Development of chemometric methodologies for supervised class-based discovery experiments using GC×GC-TOFMS: application to aerospace fuel analysis

Grant S. Ochoa

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

Dan Fu

Matthew Bush

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Development of chemometric methodologies for supervised class-based discovery experiments using GC×GC-TOFMS: application to aerospace fuel analysis

Grant S. Ochoa

Chair of the Supervisory Committee:

Robert E. Synovec

Chemistry

Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection (GC×GC-TOFMS) is a prominent instrumental technique which produces information rich datasets for the analysis of complex volatile mixtures. Due to this fact advanced data analysis methods are required and chemometric supervised discovery approaches excel at by exploiting class-based experimental design to extract useful chemical information. Analysis of aerospace kerosene-based fuels benefit from these approaches as the samples often contain thousands of unique compounds and features-of-interest may be minute in concentration. This dissertation describes several studies pertaining to the development and application of supervised discovery experiments (tile-based Fisher ratio (F-ratio) and 1v1 analysis) to fuels, focused on

extracting pertinent information. First, tile-based F-ratio results were used alongside measures of peak shape consistency to statistically determine which *m/z* were pure (comprised of a single compound) for each hit for accurate relative quantitation. Next, building on the ideas of the first project, tile-based F-ratio results are used to statistically determine which *m/z* are consistent between classes to enable the extraction of a purified mass spectrum representing the differences between classes for reliable compound identification. Next, a class-based experiment is set up to target the identities of extractable polar contaminants in rocket fuel using solid-phase extraction. This approach is then applied to multiple fuels for the purpose of inter-batch analysis for errant fuel detection. The extraction behavior of polar compounds within a fuel matrix is then investigated with tile-based 1v1 analysis. Differences in extraction profiles are investigated for several for two different stationary phases as a function of load volume. Finally, partial least squares (PLS) is combined with tile-based variance ranking and RreliefF to improve modeling of carbonaceous deposits in thermal fuel lines, a measure of thermal stability, for a dataset of rocket fuels. The results are then used to identify the compounds that are highly correlated with increased carbon deposition.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I'd like to start off by thanking Dr. Robert Synovec for accepting me into his group in 2018. I'm grateful for the this opportunity as it allowed me to discover passion for analytical research and to develop skills I may not have obtained otherwise. I'd also like to thank Dr. Derrick Gough and Dr. Sarah Prebihalo, my senior group members who helped show me the ropes when joining the lab. Their guidance in learning the basics of gas chromatography and coding helped me refine my skills to push my projects further. I am where I am today due in part to your help.

The biggest thanks must go to my wife and the love of my life, Kristy, for being with me every step of the way through my Ph.D. journey. From the get-go into looking into graduate school she was incredibly supportive and willing to uproot her life in Colorado and move with me to Seattle, for which I will forever be thankful. Thank you for your emotional support, getting me through the tough times and helping to make our apartment a home with our dogs. I couldn't ask for anyone else to share my life and experiences with and I hope to have a bright future ahead with you.

Moving out of the state for school is a big life change and I wouldn't have been able to do it without my family back home supporting me and Kristy every step of the way, my mom and dad, Kathie and Steve, and brother, Chase. Thank you for always calling Kristy and I to check up on us and keep up with us when life gets busy. Thank you for coming out of your way to visit us in Seattle, we always enjoyed the time we spent with you showing you around our new home. Mom, thank you for always being there for advice on professional emails, resumes, it always made me more confident in my abilities. Dad, thank you for helping me through tough times, listening without judgement, when I needed it. Chase, thanks for keeping up on our

collections at home and sending me great music and game recommendations, its definitely helped me in finding new musical interests.

To my friends, back home, Dylan, Codee, and Drew, thank you for being all the game nights that helped me stay entertained throughout my Ph.D., it made it feel like I'll always have a place no matter where life takes us. I know that our schedules don't always line up, but you've always made the effort to find time to hang out and for that I'll always be thankful. Those nights not only gave us much needed time to decompress with games, but also great conversation that made nights interesting, especially throughout the pandemic. I'm proud to count you all as some of my best friends.

My lab mates and friends in Seattle, Tim, Sonia, Paige, Caitlin, and Joe, also known as the "chromies", have been a huge source of support, academically, recreationally, and socially. I started my Ph.D. journey in the Synovec group at the same time as Tim and Sonia, and now here we are 4.5 years later finishing up together. Both Tim and Sonia have been wonderful friends and research collaborators to me, always willing to hear my research ideas and to hang out outside of the lab. Paige, Caitlin, and Joe similarly have been amazing friends, and while I have more to say I'll keep it brief and say thank you for being wonderful friends. Thank you for inviting me to new experiences, I don't think I would have gone to exercise classes nearly as much without you all there.

Lastly, thank you to our dogs, Marty and Machiko, for being such sweet companions to me and Kristy. Seeing you two so excited to see me everyday gave me something to look forward to everyday. Everybody loves you two and I'm so happy that we get to call you part of our family.

Chapter 1. Introduction

Parts of this chapter were reproduced from Caitlin N. Cain, Timothy J. Trinklein, Sonia Schöneich, Grant S. Ochoa, Sarah C. Rutan, Robert E. Synovec, Comprehensive Two-Dimensional Chromatography with Chemometric Data Analysis. In: Nelu Grinberg, Peter W. Carr, editors. Advances in Chromatography Vol. 59 Boca Raton, FL (USA): CRC Press 2023, p. 173-176.

1.1. Gas chromatography fundamentals

1.1.1. Principles

Chromatography is an important analytical technique for assessing the composition of chemical mixtures. It is broadly defined as the act of separating chemicals based on their interactions between a mobile and stationary phase [1]. How a compound partitions between these two phases can be described with the distribution coefficient K_D:

$$K_D = \frac{[analyte]_{sp}}{[analyte]_{mp}} \quad (1)$$

where $[analyte]_{sp}$ is the analyte concentration in the stationary phase and $[analyte]_{mp}$ is the analyte concentration in the mobile phase. Analytes with a high K_D will have a higher affinity for the stationary phase while analytes with a low K_D will have a higher affinity for the mobile phase. As the names suggest, the stationary phase is static and is generally a solid or liquid, and the mobile phase flows over the stationary phase and is either a liquid or solid. As the mobile phase flows over the stationary phase, it pulls along analytes from the chemical mixture at a rate indicated by their K_D, with analytes that have a low K_D travelling more quickly than analytes with a high K_D. By carefully selecting the phases and adjusting other parameters to obtain desirable K_D's a chemical mixture can be separated into its constituent compounds [1,2]. How these parameters are adjusted specifically will depend on the type of chromatography being performed (e.g., thin-layer chromatography, gas chromatography, ion chromatography, liquid chromatography, etc.). Chromatography is invaluable and is ubiquitous in both qualitative and

quantitative work from chemical synthesis to chemical analysis [3–6]. This dissertation will focus on gas chromatography (GC), the technique utilized in all research within, however the fundamentals described are applicable to most forms of chromatography.

Gas chromatography, first introduced in 1952, has become an invaluable multi-disciplinary technique for the analysis of volatile and semi-volatile compounds [7]. In GC the analytes partition between a liquid stationary phase coating the inside of a long silica column housed inside an air-bath oven and a gaseous mobile phase which carries the analyte(s) in the injected sample. Here the analyte K_D's are due in some part to intermolecular interactions with the stationary phase but are largely governed by analyte volatility. Thus, for a given temperature analytes with low volatility will spend more time in the stationary phase and move down the column slower than more volatile analytes. The mobile phase for flame ionization detection (FID) is generally hydrogen (H_2) gas due to its high optimal linear flow velocity over other gases. For mass spectrometry (MS) detection helium (He) is preferred as inertness avoids any unwanted fragmentation artifacts after ionization, as well as for achieving optimum vacuum conditions. The stationary phase is generally a non-polar cross-linked polydimethylsiloxane (PDMS) backbone that can be functionalized with phenyl, cyanopropyl, and trifluoropropyl groups as a liquid coated along the walls of a silica-glass column. Typical column lengths (L) are between 15-60 m, the column inner diameter (d_c) ranges from 100-530 µm, and stationary phase film thickness (d_f) can range between 0.1-1.5 µm. For polar applications polyethylene glycol (PEG) can be used to separate mixtures with predominantly polar heteroatom-containing compounds such as alcohols, ketones, and esters. However, these columns have a lower maximum operating temperature, thus must be used with care to avoid stationary phase degradation [1,8].

The elution order for a homologous series of compounds (compounds with similar structures) in GC will be in order of boiling point. However, differences in compound structure and polarity can affect the retention time, thus elution order will not necessarily match boiling point order. In GC the mobile phase gas is generally inert thus analyte retention will be affected not by analyte-mobile phase interactions but by analyte volatility and analyte-stationary phase interactions. For each stationary phase, an analyte has a specific retention factor, *k*', that the analyst can use to relate thermodynamic information to chromatographic retention. The retention factor, *k*', is defined as:

$$k' = \frac{t_r - t_0}{t_0} \tag{2}$$

where t_0 is the dead time, or the retention time of a compound which does not interact with the column. Among analytes with similar volatility, those with a low *k*' will have a low affinity for the stationary phase and have a shorter t_r while those analytes with high *k*' will have a high affinity for the stationary phase and have a longer t_r [1]. The retention factor *k* is more specifically influenced by enthalpy caused by the analyte's physical properties such as volatility, entropy by the analyte's interaction with the stationary phase, and the ratio of mobile phase to stationary phase volume, detailed by the following equation:

$$\ln k' = -\frac{\Delta H^\circ}{RT} + \frac{\Delta S^\circ}{R} - \ln \beta \tag{3}$$

Where ΔH° is the enthalpy of vaporization, *R* is the universal gas constant, *T* is the temperature of the chromatographic system, ΔS° is the entropy term, and β is the phase volume ratio.

Analytes eluting from the column are detected according to a Gaussian profile concentration distribution:

$$C(t) = \frac{V_{inj}C_{inj}}{F\sigma\sqrt{2\pi}} e^{-0.5(\frac{t-t_r}{\sigma})^2} \quad (4)$$

where t_r is the analyte retention time, σ is the distributions standard deviation, t is time, V_{inj} and C_{inj} are the volume and concentration of sample injected, respectively, and F is the column flow rate [9]. When $t = t_r$ the exponent drops out and the concentration reaches a maximum value at the analyte's retention time. Thus, the detectors will produce an electrical signal proportional to the concentration of the analyte with the most common detectors for GC being FID and MS. FID is a versatile univariate detector, which detects compounds which burn in a hydrogen flame, i.e., mainly hydrocarbons. For hydrocarbons the FID signal is proportionate to the number of carbons and hydrogen contained in the molecule, with the presence of heteroatoms (nitrogen, sulfur, oxygen) lowering the combustion efficiency of the compounds and in turn lowering their signal [10]. A MS detector on the other hand, is a universal multi-channel detector, in that it can detect any ionizable compound and provides data for several mass channels, *m/z*. These ionized compounds fragment, creating many detectable *m/z* signals, which serve as a chemical "fingerprint" which can be used for unique compound identification [11]. Thus, GC-MS chromatograms are inherently two dimensional in nature, with one separation dimension and one mass spectrum dimension.

1.1.2.Figures-of-Merit

The goal of chromatographic separations is to separate compounds effectively and efficiently, and to this end it is important to consider both the width and resolution of analyte peaks. The width of peaks is measured at their base, w_b, and is equal to 4σ, representing 95% of the total peak area. With w_b we can calculate the degree of overlap between two peaks, the resolution:

$$R_s = \frac{t_{r,2} - t_{r,1}}{\overline{w_b}} \quad (5)$$

where $t_{r,1}$ and $t_{r,2}$ are the retention times of the earlier and later eluting peaks respectively and w_b is the average peak width of base for the two peaks. For peaks to be baseline separated they must have a resolution R_s = 1.5, which is preferred for analyte identification and quantification. Unit resolution (R_s = 1.0) is when the separation between peaks is the same as their peak width at base. In practice it is easier to obtain peaks with R_s = 1 than 1.5, though it does limit the quantification of such peaks. Below a resolution of 1 the peaks start to overlap more heavily and by a R_s = 0.5 their individual peak apexes have merged into one peak. For peaks such as these, mathematical deconvolution is required to separate the combined peak into its individual components for accurate quantitative and qualitative analysis. Examples of these resolutions are demonstrated in Figure 1.1 with each peak plotted individually as red and blue with the observed chromatographic signal (their sum) plotted in black.

For optimal and efficient chromatographic separations, it is best to obtain as narrow peaks as possible, and with maximal resolution in as little as separation run time as possible. Experimental steps the analyst can take to reduce peak width include changing column dimensions, applying temperature programming to the GC oven, and adjusting the carrier gas flow rate. The three parameters of a GC column, length (L), inner diameter (d_c), and film thickness (d_f), alter how efficient the separation will be. Longer columns can allow greater separation between overlapping compounds, but at the risk of wider peaks due to separation run time. Decreasing the d_c of a column will lead to higher

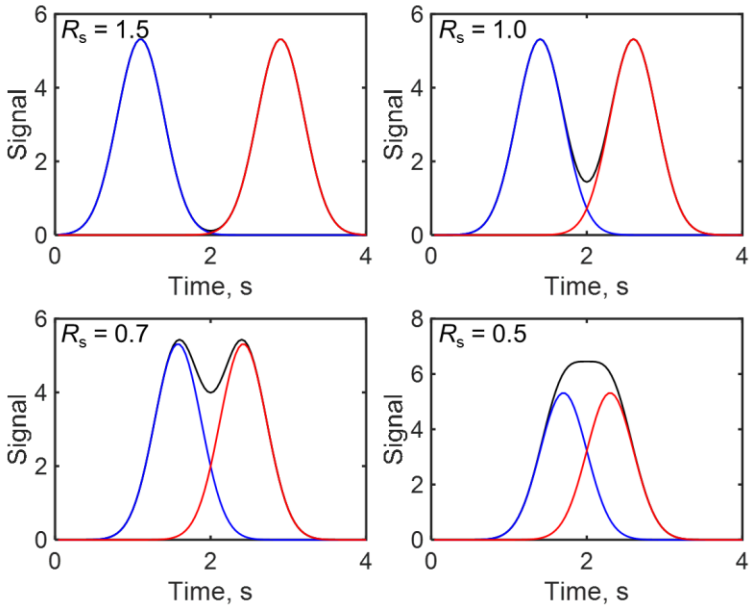


Figure 1.1. Example of overlapped chromatographic peaks at various resolutions. Individual peaks are plotted in red and blue with their sum, i.e., the observed chromatographic signal, plotted in black.

efficiency, i.e. more interactions of the analyte with the stationary phase, but at the cost of lower column capacity and increased column head pressure. Finally, the d_f affects the retention time of analytes, with a thicker d_f increasing retention time which can allow greater R_s between coeluting compounds, at the cost of wider peaks. Temperature programming can help mitigate the General Elution Problem (GEP) where earlier eluting compounds with low k' have poor resolution and narrow w_b and later eluting compounds with high k' have superfluous resolution and wide w_b. By optimally increasing the temperature of the oven over time, all compounds will elute with $k' \sim 0$ and the same w_b. Thus, when designing chromatography methods, it is important to use R_s as a metric to determine when the separation results are satisfactory and/or optimal. Finally, the carrier gas flow rate may be increased to shorten analysis time, but generally at the expense of resolution. Thus, by balancing these parameters, we can achieve efficient separations with narrow peaks and ideal resolution between analytes.

To assess the overall performance of a GC method an analyst can calculate the peak capacity n_c, estimated using the following equation for an optimized temperature programmed separation,

$$n_c = \frac{t_{sep}}{w_b} \quad (6)$$

where t_{sep} is the total separation time, w_b is the average peak width and n_c represents the maximum number of unit resolution peaks that can fit in the separation. It is desirable to achieve high peak capacities which allows for greater resolution of complex mixtures. Thus, narrow peaks will produce higher peak capacities, which also helps with detection sensitivity since narrower peaks are taller than wide peaks (holding injected concentration constant, see Eq. 4). Producing wide peaks is inefficient because not only does it hamper the resolution of closely eluting compounds, but it will also interfere with the detection of low concentration compounds. To achieve the production of narrow peaks the analyst can adjust many instrumental parameters such as the injection volume, carrier gas flow rate, temperature programming affecting the aforementioned GEP, column dimensions, and stationary phase thickness. By adjusting these parameters, we are looking to reduce band broadening of the analyte 'packet' in the column by affecting its mass transfer and longitudinal diffusion kinetics.

1.1.1. Comprehensive two-dimensional gas chromatography

One dimensional (1D) GC has long stood as an effective separation technique for volatile mixtures, serving as an invaluable tool for both qualitative and quantitative analysis. However, as mixtures grow in complexity the efficacy of 1D-GC decreases as compounds crowd the separation space, creating what are referred to as unresolved complex mixtures (UCM). As an example of this issue take diesel separations, there are potentially thousands of unique organic

compounds present in diesel with high overlap in boiling points. In 1D-GC these compounds are unable to be fully resolved on reasonable time scales and raise the baseline considerably as the compounds coelute in what is referred to as a 'petroleum hump' [12].

Various instrumental advancements have been made to improve upon the resolving power of 1D-GC, referred to collectively as multidimensional gas chromatography (MGC). The primary approach is referred to as heart-cutting, or GC-GC, collecting selected fractions of the effluent from the first GC column dimension to be reinjected onto a second GC column to achieve further separation of the coeluting compounds [13,14]. The transfer of effluent between columns in GC-GC is generally accomplished with a simple flow-switching device [13,14]. However, GC-GC is limited in scope since the extra separation is applied to targeted compounds and not the whole chromatogram. Addressing this, Liu and Phillips developed *comprehensive* two-dimensional gas chromatography (GC×GC) in 1991, achieving full transfer of all first dimension (^{1}D) effluent to a second dimension (^{2}D) column [15].

A GC×GC instrument is generally set up with a ^{1}D column of about 30 m in length connected in series by a modulator to a shorter ^{2}D column (2-5 m) of differing polarity, whereby the modulator acts as a secondary injector. The purpose of the modulator is to trap the effluent exiting the ^{1}D column into a narrow band and reinject it to the ^{2}D column, and process can be accomplished in two ways: thermal versus flow modulation. Thermal modulators use cryogens to rapidly cool the ^{1}D effluent into a narrow band which is then reinjected by heating with a puff of hot gas. Whereas flow modulators use auxiliary flow of carrier gas to direct portions of the ^{1}D effluent to the ^{2}D column. These processes are repeated continuously throughout the separation on a timing called the modulation period, P_M. Thus, after an initial ^{1}D separation the sample will experience a series of very fast ^{2}D separations, thus providing the opportunity to separate

analytes that overlapped on ^{1}D with the ^{2}D column. A typical column selection for ^{1}D could be a non-polar stationary phase (e.g. Rxi-1 100% dimethylpolysiloxane) connected to a mid-polar to polar ^{2}D column (e.g. Rxi-17Sil MS phenyl crosslinked dimethylpolysiloxane), however a reverse column configuration can be employed (polar ^{1}D and non-polar ^{2}D) which generally achieves better separation between non-polar compounds on the ^{2}D dimension for certain samples, i.e., alkanes and cycloalkanes in diesel.

The modulation process is demonstrated in Figure 1.2 for simulated GC×GC data. In Fig. 1.2A the unfolded GC×GC chromatogram of the simulated data is provided, with red dashed lines representing a P_M of 1 s. Here peaks are seen separated within each modulation, and peaks that represent the same analyte between modulations are marked by green, violet and orange asterisks. Notice that for each set of peaks a Gaussian shape is displayed following the analyte's ^{1}D separation. Cutting and folding the data along these modulations yields a contour plot revealing three analytes are separated on ^{2}D while remaining overlapped on ^{1}D (Figure 1.2B). In Figure 1.2C the ^{2}D dimension has been summed away to represent the data as if it were collected by 1D-GC, where it appears there are two unresolved peaks. This underscores the power of GC×GC to resolve complex mixtures and the benefit gained from the added second separation dimension.

With the improved resolution of GC×GC comes the ability to fit more peaks into the 2D chromatographic separation space without overlap, that is, an increase in peak capacity. Klee et. al. stated that given the same separation time GC×GC has the potential to achieve 10-fold or greater peak capacity gain over 1D-GC [16]. Figure 1.3 demonstrates this impressive gain in peak capacity, with a 1D-GC chromatogram of an RP-1 rocket fuel in Fig 1.3A with the complementary GC×GC chromatogram in Fig 1.3B. In Figure 1.3A it is clear that the separation

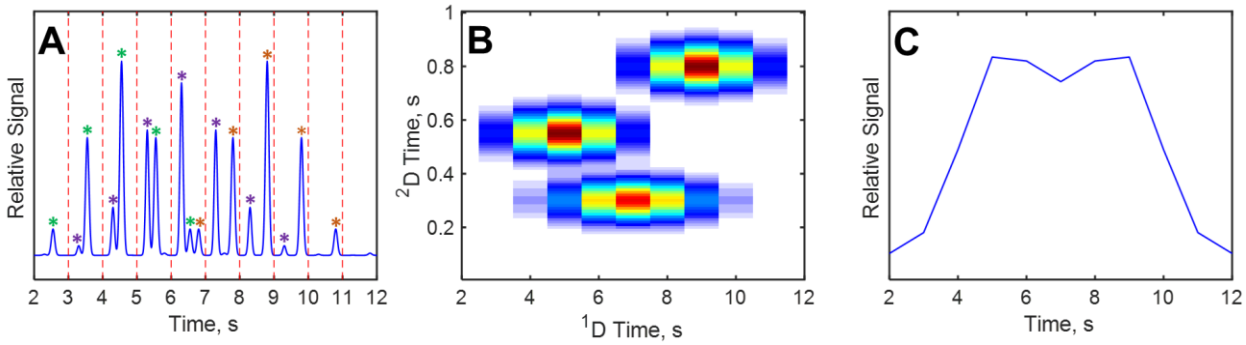


Figure 1.2. Example of GC×GC modulation process. (A) Modulated gas chromatograph signal, showing apparent rise and fall of peaks across modulations, indicated by dashed red line. (B) 2D contour plot of the modulated data, folded along the modulation periods. Here three analytes are seen separated on 2D whereas they are overlapped on 1D. (C) 1D plot with the second dimension summed away showing that without the second dimension it is difficult to determine how many analytes are present.

has many unresolved compounds, displaying the classic "petroleum hump" baseline rise indicative of the complex mixture. To improve on this for 1D-GC an analyst may need to use a longer separation time, however that is not necessary with GC×GC. Compared to Fig 1.3A, hundreds of individual analyte peaks are visible, with good use of the ^{2}D separation dimension. This GC×GC chromatogram was collected with a reverse column configuration, using a mid-polarity ^{1}D column (Rxi-17Sil MS) and a non-polar ^{2}D column (Rxi-1MS).

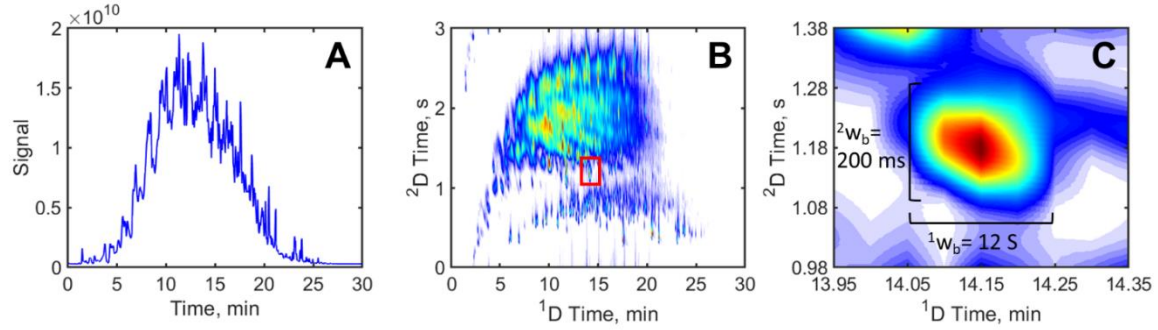


Figure 1.3. Demonstration of the increase in peak capacity gained from transitioning from 1DGC to GC×GC. (A) A 1DGC chromatogram of an RP-1 rocket fuel. (B) A GC×GC chromatogram of the same RP-1 rocket fuel. (C) Example of the peak widths associated with GC×GC.

The peak capacity of GC×GC chromatograms for an optimized temperature programmed ^{1}D separation can be calculated as the product of the peak capacity in the first and second dimension respectively, or expanded:

$$n_{c,2D} = {}^1n_c \times {}^2n_c = \frac{{}^1t\ P_m}{{}^1w_b\ {}^2w_b} \quad (7)$$

where ^1n_c and ^2n_c are the peak capacity in the first and second dimension respectively, ^1t is the total run separation time in seconds, P_M is the modulation period in s, and ^1w_b and ^2w_b are the average peak widths-at-base in the first and second dimensions in s respectively. For a typical peak from the rocket fuel chromatogram the ^1w_b is 12 s and the ^2w_b is 200 ms or 0.2 s. Given the total separation time ^1t = 30 min and P_M = 3 s that gives a ^1n_c = 150 and ^2n_c = 15 respectively, resulting in $n_{c,2D}$ = 2,250. Thus, the peak capacity gain over 1D-GC falls in line with theoretical expectations.

1.2. Chemometric data analysis

1.2.1. Challenges to address

Despite the advantages of GC×GC relative to 1D-GC, the data produced can be large and complex, especially for multi-channel detection. Because each 2D peak contains anywhere from 2-4 modulated peaks from ^{1}D to ^{2}D, conventional analysis such as quantification and identification of compounds becomes cumbersome and impractical. Thus, hands on analysis for complex samples of this size, especially if there are multiple replicates, is not practical. Therefore, more advanced analytic methods (i.e., chemometrics) are required to process the data and glean useful information, preferably involving less manual intervention and computation time. Chemometrics is the use of linear algebra and statistical methods to extract meaningful chemical information from analytical data. These techniques have generally been slow to limited

in their application and adoption, owing to the need for programming experience in software such as MATLAB, R, or Python, and specialized expertise to handle data in these programs. Fortunately, more and more commercial software has been incorporating chemometric approaches, with LECO ChromaTOF Tile and SepSolve ChromCompare+ offering both supervised and unsupervised non-target analysis [17–20].

Chemometric techniques can be categorized into two main approaches based on the analytical objective, targeted and non-targeted. Targeted methods refer to the quantification and identification of pre-selected analytes of interest and is the most basic type of analysis. If the analytes of interest are well-resolved chromatographically, or effectively so due to the detector selectivity (e.g., MS), then advanced chemometrics is generally not required. However, if the analyte is coeluting and poorly resolved, then attempts to quantify or identify analytes of interest will be hindered. In such cases, chemometric decomposition can be employed to obtain the pure signal from the contributing compounds. Two popular methods for this analysis are multivariate curve resolution alternating least squares (MCR-ALS) and parallel factor analysis (PARAFAC), and both work off linear algebra principles [21,22]. Non-targeted techniques can further be classified as either supervised or unsupervised, depending on whether the method makes use of class-based membership. Unsupervised, non-target methods include principal component analysis (PCA), clustering algorithms and unsupervised feature selection while supervised non-target method encompass the Fisher ratio method (F-ratio) for feature selection, and partial least squares (PLS) regression. While these methods will be introduced separately, they may be applied in any order to the data at hand depending upon the analytical objective. For example, quantitative results of targeted methods can be input to build non-target models for data

visualization, or non-target methods can be used to discover important analytes for targeted quantification and identification.

1.2.2. Preprocessing

1.2.2.1. *Baseline correction*

Baseline correction methods are important for removing unwanted background signal contributions to chromatographic data due to low-frequency detector drift and noise, stationary phase degradation, or temperature fluctuations. Applying baseline correction differs depending on if the data utilized single or multi-channel detection. For single channel detection, such as GC-FID data or a GC-MS total ion current (TIC), it is appropriate to apply the correction to the data as is. For multi-channel detection, such as the full range of *m/z* with GC-MS, the correction can be performed iteratively on each individual *m/z* as the background fluctuation may be mass channel dependent. Doing so benefits chemometric methods that rely on utilizing every *m/z* such as MCR-ALS and the Fisher Ratio method (further discussion later). Furthermore, GC×GC data must be unfolded along the P_M intervals prior to baseline correction.

The simplest form of baseline correction is subtracting the signal of a blank sample from the data on a pixel level, assuming that the background fluctuations due to temperature programming and stationary phase are consistent run-to-run. A method which uses only the sample's data is to leverage the identify regions of noise in the chromatogram as minimum signal values and subtract these values from the entire chromatogram. Often the minimum signal will vary with time and local analyte concentration, thus a rolling-ball minimum baseline correction method can be implemented, which records the minimum signal value in a small region defined as the ball size. This ball is then iteratively shuffled down the chromatogram, recording the minimum signal as it goes, defining the minimum signal and thus baseline for the whole

chromatogram [23]. This method works well with GC×GC data, where defining the ball size as the size of the P_M results in defining the minimum values for the chromatogram on a modulation-by-modulation basis.

While these methods are useful and perform well when correcting minor baseline fluctuations, more aberrant fluctuations require advanced baseline fitting approaches. Eilers published a method which estimates the baseline with a penalized asymmetric least squares algorithm with user defined parameters to fit a high-order polynomial using a to each data point in an effort to isolate non-analyte signal [24]. Because of the tunable parameters the model requires careful optimization to ensure that the analyte signal is properly preserved, and the baseline is not overfit, which would result in the loss of important chemical information.

1.2.2.2. Normalization

When analyte concentration is to be determined across replicates of a sample or between samples of similar composition, it is important to use normalization methods to minimize extraneous analyte signal variation whether the source of variation be due to sample preparation or injection volume. The most common and popular method of sample normalization is the internal standard method, whereby a non-native analyte is spiked into each sample at a known concentration prior to further sample preparation [25–28]. The method is completed by normalizing all samples to give the internal standard the same signal across all samples. This method assumes that each analyte in the sample can be correctly normalized with one internal standard, occasionally different sections of the chromatogram may be subject to localized variations, requiring the use of multiple internal standards. For the method to work the internal standard peak must either be chromatographically resolved or mathematically resolvable from other peaks. These constraints may make it difficult to apply to samples of unknown

composition, thus one can select an internal standard after exploratory analysis or utilize alternative methods of normalization.

When the internal standard method is not viable, one may alternatively use the sum normalization method, which sums up all the signal in each sample to use as a normalization factor, by which all signal in the chromatogram is divided [29–32]. The sum normalization method relies on the assumption that all samples are of similar composition and have similar analyte response factors such that a given volume of each sample should have similar total signal. While this method does not retain quantitative information as well as the internal standard method it does preserve relative signal information useful for qualitative comparisons [31,32]. Given this limitation, application of sum normalization to heterogenous datasets may result in unnecessary additional signal variation [29]. To solve the normalization of ^{1}H-NMR metabolomic data the probabilistic quotient normalization procedure was suggested by Dieterle et al. [33]. The method functions by identifying the median dilution value between a quotient of the chromatogram and a reference chromatogram and using this as the normalization factor for the entire sample. Thus, when applying normalization to data sets it is important to carefully consider the nature of your samples and the analysis goal when deciding what normalization technique to apply.

1.2.2.3. Retention time alignment

Variation in chromatographic parameters and conditions, such as column degradation or injection method, can result in shifts in retention times from day to day and sometimes run to run. Some chemometric methods rely on the data having reproducible retention times between samples thus it is important to have protocols to align chromatograms to ensure optimized performance of these methods. Popular alignment methods range from simple chromatogram

shifts with retention time locking to more robust algorithms such as piecewise alignment and correlated optimized warping (COW) [34–39]. These approaches, however, are difficult to implement with GC×GC data since individual ^{2}D peaks in a modulated ^{1}D peak may be aligned unevenly, introducing variability that is difficult to account for. While there have been algorithms to apply retention time alignment to GC×GC data specifically, it may be easier to simply mitigate the retention time shifting through binning.

Binning a chromatogram involves summing up data in specified regions, or "tiles" spanning specified length on both separation dimensions, of the chromatogram resulting in a chromatogram that appears coarse with many of the fine details ironed out. With a tile size that is slightly bigger than the size of the average peak, many minor retention time shifting effects can be simply summed up, eliminating the need to perform retention time alignment. Furthermore, it has been demonstrated that binning a chromatogram gives comparable chemometric results to raw chromatograms and in many cases improves the result [39]. However, a single tile grid binning system is often inadequate for chemometric discovery techniques as peaks may be split between tiles. Thus, a quad tile-based system, first described by Marney et. al in 2013, which utilizes four tile grids with half tile offsets in each dimension can be implemented to be sure that one of the four tile grids optimally encapsulates each peak (Fig. 1.4) [40]. This system has been applied to numerous other chemometric discovery techniques from F-ratio analysis to variance ranking and pairwise analysis [23,40–43].

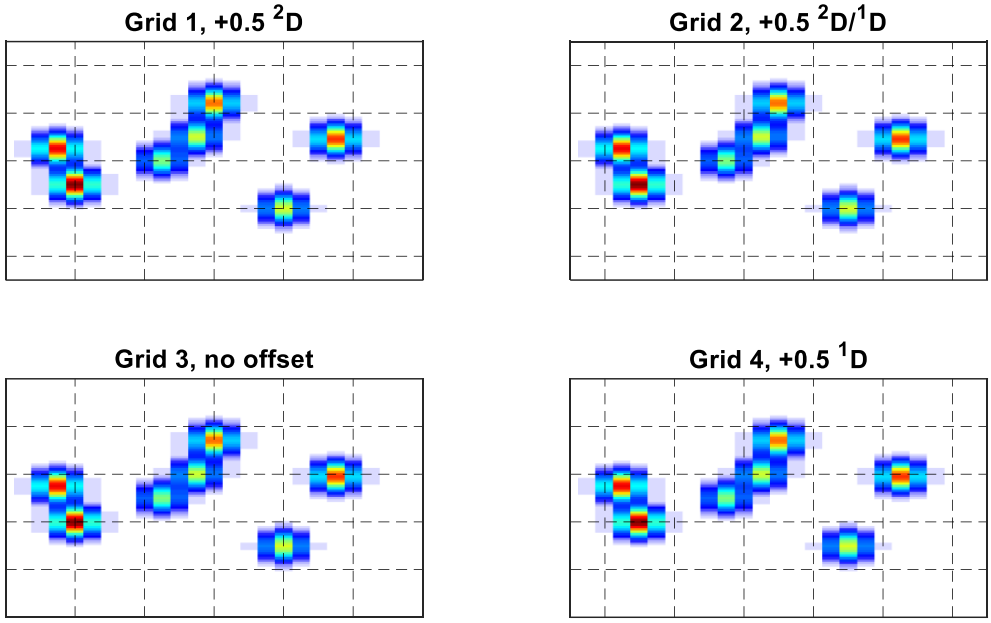


Figure 1.4. Demonstration of the quad-grid tiling system. The four grids use no tiling offset, offset by half a tile width in the first dimension, by half a tile width in the second dimension, and a grid where both offsets are applied. This ensures that for any analyte present, there is a grid that optimally encapsulates its peak volume.

1.2.3. Targeted Methods

Analysis of GC×GC data is generally classified as either targeted or non-targeted analysis based upon the analytical objective. Targeted analysis methods involve confirming the identification of targeted compounds followed by their quantification in samples when the analyte of interest is known *a priori* and are used for routine analysis in many industries and applications. However, it may not be straightforward to quantify known analytes due to coelution with interfering compounds. To address this, deconvolution (*aka* decomposition) algorithms are implemented to allow the analyst to both identify and quantify targeted analytes from component signals. Example algorithms that are commonly used are MCR-ALS and PARAFAC.

1.2.3.1. Decomposition Methods

MCR-ALS is a peak deconvolution algorithm that purifies two-way chromatographic data into its substituent components [21,44,45]. For 1D-GC-MS the two dimensions are the separation dimension and the mass spectrum dimension. For GC×GC-MS data, the separation dimensions must be unfolded into one, with the remaining dimension representing the mass spectrum. Typically, a small region around the target region is submitted to limit the number of interfering components included. The algorithm works by first making initial estimates for each of the *n* components dimensions, then applying an iterative alternating least squares algorithm to improve the estimates and fit the modeled components to the data until convergence criteria are met (number of iterations, minimized residuals, lack-of-fit improvement, etc.) [21]. In the absence of *a priori* knowledge of the number of components, the analyst may iteratively fit models with successively more components until signs of overfitting occurs (i.e. peak-splitting between components) or apply rank estimation techniques such as evolving factor analysis [46]. MCR-ALS relies on two assumptions, one, that the data is a linear combination of its components and the contribution due to chemical noise is minimal, and two, that the components are bilinear, that is having unique profiles in each dimension. Additionally, separate components must be sufficiently unique from each other for successful modeling. Constraints can be applied to MCR-ALS to aid in model fitting, such as non-negativity, unimodality, or local rank [21].

MCR-ALS can be applied either on a single sample or multiple sample basis. For single samples the decomposition yields pure spectral information for each component, which in the case of MS detection allows for the components to be identified through the use of a NIST mass spectrum library search. The components concentrations may also be quantified relative to each other. For multiple samples, the samples may be concatenated together along the separation axis,

allowing for variation in concentration between samples to distinguish between components. Furthermore, in this way standards may be concatenated to the sample allowing accurate quantification of the target analytes by their relation to the standards [47].

PARAFAC is a peak deconvolution algorithm that also utilizes alternating least squares for its model. It requires the data to be three dimensional and can model each component of trilinear data as the outer product of their three fixed vectors[22]. For GC×GC-MS data, PARAFAC decomposes the chosen chromatographic region into *n* components as the sum of all analyte and background components, with three vectors describing the ^{1}D peak profile, ^{2}D peak profile, and mass spectrum profile for each component [41,44,48–50]. In addition to allowing the analyst to isolate the mass spectrum of each component, PARAFAC can also isolate the background component(s), clearing up the analyte component profiles. Several constraints can be placed on the three dimensions to help improve the model convergence. Typical constraints would include unimodality or nonnegativity for both separation dimensions and nonnegativity for the mass spectrum dimension.

Unlike MCR-ALS, PARAFAC requires the data to be sufficiently trilinear, which adds an extra layer of complexity for GC×GC-MS data as small ^{2}D retention time shifts (e.g., due to temperature programming) can cause peaks between modulations to shift slightly. Strategies to address this could involve aligning the modulations should the shifting be great, or utilizing multiple samples and reshaping the data so each separation dimension is represented as one with mass spectra and samples making up the other two dimensions. Recently, the successor to PARAFAC, referred to as PARAFAC2 was introduced which loosened the trilinearity requirement, allowing it to handle retention time misalignment more efficiently [51–53].

1.2.4. Non-targeted methods

While targeted methods are useful for identifying and quantifying known analytes of interest in sample chromatograms, non-target methods aim to discover interesting chemical features and patterns across a wide range of chromatograms. Non-targeted methods are classified as either unsupervised or supervised depending on whether they utilize class information about the samples. Supervised methods (discussed later) which utilize class labels are suitable for regression and making specific inquiries on datasets such as cause and effect studies. Unsupervised methods do not require any *a priori* information about the samples. Therefore, unsupervised methods are well suited to exploratory analysis to identify any patterns or outliers among the samples. These methods are generally applied at the beginning of a chemometric inquiry when visualizing the data to identify such patterns can inform further analysis efforts. A common unsupervised method, principal component analysis (PCA), is applied to visualize and identify features which discriminate between samples.

PCA is a very important method for exploratory analysis as it reduces the chromatographic data to the sets of variables which describe a majority of the variance seen in the dataset [54–57]. This reduction is achieved by finding the axes of maximum variance in the dataset and projecting the data onto these new linearly orthogonal axes called principal components (PC). For analysis of a given dataset the number of PCs is equal to the number of samples, which are then ranked by the amount of variance they explain. Therefore, PC1 explains the maximum variance in the data and PC2 explains the maximum variance not explained by PC1, and so on and so forth. Generally, the first two PCs are the only ones used for visualizing the data since variance beyond these two is less likely to be important.

The outputs from PCA are called scores and loadings and are the metrics used to evaluate how samples relate to each other. The scores are how each sample ranks in each PC. These scores can be plotted against each other (i.e., PC1 on the x-axis versus PC2 on the y-axis) and allow a visual interpretation as to how the samples in the dataset relate to each other. From this PC scores plot the analyst can determine whether there are any obvious clusters of samples, patterns or outliers that present themselves. The quality of sample clusters can be quantified through metrics such as the degree-of-class separation (DCS) [58]. These differences that present themselves may be due to real differences between the samples or due to other sources of variation such as instrumental noise or time series effects (samples collected on different days) so this visualization step is important in isolating these effects and mitigating their influence before further steps of the analysis process. Next, the PC loadings describe how each variable interacts with each PC and detail which chemical features (variables) are important for describing the difference seen on the given PC. For example, if whole chromatograms are used as PCA inputs then the loadings variable can be interpreted chromatographically, detailing which peaks are important for discriminating samples in that PC [59–61].

All forms of variance will be picked up by PCA, thus it is important to apply proper pre-processing to the data to ensure chemically meaningful results. In addition to baseline correction and signal normalization, any retention time shifting should be mitigated. If the shifting is not addressed, then the first few PCs will capture variance due to shifting rather than sample related differences [62]. This issue can be addressed either through retention time alignment as previously mentioned or reduction of the input data to peak tables, which minimizes retention alignment effects. However, use of peak tables as PCA inputs is susceptible to missing features that overlap with each other. Mean centering can be applied to the dataset to obtain a PC scores

plot that is centered around zero, but otherwise does not affect the samples separation on the PC space relative to each other. Even with proper chromatographic pre-processing the PCA results may not show any separation between samples. This could be because there is no meaningful differences or that the chemically meaningful variance is dwarfed by the remaining signal. To overcome this, the analyst can implement feature selection techniques, most commonly accomplished through the use of supervised non-targeted analysis methods.

1.2.4.1. Supervised Methods

While unsupervised approaches are appropriate for initial investigations into a chromatographic data set, supervised approaches are well suited for studying cause and effect experiments by leveraging *a priori* information, specifically known sample class membership. Supervised algorithms utilize target variables such as class labels or independently measured sample properties to discover features, build regression models, and/or classify samples. Feature discovery, also known as feature selection, finds a subset of the original chromatographic data that is highly correlated with the target variable(s). For chromatographic data sets, Fisher ratio (F-ratio) analysis is typically used to discover class-distinguishing analytes. It is important to note that unsupervised methods like PCA are commonly used to visualize the results obtained from non-targeted, supervised feature selection methods. Along with identifying significant analytes (feature selection), methods used for property prediction and sample classification fall under the umbrella of supervised analysis techniques. Note that in this context, a “property” refers to either a chemical or physical quantity that was collected separately from the chromatographic data set. The most common property prediction method is partial least squares (PLS) regression, which develops a multivariate calibration model to discover which analytes

correlate with the sample property that is being modeled. This section will detail the principles and preprocessing considerations for each supervised technique.

1.2.4.2. Feature Selection

F-ratio analysis is a popular feature selection technique for chromatographic data because it inherently provides data reduction, focusing the overall data analysis before performing further targeted and non-targeted chemometric methods. This feature selection method utilizes the analysis of variance (ANOVA) statistical hypothesis test, which compares the variance of observations and discovers significant differences between groups. The total variance, defined as the squared standard deviation, can be partitioned into two contributions: variance *between* classes of samples and *within* classes of samples. The between class (BC) variance, which describes how each class mean varies from the grand mean, is defined as

$$\sigma_{\mathrm{BC}}^2 = \frac{1}{k-1}\sum(\bar{x}_{\mathrm{i}} - \bar{x})^2 n_{\mathrm{i}} \tag{8}$$

where k is the number of classes, n_{i} is the number of measurements in the i^{th} class, $\bar{x}_{\mathrm{i}}$ is the mean of the i^{th} class, and $\bar{x}$ is the grand mean. The within class (WC) variance, which indicates how much each measurement varies from its class mean, is

$$\sigma_{\mathrm{WC}}^2 = \frac{1}{N-k}\sum\sum(x_{ij} - \bar{x}_i)^2 \tag{9}$$

where N is the total number of measurements, and x_{ij} is the j^{th} measurement of the i^{th} class. Finally, the F-ratio is then obtained by taking the ratio of these two quantities:

$$F - ratio = \frac{\sigma_{\mathrm{BC}}^2}{\sigma_{\mathrm{WC}}^2} \tag{10}$$

The results from the F-ratio analysis are compiled in a "hit list," which ranks the F-ratio values in descending order. The analyst then mines the hit list in a top-down approach, identifying and quantifying peaks with larger F-ratios, since a high F-ratio generally corresponds to class-distinguishing analytes. A class-distinguishing analyte is an analyte whose concentration is statistically different between classes, which is typically based upon a *t*-test having a *p*-value < 0.05 (95% confidence limit). Thus, the results of the *t*-test show that the concentration ratio for a given analyte hit between classes sufficiently differs from one. These class-distinguishing analytes are commonly referred to as true positives. However, this data mining approach can be hindered by the presence of both false positives and negatives. A false positive refers to the discovery of an analyte that is not statistically different between classes, while a false negative is the inability to discover a class-distinguishing analyte. The presence of false positives and/or negatives can be present in all implementations of F-ratio analysis (peak table, pixel-based, and tile-based). Therefore, along with discussing the different approaches for F-ratio analysis, strategies to reduce the false discovery rate will be introduced in this section.

Peak table-based [63–65] and pixel-based [61,66,67] are the most straightforward approaches for F-ratio analysis. A peak table-based approach typically uses the instrument software to baseline correct and quantify signals for each peak in the chromatograms. These peak tables are then aligned based upon user-defined time windows and mass spectrum match criteria prior to F-ratio analysis [63]. While this approach aids in limiting the pool of potential features to only those with measurable signals, it is important to note that this approach is limited by the capability of the peak finding software and thus, potentially class-distinguishing analytes with low *S/N* can be missed. A pixel-based approach, on the other hand, calculates an F-ratio for every data point in the chromatogram (i.e., at every 2D retention time on every detector channel).

The advantage of a pixel-based approach is that it utilizes the entire data set; however, the number of false positives can increase rapidly since 2D retention time alignment on pixel-level data performs at a limited level of success. Even if misalignment is mitigated to some extent by 2D retention time alignment, random detector fluctuations can artificially inflate F-ratios and increase the number of false positives [40].

To address this challenge, a new tile-based approach for F-ratio analysis of comprehensive 2D chromatographic data was developed for GC × GC-TOFMS data [23,40] (Fig 1.4) demonstrated in Figure 1.5 with 8 chromatograms with known class membership (Fig. 1.5A). Here, the chromatogram is divided into regular repeating rectangular sections called "tiles," where the encapsulated signal is summed (binned) to a single value. This tiling procedure not only reduces the overall size of the chromatogram but also reduces the effect of retention time shifting. To ensure that peaks are not split between tiles, F-ratio analysis is performed using four different tiling grids, where each grid is offset from the original tile grid by half a tile width in either dimension. A four-grid tile scheme ensures that each peak in the chromatogram is adequately captured by at least one tile. The resulting tiled data sets are then compared using F-ratio analysis on a per-*m/z* basis. These F-ratios can be visualized as a distribution as shown in Figure 1.5B to determine the extent of class-distinguishing features. Next, a pin approximating the location of the peak for each hit is calculated using the maximum signal difference for each tile, and pins with similar retention times are clustered together. This pinning and clustering step ensures that each feature in the hit list is only represented by one pin with the highest F-ratio. Examples of these pins are shown in Figure 1.5C as red circles plotted over the GC×GC TIC chromatogram at retention time for the five highest F-ratio features from the dataset. Viewing the

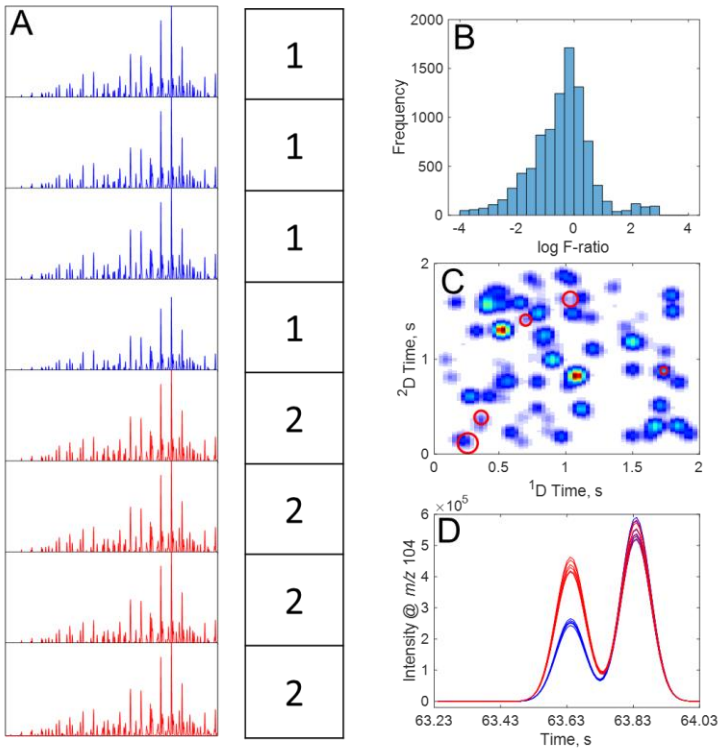


Figure 1.5. Examples of F-ratio analysis results as demonstrated on a 4 vs 4 class comparison. (A) Examples of sample chromatograms colored by class. (B) F-ratio distribution resulting from the tile-based F-ratio analysis. (C) GC×GC TIC chromatogram of one sample with the pin locations of the top 5 F-ratio hits plotted as red circles. (D) One-dimensional chromatogram of the pin location for the top F-ratio hit, demonstrating the change in concentration that was discovered by the analysis.

top F-ratio hit for all 8 samples as a one-dimensional chromatogram in Figure 1.5D demonstrates the feature discovered was a peak whose concentration increased two-fold between classes.

Proper reduction of false positives and false negatives using tile-based F-ratio analysis first requires optimization of the tile size for both dimensions, ^{1}D and ^{2}D. Ideally, the tile dimensions should encompass the average peak width along both dimensions with any misalignment [23,68]. If the tile size is too large, then interferent signals could mask the signal of true positives at low concentration [68]. Conversely, the hit list generated after using a tile smaller than optimal can result in numerous redundant hits, which may be interpreted as false positives [68]. Other strategies to reduce the presence of false positives require optimization of different parameters involved in calculating F-ratios and ranking the discovered features. Reaser

et al. demonstrated that the occurrence of false positives can be mitigated by using a *S/N* of 10 and ranking F-ratios based on the average of the top 10 *m/z* [69]. Likewise, Sudol et al. showed that in part by ranking the hit list using only the *m/z* that produced the top F-ratio for a given hit allowed for the discovery of features at concentrations as low as 1 ppm [68]. For data sets consisting of control and treatment classes, F-ratios can be calculated using solely the variance of the control class in the denominator [70]. This calculation, termed control-normalized F-ratio, was shown to discover class-distinguishing features that were initially missed by traditional F-ratio analysis due to their non-uniform/high variance in the treatment class [70]. An analyst can evaluate the effectiveness of changing these different parameters on the true and false positive rates with receiving operator characteristic (ROC) curves [69]. The area under the curve (AUC) for the ROC curve can then quantify the optimization of different parameters, where higher AUCs indicate more true positives appear at the top of the hit list [69].

Along with optimizing the calculation and ranking parameters, a reduction of the hit list can be achieved by selecting an F-ratio threshold, deeming any feature above the threshold to be important for further analysis. Traditionally, a manual cutoff can be selected by determining the point where the frequency of false positives increases. However, this method is both time-consuming and subjective toward the false positive tolerance limit defined by the analyst. An F-critical threshold can also be applied, but this approach generally picks small cutoffs that keep too many false positives. Null distribution analysis has been demonstrated to be a more robust method for cutoff determination [23]. This method develops false positive F-distributions from repeated F-ratio analyses on classes built from redistributing samples evenly and randomly into classes. A distribution of F-ratios from the null comparisons is developed at a desired null probability. The F-ratio threshold is then chosen by identifying the F-ratio that corresponds to a

user-defined confidence level. Ultimately, null distribution analysis allows for the objective determination of an F-ratio threshold, which is unique and accurate because it naturally considers the underlying noise for a given data set.

Recent developments have utilized the tile-based framework to apply different metrics to achieve different results. The 1v1 rank metric (*RM*) was introduced to enable discovery of distinguishing features between two chromatograms, vastly reducing the amount of data needed compared to the F-ratio method, which needs at least three samples per class. The 1v1 *RM* is defined as:

$$RM = \frac{abs(B-A)}{B+A} * 100 \tag{11}$$

where A and B are the signal of each sample respectively. This *RM* is unitless and ranges from 0-100 depending on how different the two signals are from each other. Studies using 1v1 analysis have demonstrated the *RM* performs better than alternative pairwise comparison techniques such as the Jensen-Shannon divergence, and/or with nearly the same efficacy as the F-ratio method [43,71]. Furthermore, the tile-based framework has been applied to unsupervised analysis with the calculation of the RSD^2 (squared relative standard deviation), a metric which considers the variance among all samples regardless of class [42]. This method, called variance rank initiated unsupervised sample indexing (VRI-USI) allows feature selection of potentially class distinguishing features without knowledge of class membership and has been demonstrated to aid in discovery of subclasses within datasets [42,72].

