Advances in instrumentation and chemometric analysis for multidimensional chromatography with mass spectrometry detection

Sonia Schöneich

A dissertation

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee

Robert E. Synovec, Chair

Ashleigh Theberge

Bo Zhang

Program authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Advances in instrumentation and chemometric analysis for multidimensional chromatography with mass spectrometry detection

Sonia Schöneich

Chair of the Supervisory Committee
Robert E. Synovec
Chemistry

Quantitative separations by gas chromatography (GC) or liquid chromatography (LC) with mass spectrometry (MS) detection are robust techniques used for the characterization of a broad variety of complex sample matrices. Due to the high complexity of samples, compounds of interest are rarely fully separated in one-dimensional (1D) separations, making quantification and identification a challenge. Two approaches to address this challenge includes using high resolution mass spectrometry (HRMS) for detection to gain selectivity, and using comprehensive two-dimensional chromatography for improving the characterization of complex samples. Each of these solutions involve increasingly more complex instrumentation and provide information-rich data sets that require efficient and effective methods to mine the data for obtaining relevant information, as manual inspection can be tedious for hundreds to thousands of analytes.

Although GC and LC are commonplace for commercial use, despite its advantages over 1D separations comprehensive multidimensional chromatography still requires optimization in instrumentation and data analysis tools to be more widely adopted. In this dissertation, developments to address challenges in LC-HRMS data processing and analysis as well as developments in instrumentation and chemometric analysis of comprehensive two-dimensional gas chromatography (GC×GC) with MS detection will be discussed. First, a workflow for processing information-rich LC-HRMS data for non-targeted analyses is introduced, with the first demonstration of tile-based Fisher ratio for LC-HRMS data. Next, a novel modulation technique termed dynamic pressure gradient modulation (DPGM) for GC×GC with time-of-flight mass spectrometry (TOFMS) is studied in multiple modes. The first mode is negative pulse mode, which was demonstrated for fast separations generating narrow peaks, producing high quality data that is amenable to chemometric decomposition. Next, DPGM is demonstrated in full modulation mode for GC×GC-TOFMS, achieving improvements in sensitivity, limit of detection, and trilinear data structure. DPGM was further investigated, demonstrating the optimization of experimental parameters for obtaining high peak capacity applied to multiple sample matrices. Finally, a novel modification to tile-based F-ratio analysis to handle highly variable metabolomics-related pacu fish samples was introduced with a computational method to remove sample derivatization artifact interferences.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Robert Synovec, for his guidance and support since I joined his group in 2018 here at the University of Washington. I am grateful to have had the opportunity to pursue many challenging and exciting projects with the support and encouragement of my advisor. I have learned so much and have grown immensely as a researcher and as a person working in this group. I would like to thank my advisor at KU, Dr. Susan Lunte, for giving me the opportunity to begin undergraduate research in analytical chemistry at KU. This is where I first became excited about research and the experience inspired my career goals.

My friends here in Seattle and from before graduate school have been such important parts of my support system. My move to Seattle was the first time I left home, so I would not have been able to push through the challenges of graduate school and navigating this big life transition without my friends and lab mates here in Seattle. Caitlin, Grant, Joe, Kristy, Paige, and Tim thank you for your friendship and for being so encouraging and supportive. I would like to specifically thank the senior students who trained me when I first joined the group, in particular Dr. Derrick Gough and Dr. Paige Sudol. Thank you for training me, encouraging me, and challenging me. From Lawrence, thank you McKenzie and Whitney for being such great friends since our undergraduate years. McKenzie, thank you for visiting me my first year here when I needed a reminder of home and for meeting up with me when I return to Lawrence even if we both have very busy schedules. Whitney, thank you for our phone calls where we could relate to each other's grad school experiences and support each other from further away.

Finally, to my family who I cannot thank enough. A brief paragraph does not do justice to the extent of the thank you to which you all deserve. Firstly, a massive thank you to my mother

and father for being the most supportive parents, encouraging me every step of the way. Thank you for being so excited about my successes and my future, as well as giving the best advice when I was faced with challenges. Antonio, Sebi, and Erin – thank you for visiting me, challenging me, and supporting me. I'm excited to hopefully be in the same time zone again. To all my extended family including grandparents, uncles, aunts, and cousins – thank you for your encouragement, support, and comedic relief. Finally, my Napoleone deserves recognition as the best and smartest dog in the world. He has been a constant source of joy and cute pictures of this pup brighten any day.

DEDICATION

To my wonderful family.

Chapter 1. Introduction to chromatographic separations, multidimensional gas chromatography, and chemometric analysis

Some parts of this chapter have been reproduced from Caitlin N. Cain[†], Sonia Schöneich[†], Robert E. Synovec, "Recent Advances in Comparative Analysis for Comprehensive Two-Dimensional-Gas Chromatography-Mass Spectrometry Data," Fundamentals and Applications of Multiway Data Analysis.

[†]These authors contributed equally.

1.1 Introduction to chromatography and multidimensional gas chromatography

1.1.1 Fundamentals of chromatography

The analytical goal in quantitative chromatographic separations is to resolve all components in the sample, allowing for accurate identification and quantification of all meaningful compounds. Chromatography is a technique used for the separation of multiple compounds in a mixture that occurs due to the interactions of the compounds, or analytes, with a mobile phase and stationary phase. Particularly for wall-coated open tubular (WCOT) gas chromatography (GC), the mobile phase is an inert gas (carrier gas), typically high purity hydrogen or helium and the stationary phase is a liquid or polymer coated onto a fused silica capillary column [1]. GC is used for the analysis of volatiles and semi-volatiles compounds and separation of analytes is dependent on polarity and boiling point, requiring thermally stable samples. For liquid chromatography (LC) separations, the mobile phase is liquid with a packed solid particle (µm scale) stationary phase in a relatively short column and is an excellent technique for a broad range of analytes [2]. Although both LC and GC are discussed herein, the main focus will be on GC as a majority of the work in this dissertation was accomplished using comprehensive multidimensional gas chromatography.

The separation of analytes in chromatographic separations is governed by their partitioning between the mobile and stationary phases and the goal is to resolve analytes such that the analyst can accurately quantify and identify the compounds of interest in the sample. Analytes with greater affinity for the stationary phase will move through the column much slower than analytes with low affinity for the stationary phase due to the intermolecular forces at play. The migration dependent upon the differences in affinities for the analytes and the column stationary phase thus results in the separation of compounds with different chemical structures. The thermodynamic partition or distribution coefficient (K_D) can be used for characterizing this system as

$$K_D=\frac{[Analyte]_s}{[Analyte]_m} \quad (1.1),$$

where the numerator is the analyte equilibrium concentration in the stationary phase and the denominator is the equilibrium concentration of the analyte in the mobile phase [3]. Thus, a greater analyte concentration in the stationary phase results in a large K_D, meaning the analyte has a longer residence time on the column, i.e., is retained to a greater extent. Note that for GC separations, the mobile phase is an inert gas and thus K_D is truly only dependent on the partitioning of analytes in the stationary phase and temperature. The retention of the analyte on a column with a specific stationary phase is related to K_D via the phase ratio (β) and is related to the analyte's Gibb's free energy of evaporation ($\Delta G°$) [4]. Using the relationship between $\Delta G°$, enthalpy ($\Delta H°$), and entropy ($\Delta S°$), where $\Delta G°=\Delta H°-T\Delta S°$, the retention factor ($k$') can be expressed in terms of $\Delta S°$, $\Delta H°$, and β as

$$\ln(k') = \frac{-\Delta H°}{RT} + \frac{\Delta S°}{R} - \ln(\beta) \quad (1.2),$$

where R is the ideal gas constant and T is the temperature of the separation in the GC oven [4]. This thermodynamic information is particularly useful for analysts to evaluate the selectivity of a column stationary phase for retaining certain analytes relative to other analytes.

For an analyte that essentially does not partition into the stationary phase, the analyte is used as a dead time marker (t_0), indicating the time that an unretained compound takes to migrate through the entire column given the experimental conditions. Each component separated then has a corresponding measurable retention time (t_R), or the time at which the analyte elutes. Thus, the *k*' of a given analyte in the separation can be expressed as

$$k' = \frac{t_R - t_0}{t_0} \quad (1.3),$$

where *k*' describes the ratio of the time a given analyte partitions in the stationary phase to the time it is in the mobile phase. The retention of the analyte can decrease by using a temperature program for the oven in GC separations and by using a mobile phase composition gradient with multiple solvents in LC separations. The use of a program/gradient rather than an isothermal GC separation or isocratic LC separation, is to address the general elution problem. The general elution problem is observed as analytes with high *k*' are retained to a longer than needed, resulting in higher resolution between peaks but also wider peak widths that limits the number of analytes that can be separated within the separation space, while the opposite is true for analytes with low *k*'. Using a program for GC or gradient for LC lowers *k*' of analytes so that ideally the *k*' approaches 0 to obtain narrow peaks of approximately the same width for all analytes. Further, the retention factor and therefore peak broadening is affected by the phase ratio [5,6]. The phase ratio (β) for open tubular columns is defined as

$$\beta = \frac{V_m}{V_s} = \frac{d_c}{4d_f} \quad (1.4),$$

or the ratio of the volume of the mobile phase (V_m) to the volume of the stationary phase (V_s). For GC separations, it is dependent upon the inner diameter of the column (d_c), typically 100-320 µm, and the film thickness (d_f), or thickness of the stationary phase layer that is typically 0.1-1.5 µm. For GC separations, when the d_f is larger relative to d_c (lower β), the analytes are retained longer thus increasing k', meaning that hotter temperatures are needed for elution. The opposite is true when β increases with lower d_c relative to d_f. The magnitude of the phase ratio affects the degree to which analytes are separated and their elution temperature [6,7]. Thus, the analyst can select the proper column specifications to manipulate the phase volume ratio and ultimately affect the elution temperature of the analyte [8].

The signal of the analyte at the end of the column is detected as a concentration-dependent profile that follows a Gaussian distribution centered about the analyte t_R, corresponding also to the analyte peak height. Detectors commonly used include mass spectrometry (MS) for both GC and LC analyses, flame ionization detection (FID) for GC analysis, and ultraviolet-visible (UV) absorbance detection for LC analysis. MS detectors are universal detectors, where analytes are ionized and the detector collects signal across multiple channels corresponding to the mass-to-charge ratio (m/z) of the fragments at each time point, providing a mass spectrum, or essentially a chemical fingerprint of the fragmentation patterns of the analytes. Thus, each analyte can be visualized as a Gaussian profile at using selective m/z or as a single Gaussian representing the sum of m/z. UV detectors can scan a range or wavelengths measuring analyte absorption or a selective wavelength for targeting specific analytes. Similarly, the FID is univariate, collecting signal for hydrocarbons at each time point providing a single Gaussian profile per analyte detected. The Gaussian profile is the ideal peak shape distribution, however deviations from this

ideal symmetric shape are experimentally observed and can pose a challenge to optimizing the separation efficiency and subsequent chemometric decomposition efforts. Such deviations from the Gaussian peak shapes may arise from issues with injection, on-column analyte degradation, and/or non-ideal interactions of the analyte with the stationary or mobile phase [3]. For the discussion of relevant figures of merit, the ideal Gaussian peak shape will be considered.

To determine the extent to which analytes are overlapped, mathematically one must consider how similar the t_R of the analytes are, and how wide the peaks are, or the peak width. The peak width-at-base (W_b) is an important metric to determine how much of the separation space each analyte peak occupies. The W_b is determined for a symmetric Gaussian peak by measuring ±2σ from the t_R, which accounts for 95% of the peak area measured at 13.4% of the peak height. The peak height is also calculated at the center of the Gaussian peak profile measuring the signal from the top of the peak profile at the t_R down to the baseline. It is ideal to optimize the chromatographic separation such that peaks are narrow, which would allow for a large number of analytes to be separated given the separation space as well as reduce the extent to which analytes overlap with each other. The resolution between two analytes is defined as

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

where $W_{b,1}$ is the peak width of analyte 1 with retention time $t_{R,1}$, and $W_{b,2}$ is the peak width of analyte 2 with retention time $t_{R,2}$, all labeled in Figure 1.1 for two simulated analyte peaks with R_s of 1. This metric is used to measure the extent of peak overlap between multiple analytes, with a R_s of 1.5 corresponding to baseline resolved analytes, which is ideal however impossible to achieve for all analytes experimentally for complex samples, and a R_s of 1 being accepted for accurate quantification (using peak heights, not areas) and identification. At lower R_s, typically chemometric deconvolution or decomposition is necessary, but common

deconvolution/decomposition techniques start to fail at R_s of 0.3. While chemometric tools can be beneficial in resolving overlapped analytes, R_s can be experimentally improved by optimizing the flow rate through the column or by using either a temperature program for GC or a mobile phase composition gradient for LC.

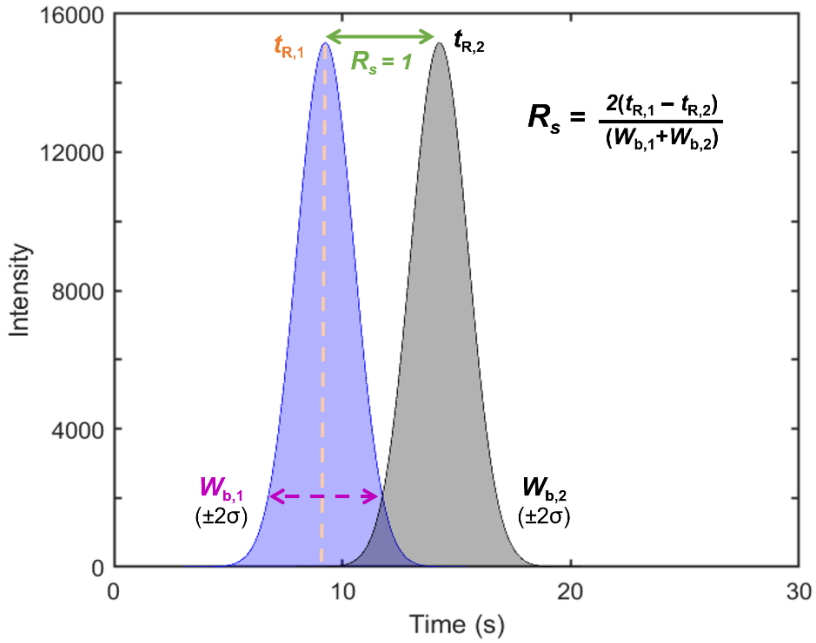


Figure 1.1. Illustration of the separation of two analyte peaks at unit resolution (R_S =1). The retention times and peak widths for each analyte n ($t_{R,n}$ and $W_{b,n}$, respectively) are provided as reference for the determination of R_S.

Certainly, the goal for applying chromatography is to obtain narrow peak widths to maximize the number of analytes that can be separated at unit R_s concurrent with optimizing analyte detectability, i.e., higher signal-to-noise ratio (*S/N*). Indeed, the peak width increases by band broadening, both from on-column and off-column sources, and thus the peak capacity and column efficiency are reduced. On-column band broadening can be mechanistically explained by contributions from eddy diffusion (not present for open tubular GC), longitudinal diffusion, and resistance to mass transfer in the mobile and stationary phases, and/or the kinetics of the analytes partitioning between the stationary and mobile phase. As these band broadening mechanisms are

dependent upon flow rate, applying the optimal flow rate reduces each of these contributions to minimize band broadening. Using too low of a flow rate increases the longitudinal diffusion contribution, while using too fast of a flow rate increases the mass transfer contribution and therefore decreases the column efficiency. While this accounts for band broadening resulting from within the column, band broadening may also be due to off-column issues with injection, detection, or dead volumes within the instrument platform [9–11]. Therefore, the analyst must consider numerous experimental parameters including the previously mentioned column specifications (dimensions, composition, stationary phase thickness for GC, particle size and distribution for LC), flow rate, temperature program for GC, and mobile phase gradient for LC, and injection pulse width to modify to minimize band broadening.

Reducing band broadening ensures narrow peak widths, which is ideal for improving R_s of analytes and maximizing the number of analytes that can be resolved within the separation time (t_{sep}), or peak capacity. The peak capacity (n_c) for an optimized temperature programmed 1D-GC separation or gradient 1D-LC separation can be estimated by

$$n_c = \frac{t_{sep}}{w_b} \quad (1.6),$$

where t_{sep} is the entire separation time. This is a measurement of the number of analytes that can be separated ($R_S = 1$) within a run (Figure 1.2) and is used as a metric to compare separation performance. However, it is unlikely that when analyzing highly complex samples by LC or GC all analytes are fully chromatographically resolved, making identification and quantification of analytes of interest a challenge [12]. [12]. One excellent approach to address this issue is by using comprehensive multidimensional chromatographic techniques [13,14] to improve upon the selectivity and attainable peak capacity of the chromatographic separation. In this dissertation, Chapter 2 will focus on 1D-LC, while the remaining chapters (3-6) will cover my work with

comprehensive two-dimensional gas chromatography (GC×GC), thus moving forward in this introductory chapter I will mainly focus on GC×GC.

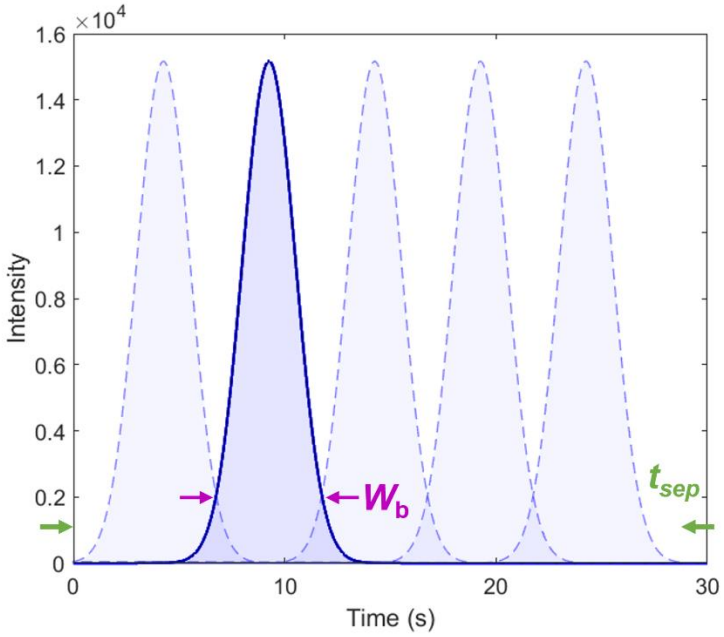


Figure 1.2. Illustration of peak capacity in 1D LC and GC separations. The peak width (W_b) for one analyte is indicated in purple, and the separation time is indicated in green.

1.1.2 Introduction to comprehensive multidimensional gas chromatography

Analyses of complex samples require efficient instrumental platforms and data handling methods for successful characterization. GC×GC was first introduced by Liu and Phillips [15] for improved resolution of volatiles and semi-volatiles in complex sample matrices. A GC×GC separation is achieved by connecting two columns with sufficiently orthogonal stationary phase selectivity in series and interfaced with a secondary injector, termed the modulator (Figure 1.3). Broadly, modulators are either thermal or flow-based and are an important consideration for the experimental design, which will be discussed further in section 1.1.3. The modulator traps and/or focuses effluent from the first dimension (^{1}D) column and re-injects it onto the second dimension (^{2}D) column at a user specified interval for secondary separation, termed the modulation period (P_M). For GC, the ^{1}D column is longer (typically 20-30 m long) than the ^{2}D column, which is

usually only a couple meters long since the ^{2}D separation is much shorter. Fast GC×GC separations report using a P_M (the ^{2}D separation time) down to 50 ms [16], however P_M are more commonly a few seconds. Since the ^{2}D separation is so quick relative to normal temperature programs, this separation is considered essentially a (pseudo) isothermal separation. GC×GC separations offer improved selectivity due to the complementary separation mechanism of the ^{2}D column, increased sensitivity, and an ideal increase in peak capacity by a factor of ~10 compared to standard 1D-GC separations [17,18]. Furthermore, statistical overlap theory for multidimensional chromatographic separations highlights that GC×GC can be advantageous in increasing the number of analytes that can be completely resolved (i.e., singlets) relative to 1D-GC [12,19,20]. These benefits allow for the better characterization of complex samples and thus, GC×GC has been used for complex applications including fuel [21,22], food [23,24], cosmetic [25,26], forensic [27,28], metabolomic [29,30], and environmental samples [31,32].

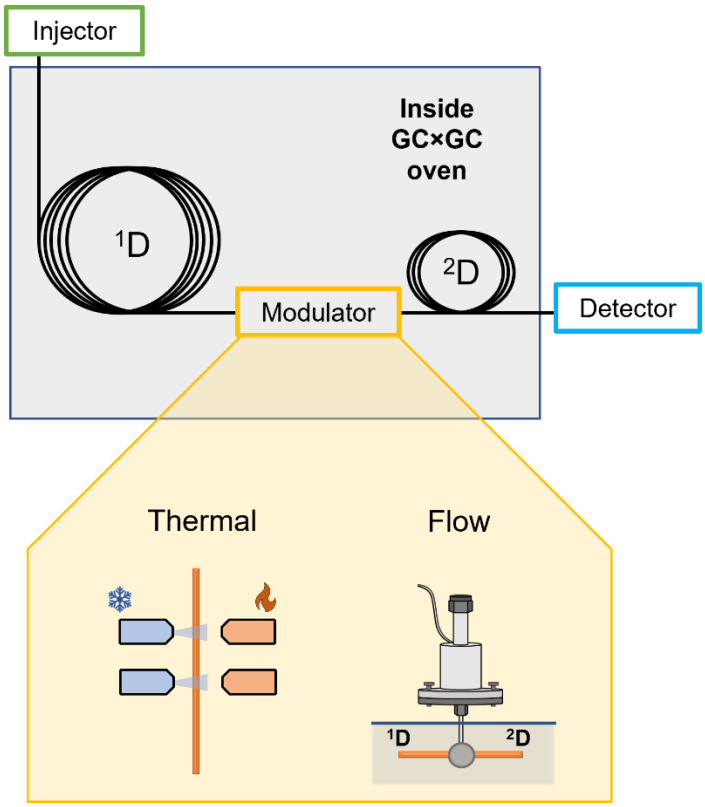


Figure 1.3. GC×GC instrument schematic, with examples of both thermal and a flow-based modulator. Note that the flow modulator shown here is mounted outside the GC×GC oven.

For GC×GC separations, the ratio of ^{1}D and ^{2}D phase volume ratios, or beta ratio (β_R), is an important metric that can be tuned to alter the retention factor on ^{2}D (^2k'). The β_R metric is defined as

$$\beta_R = \frac{^1\beta}{^2\beta} = \frac{^1d_c\ ^2d_f}{^2d_c\ ^1d_f} \quad (1.7),$$

where β is calculated first for each dimension using Eq. 1.4. Just as the ^1k' is affected by β as discussed previously, the ^2k' is affected by the β_R. With increased $^1\beta$ (higher β_R), the ^1k' is smaller with analytes eluting at a colder temperature, meaning the ^{2}D separation starts at a lower temperature and therefore the separation is slower meaning analytes have a higher ^2k'. Conversely, with lower $^1\beta$ (lower β_R), analytes are more retained (higher ^1k') meaning they elute at higher temperatures and the ^{2}D separation starts at a higher temperature leading to a faster separation and lower ^2k'. Thus, proper column selection for the pair of columns used in GC×GC is an important factor when considering separation performance due to its effect on peak capacity and resolution [7].

Prior to discussing the 2D peak capacity and resolution, the peak widths on both ^{1}D and ^{2}D should be discussed taking into consideration effects of the modulation process. The ^1W_b is calculated as the reconstructed ^{1}D profile after the modulation process and can be achieved by fitting a Gaussian curve to the tops of the ^{2}D modulated peaks. Due to the resampling of the ^{1}D effluent during the modulation process, the ^{1}D profile of the Gaussian fit to the ^{2}D modulated peaks overestimates the ^1W_b that would be calculated with a detector at the end of the ^{1}D column (just 1D separation). Therefore, the measured ^1W_b from the 2D data is the effective ^1W_b (^1W_b*) and we must consider the modulation ratio to estimate the true ^1W_b [33]. To minimize sampling-

induced band broadening the ^{1}D analyte profile should ideally be sampled 2-4 times, which also ensures the greatest improvement in sensitivity can be achieved relative to the 1D separation. Note that sampling-induced band broadening is not observed for the ^{2}D peaks since these peaks are detected with several pixels, though since the ^{2}D separation is essentially isothermal, the ^2W_b for each separation on ^{2}D increases with elution time on the second dimension, ^2t_R. Such sample-induced band-broadening will be investigated further in Chapter 5 for the flow modulation technique introduced in Chapters 3-4.

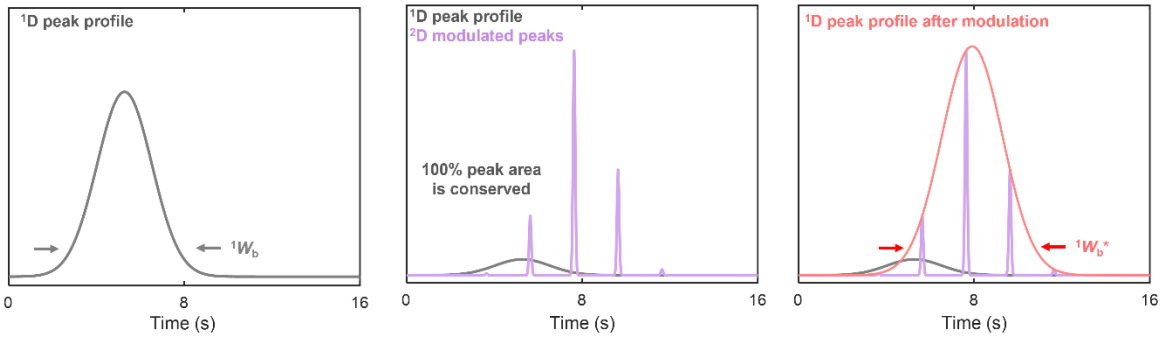


Figure 1.4. Illustration of sampling-induced peak broadening of ^1W_b. The 1D unmodulated peak profile is provided in all panels in gray, the ^{2}D modulated peak profiles are provided in purple, and the reconstructed ^{1}D profile fitting a gaussian to the ^{2}D peaks is provided in red.

The 2D resolution ($R_{s,2D}$) between two peaks is calculated as the Euclidean norm between the resolution in each dimension and is defined as

$$R_{s,2D} = \sqrt{{}^1R_s^2 + {}^2R_s^2} \tag{1.8},$$

where ^1R_s and ^2R_s are the resolution on ^{1}D and ^{2}D, respectively [34]. Thus, analytes that are highly overlapped on ^{1}D may be sufficiently resolved on ^{2}D, providing a high enough $R_{s,2D}$ that allows for accurate identification and quantification, either directly from the raw data or facilitated by chemometric decomposition. This will be investigated further in Chapter 3, for particularly overlapped analytes in an isothermal separation. Similarly, for a 2D separation, the peak capacity ($n_{c,2D}$) at unit resolution is defined as

$$n_{c,2D} = {}^{1}n_c \times {}^{2}n_c \quad (1.9),$$

with

$${}^{2}n_c = \frac{P_M}{{}^{2}W_b} \quad (1.10),$$

where ${}^{2}W_b$ is the peak width on the second dimension and ${}^{1}n_c$ is as defined in Eq. 1.6. High peak capacity and 2D resolution of analytes thus relies upon narrow peak widths on both dimensions. As previously mentioned, the ideal peak capacity obtained on ^{2}D is ~10, under conditions in which the ^{1}D peak capacity has not been compromised by using a long P_M, meaning the ^{2}D peak width should be on average a factor of 10 smaller than the P_M. Narrow peaks on ^{2}D can be achieved increasing the flow rate on ^{2}D, however especially when using flow modulation the analyst must keep in mind the compatibility of flow rate with the detector and the extent to which peaks will broaden on ^{1}D. Additionally, since the peak capacity is corrected for the amount of separation space utilized, using a high flow rate on ^{2}D may decrease the amount of separation space used, thus decreasing the peak capacity even with narrow peak widths.

Relative to GC, GC×GC provides an increase in sensitivity due to the modulation process, though the mechanism by which this is achieved differs based on thermal or flow-based modulation (discussed in the next section). The signal enhancement (SE), termed the detector response enhancement factor (*DREF*) in Chapter 4 [35], is defined as the ratio of the ^{2}D peak height (^{2}h) to the ^{1}D peak height (^{1}h), or simply

$$SE = \frac{2_h}{1_h} \quad (1.11).$$

Thus, for a modulation technique that provides 100% duty cycle, i.e., total transfer of ^{1}D effluent to the ^{2}D column, for a given ${}^{1}W_b$, an analyte with a narrower ${}^{2}W_b$ has a higher SE, since the peak area is conserved and therefore the ^{2}h is higher. Using flow modulation, higher SEs are achieved

with detectors such as the FID since high ^{2}D flow rates (~20 mL/min) into the detector are common. Such high flow rates ensure narrow $^{2}W_{b}$ are achieved, however they are incompatible with MS detector pumping capacities (6 mL/min for TOFMS), requiring further developments in total transfer flow modulation techniques compatible with MS detection. In Chapter 4, both signal and *S/N* enhancement were investigated using a novel flow modulation technique as the ratio of the *S/N* of the most intense modulated ^{2}D peak to the *S/N* of a ^{1}D peak that is corrected for the acquisition frequency. The *S/N* improvement relative to 1D-GC is indeed smaller in magnitude than the SE, however it further demonstrates the advantage of GC×GC as lower limits of detection (LOD) can be achieved. Additionally, the analyte must have sufficient signal above the noise to be quantified (limit of quantification) or identified accurately. For particularly noisy chromatograms, data pre-processing algorithms can be used to smooth the noise to improve *S/N*, or chemometric decomposition methods can be used to uncouple the background noise from the pure elution profiles of analytes. These alternative approaches will be discussed further in the next section of Chapter 1.

The quality of the data and resulting analysis workflow is heavily dependent upon the experimental design, as will be discussed in the next sections. This includes instrumentation and separation conditions used to collect the data as well as the *a priori* knowledge of the samples. Many early 2D-GC reports utilize heart-cutting, where a subset of the analytes from the 1D separation are subjected to secondary separation [36,37]. Heart-cutting is particularly useful for targeting a known set of analytes of interest and discarding the remaining information about the samples. In contrast, GC×GC is especially useful for full characterization of samples without prior knowledge of analytes of interest (non-targeted studies). GC×GC is commonly coupled to either univariate detectors, such as the flame ionization detector (FID) [10,35,38] or the electron

capture detector (ECD) [39,40], or multivariate detectors, most commonly mass spectrometry (MS) [41–45]. I will focus on GC×GC with MS detection herein as all data collected for my work was collected with MS detection due the added selectivity compared to univariate detectors. Furthermore, a greater number of studies using comprehensive three-dimensional GC (GC×GC×GC or GC^3) with MS detection have been reported in recent years demonstrating further improvements in selectivity and sensitivity [46–48]. GC^3-MS separations are achieved by adding another column in series to a GC×GC instrumental platform with a second modulator that re-injects effluent from the ^{2}D column onto the ^{3}D column at a second user-defined modulation period (^2P_M). This allows for greater peak capacity ($n_{c,3D}$), with further improvements in sensitivity, especially when 100% mass transfer is achieved, and selectivity due to the third-dimension column [46–48]. While GC×GC and GC^3 instrument platforms using FID detection have been reported as well, MS is especially useful for identification of analytes in non-targeted studies and for obtaining pure mass channels (*m/z*) that can be used for accurate analyte quantification. On top of the added dimension provided by MS detection, another degree of complexity is included when investigating samples that belong to multiple classes (i.e., control versus experimental classes) for comparative analysis. Specifically, comparative analyses aim to describe the chemical similarities and differences between multiple types of samples or sample classes. These studies may involve comparing a single chromatogram from two classes (i.e., pairwise analysis) or multiple replicates for several classes. For these data sets, where each chromatogram is comprised of hundreds to thousands of peaks, visual and manual inspection of the chemical differences between classes is tedious. Therefore, chemometric tools can be beneficial to reduce the amount of data and time required to mine.

1.1.3 Instrumentation and data processing

The data processing workflow is greatly dependent upon several key factors including the modulator, detector, and experimental design. These factors ultimately influence the final structure of the data along with the chemometric methods and/or approaches allowed. Second-order data is generated when GC×GC is coupled with univariate detectors (signal at every ^1t_R and ^2t_R). Meanwhile, coupling GC×GC with multivariate detectors produces third-order data, where the three data dimensions describes ^1t_R, ^2t_R, and spectrum (Fig. 1.5A). The added selectivity from multivariate detectors like MS is especially beneficial for analytes with low 2D resolution ($R_{s,2D}$) for increasing the likelihood of accurate statistical calculations, identification, and quantification. Furthermore, for many comparative analyses, an additional dimension is added to factor the multiple samples that comprise each class. For example, Fig. 1.5B demonstrates the augmentation of GC×GC-MS chromatograms from samples belonging to either class 1 (red) or class 2 (blue). The resulting data array is now fourth-order with the dimensions of ^1t_R, ^2t_R, *m/z*, and samples. This complex data structure can be reduced to three dimensions if the samples from all classes are concatenated or stacked, such as in Fig. 1.5C. However, while a GC×GC-MS instrument may be capable of producing third-order data, its operation may not produce trilinear data, which can exclude/compromise the use of specific chemometric methods. A trilinear data structure requires that each data dimension is linearly independent, where each analyte has sufficiently reproducible retention times and peak shapes, and the peak signals are concentration dependent. Therefore, modulator and detector selection along with its impact on the data preprocessing workflow will ultimately influence the success of the chemometric analysis and these considerations will be discussed herein.

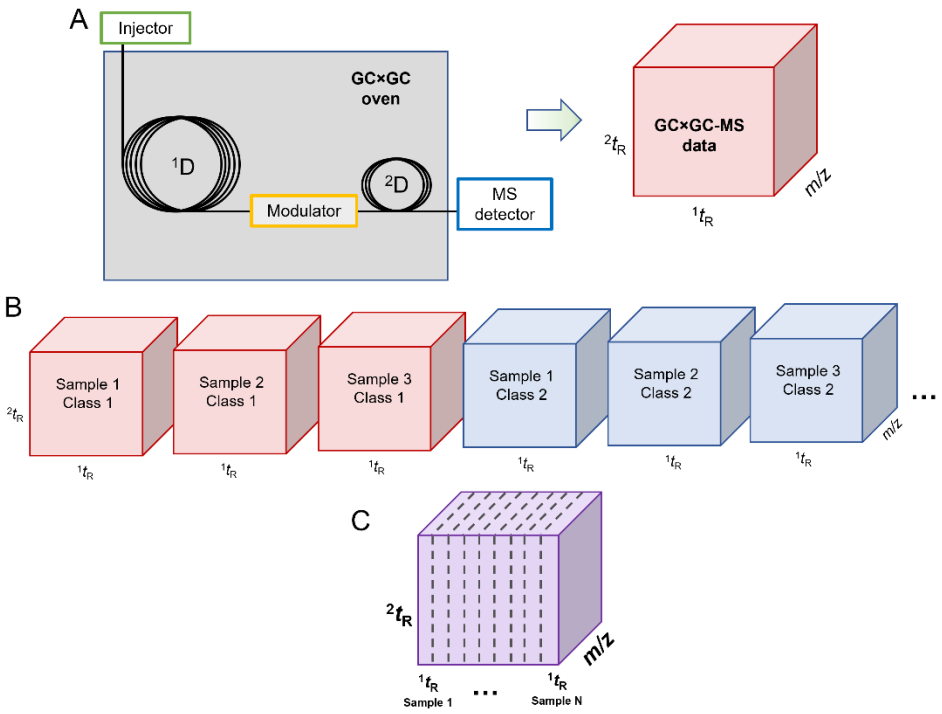


Figure 1.5. Visualization of GC×GC-MS data structure. (A) Schematic of a GC×GC-MS instrument with visualization of the data collected for a single run. At each pixel (^1t_R and ^2t_R), the mass spectrum is collected for the *m/z* specified by the analyst, resulting in the third order data structure given by the cube. (B) For many comparative analyses, multiple replicates are collected per class (red for class 1 and blue for class 2) and are augmented, resulting in a fourth-order data structure prior to chemometric analysis. (C) Samples can similarly be augmented or stacked to retain the third order dimensionality of the data prior to chemometric analysis, in this case with samples from all classes concatenated along the ^1t_R axis.

An analyst must consider both the nature of the analytes in the samples and the quality of the data required for further analysis when selecting either a thermal or flow modulator. Thermal modulators use cold and hot pulses to trap and re-inject analytes onto the ^{2}D column with increased sensitivity due to the thermal focusing of analytes in the cold zone and total transfer of effluent from the ^{1}D column to the ^{2}D column. Initially, cryogenic thermal modulation was used to achieve the short cold pulses; however, a disadvantage of this modulation technique was the cost of using liquid cryogen. Hence, cryogenic-free thermal modulation was introduced to mitigate this issue [49–54]. Commercial GC×GC-MS instruments are often equipped with thermal modulators as they are more readily compatible with the MS detector pumping capacity.

Further, thermal modulation provides reproducible peak shapes and excellent sensitivity improvement relative to 1D-GC due to the focusing of analytes from the ^{1}D column. The resulting data produced generally does not require extensive data processing prior to chemometric analysis. However, along with a higher cost of operation, thermal modulators are not able to modulate light, highly volatile compounds (< C_5) unlike flow modulators.

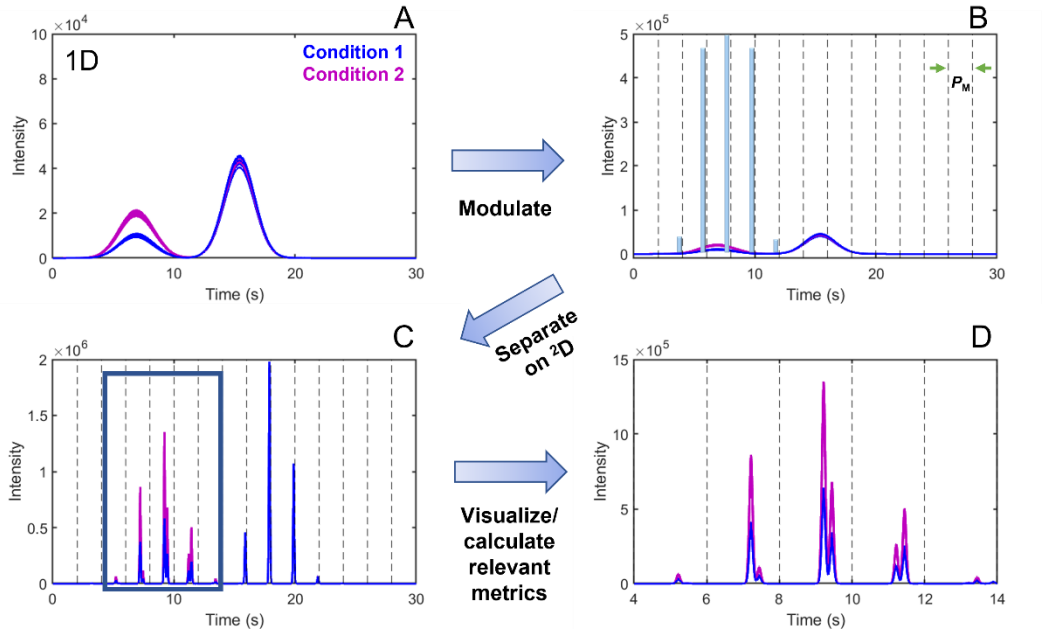


Figure 1.6. Visualization of the modulation process using thermal modulation. (A) Overlay of multiple replicates of 1D data of several simulated analytes. The ^{1}D effluent is modulated at intervals of 2 s (P_M=2 s), (B) with an illustration of the earliest eluting analyte in the blue chromatogram being preconcentrated (indicated in light blue) prior to separation on the ^{2}D column (C). (D) Finally, the separation of the analytes can be visualized further and relevant figures of merit can be calculated. The zoom-in provided in 1.6D indicates there are three analytes instead of the previously estimated two analytes in 1.6A.

Flow modulators, typically flow diversion or differential flow, use an auxiliary flow of carrier gas and re-direct, collect, or temporarily stop flow from the ^{1}D column onto the ^{2}D column [46,55–63]. These modulators are less expensive to implement than thermal modulators; however, not all flow modulators can achieve a 100% duty cycle and the high ^{2}D flow rates (^{2}F) required for some differential flow modulators to flush sample out of the sample loop onto the ^{2}D

column are not compatible with the pumping capacity of MS detectors. To address the high ^{2}F, one solution is to split flow to multiple detectors, but splitting flow will cause a decrease in sensitivity. Alternatively, significant progress has been made to ensure flow modulators can be operated at flow rates compatible with MS detectors [64–72]. Franchina et al. demonstrated a flow modulation technique operating with a ^{2}F of 4 mL/min with a quadrupole mass spectrometer (qMS), providing a 2.5-fold improvement in the *S/N* relative to 1D-GC-qMS [66]. Similarly, dynamic pressure gradient modulation (DPGM) was demonstrated using a ^{2}F of 4 mL/min with a time-of-flight mass spectrometer (TOFMS), providing a 100% duty cycle and a *S/N* improvement of 5-fold relative to GC-TOFMS [72]. A lower ^{2}F of 2 mL/min using a Deans' switch was highlighted by Ghosh et al.; however, sensitivity is lost due to the low duty cycle of this technique [71]. Additionally, the *S/N* improvement is mechanistically different for each type of modulation. During thermal modulation, analytes are cryogenically focused in the cold zones before "re-injection" on the second column using hot jets, focusing the band into a narrow pulse. During flow modulation, a higher flow rate on ^{2}D is used, increasing the analyte flux at the detector and thus the signal. Other concerns using flow modulation is the reproducibility of the ^{2}D peak shape for a given analyte [72,73] and under-sampling of the ^{1}D effluent reducing quantitative precision [74]. Both effects can negatively impact the trilinearity of the data and hinder the performance of chemometric techniques on the data set. However, the use of a high flow ratio and proper modulation ratio can mitigate these concerns.

Common MS detectors used for GC×GC instrumental platforms are the TOFMS [41,75,76], qMS [42,43,77], and HRMS [78–81], which are often HRTOF, Q-TOF, and Orbitrap mass spectrometers. The common nominal mass TOFMS detectors can achieve high collection rates (up to 500 Hz) with high sensitivity and selectivity, which is especially advantageous for

producing fast GC×GC separations with enough spectra collected across narrow peak widths. Alternatively, qMS operates at lower frequencies, but is less expensive to maintain while still being useful for a large range of applications. Additionally, qMS detectors scan at a rate dependent upon the scan cycle time, which includes the dwell time, or how long it scans each *m/z*, and the interscan delay, the time between successive scans of the *m/z* range. Thus, it is common to use non-integer scan rates and depending upon the P_M selected by the analyst, additional data processing may be required. HRMS detectors provide additional selectivity and sensitivity with high resolving power essentially improving the overall peak capacity of the instrumental system. Depending upon the experimental design for a given implementation, the analyst may use selected ion monitoring (SIM) for targeted studies, limiting the scan range to ions for specific analytes of interest, or scan across a broader range of *m/z* in full scan mode for non-targeted studies. While the high resolving power is especially advantageous for co-eluting analytes with similar mass spectra, HRMS data is incredibly information-rich requiring additional processing compared to nominal mass data

Regardless of the modulator and detector used, GC×GC-MS chromatograms can be visualized in similar ways. Naturally, since the MS detector records the analytical signal at every *m/z*, the signal from these *m/z* for a single chromatogram can be visualized in the vectorized format (Fig. 1.7A). The signals from all *m/z* can be summed together to generate the total ion current (TIC) chromatogram (Fig. 1.7B), which is the sum of all *m/z* signals. While the chromatograms are in this vectorized form, multiple chromatograms can be overlayed and compared to discover differences in signal between different classes (Fig. 1.7C). For visualizing the 2D separation and extracting relevant figures-of-merit, a single chromatogram can be cut along the modulation period (dashed lines in Fig. 1.7D) and folded into a 3D waterfall plot (Fig.

1.7E) or, more commonly, as a 2D contour plot (Fig. 1.7F). From the folded versions of the chromatogram (Figs. 1.7E-F), three fully resolved analytes can now be easily distinguished with the magnitude of the signal indicated by the color scale.

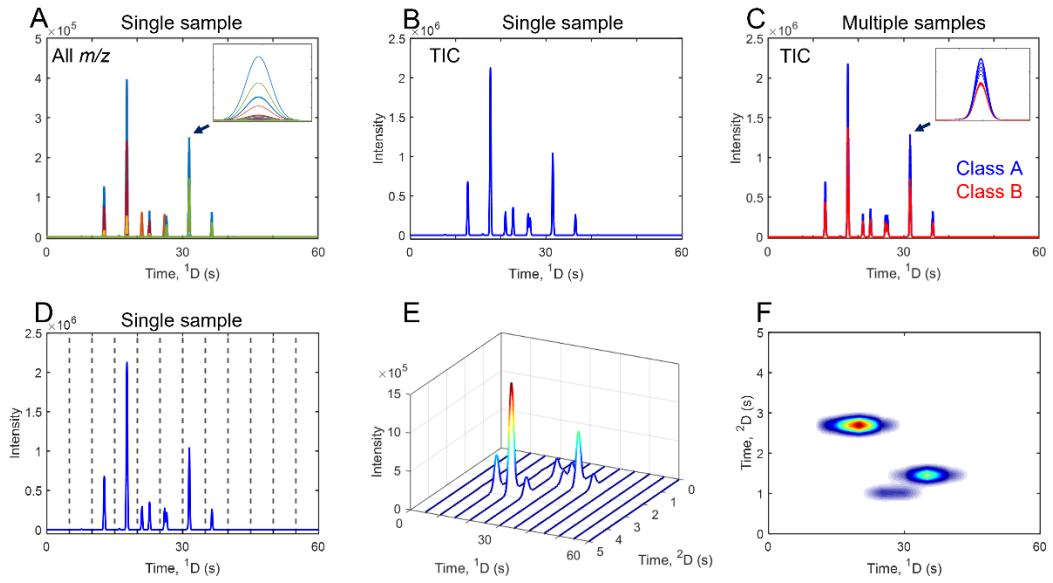


Figure 1.7. Visualization of GC×GC-MS data using simulated chromatograms for the separation of three analytes. Vectorized format of a single sample (A) with all *m/z* collected overlaid and (B) with the signal for all *m/z* summed together to form the total ion current (TIC) chromatogram. (C) Vectorized format of the TIC for all sample replicates in class A (blue) and class B (red), used to visualize similarities and differences of samples in comparative analyses. (D) The TIC chromatogram of a single sample can be cut and folded along the modulation period indicated by dashed lines in (D) to be viewed as a 2D plot including as a waterfall plot (E), or most commonly as a contour plot (F).

Prior to comparative analysis of GC×GC-MS chromatograms, several data processing steps are required to ensure successful chemometric analysis. These include baseline correction to remove low frequency detector noise and drift, which is achieved by simply subtracting a fitted function to the chromatogram or specified section of the chromatogram. For particularly noisy data, Savitsky-Golay smoothing may additionally be used to reduce noise and improve *S/N* [82]. Normalization and retention time alignment are critical processing steps to account for both sample and instrumental variation across replicates in comparative analyses. Normalization is typically accomplished either using standards (internal, external, or standard addition) or by

normalizing to the sum of the TIC chromatogram. To correct for slight run-to-run retention time shifting, several GC×GC-MS alignment algorithms exist such as 2D correlation optimized warping (2D-COW) [83,84], dynamic time warping (DTW) [85], and piecewise alignment among others [30,86–88]. Peak detection, identification, and integration are also important data processing steps, especially for the generation of accurate peak tables prior to statistical analysis (discussed in Section 21.2.3). Several peak detection approaches that distinguish analyte signal from the background noise include the enhanced TIC approach [89] and watershed-based algorithms commonly used in image analysis [90,91]. Identification may be achieved either matching analyte spectra to an in-house library or to standard open-source databases, such as NIST. For analyte quantification, the most common analyte integration approach for GC×GC is the two-step method, where the ^{2}D modulated peaks are summed together and the resulting area or peak height of the summed ^{2}D peak is calculated [92].

Selection of the mass spectrometer may also have implications on the data analysis workflow, such as using non-integer scan rates with qMS detection. Non-integer products of the scan rate and P_M may result in distortions of the chromatogram with shifts in the ^{2}D retention time of analytes. This issue may be addressed for a given non-integer scan rate by selecting a non-integer P_M such that the product between the two is an integer. While the analyst could modify the instrumental parameters to avoid missing data or non-integer acquisition rates that do not provide an integer product between the acquisition rate and modulation period [81], data processing methods can be used after data collection to correct for these concerns. Methods to address data collected with distortions due to the non-integer product of the scan rate and modulation period include interpolation and re-sampling the data [93]. Using HRMS detectors provides excellent selectivity relative to nominal mass detectors, however the data generated,

especially when multiple replicates are necessary, is very computationally expensive to analyze due to the size of the file. This requires powerful processing capabilities and some form of data reduction for simultaneous analysis of replicates. Furthermore, when considering a specific theoretical ion *m/z* collected per analyte, the experimental ion *m/z* may be slightly different for each sample replicate resulting in unique *m/z* combinations at each retention time. While the deviation from the theoretical *m/z* may not be large, many chemometric tools require that each sample have the equivalent *m/z* dimension across all classes. Methods to accomplish this prior to chemometric analysis include data reduction through binning [94,95], compiling peak tables, or discovering regions-of-interest (ROI) [96–101]. The first approach involves binning the data to a user-defined precision for non-targeted analysis followed by targeted analysis of specific sections of the chromatogram with significant features [95]. Binning quickly and effectively compresses the data improving *S/N*; however, it does not always ensure all sample chromatograms have equivalent *m/z* and the original selectivity from HRMS detection is lost. Compiling peak tables (discussed further in this section) reduces the data to chromatographic features and relies on accurate identification and quantification of all features. Accurate compilation of this peak table is challenged by low *S/N* analytes and retention time misalignment. Lastly, the ROI approach was initially developed for handling HRMS data from liquid chromatography (LC) separations, but has also been applied to comprehensive 2D-LC (LC×LC) and GC separations [96–98]. Here, ROIs on every *m/z* are extracted applying a user-defined signal threshold, allowable mass deviation, and minimum number of consecutive retention times [99–102]. The chromatogram is then reduced down to only include those ROIs [99–102]. This method is beneficial by preserving the mass accuracy from the detector and ensuring all samples have equivalent *m/z* dimension through sample augmentation steps. Hence,

comparative analysis methods can be readily applied to the compressed chromatographic data while maintaining the original fidelity of the data.

The analyst must also decide how to apply the intended analysis whether on each pixel, using a peak table, or in a tile-based manner where the chromatogram is binned into sections, which is decided depending upon the data structure, size, and quality (Figure 1.8). Pixel-based analysis (Figure 1.8 left panel) consist of calculating a statistical metric on each pixel of the 1D or 2D chromatogram [23,103–105]. While this ensures a comprehensive analysis of all the data maintaining its original fidelity, it is the most computationally expensive approach and requires careful pre-processing to ensure the number of false positives and false negatives are reduced. To reduce the computational load, the analyst may select specific *m/z* to use for analysis or use the TIC, however contributions of significant analytes may be lost by reducing the data used. The analyst must select the appropriate baseline correction and alignment procedures during pre-processing as this approach is particularly susceptible to retention time misalignment. With measurements compared across samples at each ${}^{1}t_{R}$ and ${}^{2}t_{R}$ pair, peak maxima may not ideally align at the same ${}^{1}t_{R}$ and ${}^{2}t_{R}$ for a given analyte across all sample chromatograms, resulting in a greater occurrence of false positives.

The second approach (Figure 1.8 middle panel) involves curating a peak table with an entry for each compound in all the samples that is detected above a specified *S/N* and identified by matching to a library database. Each identified compound is then quantified and the peak heights and/or areas are recorded for each entry. A list of peaks is first generated for each sample chromatogram and a resulting peak table is compiled by matching entries across samples. This relies on the accurate alignment, detection, identification, and quantification of each analyte across samples and classes. Peak tables can be created in programming platforms such as

MATLAB, Python, or R, or using commercial instrument software, which can then be exported. Following peak table curation, statistical metrics can be calculated using peak areas or heights from the table [80,106]. A benefit of this approach is that the data is reduced to just the chromatographic features, making statistical analysis less computationally expensive. However, there may be analytes in some samples within a class that do not pass the *S/N* threshold or may be missing from the sample. Recently, T.J. Davis et al. introduced an imputation-based method to account for such missingness in comparative analyses [107]. Using this approach, each analyte peak is accounted for once in the analysis. This is true if the peak table is created to accurately represent the data, which may be challenging if there is retention time misalignment or if the analyte has low *S/N*, making it difficult to detect. Similar to using pixel-based methods, pre-processing can be critical especially for aligning or smoothing to improve *S/N*.

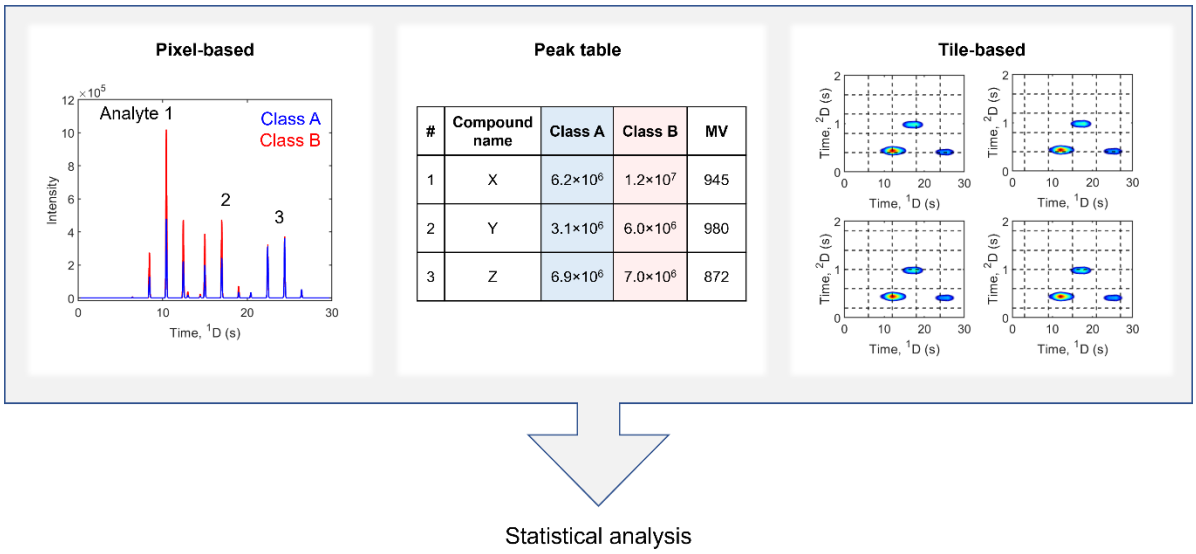


Figure 1.8. Illustration of pixel-based (left), peak table-based (middle), and tile-based analysis (right).

Finally, the last approach is the tile-based method (Figure 1.8 right panel) that was originated in the Synovec group, where the chromatogram is sectioned into rectangular tiles of user-defined dimensions and the area within the tile is summed, or binned, prior to statistical analysis [108–

110]. This method reduces the computational load by essentially reducing the data dimension by the size of the tile selected. Additionally, an appropriate tile size improves the *S/N* as well as mitigates issues with minor retention time misalignment. This is achieved by ensuring the tile size captures the entire peak width in both dimensions as well as the magnitude of retention time shift in each dimension if applicable. As seen in the right panel Figure 1.8, there are four grid schemes used for the tile-based approach seeing as there is not a single grid that ideally captures the area for each analyte peak in a single tile. Due to the multiple grids used, each analyte will be captured entirely and/or partially by multiple tiles, resulting in repeat entries of the same analyte in the compiled list of "hits" (redundant hits). This is addressed by using a redundant hit removal algorithm that removes any redundant analyte peak in the compiled list based upon retention time. In addition to redundant hits, the user must take into account the false positives or even interfering peak signals within a tile that obscure analytes that are less sensitively detected. To minimize both interfering analyte peaks as well as redundant hits, proper tile size selection is critical, with recent work using GC×GC chromatograms with simulated shifting reporting that a larger tile size is recommended [111].

1.2 Introduction to chemometric analysis

Chemometric methods mathematically and statistically extract the most relevant information that explain the chemical system. Such methods are typically categorized as either targeted, where the analytes of interest are known *a priori* and are the focus of the analysis, or non-targeted methods, where the analyst aims to extract as much chemical information as possible that explains the data set. Historically, targeted approaches were greatly used, tailoring experimental conditions to analytes of interest without considering the remainder of the sample. This may include an approach where a specific group of analytes of interest are targeted [112] or specific compound classes are the focus [113]. While targeted approaches may achieve the initial

goal of the analyst, a large amount of information about the samples remains unknown and unused. With advances in instrumentation and chemometric tools to analyze the large amount of chemical information embedded within the data sets obtained, non-targeted studies have become more utilized. Non-targeted methods aim to identify analytes that discriminate samples and/or sample classes using either an unsupervised or supervised approach, where the latter requires class membership to be known *a priori*. Additionally, a combination of non-targeted and targeted methods may also prove to be beneficial during analysis. For example, an unsupervised method may be used to cluster similar samples together, providing potential class membership that can then be used for later supervised analyses. In contrast, significant features discovered with non-targeted approaches can be identified and quantified using a targeted approach such as chemometric decomposition methods.

Compared to unsupervised methods, supervised techniques can be advantageous for comparative analyses if the target variable (i.e., class membership or sample property measurement) is known *a priori*. By leveraging information regarding the target variable, supervised methods ensure that significant features that truly discriminate or explain variances between classes are discovered. However, regardless of this *a priori* knowledge, it is important to note that the analytes of interest are not known *a priori,* and the goal is to discover the most significant information about the samples as possible. Supervised, non-targeted methods can be broken down into two groups: property prediction and feature selection. Property prediction refers to the development of a regression model that correlates the chromatographic information to an independently measured chemical or physical quantity. Conversely, feature selection techniques aim to discover peaks in the chromatographic data that correlate to class membership. Specifically for comparative analyses in GC×GC-MS studies, feature selection methods are

primarily used. Hence, these methods will be the focus of this section. The most common feature selection methods for GC×GC-MS comparative analyses include partial least squares-discriminant analysis (PLS-DA) and Fisher ratio (F-ratio) analysis. These approaches have been used for various applications, ranging from predictive modeling to discovering significant chemical changes between sample classes.

In this section, chemometric methods for GC×GC-MS data will be discussed, providing details of the performance and limitations of each approach. Unsupervised, non-targeted techniques commonly used with GC×GC data sets will first be discussed. These methods include principal component analysis (PCA), hierarchical clustering analysis (HCA), and variance rank initiated-unsupervised sample indexing (VRI-USI), though herein I will focus mainly on PCA. The following section will then focus on supervised, non-targeted comparative analysis techniques used for GC×GC data. These methods include PLS-DA, F-ratio analysis, and tile-based 1v1 analysis. Each of these methods will be discussed in the context of using the chromatographic at the pixel-level or tiled (binned) as well as using features from compiled peak tables. The use of targeted decomposition and related methods like multivariate curve resolution-alternating least squares (MCR-ALS) and parallel factor analysis (PARAFAC) to improve analyte identification and quantification within a comparative analysis workflow will also be discussed.

1.2.1 Principal component analysis (PCA)

When PCA is applied as an unsupervised, non-targeted method for initial exploratory analysis of data, no *a priori* knowledge (identity of meaningful analytes or class membership) of the samples is necessary. It is a pattern recognition technique that is used for data reduction, essentially reducing the data to linear combinations of latent variables, or principal components

(PCs), projected onto a lower dimension space[114,115]. These PCs capture the greatest amount of variation between and among all samples, or the covariance. Typically, the first few PCs capture the most variance that explains the data with the first PC (PC 1) capturing the most variance, followed by PC 2, and so on. The extent to which the original data is reduced depends upon the number of PCs selected, which can be determined using a scree plot. The scree plot indicates the amount of variance that each PC explains, thus the analyst can determine the cutoff where the added explained variance is not significant.

GC×GC-MS data is loaded to PCA in the vectorized format such that each row contains samples and the columns contain variables. The analyst may decide to apply PCA in a pixel-based fashion meaning the entire chromatogram is used such as in Figure 1.9 [23,116], or reduce the computational load by using peak areas or heights from a table [70], using a subset of features or analyte peaks that are significant, or binning the data [117]. Further, the computational load may be reduced to a greater extent if only the TIC or selective *m/z* are used, which can be concatenated prior to analysis [116], or selecting regions of the chromatogram. When using the pixel-based approach, selecting the TIC or selective *m/z* is especially important when working with many samples to reduce the computational load, making the use of binning (tile-based) or peak tables an attractive alternative. Additionally, pre-processing is crucial to ensure successful and meaningful PCA results. Specifically, normalization, baseline correction, alignment, and mean-centering should be considered when applying PCA. Especially for pixel-based approaches, proper alignment is critical to ensure the model is accurately representing the equivalent analyte peaks across samples. This can be mitigated using accurately curated peak tables or binning the data such that any slight retention time alignment is captured within the bin.

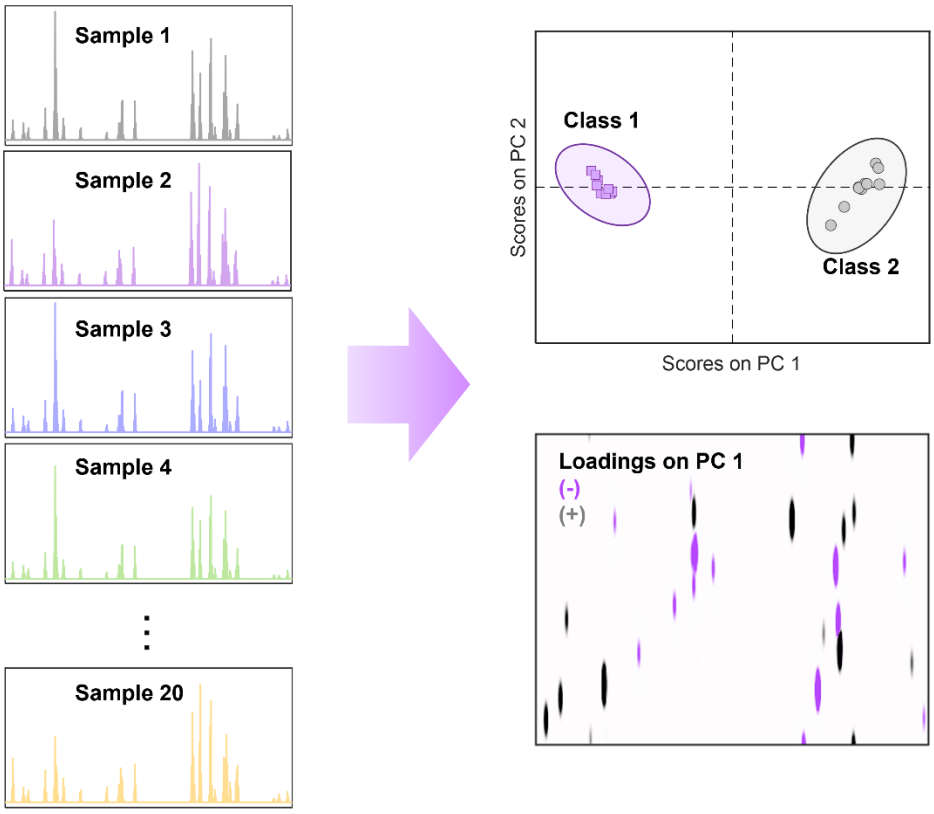


Figure 1.9. Illustration of PCA on 20 samples of unknown sample class membership, resulting in the clustering of samples into two distinct classes in the top right scores plot. The loadings on PC 1 in the bottom right scores plot indicate that analytes in purple have a higher signal in class 1, while analytes in dark gray have a higher signal in class 2.

The output from PCA are scores and loadings that aim to capture the greatest variance that explains the data. The scores hold the information of how the samples relate to each other based on the variance captured in each PC. Thus, visualization of the scores plot using PC 1 and PC 2 (PCs that capture the most variance) shows samples with the most similar chemical composition cluster into distinct groups in the PC space. The degree to which sample clusters differ from each other in the PCA scores plot can be measured, with common methods being the degree-of-class separation (DCS) [117], Mahalanobis distance [118], or confidence ellipses [119]. The loadings that correspond to each PC score contain the information of the variables or significant chemical components among the samples that contribute to the observed clustering. Variables or peaks

that are positive loaded indicate these peaks have higher signal in samples with positive scores, and the same trend is observed for negatively loaded peaks indicating higher signal in samples with negative scores. This method is excellent for capturing both differences in chemical composition between sample classes as well as accounting for differences in concentration of analytes in common between classes. The analyst may also opt to use PCA in a supervised manner for data visualization after feature selection is performed. One of the previously mentioned metrics for determining the degree to which classes cluster may then be calculated for the clusters of known sample class membership and be used as a metric for evaluating the performance of the feature selection approach.

1.2.2 Partial least squares-discriminant analysis (PLS-DA)

PLS-DA is a widely used supervised, non-targeted classification tool for data reduction and discriminant analysis of multi-class data [120,121]. While similar to PCA, PLS-DA benefits from prior knowledge of sample class membership to distinguish the variability between individual classes and the variance among all classes. Mathematically, PLS-DA aims to correlate the chromatographic data (**X** matrix) to class membership (**Y** matrix) by first reducing the dimensionality and then constructing the prediction model. The chromatographic data in the **X** matrix can be the entire chromatogram at the pixel-level [122], [122], binned chromatograms, peak tables, or features discovered from another non-targeted methods (i.e., F-ratio or 1v1 analyses) [123]. Note, pixel-based PLS-DA can be computationally expensive, especially if all collected *m/z* are used. Hence, to reduce the computational load, a subset of *m/z* or the TIC may be used at the pixel level after proper preprocessing. Otherwise, the use of either binned data, peak tables, or feature selection methods is preferred, especially with HRMS data. An illustration of PLS-DA at the pixel-level on the X matrix containing TIC chromatograms of two classes

(blue chromatograms belong to class 1 and red chromatograms belong to class 2) is provided in Figure 1.10.

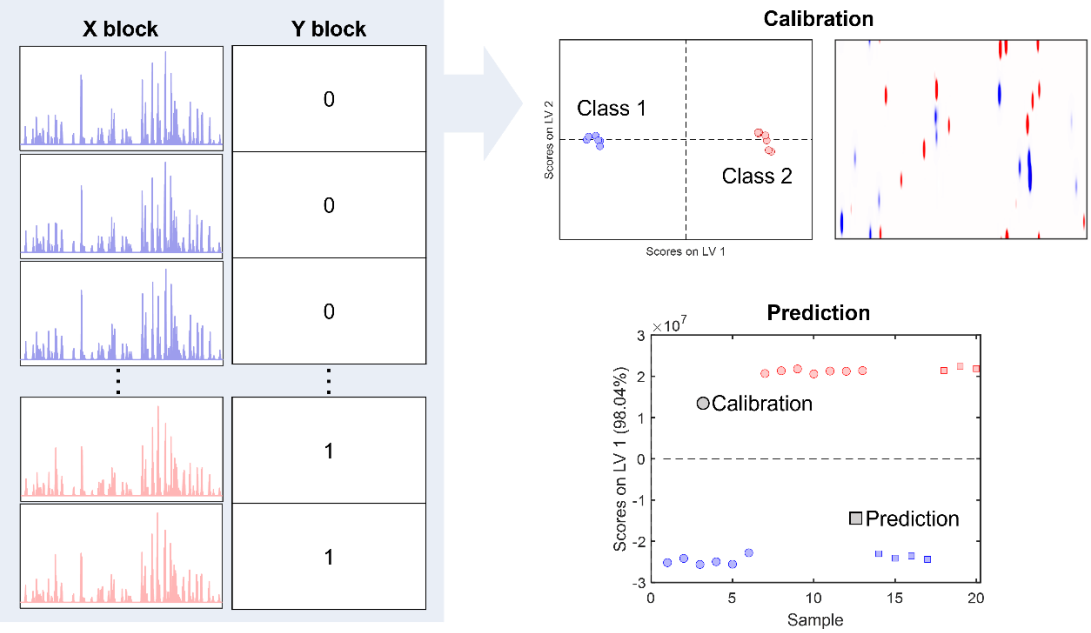


Figure 1.10. Illustration of PLS-DA applied to 20 samples (13 for calibration in circles and 7 for prediction in squares) at the pixel-level. A subset of chromatograms (X block) with assigned class membership (Y block) are used for model calibration (top right). Subsequently, the model is used for prediction of the remaining sample chromatograms (bottom right).

The input data matrix (**X**), used as a predictor, contains the samples in each row and the number of features, pixels, or bins in each column. Since PLS-DA is a supervised method, class membership must be assigned in the response vector, **Y.** Generally, sample membership is categorical (e.g., experimental or control), and therefore, each sample needs to be coded to a discrete numerical value (e.g., experimental = 0 and control = 1). The user must define these two matrices along with the number of components to model and any processing options (i.e., mean-centering). PLS-DA then reduces the dimensionality of the data by defining the latent variables (LVs) that maximizes the covariance between **X** and **Y**. To define, LVs are a combination of collinear variables measured to describe certain properties of the samples collected, like PCs for PCA. Similar to PCA, the outcomes are loadings and scores vectors (Figure 1.10 top right

calibration plots) corresponding to each LV. The **X** block scores includes the coordinates of the samples after projection; thus, when the scores plot for LV 2 versus LV 1 is viewed, the classification of the samples is visible. When using pixels or bins, the loading vectors can be essentially folded into a GC×GC chromatogram for viewing the contribution of each pixel or bin to the respective LV (positively correlated, negatively correlated, or neither). When using the peak table approach, the loadings vector can be viewed without folding to examine the correlation of each feature to the respective LV. Using the scores and loadings plots, the analyst can determine which analytes contribute the most to sample classification. For example, Rosso et al. demonstrated the use of PLS-DA on peak table features for discriminating between different hazelnut cultivars, with clear clustering of the classes in the scores plot [124]. Finally, the model produced with the training set of known class membership can then be used to predict samples with unknown class membership. After the model is built, variable importance in projection (VIP) scores or selectivity ratios can be calculated for each variable for determining variables that most contributed to the discrimination between classes [125,126].

Proper selection of the number of LVs and model validation is critical to prevent overfitting [127]. Internal validation is often accomplished using leave-one-out cross-validation (LOOCV) or Venetian blinds cross-validation. LOOCV is especially useful for data sets where a small number of samples are used to build the classification model. This method constructs a PLS-DA model after leaving one of the samples out of model generation. The sample that was excluded is then analyzed/predicted by the model and this procedure is repeated in a round-robin fashion for every sample in the data set. On the other hand, Venetian blinds cross-validation is appropriate for larger data sets, where LOOCV would result in longer computational times and model overfitting. Venetian blinds cross-validation is similar to the procedure for LOOCV, except a

subset of samples are excluded during model generation and used for prediction. The root-mean-square-error of cross-validation (RMSECV) is calculated following the cross-validation procedure. The RMSECV acts as a "goodness-of-fit" measure and is used for proper selection of the number of LVs to include in the final PLS-DA model. In comparison, when the analyst has enough samples to divide them into a calibration/training set and a test set that is representative of the entire sample set, they may select to use external calibration. This is achieved by using the training set to generate the model that is then used to predict samples in the test set for determining model performance. Paiva and Hantao provided exemplary work for demonstrating the advantages of both internal cross-validation and external validation procedures [122], where a pixel-based PLS-DA model was developed to correlate the chemical composition of lager beers to their consumer preference.

1.2.3 Fisher ratio (F-ratio) analysis

F-ratio analysis is a supervised, discovery-based method that aims to comprehensively discover chromatographic features that distinguish between known sample classes. This method is based on the analysis of variance (ANOVA) statistical test. Here, the standard F-ratio calculation is defined as the ratio of the between class variance (s_{bc}^2) to the sum of the pooled within-class variance (s_{wc}^2):

$$F - ratio = \frac{s_{bc}^2}{\sum s_{wc}^2} \quad (1.12).$$

This method prioritizes class-distinguishing ability rather than absolute analyte signal, allowing for the discovery of minor, yet significant, chemical differences between samples. Hence, the magnitude of the F-ratio indicative of the extent to which the "hit" (i.e., a chromatographic feature) discriminates between the sample classes. For a hit to be classified as class-distinguishing, it must have a significant difference (p-value < 0.05) between classes. The result

of F-ratio analysis is a list of features (i.e., a “hit list”) ranked in descending order of F-ratios. To discover these class-distinguishing analytes, the analyst then mines the hit list by identifying, quantifying, and determining statistical significance for each hit. However, the presence of false positives (i.e., statistically insignificant features with a high F-ratio) in the hit list may hinder data mining efforts by prematurely triggering a stopping point. Additionally, interfering analyte signals at the same *m/z* as a class-distinguishing analyte may decrease its discoverability or result in a false negative. With nominal mass detection, the likelihood of interfering signals from other analytes at the same *m/z* is higher than with HRMS detectors, relying to a greater extent on a higher $R_{S,2D}$ from the chromatographic separation. Since F-ratios are calculated on a per-*m/z* basis, using nominal MS detectors requires greater attention to the application of F-ratio input parameters, whereas the selectivity of the HRMS detector can be leveraged to ensure pure *m/z* are used (no interfering analyte signals) to obtain accurate F-ratios. To ensure the most accurate final list of significant features is obtained, a user-defined *S/N* threshold of 10 is typically ideal for being right at the limit of feature discovery [128,129]. Additionally, it is advised that the hit list be ranked according to the highest F-ratio *m/z* per-hit for improved discovery of trace level analytes [130]. Furthermore, the analyst must carefully consider their analysis approach (i.e., pixel-, peak table-, or tile-based) to mitigate these false positives and the computational load.

The initial approach to F-ratio analysis was pixel-based, which involves calculating F-ratios at each retention time in the data set for all *m/z* above a given *S/N* [23,86,105,131,132]. While this approach provides the benefit of using the entire data set, it introduces many redundant hits into the final hit list. Redundant hits are defined as replicate entries of the same analyte as separate hits in the final hit list. In turn, the presence of these redundant hits generates lengthy hit lists that would require tedious manual inspection to remove their presence. Additionally, this

method is sensitive to retention time misalignment across replicates, which could result in false positive F-ratios that are not truly representative of the analytes of interest. Alignment algorithms have been used widely for correcting retention time shifting across replicates and can be used as a preprocessing step prior to pixel-based F-ratio analysis [83,85,133–136]. Even with proper alignment, pixel-based F-ratio analysis is also sensitive to effects from modulation-based phasing and detector fluctuations [108].

Implementation of F-ratio analysis on peak tables can be successful in mitigating issues stemming from retention time misalignment and low *S/N* peaks affecting quantification [41,137,138]. Furthermore, using a peak table ensures that one F-ratio is calculated per feature therefore redundant hits are not an issue after generating the F-ratio hit list. This is particularly attractive for analysts that use commercial software to find peaks and build peak tables and then export the peak table prior to statistical analysis. An excellent example of this approach is provided by Dubois et al. for the characterization of decomposition odor from adipocere and soil samples at a death scene by GC×GC-HRTOFMS for supporting the use of decomposition odor as forensic evidence in court [139]. F-ratios were calculated on the peak tables obtained by GC Image software (Zoex Corporation), and for features with F-ratio higher than the F-critical threshold the responses from the different compound classes were quantified for each class. Additionally, this study highlights the use of feature selection prior to an unsupervised non-targeted method, hierarchical clustering analysis (HCA), for visualization of sample class clustering and compound class correlation.

Finally, F-ratio analysis has evolved to the current tile-based F-ratio analysis approach, illustrated in Figure 1.11. Tile-based F-ratio analysis involves sectioning the entire chromatogram into equivalent tiles of user-defined dimensions and calculating the F-ratios on the

area captured in each tile at every *m/z* [108,140]. Proper selection of both tile size and cluster window is critical for the success of F-ratio analysis in discovering all true positives and minimizing the redundant hits, especially for data with even the slightest run-to-run retention time shifting. The tile size on ^{1}D and ^{2}D should be selected based on the average width at base of the analytes and account for the magnitude of retention time shifting in ^{1}D and ^{2}D, respectively. A tile size too large may result in the loss of true positives discovered, while a tile size too small would introduce many redundant hits. Additionally, tile-based F-ratio analysis was demonstrated to be robust to simulated GC×GC-TOFMS data with shifting, concluding that a tile size too large is more beneficial than one too small [141]. The cluster window is generally ~60% of the size of the tile, with larger cluster windows reducing the amount of redundant hits in the list while possibly clustering away true positives with lower F-ratios. Conversely, while smaller cluster windows may retain more true positives in the hit list, more redundant hits are retained as well.

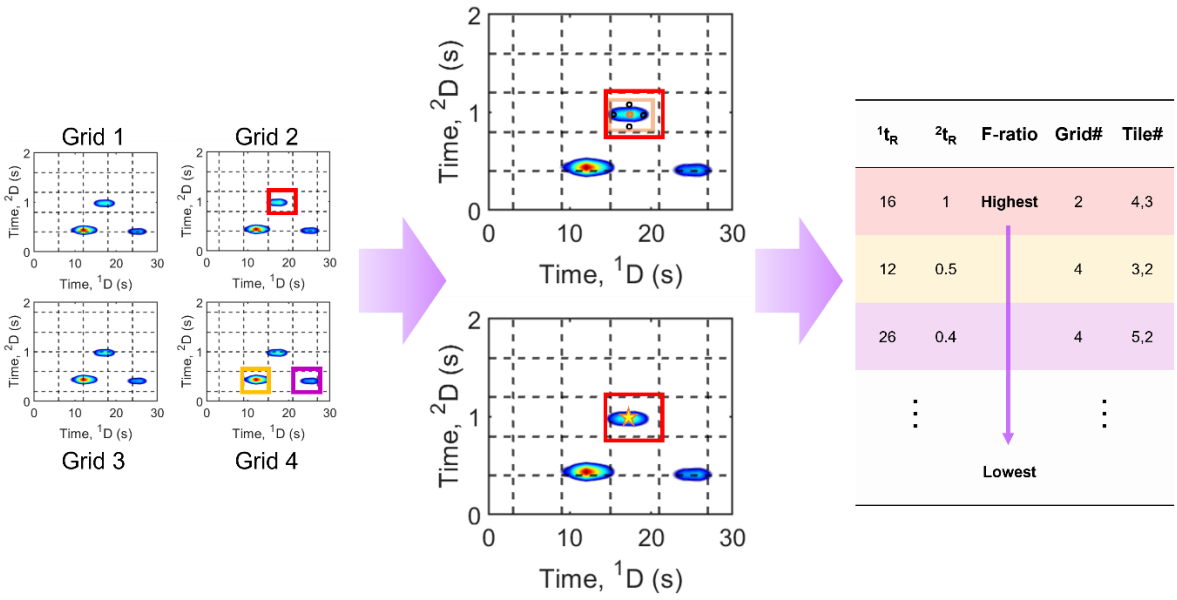


Figure 1.11. Illustration of tile-based F-ratio analysis applying the four grid schemes to pre-processed folded GC×GC data at each *m/z*, with the grid that best captures each analyte area highlighted by a red, yellow, and purple box. The next step, using the analyte in red, involves reducing the number of F-ratios calculated per analyte to one entry (pinning and clustering), and finally the remaining "hits" are compiled into a "hit list" ranked by descending F-ratio.

Even with proper mitigation of false positives using an appropriate F-ratio analysis approach, discovery of significant class-distinguishing analytes by F-ratio analysis can be challenging for multi-class metabolomics data with inherent within-class variance up to ~50%. Therefore, modifications to the F-ratio calculation shown in Eq. 1.12 have also been explored [142–144]. For instance, control-normalized F-ratio analysis uses only the within-class variance from the control group in the denominator of Eq. 1.12 [142,143]. Tile-based control-normalized F-ratio analysis was demonstrated for discovering biomarkers of knee ligament injury, with results highlighting the utility of control-normalized analysis for comparative analysis with non-normal distributions. Furthermore, for data sets with inherent large within-class variance in all sample classes, the F-ratios can be calculated using the lowest within-class variance among all classes for discovering the most true positives, termed minimum variance optimized (MVO) F-ratio analysis [144], which will be introduced in Chapter 6.

To address the sizable hit lists regardless of the approach (pixel-, peak table, or tile-based), initial F-ratio thresholds were applied to trim the hit list while monitoring the extent to which classes were distinguished applying PCA to the remaining features in the hit list [86,131]. Subsequently, combinatorial null distribution analysis was utilized as a statistical means to determine an appropriate F-ratio threshold at a given null probability and confidence level [140,145,146]. Recently, Sudol et al. proposed using the *p*-value, either from a *t*-test or one-way ANOVA, as a means for determining the F-ratio threshold [147]. Here, the F-ratio threshold was set equal to the F-ratio for the last true positive (p-value < 0.01) [147]. Features determined to be true positives (i.e., class-distinguishing analytes) can then be further studied using targeted approaches for accurate identification and quantification and related back to specific characteristics or properties of the samples.

1.2.4 Tile-based 1v1 analysis

While the F-ratio analysis requires multiple samples per class to calculate variances, the analyst may not always have the resources or opportunity to collect multiple chromatograms per-class. Hence, the development of pairwise analysis approaches can also be beneficial for feature discovery. Standard methods that have been used for pairwise analyses include Jensen-Shannon (JS) divergence and subtraction plots to discover differences between samples. JS divergence is a measure of dissimilarity and uses Bayesian statistics to determine the probability that a specific region of the chromatogram is different between classes. Barcaru et al. demonstrate using JS divergence for the analysis of a mixture of diesel and lubrication oil [148]. On the other hand, subtraction plots are generated by simply subtracting the chromatogram of one class from the other class. The remaining peaks in this subtraction plot therefore correspond to either differences in composition between chromatograms or differences in abundance. However, both standard metrics are typically implemented in a pixel-based fashion and thus, are susceptible to false positives. Therefore, Cain et al. developed a tile-based pairwise approach, termed 1v1 analysis, to discover chemical differences between classes [149].

The novel tile-based 1v1 approach utilizes the same tiling methodology described for tile-based F-ratio analysis (Section 1.2.3); however, instead of calculating variances, a rank metric (*RM*) is calculated for each tile at every *m/z*. The rank metric is defined as

$$RM = \frac{|S(m/z)_2 - S(m/z)_1|}{S(m/z)_2 + S(m/z)_1} \times 100 \tag{1.13}$$

where $s(m/z)_n$ is the summed signal within a tile at a given *m/z* for class *n*, either class 1 or class 2. The magnitude of the rank metric is representative of the extent to which the discovered analyte discriminates between the classes. Additionally, the calculation of the rank metric ensures the ability of the analyte to distinguish between classes is prioritized rather than the

absolute signal. While this method has been demonstrated solely using the tile-based approach, the rank metric could be easily calculated at every pixel of the chromatograms or on the peak areas or heights compiled into a peak table.

Tile-based 1v1 analysis was evaluated on a diesel fuel spiked with 18 compounds and demonstrated to discover biomarkers of moisture damage in cacao beans [149]. Results indicated that 1v1 analysis outperforms all other pairwise methods (JS divergence and subtraction plots) given its near ideal ranking of the hit list. Furthermore, tile-based 1v1 and F-ratio analysis results are essentially identical, indicating that the same meaningful chemical differences between classes can be discovered whether the analyst collects only one chromatogram per class or multiple replicates. Furthermore, this method was demonstrated in conjunction with the novel targeted method class comparison enabled mass spectrum purification (CCE-MSP) [150] to improve analyte identification.

