Advanced chemometrics and fundamental considerations for non-targeted analysis with comprehensive multidimensional gas chromatography coupled with time-of-flight mass spectrometry

Sarah E. Prebihalo

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Robert E. Synovec, Chair

Matthew Bush

Dan Fu

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Advanced chemometrics and fundamental considerations for non-targeted analysis with comprehensive multidimensional gas chromatography coupled with time-of-flight mass spectrometry

Sarah E. Prebihalo

Chair of the Supervisory Committee:
Robert E. Synovec
Chemistry

Comprehensive two-dimensional gas chromatography (GC×GC) coupled with time-of-flight mass spectrometry (TOFMS) is a powerful analytical technique capable of separating complex mixtures, providing valuable information about the chemical composition of samples. However, the inherent data density associated with three-dimensional data provides a unique challenge to analytical chemists. As a result, significant effort has been invested in utilizing advanced chemometrics to glean meaningful information about samples from large and complex data sets. Herein, this dissertation introduces several investigations conducted on optimizing separation conditions to be

amenable to chemometric deconvolution algorithms as well as the development, study, and application of advanced chemometric techniques applied to GC×GC-TOFMS data. To begin, the metric trilinear deviation ratio (TDR) is utilized to study the impact of experimental parameters such as column selection and modulation period, P_M, on the quantitative accuracy of parallel factor analysis (PARAFAC) deconvolution. TDR scales with increasing change in second dimension retention time, $\Delta^2 t_R$, associated with pseudo-isothermal conditions on the second dimension, ^{2}D, and quantitative accuracy decreases as TDR increases. Two column sets were utilized with varying film thickness on the first column, ^{1}D, and each column set was studied using two P_M for a total of 4 experiments. It was reported that using ^{1}D columns with larger film thicknesses allows the analyst to employ a shorter P_M, in turn lowering the $\Delta^2 t_R$, leading to higher quantitative accuracy. Many GC×GC-TOFMS studies relate to identifying class distinguishing analytes and can be tedious when performed manually. Fortunately, the use of discovery-based chemometric tools such as principal component analysis (PCA) and Fisher ratio (F-ratio) analysis has increased in popularity as less time-intensive and automated techniques for untargeted analyses. To begin, this dissertation will investigate mass channel purity obtained via the tile-based F-ratio algorithm using diesel fuel spiked with non-native analytes using GC×GC-TOFMS. The F-ratio algorithm, considered a supervised discovery technique because class membership is known *a priori*, was first used to "discover" the spiked non-native analytes. Then, using a novel signal ratio (S-ratio) algorithm, the mass channel selectivity information output by the F-ratio method was studied using three statistical metrics: null distribution analysis, p-value, and lack-of-fit (LOF). The result of this investigation revealed that a mass channel has a high likelihood of being pure when its p-value *and* LOF are sufficiently low. Finally, F-ratio analysis was applied to a dataset including patients with an anterior cruciate ligament (ACL) injury to discover potential biomarkers of post-traumatic

osteoarthritis (PTOA) post-injury. Standard F-ratios are calculated by the between class variance divided by the sum of the within-class variance, scaling up as the between class variance increases and the within-class variance remains sufficiently small. However, many biological studies involve significant biological variance (~30%) that may not be associated with disease state or injury severity, etc. Herein, the standard tile-based F-ratio algorithm was modified to use only the within-class variance associated with control samples. It was expected that the control class contained less within-class variance relative to the patient class, due to the expectation that some patient samples would be associated with increased severity of injury or the presence of coexisting conditions. Hit lists (metabolites discovered via F-ratio) from standard F-ratio and control-normalized F-ratio were studied and directly compared to establish a comprehensive metabolome of potential biomarkers for PTOA development post ACL injury. Reported in this dissertation is a discussion on the complementary nature of standard and control-normalized F-ratio, followed by demonstration of class distinguishing metabolites via PCA.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There is not a better way to start than by thanking my graduate research advisor, Rob Synovec. Thank you for your constant support and guidance during this experience, both academically and personally – I would not be here without you. To my research advisor at Penn State, Frank Dorman, thank you for convincing me that I have what it takes to "run the marathon of a PhD". Thank you to Stephen Stern of Upper Arlington High School for being the first teacher to encourage me to pursue a career in Chemistry and being my cheerleader in that endeavor since 2007.

Thank you to the members of the Synovec group I have had the privilege of working with over the last five years. Not only are you wonderful scientists and human beings, but you are hilarious. Thank you for the number of times I laughed until I cried. An extra-large thank you to Dr. Brooke Reaser, who not only answered hundreds of questions while we were in the group together, but has continued to uplift me, support me, and give me someone to look up to, even after moving across the country. Thank you to Dr. Dave Pinkerton, Dr. Nate Watson, and Dr. Chris Freye, who mentored me and answered a few too many questions as I began my time here. To Derrick and Paige, thank you for being my friends, supporting me, letting me rant, for the countless pep talks, and for sharing your dogs with me. Thank you to Brendon, Dan, Kelsey, Caitlin, Grant, Sonia, and Tim for being great group members, scientific collaborators, and for the coffee breaks.

Thank you to PAWS Cat City for providing me with kitten therapy once a week for almost two years. Volunteering quickly became one of my favorite times of the week, and I now know for sure that it IS possible to fall in love with hundreds of cats and be overjoyed when they get adopted. The opportunity to care for cats in between homes, help them find forever homes, and meet incredible people changed my life. I think I will be volunteering in some way for many years to

come. Bri, Sarah, Lynsey, Kate and Madeleine, thank you for our life chats while kittens climb up our backs and play in our hair, it has been a pleasure getting to know you. Lynsey, I sure hope you learned how to open cat food cans, because I will not be able to do it for you anymore!

Finally, thank you to my friends and family outside of the Synovec group and University of Washington for the constant support and encouragement. Thank you to Abbie and Emily for the vent sessions over bubble tea, Qdoba, and teriyaki, and for walking this journey with me. Thank you to Kari and Sarah for answering the phone when I needed a listening ear, no matter how late. Your emotional support over the last 5 years is truly something I will never forget. Mom and Dad, thank you for giving me every opportunity I could have asked for and for teaching me that "hard work beats talent unless talent works hard". To Emily, my little sister, thank you for being my best friend and my role model. I am so very excited to continue supporting you through your grad school experience. Brendan, thank you for uprooting your life and moving to Seattle with me, loving me, and supporting me in this adventure. I do not think I could have done this without you. Thank you.

DEDICATION

To my family for always believing in me;

and

To my cats for loving me even on the hardest days.

Chapter 1. Introduction to Gas Chromatography, Comprehensive Multidimensional Gas Chromatography, and Chemometrics

Some parts of this chapter have been reproduced from Sarah E. Prebihalo, Kelsey L. Berrier, Chris E. Freye, H.Daniel Bahaghighat, Nicholas R. Moore, David K. Pinkerton, and Robert E. Synovec, "Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications" *Analytical Chemistry* 90 (2018) 505-532.

1.1 Introduction to Gas Chromatography

1.1.1 *History*

The origin of chromatography is credited to Russian botanist M.S. Twsett and traces back to approximately the year 1900. In its original form, chromatography, derived from the Greek word of color (chroma) and write (graphein), consisted of a column of calcium carbonate used to separate the colored pigments of a green leaf. While chromatography was not instantly successful in this form, it became an established laboratory method in the 1930s and grew into a popular technique for a variety of applications. The advancement of modern chromatography can be, in part, attributed to A.J.P. Martin for introducing partition chromatography and gas-liquid chromatography [1]. For a complete introduction to the history behind the development of chromatography, the reader is directed to several sources that discuss major contributions [2,3].

Chromatography, by definition, encompasses a wide variety of techniques which aim to separate components through various chemical mechanisms between two phases, the first being stationary (stationary phase) and the second a "mobile phase" that moves in a definite direction [4]. Fortunately, chromatographers utilize the specific properties of each type of chromatography to effectively communicate with the greater scientific community. For example, some chromatography is classified by analyte type such as "ion chromatography", which separates ions. Other chromatographic techniques are classified by mobile phase, such as liquid chromatography

(LC), that uses a liquid mobile phase, or gas chromatography (GC) with a gaseous mobile phase. For further information on various types of chromatography, the reader is directed to several excellent sources describing the development and basic principles [3,5–7].

Gas chromatography, as stated above, refers to separations that utilize a gaseous mobile phase – typically helium, hydrogen, or nitrogen. Analytes (volatile and semi-volatile) are separated dependent on their affinity for the stationary phase, which is a liquid or polymer, and commonly operate based on boiling point or polarity. While samples separated by GC must be thermally stable but easily vaporized during injection into the inlet, fields such as environmental science [8,9], forensics [10–12], food science [13,14] and biochemistry [15,16] have exploited the separation power and detection of low concentrations that GC provides.

The expansive application of GC is in no doubt partially attributable to Marcel Golay, who introduced the capillary column that provided a monumental increase in separation power over the traditional packed columns. Capillary columns consist of the stationary phase coated on the inner wall of a narrow column, typically with inner diameters on the order of 0.1 – 0.53 mm [17]. Depending on the sample type, separations completed via capillary GC can be supervised by several detectors, such as flame ionization detection (FID) and mass spectrometry (MS). Gas chromatography coupled with mass spectrometry (GC-MS) can separate, identify, and quantify complex mixtures. Identification is performed when analytes previously separated by GC enter the MS and are ionized, fragmented, and separated based on mass-to-charge (*m/z*). For more information, the reader is directed to several sources which describe the mechanism of GC-MS [18–21].

1.1.2 *Fundamental Concepts of Gas Chromatography*

With the aim to provide context to the research reported in this dissertation, a subset of the fundamental aspects of gas chromatography will be discussed. Among these will include both qualitative and quantitative considerations as they relate to chromatographic methods and data analysis. For further information on the principles and fundamentals of chromatography, the reader is directed to a few excellent resources [1,3,6].

To begin, analytes are separated based on their affinity for the stationary phase, and as such, the identity of the stationary phase plays a significant role in a successful separation. Proceeding the injection of a mixture onto the column, the analytes equilibrate between the mobile and stationary phases, which can be described by the distribution coefficient, K_D,

$$K_D = \frac{[Analyte]_{SP}}{[Analyte]_{MP}} \tag{1.1}$$

where $[Analyte]_{SP}$ is equal to the concentration of the analyte in the stationary phase and $[Analyte]_{MP}$ is equal to the concentration of analyte in the mobile phase. Larger K_D values represent analytes which have a higher concentration in the stationary phase, i.e. have more affinity for the stationary phase, and are retained longer on the column. As mentioned above, K_D can be influenced by choice of stationary phase, as well as temperature.

Analytical chemists seek to measure the retention of an analyte on a column through the retention factor, *k'*. Mathematically, *k'* is determined by the ratio of the time the analyte occupies the stationary phase over the time the analyte spends in the mobile phase, or

$$k' = \frac{t_R - t_0}{t_0} \tag{1.2}$$

where t_R is the amount of time an analyte spends in the stationary phase, or retention time, and t_0 is the time required for an unretained analyte to elute from the column (dead time). The retention

factor is proportional to t_R such that analytes which have higher affinity for the stationary phase will have larger *k'* than analytes which have less affinity for the stationary phase.

Furthermore, the impact of stationary phase selection on *k'* can be discussed via the phase volume ratio, β, which is the ratio of the volume of the mobile phase, V_m, to the volume of the stationary phase, V_s,

$$\beta = \frac{V_m}{V_s} = \frac{d_c}{4d_f} \tag{1.3}$$

where d_c is the inner diameter of the column and d_f is the thickness of the stationary phase. By manipulating β through column selection, the analytical chemist can influence the elution temperature of analytes, T_e [22,23]. Additional discussion about the impact of β on separations will be provided in subsequent sections.

Analytes eluting from the column generate a Gaussian profile, defined as,

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1.4}$$

and σ is the standard deviation, μ is the mean, and the fraction preceding the exponent is equivalent to the peak area and is often replace by a constant, A. As mentioned above, t_R represents the time when the maximum of an analyte peak is detected by the detector and is equal to μ. Peak width is frequently measured either at baseline, W_b, or half height, W_h, where the baseline is defined as 4σ ± 2σ from t_R. Figure 1.1 demonstrates the relationship between these parameters and the percentage of the total peak height that occurs at assorted distances from t_R. Qualitative and quantitative measurements of peaks detected via mass spectrometry routinely rely upon the mass spectrum observed at t_R. However, there exists cases upon which manipulation of stationary phase identity, column length, temperature programs, etc. fails to sufficiently separate two or more analytes. The extent to which multiple analyte peaks overlap is quantified by analytical chemists through calculation of resolution, R_s. Resolution is mathematically defined as,

$$R_s = \frac{2(t_{R,2} - t_{R,1})}{W_{b,1} + W_{b,2}} \tag{1.5}$$

which can be distilled down as the difference in t_R between two peaks divided by the average W_b, demonstrated in Figure 1.2. Peak widths are routinely measured at half-height, with W_b approximated as 1.7 times W_h.

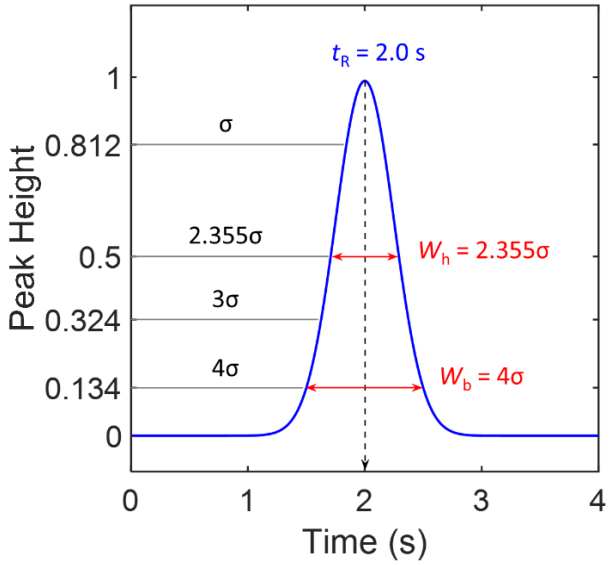

Figure 1.1. Gaussian peak with retention time, fractional peak heights, width at base and width at half-height indicated.

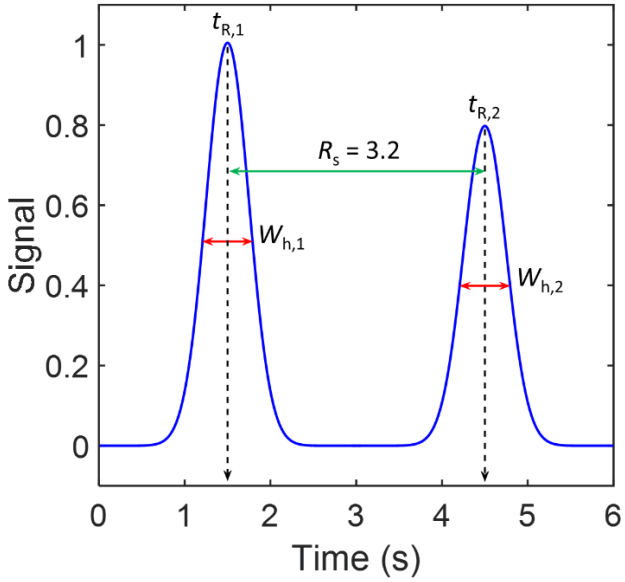

Figure 1.2. Visual depiction of the resolution between two peaks.

Such resolution calculations can give chromatographers valuable information about the quality of information. Four example R_s values are observed in Figure 1.3. The red and blue curves represent a single analyte, while the black trace provides demonstration of what a detector would visualize, and the purple area represents the area of the peaks which overlap. Baseline resolution is defined as $R_s = 1.5$, albeit $R_s = 1.0$ is commonly accepted as adequate for quantification. Two analytes are likely to appear as one peak when $Rs \leq 0.5$, and $R_s = 0.3$ is where most deconvolution methods fail. Deconvolution methods will be further discussed in Section 1.2.3 and Chapter 2.

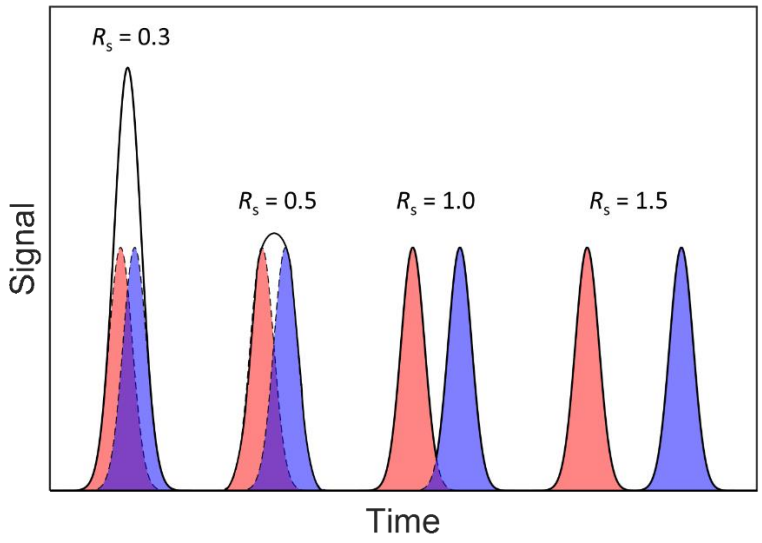

Figure 1.3. Visual representation of four resolution values.

1.1.3 *Introduction to Comprehensive Multidimensional Gas Chromatography*

Analysis of volatile and semi-volatile analytes by gas chromatography (GC) methods is an indispensable tool in the analytical chemist's toolbox. A myriad of fields of study rely upon the application of GC methods to address an ever-growing demand to provide useful chemical information from GC data. As the realm of GC application has expanded, there has been an evolution to develop more powerful instrumental and data analysis approaches to keep pace with the wealth of complex samples that require analysis. To address this challenge, advances in GC

instrumentation having evolved from one-dimensional gas chromatography (1D-GC) and heart cutting approaches such as (GC-GC), to instrumentation referred to broadly as multidimensional gas chromatography (MDGC), which can take on many forms. The principle form of MDGC that has gained wide implementation is comprehensive two-dimensional (2D) gas chromatography (GC × GC) as shown in Figure 1.4A, pioneered nearly 29 years ago by Liu and Phillips[24] When comparing 1D-GC (Figure 1.4B) relative to GC × GC (Figure 1.4C), the benefits of a secondary separation become evident. Blumberg and co-workers have theoretically determined that the 2D peak capacity provided by GC × GC compared to the peak capacity of 1D-GC is approximately an order of magnitude higher when the run times are held constant [25]. This benefit is illustrated using a relatively complex sample of diesel fuel. By adding a multivariate detector such as a time-of-flight mass spectrometry (TOFMS), another selective dimension of data is provided which may allow identification of analytes (Figure 1.4D).

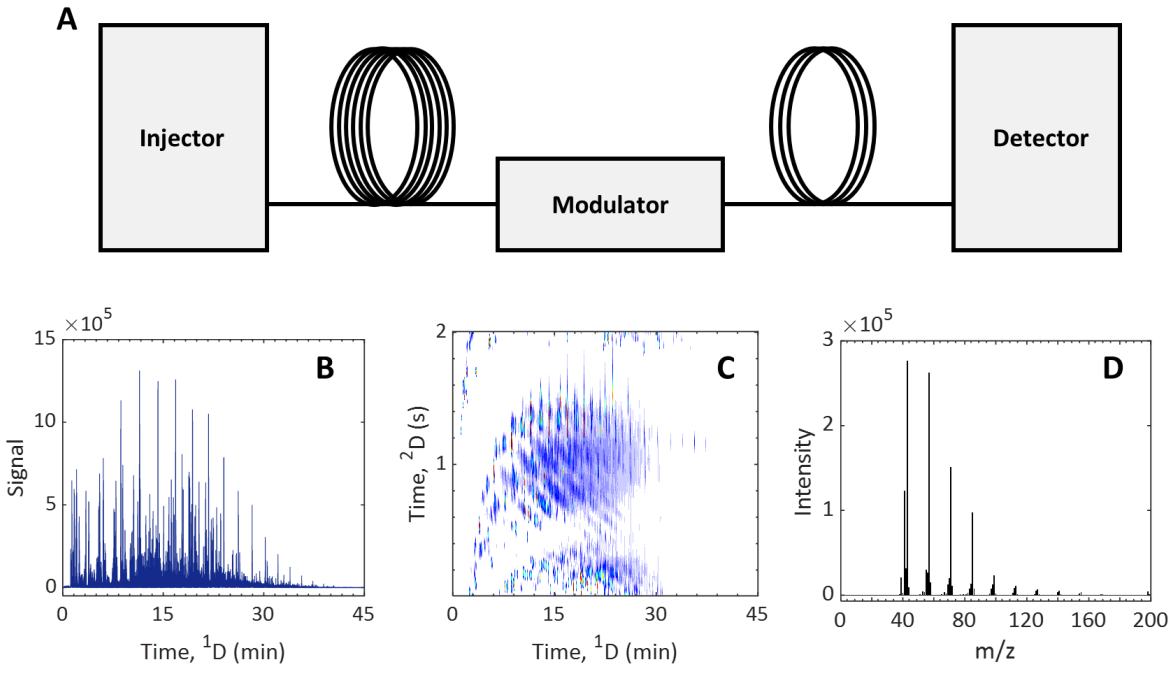

Figure 1.4. (A) Typical multidimensional GC set up. (B) 1D-GC total ion chromatogram for diesel. (C) GC×GC total ion chromatogram for diesel fuel. (D) Mass spectrum of a long-chain alkane.

In GC×GC, the two column selections ideally do not utilize the same separation mechanism and traditionally consists of a long first dimension column on the order of 20-30 m coupled to a short second dimension (1-5 m). The two columns are serially coupled with modulator that collects small amounts of eluate from the first column, focuses it into a small sample band, and reinjects it onto the second column. Traditionally, the first dimension, ^{1}D, separation method closely resembles 1D-GC, while the second dimension, ^{2}D, separation is usually only on the order of a few seconds long. The 10-fold increase in peak capacity mentioned above is due to the complementary separation mechanism of 2D, which allows analytes that would otherwise elute at the same ^{1}D retention time, $^{1}t_{R}$, to spread out on ^{2}D and result in various ^{2}D retention times, $^{2}t_{R}$. Figure 1.5 demonstrates this concept, as there is a clear coelution occurring around a $^{1}t_{R}$ of 19 minutes. The ~3 peaks present with a $^{2}t_{R}$ of ~1.65 s would otherwise mask the peak with a $^{2}t_{R}$ of ~1.52 s.

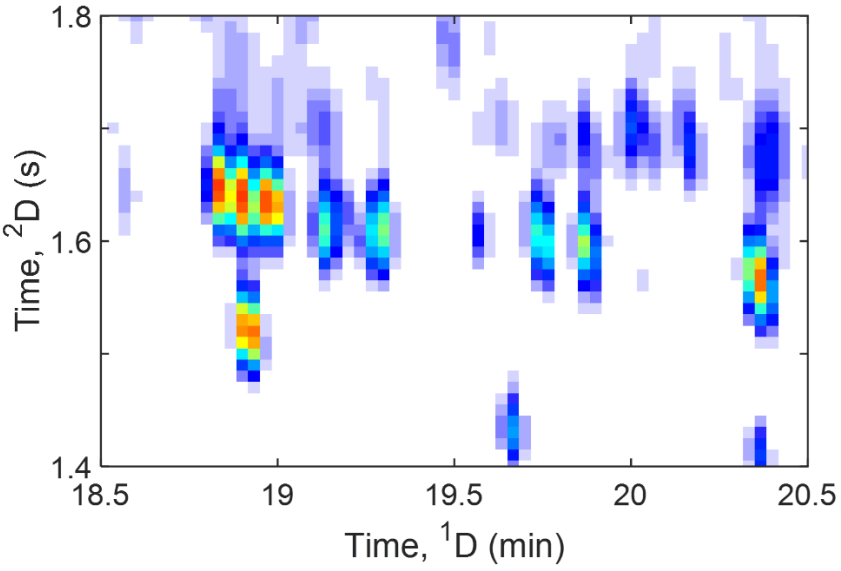

Figure 1.5. Section of a GC×GC chromatogram with each peak represented by a blue contour.

Each blue oval present in a GCGC contour plot represents one peak which has been folded and shown in two dimensions. Further example of this can be observed in Figure 1.6, where (A)

is the unfolded 2D peak, and (B) is the same peak folded and shown in a contour plot. The data demonstrated in Figure 1.6 has a modulation period, P_M, or ^{2}D separation time, of 1 s. The dotted lines in Figure 1.6A indicate where the eluate is reinjected onto the ^{2}D column, and the space in between the dotted lines represents the P_M, which is also indicated in Figure 1.6B as the length of the ^{2}D time.

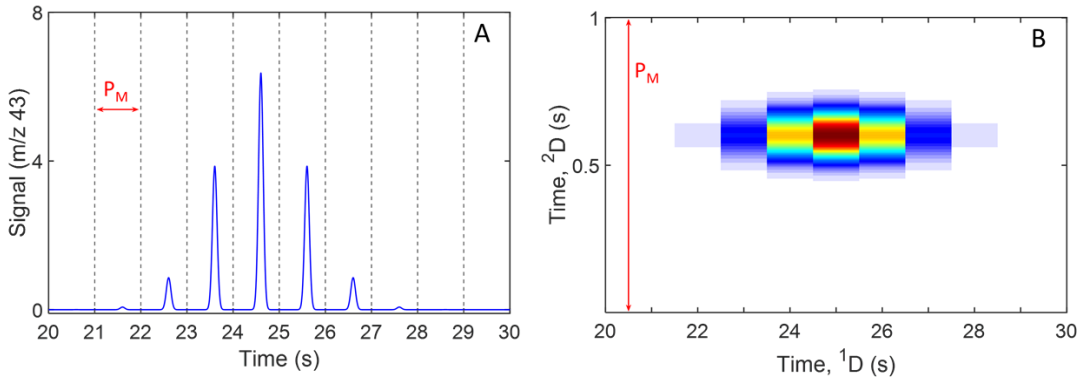

Figure 1.6. (A) GC×GC data as viewed for the detector, and (B) GC×GC data folded into ^{2}D space.

Paramount to GC×GC data quality are the concepts of sampling density and peak capacity. Sampling density, ρ_s, is mathematically defined as,

$$\rho_s = \frac{^1W_b}{P_M} \tag{1.6}$$

where ^1W_b is the ^{1}D peak width at base, and P_M is the modulation period. For a comprehensive separation, a ρ_s of 2-4 is ideal as it ensures proper sampling of the ^{1}D separation without oversampling [26].

To achieve the previously mentioned 10-fold increase in peak capacity of GC×GC as compared to 1D-GC, peak capacity, n_c, must be maximized. At unit resolution (R_s = 1), the peak capacity for a given dimension, $^x n_c$, is equal to,

$$^x n_c = \frac{^x t_{sep}}{^x W_{b,avg}} \tag{1.7}$$

where ${}^{x}t_{sep}$ is the 1D separation time, and ${}^{x}W_{b,avg}$ is the average 1D peak width at base. For GC×GC peak capacity, $n_{c,2D}$, is simply the product of the 1D peak capacity, ${}^{1}n_c$, and the 2D peak capacity, ${}^{2}n_c$,

$$n_{c,2D} = {}^{1}n_c \times {}^{2}n_c \tag{1.8}$$

While the addition of a second dimension provides powerful separation capabilities, there still exists a need to optimize the separation through column selection. To address this challenge, it has been demonstrated that the useful peak capacity can be improved by considering the β_r, or the ratio of β between the first and second dimensions [22,23,27,28]. Recalling β Eq. (1.3) and the β for an open tubular capillary column, β_r can be mathematically described as,

$$\beta_R = \frac{{}^{1}\beta}{{}^{2}\beta} = \frac{{}^{1}d_c/4\times {}^{1}d_f}{{}^{2}d_c/4\times {}^{2}d_f} = \frac{{}^{1}d_c\times {}^{2}d_f}{{}^{2}d_c\times {}^{1}d_f} \tag{1.9}$$

where ${}^{1}\beta$ is the beta of the ^{1}D column, ${}^{2}\beta$ is the beta of the ^{2}D column, ${}^{1}d_c$ is the ^{1}D column inner diameter, ${}^{1}d_f$ is the ^{1}D column film thickness, ${}^{2}d_c$ is the ^{2}D column inner diameter and ${}^{2}d_f$ is the ^{2}D column film thickness. As will be extensively discussed in Chapter 2, careful consideration of β_r of a GC×GC separation is important, as an increase in β_r results in an increase in ^{2}D retention factor, ${}^{2}k$' [22,28,29].

As demonstrated above, GC×GC data is inherently complex, especially with the addition of mass spectrometry as the detector. Therefore, an in-depth discussion of data handling of GC×GC data is warranted, as presented in Section 1.2.

1.2 Chemometrics and Data Analysis

1.2.1 *Basic Concepts*

GC × GC is generally applied to complex samples and coupled with multichannel detectors such as the TOFMS or vacuum ultraviolet spectroscopy (VUV) [27,30], resulting in immense

amounts of data. The enormity of the data is especially true when implemented for routine analyses, when large numbers of samples or replicates are analyzed. However, the task for the analyst is not over once the data is collected and the data must be transformed into useful information, via data analysis. This transformation is often provided by commercially available software; however, the analyst may find that commercially available data analysis options do not meet their needs. For these cases, many users turn to chemometric methods that take advantage of the higher order dimensionality of the GC×GC data sets. As defined by the IUPAC, chemometrics is "the application of statistics to the analysis of chemical data (from organic, analytical or medicinal chemistry) and design of chemical experiments and simulations [31]". The field of chemometrics was spearheaded largely by Bruce Kowalski and Svante Wold and has been extensively utilized in analytical chemistry [27], forensic science [32], and metabolomics [33]. Broadly speaking, chemometric methods aim to convert chemical data into information using algorithms based upon linear algebra and statistical principles.

Once the GC×GC data is collected, analysts ultimately aim to identify and quantify analytes (in either a targeted or non-targeted fashion) to address a variety of goals in the context of the experimental design: to characterize samples based upon the chemical measurements, to identify key compounds indicative of a particular cause and effect stimulus, and to relate chemical measurements to other measured properties of the samples. Targeted and non-targeted analysis refers to when the analytes of interest are either known or unknown beforehand, respectively. However, there are challenges in working with such complex and dense data sets. Some challenges include analyte peak overlap, difficult and time-consuming computations, confident discovery of important (and often subtle) chromatographic differences between samples, and extraction of other meaningful information. Optimization of the experimental and instrumental conditions is very

important in dealing with these issues. However, when instrumental capabilities have been fully optimized and have become a limiting constraint, the analyst must apply other approaches to achieve the desired goals. Data analysis approaches complement and extend the value brought by chromatographic separations, providing a means to achieve the mentioned goals from collected data while simultaneously further addressing the challenges. For example, deconvolution methods are implemented to mathematically resolve co-eluting analyte peaks with the end goal of identification and quantification.

The common classes of data analysis methods include the following: deconvolution, and pattern recognition. Before many of these data analysis methods can be applied, the GC×GC data must undergo preprocessing to facilitate accurate and meaningful implementation. Preprocessing procedures are critical in removing/minimizing artifacts in the data such as noise, baseline effects, and retention time shifts that would otherwise make performing chemometric analyses very difficult and less meaningful. Common preprocessing includes noise reduction procedures (smoothing etc.), baseline correction, normalization, and time alignment [34].

For chemometric methods, additional care is often required to ensure the GC×GC instrument generates data that is amenable to many chemometric algorithms and will be further discussed in Section 1.2.3 and Chapter 2. Although the various chemometric methods differ in their applications, data requirements are fairly consistent between them. Many methods require bilinear or trilinear data for appropriate implementation. A suitable data density across the ^{1}D and ^{2}D peaks is also required. Many of these chemometric methods require a minimum amount of chemical selectivity, in both the separation (i.e. adequate resolution) and detection dimensions. Several chemometric algorithms that were used for Chapters 2-4 will be described in the following, including deconvolution (Section 1.2.3) and pattern recognition (Section 1.2.5).

1.2.2 *Pre-processing*

For chemometric methods the data structure becomes an important consideration for successful implementation. Before many of these data analysis methods can be applied, the GC × GC data must undergo preprocessing to facilitate accurate and meaningful implementation. Preprocessing procedures are critical in removing/minimizing artifacts in the data such as noise, baseline effects, and retention time shifts that would otherwise make performing chemometric analyses very difficult and less meaningful. Common preprocessing includes noise reduction procedures (smoothing etc.), baseline correction, normalization, and time alignment [34]. Issues with the data, such as retention time misalignment, must be handled either within commercial software packages or using in-house algorithms in script language packages. Alignment algorithms can be applied within a chromatogram to improve data bilinearity or trilinearity [35] or applied across multiple samples and whole chromatograms, where approaches include peak-table based, pixel-based, and global alignment strategies[36–46]. Figure 1.7 summarizes the common data structures (unfolded versus folded, and single versus multiple samples) and issues (ideal versus misaligned) as well as the methods that can be applied to the various data manifestations. The data is often viewed as folded, either as a contour plot or pixelated. However, most chemometric methods are applied to the unfolded data, where the modulations are concatenated (Figure 1.7). After analysis, the data can be refolded. The chemometric methods referred to in Figure 1.7 are discussed in more detail below in the context of the specific analysis goal. Please take note that for equations presented in this chapter, lower case letters represent scalars, bold lower case letters portray vectors, superscript T means the variable is transposed, bold uppercase letters represent matrices, and underlined bold uppercase letters represent 3D arrays.

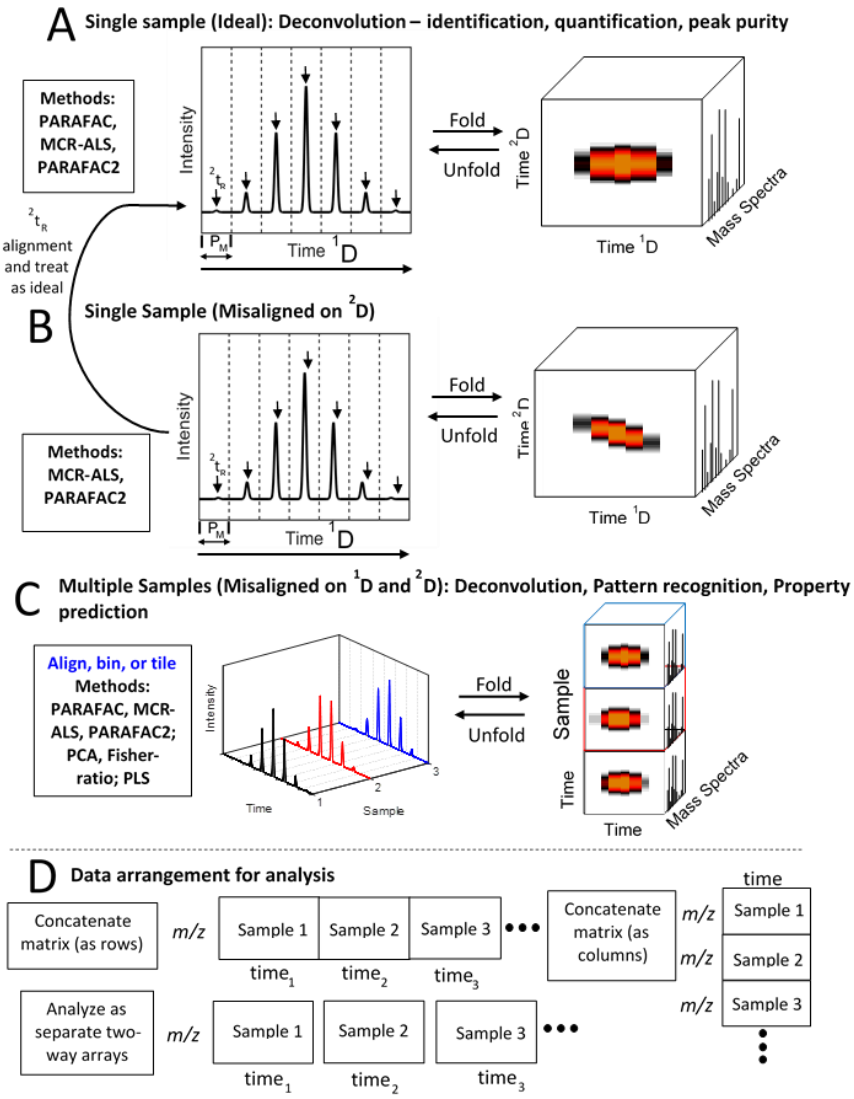

Figure 1.7. (A) Various data formats, analysis requirements, and associated chemometric methods. (A) Ideal data formats for a single sample, where the ^{2}D retention times are aligned. Deconvolution is the primary goal, with common methods listed. The arrows in the unfolded data indicate the ^{2}D retention time of the first modulated ^{2}D peaklet, with modulations indicated by dashed lines. (B) Data formats for a single sample, illustrated (for clarity) with exaggerated ^{2}D misalignment. Trilinear chemometric methods such as PARAFAC require sufficient alignment; otherwise, trilinearity is not required by PARAFAC2. (C) Data formats for cross-sample analysis of multiple samples with or without misalignment on either ^{1}D, ^{2}D, or both dimensions. Analysis goals include deconvolution, pattern recognition, and property prediction. Misalignment may exist within a single run (^{2}D misalignment), across multiple runs (^{1}D misalignment), or both. Retention alignment, or data binning/tiling may be necessary, followed by chemometric analysis of the unfolded data. (D) Analysis of multiple samples can be performed one sample at a time on the separate unfolded two-way arrays. Alternatively, multiple samples can be analyzed simultaneously by unfolding the data and concatenating the samples. This can be performed in a row-wise or column-wise approach.

1.2.3 *Deconvolution*

Despite the increased peak capacity and the additional separation dimension, analyte overlap is ubiquitous in GC×GC separations. Chromatographic peak deconvolution methods play an integral role in addressing this challenge to provide confident analyte identification, quantification, and peak purity assessment. Through the application of deconvolution methods, such as parallel factor analysis (PARAFAC) and multivariate curve resolution alternating least squares (MCR-ALS), overlapped peaks can be computationally separated on the ^{1}D time, ^{2}D time, and spectral dimensions to gain chemical information for specific analytes. With the deconvoluted peak data, analytes that were overlapped can be identified using the deconvoluted spectra, and quantified with the peak profiles obtained, thereby facilitating calibration methods [47,48]. Deconvolution methods can be applied in a targeted or non-targeted fashion, meaning that the analytes of interest are either known or unknown beforehand. In the application of these methods, the chromatographic region of interest is selected, and the data is unfolded (Figure 1.7A). Deconvolution can be applied to a single chromatogram, or multiple samples by concatenating the data (Figure 1.7). In certain cases, there is a data trilinearity or bilinearity requirement. An ideal chromatogram is one in which there is no retention time misalignment on ^{2}D (and/or ^{1}D, depending on whether one or more samples are to be analyzed), meaning that the data has sufficient trilinearity for trilinear data analysis methods (Figure 1.7A).

1.2.4 *Parallel Factor Analysis (PARAFAC)*

PARAFAC is a tensor rank decomposition method that has been applied successfully to GC×GC –TOFMS data for deconvolution, noise reduction, and calibration. Mathematically, PARAFAC can be described as:

$$\underline{\boldsymbol{R}} = \sum_{i=1}^{n} \sum x_i \otimes y_i \otimes z_i + \underline{\boldsymbol{E}} \quad (1.10)$$

Where **R** represents the data cube generated from a GC×GC –TOFMS separation. For each of the n total components, **x**, **y**, and **z** vectors represent the ^{1}D, ^{2}D and mass spectral profiles, respectively. Remaining signal is held in the residuals matrix, **E**. For more detail, the reader is directed to these references: [49,50]. PARAFAC is based on alternating least squares decomposition that requires at least three-way data, requiring a multiway detector that produces sufficiently trilinear data [49]. The instrumental design, such as the temperature programming rate and P_M selection, impacts the trilinearity of the data, particularly for compounds that are highly retained on ^{2}D separations [29]. A noticeable deviation from trilinearity may be observed as a shift in ^2t_R (^{2}D retention times) between successive modulations, as illustrated in Figure 1.7B, unless care is taken with the separation conditions. Deviation from trilinearity is observed when a relatively long P_M (~ 5 to 8 s) and high temperature programming rates are used [28,29]. If PARAFAC is to be applied to significantly nontrilinear data, then local data alignment must be performed on the data along the ^{2}D dimension to restore sufficient trilinearity. Then, PARAFAC may be applied as in the ideal case of Figure 1.2A. Otherwise, the application of PARAFAC to significantly non-trilinear data can result in considerable errors in quantification [28,29]. PARAFAC2 has been introduced as a deconvolution method capable of handling more deviation in data trilinearity. However, to date, it has not been extensively used in GCGC data analysis [51,52]. The implications of non-trilinear data on PARAFAC quantification will be extensively discussed in Chapter 2, as well as relevant experimental conditions that impact data structure.

1.2.5 *Multivariate curve resolution alternating least squares (MCR-ALS)*

MCR-ALS is an iterative method that initially estimates the pure chromatographic profiles and pure spectra and repetitively alternates and tests these values for convergence. Similar to

PARAFAC, the number of components (rank) must be provided, and this value is often varied to determine the best model when the number of chemical components is unknown; additionally, multiple constraints are also possible in the application of MCR-ALS, including unimodality, nonnegativity, selectivity, and convergence constraints. The observed data matrix, **D**, consisting of the chromatographic and mass spectral data is decomposed into the two matrices, **C**, containing the pure analyte concentration profile and $\mathbf{S^T}$, which is the pure mass spectra of n pure component species, and is mathematically described as,

$$\boldsymbol{D = CS^T + E} \tag{1.11}$$

where **E** is the residuals between the observed and predicted two-way chromatogram. One limitation to MCR-ALS is the lack of unique solutions, also referred to as the presence of rotational ambiguities, particularly when one sample is analyzed. One way to overcome this issue is to simultaneously analyze multiple samples/data sets through matrix augmentation, resulting in an extended MCR-ALS method. Matrix augmentation can be achieved by unfolding the 2D data and concatenating the individual two-way arrays as either rows or columns (Figure 1.7D). For more information, the reader is directed to several detailed resources [53–55].

1.2.6 *Comparative Analysis*

A major goal in performing GC×GC analyses is to discover relevant features in the data across multiple samples. Many pattern recognition applications fall under this category of cross-sample analysis, including sample classification, class comparison, chemical fingerprinting, and chemical/biomarker discovery. These methods are classified as non-targeted approaches, where the relevant analytes may be unknown *a priori*, and the experimental design may be either supervised or unsupervised. Supervised methods are used when information about the samples (i.e. classes) is known beforehand and thus utilized in the experimental design and operation of the

method. Unsupervised methods are applied when there is much less known about the sample classifications *a priori* and this information is desired (i.e. sample class membership). Classification is used to describe pattern recognition methods that aim to classify samples into groups based on similarity, such as principal component analysis (PCA). These methods can be used to classify unknown samples into known classes (origin, etc.) based upon the chromatographic data/features. Chemical fingerprinting falls under this category, in which the unique chemical signature of an unknown sample is used to determine identifying characteristics for that sample. Class comparison methods aim to identify significant differences between samples belonging to different classes, whereby class membership is known *a priori*, such as Fisher ratio analysis, which can assist/result in the discovery of biomarkers or chemical markers that differentiate sample classes. Pattern recognition methods can be used for feature selection or data reduction, followed by further chemometric analysis. Typically, because these methods are applied to/across (and require) whole chromatograms of multiple samples, data misalignment in both GC×GC dimensions becomes a significant issue in implementation (Figure 1.7C). This problem can be dealt with using data alignment methods but has also been addressed by data reduction techniques such as binning and tiling. Two common pattern recognition methods (PCA, Fisher ratio) will be discussed below in terms of method fundamentals, and experimental and data requirements.

1.2.7 *Principal Component Analysis (PCA)*

PCA is a popular tool utilized in many applications for the statistical separation of sample classes, providing information on the variables according to which the samples are separated (loadings) in addition to the actual separations between the classes (scores). PCA is an unsupervised non-targeted classification method that is based on rotation of the space axes to

encompass the greatest chemical variance in the measurements. The new axes are called principal components and are orthogonal to each other. Mathematically, PCA is completed via eigenvalue decomposition on the covariance matrix of the data using singular value decomposition (SVD),

$$\mathbf{X} = \mathbf{TDP}^{\mathrm{T}} + \mathbf{E} \tag{1.12}$$

Where **X** is a data matrix the size of m rows by n columns. The number of samples is represented by m, and the number of variables n. **D** is a diagonal matrix with elements equal to the square roots of the eigenvalues of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. **T** is the matrix whose columns contain the scores vectors, $\mathbf{t}_{\mathrm{i}}$, and $\mathbf{P}^{\mathrm{T}}$ is the matrix whose rows consist of the loadings vectors, $\mathbf{p}^{\mathrm{T}}{}_{\mathrm{i}}$. Such as in Eq. (1.10), **E** is the matrix consisting of the residuals. The principal components (PCs) are sorted by their eigenvalues and organized such that they span the greatest variation in the scores. PCs that contain meaningful chemical/physical information are retained, while PCs that are primarily noise are discarded. The scores plot shows the degree of sample clustering, shedding information on distinct grouping of samples. For many applications, sample classes are known beforehand: origin/source, type/variety, age/year, processing, and many other variables can be reflected in the samples. The loadings plot provides information about the basis of the groupings; in other words, the loadings can be used to determine which regions of the chromatogram (i.e. peaks/analytes) are important in differentiating the various sample classes.

Since PCA determines the sources of the greatest variance, preprocessing is necessary to obtain meaningful results. In general, mean-centering and scaling are important steps before applying PCA to a dataset; in chromatography, this is achieved through baseline correction and normalization. As seen in Figure 1.7C, PCA is often applied to the whole chromatogram across multiple samples; therefore, sufficient retention time alignment in both the ^{1}D and ^{2}D dimensions

is very important for proper implementation. Various alignment algorithms are available, and binning is another option for dealing with issues caused by misalignment (Figure 1.7C). Herein, binning (or tiling) refers to summing the area of user-specified regions of 2D separation space, whereby the data density is reduced, and misalignment is mitigated. Then, PCA is applied to the unfolded data in the form of a two-way array, where the number of rows is equal to the number of samples and the number of columns is equal to the number of variables (i.e. data points or bins) in the chromatogram. PCA can be applied to the whole chromatogram or a region of the chromatogram. Similarly, the analysis can be performed using all mass channels (*m/z*), a subset of all *m/z*, or the total ion current (TIC) chromatogram. In the former cases, the data must be completely unfolded along both time dimensions and the mass spectral dimension. In the latter case, summing to the TIC reduces a significant dimension of chemical selectivity, but the number of data points is significantly reduced.

PCA may be implemented for feature selection, due to its powerful abilities as a data reduction method. In many studies, PCA is applied for data preprocessing prior to the application of additional chemometric methods, such as deconvolution or other pattern recognition methods [56–59]. Alternatively, PCA can be applied to reduced data sets to improve class separation [58,60–65]. PCA can also be used to classify unknown samples based on the sample grouping of known samples. Multiway PCA is the extension of PCA to high order data, such as unfolded GC × GC data [56,66–68].

1.2.8 *Fisher Ratio (F-ratio) Analysis*

Non-targeted discovery-based analysis is an important focus as it allows the analyst to uncover chemical features of interest that are not known beforehand. Fisher ratio (F-ratio) analysis is a supervised non-targeted approach to discover underlying differences in samples. Supervision

refers to prior classification of samples/chromatograms as they relate to the experimental design. Originally described by Fisher in the 1920s [69], the F-ratio approach is an analysis of variance method that provides data reduction to elucidate class distinguishing features. The F-ratio is calculated as the between class variance divided by the sum of the within class variance:

$$\text{F-ratio} = \frac{s_{bc}^2}{\sum s_{wc}^2} \tag{1.13}$$

where s_{bc}^2 is the between class variance and s_{wc}^2 is the within class variance. Since the F-ratio calculation is based on the variance of the signals in the GC × GC data (raw data approach) or variance in quantified analytes (peak table approach), the F-ratio method prioritizes statistical significance over absolute signal/concentration with a higher F-ratio having a greater difference between the two classes. Similar to PCA, F-ratio analysis can be applied prior to additional chemometric analysis, serving to improve variable selection [58,61–64], or after other data reduction methods have been applied [57]. However, if sample class membership is known *a priori*, F-ratio analysis is preferred over unsupervised PCA, since within class variance in the samples severely hampers successful implementation of PCA [70].

A critical aspect of F-ratio analysis is the alignment of the data prior to the calculation of the F-ratio (Figure 1.7C). If sufficient misalignment is present, this may result in false positives or false negatives. Thus, the implementation of F-ratio analysis can often take many different forms. Pixel-based F-ratio analysis means an F-ratio is calculated at every data point within the chromatogram, i.e., at every point in which a mass spectrum has been collected in the GC×GC separation. In order to maximize the performance of pixel-based F-ratio analysis, the data should be initially aligned. F-ratio analysis can also be performed on peak tables of quantified/identified analytes [64]. After the peak tables are calculated, they are aligned so F-ratios can be calculated. While this is a simpler method then dealing with 2D retention time misalignment, peak tables may

miss analytes present at low concentrations. Another way to deal with misalignment is binning the data based upon the ^{1}D and ^{2}D peak widths in relation to the observed retention time shifting. However, this will result in splitting analytes into multiple bins and lower F-ratio values. In order to address this issue with binning, a method has been introduced to use a tiling scheme which creates bins at different locations enabling the bin to be centered on the peak resulting in a maximum F-ratio [71–74]. Ideally, the tile size should be selected as to encompass the width of the ^{1}D and ^{2}D peaks and allow for a modest amount of retention time shifting, as demonstrated in Figure 1.8. For complete details on the tile-based F-ratio method, the readers are directed to these excellent sources: [71,72,74,75].

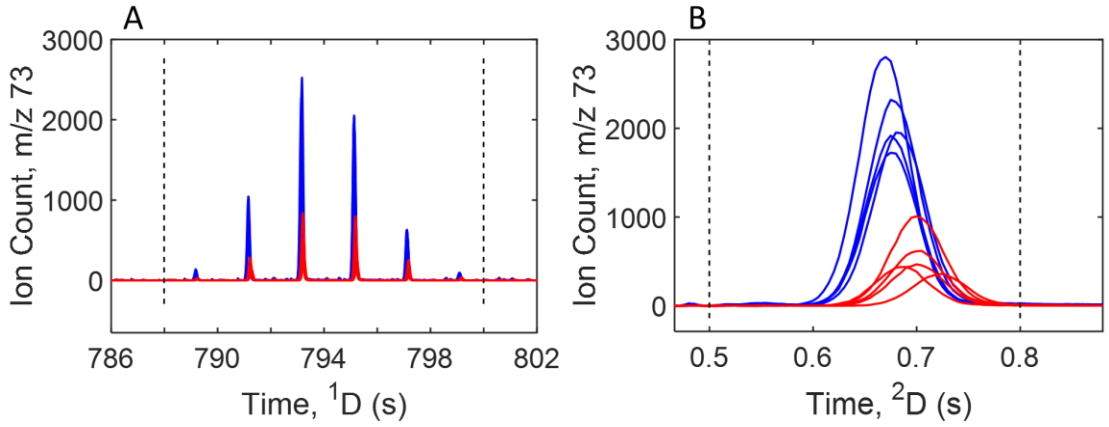

Figure 1.8. (A) ^{1}D peak profile with dashed lines indicating the size of the tile on ^{1}D. (B) ^{2}D peak profile for one modulation with dashed lines indicating the size of the tile of ^{2}D.