1.2.4.3. Regression

While the discovery of chemical compounds is very important, it is also true that their quantitative relation to either physical properties or chemical properties of a sample set can be equally important. Establishing this relationship is accomplished through chemometric regression. Partial least squares (PLS) is a multi-variate regression algorithm that correlates the sample information in a data matrix (**X**) to another data matrix (**Y**), which may be a single-column vector or multi-column matrix (i.e. n-PLS) [73]. PLS works by applying PCA individually to the **X** and **Y** block. For PCA, the direction of the loadings for each matrix is chosen to maximize the variance within their respective matrix. However, the direction of loadings for PLS is chosen to maximize the covariance between the **X** and **Y** block. This approach is commonly used to model the relationships between chromatographic data (**X**) and measured properties of samples (**Y**) such as bulk physical properties like viscosity, density, and heat of combustion or to other chemical properties such as relating biomarker concentration to IgG concentrations in blood serum [74]. Once a model has been established it can be used to predict the **Y** block response of new **X** block data.

For application with GC×GC data the chromatograms must be unfolded into a vector and stacked in a matrix (**X**) with the rows comprising separate samples and the columns containing the chromatographic variables. The data to be predicted (**Y**) should then be set up as a vector with the same number of rows as the **X** block and as many columns as there are properties to predict. A single PLS model can be constructed just using these **X** and **Y** blocks, however there is no indication as to well the model reacts to data not part of the original **X** training set. This can be addressed by either withholding data to be used as a validation set to evaluate how close the model predicts the **Y** values or by running cross-validation which performs the same check using

all of the training data. Leave-one-out cross validation (LOOCV) is a method where a PLS model is constructed using all but one sample. The resulting model is used to predict the value of the withheld sample. This process is then repeated leaving each sample out one at a time and predicting the values until each sample has been predicted once. These cross-validated predictions then give a better gauge of model performance. In addition to the calibration and prediction data, PLS also outputs linear regression vectors (LRV) which detail how each variable relates to predicting the modeled property. For GC×GC data these LRVs can be interpreted chromatographically like PCA loadings, allowing the analyst to learn for example whether variables are positively or negatively correlated with a given property [75,76].

1.3. Overview of fuel analysis, its importance to defense industries, and chemometrics place in fuel analysis

Given GC×GC-MS is an ideal approach for analyzing the chemical composition of volatile complex mixtures, this makes GC×GC-MS well suited to characterize hydrocarbon fuels such as kerosene-based propellants, such as jet fuel Jet-A and JP-5 and rocket fuel RP-1 and RP-2. Industries that are interested in propellant chemistry are fuel manufactures, airlines, and defense industries. Fuel manufactures, airlines, and government regulatory agencies are gradually replacing the existing fuel in use today with greener alternatives such as biofuel and sustainable aviation fuel (SAF) with the US Department of Energy setting a goal for companies to create fuel that achieves a minimum of 50% reduction in lifecycle greenhouse gas compared to conventional fuels to meet 100% of aviation fuel demand by 2050 [77]. GC×GC characterization of such alternative fuels will be important for understanding and furthering their development [78–84]. For defense industries the performance and reliability of propellant fuels is of utmost importance necessitating methods to analyze fuels for trace contaminants and

modelling of properties. For example, a large area of hydrocarbon fuel research is understanding the dynamics involved in carbonaceous deposit formation, an area that robust chemical composition information will undoubtedly aid [85–88]. Whether it be for characterization and property prediction of biofuels and sustainable aviation fuels, or the development of better rocket fuels, GC×GC and chemometric analysis play an important role and their continued development is important and at the heart of the motivation for this dissertation.

1.4. Overview of following chapters

1.4.1. Statistical inference of mass channel purity from Fisher ratio analysis using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry data

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

two classes (eg., 80 ppm versus 40 ppm) was applied to all mass channels (*m/z*) in the F-ratio spectrum for each hit. A lack-of-fit (*LOF*) metric was utilized as a measure of peak purity and combined with F-ratio and p-values to study the relationship of each of these metrics with *m/z* purity. Application of a *LOF* threshold coupled with a p-value threshold yielded a subset of the "most pure" *m/z* for each of the 15 spiked analytes, evident by the low deviations (< 5%) in S-ratio relative to the known spiked-in concentration ratio. A key outcome of this study was to demonstrate the isolation of pure *m/z* without the need for higher level signal decomposition algorithms such as MCR-ALS and PARAFAC.

1.4.2. Class comparison enabled mass spectrum purification for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry

Tile-based Fisher ratio (F-ratio) analysis is emerging as a versatile data analysis tool for supervised discovery-based experimentation using GC × GC-TOFMS. None the less, analyte identification can often be marred by poor 2D resolution and low analyte abundance relative to overlapping compounds. Linear algebra-based chemometric methods, in particular MCR-ALS, PARAFAC, and PARAFAC2, are often applied in an effort to address this situation. However, these chemometric methods can fail to produce an accurate spectrum when the analyte is at low 2D resolution and/or in low relative abundance. To address this challenge, we introduce class comparison enabled mass spectrum purification (CCE-MSP), a method that utilizes the underlying requirement for signal consistency of the background interference compounds between the two classes in the F-ratio analysis to purify the mass spectrum of the analyte hits. CCE-MSP is validated using a dataset obtained for a neat JP-8 jet fuel spiked with 14 sulfur containing compounds at two levels (15 ppm and 30 ppm), using the p-value and *LOF* for each analyte hit as consistency metrics. A purified mass spectrum was produced for each spiked

analyte hit and their mass spectrum match value (MV) was compared to the MV obtained by MCR-ALS, PARAFAC, and PARAFAC2. The resulting MV for CCE-MSP were found to be as good or better than these chemometric methods, e.g., for 2-butyl-5-ethylthiophene with an analyte-to-interference relative signal abundance of 1:87 and a 2D resolution of 0.2, CCE-MSP produced a MV of 831, compared to 476 for MCR-ALS, 403 for PARAFAC, and 336 for PARAFAC2. CCE-MSP is also extended to obtain the purified spectrum for more than one analyte, e.g., two analyte hits in overlapping hit locations. The spectra produced by CCE-MSP can also be utilized as estimates to facilitate quantitative signal decomposition using MCR-ALS.

1.4.3. Using solid-phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: comparative analysis of aerospace fuel composition

Tile-based Fisher ratio (F-ratio) analysis of GC×GC-TOFMS data is a powerful, supervised discovery methodology for pinpointing sample class-distinguishing analytes between two or more sample classes. Herein, we extend this analytical methodology to focus upon specific chemical groups in kerosene-based aerospace fuel using solid-phase extraction (SPE). Treating samples with SPE removes specific compounds depending on the SPE stationary phase (i.e., silica), creating an altered “pass” sample, identical to the original “neat” sample except for the extracted compounds. Application of F-ratio analysis to the neat samples against the pass samples provides global discovery with a numerically sorted hit list of all analytes affected by the SPE procedure. Sections of GC×GC-TOFMS data from the top analyte hits are reconstructed to form a “stitch” chromatogram to visualize the sample class-distinguishing compounds, revealing excellent agreement with the extract chromatogram. Additionally, utilizing the four-grid tiling scheme developed for tile-based F-ratio analysis, we demonstrate a tile-based pairwise

analysis method, referred to as 1v1 analysis, to discover analytes that differ in concentration between two fuel chromatograms. Application of 1v1 analysis is highly efficient since replicates do not necessarily need to be run on the GC×GC-TOFMS instrument, which is beneficial for sample-limited applications. The 1v1 analyses discovered most of the same features as F-ratio analysis, ranging from 69 to 81% of the features discovered by F-ratio analysis while requiring one-sixth the data. Lastly, the overall methodology is applied to three candidate rocket fuels to better understand the compound class-distinguishing differences. The separate hit lists produced for high-concentration bulk hydrocarbon differences and low-concentration level polar compound differences provided valuable insight into these candidate rocket fuels.

1.4.4. Investigating analyte breakthrough under non-linear isotherm conditions during solid phase extraction facilitated by non-targeted analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry

Solid phase extraction (SPE) sample preparation for the analysis of complex organic mixtures is often applied assuming all analytes of interest will preconcentrate on the stationary phase. This assumption ignores the reality that extraction is a dynamic interactive process and a diverse range of affinities for the stationary phase will result in equally diverse breakthrough volumes due to competitive binding processes. To study this dynamic interactive process, and further to take advantage of it, we extracted a JP-8 jet fuel spiked with 40 ppm of a polar compound mix with silica and alumina SPE cartridges and analyzed sequential extracted fractions of the fuel to both assess the shifting chemical landscape present in the extraction and the impact of both SPE stationary phases on this process. Tile-based 1v1 comparative analysis (a recently reported extension of tile-based Fisher ratio analysis) was used to discover the (polar)

compounds whose concentrations change between extracted fractions, discovering 27 compounds extracted with alumina and 21 compounds extracted with silica with at least a factor of 2 change in concentration from the neat sample relative to the first 1 mL pass fraction sample. These compounds were quantified in each fraction to construct concentration ratio profiles, defined as the concentration ratio for a given SPE fraction per analyte compound relative to the analyte concentration in the neat fuel, for which the extraction behavior for each analyte could be assessed. These analyte compounds were found to breakthrough at different rates, with some analytes remaining on the column indefinitely (until extracted with a subsequent polar solvent) and other analytes eluting before the extraction is complete. Furthermore, in comparison of the effect of selected stationary phase, alumina was found to retain oxygen-containing phenolic compounds to a greater extent than silica. PCA was used to analyze the concentration ratio profiles of the various trace analytes in the JP8 fuel (phenols, indoles, etc.) in the context of their stationary phase affinity (alumina or silica) and competitive binding behavior.

1.4.5. GC×GC-TOFMS and Chemometric Data Analysis Applied to Aerospace Fuels: Modeling Thermal Stability of Complex Hydrocarbon Mixtures

Ensuring reliability, reusability, and operability of propulsion systems motivates quantitative connections between fuel composition, properties, and performance. In turn, the need for predictive models places greater emphasis on accurate fuel property measurements and detailed compositional information, especially for multicomponent fuels like RP-1 and RP-2. To facilitate informed decisions regarding composition, specification, and fit-for-purpose behavior of complex fuels, we apply comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) and multivariate “chemometric” data analysis methods to better understand how fuel performance depends upon chemical composition in the

challenging context of fuel thermal stability. A primary objective is modeling the relationship between chemical composition of complex fuels and their application-specific thermal performance acquired in the Compact Rapid Assessment of Fuel Thermal Integrity (CRAFTI) experimental platform. The availability of multivariate thermal integrity datasets – pressure drop increase and carbonaceous deposit formation acquired in CRAFTI for several rocket and advanced aerospace fuels – offers an ideal opportunity to utilize powerful chemometric methods to advance fundamental composition-performance relationships in the pursuit of predictive models. We present a feature selection method combining variance ranking, partial least squares (PLS), and RreliefF to select chemical features from a dataset fusing GC×GC-TOFMS chromatograms of neat and solid phase extracts of fuels which optimally correlate to carbonaceous deposit data obtained via the CRAFTI testing platform. Variance ranking was used to pre-select chromatographic features and RreliefF was used rank PLS results based upon how well they correlated to the physical data and find an optimum correlation weight threshold for improving the model. This method achieved lower cross-validation modeling errors for each form of carbon investigated. The linear regression vectors (LRV) from the final PLS models were used to discover the chemical compounds highly correlated with the deposition of carbon in heated fuel lines. These compounds were found to be mainly diaromatic compounds such as naphthalene, oxygen-containing compounds such as phenols and alcohols, and nitrogen-containing compounds such as quinoline.

1.5. References

[1] K. Robards, P.R. Haddad, P.E. Jackson, Gas Chromatography, in: Princ. Pract. Mod. Chromatgor. Methods, Academic Press, San Diego, 1994.

[2] J.C. Giddings, Unified Separation Science, John Wiley & Songs, Inc., 1991.

[3] D. Guillarme, D.T.T. Nguyen, S. Rudaz, J.L. Veuthey, Recent developments in liquid chromatography—Impact on qualitative and quantitative performance, J. Chromatogr. A. 1149 (2007) 20–29. https://doi.org/10.1016/J.CHROMA.2006.11.014.

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

[5] O. Coskun, Separation techniques: Chromatography, North. Clin. Istanbul. 3 (2016) 156. https://doi.org/10.14744/NCI.2016.32757.

[6] A. Bele, An overview on thin layer chromatography, (2011). www.ijpsr.com (accessed December 29, 2022).

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

[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. https://doi.org/10.1016/J.CHROMA.2007.07.028.

[9] D.C. Harris, Statistics, in: Quant. Chem. Analyisis, Ninth, W.H. Freeman and Company, New York, 2016: pp. 64–89.

[10] C.F. Poole, Ionization-based detectors for gas chromatography, J. Chromatogr. A. 1421 (2015) 137–153. https://doi.org/10.1016/J.CHROMA.2015.02.061.

[11] F. Rey-Stolle, D. Dudzik, C. Gonzalez-Riano, M. Fernández-García, V. Alonso-Herranz, D. Rojo, C. Barbas, A. García, Low and high resolution gas chromatography-mass spectrometry for untargeted metabolomics: A tutorial, Anal. Chim. Acta. 1210 (2022) 339043. https://doi.org/10.1016/J.ACA.2021.339043.

[12] S.K. Jeon, D. Kwon, S. Lee, Identification of weathered multiple petroleum products in contaminated soils by characterizing unresolved complex mixture hump in gas chromatograph data, Sci. Total Environ. 607–608 (2017) 42–52. https://doi.org/10.1016/J.SCITOTENV.2017.06.251.

[13] K.M. Sharif, S.T. Chin, C. Kulsing, P.J. Marriott, The microfluidic Deans switch: 50 years of progress, innovation and application, TrAC Trends Anal. Chem. 82 (2016) 35–54. https://doi.org/10.1016/J.TRAC.2016.05.005.

[14] D.R. Deans, Use of heart cutting in gas chromatography: A review, J. Chromatogr. A. 203 (1981) 19–28. https://doi.org/10.1016/S0021-9673(00)80278-2.

[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.

https://doi.org/10.1093/chromsci/29.6.227.

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

[17] N.D. Spadafora, S. Mascrez, L. McGregor, G. Purcaro, Exploring multiple-cumulative trapping solid-phase microextraction coupled to gas chromatography–mass spectrometry for quality and authenticity assessment of olive oil, Food Chem. 383 (2022) 132438. https://doi.org/10.1016/J.FOODCHEM.2022.132438.

[18] 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. https://doi.org/10.1016/J.TALANTA.2022.123301.

[19] 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. https://doi.org/10.1016/J.CHROMA.2021.462735.

[20] A.C. Jimenez, C.A. Heist, M. Navaei, C. Yeago, K. Roy, Longitudinal two-dimensional gas chromatography mass spectrometry as a non-destructive at-line monitoring tool during cell manufacturing identifies volatile features correlative to cell product quality, Cytotherapy. 24 (2022) 1136–1147. https://doi.org/10.1016/J.JCYT.2022.06.001.

[21] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemom. Intell. Lab. Syst. 140 (2015) 1–12. https://doi.org/10.1016/j.chemolab.2014.10.003.

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

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

[24] P.H.C. Eilers, Parametric Time Warping, Anal. Chem. 76 (2004) 404–411. https://doi.org/10.1021/AC034800E/ASSET/IMAGES/MEDIUM/AC034800EE00007.GIF.

[25] J.M. Hogan, R.A. Engel, H.F. Stevenson, A Versatile Internal Standard Technique for the Gas Chromatographic Determination of Water in Liquids, Anal. Chem. 42 (1970) 249–252. https://doi.org/10.1021/AC60284A033/ASSET/AC60284A033.FP.PNG_V03.

[26] R.J. Strife, J.R. Simms, M.P. Lacey, Combined capillary gas chromatography/ion trap mass spectrometry quantitative methods using labeled or unlabeled internal standards, J.

Am. Soc. Mass Spectrom. 1 (1990) 265–271. https://doi.org/10.1016/1044-0305(90)85044-M.

[27] R.H. Liu, G. Foster, E.J. Cone, S.D. Kumar, Selecting an appropriate isotopic internal standard for gas chromatography/mass spectrometry analysis of drugs of abuse -- pentobarbital example, J. Forensic Sci. 40 (1995) 983–989.

[28] M. Fasciotti, T.V.C. Monteiro, A.A. Ferreira, M.N. Eberlin, L.A. Neves, Two-point normalization using internal and external standards for a traceable determination of δ13C values of fatty acid methyl esters by gas chromatography/combustion/isotope ratio mass spectrometry, Int. J. Mass Spectrom. 418 (2017) 41–50. https://doi.org/10.1016/J.IJMS.2016.12.002.

[29] P. Filzmoser, B. Walczak, What can go wrong at the data normalization step for identification of biomarkers?, J. Chromatogr. A. 1362 (2014) 194–205. https://doi.org/10.1016/J.CHROMA.2014.08.050.

[30] J.W. McIlroy, R.W. Smith, V.L. McGuffin, Assessing the effect of data pretreatment procedures for principal components analysis of chromatographic data, Forensic Sci. Int. 257 (2015) 1–12. https://doi.org/10.1016/J.FORSCIINT.2015.07.038.

[31] J. Orzel, B. Krakowska, I. Stanimirova, M. Daszykowski, Detecting chemical markers to uncover counterfeit rebated excise duty diesel oil, Talanta. 204 (2019) 229–237. https://doi.org/10.1016/J.TALANTA.2019.05.113.

[32] J.M. Baerncopf, V.L. McGuffin, R.W. Smith, Association of Ignitable Liquid Residues to Neat Ignitable Liquids in the Presence of Matrix Interferences Using Chemometric Procedures*,†, J. Forensic Sci. 56 (2011) 70–81. https://doi.org/10.1111/J.1556-4029.2010.01563.X.

[33] F. Dieterle, A. Ross, tz Schlotterbeck, H. Senn, F. Hoffman-, Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1 H NMR Metabonomics, (2006). https://doi.org/10.1021/ac051632c.

[34] N. Etxebarria, O. Zuloaga, M. Olivares, L.J. Bartolomé, P. Navarro, Retention-time locked methods in gas chromatography, J. Chromatogr. A. 1216 (2009) 1624–1629. https://doi.org/10.1016/J.CHROMA.2008.12.038.

[35] Y. Li, Q. Ruan, Y. Li, G. Ye, X. Lu, X. Lin, G. Xu, A novel approach to transforming a non-targeted metabolic profiling method to a pseudo-targeted method using the retention time locking gas chromatography/mass spectrometry-selected ions monitoring, J. Chromatogr. A. 1255 (2012) 228–236. https://doi.org/10.1016/J.CHROMA.2012.01.076.

[36] K.J. Johnson, B.W. Wright, K.H. Jarman, R.E. Synovec, High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis, J. Chromatogr. A. 996 (2003) 141–155. https://doi.org/10.1016/S0021-9673(03)00616-2.

[37] K.M. Pierce, B.W. Wright, R.E. Synovec, Unsupervised parameter optimization for automated retention time alignment of severely shifted gas chromatographic data using the

piecewise alignment algorithm, J. Chromatogr. A. 1141 (2007) 106–116. https://doi.org/10.1016/J.CHROMA.2006.11.101.

[38] G. Tomasi, F. Van Den Berg, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, J. Chemom. 18 (2004) 231–241. https://doi.org/10.1002/CEM.859.

[39] 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. https://doi.org/10.1021/AC7024317/ASSET/IMAGES/LARGE/AC7024317F00005.JPEG .

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

[41] 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. https://doi.org/10.1016/j.chroma.2020.461401.

[42] C.N. Cain, P.E. Sudol, K.L. Berrier, R.E. Synovec, Development of variance rank initiated-unsupervised sample indexing for gas chromatography-mass spectrometry analysis, Talanta. 233 (2021) 122495. https://doi.org/10.1016/J.TALANTA.2021.122495.

[43] 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. https://doi.org/10.1021/acs.analchem.2c00223.

[44] H. Parastar, J.R. Radovi, M. Jalali-Heravi, S. Diez, J. Maria Bayona, R. Tauler, Resolution and Quantification of Complex Mixtures of Polycyclic Aromatic Hydrocarbons in Heavy Fuel Oil Sample by Means of GC Â GC-TOFMS Combined to Multivariate Curve Resolution, Anal. Chem. 83 (2011) 9289–9297. https://doi.org/10.1021/ac201799r.

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

[46] M. Maeder, A. Zilian, Evolving factor analysis, a new multivariate technique in chromatography, Chemom. Intell. Lab. Syst. 3 (1988) 205–213. https://doi.org/10.1016/0169-7439(88)80051-0.

[47] L.W. Hantao, H.G. Aleme, M.M. Passador, E.L. Furtado, F.A. de L. Ribeiro, R.J. Poppi, F. Augusto, Determination of disease biomarkers in Eucalyptus by comprehensive two-dimensional gas chromatography and multivariate data analysis, J. Chromatogr. A. 1279 (2013) 86–91. https://doi.org/10.1016/j.chroma.2013.01.013.

[48] 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. 1605 (2019) 460368. https://doi.org/10.1016/J.CHROMA.2019.460368.

[49] 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. https://doi.org/10.1021/ac061710b.

[50] D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity deviation ratio: A new metric for chemometric analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data, Anal. Chim. Acta. 871 (2015) 66–76. https://doi.org/10.1016/J.ACA.2015.02.040.

[51] 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. https://doi.org/10.1002/(SICI)1099-128X(199905/08)13:3/4<295::AID-CEM547>3.0.CO;2-Y.

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

[53] J.M. Amigo, M.J. Popielarz, R.M. Callejón, M.L. Morales, A.M. Troncoso, M.A. Petersen, T.B. Toldam-Andersen, Comprehensive analysis of chromatographic data by using PARAFAC2 and principal components analysis, J. Chromatogr. A. 1217 (2010) 4422–4429. https://doi.org/10.1016/J.CHROMA.2010.04.042.

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

[55] G.T. Ventura, G.J. Hall, R.K. Nelson, G.S. Frysinger, B. Raghuraman, A.E. Pomerantz, O.C. Mullins, C.M. Reddy, Analysis of petroleum compositional similarity using multiway principal components analysis (MPCA) with comprehensive two-dimensional gas chromatographic data, J. Chromatogr. A. 1218 (2011) 2584–2592. https://doi.org/10.1016/J.CHROMA.2011.03.004.

[56] A.C. Beckstrom, E.M. Humston, L.R. Snyder, R.E. Synovec, S.E. Juul, Application of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry method to identify potential biomarkers of perinatal asphyxia in a non-human primate model, J. Chromatogr. A. 1218 (2011) 1899–1906. https://doi.org/10.1016/J.CHROMA.2011.01.086.

[57] 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. https://doi.org/10.1016/J.TALANTA.2006.01.038.

[58] B. Worley, S. Halouska, R. Powers, Utilities for quantifying separation in PCA/PLS-DA scores plots, Anal. Biochem. 433 (2013) 102–104. https://doi.org/10.1016/J.AB.2012.10.011.

[59] 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. https://doi.org/10.1016/J.TALANTA.2019.120239.

[60] G.L. Alexandrino, J. Malmborg, F. Augusto, J.H. Christensen, Investigating weathering in light diesel oils using comprehensive two-dimensional gas chromatography–High resolution mass spectrometry and pixel-based analysis: Possibilities and limitations, J. Chromatogr. A. 1591 (2019) 155–161. https://doi.org/10.1016/J.CHROMA.2019.01.042.

[61] 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. https://doi.org/10.1021/AC052106O/ASSET/IMAGES/LARGE/AC052106OF00007.JPEG.

[62] G. Malmquist, R. Danielsson, Alignment of chromatographic profiles for principal component analysis: a prerequisite for fingerprinting methods, J. Chromatogr. A. 687 (1994) 71–88. https://doi.org/10.1016/0021-9673(94)00726-8.

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

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

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

[66] 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. https://doi.org/10.1016/S0169-7439(01)00198-8.

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

[68] 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. https://doi.org/10.1016/j.chroma.2021.462092.

[69] B.C. Reaser, B.W. Wright, R.E. Synovec, Using Receiver Operating Characteristic Curves

to Optimize Discovery-Based Software with Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry, Anal. Chem. 89 (2017) 3606–3612. https://doi.org/10.1021/acs.analchem.6b04991.

[70] 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. https://doi.org/10.1021/acs.analchem.0c03456.

[71] G.S. Ochoa, M.C. Billingsley, R.E. Synovec, Using solid-phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: comparative analysis of aerospace fuel composition, Anal. Bioanal. Chem. (2022) 1–13. https://doi.org/10.1007/s00216-022-04348-1.

[72] P.E. Sudol, G.S. Ochoa, C.N. Cain, R.E. Synovec, Tile-based variance rank initiated-unsupervised sample indexing for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, Anal. Chim. Acta. 1209 (2022) 339847. https://doi.org/10.1016/j.aca.2022.339847.

[73] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: A basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.

[74] T. Miyazaki, K. Okada, T. Yamashita, M. Miyazaki, Two-dimensional gas chromatography time-of-flight mass spectrometry-based serum metabolic fingerprints of neonatal calves before and after first colostrum ingestion, J. Dairy Sci. 100 (2017) 4354–4364. https://doi.org/10.3168/jds.2017-12557.

[75] 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. https://doi.org/10.1021/acs.energyfuels.9b04108.

[76] V. Abrahamsson, N. Ristic, K. Franz, K. Van Geem, Comprehensive two-dimensional gas chromatography in combination with pixel-based analysis for fouling tendency prediction, J. Chromatogr. A. 1501 (2017) 89–98. https://doi.org/10.1016/j.chroma.2017.04.021.

[77] J. Fleming, Biden's 'Sustainable Aviation Fuel' Goals Mired in Myth: Report Diversified Efforts Needed to Decarbonize Aviation, (2022). https://biologicaldiversity.org/w/news/press-releases/bidens-sustainable-aviation-fuel-goals-mired-in-myth-report-2022-08-24/.

[78] J. V. Seeley, S.K. Seeley, E.K. Libby, J.D. McCurry, Analysis of Biodiesel/Petroleum Diesel Blends with Comprehensive Two-Dimensional Gas Chromatography, J. Chromatogr. Sci. 45 (2007) 650–656. https://doi.org/10.1093/CHROMSCI/45.10.650.

[79] N.V. Hung, C. Mohabeer, M. Vaccaro, S. Marcotte, V. Agasse-Peulon, L. Abdelouahed, B. Taouk, P. Cardinael, Development of two-dimensional gas chromatography (GC×GC)

coupled with Orbitrap-technology-based mass spectrometry: Interest in the identification of biofuel composition, J. Mass Spectrom. 55 (2020) e4495. https://doi.org/10.1002/JMS.4495.

[80] N.G.S. Mogollón, F.A.L. Ribeiro, R.J. Poppi, A.L. Quintana, J.A.G. Chávez, D.A.P. Agualongo, H.G. Aleme, F. Augusto, Exploratory Analysis of Biodiesel by Combining Comprehensive Two-Dimensional Gas Chromatography and Multiway Principal Component Analysis, Artic. J. Braz. Chem. Soc. 28 (2017) 740–746. https://doi.org/10.21577/0103-5053.20160222.

[81] J. Heyne, D. Bell, J. Feldhausen, Z. Yang, R. Boehm, Towards fuel composition and properties from Two-dimensional gas chromatography with flame ionization and vacuum ultraviolet spectroscopy, Fuel. 312 (2022) 122709. https://doi.org/10.1016/J.FUEL.2021.122709.

[82] R. van der Westhuizen, M. Ajam, P. De Coning, J. Beens, A. de Villiers, P. Sandra, Comprehensive two-dimensional gas chromatography for the analysis of synthetic and crude-derived jet fuels, J. Chromatogr. A. 1218 (2011) 4478–4486. https://doi.org/10.1016/j.chroma.2011.05.009.

[83] Z. Yang, S. Kosir, R. Stachler, L. Shafer, C. Anderson, J.S. Heyne, A GC × GC Tier α combustor operability prescreening method for sustainable aviation fuel candidates, Fuel. 292 (2021) 120345. https://doi.org/10.1016/J.FUEL.2021.120345.

[84] J. Oh, A. Oldani, T. Lee, L. Shafer, Deep Learning Algorithms for Assessing Sus-tainable Jet Fuels from Two-Dimensional Gas Chromatography, (2022). https://doi.org/10.2514/6.2022-0228.

[85] F.D. Kopinke, G. Zimmermann, S. Nowak, On the mechanism of coke formation in steam cracking-conclusions from results obtained by tracer experiments, Carbon N. Y. 26 (1988) 117–124. https://doi.org/10.1016/0008-6223(88)90027-9.

[86] S. Eser, R. Venkataraman, O. Altin, Deposition of carbonaceous solids on different substrates from thermal stressing of JP-8 and jet A fuels, Ind. Eng. Chem. Res. 45 (2006) 8946–8955. https://doi.org/10.1021/ie060968p.

[87] O. Altin, S. Eser, Carbon deposition from thermal stressing of petroleum fuels, in: Am. Chem. Soc. Div. Fuel Chem., 2004: pp. 764–766. https://www.researchgate.net/publication/267420915 (accessed June 28, 2021).

[88] R.E. Synovec, C.E. Freye, M.C. Billingsley, N. Keim, B. Hill-Lam, Recent Advances in Relating Chemical Compositional Variation in RP-1, RP-2, and Similar Fuels to Thermal Integrity Data, in: JANNAF 10th Liq. Propuls. Meet. Long Beach, CA, 2018: p. Abstract 2018-0001CV.

Chapter 2. Statistical inference of mass channel purity from fisher ratio analysis using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry data

This chapter was reproduced from Grant S. Ochoa, Sarah E. Prebihalo, Brooke C. Reaser, Luke C. Marney, Robert E. Synovec, "Statistical Inference of Mass Channel Purity from Fisher Ratio Analysis using Comprehensive Two-Dimensinoal Gas Chromatography with Time-of-Flight Mass Spectrometry Data" *J. Chromatogr. A* 1627 (2020) 461401

2.1. Introduction

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

The inherent complexity of GC×GC-TOFMS data, with its third order data structure [20,21] presents unique data analysis benefits and challenges, which are further impacted by the addition of a fourth dimension due to consideration of multiple sample classes as defined by the experimental design [20,21]. Fortunately, significant effort directed at developing chemometric tools to elucidate chemical information from GC×GC-TOFMS data have improved the quality of results obtained [18]. Broadly, chemometric methods aim to analyze analytical data qualitatively (signal decomposition (deconvolution), pattern recognition, one and multi class classification) and quantitatively (regression and projection-based methods) [22–25], while reducing

computation time concurrent with enhancing the ability to ascertain prominent differences between samples [18]. There are a wide variety of chemometric methods, including overlapped peak decomposition (mathematical resolution) methods to enable analyte identification and quantification such as Parallel Factor Analysis (PARAFAC) and Multivariate Classical Resolution Alternating Least Squares (MCR-ALS) [22–25], and discovery-based, pattern recognition-based, and property prediction methods such as Principal Component Analysis (PCA) [26–29], Fisher Ratio (F-ratio) Analysis [20,21,26,30–34], Partial Least Squares (PLS), and Partial Least Squares – Discriminant Analysis [35,36].

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

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

this issue was to require alignment prior to F-ratio analysis to overcome the retention time shifting [25,43–45]. To more effectively address the production of false positives due to retention time misalignment, tile-based F-ratio software was developed [30,31].

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

Specifically, we expand upon the computational capabilities of tile-based F-ratio analysis by introducing additional statistical-based algorithmic features that identify analyte-selective *m/z*. By doing so, analyte hits per *m/z* that exhibit high quality statistical metric outputs will provide accurate analyte quantification without *a priori* requiring analyte identification. Despite advances to data analysis software, current prevailing data analysis workflows frequently require significant analyst intervention post-discovery, such as chemometric decomposition, to perform analyte quantification [37,38]. Thus, a statistical metric-based validation of mass spectral selectivity provided via F-ratio(*m/z*) spectra will provide rapid quantification of analytes with confidence without requiring decomposition *a priori* for a large fraction of analyte hits. Herein, the capability of tile-based F-ratio analysis to provide accurate quantitative information is evaluated and validated by calculating the signal ratios (S-ratio) as a function of *m/z* between pairs of diesel samples spiked at four concentration levels, i.e., four sample classes (resulting in three different pairwise class comparisons), and comparing the results to the true concentration ratios. The S-ratio algorithm is a new feature incorporated into the tile-based F-ratio software. While the initial F-ratio analysis "discovers" the location of the spiked analytes, providing a hit list, the S-ratio algorithm provides relative quantification for each hit on a per *m/z* basis. Diesel fuel samples spiked with 15 non-native analytes at four concentration levels, then analyzed by GC×GC-TOFMS, were subjected to F-ratio analysis with S-ratio calculations performed on the 15 spiked analytes. This data set was ideal for this proof-of-principle study as it contains regions of peak overlap which often require decomposition prior to quantification. While the methodology described above is aimed to minimize the need for chemometric decomposition, the S-ratio(*m/z*) results were validated using both initial knowledge of the analyte spike level coupled with PARAFAC [20]. Based upon the F-ratio and S-ratio results obtained per analyte

hit on a *m/z* basis, statistical metric tools are implemented to ascertain mass spectral selectivity, and hence, quantitative accuracy of the S-ratios, since for a pure *m/z* the S-ratio will match the true concentration ratio.

2.2. Computational Principles

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

$$S(m/z) = R(m/z)C + \sum_{\mathrm{i}=1}^{\mathrm{n}} R(m/z)_i C_{\mathrm{i}} \tag{2.1}$$

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

$$\text{S-ratio}\ (m/z) = \frac{S_{\mathrm{A,total}}}{S_{\mathrm{B,total}}}(m/z) = \frac{R(m/z)C_{\mathrm{A}} + \sum_{\mathrm{i}=1}^{\mathrm{n}} R(m/z)_{\mathrm{i}} C_{\mathrm{i,A}}}{R(m/z)C_{\mathrm{B}} + \sum_{\mathrm{i}=1}^{\mathrm{n}} R(m/z)_{\mathrm{i}} C_{\mathrm{i,B}}} \tag{2.2}$$

with C_{A} the concentration of the analyte in class A, C_{B} the concentration of the analyte in class B. Additionally, in Eq. (2.2), $C_{\mathrm{i,A}}$ is the concentration of interferent i in class A, and $C_{\mathrm{i,B}}$ the concentration of an interferent i in class B, accompanied by the generic response factor $R(m/z)_{\mathrm{i}}$.

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

$$S_{\mathrm{A}}(m/z) = R(m/z)C_{\mathrm{A}} \tag{2.3}$$

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

$$S_{\mathrm{I,A}}\ (m/z) = \sum_{\mathrm{i}=1}^{\mathrm{n}} R(m/z)_{\mathrm{i}} C_{\mathrm{i,A}} \tag{2.4}$$

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

$$S_{\mathrm{B}}(m/z) = R(m/z)C_{\mathrm{B}} \tag{2.5}$$

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

$$S_{\mathrm{I,B}}(m/z) = \sum_{\mathrm{i}=1}^{\mathrm{n}} R(m/z)_{\mathrm{i}} C_{\mathrm{i,B}} \quad (2.6)$$

Using Eqs. (2.3)-(2.6), now Eq. (2.2) can be simplified as,

$$\text{S-ratio}(m/z) = \frac{S_{\mathrm{A}}(m/z) + S_{\mathrm{I,A}}(m/z)}{S_{\mathrm{B}}(m/z) + S_{\mathrm{I,B}}(m/z)} \quad (2.7)$$

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

$$\text{S-ratio}(m/z) = \frac{R(m/z)C_{\mathrm{A}}}{R(m/z)C_{\mathrm{B}}} = \frac{C_{\mathrm{A}}}{C_{\mathrm{B}}} \quad (2.8)$$

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

2.3. Experimental

Diesel fuel spiked with 15 non-native compounds (Table 2.1) at four nominal concentration levels (10 ppm, 20 ppm, 40 ppm, and 80 ppm) were analyzed by GC×GC-TOFMS (Pegasus III TOFMS with a quad-jet thermal modulator , LECO, St. Joseph, MI), that included an Agilent 6890N gas chromatograph with a 7683 autosampler (Agilent Technologies, Palo Alto, CA) [20]. A reverse column configuration was employed with a polar primary column, 1D

(Restek Rxi-17Sil MS, 29.75 m × 250 μm i.d. × 0.25 μm film thickness) and a nonpolar secondary column, 2D (Restek Rxi-1 MS, 1.5 m × 180 μm i.d. × 0.18 μm film thickness). Ultrahigh purity helium (Grade 5, 99.999%, Praxair, Seattle, WA, USA) was used as the carrier gas at a constant flow of 2.0 ml/min. The 1D column was maintained at 40 °C for 1 min and then increased to 300 °C using a temperature program rate of 5 °C/min, where it was held for 2 min. The 2D column and modulator block followed the same temperature program with a +15 °C and +60 °C offset, respectively. The modulation period was 2 s. The ion source was set to 225 °C, and mass channels, *m/z* 35-300, were collected at a 100 spectra/s after a 10 s acquisition delay. Six injection replicates of each of the four concentration spike levels were collected to create four sample classes containing 6 samples each.

Data analyses were performed in Matlab R2018a (The Mathworks Inc., Natick, MA, USA) and were imported from the LECO ChomaTOF software v 3.32 (LECO, St. Joseph, MI) via an in-house developed data converter [20]. The data were analyzed with in-house developed tile-based F-ratio software [8,20,21,30,31]. There were three class comparisons, the 20 ppm versus 10 ppm spike level, the 40 ppm versus the 20 ppm spike level, and the 80 ppm versus 40 ppm spike level, so for each comparison a total of 12 samples were analyzed with 6 in each class. Most of the results presented are for the 80 ppm versus 40 ppm spike level comparison, while in some instances all three comparisons are related to each other. The F-ratio tile dimensions were 12 s on 1D by 100 ms on 2D reducing the size of the data tensor from 200×1644×267 to 20×274×267 data points representing the 2D, 1D, and *m/z* respectively, and the cluster tile dimensions were 8 s on 1D by 60 ms on 2D; a *S/N* threshold of 10 was applied and at least 3 *m/z* were required to define a reported F-ratio [20].

Table 2.1. S-ratio results for each of the 15 spiked analytes in the 80/40 ppm comparison, arranged by decreasing maximum F-ratio. The S-ratio for the top m/z (which also passed the p-value and LOF thresholds) is provided and compared against the actual concentration ratio. Average concentration ratio deviation was 1.0% with a standard deviation of 1.4%.

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

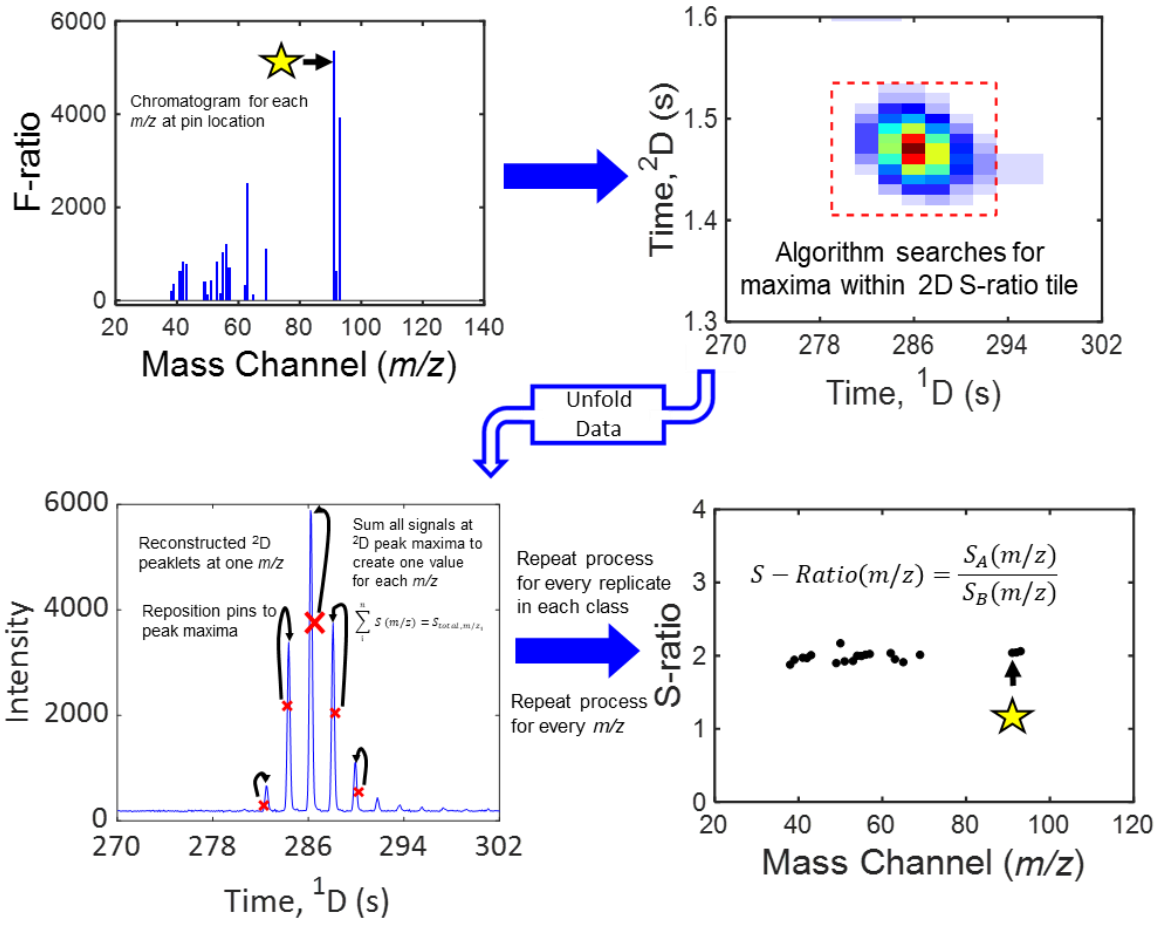


Figure 2.1. Flow chart depicting the data processing steps of the S-ratio(m/z) algorithm for 1-chlorohexane. For each F-ratio(m/z), illustrated for m/z 91 with a star (upper left), a 2D S-ratio tile is cutout around each hit (upper right) from each chromatogram centered at the average 1D × 2D pin location. The data within the 2D S-ratio tile is unfolded and concatenated into a vector laying each modulation section end to end, with the average 1D × 2D pin location represented by the large red X in the bottom left plot. Initial 2D pins are then assigned to each modulation based off the average 1D × 2D pin, separated by the modulation period, which serve as initial estimates to the peak maxima for each 2D peak, represented by the small red x's. These initial pins are then repositioned to the 2D peak maxima, indicated by the black arrows. Finally, the 2D peak heights are summed, yielding a single signal value for each selected m/z (lower left). The summed signals are then averaged for both sample classes and divided to yield the S-ratio(m/z) spectrum, with the S-ratio for m/z 91 indicated with a star (lower right).

Directly following removal of redundant hits by the pinning and clustering algorithm step [31], S-ratio(*m/z*) were obtained, for each F-ratio(*m/z*) as illustrated in Fig. 2.1, with the S-ratio algorithm flow chart provided in Fig. A1 (Appendix A). The S-ratio(*m/z*) algorithm is based in part upon a previously developed algorithm for aligned chromatograms [25,46], though the new

version does not require alignment. For each hit identified by F-ratio analysis, data for a given analyte was extracted from each chromatogram, one chromatogram at a time. First, the ^{1}D and ^{2}D pin location from the F-ratio hit list was applied to an individual chromatogram. Then, a 2D S-ratio tile with dimensions equal to the ^{1}D dimension of the F-ratio tile (6 s) and double the dimension of the ^{2}D cluster tile dimension (120 ms) was centered on the pin. Next, the algorithm determines where the average (original) F-ratio pin location is on the peak for the particular chromatogram at hand, i.e. whether it is on the right, left or top of the analyte peak maxima. Using an iterative process, the ^{2}D peak maxima for all modulations located within the 2D S-ratio tile (as illustrated in Fig. 2.1) are recorded. Then, beginning from the modulation which corresponds to the largest ^{2}D peak maximum, the ^{1}D boundaries are determined by checking subsequent modulations to the right and left for an increase in signal, which indicates the modulation belongs to another analyte and/or interferent. Finally, the signals for the ^{2}D peak maxima from all modulations identified as belonging to the analyte are summed, as to give one signal per *m/z* from the F-ratio spectrum of the hit. This algorithm is repeated across all chromatograms in the F-ratio class comparison. The process described above was repeated for all *m/z* identified as class distinguishing at each pin location. Following these signal intensity calculations, the algorithm calculated S-ratio(*m/z*) at every identified F-ratio(*m/z*) for each hit between class A (the class containing the higher spike level) and class B (lower spike level). Quantitative accuracy was evaluated through comparison of the true concentration ratio and the experimental S-ratio(*m/z*). Based upon the experimental design, all S-ratios, and hence experimentally determined concentration ratios, for the spiked analytes should nominally be 2.

Combinatorial null distribution analysis (CNDA) has been previously demonstrated to provide a statistically robust F-ratio threshold in which false positives are minimized, while

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

Independent from the inference of mass spectral purity provided by the F-ratio threshold via CNDA, the consistency of analyte peak shapes between classes was analyzed and quantified using a lack-of-fit (*LOF*) statistic using an in-house developed algorithm, with the *LOF* algorithm flow chart provided in Fig. A2 (Appendix A). To begin, as with the determinations of the S-ratio(*m/z*) signal (per Fig. 2.1), a *LOF* tile whose ^{1}D dimension matches the F-ratio tile ^{1}D dimension (6 s) and ^{2}D dimension is double the F-ratio tile ^{2}D dimension (240 ms) was pulled from a chromatogram from class A (80 ppm) and class B (40 ppm) for each *m/z* for a given hit. For a sufficiently pure *m/z*, an increase in concentration will increase the area under the ^{2}D peaks

while the peak shapes will remain constant. To test this concept, the pattern of ^{2}D peaks within a *LOF* tile for class B were normalized to the area of the corresponding modulations in class A. Next, the pattern of ^{2}D peaks in class B was aligned to class A by interpolation such that the original ^{2}D peak shapes remain intact, minimizing the squared residuals between classes. The data were then unfolded into a concatenated vector and the *LOF* was calculated according to Eq. (2.9),

$$LOF\ (m/z)\ = 100\sqrt{\frac{\sum_i (y-x)_i^2}{\sum_i x_i^2}} \tag{2.9}$$

where *x* is the concatenated pattern of ^{2}D peaks for class B, and *y* is the concatenated pattern of ^{2}D peaks for class A [47]. Purer *m/z* result in smaller residuals indicated by a lower *LOF*, while a larger *LOF* is indicative of a change in the pattern of ^{2}D peaks between classes, indicative of interfering compounds. However, a low *LOF* can also be obtained when there is only an interferent present in the *LOF* bin at a particular *m/z*. This results from a fully non-selective *m/z* that produces an extremely low F-ratio. So, in combination with the *LOF* metric, the S-ratio algorithm also calculates the p-value for each hit per *m/z*, and a combination of a sufficiently low *LOF* and low p-value are required to infer identification of pure *m/z*.

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

2.4. Results and Discussion

The GC×GC separation of the diesel fuel spiked with 15 analytes at the 80 ppm level is presented in Fig. 2.2A. The locations of the spiked analytes span the 2D separation space and are indicated by circles. Tile-based F-ratio analysis was employed for the class comparison using six injection replicates of 80 ppm spiked diesel (class A) and six injection replicates of 40 ppm spiked diesel (class B) to generate a hit list on a per *m/z* basis for each hit. Comparisons were also made between the 40 and 20 ppm spiked diesel samples and between the 20 and 10 ppm spiked diesel samples, yielding an F-ratio(*m/z*) distribution for all three two-class comparisons, provided in Fig. 2.2B. Note that most of the F-ratio(*m/z*) are relatively low, and the concentration dependence, i.e., *S/N* dependence, is evident with the 80/40 ppm F-ratio(*m/z*) distribution shifted preferentially to higher F-ratios, followed by the 40/20 ppm comparison, and finally the 20/10 ppm produced the lowest F-ratios. Combinatorial null distribution analysis (CNDA) was then applied to the three null class comparisons, to provide insight into determining a statistically robust F-ratio threshold in which one would anticipate that false positives are minimized for the true class comparisons in Fig. 2.2B. The CNDA workflow and results are presented in Appendix A (Fig. A3) for all three comparisons, and an average F-ratio threshold of 97 was obtained. Based upon the probability levels applied, an F-ratio(*m/z*) exceeding this threshold has a 1 in 100,000 chance of being a false positive at a 95% confidence level, to purposely provide a very restrictive “statistically implied quantitative accuracy” filter via CNDA. Note that this implied quantitative accuracy filter is further investigated by more direct *m/z* purity metrics, specifically the *LOF* and p-values for the S-ratio(*m/z*) outputs.

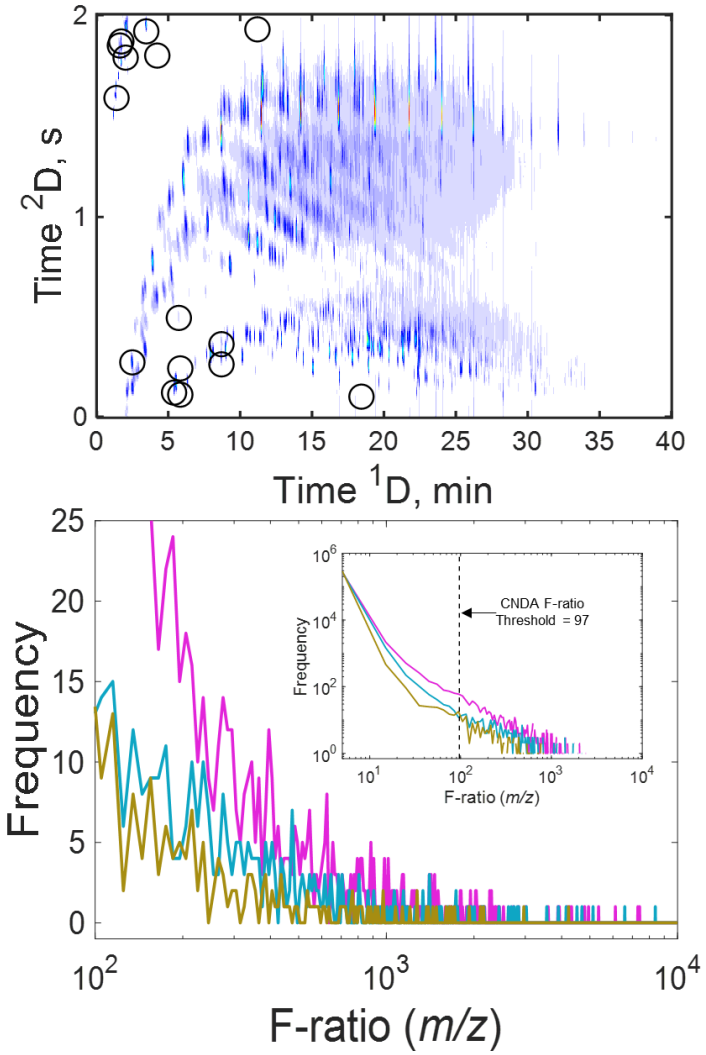


Figure 2.2. (A) GC×GC chromatogram (TIC) of the 80 ppm spiked diesel. Alkanes (top), cycloalkanes and branched alkanes (middle), and aromatic (bottom) compound classes are distinctively separated. The 15 spiked analyte locations are denoted with circles. (B) F-ratio(m/z) distribution of spiked diesel at three different concentration level comparisons: 20/10 ppm in gold, 40/20 ppm in cyan, and 80/40 ppm in magenta. As the concentration levels of each comparison increase, the overall frequency of F-ratio(m/z) > 100 increases, due to the increased signal per m/z overcoming the user-selected S/N threshold. The majority of the F-ratio(m/z) > 100 will be shown to be attributed to the spiked analytes.

F-ratio(*m/z*) and S-ratio(*m/z*) spectra were calculated for each spiked analyte in the 80/40 ppm comparison to begin to visualize how these two outputs are related in the context of *m/z* purity, beginning with a representative pure analyte (1-chlorohexane) in Fig. 2.3A, and a representative overlapped analyte (methyl caproate) in Fig. 2.3B. The signal for 1-chlorohexane in the 80 ppm sample is presented for both the top F-ratio at *m/z* = 91 in Fig. 2.3A(i), and the total ion current (TIC) in Fig. 2.3A(ii) with a red box outlining the 2D bin used by the S-ratio algorithm. The 1-chlorohexane peak is not overlapped with any interferences. Likewise, the top F-ratio at *m/z* = 87 and TIC are presented for methyl caproate in Fig. 2.3B(i-ii). While the signal for the top F-ratio(*m/z*) for methyl caproate is framed by the red box, suggesting *m/z* = 91 might be pure, most of the TIC signal falls outside of the red box, suggesting interferences are overlapping with methyl caproate. Indication that 1-chlorohexane is a pure peak, while methyl caproate is not a pure peak is provided by the visual comparison of the mass spectrum of the signal in the 2D bin versus the library spectrum for each, in Figs 2.3A(v) and 2.3B(v), respectively. Most, if not all, of the *m/z* match for 1-chlorohexane between the 2D bin spectrum and the library spectrum, while for methyl caproate there are several *m/z* in the 2D spectrum that are not in the library spectrum. Note that providing the library spectrum is not needed for application of the method presented herein, rather, the library spectra are merely provided in Fig. 2.3 to aid in understanding the interrelationships of the F-ratio(*m/z*), S-ratio(*m/z*), and other quantitative metrics.