1.2.5 Targeted decomposition methods

The major analytical goal is to identify and quantify analytes of interest in samples and in targeted studies the analyte identity is known *a priori*. Accurate identification and quantification of analytes relies on pure spectra and resolved peak profiles. To quantify analytes, the two-step method is commonly employed for two-dimensional chromatographic data using either peak heights or areas [92]. Ideally all analytes in the sample are resolved during the GC×GC separations providing improved peak capacity over 1D separations and making quantification and identification much easier, however fully resolving all analytes in complex mixtures is not likely due to the high complexity of the samples [18]. While HRMS data has additional selectivity and sensitivity for more accurate identification and quantification using pure *m/z*, chemometric tools may still be used to aide with unresolved analytes and improve upon analyte

S/N. The overlapped analyte signals pose a challenge for accurate quantification of target analytes while mass spectra with interferent fragment ions make accurate identification a challenge, thus requiring chemometric tools to aid in decomposition of compounds. For these unresolved analytes, MCR-ALS and PARAFAC are excellent decomposition (often referred to as deconvolution) methods to model the section as a linear combination of the analyte signals. All dimensions of the data are leveraged by chemometric tools to fully resolve analytes and obtain pure peak profiles and spectra that would otherwise be lost to the low 2D resolution. Overlapped analytes with similar spectra rely on decomposition that leverages the difference in retention time from the chromatographic separation, while analytes that have 2D resolution approaching zero rely on differences in spectra to provide pure signals. The experimental design for collecting the data is critical for ensuring the bilinear or trilinear structure of the data. While a non-trilinear data structure is not ideal for chemometric tools such as PARAFAC [151], PARAFAC2 [152,153] has less strict data structure requirements and MCR-ALS is capable of decomposing analyte profiles and spectra with sufficiently bilinear data structure as it relies more on the spectral dimension [154]. These chemometric techniques offer the advantage that the analyst is not required to know the identity of analytes prior to decomposition.

Data that has a sufficiently bilinear data structure is ideal for decomposition by MCR-ALS [155–157], including third order GC×GC [158,159] with a multivariate detector and concatenated sample chromatograms of univariate detectors (GC×GC-FID). MCR-ALS can handle this data as it is less sensitive to differences in peak shapes of modulated peaks and slight retention time shifts. If the data set does not have a sufficiently bilinear data structure, alignment algorithms may be used to address retention time shifting [31,83,84]. This is especially important when using a univariate detector rather than a multivariate detector that can leverage the analyte

spectrum for decomposition. When considering third order data, generally the chromatographic dimension can be reduced prior to MCR-ALS either by summing away in one dimension or by unfolding the data and leveraging both ^{1}D profiles and modulated ^{2}D peaks in one dimension, as illustrated in Figure 1.12. As could be expected, using MCR-ALS, or any decomposition method, on the entire chromatogram is computationally expensive. To use MCR-ALS more efficiently, the analyst must select the region of the chromatogram to decompose and enter an initial estimate of the number of components in the chromatographic region that the algorithm must model. These decomposition methods require minimal pre-processing of the data as they remove background and noise contributions, providing baseline corrected vectors of the analyte signals. Typically, this number of components should be estimated including an additional component than the number of analytes present, as the algorithm models the noise as one component, removing noise from the loadings, or the pure elution profiles of the analytes and improving the signal-to-noise ratio. MCR-ALS was initially described for second order data and can be represented in the following expression:

$$X=CS^{T} + E \quad (1.14),$$

where for comprehensive two-dimensional chromatographic data using a univariate detector X is the raw data matrix, C is the matrix containing the one chromatographic dimension with the resolved elution profiles for each component modeled, S is the matrix containing the second chromatographic dimension elution profiles associated with C, and E is an error matrix containing the unexplained modeled data [160]. For third order data, the loadings depend on the data structure MCR-ALS is applied to. If the two chromatographic separations are unfolded into one dimension and the other dimension contains the spectral information, the resulting loadings would be pure elution profiles with both ^{1}D and ^{2}D peaks in one vector per component and the

other loadings would include the corresponding spectrum for each component modeled (Figure 1.12 right panels). This can be extended as follows

$$X_{aug} = C_{aug}S^T + E_{aug} \quad (1.15),$$

where X_{aug}, C_{aug}, and E_{aug} are column-wise augmented matrices containing multiple replicates [160], which is especially considered for non-trilinear data and to reduce rotational ambiguities. The algorithm uses initial estimates of each dimension to test for convergence and iterates the values alternating between dimensions until convergence is met. The ease with which the algorithm can decompose each component is dependent on adequate signal-to-noise ratio. The resolved elution profiles retain concentration information from the sample and may then be used to quantify analytes, while decomposed spectra are valuable for more accurate identification. The alternating least squares will improve the fit of the model with each iteration, however the amount of time it takes for the solution to be determined is dependent on the number of components selected to be modeled. Additionally, selecting the correct number of components to model is important for obtaining true representations of the analytes of interest. For example, selecting too few components to model by MCR-ALS will result in loading vectors comprised of two inseparable analyte signals and convoluted spectra. When the appropriate number of components are selected, the resulting loadings provide pure elution profiles that can be used to quantify analytes in the selected region. Decomposed spectra are also provided that can be used in a non-targeted fashion to tentatively identify analytes or in a targeted fashion to confirm identification by matching to a library database.

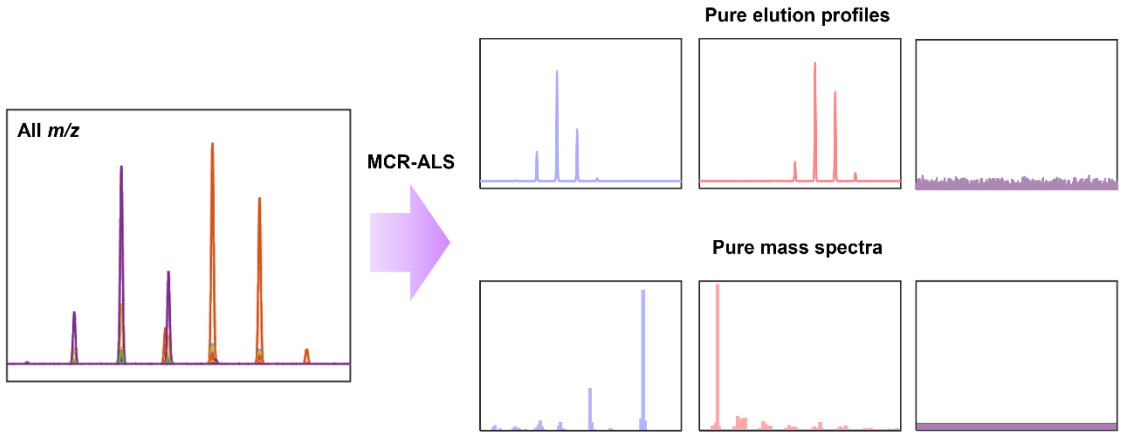


Figure 1.12. Demonstration of MCR-ALS on GC×GC-MS data, modeling three components: two analytes and the background noise. Outputs are pure elution profiles for each modeled component and pure mass spectra.

For optimal decomposition results, MCR-ALS initialization, constraints, and rotational ambiguities must be considered. Several methods for initialization are proposed including initial estimates of the concentration profile or the spectra, SIMPle-to-use Interactive Self-modeling Mixture Analysis (SIMPLISMA) for obtaining pure variables [161], and Iterative Key Set Factor Analysis (IKSFA), based upon KSFA, which determines a set of spectra of the minimum number of components to model that are the most unlike, or orthogonal, to the raw matrix spectra [162]. Next, constraints are selected to reduce the possible range of solutions. The most common applied constraint is non-negativity, ensuring the elution profiles have non-negative concentrations. Other available constraints include unimodality, closure, selectivity of pure spectra, and selectivity of concentration profiles if location of analytes of interest and pure spectra are known *a priori*. While these constraints and proper initialization reduces the number of possible solutions to the most appropriate, one remaining concern is the rotational ambiguities, or the different solutions that may be produced for the same matrix input and fit the data equally well. This issue can be evaluated using MCR-BANDS [163] and can be addressed by local rank and selective constraints, concatenating replicates prior to decomposition, or using

hard modeling to obtain unique solutions. Additionally, a trilinearity constraint may be used to obtain essentially the same unique solution as would be obtained by PARAFAC [164]. The analyst must also consider that decomposition may be challenged by interfering peaks that overlap with the analyte of interest on multiple dimensions. This is especially challenging when the interfering peak signal is much larger than the analyte of interest.

Specifically for HRMS data sets, the data requires processing prior to MCR to compress the data as they are information-rich and are computationally expensive. One method involves binning the spectral dimension to lower resolution, determining interesting regions to use MCR-ALS on, and then using the high resolution data for further analysis [95]. This method does ensure that each scan and each replicate share the same *m/z* dimension depending on the level of binning at the cost of losing selectivity. The second method, termed ROIMCR, involves first compressing the data by finding regions of interest (ROI), or simply chromatographic peaks, in a non-targeted fashion while conserving the high resolution mass spectral dimension [96,99,100,165]. Next, segments of the ROI data matrix may be input to MCR-ALS to decompose and obtain pure elution profiles and mass spectra for subsequent identification and quantification. The ROIMCR method has been evaluated with LC-HRMS and LC×LC-HRMS data, though it can easily be extended to GC×GC-HRMS data sets. Whether the data was binned or the ROIMCR software was used, the MCR-ALS loadings obtained for all types of data can be subsequently used for accurate identification and quantification as well as other supervised data analyses, such as PCA and PLS-DA [166,167]. Recently, the ROIMCR method was used to compress LC×LC-MS and LC×LC-UV data prior to evaluating the bilinearity and trilinearity of the data structure with MCR-ALS and PARAFAC, respectively, with results suggesting MCR-

ALS was more appropriate tool as the data structure deviated from an ideal trilinear behavior [168].

For data with a sufficiently trilinear data structure, PARAFAC [152,169,170] is an excellent chemometric tool to facilitate the identification and quantification of analytes leveraging all three dimensions of GC×GC data sets [105,171]. Prior to PARAFAC decomposition, the data must be a 3-dimensional matrix, typically a combination of ${}^{1}t_{R}$, ${}^{2}t_{R}$, and spectral dimension of a single sample [172] or concatenated samples [173]. As mentioned previously, using the entire chromatogram is computationally expensive, however there are methods that may be used to apply PARAFAC to large sections of data [171]. The number of components to be modeled must be selected prior to PARAFAC decomposition taking into consideration the noise being modeled as one component. PARAFAC operates in a similar manner to MCR-ALS, producing three loading vectors including peak profiles that retain the concentration information of each modeled analyte and the corresponding mass spectrum. The PARAFAC model for a three-way chromatogram ($\underline{\mathbf{X}}$) and F number of components can be expressed as

$$\underline{\boldsymbol{X}} = \sum_{f=1}^{F} \boldsymbol{a}_f \otimes \boldsymbol{b}_f \otimes \boldsymbol{c}_f + \underline{\boldsymbol{E}} \qquad (1.16),$$

where **a**, **b**, and **c** are the purified loadings in each data dimension for each component and $\underline{\mathbf{E}}$ is the residual error. [169]. The model is achieved by using initial estimates for two of the modes and then applying alternating least squares to fit the remaining mode to obtain the solution where the residuals are minimized. Usually, random initial estimates from trilinear decomposition are used for initialization and minimizing residuals is the stopping point, with the fit of the model improving with each iteration. An advantage of PARAFAC is the uniqueness of the model. While rotational ambiguities are a main concern for MCR-ALS models, PARAFAC models cannot be rotated without altering the fit, so this is an advantage for PARAFAC.

The loadings matrices contain the vectors corresponding to the peak profiles and mass spectra of each modeled component. For a raw data matrix (m×n×l) with m containing the ^{2}D separation, n containing the ^{1}D separation, and l containing the *m/z* dimension, the PARAFAC loadings obtained will contain the pure ^{1}D peak profiles that are normalized to the ^{1}D peak profile with the greatest signal, pure ^{2}D peak profiles with signal reflecting the concentration of the component in the sample, and finally the pure mass spectra normalized to the *m/z* with greatest intensity. Reproducing the peak profile for one component can then be achieved by the outer product of the vectors corresponding to the component of each loading matrix. The peak profiles generated from this decomposition method are more trilinear than the raw data as the algorithm ensures the modulated peaks have the same retention time and more consistent peak shapes when modeled. Similar to MCR-ALS, the pure analyte loadings from PARAFAC can be used for identification, quantification, or other analyses such as evaluating alignment algorithms that may have been used to improve data trilinearity [85,174,175].

The trilinear structure condition requires that the data be of high quality with modulated peaks for a single analyte retaining the same peak shape and having the same ^{2}D retention time (^2t_R), or at least that it does not deviate greatly [176]. The magnitude of the retention time shifting between modulations as well as peak width on the second dimension (^2W_b) are the main considerations for the trilinearity deviation ratio (TDR) metric for evaluating the accuracy of PARAFAC models for quantification [176]. This metric was demonstrated for PARAFAC applied to GC×GC-TOFMS data due to the negative quantification bias that is due to nontrilinear data, though it may be extended to other chromatographic methods and/or chemometric methods such as MCR-ALS. Data with shifting retention times, mainly due to the temperature program in GC, can easily be corrected with alignment algorithms [31,83,84,134] during data pre-processing

steps, improving the trilinearity of the data set and making the data more amenable to PARAFAC. As PARAFAC has more strict requirements, it may additionally be used as a tool to evaluate the quality of third order data produced [63,168], however third order data that does not meet this requirement may employ PARAFAC2 to still leverage all three dimensions of the data [152,170]. PARAFAC2 is less sensitive to retention time shifts of the modulated ^{2}D peaks that may affect the trilinear data structure, though is still sensitive to differences in peak shapes between modulations

Similar to MCR-ALS, the performance of PARAFAC to obtain the correct model of the data is dependent on the selection of the correct number of components to model and appropriate signal-to-noise ratio of the raw data [177]. Examining the raw data generally gives an indication of how many components should be modeled taking into consideration that the noise will be modeled as well. In addition to visual inspection of the raw data, split-half experiments can be used to determine the correct number of components. In split-half experiments the section to be decomposed is split in half and a PARAFAC model is made for both halves [178]. If the number of components is selected correctly, the same loadings should be evident in the models of both data sets. Increasing the number of components to be modeled may slow down the algorithm and result in more noise being modeled as well as true analytes to be modeled by multiple components, or splitting. When too few components are selected to be modeled, several true analytes will be modeled by a single component and the goal to decompose the analytes is not accomplished. The initialization step ensures the algorithm reaches a solution quicker and that it helps make sure the algorithm does not get stuck on one iteration of the model. Initialization for PARAFAC decomposition typically involves either random values or initial estimates from trilinear decomposition. In addition to initialization, there are several constraints that may be

selected. Constraints include unimodality, orthogonality that may help stabilize the solution, and non-negativity that ensures the loading vector should have positive signal [169]. While the total explained variance may be less than what may be achieved with an unconstrained model, applying the constraints may be beneficial for the interpretation of the data, i.e. non-negativity constraint ensures that signals of analytes with non-zero concentration are positive.

1.3 Overview of following chapters

1.3.1 Chapter 2: Optimization of Parameters for ROI Data Compression for Nontargeted Analyses Using LC−HRMS

Non-targeted analyses of low concentration analytes in the information rich data collected by liquid chromatography with high resolution mass spectrometry detection can be challenging to accomplish in an efficient and comprehensive manner. The aim of this study is to demonstrate a workflow involving targeted parameter optimization for entire chromatograms using region of interest (ROI) data compression uncoupled from a subsequent tile-based F-ratio analysis, a supervised discovery-based method, for the discovery of low concentration analytes. Soil samples spiked with 18 pesticides at nominal concentrations ranging from 0.1 ppb to 50 ppb for a total of six sample classes served as challenging samples to demonstrate the overall workflow. Optimization of two parameters proved to be the most critical for ROI data compression: the signal threshold parameter and the admissible mass deviation parameter. The parameter optimization method workflow we introduce is based upon spiking known analytes into a representative sample and determining the number of detectable spikes and the Δppm for various combinations of signal threshold and admissible mass deviation, where Δppm is the absolute value of the difference between the theoretical *m/z* and the ROI *m/z*. Once optimal parameters are determined providing the lowest average Δppm and the greatest number of detectable analytes,

the optimized parameters can be utilized for the intended analysis. Herein, tile-based F-ratio analysis was performed on the ROI compressed data of all spiked soil samples first by applying ROI parameters recommended in the literature, referred to herein as the initial ROI parameters, and finally by the combination of the two optimized parameters. Using the initial ROI parameters, 3 pesticides were discovered, where-as all 18 spiked pesticides were discovered by optimizing both ROI parameters.

1.3.2 Chapter 3: Chemometric decomposition of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data employing partial modulation in the negative pulse mode

Partial modulation in the negative pulse mode was optimized for GC×GC-TOFMS. With partial modulation in the negative pulse mode, a flow of auxiliary carrier gas is applied through a pulse valve to a T-junction joining the ^{1}D and ^{2}D columns for nearly all of the P_M. This results in dilution of the ^{1}D eluate, followed by briefly turning off the pulse valve for a specified pulse width (p_w), effectively "injecting" undiluted ^{1}D eluate onto the ^{2}D column. The raw data has the appearance of ^{2}D separations superimposed on top of the ^{1}D separation. While high peak capacity GC×GC data are produced, there are challenges that needed to be addressed regarding the use of chemometrics for analyte decomposition, identification, and quantification. Herein, these data analysis challenges are addressed using multivariate curve resolution - alternating least squares (MCR-ALS). An isothermal separation of a 15-component mixture of similar compounds is obtained in 20 s using a P_M = 250 ms and a p_w = 6 ms. Various peak overlap situations were purposely produced to facilitate the chemometric method demonstration. The MCR-ALS chromatographic loadings (peak profiles) were used to determine t_R and W_b in both separation dimensions. MCR-ALS readily decomposed ^{1}D×^{2}D regions for analytes with severe

^{1}D overlap if they were fully resolved on ^{2}D. Decomposition was more challenging for analytes severely overlapped on both GC×GC dimensions when their spectra were similar, though ultimately all 15 compounds were successfully decomposed. Additionally, the ^{2}D concentration peak profiles obtained via MCR-ALS are demonstrated to be reliable for identification and quantifiable. The predicted versus prepared concentration values are in good agreement for two representative analytes, with percent deviation values of -5.6% (± 2.2%) for 1-hexene, and 1.8% (± 3.4%) for 2-pentanone.

1.3.3 Chapter 4: Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection

Dynamic pressure gradient modulation (DPGM) is investigated for GC×GC-TOFMS. With DPGM, a commercial pneumatic "pulse" valve is opened to introduce a suitably high auxiliary gas pressure at a T-junction connecting the ^{1}D and ^{2}D columns during the P_M, temporarily stopping the ^{1}D flow. The valve is then closed for the duration of a pulse width (p_w) to "re-inject" temporally focused ^{1}D eluate onto the ^{2}D column for separation. This flow modulation technique is observed to be compatible with TOFMS detection using a ^{2}D flow rate of 4 ml/min for the separation of a 90-component test mixture. A 25 min separation window using a P_M = 1 s and p_w = 200 ms for full modulation (and 100% duty cycle) provided an average ^1W_b = 4.5 s and ^2W_b = 130 ms for a 2D peak capacity of $n_{c,2D}$ = 2700 (100 peaks per min). The detector response enhancement factor (DREF) serves as a metric for the enhanced sensitivity of the modulated relative to the unmodulated ^{1}D peaks, with DREFs ranging between 10 and 20 and about a 5-fold improvement in *S/N*. The bilinear "quality" of the GC×GC data is studied using PARAFAC. Since PARAFAC requires sufficiently trilinear data, the reproducibility of the ^{2}D peak shape for a given analyte is confirmed using lack-of-fit (*LOF*) and percent variation (R^2) metrics. The

limit-of-detection (*LOD*) for the representative analyte hexadecane is determined using PARAFAC, providing an *LOD* of 0.7 ppb (±0.03 ppb) for three replicates. Seven heavily overlapped analytes are also fully resolved by PARAFAC down to the part-per-million (ppm) concentration level, producing reproducible spectra with a majority of spectral match values (MV) over 800 ($RSD \leq 7.1\%$). This study provides promising results for DPGM as a flow modulation technique compatible with GC×GC-TOFMS, providing high sensitivity data suitable for chemometric analysis.

1.3.4 Chapter 5: A systematic investigation of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry with dynamic pressure gradient modulation for high peak capacity separations

Dynamic pressure gradient modulation (DPGM) in full modulation mode is optimized for GC×GC-TOFMS) to obtain high peak capacity separations and demonstrate broad applicability for complex samples. A pulse valve introduces an auxiliary carrier gas flow at a T-union connecting the ^{1}D column to the ^{2}D column. At a sufficiently high auxiliary pressure (P_{aux}) the ^{1}D flow is temporarily stopped. Then, during each P_M the valve is turned off briefly, a period termed the pulse width (p_w), allowing the ^{1}D effluent to essentially be reinjected onto the ^{2}D column for the modulated separations. Modifications to the modulator assembly are provided to improve performance. Method optimization is demonstrated for a 116-component test mixture by tuning the P_{aux} and the p_w. For a P_M = 2 s and ^{1}F of 0.10 ml/min, the optimal p_w and initial P_{aux} selected were 200 ms and 330.9 kPa (33 psig), respectively. The 30 min separation of the test mixture provided a ^{1}D peak capacity of $^{1}n_c$ = 330 and a ^{2}D peak capacity of $^{2}n_c$ = 15, hence an ideal 2D peak capacity $n_{c,2D} = {}^{1}n_c \times {}^{2}n_c$ = 4950. Likewise, the 2D peak capacity corrected for undersampling of the ^{1}D separation was 4500 and corrected for both undersampling and sampling variation via statistical overlap theory was 4090. These results provide a 2-fold improvement in peak capacity relative to the previous DPGM study in full modulation mode for

GC×GC-TOFMS. The optimized conditions were applied for a variety of applications: diesel fuel, derivatized cow serum, solid phase microextraction (SPME) of coffee headspace, and SPME of river water headspace. Additionally, the fraction of 2D separation space utilized ($f_{coverage}$), as defined by the minimum convex hull method, ranged from 0.60 – 0.85. We observed that any $f_{coverage}$ correction to 2D peak capacity is highly sample dependent, since all samples, except for the diesel sample, were run with the same separation conditions, and yet the $f_{coverage}$ ranged from 0.60 – 0.80.

1.3.5 Chapter 6: Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography / mass spectrometry data: Study of the pacu fish metabolome

Integration of rice and fish farming, eg., pacu fish in Argentina, has raised concern that herbicides used for rice paddies may adversely affect the fish metabolome. To study this issue, tile-based F-ratio analysis was applied to GC×GC-TOFMS data of pacu fish raised in an integrated rice-fish farming system (farmed class) versus fish raised in tanks (tank-raised class) to discover class-distinguishing analytes. F-ratio analysis resulted in hit lists initially dominated by artifact peaks from the sample derivatization process, as well as some redundant hits. These challenges were addressed by developing an automated artifact removal algorithm and an improvement to redundant hit removal in the tile-based F-ratio analysis using a pooled farmed fish sample, either spiked with 29 metabolites (spiked class) or unspiked (i.e., the background serving as the control class). Of the 29 spiked metabolites, 23 were discovered by standard F-ratio analysis, improving to 28 discovered using control-normalized F-ratio analysis. Standard and control-normalized F-ratio hit lists initially with 185 and 246 hits, were reduced to 56 and 49 hits, respectively, after artifact removal and removing redundant hits. Next, we returned to the F-ratio analysis of the farmed fish versus the control fish. Here, we introduce a minimum variance optimized (MVO) F-ratio calculation (^{MVO}F-ratio) that provides a comprehensive hit list ranking. The initial ^{MVO}F-ratio analysis hit list of 537 hits was reduced to 110 hits following artifact

removal. Of the 110 hits, 70 expressed a concentration ratio statistically different than 1 ($p < 0.05$). The ^{MVO}F-ratio hit list discovered more true positives compared to the standard, tank-normalized, and farm-normalized F-ratio hit lists, providing the combination of results from the farm-normalized and tank-normalized hit lists. A majority of analytes (54 out of 70) important for normal biological functioning of pacu fish were significantly downregulated in the farmed fish, suggesting the integrated farming system may negatively impact pacu fish quality.

1.3.6 References

[1] A.T. James, A.J.P. Martin, Gas-liquid partition chromatography: the separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid, Biochem. J. 50 (1952) 679–690. doi:10.1042/bj0500679.

[2] K. Robards, D. Ryan, Chapter 5 - High performance liquid chromatography: Instrumentation and techniques, in: K. Robards, D. Ryan (Eds.), Princ. Pract. Mod. Chromatogr. Methods (Second Ed., Second Edition, Academic Press, 2022: pp. 247–282. doi:https://doi.org/10.1016/B978-0-12-822096-2.00009-8.

[3] K. Robards, P.R. Haddad, P.E. Jackson, 2 - Theory of Chromatography, in: K. Robards, P.R. Haddad, P.E. Jackson (Eds.), Princ. Pract. Mod. Chromatogr. Methods, Academic Press, Boston, 2004: pp. 35–74. doi:https://doi.org/10.1016/B978-0-08-057178-2.50005-4.

[4] V.R. Reid, R.E. Synovec, High-speed gas chromatography: The importance of instrumentation optimization and the elimination of extra-column band broadening, Talanta. 76 (2008) 703–717. doi:10.1016/j.talanta.2008.05.012.

[5] 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. doi:10.1016/j.chroma.2019.460368.

[6] 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. 195 (2019) 822–829. doi:10.1016/j.talanta.2018.12.007.

[7] B.A. Parsons, D.K. Pinkerton, R.E. Synovec, Implications of phase ratio for maximizing peak capacity in comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A. 1536 (2018) 16–26. doi:10.1016/j.chroma.2017.07.018.

[8] C.F. Poole, S.K. Poole, Separation characteristics of wall-coated open-tubular columns for gas chromatography, J. Chromatogr. A. 1184 (2008) 254–280. doi:10.1016/j.chroma.2007.07.028.

[9] J.S. Nadeau, R.B. Wilson, B.D. Fitz, J.T. Reed, R.E. Synovec, Utilizing a constant peak width transform for isothermal gas chromatography, J. Chromatogr. A. 1218 (2011) 3718–

3724. doi:10.1016/j.chroma.2011.04.007.

[10] R.B. Wilson, W.C. Siegler, J.C. Hoggard, B.D. Fitz, J.S. Nadeau, R.E. Synovec, Achieving high peak capacity production for gas chromatography and comprehensive two-dimensional gas chromatography by minimizing off-column peak broadening, J. Chromatogr. A. 1218 (2011) 3130–3139. doi:10.1016/j.chroma.2010.12.108.

[11] B.D. Fitz, R.B. Wilson, B.A. Parsons, J.C. Hoggard, R.E. Synovec, Fast, high peak capacity separations in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A. 1266 (2012) 116–123. doi:10.1016/j.chroma.2012.09.096.

[12] J.M. Davis, J.C. Giddings, Statistical Theory of Component Overlap in Multicomponent Chromatograms, Anal. Chem. 55 (1983) 418–424. doi:10.1021/ac00254a003.

[13] M. Edwards, H. Boswell, T. Górecki, Comprehensive Multidimensional Chromatography, Curr. Chromatogr. 2 (2015) 80–109. doi:10.2174/2213240602666150722232236.

[14] J.C. Giddings, Two-dimensional separations: concept and promise., Anal. Chem. 56 (1984). doi:10.1021/ac00276a003.

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

[16] H.D. Bahaghighat, C.E. Freye, D. V. Gough, P.E. Sudol, R.E. Synovec, Ultrafast separations via pulse flow valve modulation to enable high peak capacity multidimensional gas chromatography, J. Chromatogr. A. 1573 (2018). doi:10.1016/j.chroma.2018.08.001.

[17] L.M. Blumberg, F. David, M.S. Klee, P. Sandra, Comparison of one-dimensional and comprehensive two-dimensional separations by gas chromatography, J. Chromatogr. A. 1188 (2008) 2–16. doi:10.1016/j.chroma.2008.02.044.

[18] J.M. Davis, D.R. Stoll, P.W. Carr, Dependence of effective peak capacity in comprehensive two-dimensional separations on the distribution of peak capacity between the two dimensions, Anal. Chem. 80 (2008) 8122–8134. doi:10.1021/ac800933z.

[19] F.J. Oros, J.M. Davis, Comparison of statistical theories of spot overlap in two-dimensional separations and verification of means for estimating the number of zones, J. Chromatogr. A. 591 (1992) 1–18. doi:10.1016/0021-9673(92)80218-J.

[20] J.M. Davis, Statistical theory of spot overlap for n-dimensional separations, Anal. Chem. 65 (1993) 2014–2023. https://pubs.acs.org/doi/abs/10.1021/ac00019a014.

[21] 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.

[22] B. Giocastro, M. Zoccali, P.Q. Tranchida, L. Mondello, Evaluation of different internal diameter coated modulation columns within the context of solid-state modulation, J. Sep.

Sci. (2021) 1–20. doi:10.1002/jssc.202001031.

[23] J. Crucello, L.F.O. Miron, V.H.C. Ferreira, H. Nan, M.O.M. Marques, P.S. Ritschel, M.C. Zanus, J.L. Anderson, R.J. Poppi, L.W. Hantao, Characterization of the aroma profile of novel Brazilian wines by solid-phase microextraction using polymeric ionic liquid sorbent coatings, Anal. Bioanal. Chem. 410 (2018) 4749–4762. doi:10.1007/s00216-018-1134-3.

[24] F. Stilo, C. Bicchi, A. Robbat, S.E. Reichenbach, C. Cordero, Untargeted approaches in food-omics: The potential of comprehensive two-dimensional gas chromatography/mass spectrometry, TrAC - Trends Anal. Chem. 135 (2021) 116162. doi:10.1016/j.trac.2020.116162.

[25] V.G. Van Mispelaar, A.C. Tas, A.K. Smilde, P.J. Schoenmakers, A.C. Van Asten, Quantitative analysis of target components by comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1019 (2003) 15–29. doi:10.1016/j.chroma.2003.08.101.

[26] J.F. Griffith, W.L. Winniford, K. Sun, R. Edam, J.C. Luong, A reversed-flow differential flow modulator for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1226 (2012) 116–123. doi:10.1016/j.chroma.2011.11.036.

[27] 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. doi:10.1016/j.trac.2018.05.017.

[28] P.E. Sudol, K.M. Pierce, S.E. Prebihalo, K.J. Skogerboe, B.W. Wright, R.E. Synovec, Development of gas chromatographic pattern recognition and classification tools for compliance and forensic analyses of fuels: A review, Anal. Chim. Acta. (2020). doi:10.1016/j.aca.2020.07.027.

[29] G. Ebadzadsahrai, E.A. Higgins Keppler, S.D. Soby, H.D. Bean, Inhibition of Fungal Growth and Induction of a Novel Volatilome in Response to Chromobacterium vaccinii Volatile Organic Compounds, Front. Microbiol. 11 (2020). doi:10.3389/fmicb.2020.01035.

[30] H.D. Bean, J.E. Hill, J.M.D. Dimandja, Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography-mass spectrometry data, J. Chromatogr. A. 1394 (2015) 111–117. doi:10.1016/j.chroma.2015.03.001.

[31] 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. doi:10.1016/j.chroma.2017.05.065.

[32] K. Ralston-Hooper, A. Hopf, C. Oh, X. Zhang, J. Adamec, M.S. Sepúlveda, Development of GCxGC/TOF-MS metabolomics for use in ecotoxicological studies with invertebrates, Aquat. Toxicol. 88 (2008) 48–52. doi:10.1016/j.aquatox.2008.03.002.

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

[34] R.E. Murphy, M.R. Schure, J.P. Foley, Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography, Anal. Chem. 70 (1998) 1585–1594. doi:10.1021/ac971184b.

[35] J. Krupčík, P. Májek, R. Gorovenko, I. Špánik, P. Sandra, D.W. Armstrong, On the determination of a detector response enhancement factor for flow modulated comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1286 (2013) 235–240. doi:10.1016/j.chroma.2013.02.068.

[36] D.R. Deans, A new technique for heart cutting in gas chromatography, Chromatographia. 1 (1968) 18–22. doi:10.1007/BF02259005.

[37] H.D. Bahaghighat, C.E. Freye, R.E. Synovec, Recent advances in modulator technology for comprehensive two dimensional gas chromatography, TrAC - Trends Anal. Chem. 113 (2019) 379–391. doi:10.1016/j.trac.2018.04.016.

[38] M.W. Amer, B. Mitrevski, W. Roy Jackson, A.L. Chaffee, P.J. Marriott, Multidimensional and comprehensive two-dimensional gas chromatography of dichloromethane soluble products from a high sulfur Jordanian oil shale, Talanta. 120 (2014) 55–63. doi:10.1016/j.talanta.2013.11.069.

[39] P. Korytár, P.E.G. Leonards, J. De Boer, U.A.T. Brinkman, Group separation of organohalogenated compounds by means of comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1086 (2005) 29–44. doi:10.1016/j.chroma.2005.05.087.

[40] P. Haglund, P. Korytár, C. Danielsson, J. Diaz, K. Wiberg, P. Leonards, U.A.T. Brinkman, J. De Boer, GC×GC-ECD: A promising method for the determination of dioxins and dioxin-like PCBs in food and feed, Anal. Bioanal. Chem. 390 (2008) 1815–1827. doi:10.1007/s00216-008-1896-0.

[41] 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. doi:10.1016/j.chroma.2017.05.064.

[42] 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. doi:10.1016/j.talanta.2013.08.033.

[43] J. Krupčík, R. Gorovenko, I. Špánik, P. Sandra, D.W. Armstrong, Flow-modulated comprehensive two-dimensional gas chromatography with simultaneous flame ionization and quadrupole mass spectrometric detection, J. Chromatogr. A. 1280 (2013) 104–111. doi:10.1016/j.chroma.2013.01.015.

[44] K. Murtada, D. Bowman, M. Edwards, J. Pawliszyn, Thin-film microextraction combined

with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry screening for presence of multiclass organic pollutants in drinking water samples, Talanta. 242 (2022) 123301. doi:10.1016/j.talanta.2022.123301.

[45] A. Moreira de Oliveira, C.A. Teixeira, L.W. Hantao, Advanced tuning of the ion management parameters in GC × GC-HRMS using a Fourier transform Orbitrap mass analyzer for pixel-based data handling and multivariate analysis, Anal. Methods. (2022). doi:10.1039/D2AY00314G.

[46] T.J. Trinklein, S. Schöneich, P.E. Sudol, C.G. Warren, D. V. Gough, R.E. Synovec, Total-transfer comprehensive three-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A. 1634 (2020) 461654. doi:10.1016/j.chroma.2020.461654.

[47] T.J. Trinklein, C.G. Warren, R.E. Synovec, Determination of the Signal-To-Noise Ratio Enhancement in Comprehensive Three-Dimensional Gas Chromatography, Anal. Chem. 93 (2021) 8526–8535. doi:10.1021/acs.analchem.1c01190.

[48] P.E. Sudol, S. Schöneich, R.E. Synovec, Principal component analysis of comprehensive three-dimensional gas chromatography time-of-flight mass spectrometry data, J. Chromatogr. Open. 2 (2022) 100043. doi:10.1016/j.jcoa.2022.100043.

[49] O. Panić, T. Górecki, C. McNeish, A.H. Goldstein, B.J. Williams, D.R. Worton, S. V. Hering, N.M. Kreisberg, Development of a new consumable-free thermal modulator for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1218 (2011) 3070–3079. doi:10.1016/j.chroma.2011.03.024.

[50] S.J. Kim, G. Serrano, K.D. Wise, K. Kurabayashi, E.T. Zellers, Evaluation of a microfabricated thermal modulator for comprehensive two-dimensional microscale gas chromatography, Anal. Chem. 83 (2011) 5556–5562. doi:10.1021/ac200336e.

[51] J. Luong, X. Guan, S. Xu, R. Gras, R.A. Shellie, Thermal Independent Modulator for Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 88 (2016) 8428–8432. doi:10.1021/acs.analchem.6b02525.

[52] A.M. Muscalu, M. Edwards, T. Górecki, E.J. Reiner, Evaluation of a single-stage consumable-free modulator for comprehensive two-dimensional gas chromatography: Analysis of polychlorinated biphenyls, organochlorine pesticides and chlorobenzenes, J. Chromatogr. A. 1391 (2015) 93–101. doi:10.1016/j.chroma.2015.02.074.

[53] B. Savareear, M. Brokl, C. Wright, J.F. Focant, Thermal desorption comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry for vapour phase mainstream tobacco smoke analysis, J. Chromatogr. A. 1525 (2017) 126–137. doi:10.1016/j.chroma.2017.10.013.

[54] B. V. Burger, T. Snyman, W.J.G. Burger, W.F. Van Rooyen, Thermal modulator array for analyte modulation and comprehensive two-dimensional gas chromatography, J. Sep. Sci. 26 (2003) 123–128. doi:10.1002/jssc.200390002.

[55] A. Ghosh, C.T. Bates, S.K. Seeley, J. V. Seeley, High speed Deans switch for low duty cycle comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1291 (2013) 146–154. doi:10.1016/j.chroma.2013.04.003.

[56] J. V. Seeley, N.E. Schimmel, S.K. Seeley, The multi-mode modulator: A versatile fluidic device for two-dimensional gas chromatography, J. Chromatogr. A. 1536 (2018) 6–15. doi:10.1016/j.chroma.2017.06.030.

[57] P.A. Bueno, J. V. Seeley, Flow-switching device for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1027 (2004) 3–10. doi:10.1016/j.chroma.2003.10.033.

[58] J. V. Seeley, N.J. Micyus, S. V. Bandurski, S.K. Seeley, J.D. McCurry, Microfluidic deans switch for comprehensive two-dimensional gas chromatography, Anal. Chem. 79 (2007) 1840–1847. doi:10.1021/ac061881g.

[59] C.E. Freye, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta. 161 (2016) 675–680. doi:10.1016/j.talanta.2016.09.002.

[60] R.E. Mohler, B.J. Prazen, R.E. Synovec, Total-transfer, valve-based comprehensive two-dimensional gas chromatography, Anal. Chim. Acta. 555 (2006) 68–74. doi:10.1016/j.aca.2005.08.072.

[61] P.Q. Tranchida, G. Purcaro, A. Visco, L. Conte, P. Dugo, P. Dawes, L. Mondello, A flexible loop-type flow modulator for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1218 (2011) 3140–3145. doi:10.1016/j.chroma.2010.11.082.

[62] P.Q. Tranchida, S. Salivo, F.A. Franchina, L. Mondello, Flow-modulated comprehensive two-dimensional gas chromatography combined with a high-resolution time-of-flight mass spectrometer: A proof-of-principle study, Anal. Chem. 87 (2015) 2925–2930. doi:10.1021/ac5044175.

[63] S. Schöneich, D. V Gough, T.J. Trinklein, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection, J. Chromatogr. A. (2020) 460982. doi:https://doi.org/10.1016/j.chroma.2020.460982.

[64] P.Q. Tranchida, F.A. Franchina, P. Dugo, L. Mondello, Flow-modulation low-pressure comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1372 (2014) 236–244. https://doi.org/10.1016/j.chroma.2014.10.097.

[65] P.Q. Tranchida, F.A. Franchina, P. Dugo, L. Mondello, Use of greatly-reduced gas flows in flow-modulated comprehensive two-dimensional gas chromatography-mass spectrometry, J. Chromatogr. A. 1359 (2014) 271–276.

[66] F.A. Franchina, M. Maimone, P.Q. Tranchida, L. Mondello, Flow modulation comprehensive two-dimensional gas chromatography-mass spectrometry using ≈ 4 mL min-1 gas flows, J. Chromatogr. A. 1441 (2016) 134–139. doi:10.1016/j.chroma.2016.02.041.

[67] P.Q. Tranchida, F.A. Franchina, M. Zoccali, I. Bonaccorsi, F. Cacciola, L. Mondello, A direct sensitivity comparison between flow-modulated comprehensive 2D and 1D GC in untargeted and targeted MS-based experiments, J. Sep. Sci. 36 (2013) 2746–2752. doi:10.1002/jssc.201300423.

[68] T.J. Trinklein, D. V. Gough, C.G. Warren, G.S. Ochoa, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1609 (2020) 460488. doi:10.1016/j.chroma.2019.460488.

[69] S. Schöneich, T.J. Trinklein, C.G. Warren, R.E. Synovec, A systematic investigation of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry with dynamic pressure gradient modulation for high peak capacity separations, Anal. Chim. Acta. 1134 (2020) 115–124. doi:10.1016/j.aca.2020.08.023.

[70] P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis, J. Chromatogr. A. 1662 (2022) 462735. doi:10.1016/j.chroma.2021.462735.

[71] A. Ghosh, C.T. Bates, S.K. Seeley, J. V. Seeley, High speed Deans switch for low duty cycle comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1291 (2013) 146–154. doi:10.1016/j.chroma.2013.04.003.

[72] S. Schöneich, D. V Gough, T.J. Trinklein, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection, J. Chromatogr. A. (2020) 460982. doi:https://doi.org/10.1016/j.chroma.2020.460982.

[73] P.M. Harvey, R.A. Shellie, Factors affecting peak shape in comprehensive two-dimensional gas chromatography with non-focusing modulation, J. Chromatogr. A. 1218 (2011) 3153–3158. doi:10.1016/j.chroma.2010.08.029.

[74] W.C. Siegler, B.D. Fitz, J.C. Hoggard, R.E. Synovec, Experimental study of the quantitative precision for valve-based comprehensive two-dimensional gas chromatography, Anal. Chem. 83 (2011) 5190–5196. doi:10.1021/ac200302b.

[75] K. Murtada, D. Bowman, M. Edwards, J. Pawliszyn, Thin-film microextraction combined with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry screening for presence of multiclass organic pollutants in drinking water samples, Talanta. 242 (2022) 123301. doi:10.1016/j.talanta.2022.123301.

[76] G. Ebadzadsahrai, E.A. Higgins Keppler, S.D. Soby, H.D. Bean, Inhibition of Fungal Growth and Induction of a Novel Volatilome in Response to Chromobacterium vaccinii Volatile Organic Compounds, Front. Microbiol. 11 (2020). doi:10.3389/fmicb.2020.01035.

[77] 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.

[78] U. Käfer, T. Gröger, C.J. Rohbogner, D. Struckmeier, M.R. Saraji-Bozorgzad, T. Wilharm, R. Zimmermann, Detailed Chemical Characterization of Bunker Fuels by High-Resolution Time-of-Flight Mass Spectrometry Hyphenated to GC × GC and Thermal Analysis, Energy and Fuels. 33 (2019) 10745–10755.

doi:10.1021/acs.energyfuels.9b02626.

[79] P.H. Stefanuto, D. Zanella, J. Vercammen, M. Henket, F. Schleich, R. Louis, J.F. Focant, Multimodal combination of GC × GC-HRTOFMS and SIFT-MS for asthma phenotyping using exhaled breath, Sci. Rep. 10 (2020) 1–11. doi:10.1038/s41598-020-73408-2.

[80] L.M. Dubois, P.H. Stefanuto, L. Heudt, J.F. Focant, K.A. Perrault, Characterizing decomposition odor from soil and adipocere samples at a death scene using HS-SPME-GC×GC-HRTOFMS, Forensic Chem. 8 (2018) 11–20. doi:10.1016/j.forc.2018.01.001.

[81] A. Moreira de Oliveira, C.A. Teixeira, L.W. Hantao, Advanced tuning of the ion management parameters in GC × GC-HRMS using a Fourier transform Orbitrap mass analyzer for pixel-based data handling and multivariate analysis, Anal. Methods. (2022). doi:10.1039/D2AY00314G.

[82] L. Mikaliunaite, P.E. Sudol, C.N. Cain, R.E. Synovec, Baseline correction method for dynamic pressure gradient modulated comprehensive two-dimensional gas chromatography with flame ionization detection, J. Chromatogr. A. 1652 (2021).

[83] D. Zhang, X. Huang, F.E. Regnier, M. Zhang, Two-dimensional correlation optimized warping algorithm for aligning GCxGC-MS data, Anal. Chem. 80 (2008) 2664–2671. doi:10.1021/ac7024317.

[84] 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. doi:10.1016/j.chemolab.2012.02.003.

[85] D. Bylund, R. Danielsson, G. Malmquist, K.E. Markides, Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modelling of liquid chromatography-mass spectrometry data, J. Chromatogr. A. 961 (2002) 237–244. doi:10.1016/S0021-9673(02)00588-5.

[86] K.M. Pierce, J.L. Hope, K.J. Johnson, B.W. Wright, R.E. Synovec, Classification of gasoline data obtained by gas chromatography using a piecewise alignment algorithm combined with feature selection and principal component analysis, J. Chromatogr. A. 1096 (2005) 101–110. doi:10.1016/j.chroma.2005.04.078.

[87] 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. doi:10.1021/ac301367s.

[88] 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. doi:10.1016/j.chroma.2017.05.065.

[89] C.N. Cain, S. Schöneich, R.E. Synovec, Development of an Enhanced Total Ion Current Chromatogram Algorithm to Improve Untargeted Peak Detection, Anal. Chem. 92 (2020) 11365–11373. doi:10.1021/acs.analchem.0c02136.

[90] B. Li, S.E. Reichenbach, Q. Tao, R. Zhu, A streak detection approach for comprehensive two-dimensional gas chromatography based on image analysis, Neural Comput. Appl. 32 (2020) 649–663. doi:10.1007/s00521-018-3917-z.

[91] A. Bieniek, A. Moga, An efficient watershed algorithm based on connected components, Pattern Recognit. 33 (2000) 907–916. doi:10.1016/S0031-3203(99)00154-5.

[92] O. Amador-Muñoz, P.J. Marriott, Quantification in comprehensive two-dimensional gas chromatography and a model of quantification based on selected summed modulated peaks, J. Chromatogr. A. 1184 (2008) 323–340. doi:10.1016/j.chroma.2007.10.041.

[93] L.W. Hantao, C.A. Teixeira, V.H.C. Ferreira, CDF2MAT Automated SCRIPT to import NETCDF files to MATLAB | RESAMPLING added to correct RESHAPE for non-integer MS acquisition rates in GCxGC-MS data, (2022). doi:doi/10.7910/DVN/WMTEMF.

[94] P.E. Sudol, D. V. Gough, S.E. Prebihalo, R.E. Synovec, Impact of data bin size on the classification of diesel fuels using comprehensive two-dimensional gas chromatography with principal component analysis, Talanta. 206 (2020) 120239. doi:10.1016/j.talanta.2019.120239.

[95] M.M. Sinanian, D.W. Cook, S.C. Rutan, D.S. Wijesinghe, Multivariate curve resolution-alternating least squares analysis of high-resolution liquid chromatography-mass spectrometry data, Anal. Chem. 88 (2016) 11092–11099. doi:10.1021/acs.analchem.6b03116.

[96] M. Navarro-Reig, J. Jaumot, R. Tauler, An untargeted lipidomic strategy combining comprehensive two-dimensional liquid chromatography and chemometric analysis, J. Chromatogr. A. 1568 (2018) 80–90. doi:10.1016/j.chroma.2018.07.017.

[97] M. Pérez-Cova, J. Jaumot, R. Tauler, Untangling comprehensive two-dimensional liquid chromatography data sets using regions of interest and multivariate curve resolution approaches, TrAC Trends Anal. Chem. 137 (2021) 116207. doi:10.1016/j.trac.2021.116207.

[98] R.S.M. Pourasil, J. Cristale, S. Lacorte, R. Tauler, Non-targeted Gas Chromatography Orbitrap Mass Spectrometry qualitative and quantitative analysis of semi-volatile organic compounds in indoor dust using the Regions of Interest Multivariate Curve Resolution chemometrics procedure, J. Chromatogr. A. 1668 (2022) 462907. doi:10.1016/j.chroma.2022.462907.

[99] E. Gorrochategui, J. Jaumot, R. Tauler, A protocol for LC-MS metabolomic data processing using chemometric tools, Protoc. Exch. (2015) 1–41. doi:10.1038/protex.2015.102.

[100] E. Gorrochategui, J. Jaumot, R. Tauler, ROIMCR: A powerful analysis strategy for LC-MS metabolomic datasets, BMC Bioinformatics. 20 (2019) 1–17. doi:10.1186/s12859-019-2848-8.

[101] M. Pérez-Cova, S. Platikanov, R. Tauler, J. Jaumot, Quantification strategies for two-dimensional liquid chromatography datasets using regions of interest and multivariate curve resolution approaches, Talanta. 247 (2022). doi:10.1016/j.talanta.2022.123586.

[102] R.S.M. Pourasil, J. Cristale, S. Lacorte, R. Tauler, Non-targeted Gas Chromatography Orbitrap Mass Spectrometry qualitative and quantitative analysis of semi-volatile organic compounds in indoor dust using the Regions of Interest Multivariate Curve Resolution chemometrics procedure, J. Chromatogr. A. 1668 (2022) 462907. doi:10.1016/j.chroma.2022.462907.

[103] 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. doi:10.1021/ac403650d.

[104] C.E. Freye, P.R. Bowden, M.T. Greenfield, B.C. Tappan, Non-targeted discovery-based analysis for gas chromatography with mass spectrometry: A comparison of peak table, tile, and pixel-based Fisher ratio analysis, Talanta. 211 (2020) 120668. doi:10.1016/j.talanta.2019.120668.

[105] 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.

[106] Y. Zou, M. Gaida, F.A. Franchina, P.H. Stefanuto, J.-F. Focant, Distinguishing between Decaffeinated and Regular Coffee by HS-SPME-GC×GC-TOFMS, Chemometrics, and Machine Learning, Molecules. 27 (2022) 1806. doi:10.3390/molecules27061806.

[107] T.J. Davis, T.R. Firzli, E.A. Higgins Keppler, M. Richardson, H.D. Bean, Addressing Missing Data in GC × GC Metabolomics: Identifying Missingness Type and Evaluating the Impact of Imputation Methods on Experimental Replication, Anal. Chem. 94 (2022) 10912–10920. doi:10.1021/acs.analchem.1c04093.

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

[109] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC-TOFMS) Data Using a Null Distribution Approach, Anal. Chem. 87 (2015) 3812–3819. doi:10.1021/ac504472s.

[110] 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.

[111] T.J. Trinklein, R.E. Synovec, Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based Fisher ratio analysis, J. Chromatogr. A. 1677 (2022) 463321. doi:10.1016/j.chroma.2022.463321.

[112] D. Zanella, A. Henin, S. Mascrez, P.H. Stefanuto, F.A. Franchina, J.F. Focant, G. Purcaro, Comprehensive two-dimensional gas chromatographic platforms comparison for exhaled breath metabolites analysis, J. Sep. Sci. 45 (2022) 3542–3555.

doi:10.1002/jssc.202200164.

[113] E. Hoh, N.G. Dodder, S.J. Lehotay, K.C. Pangallo, C.M. Reddy, K.A. Maruya, Nontargeted comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry method and software for inventorying persistent and bioaccumulative contaminants in marine environments, Environ. Sci. Technol. 46 (2012) 8001–8008. doi:10.1021/es301139q.

[114] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1987) 37–52. doi:https://doi.org/10.1016/0169-7439(87)80084-9.

[115] I.T. Jolliffe, Graphical Representation of Data Using Principal Components, in: Princ. Compon. Anal., Springer New York, New York, NY, 1986: pp. 64–91. doi:10.1007/978-1-4757-1904-8_5.

[116] C. Monzón, S. Schöneich, R.E. Synovec, Non-targeted discovery of class-distinguishing metabolites in Argentinian pacu fish by comprehensive two-dimensional gas chromatography with principal component analysis, Microchem. J. 164 (2021) 106004. doi:10.1016/j.microc.2021.106004.

[117] P.E. Sudol, D. V. Gough, S.E. Prebihalo, R.E. Synovec, Impact of data bin size on the classification of diesel fuels using comprehensive two-dimensional gas chromatography with principal component analysis, Talanta. 206 (2020) 120239. doi:10.1016/j.talanta.2019.120239.

[118] R. De Maesschalck, D.L. Massart, The Mahalanobis distance, Chemom. Intell. Lab. Syst. 50 (2000) 1–18.

[119] N.A. Sinkov, J.J. Harynuk, Cluster resolution: A metric for automated, objective and optimized feature selection in chemometric modeling, Talanta. 83 (2011) 1079–1087. doi:10.1016/j.talanta.2010.10.025.

[120] D. Ballabio, V. Consonni, Classification tools in chemistry. Part 1: Linear models. PLS-DA, Anal. Methods. 5 (2013) 3790–3798. doi:10.1039/c3ay40582f.

[121] L.C. Lee, C.Y. Liong, A.A. Jemain, Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps, Analyst. 143 (2018) 3526–3539. doi:10.1039/c8an00599k.

[122] A.C. Paiva, L.W. Hantao, Exploring a public database to evaluate consumer preference and aroma profile of lager beers by comprehensive two-dimensional gas chromatography and partial least squares regression discriminant analysis, J. Chromatogr. A. 1630 (2020) 461529. doi:10.1016/j.chroma.2020.461529.

[123] L.A. Adutwum, J.K. Kwao, J.J. Harynuk, Unique ion filter—A data reduction tool for chemometric analysis of raw comprehensive two-dimensional gas chromatography-mass spectrometry data, J. Sep. Sci. 44 (2021) 2773–2784. doi:10.1002/jssc.202001127.

[124] M.C. Rosso, F. Stilo, C. Bicchi, M. Charron, G. Rosso, R. Menta, S.E. Reichenbach, C.H. Weinert, C.I. Mack, S.E. Kulling, C. Cordero, Combined untargeted and targeted fingerprinting by comprehensive two-dimensional gas chromatography to track compositional changes on hazelnut primary metabolome during roasting, Appl. Sci. 11

(2021) 525. doi:10.3390/app11020525.

[125] I.G. Chong, C.H. Jun, Performance of some variable selection methods when multicollinearity is present, Chemom. Intell. Lab. Syst. 78 (2005) 103–112. doi:10.1016/j.chemolab.2004.12.011.

[126] T. Rajalahti, R. Arneberg, F.S. Berven, K.M. Myhr, R.J. Ulvik, O.M. Kvalheim, Biomarker discovery in mass spectral profiles by means of selectivity ratio plot, Chemom. Intell. Lab. Syst. 95 (2009) 35–48. doi:10.1016/j.chemolab.2008.08.004.

[127] J.A. Westerhuis, H.C.J. Hoefsloot, S. Smit, D.J. Vis, A.K. Smilde, E.J.J. Velzen, J.P.M. Duijnhoven, F.A. Dorsten, Assessment of PLSDA cross validation, Metabolomics. 4 (2008) 81–89. doi:10.1007/s11306-007-0099-6.

[128] 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.

[129] P.E. Sudol, G.S. Ochoa, R.E. Synovec, Investigation of the limit of discovery using tile-based Fisher ratio analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A. 1644 (2021) 462092. doi:10.1016/j.chroma.2021.462092.

[130] P.E. Sudol, G.S. Ochoa, R.E. Synovec, Investigation of the limit of discovery using tile-based Fisher ratio analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A. 1644 (2021) 462092. doi:10.1016/j.chroma.2021.462092.

[131] K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: Comprehensive GC × GC with ANOVA-based feature selection and principal component analysis, Chemom. Intell. Lab. Syst. 60 (2002) 225–237. doi:10.1016/S0169-7439(01)00198-8.

[132] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher ratio method applied to third-order separation data to identify significant chemical components of metabolite extracts, Anal. Chem. 78 (2006) 5068–5075. doi:10.1021/ac0602625.

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

[134] C.G. Fraga, B.J. Prazen, R.E. Synovec, Objective data alignment and chemometric analysis of comprehensive two-dimensional separations with run-to-run peak shifting on both dimensions, Anal. Chem. 73 (2001) 5833–5840. doi:10.1021/ac010656q.

[135] H. Parastar, N. Akvan, Multivariate curve resolution based chromatographic peak alignment combined with parallel factor analysis to exploit second-order advantage in complex chromatographic measurements, Anal. Chim. Acta. 816 (2014) 18–27. doi:10.1016/j.aca.2014.01.051.

[136] 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. doi:10.1016/j.chemolab.2012.02.003.

[137] 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 Anal. Technol. Biomed. Life Sci. 1114–1115 (2019) 146–153. doi:10.1016/j.jchromb.2019.01.029.

[138] S.E. Reichenbach, X. Tian, Q. Tao, E.B. Ledford, Z. Wu, O. Fiehn, Informatics for cross-sample analysis with comprehensive two-dimensional gas chromatography and high-resolution mass spectrometry (GCxGC-HRMS), Talanta. 83 (2011) 1279–1288. doi:10.1016/j.talanta.2010.09.057.

[139] L.M. Dubois, P.H. Stefanuto, L. Heudt, J.F. Focant, K.A. Perrault, Characterizing decomposition odor from soil and adipocere samples at a death scene using HS-SPME-GC×GC-HRTOFMS, Forensic Chem. 8 (2018) 11–20. doi:10.1016/j.forc.2018.01.001.

[140] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach, Anal. Chem. 87 (2015) 3812–3819. doi:10.1021/ac504472s.

[141] T.J. Trinklein, R.E. Synovec, Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based Fisher ratio analysis, J. Chromatogr. A. 1677 (2022) 463321. doi:10.1016/j.chroma.2022.463321.

[142] S.E. Prebihalo, G.S. Ochoa, K.L. Berrier, K.J. Skogerboe, K.L. Cameron, J.R. Trump, S.J. Svoboda, J.K. Wickiser, R.E. Synovec, 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, Anal. Chem. 92 (2020) 15526–15533. doi:10.1021/acs.analchem.0c03456.

[143] 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 (2019) 259–266.

[144] S. Schöneich, G.S. Ochoa, C.M. Monzón, R.E. Synovec, Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography / mass spectrometry data: Study of the pacu fish metabolome, J. Chromatogr. A. 1667 (2022) 462868. doi:10.1016/j.chroma.2022.462868.

[145] B.A. Parsons, D.K. Pinkerton, B.W. Wright, R.E. Synovec, Chemical characterization of the acid alteration of diesel fuel: Non-targeted analysis by two-dimensional gas chromatography coupled with time-of-flight mass spectrometry with tile-based Fisher ratio and combinatorial threshold determination, J. Chromatogr. A. 1440 (2016) 179–190. doi:10.1016/j.chroma.2016.02.067.

[146] G.S. Ochoa, S.E. Prebihalo, B.C. Reaser, L.C. Marney, R.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, J. Chromatogr. A. 1627 (2020) 461401. doi:10.1016/j.chroma.2020.461401.

[147] P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis, J. Chromatogr. A. 1662 (2022) 462735. doi:10.1016/j.chroma.2021.462735.

[148] A. Barcaru, G. Vivó-Truyols, Use of Bayesian Statistics for Pairwise Comparison of Megavariate Data Sets: Extracting Meaningful Differences between GCxGC-MS Chromatograms Using Jensen-Shannon Divergence, Anal. Chem. 88 (2016) 2096–2104. doi:10.1021/acs.analchem.5b03506.

[149] C.N. Cain, T.J. Trinklein, G.S. Ochoa, R.E. Synovec, Tile-Based Pairwise Analysis of GC × GC-TOFMS Data to Facilitate Analyte Discovery and Mass Spectrum Purification, Anal. Chem. 94 (2022) 5658–5666. doi:10.1021/acs.analchem.2c00223.

[150] G.S. Ochoa, P.E. Sudol, T.J. Trinklein, R.E. Synovec, Class comparison enabled mass spectrum purification for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta. 236 (2022) 122844. doi:10.1016/j.talanta.2021.122844.

[151] H.P. Bailey, S.C. Rutan, Chemometric resolution and quantification of four-way data arising from comprehensive 2D-LC-DAD analysis of human urine, Chemom. Intell. Lab. Syst. 106 (2011) 131–141. doi:10.1016/j.chemolab.2010.07.008.

[152] R. Bro, C.A. Andersson, H.A.L. Kiers, PARAFAC2 - Part II. Modeling chromatographic data with retention time shifts, J. Chemom. 13 (1999) 295–309. doi:10.1002/(SICI)1099-128X(199905/08)13:3/4<295::AID-CEM547>3.0.CO;2-Y.

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

[154] 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.

[155] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146.

[156] M. Alier, R. Tauler, Multivariate Curve Resolution of incomplete data multisets, Chemom. Intell. Lab. Syst. 127 (2013) 17–28. doi:10.1016/j.chemolab.2013.05.006.

[157] A. De Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods. 6 (2014) 4964–4976. doi:10.1039/c4ay00571f.

[158] L.A.F. de Godoy, L.W. Hantao, M.P. Pedroso, R.J. Poppi, F. Augusto, Quantitative

analysis of essential oils in perfume using multivariate curve resolution combined with comprehensive two-dimensional gas chromatography, Anal. Chim. Acta. 699 (2011) 120–125. doi:10.1016/j.aca.2011.05.003.

[159] 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.

[160] 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.

[161] W. Windig, C.E. Heckler, F.A. Agblevor, R.J. Evans, Self-modeling mixture analysis of categorized pyrolysis mass spectral data with the SIMPLISMA approach, Chemom. Intell. Lab. Syst. 14 (1992) 195–207. doi:10.1016/0169-7439(92)80104-C.

[162] K.J. Schostack, E.R. Malinowski, Preferred set selection by iterative key set factor analysis, Chemom. Intell. Lab. Syst. 6 (1989) 21–29. doi:10.1016/0169-7439(89)80062-0.

[163] J. Jaumot, R. Tauler, MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution, Chemom. Intell. Lab. Syst. 103 (2010) 96–107. doi:10.1016/j.chemolab.2010.05.020.

[164] M. Navarro-Reig, J. Jaumot, T.A. van Beek, G. Vivó-Truyols, R. Tauler, Chemometric analysis of comprehensive LC×LC-MS data: Resolution of triacylglycerol structural isomers in corn oil, Talanta. 160 (2016) 624–635. doi:10.1016/j.talanta.2016.08.005.

[165] M. Navarro-Reig, J. Jaumot, A. Baglai, G. Vivó-Truyols, P.J. Schoenmakers, R. Tauler, Untargeted Comprehensive Two-Dimensional Liquid Chromatography Coupled with High-Resolution Mass Spectrometry Analysis of Rice Metabolome Using Multivariate Curve Resolution, Anal. Chem. 89 (2017) 7675–7683. doi:10.1021/acs.analchem.7b01648.

[166] C. Perez-Lopez, A. Ginebreda, M. Carrascal, D. Barcelò, J. Abian, R. Tauler, Non-target protein analysis of samples from wastewater treatment plants using the regions of interest-multivariate curve resolution (ROIMCR) chemometrics method, J. Environ. Chem. Eng. (2021) 105752. doi:10.1016/j.jece.2021.105752.

[167] L.L. Hohrenk, M. Vosough, T.C. Schmidt, Implementation of Chemometric Tools to Improve Data Mining and Prioritization in LC-HRMS for Nontarget Screening of Organic Micropollutants in Complex Water Matrixes, Anal. Chem. 91 (2019) 9213–9220. doi:10.1021/acs.analchem.9b01984.

[168] M. Pérez-Cova, R. Tauler, J. Jaumot, Chemometrics in comprehensive two-dimensional liquid chromatography: A study of the data structure and its multilinear behavior, Chemom. Intell. Lab. Syst. 201 (2020). doi:10.1016/j.chemolab.2020.104009.

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

[170] H.A.L. Kiers, J.M.F. ten Berge, R. Bro, PARAFAC2 - Part I. A direct fitting algorithm for

the PARAFAC2 model, J. Chemom. 13 (1999) 275–294. doi:10.1002/(sici)1099-128x(199905/08)13:3/4<275::aid-cem543>3.3.co;2-2.

[171] J.C. Hoggard, W.C. Siegler, R.E. Synovec, Toward automated peak resolution in complete GC × GC-TOFMS chromatograms by PARAFAC, J. Chemom. 23 (2009) 421–431. doi:10.1002/cem.1239.

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

[173] T.J. Trinklein, S.E. Prebihalo, C.G. Warren, G.S. Ochoa, R.E. Synovec, Discovery-based analysis and quantification for comprehensive three-dimensional gas chromatography flame ionization detection data, J. Chromatogr. A. 1623 (2020) 461190. doi:10.1016/j.chroma.2020.461190.

[174] 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.

[175] R.C. Allen, S.C. Rutan, Semi-automated alignment and quantification of peaks using parallel factor analysis for comprehensive two-dimensional liquid chromatography-diode array detector data sets, Anal. Chim. Acta. 723 (2012) 7–17. doi:10.1016/j.aca.2012.02.019.

[176] 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.

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

[178] R.A. Harshman, M.E. Lundy, PARAFAC : Parallel factor analysis, Comput. Stat. Data Anal. 18 (1994) 39–72.

[179] C.E. Freye, H.D. Bahaghighat, R.E. Synovec, Comprehensive two-dimensional gas chromatography using partial modulation via a pulsed flow valve with a short modulation period, Talanta. 177 (2018) 142–149. doi:10.1016/j.talanta.2017.08.095.

[180] C.E. Freye, L. Mu, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1424 (2015) 127–133. doi:10.1016/j.chroma.2015.10.098.

[181] 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. doi:10.1016/j.chroma.2019.460368.

[182] H.D. Bahaghighat, C.E. Freye, D. V. Gough, R.E. Synovec, Comprehensive two-

dimensional gas chromatography and time-of-flight mass spectrometry detection with a 50 ms modulation period, J. Chromatogr. A. 1583 (2019) 117–123. doi:10.1016/j.chroma.2018.11.027.

[183] F.A. Franchina, M.E. Machado, P.Q. Tranchida, C.A. Zini, E.B. Caramão, L. Mondello, Determination of aromatic sulphur compounds in heavy gas oil by using (low-)flow modulated comprehensive two-dimensional gas chromatography–triple quadrupole mass spectrometry, J. Chromatogr. A. 1387 (2015) 86–94. doi:10.1016/j.chroma.2015.01.082.

[184] R.B. Wilson, J.C. Hoggard, R.E. Synovec, High throughput analysis of atmospheric volatile organic compounds by thermal injection - Isothermal gas chromatography - Time-of-flight mass spectrometry, Talanta. 103 (2013) 95–102. doi:10.1016/j.talanta.2012.10.013.

[185] 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.

[186] 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.

[187] I. Koo, S. Kim, X. Zhang, Comparative analysis of mass spectral matching-based compound identification in gas chromatography-mass spectrometry, J. Chromatogr. A. 1298 (2013) 132–138. doi:10.1016/j.chroma.2013.05.021.

[188] S. Kim, I. Koo, X. Wei, X. Zhang, A method of finding optimal weight factors for compound identification in gas chromatography-mass spectrometry, Bioinformatics. 28 (2012) 1158–1163. doi:10.1093/bioinformatics/bts083.

[189] B. Kehimkar, J.C. Hoggard, J.S. Nadeau, R.E. Synovec, Targeted mass spectral ratio analysis: A new tool for gas chromatography - Mass spectrometry, Talanta. 103 (2013) 267–275. doi:10.1016/j.talanta.2012.10.043.

[190] C.D. Brown, H.T. Davis, Receiver operating characteristics curves and related decision measures: A tutorial, Chemom. Intell. Lab. Syst. 80 (2006) 24–38. doi:10.1016/j.chemolab.2005.05.004.

[191] M. Sadeghi, T.K. Lee, D. McLean, H. Lui, M.S. Atkins, Detection and analysis of irregular streaks in dermoscopic images of skin lesions, IEEE Trans. Med. Imaging. 32 (2013) 849–861. doi:10.1109/TMI.2013.2239307.

[192] G.S. Ochoa, S.E. Prebihalo, B.C. Reaser, L.C. Marney, R.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, J. Chromatogr. A. 1627 (2020) 461401. doi:10.1016/j.chroma.2020.461401.

Chapter 2: Optimization of parameters for ROI data compression for non-targeted analyses using LC-HRMS

This chapter was reproduced from S. Schöneich, C.N. Cain, C.E. Freye, R.E. Synovec, Optimization of Parameters for ROI Data Compression for Nontargeted Analyses Using LC–HRMS, Anal. Chem. 95 (2022) 1513–1521.

2.1 Introduction

Liquid chromatography (LC) with high resolution mass spectrometry (HRMS) detection is a powerful separation tool implemented numerous applications [1–6]. LC-HRMS is well suited for studies targeting known analytes in complex samples [7,8] and for leveraging the detector selectivity in full scan mode for non-targeted studies [1,4,9], especially for trace level analytes possessing similar mass spectra along with chromatographic overlap. Non-targeted chemometric methods aim to discover the most chemically relevant information in a data set without prior knowledge of the analytes of interest. To mine the information-rich LC-HRMS data sets acquired in full scan mode in a non-targeted fashion, efficient chemometric techniques are required. Non-targeted methods generally fall under supervised methods, where class membership is known a priori, or unsupervised methods. Common unsupervised chemometric methods include principal component analysis (PCA) [4,10,11] and hierarchical cluster analysis (HCA) [6,12,13], while partial least squares-discriminant analysis (PLS-DA) [6,14,15] is commonly used for supervised analyses. Tile-based Fisher ratio (F-ratio) analysis [16–19] is another supervised method based upon the analysis of variance (ANOVA) [20,21], which has been developed for multidimensional gas chromatography [22,23] data but has yet to be evaluated for LC-HRMS data. Additionally, prior to using these chemometric methods, LC-HRMS data requires significant pre-processing. Thus, a workflow for pre-processing entire LC-HRMS chromatograms that is uncoupled from the subsequent intended analysis (targeted or non-

targeted, where the non-targeted analysis could be supervised or unsupervised) is a challenging issue that needs to be addressed and is the focus of this report. Herein, F-ratio analysis is demonstrated in the context of "entire chromatogram" parameter optimization.

Data collected by LC-HRMS requires pre-processing prior to chemometric analysis due to the considerable number of *m/z* collected [24]. Additionally, since HRMS data provides an irregular number of measured *m/z* and signal intensity pairs at each retention time, the raw data for each sample contains a unique array of *m/z* collected, requiring further processing prior to F-ratio analysis. To address this challenge, some form of data compression is required to reduce the computational load. Common data reduction approaches used for LC-HRMS data include binning [2,25], open-source tools that often use wavelet transforms with peak detection, alignment, normalization, and integration [6,24,26–28], or a recently developed region of interest (ROI) procedure [29,30]. The binning approach initially reduces the data to a user-defined precision for feature discovery and subsequently uses the high-resolution data for a targeted analysis of features. While this is effective and simple to apply, the mass accuracy and chromatographic resolution are not preserved, and the binned *m/z* dimension may not be equivalent across all replicates. In addition, the high sensitivity of the instrument is beneficial for analyzing analyte peaks while also collecting a substantial amount of noise or baseline. Such noisy bins are not useful to the analysis and may obscure the signal of low intensity analytes while increasing computational expense. The second approach typically utilizes a final peak table with peak areas or heights that could be used for statistical analysis rather than reduced chromatographic data, making it incompatible with tile-based analyses and susceptible to errors in the compilation of an accurate peak table [31,32]. The last approach has some similar steps as a few open-source tools [26,28], without requiring alignment and peak shaping, and reduces the

raw data to ROIs based on user-defined parameters that correspond to chromatographic peaks while preserving the mass accuracy. Through data augmentation steps, a single data cube is obtained with replicates across all sample classes with the same *m/z* dimension. Thus, the preserved mass accuracy provided by the ROI data compression method increases the likelihood that pure analyte *m/z* will be available for the intended chemometric method application. While binning and open source tools have been used extensively for chemometric decomposition [2,25] and PCA and/or PLS-DA [6,15], the ROI method has been implemented to a lesser extent, i.e., for chemometric decomposition [30,33,34]. However, ROI data compression has not been evaluated for tile-based non-targeted methods, which mitigates alignment issues and provides signal-to-noise (*S/N*) enhancement (both concerns with pixel-based and peak table methods). To address this challenging area, we introduce an effective workflow for optimizing ROI data compression in a targeted manner, independent of its implementation in a subsequent non-targeted analysis. Herein, we demonstrate the application of this parameter optimization workflow with the first application of tile-based F-ratio analysis to LC-HRMS data.

F-ratio analysis aims to discover the chromatographic features that distinguish between known sample classes, ranking features by their class-distinguishing ability. The standard F-ratio (^{S}F-ratio) is defined as the ratio of the between-class variance (s^2_{bc}) to the sum of the within-class variance (Σs^2_{wc}) and is calculated at each mass channel (*m/z*). Since the ^{S}F-ratio is nominally independent of the peak intensity by normalizing to Σs^2_{wc}, the ^{S}F-ratio magnitude indicates to what extent the discovered analyte "hits" discriminate between classes. Rather than calculating the F-ratio at every data pixel [35–37] that is susceptible to retention time misalignment, with tile-based F-ratio analysis the entire chromatogram is sectioned into tiles of user-defined dimensions and the F-ratios are calculated for each summed tile [16–18,31]. With proper

attention to tile size selection, peaks with retention time misalignment across replicates are captured within a single tile, requiring no additional alignment procedure [38]. Proper tile size selection is critical for the discovery of trace level changes across classes as analyte signal could be obscured by noise or high intensity interferent peaks as well as a lower likelihood of analyte *m/z* that are pure [19] meeting signal threshold requirements for accurate quantification. Using HRMS increases the probability of obtaining pure *m/z* per analyte as the mass accuracy improves, though not all *m/z* necessarily could pass the signal threshold to be considered by F-ratio analysis. To improve discovery of true positive hits, a recent development was applied herein termed the minimum variance optimized (MVO) F-ratio analysis [39], replacing the denominator with the minimum within class variance. For every tile, the within class variance is calculated for each class and the lowest variance is used. While this is not a true F-ratio, this modification to the ^{S}F-ratio calculation is valuable as it improves discoverability ranking hits more effectively. However, with F-ratios calculated at each *m/z* and the necessity of multiple replicates per class, F-ratio analysis cannot be readily performed on raw LC-HRMS data.

Specifically, herein we optimize the workflow for tile-based F-ratio analysis, applying the ^{MVO}F-ratio calculation, of low concentration spiked pesticides in soil samples analyzed by LC-HRMS compressed by the ROI procedure. Spiked versus neat samples are used to evaluate the workflow, and we envisage this methodology will be beneficial for the implementation of a wide variety of non-targeted studies using chemical separations with HRMS detection. To begin, F-ratio analysis was performed on samples compressed using the previously reported ROI signal threshold and admissible mass deviation parameters, referred to herein as the *initial* ROI parameter [30,40]. The ROI compression method was first applied to the raw data using initial parameters including an admissible mass deviation that is a multiple of the detector accuracy,

with ±0.05 Da previously selected based upon minimizing computation time and the number of ROIs extracted [30,40]. The recommended signal threshold was between 0.1% and 1% of the maximum intensity, with 0.1% selected in an attempt to extract analyte signals that are the least sensitively detected and/or present at the lowest concentrations. Due to the low spike concentrations in our present study, the initial ROI parameters required further optimization to ensure analyte discovery with the subsequent F-ratio analysis. To demonstrate this workflow, parameter selection was accomplished in a targeted fashion where a known set of analytes is spiked in a representative soil sample. The pair of parameters that allows for the detection of most analytes with the lowest average Δppm (< 20 ppm) is selected as optimal for the subsequent non-targeted studies, where Δppm is the absolute value of the difference between the theoretical *m/z* and the ROI *m/z*.

2.2 Experimental

To demonstrate the optimization of this workflow for trace-level analyses, soil samples spiked with pesticides at several concentrations were prepared. Soil samples were obtained from the area surrounding the Los Alamos National Laboratory (New Mexico, USA) and spiked with the GC Multiresidue Pesticide Standard #5 obtained from Restek (Bellefonte, PA, USA) to obtain nominal concentrations of 0.1 parts-per-billion (ppb), 0.5 ppb, 1 ppb, 10 ppb, 25 ppb, and 50 ppb for each analyte, comprising the six sample classes. Five LC-QTOF chromatograms were collected per spiked class and neat soil sample. Details of sample preparation, data collection, and identities of spiked pesticides are provided in Table A.1 in Appendix A.

All data was imported to Matlab2020b (MathWorks Inc., Natick, MA, USA) where it was processed using the ROI procedure [29,30,41] prior to performing in-house tile-based F-ratio analysis [16,17,39]. The ROI workflow is presented in Figure 2.1, with a schematic describing

each step of the data processing prior to F-ratio analysis. This workflow requires no additional pre-processing and effectively compresses HRMS data without losing the detector mass accuracy. First, regions of interest are found within an individual replicate (step 1 in Figure 2.1). Since HRMS data provides a unique number of measured *m/z* and signal intensity pairs at each retention time, the workflow searches for ROIs in the raw data in regions of high-density signals. The user defines the minimum number of retention times that make up a ROI, the admissible mass deviation (*m/z* error), and a signal threshold as parameters for defining ROIs.

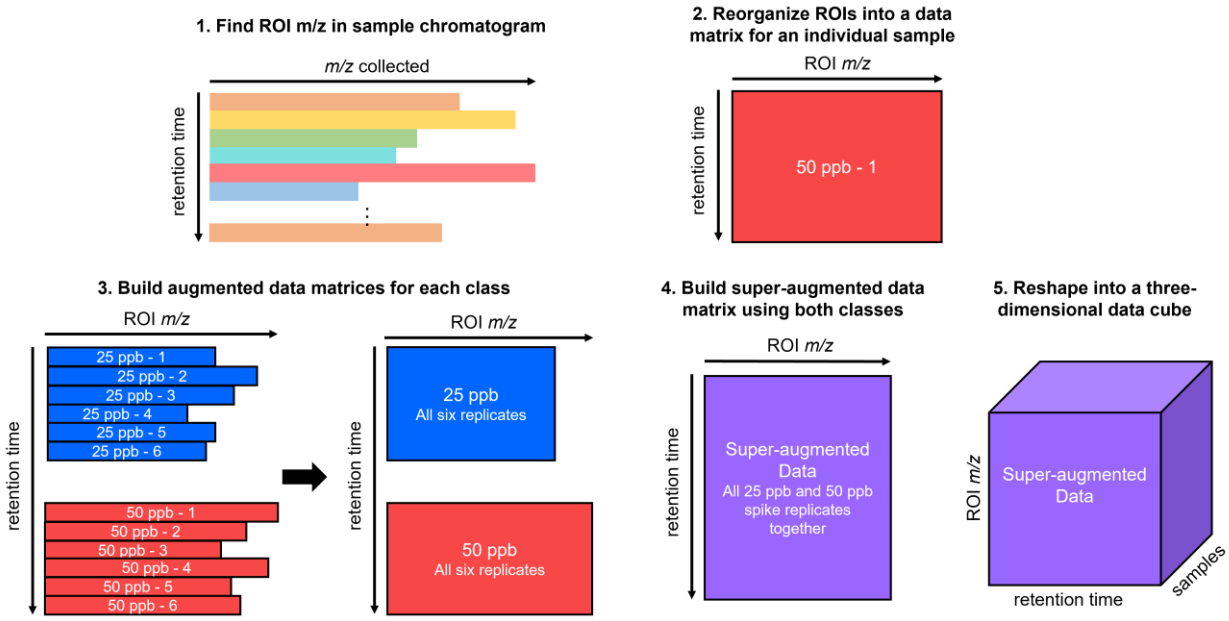


Figure 2.1. Schematic demonstrating the steps of the region of interest (ROI) methodology used to compress LC-HRMS data prior to F-ratio analysis.

Visual representation of these parameters is provided in Figure 2.2 for a simulated analyte. For a single analyte, with Figure 2.2A representing the raw data, only *m/z* with signal above the threshold will be extracted. Furthermore, for each *m/z* only the signal intensity pairs that also meet the minimum number of consecutive retention times and admissible mass deviation (*m/z* error) requirements (Figure 2.2B) are used to calculate the average ROI *m/z*. The result (step 2) is a reduced data matrix with all ROIs that meet the user-defined parameters, containing the new

ROI *m/z*. Note that individual ROI matrices can have a unique combination of ROI *m/z*; therefore, an augmentation step is required to provide the same *m/z* dimension across replicates within a sample class (step 3). To accomplish this step, the user must again define an appropriate signal threshold and *m/z* error. Each augmented data matrix per class may have a unique set of ROI *m/z*, requiring a second augmentation step (step 4) to build a super-augmented data matrix with the same *m/z* dimension prior to F-ratio analysis. This step was achieved by modifying the algorithm to augment all six classes simultaneously rather than sequentially. Finally, the super-augmented data matrix can be folded into a three-way array (step 5) for chemometric analysis including PCA and tile-based F-ratio analysis. While previous studies using ROI compression did not require extensive parameter optimization for the intended analyses, the comparative analysis herein to discover trace level analytes relied on optimizing parameter selection.

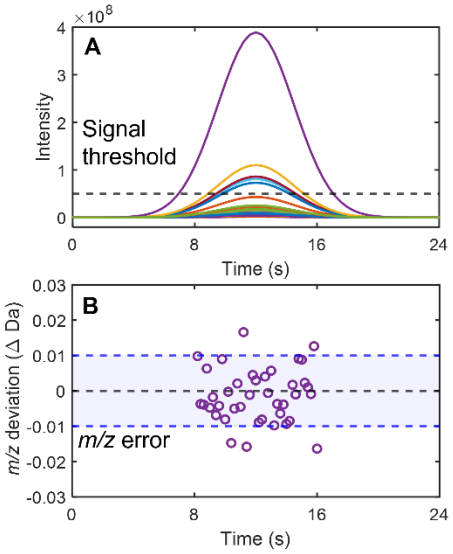


Figure 2.2. Illustration of parameters to define for ROI data compression. (A) The analyte is plotted as an overlay of all *m/z*. (B) Using the most sensitive *m/z* (purple trace) as an example, the signal from all *m/z* within the *m/z* error of ±0.01 Da (dashed blue lines) are summed to form the trace in (A).

Parameter optimization (Figure 2.1) involves spiking known analytes into a sample that represents the overall complexity of the sample set, in this case a soil sample, and applying the

ROI procedure to determine optimal parameters in a targeted fashion. Subsequently, the optimized parameters are applied to the data set in the intended non-targeted analysis. Specifically, the user-defined signal threshold is examined in step 1 using an *m/z* error of ±0.01 Da which was selected based on the detector mass accuracy. For all data compression, a minimum of 10 consecutive retention times was selected for defining a ROI. The next step that requires parameter inputs is step 3, the first augmentation step, in which the same signal threshold and *m/z* error from step 1 are applied. Next, in step 4, *m/z* error optimization occurs with the user-defined *m/z* error examined. For each pair of signal threshold and *m/z* error examined, the number of spiked analytes detected and average Δppm were calculated for each combination of *m/z* errors (±0.001 Da to ±0.05 Da) and signal thresholds (62.5 to 0.1% of max intensity). From a fundamental standpoint, one might be concerned with a fixed signal threshold due to baseline changes throughout a chromatogram, however the fixed signal threshold allows for simultaneous compression of the entire chromatograms. ROI compression is generally used in time segments, with different thresholds selected based on the number of ROIs extracted, and with chemometric decomposition for obtaining pure spectra an elution profiles to aide quantitation and identification efforts [42]. However, this method requires an initial guess of the number of components to model and is prone to splitting peaks, requiring time segment overlap and additional processing steps (i.e. removing redundant analyte peaks, augmentation to rebuild chromatograms). Additionally, *m/z* that simply capture noise in the chromatogram using a fixed signal threshold are not discovered as significant features if non-targeted methods such as when F-ratio analysis is applied. The Δppm (average of all detected compounds) is calculated as the absolute value of the difference between the theoretical *m/z* and the ROI *m/z*.

First, parameter optimization was evaluated using only the 50 ppb and 1 ppb spiked samples relative to the neat soil. Subsequently, F-ratio analysis before and after optimization was applied to all spiked classes to challenge the discovery of low concentration spikes. The F-ratio tile size is selected to fully capture the analyte peak and account for retention time misalignment. A tile size of 7.5 s was selected based on the average peak width-at-base of the analytes, which is described in greater detail in Figure A.1 in Appendix A. Hit lists were ranked by F-ratios calculated using only the top F-ratio *m/z* passing a *S/N* threshold of 10, as this is beneficial for low concentration comparisons [43]. PCA was then applied on mean-centered data of all six sample classes for visualizing the success of applying optimal parameters.

