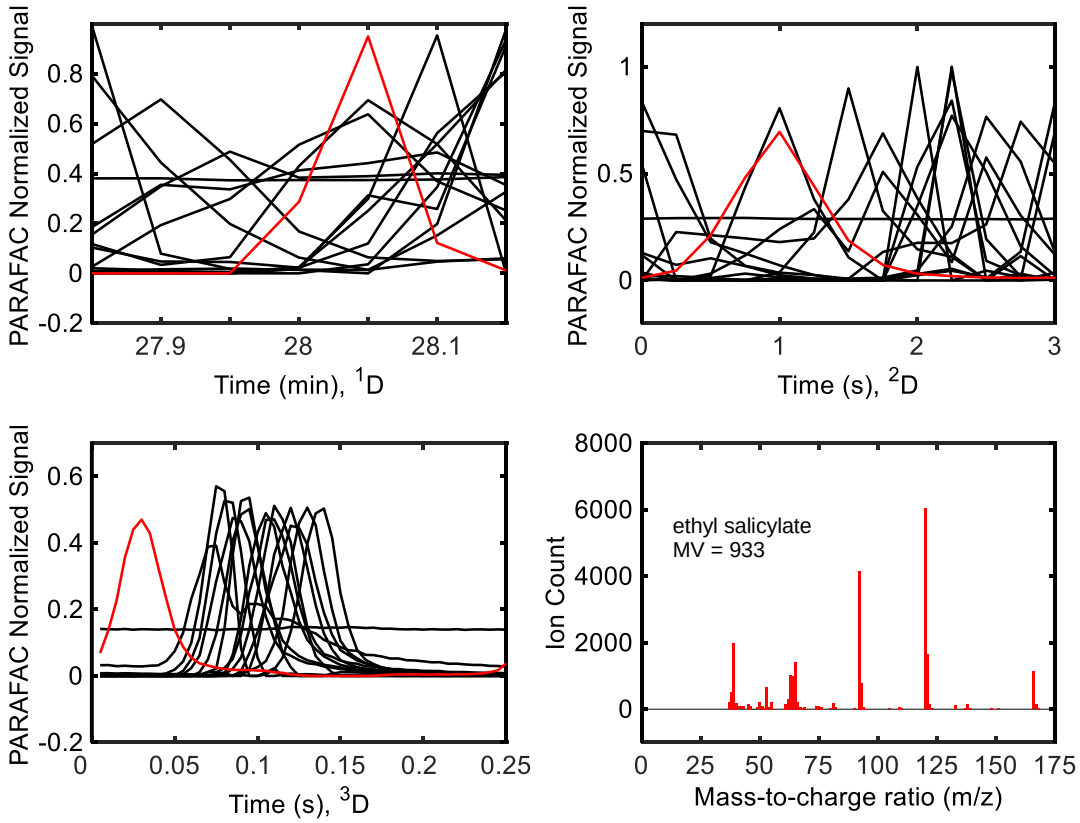


Figure 4.6. Four-way PARAFAC results for ethyl salicylate

other mass spectrometers. Even considering this reality, 34/36 analytes reported showed match values greater than 900 with or without in-house standard spectra. Additionally, these analyses

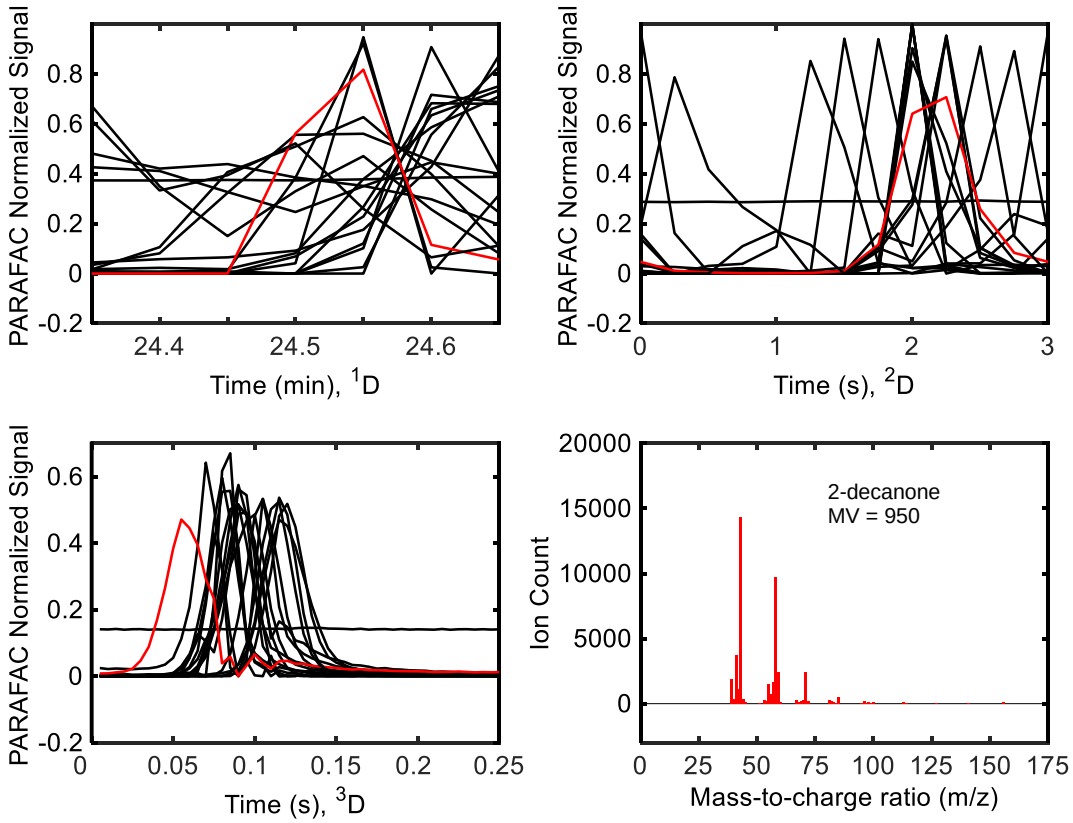


Figure 4.7. Four-way PARAFAC results for 2-decanone.

were conducted in triplicate with nearly exact correlation between injection replicates as quantified by standard deviations in Table 4.2. Figure 4.6 and Figure 4.7 depict the loadings analogous to Figure 4.5 for two of the more challenging target analyte cases: ethyl salicylate and 2-decanone.

Additionally, three-way PARAFAC analysis was also conducted on cyclohexyl benzene and 2,5-dimethyl thiophene; results are not shown for brevity. These studies considered the same seven-slab $^{1}$D windows on a single slab-by-slab basis (each slab being a $^{2}$D by $^{3}$D separation of GC × GC-TOFMS data). The results were congruent. A speed optimized temperature

programming rate of 4 °C/min was employed resulting in minimal to no retention time shifting in $^{2}$D and $^{3}$D. The four-way method presented here will be generally adequate and faster in all cases where $^{2}$D and $^{3}$D retention time shifting is sufficiently small [29]. In cases of extreme retention time shifting in $^{2}$D and $^{3}$D, using slab-by-slab $^{2}$D by $^{3}$D separations may be preferable, but was not necessary in the current study.

## 4.4 CONCLUSION

The true opportunities afforded by higher order (> 2) chromatographic methods, such as GC$^{3}$, are poorly understood at best, largely due to their rarity. A targeted analyte discovery and deconvolution method was presented which embraces the complexity of three-dimensional chromatographic separations. Here we decouple the enthalpy of vaporization component of gas chromatography from the other dimensions and utilize a GC$^{3}$-TOFMS instrument which effectively distills the sample and periodically analyzes the distillate by GC × GC-TOFMS. The results, which address challenges in the analysis of complex samples, are powerful. Both simple and complex sample types are presented resulting in successful deconvolution in each case as quantified by mass spectral match value. These data suggest that GC$^{3}$-TOFMS provides a platform to resolve components of complex samples both chemically and mathematically. Interestingly, this analytical approach appears to provide a technique for native PARAFAC analysis without the tedium of windowing sections of the GC × GC chromatograms provided by the $^{2}$D by $^{3}$D separations. In other words, the separation conditions and modulation parameters create data naturally amenable to PARAFAC. Additionally, the fast modulation from $^{2}$D to $^{3}$D produces trilinear GC × GC-TOFMS data which are immune to the negative effects of trilinearity deviation which may otherwise be deleterious to PARAFAC analysis [21,29]. PARAFAC2 [30], MCL-ALS [21], or other three-way approaches provide additional options in cases where the

trilinearity is sufficiently violated. Certainly, the results presented suggest the necessity for further study, application, and advancement toward non-targeted methods.

## 4.5 REFERENCES


[1] J.V. Seeley, S.K. Seeley, Multidimensional gas chromatography: fundamental advances and new applications, Anal. Chem. 85 (2013) 557–578. doi:10.1021/ac303195u.

[2] J.C. Giddings, Unified Separation Science, John Wiley & Sons, Inc., New York, NY, 1991.

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

[4] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells, Anal. Chem. 78 (2006) 2700–2709. doi:10.1021/ac052106o.

[5] P.-H. Stefanuto, K.A. Perrault, S. Stadler, R. Pesesse, H.N. LeBlanc, S.L. Forbes, J.-F. Focant, GC × GC–TOFMS and supervised multivariate approaches to study human cadaveric decomposition olfactive signatures, Anal. Bioanal. Chem. 407 (2015) 4767–4778. doi:10.1007/s00216-015-8683-5.

[6] Y.F. Wong, S.-T. Chin, P. Perlmutter, P.J. Marriott, Evaluation of comprehensive two-dimensional gas chromatography with accurate mass time-of-flight mass spectrometry for the metabolic profiling of plant–fungus interaction in Aquilaria malaccensis, J. Chromatogr. A. 1387 (2015) 104–115. doi:10.1016/j.chroma.2015.01.096.

[7] M. Jiang, C. Kulsing, Y. Nolvachai, P.J. Marriott, Two-Dimensional Retention Indices Improve Component Identification in Comprehensive Two-Dimensional Gas Chromatography of Saffron, Anal. Chem. (2015) 150515091830003. doi:10.1021/acs.analchem.5b00953.

[8] R. Bro, Review on Multiway Analysis in Chemistry—2000–2005, Crit. Rev. Anal. Chem. 36 (2006) 279–293. doi:10.1080/10408340600969965.

[9] H.-L. Wu, Y. Li, R.-Q. Yu, Recent developments of chemical multiway calibration methodologies with second-order or higher-order advantages: Chemical multiway calibration methodologies, J. Chemom. 28 (2014) 476–489. doi:10.1002/cem.2570.

[10] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149–171. doi:10.1016/S0169-7439(97)00032-4.

[11] C.A. Andersson, R. Bro, The N-way Toolbox for MATLAB, Chemom. Intell. Lab. Syst. 52 (2000) 1–4.

[12] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, J. Chemom. 17 (2003) 274–286. doi:10.1002/cem.801.

[13] A.E. Sinha, C.G. Fraga, B.J. Prazen, R.E. Synovec, Trilinear chemometric analysis of two-dimensional comprehensive gas chromatography–time-of-flight mass spectrometry data, J. Chromatogr. A. 1027 (2004) 269–277. doi:10.1016/j.chroma.2003.08.081.

[14] J.C. Hoggard, R.E. Synovec, Parallel Factor Analysis (PARAFAC) of target analytes in GC × GC−TOFMS data: automated selection of a model with an appropriate number of factors, Anal. Chem. 79 (2007) 1611–1619. doi:10.1021/ac061710b.

[15] J.C. Hoggard, R.E. Synovec, Automated Resolution of Nontarget Analyte Signals in GC × GC-TOFMS Data Using Parallel Factor Analysis, Anal. Chem. 80 (2008) 6677–6688. doi:10.1021/ac800624e.
[16] J.C. Hoggard, W.C. Siegler, R.E. Synovec, Toward automated peak resolution in complete GC × GC–TOFMS chromatograms by PARAFAC, J. Chemom. 23 (2009) 421–431. doi:10.1002/cem.1239.
[17] N.E. Watson, W.C. Siegler, J.C. Hoggard, R.E. Synovec, Comprehensive Three-Dimensional Gas Chromatography with Parallel Factor Analysis, Anal. Chem. 79 (2007) 8270–8280. doi:10.1021/ac070829x.
[18] W.C. Siegler, J.A. Crank, D.W. Armstrong, R.E. Synovec, Increasing selectivity in comprehensive three-dimensional gas chromatography via an ionic liquid stationary phase column in one dimension, J. Chromatogr. A. 1217 (2010) 3144–3149. doi:10.1016/j.chroma.2010.02.082.
[19] N.E. Watson, H.D. Bahaghighat, K. Cui, R.E. Synovec, Comprehensive Three-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry, Anal. Chem. 89 (2017) 1793–1800. doi:10.1021/acs.analchem.6b04112.
[20] S.E.G. Porter, D.R. Stoll, S.C. Rutan, P.W. Carr, J.D. Cohen, Analysis of Four-Way Two-Dimensional Liquid Chromatography-Diode Array Data: Application to Metabolomics, Anal. Chem. 78 (2006) 5559–5569. doi:10.1021/ac0606195.
[21] H. Parastar, J.R. Radović, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Resolution and Quantification of Complex Mixtures of Polycyclic Aromatic Hydrocarbons in Heavy Fuel Oil Sample by Means of GC × GC-TOFMS Combined to Multivariate Curve Resolution, Anal. Chem. 83 (2011) 9289–9297. doi:10.1021/ac201799r.
[22] V. Pacakova, L. Feltl, Chromatographic Retention Indices: an aid to identification of organic compounds, Ellis Horwood, New York, NY, 1992.
[23] S.E. Stein, An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data, J. Am. Soc. Mass Spectrom. 10 (1999) 770–781. doi:10.1016/S1044-0305(99)00047-1.
[24] K.M. Pierce, J.C. Hoggard, Chromatographic data analysis. Part 3.3.4: handling hyphenated data in chromatography, Anal. Methods. 6 (2014) 645–653. doi:10.1039/c3ay40965a.
[25] E. de Hoffmann, V. Stroobant, Mass Spectrometry, John Wiley & Sons, Inc., London, 2002.
[26] S.E. Stein, Estimating probabilities of correct identification from results of mass spectral library searches, J. Am. Soc. Mass Spectrom. 5 (1994) 316–323. doi:10.1016/1044-0305(94)85022-4.
[27] W. Khummueng, J. Harynuk, P.J. Marriott, Modulation Ratio in Comprehensive Two-dimensional Gas Chromatography, Anal. Chem. 78 (2006) 4578–4587. doi:10.1021/ac052270b.
[28] B.A. Parsons, D.K. Pinkerton, B.W. Wright, R.E. Synovec, Chemical characterization of the acid alteration of diesel fuel: non-targeted analysis by two-dimensional gas chromatography coupled with time-of-flight mass spectrometry with tile-based Fisher ratio and combinatorial threshold determination, J. Chromatogr. A. 1440 (2016) 179–190. doi:10.1016/j.chroma.2016.02.067.
[29] D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity deviation ratio: A new metric for chemometric analysis of comprehensive two-dimensional gas

chromatography time-of-flight mass spectrometry data, Anal. Chim. Acta. 871 (2015) 66–76. doi:10.1016/j.aca.2015.02.040.
[30] J.M. Amigo, T. Skov, R. Bro, J. Coello, S. Maspoch, Solving GC-MS problems with PARAFAC2, TrAC Trends Anal. Chem. 27 (2008) 714–725. doi:10.1016/j.trac.2008.05.011.

# Chapter 5. Conclusion

Multi-dimensional chromatography continues to be a central technique in the modern analytical laboratory. It is applied in a multitude of situations for both targeted and non-targeted analyses. Three new techniques were explored which address both of these approaches. Initially delving into the non-targeted F-ratio. Later, modifications to the $GC^3$ were considered and applied, in a targeted way, to a complex sample set.