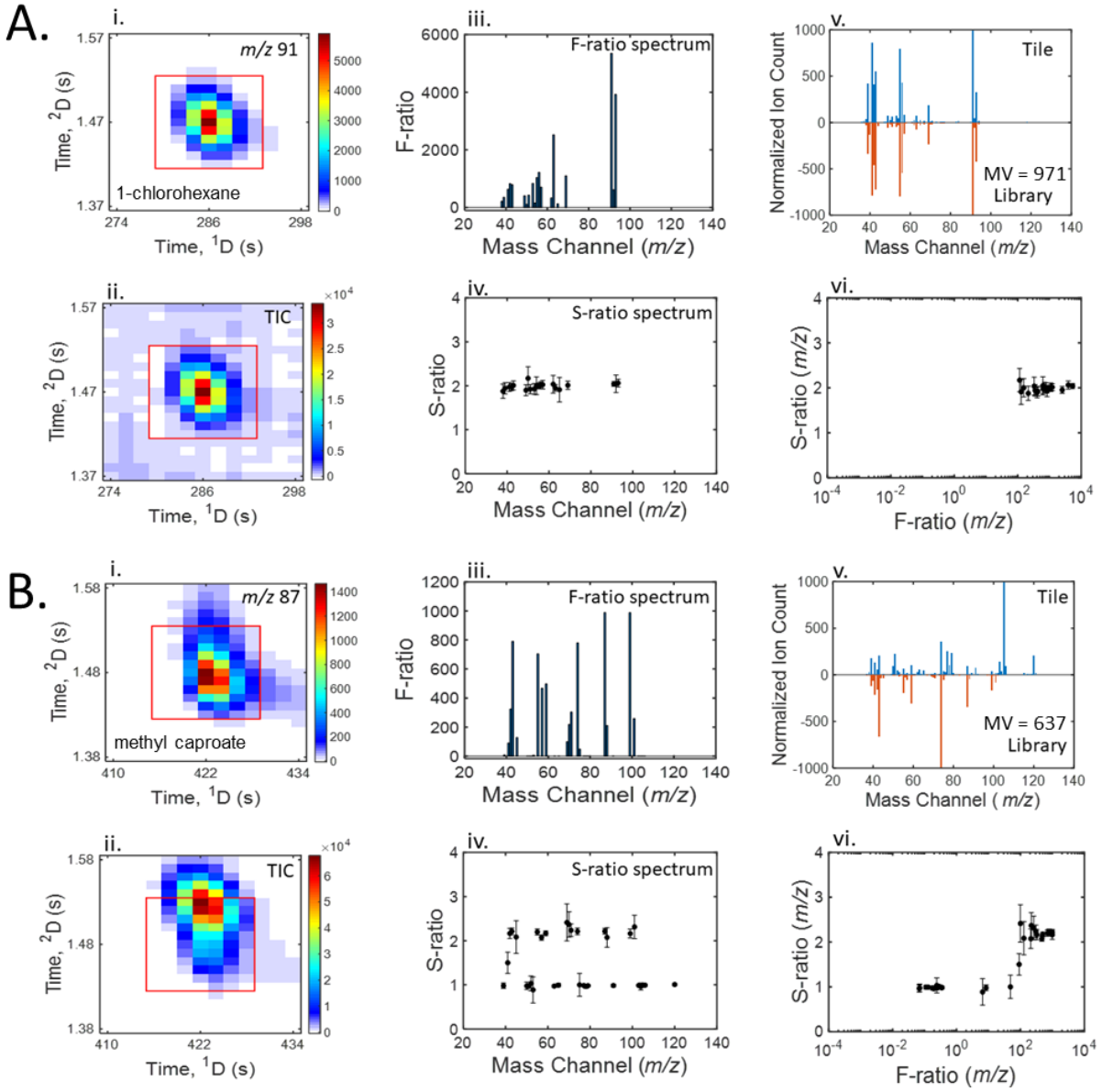


Figure 2.3. Illustration of the F-ratio(m/z) and S-ratio(m/z) spectra from the 80/40 ppm comparison for two representative analytes: (A) one pure, 1-chlorohexane, and (B) one overlapped, methyl caproate. Two-dimensional data for the region around each analyte for the top F-ratio m/z (i), inferred to be pure, and the total ion current (TIC) (ii). The red box outlines the tile space used by the S-ratio algorithm. Each compound F-ratio(m/z) and S-ratio(m/z) spectrum is provided in (iii) and (iv), respectively. The mass spectrum from the red box region is also provided (v) for both analytes and compared to the library spectrum. Finally, a plot of S-ratio(m/z) as a function of F-ratio(m/z) in (vi) provides insight into the correlation between high F-ratio(m/z) and having the S-ratio(m/z) correspond to the true concentration ratio of 2.

The F-ratio(*m/z*) spectrum for each hit provides a "chemical fingerprint" to determine which *m/z* are most likely to distinguish the two sample classes. For example, Figs. 2.3A(iii) and 3B(iii) provide the F-ratio(*m/z*) spectra for 1-chlorohexane and methyl caproate, respectively. A lower F-ratio(*m/z*) implies lower spectral selectivity (and/or lower *S/N*), with an adverse impact on quantitative accuracy anticipated, while a higher F-ratio(*m/z*) implies higher spectral selectivity. However, since the F-ratio(*m/z*) is principally a tool to prioritize which *m/z* per hit may likely be statistically different in concentration between the two classes, we turn to the S-ratio(*m/z*) to bridge this gap between analyte discovery and quantification. For 1-chlorohexane and methyl caproate, the S-ratio(*m/z*) spectra are provided in Figs. 2.3A(iv) and 2.3B(iv), respectively. For a pure analyte (1-chlorohexane), the S-ratio(*m/z*) spectrum exhibits a consistent S-ratio(*m/z*) around 2 in Fig. 2.3A(iv), while the presence of interferences (with methyl caproate) results in two distinct S-ratio(*m/z*) levels in Fig. 2.3B(iv), belonging to both the analyte and an interferent. An S-ratio(*m/z*) of 2 is obtained for nominally pure *m/z* for methyl caproate, while an S-ratio(*m/z*) of about 1 suggests the signal obtained is dominated by an interferent that is invariant between the two sample classes. Finally, in Figs. 2.3A(vi) and 2.3B(vi) the S-ratio(*m/z*) is plotted as a function of F-ratio(*m/z*), for 1-chlorohexane and methyl caproate, respectively. Here, the relationship between obtaining an accurate S-ratio(*m/z*) and its connection to requiring a sufficiently high F-ratio(*m/z*) provides two key observations. First, the S-ratio(*m/z*) trends toward the true concentration ratio for the spiked analyte for a high F-ratio(*m/z*), indicating quantitative accuracy. Second, as the F-ratio(*m/z*) magnitude decreases the S-ratio(*m/z*) deviate from the true concentration ratio, leveling off at 1. Indeed, the transition from obtaining the true concentration to a biased concentration ratio appears to occur at an F-ratio $\approx$ 100, which is fully supported by the F-ratio threshold of 97 provided by the CNDA (Fig. A3). Similar trends were

also observed for the remaining the 13 spiked analytes in all three F-ratio comparisons, with results such as in Fig. 2.3 provided for 1-nonanol in Fig. A4 in Appendix A.

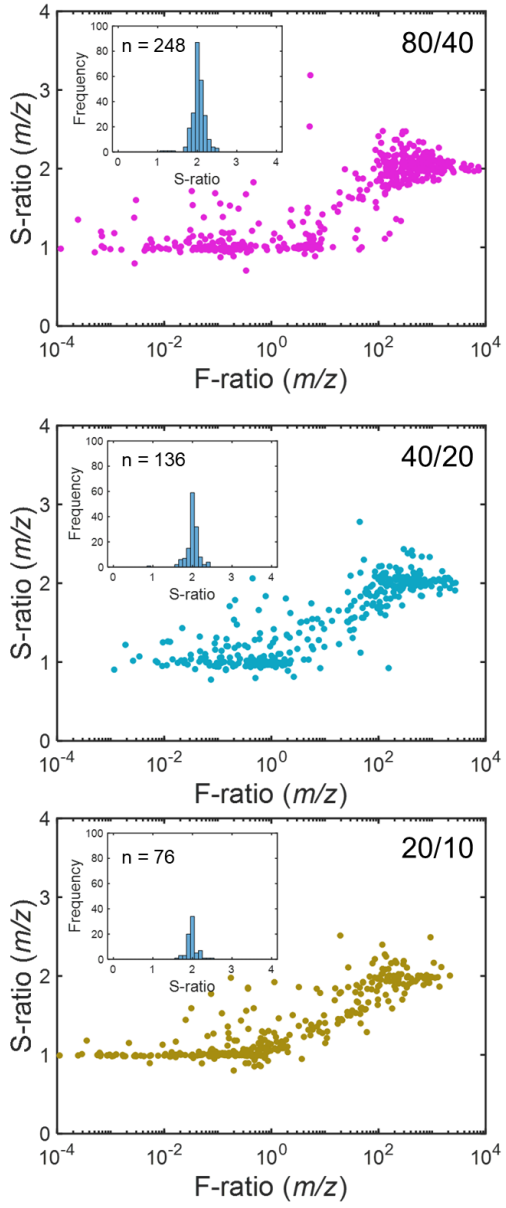


Figure 2.4. The S-ratio(m/z) for all m/z selected by F-ratio analysis for the 15 spiked analytes plotted versus their respective F-ratio(m/z) for each comparison: (A) 80/40 ppm, (B) 40/20 ppm, and (C) 20/10 ppm. Each plot shows a similar pattern of the S-ratio(m/z) tending toward an accurate concentration ratio of 2 at high F-ratio(m/z). The inset in each plot shows the distribution of S-ratio(m/z) above the F-ratio threshold of 97 determined by combinatorial null distribution analysis (Supplementary Material). The occurrence of accurate S-ratio(m/z) ≈ 2) increases with higher concentration comparisons.

A summary of the S-ratio(*m/z*) as a function of F-ratio(*m/z*) for all three comparisons is provided in Fig. 2.4, where an S-ratio(*m/z*) ≈ 2 is obtained for an F-ratio(*m/z)* ≥ 97. As the *S/N* of the analyte signal within each comparison increases, the number of *m/z* that exceed the F-ratio threshold of 97 also increases, since more *m/z* have passed the *S/N* threshold. Specifically, for the 15 spiked analytes, 248 of the F-ratio(*m/z*) are ≥ 97 for the 80/40 ppm comparison, while there were 136 for the 40/20 ppm comparison, and then dropping down to only 76 for the 20/10 ppm comparison. While the results in Fig. 2.4 suggest that an F-ratio(*m/z)* ≥ 97 may very well serve as a suitable metric to infer quantitative accuracy for determining the analyte concentration ratio via the S-ratio(*m/z*), a firmer connection to the *LOF* and p-value statistical metrics was warranted.

For each hit, the *LOF* statistic was calculated to determine the extent of *m/z* purity. Figure 2.5A-D provides an example of applying the *LOF* algorithm, via Eq. (2.9), for four *m/z* "discovered" by F-ratio analysis for methyl caproate. Here the chromatographic data provided by the S-ratio(*m/z*) algorithm are utilized before and after normalization. In Fig. 2.5A for *m/z* = 43, doubling the analyte concentration between sample classes results in a consistent peak shape. Following the normalization step for the *LOF* algorithm, low residuals are obtained, which experience some variance due to sampling frequency, ultimately resulting in a low *LOF* of 5.5% indicative of a sufficiently pure *m/z*. In contrast, as the interferences become a more dominant contributor to the signal, the consistency in peak shape pattern suffers when the analyte concentration is doubled between sample classes. This is recognized with both *m/z* = 101 (Fig. 2.5B) and *m/z* = 41 (Fig. 2.5C), where after area normalization the presence of interferences causes the peak pattern to misalign and subsequently increases the size of the residuals, resulting in large *LOF* values of 22.4% and 29.1%, respectively. If the interference signal dominates to the extent that the analyte signal is negligible, the *LOF* returns to a smaller value, indicating that the

m/z is essentially pure, or at least a constant contribution from class to class, for the interference. This situation is illustrated in Fig. 2.5D for *m/z* = 105, where a *LOF* of 1.6% is obtained. However, there is a key distinction between the two low *LOF* conditions, Fig. 2.5A versus Fig 2.5D. For *m/z* = 43, the p-value = 1.6×10^{-12}, indicating the concentration ratio is significantly different than 1, while for *m/z* = 105, the p-value = 0.47, indicating the concentration ratio is not significantly different than 1. Thus, it is the combination of using the *LOF* and p-value metrics, both provided directly from the S-ratio(*m/z*) output, that provide an inferred, but statistically confident, assessment of pure *m/z*. For the purpose of this study, we elected to require the *LOF* ≤ 20% concurrent with a p-value ≤ 0.001 in order to deem a *m/z* to be sufficiently pure to provide accurate quantitative analysis, i.e., an accurate S-ratio(*m/z*) corresponding with the true concentration ratio.

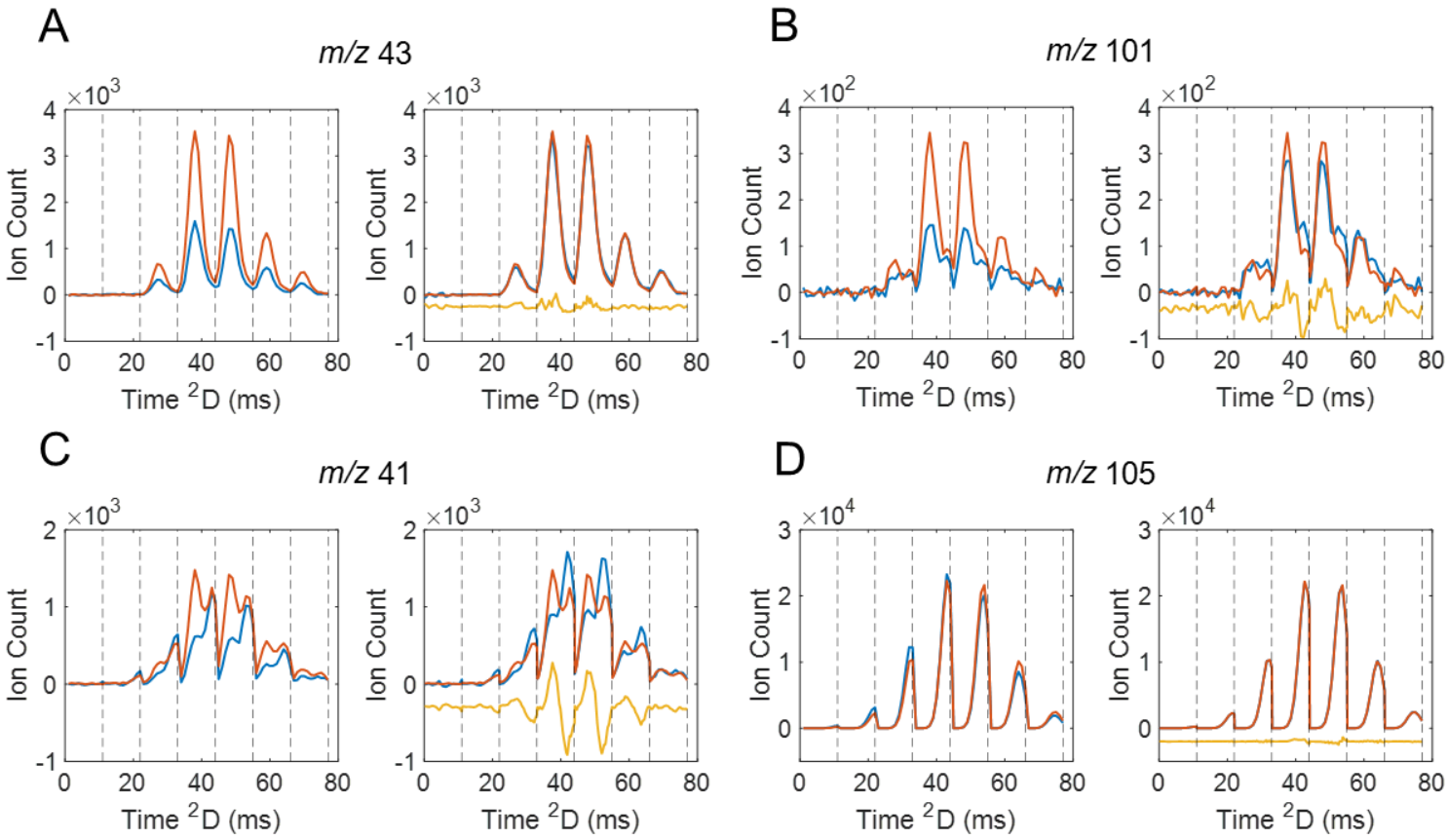


Figure 2.5. Demonstration of the LOF metric determination for four selected m/z from methyl caproate. An overlay of the original concatenated ^{2}D data for a chromatogram from class A (80 ppm) and one from class B (40 ppm) is provided for each m/z in the left panel. The right panel for each m/z shows the same concatenated ^{2}D data after normalization and alignment, with the residuals plotted underneath, offset for clarity. (A) A pure m/z = 43, yields excellent consistency between classes with low residuals, LOF = 5.5%. (B) For m/z = 101, a small relative contribution from an overlapped compound, increases the residuals, LOF = 22.4%. (C) For m/z = 41, more significant contribution from the overlapped compound is indicated, and the residuals become very large relative to the analyte signal, LOF = 29.1%. (D) At m/z = 105, essentially a pure m/z for the overlapped compound, with negligible analyte contribution, with no change between classes the residuals are very low, yielding a low LOF of 1.6% (but the p-value would be high).

These *LOF* and p-value metrics are applied to all F-ratio(*m/z*) in Fig. 2.6 for the 80/40 ppm comparison. The first statistical filter applied is the p-value. All F-ratio(*m/z*) with a p-value > 0.001 are indicated as blue dots. Next the *LOF* filter was applied, and all F-ratio(*m/z*) with a *LOF* > 20% are indicated as red dots. This leaves a region filled with green dots indicating pure *m/z*, with representations of these results put into the context of the S-ratio(*m/z*) in Fig. 2.7 for all three comparisons. It is interesting to relate the S-ratio(*m/z*) results in Fig. 2.4 prior to applying these metrics, in contrast to in Fig. 2.7 after their application. There are 170 pure *m/z* for the 80/40 comparison as indicated in Fig. 2.7A. Note that use of the *LOF* and p-value metrics to determine pure *m/z* does not rely upon the F-ratio ≥ 97 threshold applied in Fig. 2.4, where there were 248 *m/z* for the 80/40 comparison that were deemed good *candidates* to be determined to be pure *m/z*. Hence, the combination of the *LOF* and p-value metrics filter the candidate *m/z*, forming a more robust cohort of pure *m/z*. The *LOF* and p-value metrics applied to all F-ratio(*m/z*) for the other two comparisons are provided in Appendix A (Fig. A5). Likewise, for the other two comparisons, substantially fewer *m/z* met the *LOF* and p-value metrics relative to the F-ratio ≥ 97 threshold. For the 40/20 ppm comparison there were 91 *m/z* (Fig. 2.7B) versus 136 *m/z* (Fig. 2.4), and for the 20/10 comparison even fewer with 70 *m/z* (Fig. 2.7C) versus 76 *m/z* (Fig. 2.4).

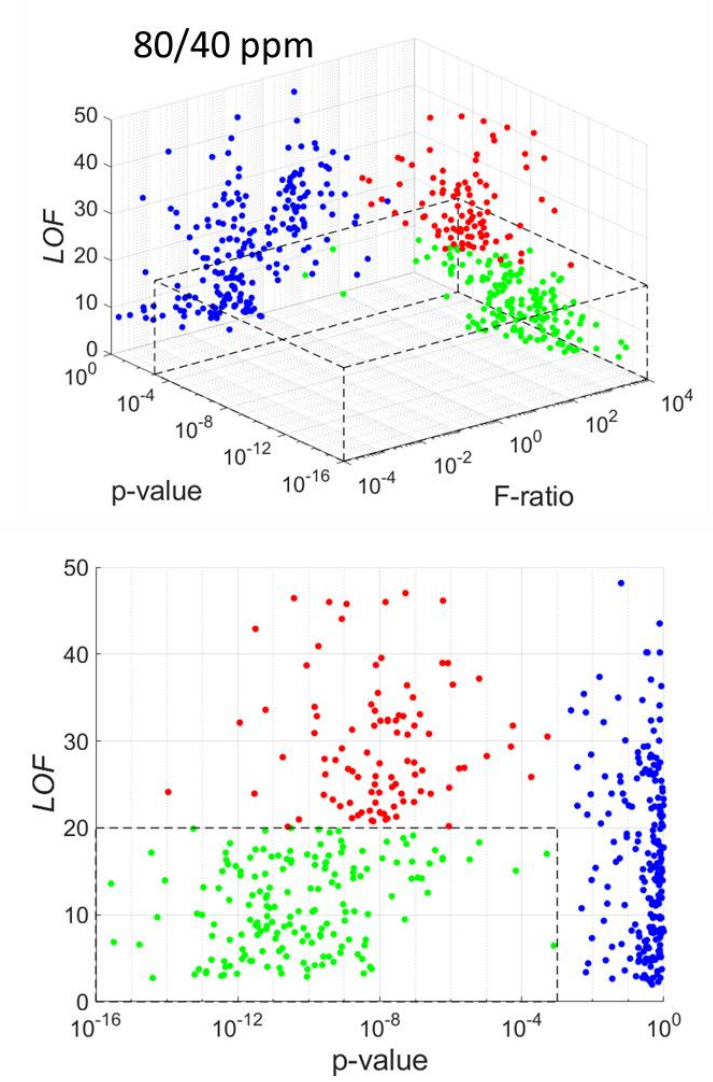


Figure 2.6. F-ratio(m/z), p-value(m/z), and LOF(m/z) for all m/z of all 15 spiked analytes. The dashed lines represent the regions defined by the p-value ≤ 0.001 and LOF ≤ 20 % thresholds (A) Blue m/z have not passed either threshold, red have only passed the p-value threshold, and green m/z have passed both thresholds. (B) A 2D projection of the LOF and p-value dimensions illustrates these two metrics independent of the F-ratio dimension. The blue m/z carry insignificant m/z belonging to interfering compounds. Red m/z have statistically significant signal differences between classes however they are interfered to a large extent with overlapping compounds. Lastly the green m/z are considered both statistically significant and sufficiently pure to be used for signal quantification, while not relying upon the F-ratio(m/z).

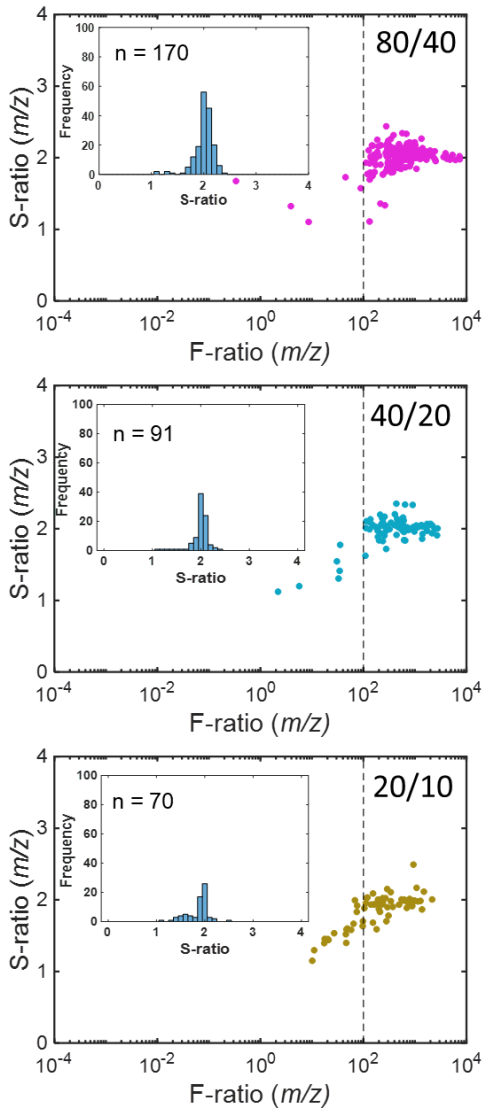


Figure 2.7. S-ratio versus F-ratio plots for all three concentration comparisons after applying p-value and LOF metric thresholds in Fig. 2.6 with the top F-ratio(m/z). The combinatorial null distribution analysis F-ratio threshold = 97 is marked for context. Inset in each plot are histograms of the m/z that pass these two metrics along with a count.

The S-ratio(*m/z*) as a function of F-ratio(*m/z*) in Fig. 2.7 also reveals that the application of the *LOF* and p-value metrics did not rid the *m/z* selected of all apparent outliers. There were a small number of *m/z* that while meeting the *LOF* and p-value metrics, they exhibited an F-ratio < 97. For example, with the 80/40 ppm comparison, there were 6 *m/z* with an F-ratio < 97 that met the *LOF* and p-value metrics. One approach to improve this situation is to lower the *LOF* threshold, but this is at the expense of discarding an *m/z* that may be sufficiently pure for the purpose of a given quantitative application (see Appendix A, Fig. A6 for application of a *LOF* thresholds of 10% and 15%). None the less, it is tempting to consider the notion of inferring *m/z* purity strictly by applying the F-ratio(*m/z*) ≥ 97 threshold provided by the CNDA. By doing so, the analyst would be relying upon the correlation between how well the LOF and p-value metrics performed their task as manifested in the S-ratio(*m/z*) dependency on F-ratio(*m/z*) in Fig. 2.7. This approach involves taking the S-ratio(*m/z*) at the maximum F-ratio(*m/z*) as the best estimate of the analyte concentration ratio. For the 80/40 ppm comparison in Table 2.1, all of the concentration ratio estimates are reasonably accurate with an average deviation of 1.0%, albeit benefited by the maximum F-ratio(*m/z*) being much greater than the 97 threshold for all 15 analytes. Indeed, all 15 of the maximum F-ratio(*m/z*) also passed the *LOF* and p-value threshold metrics. Tables for the other two comparisons are provided in the Supplementary Material, Tables A1 and A2, with diminished accuracy due to the maximum F-ratio(*m/z*) decreasing as the analyte concentration levels decrease. While this approach does not require the analyte to be identified *a priori* in the process, the analyst must realize that other *m/z* purity metrics are needed to ensure an accurate quantification, as demonstrated in this study.

2.5. Conclusion

In this study, we evaluated the level of mass channel selectivity afforded by tile-based F-ratio analysis through statistical inference. Following F-ratio analysis of spiked diesel at two concentration levels (80/40 ppm), a novel S-ratio algorithm and novel *LOF* algorithm were implemented to study the statistical relationship between F-ratio, p-value, *LOF*, and *m/z* purity, as they relate to quantitative accuracy. Application of a suitable *LOF* threshold coupled with a p-value threshold yielded a subset of the most pure *m/z* for each of the 15 spiked analytes, evident by the low deviations (< 5%) in S-ratio relative to the true concentration ratio**,** for the top *m/z* for each analyte. Thus, the ability to isolate pure *m/z* without the need for higher level signal decomposition algorithms was demonstrated. The analyst may also consider inferring *m/z* purity based solely on the F-ratio (*m/z*) threshold provided by CNDA. The *LOF* metric was applied for the case where the analyte was present in both classes at sufficient concentrations. In the case that an analyte is not present in one class, a separate algorithm step would need to be included to identify this situation, e.g., any *m/z* present in one class but not the other (below the *S/N* threshold) would identify that *m/z* as belonging to an analyte of interest. Additionally, samples with analytes of interest that are native to the unspiked samples are also amenable to this method. However, spiking analytes with appreciable pre-existing concentration will result in smaller S-ratios. A future study is warranted to discover the limits of S-ratio quantification using this method, as the current study focused on non-native analytes at a factor of 2 concentration ratio. Future application to discover spectrally pure *m/z* will be utilized not only for analyte identification but as an advanced feature selection for inputs to methods like PCA and PLS.

2.6. References

[1] Z. Liu, J.B. Phillips, Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface, J. Chromatogr. Sci. 29 (1991) 227–231. https://doi.org/10.1093/chromsci/29.6.227.
[2] R.B. Gaines, G.S. Frysinger, M.S. Hendrick-Smith, J.D. Stuart, Oil Spill Source Identification by Comprehensive Two-Dimensional Gas Chromatography, Environ. Sci. Technol. 33 (1999) 2106–2112. https://doi.org/10.1021/es9810484.
[3] J.F. Focant, A. Sjödin, D.G. Patterson, Improved separation of the 209 polychlorinated biphenyl congeners using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, J. Chromatogr. A 1040 (2004) 227–238. https://doi.org/10.1016/j.chroma.2004.04.003.
[4] P.J. Marriott, S.T. Chin, B. Maikhunthod, H.G. Schmarr, S. Bieri, Multidimensional gas chromatography, TrAC - Trends Anal. Chem. 34 (2012) 1–21. https://doi.org/10.1016/j.trac.2011.10.013.
[5] D. Megson, R. Kalin, P.J. Worsfold, C. Gauchotte-Lindsay, D.G. Patterson, M.C. Lohan, S. Comber, T.A. Brown, G. O'Sullivan, Fingerprinting polychlorinated biphenyls in environmental samples using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, J. Chromatogr. A 1318 (2013) 276–283. https://doi.org/10.1016/j.chroma.2013.10.016.
[6] K.L. Organtini, A.L. Myers, K.J. Jobst, J. Cochran, B. Ross, B. McCarry, E.J. Reiner, F.L. Dorman, Comprehensive characterization of the halogenated dibenzo-p-dioxin and dibenzofuran contents of residential fire debris using comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry, J. Chromatogr. A 1369 (2014) 138–146. https://doi.org/10.1016/j.chroma.2014.09.088.
[7] M.R. Jacobs, M. Edwards, T. Górecki, P.N. Nesterenko, R.A. Shellie, Evaluation of a miniaturised single-stage thermal modulator for comprehensive two-dimensional gas chromatography of petroleum contaminated soils, J. Chromatogr. A 1463 (2016) 162–168. https://doi.org/10.1016/j.chroma.2016.08.009.
[8] N.E. Watson, B.A. Parsons, R.E. Synovec, Performance evaluation of tile-based Fisher Ratio analysis using a benchmark yeast metabolome dataset, J. Chromatogr. A 1459 (2016) 101–111. https://doi.org/10.1016/j.chroma.2016.06.067.
[9] F. Magagna, A. Guglielmetti, E. Liberto, S.E. Reichenbach, E. Allegrucci, G. Gobino, C. Bicchi, C. Cordero, Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination, J. Agric. Food Chem. 65 (2017) 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167.
[10] E.A. Higgins Keppler, C.L. Jenkins, T.J. Davis, H.D. Bean, Advances in the application of comprehensive two-dimensional gas chromatography in metabolomics, TrAC - Trends Anal. Chem. 109 (2018) 275–286. https://doi.org/10.1016/j.trac.2018.10.015.
[11] B. Gruber, B.A. Weggler, R. Jaramillo, K.A. Murrell, P.K. Piotrowski, F.L. Dorman, Comprehensive two-dimensional gas chromatography in forensic science: A critical review of recent trends, TrAC - Trends Anal. Chem. 105 (2018) 292–301. https://doi.org/10.1016/j.trac.2018.05.017.
[12] A.M. Muscalu, T. Górecki, Comprehensive two-dimensional gas chromatography in environmental analysis, TrAC - Trends Anal. Chem. 106 (2018) 225–245. https://doi.org/10.1016/j.trac.2018.07.001.
[13] M. Zoccali, S. Cappello, L. Mondello, Multilevel characterization of marine microbial biodegradation potentiality by means of flow-modulated comprehensive two-dimensional gas chromatography combined with a triple quadrupole mass spectrometer, J. Chromatogr. A 1547 (2018) 99–106. https://doi.org/10.1016/j.chroma.2018.03.013.
[14] R. Pesesse, P.H. Stefanuto, F. Schleich, R. Louis, J.F. Focant, Multimodal chemometric approach for the analysis of human exhaled breath in lung cancer patients by TD-GC × GC-TOFMS, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 1114–1115 (2019) 146–153. https://doi.org/10.1016/j.jchromb.2019.01.029.

[15] J.B. Phillips, J. Beens, Comprehensive two-dimensional gas chromatography: A hyphenated method with strong coupling between the two dimensions, J. Chromatogr. A 856 (1999) 331–347. https://doi.org/10.1016/S0021-9673(99)00815-8.
[16] M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, J. Chromatogr. A 1383 (2015) 151–159. https://doi.org/10.1016/j.chroma.2015.01.031.
[17] L.M. Blumberg, Multidimensional Gas Chromatography: Theoretical Considerations, in: Compr. Chromatogr. Comb. with Mass Spectrom., John Wiley and Sons, 2011: pp. 13–63. https://doi.org/10.1002/9781118003466.ch2.
[18] S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, R.E. Synovec, Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications, Anal. Chem. 90 (2018) 505–532. https://doi.org/10.1021/acs.analchem.7b04226.
[19] J. V. Seeley, S.K. Seeley, Multidimensional gas chromatography: Fundamental advances and new applications, Anal. Chem. 85 (2013) 557–578. https://doi.org/10.1021/ac303195u.
[20] B.C. Reaser, B.W. Wright, R.E. Synovec, Using Receiver Operating Characteristic Curves to Optimize Discovery-Based Software with Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry, Anal. Chem. 89 (2017) 3606–3612. https://doi.org/10.1021/acs.analchem.6b04991.
[21] B.A. Parsons, D.K. Pinkerton, B.W. Wright, R.E. Synovec, Chemical characterization of the acid alteration of diesel fuel: Non-targeted analysis by two-dimensional gas chromatography coupled with time-of-flight mass spectrometry with tile-based Fisher ratio and combinatorial threshold determination, J. Chromatogr. A 1440 (2016) 179–190. https://doi.org/10.1016/j.chroma.2016.02.067.
[22] R. Bro, PARAFAC. Tutorial and applications, in: Chemom. Intell. Lab. Syst., 1997: pp. 149–171. https://doi.org/10.1016/S0169-7439(97)00032-4.
[23] A. De Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, J. Chemom. 15 (2001) 749–772. https://doi.org/10.1002/cem.662.
[24] R. Gargallo, R. Tauler, F. Cuesta-Sánchez, D.L. Massart, Validation of alternating least-squares multivariate curve resolution for chromatographic resolution and quantitation, TrAC - Trends Anal. Chem. 15 (1996) 279–286. https://doi.org/10.1016/0165-9936(96)00048-9.
[25] R.E. Mohler, B.P. Tu, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography-time-of-flight-mass spectrometry data, J. Chromatogr. A 1186 (2008) 401–411. https://doi.org/10.1016/j.chroma.2007.10.063.
[26] J.E. Welke, V. Manfroi, M. Zanus, M. Lazzarotto, C.A. Zini, Differentiation of wines according to grape variety using multivariate analysis of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection data, Food Chem. 141 (2013) 3897–3905. https://doi.org/10.1016/j.foodchem.2013.06.100.
[27] P.H. Stefanuto, K.A. Perrault, L.M. Dubois, B. L'Homme, C. Allen, C. Loughnane, N. Ochiai, J.F. Focant, Advanced method optimization for volatile aroma profiling of beer using two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A 1507 (2017) 45–52. https://doi.org/10.1016/j.chroma.2017.05.064.
[28] M. Brokl, L. Bishop, C.G. Wright, C. Liu, K. McAdam, J.F. Focant, Multivariate analysis of mainstream tobacco smoke particulate phase by headspace solid-phase micro extraction coupled with comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, J. Chromatogr. A 1370 (2014) 216–229. https://doi.org/10.1016/j.chroma.2014.10.057.
[29] L.M. Dubois, K.A. Perrault, P.H. Stefanuto, S. Koschinski, M. Edwards, L. McGregor, J.F. Focant, Thermal desorption comprehensive two-dimensional gas chromatography coupled to variable-energy electron ionization time-of-flight mass spectrometry for monitoring subtle changes in volatile organic compound profiles of human blood, J. Chromatogr. A 1501 (2017) 117–127. https://doi.org/10.1016/j.chroma.2017.04.026.

[30] L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry data, Talanta. 115 (2013) 887–895. https://doi.org/10.1016/j.talanta.2013.06.038.
[31] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC-TOFMS) Data Using a Null Distribution Approach, Anal. Chem. 87 (2015) 3812–3819. https://doi.org/10.1021/ac504472s.
[32] P.H. Stefanuto, K.A. Perrault, R.M. Lloyd, B. Stuart, T. Rai, S.L. Forbes, J.F. Focant, Exploring new dimensions in cadaveric decomposition odour analysis, Anal. Methods. 7 (2015) 2287–2294. https://doi.org/10.1039/c5ay00371g.
[33] L.W. Hantao, B.R. Toledo, F.A. De Lima Ribeiro, M. Pizetta, C.G. Pierozzi, E.L. Furtado, F. Augusto, Comprehensive two-dimensional gas chromatography combined to multivariate data analysis for detection of disease-resistant clones of Eucalyptus, Talanta 116 (2013) 1079–1084. https://doi.org/10.1016/j.talanta.2013.08.033.
[34] K.M. Pierce, J.C. Hoggard, Chromatographic data analysis. Part 3.3.4: Handling hyphenated data in chromatography, Anal. Methods. 6 (2014) 645–653. https://doi.org/10.1039/c3ay40965a.
[35] T. Gröger, R. Zimmermann, Application of parallel computing to speed up chemometrics for GC × GC-TOFMS based metabolic fingerprinting, Talanta 83 (2011) 1289–1294. https://doi.org/10.1016/j.talanta.2010.09.015.
[36] E. Aprea, H. Gika, S. Carlin, G. Theodoridis, U. Vrhovsek, F. Mattivi, Metabolite profiling on apple volatile content based on solid phase microextraction and gas-chromatography time of flight mass spectrometry, J. Chromatogr. A 1218 (2011) 4517–4524. https://doi.org/10.1016/j.chroma.2011.05.019.
[37] K.M. Pierce, R.E. Mohler, A review of chemometrics applied to comprehensive two-dimensional separations from 2008-2010, Sep. Purif. Rev. 41 (2012) 143–168. https://doi.org/10.1080/15422119.2011.591868.
[38] Z. Zeng, J. Li, H.M. Hugel, G. Xu, P.J. Marriott, Interpretation of comprehensive two-dimensional gas chromatography data using advanced chemometrics, TrAC - Trends Anal. Chem. 53 (2014) 150–166. https://doi.org/10.1016/j.trac.2013.08.009.
[39] S. Stadler, P.H. Stefanuto, M. Brokl, S.L. Forbes, J.F. Focant, Characterization of volatile organic compounds from human analogue decomposition using thermal desorption coupled to comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, Anal. Chem. 85 (2013) 998–1005. https://doi.org/10.1021/ac302614y.
[40] K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: Comprehensive GC × GC with ANOVA-based feature selection and principal component analysis, in: Chemom. Intell. Lab. Syst., 2002: pp. 225–237. https://doi.org/10.1016/S0169-7439(01)00198-8.
[41] F. Magagna, E. Liberto, S.E. Reichenbach, Q. Tao, A. Carretta, L. Cobelli, M. Giardina, C. Bicchi, C. Cordero, Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from thermal to differential-flow modulated comprehensive two dimensional gas chromatography, J. Chromatogr. A 1536 (2018) 122–136. https://doi.org/10.1016/j.chroma.2017.07.014.
[42] M. Schäffer, T. Gröger, M. Pütz, S. Dieckmann, R. Zimmermann, Comparative Analysis of the Chemical Profiles of 3,4-Methylenedioxymethamphetamine Based on Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry (GC×GC-TOFMS), J. Forensic Sci. 57 (2012) 1181–1189. https://doi.org/10.1111/j.1556-4029.2012.02137.x.
[43] H.D. Bean, J.E. Hill, J.M.D. Dimandja, Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography-mass spectrometry data, J. Chromatogr. A 1394 (2015) 111–117. https://doi.org/10.1016/j.chroma.2015.03.001.
[44] J.S. Nadeau, B.W. Wright, R.E. Synovec, Chemometric analysis of gas chromatography-mass spectrometry data using fast retention time alignment via a total ion current shift function, Talanta 81 (2010) 120–128. https://doi.org/10.1016/j.talanta.2009.11.046.

[45] K.M. Pierce, L.F. Wood, B.W. Wright, R.E. Synovec, A comprehensive two-dimensional retention time alignment algorithm to enhance chemometric analysis of comprehensive two-dimensional separation data, Anal. Chem. 77 (2005) 7735–7743. https://doi.org/10.1021/ac0511142.
[46] B.P. Tu, R.E. Mohler, J.C. Liu, K.M. Dombek, E.T. Young, R.E. Synovec, S.L. McKnight, Cyclic changes in metabolic state during the life of a yeast cell, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 16886–16891. https://doi.org/10.1073/pnas.0708365104.
[47] Y. Izadmanesh, E. Garreta-Lara, J.B. Ghasemi, S. Lacorte, V. Matamoros, R. Tauler, Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data, J. Chromatogr. A 1488 (2017) 113–125. https://doi.org/10.1016/J.CHROMA.2017.01.052.

Chapter 3. Class comparison enabled mass spectrum purification for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry

This chapter was reproduced from Grant S. Ochoa, Paige E. Sudol, Timothy J. Trinklein, Robert E. Synovec, "Class comparison enabled mass spectrum purification for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry" *Talanta* 236 (2022) 461401

3.1. Introduction

Comprehensive two-dimensional (2D) gas chromatography (GC×GC), has proven to be a powerful method for the separation and analysis of complex samples containing volatile/semi-volatile compounds [1–6]. Utilization of time-of-flight mass spectrometry (TOFMS) detection provides a third dimension to the GC×GC-TOFMS data structure, facilitating a confident analyte identification and accurate quantitative analysis [7–20]. Despite the many advantages for GC×GC-TOFMS, chromatographic overlap is generally inevitable, and lower abundance analyte peaks are often overshadowed by higher abundance co-eluting interference peaks [21,22]. These difficulties hinder identification and quantification of analytes of interest. Thus, advances in data analysis are an ongoing challenge in this field.

Fisher ratio (F-ratio) analysis is emerging as a powerful method for analyte discovery in "supervised" sample class comparisons that discern analyte concentration differences amid complex GC×GC-TOFMS datasets [11,12,15–20,23–25]. The F-ratio, defined as the ratio of the class-to-class variance to the sum of the within-class variance, prioritizes class distinguishing significance over absolute signal amplitude [11]. Application to GC×GC-TOFMS datasets with two (or more) sample classes results in the ranking of class distinguishing features in a hitlist. This list provides analyte hit location, mass channels (*m/z*) of interest at each hit location, and the

F-ratio for each *m/z* discovered, allowing for significant data reduction. This focused dataset is readily analyzed, producing statistics that are useful for exploiting the benefits of the "hypothesis driven" class-based experimental design; namely, which regions of the dataset statistically differ.

Two F-ratio analysis approaches have been primarily used, peak table- based [15–20] and tile-based [11,12,23–25], which is now implemented in commercial software [26,27] and can readily discover changes at the low ppm concentration level between two sample classes, while providing the 2D retention time(s) and spectrum for each analyte hit [12,23–25,28]. Recently, we explored using the tile-based F-ratio analysis outputs to obtain pure analyte *m/z*, which in turn determined the concentration ratio via the signal-ratio (S-ratio) across the two sample classes [29]. By calculating p-value and lack of fit (*LOF*) statistical metrics, sufficiently pure *m/z* for the analyte were identified from background interference *m/z* allowing for an accurate concentration ratio determination without the need for chemometric or other software-based decomposition (i.e., mathematical resolution) of the peak profiles. Beyond obtaining the concentration ratio for analyte hits, we now recognize other hidden benefits from the use of these statistical metrics to aid in analyte identification, which we explore herein. Specifically, by also focusing on the *m/z* whose statistical metrics imply unchanging signal indicative of non-analyte background components, these *m/z* signals can be used as a fine-tuning step with background normalization and subtraction to obtain a high-quality analyte mass spectrum. We refer to this new method as class comparison enabled mass spectrum purification (CCE-MSP).

Obtaining a pure analyte mass spectrum is vital to its identification. Chemometric methods are often implemented for this purpose, such as multivariate curve resolution alternating least squares (MCR-ALS), parallel factor analysis (PARAFAC), PARAFAC2, and related methods [30–40]. However, methods such as MCR-ALS can have difficulty modeling

components that are at low chromatographic resolution and/or if the desired components are minor contributors to the net signal, even in low rank environments [30]. The shortcomings in the modeling can result in a low-quality extracted spectrum. To overcome these challenges, we demonstrate herein that the class-based experimental design employed by tile-based F-ratio analysis coupled with CCE-MSP provides an alternative route to spectrum extraction, which can then serve as input constraints to improve peak profile extraction using chemometric methods.

CCE-MSP leverages information provided by tile-based F-ratio analysis whereby the "hit" spectra are initially handled as though they are composed of two components, the analyte spectrum and the background interference spectrum. At a basic level, CCE-MSP is the subtraction of the hit spectrum for a run in one sample class from the hit spectrum for an arbitrarily paired run in the other sample class. Sample-to-sample variation can alter the intensity of the underlying background interference spectrum, even after a dataset has been initially normalized via the total ion current [7,18–20] or an internal standard [9], making the quality of simple spectrum subtraction dubious. Thus, signal consistency metrics (p-value and *LOF* per *m/z*) are leveraged to optimize the hit spectrum subtraction that isolates the analyte spectrum from the background interference spectrum. The core of the CCE-MSP method is to find the pure *m/z* in the *background interference spectrum* (or simply the *interference spectrum*) using the p-value and *LOF* metrics [29], then normalize the hit spectrum from the two classes to these pure interference *m/z* signals prior to subtracting the two hit spectra to obtain the "purified" mass spectrum for the analyte hit. This process is repeated for all qualifying "pure" interference *m/z* to obtain an average analyte spectrum that has been optimally purified.

Herein, the principles of CCE-MSP are presented using JP-8 jet fuel spiked with 14 sulfur containing compounds at two concentration levels (30 ppm and 15 ppm) to go along with

the neat fuel sample, providing a dataset that expresses a wide range of interference overlap conditions. The MV for each of the 14 sulfur-containing compound "hits" is compared to the MV obtained using MCR-ALS, PARAFAC and PARAFAC2, to demonstrate that CCE-MSP provides a distinct advantage for low relative intensity analytes coupled with low 2D resolution. In doing so, we note that with CCE-MSP, unlike with similar methods, it is not necessary to know the identity of discovered analytes *a priori* [41]. Finally, for the case in which two or more analyte hits are in overlapping hit locations, we demonstrate that CCE-MSP can also be readily implemented and extended to obtain the purified spectrum for each of the analyte hits involved.

3.2. Class Comparison Enabled Spectrum Purification Principles

Application of CCE-MSP takes advantage of the sample class experimental design of tile-based F-ratio analysis and initially operates under the condition that an analyte "hit" is changing between the two classes while all other components in the hit tile are constant. This creates a simplified situation where no matter how many components are present, we can reduce the problem to two components, the analyte (A) and the background interference (Int),

$$S(m/z)_1 = k_1[R(m/z)_{\mathrm{A}}C_{\mathrm{A},1} + \sum_{\mathrm{i}=1}^{n} R(m/z)_{\mathrm{Int}}C_{\mathrm{Int},1}] \quad (3.1)$$

$$S(m/z)_2 = k_2[R(m/z)_{\mathrm{A}}C_{\mathrm{A},2} + \sum_{\mathrm{i}=1}^{n} R(m/z)_{\mathrm{Int}}C_{\mathrm{Int},2}] \quad (3.2)$$

where $S(m/z)_1$ and $S(m/z)_2$ are the hit spectra, $R(m/z)_{\mathrm{A}}$ is the analyte mass spectrum, and $R(m/z)_{\mathrm{Int}}$ is the interference spectrum, for class 1 and class 2, respectively. Likewise, the analyte concentration is $C_{\mathrm{A},1}$ and $C_{\mathrm{A},2}$, and the overall interference concentration is $C_{\mathrm{Int},1}$ and $C_{\mathrm{Int},2}$, for class 1 and class 2, respectively. Thus, under the conditions stated above, $C_{\mathrm{Int},1} = C_{\mathrm{Int},2}$. However, $C_{\mathrm{A},1} \neq C_{\mathrm{A},2}$ in order for the analyte to be discovered by the F-ratio calculation. Even though the dataset has been initially normalized via application of tile-based F-ratio analysis, relatively small yet significant differences in k_1 and k_2 initially exist in the dataset for each hit,

making the quality of simple spectrum subtraction dubious. Specifically, by assuming $k_1 = k_2$ the "initial difference" of Eq. (3.1) from Eq. (3.2) produces a rough estimate of the analyte spectrum, which can be severely contaminated by the interference spectrum,

$$S(m/z)_2 - S(m/z)_1 = \textit{Analyte Spectrum + Residual Interference Spectrum} =$$

$$= \{k_2 R(m/z)_{\mathrm{A}} C_{\mathrm{A},2} - k_1 R(m/z)_{\mathrm{A}} C_{\mathrm{A},1}\} + \{k_2 \textstyle\sum_{\mathrm{i}=1}^{n} R(m/z)_{\mathrm{Int}} C_{\mathrm{Int},2} - k_1 \sum_{\mathrm{i}=1}^{n} R(m/z)_{\mathrm{Int}} C_{\mathrm{Int},1}\} \quad (3.3)$$

where the analyte spectrum is given by,

$$\textit{Analyte Spectrum} \; = \; k_2 R(m/z)_{\mathrm{A}} C_{\mathrm{A},2} - k_1 R(m/z)_{\mathrm{A}} C_{\mathrm{A},1} \quad (3.4)$$

and the residual interference spectrum is given by,

$$\textit{Residual Interference Spectrum} = \; k_2 \textstyle\sum_{\mathrm{i}=1}^{n} R(m/z)_{\mathrm{Int}} C_{\mathrm{Int},2} - k_1 \sum_{\mathrm{i}=1}^{n} R(m/z)_{\mathrm{Int}} C_{\mathrm{Int},1} \quad (3.5)$$

Note that the spectrum in Eq. (3.4) is indeed that of the analyte, $R(m/z)_{\mathrm{A}}$, which has simply been scaled by the quantity $k_2 C_{\mathrm{A},2} - k_1 C_{\mathrm{A},1}$. Furthermore, when the interference spectrum is large relative to the analyte spectrum in the context of their relative concentration, $R(m/z)_{\mathrm{A}} C_{\mathrm{A}} << R(m/z)_{\mathrm{Int}} C_{\mathrm{Int}}$, then the residual interference spectrum quantified by Eq. (3.5) can produce a contaminated analyte spectrum via Eq. (3.3) unless either $k_1 = k_2$, or additional normalization is applied (as is performed by CCE-MSP). The severity of the spectrum contamination will depend upon the ratio of the interference signal, s_{Int}, to the analyte signal, s_{A}, both scalars, obtained by summing the spectrum contributions defined in Eqs. (3.1) and/or (3.2),

$$\frac{s_{Int}}{s_{\mathrm{A}}} = \frac{\sum \sum_{\mathrm{i}=1}^{n} R(m/z)_{\mathrm{Int}} C_{\mathrm{Int}}}{\sum R(m/z)_{\mathrm{A}} C_{\mathrm{A}}} \quad (3.6)$$

Thus, to obtain the analyte spectrum, pure interference *m/z* are initially identified using the p-value and *LOF* information. Then, the ratio k_1/k_2 is calculated using the signal at the pure interference *m/z* from the hit spectra $S(m/z)_1$ and $S(m/z)_2$,

$$\textit{Normalization Factor} = \frac{s(m/z)_1}{s(m/z)_2} = \frac{k_1}{k_2} \quad (3.7)$$

where s(*m/z*) is a scalar. Normalization factors from Eq. (3.7) can be applied to $S(m/z)_2$ to eliminate the residual interference spectrum contamination from Eq. (3.5),

$$\textit{Residual Int. Spectrum} = \frac{k_1}{k_2} k_2 \sum_{i=1}^{n} R(m/z)_{\mathrm{Int}} C_{\mathrm{Int},2} - k_1 \sum_{i=1}^{n} R(m/z)_{\mathrm{Int}} C_{\mathrm{Int},1} = 0 \quad (3.8)$$

The ratio k_1/k_2 is determined for all pure interference *m/z* that qualify via p-value and *LOF* metrics. When such conditions in Eq. (3.8) are met, the normalized hit spectra may be subtracted from each other to reveal the purified analyte spectrum,

$$(\frac{k_1}{k_2}) S(m/z)_2 - S(m/z)_1 = R(m/z)_{\mathrm{A}} k_1 (C_{\mathrm{A},2} - C_{\mathrm{A},1}) \quad (3.9)$$

whereby the analyte spectrum $R(m/z)_{\mathrm{A}}$ is scaled by the quantity $k_1(C_{\mathrm{A},2} - C_{\mathrm{A},1})$. Note that this scaling of the analyte spectrum does not impact its utility for identification purposes or for input as a constraint into other computational tools, e.g., MCR-ALS. Equation (3.9) is derived with the condition that $C_{\mathrm{A},1} < C_{\mathrm{A},2}$, to obtain a "positive" mass spectrum. However, the signal ratio (S-ratio) for each analyte hit, i.e., the ratio of Eq. (3.2) to Eq. (3.1), yields an accurate concentration ratio, $C_{\mathrm{A},2}/C_{A,1}$, using pure analyte *m/z* based upon p-value and *LOF* metrics [29],

$$S - ratio = \frac{s(m/z)_2}{s(m/z)_1} = (\frac{C_{\mathrm{A},2}}{C_{\mathrm{A},1}})/(\frac{k_1}{k_2}) \quad (3.10)$$

Thus, since the concentration ratio for each hit is determined, the assignment of class 1 versus class 2 when applying Eq. (3.9) can be flipped such that $C_{\mathrm{A},1} < C_{\mathrm{A},2}$, to obtain a positive mass spectrum. We will experimentally observe that k_1/k_2 varies from 1 by about ± 0.25 (~25%). This variation in k_1/k_2 will be demonstrated to not impact the accuracy of the concentration ratios determined by the S-ratio via Eq. (3.10), nearly as much as when the same variation is manifested in the residual interference spectrum in Eq. (3.5), especially when the analyte signal is much less relative to the interference signal as expressed by Eq. (3.6).