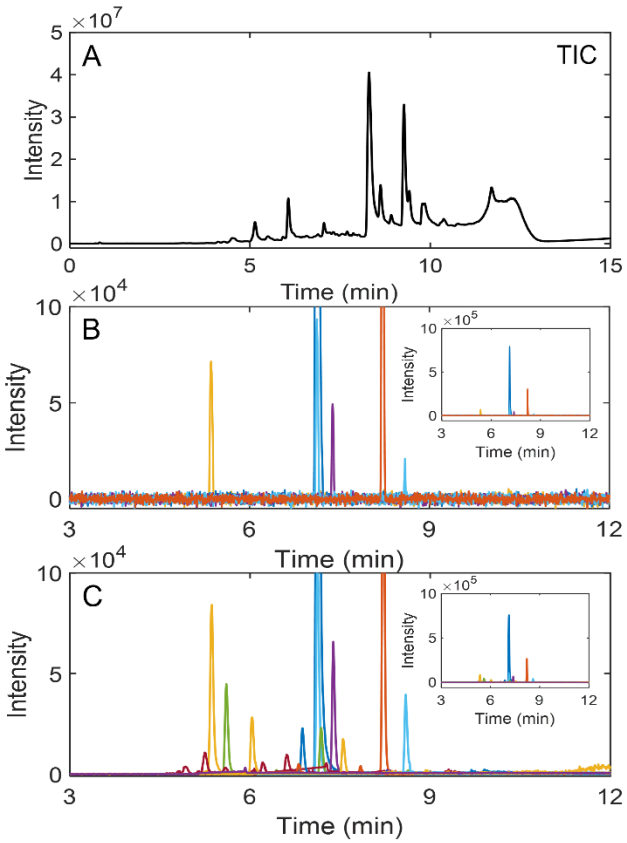


Figure 2.3. Sample chromatograms before and after parameter optimization. (A) Average TIC chromatogram of all the neat soil samples after the ROI workflow using the optimal parameters. An overlay of the spiked pesticides using selective *m/z* that were discovered as ROIs using (B) the initial parameters and (C) optimal parameters during the ROI workflow for a typical 50 ppb spiked replicate.

2.3 Results and Discussion

We begin by demonstrating the initial parameters versus the optimized parameters for ROI data compression of an entire LC-HRMS chromatogram for a soil sample spiked with pesticides at trace levels (Figure 2.3). We do so in order to frame the issue of parameter optimization before getting into the details of how the optimization was performed. A total ion current (TIC) chromatogram of the neat soil sample obtained by LC-QTOF using the ROI workflow applying the optimal signal threshold of 250 and *m/z* error ±0.003 Da (vide infra) is provided (Figure 2.3A) to visualize the high intensity components native to the soil sample that may obscure discovery of lower intensity analytes. The pesticides discovered by the ROI data compression workflow using the initial parameters (*m/z* error ±0.05 Da, and a signal threshold of 0.1% of the maximum chromatogram) are overlaid for a single 50 ppb spiked replicate using analyte selective *m/z* (Figure 2.3B), with results for the same replicate provided in Figure 2.3C using the optimal parameters. Ideally, using optimized parameters for ROI data compression all spiked pesticides should be observed in the chromatograms. However, the pesticides observed in Figure 2.3B constitute only a small subset of those observed in Figure 2.3C. Most of the pesticides that are not ROIs in Figure 2.3B but are ROIs in Figure 2.3C are the lower intensity analytes and the signal for some of the other analytes in Figure 2.3B is lower than their signal in Figure 2.3C. Lowering the initial ROI parameters to an optimal level allowed for improved detection of low intensity spiked pesticides with ROI *m/z* closer to the theoretical *m/z*. It appears that the initial parameters [30,40] are better suited for either targeted approaches using subsections of chromatograms or of analytes with higher signal.

Using the initial ROI parameters for data compression, we generated the hit list for the F-ratio analysis for all spiked sample classes (0.5 ppb, 1 ppb, 10 ppb, 25 ppb, and 50 ppb) in order to highlight the poor obtained (Table 2.1) due to the shortcomings of the data pre-processing. The

hits are numbered, ranked by F-ratio, and the *m/z* that was used for analyte identification with corresponding Δppm. The complete hit list using initial parameters is provided in Table A.2 in Appendix A. Only three pesticides were discovered with calculated p-values < 0.05 using a student's t-test using the experimental *m/z*. Additionally, while the identification *m/z* for both fludioxonil (Hit# 1) and fipronil (Hit# 3) are very similar to the expected $[M-H]^-$ with Δppm < 8 ppm, identification of tricyclazole in a truly non-targeted study would be less confident with a Δppm of 56.0 ppm. It would be expected that a large signal threshold and large allowable mass deviation would greatly reduce the discoverability of low signal analytes and inflate the Δppm due to a larger range of *m/z* being averaged for each ROI. These results indicate that non-targeted analyses would benefit greatly from optimization of ROI parameters.

Table 2.1. F-ratio hit list for the comparison of all spiked classes using data compressed by ROI using initial parameters. Hits are ranked by Hit# (high to low F-ratios) with *m/z* used for confirming analyte identity and corresponding Δppm.

Hit#	Compound name	Identification *m/z* (Δppm)
1	fludioxonil	247.0343 (7.6)
2	tricyclazole	248.0361 (56.0)
3	fipronil	434.9322 (1.8)

Parameter optimization relies upon using a neat, representative sample versus the same representative nest sample spiked with known analytes. Specifically, in this study the neat soil and soil spiked with pesticides at a concentration of 50 ppb and 1 ppb (Figure 2.4) so the process can be demonstrated twice. First, using data from the 50 ppb spiked and neat samples, for each set of parameters the number of detected spikes is calculated (Figure 2.4A) along with the average Δppm for all detected spikes (Figure 2.4B). These metrics were calculated for all *m/z* error and signal threshold combinations in Figure 2.4. The optimal parameters correspond to the detection of most spiked analytes and an average Δppm < 20 ppm, with the lowest Δppm being optimal. This ensures the greatest chance of detecting all other unknown analytes in the sample

as well as improved confidence in identification in non-targeted analyses. As this may be time-consuming to calculate values for each combination, a more direct approach can be used. The number of detected spikes and Δppm metric are calculated using an initial set of parameters and this is repeated decreasing the signal threshold and *m/z* error until no additional analytes are detected and the Δppm does not improve (lowest Δppm). In this this case, with initial parameters ±0.05 Da and 0.1% of the maximum signal, only 6/18 spiked analytes were detected and the average Δppm was 22.4 ppm, which is not ideal for non-targeted identification. Next, the nearest three cells are investigated, moving toward lower *m/z* error and lower signal threshold: (1) ±0.01 Da and 0.1% maximum intensity, (2) ±0.05 Da and 0.05% maximum intensity, and (3) ±0.01 Da and 0.05% maximum intensity. For (1) and (2) the average Δppm was below 20 ppm but 4 additional spiked compounds were detected using (1), making it the next best set. This same process was repeated until the optimal parameters of ±0.003 Da and signal threshold of 250 were selected with all analytes detected with an average Δppm of 6.6 ppm. Although using lower signal thresholds and *m/z* errors still allows for the discovery of 18 analytes, the neighboring cells do not provide lower average Δppm. This substantially reduces the number of metrics that need to be calculated, from 35 signal threshold to *m/z* error combinations to only 15. To reinforce selection of this set of parameters, equivalent plots are provided in Figures 2.4C and 2.4D using the 1 ppb and neat samples. Using the same initial set of parameters there are no analytes detected, most likely due to the signal threshold being too high, and therefore an average Δppm could not be calculated. Therefore, the signal threshold was lowered until an analyte is detected (±0.05 Da and signal threshold 1000). Applying the same method, with 4 steps the optimal parameters are again determined to be ±0.003 Da *m/z* error and signal threshold 250. The maximum number of analytes (11) were detected with Δppm 15.6 ppm. Note that using the

lowest *m/z* error and signal threshold combination may provide similar results, however a larger number of ROI *m/z* are extracted making chemometric analysis more computationally expensive.

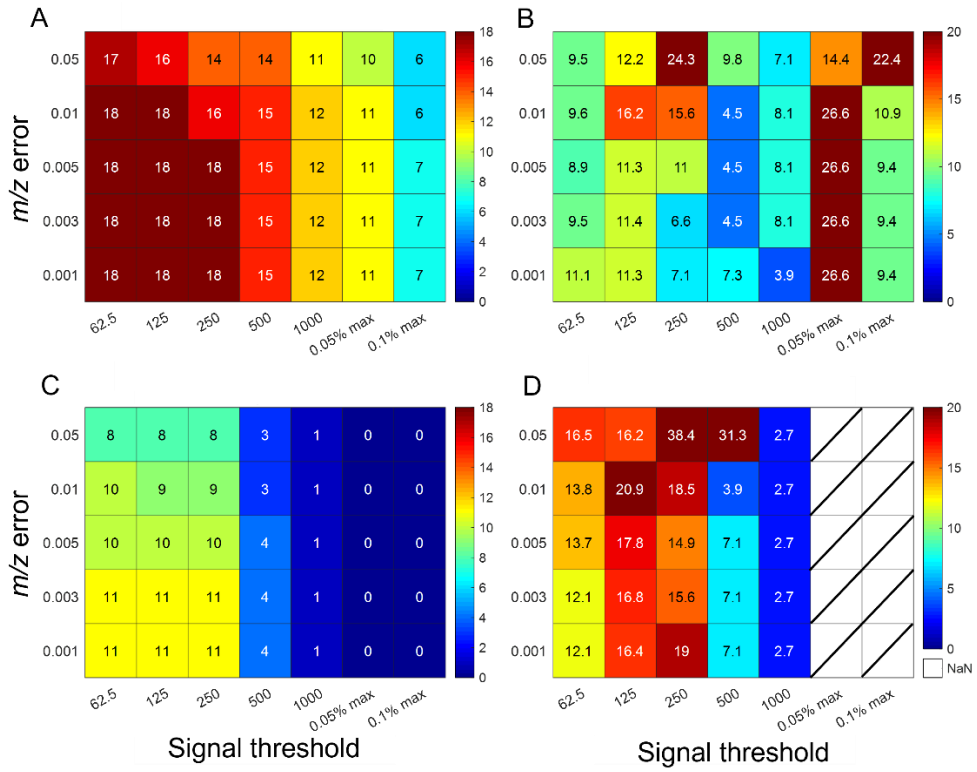


Figure 2.4. ROI parameter optimization. (A) The number of spikes detected and (B) average Δppm based upon the 50 ppb spiked and neat sample chromatograms using the ROI data compression for each combination of signal threshold and *m/z* error. (C) and (D) are the equivalent plots based upon the 1 ppb and neat sample chromatograms.

Following optimization of the admissible mass deviation and signal threshold, F-ratio analysis of all spiked sample classes was investigated to confirm the improvement in analyte discoverability. The F-ratio results using ROI data obtained applying an optimal *m/z* error of ±0.003 Da in the final augmentation step and signal threshold of 250 are provided in Table 2.2 (more details provided in the hit list in Table A.3 in Appendix A). The hit list is reduced to the true positives discovered by F-ratio analysis and is organized in the same manner as Table 2.1. Using this set of parameters, the eighteen spiked pesticides are all discovered in the top 30 hits

and are statistically significant using the student's t-test (p-value < 0.05). The remaining hits in the list excluded from Table 2.2 were of unknown identity, with no *m/z* in the tile that corresponded to expected $[M-H]^-$, $[M+Hac-H]^-$, $[M-CO_2-H]^-$, or $[M-H_2O-H]^-$ *m/z* for any of the spiked pesticides. The average Δppm is 5 ppm, demonstrating that the ROI workflow using optimal conditions provided high quality data for improved confidence in analyte identification for these low concentration analytes. A closer look at some of the hits that range from the top to the bottom of the hit list gives insight to the success of F-ratio analysis using these optimal parameters. Even though in a truly non-targeted study analyte identification is not known a priori, since the spiked analytes are known herein this F-ratio analysis provides validation of the ROI workflow.

Table 2.2. Tile-based F-ratio analysis of all spiked soil samples processed by the ROI workflow using parameters that were determined to be optimal.

Hit#	Compound name	Identification *m/z* (Δppm)
1	fludioxonil	247.0337 (5.2)
2	fenarimol	329.0260 (1.8)
3	lenacil	233.1292 (1.4)
4	terbacil	215.0604 (5.0)
5	flutriafol	300.0966 (3.9)
6	tricylcazole	248.0499 (0.2)
7	paclobutrazol	352.1440 (1.8)
8	triadimefon	292.0858 (0.1)
10	triadimenol	354.1245 (5.2)
11	firponil	434.9318 (0.9)
12	tebuconazole	366.1657 (18.4)
13	triflumizole	344.0783 (0.7)
14	myclobutanil	347.1295 (4.2)
15	hexazinone	251.1512 (0.5)
17	propargite	409.1769 (19.2)
18	flusilazole	314.0932 (0.6)
19	procymidone	282.0103 (3.1)
23	pyrimethanil	258.1298 (19.6)

A subset of six hits from Table 2.2 is provided in Figure 2.5 as an overlay of the 50 ppb replicates. More specifically, Hit# 1 fludioxonil (Figure 2.5A), Hit# 7 paclobutrazol (Figure 2.5B), Hit# 10 triadimenol (Figure 2.5C), Hit# 15 hexazinone (Figure 2.5D), Hit# 18 flusilazole (Figure 2.5E), and Hit# 23 pyrimethanil (Figure 2.5F) are provided to represent a broad range of sensitivity. The most sensitive analyte, fludioxonil, is discovered as Hit# 3 in Table 2.2 and was discovered in all F-ratio hit lists as the raw signal was high enough to meet the signal threshold regardless of the admissible mass deviation selected. However, many pesticides were not as sensitively detected, which can be seen in calibration curves for representative analytes (Figure A.2 in Appendix A). Hexazinone, triadimenol, paclobutrazol, pyrimethanil, and flusilazole are examples of less sensitively detected pesticides newly discovered in Table 2.2 compared to using the initial parameters. These newly discovered hits are identified with $\Delta ppm < 20$ ppm, indicating confident identification can be achieved using this workflow in a non-targeted study. F-ratio hit lists tuning one parameter at a time were also compared (Tables A.4-A.7 in Appendix A), with the optimized workflow providing the best results, indicating optimization of both parameters was critical for successful F-ratio analysis.

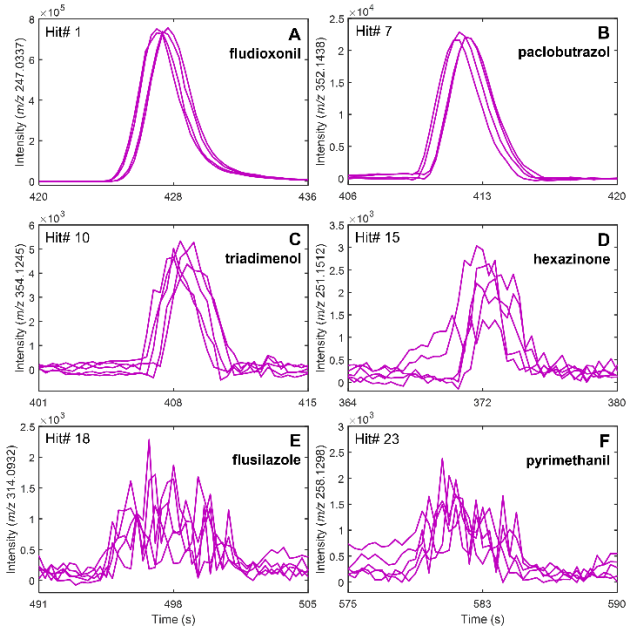


Figure 2.5. Using selective m/z, peak profiles of a subset of true positives are provided. The overlay of all five replicates of the 50 ppb replicates are provided for fludioxonil (A), paclobutrazol (B), triadimenol (C), hexazinone (D), flusilazole (E), and pyrimethanil (F). Hit# is provided in the top left corners.

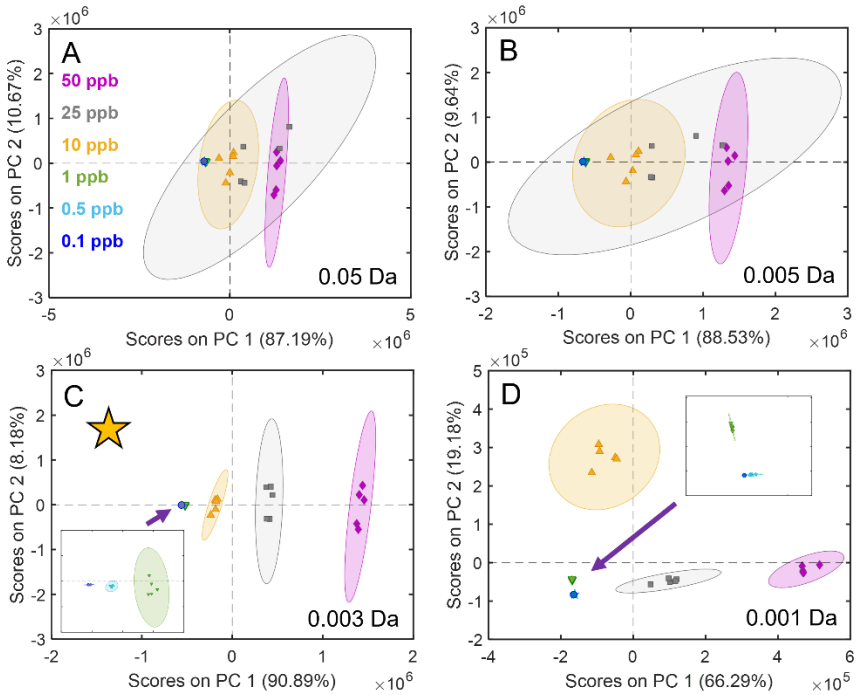


Figure 2.6. Visualization of optimal m/z error selection. (A) Optimal *m/z* error selection was visualized by PCA of all six sample classes processed with *m/z* error values (A) ±0.03 Da, (B) ±0.005 Da, (C) ±0.003 Da, and (D) ±0.001 Da to capture a large range of values. Confidence ellipses are provided (95% Confidence Interval) with excellent sample clustering using an optimal *m/z* error of ±0.003 Da (starred).

To better understand why the optimal parameters were successful for the subsequent non-targeted analysis, PCA was applied to the hits in Table 2.2 using all spiked samples and scores plots were used for visualization. This was done varying the *m/z* error while using the optimal signal threshold of 250. Using all replicates per class, the peak profiles for all discoverable pesticides using selective *m/z* were used for PCA using an *m/z* error of ±0.05 Da (Figure 2.6A), ±0.005 Da (Figure 2.6B), ±0.003 Da (Figure 2.6C), and ±0.001 Da (Figure 2.6D). Higher concentration samples are best clustered using ±0.003 Da (Figure 2.6C) and ±0.001 Da (Figure 2.6D), however examination of the low concentration samples indicates using ±0.003 Da (inset of Figure 2.6C) as the optimal *m/z* error (starred) with all classes well clustered, while the 0.5 ppb and 0.1 ppb classes are not easily distinguishable using ±0.001 Da (inset of Figure 2.6D). Additionally, while virtually all of the variance is captured by PC 1 and PC 2 for most of the scores plots, in Figure 2.6C over 90% of the variance is explained by PC 1 and can be related to the concentration of the sample classes (increasing concentration from left to right). PC 2 captures the variance within a class including slight misalignment of peaks, which the tile-based approach addresses for effective F-ratio analysis. The trend on PC 1 using ±0.003 Da is no longer clear when deviating from this *m/z* error. Using an *m/z* error above ±0.003 Da results in summing the signal contributions from a broader range of *m/z* that may correspond to analytes other than the one of interest and therefore the summed signal no longer reproducibly reflects the actual concentration of the analyte in the class. This is evident in (Figures 2.6A-2.6B) where there is no clear distinction between sample classes, particularly the higher concentration classes (Figure A.3 in Appendix A), hindering accurate quantification in non-targeted studies. Alternatively, using a lower than optimal *m/z* error of ±0.001 Da (Figure 2.6D), PC 1 captures 66% of the variance and there is no longer a clear trend in concentration, resulting in peak signal

splitting among several *m/z* (Figure A.4 in Appendix A).While this was visualized tuning only one parameter, it is advised to optimize both parameters simultaneously with spiked standards as in Figure 2.4 to ensure that the data is accurately represented for the intended analysis, especially for low concentration analytes of interest. Furthermore, we expect this workflow to be readily adapted to comprehensive two-dimensional liquid chromatography (LC×LC) [44] with HRMS data as both the ROI procedure and tile-based F-ratio analysis have been demonstrated for various comprehensive two-dimensional chromatography data applications [33,34,39,45–47].

2.4 Conclusion

Optimization of parameter selection for entire chromatograms using for ROI data compression was demonstrated, and successfully applied to a subsequent F-ratio analysis of trace level analytes. Specifically, a targeted parameter optimization method is described using Δppm and the number of detected spiked compound as metrics to select optimal ROI parameters independent of the non-targeted analysis. While demonstrated successfully for a supervised non-targeted study, we envisage that for an unsupervised non-targeted analysis the analyst could utilize a pooled sample capturing the complexity of all the samples and then spike standards into the pooled sample for parameter optimization prior to non-targeted analysis on the entire data set for all of the samples. Proper selection of the ROI signal threshold and admissible mass deviation was most essential for analyte discovery, while the selection of the F-ratio tile size was relatively straight-forward due to obtaining pure high-resolution *m/z* for each spiked pesticide. This ROI procedure optimization was demonstrated for samples spiked as low as 0.5 ppb, with no analytes detectable at the 0.1 ppb level. Using the initial parameters only three pesticides were discovered by F-ratio analysis (Δppm = 22 ppm) compared to 18 discovered using optimal parameters (Δppm = 5 ppm). The hit list obtained using the optimized workflow ranked all spiked analytes more effectively with greatest confidence in identification due to low Δppm. For similar trace level studies, we emphasize the importance of proper signal threshold and admissible mass deviation parameter selection achieved simultaneously for ensuring the meaningful chemical differences may be discoverable.

2.5 Acknowledgment

C.N. Cain acknowledges the U.S. National Science Foundation Graduate Research Fellowship (DGE-1762114).

2.6 References

[1] L.L. Hohrenk, M. Vosough, T.C. Schmidt, Implementation of Chemometric Tools to Improve Data Mining and Prioritization in LC-HRMS for Nontarget Screening of Organic Micropollutants in Complex Water Matrixes, Anal. Chem. 91 (2019) 9213–9220. doi:10.1021/acs.analchem.9b01984.

[2] M.M. Sinanian, D.W. Cook, S.C. Rutan, D.S. Wijesinghe, Multivariate curve resolution-alternating least squares analysis of high-resolution liquid chromatography-mass spectrometry data, Anal. Chem. 88 (2016) 11092–11099. doi:10.1021/acs.analchem.6b03116.

[3] J.E. Schollée, E.L. Schymanski, S.E. Avak, M. Loos, J. Hollender, Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic, Anal. Chem. 87 (2015) 12121–12129. doi:10.1021/acs.analchem.5b02905.

[4] K. Purschke, M. Vosough, J. Leonhardt, M. Weber, T.C. Schmidt, Evaluation of Nontarget Long-Term LC-HRMS Time Series Data Using Multivariate Statistical Approaches, Anal. Chem. 92 (2020) 12273–12281. doi:10.1021/acs.analchem.0c01897.

[5] A.Y. Mohd Yusop, L. Xiao, S. Fu, Determination of phosphodiesterase 5 (PDE5)inhibitors in instant coffee premixes using liquid chromatography-high-resolution mass spectrometry (LC-HRMS), Talanta. 204 (2019) 36–43. doi:10.1016/j.talanta.2019.05.078.

[6] M.C. Alexandre-Gouabau, F. Courant, T. Moyon, A. Küster, G. Le Gall, I. Tea, J.P. Antignac, D. Darmaun, Maternal and cord blood LC-HRMS metabolomics reveal alterations in energy and polyamine metabolism, and oxidative stress in very-low birth weight infants, J. Proteome Res. 12 (2013) 2764–2778. doi:10.1021/pr400122v.

[7] A.E. Schwaninger, M.R. Meyer, M.A. Huestis, H.H. Maurer, Development and validation of LC-HRMS and GC-NICI-MS methods for stereoselective determination of MDMA and its phase i and II metabolites in human urine, J. Mass Spectrom. 46 (2011) 603–614. doi:10.1002/jms.1929.

[8] Y.Y. Fan, H. Neubert, Quantitative Analysis of Human Neonatal Fc Receptor (FcRn) Tissue Expression in Transgenic Mice by Online Peptide Immuno-Affinity LC-HRMS, Anal. Chem. 88 (2016) 4239–4247. doi:10.1021/acs.analchem.5b03900.

[9] E.L. Jamin, N. Bonvallot, M. Tremblay-Franco, J.P. Cravedi, C. Chevrier, S. Cordier, L. Debrauwer, Untargeted profiling of pesticide metabolites by LC-HRMS: An exposomics tool for human exposure evaluation, Anal. Bioanal. Chem. 406 (2014) 1149–1161. doi:10.1007/s00216-013-7136-2.

[10] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1987) 37–52. doi:https://doi.org/10.1016/0169-7439(87)80084-9.

[11] N. Núñez, O. Vidal-Casanella, S. Sentellas, J. Saurina, O. Núñez, Characterization, classification and authentication of turmeric and curry samples by targeted LC-HRMS polyphenolic and curcuminoid profiling and chemometrics, Molecules. 25 (2020) 1–16. doi:10.3390/molecules25122942.

[12] J.H. Ward, Hierarchical Grouping to Optimize an Objective Function, J. Am. Stat. Assoc. 58 (1963) 236–244. doi:10.1080/01621459.1963.10500845.

[13] E. Ortiz-Villanueva, J. Jaumot, R. Martínez, L. Navarro-Martín, B. Piña, R. Tauler, Assessment of endocrine disruptors effects on zebrafish (Danio rerio) embryos by untargeted LC-HRMS metabolomic analysis, Sci. Total Environ. 635 (2018) 156–166. doi:10.1016/j.scitotenv.2018.03.369.

[14] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemom. 17 (2003) 166–173. doi:10.1002/cem.785.

[15] F. Souard, C. Delporte, P. Stoffelen, E.A. Thévenot, N. Noret, B. Dauvergne, J.-M. Kauffmann, P. Van Antwerpen, C. Stévigny, Metabolomics fingerprint of coffee species determined by untargeted-profiling study using LC-HRMS, Food Chem. 245 (2018) 603–612. doi:10.1016/j.foodchem.2017.10.022.

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

[17] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC-TOFMS) Data Using a Null Distribution Approach, Anal. Chem. 87 (2015) 3812–3819. doi:10.1021/ac504472s.

[18] 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.

[19] G.S. Ochoa, S.E. Prebihalo, B.C. Reaser, L.C. Marney, R.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, J. Chromatogr. A. 1627 (2020) 461401. doi:10.1016/j.chroma.2020.461401.

[20] L. Stahle, S. Wold, Analysis of variance (ANOVA), Chemom. Intell. Lab. Syst. 6 (1989) 259–272. doi:10.1016/0169-7439(89)80095-4.

[21] B.J. Feir-walsh, An Emprircal Comparison of the ANOVA F-test Normal Scores Test and Krsukal-Wallis Test Under Violation of Assumptions, Educ. Psychol. Meas. 34 (1974) 789–799.

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

[23] 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.

[24] T.S. Bos, W.C. Knol, S.R.A. Molenaar, L.E. Niezen, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Recent applications of chemometrics in one- and two-dimensional chromatography, J. Sep. Sci. 43 (2020) 1678–1727. doi:10.1002/jssc.202000011.

[25] D.W. Cook, S.C. Rutan, Analysis of Liquid Chromatography-Mass Spectrometry Data with an Elastic Net Multivariate Curve Resolution Strategy for Sparse Spectral Recovery, Anal. Chem. 89 (2017) 8405–8412. doi:10.1021/acs.analchem.7b01832.

[26] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinformatics. 11 (2010) 395. doi:10.1186/1471-2105-11-395.

[27] A. Lommen, Metalign: Interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing, Anal. Chem. 81 (2009) 3079–3086. doi:10.1021/ac900036d.

[28] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, Anal. Chem. 78 (2006) 779–787. doi:10.1021/ac051437y.

[29] E. Gorrochategui, J. Jaumot, R. Tauler, A protocol for LC-MS metabolomic data processing using chemometric tools, Protoc. Exch. (2015) 1–41. doi:10.1038/protex.2015.102.

[30] E. Gorrochategui, J. Jaumot, R. Tauler, ROIMCR: A powerful analysis strategy for LC-MS metabolomic datasets, BMC Bioinformatics. 20 (2019) 1–17. doi:10.1186/s12859-019-2848-8.

[31] C.E. Freye, P.R. Bowden, M.T. Greenfield, B.C. Tappan, Non-targeted discovery-based analysis for gas chromatography with mass spectrometry: A comparison of peak table, tile, and pixel-based Fisher ratio analysis, Talanta. 211 (2020) 120668. doi:10.1016/j.talanta.2019.120668.

[32] O.D. Myers, S.J. Sumner, S. Li, S. Barnes, X. Du, One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks, Anal. Chem. 89 (2017) 8696–8703. doi:10.1021/acs.analchem.7b00947.

[33] M. Navarro-Reig, J. Jaumot, R. Tauler, An untargeted lipidomic strategy combining comprehensive two-dimensional liquid chromatography and chemometric analysis, J. Chromatogr. A. 1568 (2018) 80–90. doi:10.1016/j.chroma.2018.07.017.

[34] M. Navarro-Reig, J. Jaumot, A. Baglai, G. Vivó-Truyols, P.J. Schoenmakers, R. Tauler, Untargeted Comprehensive Two-Dimensional Liquid Chromatography Coupled with High-Resolution Mass Spectrometry Analysis of Rice Metabolome Using Multivariate Curve Resolution, Anal. Chem. 89 (2017) 7675–7683. doi:10.1021/acs.analchem.7b01648.

[35] K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: Comprehensive GC × GC with ANOVA-based feature selection and principal component analysis, Chemom. Intell. Lab. Syst. 60 (2002) 225–237. doi:10.1016/S0169-7439(01)00198-8.

[36] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher ratio method applied to third-order separation data to identify significant chemical components of metabolite extracts, Anal. Chem. 78 (2006) 5068–5075. doi:10.1021/ac0602625.

[37] C.N. Cain, N.J. Haughn, H.J. Purcell, L.C. Marney, R.E. Synovec, C.T. Thoumsin, S.C. Jackels, K.J. Skogerboe, Analytical Determination of the Severity of Potato Taste Defect in Roasted East African Arabica Coffee, J. Agric. Food Chem. 69 (2021) 2253–2261. doi:10.1021/acs.jafc.1c00605.

[38] T.J. Trinklein, R.E. Synovec, Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based Fisher ratio analysis, J. Chromatogr. A. 1677 (2022) 463321. doi:10.1016/j.chroma.2022.463321.

[39] S. Schöneich, G.S. Ochoa, C.M. Monzón, R.E. Synovec, Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography / mass spectrometry data: Study of the pacu fish metabolome, J. Chromatogr. A. 1667 (2022) 462868. doi:10.1016/j.chroma.2022.462868.

[40] M. Pérez-Cova, C. Bedia, D.R. Stoll, R. Tauler, J. Jaumot, MSroi: A pre-processing tool for mass spectrometry-based studies, Chemom. Intell. Lab. Syst. 215 (2021) 104333. doi:10.1016/j.chemolab.2021.104333.

[41] C. Perez-Lopez, A. Ginebreda, M. Carrascal, D. Barcelò, J. Abian, R. Tauler, Non-target protein analysis of samples from wastewater treatment plants using the regions of interest-multivariate curve resolution (ROIMCR) chemometrics method, J. Environ. Chem. Eng. (2021) 105752. doi:10.1016/j.jece.2021.105752.

[42] R.S.M. Pourasil, J. Cristale, S. Lacorte, R. Tauler, Non-targeted Gas Chromatography Orbitrap Mass Spectrometry qualitative and quantitative analysis of semi-volatile organic compounds in indoor dust using the Regions of Interest Multivariate Curve Resolution chemometrics procedure, J. Chromatogr. A. 1668 (2022) 462907. doi:10.1016/j.chroma.2022.462907.

[43] P.E. Sudol, G.S. Ochoa, R.E. Synovec, Investigation of the limit of discovery using tile-based Fisher ratio analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A. 1644 (2021) 462092. doi:10.1016/j.chroma.2021.462092.

[44] B.W.J. Pirok, D.R. Stoll, P.J. Schoenmakers, Recent Developments in Two-Dimensional Liquid Chromatography: Fundamental Improvements for Practical Applications, Anal. Chem. 91 (2019) 240–263. doi:10.1021/acs.analchem.8b04841.

[45] M. Pérez-Cova, J. Jaumot, R. Tauler, Untangling comprehensive two-dimensional liquid chromatography data sets using regions of interest and multivariate curve resolution approaches, TrAC Trends Anal. Chem. 137 (2021) 116207. doi:10.1016/j.trac.2021.116207.

[46] T.J. Trinklein, J. Jiang, R.E. Synovec, Profiling Olefins in Gasoline by Bromination Using GC×GC-TOFMS Followed by Discovery-Based Comparative Analysis, Anal. Chem. 94 (2022) 9407–9414. doi:10.1021/acs.analchem.2c01549.

[47] P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis, J. Chromatogr. A. 1662 (2022) 462735. doi:10.1016/j.chroma.2021.462735.

Chapter 3: Chemometric decomposition of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data employing partial modulation in the negative pulse mode

This chapter was reproduced from D. V. Gough[†], S. Schöneich[†], R.E. Synovec, Chemometric decomposition of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data employing partial modulation in the negative pulse mode, Talanta. 210 (2020) 120670.

[†] These authors contributed equally to this work

3.1 Introduction

Gas chromatography (GC) is a powerful separation tool that enables the gas phase separation of volatile and semi-volatile analytes. The efficacy of GC has significantly increased with the advent of comprehensive, two-dimensional (2D) gas chromatography (GC×GC) pioneered by Liu and Phillips [1], since significantly more peak capacity [2] and chemical selectivity [3–5] are provided by GC×GC relative to one-dimensional GC. GC×GC instruments require a modulator between two successive capillary columns in order to comprehensively couple the second dimension (^{2}D) separation to the first dimension (^{1}D) separation [6–8]. Furthermore, the ^{2}D column must be equipped with a stationary phase sufficiently different from the ^{1}D column so as to provide orthogonal selectivity [5]. The growth of GC×GC has been significant in recent years, especially when coupled with time-of-flight mass spectrometry (TOFMS) detection [5,6,9–16].

The modulator is the key instrument component that enabled the jump from one-dimensional GC to GC×GC. Currently, there are two broad categories of modulators: thermal and flow-based modulators [6,8,12,17–21]. Both modulator categories generally provide “full” modulation [17], resulting in a series of ^{2}D peaks for a given ^{1}D peak in which baseline is

observed between adjacent ^{2}D peaks for a given analyte. The series of narrow ^{2}D peaks follow the same general shape of the ^{1}D peak. When each analyte is modulated at least 2 or more times, the GC×GC separation is considered "comprehensive" as it retains the ^{1}D separation peak profiles [5,22,23]. The number of times an analyte is modulated, defined as the ^{1}D width-at-base ($^{1}W_b$) divided by the modulation period (P_M), is termed the sampling density (ρ_s) [23], or alternatively the modulation ratio (M_R) [5,22]. Currently, modulators typically run in the seconds timeframe [5,6,8,9,12], though some have been able to operate as quickly as every 250-500 ms [8,24,25]. Further research towards shorter and more efficient GC×GC separations to expand applicability requires a modulator capable of operating at a suitably higher frequency.

Recently, we have been exploring an alternative modulation approach, referred to as partial modulation [17], that employs a pulse valve and is capable of higher frequency sampling rates with P_M as low as 50 ms [26–29]. As the name implies, partial modulation differs from full modulation by only modulating a portion of the ^{1}D eluate. The fraction of a modulated analyte becomes a series of ^{2}D peaks superimposed on the ^{1}D peak profile. Overall, the ^{1}D separation is comprehensively obtained by ^{2}D separations, simultaneous with also measuring the ^{1}D separation. Note that although the ^{1}D eluate is only partially modulated, the duty cycle is technically 100% since all of the ^{1}D eluate is detected.

Partial modulation can be operated in two modes, either the negative pulse mode (NPM), or the positive pulse mode (PPM) [17,29]. With partial modulation in the NPM, a flow of auxiliary carrier gas at a suitable pressure (P_{aux}) is applied through a pulse valve to a T-junction joining the ^{1}D and ^{2}D columns for nearly all of the P_M, resulting in dilution of the ^{1}D eluate. This is followed by briefly turning off the pulse valve for a specified pulse width (p_w), effectively "injecting" undiluted ^{1}D eluate onto the ^{2}D column. The resulting raw data has the appearance of

^{2}D separations superimposed on top of the ^{1}D separation [29]. Application of a rolling ball minimum baseline correction to the raw data uncouples the ^{2}D separation from the ^{1}D separation, and the uncoupled ^{2}D chromatograms can then be cut and folded into a GC×GC chromatogram. Extremely narrow ^{2}D peak widths-at-base ^2W_b (4σ) of about 10 ms with a P_M of 100 ms were obtained [29]. In contrast, in the PPM, the flow of auxiliary carrier gas from the pulse valve is turned off for nearly all of the P_M (p_w approaching P_M), and then is briefly turned on for a short time interval of P_M - p_w, creating a local analyte vacancy in the ^{1}D eluate as it enters the ^{2}D column. Although the raw data collected in the PPM can be converted into traditional appearing GC×GC chromatograms, the data analysis steps were overly complex and the resulting signal-to-noise ratio (*S/N*) suffered [26–28]. The overall consensus was that the partial modulation in the NPM has superior potential relative to the PPM, by achieving extremely narrow peaks using a very short P_M, concurrent with a much higher *S/N* post data processing. In the recent report, we introduced the potential for coupling GC×GC using partial modulation in the NPM to a TOFMS [29]. In this initial report we did not address the challenges that would arise due to the necessity to uncouple the ^{1}D and ^{2}D contributions to the overall detected signal at each mass channel (*m/z*) for overlapped peaks. In this current study we focus our attention on this challenge by employing chemometric decomposition [10,30–33].

The data produced by GC×GC-TOFMS instruments is inherently complex. Numerous computational tools (broadly termed chemometrics) have been developed to adequately and quickly process and analyze the data. There are numerous useful chemometric tools available for use [10,33–38]. For GC×GC-TOFMS, using partial modulation in the NPM, the data are inherently not trilinear because of the nature in which the ^{1}D separation is coupled to the ^{2}D separations prior to applying chemometric decomposition. Essentially, in the workflow the signal

as a function of *m/z* for each analyte must first be separated from the other analytes and/or interferences. Following this step, the ^{1}D separation can be uncoupled from the ^{2}D separations using a rolling ball minimum baseline correction step. For this purpose multivariate curve resolution – alternating least squares (MCR-ALS) is well suited for the chemometric decomposition step of the workflow [31,38–44]. While partial modulation in the NPM is a chromatographically efficient method that produces narrow peaks, the potential drawback is that the GC×GC data produced is intrinsically non-bilinear. MCR-ALS is robust to shifts in ^{2}D retention time shifts and less sensitive to peak shape differences than other chemometric techniques that require the data to be bilinear, which makes it an advantageous algorithm to use for the non-bilinear data produced by this modulation technique. MCR-ALS also allows for the use of common constraints such as non-negativity, unimodality, convergence, and selectivity across the entire data set or just selected components, as desired. An important consideration when applying MCR-ALS is that the solutions are subject to rotational ambiguity. This can often be minimized by applying harsher constraints, building augmented data matrices, and hard modeling [45]. Augmented data matrices can be built by concatenating each unfolded 2D data array either column-wise or row-wise. Uniqueness of the solutions may be evaluated by various methods including the MCR-BANDS [45] software or other soft modelling methods [46].

MCR-ALS splits the instrumental response of the sample matrix into the concentration peak profiles and the spectral information, and uses the alternating least-squares process to fit the data [42–44]. The algorithm calculates the total variance explained in the data set as well as the residuals, or the part of the data set that was not modeled. An initial estimate of the pure analytes is first determined and the ALS process adjusts the elution profiles and spectra iteratively to maximize the explained variance of the data being modeled [40]. GC×GC-TOFMS data can be

decomposed computationally by MCR-ALS along the ^{1}D elution time (^1t_R), ^{2}D elution time (^2t_R), and mass spectral data dimensions. In addition to selecting a time window of the raw sample data, the number of chemical components present in this selected region must be estimated prior to decomposition, though knowing the identity of the components is not necessary [44]. This assigns the number of response profiles the model generates and the number of predicted chemical components that can be edited according to the modeled results. Computing an MCR-ALS model decomposes the raw data matrix into a matrix containing the ^{1}D and ^{2}D elution peak profiles, herein referred to as the loadings, for each analyte [28,40,45]. These loadings contain the concentration information of the pure response peak profiles that allows for the quantification of analytes that are susceptible to co-elution with similar compounds. The MCR-ALS model also provides a matrix containing the mass spectral profiles, corresponding to each loading, that can be matched to established libraries to identify analytes. An advantage of including spectral data as one of the model variables includes the ability to chemometrically decompose complex samples based upon selective *m/z*. The details of this algorithm and its applications are described more extensively in the literature [31,39,42–44].

GC×GC-TOFMS using more "traditional" modulation approaches produces a series of modulated peaks that follow the same general shape of the ^{1}D analyte peak profile. From a single ^{1}D analyte peak, a series of ^{2}D peaks are produced, which can be summed for quantification. Partial modulation in the NPM creates a similar set of modulated ^{2}D peaks, but also retains the ^{1}D peak profile underneath the ^{2}D peak profile. These ^{2}D peaks were shown to be reproducible with very narrow ^2W_b, and the ^{2}D peaks are relatively easily isolated from the ^{1}D peaks for comparison and calculation of figures-of-merit [29]. However, analyte quantification, either with or without peak overlap, was not previously addressed. A goal of the current study is to

determine if data produced using partial modulation in the NPM is quantifiable using a chemometric decomposition approach. The quantification task at hand bares similarity to that presented by Carabahal et al. [47] that describes an elegant method using MCR-ALS to quantify data with two coupled instrumental modes. They reported an approach that models augmented data matrices containing concatenated test samples and calibration samples, whereby the peak area under the entire resolved concatenated profiles is calculated for quantification. In our study, the ^{1}D peak for each analyte is modulated by the same ratio, the resulting ^{2}D peak areas can be summed (without concern for the remaining, overlapped ^{1}D signals) in a similar manner as traditional one-dimensional GC and GC×GC data, and yield reliable identification and accurate quantification [48]. For this purpose, an isothermal separation of a 15-component mixture of similar compounds is studied, with a 20 s separation time window using a P_M of 250 ms and a short pulse width p_w of 6 ms in order to produce very narrow ^{2}D peaks. Various peak overlap situations were purposely produced to facilitate the chemometric method demonstration.

3.2 Experimental

The GC×GC-TOFMS instrument using partial modulation in the NPM was based upon an Agilent 6890N GC (Agilent Technologies, Palo Alto, CA, USA) and a LECO Pegasus III TOFMS (LECO Corporation, St. Joseph, MI), with the schematic provided in Figure 3.1. The carrier gas (Praxair, Seattle, WA, USA) was ultra-high purity helium (Grade 5, 99.999%). The inlet temperature was 175 °C and the transfer line temperature was 240 °C. The oven temperature was set to 100 °C. Post-run data processing was performed in MATLAB R2017b (The Mathworks, Inc., Natick, MA, USA). The Agilent 6890N GC was fitted with a high-speed pulse valve (Model 009-0347-900, Parker Hannifin, Hollis, NH, USA) to facilitate the GC×GC operation. The pulse valve was mounted outside the oven (Figure 3.1), with a custom-built fitting

to connect a 0.005-inch inner diameter (ID) stainless steel tubing (VICI model T5C5D, Valco Instruments Company Inc., Houston, TX, USA) to a 3-way T-union (Model MT.5CXS6, Valco Instruments Company Inc., Houston, TX, USA) mounted inside the oven. The 3-way T-union served as the point of modulation for the pulse valve, coupling the ^{1}D and ^{2}D columns. The pulse valve was controlled by an Iota One Microfluidic Valve Driver (Model 060-0010-900, Parker Hannifin, Hollis, NH, USA). The auxiliary pressure was controlled by the Agilent 6890N GC.

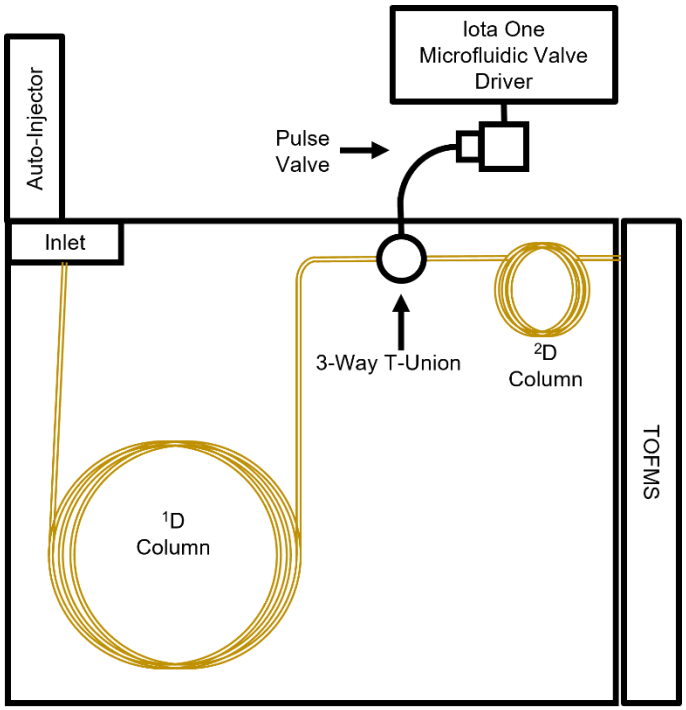


Figure 3.1. Schematic of the GC×GC-TOFMS instrumental setup used for all data collection. Partial modulation in the negative pulse mode by the pulse valve occurs at the 3-way T-union, where eluate from the ^{1}D column is "re-injected" onto the ^{2}D column.

The study was conducted with a 15-component mixture (Table 3.1) measured by volume in octane. Octane served as a common solvent that did not elute in the same ^{1}D time window as the rest of the analytes. For the quantification study, a second test mixture with the same compounds was prepared varying the concentrations of 1-hexene (4), and 2-pentanone (13). The

concentration of 1-hexene (4) injected for one test mixture was 34 parts-per-thousand by volume (µl/ml) and 39 µl/ml for the other mixture, and concentrations for 2-pentanone (13) were 48 µl/ml and 81 µl/ml, respectively. The studies used a "normal" column configuration: a non-polar ^{1}D column coated with DB-5 stationary phase (3.0 m length × 100 µm $^{1}d_{c}$ × 0.10 µm $^{1}d_{f}$), followed by a polar ^{2}D column coated with DB-WAX stationary phase (1.0 m length × 100 µm $^{2}d_{c}$ × 0.10 µm $^{2}d_{f}$). The inlet flow rate was set to 1.5 ml/min, equating to a P_{inlet} of 384.7 kPa (55.8 psig) at the inlet. The auxiliary flow to the pulse flow valve was set to a P_{aux} of 393.7 kPa (57.1 psig) with a p_{w} of 6 ms and a P_{M} of 250 ms. A sample volume of 1 µl was injected with a split of 150:1 at the inlet. The TOFMS was set to collect *m/z* 15-150 at a collection frequency of 500 Hz. The heaviest component of the test mixture was octane, with a molar mass of 114 g/mol, allowing for use of the narrow *m/z* range. A relatively narrow *m/z* range was also beneficial to reduce the size of the data obtained with this high collection frequency. An electron impact ionization voltage of -70 eV was used with a detector voltage of 1562 V.

Table 3.1. The fifteen-component test mixture is listed here, with peak number and boiling point for each analyte. The analytes are ordered by $^{2}t_{R}$ within the time window input for MCR-ALS decomposition. The compounds were selected to create separation regions of overlap that required chemometric decomposition.

Peak Number	Analyte Name	Boiling Point (°C)
1	Pentane	36.1
2	Acetone	56
3	Isopropyl Alcohol	82.6
4	1-Hexene	62-63
5	1-Hexyne	69-71
6	2-Butanone	79.6
7	2-Butanol	98
8	Methylcyclopentane	71.8
9	1,2-Dichloroethane	83
10	Heptane	98.4
11	Cyclohexene	83
12	Benzene	80.1
13	2-Pentanone	101

14	Thiophene	84
15	1-Heptyne	98-99

3.3 Results and Discussion

The compounds in the 15-component mixture were chosen to create a heavily overlapped GC×GC chromatogram to evaluate the ability to analyze the data using MCR-ALS. The raw total ion current (TIC) chromatographic data obtained from a 20 s window of the separation of the 15-component mixture is shown in Figure 3.2A, while an overlay of the individual *m/z* raw chromatographic data are shown in Figure 3.2B. In order to provide an *approximate* GC×GC chromatogram, the raw TIC chromatographic data in Figure 3.2A was baseline corrected using a rolling ball minimum algorithm [26,29], with the result provided in Figure 3.2C. The baseline correction effectively subtracted the underlying ^{1}D chromatogram from the raw TIC data to isolate the ^{2}D data (albeit with some artifacts at this stage due to peak overlaps). Analogous to full modulation methods, the data in Figure 3.2C was then cut-and-folded into a 2D contour plot (Figure 3.2D), which has the appearance of a typical GC×GC chromatogram. Although the image in Figure 3.2D contains artifacts due to the peak overlaps that have not been properly handled via chemometric decomposition, and hence is an *approximate* GC×GC chromatogram, the image provides a suitable starting point to initiate the chemometric analysis. One can see that the 15 analytes are heavily overlapped along the ^{1}D separation, with only one analyte, 1-heptyne, having any substantial amount of its ^{1}D peak elute alone ($^{1}t_{R}$ = 98.75 s). Due to this heavy overlap, and in order to apply MCR-ALS with a suitable rank for any given decomposition analysis, the data was divided into five "time windows" along the ^{1}D axis. The accuracy of the decomposition increased along with a reduction of the computational load. The five time windows were: 80.0-87.5 s (pentane, acetone, isopropyl alcohol), 86.5-92.0 s (1-hexene, 1-hexyne, 2-butanone, and 2-butanol), 89.0-97.5 s (methylcyclopentane and 1,2-dichlororethane),

92.0-101.0 s (heptane, cyclohexene, benzene, 2-pentanone, and thiophene), and 96.0-102.0 s (1-heptyne). The five time windows overlap such that each analyte is completely decomposed in the window indicated, although some analytes are also partially observed in the decomposition of an adjacent window.

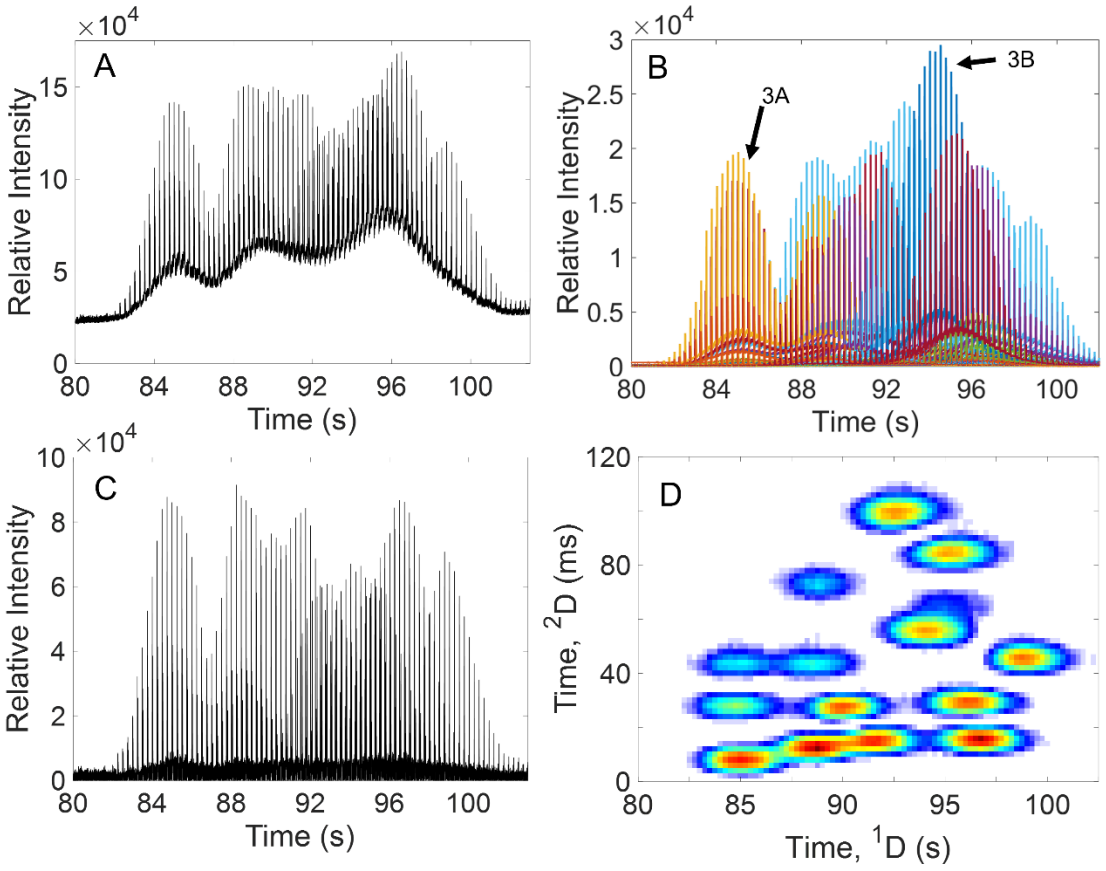


Figure 3.2. Visualization of raw data prior to chemometric decomposition. (A) The total ion current (TIC) raw data for the separation of the 15-component mixture by GC×GC-TOFMS using partial modulation in the negative pulse mode. (B) The raw data in (A) showing mass channels (*m/z*) 15-150 with the regions 3.3A and 3.3B corresponding to heavily overlapped regions selected for demonstrating decomposition by MCR-ALS. (C) Baseline corrected TIC obtained by subtracting the underlying ^{1}D profile from (A) prior to decomposition. (D) Two-dimensional contour plot for visualization purposes, obtained by treating the baseline corrected data in (C) as raw GC×GC data, and cutting and folding on the P_M = 250 ms intervals (only 120 of 250 ms on ^{2}D is shown).

In order to better visualize the ^{2}D separations in Figure 3.2B, the signals for all *m/z* for specific modulations along the ^{1}D separation for two of the time windows are provided in Figure

3.3A for pentane (1), acetone (2), and isopropyl alcohol (3) nominally centered at a ^1t_R = 84.75 s, and in Figure 3.3B for heptane (10), cyclohexene (11), benzene (12), 2-pentanone (13), and thiophene (14) centered at a ^1t_R = 94.25 s. In Figure 3.3A, pentane (1), acetone (2), and isopropyl alcohol (3) appear within a 60 ms window on the ^{2}D separation. Due to the narrow ^2W_b achieved by the partial modulation in the NPM, the peaks in this window should be readily decomposed by MCR-ALS. A more challenging region for MCR-ALS decomposition is shown in Figure 3.3B, where there is less resolution on the ^{2}D separation and apparently fewer selective *m/z*. Both of these time windows will serve to illustrate the chemometric data analysis workflow.

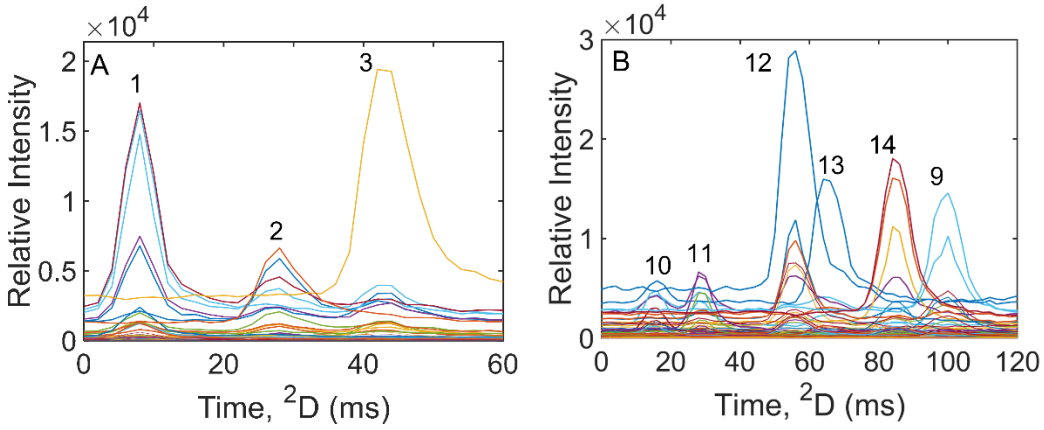


Figure 3.3. Sections highlighted in Figure 3.2B for demonstration of decomposition by MCR-ALS. (A) Plot of *m/z* the first 60 ms in a single ^{2}D modulation at ^1t_R = 84.75 s in the first time window (80.0-87.5 s). From left to right: pentane (1), acetone (2), and isopropyl alcohol (3). (B) Plot of *m/z* the first 120 ms in a single ^{2}D modulation at ^1t_R = 94.25 s in the fourth time window (92.0-101.0 s). From left to right: heptane (10), cyclohexene (11), benzene (12), 2-pentanone (13), thiophene (14), and 1,2-dichloroethane (9).

The raw, three-dimensional (3D) data (^1t_R, ^2t_R, *m/z*), as illustrated in Figures 3.2 and 3.3A, is fed into MCR-ALS in a series of the defined ^{1}D time windows. The first time window, from 80.0 to 87.5 s, covering the first three analytes (pentane, acetone, isopropyl alcohol) is shown in Figure 3.4A. The data was entered into MCR-ALS as a 2-column array for decomposition after *m/z* 18, 28, and 32, that are affected by atmospheric contamination, were set to zero. To avoid

negative elution profiles and spectra, the algorithm was initialized by applying non-negativity constraints. All MCR-ALS analyses were non-targeted with no additional filters or thresholds applied using the functions available in MATLAB R2017b.The MCR-ALS algorithm returns the decomposed analyte peak profiles i.e., loadings (Figure 3.4B), yet still in the form provided by partial modulation in the NPM with the ^{2}D peaks superimposed on the ^{1}D peak background (which is relatively small). Not shown are the noise loadings, or the sum of all *m/z* with essentially zero analyte signal. MCR-ALS also returns the mass spectral profiles for each of the decomposed analytes (Figure 3.4C). The mass spectral profiles can be directly compared with NIST library standards (or in-house collected mass spectral standards) for match value (MV) calculations to assist analyte identification. This MV was calculated by the match spectral matching algorithm defined by Stein [49] using MS Search 2.0 (NIST, Gaithersburg, MD, U.S.A.) applying no limits to the mass range of NIST that was searched. A forward match value of 800 was accepted for identification. To address the issue of rotational ambiguity when using MCR-ALS for a single data matrix, augmented data matrices were created containing each replicate for each time window. MCR-ALS analysis of the augmented data matrices was performed four times, providing a unique solution that was essentially the same as each single data matrix solution. This included the explained variance (R^2), with values 99.93% or higher for each, and the percent fit for each component that was modeled in the given time window.

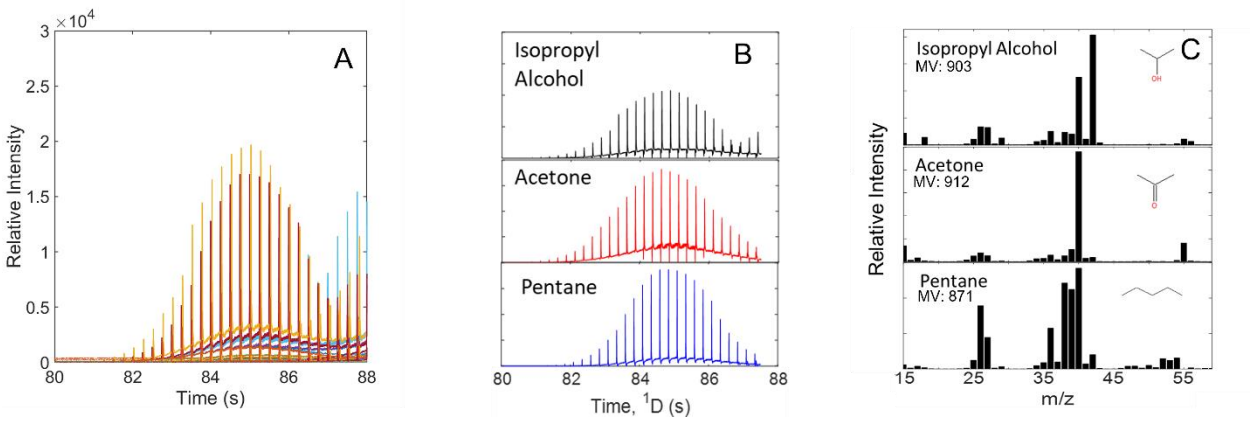


Figure 3.4. MCR-ALS decomposition of first time window. (A) The raw data (all *m/z*) of the first time window (80.0-87.5 s), input into MCR-ALS for decomposition. (B) Loadings obtained, providing concurrent ^{1}D and ^{2}D separation information for each individual analyte: pentane (1), acetone (2), and isopropyl alcohol (3). (C) Mass spectral profiles obtained. These were compared against the NIST library for match value (MV) comparisons. The average MV obtained with excellent accuracy (RSD ≤ 1.35 %) are included in the mass spectral profile for the respective analyte.

The individual analyte peak loadings in Figure 3.4B contain both the naturally coupled ^{1}D and ^{2}D peak profiles, which must be uncoupled for further analysis. Application of a rolling minimum baseline subtraction algorithm as previously described (written in-house) [26,29], and Savitsky-Golay smoothing readily provides the uncoupling, producing two "sub-loadings," referred to as loading 1.1 (^{2}D peak profiles) in Figure 3.5A, and loading 1.2 (^{1}D peak profile) in Figure 3.5B. A zoom in of one modulation from each individual analyte in Figure 3.5A (loading 1.1) is provided in Figure 3.5C to assist in visualizing the narrow ^{2}D peaks. All peaks in Figure 3.5C are taken at 84 s in Figure 3.5A. From here, pertinent figures-of-merit for both dimensions, such as t_R and W_b, can be readily obtained without any mathematical peak reconstruction. After uncoupling the elution time modes by baseline subtraction, the loading 1.1 (Figure 3.5A) can be treated analogous to traditional GC×GC data, and can be cut-and-folded into a contour plot (Figure 3.5D) for visual interpretation [28,29]. Note that Figure 3.5D is an overlay of the individual contour plots for each analyte from Figure 3.5A.

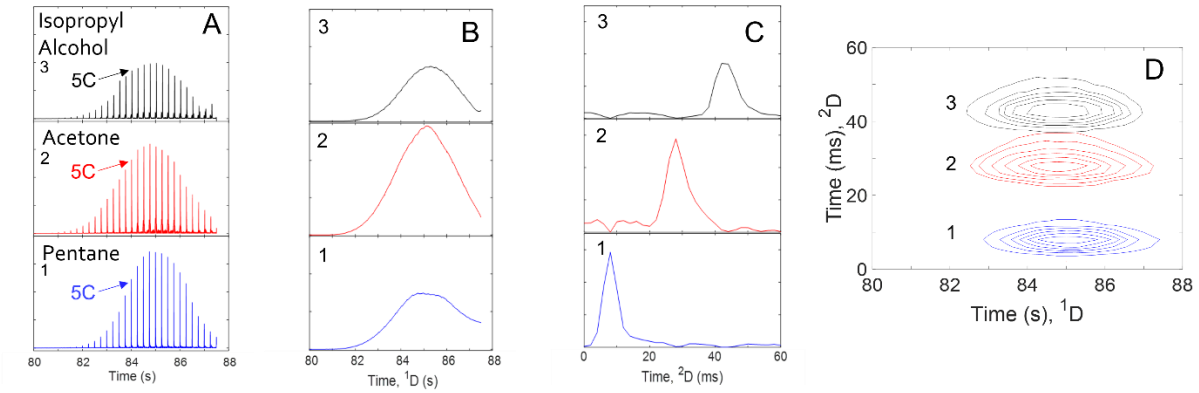


Figure 3.5. Presentation of data processing steps, starting with the MCR-ALS loadings from Figure 3.4B. (A) Stacked baseline-corrected response profiles (loading 1.1) for pentane (1, blue), acetone (2, red), and isopropyl alcohol (3, black) achieved by baseline correcting the loadings in Figure 3.4B. (B) The ^{1}D analyte profiles (loading 1.2) are obtained for each analyte by Savitsky-Golay smoothing the baseline subtracted data from (A), or (3.4B)-(3.5A). These plots were used to determine ^1t_R and ^1W_b (4σ), presented in Table 3.2. (C) The first 60 ms of a single ^{2}D modulation, showing the differences in ^2t_R and narrow ^2W_b obtained, obtained by zooming in on a single modulation at 84 s in (A) as indicated by the arrows. These plots were used to determine ^2t_R and ^2W_b (4σ) in Table 2. (D) Overlay of individual analyte 2D contour plots created by cutting and folding the loading 1.1 profiles in (A).

The first three analytes, pentane (1), acetone (2), and isopropyl alcohol (3), are severely overlapped in the ^{1}D separation, as can be seen in Figure 3.5B. The difference in ^1t_R is 0.25 s between analytes, and with a ^1W_b of 5 s, these analytes would normally remain inseparable to chemometric decomposition due to their low ^{1}D resolution (^1R_s) of 0.05. Due to the short, 1.0 m ^{2}D column, the difference in ^2t_R between each of these three analytes is never greater than 20 ms. However, due to the high ^{2}D separating power facilitated by partial modulation in the NPM, an average ^2W_b of 15 ms was produced. The resulting ^2R_s for pentane (1) and acetone (2) is 1.6 and for acetone (2) and isopropyl alcohol (3) the ^2R_s is 1.1, i.e., there is near baseline separation on ^{2}D between adjacent analyte pairs. The ^2R_s was sufficiently high that it compensated for the low ^1R_s, of 0.05, for the two analyte pairs, resulting in a 2D resolution ($R_{s,2D}$) of 1.6 between pentane (1) and acetone (2), and $R_{s,2D}$ of 1.1 between acetone (2) and isopropyl alcohol (3). Here, the 2D resolution, $R_{s,2D}$, is defined as the Euclidean norm of the resolution in each dimension [50]. The

significantly higher ${}^{2}R_s$ enabled a successful MCR-ALS decomposition, albeit with some noise, of the individual analyte peak profiles, observed in Figure 3.5C.

Next, the MCR-ALS analysis of the most complicated ^{1}D time window (92.0-101.0 s) that fully contained heptane (10), cyclohexene (11), benzene (12), 2-pentanone (13), and thiophene (14), and partially contained 1,2-dichloroethane (9) is presented. The most challenging analytes to separate in this time window were benzene (12) and 2-pentanone (13) due the severe overlap on both dimensions. The ${}^{1}R_s$ was determined to be 0.19 and the ${}^{2}R_s$ is 0.56, resulting in a $R_{s,2D}$ of 0.6. The lack of ^{1}D resolution was overcome by a sufficiently high ^{2}D resolution, facilitating decomposition by MCR-ALS. The raw, 3D data (${}^{1}t_R$, ${}^{2}t_R$, *m/z*) for this section was illustrated in Figures 3.2 and 3.3B. The MCR-ALS results are summarized in Figure 3.6 at the same level of processing as in Figure 3.5 for the first ^{1}D time window. In Figure 3.6A are ^{2}D peak profiles (loading 1.1) following the baseline correction of the full peak loadings, and in Figure 3.6B are the ^{1}D peak profiles (loading 1.2). A zoom in of one of the ^{2}D peak profiles from Figure 3.6A is provided in Figure 3.6C to better judge the performance of the decomposition and to assess ^{2}D peak widths. Once again, the ^{2}D peak profiles in Figure 3.6A (loading 1.1) were cut-and-folded into individual contour plots and added together to produce a traditional appearing GC×GC plot for this time window, presented in Figure 3.6D. Although 1,2-dichloroethane (9) appears in Figure 3.6D, note that part of the peak is missing in this time window. For the purpose of providing an accurate analysis for 1,2-dichloroethane (9), it is fully analyzed in the ^{1}D time window of 89.0-97.5 s that fully contains methylcyclopentane (8) and 1,2-dichlororethane (9).

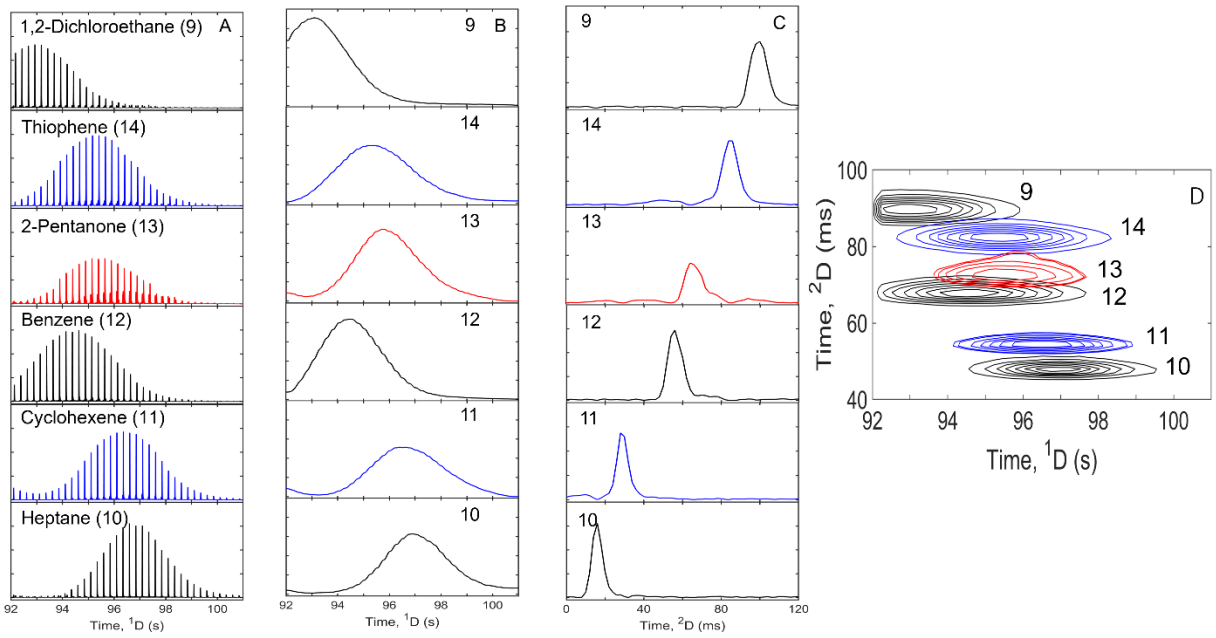


Figure 3.6. MCR-ALS decomposition of the fourth time window (Figure 3.3B), illustrating a more complex separation region. The same processing approach applied in Figures 3.4 and 3.5 was used. (A) Stacked plots of each baseline corrected loading 1.1 of heptane (10), cyclohexene (11), benzene (12), 2-pentanone (13), thiophene (14), and 1,2-dichloroethane (9). (B) The ^{1}D analyte profiles (loading 1.2) are obtained for each analyte by Savitsky-Golay smoothing the baseline subtracted data from (A). The $^{1}R_s$ between adjacent peaks is never greater than 0.35 (1,2-dichloroethane and benzene). (C) Stacked plots of the first 120 ms of a single ^{2}D modulation. The ^{1}D and ^{2}D overlap between 2-pentanone (13) and benzene (12) is evident in (B) and (C). (D) Overlay of the 2D contour plots for each analyte obtained by cutting and folding each loading 1.1 in (A).

The same analytical workflow was applied to the three remaining ^{1}D time windows. Using the isolated loadings for each analyte, each analyte was individually processed to form an overlay 2D contour plot presented in Figure 3.7, providing clear detail regarding the extent of overlap in both separation dimensions. As previously mentioned, the only analyte to have any significant amount of its peak profile elute alone is 1-heptyne (15). The most challenging analyte to fully resolve was 1-hexene (4) due to ^{2}D overlap with two other analytes, pentane (1) and methylcyclopentane (8), with both adjacent pairs producing a $^{2}R_s$ of 0.36. Similar to previous examples, the analytes were resolved due to the relatively higher $^{1}R_s$ of 0.8 resulting in a $R_{s,2D}$ of 0.85 for both adjacent analyte pairs. Despite this high degree of overlap, all 15 analytes were

successfully decomposed by MCR-ALS, with each mass spectral profile subsequently compared to the NIST library for MV comparison, as provided in Table 3.2. The average MVs obtained from the mass spectral profiles of the decomposed analytes matched to the NIST library are remarkably all ≥ 800, with average values for four analytes above 900. These high MVs obtained from MCR-ALS decomposition prove to be excellent for accurate identification. Table 3.2 shows the calculated ${}^{1}t_{R}$, ${}^{1}W_{b}$, ${}^{2}t_{R}$, and ${}^{2}W_{b}$ values, obtained by using sub-loadings 1.1 and 1.2 for each individual analyte, reported for a single representative run. We found the average ${}^{1}W_{b}$ of all analytes is 5 s (± 0.5 s), and the average ${}^{2}W_{b}$ is 15 ms (± 3 ms). Fully resolving all 15 compounds with such narrow ${}^{2}W_{b}$, despite low R_s values, produced by the partially modulated data in the NPM highlights the quality of the response profiles obtained by MCR-ALS.

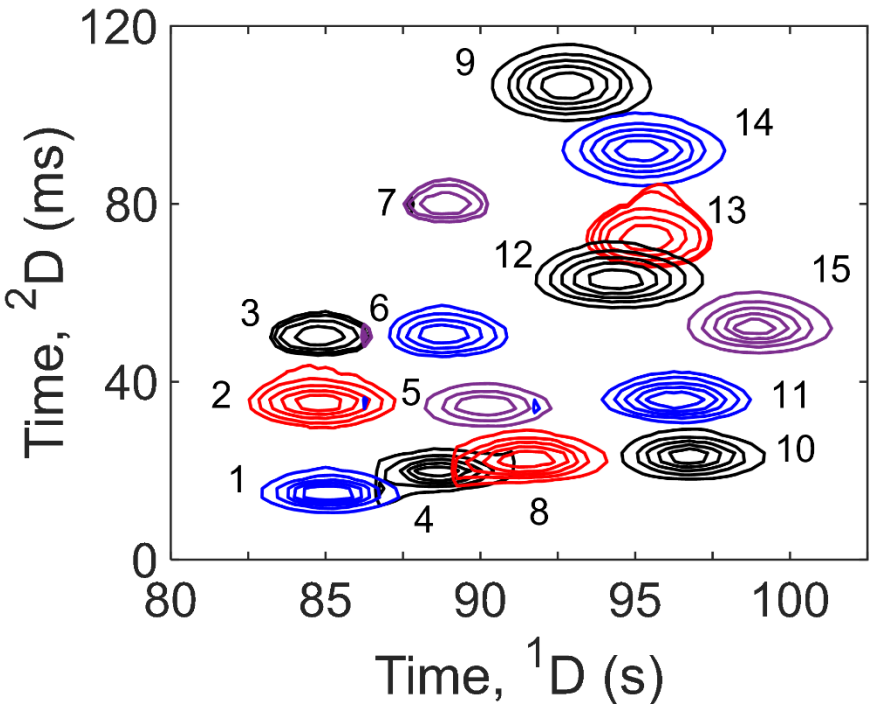


Figure 3.7. Overlay of the 2D contour plots for each analyte of the 15-component mixture obtained by MCR-ALS decomposition. All components were decomposed in a 20 s ^{1}D window with the analytes ordered according to Table 3.2. The 2D contour plot clearly shows the severe overlap on both dimensions.

Table 3.2. All 15 components in the mixture are listed in order by 2tR within the time window used for MCR-ALS decomposition. Average MV from comparison of mass spectral profiles to NIST library spectra and corresponding figures-of-merit are listed, with standard deviation (sd). All MVs are for six runs and are ≥ 800. All t_R and W_b values reported are for a representative replicate. The average ^1W_b is 5 s (± 0.5 s) and the average ^2W_b is 15 ms (± 3 ms).

Peak Number	Analyte Name	MV average (± sd)	^1t_R (s)	^1W_b (s)	^2t_R (ms)	^2W_b (ms)
1	Pentane	871 (± 5)	84.50	4.7	8	10
2	Acetone	912 (± 7)	84.75	5.0	28	15
3	Isopropyl Alcohol	903 (± 12)	85.00	4.7	44	15
4	1-Hexene	800 (± 27)	88.00	4.5	12	10
5	1-Hexyne	874 (± 9)	90.00	4.5	28	12
6	2-Butanone	876 (± 16)	88.50	4.8	44	18
7	2-Butanol	813 (± 45)	88.75	5.0	74	18
8	Methylcyclopentane	910 (± 5)	91.75	5.0	16	12
9	1,2-Dichloroethane	885 (± 14)	92.50	5.9	100	17
10	Heptane	827 (± 11)	96.75	5.5	16	12
11	Cyclohexene	882 (± 27)	96.25	5.6	28	14
12	Benzene	906 (± 6)	94.25	5.5	56	16
13	2-Pentanone	837 (± 24)	95.25	5.0	66	20
14	Thiophene	884 (± 6)	95.00	5.8	86	18
15	1-Heptyne	863 (± 11)	98.75	4.9	46	15

Last, we report on the ability to use the isolated ^{2}D peaks (i.e. loadings 1.1) to demonstrate how quantitative analysis is performed. Two test mixtures composed of the analytes in Table 3.1 were prepared for the quantification study with two different concentrations of 1-hexene (4) and 2-pentanone (13) injected. Peak areas were calculated for a given analyte by the "two-step method" for GC×GC, where the ^{2}D peaks were uncoupled from the ^{1}D profile and the area peak area for the analyte was calculated as the sum of the area of all ^{2}D peaks [92,179,180]. The average peak areas (n=3 replicates) calculated for the highest concentration for each analyte were fitted to the origin, generating a calibration line used to predict their lower concentration.

As can be seen in Figure 3.7, 1-hexene is severely overlapped on ^{2}D between pentane (1) and methylcyclopentane (8), and all of these analytes share several *m/z* with high signal. The predicted versus prepared concentration for 1-hexene was found to be in relatively good agreement, albeit with a slight negative bias quantified by the percent deviation of -5.6% (± 2.2%). The challenge in resolving the 1-hexene peak on ^{2}D and many shared *m/z* may have contributed to the slight negative bias observed. Meanwhile, the percent deviation for the prediction of the concentration of 2-pentanone was 1.8% (± 3.4%), possibly due to the ^{1}D and ^{2}D overlap with benzene (12) as seen in the decomposition of the time window in Figure 3.6. Overall, the predicted and prepared concentration values are in good agreement with percent deviation values of -5.6% (± 2.2%) for 1-hexene and 1.8% (± 3.4%) for 2-pentanone, providing promising results supporting the notion that NPM produces quantifiable data.

3.4 Conclusion

Partial modulation in the NPM is demonstrated to be a powerful form of modulation for GC×GC-TOFMS. The method is amenable to commercial chemometric tools, demonstrated with MCR-ALS, for analyte decomposition, identification, and quantification. This fast modulation technique readily separated similar, volatile organic compounds over two short separation columns and provided sufficient R_s in the 2D separation space. The 15-component mixture eluted within a 20 s time window with no R_s in the ^{1}D separation being greater than 0.38 (1-heptyne to heptane). Despite the severe overlap in the ^{1}D, the excellent separation efficiency on the ^{2}D separation produced narrow ^{2}D peaks (average ^2W_b = 15 ms) with TOFMS detection, facilitating successful MCR-ALS decomposition of the peak profiles. Further, we demonstrated how the loadings produced by MCR-ALS can be further analyzed to extract the ^{1}D and ^{2}D peak profiles in order to obtain pertinent chemical measurements (retention times and peak widths) without the typical mathematical reconstruction of the ^{1}D peak profile from modulated data.

Last, the ^{2}D peak profiles isolated from the MCR-ALS data were shown to be quantifiable and accurate with excellent mass spectral match values. This study provides further impetus into fast GC research for high throughput analysis as modulation techniques are beginning to easily achieve sub-second separations and push the limit of detector collection frequencies.

3.5 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. doi:10.1093/chromsci/29.6.227.

[2] 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.

[3] L.M. Blumberg, F. David, M.S. Klee, P. Sandra, Comparison of one-dimensional and comprehensive two-dimensional separations by gas chromatography, J. Chromatogr. A 1188 (2008) 2–16. doi:10.1016/j.chroma.2008.02.044.

[4] J.A. Murray, Qualitative and quantitative approaches in comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1261 (2012) 58–68. doi:10.1016/j.chroma.2012.05.012.

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

[6] J. V. Seeley, Recent advances in flow-controlled multidimensional gas chromatography, J. Chromatogr. A 1255 (2012) 24–37. doi:10.1016/j.chroma.2012.01.027.

[7] P.Q. Tranchida, G. Purcaro, P. Dugo, L. Mondello, Modulators for comprehensive two-dimensional gas chromatography, TrAC - Trends Anal. Chem. 30 (2011) 1437–1461. doi:10.1016/j.trac.2011.06.010.

[8] H.D. Bahaghighat, C.E. Freye, R.E. Synovec, Recent advances in modulator technology for comprehensive two dimensional gas chromatography, TrAC - Trends Anal. Chem. 113 (2018) 379–391. doi:10.1016/j.trac.2018.04.016.

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

[10] K.M. Pierce, B. Kehimkar, L.C. Marney, J.C. Hoggard, R.E. Synovec, Review of chemometric analysis techniques for comprehensive two dimensional separations data, J. Chromatogr. A 1255 (2012) 3–11. doi:10.1016/j.chroma.2012.05.050.

[11] M. Edwards, H. Boswell, T. Górecki, Comprehensive Multidimensional Chromatography, Curr. Chromatogr. 2 (2015) 80–109. doi:10.2174/2213240602666150722232236.

[12] 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.

[13] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.Th. Brinkman, Recent developments in comprehensive two-dimensional gas chromatography (GC × GC). I. Introduction and instrumental set-up, TrAC - Trends Anal. Chem. 25 (2006) 438–454. doi:10.1016/j.trac.2006.03.002.

[14] 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. doi:10.1016/j.trac.2013.08.009.
[15] P.Q. Tranchida, S. Salivo, F.A. Franchina, L. Mondello, Flow-modulated comprehensive two-dimensional gas chromatography combined with a high-resolution time-of-flight mass spectrometer: A proof-of-principle study, Anal. Chem. 87 (2015) 2925–2930. doi:10.1021/ac5044175.
[16] N.S. Tessarolo, L.R.M. dos Santos, R.S.F. Silva, D.A. Azevedo, Chemical characterization of bio-oils using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A 1279 (2013) 68–75. doi:10.1016/j.chroma.2012.12.052.
[17] H. Cai, S.D. Stearns, Partial modulation method via pulsed flow modulator for comprehensive two-dimensional gas chromatography, Anal. Chem. 76 (2004) 6064–6076. doi:10.1021/ac0492463.
[18] J. Harynuk, T. Górecki, New liquid nitrogen cryogenic modulator for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1019 (2003) 53–63. doi:10.1016/j.chroma.2003.08.097.
[19] J. V. Seeley, N.J. Micyus, S. V. Bandurski, S.K. Seeley, J.D. McCurry, Microfluidic deans switch for comprehensive two-dimensional gas chromatography, Anal. Chem. 79 (2007) 1840–1847. doi:10.1021/ac061881g.
[20] J. Krupčík, R. Gorovenko, I. Špánik, P. Sandra, D.W. Armstrong, Flow-modulated comprehensive two-dimensional gas chromatography with simultaneous flame ionization and quadrupole mass spectrometric detection, J. Chromatogr. A 1280 (2013) 104–111. doi:10.1016/j.chroma.2013.01.015.
[21] H.J. De Geus, J. De Boer, U.A.Th. Brinkman, Development of a thermal desorption modulator for gas chromatography, J. Chromatogr. A 767 (1997) 137–151. doi:10.1016/S0021-9673(97)00038-1.
[22] 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.
[23] 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.
[24] C.E. Freye, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta 161 (2016) 675–680. doi:10.1016/j.talanta.2016.09.002.
[25] C.E. Freye, L. Mu, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1424 (2015) 127–133. doi:10.1016/j.chroma.2015.10.098.
[26] C.E. Freye, H.D. Bahaghighat, R.E. Synovec, Comprehensive two-dimensional gas chromatography using partial modulation via a pulsed flow valve with a short modulation period, Talanta 177 (2018) 142–149. doi:10.1016/j.talanta.2017.08.095.
[27] H.D. Bahaghighat, C.E. Freye, D. V. Gough, P.E. Sudol, R.E. Synovec, Ultrafast separations via pulse flow valve modulation to enable high peak capacity multidimensional gas chromatography, J. Chromatogr. A 1573 (2018). doi:10.1016/j.chroma.2018.08.001.
[28] H.D. Bahaghighat, C.E. Freye, D. V. Gough, R.E. Synovec, Comprehensive two-dimensional gas chromatography and time-of-flight mass spectrometry detection with a 50 ms modulation period, J. Chromatogr. A 1583 (2019) 117–123. doi:10.1016/j.chroma.2018.11.027.