In Chapter 2 the tile-based F-ratio was studied in detail and validated against a well described dataset. The previously analyzed data (referred to herein as the benchmark dataset) were intracellular extracts from *Saccharomyces cerevisiae* (yeast), either metabolizing glucose (repressed) or ethanol (derepressed), which defined the two classes in the discovery-based analysis to find metabolites that were statistically different in concentration between the two classes. Beneficially, this previously analyzed dataset provided a concrete means to validate the tile-based F-ratio software. Significant benefits of applying tile-based F-ratio analysis were demonstrated. The yeast metabolomics data were analyzed more rapidly, in about one week versus one year for the prior studies with this dataset. Furthermore, a null distribution analysis was implemented to statistically determine an adequate F-ratio threshold, whereby the variables with F-ratio values below the threshold could be ignored as not class distinguishing, which provides the analyst with confidence when analyzing the hit table. Forty-six of the fifty-four benchmarked changing metabolites were discovered by the new methodology while consistently excluding all but one of the benchmarked nineteen false positive metabolites previously identified.

Chapter 3 described further development of the $GC^3$ to include a TOFMS. This instrument provides four dimensions (4D) of chemical selectivity and includes significant

improvements to total selectivity (mass spectrometric and chromatographic), peak identification, and operational temperature range relative to previous models of the $GC^3$ reported. The new instrumental design and data output were evaluated and illustrated via two samples, a 115-component test mixture and a diesel fuel spiked with several compounds, for the purpose of illustrating the chemical selectivity benefits of this instrumental platform. Useful approaches to visualize the 4D data were presented. The $GC^3$ – TOFMS instrument experimentally achieved total peak capacity, $n_{c,3D}$, ranging from 5000 to 9600 ($\bar{x}$= 7000, $s$ = 1700) for 10 representative analytes for 50 min separations with component dimensional peak capacities averaging 406, 3.6, and 4.9 for $^1$D, $^2$D, and $^3$D, respectively. Particularly, $GC^3$ – TOFMS achieved a combined $^2$D × $^3$D peak capacity ranging from 10 to 26 ($\bar{x}$= 17.6, $s$ = 5.0), which is similar to what is achieved by $^2$D alone in a GC × GC operating at equivalent modulation period conditions. The analytical benefits of employing three varied chemical selectivities in the 3D separation coupled with TOFMS were illustrated through the separation and detection of 1,6-dichlorohexane and cyclohexyl isothiocyanate as part of the diesel fuel analysis.

Finally, Comprehensive three-dimensional gas chromatography with time-of-flight mass spectrometry ($GC^3$-TOFMS), which creates an opportunity to explore a new paradigm in chemometric analysis, was probed in Chapter 4. Using this newly described instrument and the well understood Parallel Factor Analysis (PARAFAC) model one option for utilization of the novel $GC^3$-TOFMS data structure was presented. The method builds upon previous work in both $GC^3$ and targeted analysis using PARAFAC to simplify some of the implementation challenges previously discovered. Conceptualizing the $GC^3$-TOFMS as a one-dimensional gas chromatograph with GC × GC-TOFMS detection the instrument was allowed to create the PARAFAC target window natively. Each first dimension modulation thus created a full GC ×

GC-TOFMS chromatogram totally amenable to PARAFAC. A simple mixture of 115 compounds and a diesel sample were interrogated through this methodology. All test analyte targets were successfully identified in both mixtures. In addition, mass spectral matching of the PARAFAC loadings to library spectra yielded results greater than 900 in 40 of 42 test analyte cases. Twenty-nine of these cases produced match values greater than 950.

The successes described in also produce opportunities for future work. Particularly, modification of the F-ratio for application to the $GC^3$ –TOFMS seems promising. Additionally, the initial targeted PARAFAC work suggests the potential for even greater automation of the method. Naturally, these developments can be combined with application of the methods to real scientific problems in the other scientific realms which analytical chemistry supports.

# BIBLIOGRAPHY


A, Jiye, Johan Trygg, Jonas Gullberg, Annika I. Johansson, Pär Jonsson, Henrik Antti, Stefan L. Marklund, and Thomas Moritz. "Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome." *Analytical Chemistry* 77, no. 24 (December 1, 2005): 8086–94. doi:10.1021/ac051211v.

Adams, S.B., L.A. Setton, E. Kensicki, M.P. Bolognesi, A.P. Toth, and D.L. Nettles. "Global Metabolic Profiling of Human Osteoarthritic Synovium." *Osteoarthritis and Cartilage* 20, no. 1 (January 2012): 64–67. doi:10.1016/j.joca.2011.10.010.

Adcock, Jacqui L., Mike Adams, Blagoj S. Mitrevski, and Philip J. Marriott. "Peak Modeling Approach to Accurate Assignment of First-Dimension Retention Times in Comprehensive Two-Dimensional Chromatography." *Analytical Chemistry* 81, no. 16 (August 15, 2009): 6797–6804. doi:10.1021/ac900960n.

Alberts, Bruce, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell*. Fifth. New York, NY: Garland Science, 2008.

Alves, Zélia, André Melo, Ana Raquel Figueiredo, Manuel A. Coimbra, Ana C. Gomes, and Sílvia M. Rocha. "Exploring the Saccharomyces Cerevisiae Volatile Metabolome: Indigenous versus Commercial Strains." *PLoS ONE* 10, no. 11 (November 24, 2015): e0143641. doi:10.1371/journal.pone.0143641.

Amigo, José Manuel, Thomas Skov, and Rasmus Bro. "ChroMATHography: Solving Chromatographic Issues with Mathematical Models and Intuitive Graphics." *Chemical Reviews* 110, no. 8 (August 11, 2010): 4582–4605. doi:10.1021/cr900394n.

Amigo, José Manuel, Thomas Skov, Rasmus Bro, Jordi Coello, and Santiago Maspoch. "Solving GC-MS Problems with PARAFAC2." *TrAC Trends in Analytical Chemistry* 27, no. 8 (September 2008): 714–25. doi:10.1016/j.trac.2008.05.011.

Amster, I. Jonathan, and others. "Fourier Transform Mass Spectrometry." *Journal of Mass Spectrometry* 31, no. 12 (1996): 1325–1337.

Anderson, Jared L., and Daniel W. Armstrong. "Immobilized Ionic Liquids as High-Selectivity/High-Temperature/High-Stability Gas Chromatography Stationary Phases." *Analytical Chemistry* 77, no. 19 (October 2005): 6453–62. doi:10.1021/ac051006f.

Anderson, Jared L., Jie Ding, Thomas Welton, and Daniel W. Armstrong. "Characterizing Ionic Liquids On the Basis of Multiple Solvation Interactions." *Journal of the American Chemical Society* 124, no. 47 (November 2002): 14247–54. doi:10.1021/ja028156h.

Anderson, Jared L., Rongfang Ding, Arkady Ellern, and Daniel W. Armstrong. "Structure and Properties of High Stability Geminal Dicationic Ionic Liquids." *Journal of the American Chemical Society* 127, no. 2 (January 2005): 593–604. doi:10.1021/ja046521u.

Andersson, Claus A, and Rasmus Bro. "The N-Way Toolbox for MATLAB." *Chemometrics and Intelligent Laboratory Systems* 52, no. 1 (August 14, 2000): 1–4. doi:10.1016/S0169-7439(00)00071-X.

"Applied Chemometrics," December 9, 2016. http://www.chemometrics.com/.

Armstrong, Daniel W., Lingfeng He, and Yan-Song Liu. "Examination of Ionic Liquids and Their Interaction with Molecules, When Used as Stationary Phases in Gas

Chromatography." *Analytical Chemistry* 71, no. 17 (September 1999): 3873–76. doi:10.1021/ac990443p.

Arthur, Catherine L., and Janusz Pawliszyn. "Solid Phase Microextraction with Thermal Desorption Using Fused Silica Optical Fibers." *Analytical Chemistry* 62, no. 19 (1990): 2145–2148.

Auernik, K. S., and R. M. Kelly. "Identification of Components of Electron Transport Chains in the Extremely Thermoacidophilic Crenarchaeon Metallosphaera Sedula through Iron and Sulfur Compound Oxidation Transcriptomes." *Applied and Environmental Microbiology* 74, no. 24 (December 15, 2008): 7723–32. doi:10.1128/AEM.01545-08.

———. "Impact of Molecular Hydrogen on Chalcopyrite Bioleaching by the Extremely Thermoacidophilic Archaeon Metallosphaera Sedula." *Applied and Environmental Microbiology* 76, no. 8 (April 15, 2010): 2668–72. doi:10.1128/AEM.02016-09.

———. "Physiological Versatility of the Extremely Thermoacidophilic Archaeon Metallosphaera Sedula Supported by Transcriptomic Analysis of Heterotrophic, Autotrophic, and Mixotrophic Growth." *Applied and Environmental Microbiology* 76, no. 3 (February 1, 2010): 931–35. doi:10.1128/AEM.01336-09.

Auernik, K. S., Y. Maezato, P. H. Blum, and R. M. Kelly. "The Genome Sequence of the Metal-Mobilizing, Extremely Thermoacidophilic Archaeon Metallosphaera Sedula Provides Insights into Bioleaching-Associated Metabolism." *Applied and Environmental Microbiology* 74, no. 3 (February 1, 2008): 682–92. doi:10.1128/AEM.02019-07.

Auernik, Kathryne S, Charlotte R Cooper, and Robert M Kelly. "Life in Hot Acid: Pathway Analyses in Extremely Thermoacidophilic Archaea." *Current Opinion in Biotechnology* 19, no. 5 (October 2008): 445–53. doi:10.1016/j.copbio.2008.08.001.

Bailey, Hope P., Sarah C. Rutan, and Peter W. Carr. "Factors That Affect Quantification of Diode Array Data in Comprehensive Two-Dimensional Liquid Chromatography Using Chemometric Data Analysis." *Journal of Chromatography A* 1218, no. 46 (November 18, 2011): 8411–22. doi:10.1016/j.chroma.2011.09.057.

Bean, Heather D., Jane E. Hill, and Jean-Marie D. Dimandja. "Improving the Quality of Biomarker Candidates in Untargeted Metabolomics via Peak Table-Based Alignment of Comprehensive Two-Dimensional Gas Chromatography–mass Spectrometry Data." *Journal of Chromatography A* 1394 (May 15, 2015): 111–17. doi:10.1016/j.chroma.2015.03.001.

Beckstrom, Andrew C., Elizabeth M. Humston, Laura R. Snyder, Robert E. Synovec, and Sandra E. Juul. "Application of Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry Method to Identify Potential Biomarkers of Perinatal Asphyxia in a Non-Human Primate Model." *Journal of Chromatography A* 1218, no. 14 (April 8, 2011): 1899–1906. doi:10.1016/j.chroma.2011.01.086.

Beebe, Kenneth R., Randy J. Pell, and Mary Beth Seasholtz. *Chemometrics: A Practical Guide*. John Wiley & Sons, Inc., 1998.

Begnaud, Frédéric, Christian Debonneville, Jean-Pierre Probst, Alain Chaintreau, Paul D. Morrison, Jacqui L. Adcock, and Philip J. Marriott. "Effects of Variation in Modulator Temperature during Cryogenic Modulation in Comprehensive Two-Dimensional Gas Chromatography." *Journal of Separation Science* 32, no. 18 (September 1, 2009): 3144–51. doi:10.1002/jssc.200900290.

Bereman, Michael S., Leonard Nyadong, Facundo M. Fernandez, and David C. Muddiman. "Direct High-Resolution Peptide and Protein Analysis by Desorption Electrospray

Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry." *Rapid Communications in Mass Spectrometry* 20, no. 22 (November 30, 2006): 3409–11. doi:10.1002/rcm.2759.

Blumberg, Leonid M., Frank David, Matthew S. Klee, and Pat Sandra. "Comparison of One-Dimensional and Comprehensive Two-Dimensional Separations by Gas Chromatography." *Journal of Chromatography A* 1188, no. 1 (April 2008): 2–16. doi:10.1016/j.chroma.2008.02.044.

Bogdanov, Bogdan, and Richard D. Smith. "Proteomics by FTICR Mass Spectrometry: Top down and Bottom up." *Mass Spectrometry Reviews* 24, no. 2 (March 2005): 168–200. doi:10.1002/mas.20015.

Box, George E.P., and George C. Tiao. *Bayesian Inference In Statistical Analysis*. Addison-Wesley Series In Behavioral Science: Quantitative Methods. Reading, MA: Addison-Wesley Publishing Company, 1973.

Bressanello, Davide, Erica Liberto, Massimo Collino, Stephen E. Reichenbach, Elisa Benetti, Fausto Chiazza, Carlo Bicchi, and Chiara Cordero. "Urinary Metabolic Fingerprinting of Mice with Diet-Induced Metabolic Derangements by Parallel Dual Secondary Column-Dual Detection Two-Dimensional Comprehensive Gas Chromatography." *Journal of Chromatography A* 1361 (September 26, 2014): 265–76. doi:10.1016/j.chroma.2014.08.015.

Bro, Rasmus. "PARAFAC. Tutorial and Applications." *Chemometrics and Intelligent Laboratory Systems* 38, no. 2 (October 1997): 149–71. doi:10.1016/S0169-7439(97)00032-4.

———. "Review on Multiway Analysis in Chemistry—2000–2005." *Critical Reviews in Analytical Chemistry* 36, no. 3–4 (December 2006): 279–93. doi:10.1080/10408340600969965.

Bro, Rasmus, and Henk A. L. Kiers. "A New Efficient Method for Determining the Number of Components in PARAFAC Models." *Journal of Chemometrics* 17, no. 5 (May 1, 2003): 274–86. doi:10.1002/cem.801.

Brokl, Michał, Louise Bishop, Christopher G. Wright, Chuan Liu, Kevin McAdam, and Jean-François Focant. "Multivariate Analysis of Mainstream Tobacco Smoke Particulate Phase by Headspace Solid-Phase Micro Extraction Coupled with Comprehensive Two-Dimensional Gas Chromatography–time-of-Flight Mass Spectrometry." *Journal of Chromatography A* 1370 (November 28, 2014): 216–29. doi:10.1016/j.chroma.2014.10.057.

Brubaker, W. M. "Influence of Space Charge on the Potential Distribution in Mass Spectrometer Ion Sources." *Journal of Applied Physics* 26, no. 8 (August 1, 1955): 1007–12. doi:10.1063/1.1722124.

Bruckner, Carsten A., Bryan J. Prazen, and Robert E. Synovec. "Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis." *Analytical Chemistry* 70, no. 14 (July 1, 1998): 2796–2804. doi:10.1021/ac980164m.

Bryant, Chet K., Peter T. LaPuma, Gary L. Hook, and Eric J. Houser. "Chemical Agent Identification by Field-Based Attenuated Total Reflectance Infrared Detection and Solid-Phase Microextraction." *Analytical Chemistry* 79, no. 6 (March 2007): 2334–40. doi:10.1021/ac062047u.

Bueno Jr., Pedro A., and John V. Seeley. "Flow-Switching Device for Comprehensive Two-Dimensional Gas Chromatography." *Journal of Chromatography A*, 26th International

Symposium on Capillary Chromatography and Electrophoresis, 1027, no. 1–2 (February 20, 2004): 3–10. doi:10.1016/j.chroma.2003.10.033.

Burgert, Ralf, Sarah T. Stokes, Kit H. Bowen, and Hansgeorg Schnöckel. "Primary Reaction Steps of $Al_{13}^-$ Clusters in an HCl Atmosphere: Snapshots of the Dissolution of a Base Metal." *Journal of the American Chemical Society* 128, no. 24 (June 2006): 7904–8. doi:10.1021/ja060613x.