After the F-ratios are calculated, a hit list is assembled indicating the 2D retention times, with the highest F-ratio indicative of the statistically most prominent feature or "hit". Often, a F-ratio cutoff in the hit list is selected to determine the point at which the analysis becomes unreliable. Although it is sometimes possible to examine the entire hit list to determine which hits are true positives and which are false positives, this is a time-consuming step. Manual selection of the cutoff is possible but may require the user to wade through the hit list until too many false positives are discovered. Two different methods have been purposed to define a F-ratio threshold cutoff.

The first is using the F-critical value [62,63], and the second method takes a null distribution approach, which takes advantage of the inherent noise (instrumental and chemical) in the data set [71,73,74]. Tile-based F-ratio has been successfully applied to acid-altered diesel fuel [74], yeast metabolomics [71] and spiked diesel fuel [75]. The tile-based F-ratio method was employed for the investigations in Chapters 3 and 4.

1.3 Overview of Following Chapters

1.3.1 *Chapter 2: Impact of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Experimental Design on Data Trilinearity and Parallel Factor Analysis Deconvolution*

Comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC×GC-TOFMS) is a powerful instrument for the analysis of complex samples. Deconvolution of overlapped analytes using a suitable chemometric data analysis method such as Parallel Factor Analysis (PARAFAC) is often required. However, PARAFAC is designed to require a strict data trilinearity requirement. In this study we examine how strict this requirement is in the context of GC×GC experimental conditions and demonstrate that under suitable conditions the data is sufficiently trilinear to achieve accurate deconvolution. The term trilinear deviation ratio (*TDR*) was previously introduced as a quantitative metric to predict the accuracy of PARAFAC deconvolution. Trilinear deviation ratio is defined as the run-to-run retention time shift, $\Delta^2 t_R$, for a given analyte on the second dimension (^{2}D) separation, divided by the ^{2}D analyte peak width-at-base, ^2W_b. We demonstrate that experimental conditions impact the *TDR* range produced and PARAFAC performance. Column selection and modulation period, P_M, are shown to significantly influence the *TDR* range. Two column sets were evaluated, giving rise to different *k'* ranges for the ^{2}D separations. Each column set was used with an optimum P_M as well as a longer

P_M to demonstrate the effect of P_M selection on the *TDR* range and PARAFAC quantification. It is hypothesized that the selection of a column set which provides a sufficient *k*' range and a P_M which maximizes the use of the ^{2}D separation will result in the lowest *TDR* and correspondently low PARAFAC quantification error.

1.3.2 *Statistical Inference of Mass Channel Purity from Fisher Ratio Analysis Using Comprehensive Two-Dimensional Gas Chromatography with Time of Flight Mass Spectrometry Data*

Tile-based Fisher ratio (F-ratio) analysis has recently been developed and validated for discovery-based studies of highly complex data collected using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC×GC-TOFMS). In previous studies, interpretation and utilization of F-ratio hit lists has relied upon manual decomposition and quantification performed by chemometric methods such as parallel factor analysis (PARAFAC), or via manual translation of the F-ratio hit list information to peak table quantitative information provided by the instrument software (ChromaTOF). Both of these quantification approaches are bottlenecks in the overall workflow. In order to address this issue, a more automatable approach to provide accurate relative quantification for F-ratio analyses was investigated, based upon the mass spectral selectivity provided via the F-ratio spectral output. Diesel fuel spiked with 15 analytes at four concentration levels (80, 40, 20, and 10 ppm) produced three sets of two class comparisons that were submitted to tile-based F-ratio analysis to obtain three hit lists, with an F-ratio spectrum for each hit. A novel algorithm which calculates the signal ratio (S-ratio) between two classes (eg., 80 ppm versus 40 ppm) was applied to all mass channels (*m/z*) in the F-ratio spectrum for each hit. A lack of fit (*LOF*) metric was utilized as a measure of peak purity and combined with F-ratio and p-values to study the relationship of each of these

metrics with *m/z* purity. Application of a *LOF* threshold coupled with a p-value threshold yielded a subset of the most pure *m/z* for each of the 15 spiked analytes, evident by the low deviations (< 5%) in S-ratio relative to the true concentration ratio. A key outcome of this study was to demonstrate the isolation of pure *m/z* without the need for higher level signal decomposition algorithms.

1.3.3 *Control-Normalized Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data for Enhanced Biomarker Discovery in a Metabolomic Study of Orthopedic Injury*

Comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC×GC-TOFMS) is followed by discovery-based chemometrics to investigate the changes in the metabolome in human plasma for patients with injury to their anterior cruciate ligament (ACL) as compared to healthy controls. Fisher ratio (F-ratio) analysis provides a supervised untargeted discovery of metabolites that express a statistically significant variance between the two sample classes: patients and controls. Generally, the standard F-ratio is applied, defined as the class-to-class variance relative to the pooled within-class variance. However, it is hypothesized that the standard F-ratio will be adversely impacted by metabolites expressed with an unusually large within-class variance in the patient class. Essentially, while the analyst seeks to understand what the overarching differences in the metabolome are between the patient and control classes, they also seek to uncover interesting and potentially relevant patient-to-patient differences within that class. The standard F-ratio does not adequately provide insight into this later concern. To address this issue, we also apply a modified F-ratio (control-normalized) definition, which provides complementary information to the standard F-ratio, by normalizing the class-to-class variance by the variance of the control class only, instead of normalizing to the pooled within-

class variance. Thirty plasma samples were collected from patients with ACL injuries and thirty plasma samples from matched controls were collected at two time points, time-of-injury and 12 months post-surgery. Following sample preparation, the data collected using GC×GC-TOFMS was submitted to both standard and control-normalized F-ratio analysis, and t-tested for statistical significance. Twenty-one metabolites were discovered at the time-of-injury via standard F-ratio analysis, with the addition of fourteen metabolites from control-normalized F-ratio analysis. Similarly, at 12 months post-surgery, twenty-three metabolites were discovered by standard F-ratio analysis, and fifteen additional metabolites were uncovered by control-normalized F-ratio analysis. Collectively, the complementary metabolite information ascertained by way of standard and control-normalized F-ratio analysis culminates in a more informative patients versus controls class separation.

1.4 REFERENCES

[1] C. Poole, The Essence of Chromatography, 1st ed., Elsevier, 2003.

[2] L.S. ETTRE, The Development of Chromatography, Anal. Chem. 43 (1971) 20A-31A. https://doi.org/10.1021/ac60308a718.

[3] K. Robards, P.R. Haddad, P.E. Jackson, Principles and Practice of Modern Chromatographic Methods, 1st ed., Elsevier, 2004. https://www.elsevier.com/books/principles-and-practice-of-modern-chromatographic-methods/robards/978-0-08-057178-2 (accessed June 17, 2020).

[4] IUPAC chromatography (C01075), (n.d.). https://doi.org/10.1351/goldbook.C01075.

[5] Liquid Chromatography - 2nd Edition, (n.d.). https://www.elsevier.com/books/liquid-chromatography/fanali/978-0-12-805393-5 (accessed June 17, 2020).

[6] J.C. Giddings, Unified Separation Science, John Wiley & Sons, Inc., 1991. https://www.wiley.com/en-us/Unified+Separation+Science-p-9780471520894 (accessed June 17, 2020).

[7] M.C. Henry, C.R. Yonker, Supercritical Fluid Chromatography, Pressurized Liquid Extraction, and Supercritical Fluid Extraction, Anal. Chem. 78 (2006) 3909–3916. https://doi.org/10.1021/ac0605703.

[8] X. Liu, A. Zhao, A. Zhang, H. Liu, W. Xiao, C. Wang, X. Wang, Dispersive liquid–liquid microextraction and gas chromatography–mass spectrometry determination of polychlorinated biphenyls and polybrominated diphenyl ethers in milk, J. Sep. Sci. 34 (2011) 1084–1090. https://doi.org/10.1002/jssc.201000767.

[9] M. Chai, J. Pawliszyn, Analysis of Environmental Air Samples by Solid-Phase Microextraction and Gas Chromatography/Ion Trap Mass Spectrometry, Environ. Sci. Technol. 29 (1995) 693–701. https://doi.org/10.1021/es00003a017.

[10] A. Rodrigues, M. Yegles, N. Van Elsué, S. Schneider, Determination of cannabinoids in hair of CBD rich extracts consumers using gas chromatography with tandem mass spectrometry (GC/MS–MS), Forensic Sci. Int. 292 (2018) 163–166. https://doi.org/10.1016/j.forsciint.2018.09.015.

[11] K.D.B. Bezemer, M. Koeberg, A.E.D.M. van der Heijden, C.A. van Driel, C. Blaga, J. Bruinsma, A.C. van Asten, The Potential of Isotope Ratio Mass Spectrometry (IRMS) and Gas Chromatography-IRMS Analysis of Triacetone Triperoxide in Forensic Explosives Investigations, J. Forensic Sci. 61 (2016) 1198–1207. https://doi.org/10.1111/1556-4029.13135.

[12] L.A. Nisbet, F.M. Wylie, B.K. Logan, K.S. Scott, Gas Chromatography-Mass Spectrometry Method for the Quantitative Identification of 23 New Psychoactive Substances in Blood and Urine, J. Anal. Toxicol. 43 (2019) 346–352. https://doi.org/10.1093/jat/bky109.

[13] S. Mao, C. Lu, M. Li, Y. Ye, X. Wei, H. Tong, Identification of key aromatic compounds in Congou black tea by partial least-square regression with variable importance of projection scores and gas chromatography–mass spectrometry/gas chromatography–olfactometry, J. Sci. Food Agric. 98 (2018) 5278–5286. https://doi.org/10.1002/jsfa.9066.

[14] N. Chung, Y. Jo, M.-H. Joe, M.-H. Jeong, Y.-J. Jeong, J.-H. Kwon, Rice vinegars of different origins: discriminative characteristics based on solid-phase microextraction and gas chromatography with mass spectrometry, an electronic nose, electronic tongue and sensory evaluation, J. Inst. Brew. 123 (2017) 159–166. https://doi.org/10.1002/jib.406.

[15] N. Homer, S. Kothiya, A. Rutter, B.R. Walker, R. Andrew, Gas chromatography tandem mass spectrometry offers advantages for urinary steroids analysis, Anal. Biochem. 538 (2017) 34–37. https://doi.org/10.1016/j.ab.2017.09.002.

[16] A. Leghissa, Z.L. Hildenbrand, F.W. Foss, K.A. Schug, Determination of the metabolites of Δ9-tetrahydrocannabinol in urine and plasma using multiple reaction monitoring gas chromatography with triple quadrupole mass spectrometry, Sep. Sci. PLUS. 1 (2018) 43–47. https://doi.org/10.1002/sscp.201700006.

[17] Nicholas H. Snow, Basic Multidimensional Gas Chromatography, Elsevier, 2020.

[18] K.D. Bartle, P. Myers, History of gas chromatography, TrAC Trends Anal. Chem. 21 (2002) 547–557. https://doi.org/10.1016/S0165-9936(02)00806-3.

[19] Hans-Joachim Hubsschmann, Handbook of GC-MS: Fundamentals and Applications, Wiley, 2015.

[20] Fulton G. Kitson, Barbara S. Larsen, Charles N. McEwen, Gas Chromatography and Mass Spectrometry: A Practical Guide, Academic Press, 1996.

[21] Christopher G. Herbert, Robert A.W. Johnstone, Mass Spectrometry Basics, CRC Press, 2003.

[22] B.A. Parsons, D.K. Pinkerton, R.E. Synovec, Implications of Phase Ratio for Maximixing Peak Capacity in Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry, J. Chromatogr. A. (2016).

[23] D.V. Gough, H.D. Bahaghighat, R.E. Synovec, Column Selection Approach to Achieve a High Peak Capacity in Comprehensive Three-Dimensional Gas Chromatography, Talanta. (2018). https://doi.org/10.1016/j.talanta.2018.12.007.

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

[25] 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. https://doi.org/10.1016/j.chroma.2015.01.031.

[26] 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. https://doi.org/10.1016/j.chroma.2016.11.015.

[27] S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, R.E. Synovec, Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications, Anal. Chem. 90 (2018) 505–532. https://doi.org/10.1021/acs.analchem.7b04226.

[28] S.E. Prebihalo, D.K. Pinkerton, R.E. Synovec, Impact of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry experimental design on data trilinearity and parallel factor analysis deconvolution, J. Chromatogr. A. (2019) 460368. https://doi.org/10.1016/j.chroma.2019.460368.

[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. https://doi.org/10.1016/j.aca.2015.02.040.

[30] A. Lelevic, V. Souchon, M. Moreaud, C. Lorentz, C. Geantet, Gas chromatography vacuum ultraviolet spectroscopy: A review, J. Sep. Sci. 43 (2020) 150–173. https://doi.org/10.1002/jssc.201900770.

[31] A.D. McNaught, A. Wilkinson, IUPAC. Compendium of Chemical Terminology (the "Gold Book"), 2nd ed., Blackwell Scientific Publications, Oxford, 1997.

[32] J. Dekeirsschieter, P.-H. Stefanuto, C. Brasseur, E. Haubruge, J.-F. Focant, Enhanced Characterization of the Smell of Death by Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry (GCxGC-TOFMS), PLOS ONE. 7 (2012) e39005. https://doi.org/10.1371/journal.pone.0039005.

[33] R. Madsen, T. Lundstedt, J. Trygg, Chemometrics in metabolomics—A review in human disease diagnosis, Anal. Chim. Acta. 659 (2010) 23–33. https://doi.org/10.1016/j.aca.2009.11.042.

[34] 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. https://doi.org/10.1016/j.chroma.2012.05.050.

[35] H. Parastar, M. Jalali-Heravi, R. Tauler, Comprehensive two-dimensional gas chromatography (GC×GC) retention time shift correction and modeling using bilinear peak alignment, correlation optimized shifting and multivariate curve resolution, Chemom. Intell. Lab. Syst. 117 (2012) 80–91. https://doi.org/10.1016/j.chemolab.2012.02.003.

[36] C. Couprie, L. Duval, M. Moreaud, S. Hénon, M. Tebib, V. Souchon, BARCHAN: Blob Alignment for Robust CHromatographic ANalysis, J. Chromatogr. A. 1484 (2017) 65–72. https://doi.org/10.1016/j.chroma.2017.01.003.

[37] T.-F. Tian, S.-Y. Wang, T.-C. Kuo, C.-E. Tan, G.-Y. Chen, C.-H. Kuo, C.-H.S. Chen, C.-C. Chan, O.A. Lin, Y.J. Tseng, Web Server for Peak Detection, Baseline Correction, and

Alignment in Two-Dimensional Gas Chromatography Mass Spectrometry-Based Metabolomics Data, Anal. Chem. 88 (2016) 10395–10403. https://doi.org/10.1021/acs.analchem.6b00755.

[38] J. Gros, D. Nabi, P. Dimitriou-Christidis, R. Rutler, J.S. Arey, Robust Algorithm for Aligning Two-Dimensional Chromatograms, Anal. Chem. 84 (2012) 9033–9040. https://doi.org/10.1021/ac301367s.

[39] Y. Zushi, J. Gros, Q. Tao, S.E. Reichenbach, S. Hashimoto, J.S. Arey, Pixel-by-pixel correction of retention time shifts in chromatograms from comprehensive two-dimensional gas chromatography coupled to high resolution time-of-flight mass spectrometry, J. Chromatogr. A. 1508 (2017) 121–129. https://doi.org/10.1016/j.chroma.2017.05.065.

[40] 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. https://doi.org/10.1016/j.chroma.2015.03.001.

[41] B. Egert, C.H. Weinert, S.E. Kulling, A peaklet-based generic strategy for the untargeted analysis of comprehensive two-dimensional gas chromatography mass spectrometry data sets, J. Chromatogr. A. 1405 (2015) 168–177. https://doi.org/10.1016/j.chroma.2015.05.056.

[42] S.E. Reichenbach, D.W. Rempe, Q. Tao, D. Bressanello, E. Liberto, C. Bicchi, S. Balducci, C. Cordero, Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors, Anal. Chem. 87 (2015) 10056–10063. https://doi.org/10.1021/acs.analchem.5b02718.

[43] D.W. Rempe, S.E. Reichenbach, Q. Tao, C. Cordero, W.E. Rathbun, C.A. Zini, Effectiveness of Global, Low-Degree Polynomial Transformations for GCxGC Data Alignment, Anal. Chem. 88 (2016) 10028–10035. https://doi.org/10.1021/acs.analchem.6b02254.

[44] S. Furbo, A.B. Hansen, T. Skov, J.H. Christensen, Pixel-Based Analysis of Comprehensive Two-Dimensional Gas Chromatograms (Color Plots) of Petroleum: A Tutorial, Anal. Chem. 86 (2014) 7160–7170. https://doi.org/10.1021/ac403650d.

[45] N. Hoffmann, M. Wilhelm, A. Doebbe, K. Niehaus, J. Stoye, BiPACE 2D--graph-based multiple alignment for comprehensive 2D gas chromatography-mass spectrometry, Bioinforma. Oxf. Engl. 30 (2014) 988–995. https://doi.org/10.1093/bioinformatics/btt738.

[46] B. Deng, S. Kim, H. Li, E. Heath, X. Zhang, Global peak alignment for comprehensive two-dimensional gas chromatography mass spectrometry using point matching algorithms, J. Bioinform. Comput. Biol. 14 (2016) 1650032. https://doi.org/10.1142/S0219720016500323.

[47] A.C. Olivieri, H.-L. Wu, R.-Q. Yu, MVC3: A MATLAB graphical interface toolbox for third-order multivariate calibration, Chemom. Intell. Lab. Syst. 116 (2012) 9–16. https://doi.org/10.1016/j.chemolab.2012.03.018.

[48] A. Eftekhari, H. Parastar, Multivariate analytical figures of merit as a metric for evaluation of quantitative measurements using comprehensive two-dimensional gas chromatography–mass spectrometry, J. Chromatogr. A. 1466 (2016) 155–165. https://doi.org/10.1016/j.chroma.2016.09.016.

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

[50] 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. https://doi.org/10.1016/j.trac.2008.05.011.

[51] 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. https://doi.org/10.1016/j.chroma.2009.02.049.

[52] L.G. Johnsen, J.M. Amigo, T. Skov, R. Bro, Automated resolution of overlapping peaks in chromatographic data: Chromatographic data analysis, J. Chemom. 28 (2014) 71–82. https://doi.org/10.1002/cem.2575.

[53] H. Parastar, R. Tauler, Multivariate Curve Resolution of Hyphenated and Multidimensional Chromatographic Measurements: A New Insight to Address Current Chromatographic Challenges, Anal. Chem. 86 (2014) 286–297. https://doi.org/10.1021/ac402377d.

[54] A. de Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods. 6 (2014) 4964–4976. https://doi.org/10.1039/C4AY00571F.

[55] R.J. Pell, X. Chen, Multivariate curve resolution for understanding complex reactions: MCR for understanding complex reactions, J. Chemom. 28 (2014) 411–419. https://doi.org/10.1002/cem.2507.

[56] N.G.S. Mogollon, F.A. de L. Ribeiro, M.M. Lopez, L.W. Hantao, R.J. Poppi, F. Augusto, Quantitative analysis of biodiesel in blends of biodiesel and conventional diesel by comprehensive two-dimensional gas chromatography and multivariate curve resolution, Anal. Chim. Acta. 796 (2013) 130–136. https://doi.org/10.1016/j.aca.2013.07.071.

[57] 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. https://doi.org/10.1016/j.chroma.2014.10.057.

[58] J.E. Welke, V. Manfroi, M. Zanus, M. Lazzarotto, C. Alcaraz Zini, Differentiation of wines according to grape variety using multivariate analysis of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection data, Food Chem. 141 (2013) 3897–3905. https://doi.org/10.1016/j.foodchem.2013.06.100.

[59] L.W. Hantao, B.R. Toledo, F.A. de Lima Ribeiro, M. Pizetta, C.G. Pierozzi, E.L. Furtado, F. Augusto, Comprehensive two-dimensional gas chromatography combined to multivariate data analysis for detection of disease-resistant clones of Eucalyptus, Talanta. 116 (2013) 1079–1084. https://doi.org/10.1016/j.talanta.2013.08.033.

[60] A.P. de la Mata, R.H. McQueen, S.L. Nam, J.J. Harynuk, Comprehensive two-dimensional gas chromatographic profiling and chemometric interpretation of the volatile profiles of sweat in knit fabrics, Anal. Bioanal. Chem. 409 (2017) 1905–1913. https://doi.org/10.1007/s00216-016-0137-1.

[61] L.F. Oliveira, S.C.G.N. Braga, F. Augusto, J.C. Hashimoto, P. Efraim, R.J. Poppi, Differentiation of cocoa nibs from distinct origins using comprehensive two-dimensional gas chromatography and multivariate analysis, Food Res. Int. 90 (2016) 133–138. https://doi.org/10.1016/j.foodres.2016.10.047.

[62] P. Armstrong, K.D. Nizio, K.A. Perrault, S.L. Forbes, Establishing the volatile profile of pig carcasses as analogues for human decomposition during the early postmortem period, Heliyon. 2 (2016) e00070. https://doi.org/10.1016/j.heliyon.2016.e00070.

[63] K.D. Nizio, M. Ueland, B.H. Stuart, S.L. Forbes, The analysis of textiles associated with decomposing remains as a natural training aid for cadaver-detection dogs, Forensic Chem. 5 (2017) 33–45. https://doi.org/10.1016/j.forc.2017.06.002.

[64] L.M. Dubois, K.A. Perrault, P.-H. Stefanuto, S. Koschinski, M. Edwards, L. McGregor, J.-F. Focant, Thermal desorption comprehensive two-dimensional gas chromatography coupled to variable-energy electron ionization time-of-flight mass spectrometry for monitoring subtle changes in volatile organic compound profiles of human blood, J. Chromatogr. A. 1501 (2017) 117–127. https://doi.org/10.1016/j.chroma.2017.04.026.

[65] T.R. Mellors, L. Blanchet, J.L. Flynn, J. Tomko, M. O'Malley, C.A. Scanga, P.L. Lin, J.E. Hill, A new method to evaluate macaque health using exhaled breath: A case study of *M. tuberculosis* in a BSL-3 setting, J. Appl. Physiol. 122 (2017) 695–701. https://doi.org/10.1152/japplphysiol.00888.2016.

[66] P.S. Prata, G.L. Alexandrino, N.G.S. Mogollón, F. Augusto, Discriminating Brazilian crude oils using comprehensive two-dimensional gas chromatography–mass spectrometry and multiway principal component analysis, J. Chromatogr. A. 1472 (2016) 99–106. https://doi.org/10.1016/j.chroma.2016.10.044.

[67] P.F. de Lima, M.F. Furlan, F.A. de Lima Ribeiro, S.F. Pascholati, F. Augusto, In vivo determination of the volatile metabolites of saprotroph fungi by comprehensive two-dimensional gas chromatography: Gas Chromatography, J. Sep. Sci. 38 (2015) 1924–1932. https://doi.org/10.1002/jssc.201401404.

[68] N.G.S. Mogollón, F.A. de L. Ribeiro, R.J. Poppi, A. Quintana, J. Chávez, D. Agualongo, H. Aleme, F. Augusto, Exploratory Analysis of Biodiesel by Combining Comprehensive Two-Dimensional Gas Chromatography and Multiway Principal Component Analysis, J. Braz. Chem. Soc. (2016). https://doi.org/10.21577/0103-5053.20160222.

[69] R.A. Fisher, Statistical Methods for Research Workers, in: S. Kotz, N.L. Johnson (Eds.), Breakthr. Stat. Methodol. Distrib., Springer, New York, NY, 1992: pp. 66–70. https://doi.org/10.1007/978-1-4612-4380-9_6.

[70] 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. https://doi.org/10.1021/ac0602625.

[71] 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. https://doi.org/10.1016/j.chroma.2016.06.067.

[72] 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. https://doi.org/10.1016/j.talanta.2013.06.038.

[73] 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. https://doi.org/10.1021/ac504472s.

[74] 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. https://doi.org/10.1016/j.chroma.2016.02.067.

[75] B.C. Reaser, B.W. Wright, R.E. Synovec, Using Receiver Operating Characteristic Curves To Optimize Discovery-Based Software with Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry, Anal. Chem. 89 (2017) 3606–3612. https://doi.org/10.1021/acs.analchem.6b04991.

Chapter 2. Impact of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Experimental Design on Data Trilinearity and Parallel Factor Analysis Deconvolution

This chapter was reproduced from Sarah E. Prebihalo, David K. Pinkerton, and Robert E. Synovec, "Impact of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry experimental design on data trilinearity and parallel factor analysis deconvolution" *J. Chromatogr. A* 1605 (2019).

2.1 INTRODUCTION

Comprehensive two-dimensional gas chromatography (GC×GC) is an advantageous method for the analysis of a variety of complex sample types including diesel fuel, waste water, and fire debris [1–14]. While one-dimensional GC (1D-GC) has many beneficial applications [15–17], it often does not provide sufficient chemical selectivity and resolving power for complex samples. As a result, there is an increased potential to have a significantly higher fraction of chromatographically overlapped analytes within each separation. With the addition of another chromatographic dimension, GC×GC can produce a higher peak capacity by providing another degree of chemical selectivity through a second column containing a complementary stationary phase [18]. When coupled with time-of-flight mass spectrometry (TOFMS), GC×GC-TOFMS provides improved identification and quantification of trace level analytes, resulting in an information-rich three-dimensional data cube for each sample.

Due to the complexity of GC×GC-TOFMS data, data processing software has been developed to implement chemometric techniques [19–21], providing analyte deconvolution, identification and quantification. Parallel Factor Analysis (PARAFAC) has been demonstrated as a useful deconvolution technique for sufficiently trilinear, higher order dimensional data (third-order or higher) [20–22]. For GC×GC-TOFMS, the third-order data is traditionally the two

separation dimensions, with the third dimension the mass spectral dimension. Despite numerous advantages, PARAFAC is based upon a strict trilinearity requirement [20], and quantification is adversely impacted for non-trilinear data [23–25]. On the other hand, PARAFAC has also been successfully applied in numerous studies using GC×GC-TOFMS [5,7,21,26]. Hence, a cause-and-effect study of how data trilinearity impacts PARAFAC performance such as quantitative accuracy is needed in order to shed light on this issue [27]. Furthermore, how the data trilinearity relates to the applied GC×GC separation conditions needs to be critically assessed.

Trilinear and/or non-trilinear data in GC×GC will primarily arise from issues associated with the chromatographic conditions. Much like 1D-GC, GC×GC is affected by the general elution problem (GEP), whereby later eluting peaks are more broadened unless a temperature program is applied [28]. A temperature program (i.e. 5-10 °C/min) is used to address the GEP on the first dimension, ^{1}D, separation. Due to the nature of how GC×GC separations are performed, the second dimension, ^{2}D, separation is subject to the same oven temperature program as ^{1}D, or an offset temperature from ^{1}D. However, the time of each ^{2}D separation is relatively short in comparison with each modulation period, P_M, typically on the order of 1 to 10 s. This results in pseudo-isothermal ^{2}D separations as there is a relatively small temperature change during each P_M. Consequently, the ^{2}D peak widths, $^{2}W_b$, increase with increasing retention, $^{2}k'$.

While each individual ^{2}D separation is pseudo-isothermal, each successive modulation also occurs at a temperature slightly greater than the one preceding it. Since the ^{2}D retention time, $^{2}t_R$, is dependent on temperature, there is a negative shift in ^{2}D retention time, $\Delta^{2}t_R$, from one modulation to the next. Due to the $\Delta^{2}t_R$, the data are no longer fully trilinear, potentially introducing bias in utilizing PARAFAC as a deconvolution method. Thus, it is important to study the relationship between data trilinearity and the accuracy of PARAFAC-based quantification. The

quantity referred to as the trilinearity deviation ratio (*TDR*) was previously introduced by Pinkerton et al. as a metric to study this issue [23], whereby the *TDR* is defined as $\Delta^2 t_R$ divided by ${}^2W_b$ for a given analyte. The theory for *TDR* is provided below and related to an experimental evaluation of how *TDR* relates to PARAFAC accuracy. Our underlying hypothesis is that although PARAFAC strictly speaking is based upon a trilinear data requirement, if the shifting introduced by the GC×GC experimental design is sufficiently small (for small *TDR*), then the data will be "sufficiently trilinear" for the purposes of applying PARAFAC, and any quantitative bias introduced by *TDR* will be essentially insignificant in relation to other sources of noise in the data, eg., *S/N*, sample-to-sample injection precision, etc.

This study is aimed to experimentally validate the use of *TDR* as a prediction tool for PARAFAC performance, and to provide the GC×GC practitioner with guidance toward selecting separation conditions to achieve accurate quantification with PARAFAC. One issue studied is the ^{1}D and ^{2}D column selection. Column selection influences the resulting ${}^2k'$ range, which in turn influences the P_M selection, and thus the *TDR* range obtained. The P_M which most efficiently utilizes the ^{2}D separation space without causing significant wraparound was considered the benchmark, for which longer P_M were compared. In turn, how temperature change, ΔT, from one modulation to the next influences the *TDR* range obtained is also investigated. Here, wraparound is defined as the condition in which a ^{2}D peak elutes later than less retained ^{2}D peaks from a subsequent modulation. All of these considerations are ultimately described in terms of PARAFAC quantification accuracy.

2.2 THEORY

2.2.1 *Trilinearity Deviation Ratio*

PARAFAC quantification accuracy has been demonstrated to be related to both $\Delta^2 t_R$ and ${}^2W_b$, such that peaks with a wider W_b are less influenced by shifts in retention time [23]. The trilinear deviation ratio can be mathematically defined as,

$$TDR = \frac{\Delta^2 t_R}{{}^2W_b} \tag{2.1}$$

The $\Delta^2 t_R$ is derived starting with the van't Hoff equation [28] for ^{2}D separations,

$$\text{Ln } {}^2k' = \ln\left(\frac{{}^2t_R - {}^2t_0}{{}^2t_0}\right) = \frac{\Delta H}{RT} + \frac{\Delta S^\circ}{R} + \ln {}^2\beta \tag{2.2}$$

where ${}^2k$' is the analyte retention factor, ${}^2t_R$ is the analyte retention time, ${}^2t_0$ is the dead time, ΔH is the analyte enthalpy of vaporization, T is the temperature, R is the gas constant, ΔS° is the standard entropy of transfer of the analyte from the mobile to stationary phase, and ${}^2\beta$ is the ^{2}D phase volume ratio. For successive modulations, N+1 relative to N, from ^{1}D onto ^{2}D,

$$\ln {}^2k'_{N+1} - \ln {}^2k'_N = \left(\frac{\Delta H}{RT_{N+1}} + \frac{\Delta S^\circ}{R} + \ln {}^2\beta\right) - \left(\frac{\Delta H}{RT_N} + \frac{\Delta S^\circ}{R} + \ln {}^2\beta\right) = \frac{\Delta H}{R}\left(\frac{1}{T_{N+1}} - \frac{1}{T_N}\right) \tag{2.3}$$

Combining Eqs. (2.2) and (2.3) for the successive modulations,

$$\text{Ln}\left(\frac{{}^2t_{R,N+1} - {}^2t_0}{{}^2t_0}\right) - \ln\left(\frac{{}^2t_{R,N} - {}^2t_0}{{}^2t_0}\right) = \frac{\Delta H}{R}\left(\frac{T_N - T_{N+1}}{T_{N+1}T_N}\right) \tag{2.4}$$

Using the laws of logarithms, rearranging, and taking the exponential function,

$$\frac{{}^2t_{R,N+1} - {}^2t_0}{{}^2t_{R,N} - {}^2t_0} = e^{\left(\frac{\Delta H}{R}\left(\frac{T_N - T_{N+1}}{T_{N+1}T_N}\right)\right)} \tag{2.5}$$

Since $T_{N+1} > T_N$, then ${}^2t_{R,\,N+1} < {}^2t_{R,\,N}$. We define $\Delta T = T_{N+1} - T_N$ as the temperature change, and $\Delta^2 t_R = {}^2t_{R,\,N+1} - {}^2t_{R,\,N}$ as the analyte retention time change for a successive modulation. Since $\Delta T <<$

$T_{N+1}T_N$, then $T_{N+1}T_N \approx T_N^2$ in the denominator of the exponential term, and then substitution of ΔT and $\Delta^2 t_R$ into Eq. (2.5) yields,

$$\frac{\left({}^2t_{R,N} + {}^2\Delta t_R - {}^2t_0\right)}{{}^2t_{R,N} - {}^2t_0} = e^{\left(\frac{-\Delta H}{R}\left(\frac{\Delta T}{T_N^2}\right)\right)} \tag{2.6}$$

Applying the approximation $e^x \approx 1 + x$, that is justified since ΔT is sufficiently small, and consolidating terms on the left side of Eq. (2.6),

$$1 + \left(\frac{\Delta^2 t_R}{{}^2t_{R,N} - {}^2t_0}\right) = 1 - \frac{\Delta H}{R} \cdot \frac{\Delta T}{T_N^2} \tag{2.7}$$

Solving for $\Delta^2 t_R$, and since ${}^2t_{R,N} - {}^2t_0 = {}^2k_N' \cdot {}^2t_0$

$$\Delta^2 t_R = \frac{-\Delta H}{R}\left(\frac{\Delta T}{T_N^2}\right)\left({}^2k_N' \cdot {}^2t_0\right) \tag{2.8}$$

Note that ΔT, is also experimentally equivalent to the product of the ^{1}D oven temperature program T_{ramp} in °C/min, and the modulation period, P_M,

$$\Delta T = T_{ramp} \cdot P_M \tag{2.9}$$

Finally, since each ^{2}D separation is nominally pseudo-isothermal, ${}^2W_b$ is observed experimentally to follow the expression,

$${}^2W_b \approx C\ {}^2k' + {}^2W_{b,0} \tag{2.10}$$

where C is the slope of the ${}^2W_b$ versus ${}^2k'$ plot, and is dominated by on-column band broadening due to analyte mass transfer in the mobile phase, and ${}^2W_{b,0}$ is the minimum peak width of the analyte at $k' = 0$ [100].

2.2.2 *Beta Ratio, the Ratio of Column Phase Volume Ratios*

The *TDR* depends on ${}^2k'$ such that analytes with higher ${}^2k'$ are subjected to increased ${}^2W_b$ via Eq. (2.10) and more substantial $\Delta^2 t_R$ as the oven temperature rises via Eq. (2.8). It is therefore necessary to study how GC×GC column sets, that produce varying ${}^2k'$ ranges, relate to PARAFAC

quantification accuracy. The *k*' for a given analyte relates to the phase volume ratio, β, which is the ratio of the mobile phase volume, V_m, to the stationary phase volume, V_s,

$$\beta = \frac{V_m}{V_s} = \frac{d_c}{4 \cdot d_f} \tag{2.11}$$

where d_c is the inner diameter of the column and d_f is the stationary phase film thickness. Common β for commercially available columns include 250 (d_c: 250 μm, d_f: 0.25 μm), 125 (d_c: 250 μm, d_f: 0.50 μm) and 225 (d_c: 180 μm, d_f: 0.2 μm), which are the β utilized in this study.

For GC×GC separations, it has been demonstrated that manipulation of β for successive dimensions allows the analyst to tune the *k'* range and improve the peak capacity of their GC×GC separation [30–32]. Parsons et al. introduced the variable β_R, or the ratio of the phase volume ratios, between successive dimensions,

$$\beta_R = \frac{^1\beta}{^2\beta} \tag{2.12}$$

where $^1\beta$ is for the ^{1}D separation and $^2\beta$ is for the ^{2}D separation [30]. Through Eqs. (2.1)-(2.12), the interrelationships of $^2k'$, ^2W_b, Δ^2t_R on *TDR* will be elucidated. Selection of an appropriate β_R impacts Δ^2t_R through the $^2k'$ range produced. A larger β_R prompts a larger starting and ending $^2k'$ and longer P_M in order to avoid excessive wraparound. In return, the P_M and T_{ramp} dictate the ΔT per modulation [31]. Note that although Δ^2t_R in Eq. (2.8) is fundamentally a negative value, the absolute value of Δ^2t_R is used in Eq. (2.1) in order to express *TDR* as a positive value, to be consistent with our prior report [23]. A couple of additional points from this previous β_R study pertain to the current study [30]. First, a larger β_R translates into not only a larger $^2k'$ range, but also a larger average ^2W_b, with the larger $^2k'$ range matched with the appropriate P_M to avoid wraparound. The P_M selection impacts the sampling density, ρ_s, defined as the P_M divided by the ^{1}D width-at-base,^1W_b [33–36]. If the $\rho_s \leq 2$, the ^{1}D peaks are "undersampled," which very noticeably increases their apparent ^1W_b in the GC×GC plot, and reduces the overall 2D peak

capacity. Second, and in contrast, application of a smaller β_R results in a smaller ^2k' range and a smaller average ^2W_b, allowing for use of a shorter P_M and a higher and more appropriate ρ_s of the ^{1}D peaks, concurrent with providing a high overall 2D peak capacity [30].

2.3 EXPERIMENTAL

Experimental data utilizing a mixture containing 115 analytes were previously collected using a Pegasus 4D system (LECO, St. Joseph, MI), consisting of an Agilent 6890 gas chromatograph with a 7683 autosampler (Agilent Technologies, Palo Alto, CA) and a Pegasus III TOFMS with a quad-jet thermal modulator and secondary oven [30]. Of the six possible column combinations previously reported [30], two representative column sets were chosen and analyzed in duplicate. For the first column set, herein referred to as "column set A", the ^{1}D column contained a Restek Rtx-5MS stationary phase (20.0 m × 250 µm i.d. ^1d_c × 0.25 µm ^1d_f) and the ^{2}D column contained a Restek Rtx-200 stationary phase (2.0 m × 180 µm i.d. ^2d_c × 0. 2 µm ^2d_f). Utilizing Eqs. (2.11) and (2.12), $^1\beta$ is 250, while $^2\beta$ is 225, yielding a β_R of 1.11. The second column set, herein referred to as "column set B" was identical to column set A with the exception that the ^{1}D column film thickness was doubled with a ^1d_f of 0.50 µm. Column set B has the identical ^{2}D column, so $^2\beta$ is 225. However, use of a thicker film provides a smaller $^1\beta$ of 125. The subsequent β_R of column set B is thus 0.56. More detailed instrumental parameters are listed in the previous report [30]. Relevant conditions for the current study are the ^{1}D oven program (held constant at 40 °C for 1 min, T_{ramp} = 5 °C/min). The secondary oven followed the same temperature program as the primary oven, with a +5°C offset. Mass channels *m/z* 33-300 were collected at 500 spectra/s. Modulation periods P_M of 3 and 6 s were used for column set A (β_R = 1.11), while P_M for column set B (β_R = 0.56) were 1.5 and 3 s, resulting in four sets of GC×GC conditions (Table 2.1), referred to as Expt1, Expt2, Expt3, and Expt4, respectively.

Table 2.1 Summary of relevant experimental separation conditions for Expt1-4. The phase volume ratio, β, and ratio of phase volume ratios, β_R, are defined in Eqs. (2.11) and (2.12), respectively. ^{1}D column length and inner diameter were maintained at 20 m and 250 μm, respectively, for all experiments. Also constant across experiments were the following: ^{2}D column dimensions (2 m × 180 μm × 0.2 μm), T_{ramp} (5 °C/min), *m/z* range (33-300), and data collection frequency (500 spectra/s).

	Column Set	**^{1}D Column Film Thickness (μm)**	**$^1\beta$**	**$^2\beta$**	**β_R**	**P_M (s)**	**ΔT (°C)**
Expt1	A	0.25	250	225	1.11	3	0.250
Expt2	A	0.25	250	225	1.11	6	0.500
Expt3	B	0.50	125	225	0.56	1.5	0.125
Expt4	B	0.50	125	225	0.56	3	0.250

Data analysis and chemometric modeling were performed in Matlab R2016b (The Mathworks Inc., Natick, MA, USA). To accurately measure Δ^2t_R and ^2W_b for *TDR* calculations, curve fitting was performed using the Matlab curve fitting toolbox 8.1.1 on each analyte at selective *m/z*. Twenty-one representative test analytes (Table 2.2) were selected from the total 115 in the test mixture so as to span the full range of ^1t_R and ^2t_R (both measured at the peak maximum for the purposes of the table), and corresponding $^2k'$ for column sets A and B. Each analyte was given a peak number (1-21) for ease of reference. The ^{2}D dead time, $^2t_0$, was calculated for both column sets using the ChromaTOF flow calculator, and determined to be 1.00 s at an oven temperature of 150°C. To obtain the *TDR* via Eq. (2.1), curve fitting was performed on a pair of ^{2}D peaks for each analyte by fitting a Gaussian model to the largest ^{2}D peak and the next largest ^{2}D peak for the purpose of obtaining ^2W_b and Δ^2t_R. Each ^2W_b was obtained by determining the standard deviation provided by each Gaussian fit. Each Δ^2t_R was determined by the following procedure. First, the Gaussian peak maximum represented the ^2t_R of each ^{2}D peak selected for a given analyte. Next, each Δ^2t_R was calculated by taking the difference of the ^2t_R from the largest ^{2}D peak from the ^2t_R of the adjacent ^{2}D peak (the next largest ^{2}D peak, either directly before or after the largest ^{2}D peak) in order to provide the most accuracy in the Δ^2t_R measurement.

The impact of data trilinearity on PARAFAC quantification was studied in the following way. The 21 test analytes over each of the four GC×GC conditions were first analyzed by PARAFAC to obtain the full loadings. Next a selective *m/z* was determined for each analyte strictly for the purpose of comparing the PARAFAC deconvolution results to quantification values obtained for the benchmark peak area using the raw data and the same selective *m/z*. PARAFAC was performed using the PLS Toolbox, version 8.2.1 (Eigenvector Research Inc., Wenatchee, WA, USA) on each analyte, with the data cube consisting of the two separation dimensions, ^{1}D and ^{2}D, and the full mass spectral dimension. Non-negativity constraints were applied to both the ^{2}D peak and mass spectral dimensions, with the stricter unimodality constraint applied to the ^{1}D peak. The number of factors applied corresponded to the most factors necessary to describe the peak, without allowing for splitting (i.e. two or more factors describing the same component). For the sake of comparison, PARAFAC quantification was then performed by reconstructing the trilinear data cube via the outer product of the PARAFAC loadings and summing across all dimensions at a selective *m/z* appropriate for each analyte.

The benchmark peak area for each of the 21 test analytes over each of the four GC×GC conditions was calculated using the raw data and the same selective *m/z* as for PARAFAC quantification. For each replicate and column set, the chromatograms were unfolded with each analyte selected to include all modulations. In order to minimize quantification error in the benchmark peak area, all modulations across the ^{1}D dimension were summed into a single ^{2}D peak that was then baseline corrected and summed. Misalignment of consecutive modulations due to ΔT, as well as baseline effects, were mitigated using this procedure. Finally, a quantity referred to as the PARAFAC %Error was calculated as follows. For each of the 21 test analytes the benchmark peak area was subtracted from the PARAFAC peak area (at the same *m/z*), and then this difference

was divided by the benchmark peak area and the percentage was reported. The PARAFAC %Error is deemed principally due to any error introduced by insufficient trilinearity during PARAFAC computation. Thus, the PARAFAC %Error results can be directly compared to experimentally obtained *TDR* data to study their underlying interrelationship and compare the findings to our prior theoretical study [23].

2.4 Results and Discussion

2.4.1 *GC×GC Chromatograms for Evaluation of TDR Relationship to PARAFAC*

We begin by qualitatively describing and comparing chromatograms in Fig. 2.1 obtained with the four sets of GC×GC conditions (Table 2.1). All chromatograms in Fig. 2.1 except for Fig. 2.1(B) were reregistered on the ^{2}D axis to visually adjust them for partial and/or full wraparound. Hence, the apparent ^2t_R in Fig. 2.1(B) are equal to the true ^2t_R. However, in Figs. 2.1(A) and (C) the true ^2t_R are equal to the apparent ^2t_R plus 1.7 s (i.e., the reregistration adjustment), and in Fig. 2.1(D) the true ^2t_R are equal to the apparent ^2t_R plus 0.8 s. For Expt1, with a β_R of 1.11, a P_M of 3 s and a ΔT of 0.25 °C/modulation, efficient use of the ^{2}D dimension was observed in Fig. 2.1(A), with minimal wraparound.

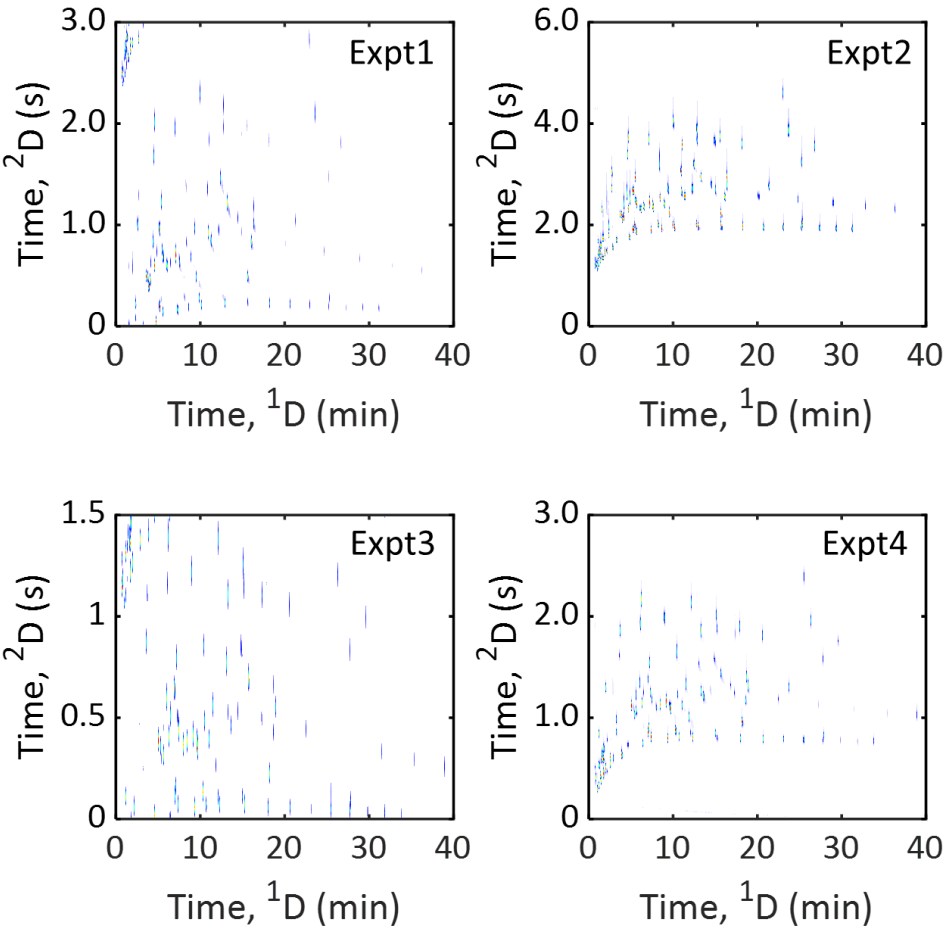

Figure 2.1. GC×GC-TIC chromatograms of the 115 component mixture. Each chromatogram has been registered on ^{2}D to restore the latest eluting analyte (diethyl phthalate) to the top of the chromatogram. Data acquisition was delayed 10 s so peaks appear to elute earlier than they actually do on ^{1}D. For column set A (β_R = 1.11) and column set B (β_R = 0.56), two P_M were studied with each column set (Table 1). (A) Column set A with a P_M of 3 s. (B) Column set A with a P_M of 6 s. (C) Column set B with a P_M of 1.5 s. (D) Column set B with a P_M of 3 s.

It is also instructive to see how many modulations are observed for a representative analyte for each set of conditions. Adamantane was chosen as a representative analyte, since its $^{2}k'$ falls roughly in the middle of the $^{2}k'$ range for all four sets of GC×GC conditions. Expt1 yielded two appreciable modulations, i.e., a sampling density of $\rho_s \sim 2$, as observed in Fig. 2.2(A). In contrast, Expt2 again utilized column set A (β_R = 1.11), but the P_M was doubled relative to Expt1, from 3 s to 6 s, which resulted in ΔT increasing from 0.25 °C/modulation in Expt1 to 0.50 °C/modulation. The resulting chromatogram, provided in Fig. 2.1(B), demonstrated no wraparound but there was

less efficient use of the ^{2}D dimension. Due to the longer P_M, adamantane was observed in Fig. 2.2(B) to contain only one appreciable modulation ($\rho_s \sim 1$), with a second minor modulation. While this sampling density is too low to maintain a truly comprehensive separation [33–37], it is common for practitioners to use such a long P_M. More importantly in the context of this study, we shall see that the long P_M will lead to the least accurate PARAFAC results.

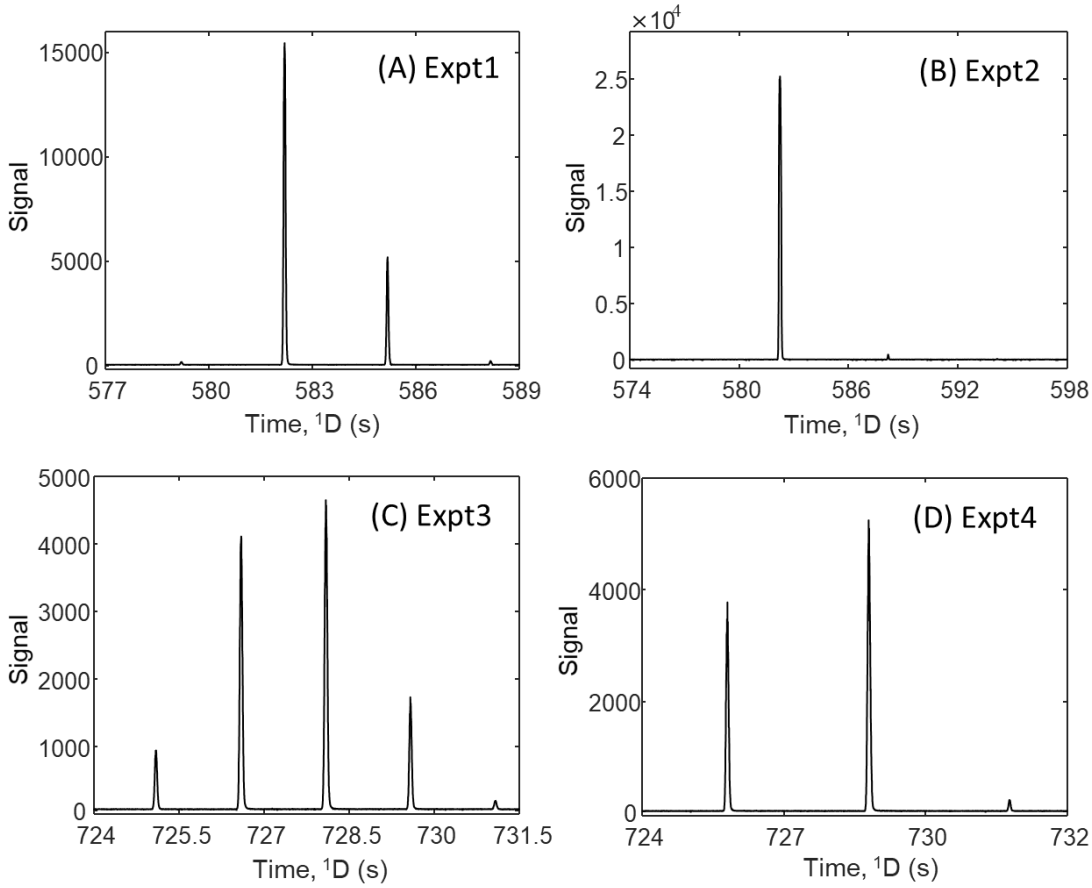

Figure 2.2. Unfolded GC×GC data for adamantane to illustrate impact of the separation conditions on the number of modulations observed. (A) Expt1 conditions yielded two appreciable modulations, so a sampling density $\rho_s \sim 2$. (B) Expt2 conditions produced only one appreciable modulation with a very minor second modulation, $\rho_s \sim 1$. (C) Expt3 conditions resulted in four appreciable modulations, $\rho_s \sim 4$. (D) Expt4 conditions utilized the same P_M as Expt1 and thus had a similar $\rho_s \sim 2$.