3.3. Experimental

JP8 jet fuel was spiked with a mixture of 14 sulfur-containing compounds at concentrations of 30 ppm (mg/kg) and 15 ppm per compound. Sulfur is present in aviation fuels at a typical level of ~ 600 ppm sulfur, while the content in ultra-low sulfur jet fuel (ULSJ) is 15 ppm, thus, it is reasonable to find sulfur compounds in jet fuel at the spiked levels [42]. The GC×GC-TOFMS method for the analysis of the spiked and neat jet fuel dataset has been previously reported [28], so the details are provided in the Supplementary Material. The GC×GC-TOFMS chromatograms were imported into MATLAB 2019a (The Mathworks Inc., Natick, MA) from the LECO ChromaTOF for BT software v5.20 (LECO, St. Joseph, MI) via a ChromaTOF data converter. The chromatograms were normalized to the total ion current (TIC) intensity, baseline corrected, followed by tile-based F-ratio analysis to "discover" the spiked analytes [11,12,23–25]. This was performed for three "4 versus 4" sample class comparisons: 30-ppm versus 15-ppm, 30-ppm versus neat, and 15-ppm versus neat, although the primary focus is on the 30-ppm versus 15-ppm comparison. The following conditions and steps were applied to all three F-ratio analyses. The tile dimensions were 12 s on ^{1}D by 600 ms on ^{2}D with a *S/N* threshold of 10, and the cluster tile dimensions were 12 s on ^{1}D by 350 ms on ^{2}D. The maximum F-ratio (1 *m/z*) was then used to rank the hitlist [28]. However, all *m/z* signals passing the *S/N* threshold for each of the 14 sulfur-containing analytes were saved and further analyzed to obtain their p-value, *LOF*, and S- ratio [29]. Also, all *m/z* signals that pass the *S/N* threshold define the "F-ratio spectrum," which is the F-ratio as a function of *m/z*. The peak shape consistency statistic, *LOF*, was calculated as follows. The data within a "*LOF* tile" centered upon the hit location (with a ^{1}D dimension of 9 s by ^{2}D dimension of 600 ms) was extracted from each *m/z* for a given hit from both sample classes. The data were normalized such that corresponding modulations between classes have the same area. Each modulation was interpolated such that the

peak shape remains intact and aligned between classes. The data were then unfolded and concatenated into a single vector and the *LOF* was calculated,

$$LOF\ (m/z)\ = 100\sqrt{\frac{\sum_i (y-x)_i^2}{\sum_i x_i^2}} \tag{3.11}$$

where *x* is the concatenated pattern of ^{2}D peaks for one class and *y* is the concatenated pattern of ^{2}D peaks for the other class [40]. These were calculated for four pairs of samples which were then averaged yielding one *LOF* per *m/z*. Peaks that are consistent between classes result in a low *LOF*, while perturbations in the peak shape between classes results in a higher *LOF*. A low *LOF* can be obtained for pure analyte *m/z* that experience a concentration change between classes (i.e., the hit with a high F-ratio), as well as for pure *m/z* for the interference species that remain unchanged between classes. Indeed, it is the p-value statistic that distinguishes between these two cases. A high *LOF* with an intermediate p-value is generally observed for interference species *m/z* that are interfering with the analyte.

The combination of the p-value and *LOF* statistics facilitate CCE-MSP. On the heels of the F-ratio analysis, for each hit of interest the signal was examined, with the p-value and *LOF* calculated on a *m/z* basis. The p-value was calculated from the signals of all samples per class. However, the *LOF* was calculated by arbitrarily pairing the signals between the two classes. Based upon the p-value and *LOF* per *m/z*, "pure" interference *m/z* were identified for each hit; this could be several *m/z*. To qualify as a pure interference *m/z*, a p-value ≥ 0.1 and a *LOF* $\leq 20\%$ was required. The k_1/k_2 normalization factor in Eq. (3.7) was then calculated for the qualifying *m/z*. The four hit spectra per class were then arbitrarily paired, and the class 2 hit spectrum was then normalized to the class 1 hit spectrum by each individual k_1/k_2 of a given *m/z*, equalizing the interference component between classes per paired spectra per Eq. (3.8). The paired hit normalized spectra were subtracted from each other, resulting in the purified analyte spectrum

per Eq. (3.9), with their mass spectrum match value (MV) calculated using an in-house library spectrum collected on the same instrument using a standard mixture of the 14 sulfur containing analytes. A flow chart for the method workflow is provided in the Supplementary Material (Fig. B1).

These spectra obtained by CCE-MSP were compared against three chemometric methods, whereby the analyte hits were processed using MCR-ALS, PARAFAC, and PARAFAC2 as part of the PLS Toolbox 8.8 (Eigenvector Research Inc., Wenatchee, WA) to obtain the isolated analyte spectra. Starting with MCR-ALS, a section of data with $^{1}D \times {}^{2}D$ dimensions of three modulations (9 s) by 600 ms was cut out around each hit location for each sample which are then augmented and submitted to MCR-ALS. The number of components input to MCR-ALS was determined by visual inspection of the chromatographic section in both the TIC and selective analyte *m/z* found by F-ratio analysis, and then testing ± 2 components to optimize the rank determination. The MCR-ALS model with the lowest number of components and whose spectrum loadings best matched the in-house analyte library spectrum was selected and a MV recorded. Additionally, using the rank determined from this step, data was submitted for each hit to both PARAFAC and PARAFAC2 with the three-way data structure retention × spectra × samples, obtaining modeled spectra for each hit. A MV was also calculated for each of these spectra for comparison with CCE-MSP and MCR-ALS.

3.4. Results and Discussion

The GC×GC-TIC chromatogram of the JP-8 jet fuel spiked with the 14 sulfur containing compounds at the 15-ppm level is presented in Fig. 3.1A. Tile-based F-ratio analysis was applied for the “discovery” of the spiked sulfur containing compounds for the comparison of 30 ppm versus 15 ppm spiked JP-8 jet fuel. All 14 spiked analytes possessed a high F-ratio as indicated in the F-ratio distribution in Fig. 3.1B, as summarized in the hitlist in Table 3.1. A mass

spectrum MV was calculated for each of the initial pin locations in the GC×GC chromatogram, by comparison to an in-house library reference spectrum for each analyte. All of the hits initially had a very low MV with large variation indicating a poor mass spectrum quality at each hit location in the 15 ppm class chromatograms. The S-ratio, p-value, and *LOF* were calculated on a per *m/z* basis for each analyte hit. Pure analyte *m/z* result in a high F-ratio and an S-ratio that is a good estimate of the concentration ratio per Eq. (3.10) [29]. This relationship between S-ratio and F-ratio is observed in Fig. 3.1C where a majority of the spiked analytes have the expected concentration ratio ≈ 2 (30 ppm/15 ppm) for pure analyte *m/z* at higher F-ratio, then dropping down to a concentration ratio ≈ 1 for interference *m/z* that also exhibit a lower F-ratio.

Table 3.1. Analyte hitlist obtained from tile-based F-ratio analysis. All hits are ranked by their top F-ratio m/z alongside the p-value and LOF statistics from the S-ratio analysis as well as the average match value (MV) of the hit spectrum (with standard deviation) at the pin location against in-house library spectra. Redundant hits per Fig. 3.1B have been omitted for clarity.

Hit	Analyte	^1t_R, min	^2t_R, s	F-ratio	Match Value at Pin location
1	2,5-dimethylthiophene	7.25	0.50	27873	482 ± 126
2	2-methylthiophene	5.00	0.14	17471	643 ± 205
3	thiophene	3.20	2.69	14225	611 ± 233
4	2-chloroethyl phenyl sulfide	23.85	0.18	12192	455 ± 58
5	3-methylthiophene	5.20	0.21	8576	637 ± 198
6	2-butyl-5-ethylthiophene	17.85	0.63	6731	276 ± 65
7	2-methylbenzothiophene	21.8	0.22	6289	330 ± 47
8	2-hexylthiophene	18.40	0.70	4563	427 ± 146
9	tetrahydrothiophene	6.30	0.25	3289	544 ± 65
10	2-propylthiophene	9.75	0.42	2314	368 ± 82
11	benzothiophene	19.30	0.13	1441	418 ± 53
12	1,4-thioxane	9.35	0.12	1257	507 ± 50
13	3-methylbenzothiophene	22.45	0.13	1060	381 ± 40
14	3-acetyl-2,5-dimethylthiophene	20.45	0.27	870	413 ± 55

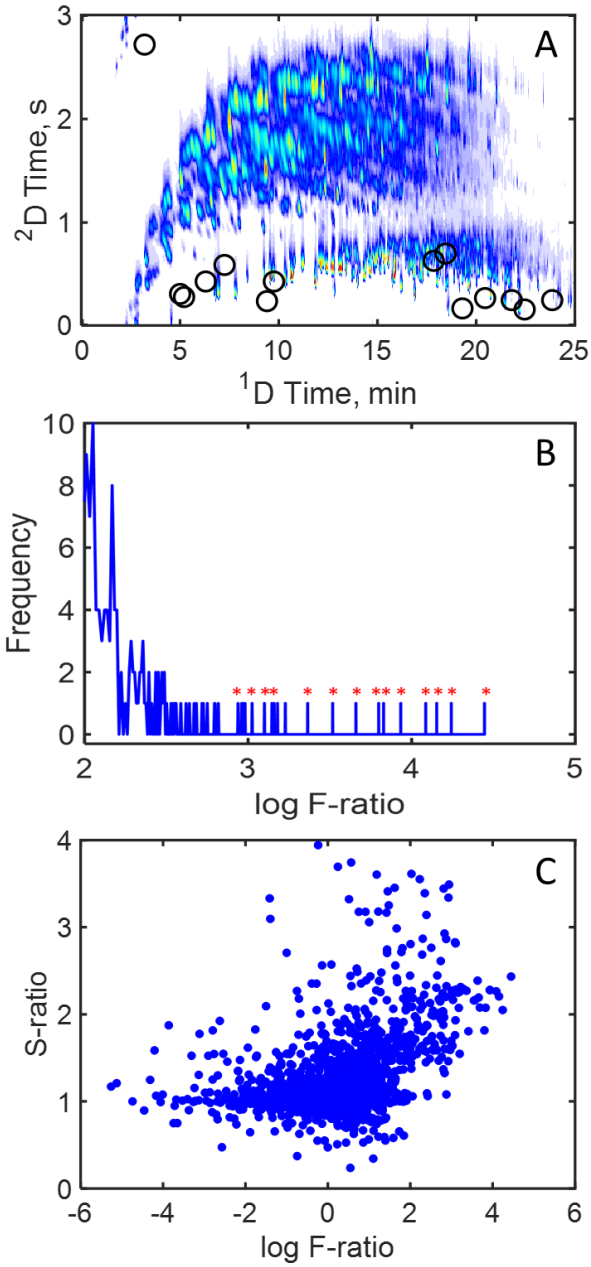


Figure 3.1. (A) GC×GC-TIC chromatogram of JP-8 jet fuel spiked with 14 sulfur containing analytes spiked at 15 ppm. Circles indicate the analyte locations. (B) F-ratio distribution for the 30 ppm versus 15 ppm spike level comparison, ranked by the top F-ratio per m/z for each hit. The 14 sulfur containing analytes are indicated by asterisk; the F-ratio for each analyte hit exceeded 800, lifting them to the top of the hitlist. A few redundant hits from the spiked analytes are interspersed with the primary hits. (C) S-ratio for each of the 14 spiked sulfur containing analytes plotted against their respective F-ratio, on a per m/z basis, i.e., all m/z with S/N > 10. The nominal concentration ratio was ~ 2, i.e., 30 ppm/15 ppm, which can be seen is reflected by the S-ratios at the higher F-ratios which are likely to be pure analyte m/z [29].

The CCE-MSP method is demonstrated for 2-butyl-5-ethylthiophene (hit 6) in Fig. 3.2. The TIC at the hit location for the 30 ppm (Fig. 3.2A) and 15 ppm (Fig. 3.2B) spike chromatograms look identical since the analyte concentration level and change is insignificant in the TIC. Figure 3.2C shows this location for analyte selective *m/z* 125, while Fig. 3.2D shows *m/z* 145 which is selective for the most intense overlapping interference peak. Comparison of Figs. 3.2C,D indicates the analyte and interference are highly overlapped with a 2D resolution of ~ 0.2, determined based upon the Euclidean distance separating the two peaks utilizing their respective 2D retention times and peak widths [43]. Furthermore, the maximum intensity on the color scale for Fig. 3.2C for the analyte is 10-fold less than that of Fig. 3.2D for the interference peak, and 100-fold less than that of Figs. 3.2A and 3.2B for the TIC data, indicating the s_{Int}/s_A ratio per Eq. (3.6) for 2-butyl-5-ethylthiophene is approaching 100. Due to the low 2D resolution and low relative analyte abundance, this is a severely challenging situation to provide a high-quality analyte mass spectrum. Testing this, MCR-ALS was applied in an effort to obtain the mass spectrum for 2-butyl-5-ethylthiophene. Unfortunately, MCR-ALS was unable to accurately model the analyte, with the closest matching component having on average a MV of 476 with a concentration ratio of 1.01 ± 0.08. In this case, MCR-ALS is modeling the background interference in the data with negligible contribution from the analyte. In contrast, as we shall see, CCE-MSP leverages the class-based experimental design of the F-ratio analysis in favor of the analyte to obtain a pure spectrum even in these circumstances. Additional plots for three other analytes that significantly benefited from CCE-MSP are provided in the Supplementary Material (Figs. B2-B4).

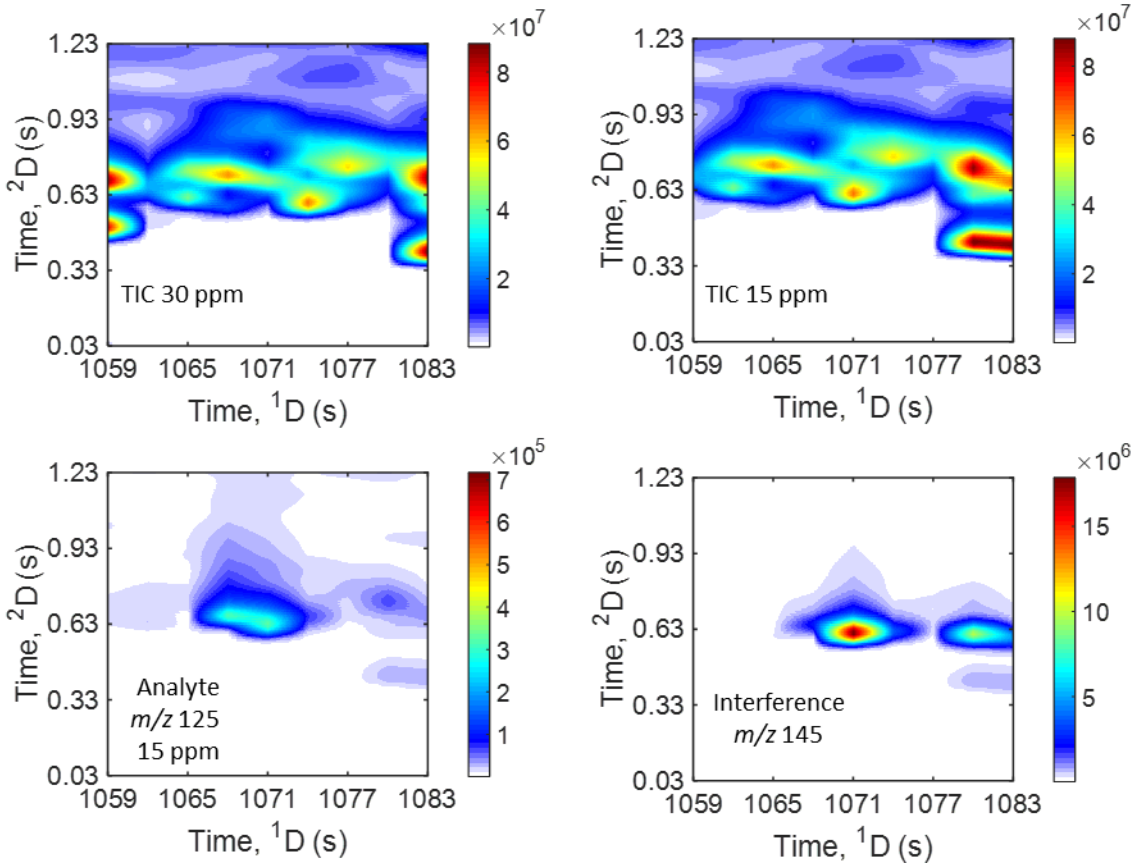


Figure 3.2. (A) Section of the JP-8 jet fuel GC×GC-TIC chromatogram showing the location of 2-butyl-5-ethylthiophene (hit 6) spiked at 30 ppm. (B) The same section as (A), now at the 15-ppm spike level. (C) The same section of the 15-ppm spiked JP-8 jet fuel chromatogram is displayed for a selective and most sensitive m/z 125 for the analyte. (D) A highly overlapped, and intense interference peak at a selective m/z 145 in the 15-ppm spiked JP-8 jet fuel. The difference in the intensity scales indicate the analyte is at much lower signal than the background interference.

For CCE-MSP, only *m/z* experiencing a signal change between classes belong to the analyte hit (unless there is more than one analyte changing between classes in overlapping hit tiles which can also be handled as will be demonstrated), all other *m/z* contribute to the interference. This is demonstrated for data obtained for 2-butyl-5-ethylthiophene, where the hit spectrum at the 30-ppm spike level (Fig. 3.3A) is subtracted from the hit spectrum at the 15-ppm spike level (Fig. 3.3B), which in principle provides the analyte spectrum (Fig. 3.3C). However, in practice, run-to-run background variation due to detector noise, injection variation, and

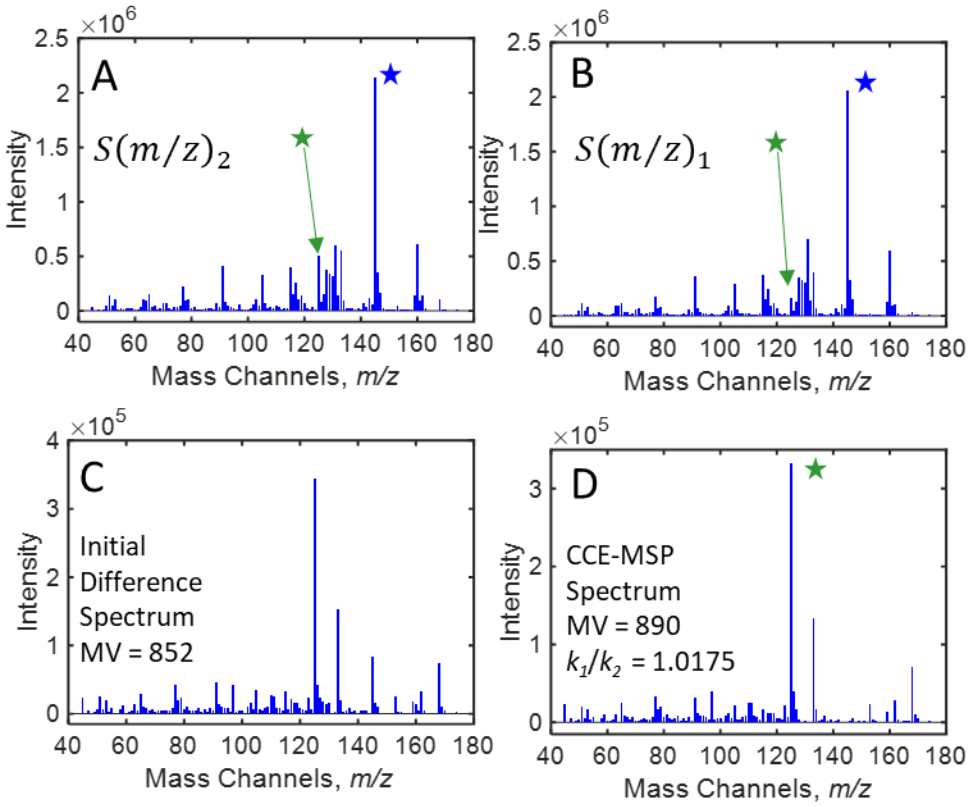


Figure 3.3. Illustration of implementation of CCE-MSP using data for 2-butyl-5-ethylthiophene (hit 6). (A) The hit spectrum at the 15-ppm spike level, $S(m/z)_2$. (B) The hit spectrum at the 30-ppm spike level, $S(m/z)_1$. (C) Initial difference spectrum using Eq. (3.3) is obtained by subtracting the spectrum $S(m/z)_1$ in (B) from the spectrum $S(m/z)_2$ in (A) without normalization ($k_1/k_2 = 1$), resulting in a MV = 852 relative to the in-house library spectrum. (D) The purified analyte spectrum is obtained by applying CCE-MSP via Eq. (3.9). The spectrum $S(m/z)_2$ is normalized by k_1/k_2, in this example equal to 1.0175, such that the signal of the blue starred pure interference m/z 145 is the same between both hit spectra. The two normalized hit spectra are then subtracted from each other leaving the purified analyte spectrum that has a MV = 890 relative to an in-house library spectrum. The green starred m/z 125 is the most intense m/z for 2-butyl-5-ethylthiophene, initially a small feature in either hit spectrum.

differences in the baseline correction must be corrected. This is accomplished by CCE-MSP by normalizing to "pure" interference *m/z* via Eqs. (3.7) and (3.8), illustrated in Fig. 3.3D, whereby in the hit spectra the blue star denotes the interference *m/z* 145 used for normalization to obtain k_1/k_2 (equal to 1.0175 in this example) while the green star denotes the top F-ratio *m/z* 125 for 2-butyl-5-ethylthiophene. When k_1/k_2 normalization is applied via Eq. (3.9), the purified analyte spectrum is obtained (Fig. 3.3D). The purified analyte spectrum has a MV = 890 compared with the in-house library spectrum, whereas the initial difference spectrum via Eq. (3.3) in Fig. 3.3C

had a MV = 852 (prior to k_1/k_2 normalization). An example of CCE-MSP applied to 2-butyl-5-ethylthiophene with a k_1/k_2 much less than 1 is provided in the Supplementary Material (Fig. B5). Beyond this illustration in Fig. 3.3, objective selection of pure interference *m/z* on a larger scale is required. Qualifying *m/z* must meet criteria based upon peak signal intensity and shape consistency. Signal intensity consistency is handled by the p-value, a statistic that determines how likely the two populations are the same, with smaller numbers indicating a higher likelihood to be different. Class-to-class peak shape differences are quantified by calculating the *LOF* metric per Eq. (3.11).

The *LOF* statistic is demonstrated for the selective analyte *m/z* 125 and interference *m/z* 145 for 2-butyl-5-ethylthiophene in Fig. 3.4A-B, respectively. Peak shape comparisons between classes in Fig. 3.4A for *m/z* 125 results in a relatively low *LOF* = 6.2%, with a p-value = 1.9×10^{-9}. The combination of the low *LOF* and low p-value is indicative of a pure analyte *m/z*, thus, the S-ratio = 1.99 is an accurate estimate of the concentration ratio [29]. Likewise, for *m/z* 145 in Fig. 3.4B, the peak profiles are nearly identical, resulting in a very low *LOF* = 1.8%, however the p-value = 0.54. Now the combination of the low *LOF* and high p-value (S-ratio = 0.9916) is indicative of a pure interference *m/z*. Identification of pure interference *m/z* using the p-value and *LOF* facilitates spectrum normalization by k_1/k_2 to obtain a purified analyte spectrum per Fig. 3.3.

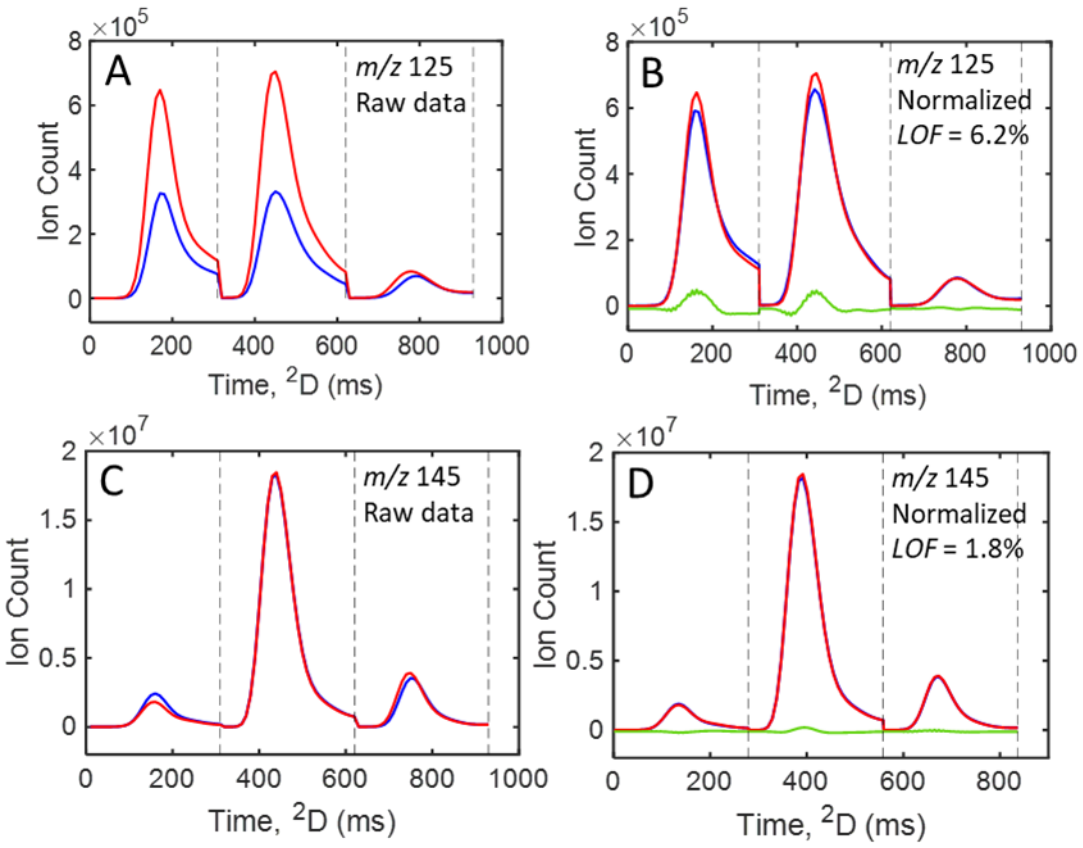


Figure 3.4. Demonstration of the LOF statistic calculations for 2-butyl-5-ethylthiophene (hit 6) for the unfolded and concatenated tile of data around the hit location. The red trace is the 30-ppm analyte spike level, while the blue trace represents the 15-ppm analyte spike level. Two m/z are presented, m/z 125, selective for the analyte (top), and m/z 145, selective for the interference (bottom). For each m/z, the left plot provides the original chromatographic data, while the right plot provides the data after they have been normalized to the same area between samples classes, while the green traces are the residuals between the normalized data.

Utilizing 2-butyl-5-ethylthiophene, the relationship between p-value, *LOF*, and S-ratio is interrogated in Fig. 3.5. Sets of plots as in Fig. 3.5 are provided for all 14 analytes in Fig. B6 in the Supplementary Material. By examining various metric combinations, the requirements to identify pure analyte *m/z* and pure interference *m/z* are illuminated, and appropriate thresholds can be established. The S-ratio is plotted in relation to the *LOF* (Fig. 3.5A), the p-value (Fig. 3.5B), and the p-value after filtering *m/z* with high *LOF* and low signal (Fig. 3.5C). Because the spike level was a 2-fold concentration change, pure analyte *m/z* exhibit an S-ratio ≈ 2, while pure

interference *m/z* exhibit an S-ratio ≈ 1. Any *m/z* composed of analyte and interference will have an intermediate S-ratio. Likewise, pure analyte *m/z* and pure interference *m/z* exhibit a low *LOF*, as in Fig. 3.4, while *m/z* composed of analyte and interference have a high *LOF*. Based upon previous studies, only *m/z* with $LOF \leq 20\%$ were selected as good candidates for either pure analyte or pure interference *m/z* [29,40]. The S-ratio versus *LOF* plot in Fig. 3.5A indicates there are 19 *m/z* with a LOF > 20%, which are ultimately discarded during CCE-MSP. The S-ratio versus p-value plot in Fig. 3.5B identifies *m/z* that indicate whether or not the two sample classes are statistically different in "apparent" concentration, expressed as the S-ratio. While the S-ratios for 2-butyl-5-ethylthiophene spanned the range from 1 to 2 along a continuum due to various degrees of analyte-to-interference overlap, p-value thresholds were also implemented for the purpose of categorizing sufficiently pure analyte and interference *m/z*. Taking into account the p-value and *LOF* results for all 14 analytes, a p-value $\leq 10^{-6}$ threshold was selected to be categorized as a pure analyte *m/z* (S-ratio ≈ 2), while a p-value $\geq 10^{-1}$ was selected to be categorized as a pure interference *m/z* (S-ratio ≈ 1). Putting the *LOF* and p-value thresholds to work, a refined version of Fig. 3.5B is provided in Fig. 3.5C. Here, a $LOF \leq 20\%$ threshold was applied first to filter high variability peaks followed by a p-value filter to separate pure analyte *m/z* (p-value $\leq 10^{-6}$, green dots) from pure interference *m/z* (p-value $\geq 10^{-1}$, blue dots). Furthermore, a 5% signal threshold of the maximum signal among the interference *m/z* was applied to remove inclusion of *m/z* with low *S/N* whose variability can cause highly errant k_1/k_2 normalization ratios. Next, all of the qualifying pure interference *m/z* were used to normalize the paired spectra between classes, via the k_1/k_2 applying Eqs. (3.7) – (3.9), to obtain a purified analyte spectrum as illustrated in Fig. 3.4 for just one pure interference *m/z*. For example, if there are *n* pure interference *m/z* for a hit, there will be 4*n* purified analyte spectra obtained, since this

is a 4 sample versus 4 sample F-ratio comparison. The resulting spectra for each sample pair are then averaged to obtain 4 analyte spectra. The MV was determined for each of these 4 average spectra, with a final average reported for the MV. More details on the k_1/k_2 normalization are provided in the Supplemental Material (Fig. B7).

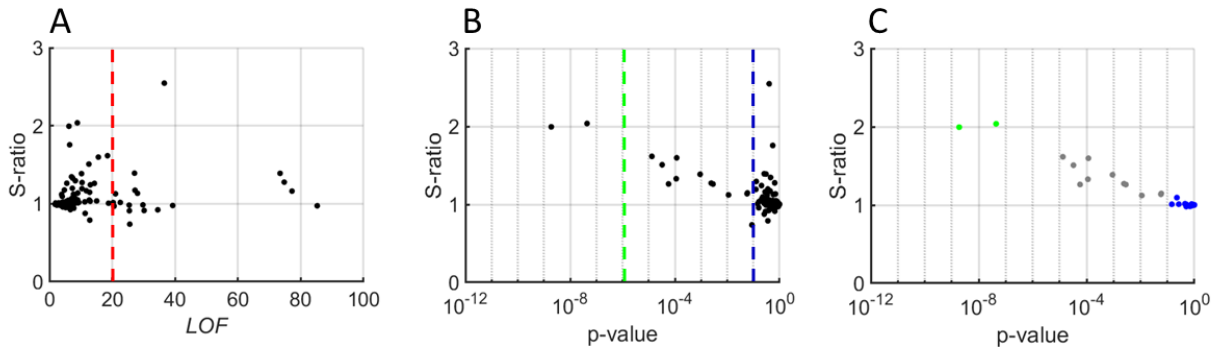


Figure 3.5. Statistical comparison plots of S-ratio, p-value, and LOF for 2-butyl-5-ethylthiophene (hit 6) where each dot corresponds to a specific m/z. (A) The S-ratio versus LOF plot separates the m/z depending upon the relative analyte-to-interference contributions. Pure interference m/z exhibit a low LOF with S-ratios ≈ 1, pure analyte m/z also exhibit a low LOF but with S-ratios ≈ 2, and m/z with intermediate S-ratios ranging from 1 to 2 exhibit a much higher LOF. A LOF ≤ 20% was selected as the threshold for pure analyte or pure interference m/z (dashed red line) [30]. (B) The S-ratio versus p-value plot identifies m/z that indicate whether or not the two sample classes are statistically different in "apparent" concentration, expressed as the S-ratio. While the S-ratios spanned the range from 1 to 2 along a continuum due to various degrees of analyte-to-interference overlap, p-value thresholds can be implemented to categorize sufficiently pure analyte and interference m/z. A p-value ≤ 10^{-6} was selected as a threshold for pure analyte m/z (dashed green line) while p-value ≥ 10^{-1} was selected for pure interference m/z (dashed blue line). (C) The S-ratio versus p-value plot from (B) has been filtered, keeping m/z from (A) with a LOF ≤ 20% and distinguishing the remaining m/z according to p-value: analyte m/z (p-value ≤ 10^{-6}, green dots) and interference m/z (p-value ≥ 10^{-1}, blue dots), and mixed component m/z (gray dots). Additionally, a 5% signal threshold was applied to the interference m/z to remove inclusion of low S/N m/z.

Table 3.2. Summary of CCE-MSP results for the 14 sulfur containing analytes in the 30 ppm versus 15 ppm spike level comparison. For each analyte, from left to right, the 2D resolution, the interference-to-analyte signal ratio, s_{Int}/s_A via Eq. (3.6), using the 15 ppm spike level data, and the average match value (MV) for the initial difference spectrum via Eq. (3.3), CCE-MSP spectrum via Eq. (3.9), MCR-ALS spectrum, PARAFAC spectrum, and PARAFAC2 spectrum using the in-house library spectrum are provided.

Hit	Analyte	2D Resolution	s_{Int}/s_A	MV Initial	MV CCE-MSP	MV MCR-ALS	MV PARAFAC	MV PARAFAC2
1	2,5-dimethylthiophene	0.89	1.98	920 ± 17	920 ± 10	876	965	956
2	2-methylthiophene	1.34	1.29	953 ± 6	955 ± 6	941	959	954
3	thiophene	1.87	0.38	989 ± 1	989 ± 1	956	952	945
4	2-chloroethyl phenyl sulfide	0.36	1.13	891 ± 89	942 ± 13	854	920	892
5	3-methylthiophene	1.34	1.49	901 ± 16	902 ± 5	896	855	912
6	2-butyl-5-ethylthiophene	0.2	86.78	759 ± 73	831 ± 11	476	403	336
7	2-methylbenzothiophene	0.5	0.32	974 ± 3	968 ± 7	940	945	931
8	2-hexylthiophene	0.49	5.53	899 ± 22	928 ± 20	674	670	705
9	tetrahydrothiophene	0.89	2.49	936 ± 9	937 ± 11	707	810	828
10	2-propylthiophene	0.27	37.22	786 ± 131	965 ± 7	803	812	649
11	benzothiophene	1.1	0.28	986 ± 2	982 ± 5	953	979	955
12	1,4-thioxane	1.13	3.87	906 ± 21	918 ± 11	845	841	771
13	3-methylbenzothiophene	1.78	0.14	989 ± 3	991 ± 3	974	987	991
14	3-acetyl-2,5-dimethylthiophene	0.67	1.97	935 ± 5	939 ± 6	858	808	883

The CCE-MSP method was also carried out for the remaining 13 sulfur containing analytes, with a MV summary provided in Table 3.2, along with the 2D resolution between the analyte and primary interference peak, and the interference-to-analyte signal ratios, s_{Int}/s_A. The MV relative to the in-house library was determined for each analyte spectrum obtained by five methods: the initial difference spectrum by Eq. (3.3), the purified mass spectrum using CCE-MSP by Eq. (3.9), and the mass spectrum obtained using the three chemometric methods: MCR-ALS, PARAFAC and PARAFAC2. The interference-to-analyte signal ratios, s_{Int}/s_A, were obtained by summing signals at each *m/z* for the interference spectrum and the analyte spectrum using Eq. (3.6). To measure the 2D resolution for each analyte hit, the retention time difference along both

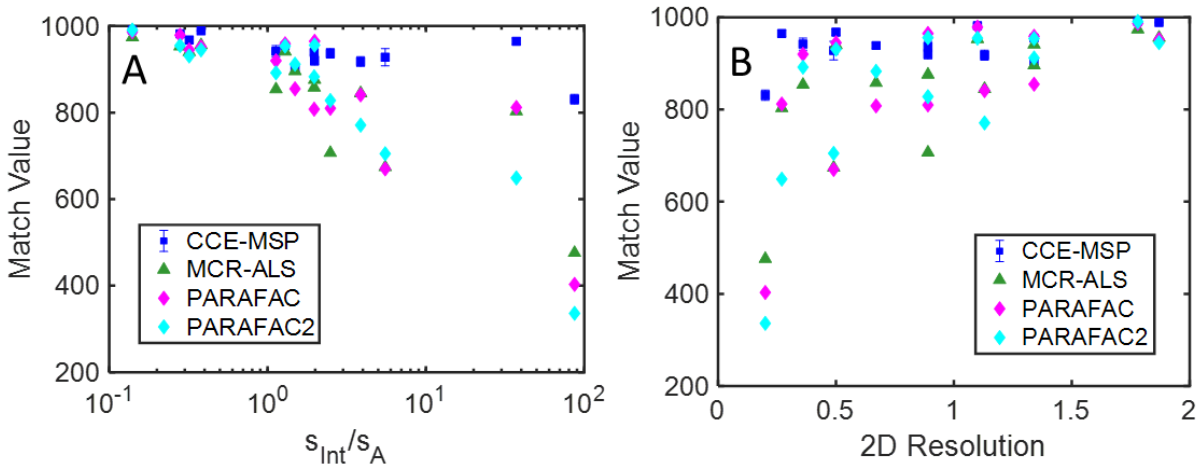


Figure 3.6. (A) Match values (MV) of each analyte provided in Table 3.2 plotted against their respective interference-to-analyte signal ratios, s_{Int}/s_A. Blue circles represent the initial subtraction spectrum results without normalization, light and dark green triangles represent the result of the best matching component from MCR-ALS for 15 ppm and 30 ppm spike level samples respectively, and red squares represent the CCE-MSP method results. (B) Match values (MV) of each analyte provided in Table 3.2 plotted against their respective 2D resolution with the primary interference peak.

GC dimensions was obtained for the interference peak nearest to the spiked analyte. Peak widths were measured using the peak profile for the top F-ratio *m/z* for the analyte and one of the selected pure interference *m/z*. The peak widths and retention time measures were used to calculate the 2D resolution [43]. The results in Table 3.2 are plotted in Fig. 3.6, with the MV versus s_{Int}/s_A in Fig. 3.6A, and the MV versus 2D resolution in Fig. 3.6B. The initial difference spectrum results reported in Table 3.2 are omitted in Fig. 3.6 for clarity. CCE-MSP produces a higher quality analyte mass spectrum relative to the three chemometric methods, approximately when $s_{Int}/s_A > 1$ and/or when 2D resolution < 0.5. Furthermore, the theoretical principles for when the analyte hit is in only one of the two classes is provided in Supplementary Material, along with complementary table and plots comparing MVs for the 30 versus neat and 15 ppm versus neat F-ratio comparisons (Table B1, Fig. B8), which provide the same conclusions as Fig. 3.6.

While this proof-of-principle study has not definitively uncoupled the relative impacts of s_{Int}/s_A and 2D resolution, the results are highly supportive of the benefit of applying CCE-MSP. The trend in mass spectrum quality deterioration using MCR-ALS when the relative contribution from the background interference increases has been previously reported [30]. This is due to the tendency of MCR-ALS to return one of many feasible solutions for samples with heavy spectrum and temporal overlap due to rotational ambiguity of the decomposition matrices [44]. PARAFAC and PARAFAC2 also struggled to provide a high-quality analyte spectrum when s_{Int}/s_A and 2D resolution were unfavorable. The ^{2}D retention times are reproducible to within ±30 ms, while the ^{1}D retention times were very reproducible with peaks shifting no more than a fraction of the modulation period, resulting in a slight change in the phasing of the modulated ^{2}D peak patterns. Thus, the performance of all these chemometric methods was not hampered by retention time reproducibility. Instead, we assert that these shortcomings to be due to a "chemometric" multiplex disadvantage, similar to the well-known multiplex disadvantage observed with some spectroscopic instrumentation, whereby the noise of large signals at one wavelength interferes with measuring small signals in other wavelengths for simultaneous collection techniques such as Fourier and Hadamard transform [45–48]. For the chemometric methods examined herein, this disadvantage presents itself since the entire *m/z* dimension is simultaneously analyzed and minimizing residuals for the modeled components introduces adverse effects such as modeling noise and/or the signal from all components from one *m/z* into another. In contrast, CCE-MSP takes its weakness which is the analyte at low signal abundance and turns it into a strength to readily determine sufficiently pure interference *m/z* to provide an accurate normalization step. Overall, CCE-MSP serves as a useful tool to obtain a high-quality analyte spectrum when there

is moderate to severe interference overlap where the use of typically applied chemometric decomposition approaches may not effectively isolate the analyte spectrum.

Table 3.3. Concentration ratio (CR), defined as the analyte concentration in the 30 ppm spike level data relative to the 15 ppm spike level data for the 14 spiked sulfur containing analytes using the S-ratio, and MCR-ALS both unconstrained and constrained using the spectrum obtained by CCE-MSP. Analyte 2D resolution with the closest interference peak, along with the interference-to-analyte signal ratio, s_{Int}/s_A, using the 15 ppm spike level data to give chromatographic context. Analytes that did not possess any pure analyte m/z (labeled as N/A, so an accurate S-ratio was not provided), did possess pure interference m/z so were quantified using CCE-MSP by initially obtaining the purified analyte spectrum which was then used as a constraint for MCR-ALS.

Hit	Analyte	2D Resolution	s_{Int}/s_A	CR using S-ratio	MCR-ALS Unconstrained CR	MCR-ALS Constrained CR
1	2,5-dimethylthiophene	0.89	1.98	2.38 ± 0.06	2.45 ± 0.14	2.38 ± 0.12
2	2-methylthiophene	1.34	1.29	1.99 ± 0.09	2.08 ± 0.10	2.07 ± 0.04
3	thiophene	1.87	0.38	2.17 ± 0.05	2.19 ± 0.08	2.19 ± 0.07
4	2-chloroethyl phenyl sulfide	0.36	1.13	2.26 ± 0.11	2.13 ± 0.10	2.43 ± 0.08
5	3-methylthiophene	1.34	1.49	2.28 ± 0.06	2.15 ± 0.09	2.26 ± 0.10
6	2-butyl-5-ethylthiophene	0.2	86.78	2.05 ± 0.07	1.01 ± 0.08	1.8 ± 0.44
7	2-methylbenzothiophene	0.5	0.32	1.78 ± 0.05	1.75 ± 0.05	1.81 ± 0.03
8	2-hexylthiophene	0.49	5.53	2.26 ± 0.05	0.93 ± 0.17	2.21 ± 0.27
9	tetrahydrothiophene	0.89	2.49	N/A	3.33 ± 0.03	3.86 ± 0.16
10	2-propylthiophene	0.27	37.22	1.94 ± 0.03	1 ± 0.04	2.49 ± 0.10
11	benzothiophene	1.1	0.28	2.23 ± 0.07	2.18 ± 0.02	2.19 ± 0.03
12	1,4-thioxane	1.13	3.87	N/A	2.59 ± 0.05	2.07 ± 0.04
13	3-methylbenzothiophene	1.78	0.14	1.6 ± 0.03	1.6 ± 0.07	1.62 ± 0.03
14	3-acetyl-2,5-dimethylthiophene	0.67	1.97	N/A	2.1 ± 0.05	2.39 ± 0.13

The spectra obtained by CCE-MSP may then be used as constraints for MCR-ALS to model the analyte more accurately allowing for quantitative analysis. Table 3.3 gives the quantification results for each analyte using the S-ratio method via Eq. (3.10) and MCR-ALS both unconstrained and constrained using the CCE-MSP spectra as inputs. The S-ratio provides the analyte concentration ratio and was calculated using the signal for all pure analyte *m/z* defined as having a p-value $\leq 10^{-6}$ and a $LOF \leq 20\%$. In the S-ratio calculation, the signals were

weighted by the inverse of the p-value, because *m/z* purity for the analyte is expected to increase as the p-value decreases. For 2-butyl-5-ethylthiophene in Fig. 3.5C, there are two pure analyte *m/z* (green dots), and an S-ratio = 2.05 was obtained. Overall, other than the 3 analytes (1,4-thioxane, tetrahydrothiophene, 3-acetyl-2,5-dimethylthiophene) that did not have pure analyte *m/z* (green dots) and considering that two analytes were natively present (3-methylbenzothiophene, 2-methylbenzothiophene), the S-ratios in Table 3.3 are in good agreement with the expected concentration ratio of 2 for the 30 ppm versus 15 ppm analyte spike levels. In contrast, before constraints are applied, MCR-ALS quantification is poor or fails approximately when $s_{Int}/s_A > 1$ and/or when 2D resolution < 0.5, consistent with the MV results in Fig. 3.6. Using the analyte spectra obtained by CCE-MSP as input constraints to MCR-ALC allowed for concentration profiles to be obtained from which concentration ratios were calculated which agreed well with those calculated by the S-ratio method.

Furthermore, when two or more analyte hits are in overlapping hit locations, CCE-MSP can also be readily implemented to obtain the purified spectrum for each of the analyte hits involved. Application of CCE-MSP to two overlapping hit locations produces two spectra that are linear combinations of the two analytes that are changing in concentration between the two sample classes. Since all background interference has been removed, these two spectra can be processed with ALS to obtain a purified spectrum for each changing analyte. To demonstrate this, as the spiked fuel data did not produce this situation, we simulated GC×GC-TOFMS data such that the 2D chromatographic space was highly overlapped for five components (2 analytes changing between classes and 3 interference components) as explained in the Supplementary material and summarized in Table B2. Figure 3.7A shows the location of the 5 components in the 2D space

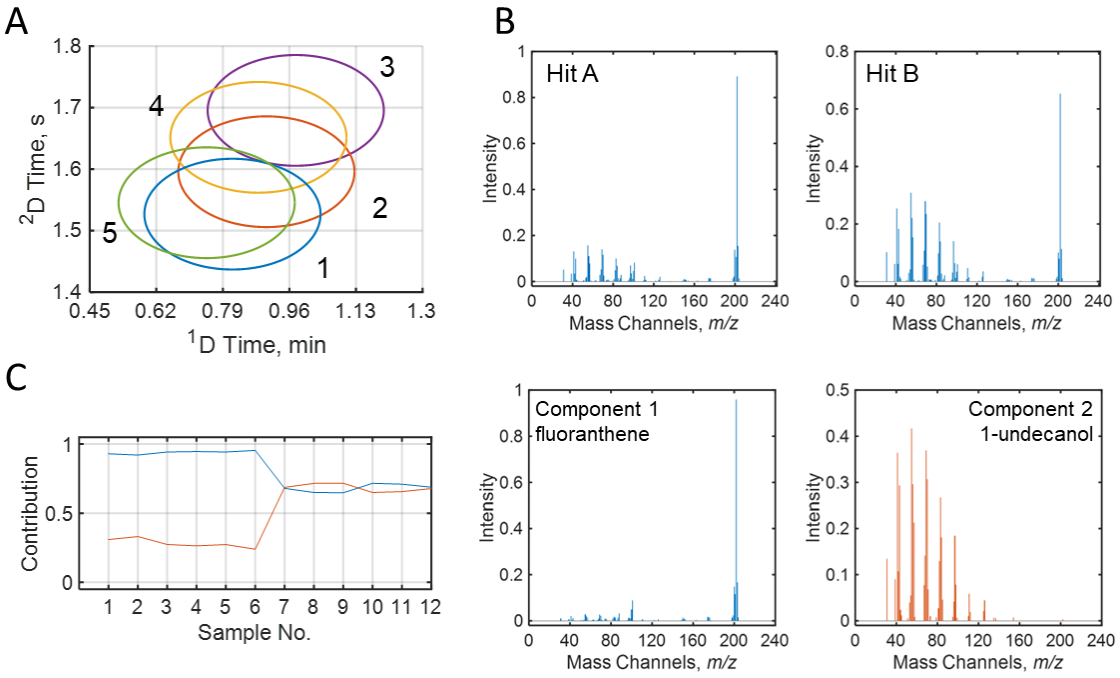


Figure 3.7. (A) Simulated GC×GC-TOFMS chromatogram showing the locations of the 5 components whereby the ellipse signifies TIC outline of each peak width-at-base (4 standard deviations): (1) fluoranthene, (2) 1-undecanol, (3) 1-iodohexane, (4) 1-bromohexane, and (5) pentane. (B) The CCE-MSP obtained spectra obtained from processing the chromatogram, labeled Hits A and B. Each spectrum contains different proportions of the two analyte spectra as they were obtained from nearby (overlapping) hit locations. (C) The results of applying ALS using the CCE-MSP spectra from (B) yields two outputs: the signal contributions for the two components in each of the CCE-MSP spectra and the purified spectrum for the two components.

for 1 of the 10 chromatograms simulated, where the concentration of components 1 and 2 change 2-fold between classes, producing two analyte hits when F-ratio analysis was applied. Next, application of CCE-MSP produced analyte "combination" spectra for these two hits, from the two hit locations that are scaled by the different relative amounts of each analyte component present (Fig. 3.7B). Note that the spectra in Fig. 3.7B only contain the analyte contributions as the background contributions have been removed by the CCE-MSP method. These spectra are linear combinations of the two analytes that are readily separated using an ALS algorithm, obtaining both the contribution (relative concentrations) of the six replicates per sample class and the purified spectrum for each analyte (Fig 3.7C). This approach should be readily extended to situations with more than 2 analyte hits in overlapping hit locations, which is likely to be rare.

3.5. Conclusion

Class comparison enabled mass spectrum purification (CCE-MSP) extracts an analyte mass spectrum relatively free of background interferences for analytes discovered in class-based experiments such as tile-based F-ratio analysis. Following supervised tile-based F-ratio analysis of JP-8 jet fuel spiked at the 30 ppm and 15 ppm level for 14 sulfur containing compounds, the statistical metrics of p-value and *LOF* were calculated for each *m/z* for each analyte. Applying thresholds on these statistics to select the most consistent *m/z* purified spectra were readily obtained yielding a high MV for each analyte. The CCE-MSP spectra quality was generally superior to MCR-ALS, PARAFAC, and PARFAC2 decomposition spectra when $s_{Int}/s_A > 1$, i.e., when the interference signal outweighed the analyte signal, and/or when 2D resolution < 0.5. The CCE-MSP method provides an effective way to extract purified spectra for severely interfered hits without *a priori* knowledge of analyte identity.

3.6. References

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

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

[3] R. van der Westhuizen, M. Ajam, P. De Coning, J. Beens, A. de Villiers, P. Sandra, Comprehensive two-dimensional gas chromatography for the analysis of synthetic and crude-derived jet fuels, J. Chromatogr. A 1218 (2011) 4478–4486. https://doi.org/10.1016/j.chroma.2011.05.009.

[4] T. Dymerski, J. Namieśnik, K. Vearasilp, P. Arancibia-Avila, F. Toledo, M. Weisz, E. Katrich, S. Gorinstein, Comprehensive two-dimensional gas chromatography and three-dimensional fluorometry for detection of volatile and bioactive substances in some berries, Talanta 134 (2015) 460–467. https://doi.org/10.1016/j.talanta.2014.11.061.

[5] L.W. Hantao, H.G. Aleme, M.M. Passador, E.L. Furtado, F.A. de L. Ribeiro, R.J. Poppi, F. Augusto, Determination of disease biomarkers in Eucalyptus by comprehensive two-dimensional gas chromatography and multivariate data analysis, J. Chromatogr. A 1279 (2013) 86–91. https://doi.org/10.1016/j.chroma.2013.01.013.