[29] D. V. Gough, D.H. Song, S. Schöneich, S.E. Prebihalo, R.E. Synovec, Development of Ultrafast Separations Using Negative Pulse Partial Modulation to Enable New Directions in Gas Chromatography, Anal. Chem. 91 (2019) 7328–7335. doi:10.1021/acs.analchem.9b01085.
[30] T. Rajalahti, O.M. Kvalheim, Multivariate data analysis in pharmaceutics: A tutorial review, Int. J. Pharm. 417 (2011) 280–290. doi:10.1016/j.ijpharm.2011.02.019.
[31] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146.
[32] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, J. Chemom. 9 (1995) 31–58. doi:10.1002/cem.1180090105.
[33] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149–171. doi:10.1016/S0169-7439(97)00032-4.
[34] 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.
[35] A.E. Sinha, C.G. Fraga, B.J. Prazen, R.E. Synovec, Trilinear chemometric analysis of two-dimensional comprehensive gas chromatography-time-of-flight mass spectrometry data, J. Chromatogr. A 1027 (2004) 269–277. doi:10.1016/j.chroma.2003.08.081.
[36] J.M. Amigo, T. Skov, R. Bro, J. Coello, S. Maspoch, Solving GC-MS problems with PARAFAC2, TrAC - Trends Anal. Chem. 27 (2008) 714–725. doi:10.1016/j.trac.2008.05.011.
[37] H.A.L. Kiers, J.M.F. ten Berge, R. Bro, PARAFAC2 - Part I. A direct fitting algorithm for the PARAFAC2 model, J. Chemom. 13 (1999) 275–294. doi:10.1002/(sici)1099-128x(199905/08)13:3/4<275::aid-cem543>3.3.co;2-2.
[38] 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.
[39] 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.
[40] 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. doi:10.1016/j.aca.2013.07.071.
[41] J. Omar, A. Vallejo, M. Olivares, A. Usobiaga, O. Zuloaga, N. Etxebarria, Resolution and identification of co-eluting alkylphenols in comprehensive two-dimensional gas chromatography-mass spectrometry by multivariate curve resolution-alternating least squares, J. Chemom. 29 (2015) 237–244. doi:10.1002/cem.2701.
[42] A. De Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution, Anal. Chim. Acta 500 (2003) 195–210. doi:10.1016/S0003-2670(03)00724-4.
[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.

[44] L.A.F. de Godoy, L.W. Hantao, M.P. Pedroso, R.J. Poppi, F. Augusto, Quantitative analysis of essential oils in perfume using multivariate curve resolution combined with comprehensive two-dimensional gas chromatography, Anal. Chim. Acta 699 (2011) 120–125. doi:10.1016/j.aca.2011.05.003.
[45] J. Jaumot, R. Tauler, MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution, Chemom. Intell. Lab. Syst. 103 (2010) 96–107. doi:10.1016/j.chemolab.2010.05.020.
[46] A. Golshan, H. Abdollahi, S. Beyramysoltan, M. Maeder, K. Neymeyr, R. Rajkó, M. Sawall, R. Tauler, A review of recent methods for the determination of ranges of feasible solutions resulting from soft modelling analyses of multivariate data, Anal. Chim. Acta 911 (2016) 1–13. doi:10.1016/j.aca.2016.01.011.
[47] M.D. Carabajal, J.A. Arancibia, G.M. Escandar, Multivariate curve resolution strategy for non-quadrilinear type 4 third-order/four way liquid chromatography–excitation-emission fluorescence matrix data, Talanta 189 (2018) 509–516. doi:10.1016/j.talanta.2018.07.017.
[48] O. Amador-Muñoz, P.J. Marriott, Quantification in comprehensive two-dimensional gas chromatography and a model of quantification based on selected summed modulated peaks, J. Chromatogr. A 1184 (2008) 323–340. doi:10.1016/j.chroma.2007.10.041.
[49] S.E. Stein, D.R. Scott, ScienceDirect - Journal of the American Society for Mass Spectrometry : Optimization and testing of mass spectral library search algorithms for compound identification, J. Am. Soc. Mass Spectrom. 5 (1994) 859–866. doi:10.1016/1044-0305(94)87009-8.
[50] R.E. Murphy, M.R. Schure, J.P. Foley, Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography, Anal. Chem. 70 (1998) 1585–1594. doi:10.1021/ac971184b.

Chapter 4: Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection

This chapter was reproduced from S. Schöneich, D. V Gough, T.J. Trinklein, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection, J. Chromatogr. A. (2020) 460982.

4.1 Introduction

Comprehensive two-dimensional (2D) gas chromatography (GC×GC) for the separation of volatile and semi-volatile compounds was pioneered by Liu and Phillips [1]. Significant advances have been made in instrumentation and chemometric analysis tools for a diverse range of applications [2–9]. The major GC×GC advantages over one-dimensional (1D) GC include about a 10-fold gain in peak capacity and increased selectivity [10]. A GC×GC separation is accomplished using two columns with sufficiently different polarities coupled in series, separated by the modulator [2], so analytes overlapped on the first dimension (^{1}D) separation have the opportunity to be separated on the second dimension (^{2}D) separation [11–13]. Coupled with detection by mass spectrometry (MS), additional sensitivity and selectivity is attained with a third dimension of data, providing mass spectra that can be used for analyte identification.

Modulators are generally classified as either thermal [14–20] or flow modulators [21–29]. Thermal modulators commonly involve the use of cryogens [20]. An advantage of thermal modulation is that a 100% duty cycle is achieved when the ^{1}D eluate is trapped and focused, meaning all the analyte within the ^{1}D eluate is transferred to the ^{2}D column. Advances in thermal modulation have led to commercially available consumable-free thermal modulators [15,30]. Thermal modulation has been applied with a modulation period (P_M), i.e., the ^{2}D separation time,

down to 250 ms [31]. In contrast, flow modulation can be classified as either flow diversion or differential flow modulation, redirecting the flow from the ^{1}D column to the ^{2}D column. Flow modulators achieve a duty cycle ≤ 100%, depending on the design. Flow diversion modulators [32] sample the ^{1}D eluate for a small fraction of the P_M and re-inject it onto the ^{2}D column, resulting in a low duty cycle [25]. Differential flow modulation based on diaphragm valves involves an initial loading step in which a fraction of the ^{1}D eluate is collected in a sample loop, followed by an injection step where the eluate is flushed out using high ^{2}D volumetric flow rates (^{2}F), eg. 20 ml/min. A high ^{2}F is not an issue for flame ionization detection (FID), however, MS detectors are limited to operating under lower pressures to not exceed the pumping capacity. For GC×GC with time-of-flight mass spectrometry detection (GC×GC-TOFMS), flow splitting can be used to reduce the flow rate entering the detector at the cost of reducing sensitivity [6,33]. Additionally, diaphragm valves require few consumables and can now operate at temperatures up to 325 °C [24,34]. Considering all of the issues, the ongoing challenge for modulator development includes providing a total analyte transfer (100% duty cycle) and consumable-free design that is easy to implement that can modulate at high frequencies (low P_M) concurrent with producing narrow ^{2}D peaks ($^{2}W_b$) for high peak capacity on the ^{2}D separations ($^{2}n_c$).

To address this challenge, a novel flow modulation approach was introduced in 2004 by Cai & Stearns [28], and recently expanded upon by our group to achieve rapid separations with short modulation periods ($P_M \geq 50$ ms) and narrow peaks ($^{2}W_b \geq 10$ ms) using a commercially available pneumatic pulse valve [35–38]. The most recent development is referred to as dynamic pressure gradient modulation (DPGM) in the full modulation mode (FMM), due to a pseudo stop-flow modulation mechanism that provides a 100% duty cycle [39]. Indeed, in many ways DPGM is similar to some existing stop-flow designs applied to GC×GC-FID [40,41]. Dynamic

pressure gradient modulation in the FMM was recently demonstrated for GC×GC-FID providing excellent signal enhancement relative to GC-FID, 100% duty cycle, narrow ^{2}D peak widths ($^2W_b \geq 22$ ms), and high peak capacity on the ^{2}D separations [39]. While the ^2F applied were relatively high (20 ml/min), since an FID was implemented, there remained the concern that DPGM in the FMM must be evaluated with TOFMS detection.

Advances in flow modulation using a reduced ^2F (under 10 ml/min) for MS detector compatibility has made remarkable progress [26,42–44]. Tranchida et al. described a flow modulation technique using a ^{2}D flow of 6 ml/min applicable to GC×GC-MS by extending the re-injection step, providing ^2W_b of ~ 780 ms and a 1.5-fold increase in sensitivity over GC-MS [44]. Franchina et al. used a ^{2}D flow rate of 4 ml/min with a quadrupole mass spectrometer (qMS), obtaining ~ 2.5-fold increase in signal-to-noise ratio (*S/N*) over GC-qMS [43]. Finally, a 2 ml/min ^{2}D flow rate was demonstrated using a Deans' switch generating short pulses (≥ 50 ms) at the cost of decreased duty cycle and sensitivity [26].

Building upon these low-pressure flow modulation approaches, herein we demonstrate the compatibility of DPGM in the FMM for GC×GC-TOFMS detection operating under low ^2F conditions, i.e., 4 ml/min. Briefly, DPGM functions by temporally controlling an auxiliary carrier gas flow introduced at a T-union joining the ^{1}D and ^{2}D columns using a pneumatic "pulse" valve (Figure 4.1). An auxiliary gas pressure (P_{aux}) slows the ^{1}D flow while increasing the ^{2}D flow, and when tuned properly, P_{aux} can temporarily stop the ^{1}D flow before the T-union. During each P_M, the valve is closed for a brief time interval defined as the pulse width (p_w), in which ^{1}D eluate is modulated (injected) onto the ^{2}D column. Manipulating the p_w for a given P_M and ratio of inlet pressure (P_{inlet}) to P_{aux} determines the level of modulation achieved. Partial modulation in the positive pulse mode (PPM) is achieved using long pulse widths, p_w

approaching P_M [37], while partial modulation in the negative pulse mode (NPM) is achieved using very short pulse widths, p_w approaching 0 ms [38]. In the PPM the valve is closed for most of the P_M, and opened very briefly to inject carrier gas at the T-union. The resulting ^{2}D analyte peaks appear as vacancies superimposed upon the ^{1}D peak profile, requiring multiple data processing steps before chromatographic information is revealed [13,35–37]. In the NPM, the valve is open for most of the P_M, diluting the ^{1}D eluate, and then is closed very briefly allowing for a short pulse of undiluted ^{1}D eluate to be modulated. The signal for a given analyte in the NPM appears as a series of ^{2}D peaks superimposed on top of a diluted ^{1}D peak, requiring less data processing than using the PPM [38,39]. Notably, for a wide range of p_w between the two extremes of the PPM and NPM, the ^{1}D peak is fully modulated (100% duty cycle) and the ^{2}D peaks return to baseline between modulations, referred to as DPGM in the FMM [39]. In the FMM, a suitably high auxiliary pressure is necessary to slow down and then temporarily stop the ^{1}D flow immediately before the T-union during the interval of P_M - p_w, and injection onto ^{2}D occurs during p_w.

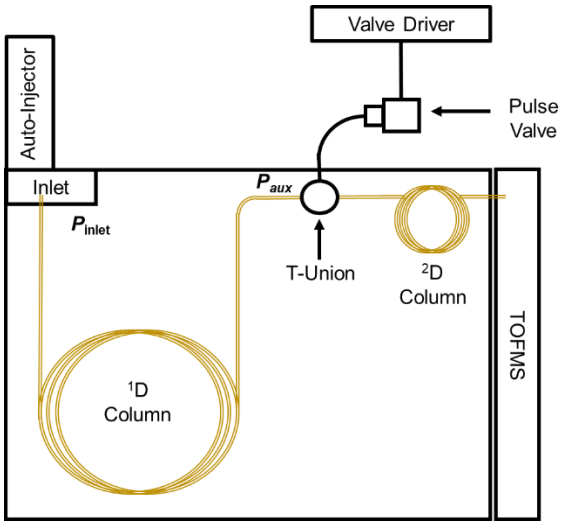


Figure 4.1. GC×GC-TOFMS instrument employing dynamic pressure gradient modulation (DPGM). Modulation occurs at the T-union where the auxiliary carrier gas flow is introduced by the pulse valve.

This report provides an initial evaluation of the compatibility of DPGM in the FMM for GC×GC-TOFMS, with an assessment of the data quality produced. Representative analytes from a 90-component test mixture separation are used for this purpose. Several important aspects were investigated. First, detector response enhancement factors (*DREF*) [45] provide a figure-of-merit as to how well DPGM in the FMM focuses the ^{2}D peaks in time relative to their ^{1}D peak width. Second, we investigate the extent to which the ^{2}D peaks produced adhere to a bilinear data structure while using this flow modulator, since flow modulation is known to potentially impact the reproducibility of the ^{2}D peak shapes for a given modulated ^{1}D peak [46]. To study the bilinearity of the GC×GC data, we selected parallel factor analysis (PARAFAC) due to it requiring the most strict trilinearity requirement for chemometric data analysis of GC×GC-TOFMS data [47–50]. PARAFAC provides a very high bar to assess the data quality when DPGM in the FMM is used. Third, the limit of detection (*LOD*) is studied using PARAFAC [47]. Finally, mass spectra obtained from the PARAFAC decomposition of several highly overlapped analytes are compared to an in-house library of mass spectra, and the match values (MV) are used as a metric to support the trilinearity of the GC×GC-TOFMS data produced using DPGM.

4.2 Experimental

All data were collected on a GC×GC-TOFMS instrument configured with an Agilent 6890N GC (Agilent Technologies, Palo Alto, CA, USA) and LECO Pegasus III TOFMS (LECO Corporation, St. Joseph, MI), as shown in Figure 4.1. A pneumatic valve served as the modulator, i.e, the “pulse valve” (Model 009-0347-900, Parker Hannifin, Hollis, NH, USA), formatted outside the GC oven and controlled by an Iota One Microfluidic Valve Driver (Model 060-0010-900, Parker Hannifin, Hollis, NH, USA) to introduce an auxiliary gas flow controlled by the Agilent GC. The valve was incorporated at the junction point of the ^{1}D and ^{2}D columns at

a T-union (Model MT.5CXS6, Valco Instruments Company Inc., Houston, TX, USA) via 0.005 inch inner diameter (i.d.) stainless steel tubing (VICI model T5C5D, Valco Instruments Company Inc., Houston, TX, USA). The carrier gas used was ultra-high purity helium (Grade 5, 99.999%, Praxair, Seattle, WA, USA). The inlet and transfer line temperatures were maintained at 200 °C and 240 °C, respectively. A reverse column set [51,52] was chosen with the ^{1}D column (19 m length, 180 µm $^{1}d_{c}$, 0.20 µm $^{1}d_{f}$) coated with a Rtx-200 (polar) stationary phase (Restek, Bellefonte, PA, USA), and the ^{2}D column (2 m length, 180 µm $^{2}d_{c}$, 0.18 µm $^{2}d_{f}$) coated with a Rxi-1ms (non-polar) stationary phase (Restek, Bellefonte, PA, USA). The TOFMS was set to collect from mass channels 34 *m/z* to 338 *m/z* at a collection rate of 200 *Hz*. The electron impact ionization energy was -70 eV and the detector voltage was 1562 V. Following data collection, all analyses were performed using MATLAB R2018b (The Mathworks, Inc., Natick, MA, USA).

To investigate DPGM with GC×GC-TOFMS, a 90-component test mixture composed of a wide range of boiling points and chemical classes was used (Table 4.1) [38]. A 0.2 µl volume of the 10 part-per-thousand (ppth) test mixture was injected with a split of 150:1. Serial dilution of this test mixture in methanol was performed, ranging from 10 parts-per-million (ppm) to 10 parts-per-billion (ppb) by factors of ten, in order to evaluate the *DREF* [45], *LOD*, calibration linearity, and a chemometric decomposition study [48]. The diluted test mixtures were injected splitless in a 1 µl volume. A modulation period (P_{M}) of 1 s was used for all runs with the 90-component test mixture. The pulse width (p_{w}) activating the pulse valve (Figure 4.1), defined as the length of time that the flow is stopped from the pulse valve to the junction point of the ^{1}D and ^{2}D columns at a T-union [39], was varied to demonstrate different levels of modulation. The oven temperature was initially held at 40 °C for 1 min, then ramped at a rate of 10 °C/min to 250 °C, where it was held for 1 min. The auxiliary pressure (P_{aux}) on the valve was held at 124.1 kPa

(18.0 psig) for 1 min, then raised at a rate of 5.91 kPa/min (0.857 psi/min) to 248.2 kPa (36.0 psig), where it was held for 1 min. An initial inlet pressure (P_{inlet}) of 141 kPa (20.5 psig) was applied for the first minute of the run and then increased by 6.38 kPa/min (0.923 psi/min) to 275.1 kPa (39.9 psig) where it was held for 1 min, maintaining a similar ratio between the inlet and the auxiliary pressure throughout the run. Although the inlet flow rate was set to 1 ml/min in the GC experimental method (P_{inlet}), a linear velocity of 7.9 cm/s was calculated based on the experimentally observed 4 min dead time, which corresponds to a ^{1}D volumetric flow rate of 0.12 ml/min. Similarly, the flow rate departing the ^{2}D column was determined to be 4 ml/min.

Table 4.1. The 90-component test mixture used for the evaluation of DPGM in the FMM with GC×GC-TOFMS.

Alkanes	**Boiling Point**	**Vendor**	**Esters**	**Boiling Point**	**Vendor**
hexane	69	Sigma-Aldrich	ethyl formate	54	Sigma-Aldrich
heptane	98	Sigma-Aldrich	methyl decanoate	224	Eastman Chemicals
octane	126	Sigma-Aldrich	methyl caprylate	193	Sigma-Aldrich
nonane	151	Sigma-Aldrich	methyl salicylate	223	Alfa-Aesar
decane	174	Sigma-Aldrich	ethyl salicylate	233	Alfa-Aesar
undecane	196	Sigma-Aldrich	methyl laurate	262	Sigma-Aldrich
dodecane	216	Sigma-Aldrich	methly caproate	151	Sigma-Aldrich
tridecane	235	Fluka	diethyl phthalate	299	Sigma-Aldrich
tetradecane	254	Fluka	**Ketones**	**Boiling Point**	**Vendor**
pentadecane	271	Alfa-Aesar	2-butanone	80	Sigma-Aldrich
hexadecane	287	Sigma-Aldrich	2-pentanone	102	Sigma-Aldrich
pristane	296	Sigma-Aldrich	3-hexanone	128	Sigma-Aldrich
octadecane	316	Fluka	2-hexanone	128	Sigma-Aldrich
eicosane	343	Sigma-Aldrich	2-heptanone	152	Sigma-Aldrich
Halogenated Alkanes	**Boiling Point**	**Vendor**	3-heptanone	141	Sigma-Aldrich
1,5-dichloropentane	179	Sigma-Aldrich	3-octanone	167	Sigma-Aldrich
1-chlorohexane	135	Sigma-Aldrich	2-decanone	209	Sigma-Aldrich
1-bromohexane	155	Sigma-Aldrich	2-undecanone	231	Sigma-Aldrich
1-bromoheptane	179	Sigma-Aldrich	2-dodecanone	245	Sigma-Aldrich
1-bromooctane	200	Sigma-Aldrich	**Aromatics**	**Boiling Point**	**Vendor**
1-chlorobutane	78	Sigma-Aldrich	benzene	80	Fischer
1,1,1-trichloroethane	74	Sigma-Aldrich	toluene	111	Sigma-Aldrich
1,2-dichloroethane	84	Sigma-Aldrich	3-ethyltoluene	161	Sigma-Aldrich

carbon tetrachloride	77	Sigma-Aldrich	4-ethyltoluene	162	Fluka
Cyclics	**Boiling Point**	**Vendor**	mesitylene	165	Sigma-Aldrich
methylcyclopentane	72	Fluka	ethylbenzene	136	Sigma-Aldrich
cyclohexane	81	Sigma-Aldrich	butylbenzene	183	Sigma-Aldrich
cyclooctane	150	Sigma-Aldrich	isobutylbenzene	170	Sigma-Aldrich
butylcyclohexane	181	Sigma-Aldrich	tert-butyl benzene	167	Sigma-Aldrich
bicyclohexyl	227	Sigma-Aldrich	propylbenzene	159	Sigma-Aldrich
2,2,4-trimethylpentane	99	Sigma-Aldrich	1-ethylnapthelene	260	Sigma-Aldrich
Alkenes	**Boiling Point**	**Vendor**	bromobenzene	155	Sigma-Aldrich
1-hexene	63	Sigma-Aldrich	cyclohexylbenzene	239	Sigma-Aldrich
cyclohexene	83	Sigma-Aldrich	diphenylmethane	265	Sigma-Aldrich
dodecene	214	Sigma-Aldrich	p-xylene	189	Sigma-Aldrich
1-undecene	194	Fluka	o-xylene	145	Sigma-Aldrich
Alkynes	**Boiling Point**	**Vendor**	m-xylene	128	Sigma-Aldrich
1-hexyne	71	Sigma-Aldrich	1,2,4-trimethylbenzene	169	Sigma-Aldrich
1-heptyne	109	Sigma-Aldrich	anisole	154	Sigma-Aldrich
1-nonyne	151	Sigma-Aldrich	dibutyl phthalate	340	Sigma-Aldrich
5-decyne	177	Sigma-Aldrich	a-terpineol (90%)	219	Sigma-Aldrich
Alcohols	**Boiling Point**	**Vendor**	**Alcohols**	**Boiling Point**	**Vendor**
1-propanol	97	Sigma-Aldrich	hexyl alcohol	158	Sigma-Aldrich
2-butanol	117	Sigma-Aldrich	2-heptanol	160	Fluka
1-pentanol	137	Sigma-Aldrich	1-octanol	194	Sigma-Aldrich
2-pentanol	158	Sigma-Aldrich	1-nonanol	215	Sigma-Aldrich
1-decanol	229	Sigma-Aldrich	1-eicosanol	372	Sigma-Aldrich
1-tetradecanol	289	Sigma-Aldrich	benzyl alcohol	205	Sigma-Aldrich
1-octadecanol	210	Sigma-Aldrich	2-ethyl-1-hexanol	185	Fluka

The following steps were taken to tune the system to achieve full modulation. While maintaining the same P_M, P_{inlet}, and P_{aux}, so the P_{inlet}/P_{aux} is slightly greater than 1, the p_w can be varied to achieve different levels of modulation (Figure 4.2). The tallest analyte peak in this section of the chromatogram of the separation of the 90-component test mixture was 1,1-bicyclohexyl, which was used for the comparison of modulation modes. To obtain the unmodulated chromatographic data (Figure 4.2A), the pulse valve was maintained open throughout the entire run, allowing for the auxiliary gas flow to continuously dilute the effluent from the ^{1}D column to the ^{2}D column, so only ^{1}D peak profiles are obtained (p_w = 0 ms). For the partially modulated analyte in the NPM (Figure 4.2B), the pulse valve was open for 990 ms (P_M

– p_w) of the 1 s P_M, introducing the auxiliary gas flow and diluting the sample transferred from the ^{1}D column to the ^{2}D column. The valve was then closed for 10 ms (p_w), allowing the effluent to flow from the ^{1}D column to the ^{2}D column undiluted during this short pulse. In the NPM, the diluted ^{1}D peak is maintained with the modulated ^{2}D peaks superimposed on the ^{1}D peak profile. The ^{2}D peaks could then be isolated from the ^{1}D peak profile by baseline subtraction [38]. Partial modulation in the PPM was compared to the other modulation methods in the previous DPGM report [39], and is not presented herein.

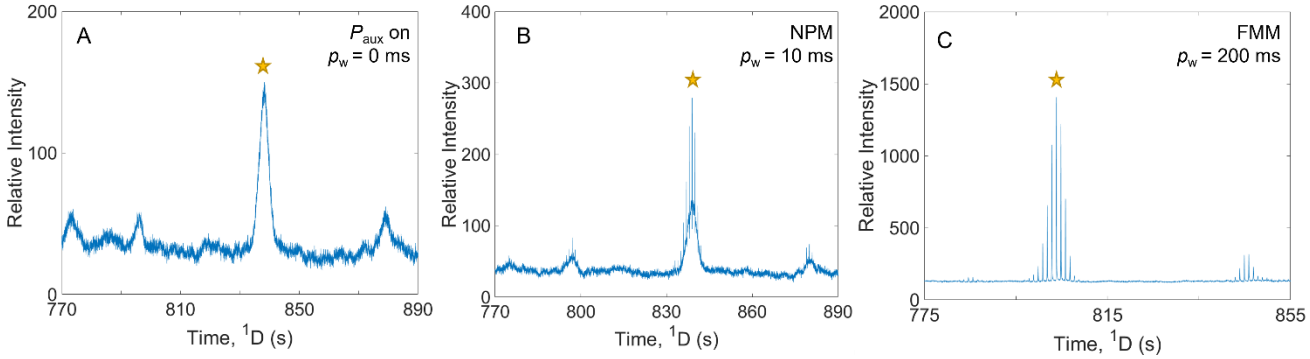


Figure 4.2 The different modes of dynamic pressure gradient modulation (DPGM). DPGM in the full modulation mode for GC×GC-TOFMS is compared to the unmodulated data and partial modulation in the negative pulse mode. For each modulation method presented P_M = 1 s and each panel presents a section of the 90-component test mixture separation containing the analyte 1,1-bicyclohexyl (labeled with *) for the sensitive *m/z* 55. (A) The section of the unmodulated chromatogram with the auxiliary flow continually on, so p_w = 0 ms. (B) The same section of the chromatogram, partially modulated in the negative pulse mode (NPM), with p_w = 10 ms. (C) The same section of the chromatogram in the full modulation mode (FMM), with p_w = 200 ms.

Selection of the appropriate p_w for the applied P_M in order to achieve full modulation followed the approach previously described [39]. Full modulation (Figure 4.2C) was achieved by keeping the pulse valve open for 800 ms (P_M – p_w) and closing it for 200 ms (p_w = 200 ms), allowing for the effluent from the ^{1}D column to be transferred to the ^{2}D column with the signal returning to baseline between modulations, i.e, the full modulation mode (FMM). In the context of the selected p_w of 200 ms, the auxiliary gas pressure applied to the pulse valve temporarily

stops the flow on the ^{1}D column during each modulation, followed by an injection onto the ^{2}D column during each p_w interval. The signal area of the entire ^{1}D peak profile of the analyte is fully distributed into the signal areas for each of the modulated ^{2}D peaks (100% duty cycle). It is noticeable that the signal scale for the full modulation case (Figure 4.2C) is ~10-fold greater compared to the unmodulated (Figure 4.2A) and partially modulated peaks in the NPM (Figure 4.2B) for this analyte. Similarly, the DPGM study for GC×GC-FID using higher auxiliary carrier gas pressures reported a 20-fold increase in signal scale for full modulation relative to both unmodulated and partially modulated peaks in NPM [39]. The narrower peaks generated using higher pressures for FID achieved this higher increase in signal. In addition, the $^{1}t_R$ for 1,1-bicyclohexyl remains ~ 840 s for the unmodulated and partially modulated peaks in the NPM, while the $^{1}t_R$ decreases to 805 s when fully modulated. This is explained by applying the auxiliary carrier gas flow for a shorter period of time due to the longer p_w for full modulation, slowing down the flow on the ^{1}D column for a shorter period of time between modulations. The same trend was observed for the FID study, where the fully modulated peaks had lower $^{1}t_R$ than the unmodulated peaks for p_w = 0 ms (P_{aux} on) [39].

4.3 Results and Discussion

Dynamic pressure gradient modulation for GC×GC-TOFMS was evaluated using the 90-component test mixture. The 10 ppth test mixture was selected for ease of visualization and determining width-at-base (W_b) values, i.e., 4 standard deviations (s.d.). The unfolded chromatogram (raw data like Figure 4.2C) of the fully modulated test mixture is shown in Figure 4.3A for the 25 min separation using a temperature program of 10 °C/min. This data produced by full modulation required no pre-processing steps prior to being cut and folded to produce the traditional 2D contour plot chromatogram (Figure 4.3B). A section from 7 min to 11 min is displayed in Figure 4.3C, providing a closer visual inspection of the separation performance and

adequate use of 2D separation space. All peaks in Figure 4.3C are numbered according to ${}^{1}t_{R}$ and are identified in Table 4.2 with a mass spectral MV provided. The cluster of analytes numbered 11 in Figure 4.3C includes several overlapped components in this ^{1}D interval that will be the focus of a chemometric decomposition study to demonstrate the suitability of DPGM to produce data amenable to chemometric analysis. In terms of separation metrics, an average ${}^{1}W_{b}$ of about 4.5 s was determined for a ^{1}D peak capacity (${}^{1}n_{c}$), of about 330 for the 25 min separation. An average ${}^{2}W_{b}$ of 130 ms was determined for the ^{2}D separation time (P_{M}) of 1 s, providing a ^{2}D peak capacity (${}^{2}n_{c}$) of 8, resulting in a 2D peak capacity ($n_{c,2D}$) of about 2700 (roughly 100 peaks/min peak capacity production). While this study was primarily intended to demonstrate that DPGM in the FMM was compatible with the TOFMS detector, further optimization of the experimental parameters should be readily achievable to improve upon this peak capacity, although the peak capacity production achieved was very satisfactory.

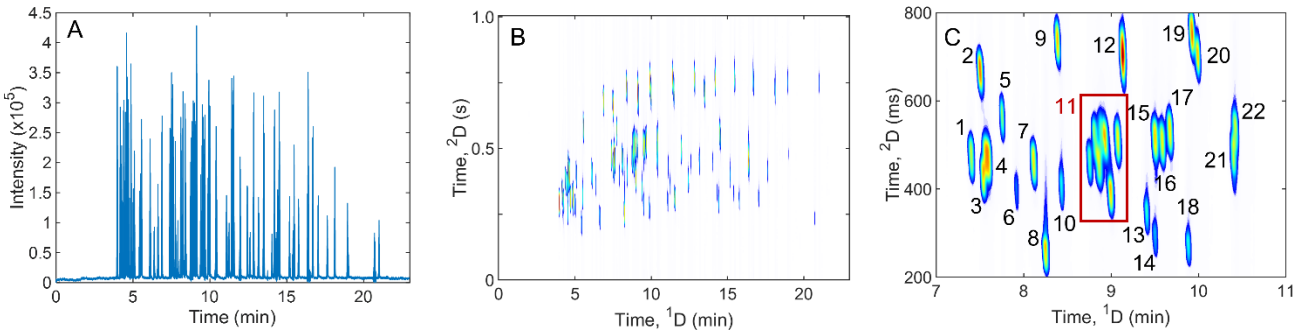


Figure 4.3. Visualization of GC×GC data using DPGM. (A) The unfolded chromatogram (total ion current, TIC) of the 10 part-per-thousand (ppth) 90-component test mixture using DPGM in the FMM (P_{M} = 1 s and p_{w} = 200 ms). (B) The 2D chromatogram created by cutting and folding (A). (C) Zoom in of the 2D chromatogram in (B) highlighting a region of overlapped components. All components are numbered according to ${}^{1}t_{R}$ and are identified in Table 4.2 with corresponding MV.

Table 4.2. Identification of numbered analytes in Figure 4.3C with corresponding MV.

Peak Number	Analyte	MV	Peak Number	Analyte	MV
1	ethylbenzene	926	12	mesitylene	849
2	cyclooctane	889	13	methyl caproate	906
3	1-chlorohexane	877	14	3-heptanone	945
4	o-xylene	953	15	tert-butylbenzene	897
5	1-nonyne	822	16	isopropylbenzene	917
6	hexyl alcohol	822	17	isobutylbenzene	945
7	p-xylene	811	18	3-octanone	818
8	2-hexanone	871	19	undecane	921
9	decane	914	20	1-undecene	901
10	2-heptanol	839	21	1-bromoheptane	924
11	Cluster – See Table 4		22	butylbenzene	912

Next, the signal enhancement achieved using DPGM with GC×GC-TOFMS was assessed by the detector response enhancement factor (*DREF*) described by Krupčik et al. [45]. The *DREF* is defined as the ratio of the height of the tallest ^{2}D peak of the modulated analyte (^2h) relative to its corresponding unmodulated peak height (^1h), or simply $DREF = {}^2h/{}^1h$. For the determination of the unmodulated peak height (^1h), data as in Figure 4.2A were used, whereas for the tallest modulated ^{2}D peak (^2h), data as in Figure 4.2C were used. An overlay of the modulated ^{2}D peaks and the unmodulated ^{1}D peak for two representative analytes for the 10 ppm test mixture is provided in Figure 4.4. For hexadecane a *DREF* of 14.9 was obtained using *m/z* 57 (Figure 4.4A), while the inset highlights that ^2W_b was 130 ms. Similar results for eicosane were obtained using the data in Figure 4.4B; a *DREF* of 17.6 was obtained using *m/z* 43, with the inset displaying a ^2W_b = 125 ms. In addition to similar ^2W_b, these analytes have similar ^1W_b of about 4.5 s. The *DREF* study for GC×GC-FID reported a much larger range of DREF values

between 7 – 87 with a greater range in ^2W_b [39]. With similar ^2W_b of about 130 ms and ^1W_b of about 4.5 s obtained for most analytes using TOFMS for detection, the *DREF* did not vary greatly. Overall, *DREFs* for several analytes ranged from 10 to 20 under the specified experimental conditions.

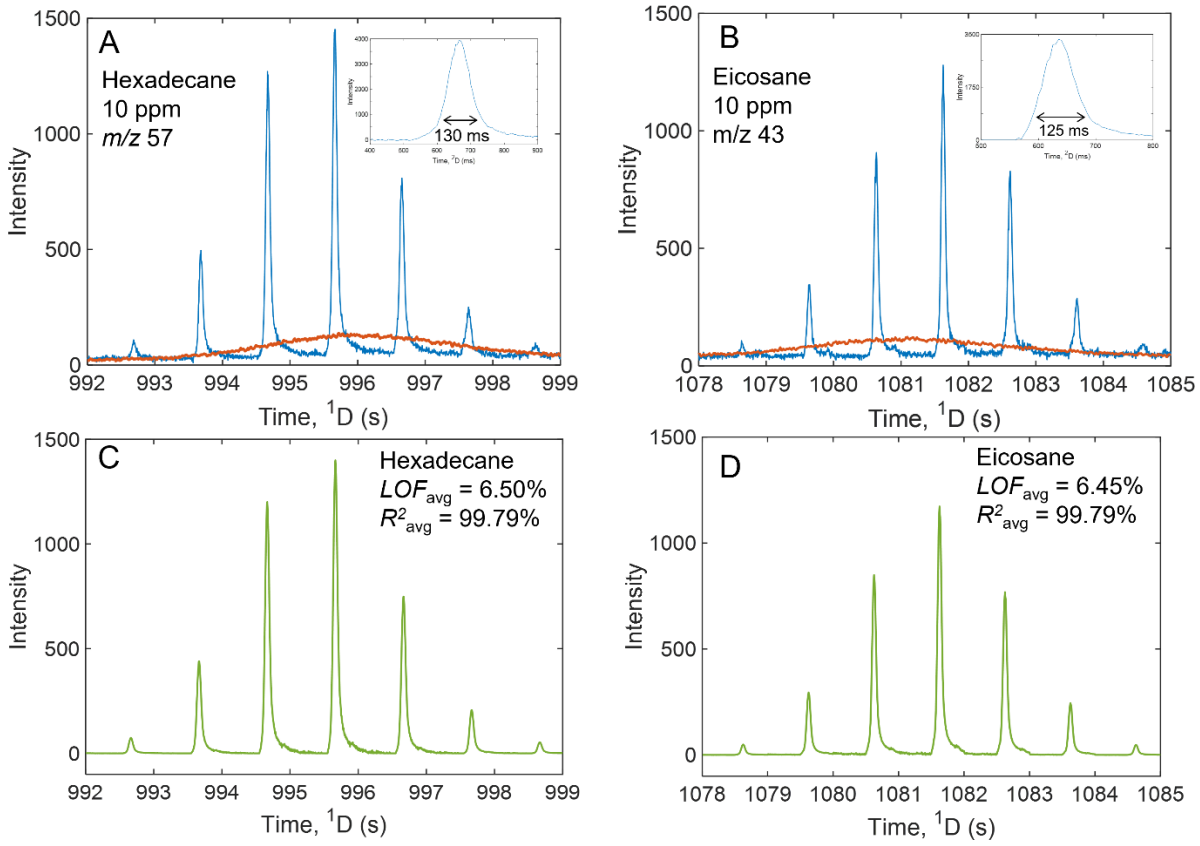


Figure 4.4. The detector response enhancement factor (DREF) using the TOFMS detector obtained by DPGM in the FMM is demonstrated for two representative analytes from the 10 ppm test mixture. The overlay of the unmodulated (orange) ^{1}D peak onto the fully modulated (blue) ^{2}D peaks is shown for (A) hexadecane (*m/z* 57) and (B) eicosane (*m/z* 43). The inset in the top right corner of each panel is included to show the ^{2}D peak profile plotted as the sum of the ^{2}D peaks. For hexadecane, ^1W_b = 4.5 s and ^2W_b = 130 ms, and for eicosane ^1W_b = 4.5 s and ^2W_b = 125 ms. (C) Hexadecane peak reconstructed from PARAFAC using *m/z* 57. Average lack-of-fit (*LOF*) and explained variance (R^2) are included from fitting (i.e., comparing) the ^{2}D raw data peaks to the PARAFAC ^{2}D peaks over ±3 s.d. intervals (1.5×^2W_b, or ~99% peak area). (D) The same workflow used to obtain (C) was applied for eicosane using *m/z* 43.

The *DREF*s were then used to determine the relative signal-to-noise $(S/N)_{Rel} = ({}^2h/{}^2s_N)/({}^1h/{}^1s_N) = DREF\times({}^1s_N/{}^2s_N)$, where ${}^1s_N$ is one s.d. of the baseline noise of the unmodulated data, and ${}^2s_N$ is one s.d. of the baseline noise of the modulated data. With a 200 *Hz* spectral scan rate, for an unmodulated ^{1}D peak, a ${}^1W_b = 4.5$ s corresponds to 900 spectra (data points), and for the modulated peak a ${}^2W_b = 130$ ms and is comprised of 26 spectra. In order to provide an objective determination of the $(S/N)_{Rel}$ prior to calculating ${}^1s_N$ and ${}^2s_N$, the unmodulated data was averaged using a 35 point boxcar average (900 spectra/26 spectra) so the ^{1}D peak would also be comprised of 26 spectra for the ${}^1s_N$ calculation, matching the number of spectra comprising a modulated ^{2}D peak for the ${}^2s_N$ calculation. Additionally, further taking the ${}^1W_b$ and ${}^2W_b$ peak widths into account, the ${}^1s_N$ and ${}^2s_N$ calculations were performed on baseline intervals of 4.5 s and 130 ms, respectively. A representative ~ 5 min section of baseline was used to calculate an average ${}^1s_N = 1.36$ (± 0.32) for the boxcar averaged baseline data, while an average ${}^2s_N = 4.41$ (± 0.68) was determined for the raw baseline data. Finally, the $(S/N)_{Rel} = 4.6$ for hexadecane (DREF = 14.9) and the $(S/N_{Rel}) = 5.4$ for eicosane (DREF of 17.6). Thus, the *S/N* of the GC×GC improves by a factor of about 5 relative to unmodulated data after adjusting the spectral acquisition frequency to put the ^{1}D and ^{2}D peaks on a level playing field.

Next, we investigated the extent to which the ^{2}D peaks produced a bilinear GC×GC data structure while using this flow modulator, since flow modulation is known to potentially impact the reproducibility of the ^{2}D peak shapes for a given modulated ^{1}D peak [46]. Having a bilinear GC×GC data structure is often critical for applying chemometric data analysis tools. PARAFAC was selected as the chemometric tool for this study due to how sensitive it is to shifts in ^{2}D retention time and differences in ^{2}D peak profile shapes. In order to provide a fully trilinear GC×GC-TOFMS data set for PARAFAC, the ^{2}D peak shapes need to be reproducible from one

^{2}D peak to the next for a given ^{1}D peak, and the ^{2}D retention times must not shift significantly [176,181]. In our data set, a slight ^{2}D retention time decrease from one ^{2}D peak to the next ($\Delta^2 t_R$) of 10 ms was observed for the two representative analytes studied (hexadecane and eicosane), due to the relatively fast temperature ramp for the primary oven (10 °C/min). For the data produced by this modulation technique, we are more concerned with examining the peak profiles rather than the ^2t_R, which can easily be corrected. In order to evaluate to what extent DPGM in the FMM impacted the reproducibility of the ^{2}D peak shapes for a given ^{1}D peak, the raw data for hexadecane and eicosane in Figure 4.4 (all of the visible 7 modulations) were initially corrected for the 10 ms retention time shift prior to comparing the ^{2}D peak shapes using PARAFAC. For this purpose we applied lack-of-fit (*LOF*) and percent variation (R^2) metrics defined by Izadmanesh et al. [50]. The PARAFAC model fits the raw data ($\Delta^2 t_R$ corrected) by forcing each modulated ^{2}D peak to retain the same peak profile, determined by the shape of each modulation from the raw data. As a point of reference, if the raw data ^{2}D peak profiles (shape, height and width) are fully reproducible, a *LOF* and R^2 comparison of the raw data of hexadecane and eicosane in Figures 4.4A,B would completely match on a per ^{2}D peak basis in Figures 4.4C,D, resulting in a *LOF* of 0% and an R^2 of 100% for each analyte.

All 7 modulations of the raw and PARAFAC modeled peaks were compared for the ^{2}D peaks over ±3 s.d. intervals (1.5×^2W_b, or ~99% peak area) centered on the corresponding ^2t_R, in triplicate, with average *LOF* and average R^2 values reported in Table 4.3. The *LOFs* are very similar and relatively low with the average *LOF* of 6.50% (± 0.90%) for hexadecane and an average *LOF* of 6.45% (± 0.12%) for eicosane [50]. The R^2 values are high for both analytes with a $R^2$of 99.79% (± 0.05%) for hexadecane and an R^2 of 99.79% (± 0.01%) for eicosane. These *LOF* and R^2 metrics indicate that the ^{2}D peaks from the PARAFAC modeling (Figures 4.4C,D)

are essentially the same shape, height and width as their counterpart peaks in the raw data (Figures 4.4A,B). In addition, in the remaining two columns of Table 4.3, the total peak area captured by PARAFAC was compared to the total peak area from the raw data before and after ^2t_R correction, respectively. To be considered an accurate model of the raw data, the area of the peaks modeled by PARAFAC should theoretically be equal to the area of the raw peaks being modeled. The sum of the peak area of the raw ^{2}D peaks was subtracted from the sum of the peak area of the ^{2}D peaks modeled by PARAFAC (PARAFAC-raw) and then divided by the sum of the raw ^{2}D peaks prior to reporting as a percentage deviation. Before the retention time correction, the total peak area calculated for the raw data and those modeled by PARAFAC were nearly equal with an average percent error of -0.69% (± 0.09%) for hexadecane and -0.60% (± 0.11%) for eicosane. The slight negative bias is likely contributed to by the slight shift in ^2t_R of 10 ms [53,54], although a negative bias of less than 1% is generally tolerable. The peak areas were calculated again for each replicate after the 10 ms ^2t_R correction with similar results. The peak areas of the data modeled by PARAFAC were essentially equal to those of the raw data with an average percent error of -0.30% (± 0.65%) for hexadecane and 0.16% (± 0.43%) for eicosane. The area is essentially conserved in both cases, suggesting that, in conjunction with the *LOF* and R^2 evaluation, the data produced by this flow modulation technique is sufficiently trilinear for the purpose of implementing a wide range of chemometric tools.

Table 4.3. Measurements to investigate quality of DPGM data. Lack-of-fit (*LOF*) and percent of the variance explained (R^2) calculated for three replicates of the 10 ppm test mix for hexadecane (*m/z* 57) and eicosane (*m/z* 43). Both metrics were calculated for the ^{2}D peaks over ±3 s.d. intervals ($1.5\times{}^2W_b$, or ~99% peak area), and the average values (±1 s.d.) are reported. The second to last and the last columns contain the percent error calculated as the relative difference in sum of the ^{2}D peak areas of the PARAFAC and raw peaks (PARAFAC-raw)/raw, prior to ^2t_R correction and after ^2t_R correction, respectively.

Hexadecane	*LOF*	R^2	Percent error in sum of ^{2}D peak areas before ^2t_R correction	Percent error in sum of ^{2}D peak areas after ^2t_R correction
Run 1	7.24%	99.75%	-0.75%	0.45%
Run 2	5.50%	99.85%	-0.71%	-0.63%
Run 3	6.76%	99.77%	-0.59%	-0.72%
Ave (±1 s.d.)	6.50% (± 0.90%)	99.79% (± 0.053%)	-0.69% (± 0.09%)	-0.30% (± 0.65%)
Eicosane				
Run 1	6.42%	99.79%	-0.48%	-0.32%
Run 2	6.34%	99.80%	-0.68%	0.52%
Run 3	6.58%	99.78%	-0.65%	0.27%
Ave (±1 s.d.)	6.45% (± 0.12%)	99.79% (± 0.009%)	-0.60% (± 0.11%)	0.16% (± 0.43%)

Next, the quantitative performance of DPGM in FMM with TOFMS detection was evaluated using the ^{2}D peak areas for the same two representative analytes hexadecane and eicosane. A serial dilution of the 90-component test mixture was carried out as described in the Experimental section (4.2). Peak areas were calculated for three replicates of hexadecane and eicosane for their highest signal *m/z* for the 10 ppb, 100 ppb, 1 ppm, and 10 ppm test mixtures. The peak area was calculated as the sum of the area of the modulated ^{2}D peaks. If the peak area (y) is linear with concentration (x), a slope of 1 is expected for the log-log calibration plot. A linear log-log calibration was obtained with a slope of 1.02 ($y = 1.02x + 2.45$) and an R^2 of 0.997 for hexadecane, and a slope of 0.98 ($y = 0.98x + 2.65$) with an R^2 of 0.996 for eicosane,

indicating a good calibration was achieved for both analytes. These are promising results supporting the use of the data obtained by DPGM in FMM for quantitative studies.

To demonstrate the detection sensitivity achieved using TOFMS detection with DPGM in the FMM, a representative analyte was studied from the test mix at a concentration of 10 ppb to determine the *LOD* facilitated by chemometric decomposition. With the high signal enhancement achieved as demonstrated by the *DREF* in Figure 4.4A, the fully modulated hexadecane *m/z* 57 peak for a representative of three replicates is shown in Figure 4.5A, with sufficient signal above the noise for this concentration to extrapolate to the *LOD*. Furthermore, using the trilinear data structure provided by GC×GC-TOFMS, PARAFAC modeling resulted in the modulated peaks in Figure 4.5B. Here, following PARAFAC, the individual ^{1}D and ^{2}D peak loadings are used to reconstruct the data in the unfolded format. Significant noise from the baseline in Figure 4.5A has been removed relative to Figure 4.5B, as one of the modeled components via PARAFAC. Analogous to applying the "two-step" method for quantification, the sum of the ^{2}D peaks in Figure 4.5B was plotted in Figure 4.5C for calculating the *LOD*, herein defined as three times one s.d. of the baseline noise signal ($LOD = 3s_N$). The *S/N* used was the ratio of the peak signal to one s.d. of the baseline noise (s_N). For three replicates of hexadecane in the 10 ppb test mix, the average *S/N* was found to be 14-fold greater than the calculated *LOD* of 0.70 ppb (± 0.03 ppb), demonstrating the high sensitivity that is achieved. Replicates of data as presented in Figure 4.5C were also used to examine quantitative reproducibility. The PARAFAC loadings that produce the ^{2}D peak profiles retain the concentration information, so the sum of the modulated peaks is proportional to the analyte concentration. An average peak area of 3180 (± 125 s.d.) was obtained for hexadecane, corresponding to a RSD of 3.93%.

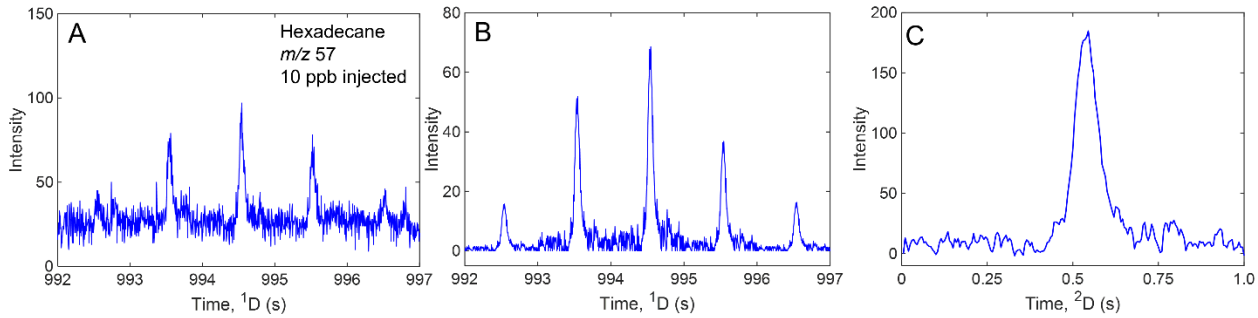


Figure 4.5. Visualization of analyte quantification. (A) The fully modulated hexadecane ^{2}D peak data (*m/z* 57) for a representative of three replicates from the 10 ppb test mix. (B) The reconstructed ^{2}D peaks following PARAFAC of the raw data in (A). (C) The single ^{2}D peak obtained by summing all of the ^{2}D peaks from (B).

PARAFAC decomposition was performed to further evaluate the quality of the data regarding the issue of trilinearity, building upon the positive findings in Figure 4.4 and Table 4.3. For this evaluation, the severely overlapped region in Figure 4.3C centered at about 9 min on ^{1}D and 450 ms on ^{2}D was subject to PARAFAC, in order to resolve the overlapped analytes and provide their mass spectra for identification. If the data are suitably trilinear, the mass spectra obtained by PARAFAC should provide a high MV, even for low concentration, highly overlapped analytes. The seven analytes from the 90-component test mixture that were resolved by PARAFAC are the following: 1-bromohexane, propylbenzene, bromobenzene, 4-ethyltoluene, 3-ethyltoluene, anisole, and 1,2,4-trimethylbenzene (Table 4.4). From Figure 4.3C, the 2D region selected for decomposition is presented in Figure 4.6A for the 10 ppth test mixture displaying the TIC chromatogram. The 10 ppth test mixture data is provided for reference, since at this higher concentration, the analytes are more readily visible, albeit severely overlapped. The same section of the 2D chromatogram is displayed in Figure 4.6B for the 1 ppm test mixture, which was analyzed by PARAFAC. In Figure 4.6B, most of the seven analytes appear indistinguishable prior to PARAFAC modeling. Following PARAFAC, the analyte peak profiles were reconstructed [47]. Here, the outer product of the loadings (^{1}D and ^{2}D peak profiles) was

taken for the *m/z* with the highest intensity of the decomposed mass spectrum of the individual fully resolved peaks. The overlay of the contour plot of each analyte (Figure 4.6C) was produced by cutting and folding the analyte response profiles generated by the outer product of the loadings for the selected *m/z*. The poor chromatographic resolution of the analytes was expected to present a challenge for decomposition and reliable identification, especially for the low concentration of the 1 ppm test mixture. In particular, propylbenzene (2), bromobenzene (3), 4-ethyltoluene (4), and 3-ethyltoluene (5) were all severely overlapped. The poor ^{1}D chromatographic resolution ($^{1}R_s$) for 4-ethyltoluene (4) and bromobenzene (3) is evident in Figure 4.6C, and retention time and peak width information from this contour plot was used to calculate the $^{1}R_s = 0.063$. A higher ^{2}D chromatographic resolution ($^{2}R_s$) of 0.61 made up for the lack of ^{1}D resolution, giving a $R_{s,2D} = 0.61$ and facilitating PARAFAC decomposition. The 2D resolution ($R_{s,2D}$) reported here is calculated as the Euclidean norm of the $^{1}R_s$ and $^{2}R_s$ [55]. Similarly, a $^{1}R_s = 0.44$ for 4-ethyltoluene (4) and propylbenzene (2) compensated for the poor ^{2}D resolution ($^{2}R_s = 0.16$), resulting in a 2D resolution of $R_{s,2D} = 0.47$ that was sufficiently high to resolve this pair of analytes. In addition to both $^{1}R_s$ and $^{2}R_s$ of ~ 0.3 ($R_{s,2D} = 0.47$), the similar mass spectra of 4-ethyltoluene (4) and 3-ethyltoluene (5) made decomposition very challenging. Overall, successful PARAFAC decomposition was achieved for all seven components for the 10 ppm and 1 ppm test mixtures. In addition, the mass spectra obtained by PARAFAC that corresponded to each loading was individually matched to the NIST spectral library to obtain MV (Table 4.4). The MV were reproducible (RSD $\leq$ 7.1%) for each component, with a majority of MV over 800. The ability to fully resolve all seven components despite the low chromatographic resolution, and low sample concentration, while providing reproducible spectra for identification, provides strong support that DPGM with GC×GC-TOFMS produces

sufficiently trilinear data, suitable for chemometric analysis with methods such as PARAFAC.

This finding significantly broadens the scope of the utility of this novel flow modulator.

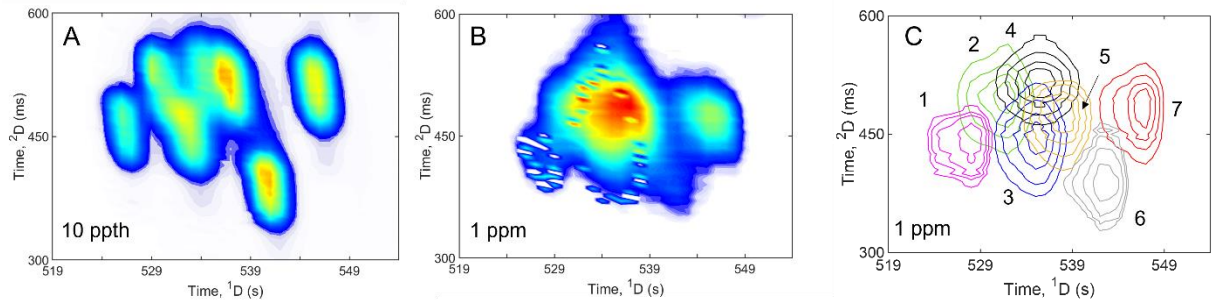


Figure 4.6. PARAFAC was used for the decomposition and identification of a heavily overlapped 2D region containing seven analytes, centered at about 9 min on ^{1}D and 450 ms on ^{2}D in Figure 4.3C. (A) The section of the TIC chromatogram of the 10 part-per-thousand (ppth) 90-component mix from Figure 4.3C provided for visual reference in order to more clearly see the peaks. (B) The same section as in (A) used for chemometric decomposition for a representative replicate of the more challenging 1 ppm test mixture. (C) The 2D contour plot created from the PARAFAC loadings from data in (B). All seven components were successfully decomposed for the 10 ppm and 1 ppm test mixtures and are numbered according to Table 4: (1) 1-bromohexane, (2) propylbenzene, (3) bromobenzene, (4) 4-ethyltoluene, (5) 3-ethyltoluene, (6) anisole, (7) 1,2,4-trimethylbenzene.

Table 4.4. The match values (MV) for the seven components following PARAFAC are shown for the 10 ppm and 1 ppm test mixtures. In parentheses next to the MV is the RSD value (%) for three replicates of each analyte.

Peak Number	Analyte	Avg MV 10 ppm	Avg MV 1 ppm
1	1-bromohexane	866 (±6.5%)	741 (±7.1%)
2	propylbenzene	868 (±0.7%)	830 (±0.8%)
3	bromobenzene	800 (±4.2%)	767 (±5.8%)
4	4-ethyltoluene	914 (±1.0%)	859 (±0.9%)
5	3-ethyltoluene	894 (±1.8%)	823 (±5.9%)
6	anisole	811 (±3.6%)	607 (±2.6%)
7	1,2,4-trimethylbenzene	931 (±0.8%)	847 (±2.5%)

4.4 Conclusion

The new differential flow modulation technique “dynamic pressure gradient modulation” (DPGM) in the full modulation mode (FMM) is demonstrated to be compatible for GC×GC-TOFMS. A ^{2}D flow rate lower than typical flow modulation techniques (4 ml/min) was employed, and the quality of the data produced was examined. Full modulation was initially compared to partial modulation in the negative pulse mode by modifying the pulse width for a given P_M, resulting in higher signal intensity and slightly earlier elution of analytes. A P_M of 1 s and p_w of 200 ms were selected for the 25 min separation of a test mixture, producing fully modulated peaks with an average $^{1}W_b$ of 4.5 s and $^{2}W_b$ of 130 ms for a 2D peak capacity ($n_{c,2D}$) of 2700, or a peak production of 100 peaks/min. The raw data required no preprocessing prior to analysis and provided a 10 to 20-fold enhancement in signal and 5-fold increase in *S/N* relative to unmodulated peaks. Representative analytes were selected to demonstrate that the data produced is sufficiently trilinear and suitable for chemometric tools and quantification. Assisted by chemometric decomposition, the *LOD* for DPGM coupled to the TOFMS was determined to be 0.7 ppb (± 0.03 ppb). Finally, the spectra obtained by decomposition of several heavily overlapped analytes at the low ppm concentration level provided reproducible identification, further supporting that this technique produces high quality data. This report presents promising results for a flow modulation technique compatible with MS detection and warrants further investigation to optimize separation efficiency.

4.5 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. doi:10.1093/chromsci/29.6.227.

[2] 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.

[3] M. Edwards, H. Boswell, T. Górecki, Comprehensive Multidimensional Chromatography, Curr. Chromatogr. 2 (2015) 80–109. doi:10.2174/2213240602666150722232236.

[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. doi:10.1016/j.trac.2011.10.013.

[5] R. Shellie, P. Marriott, A. Chaintreau, Quantitation of suspected allergens in fragrances (Part I): Evaluation of comprehensive two-dimensional gas chromatography for quality control, Flavour Fragr. J. 19 (2004) 91–98. doi:10.1002/ffj.1334.

[6] P.Q. Tranchida, G. Purcaro, A. Visco, L. Conte, P. Dugo, P. Dawes, L. Mondello, A flexible loop-type flow modulator for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1218 (2011) 3140–3145. doi:10.1016/j.chroma.2010.11.082.

[7] 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. doi:10.1016/j.trac.2018.05.017.

[8] Y. Nolvachai, C. Kulsing, P.J. Marriott, Multidimensional gas chromatography in food analysis, TrAC - Trends Anal. Chem. 96 (2017) 124–137. doi:10.1016/j.trac.2017.05.001.

[9] G. Semard, M. Adahchour, J.-F. Focant, Chapter 2 Basic Instrumentation for GC×GC, Compr. Anal. Chem. 55 (2009) 15–48. doi:10.1016/S0166-526X(09)05502-0.

[10] 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.

[11] J. V. Seeley, Recent advances in flow-controlled multidimensional gas chromatography, J. Chromatogr. A 1255 (2012) 24–37. doi:10.1016/j.chroma.2012.01.027.

[12] P.Q. Tranchida, G. Purcaro, P. Dugo, L. Mondello, Modulators for comprehensive two-dimensional gas chromatography, TrAC - Trends Anal. Chem. 30 (2011) 1437–1461. doi:10.1016/j.trac.2011.06.010.

[13] H.D. Bahaghighat, C.E. Freye, R.E. Synovec, Recent advances in modulator technology for comprehensive two dimensional gas chromatography, TrAC - Trends Anal. Chem. 113 (2018) 379–391. doi:10.1016/j.trac.2018.04.016.

[14] M. Libardoni, J.H. Waite, R. Sacks, Electrically heated, air-cooled thermal modulator and at-column heating for comprehensive two-dimensional gas chromatography, Anal. Chem. 77 (2005) 2786–2794. doi:10.1021/ac040161b.

[15] O. Panić, T. Górecki, C. McNeish, A.H. Goldstein, B.J. Williams, D.R. Worton, S. V. Hering, N.M. Kreisberg, Development of a new consumable-free thermal modulator for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1218 (2011) 3070–3079. doi:10.1016/j.chroma.2011.03.024.

[16] J. Luong, X. Guan, S. Xu, R. Gras, R.A. Shellie, Thermal Independent Modulator for Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 88 (2016) 8428–8432. doi:10.1021/acs.analchem.6b02525.

[17] V. Mucédola, L.C.S. Vieira, D. Pierone, A.L. Gobbi, R.J. Poppi, L.W. Hantao, Thermal desorption modulation for comprehensive two-dimensional gas chromatography using a simple and inexpensive segmented-loop fluidic interface, Talanta 164 (2017) 470–476. doi:10.1016/j.talanta.2016.12.005.

[18] H.J. De Geus, J. De Boer, U.A.Th. Brinkman, Development of a thermal desorption modulator for gas chromatography, J. Chromatogr. A 767 (1997) 137–151. doi:10.1016/S0021-9673(97)00038-1.

[19] B. V. Burger, T. Snyman, W.J.G. Burger, W.F. Van Rooyen, Thermal modulator array for analyte modulation and comprehensive two-dimensional gas chromatography, J. Sep. Sci. 26 (2003) 123–128. doi:10.1002/jssc.200390002.

[20] J. Harynuk, T. Górecki, New liquid nitrogen cryogenic modulator for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1019 (2003) 53–63. doi:10.1016/j.chroma.2003.08.097.

[21] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis, Anal. Chem. 70 (1998) 2796–2804. doi:10.1021/ac980164m.

[22] J. V. Seeley, F. Kramp, C.J. Hicks, Comprehensive two-dimensional gas chromatography via differential flow modulation, Anal. Chem. 72 (2000) 4346–4352. doi:10.1021/ac000249z.

[23] J. V. Seeley, N.E. Schimmel, S.K. Seeley, The multi-mode modulator: A versatile fluidic device for two-dimensional gas chromatography, J. Chromatogr. A 1536 (2018) 6–15. doi:10.1016/j.chroma.2017.06.030.

[24] C.E. Freye, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta 161 (2016) 675–680. doi:10.1016/j.talanta.2016.09.002.

[25] J. V. Seeley, N.J. Micyus, S. V. Bandurski, S.K. Seeley, J.D. McCurry, Microfluidic deans switch for comprehensive two-dimensional gas chromatography, Anal. Chem. 79 (2007) 1840–1847. doi:10.1021/ac061881g.

[26] A. Ghosh, C.T. Bates, S.K. Seeley, J. V. Seeley, High speed Deans switch for low duty cycle comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1291 (2013) 146–154. doi:10.1016/j.chroma.2013.04.003.

[27] J.F. Griffith, W.L. Winniford, K. Sun, R. Edam, J.C. Luong, A reversed-flow differential flow modulator for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1226 (2012) 116–123. doi:10.1016/j.chroma.2011.11.036.

[28] H. Cai, S.D. Stearns, Partial modulation method via pulsed flow modulator for comprehensive two-dimensional gas chromatography, Anal. Chem. 76 (2004) 6064–6076. doi:10.1021/ac0492463.

[29] T. Veriotti, M. McGuigan, R. Sacks, Pulsed flow modulation for high-speed GC using a pressure-tunable column ensemble, Anal. Chem. 73 (2001) 279–285. doi:10.1021/ac000665j.

[30] A.M. Muscalu, M. Edwards, T. Górecki, E.J. Reiner, Evaluation of a single-stage consumable-free modulator for comprehensive two-dimensional gas chromatography: Analysis of polychlorinated biphenyls, organochlorine pesticides and chlorobenzenes, J. Chromatogr. A 1391 (2015) 93–101. doi:10.1016/j.chroma.2015.02.074.

[31] B.D. Fitz, R.B. Wilson, B.A. Parsons, J.C. Hoggard, R.E. Synovec, Fast, high peak capacity separations in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A 1266 (2012) 116–123. doi:10.1016/j.chroma.2012.09.096.

[32] D.R. Deans, A new technique for heart cutting in gas chromatography [1], Chromatographia. 1 (1968) 18–22. doi:10.1007/BF02259005.

[33] P.Q. Tranchida, S. Salivo, F.A. Franchina, L. Mondello, Flow-modulated comprehensive two-dimensional gas chromatography combined with a high-resolution time-of-flight mass spectrometer: A proof-of-principle study, Anal. Chem. 87 (2015) 2925–2930. doi:10.1021/ac5044175.

[34] C.E. Freye, L. Mu, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1424 (2015) 127–133. doi:10.1016/j.chroma.2015.10.098.

[35] H.D. Bahaghighat, C.E. Freye, D. V. Gough, P.E. Sudol, R.E. Synovec, Ultrafast separations via pulse flow valve modulation to enable high peak capacity multidimensional gas chromatography, J. Chromatogr. A 1573 (2018). doi:10.1016/j.chroma.2018.08.001.

[36] C.E. Freye, H.D. Bahaghighat, R.E. Synovec, Comprehensive two-dimensional gas chromatography using partial modulation via a pulsed flow valve with a short modulation period, Talanta 177 (2018) 142–149. doi:10.1016/j.talanta.2017.08.095.

[37] H.D. Bahaghighat, C.E. Freye, D. V. Gough, R.E. Synovec, Comprehensive two-dimensional gas chromatography and time-of-flight mass spectrometry detection with a 50 ms modulation period, J. Chromatogr. A 1583 (2019) 117–123. doi:10.1016/j.chroma.2018.11.027.

[38] D. V. Gough, D.H. Song, S. Schöneich, S.E. Prebihalo, R.E. Synovec, Development of Ultrafast Separations Using Negative Pulse Partial Modulation to Enable New Directions in Gas Chromatography, Anal. Chem. 91 (2019) 7328–7335. doi:10.1021/acs.analchem.9b01085.

[39] T.J. Trinklein, D. V. Gough, C.G. Warren, G.S. Ochoa, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography, J. Chromatogr. A (2019) 460488. doi:10.1016/j.chroma.2019.460488.

[40] R.E. Mohler, B.J. Prazen, R.E. Synovec, Total-transfer, valve-based comprehensive two-dimensional gas chromatography, Anal. Chim. Acta 555 (2006) 68–74. doi:10.1016/j.aca.2005.08.072.

[41] N. Oldridge, O. Panic, T. Górecki, Stop-flow comprehensive two-dimensional gas chromatography with pneumatic switching, J. Sep. Sci. 31 (2008) 3375–3384. doi:10.1002/jssc.200800265.

[42] P.Q. Tranchida, F.A. Franchina, P. Dugo, L. Mondello, Flow-modulation low-pressure comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1372 (2014) 236–244.

[43] F.A. Franchina, M. Maimone, P.Q. Tranchida, L. Mondello, Flow modulation comprehensive two-dimensional gas chromatography-mass spectrometry using ≈ 4 mL min-1 gas flows, J. Chromatogr. A 1441 (2016) 134–139. doi:10.1016/j.chroma.2016.02.041.

[44] P.Q. Tranchida, F.A. Franchina, P. Dugo, L. Mondello, Use of greatly-reduced gas flows in flow-modulated comprehensive two-dimensional gas chromatography-mass spectrometry, J. Chromatogr. A 1359 (2014) 271–276.

[45] J. Krupčík, P. Májek, R. Gorovenko, I. Špánik, P. Sandra, D.W. Armstrong, On the determination of a detector response enhancement factor for flow modulated comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1286 (2013) 235–240. doi:10.1016/j.chroma.2013.02.068.

[46] P.M. Harvey, R.A. Shellie, Factors affecting peak shape in comprehensive two-dimensional gas chromatography with non-focusing modulation, J. Chromatogr. A 1218 (2011) 3153–3158. doi:10.1016/j.chroma.2010.08.029.

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

[48] 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.

[49] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146.

[50] 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.

[51] T.C. Tran, G.A. Logan, E. Grosjean, J. Harynuk, D. Ryan, P. Marriott, Comparison of column phase configurations for comprehensive two dimensional gas chromatographic analysis of crude oil and bitumen, Org. Geochem. 37 (2006) 1190–1194. doi:10.1016/j.orggeochem.2006.05.006.

[52] M. Jennerwein, M. Eschner, T. Wilharm, T. Gröger, R. Zimmermann, Evaluation of reversed phase versus normal phase column combination for the quantitative analysis of common commercial available middle distillates using GC × GC-TOFMS and Visual Basic Script, Fuel 235 (2019) 336–338. doi:10.1016/j.fuel.2018.07.081.

[53] 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.

[54] 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. doi:10.1016/j.chroma.2019.460368.

[55] R.E. Murphy, M.R. Schure, J.P. Foley, Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography, Anal. Chem. 70 (1998) 1585–1594. doi:10.1021/ac971184b.

Chapter 5: A systematic investigation of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry with dynamic pressure gradient modulation for high peak capacity separations

This chapter was reproduced from S. Schöneich[†], T.J. Trinklein[†], C.G. Warren, R.E. Synovec, A systematic investigation of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry with dynamic pressure gradient modulation for high peak capacity separations, Anal. Chim. Acta 1134 (2020) 115–124.

[†] These authors contributed equally to this work

5.1 Introduction

The emergence of comprehensive two-dimensional (2D) gas chromatography (GC×GC) [1] has been one of the most enabling developments in the GC field since the invention of capillary columns [2]. There are several benefits of GC×GC relative to one-dimensional GC, including about a 10-fold gain in peak capacity [3], added chemical selectivity, and the potential for lower detection limits due to zone compression in the modulator. All of these benefits facilitate a more informative analysis of complex samples in a variety of studies: petroleum [4], metabolomics [5–7], food science [8,9], and environmental [10], among others [11].

In order to perform a comprehensive two-dimensional separation, two capillary columns of sufficiently different selectivity must be interfaced such that a longer first dimension (^{1}D) separation can be performed simultaneous with obtaining rapid separations on the second dimension (^{2}D) column. This is accomplished using a modulator, which collects portions of ^{1}D effluent, at a user-selected modulation period (P_M), then rapidly injects each portion onto the ^{2}D column. The effluent leaving the ^{2}D column is then monitored in real time by a suitable detector. The resulting one-dimensional data array is reshaped into an $m \times n$ matrix, where the length of n

is the number of modulations, and the length of *m* is the number of data points per modulation, which is dependent on the data collection frequency and the P_M applied.

Modulators can be broadly categorized into two classes: thermal [1,12] and flow-based [13–17]. Thermal modulators collect analyte by cold or phase-ratio trapping followed by rapid injection onto ^{2}D by thermal desorption. Detectability improves by this analyte preconcentration, with total transfer of ^{1}D effluent to the ^{2}D column, referred to as a 100% duty cycle [18,19]. While undeniably a powerful approach, cryogen use necessitates higher operating costs. More recently, consumable-free thermal modulation shows promising performance at a fraction of the operating cost of cryogenic systems [20–23].

Flow modulation (i.e., valve-based, pneumatic, or fluidic modulation) was introduced by our group [13] and further developed by Seeley [14,17,24] and others [25–27] as a simple, low-cost alternative to thermal modulation. Flow modulators are capable of operating at a broad P_M range (50 ms $\leq P_M \leq$ 6 s) [182,183]. Also, light compounds (down to C_1) are efficiently modulated, which are difficult to trap using thermal modulation. Flow modulation does not truly "focus" analyte by concentrating analyte molecules relative to the carrier gas molecules. Instead, a sensitivity enhancement is provided by compressing the ^{1}D effluent in *time* since a much higher flow rate is generally used along ^{2}D relative to ^{1}D [30,31]. The higher analyte flux increases detection sensitivity with mass-sensitive detectors such as flame ionization detection (FID), time-of-flight mass spectrometry (TOFMS), and quadrupole mass spectrometry (qMS) [32].

Despite the potential benefits of flow modulation, their use is not widespread and only recently have they been commercially introduced. Indeed, flow modulation often results in discarding a portion of ^{1}D effluent to waste (duty cycle < 100%), which decreases detection

sensitivity, or involves use of a high ^{2}D flow rate, which is a compatibility challenge with MS detection. Recent advances have aimed to address these challenges. High duty cycle flow modulation has been interfaced with atmospheric pressure chemical ionization (APCI) which reduces the need to slow down the ^{2}D flow [33]. Conversely, low duty cycle designs that do not require a high ^{2}D flow are gaining acceptance. One such design, the Flux Flow modulator, was commercialized, based on the multi-mode modulator [34]. It is simple to use and effective but discards a large portion of ^{1}D effluent to waste. Despite these recent advances, total-transfer flow modulation with MS detection (without flow-splitting) is not commercially available or widely implemented. Therefore, there is an ongoing need for the development of high-performance GC×GC separations with MS using total transfer differential flow modulation.

To address this need, we introduced a differential flow modulation technique referred to as dynamic pressure gradient modulation (DPGM), which provides a 100% duty cycle and is compatible with TOFMS [35,36]. Herein, we provide optimization steps and demonstrate improved modulator performance in terms of 2D peak capacity ($n_{c,2D}$) for GC×GC-TOFMS relative to our previous report [36]. An improved design for the modulator interface is described which minimizes dead volumes and increases modulator longevity. Furthermore, we investigate the impact of sampling-induced band broadening of the ^{1}D peaks at various levels, to assess the impact on $n_{c,2D}$. The effect of undersampling alone due to the modulated ^{1}D peak widths [37] is related to the effective ^{1}D peak widths that considers both undersampling and sampling variation via statistical overlap theory (SOT) [38]. Next, we evaluate the modulation conditions that provide the maximum usage of the 2D separation space in the context of the fractional space occupied, which is also used to correct $n_{c,2D}$ [39]. Finally, broad applicability is demonstrated

with the analysis of diesel fuel, derivatized cow serum, coffee headspace using head-space solid-phase microextraction (HS-SPME), and contaminated river water (with HS-SPME).

5.2. Theory

The ideal 2D peak capacity [14] is described by

$$n_{c,2D} = {}^{1}n_c \times {}^{2}n_c \tag{5.1}$$

where ${}^{1}n_c$ is the temporal portion of the ^{1}D separation (t_{sep}) occupied by analytes divided by the experimentally measured true ^{1}D peak width ($^{1}W_b$), obtained by running the GC×GC separation without modulating. The peak capacity on ^{2}D ($^{2}n_c$) is determined as previously described [184]. Beyond Eq. (5.1), we seek to compare the 2D peak capacity that is impacted by undersampling alone, versus the 2D peak capacity that considers both undersampling and sampling variation via SOT, in order to assess their relative contribution. The 2D peak capacity impacted by undersampling alone is given by

$$n_{c,2D}{}^{*} = {}^{1}n_c{}^{*} \times {}^{2}n_c \tag{5.2}$$

where ${}^{1}n_c$* is given by ${}^{1}n_c = t_{sep}/{}^{1}W_b$*, with ${}^{1}W_b$* the ^{1}D peak width measured directly from the modulator broadened ^{1}D peak profiles in the unfolded GC×GC data [33]. In our comparison of results from Eqs. (5.1) and (5.2), we also relate the experimentally measured ${}^{1}W_b$ and ${}^{1}W_b$*, to the predicted ^{1}D peak width ($^{1}W_b$**) that should in principle match ${}^{1}W_b$ [33].

The 2D peak capacity that considers both undersampling and sampling variation via SOT is given by [18]

$$n_{c,2D,\beta} = {}^{1}n_{c,\beta} \times {}^{2}n_c \tag{5.3}$$

where ${}^{1}n_{c,\beta}$ is $t_{sep}/{}^{1}W_{b,\beta}$, where ${}^{1}W_{b,\beta}$ is the ^{1}D peak width from Eq. (5.1), ${}^{1}W_b$, multiplied by $\langle\beta\rangle$,

$$\langle\beta\rangle = \sqrt{1 + 3.4\left(\frac{1}{M_R}\right)^2} \tag{5.4}$$

with M_R the modulation ratio (*aka* sampling density [185,186]), given by ${}^1W_b/P_M$. While it is anticipated that $n_{c,2D} > n_{c,2D}^* > n_{c,2D,\beta}$, we seek to examine the relative contributions of these various effects. Additionally, since Eqs. (5.1-5.3) imply that the entire 2D space along t_{sep} is occupied by analytes, which is seldom achieved for a given sample, the 2D peak capacities can be corrected by the fraction of the 2D space that is occupied ($f_{coverage}$). A recent evaluation of several area-based metrics for determining $f_{coverage}$ determined that the minimum convex hull method was the most independent of sample dimensionality [39],

$$f_{\text{coverage}} = \frac{A_{\text{enclosed}}}{t_{\text{sep}} \cdot P_{\text{M}}} \tag{5.5}$$

where $A_{enclosed}$ is the area within the minimum convex hull. The minimum convex hull, in the context of a 2D separation, is the smallest polygon that be drawn around a set of retention time coordinates with no interior angle exceeding 180°. The method can be visualized as snapping a rubber band around the retention time coordinates and calculating the area enclosed.

5.3. Experimental

5.3.1 Instrumentation

The instrumental platform in Fig. 5.1A was based on a modified commercial GC×GC-TOFMS, comprised of an Agilent 6890 gas chromatograph (Agilent Technologies, Palo Alto, CA, USA) coupled to a LECO Pegasus III TOFMS (LECO Corporation, St. Joseph, MI). The ^{1}D and ^{2}D columns were joined via a T-union with the third port interfaced to a commercial 2-way solenoid "pulse" valve (Model 0 09–0347–900, Parker Hannifin, Hollis, NH) using an in-house made interface. Modulation is provided by rapidly pulsing an auxiliary pressure (P_{aux}) at the T-union joining the ^{1}D and ^{2}D columns. Injection onto ^{2}D occurs when auxiliary flow is not applied, and the ^{1}D effluent previously stopped, or "loaded," at the end of the ^{1}D column can pass the T-union junction. Then, when suitable pressure is reapplied to the T-union, flow along

^{1}D temporarily stops simultaneous with separation of the injection onto ^{2}D. We describe modulation optimization in terms of the pulse width (p_w), i.e., the time the valve is closed for every P_M, and the applied P_{aux}.

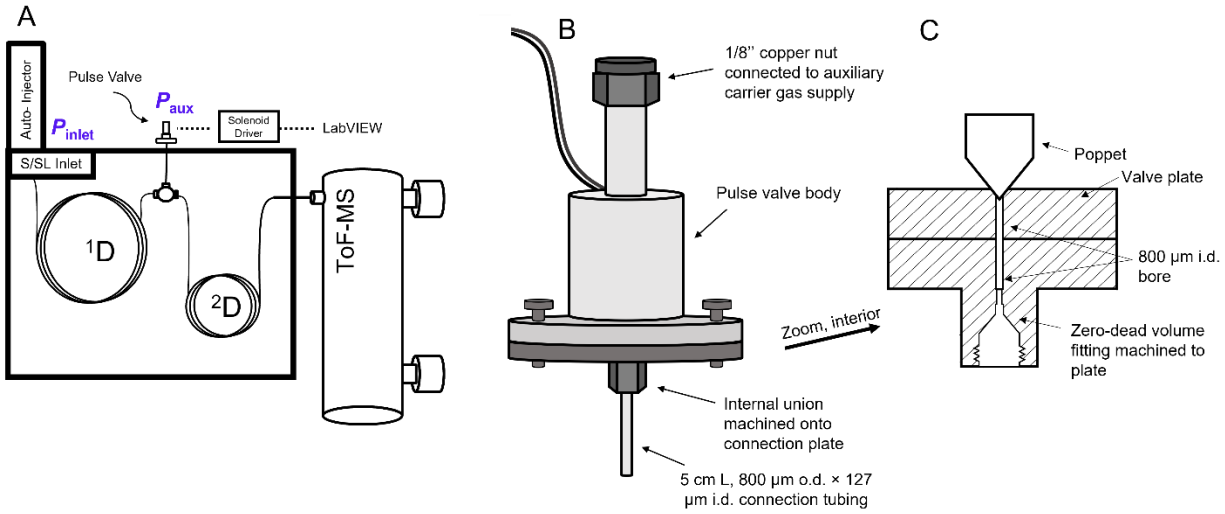


Figure 5.1. Schematic of the instrument with a zoom-in of the modulator. (A) Schematic diagram of the GC×GC-TOFMS instrument. P_{inlet} refers to the pressure at the inlet of the ^{1}D column, and P_{aux} refers to the auxiliary gas pressure applied to the valve, which is tuned to achieve full modulation. (B) Schematic of the exterior of the pulse valve including the connection to the auxiliary gas supply, the pulse valve body, the internal union, and the tubing that connects the valve to the T-union at the junction of the ^{1}D and ^{2}D columns. (C) Zoom-in of the interior of the valve including the custom fitting at the bottom of the interior of the valve made to eliminate dead volume.

A schematic of the solenoid valve and interface exterior is shown in Fig. 5.1B. The interface is constructed from a circular stainless-steel plate machined to a zero-dead volume 1/32" union. A zoom-in of the interior region showing the bottom portion of the solenoid valve and the in-house fitting is given in Fig. 5.1C. The interface is constructed such that the connection tubing can be seated just inside the valve interior, touching the poppet during the normally closed (NC) position. This construction eliminates the potential for dead volumes which may cause injection-induced band broadening since ^{2}D peak widths have been shown to be dependent on the time of the pressure pulse applied to the junction [35]. To accomplish this,

the base plate of the pulse valve was removed, and the exit orifice was reamed out to 800 µm, thus matching the o.d. of the connection tubing. In previous reports, the pulse valve orifice was either much larger/or smaller than the connection tubing or employed a "cone" exit which caused leaks along with the potential for a dead volume(s). Additionally, we have observed better performance of the modulator, in terms of the observed ^{2}D peak widths, using very short and/or narrow connection tubing with a small active volume between the valve and the T-union. Therefore, a 5 cm length of 127 µm i.d. length of tubing was used, which was the shortest length that could be practically used while keeping the pulse valve outside of the oven. More detail of the modulator assembly, interface, and software control is provided in Fig. B.1 in Appendix B.

One column set was applied for the analysis of a 116-component test mixture and various complex samples (cow serum, river water headspace, and coffee headspace), and another column set was applied for a diesel sample. Sample preparation and specific analysis conditions for all the complex samples are provided in the Appendix B. For the separations of the 116-component test mixture (Table B.1 in Appendix B), derivatized cow serum, coffee headspace (using HS-SPME), and contaminated river water (using HS-SPME), a standard non-polar × polar column set was applied. The ^{1}D column had a DB-5 stationary phase (5% phenyl polydimethylsiloxane): 18.5 m with a 180 µm capillary i.d. (d_c) and 0.18 µm film thickness (d_f). The ^{2}D column had a DB-WAX stationary phase (polyethylene glycol): 4 m with a 180 µm d_c and 0.18 µm d_f. This relatively long ^{2}D column [44] optimally used the separation space, enabling use of a longer P_M of 2 s relative to previous work [63], thus maximizing the peak capacity. For the diesel fuel, a "reverse format" polar × non-polar column configuration was applied. The ^{1}D column had an Rxi-17 stationary phase (50% phenyl-methylpolysiloxane): 20 m with a 180 µm d_c and 0.18 µm d_f. The ^{2}D column had a DB-5 stationary phase (5% phenyl-polydimethylpolysiolxane): 4 m with

a 180 µm d_c and 0.18 µm d_f. When employing a "reverse format" polar × non-polar column configuration, analytes of widely varying boiling points elute at the same time on ^{1}D due to polar interactions, causing analytes to be separated on ^{2}D largely by vapor pressure, even when using a mid-polar ^{2}D column [45]. Reverse column configurations enable excellent use of the ^{2}D separation space for fuel samples [46].