Next, we move from a β_R of 1.11 to a β_R of 0.56 using column set B for Expt3 and Expt4 (Table 2.1). To do so, the $^{1}d_f$ was increased from 0.25 μm (column set A) to 0.5 μm. As a result,

analytes require a higher elution temperature for the ^{1}D separation, resulting in less retention on the ^{2}D separation since the ^{2}D column was held constant for the two column sets. Less ^{2}D retention translates into a smaller $^{2}k'$ range that is satisfied by a shorter P_M. Figure 2.1(C) provides a representative chromatogram with a P_M of 1.5 s (ΔT = 0.125 °C/modulation), while demonstrating efficient use of the ^{2}D dimension with modest wraparound. The decrease in P_M with Expt3 relative to Expt1 results in an increase in ρ_s to ~4, with the 4 modulations for adamantane, as presented in Fig. 2.2(C). Thus, these conditions provide the most "comprehensive" GC×GC separation. Again, more importantly, the short P_M of 1.5 s with a concurrent narrow $^{2}k'$ range will be beneficial to provide small *TDRs*, leading to accurate PARAFAC results. Finally, Expt4 also used column set B, but a P_M of 3 s (ΔT = 0.25 °C/modulation as in Expt1) was used instead of 1.5 s with Expt3, such that the ^{2}D separation contained empty chromatographic space, as observed in Fig. 2.1(D). In addition to an increase in ΔT relative to Expt3, the ρ_s for adamantane decreased from ~4 to ~2 with Expt4, as shown in Fig. 2.2(D), with a similar ρ_s to Expt1 in Fig. 2.2(A) since ΔT was the same for Expt1 and Expt4.

Using the separation data presented in Fig. 2.1 for the 21 test analytes in the mixture (Table 2.2), we explore the interconnectivity of β_R and P_M on *TDR*. The magnitude and range of $^{2}k'$ obtained, coupled with the P_M applied, can be critically evaluated for each set of GC×GC conditions via measurements of $\Delta^{2}t_R$ and $^{2}W_{b,}$ per Eqs. (2.8) and (2.10) to ultimately determine the *TDR*s obtained using Eq. (2.1). The *TDR*s will, in turn, be related to the accuracy of PARAFAC quantification for the 21 test analytes.

Table 2.2 Summary of the ^{1}D and ^{2}D retention times as well ^2k' values for the 21 representative test analytes. Each analyte was given a peak number (1-21) for ease of reference and are listed in order of increasing ^1t_R.

		Column Set A			Column Set B		
Peak Number	**Analyte**	**^1t_R (s)**	**^2t_R (s)**	**^2k'**	**^1t_R (s)**	**^2t_R (s)**	**^2k'**
1	2-heptanone	292	3.70	2.700	385.0	2.90	1.906
2	nonane	298	1.74	0.740	400.0	1.50	0.510
3	methylhexanoate	337	2.92	1.920	443.5	2.29	1.300
4	bromobenzene	340	2.46	1.460	454.0	1.99	0.996
5	2,6-dimethlyoctane	346	1.85	0.840	457.0	1.56	0.570
6	1-bromohexane	349	2.38	1.380	461.5	1.93	0.940
7	adamantane	583	2.19	1.190	727.0	1.76	0.770
8	1-undecene	607	1.97	0.970	739.0	1.61	0.620
9	2-nonanone	613	3.98	2.980	740.5	2.88	1.870
10	methyl octanoate	667	3.05	2.020	799.0	2.27	1.280
11	1-bromooctane	691	2.54	1.540	830.5	1.96	0.970
12	napthalene	757	3.18	2.180	904.0	2.34	1.340
13	1-decanol	910	2.61	1.610	1049.5	2.00	1.010
14	2-undecanone	946	3.71	2.710	1087.0	2.65	1.660
15	bicyclohexyl	949	2.18	1.180	1105.0	1.72	0.720
16	2-dodecanone	1102	3.56	2.550	1246.0	2.55	1.570
17	methyl laurate	1288	2.74	1.740	1438.0	2.05	1.060
18	diethyl phthalate	1390	4.54	3.540	1544.5	3.11	2.130
19	hexadecane	1390	1.91	0.900	1543.0	1.54	0.550
20	2-pentadecanone	1525	3.19	2.190	1678.0	2.32	1.330
21	phenathrene	1618	3.49	2.490	1789.0	2.50	1.510

2.4.2 *Trilinear Deviation Ratio (TDR) Measurement*

Measurement of *TDR* is based upon taking the ratio of $\Delta^2 t_R$ by ^2W_b defined in Eq. (2.1). We examine the dependencies of $\Delta^2 t_R$ (and thus ΔT) on $^2k'$ and ^2W_b on $^2k'$ separately, in order to evaluate Eqs. (2.8) and (2.10), respectively. We start with the dependence of ^2W_b on $^2k'$, with the results provided in Fig. 2.3. The two-fold thinner ^1d_f for column set A (β_R = 1.11) relative to column set B (β_R = 0.56) resulted in analytes eluting from the ^{1}D column at a lower temperature. Consequently, analytes enter the ^{2}D column at a lower temperature with column set A, which increases the $^2k'$ range. Since each ^{2}D separation is pseudo-isothermal, the ^2W_b trend with $^2k'$

observed in Fig. 2.3 followed Eq. (2.10) for all four sets of GC×GC conditions [23,29,30]. Expt1 produced ^2W_b from 81 to 245 ms, while with Expt2 the ^2W_b were slightly elevated. The effect of lowering β_R to 0.56 by using a thicker ^1d_f is also evident, with Expt3 and Expt4 producing ^2W_b ranges shifted toward lower $^2k'$ as a result of less time spent on ^{2}D, as well as lower maximum ^2W_b due to the lower $^2k'$ values. Within experimental uncertainty the four sets of experimental conditions produced roughly the same linear ^2W_b dependence on $^2k'$, with the two column set A conditions extended over a two-fold larger $^2k'$ range. Significantly, Δ^2t_R is observed to be linearly dependent on $^2k'$ for all four sets of GC×GC conditions, supporting Eq. (2.8). In studying Δ^2t_R and how it relates to $^2k'$ through the selection of β_R and P_M, a clearer picture of how appropriate selection of ^{1}D and ^{2}D columns affects data trilinearity is revealed. That is, smaller β_R (smaller $^2k'$ range) allows for the use of shorter P_M (smaller ΔT), and results in smaller Δ^2t_R.

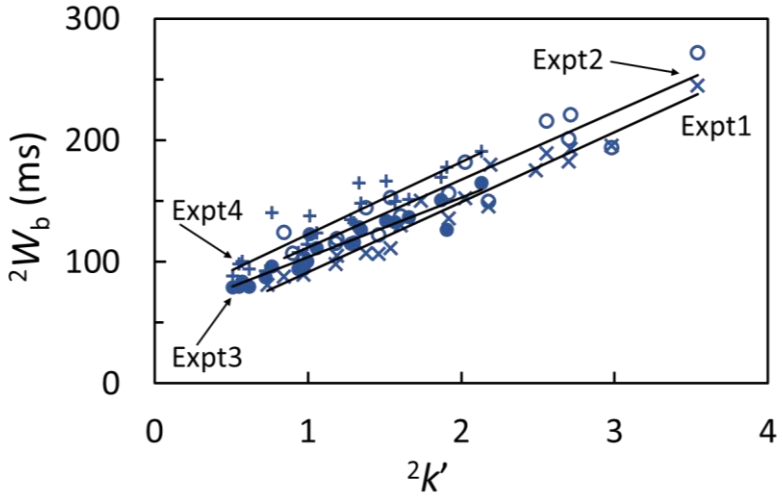

Figure 2.3. ^2W_b measurements as a function of $^2k'$ for Expt1-4, following the linear form of Eq. (2.10). Expt1 (×) produced a ^2W_b range of 81 ms to 245. Expt2 (o) produced a ^2W_b range of 95 ms to 272 ms. Expt3 (●) produced a ^2W_b range of 79 ms to 164 ms. Expt4 (+) produced a ^2W_b range of 88 ms to 191 ms.

Next we examine the dependencies of Δ^2t_R on $^2k'$. Using curve fitting as described in the Experimental section, Δ^2t_R was calculated and plotted against $^2k'$ in Fig. 2.4. Expt1 (β_R = 1.11, P_M = 3 s, ΔT = 0.25 °C/modulation) yielded a Δ^2t_R range of -8.3 to -33.8 ms. With Expt2, when the

P_M is increased to 6 s relative to 3 s (all else constant relative to Expt1), a visual increase in slope and $\Delta^2 t_R$ is observed, and the $\Delta^2 t_R$ range obtained in Expt2 is -19.5 to -98 ms. The increase in $\Delta^2 t_R$ range from Expt1 to Expt2 is due to the doubling of the temperature difference between adjacent modulations (ΔT from 0.25 to 0.5 °C), caused by doubling the P_M, Eqs. (2.8) and (2.9). Decreasing the β_R from 1.11 in Expt1 and Expt2, to 0.56 in Expt3 and Expt4 allows for the use of shorter P_M since the $^2k'$ range was reduced two-fold, and in turn, a smaller ΔT (0.125 and 0.25 °C, respectively). For Expt3 with a P_M of 1.5 s, the $\Delta^2 t_R$ values were extremely small, ranging from -1.1 ms to -8.8 ms. Increasing the P_M from 1.5 s to 3 s with Expt4 produced $\Delta^2 t_R$ ranging from -5.3 to -24.4 ms. Overall, the $^2k'$ range doubled due to doubling β_R from 0.56 to 1.11, and the $\Delta^2 t_R$ range increased when β_R and/or P_M increase.

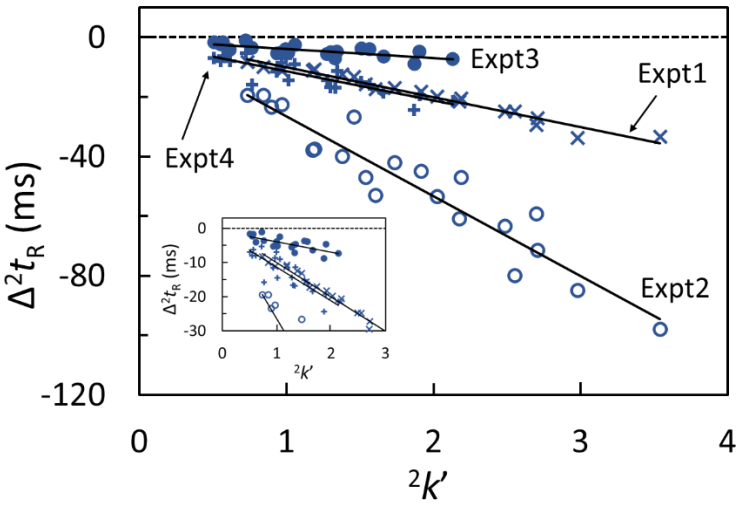

Figure 2.4. $\Delta^2 t_R$ measurements as a function of $^2k'$ for Expt1-4, following Eq. (2.8). Expt1 (×) produced a linear $\Delta^2 t_R$ range (R^2=0.96) from -8.3 ms to -33.8 ms. Expt2 (o) produced a linear $\Delta^2 t_R$ range from -19.5 ms to -98 ms (R^2=0.90). Expt3 (●) gave rise to the smallest $\Delta^2 t_R$ range from -1.1 ms to -8.8 ms (R^2=0.54). Expt4 (+) produced a linear $\Delta^2 t_R$ range (R^2=0.74) from -5.3 to -24.4 ms. The inset figure provides finer detail of Expt3 (●) and Expt4 (+).

Significantly, $\Delta^2 t_R$ is observed to be linearly dependent on $^2k'$ for all four sets of GC×GC conditions, supporting Eq. (2.8). In studying $\Delta^2 t_R$ and how it relates to $^2k'$ through the selection of β_R and P_M, a clearer picture of how appropriate selection of ^{1}D and ^{2}D columns affects data

trilinearity is revealed. That is, smaller β_R (smaller $^2k'$ range) allows for the use of shorter P_M (smaller ΔT), and results in smaller $\Delta^2 t_R$.

Finally, applying Eq. (2.1) using the ^2W_b and $\Delta^2 t_R$ data presented in Figs. 2.3 and 2.4, respectively, the *TDR*s for the four sets of GC×GC conditions are provided in Fig. 2.5. First, we compare Expt1 and Expt2, where Expt1 exhibits a *TDR* range from 0.100 to 0.170, while the *TDR* range is roughly doubled for Expt2, from 0.157 to 0.439. The elevated scatter of *TDR*s for Expt2 is due to inaccuracies in measuring ^2t_R and ^2W_b due to the consequences of undersampling the analytes as presented in Fig. 2.2(B). The *TDR*s double in Expt2 relative to Expt1 due to roughly doubling the $\Delta^2 t_R$ (Fig. 2.4), due to doubling the P_M, which in turn doubled the ΔT. Similarly, the *TDR*s double in Expt4 relative to Expt3 for the same reason. However, with Expt3 and Expt4 the smaller β_R of 0.56 (relative to the β_R of 1.11 with Expt 1 and Expt2) provided two-fold smaller $^2k'$ ranges, which allowed the P_M and ΔT to be cut in half. Notably, the *TDR*s determined in Expt3 were the smallest observed of the four sets of GC×GC conditions, ranging from about 0.013 to 0.057. This is significant because our prior theoretical study indicated that for *TDR* < 0.05 the bias in PARAFAC quantification was less than 1.5% [23]. Also of note is that the *TDR* plots for Expt1 and Expt4 are essentially superimposed with the results for Expt4 extended to higher $^2k'$ and higher *TDR* due to the two-fold difference in β_R.

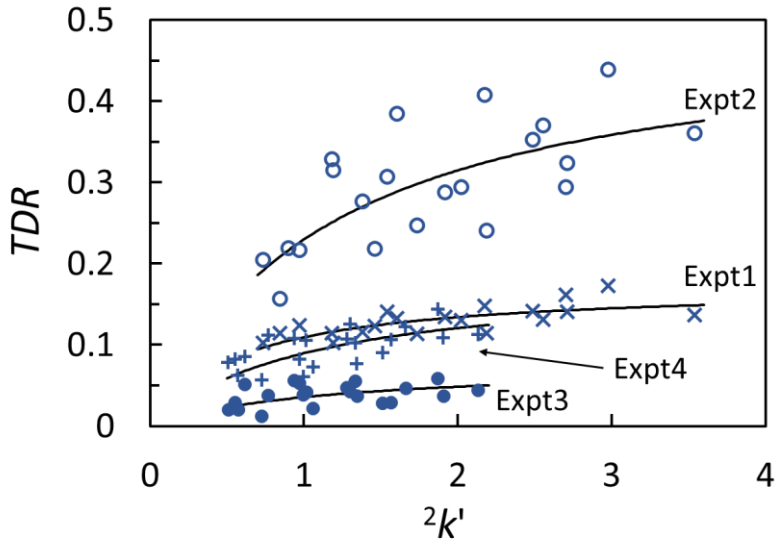

Figure 2.5. *TDR* as a function of ^2k' for Expt1-4, as defined in Eq. (2.1), with each measured $\Delta^2 t_R$ divided by measured ^2W_b. Expt1 (×) yielded *TDR*s ranging 0.100 to 0.170. Expt2 (o) produced *TDR*s considerably higher with a range of 0.157 to 0.439. Expt3 (●) achieved the lowest overall *TDR* range (0.013 to 0.057) as a result of the smaller β_R (relative to Expt1 and Expt2), which allowed P_M and ΔT to be cut in half. Expt4 (+) demonstrated similar *TDR*s to Expt1, with a smaller maximum due to a shorter ^2k' range resulting from a decrease in β_R. The *TDR* curves drawn through each Expt data set were obtained using Eq. (2.1) by taking the function of the best fit line of $\Delta^2 t_R$ versus ^2k' in Fig. 2.4 and dividing that by the function of the best fit line of ^2W_b versus ^2k'in Fig. 2.3.

2.4.3 *PARAFAC Results in the Context of TDR*

The relationship between *TDR* and PARAFAC quantification results are provided in Fig. 2.6, where the PARAFAC %Error as defined in the Experimental section is plotted relative to *TDR* for all four sets of GC×GC conditions. Generally, all of the %Errors are either near zero or are negative as expected, with the %Error worsening with *TDR*, following the theoretical trend previously reported [23]. However, there are some positive values that point to the challenges in accurately quantifying the benchmark peak areas. Additionally, within a given set of GC×GC conditions, there are challenges in accurately determining the *TDR* for each analyte, as previously discussed in Figs. 2.3-2.5, which result in additional scatter in Fig. 2.6. Using column set A, Expt1 and Expt 2 produced the worst %Errors, ranging from +1.3% to -12.5% and from +1.6% to -13.5%,

respectively. The large negative biases correlate to the lack of data trilinearity, with *TDR* ranges from 0.100 to 0.170, and from 0.157 to 0.439, respectively.

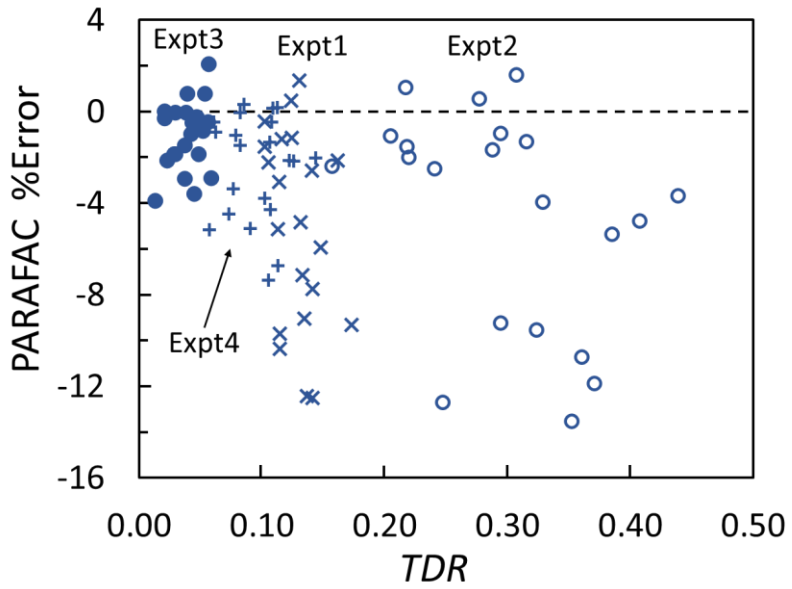

Figure 2.6. PARAFAC %Error as a function of *TDR*. % Error was calculated by comparing the peak area obtained via PARAFAC quantification of a selective *m/z* to that of the benchmark peak area. Expt1 (×) and Expt2 (o) produced the largest PARAFAC %Errors due to the higher *TDR*s, +1.3% to -12.5% and +1.6% to -13.5%, respectively. Expt3 (●) and Expt (+) yielded much lower PARAFAC %Errors. Indeed, with Expt 3 the %Error range was +2.1% to -3.9%, with an average of -1.1% ± 1.4.

The %Error ranges are significantly decreased when using column set B for Expt3 and Expt4. For Expt3, with the shortest P_M of 1.5 s, the lowest overall %Errors were produced for all of the four sets of GC×GC conditions. Additionally, as observed in Fig. 2.6, the range of %Errors for Expt3 is from +2.1% to -3.9%, with an average of -1.1% ± 1.4, and no apparent dependence on *TDR*. This suggests that Expt3 is in a regime where other sources of quantitative uncertainty take an equal if not greater role than *TDR* in determining the quantification error. This is not surprising since the average %Error of -1.1% ± 1.4 is extremely close to zero, i.e., nearly unbiased. Based on this finding we suggest that the GC×GC-TOFMS data are sufficiently trilinear for the purposes of confidently applying PARAFAC for GC×GC conditions as in Expt3. For Expt4, due

to the longer P_M, the %Error range of +0.3% to -7.4% (average of -2.5% ± 2.3) is slightly worse than Expt3, but still much better than Expt1 and Expt2. Also, use of a P_M which efficiently utilizes the ^{2}D separation dimension, as demonstrated for Expt3 in Fig. 2.1(C), has been recognized as an optimal approach for GC×GC [30,31]. The theoretical peak capacity in GC×GC has been determined to be about 10-fold higher than 1D-GC [31,32,38,39]. If the ^{1}D separation has been initially optimized and an appropriate P_M has been selected to not undersample the ^{1}D peaks (i.e., ρ_s ~4 as in Fig. 2.2(C) for Expt3), then the resulting ^{2}D separations will each have a peak capacity of ~10. Indeed, based upon the ^2W_b results in Fig. 2.3 for Expt3, one can readily determine the ^{2}D peak capacity provided in Fig. 2.1(C) is ~12, consistent with experimentally providing the theoretical optimal benefit to total peak capacity with GC×GC while also maintaining data trilinearity.

2.4.4 *Conclusion*

Based upon the results presented in Fig. 2.6, we provide some general guidelines for the practitioner, so PARAFAC can be more optimally applied. The β_R between the two columns in GC×GC [30,33] directly impacts the P_M selection and observed $^2k'$ range, which both impact the Δ^2t_R and ^2W_b range obtained, which in turn determine the *TDR* range obtained. In order to obtain a maximum %Error ≤ -7%, experimental conditions with a *TDR* ≤ 0.1 should be applied. The GC×GC separation conditions to achieve this %Error require a $\beta_R \approx 0.5$, or less, as in Expt3 and Expt4 [5,23,24]. A short P_M is advised (eg., 1 s to 2 s), as this reduces the ΔT. Indeed, this is why Expt 3 performed better than Expt4, and provided a maximum %Error ≤ -4%. However, if the analyst applies GC×GC conditions with $\beta_R \approx 1$, as in Expt1 and Expt 2, then the *TDR* range will be elevated (especially for the longer P_M), and data trilinearity will be impaired for the purpose of

using PARAFAC. Other methods would be advised, such as multivariate curve resolution alternating least squares (MCR-ALS) and PARAFAC2 [24,25,40–43].

2.5 ACKNOWLEDGEMENTS

This study was previously published in the *Journal of Chromatography A*, 1605 (2019).

2.6 REFERENCES

[1] R. Pesesse, P.-H. Stefanuto, F. Schleich, R. Louis, J.-F. Focant, Multimodal chemometric approach for the analysis of human exhaled breath in lung cancer patients by TD-GC × GC-TOFMS, J. Chromatogr. B. 1114–1115 (2019) 146–153. doi:10.1016/j.jchromb.2019.01.029.

[2] M.R. Jacobs, M. Edwards, T. Górecki, P.N. Nesterenko, R.A. Shellie, Evaluation of a miniaturised single-stage thermal modulator for comprehensive two-dimensional gas chromatography of petroleum contaminated soils, J. Chromatogr. A. 1463 (2016) 162–168. doi:10.1016/j.chroma.2016.08.009.

[3] F. Magagna, E. Liberto, S.E. Reichenbach, Q. Tao, A. Carretta, L. Cobelli, M. Giardina, C. Bicchi, C. Cordero, Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from thermal to differential-flow modulated comprehensive two dimensional gas chromatography, J. Chromatogr. A. 1536 (2018) 122–136. doi:10.1016/j.chroma.2017.07.014.

[4] 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.

[5] 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.

[6] B.C. Reaser, B.W. Wright, R.E. Synovec, Using Receiver Operating Characteristic Curves To Optimize Discovery-Based Software with Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry, Anal. Chem. 89 (2017) 3606–3612. doi:10.1021/acs.analchem.6b04991.

[7] C.E. Freye, N.R. Moore, R.E. Synovec, Enhancing the chemical selectivity in discovery-based analysis with tandem ionization time-of-flight mass spectrometry detection for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1537 (2018) 99–108. doi:10.1016/j.chroma.2018.01.008.

[8] S. Prebihalo, A. Brockman, J. Cochran, F.L. Dorman, Determination of emerging contaminants in wastewater utilizing comprehensive two-dimensional gas-chromatography coupled with time-of-flight mass spectrometry, J. Chromatogr. A. 1419 (2015) 109–115. doi:10.1016/j.chroma.2015.09.080.

[9] K.L. Organtini, A.L. Myers, K.J. Jobst, J. Cochran, B. Ross, B. McCarry, E.J. Reiner, F.L. Dorman, Comprehensive characterization of the halogenated dibenzo-p-dioxin and

dibenzofuran contents of residential fire debris using comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry, J. Chromatogr. A. 1369 (2014) 138–146. doi:10.1016/j.chroma.2014.09.088.

[10] D. Megson, R. Kalin, P.J. Worsfold, C. Gauchotte-Lindsay, D.G. Patterson, M.C. Lohan, S. Comber, T.A. Brown, G. O'Sullivan, Fingerprinting polychlorinated biphenyls in environmental samples using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A. 1318 (2013) 276–283. doi:10.1016/j.chroma.2013.10.016.

[11] J.-F. Focant, A. Sjödin, D.G. Patterson, Improved separation of the 209 polychlorinated biphenyl congeners using comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry, J. Chromatogr. A. 1040 (2004) 227–238. doi:10.1016/j.chroma.2004.04.003.

[12] R.B. Gaines, G.S. Frysinger, M.S. Hendrick-Smith, J.D. Stuart, Oil Spill Source Identification by Comprehensive Two-Dimensional Gas Chromatography, Environ. Sci. Technol. 33 (1999) 2106–2112. doi:10.1021/es9810484.

[13] O. Pani, T. Górecki, Comprehensive two-dimensional gas chromatography (GC×GC) in environmental analysis and monitoring, Anal. Bioanal. Chem. 386 (2006) 1013–1023. doi:10.1007/s00216-006-0568-1.

[14] E. Skoczyńska, P. Korytár, J. de Boer, Maximizing Chromatographic Information from Environmental Extracts by GCxGC-ToF-MS, Environ. Sci. Technol. 42 (2008) 6611–6618. doi:10.1021/es703229t.

[15] M. Uhl, Determination of drugs in hair using GC/MS/MS, Forensic Sci. Int. 84 (1997) 281–294. doi:10.1016/S0379-0738(96)02072-5.

[16] C. Hao, X. Zhao, P. Yang, GC-MS and HPLC-MS analysis of bioactive pharmaceuticals and personal-care products in environmental matrices, TrAC Trends Anal. Chem. 26 (2007) 569–580. doi:10.1016/j.trac.2007.02.011.

[17] M.J. Shreve, A. Brockman, M. Hartleb, S. Prebihalo, F.L. Dorman, R.A. Brennan, The white-rot fungus Trametes versicolor reduces the estrogenic activity of a mixture of emerging contaminants in wastewater treatment plant effluent, Int. Biodeterior. Biodegrad. 109 (2016) 132–140. doi:10.1016/j.ibiod.2016.01.018.

[18] J.B. Phillips, J. Beens, Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions, J. Chromatogr. A. 856 (1999) 331–347. doi:10.1016/S0021-9673(99)00815-8.

[19] J.L. Hope, A.E. Sinha, B.J. Prazen, R.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, J. Chromatogr. A. 1086 (2005) 185–192.

[20] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.

[21] A.E. Sinha, J.L. Hope, B.J. Prazen, C.G. Fraga, E.J. Nilsson, R.E. Synovec, Multivariate selectivity as a metric for evaluating comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry subjected to chemometric peak deconvolution, J. Chromatogr. A. 1056 (2004) 145–154. doi:10.1016/j.chroma.2004.06.110.

[22] A. de Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, J. Chemom. 15 (2001) 749–771. doi:10.1002/cem.662.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] H. Parastar, R. Tauler, Multivariate Curve Resolution of Hyphenated and Multidimensional Chromatographic Measurements: A New Insight to Address Current Chromatographic Challenges, Anal. Chem. 86 (2014) 286–297. doi:10.1021/ac402377d.

[28] C. Poole, The Essence of Chromatography, 1st ed., Elsevier, 2003. Pgs. 20, 23-24.

[29] G.M. Gross, B.J. Prazen, J.W. Grate, R.E. Synovec, High-Speed Gas Chromatography Using Synchronized Dual-Valve Injection, Anal Chem. 76 (2004) 3517–3524. doi:10.1021/ac049909g.

[30] B.A. Parsons, D.K. Pinkerton, R.E. Synovec, Implications of Phase Ratio for Maximixing Peak Capacity in Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry, J. Chromatogr. A. (2016).

[31] 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.

[32] L.M. Blumberg, Multidimensional Gas Chromatography: Theoretical Considerations, in: L. Mondello (Ed.), Compr. Chromatogr. Comb. Mass Spectrom., John Wiley & Sons, Inc., Hoboken, NJ, USA, 2011: pp. 13–63. doi:10.1002/9781118003466.ch2.

[33] D.V. Gough, H.D. Bahaghighat, R.E. Synovec, Column Selection Approach to Achieve a High Peak Capacity in Comprehensive Three-Dimensional Gas Chromatography, Talanta. (2018). doi:10.1016/j.talanta.2018.12.007.

[34] L.M. Blumberg, Accumulating resampling (modulation) in comprehensive two-dimensional capillary GC (GC×GC), J. Sep. Sci. 31 (2008) 3358–3365. doi:10.1002/jssc.200800424.

[35] 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.

[36] 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.

[37] R. Ong, P. Marriott, P. Morrison, P. Haglund, Influence of chromatographic conditions on separation in comprehensive gas chromatography, J. Chromatogr. A. 962 (2002) 135–152. doi:10.1016/S0021-9673(02)00458-2.

[38] S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, R.E. Synovec, Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications, Anal. Chem. 90 (2018) 505–532. doi:10.1021/acs.analchem.7b04226.

[39] J.V. Seeley, S.K. Seeley, Multidimensional Gas Chromatography: Fundamental Advances and New Applications, Anal. Chem. 85 (2013) 557–578. doi:10.1021/ac303195u.

[40] L.G. Johnsen, J.M. Amigo, T. Skov, R. Bro, Automated resolution of overlapping peaks in chromatographic data: Chromatographic data analysis, J. Chemom. 28 (2014) 71–82. doi:10.1002/cem.2575.

[41] Y. Izadmanesh, E. Garreta-Lara, J.B. Ghasemi, S. Lacorte, V. Matamoros, R. Tauler, Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data, J. Chromatogr. A. 1488 (2017) 113–125. doi:10.1016/j.chroma.2017.01.052.

[42] C. Hurtado, H. Parastar, V. Matamoros, B. Piña, R. Tauler, J.M. Bayona, Linking the morphological and metabolomic response of Lactuca sativa L exposed to emerging contaminants using GC × GC-MS and chemometric tools, Sci. Rep. 7 (2017). doi:10.1038/s41598-017-06773-0.

[43] J. Omar, M. Olivares, J.M. Amigo, N. Etxebarria, Resolution of co-eluting compounds of Cannabis Sativa in comprehensive two-dimensional gas chromatography/mass spectrometry detection with Multivariate Curve Resolution-Alternating Least Squares, Talanta. 121 (2014) 273–280. doi:10.1016/j.talanta.2013.12.044.

Chapter 3. Statistical Inference of Mass Channel Purity from Fisher Ratio Analysis Using Comprehensive Two-Dimensional Gas Chromatography with Time of Flight Mass Spectrometry Data

This chapter was reproduced from Grant S. Ochoa[†], Sarah E. Prebihalo[†], Brooke C. Reaser, Luke C. Marney, Robert E. Synovec, "Statistical inference of mass channel purity from Fisher ratio analysis using comprehensive two-dimensional gas chromatography with time of flight mass spectrometry data" Accepted to *J. Chromatogr. A* (2020).

[†]These authors contributed equally to this work.

3.1 INTRODUCTION

Comprehensive two-dimensional (2D) gas chromatography (GC×GC), pioneered in 1991 by Liu and Phillips [1], has been extensively developed and is a powerful tool for the analysis of complex mixtures [2–14]. While traditional one-dimensional gas chromatography (1D-GC) can efficiently separate, identify and quantify volatile and semi-volatile analytes, the separation power may be inadequate for many complex samples [15]. Indeed, addition of the second chromatographic dimension with a complementary stationary phase can increase the separation and resolving power of GC×GC over 1D-GC about 10-fold [16–19]. When coupled with time-of-flight mass spectrometry (TOFMS), GC×GC-TOFMS facilitates improved identification and quantification of trace level analytes [3,5,6,14,20].

The inherent complexity of GC×GC-TOFMS data, with its third order data structure [20,21] presents unique data analysis benefits and challenges, which are further impacted by the addition of a fourth dimension due to consideration of multiple sample classes as defined by the experimental design [20,21]. Fortunately, significant effort directed at developing chemometric tools to elucidate chemical information from GC×GC-TOFMS data have improved the quality of

results obtained [18]. Broadly, chemometric methods aim to analyze analytical data qualitatively (signal decomposition (deconvolution), pattern recognition, one and multi class classification) and quantitatively (regression and projection-based methods) [22-25]. There are a wide variety of chemometric methods, including overlapped peak decomposition (mathematical resolution) methods to enable analyte identification and quantification such as Parallel Factor Analysis (PARAFAC) and Multivariate Classical Resolution Alternating Least Squares (MCR-ALS) [22–25], and discovery-based, pattern recognition-based, and property prediction methods such as Principal Component Analysis (PCA) [26–29], Fisher Ratio (F-ratio) Analysis [20,21,26,30–34], Partial Least Squares (PLS), and Partial Least Squares – Discriminant Analysis [35,36].

Chemometric methods are often implemented to extract useful class distinguishing chemical features and can be further distilled into supervised or unsupervised methods, depending on whether sample class membership is known *a priori* [18,19,37,38]. Supervised pixel-based F-ratio analysis was first applied to GC×GC-TOFMS data in 2006 and operates on the conditions that class membership is known, and between class (chemical) variations of discoverable analytes will be larger than within-class (random noise) variations [34]. The F-ratio is mathematically defined as the class-to-class variance divided by the sum of the within-class variance and prioritizes statistical significance over absolute signal [18].

F-ratio analysis has been applied in diverse fields such as metabolomics [8,39], fuel [21,40], food [41], and forensics [42]. However, the nature of differentiating chromatographic regions via F-ratio analysis necessitates having sufficiently aligned chromatograms without run-to-run retention time shifting [30,31]. At the initial stage in the F-ratio software development, due to the limitations of the pixel-based data structure, retention time misalignment resulted in the true positives (class distinguishing features) being overly intermingled with false positives (features

that are falsely recorded as class distinguishing); a cumbersome approach to address this issue was to require alignment prior to F-ratio analysis to overcome the retention time shifting [43–46]. To more effectively address the production of false positives due to retention time misalignment, tile-based F-ratio software was developed [30,31].

Briefly, tile-based F-ratio analysis implements four spatially offset 2D grids, referred to as "tile grids" [30]. Summing the baseline corrected pixel-based signal within each tile per *m/z* optimally captures a given analyte peak feature within a tile from one of the four tile grids for the purpose of calculating F-ratios as a function of mass channel (*m/z*), written as F-ratio(*m/z*). The user-selected 2D tile dimension not only captures the entire analyte peak within a given tile from one of the four tile grids, but also the tile dimension is selected to be large enough to mitigate the sample run-to-sample run retention time shifting issue. Since the four tile grids initially produce redundant analyte "hit" information, the redundant hits are removed using a pinning and clustering algorithm that preserves the single "optimal" tile per analyte hit with the maximum F-ratio(*m/z*), and focuses the multiple 2D tile locations back to the original high-resolution 2D chromatographic data for each analyte hit [31]. In a previous study, for a given hit, the average F-ratio for the top 10 *m/z* was reported per 2D tile location, each hit requiring at least 3 *m/z* above the signal-to-noise ratio (*S/N*) and F-ratio thresholds [20]. However, further exploration of the most appropriate use of the F-ratio(*m/z*) output is warranted. The initial development and several performance and validation studies demonstrate the capabilities of tile-based F-ratio analysis for the discovery of trace class distinguishing characteristics while also reducing false positive rates [8,20,21,30,31]. Regrettably, F-ratio "spectra", i.e., F-ratio(*m/z*), for a given analyte hit have been an underutilized feature. Hence, one of the primary goals of the current study is to carefully examine the chemical

information that can be unmasked and better utilized by capitalizing on the *m/z* selectivity provided by F-ratio analysis.

Specifically, we expand upon the computational capabilities of tile-based F-ratio analysis by introducing additional statistical-based algorithmic features that identify analyte-selective *m/z*. By doing so, analyte hits per *m/z* that exhibit high quality statistical metric outputs will provide accurate analyte quantification without *a priori* requiring analyte identification. Despite advances to data analysis software, current prevailing data analysis workflows frequently require significant analyst intervention post-discovery, such as chemometric decomposition, to perform analyte quantification [37,38]. Thus, a statistical metric-based validation of mass spectral selectivity provided via F-ratio(*m/z*) spectra will provide rapid quantification of analytes with confidence without requiring decomposition *a priori* for a large fraction of analyte hits.

Herein, the capability of tile-based F-ratio analysis to provide accurate quantitative information is evaluated and validated by calculating the signal ratios (S-ratio) as a function of *m/z* between pairs of diesel samples spiked at four concentration levels, i.e., four sample classes (resulting in three different pairwise class comparisons), and comparing the results to the true concentration ratios. The S-ratio algorithm is a new feature incorporated into the tile-based F-ratio software. While the initial F-ratio analysis "discovers" the location of the spiked analytes, providing a hit list, the S-ratio algorithm provides relative quantification for each hit on a per *m/z* basis. Diesel fuel samples spiked with 15 non-native analytes at four concentration levels, then analyzed by GC×GC-TOFMS, were subjected to F-ratio analysis with S-ratio calculations performed on the 15 spiked analytes. This data set was ideal for this proof-of-principle study as it contains regions of peak overlap which often require decomposition prior to quantification. While the methodology described above is aimed to minimize the need for chemometric decomposition,

the S-ratio(*m/z*) results were validated using both initial knowledge of the analyte spike level coupled with PARAFAC [20]. Based upon the F-ratio and S-ratio results obtained per analyte hit on a *m/z* basis, statistical metric tools are implemented to ascertain mass spectral selectivity, and hence, quantitative accuracy of the S-ratios, since for a pure *m/z* the S-ratio will match the true concentration ratio.

3.2 COMPUTATIONAL PRINCIPLES

Following "discovery" based upon the F-ratio(*m/z*) output, the total signal for either a data pixel or F-ratio tile, as a function of *m/z*, can be defined as,

$$S(m/z) = R(m/z)C + \sum_{i=1}^{n} R(m/z)_i C_I \tag{3.1}$$

where $R(m/z)$ is the analyte response factor, C is the analyte concentration, and $R(m/z)_i$ and C_i are the interferent i response factor and concentration, respectively, for n interferences. For two sample classes, A and B, the S-ratio as a function of mass channel, S-ratio(*m/z*), is defined as the total signal of class A, $S_{A,total}$, divided by the total signal of class B, $S_{B,total}$, where "total" implies contributions from the analyte and all interferences,

$$\text{S-ratio}\ (m/z) = \frac{S_{A,total}}{S_{B,total}}(m/z) = \frac{R\left(\frac{m}{z}\right)C_A + \sum_{i=1}^{n} R(m/z)_i C_{I,A}}{R\left(\frac{m}{z}\right)C_B + \sum_{i=1}^{n} R(m/z)_i C_{I,B}} \tag{3.2}$$

with C_A the concentration of the analyte in class A, C_B the concentration of the analyte in class B. Additionally, in Eq. (3.2), $C_{i,A}$ is the concentration of interferent i in class A, and $C_{i,B}$ the concentration of an interferent i in class B, accompanied by the generic response factor $R(m/z)_i$.

The signal of the analyte in class A can be written as,

$$S_A(m/z) = R(m/z)C_A \tag{3.3}$$

and the total signal from all interferences in class A, denoted by subscript i, written as,

$$S_{I,A}\,(m/z) = \sum_{i=1}^{n} R(m/z)_i C_{I,A} \tag{3.4}$$

Likewise, the signal of the analyte in class B can be written as,

$$S_{\mathrm{B}}(m/z) = R(m/z)C_{\mathrm{B}} \tag{3.5}$$

and the signal from all interferences in class B, denoted by subscript i, written as,

$$S_{\mathrm{I,B}}(m/z) = \sum_{\mathrm{i}=1}^{\mathrm{n}} R(m/z)_{\mathrm{i}} C_{\mathrm{I,B}} \tag{3.6}$$

Using Eqs. (3.3)-(3.6), now Eq. (3.2) can be simplified as,

$$\text{S-ratio}(m/z) = \frac{S_{\mathrm{A}}(m/z) + S_{\mathrm{I,A}}(m/z)}{S_{\mathrm{B}}(m/z) + S_{\mathrm{I,B}}(m/z)} \tag{3.7}$$

In the current study, the two sample classes are spiked with analytes so the number and concentration of interferences in class A are equal to the number and concentration of interferences in class B, such that, $S_{\mathrm{I,A}}(m/z) = S_{\mathrm{I,B}}(m/z)$. However, this condition is not a method requirement. Under any circumstances, when $S_{\mathrm{I,A}}(m/z)$ and $S_{\mathrm{I,B}}(m/z)$ both go to 0, such as in the case of pure *m/z* for an analyte, then $R(m/z)$ cancels and the S-ratio(*m/z*) is equal to the true concentration ratio of the analyte, $C_{\mathrm{A}}/C_{\mathrm{B}}$, between the two classes,

$$\text{S-ratio}(m/z) = \frac{R(m/z)C_{\mathrm{A}}}{R(m/z)C_{\mathrm{B}}} = \frac{C_{\mathrm{A}}}{C_{\mathrm{B}}} \tag{3.8}$$

However, in the case of non-selective *m/z*, when $S_{\mathrm{I,A}}(m/z)$ and $S_{\mathrm{I,B}}(m/z)$ are equal and much larger relative to $S_{\mathrm{A}}(m/z)$ and $S_{\mathrm{B}}(m/z)$, then the S-ratio(*m/z*) goes to 1. Additionally, when $S_{\mathrm{I,A}}(m/z)$ and $S_{\mathrm{I,B}}(m/z)$ are either sufficiently large relative to $S_{\mathrm{A}}(m/z)$ and $S_{\mathrm{B}}(m/z)$, or when $S_{\mathrm{I,A}}(m/z)$ and $S_{\mathrm{I,B}}(m/z)$ are not equal, then the S-ratio(*m/z*) will be intermediate between $C_{\mathrm{A}}/C_{\mathrm{B}}$, and a value of 1. Statistical metrics implemented with the S-ratio algorithm seek to explore this transition so analyte selective *m/z* can be identified with confidence to obtain an accurate $C_{\mathrm{A}}/C_{\mathrm{B}}$.

3.3 EXPERIMENTAL

Diesel fuel spiked with 15 non-native compounds (Table 3.1) at four nominal concentration levels (10 ppm, 20 ppm, 40 ppm, and 80 ppm) were analyzed by GC×GC-TOFMS (Pegasus III

TOFMS with a quad-jet thermal modulator , LECO, St. Joseph, MI), that included an Agilent 6890N gas chromatograph with a 7683 autosampler (Agilent Technologies, Palo Alto, CA) [130]. A reverse column configuration was employed with a polar primary column, ^{1}D (Restek Rxi-17Sil MS, 29.75 m × 250 μm i.d. × 0.25 μm film thickness) and a nonpolar secondary column, ^{2}D (Restek Rxi-1 MS, 1.5 m × 180 μm i.d. × 0.18 μm film thickness). Ultrahigh purity helium (Grade 5, 99.999%, Praxair, Seattle, WA, USA) was used as the carrier gas at a constant flow of 2.0 mL/min. The ^{1}D column was maintained at 40 °C for 1 min and then increased to 300 °C using a temperature program rate of 5 °C/min, where it was held for 2 min. The ^{2}D column and modulator block followed the same temperature program with a +15 °C and +60 °C offset, respectively. The modulation period was 2 s. The ion source was set to 225 °C, and mass channels, *m/z* 35-300, were collected at a 100 spectra/s after a 10 s acquisition delay. Six injection replicates of each of the four concentration spike levels were collected to create four sample classes containing 6 samples each.

Table 3.1 S-ratio results for each of the 15 spiked analytes in the 80/40 ppm comparison, arranged by decreasing maximum F-ratio. The S-ratio for the top *m/z* (which also passed the p-value and *LOF* thresholds) is provided and compared against the actual concentration ratio. Average concentration ratio deviation was 1.0% with a standard deviation of 1.4%.

Analyte Name	**True Concentration Ratio**	**F-ratio**	***m/z***	**S-ratio**	**Deviation, %**
bromobenzene	1.97	7404	38	1.98	0.5
1,6-dichlorohexane	1.99	6112	67	1.97	-1.0
1-chlorohexane	2.04	5351	91	2.04	0
5-decyne	2.12	3597	77	2.09	-1.4
2,5-dimethylthiophene	2.07	2173	111	2.07	0
1-nonanol	2.05	1979	43	2.05	0
limonene	2.07	1533	53	2.06	-0.5
ethyl salicylate	2.15	1404	92	2.13	-0.9
aniline	2.27	1337	92	2.27	0
2-decanone	2.28	1318	41	2.17	-4.8
methyl caproate	2.23	989	87	2.21	-0.9
butyrophenone	2.12	972	106	2.14	0.9
2-heptanol	2.25	682	45	2.33	3.6
3-octanone	1.92	529	42	1.92	0
cyclohexyl isothiocyanate	1.96	298	82	1.94	-1.0

Data analyses were performed in Matlab R2018a (The Mathworks Inc., Natick, MA, USA) and were imported from the LECO ChomaTOF software v 3.32 (LECO, St. Joseph, MI) via an in-house developed data converter [20]. The data were analyzed with in-house developed tile-based F-ratio software [8,20,21,30,31]. There were three class comparisons, the 20 ppm versus 10 ppm spike level, the 40 ppm versus the 20 ppm spike level, and the 80 ppm versus 40 ppm spike level, so for each comparison a total of 12 samples were analyzed with 6 in each class. Most of the results presented are for the 80 ppm versus 40 ppm spike level comparison, while in some instances all three comparisons are related to each other. The F-ratio tile dimensions were 12 s on ^{1}D by 100 ms on ^{2}D reducing the size of the data tensor from 200×1644×267 to 20×274×267 data points representing the ^{2}D, ^{1}D and m/z respectively, and the cluster tile dimensions were 8 s on ^{1}D by 60

ms on ^{2}D; a *S/N* threshold of 10 was applied and at least 3 *m/z* were required to define a reported F-ratio [20].

Directly following removal of redundant hits by the pinning and clustering algorithm step [31], S-ratio(*m/z*) were obtained, for each F-ratio(*m/z*) as illustrated in Fig. 3.1, with the S-ratio algorithm flow chart provided in Fig. 3.2. The S-ratio(*m/z*) algorithm is based in part upon a previously developed algorithm for aligned chromatograms [43,47], though the new version does not require alignment. For each hit identified by F-ratio analysis, data for a given analyte was extracted from each chromatogram, one chromatogram at a time. First, the ^{1}D and ^{2}D pin location from the F-ratio hit list was applied to an individual chromatogram. Then, a S-ratio tile with dimensions equal to the ^{1}D dimension of the F-ratio tile (12 s) and double the dimension of the ^{2}D cluster tile dimension (120 ms) was centered on the pin. Next, the algorithm determines where the average (original) F-ratio pin location is on the peak for the particular chromatogram at hand, i.e. whether it is on the right, left or top of the analyte peak maxima. Using an iterative process, the ^{2}D peak maxima for all modulations located within the S-ratio tile (as illustrated in Fig. 3.1) are recorded. Then, beginning from the modulation which corresponds to the largest ^{2}D peak maximum, the ^{1}D boundaries are determined by checking subsequent modulations to the right and left for an increase in signal, which indicates the modulation belongs to another analyte and/or interferent. Finally, the signals for the ^{2}D peak maxima from all modulations identified as belonging to the analyte are summed, as to give one signal per *m/z* from the F-ratio spectrum of the hit. This algorithm is repeated across all chromatograms in the F-ratio class comparison.

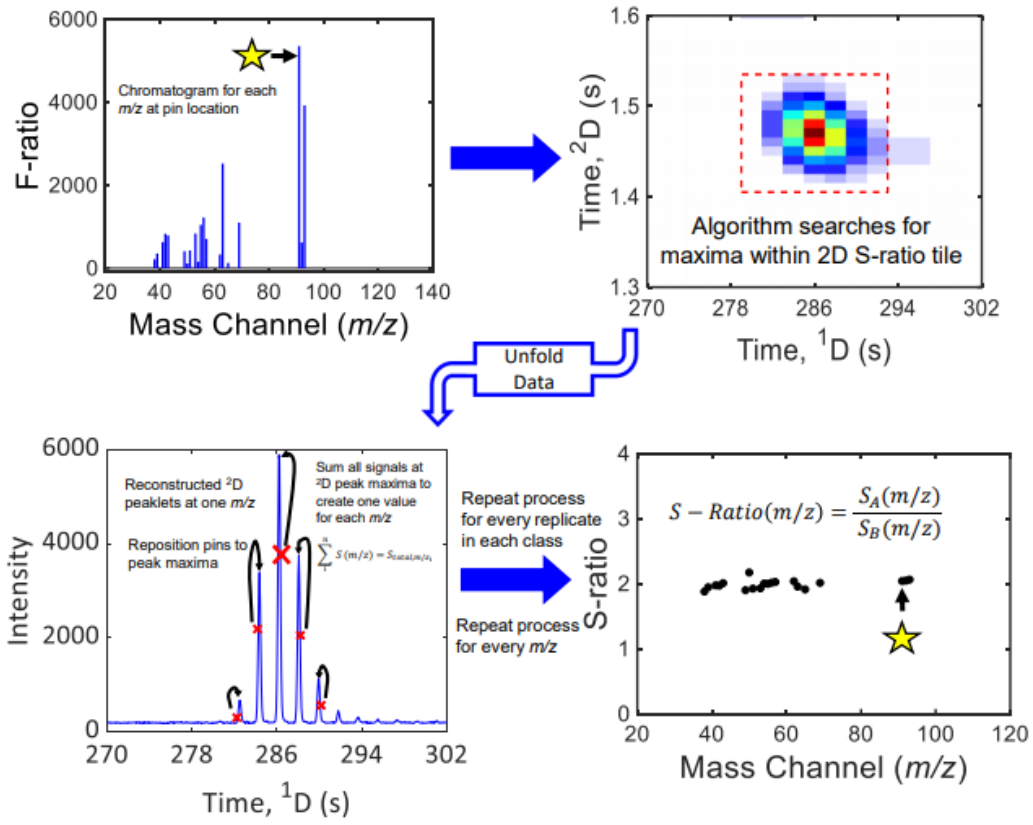

Figure 3.1. Flowchart depicting the data processing steps of the S-ratio(*m/z*) algorithm for 1-chlorohexane. For each F-ratio(*m/z*), illustrated for *m/z* 91 with a star (upper left), a 2D S-ratio tile is cutout around each hit (upper right) from each chromatogram centered at the average $^{1}D \times {}^{2}D$ pin location. The data within the 2D S-ratio tile is unfolded and concatenated into a vector laying each modulation section end to end, with the average $^{1}D \times {}^{2}D$ pin location represented by the large red X in the bottom left plot. Initial ^{2}D pins are then assigned to each modulation based off the average $^{1}D \times {}^{2}D$ pin, separated by the modulation period, which serve as initial estimates to the peak maxima for each ^{2}D peak, represented by the small red x's. These initial pins are then repositioned to the ^{2}D peak maxima, indicated by the black arrows. Finally, the ^{2}D peak heights are summed, yielding a single signal value for each selected *m/z* (lower left). The summed signals are then averaged for both sample classes and divided to yield the S-ratio(*m/z*) spectrum, with the S-ratio for *m/z* 91 indicated with a star (lower right).