Bushey, Michelle M., and James W. Jorgenson. "Automated Instrumentation for Comprehensive Two-Dimensional High-Performance Liquid Chromatography of Proteins." *Analytical Chemistry* 62, no. 2 (1990): 161–167.

———. "Automated Instrumentation for Comprehensive Two-Dimensional High-Performance Liquid Chromatography/capillary Zone Electrophoresis." *Analytical Chemistry* 62, no. 10 (May 1, 1990): 978–84. doi:10.1021/ac00209a002.

Cai, Huamin, and Stanley D. Stearns. "Partial Modulation Method via Pulsed Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography." *Analytical Chemistry* 76, no. 20 (October 2004): 6064–76. doi:10.1021/ac0492463.

Carlin, Bradley P., and Thomas A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd ed. Texts in Statistical Science Series. Boca Raton, FL: Chapman & Hall/CRC, 2000.

Castillo, Sandra, Ismo Mattila, Jarkko Miettinen, Matej Orešič, and Tuulia Hyötyläinen. "Data Analysis Tool for Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry." *Analytical Chemistry* 83, no. 8 (April 15, 2011): 3058–67. doi:10.1021/ac103308x.

Castrillo, Juan I, Andrew Hayes, Shabaz Mohammed, Simon J Gaskell, and Stephen G Oliver. "An Optimized Protocol for Metabolome Analysis in Yeast Using Direct Infusion Electrospray Mass Spectrometry." *Phytochemistry*, Plant Metabolomics, 62, no. 6 (March 2003): 929–37. doi:10.1016/S0031-9422(02)00713-6.

Chen, Di, Jung Hwan Seo, Jing Liu, Katsuo Kurabayashi, and Xudong Fan. "Smart Three-Dimensional Gas Chromatography." *Analytical Chemistry* 85, no. 14 (July 16, 2013): 6871–75. doi:10.1021/ac401152v.

Chen, Xueguo, Liang Kong, Xingye Su, Hongjing Fu, Jianyi Ni, Ruihuan Zhao, and Hanfa Zou. "Separation and Identification of Compounds in Rhizoma Chuanxiong by Comprehensive Two-Dimensional Liquid Chromatography Coupled to Mass Spectrometry." *Journal of Chromatography A* 1040, no. 2 (June 2004): 169–78. doi:10.1016/j.chroma.2004.04.002.

Chen, Zhongxue, Monnie McGee, Qingzhong Liu, and Richard H. Scheuermann. "A Distribution Free Summarization Method for Affymetrix GeneChip® Arrays." *Bioinformatics* 23, no. 3 (February 1, 2007): 321–27. doi:10.1093/bioinformatics/btl609.

Chun, Pattaraporn T., Ronald J. McPherson, Luke C. Marney, Sahar Z. Zangeneh, Brendon A. Parsons, Ali Shojaie, Robert E. Synovec, and Sandra E. Juul. "Serial Plasma Metabolites Following Hypoxic-Ischemic Encephalopathy in a Nonhuman Primate Model." *Developmental Neuroscience* 37, no. 2 (February 27, 2015): 161–71. doi:10.1159/000370147.

Clegg, D. E., and D. L. Massart. "Information Theory and Its Application to Analytical Chemistry." *Journal of Chemical Education* 70, no. 1 (1993): 19.

Cordeiro de Amorim, Renato, Vladimir Makarenkov, and Boris Mirkin. "A-Wardpβ: Effective Hierarchical Clustering Using the Minkowski Metric and a Fast K-Means Initialisation."

*Information Sciences* 370–371 (November 20, 2016): 343–54. doi:10.1016/j.ins.2016.07.076.

Cordero, Chiara, Erica Liberto, Carlo Bicchi, Patrizia Rubiolo, Peter Schieberle, Stephen E. Reichenbach, and Qingping Tao. "Profiling Food Volatiles by Comprehensive Two-Dimensional Ga Schromatography Coupled with Mass Spectrometry: Advanced Fingerprinting Approaches for Comparative Analysis of the Volatile Fraction of Roasted Hazelnuts (Corylus Avellana L.) from Different Origins." *Journal of Chromatography A* 1217, no. 37 (September 10, 2010): 5848–58. doi:10.1016/j.chroma.2010.07.006.

Cottingham, Katie. "Product Review: FTMS: Overcoming Challenges." *Analytical Chemistry* 78, no. 3 (2006): 655–657.

Cuttler, Jerry M. "Commentary on the Appropriate Radiation Level for Evacuations." *Dose-Response* 10, no. 4 (January 1, 2012): 473–79. doi:10.2203/dose-response.12-013.Cuttler.

Daiichi, Fukushima. "ANS Committee Report." *Rep. American Nuclear Society*, 2012. http://205.254.131.69/nuclearsafety/nfsp/FacRep/Workshop2012/Presentations/FR/11-KevinGray_Fukushima_Refs/ANS_SpecialReport.pdf.

Dallüge, Jens, Martijn van Rijn, Jan Beens, René J. J Vreuls, and Udo A. Th Brinkman. "Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometric Detection Applied to the Determination of Pesticides in Food Extracts." *Journal of Chromatography A,* Application of Theory to the Practice and Understanding of Chromatography, 965, no. 1–2 (August 2, 2002): 207–17. doi:10.1016/S0021-9673(01)01324-3.

Davis, Joe M. "Statistical Theory of Spot Overlap for N-Dimensional Separations." *Analytical Chemistry* 65, no. 15 (1993): 2014–2023.

———. "Statistical Theory of Spot Overlap in Two-Dimensional Separations." *Analytical Chemistry* 63, no. 19 (1991): 2141–2152.

———. "Statistical-Overlap Theory for Elliptical Zones of High Aspect Ratio in Comprehensive Two-Dimensional Separations." *Journal of Separation Science* 28, no. 4 (March 2005): 347–59. doi:10.1002/jssc.200401798.

Davis, Joe M., and J. Calvin Giddings. "Statistical Theory of Component Overlap in Multicomponent Chromatograms." *Analytical Chemistry* 55, no. 3 (1983): 418–424.

Davis, Joe M., Dwight R. Stoll, and Peter W. Carr. "Dependence of Effective Peak Capacity in Comprehensive Two-Dimensional Separations on the Distribution of Peak Capacity between the Two Dimensions." *Analytical Chemistry* 80, no. 21 (November 1, 2008): 8122–34. doi:10.1021/ac800933z.

———. "Effect of First-Dimension Undersampling on Effective Peak Capacity in Comprehensive Two-Dimensional Separations." *Analytical Chemistry* 80, no. 2 (January 2008): 461–73. doi:10.1021/ac071504j.

Davydova, Ekaterina, Peter J. Schoenmakers, and Gabriel Vivó-Truyols. "Study on the Performance of Different Types of Three-Dimensional Chromatographic Systems." *Journal of Chromatography A* 1271, no. 1 (January 2013): 137–43. doi:10.1016/j.chroma.2012.11.043.

Dettmer, Katja, Pavel A. Aronov, and Bruce D. Hammock. "Mass Spectrometry-Based Metabolomics." *Mass Spectrometry Reviews* 26, no. 1 (January 1, 2007): 51–78. doi:10.1002/mas.20108.

Dillner, Debra K., Robert F. Ferrante, Jeffrey P. Fitzgerald, William B. Heuer, and Maria J. Schroeder. “Integrated Laboratories: Crossing Traditional Boundaries.” *Journal of Chemical Education* 84, no. 10 (2007): 1706.

Dimandja, Jean-Marie D., Garrick C. Clouden, Ivelisse Colón, Jean-François Focant, Whitney V. Cabey, and Ritchard C. Parry. “Standardized Test Mixture for the Characterization of Comprehensive Two-Dimensional Gas Chromatography Columns: The Phillips Mix.” *Journal of Chromatography A*, First International Symposium on Comprehensive Multidimensional Gas Chromatography, 1019, no. 1–2 (November 26, 2003): 261–72. doi:10.1016/j.chroma.2003.09.027.

Duda, Richard O., Peter E. Hart, and David G. Stork. *Pattern Classification*. New York, NY: John Wiley & Sons, Inc., 2001.

Dunlop, Peter J., C. M. Bignell, J. F. Jackson, and D. Brynn Hibbert. “Chemometric Analysis of Gas Chromatographic Data of Oils from Eucalyptus Species.” *Chemometrics and Intelligent Laboratory Systems* 30, no. 1 (November 1995): 59–67. doi:10.1016/0169-7439(95)00036-4.

Dunn, Warwick B, David Broadhurst, Paul Begley, Eva Zelena, Sue Francis-McIntyre, Nadine Anderson, Marie Brown, et al. “Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry.” *Nature Protocols* 6, no. 7 (June 30, 2011): 1060–83. doi:10.1038/nprot.2011.335.

Dunn, Warwick B., David Broadhurst, David I. Ellis, Marie Brown, Anthony Halsall, Steven O’Hagan, Irena Spasic, Andrew Tseng, and Douglas B. Kell. “A GC-TOF-MS Study of the Stability of Serum and Urine Metabolomes during the UK Biobank Sample Collection and Preparation Protocols.” *International Journal of Epidemiology* 37, no. suppl_1 (April 1, 2008): i23–30. doi:10.1093/ije/dym281.

Edam, R., J. Blomberg, H.-G. Janssen, and P.J. Schoenmakers. “Comprehensive Multi-Dimensional Chromatographic Studies on the Separation of Saturated Hydrocarbon Ring Structures in Petrochemical Samples.” *Journal of Chromatography A* 1086, no. 1–2 (September 2005): 12–20. doi:10.1016/j.chroma.2005.02.048.

Efron, Bradley. “Bayesians, Frequentists, and Scientists.” *Journal of the American Statistical Association* 100, no. 469 (March 1, 2005): 1–5. doi:10.1198/016214505000000033.

———. “Correlated Z-Values and the Accuracy of Large-Scale Statistical Estimates.” *Journal of the American Statistical Association* 105, no. 491 (September 1, 2010): 1042–55. doi:10.1198/jasa.2010.tm09129.

———. “Correlation and Large-Scale Simultaneous Significance Testing.” *Journal of the American Statistical Association* 102, no. 477 (March 1, 2007): 93–103. doi:10.1198/016214506000001211.

Efron, Bradley, and Robert Tibshirani. “Empirical Bayes Methods and False Discovery Rates for Microarrays.” *Genetic Epidemiology* 23, no. 1 (June 1, 2002): 70–86. doi:10.1002/gepi.1124.

Eiceman, Gary A., Jorge Gardea-Torresdey, Ed Overton, Kenneth Carney, and Frank Dorman. “Gas Chromatography.” *Analytical Chemistry* 74, no. 12 (June 2002): 2771–80. doi:10.1021/ac020210p.

“Eigenvector Research: Chemometrics Software, Consulting and Training,” December 9, 2016. http://eigenvector.com/.

Erney, D. R., A. M. Gillespie, D. M. Gilvydis, and C. F. Poole. “Explanation of the Matrix-Induced Chromatographic Response Enhancement of Organophosphorus Pesticides during Open Tubular Column Gas Chromatography with Splitless or Hot on-Column Injection and Flame Photometric Detection.” *Journal of Chromatography A* 638, no. 1 (May 21, 1993): 57–63. doi:10.1016/0021-9673(93)85007-T.

Fang, Mingliang, Julijana Ivanisevic, H. Paul Benton, Caroline H. Johnson, Gary J. Patti, Linh T. Hoang, Winnie Uritboonthai, Michael E. Kurczy, and Gary Siuzdak. “Thermal Degradation of Small Molecules: A Global Metabolomic Investigation.” *Analytical Chemistry*, October 4, 2015. doi:10.1021/acs.analchem.5b03003.

Fernandez, Facundo M., Vicki H. Wysocki, Jean H. Futrell, and Julia Laskin. “Protein Identification via Surface-Induced Dissociation in an FT-ICR Mass Spectrometer and a Patchwork Sequencing Approach.” *Journal of the American Society for Mass Spectrometry* 17, no. 5 (May 2006): 700–709. doi:10.1016/j.jasms.2006.01.012.

Ferreira, Sergio Luis Costa, Roy Edward Bruns, Erik Galvão Paranhos da Silva, Walter Nei Lopes dos Santos, Cristina Maria Quintella, Jorge Mauricio David, Jailson Bittencourt de Andrade, Marcia Cristina Breitkreitz, Isabel Cristina Sales Fontes Jardim, and Benicio Barros Neto. “Statistical Designs and Response Surface Techniques for the Optimization of Chromatographic Systems.” *Journal of Chromatography A*, Data Analysis in Chromatography, 1158, no. 1–2 (July 27, 2007): 2–14. doi:10.1016/j.chroma.2007.03.051.

Fiehn, Oliver, Joachim Kopka, Richard N. Trethewey, and Lothar Willmitzer. “Identification of Uncommon Plant Metabolites Based on Calculation of Elemental Compositions Using Gas Chromatography and Quadrupole Mass Spectrometry.” *Analytical Chemistry* 72, no. 15 (August 1, 2000): 3573–80. doi:10.1021/ac991142i.

Fisher, Ronald A. *Statistical Methods for Research Workers*. 14th ed. Oliver and Boyd, 1970.

Fitz, Brian D., Brandyn C. Mannion, Khang To, Trinh Hoac, and Robert E. Synovec. “Evaluation of Injection Methods for Fast, High Peak Capacity Separations with Low Thermal Mass Gas Chromatography.” *Journal of Chromatography A* 1392 (May 1, 2015): 82–90. doi:10.1016/j.chroma.2015.03.009.

Fitz, Brian D., Brooke C. Reaser, David K. Pinkerton, Jamin C. Hoggard, Kristen J. Skogerboe, and Robert E. Synovec. “Enhancing Gas Chromatography–Time of Flight Mass Spectrometry Data Analysis Using Two-Dimensional Mass Channel Cluster Plots.” *Analytical Chemistry* 86, no. 8 (April 15, 2014): 3973–79. doi:10.1021/ac5004344.

Fitz, Brian D., and Robert E. Synovec. “Extension of the Two-Dimensional Mass Channel Cluster Plot Method to Fast Separations Utilizing Low Thermal Mass Gas Chromatography with Time-of-Flight Mass Spectrometry.” *Analytica Chimica Acta* 913 (March 2016): 160–70. doi:10.1016/j.aca.2016.01.045.

Fitz, Brian D., Ryan B. Wilson, Brendon A. Parsons, Jamin C. Hoggard, and Robert E. Synovec. “Fast, High Peak Capacity Separations in Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry.” *Journal of Chromatography A* 1266 (November 2012): 116–23. doi:10.1016/j.chroma.2012.09.096.

Fraga, Carlos G., Gabriel A. Pérez Acosta, Michael D. Crenshaw, Krys Wallace, Gary M. Mong, and Heather A Colburn. “Impurity Profiling to Match a Nerve Agent to Its Precursor Source for Chemical Forensics Applications.” *Analytical Chemistry* 83, no. 24 (December 15, 2011): 9564–72. doi:10.1021/ac202340u.

Freye, Chris E., Lan Mu, and Robert E. Synovec. “High Temperature Diaphragm Valve-Based Comprehensive Two-Dimensional Gas Chromatography.” *Journal of Chromatography A* 1424 (December 11, 2015): 127–33. doi:10.1016/j.chroma.2015.10.098.

Freye, Chris E., and Robert E. Synovec. “High Temperature Diaphragm Valve-Based Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry.” *Talanta* 161 (December 1, 2016): 675–80. doi:10.1016/j.talanta.2016.09.002.

Geladi, Paul, and Bruce R. Kowalski. “Partial Least-Squares Regression: A Tutorial.” *Analytica Chimica Acta* 185 (1986): 1–17.