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

[7] M.K. Das, S.C. Bishwal, A. Das, D. Dabral, A. Varshney, V.K. Badireddy, R. Nanda, Investigation of gender-specific exhaled breath volatome in humans by GCxGC-TOF-MS, Anal. Chem. 86 (2014) 1229–1237. https://doi.org/10.1021/ac403541a.

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

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

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

[11] 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. https://doi.org/10.1016/S0169-7439(01)00198-8.

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

[13] M.K. Jennerwein, M. Eschner, T. Gröger, T. Wilharm, R. Zimmermann, Complete group-type quantification of petroleum middle distillates based on comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC-TOFMS) and visual basic scripting, Energy and Fuels 28 (2014) 5670–5681. https://doi.org/10.1021/ef501247h.

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

[15] 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. https://doi.org/10.1016/j.jchromb.2019.01.029.

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

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

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

[19] 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. https://doi.org/10.1016/j.chroma.2017.10.013.

[20] S.E. Reichenbach, C.A. Zini, K.P. Nicolli, J.E. Welke, C. Cordero, Q. Tao, Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern

recognition, J. Chromatogr. A 1595 (2019) 158–167. https://doi.org/10.1016/j.chroma.2019.02.027.

[21] J.M. Davis, Statistical Theory of Spot Overlap in Two-Dimensional Separations, Anal. Chem. 63 (1991) 2141–2152. https://doi.org/10.1021/ac00019a014.

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

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

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

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

[26] ChromaTOF® Tile Analytical Software, (n.d.). https://www.leco.com/product/chromatof-tile (accessed May 6, 2021).

[27] ChromCompare+, (n.d.). https://www.sepsolve.com/chromcompare/ (accessed May 6, 2021).

[28] 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. https://doi.org/10.1016/j.chroma.2021.462092.

[29] 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. https://doi.org/10.1016/j.chroma.2020.461401.

[30] I.H.M. van Stokkum, K.M. Mullen, V. V. Mihaleva, Global analysis of multiple gas chromatography-mass spectrometry (GC/MS) data sets: A method for resolution of co-eluting components with comparison to MCR-ALS, Chemom. Intell. Lab. Syst. 95 (2009) 150–163. https://doi.org/10.1016/j.chemolab.2008.10.004.

[31] M.A. Sharaf, B.R. Kowalski, Extraction of Individual Mass Spectra from Gas Chromatography-Mass Spectrometry Data of Unseparated Mixtures, Anal. Chem. 53 (1981) 518–522. https://doi.org/10.1021/ac00226a031.

[32] X. Domingo-Almenara, A. Perera, N. Ramírez, N. Cañellas, X. Correig, J. Brezmes, Compound identification in gas chromatography/mass spectrometry-based metabolomics by blind source separation, J. Chromatogr. A 1409 (2015) 226–233. https://doi.org/10.1016/j.chroma.2015.07.044.

[33] M.D. King, G.S. King, A Numerical Method for Extracting Mass Spectra from Gas Chromatography/Mass Spectrometry Data Arrays, Anal. Chem. 57 (1985) 1049–1056. https://doi.org/10.1021/ac00283a020.

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

[35] 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. https://doi.org/10.1002/(SICI)1099-128X(199905/08)13:3/4<295::AID-CEM547>3.0.CO;2-Y.

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

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

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

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

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

[41] B.L. Atwater, R. Venkataraghavan, F.W. McLafferty, Matching of Mixture Mass Spectra by Subtraction of Reference Spectra, Anal. Chem. 51 (1979) 1945–1949. https://pubs.acs.org/doi/pdf/10.1021/ac50048a013 (accessed August 13, 2020).

[42] Z.Z. Kapadia, D. V Spracklen, S.R. Arnold, D.J. Borman, G.W. Mann, K.J. Pringle, S.A. Monks, C.L. Reddington, F. Benduhn, A. Rap, C.E. Scott, E.W. Butt, M. Yoshioka, Impacts of aviation fuel sulfur content on climate and human health, Atmos. Chem. Phys. 16 (2016) 10521–10541. https://doi.org/10.5194/acp-16-10521-2016.

[43] 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. https://doi.org/10.1021/ac971184b.

[44] R.B. Pellegrino Vidal, A.C. Olivieri, R. Tauler, Quantifying the Prediction Error in Analytical Multivariate Curve Resolution Studies of Multicomponent Systems, Anal. Chem. 90 (2018) 7040–7047. https://doi.org/10.1021/acs.analchem.8b01431.

[45] F.W. Plankey, T.H. Glenn, L.P. Hart, J.D. Winefordner, Hadamard Spectrometer for Ultraviolet-Visible Spectrometry, Anal. Chem. 46 (1974) 1000–1005.

[46] A.P. Thorne, Fourier Transform Spectrometry in the Ultraviolet, Anal. Chem. 63 (1991) 57 A-65 A.

[47] S.A. Dyer, Hadamard transform spectrometry, Chemom. Intell. Lab. Syst. 12 (1991) 101–115.

[48] M.R. Webb, C.N. LaFratta, D.R. Walt, Chromatically Resolved Optical Microscope (CROMoscope): A Grating-Based Instrument for Spectral Imaging, Anal. Chem. 81 (2009) 7309–7313.

Chapter 4. Using solid phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: Comparative analysis of aerospace fuel composition

This chapter was reproduced from Grant S. Ochoa, Matthew C. Billingsley, Robert E. Synovec, "Using solid phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: Comparative analysis of aerospace fuel composition" *Anal. Bioanal. Chem.* (2022) 1-13.

4.1. Introduction

To ensure consistent and optimal performance in aerospace applications, the chemical composition of kerosene-based rocket propellants is either directly controlled by maximum allowable limits for certain heteroatomic and unsaturated species, or indirectly controlled by limits on physical and thermochemical properties. Small variations in chemical composition can have a demonstrable effect on fuel physical and thermal behavior in engine systems [1–7]. For example, sulfur-, nitrogen-, and oxygen-containing compounds contribute to fuel reactivity at elevated temperatures and surface fouling such as corrosion and carbonaceous deposit formation in regenerative cooling systems [8–12]. Fuel specifications are in place to ensure that procured fuels meet required property and performance metrics before use [13]. However, analytical-computational methods capable of resolving chemical differences (variations in compound identity and/or concentration between fuels) offer a rapid, cost-effective solution for digital fuel design and development.

Gas chromatography coupled with mass spectrometry (GC-MS) is an effective technique for fuel analysis [14–16]. Information gained from the mass spectrum dimension not only benefits compound identification but also improves the quality of chemometric data analysis [17–20]. Comprehensive two-dimensional (2D) gas chromatography with time-of-flight mass

spectrometry (GC×GC-TOFMS) offers significantly improved separation power that benefits the analysis of highly complex samples such as kerosene-based fuels. When equipped with a reverse column configuration consisting of a polar first dimension (^{1}D) column coupled to a non-polar second dimension (^{2}D) column, GC×GC achieves near complete compound class separation for kerosene-based fuels [15, 20–22]. Thus, the wealth of information afforded by GC×GC-TOFMS can be utilized to help discriminate the differences between fuels through comparative analysis.

Relevant specification property measurements and fit-for-purpose performance testing reveal differences between fuels but do not provide insight into the underlying cause of observed physical behaviors. This is precisely what a chemical composition-based comparative analysis aims to accomplish. The supervised discovery method referred to as tile-based Fisher ratio (F-ratio) analysis is well suited for this task. F-ratio analysis utilizes column separation (temporal) and mass channel (spectrum) dimensions to find sample-class-distinguishing features among the fuels being compared [17, 21–29]. This method calculates the F-ratio, the ratio of between-class variance to within-class variance, for regions of the 2D chromatogram at each mass channel (*m/z*). Thus, the magnitude of the F-ratio reflects the degree of chemical difference present for a given sample feature [23], facilitating a complete interrogation of the chemical compositional differences across the sample classes. Each neat (untreated) fuel may constitute a sample class. However, new sample classes may be created by treating the fuel to alter specific compounds (i.e., compound types within a fuel). Solid phase extraction (SPE) is commonly utilized for analyzing polar constituents and other compound types in jet and rocket fuel [30–36]. Application of SPE with the appropriate stationary phase creates a fuel deficient in the selected target compounds, which may then be treated as a new sample class; in turn, F-ratio analysis readily locates compounds affected by SPE treatment. Leveraging SPE facilitates comparison of

compositionally complex fuels, allowing deviations in fuel chemistry – whether low-level contaminants or bulk hydrocarbon types – to be quickly identified and correlated with observed physical property or thermal performance behavior.

Herein, we demonstrate application of tile-based F-ratio analysis in a methodology that utilizes SPE to isolate and investigate key compound types. In the first study, a proof-of-concept, an RP-1 (rocket-grade kerosene) fuel with exhibited poor thermal performance is examined [20]. For this study, three samples are analyzed by GC×GC-TOFMS: (1) the untreated (neat) fuel; (2) the SPE-treated (pass) fuel, which is collected during extraction of polar compounds; and (3) the extract eluted from the SPE column stationary phase with methanol. The neat and pass fuels are examined using tile-based F-ratio analysis, resulting in a hit list comprising compounds extracted by SPE and retained on the cartridge stationary phase as the pass sample is collected. This F-ratio comparison reveals the polar contaminant profile, which is subsequently validated by the GC×GC-TOFMS data acquired from the extract sample. While direct analysis of the SPE extract is possible and is commonly practiced, one objective of this study is establishing a generalized framework for using tile-based F-ratio analysis in the discovery of low-level chemical compositional differences among two or more samples.

The framework described above is then applied in a case study involving three similar but compositionally distinct formulations of a multicomponent hydrocarbon product being evaluated as a rocket fuel. These formulations were produced by the same process with one exception: measures were taken to reduce detrimental compound types in one fuel. By applying SPE coupled with F-ratio analysis as described above, the neat and pass samples of these fuels are examined to determine the extent and nature of compositional differences in their bulk and contaminant profiles. This practical demonstration of the analytical methodology introduced and

discussed herein reveals critical compositional differences that are not evident at first glance in the GC×GC-TOFMS data.

One shortcoming of tile-based F-ratio analysis in rapidly comparing sample classes is that F-ratio calculation of hit list ranking necessitates collection of multiple GC×GC-TOFMS separation replicates for each sample, i.e., often six replicates per sample class [21,22,24]. Data collection to apply tile-based F-ratio analysis can therefore be time consuming and prohibitive for some applications. To address this issue, we also implement a tile-based pairwise analysis method, referred to as 1v1 analysis [37]. This method discovers analytes that are different in concentration between two chromatograms while utilizing the same four-grid tiling scheme as developed for tile-based F-ratio analysis [22–24,28,38,39]. However, instead of calculating F-ratios, 1v1 analysis utilizes a rank metric based upon a sum-normalized absolute difference between the signals in the two chromatograms being compared. Implementation of the tiling algorithm mitigates run-to-run retention shifting that would otherwise imperil simply taking the difference between two chromatograms on the pixel-level data [38]. We envisage widespread implementation of tile-based 1v1 analysis, thereby broadening the application and usefulness of comparative analyses employing GC×GC-TOFMS data. Thus, a primary aim of the studies presented is to develop an analytical methodology that provides a comprehensive list of chemical differences for any fuel pair quickly and efficiently to improve fuel performance and compliance.

4.2. Experimental

Samples

Four rocket fuels were supplied and analyzed in the studies communicated here: RP-1 (Batch B0112868) was the subject of the first study, and three highly paraffinic formulations referred to as RF-A, RF-B, and RF-C were used for the second study (Edwards AFB, CA). Each

fuel was subjected to the following experimental procedures followed by tile-base d F-ratio analysis [21,22,24,28,38,39].

Solid phase extraction

SPE cartridges were used to isolate polar compounds from each fuel sample. A nominally ~1 mL silica phase SPE cartridge (SepPak, Waters) was rinsed with 5 mL methanol to clean the cartridge. Next, the cartridge was conditioned with 15 mL of hexanes (Sigma-Aldrich, St. Louis, MO), followed by loading the cartridge with 8 mL of sample at a flow rate of 1-3 mL/min. The first mL of sample to pass through the cartridge was collected and set aside for analysis as the "pass" sample. The cartridge was then rinsed with 15 mL of hexanes to remove any non-polar compounds left on the cartridge. Finally, the pre-concentrated polar compounds were eluted with 0.5 mL of methanol. This extract was then analyzed alongside the neat and pass fuel by GC×GC-TOFMS. The process is summarized in Fig. C1 in the Supplementary Material.

GC×GC-TOFMS procedure

GC×GC-TOFMS data were collected for the neat fuel, pass fuel, and fuel extract samples as previously reported, with the details in Supplementary Material [29,40]. The GC×GC-TOFMS chromatograms were imported into MATLAB (version 2019a; The MathWorks, Inc., Natick, MA) from the LECO ChromaTOF for BT software v5.20 (LECO, St. Joseph, MI). All chromatograms were baseline corrected prior to one-point normalization using the sum of the total ion current (TIC) relative to the average TIC, which was applied on a comparison basis. In short, samples belonging to the same comparison were normalized together. Tile-based F-ratio analysis was performed for each neat versus pass fuel comparison to discover analytes affected by the SPE procedure. For all F-ratio analyses, six-versus-six (6v6) comparisons were used, with six replicates of each sample per class. The following conditions were applied to all F-ratio

analyses. The tile dimensions were 12 s on ^{1}D by 300 ms on ^{2}D with a signal-to-noise ratio (*S/N*) threshold of 10. Two samples must pass the *S/N* threshold to be included in the final hit list. The mass channel (*m/z*) providing the maximum F-ratio was used to rank the hit list [29].

Tile-based 1v1 analyses were applied for the direct comparison of two chromatograms, utilizing the same tiling scheme and pinning and clustering algorithms that form the basis of tile-based F-ratio analysis. However, instead of calculating F-ratios, the sum-normalized absolute difference rank metric (*RM*) is calculated for each tile-based 1v1 analysis [37],

$$RM = \frac{|s(m/z)_2 - s(m/z)_1|}{s(m/z)_2 + s(m/z)_1} \times 100 \quad (4.1)$$

where $s(m/z)_1$ and $s(m/z)_2$ are the summed signals at a given *m/z* for a given tile in one sample (class 1) versus another sample (class 2). This dimensionless metric ranges from 0% to 100% and allows the ranking of significant differences high on a hit list regardless of the direction of change. The tile size parameters for tile-based 1v1 analysis were the same as for tile-based F-ratio analysis. Similarly, the maximum *RM* at a specific *m/z* was used to rank the hit list. As with F-ratio analysis, with 1v1 analysis if the *S/N* threshold is raised then the number of false positives is reduced at the expense of reducing the true positives as well. Indeed, the performance of the F-ratio and 1v1 analyses were previously demonstrated to produce nearly identical results using ROC curves [37].

Chromatographic visualization of the hits from the F-ratio and 1v1 analyses is achieved by piecing together data from each of the hit locations into a single ‘stitch’ chromatogram via the following algorithmic process, based on the principles for the enhanced total ion current chromatogram algorithm [41]. For each hit selected from the top of a given hit list, a tile of data with the same dimensions as above is cut around its pin location from the *m/z* with the top F-ratio or top *RM* (for 1v1 analysis). The algorithm then unfolds the tile and locates all the ^{2}D peak

maxima present in the tile above a *S/N* threshold of 10, keeping one peak per modulation that is detected. Thus, the maximum number of modulations kept per sample is the ^{1}D tile size plus one. A window of data is then cut around each ^{2}D peak based upon the average $^{2}w_{b}$ from the relevant chromatogram, in this case that ^{2}D window corresponded to 180 ms, i.e., 18 data points (spectra) wide with the mass spectrum collection frequency of 100 Hz. These windows are then pasted into an array of zeros the same size of the 2D chromatogram from the same coordinates from which they were cut. This process is repeated for each hit, creating a stitch chromatogram highlighting the class distinguishing features from the rest of the data. If a peak is not found there will be no data cut from the tile for that hit. To minimize scale issues, each hit may be auto scaled making each hit the same height for broad class visualization. One can imagine the finished stitch chromatogram to be a "quilt" of peaks obtained from the top of a given hit list, constructed from the best *m/z* for each hit obtained from the sample-class comparison software.

4.3. Results and Discussion

Low-level polar compounds analysis in the RP-1 fuel: F-ratio of neat vs pass samples

The GC×GC-TOFMS total ion current (TIC) chromatograms of the RP-1 fuel (B0112868) are presented in Fig. 4.1, with the neat fuel sample shown in Fig. 4.1a, the pass fuel collected during the extraction of polar compounds from the fuel using the silica gel SPE cartridge shown in Fig. 4.1b, and the methanol extract obtained from the SPE cartridge shown in Fig. 4.1c. At first glance, there are no discernable differences between the neat and pass fuel samples since any affected polar compounds are present at low concentration in the predominantly non-polar fuel matrix. However, the extract shows several peaks eluting in the 2D section between 15 min to 25 min on ^{1}D, and 0 s to 1 s on ^{2}D. Since the extracted-then-eluted compounds were removed from the neat fuel, they are significantly reduced or undetectable in

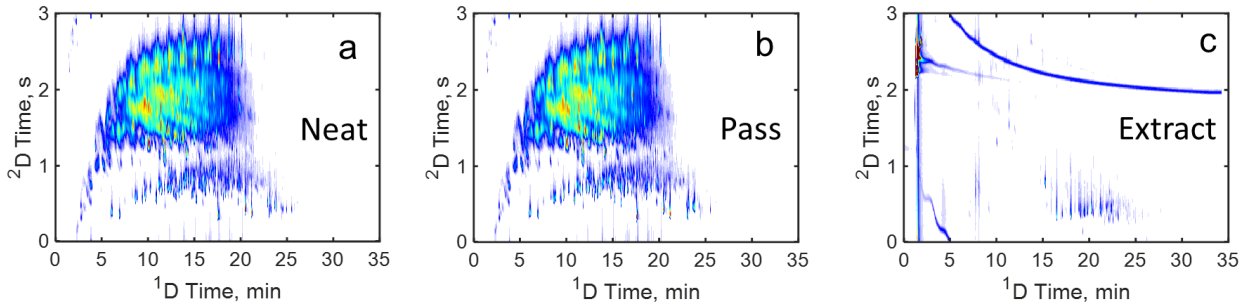


Figure 4.1. (a) GC×GC-TOFMS TIC chromatograms for neat RP-1 fuel (batch B0112868), (b) the post silica-SPE extracted pass RP-1, and (c) the extract eluted from the silica-SPE cartridge. There are initially no visible differences between the neat and pass fuels in (a) versus (b). However, the extract in (c) shows a collection of peaks between 15-25 min on ^{1}D and around 0.5 s on ^{2}D, indicating that the compounds producing these peaks must be present in the neat fuel and therefore significantly reduced or undetectable in the pass fuel, albeit at too low of a concentration to be visible in (a) versus (b).

the pass fuel. Thus, comparison of the neat and pass fuel via tile-based F-ratio analysis should reveal focused information on the compounds affected by the extraction.

Tile-based F-ratio analysis was performed on a 6v6 sample comparison of the neat and pass samples of RP-1, yielding a hit list with the location of significant differences between the classes. The top fifty hits were inspected, followed by false positive and redundant hit removal by visual inspection of chromatographic profiles and retention time association respectively, leaving thirty-nine true positive hits that are plotted as an overlay on the GC×GC-TOFMS TIC chromatogram of the neat fuel in Fig. 4.2a. The true positives of the hit list correspond to compounds extracted from the neat fuel and indicate areas of the chromatogram to investigate for these compounds. The majority of the thirty-nine hits lie in a band underneath the aromatic compounds (15-25 min on ^{1}D and 0-1 s on ^{2}D), which for the reverse column GC×GC configuration implies the hit compounds are greater in polarity than aromatic hydrocarbons. Figure 4.2b magnifies a region of the chromatogram encapsulating thirty-two of the thirty-nine plotted hits, revealing that most of the hits have too small a signal to appear in the TIC chromatogram of the neat fuel. To remedy this situation, the stitch chromatogram tool enables

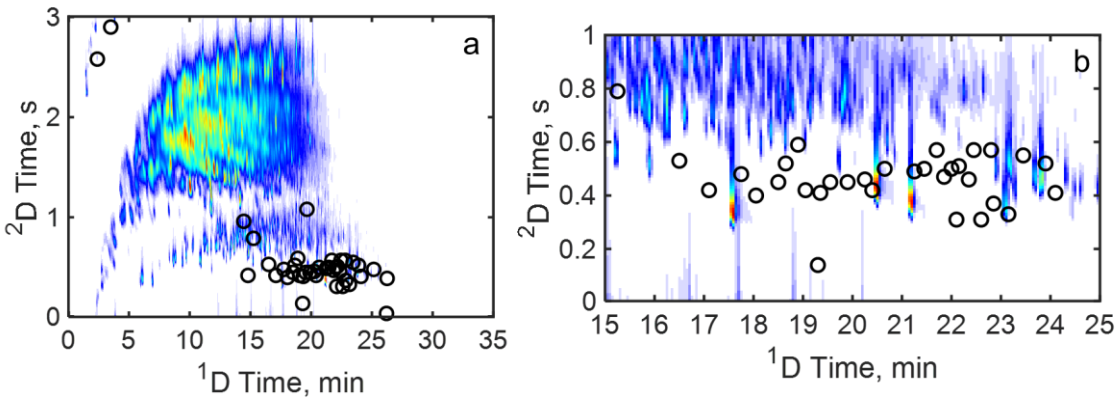


Figure 4.2. (a) GC×GC-TOFMS TIC chromatogram of RP-1 (batch B0112868) with circles indicating the location of the top 39 true positive F-ratio hits (from top 50 hits in the hit list). These hits mainly cluster along a band between 15-25 min on ^{1}D and 0.5 s on ^{2}D, underneath the aromatics band. These hit locations are in the same 2D separation region in which peaks were found in the extract chromatogram in Fig. 4.1c. (b) A zoom in of this region (15-25 min on ^{1}D and 0-1 s on ^{2}D) where the majority of the 39 hits from the F-ratio analysis are observed.

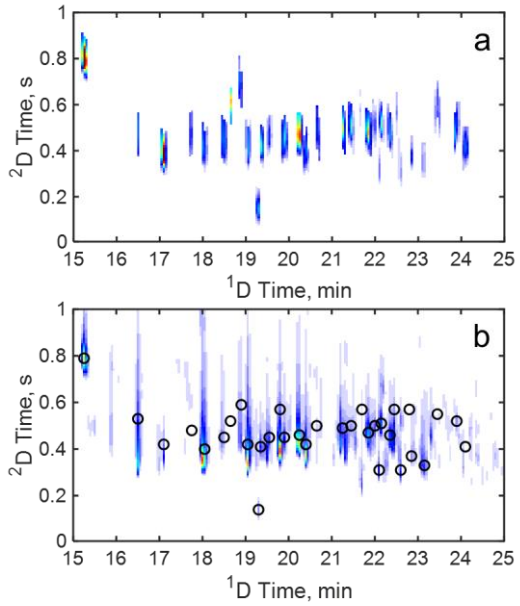


Figure 4.3. (a) Stitch chromatogram constructed using the top F-ratio hits pulling data from pure m/z for each hit from the neat RP-1 (batch B0112868) chromatogram. (b) The top F-ratio hit markers, which were obtained without information of the extract, have been included over the extract chromatogram, demonstrating good agreement with the extract peaks. Visually, the stitch chromatogram matches well to the extract chromatogram from Fig. 4.1c.

visualization of the hits from neat-versus-pass fuel analysis in chromatographic space, bringing together data from selective *m/z* for each hit into a single chromatogram absent of superfluous compounds that were unaffected by the SPE treatment.

The stitch chromatogram resulting from F-ratio analysis for the RP-1 fuel (B0112868) is provided in Fig. 4.3a. Several peaks appear in the same 2D separation region that was inspected in Fig. 4.2b. Most of the peaks have nearly the same ^{2}D retention time, indicating a similarity in polarity resulting from their chemical structure. To assess the accuracy of this visualization tool, we compare the stitch chromatogram against the TIC of the methanol SPE-extract chromatogram for the same region in Fig. 4.3b. There is remarkable likeness between both chromatographic regions (2D section between 15 min to 25 min on ^{1}D, and 0 s to 1 s on ^{2}D). At each location that a peak appears in the stitch chromatogram, a corresponding peak is found in the extract chromatogram, albeit not with the same intensities due to SPE preconcentration and stitch-chromatogram-algorithm effects. This demonstrates the effectiveness of utilizing the information obtained from tile-based F-ratio analysis to visualize compound types affected by the SPE procedure. Similarly, the F-ratio hit markers in Fig. 4.3b correspond to peaks from the extract sample, covering the majority of the features from the extract chromatogram. Table 4.1 shows the tentative mass spectrum identifications using a NIST library along with both their match value (MV) and reverse match value (RMV) determined for the top ten hits of the F-ratio comparison, suggesting these compounds are oxygenates with the majority phenolic in character. These compounds are likely primary contributors to the observed poor thermal performance for this RP-1 fuel [11].

Table 4.1. Top 10 hits from the neat versus pass F-ratio comparison of RP-1 fuel (batch B0112868) providing compound identification as well as contrasting the performance of the tile-based F-ratio and 1v1 analysis methods.

F-ratio Hit No.	Name	^{1}D Time, min	^{2}D Time, s	MV	RMV	Top *m/z*	F-ratio	1v1 Hit No.	Rank Metric
1	1-(2-isopropyl-5-methylcyclopentyl)ethanone	15.2	0.79	803	816	110	4638	6	78.45
2	m-isopropylphenol	18.0	0.40	870	956	136	1526	1	92.77
3	2,5-diethylphenol	19.3	0.41	818	891	135	1381	9	68.01
4	carvacrol	19.85	0.45	860	878	135	1324	3	89.30
5	2-isopropylphenyl ester formic acid	22.4	0.57	704	830	107	1314	13	53.90
6	thymol	20.35	0.42	863	907	150	1221	19	40.50
7	carvacrol isomer	19.5	0.45	853	871	150	1088	8	72.45
8	4-secbutylphenol	20.2	0.46	886	910	121	1060	5	86.28
9	1-methoxy-4-sec-butylbenzene	21.8	0.47	809	841	135	1032	12	56.37
10	3-methyl-6-propylphenol	21.2	0.49	747	867	121	1021	10	66.60

Application of 1v1 analysis to the neat vs pass RP-1 fuel samples

The preceding F-ratio comparison provided in-depth and illuminating information regarding subtle changes between the neat and pass fuel samples of the RP-1 formulation. However, a shortcoming of the tile-based F-ratio analysis methodology is the need to collect replicates. Collecting several (i.e., six) replicates for tile-based F-ratio analysis is often prohibitive, especially when working with a large number of samples. Thus, implementation of tile-based 1v1 analysis may be more appealing. Accordingly, we demonstrate application of tile-based 1v1 analysis on the same set of six replicates each of the neat and pass RP-1 fuel samples, arbitrarily arranged to make six neat versus pass RP-1 fuel sample pairs. The *RM* calculated via Eq. (4.1) for 1v1 analysis yielded high values for the features whose signal differed greatly between the two sample classes, even if that difference was due to natural variance. Thus, interspersed with the true positives are some false positives. These were identified using a follow-up peak detection step that removed hits that did not have a peak present within the given hit tile. The top fifty hits are shown in Fig. 4.4 for the six stitch chromatograms resulting from

the 1v1 analyses for the analyte region of interest (the entire chromatographic region is provided in Fig. C2). The 1v1 analysis hit lists from each pair of chromatograms vary slightly due to the nature of the rank metric employed but are generally very similar to the hit list from tile-based F-ratio analysis (Fig. 4.3b). The fidelity of each hit list compared to the F-ratio hit list was quantified as the proportion of hits represented by the 1v1 hit lists, resulting in values ranging from 69 - 81%. Of the thirty-two hits from the F-ratio analysis, thirty-one were present in at least one of the 1v1 analyses. Twenty-seven were present in at least four analyses, with most of these belonging to peaks along the 0.4 s ^{2}D retention band. Four of the remaining hits were present in three or fewer analyses with one hit absent entirely. The less-frequently discovered hits by 1v1 analysis were generally of much lower signal compared to the other hits, thus, their *RM* was not significant enough to rank higher than other similarly ranked false positive hits within the top fifty hits. Despite the modest variability of the 1v1 analyses, these results demonstrate that 1v1 analysis provides sufficient information to capture the major differences between sample classes (for SPE-treated multicomponent distillate fuels).

In the previous proof-of-concept study for the RP-1 fuel, the tile-based 1v1 analysis method was shown to be useful for the focused comparison of analyte classes resulting from SPE treatment of a *single* fuel formulation. However, comparing the chemical composition between *different* fuels is desirable, for example, when evaluating against a reference (baseline) fuel with known performance, or when characterizing the effects of changes in production (e.g., feedstock source and refinery conditions) and post-production processes (e.g., chemical treatment, filtration, etc.). Comparative analysis can also be performed to determine the compositional differences between samples for both the “bulk” hydrocarbons and “low-level” polar species.

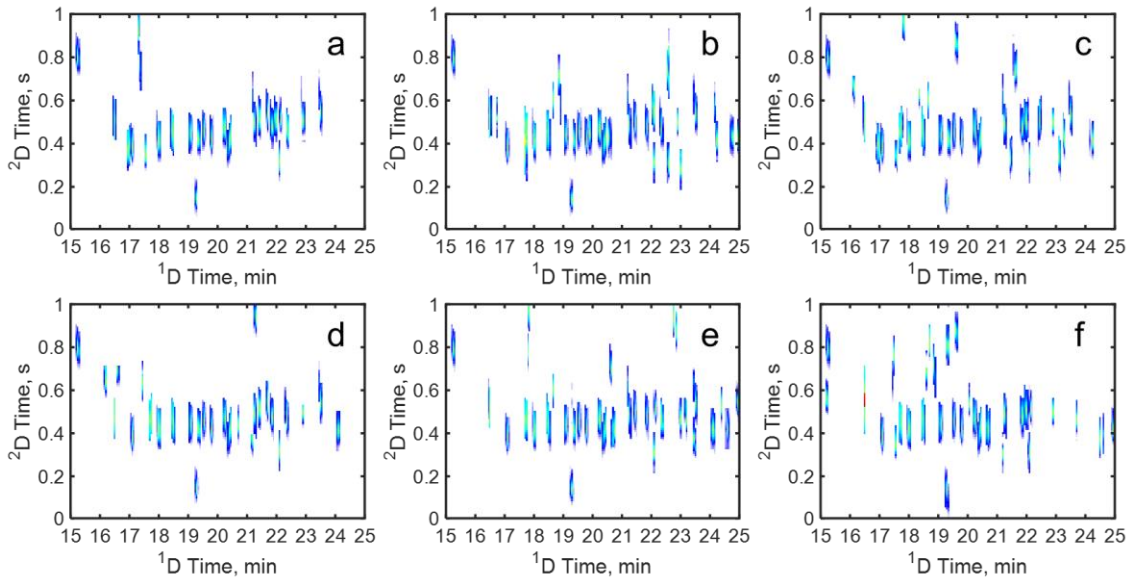


Figure 4.4. Stitch chromatograms of the top fifty hits for the tile-based 1v1 analyses for each of the six chromatogram neat versus pass fuel pairs displaying the region of interest (15-25 min on ^{1}D and 0-1 s on ^{2}D). Even though there is some experimental variation from one 1v1 analysis to the next, each of the sample pairs captured a majority of the top 39 true positive hits in the F-ratio hit list with the following capture percentages of these true positives: (a) Pair 1 at 68.8%, (b) pair 2 at 81.3%, (c) pair 3 at 78.1%, (d) pair 4 at 75.0%, (e) pair 5 at 75.0%, and (f) pair 6 at 68.8%.

While bulk hydrocarbons typically account for >99% of fuel mass (e.g., *n*-alkanes and aromatics), low concentration level polar species include oxygenates and other heteroatom containing compounds. F-ratio and/or 1v1 analysis of neat fuels is primarily useful for comparing bulk hydrocarbon differences along with the low-level polar species, eg., oxygenates and other heteroatom containing compound differences between samples. In contrast, analysis of the pass fuel samples resulting from using the SPE step will yield information regarding only the bulk hydrocarbon differences since the low-level polar species have been removed from the pass fuel samples. These two strategies were applied next in the comparative analysis of three candidate fuel formulations produced in a gas-to-liquid process: RF-A, RF-B, and RF-C.

InterComparative analyses for three fuel formulations

Figure 4.5a-c shows the TIC chromatograms for the three fuels: RF-A, RF-B, and RF-C. First-look visual comparison suggests they are multicomponent fuels with similar composition. However, adjusting the color scale to emphasize lower concentration peaks reveals subtle differences between them, namely, compounds present in RF-A and RF-B that are absent from RF-C (Fig. 4.5d-f). The same silica-based SPE procedure applied to RP-1 in the previous proof-of-principle study was applied to each of these, again generating a pass fuel for each neat fuel formulation. Meanwhile, the extract chromatograms (Fig. 4.5g-i) were obtained from application of SPE of the fuels diluted 1:10 with hexane in order to obtain a better extraction of the low-level polar species. The extract chromatograms indicate that RF-A and RF-B contain similar polar compounds between 0 and 1.8 s on ^{2}D, whereas RF-C is a relatively clean fuel in this regard. These extract chromatograms qualitatively highlight the differences in the low-level polar species between these fuels.

To facilitate a deeper examination of the differences in the low-level polar species present in the neat fuels, each of the three pass fuels, depleted in polar compounds due to SPE treatment, was compared to its corresponding neat fuel using tile-based F-ratio analysis. The resulting F-ratio distributions for hit list ranking are provided in Fig. 4.6a for each fuel. In contrast to the other two fuels, RF-C has essentially no F-ratios > 100 (i.e., above the primary distribution of F-ratios), as evidenced by the sharp decrease in the red trace compared with the blue and green traces. When comparing the neat and pass samples, if concentration changes between the sample classes are insignificant, the majority of F-ratios are present in the hit list merely because they generate sufficient signal above the *S/N* threshold to calculate a (low) F-ratio. Hits that are likely indicative of significant concentration differences between the neat

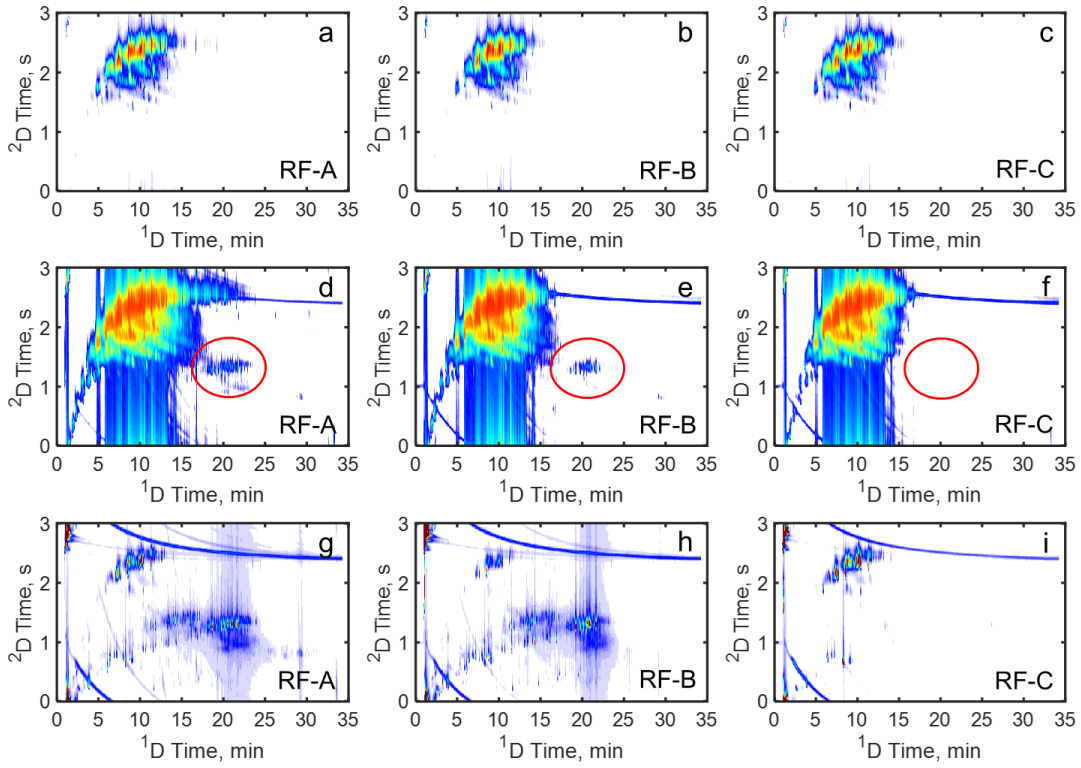


Figure 4.5. GC×GC-TOFMS TIC chromatograms for the three neat fuel formulations (a) RF-A, (b) RF-B, and (c) RF-C, indicating apparent compositional similarity. Use of a different color scale to emphasize low-level signals, compositional differences are revealed for the same GC×GC-TOFMS TIC chromatograms (circled regions in red, for (d) RF-A, (e) RF-B, and (f) RF-C, suggesting one limitation of relying on visual inspection when comparing complex fuels. Chromatograms of the SPE extract of the fuels diluted 1:10 in hexane are provided for (g) RF-A, (h) RF-B, (i) RF-C. Fuels RF-A and RF-B exhibit relatively high concentrations of polar compounds relative to RF-C.

versus pass fuel classes will have F-ratios that extend beyond the cluster of low F-ratios. Thus, the lack of high F-ratio hits for the RF-C distribution indicates a similar composition between the neat and pass fuel samples. The implication is that RF-C is a cleaner fuel compared with the others in terms of polar compounds, as indicated by the RF-C extract chromatograms.

Based upon the distribution of the neat versus pass RF-C analysis, an F-ratio threshold of 100 was selected to inspect all three hit lists. This threshold retains 148 hits for RF-A, 123 hits for RF-B, and 1 hit for RF-C. The locations for these hits are plotted on their respective chromatograms in Fig. 4.6b-d. The pattern of hits is similar for both RF-A and RF-B (Fig.

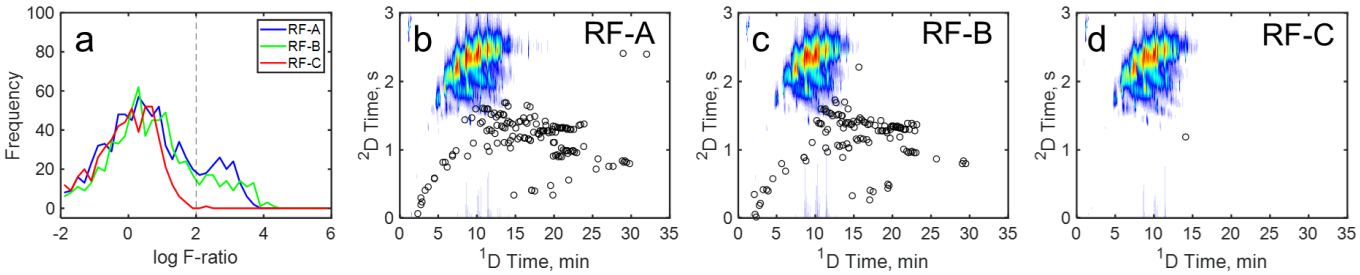


Figure 4.6. (a) Histogram of the intra-fuel neat versus pass F-ratio comparisons alongside GC×GC-TOFMS -TIC chromatograms of each of the three rocket fuels, (b) RF-A, (c) RF-B, and (d) RF-C, with circles indicating the locations of all hits with a F-ratio > 100. Fuels RF-A, RF-B, and RF-C had 148 hits, 123 hits, and one hit with an F-ratio > 100, respectively. The patterns of these hit markers demonstrate the ability of the tile-based F-ratio analysis to discover the minute signal changes affected by the SPE treatment. The hit markers for RF-A and RF-B show very similar patterns indicating a similar profile of polar compounds present in these two fuels.

4.6b,c), suggesting the polar compounds present in these fuels may be a byproduct of the production process of these paraffin-rich fuels. In contrast, RF-C had one extracted hit passing the F-ratio threshold (Fig. 4.6d). Visual inspection of these regions in each chromatogram at low signal levels confirms the lack of signal for RF-C compared to the other fuels (Fig. 4.5d-f). Table 4.2 lists several hit locations spread about the 2D separation space and their tentative identification. Each of these compounds was determined to be oxygenates, including alcohols, aldehydes, ketones, and furanones. Several of the polar compounds eluting after 18 min on ^{1}D could not be tentatively identified due to the similarity in their mass spectra. However, their mass spectra were indicative of long chain alkanes likely with an oxirane or alcohol group. Next, we performed F-ratio analysis on inter-fuel pairs of fuel formulations, investigating both neat versus neat as well as pass vs pass comparisons. Through this we aim to reveal a comprehensive perspective to uncouple the bulk hydrocarbon compositional differences from the low-level polar species present for these three fuel formulations.

Inter-fuel comparative analysis for the three fuel formulations

Next, the bulk hydrocarbon compositional differences between fuel formulations were examined alongside the low-level polar compounds through inter-fuel analysis of the neat versus neat and pass versus pass samples. Performing tile-based F-ratio analysis on the three inter-fuel pairs (RF-A versus RF-B, RF-A versus RF-C, and RF-B versus RF-C) results in F-ratio distributions for the neat versus neat fuel comparisons (Fig. 4.7a) and pass versus pass fuel comparisons (Fig. 4.7b) that are shifted toward higher values than the intra-fuel (neat versus pass fuel) comparisons (Fig. 4.6a). Using the previously selected F-ratio threshold of 100, the comparisons retained 344 (RF-A versus RF-B), 389 (RF-A versus RF-C), and 270 hits (RF-B versus RF-C) for the inter-fuel neat versus neat fuel comparisons (Fig. 4.7a). In these hit lists are a mix of hits due to bulk hydrocarbon and low-level polar species differences sporadically distributed among the hits passing the threshold, initially complicating the analysis. Fortunately, the F-ratio analysis performed on the inter-fuel pass versus pass fuel pairs (Fig. 4.7b), yield hit lists that retain the bulk hydrocarbon compositional differences but are simpler since they lack the low-level polar-compound differences. Applying the F-ratio threshold of 100, these pass versus pass inter-fuel comparisons retained 310 (RF-A versus RF-B), 314 (RF-A versus RF-C), and 200 hits (RF-B versus RF-C). For the sake of brevity, the most compositionally extreme pair (RF-A versus RF-C) will be discussed with the other two fuel pairs available in the Supplementary Material (Fig. C3-4).

F-ratio analysis of fuels RF-A and RF-C, being the most dissimilar fuels per Fig. 4.7a,b, resulted in a multitude of high F-ratio values for the inter-fuel comparison of both neat (untreated, 344 hits) and pass (SPE treated, 310 hits) fuel as indicated by the markers in Fig. 4.8a,b. The hit markers in Fig. 4.8a (neat versus neat fuel comparison) encompass both bulk

Table 4.2. NIST identification of compounds representing the polar compound band present in fuels RF-A and RF-B. Mass spectra of compounds after 20 min on ^{1}D were very similar and did not provide confidence to their identification.

^{1}D Time, min	^{2}D Time, s	Compound Name	MV	RMV
2.75	0.19	2,2-dimethoxybutane	744	754
3.9	0.25	2-hexanone	889	921
6.1	0.54	2-heptanone	885	910
6.4	0.72	3-methyl-4-heptanol	733	737
8.3	0.66	(Z)-2-methyl-2-penten-1-ol	734	743
8.75	0.75	2-octanone	916	928
10.65	1.39	3,7-dimethyl-6-nonenal	727	729
11.3	1.26	4,7-dimethyl-4-octanol	809	894
11.55	0.88	2-nonanone	912	946
11.55	1.26	2-methyl-2-nonanol	810	856
11.65	0.22	4-methl-3-pentenoic acid	878	898
12.95	1.34	2-methanol-2-decanol	730	824
13.2	1.06	2-methyl-2-undecanal	779	840
13.8	1.27	2-decanol	848	932
14.3	1.39	2-methyl-2-decanol	789	845
14.3	0.99	2-decanone	911	925
14.55	0.31	5-ethyldihydro-3-methyl-2(3H)-furanone	897	982
15.8	1.15	dodecyloxirane	738	774
16.05	1.42	3-dodecanol	769	849
16.95	1.13	2-undecanone	842	902
17.05	0.39	trans-3-methyl-4-octanolide (whiskey lactone)	797	874
19.55	0.47	dihydro-5-methyl-5-(2-methylpropyl)-2(3H)-furanone	780	823

hydrocarbon and low-level polar species differences, while the hit markers in Fig. 4.8b (pass versus pass fuel comparison) represent primarily the bulk hydrocarbon differences. By subtracting the set of hits from Fig. 4.8b from the set of hits from Fig. 4.8a, a refined list of hits due to the low-level polar species is obtained. This hit list is visualized in Fig. 4.8c where these hits are plotted on top of the RF-A extract chromatogram (Fig. 4.5g), exhibiting close agreement with the pattern of peaks. Additionally, the bulk hydrocarbon hits from the pass versus pass inter-fuel comparison plotted in Fig. 4.8b are more visible in some portions of the separation

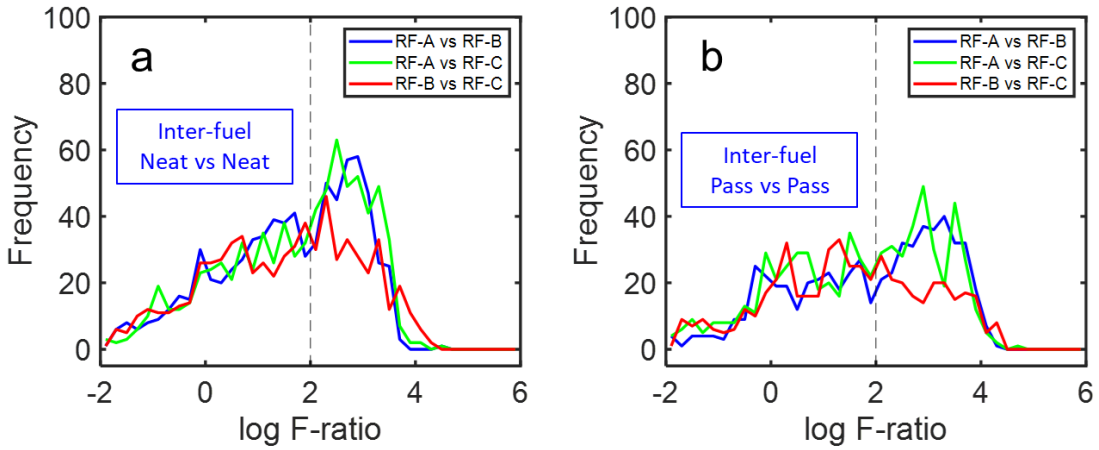


Figure 4.7. F-ratio distributions for the (a) inter-fuel neat versus neat F-ratio comparisons and (b) inter-fuel pass versus pass F-ratio comparisons. Dashed line indicates the applied F-ratio threshold of 100. The inter-fuel comparison distributions are shifted to higher F-ratios compared to the intra-fuel neat versus pass comparisons in Fig. 4.6a.

space when plotted on a more intense color scale in Fig. 4.8d. Numerous hits are present on the n-alkane band (after 15 min on ^{1}D and ~ 2.5 s on ^{2}D) in Fig. 4.8d, suggesting they result from RF-A having a higher concentration of these compounds relative to RF-C. Several other hit markers can be seen tracing the branched alkane bands between 10-20 min on ^{1}D and 1.5-2.2 s on ^{2}D, indicating the distribution of these compounds varies between the fuels. The hit markers from 7-15 min on ^{1}D and 0.8-1 s on ^{2}D indicate the presence of aromatic compounds present in RF-A but absent from RF-C. These differences are readily visualized by utilizing the stitch chromatogram approach to generate chromatograms of both the bulk hydrocarbon and low-level polar compound hits separately for both fuels in each pair. Furthermore, dividing these stitch signal data into each other yields a signal-ratio stitch chromatogram, which is essentially the concentration ratio for each peak feature, provided in Fig. C5 [28].

The inter-fuel comparisons (neat versus neat and pass versus pass) and extract chromatograms allow determination of individual fuel polar profiles, facilitating the interpretation of both the bulk hydrocarbon and low-level polar species differences via the F-

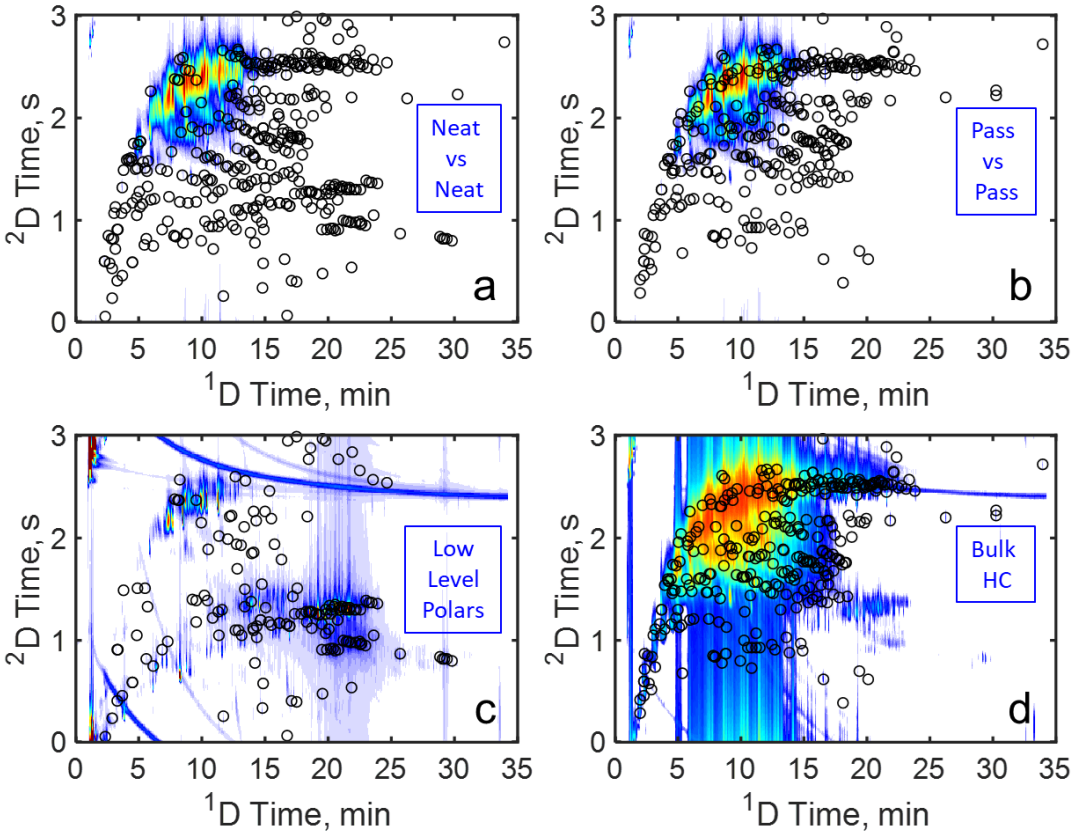


Figure 4.8. (a) GC×GC-TOFMS TIC chromatograms of rocket fuel RF-A with 344 hit markers with an F-ratio > 100 for the inter-fuel neat RF-A versus neat RF-C comparison, and (b) 310 hit markers with an F-ratio > 100 for the inter-fuel pass RF-A versus pass RF-C comparison. Both comparisons contain many of the same hits (due to the bulk hydrocarbon differences), however the pass versus pass comparison will lack hits due to compounds extracted from the neat fuels (low-level polar species). Thus, by comparing the two hit lists, a hit list representing the low-level polar species is created from the hits only present in the neat versus neat comparison. (c) These low-level polar species hits have been plotted over the 1:10 diluted RF-A extract chromatogram. (d) The hits of the pass versus pass hit list, representing the bulk hydrocarbon (HC) compound hits, are plotted over the TIC chromatogram of fuel RF-A with a more intense color scale to better visualize the location of relevant compounds.

ratio comparisons. By combining the two comparisons, low-level polar species differences are discriminated from the bulk hydrocarbon differences between fuels. Since discriminated compounds are determined by the SPE cartridge silica stationary phase, an analyst may target desired compound types through careful selection of the SPE phase and procedure. This process may be generalized beyond extraction to identify compounds using other sample treatments.

4.4. Conclusion

In this work, we have introduced a procedure that advances both targeted polar compound discovery and comprehensive fuel comparison. Fuels treated with SPE to remove polar compounds allows their discovery with supervised F-ratio analysis, and visualization with the stitch chromatogram method. Moreover, this process can be completed with a fraction of the data utilizing tile-based 1v1 analysis. The procedure may be coupled with comparisons of untreated (neat) and SPE-treated (pass) fuel to generate hit lists detailing the comprehensive differences in high-concentration compounds (bulk hydrocarbon) and low-concentration (SPE-extracted) compounds. The comparative analytical methods introduced and demonstrated herein are a significant advancement toward a deeper understanding of the chemical compositional reasons behind variations in physical properties and thermal performance.