The process described above was repeated for all *m/z* identified as class distinguishing at each pin location. Following these signal intensity calculations, the algorithm calculated S-ratio(*m/z*) at every identified F-ratio(*m/z*) for each hit between class A (the class containing the higher spike level) and class B (lower spike level). Quantitative accuracy was evaluated through comparison of the true concentration ratio and the experimental S-ratio(*m/z*). Based upon the

experimental design, all S-ratios, and hence experimentally determined concentration ratios, for the spiked analytes should nominally be 2.

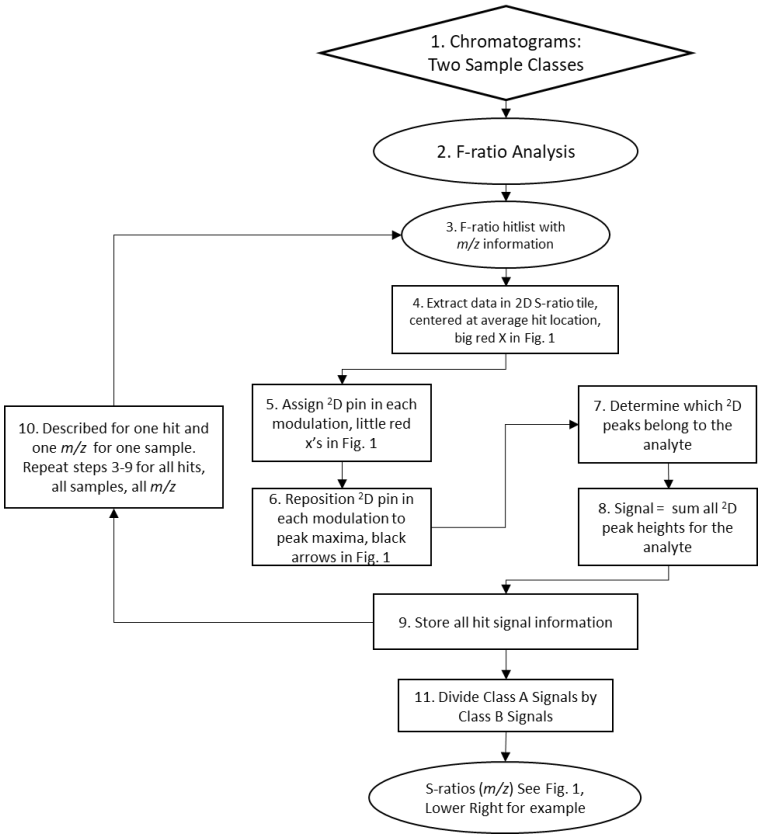

Figure 3.2. Flow chart depicting the workflow for the S-ratio algorithm. First, F-ratio analysis is performed on two classes of chromatograms yielding a hitlist ordered by decreasing F-ratio with hit locations along with a list of *m/z* that passed the *S/N* threshold for each hit (steps 1-3). Utilizing this information, the chromatographic data in the 2D S-ratio tile is extracted for a given hit around the average $^{1}D \times {}^{2}D$ pin represented by the large red X seen in Fig. 3.1 for the signal at one *m/z* (step 4). Initial ^{2}D pins are then assigned to each modulation based off the average $^{1}D \times {}^{2}D$ pin for the hit which serve as initial estimates to the peak maxima for the ^{2}D peak in each modulation, represented by the small red x's in Fig. 3.1 (step 5). These initial pins are then repositioned to the maxima of their respective ^{2}D peaks, black arrows (step 6). In step 7 we determine which ^{2}D peaks belong to the analyte by scanning for locations where consecutive peak maxima switch from decreasing to increasing in signal, indicating the presence of an interfering compound. Peak maxima that start to increase in signal are discarded and the signals of the ^{2}D peaks determined to belong to the analyte are summed, yielding a single signal value for the hit at the selected *m/z* (step 8). The signal information is stored in an array (step 9) and the process from steps 3-9 is repeated (step 10) for each *m/z* for each selected hit from the F-ratio hitlist for each sample. At the end of the process we have now obtained the signal values for each software selected *m/z* for each hit for each sample. The signal values for class A samples are then divided by class B sample signals yielding six S-ratio values given a six versus six two class experimental design (step 11). The S-ratios can then be averaged yielding a mean S-ratio along with the standard deviation information.

Combinatorial null distribution analysis (CNDA) has been previously demonstrated to provide a statistically robust F-ratio threshold in which false positives are minimized, while maximizing the true positive discovery rate, defined as the "limit of discovery" [8,20,30,31]. An extension of CNDA to study F-ratio(*m/z*) distributions for the three class comparisons (20 ppm versus 10 ppm spike, 40 ppm versus the 20 ppm spike, and 80 ppm versus 40 ppm spike) is provided to determine an F-ratio threshold that correlates with quantitative accuracy via the S-ratio(*m/z*) calculations. Null class comparisons were analyzed by creating two "null classes," which consist of intermingling samples from class A and class B and submitting the null classes to the tile-based F-ratio analysis algorithm. For a six versus six sample class comparison, there are 200 possible null permutations of sample class membership that are obtained. Unlike previously demonstrated applications of CNDA, where all 200 hit lists were constructed to create an average null probability distribution for threshold determination, this study used a null probability distribution of all F-ratio(*m/z*) as a method for determining a threshold above which quantitative accuracy is anticipated to be maximized. A null probability level of 99.999% was applied to each probability plot, yielding three distributions with 200 F-ratio threshold values. Then a 95% confidence interval was applied to each distribution to select a single F-ratio threshold, from there an average of the three class comparisons was calculated and applied.

Independent from the inference of mass spectral purity provided by the F-ratio threshold via CNDA, the consistency of analyte peak shapes between classes was analyzed and quantified using a lack-of-fit (*LOF*) statistic using an in-house developed algorithm, with the LOF algorithm flow chart provided in Fig. 3.3. To begin, as with the determinations of the S-ratio(*m/z*) signal (per Fig. 3.1), a *LOF* tile whose ^{1}D dimension matches the F-ratio tile ^{1}D dimension (12 s) and ^{2}D dimension is double the F-ratio tile ^{2}D dimension (200 ms) was pulled from a chromatogram from

class A (80 ppm) and class B (40 ppm) for each *m/z* for a given hit. For a sufficiently pure *m/z*, an increase in concentration will increase the area under the ^{2}D peaks while the peak shapes will remain constant. To test this concept, the pattern of ^{2}D peaks within a *LOF* tile for class B were normalized to the area of the corresponding modulations in class A. Next, the pattern of ^{2}D peaks in class B was aligned to class A by interpolation such that the original ^{2}D peak shapes remain intact, minimizing the squared residuals between classes. The data were then unfolded into a concatenated vector and the *LOF* was calculated according to Eq. (3.9),

$$LOF\ \left(\frac{m}{z}\right) = 100\% \sqrt{\frac{\sum_i (y-x)_i^2}{\sum_i x_i^2}} \tag{3.9}$$

where *x* is the concatenated pattern of ^{2}D peaks for class B, and *y* is the concatenated pattern of ^{2}D peaks for class A [48]. Purer *m/z* result in smaller residuals indicated by a lower *LOF*, while a larger *LOF* is indicative of a change in the pattern of ^{2}D peaks between classes, indicative of interfering compounds. However, a low *LOF* can also be obtained when there is only an interferent present in the *LOF* bin at a particular *m/z*. This results from a fully non-selective *m/z* that produces an extremely low F-ratio. So, in combination with the *LOF* metric, the S-ratio algorithm also calculates the p-value for each hit per *m/z*, and a combination of a sufficiently low *LOF* and low p-value are required to infer identification of pure *m/z*.

While the methodology described above is aimed to minimize the need for chemometric decomposition, the S-ratio(*m/z*) results were validated using both initial knowledge of the analyte spike level coupled with PARAFAC [20]. Small 2D chromatographic regions surrounding the hit locations were input into PARAFAC, and the number of components were iteratively adjusted until explained variance was maximized. Profiles belonging to the spiked analytes were confirmed through NIST library search. The pure analyte profiles in each class were then integrated to obtain the true concentration ratio.

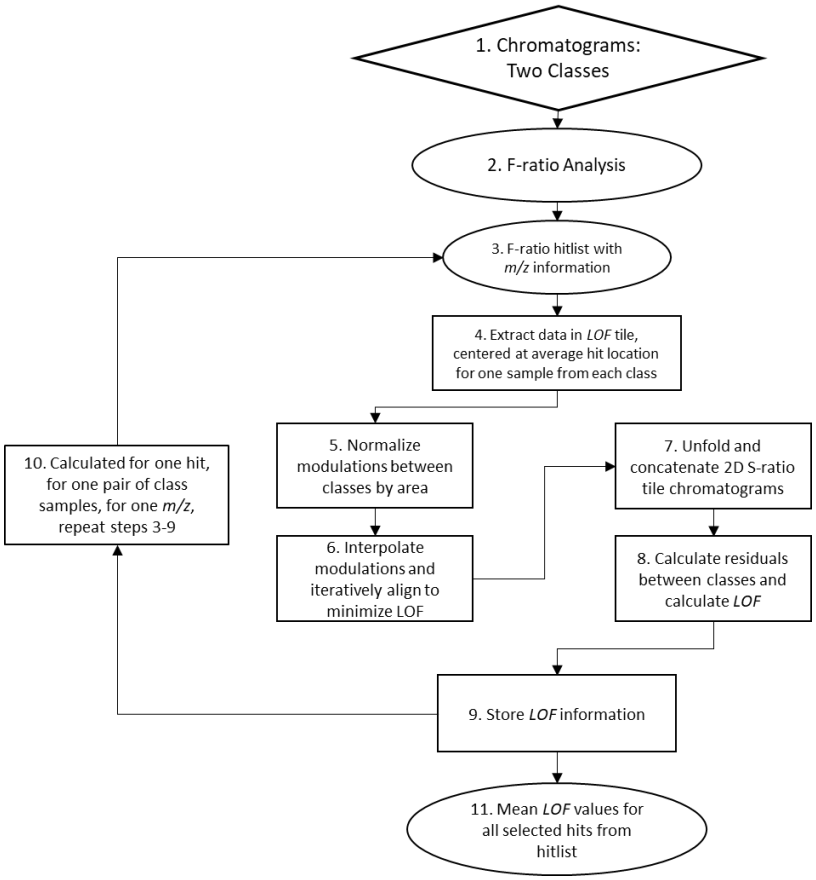

Figure 3.3. Flow chart depicting the workflow for the lack-of-fit (*LOF*) algorithm. Steps 1-3 are identical to the S-ratio algorithm, using complementary outputs. Utilizing the output information, we extract the chromatographic data in the *LOF* tile for a given hit around the average $^{1}D \times {}^{2}D$ pin represented by the large red X seen in Fig. 3.1 for one *m/z* for one sample from each class (step 4). The modulations of the extracted data are then normalized such that each modulation between has the same area between classes so that the peak shape can be evaluated with concentration effects removed (step 5). Next the data in each modulation is interpolated such that more points are added in between the existing points without altering the peak shape. These modulations are then iteratively aligned between classes calculating a *LOF* value for each iteration in an attempt to minimize this value so that the resulting *LOF* value is due to peak shape differences and not retention time shifting (step 6). The aligned modulations are then unfolded and concatenated, laid end to end, and the residuals are calculated between classes which are in turn used to calculate a single *LOF* value for the chromatogram pair (steps 7-8). This value is then stored in an array (step 9) and the process from steps 3-9 is repeated for each sample pair for each software selected *m/z* for each hit (step 10). At the end of this process mean *LOF*s and standard deviations can be calculated from the six sample pairs from each hit's mass channels (step 11).

3.4 RESULTS AND DISCUSSION

The GC×GC separation of the diesel fuel spiked with 15 analytes at the 80 ppm level is presented in Fig. 3.4A. The locations of the spiked analytes span the 2D separation space and are indicated by circles. Tile-based F-ratio analysis was employed for the class comparison using six injection replicates of 80 ppm spiked diesel (class A) and six injection replicates of 40 ppm spiked diesel (class B) to generate a hit list on a per *m/z* basis for each hit. Comparisons were also made between the 40 and 20 ppm spiked diesel samples and between the 20 and 10 ppm spiked diesel samples, yielding an F-ratio(*m/z*) distribution for all three two-class comparisons, provided in Fig. 3.4B. Note that most of the F-ratio(*m/z*) are relatively low, and the concentration dependence, i.e., *S/N* dependence, is evident with the 80/40 ppm F-ratio(*m/z*) distribution shifted preferentially to higher F-ratios, followed by the 40/20 ppm comparison, and finally the 20/10 ppm produced the lowest F-ratios. Combinatorial null distribution analysis (CNDA) was then applied to the three null class comparisons, to provide insight into determining a statistically robust F-ratio threshold in which one would anticipate that false positives are minimized for the true class comparisons in Fig. 3.4B. The CNDA workflow and results are presented in Fig. 3.5 for all three comparisons, and an average F-ratio threshold of 97 was obtained. Based upon the probability levels applied, an F-ratio(*m/z*) exceeding this threshold has a 1 in 100,000 chance of being a false positive at a 95% confidence level, to purposely provide a very restrictive “statistically implied quantitative accuracy” filter via CNDA. Note that this implied quantitative accuracy filter is further investigated by more direct *m/z* purity metrics, specifically the *LOF* and p-values for the S-ratio(*m/z*) outputs.

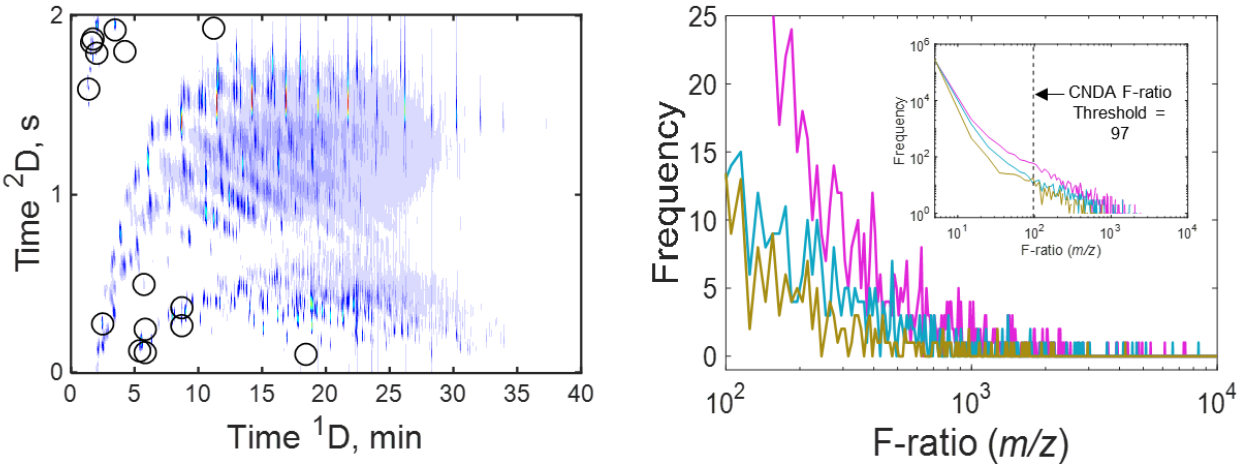

Figure 3.4. (A) GC×GC chromatogram (TIC) of the 80 ppm spiked diesel. Alkanes (top), cycloalkanes and branched alkanes (middle), and aromatic (bottom) compound classes are distinctively separated. The 15 spiked analyte locations are denoted with circles. (B) F-ratio(*m/z*) distribution of spiked diesel at three different concentration level comparisons: 20/10 ppm in gold, 40/20 ppm in cyan, and 80/40 ppm in magenta. As the concentration levels of each comparison increase, the overall frequency of F-ratio(*m/z*) > 100 increases, due to the increased signal per *m/z* overcoming the user-selected *S/N* threshold. The majority of the F-ratio(*m/z*) > 100 will be shown to be attributed to the spiked analytes.

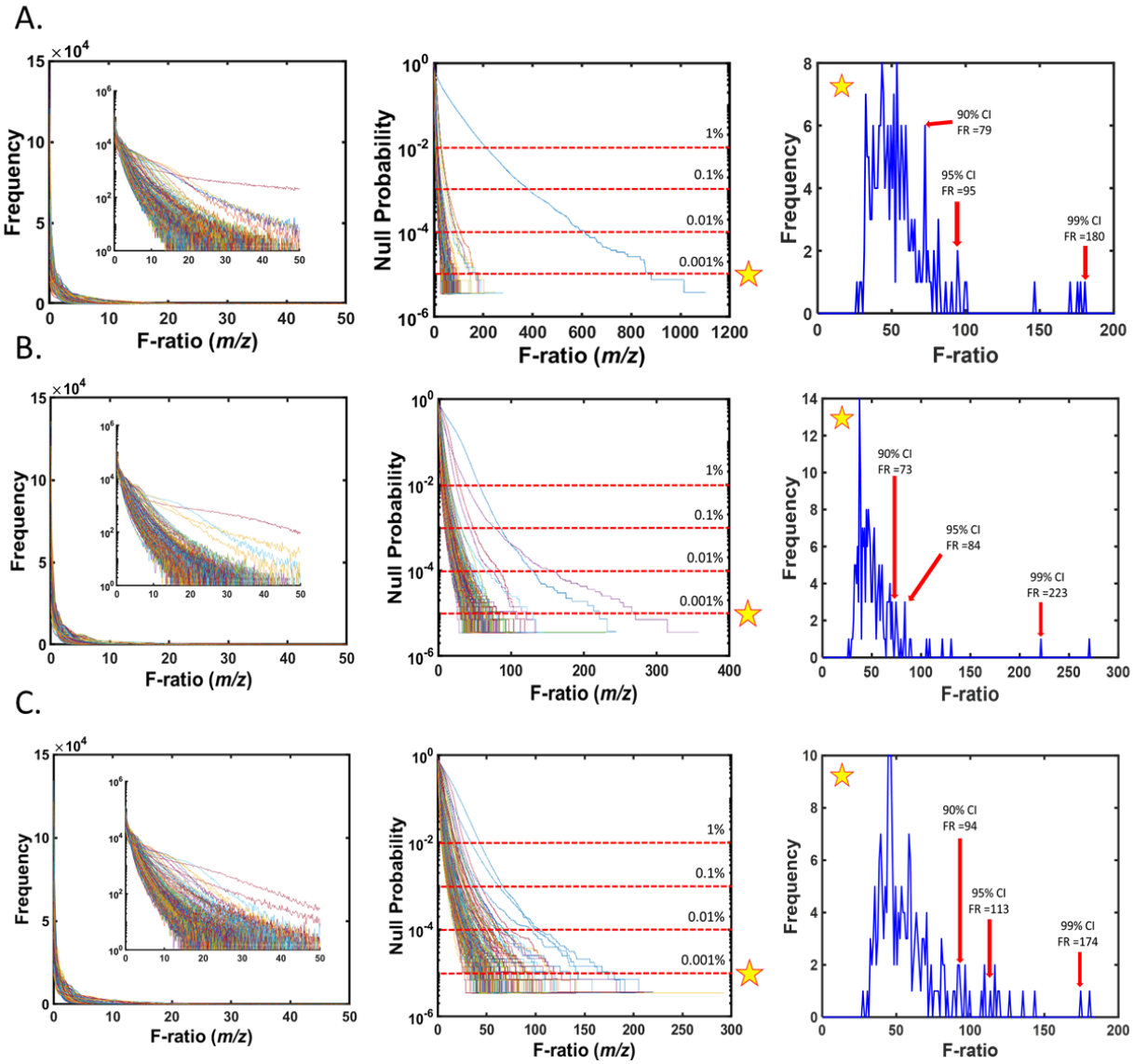

Figure 3.5. Null distributions using CNDA for each of the concentration comparisons: (A) 80/40 ppm, (B) 40/20 ppm, and (C) 20/10 ppm, and the inset displays a zoom in of distributions (Left). Null probability plots for each of the 200 combinatorial null class comparisons (Middle). Probability levels marked off by dashed red lines indicate the respective percentage of features lying below that line. For instance, for the probability curve passing the starred 0.001% line, the F-ratio value where it crossed represents the threshold that only 0.001% of features in that respective distribution lie above. By finding this value for each probability curve, we can create a distribution of 0.001% null probability F-ratio values (Right). The 80/40 comparison 0.001% probability distribution contains one F-ratio that is off scale at 881, excluded for distribution clarity in (A). The 90, 95, and 99% confidence levels for the 0.001% probability level are marked. Using the 95% level as an example, we can be 95% confident that an F-ratio in the true class comparison that exceeds 63 has a 1 in 100,000 (0.001%) chance of being a true positive.

F-ratio(*m/z*) and S-ratio(*m/z*) spectra were calculated for each spiked analyte in the 80/40 ppm comparison to begin to visualize how these two outputs are related in the context of *m/z* purity, beginning with a representative pure analyte (1-chlorohexane) in Fig. 3.6A, and a representative overlapped analyte (methyl caproate) in Fig. 3.6B. The signal for 1-chlorohexane in the 80 ppm sample is presented for both the top F-ratio at *m/z* = 91 in Fig. 3.6A(i), and the total ion current (TIC) in Fig. 3.6A(ii) with a red box outlining the 2D bin used by the S-ratio algorithm. The 1-chlorohexane peak is not overlapped with any interferences. Likewise, the top F-ratio at *m/z* = 87 and TIC are presented for methyl caproate in Fig. 3.6B(i-ii). While the signal for the top F-ratio(*m/z*) for methyl caproate is framed by the red box, suggesting *m/z* = 91 might be pure, most of the TIC signal falls outside of the red box, suggesting interferences are overlapping with methyl caproate. Indication that 1-chlorohexane is a pure peak, while methyl caproate is not a pure peak is provided by the visual comparison of the mass spectrum of the signal in the 2D bin versus the library spectrum for each, in Figs 3.6A(v) and 3.6B(v), respectively. Most, if not all, of the *m/z* match for 1-chlorohexane between the 2D bin spectrum and the library spectrum, while for methyl caproate there are several *m/z* in the 2D spectrum that are not in the library spectrum. Note that providing the library spectrum is not needed for application of the method presented herein, rather, the library spectra are merely provided in Fig. 3.6 to aid in understanding the interrelationships of the F-ratio(*m/z*), S-ratio(*m/z*), and other quantitative metrics.

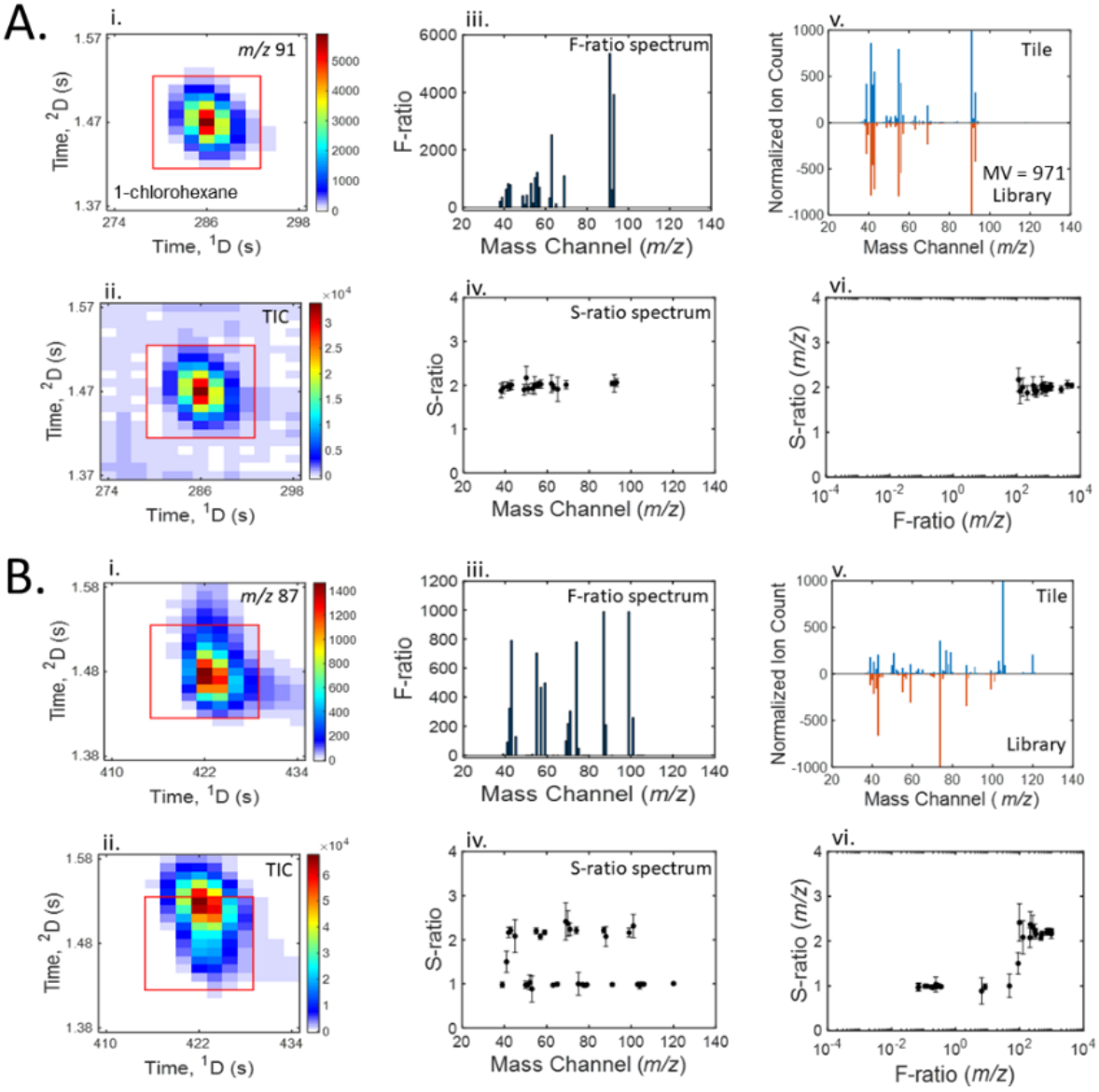

Figure 3.6. Illustration of the F-ratio(*m/z*) and S-ratio(*m/z*) spectra from the 80/40 ppm comparison for two representative analytes: (A) one pure, 1-chlorohexane, and (B) one overlapped, methyl caproate. Two-dimensional data for the region around each analyte for the top F-ratio *m/z* (i), inferred to be pure, and the total ion current (TIC) (ii). The red box outlines the tile space used by the S-ratio algorithm. Each compound F-ratio(m/z) and S-ratio(m/z) spectrum is provided in (iii) and (iv), respectively. The mass spectrum from the red box region is also provided (v) for both analytes and compared to the library spectrum. Finally, a plot of S-ratio(*m/z*) as a function of F-ratio(*m/z*) in (vi) provides insight into the correlation between high F-ratio(*m/z*) and having the S-ratio(*m/z*) correspond to the true concentration ratio of 2.

The F-ratio(*m/z*) spectrum for each hit provides a "chemical fingerprint" to determine which *m/z* are most likely to distinguish the two sample classes. For example, Figs. 3.6A(iii) and 3.6B(iii) provide the F-ratio(*m/z*) spectra for 1-chlorohexane and methyl caproate, respectively. A lower F-ratio(*m/z*) implies lower spectral selectivity (and/or lower *S/N*), with an adverse impact on quantitative accuracy anticipated, while a higher F-ratio(*m/z*) implies higher spectral selectivity. However, since the F-ratio(*m/z*) is principally a tool to prioritize which *m/z* per hit may likely be statistically different in concentration between the two classes, we turn to the S-ratio(*m/z*) to bridge this gap between analyte discovery and quantification. For 1-chlorohexane and methyl caproate, the S-ratio(*m/z*) spectra are provided in Figs. 3.6A(iv) and 3.6B(iv), respectively. For a pure analyte (1-chlorohexane), the S-ratio(*m/z*) spectrum exhibits a consistent S-ratio(*m/z*) around 2 in Fig. 3.6A(iv), while the presence of interferences (with methyl caproate) results in two distinct S-ratio(*m/z*) levels in Fig. 3.6B(iv), belonging to both the analyte and an interferent. An S-ratio(*m/z*) of 2 is obtained for nominally pure *m/z* for methyl caproate, while an S-ratio(*m/z*) of about 1 suggests the signal obtained is dominated by an interferent that is invariant between the two sample classes. Finally, in Figs. 3.6A(vi) and 3.6B(vi) the S-ratio(*m/z*) is plotted as a function of F-ratio(*m/z*), for 1-chlorohexane and methyl caproate, respectively. Here, the relationship between obtaining an accurate S-ratio(*m/z*) and its connection to requiring a sufficiently high F-ratio(*m/z*) provides two key observations. First, the S-ratio(*m/z*) trends toward the true concentration ratio for the spiked analyte for a high F-ratio(*m/z*), indicating quantitative accuracy. Second, as the F-ratio(*m/z*) magnitude decreases the S-ratio(*m/z*) deviate from the true concentration ratio, leveling off at 1. Indeed, the transition from obtaining the true concentration to a biased concentration ratio appears to occur at an F-ratio $\approx$ 100, which is fully supported by the F-ratio threshold of 97 provided by the CNDA (Fig. 3.5). Similar trends were also observed for the remaining the 13

spiked analytes in all three F-ratio comparisons, with results such as in Fig. 3.6 provided for 1-nonanol in Fig. 3.7.

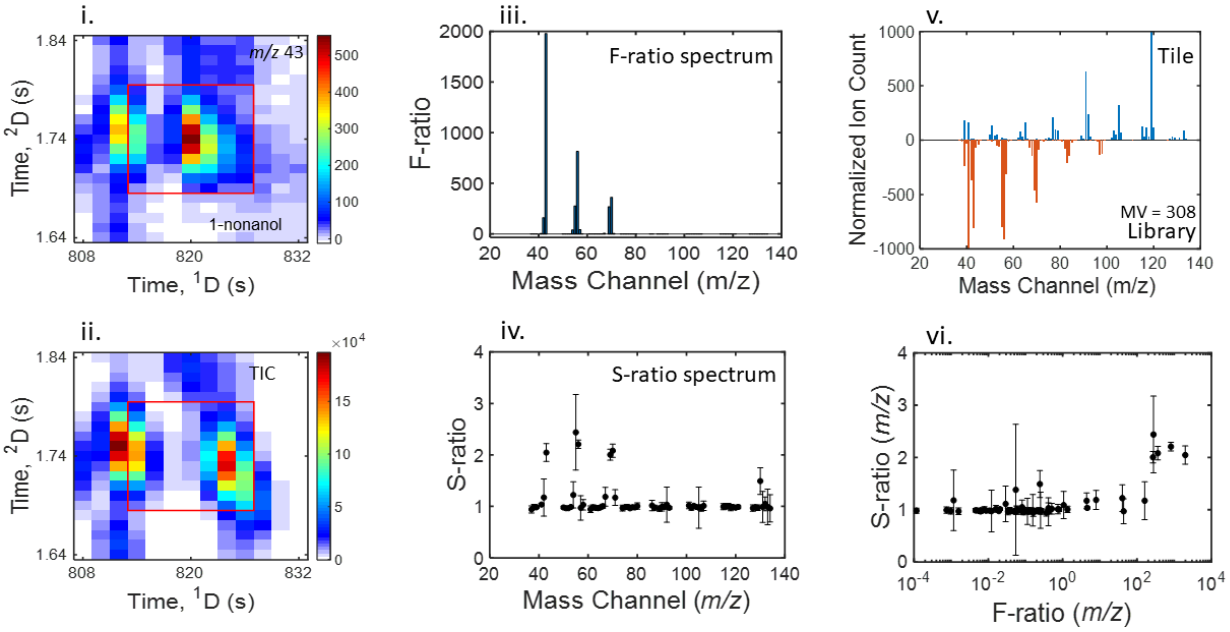

Figure 3.7. Illustration of the F-ratio(*m/z*) and S-ratio(*m/z*) spectra from the 80/40 ppm comparison for a severely overlapped analyte, 1-nonanol. Two-dimensional data for the region around each analyte for the top F-ratio *m/z* (i), inferred to be relatively pure, and the total ion current (TIC) (ii) from an average 80 ppm chromatogram. The red box outlines the tile space used by the S-ratio algorithm. The F-ratio(m/z) and S-ratio(m/z) spectrum is provided in (iii) and (iv), respectively. The mass spectrum from the red box region is also provided (v) and compared to the library spectrum. Finally, a plot of S-ratio(*m/z*) as a function of F-ratio(*m/z*) in (vi) provides insight into the correlation between high F-ratio(*m/z*) and having the S-ratio(*m/z*) correspond to the true concentration ratio of 2. Since very few F-ratio(*m/z*) correspond to results with an S-ratio(*m/z*) ≈ 2, this demonstrates a very challenging case.

A summary of the S-ratio(*m/z*) as a function of F-ratio(*m/z*) for all three comparisons is provided in Fig. 3.8, where an S-ratio(*m/z*) ≈ 2 is obtained for an F-ratio(*m/z)* ≥ 97. As the *S/N* of the analyte signal within each comparison increases, the number of *m/z* that exceed the F-ratio threshold of 97 also increases, since more *m/z* have passed the *S/N* threshold. Specifically, for the 15 spiked analytes, 248 of the F-ratio(*m/z*) are ≥ 97 for the 80/40 ppm comparison, while there were 136 for the 40/20 ppm comparison, and then dropping down to only 76 for the 20/10 ppm comparison. While the results in Fig. 3.8 suggest that an F-ratio(*m/z)* ≥ 97 may very well serve as

a suitable metric to infer quantitative accuracy for determining the analyte concentration ration via the S-ratio(*m/z*), a firmer connection to the *LOF* and p-value statistical metrics was warranted.

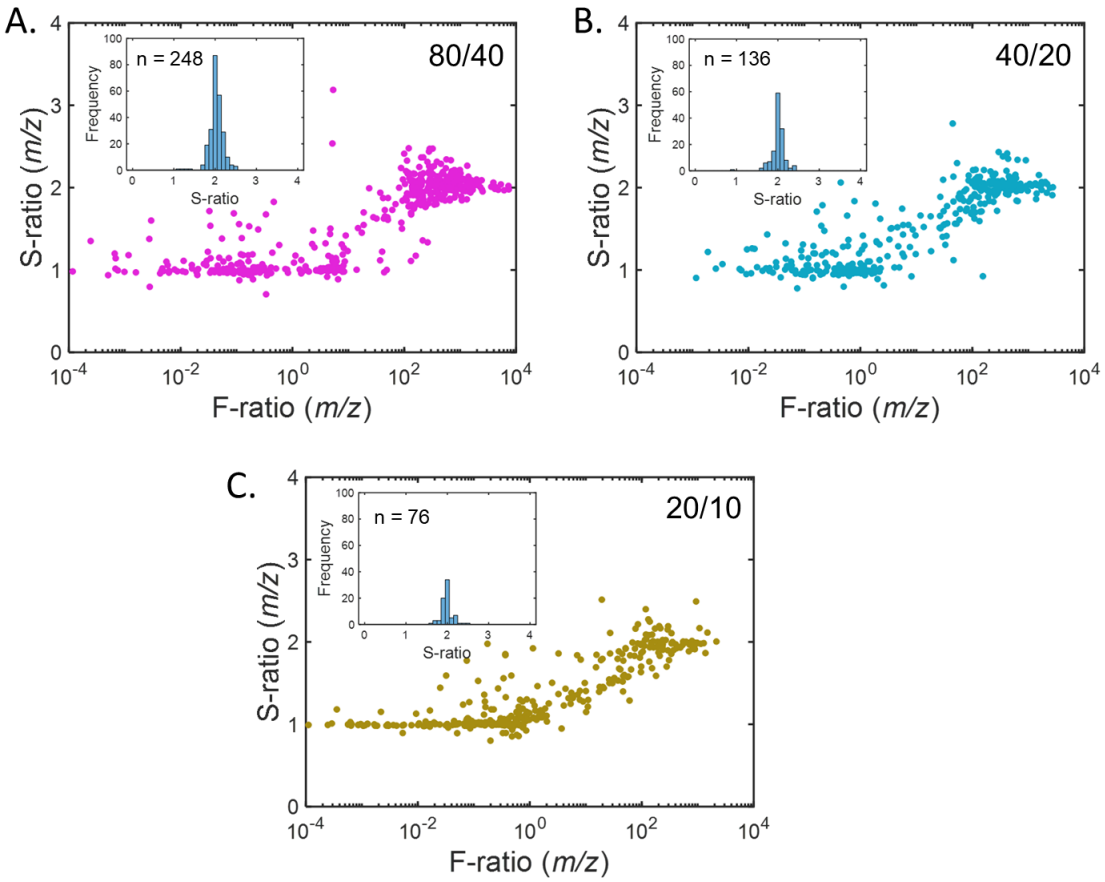

Figure 3.8. The S-ratio(*m/z*) for all *m/z* selected by F-ratio analysis for the 15 spiked analytes plotted versus their respective F-ratio(*m/z*) for each comparison: (A) 80/40 ppm, (B) 40/20 ppm, and (C) 20/10 ppm. Each plot shows a similar pattern of the S-ratio(*m/z*) tending toward an accurate concentration ratio of 2 at high F-ratio(*m/z*). The inset in each plot shows the distribution of S-ratio(m/z) above the F-ratio threshold of 97 determined by combinatorial null distribution analysis. The occurrence of accurate S-ratio(m/z) ≈ 2) increases with higher concentration comparisons.

For each hit, the *LOF* statistic was calculated to determine the extent of *m/z* purity. Figure 3.9A-D provides an example of applying the *LOF* algorithm, via Eq. (3.9), for four *m/z* "discovered" by F-ratio analysis for methyl caproate. Here the chromatographic data provided by the S-ratio(*m/z*) algorithm are utilized before and after normalization. In Fig. 3.9A for *m/z* = 43, doubling the analyte concentration between sample classes results in a consistent peak shape.

Following the normalization step for the *LOF* algorithm, low residuals are obtained, which experience some variance due to sampling frequency, ultimately resulting in a low *LOF* of 5.5% indicative of a sufficiently pure *m/z*. In contrast, as the interferences become a more dominant contributor to the signal, the consistency in peak shape pattern suffers when the analyte concentration is doubled between sample classes. This is recognized with both *m/z* = 101 (Fig. 3.9B) and *m/z* = 41 (Fig. 3.9C), where after area normalization the presence of interferences causes the peak pattern to misalign and subsequently increases the size of the residuals, resulting in large *LOF* values of 22.4% and 29.1%, respectively. If the interference signal dominates to the extent that the analyte signal is negligible, the *LOF* returns to a smaller value, indicating that the *m/z* is essentially pure, or at least a constant contribution from class to class, for the interference. This situation is illustrated in Fig. 3.9D for *m/z* = 105, where a *LOF* of 1.6% is obtained. However, there is a key distinction between the two low *LOF* conditions, Fig. 3.9A versus Fig 3.9D. For *m/z* = 43, the p-value = $1.6*10^{-12}$, indicating the concentration ratio is significantly different than 1, while for *m/z* = 105, the p-value = 0.47, indicating the concentration ratio is not significantly different than 1. Thus, it is the combination of using the *LOF* and p-value metrics, both provided directly from the S-ratio(*m/z*) output, that provide an inferred, but statistically confident, assessment of pure *m/z*. For the purpose of this study, we elected to require the *LOF* $\leq$ 20% concurrent with a p-value $\leq$ 0.001 in order to deem a *m/z* to be sufficiently pure to provide accurate quantitative analysis, i.e., an accurate S-ratio(*m/z*) corresponding with the true concentration ratio.

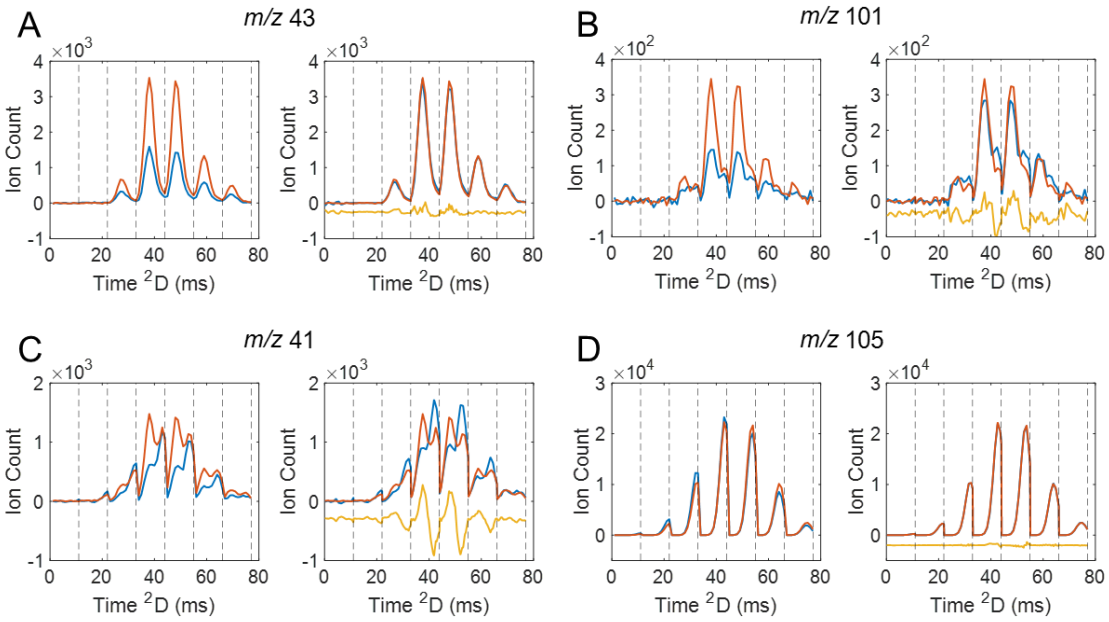

Figure 3.9. Demonstration of the *LOF* metric determination for four selected *m/z* from methyl caproate. An overlay of the original concatenated ^{2}D data for a chromatogram from class A (80 ppm) and one from class B (40 ppm) is provided for each *m/z* in the left panel. The right panel for each *m/z* shows the same concatenated ^{2}D data after normalization and alignment, with the residuals plotted underneath, offset for clarity. (A) A pure *m/z* = 43, yields excellent consistency between classes with low residuals, *LOF* = 5.5%. (B) For *m/z* = 101, a small relative contribution from an overlapped compound, increases the residuals, *LOF* = 22.4%. (C) For *m/z* = 41, more significant contribution from the overlapped compound is indicated, and the residuals become very large relative to the analyte signal, *LOF* = 29.1%. (D) At *m/z* = 105, essentially a pure *m/z* for the overlapped compound, with negligible analyte contribution, with no change between classes the residuals are very low, yielding a low *LOF* of 1.6% (but the p-value would be high).

These *LOF* and p-value metrics are applied to all F-ratio(*m/z*) in Fig. 3.10 for the 80/40 ppm comparison. The first statistical filter applied is the p-value. All F-ratio(*m/z*) with a p-value > 0.001 are indicated as blue dots. Next the *LOF* filter was applied, and all F-ratio(*m/z*) with a *LOF* > 20% are indicated as red dots. This leaves a region filled with green dots indicating pure *m/z*, with representations of these results put into the context of the S-ratio(*m/z*) in Fig. 3.11 for all three comparisons.

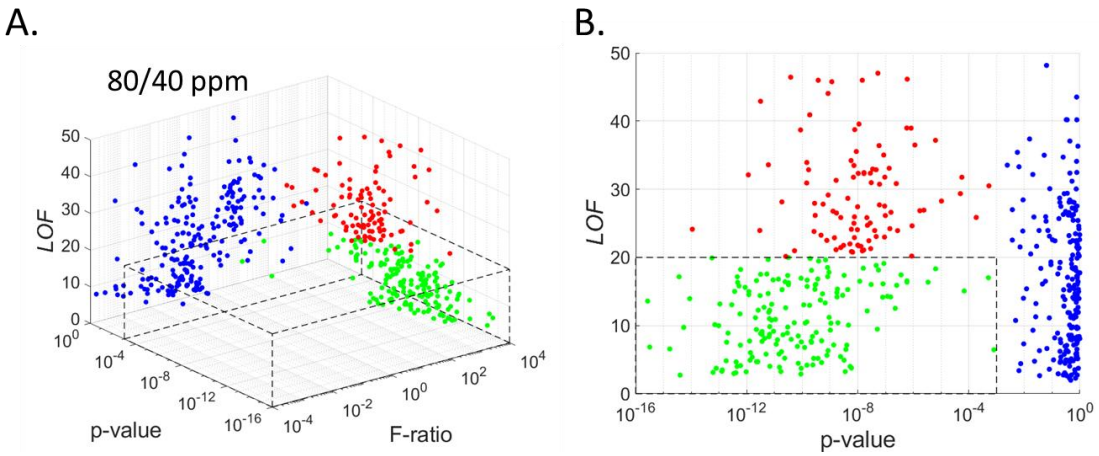

Figure 3.10. F-ratio(m/z), p-value(m/z), and *LOF(m/z)* for all *m/z* of all 15 spiked analytes. The dashed lines represent the regions defined by the p-value ≤ 0.001 and *LOF* ≤ 20 % thresholds (A) Blue *m/z* have not passed either threshold, red have only passed the p-value threshold, and green *m/z* have passed both thresholds. (B) A 2D projection of the *LOF* and p-value dimensions illustrates these two metrics independent of the F-ratio dimension. The blue *m/z* carry insignificant *m/z* belonging to interfering compounds. Red *m/z* have statistically significant signal differences between classes however they are interfered to a large extent with overlapping compounds. Lastly the green *m/z* are considered both statistically significant and sufficiently pure to be used for signal quantification, while not relying upon the F-ratio(*m/z*).

It is interesting to relate the S-ratio(*m/z*) results in Fig. 3.8 prior to applying these metrics, in contrast to in Fig. 3.11 after their application. There are 170 pure *m/z* for the 80/40 comparison as indicated in Fig. 3.11A. Note that use of the *LOF* and p-value metrics to determine pure *m/z* does not rely upon the F-ratio ≥ 97 threshold applied in Fig. 3.8, where there were 248 *m/z* for the 80/40 comparison that were deemed good *candidates* to be determined to be pure *m/z*. Hence, the combination of the *LOF* and p-value metrics filter the candidate *m/z*, forming a more robust cohort of pure *m/z*. The *LOF* and p-value metrics applied to all F-ratio(*m/z*) for the other two comparisons are provided in Fig. 12. Likewise, for the other two comparisons, substantially fewer *m/z* met the *LOF* and p-value metrics relative to the F-ratio ≥ 97 threshold. For the 40/20 ppm comparison there were 91 *m/z* (Fig. 3.11B) versus 136 *m/z* (Fig. 3.8), and for the 20/10 comparison even fewer with 70 *m/z* (Fig. 3.11C) versus 76 *m/z* (Fig. 3.8).

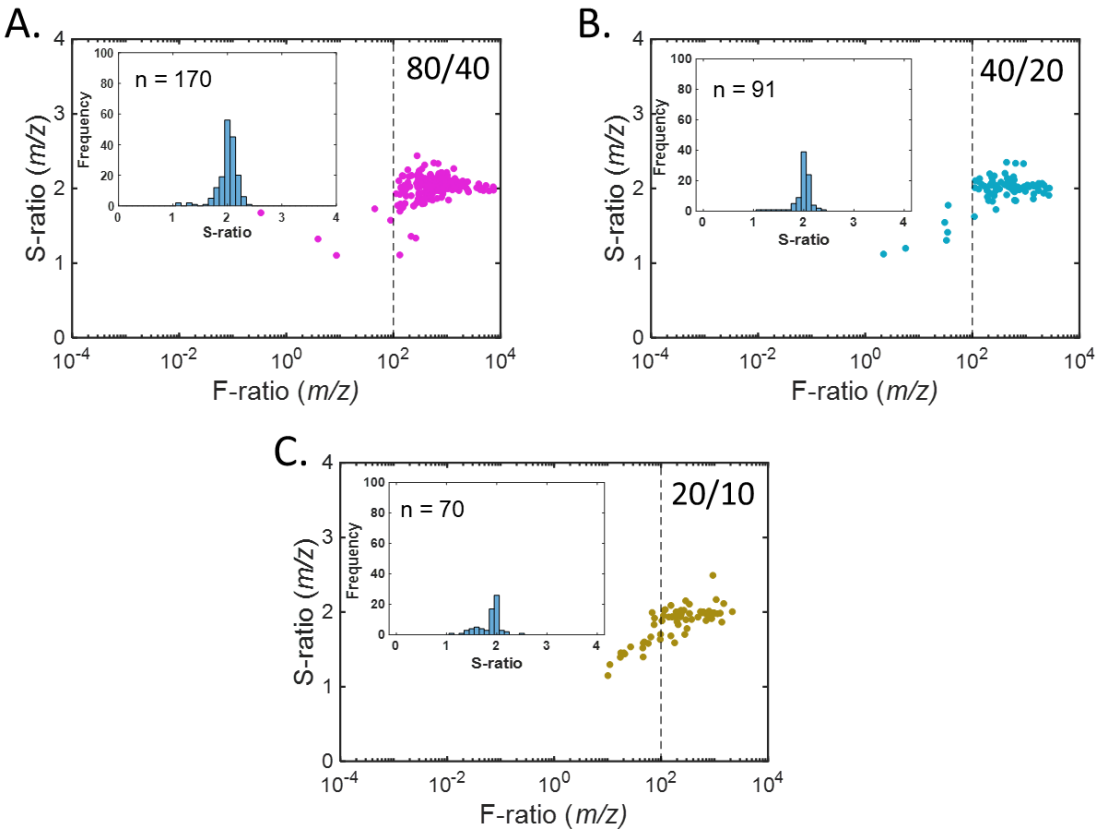

Figure 3.11. S-ratio versus F-ratio plots for all three concentration comparisons after applying p-value and *LOF* metric thresholds in Fig. 3.6 with the top F-ratio(*m/z*). The combinatorial null distribution analysis F-ratio threshold = 97 is marked for context. Inset in each plot are histograms of the *m/z* that pass these two metrics along with a count.