5.3.2 Experimental parameters

For the separations of the 116-component test mixture, derivatized cow serum, coffee headspace (using HS-SPME), and contaminated river water (using HS-SPME), the oven temperature was held at 50 °C for 5 min, then ramped at 8 °C/min to 290 °C where it was held for 5 min for a total run time of 40 min. The inlet was operated in constant flow mode at 1.55 ml/min by setting the ^{1}D column length at 20 m and the ^{2}D column length at 4 m in ChromaTOF (LECO, St. Joseph, MI) with vacuum outlet pressure. This corresponded to a ramp of P_{inlet} from 323 kPa (5 min hold) to 513 kPa (3 min hold) at 7 kPa/min. Since the d_c and d_f between ^{1}D and ^{2}D was constant, calculation of an equivalent column length to input into the GC software was unnecessary. Although the flow was set to 1.55 ml/min, the average ^{1}D volumetric flow rate (^1F) was estimated at ~0.10 ml/min by converting the observed linear velocity, as calculated from the dead time and known column length, to a flow rate using the Agilent Pressure Flow Calculator (Agilent Technologies, Palo Alto, CA, USA). The decreased flow rate is due to the temporary stop flow conditions because of applying DPGM. The optimized auxiliary pressure, P_{aux}, applied to the pulse valve was 330.9 kPa (33 psig), where it was held for 3 min, followed by a ramp at 4.1 kPa/min to 379.7 kPa (55 psig). This corresponded to an average flow at the ^{2}D column outlet of ~6 ml/min. Although an exit flow of 6 ml/min is relatively high compared to most GC-based methods using MS detection, overheating of the turbopump was not observed and the

quality of the mass spectra obtained was not affected in any readily discernable way. The TOFMS was set to collect from mass channels 33 – 223 *m/z* at a collection rate of 200 Hz. The electron impact ionization energy was −70 eV and the detector voltage was 1562 V.

The 116-component test mixture used to evaluate the instrumental platform contained a wide variety of analytes (Table B.1 in Appendix B) which were all ~1 parts-per-thousand in methanol. Triplicate chromatograms were obtained by 1 µl injections at a 1:100 split ratio. First, the effect of the applied P_{aux} on separation performance was studied by changing the initial pressure and final pressures, while keeping all other parameters the same, including the P_{aux} ramp rate. Separations at an initial P_{aux} from 296.5 kPa to 341.3 kPa were obtained in triplicate, at intervals of 3.5 kPa. The effect of pulse width (p_w) was investigated by varying p_w while keeping all other parameters constant, using the optimized initial P_{aux} of 330.9 kPa (33 psig). Triplicate chromatograms were obtained at p_w of 0, 40, 80, 120, 160, 200, 240, 280, 360, 440, 520, 600, 680, 840, 1000, 1320, and 1960 ms. All calculations were performed in MATLAB 2019a (The Mathworks, Inc., Natick, MA, USA)., with Gaussian curve fitting obtained using the Curve Fitting Toolbox.

5.4 Results and Discussion

Optimization of the DPGM conditions for GC×GC-TOFMS was evaluated with the 116-component test mixture. The total ion current (TIC) chromatogram (Fig. 5.2A) displays the optimized 30 min separation using a P_M of 2 s and p_w of 200 ms. A zoom in (Fig. 5.2B) highlights the excellent use of 2D separation space, concurrent with the 2D peak widths (^2W_b) at ±2 standard deviations (s.d.), remaining narrow relative to the P_M throughout the ^{2}D separations. The peak capacity on ^{2}D, was determined as previously described [41] using 10 analytes with a wide range of ^2k' with ^2W_b ranging from 102 to 221 ms (see Appendix B), resulting in a ^2n_c = 15

with P_M = 2 s (Fig. 5.2C). The 10 analytes used for this ${}^2n_c$ determination and their ${}^2W_b$ are provided in Table B.2 in Appendix B. The analyte 1-bromoheptane (Fig. 5.2D) is used to demonstrate how the various ^{1}D peak widths in the context of Eqs. (5.1-5.3) were determined: experimentally measured ${}^1W_b$ and ${}^1W_b$* at ±2 s.d., predicted ${}^1W_b$** [37], and <β> corrected ${}^1W_{b,\beta}$ [38]. Pinkerton et al. demonstrated that while the true ^{1}D peak width (${}^1W_b$) has to be measured experimentally by not modulating the GC×GC instrument (essentially the same as placing a separate detector prior to the modulator), ${}^1W_b$ can also be predicted from modulated data via knowledge of the true M_R, herein referred to as ${}^1W_b$** [37]. First, the experimental peak width (${}^1W_b$*) was measured by fitting a Gaussian curve to the modulated data. For 1-bromoheptane the ${}^1W_b$*= 5.50 s, which is inflated from undersampling. For a P_M = 2 s, the effective modulation ratio (M_R*=${}^1W_b$*/P_M) is 2.75. Using the information in Table 5.2 from the study conducted by Pinkerton et al., the true modulation ratio was determined to be M_R = 2.48, corresponding to a ${}^1W_b$ ** = P_M×M_R= 4.96 s. Next, the peak width of unmodulated 1-bromoheptane was measured and determined to be ${}^1W_b$ = 4.95 s, which is essentially identical to the predicted ${}^1W_b$**. Next, the 2D peak capacity that considers both undersampling and sampling variation via SOT was investigated per Eq. (5.3). The <β> was calculated and multiplied by the ${}^1W_b$ to determine ${}^1W_{b,\beta}$. For 1-bromoheptane, <β> = 1.24 and ${}^1W_{b,\beta}$= 6.10 s.

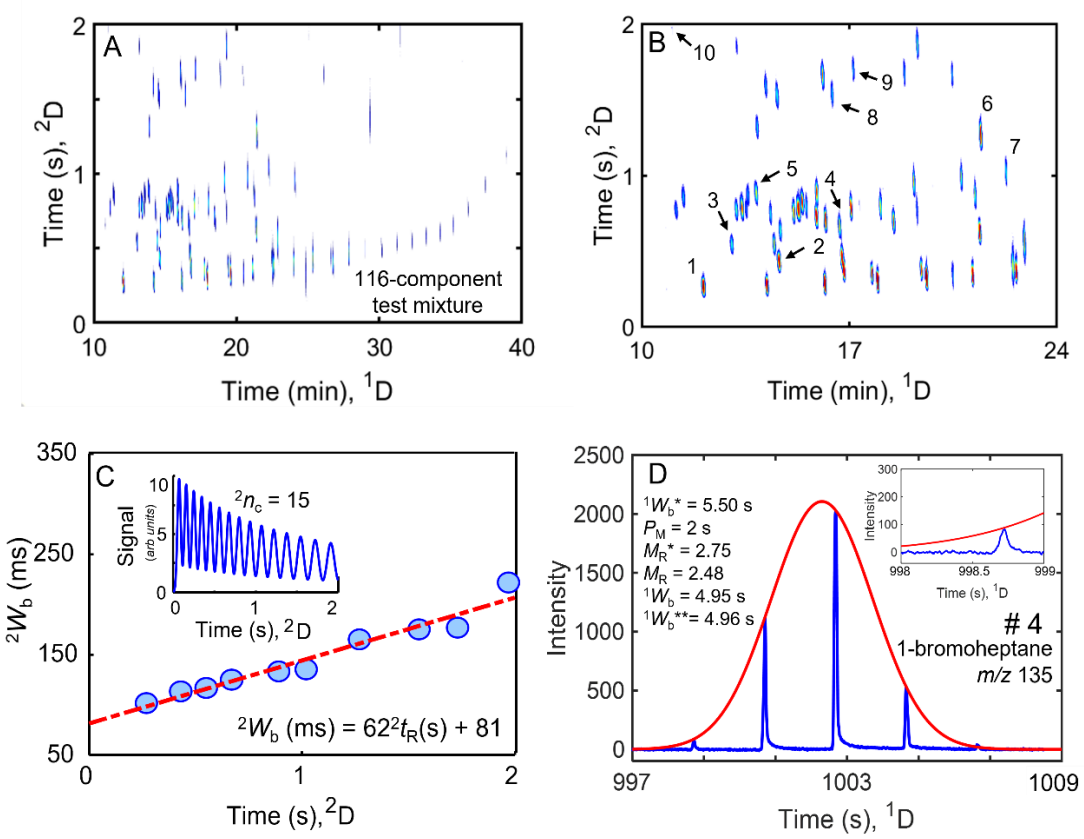


Figure 5.2. 116-component sample separation and visualization of sampling-induced band broadening. (A) TIC chromatogram of the 30 min GC×GC-TOFMS separation of the 116-component test mixture. (B) Zoom-in from 10 – 24 min of the separation in (A). Analytes spanning all ^2k' used for the calculation of ^2n_c in (C) are numbered. (C) Linear relationship between ^2W_b and ^2k' used for the iterative calculation of retention times and ^2W_b (plotted in inset) to obtain ^2n_c = 15 for the 116-component test mixture [184]. (D) A Gaussian curve (red) was fit to the modulated 1-bromoheptane peak to determine the effective ^{1}D peak width (^1W_b*). From the ^1W_b*, the M_R*, M_R, and ^1W_b** were using Table 5.2 in reference [37]. The inset demonstrates how well the Gaussian curve was fit to the modulated data for the small ^{2}D peak between 998-999 s. The ^1W_b measured from the unmodulated data is also reported here for comparison.

Optimization of the separation and modulation conditions in Fig. 5.2 was achieved by tuning two main parameters, P_{aux} (the pressure applied to the valve), and p_w (how long the valve is closed). For demonstration purposes, P_{aux} and p_w were tuned individually, however, in practice it is recommended to tune them simultaneously. For a user-defined P_M and p_w, the P_{aux} must be suitably high relative to P_{inlet} to achieve full modulation but not too high to diminish the 2D separation space used. The effect of P_{aux} applied to the valve on the resulting modulation depth

[47], which describes the extent to which the signal reaches the baseline between modulations, (Fig. 5.3A) and $^{2}W_{b}$ (Fig. 5.3B) was studied first. To do so, the initial P_{aux} of 296.5 kPa (28 psig) was raised by increments of 3.5 kPa (0.5 psig) to 330.9 kPa (33 psig) for each successive run while maintaining a constant P_{aux} ramp rate of 4.1 kPa/min. A normalized modulated nonane peak (*m/z* 57) and the beginning of the next modulated peak is shown in Fig. 5.3A for each pressure program for visual interpretation of the extent to which the analyte is modulated. Note that with an initial P_{aux} of 330.9 kPa full modulation is achieved and the ^{2}D peak width, $^{2}W_{b}$, is narrower, and for lower P_{aux} the ^{2}D peak is slightly wider and no longer returns to baseline between modulations, rather the modulated ^{2}D peaks are superimposed on the underlying ^{1}D profiles referred to as the negative pulse mode regime, observed here in Fig. 5.3A with an initial P_{aux} = 296.5 kPa [47–49]. A greater range of P_{aux} is provided to demonstrate the effect on $^{2}W_{b}$ (Fig. 5.3B) for representative analytes nonane (*m/z* 57) and p-xylene (*m/z* 91), with the optimal initial P_{aux} indicated with a star. Error bars are representative of ±1 s.d. for $^{2}W_{b}$ measurements obtained from triplicate chromatograms. The optimal initial P_{aux} of 330.9 kPa was selected due to the narrow $^{2}W_{b}$ achieved concurrent with a maximum amount of 2D space used, whereby the $f_{coverage}$ obtained via Eq. (5.5) was a quantitative metric. For the optimized test mixture separation in Fig. 5.2A, application of the minimum convex hull method is shown in Fig. 5.3C, with an $f_{coverage}$ = 0.78 (P_{aux} = 330.9 kPa). Similarly, $f_{coverage}$ was calculated for all chromatograms used to prepare Fig. 5.3B, with the result given in Fig. 5.3D. For P_{aux} above the optimum, slightly narrower $^{2}W_{b}$ were obtained, though the $f_{coverage}$ was significantly lower; since analytes have less residence time on the ^{2}D column resulting in less use of the 2D space. Concurrently, further increasing P_{aux} slows flow on ^{1}D to greater extent, resulting in increasing ^{1}D peak widths. Note that the ^{1}D flow rate could be increased by increasing P_{inlet}, however, this would call for further

increasing P_{aux} to maintain full modulation. For P_{aux} below the optimum, the $f_{coverage}$ was near that at the optimum, but the chromatographic data was in the negative pulse mode regime.

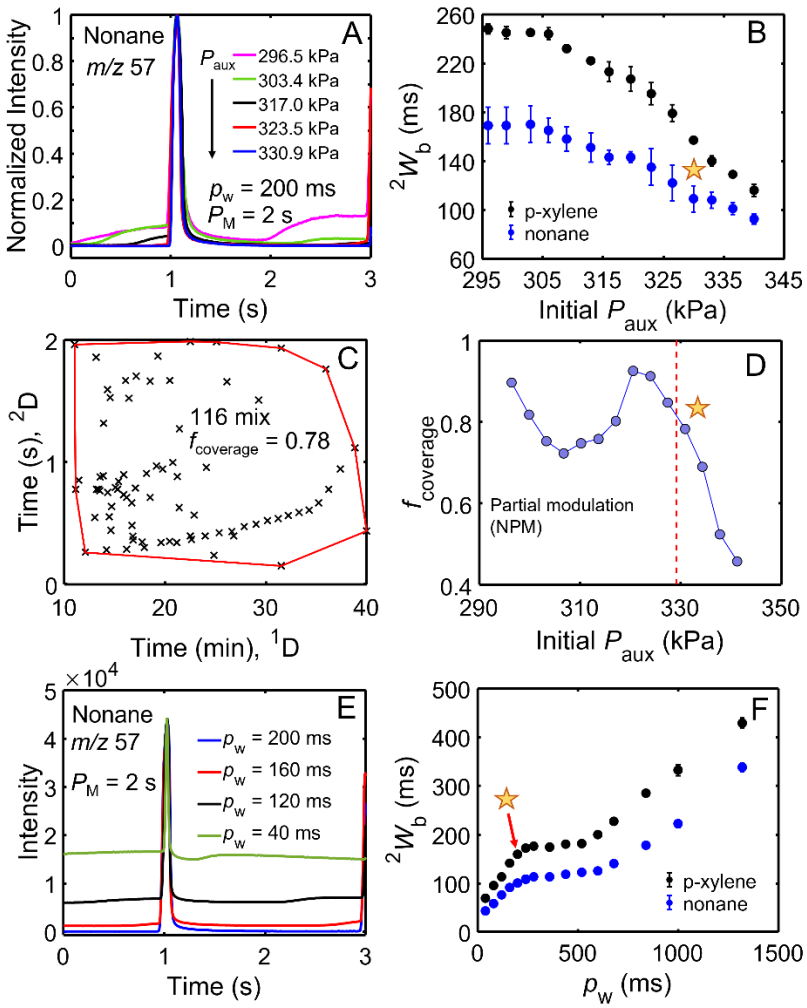


Figure 5.3. Demonstration of tuning P_{aux} and p_w. (A) Overlay of normalized single ^{2}D nonane peak at P_{aux} = 296.5 kPa, 303.4 kPa, 317 kPa, 323.9 kPa, and 330.9 kPa for using *m/z* 57. The P_M = 2 s and p_w = 200 ms. (B) Calculated average ^2W_b values for three replicates of modulated nonane (*m/z* 57) and p-xylene (*m/z* 91) peaks for P_{aux} ranging from 296.5 kPa to 341.3 kPa increasing in increments of 3.5 kPa. The optimal initial P_{aux} for all applications was 330.9 kPa and is indicated by the star. (C) Minimum convex hull drawn around the optimized test mixture separation (Fig. 2A), and $f_{coverage}$ metric. (D) Using the minimum convex hull method, $f_{coverage}$ is plotted for the pressure optimization study in (B), with the optimal P_{aux} (starred). (E) Overlay of single ^{2}D peak at p_w = 200 ms, 160 ms, 120 ms, and 40 ms for nonane using *m/z* 57. (F) Calculated average ^2W_b values for three replicates of modulated nonane (*m/z* 57) and p-xylene (*m/z* 91) peaks for P_M = 2 s and p_w ranging from 40 ms to 1320 ms. The star indicates the optimal p_w of 200 ms. Error bars for (B) and (F) represent one s.d. for three replicates.

Next, the optimal p_w was selected by tuning p_w and observing the modulation depth [47] (Fig. 5.3E) concurrent with measuring ${}^2W_b$ (Fig. 5.3F). The optimal p_w must provide full modulation (i.e., no breakthrough) and produce narrow ${}^2W_b$ for high peak capacity. A series of p_w ranging from 0 ms to 1960 ms was evaluated so that data could be compared across all modes of DPGM: the full modulation mode, and partial modulation in the negative and positive modes [35,47,49,50], wherein the positive pulse mode indicates baseline between modulations concurrent with p_w approaching P_M. A P_M of 2 s and initial P_{aux} of 330.9 kPa were applied for each p_w studied. Low p_w near 0 ms produce negative pulse mode data shown in Fig. 5.3E for p_w = 40 ms (green), 120 ms (black), and 160 ms (red). The lowest p_w studied that achieved full modulation was p_w = 200 ms, displayed as the blue trace in Fig. 5.3E. The effect of p_w on ${}^2W_b$ for nonane and p-xylene for p_w ranging from 40 ms to 1320 ms is presented in Fig. 5.3F. The lowest few p_w ranging from 40 ms to 160 ms provide negative pulse mode data with ${}^2W_b \approx 100$ ms, with increasing ${}^2W_b$ as p_w is increased. For p_w in the range of 200 ms to 1000 ms, full modulation is achieved. Since ${}^2W_b$ for a p_w of 200 ms is the narrowest in this range, this was selected as the optimal p_w indicated by a star. It is noteworthy that for ease of implementation, a large range of p_w provides full modulation (200 ms $\leq p_w \leq$ 680 ms) concurrent with ${}^2W_b$ not changing significantly. At a p_w = 1320 ms in Fig. 5.3D, the peaks just begin to be partially modulated in the positive pulse mode. The $p_w > 1320$ examined (up to 1960 ms) are also in the positive pulse mode, not shown for brevity.

To better visualize the effect of tuning p_w and support the selection of the optimal p_w of 200 ms, Fig. 5.4 highlights the raw data obtained for nonane at *m/z* 57 for a few select p_w ranging across all DPGM modes included in Fig. 5.3F. For ease of interpretation, all nonane peaks are plotted for 12 s of the separation and are on the same intensity scale. The unmodulated nonane

peak is shown in Fig. 5.4A corresponding to a p_w = 0 ms, i.e., when the valve is open throughout the entire P_M, and the auxiliary flow dilutes the ^{1}D effluent entering the ^{2}D column. This produces standard ^{1}D peak profiles at much lower signal than that of the modulated ^{2}D peaks. For short p_w approaching 0 ms, such as p_w = 40 ms in Fig. 5.4B, data in the negative pulse mode regime is produced [47–49]. The raw data requires a cumbersome baseline subtraction step to isolate the narrow ^{2}D peaks (less than 100 ms) prior to cutting and folding to produce a traditional 2D chromatogram. Increasing the p_w further, the data produced is fully modulated, beginning at p_w = 200 ms in Fig. 5.4C, to p_w = 600 ms in Fig. 5.4D, and with just a hint of going into the positive pulse mode at p_w = 1320 ms (p_w approaching the P_M) in Fig. 5.4E. Note that the heights of the fully modulated peaks for a p_w of 200 ms are slightly greater than that of the fully modulated peaks for p_w = 600 ms and p_w = 1320 ms, and the partially modulated peaks. A p_w of 200 ms produces narrower peaks relative to the fully modulated peaks with a p_w = 600 ms and 1320 ms, which accounts for the difference in peak heights since the sum of the ^{2}D peak areas is conserved. Similarly, the signal decreases at p_w < 200 ms as the analyte is not fully modulated and more signal is retained in the ^{1}D profile. This further supports the selection of p_w = 200 ms to be the optimal p_w. As p_w was increased to 1960 ms (Fig. 5.4F), almost the entire P_M of 2 s, the data is significantly more partially modulated in the positive pulse mode where the auxiliary gas flow creates vacancies in the peak profile [28,47,50]. It is also of note that as p_w was increased, the ^{1}D retention time of the nonane peak decreased. A longer p_w corresponds to less time that the auxiliary gas flow is applied, so the flow on ^{1}D is slowed down for a shorter period of time.

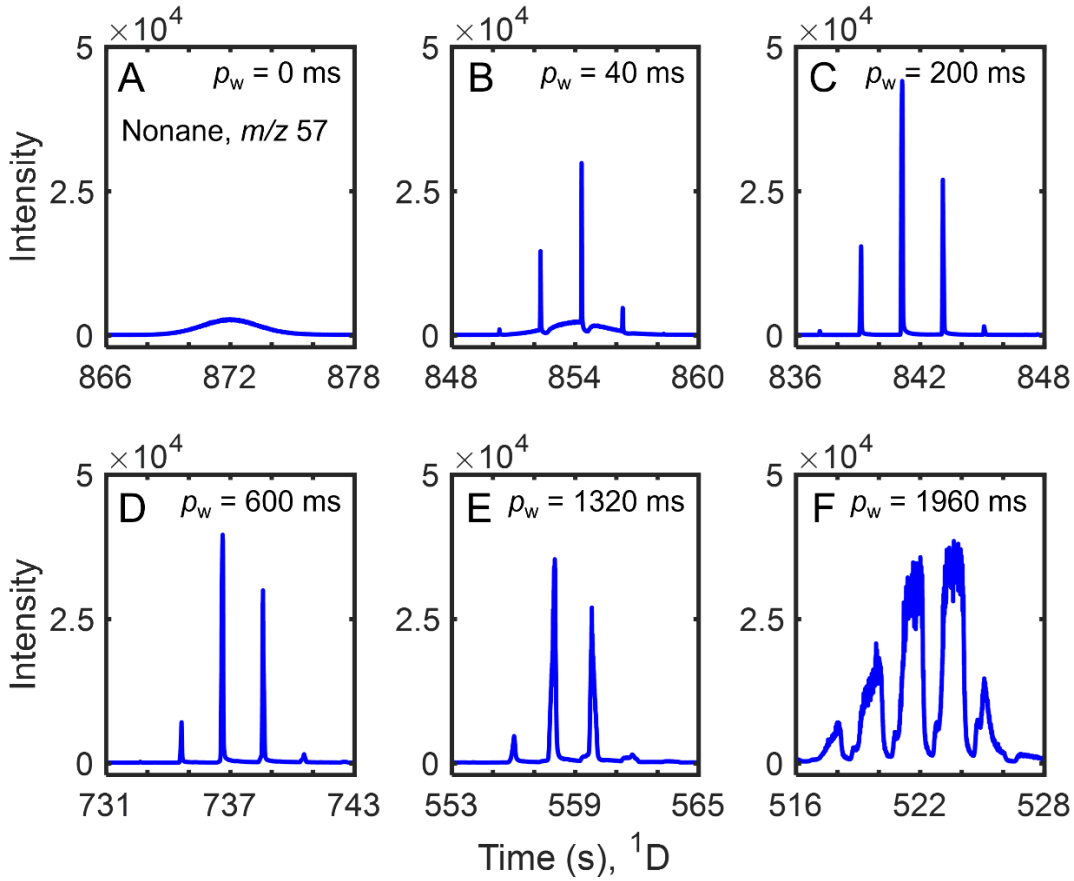


Figure 5.4. Raw data for nonane (*m/z* 57) from the test mixture separation. (A) Unmodulated ^{1}D peak with p_w = 0 ms, (B) Partial modulation of the ^{1}D peak in the negative pulse mode with p_w = 40 ms, (C) Full modulation mode of the ^{1}D peak with p_w = 200 ms. (D) Full modulation mode with p_w = 600 ms. (E) Partial modulation in the positive pulse mode with p_w = 1320 ms begins to emerge. (F) Partial modulation in the positive pulse mode for p_w = 1960 ms.

We now turn our focus on the evaluation of the 2D peak capacity performance of the optimized separation of the test mixture in Fig. 5.2. The same 10 analytes from Fig. 5.2C were used to determine the ^{1}D peak widths as defined in the context of Eqs. (5.1-5.3): experimentally measured ^1W_b and ^1W_b* at ±2 s.d., predicted ^1W_b** [37], and <β> corrected $^1W_{b,\beta}$ [38]. With analyte 1-bromoheptane (Fig. 5.2D) as the prior example, ^{1}D peak width measurements for all 10 analytes is provided in Fig. B.2 in Appendix B, with a summary in Table 5.1. The average ^{1}D peak widths for all 10 analytes are as follows: ^1W_b* = 6.0 (± 1.0) s, ^1W_b = 5.5 (± 1.0) s, ^1W_b** = 5.5 (± 1.0) s, and $^1W_{b,\beta}$ = 6.6 (± 0.9) s. Note that the agreement between ^1W_b and ^1W_b** serves to

experimentally validate the previous theoretical study [37]. Additionally, an average ${}^{1}W_{b,\beta} = 6.6$ s, indicates ~10% increase relative to ${}^{1}W_{b}$* due to only the statistical aspect of the SOT, since both contain the penalty for undersampling. Using Eqs. (5.1-5.3), with the average ^{1}D peak widths in Table 5.1, ${}^{2}n_{c} = 15$, and a $t_{sep} = 30$ min, the following 2D peak capacities were obtained: $n_{c,2D} = 4950$, $n_{c,2D}$* = 4500, and $n_{c,2D,\beta} = 4090$. For this test mixture, the overall 2D peak capacities are 22% less when considering the $f_{coverage} = 0.78$ (see Fig. 5.3C). Overall, the $n_{c,2D}$ provides about a 2-fold improvement relative to a previous GC×GC-TOFMS study [36], and corresponds to a high ideal peak production rate of 164 peaks/min ($n_{c,2D}/t_{sep}$). Factors contributing to this improvement include using a high ${}^{2}F$ (6 ml/min, versus 4 ml/min before) relative to the ${}^{1}F$, and installing a longer ^{2}D column (4 m, versus 2 m before). The higher ^{2}D flow rate produces narrower peaks, and the longer ^{2}D column increases ^{2}D retention time, resulting in a greater use of the 2D space and $f_{coverage}$, and therefore a smaller peak capacity correction.

Table 5.1. Summary for 1D peak width study using analytes labeled in Fig. 5.2B. The experimental ^1D peak widths, ^1W_b*, include modulation-induced band broadening, and were measured across ±2 s.d. of the Gaussian fit (Fig. 5.2D and Fig. B.2). The measured true ^1W_b were obtained from unmodulated data. The ^1W_b* were used to determine M_R*, M_R, and finally ^1W_b**, the predicted true ^1W_b using Table 5.2 in reference [37]. Finally, using ^1W_b as input, the $^1W_{b,\beta}$ were calculated, that takes into account band broadening due to modulation and the statistical aspect of SOT [38].

Analyte	^1W_b* (s)	^1W_b (s)	M_R*	M_R	^1W_b** (s)	$^1W_{b,\beta}$ (s)
1. Octane	6.5	5.9	3.25	3.02	6.0	6.9
2. Cyclooctane	5.9	5.5	2.96	2.76	5.5	6.6
3. 1-Chlorohexane	7.5	7.1	3.73	3.56	7.1	8.0
4. 1-Bromoheptane	5.5	5.0	2.75	2.48	5.0	6.1
5. P-xylene	7.5	7.1	3.75	3.56	7.1	8.0
6. Chlorohexylbenzene	5.2	4.6	2.60	2.30	4.6	5.9
7. 2-Undecanone	4.9	4.4	2.47	2.20	4.4	5.7
8. 2-ethyl-1-hexanol	5.5	5.0	2.77	2.48	5.0	6.2
9. 1-Octanol	5.0	4.4	2.52	2.20	4.4	5.7
10. 1-pentanol	6.5	6.0	3.24	3.02	6.0	7.0
Average (± 1 s.d.)	**6.0 (± 1.0)**	**5.5 (± 1.0)**	**3.00 (± 0.5)**	**2.75 (± 0.5)**	**5.5 (± 1.0)**	**6.6 (± 0.9)**

Finally, applicability of GC×GC-TOFMS with optimized DPGM in the full modulation mode was investigated for a broad range of complex samples: diesel, derivatized cow serum, coffee headspace, and river water. Inspired by other researcher's excellent results with reverse

column sets for fuel samples [51,52], we present the diesel fuel separation first. The TIC chromatogram for the entire 45 min run of this diesel is displayed in Fig. 5.5A. A section of the separation from 15 – 25 min is displayed in Fig. 5.5B to better visualize the W_b on both dimensions and the excellent use of 2D space. Indeed, the separation had an $f_{coverage}$ = 0.85 (Fig. B.3 in Appendix B). The relatively long ^{2}D column (4 m) allowed for the different compound classes to be separated well on ^{2}D and the high ^2F was essential to maintain narrow peaks for a high peak capacity that did not vary greatly in ^2W_b for analytes with varying ^2k'. Next, the separation of a derivatized cow serum sample was obtained using the same standard column set as the test mixture. The TIC of the separation of this cow serum metabolomic sample is provided in Fig. 5.5C with a zoom-in of 7.5 – 14.5 min (Fig. 5.5D) to highlight good use of 2D space, $f_{coverage}$ = 0.60 (Fig. B.3 in Appendix B). Next, SPME of coffee headspace was sampled, with the TIC of the separation provided in Fig. 5.6A, $f_{coverage}$ = 0.80 (Fig. B3 in Appendix B). Excellent use of 2D space is evident in the zoom-in from 15 – 25 min (Fig. 5.6B). The final complex sample analyzed using the optimized DPGM parameters was the SPME of the headspace of a locally sourced river water sample. An analytical ion current (AIC) chromatogram for a section of 10 – 40 min of the river water sample is displayed in Fig. 5.6C, $f_{coverage}$ = 0.60 (Fig. B.3 in Appendix B). As most peaks had low signal relative to a few peaks at the beginning of the separation, the chromatogram is shown starting at 10 min using selective *m/z* 41, 71, and 73 instead of the TIC. The TIC was not representative of all the analytes present in the sample, so these three *m/z* were selected as they captured most of the peaks in the chromatogram. To better visualize some of the detail in the separation, a zoom-in from 20 – 30 min is provided in Fig. 5.6D. A few components in the river headspace and coffee headspace were in relatively high concentration, resulting in non-Gaussian peaks as a result of non-linear isotherm conditions.

Overall, the $^{2}n_{c}$ for these four complex samples was just slightly less than that of the test mixture, ranging from 12 to 14, and the t_{sep} and ^{1}D peak widths were also essentially the same as that of the test mixture (results not shown for brevity). Likewise, while the test mixture had an $f_{coverage}$ = 0.78, the four complex samples ranged from 0.60 – 0.85. Therefore, the overall separation performance for the test mixture was found to translate over to the complex sample separations.

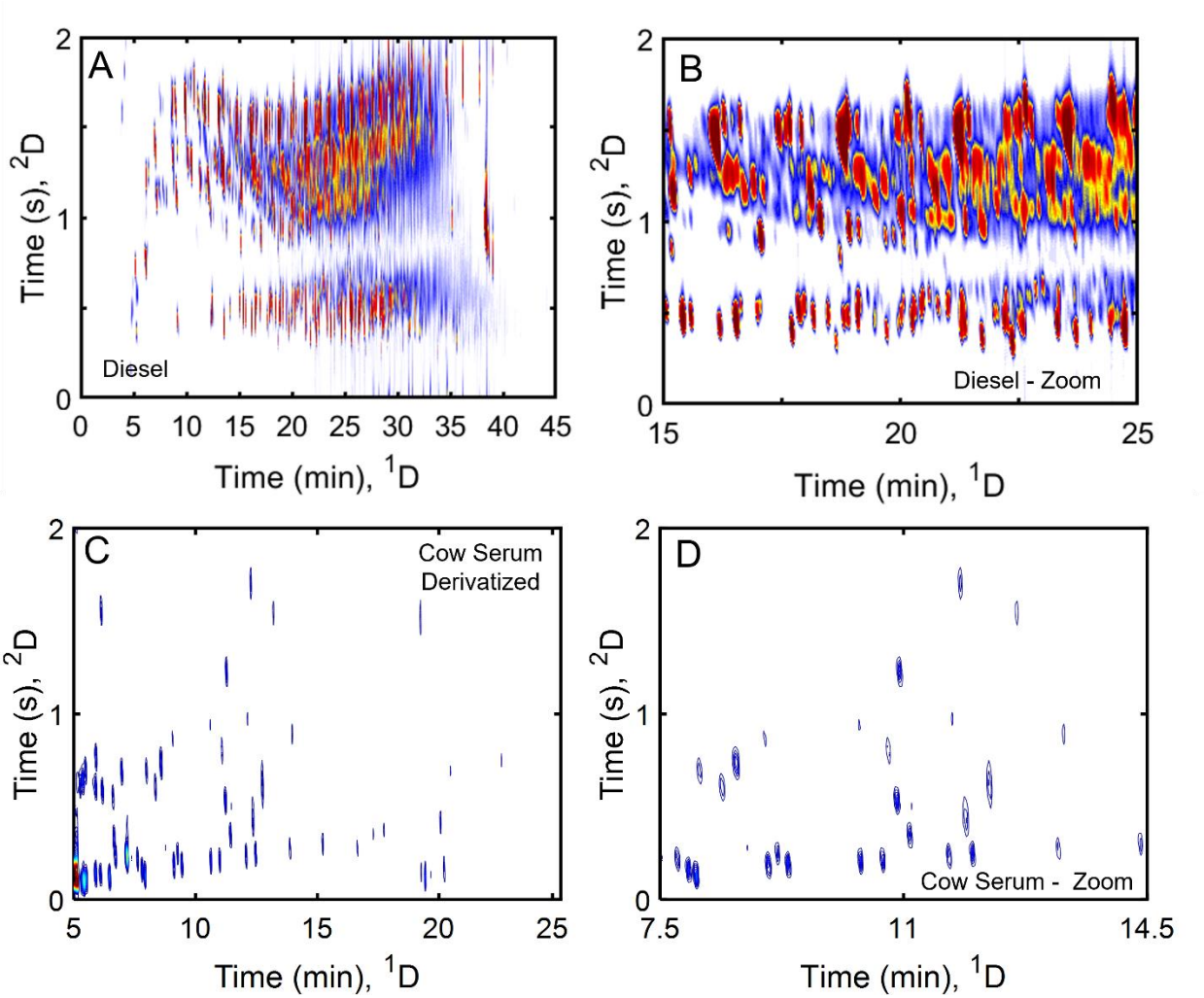


Figure 5.5. GC×GC chromatograms of diesel and derivatized cow serum. (A) The TIC chromatogram of diesel fuel using the reverse column set. (B) Diesel separation zoom-in from 15 – 25 min. (C) The TIC chromatogram of derivatized cow serum using the standard column set. (D) Zoom-in of the cow serum separation from 7.5 – 14.5 min.

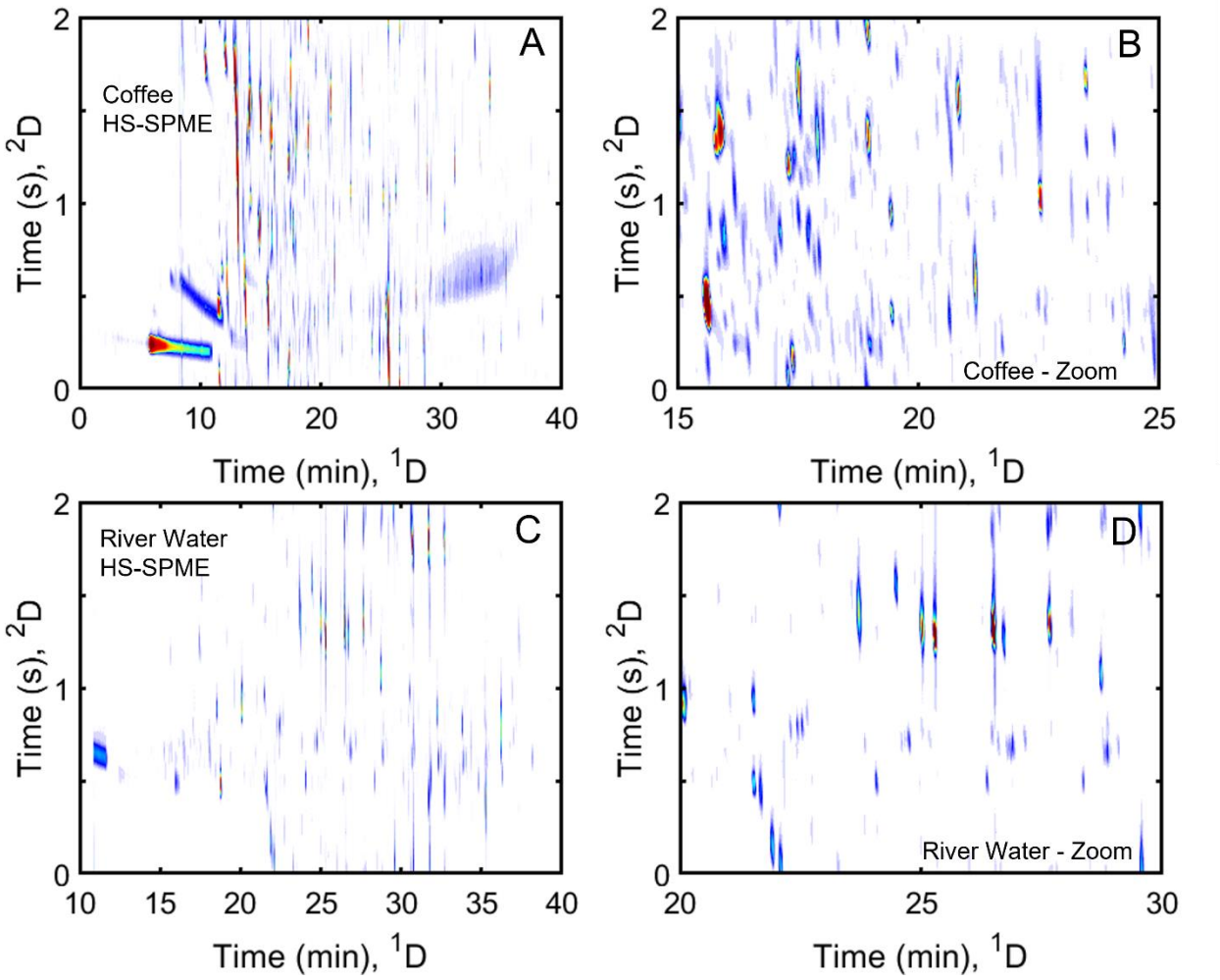


Figure 5.6. GC×GC chromatograms of coffee and river water headspace. (A) The TIC chromatogram of coffee headspace sampled by SPME. (B) Zoom-in of the coffee headspace separation from 15 – 25 min. (C) The analytical ion current (AIC) chromatogram of river water headspace, sampled by SPME, using *m/z* 41, 71, and 73. (D) Zoom-in from 20 – 30 min of the river water headspace separation.

5.5 Conclusion

The detailed assembly and experimental tuning for the optimization of DPGM for GC×GC-TOFMS was demonstrated. The optimized conditions were applied to a broad range of complex samples: diesel fuel, derivatized cow serum, coffee, and river water. For a user-defined P_M, ^{1}D flow conditions, and temperature program, optimal separations are achieved by tuning p_w and P_{aux} in concert. Both parameters affect the extent of ^{1}D peak modulation and detection

sensitivity. The p_w must be large enough to produce full modulation, but small enough that detection sensitivity is not compromised. The P_{aux} must be high enough to provide the stop-flow conditions necessary to provide full modulation concurrent with narrow ^{2}D peaks, but not too high as to significantly broaden ^{1}D peaks and reduce the 2D space used. Using optimized conditions, peak capacity was determined for all sample types. For the test mixture, 2D peak capacities of $n_{c,2D}$ = 4950, $n_{c,2D}$* = 4500, and $n_{c,2D,\beta}$ = 4090 were obtained. Thus, band broadening from the undersampling effect resulted in a 10% increase in peak width (^1W_b* relative to ^1W_b), and an additional 10% increase when corrected for the statistical aspect of $^1W_{b,\beta}$. The $f_{coverage}$ was calculated for all chromatograms using the minimum convex hull method, ranging from 0.60 – 0.85, and this should be considered for the net 2D peak capacity for a given sample. However, any $f_{coverage}$ correction to 2D peak capacity is highly sample dependent, since all samples, except for the diesel sample ($f_{coverage}$ = 0.85), were run with the same separation conditions, yet the $f_{coverage}$ ranged from 0.60 – 0.80. Expanding on findings of previous reports on DPGM for GC×GC-TOFMS, the demonstrated applicability to various complex samples with high peak capacity further broadens the scope and utility of this flow modulator.

5.6 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. doi:10.1093/chromsci/29.6.227.

[2] M.J.E. Golay, others, Theory of chromatography in open and coated tubular columns with round and rectangular cross-sections, Gas Chromatogr. 2 (1958) 36–55.

[3] 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.

[4] 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. doi:10.1016/j.trac.2018.05.007.

[5] 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. doi:10.1016/j.trac.2018.10.015.

[6] H.D. Bean, J.E. Hill, J.M.D. Dimandja, Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography-mass spectrometry data, J. Chromatogr. A 1394 (2015) 111–117. doi:10.1016/j.chroma.2015.03.001.

[7] F.A. Franchina, G. Purcaro, A. Burklund, M. Beccaria, J.E. Hill, Evaluation of different adsorbent materials for the untargeted and targeted bacterial VOC analysis using GC×GC-MS, Anal. Chim. Acta 1066 (2019) 146–153. doi:10.1016/j.aca.2019.03.027.

[8] P.Q. Tranchida, G. Purcaro, M. Maimone, L. Mondello, Impact of comprehensive two-dimensional gas chromatography with mass spectrometry on food analysis, J. Sep. Sci. 39 (2016) 149–161. doi:10.1002/jssc.201500379.

[9] 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. doi:10.1016/j.chroma.2017.05.064.

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

[11] K.A. Perrault, P.H. Stefanuto, L. Dubois, D. Cnuts, V. Rots, J.F. Focant, A new approach for the characterization of organic residues from stone tools using GC×GC-TOFMS, Separations 3 (2016) 8–10. doi:10.3390/separations3020016.

[12] J.B. Phillips, R.B. Gaines, J. Blomberg, F.W.M. van der Wielen, J.M. Dimandja, V. Green, J. Granger, D. Patterson, L. Racovalis, H.J. de Geus, J. de Boer, P. Haglund, J. Lipsky, V. Sinha, E.B. Ledford, A robust thermal modulator for comprehensive two-dimensional gas chromatography, HRC J. High Resolut. Chromatogr. 22 (1999) 3–10. https://doi.org/10.1002/(SICI)1521-4168(19990101)22:1%3C3::AID-JHRC3%3E3.0.CO;2-U.

[13] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive two-dimensional high-speed gas chromatography with chemometric analysis, Anal. Chem. 70 (1998) 2796–2804. doi:10.1021/ac980164m.

[14] J. V. Seeley, F. Kramp, C.J. Hicks, Comprehensive two-dimensional gas chromatography via differential flow modulation, Anal. Chem. 72 (2000) 4346–4352. doi:10.1021/ac000249z.

[15] R.W. LaClair, P.A. Bueno, J. V. Seeley, A systematic analysis of a flow-switching modulator for comprehensive two-dimensional gas chromatography, J. Sep. Sci. 27 (2004) 389–396. doi:10.1002/jssc.200301668.

[16] A. Ghosh, C.T. Bates, S.K. Seeley, J. V. Seeley, High speed Deans switch for low duty cycle comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1291 (2013) 146–154. doi:10.1016/j.chroma.2013.04.003.

[17] J. V. Seeley, N.J. Micyus, S. V. Bandurski, S.K. Seeley, J.D. McCurry, Microfluidic deans switch for comprehensive two-dimensional gas chromatography, Anal. Chem. 79 (2007) 1840–1847. doi:10.1021/ac061881g.

[18] P.J. Marriott, R.M. Kinghorn, Longitudinally modulated cryogenic system. A generally applicable approach to solute trapping and mobilization in gas chromatography, Anal. Chem. 69 (1997) 2582–2588. doi:10.1021/ac961310w.

[19] R.M. Kinghorn, P.J. Marriott, P.A. Dawes, Design and implementation of comprehensive gas chromatography with cryogenic modulation, HRC J. High Resolut. Chromatogr. 23 (2000) 245–252. doi:10.1002/(SICI)1521-4168(20000301)23:3<245::AID-JHRC245>3.0.CO;2-E.

[20] S.J. Kim, G. Serrano, K.D. Wise, K. Kurabayashi, E.T. Zellers, Evaluation of a microfabricated thermal modulator for comprehensive two-dimensional microscale gas chromatography, Anal. Chem. 83 (2011) 5556–5562. doi:10.1021/ac200336e.

[21] J. Luong, X. Guan, S. Xu, R. Gras, R.A. Shellie, Thermal independent modulator for comprehensive two-dimensional gas chromatography, Anal. Chem. 88 (2016) 8428–8432. doi:10.1021/acs.analchem.6b02525.

[22] S. Wohlfahrt, M. Fischer, J. Varga, M.R. Saraji-Bozorgzad, G. Matuschek, T. Denner, R. Zimmermann, Dual-stage consumable-free thermal modulator for the hyphenation of thermal analysis, gas chromatography, and mass spectrometry, Anal. Chem. 88 (2016) 640–644. doi:10.1021/acs.analchem.5b04183.

[23] V. Mucédola, L.C.S. Vieira, D. Pierone, A.L. Gobbi, R.J. Poppi, L.W. Hantao, Thermal desorption modulation for comprehensive two-dimensional gas chromatography using a simple and inexpensive segmented-loop fluidic interface, Talanta 164 (2017) 470–476. doi:10.1016/j.talanta.2016.12.005.

[24] P.A. Bueno, J. V. Seeley, Flow-switching device for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1027 (2004) 3–10. doi:10.1016/j.chroma.2003.10.033.

[25] J.F. Griffith, W.L. Winniford, K. Sun, R. Edam, J.C. Luong, A reversed-flow differential flow modulator for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1226 (2012) 116–123. doi:10.1016/j.chroma.2011.11.036.

[26] F.A. Franchina, M. Maimone, P.Q. Tranchida, L. Mondello, Flow modulation comprehensive two-dimensional gas chromatography-mass spectrometry using ≈ 4 mL min-1 gas flows, J. Chromatogr. A 1441 (2016) 134–139. doi:10.1016/j.chroma.2016.02.041.

[27] P.Q. Tranchida, F.A. Franchina, P. Dugo, L. Mondello, Flow-modulation low-pressure comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1372 (2014) 236–244. https://doi.org/10.1016/j.chroma.2014.10.097.

[28] H.D. Bahaghighat, C.E. Freye, D. V. Gough, R.E. Synovec, Comprehensive two-dimensional gas chromatography and time-of-flight mass spectrometry detection with a 50 ms modulation period, J. Chromatogr. A 1583 (2019) 117–123. doi:10.1016/j.chroma.2018.11.027.

[29] F.A. Franchina, M.E. Machado, P.Q. Tranchida, C.A. Zini, E.B. Caramão, L. Mondello, Determination of aromatic sulphur compounds in heavy gas oil by using (low-)flow modulated comprehensive two-dimensional gas chromatography-triple quadrupole mass spectrometry, J. Chromatogr. A 1387 (2015) 86–94. doi:10.1016/j.chroma.2015.01.082.

[30] J.F. Hamilton, A.C. Lewis, K.D. Bartle, Peak amplitude and resolution in comprehensive gas chromatography using valve modulation, J. Sep. Sci. 26 (2003) 578–584. doi:10.1002/jssc.200390079.

[31] P.Q. Tranchida, F.A. Franchina, M. Zoccali, I. Bonaccorsi, F. Cacciola, L. Mondello, A direct sensitivity comparison between flow-modulated comprehensive 2D and 1D GC in untargeted and targeted MS-based experiments, J. Sep. Sci. 36 (2013) 2746–2752. doi:10.1002/jssc.201300423.

[32] P.L. Urban, Clarifying Misconceptions about Mass and Concentration Sensitivity, J. Chem. Educ. 93 (2016) 984–987. doi:10.1021/acs.jchemed.5b00986.

[33] K.J. Jobst, J.V. Seeley, E.J. Reiner, L. Mullin, A. Ladak, Enhancing the sensitivity of atmospheric pressure ionization mass spectrometry using flow modulated gas chromatography, LCGC North Am. 14 (2018) 23–28.

[34] J. V. Seeley, N.E. Schimmel, S.K. Seeley, The multi-mode modulator: A versatile fluidic device for two-dimensional gas chromatography, J. Chromatogr. A 1536 (2018) 6–15. doi:10.1016/j.chroma.2017.06.030.

[35] T.J. Trinklein, D. V. Gough, C.G. Warren, G.S. Ochoa, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1609 (2019) . doi:10.1016/j.chroma.2019.460488.

[36] S. Schöneich, D. V Gough, T.J. Trinklein, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection, J. Chromatogr. A 1620 (2020). doi:10.1016/j.chroma.2020.460982.

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

[38] J.M. Davis, D.R. Stoll, P.W. Carr, Dependence of effective peak capacity in comprehensive two-dimensional separations on the distribution of peak capacity between the two dimensions, Anal. Chem. 80 (2008) 8122–8134. doi:10.1021/ac800933z.

[39] S.C. Rutan, J.M. Davis, P.W. Carr, Fractional coverage metrics based on ecological home range for calculation of the effective peak capacity in comprehensive two-dimensional separations, J. Chromatogr. A 1255 (2012) 267–276. doi:10.1016/j.chroma.2011.12.061.

[40] J.C. Giddings, Two-dimensional separations: concept and promise., Anal. Chem. 56 (1984). doi:10.1021/ac00276a003.

[41] R.B. Wilson, J.C. Hoggard, R.E. Synovec, High throughput analysis of atmospheric volatile organic compounds by thermal injection - Isothermal gas chromatography - Time-of-flight mass spectrometry, Talanta 103 (2013) 95–102. doi:10.1016/j.talanta.2012.10.013.

[42] 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.

[43] 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.

[44] B.A. Parsons, D.K. Pinkerton, R.E. Synovec, Implications of phase ratio for maximizing peak capacity in comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A 1536 (2018) 16–26. doi:10.1016/j.chroma.2017.07.018.

[45] B. Omais, M. Courtiade, N. Charon, S. Esnault, D. Thiébaut, Reversal of elution order in a single second dimension by changing the first column nature in comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1255 (2012) 196–201. doi:10.1016/j.chroma.2012.07.042.

[46] B. Omais, M. Courtiade, N. Charon, J. Ponthus, D. Thiébaut, Considerations on orthogonality duality in comprehensive two-dimensional gas chromatography., Anal. Chem. 83 (2011) 7550–7554. doi:10.1021/ac201103e.

[47] H. Cai, S.D. Stearns, Partial modulation method via pulsed flow modulator for comprehensive two-dimensional gas chromatography, Anal. Chem. 76 (2004) 6064–6076. doi:10.1021/ac0492463.

[48] D.V. Gough, S. Schöneich, R.E. Synovec, Chemometric decomposition of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data employing partial modulation in the negative pulse mode, Talanta 210 (2020). doi:10.1016/j.talanta.2019.120670.

[49] D. V. Gough, D.H. Song, S. Schöneich, S.E. Prebihalo, R.E. Synovec, Development of ultrafast separations using negative pulse partial modulation to enable new directions in gas chromatography, Anal. Chem. 91 (2019) 7328–7335. doi:10.1021/acs.analchem.9b01085.

[50] H.D. Bahaghighat, C.E. Freye, D. V. Gough, P.E. Sudol, R.E. Synovec, Ultrafast separations via pulse flow valve modulation to enable high peak capacity multidimensional gas chromatography, J. Chromatogr. A 1573 (2018). doi:10.1016/j.chroma.2018.08.001.

[51] M. Jennerwein, M. Eschner, T. Wilharm, T. Gröger, R. Zimmermann, Evaluation of reversed phase versus normal phase column combination for the quantitative analysis of common commercial available middle distillates using GC × GC-TOFMS and Visual Basic Script, Fuel 235 (2019) 336–338. doi:10.1016/j.fuel.2018.07.081.

[52] T.C. Tran, G.A. Logan, E. Grosjean, J. Harynuk, D. Ryan, P. Marriott, Comparison of column phase configurations for comprehensive two dimensional gas chromatographic analysis of crude oil and bitumen, Org. Geochem. 37 (2006) 1190–1194. doi:10.1016/j.orggeochem.2006.05.006.

Chapter 6: Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography / mass spectrometry data: Study of the pacu fish metabolome

This chapter was reproduced from S. Schöneich, G.S. Ochoa, C.M. Monzón, R.E. Synovec, Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography / mass spectrometry data: Study of the pacu fish metabolome, J. Chromatogr. A. 1667 (2022) 462868.

6.1 Introduction

In the last decade, pacu fish have become one of the most produced fish in Argentina as an important source of revenue and sustenance [1]. These fish are often raised in an integrated system with rice fields so that insects and weeds can be controlled, and the fields can continue to produce throughout the year [2]. While this agricultural approach may be economically beneficial, it raises concerns that herbicides and/or pesticides used for maintaining the rice fields may alter the fish metabolome prior to human consumption [3]. To study the impact on the pacu fish metabolome, comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) is an excellent instrument platform, as it has been used for the separation of complex samples in many applications [4–9]. Introduced by Liu and Philips [10], GC×GC achieves roughly a 10-fold increase in peak capacity relative to 1D-GC in addition to increased sensitivity [11,12], ideal for metabolomic applications and discovering trace analytes. Due to the complexity of the samples and separations produced, chemometric tools are necessary to interpret the data efficiently.

Non-targeted methods of feature discovery, where the analytes of interest are not known *a priori*, that can handle complex samples include principal component analysis (PCA) [13–16] and Fisher ratio (F-ratio) analysis [17–21], which will be the focus herein. Other methods include

multivariate analysis of variance (MANOVA) [22,23] that assesses multiple dependent variables to determine if sample class means are equal, partial least squares – discriminant analysis (PLS-DA) [24–26] for predictive modeling using a determined set of latent variables, Wilcoxon rank sum test (Mann-Whitney U test) [27–29] that tests if independent samples derive from the same population, and the Kruskal-Wallis test [29–31], which is an extension of the Wilcoxon rank sum test to more than two sample classes. Most of these methods were not implemented herein as the data set collected was not be suitable for these tests. PCA is an unsupervised method since sample class membership is not known prior to analysis, while F-ratio analysis and the Wilcoxon rank sum test are supervised as the sample class membership is assigned based upon the experimental design. In the current study, GC×GC-TOFMS data of pacu fish raised in an integrated rice-fish farming system (farmed class) versus fish raised in tanks (tank-raised class) were collected, with a preliminary study of this pacu fish data set achieved using PCA [32]. Although the PCA loadings revealed several class distinguishing analytes and provided some insight into class differences, this prior work was not a comprehensive study of the samples as only the total ion current (TIC) and *m/z* 217 were used for PCA. Only 34/50 analytes discovered had significant differences in concentrations (p-value < 0.05) and derivatization artifact peaks made it difficult to discover analyte peaks with lower signal. To more comprehensively discover class-distinguishing metabolites between pacu fish raised in tanks and pacu fish farmed in the integrated rice-fish farming system, for the current study tile-based F-ratio analysis will be implemented [19–21]. The standard F-ratio (^{S}F-ratio) discovers hits based on the ratio of the between-class variance, s^2_{bc}, to the sum of the pooled within-class variance, $\sum s^2_{wc}$,

$$^{S}\text{F-ratio} = \frac{s^2_{bc}}{\sum s^2_{wc}} \quad (6.1)$$

Recently, the control-normalized F-ratio (^{CN}F-ratio) was introduced, that only uses the within-class variance of the control class, $s^2_{wc,\ controls}$, to normalize to between-class variance, s^2_{bc}, since the control class may have a smaller within-class variance than the patient class [33,34],

$$^{CN}\text{F-ratio} = \frac{s^2_{bc}}{s^2_{wc,controls}} \tag{6.2}$$

The goal was to discover significant analytes with non-normal class distributions in the patient class, allowing hits with large between-class variance yet high within-class variance in the patient class due to extreme biological variation to be discovered higher on the hit list than if ^{S}F-ratio analysis were used. This approach does not rely on correcting the non-normality of entire GC×GC-TOFMS data sets prior to analysis as it normalizes to the class with a normal distribution during F-ratio analysis. ^{CN}F-ratio analysis works well for a well-behaving control class sample set with a normal distribution; however, in the case that discoverable analytes in both classes may have non-normal distributions, the ^{CN}F-ratio calculation would likely miss some significant analytes. As was observed in the preliminary study of the pacu fish data sets by PCA, the %RSD was not consistently higher in one class for all analytes [32]. This suggested that neither the ^{S}F-ratio or ^{CN}F-ratio calculations would be ideal for the pacu fish data set and the F-ratio calculation needed to be optimized to optimally discover the most true positives. To address this challenge, we propose the minimum variance optimized (MVO) F-ratio (^{MVO}F-ratio) where the within-class variance is calculated for each hit for both classes and the smallest of the two is used as the denominator in the F-ratio calculation,

$$^{MVO}\text{F-ratio} = \frac{s^2_{bc}}{s^2_{wc,minimum}} \tag{6.3}$$

The ^{MVO}F-ratio should discover more true positives than either the ^{S}F-ratio or the ^{CN}F-ratio, as this concept will be explored in this study.

Metabolomic samples are not inherently amenable to GC×GC-TOFMS analyses for detecting metabolites due to their low volatility, so samples are typically derivatized prior to injection. One common method of metabolite derivatization is a two-step process of methoximation and trimethylsilylation, replacing amine and hydroxyl hydrogen atoms with a trimethylsilyl group, increasing analyte volatility [35,36]. This derivatization step produces artifact peaks during the GC×GC-TOFMS data collection that are subsequently discovered by F-ratio analysis, that may dominate hit lists and obscure important analytes. These spurious artifact entries make hit list mining to obtain a subset of meaningful features very tedious work. Such artifact peaks include trimethylsilanol, pyridine, glycerol, and phosphoric acid, which may result from the decomposition of several other species in the sample. This multi-origin phenomenon was demonstrated for derivatized metabolomic samples analyzed by GC-MS with observed trimethylsilylated glycerol and phosphoric acid peaks deriving from multiple phosphocholines and lysophosphatidylcholines [37], although the mechanisms are not fully understood and can add some sample-to-sample variance. Other streaks that are commonly present in GC×GC-TOFMS analyses include solvent peaks and column bleed, all of which potentially interfere with discovering analytes of interest, making accurate identification and quantification more challenging.

Current methods for tackling this issue of artifact removal include mass spectral filtering or a streak detection [8,38]. Especially useful for targeted approaches, filtering based on mass spectra and retention times can be used to discover compounds of certain classes such as halogenated contaminants in fish [8,39]. For nontargeted studies, a streak detection approach used a steerable Gaussian filtering and a marker-controlled watershed algorithm to detect and remove streaks in GC×GC-FID chromatograms to aide in accurate analyte quantification [38],

with the performance of this method relying on the streak tilt degree. Other streak detection approaches have been reported for other imaging applications, however none were adapted for chromatographic data [40,41]. While these are excellent possibilities, a method is necessary for non-targeted studies with GC×GC-TOFMS that can handle the full range of ^{2}D retention time, and that do not rely on streak shape or orientation while discovering hits that may be overlapped with streaks in the chromatogram. Herein, to address this challenge, we propose a derivatization artifact removal step for tile-based F-ratio analysis based upon mass spectral similarity between all of the hits on the hit list and a set of reference mass spectra from the major derivatization artifacts observed. Hits that spectrally match the derivatization artifact library will be automatically removed from the final hit list.

For development of the artifact removal algorithm, both ^{S}F-ratio and ^{CN}F-ratio are used, via Eqs. (6.1) and (6.2), on a pooled farmed pacu fish sample spiked with 29 metabolites and a pooled farmed pacu fish sample diluted in pyridine (control). Next, the 10 farmed pacu and 10 tank-raised pacu fish called for the implementation of ^{MVO}F-ratio analysis via Eq. (6.3), evaluating its performance relative to the standard F-ratio via Eq. (6.1), and the tank-normalized and farm-normalized F-ratio, both calculated via Eq. (6.2). Discovery of meaningful hits was aided by the artifact removal algorithm and were finally determined to be true positives by t-test at the 95% confidence level. Concentration ratios for each hit were calculated using the most pure *m/z* determined by p-value and percent lack-of-fit (*LOF*) statistical metrics [21]. These results were used to interpret differences in metabolomic profiles of the fish raised in the rice-fish farming method relative to fish raised in a controlled environment.

6.2 Experimental

Tank-raised and farmed pacu fish samples (10 fish per class) were obtained from Argentina as freeze-dried fish flakes divided into 5 g portions in 2-3 vials per fish. To minimize

heterogeneity across vials for an individual fish, four 20 mg extracts per vial were pooled when samples for each fish were prepared prior to derivatization (Fig. C.1 in Appendix C). For both studies, samples were prepared and analyzed by GC×GC-TOFMS on the same day, with the sample preparation steps previously described [32,36]. For the first study, the 10 pooled samples (1 per fish) of the farmed fish class were pooled together to make a single pooled farmed fish sample. Next, a mixture of 29 metabolites was spiked into the pooled sample of all farmed fish at a concentration of 50 ppm prior to derivatization. Although only tested on a 50-ppm spike mixture, the signals ranged from roughly 1×10^4 to 9×10^6 with observable differences in signal-to-noise (*S/N*) (Fig. C.2 in Appendix C), demonstrating that F-ratio analysis and the artifact removal algorithm work well for a range of *S/N*. The unspiked pooled farmed fish sample was diluted in pyridine, since pyridine was already used in the sample preparation, to ensure native analytes in both classes would be diluted by the same amount. Six replicates each of the spiked and unspiked pooled farmed samples were collected for development of the artifact removal algorithm. Four replicates of the neat standard mixture (Table C.1 in Appendix C) were also collected under the same instrumental parameters to create an in-house library and to use for retention time matching with spiked fish samples. For the second study aimed at the study of the changes in the pacu fish metabolome, which included evaluation of the performance of ^{MVO}F-ratio analysis, pooled samples for each of the ten tank-raised fish and ten farmed fish (20 total samples) described previously were used. Two replicates were collected per farmed and tank-raised fish for a total of 20 chromatograms per class.

The LECO Pegasus BT 4D GC×GC-TOFMS instrument (LECO Corporation, St. Joseph, MI) was used for data collection using ultra-high purity helium (Grade 5, 99.999%, Praxair, Seattle, WA, USA) carrier gas and equipped with a thermal modulator. The ^{1}D column was

coated with a polar Rxi-17 stationary phase (30 m length, 250 μm $^{1}d_{c}$, 0.25 μm $^{1}d_{f}$) and the ^{2}D column was coated with a nonpolar Rxi-1ms stationary phase (2 m length, 180 μm $^{1}d_{c}$, 0.18 μm $^{1}d_{f}$). Although a reverse column set was used based on past metabolomic studies [34], future work using a normal column set may be beneficial for identification using retention time indices. The inlet and transfer line temperatures were both set to 250 °C. The oven temperature was programmed to be held at 50 °C for 0.25 min, then ramped to 300 °C at 8 °C/min with a secondary oven temperature offset of 5 °C and a modulator temperature offset of 15 °C. A 1.5 ml/min flow was selected with a modulation period, P_{M}, of 2 s. Mass spectra (*m/z* 40-500) were collected with an electron ionization energy of 70 eV at a rate of 100 Hz. A 3.5 min acquisition delay was selected to avoid solvent peaks and maintain the instrument condition.

All fish data from the LECO ChromaTOF software (v5.20) (LECO, St. Joseph, MI) were imported to Matlab v 2019a (MathWorks, Inc., Natick, MA) for analysis. Our in-house tile-based F-ratio software [19–21] was used for standard F-ratio analysis per Eq. (6.1) prior to modification to implement the automated control-normalized (i.e., background-normalized) [34], farm-normalized both per Eq. (6.2), and ^{MVO}F-ratio analyses per Eq. (6.3). We note that the tile-based ^{S}F-ratio analysis is functionally equivalent to the software now commercially available as ChromaTOF Tile, whereas the F-ratio variations examined herein are not presently part of this commercial software. Additionally, the artifact removal step was written to be mostly automated and limiting the manual input to the reference spectra of the artifact peaks, the match value (MV) threshold, and the number of *m/z* to use for the MV calculation. For the first study involving the spiked fish, a ^{1}D tile size of 12 s (6 modulations) and ^{2}D tile size of 500 ms were used based upon the average peak width of the spiked analytes, and the cluster window size selected was 10 s on ^{1}D and 400 ms on ^{2}D. Due to smaller peak widths of the native analytes in the tank-raised

and farmed samples and in an attempt to discover the most true positives, a smaller tile size of 8 s (4 modulations) on ^{1}D was used with 400 ms on the ^{2}D with a cluster window size of 8 s on ^{1}D and 300 ms on ^{2}D. Finally, the reported F-ratios are an average of the top 5 *m/z* with a minimum of 1 *m/z* per hit that pass a *S/N* threshold of 10.

Hits discovered in the second study examining the changes in fish metabolome were identified by a MV [42] using the NIST library (2011) with an MV ≥ 800 required for tentative identification. The forward MV was selected over reverse MV or probability values to avoid any bias due to ignoring potentially relevant *m/z* in the hit spectra. To aid in analyte identification, PARAFAC was used to deconvolute hit mass spectra from noise and/or interferents prior to mass spectral matching [43,44]. By visual inspection of hits with mass spectra with clear fragmentation patterns corresponding to derivatized metabolites that could otherwise not be tentatively identified due to MV < 800, these were listed as "unknown metabolites." Concentration ratios were calculated as [Farmed]/[Tank-raised], or simply [F]/[T], for each hit in the hit list by the "two step" method [45]. An in-house algorithm sums together the maximum signal of each modulated ^{2}D peak per analyte hit to obtain signal measurements of all samples at all *m/z* that pass the *S/N* threshold. Using the *LOF* metric and a t-test assuming variances are unequal, the most pure *m/z* was determined with the lowest *LOF* (minimum *LOF* < 20%) and the lowest p-value (minimum p-value < 0.05) for accurate concentration ratio calculations [21].

6.3. Results and Discussion

6.3.1 Artifact removal development using spiked versus unspiked fish samples

Algorithm development to remove artifact peaks due to metabolite derivatization was performed on a spiked sample of the pooled farmed pacu fish and an unspiked farmed pacu sample (serving as the control class). The GC×GC chromatogram using *m/z* 73 for the unspiked sample is provided in Fig. 6.1A and that of the sample spiked with 29 metabolites is provided in

Fig. 6.1B. The *m/z* 73 was plotted instead of the TIC to reduce the presence of the artifact peaks with high signal in the TIC and due to its universal selectivity for derivatized metabolites that contain the trimethylsilyl groups. A zoom-in of both the unspiked (Fig. 6.1C) and spiked (Fig. 6.1D) from 5 to 11 min displays the presence of spiked metabolite peaks in Fig. 6.1D that are not present in Fig. 6.1C as well as other peaks that are native to the sample. All spiked compounds are provided with retention times and signal in the standard mixture in the Table C.1 in Appendix C; there is roughly a 1×10^4 to 9×10^6 range in detected peak height signal at *m/z* 73 for the metabolites, all spiked at 50 ppm. The spiked metabolites discovered by F-ratio analysis in the 5 to 11 min region are circled with confirmed identification by retention time matching to the standard mixture. It is clear in Fig. 6.1D that some of the circles do not show definitive peaks as a consequence of natural variation in detection sensitivity for the metabolites and the color scale applied. Selective *m/z* could be used for visualizing these specific metabolites. We note that there is a faint streak throughout the entire chromatogram with ${}^2t_R$ between 0.5 to 0.8 s highlighted in the red region in Fig. 6.1A identified as trimethylsilanol (A1), an artifact of the derivatization process. Two other reoccurring artifacts from the sample preparation process that are not as visibly evident in Figs. 6.1A and 6.1B were identified as pyridine (A2) as a streak with ${}^2t_R$ between 0.80 s and 1.05 s (green region in Fig. 6.1A) and phosphoric acid fragment (A3) with varying ${}^2t_R$ as shown in the blue region in Fig. 6.1A. These interfering artifacts pose a challenge to tile-based F-ratio analysis as they may be discovered easily and fill the hit list with false positives.

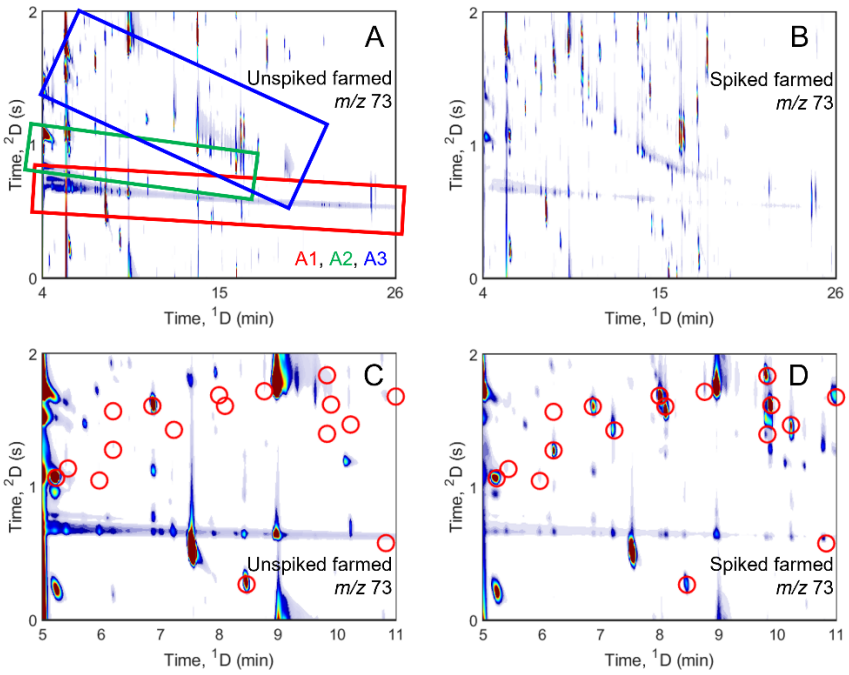


Figure 6.1. Analytical ion current (AIC) contour plots of the unspiked (A) and spiked (B) pooled farmed samples using *m/z* 73, selective for trimethylsililated groups for the derivatized spiked analytes. The retention time windows that contain the three artifacts are indicated in (A): trimethylsilanol (A1), (B) spectrum for pyridine (A2), and (C) spectrum for phosphoric acid fragment (A3). A zoom-in of (A) from 5 to 11 min is displayed in (C) and a zoom-in of the same time window for (B) is shown in (D) to better view differences between classes. All spiked analytes discovered by ^{CN}F-ratio analysis in this time window are circled in red.

To address the issue of artifact peak removal, we demonstrate the principles with the ^{CN}F-ratio analysis implementing Eq. (6.2), with the ^{S}F-ratio analysis performed the same way using Eq. (6.1). Our approach involves first obtaining a hit list by tile-based F-ratio analysis and then removing the artifact peak hits by MV similarity to reference spectra of the same artifacts. This approach ensures comprehensive analyte discovery, including any that may be overlapped with an artifact peak which could otherwise be removed if the section of the chromatogram with the artifact streak was excluded, and does not rely on the retention times of the artifact peaks which may significantly change. To apply this artifact removal approach, a reference mass spectrum was obtained for trimethylsilanol (Fig. 6.2A), pyridine (Fig. 6.2B), and phosphoric acid fragment (Fig. 6.2C). While it is possible to use the algorithm for as many artifact peaks as the user

defines, these three were selected for this data set as it was apparent they were obscuring the discovery of true positives, defined as the spiked analytes, and false positives being any peak that is not a spiked analyte in this first study. In addition, this approach may be used to remove column bleed peaks in a similar manner, however this was not demonstrated herein as column bleed was not a problem for this data set. To select the appropriate reference spectrum for each artifact peak, PCA was performed on 15 mass spectra for each A1 (Fig. 6.2D), A2 (Fig. 6.2E), and A3 (Fig. 6.2F). These 15 mass spectra used to select the reference spectrum for each artifact were obtained from various $^{1}t_{R}$ in the regions highlighted in the chromatograms in Fig. 6.1A. Once the scores were obtained, the centermost spectrum from each cluster was selected as the most representative spectrum using the Mahalanobis distance [46]. Next, the artifact removal algorithm calculated the MV between the mass spectrum at the 2D retention time coordinates of each hit in the hit list and each of the artifact reference spectra. A visual representation of the MVs calculated between each of the hit mass spectra in the ^{CN}F-ratio analysis hit list and A1, A2, and A3 mass spectra are provided in Fig. 6.3A, 6.3B, and 6.3C, respectively. To obtain a MV, a weighted cosine similarity calculation [187,188] with weight factor (1, 1.3) for the *m/z* intensity and *m/z* value, respectively, was performed using a subset of the 10 most intense *m/z* to avoid including noisy *m/z*. The 10 most intense *m/z* of each hit in the hit list were used for the MV calculation instead of the reference spectra to avoid using low *m/z* that would artificially inflate the MV. To remove these artifacts from the hit list, a threshold of MV $\geq$ 800 was applied as this is generally accepted to tentatively identify compounds, represented by the dotted red lines in Fig. 6.3. All hits with a MV $\geq$ 800 were subsequently removed from the original hit list, ideally reducing the hit list to only true positives and additional artifacts that are not included in the three major artifacts. To confirm that artifacts were removed correctly, indexed mass purity plots

[189] were calculated for hits that were removed by the algorithm (Fig. C.3 in Appendix C). The artifact removal algorithm eliminated a total of 172 false positives from the ^{CN}F-ratio analysis hit list of which 104 were of A1, 29 were of A2, and 39 were of A3. Next, 25 redundant hits were removed for a total of 197 false positives removed, nominally from the bottom of the hit list, leaving a final hit list of 49 hits for the ^{CN}F-ratio analysis. Similarly, the ^{S}F-ratio analysis produced a final hit list with 56 hits.

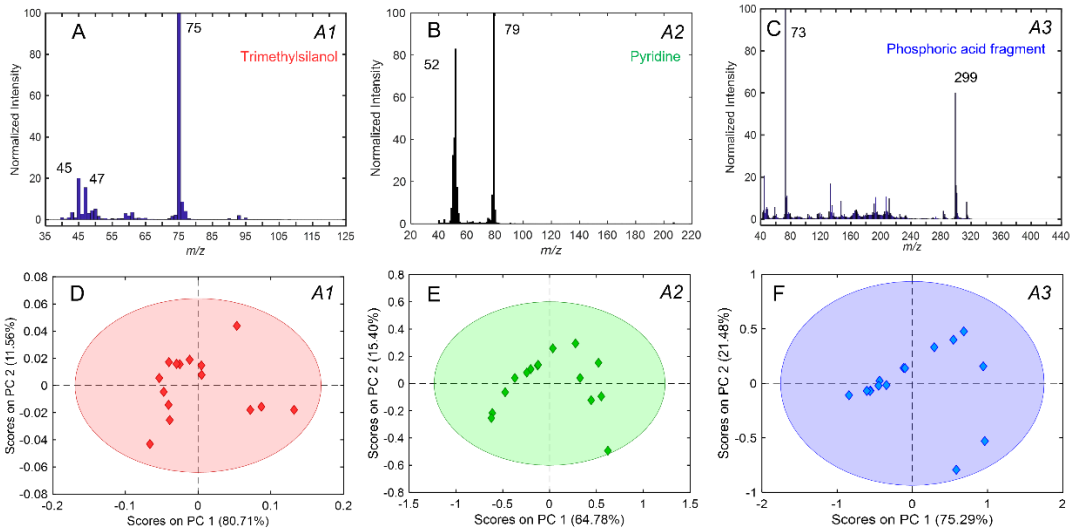


Figure 6.2. Selection of reference spectra to use for artifact removal. (A) The mass spectrum for the first artifact peak trimethylsilanol (A1), (B) spectrum for pyridine (A2), and (C) spectrum for phosphoric acid fragment (A3). A few of the most intense *m/z* are labeled for each spectrum. PCA scores plot used to select the most representative spectrum from each cluster for each artifact peak A1 (D), A2 (E), and A3 (F) to use in the artifact removal algorithm. For each artifact, 15 spectra were used to generate each PCA scores plot.

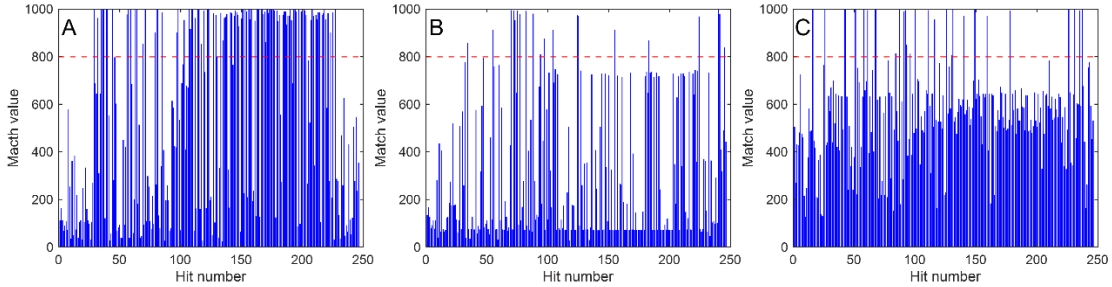


Figure 6.3. Illustration of hits removed by artifact removal. Bar plot displaying the MV between the mass spectrum of each hit in the ^{CN}F-ratio original hit list and the spectrum of A1 (A), A2 (B), A3 (C). The threshold selected for removing artifact peaks from the hit list is a MV ≥ 800, represented by the red dotted lines.

The results of the development of the artifact removal algorithm are summarized in Table 6.1 for the spiked and unspiked farmed pacu fish samples. All hits are ranked by their ^{CN}F-ratio with ascending hit number in the ^{CN}F-ratio hit list. In the ^{CN}F-ratio hit list 26/29 spiked metabolites were discovered, noting that there still were 23 false positives in the 49 total hits. In contrast, for the ^{S}F-ratio analysis, 23/29 of the spiked metabolites were discovered with 33 false positives in the 56 total hits, indicating the ^{CN}F-ratio analysis was more effective in discovering true positives. Twenty-two hits were discovered by both the ^{S}F-ratio and ^{CN}F-ratio analyses, while aspartic acid was only discovered by ^{S}F-ratio analysis, and six additional metabolites were discovered only by ^{CN}F-ratio analysis. During artifact removal, glycine was removed due to a high MV (927) calculated with the A1 mass spectrum. Fortunately, most of the MVs calculated for each of the hits with the three artifact reference spectra were very low, as reported in the three right-most columns of Table 6.1. Additionally, acetic acid and creatinine were not originally identified, however spectra from PARAFAC deconvolution aided in their identification (Fig. C.4 in Appendix C).

Table 6.1. Hit list generated using the ^{C}NF-ratio and ^{S}F-ratio calculations, via Eqs. (6.2) and (6.1) respectively. The analytes are listed according to rank in the ^{CN}F-ratio hit list. Along with the identity of the hits, hit number in the final hit list is provided for both ^{CN}F-ratio and ^{S}F-ratio. Finally, MVs are listed for each hit with respect to the three artifacts used in the artifact removal step.

Compound	^{CN}Hit #	^{CN}F-ratio	^{S}Hit #	^{S}F-ratio	A1MV	A2MV	A3MV
gluconic acid	1	2.72×10^{10}	9	593	113	134	506
fumaric acid	2	4.18×10^{9}	1	1059	164	166	270
D-galactose	3	4.05×10^{9}	12	574	113	127	432
glycerol-3-phosphate	4	3.60×10^{9}	11	582	67	78	429
malonic acid	5	3.14×10^{9}	4	716	90	91	481
succinic acid	6	2.01×10^{9}	--	--	108	113	724
methylmalonic acid	7	7.56×10^{8}	6	636	72	77	476
4-methylvaleric acid	8	2.20×10^{8}	--	--	577	112	462
L-glutamic acid	9	1.54×10^{8}	3	730	254	280	216
pyroglutamic acid	10	1.47×10^{8}	13	564	35	38	126
uracil	11	5.10×10^{7}	5	684	362	436	248
myo-inositol	12	3.98×10^{7}	14	560	51	66	376
D-glucose-6-phosphate	13	3.89×10^{7}	15	533	382	407	582
maleic acid	14	3.23×10^{7}	10	590	74	75	487
D-glucose	15	1.02×10^{7}	16	507	112	126	418
pyruvate	17	3.34×10^{6}	18	424	105	175	367
oxalic acid	18	2.28×10^{6}	8	612	98	117	764
maltose	19	9.66×10^{5}	20	416	104	111	427
acetic acid	21	1.01×10^{5}	--	--	94	98	240
creatinine	22	8.01×10^{4}	--	--	39	55	221
L-cysteine	23	7.88×10^{4}	23	379	93	117	693
isovaleric acid	24	6.57×10^{4}	33	129	684	764	697
L-valine	25	6.35×10^{4}	--	450	199	585	394
lactic acid	28	3.28×10^{4}	--	--	76	94	635
L-serine	29	2.87×10^{4}	27	274	264	277	289
glycine	33	1.05×10^{4}	--	--	26	29	163
aspartic acid	--	--	17	485	248	276	206
alanine	--	1.82×10^{4}	--	182	424	182	453

glycine	--	5117	33	1.05×10^4	26	29	163
Total hits	49		56				

The results in Table 6.1 can be further evaluated using a receiver operating characteristic (ROC) curve [128,190] for each final hit list (Fig. 6.4). Using the ROC curves, the area under the curve (AUC) may be calculated as a quantitative metric to compare the performance of the F-ratio analyses. The AUC represents the probability that a randomly selected hit will correctly be assigned a true positive or false positive [128,191]. An ideal ROC curve exhibits an AUC of 1, indicating all true positives were discovered as the top hits in the hit list. The ranking worsens as the AUC deviates from 1. To generate the ROC curves for the artifact removal development, the true positive rate (sensitivity) is calculated as the running sum of true positives discovered out of the total number of possible true positives (29), while the false positive rate is the running sum of the discovered false positives out of the total number of false positives observed up to a suitable stopping point beyond the last observed true positive. We selected to prepare the ROC curves, provided in Fig. 6.4, using the top 40 hits from Table 6.1 as a reasonable stopping point, as both the ^{CN}F-ratio and ^{S}F-ratio analyses had Hit# 33 as the last true positive. The true positive rate was calculated relative to the 29 potentially discoverable spiked metabolites, with 26 true positives for the ^{CN}F-ratio analysis, 23 true positives for the ^{S}F-ratio analysis. An AUC of 0.83 was determined for the ^{CN}F-ratio analysis, indicating that the hits were ranked much better than the ^{S}F-ratio analysis with an AUC of only 0.59. The implications of this first study were two-fold. First, it was established that the artifact removal algorithm was very effective at cleaning up the hit lists to benefit true positive discovery. Second, a substantial improvement in true positive discovery was obtained using ^{CN}F-ratio analysis relative to using ^{S}F-ratio analysis [142]. Both of these outcomes have important implications for the second study of the farmed and tank-raised

pacu data set aimed to provide insight into the implications of farming fish on the same land that rice is farmed.

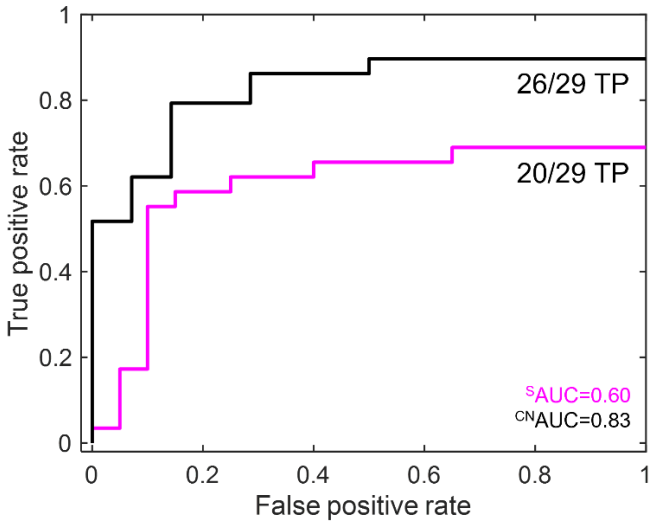


Figure 6.4. Receiver operating characteristic (ROC) curves for evaluating the performance of the ^{S}F-ratio and ^{C}NF-ratio analyses using the top 40 hits in each hit list. The true positive rate was calculated as the running sum of the true positive hits that were not removed in any step out of a total 29 discoverable hits. The false positive rate was calculated as the running sum of the false positive out of a total number of remaining hits in the hit list. AUC values of 0.60 and 0.83 are provided for ^{S}F-ratio and ^{CN}F-ratio, respectively.