Acknowledgements

The work at the University of Washington was performed under subcontract to Jacobs Technology (RAPT1-0000000061), Air Force Research Laboratory, Edwards AFB, CA. Distribution Statement A: Approved for Public Release; Distribution is Unlimited. AFRL-2022-3894. The fuels were provided by the Air Force Research Laboratory/RQRC, Edwards AFB, CA. Certain commercial equipment, instruments, or materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by the University of Washington or the United States Air Force, nor does it imply the materials or equipment identified are necessarily the best available for that purpose.

4.5. References

[1] D.J. Cookson, J.L. Latten, I.M. Shaw, B.E. Smith, Property-composition relationships for diesel and kerosene fuels, Fuel 64 (1985) 509–519. https://doi.org/10.1016/0016-2361(85)90086-9.

[2] D.J. Cookson, C. Paul Lloyd, B.E. Smith, Investigation of the Chemical Basis of Diesel Fuel Properties, Energy and Fuels 2 (1988) 854–860. https://doi.org/10.1021/ef00012a021.

[3] D.J. Cookson, P. Iliopoulos, B.E. Smith, Composition-property relations for jet and diesel fuels of variable boiling range, Fuel 74 (1995) 70–78. https://doi.org/10.1016/0016-2361(94)P4333-W.

[4] T.J. Bruno, M.L. Huber, E.W. Lemmon, Effect of RP-1 compositional variability on thermophysical properties, Energy and Fuels 23 (2009) 5550–5555. https://doi.org/10.1021/ef900597q.

[5] G. Liu, L. Wang, H. Qu, H. Shen, X. Zhang, S. Zhang, Z. Mi, Artificial neural network approaches on composition-property relationships of jet fuels based on GC-MS, Fuel 86 (2007) 2551–2559. https://doi.org/10.1016/j.fuel.2007.02.023.

[6] M.J. DeWitt, T. Edwards, L. Shafer, D. Brooks, R. Striebich, S.P. Bagley, M.J. Wornat, Effect of Aviation Fuel Type on Pyrolytic Reactivity and Deposition Propensity under Supercritical Conditions, Ind. Eng. Chem. Res. 50 (2011) 10434–10451. https://doi.org/10.1021/IE200257B.

[7] R.R. Mallepally, B.A. Bamgbade, M.A. McHugh, H.O. Baled, R.M. Enick, M.C. Billingsley, Measurements and modeling of the density of rocket propellant RP-2 at temperatures to 573 K and pressures to 100 MPa, Fuel 253 (2019) 1193–1203. https://doi.org/10.1016/J.FUEL.2019.05.089.

[8] G. Datschefski, Role of the jftot in aviation fuel stability research, Fuel Sci. Technol. Int. 6 (1988) 609–631. https://doi.org/10.1080/08843758808915906.

[9] B. Beaver, L. Gao, C. Burgess-Clifford, M. Sobkowiak, On the mechanisms of formation of thermal oxidative deposits in jet fuels. Are unified mechanisms possible for both storage and thermal oxidative deposit formation for middle distillate fuels?, Energy and Fuels 19 (2005) 1574–1579. https://doi.org/10.1021/ef040090j.

[10] I.C. Lee, H.C. Ubanyionwu, Determination of sulfur contaminants in military jet fuels, Fuel 87 (2008) 312–318. https://doi.org/10.1016/J.FUEL.2007.05.010.

[11] E. Alborzi, P. Gadsby, M.S. Ismail, A. Sheikhansari, M.R. Dwyer, A.J.H.M. Meijer, S.G. Blakey, M. Pourkashanian, Comparative Study of the Effect of Fuel Deoxygenation and Polar Species Removal on Jet Fuel Surface Deposition, Energy and Fuels 33 (2019) 1825–1836. https://doi.org/10.1021/acs.energyfuels.8b03468.

[12] Z. Liu, S. Tang, Z. Li, Z. Qin, S. Yuan, L. Wang, L. Wang, X. Zhang, G. Liu, An improved kinetic model for deposition by thermal oxidation of aviation hydrocarbon fuels, Fuel 258 (2019) 116139. https://doi.org/10.1016/J.FUEL.2019.116139.

[13] P.M. Paniv, S. V. Pysh'ev, V.I. Gaivanovich, O.I. Lazorko, Noncatalytic oxidation desulfurization of the kerosene cut, Chem. Technol. Fuels Oils 42 (2006) 159–166. https://doi.org/10.1007/s10553-006-0049-4.

[14] B. Omais, M. Courtiade, N. Charon, D. Thiébaut, A. Quignard, M.C. Hennion, Investigating comprehensive two-dimensional gas chromatography conditions to optimize the separation of oxygenated compounds in a direct coal liquefaction middle distillate, J. Chromatogr. A 1218 (2011) 3233–3240. https://doi.org/10.1016/J.CHROMA.2010.12.049.

[15] B. Kehimkar, J.C. Hoggard, L.C. Marney, M.C. Billingsley, C.G. Fraga, T.J. Bruno, R.E. Synovec, Correlation of rocket propulsion fuel properties with chemical composition using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry followed by partial least squares regression analysis, J. Chromatogr. A 1327 (2014) 132–140. https://doi.org/10.1016/J.CHROMA.2013.12.060.

[16] 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. https://doi.org/10.1016/J.FUEL.2018.07.081.

[17] 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. https://doi.org/10.1016/S0169-7439(01)00198-8.

[18] K.J. Johnson, S.L. Rose-Pehrsson, R.E. Morris, Characterization of fuel blends by GC-MS and multi-way chemometric tools, Pet. Sci. Technol. 24 (2006) 1175–1186. https://doi.org/10.1081/LFT-200048192.

[19] T.A. Bolotnik, Y. V. Timchenko, I. V. Plyushchenko, V. V. Levkina, A. V. Pirogov, A.D. Smolenkov, M. V. Popik, O.A. Shpigun, Use of Chemometric Methods of Data Analysis for the Identification and Typification of Petroleum and Petroleum Products, J. Anal. Chem. 2019 7413. 74 (2019) 1336–1340. https://doi.org/10.1134/S1061934819130045.

[20] 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. https://doi.org/10.1021/acs.energyfuels.9b04108.

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

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

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

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

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

[26] 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. https://doi.org/10.1016/j.jchromb.2019.01.029.

[27] S.E. Reichenbach, C.A. Zini, K.P. Nicolli, J.E. Welke, C. Cordero, Q. Tao, Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition, J. Chromatogr. A 1595 (2019) 158–167. https://doi.org/10.1016/j.chroma.2019.02.027.

[28] 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. https://doi.org/10.1016/j.chroma.2020.461401.

[29] 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. https://doi.org/10.1016/j.chroma.2021.462092.

[30] B. Bennett, S.R. Larter, Quantitative separation of aliphatic and aromatic hydrocarbons using silver ion-silica solid-phase extraction, Anal. Chem. 72 (2000) 1039–1044. https://doi.org/10.1021/ac9910482.

[31] L.M. Balster, S. Zabarnick, R.C. Striebich, L.M. Shafer, Z.J. West, Analysis of polar species in jet fuel and determination of their role in autoxidative deposit formation, Energy and Fuels 20 (2006) 2564–2571. https://doi.org/10.1021/ef060275l.

[32] R.C. Striebich, J. Contreras, L.M. Balster, Z. West, L.M. Shafer, S. Zabarnick, Identification of polar species in aviation fuels using multidimensional gas chromatography-time of flight mass spectrometry, Energy and Fuels 23 (2009) 5474–5482. https://doi.org/10.1021/ef900386x.

[33] D. Jones, C.E. West, A.G. Scarlett, R.A. Frank, S.J. Rowland, Isolation and estimation of the “aromatic” naphthenic acid content of an oil sands process-affected water extract, J. Chromatogr. A 1247 (2012) 171–175. https://doi.org/10.1016/j.chroma.2012.05.073.

[34] V. V. Lobodin, W.K. Robbins, J. Lu, R.P. Rodgers, Separation and Characterization of Reactive and Non-Reactive Sulfur in Petroleum and Its Fractions, Energy and Fuels 29 (2015) 6177–6186. https://doi.org/10.1021/acs.energyfuels.5b00780.

[35] K. Mainali, M. Garcia-Perez, Identification and quantification of trace oxygenated compounds in alternative jet fuels: Fluorescence methods for fast detection of phenolic compounds in operational field conditions, Fuel 271 (2020) 117652. https://doi.org/10.1016/j.fuel.2020.117652.

[36] S. Yuan, H. Li, Z. Liu, Y. Wang, L. Wang, X. Zhang, G. Liu, Measurement of non-hindered and hindered phenolic species in aviation fuels via tandem-SPE with comprehensive GC × GC–MS/FID, Fuel 287 (2021) 119561. https://doi.org/10.1016/j.fuel.2020.119561.

[37] 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. https://doi.org/10.1021/acs.analchem.2c00223.

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

[39] 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. https://doi.org/10.1021/acs.analchem.0c03456.

[40] 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 (2021) 122844. https://doi.org/10.1016/J.TALANTA.2021.122844.

[41] 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. https://doi.org/10.1021/ACS.ANALCHEM.0C02136.

Chapter 5. Investigating analyte breakthrough under non-linear isotherm conditions during solid phase extraction facilitated by non-targeted analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry

This chapter was reproduced from Grant S. Ochoa, Robert E. Synovec, "Investigating analyte breakthrough under non-linear isotherm conditions during solid phase extraction facilitated by non-targeted analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry" *Talanta*, Submitted (2023)

5.1. Introduction

Sample preparation is a critical step to any instrumental analysis technique, and solid phase extraction (SPE) serves as a rapid and popular method for isolating and preconcentrating analytes from complex sample matrices for various applications [1–14]. With SPE, analytes are adsorbed/absorbed to the solid stationary phase from the mobile phase, isolating them from the majority of the initial sample matrix. The extracted analytes are then typically recovered by eluting them into a minimal volume of a suitable solvent to achieve preconcentration of these extracted compounds. However, the effectiveness of SPE can be hampered by limited adsorption capacity of the stationary phase, leading to non-linear isotherm elution conditions and/or competitive binding processes resulting in premature breakthrough of many analyte compounds in complex matrices where the amount of compounds extractable by the stationary phase may be unknown [15–18]. The limited adsorption capacity is a result of the dead volume of the SPE cartridge being much smaller than the volume of sample passed through the cartridge in a typical SPE experiment.

Breakthrough elution of compounds from an SPE cartridge can be described using displacement chromatography principles, where the various components in the sample act as

displacers of other less retained components. Thus, the concentration of compounds adsorbed to the cartridge at any point during the initial sample *loading step* will be a function of the volume of sample passed through the cartridge. This opens the opportunity to study SPE as a chromatographic process, by sequentially analyzing fractions to monitor the retention of all compounds. Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) is particularly well suited to this task with its ability to provide rich chemical information for diverse sample matrices [19–26]. Indeed, GC×GC-TOFMS has often been paired with SPE for complex sample preparation and subsequent analysis of the preconcentrated fractions such as petroleum-based fuels. Fuels such as kerosene-based jet fuel, such as JP-8, have thousands of organic compounds and the fuels must meet specification limits on heteroatom containing or polar compounds, necessitating the use of SPE to facilitate a focused analysis of the trace polar compounds [14,27–29]. This makes kerosene-based jet fuel samples well suited for the investigation of the retention behavior of extracted polar compounds using SPE.

Two general approaches can be taken to study (and perhaps implement) SPE as a chromatographic process in which sequential fractions of an introduced fuel sample can be monitored to assess the dynamic retention behavior of a broad range of analyte compounds, either via a targeted approach or a non-targeted approach. The targeted approach involves spiking the fuel with curated polar compounds to study their interaction with the SPE cartridge. This provides points of interest in the data set which are known and can be quantified directly. In contrast, a non-targeted approach must consider all the data impartially to find the extraction differences. A technique fitting of the task is tile-based 1v1 analysis, a non-targeted discovery technique for locating distinguishing analyte “features” between two chromatograms [30].

Based on the algorithmic framework of the tile-based F-ratio analysis software, the 1v1 analysis algorithms tile the 2D chromatographic space using four overlapping tile grids to minimize the effect of misalignment, followed by redundant hit removal; however, unlike tile-based F-ratio which requires several samples per sample class, with 1v1 analysis only two chromatograms are required [31–34]. In previous work we have demonstrated the 1v1 analysis approach to be more effective than other analysis approaches to compare two chromatograms. Indeed, tile-based 1v1 analysis has been found to discover analytes in complex samples with a similar fidelity to tile-based F-ratio analysis [30].

In the present study, we aim to gain insight into what extent SPE can be used as a "comprehensive" sample preparation method when applied to a complex sample matrix such as jet fuel, and further to explore the possibility of using the sequential fractions obtained from a SPE cartridge as another chromatographic dimension. Specifically, we investigate the retention behavior of extractable compounds in JP-8 jet fuel by sequential collection of the pass fractions using silica-based SPE, and then alumina-based SPE. The collected sequential pass fractions from the two types of SPE cartridges are then separated and detected by GC×GC-TOFMS, producing a data set for the neat fuel and 9 sequential pass fractions for each SPE stationary phase. To facilitate this, we spike the fuel with 10 heteroatom-containing polar compounds and analyze the data set with tile-based 1v1 analysis to discover a broad range of native compounds that are also affected by the SPE process. Regarding the experimental design, the spiked compounds are introduced at a very low fraction of the fuel itself so as not to impact the extraction of compounds native to the fuel. Thus, the spiked compounds serve to provide an *in situ* observation and calibration of the dynamic SPE process. The two SPE stationary phases,

alumina and silica, facilitate the investigation of the differences in polar compound extractability and their impact on the final extract obtained from both cartridge types.

5.2. Experimental

5.2.1. Solid phase extraction

A JP-8 jet fuel spiked with 10 polar compounds at 40 ppm each (Table 5.1) was prepared for extraction using SPE. A nominally ~1 mL alumina phase SPE cartridge with a dead volume ~0.4 mL (SepPak, Waters, Milford, MA) was rinsed with 5 mL methanol followed by 15 mL of hexanes (Sigma-Aldrich, St. Louis, MO) to condition the cartridge. Next, 9 mL of spiked fuel sample were loaded onto the cartridge at a flow rate at about 2 mL/min. During this step, every 1 mL of sample exiting the cartridge is kept and saved as a 'pass fraction' resulting in 9 total pass fraction samples. The total sample load volume, 9 mL, being much larger than the cartridge dead volume, 0.4 mL, results in nonlinear isotherm retention behavior, due to the limited adsorption capacity of a cartridge. Finally, polar compounds that were preconcentrated onto a SPE cartridge without eluting in any of pass fractions are extracted from the SPE cartridge with 0.5 mL methanol. This overall process was repeated with a silica phase SPE cartridge (SepPak, Waters). The neat sample (fuel + spiked compounds), 9 pass fractions, and SPE extract samples, collected using both SPE phases, were then analyzed by GC×GC-TOFMS.

5.2.2. GC×GC-TOFMS procedure

From the SPE experiments using both stationary phases, the spiked JP8 jet fuel, 9 pass fractions, and SPE extract samples were analyzed by GC×GC-TOFMS (Pegasus BT TOFMS, LECO, St. Joseph, MI), that included an Agilent 7890 gas chromatograph with an L-PAL3 autosampler (LECO, St. Joseph, MI). A "reverse" column GC×GC configuration was used with a polar primary column, ^{1}D (Restek Rxi-17SilMS 30 m × 0.250 mm i.d. × 0.25 µm film

thickness) and a nonpolar secondary column, ^{2}D (Restek Rxi-1MS 2 m × 0.180 mm i.d. × 0.18 μm film thickness) connected by a SilTite μ-Union (Trajan Scientific Americas, Austin, TX). Ultrahigh purity helium (Grade 5, 99.999%, Praxair, Seattle, WA, USA) was used as the carrier gas at a constant flow of 2.0 mL/min. The temperature program held the oven at 40°C for 1 min, ramping to 200°C at a rate of 5°C/min, where it was held for 1 min. The ^{2}D oven and modulator had a temperature offset from the oven of +18°C and +20°C, respectively. The modulation period was 3 s. The ion source was set to 225 °C, and mass channels, *m/z* 30-334, were collected at a 100 spectra/s after a 2 min acquisition delay. One injection replicate was collected per fraction, yielding 11 chromatograms per SPE phase, totaling 22 chromatograms.

5.2.3. Data Processing

The GC×GC-TOFMS chromatograms were imported into MATLAB 2021a (The Mathworks Inc., Natick, MA) from the LECO ChromaTOF for BT software v5.20 (LECO, ST. Joseph, MI). All subsequent data analyses were performed in MATLAB, including principal component analysis (PCA) using the built-in singular value decomposition (SVD) function. The signal of six native compounds in all of the samples (adamantane, 1,4-dimethylcyclohexane, tetradecane, m-cymene, tetralin, and 6-methyltetralin) that were not extracted by either SPE stationary phase to any noticeable extent were averaged for each of the 11 chromatograms and used as *in situ* internal standards for normalization purposes to account for sample-to-sample injection variation. Tile-based 1v1 analysis, which is based upon the same framework as tile-based F-ratio analysis, was utilized for the direct comparison of two chromatograms in this work [30]. Instead of F-ratios, the sum-normalized absolute difference rank metric (*RM*) is calculated for tile-based 1v1 analysis,

$$RM_{\mathrm{i}} = \frac{|s(m/z)_{\mathrm{neat}} - s(m/z)_{\mathrm{i}}|}{s(m/z)_{\mathrm{neat}} + s(m/z)_{\mathrm{i}}} * 100\% \quad (5.1)$$

where $s(m/z)_{\mathrm{i}}$ and $s(m/z)_{\mathrm{neat}}$ are the summed signals for a given hit tile per *m/z* in a given pass fraction (sample class i) versus the neat spiked fuel (neat sample class). This unitless rank metric ranges from *RM* of 0% to 100% and allows the ranking of significant differences high on a hitlist regardless of the direction of signal change. Tile-based 1v1 analysis was applied to the sequential comparison of chromatograms to find differences between the spiked neat fuel and the various pass fractions. Initially, comparisons were performed on the spiked neat fuel versus each pass fraction for both SPE phases, yielding nine 1v1 analysis comparisons per SPE phase (neat versus 1 mL pass fraction, neat versus 2 mL pass fraction, and so on to neat fuel versus 9 mL pass fraction). However, it was observed that nearly all of the compounds discovered from these nine 1v1 analyses were found in the neat versus 1 mL pass fraction comparison, and further analysis of the remaining pass fractions did not discover any new analytes to consider, thus the subsequent data analysis steps were based upon the hitlist generated from the neat sample versus 1 mL pass fraction sample comparison for each SPE phase. Thus, the rank metric calculation used for the 1v1 analysis hitlist generation were obtained using the following equation,

$$RM_1 = \frac{|s(m/z)_{\mathrm{neat}} - s(m/z)_1|}{s(m/z)_{\mathrm{neat}} + s(m/z)_1} * 100\% \quad (5.2)$$

where $s(m/z)_1$ is the signal for a given analyte hit from the first 1 mL pass fraction chromatogram and $s(m/z)_{\mathrm{neat}}$ is the signal for a the same analyte hit from the neat spiked fuel chromatogram.

To evaluate the change in concentration of each discovered analyte compound via the hitlist generated using Eq. (5.2) ranked by RM_1 throughout the entire extraction-elution process, concentration ratios (CR_{i}) are calculated for each analyte in the neat versus *each* pass fraction

comparison using Eqs. (1) to provide the RM_i for each pass fraction. Each CR_i may be obtained directly from the RM_i by the following two equations,

$$CR_i = \frac{s(m/z)_i}{s(m/z)_{neat}} = \frac{\left(1+\frac{1}{RM_i}\right)}{\left(\frac{1}{RM_i}-1\right)}, \text{ where } s(m/z)_i > s(m/z)_{neat} \quad (5.3)$$

$$CR_i = \frac{s(m/z)_i}{s(m/z)_{neat}} = \frac{\left(\frac{1}{RM_i}-1\right)}{\left(1+\frac{1}{RM_i}\right)}, \text{ where } s(m/z)_{neat} > s(m/z)_i \quad (5.4)$$

where $s(m/z)_{neat}$ is the signal of a given analyte hit in the spiked neat fuel and $s(m/z)_i$ is the signal of the same analyte hit in the i^{th} pass fraction. If $s(m/z)_{neat}$ is equal to $s(m/z)_i$ than CR_i is exactly 1. A plot of CR_i against their extract fraction (neat, 1 mL fraction, 2 mL fraction….9 mL fraction) provides *concentration ratio* profiles (*CR* profiles), allowing for interpretation of the discovered analyte extraction behavior on each SPE phase.

The following computational conditions were applied to all 1v1 analyses to generate hitlists, to analyze the hitlists, and to prepare stitch chromatograms. The tile dimensions for the initial 1v1 analysis discovery were 12 s on ^{1}D by 300 ms on ^{2}D with a *S/N* threshold of 10. The *m/z* providing the maximum *RM* was used to rank the hits in each hitlist [35]. Visualization of the hits from each neat versus 1 mL fraction 1v1 analysis is accomplished by piecing together data from each of the hit locations into a single 'stitch' GC×GC chromatogram via the following algorithmic process based off the principles of the enhanced total ion current chromatogram algorithm [27,36]. Starting from the top of the hitlist, on a per hit basis a tile of data with the same dimensions as stated above used for 1v1 analysis hit discovery is cut around each pin location using the signal from the *m/z* with the top *RM*. These tiles of signal are unfolded and then all the ^{2}D peak maxima are detected above a *S/N* threshold of 10, keeping one peak per

modulation, thus the maximum number of modulations kept per sample is the ^{1}D tile size plus one. A window of data is cut around each ^{2}D peak equaling the average ^2w_b (4σ width-at-base) of the chromatogram. In this study this ^{2}D window corresponded to 180 ms, i.e., 18 data points (spectra) wide with the mass spectrum collection frequency of 100 Hz. These peak windows are then pasted into an array of zeros the same size of the 2D chromatogram from the same coordinates from which they were cut. This process is repeated for each hit, yielding a 2D chromatogram with all of the peaks from the hits of interest 'stitched' together, like making a quilt, hence the name 'stitch' chromatogram. The resulting stitch chromatogram highlights the class distinguishing features without being encumbered by a significant amount of interfering signal. If there is signal in a hit location without a peak maximum, there will be no data cut from the hit tile for that location for the purposes of constructing the stitch chromatogram.

Table 5.1. Polar compounds spiked into the JP-8 jet fuel along with their ^{1}D and ^{2}D retention times (RT) for the alumina phase separation. Characteristic m/z represent base and other major peaks of the compounds mass spectrum. The type of heteroatom present in each spiked analyte is also indicated.

Name	^{2}D RT, s	^{1}D RT, min	Char. *m/z*	Type
1-decanol	1.15	16.20	70, 112	Oxygen
dibenzofuran	0.50	25.80	168, 139	Oxygen
methyl decanoate	1.16	17.60	74, 87, 143	Oxygen
phenol	0.32	10.05	66, 94	Oxygen
p-cresol	0.38	12.90	77, 107	Oxygen
2-butyl-5-ethylthiophene	1.02	16.00	125, 168	Sulfur
2-hexylthiophene	1.04	16.60	97, 168	Sulfur
2-methylbenzothiophene	0.50	19.85	147, 148	Sulfur
o-toluidine	0.25	13.95	106, 107	Nitrogen
indole	0.14	21.30	90, 117	Nitrogen

5.3. Results and Discussion

Ten compounds containing heteroatoms (oxygen, sulfur, and/or nitrogen) were spiked into JP-8 fuel for method development (Table 5.1). Figure 5.1 shows GC×GC-TOFMS chromatograms (TIC) of JP-8 jet fuel collected for both the alumina (Fig. 5.1A) and silica (Fig. 5.1B) phase data with the location of the 10 spiked compounds overlaid. These chromatograms, collected with a reverse column GC×GC configuration (polar ^{1}D by non-polar ^{2}D) separates the compound classes by relative polarity on the ^{2}D dimension, with alkanes at the top, followed by cycloalkanes beneath them and then aromatic compounds toward the bottom. The oxygen and nitrogen containing spiked compounds are situated at shorter ^{2}D retention times than the aromatics due to the polarity imparted by their heteroatoms, while the sulfur containing spiked compounds are located within the aromatic region. Note, the ^{2}D retention times between the alumina and silica-based chromatograms differ because they were collected months apart on GC×GC columns that were of the same phase but slightly different lengths. Figure 5.1C-D show the methanol extract chromatograms obtained from the SPE procedure outlined in the methods for alumina and silica, respectively. Each chromatogram shows native compound peaks that are not readily visible in the initial fuel TIC chromatograms from Fig. 5.1A-B that fall in a region beneath the aromatics, suggesting higher polarity. However, the pattern of native compound peaks between the two phases differs, with the alumina extract compounds eluting between 10 and 23 min, while the silica extract compounds eluting later from 23-30 min. This suggests that each phase interacts with the spiked fuel matrix differently, as anticipated.

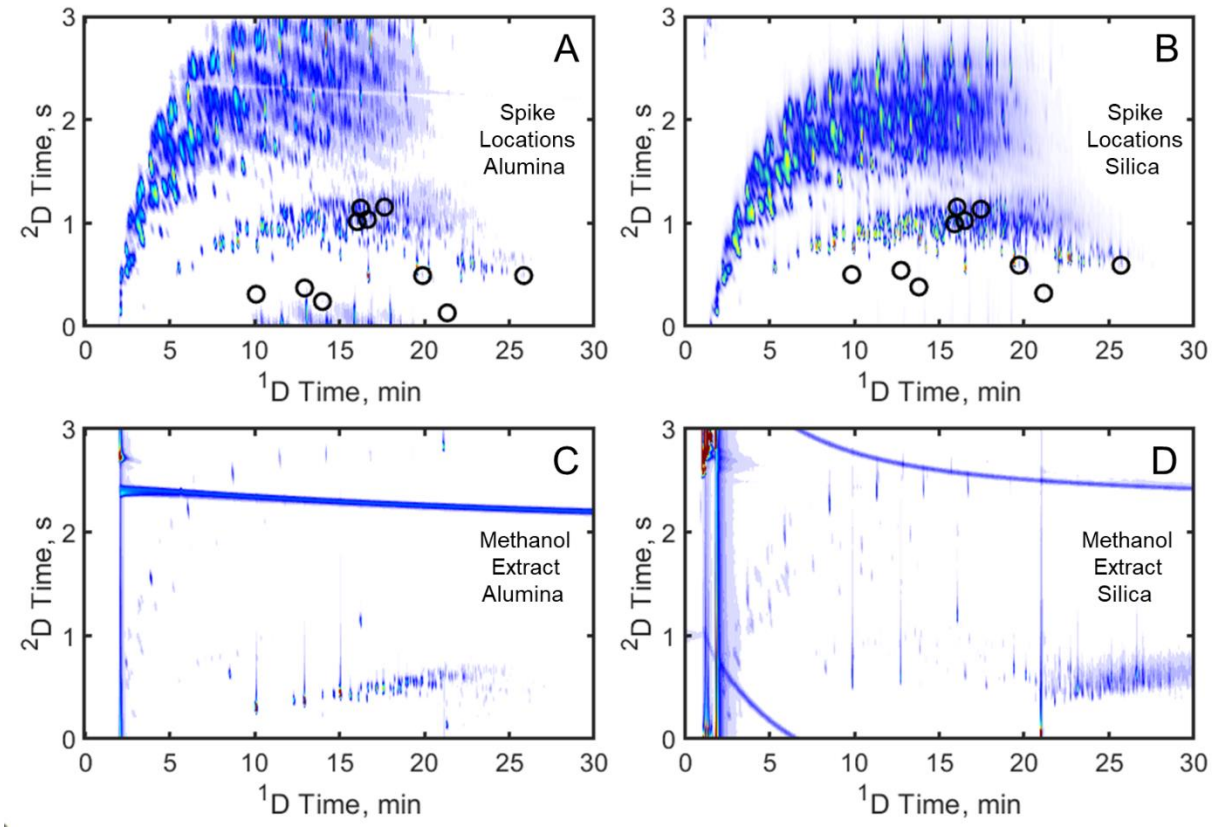


Figure 5.1. GC×GC-TOFMS chromatograms of the TIC for the JP-8 jet fuel, collected for both the alumina (A), and the silica phase data sets (B). The locations of the 10 polar compounds spiked at 40 ppm are overlaid. Chromatograms of the methanol extracts for both the alumina (C) and silica (D) extractions show differences in compounds retained by each phase.

Next, in order to study analyte breakthrough under non-linear isotherm conditions via competitive binding during SPE, alumina and silica SPE was performed on the JP-8 jet fuel spiked with the 10 polar compounds at 40 ppm each, sequestering each milliliter of fuel passing through the SPE cartridge as a pass fuel fraction. GC×GC-TOFMS data was collected on the neat, spiked fuel, and each of the pass fuel fractions, and followed by non-targeted 1v1 analysis of the neat, spiked fuel versus the first pass fuel fraction. To investigate how the concentration of each spiked compound changes throughout the SPE elution process, each spiked compound was quantified in each pass fraction by the method outlined in Fig. 5.2. A tile with a ^{1}D interval of 3 modulations and ^{2}D interval of 31 mass spectra is drawn around the retention time of each spiked

compound. The signal in this region at the *m/z* that provided the highest *RM* for the given spiked compound in the 1v1 analysis (spiked fuel versus the first pass fuel fraction) to rank the hits is summed for the neat fuel and each pass fuel fraction. Normalizing these summed signals for each pass fraction to the summed signal of the given analyte in the neat fuel yields an estimate of the CR_i, applying either Eq. (5.3) or (5.4), showing how the concentration changes during the SPE elution compared to the concentration in the spiked neat fuel (initial state). Plotting the CR_i as a function of volume passed through the SPE cartridge yields *CR* profiles that provide visualization of each spiked compound interaction with the SPE stationary phase throughout the extraction process.

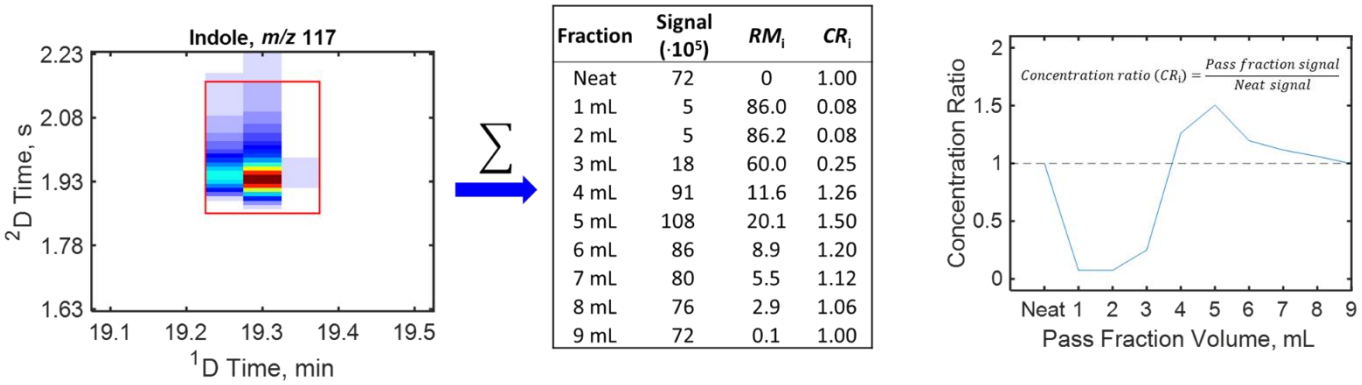


Figure 5.2. Process for calculating CR profiles using data from the 1v1 analysis, demonstrated with a spiked compound, indole. Signal for the analyte peak is summed in a tile centered on the peak at the top RM_1 m/z for each pass fraction volume as well as for the spiked neat fuel. These signals are divided by the analyte signal in the spiked neat fuel and plotted against the pass fraction volume to yield concentration ratio profiles (CR profiles). Rank metrics (RM_i) are provided that are used in turn to calculate each concentration ratio (CR_i) using data from the 1v1 analysis. The CR_i values are then used to form the CR profile vector.

While 1v1 analysis "discovery" of these ten spiked compounds was not necessary, due to previous knowledge of their retention times when the mix was run alone, it is important to treat the GC×GC-TOFMS data as if the spiked compound identity were unknown to verify the quality of *RM*s obtained by 1v1 analysis and overall method development for untargeted analysis. Figure 5.3 shows the RM_1 distributions via Eq. (5.2), and the CR_1 distributions using either Eq. (5.3) or (5.4) for the ten spiked compounds for the neat versus 1 mL comparison for both the alumina

(top row) and silica (bottom row) extractions. The RM_1 distribution for the alumina SPE experiment for the neat versus 1 mL comparison in Fig. 5.3A shows several compounds exhibit high RM_1, a trend seen for the silica SPE experiment comparison as well (Fig. 5.3C). Similarly, the CR_1 distribution of the spiked compounds shows several CR_1 close to zero for both the alumina and silica SPE comparisons, while some remain around a CR_1 of 1 (Fig. 5.3B, D). For this study to focus on the hits which change the most in the subsequent analysis of the native polar compounds likely present in the JP-8 jet fuel we impose a RM_1 threshold of > 33.33 % (equivalent to a CR_1 < 0.5), indicated by the dashed vertical lines in Fig. 5.3. This threshold excludes compounds that do not change in concentration appreciably during the extraction process, however, these compounds possess sufficient signal to be initially discovered by the tile-based 1v1 analysis software; such hits are simply further down the hitlist.

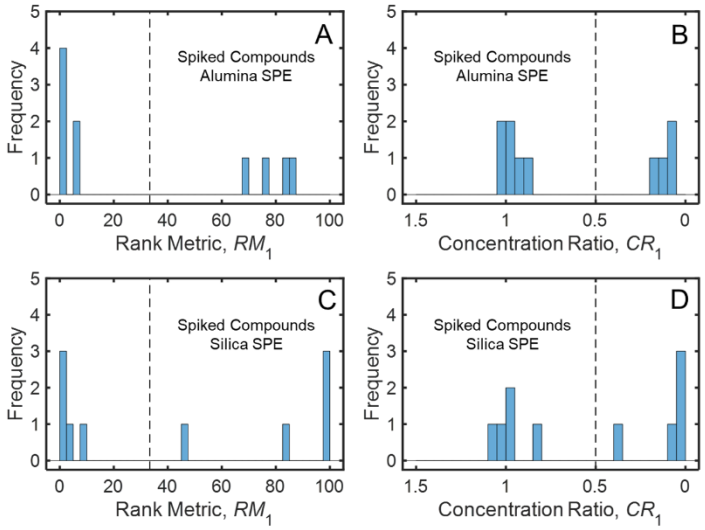


Figure 5.3. (A) Rank metric (RM_1) distribution and (B) concentration ratio (CR_1) distribution of the spiked polar compounds for the neat versus 1 mL pass fraction 1v1 comparison in the alumina data. The RM_1 distribution and CR_1 distribution for the spiked compounds in the silica data are provided in (C) and (D), respectively. Thresholds for determining if a given hit makes the final hitlist are indicated by the vertical dashed lines, at a RM_1 of 33.33 % corresponding to a CR_1 of 0.5.

Next, all ten spiked compounds *CR* profiles are compared between the alumina and silica SPE phases in Figure 5.4. The *CR* profile comparisons for phenol and p-cresol indicate that both compounds interact similarly on alumina relative to silica. Each compound is well retained by the alumina phase as evident by the CR_i of each compound dropping and remaining low for all of the pass fractions. In contrast, for the silica phase the CR_i for each compound drops low but starts to return to neat levels for the later pass fractions, eventually exceeding a concentration ratio of 1 indicating these compounds are more concentrated in these pass fractions than the spiked neat sample. These differences are likely due to acid-base interactions of the phenolic compounds with the terminal hydroxyl groups on the stationary phase surface. Silica is slightly acidic while alumina is slightly basic leading to stronger interactions of the slightly acidic phenolic compounds with alumina. The nitrogen containing compounds, o-toluidine and indole, also interact with the two phases differently, though seemingly less drastically as the *CR* profile of each returns to and eventually exceeds 1 albeit at a different pace. The four spiked compounds that extracted well for both SPE phases (phenol, p-cresol, o-toluidine and indole) serve as examples of the *CR* profiles to anticipate for the true positive hits likely to be (simultaneously) obtained for the native (trace level) polar compounds in the JP8 fuel via tile-based 1v1 analysis. Another spiked compound, methyl decanoate, was slightly extracted by the silica phase initially but remained unaffected by the alumina phase. However, its *CR* profile returns to 1 more rapidly than the previous four compounds, possibly due to its lack of hydrogen bonding capability. 1-decanol was slightly retained by silica, as its *CR* profile decreases and increases 6 mL into the extraction. Lastly, dibenzofuran and the spiked sulfur compounds, 2-hexylthiophene, 2-butyl-5-ethylthiophene, and 2-methylbenzothiophene were not significantly extracted by either phase. Thus, alumina and silica phase SPE do not appear to be suitable for the general extraction of

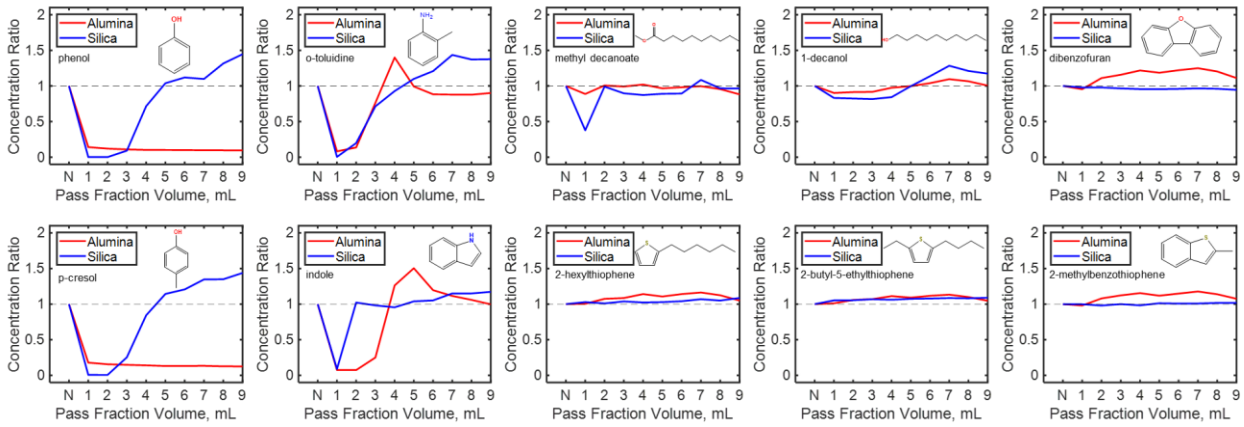


Figure 5.4. Plots comparing the CR profiles of the 10 spiked polar compounds for the alumina and silica extractions as a function of the pass fraction volume. Structures are provided for each spike to aid in interpretation.

aromatic sulfur compounds in this JP-8 jet fuel under the conditions employed herein. Using the threshold of $RM_1 > 33.33$ % which equates to a $CR_1 < 0.5$, as intended naturally excludes the spikes which did not interact with the stationary phases, retaining four spiked analytes for alumina and five spiked analytes for silica.

While these *CR* profiles of the ten spiked compounds demonstrate differences in extraction behavior between the two SPE phases that are discoverable by 1v1 analysis, a complete explanation cannot be provided based upon the stationary phase interactions alone. While these compounds are all spiked at 40 ppm they only represent ≈ 2% of the polar compounds found in the extract (Fig 5.1C-D). The dynamic retentivity of each compound is affected by the other compounds present in the overall fuel sample matrix; thus, it is important to contextualize the extraction behavior of the spiked compounds relative to native polar compounds likely present in the JP-8 jet fuel. For the purposes of this study, we accept that the presence of the spiked compounds may slightly affect the retention behavior of the native compounds, but without significantly impacting the overall interpretation of the native

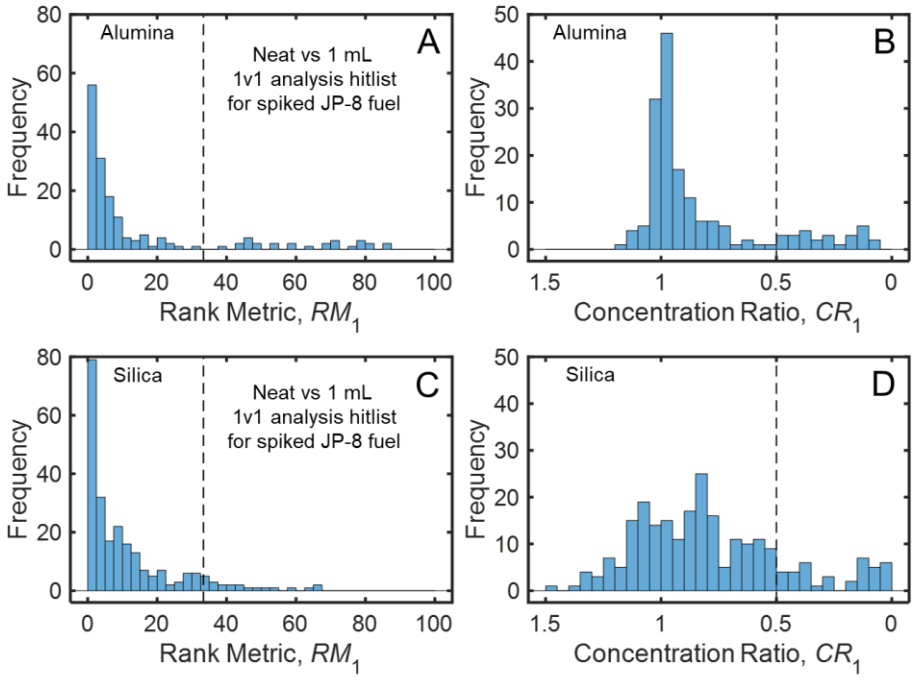


Figure 5.5. (A) Rank metric (RM_1) distribution and (B) concentration ratio (CR_1) distribution for the neat versus 1 mL pass fraction 1v1 analysis hitlists of the alumina data. The RM_1 distribution and CR_1 distribution for the neat versus 1 mL pass fraction 1v1 analysis hitlist of the silica data are provided in (C) and (D), respectively. Thresholds for determining if a given hit makes the final hitlist are indicated by the vertical dashed lines at a RM_1 of 33.33%, corresponding to a CR_1 of 0.5. The rank metric distributions (A, C) have been marked with a threshold at a RM_1 of 33.33 %. This corresponds to the concentration ratio threshold of 0.5 (B, D) (vertical dashed lines).

compound *CR* profiles. Indeed, the spiked compounds serve as *in situ* calibrants of the overall SPE extraction process. To discover the host of other compounds extracted by each phase tile-based 1v1 analysis is applied to comprehensively discover all compounds changing in concentration for the neat versus 1 mL pass fraction comparison for the alumina and silica GC×GC-TOFMS data. In this study, for each phase, note that the hitlist for the native compounds and spiked analytes is a single list with the spiked analyte hits interspersed in the native compound hits. Based upon the hitlist, the RM_1 distributions and CR_1 distributions for the tile-based 1v1 analysis of the neat versus 1 mL pass fraction for both SPE phases are provided in Figure 5.5. The RM_1 distributions in Fig. 5.5A and C show a large abundance of hits at high RM_1

for both the data collected with alumina and silica SPE, respectively. Likewise, the CR_1 distributions for the two phases is provided in Fig. 5.5B and D. Each comparison has many hits with a CR_1 around one with several hits occurring at lower CR_1. These plots demonstrate that there are many hits from the JP8 jet fuel that occur due to the extraction process, and with low CR_1 occurring in the first pass fraction (1 mL) that are likely polar compounds. Again, the RM_1 and thus CR_1 threshold criteria defined in Fig. 5.3 for the spiked analytes (method development) are invoked to focus on the most highly extracted compounds in the JP8 jet fuel. Table 5.2 lists the hits from the alumina neat versus 1 mL pass fraction comparison hitlist which meet the threshold criteria along with their associated location in the silica neat versus 1 mL pass fraction hitlist. Identities are provided along with forward and reverse match values (MVs) obtained from a NIST library search. Table 5.3 lists the same type of information for the hits meeting the RM_1 and CR_1 criteria for the silica SPE hitlist, with their associated location in the alumina SPE hitlist. There are 27 hits for the alumina hitlist and 21 hits for the silica hitlist with a $RM_1 > 33.33$ % ($CR_1 < 0.5$). Between the two hitlists there were 38 unique hits, with 28 hits shared between the alumina and silica hitlists which had a $RM_1 > 33.33$ % ($CR_1 < 0.5$) in at least one of the two hitlists.

The $^1D \times {}^2D$ retention time location of the hits with a $CR_1 < 0.5$ are plotted over the TIC for both the alumina and silica SPE data in Fig. 5.6A,D. The majority of hits for both phases are located beneath the aromatic band on ^2D, indicating these compounds are more polar than aromatic compounds. To visualize these hits directly (analyte hits in Tables 5.2 and 5.3), we construct stitch chromatograms which displays the signal for each peak using the top RM_1 *m/z* for each hit. Figures 5.6B,E display these stitch chromatograms for the alumina and silica SPE phase results, respectively, free of interfering compounds. Each phase had several compounds

Table 5.2. Ranked hitlist of the neat versus 1 mL pass fraction 1v1 analysis comparison for the alumina data. Only hits with a RM_1 > 33.33 % (CR_1 < 0.5) are reported. Compounds are tentatively identified with the NIST match value (MV) and reverse match value (RMV) provided, along with their ^{1}D and ^{2}D retention times (RT). Corresponding hits in the silica hitlist are also provided, if present, otherwise indicated as N/A.

Alumina								Silica				
Hit Number	Compound ID	MV	RMV	^{1}D RT, min	^{2}D RT, s	Alumina RM_1, %	CR_1	Silica Hit No	^{1}D RT, min	^{2}D RT, s	Silica RM_1, %	CR_1
1	thymol	847	867	19.00	0.56	87.47	0.07	14	19.02	0.65	37.65	0.45
2	2,6-dimethylphenol	846	846	13.95	0.44	87.32	0.07	43	13.92	0.40	22.39	0.63
3	2,3,5-trimethylphenol	885	885	17.60	0.50	82.10	0.10	5	17.52	0.64	53.14	0.31
4	2-ethyl-6-methylphenol	881	921	16.15	0.48	82.09	0.10	8	16.07	0.59	46.07	0.37
5	3,4,5-trimethylphenol	875	880	18.45	0.50	78.87	0.12	2	18.47	0.65	66.05	0.20
6	2-methyl-6-propylphenol	802	861	18.15	0.58	77.82	0.12	106	18.17	0.67	8.22	0.85
7	2,5-diethylphenol	847	879	18.60	0.55	77.67	0.13	N/A	N/A	N/A	N/A	N/A
8	2,3-dimethylphenol	888	888	15.00	0.45	76.03	0.14	4	14.97	0.63	58.94	0.26
9	2,4,6-trimethylphenol	840	933	16.65	0.53	71.84	0.16	3	16.62	0.64	63.63	0.22
10	indole	898	907	21.30	0.14	71.57	0.17	12	21.32	0.34	41.06	0.42
11	3-ethyl-5-methylphenol	865	869	17.15	0.50	71.13	0.17	129	17.17	0.65	4.45	0.91
12	o-toluidine	934	938	13.95	0.25	68.35	0.19	43	13.92	0.40	22.39	0.63
13	3,5-dimethylphenol	762	801	15.65	0.44	68.19	0.19	138	15.62	0.68	3.84	0.93
14	p-sec-butylphenol	882	893	19.30	0.58	64.06	0.22	76	19.32	0.73	11.34	0.80
15	4-methyl-2-propylphenol	784	878	19.15	0.58	59.58	0.25	N/A	N/A	N/A	N/A	N/A
16	2-ethyl-4,5-dimethylphenol	863	944	20.65	0.54	59.43	0.25	6	20.67	0.65	50.61	0.33
17	sulfolane	941	951	21.10	2.83	54.88	0.29	15	21.12	0.07	35.98	0.47
18	p-cresol	940	940	12.90	0.38	53.45	0.30	46	12.87	0.57	20.34	0.66
19	phenol	898	915	10.05	0.32	49.42	0.34	30	10.02	0.55	28.76	0.55
20	2-propylphenol	903	910	16.85	0.53	47.58	0.36	N/A	N/A	N/A	N/A	N/A
21	3-sec-butylanisole	844	866	21.40	0.63	46.20	0.37	116	21.42	0.74	6.41	0.88
22	beta-Isopropyltropolone	796	812	21.60	0.62	46.01	0.37	89	21.62	0.74	9.95	0.82
23	4-tert-butyl-o-cresol	810	837	20.40	0.64	45.92	0.37	31	20.27	0.68	28.50	0.56
24	2-ethylphenol	911	914	14.65	0.45	45.88	0.37	28	14.62	0.65	30.22	0.54
25	o-cresol	932	937	12.30	0.39	43.39	0.39	83	12.22	0.71	10.76	0.81
26	2-methyl-4-propylphenol	857	925	19.75	0.57	42.98	0.40	16	19.82	0.67	35.98	0.47
27	2-ethyl-5-n-propylphenol	816	899	20.20	0.63	39.79	0.43	N/A	N/A	N/A	N/A	N/A

extracted between 10 and 25 min on ^{1}D and 0 to 1 s on ^{2}D, representing phenolic compounds. The silica SPE data contained hits outside this region, between 5 and 10 min on ^{1}D and 1 to 2 s on ^{2}D. The top RM_1 *m/z* for these hits in this cycloalkane region are characteristic of cycloalkene fragmentation patterns. It is noted that the compounds seen eluting between 23 to 30 min on ^{1}D in Fig. 5.1D do not appear in the hitlist. The likely reason for this is that the compounds seen eluting are below the limit of detection for the GC×GC and are not discovered. However, they still preconcentrate on the silica SPE cartridge and elute in the extract, giving rise to the peaks

Table 5.3. Ranked hitlist of the neat versus 1 mL pass fraction 1v1 analysis comparison for the silica data. Only hits with a RM_1 > 33.33 % (CR_1 < 0.5) are reported. Compounds are tentatively identified with the NIST match value (MV) and reverse match value (RMV) provided, along with their ^{1}D and ^{2}D retention times (RT). Corresponding hits in the alumina hitlist are also provided, if present, otherwise indicated as N/A.