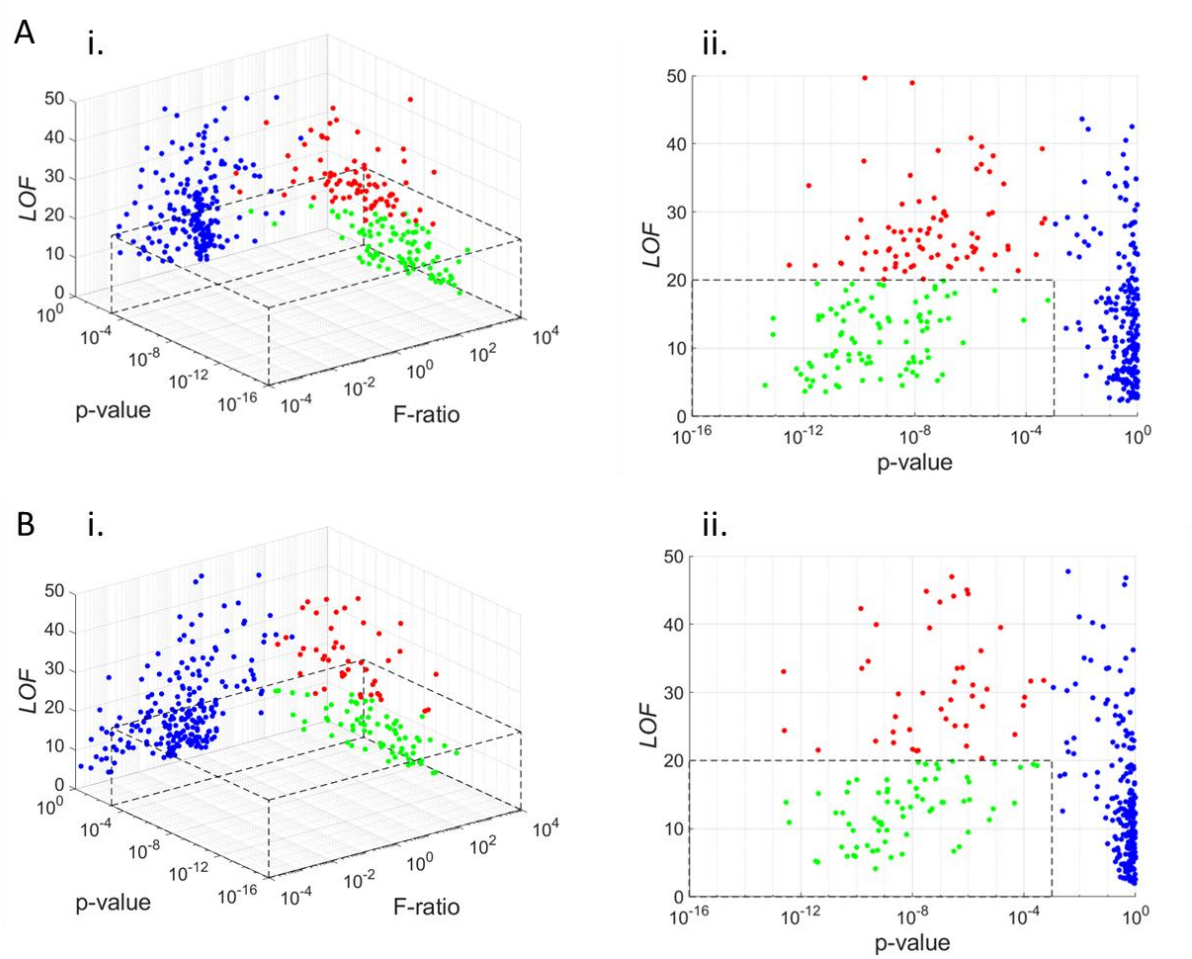

Figure 3.12. F-ratio(m/z), p-value(m/z), and *LOF(m/z)* for all *m/z* of all 15 spiked analytes for the for the 40/20 and 20/10 ppm comparisons, (A) and (B), respectively The dashed lines represent the regions defined by the p-value ≤ 0.001 and *LOF* ≤ 20 % thresholds (A) Blue *m/z* have not passed either threshold, red have only passed the p-value threshold, and green *m/z* have passed both thresholds. (B) A 2D projection of the *LOF* and p-value dimensions illustrates these two metrics independent of the F-ratio dimension. The blue *m/z* carry insignificant *m/z* belonging to interfering compounds. Red *m/z* have statistically significant signal differences between classes however they are interfered to a large extent with overlapping compounds. Lastly the green *m/z* are considered both statistically significant and sufficiently pure to be used for signal quantification, while not relying upon the F-ratio(*m/z*). We see that as the average concentration of the comparisons decreases there are fewer *m/z* that pass the F-ratio *S/N* filter and/or have higher *LOF* values.

The S-ratio(*m/z*) as a function of F-ratio(*m/z*) in Fig. 3.11 also reveals that the application of the *LOF* and p-value metrics did not rid the *m/z* selected of all apparent outliers. There were a small number of *m/z* that while meeting the *LOF* and p-value metrics, they exhibited an F-ratio <

97. For example, with the 80/40 ppm comparison, there were 6 *m/z* with an F-ratio < 97 that met the *LOF* and p-value metrics. One approach to improve this situation is to lower the *LOF* threshold, but this is at the expense of discarding an *m/z* that may be sufficiently pure for the purpose of a given quantitative application. For application of LOF thresholds of 10% and 15%, see Fig. 3.13.

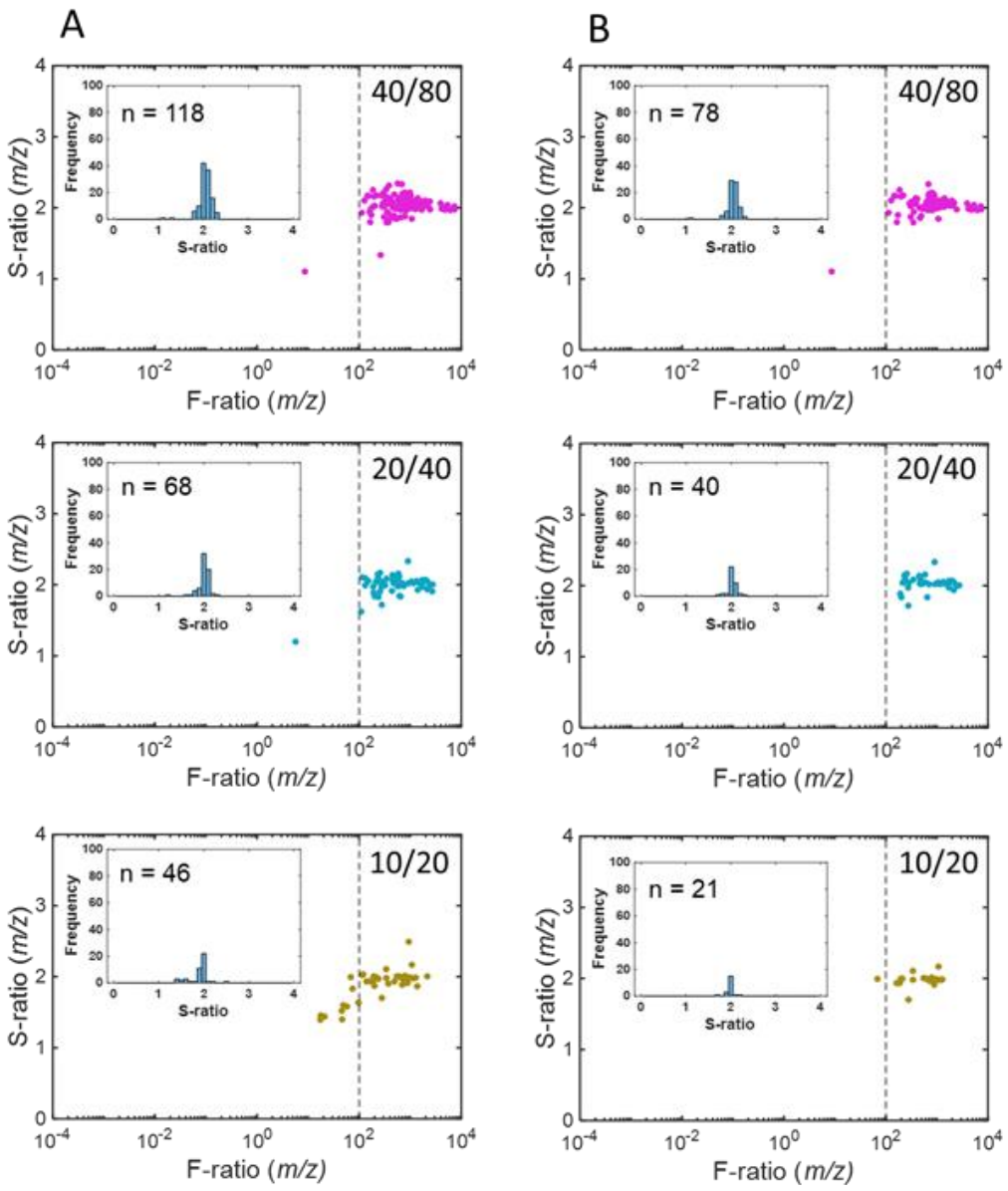

Figure 3.13. Plots of S-ratio vs F-ratio for each concentration comparison of the *m/z* that passed a p-value threshold of < 0.001, but with with a *LOF* threshold of <15% (A) and <10% (B). The CNDA F-ratio threshold = 97 is marked for context. Inset in each plot are histograms of the *m/z* that pass these two metrics along with a count. Compared to the *LOF* threshold of < 20 in Figure 7 there are much fewer *m/z* in each comparison, reflecting the tightening of the parameter.

None the less, it is tempting to consider the notion of inferring *m/z* purity strictly by applying the F-ratio(*m/z*) ≥ 97 threshold provided by the CNDA. By doing so, the analyst would be relying upon the correlation between how well the LOF and p-value metrics performed their task as manifested in the S-ratio(*m/z*) dependency on F-ratio(*m/z*) in Fig. 3.11. This approach involves taking the S-ratio(*m/z*) at the maximum F-ratio(*m/z*) as the best estimate of the analyte concentration ratio. For the 80/40 ppm comparison in Table 3.1, all of the concentration ratio estimates are reasonably accurate with an average deviation of 1.0%, albeit benefited by the maximum F-ratio(*m/z*) being much greater than the 97 threshold for all 15 analytes. Indeed, all 15 of the maximum F-ratio(*m/z*) also passed the *LOF* and p-value threshold metrics. Tables for the other two comparisons are provided below, Tables 3.2 and 3.3, with diminished accuracy due to the maximum F-ratio(*m/z*) decreasing as the analyte concentration levels decrease. While this approach does not require the analyte to be identified *a priori* in the process, the analyst must realize that other *m/z* purity metrics are needed to ensure an accurate quantification, as demonstrated in this study.

Table 3.2 S-ratio results for each of the 15 spiked analytes in the 40/20 ppm comparison, arranged by decreasing F-ratio. The S-ratio for the top *m/z* which passed the p-value and *LOF* thresholds is provided and compared against the actual concentration ratio. Cyclohexyl isothiocyanate did not have any *m/z* passing these thresholds. Average concentration ratio deviation was 1.9% with a standard deviation of 2.1%.

Analyte Name	True Concentration Ratio	F-ratio	*m/z*	S-ratio	Deviation, %
bromobenzene	2.02	2744	77	2.00	-0.7
1,6-dichlorohexane	2.03	2684	67	1.91	-6.2
1-chlorohexane	2.05	2169	91	2.04	-0.3
2,5-dimethylthiophene	2.04	2006	111	2.07	1.6
2-heptanol	2.31	902	45	2.33	0.9
5-decyne	2.13	805	68	2.00	-6.2
ethyl salicylate	2.15	604	120	2.12	-1.7
methyl caproate	2.18	587	43	2.16	-1.0
butyrophenone	2.10	542	148	2.13	1.1
2-decanone	2.06	455	59	2.07	0.3
3-octanone	2.02	438	43	1.91	-5.1
aniline	2.16	236	93	2.14	-0.8
1-nonanol	2.24	211	56	2.20	-1.8
limonene	2.03	169	92	2.03	-0.3
cyclohexyl isothiocyanate	1.77	-	-	-	-

Table 3.3 S-ratio results for each of the 15 spiked analytes in the 20/10 ppm comparison, arranged by decreasing F-ratio. The S-ratio for the top *m/z* which passed the p-value and *LOF* thresholds is provided and compared against the actual concentration ratio. Limonene had no *m/z* pass both of these filters. 1-Nonanol had 1 passing *m/z* which appears to belong to an interferent. Average concentration ratio deviation was 6.1% with a standard deviation of 11%. If 1-nonanol is excluded, the resulting average deviation was 2.8% with a standard deviation of 2.4%.

Analyte Name	True Concentration Ratio	F-ratio	*m/z*	S-ratio	Deviation, %
1-chlorohexane	2.03	2155	43	2.01	-1.0
2-decanone	2.03	1488	58	2.11	4.1
1,6-dichlorohexane	1.97	1372	67	1.87	-5.2
butyrophenone	2.05	1278	105	1.99	-3.0
2,5-dimethylthiophene	1.97	1134	111	1.98	0.4
aniline	2.16	1075	93	2.17	0.3
bromobenzene	1.99	976	77	1.99	0.2
2-heptanol	2.60	933	45	2.49	-4.3
3-octanone	2.17	582	72	1.99	-8.5
ethyl salicylate	1.95	340	120	1.98	1.7
methyl caproate	2.21	289	43	2.15	-2.5
5-decyne	2.08	236	41	2.02	-2.9
cyclohexyl isothiocyanate	1.51	46	141	1.59	4.7
1-nonanol	2.15	10	42	1.15	-46.5
limonene	1.88	-	-	-	-

3.5 CONCLUSION

In this study, we evaluated the level of mass channel selectivity afforded by tile-based F-ratio analysis through statistical inference. Following F-ratio analysis of spiked diesel at two concentration levels (80/40 ppm), a novel S-ratio algorithm and novel *LOF* algorithm were implemented to study the statistical relationship between F-ratio, p-value, *LOF*, and *m/z* purity, as they relate to quantitative accuracy. Application of a suitable *LOF* threshold coupled with a p-value threshold yielded a subset of the most pure *m/z* for each of the 15 spiked analytes, evident by the low deviations (< 5%) in S-ratio relative to the true concentration ratio**,** for the top *m/z* for each analyte. Thus, the ability to isolate pure *m/z* without the need for higher level signal

decomposition algorithms was demonstrated. The *LOF* metric was applied for the case where the analyte was present in both classes at sufficient concentrations. In the case that an analyte is not present in one class, a separate algorithm step would need to be included to identify this situation, eg., any *m/z* present in one class but not the other (below the *S/N* threshold) would identify that *m/z* as belonging to an analyte of interest. Future application to discover spectrally pure *m/z* will be utilized not only for analyte identification but as an advanced feature selection for inputs to methods like PCA and PLS.

3.6 ACKNOWLEDGEMENTS

S-ratio algorithm development, conceptualization, and data analysis was performed by Sarah E. Prebihalo. Lack-of-fit algorithm development, conceptualization, and data analysis was performed by Grant S. Ochoa.

3.7 REFERENCES

[1] Z. Liu, J.B. Phillips, Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface, J. Chromatogr. Sci. 29 (1991) 227–231. https://doi.org/10.1093/chromsci/29.6.227.

[2] R.B. Gaines, G.S. Frysinger, M.S. Hendrick-Smith, J.D. Stuart, Oil Spill Source Identification by Comprehensive Two-Dimensional Gas Chromatography, Environ. Sci. Technol. 33 (1999) 2106–2112. https://doi.org/10.1021/es9810484.

[3] J.F. Focant, A. Sjödin, D.G. Patterson, Improved separation of the 209 polychlorinated biphenyl congeners using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, J. Chromatogr. A. 1040 (2004) 227–238. https://doi.org/10.1016/j.chroma.2004.04.003.

[4] P.J. Marriott, S.T. Chin, B. Maikhunthod, H.G. Schmarr, S. Bieri, Multidimensional gas chromatography, TrAC - Trends Anal. Chem. 34 (2012) 1–21. https://doi.org/10.1016/j.trac.2011.10.013.

[5] D. Megson, R. Kalin, P.J. Worsfold, C. Gauchotte-Lindsay, D.G. Patterson, M.C. Lohan, S. Comber, T.A. Brown, G. O'Sullivan, Fingerprinting polychlorinated biphenyls in environmental samples using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A. 1318 (2013) 276–283. https://doi.org/10.1016/j.chroma.2013.10.016.

[6] K.L. Organtini, A.L. Myers, K.J. Jobst, J. Cochran, B. Ross, B. McCarry, E.J. Reiner, F.L. Dorman, Comprehensive characterization of the halogenated dibenzo-p-dioxin and dibenzofuran contents of residential fire debris using comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry, J. Chromatogr. A. 1369 (2014) 138–146. https://doi.org/10.1016/j.chroma.2014.09.088.

[7] M.R. Jacobs, M. Edwards, T. Górecki, P.N. Nesterenko, R.A. Shellie, Evaluation of a miniaturised single-stage thermal modulator for comprehensive two-dimensional gas chromatography of petroleum contaminated soils, J. Chromatogr. A. 1463 (2016) 162–168. https://doi.org/10.1016/j.chroma.2016.08.009.

[8] 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. https://doi.org/10.1016/j.chroma.2016.06.067.

[9] F. Magagna, A. Guglielmetti, E. Liberto, S.E. Reichenbach, E. Allegrucci, G. Gobino, C. Bicchi, C. Cordero, Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination, J. Agric. Food Chem. 65 (2017) 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167.

[10] E.A. Higgins Keppler, C.L. Jenkins, T.J. Davis, H.D. Bean, Advances in the application of comprehensive two-dimensional gas chromatography in metabolomics, TrAC - Trends Anal. Chem. 109 (2018) 275–286. https://doi.org/10.1016/j.trac.2018.10.015.

[11] B. Gruber, B.A. Weggler, R. Jaramillo, K.A. Murrell, P.K. Piotrowski, F.L. Dorman, Comprehensive two-dimensional gas chromatography in forensic science: A critical review of recent trends, TrAC - Trends Anal. Chem. 105 (2018) 292–301. https://doi.org/10.1016/j.trac.2018.05.017.

[12] A.M. Muscalu, T. Górecki, Comprehensive two-dimensional gas chromatography in environmental analysis, TrAC - Trends Anal. Chem. 106 (2018) 225–245. https://doi.org/10.1016/j.trac.2018.07.001.

[13] M. Zoccali, S. Cappello, L. Mondello, Multilevel characterization of marine microbial biodegradation potentiality by means of flow-modulated comprehensive two-dimensional gas chromatography combined with a triple quadrupole mass spectrometer, J. Chromatogr. A. 1547 (2018) 99–106. https://doi.org/10.1016/j.chroma.2018.03.013.

[14] R. Pesesse, P.H. Stefanuto, F. Schleich, R. Louis, J.F. Focant, Multimodal chemometric approach for the analysis of human exhaled breath in lung cancer patients by TD-GC × GC-TOFMS, J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci. 1114–1115 (2019) 146–153. https://doi.org/10.1016/j.jchromb.2019.01.029.

[15] J.B. Phillips, J. Beens, Comprehensive two-dimensional gas chromatography: A hyphenated method with strong coupling between the two dimensions, J. Chromatogr. A. 856 (1999) 331–347. https://doi.org/10.1016/S0021-9673(99)00815-8.

[16] 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. https://doi.org/10.1016/j.chroma.2015.01.031.

[17] L.M. Blumberg, Multidimensional Gas Chromatography: Theoretical Considerations, in: Compr. Chromatogr. Comb. Mass Spectrom., John Wiley and Sons, 2011: pp. 13–63. https://doi.org/10.1002/9781118003466.ch2.

[18] S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, R.E. Synovec, Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications, Anal. Chem. 90 (2018) 505–532. https://doi.org/10.1021/acs.analchem.7b04226.

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

[20] B.C. Reaser, B.W. Wright, R.E. Synovec, Using Receiver Operating Characteristic Curves to Optimize Discovery-Based Software with Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry, Anal. Chem. 89 (2017) 3606–3612. https://doi.org/10.1021/acs.analchem.6b04991.

[21] 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. https://doi.org/10.1016/j.chroma.2016.02.067.

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

[23] A. De Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, J. Chemom. 15 (2001) 749–772. https://doi.org/10.1002/cem.662.

[24] R. Gargallo, R. Tauler, F. Cuesta-Sánchez, D.L. Massart, Validation of alternating least-squares multivariate curve resolution for chromatographic resolution and quantitation, TrAC - Trends Anal. Chem. 15 (1996) 279–286. https://doi.org/10.1016/0165-9936(96)00048-9.

[25] R.E. Mohler, B.P. Tu, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography-time-of-flight-mass spectrometry data, J. Chromatogr. A. 1186 (2008) 401–411. https://doi.org/10.1016/j.chroma.2007.10.063.

[26] J.E. Welke, V. Manfroi, M. Zanus, M. Lazzarotto, C.A. Zini, Differentiation of wines according to grape variety using multivariate analysis of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection data, Food Chem. 141 (2013) 3897–3905. https://doi.org/10.1016/j.foodchem.2013.06.100.

[27] P.H. Stefanuto, K.A. Perrault, L.M. Dubois, B. L'Homme, C. Allen, C. Loughnane, N. Ochiai, J.F. Focant, Advanced method optimization for volatile aroma profiling of beer using two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A. 1507 (2017) 45–52. https://doi.org/10.1016/j.chroma.2017.05.064.

[28] 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. https://doi.org/10.1016/j.chroma.2014.10.057.

[29] L.M. Dubois, K.A. Perrault, P.H. Stefanuto, S. Koschinski, M. Edwards, L. McGregor, J.F. Focant, Thermal desorption comprehensive two-dimensional gas chromatography coupled to variable-energy electron ionization time-of-flight mass spectrometry for monitoring subtle changes in volatile organic compound profiles of human blood, J. Chromatogr. A. 1501 (2017) 117–127. https://doi.org/10.1016/j.chroma.2017.04.026.

[30] 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. https://doi.org/10.1016/j.talanta.2013.06.038.

[31] 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. https://doi.org/10.1021/ac504472s.

[32] 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. https://doi.org/10.1039/c5ay00371g.

[33] L.W. Hantao, B.R. Toledo, F.A. De Lima Ribeiro, M. Pizetta, C.G. Pierozzi, E.L. Furtado, F. Augusto, Comprehensive two-dimensional gas chromatography combined to multivariate data analysis for detection of disease-resistant clones of Eucalyptus, Talanta. 116 (2013) 1079–1084. https://doi.org/10.1016/j.talanta.2013.08.033.

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

[35] T. Gröger, R. Zimmermann, Application of parallel computing to speed up chemometrics for GC × GC-TOFMS based metabolic fingerprinting, Talanta. 83 (2011) 1289–1294. https://doi.org/10.1016/j.talanta.2010.09.015.

[36] E. Aprea, H. Gika, S. Carlin, G. Theodoridis, U. Vrhovsek, F. Mattivi, Metabolite profiling on apple volatile content based on solid phase microextraction and gas-chromatography time of flight mass spectrometry, J. Chromatogr. A. 1218 (2011) 4517–4524. https://doi.org/10.1016/j.chroma.2011.05.019.

[37] K.M. Pierce, R.E. Mohler, A review of chemometrics applied to comprehensive two-dimensional separations from 2008-2010, Sep. Purif. Rev. 41 (2012) 143–168. https://doi.org/10.1080/15422119.2011.591868.

[38] Z. Zeng, J. Li, H.M. Hugel, G. Xu, P.J. Marriott, Interpretation of comprehensive two-dimensional gas chromatography data using advanced chemometrics, TrAC - Trends Anal. Chem. 53 (2014) 150–166. https://doi.org/10.1016/j.trac.2013.08.009.

[39] S. Stadler, P.H. Stefanuto, M. Brokl, S.L. Forbes, J.F. Focant, Characterization of volatile organic compounds from human analogue decomposition using thermal desorption coupled to comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, Anal. Chem. 85 (2013) 998–1005. https://doi.org/10.1021/ac302614y.

[40] K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: Comprehensive GC × GC with ANOVA-based feature selection and principal component analysis, in: Chemom. Intell. Lab. Syst., 2002: pp. 225–237. https://doi.org/10.1016/S0169-7439(01)00198-8.

[41] F. Magagna, E. Liberto, S.E. Reichenbach, Q. Tao, A. Carretta, L. Cobelli, M. Giardina, C. Bicchi, C. Cordero, Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from thermal to differential-flow modulated comprehensive two dimensional gas chromatography, J. Chromatogr. A. 1536 (2018) 122–136. https://doi.org/10.1016/j.chroma.2017.07.014.

[42] M. Schäffer, T. Gröger, M. Pütz, S. Dieckmann, R. Zimmermann, Comparative Analysis of the Chemical Profiles of 3,4-Methylenedioxymethamphetamine Based on Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry (GC×GC-

TOFMS), J. Forensic Sci. 57 (2012) 1181–1189. https://doi.org/10.1111/j.1556-4029.2012.02137.x.

[43] R.E. Mohler, B.P. Tu, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography-time-of-flight-mass spectrometry data, J. Chromatogr. A. 1186 (2008) 401–411. https://doi.org/10.1016/j.chroma.2007.10.063.

[44] 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. https://doi.org/10.1016/j.chroma.2015.03.001.

[45] 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. https://doi.org/10.1016/j.talanta.2009.11.046.

[46] 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. https://doi.org/10.1021/ac0511142.

[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. U. S. A. 104 (2007) 16886–16891. https://doi.org/10.1073/pnas.0708365104.

[48] Y. Izadmanesh, E. Garreta-Lara, J.B. Ghasemi, S. Lacorte, V. Matamoros, R. Tauler, Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data, J. Chromatogr. A. 1488 (2017) 113–125. https://doi.org/10.1016/J.CHROMA.2017.01.052.

Chapter 4. Control-Normalized Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data for Enhanced Biomarker Discovery in a Metabolomic Study of Orthopedic Knee-Ligament Injury

This chapter was partially reproduced from Sarah E. Prebihalo[1], Grant S. Ochoa[1], K.L. Berrier[1], K.J. Skogerboe[2], Kenneth L. Cameron[3], Jesse R. Trump[3], Steven J. Svoboda[3], J. Kennth Wickisier[3], Robert E. Synovec[1], "Control-normalized Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data for enhanced biomarker discovery in a metabolomic study of orthopedic injury" Submitted to *Analytical Chemistry* (2020).

[1]Department of Chemistry, University of Washington, Box 351700, Seattle, WA 98195, USA
[2]Department of Chemistry, Seattle University, Seattle. WA 98122, USA
[3]Keller Army Community Hospital, West Point, NY, USA

4.1 INTRODUCTION

Since its introduction in 1991 by Liu and Phillips, comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC×GC-TOFMS) has become an indispensable instrumental platform capable of separating highly complex mixtures [1–7]. Indeed, GC×GC provides a 10-fold or more increase in peak capacity as compared to traditional one-dimensional chromatography, 1D-GC, and simultaneously improves identification and quantification of trace level analytes [2,8,9]. While GC×GC-TOFMS first gained popularity in the petroleum industry [10–18], it has since been extended to other fields such as, but not limited to, food science [5,19], forensics [20,21], wastewater [22,23], and metabolomics [24–26].

Application of GC×GC-TOFMS to highly complex samples is often limited by the availability of chemometric tools to handle the challenging data analysis expectations [9], thus there is a significant effort to address this challenge [6,24–36]. Many chemometric software tools rely upon identifying and quantifying analytes of interest in either a targeted or nontargeted

fashion, where nontargeted implies that the analytes of interest are not known *a priori* [2,9,30]. Nontargeted "discovery" based methods aim to expose meaningful class distinguishing analyte features with limited manual intervention and interpretation necessary. Nontargeted "discovery" methods are further classified as either supervised or unsupervised, depending upon whether class membership is known *a priori* [9,30]. With unsupervised discovery, class membership information is not known *a priori*. An example of this is principal component analysis (PCA), which can either be utilized as standalone chemometric tool for visualization of sample class-based differences, or as a preprocessing method for data reduction prior to analysis with other chemometric methods [31–34]. While PCA is capable of identifying important analyte features, it can be severely hampered by within-class variance [9]. Therefore, if class membership is known, such as when the analyst can design their experiment to define sample classes, supervised methods that account for within-class variance, such as Fisher ratio analysis (F-ratio), are often preferred [6,21,26,27,29,31,33–35,37–40].

Mathematically, the standard F-ratio (^{S}F-ratio) is defined as

$$^{S}\text{F-ratio} = \frac{s_{bc}^2}{s_{wc}^2} \tag{4.1}$$

where s^2_{bc} is the between-class variance and Σs^2_{wc} is the pooled within-class variance [41]. As defined, the standard F-ratio naturally is normalized by the pooled within-class variance, so the ^{S}F-ratio is not biased toward higher intensity chromatographic peaks. Indeed, F-ratio analysis has been successfully applied as a supervised discovery-based tool with GC×GC-TOFMS data sets for a variety of applications such as fuel [6,11,35], forensics [36,40,42,43], and metabolomics [24,26,34]. However, retention time misalignment between chromatograms inflates the false positive rate, and features falsely identified as class distinguishing are intermingled among true positives, which unnecessarily burdens the data analysis work flow [27,29]. While sample-to-

sample retention time alignment across all chromatograms should in principle address this issue, sparse sampling of the first dimension (^{1}D) separation in GC×GC obfuscates the effectiveness of alignment algorithms due to the random nature of the in-phase to out-of-phase sampling of the ^{1}D peaks during modulation that produces the second dimension (^{2}D) separations [24,27,29,44]. Recently, tile-based F-ratio analysis has been introduced as an alternative to retention time alignment, and has been demonstrated to be very effective at optimizing the discovery of class distinguishing features while reducing false positives [6,26,27,29,35]; indeed, ^{S}F-ratio analysis employing the standard definition in eq 1 has been readily implemented when the within-class variance for all samples classes does not dominate the between-class variances of interest.

However, for applications in which there is a large within-class variance for one of the two class in a two-class comparison, the standard F-ratio may struggle to elucidate chemically meaningful features [25,45–47]. While the analyst seeks to discover the overarching differences between two sample classes, they may concurrently seek to discover interesting and potentially relevant sample-to-sample differences within a single class. Unfortunately, this later goal would be severely hampered using the standard F-ratio calculation. For example, one can imagine the standard F-ratio being perilously challenged in the study of clinically obtained samples where one class (eg., a patient class) exhibits significantly greater within-class variance due to injury or illness severity, than the other class (eg., control class). Unquestionably, in studies where mean chemical differences between classes are largely influenced by increased variance of this nature, there exists a need to modify existing metrics so as not to penalize such data sets [48].

To address this need, herein we implement an F-ratio definition using a control-normalized analysis of variance (ANOVA) calculation (^{CN}F-ratio) [48],

$$^{CN}\text{F-ratio} = \frac{s_{bc}^2}{s_{wc,controls}^2} \tag{4.2}$$

where s^2_{bc} is the between-class variance and $s^2_{wc,\ controls}$ is the within-class variance for the control class. The only difference between eqs 4.1 and 4.2 is the omission of the $s^2_{wc,\ patients}$ in the denominator. Both eqs 4.1 and 4.2 are implemented in a complementary fashion to obtain a more comprehensive discovery of biomarkers associated with anterior cruciate ligament (ACL) injuries in human plasma. Due to injury severity and coexisting injury, within-class variance among patients will likely be elevated in comparison to healthy controls, translating into lower biomarker discovery rates when pooled within-class variance is utilized via eq 4.1. As such, control-normalized F-ratio analysis via eq 4.2 should leverage the comparatively lower within-class variance of the controls to facilitate discover of interesting metabolites that exhibit large with-in class variance. The hypothesis is that a control-normalized F-ratio analysis would likely enhance the capability of finding additional biomarkers when used in parallel with standard F-ratio analysis.

One-hundred twenty human plasma samples were obtained from the Keller Army Community Hospital (West Point, NY, USA), consisting of patient samples with matched controls at two time points: time-of-injury and 12 months post-surgery. Previously validated tile-based F-ratio software [26,27,29,35] was modified to use the within-class variance of controls only in the denominator of eq 4.2, maintaining the appropriate degrees-of-freedom [48]. Traditionally, analysts mine the hit list by carefully investigating a subset of hits in an individual fashion working from the top down to a predetermined stopping point to reduce analyst time and effort [26,27,33,40,49]. Herein, control-normalized F-ratio analysis was used in conjunction with standard F-ratio analysis, both with top down data mining, to provide a total of four class comparisons: time-of-injury patients versus controls, and 12 months post-surgery versus controls, using both the standard and control-normalized F-ratio software. We describe the complementary

nature of standard and control-normalized F-ratio analyses to provide a more comprehensive list of potential biomarkers in the case where the patient class is expected to have increased variability.

4.2 EXPERIMENTAL

A total of 120 human plasma samples (30 patients, 30 matched controls for time-of-injury and 12 months post-surgery) were obtained from the Keller Army Community Hospital (West Point, NY, USA) for analysis via GC×GC-TOFMS. The following extraction and derivatization procedure was developed based upon a previous study using non-human primate plasma and performed in batches of 20 samples to maintain sample stability prior to data collection [50]. Briefly, 100 µL of room temperature plasma was transferred to a 1.5 mL microcentrifuge tube and extracted with 900 µL of methanol: water (8:1 v/v). The microcentrifuge tube was then cooled on ice for 10 min, followed by 10 min of sample homogenization using a Vortex-Genie 2 SI-0236 (SI Industries, Inc. Bohemia, New York). Each sample was again cooled on ice for 10 min before centrifugation using a MicoCL 17 centrifuge (Thermo Scientific, Waltham, MA) at 15,000 g. Following centrifugation, 200 µL of supernatant was transferred to a 300 µL GC insert (National Scientific, Rockwood, TN), which was inserted into a clean microcentrifuge tube. Solvent was evaporated until dry using a SpeedVac SVC 100 (Savant). Dried sample extract was immediately stored at -4°C for up to 1 week.

Sample derivatization was performed no more than 24 hr prior to data collection. Each previously dried sample was again dried on the SpeedVac for 40 min prior to the first derivatization step to ensure no water was present. The sample insert containing the dried sample was transferred to a GC vial for the remaining preparation and data collection. A 20 mg/mL solution of methoxyamine hydrochloride (Aldrich, Burlington, MA) in ACS grade pyridine (Fisher Chemical, Waltham, MA) was prepared on the day of derivatization. A 30 µL aliquot of

methoxyamine:pyridine solution was added to each dry sample extract and incubated at 30°C for 90 min with the GC vial capped. Upon completion, the samples were removed from the oven and briefly cooled. A 70 µL aliquot of MSTFA + 0.1 TMCS (Thermo Scientific, Waltham, MA) solution was added prior to incubation at 60°C for 90 min. A laboratory blank for each day of sample preparation was prepared using an identical procedure, except with the plasma sample replaced by extraction solution, to enable differentiation between matrix compounds and true metabolites during data analysis.

A metabolite standard containing 45common metabolites was prepared for the purpose of identifying metabolites in the sample runs (Table 4.1). Six standard solutions were prepared at a final concentration of ~5000 ppm per metabolite by measuring 50 mg of each individual metabolite standard into a vial and diluting with 10 g of HPLC grade water. A 100 µL aliquot of each of the six standard solutions were combined to form the final metabolite standard solution with a concentration of ~500 ppm per metabolite. Next, 200 µL of the ~500 ppm standard solution was spiked into 100 µL plasma and 700 µL 8:1 methanol/water extraction solution to yield a final spike concentration of ~100 ppm per metabolite standard. The spiked plasma sample subsequently underwent preparation via the extraction and derivatization method.

GC×GC–TOFMS data were collected using a Pegasus 4D system (LECO, St. Joseph, MI), consisting of an Agilent 6890N gas chromatograph with a 7683 autosampler (Agilent Technologies, Palo Alto, CA) and a Pegasus III TOFMS with a quad-jet thermal modulator and secondary oven. Capillary columns were obtained from Restek (Bellefonte, PA). A reverse column configuration was employed with a Rxi-17Sil MS (30 m × 250 µm i.d. × 0.25 µm df) primary column, ^{1}D, and Rxi-1 MS (1.5 m × 180 µm i.d. × 0.18 µm df) secondary column, ^{2}D. Instrumental parameters were based on previous studies and summarized below [50]. The GC inlet and transfer

line were each set to 280°C. A 1 µL injection of each sample was made with a 5 µL autosampler syringe (Hamilton, Reno, NV) in split mode with a split ratio of 10:1. The ^{1}D oven was held at 60°C for 1 min and then increased at 5°C/min to a final temperature of 290°C, which was maintained for 10 min. The ^{2}D oven and modulator block followed the same program as the primary oven, with +10°C and +60°C offsets, respectively. The modulation period, P_M, was 2 s with 0.50 s hot and cold pulses for each stage. The carrier gas consisted of ultrahigh purity helium (Grade 5, 99.999%, Praxair, Seattle, WA, USA) maintained at a constant flow of 2 mL/min. The ion source was set to 225°C. Mass spectra were collected for all mass channels (*m/z*) from 33 to 500 at a scan rate of 100 spectra/s, following a 90 s acquisition delay.

Table 4.1 45 metabolite standards analyzed for identification verification.

#	**Compound**	#	**Compound**
1	D-glucose 6-phosphate	24	Butyric acid
2	Methylmalonic acid	25	Formic acid
3	Myo-inositol	26	Heptanoic acid
4	Malonic acid	27	Hexanoic acid
5	Pyruvate	28	Isobutyric acid
6	Maltose	29	Isovaleric acid
7	Malic Acid	30	4-methylvaleric acid
8	Succinic Acid	31	Propionic acid
9	Glycerol-3-phosphate	32	Valeric acid
10	Uracil	33	L-isoleucine
11	Pyroglutamic acid	34	L-leucine
12	Oxaloacetic acid	35	L-lysine
13	Maleic acid	36	L-methionine
14	Glutamic acid	37	L-phenylalanine
15	Fumaric acid	38	L-serine
16	Gluconic acid	39	L-tyrosine
17	D-glucose	40	L-valine
18	D-galactose	41	L-Alanine
19	Adenine	42	L-cystine
20	Palmitic acid	43	L-arginine
21	Stearic acid	44	DL-lactic acid
22	Glycerol	45	Oxalic acid
23	Acetic acid		

GC×GC–TOFMS data were imported from the LECO ChromaTOF software v 3.32 (LECO, St. Joseph, MI) to Matlab v R2019a (MathWorks, Inc., Natick, MA), and the imported data were analyzed on a high-level personal computer, having an Intel Core™ i7-7700K (4.20GHz), a 512 GB Samsung 950 Pro solid-state hard drive, and 64 GB dual-channel DDR4 RAM. For both time points, time-of-injury and 12 months post-surgery, a class comparison was obtained by standard F-ratio analysis using pooled within-class variance (using the tile-based F-ratio software) [27,29], followed by the implementation of a control-normalized F-ratio analysis using control variance only, for a total of four class comparisons. A ^{1}D tile size of 12 s, a ^{2}D tile size of 400 ms, a ^{1}D cluster tile size of 8 s, and a ^{2}D cluster size tile of 200 ms were selected to encompass average peak widths and allow for moderate retention time shifting in both dimensions. Based on previous optimization and validation studies, the average F-ratio was calculated from a maximum of 10 *m/z*, with at least 3 *m/z* required per hit. All *m/z* utilized were above a signal-to-noise (*S/N*) threshold of 10 [26,35]. The chromatographic features (key metabolites) discovered by F-Ratio analysis were arranged in a ranked table "hit list" sorted by decreasing F-Ratio. Two hit lists were obtained for each time point, time-of-injury and 12 months post-surgery, for the standard and modified F-ratio analyses, respectively.

Top down quantitative analysis of metabolites in the F-ratio hit list was performed using an in-house developed signal measurement algorithm described in Figure 4.1 that functionally applies the "two-step" method on the GC×GC data in the unfolded form [42], by performing ^{2}D peak apex finding for pure *m/z* followed by summing the peaks heights for each analyte. The algorithm uses key information gleaned from each F-ratio analysis to determine the signal for given hits for each chromatogram. First, each hit "pin" location in the 2D separation space is used to target the quantitative analysis for each analyte. Second, since the magnitude of the F-ratio

correlates with *m/z* purity, this aids in *m/z* selection for quantification. For each hit, a pure *m/z*, generally associated with the largest F-ratio, was used to calculate the relative signal for each sample. Using the two-step method, determination of *m/z* purity was further assessed by visual inspection of the raw unfolded data using the ChromaTOF software provided with the instrument platform. For the signal at a selected pure *m/z*, a 2D bin with dimensions equal to the ^{1}D tile size (6 s) and 2-times the ^{2}D cluster size (400 ms) is centered on the ^{1}D retention time ($^{1}t_{R}$) and ^{2}D retention time ($^{2}t_{R}$) pin for each hit. Next, the algorithm determines where the original F-ratio pin location is on the peak, i.e. whether it is on the right, left, or top of the ^{2}D peak maxima. Finally, the algorithm uses an iterative process to determine the true peak maxima of each ^{2}D peak and record the signal. For every chromatogram, all ^{2}D peak heights for a given analyte hit are summed to produce one signal per hit. The algorithm then calculates average signals and standard deviations for the control and patient classes, followed by a concentration ratio between patients and controls. Statistical significance of each hit was calculated using Welch's t-test [51,52], which does not assume the variance of the two classes are equal.

Principal component analysis (PCA) was performed on the calculated metabolite signals for the time-of-injury and 12 months post-surgery time points using PLS Toolbox 8.8.1 (Eigenvector Research, Inc., Wenatchee, WA, USA). Only signals which passed the t-test at a 95% confidence level were submitted to PCA [53]. More specifically, PCA was performed on three F-ratio data sets for the time-of-injury and 12 months post-surgery time points: hits from the standard F-ratio analysis alone, control-normalized F-ratio analysis alone, and a combined list of metabolites which passed the t-test in both F-ratio analyses. The data was submitted to PCA as a 2D array, whereby vectors of calculated metabolite signals for each sample chromatogram were concatenated row-wise. For illustration of the complete data analysis workflow see Figure 4.2.

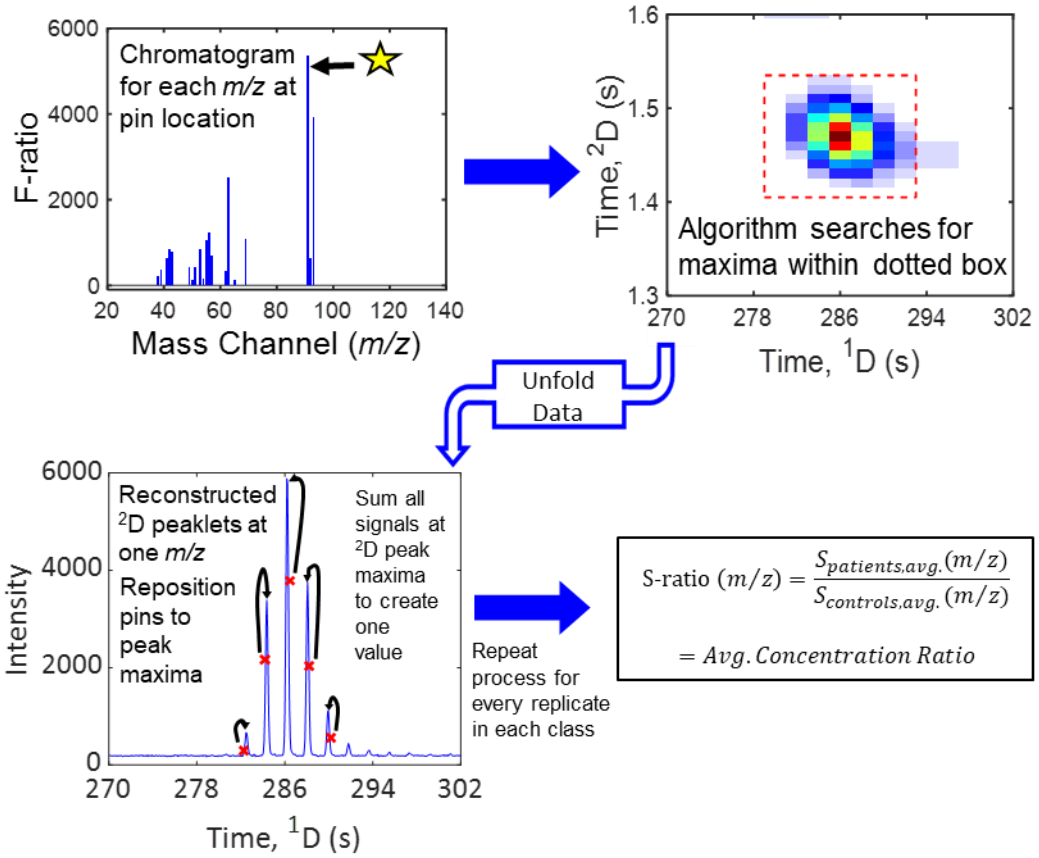

Figure 4.1. Flow chart describing data analysis steps for the algorithm used to calculate average signals, and then concentration ratios of patients vs. controls. Using the F-ratio spectrum output (upper left) as a guide, a selective *m/z* was identified through visual inspection of the peak in ChromaTOF. A 2D tile is cutout around each hit from each chromatogram (upper right) and unfolded to produce a vector containing the hit. Within this window, the ^{2}D peak maxima are located (lower left) and their ^{2}D peak heights summed to yield a single value (lower right), per analyte hit, per sample. Following the calculation of total signal (*S*) for a hit in every chromatogram, the patient signals were averaged and divided by the average of the control signals to give an average Signal ratio (S-ratio). For a sufficiently pure *m/z*, the average S-ratio equals the average concentration ratio (Ochoa G.S., Prebihalo, S.E., Reaser, B.C., Marney. L.C., Synovec, R.E., J. Chromatogr. A. **2020**, 1627.)

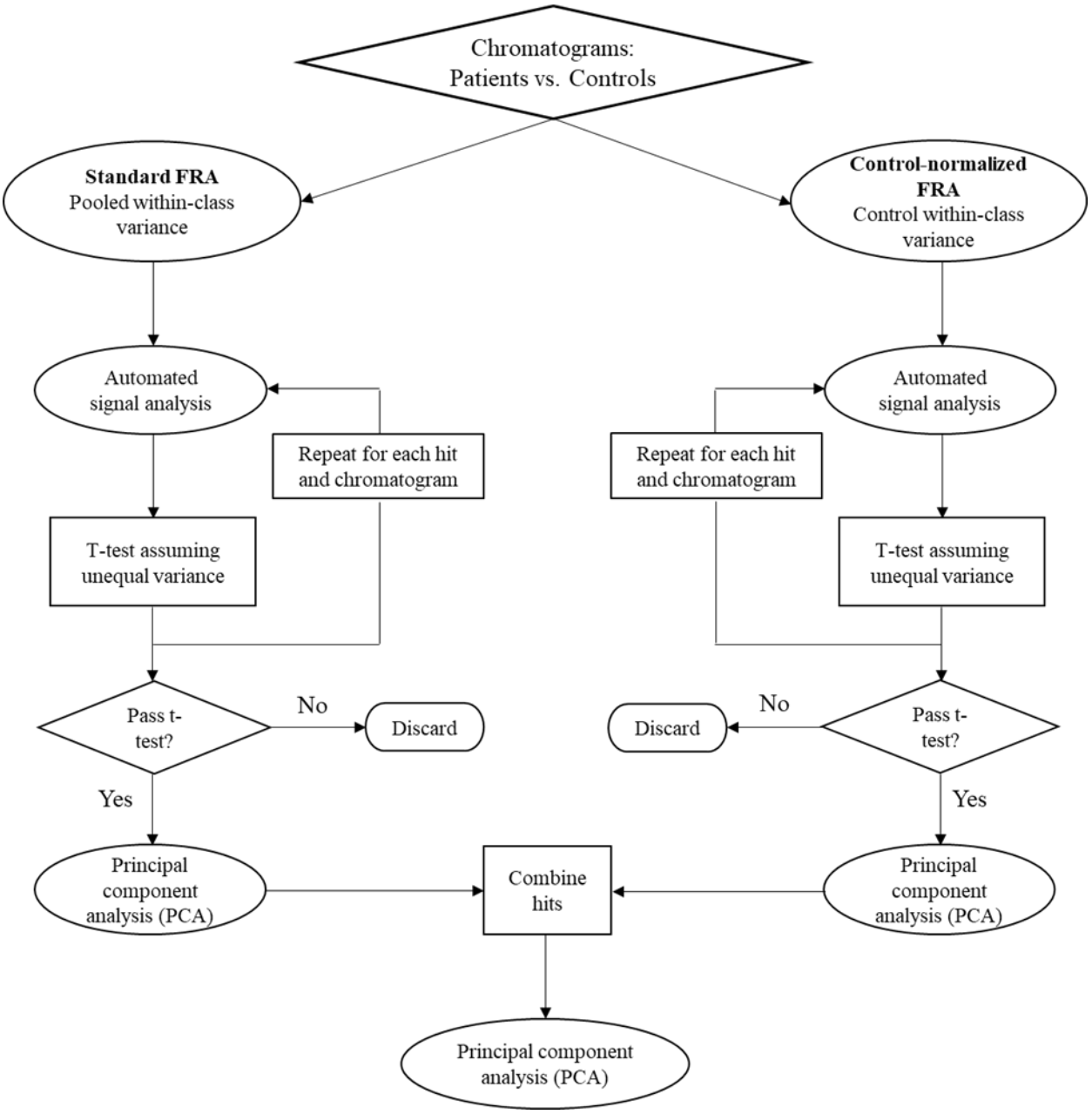

Figure 4.2. Workflow depicting data analysis for patients and controls using the standard and control-normalized F-ratio methods. Following F-ratio analysis, an automated signal algorithm was employed to rapidly estimate signal of all hits in every chromatogram. Statistical analysis was calculated via the t-test assuming unequal variance. Hits that passed a t-value of 2 (according to 30 degrees of freedom and 95% confidence interval) were then input to PCA. Finally, a combined hit list of all hits found using the standard and control-normalized F-ratio methods was also fed into PCA.

4.3 Results and Discussion

The 30 patient and 30 control human plasma samples proved to be an excellent sample set for investigation by control-normalized F-ratio analysis, as the 30 patient samples were expected to have large within-class variance associated with various factors, particularly the injury severity. However, the control samples primarily contained normal biological variation associated with gender, age, and body mass index (BMI), and so on [54]. The complexity of the human metabolome provided a suitably complex sample for GC×GC–TOFMS analysis, where a considerable portion of the 2D space was utilized, presented in Figure 4.3. The presentation of both chromatograms begins at a ^{1}D retention time, ^1t_R, of 5 min due to the presence of a large solvent peak at the beginning of the chromatograms. Further observation of the chromatograms revealed no analytes of interest past 35 min. Subsequently, the data were reduced to the time period 5 to 35 min prior to data analysis.