Giddings, J. Calvin. *Unified Separation Science*. New York, NY: John Wiley & Sons, Inc., 1991.

Gilar, Martin, Petra Olivova, Amy E. Daly, and John C. Gebler. “Orthogonality of Separation in Two-Dimensional Liquid Chromatography.” *Analytical Chemistry* 77, no. 19 (October 2005): 6426–34. doi:10.1021/ac050923i.

Górecki, Tadeusz, James Harynuk, and Ognjen Panić. “The Evolution of Comprehensive Two-Dimensional Gas Chromatography (GC×GC).” *Journal of Separation Science* 27, no. 5–6 (April 1, 2004): 359–79. doi:10.1002/jssc.200301650.

Gorrochategui, Eva, Joaquim Jaumot, Sílvia Lacorte, and Romà Tauler. “Data Analysis Strategies for Targeted and Untargeted LC-MS Metabolomic Studies: Overview and Workflow.” *TrAC Trends in Analytical Chemistry* 82 (September 2016): 425–42. doi:10.1016/j.trac.2016.07.004.

Grannas, Amanda M., William C. Hockaday, Patrick G. Hatcher, Lonnie G. Thompson, and Ellen Mosley-Thompson. “New Revelations on the Nature of Organic Matter in Ice Cores.” *Journal of Geophysical Research* 111, no. D4 (2006). doi:10.1029/2005JD006251.

Greiner, Thomas U., Tuulia Hyötyläinen, Mikael Knip, Fredrik Bäckhed, and Matej Orešič. “The Gut Microbiota Modulates Glycaemic Control and Serum Metabolite Profiles in Non-Obese Diabetic Mice.” *PLoS ONE* 9, no. 11 (November 12, 2014). doi:10.1371/journal.pone.0110359.

Gröger, Th., M. Schäffer, M. Pütz, B. Ahrens, K. Drew, M. Eschner, and R. Zimmermann. “Application of Two-Dimensional Gas Chromatography Combined with Pixel-Based Chemometric Processing for the Chemical Profiling of Illicit Drug Samples.” *Journal of Chromatography A* 1200, no. 1 (July 18, 2008): 8–16. doi:10.1016/j.chroma.2008.05.028.

Gromski, Piotr S., Howbeer Muhamadali, David I. Ellis, Yun Xu, Elon Correa, Michael L. Turner, and Royston Goodacre. “A Tutorial Review: Metabolomics and Partial Least Squares-Discriminant Analysis – a Marriage of Convenience or a Shotgun Wedding.” *Analytica Chimica Acta* 879 (June 16, 2015): 10–23. doi:10.1016/j.aca.2015.02.012.

Gross, Gwen M., Bryan J. Prazen, Jay W. Grate, and Robert E. Synovec. “High-Speed Gas Chromatography Using Synchronized Dual-Valve Injection.” *Analytical Chemistry* 76, no. 13 (2004): 3517–24. doi:10.1021/ac049909g.

Harley, J., and V. Pretorius. “A New Detector for Vapour Phase Chromatography.” *Nature* 178, no. 4544 (December 1, 1956): 1244–1244. doi:10.1038/1781244b0.

Harris, Daniel C. *Quantitative Chemical Analysis*. 6th ed. New York, NY: W. H. Freeman and Company, 2003.

Harshman, Richard A., and Margaret E. Lundy. “PARAFAC: Parallel Factor Analysis.” *Computational Statistics & Data Analysis* 18, no. 1 (August 1994): 39–72. doi:10.1016/0167-9473(94)90132-5.

Harynuk, James, and Tadeusz Górecki. “Comprehensive Two-Dimensional Gas Chromatography in Stop-Flow Mode.” *Journal of Separation Science* 27, no. 5–6 (April 2004): 431–41. doi:10.1002/jssc.200301649.

He, Fei, Mark R. Emmett, Kristina Håkansson, Christopher L. Hendrickson, and Alan G. Marshall. “Theoretical and Experimental Prospects for Protein Identification Based Solely on Accurate Mass Measurement.” *Journal of Proteome Research* 3, no. 1 (February 2004): 61–67. doi:10.1021/pr034058z.

Head, Megan L., Luke Holman, Rob Lanfear, Andrew T. Kahn, and Michael D. Jennions. “The Extent and Consequences of P-Hacking in Science.” *PLoS Biol* 13, no. 3 (March 13, 2015): e1002106. doi:10.1371/journal.pbio.1002106.

Henriksen, Thormod, and H. David. Maillie. *Radiation and Health*. London; New York: Taylor & Francis, 2003.

Hiller, Karsten, Jasper Hangebrauk, Christian Jäger, Jana Spura, Kerstin Schreiber, and Dietmar Schomburg. “MetaboliteDetector: Comprehensive Analysis Tool for Targeted and Nontargeted GC/MS Based Metabolome Analysis.” *Analytical Chemistry* 81, no. 9 (May 1, 2009): 3429–39. doi:10.1021/ac802689c.

Hiller, Karsten, Christian M. Metallo, Joanne K Kelleher, and Gregory Stephanopoulos. “Nontargeted Elucidation of Metabolic Pathways Using Stable-Isotope Tracers and Mass Spectrometry.” *Analytical Chemistry* 82, no. 15 (August 2010): 6621–28. doi:10.1021/ac1011574.

Hippel, Frank von. “The Uncertain Future of Nuclear Energy.” *International Panel on Fissile Materials (IPFM), Research Report*, no. 9 (2010): 3.

Ho, Tien D., Cheng Zhang, Leandro W. Hantao, and Jared L. Anderson. “Ionic Liquids in Analytical Chemistry: Fundamentals, Advances, and Perspectives.” *Analytical Chemistry* 86, no. 1 (January 7, 2014): 262–85. doi:10.1021/ac4035554.

Hoffmann, Edmond de, and Vincent Stroobant. *Mass Spectrometry*. London: John Wiley & Sons, Inc., 2002.

Hoggard, Jamin C., W. Christopher Siegler, and Robert E. Synovec. “Toward Automated Peak Resolution in Complete GC × GC–TOFMS Chromatograms by PARAFAC.” *Journal of Chemometrics* 23, no. 7–8 (July 1, 2009): 421–31. doi:10.1002/cem.1239.

Hoggard, Jamin C., and Robert E. Synovec. “Automated Resolution of Nontarget Analyte Signals in GC × GC-TOFMS Data Using Parallel Factor Analysis.” *Analytical Chemistry* 80, no. 17 (September 1, 2008): 6677–88. doi:10.1021/ac800624e.

———. “Parallel Factor Analysis (PARAFAC) of Target Analytes in GC × GC−TOFMS Data: Automated Selection of a Model with an Appropriate Number of Factors.” *Analytical Chemistry* 79, no. 4 (February 1, 2007): 1611–19. doi:10.1021/ac061710b.

Hoggard, Jamin C., Jon H. Wahl, Robert E. Synovec, Gary M. Mong, and Carlos G. Fraga. “Impurity Profiling of a Chemical Weapon Precursor for Possible Forensic Signatures by Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry and Chemometrics.” *Analytical Chemistry* 82, no. 2 (January 15, 2010): 689–98. doi:10.1021/ac902247x.

Hope, Janiece L., Bryan J. Prazen, Erik J. Nilsson, Mary E. Lidstrom, and Robert E. Synovec. “Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry Detection: Analysis of Amino Acid and Organic Acid Trimethylsilyl Derivatives, with Application to the Analysis of Metabolites in Rye Grass Samples.”

*Talanta*, Evaluation/Validation of Novel Biosensors in Real Environmental and Food Samples, 65, no. 2 (January 30, 2005): 380–88. doi:10.1016/j.talanta.2004.06.025.

Hope, Janiece L., Amanda E. Sinha, Bryan J. Prazen, and Robert E. Synovec. "Evaluation of the DotMap Algorithm for Locating Analytes of Interest Based on Mass Spectral Similarity in Data Collected Using Comprehensive Two-Dimensional Gas Chromatography Coupled with Time-of-Flight Mass Spectrometry." *Journal of Chromatography A*, 2nd International Symposium on Comprehensive Multidimensional Gas Chromatography 2nd International Symposium on Comprehensive Multidimensional Gas Chromatography, 1086, no. 1–2 (September 9, 2005): 185–92. doi:10.1016/j.chroma.2005.06.026.

Humston, Elizabeth M., Kenneth M. Dombek, Jamin C. Hoggard, Elton T. Young, and Robert E. Synovec. "Time-Dependent Profiling of Metabolites from Snf1 Mutant and Wild Type Yeast Cells." *Analytical Chemistry* 80, no. 21 (November 2008): 8002–11. doi:10.1021/ac800998j.

Humston, Elizabeth M., Jamin C. Hoggard, and Robert E. Synovec. "Utilizing the Third Order Advantage with Isotope Dilution Mass Spectrometry." *Analytical Chemistry* 82, no. 1 (January 1, 2010): 41–43. doi:10.1021/ac902184b.

Humston, Elizabeth M., Yan Zhang, Gregory F. Brabeck, Andrew McShea, and Robert E. Synovec. "Development of a GC×GC–TOFMS Method Using SPME to Determine Volatile Compounds in Cacao Beans." *Journal of Separation Science* 32, no. 13 (July 1, 2009): 2289–95. doi:10.1002/jssc.200900143.

Humston, Elizabeth M., and Synovec, Robert E. "Comprehensive Chromatography Data Interpretation Technologies." In *Comprehensive Chromatography in Combination with Mass Spectrometry*, 449–75. Wiley-Interscience Series in Mass Spectrometry. Hoboken, NJ: John Wiley & Sons, Inc., 2011.

Ibáñez, Clara, Carolina Simó, Dinesh K. Barupal, Oliver Fiehn, Miia Kivipelto, Angel Cedazo-Mínguez, and Alejandro Cifuentes. "A New Metabolomic Workflow for Early Detection of Alzheimer's Disease." *Journal of Chromatography A* 1302 (August 9, 2013): 65–71. doi:10.1016/j.chroma.2013.06.005.

"Identification of Disease- and Nutrient-Related Metabolic Fingerprints in Osteoarthritic Guinea Pigs," n.d.

Irvine, Solveig A., Amanda K. Schoettmer, Ronald W. Bates, and Michael L. Meyer. "History of Sulfur Content Effects on the Thermal Stability of RP-1 under Heated Conditions." *History* 11 (2004): 14.

Jaumot, Joaquim, Anna de Juan, and Romà Tauler. "MCR-ALS GUI 2.0: New Features and Applications." *Chemometrics and Intelligent Laboratory Systems* 140 (January 15, 2015): 1–12. doi:10.1016/j.chemolab.2014.10.003.

Jiang, Ming, Chadin Kulsing, Yada Nolvachai, and Philip J. Marriott. "Two-Dimensional Retention Indices Improve Component Identification in Comprehensive Two-Dimensional Gas Chromatography of Saffron." *Analytical Chemistry*, May 15, 2015, 150515091830003. doi:10.1021/acs.analchem.5b00953.

Johnson, Kevin J, and Robert E Synovec. "Pattern Recognition of Jet Fuels: Comprehensive GC×GC with ANOVA-Based Feature Selection and Principal Component Analysis." *Chemometrics and Intelligent Laboratory Systems*, Fourth International Conference on Environ metrics and Chemometrics held in Las Vegas, NV, USA, 18-20 September 2000, 60, no. 1–2 (January 28, 2002): 225–37. doi:10.1016/S0169-7439(01)00198-8.

Junge, Melanie, Stefan Bieri, Helmut Huegel, and Philip J. Marriott. “Fast Comprehensive Two-Dimensional Gas Chromatography with Cryogenic Modulation.” *Analytical Chemistry* 79, no. 12 (June 2007): 4448–54. doi:10.1021/ac062351n.

Kaiser, Theodore J., J. Will Thompson, J. Scott Mellors, and James W. Jorgenson. “Capillary-Based Instrument for the Simultaneous Measurement of Solution Viscosity and Solute Diffusion Coefficient at Pressures up to 2000 Bar and Implications for Ultrahigh Pressure Liquid Chromatography.” *Analytical Chemistry* 81, no. 8 (April 15, 2009): 2860–68. doi:10.1021/ac802467k.

Kehimkar, Benjamin, Jamin C. Hoggard, Luke C. Marney, Matthew C. Billingsley, Carlos G. Fraga, Thomas J. Bruno, and Robert E. Synovec. “Correlation of Rocket Propulsion Fuel Properties with Chemical Composition Using Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry Followed by Partial Least Squares Regression Analysis.” *Journal of Chromatography A* 1327 (January 2014): 132–40. doi:10.1016/j.chroma.2013.12.060.

Kehimkar, Benjamin, Jamin C. Hoggard, Jeremy S. Nadeau, and Robert E. Synovec. “Targeted Mass Spectral Ratio Analysis: A New Tool for Gas Chromatography—mass Spectrometry.” *Talanta* 103 (January 15, 2013): 267–75. doi:10.1016/j.talanta.2012.10.043.

Kelly, Jeffrey J., Clyde H. Barlow, Thomas M. Jinguji, and James B. Callis. “Prediction of Gasoline Octane Numbers from near-Infrared Spectral Features in the Range 660-1215 Nm.” *Analytical Chemistry* 61, no. 4 (1989): 313–320.

Khummueng, Weeraya, James Harynuk, and Philip J. Marriott. “Modulation Ratio in Comprehensive Two-Dimensional Gas Chromatography.” *Analytical Chemistry* 78, no. 13 (July 2006): 4578–87. doi:10.1021/ac052270b.

Kim, Sunghwan, Ryan P. Rodgers, and Alan G. Marshall. “Truly ‘exact’ Mass: Elemental Composition Can Be Determined Uniquely from Molecular Mass Measurement at ~0.1mDa Accuracy for Molecules up to ~500Da.” *International Journal of Mass Spectrometry* 251, no. 2–3 (April 2006): 260–65. doi:10.1016/j.ijms.2006.02.001.

Klee, Matthew S., Jack Cochran, Mark Merrick, and Leonid M. Blumberg. “Evaluation of Conditions of Comprehensive Two-Dimensional Gas Chromatography That Yield a near-Theoretical Maximum in Peak Capacity Gain.” *Journal of Chromatography A* 1383 (February 27, 2015): 151–59. doi:10.1016/j.chroma.2015.01.031.

Kolwicz, Stephen C., David P. Olson, Luke C. Marney, Lorena Garcia-Menendez, Robert E. Synovec, and Rong Tian. “Cardiac-Specific Deletion of Acetyl CoA Carboxylase 2 Prevents Metabolic Remodeling During Pressure-Overload Hypertrophy.” *Circulation Research* 111, no. 6 (August 31, 2012): 728–38. doi:10.1161/CIRCRESAHA.112.268128.

Kong, Weijun, Jie Xiong, Huan Yue, and Zhifeng Fu. “Sandwich Fluorimetric Method for Specific Detection of Staphylococcus Aureus Based on Antibiotic-Affinity Strategy.” *Analytical Chemistry*, September 9, 2015. doi:10.1021/acs.analchem.5b02301.

Koning, Sjaak de, Hans-Gerd Janssen, Marieke van Deursen, and Udo A. Th. Brinkman. “Automated on-Line Comprehensive Two-Dimensional LC×GC and LC×GC–ToF MS: Instrument Design and Application to Edible Oil and Fat Analysis.” *Journal of Separation Science* 27, no. 5–6 (April 1, 2004): 397–409. doi:10.1002/jssc.200301676.

Kramer, Richard. *Chemometric Techniques for Quantitative Analysis*. Marcel Dekker, Inc., 1998.

Kuiper, K. F., A. Deino, F. J. Hilgen, W. Krijgsman, P. R. Renne, and J. R. Wijbrans. "Synchronizing Rock Clocks of Earth History." *Science* 320, no. 5875 (April 25, 2008): 500–504. doi:10.1126/science.1154339.