6.3.2 Evaluation of ^{MVO}F-ratio analysis for tank-raised and farmed fish comparison

For the second study, contour plots of analytical ion chromatograms (AIC) are provided in Figs. 6.5A and 6.5B for visual inspection of the tank-raised and farmed classes, respectively. The AIC contour plots of the average chromatogram per class were plotted using *m/z* 73, 117, 174, and 217 which are representative fragment ions of TMS derivatives, alcohols, carboxylic acids, primary amines, and sugars. The AIC is displayed instead of the TIC (Fig. C.5 in Appendix C) as to show the complexity of the samples without the artifact peaks drowning out the signal of smaller analytes. There are some differences between the classes that are visible between contour plots in in Figs. 6.5A and 6.5B, which can be readily seen in the zoom-in plots of the tank-raised fish (Fig. 6.5C) and farmed fish (Fig. 6.5D) chromatograms between 14 and 20

min with hits discovered by ^{MVO}F-ratio analysis in this time range circled. Four of the hits are circled in green as examples of hits that are visibly different between the classes. Closer inspection of the data in the preliminary PCA study [32] revealed that within-class variance could be high for either class depending on the peak being considered (%*RSD* in average peak signal ranged from 4 to 115%), thus using either ^{S}F-ratio analysis or ^{CN}F-ratio analysis would fail to discover all true positives. To account for this, ^{MVO}F-ratio analysis via Eq. (6.3) was devised and implemented whereby the lowest within-class variance would be used per each discovered hit, allowing for the discovery of the most true positives. While this approach was demonstrated for a two-class comparison, it may be easily adjusted to handle multiple classes, since the ^{S}F-ratio analysis has been modified to handle multiple classes [70].

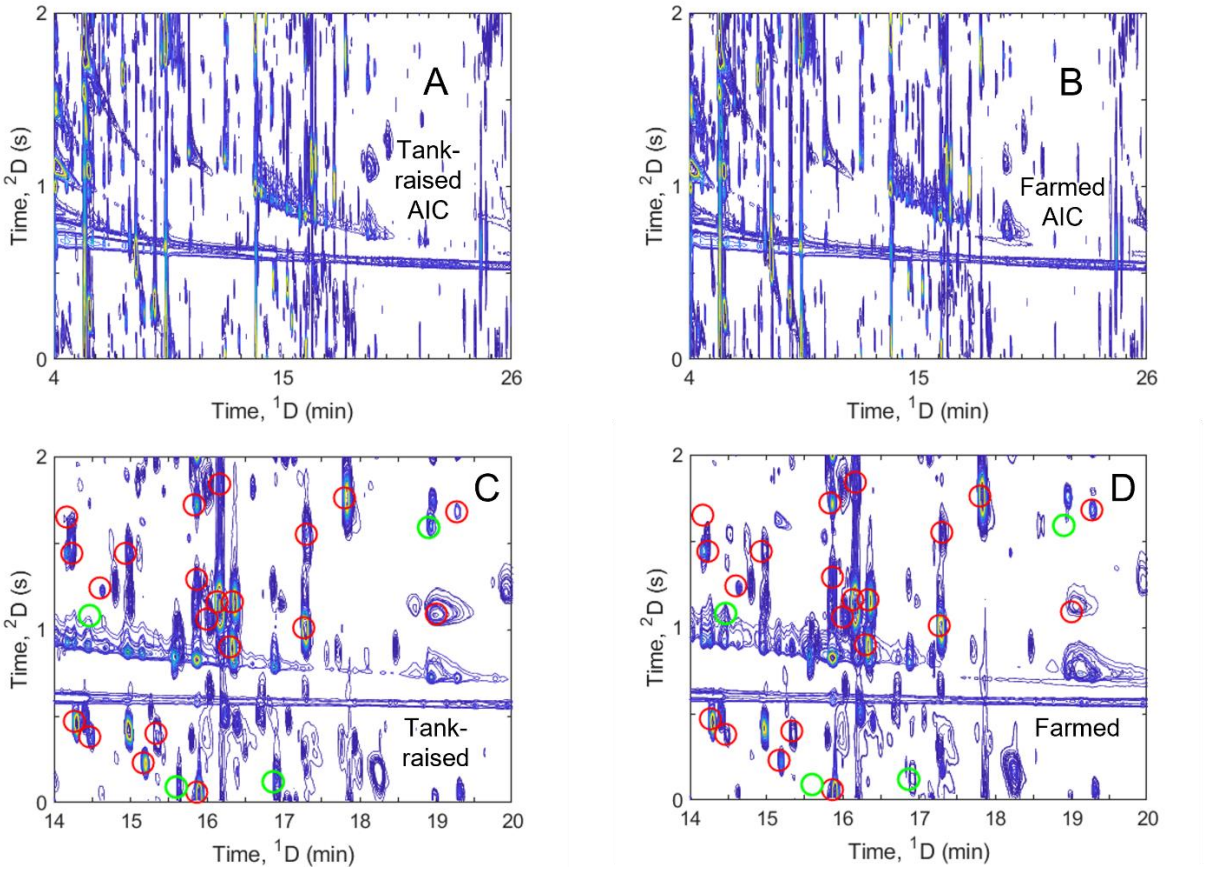


Figure 6.5. Analytical ion chromatogram contour plots of the tank-raised (A) and farmed (B) fish samples using *m/z* 73, 117, 174, and 217. These mass fragments correspond to trimethylsilyl derivatives, alcohols, carboxylic acids, primary amines, and sugars. To better view the differences between classes, a zoom-in of 14 to 20 min is provided for the tank-raised (C) and farmed (D) samples with all the hits discovered by ^{MVO}F-ratio analysis in this region circled. Four hits are circled in green in both (C) and (D) as the differences are readily visible. The left-most green circle at about 14.5 min is an unknown metabolite (Hit# 95) which is visibly in higher concentration in the farmed class ([F]/[T]=1.3), the next hit circled in green is at around 15.6 min is DL-ornithine which is present in the tank-raised class but does not appear in the farmed class ([F]/[T]=0.27), the third hit circled in green is L-ornithine at 16.9 min and is greater in concentration in the tank-raised class ([F]/[T]=0.029), and the final hit circled in green is tyrosine at 18.9 min which is present in the tank-raised class but not visible in the farmed class contour plot ([F]/[T]=0.051). As can be seen, most of the hits in the hit list were downregulated in the farmed pacu fish.

To evaluate the performance of ^{MVO}F-ratio analysis Eq. (6.3), the ^{S}F-ratio (Eq. (6.1)), the tank-normalized ^{TN}F-ratio and farm-normalized ^{FN}F-ratio (both using Eq. (6.2)) analyses were also performed, and the four hit lists after artifact removal were compared. Following metabolite hit identification, a *LOF* was calculated for all *m/z* that passed the *S/N* threshold along with a p-

value to determine the most pure *m/z* to use for the concentration ratio calculation [192]. To facilitate this process, all *m/z* with a calculated *LOF* < 20% were initially retained, then the *m/z* with the lowest p-value was selected for quantification [54]. For several analytes a *LOF* is not reported as the analyte was not present in all samples, resulting in an inflated *LOF*. The concentration ratio was then calculated as the [Farmed]/[Tank-raised], or simply the [F]/[T]. First, using the selected most pure *m/z* all of the ^{2}D modulated peaks are summed together to get a single analyte peak which is quantified by the peak height for each replicate. The average analyte signal of the two replicates collected per fish within a class was then calculated. Next, the average signal of the analyte was calculated per class and the ratio of these 2 averages was the reported [F]/[T] with an accompanying p-value. The 20 average signals per fish (10 per class) were used in a t-test assuming unequal variances to calculate a p-value with 95% confidence. Full hit lists are provided in Tables C.2-C.4 in Appendix C, however only the ^{MVO}F-ratio analysis hit list is provided in Table 2, with tentative compound identification, F-ratio, final hit number after artifact removal, the pure *m/z* used for the [F]/[T] calculation followed by the [F]/[T] value, and the salient p-value and *LOF* statistical metrics. A few unknown metabolites are listed in Table 6.2 as they could not be identified confidently but the mass spectrum includes fragments corresponding to derivatized metabolites. In addition, the original ^{MVO}F-ratio hit list with 541 hits was reduced to 110 hits using the artifact removal algorithm and removing a few redundant hits, allowing for a much smaller hit list to mine and discovery of true positives higher on the hit list. Of these 110 hits, 12 were unidentified artifacts not removed by the artifact removal algorithm, and an additional 28 were metabolites possessing a p-value ≥ 0.05, leaving a total of 70 hits with a significant concentration ratio (p-value < 0.05), which is double the number of analytes with significantly different concentrations discovered in the preliminary study using

PCA [116]. An additional 7 metabolites possess a p-value < 0.10 for a total of 77 hits at the 90% confidence level.

Table 6.2. The minimum variance optimized (MVO) Fisher ratio analysis hit list of the tank-raised and farmed pacu fish samples. This hit list includes all true positive and false positive hits. The hits are ranked in order of descending F-ratios with the final hit list numbers reported. Next, the pure *m/z* used to calculate the concentration ratio is reported along with the concentration ratio calculated as [farmed]/[tank-raised] ([F]/[T]). Finally, the p-value and *LOF* are reported as the metrics for determining the statistical significance of concentration ratios and pure *m/z*. The lowest *LOF* (minimally a *LOF* < 20%) coupled with the lowest p-value (minimum p-value < 0.05) corresponds to the most pure *m/z*. Some *LOF* are not reported as the analytes may not be present in all samples, inflating the *LOF* values. Only hit numbers and F-ratio values are reported for unidentified artifact peaks.

Compound	**F-ratio**	**Final Hit#**	***m/z***	**[F]/[T]**	**p-value**	***LOF (%)***
propylene glycol	27802	1	117	0.17	5.8E-04	11.8
2-butanol	19113	2	117	0.029	1.8E-05	--
γ-butyrolactone	13456	3	86	0.12	0.011	--
tyrosine	5951	4	179	0.051	0.021	--
DL-phenylalanine	2098	5	74	0.080	1.6E-04	--
scyllo-inositol	1222	6	318	0.15	0.0022	9.5
L-ornithine	816	7	174	0.029	0.049	--
glycerol	751	8	74	1.9	1.1E-05	8.6
D-glucose	750	9	135	0.67	5.9E-05	3.8
succinic acid	702	10	147	0.15	0.096	--
glycolic acid	687	11	204	0.50	9.9E-07	2.7
unidentified artifact 1	634	12	--	--	--	--
2,3-butanediol	591	13	117	2.9	2.3E-07	8.0
caproic acid	551	14	73	0.24	9.4E-07	--
glycerol-3-phosphate	465	15	357	0.45	0.065	3.2
creatinine	428	16	204	0.74	6.3E-05	6.0
N-(hydroxymethyl)trifluoroacetamide	390	17	170	3.1	1.7E-05	12.7
2,2'-methylenediphenol	389	18	171	0.27	0.0041	5.6
mannose	372	19	160	0.63	2.2E-04	2.3
sarcosine	310	20	116	0.14	0.0088	--
pyruvate	289	21	174	0.47	0.012	12.2

lactic acid	280	22	129	1.0	0.49	10.0
L-cysteine	275	23	146	0.65	7.7E-04	1.9
methylaminoheptane	271	24	170	3.0	5.2E-06	6.8
D-alloisoleucine	239	25	86	0.41	0.0023	17.7
L-iosleucine	226	26	218	0.18	9.6E-04	--
hexadecenitrile	213	27	136	0.88	0.38	--
4-hydroxybutyric acid	211	28	147	2.7	3.1E-06	10.3
hypoxanthine	211	29	136	0.45	0.019	9.6
bromosuccinate	204	30	73	0.33	8.3E-06	10.0
L-leucine	200	31	73	0.27	6.1E-05	--
methionine	190	32	147	0.28	0.017	19.9
L-serine	187	33	116	0.41	5.3E-04	19.1
L-valine	158	34	72	0.52	0.0030	18.8
D-mannopyranose	145	35	191	0.83	0.22	3.5
unknown metabolite 1	143	36	73	1.4	1.9E-04	16.4
alanine	136	37	44	0.69	8.3E-04	14.2
N,N-dimethylethanolamine	136	38	58	2.1	6.7E-07	--
mercaptoacetic acid	132	39	221	0.44	9.1E-04	17.2
β-D-glucopyranose	122	40	191	0.73	0.073	2.1
asparagine	122	41	80	0.53	2.8E-06	5.1
L-isoleucine	118	42	86	0.26	0.0066	--
DL-norleucine	115	43	74	0.62	0.014	--
L-arginine	105	44	257	0.45	0.0019	--
myo-inositol	104	45	75	1.9	3.0E-05	4.7
malic acid	103	46	147	0.47	0.13	6.4
unknown metabolite 2	98	47	73	0.35	0.0082	--
acetic acid	95	48	75	1.2	0.38	7.0
unidentified artifact 2	93	49	--	--	--	--
N-trifluoroacetylmorpholine	92	50	113	0.51	0.0050	4.2
unknown metabolite 3	91	51	117	0.89	0.75	15.9
unknown metabolite 4	87	52	188	0.74	0.011	1.9
D-glucose-1-phosphate	83	53	217	0.32	1.5E-04	6.7
maltose	80	54	361	0.52	0.13	4.0

unknown metabolite 5	75	55	84	0.37	0.0028	--
niacinamide	71	56	179	0.77	0.34	8.6
erythrose	62	57	72	0.53	0.0034	12.4
urea	62	58	171	1.6	0.18	--
diethylcarbamate	61	59	174	0.80	0.032	2.6
unidentified artifact 3	58	60	--	--	--	--
nonadecane	54	61	57	1.5	0.25	--
L-fucitol	53	62	117	0.33	5.7E-04	11.7
1,2-propanediol-1-phosphate	53	63	73	1.4	0.060	10.2
norvaline	50	64	144	0.32	0.0017	--
DL-ornithine	45	65	147	0.27	0.21	--
unidentified artifact 4	38	66	--	--	--	--
unidentified artifact 5	37	67	--	--	--	--
β-amino isobutyric acid	34	68	73	1.1	0.49	2.5
glycylisoleucine	34	69	86	0.68	0.098	18.9
L-leucine	33	70	86	0.55	0.031	13.4
unidentified artifact 6	31	71	--	--	--	--
isovaleric acid	31	72	142	1.3	0.011	2.9
2-hydroxyisovaleric acid	30	73	147	0.48	0.049	--
malonic acid	29	74	233	0.32	0.035	--
glycine	29	75	102	1.2	0.31	3.0
4-methyl-3-heptanol	29	76	73	0.57	0.037	--
unidentified artifact 7	28	77	--	--	--	--
pyrocatechol	27	78	254	0.25	0.041	--
D-arabinose	27	79	318	2.9	4.5E-06	8.1
galactose	26	80	73	1.3	0.054	13.4
unidentified artifact 8	26	81	--	--	--	--
stearic acid	25	82	73	1.9	0.050	14.3
2-aminobenzoxazole	25	83	263	1.3	0.37	7.4
salicylic acid	24	84	267	0.16	0.018	--
unidentified artifact 9	24	85	--	--	--	--
L-glutamine	22	86	84	0.79	0.045	17.4
threonic acid	21	87	117	0.50	8.6E-04	--

unknown metabolite 6	21	88	221	0.86	0.51	8.6
citrulline	21	89	188	0.78	0.090	3.7
undecanoic acid	20	90	43	1.6	0.011	9.2
unidentified artifact 10	18	91	--	--	--	--
methylmalonic	18	92	218	0.40	0.034	--
xylitol	18	93	217	2.6	2.4E-05	--
2-ethylhexanoic acid	18	94	201	0.41	0.046	--
unknown metabolite 7	18	95	103	1.3	3.7E-04	--
unidentified artifact 11	17	96	--	--	--	--
unidentified artifact 12	17	97	--	--	--	--
unknown metabolite 8	17	98	73	3.0	2.9E-07	--
creatinine enol	17	99	115	0.15	0.029	--
aspartic acid	16	100	232	0.66	2.5E-04	6.2
unknown metabolite 9	15	101	117	1.1	0.44	--
2-pentanol	14	102	73	1.7	9.9E-04	--
unidentified artifact 13	14	103	--	--	--	--
L-valine	12	104	72	0.55	0.041	--
unknown metabolite 10	9	105	186	0.64	1.6E-05	--
lactose	9	106	204	1.0	0.98	4.5
n-butylamine	8	107	174	4.0	0.38	--
D-xylose	6	108	73	1.5	0.11	6.1
unknown metabolite 11	4	109	147	0.75	0.0017	8.4
2,2-dimethyl-3-pentanol	4	110	73	1.1	0.73	14.3

For the purpose of examining the effectiveness of the artifact peak removal algorithm, we define a true positive as any discovered metabolite regardless of the calculated p-value or known identity and a false positive as any non-metabolite artifact peak. By this definition the ^{MVO}F-ratio analysis discovered a total of 97 true positives. For visual interpretation of these results, PCA was performed on all 40 replicates of the TIC (Fig. 6.6A), all *m/z* concatenated and binned by 8 s on ^{1}D and 100 ms on ^{2}D (Fig. 6.6B), and finally on all the true positives discovered by ^{MVO}F-ratio analysis using the signal from the purest *m/z* withing each hit tile reported in Table 2 (Fig.

6.6C). Roughly the same amount of total variance was captured by PC1 and PC2 in all three scores plots (71-77%), however the loadings on the PC that best capture the chemically relevant between-class variance provide more insight. Loadings on PC2 (15%) were plotted for the TIC (Fig. 6.6D), PC1 (45%) for the concatenated *m/z* (Fig. 6.6E), and PC1 (45%) from the ^{MVO}F-ratio analysis (Fig. 6.6F). Visibly, a smaller set of analytes contribute to distinguishing sample classes using all binned *m/z* than when using either the TIC or the 97 true positives from ^{MVO}F-ratio analysis. Additionally, due to the artifact removal procedure, the loadings for the PCA of the ^{MVO}F-ratio analysis results were not hampered by artifact features while the loadings for the other two PCA results were adversely impacted.

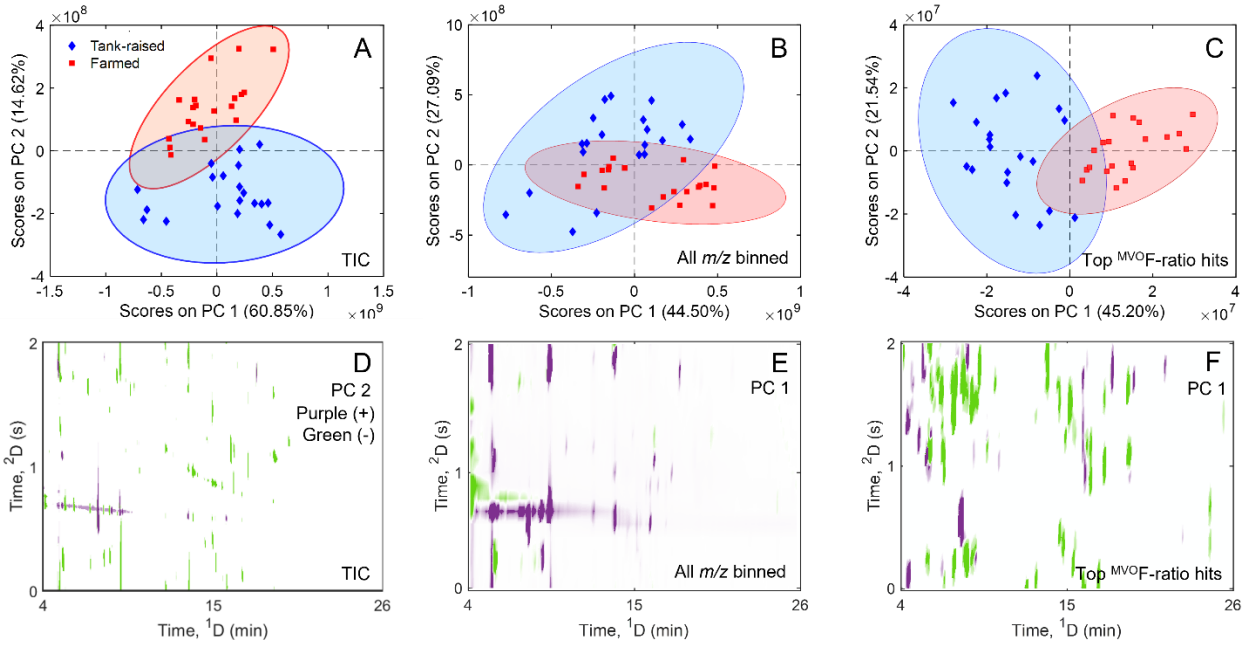


Figure 6.6. Visualization of F-ratio features by PCA. (A) PCA scores plot generated using all 40 replicates of the unfolded TIC. (B) PCA scores plot using all concatenated *m/z*. (C) PCA scores plot using the true positives discovered by ^{MVO}F-ratio using the signal from the purest *m/z* withing each hit tile reported in Table 6.2. The blue diamonds represent the tank-raised fish and red squares represent the farmed fish, clustered with confidence ellipses at the 95% confidence level. Corresponding contour plots of the loadings on PC2 using the TIC (D), PC1 using all *m/z* (E), and PC1 using the true positives from Table 2 (F) were plotted for visual interpretation of

analytes that contributed to distinguishing between the sample classes. All purple features represent positive loadings and all green features represent negative loadings.

In contrast to ^{MVO}F-ratio analysis, ^{S}F-ratio analysis discovered 54, ^{TN}F-ratio analysis discovered 58, and ^{FN}F-ratio analysis discovered 93 (per Tables C.2-C.4 in Appendix C). These F-ratio analyses discovered more true positives than the non-parametric Wilcoxon rank sum test, which discovered only 36 metabolites using a two-tailed test at the 95% confidence level (Table C.6 in Appendix C). The results from these four F-ratio analyses are summarized in Fig. 6.7 with the ROC curves and calculated AUC values. As the ^{MVO}F-ratio analysis discovers the most true positives, the total number of true positives that was used to calculate the true positive rate was 97. As the ^{S}F-ratio analysis discovered the most false positives, the total number of false positives used to calculate the false positive rate was 20. The number of total hits did not vary greatly between original hit lists, ranging from 494 to 546 total hits, indicating that the false positives were still discovered by ^{MVO}F-ratio analysis, but it was beneficial in preserving true positives by ranking them higher than false positives on the hit list. The ^{MVO}F-ratio analysis original hit list was comprised of 541 hits and was reduced to 110 after artifacts and redundant hits were removed, ^{S}F-ratio analysis was reduced from 494 to 74, ^{CN}F-ratio analysis was reduced from 536 to 69, and ^{FN}F-ratio analysis was reduced from 546 to 105 hits. Note that the ^{MVO}F-ratio and ^{FN}F-ratio analyses performed nearly equally, as the ^{MVO}F-ratio analysis only discovered 4 more true positives than the ^{FN}F-ratio analysis resulting in ^{MVO}AUC = 0.82 and ^{FN}AUC = 0.79. The similar performance is due to the overall higher within-class variance in the tank-raised fish class for the true positives that were discovered for this data set. The ^{TN}F-ratio and ^{S}F-ratio AUC are both low (^{TN}AUC = 0.47 and ^{S}AUC = 0.35); while this does not mean they are not effective versions of the F-ratio calculation for discovering class-distinguishing analytes, it does suggest that neither was suited to discover the most true positives for this data set. For a data set that has

greater variability in all of the sample classes, ^{MVO}F-ratio analysis is expected to be even more beneficial. As ^{MVO}F-ratio analysis essentially includes the combined results from ^{FN}F-ratio and ^{TN}F-ratio analyses, the ^{MVO}F-ratio analysis provides a comprehensive analysis coupled with reducing the time required for data mining. More broadly, the analyst does not know ahead of time which analytes and/or sample class will exhibit the lowest within-class variance, so a ^{MVO}F-ratio analysis is the most prudent choice.

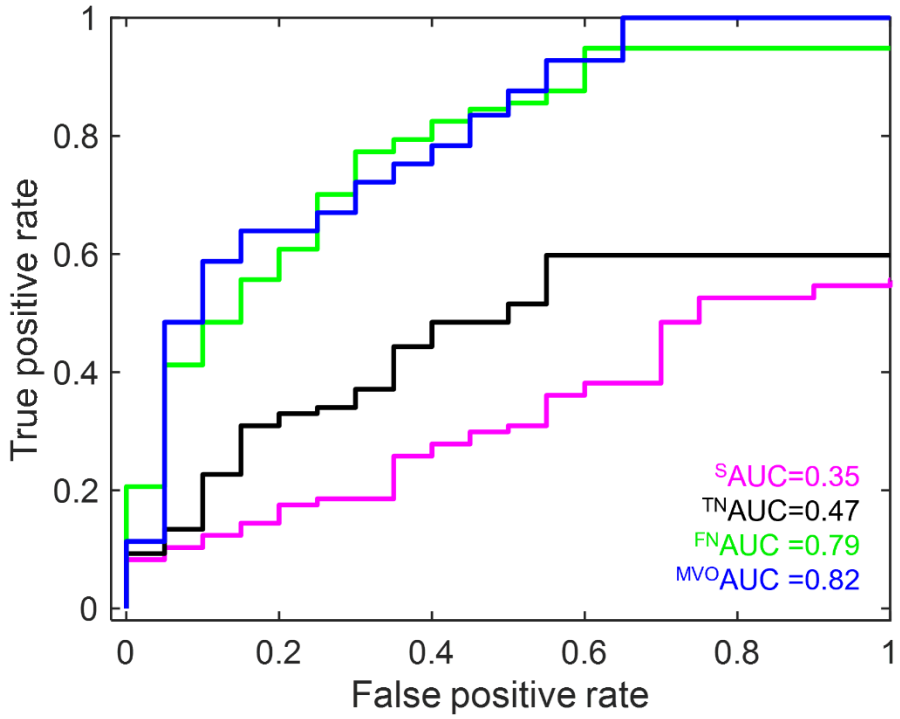


Figure 6.7. Receiver operating characteristic (ROC) curves for evaluating the performance of the four F-ratio calculations for the tank-raised and farmed pacu fish data set. An overlay of ROC curves calculated for the final ^{MVO}F-ratio analysis hit list (blue), ^{S}F-ratio hit list (magenta), ^{TN}F-ratio hit list (black), and ^{FN}F-ratio hit list (green). The true positive rate for each ROC curve is calculated as the running sum of the true positives out of a total of 97 true positives. AUC values of 0.35, 0.47, 0.79, and 0.82 are reported for the ^{S}F-ratio, ^{TN}F-ratio, ^{FN}F-ratio, and ^{MVO}F-ratio analyses, respectively.

We next looked into gaining a more in-depth understanding of *why* ^{MVO}F-ratio analysis provides optimal discovery in the context of how the calculations are performed for the tile-based F-ratio algorithms. For this purpose, the advantage of the ^{MVO}F-ratio calculation relative to

the ^{S}F-ratio calculation is considered, as visualized in Fig. 6.8 whereby the higher F-ratio values calculated by ^{MVO}F-ratio calculation are leveraged during the pin and cluster algorithm [55]. To illustrate this advantage, two representative analytes, γ-butyrolactone (Hit# 3) and N-(hydroxymethyl) trifluoroacetamide (Hit# 17), that were discovered by ^{MVO}F-ratio analysis but not by ^{S}F-ratio analysis were selected. The pin locations of hits discovered by the ^{S}F-ratio analysis are circled in magenta and those discovered by ^{MVO}F-ratio analysis are circled in blue with $^{1}t_{R}$ between 6.2 and 6.8 min (Fig. 6.8A) which includes γ-butyrolactone and between 4.0 and 4.7 min (Fig. 6.8B) that includes N-(hydroxymethyl) trifluoroacetamide. The size of the circle around the pin location corresponds to log(F-ratio) of the hit, so larger circles correspond to hits with a higher F-ratio, the dashed lines (red and black for ease of visualization) outline the four different tile grid boundaries, and the green rectangles highlights the tile around the respective hit indicated by an arrow. Following the pinning step, redundant hits are removed using the clustering part of the algorithm described in detail elsewhere [55], where analytes that are in the same tile and have lower F-ratios are removed from the hit list. The goal of this step is to avoid multiple hits in the hit list that belong to the same analyte peak, referred to as redundant hits. The remaining hits after clustering with $^{1}t_{R}$ between 6.2 and 6.8 min (Fig. 6.8C) and 4.0 and 4.7 min (Fig. 6.8D) are provided. While γ-butyrolactone and N-(hydroxymethyl) trifluoroacetamide are discovered by both F-ratio calculations initially, note that both are removed by the clustering step when the ^{S}F-ratio calculation is used. This occurs due to the A2 (Fig. 6.8C) and A3 (Fig. 6.8D) peaks that have higher F-ratio values when using the ^{S}F-ratio and cluster away the γ-butyrolactone and N-(hydroxymethyl) trifluoroacetamide peaks. Instead, when the ^{MVO}F-ratio calculation is used, the γ-butyrolactone and N-(hydroxymethyl) trifluoroacetamide peaks have a larger F-ratio than the surrounding peaks, evident by the larger

circle size. These examples demonstrate that the ^{MVO}F-ratio is beneficial in the true positives are higher on the F-ratio hit list so that they are not removed in the clustering step of the tile-based F-ratio algorithm.

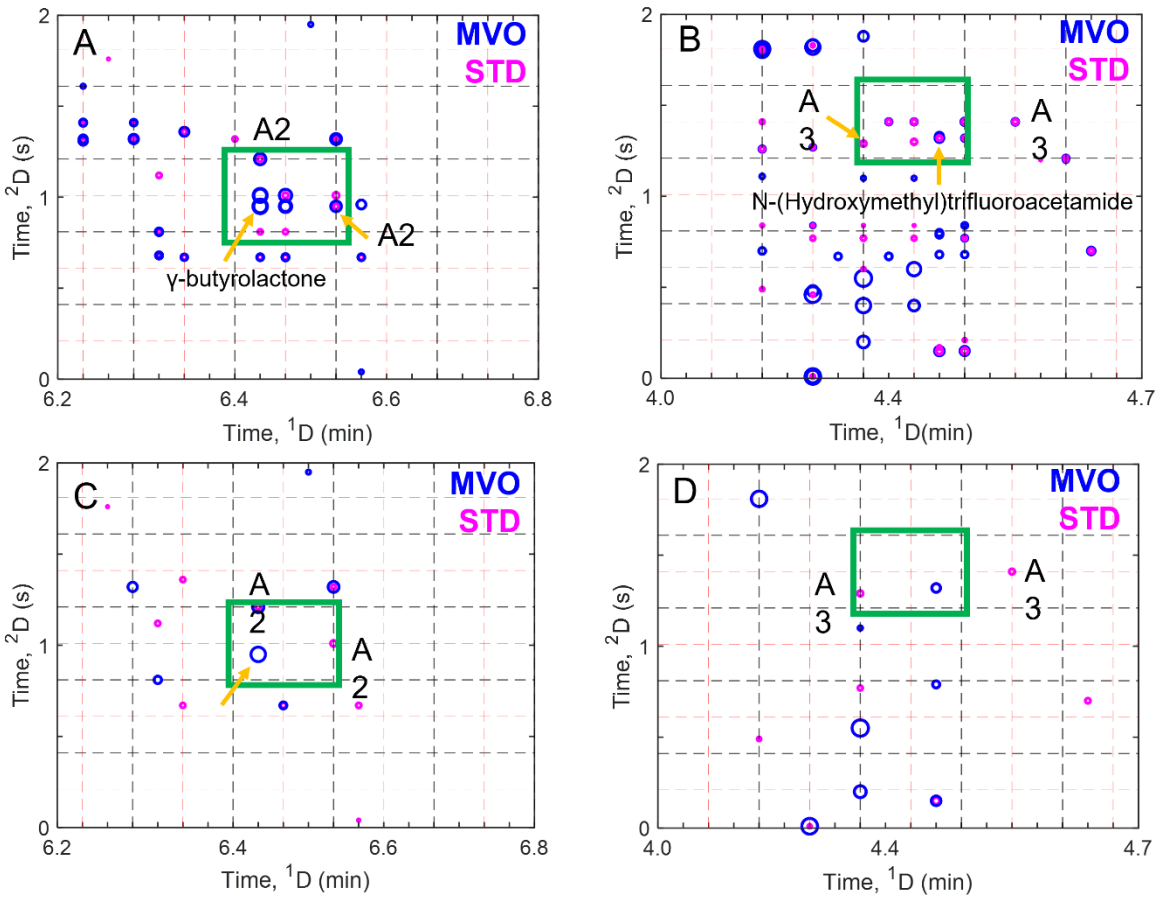


Figure 6.8. Demonstration of the advantage of MVOF-ratio analysis compared to SF-ratio analysis. The pin locations of hits discovered by ^{S}F-ratio (magenta) and ^{MVO}F-ratio (blue) before clustering around γ-butyrolactone (A) and N-(hydroxymethyl) trifluoroacetamide (B) with circle size corresponding to log(F-ratio). The dashed lines represent the tiling grids and the green box highlights the tile that corresponds to the hit indicated by the arrow. Following the clustering step, the remaining pins of hits are plotted in the same time window for γ-butyrolactone (C) and N-(hydroxymethyl) trifluoroacetamide (D) demonstrating that the higher F-ratio values calculated by ^{MVO}F-ratio analysis do not cluster away these true positives while the lower F-ratios for the ^{S}F-ratio analysis results in losing these hits.

6.3.3 Discussion on significant metabolites

It is interesting that most of the true positive hits with p-value < 0.05 (54/70) had a [F]/[T] < 1, indicating that most of the discovered compounds including important metabolites

necessary for sustaining regular metabolic processes were significantly downregulated in the farmed fish. A few examples of the hits with significant differences in concentration between the classes are starred in Fig. 6.9 for visual interpretation using the corresponding *m/z* in Table 2. These plots are summed ^{2}D plots, where all the modulations are summed together in one representative peak per analyte. Figures 6.9A, 6.9B, and 6.9D are examples of hits with [F]/[T] < 1, or hits that were downregulated in the farmed class, tentatively identified as 2-butanol (Hit# 2), γ-butyrolactone (Hit# 3), and L-isoleucine (Hit# 26), respectively. 2-butanol has been associated to a sweet, flowery odor in fish paste samples [56] and along with L-ornithine, involved in tissue repair and immune response in fish [57], 2-butanol had the lowest calculated [F]/[T] of 0.029 (34-fold decrease relative to tank-raised class). The top hit, propylene glycol with [F]/[T] = 0.17, and glycerol with [F]/[T] = 1.9, also pictured in Fig. 6.9A, are proposed as a dietary supplements for several types of fish [58]. γ-butyrolactone is sold commercially as a solvent labeled as fish tank cleaner and has been used in the synthesis of pesticides [59,60], with an 8-fold higher concentration in the tank-raised fish with [F]/[T] = 0.12. L-isoleucine, downregulated in the farmed fish with a [F]/[T] = 0.18, plays an important role in purine biosynthesis as a precursor to aspartic acid, which is important for cell proliferation and immune responses [61]. Examples of the 16/70 hits significantly upregulated in the farmed fish include 2,3-butanediol (Fig. 9C), 4-hydroxybutyric acid (Fig. 6.9E), and myo-inositol (Fig. 6.9F). 2,3-butanediol was one of the most upregulated compounds in the farmed fish with a [F]/[T] = 2.9 and has been found in cuttlefish, contributing to a fishy odor [62]. In addition, a study on identifying potential markers of microbial degradation in cod, mackerel, and whiting fish determined 2,3-butanediol an indicator of microbial degradation as its concentration increased after a 10-day storage period for all fish types [63]. γ-butyrolactone is metabolized to 4-

hydroxybutyric acid, which is higher in concentration in the farmed fish with a [F]/[T] = 2.7 and is known to have a sedative effect [64]. Myo-inositol, an essential vitamin for animals contributing to growth and nutrition of fish, was upregulated in the farmed fish with a [F]/[T] of 1.9 [59,65]. Several amino acids were identified with significantly different concentrations between the sample classes, which may contribute to different flavor profiles in fish [66]. Another study found that lactic acid, creatinine and phosphoric acid contribute to the taste profile of different fish species by GC-MS and electronic tongue [67]. Specifically, lactic acid ([F]/[T]=1) and creatinine ([F]/[T]=0.74) were associated with *Pagrus major* fish that measured higher in intensity of sourness and richness than four other fish species. Stearic acid was downregulated in the farmed fish, which has been associated to increase in fatty acid synthesis and degradation consistent with higher lipid metabolism as a primary energy pathway in fish under acute hypoxia stress [68]. While these are only a few examples, many more metabolites that are essential for regular functioning were significantly downregulated in the farmed fish, and ultimately, the artifact removal algorithm coupled to ^{MVO}F-ratio analysis made mining metabolomic data hit lists to discover true positives more efficient than ^{S}F-ratio analysis.

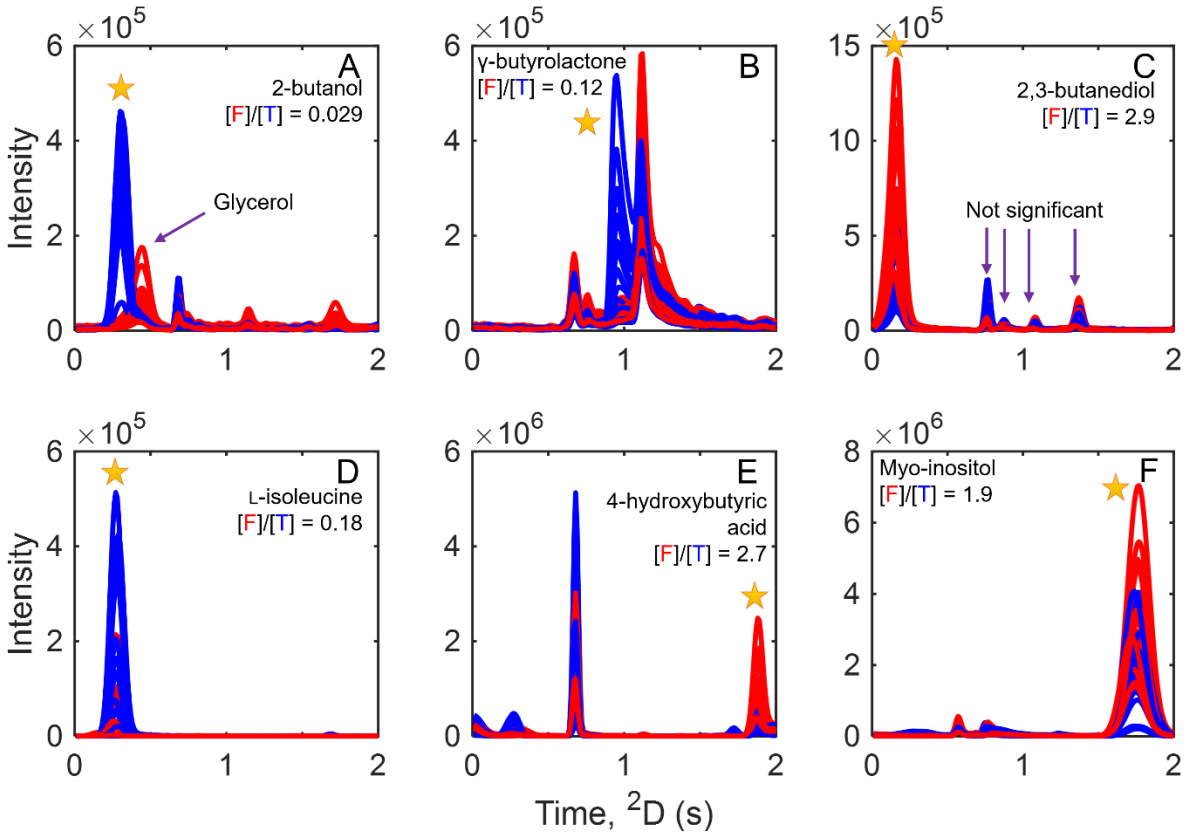


Figure 6.9. Overlays of the summed 2D plots of several metabolites. (A) 2-butanol, (B) γ-butyrolactone, (C) 2,3-butanediol, (D) L-isoleucine, (E) 4-hydroxybutyric acid, and (F) myo-inositol indicated by stars in the respective panels. The analytes are plotted using the associated *m/z* reported in Table 2. All six analytes had significantly different concentrations between classes with 2-butanol, γ-butyrolactone, and L-isoleucine all downregulated in the farmed fish with [F]/[T] of 0.029, 0.12, and 0.18, respectively. On the other hand, 2,3-butanediol, 4-hydroxybutyric acid, and myo-inositol are examples of hits upregulated in the farmed fish class with [F]/[T] of 2.9, 2.7, and 1.9, respectively.

6.4 Conclusion

Discovery of class distinguishing analytes in pacu fish raised in an integrated rice-fish farming system was accomplished in a non-targeted approach by F-ratio analysis of GC×GC-TOFMS data to reveal potential benefits and/or risks of this farming method on the fish metabolome. Initial examination of the data revealed many peaks resulting from the derivatization process were discovered as false positives in the F-ratio hit list. To address this challenge, an artifact removal algorithm was developed and implemented, which facilitated the

discovery of true positives while mining the hit list more efficiently. This algorithm was developed and initially evaluated in the first study on a spiked fish data set, reducing hit lists by 70-80% while only sacrificing a few true positives. For the second study evaluating the metabolome differences of the tank-raised and farmed fish sample classes, due to non-normal distributions for both classes, the ^{MVO}F-ratio implementation was introduced and applied. The ^{MVO}F-ratio implementation discovered more true positives than the standard, tank-normalized, or farm-normalized alone and proved to provide equivalent results as combining the ^{TN}F-ratio and ^{FN}F-ratio analyses, while requiring only one hit list to be mined. ^{MVO}F-ratio discovers more true positives than ^{S}F-ratio due to the higher calculated F-ratio values that decreases their likelihood of being removed in the clustering step of the algorithm. ROC curves and AUC were calculated for the original ^{MVO}F-ratio hit list (AUC=0.82), ^{S}F-ratio (AUC=0.35), ^{TN}F-ratio (AUC=0.47) and ^{FN}F-ratio (AUC=0.79), indicating that ^{MVO}F-ratio and ^{FN}F-ratio performed almost equally well with higher within-class variance in the tank-raised fish class overall. Using ^{MVO}F-ratio, 70 true positives out of a total of 97 discovered metabolites possessed a significant concentrations ratio $p < 0.05$. Most of these true positives with $p < 0.05$ (54 out of 70) were downregulated in the farmed fish that are necessary to maintain regular metabolic processes and fish quality, suggesting the integrated farming system may negatively impact the pacu fish.

6.5 Acknowledgements

Dr. Celina Monzón thanks the Fulbright Scholarship Program that funded her appointment as a Visiting Researcher at the University of Washington.

6.6 References

[1] G.M.R. Valladão, S.U. Gallani, F. Pilarski, South American fish for continental aquaculture, Rev. Aquac. 10 (2018) 351–369. doi:10.1111/raq.12164.

[2] M. Halwart, Trends in rice-fish farming, FAO Aquac. Newsl. (2019).

[3] N.F. Schenone, L. Vackova, A.F. Cirelli, Fish-farming water quality and environmental concerns in Argentina: A regional approach, Aquac. Int. 19 (2011) 855–863. doi:10.1007/s10499-010-9404-x.

[4] P.E. Sudol, K.M. Pierce, S.E. Prebihalo, K.J. Skogerboe, B.W. Wright, R.E. Synovec, Development of gas chromatographic pattern recognition and classification tools for compliance and forensic analyses of fuels: A review, Anal. Chim. Acta (2020). doi:10.1016/j.aca.2020.07.027.

[5] C. Cordero, P. Rubiolo, L. Cobelli, G. Stani, A. Miliazza, M. Giardina, R. Firor, C. Bicchi, Potential of the reversed-inject differential flow modulator for comprehensive two-dimensional gas chromatography in the quantitative profiling and fingerprinting of essential oils of different complexity, J. Chromatogr. A 1417 (2015) 79–95. doi:10.1016/j.chroma.2015.09.027.

[6] 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. doi:10.1016/j.trac.2018.10.015.

[7] 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. doi:10.1016/j.trac.2018.05.007.

[8] S. Fernando, A. Renaguli, M.S. Milligan, J.J. Pagano, P.K. Hopke, T.M. Holsen, B.S. Crimmins, Comprehensive analysis of the Great Lakes top predator fish for novel halogenated organic contaminants by GC×GC-HR-ToF mass spectrometry, Environ. Sci. Technol. 52 (2018) 2909–2917. doi:10.1021/acs.est.7b05999.

[9] H.R. Yamanaka, C. Cheung, J.S. Mendoza, D.J. Oliva, K. Elzey-Aberilla, K.A. Perrault, Pilot study on exhaled breath analysis for a healthy adult population in Hawaii, Molecules 26 (2021) 3726. doi:10.3390/molecules26123726.

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

[11] L.M. Blumberg, F. David, M.S. Klee, P. Sandra, Comparison of one-dimensional and comprehensive two-dimensional separations by gas chromatography, J. Chromatogr. A 1188 (2008) 2–16. doi:10.1016/j.chroma.2008.02.044.

[12] 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.

[13] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1987) 37–52. doi:https://doi.org/10.1016/0169-7439(87)80084-9.

[14] N.G.S. Mogollón, G.L. Alexandrino, J.R. de Almeida, Z. Niño-Ruiz, J.G. Peña-Delgado, R. Torres-Gutiérrez, F. Augusto, Comprehensive two-dimensional gas chromatography-mass spectrometry combined with multivariate data analysis for pattern recognition in Ecuadorian spirits, Chem. Cent. J. 12 (2018) 1–10. doi:10.1186/s13065-018-0470-x.

[15] 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. doi:10.1016/j.chroma.2017.05.064.

[16] K.M. Pierce, J.L. Hope, J.C. Hoggard, R.E. Synovec, A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) separations of metabolites in plant samples, Talanta 70 (2006) 797–804. doi:10.1016/j.talanta.2006.01.038.

[17] 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.

[18] 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. doi:10.1016/j.chroma.2011.01.086.

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

[20] 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.

[21] G.S. Ochoa, S.E. Prebihalo, B.C. Reaser, L.C. Marney, R.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, J. Chromatogr. A 1627 (2020). doi:10.1016/j.chroma.2020.461401.

[22] A.H. Smith, R. Gnanadesikan, J.B. Hughes, Multivariate analysis of variance (MANOVA), Biometrics 18 (1962) 22–41.

[23] C.J. Huberty, S. Olejnik, Applied MANOVA and discriminant analysis, Wiley, 2006. https://books.google.com/books?id=Cy%5C_IoTEKkngC.

[24] 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 and Fuels 34 (2020) 4084–4094. doi:10.1021/acs.energyfuels.9b04108.

[25] E. Szymańska, J. Gerretzen, J. Engel, B. Geurts, L. Blanchet, L.M.C. Buydens, Chemometrics and qualitative analysis have a vibrant relationship, TrAC - Trends Anal. Chem. 69 (2015) 34–51. doi:10.1016/j.trac.2015.02.015.

[26] K.K. Pasikanti, J. Norasmara, S. Cai, R. Mahendran, K. Esuvaranathan, P.C. Ho, E.C.Y. Chan, Metabolic footprinting of tumorigenic and nontumorigenic uroepithelial cells using two-dimensional gas chromatography time-of-flight mass spectrometry, Anal. Bioanal. Chem. 398 (2010) 1285–1293. doi:10.1007/s00216-010-4055-3.

[27] A. Willse, A.M. Belcher, G. Preti, J.H. Wahl, M. Thresher, P. Yang, K. Yamazaki, G.K. Beauchamp, Identification of major histocompatibility complex-regulated body odorants by

statistical analysis of a comparative gas chromatography/mass spectrometry experiment, Anal. Chem. 77 (2005) 2348–2361. doi:10.1021/ac048711t.

[28] K. Toustrup, B.S. Sørensen, M. Nordsmark, M. Busk, C. Wiuf, J. Alsner, J. Overgaard, Development of a hypoxia gene expression classifier with predictive impact for hypoxic modification of radiotherapy in head and neck cancer, Cancer Res. 71 (2011) 5923–5931. doi:10.1158/0008-5472.CAN-11-1182.

[29] S. Datta, G.A. Satten, Rank-sum tests for clustered data, J. Am. Stat. Assoc. 100 (2005) 908–915. doi:10.1198/016214504000001583.

[30] B.J. Feir-walsh, An empirical comparison of the ANOVA F-test, normal scores test and Kruskal-Wallis test under violation of assumptions, Educ. Psychol. Meas. 34 (1974) 789–799.

[31] B. Buszewski, T. Ligor, T. Jezierski, A. Wenda-Piesik, M. Walczak, J. Rudnicka, Identification of volatile lung cancer markers by gas chromatography-mass spectrometry: Comparison with discrimination by canines, Anal. Bioanal. Chem. 404 (2012) 141–146. doi:10.1007/s00216-012-6102-8.

[32] C. Monzón, S. Schöneich, R.E. Synovec, Non-targeted discovery of class-distinguishing metabolites in Argentinian pacu fish by comprehensive two-dimensional gas chromatography with principal component analysis, Microchem. J. 164 (2021) 106004. doi:10.1016/j.microc.2021.106004.

[33] 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 (2019) 259–266.

[34] S.E. Prebihalo, G.S. Ochoa, K.L. Berrier, K.J. Skogerboe, K.L. Cameron, J.R. Trump, S.J. Svoboda, J.K. Wickiser, R.E. Synovec, 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, Anal. Chem. 92 (2020) 15526–15533. doi:10.1021/acs.analchem.0c03456.

[35] O. Fiehn, J. Kopka, R.N. Trethewey, L. Willmitzer, Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry, Anal. Chem. 72 (2000) 3573–3580. doi:10.1021/ac991142i.

[36] L.C. Marney, S.C. Kolwicz, R. Tian, R.E. Synovec, Sample preparation methodology for mouse heart metabolomics using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry, Talanta 108 (2013) 123–130. doi:10.1016/j.talanta.2013.03.005.

[37] F. Xu, L. Zou, C.N. Ong, Multiorigination of chromatographic peaks in derivatized GC/MS metabolomics: A confounder that influences metabolic pathway interpretation, J. Proteome Res. 8 (2009) 5657–5665. doi:10.1021/pr900738b.

[38] B. Li, S.E. Reichenbach, Q. Tao, R. Zhu, A streak detection approach for comprehensive two-dimensional gas chromatography based on image analysis, Neural Comput. Appl. 32 (2020) 649–663. doi:10.1007/s00521-018-3917-z.

[39] S.L. Nam, A.P. de la Mata, J.J. Harynuk, Automated screening and filtering scripts for GC×GC-TOFMS metabolomics data, Separations 8 (2021) 84. doi:10.3390/separations8060084.

[40] E. Saber, Streak detection in mottled and noisy images, J. Electron. Imaging 16 (2007) 043005. doi:10.1117/1.2816444.

[41] G. Nir, B. Zackay, E.O. Ofek, Optimal and efficient streak detection in astronomical images, Astron. J. 156 (2018) 229. doi:10.3847/1538-3881/aaddff.

[42] S.E. Stein, D.R. Scott, Optimization and testing of mass spectral library search algorithms for compound identification, J. Am. Soc. Mass Spectrom. 5 (1994) 859–866. doi:10.1016/1044-0305(94)87009-8.

[43] 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.

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

[45] O. Amador-Muñoz, P.J. Marriott, Quantification in comprehensive two-dimensional gas chromatography and a model of quantification based on selected summed modulated peaks, J. Chromatogr. A 1184 (2008) 323–340. doi:10.1016/j.chroma.2007.10.041.

[46] R. De Maesschalck, D.L. Massart, The Mahalanobis distance, Chemom. Intell. Lab. Syst. 50 (2000) 1–18.

[47] I. Koo, S. Kim, X. Zhang, Comparative analysis of mass spectral matching-based compound identification in gas chromatography-mass spectrometry, J. Chromatogr. A 1298 (2013) 132–138. doi:10.1016/j.chroma.2013.05.021.

[48] S. Kim, I. Koo, X. Wei, X. Zhang, A method of finding optimal weight factors for compound identification in gas chromatography-mass spectrometry, Bioinformatics 28 (2012) 1158–1163. doi:10.1093/bioinformatics/bts083.

[49] B. Kehimkar, J.C. Hoggard, J.S. Nadeau, R.E. Synovec, Targeted mass spectral ratio analysis: A new tool for gas chromatography - Mass spectrometry, Talanta 103 (2013) 267–275. doi:10.1016/j.talanta.2012.10.043.

[50] 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.

[51] C.D. Brown, H.T. Davis, Receiver operating characteristics curves and related decision measures: A tutorial, Chemom. Intell. Lab. Syst. 80 (2006) 24–38. doi:10.1016/j.chemolab.2005.05.004.

[52] M. Sadeghi, T.K. Lee, D. McLean, H. Lui, M.S. Atkins, Detection and analysis of irregular streaks in dermoscopic images of skin lesions, IEEE Trans. Med. Imaging 32 (2013) 849–861. doi:10.1109/TMI.2013.2239307.

[53] P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis, J. Chromatogr. A 1662 (2022) 462735. doi:10.1016/j.chroma.2021.462735.

[54] G.S. Ochoa, P.E. Sudol, T.J. Trinklein, R.E. Synovec, Class comparison enabled mass spectrum purification for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta 236 (2022) 122844. doi:10.1016/j.talanta.2021.122844.

[55] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS) data using a null distribution approach, Anal. Chem. 87 (2015) 3812–3819. doi:10.1021/ac504472s.

[56] J. Wichaphon, C. Thongthai, A. Assavanig, S. Lertsiri, Volatile aroma components of Thai fish sauce in relation to product categorization, Flavour Fragr. J. 27 (2012) 149–156. doi:10.1002/ffj.2095.

[57] T.C. Clark, J. Tinsley, T. Sigholt, D.J. Macqueen, S.A.M. Martin, Arginine, ornithine and citrulline supplementation in rainbow trout: Free amino acid dynamics and gene expression responses to bacterial infection, Fish Shellfish Immunol. 98 (2020) 374–390. doi:10.1016/j.fsi.2020.01.026.

[58] M.R. Soaudy, E.Y. Mohammady, M.A. Ashry, M.M. Ali, N.M. Ahmed, M.H. Hegab, H.A.S. Elgarhy, E.R. El-Haroun, M.S. Hassaan, Possibility mitigation of cold stress in Nile tilapia under biofloc system by dietary propylene glycol: Performance feeding status, immune, physiological responses and transcriptional response of delta-9-desaturase gene, Aquaculture 538 (2021) 736519. doi:10.1016/j.aquaculture.2021.736519.

[59] D. Pazos, P. Giannasi, Q. Rossy, P. Esseiva, Combining Internet monitoring processes, packaging and isotopic analyses to determine the market structure: Example of Gamma Butyrolactone, Forensic Sci. Int. 230 (2013) 29–36. doi:10.1016/j.forsciint.2013.02.033.

[60] K.E. Davis, L.D. Hickey, J. V. Goodpaster, Detection of ɣ-hydroxybutyric acid (GHB) and ɣ-butyrolactone (GBL) in alcoholic beverages via total vaporization solid-phase microextraction (TV-SPME) and gas chromatography–mass spectrometry, J. Forensic Sci. 66 (2021) 846–853. doi:10.1111/1556-4029.14660.

[61] B.T. Yang, B. Wen, Y. Ji, Q. Wang, H.R. Zhang, Y. Zhang, J.Z. Gao, Z.Z. Chen, Comparative metabolomics analysis of pigmentary and structural coloration in discus fish (Symphysodon haraldi), J. Proteomics 233 (2021). doi:10.1016/j.jprot.2020.104085.

[62] C. Shen, J. Xie, X. Xu, The components of cuttlefish (Sepiella maindroni de Rochebruns) oil, Food Chem. 102 (2007) 210–214. doi:10.1016/j.foodchem.2006.05.014.

[63] G. Duflos, V.M. Coin, M. Cornu, J.F. Antinelli, P. Malle, Determination of volatile compounds to characterize fish spoilage using headspace/mass spectrometry and solid-phase microextraction/gas chromatography/mass spectrometry, J. Sci. Food Agric. 86 (2006) 600–611. doi:10.1002/jsfa.2386.

[64] A.K. Goodwin, P.R. Brown, E.E.W. Jansen, C. Jakobs, K.M. Gibson, E.M. Weerts, Behavioral effects and pharmacokinetics of gamma-hydroxybutyrate (GHB) precursors gamma-butyrolactone (GBL) and 1,4-butanediol (1,4-BD) in baboons, Psychopharmacology (Berl). 204 (2009) 465–476. doi:10.1007/s00213-009-1477-8.

[65] D.F. Deng, G.I. Hemre, R.P. Wilson, Juvenile sunshine bass (Morone chrysops ♀x Morone saxatilis ♂ do not require dietary myo-inositol, Aquaculture 213 (2002) 387–393. doi:10.1016/S0044-8486(02)00119-9.

[66] Z. Duan, S. Dong, Y. Dong, Q. Gao, Geographical origin identification of two salmonid species via flavor compound analysis using headspace-gas chromatography-ion mobility spectrometry combined with electronic nose and tongue, Food Res. Int. 145 (2021) 110385. doi:10.1016/j.foodres.2021.110385.

[67] R. Mabuchi, A. Ishimaru, M. Tanaka, O. Kawaguchi, S. Tanimoto, Metabolic profiling of fish meat by gc-ms analysis, and correlations with taste attributes obtained using an electronic tongue, Metabolites 9 (2019) 1–11. doi:10.3390/metabo9010001.

[68] J.L. Ma, J. Qiang, Y.F. Tao, J.W. Bao, H.J. Zhu, L.G. Li, P. Xu, Multi-omics analysis reveals the glycolipid metabolism response mechanism in the liver of genetically improved farmed Tilapia (GIFT, Oreochromis niloticus) under hypoxia stress, BMC Genomics 22 (2021) 1–16. doi:10.1186/s12864-021-07410-x.

Chapter 7. Conclusion and future directions

Developments in instrumentation and data analysis tools for liquid chromatography with high resolution mass spectrometry and comprehensive two-dimensional gas chromatography (GC×GC) with mass spectrometry detection have made these techniques excellent solutions to tackling characterization of complex samples. Despite their advantages over standard 1D separations with unit resolution MS, the increased complexity of the data and experimental optimization make these techniques less utilized. The goal herein was to provide improvements to data handling tools for complex data (matrices and structures) as well as to explore the use of a more cost-effective modulator for GC×GC analysis that is compatible with MS detectors and provides high quality data. Particularly for addressing chemometric data analysis challenges, data analysis algorithms first developed for GC-MS (and GC×GC-MS) data were modified for LC-HRMS data, demonstrating that collaboration between the LC and GC (and LC×LC and GC×GC) communities can be beneficial for addressing current challenges.

7.1 Chapter 2 Summary and Future Directions

In chapter 2, a method for optimizing the pre-processing of large LC-HRMS data sets uncoupled to subsequent tile-based non-targeted analysis was introduced. The region of interest (ROI) data compression method was utilized for pre-processing five replicate LC-HRMS chromatograms of soil samples spiked at trace levels with pesticides at varying concentrations (0.1 ppb to 50 ppb) to form six distinct classes. This data compression method was used in a targeted manner on entire chromatograms, using the number of detectable analytes and the difference between experimental *m/z* and theoretical *m/z* (Δppm) as metrics for optimal parameter selection, namely admissible mass deviation and *S/N* threshold. Following this pre-processing, the first use of tile-based F-ratio analysis on LC-HRMS data was performed,

successfully discovering all analytes using the optimal parameters. Results suggest the analyst must spike a set of analytes in a sample to use the optimization workflow successfully and ensure proper pre-processing of data prior to non-targeted analysis of choice.

The benefit of this data compression method and the parameter optimization workflow being uncoupled from the subsequent analysis means the analyst is not limited to a specific subset of options. Furthermore, the analyst may select to use pixel-based methods, peak tables, or tile-based methods. This study demonstrated the workflow using tile-based F-ratio analysis as the non-targeted method of choice, however it would be beneficial to demonstrate this method using an unsupervised, non-targeted technique. Additionally, since the method needed to be evaluated using a spiked data set, it should be applied to a real data set for a truly non-targeted study. This workflow could be repeated for studies using unsupervised non-targeted methods such as PCA or HCA for investigating patterns in composition of samples or for PLS to model sample characteristics. Furthermore, the novel unsupervised technique VRI-USI could be used to determine sample class membership prior to supervised, non-targeted analysis. Additionally, it would be beneficial to extend this workflow to non-targeted analyses using multidimensional chromatography with HRMS detection, as data processing is computationally expensive and often analysts are limited to peak tables prior to analysis.

7.2 Chapter 3 Summary and Future Directions

In chapter 3, a cheap and fast flow modulation technique, DPGM in negative pulse mode, was used to study the quality of GC×GC-MS data produced for chemometric decomposition, identification, and quantification. To challenge chemometric decomposition, an isothermal separation of 15 similar analytes in 20 s was selected using a P_M of 250 ms, providing narrow ^2W_b averaging 15 ms and severe overlap on ^{1}D with ^1R_S as low as 0.05. Chemometric

decomposition of all analytes was successful due to sufficiently high ^2R_S between analytes allowing for $R_{S,2D} \geq 0.56$, and accurate analyte quantification and identification with MV≥ 800 was achieved. Results demonstrate this is a powerful modulation technique for fast GC×GC separations that provides high peak capacity and high-quality bilinear data for accurate quantification and identification.

A logical next step would be to apply this method to a real sample, that may have regions of high saturation, or severe analyte overlap, that may benefit from the use of chemometric decomposition tools. While this would be a useful next step, further work using spiked samples would also be beneficial to evaluate the accuracy of identification and quantification efforts. Following this work, chemometric decomposition of GC×GC-MS data collected using DPGM in negative pulse mode could be challenged further as chemometric tools typically start to fail at a R_S of 0.3, thus attempting to collect data with regions of ^2R_S lower than was achieved in this study may be useful to investigate the limit. Additionally, success of such decomposition methods rely upon the analyte *S/N*, the acquisition rate, and identity. Thus, lowering the analyte concentration, using lower acquisition rates, or selecting analytes with more similar mass spectra may further challenge chemometric decomposition. Furthermore, as PARAFAC has more strict trilinear data quality requirements relative to MCR-ALS, a more rigorous study of the data quality could be evaluated using PARAFAC for decomposition. Since this modulation method produces data that requires pre-processing and a loss in sensitivity due to baseline subtraction of the ^{1}D profile, a development that is discussed in chapter 4 involves operating this modulation technique in full modulation mode.

7.3 Chapter 4 Summary and Future Directions

In chapter 4, DPGM in full modulation mode is operated using low flows compatible with MS detection to provide high quality data. Following the previous demonstration of DPGM in partial modulation mode, in this chapter the signal enhancement up to a factor of ~20 and 5-fold *S/N* enhancement relative to GC-MS was demonstrated. Furthermore, the data quality was examined using PARAFAC, as flow modulation techniques can often affect ^{2}D peak shapes. *LOF* and R^2 metrics calculated between raw peaks and PARAFAC loadings suggested excellent trilinear data structure that is amenable to chemometric analysis. Due to the enhancement in sensitivity and facilitated by chemometric decomposition, an LOD of 0.7 ppb was determined for DPGM coupled to TOFMS with accurate identification achieved for highly overlapped analytes.

As this report was the first investigation of using DPGM in full modulation mode for GC×GC with MS detection, further optimization of the separation efficiency is warranted. Particularly, for achieving narrower ^2W_b for higher peak capacity and higher usage of the separation space. Further investigation of the optimization of experimental parameters would be useful for providing other GC×GC users an understanding of how to optimize the system for their own use. Optimization could be investigated on multiple column sets of differing lengths, inner diameters, stationary phase combinations and thicknesses used for different types of applications. Applying this modulation technique for a variety of real, complex samples would be beneficial to demonstrate its broad utility. Additionally, only PARAFAC was applied as a targeted chemometric decomposition tool, and due to the high quality of the data produced it would be worthwhile to use this instrument for the non-targeted analysis of real samples. Many of these points are addressed in chapter 5.

7.4 Chapter 5 Summary and Future Directions

The optimization of DPGM in full modulation mode for GC×GC-TOFMS is investigated in chapter 5 to achieve high peak capacities. The effect of tuning the P_{aux} and the p_w for a given P_M and ${}^2F$ were studied initially on a test mixture and the optimal parameters were selected based upon the average ${}^2W_b$ and $f_{coverage}$ metrics using representative analytes. Furthermore, the effect of sampling induced band broadening on the peak capacity was investigated, thus peak capacities were corrected for $f_{coverage,}$ undersampling by estimating true ${}^1W_b$ from the ${}^1W_b$* and the modulation ratio, and sampling variation via SOT. Results revealed that undersampling broadened ${}^1W_b$ by 10%, therefore decreasing the peak capacity, and was further broadened by 10% due to the statistical aspect of SOT. Broad applicability of this flow modulation technique compatible with MS to achieve high peak capacity was demonstrated for diesel, derivatized cow serum, headspace of coffee, and headspace of a local river water sample.

Due to the ability of this flow modulation technique to operate at fast P_M (down to 50 ms), the low cost of operation, and its compatibility with MS detection make it an excellent modulator for a broad range of applications and warrants greater use and further investigation. Thus far, only targeted chemometric decomposition by PARAFAC has been applied to GC×GC-MS using DPGM. While it was demonstrated that this technique is suitable for achieving high peak capacity separations of a range of applications with great enhancement in *S/N*, it would be beneficial to apply chemometric tools, both targeted and non-targeted, to data collected with this instrumental platform. Additionally, DPGM is an excellent option for secondary modulation on a GC×GC×GC (GC^3) instrument platform. Since ${}^2W_b$ are generally on the order of tens to hundreds of milliseconds, a fast modulator is necessary to adequately sample the effluent off the ^{2}D column, which can be achieved using this modulator. Furthermore, with the increased

complexity of the data collected using a GC^3-TOFMS instrument, chemometric analysis methods must be modified to handle the 4D data structure (^1t_R, ^2t_R, ^3t_R, *m/z*) and leverage the added selectivity from the ^3D.

7.5 Chapter 6 Summary and Future Directions

In this chapter, a comprehensive study of the differences in metabolome of pacu fish raised in tanks versus raised in an integrated rice-fish farming practice were studied by GC×GC-TOFMS with tile-based F-ratio analysis. For metabolomics studies with high within class variation, significant analytes may not be easily discovered by standard F-ratio analysis, therefore a modification to the standard F-ratio analysis termed ^{MVO}F-ratio analysis was introduced for improving significant analyte discovery in metabolomic applications. In fact, ^{MVO}F-ratio outperformed standard F-ratio, as well as F-ratio calculation normalized to each class within class variance as determined by area under the ROC curves. Furthermore, as many metabolomic samples are not readily amenable to GC×GC analysis and require derivatization, often artifacts from the sample preparation procedure makes discovery of significant analytes tedious due to manual investigation of all hits. Thus, a derivatization artifact removal procedure was introduced for redundant peaks of artifacts from the sample preparation procedure, reducing the hitlists by 70-80% and making inspection of truly significant hits more manageable. Furthermore, results of the study (110 hits discovered, 70 statistically significant with $p<0.5$) indicated that a majority of metabolites that play significant roles in regular biological processes were downregulated in the fish raised in the newer integrated farming practice.

Using this approach, twice as many significant analytes were discovered relative to a previous study using PCA. However, more confidence in differences between classes may be achieved using a larger sample size or normalizing signals to fish weight. Furthermore, it may be

interesting to study fish development in both classes and tie behavioral and developmental information to identified analytes from comparative analysis to better understand the impact of the farming environment on the fish. Due to the large number of analytes detected and the similarity in mass spectrum between many metabolites, collecting GC×GC data using a HRMS may also prove beneficial for improved confidence in identification, especially of the unknown analytes with the fragmentation pattern of derivatized metabolites. Additionally, using LC or LC×LC with MS detection could be beneficial for discovering complementary information about samples, potentially detecting residual herbicides that are still used in Argentinian farming practices. Data visualization was provided herein using PCA before and after feature selection using ^{MVO}F-ratio, though it would be interesting to compare PCA results using ^{MVO}F-ratio and ^{S}F-ratio for feature selection. ^{MVO}F-ratio could also be investigated for feature selection of metabolomic samples prior to other non-targeted methods such as PLS-DA, PLS, or HCA and further work using ^{MVO}F-ratio is warranted to get a better understanding of all its benefits and limitations. Furthermore, the artifact removal procedure was used for removing derivatization artifacts, however it could be useful to test for addressing streaks from column bleed or solvent peaks.

Bibliography

Adahchour, M., J. Beens, R. J.J. Vreuls, and U. A.Th Brinkman. 2006. “Recent Developments in Comprehensive Two-Dimensional Gas Chromatography (GC × GC). I. Introduction and Instrumental Set-Up.” TrAC - Trends in Analytical Chemistry 25 (5): 438–54. https://doi.org/10.1016/j.trac.2006.03.002.

Adutwum, Lawrence A., Joanna Koryo Kwao, and James J. Harynuk. 2021. “Unique Ion Filter—A Data Reduction Tool for Chemometric Analysis of Raw Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Data.” Journal of Separation Science 44 (14): 2773–84. https://doi.org/10.1002/jssc.202001127.

Alexandre-Gouabau, Marie Cécile, Frédérique Courant, Thomas Moyon, Alice Küster, Gwénaëlle Le Gall, Illa Tea, Jean Philippe Antignac, and Dominique Darmaun. 2013. “Maternal and Cord Blood LC-HRMS Metabolomics Reveal Alterations in Energy and Polyamine Metabolism, and Oxidative Stress in Very-Low Birth Weight Infants.” Journal of Proteome Research 12 (6): 2764–78. https://doi.org/10.1021/pr400122v.

Alier, Marta, and Romà Tauler. 2013. “Multivariate Curve Resolution of Incomplete Data Multisets.” Chemometrics and Intelligent Laboratory Systems 127: 17–28. https://doi.org/10.1016/j.chemolab.2013.05.006.

Allen, Robert C., and Sarah C. Rutan. 2012. “Semi-Automated Alignment and Quantification of Peaks Using Parallel Factor Analysis for Comprehensive Two-Dimensional Liquid Chromatography-Diode Array Detector Data Sets.” Analytica Chimica Acta 723: 7–17. https://doi.org/10.1016/j.aca.2012.02.019.

Amador-Muñoz, Omar, and Philip J. Marriott. 2008. “Quantification in Comprehensive Two-Dimensional Gas Chromatography and a Model of Quantification Based on Selected Summed Modulated Peaks.” Journal of Chromatography A 1184 (1–2): 323–40. https://doi.org/10.1016/j.chroma.2007.10.041.

Amer, Mohammad W., Blagoj Mitrevski, W. Roy Jackson, Alan L. Chaffee, and Philip J. Marriott. 2014. “Multidimensional and Comprehensive Two-Dimensional Gas Chromatography of Dichloromethane Soluble Products from a High Sulfur Jordanian Oil Shale.” Talanta 120: 55–63. https://doi.org/10.1016/j.talanta.2013.11.069.

Amigo, José Manuel, Thomas Skov, Rasmus Bro, Jordi Coello, and Santiago Maspoch. 2008. “Solving GC-MS Problems with PARAFAC2.” TrAC - Trends in Analytical Chemistry 27 (8): 714–25. https://doi.org/10.1016/j.trac.2008.05.011.

Bahaghighat, H. Daniel, Chris E. Freye, Derrick V. Gough, Paige E. Sudol, and Robert E. Synovec. 2018. “Ultrafast Separations via Pulse Flow Valve Modulation to Enable High Peak Capacity Multidimensional Gas Chromatography.” Journal of Chromatography A 1573. https://doi.org/10.1016/j.chroma.2018.08.001.

Bahaghighat, H. Daniel, Chris E. Freye, Derrick V. Gough, and Robert E. Synovec. 2019. “Comprehensive Two-Dimensional Gas Chromatography and Time-of-Flight Mass Spectrometry Detection with a 50 Ms Modulation Period.” Journal of Chromatography A 1583: 117–23. https://doi.org/10.1016/j.chroma.2018.11.027.

Bahaghighat, H. Daniel, Chris E. Freye, and Robert E. Synovec. 2018. “Recent Advances in Modulator Technology for Comprehensive Two Dimensional Gas Chromatography.” TrAC - Trends in Analytical Chemistry 113: 379–91. https://doi.org/10.1016/j.trac.2018.04.016.

Bailey, Hope P., and Sarah C. Rutan. 2011. “Chemometric Resolution and Quantification of Four-Way Data Arising from Comprehensive 2D-LC-DAD Analysis of Human Urine.” Chemometrics and Intelligent Laboratory Systems 106 (1): 131–41. https://doi.org/10.1016/j.chemolab.2010.07.008.

Ballabio, Davide, and Viviana Consonni. 2013. “Classification Tools in Chemistry. Part 1: Linear Models. PLS-DA.” Analytical Methods 5 (16): 3790–98. https://doi.org/10.1039/c3ay40582f.
Barcaru, Andrei, and Gabriel Vivó-Truyols. 2016. “Use of Bayesian Statistics for Pairwise Comparison of Megavariate Data Sets: Extracting Meaningful Differences between GCxGC-MS Chromatograms Using Jensen-Shannon Divergence.” Analytical Chemistry 88 (4): 2096–2104. https://doi.org/10.1021/acs.analchem.5b03506.
Barker, Matthew, and William Rayens. 2003. “Partial Least Squares for Discrimination.” Journal of Chemometrics 17 (3): 166–73. https://doi.org/10.1002/cem.785.
Bean, Heather D., Jane E. Hill, and Jean Marie D. Dimandja. 2015. “Improving the Quality of Biomarker Candidates in Untargeted Metabolomics via Peak Table-Based Alignment of Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Data.” Journal of Chromatography A 1394: 111–17. https://doi.org/10.1016/j.chroma.2015.03.001.
Beckstrom, Andrew C., Elizabeth M. Humston, Laura R. Snyder, Robert E. Synovec, and Sandra E. Juul. 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.” Journal of Chromatography A 1218 (14): 1899–1906. https://doi.org/10.1016/j.chroma.2011.01.086.
Berrier, Kelsey L., Chris E. Freye, Matthew C. Billingsley, and Robert E. Synovec. 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 and Fuels 34 (4): 4084–94. https://doi.org/10.1021/acs.energyfuels.9b04108.
Bieniek, A., and A. Moga. 2000. “An Efficient Watershed Algorithm Based on Connected Components.” Pattern Recognition 33 (6): 907–16. https://doi.org/10.1016/S0031-3203(99)00154-5.
Blumberg, Leonid M. 2008. “Accumulating Resampling (Modulation) in Comprehensive Two-Dimensional Capillary GC (GC×GC).” Journal of Separation Science 31 (19): 3358–65. https://doi.org/10.1002/jssc.200800424.
Blumberg, Leonid M., Frank David, Matthew S. Klee, and Pat Sandra. 2008. “Comparison of One-Dimensional and Comprehensive Two-Dimensional Separations by Gas Chromatography.” Journal of Chromatography A 1188 (1): 2–16. https://doi.org/10.1016/j.chroma.2008.02.044.
Bos, Tijmen S., Wouter C. Knol, Stef R.A. Molenaar, Leon E. Niezen, Peter J. Schoenmakers, Govert W. Somsen, and Bob W.J. Pirok. 2020. “Recent Applications of Chemometrics in One- and Two-Dimensional Chromatography.” Journal of Separation Science 43 (9–10): 1678–1727. https://doi.org/10.1002/jssc.202000011.
Bro, Rasmus. 1997. “PARAFAC. Tutorial and Applications.” Chemometrics and Intelligent Laboratory Systems 38 (2): 149–71. https://doi.org/10.1016/S0169-7439(97)00032-4.
Bro, Rasmus, Claus A. Andersson, and Henk A.L. Kiers. 1999. “PARAFAC2 - Part II. Modeling Chromatographic Data with Retention Time Shifts.” Journal of Chemometrics 13 (3–4): 295–309. https://doi.org/10.1002/(SICI)1099-128X(199905/08)13:3/4<295::AID-CEM547>3.0.CO;2-Y.
Brown, Christopher D., and Herbert T. Davis. 2006. “Receiver Operating Characteristics Curves and Related Decision Measures: A Tutorial.” Chemometrics and Intelligent Laboratory Systems 80 (1): 24–38. https://doi.org/10.1016/j.chemolab.2005.05.004.
Brownie, Cavell, Dennis D Boos, and Jacqueline Hughes-oliver. 2019. “Modifying the t and ANOVA F Tests When Treatment Is Expected to Increase Variability Relative to Controls.” Biometrics 46 (1): 259–66.

Bruckner, Carsten A., Bryan J. Prazen, and Robert E. Synovec. 1998. “Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis.” Analytical Chemistry 70 (14): 2796–2804. https://doi.org/10.1021/ac980164m.

Bueno, Pedro A., and John V. Seeley. 2004. “Flow-Switching Device for Comprehensive Two-Dimensional Gas Chromatography.” Journal of Chromatography A 1027 (1–2): 3–10. https://doi.org/10.1016/j.chroma.2003.10.033.

Burger, Ben V., Tertia Snyman, Wina J G Burger, and Willie F. Van Rooyen. 2003. “Thermal Modulator Array for Analyte Modulation and Comprehensive Two-Dimensional Gas Chromatography.” Journal of Separation Science 26 (1–2): 123–28. https://doi.org/10.1002/jssc.200390002.

Buszewski, Bogusław, Tomasz Ligor, Tadeusz Jezierski, Anna Wenda-Piesik, Marta Walczak, and Joanna Rudnicka. 2012. “Identification of Volatile Lung Cancer Markers by Gas Chromatography-Mass Spectrometry: Comparison with Discrimination by Canines.” Analytical and Bioanalytical Chemistry 404 (1): 141–46. https://doi.org/10.1007/s00216-012-6102-8.

Bylund, Dan, Rolf Danielsson, Gunnar Malmquist, and Karin E. Markides. 2002. “Chromatographic Alignment by Warping and Dynamic Programming as a Pre-Processing Tool for PARAFAC Modelling of Liquid Chromatography-Mass Spectrometry Data.” Journal of Chromatography A 961 (2): 237–44. https://doi.org/10.1016/S0021-9673(02)00588-5.

Cai, Huamin, and Stanley D. Stearns. 2004. “Partial Modulation Method via Pulsed Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography.” Analytical Chemistry 76 (20): 6064–76. https://doi.org/10.1021/ac0492463.

Cain, Caitlin N., Noah J. Haughn, Hayley J. Purcell, Luke C. Marney, Robert E. Synovec, Chelsea T. Thoumsin, Susan C. Jackels, and Kristen J. Skogerboe. 2021. “Analytical Determination of the Severity of Potato Taste Defect in Roasted East African Arabica Coffee.” Journal of Agricultural and Food Chemistry 69 (7): 2253–61. https://doi.org/10.1021/acs.jafc.1c00605.

Cain, Caitlin N., Sonia Schöneich, and Robert E. Synovec. 2020. “Development of an Enhanced Total Ion Current Chromatogram Algorithm to Improve Untargeted Peak Detection.” Analytical Chemistry 92 (16): 11365–73. https://doi.org/10.1021/acs.analchem.0c02136.

Cain, Caitlin N, Timothy J. Trinklein, Grant S Ochoa, and Robert E. Synovec. 2022. “Tile-Based Pairwise Analysis of GC × GC-TOFMS Data to Facilitate Analyte Discovery and Mass Spectrum Purification.” Analytical Chemistry 94 (14): 5658–66. https://doi.org/10.1021/acs.analchem.2c00223.

Carabajal, Maira D., Juan A. Arancibia, and Graciela M. Escandar. 2018. “Multivariate Curve Resolution Strategy for Non-Quadrilinear Type 4 Third-Order/Four Way Liquid Chromatography–Excitation-Emission Fluorescence Matrix Data.” Talanta 189 (July): 509–16. https://doi.org/10.1016/j.talanta.2018.07.017.

Chong, Il Gyo, and Chi Hyuck Jun. 2005. “Performance of Some Variable Selection Methods When Multicollinearity Is Present.” Chemometrics and Intelligent Laboratory Systems 78 (1): 103–12. https://doi.org/10.1016/j.chemolab.2004.12.011.

Clark, T. C., J. Tinsley, T. Sigholt, D. J. Macqueen, and S. A.M. Martin. 2020. “Arginine, Ornithine and Citrulline Supplementation in Rainbow Trout: Free Amino Acid Dynamics and Gene Expression Responses to Bacterial Infection.” Fish and Shellfish Immunology 98 (January): 374–90. https://doi.org/10.1016/j.fsi.2020.01.026.

Cook, Daniel W., and Sarah C. Rutan. 2017. “Analysis of Liquid Chromatography-Mass Spectrometry Data with an Elastic Net Multivariate Curve Resolution Strategy for Sparse Spectral Recovery.” Analytical Chemistry 89 (16): 8405–12. https://doi.org/10.1021/acs.analchem.7b01832.

Cordero, Chiara, Patrizia Rubiolo, Luigi Cobelli, Gianluca Stani, Armando Miliazza, Matthew Giardina, Roger Firor, and Carlo Bicchi. 2015. “Potential of the Reversed-Inject Differential Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography in the Quantitative Profiling and Fingerprinting of Essential Oils of Different Complexity.” Journal of Chromatography A 1417: 79–95. https://doi.org/10.1016/j.chroma.2015.09.027.

Crucello, Juliana, Luiz F.O. Miron, Victor H.C. Ferreira, He Nan, Marcia O.M. Marques, Patricia S. Ritschel, Mauro C. Zanus, Jared L. Anderson, Ronei J. Poppi, and Leandro W. Hantao. 2018. “Characterization of the Aroma Profile of Novel Brazilian Wines by Solid-Phase Microextraction Using Polymeric Ionic Liquid Sorbent Coatings.” Analytical and Bioanalytical Chemistry 410 (19): 4749–62. https://doi.org/10.1007/s00216-018-1134-3.

Datta, Somnath, and Glen A. Satten. 2005. “Rank-Sum Tests for Clustered Data.” Journal of the American Statistical Association 100 (471): 908–15. https://doi.org/10.1198/016214504000001583.

Davis, Joe M. 1993. “Statistical Theory of Spot Overlap for N-Dimensional Separations.” Analytical Chemistry 65 (October): 2014–23. https://pubs.acs.org/doi/abs/10.1021/ac00019a014.

Davis, Joe M., and J. Calvin Giddings. 1983. “Statistical Theory of Component Overlap in Multicomponent Chromatograms.” Analytical Chemistry 55 (3): 418–24. https://doi.org/10.1021/ac00254a003.

Davis, Joe M., Dwight R. Stoll, and Peter W. Carr. 2008. “Dependence of Effective Peak Capacity in Comprehensive Two-Dimensional Separations on the Distribution of Peak Capacity between the Two Dimensions.” Analytical Chemistry 80 (21): 8122–34. https://doi.org/10.1021/ac800933z.

Davis, Kymeri E., Logan D. Hickey, and John V. Goodpaster. 2021. “Detection of Ɣ-Hydroxybutyric Acid (GHB) and ɣ-Butyrolactone (GBL) in Alcoholic Beverages via Total Vaporization Solid-Phase Microextraction (TV-SPME) and Gas Chromatography–Mass Spectrometry.” Journal of Forensic Sciences 66 (3): 846–53. https://doi.org/10.1111/1556-4029.14660.

Davis, Trenton J., Tarek R. Firzli, Emily A. Higgins Keppler, Matthew Richardson, and Heather D. Bean. 2022. “Addressing Missing Data in GC × GC Metabolomics: Identifying Missingness Type and Evaluating the Impact of Imputation Methods on Experimental Replication.” Analytical Chemistry 94 (31): 10912–20. https://doi.org/10.1021/acs.analchem.1c04093.

Deans, D R. 1968. “A New Technique for Heart Cutting in Gas Chromatography.” Chromatographia 1 (1–2): 18–22. https://doi.org/10.1007/BF02259005.

Deng, Dong Fang, Gro Ingunn Hemre, and Robert P. Wilson. 2002. “Juvenile Sunshine Bass (Morone Chrysops ♀x Morone Saxatilis ♂ Do Not Require Dietary Myo-Inositol.” Aquaculture 213 (1–4): 387–93. https://doi.org/10.1016/S0044-8486(02)00119-9.

Duan, Zelin, Shuanglin Dong, Yunwei Dong, and Qinfeng Gao. 2021. “Geographical Origin Identification of Two Salmonid Species via Flavor Compound Analysis Using Headspace-Gas Chromatography-Ion Mobility Spectrometry Combined with Electronic Nose and Tongue.” Food Research International 145 (November 2020): 110385. https://doi.org/10.1016/j.foodres.2021.110385.