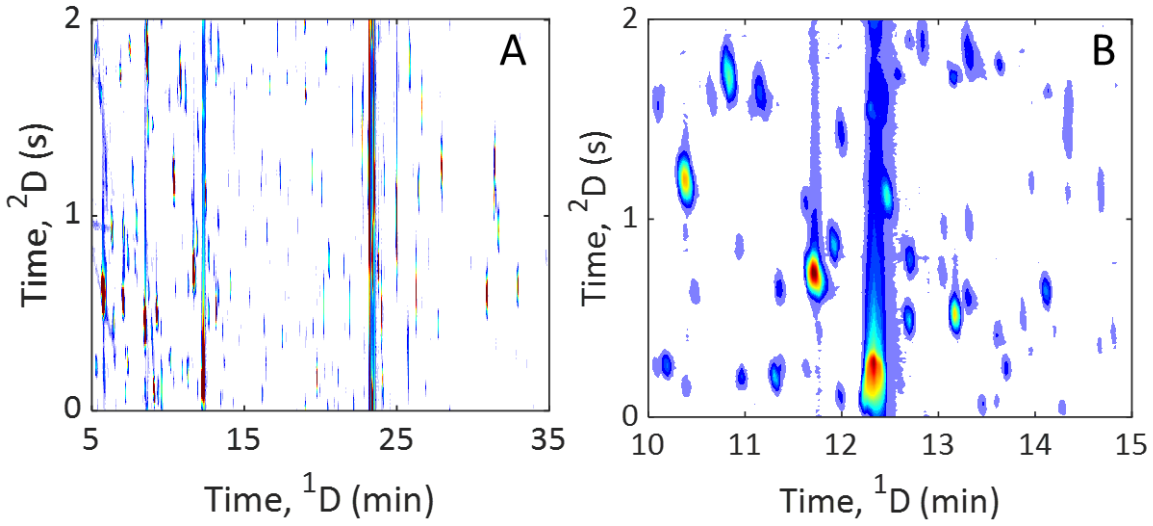

Figure 4.3. Representative GC×GC chromatogram for human plasma samples at *m/z* 73. (A) Total separation time used for F-ratio analysis. (B) Zoom of chromatogram in (A) between 10 and 15 minutes.

Standard tile-based F-ratio analysis via eq 1 was employed for the sample class comparison for the patients versus controls at the time-of-injury time point to generate a hit list of potentially

class distinguishing analytes. The F-ratio distribution in Figure 4.4A was produced with the majority of the ^{S}F-ratios below 10, with the distribution dropping to baseline between an ^{S}F-ratio of 10 and 20 as it moves towards hits with a higher ^{S}F-ratio. Note that the software is designed to produce an F-ratio for every chromatographic peak feature that satisfies the *S/N* threshold and number of *m/z* required as specified in the Experimental. Manual evaluation of all hits in Figure 2A would be computationally prohibitive, as there are a total of 375 hits identified by standard F-ratio analysis. Therefore, careful top-down curation of the hit list is preferred and a summary of the top 30 hits in the hit list is presented in Table 4.2. The largest ^{S}F-ratio observed using standard F-ratio analysis was 98, corresponding to the metabolite L-methionine. The lowest ^{S}F-ratio in this hit list was L-histidine was 23 at hit #30. Going below an ^{S}F-ratio ≈ 20 required the analysis of too many false positives via subsequent t-testing, therefore only the top 30 hits are presented for the purpose of this investigation.

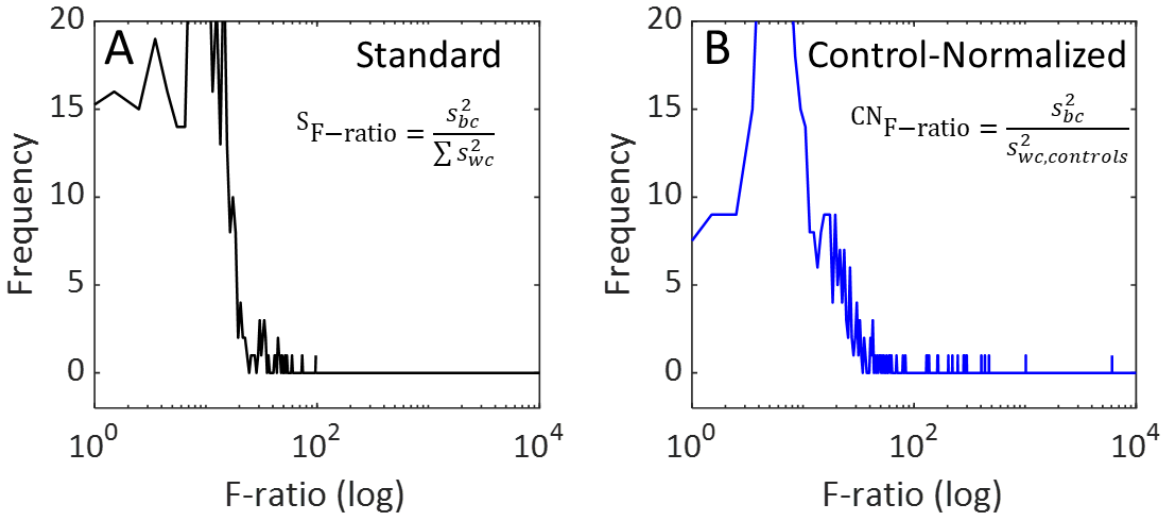

Figure 4.4. (A) F-ratio distribution for the standard F-ratio hit list at time-of-injury, with a maximum F-ratio of 98. Inset is eq 4.1 for standard F-ratio analysis. (B) F-ratio distribution for control-normalized F-ratio hit list at time-of-injury. Seventeen hits are observed to be above an ^{CN}F-ratio of 100. Inset is eq 4.2 for control-normalized F-ratio analysis.

Table 4.2 Top 30 hits using standard F-ratio analysis and the corresponding control-normalized F-ratio for time-of-injury. For a given hit and *m/z*, [P]/[C] is the average signal of patient samples divided by the average signal of control samples. *Verified via 2D retention time and mass spectral match using metabolite standard mix. Green highlighted hits are unique to standard F-ratio analysis; blue highlighted hits are common between control-normalized and standard F-ratio analysis.

ID	^{S}Hit#	^{S}F-ratio	^{CN}Hit#	^{CN}F-ratio	m/z	Average Patient Signal	Average Control Signal	[P]/[C]	t_{expt}
L-methionine*	1	98	29	47	155	317 ± 152	694 ± 184	0.46	8.67
Mannose	2	74	6	299	49	1831 ± 814	609 ± 241	3.01	7.88
Malonic acid*	3	59	26	52	55	171 ± 71	263 ± 66	0.65	5.19
Metabolite Unk. 1	4	54	14	162	49	4360 ± 1675	1268 ± 398	3.44	9.84
Glycine	5	51	8	285	77	4643 ± 2473	3621 ± 856	1.28	2.14
Glucose Fragment*	6	49	19	80	73	52917 ± 24620	50653 ± 15618	1.04	0.43
Alanine*	7	46	32	42	116	5178 ± 3819	4649 ± 1757	1.11	0.69
Pyroglutamic acid*	8	46	206	8	77	932 ± 447	383 ± 237	2.44	5.95
Glycerol*	9	45	36	40	152	730 ± 446	1213 ± 307	0.60	4.89
L-lysine*	10	44	13	165	77	920 ± 474	207 ± 132	4.46	7.94
Urea	11	43	38	36	47	3390 ± 779	3095 ± 685	1.10	1.56
Phenylalanine*	12	42	3	473	73	776 ± 395	319 ± 201	2.44	5.66
Glycopyranose	13	37	66	23	73	18304 ± 14662	29659 ± 15379	0.62	2.93
Oxoisocaproic acid	14	34	43	32	75	346 ± 220	376 ± 80	0.92	0.69
CoA Fragment	15	34	16	133	77	1616 ± 2197	502 ± 185	3.22	2.77
Linoliec acid	16	34	9	280	77	963 ± 638	220 ± 112	4.38	6.28
Ornithine	17	34	18	84	142	652 ± 475	459 ± 342	1.42	1.81
Glutamic acid*	18	33	12	203	77	1583 ± 623	552 ± 351	2.87	7.90
Stearic acid*	19	33	7	295	77	445 ± 217	119 ± 46	3.73	8.05
Palmitic acid*	20	32	5	403	49	881 ± 520	279 ± 179	3.16	6.00
Lactic acid*	21	32	23	60	49	5462 ± 2195	2260 ± 655	2.42	7.65
Metabolite Unk. 2	22	31	85	21	73	583 ± 363	1652 ± 1087	0.35	5.11
Metabolite Unk. 3	23	30	22	60	77	603 ± 373	436 ± 381	1.38	1.71
Glutamine	24	30	11	222	47	135 ± 57	57 ± 27	2.36	6.72
L-tyrosine*	25	30	2	1016	73	2191 ± 1209	1602 ± 1480	1.37	1.69
Metabolite Unk. 4	26	28	168	10	77	197 ± 90	62 ± 24	3.19	7.94
L-tryptophan	27	26	307	5	73	382 ± 422	253 ± 166	1.51	1.57
L-threonine	28	25	30	44	77	831 ± 344	333 ± 170	2.50	7.10
Metabolite Unk. 5	29	23	110	17	73	912 ± 529	850 ± 247	1.07	0.57
L-histidine	30	23	27	49	154	948 ± 841	484 ± 420	1.96	2.71

Likewise, for the time-of-injury patient versus control class comparison using control-normalized F-ratio analysis via eq 4.2, replacing the pooled within-class variance with the variance of controls results in an overall increase of larger ^{CN}F-ratios, as evident in the F-ratio distribution in Figure 4.4B. A total of seventeen hits have an ^{CN}F-ratio larger than the maximum ^{S}F-ratio using the standard F-ratio approach. More importantly, the control-normalized F-ratio analysis ranks hits

in a complementary fashion relative to the standard F-ratio analysis, indicated by inspection of Table 4.3. Here only the top 40 hits (of 358 total hits) are presented since an F-ratio ≈ 30 for the modified F-ratio analysis was roughly the point at which too many false positives indicated via subsequent t-testing began to be intermingled amongst the true positives. The F-ratio of 30 also appears in Figure 4.4B to be the ^{CN}F-ratio at which the frequency of hits with similar ^{CN}F-ratios begins to significantly increase, suggesting an increase in the false positive rate. The maximum ^{CN}F-ratio, 6075, observed as the furthest right F-ratio in Figure 4.4B, corresponds to naproxen, which was significantly upregulated in the patient class with a concentration ratio of ≈ 23. While naproxen is not strictly speaking a metabolite, it does serve to highlight the potential contribution of control-normalized F-ratio analysis towards the uncovering of metabolites which vary among patients. Furthermore, for the seventeen additional hits that were discovered as compared to the standard F-ratio hit list, fourteen passed the t-test at a 95% confidence interval.

Notably, of the top 30 hits discovered via standard F-ratio analysis, only 4 statistically significant hits were not subsequently discovered by control-normalized F-ratio analysis. Per Tables 4.2 and 4.3, two of these metabolites, glycopyranose and metabolite unknown 2, corresponded to ^{CN}Hit #66 and ^{CN}Hit #85 in the control-normalized F-ratio hit list, respectively. Both hits produced ^{CN}F-ratios above 20 but were likely pushed further down in the control-normalized hit list due to the appearance of metabolites with higher ^{CN}F-ratios. This is in stark contrast to the 14 statistically significant hits discovered exclusively through control-normalized F-ratio analysis. Remarkably, arachidonic acid, succinic acid, which have been implicated in inflammation studies [50,55], were discovered via the control-normalized F-ratio approach, yet not found in the top 30 hits using standard F-ratio analysis.

Table 4.3 Top 40 hits using control-normalized F-ratio analysis and the corresponding standard F-ratio for time-of-injury. For a given hit and *m/z*, [P]/[C] is the average signal of patient samples divided by the average signal of control. *Verified via 2D retention time and mass spectral match using metabolite standard mix. Yellow highlighted hits are unique to control-normalized F-ratio analysis; blue highlighted hits are common between control-normalized and standard F-ratio analysis.

ID	^{CN}Hit#	^{CN}F-ratio	^{S}Hit#	^{S}F-ratio	*m/z*	Average Patient Signal	Average Control Signal	[P]/[C]	t_{expt}
Naproxen	1	6075	254	7	73	1566 ± 3050	69 ± 33	22.83	2.69
L-Tyrosine*	2	1016	25	30	73	2191 ± 1209	1602 ± 1480	1.37	1.69
Phenylalanine*	3	473	12	42	73	776 ± 395	319 ± 201	2.44	5.66
Arachidonic acid	4	435	123	18	73	375 ± 378	68 ± 26	5.51	4.43
Palmitic acid*	5	403	20	32	49	881 ± 520	279 ± 179	3.16	6.00
Mannose	6	299	2	74	49	1831 ± 814	609 ± 241	3.01	7.88
Stearic acid*	7	295	19	33	77	445 ± 217	119 ± 46	3.73	8.05
Glycine	8	285	5	51	73	4643 ± 2473	3621 ± 856	1.28	2.14
Linoliec acid	9	280	16	34	77	963 ± 638	220 ± 112	4.38	6.28
Metabolite Unk. 6	10	247	147	11	204	1184 ± 940	564 ± 151	2.10	3.57
Glutamine	11	222	24	30	47	135 ± 57	57 ± 27	2.36	6.72
Glutamic acid*	12	203	18	33	77	1583 ± 623	552 ± 351	2.87	7.90
L-lysine*	13	165	10	44	77	920 ± 474	207 ± 132	4.46	7.94
Metabolite Unk. 1	14	162	4	54	49	4360 ± 1675	1268 ± 398	3.44	9.84
Succinic acid*	15	137	37	20	77	3293 ± 2250	995 ± 291	3.31	5.55
CoA fragment	16	133	15	34	77	1616 ± 2197	502 ± 185	3.22	2.77
Metabolite Unk. 7	17	130	34	21	144	1354 ± 1299	1537 ± 930	0.88	0.63
Ornithine	18	84	17	36	142	652 ± 475	459 ± 342	1.42	1.81
Glucose Fragment*	19	80	6	49	73	52917 ± 24620	50653 ± 15618	1.04	0.43
Metabolite Unk. 8	20	69	68	16	47	1793 ± 1529	1761 ± 566	1.02	0.11
Metabolite Unk. 9	21	63	109	13	77	298 ± 192	79 ± 33	3.77	6.15
Metabolite Unk. 3	22	60	23	30	77	603 ± 373	436 ± 381	1.38	1.71
Lactic acid*	23	60	21	32	49	5462 ± 2195	2260 ± 655	2.42	7.65
6-phospho-D-gluconate	24	56	31	23	58	364 ± 161	265 ± 140	1.37	2.52
Metabolite Unk. 10	25	53	36	21	73	2243 ± 1634	1439 ± 800	1.56	2.42
Malonic acid*	26	52	3	59	55	171 ± 71	263 ± 66	0.65	5.19
L-histidine	27	49	30	23	154	948 ± 841	484 ± 420	1.96	2.71
Metabolite Unk. 11	28	48	130	12	73	313 ± 179	427 ± 109	0.73	2.97
L-methionine*	29	47	1	98	155	317 ± 152	694 ± 184	0.46	8.67
L-threonine	30	44	28	25	77	831 ± 344	333 ± 170	2.50	7.10
Proline	31	43	271	7	77	314 ± 210	90 ± 73	3.48	5.50
Alanine*	32	43	7	46	116	5178 ± 3819	4649 ± 1757	1.11	0.69
Metabolite Unk. 12	33	42	40	19	77	435 ± 254	123 ± 50	3.55	6.62
Aspartic acid	34	42	115	13	49	206 ± 113	87 ± 62	2.37	5.07
Picolinc acid	35	40	187	10	77	2643 ± 1402	1257 ± 312	2.10	5.29
Glycerol*	36	40	9	45	152	730 ± 446	1213 ± 307	0.60	4.89
Metabolite Unk. 13	37	37	36	21	73	345 ± 229	722 ± 541	0.48	3.52
Urea	38	36	11	43	47	3390 ± 779	3095 ± 685	1.10	1.56
Myristate*	39	36	41	19	77	225 ± 99	81 ± 63	2.77	6.71
Maleic acid*	40	33	39	20	73	1643 ± 1443	1885 ± 851	0.87	0.79

Table 4.4 Summary of hits that pass t-test from time-of-injury hit lists using both standard and control-normalized F-ratio. For a given hit and *m/z*, [P]/[C] is the average signal of patient samples divided by the average signal of control. *Verified via 2D retention time and mass spectral match using metabolite standard mix. Yellow highlighted hits are unique to control-normalized F-ratio analysis; green highlighted hits are unique to standard F-ratio analysis; blue highlighted hits are common between control-normalized and standard F-ratio analysis.

ID	^{CN}Hit#	^{CN}F-ratio	^{S}Hit#	^{S}F-ratio	*m/z*	Average Patient Signal	Average Control Signal	[P]/[C]	t_{expt}
Naproxen	1	6075	254	7	73	1566 ± 3050	69 ± 33	22.8	2.7
Phenylalanine*	3	473	12	42	73	776 ± 395	319 ± 201	2.44	5.7
Arachidonic acid	4	435	123	13	73	375 ± 378	68 ± 26	5.51	4.4
Palmitic acid*	5	403	20	32	49	881 ± 520	279 ± 179	3.16	6.0
Mannose	6	299	2	74	49	1831 ± 814	609 ± 241	3.01	7.9
Stearic acid*	7	295	19	33	77	445 ± 217	119 ± 46	3.73	8.1
Glycine	8	285	5	51	73	4643 ± 2473	3621 ± 856	1.28	2.1
Linoliec acid	9	280	16	34	77	963 ± 638	220 ± 112	4.38	6.3
Metabolite Unk. 6	10	247	147	11	204	1184 ± 940	564 ± 151	2.10	3.6
Glutamine	11	222	24	30	47	135 ± 57	57 ± 27	2.36	6.7
Glutamic acid*	12	203	18	33	77	1583 ± 623	552 ± 351	2.87	7.9
L-lysine*	13	165	10	44	77	920 ± 474	207 ± 132	4.46	7.9
Metabolite Unk. 1	14	163	4	54	49	4360 ± 1675	1268 ± 398	3.44	9.8
Succinic acid*	15	137	37	20	77	3293 ± 2250	995 ± 291	3.31	5.6
CoA fragment	16	133	15	34	77	1616 ± 2197	502 ± 185	3.22	2.8
Metabolite Unk. 9	21	63	109	13	77	298 ± 192	79 ± 33	3.77	6.2
Lactic acid*	23	60	21	32	49	5462 ± 2195	2260 ± 655	2.42	7.7
6-phospho-D-gluconate	24	56	31	23	58	364 ± 161	265 ± 140	1.37	2.5
Metabolite Unk. 10	25	53	36	21	73	2243 ± 1634	1439 ± 800	1.56	2.4
Malonic acid*	26	52	3	59	55	171 ± 71	263 ± 66	0.65	5.2
L-histidine	27	49	30	23	154	948 ± 841	484 ± 420	1.96	2.7
Metabolite Unk. 11	28	48	130	12	73	313 ± 179	427 ± 109	0.73	3.0
L-methionine*	29	47	1	98	155	317 ± 152	694 ± 184	0.46	8.7
L-threonine	30	44	28	25	77	831 ± 344	333 ± 170	2.50	7.1
Proline	31	43	271	7	77	314 ± 210	90 ± 73	3.48	5.5
Metabolite Unk. 12	33	42	40	19	77	435 ± 254	123 ± 50	3.55	6.6
Aspartic acid	34	42	115	13	49	206 ± 113	87 ± 62	2.37	5.1
Picolinc acid	35	40	187	10	77	2643 ± 1402	1257 ± 312	2.10	5.3
Glycerol*	36	40	9	45	152	730 ± 446	1213 ± 307	0.60	4.9
Metabolite Unk. 13	37	37	36	21	73	345 ± 229	722 ± 541	0.48	3.5
Myristate*	39	36	41	19	77	225 ± 99	81 ± 63	2.77	6.7
Glycopyranose	66	23	13	37	73	18304 ± 14662	29659 ± 15379	0.62	2.9
Metabolite Unk. 2	85	21	22	31	73	583 ± 363	1652 ± 1087	0.35	5.1
Metabolite Unk. 4	168	10	26	28	77	197 ± 90	62 ± 24	3.19	7.9
Pyroglutamic acid*	206	8	8	46	77	932 ± 447	383 ± 237	2.44	6.0

The interconnected information provided via standard and control-normalized F-ratio is ascertained through PCA of the individual hit lists as well as by inspection of the combined hit list that represents all metabolites discovered by standard and control-normalized F-ratio (Table 4.4). In each case, PCA was performed on exclusively the hits which pass the t-test, which culminated in 21 metabolites (out of the 30 top hits) by standard F-ratio analysis, and 31 metabolites (out of the 40 top hits) by control-normalized F-ratio analysis (for PCA completed with top 30 hits from control-normalized F-ratio analysis, see Appendix A). The standard F-ratio PCA scores plot presented in Figure 4.5A demonstrates modest class separation between patients (green diamonds) and controls (gray squares), with 5 patient samples intermingled with control samples. This same observation holds true in the PCA scores plot representing the 31 metabolites discovered via the control-normalized F-ratio; however, naproxen was excluded as it was not strictly speaking a metabolite (Fig. 4.5(B)). Notably, the controls are tightly clustered in Figure 4.5B, likely corresponding to metabolites with little within-class variance among controls, while the patients exhibit much larger within-class variance. For PCA results including naproxen, see Figure 4.6.

Remarkably, using a hit list curated by combining the results from both the standard and control-normalized F-ratio analyses (35 total hits passing the t-test at 95% confidence) resulted in a PCA scores plot (Figure 4.5C) that visually appears almost identical to that obtained with exclusively the hits from the control-normalized F-ratio analysis. Referring to Tables 4.2-4.4, 14 statistically significant metabolites were discovered via control-normalized F-ratio analysis that were not discovered using standard F-ratio analysis. Comparatively, 4 metabolites were unique to the standard F-ratio hit list and not discovered using the control-normalized F-ratio. Furthermore, the 14 additional metabolites identified by means of the control-normalized F-ratio hit list appear to be significantly hampered by the patient within-class variance, as evident by their corresponding

F-ratio using the standard F-ratio (Tables 4.3-4.4). For the time-of-injury time point, it appears as though the metabolites identified via control-normalized F-ratio are the driving force behind the class separation present in Figure 4.5C. PCA loadings for the PCA scores plot in Figure 4.5C are presented in Figure 4.5D, illustrating the metabolites responsible for class separation along PC 1 and PC 2. Further insight into how the metabolites separate samples in the PC space can be observed in Appendix A.

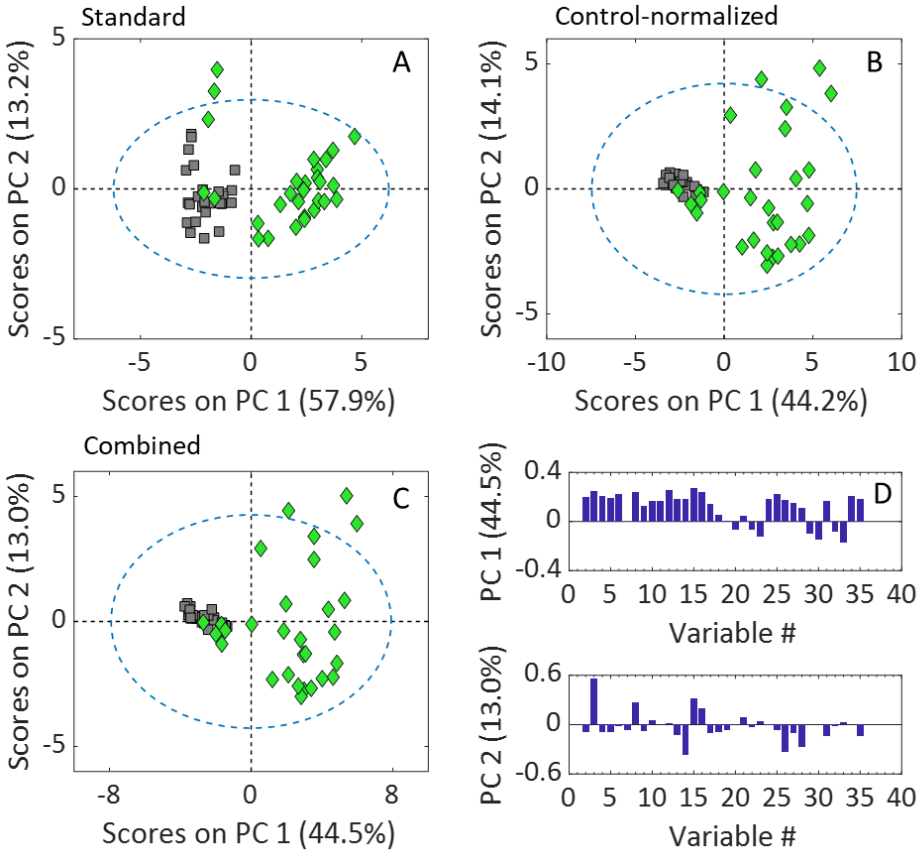

Figure 4.5. Scores on PC 1 versus PC 2 for time-of-injury samples. (A) PCA scores plot for 21 metabolites that pass the t-test from the time-of-injury standard F-ratio hit list. The classes (controls, gray squares; patients, green diamonds) are well separated by PC 1, apart from 5 patients which appear similar to controls. (B) PCA scores plot for 31 metabolites which pass the t-test from the time-of-injury control-normalized F-ratio hit list, excluding naproxen. Controls are tightly clustered indicating the metabolites have minor within-class variance. (C) PCA scores plot prepared from the combined standard and control-normalized F-ratio hit lists, excluding naproxen. The class separation appears functionally the same as the PCA scores plot for the modified hit list in (B). (D) PCA loadings for the scores plot presented in (C), with 35 variables corresponding to the 35 metabolites presented in Table 3 (naproxen excluded).

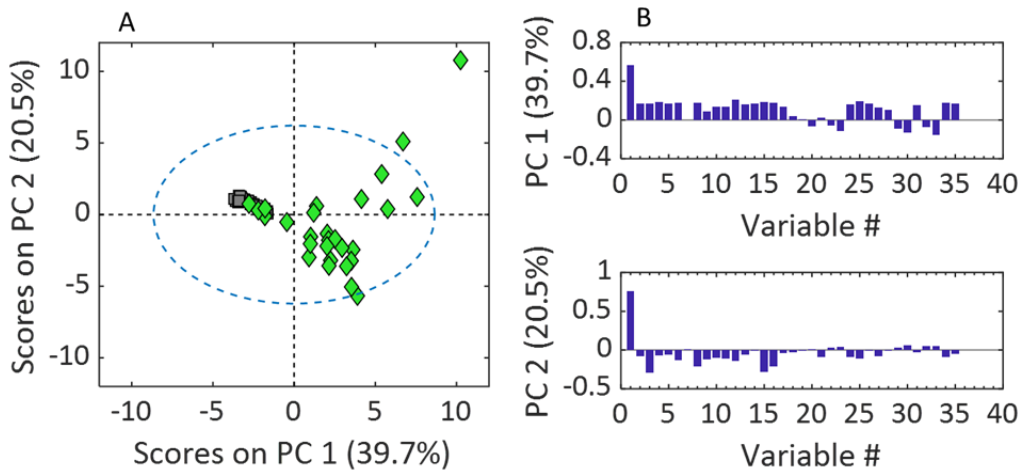

Figure 4.6. (A) PCA scores plot of patients vs. controls for the metabolites that pass the t-test in the top 40 of the time-of-injury control-normalized hit list when naproxen is included. (B) PCA loadings for PC 1 and PC 2 when naproxen (variable #1) is included.

The data analysis workflow presented in Fig. 4.2 was replicated using the 12 months post-surgery time point data. The data were subjected to standard and control-normalized F-ratio analyses, which generated the F-ratio distributions presented in Fig. 4.7.

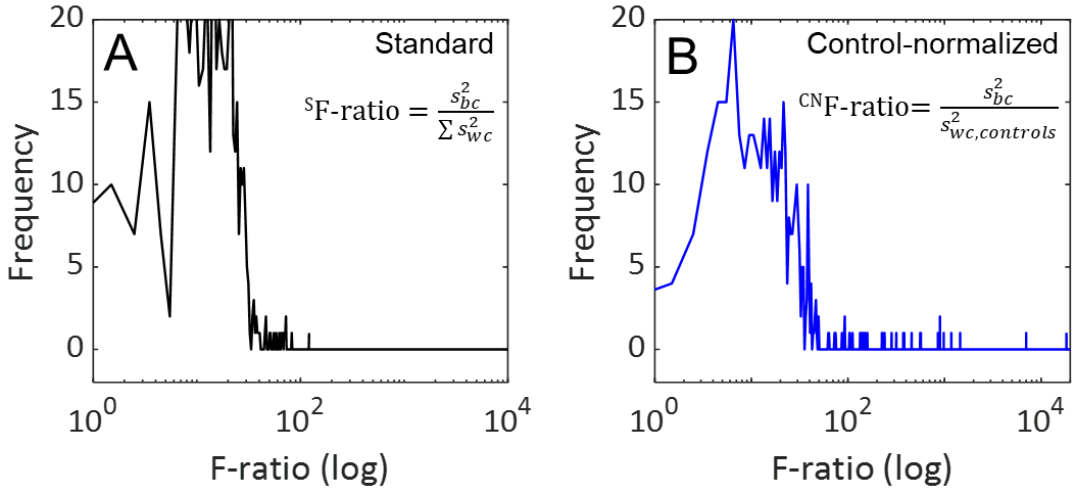

Figure 4.7. F-ratio distributions for patient vs. control class comparison at 12 months post-surgery (A) F-ratio distribution using standard F-ratio software and (B) F-ratio distribution using control-normalized F-ratio algorithm.

For both F-ratio analyses, the distributions were centered approximately at an F-ratio of 6 to 8. However, the tail of the distribution for standard F-ratio analysis extends from an ^{S}F-ratio of

30 to 122, equating to roughly the top 30 hits (of 498 total hits) that were investigated and presented in Table 4.5. Comparatively, the tail of the modified F-ratio distribution extends from roughly 50 to 1.8×10^4, which corresponds to roughly the top 40 hits (419 total hits). Unquestionably, some of these hits were not discovered using exclusively a standard F-ratio analysis. In fact, 21 of the 40 hits discovered using control-normalized F-ratio analysis were not discovered using the standard F-ratio, 15 of which were determined to be statistically significant (Table 4.6). Consistent with the time-of-injury time point, PCA was performed individually on the top 30 hits from the hit lists obtained via the standard and control-normalized F-ratio analyses which passed the t-test, as presented in Figure 4.8. For an updated table containing only the top 30 hits from control-normalized used in PCA, see Appendix A. Further information related to the loadings of all three PCA scores plots can be found in Appendix A.

Table 4.5 Top 30 hits using standard F-ratio analysis and the corresponding control-normalized F-ratio for 12 months post-surgery. For a given hit and *m/z*, [P]/[C] is the average signal of patient samples divided by the average signal of control samples. *Verified via 2D retention time and mass spectral match using metabolite standard mix. Green highlighted hits are unique to standard F-ratio analysis; blue highlighted hits are common between control-normalized and standard F-ratio analysis.

ID	^{S}Hit#	^{S}F-ratio	^{CN}Hit#	^{CN}F-ratio	m/z	Average Patient Signal	Average Control Signal	[P]/[C]	t_{expt}
L-methionine*	1	122	22	147	155	365 ± 216	626 ± 170	0.58	5.2
Metabolite Unk. 8	2	83	15	283	47	1520 ± 757	1943 ± 697	0.78	2.3
CoA fragment	3	73	23	142	77	5768 ± 6657	800 ± 312	7.21	4.1
Metabolite Unk. 14	4	72	17	229	73	1760 ± 1202	2489 ± 1344	0.71	2.2
Metabolite Unk. 15	5	71	48	42	73	82 ± 48	66 ± 29	1.24	1.6
Alanine*	6	70	30	93	116	2628 ± 2061	5402 ± 2419	0.49	4.8
Metabolite Unk. 16	7	69	44	46	73	8826 ± 5823	17764 ± 5266	0.50	6.2
Mannose	8	67	6	898	49	2414 ± 620	741 ± 212	3.26	14.0
Oxoisocaproic acid	9	67	39	49	75	353 ± 186	385 ± 148	0.91	0.8
Malonic acid*	10	66	46	44	55	167 ± 96	276 ± 81	0.61	4.7
Lactic acid*	11	62	9	568	49	5187 ± 2518	2589 ± 835	2.00	5.4
Stearic acid*	12	59	2	7008	77	556 ± 265	133 ± 67	4.19	8.5
Metabolite Unk. 7	13	56	14	317	144	1342 ± 695	1761 ± 773	0.75	2.4
Metabolite Unk. 13	14	52	161	22	73	292 ± 197	708 ± 350	0.41	5.7
Urea	15	50	43	46	47	3479 ± 1705	3234 ± 572	1.08	0.8
Metabolite Unk. 17	16	46	69	37	73	390 ± 237	702 ± 205	0.56	5.4
Oxalic acid*	17	46	32	91	73	10273 ± 7715	21428 ± 7146	0.48	5.8
Metabolite Unk. 18	18	46	59	39	73	68 ± 39	78 ± 20	0.88	1.2
Myo-inositol	19	41	12	383	73	1954 ± 763	1688 ± 554	1.16	1.6
Metabolite Unk. 19	20	39	26	137	73	1719 ± 848	2086 ± 837	0.82	1.7
Glutamine	21	39	89	31	47	116 ± 46	58 ± 25	2.00	6.1
Metabolite Unk. 4	22	38	19	158	77	163 ± 107	61 ± 28	2.67	5.1
Metabolite Unk. 1	23	37	3	1450	49	5118 ± 1926	1628 ± 526	3.14	9.6
Glycerol*	24	37	146	24	152	600 ± 312	1153 ± 344	0.52	6.5
Succinic acid*	25	36	4	1175	77	5018 ± 2700	1155 ± 398	4.34	7.8
Metabolite Unk. 2	26	35	199	19	73	612 ± 431	1560 ± 783	0.39	5.8
Glutamic acid*	27	35	10	559	77	1283 ± 683	712 ± 342	1.80	4.1
Glycine	28	34	97	30	73	109 ± 83	99 ± 51	0.94	0.7
Linoliec acid	29	34	11	456	77	998 ± 638	332 ± 410	3.01	4.8
Maleic acid*	30	32	16	239	73	3237 ± 1554	2544 ± 855	1.27	2.1

Table 4.6 Top 40 hits using control-normalized F-ratio analysis for 12 months post-surgery. For a given hit and *m/z*, [P]/[C] is the average signal of patient samples divided by the average signal of control samples. Yellow highlighted hits are unique to control-normalized F-ratio analysis; blue highlighted hits are common between control-normalized and standard F-ratio analysis.

ID	^{CN}Hit#	^{CN}F-ratio	^{S}Hit#	^{S}F-ratio	m/z	Average Patient Signal	Average Control Signal	[P]/[C]	t_{expt}
Palmitic acid*	1	18290	185	20	49	1063 ± 494	305 ± 205	3.48	7.75
Stearic acid*	2	7008	12	59	77	556 ± 265	133 ± 67	4.19	8.47
Metabolite Unk. 1	3	1450	23	37	49	5118 ± 1926	1628 ± 526	3.14	9.57
Succinic acid*	4	1175	25	36	77	5018 ± 2700	1155 ± 398	4.34	7.75
6-phospho-D-gluconate	5	981	209	18	58	426 ± 403	281 ± 150	1.52	1.85
Mannose	6	898	8	67	49	2414 ± 620	741 ± 212	3.26	13.99
Picolinc acid	7	898	69	27	77	3409 ± 1938	1561 ± 457	2.18	5.08
Pyroglutamic acid*	8	854	88	25	77	1203 ± 664	415 ± 244	2.90	6.10
Lactic acid*	9	568	11	62	49	5187 ± 2518	2589 ± 835	2.00	5.36
Glutamic acid*	10	559	27	35	77	1283 ± 683	712 ± 342	1.80	4.09
Linoliec acid	11	456	29	34	77	998 ± 638	332 ± 410	3.01	4.81
Myo-inositol	12	383	19	41	73	1954 ± 763	1688 ± 554	1.16	1.55
Metabolite Unk. 20	13	374	439	7	73	6979 ± 5268	2910 ± 1374	2.40	4.09
Metabolite Unk. 7	14	317	13	56	144	1342 ± 695	1761 ± 773	0.75	2.36
Metabolite Unk. 8	15	283	2	83	47	1520 ± 757	1943 ± 697	0.78	2.25
Maleic acid*	16	239	30	32	73	3237 ± 1554	2544 ± 855	1.27	2.14
Metabolite Unk. 14	17	229	4	72	73	1760 ± 1202	2489 ± 1344	0.71	2.21
Metabolite Unk. 21	18	225	451	6	73	165 ± 312	21 ± 14	7.97	2.53
Metabolite Unk. 4	19	158	22	38	77	163 ± 107	61 ± 28	2.67	5.05
Metabolite Unk. 22	20	157	212	18	73	1560 ± 1001	1696 ± 988	0.92	0.53
Myristate*	21	150	82	26	77	286 ± 190	93 ± 57	3.08	5.35
L-methionine*	22	147	1	122	155	365 ± 216	626 ± 170	0.58	5.20
CoA fragment	23	142	3	73	77	5768 ± 6657	800 ± 312	7.21	4.08
Metabolite Unk. 23	24	140	161	21	73	2507 ± 1051	1803 ± 551	1.39	3.24
Glycopyranose	25	138	309	13	73	25781 ± 30604	24360 ± 17841	1.06	0.22
Metabolite Unk. 19	26	134	20	39	73	1719 ± 848	2086 ± 837	0.82	1.69
L-tryptophan	27	112	113	23	73	374 ± 306	197 ± 171	1.90	2.77
Metabolite Unk. 12	28	105	31	31	77	594 ± 404	181 ± 72	3.28	5.51
Metabolite Unk. 24	29	103	297	14	73	1297 ± 646	2114 ± 787	0.61	4.4
Alanine*	30	93	6	70	116	2628 ± 2061	5402 ± 2419	0.49	4.78
Arachidonic acid	31	93	36	31	73	225 ± 145	85 ± 39	2.64	5.11
Oxalic acid*	32	91	17	46	73	10273 ± 7715	21428 ± 7146	0.48	5.81
Metabolite Unk. 6	33	87	403	8	204	987 ± 510	603 ± 206	1.62	3.74
L-histidine	34	75	136	22	154	599 ± 539	420 ± 254	1.43	1.65
Metabolite Unk. 25	35	73	270	15	73	282 ± 321	31 ± 15	9.09	4.28
Metabolite Unk. 26	36	64	410	8	73	97 ± 118	39 ± 19	2.47	2.66
Metabolite Unk. 27	37	63	422	8	73	280 ± 270	529 ± 288	0.53	3.46
Metabolite Unk. 28	38	50	452	5	75	60 ± 34	45 ± 47	1.33	1.42
Oxoisocaproic acid	39	49	9	67	75	353 ± 186	385 ± 148	0.91	0.76
Metabolite Unk. 10	40	48	165	21	73	2170 ± 1445	1907 ± 852	1.14	0.86

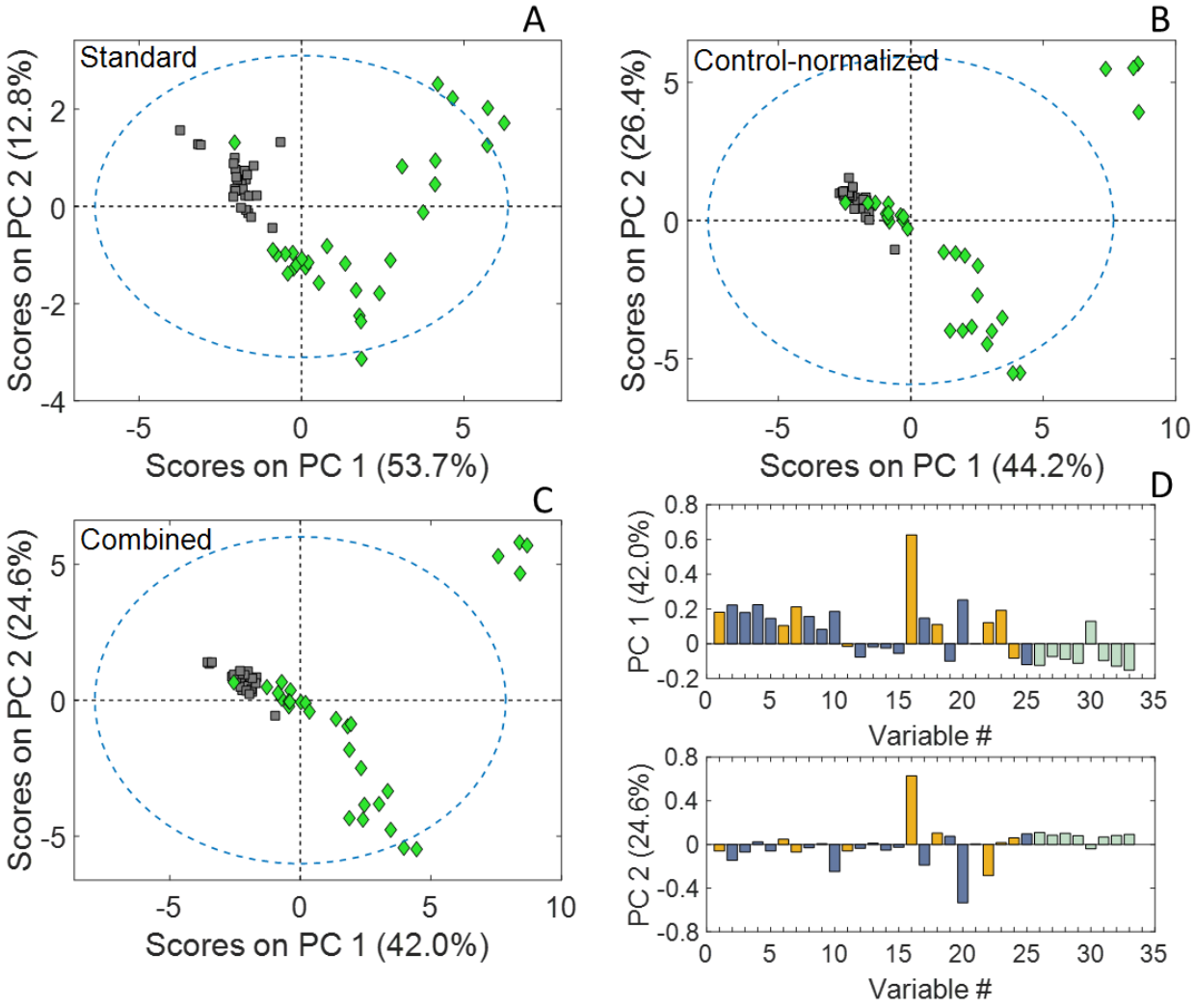

Figure 4.8. Scores on PC 1 versus PC 2 for 12 months post-surgery samples for the top 30 hits of each method (Appendix A, Tables A.2 and A.3). (A) PCA scores plot prepared using hits discovered via standard F-ratio analysis (which pass the t-test) shows modest class separation between controls (gray squares) and patients (green diamonds). (B) PCA scores plot of control-normalized F-ratio hit list. Twelve patient samples appear to be intermingled with control samples, while 4 patient samples are significantly separated along PC 2. (C) PCA scores plot prepared with combined hit list (Appendix A, Table A.4). As at the time-of-injury time point, the combined PCA scores plot looks remarkably like the PCA scores plot prepared using exclusively the control-normalized F-ratio hit list. (D) Loadings plot for PCA scores plot presented in (C) of the combined hit list. Variable 16 is evident to be causing most of the separation of 4 patient samples on PC 2.

Table 4.7 Summary of hits that pass t-test from 12 months post-surgery hit lists using both standard and control-normalized F-ratio. For a given hit and *m/z*, [P]/[C] is the average signal of patient samples divided by the average signal of control. *Verified via 2D retention time and mass spectral match using metabolite standard mix. Yellow highlighted hits are unique to control-normalized F-ratio analysis; green highlighted hits are unique to standard F-ratio analysis; blue highlighted hits are common between control-normalized and standard F-ratio analysis.

ID	^{CN}Hit#	^{CN}F-ratio	^{S}Hit#	^{S}F-ratio	m/z	Average Patient Signal	Average Control Signal	[P]/[C]	t_{expt}
Palmitic acid*	1	18290	185	20	49	1063 ± 494	305 ± 205	3.48	7.75
Stearic acid*	2	7008	12	59	77	556 ± 265	133 ± 67	4.19	8.47
Metabolite Unk. 1	3	1450	23	37	49	5118 ± 1926	1628 ± 526	3.14	9.57
Succinic acid*	4	1175	25	36	77	5018 ± 2700	1155 ± 398	4.34	7.75
Mannose	6	898	8	67	49	2414 ± 620	741 ± 212	3.26	14.0
Picolinc acid	7	898	69	27	77	3049 ± 1938	1561 ± 457	2.18	5.08
Pyroglutamic acid*	8	854	88	25	77	1203 ± 664	415 ± 244	2.90	6.10
Lactic acid*	9	568	11	62	49	5187 ± 2518	2589 ± 835	2.00	5.36
Glutamic acid*	10	559	27	35	77	1283 ± 683	712 ± 342	1.80	4.09
Linoliec acid	11	456	29	34	77	998 ± 638	332 ± 410	3.01	4.81
Metabolite Unk. 20	13	374	439	7	73	6979 ± 5268	2910 ± 1374	2.40	4.09
Metabolite Unk. 7	14	317	13	56	144	1342 ± 695	1761 ± 773	0.75	2.36
Metabolite Unk. 8	15	283	2	83	47	1520 ± 757	1943 ± 697	0.78	2.25
Maleic acid*	16	239	30	32	73	3237 ± 1554	2544 ± 855	1.27	2.14
Metabolite Unk. 14	17	229	4	72	73	1760 ± 1202	2489 ± 1344	0.71	2.21
Metabolite Unk. 21	18	225	451	6	73	165 ± 312	21 ± 14	7.97	2.53
Metabolite Unk. 4	19	158	22	38	77	163 ± 107	61 ± 28	2.67	5.05
Myristate*	21	150	82	26	77	286 ± 190	93 ± 57	3.08	5.35
L-methionine*	22	147	1	122	155	365 ± 216	626 ± 170	0.58	5.20
CoA fragment	23	142	3	73	77	5768 ± 6657	800 ± 312	7.21	4.08
Metabolite Unk. 23	24	140	161	21	73	2507 ± 1051	1803 ± 551	1.39	3.24
L-tryptophan	27	112	113	23	73	374 ± 306	197 ± 171	1.90	2.77
Metabolite Unk. 12	28	105	31	31	77	594 ± 404	181 ± 72	3.28	5.51
Metabolite Unk. 24	29	103	297	14	73	1297 ± 646	2114 ± 787	0.61	4.4
Alanine*	30	93	6	70	116	2628 ± 2061	5402 ± 2419	0.49	4.78
Arachidonic acid	31	93	36	31	73	225 ± 145	85 ± 39	2.64	5.11
Oxalic acid*	32	91	17	46	73	10273 ± 7715	21428 ± 7146	0.48	5.81
Metabolite Unk. 6	33	87	403	8	204	978 ± 510	603 ± 206	1.62	3.74
Metabolite Unk. 25	35	73	270	15	73	282 ± 321	31 ± 15	9.09	4.28
Metabolite Unk. 26	36	64	410	8	73	97 ± 118	39 ± 19	2.47	2.66
Metabolite Unk. 27	37	63	422	8	73	280 ± 270	529 ± 288	0.53	3.46
Metabolite Unk. 16	44	46	7	69	73	8826 ± 5823	17764 ± 5266	0.50	6.24
Malonic acid*	46	44	10	66	55	167 ± 96	276 ± 81	0.61	4.72
Metabolite Unk. 17	69	37	16	46	73	390 ± 237	702 ± 205	0.56	5.44
Glutamine	89	31	21	39	47	116 ± 46	58 ± 25	2.00	6.09
Glycerol*	146	24	24	37	152	600 ± 312	1153 ± 344	0.52	6.53
Metabolite Unk. 13	161	22	14	52	73	292 ± 197	708 ± 350	0.41	5.66
Metabolite Unk. 2	199	19	26	35	73	612 ± 431	1560 ± 783	0.39	5.81

According to the workflow used for time-of-injury, a combined hit list consisting of hits discovered using the standard F-ratio and control-normalized F-ratio software in parallel was prepared, of which 38 passed the t-test and are reported in Table 4.7. Interestingly, there appear to be 4 patients which are heavily loaded on PC 2, observed in Figure 4.8C. Examination of the loadings provided in Figure 4.8D for the combined hit list results in the emergence of variable 16 (Unknown #21) as the likely source of this separation. Tantamount to the time-of-injury time point, the 15 metabolites discovered via control-normalized F-ratio analysis that pass the t-test emerge as the dominating metabolites that separate patients from controls, as the PCA scores plots obtained using the control-normalized and combined hit lists are indistinguishable. To serve as a baseline check, the metabolites in common between Tables 4.2 and 4.3, and Tables 4.5 and 4.6, as well as urea, PCA was performed using only the time-of-injury controls and 12 months post-surgery controls (Figure 4.9). Results indicate no discernable separation between controls at the two time points.

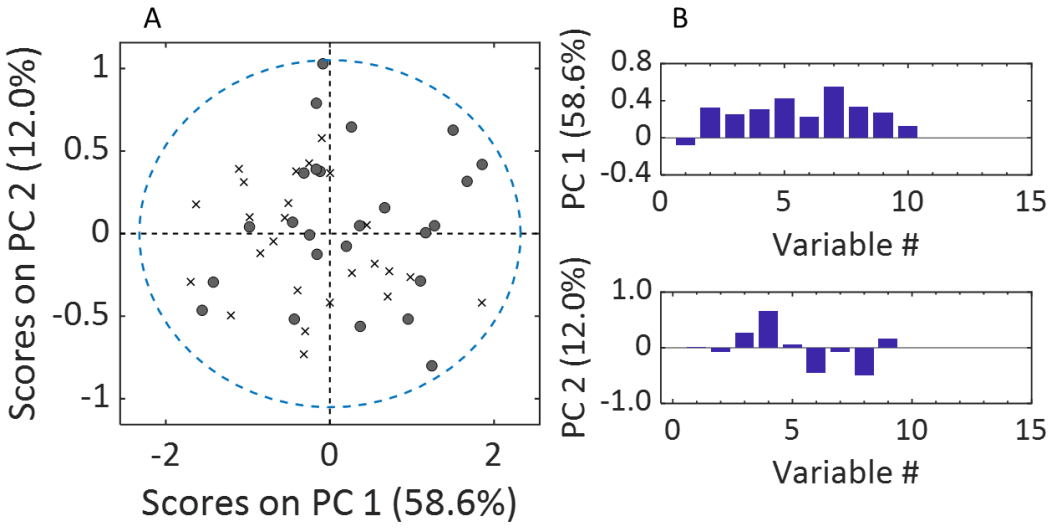

Figure 4.9. (A) PCA scores plot of time-of-injury controls (gray circles) vs. 12 months post-surgery controls (gray x) using metabolites common between 4 hit lists: time-of-injury standard, time-of-injury control-normalized, 12 months post-surgery standard, and 12 months post-surgery control-normalized. (B) PC 1 and PC 2 loadings for scores plot in (A).