Silica								Alumina				
Hit Number	Compound ID	MV	RMV	^{1}D RT, min	^{2}D RT, s	Silica RM_1, %	CR_1	Alumina Hit No	^{1}D RT, min	^{2}D RT, s	Alumina $RM1$, %	CR_1
1	3-methylindene	845	918	17.92	0.53	66.84	0.20	30	17.95	0.41	24.27	0.61
2	3,4,5-trimethylphenol	817	859	18.47	0.65	66.05	0.20	5	18.45	0.50	78.87	0.12
3	2,4,6-trimethylphenol	797	835	16.62	0.64	63.63	0.22	9	16.65	0.53	71.84	0.16
4	2,3-dimethylphenol	881	905	14.97	0.63	58.94	0.26	8	15.00	0.45	76.03	0.14
5	2,3,5-trimethylphenol	821	828	17.52	0.64	53.14	0.31	3	17.60	0.50	82.10	0.10
6	2-ethyl-4,5-dimethylphenol	863	944	20.67	0.65	50.61	0.33	16	20.65	0.54	59.43	0.25
7	4,8-dimethyl-7-nonen-2-one	755	797	14.37	0.97	47.73	0.35	105	14.40	1.01	2.63	0.95
8	2-ethyl-6-methylphenol	881	921	16.07	0.59	46.07	0.37	4	16.15	0.48	82.09	0.10
9	4-sec-butylanisole	807	851	21.12	0.71	43.06	0.40	47	21.10	0.63	10.84	0.80
10	o-isopropylphenol	787	872	16.22	0.64	41.70	0.41	N/A	N/A	N/A	N/A	N/A
11	indole	898	907	21.32	0.34	41.06	0.42	10	21.30	0.14	71.57	0.17
12	limonene	775	817	6.22	1.43	39.23	0.44	N/A	N/A	N/A	N/A	N/A
13	thymol	832	874	19.02	0.65	37.65	0.45	1	19.00	0.56	87.47	0.07
14	sulfolane	941	955	21.12	0.07	35.98	0.47	17	21.10	2.83	54.88	0.29
15	2-methyl-6-propylphenol	857	920	19.82	0.67	35.98	0.47	26	19.75	0.57	42.98	0.40
16	cyclopropylidenecyclohexane	715	717	5.22	1.1	35.27	0.48	N/A	N/A	N/A	N/A	N/A
17	1-methyl-2-methylene-cycloheptanol	696	708	4.62	1.01	34.96	0.48	N/A	N/A	N/A	N/A	N/A
18	1-methylenespiro[4.4]nonane	751	877	7.97	1.46	34.58	0.49	N/A	N/A	N/A	N/A	N/A
19	methyl decanoate	728	738	17.62	1.17	34.10	0.49	95	17.60	1.19	3.37	0.93
20	2,2-dimethyl-3-methylenebicyclo[2.2.1]heptane	733	805	5.92	1.4	33.81	0.49	N/A	N/A	N/A	N/A	N/A
21	8,8-dimethyl-4-methylene-1-oxaspiro[2.5]oct-5-ene	791	865	7.27	1.73	33.33	0.50	43	7.35	2.02	14.16	0.75

observed. Finally, *CR* profiles for the hits that passed the applied RM_1 threshold for each phase are provided in Fig. 5.6C,F (i.e., analytes in Tables 5.2 and 5.3), for alumina and silica, respectively. By careful observation, one will note that the CR_1 values for the hits in Tables 5.2 and 5.3 are slightly larger for some hits relative to the same given hit CR_1 value in the *CR* profiles in Fig. 5.6C,F. This is a result of using a tile size of 3 modulation on ^{1}D to capture the hit signal for the preparation of Fig. 5.6C,F, however the 1v1 analysis used to arrive at the hitlists in Tables 5.2 and 5.3 used a tile size of 5 modulations, thus the resulting RM_1 in the tables for

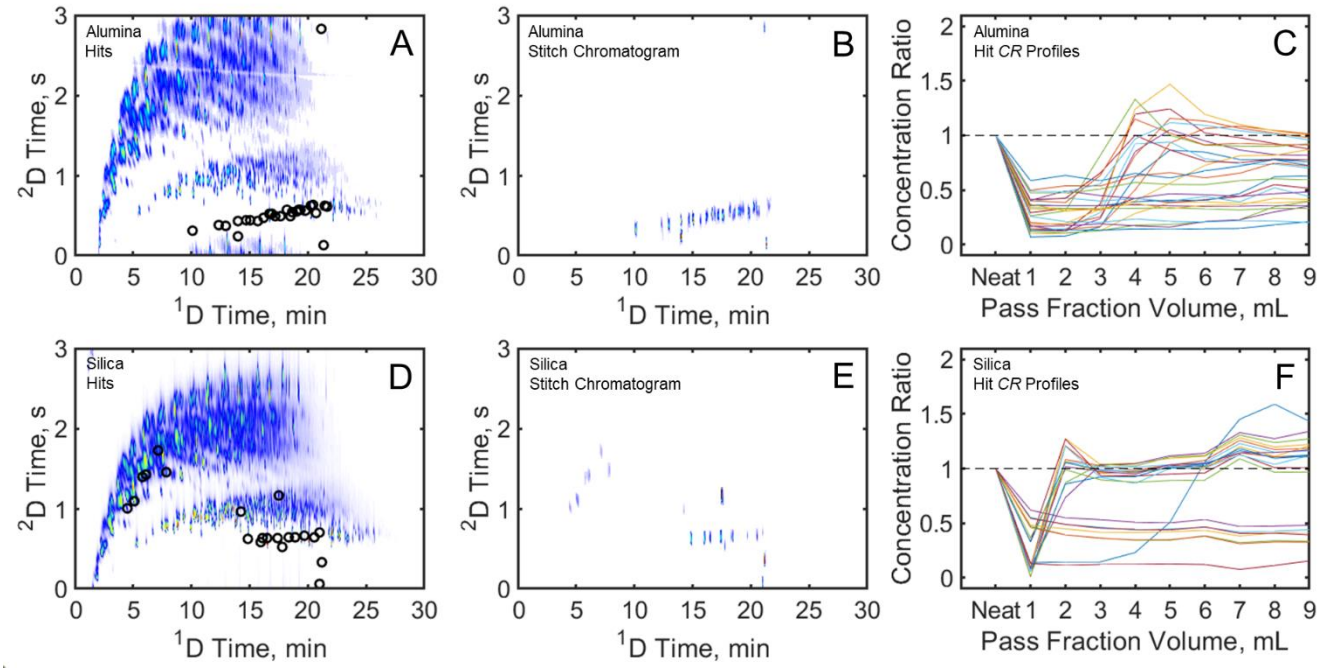


Figure 5.6. (A) The hits with a RM_1 > 33.33 % (CR_1 < 0.5) for the neat versus 1 mL pass fraction 1v1 analysis of the alumina data are plotted over the TIC GC×GC-TOFMS chromatogram of the alumina phase data. (B) These hits in (A) are then used to create the stitch chromatogram which allows the hits to be visualized free of interferences. (C) The CR profiles for the alumina hits show that all of the hits experience an initial decrease in concentration in the first pass fraction, being extracted by the alumina phase cartridge. (D) The hits with a RM_1 > 33.33 % (CR_1 < 0.5) in the silica hitlist are plotted over the TIC chromatogram of the silica phase data are provided, along with (E) the stitch chromatogram of these same hits. (F) Many of the CR profiles of the silica hits possess very different shapes compared to the alumina profiles, with the majority rising to a CR of 1 by a pass fraction volume of 2 mL after an initial decrease in the first mL pass fraction.

some hits may be lower. The larger tiles were used for hit discovery to ensure mitigation of any retention time shifting, but once discovered the hits were quantified using a smaller tile that has been centered on each hit peak on a per sample basis in order to optimize the quantification accuracy by minimizing potential inclusion of interference signal. Thus, the *CR* profiles reflect the effort to provide accurate quantification at the top RM_1 *m/z*.

Broadly comparing the *CR* profiles in Fig. 5.6C,F for alumina and silica we see that the polar compounds native in the JP8 jet fuel had stronger affinity for the alumina phase, because the earliest eluting compounds in the silica SPE data return to a *CR* of 1 by the second mL pass

fraction of fuel, whereas for the alumina extraction it takes until the fourth mL pass fraction to see any compound return to a *CR* of 1. Furthermore, the polar compounds appear to have a wider range of affinities for the silica phase, as evident by the wider range of compound breakthrough volumes. A common pattern seen in these *CR* profiles are for the CR_i to drop below 1 before increasing above 1, then dropping back to 1, which demonstrates and underpinning mechanism of analyte retention in SPE. This *CR* profile pattern suggests that the SPE cartridge has begun to reach its adsorbent capacity leading to the compounds to begin to elute off the cartridge, breaking through in the later pass fuel fractions. For some analytes, this breakthrough adds to the analyte concentration in the neat fuel, resulting in a rise in signal relative to the neat fuel causing the *CR* profile to increase over 1. In this instance, such analyte compounds may be initially strongly adsorbed (CR_i < to << 1), but then are being displaced by compounds with stronger affinity for the SPE phase which are being supplied by the influx of new sample to the cartridge. Analytes with weaker affinity naturally elute first followed by the next weakest, and so on. Thus, the breakthrough volume for each compound depends upon how much sample is introduced to the SPE cartridge, how many compounds from the samples are extractable by the selected stationary phase, and their stationary phase affinity relative to each other. To further evaluate the differences in retention behavior both by phase and compound we analyze the *CR* profiles individually using PCA, followed by augmenting the *CR* profiles from both phases prior to PCA.

Scores and loading plots of the PCA of the *CR* profiles for all analyte 28 hits shared between the alumina SPE and silica SPE hitlists are provided in Figure 5.7. These scores demonstrate that the compounds interacted with each phase to different degrees, as the profiles tend to separate by phase, with most of the silica SPE *CR* profiles clustered on the negative side of PC1 with the alumina SPE *CR* profiles spread out throughout the remainder of PC1. The PC1

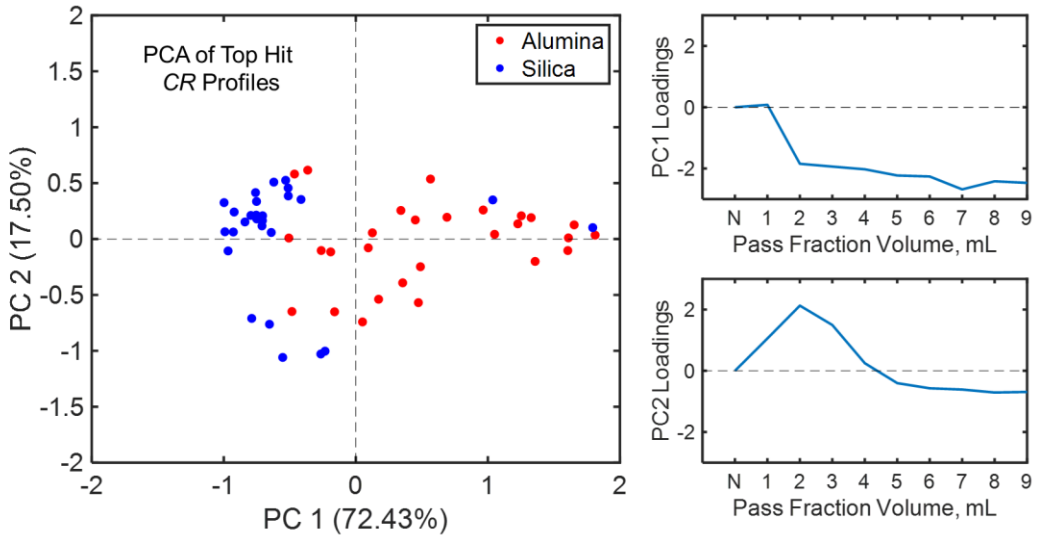


Figure 5.7. PCA scores plot of the alumina and silica CR profiles analyzed separately for hits present in both hitlists. The profiles generally cluster by extraction phase, with two "outlier" silica profiles. Loadings for PC1 and PC2 are plotted against the pass fraction volume and give insight to the features responsible for separating the phases.

loadings indicate that *CR* profiles which remain at low CR_i longer beyond the first mL pass fraction (CR_1) were scored positively on PC1, with the right most samples on PC1 having *CR* profiles which did not return to neat concentration levels and the left most returning to neat concentration levels right after CR_1. Conversely, PC2 captured variance in the extent to which each *CR* profile decreased in concentration relative to the neat concentration level, with the second mL pass fraction concentration (CR_2) weighting the effect most heavily. Thus, the PC2 loading indicates that the closer a given *CR* profile approaches zero between the first and third mL pass fractions, the more negative they are scored on PC2 (e.g. as in Figure 5.6C,F). Thus, the region on PC2 between -1 and -0.5 contains both alumina and silica *CR* profiles at similar PC1 scores, making these *CR* profiles the most similar between the two phases. However, because each analyte is represented twice in this PCA scores plot it is difficult to infer the differences in the stationary phase interactions between the analytes. To accomplish this, PCA can be carried

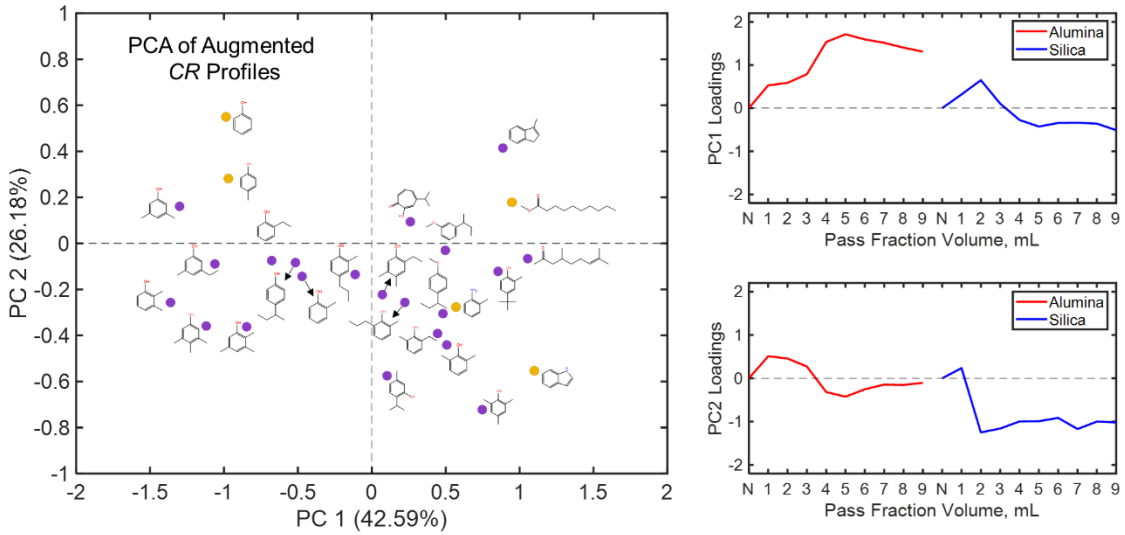


Figure 5.8. PCA scores plot of the alumina and silica CR profiles augmented together for hits present in both hitlists. Spike compounds are denoted with yellow dots while native compounds are denoted with purple dots. Structures have been provided for each dot for reference and ease of interpretation of the variation captured by the PCs. Two dots are left offscreen for clarity. Phenolic compounds tend to increase in alkyl substitution as the PC1 score increases. Loadings for PC1 and PC2, plotted against each phase pass fraction volume, respectively, are given for the augmented profiles to aid in description of the underlying reasons for the variance captured by each PC.

out on the augmented *CR* profiles, joining alumina CR profile and silica CR profile for a given analyte hit one after the other in a single vector.

Scores and loading plots of the PCA analysis of augmented alumina and silica *CR* profiles for all 28 hits shared between the alumina and silica hitlists are provided in Figure 5.8. Structures of each analyte have been added to provide molecular context for the PC separation. Based on the loadings vectors the variance captured by PC1 describes differences in alumina affinity, with affinity decreasing from left to right. Thus, the analytes on the far left of the scores plot have alumina *CR* profiles which decrease and remain at low CR_i for the duration of the extraction process, while the alumina *CR* profiles of analytes on the right side of the scores plot decrease and return to neat concentrations in the first few pass fractions. Conversely, the

variance captured by PC2 describes differences in silica affinity, with affinity increasing from bottom to top, where the silica *CR* profiles of the analytes in the negative PC2 scores space decrease and return to neat concentrations promptly while the analytes in the positive PC2 scores space remain at low CR_i longer. Note that two analytes are outside the limits of the plot, sulfolane and 8,8-dimethyl-4-methylene-1-oxaspiro[2.5]oct-5-ene, at PC coordinates of (0.43,2.18) and (0.32,1.65) respectively. Both analytes have silica *CR* profiles which decrease and do not return to neat concentrations.

In general, based upon the trends in the scores plot in Fig. 5.8 per the compound identification, the affinity for the alumina phase appears to correlate with a decrease in both an increase in the degree of alkyl substitution for the phenolic compounds and increase in stearic hindrance of the hydroxyl group. Additionally, there are interesting trends for analytes with similar degree of alkyl substitution. Take 2,3-dimethylphenol, the leftmost analyte at coordinates (PC1, PC2) of (-1.36, -0.26), and 2,6-dimethylphenol at coordinates (0.51, -0.44). Both of these compounds have the same degree of alkyl substitution, but 2,3-dimethylphenol has a high affinity for alumina while 2,6-dimethylphenol has lower affinity. This difference could possibly be due to the proximity of the methyl groups to the hydroxyl group. A similar trend is seen with the rest of the analytes, with the phenolic compounds on the right side of the scores plot possessing alkyl substitutions directly adjacent to the hydroxyl group while the hydroxyl groups of the compounds on the left side of the scores plot are unencumbered. Because the retention mechanism of the SPE cartridges is an adsorption of the molecules with the stationary phase surface the presence of alkyl substituents adjacent to the hydroxy group imparts a steric hindrance effect, limiting the affinity the molecule can exhibit toward the stationary phase.

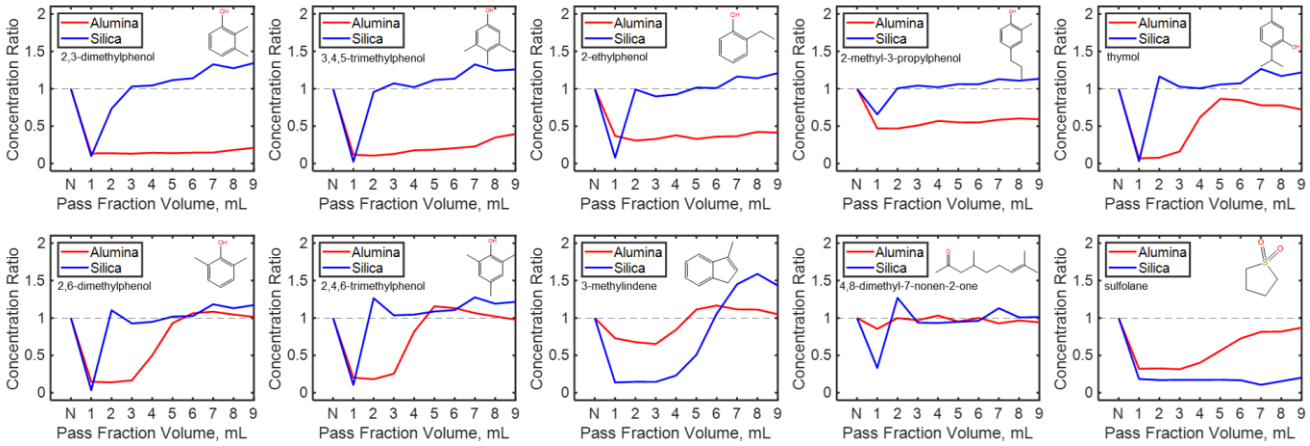


Figure 5.9. Plots directly comparing the CR profiles of 10 discovered analytes against pass fraction volume, chosen from across the PC1 space. Structures are provided for each analyte.

Plots directly comparing the extraction behavior of 10 compounds between alumina and silica phases, selected from across the PC space of the augmented profiles in the Fig. 5.8 scores plot, are shown in Fig. 5.9. The first four compounds (A-D) all extract strongly and remain extracted at a steady-state CR_i for the alumina while they begin to elute from the silica cartridges by the second mL of fuel (CR_2). These compounds were all selected from the left side of the PC1 space and represent compounds that have high affinity for the alumina phase and have unhindered hydroxyl groups. The next three compounds (E-G) display a lower affinity for the alumina phase than the preceding compounds, with their concentration returning to neat levels around halfway through the extraction. These compounds were selected from the right side of the PC1 space, and all have alkyl substituents adjacent to their hydroxyl groups which may contribute to their lower affinity with the stationary phase. The eighth compound (H), 3-methylindene, was well extracted and retained by both phases, though interestingly this was the one analyte discovered which did not contain a heteroatom. In both phases the affinity was sufficiently low that 3-methylindene began to rapidly elute partway through the extraction. The ninth compound (I), 4,8-dimethyl-7-nonen-2-one, is an example of a native compound which

was well extracted by the silica phase but not the alumina phase. This can be seen in the PCA scores plot in Fig. 5.8 at the far right of PC1, where the compounds with lowest affinity for alumina reside. Finally, the tenth compound (J), sulfolane, was extracted well by both phases but displayed higher affinity for the silica phase. Sulfolane is an industrial solvent commonly used for the desulfurization of hydrocarbon mixtures so its presence could be due to that use.

In the context of this JP-8 jet fuel the alumina phase SPE appears to be more effective at extracting the phenolic compounds than silica phase SPE. Because alumina is slightly basic it more strongly interacts with the slightly acidic phenols than the relatively more acidic silica phase. This does not necessarily mean silica is a poor choice for phenol extraction, but that the phenols have less affinity for silica than the late eluting compounds, tentatively identified as cyclothianes by their RMV, in silica SPE extract in Fig. 5.1D. The preference for these compounds, eluting between 23-30 min on ^{1}D, outcompeted the phenols, causes them to elute from the cartridge earlier than for alumina for the given sample volume. Had the extraction ended after passing the first mL of sample, the silica extract would have likely been very similar to the alumina extract, given most phenol compounds would still be adhering to the cartridge. Thus, given a sample whose polar constituents are unknown, it would be beneficial to perform extractions at several different volumes to explore the range of compounds extractable by the chosen SPE phase [37].

5.4. Conclusion

A workflow is provided for the discovery of analytes in a complex sample (JP8 jet fuel) subjected to dynamic SPE extraction using two phases (alumina and silica) following by tile-based 1v1 analysis of GC×GC-TOFMS. We demonstrate that the extractable analytes breakthrough at different rates and that their breakthrough volumes are due in part to analyte concentration in the sample, analyte affinity for the stationary phase, and the chemical complexity of the sample resulting in a competitive extraction environment. Furthermore, the extraction affinity of discovered compounds were compared directly, discovering that among phenols a higher degree of alkyl substitution and stearic hindrance led to decreased affinity for both SPE phases. Thus, it is important to consider the overall chemical environment of possible extractable and other interactive compounds to ensure your target compound classes are efficiently analyzed. More broadly, this research provides a platform for more comprehensively analyzing the trace compounds in complex samples by coupling SPE prior to GC×GC-TOFMS, with the 1v1 analysis being integral for the discovery and thorough analysis of the analytes with interesting extraction behavior, so the SPE can be thought of as another separation dimension providing additional chemical information.

5.5. References

[1] R.C. Striebich, J. Contreras, L.M. Balster, Z. West, L.M. Shafer, S. Zabarnick, Identification of polar species in aviation fuels using multidimensional gas chromatography-time of flight mass spectrometry, Energy and Fuels 23 (2009) 5474–5482. https://doi.org/10.1021/ef900386x.

[2] Z. Yang, B.P. Hollebone, Z. Wang, C. Yang, M. Landriault, Determination of polar impurities in biodiesels using solid-phase extraction and gas chromatography-mass spectrometry, J. Sep. Sci. 34 (2011) 409–421. https://doi.org/10.1002/JSSC.201000533.

[3] S.M. Rowland, W.K. Robbins, Y.E. Corilo, A.G. Marshall, R.P. Rodgers, Solid-phase extraction fractionation to extend the characterization of naphthenic acids in crude oil by electrospray ionization fourier transform ion cyclotron resonance mass spectrometry, Energy and Fuels 28 (2014) 5043–5048. https://doi.org/10.1021/EF5015023/ASSET/IMAGES/LARGE/EF-2014-015023_0008.JPEG.

[4] K. Pyrzynska, A. Kubiak, I. Wysocka, Application of solid phase extraction procedures for rare earth elements determination in environmental samples, Talanta 154 (2016) 15–22. https://doi.org/10.1016/J.TALANTA.2016.03.022.

[5] S. Ma, M. He, B. Chen, W. Deng, Q. Zheng, B. Hu, Magnetic solid phase extraction coupled with inductively coupled plasma mass spectrometry for the speciation of mercury in environmental water and human hair samples, Talanta 146 (2016) 93–99. https://doi.org/10.1016/J.TALANTA.2015.08.036.

[6] M.S. Shahriman, M.R. Ramachandran, N.N.M. Zain, S. Mohamad, N.S.A. Manan, S.M. Yaman, Polyaniline-dicationic ionic liquid coated with magnetic nanoparticles composite for magnetic solid phase extraction of polycyclic aromatic hydrocarbons in environmental samples, Talanta 178 (2018) 211–221. https://doi.org/10.1016/J.TALANTA.2017.09.023.

[7] E.J. Mercer, F. Halaweish, Determination of free glycerol in biodiesel via solid-phase extraction and spectrophotometric analysis, JAOCS, J. Am. Oil Chem. Soc. 88 (2011) 655–659. https://doi.org/10.1007/S11746-010-1707-6.

[8] S.L. Anjos, J.S. Almeida, L.S.G. Teixeira, A.C.M. da Silva, A.P. Santos, A.F.S. Queiroz, S.L.C. Ferreira, S. Mattedi, Determination of Cu, Ni, Mn and Zn in diesel oil samples using energy dispersive X-ray fluorescence spectrometry after solid phase extraction using sisal fiber, Talanta 225 (2021). https://doi.org/10.1016/J.TALANTA.2020.121910.

[9] M. Barroso, S. Costa, M. Dias, D.N. Vieira, J.A. Queiroz, M. López-Rivadulla, Analysis of phenylpiperazine-like stimulants in human hair as trimethylsilyl derivatives by gas chromatography–mass spectrometry, J. Chromatogr. A 1217 (2010) 6274–6280. https://doi.org/10.1016/J.CHROMA.2010.08.001.

[10] G. Milman, A.J. Barnes, R.H. Lowe, M.A. Huestis, Simultaneous quantification of cannabinoids and metabolites in oral fluid by two-dimensional gas chromatography mass spectrometry, J. Chromatogr. A 1217 (2010) 1513–1521. https://doi.org/10.1016/J.CHROMA.2009.12.053.

[11] T. Murakami, Y. Iwamuro, R. Ishimaru, S. Chinaka, H. Hasegawa, Molecularly imprinted polymer solid-phase extraction of synthetic cathinones from urine and whole blood samples, J. Sep. Sci. 41 (2018) 4506–4514. https://doi.org/10.1002/JSSC.201800874.

[12] L. Fernandez-Lopez, A. Luna-Maldonado, M. Falcon, L. Mastrobattista, J. Navarro-Zaragoza, R. Mancini, Development and validation of a gas chromatography–mass spectrometry method for opiates and cocaine in human bone, J. Pharm. Biomed. Anal. 164 (2019) 636–641. https://doi.org/10.1016/J.JPBA.2018.11.015.

[13] J. Ma, G. Wu, S. Li, W. Tan, X. Wang, J. Li, L. Chen, Magnetic solid-phase extraction of heterocyclic pesticides in environmental water samples using metal-organic frameworks coupled to high performance liquid chromatography determination, J. Chromatogr. A 1553 (2018) 57–66. https://doi.org/10.1016/J.CHROMA.2018.04.034.

[14] V. V. Lobodin, W.K. Robbins, J. Lu, R.P. Rodgers, Separation and Characterization of Reactive and Non-Reactive Sulfur in Petroleum and Its Fractions, Energy and Fuels 29 (2015) 6177–6186. https://doi.org/10.1021/acs.energyfuels.5b00780.

[15] M. Moors, D.L. Massart, R.D. McDowall, Analyte isolation by solid phase extraction (SPE) on silica-bonded phases: Classification and recommended practices, Pure Appl. Chem. 66 (1994) 277–304. https://doi.org/10.1351/pac199466020277.

[16] C.F. Poole, New trends in solid-phase extraction, TrAC Trends Anal. Chem. 22 (2003) 362–373. https://doi.org/10.1016/S0165-9936(03)00605-8.

[17] G.Z. Fang, J. Tan, X.P. Yan, An ion-imprinted functionalized silica gel sorbent prepared by a surface imprinting technique combined with a sol-gel process for selective solid-phase extraction of cadmium(II), Anal. Chem. 77 (2005) 1734–1739. https://doi.org/10.1021/AC048570V/ASSET/IMAGES/LARGE/AC048570VF00002.JPEG.

[18] A.K. Callesen, J.S. Madsen, W. Vach, T.A. Kruse, O. Mogensen, O.N. Jensen, Serum protein profiling by solid phase extraction and mass spectrometry: A future diagnostics tool?, Proteomics 9 (2009) 1428–1441. https://doi.org/10.1002/PMIC.200800382.

[19] 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. https://doi.org/10.1016/j.jchromb.2019.01.029.

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

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

[22] 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. https://doi.org/10.1016/j.chroma.2020.461401.

[23] C. Cordero, J. Kiefl, S.E. Reichenbach, C. Bicchi, Characterization of odorant patterns by comprehensive two-dimensional gas chromatography: A challenge in omic studies, TrAC Trends Anal. Chem. 113 (2019) 364–378. https://doi.org/10.1016/J.TRAC.2018.06.005.

[24] D.T. Bowman, K.J. Jobst, X. Ortiz, E.J. Reiner, L.A. Warren, B.E. McCarry, G.F. Slater, Improved coverage of naphthenic acid fraction compounds by comprehensive two-dimensional gas chromatography coupled with high resolution mass spectrometry, J. Chromatogr. A 1536 (2018) 88–95. https://doi.org/10.1016/J.CHROMA.2017.07.017.

[25] N. Di Giovanni, M.A. Meuwis, E. Louis, J.-F.F. Focant, Untargeted Serum Metabolic Profiling by Comprehensive Two-Dimensional Gas Chromatography-High-Resolution Time-of-Flight Mass Spectrometry, J. Proteome Res. 19 (2020) 1013–1028. https://doi.org/10.1021/acs.jproteome.9b00535.

[26] 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. https://doi.org/10.1016/j.forc.2018.01.001.

[27] G.S. Ochoa, M.C. Billingsley, R.E. Synovec, Using solid-phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: comparative analysis of aerospace fuel composition, Anal. Bioanal. Chem. (2022) 1–13. https://doi.org/10.1007/s00216-022-04348-1.

[28] S. Yuan, H. Li, Z. Liu, Y. Wang, L. Wang, X. Zhang, G. Liu, Measurement of non-hindered and hindered phenolic species in aviation fuels via tandem-SPE with comprehensive GC × GC–MS/FID, Fuel 287 (2021) 119561. https://doi.org/10.1016/j.fuel.2020.119561.

[29] L.M. Balster, S. Zabarnick, R.C. Striebich, L.M. Shafer, Z.J. West, Analysis of polar species in jet fuel and determination of their role in autoxidative deposit formation, Energy and Fuels 20 (2006) 2564–2571. https://doi.org/10.1021/ef060275l.

[30] 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. https://doi.org/10.1021/ACS.ANALCHEM.2C00223/ASSET/IMAGES/LARGE/AC2C00223_0008.JPEG.

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

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

[33] 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. https://doi.org/10.1021/acs.analchem.0c03456.

[34] 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. https://doi.org/10.1016/J.TALANTA.2021.122844.

[35] 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. https://doi.org/10.1016/j.chroma.2021.462092.

[36] 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. https://doi.org/10.1021/ACS.ANALCHEM.0C02136.

[37] U. Tilstam, Sulfolane: A versatile dipolar aprotic solvent, Org. Process Res. Dev. 16 (2012) 1273–1278. https://doi.org/10.1021/op300108w.

Chapter 6. Property correlation-driven chemometric feature selection: modeling thermal stability of aerospace fuels with GC×GC-TOFMS

6.1. Introduction

The presence of impurities (i.e., sulfur, olefins, oxygenates, and aromatics) in kerosene-based fuels has known impacts on the performance and system reliability [1–4]. For instance, the presence of these impurities in kerosene-based fuels can cause the decomposition of hydrocarbons into carbonaceous deposits (“coking”), which in turn, can effect the regenerative cooling in rocket and jet engine s [5–7]. It is generally accepted that there are three basic mechanisms governing the formation of carbon deposits [1–3,8]. These mechanisms include (a) production of condensation (amorphous) carbon via formation of polycyclic aromatic hydrocarbons (PAHs) in the hydrocarbon stream which condense on the tubing, catalyst, and/or reactor wall surfaces, (b) free radical growth reactions between the hydrocarbon stream and the surface coke, and (c) formation of filamentous carbon via catalytic reactions with surface metals found in the tubing, catalyst, and/or reactor wall surfaces. Given the detrimental impact of these impurities on fuel performance, the chemical composition of candidate fuels must be carefully assessed prior to use in in systems imposing high levels of thermal stress and surface reactivity.

Investigation into the thermal stability of fuels has taken a three-fold approach: (1) fuel composition [9], (2) cooling surface substrates (i.e., copper, stainless steel, etc.) [10–12], and (3) testing conditions (i.e., pressure and flow rates) [13].Compact Rapid Assessment of Fuel Thermal Integrity (CRAFTI) is a laboratory scale experiment that follows this three-fold appro ach, providing quantitative data under conditions relevant to rocket regenerative cooling systems [14–16]. Briefly, fuel flows through a copper test article, a section which is resistively heated (the heated zone) to temperatures that promote the onset of fuel chemical degradation. Thermal

(and catalytic) conditions at the cooling channel inner surface accelerate reactivity with fuel constituents, resulting in the formation of carbonaceous deposits in the heated and unheated downstream surfaces of the test article. The physical metrics obtained from CRAFTI analysis, and the subsequent temperature-programmed oxidation of carbonaceous test article deposits are (a) the change in test article pressure drop as function of time, and (b) the carbon deposition in mass as a function of test article position. Pressure change in the test article has been identified as a critical physical property since it may be indicative of significant deposition leading to local wall overheating and, in the worst case, catastrophic failure of a rocket engine. Coupling the information gained from CRAFTI analysis to analytical information about the chemical composition of these fuels offers an experimental platform that can interrelate the chemical and physical properties of these fuels.

Gas chromatography (GC) is a traditional analytical technique that is amenable for the separation and analysis of volatile and semi-volatile mixtures. When GC is coupled with mass spectrometry (MS), spectral information can be gathered allowing for further selectivity and the ability to identify chemical compounds. GC-MS has been shown to be a powerful tool for the analysis of kerosene-based fuel [17–21]. Comprehensive two-dimensional (2D) gas chromatography coupled with time-of-flight mass spectrometry (GC×GC–TOFMS) can further improve upon the separation power of one dimensional GC and provide additional insight into complex mixtures of volatile compounds such as those present in kerosene-based fuels [21–24]. With GC×GC, the two separation columns should provide “complementary” chemical selectivity. This means either the first separation dimension uses a non-polar stationary phase while the second dimension uses a polar stationary phase, which is referred to as a standard-column configuration, or alternatively, a reverse-column configuration (polar first dimension and

non-polar second dimension) has also been shown to provide good selectivity for petroleum based samples [24].Despite the obvious analytical advantages of applying GC×GC to the analysis of kerosene-based fuels, the large amount of information-rich data produced can be challenging to manually interpret. These challenges are further compounded when analyzing multiple injection replicates for a large set of fuels. Yet, the challenge of obtaining meaningful information can be overcome through application of powerful chemometric software methods that can aid in the interpretation of such complex data sets.

Non-targeted chemometric methods can be beneficial in explaining differences in fuel performance by providing insight into the underlying differences in chemical composition. Partial least squares (PLS) analysis is a chemometric method that associates the differences in measurable information for two different data sets [25].For example, PLS has been used to associate the chemical information obtained from GC×GC chromatograms to physical properties of kerosene-based fuels [26–31].Briefly, PLS analysis mathematically relates, via linear algebra, two data matrices (X- and Y-block) through extraction of factors referred to as latent variables (LVs). Using PLS, in the study herein, models are constructed to account for the variance (ideally, the relevant chemical differences) in both the GC×GC–TOFMS data for a fuel sample set, i.e., the signal intensities (which constitute the X-block) and the respective measured property values, for the same fuel sample set (which constitute the Y-block). PLS yields two important outcomes: (1) a linear correspondence of the chemical/physical properties to the GC×GC–TOFMS data, which can subsequently be used to predict chemical/physical properties without having to directly measure these properties in new samples, and (2) the underlying relationship between the chemical composition of the samples and the predicted chemical/physical measurements, provided by the linear regression vectors (LRVs).

However, PLS analysis can be computationally expensive and prone to prediction errors when performed on large data sets with a high degree of noisy and/or irrelevant variables [30]. Feature selection, also known as variable selection, can be beneficial in improving the performance of PLS models by removing these uninformative features from the data set. Recently, tile-based Fisher ratio (F-ratio) analysis is used as a feature selection method for GC×GC–TOFMS data [32,33]. This method calculates the F-ratio metric for the summed chromatographic signal within a small, rectangular section (i.e., tile) on a per-mass channel (*m/z*) basis [32,33].The tile-based algorithm offers two advantages for feature discovery: (1) mitigation of retention time misalignment across chromatograms and (2) enhancement of the signal-to-noise ratio [32,33].Tile-based F-ratio analysis has been beneficial in comparing the chemical composition of petroleum-based fuels [34–36].However, when performing a large-scale analysis of many candidate fuels, the experimental design may not be conducive to F-ratio analysis due to sample and/or time limitations. An alternative feature discovery approach, termed tile-based variance ranking, has recently been developed to discover compositional differences in large data sets without the need for multiple injection replicates per sample [37].Using the same tile-based algorithm originally developed for F-ratio analysis, tile-based variance ranking calculates the relative signal variance (RSD2) between samples [37].This algorithm showed promising results in discovering both non-native compounds spiked at different concentrations and larger-scale compositional differences between three jet fuels [37].

Using these analytical and computational methods, we aim to establish the relationship between the thermal stability of 30 kerosene-based fuels evaluated via CRAFTI analysis and their differences in chemical composition. The GC×GC-TOFMS analysis herein leverages the use of solid phase extraction (SPE), which allows for deviations in fuel chemistry – whether low-

level contaminants or bulk hydrocarbon types – to be quickly identified and correlated with the observed thermal performance behavior [36].Next, tile-based variance ranking is utilized to narrow down the chromatographic analysis to a key list of analyte (chemical features) with significant (meaningful) deviation between the fuels. The signal variances observed for each discovered analyte is then correlated with the measured physical properties of the fuels using two methods: (1) by individually regressing each feature against the physical property of interest [15,38], and (2) through the RReliefF machine learning algorithm [30].Ideally, this feature selection workflow will be beneficial in improving the PLS modeling of these fuels to ensure informative connections are made between the chemical information provided by GC×GC–TOFMS and the measured physical properties from thermal stressing in the CRAFTI apparatus.

6.2. Experimental

SPE Extraction

SPE cartridges with alumina and silica phases were used to isolate polar compounds from the fuels. Each cartridge was rinsed with 5 mL methanol followed by conditioning with 15 mL hexanes. The fuel samples were then diluted 1:10 with hexane and 8 mL were loaded at a flow rate of 1-3 mL/min. The cartridges were then rinsed with 15 mL of hexanes to wash non-polar compounds from the cartridge. Finally, the pre-concentrated polar compounds were eluted from the cartridge with 0.5 mL of methanol. The neat fuels and their corresponding SPE methanol extracts (alumina and silica) were analyzed by GC×GC–TOFMS.

GC×GC Analysis

To investigate the chemical composition of the fuels and fuel SPE extract samples, a GC×GC–TOFMS instrumental platform was used, consisting of an Agilent 7890N GC (Agilent Technologies, Palo Alto, CA, USA), and a Pegasus 4D BT with a thermal modulator (LECO, St. Joseph, MI, USA). Aliquots of the fuel samples were introduced to the GC×GC–TOFMS instrument via a 7683B auto-injector (Agilent Technologies, Palo Alto, CA, USA). The auto-injector was set to inject 0.5 µl of sample at a 100:1 split for the fuels and 30:1 split ratio for the fuel extracts, both at an inlet temperature of 275 °C. Prior to injection, HPLC grade acetone and hexane (Fisher Scientific) were used as solvent rinses. The primary GC×GC column (^{1}D) was a Rxi-17Sil MS: 24 m × 250 µm inner diameter (i.d.) × 0.25 µm film thickness, and the secondary GC×GC column (^{2}D) was a Rxi-1MS 2 m × 180 µm I.D. × 0.18 µm film thickness. Ultrahigh purity helium (Grade 5, 99.999%, Praxair, Seattle, WA, USA) was used as the carrier gas at a constant flow rate of 2.0 ml/min. The primary ^{1}D oven was held at 40 °C for 1.5 min before being ramped at 5 °C/min to 200 °C where it was held for 1 min. The secondary ^{2}D oven was held at a +25 °C offset relative to the primary oven and the modulator block was held at a +18 °C offset to the primary oven. The modulation period was 3 s (separation run time of the ^{2}D column) with 0.75 s hot and cold pulses for each stage. The transfer line was set to 285 °C and the ion source was 225 °C. Mass channels, *m/z*, 40-334 at unit resolution were collected with an ionization voltage of 70 eV at 100 spectra/s after a 120 s acquisition delay. Two replicates were collected for each fuel. A summary of the fuels is provided in Table 6.1. Note that while 38 fuels were analyzed by the CRAFTI platform, two fuels were excluded from PLS modeling due to contamination issues, a further six fuels were excluded due to insufficient available fuel sample

to perform SPE, leaving a total of 30 fuels in the current study. Thus, PLS modeling was performed on the remaining 30 fuels, as indicated in the second column of Table 6.1.

Fuel Thermal Performance in CRAFTI Apparatus

Previously, the thermal stability of 38 fuels was assessed using the CRAFTI experimental platform at standard run conditions, complemented by subsequent analysis of the types and amount of carbonaceous deposits via temperature-programmed oxidation (TPO) using an RC612 Carbon Determinator (LECO Corporation), as summarized in Table 6.1. Reduction of the data set to 30 fuels is described in Table 6.1. During the course of the experiment, the pressure drop ΔP, the difference between the test article inlet and outlet pressure, was measured as a function of time. The pressure data for repeat experimental tests were averaged together to obtain a single ΔP versus time vector for each fuel. To account for the variation in starting (run time = 0) ΔP values for each fuel, we determined the *change* in ΔP over the time period of 300 to 900 s, or $\Delta(\Delta P)$, defined as ΔP at 900 s minus the ΔP at 300 s. Note that the ΔP values used to calculate $\Delta(\Delta P)$ are the average of 10 s intervals surrounding 300 s, and just prior to and including 900s. Following each 900 s duration thermal integrity test, the test article was cut into 21 roughly equivalent length segments (~1.27 cm) for TPO analysis. In previous reports, carbon deposition was reported in counts as obtained directly from the Carbon Determinator measurements. To provide a chemically and physically more meaningful measurement, the signal in counts corresponding to the measurement of CO_2 was converted to carbon mass via calibration. For each type of carbon (determined for specified time ranges during which CO_2 was detected in the TPO analysis), the mass was normalized by the area of the corresponding test article segment to

yield mass/area for each segment of the test article. For each fuel analysis run, the normalized mass/area was calculated for each zone (i.e., heated and exit) by calculating the total mass deposited in the zone and dividing by the total area of the segments that comprise the zone. The two zones of the test article were defined as follows: the heated zone (segments 4-11), and the exit zone (segments 12-21). Multiple runs were averaged to provide a single mass/area value for each sample (for each form of carbon and test article zone). Table 6.1 provides ΔP values at 900 s, and Δ(ΔP) values, along with the deposit mass/area of three types of carbon (chemisorbed, amorphous, and filamentous) measured in the heated and exit zones of the test article via TPO analysis, respectively. We refer the reader to our previous reports for additional information regarding experimental conditions, instrumentation, and test article analysis along with the carbonaceous deposit measurements in counts [15,38].

Data Analysis

The GC×GC–TOFMS chromatograms were imported into MATLAB 2021a (MathWorks, Natick, MA) from ChromaTOF (LECO Corporation) as CDF files. To reduce modeling errors and discover chromatographic features associated with measures of thermal stability, a three-step feature selection protocol leveraging the property-chemical correlation of PLS was implemented. For the first step, the fuel and fuel extract chromatograms were reduced to features that have significant variation amongst the whole dataset using tile-based variance ranking, which is based on the four tile-grid framework of the tile-based F-ratio algorithm. For each dataset, a tile size of 4 modulations on ^{1}D and 30 spectra on ^{2}D was used alongside a cluster window size of 4 modulations on ^{1}D and 20 spectra on ^{2}D. A signal-to-noise ratio (*S/N*)

threshold of 30 was implemented to remove low signal false positive hits. On each tile and each *m/z* the RSD^2 was calculated, and redundant hits were removed. The final hitlist was populated with hits which had at least three *m/z* pass the *S/N* threshold and were ranked by value of the top RSD^2 *m/z*. This hitlist was then cleaned up to both remove false positives and readjust the ranking to account for inconsistent baseline between samples. First, an area of data equal to the tile size used for variance ranking was cut out around the top RSD^2 *m/z* for each hit and unfolded into a vector. These data were then baseline corrected by fitting a line between the first and last data points of each modulation and subtracting it from the data vector. Now that the data has been readjusted the RSD^2 is recalculated and the hitlist reranked. Lastly, a one-dimensional peak detection algorithm was run on each hit and every hit which did not have a peak apex present in the retention window was excluded from the hitlist. This procedure was run for each chromatographic dataset (neat, alumina extract, and silica extract), resulting in three final hitlists containing the regions of each dataset with high signal variation, cutting out superfluous noise and reducing the amount of data needed to model.

The second step of the feature selection protocol is to use the data from each hit's multiple *m/z* from the three hitlists to model the carbonaceous deposits and pressure data using PLS. PLS Toolbox 8.61 (Eigenvector Research Inc., Wenatchee, WA, USA) was used to build models for relating the mean-centered chromatographic features to the property data. PLS analysis was performed on the 30 fuel and fuel extract samples in Table 6.1, using Venetian blinds cross validation with 10 splits and a blind thickness of 1 (i.e. grouping together the replicates for each fuel). To determine the "goodness of fit" of the PLS models, the root mean square error of cross validation (RMSECV) was calculated as defined by

$$RMSECV = \left[\frac{1}{N} * \sum (y_{i,cv} - y_{i,meas})^2\right]^{0.5} \quad (6.1)$$

This analysis was performed for each hit and each *m/z* from each hitlist, modeling against each of the two properties individually, and recording an NRMSECV and slope of the calibration plot for each one. The slope was then plotted against the NRMSECV values, spreading the well correlated features from the poorly correlated features. Then, NRMSECV thresholds were set to retain the top 100-200 features (*m/z* from hits) with the lowest NRMSECV for modeling the given property. These thresholds were set for each hitlist (neat, alumina, and silica) for each property, resulting in 6 feature sets.

Finally, these 6 feature sets were ranked by their importance in describing the modeled properties using the RReliefF algorithm. RReliefF uses *k*-means to weight features based on their association with a property, giving features whose neighbors correlate to the property similarly high weights while giving features whose neighbors correlate to the property differently lower weights [39–41]. Then, PLS models were constructed using the 10 features with the highest weights, recording the NRMSECV, and iteratively adding more features and repeating the PLS process until models using all features were tested. An RReliefF weight threshold was then selected at the value which resulted in a minimum of the NRMSECV for each of the 6 feature sets. The resulting feature sets for the neat, alumina, and silica data were fused for each modeled property, resulting in two fused-feature sets. These fused sets were then used to produce a final PLS model for each modeled property, producing both an optimized NRMSECV and linear regression vectors from which the features relationship to the property could be investigated.

Table 6.1. Summary of the fuel properties measured via CRAFTI and the LECO RC612 Carbon Determinator. The change in pressure ΔP, at 900 s, and Δ(ΔP), the difference of ΔP at 900 s and 300 s, is measured in psi, and carbon measurements are provided in mass per unit area (μg/cm^2). The heated zone is defined as segments 4-11, and the exit zone is defined as segments 12-21, of the test article. Omitted samples are indicated in the Sample Number column: (a) omitted due to contamination, (b) omitted due to insufficient sample to perform SPE.

Sample Number	Sample Name	ΔP	Δ(ΔP)	Chemisorbed (Heated)	Chemisorbed (Exit)	Amorphous (Heated)	Amorphous (Exit)	Filamentous (Heated)	Filamentous (Exit)
1	YA2921HW10	61	2.7	5.9	13.6	11.7	45.9	2.2	3.4
2	BG1121GP04	139	54.3	397.2	16.7	33.2	106.1	6.7	4.3
3	GRC RP-1	68	4.0	4.7	9.7	14.4	58.0	4.0	3.7
4	WC0721HW01	62	3.1	3.7	7.1	6.7	24.2	1.5	2.0
5	LB073009-05	125	40.7	20.3	7.2	15.9	107.1	1.9	3.7
6	ZI1521HW10	65	5.0	3.7	8.2	7.0	27.5	2.0	3.6
7	CG0721HW10	60	3.4	-0.9	4.8	7.4	28.1	1.3	2.7
8	LB073009-08	117	39.5	53.1	14.6	28.3	151.3	6.9	5.4
9	BB0821HW10	56	3.6	1.1	7.1	8.2	40.2	1.9	2.8
10	LB080409-05	57	3.5	-0.1	9.1	10.7	57.2	1.9	4.0
11	ZI2621HW01	62	4.4	5.5	14.0	14.6	41.7	2.3	4.7
12	ZJ1321GP01	60	3.5	6.0	13.6	12.7	32.2	2.4	3.7
13[b]	RG3021LS06	55	1.4	6.1	13.0	83.4	96.9	241.0	3.9
14	RG3021LS05	68	6.8	8.6	13.1	16.6	83.3	4.2	4.8
15[b]	POSF 3327	80	14.2	20.5	12.8	11.8	65.0	2.5	3.7
16[b]	POSF 4765	76	8.0	7.0	6.6	9.7	41.3	2.5	2.4
17	LB073009-02	142	65.9	9.3	15.4	26.5	208.0	0.2	1.9
18[a]	B0112868	221	88.6	1294.0	43.4	67.1	287.9	43.5	4.0
19[b]	VI2621LS01	57	4.4	4.6	8.6	8.3	42.5	3.3	4.3
20[b]	DB131014	55	3.9	3.5	9.5	20.7	41.8	87.5	2.8
21	DC310925	64	3.1	23.8	24.6	38.2	81.2	666.0	4.5
22	DC310923	59	2.3	2.2	11.2	21.2	89.4	140.8	4.4
23	DB131013	52	3.8	3.2	7.7	6.9	27.4	2.1	2.9
24	DB131015	56	4.7	6.6	15.4	44.7	89.6	7.5	4.1
25[b]	CL031236	57	3.3	3.3	8.6	9.0	41.2	2.5	4.5
26[a]	CB1121HW10	181	84.9	305.5	17.9	69.4	81.8	103.7	4.1
27	EA130720	85	14.0	34.7	22.6	33.7	187.6	11.7	15.4
28	EB220705	58	3.6	5.1	15.0	6.7	45.7	1.2	-1.9
29	CHC JP-5	65	4.7	3.3	7.5	4.4	43.2	1.5	4.2
30	LB080409-01	60	3.7	5.6	11.7	11.7	69.8	3.1	3.7
31	LB073009-01	65	4.7	6.7	11.5	13.0	72.8	3.2	4.2
32	A0072256	68	4.5	8.6	12.2	13.0	76.0	2.9	3.4
33	CL11-3089	64	3.2	4.2	7.5	9.4	35.0	3.3	3.4
34	LB100413-40	61	2.1	4.1	11.2	172.4	53.4	47.7	3.9
35	LB073009-03	65	4.4	6.3	12.9	12.5	97.0	2.5	4.1
36	LB073009-10	63	3.9	5.3	9.7	6.7	61.1	1.4	2.2
37	SA1421LS03	58	4.9	2.7	7.6	8.4	31.7	2.8	2.1
38	ED060739	86	14.3	10.3	24.2	42.1	225.7	6.9	32.2

6.3. Results and Discussion

Chemical Composition

Using GC×GC–TOFMS in the reverse-column configuration, an excellent two-dimensional (2D) chromatographic separation of the compound classes (alkanes, cycloalkanes, and aromatics) was achieved. Figure 6.1 shows the 2D separation of two representative fuels: Sample 3: GRC RP-1 and Sample 8: LB073009-08. The alkanes are located from 2.0-3.0 s on the ^{2}D dimension, the cycloalkanes are located from 1.0-2.0 s on the ^{2}D dimension, and the aromatics are located from 0-1.0 s on ^{2}D. In each chromatogram, one can visually discern subclasses for the cycloalkanes and aromatic classes (i.e. monocyclics, dicyclics, tricyclics, mono-aromatics, and di-aromatics). Each fuel was extracted with alumina and silica phase SPE cartridges to extract and preconcentrate polar compounds. Figure 6.2 shows an example of a chromatogram resulting from each phase for Sample 8: LB073009-08. Several compounds have been pre-concentrated via SPE and now appear in the region between 0.3-1.2 s on ^{2}D. In principle, every hydrocarbon class (indeed every hydrocarbon) could be identified and quantified (utilizing retention indices and mass-to-charge ratios, *m/z*). However, in order to create compositional descriptions of the fuels, we have elected to approach the chemical analysis of the fuels using a PLS-based approach. The identification of the general elution times for the three classes (alkanes, cycloalkanes, aromatics) in the 2D chromatographic space serves a more instructive purpose of correlating general fuel chemical composition with differences in measured performance behavior. The broad range of fuels exhibited large compositional variation leading to their various physical properties and thermal stability.

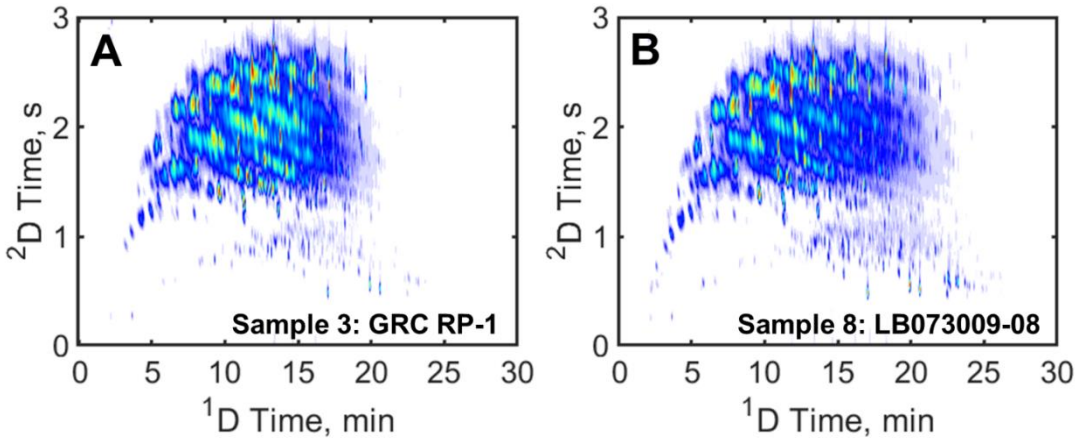


Figure 6.1. Two total ion current (TIC) GC×GC chromatograms representing examples of a well-behaving fuel and a poorly behaving fuel in terms of CRAFTI performance. (A) Sample 3: GRC RP-1 (well behaving), (B) Sample 8: LB073009-08 (poor behaving).