Kumar, Naveen, Ankit Bansal, G.S. Sarma, and Ravindra K. Rawal. "Chemometrics Tools Used in Analytical Chemistry: An Overview." *Talanta* 123 (June 2014): 186–99. doi:10.1016/j.talanta.2014.02.003.

Kuriki, Ayako, Takeshi Kumazawa, Xiao-Pen Lee, Chika Hasegawa, Mitsuru Kawamura, Osamu Suzuki, and Keizo Sato. "Simultaneous Determination of Selegiline and Desmethylselegiline in Human Body Fluids by Headspace Solid-Phase Microextraction and Gas Chromatography–mass Spectrometry." *Journal of Chromatography B* 844, no. 2 (December 2006): 283–91. doi:10.1016/j.jchromb.2006.07.019.

La Rosa, Ruggero, Volker Behrends, Huw D. Williams, Jacob G. Bundy, and Fernando Rojo. "Influence of the Crc Regulator on the Hierarchical Use of Carbon Sources from a Complete Medium in Pseudomonas." *Environmental Microbiology* 18, no. 3 (March 1, 2016): 807–18. doi:10.1111/1462-2920.13126.

LaClair, Russell W., Pedro A. Bueno, and John V. Seeley. "A Systematic Analysis of a Flow-Switching Modulator for Comprehensive Two-Dimensional Gas Chromatography." *Journal of Separation Science* 27, no. 5–6 (April 2004): 389–96. doi:10.1002/jssc.200301668.

Lamers, R.J.A.N., J.H.J. van Nesselrooij, V.B. Kraus, J.M. Jordan, J.B. Renner, A.D. Dragomir, G. Luta, J. van der Greef, and J. DeGroot. "Identification of an Urinary Metabolite Profile Associated with Osteoarthritis." *Osteoarthritis and Cartilage* 13, no. 9 (September 2005): 762–68. doi:10.1016/j.joca.2005.04.005.

Lämmerhofer, Michael, and Wolfram Weckwerth. *Metabolomics in Practice: Successful Strategies to Generate and Analyze Metabolic Data*. John Wiley & Sons, 2013. http://books.google.com/books?hl=en&lr=&id=bAmCNITLwGkC&oi=fnd&pg=PA2007&dq=%22Michael,+and+Weckwerth,+Wolfram,+eds.+Metabolomics+in+Practice+:+Successful+Strategies+to+Generate+and+Analyze+Metabolic+Data.+Somerset,+NJ,+USA:+John+Wiley+%26+Sons,+2013.+ProQuest+ebrary.+Web.+26+October%22+&ots=s-r0EfkcUF&sig=gz9ax1YH83yJGsvPSFlVK_OQyho.

Lander, J. J. "Auger Peaks in the Energy Spectra of Secondary Electrons from Various Materials." *Physical Review* 91, no. 6 (1953): 1382.

Lavine, Barry, and Jerome Workman. "Chemometrics." *Analytical Chemistry* 80, no. 12 (June 2008): 4519–31. doi:10.1021/ac800728t.

Ledford, Edward B., Chris A. Billesbach, and Quanyu Zhu. "GC3: Comprehensive Three-Dimensional Gas Chromatography." *Journal of High Resolution Chromatography* 23, no. 3 (March 1, 2000): 205–7. doi:10.1002/(SICI)1521-4168(20000301)23:3<205::AID-JHRC205>3.0.CO;2-U.

Lee, Andrew L., Keith D. Bartle, and Alastair C. Lewis. "A Model of Peak Amplitude Enhancement in Orthogonal Two-Dimensional Gas Chromatography." *Analytical Chemistry* 73, no. 6 (March 2001): 1330–35. doi:10.1021/ac001120s.

Lee, Ben H., Felipe D. Lopez-Hilfiker, Claudia Mohr, Theo Kurtén, Douglas R. Worsnop, and Joel A. Thornton. "An Iodide-Adduct High-Resolution Time-of-Flight Chemical-Ionization Mass Spectrometer: Application to Atmospheric Inorganic and Organic Compounds." *Environmental Science & Technology* 48, no. 11 (June 3, 2014): 6309–17. doi:10.1021/es500362a.

Link, Jeanne M., and Robert E. Synovec. “Whole-Column Radioactivity Detection: Simultaneous Separation and Enhanced Detectability.” *Analytical Chemistry* 71, no. 14 (July 1, 1999): 2700–2707. doi:10.1021/ac981401o.

Liu, Siyuan, and Joe M. Davis. “Dependence on Saturation of Average Minimum Resolution in Two-Dimensional Statistical-Overlap Theory: Peak Overlap in Saturated Two-Dimensional Separations.” *Journal of Chromatography A* 1126, no. 1–2 (September 2006): 244–56. doi:10.1016/j.chroma.2006.05.064.

Liu, Zaiyou, Donald G. Patterson Jr, and Milton L. Lee. “Geometric Approach to Factor Analysis for the Estimation of Orthogonality and Practical Peak Capacity in Comprehensive Two-Dimensional Separations.” *Analytical Chemistry* 67, no. 21 (1995): 3840–3845.

Liu, Z., and J. B. Phillips. “Comprehensive Two-Dimensional Gas Chromatography Using an On-Column Thermal Modulator Interface.” *Journal of Chromatographic Science* 29, no. 6 (June 1, 1991): 227–31. doi:10.1093/chromsci/29.6.227.

Lohmander, L. S., A. ﷺstenberg, M. Englund, and H. Roos. “High Prevalence of Knee Osteoarthritis, Pain, and Functional Limitations in Female Soccer Players Twelve Years after Anterior Cruciate Ligament Injury.” *Arthritis & Rheumatism* 50, no. 10 (October 2004): 3145–52. doi:10.1002/art.20589.

Long, Gary L., and James D. Winefordner. “Limit of Detection. A Closer Look at the IUPAC Definition.” *Analytical Chemistry* 55, no. 7 (1983): 712A–724A.

Lopez-Hilfiker, F. D., S. Iyer, C. Mohr, B. H. Lee, E. L. D’Ambro, T. Kurtén, and J. A. Thornton. “Constraining the Sensitivity of Iodide Adduct Chemical Ionization Mass Spectrometry to Multifunctional Organic Molecules Using the Collision Limit and Thermodynamic Stability of Iodide Ion Adducts.” *Atmospheric Measurement Techniques Discussions* 8, no. 10 (October 23, 2015): 10875–96. doi:10.5194/amtd-8-10875-2015.

Lopez-Hilfiker, F. D., C. Mohr, M. Ehn, F. Rubach, E. Kleist, J. Wildt, Th. F. Mentel, et al. “A Novel Method for Online Analysis of Gas and Particle Composition: Description and Evaluation of a Filter Inlet for Gases and AEROsols (FIGAERO).” *Atmospheric Measurement Techniques* 7, no. 4 (April 15, 2014): 983–1001. doi:10.5194/amt-7-983-2014.

Lu, Hongmei, Yizeng Liang, Warwick B. Dunn, Hailin Shen, and Douglas B. Kell. “Comparative Evaluation of Software for Deconvolution of Metabolomics Data Based on GC-TOF-MS.” *TrAC Trends in Analytical Chemistry*, Metabolomics, 27, no. 3 (March 2008): 215–27. doi:10.1016/j.trac.2007.11.004.

Luong, Jim, Ronda Gras, Robert Mustacich, and Hernan Cortes. “Low Thermal Mass Gas Chromatography: Principles and Applications.” *Journal of Chromatographic Science* 44, no. 5 (2006): 253–261.

Magagna, Federico, Lucia Valverde-Som, Cristina Ruíz-Samblás, Luis Cuadros-Rodríguez, Stephen E. Reichenbach, Carlo Bicchi, and Chiara Cordero. “Combined Untargeted and Targeted Fingerprinting with Comprehensive Two-Dimensional Chromatography for Volatiles and Ripening Indicators in Olive Oil.” *Analytica Chimica Acta* 936 (September 14, 2016): 245–58. doi:10.1016/j.aca.2016.07.005.

Marney, Luke C., W. Christopher Siegler, Brendon A. Parsons, Jamin C. Hoggard, Bob W. Wright, and Robert E. Synovec. “Tile-Based Fisher-Ratio Software for Improved Feature Selection Analysis of Comprehensive Two-Dimensional Gas Chromatography–time-of-

Flight Mass Spectrometry Data." *Talanta* 115 (October 2013): 887–95. doi:10.1016/j.talanta.2013.06.038.
Marney, Luke C., Stephen C. Kolwicz Jr., Rong Tian, and Robert E. Synovec. "Sample Preparation Methodology for Mouse Heart Metabolomics Using Comprehensive Two-Dimensional Gas Chromatography Coupled with Time-of-Flight Mass Spectrometry." *Talanta* 108 (April 15, 2013): 123–30. doi:10.1016/j.talanta.2013.03.005.
Marshall, Alan G. "Fourier Transform Ion Cyclotron Resonance Mass Spectrometry." *Accounts of Chemical Research* 18, no. 10 (1985): 316–322.
Marshall, Alan G., Christopher L. Hendrickson, and George S. Jackson. "Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer." *Mass Spectrometry Reviews* 17, no. 1 (1998): 1–35.
Massart, D.L., B.G.M. VanDeginste, S.N. Deming, Y. Michotte, and L. Kaufman. *Chemometrics: A Textbook*. Data Handling In Science and Technology 2. Amsterdam: Elsevier Academic Press, 1988.
McLafferty, Fred W. "A Century of Progress in Molecular Mass Spectrometry." *Annual Review of Analytical Chemistry* 4, no. 1 (July 19, 2011): 1–22. doi:10.1146/annurev-anchem-061010-114018.
McLafferty, Fred W., and Frantisek Turecek. *Interpretation of Mass Spectra*. 4th ed. Sausalito, CA: University Science Books, 1993.
Metallo, Christian M., Jason L. Walther, and Gregory Stephanopoulos. "Evaluation of 13C Isotopic Tracers for Metabolic Flux Analysis in Mammalian Cells." *Journal of Biotechnology* 144, no. 3 (November 2009): 167–74. doi:10.1016/j.jbiotec.2009.07.010.
Mi, Xiaoxia, Sicong Li, Yanhua Li, Kaiqiang Wang, Dan Zhu, and Gang Chen. "Quantitative Determination of 26 Steroids in Eggs from Various Species Using Liquid Chromatography–triple Quadrupole-Mass Spectrometry." *Journal of Chromatography A* 1356 (August 22, 2014): 54–63. doi:10.1016/j.chroma.2014.05.084.
Michels, David A., Shen Hu, K. Amy Dambrowitz, Michael J. Eggertson, Kevin Lauterbach, and Norman J. Dovichi. "Capillary Sieving Electrophoresis-Micellar Electrokinetic Chromatography Fully Automated Two-Dimensional Capillary Electrophoresis Analysis of Deinococcus Radiodurans Protein Homogenate." *ELECTROPHORESIS* 25, no. 18–19 (October 1, 2004): 3098–3105. doi:10.1002/elps.200405939.
Mitrevski, Blagoj, and Philip J. Marriott. "Evaluation of Quadrupole-Time-of-Flight Mass Spectrometry in Comprehensive Two-Dimensional Gas Chromatography." *Journal of Chromatography A* 1362 (October 3, 2014): 262–69. doi:10.1016/j.chroma.2014.08.053.
———. "Novel Hybrid Comprehensive 2D – Multidimensional Gas Chromatography for Precise, High-Resolution Characterization of Multicomponent Samples." *Analytical Chemistry* 84, no. 11 (June 5, 2012): 4837–43. doi:10.1021/ac300429y.
Mohler, Rachel E., Kenneth M. Dombek, Jamin C. Hoggard, Karisa M. Pierce, Elton T. Young, and Robert E. Synovec. "Comprehensive Analysis of Yeast Metabolite GC×GC–TOFMS Data: Combining Discovery-Mode and Deconvolution Chemometric Software." *Analyst* 132, no. 8 (July 23, 2007): 756–67. doi:10.1039/B700061H.
Mohler, Rachel E., Kenneth M. Dombek, Jamin C. Hoggard, Elton T. Young, and Robert E. Synovec. "Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells." *Analytical Chemistry* 78, no. 8 (April 1, 2006): 2700–2709. doi:10.1021/ac052106o.

Mohler, Rachel E., Benjamin P. Tu, Kenneth M. Dombek, Jamin C. Hoggard, Elton T. Young, and Robert E. Synovec. "Identification and Evaluation of Cycling Yeast Metabolites in Two-Dimensional Comprehensive Gas Chromatography–time-of-Flight-Mass Spectrometry Data." *Journal of Chromatography A*, Trends and Developments in Gas Chromatography, 1186, no. 1–2 (April 4, 2008): 401–11. doi:10.1016/j.chroma.2007.10.063.

Mondello, L. *Comprehensive Chromatography in Combination with Mass Spectrometry*. Wiley Series on Mass Spectrometry. Hoboken, NJ: John Wiley & Sons, Inc., 2011.

Montenegro-Burke, J. Rafael, Aries E. Aisporna, H. Paul Benton, Duane Rinehart, Mingliang Fang, Tao Huan, Benedikt Warth, et al. "Data Streaming for Metabolomics: Accelerating Data Processing and Analysis from Days to Minutes." *Analytical Chemistry* 89, no. 2 (January 17, 2017): 1254–59. doi:10.1021/acs.analchem.6b03890.

Moore Jr, Alvin W., and James W. Jorgenson. "Comprehensive Three-Dimensional Separation of Peptides Using Size Exclusion Chromatography/reversed Phase Liquid Chromatography/optically Gated Capillary Zone Electrophoresis." *Analytical Chemistry* 67, no. 19 (1995): 3456–3463.

Muddiman, David C., and Ann L. Oberg. "Statistical Evaluation of Internal and External Mass Calibration Laws Utilized in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry." *Analytical Chemistry* 77, no. 8 (April 2005): 2406–14. doi:10.1021/ac048258l.

Murphy, Kathleen R., Philip Wenig, Gavin Parcsi, Thomas Skov, and Richard M. Stuetz. "Characterizing Odorous Emissions Using New Software for Identifying Peaks in Chemometric Models of Gas Chromatography–mass Spectrometry Datasets." *Chemometrics and Intelligent Laboratory Systems* 118 (August 15, 2012): 41–50. doi:10.1016/j.chemolab.2012.07.006.

Murphy, Robert E., Mark R. Schure, and Joe P. Foley. "Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography." *Analytical Chemistry* 70, no. 8 (April 1, 1998): 1585–94. doi:10.1021/ac971184b.

Nadeau, Jeremy S., Ryan B. Wilson, Brian D. Fitz, Jason T. Reed, and Robert E. Synovec. "Utilizing a Constant Peak Width Transform for Isothermal Gas Chromatography." *Journal of Chromatography A* 1218, no. 23 (June 2011): 3718–24. doi:10.1016/j.chroma.2011.04.007.

Nadeau, Jeremy S., Ryan B. Wilson, Jamin C. Hoggard, Bob W. Wright, and Robert E. Synovec. "Study of the Interdependency of the Data Sampling Ratio with Retention Time Alignment and Principal Component Analysis for Gas Chromatography." *Journal of Chromatography A* 1218, no. 50 (December 16, 2011): 9091–9101. doi:10.1016/j.chroma.2011.10.031.

Nadeau, Jeremy S., Bob W. Wright, and Robert E. Synovec. "Chemometric Analysis of Gas Chromatography–mass Spectrometry Data Using Fast Retention Time Alignment via a Total Ion Current Shift Function." *Talanta* 81, no. 1–2 (April 15, 2010): 120–28. doi:10.1016/j.talanta.2009.11.046.