Table 4.8 Ten variables (metabolites) in common between Tables 4.2 and 4.3, and Tables 4.5 and 4.6, as well as urea, for PCA.

Variable #	ID	Variable #	ID
1	L-methionine	6	Linoleic acid
2	Mannose	7	Glutamic acid
3	Metabolite Unknown 1	8	Stearic acid
4	Alanine	9	Lactic acid
5	CoA Fragment	10	Urea

Beyond the broad investigation of the class distinguishing ability of metabolites discovered via standard and control-normalized F-ratio analyses, a more detailed examination of specific metabolites provided in Figure 4.10. Four notable metabolites provide a visual perspective on the benefits and challenges of both F-ratio analyses and the potential for complementary information obtained through using the F-ratio analyses in parallel. Figures 4.10A, 4.10C, 4.10E, and 4.10G correspond to the time-of-injury time point, and each are paired with their 12 months post-surgery response (Figures 4.10B, 4.10D, 4.10F, and 4.10H). To begin, while naproxen is not a metabolite indicative of injury, it does provide an example of the capacity of the control-normalized F-ratio to enhance the discovery of potential biomarkers. As observed in Figure 4.10A, naproxen is not present in any appreciable level in the control samples (gray circles) but is elevated in nine patient samples (green circles). Relating back to Tables 4.2 and 4.3, naproxen would not be discovered using the standard F-ratio software, as its F-ratio was only 7, despite the average concentration ratio of patients relative to controls of ~ 23. On the other hand, naproxen was the top hit when using a control-normalized F-ratio, with an F-ratio of 6075. It is important to note that since naproxen was only elevated in two samples at the 12 months post-surgery time point, it was not "discovered" since the F-ratio analysis was constrained by a *S/N* threshold that that required analyte presence in more than two samples [27,36]. This constraint could be changed to meet the needs of a given study.

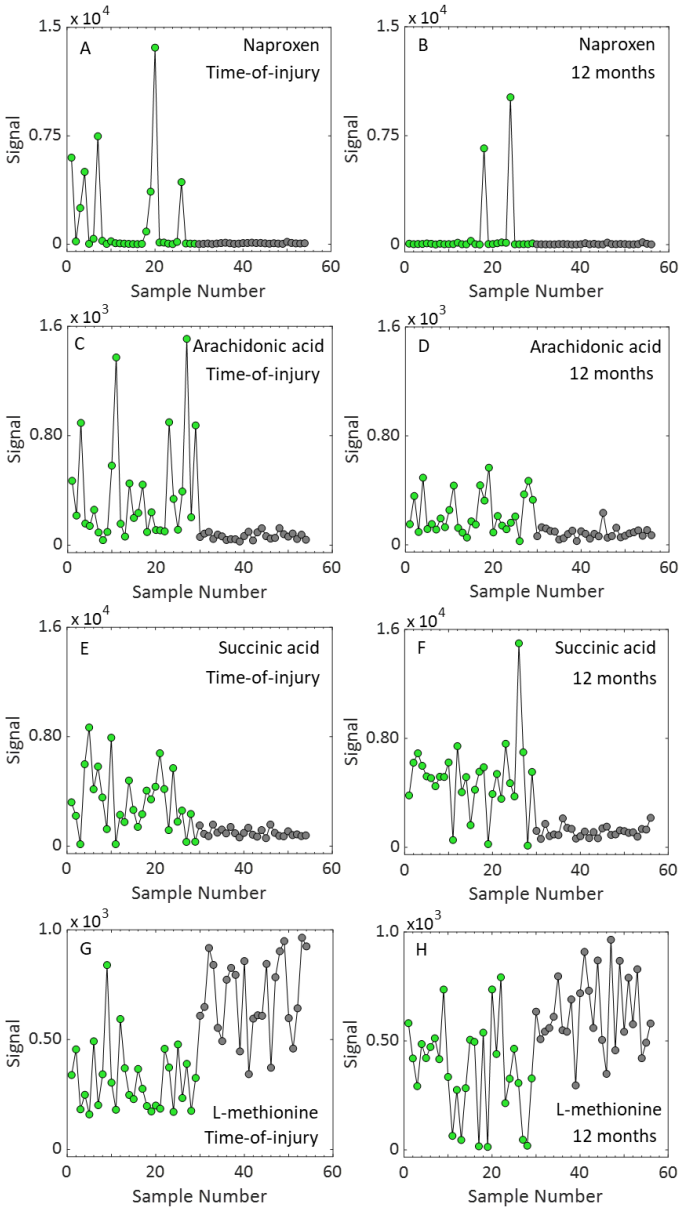

Figure 4.10. Signal plots for four notable metabolites. (A, C, E, G) Time-of-injury and (B, D, F, H) 12 months post-surgery. Each signal plot contains patient samples (green circles) and control samples (gray circles). (A) Naproxen, *m/z* 73 (^{CN}Hit #1), elevated in 9 patients. (B) Naproxen, *m/z* 73, upregulated in two patients. (C) Arachidonic acid, *m/z* 73, ^{CN}Hit #15, ^{CN}F-ratio = 137; ^{S}Hit #37, ^{S}F-ratio = 20upregulated in most patients (^{CN}Hit #4, ^{CN}F-ratio = 435; ^{S}Hit #123, ^{S}F-ratio = 13). (D) Arachidonic acid, *m/z* 73, compared to time-of-injury, signal in patients is approaching that of controls. (E) Succinic acid, *m/z* 77, modestly upregulated in patients (^{CN}Hit #15, ^{CN}F-ratio = 137; ^{S}Hit #37, ^{S}F-ratio = 20). (F) Succinic acid, *m/z* 77, remain modestly upregulated, with one exception. (G) L-methionine, *m/z* 155, ^{S}Hit #1 in standard F-ratio analysis and modestly downregulated in patients. (H) L-methionine, *m/z* 155, remains as ^{S}Hit #1 and is modestly downregulated in patients.

The superiority of the control-normalized F-ratio analysis is further highlighted by two metabolites which have been previously implicated in inflammation studies [55,56], arachidonic acid and succinic acid. Both were discovered using the control-normalized F-ratio with an ^{CN}F-ratio of 435 (^{CN}Hit #3) and 137 (^{CN}Hit #15), respectively. In contrast, due to the high within-class variance in the patient class for both metabolites, observed in Figures 4.10C and 4.10E, their ^{S}F-ratios were significantly hampered when using standard F-ratio analysis (arachidonic acid, ^{S}Hit #123, ^{S}F-ratio 13; succinic acid, ^{S}Hit #37, ^{S}F-ratio 20). While succinic acid remained moderately upregulated in patients at 12 months post-surgery (Figure 4.10F), as whole, arachidonic acid returned closer in signal to the control class over time (Figure 4.10D).

To finish, L-methionine was identified as the top hit in both time points when using the standard F-ratio approach (Table 4.2 and Table 4.5), serving to highlight the utility of using standard F-ratio analysis in parallel with modified F-ratio analysis. For instance, at the time-of-injury (Figure 4.10G), L-methionine is observed to be moderately downregulated in patient samples, which is a trend maintained when moving to the 12 months post-surgery time point (Figure 4.10H). In both cases, while L-methionine is discovered via control-normalized F-ratio analysis, its rank in the hit list (time-of-injury, ^{CN}Hit #29; 12 months post-surgery, ^{CN}Hit #22) is diminished presumably due to the emergence of other hits such as succinic acid and arachidonic acid.

4.4 CONCLUSION

Control-normalized F-ratio analysis leverages the relatively low within-class variance associated with a control class, to enhance the discovery of potentially important metabolites that exhibit high within-class variance in the patient class. The control-normalized F-ratio calculation is a simple modification of the standard F-ratio calculation, by using only the within-class variance

for a control class for F-ratio normalization instead of using the pooled within-class variance for both classes as with the standard F-ratio calculation. Control-normalized F-ratio analysis was complementary to applying standard F-ratio analysis by discovering many metabolites that exhibited a statistically significant concentration difference: 14 additional metabolites at a time-of-injury time point and 15 metabolites at 12 months post-surgery. Additionally, PCA visualization of the combined hit list of statistically significant metabolites indicated that the scores plot clustering of the two sample classes was dominated by the metabolites discovered using the control-normalized F-ratio analysis, with the standard F-ratio analysis playing only a modest supportive role. The results supported the hypothesis that the control-normalized F-ratio analysis coupled to the standard F-ratio analysis has the capacity to enhance the discovery of biomarkers in highly complex samples, such as human plasma. Finally, it is anticipated that control-normalized F-ratio analysis could be applied to other applications that contain one class which is expected to have significantly less within-class variance, effectively increasing the discovery power of F-ratio analysis.

4.5 ACKNOWLEDGMENTS

The author would like to acknowledge the support of Keller Army Community Hospital and for providing the human plasma samples.

4.6 REFERENCES

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

[2] J.V. Seeley, S.K. Seeley, Multidimensional Gas Chromatography: Fundamental Advances and New Applications, Anal. Chem. 85 (2013) 557–578. https://doi.org/10.1021/ac303195u.

[3] D. Megson, R. Kalin, P.J. Worsfold, C. Gauchotte-Lindsay, D.G. Patterson, M.C. Lohan, S. Comber, T.A. Brown, G. O'Sullivan, Fingerprinting polychlorinated biphenyls in environmental samples using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A. 1318 (2013) 276–283. https://doi.org/10.1016/j.chroma.2013.10.016.

[4] O. Pani, T. Górecki, Comprehensive two-dimensional gas chromatography (GC×GC) in environmental analysis and monitoring, Anal. Bioanal. Chem. 386 (2006) 1013–1023. https://doi.org/10.1007/s00216-006-0568-1.

[5] S.-T. Chin, G.T. Eyres, P.J. Marriott, Application of integrated comprehensive/multidimensional gas chromatography with mass spectrometry and olfactometry for aroma analysis in wine and coffee, Food Chem. 185 (2015) 355–361. https://doi.org/10.1016/j.foodchem.2015.04.003.

[6] 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. https://doi.org/10.1016/j.chroma.2016.02.067.

[7] K.L. Berrier, C.E. Freye, M.C. Billingsley, R.E. Synovec, Predictive Modeling of Aerospace Fuel Properties Using Comprehensive Two-Dimensional Gas Chromatography with Time-Of-Flight Mass Spectrometry and Partial Least Squares Analysis, Energy Fuels. 34 (2020) 4084–4094. https://doi.org/10.1021/acs.energyfuels.9b04108.

[8] 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. https://doi.org/10.1016/j.chroma.2015.01.031.

[9] S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, R.E. Synovec, Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications, Anal. Chem. 90 (2018) 505–532. https://doi.org/10.1021/acs.analchem.7b04226.

[10] I. Aloisi, M. Zoccali, P.Q. Tranchida, L. Mondello, Analysis of Organic Sulphur Compounds in Coal Tar by Using Comprehensive Two-Dimensional Gas Chromatography-High Resolution Time-of-Flight Mass Spectrometry, Separations. 7 (2020) 26. https://doi.org/10.3390/separations7020026.

[11] B.J. Pollo, G.L. Alexandrino, F. Augusto, L.W. Hantao, The impact of comprehensive two-dimensional gas chromatography on oil & gas analysis: Recent advances and applications in petroleum industry, TrAC Trends Anal. Chem. 105 (2018) 202–217. https://doi.org/10.1016/j.trac.2018.05.007.

[12] M.R. Jacobs, M. Edwards, T. Górecki, P.N. Nesterenko, R.A. Shellie, Evaluation of a miniaturised single-stage thermal modulator for comprehensive two-dimensional gas chromatography of petroleum contaminated soils, J. Chromatogr. A. 1463 (2016) 162–168. https://doi.org/10.1016/j.chroma.2016.08.009.

[13] K. Qian, F.C. Wang, Compositional Analysis of Heavy Petroleum Distillates by Comprehensive Two-dimensional Gas Chromatography, Field Ionization and High-resolution Mass Spectrometry, J. Am. Soc. Mass Spectrom. 30 (2019) 2785–2794. https://doi.org/10.1021/jasms.8b06302.

[14] D.M. Coutinho, D. França, G. Vanini, L.A.N. Mendes, A.O. Gomes, V.B. Pereira, B.M.F. Ávila, D.A. Azevedo, Rapid hydrocarbon group-type semi-quantification in crude oils by comprehensive two-dimensional gas chromatography, Fuel. 220 (2018) 379–388. https://doi.org/10.1016/j.fuel.2018.02.009.

[15] M.E. Machado, Comprehensive two-dimensional gas chromatography for the analysis of nitrogen-containing compounds in fossil fuels: A review, Talanta. 198 (2019) 263–276. https://doi.org/10.1016/j.talanta.2019.02.031.

[16] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, Recent developments in comprehensive two-dimensional gas chromatography (GC×GC): III. Applications for petrochemicals and organohalogens, TrAC Trends Anal. Chem. 25 (2006) 726–741. https://doi.org/10.1016/j.trac.2006.03.005.

[17] L.W. Hantao, A. Najafi, C. Zhang, F. Augusto, J.L. Anderson, Tuning the Selectivity of Ionic Liquid Stationary Phases for Enhanced Separation of Nonpolar Analytes in Kerosene Using Multidimensional Gas Chromatography, Anal. Chem. 86 (2014) 3717–3721. https://doi.org/10.1021/ac5004129.

[18] N.G.S. Mogollon, F.A. de L. Ribeiro, M.M. Lopez, L.W. Hantao, R.J. Poppi, F. Augusto, Quantitative analysis of biodiesel in blends of biodiesel and conventional diesel by comprehensive two-dimensional gas chromatography and multivariate curve resolution, Anal. Chim. Acta. 796 (2013) 130–136. https://doi.org/10.1016/j.aca.2013.07.071.

[19] G. Purcaro, L. Barp, M. Beccaria, L.S. Conte, Characterisation of minor components in vegetable oil by comprehensive gas chromatography with dual detection, Food Chem. 212 (2016) 730–738. https://doi.org/10.1016/j.foodchem.2016.06.048.

[20] A.A.S. Sampat, M. Lopatka, G. Vivó-Truyols, P.J. Schoenmakers, A.C. van Asten, Towards chemical profiling of ignitable liquids with comprehensive two-dimensional gas chromatography: Exploring forensic application to neat white spirits, Forensic Sci. Int. 267 (2016) 183–195. https://doi.org/10.1016/j.forsciint.2016.08.006.

[21] P. Armstrong, K.D. Nizio, K.A. Perrault, S.L. Forbes, Establishing the volatile profile of pig carcasses as analogues for human decomposition during the early postmortem period, Heliyon. 2 (2016) e00070. https://doi.org/10.1016/j.heliyon.2016.e00070.

[22] N. Ochiai, T. Ieda, K. Sasamoto, Y. Takazawa, S. Hashimoto, A. Fushimi, K. Tanabe, Stir bar sorptive extraction and comprehensive two-dimensional gas chromatography coupled to high-resolution time-of-flight mass spectrometry for ultra-trace analysis of organochlorine pesticides in river water, J. Chromatogr. A. 1218 (2011) 6851–6860. https://doi.org/10.1016/j.chroma.2011.08.027.

[23] S. Prebihalo, A. Brockman, J. Cochran, F.L. Dorman, Determination of emerging contaminants in wastewater utilizing comprehensive two-dimensional gas-chromatography coupled with time-of-flight mass spectrometry, J. Chromatogr. A. 1419 (2015) 109–115. https://doi.org/10.1016/j.chroma.2015.09.080.

[24] 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. https://doi.org/10.1016/j.chroma.2015.03.001.

[25] R. Pesesse, P.-H. Stefanuto, F. Schleich, R. Louis, J.-F. Focant, Multimodal chemometric approach for the analysis of human exhaled breath in lung cancer patients by TD-GC × GC-TOFMS, J. Chromatogr. B. 1114–1115 (2019) 146–153. https://doi.org/10.1016/j.jchromb.2019.01.029.

[26] 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. https://doi.org/10.1016/j.chroma.2016.06.067.

[27] 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. https://doi.org/10.1021/ac504472s.

[28] 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. https://doi.org/10.1016/j.chroma.2012.05.050.

[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. https://doi.org/10.1016/j.talanta.2013.06.038.

[30] 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. https://doi.org/10.1016/j.chroma.2011.07.046.

[31] J.E. Welke, V. Manfroi, M. Zanus, M. Lazzarotto, C. Alcaraz Zini, Differentiation of wines according to grape variety using multivariate analysis of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection data, Food Chem. 141 (2013) 3897–3905. https://doi.org/10.1016/j.foodchem.2013.06.100.

[32] P.-H. Stefanuto, K.A. Perrault, L.M. Dubois, B. L'Homme, C. Allen, C. Loughnane, N. Ochiai, J.-F. Focant, Advanced method optimization for volatile aroma profiling of beer using two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A. 1507 (2017) 45–52. https://doi.org/10.1016/j.chroma.2017.05.064.

[33] 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. https://doi.org/10.1016/j.chroma.2014.10.057.

[34] L.M. Dubois, K.A. Perrault, P.-H. Stefanuto, S. Koschinski, M. Edwards, L. McGregor, J.-F. Focant, Thermal desorption comprehensive two-dimensional gas chromatography coupled to variable-energy electron ionization time-of-flight mass spectrometry for monitoring subtle changes in volatile organic compound profiles of human blood, J. Chromatogr. A. 1501 (2017) 117–127. https://doi.org/10.1016/j.chroma.2017.04.026.

[35] B.C. Reaser, B.W. Wright, R.E. Synovec, Using Receiver Operating Characteristic Curves To Optimize Discovery-Based Software with Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry, Anal. Chem. 89 (2017) 3606–3612. https://doi.org/10.1021/acs.analchem.6b04991.

[36] M. Schäffer, T. Gröger, M. Pütz, S. Dieckmann, R. Zimmermann, Comparative Analysis of the Chemical Profiles of MDMA based on GCxGCTOFMS.pdf, J. Forensic Sci. 57 (2012) 1181–1189.

[37] K.D. Nizio, M. Ueland, B.H. Stuart, S.L. Forbes, The analysis of textiles associated with decomposing remains as a natural training aid for cadaver-detection dogs, Forensic Chem. 5 (2017) 33–45. https://doi.org/10.1016/j.forc.2017.06.002.

[38] L.F. Oliveira, S.C.G.N. Braga, F. Augusto, J.C. Hashimoto, P. Efraim, R.J. Poppi, Differentiation of cocoa nibs from distinct origins using comprehensive two-dimensional gas chromatography and multivariate analysis, Food Res. Int. 90 (2016) 133–138. https://doi.org/10.1016/j.foodres.2016.10.047.

[39] F. Magagna, A. Guglielmetti, E. Liberto, S.E. Reichenbach, E. Allegrucci, G. Gobino, C. Bicchi, C. Cordero, Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination, J. Agric. Food Chem. 65 (2017) 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167.

[40] P.-H. Stefanuto, K. Perrault, S. Stadler, R. Pesesse, M. Brokl, S. Forbes, J.-F. Focant, Reading Cadaveric Decomposition Chemistry with a New Pair of Glasses, ChemPlusChem. 79 (2014) 786–789. https://doi.org/10.1002/cplu.201402003.

[41] 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. https://doi.org/10.1021/ac0602625.

[42] S. Stadler, P.-H. Stefanuto, M. Brokl, S.L. Forbes, J.-F. Focant, Characterization of Volatile Organic Compounds from Human Analogue Decomposition Using Thermal Desorption Coupled to Comprehensive Two-Dimensional Gas Chromatography–Time-of-Flight Mass Spectrometry, Anal. Chem. 85 (2013) 998–1005. https://doi.org/10.1021/ac302614y.

[43] 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. https://doi.org/10.1007/s00216-015-8683-5.

[44] 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. https://doi.org/10.1039/B700061H.

[45] A. Doebbe, M. Keck, M.L. Russa, J.H. Mussgnug, B. Hankamer, E. Tekçe, K. Niehaus, O. Kruse, The Interplay of Proton, Electron, and Metabolite Supply for Photosynthetic H2 Production in Chlamydomonas reinhardtii, J. Biol. Chem. 285 (2010) 30247–30260. https://doi.org/10.1074/jbc.M110.122812.

[46] M.C. Rosso, E. Liberto, N. Spigolon, M. Fontana, M. Somenzi, C. Bicchi, C. Cordero, Evolution of potent odorants within the volatile metabolome of high-quality hazelnuts (Corylus avellana L.): evaluation by comprehensive two-dimensional gas chromatography coupled with mass spectrometry, Anal. Bioanal. Chem. 410 (2018) 3491–3506. https://doi.org/10.1007/s00216-017-0832-6.

[47] L.A. Adutwum, A.P. de la Mata, H.D. Bean, J.E. Hill, J.J. Harynuk, Estimation of start and stop numbers for cluster resolution feature selection algorithm: an empirical approach using null distribution analysis of Fisher ratios, Anal. Bioanal. Chem. 409 (2017) 6699–6708. https://doi.org/10.1007/s00216-017-0628-8.

[48] C. Brownie, D.D. Boos, J. Hughes-Oliver, Modifying the t and ANOVA F Tests When Treatment Is Expected to Increase Variability Relative to Controls, Biometrics. 46 (1990) 259–266. https://doi.org/10.2307/2531650.

[49] 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. https://doi.org/10.1039/C5AY00371G.

[50] A.C. Beckstrom, E.M. Humston, L.R. Snyder, R.E. Synovec, S.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, J. Chromatogr. A. 1218 (2011) 1899–1906. https://doi.org/10.1016/j.chroma.2011.01.086.

[51] B.L. Welch, The Generalization of `Student's' Problem when Several Different Population Variances are Involved, Biometrika. 34 (1947) 28–35. https://doi.org/10.2307/2332510.

[52] B. Derrick, P. White, Why Welch's test is Type I error robust, Quant. Methods Psychol. 12 (2016) 30–38. https://doi.org/10.20982/tqmp.12.1.p030.

[53] Appendix 04: Critical Values for t-Test, Chem. Libr. (2013). https://chem.libretexts.org/Bookshelves/Ancillary_Materials/Reference/Reference_Tables/Analytic_References/Appendix_04%3A_Critical_Values_for_t-Test (accessed June 12, 2020).

[54] E.A. Higgins Keppler, C.L. Jenkins, T.J. Davis, H.D. Bean, Advances in the application of comprehensive two-dimensional gas chromatography in metabolomics, TrAC Trends Anal. Chem. 109 (2018) 275–286. https://doi.org/10.1016/j.trac.2018.10.015.

[55] E. Mills, L.A.J. O'Neill, Succinate: a metabolic signal in inflammation, Trends Cell Biol. 24 (2014) 313–320. https://doi.org/10.1016/j.tcb.2013.11.008.

[56] G.S. Jutley, S.P. Young, Metabolomics to identify biomarkers and as a predictive tool in inflammatory diseases, Best Pract. Res. Clin. Rheumatol. 29 (2015) 770–782. https://doi.org/10.1016/j.berh.2016.02.010.

Chapter 5. Conclusions and Future Directions

Comprehensive two-dimensional gas chromatography coupled with mass spectrometry detection (GC×GC-TOFMS) has been a substantial advancement in modern chromatography over the last 30 years and is well-suited for a variety of samples across many applications. Considering that this analytical technique is commonly applied to study complex mixtures, the resulting data contains a wealth of information. Chemometric methods are often applied to address the associated analytical challenges, such as the discovery of class distinguishing analytes or the decomposition, quantification, and identification of overlapped analyte peaks. The goal of the work presented herein was aimed at enhancing the chemical information gleaned from GC×GC data in the context of chemometric benefits and challenges, as well as the investigation of the impact of experimental design on data structure and quantification accuracy.

5.1 Chapter 2 Summary and Future Directions

Presented in Chapter 2 was the investigation of experimental conditions on quantification accuracy in the context of providing some general guidelines for the practitioner, so parallel factor analysis (PARAFAC) can be more optimally applied. The term Trilinear Deviation Ratio, *TDR*, can be used to predict the quantitative accuracy of PARAFAC, and relates to the deviations in trilinearity that can occur due to retention time shifting from one modulation to the next, $\Delta^2 t_R$, normalized to the width-at-base of the ^{2}D peak, ^2W_b. Both figures-of-merit are influenced by experimental parameters such as column selection, temperature ramp, and modulation period, P_M. Four sets of experimental conditions were explored, aimed at encompassing the common experimental set-ups within the field of GC×GC. It was demonstrated in Chapter 2 that column selection directly impacts the P_M selection and observed ^2k' range. A β_R of less than ~0.5 and a

short P_M on the order of 2-3 seconds reduces the change in temperature between modulations, ΔT, which translates into smaller $\Delta^2 t_R$. As a result of smaller $\Delta^2 t_R$ contributes to *TDR* values less than 0.1, corresponding to PARAFAC quantification % error less than 4%. In contrast, Chapter 2 also demonstrates that adverse impact of longer P_M on quantification accuracy. When an analyst applies GC×GC conditions with $\beta_R \approx 1$, such as two of the previously mentioned experiments, a longer P_M is necessary to elute all analytes. As a result, the range of possible *TDR*s increases to upwards of ~15% for heavily retained analytes, especially for long modulation periods which do not efficiently use the ^{2}D separation.

While it was beyond the scope of the investigation presented herein, an efficacious continuation of the previous study would be to study the ability of measured $\Delta^2 t_R$ to aid in retention time shifting of data prior to PARAFAC decomposition and quantification. The hypothesis is that it would be possible for an analyst to analyze a subset of analytes which span the ^2k' range for the chosen column set as a guide for the magnitude of retention time shifting that is necessary to provide accurate quantification results. In practice, an analyst could select an analyte from the representative standard which displays similar ^2k' for a given set of experimental conditions, and shift the ^{2}D peaks from the analyte of interest by the measured $\Delta^2 t_R$ in an effort to decrease quantification error. Such a procedure could benefit an analyst that chooses to use PARAFAC over another decomposition method, e.g. MCR-ALS, due to rotational ambiguity. However, established in Chapter 2 is the fact that an analyst can reduce deviations in trilinearity that lead to quantification errors through careful consideration of experimental parameters.

5.2 Chapter 3 Summary and Future Directions

Presented in Chapter 3 was the evaluation of the capability of F-ratio analysis to provide quantification information and validated by calculating S-ratios (and true concentration ratios) as

a function of *m/z* between diesel spiked with 15 non-native analytes at four concentration levels, resulting in three different pairwise class comparisons. Introduced in Chapter 3 is the development of a novel S-ratio algorithm, which provides relative quantification information for each hit discovered by tile-based F-ratio analysis while minimizing the need for chemometric decomposition. Reported in Chapter 3 is the use of statistical metric tools for ascertaining mass selectivity (and quantitative accuracy). Lack-of-fit was introduced as a measure of peak purity and when combined with the use of a p-value threshold, resulted in deviations < 5% of s-ratio values relative to the true concentration ratios for the top *m/z* for each analyte.

Building upon the work presented herein, a valuable continuation of the study presented in Chapter 3 is two-fold. First, the analytes studied in Chapter 3 were not natively found in diesel fuel, resulting in relatively high concentration ratios of ~2. However, analytes that are natively present in diesel fuel are amenable to this method as well. The evaluation of spiked analytes that are natively found in diesel at appreciable concentrations would be advantageous, as the pre-existing concentration will result in smaller S-ratios. Additional study of these analyte types is warranted to discover the limits of S-ratio quantification. The second extension of the work presented herein is the application to discover spectrally pure m/z as an advanced feature selection step prior to methods such as PCA and partial-least squares (PLS). Current research building on these principles is being conducted by G.S. Ochoa.

5.3 Chapter 4 Summary and Future Directions

Chapter 4 included the evaluation of a control-normalized F-ratio calculation to provide complementary information as standard F-ratio for a comprehensive metabolite profile associated with anterior cruciate ligament (ACL) injury. Simply, the control-normalized F-ratio leverages the relatively low within-class variance associated with a control class by replacing the pooled within-

class variance in the denominator of the standard F-ratio calculation with the within-class variance computed from controls only. Thirty human plasma samples from patients with ACL injuries were compared to thirty matched controls at two time points to result in a total of four F-ratio comparisons: time-of-injury and 12 months post-surgery using both the standard and control-normalized F-ratio calculations. It was hypothesized in Chapter 4 that the use of a control-normalized F-ratio would provide valuable class distinguishing information as the F-ratio values would not be inversely impacted by the high patient within-class variance. Indeed, Chapter 4 presented evidence that the metabolites discovered via control-normalized F-ratio dominated the class separation when all metabolite features discovered using both standard and control-normalized F-ratio were submitted to PCA. Additionally, the standard F-ratio analysis played only a modest supporting role in the separation of patients from controls. Specifically, 14 additional metabolites were exclusively discovered at the time-of-injury time point when the control-normalized F-ratio was used, and 15 metabolites were exclusively discovered at 12 months post-surgery when the control-normalized F-ratio calculation was utilized.

Further study into the control-normalized F-ratio method presented in Chapter 4 would be beneficial to increased understanding of the benefits and limitations of the control-normalized F-ratio. For instance, one can imagine applying the method to other applications which contains one class that is expected to a have less within-class variance than the second class. However, an efficacious continuation of this work would be to evaluate the limitation of the software in the context of a spiked data set which gives the analyst a baseline for the class distinguishing analytes, as well as the use of a control-normalized F-ratio feature selection step prior to other chemometric methods, such as PLS.

BIBLIOGRAPHY

A.D. McNaught, A. Wilkinson, 1997. IUPAC. Compendium of Chemical Terminology (the "Gold Book"), 2nd ed. Blackwell Scientific Publications, Oxford.

Adahchour, M., Beens, J., Vreuls, R.J.J., Brinkman, U.A.T., 2006. Recent developments in comprehensive two-dimensional gas chromatography (GC×GC): III. Applications for petrochemicals and organohalogens. TrAC Trends Anal. Chem., On-site Instrumentation and Analysis 25, 726–741. https://doi.org/10.1016/j.trac.2006.03.005

Adutwum, L.A., de la Mata, A.P., Bean, H.D., Hill, J.E., Harynuk, J.J., 2017. Estimation of start and stop numbers for cluster resolution feature selection algorithm: an empirical approach using null distribution analysis of Fisher ratios. Anal. Bioanal. Chem. 409, 6699–6708. https://doi.org/10.1007/s00216-017-0628-8

Aloisi, I., Zoccali, M., Tranchida, P.Q., Mondello, L., 2020. Analysis of Organic Sulphur Compounds in Coal Tar by Using Comprehensive Two-Dimensional Gas Chromatography-High Resolution Time-of-Flight Mass Spectrometry. Separations 7, 26. https://doi.org/10.3390/separations7020026

Amigo, J.M., Skov, T., Bro, R., Coello, J., Maspoch, S., 2008. Solving GC-MS problems with PARAFAC2. TrAC Trends Anal. Chem. 27, 714–725. https://doi.org/10.1016/j.trac.2008.05.011

Appendix 04: Critical Values for t-Test [WWW Document], 2013. . Chem. Libr. URL https://chem.libretexts.org/Bookshelves/Ancillary_Materials/Reference/Reference_Tables/Analytic_References/Appendix_04%3A_Critical_Values_for_t-Test (accessed 6.12.20).

Aprea, E., Gika, H., Carlin, S., Theodoridis, G., Vrhovsek, U., Mattivi, F., 2011. Metabolite profiling on apple volatile content based on solid phase microextraction and gas-chromatography time of flight mass spectrometry. J. Chromatogr. A 1218, 4517–4524. https://doi.org/10.1016/j.chroma.2011.05.019

Armstrong, P., Nizio, K.D., Perrault, K.A., Forbes, S.L., 2016. Establishing the volatile profile of pig carcasses as analogues for human decomposition during the early postmortem period. Heliyon 2, e00070. https://doi.org/10.1016/j.heliyon.2016.e00070

Bartle, K.D., Myers, P., 2002. History of gas chromatography. TrAC Trends Anal. Chem. 21, 547–557. https://doi.org/10.1016/S0165-9936(02)00806-3

Bean, H.D., Hill, J.E., Dimandja, J.-M.D., 2015a. 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, 111–117. https://doi.org/10.1016/j.chroma.2015.03.001

Beckstrom, A.C., Humston, E.M., Snyder, L.R., Synovec, R.E., Juul, S.E., 2011. 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. J. Chromatogr. A 1218, 1899–1906. https://doi.org/10.1016/j.chroma.2011.01.086

Berrier, K.L., Freye, C.E., Billingsley, M.C., Synovec, R.E., 2020. Predictive Modeling of Aerospace Fuel Properties Using Comprehensive Two-Dimensional Gas Chromatography with Time-Of-Flight Mass Spectrometry and Partial Least Squares Analysis. Energy Fuels 34, 4084–4094. https://doi.org/10.1021/acs.energyfuels.9b04108

Bezemer, K.D.B., Koeberg, M., Heijden, A.E.D.M. van der, Driel, C.A. van, Blaga, C., Bruinsma, J., Asten, A.C. van, 2016. The Potential of Isotope Ratio Mass Spectrometry

(IRMS) and Gas Chromatography-IRMS Analysis of Triacetone Triperoxide in Forensic Explosives Investigations. J. Forensic Sci. 61, 1198–1207. https://doi.org/10.1111/1556-4029.13135

Blumberg, L.M., 2011a. Multidimensional Gas Chromatography: Theoretical Considerations, in: Mondello, L. (Ed.), Comprehensive Chromatography in Combination with Mass Spectrometry. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 13–63. https://doi.org/10.1002/9781118003466.ch2

Blumberg, L.M., 2008. Accumulating resampling (modulation) in comprehensive two-dimensional capillary GC (GC×GC). J. Sep. Sci. 31, 3358–3365. https://doi.org/10.1002/jssc.200800424

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

Brokl, M., Bishop, L., Wright, C.G., Liu, C., McAdam, K., Focant, J.-F., 2014a. 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, 216–229. https://doi.org/10.1016/j.chroma.2014.10.057

Brownie, C., Boos, D.D., Hughes-Oliver, J., 1990. Modifying the t and ANOVA F Tests When Treatment Is Expected to Increase Variability Relative to Controls. Biometrics 46, 259–266. https://doi.org/10.2307/2531650

Chai, M., Pawliszyn, J., 1995. Analysis of Environmental Air Samples by Solid-Phase Microextraction and Gas Chromatography/Ion Trap Mass Spectrometry. Environ. Sci. Technol. 29, 693–701. https://doi.org/10.1021/es00003a017

Chin, S.-T., Eyres, G.T., Marriott, P.J., 2015. Application of integrated comprehensive/multidimensional gas chromatography with mass spectrometry and olfactometry for aroma analysis in wine and coffee. Food Chem. 185, 355–361. https://doi.org/10.1016/j.foodchem.2015.04.003

Christopher G. Herbert, Robert A.W. Johnstone, 2003. Mass Spectrometry Basics. CRC Press.

Chung, N., Jo, Y., Joe, M.-H., Jeong, M.-H., Jeong, Y.-J., Kwon, J.-H., 2017. Rice vinegars of different origins: discriminative characteristics based on solid-phase microextraction and gas chromatography with mass spectrometry, an electronic nose, electronic tongue and sensory evaluation. J. Inst. Brew. 123, 159–166. https://doi.org/10.1002/jib.406

Couprie, C., Duval, L., Moreaud, M., Hénon, S., Tebib, M., Souchon, V., 2017. BARCHAN: Blob Alignment for Robust CHromatographic ANalysis. J. Chromatogr. A 1484, 65–72. https://doi.org/10.1016/j.chroma.2017.01.003

Coutinho, D.M., França, D., Vanini, G., Mendes, L.A.N., Gomes, A.O., Pereira, V.B., Ávila, B.M.F., Azevedo, D.A., 2018. Rapid hydrocarbon group-type semi-quantification in crude oils by comprehensive two-dimensional gas chromatography. Fuel 220, 379–388. https://doi.org/10.1016/j.fuel.2018.02.009

Davis, J.M., Stoll, D.R., Carr, P.W., 2008. Effect of First-Dimension Undersampling on Effective Peak Capacity in Comprehensive Two-Dimensional Separations. Anal. Chem. 80, 461–473. https://doi.org/10.1021/ac071504j

de Juan, A., Tauler, R., 2001. Comparison of three-way resolution methods for non-trilinear chemical data sets. J. Chemom. 15, 749–771. https://doi.org/10.1002/cem.662

de la Mata, A.P., McQueen, R.H., Nam, S.L., Harynuk, J.J., 2017. Comprehensive two-dimensional gas chromatographic profiling and chemometric interpretation of the volatile

profiles of sweat in knit fabrics. Anal. Bioanal. Chem. 409, 1905–1913. https://doi.org/10.1007/s00216-016-0137-1

de Lima, P.F., Furlan, M.F., de Lima Ribeiro, F.A., Pascholati, S.F., Augusto, F., 2015. In vivo determination of the volatile metabolites of saprotroph fungi by comprehensive two-dimensional gas chromatography: Gas Chromatography. J. Sep. Sci. 38, 1924–1932. https://doi.org/10.1002/jssc.201401404

Dekeirsschieter, J., Stefanuto, P.-H., Brasseur, C., Haubruge, E., Focant, J.-F., 2012. Enhanced Characterization of the Smell of Death by Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry (GCxGC-TOFMS). PLOS ONE 7, e39005. https://doi.org/10.1371/journal.pone.0039005

Deng, B., Kim, S., Li, H., Heath, E., Zhang, X., 2016. Global peak alignment for comprehensive two-dimensional gas chromatography mass spectrometry using point matching algorithms. J. Bioinform. Comput. Biol. 14, 1650032. https://doi.org/10.1142/S0219720016500323

Derrick, B., White, P., 2016. Why Welch's test is Type I error robust. Quant. Methods Psychol. 12, 30–38. https://doi.org/10.20982/tqmp.12.1.p030

Doebbe, A., Keck, M., Russa, M.L., Mussgnug, J.H., Hankamer, B., Tekçe, E., Niehaus, K., Kruse, O., 2010. The Interplay of Proton, Electron, and Metabolite Supply for Photosynthetic H2 Production in Chlamydomonas reinhardtii. J. Biol. Chem. 285, 30247–30260. https://doi.org/10.1074/jbc.M110.122812

Dubois, L.M., Perrault, K.A., Stefanuto, P.-H., Koschinski, S., Edwards, M., McGregor, L., Focant, J.-F., 2017a. Thermal desorption comprehensive two-dimensional gas chromatography coupled to variable-energy electron ionization time-of-flight mass spectrometry for monitoring subtle changes in volatile organic compound profiles of human blood. J. Chromatogr. A 1501, 117–127. https://doi.org/10.1016/j.chroma.2017.04.026

Eftekhari, A., Parastar, H., 2016. Multivariate analytical figures of merit as a metric for evaluation of quantitative measurements using comprehensive two-dimensional gas chromatography–mass spectrometry. J. Chromatogr. A 1466, 155–165. https://doi.org/10.1016/j.chroma.2016.09.016

Egert, B., Weinert, C.H., Kulling, S.E., 2015. A peaklet-based generic strategy for the untargeted analysis of comprehensive two-dimensional gas chromatography mass spectrometry data sets. J. Chromatogr. A 1405, 168–177. https://doi.org/10.1016/j.chroma.2015.05.056

E. Mohler, R., M. Dombek, K., C. Hoggard, J., M. Pierce, K., T. Young, E., E. Synovec, R., 2007. Comprehensive analysis of yeast metabolite GC×GC–TOFMS data: combining discovery-mode and deconvolution chemometric software. Analyst 132, 756–767. https://doi.org/10.1039/B700061H

Ettre, L.S., 1971. The Development of Chromatography. Anal. Chem. 43, 20A-31A. https://doi.org/10.1021/ac60308a718

Fisher, R.A., 1992. Statistical Methods for Research Workers, in: Kotz, S., Johnson, N.L. (Eds.), Breakthroughs in Statistics: Methodology and Distribution, Springer Series in Statistics. Springer, New York, NY, pp. 66–70. https://doi.org/10.1007/978-1-4612-4380-9_6

Focant, J.-F., Sjödin, A., Patterson, D.G., 2004. Improved separation of the 209 polychlorinated biphenyl congeners using comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry. J. Chromatogr. A 1040, 227–238. https://doi.org/10.1016/j.chroma.2004.04.003

Freye, C.E., Moore, N.R., Synovec, R.E., 2018. Enhancing the chemical selectivity in discovery-based analysis with tandem ionization time-of-flight mass spectrometry detection for comprehensive two-dimensional gas chromatography. J. Chromatogr. A 1537, 99–108. https://doi.org/10.1016/j.chroma.2018.01.008

Fulton G. Kitson, Barbara S. Larsen, Charles N. McEwen, 1996. Gas Chromatography and Mass Spectrometry: A Practical Guide. Academic Press.

Furbo, S., Hansen, A.B., Skov, T., Christensen, J.H., 2014. Pixel-Based Analysis of Comprehensive Two-Dimensional Gas Chromatograms (Color Plots) of Petroleum: A Tutorial. Anal. Chem. 86, 7160–7170. https://doi.org/10.1021/ac403650d

Gaines, R.B., Frysinger, G.S., Hendrick-Smith, M.S., Stuart, J.D., 1999a. Oil Spill Source Identification by Comprehensive Two-Dimensional Gas Chromatography. Environ. Sci. Technol. 33, 2106–2112. https://doi.org/10.1021/es9810484

Gargallo, R., Tauler, R., Cuesta-Sánchez, F., Massart, D.L., 1996. Validation of alternating least-squares multivariate curve resolution for chromatographic resolution and quantitation. TrAC - Trends Anal. Chem. 15, 279–286. https://doi.org/10.1016/0165-9936(96)00048-9

Gough, D.V., Bahaghighat, H.D., Synovec, R.E., 2018. Column Selection Approach to Achieve a High Peak Capacity in Comprehensive Three-Dimensional Gas Chromatography. Talanta. https://doi.org/10.1016/j.talanta.2018.12.007

Gröger, T., Zimmermann, R., 2011. Application of parallel computing to speed up chemometrics for GC × GC-TOFMS based metabolic fingerprinting. Talanta 83, 1289–1294. https://doi.org/10.1016/j.talanta.2010.09.015

Gros, J., Nabi, D., Dimitriou-Christidis, P., Rutler, R., Arey, J.S., 2012. Robust Algorithm for Aligning Two-Dimensional Chromatograms. Anal. Chem. 84, 9033–9040. https://doi.org/10.1021/ac301367s

Gross, G.M., Prazen, B.J., Grate, J.W., Synovec, R.E., 2004. High-Speed Gas Chromatography Using Synchronized Dual-Valve Injection. Anal Chem 76, 3517–3524. https://doi.org/10.1021/ac049909g

Gruber, B., Weggler, B.A., Jaramillo, R., Murrell, K.A., Piotrowski, P.K., Dorman, F.L., 2018. Comprehensive two-dimensional gas chromatography in forensic science: A critical review of recent trends. TrAC - Trends Anal. Chem. https://doi.org/10.1016/j.trac.2018.05.017

Hans-Joachim Hubsschmann, 2015. Handbook of GC-MS: Fundamentals and Applications. Wiley.

Hantao, L.W., Najafi, A., Zhang, C., Augusto, F., Anderson, J.L., 2014. Tuning the Selectivity of Ionic Liquid Stationary Phases for Enhanced Separation of Nonpolar Analytes in Kerosene Using Multidimensional Gas Chromatography. Anal. Chem. 86, 3717–3721. https://doi.org/10.1021/ac5004129

Hantao, L.W., Toledo, B.R., de Lima Ribeiro, F.A., Pizetta, M., Pierozzi, C.G., Furtado, E.L., Augusto, F., 2013a. Comprehensive two-dimensional gas chromatography combined to multivariate data analysis for detection of disease-resistant clones of Eucalyptus. Talanta 116, 1079–1084. https://doi.org/10.1016/j.talanta.2013.08.033

Hao, C., Zhao, X., Yang, P., 2007. GC-MS and HPLC-MS analysis of bioactive pharmaceuticals and personal-care products in environmental matrices. TrAC Trends Anal. Chem., Pharmaceutical-residue analysis 26, 569–580. https://doi.org/10.1016/j.trac.2007.02.011

Henry, M.C., Yonker, C.R., 2006. Supercritical Fluid Chromatography, Pressurized Liquid Extraction, and Supercritical Fluid Extraction. Anal. Chem. 78, 3909–3916. https://doi.org/10.1021/ac0605703

Higgins Keppler, E.A., Jenkins, C.L., Davis, T.J., Bean, H.D., 2018a. Advances in the application of comprehensive two-dimensional gas chromatography in metabolomics. TrAC - Trends Anal. Chem. https://doi.org/10.1016/j.trac.2018.10.015

Hoffmann, N., Wilhelm, M., Doebbe, A., Niehaus, K., Stoye, J., 2014. BiPACE 2D--graph-based multiple alignment for comprehensive 2D gas chromatography-mass spectrometry. Bioinforma. Oxf. Engl. 30, 988–995. https://doi.org/10.1093/bioinformatics/btt738

Homer, N., Kothiya, S., Rutter, A., Walker, B.R., Andrew, R., 2017. Gas chromatography tandem mass spectrometry offers advantages for urinary steroids analysis. Anal. Biochem. 538, 34–37. https://doi.org/10.1016/j.ab.2017.09.002

Hope, J.L., Sinha, A.E., Prazen, B.J., Synovec, R.E., 2005. 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. J. Chromatogr. A 1086, 185–192.

Hurtado, C., Parastar, H., Matamoros, V., Piña, B., Tauler, R., Bayona, J.M., 2017. Linking the morphological and metabolomic response of Lactuca sativa L exposed to emerging contaminants using GC × GC-MS and chemometric tools. Sci. Rep. 7. https://doi.org/10.1038/s41598-017-06773-0

IUPAC - chromatography (C01075) [WWW Document], n.d. https://doi.org/10.1351/goldbook.C01075

Izadmanesh, Y., Garreta-Lara, E., Ghasemi, J.B., Lacorte, S., Matamoros, V., Tauler, R., 2017a. Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data. J. Chromatogr. A 1488, 113–125. https://doi.org/10.1016/j.chroma.2017.01.052

Jacobs, M.R., Edwards, M., Górecki, T., Nesterenko, P.N., Shellie, R.A., 2016a. Evaluation of a miniaturised single-stage thermal modulator for comprehensive two-dimensional gas chromatography of petroleum contaminated soils. J. Chromatogr. A 1463, 162–168. https://doi.org/10.1016/j.chroma.2016.08.009

J.C. Giddings, 1991. Unified Separation Science. John Wiley & Sons, Inc.

Johnsen, L.G., Amigo, J.M., Skov, T., Bro, R., 2014. Automated resolution of overlapping peaks in chromatographic data: Chromatographic data analysis. J. Chemom. 28, 71–82. https://doi.org/10.1002/cem.2575

Johnson, K.J., Synovec, R.E., 2002. Pattern recognition of jet fuels: Comprehensive GC × GC with ANOVA-based feature selection and principal component analysis, in: Chemometrics and Intelligent Laboratory Systems. pp. 225–237. https://doi.org/10.1016/S0169-7439(01)00198-8

Jutley, G.S., Young, S.P., 2015. Metabolomics to identify biomarkers and as a predictive tool in inflammatory diseases. Best Pract. Res. Clin. Rheumatol., Musculoskeletal Science 29, 770–782. https://doi.org/10.1016/j.berh.2016.02.010

K. Robards, P.R. Haddad, P.E. Jackson, 2004. Principles and Practice of Modern Chromatographic Methods, 1st ed. Elsevier.

Khummueng, W., Harynuk, J., Marriott, P.J., 2006. Modulation Ratio in Comprehensive Two-dimensional Gas Chromatography. Anal. Chem. 78, 4578–4587. https://doi.org/10.1021/ac052270b

Klee, M.S., Cochran, J., Merrick, M., Blumberg, L.M., 2015a. Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain. J. Chromatogr. A 1383, 151–159. https://doi.org/10.1016/j.chroma.2015.01.031

Leghissa, A., Hildenbrand, Z.L., Foss, F.W., Schug, K.A., 2018. Determination of the metabolites of Δ9-tetrahydrocannabinol in urine and plasma using multiple reaction monitoring gas chromatography with triple quadrupole mass spectrometry. Sep. Sci. PLUS 1, 43–47. https://doi.org/10.1002/sscp.201700006

Lelevic, A., Souchon, V., Moreaud, M., Lorentz, C., Geantet, C., 2020. Gas chromatography vacuum ultraviolet spectroscopy: A review. J. Sep. Sci. 43, 150–173. https://doi.org/10.1002/jssc.201900770

Liquid Chromatography - 2nd Edition [WWW Document], n.d. URL https://www.elsevier.com/books/liquid-chromatography/fanali/978-0-12-805393-5 (accessed 6.17.20).