Dubois, Lena M., Pierre Hugues Stefanuto, Laetitia Heudt, Jean François Focant, and Katelynn A. Perrault. 2018a. “Characterizing Decomposition Odor from Soil and Adipocere Samples at a Death Scene Using HS-SPME-GC×GC-HRTOFMS.” Forensic Chemistry 8 (48): 11–20. https://doi.org/10.1016/j.forc.2018.01.001.

Duflos, Guillaume, Vincent Marcel Coin, Marie Cornu, Jean Francois Antinelli, and Pierre Malle. 2006. “Determination of Volatile Compounds to Characterize Fish Spoilage Using Headspace/Mass Spectrometry and Solid-Phase Microextraction/Gas Chromatography/Mass

Spectrometry." Journal of the Science of Food and Agriculture 86 (4): 600–611. https://doi.org/10.1002/jsfa.2386.
Ebadzadsahrai, Ghazal, Emily A. Higgins Keppler, Scott D. Soby, and Heather D. Bean. 2020. "Inhibition of Fungal Growth and Induction of a Novel Volatilome in Response to Chromobacterium Vaccinii Volatile Organic Compounds." Frontiers in Microbiology 11 (May). https://doi.org/10.3389/fmicb.2020.01035.
Edwards, Matthew, Haleigh Boswell, and Tadeusz Górecki. 2015. "Comprehensive Multidimensional Chromatography." Current Chromatography 2 (2): 80–109. https://doi.org/10.2174/2213240602666150722232236.
Fan, Yao Yun, and Hendrik Neubert. 2016. "Quantitative Analysis of Human Neonatal Fc Receptor (FcRn) Tissue Expression in Transgenic Mice by Online Peptide Immuno-Affinity LC-HRMS." Analytical Chemistry 88 (8): 4239–47. https://doi.org/10.1021/acs.analchem.5b03900.
Feir-walsh, Betty J. 1974. "An Emprircal Comparison of the ANOVA F-Test Normal Scores Test and Krsukal-Wallis Test Under Violation of Assumptions." Educational and Psychological Measurement 34 (4): 789–99.
Fernando, Sujan, Aikebaier Renaguli, Michael S. Milligan, James J. Pagano, Philip K. Hopke, Thomas M. Holsen, and Bernard S. Crimmins. 2018. "Comprehensive Analysis of the Great Lakes Top Predator Fish for Novel Halogenated Organic Contaminants by GC×GC-HR-ToF Mass Spectrometry." Environmental Science and Technology 52 (5): 2909–17. https://doi.org/10.1021/acs.est.7b05999.
Fiehn, Oliver, Joachim Kopka, Richard N. Trethewey, and Lothar Willmitzer. 2000. "Identification of Uncommon Plant Metabolites Based on Calculation of Elemental Compositions Using Gas Chromatography and Quadrupole Mass Spectrometry." Analytical Chemistry 72 (15): 3573–80. https://doi.org/10.1021/ac991142i.
Fitz, Brian D., Ryan B. Wilson, Brendon A. Parsons, Jamin C. Hoggard, and Robert E. Synovec. 2012. "Fast, High Peak Capacity Separations in Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry." Journal of Chromatography A 1266: 116–23. https://doi.org/10.1016/j.chroma.2012.09.096.
Fraga, C. G., B. J. Prazen, and Robert E. Synovec. 2001. "Objective Data Alignment and Chemometric Analysis of Comprehensive Two-Dimensional Separations with Run-to-Run Peak Shifting on Both Dimensions." Analytical Chemistry 73 (24): 5833–40. https://doi.org/10.1021/ac010656q.
Franchina, Flavio A., Mariarosa Maimone, Peter Q. Tranchida, and Luigi Mondello. 2016. "Flow Modulation Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Using ≈ 4 ML Min-1 Gas Flows." Journal of Chromatography A 1441: 134–39. https://doi.org/10.1016/j.chroma.2016.02.041.
Franchina, Flavio A., Giorgia Purcaro, Alison Burklund, Marco Beccaria, and Jane E. Hill. 2019. "Evaluation of Different Adsorbent Materials for the Untargeted and Targeted Bacterial VOC Analysis Using GC×GC-MS." Analytica Chimica Acta 1066: 146–53. https://doi.org/10.1016/j.aca.2019.03.027.
Franchina, Flavio Antonio, Maria Elisabete Machado, Peter Q. Tranchida, Cláudia Alcaraz Zini, Elina Bastos Caramão, and Luigi Mondello. 2015. "Determination of Aromatic Sulphur Compounds in Heavy Gas Oil by Using (Low-)Flow Modulated Comprehensive Two-Dimensional Gas Chromatography–Triple Quadrupole Mass Spectrometry." Journal of Chromatography A 1387 (March): 86–94. https://doi.org/10.1016/j.chroma.2015.01.082.
Freye, Chris E., H. Daniel Bahaghighat, and Robert E. Synovec. 2018. "Comprehensive Two-Dimensional Gas Chromatography Using Partial Modulation via a Pulsed Flow Valve with a

Short Modulation Period." Talanta 177 (September 2017): 142–49. https://doi.org/10.1016/j.talanta.2017.08.095.
Freye, Chris E., Patrick R Bowden, Margo T Greenfield, and Bryce C Tappan. 2020. "Non-Targeted Discovery-Based Analysis for Gas Chromatography with Mass Spectrometry: A Comparison of Peak Table, Tile, and Pixel-Based Fisher Ratio Analysis." Talanta 211 (May): 120668. https://doi.org/10.1016/j.talanta.2019.120668.
Freye, Chris E., Nicholas R. Moore, and Robert E. Synovec. 2018. "Enhancing the Chemical Selectivity in Discovery-Based Analysis with Tandem Ionization Time-of-Flight Mass Spectrometry Detection for Comprehensive Two-Dimensional Gas Chromatography." Journal of Chromatography A 1537: 99–108. https://doi.org/10.1016/j.chroma.2018.01.008.
Freye, Chris E., Lan Mu, and Robert E. Synovec. 2015. "High Temperature Diaphragm Valve-Based Comprehensive Two-Dimensional Gas Chromatography." Journal of Chromatography A 1424: 127–33. https://doi.org/10.1016/j.chroma.2015.10.098.
Freye, Chris E., and Robert E. Synovec. 2016. "High Temperature Diaphragm Valve-Based Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry." Talanta 161: 675–80. https://doi.org/10.1016/j.talanta.2016.09.002.
Furbo, Søren, Asger B. Hansen, Thomas Skov, and Jan H. Christensen. 2014. "Pixel-Based Analysis of Comprehensive Two-Dimensional Gas Chromatograms (Color Plots) of Petroleum: A Tutorial." Analytical Chemistry 86 (15): 7160–70. https://doi.org/10.1021/ac403650d.
Geus, Henk Jan De, Jacob De Boer, and Udo A.Th Brinkman. 1997. "Development of a Thermal Desorption Modulator for Gas Chromatography." Journal of Chromatography A 767 (1–2): 137–51. https://doi.org/10.1016/S0021-9673(97)00038-1.
Ghosh, Abhijit, Carly T. Bates, Stacy K. Seeley, and John V. Seeley. 2013. "High Speed Deans Switch for Low Duty Cycle Comprehensive Two-Dimensional Gas Chromatography." Journal of Chromatography A 1291: 146–54. https://doi.org/10.1016/j.chroma.2013.04.003.
Giddings, J. C. 1984. "Two-Dimensional Separations: Concept and Promise." Analytical Chemistry 56 (12). https://doi.org/10.1021/ac00276a003.
Giocastro, Barbara, Mariosimone Zoccali, Peter Q. Tranchida, and Luigi Mondello. 2021. "Evaluation of Different Internal Diameter Coated Modulation Columns within the Context of Solid-State Modulation." Journal of Separation Science, 1–20. https://doi.org/10.1002/jssc.202001031.
Godoy, Luiz Antonio Fonseca de, Leandro W. Hantao, Marcio Pozzobon Pedroso, Ronei Jesus Poppi, and Fabio Augusto. 2011. "Quantitative Analysis of Essential Oils in Perfume Using Multivariate Curve Resolution Combined with Comprehensive Two-Dimensional Gas Chromatography." Analytica Chimica Acta 699 (1): 120–25. https://doi.org/10.1016/j.aca.2011.05.003.
Golay, Marcel J E, and others. 1958. "Theory of Chromatography in Open and Coated Tubular Columns with Round and Rectangular Cross-Sections." Gas Chromatography 2: 36–55.
Golshan, Azadeh, Hamid Abdollahi, Samira Beyramysoltan, Marcel Maeder, Klaus Neymeyr, Robert Rajkó, Mathias Sawall, and Romà Tauler. 2016. "A Review of Recent Methods for the Determination of Ranges of Feasible Solutions Resulting from Soft Modelling Analyses of Multivariate Data." Analytica Chimica Acta 911: 1–13. https://doi.org/10.1016/j.aca.2016.01.011.
Goodwin, A. K., P. R. Brown, E. E.W. Jansen, C. Jakobs, K. M. Gibson, and E. M. Weerts. 2009. "Behavioral Effects and Pharmacokinetics of Gamma-Hydroxybutyrate (GHB) Precursors Gamma-Butyrolactone (GBL) and 1,4-Butanediol (1,4-BD) in Baboons." Psychopharmacology 204 (3): 465–76. https://doi.org/10.1007/s00213-009-1477-8.

Gorrochategui, Eva, Joaquim Jaumot, and Romà Tauler. 2015. “A Protocol for LC-MS Metabolomic Data Processing Using Chemometric Tools.” Protocol Exchange, 1–41. https://doi.org/10.1038/protex.2015.102.

Gorrochategui, Eva, Joaquim Jaumot, and Romà Tauler. 2019. “ROIMCR: A Powerful Analysis Strategy for LC-MS Metabolomic Datasets.” BMC Bioinformatics 20 (1): 1–17. https://doi.org/10.1186/s12859-019-2848-8.

Gough, D.V., S. Schöneich, and R.E. Synovec. 2020. “Chemometric Decomposition of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data Employing Partial Modulation in the Negative Pulse Mode.” Talanta 210. https://doi.org/10.1016/j.talanta.2019.120670.

Gough, Derrick V., H. Daniel Bahaghighat, and Robert E. Synovec. 2019. “Column Selection Approach to Achieve a High Peak Capacity in Comprehensive Three-Dimensional Gas Chromatography.” Talanta 195 (October 2018): 822–29. https://doi.org/10.1016/j.talanta.2018.12.007.

Gough, Derrick V., Dong H. Song, Sonia Schöneich, Sarah E. Prebihalo, and Robert E. Synovec. 2019. “Development of Ultrafast Separations Using Negative Pulse Partial Modulation to Enable New Directions in Gas Chromatography.” Analytical Chemistry 91 (11): 7328–35. https://doi.org/10.1021/acs.analchem.9b01085.

Griffith, James F., William L. Winniford, Kefu Sun, Rob Edam, and Jim C. Luong. 2012. “A Reversed-Flow Differential Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography.” Journal of Chromatography A 1226: 116–23. https://doi.org/10.1016/j.chroma.2011.11.036.

Gros, Jonas, Deedar Nabi, Petros Dimitriou-Christidis, Rebecca Rutler, and J. Samuel Arey. 2012. “Robust Algorithm for Aligning Two-Dimensional Chromatograms.” Analytical Chemistry 84 (21): 9033–40. https://doi.org/10.1021/ac301367s.

Gruber, B., B. A. Weggler, R. Jaramillo, K. A. Murrell, P. K. Piotrowski, and F. L. Dorman. 2018. “Comprehensive Two-Dimensional Gas Chromatography in Forensic Science: A Critical Review of Recent Trends.” TrAC - Trends in Analytical Chemistry 105: 292–301. https://doi.org/10.1016/j.trac.2018.05.017.

Haglund, Peter, Peter Korytár, Conny Danielsson, Jordí Diaz, Karin Wiberg, Pim Leonards, Udo A.T. Brinkman, and Jacob De Boer. 2008. “GC×GC-ECD: A Promising Method for the Determination of Dioxins and Dioxin-like PCBs in Food and Feed.” Analytical and Bioanalytical Chemistry 390 (7): 1815–27. https://doi.org/10.1007/s00216-008-1896-0.

Halwart, Matthias. 2019. “Trends in Rice-Fish Farming.” FAO Aquaculture Newsletter, no. April.

Hamilton, Jacqueline F., Alastair C. Lewis, and Keith D. Bartle. 2003. “Peak Amplitude and Resolution in Comprehensive Gas Chromatography Using Valve Modulation.” Journal of Separation Science 26 (6–7): 578–84. https://doi.org/10.1002/jssc.200390079.

Hantao, Leandro Wang, Carlos Alberto Teixeira, and Victor Hugo Cavalcanti Ferreira. 2022. “CDF2MAT Automated SCRIPT to Import NETCDF Files to MATLAB | RESAMPLING Added to Correct RESHAPE for Non-Integer MS Acquisition Rates in GCxGC-MS Data.” Harvard Dataverse. https://doi.org/doi/10.7910/DVN/WMTEMF.

Hantao, Leandro Wang, Bruna Regina Toledo, Fabiana Alves De Lima Ribeiro, Marilia Pizetta, Caroline Geraldi Pierozzi, Edson Luiz Furtado, and Fabio Augusto. 2013. “Comprehensive Two-Dimensional Gas Chromatography Combined to Multivariate Data Analysis for Detection of Disease-Resistant Clones of Eucalyptus.” Talanta 116: 1079–84. https://doi.org/10.1016/j.talanta.2013.08.033.

Harshman, Richard A, and Margaret E Lundy. 1994. “PARAFAC : Parallel Factor Analysis.” Computational Statistics & Data Analysis 18: 39–72.
Harvey, Paul McA, and Robert A. Shellie. 2011. “Factors Affecting Peak Shape in Comprehensive Two-Dimensional Gas Chromatography with Non-Focusing Modulation.” Journal of Chromatography A 1218 (21): 3153–58. https://doi.org/10.1016/j.chroma.2010.08.029.
Harynuk, James, and Tadeusz Górecki. 2003. “New Liquid Nitrogen Cryogenic Modulator for Comprehensive Two-Dimensional Gas Chromatography.” Journal of Chromatography A 1019 (1–2): 53–63. https://doi.org/10.1016/j.chroma.2003.08.097.
Higgins Keppler, Emily A., Carrie L. Jenkins, Trenton J. Davis, and Heather D. Bean. 2018. “Advances in the Application of Comprehensive Two-Dimensional Gas Chromatography in Metabolomics.” TrAC - Trends in Analytical Chemistry 109: 275–86. https://doi.org/10.1016/j.trac.2018.10.015.
Hoggard, Jamin C., W. Christopher Siegler, and Robert E. Synovec. 2009. “Toward Automated Peak Resolution in Complete GC × GC-TOFMS Chromatograms by PARAFAC.” Journal of Chemometrics 23 (7–8): 421–31. https://doi.org/10.1002/cem.1239.
Hoggard, Jamin C., and Robert E. Synovec. 2007. “Parallel Factor Analysis (PARAFAC) of Target Analytes in GC × GC-TOFMS Data: Automated Selection of a Model with an Appropriate Number of Factors.” Analytical Chemistry 79 (4): 1611–19. https://doi.org/10.1021/ac061710b.
Hoh, Eunha, Nathan G. Dodder, Steven J. Lehotay, Kristin C. Pangallo, Christopher M. Reddy, and Keith A. Maruya. 2012. “Nontargeted Comprehensive Two-Dimensional Gas Chromatography/Time-of-Flight Mass Spectrometry Method and Software for Inventorying Persistent and Bioaccumulative Contaminants in Marine Environments.” Environmental Science and Technology 46 (15): 8001–8. https://doi.org/10.1021/es301139q.
Hohrenk, Lotta L., Maryam Vosough, and Torsten C. Schmidt. 2019. “Implementation of Chemometric Tools to Improve Data Mining and Prioritization in LC-HRMS for Nontarget Screening of Organic Micropollutants in Complex Water Matrixes.” Analytical Chemistry 91 (14): 9213–20. https://doi.org/10.1021/acs.analchem.9b01984.
Huberty, Carl J, and Stephen Olejnik. 2006. Applied MANOVA and Discriminant Analysis. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/047178947X.
Izadmanesh, Yahya, Elba Garreta-Lara, Jahan B. Ghasemi, Silvia Lacorte, Victor Matamoros, and Roma Tauler. 2017. “Chemometric Analysis of Comprehensive Two Dimensional Gas Chromatography–Mass Spectrometry Metabolomics Data.” Journal of Chromatography A 1488: 113–25. https://doi.org/10.1016/j.chroma.2017.01.052.
James, A T, and A J P Martin. 1952. “Gas-Liquid Partition Chromatography: The Separation and Micro-Estimation of Volatile Fatty Acids from Formic Acid to Dodecanoic Acid.” Biochemical Journal 50 (5): 679–90. https://doi.org/10.1042/bj0500679.
Jamin, Emilien L., Nathalie Bonvallot, Marie Tremblay-Franco, Jean Pierre Cravedi, Cécile Chevrier, Sylvaine Cordier, and Laurent Debrauwer. 2014. “Untargeted Profiling of Pesticide Metabolites by LC-HRMS: An Exposomics Tool for Human Exposure Evaluation.” Analytical and Bioanalytical Chemistry 406 (4): 1149–61. https://doi.org/10.1007/s00216-013-7136-2.
Jaumot, Joaquim, and Romà Tauler. 2010. “MCR-BANDS: A User Friendly MATLAB Program for the Evaluation of Rotation Ambiguities in Multivariate Curve Resolution.” Chemometrics and Intelligent Laboratory Systems 103 (2): 96–107. https://doi.org/10.1016/j.chemolab.2010.05.020.
Jennerwein, Maximilian, Markus Eschner, Thomas Wilharm, Thomas Gröger, and Ralf Zimmermann. 2019. “Evaluation of Reversed Phase versus Normal Phase Column Combination for the Quantitative Analysis of Common Commercial Available Middle Distillates Using GC × GC-

TOFMS and Visual Basic Script." Fuel 235 (January 2017): 336–38. https://doi.org/10.1016/j.fuel.2018.07.081.
Jobst, K.J., J.V. Seeley, Eric J. Reiner, L. Mullin, and A. Ladak. 2018. "Enhancing the Sensitivity of Atmospheric Pressure Ionization Mass Spectrometry Using Flow Modulated Gas Chromatography." LCGC North Am. 14 (6): 23–28.
Johnson, Kevin J., and Robert E. Synovec. 2002. "Pattern Recognition of Jet Fuels: Comprehensive GC × GC with ANOVA-Based Feature Selection and Principal Component Analysis." Chemometrics and Intelligent Laboratory Systems 60 (1–2): 225–37. https://doi.org/10.1016/S0169-7439(01)00198-8.
Jolliffe, I T. 1986. "Graphical Representation of Data Using Principal Components." In Principal Component Analysis, 64–91. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4757-1904-8_5.
Juan, A. De, and Romà Tauler. 2003. "Chemometrics Applied to Unravel Multicomponent Processes and Mixtures: Revisiting Latest Trends in Multivariate Resolution." Analytica Chimica Acta 500 (1–2): 195–210. https://doi.org/10.1016/S0003-2670(03)00724-4.
Juan, Anna De, Joaquim Jaumot, and Romà Tauler. 2014. "Multivariate Curve Resolution (MCR). Solving the Mixture Analysis Problem." Analytical Methods 6 (14): 4964–76. https://doi.org/10.1039/c4ay00571f.
Käfer, Uwe, Thomas Gröger, Christoph J. Rohbogner, Daniel Struckmeier, Mohammad Reza Saraji-Bozorgzad, Thomas Wilharm, and Ralf Zimmermann. 2019. "Detailed Chemical Characterization of Bunker Fuels by High-Resolution Time-of-Flight Mass Spectrometry Hyphenated to GC × GC and Thermal Analysis." Energy and Fuels 33 (11): 10745–55. https://doi.org/10.1021/acs.energyfuels.9b02626.
Kehimkar, Benjamin, Jamin C. Hoggard, Jeremy S. Nadeau, and Robert E. Synovec. 2013. "Targeted Mass Spectral Ratio Analysis: A New Tool for Gas Chromatography - Mass Spectrometry." Talanta 103: 267–75. https://doi.org/10.1016/j.talanta.2012.10.043.
Khummueng, Weeraya, James Harynuk, and Philip J. Marriott. 2006. "Modulation Ratio in Comprehensive Two-Dimensional Gas Chromatography." Analytical Chemistry 78 (13): 4578–87. https://doi.org/10.1021/ac052270b.
Kiers, Henk A. L., Jos M. F. ten Berge, and Rasmus Bro. 1999. "PARAFAC2 - Part I. A Direct Fitting Algorithm for the PARAFAC2 Model." Journal of Chemometrics 13 (3–4): 275–94. https://doi.org/10.1002/(sici)1099-128x(199905/08)13:3/4<275::aid-cem543>3.3.co;2-2.
Kim, Seongho, Imhoi Koo, Xiaoli Wei, and Xiang Zhang. 2012. "A Method of Finding Optimal Weight Factors for Compound Identification in Gas Chromatography-Mass Spectrometry." Bioinformatics 28 (8): 1158–63. https://doi.org/10.1093/bioinformatics/bts083.
Kim, Sung Jin, Gustavo Serrano, Kensall D. Wise, Katsuo Kurabayashi, and Edward T. Zellers. 2011. "Evaluation of a Microfabricated Thermal Modulator for Comprehensive Two-Dimensional Microscale Gas Chromatography." Analytical Chemistry 83 (14): 5556–62. https://doi.org/10.1021/ac200336e.
Kinghorn, Russell M., Philip J. Marriott, and Peter A. Dawes. 2000. "Design and Implementation of Comprehensive Gas Chromatography with Cryogenic Modulation." HRC Journal of High Resolution Chromatography 23 (3): 245–52. https://doi.org/10.1002/(SICI)1521-4168(20000301)23:3<245::AID-JHRC245>3.0.CO;2-E.
Klee, Matthew S., Jack Cochran, Mark Merrick, and Leonid M. Blumberg. 2015. "Evaluation of Conditions of Comprehensive Two-Dimensional Gas Chromatography That Yield a near-

Theoretical Maximum in Peak Capacity Gain." Journal of Chromatography A 1383: 151–59. https://doi.org/10.1016/j.chroma.2015.01.031.
Koo, Imhoi, Seongho Kim, and Xiang Zhang. 2013. "Comparative Analysis of Mass Spectral Matching-Based Compound Identification in Gas Chromatography-Mass Spectrometry." Journal of Chromatography A 1298: 132–38. https://doi.org/10.1016/j.chroma.2013.05.021.
Korytár, P., P. E.G. Leonards, Jacob De Boer, and U. A.Th Brinkman. 2005. "Group Separation of Organohalogenated Compounds by Means of Comprehensive Two-Dimensional Gas Chromatography." Journal of Chromatography A 1086 (1–2): 29–44. https://doi.org/10.1016/j.chroma.2005.05.087.
Krupčík, Ján, Roman Gorovenko, Ivan Špánik, Pat Sandra, and Daniel W. Armstrong. 2013. "Flow-Modulated Comprehensive Two-Dimensional Gas Chromatography with Simultaneous Flame Ionization and Quadrupole Mass Spectrometric Detection." Journal of Chromatography A 1280: 104–11. https://doi.org/10.1016/j.chroma.2013.01.015.
Krupčík, Ján, Pavel Májek, Roman Gorovenko, Ivan Špánik, Pat Sandra, and Daniel W. Armstrong. 2013. "On the Determination of a Detector Response Enhancement Factor for Flow Modulated Comprehensive Two-Dimensional Gas Chromatography." Journal of Chromatography A 1286: 235–40. https://doi.org/10.1016/j.chroma.2013.02.068.
LaClair, Russell W., Pedro A. Bueno, and John V. Seeley. 2004. "A Systematic Analysis of a Flow-Switching Modulator for Comprehensive Two-Dimensional Gas Chromatography." Journal of Separation Science 27 (5–6): 389–96. https://doi.org/10.1002/jssc.200301668.
Lee, Loong Chuen, Choong Yeun Liong, and Abdul Aziz Jemain. 2018. "Partial Least Squares-Discriminant Analysis (PLS-DA) for Classification of High-Dimensional (HD) Data: A Review of Contemporary Practice Strategies and Knowledge Gaps." Analyst 143 (15): 3526–39. https://doi.org/10.1039/c8an00599k.
Li, Bo, Stephen E. Reichenbach, Qingping Tao, and Rongbo Zhu. 2020. "A Streak Detection Approach for Comprehensive Two-Dimensional Gas Chromatography Based on Image Analysis." Neural Computing and Applications 32 (3): 649–63. https://doi.org/10.1007/s00521-018-3917-z.
Libardoni, Mark, J. Hunter Waite, and Richard Sacks. 2005. "Electrically Heated, Air-Cooled Thermal Modulator and at-Column Heating for Comprehensive Two-Dimensional Gas Chromatography." Analytical Chemistry 77 (9): 2786–94. https://doi.org/10.1021/ac040161b.
Liu, Zaiyou, and John B Phillips. 1991. "Comprehensive Two-Dimensional Gas Chromatography Using an On-Column Thermal Modulator Interface." Journal of Chromatographic Science 29 (6): 227–31. https://doi.org/10.1093/chromsci/29.6.227.
Lommen, Arjen. 2009. "Metalign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing." Analytical Chemistry 81 (8): 3079–86. https://doi.org/10.1021/ac900036d.
Luong, Jim, Xiaosheng Guan, Shifen Xu, Ronda Gras, and Robert A. Shellie. 2016. "Thermal Independent Modulator for Comprehensive Two-Dimensional Gas Chromatography." Analytical Chemistry 88 (17): 8428–32. https://doi.org/10.1021/acs.analchem.6b02525.
Ma, Jun Lei, Jun Qiang, Yi Fan Tao, Jing Wen Bao, Hao Jun Zhu, Lian Ge Li, and Pao Xu. 2021. "Multi-Omics Analysis Reveals the Glycolipid Metabolism Response Mechanism in the Liver of Genetically Improved Farmed Tilapia (GIFT, Oreochromis Niloticus) under Hypoxia Stress." BMC Genomics 22 (1): 1–16. https://doi.org/10.1186/s12864-021-07410-x.
Mabuchi, Ryota, Ayaka Ishimaru, Mao Tanaka, Osamu Kawaguchi, and Shota Tanimoto. 2019. "Metabolic Profiling of Fish Meat by Gc-Ms Analysis, and Correlations with Taste Attributes

Obtained Using an Electronic Tongue." Metabolites 9 (1): 1–11. https://doi.org/10.3390/metabo9010001.
Maesschalck, R De, and D L Massart. 2000. "The Mahalanobis Distance." Chemometrics and Intelligent Laboratory Systems 50: 1–18.
Marney, Luke C., W. Christopher Siegler, Brendon A. Parsons, Jamin C. Hoggard, Bob W. Wright, and Robert E. Synovec. 2013. "Tile-Based Fisher-Ratio Software for Improved Feature Selection Analysis of Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry Data." Talanta 115: 887–95. https://doi.org/10.1016/j.talanta.2013.06.038.
Marney, Luke C., Stephen C. Kolwicz, Rong Tian, and Robert E. Synovec. 2013. "Sample Preparation Methodology for Mouse Heart Metabolomics Using Comprehensive Two-Dimensional Gas Chromatography Coupled with Time-of-Flight Mass Spectrometry." Talanta 108: 123–30. https://doi.org/10.1016/j.talanta.2013.03.005.
Marriott, Philip J., Sung Tong Chin, Bussayarat Maikhunthod, Hans Georg Schmarr, and Stefan Bieri. 2012. "Multidimensional Gas Chromatography." TrAC - Trends in Analytical Chemistry 34: 1–21. https://doi.org/10.1016/j.trac.2011.10.013.
Marriott, Philip J., and Russell M. Kinghorn. 1997. "Longitudinally Modulated Cryogenic System. A Generally Applicable Approach to Solute Trapping and Mobilization in Gas Chromatography." Analytical Chemistry 69 (13): 2582–88. https://doi.org/10.1021/ac961310w.
Mikaliunaite, Lina, Paige E. Sudol, Caitlin N. Cain, and Robert E. Synovec. 2021. "Baseline Correction Method for Dynamic Pressure Gradient Modulated Comprehensive Two-Dimensional Gas Chromatography with Flame Ionization Detection." Journal of Chromatography A 1652.
Mispelaar, Valentijn G. Van, Albert C. Tas, Age K. Smilde, Peter J. Schoenmakers, and Arian C. Van Asten. 2003. "Quantitative Analysis of Target Components by Comprehensive Two-Dimensional Gas Chromatography." Journal of Chromatography A 1019 (1–2): 15–29. https://doi.org/10.1016/j.chroma.2003.08.101.
Mogollón, Noroska Gabriela Salazar, Guilherme Lionello Alexandrino, José Rafael de Almeida, Zulay Niño-Ruiz, José Gregorio Peña-Delgado, Roldán Torres-Gutiérrez, and Fabio Augusto. 2018. "Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry Combined with Multivariate Data Analysis for Pattern Recognition in Ecuadorian Spirits." Chemistry Central Journal 12 (1): 1–10. https://doi.org/10.1186/s13065-018-0470-x.
Mogollon, Noroska Gabriela Salazar, Fabiana Alves de Lima Ribeiro, Monica Mamian Lopez, Leandro W. Hantao, Ronei Jesus Poppi, and Fabio Augusto. 2013. "Quantitative Analysis of Biodiesel in Blends of Biodiesel and Conventional Diesel by Comprehensive Two-Dimensional Gas Chromatography and Multivariate Curve Resolution." Analytica Chimica Acta 796: 130–36. https://doi.org/10.1016/j.aca.2013.07.071.
Mohd Yusop, Ahmad Yusri, Linda Xiao, and Shanlin Fu. 2019. "Determination of Phosphodiesterase 5 (PDE5)Inhibitors in Instant Coffee Premixes Using Liquid Chromatography-High-Resolution Mass Spectrometry (LC-HRMS)." Talanta 204 (May): 36–43. https://doi.org/10.1016/j.talanta.2019.05.078.
Mohler, Rachel E., Kenneth M. Dombek, Jamin C. Hoggard, Karisa M. Pierce, Elton T. Young, and Robert E. Synovec. 2007. "Comprehensive Analysis of Yeast Metabolite GC×GC-TOFMS Data: Combining Discovery-Mode and Deconvolution Chemometric Software." Analyst 132 (8): 756–67. https://doi.org/10.1039/b700061h.
Mohler, Rachel E., Kenneth M. Dombek, Jamin C. Hoggard, Elton T. Young, and Robert E. Synovec. 2006. "Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass

Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells." Analytical Chemistry 78 (8): 2700–2709. https://doi.org/10.1021/ac052106o.
Mohler, Rachel E., Bryan J. Prazen, and Robert E. Synovec. 2006. "Total-Transfer, Valve-Based Comprehensive Two-Dimensional Gas Chromatography." Analytica Chimica Acta 555 (1): 68–74. https://doi.org/10.1016/j.aca.2005.08.072.
Monzón, Celina, Sonia Schöneich, and Robert E. Synovec. 2021. "Non-Targeted Discovery of Class-Distinguishing Metabolites in Argentinian Pacu Fish by Comprehensive Two-Dimensional Gas Chromatography with Principal Component Analysis." Microchemical Journal 164 (February): 106004. https://doi.org/10.1016/j.microc.2021.106004.
Moreira de Oliveira, Amilton, Carlos Alberto Teixeira, and Leandro W. Hantao. 2022. "Advanced Tuning of the Ion Management Parameters in GC × GC-HRMS Using a Fourier Transform Orbitrap Mass Analyzer for Pixel-Based Data Handling and Multivariate Analysis." Analytical Methods. https://doi.org/10.1039/D2AY00314G.
Mucédola, Vanessa, Luis C.S. Vieira, Danilo Pierone, Angelo L. Gobbi, Ronei J. Poppi, and Leandro W. Hantao. 2017. "Thermal Desorption Modulation for Comprehensive Two-Dimensional Gas Chromatography Using a Simple and Inexpensive Segmented-Loop Fluidic Interface." Talanta 164 (September 2016): 470–76. https://doi.org/10.1016/j.talanta.2016.12.005.
Murphy, Robert E., Mark R. Schure, and Joe P. Foley. 1998. "Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography." Analytical Chemistry 70 (8): 1585–94. https://doi.org/10.1021/ac971184b.
Murray, Jacolin A. 2012. "Qualitative and Quantitative Approaches in Comprehensive Two-Dimensional Gas Chromatography." Journal of Chromatography A 1261: 58–68. https://doi.org/10.1016/j.chroma.2012.05.012.
Murtada, Khaled, David Bowman, Matthew Edwards, and Janusz Pawliszyn. 2022. "Thin-Film Microextraction Combined with Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Screening for Presence of Multiclass Organic Pollutants in Drinking Water Samples." Talanta 242 (December 2021): 123301. https://doi.org/10.1016/j.talanta.2022.123301.
Muscalu, Alina M., Matthew Edwards, Tadeusz Górecki, and Eric J. Reiner. 2015. "Evaluation of a Single-Stage Consumable-Free Modulator for Comprehensive Two-Dimensional Gas Chromatography: Analysis of Polychlorinated Biphenyls, Organochlorine Pesticides and Chlorobenzenes." Journal of Chromatography A 1391 (1): 93–101. https://doi.org/10.1016/j.chroma.2015.02.074.
Muscalu, Alina M., and Tadeusz Górecki. 2018. "Comprehensive Two-Dimensional Gas Chromatography in Environmental Analysis." TrAC - Trends in Analytical Chemistry 106: 225–45. https://doi.org/10.1016/j.trac.2018.07.001.
Myers, Owen D., Susan J. Sumner, Shuzhao Li, Stephen Barnes, and Xiuxia Du. 2017. "One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks." Analytical Chemistry 89 (17): 8696–8703. https://doi.org/10.1021/acs.analchem.7b00947.
Nadeau, Jeremy S., Ryan B. Wilson, Brian D. Fitz, Jason T. Reed, and Robert E. Synovec. 2011. "Utilizing a Constant Peak Width Transform for Isothermal Gas Chromatography." Journal of Chromatography A 1218 (23): 3718–24. https://doi.org/10.1016/j.chroma.2011.04.007.

Nam, Seo Lin, A. Paulina de la Mata, and James J. Harynuk. 2021. “Automated Screening and Filtering Scripts for GC×GC-TOFMS Metabolomics Data.” Separations 8 (6): 84. https://doi.org/10.3390/separations8060084.

Navarro-Reig, Meritxell, Joaquim Jaumot, Anna Baglai, Gabriel Vivó-Truyols, Peter J. Schoenmakers, and Romà Tauler. 2017. “Untargeted Comprehensive Two-Dimensional Liquid Chromatography Coupled with High-Resolution Mass Spectrometry Analysis of Rice Metabolome Using Multivariate Curve Resolution.” Analytical Chemistry 89 (14): 7675–83. https://doi.org/10.1021/acs.analchem.7b01648.

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

Navarro-Reig, Meritxell, Joaquim Jaumot, and Romà Tauler. 2018. “An Untargeted Lipidomic Strategy Combining Comprehensive Two-Dimensional Liquid Chromatography and Chemometric Analysis.” Journal of Chromatography A 1568: 80–90. https://doi.org/10.1016/j.chroma.2018.07.017.

Nir, Guy, Barak Zackay, and Eran O. Ofek. 2018. “Optimal and Efficient Streak Detection in Astronomical Images.” The Astronomical Journal 156 (5): 229. https://doi.org/10.3847/1538-3881/aaddff.

Nolvachai, Yada, Chadin Kulsing, and Philip J. Marriott. 2017. “Multidimensional Gas Chromatography in Food Analysis.” TrAC - Trends in Analytical Chemistry 96: 124–37. https://doi.org/10.1016/j.trac.2017.05.001.

Núñez, Nerea, Oscar Vidal-Casanella, Sonia Sentellas, Javier Saurina, and Oscar Núñez. 2020. “Characterization, Classification and Authentication of Turmeric and Curry Samples by Targeted LC-HRMS Polyphenolic and Curcuminoid Profiling and Chemometrics.” Molecules 25 (12): 1–16. https://doi.org/10.3390/molecules25122942.

Ochoa, Grant S., Sarah E. Prebihalo, Brooke C. Reaser, Luke C. Marney, and Robert E. Synovec. 2020. “Statistical Inference of Mass Channel Purity from Fisher Ratio Analysis Using Comprehensive Two-Dimensional Gas Chromatography with Time of Flight Mass Spectrometry Data.” Journal of Chromatography A 1627 (September): 461401. https://doi.org/10.1016/j.chroma.2020.461401.

Ochoa, Grant S., Paige E. Sudol, Timothy J. Trinklein, and Robert E. Synovec. 2022. “Class Comparison Enabled Mass Spectrum Purification for Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry.” Talanta 236 (May 2021): 122844. https://doi.org/10.1016/j.talanta.2021.122844.

Oldridge, Nathan, Ognjen Panic, and Tadeusz Górecki. 2008. “Stop-Flow Comprehensive Two-Dimensional Gas Chromatography with Pneumatic Switching.” Journal of Separation Science 31 (19): 3375–84. https://doi.org/10.1002/jssc.200800265.

Omais, Badaoui, Marion Courtiade, Nadège Charon, Sébastien Esnault, and Didier Thiébaut. 2012. “Reversal of Elution Order in a Single Second Dimension by Changing the First Column Nature in Comprehensive Two-Dimensional Gas Chromatography.” Journal of Chromatography A 1255: 196–201. https://doi.org/10.1016/j.chroma.2012.07.042.

Omais, Badaoui, Marion Courtiade, Nadège Charon, Jérémie Ponthus, and Didier Thiébaut. 2011. “Considerations on Orthogonality Duality in Comprehensive Two-Dimensional Gas Chromatography.” Analytical Chemistry 83 (19): 7550–54. https://doi.org/10.1021/ac201103e.

Omar, Jone, Maitane Olivares, José Manuel Amigo, and Nestor Etxebarria. 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: 273–80. https://doi.org/10.1016/j.talanta.2013.12.044.

Omar, Jone, Asier Vallejo, Maitane Olivares, Aresatz Usobiaga, Olatz Zuloaga, and Nestor Etxebarria. 2015. “Resolution and Identification of Co-Eluting Alkylphenols in Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry by Multivariate Curve Resolution-Alternating Least Squares.” Journal of Chemometrics 29 (4): 237–44. https://doi.org/10.1002/cem.2701.

Oros, Frances J., and Joe M. Davis. 1992. “Comparison of Statistical Theories of Spot Overlap in Two-Dimensional Separations and Verification of Means for Estimating the Number of Zones.” Journal of Chromatography A 591 (1–2): 1–18. https://doi.org/10.1016/0021-9673(92)80218-J.

Ortiz-Villanueva, Elena, Joaquim Jaumot, Rubén Martínez, Laia Navarro-Martín, Benjamin Piña, and Romà Tauler. 2018. “Assessment of Endocrine Disruptors Effects on Zebrafish (Danio Rerio) Embryos by Untargeted LC-HRMS Metabolomic Analysis.” Science of the Total Environment 635: 156–66. https://doi.org/10.1016/j.scitotenv.2018.03.369.

Paiva, Andre Cunha, and Leandro W. Hantao. 2020. “Exploring a Public Database to Evaluate Consumer Preference and Aroma Profile of Lager Beers by Comprehensive Two-Dimensional Gas Chromatography and Partial Least Squares Regression Discriminant Analysis.” Journal of Chromatography A 1630: 461529. https://doi.org/10.1016/j.chroma.2020.461529.

Panić, Ognjen, Tadeusz Górecki, Christopher McNeish, Allen H. Goldstein, Brent J. Williams, David R. Worton, Susanne V. Hering, and Nathan M. Kreisberg. 2011. “Development of a New Consumable-Free Thermal Modulator for Comprehensive Two-Dimensional Gas Chromatography.” Journal of Chromatography A 1218 (20): 3070–79. https://doi.org/10.1016/j.chroma.2011.03.024.

Parastar, Hadi, and Nadia Akvan. 2014. “Multivariate Curve Resolution Based Chromatographic Peak Alignment Combined with Parallel Factor Analysis to Exploit Second-Order Advantage in Complex Chromatographic Measurements.” Analytica Chimica Acta 816: 18–27. https://doi.org/10.1016/j.aca.2014.01.051.

Parastar, Hadi, Mehdi Jalali-Heravi, and Romà Tauler. 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.” Chemometrics and Intelligent Laboratory Systems 117: 80–91. https://doi.org/10.1016/j.chemolab.2012.02.003.

Parastar, Hadi, Jagoš R. Radović, Mehdi Jalali-Heravi, Sergi Diez, Josep Maria Bayona, and Romà Tauler. 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.” Analytical Chemistry 83 (24): 9289–97. https://doi.org/10.1021/ac201799r.

Parastar, Hadi, and Romà Tauler. 2014. “Multivariate Curve Resolution of Hyphenated and Multidimensional Chromatographic Measurements: A New Insight to Address Current Chromatographic Challenges.” Analytical Chemistry 86 (1): 286–97. https://doi.org/10.1021/ac402377d.

Parsons, Brendon A., Luke C. Marney, W. Christopher Siegler, Jamin C. Hoggard, Bob W. Wright, and Robert E. Synovec. 2015. “Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC-TOFMS) Data

Using a Null Distribution Approach." Analytical Chemistry 87 (7): 3812–19. https://doi.org/10.1021/ac504472s.
Parsons, Brendon A., David K. Pinkerton, and Robert E. Synovec. 2018. "Implications of Phase Ratio for Maximizing Peak Capacity in Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry." Journal of Chromatography A 1536: 16–26. https://doi.org/10.1016/j.chroma.2017.07.018.
Parsons, Brendon A., David K. Pinkerton, Bob W. Wright, and Robert E. Synovec. 2016. "Chemical Characterization of the Acid Alteration of Diesel Fuel: Non-Targeted Analysis by Two-Dimensional Gas Chromatography Coupled with Time-of-Flight Mass Spectrometry with Tile-Based Fisher Ratio and Combinatorial Threshold Determination." Journal of Chromatography A 1440: 179–90. https://doi.org/10.1016/j.chroma.2016.02.067.
Pasikanti, Kishore Kumar, Juwita Norasmara, Shirong Cai, Ratha Mahendran, Kesavan Esuvaranathan, Paul C. Ho, and Eric Chun Yong Chan. 2010. "Metabolic Footprinting of Tumorigenic and Nontumorigenic Uroepithelial Cells Using Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry." Analytical and Bioanalytical Chemistry 398 (3): 1285–93. https://doi.org/10.1007/s00216-010-4055-3.
Pazos, Diego, Pauline Giannasi, Quentin Rossy, and Pierre Esseiva. 2013. "Combining Internet Monitoring Processes, Packaging and Isotopic Analyses to Determine the Market Structure: Example of Gamma Butyrolactone." Forensic Science International 230 (1–3): 29–36. https://doi.org/10.1016/j.forsciint.2013.02.033.
Pérez-Cova, Miriam, Carmen Bedia, Dwight R. Stoll, Romà Tauler, and Joaquim Jaumot. 2021. "MSroi: A Pre-Processing Tool for Mass Spectrometry-Based Studies." Chemometrics and Intelligent Laboratory Systems 215 (August): 104333. https://doi.org/10.1016/j.chemolab.2021.104333.
Pérez-Cova, Miriam, Joaquim Jaumot, and Romà Tauler. 2021. "Untangling Comprehensive Two-Dimensional Liquid Chromatography Data Sets Using Regions of Interest and Multivariate Curve Resolution Approaches." TrAC Trends in Analytical Chemistry 137 (April): 116207. https://doi.org/10.1016/j.trac.2021.116207.
Pérez-Cova, Miriam, Stefan Platikanov, Romà Tauler, and Joaquim Jaumot. 2022. "Quantification Strategies for Two-Dimensional Liquid Chromatography Datasets Using Regions of Interest and Multivariate Curve Resolution Approaches." Talanta 247 (May). https://doi.org/10.1016/j.talanta.2022.123586.
Pérez-Cova, Miriam, Romà Tauler, and Joaquim Jaumot. 2020. "Chemometrics in Comprehensive Two-Dimensional Liquid Chromatography: A Study of the Data Structure and Its Multilinear Behavior." Chemometrics and Intelligent Laboratory Systems 201 (March). https://doi.org/10.1016/j.chemolab.2020.104009.
Perez-Lopez, Carlos, Antoni Ginebreda, Montserrat Carrascal, Damiá Barcelò, Joaquin Abian, and Romà Tauler. 2021. "Non-Target Protein Analysis of Samples from Wastewater Treatment Plants Using the Regions of Interest-Multivariate Curve Resolution (ROIMCR) Chemometrics Method." Journal of Environmental Chemical Engineering, June, 105752. https://doi.org/10.1016/j.jece.2021.105752.
Perrault, Katelynn A., Pierre Hugues Stefanuto, Lena Dubois, Dries Cnuts, Veerle Rots, and Jean François Focant. 2016. "A New Approach for the Characterization of Organic Residues from Stone Tools Using GC×GC-TOFMS." Separations 3 (2): 8–10. https://doi.org/10.3390/separations3020016.

Pesesse, R., Pierre Hugues Stefanuto, F. Schleich, R. Louis, and J. F. Focant. 2019. “Multimodal Chemometric Approach for the Analysis of Human Exhaled Breath in Lung Cancer Patients by TD-GC × GC-TOFMS.” Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences 1114–1115 (December 2018): 146–53. https://doi.org/10.1016/j.jchromb.2019.01.029.

Phillips, J.B., R.B. Gaines, J. Blomberg, F. W. M. van der Wielen, Jean Marie D. Dimandja, V. Green, J. Granger, et al. 1999. “A Robust Thermal Modulator for Comprehensive Two-Dimensional Gas Chromatography.” HRC Journal of High Resolution Chromatography 22: 3–10. https://doi.org/10.1002/(SICI)1521-4168(19990101)22:1%3C3::AID-JHRC3%3E3.0.CO;2-U.

Pierce, Karisa M., Jamin C. Hoggard, Janiece L. Hope, Petrie M. Rainey, Andrew N. Hoofnagle, Rhona M. Jack, Bob W. Wright, and Robert E. Synovec. 2006. “Fisher Ratio Method Applied to Third-Order Separation Data to Identify Significant Chemical Components of Metabolite Extracts.” Analytical Chemistry 78 (14): 5068–75. https://doi.org/10.1021/ac0602625.

Pierce, Karisa M., Janiece L. Hope, Jamin C. Hoggard, and Robert E. Synovec. 2006. “A Principal Component Analysis Based Method to Discover Chemical Differences in Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry (GC × GC-TOFMS) Separations of Metabolites in Plant Samples.” Talanta 70 (4): 797–804. https://doi.org/10.1016/j.talanta.2006.01.038.

Pierce, Karisa M., Janiece L. Hope, Kevin J. Johnson, Bob W. Wright, and Robert E. Synovec. 2005. “Classification of Gasoline Data Obtained by Gas Chromatography Using a Piecewise Alignment Algorithm Combined with Feature Selection and Principal Component Analysis.” Journal of Chromatography A 1096 (1–2): 101–10. https://doi.org/10.1016/j.chroma.2005.04.078.

Pierce, Karisa M., Benjamin Kehimkar, Luke C. Marney, Jamin C. Hoggard, and Robert E. Synovec. 2012. “Review of Chemometric Analysis Techniques for Comprehensive Two Dimensional Separations Data.” Journal of Chromatography A 1255: 3–11. https://doi.org/10.1016/j.chroma.2012.05.050.

Pierce, Karisa M., Lianna F. Wood, Bob W. Wright, and Robert E. Synovec. 2005. “A Comprehensive Two-Dimensional Retention Time Alignment Algorithm to Enhance Chemometric Analysis of Comprehensive Two-Dimensional Separation Data.” Analytical Chemistry 77 (23): 7735–43. https://doi.org/10.1021/ac0511142.

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

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

Pirok, Bob W.J., Dwight R. Stoll, and Peter J. Schoenmakers. 2019. “Recent Developments in Two-Dimensional Liquid Chromatography: Fundamental Improvements for Practical Applications.” Analytical Chemistry 91 (1): 240–63. https://doi.org/10.1021/acs.analchem.8b04841.

Pluskal, Tomáš, Sandra Castillo, Alejandro Villar-Briones, and Matej Orešič. 2010. “MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data.” BMC Bioinformatics 11 (1): 395. https://doi.org/10.1186/1471-2105-11-395.

Pollo, Breno J., Guilherme L. Alexandrino, Fabio Augusto, and Leandro W. Hantao. 2018. “The Impact of Comprehensive Two-Dimensional Gas Chromatography on Oil & Gas Analysis: Recent Advances and Applications in Petroleum Industry.” TrAC - Trends in Analytical Chemistry 105: 202–17. https://doi.org/10.1016/j.trac.2018.05.007.

Poole, Colin F., and Salwa K. Poole. 2008. “Separation Characteristics of Wall-Coated Open-Tubular Columns for Gas Chromatography.” Journal of Chromatography A 1184 (1–2): 254–80. https://doi.org/10.1016/j.chroma.2007.07.028.

Pourasil, Roudabeh Sadat Moazeni, Joyce Cristale, Silvia Lacorte, and Romà Tauler. 2022. “Non-Targeted Gas Chromatography Orbitrap Mass Spectrometry Qualitative and Quantitative Analysis of Semi-Volatile Organic Compounds in Indoor Dust Using the Regions of Interest Multivariate Curve Resolution Chemometrics Procedure.” Journal of Chromatography A 1668: 462907. https://doi.org/10.1016/j.chroma.2022.462907.

Prebihalo, Sarah E., Kelsey L. Berrier, Chris E. Freye, H. Daniel Bahaghighat, Nicholas R. Moore, David K. Pinkerton, and Robert E. Synovec. 2018. “Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications.” Analytical Chemistry 90 (1): 505–32. https://doi.org/10.1021/acs.analchem.7b04226.

Prebihalo, Sarah E., Grant S. Ochoa, Kelsey L. Berrier, Kristen J. Skogerboe, Kenneth L. Cameron, Jesse R. Trump, Steven J. Svoboda, J. Kenneth Wickiser, and Robert E. Synovec. 2020. “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.” Analytical Chemistry 92 (23): 15526–33. https://doi.org/10.1021/acs.analchem.0c03456.

Prebihalo, Sarah E., David K. Pinkerton, and Robert E. Synovec. 2019. “Impact of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Experimental Design on Data Trilinearity and Parallel Factor Analysis Deconvolution.” Journal of Chromatography A, no. xxxx: 460368. https://doi.org/10.1016/j.chroma.2019.460368.

Purschke, Kirsten, Maryam Vosough, Juri Leonhardt, Markus Weber, and Torsten C. Schmidt. 2020. “Evaluation of Nontarget Long-Term LC-HRMS Time Series Data Using Multivariate Statistical Approaches.” Analytical Chemistry 92 (18): 12273–81. https://doi.org/10.1021/acs.analchem.0c01897.

Rajalahti, Tarja, Reidar Arneberg, Frode S. Berven, Kjell Morten Myhr, Rune J. Ulvik, and Olav M. Kvalheim. 2009. “Biomarker Discovery in Mass Spectral Profiles by Means of Selectivity Ratio Plot.” Chemometrics and Intelligent Laboratory Systems 95 (1): 35–48. https://doi.org/10.1016/j.chemolab.2008.08.004.

Rajalahti, Tarja, and Olav M. Kvalheim. 2011. “Multivariate Data Analysis in Pharmaceutics: A Tutorial Review.” International Journal of Pharmaceutics 417 (1–2): 280–90. https://doi.org/10.1016/j.ijpharm.2011.02.019.

Ralston-Hooper, Kimberly, Amber Hopf, Cheolhwan Oh, Xiang Zhang, Jiri Adamec, and Maria S. Sepúlveda. 2008. “Development of GCxGC/TOF-MS Metabolomics for Use in Ecotoxicological Studies with Invertebrates.” Aquatic Toxicology 88 (1): 48–52. https://doi.org/10.1016/j.aquatox.2008.03.002.

Reaser, Brooke C., Bob W. Wright, and Robert E. Synovec. 2017. “Using Receiver Operating Characteristic Curves to Optimize Discovery-Based Software with Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry.” Analytical Chemistry 89 (6): 3606–12. https://doi.org/10.1021/acs.analchem.6b04991.

Reichenbach, Stephen E., Xue Tian, Qingping Tao, Edward B. Ledford, Zhanpin Wu, and Oliver Fiehn. 2011. “Informatics for Cross-Sample Analysis with Comprehensive Two-Dimensional Gas Chromatography and High-Resolution Mass Spectrometry (GCxGC-HRMS).” Talanta 83 (4): 1279–88. https://doi.org/10.1016/j.talanta.2010.09.057.

Reid, Vanessa R., and Robert E. Synovec. 2008. “High-Speed Gas Chromatography: The Importance of Instrumentation Optimization and the Elimination of Extra-Column Band Broadening.” Talanta 76 (4): 703–17. https://doi.org/10.1016/j.talanta.2008.05.012.

Robards, K, P R Haddad, and P E Jackson. 2004. “2 - Theory of Chromatography.” In Principles and Practice of Modern Chromatographic Methods, edited by K Robards, P R Haddad, and P E Jackson, 35–74. Boston: Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-08-057178-2.50005-4.

Robards, Kevin, and Danielle Ryan. 2022. “Chapter 5 - High Performance Liquid Chromatography: Instrumentation and Techniques.” In Principles and Practice of Modern Chromatographic Methods (Second Edition), edited by Kevin Robards and Danielle Ryan, Second Edition, 247–82. Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-822096-2.00009-8.

Rosso, Marta Cialiè, Federico Stilo, Carlo Bicchi, Melanie Charron, Ginevra Rosso, Roberto Menta, Stephen E. Reichenbach, et al. 2021. “Combined Untargeted and Targeted Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography to Track Compositional Changes on Hazelnut Primary Metabolome during Roasting.” Applied Sciences 11 (January): 525. https://doi.org/10.3390/app11020525.

Rutan, Sarah C., Joe M. Davis, and Peter W. Carr. 2012. “Fractional Coverage Metrics Based on Ecological Home Range for Calculation of the Effective Peak Capacity in Comprehensive Two-Dimensional Separations.” Journal of Chromatography A 1255: 267–76. https://doi.org/10.1016/j.chroma.2011.12.061.

Saber, Eli. 2007. “Streak Detection in Mottled and Noisy Images.” Journal of Electronic Imaging 16 (4): 043005. https://doi.org/10.1117/1.2816444.

Sadeghi, Maryam, Tim K. Lee, David McLean, Harvey Lui, and M. Stella Atkins. 2013. “Detection and Analysis of Irregular Streaks in Dermoscopic Images of Skin Lesions.” IEEE Transactions on Medical Imaging 32 (5): 849–61. https://doi.org/10.1109/TMI.2013.2239307.

Savareear, Benjamin, Michał Brokl, Chris Wright, and Jean Francois Focant. 2017. “Thermal Desorption Comprehensive Two-Dimensional Gas Chromatography Coupled to Time of Flight Mass Spectrometry for Vapour Phase Mainstream Tobacco Smoke Analysis.” Journal of Chromatography A 1525: 126–37. https://doi.org/10.1016/j.chroma.2017.10.013.

Schenone, Nahuel Francisco, Lenka Vackova, and Alicia Fernández Cirelli. 2011. “Fish-Farming Water Quality and Environmental Concerns in Argentina: A Regional Approach.” Aquaculture International 19 (5): 855–63. https://doi.org/10.1007/s10499-010-9404-x.

Schollée, Jennifer E., Emma L. Schymanski, Sven E. Avak, Martin Loos, and Juliane Hollender. 2015. “Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic.” Analytical Chemistry 87 (24): 12121–29. https://doi.org/10.1021/acs.analchem.5b02905.

Schostack, Kathleen J., and Edmund R. Malinowski. 1989. “Preferred Set Selection by Iterative Key Set Factor Analysis.” Chemometrics and Intelligent Laboratory Systems 6 (1): 21–29. https://doi.org/10.1016/0169-7439(89)80062-0.

Schwaninger, Andrea E., Markus R. Meyer, Marilyn A. Huestis, and Hans H. Maurer. 2011. “Development and Validation of LC-HRMS and GC-NICI-MS Methods for Stereoselective

Determination of MDMA and Its Phase i and II Metabolites in Human Urine." Journal of Mass Spectrometry 46 (7): 603–14. https://doi.org/10.1002/jms.1929.
Schöneich, Sonia, Derrick V Gough, Timothy J. Trinklein, and Robert E Synovec. 2020. "Dynamic Pressure Gradient Modulation for Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry Detection." Journal of Chromatography A, 460982. https://doi.org/https://doi.org/10.1016/j.chroma.2020.460982.
Schöneich, Sonia, Grant S. Ochoa, Celina M. Monzón, and Robert E. Synovec. 2022. "Minimum Variance Optimized Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography / Mass Spectrometry Data: Study of the Pacu Fish Metabolome." Journal of Chromatography A 1667 (March): 462868. https://doi.org/10.1016/j.chroma.2022.462868.
Schöneich, Sonia, Timothy J. Trinklein, Cable G. Warren, and Robert E. Synovec. 2020. "A Systematic Investigation of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry with Dynamic Pressure Gradient Modulation for High Peak Capacity Separations." Analytica Chimica Acta 1134: 115–24. https://doi.org/10.1016/j.aca.2020.08.023.
Seeley, John V. 2012. "Recent Advances in Flow-Controlled Multidimensional Gas Chromatography." Journal of Chromatography A 1255: 24–37. https://doi.org/10.1016/j.chroma.2012.01.027.
Seeley, John V., Frederick Kramp, and Christine J. Hicks. 2000. "Comprehensive Two-Dimensional Gas Chromatography via Differential Flow Modulation." Analytical Chemistry 72 (18): 4346–52. https://doi.org/10.1021/ac000249z.
Seeley, John V., Nicole J. Micyus, Steven V. Bandurski, Stacy K. Seeley, and James D. McCurry. 2007. "Microfluidic Deans Switch for Comprehensive Two-Dimensional Gas Chromatography." Analytical Chemistry 79 (5): 1840–47. https://doi.org/10.1021/ac061881g.
Seeley, John V., Nicolaas E. Schimmel, and Stacy K. Seeley. 2018. "The Multi-Mode Modulator: A Versatile Fluidic Device for Two-Dimensional Gas Chromatography." Journal of Chromatography A 1536: 6–15. https://doi.org/10.1016/j.chroma.2017.06.030.
Seeley, John V., and Stacy K. Seeley. 2013. "Multidimensional Gas Chromatography: Fundamental Advances and New Applications." Analytical Chemistry 85 (2): 557–78. https://doi.org/10.1021/ac303195u.
Semard, Gaëlle, Mohamed Adahchour, and Jean-François Focant. 2009. "Chapter 2 Basic Instrumentation for GC×GC." Comprehensive Analytical Chemistry 55 (January): 15–48. https://doi.org/10.1016/S0166-526X(09)05502-0.
Shellie, Robert, Philip Marriott, and Alain Chaintreau. 2004. "Quantitation of Suspected Allergens in Fragrances (Part I): Evaluation of Comprehensive Two-Dimensional Gas Chromatography for Quality Control." Flavour and Fragrance Journal 19 (2): 91–98. https://doi.org/10.1002/ffj.1334.
Shen, Caiying, Jingli Xie, and Xueshu Xu. 2007. "The Components of Cuttlefish (Sepiella Maindroni de Rochebruns) Oil." Food Chemistry 102 (1): 210–14. https://doi.org/10.1016/j.foodchem.2006.05.014.
Siegler, W. Christopher, Brian D. Fitz, Jamin C. Hoggard, and Robert E. Synovec. 2011. "Experimental Study of the Quantitative Precision for Valve-Based Comprehensive Two-Dimensional Gas Chromatography." Analytical Chemistry 83 (13): 5190–96. https://doi.org/10.1021/ac200302b.
Sinanian, Melanie M., Daniel W. Cook, Sarah C. Rutan, and Dayanjan S. Wijesinghe. 2016. "Multivariate Curve Resolution-Alternating Least Squares Analysis of High-Resolution Liquid Chromatography-Mass Spectrometry Data." Analytical Chemistry 88 (22): 11092–99. https://doi.org/10.1021/acs.analchem.6b03116.
Sinha, Amanda E., Carlos G. Fraga, Bryan J. Prazen, and Robert E. Synovec. 2004. "Trilinear Chemometric Analysis of Two-Dimensional Comprehensive Gas Chromatography-Time-of-

Flight Mass Spectrometry Data." Journal of Chromatography A 1027 (1–2): 269–77. https://doi.org/10.1016/j.chroma.2003.08.081.
Sinha, Amanda E., Janiece L. Hope, Bryan J. Prazen, Carlos G. Fraga, Erik J. Nilsson, and Robert E. Synovec. 2004. "Multivariate Selectivity as a Metric for Evaluating Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry Subjected to Chemometric Peak Deconvolution." Journal of Chromatography A 1056 (1-2 SPEC.ISS.): 145–54. https://doi.org/10.1016/j.chroma.2004.06.110.
Sinkov, Nikolai A., and James J. Harynuk. 2011. "Cluster Resolution: A Metric for Automated, Objective and Optimized Feature Selection in Chemometric Modeling." Talanta 83 (4): 1079–87. https://doi.org/10.1016/j.talanta.2010.10.025.
Smith, Author H, R Gnanadesikan, and J B Hughes. 1962. "Multivariate Analysis of Variance (MANOVA)." Biometrics 18 (1): 22–41.
Smith, Colin A., Elizabeth J. Want, Grace O'Maille, Ruben Abagyan, and Gary Siuzdak. 2006. "XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification." Analytical Chemistry 78 (3): 779–87. https://doi.org/10.1021/ac051437y.
Soaudy, Mohamed R., Eman Y. Mohammady, Mohamed A. Ashry, Marwa M. Ali, Nasr M. Ahmed, Mahmoud H. Hegab, Hoda A.S. Elgarhy, Ehab R. El-Haroun, and Mohamed S. Hassaan. 2021. "Possibility Mitigation of Cold Stress in Nile Tilapia under Biofloc System by Dietary Propylene Glycol: Performance Feeding Status, Immune, Physiological Responses and Transcriptional Response of Delta-9-Desaturase Gene." Aquaculture 538 (February): 736519. https://doi.org/10.1016/j.aquaculture.2021.736519.
Souard, Florence, Cédric Delporte, Piet Stoffelen, Etienne A. Thévenot, Nausicaa Noret, Bastien Dauvergne, Jean-Michel Kauffmann, Pierre Van Antwerpen, and Caroline Stévigny. 2018. "Metabolomics Fingerprint of Coffee Species Determined by Untargeted-Profiling Study Using LC-HRMS." Food Chemistry 245 (April): 603–12. https://doi.org/10.1016/j.foodchem.2017.10.022.
Stahle, Lars, and Svante Wold. 1989. "Analysis of Variance (ANOVA)." Chemometrics and Intelligent Laboratory Systems 6 (4): 259–72. https://doi.org/10.1016/0169-7439(89)80095-4.
Stefanuto, Pierre Hugues, Katelynn A. Perrault, Lena M. Dubois, Benjamin L'Homme, Catherine Allen, Caitriona Loughnane, Nobuo Ochiai, and Jean François Focant. 2017. "Advanced Method Optimization for Volatile Aroma Profiling of Beer Using Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry." Journal of Chromatography A 1507: 45–52. https://doi.org/10.1016/j.chroma.2017.05.064.
Stefanuto, Pierre Hugues, Delphine Zanella, Joeri Vercammen, Monique Henket, Florence Schleich, Renaud Louis, and Jean François Focant. 2020. "Multimodal Combination of GC × GC-HRTOFMS and SIFT-MS for Asthma Phenotyping Using Exhaled Breath." Scientific Reports 10 (1): 1–11. https://doi.org/10.1038/s41598-020-73408-2.
Stein, Stephen E, and Donald R Scott. 1994. "Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification." Journal of the American Society for Mass Spectrometry 5: 859–66. https://doi.org/10.1016/1044-0305(94)87009-8.
Stilo, Federico, Carlo Bicchi, Albert Robbat, Stephen E. Reichenbach, and Chiara Cordero. 2021. "Untargeted Approaches in Food-Omics: The Potential of Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry." TrAC - Trends in Analytical Chemistry 135: 116162. https://doi.org/10.1016/j.trac.2020.116162.

Sudol, Paige E., Micaela Galletta, Peter Q. Tranchida, Mariosimone Zoccali, Luigi Mondello, and Robert E. Synovec. 2022. “Untargeted Profiling and Differentiation of Geographical Variants of Wine Samples Using Headspace Solid-Phase Microextraction Flow-Modulated Comprehensive Two-Dimensional Gas Chromatography with the Support of Tile-Based Fisher Ratio Analysis.” Journal of Chromatography A 1662: 462735. https://doi.org/10.1016/j.chroma.2021.462735.

Sudol, Paige E., Derrick V. Gough, Sarah E. Prebihalo, and Robert E. Synovec. 2020. “Impact of Data Bin Size on the Classification of Diesel Fuels Using Comprehensive Two-Dimensional Gas Chromatography with Principal Component Analysis.” Talanta 206 (July 2019): 120239. https://doi.org/10.1016/j.talanta.2019.120239.

Sudol, Paige E., Grant S. Ochoa, and Robert E. Synovec. 2021. “Investigation of the Limit of Discovery Using Tile-Based Fisher Ratio Analysis with Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry.” Journal of Chromatography A 1644 (May): 462092. https://doi.org/10.1016/j.chroma.2021.462092.

Sudol, Paige E., Karisa M. Pierce, Sarah E Prebihalo, Kristen J Skogerboe, Bob W. Wright, and Robert E. Synovec. 2020. “Development of Gas Chromatographic Pattern Recognition and Classification Tools for Compliance and Forensic Analyses of Fuels: A Review.” Analytica Chimica Acta, July. https://doi.org/10.1016/j.aca.2020.07.027.

Sudol, Paige E., Sonia Schöneich, and Robert E. Synovec. 2022. “Principal Component Analysis of Comprehensive Three-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data.” Journal of Chromatography Open 2 (February): 100043. https://doi.org/10.1016/j.jcoa.2022.100043.

Szymańska, Ewa, Jan Gerretzen, Jasper Engel, Brigitte Geurts, Lionel Blanchet, and Lutgarde M.C. Buydens. 2015. “Chemometrics and Qualitative Analysis Have a Vibrant Relationship.” TrAC - Trends in Analytical Chemistry 69: 34–51. https://doi.org/10.1016/j.trac.2015.02.015.

Tauler, Romà. 1995. “Multivariate Curve Resolution Applied to Second Order Data.” Chemometrics and Intelligent Laboratory Systems 30 (95): 133–46.

Tauler, Romà, Age Smilde, and Bruce Kowalski. 1995. “Selectivity, Local Rank, Three-Way Data Analysis and Ambiguity in Multivariate Curve Resolution.” Journal of Chemometrics 9 (1): 31–58. https://doi.org/10.1002/cem.1180090105.

Tessarolo, Nathalia S., Luciana R.M. dos Santos, Raphael S.F. Silva, and Débora A. Azevedo. 2013. “Chemical Characterization of Bio-Oils Using Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry.” Journal of Chromatography A 1279 (2013): 68–75. https://doi.org/10.1016/j.chroma.2012.12.052.

Toustrup, Kasper, Brita Singers Sørensen, Marianne Nordsmark, Morten Busk, Carsten Wiuf, Jan Alsner, and Jens Overgaard. 2011. “Development of a Hypoxia Gene Expression Classifier with Predictive Impact for Hypoxic Modification of Radiotherapy in Head and Neck Cancer.” Cancer Research 71 (17): 5923–31. https://doi.org/10.1158/0008-5472.CAN-11-1182.

Tran, Tin C., Graham A. Logan, Emmanuelle Grosjean, James Harynuk, Danielle Ryan, and Philip Marriott. 2006. “Comparison of Column Phase Configurations for Comprehensive Two Dimensional Gas Chromatographic Analysis of Crude Oil and Bitumen.” Organic Geochemistry 37 (9): 1190–94. https://doi.org/10.1016/j.orggeochem.2006.05.006.

Tranchida, Peter Q., Flavio A. Franchina, Mariosimone Zoccali, Ivana Bonaccorsi, Francesco Cacciola, and Luigi Mondello. 2013. “A Direct Sensitivity Comparison between Flow-Modulated Comprehensive 2D and 1D GC in Untargeted and Targeted MS-Based Experiments.” Journal of Separation Science 36 (17): 2746–52. https://doi.org/10.1002/jssc.201300423.

Tranchida, Peter Q., Flavio A Franchina, Paola Dugo, and Luigi Mondello. 2014. “Flow-Modulation Low-Pressure Comprehensive Two-Dimensional Gas Chromatography.” Journal of Chromatography A 1372: 236–44. https://doi.org/10.1016/j.chroma.2014.10.097.
Tranchida, Peter Q., Flavio A Franchina, Paola Dugo, and Luigi Mondello. 2014. “Use of Greatly-Reduced Gas Flows in Flow-Modulated Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry.” Journal of Chromatography A 1359: 271–76.
Tranchida, Peter Q., Giorgia Purcaro, Paola Dugo, and Luigi Mondello. 2011. “Modulators for Comprehensive Two-Dimensional Gas Chromatography.” TrAC - Trends in Analytical Chemistry 30 (9): 1437–61. https://doi.org/10.1016/j.trac.2011.06.010.
Tranchida, Peter Q., Giorgia Purcaro, Mariarosa Maimone, and Luigi Mondello. 2016. “Impact of Comprehensive Two-Dimensional Gas Chromatography with Mass Spectrometry on Food Analysis.” Journal of Separation Science 39 (1): 149–61. https://doi.org/10.1002/jssc.201500379.
Tranchida, Peter Q., Giorgia Purcaro, Alessandro Visco, Lanfranco Conte, Paola Dugo, Peter Dawes, and Luigi Mondello. 2011. “A Flexible Loop-Type Flow Modulator for Comprehensive Two-Dimensional Gas Chromatography.” Journal of Chromatography A 1218 (21): 3140–45. https://doi.org/10.1016/j.chroma.2010.11.082.
Tranchida, Peter Q., Simona Salivo, Flavio A. Franchina, and Luigi Mondello. 2015. “Flow-Modulated Comprehensive Two-Dimensional Gas Chromatography Combined with a High-Resolution Time-of-Flight Mass Spectrometer: A Proof-of-Principle Study.” Analytical Chemistry 87 (5): 2925–30. https://doi.org/10.1021/ac5044175.
Trinklein, Timothy J., Derrick V. Gough, Cable G. Warren, Grant S. Ochoa, and Robert E. Synovec. 2020. “Dynamic Pressure Gradient Modulation for Comprehensive Two-Dimensional Gas Chromatography.” Journal of Chromatography A 1609: 460488. https://doi.org/10.1016/j.chroma.2019.460488.
Trinklein, Timothy J., Jiaxin Jiang, and Robert E. Synovec. 2022. “Profiling Olefins in Gasoline by Bromination Using GC×GC-TOFMS Followed by Discovery-Based Comparative Analysis.” Analytical Chemistry 94 (26): 9407–14. https://doi.org/10.1021/acs.analchem.2c01549.
Trinklein, Timothy J., Sarah E. Prebihalo, Cable G. Warren, Grant S. Ochoa, and Robert E. Synovec. 2020. “Discovery-Based Analysis and Quantification for Comprehensive Three-Dimensional Gas Chromatography Flame Ionization Detection Data.” Journal of Chromatography A 1623: 461190. https://doi.org/10.1016/j.chroma.2020.461190.
Trinklein, Timothy J., Sonia Schöneich, Paige E. Sudol, Cable G. Warren, Derrick V. Gough, and Robert E. Synovec. 2020. “Total-Transfer Comprehensive Three-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry.” Journal of Chromatography A 1634: 461654. https://doi.org/10.1016/j.chroma.2020.461654.
Trinklein, Timothy J., and Robert E. Synovec. 2022. “Simulating Comprehensive Two-Dimensional Gas Chromatography Mass Spectrometry Data with Realistic Run-to-Run Shifting to Evaluate the Robustness of Tile-Based Fisher Ratio Analysis.” Journal of Chromatography A 1677 (August): 463321. https://doi.org/10.1016/j.chroma.2022.463321.
Trinklein, Timothy J., Cable G. Warren, and Robert E. Synovec. 2021. “Determination of the Signal-To-Noise Ratio Enhancement in Comprehensive Three-Dimensional Gas Chromatography.” Analytical Chemistry 93 (24): 8526–35. https://doi.org/10.1021/acs.analchem.1c01190.
Urban, Pawel L. 2016. “Clarifying Misconceptions about Mass and Concentration Sensitivity.” Journal of Chemical Education 93 (6): 984–87. https://doi.org/10.1021/acs.jchemed.5b00986.

Valladão, Gustavo Moraes Ramos, Sílvia Umeda Gallani, and Fabiana Pilarski. 2018. “South American Fish for Continental Aquaculture.” Reviews in Aquaculture 10 (2): 351–69. https://doi.org/10.1111/raq.12164.

Veriotti, T., M. McGuigan, and R. Sacks. 2001. “Pulsed Flow Modulation for High-Speed GC Using a Pressure-Tunable Column Ensemble.” Analytical Chemistry 73 (2): 279–85. https://doi.org/10.1021/ac000665j.

Ward, Joe H. 1963. “Hierarchical Grouping to Optimize an Objective Function.” Journal of the American Statistical Association 58 (301): 236–44. https://doi.org/10.1080/01621459.1963.10500845.

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

Westerhuis, Johan A., Huub C.J. Hoefsloot, Suzanne Smit, Daniel J. Vis, Age K. Smilde, Ewoud J.J. Velzen, John P.M. Duijnhoven, and Ferdi A. Dorsten. 2008. “Assessment of PLSDA Cross Validation.” Metabolomics 4 (1): 81–89. https://doi.org/10.1007/s11306-007-0099-6.

Wichaphon, Jetsada, Chaufah Thongthai, Apinya Assavanig, and Sittiwat Lertsiri. 2012. “Volatile Aroma Components of Thai Fish Sauce in Relation to Product Categorization.” Flavour and Fragrance Journal 27 (2): 149–56. https://doi.org/10.1002/ffj.2095.

Willse, Alan, Anne M. Belcher, George Preti, Jon H. Wahl, Miranda Thresher, Peter Yang, Kunio Yamazaki, and Gary K. Beauchamp. 2005. “Identification of Major Histocompatibility Complex-Regulated Body Odorants by Statistical Analysis of a Comparative Gas Chromatography/Mass Spectrometry Experiment.” Analytical Chemistry 77 (8): 2348–61. https://doi.org/10.1021/ac048711t.

Wilson, Ryan B., Jamin C. Hoggard, and Robert E. Synovec. 2013. “High Throughput Analysis of Atmospheric Volatile Organic Compounds by Thermal Injection - Isothermal Gas Chromatography - Time-of-Flight Mass Spectrometry.” Talanta 103: 95–102. https://doi.org/10.1016/j.talanta.2012.10.013.

Wilson, Ryan B., W. Christopher Siegler, Jamin C. Hoggard, Brian D. Fitz, Jeremy S. Nadeau, and Robert E. Synovec. 2011. “Achieving High Peak Capacity Production for Gas Chromatography and Comprehensive Two-Dimensional Gas Chromatography by Minimizing off-Column Peak Broadening.” Journal of Chromatography A 1218 (21): 3130–39. https://doi.org/10.1016/j.chroma.2010.12.108.

Windig, W., C. E. Heckler, F. A. Agblevor, and R. J. Evans. 1992. “Self-Modeling Mixture Analysis of Categorized Pyrolysis Mass Spectral Data with the SIMPLISMA Approach.” Chemometrics and Intelligent Laboratory Systems 14 (1–3): 195–207. https://doi.org/10.1016/0169-7439(92)80104-C.

Wohlfahrt, Sebastian, Michael Fischer, Janos Varga, Mohammad Reza Saraji-Bozorgzad, Georg Matuschek, Thomas Denner, and Ralf Zimmermann. 2016. “Dual-Stage Consumable-Free Thermal Modulator for the Hyphenation of Thermal Analysis, Gas Chromatography, and Mass Spectrometry.” Analytical Chemistry 88 (1): 640–44. https://doi.org/10.1021/acs.analchem.5b04183.

Wold, Svante, Kim Esbensen, and Paul Geladi. 1987. “Principal Component Analysis.” Chemometrics and Intelligent Laboratory Systems 2 (1): 37–52. https://doi.org/https://doi.org/10.1016/0169-7439(87)80084-9.

Xu, Fengguo, Li Zou, and Choon Nam Ong. 2009. “Multiorigination of Chromatographic Peaks in Derivatized GC/MS Metabolomics: A Confounder That Influences Metabolic Pathway

Interpretation." Journal of Proteome Research 8 (12): 5657–65. https://doi.org/10.1021/pr900738b.
Yamanaka, Hunter R., Cynthia Cheung, Jireh S. Mendoza, Danson J. Oliva, Kealina Elzey-Aberilla, and Katelynn A. Perrault. 2021. "Pilot Study on Exhaled Breath Analysis for a Healthy Adult Population in Hawaii." Molecules 26 (12): 3726. https://doi.org/10.3390/molecules26123726.
Yang, Bo Tian, Bin Wen, Yu Ji, Qin Wang, Hao Ran Zhang, Yuan Zhang, Jian Zhong Gao, and Zai Zhong Chen. 2021. "Comparative Metabolomics Analysis of Pigmentary and Structural Coloration in Discus Fish (Symphysodon Haraldi)." Journal of Proteomics 233 (September 2020). https://doi.org/10.1016/j.jprot.2020.104085.
Zanella, Delphine, Adèle Henin, Steven Mascrez, Pierre Hugues Stefanuto, Flavio Antonio Franchina, Jean François Focant, and Giorgia Purcaro. 2022. "Comprehensive Two-Dimensional Gas Chromatographic Platforms Comparison for Exhaled Breath Metabolites Analysis." Journal of Separation Science 45 (18): 3542–55. https://doi.org/10.1002/jssc.202200164.
Zeng, Zhongda, Jia Li, Helmut M. Hugel, Guowang Xu, and Philip J. Marriott. 2014. "Interpretation of Comprehensive Two-Dimensional Gas Chromatography Data Using Advanced Chemometrics." TrAC - Trends in Analytical Chemistry 53: 150–66. https://doi.org/10.1016/j.trac.2013.08.009.
Zhang, Dabao, Xiaodong Huang, Fred E. Regnier, and Min Zhang. 2008. "Two-Dimensional Correlation Optimized Warping Algorithm for Aligning GCxGC-MS Data." Analytical Chemistry 80 (8): 2664–71. https://doi.org/10.1021/ac7024317.
Zou, Yun, Meriem Gaida, Flavio A Franchina, Pierre Hugues Stefanuto, and Jean-François Focant. 2022. "Distinguishing between Decaffeinated and Regular Coffee by HS-SPME-GC×GC-TOFMS, Chemometrics, and Machine Learning." Molecules 27 (6): 1806. https://doi.org/10.3390/molecules27061806.
Zushi, Yasuyuki, Jonas Gros, Qingping Tao, Stephen E. Reichenbach, Shunji Hashimoto, and J. Samuel Arey. 2017a. "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." Journal of Chromatography A 1508: 121–29. https://doi.org/10.1016/j.chroma.2017.05.065.

Appendix A

This appendix is reproduced from the Electronic Supplementary content of S. Schöneich, C.N. Cain, C.E. Freye, R.E. Synovec, Optimization of Parameters for ROI Data Compression for Nontargeted Analyses Using LC–HRMS, Anal. Chem. 95 (2022) 1513–1521.

Table A.1. Eighteen pesticides spiked into the soil samples at nominal concentrations of 0.1, 0.5, 1, 10, 25, and 50 ppb to demonstrate the workflow for applying F-ratio analysis to LC-HRMS data. The compound name and retention time (t_R) is provided for each spiked pesticide. The theoretical *m/z*, identification *m/z* along with the ppm difference (Δppm) between the theoretical *m/z* and ROI *m/z* used to identify these compounds are reported with the adduct based on the data obtained using the optimal ROI workflow parameters.

Pesticide ID	t_R (min)	Theoretical *m/z*	Identification *m/z* (Δ ppm)	**Adduct**
terbacil	5.3	215.0593	215.0604 (5.0)	$[M-H]^-$
lenacil	5.6	233.1295	233.1292 (1.4)	$[M-H]^-$
propargite	5.9	409.1691	409.1769 (19.2)	$[M+Hac-H]^-$
flutriafol	6.0	300.0954	300.0966 (3.9)	$[M-H]^-$
hexazinone	6.2	251.1513	251.1512 (0.5)	$[M-H]^-$
triadimenol	6.8	354.1227	354.1245 (5.2)	$[M+Hac-H]^-$
paclobutrazol	6.9	352.1434	352.1440 (1.8)	$[M+Hac-H]^-$
fludioxonil	7.1	247.0324	247.0337 (5.2)	$[M-H]^-$
tricyclazole	7.1	248.0500	248.0499 (0.2)	$[M+Hac-H]^-$
fenarimol	7.2	329.0254	329.0260 (1.8)	$[M-H]^-$
myclobutanil	7.3	347.1281	347.1295 (4.2)	$[M+Hac-H]^-$
triadimefon	7.4	292.0858	292.0858 (0.1)	$[M-H]^-$
tebuconazole	7.5	366.1590	366.1657 (18.4)	$[M+Hac-H]^-$
procymidone	7.8	282.0094	282.0103 (3.1)	$[M-H]^-$
fipronil	8.2	434.9314	434.9318 (0.9)	$[M-H]^-$
flusilazole	8.3	314.0930	314.0932 (0.6)	$[M-H]^-$
triflumizole	8.6	344.0785	344.0783 (0.7)	$[M-H]^-$
pyrimethanil	9.7	258.1248	258.1298 (19.6)	$[M+Hac-H]^-$

Experimental details

Acetonitrile (HPLC Plus) and deionized water (HPLC Plus) were obtained from Sigma-Aldrich. Ammonium acetate was obtained from VWR and was diluted to 10 mM with deionized water. Soil samples were obtained from the area surrounding the Los Alamos National Laboratory (New Mexico, USA) and were initially sifted with 1.2 mm mesh followed by 500 µm mesh to remove debris. A soil sample stock solution was prepared by dissolving 36 g of the sifted soil in 54 mL acetonitrile and filtered through a 25 mm diameter Whatman Puradisc 25TF filter (GE Healthcare, Marlborough, MA, USA) with a 0.45 µm PFTE membrane in a polypropylene housing. For the evaluation of the workflow presented herein, GC Multiresidue Pesticide Standard #5 obtained from Restek was serially diluted with the stock solution to obtain nominal concentrations of 0.1 ppb, 0.5 ppb, 1 ppb, 10 ppb, 25 ppb, and 50 ppb for each spiked analyte, to comprise the six sample classes. The identities of spiked pesticides discoverable in the soil matrix are provided in Table 1 with their retention time and the *m/z* used for identification based on the data obtained with the optimal ROI parameters. Identification of spiked pesticides was confirmed by comparison to a neat standard mix and library matching with available spectra from MassBank.

All data were collected on a Shimadzu UHPLC system (Shimadzu, Japan) coupled to a SCIEX X500R QTOF (SCIEX, Framingham, MA, USA) controlled by SCIEX OS 1.6.1 software. This UHPLC system is equipped with two binary pumps (LC30-AD), a degasser (DGU-30A), column oven (CTO-30A), and autosampler (SIL-30A). The QTOF was operated with electrospray ionization (ESI) in the negative mode with an ion spray voltage of -4.5 kV, temperature of 275 °C, curtain gas 25 psi, ion source gas #1 at 35 psi, ion source gas #2 at 40 psi, declustering potential of -50 V, and CAD gas 3 arbitrary units (a.u). Spectra were collected using *m/z* 100-1200 at an accumulation time of 0.373 s.

A Kinetex (Phenomenex, Torrance, CA, USA) C_{18} column (100 mm × 2.1 mm, 1.7 µm) was used for all soil sample separations. A 5 µL sample volume was injected with the column oven set to 40 °C and a 0.3 mL/min flow rate. The mobile phase was composed of (A) 10 mM ammonium acetate and (B) acetonitrile. The gradient program applied was initially isocratic at 0% B at 0 min, followed by a linear gradient to 100% B from 0 to 12 min, and finally isocratic at 100% B for 3 min (12-15 min). For each of the six sample classes, five replicates were obtained for 30 total chromatograms.