Navarro-Reig, Meritxell, Joaquim Jaumot, Teris A. van Beek, Gabriel Vivó-Truyols, and Romà Tauler. "Chemometric Analysis of Comprehensive LC×LC-MS Data: Resolution of Triacylglycerol Structural Isomers in Corn Oil." *Talanta* 160 (November 1, 2016): 624–35. doi:10.1016/j.talanta.2016.08.005.

Niu, Weihuan, Elisa Knight, Qingyou Xia, and Brian D. McGarvey. "Comparative Evaluation of Eight Software Programs for Alignment of Gas Chromatography–mass Spectrometry

Chromatograms in Metabolomics Experiments." *Journal of Chromatography A* 1374 (December 29, 2014): 199–206. doi:10.1016/j.chroma.2014.11.005.
Nolvachai, Yada, Chadin Kulsing, and Philip J. Marriott. "In Silico Modeling of Hundred Thousand Experiments for Effective Selection of Ionic Liquid Phase Combinations in Comprehensive Two-Dimensional Gas Chromatography." *Analytical Chemistry*, January 17, 2016. doi:10.1021/acs.analchem.5b03688.
Norrby, Lars J. "Why Is Mercury Liquid? Or, Why Do Relativistic Effects Not Get into Chemistry Textbooks?" *Journal of Chemical Education* 68, no. 2 (February 1, 1991): 110. doi:10.1021/ed068p110.
O'Callaghan, Sean, David P. De Souza, Andrew Isaac, Qiao Wang, Luke Hodkinson, Moshe Olshansky, Tim Erwin, et al. "PyMS: A Python Toolkit for Processing of Gas Chromatography-Mass Spectrometry (GC-MS) Data. Application and Comparative Study of Selected Tools." *BMC Bioinformatics* 13, no. 1 (May 30, 2012): 115. doi:10.1186/1471-2105-13-115.
Olivier, Ilse, and Du Toit Loots. "A Metabolomics Approach to Characterise and Identify Various Mycobacterium Species." *Journal of Microbiological Methods* 88, no. 3 (March 2012): 419–26. doi:10.1016/j.mimet.2012.01.012.
Olver, Frank W. J., and National Institute of Standards and Technology (U.S.), eds. *NIST Handbook of Mathematical Functions*. Cambridge ; New York: Cambridge University Press : NIST, 2010.
Omais, Badaoui, Marion Courtiade, Nadège Charon, Jérémie Ponthus, and Didier Thiébaut. "Considerations on Orthogonality Duality in Comprehensive Two-Dimensional Gas Chromatography." *Analytical Chemistry* 83, no. 19 (October 2011): 7550–54. doi:10.1021/ac201103e.
Pacakova, Vera, and Ladislav Feltl. *Chromatographic Retention Indices: An Aid to Identification of Organic Compounds*. Ellis Horwood Series in Analytical Chemistry. New York, NY: Ellis Horwood, 1992.
Parastar, Hadi, and Nadia Akvan. "Multivariate Curve Resolution Based Chromatographic Peak Alignment Combined with Parallel Factor Analysis to Exploit Second-Order Advantage in Complex Chromatographic Measurements." *Analytica Chimica Acta* 816 (March 13, 2014): 18–27. doi:10.1016/j.aca.2014.01.051.
Parastar, Hadi, Jagoš R. Radović, Mehdi Jalali-Heravi, Sergi Diez, Josep Maria Bayona, and Roma Tauler. "Resolution and Quantification of Complex Mixtures of Polycyclic Aromatic Hydrocarbons in Heavy Fuel Oil Sample by Means of GC × GC-TOFMS Combined to Multivariate Curve Resolution." *Analytical Chemistry* 83, no. 24 (December 15, 2011): 9289–97. doi:10.1021/ac201799r.
Parsons, Brendon A., Luke C. Marney, W. Christopher Siegler, Jamin C. Hoggard, Bob W. Wright, and Robert E. Synovec. "Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach." *Analytical Chemistry* 87, no. 7 (April 7, 2015): 3812–19. doi:10.1021/ac504472s.
Parsons, Brendon A., David K. Pinkerton, Bob W. Wright, and Robert E. Synovec. "Chemical Characterization of the Acid Alteration of Diesel Fuel: Non-Targeted Analysis by Two-Dimensional Gas Chromatography Coupled with Time-of-Flight Mass Spectrometry with Tile-Based Fisher Ratio and Combinatorial Threshold Determination." *Journal of Chromatography A* 1440 (April 1, 2016): 179–90. doi:10.1016/j.chroma.2016.02.067.

Passey, Benjamin H. “Biogeochemical Tales Told by Isotope Clumps.” *Science* 348, no. 6233 (April 24, 2015): 394–95. doi:10.1126/science.aab1604.
Pauling, Linus. “The Nature of the Chemical Bond. Application of Results Obtained from the Quantum Mechanics and from a Theory of Paramagnetic Susceptibility to the Structure of Molecules.” *Journal of the American Chemical Society* 53, no. 4 (1931): 1367–1400.
Payagala, Tharanga, Ying Zhang, Eranda Wanigasekara, Ke Huang, Zachary S. Breitbach, Pritesh S. Sharma, Leonard M. Sidisky, and Daniel W. Armstrong. “Trigonal Tricationic Ionic Liquids: A Generation of Gas Chromatographic Stationary Phases.” *Analytical Chemistry* 81, no. 1 (January 1, 2009): 160–73. doi:10.1021/ac8016949.
Peppe, Daniel J., and Dana L. Royer. “Can Climate Feel the Pressure?” *Science* 348, no. 6240 (June 12, 2015): 1210–11. doi:10.1126/science.aac5264.
Phelan, M. Kathleen, Clyde H. Barlow, Jeffrey J. Kelly, Thomas M. Jinguji, and James B. Callis. “Measurement of Caustic and Caustic Brine Solutions by Spectroscopic Detection of the Hydroxide Ion in the near-Infrared Region, 700-1150 Nm.” *Analytical Chemistry* 61, no. 13 (1989): 1419–1424.
Pierce, Karisa M., and Jamin C. Hoggard. “Chromatographic Data Analysis. Part 3.3.4: Handling Hyphenated Data in Chromatography.” *Analytical Methods* 6, no. 3 (2014): 645–53. doi:10.1039/c3ay40965a.
Pierce, Karisa M., Jamin C. Hoggard, Janiece L. Hope, Petrie M. Rainey, Andrew N. Hoofnagle, Rhona M. Jack, Bob W. Wright, and Robert E. Synovec. “Fisher Ratio Method Applied to Third-Order Separation Data to Identify Significant Chemical Components of Metabolite Extracts.” *Analytical Chemistry* 78, no. 14 (July 2006): 5068–75. doi:10.1021/ac0602625.
Pierce, Karisa M., Jamin C. Hoggard, Rachel E. Mohler, and Robert E. Synovec. “Recent Advancements in Comprehensive Two-Dimensional Separations with Chemometrics.” *Journal of Chromatography A* 1184, no. 1–2 (March 2008): 341–52. doi:10.1016/j.chroma.2007.07.059.
Pierce, Karisa M., Janiece L. Hope, Jamin C. Hoggard, and Robert E. Synovec. “A Principal Component Analysis Based Method to Discover Chemical Differences in Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry (GC × GC-TOFMS) Separations of Metabolites in Plant Samples.” *Talanta*, A collection of Papers Presented at the 1st Workshop of the European Union: Analysis and Removal of Contaminants from Wastewaters for the Implementation of the Water Framework Directive1st EMCO 2005, 70, no. 4 (November 15, 2006): 797–804. doi:10.1016/j.talanta.2006.01.038.
Pierce, Karisa M., Janiece L. Hope, Kevin J. Johnson, Bob W. Wright, and Robert E. Synovec. “Classification of Gasoline Data Obtained by Gas Chromatography Using a Piecewise Alignment Algorithm Combined with Feature Selection and Principal Component Analysis.” *Journal of Chromatography A*, Chemical Separations and Chemometrics, 1096, no. 1–2 (November 25, 2005): 101–10. doi:10.1016/j.chroma.2005.04.078.
Pierce, Karisa M., Benjamin Kehimkar, Luke C. Marney, Jamin C. Hoggard, and Robert E. Synovec. “Review of Chemometric Analysis Techniques for Comprehensive Two Dimensional Separations Data.” *Journal of Chromatography A* 1255 (September 2012): 3–11. doi:10.1016/j.chroma.2012.05.050.
Pierce, Karisa M., Lianna F. Wood, Bob W. Wright, and Robert E. Synovec. “A Comprehensive Two-Dimensional Retention Time Alignment Algorithm to Enhance Chemometric

Analysis of Comprehensive Two-Dimensional Separation Data." *Analytical Chemistry* 77, no. 23 (December 2005): 7735–43. doi:10.1021/ac0511142.

Pinkerton, David K., Brendon A. Parsons, Todd J. Anderson, and Robert E. Synovec. "Trilinearity Deviation Ratio: A New Metric for Chemometric Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data." *Analytica Chimica Acta* 871 (April 29, 2015): 66–76. doi:10.1016/j.aca.2015.02.040.

Pinkerton, David K., Brendon A. Parsons, and Robert E. Synovec. "Method to Determine the True Modulation Ratio for Comprehensive Two-Dimensional Gas Chromatography." *Journal of Chromatography A* 1476 (December 9, 2016): 114–23. doi:10.1016/j.chroma.2016.11.015.

Plutzer, Eric, Mark McCaffrey, A. Lee Hannah, Joshua Rosenau, Minda Berbeco, and Ann H. Reid. "Climate Confusion among U.S. Teachers." *Science* 351, no. 6274 (February 12, 2016): 664–65. doi:10.1126/science.aab3907.

Poe, Russel B., and Sarah C. Rutan. "Effects of Resolution, Peak Ratio and Sampling Frequency in Diode-Array Fluorescence Detection in Liquid Chromatography." *Analytica Chimica Acta* 283, no. 2 (November 26, 1993): 845–53. doi:10.1016/0003-2670(93)85298-X.

Poole, Colin F. "Ionization-Based Detectors for Gas Chromatography." *Journal of Chromatography A*. Accessed May 26, 2015. doi:10.1016/j.chroma.2015.02.061.

Porter, Sarah E. G., Dwight R. Stoll, Sarah C. Rutan, Peter W. Carr, and Jerry D. Cohen. "Analysis of Four-Way Two-Dimensional Liquid Chromatography-Diode Array Data: Application to Metabolomics." *Analytical Chemistry* 78, no. 15 (August 2006): 5559–69. doi:10.1021/ac0606195.

Poulsen, Christopher J., Clay Tabor, and Joseph D. White. "Long-Term Climate Forcing by Atmospheric Oxygen Concentrations." *Science* 348, no. 6240 (June 12, 2015): 1238–41. doi:10.1126/science.1260670.

Prazen, Bryan J., Carsten A. Bruckner, Robert E. Synovec, and Bruce R. Kowalski. "Second-Order Chemometric Standardization for High-Speed Hyphenated Gas Chromatography: Analysis of GC/MS and Comprehensive GC×GC Data." *Journal of Microcolumn Separations* 11, no. 2 (January 1, 1999): 97–107. doi:10.1002/(SICI)1520-667X(1999)11:2<97::AID-MCS2>3.0.CO;2-Z.

Prazen, Bryan J., Robert E. Synovec, and Bruce R. Kowalski. "Standardization of Second-Order Chromatographic/spectroscopic Data for Optimum Chemical Analysis." *Analytical Chemistry* 70, no. 2 (1998): 218–225.

Prebihalo, Sarah, Adrienne Brockman, Jack Cochran, and Frank L. Dorman. "Determination of Emerging Contaminants in Wastewater Utilizing Comprehensive Two-Dimensional Gas-Chromatography Coupled with Time-of-Flight Mass Spectrometry." *Journal of Chromatography A* 1419 (November 6, 2015): 109–15. doi:10.1016/j.chroma.2015.09.080.

Pronschinske, Alex, Philipp Pedevilla, Colin J. Murphy, Emily A. Lewis, Felicia R. Lucci, Garth Brown, George Pappas, Angelos Michaelides, and E. Charles H. Sykes. "Enhancement of Low-Energy Electron Emission in 2D Radioactive Films." *Nature Materials* advance online publication (June 15, 2015). doi:10.1038/nmat4323.

Psychogios, Nikolaos, David D. Hau, Jun Peng, An Chi Guo, Rupasri Mandal, Souhaila Bouatra, Igor Sinelnikov, et al. "The Human Serum Metabolome." *PLOS ONE* 6, no. 2 (February 16, 2011): e16957. doi:10.1371/journal.pone.0016957.

Purcell, Jeremiah M., Christopher L. Hendrickson, Ryan P. Rodgers, and Alan G. Marshall. "Atmospheric Pressure Photoionization Fourier Transform Ion Cyclotron Resonance Mass

Spectrometry for Complex Mixture Analysis." *Analytical Chemistry* 78, no. 16 (August 2006): 5906–12. doi:10.1021/ac060754h.

Qian, Rong, Yuan -Xi Liao, Yin -Long Guo, and Hao Guo. "ESI-FTICR-MS Studies on Gas Phase Fragmentation Reactions of ArPd(PPh3)2I Complexes." *Journal of the American Society for Mass Spectrometry* 17, no. 11 (November 2006): 1582–89. doi:10.1016/j.jasms.2006.07.009.

Quigley, Wes W. C., Carlos G. Fraga, and Robert E. Synovec. "Comprehensive LC×GC for Enhanced Headspace Analysis." *Journal of Microcolumn Separations* 12, no. 3 (January 1, 2000): 160–66. doi:10.1002/(SICI)1520-667X(2000)12:3<160::AID-MCS5>3.0.CO;2-8.

Rainville, S. "An Ion Balance for Ultra-High-Precision Atomic Mass Measurements." *Science* 303, no. 5656 (January 16, 2004): 334–38. doi:10.1126/science.1092320.

Reaser, Brooke C., Song Yang, Brian D. Fitz, Brendon A. Parsons, Mary E. Lidstrom, and Robert E. Synovec. "Non-Targeted Determination of 13C-Labeling in the Methylobacterium Extorquens AM1 Metabolome Using the Two-Dimensional Mass Cluster Method and Principal Component Analysis." *Journal of Chromatography A* 1432 (February 5, 2016): 111–21. doi:10.1016/j.chroma.2015.12.088.

Reichenbach, Stephen E., Mingtian Ni, Visweswara Kottapalli, and Arvind Visvanathan. "Information Technologies for Comprehensive Two-Dimensional Gas Chromatography." *Chemometrics and Intelligent Laboratory Systems* 71, no. 2 (May 28, 2004): 107–20. doi:10.1016/j.chemolab.2003.12.009.

Reichenbach, Stephen E., Xue Tian, Chiara Cordero, and Qingping Tao. "Features for Non-Targeted Cross-Sample Analysis with Comprehensive Two-Dimensional Chromatography." *Journal of Chromatography A* 1226 (February 24, 2012): 140–48. doi:10.1016/j.chroma.2011.07.046.

Reid, Vanessa R., Jeffery A. Crank, Daniel W. Armstrong, and Robert E. Synovec. "Characterization and Utilization of a Novel Triflate Ionic Liquid Stationary Phase for Use in Comprehensive Two-Dimensional Gas Chromatography." *Journal of Separation Science* 31, no. 19 (October 2008): 3429–36. doi:10.1002/jssc.200800251.

Reid, V, and R Synovec. "High-Speed Gas Chromatography: The Importance of Instrumentation Optimization and the Elimination of Extra-Column Band Broadening." *Talanta* 76, no. 4 (August 15, 2008): 703–17. doi:10.1016/j.talanta.2008.05.012.

Reshef, D. N., Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti. "Detecting Novel Associations in Large Data Sets." *Science* 334, no. 6062 (December 16, 2011): 1518–24. doi:10.1126/science.1205438.