Liu, X., Zhao, A., Zhang, A., Liu, H., Xiao, W., Wang, C., Wang, X., 2011. Dispersive liquid–liquid microextraction and gas chromatography–mass spectrometry determination of polychlorinated biphenyls and polybrominated diphenyl ethers in milk. J. Sep. Sci. 34, 1084–1090. https://doi.org/10.1002/jssc.201000767

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

Machado, M.E., 2019. Comprehensive two-dimensional gas chromatography for the analysis of nitrogen-containing compounds in fossil fuels: A review. Talanta 198, 263–276. https://doi.org/10.1016/j.talanta.2019.02.031

Madsen, R., Lundstedt, T., Trygg, J., 2010. Chemometrics in metabolomics—A review in human disease diagnosis. Anal. Chim. Acta 659, 23–33. https://doi.org/10.1016/j.aca.2009.11.042

Magagna, F., Guglielmetti, A., Liberto, E., Reichenbach, S.E., Allegrucci, E., Gobino, G., Bicchi, C., Cordero, C., 2017a. Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination. J. Agric. Food Chem. 65, 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167

Magagna, F., Liberto, E., Reichenbach, S.E., Tao, Q., Carretta, A., Cobelli, L., Giardina, M., Bicchi, C., Cordero, C., 2018a. Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from thermal to differential-flow modulated comprehensive two dimensional gas chromatography. J. Chromatogr. A, Comprehensive two-dimensional chromatography 1536, 122–136. https://doi.org/10.1016/j.chroma.2017.07.014

Mao, S., Lu, C., Li, M., Ye, Y., Wei, X., Tong, H., 2018. Identification of key aromatic compounds in Congou black tea by partial least-square regression with variable importance of projection scores and gas chromatography–mass spectrometry/gas chromatography–olfactometry. J. Sci. Food Agric. 98, 5278–5286. https://doi.org/10.1002/jsfa.9066

Marney, L.C., Christopher Siegler, W., Parsons, B.A., Hoggard, J.C., Wright, B.W., Synovec, R.E., 2013a. Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data. Talanta 115, 887–895. https://doi.org/10.1016/j.talanta.2013.06.038

Marriott, P.J., Chin, S.T., Maikhunthod, B., Schmarr, H.G., Bieri, S., 2012. Multidimensional gas chromatography. TrAC - Trends Anal. Chem. https://doi.org/10.1016/j.trac.2011.10.013

Megson, D., Kalin, R., Worsfold, P.J., Gauchotte-Lindsay, C., Patterson, D.G., Lohan, M.C., Comber, S., Brown, T.A., O'Sullivan, G., 2013a. Fingerprinting polychlorinated biphenyls in environmental samples using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry. J. Chromatogr. A 1318, 276–283. https://doi.org/10.1016/j.chroma.2013.10.016

Mellors, T.R., Blanchet, L., Flynn, J.L., Tomko, J., O'Malley, M., Scanga, C.A., Lin, P.L., Hill, J.E., 2017. A new method to evaluate macaque health using exhaled breath: A case study of *M. tuberculosis* in a BSL-3 setting. J. Appl. Physiol. 122, 695–701. https://doi.org/10.1152/japplphysiol.00888.2016

Mills, E., O'Neill, L.A.J., 2014. Succinate: a metabolic signal in inflammation. Trends Cell Biol. 24, 313–320. https://doi.org/10.1016/j.tcb.2013.11.008

Mogollon, N.G.S., Ribeiro, F.A. de L., Lopez, M.M., Hantao, L.W., Poppi, R.J., Augusto, F., 2013b. Quantitative analysis of biodiesel in blends of biodiesel and conventional diesel by comprehensive two-dimensional gas chromatography and multivariate curve resolution. Anal. Chim. Acta 796, 130–136. https://doi.org/10.1016/j.aca.2013.07.071

Mogollón, N.G.S., Ribeiro, F.A. de L., Poppi, R.J., Quintana, A., Chávez, J., Agualongo, D., Aleme, H., Augusto, F., 2016. Exploratory Analysis of Biodiesel by Combining Comprehensive Two-Dimensional Gas Chromatography and Multiway Principal Component Analysis. J. Braz. Chem. Soc. https://doi.org/10.21577/0103-5053.20160222

Mohler, R.E., Dombek, K.M., Hoggard, J.C., Pierce, K.M., Young, E.T., Synovec, R.E., 2007. Comprehensive analysis of yeast metabolite GC×GC–TOFMS data: combining discovery-mode and deconvolution chemometric software. Analyst 132, 756–767. https://doi.org/10.1039/B700061H

Mohler, R.E., Dombek, K.M., Hoggard, J.C., Young, E.T., Synovec, R.E., 2006. Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells. Anal. Chem. 78, 2700–2709. https://doi.org/10.1021/ac052106o

Mohler, R.E., Tu, B.P., Dombek, K.M., Hoggard, J.C., Young, E.T., Synovec, R.E., 2008a. Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography-time-of-flight-mass spectrometry data. J. Chromatogr. A 1186, 401–411. https://doi.org/10.1016/j.chroma.2007.10.063

Muscalu, A.M., Górecki, T., 2018. Comprehensive two-dimensional gas chromatography in environmental analysis. TrAC - Trends Anal. Chem. https://doi.org/10.1016/j.trac.2018.07.001

Nadeau, J.S., Wright, B.W., Synovec, R.E., 2010. Chemometric analysis of gas chromatography-mass spectrometry data using fast retention time alignment via a total ion current shift function. Talanta 81, 120–128. https://doi.org/10.1016/j.talanta.2009.11.046

Nicholas H. Snow, 2020. Basic Multidimensional Gas Chromatography, Separation Science and Technology. Elsevier.

Nisbet, L.A., Wylie, F.M., Logan, B.K., Scott, K.S., 2019. Gas Chromatography-Mass Spectrometry Method for the Quantitative Identification of 23 New Psychoactive Substances in Blood and Urine. J. Anal. Toxicol. 43, 346–352. https://doi.org/10.1093/jat/bky109

Nizio, K.D., Ueland, M., Stuart, B.H., Forbes, S.L., 2017. The analysis of textiles associated with decomposing remains as a natural training aid for cadaver-detection dogs. Forensic Chem. 5, 33–45. https://doi.org/10.1016/j.forc.2017.06.002

Ochiai, N., Ieda, T., Sasamoto, K., Takazawa, Y., Hashimoto, S., Fushimi, A., Tanabe, K., 2011. Stir bar sorptive extraction and comprehensive two-dimensional gas chromatography coupled to high-resolution time-of-flight mass spectrometry for ultra-trace analysis of organochlorine pesticides in river water. J. Chromatogr. A 1218, 6851–6860. https://doi.org/10.1016/j.chroma.2011.08.027

Oliveira, L.F., Braga, S.C.G.N., Augusto, F., Hashimoto, J.C., Efraim, P., Poppi, R.J., 2016a. Differentiation of cocoa nibs from distinct origins using comprehensive two-dimensional gas chromatography and multivariate analysis. Food Res. Int. 90, 133–138. https://doi.org/10.1016/j.foodres.2016.10.047

Olivieri, A.C., Wu, H.-L., Yu, R.-Q., 2012. MVC3: A MATLAB graphical interface toolbox for third-order multivariate calibration. Chemom. Intell. Lab. Syst. 116, 9–16. https://doi.org/10.1016/j.chemolab.2012.03.018

Omar, J., Olivares, M., Amigo, J.M., Etxebarria, N., 2014. Resolution of co-eluting compounds of Cannabis Sativa in comprehensive two-dimensional gas chromatography/mass spectrometry detection with Multivariate Curve Resolution-Alternating Least Squares. Talanta 121, 273–280. https://doi.org/10.1016/j.talanta.2013.12.044

Ong, R., Marriott, P., Morrison, P., Haglund, P., 2002. Influence of chromatographic conditions on separation in comprehensive gas chromatography. J. Chromatogr. A 962, 135–152. https://doi.org/10.1016/S0021-9673(02)00458-2

Organtini, K.L., Myers, A.L., Jobst, K.J., Cochran, J., Ross, B., McCarry, B., Reiner, E.J., Dorman, F.L., 2014a. Comprehensive characterization of the halogenated dibenzo-p-dioxin and dibenzofuran contents of residential fire debris using comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry. J. Chromatogr. A 1369, 138–146. https://doi.org/10.1016/j.chroma.2014.09.088

Pani, O., Górecki, T., 2006. Comprehensive two-dimensional gas chromatography (GC×GC) in environmental analysis and monitoring. Anal. Bioanal. Chem. 386, 1013–1023. https://doi.org/10.1007/s00216-006-0568-1

Parastar, H., Jalali-Heravi, M., Tauler, R., 2012. Comprehensive two-dimensional gas chromatography (GC×GC) retention time shift correction and modeling using bilinear peak alignment, correlation optimized shifting and multivariate curve resolution. Chemom. Intell. Lab. Syst., Special Issue Section: Selected Papers from the 1st African-European Conference on Chemometrics, Rabat, Morocco, September 2010 Special Issue Section: Preprocessing methods Special Issue Section: Spectroscopic imaging 117, 80–91. https://doi.org/10.1016/j.chemolab.2012.02.003

Parastar, H., Radović, J.R., Jalali-Heravi, M., Diez, S., Bayona, J.M., Tauler, R., 2011. 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, 9289–9297. https://doi.org/10.1021/ac201799r

Parastar, H., Tauler, R., 2014. Multivariate Curve Resolution of Hyphenated and Multidimensional Chromatographic Measurements: A New Insight to Address Current Chromatographic Challenges. Anal. Chem. 86, 286–297. https://doi.org/10.1021/ac402377d

Parsons, B.A., Marney, L.C., Siegler, W.C., Hoggard, J.C., Wright, B.W., Synovec, R.E., 2015a. 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, 3812–3819. https://doi.org/10.1021/ac504472ss

Parsons, B.A., Pinkerton, D.K., Synovec, R.E., 2016a. Implications of Phase Ratio for Maximixing Peak Capacity in Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry. J. Chromatogr. A.

Parsons, B.A., Pinkerton, D.K., Wright, B.W., Synovec, R.E., 2016b. 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, 179–190. https://doi.org/10.1016/j.chroma.2016.02.067

Pesesse, R., Stefanuto, P.-H., Schleich, F., Louis, R., Focant, J.-F., 2019a. Multimodal chemometric approach for the analysis of human exhaled breath in lung cancer patients by TD-GC × GC-TOFMS. J. Chromatogr. B 1114–1115, 146–153. https://doi.org/10.1016/j.jchromb.2019.01.029

Phillips, J.B., Beens, J., 1999a. Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions. J. Chromatogr. A 856, 331–347. https://doi.org/10.1016/S0021-9673(99)00815-8

Pierce, K.M., Hoggard, J.C., 2014. Chromatographic data analysis. Part 3.3.4: Handling hyphenated data in chromatography. Anal. Methods. https://doi.org/10.1039/c3ay40965a

Pierce, K.M., Hoggard, J.C., Hope, J.L., Rainey, P.M., Hoofnagle, A.N., Jack, R.M., Wright, B.W., Synovec, R.E., 2006a. Fisher Ratio Method Applied to Third-Order Separation Data To Identify Significant Chemical Components of Metabolite Extracts. Anal. Chem. 78, 5068–5075. https://doi.org/10.1021/ac0602625

Pierce, K.M., Kehimkar, B., Marney, L.C., Hoggard, J.C., Synovec, R.E., 2012a. Review of chemometric analysis techniques for comprehensive two dimensional separations data. J. Chromatogr. A, Hyphenated and Multidimensional Chromatography Techniques 1255, 3–11. https://doi.org/10.1016/j.chroma.2012.05.050

Pierce, K.M., Mohler, R.E., 2012. A review of chemometrics applied to comprehensive two-dimensional separations from 2008-2010. Sep. Purif. Rev. https://doi.org/10.1080/15422119.2011.591868

Pierce, K.M., Wood, L.F., Wright, B.W., Synovec, R.E., 2005. A comprehensive two-dimensional retention time alignment algorithm to enhance chemometric analysis of comprehensive two-dimensional separation data. Anal. Chem. 77, 7735–7743. https://doi.org/10.1021/ac0511142

Pinkerton, D.K., Parsons, B.A., Anderson, T.J., Synovec, R.E., 2015a. 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, 66–76. https://doi.org/10.1016/j.aca.2015.02.040

Pinkerton, D.K., Parsons, B.A., Synovec, R.E., 2016. Method to determine the true modulation ratio for comprehensive two-dimensional gas chromatography. J. Chromatogr. A 1476, 114–123. https://doi.org/10.1016/j.chroma.2016.11.015

Pollo, B.J., Alexandrino, G.L., Augusto, F., Hantao, L.W., 2018. The impact of comprehensive two-dimensional gas chromatography on oil & gas analysis: Recent advances and

applications in petroleum industry. TrAC Trends Anal. Chem. 105, 202–217. https://doi.org/10.1016/j.trac.2018.05.007

Poole, C., 2003. The Essence of Chromatography, 1st ed. Elsevier.

Prata, P.S., Alexandrino, G.L., Mogollón, N.G.S., Augusto, F., 2016. Discriminating Brazilian crude oils using comprehensive two-dimensional gas chromatography–mass spectrometry and multiway principal component analysis. J. Chromatogr. A 1472, 99–106. https://doi.org/10.1016/j.chroma.2016.10.044

Prebihalo, S., Brockman, A., Cochran, J., Dorman, F.L., 2015. Determination of emerging contaminants in wastewater utilizing comprehensive two-dimensional gas-chromatography coupled with time-of-flight mass spectrometry. J. Chromatogr. A 1419, 109–115. https://doi.org/10.1016/j.chroma.2015.09.080

Prebihalo, S.E., Berrier, K.L., Freye, C.E., Bahaghighat, H.D., Moore, N.R., Pinkerton, D.K., Synovec, R.E., 2018a. Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications. Anal. Chem. 90, 505–532. https://doi.org/10.1021/acs.analchem.7b04226

Purcaro, G., Barp, L., Beccaria, M., Conte, L.S., 2016. Characterisation of minor components in vegetable oil by comprehensive gas chromatography with dual detection. Food Chem. 212, 730–738. https://doi.org/10.1016/j.foodchem.2016.06.048

Qian, K., Wang, F.C., 2019. Compositional Analysis of Heavy Petroleum Distillates by Comprehensive Two-dimensional Gas Chromatography, Field Ionization and High-resolution Mass Spectrometry. J. Am. Soc. Mass Spectrom. 30, 2785–2794. https://doi.org/10.1021/jasms.8b06302

Reaser, B.C., Wright, B.W., Synovec, R.E., 2017a. Using Receiver Operating Characteristic Curves To Optimize Discovery-Based Software with Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry. Anal. Chem. 89, 3606–3612. https://doi.org/10.1021/acs.analchem.6b04991

Reichenbach, S.E., Rempe, D.W., Tao, Q., Bressanello, D., Liberto, E., Bicchi, C., Balducci, S., Cordero, C., 2015. Alignment for Comprehensive Two-Dimensional Gas Chromatography with Dual Secondary Columns and Detectors. Anal. Chem. 87, 10056–10063. https://doi.org/10.1021/acs.analchem.5b02718

Reichenbach, S.E., Tian, X., Cordero, C., Tao, Q., 2012. Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography. J. Chromatogr. A, Selected Papers from the 35th International Symposium on Capillary Chromatography, the 26th International Symposium on MicroScale Bioseparations and the 8th GC×GC Symposium, San Diego, CA, USA, 1-5 May 2011 1226, 140–148. https://doi.org/10.1016/j.chroma.2011.07.046

Rempe, D.W., Reichenbach, S.E., Tao, Q., Cordero, C., Rathbun, W.E., Zini, C.A., 2016. Effectiveness of Global, Low-Degree Polynomial Transformations for GCxGC Data Alignment. Anal. Chem. 88, 10028–10035. https://doi.org/10.1021/acs.analchem.6b02254

Rodrigues, A., Yegles, M., Van Elsué, N., Schneider, S., 2018. Determination of cannabinoids in hair of CBD rich extracts consumers using gas chromatography with tandem mass spectrometry (GC/MS–MS). Forensic Sci. Int. 292, 163–166. https://doi.org/10.1016/j.forsciint.2018.09.015

Rosso, M.C., Liberto, E., Spigolon, N., Fontana, M., Somenzi, M., Bicchi, C., Cordero, C., 2018. Evolution of potent odorants within the volatile metabolome of high-quality hazelnuts (Corylus avellana L.): evaluation by comprehensive two-dimensional gas chromatography

coupled with mass spectrometry. Anal. Bioanal. Chem. 410, 3491–3506. https://doi.org/10.1007/s00216-017-0832-6

Sampat, A.A.S., Lopatka, M., Vivó-Truyols, G., Schoenmakers, P.J., van Asten, A.C., 2016. Towards chemical profiling of ignitable liquids with comprehensive two-dimensional gas chromatography: Exploring forensic application to neat white spirits. Forensic Sci. Int. 267, 183–195. https://doi.org/10.1016/j.forsciint.2016.08.006

Schäffer, M., Gröger, T., Pütz, M., Dieckmann, S., Zimmermann, R., 2012a. Comparative Analysis of the Chemical Profiles of 3,4-Methylenedioxymethamphetamine Based on Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry (GC×GC-TOFMS). J. Forensic Sci. 57, 1181–1189. https://doi.org/10.1111/j.1556-4029.2012.02137.x.

Seeley, J.V., Seeley, S.K., 2013a. Multidimensional Gas Chromatography: Fundamental Advances and New Applications. Anal. Chem. 85, 557–578. https://doi.org/10.1021/ac303195u

Shreve, M.J., Brockman, A., Hartleb, M., Prebihalo, S., Dorman, F.L., Brennan, R.A., 2016. The white-rot fungus Trametes versicolor reduces the estrogenic activity of a mixture of emerging contaminants in wastewater treatment plant effluent. Int. Biodeterior. Biodegrad. 109, 132–140. https://doi.org/10.1016/j.ibiod.2016.01.018

Sinha, A.E., Hope, J.L., Prazen, B.J., Fraga, C.G., Nilsson, E.J., Synovec, R.E., 2004. Multivariate selectivity as a metric for evaluating comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry subjected to chemometric peak deconvolution. J. Chromatogr. A 1056, 145–154. https://doi.org/10.1016/j.chroma.2004.06.110

Skoczyńska, E., Korytár, P., Boer, J. de, 2008. Maximizing Chromatographic Information from Environmental Extracts by GCxGC-ToF-MS. Environ. Sci. Technol. 42, 6611–6618. https://doi.org/10.1021/es703229t

Skov, T., Hoggard, J.C., Bro, R., Synovec, R.E., 2009. Handling within run retention time shifts in two-dimensional chromatography data using shift correction and modeling. J. Chromatogr. A 1216, 4020–4029. https://doi.org/10.1016/j.chroma.2009.02.049

Stadler, S., Stefanuto, P.H., Brokl, M., Forbes, S.L., Focant, J.F., 2013a. Characterization of volatile organic compounds from human analogue decomposition using thermal desorption coupled to comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. Anal. Chem. 85, 998–1005. https://doi.org/10.1021/ac302614y

Stefanuto, P.-H., A. Perrault, K., M. Lloyd, R., Stuart, B., Rai, T., L. Forbes, S., Focant, J.-F., 2015. Exploring new dimensions in cadaveric decomposition odour analysis. Anal. Methods 7, 2287–2294. https://doi.org/10.1039/C5AY00371G

Stefanuto, P.-H., Perrault, K., Stadler, S., Pesesse, R., Brokl, M., Forbes, S., Focant, J.-F., 2014. Reading Cadaveric Decomposition Chemistry with a New Pair of Glasses. ChemPlusChem 79, 786–789. https://doi.org/10.1002/cplu.201402003

Stefanuto, P.H., Perrault, K.A., Dubois, L.M., L'Homme, B., Allen, C., Loughnane, C., Ochiai, N., Focant, J.F., 2017. Advanced method optimization for volatile aroma profiling of beer using two-dimensional gas chromatography time-of-flight mass spectrometry. J. Chromatogr. A 1507, 45–52. https://doi.org/10.1016/j.chroma.2017.05.064

Stefanuto, P.H., Perrault, K.A., Lloyd, R.M., Stuart, B., Rai, T., Forbes, S.L., Focant, J.F., 2015. Exploring new dimensions in cadaveric decomposition odour analysis. Anal. Methods 7, 2287–2294. https://doi.org/10.1039/c5ay00371g

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

Tian, T.-F., Wang, S.-Y., Kuo, T.-C., Tan, C.-E., Chen, G.-Y., Kuo, C.-H., Chen, C.-H.S., Chan, C.-C., Lin, O.A., Tseng, Y.J., 2016. Web Server for Peak Detection, Baseline Correction, and Alignment in Two-Dimensional Gas Chromatography Mass Spectrometry-Based Metabolomics Data. Anal. Chem. 88, 10395–10403. https://doi.org/10.1021/acs.analchem.6b00755

Tu, B.P., Mohler, R.E., Liu, J.C., Dombek, K.M., Young, E.T., Synovec, R.E., McKnight, S.L., 2007. Cyclic changes in metabolic state during the life of a yeast cell. Proc. Natl. Acad. Sci. U. S. A. 104, 16886–16891. https://doi.org/10.1073/pnas.0708365104

Uhl, M., 1997. Determination of drugs in hair using GC/MS/MS. Forensic Sci. Int., Proceedings of the 1st European Meeting on Hair Analysis. Clinical, Occupational and Forensic Application 84, 281–294. https://doi.org/10.1016/S0379-0738(96)02072-5

Watson, N.E., Parsons, B.A., Synovec, R.E., 2016a. Performance evaluation of tile-based Fisher Ratio analysis using a benchmark yeast metabolome dataset. J. Chromatogr. A 1459, 101–111. https://doi.org/10.1016/j.chroma.2016.06.067

Welch, B.L., 1947. The Generalization of `Student's' Problem when Several Different Population Variances are Involved. Biometrika 34, 28–35. https://doi.org/10.2307/2332510

Welke, J.E., Manfroi, V., Zanus, M., Lazzarotto, M., Alcaraz Zini, C., 2013a. Differentiation of wines according to grape variety using multivariate analysis of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection data. Food Chem. 141, 3897–3905. https://doi.org/10.1016/j.foodchem.2013.06.100

Zeng, Z., Li, J., Hugel, H.M., Xu, G., Marriott, P.J., 2014. Interpretation of comprehensive two-dimensional gas chromatography data using advanced chemometrics. TrAC - Trends Anal. Chem. https://doi.org/10.1016/j.trac.2013.08.009

Zoccali, M., Cappello, S., Mondello, L., 2018. Multilevel characterization of marine microbial biodegradation potentiality by means of flow-modulated comprehensive two-dimensional gas chromatography combined with a triple quadrupole mass spectrometer. J. Chromatogr. A 1547, 99–106. https://doi.org/10.1016/j.chroma.2018.03.013

Zushi, Y., Gros, J., Tao, Q., Reichenbach, S.E., Hashimoto, S., Arey, J.S., 2017. Pixel-by-pixel correction of retention time shifts in chromatograms from comprehensive two-dimensional gas chromatography coupled to high resolution time-of-flight mass spectrometry. J. Chromatogr. A 1508, 121–129. https://doi.org/10.1016/j.chroma.2017.05.065

APPENDIX A

Table A.1. Time-of-injury table including hits found via standard and control-normalized F-ratio that passed the t-test in the top 30 (to maintain consistency with standard F-ratio analysis, control-normalized F-ratio analysis was updated to be limited to the top 30).

ID	^{CN}Hit#	^{CN}F-ratio	^{S}Hit#	^{S}F-ratio	[P]/[C]	Var. #
Naproxen	1	6075	254	7	22.8	1
Phenylalanine*	3	473	12	42	2.44	2
Arachidonic acid	4	435	123	13	5.51	3
Palmitic acid*	5	403	20	32	3.16	4
Mannose	6	299	2	74	3.01	5
Stearic acid*	7	295	19	33	3.73	6
Glycine	8	285	5	51	1.28	7
Linoleic acid	9	280	16	34	4.38	8
Metabolite Unk 6	10	247	147	11	2.10	9
Glutamine	11	222	24	30	2.36	10
Glutamic acid*	12	203	18	33	2.87	11
L-lysine*	13	165	10	44	4.46	12
Metabolite Unk 1	14	162	4	54	3.44	13
Succinic acid*	15	137	37	20	3.31	14
CoA fragment	16	133	15	34	3.22	15
Metabolite Unk 9	21	63	109	13	3.77	16
Lactic acid*	23	60	21	32	2.42	17
Phospho-D-gluconate	24	56	31	23	1.37	18
Metabolite Unk 10	25	53	36	21	1.56	19
Malonic acid*	26	52	3	59	0.65	20
L-histidine	27	49	30	23	1.96	21
Metabolite Unk 11	28	48	130	12	0.73	22
L-methionine*	29	47	1	98	0.46	23
L-threonine	30	44	28	25	2.50	24
Glycerol*	36	40	9	45	0.60	25
Glycopyranose	66	23	13	37	0.62	26
Metabolite Unk 2	85	21	22	31	0.35	27
Metabolite Unk 4	168	10	26	28	3.19	28
Pyroglutamic acid*	206	8	8	46	2.44	29

Time-of-injury PCA scores and loadings plots for standard F-ratio analysis, control-normalized F-ratio analysis, and combined.

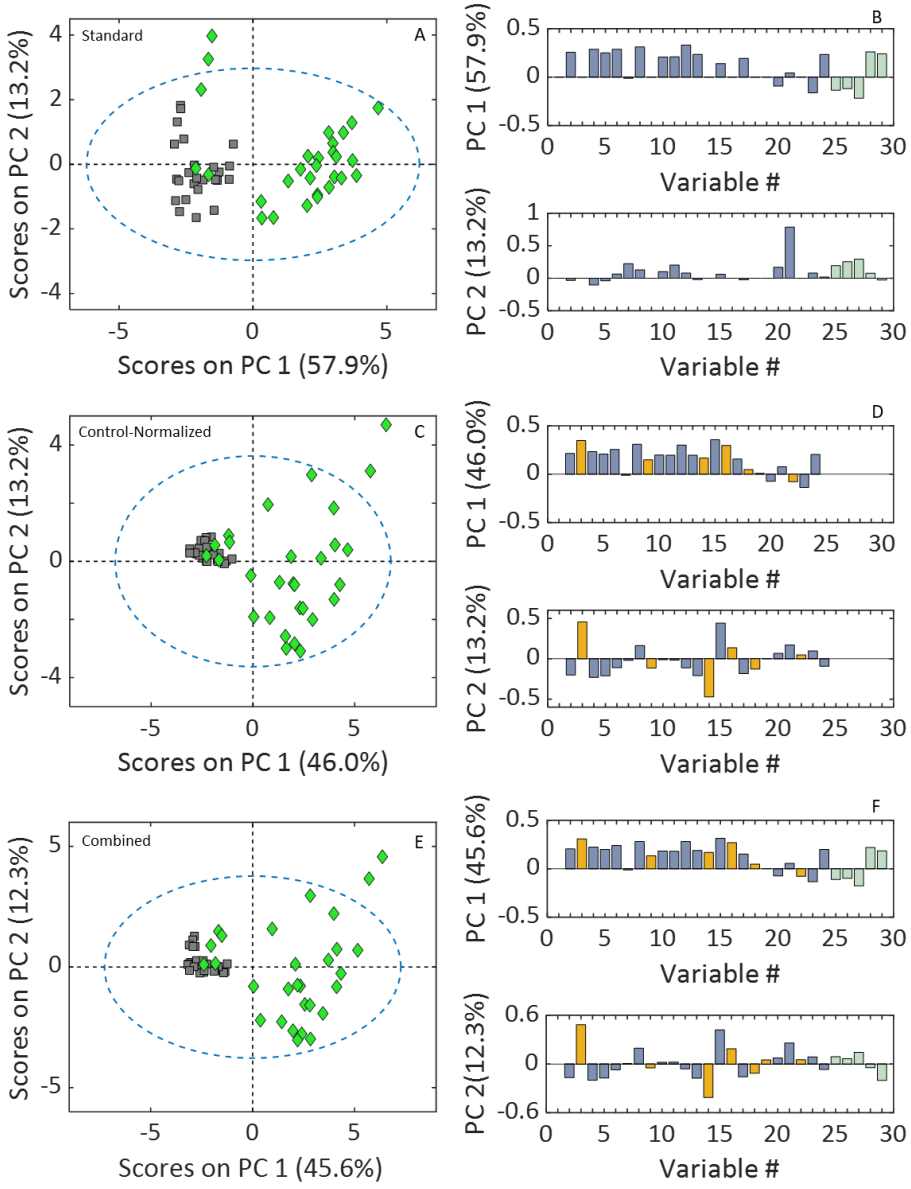

Figure A.1. Scores and loadings plots for sample and metabolite data summarized in Table A.1, where the 29 variables refer to the metabolites labeled as "Variable #" in the far-right table column (naproxen excluded). (A) PCA scores and (B) loadings plots for 21 metabolites that pass the t-test from the standard FRA hit list. The classes (controls, gray squares; patients, green diamonds) are well separated by PC 1, apart from 5 patients which appear like controls. (C) PCA scores and (D) loadings plots for 24 metabolites which pass the t-test from the control-normalized FRA hit list. Controls are tightly clustered indicating the metabolites have minor within-class variance. (E) PCA scores and (F) loadings plots from the combined standard and control-normalized FRA hit lists. The class separation appears functionally the same as the PCA scores plot for the control-normalized hit list in (C).

Metabolite profiles for 6 representative samples at time-of-injury

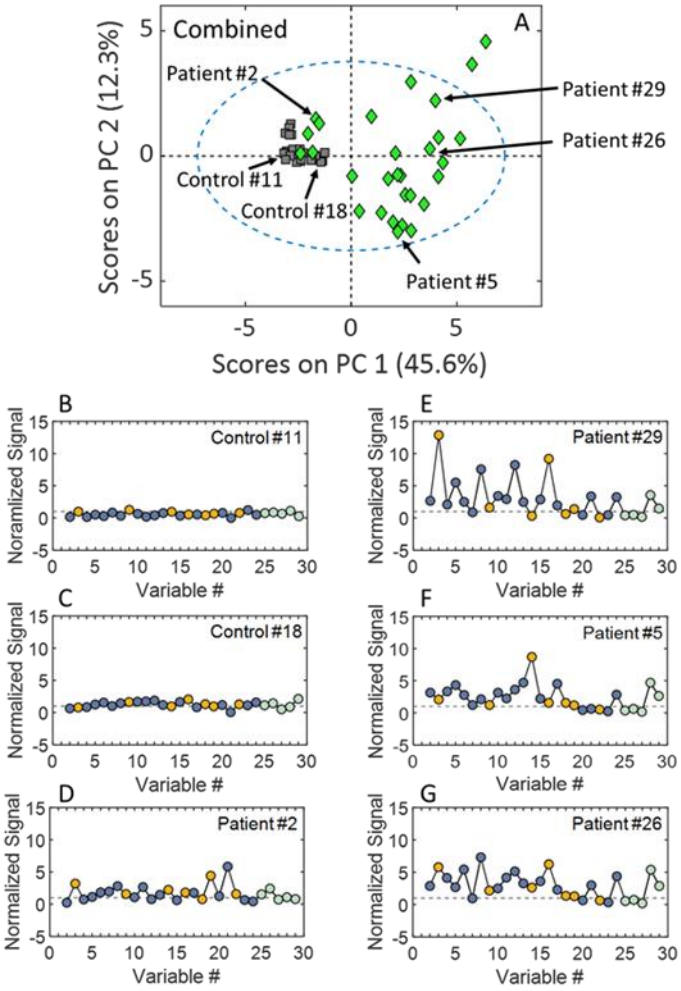

Figure A.2. Metabolite profiles for 6 representative samples and position in combined PCA scores plot (Figure A.1E). The normalized signal is calculated by taking the signal for each metabolite for the given sample and dividing by the mean for that metabolite within the control group. Thus, the metabolite signals for the control samples will be centered near 1. (A) PCA scores plot with sample # labeled for two control samples and 4 patient samples. (B) Metabolite profile for control #11, located furthest to the left on PC 1 and centered on PC 2. (C) Control #18, located just to the left of center for PC 1. (D) Patient #2, which is somewhat clustered with control samples and negatively loaded on PC 1 and positively loaded on PC 2. It appears variables #3, #19, and #21 are responsible for the separation from controls on PC 2. (E) Patient #29, heavily loaded on PC 1 and PC 2, with significant upregulation of variables #3, #5, #8, #16. (F) Patient #5, positively loaded on PC 1 and negatively loaded on PC 2, with significant upregulation of variables #14, #17, and #28. Interestingly, it appears as though upregulation of arachidonic acid (variable #3) drives the separation into positive PC 2 space, and upregulation of succinic acid (variable #14) is the driving force of separation into negative PC 2 space. (G) Patient #26, which is neutral on PC 2, has modest upregulation of both arachidonic acid (variable #3) and succinic acid (variable #14).

Table A.2. Hit list for top 30 hits using standard F-ratio analysis at 12 months post-surgery, including signal information.

12 months - Standard FRA									
ID	**^{S}Hit#**	**^{S}F-ratio**	**^{CN}Hit#**	**^{CN}F-ratio**	**m/z**	**Average Patient Signal**	**Average Control Signal**	**[P]/[C]**	**texpt**
L-methionine*	1	122	22	147	155	365 ± 216	626 ± 170	0.58	5.2
Metabolite Unk. 8	2	83	15	283	47	1520 ± 757	1943 ± 697	0.78	2.3
CoA fragment	3	73	23	142	77	5768 ± 6657	800 ± 312	7.21	4.1
Metabolite Unk. 12	4	72	17	229	73	1760 ± 1202	2489 ± 1344	0.71	2.2
Metabolite Unk. 13	5	71	48	42	73	82 ± 48	66 ± 29	1.24	1.6
Alanine*	6	70	30	93	116	2628 ± 2061	5402 ± 2419	0.49	4.8
Metabolite Unk. 14	7	69	44	46	73	8826 ± 5823	17764 ± 5266	0.5	6.2
Mannose	8	67	6	898	49	2414 ± 620	741 ± 212	3.26	14
Oxoisocaproic acid	9	67	39	49	75	353 ± 186	385 ± 148	0.91	0.8
Malonic acid*	10	66	46	44	55	167 ± 96	276 ± 81	0.61	4.7
Lactic acid*	11	62	9	568	49	5187 ± 2518	2589 ± 835	2	5.4
Stearic acid*	12	59	2	7008	77	556 ± 265	133 ± 67	4.19	8.5
Metabolite Unk. 7	13	56	14	317	144	1342 ± 695	1761 ± 773	0.75	2.4
Metabolite Unk. 15	14	52	161	22	73	292 ± 197	708 ± 350	0.41	5.7
Urea	15	50	43	46	47	3479 ± 1705	3234 ± 572	1.08	0.8
Metabolite Unk. 16	16	46	69	37	73	390 ± 237	702 ± 205	0.56	5.4
Oxalic acid*	17	46	32	91	73	10273 ± 7715	21428 ± 7146	0.48	5.8
Metabolite Unk. 17	18	46	59	39	73	68 ± 39	78 ± 20	0.88	1.2
Myo-inositol	19	41	12	383	73	1954 ± 763	1688 ± 554	1.16	1.6
Metabolite Unk. 18	20	39	26	137	73	1719 ± 848	2086 ± 837	0.82	1.7
Glutamine	21	39	89	31	47	116 ± 46	58 ± 25	2	6.1
Metabolite Unk. 4	22	38	19	158	77	163 ± 107	61 ± 28	2.67	5.1
Metabolite Unk. 1	23	37	3	1450	49	5118 ± 1926	1628 ± 526	3.14	9.6
Glycerol*	24	37	146	24	152	600 ± 312	1153 ± 344	0.52	6.5
Succinic acid*	25	36	4	1175	77	5018 ± 2700	1155 ± 398	4.34	7.8
Metabolite Unk. 2	26	35	199	19	73	612 ± 431	1560 ± 783	0.39	5.8
Glutamic acid*	27	35	10	559	77	1283 ± 683	712 ± 342	1.8	4.1
Glycine	28	34	97	30	73	109 ± 83	99 ± 51	0.94	0.7
Linoliec acid	29	34	11	456	77	998 ± 638	332 ± 410	3.01	4.8
Maleic acid*	30	32	16	239	73	3237 ± 1554	2544 ± 855	1.27	2.1

Table A.3. Hit list for top 30 hits using control-normalized F-ratio analysis at 12 months post-surgery, including signal information.

12 months - Control-normalized FRA									
ID	**^{CN}Hit#**	**^{CN}F-ratio**	**^{S}Hit#**	**^{S}F-ratio**	**m/z**	**Average Patient Signal**	**Average Control Signal**	**[P]/[C]**	**texpt**
Palmitic acid*	1	18290	185	20	49	1063 ± 494	305 ± 205	3.48	7.75
Stearic acid*	2	7008	12	59	77	556 ± 265	133 ± 67	4.19	8.47
Metabolite Unk. 1	3	1450	23	37	49	5118 ± 1926	1628 ± 526	3.14	9.57
Succinic acid*	4	1175	25	36	77	5018 ± 2700	1155 ± 398	4.34	7.75
6-phospho-D-gluconate	5	981	209	18	58	426 ± 403	281 ± 150	1.52	1.85
Mannose	6	898	8	67	49	2414 ± 620	741 ± 212	3.26	13.99
Picolinc acid	7	898	69	27	77	3409 ± 1938	1561 ± 457	2.18	5.08
Pyroglutamic acid*	8	854	88	25	77	1203 ± 664	415 ± 244	2.9	6.1
Lactic acid*	9	568	11	62	49	5187 ± 2518	2589 ± 835	2	5.36
Glutamic acid*	10	559	27	35	77	1283 ± 683	712 ± 342	1.8	4.09
Linoliec acid	11	456	29	34	77	998 ± 638	332 ± 410	3.01	4.81
Myo-inositol	12	383	19	41	73	1954 ± 763	1688 ± 554	1.16	1.55
Metabolite Unk. 19	13	374	439	7	73	6979 ± 5268	2910 ± 1374	2.4	4.09
Metabolite Unk. 7	14	317	13	56	144	1342 ± 695	1761 ± 773	0.75	2.36
Metabolite Unk. 8	15	283	2	83	47	1520 ± 757	1943 ± 697	0.78	2.25
Maleic acid*	16	239	30	32	73	3237 ± 1554	2544 ± 855	1.27	2.14
Metabolite Unk. 12	17	229	4	72	73	1760 ± 1202	2489 ± 1344	0.71	2.21
Metabolite Unk. 20	18	225	451	6	73	165 ± 312	21 ± 14	7.97	2.53
Metabolite Unk. 4	19	158	22	38	77	163 ± 107	61 ± 28	2.67	5.05
Metabolite Unk. 21	20	157	212	18	73	1560 ± 1001	1696 ± 988	0.92	0.53
Myristate*	21	150	82	26	77	286 ± 190	93 ± 57	3.08	5.35
L-methionine*	22	147	1	122	155	365 ± 216	626 ± 170	0.58	5.2
CoA fragment	23	142	3	73	77	5768 ± 6657	800 ± 312	7.21	4.08
Metabolite Unk. 22	24	140	161	21	73	2507 ± 1051	1803 ± 551	1.39	3.24
Glycopyranose	25	138	309	13	73	25781 ± 30604	24360 ± 17841	1.06	0.22
Metabolite Unk. 23	26	134	20	39	73	1719 ± 848	2086 ± 837	0.82	1.69
L-tryptophan	27	112	113	23	73	374 ± 306	197 ± 171	1.9	2.77
Metabolite Unk. 8	28	105	31	31	77	594 ± 404	181 ± 72	3.28	5.51
Metabolite Unk. 24	29	103	297	14	73	1297 ± 646	2114 ± 787	0.61	4.4
Alanine*	30	93	6	70	116	2628 ± 2061	5402 ± 2419	0.49	4.78

Table A.4. Combined hit list for top 30 hits using control-normalized and standard F-ratio analysis at 12 months post-surgery, including signal information. Far right column contains variable # as input to PCA.

12 months post-surgery - combined										
ID	**^{CN}Hit#**	**^{CN}F-ratio**	**^{S}Hit#**	**^{S}F-ratio**	**m/z**	**Average Patient Signal**	**Average Control Signal**	**[P]/[C]**	**texpt**	**Var. #**
Palmitic acid*	1	18290	185	20	49	1063 ± 494	305 ± 205	3.48	7.75	1
Stearic acid*	2	7008	12	59	77	556 ± 265	133 ± 67	4.19	8.47	2
Metabolite Unk. 1	3	1450	23	37	49	5118 ± 1926	1628 ± 526	3.14	9.57	3
Succinic acid*	4	1175	25	36	77	5018 ± 2700	1155 ± 398	4.34	7.75	4
Mannose	6	898	8	67	49	2414 ± 620	741 ± 212	3.26	13.99	5
Picolinc acid	7	898	69	27	77	3409 ± 1938	1561 ± 457	2.18	5.08	6
Pyroglutamic acid*	8	854	88	25	77	1203 ± 664	415 ± 244	2.9	6.1	7
Lactic acid*	9	568	11	62	49	5187 ± 2518	2589 ± 835	2	5.36	8
Glutamic acid*	10	559	27	35	77	1283 ± 683	712 ± 342	1.8	4.09	9
Linoliec acid	11	456	29	34	77	998 ± 638	332 ± 410	3.01	4.81	10
Metabolite Unk. 19	13	374	439	7	73	6978.5	2909.8	2.4	4.09	11
Metabolite Unk. 7	14	317	13	56	144	1342 ± 695	1761 ± 773	0.75	2.36	12
Metabolite Unk. 8	15	283	2	83	47	1520 ± 757	1943 ± 697	0.78	2.25	13
Maleic acid*	16	239	30	32	73	3237 ± 1554	2544 ± 855	1.27	2.14	14
Metabolite Unk. 12	17	229	4	72	73	1760 ± 1202	2489 ± 1344	0.71	2.21	15
Metabolite Unk. 20	18	225	451	6	73	165 ± 312	21 ± 14	7.97	2.53	16
Metabolite Unk. 4	19	158	22	38	77	163 ± 107	61 ± 28	2.67	5.05	17
Myristate*	21	150	82	26	77	286 ± 190	93 ± 57	3.08	5.35	18
L-methionine*	22	147	1	122	155	365 ± 216	626 ± 170	0.58	5.2	19
CoA fragment	23	142	3	73	77	5768 ± 6657	800 ± 312	7.21	4.08	20
Metabolite Unk. 22	24	140	161	21	73	2507 ± 1051	1803 ± 551	1.39	3.24	21
L-tryptophan	27	112	113	23	73	374 ± 306	197 ± 171	1.9	2.77	22
Metabolite Unk. 8	28	105	31	31	77	594 ± 404	181 ± 72	3.28	5.51	23
Metabolite Unk. 24	29	103	297	14	73	1297 ± 646	2114 ± 787	0.61	4.4	24
Alanine*	30	93	6	70	116	2628 ± 2061	5402 ± 2419	0.49	4.78	25
Oxalic acid*	32	91	17	46	73	10273 ± 7715	21428 ± 7146	0.48	5.8	26
Metabolite Unk. 14	44	46	7	69	73	8826 ± 5823	17764 ± 5266	0.5	6.2	27
Malonic acid*	46	44	10	66	55	167 ± 96	276 ± 81	0.61	4.7	28
Metabolite Unk. 16	69	37	16	46	73	390 ± 237	702 ± 205	0.56	5.4	29
Glutamine	89	31	21	39	47	116 ± 46	58 ± 25	2	6.1	30
Glycerol*	146	24	24	37	152	600 ± 312	1153 ± 344	0.52	6.5	31
Metabolite Unk. 15	161	22	14	52	73	292 ± 197	708 ± 350	0.41	5.7	32
Metabolite Unk. 2	199	19	26	35	73	612 ± 431	1560 ± 783	0.39	5.8	33

PCA scores plots and loadings for top 30 hits discovered with standard F-ratio analysis, control-normalized F-ratio analysis, and combined.

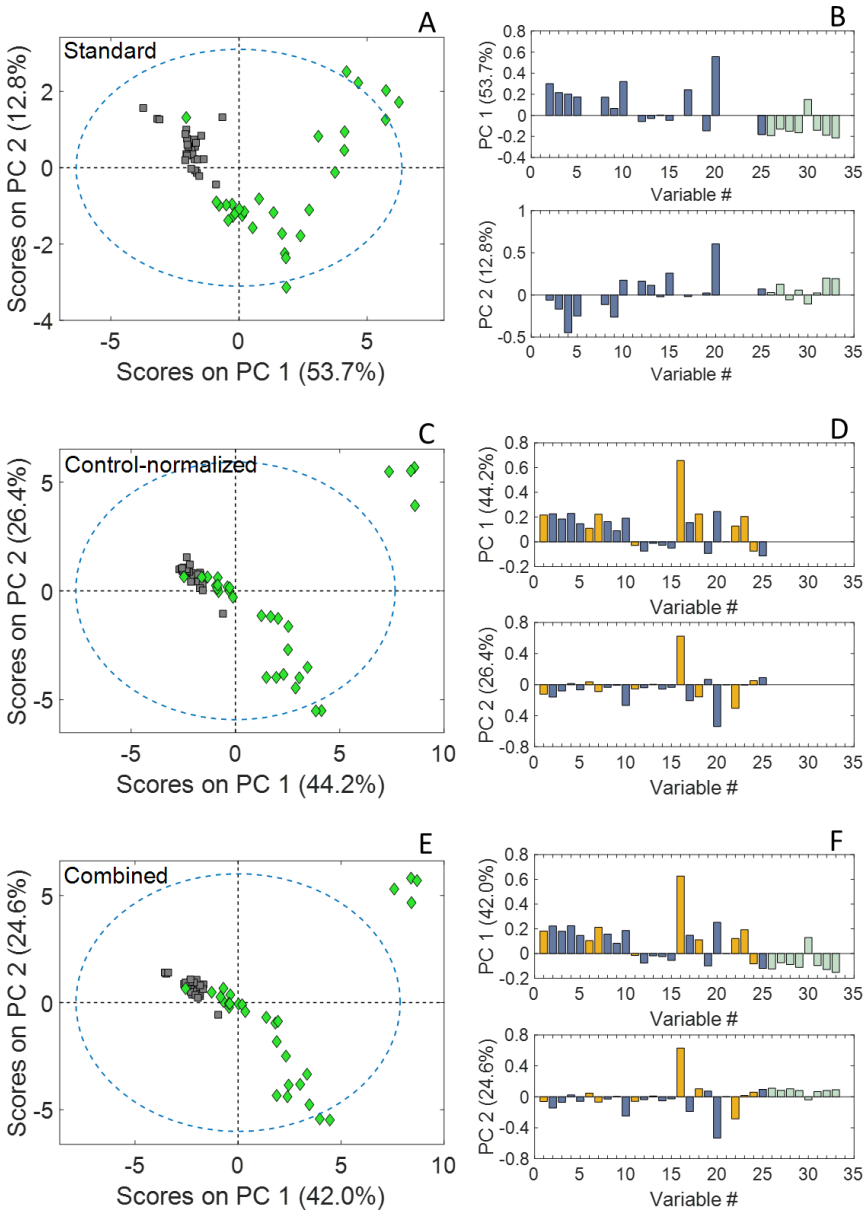

Figure A.3. Scores and loadings plots for sample and metabolite data summarized in Tables A.2-A.4, where the variables refer to the metabolites labeled as "Var. #" in the far-right table column. (A) PCA scores and (B) loadings plots for 23 metabolites that pass the t-test from the standard FRA hit list. The classes (controls, gray squares; patients, green diamonds) are well separated by PC 1.. (C) PCA scores and (D) loadings plots for 25 metabolites which pass the t-test from the control-normalized FRA hit list. Controls are tightly clustered indicating the metabolites have minor within-class variance. (E) PCA scores and (F) loadings plots from the combined standard and control-normalized FRA hit lists. The class separation appears functionally the same as the PCA scores plot for the control-normalized hit list in (C).

Metabolite profiles for 2 representative controls and 4 representative patient samples

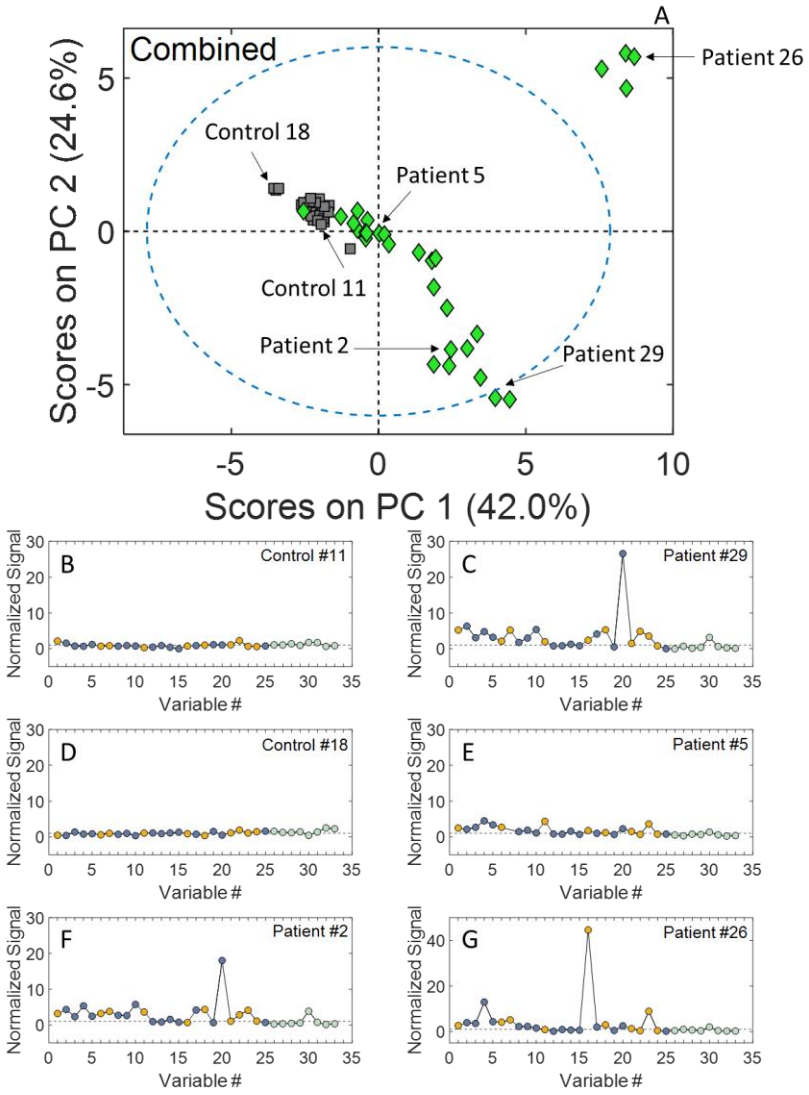

Figure A.4. Metabolite profiles for 6 representative samples and position in combined PCA scores plot (Figure A.3E). The normalized signal is calculated by taking the signal for each metabolite for the given sample and dividing by the mean for that metabolite within the control group. Thus, the metabolite signals for the control samples will be centered near 1. (A) PCA scores plot with sample # labeled for two control samples and 4 patient samples. (B) Metabolite profile for control #11, located just left of center on PC 1 and centered on PC 2. (C) Control #18, located on the far left of PC 1. (D) Patient #2, negatively loaded on PC 2 and positively loaded on PC 1. It appears variable #20 plays a significant role in the separation on PC 2. (E) Patient #29, heavily loaded on PC 1 and PC 2, with significant upregulation of variables #20 and #6, among others. (F) Patient #5, centered on PC1 and PC2, with minor upregulation of variables #11 and #23. (G) Patient #26, which is heavily loaded on PC1 and PC2, has significant upregulation of variables #4, #16, and #23.

VITA

Sarah Prebihalo was born and raised in Columbus, Ohio in 1989. After graduating from Upper Arlington High School in 2007, she relocated to Cincinnati, Ohio to attend Xavier University for a Bachelor of Science degree in Chemistry. As an undergraduate, Sarah was named as one of two Clare Boothe Luce Scholar's, awarded to female students in science, technology, engineering and mathematics, which provided her with the opportunity to perform summer research with Dr. Barbara Hopkins. She also had the opportunity to participate in a summer research program with Dr. Patrick Woodward at the Ohio State University, where she synthesized double perovskites.

Upon graduating with her chemistry degree in 2011, Sarah was hired as an analytical chemist with Advanced Testing Laboratory in Cincinnati, Ohio. During her two-year tenure, she was involved in ion chromatography and liquid chromatography method development for a major consumer product brand. In the fall of 2013, Sarah moved to State College, Pennsylvania to attend Pennsylvania State University as a master's student in Forensic Science. Under the guidance of Dr. Frank Dorman, Sarah investigated the use of multidimensional gas chromatography towards the discovery of emerging environmental contaminants and earned her MPS degree in 2015.

In the fall of 2015, Sarah began her research at the University of Washington with Dr. Robert Synovec. While at the University of Washington, in addition to her individual research, Sarah was involved in the commercialization of chemometric software for multidimensional gas chromatographic data. Following completion of her doctorate, Sarah will relocate to Washington D.C. to continue research as a Postdoctoral Fellow at the Food and Drug Administration.