ROI user-defined parameter details

For a single analyte, with Figure 2.2A in the main paper representing the raw data, only *m/z* with signal above the threshold are extracted. In this case, only the five most intense *m/z* pass the signal threshold and the peak profiles are truncated, however if the signal threshold was doubled only the single most intense *m/z* would pass the ROI criteria. Furthermore, focusing on the most intense *m/z* in Figure 2.2A, it is visible in Figure 2.2B that 40 *m/z* and signal intensity pairs indicated by the purple circles passed the signal threshold for the minimum number of consecutive retention times, though five of these *m/z* are not within the bounds of the *m/z* error (dashed blue lines). Ultimately it is the 35 *m/z* and signal intensity pairs that make up the signal of the purple trace and the average *m/z* of these 35 pairs is the final ROI *m/z*. Since the ROI *m/z* is the average of the raw *m/z* values, this parameter is critical to ensure the ROI *m/z* can be used for accurate analyte identification. The minimum number of retention times to be considered an ROI is typically based on the average peak width in the chromatogram. These parameters all work simultaneously to reduce the data to the chemically relevant information while conserving the mass accuracy of the detector.

F-ratio analysis methodology

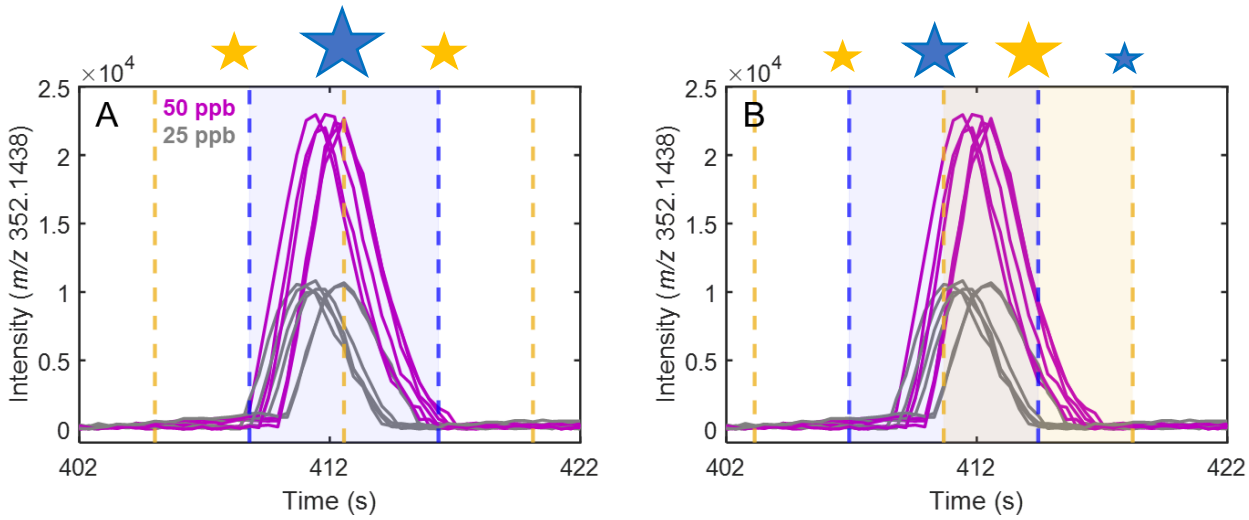


Figure A.1. Demonstration of tiling used for tile-based F-ratio analysis. (A) An ideal example of how the tiles capture the peak area when one of tiles captures the entire analyte peak. (B) An example of how tiles capture the peak area when no tiles ideally capture the entire peak area, but still discover the analyte as a hit.

For tile-based F-ratio analysis, the entire chromatogram is sliced into tiles of user-defined dimensions with the aim to capture the entire analyte signal within a tile for each peak. While dividing the chromatogram into these tiles does not guarantee that each peak is captured ideally within a tile and therefore are split between tiles, a second grid is applied that shifts the tiles by one half of a tile width. It is of note that it is not necessary for a peak to be ideally captured in the center of a tile for the F-ratio analysis software to discover the analyte as a "hit." Figure A.1A provides one example of how the F-ratio tiling schemes aims to capture the entire signal of an analyte peak. A tile size of 7.5 s was selected as it best captures the peak area of the analyte without including too much noise. A smaller tile size of 3.7 s was not selected as it introduced many spurious false positives and redundant hits, making mining the hit list tedious. A larger tile size of 15 s was not ideal as it discovered fewer spiked pesticides than the 7.5 s tile. Due to the ample selectivity of the detector, tile size was easily optimized as pure *m/z* for each analyte were readily available for F-ratio analysis.

In this example, one of the tiles highlighted in blue captures the entire peak signal. The tiles in gold, shifted by half of a tile dimension, partially capture the peak in both tiles, however the magnitude of the F-ratios calculated for these two tiles relative to that in the blue centered tile is much lower (magnitude of F-ratios are indicated by the stars). Figure A.1B demonstrates an analyte that does not get captured completely by any of the tiles, but still gets discovered in the hit list. In this case, both the gold and blue tiles that capture the most peak area will have higher F-ratios calculated than the other tiles with less peak area captured. While the same analyte gets discovered in multiple tiles using both grids, these redundant hits are algorithmically removed. After the location of the hit is determined, the redundant hit removal algorithm is applied to the hit list ranked by descending F-ratios. Starting with the first hit, the retention time is compared to the rest of the hits on the list. Any other hits with lower F-ratios and retention times that fall within a user-defined time window are determined to be the same analyte and are removed from the hit list. Therefore, taking Figure A.1A as an example, only the middle blue star with the largest F-ratio would be retained in the hit list for this analyte, ideally reducing the hit list to only one hit per analyte peak. To ensure the analytes are ideally ranked, the tiles are re-centered around the retention time of each analyte and F-ratios are calculated again.

F-ratio hit list using initial parameters for ROI compression (all spikes)

Table A.2. F-ratio hit list using initial parameters. Hit list was obtained using ^{MVO}F-ratio analysis for the comparison of the spiked classes using parameters recommended in the literature for the ROI methodology as an initial attempt to discover the pesticides spiked into the samples. The hit list is organized with Hit#, compound identity, the calculated F-ratio using the top F-ratio *m/z* (experimental *m/z*), identification *m/z* that was used for confirming analyte identification with the Δppm (average Δppm = 22 ppm) and adduct. Using these parameters, a smaller data set with only 212 ROI extracted for a range of 100-1200 *m/z* collected and only 3 spiked pesticides were discovered in the F-ratio hit list (all p-values < 0.05).

Hit#	ID	F-ratio	Experimental *m/z*	Identification *m/z* (Δppm)	**Adduct**
1	fludioxonil	608	247.0343	247.0343 (7.6)	$[M-H]^-$
2	tricyclazole	203	248.0361	248.0361 (56.0)	$[M+Hac-H]^-$
3	fipronil	200	436.9290	434.9322 (1.8)	$[M-H]^-$

F-ratio hit list using optimal parameters (all spikes)

Table A.3. F-ratio hit list using optimal parameters. Tile-based F-ratio results are provided for the comparison of all spiked soil samples processed by the ROI workflow using parameters that were determined to be optimal. The same tile size as the other F-ratio analyses was used (7.5 s). The hit list is ranked according to descending F-ratio with the hit number in the first column followed by the compound identity, F-ratio value calculated using the top F-ratio *m/z* (experimental *m/z*), *m/z* used for confirming compound identity with Δppm followed by the adduct.

Hit#	ID	F-ratio	Experimental *m/z*	Identification *m/z* (Δppm)	Adduct
1	fludioxonil	8377	247.0337	247.0337 (5.2)	$[M-H]^-$
2	fenarimol	7709	329.0260	329.0260 (1.8)	$[M-H]^-$
3	lenacil	2522	233.1292	233.1292 (1.4)	$[M-H]^-$
4	terbacil	2185	215.0604	215.0604 (5.0)	$[M-H]^-$
5	flutriafol	1697	204.0593	300.0966 (3.9)	$[M-H]^-$
6	tricyclazole	1288	248.0357	248.0499 (0.2)	$[M+Hac-H]^-$
7	paclobutrazol	1271	352.1440	352.1440 (1.8)	$[M+Hac-H]^-$
8	triadimefon	735	292.0858	292.0858 (0.1)	$[M-H]^-$
10	triadimenol	627	298.1527	354.1245 (5.2)	$[M+Hac-H]^-$
11	fipronil	432	305.1951	434.9318 (0.9)	$[M-H]^-$
12	tebuconazole	392	311.1663	366.1657 (18.4)	$[M+Hac-H]^-$
13	triflumizole	385	344.0783	344.0783 (0.7)	$[M-H]^-$
14	myclobutanil	336	347.1295	347.1295 (4.2)	$[M+Hac-H]^-$
15	hexazinone	280	295.1930	251.1512 (0.5)	$[M-H]^-$
17	propargite	85	291.1610	409.1769 (19.2)	$[M+Hac-H]^-$
18	flusilazole	81	314.4703	314.0932 (0.6)	$[M-H]^-$
19	procymidone	70	217.0622	282.0103 (3.1)	$[M-H]^-$
23	pyrimethanil	40	305.1951	258.1298 (19.6)	$[M+Hac-H]^-$

Calibration curves of representative analytes require low signal threshold

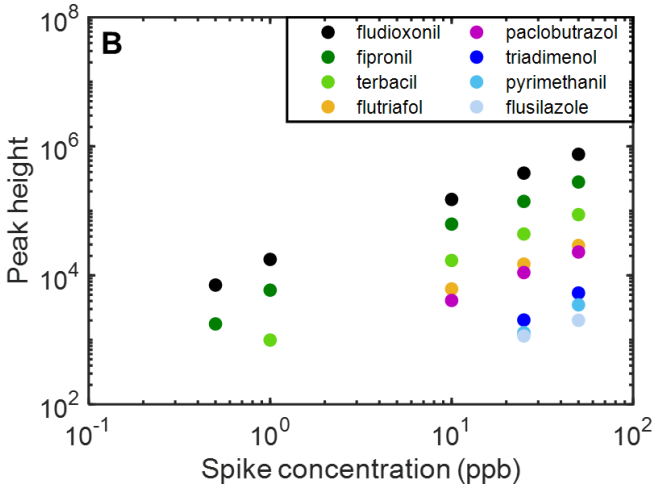


Figure A.2. Calibration curves of representative analytes. Using the optimal signal threshold of 250, the average signal using the identification *m/z* for all replicates per class of a subset of eight spiked pesticides was calculated for all six spike sample classes.

Using the optimal signal threshold of 250, the average peak height was calculated for a subset of eight analytes of differing sensitivity in all six sample classes using the identification *m/z* provided in Table A.1. Not all pesticides are discoverable in all sample classes by the ROI procedure and therefore by F-ratio analysis, with the peak heights spanning a wide range both between and within a single sample class, which may be due to differences in ionization efficiencies and ion suppression. The analytes with low signal in the higher concentration sample classes present a challenge not only to discovery by ROI methodology, but to the F-ratio analysis as proper tile selection is critical to ensure noise does not obscure the analyte signal. As observed previously, without applying a lower than recommended threshold, pyrimethanil and flusilazole and the [M+Hac-H]- *m/z* for triadimenol would not be discovered as ROIs or by F-ratio analysis. This could be expected for analytes present in trace levels in samples, however it is important to ensure that the data is truly reduced to a manageable size while providing the most important and chemically relevant information for the intended study. With a known analyte spiked into samples at a defined concentration, in addition to computational time, Δppm, and number of

ROIs, the expected peak profile ratio reflecting the spike concentration in the samples can be used as confirmation of effective data reduction that accurately represents the raw data.

Table A.4. F-ratio hit list using initial parameters (25 ppb vs 50 ppb). F-ratio results for the 25 ppb and 50 ppb class comparison using the initial parameters for ROI data compression. The hit list is organized in the same manner as previous hit lists.

Hit#	ID	F-ratio	Experimental *m/z*	Identification *m/z* (Δppm)	Adduct
1	fludioxonil	4.80E+04	247.0343	247.0343 (7.6)	$[M-H]^-$
2	tricyclazole	1.60E+03	248.0361	248.0361 (56.0)	$[M+Hac-H]^-$
3	fipronil	1.30E+03	434.9323	434.9323 (1.9)	$[M-H]^-$

Table A.5. F-ratio hit list using optimal m/z error, initial signal threshold (25 ppb vs 50 ppb). The F-ratio hit list using the suggested signal threshold, 0.1% of the maximum intensity, and the optimal *m/z* error of ±0.003 Da in the super-augmentation step using the same tile size as in Table A.2. The hit list is organized with hit#, compound identity, the F-ratio calculated for the top F-ratio *m/z* (experimental *m/z*), the identification *m/z* used to confirm compound identity with corresponding Δppm followed by the adduct.

Hit#	ID	F-ratio	Experimental *m/z*	Identification *m/z* (Δppm)	Adduct
1	fludioxonil	6589.7	247.0336	247.0336 (4.7)	$[M-H]^-$
2	tricyclazole	290.7	248.0361	248.0361 (56.0)	$[M+Hac-H]^-$
3	fipronil	72.7	436.9279	434.9305 (2.2)	$[M-H]^-$
6	tebuconazole	22.3	325.1826	366.2369 (212.9)	$[M+Hac-H]^-$
10	triadimenol	10.3	297.1518	--	--
11	paclobutrazol	7.5	297.1518	--	--

To evaluate the workflow after optimizing the admissible mass deviation, F-ratio analysis was applied to the data using the recommended signal threshold (0.1% of maximum intensity) and the optimal *m/z* error of ±0.003 Da for the 50 ppb and 25 ppb class comparison. The hit list is organized in the same manner as Table A.4 with Hit#, compound identity, F-ratio value calculated using the top F-ratio *m/z* (Experimental *m/z*), the identification *m/z* used for confidence in analyte identification with corresponding Δppm and adduct. With this initial optimization step, the number of true positives discovered doubled, all statistically significant (p-value < 0.05), and F-ratios decreased greatly across the top three hits. The discovery of more spiked pesticides could be explained by adding signal contributions from fewer *m/z* of the raw data to obtain ROI with corresponding experimental *m/z* that better represent the true peak profiles and theoretical *m/z*. Interestingly, for the two true positives with lowest F-ratios (Hit# 10 and Hit# 11) there was no identification *m/z* discovered by inspection of expected $[M-H]^-$, $[M+Hac-H]^-$, $[M-CO_2-H]^-$, or $[M-H_2O-H]^-$ *m/z* for triadimenol and paclobutrazol and were therefore primarily identified based on retention time. A possible explanation for this may be that the signal threshold is not high enough for signal from these analytes to be considered ROIs for

the identification *m/z*. Additionally, the Δppm between some of the experimental ROI *m/z* and theoretical *m/z* were greater resulting in an average absolute Δppm of 69 ppm (21 ppm excluding tebuconazole) indicating that to obtain the optimal results the *m/z* error should not be optimized alone, rather it should be optimized in conjunction with the signal threshold.

Table A.6. F-ratio hit list using optimal signal threshold, initial m/z error (25 ppb vs 50 ppb). The F-ratio hit list is provided using a signal threshold of 250 and using an admissible mass deviation of ±0.05 Da and the same tile size as in Table 2.2. The hit list is ranked by descending F-ratios with reported Hit#, compound identity, F-ratio value (all statistically significant with p-values < 0.05), experimental *m/z* used to calculate the F-ratios, the identification *m/z* used for hit identification with Δppm, and finally the adduct.

Hit#	ID	F-ratio	Experimental *m/z*	Identification *m/z* (Δppm)	Adduct
1	terbacil	1.15E+05	215.0661	215.0661 (31.4)	$[M-H]^-$
3	tebuconazole	2.03E+04	308.2068	366.1659 (18.9)	$[M+Hac-H]^-$
4	lenacil	1.57E+04	233.1564	233.1564 (115.4)	$[M-H]^-$
6	flutriafol	1.55E+04	204.0486	300.0901 (17.8)	$[M-H]^-$
7	tricyclazole	1.47E+04	248.0393	248.0393 (42.8)	$[M+Hac-H]^-$
8	fludioxonil	1.28E+04	247.0425	247.0425 (40.6)	$[M-H]^-$
9	triadimefon	1.10E+04	292.0907	292.0907 (16.8)	$[M-H]^-$
10	paclobutrazol	1.06E+04	352.1355	352.1355 (22.2)	$[M+Hac-H]^-$
12	triflumizole	8.84E+03	344.0722	344.0770 (4.3)	$[M-H]^-$
13	fipronil	7.86E+03	329.9730	434.9371 (13.1)	$[M-H]^-$
23	fenarimol	2.96E+03	331.0077	329.0272 (5.3)	$[M-H]^-$
35	triadimenol	693	354.1295	354.1295 (19.4)	$[M+Hac-H]^-$
61	myclobutanil	208	379.2432	347.1267 (4.0)	$[M+Hac-H]^-$
97	hexazinone	31	249.1721	251.1323 (75.6)	$[M-H]^-$

The hit list obtained from F-ratio analysis of the ROI data obtained using the optimal signal threshold and the recommended *m/z* error is provided in Table A.6, organized in the same manner as Table A.4. In comparison to the hit list obtained using optimal parameters, one less spiked pesticide was discovered for a total of 14 pesticides discovered (11 of these within the top 30 hits on the list). Though a greater number of pesticides were discovered, many of the hit numbers were much higher than those in Table A.4 with many more spurious hits in the hit list resulting in the pesticides being less discoverable. Additionally, the calculated Δppm for the identification *m/z* is as high as 115 ppm (average of 31 ppm) with numerous analytes with higher Δppm relative to Table A.4, indicating the importance of using a smaller admissible mass deviation to improve identification of analytes in a non-targeted study. It is evident that lowering

the signal threshold was responsible for discovering a larger amount of pesticides than simply optimizing the admissible mass deviation, however, to maximize analyte discovery and confident identification in non-targeted studies, it is advised to optimize both *m/z* error and signal threshold parameters simultaneously with spiked standards to ensure that the data is accurately represented for the intended analysis, especially for low concentration analytes of interest.

Table A.7. F-ratio hit list using optimal parameters (25 ppb vs 50 ppb). Tile-based F-ratio results are provided for the comparison of the 50 ppb and 25 ppb spiked soil samples processed by the ROI workflow using parameters that were determined to be optimal. The same tile size as the other F-ratio analyses was used (7.5 s). The hit list is ranked according to descending F-ratio with the hit number in the first column followed by the compound identity, F-ratio value calculated using the top F-ratio *m/z* (experimental m/z), *m/z* used for confirming compound identity with Δppm followed by the adduct.

Hit#	ID	F-ratio	Experimental *m/z*	Identification *m/z* (Δppm)	**Adduct**
1	fipronil	11751	399.9584	434.9312 (0.5)	$[M-H]^-$
2	myclobutanil	8368	347.1289	347.1289 (2.6)	$[M+Hac-H]^-$
3	fludioxonil	7694	247.0333	247.0333 (3.6)	$[M-H]^-$
4	tricyclazole	7694	249.0390	248.0501 (0.5)	$[M+Hac-H]^-$
5	triadimefon	7176	235.0160	292.0852 (2.1)	$[M-H]^-$
7	tebuconazole	3020	366.1660	366.1660 (19.2)	$[M+Hac-H]^-$
8	terbacil	2878	217.0691	215.0610 (7.8)	$[M-H]^-$
9	fenarimol	2074	331.0172	329.0262 (2.4)	$[M-H]^-$
10	hexazinone	1540	295.9131	251.1512 (0.5)	$[M-H]^-$
11	triflumizole	1388	276.0418	344.0826 (11.8)	$[M-H]^-$
12	triadimenol	993	354.1240	354.1240 (3.7)	$[M+Hac-H]^-$
13	flutriafol	846	301.0989	300.0967 (4.2)	$[M-H]^-$
14	lenacil	628	233.1294	233.1294 (0.5)	$[M-H]^-$
15	paclobutrazol	537	352.1438	352.1438 (1.3)	$[M+Hac-H]^-$
16	procymidone	488	217.0691	282.0104 (3.5)	$[M-H]^-$
17	propargite	410	303.1591	409.1774 (20.4)	$[M+Hac-H]^-$
18	pyrimethanil	332	258.1295	258.1295 (18.2)	$[M+Hac-H]^-$
28	flusilazole	103	326.1850	314.0990 (18.7)	$[M-H]^-$

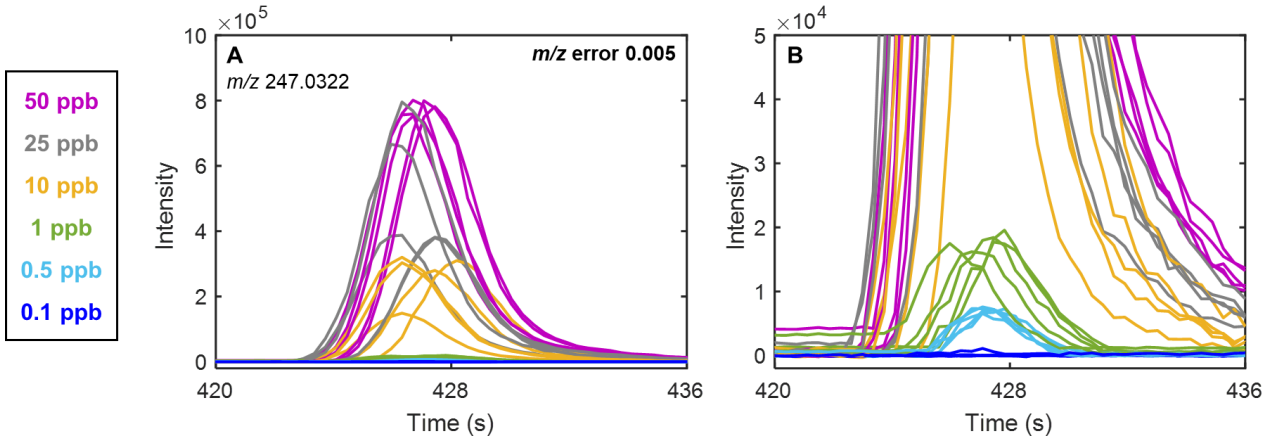


Figure A.3. Effect of tolerant m/z error on data compression by ROI. Further investigation of the ROI data produced from Figure 2.6B, illustrates the issue of using a higher than optimal *m/z* error that is too tolerant (±0.005 Da). (A) The overlay of all replicates for all sample classes of the fludioxonil peak is provided with a zoom-in of the three lowest spike concentration classes (B) using *m/z* 247.0322. While inspection of (B) indicates that the lowest three sample classes are easily distinguishable with relative signal that is consistent with their concentrations, several of the 25 ppb replicates are indistinguishable from the 50 ppb replicates and similarly the remaining 25 ppb replicates are indistinguishable from the 10 ppb replicates based on peak heights.

An overlay of all sample classes for fludioxonil using *m/z* 247.0322 is provided in Figure A.3A for visual interpretation. Notably, two of the 25 ppb replicates are indistinguishable from the 50 ppb replicates and the remaining 3 replicates of the 25 ppb sample class are indistinguishable from a majority of the 10 ppb sample class. The lower concentration sample classes are provided in Figure A.3B and appear to be more representative of the true class comparisons. These observations reinforce the interpretation of the poor sample clustering in Figure 2.6B. Without a priori knowledge of class membership, one could mistakenly interpret the data to only have five sample classes rather than six. With the signals across classes no longer distinguishing, F-ratio analysis would become more challenging as the high within class variance of the 25 ppb replicates and lower between-class variance would result in a lower standard F-ratio, making this hit less discoverable or class-distinguishing.

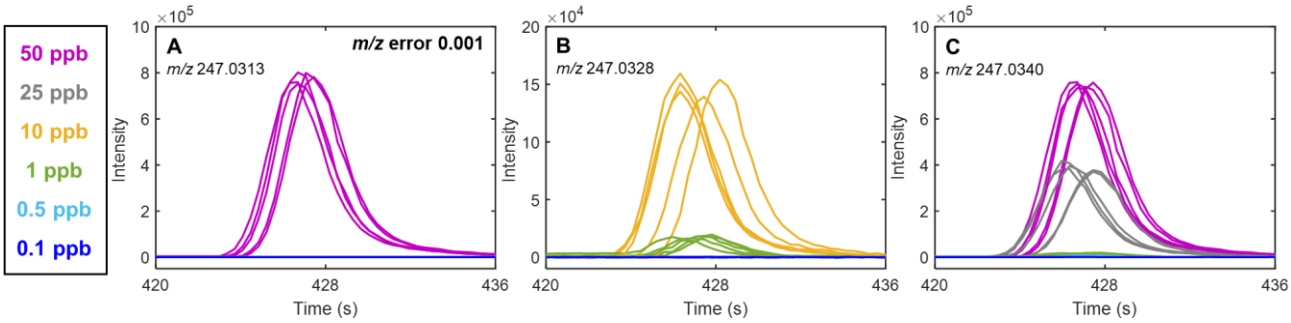


Figure A.4. Effect of strict m/z error on data compression by ROI. Further investigation of the ROI data produced from Figure 2.6D, illustrates the issue of using an *m/z* error that is too restrictive (±0.001 Da).Using the most sensitive analyte, fludioxonil (theoretical *m/z* 247.0324), it becomes evident that the analyte signal gets split among *m/z* 247.0313 (A), *m/z* 247.0328 (B), and *m/z* 247.0340 (C) that are all within 7 ppm of the theoretical $[M-H]^-$ *m/z*. The ROI in (A) only captures the signal for the 50 ppb class while the ROI in (B) captures signal for the 10 ppb and 1 ppb samples and the ROI in (C) captures the 50 ppb, 25 ppb, 1 ppb, and 0.5 ppb sample signal. This indicates that an admissible mass deviation as low as ±0.001 Da is not optimal for this data set and would be problematic for F-ratio analysis.

The overlay of all sample classes is provided for *m/z* 247.0313 (Δppm = 4.6 ppm) in Figure A.4A, *m/z* 247.0328 (Δppm = 1.5 ppm) in Figure A.4B, and *m/z* 247.0340 (Δppm = 6.4 ppm) in Figure A.4C. In Figure A.4A fludioxonil is discovered as an ROI only in the 50 ppb sample class, whereas in Figure A.4B it is considered an ROI only in the 10 ppb and 1 ppb classes, and in Figure A.4C it is an ROI in all but the 10 ppb sample class. Though the number of ROIs discovered is greater than the number using a larger *m/z* error, the ROI data produced is not ideal for F-ratio analysis as it calculates F-ratio values across samples per *m/z* and relies on signal for a given analyte to be present at the *m/z* across all replicates. Any analyte native to the sample with split signal would be discovered as a false positive by F-ratio and make true positive discovery more challenging for non-targeted analyses. In addition, the true positives could be discovered across many *m/z* with low Δppm and result in inaccurate quantification using techniques such as S-ratio analysis in subsequent steps. In fact, F-ratio analysis using this *m/z* error would be successful in discovering fludioxonil as a class-distinguishing analyte, however it would introduce many redundant hits that do not correctly represent the data.

Appendix B

This chapter was reproduced from S. Schöneich[†], T.J. Trinklein[†], C.G. Warren, R.E. Synovec, A systematic investigation of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry with dynamic pressure gradient modulation for high peak capacity separations, Anal. Chim. Acta 1134 (2020) 115–124.

[†] These authors contributed equally to this work

Pulse valve assembly and operation

The pulse valve is interfaced via an in-house machined face plate which enables a screw-tight and leak free connection between the valve poppet and the T-union via a 5 cm long connection tubing (Figs 1B in main manuscript). The in-house made face plate (shown in darker grey) is made of stainless steel of equal diameter and thickness to the bottom plate of the pulse valve. A zero-dead volume internal union was then machined onto the bottom of the face plate (Valco instruments, ZU.5T). The connection tubing is then attached using a 1/32” screw and stainless-steel ferrule and inserted slightly into the pulse valve body plate.

The valve is controlled via an in-house developed LABView vi and driven by an in-house made valve driver. Using LABView enables the user to synchronize the valve sequence with sample injection via Agilent’s remote out connection. The valve driver can be modified to deliver only the minimum required hit and hold voltage which reduces the wattage delivered to valve. This is important when operating the valve at extremely fast modulation periods (P_M < 250 ms) for long runs (>10 min), when valve overheating can occur. Additionally, a computer fan is used to cool the valve when performing runs with very fast P_M. When operating at longer modulation periods, (P_M > 250 ms), valve overheating has not been observed to be an issue. Indeed, we have observed no valve failures when operating at P_M = 2 s over several months.

Figure B.1. Images of the valve face plate fitted with the connection tubing (A) and the assembled pulse valve (B).

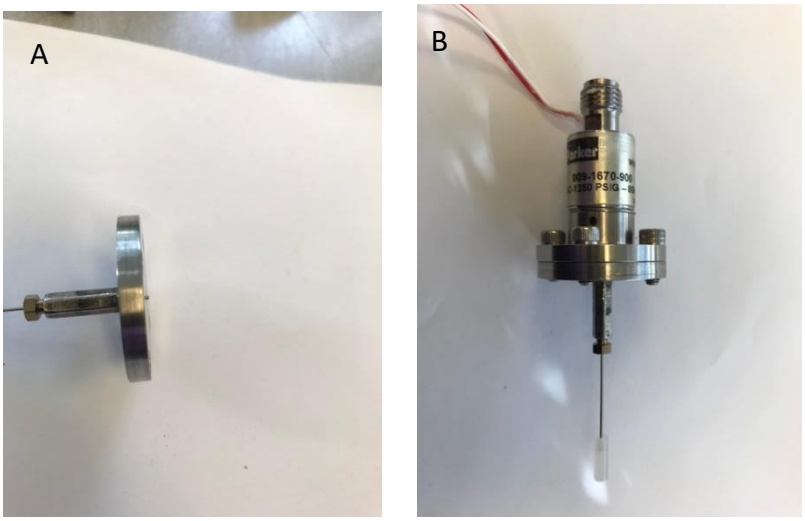


Table B.1. List of the 116-component test mixture

Alkanes	**Boiling Point**	**Vendor**	**Alcohols**	**Boiling Point**	**Vendor**
hexane	69	Sigma-Aldrich	1-propanol	97	Sigma-Aldrich
heptane	98	Sigma-Aldrich	2-butanol	117	Sigma-Aldrich
octane	126	Sigma-Aldrich	1-pentanol	137	Sigma-Aldrich
nonane	151	Sigma-Aldrich	2-pentanol	158	Sigma-Aldrich
decane	174	Sigma-Aldrich	hexyl alcohol	158	Sigma-Aldrich
undecane	196	Sigma-Aldrich	2-heptanol	160	Fluka
dodecane	216	Sigma-Aldrich	1-octanol	194	Sigma-Aldrich
tridecane	235	Fluka	1-nonanol	215	Sigma-Aldrich
tetradecane	254	Fluka	1-decanol	229	Sigma-Aldrich
pentadecane	271	Alfa-Aesar	1-tetradecanol	289	Sigma-Aldrich
hexadecane	287	Sigma-Aldrich	1-octadecanol	210	Sigma-Aldrich
heptadecane	302	Sigma-Aldrich	1-eicosanol	372	Sigma-Aldrich
octadecane	316	Fluka	benzyl alcohol	205	Sigma-Aldrich
nonadecane	330	Sigma-Aldrich	2-ethyl-1-hexanol	185	Fluka
eicosane	343	Sigma-Aldrich	**Esters**	**Boiling Point**	**Vendor**
henicosane	356	Sigma-Aldrich	ethyl formate	54	Sigma-Aldrich
docosane	370	Sigma-Aldrich	methyl decanoate	224	Eastman Chemicals
tricosane	380	Sigma-Aldrich	methyl caprylate	193	Sigma-Aldrich
tetracosane	391	Sigma-Aldrich	methyl salicylate	223	Alfa-Aesar
pentacosane	401	Sigma-Aldrich	ethyl salicylate	233	Alfa-Aesar
hexacosane	412	Sigma-Aldrich	methyl laurate	262	Sigma-Aldrich
heptacostane	422	Sigma-Aldrich	methly caproate	151	Sigma-Aldrich
octacosane	431	Sigma-Aldrich	diethyl phthalate	299	Sigma-Aldrich
nonacosane	440	Sigma-Aldrich	**Ketones**	**Boiling Point**	**Vendor**
triacontane	449	Sigma-Aldrich	2-butanone	80	Sigma-Aldrich
hentriacontane	458	Sigma-Aldrich	2-pentanone	102	Sigma-Aldrich
dotriacontane	467	Sigma-Aldrich	3-hexanone	128	Sigma-Aldrich
titriacontane	474	Sigma-Aldrich	2-hexanone	128	Sigma-Aldrich
pentatriacontane	491	Sigma-Aldrich	2-heptanone	152	Sigma-Aldrich
hexatriacontane	498	Sigma-Aldrich	3-heptanone	141	Sigma-Aldrich
heptatriacontane	505	Sigma-Aldrich	3-octanone	167	Sigma-Aldrich
octatriacontane	512	Sigma-Aldrich	2-decanone	209	Sigma-Aldrich
nonatriacontane	518	Sigma-Aldrich	2-undecanone	231	Sigma-Aldrich

tetracontane	525	Sigma-Aldrich	2-dodecanone	245	Sigma-Aldrich
pristane	296	Sigma-Aldrich	**Aromatics**	**Boiling Point**	**Vendor**
Halogenated Alkanes	**Boiling Point**	**Vendor**	benzene	80	Fischer
1,5-dichloropentane	179	Sigma-Aldrich	toluene	111	Sigma-Aldrich
1-chlorohexane	135	Sigma-Aldrich	3-ethyltoluene	161	Sigma-Aldrich
1-bromohexane	155	Sigma-Aldrich	4-ethyltoluene	162	Fluka
1-bromoheptane	179	Sigma-Aldrich	mesitylene	165	Sigma-Aldrich
1-bromooctane	200	Sigma-Aldrich	ethylbenzene	136	Sigma-Aldrich
1-chlorobutane	78	Sigma-Aldrich	butylbenzene	183	Sigma-Aldrich
1,1,1-trichloroethane	74	Sigma-Aldrich	isobutylbenzene	170	Sigma-Aldrich
1,2-dichloroethane	84	Sigma-Aldrich	tert-butyl benzene	167	Sigma-Aldrich
carbon tetrachloride	77	Sigma-Aldrich	propylbenzene	159	Sigma-Aldrich
Cyclics	**Boiling Point**	**Vendor**	1-ethylnapthelene	260	Sigma-Aldrich
methylcyclopentane	72	Fluka	bromobenzene	155	Sigma-Aldrich
cyclohexane	81	Sigma-Aldrich	cyclohexylbenzene	239	Sigma-Aldrich
cyclooctane	150	Sigma-Aldrich	diphenylmethane	265	Sigma-Aldrich
butylcyclohexane	181	Sigma-Aldrich	p-xylene	189	Sigma-Aldrich
bicyclohexyl	227	Sigma-Aldrich	o-xylene	145	Sigma-Aldrich
2,2,4-trimethylpentane	99	Sigma-Aldrich	m-xylene	128	Sigma-Aldrich
Alkenes	**Boiling Point**	**Vendor**	1,2,4-trimethylbenzene	169	Sigma-Aldrich
1-hexene	63	Sigma-Aldrich	anisole	154	Sigma-Aldrich
cyclohexene	83	Sigma-Aldrich	dibutyl phthalate	340	Sigma-Aldrich
dodecene	214	Sigma-Aldrich	a-terpineol (90%)	219	Sigma-Aldrich
1-undecene	194	Fluka	**PAHs**	**Boiling Point**	**Vendor**
Alkynes	**Boiling Point**	**Vendor**	fluoranthene	375	Supelco
1-hexyne	71	Sigma-Aldrich	fluorene	295	Supelco
1-heptyne	109	Sigma-Aldrich	chrysene	448	Supelco
1-nonyne	151	Sigma-Aldrich	pyrene	404	Supelco
5-decyne	177	Sigma-Aldrich	anthracene	340	Supelco

Sample preparation and specific analysis conditions for the complex samples

Diesel Fuel

A diesel sample was sourced from a local fueling station. The oven temperature was held at 50 °C for 5 min, then ramped at 6 °C/min to 300 °C, where it was held for 2.5 min. The inlet was operated at constant flow mode at 1.40 ml/min, though the average flow rate was estimated at 0.15 ml/min on ^{1}D due to the DPGM effect on the resulting ^{1}D flow rate. The P_{aux} applied to the pulse valve was 296.5 kPa and was immediately ramped at 3.6 kPa/min to 447.3 kPa where it was held for 2.5 min. A 1 µl sample was injected at a 10:1 split ratio.

Cow serum

The cow serum was derivatized as described previously to provide trimethylsilation and methoximation [J. Chromatogr. A 1218 (2011) 1899–1906]. Briefly, 0.1 ml aliquots of room temperature cow serum were transferred to 2 ml microcentrifuge tubes and combined with 0.9 ml of an 8:1 methanol:water solution. The tubes were cooled on ice for 10 min, and then homogenized for 10 min. The samples were cooled again for 10 min, and then centrifuged at 15,000 G for 10 min. Next, 0.2 ml aliquots of the supernatant were transferred to 300 µl GC vial inserts and evaporated for 1 hr using a SpeedVac SVC 100. Dried samples were immediately stored at -4 °C. Prior to derivatization, the samples were again evaporated for 1 hr. A 30 µl aliquot of 20 mg/ml methoxyamine hydrochloride in ACS grade pyridine (prepared the same day) was added to each dry sample extract and incubated at 30 °C for 90 min. Next, a 70 µl aliquot of MSTFA with 0.1% TMCS (Thermo Scientific, Waltham, MA) solution was introduced followed by incubation at 60 °C for 90 min. Samples were analyzed immediately after completion of the derivatization procedure. One µl samples were injected in splitless mode.

SPME of river water headspace

Samples of contaminated river water containing sediment were collected from Ravenna Creek in Seattle, Washington, USA. A ~4 ml sample of river water was placed in a crimp vial without filtration and equilibrated at 80 °C for 30 min prior to extraction. River water headspace was sampled via SPME with a 65 µm polydimethylsiloxane / divinylbenzene (Supelco, PA, USA) fiber in a manual fiber holder apparatus (Supelco, Bellefonte, USA) without stirring. The fiber was conditioned at 250 °C for 30 min with a GC inlet flow of 0.5 ml/min before exposure to the water headspace for 2 hr. The fiber was desorbed in the inlet at 250 °C for 5 min.

SPME of coffee headspace

A 2 g sample of ground coffee was placed into a headspace vial and equilibrated for 30 min. A 65 µm polydimethylsiloxane / divinylbenzene fiber (Supelco, PA, USA) was conditioned at 250 °C for 30 min with a GC inlet flow of 0.5 ml/min and then exposed to the coffee headspace for 30 min. The fiber was desorbed in the inlet at 250 °C for 5 min.

Table B.2. List of analytes indicated in Fig. 5.2B
List of the 10 representative analytes labeled in Fig. 5.2B used to calculate the ^{2}D peak capacity, ^2n_c, as shown in Fig. 5.2C.

Peak number	Analyte	*m/z* used	^1t_R (min), ^2t_R (ms)	^2W_b (ms)
1	octane	57	12.1, 265	102
2	cyclooctane	41	14.6, 430	113
3	1-chlorohexane	91	13.0, 550	116
4	1-bromoheptane	135	16.7, 665	125
5	p-xylene	91	13.9, 880	133
6	chlorohexylbenzene	91	21.4, 1265	134
7	2-undecanone	58	22.3, 1025	165
8	2-ethyl-1-hexanol	55	16.4, 1535	175
9	1-octanol	41	17.2, 1715	176
10	1-pentanol	42	11.1, 1950	221

Figures B.2(a-j). Study of modulator induced band broadening on 1D peak widths. Figures for each of the 10 analytes labeled in Fig. 5.2B showing the Gaussian curve fit.

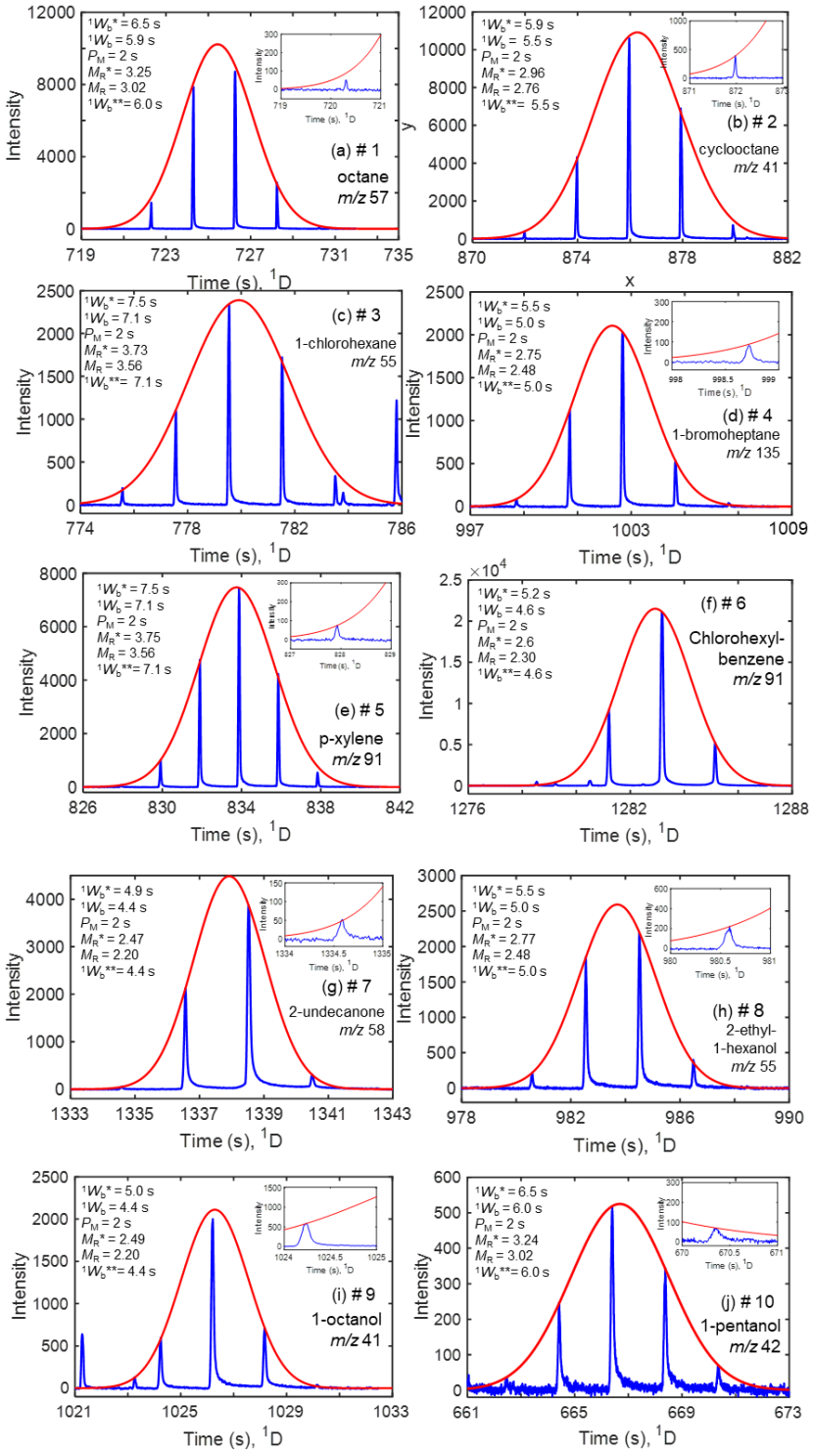
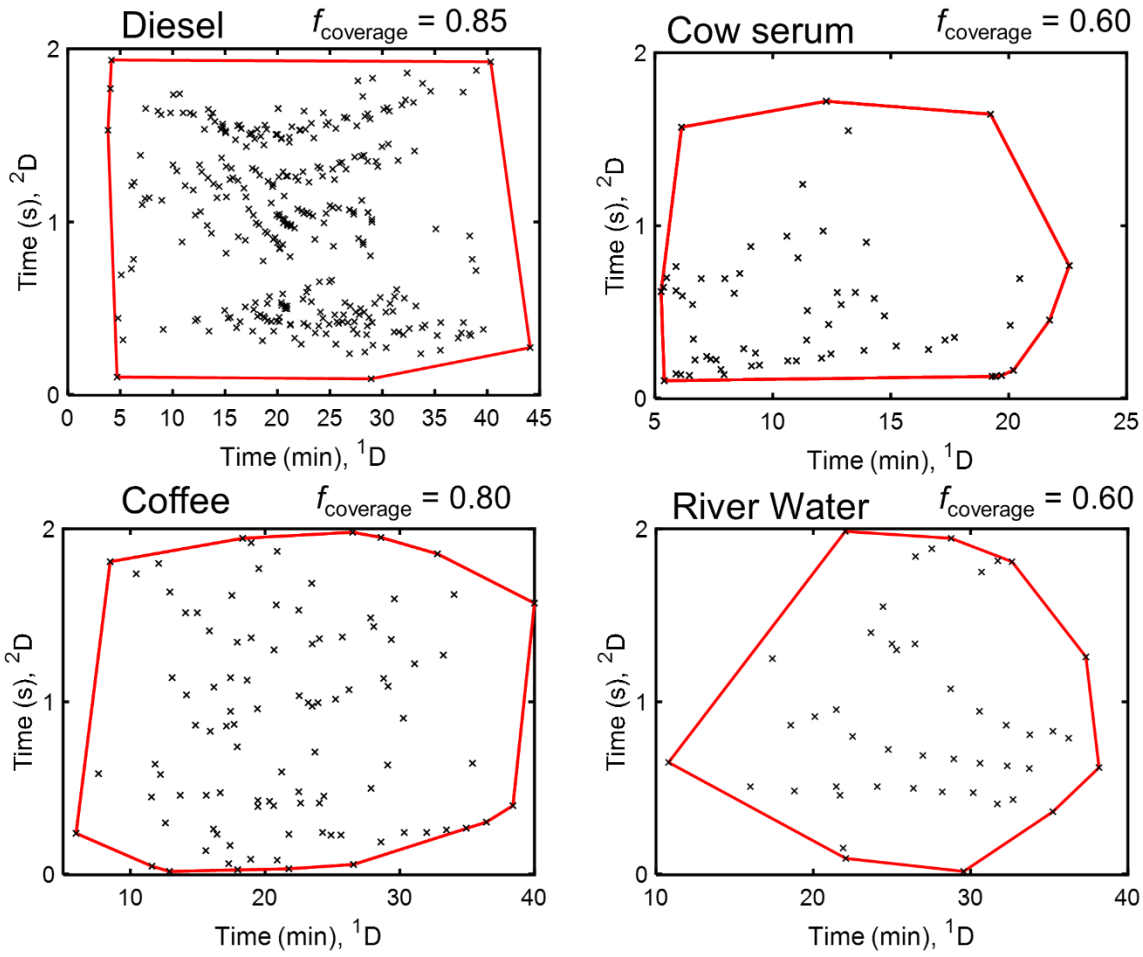


Figure B.3. Evaluation of $f_{coverage}$ for the complex samples. The fractional space covered for each of the samples is outlined in the red trace and analytes are represented by the × symbols with the $f_{coverage}$ reported above each panel.



Appendix C

This appendix was reproduced from S. Schöneich, G.S. Ochoa, C.M. Monzón, R.E. Synovec, Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography / mass spectrometry data: Study of the pacu fish metabolome, J. Chromatogr. A. 1667 (2022) 462868.

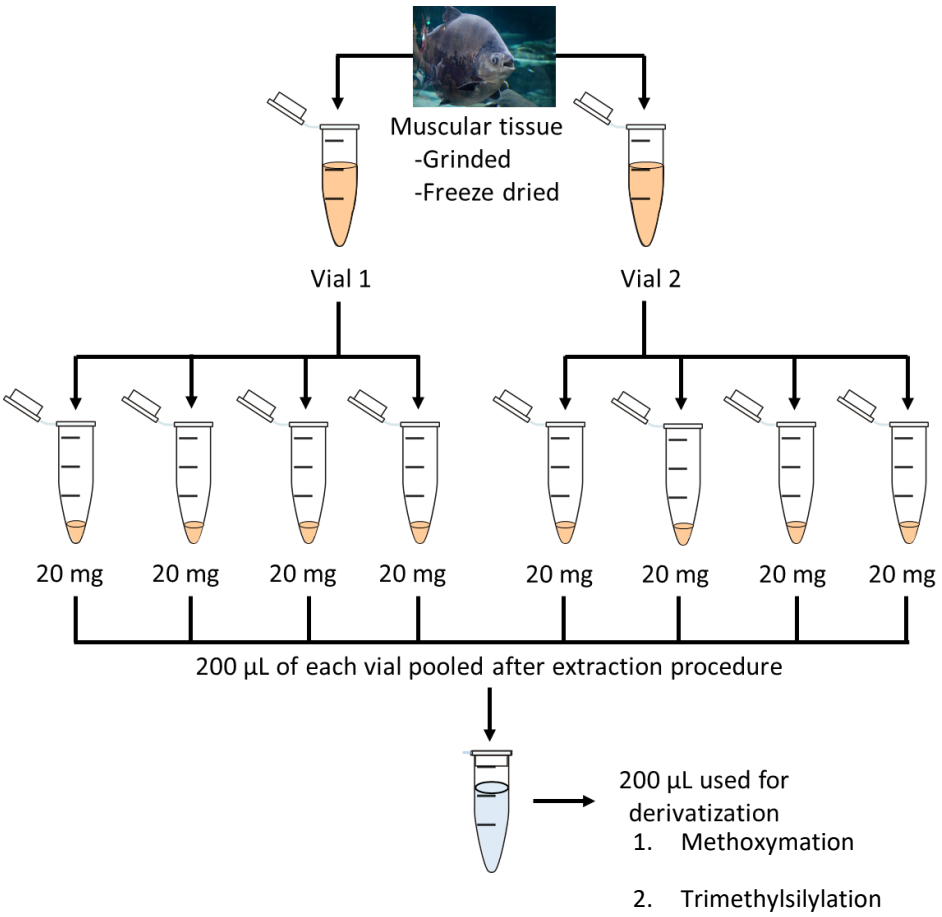


Figure C.1. Flowchart of the sample preparation steps. For each fish, 2 vials containing 5 g of freeze-dried fish tissue was available. To reduce sample heterogeneity, 4 extracts weighing 20 mg were pooled from each vial per fish. After the extraction procedure, 200 µL of each extract was pooled (1.6 mL) and 200 µL of the pooled extracts were used for derivatization.

Table C.1. The 29 metabolites spiked into the farmed pacu fish for the initial evaluation of the artifact removal and redundant hit removal algorithms is provided with retention times and corresponding signal using *m/z* 73 in the neat standard mixture. Retention times were used to verify identification of these analytes in the spiked fish samples.

Compound name	**^1t_R (min)**	**^2t_R (s)**	**Signal**
lactic acid	5.0	1.79	4.24×10^5
isovaleric acid	5.2	1.07	5.93×10^4
alanine	5.4	1.14	8.10×10^5
glycine	6.0	1.05	4.67×10^4
aspartic acid	6.2	1.92	1.32×10^4
L-valine	6.2	1.57	1.61×10^4
pyruvate	6.2	1.28	9.16×10^5
L-cysteine	6.9	1.61	3.11×10^6
oxalic acid	7.2	1.43	2.49×10^5
methylmalonic acid	8.0	1.68	2.26×10^6
malonic acid	8.1	1.61	1.58×10^6
glycine	8.5	0.27	2.51×10^4
L-serine	8.8	1.72	1.39×10^4
fumaric acid	9.8	1.84	8.33×10^6
4-methylvaleric acid	9.8	1.4	7.32×10^4
succinic acid	9.9	1.62	3.73×10^6
acetic acid	10.1	1.18	3.99×10^4
maleic acid	10.2	1.47	1.33×10^6
uracil	10.8	0.58	1.79×10^5
creatinine	13.4	1.84	5.90×10^5
L-glutamic acid	13.7	1.54	5.89×10^4
pyroglutamic acid	14.2	1.43	6.82×10^5
glycerol-3-phosphate	15.3	0.03	3.77×10^5
D-galactose	16.1	1.09	4.18×10^5
D-glucose	16.2	1.07	2.74×10^5
gluconic acid	17.2	1.29	1.63×10^5
D-glucose-6-phosphate	17.3	0.14	1.63×10^5
myo-inositol	17.8	1.71	9.06×10^5
maltose	24.5	0.53	1.84×10^4

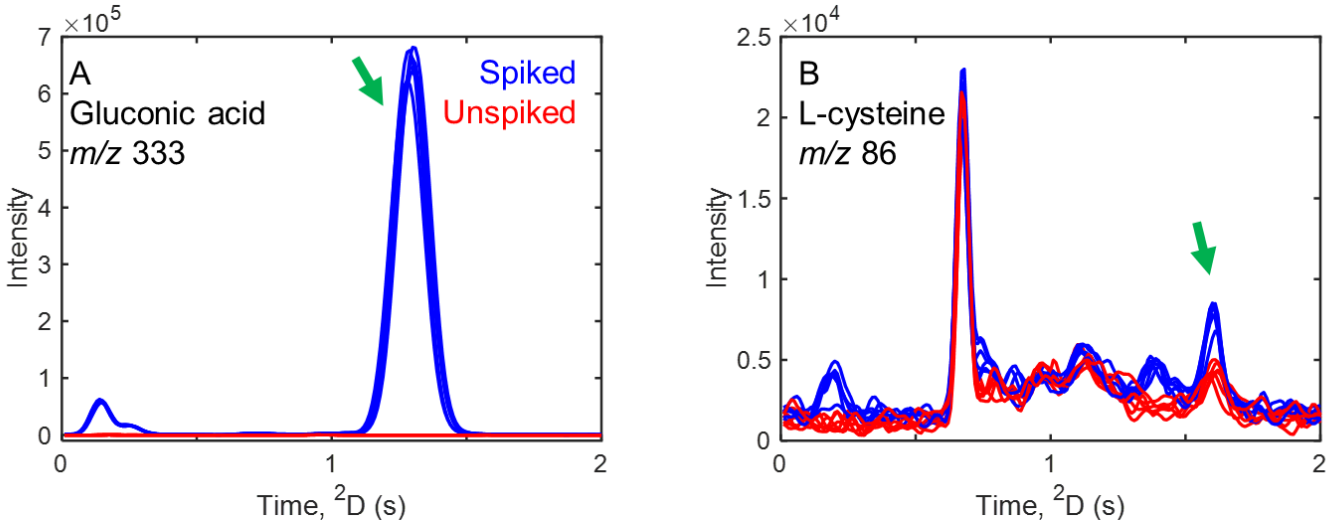


Figure C.2. Two examples of analytes spiked at a concentration of 50 ppm in the pooled farmed fish samples (indicated by green arrows), which ranged in signal and in observable signal-to-noise (*S/N*) ratio. (A) A high *S/N* hit example represented by the overlay of all spiked (blue) and unspiked (red) traces of gluconic acid using selective *m/z* 333, and (B) a low *S/N* hit example of L-cysteine using selective *m/z* 86. Both were successfully discovered and identified by ^{CN}F-ratio analysis.

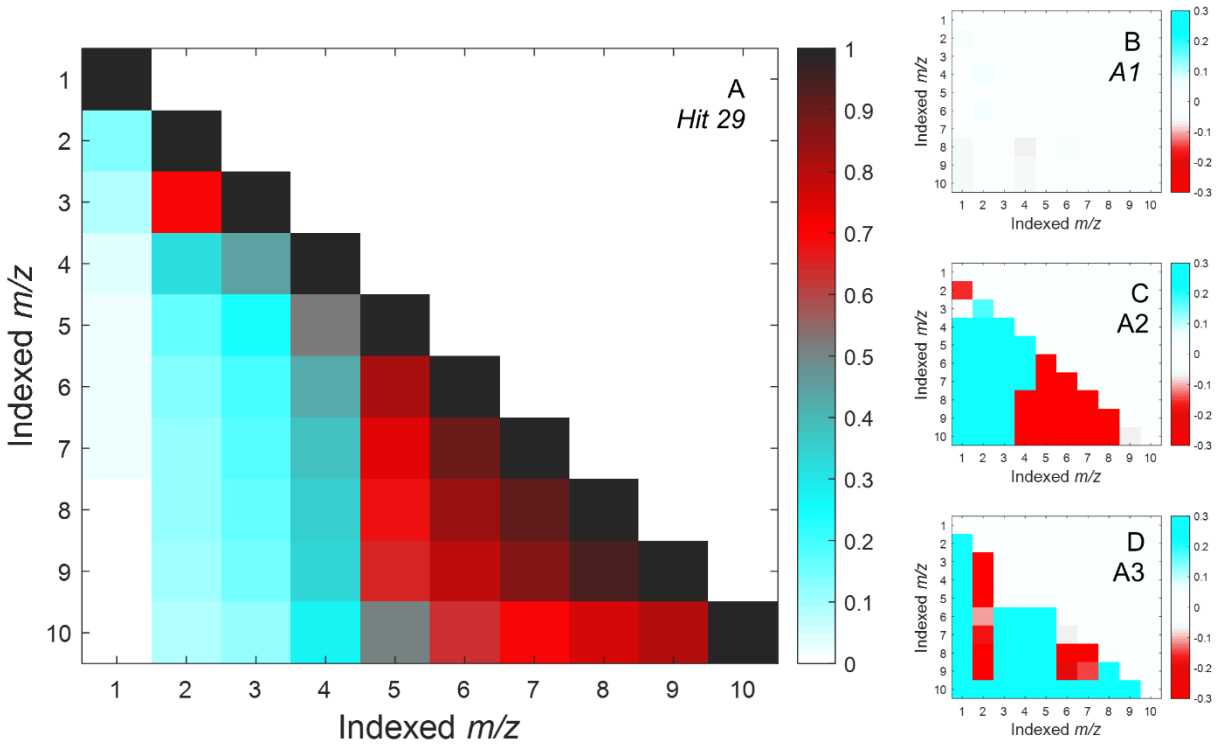


Figure C.3. Indexed mass purity plots show the deviation between the hit mass spectrum and the reference mass spectra (A1-A3). (A) Indexed mass ratio plot is the ratio of intensities at the indexed *m/z*, in this case the 10 most intense *m/z*, with the diagonal all equal to 1. The same indexed mass ratio plot can be calculated for each of the reference spectra using the same *m/z* as for the hit and the deviation between the mass ratio plots is represented by the indexed mass purity plot. The indexed mass purity plot showing the deviation between the hit spectrum and A1 (B), A2 (C), and A3 (D). If there is no deviation, or the spectra are the same, the plot should be mainly white such as B. In this case, hit 29 was confirmed to be A1 and correctly removed by the artifact removal algorithm.

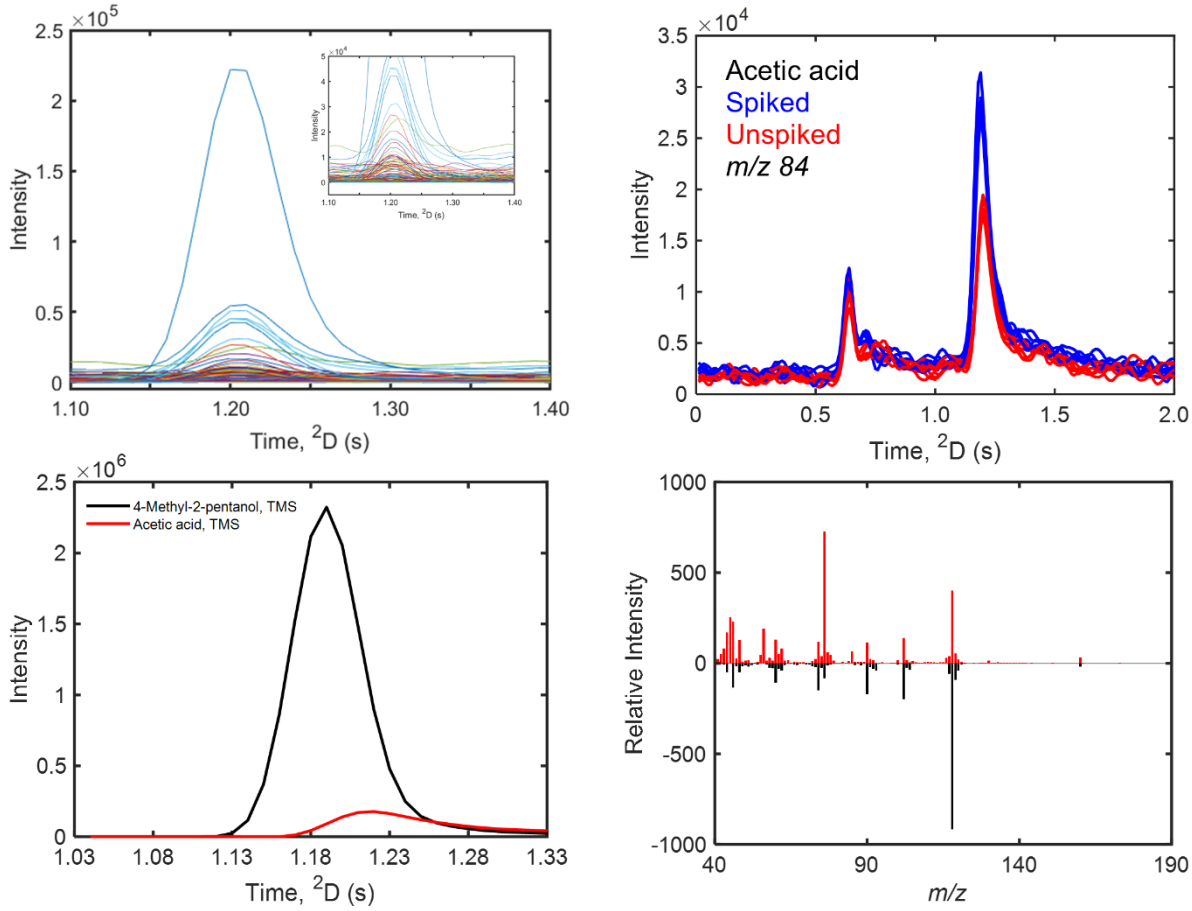


Figure C.4. Identification facilitated by PARAFAC decomposition. (A) Overlay of all *m/z* at the ^2t_R of hit 21 in the background normalized hit list of the spiked data set after artifact removal and redundant hit removal. An inset is provided for visualization of lower intensity *m/z* that correspond to acetic acid peak. (B) Overlay of the summed 2D plots for all spiked samples (blue) and unspiked samples (red) using selective *m/z* 84 determined from the PARAFAC loadings. (C) PARAFAC loadings for 4-methyl-2-pentanol (black) and acetic acid (red) overlay showing slightly different ^2t_R and corresponding mass spectrum using the same color traces (D).

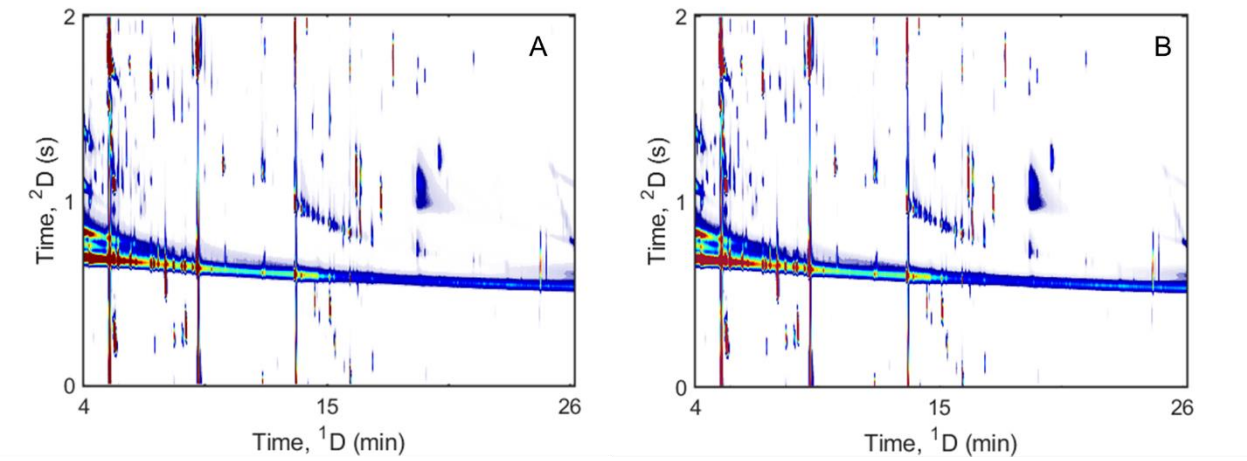


Figure C.5. TIC chromatograms for tank-raised and farmed pacu fish. (A) Tank-raised TIC chromatogram using the average chromatogram of all tank-raised replicates and (B) the farmed TIC chromatogram using the average chromatogram of all farmed replicates. The high signal peaks that make up the streaks drown out the signal for many of the lower signal analytes and make it difficult to notice differences between chromatograms.

Table C.2. Standard F-ratio hit list including compound name, F-ratio, final hit numbers, [F]/[T], p-value calculated for determining the significance of the concentration ratio, pure *m/z* used for [F]/[T] calculation, and *LOF*.

#	Compound name	F-ratio	Hit #	[F]/[T]	*p*-value	*m/z*	*LOF*(%)
1	caproic acid	98	1	0.24	9.4E-07	73	--
2	2,3-butanediol	95	2	2. 9	2.3E-07	117	7.9
3	glycerol	92	3	1.9	1.1E-05	74	8.6
4	glycylisoleucine	76	4	0.68	0.098	86	18.9
5	glycolic acid	61	5	0.50	9.9E-07	204	2.6
6	unknown metabolite 1	58	6	1.4	1.9E-04	73	16.4
7	2-butanol	51	7	0.029	1.8E-05	117	--
8	4-hydroxybutyric acid	50	8	2.7	3.1E-06	147	10.3
9	unidentified artifact 1	48	9	--	--	--	--
10	D-glucose	46	10	0.67	5.9E-05	135	3.8
11	myo-inositol	45	11	1.9	3.0E-05	75	4.7
12	unidentified artifact 2	42	12	--	--	--	--
13	methylaminoheptane	40	13	3.0	5.2E-06	170	6.8
14	DL-phenylalanine	38	14	0.080	1.6E-04	74	--
15	unidentified artifact 3	37	15	--	--	--	--
16	L-cysteine	35	16	0.65	7.7E-04	146	1.9
17	mannose	32	17	0.63	2.2E-04	160	2.3
18	unidentified artifact 4	31	18	--	--	--	--
19	unknown metabolite 2	30	19	0.35	0.0082	73	--
20	bromosuccinate	28	20	0.33	8.3E-06	73	10.0
21	L-serine	28	21	0.41	5.3E-04	116	19.1
22	unidentified artifact 5	28	22	--	--	--	--
23	scyllo-inositol	27	23	0.15	0.0022	318	9.5
24	unidentified artifact 6	26	24	--	--	--	--
25	unidentified artifact 7	25	25	--	--	--	--
26	L-leucine	25	26	0.27	6.1E-05	73	--
27	tyrosine	25	27	0.051	0.021	179	--
28	L-valine	25	28	0.52	0.0030	72	18.8
29	D-alloisoleucine	24	29	0.41	0.0023	86	17.7

30	N-trifluoroacetylmorpholine	23	30	0.51	0.0050	113	4.2
31	unknown metabolite 3	21	31	0.89	0.75	117	15.9
32	propylene glycol	20	32	0.17	5.8E-04	117	11.8
33	unidentified artifact 8	19	33	--	--	--	--
34	erythrose	19	34	0.53	0.0034	72	12.4
35	acetic acid	19	35	1.2	0.38	75	7.0
36	unidentified artifact 9	19	36	--	--	--	--
37	1,2-propanediol-1-phosphate	19	37	1.4	0.060	73	10.2
38	creatinine	17	38	0.74	6.3E-05	204	6.0
39	unidentified artifact 10	16	39	--	--	--	--
40	2,2'-methylenediphenol	16	40	0.27	0.0041	171	5.6
41	unidentified artifact 11	16	41	--	--	--	--
42	isovaleric acid	16	42	1.3	0.011	142	2.9
43	methionine	16	43	0.28	0.017	147	19.9
44	L-isoleucine	15	44	0.18	9.6E-04	159	--
45	unknown metabolite 5	15	45	0.37	0.0028	84	--
46	DL-norleucine	15	46	0.62	0.014	74	--
47	unidentified artifact 12	14	47	--	--	--	--
48	sarcosine	14	48	0.14	0.0088	116	--
49	succinic acid	13	49	0.15	0.096	147	--
50	unidentified artifact 13	13	50	--	--	--	--
51	unidentified artifact 14	13	51	--	--	--	--
52	glycerol-3-phosphate	13	52	0.45	0.065	357	3.2
53	glycine	13	53	1.2	0.31	102	3.0
54	norvaline	12	54	0.32	0.0017	144	--
55	L-fucitol	12	55	0.33	5.7E-04	117	11.7
56	L-leucine	11	56	0.55	0.031	86	13.4
57	unknown metabolite 4	11	57	0.74	0.011	188	1.9
58	β-D-glucopyranose	10	58	0.73	0.073	191	2.1
59	undecanoic acid	9	59	1.6	0.011	43	9.2
60	L-glutamine	9	60	0.79	0.045	84	17.4
61	L-ornithine	9	61	0.029	0.049	174	--
62	unidentified artifact 15	9	62	--	--	--	--

63	salicylic acid	9	63	0.16	0.018	267	--
64	2-aminobenzoxazole	8	64	1.3	0.37	263	7.4
65	xylitol	8	65	2.6	2.4E-05	217	--
66	hypoxanthine	8	66	0.45	0.019	136	9.6
67	unidentified artifact 16	8	67	--	--	--	--
68	unidentified artifact 17	7	68	--	--	--	--
69	unidentified artifact 18	7	69	--	--	--	--
70	2-hydroxyisovaleric acid	6	70	0.48	0.049	147	--
71	niacinamide	6	71	0.77	0.34	179	8.6
72	unidentified artifact 19	6	72	--	--	--	--
73	unidentified artifact 20	5	73	--	--	--	--
74	D-xylose	5	74	1.5	0.11	73	6.1

Table C.3. Tank-normalized F-ratio hit list including compound name, F-ratio, final hit numbers, [F]/[T], p-value calculated to determine significance of concentration ratio, pure *m/z* used for [F]/[T] calculation, and *LOF*.

#	Compound name	F-ratio	Hit #	[F]/[T]	*p-value*	*m/z*	*LOF*(%)
1	glycerol	751	1	1.9	1.1E-05	74	8.6
2	2,3-butanediol	591	2	2.9	2.3E-07	117	7.9
3	N-(hydroxymethyl)trifluoroacetamide	390	3	3.1	1.7E-05	170	12.7
4	methylaminoheptane	271	4	3.0	5.2E-06	170	6.8
5	4-hydroxybutyric acid	211	5	2.7	3.1E-06	147	10.3
6	caproic acid	185	6	0.24	9.4E-07	73	--
7	mercaptoacetic acid	132	7	0.44	9.1E-04	221	17.2
8	N,N-dimethylethanolamine	98	8a	2.1	6.7E-07	58	--
9	alanine	98	9	0.69	8.3E-04	44	14.2
10	unidentified artifact 1	93	10	--	--	--	--
11	myo-inositol	92	11	1.9	3.0E-05	75	4.7
12	glycolic acid	69	12	0.50	9.9E-07	204	2.7
13	D-glucose	54	13	0.67	5.9E-05	135	3.8
14	2-butanol	54	14	0.029	1.8E-05	117	--
15	unidentified artifact 2	53	15	--	--	--	--
16	L-cysteine	52	16	0.65	7.7E-04	146	1.9
17	lactic acid	41	17	1.0	0.49	129	10.0
18	asparagine	39	18	0.53	2.8E-06	80	5.1
19	L-valine	38	19	0.52	0.0030	72	18.8
20	galactose	36	20	1.3	0.054	73	13.4
21	bromosuccinate	34	21	0.33	8.3E-06	73	9.9
22	L-serine	34	22	0.41	5.3E-04	116	19.1
23	L-leucine	30	23	0.27	6.1E-05	73	--
24	tyrosine	30	24	0.051	0.021	179	--
25	unidentified artifact 3	29	25	--	--	--	--
26	1,2-propanediol-1-phosphate	29	26	1.4	0.060	73	10.2
27	diethylcarbamate	28	27	0.80	0.032	174	2.6
28	isovaleric acid	28	28	1.3	0.011	142	2.9
29	scyllo-inositol	28	29	0.15	0.0022	318	9.5

30	acetic acid	27	30	1.2	0.38	75	7.0
31	D-alloisoleucine	27	31	0.41	0.0023	86	17.7
32	N-trifluoroacetylmorpholine	26	32	0.51	0.0050	113	4.2
33	stearic acid	25	33	1.9	0.050	73	14.3
34	unidentified artifact 4	25	34	--	--	--	--
35	unknown metabolite 3	25	35	0.89	0.75	117	15.9
36	pyruvate	24	36	0.47	0.012	174	12.2
37	unidentified artifact 5	23	37	--	--	--	--
38	niacinamide	22	38	0.77	0.34	179	8.6
39	unidentified artifact 6	20	39	--	--	--	--
40	unknown metabolite 5	20	40	0.37	0.0028	84	--
41	creatinine	19	41	0.74	6.3E-05	204	6.0
42	glycerol-3-phosphate	18	42	0.45	0.065	357	3.2
43	unidentified artifact 7	18	43	--	--	--	--
44	salicylic acid	18	44	0.16	0.018	267	--
45	2,2'-methylenediphenol	17	45	0.27	0.0041	171	5.6
46	L-isoleucine	17	46	0.19	9.6E-04	159	--
47	sarcosine	16	47	0.14	0.0088	116	-
48	succinic acid	16	48	0.15	0.096	147	--
49	L-isoleucine	15	49	0.26	0.0066	86	--
50	L-fucitol	15	50	0.29	5.7E-04	117	11.7
51	unidentified artifact 8	15	51	--	--	--	--
52	L-leucine	15	52	0.55	0.031	86	13.4
53	norvaline	15	53	0.32	0.0017	144	--
54	unknown metabolite 4	14	54	0.74	0.011	188	1.9
55	methionine	14	55	0.28	0.017	147	19.9
56	unidentified artifact 9	13	56	--	--	--	--
57	unidentified artifact 10	12	57	--	--	--	--
58	malic acid	10	58	0.47	0.13	147	6.4
59	2-hydroxyisovaleric acid	10	59	0.48	0.049	147	--
60	D-xylose	10	60	1.5	0.11	73	6.1
61	unidentified artifact 11	10	61	--	--	--	--
62	L-ornithine	9	62	0.029	0.049	174	--

63	creatinine enol	8	63	0.15	0.029	115	--
64	hypoxanthine	8	64	0.45	0.019	136	9.6
65	maltose	8	65	0.52	0.13	361	4.0
66	2-pentanol	7	66	1.7	9.9E-04	73	--
67	L-valine	6	67	0.55	0.041	72	--
68	lactose	6	68	1.0	0.98	204	4.5
69	glycine	6	69	1.2	0.31	102	3.0

Table C.4. Farm-normalized F-ratio hit list including compound name, F-ratio, final hit numbers, [F]/[T], p-value calculated to determine statistical significance of concentration ratio, pure *m/z* used for [F]/[T] calculation, and *LOF*.

#	Compound name	F-ratio	Hit #	[F]/[T]	*p*-value	*m/z*	*LOF*(%)
1	2-butanol	19113	1	0.029	1.8E-05	117	--
2	γ-butyrolactone	13456	2	0.12	0.011	86	--
3	propylene glycol	6045	3	0.17	5.8E-04	117	11.8
4	tyrosine	5951	4	0.051	0.021	179	--
5	DL-phenylalanine	2098	5	0.080	1.6E-04	74	--
6	scyllo-inositol	1222	6	0.15	0.0022	318	9.5
7	L-ornithine	816	7	0.029	0.049	174	--
8	D-glucose	750	8	0.67	5.9E-05	135	3.8
9	succinic acid	702	9	0.15	0.096	147	--
10	glycolic acid	687	10	0.50	9.9E-07	204	2.7
11	caproic acid	551	11	0.24	9.4E-07	73	--
12	glycerol-3-phosphate	465	12	0.45	0.065	357	3.2
13	creatinine	428	13	0.74	6.3E-05	204	6.0
14	2,2'-methylenediphenol	389	14	0.27	0.0041	171	5.6
15	mannose	372	15	0.52	2.2E-04	160	2.3
16	sarcosine	310	16	0.14	0.0088	116	--
17	pyruvate	289	17	0.47	0.012	174	12.2
18	L-cysteine	275	18	0.65	7.7E-04	146	1.9
19	D-alloisoleucine	239	19	0.41	0.0023	86	17.7
20	L-iosleucine	226	20	0.19	9.6E-04	159	--
21	unidentified artifact 1	225	21	--	--	--	--
22	hexadecenitrile	213	22	0.88	0.38	136	--
23	hypoxanthine	211	23	0.45	0.019	136	9.6
24	glycerol	207	24	1.9	1.1E-05	74	8.6
25	bromosuccinate	204	25	0.33	8.3E-06	73	10.0
26	L-leucine	200	26	0.27	6.1E-05	73	--
27	methionine	190	27	0.28	0.017	147	19.9
28	L-serine	187	28	0.41	5.3E-04	116	19.1
29	L-valine	158	29	0.52	0.0030	72	18.8

30	D-mannopyranose	145	30	0.83	0.22	191	3.5
31	2,3-butanediol	145	31	2.9	2.3E-07	117	8.0
32	alanine	123	32	0.69	8.3E-04	44	14.2
33	N,N-dimethylethanolamine	123	33	2.1	6.7E-07	58	--
34	β-D-glucopyranose	122	34	0.73	0.073	191	2.1
35	asparagine	122	35	0.53	2.8E-06	80	5.1
36	L-isoleucine	118	36	0.26	0.0066	86	--
37	DL-norleucine	115	37	0.62	0.014	74	--
38	unknown metabolite 1	111	38	1.4	1.9E-04	73	16.4
39	L-arginine	104	39	0.45	0.0019	257	--
40	myo-inositol	104	40	1.9	3.0E-05	75	4.7
41	malic acid	103	41	0.47	0.13	147	6.7
42	unidentified artifact 2	102	42	--	--	--	--
43	mercaptoacetic acid	98	43	0.44	9.1E-04	221	17.2
44	unknown metabolite 2	98	44	0.35	0.0082	73	--
45	acetic acid	95	45	1.2	0.38	75	7.0
46	N-trifluoroacetylmorpholine	92	46	0.51	0.0050	113	4.2
47	unknown metabolite 3	91	47	0.89	0.75	117	15.9
48	unknown metabolite 4	87	48	0.74	0.011	188	1.9
49	D-glucose-1-phosphate	83	49	0.32	1.5E-04	217	6.7
50	unidentified artifact 3	81	50	--	--	--	--
51	maltose	80	51	0.52	0.13	361	4.0
52	unknown metabolite 5	75	52	0.37	0.0028	84	--
53	niacinamide	71	53	0.77	0.34	179	8.6
54	methylaminoheptane	66	54	3.0	5.2E-06	77	--
55	erythrose	62	55	0.53	0.0034	72	12.4
56	urea	62	56	1.6	0.18	171	--
57	diethylcarbamate	61	57	0.80	0.032	174	2.6
58	unidentified artifact 4	55	58	--	--	--	--
59	nonadecane	54	59	1.5	0.25	57	--
60	L-fucitol	53	60	0.33	5.7E-04	117	11.7
61	1,2-propanediol-1-phosphate	53	61	1.4	0.060	73	10.2

62	norvaline	50	62	0.32	0.0017	144	--
63	DL-ornithine	45	63	0.27	0.21	147	--
65	unidentified artifact 5	38	65	--	--	--	--
64	β-amino isobutyric acid	34	64	1.1	0.49	73	2.5
66	glycylisoleucine	34	66	0.68	0.098	86	18.9
67	L-leucine	33	67	0.55	0.031	86	13.4
68	2-hydroxyisovaleric acid	30	68	0.48	0.049	147	--
69	malonic acid	29	69	0.32	0.035	233	--
70	glycine	29	70	1.2	0.31	102	3.0
71	4-methyl-3-heptanol	29	71	0.57	0.037	73	--
72	isovaleric acid	28	72	1.3	0.011	142	2.9
73	pyrocatechol	27	73	0.25	0.041	254	--
74	D-arabinose	27	74	2.9	4.5E-06	318	8.1
75	2-aminobenzoxazole	25	75	1.3	0.37	263	7.4
76	galactose	24	76	1.3	0.054	73	13.4
77	L-glutamine	22	77	0.79	0.045	84	17.4
78	threonic acid	21	78	0.50	8.6E-04	117	--
79	unknown metabolite 6	21	79	0.86	0.51	221	8.6
80	citrulline	21	80	0.78	0.090	188	3.7
81	salicylic acid	20	81	0.16	0.018	267	--
82	unidentified artifact 6	18	82	--	--	--	--
83	undecanoic acid	18	83	1.6	0.011	43	9.2
84	methylmalonic	18	84	0.40	0.034	218	--
85	unidentified artifact 7	18	85	--	--	--	--
86	xylitol	18	86	2.6	2.4E-05	217	--
87	2-ethylhexanoic acid	18	87	0.41	0.046	201	--
88	unknown metabolite 7	18	88	1.3	3.7E-04	103	--
89	unidentified artifact 8	17	89	--	--	--	--
90	unidentified artifact 9	17	90	--	--	--	--
91	unknown metabolite 8	17	91	3.0	2.9E-07	73	--
92	creatinine enol	17	92	0.15	0.029	115	--
93	unidentified artifact 10	16	93	--	--	--	--

94	aspartic acid	16	94	0.66	2.5E-04	232	6.2
95	unidentified artifact 11	16	95	--	--	--	--
96	unknown metabolite 9	15	96	1.1	0.44	117	--
97	2-pentanol	14	97	1.7	9.9E-04	73	--
98	unidentified artifact 12	14	98	--	--	--	--
99	L-valine	12	99	0.55	0.041	72	--
100	unknown metabolite 10	9	100	0.64	1.6E-05	186	--
101	lactose	9	101	1.0	0.98	204	4.5
102	n-butylamine	8	102	4.0	0.38	174	--
103	D-xylose	6	103	1.5	0.11	73	6.1
104	unknown metabolite 11	4	104	0.75	0.0017	147	8.4
105	2,2-dimethyl-3-pentanol	4	104	1.1	0.73	73	14.3

Table C.5. Wilcoxon rank sum test hit list including compound name, z-statistic, and final hit numbers. The hit list is organized by descending z-statistic with an applied z-critical of 1.96 for a two-tailed test at the 95% confidence level.

#	Compound name	z-statistic	Hit #
1	2,3-butanediol	5.38	1
2	glycerol	5.33	2
3	caproic acid	5.23	3
4	N-(hydroxymethyl)trifluoroacetamide	5.09	4
5	gamma-butyrolactone	5.02	5
6	methylaminoheptane	4.86	6
7	glycolic acid	4.64	7
8	D-glucose	4.60	8
9	2-butanol	4.52	9
10	myo-inositol	4.48	10
11	propylene glycol	4.42	11
12	L-cysteine	4.36	12
13	mannose	4.30	13
14	L-serine	4.21	14
15	DL-phenylalanine	4.21	15
16	unidentified artifact 1	4.13	16
17	tyrosine	4.10	17
18	L-alanine	4.04	18
19	asparagine	4.01	19
20	unidentified artifact 2	3.98	20
21	sarcosine	3.87	21
22	L-isoleucine	3.78	22
23	D-alloisoleucine	3.72	23
24	lactic acid	3.67	24
25	creatinine	3.63	25
26	L-leucine	3.61	26
27	unidentified artifact 3	3.52	27
28	N-trifluoroacetylmorpholine	3.51	28

29	L-valine	3.51	29
30	pyruvate	3.39	30
31	2,2'-methylenediphenol	3.32	31
32	unidentified artifact 4	3.29	32
33	acetic acid	3.29	33
34	unidentified artifact 5	2.95	34
35	DL-norleucine	2.87	35
36	unidentified artifact 6	2.76	36
37	isovaleric acid	2.75	37
38	unidentified artifact 7	2.74	38
39	unknown metabolite 4	2.71	39
40	1,2-propanediol-1-phosphate	2.66	40
41	unidentified artifact 8	2.66	41
42	β-D-glucopyranose	2.65	42
43	unidentified artifact 9	2.63	43
44	unidentified artifact 10	2.62	44
45	unidentified artifact 11	2.56	45
46	unknown metabolite 3	2.34	46
47	methionine	2.34	47