Robards, Kevin, Paul Haddad, and Peter Jackson. *Principles and Practice of Modern Chromatographic Methods*. Amsterdam: Elsevier Academic Press, 2004.

Rontani, Jean-François, and Claude Aubert. "Hydrogen and Trimethylsilyl Transfers during EI Mass Spectral Fragmentation of Hydroxycarboxylic and Oxocarboxylic Acid Trimethylsilyl Derivatives." *Journal of the American Society for Mass Spectrometry* 19, no. 1 (January 2008): 66–75. doi:10.1016/j.jasms.2007.10.014.

Rubio, L., L.A. Sarabia, and M.C. Ortiz. "Standard Addition Method Based on Four-Way PARAFAC Decomposition to Solve the Matrix Interferences in the Determination of Carbamate Pesticides in Lettuce Using Excitation–emission Fluorescence Data." *Talanta* 138 (June 2015): 86–99. doi:10.1016/j.talanta.2015.01.042.

Ruckebusch, Cyril, ed. *Resolving Spectral Mixtures: With Applications from Ultrafast Time-Resolved Spectroscopy to Super-Resolution Imaging*. First edition. Data Handling in Science and Technology, Volume 30. Amsterdam: Elsevier, 2016.

Ryan, Danielle, Robert Shellie, Peter Tranchida, Alessandro Casilli, Luigi Mondello, and Philip Marriott. "Analysis of Roasted Coffee Bean Volatiles by Using Comprehensive Two-Dimensional Gas Chromatography–time-of-Flight Mass Spectrometry." *Journal of Chromatography A*, Food Science, 1054, no. 1–2 (October 29, 2004): 57–65. doi:10.1016/j.chroma.2004.08.057.

Sampson, Jason S., Adam M. Hawkridge, and David C. Muddiman. "Generation and Detection of Multiply-Charged Peptides and Proteins by Matrix-Assisted Laser Desorption Electrospray Ionization (MALDESI) Fourier Transform Ion Cyclotron Resonance Mass Spectrometry." *Journal of the American Society for Mass Spectrometry* 17, no. 12 (December 2006): 1712–16. doi:10.1016/j.jasms.2006.08.003.

Savareear, Benjamin, Matthew R. Jacobs, and Robert A. Shellie. "Multiplexed Dual First-Dimension Comprehensive Two-Dimensional Gas Chromatography–mass Spectrometry with Contra-Directional Thermal Modulation." *Journal of Chromatography A* 1365 (October 24, 2014): 183–90. doi:10.1016/j.chroma.2014.09.014.

Savareear, Benjamin, and Robert A. Shellie. "Multiplexed Dual Second-Dimension Column Comprehensive Two-Dimensional Gas Chromatography (GC × 2GC) Using Thermal Modulation and Contra-Directional Second-Dimension Columns." *Analytica Chimica Acta*, 20th Research and Development Topics Meeting in Analytical and Environmental Chemistry, 803 (November 25, 2013): 160–65. doi:10.1016/j.aca.2013.05.013.

Savitzky, Abraham, and Marcel JE Golay. "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." *Analytical Chemistry* 36, no. 8 (1964): 1627–1639.

S. Yang, J.C. Hoggard, M.E. Lidstrom, and R.E. Synovec. "Gas Chromatography and Comprehensive Two-Dimensional Gas Chromatography Hyphenated with Mass Spectrometry for Targeted and Nontargeted Metabolomics." In *Metabolomics in Practice: Successful Strategies to Generate and Analyze Metabolic Data*, 69–92. Weinheim, Germany: Wiley-VCH, 2013.

Scheffe, Henry. *The Analysis of Variance*. New York, NY: John Wiley & Sons, Inc., 1959.

Schmarr, Hans-Georg, and Jörg Bernhardt. "Profiling Analysis of Volatile Compounds from Fruits Using Comprehensive Two-Dimensional Gas Chromatography and Image Processing Techniques." *Journal of Chromatography A* 1217, no. 4 (January 22, 2010): 565–74. doi:10.1016/j.chroma.2009.11.063.

Schoenmakers, Peter, Philip Marriott, and Jan Beens. "Nomenclature and Conventions in Comprehensive Multidimensional Chromatography." *LC-GC Europe* 16 (June 2004): 335–39.

Scott, Bobby R., and Ludwik Dobrzyński. "Special Issue Introduction." *Dose-Response* 10, no. 4 (January 1, 2012): 462–66. doi:10.2203/dose-response.12-047.Scott.

Seeley, John V. "Theoretical Study of Incomplete Sampling of the First Dimension in Comprehensive Two-Dimensional Chromatography." *Journal of Chromatography A* 962, no. 1–2 (July 2002): 21–27. doi:10.1016/S0021-9673(02)00461-2.

Seeley, John V., Frederick Kramp, and Christine J. Hicks. "Comprehensive Two-Dimensional Gas Chromatography via Differential Flow Modulation." *Analytical Chemistry* 72, no. 18 (September 2000): 4346–52. doi:10.1021/ac000249z.

Seeley, John V., Nicole J. Micyus, Steven V. Bandurski, Stacy K. Seeley, and James D. McCurry. “Microfluidic Deans Switch for Comprehensive Two-Dimensional Gas Chromatography.” *Analytical Chemistry* 79, no. 5 (March 2007): 1840–47. doi:10.1021/ac061881g.

Seeley, John V., and Stacy K. Seeley. “Multidimensional Gas Chromatography: Fundamental Advances and New Applications.” *Analytical Chemistry* 85, no. 2 (January 15, 2013): 557–78. doi:10.1021/ac303195u.

Sgorbini, Barbara, Carlo Bicchi, Cecilia Cagliero, Chiara Cordero, Erica Liberto, and Patrizia Rubiolo. “Herbs and Spices: Characterization and Quantitation of Biologically-Active Markers for Routine Quality Control by Multiple Headspace Solid-Phase Microextraction Combined with Separative or Non-Separative Analysis.” *Journal of Chromatography A* 1376 (January 9, 2015): 9–17. doi:10.1016/j.chroma.2014.12.007.

Shellie, Robert A., Werner Welthagen, Jitka Zrostliková, Joachim Spranger, Michael Ristow, Oliver Fiehn, and Ralf Zimmermann. “Statistical Methods for Comparing Comprehensive Two-Dimensional Gas Chromatography–time-of-Flight Mass Spectrometry Results: Metabolomic Analysis of Mouse Tissue Extracts.” *Journal of Chromatography A*, 2nd International Symposium on Comprehensive Multidimensional Gas Chromatography2nd International Symposium on Comprehensive Multidimensional Gas Chromatography, 1086, no. 1–2 (September 9, 2005): 83–90. doi:10.1016/j.chroma.2005.05.088.

Siegler, W. Christopher, Jeffery A. Crank, Daniel W. Armstrong, and Robert E. Synovec. “Increasing Selectivity in Comprehensive Three-Dimensional Gas Chromatography via an Ionic Liquid Stationary Phase Column in One Dimension.” *Journal of Chromatography A* 1217, no. 18 (April 2010): 3144–49. doi:10.1016/j.chroma.2010.02.082.

Siegler, W. Christopher, Brian D. Fitz, Jamin C. Hoggard, and Robert E. Synovec. “Experimental Study of the Quantitative Precision for Valve-Based Comprehensive Two-Dimensional Gas Chromatography.” *Analytical Chemistry* 83, no. 13 (July 2011): 5190–96. doi:10.1021/ac200302b.

Sinanian, Melanie M., Daniel W. Cook, Sarah C. Rutan, and Dayanjan S. Wijesinghe. “Multivariate Curve Resolution-Alternating Least Squares Analysis of High-Resolution Liquid Chromatography–Mass Spectrometry Data.” *Analytical Chemistry* 88, no. 22 (November 15, 2016): 11092–99. doi:10.1021/acs.analchem.6b03116.

Sinha, Amanda E, Carlos G Fraga, Bryan J Prazen, and Robert E Synovec. “Trilinear Chemometric Analysis of Two-Dimensional Comprehensive Gas Chromatography–time-of-Flight Mass Spectrometry Data.” *Journal of Chromatography A*, 26th International Symposium on Capillary Chromatography and Electrophoresis, 1027, no. 1–2 (February 20, 2004): 269–77. doi:10.1016/j.chroma.2003.08.081.

Sinha, Amanda E., Janiece L. Hope, Bryan J. Prazen, Carlos G. Fraga, Erik J. Nilsson, and Robert E. Synovec. “Multivariate Selectivity as a Metric for Evaluating Comprehensive Two-Dimensional Gas Chromatography–time-of-Flight Mass Spectrometry Subjected to Chemometric Peak Deconvolution.” *Journal of Chromatography A*, 8th International Symposium on Hyphenated Techniques in Chromatography and Hyphenated Chromatographic Analyzers, 1056, no. 1–2 (November 12, 2004): 145–54. doi:10.1016/j.chroma.2004.06.110.

Sinha, Amanda E., Janiece L. Hope, Bryan J. Prazen, Erik J. Nilsson, Rhona M. Jack, and Robert E. Synovec. “Algorithm for Locating Analytes of Interest Based on Mass Spectral Similarity in GC × GC–TOF-MS Data: Analysis of Metabolites in Human Infant Urine.”

*Journal of Chromatography A*, Mass Spectrometry: Innovation and Application. Part III, 1058, no. 1–2 (November 26, 2004): 209–15. doi:10.1016/j.chroma.2004.08.064.

Sinha, Amanda E, Kevin J Johnson, Bryan J Prazen, Samuel V Lucas, Carlos G Fraga, and Robert E Synovec. "Comprehensive Two-Dimensional Gas Chromatography of Volatile and Semi-Volatile Components Using a Diaphragm Valve-Based Instrument." *Journal of Chromatography A* 983, no. 1–2 (January 2003): 195–204. doi:10.1016/S0021-9673(02)01651-5.

Sinha, Amanda E., Bryan J. Prazen, Carlos G. Fraga, and Robert E. Synovec. "Valve-Based Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometric Detection: Instrumentation and Figures-of-Merit." *Journal of Chromatography A* 1019, no. 1–2 (November 2003): 79–87. doi:10.1016/j.chroma.2003.08.047.

Sinkov, Nikolai A., and James J. Harynuk. "Three-Dimensional Cluster Resolution for Guiding Automatic Chemometric Model Optimization." *Talanta* 103 (January 15, 2013): 252–59. doi:10.1016/j.talanta.2012.10.040.

Sinkov, Nikolai A., Brandon M. Johnston, P. Mark L. Sandercock, and James J. Harynuk. "Automated Optimization and Construction of Chemometric Models Based on Highly Variable Raw Chromatographic Data." *Analytica Chimica Acta* 697, no. 1–2 (July 4, 2011): 8–15. doi:10.1016/j.aca.2011.04.029.

Skov, Thomas, Jamin C. Hoggard, Rasmus Bro, and Robert E. Synovec. "Handling within Run Retention Time Shifts in Two-Dimensional Chromatography Data Using Shift Correction and Modeling." *Journal of Chromatography A* 1216, no. 18 (May 1, 2009): 4020–29. doi:10.1016/j.chroma.2009.02.049.

Skrobot, Vinicius L., Eustáquio V. R. Castro, Rita C. C. Pereira, Vânya M. D. Pasa, and Isabel C. P. Fortes. "Identification of Adulteration of Gasoline Applying Multivariate Data Analysis Techniques HCA and KNN in Chromatographic Data." *Energy & Fuels* 19, no. 6 (November 1, 2005): 2350–56. doi:10.1021/ef050031l.

Slonecker, Patrick J., Xiaodong Li, Thomas H. Ridgway, and John G. Dorsey. "Informational Orthogonality of Two-Dimensional Chromatographic Separations." *Analytical Chemistry* 68, no. 4 (1996): 682–689.

Snijders, Henri, Hans-Gerd Janssen, and Carel Cramers. "Optimization of Temperature-Programmed Gas Chromatographic Separations I. Prediction of Retention Times and Peak Widths from Retention Indices." *Journal of Chromatography A* 718, no. 2 (December 22, 1995): 339–55. doi:10.1016/0021-9673(95)00692-3.

———. "Optimization of Temperature-Programmed Gas Chromatographic Separations II. Off-Line Simplex Optimization and Column Selection." *Journal of Chromatography A* 756, no. 1 (December 20, 1996): 175–83. doi:10.1016/S0021-9673(96)00626-7.

Snyder, L. R_, and H. J. Adler. "Dispersion in Segmented Flow through Glass Tubing in Continuous-Flow Analysis: The Nonideal Model." *Analytical Chemistry* 48, no. 7 (1976): 1022–1027.

Snyder, Laura R., Jamin C. Hoggard, Thomas J. Montine, and Robert E. Synovec. "Development and Application of a Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry Method for the Analysis of L-β-Methylamino-Alanine in Human Tissue." *Journal of Chromatography A* 1217, no. 27 (July 2, 2010): 4639–47. doi:10.1016/j.chroma.2010.04.065.

Song, Shin Miin, Philip Marriott, Alex Kotsos, Olaf H Drummer, and Paul Wynne. "Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry (GC × GC-TOFMS) for Drug Screening and Confirmation." *Forensic Science International* 143, no. 2–3 (July 2004): 87–101. doi:10.1016/j.forsciint.2004.02.042.

Stearns, Stanley D., Huamin Cai, J. Art Koehn, Martin Brisbin, Chris Cowles, Chris Bishop, Santos Puente, and Dale Ashworth. "A Direct Resistively Heated Gas Chromatography Column with Heating and Sensing on the Same Nickel Element." *Journal of Chromatography A* 1217, no. 27 (July 2010): 4629–38. doi:10.1016/j.chroma.2010.04.050.

Stefanuto, P.-H., K. A. Perrault, R. M. Lloyd, B. Stuart, T. Rai, S. L. Forbes, and J.-F. Focant. "Exploring New Dimensions in Cadaveric Decomposition Odour Analysis." *Analytical Methods* 7, no. 6 (March 12, 2015): 2287–94. doi:10.1039/C5AY00371G.

Stefanuto, Pierre-Hugues, Katelynn A. Perrault, Sonja Stadler, Romain Pesesse, Helene N. LeBlanc, Shari L. Forbes, and Jean-François Focant. "GC × GC–TOFMS and Supervised Multivariate Approaches to Study Human Cadaveric Decomposition Olfactive Signatures." *Analytical and Bioanalytical Chemistry* 407, no. 16 (April 25, 2015): 4767–78. doi:10.1007/s00216-015-8683-5.

Stein, S. E. "An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/mass Spectrometry Data." *Journal of the American Society for Mass Spectrometry* 10, no. 8 (August 1999): 770–81. doi:10.1016/S1044-0305(99)00047-1.

Stein, Stephen E. "Estimating Probabilities of Correct Identification from Results of Mass Spectral Library Searches." *Journal of the American Society for Mass Spectrometry* 5, no. 4 (April 1, 1994): 316–23. doi:10.1016/1044-0305(94)85022-4.

Stein, Stephen E., and David N. Heller. "On the Risk of False Positive Identification Using Multiple Ion Monitoring in Qualitative Mass Spectrometry: Large-Scale Intercomparisons with a Comprehensive Mass Spectral Library." *Journal of the American Society for Mass Spectrometry* 17, no. 6 (June 2006): 823–35. doi:10.1016/j.jasms.2006.02.021.

Stein, Stephen E., and Donald R. Scott. "Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification." *Journal of the American Society for Mass Spectrometry* 5, no. 9 (September 1994): 859–66. doi:10.1016/1044-0305(94)87009-8.

Stern, S. A., F. Bagenal, K. Ennico, G. R. Gladstone, W. M. Grundy, W. B. McKinnon, J. M. Moore, et al. "The Pluto System: Initial Results from Its Exploration by New Horizons." *Science* 350, no. 6258 (October 16, 2015): aad1815. doi:10.1126/science.aad1815.

Stoll, Dwight R., and Peter W. Carr. "Fast, Comprehensive Two-Dimensional HPLC Separation of Tryptic Peptides Based on High-Temperature HPLC." *Journal of the American Chemical Society* 127, no. 14 (April 2005): 5034–35. doi:10.1021/ja050145b.

Strittmatter, Eric F., P. Lee Ferguson, Keqi Tang, and Richard D. Smith. "Proteome Analyses Using Accurate Mass and Elution Time Peptide Tags with Capillary LC Time-of-Flight Mass Spectrometry." *Journal of the American Society for Mass Spectrometry* 14, no. 9 (September 2003): 980–91. doi:10.1016/S1044-0305(03)00146-6.

Synovec, Robert E., and Edward S. Yeung. "Improvement of the Limit of Detection in Chromatography by an Integration Method." *Analytical Chemistry* 57, no. 12 (October 1, 1985): 2162–67. doi:10.1021/ac00289a001.

Taylor, Jonathan, and Robert J. Tibshirani. “Statistical Learning and Selective Inference.” *Proceedings of the National Academy of Sciences* 112, no. 25 (June 23, 2015): 7629–34. doi:10.1073/pnas.1507583112.

Tomasi, Giorgio, Frans van den Berg, and Claus Andersson. “Correlation Optimized Warping and Dynamic Time Warping as Preprocessing Methods for Chromatographic Data.” *Journal of Chemometrics* 18, no. 5 (May 1, 2004): 231–41. doi:10.1002/cem.859.

Tomasi, Giorgio, and Rasmus Bro. “A Comparison of Algorithms for Fitting the PARAFAC Model.” *Computational Statistics & Data Analysis* 50, no. 7 (April 1, 2006): 1700–1734. doi:10.1016/j.csda.2004.11.013.

Tondeur, Yves, Phillip W. Albro, J. Ronald Hass, Donald J. Harvan, and Joanna L. Schroeder. “Matrix Effect in Determination of 2, 3, 7, 8-Tetrachlorodibenzodioxin by Mass Spectrometry.” *Analytical Chemistry* 56, no. 8 (1984): 1344–1347.

Tranchida, Peter Quinto, and Luigi Mondello. “Current-Day Employment of the Micro-Bore Open-Tubular Capillary Column in the Gas Chromatography Field.” *Journal of Chromatography A*, High Speed and High Resolution Separations. In Honour of Professor Milton Lee on the Occasion of his 65th Birthday, 1261 (October 26, 2012): 23–36. doi:10.1016/j.chroma.2012.05.074.

Tu, B. P. “Logic of the Yeast Metabolic Cycle: Temporal Compartmentalization of Cellular Processes.” *Science* 310, no. 5751 (November 18, 2005): 1152–58. doi:10.1126/science.1120499.

Tu, Benjamin P., Rachel E. Mohler, Jessica C. Liu, Kenneth M. Dombek, Elton T. Young, Robert E. Synovec, and Steven L. McKnight. “Cyclic Changes in Metabolic State during the Life of a Yeast Cell.” *Proceedings of the National Academy of Sciences* 104, no. 43 (2007): 16886–16891.

Uarrota, Virgílio Gavicho, Rodolfo Moresco, Bianca Coelho, Eduardo da Costa Nunes, Luiz Augusto Martins Peruch, Enilto de Oliveira Neubert, Miguel Rocha, and Marcelo Maraschin. “Metabolomics Combined with Chemometric Tools (PCA, HCA, PLS-DA and SVM) for Screening Cassava (Manihot Esculenta Crantz) Roots during Postharvest Physiological Deterioration.” *Food Chemistry* 161 (October 15, 2014): 67–78. doi:10.1016/j.foodchem.2014.03.110.

Vasquez, N. Pérez, M. Crosnier de bellaistre-Bonose, N. Lévêque, E. Thioulouse, D. Doummar, T. Billette de Villemeur, D. Rodriguez, et al. “Advances in the Metabolic Profiling of Acidic Compounds in Children’s Urines Achieved by Comprehensive Two-Dimensional Gas Chromatography.” *Journal of Chromatography B* 1002 (October 1, 2015): 130–38. doi:10.1016/j.jchromb.2015.08.006.

Vis, Daniel J., Johan A. Westerhuis, Age K. Smilde, and Jan van der Greef. “Statistical Validation of Megavariate Effects in ASCA.” *BMC Bioinformatics* 8 (2007): 322. doi:10.1186/1471-2105-8-322.

Walpole, Ronald E., and Raymond H. Myers. *Probability and Statistics for Engineers and Scientists*. 2nd ed. New York, NY: MacMillan Publishing Co., Inc., 1978.

Wang, David T., Danielle S. Gruen, Barbara Sherwood Lollar, Kai-Uwe Hinrichs, Lucy C. Stewart, James F. Holden, Alexander N. Hristov, et al. “Nonequilibrium Clumped Isotope Signals in Microbial Methane.” *Science* 348, no. 6233 (April 24, 2015): 428–31. doi:10.1126/science.aaa4326.

Ward, Joe. H. “Hierarchical Grouping to Optimize an Objective Function.” *Journal of the American Statistical Association* 58, no. 301 (March 1963): 236–44.

Wasserstein, Ronald L., and Nicole A. Lazar. “The ASA’s Statement on P-Values: Context, Process, and Purpose.” *The American Statistician*, March 7, 2016, 00–00. doi:10.1080/00031305.2016.1154108.

Watson, Nathanial E., H. Daniel Bahaghighat, Ke Cui, and Robert E. Synovec. “Comprehensive Three-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry.” *Analytical Chemistry* 89, no. 3 (February 7, 2017): 1793–1800. doi:10.1021/acs.analchem.6b04112.

Watson, Nathanial E., Joe M. Davis, and Robert E. Synovec. “Observations on ‘Orthogonality’ in Comprehensive Two-Dimensional Separations.” *Analytical Chemistry* 79, no. 20 (2007): 7924–27. doi:10.1021/ac0710578.

Watson, Nathanial E., Brendon A. Parsons, and Robert E. Synovec. “Performance Evaluation of Tile-Based Fisher Ratio Analysis Using a Benchmark Yeast Metabolome Dataset.” *Journal of Chromatography A* 1459 (August 12, 2016): 101–11. doi:10.1016/j.chroma.2016.06.067.

Watson, Nathanial E., W. Christopher Siegler, Jamin C. Hoggard, and Robert E. Synovec. “Comprehensive Three-Dimensional Gas Chromatography with Parallel Factor Analysis.” *Analytical Chemistry* 79, no. 21 (November 1, 2007): 8270–80. doi:10.1021/ac070829x.

Watson, Nathanial E., Matthew M. VanWingerden, Karisa M. Pierce, Bob W. Wright, and Robert E. Synovec. “Classification of High-Speed Gas Chromatography–mass Spectrometry Data by Principal Component Analysis Coupled with Piecewise Alignment and Feature Selection.” *Journal of Chromatography A* 1129, no. 1 (September 2006): 111–18. doi:10.1016/j.chroma.2006.06.087.

Weaver, Gabriela C., and Hannah G. Sturtevant. “Design, Implementation, and Evaluation of a Flipped Format General Chemistry Course.” *Journal of Chemical Education* 92, no. 9 (September 8, 2015): 1437–48. doi:10.1021/acs.jchemed.5b00316.

Welthagen, Werner, Robert A. Shellie, Joachim Spranger, Michael Ristow, Ralf Zimmermann, and Oliver Fiehn. “Comprehensive Two-Dimensional Gas Chromatography–time-of-Flight Mass Spectrometry (GC × GC-TOF) for High Resolution Metabolomics: Biomarker Discovery on Spleen Tissue Extracts of Obese NZO Compared to Lean C57BL/6 Mice.” *Metabolomics* 1, no. 1 (March 2005): 65–73. doi:10.1007/s11306-005-1108-2.

Wilson, Ryan B., Brian D. Fitz, Brandyn C. Mannion, Tina Lai, Roy K. Olund, Jamin C. Hoggard, and Robert E. Synovec. “High-Speed Cryo-Focusing Injection for Gas Chromatography: Reduction of Injection Band Broadening with Concentration Enrichment.” *Talanta* 97 (August 2012): 9–15. doi:10.1016/j.talanta.2012.03.054.

Wilson, Ryan B., Jamin C. Hoggard, and Robert E. Synovec. “Fast, High Peak Capacity Separations in Gas Chromatography–Time-of-Flight Mass Spectrometry.” *Analytical Chemistry* 84, no. 9 (May 2012): 4167–73. doi:10.1021/ac300481k.

———. “High Throughput Analysis of Atmospheric Volatile Organic Compounds by Thermal Injection – Isothermal Gas Chromatography – Time-of-Flight Mass Spectrometry.” *Talanta* 103 (January 2013): 95–102. doi:10.1016/j.talanta.2012.10.013.

Wilson, Ryan B., W. Christopher Siegler, Jamin C. Hoggard, Brian D. Fitz, Jeremy S. Nadeau, and Robert E. Synovec. “Achieving High Peak Capacity Production for Gas Chromatography and Comprehensive Two-Dimensional Gas Chromatography by Minimizing off-Column Peak Broadening.” *Journal of Chromatography A* 1218, no. 21 (May 2011): 3130–39. doi:10.1016/j.chroma.2010.12.108.

Wise, Barry M., and Neal B. Gallagher. “The Process Chemometrics Approach to Process Monitoring and Fault Detection.” *Journal of Process Control* 6, no. 6 (December 1, 1996): 329–48. doi:10.1016/0959-1524(96)00009-1.

Wold, Svante, Kim Esbensen, and Paul Geladi. “Principal Component Analysis.” *Chemometrics and Intelligent Laboratory Systems* 2, no. 1–3 (1987): 37–52.

Wong, Yong Foo, Sung-Tong Chin, Patrick Perlmutter, and Philip J. Marriott. “Evaluation of Comprehensive Two-Dimensional Gas Chromatography with Accurate Mass Time-of-Flight Mass Spectrometry for the Metabolic Profiling of Plant–fungus Interaction in Aquilaria Malaccensis.” *Journal of Chromatography A* 1387 (March 27, 2015): 104–15. doi:10.1016/j.chroma.2015.01.096.

Wu, Hai-Long, Yong Li, and Ru-Qin Yu. “Recent Developments of Chemical Multiway Calibration Methodologies with Second-Order or Higher-Order Advantages: Chemical Multiway Calibration Methodologies.” *Journal of Chemometrics* 28, no. 5 (May 2014): 476–89. doi:10.1002/cem.2570.

Xu, Feng, Wenna Guan, Guiyan Yao, and Yafeng Guan. “Fast Temperature Programming on a Stainless-Steel Narrow-Bore Capillary Column by Direct Resistive Heating for Fast Gas Chromatography.” *Journal of Chromatography A*, Trends and Developments in Gas Chromatography, 1186, no. 1–2 (April 4, 2008): 183–88. doi:10.1016/j.chroma.2007.08.063.

Yang, Song, Jamin C. Hoggard, Mary E. Lidstrom, and Robert E. Synovec. “Comprehensive Discovery of 13C Labeled Metabolites in the Bacterium Methylobacterium Extorquens AM1 Using Gas Chromatography–mass Spectrometry.” *Journal of Chromatography A* 1317 (November 2013): 175–85. doi:10.1016/j.chroma.2013.08.059.

Yang, Song, Jeremy S. Nadeau, Elizabeth M. Humston-Fulmer, Jamin C. Hoggard, Mary E. Lidstrom, and Robert E. Synovec. “Gas Chromatography–mass Spectrometry with Chemometric Analysis for Determining 12C and 13C Labeled Contributions in Metabolomics and 13C Flux Analysis.” *Journal of Chromatography A* 1240 (June 2012): 156–64. doi:10.1016/j.chroma.2012.03.072.

Yang, Song, Martin Sadilek, Robert E. Synovec, and Mary E. Lidstrom. “Liquid Chromatography–tandem Quadrupole Mass Spectrometry and Comprehensive Two-Dimensional Gas Chromatography–time-of-Flight Mass Spectrometry Measurement of Targeted Metabolites of Methylobacterium Extorquens AM1 Grown on Two Different Carbon Sources.” *Journal of Chromatography A* 1216, no. 15 (April 2009): 3280–89. doi:10.1016/j.chroma.2009.02.030.

Yeung, Laurence Y., Jeanine L. Ash, and Edward D. Young. “Biological Signatures in Clumped Isotopes of O2.” *Science* 348, no. 6233 (April 24, 2015): 431–34. doi:10.1126/science.aaa6284.

Zeng, Zhong-Da, Helmut M. Hugel, and Philip J. Marriott. “A Modeling Approach for Orthogonality of Comprehensive Two-Dimensional Separations.” *Analytical Chemistry* 85, no. 13 (July 2, 2013): 6356–63. doi:10.1021/ac400736v.

Zhai, G., R. Wang-Sattler, D. J. Hart, N. K. Arden, A. J. Hakim, T. Illig, and T. D. Spector. “Serum Branched-Chain Amino Acid to Histidine Ratio: A Novel Metabolomic Biomarker of Knee Osteoarthritis.” *Annals of the Rheumatic Diseases* 69, no. 6 (June 1, 2010): 1227–31. doi:10.1136/ard.2009.120857.

Zhang, Lei, Zhongda Zeng, Chunxia Zhao, Hongwei Kong, Xin Lu, and Guowang Xu. “A Comparative Study of Volatile Components in Green, Oolong and Black Teas by Using

Comprehensive Two-Dimensional Gas Chromatography–time-of-Flight Mass Spectrometry and Multivariate Data Analysis." *Journal of Chromatography A* 1313 (October 25, 2013): 245–52. doi:10.1016/j.chroma.2013.06.022.

Zhang, Li-Kang, Don Rempel, Birendra N. Pramanik, and Michael L. Gross. "Accurate Mass Measurements by Fourier Transform Mass Spectrometry." *Mass Spectrometry Reviews* 24, no. 2 (March 2005): 286–309. doi:10.1002/mas.20013.

Zhu, Jiangjiang, Danijel Djukovic, Lingli Deng, Haiwei Gu, Farhan Himmati, E. Gabriela Chiorean, and Daniel Raftery. "Colorectal Cancer Detection Using Targeted Serum Metabolic Profiling." *Journal of Proteome Research* 13, no. 9 (September 5, 2014): 4120–30. doi:10.1021/pr500494u.

Zimmer, Jennifer S.D., Matthew E. Monroe, Wei-Jun Qian, and Richard D. Smith. "Advances in Proteomics Data Analysis and Display Using an Accurate Mass and Time Tag Approach." *Mass Spectrometry Reviews* 25, no. 3 (May 2006): 450–82. doi:10.1002/mas.20071.

# VITA

Nathanial Watson was raised as a military dependent and spent his childhood years in Hawaii, Virginia, Florida and, for the greatest majority, Southeastern Germany. He graduated high school in 1995 as part of the final graduating class of Nürnberg American High School. Following high school he attended the University of Florida and graduated *cum laude* in 1999 with a degree in Soil and Water Science. Having accepted a US Army ROTC scholarship to fund his undergraduate studies, he was commissioned as a Second Lieutenant of Field Artillery that same year and proceeded to Fort Sill, OK. He retired from active service in 2016 after serving in Oklahoma, Colorado, Washington, New York, New Mexico, Bosnia and Herzegovina, South Korea and Afghanistan. He was married in 2002 and is the proud father of four. Having completed his PhD and military service all in the same six month window, he and his family are taking an extended sabbatical and touring across North America on two triple tandem bicycles. Someday he will likely settle down and get a regular job.