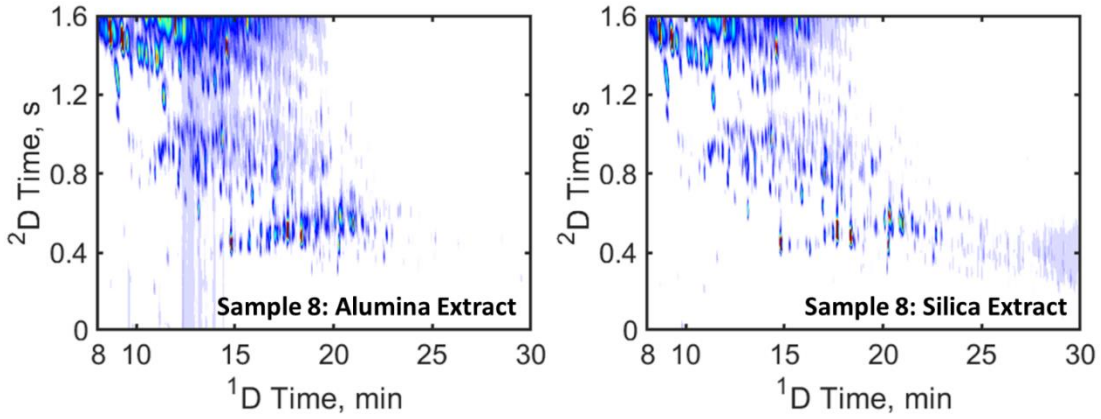


Figure 6.2. Two TIC GC×GC chromatograms of the alumina and silica extracts of Sample 8: LB073009-08, a poorly behaving fuel. The chromatograms have been limited to the polar-eluting region between 8 to 30 min on 1D and 0 to 1.6 s on 2D.

Fuel Thermal Performance

The time region of 300 s to 900 s was chosen to calculate Δ(ΔP) because the ΔP of the well-behaving fuels reaches pseudo-steady state at roughly 300 s, producing Δ(ΔP) values approaching 0. In contrast, the ΔP of the poorly behaving fuels continues to increase, resulting in higher Δ(ΔP) values. Figure 6.2 shows CRAFTI test article pressure drop profiles as a function of experimental time of the 30 fuels. The relationship between Δ(ΔP) and ΔP measured at 900 s as reported in the previous JANNAF report is provided in Figure 6.3. The data in Figure 6.3 indicates that four fuels have a relatively large Δ(ΔP), Samples 2, 5, 8, and 17, with an additional four fuels that have a moderate Δ(ΔP) relative to the rest of the fuels that have relatively small Δ(ΔP). Figure 6.4 shows the 2D TPO data, detailing the deposition of the various types of carbon and location for two different fuels. Sample 3: GRC RP-1 deposits a relatively small amount of chemisorbed carbon in the heated zone and modest amount of amorphous carbon in the exit zone (ACE), while Sample 8: LB073007-08 deposits a relatively significant amount of chemisorbed carbon, which was concentrated in the heated zone, and more amorphous carbon in the exit zone. Sample 3 is representative of fuels that perform "well" in terms of yielding a low Δ(ΔP) and depositing a limited amount of carbon, while Sample 8 is representative of fuels that perform "poorly" yielding a large Δ(ΔP) and depositing relatively high amounts of carbon. This suggests that these forms of carbon may contribute to the poor fuel thermal performance. The total mass of amorphous and chemisorbed carbon deposited was determined by summing the carbon counts within each section for the times corresponding to the different allotropes. These sections were then summed to represent carbon from the heated zone and exit zone. Initially, PLS modeling

was used to predict ACE and Δ(ΔP) from the CRAFTI study utilizing the entire GC×GC-TOFMS chromatograms of the 30 fuels. These models provided relatively high NRMSECV and subsequent efforts to implement feature selection improved the NRMSECV for Δ(ΔP) [38]. However, these models were only marginally informative as the models only identified aromatic compounds as the main contributors whereas heteroatom containing polar compounds, which are considered to be the main driver of carbonaceous deposits, were not represented [42]. Thus, to improve this aspect the ACE and Δ(ΔP) were modeled using a data fusion strategy, joining the neat fuel GC×GC–TOFMS data to the SPE extraction GC×GC–TOFMS data, in an effort to unravel polar compound contributions to these properties (Figure 6.5). An improved feature selection approach was implemented, applying a tile-based variance ranking initial step to preselect features of the dataset which vary significantly between samples, eliminating superfluous data. An NRMSECV threshold was then set for each property, after which the threshold passing features importance in describing the properties was ranked by the RreliefF algorithm. From these ranked features a final PLS model can be created, yielding linear regression vectors, describing how each feature contributes to predicting the modeled property.

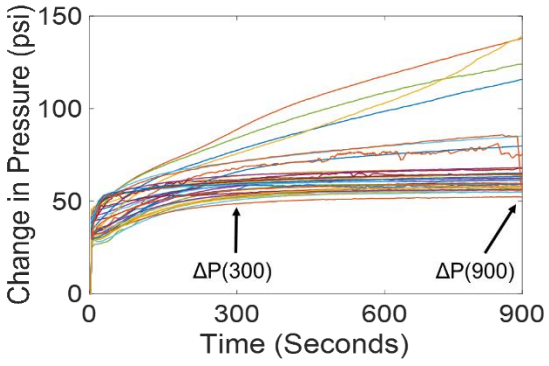


Figure 6.3. Pressure vectors for the fuels obtained from the CRAFTI analysis. Pressure drop as a function of time is shown, where ΔP is defined as the pressure of the inlet minus pressure of the outlet. A corrected change in pressure, Δ(ΔP), was calculated by taking the ΔP at 900 s minus ΔP at 300 s to account for variation in starting ΔP between fuels.

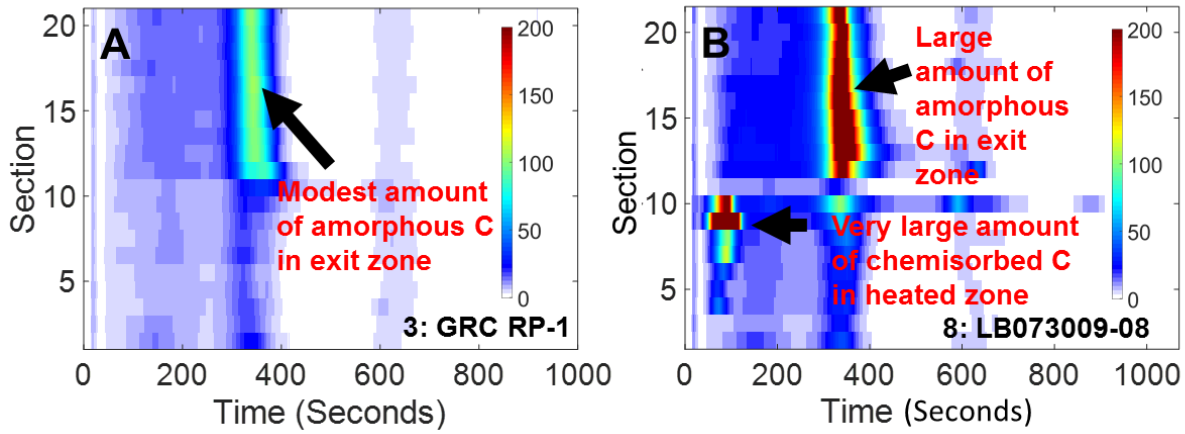


Figure 6.4. Instrument data from Temperature-Programmed Oxidation (TPO) is shown for two fuels that deposit different amounts of carbon. (A) Carbon deposition for fuel GRC RP-1 is representative of fuels that exhibit low pressure changes. These fuels deposit a relatively small amount of chemisorbed and amorphous carbon. (B) Carbon deposition for fuel LB073007-08 is representative of fuels that perform poorly. These fuels deposit a relatively large amount of chemisorbed carbon in the heated zone (segments 4-11) as well as more amorphous carbon in the exit zone (segments 12-21).

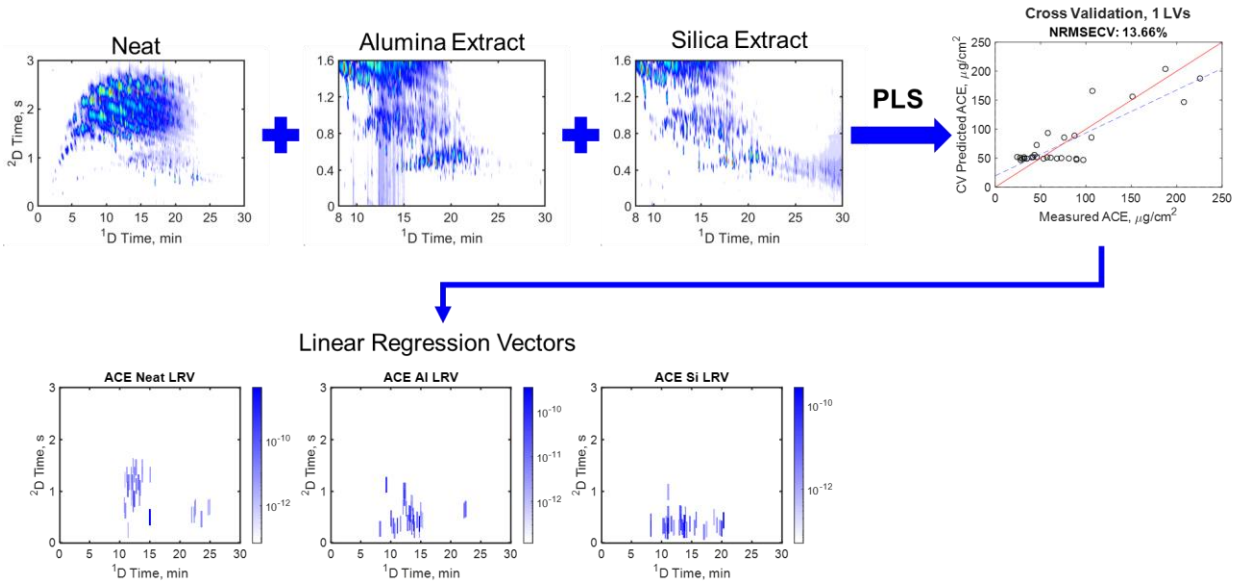


Figure 6.5. Schematic showing the strategy for data fusion of the chromatographic data for input into PLS. Linear regression vectors (LRVs) can then be obtained from the final PLS model representing the chromatographic features which are important for predicting the modeled property.

Figure 6.6A shows the PLS cross validated regression using the feature selection results for the amorphous carbon in the exit zone (ACE) data. Initial PLS models utilizing the whole chromatogram and data fusion without feature selection gave an NRMSECV of 22.73%. The NRMSECV has been improved to 13.66% demonstrating that feature selection is effective at focusing upon features which are closely related to the modeled property, eliminating the noise contribution that irrelevant data adds to the model. More valuable than the accuracy of the prediction is the PLS model linear regression vector (LRV), which describes how each feature contributes to the property. By utilizing the retention times of these feature, we can create "stitch" chromatograms which give a visual interpretation of the LRVs, allowing for compound classes to be inferred based on the 2D chromatogram structure [36]. Figure 6.6B shows the 2D LRV for the neat fuel features that contribute to ACE modeling. The features tend to cluster in the aromatic region, the majority of which are alkyl-substituted monoaromatics and a few diaromatics such as naphthalene and acenaphthene (Table 6.2). In contrast, the LRVs for the alumina and silica extracts in Figure 6.6C and 6.6D pull out sets of features which are unique from those found in the neat. These compounds elute in a broad band in a more polar region than the aromatic compounds. Indeed, several of the compounds from the alumina LRV in Figure 6.6C were identified as oxygen containing compounds, such as 2,5-dimethylphenol and benzyl alcohol (Table 6.2). Several partial aromatic indene compounds were also identified through the alumina LRV. The silica extract LRV selected similar compounds that the alumina selected such as m-cresol and 2-benzenemethanol, but it also uniquely selected 2.3-dihydrobenzofuran and quinoline, a nitrogen containing compound (Table 6.2). These plots demonstrate that the strategy of SPE data fusion with the neat fuel data uniquely allowed for polar compounds as contributors to the ACE deposit phenomenon. Several more tentative compound identity assignments from

the ACE LRVs are listed in Table 6.2 alongside their spectral match values to the NIST library and *m/z* selected by the feature selection method. Notice that in Figure 6.6A only a subset of samples are modeled well while several of the samples form a flat line. This occurs because the selected polar compound features are not present in every sample and thus do not describe the deposit formation formed in their CRAFTI experiments. It can be reasoned then that higher concentrations of polar compounds contribute to an increase in amorphous carbon deposition in the exit zone, however they are not the sole drivers. Further investigation on the subset of samples is warranted to unravel this unexplained variance more deeply.

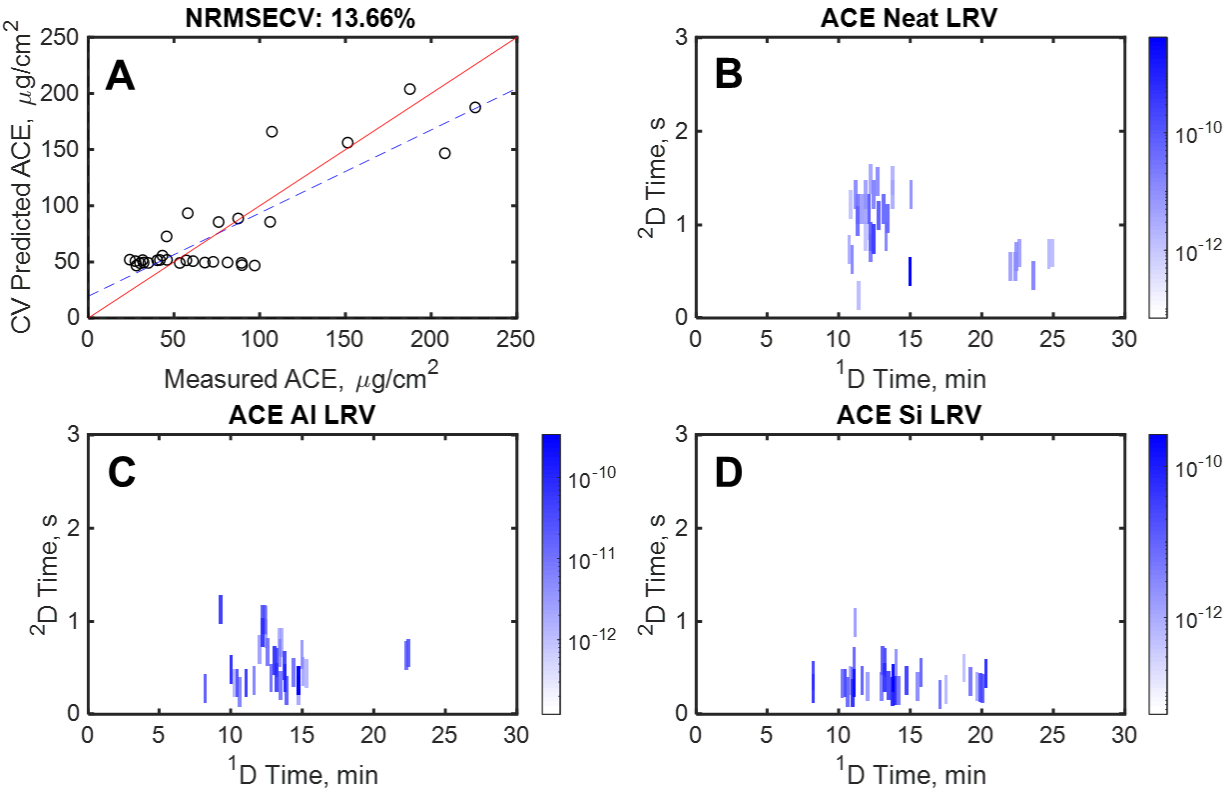


Figure 6.6. (A) The PLS prediction of the amorphous carbon in the exit zone using each chromatographic feature which was selected by the multi-step feature selection method. (B-D) LRV chromatogram displaying the chromatographic location and regression magnitude for the features from the neat samples (B), alumina extracts (C), and silica extracts (D) which are important for modeling amorphous carbon in the exit zone.

Figure 6.7A shows the PLS cross validated regression using the feature selection results for the Δ(ΔP) data. Like the PLS modeling for ACE, Δ(ΔP) was initially modeled with whole chromatograms giving a NRMSECV of 18.44% which has been improved to 12.68%. Like for ACE, the regression performs well for describing the increase in Δ(ΔP) for the samples with the highest values which makes the LRVs important for discovering the compounds that contribute to an increased pressure drop. Interestingly, the LRVs in Figures 6.7B-6.7D capture similar yet distinct features which describe Δ(ΔP) when compared to ACE. For example, the neat LRV in Figure 6.7B captures dimethylnaphthalenes, i.e., the same as ACE, however, it also selects higher alkyl-substituted naphthalenes, different forms of diaromatics like biphenyl and dibenzofuran, an oxygen containing compound, which is unique to the Δ(ΔP) neat LRV (Table 6.3). The alumina LRV for Δ(ΔP) selected a smaller range of features than ACE, but very similar composition, selecting several higher alkyl-substituted phenols, such as p-isopropylphenol and p-sec-butylphenol, as contributors to explaining Δ(ΔP) (Table 6.3). In contrast, the silica extract LRV was unique in that despite extracting similar compounds as the alumina extract the feature selection method isolated a completely different set of compounds. A nitrogen containing compound, isoquinoline, and dihydro methyl napthalenes were identified as important for describing Δ(ΔP) (Table 6.3). Further tentative compound identities from the Δ(ΔP) LRVs are provided in Table 6.3 with their spectral match values. From the range of compounds identified by the feature selection in Tables 6.2 and 6.3 it appears that the main drivers of amorphous carbon deposition and pressure increase in the CRAFTI experiment is aromatic content, particularly di- and highly substituted aromatics, and heteroatom containing aromatic compounds. These results offer insights for what compounds to target and monitor in order to gauge a given fuel's propensity for passing or failing thermal performance testing.

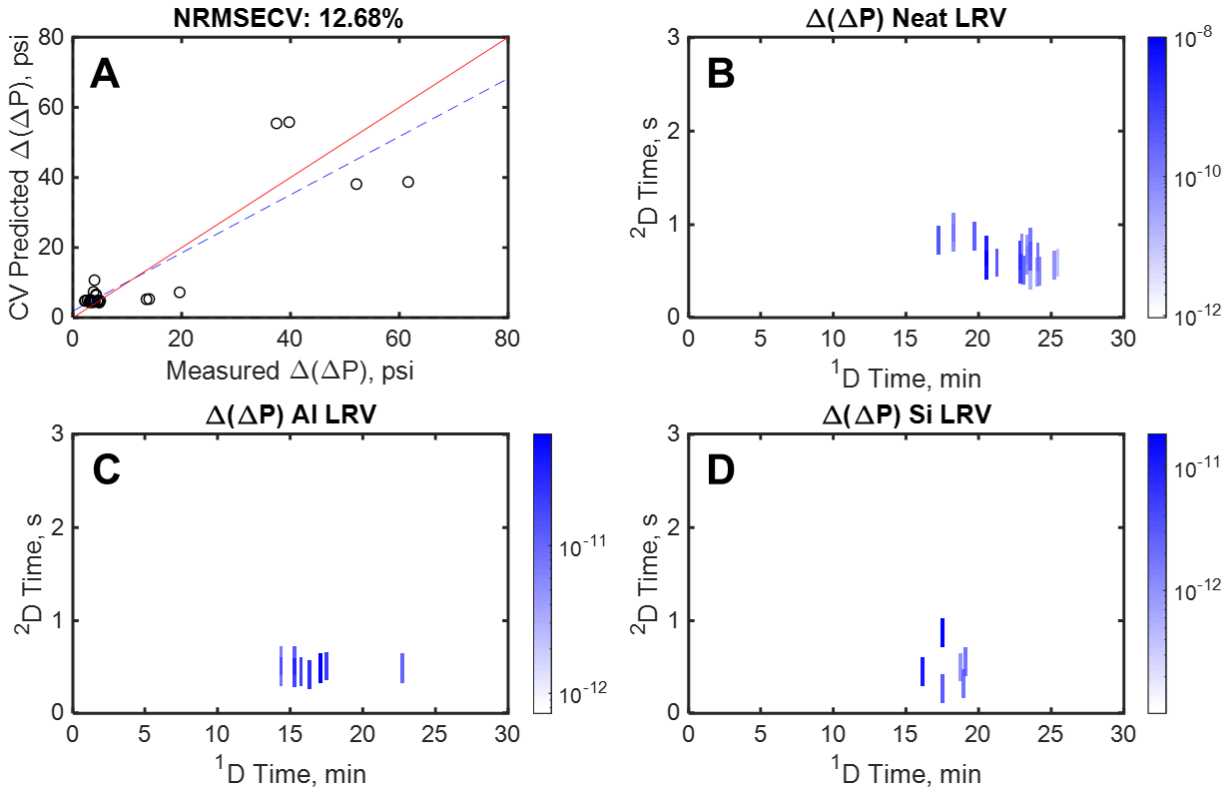


Figure 6.7. (A) The PLS prediction of the Δ(ΔP) using each chromatographic feature which was selected by the multi-step feature selection method. (B-D) LRV chromatogram displaying the chromatographic location and regression magnitude for the features from the neat samples (B), alumina extracts (C), and silica extracts (D) which are important for modeling Δ(ΔP).

Table 6.2. List of compounds tentatively identified by the feature selection method for modeling ACE. Ten compounds are provided per dataset are provided alongside their spectral match values and the m/z identified by the feature selection. Match value (MV) and reverse match value (RMV) for the analyte mass spectrum relative to a compound library, ranging from 0-1000, where 1000 is a perfect match.

Sample Set	Name	[1]t_r, min	[2]t_r, s	RSD[2]	*m/z* selected	MV	RMV
Neat	naphthalene	14.95	0.49	3.82	64	921	939
	acenaphthene	23.55	0.45	3.19	152	922	955
	1,4-dimethylnaphthalene	22.3	0.54	4.73	89	902	950
	1,7-dimethylnaphthalene	21.95	0.54	4.93	118	913	939
	1,2,4,5-tetramethylbenzene	12.4	0.84	4.41	134	869	897
	m-diisopropylbenzene	11.3	1.16	2.45	141	893	909
	2,3-dihydro-4-methyl-1H-indene	12.2	0.86	3.49	121	857	883
	3-methyl-1H-indene	10.9	0.62	10.27	52	893	904
	1,4-dimethyladamantane	12.4	1.31	0.28	134	863	936
	4-ethenyl-1,2-dimethylbenzene	10.7	0.73	8.36	50	884	959
Alumina	indene	10	0.47	11.58	115	909	928
	2,5-dimethylphenol	14.7	0.36	5.98	128	899	915
	m-cresol	11.05	0.33	7.85	104	836	903
	2-phenyl-2-propanol	11.6	0.36	0.96	118	812	833
	benzyl alcohol	10.6	0.23	1.79	108	880	909
	2-ethylphenol	12.8	0.38	5.92	122	905	911
	2-methylbenzenemethanol	13.85	0.25	0.69	107	793	847
	3-methyl-1H-indene	13.05	0.57	12.09	122	922	944
	2,3-dihydro-4-methyl-1H-indene	12.5	0.66	8.98	130	908	909
	2,3-dihydro-1,6-dimethyl-1H-indene	13.5	0.76	3.75	144	883	883
Silica	m-cresol	11.05	0.33	7.65	104	848	906
	1-naphthalenol	19.85	0.28	7.10	144	772	864
	benzyl methyl ketone	13.7	0.24	0.28	107	878	886
	2-methylbenzenemethanol	13.85	0.25	0.69	107	818	843
	2,3-dihydrobenzofuran	15.45	0.29	2.23	120	838	857
	indano[2,1-d]1,3-dioxane	20	0.27	4.78	146	841	856
	1,2,3,4-tetrahydronaphthalene	13.05	0.57	8.73	107	898	906
	1,2-epoxyindan	17.05	0.21	3.32	104	917	939
	quinoline	17.45	0.26	1.96	130	882	889
	3-methyl-2-hexanone	10.75	0.36	0.51	107	835	848

Table 6.3. List of compounds tentatively identified by the feature selection method for modeling Δ(ΔP). Ten compounds are provided per dataset are provided alongside their spectral match values and the m/z identified by the feature selection. Datasets with fewer compounds had less than ten unique compounds identified by the feature selection. Match value (MV) and reverse match value (RMV) for the analyte mass spectrum relative to a compound library, ranging from 0-1000, where 1000 is a perfect match.

Sample set	Name	$^{1}t_r$, min	$^{2}t_r$, s	RSD[2]	*m/z* selected	MV	RMV
Neat	2-isopropylnaphthalene	22.8	0.66	4.70	154	862	875
	1,7-dimethylnaphthalene	20.5	0.56	18.02	156	902	931
	3-methyl-1,1'-biphenyl	22.85	0.52	5.10	76	929	947
	2,3,6-trimethylnaphthalene	23.55	0.45	3.19	170	892	903
	cyclohexyl-benzene	17.2	0.82	7.50	104	882	908
	1,6,7-trimethylnaphthalene	23.55	0.65	5.36	170	886	904
	1-methyl-3-(phenylmethyl)-benzene	24	0.48	2.77	170	911	915
	dibenzofuran	24.15	0.49	5.28	170	876	913
	1-cyclohexyl-3-methylbenzene	19.65	0.86	4.13	118	871	878
	3,3'-dimethylbiphenyl	25.15	0.56	3.94	152	890	906
Alumina	3-methyl-1,1'-biphenyl	22.7	0.48	4.79	168	911	920
	p-isopropylphenol	15.25	0.43	17.83	121	887	899
	o-isopropylphenol	14.35	0.44	17.87	121	890	896
	3-ethyl-5-methylphenol	15.7	0.44	10.05	121	875	890
	thymol	17.05	0.48	8.89	135	889	914
	2-ethyl-5-methylphenol	16.3	0.41	13.68	121	885	915
	p-sec-butylphenol	17.45	0.5	10.86	121	875	894
Silica	isoquinoline	17.45	0.26	1.96	129	811	887
	3-hydroxypropanoic acid 1 -butyl ester	17.45	0.86	2.03	89	734	837
	1,2-dihydro-3-methylnaphthalene	18.9	0.31	4.12	144	807	853
	1,2-dihydro-6-methylnaphthalene	19.05	0.55	4.67	144	807	857
	3-propylphenol	16.1	0.44	11.70	107	922	957

6.4. Conclusions

The overarching goal of this research is to better understand and enhance fuel performance through chemometric modeling by relating fuel property data (ASTM measurements) and thermal performance (CRAFTI data) to chemical composition data (GC×GC–TOFMS). The predictive PLS modeling approach presented herein may ultimately be used to educate fuel selection. A feature selection method based on PLS analysis was demonstrated in the context of fuel thermal stability (test article pressure drop increase and carbon deposition data) in relation to chemical information present in GC×GC–TOFMS chromatograms. Using a subset of 30 fuels and selected features, PLS modeling was successful in predicting fuel properties based on chemical composition and determining chemical species responsible for large changes in test article pressure over time and carbonaceous deposition, particularly of amorphous carbon in the exit zone. This information can be used to further tailor chemical composition of kerosene-based rocket fuels to achieve optimal fuel performance. Future work should focus on investigating chemical compound identification that correlates to each physical property.

6.5. References

[1] F.D. Kopinke, G. Zimmermann, S. Nowak, On the mechanism of coke formation in steam cracking-conclusions from results obtained by tracer experiments, Carbon N. Y. 26 (1988) 117–124. https://doi.org/10.1016/0008-6223(88)90027-9.

[2] F.D. Kopinke, G. Zimmermann, G.C. Reyniers, G.F. Froment, Relative Rates of Coke Formation from Hydrocarbons in Steam Cracking of Naphtha. 3. Aromatic Hydrocarbons, Ind. Eng. Chem. Res. 32 (1993) 2620–2625. https://doi.org/10.1021/ie00023a027.

[3] M.-F.F.S.G. Reyniers, G.F. Froment, Influence of Metal Surface and Sulfur Addition on Coke Deposition in the Thermal Cracking of Hydrocarbons, Ind. Eng. Chem. Res. 34 (1995) 773–785. https://doi.org/10.1021/ie00042a009.

[4] J. Wang, M.F. Reyniers, G.B. Marin, Influence of dimethyl disulfide on coke formation during steam cracking of hydrocarbons, Ind. Eng. Chem. Res. 46 (2007) 4134–4148. https://doi.org/10.1021/ie061096u.

[5] O. Altin, S. Eser, Analysis of Carboneceous Deposits from Thermal Stressing of a JP-8 Fuel on Superalloy Foils in a Flow Reactor, Ind. Eng. Chem. Res. 40 (2001) 589–595. https://doi.org/10.1021/ie0004489.

[6] S. Eser, R. Venkataraman, O. Altin, Deposition of carbonaceous solids on different substrates from thermal stressing of JP-8 and jet A fuels, Ind. Eng. Chem. Res. 45 (2006) 8946–8955. https://doi.org/10.1021/ie060968p.

[7] L. Guozhu, H. Yongjin, W. Li, Z. Xiangwen, M. Zhentao, Solid deposits from thermal stressing of n - Dodecane and chinese RP-3 jet fuel in the presence of several initiators, Energy and Fuels. 23 (2009) 356–365. https://doi.org/10.1021/ef800657z.

[8] O. Altin, S. Eser, Carbon deposition from thermal stressing of petroleum fuels, in: Am. Chem. Soc. Div. Fuel Chem., 2004: pp. 764–766. https://www.researchgate.net/publication/267420915 (accessed June 28, 2021).

[9] M.J. DeWitt, T. Edwards, L. Shafer, D. Brooks, R. Striebich, S.P. Bagley, M.J. Wornat, Effect of Aviation Fuel Type on Pyrolytic Reactivity and Deposition Propensity under Supercritical Conditions, Ind. Eng. Chem. Res. 50 (2011) 10434–10451. https://doi.org/10.1021/IE200257B.

[10] R. Venkataraman, S. Eser, Characterization of solid deposits formed from jet fuel degradation under pyrolytic conditions: Metal sulfides, Ind. Eng. Chem. Res. 47 (2008) 9351–9360. https://doi.org/10.1021/ie801007r.

[11] A.R. Mohan, S. Eser, Analysis of carbonaceous solid deposits from thermal oxidative stressing of Jet-A fuel on iron- and nickel-based alloy surfaces, Ind. Eng. Chem. Res. 49 (2010) 2722–2730. https://doi.org/10.1021/ie901283r.

[12] S. Tang, N. Shi, J. Wang, A. Tang, Comparison of the anti-coking performance of CVD TiN, TiO2 and TiC coatings for hydrocarbon fuel pyrolysis, Ceram. Int. 43 (2017) 3818–3823. https://doi.org/10.1016/j.ceramint.2016.12.036.

[13] B. Jin, K. Jing, J. Liu, X. Zhang, G. Liu, Pyrolysis and coking of endothermic hydrocarbon fuel in regenerative cooling channel under different pressures, J. Anal. Appl. Pyrolysis. 125 (2017) 117–126. https://doi.org/10.1016/j.jaap.2017.04.010.

[14] R.E. Synovec, C.E. Freye, B.A. Parsons, M.C. Billingsley, N. Keim, B. Hill-Lam, J.C. Wilhelm, Recent Advances in the Development of Chemical Analysis Tools to Relate Compositional Variation to Thermal Integrity Data for RP-1, RP-2, and Related Fuels, in: JANNAF 8th Liq. Propuls. Meet., Nashville, TN, 2015.

[15] R.E. Synovec, C.E. Freye, M.C. Billingsley, N. Keim, B. Hill-Lam, Recent Advances in the Development of Chemical Analysis Tools to Relate Compositional Variation to Thermal Integrity Data for RP-1, RP-2, and Related Fuels, in: JANNAF 9th Liq. Propuls. Meet. Phoenix, AZ, 2016.

[16] R.E. Synovec, C.E. Freye, M.C. Billingsley, N. Keim, B. Hill-Lam, Recent Advances in Relating

Chemical Compositional Variation in RP-1, RP-2, and Similar Fuels to Thermal Integrity Data, in: JANNAF 10th Liq. Propuls. Meet. Long Beach, CA, 2018: p. Abstract 2018-0001CV.
[17] T.M. Lovestead, B.C. Windom, J.R. Riggs, C. Nickell, T.J. Bruno, Assessment of the compositional variability of RP-1 and RP-2 with the advanced distillation curve approach, Energy and Fuels. 24 (2010) 5611–5623. https://doi.org/10.1021/ef100994w.
[18] N.J. Begue, J.A. Cramer, C. Von Bargen, K.M. Myers, K.J. Johnson, R.E. Morris, Automated method for determining hydrocarbon distributions in mobility fuels, Energy and Fuels. 25 (2011) 1617–1623. https://doi.org/10.1021/ef101635a.
[19] R. V. Gough, T.J. Bruno, Composition-explicit distillation curves of alternative turbine fuels, Energy and Fuels. 27 (2013) 294–302. https://doi.org/10.1021/ef3016848.
[20] P.Y. Hsieh, K.R. Abel, T.J. Bruno, Analysis of marine diesel fuel with the advanced distillation curve method, Energy and Fuels. 27 (2013) 804–810. https://doi.org/10.1021/ef3020525.
[21] 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. 1132 (2020) 157–186. https://doi.org/10.1016/j.aca.2020.07.027.
[22] M.K. Jennerwein, M. Eschner, T. Gröger, T. Wilharm, R. Zimmermann, Complete group-type quantification of petroleum middle distillates based on comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC-TOFMS) and visual basic scripting, Energy and Fuels. 28 (2014) 5670–5681. https://doi.org/10.1021/ef501247h.
[23] M.K. Jennerwein, A.C. Sutherland, M. Eschner, T. Gröger, T. Wilharm, R. Zimmermann, Quantitative analysis of modern fuels derived from middle distillates – The impact of diverse compositions on standard methods evaluated by an offline hyphenation of HPLC-refractive index detection with GC×GC-TOFMS, Fuel. 187 (2017) 16–25. https://doi.org/10.1016/j.fuel.2016.09.033.
[24] 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. https://doi.org/10.1016/J.FUEL.2018.07.081.
[25] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: A basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.
[26] B.J. Prazen, K.J. Johnson, A. Weber, R.E. Synovec, Two-dimensional gas chromatography and trilinear partial least squares for the quantitative analysis of aromatic and naphthene content in naphtha, Anal. Chem. 73 (2001) 5677–5682. https://doi.org/10.1021/ac010637g.
[27] B. Kehimkar, J.C. Hoggard, L.C. Marney, M.C. Billingsley, C.G. Fraga, T.J. Bruno, R.E. Synovec, Correlation of rocket propulsion fuel properties with chemical composition using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry followed by partial least squares regression analysis, J. Chromatogr. A. 1327 (2014) 132–140. https://doi.org/10.1016/J.CHROMA.2013.12.060.
[28] B. Kehimkar, B.A. Parsons, J.C. Hoggard, M.C. Billingsley, T.J. Bruno, R.E. Synovec, Modeling RP-1 fuel advanced distillation data using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry and partial least squares analysis, (n.d.). https://doi.org/10.1007/s00216-014-8233-6.
[29] C.E. Freye, B.D. Fitz, M.C. Billingsley, R.E. Synovec, Partial least squares analysis of rocket propulsion fuel data using diaphragm valve-based comprehensive two-dimensional gas chromatography coupled with flame ionization detection, Talanta. 153 (2016) 203–210. https://doi.org/10.1016/j.talanta.2016.03.016.
[30] V. Abrahamsson, N. Ristic, K. Franz, K. Van Geem, Comprehensive two-dimensional gas chromatography in combination with pixel-based analysis for fouling tendency prediction, J. Chromatogr. A. 1501 (2017) 89–98. https://doi.org/10.1016/j.chroma.2017.04.021.
[31] 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. https://doi.org/10.1021/acs.energyfuels.9b04108.

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

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

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

[35] 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. https://doi.org/10.1021/acs.analchem.2c01549.

[36] G.S. Ochoa, M.C. Billingsley, R.E. Synovec, Using solid-phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: comparative analysis of aerospace fuel composition, Anal. Bioanal. Chem. (2022) 1–13. https://doi.org/10.1007/s00216-022-04348-1.

[37] P.E. Sudol, G.S. Ochoa, C.N. Cain, R.E. Synovec, Tile-based variance rank initiated-unsupervised sample indexing for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, Anal. Chim. Acta. 1209 (2022) 339847. https://doi.org/10.1016/j.aca.2022.339847.

[38] R.E. Synovec, K.L. Berrier, C.E. Freye, M.C. Billingsley, N. Keim, B. Hill-Lam, A. Bishop, Improvements to GC×GC-TOFMS Composition-Based Models for Hydrocarbon Fuel Thermal Integrity, in: JANNAF 11th Liq. Propuls. Meet., Tampa, FL, 2019.

[39] M. Robnik-Šikonja, I. Kononenko, An adaptation of Relief for attribute estimation in regression, in: Mach. Learn. Proc. Fourteenth Int. Conf., 1997: pp. 296–304. https://www.researchgate.net/publication/2635627 (accessed January 11, 2023).

[40] M. Robnik-ˇ Sikonja, I. Kononenko, Theoretical and Empirical Analysis of ReliefF and RReliefF, 53 (2003) 23–69.

[41] K.C. Patchava, M. Benaissa, H. Behairy, Improving the prediction performance of PLSR using RReliefF and FSD for the quantitative analysis of glucose in Near Infrared spectra, Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS. 2015-November (2015) 2379–2382. https://doi.org/10.1109/EMBC.2015.7318872.

[42] E. Alborzi, P. Gadsby, M.S. Ismail, A. Sheikhansari, M.R. Dwyer, A.J.H.M. Meijer, S.G. Blakey, M. Pourkashanian, Comparative Study of the Effect of Fuel Deoxygenation and Polar Species Removal on Jet Fuel Surface Deposition, Energy and Fuels. 33 (2019) 1825–1836. https://doi.org/10.1021/acs.energyfuels.8b03468.

Chapter 7. Conclusion and future directions

Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection (GC×GC-TOFMS) is a powerful and versatile tool for the analysis of volatile samples, a particular example of interest being aerospace fuels. Chemometric analysis is indispensable for efficiently extracting relevant information from the chemically and data rich GC×GC-TOFMS datasets, however, the unique and creative application of chemometric results remains an active area of research. The goal of this dissertation then is to disseminate the furthering development of supervised chemometric methods and application of the results to facilitate improved analyte discovery, identification, and quantification. Thus, the work presented in this dissertation will ideally be utilized as tools and methods to facilitate the comprehensive analysis of unique chemical and biological applications, such as fuels, environmental samples, and metabolomics studies.

7.1. Chapter 2 Summary and Future Directions

In Chapter 2, a method for determining the selectivity of *m/z* using tile-based F-ratio analysis results for GC×GC-TOFMS data was explored. A dataset of diesel, spiked with 15 non-native compounds at 10, 20, 40, and 80 ppm, was analyzed with tile-based F-ratio analysis comparing the 10 vs 20 ppm samples, 20 vs 40 ppm, and 40 vs 80 ppm samples. A unique metric, the lack-of-fit (*LOF*), was used to compare the consistency of peak shapes for discovered F-ratio hits between classes. The results demonstrated that by combining appropriate thresholds for the p-value of discovered hits alongside *LOF* thresholds the purest *m/z* may be obtained which are ideal for quantification as evident by the low deviation in signal ratio (S-ratio) from the expected values. Furthermore, the use of combinatorial null distribution analysis (CNDA) proved to be a useful technique to set an appropriate threshold to infer *m/z* purity when compared

to the statistical thresholds. The method then may aid in the quantification of subtle concentration differences between sample classes, by giving increased confidence in the selectivity of the discovered *m/z*.

The current study was performed investigating S-ratios of 2 to aid in the method development, however there may exist a minimum detectable S-ratio through this method. Therefore, it would be interesting to look further into this method with spiked native compounds whose S-ratios could vary between 1 and 2, to find this limit. Furthermore, the variance of samples in this dataset were injection limited, which necessarily aids in improving both the F-ratio and p-value between classes. Testing this method on a metabolomics dataset with more complex sources of variance would help find the limits of the method.

7.2. Chapter 3 Summary and Future Directions

In Chapter 3, the methodology from Chapter 2 used to infer *m/z* purity was instead used to identify invariant *m/z* between classes to aid in obtaining pure mass spectra of class distinguishing compounds. Tile-based F-ratio was used to discover the hits in a dataset of JP-8 jet fuel spiked with sulfur-containing compounds at 30 and 15 ppm. Then p-value and *LOF* thresholds were applied to select *m/z* among the hits that do not change, allowing normalization ratios to be obtained for informed subtraction of the classes' spectra, obtaining the purified spectra for the given hit in a method referred to as class comparison enabled mass spectrum purification (CCE-MSP). The quality of the spectra were generally superior to those obtained with other signal decomposition algorithms. This method, combined with the method outlined in Chapter 2, gives both the tools to confidently identify and quantify analytes of interest using the results from tile-based F-ratio analysis.

The present study is not without its limitations. The investigated sulfur compounds were relatively unabated by overlapping compounds, resulting in few instances to stress test the CCE-MSP method. Furthermore, the method works best for chromatographic regions changing by themselves as two or more changing compounds results in combined purified spectra. Thus, like for the method from Chapter 2, it would benefit future studies to test against non-contrived datasets such as a metabolomics dataset where the variations are more complex and may offer a wider range of resolutions and relative signal intensities to test. This line of experimentation may lead to further improvements in the workflow of CCE-MSP to handle fringe cases.

7.3. Chapter 4 Summary and Future Directions

In Chapter 4 tile-based F-ratio was used to discover polar compounds and facilitate comprehensive fuel comparisons via solid phase extraction (SPE). Jet and rocket fuel samples were passed through a SPE cartridge, after which it was compared against the original sample to discover all compounds extracted by the cartridge. By using F-ratio to compare the neat and pass samples between fuel types hitlists detailing both the high-concentration (bulk hydrocarbon) and low-concentration (SPE-extracted) compounds may be obtained. This process aids exceptionally in batch comparison, allowing new formulations to be assessed comparatively against a reference fuel without the need to run every standard method for regular fuel assessment. Furthermore, the SPE pass fuel comparison approach serves as a demonstration of applying F-ratio to discover the changes brought on by a physical/chemical alteration to a sample.

This study used one extraction phase, silica, to assess the polar compound environment in fuels, which may be an incomplete picture. In future studies of the fuels, it would be beneficial to apply more varied extraction phases to find the optimal conditions for general polar compound extraction. The more promising point of the chapter, applying a physical/chemical change to a

sample and gaining full assessment of its effects, is worthy of further application study. For example, applying an electrochemical reaction using a gold electrode in fuels may serve for a selective reaction of sulfur-containing compounds. For batch comparison it would be beneficial to study a larger set of fuels with complementary performance data to assess whether bad fuels could be reliably isolated based on the GC×GC-TOFMS data alone.

7.4. Chapter 5 Summary and Future Directions

In Chapter 5, the retention behavior of polar compounds in JP-8 jet fuel was explored for two extraction phases, alumina and silica, by extracting the fuels with SPE cartridges and isolating fractions of the pass fuel for sequential analysis. Tile-based 1v1 analysis is then applied to the neat and pass fractions to discover the compounds that are changing in concentration between samples. Plotting the concentration of these compounds as a function of extraction volume gives insight to each compound's residence on the SPE cartridges. For the polar compounds present in JP-8 jet fuel, primarily alkylated phenols, alumina was found to retain them more strongly than silica, which isolated different compound classes preferentially. Furthermore, among the phenols the affinity for the stationary phase was found to be influenced by stearic effects of the molecules. This research gives insight into the appropriate selection of stationary phases for polar compound extraction for fuels and serves as a proof-of-principle for applying tile-based 1v1 analysis for the evaluation of a chemical process.

Future studies of polar compounds in fuels may include testing further extraction phases, such as fluorosil, zirconia, etc., to fully assess the extraction environment. This data should then be used to supplement chemometric modeling efforts (e.g., partial least-squares) to understand their effects on fuel properties and thermal performance. These models will aid not only in identifying poorly behaved fuels but also for learning the acceptable tolerance of such

compounds in fuels more definitively. Additionally, the sequential analysis framework should be tested with new applications, such as the continuous sampling of a chemical reaction by GC×GC with a valve-based injection system. Furthermore, the approach could be applied to a continuous process for the determination of in-specification versus out-of-specification samples which could populate chemometric models.

7.5. Chapter 6 Summary and Future Directions

In Chapter 6, a dataset of rocket fuels was analyzed with GC×GC-TOFMS to model the fuel's carbonaceous deposits (coke) and change in pressure in regenerative cooling lines and to identify the compounds associated. Partial least-squares (PLS) was used in conjunction with tile-based variance ranking and RreliefF to select the features from the dataset that most closely correlate with the metrics of thermal integrity. The approach was successful in reducing the modeling error obtained, yielding good association between measured and predicted property values. The linear regression vectors (LRV) were then investigated, and the discovered compounds identified, revealing an increasing linear relationship between compound concentration and degree of coking. The discovered compounds were identified as oxygenated and nitrogenated compounds, all present in low concentrations. Stitch chromatograms were utilized for the visualization of these LRVs which helps to interpret the information in relation to chromatogram retention times. These results give us great insight to contributing factors to poor thermal performance to monitor in fuels and the feature selection method as a unique way to combine supervised regression with untargeted feature discovery.

This study is not without its shortcomings, for example, the fuel dataset is relatively limited, with the fuels properties unevenly distributed over the range of property values. This means the fuels at the highest end of the range are given more weight for error reduction and

makes the model more limited in scope. Furthermore, the model for amorphous carbon in the exit zone displayed several fuels that held no correlation, because they did not contain the selected features that were prominent among the samples with the highest amounts of coke deposition. Thus, for future studies it would be beneficial to perform modeling of this sample subset separately and combine the results later with the original model to more fully model the fuels in the dataset. Expanding the fuels dataset will also be a priority, though this avenue may be difficult as the number of poor performing rocket fuels is generally low due to high manufacturing standards.

Bibliography

A. Bele, An overview on thin layer chromatography, (2011). www.ijpsr.com (accessed December 29, 2022).

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

A.C. Beckstrom, E.M. Humston, L.R. Snyder, R.E. Synovec, S.E. Juul, Application of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry method to identify potential biomarkers of perinatal asphyxia in a non-human primate model, J. Chromatogr. A 1218 (2011) 1899–1906. https://doi.org/10.1016/J.CHROMA.2011.01.086.

A.C. Jimenez, C.A. Heist, M. Navaei, C. Yeago, K. Roy, Longitudinal two-dimensional gas chromatography mass spectrometry as a non-destructive at-line monitoring tool during cell manufacturing identifies volatile features correlative to cell product quality, Cytotherapy 24 (2022) 1136–1147. https://doi.org/10.1016/J.JCYT.2022.06.001.

A.K. Callesen, J.S. Madsen, W. Vach, T.A. Kruse, O. Mogensen, O.N. Jensen, Serum protein profiling by solid phase extraction and mass spectrometry: A future diagnostics tool?, Proteomics 9 (2009) 1428–1441. https://doi.org/10.1002/PMIC.200800382.

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

A.P. Thorne, Fourier Transform Spectrometry in the Ultraviolet, Anal. Chem. 63 (1991) 57 A-65 A.

A.R. Mohan, S. Eser, Analysis of carbonaceous solid deposits from thermal oxidative stressing of Jet-A fuel on iron- and nickel-based alloy surfaces, Ind. Eng. Chem. Res. 49 (2010) 2722–2730. https://doi.org/10.1021/ie901283r.

B. Beaver, L. Gao, C. Burgess-Clifford, M. Sobkowiak, On the mechanisms of formation of thermal oxidative deposits in jet fuels. Are unified mechanisms possible for both storage and thermal oxidative deposit formation for middle distillate fuels?, Energy and Fuels 19 (2005) 1574–1579. https://doi.org/10.1021/ef040090j.

B. Bennett, S.R. Larter, Quantitative separation of aliphatic and aromatic hydrocarbons using silver ion-silica solid-phase extraction, Anal. Chem. 72 (2000) 1039–1044. https://doi.org/10.1021/ac9910482.

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

B. Jin, K. Jing, J. Liu, X. Zhang, G. Liu, Pyrolysis and coking of endothermic hydrocarbon fuel in regenerative cooling channel under different pressures, J. Anal. Appl. Pyrolysis 125 (2017) 117–126. https://doi.org/10.1016/j.jaap.2017.04.010.

B. Kehimkar, B.A. Parsons, J.C. Hoggard, M.C. Billingsley, T.J. Bruno, R.E. Synovec, Modeling RP-1 fuel advanced distillation data using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry and partial least squares analysis, Anal. Bioanal. Chem. 407 (2015) 321–330. https://doi.org/10.1007/s00216-014-8233-6.

B. Kehimkar, J.C. Hoggard, L.C. Marney, M.C. Billingsley, C.G. Fraga, T.J. Bruno, R.E. Synovec, Correlation of rocket propulsion fuel properties with chemical composition using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry followed by partial least squares regression analysis, J. Chromatogr. A 1327 (2014) 132–140. https://doi.org/10.1016/J.CHROMA.2013.12.060.

B. Omais, M. Courtiade, N. Charon, D. Thiébaut, A. Quignard, M.C. Hennion, Investigating comprehensive two-dimensional gas chromatography conditions to optimize the separation of oxygenated compounds in a direct coal liquefaction middle distillate, J. Chromatogr. A 1218 (2011) 3233–3240. https://doi.org/10.1016/J.CHROMA.2010.12.049.

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. https://doi.org/10.1016/j.chroma.2017.10.013.

B. Worley, S. Halouska, R. Powers, Utilities for quantifying separation in PCA/PLS-DA scores plots, Anal. Biochem. 433 (2013) 102–104. https://doi.org/10.1016/J.AB.2012.10.011.

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

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

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

B.J. Prazen, K.J. Johnson, A. Weber, R.E. Synovec, Two-dimensional gas chromatography and trilinear partial least squares for the quantitative analysis of aromatic and naphthene content in naphtha, Anal. Chem. 73 (2001) 5677–5682. https://doi.org/10.1021/ac010637g.

B.L. Atwater, R. Venkataraghavan, F.W. McLafferty, Matching of Mixture Mass Spectra by Subtraction of Reference Spectra, Anal. Chem. 51 (1979) 1945–1949. https://pubs.acs.org/doi/pdf/10.1021/ac50048a013 (accessed August 13, 2020).

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

C. Cordero, J. Kiefl, S.E. Reichenbach, C. Bicchi, Characterization of odorant patterns by comprehensive two-dimensional gas chromatography: A challenge in omic studies, TrAC Trends Anal. Chem. 113 (2019) 364–378. https://doi.org/10.1016/J.TRAC.2018.06.005.

C.E. Freye, B.D. Fitz, M.C. Billingsley, R.E. Synovec, Partial least squares analysis of rocket propulsion fuel data using diaphragm valve-based comprehensive two-dimensional gas chromatography coupled with flame ionization detection, Talanta 153 (2016) 203–210. https://doi.org/10.1016/j.talanta.2016.03.016.

C.F. Poole, Ionization-based detectors for gas chromatography, J. Chromatogr. A 1421 (2015) 137–153. https://doi.org/10.1016/J.CHROMA.2015.02.061.

C.F. Poole, New trends in solid-phase extraction, TrAC Trends Anal. Chem. 22 (2003) 362–373. https://doi.org/10.1016/S0165-9936(03)00605-8.

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

C.N. Cain, P.E. Sudol, K.L. Berrier, R.E. Synovec, Development of variance rank initiated-unsupervised sample indexing for gas chromatography-mass spectrometry analysis, Talanta 233 (2021) 122495. https://doi.org/10.1016/J.TALANTA.2021.122495.

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. https://doi.org/10.1021/ACS.ANALCHEM.0C02136.

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. https://doi.org/10.1021/acs.analchem.2c00223.

ChromaTOF® Tile Analytical Software, (n.d.). https://www.leco.com/product/chromatof-tile (accessed May 6, 2021).

ChromCompare+, (n.d.). https://www.sepsolve.com/chromcompare/ (accessed May 6, 2021).

D. Guillarme, D.T.T. Nguyen, S. Rudaz, J.L. Veuthey, Recent developments in liquid chromatography—Impact on qualitative and quantitative performance, J. Chromatogr. A 1149 (2007) 20–29. https://doi.org/10.1016/J.CHROMA.2006.11.014.

D. Jones, C.E. West, A.G. Scarlett, R.A. Frank, S.J. Rowland, Isolation and estimation of the “aromatic” naphthenic acid content of an oil sands process-affected water extract, J. Chromatogr. A 1247 (2012) 171–175. https://doi.org/10.1016/j.chroma.2012.05.073.

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

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. https://doi.org/10.1021/AC7024317/ASSET/IMAGES/LARGE/AC7024317F00005.JPEG.

D.C. Harris, Statistics, in: Quant. Chem. Analyisis, Ninth, W.H. Freeman and Company, New York, 2016: pp. 64–89.

D.J. Cookson, C. Paul Lloyd, B.E. Smith, Investigation of the Chemical Basis of Diesel Fuel Properties, Energy and Fuels 2 (1988) 854–860. https://doi.org/10.1021/ef00012a021.

D.J. Cookson, J.L. Latten, I.M. Shaw, B.E. Smith, Property-composition relationships for diesel and kerosene fuels, Fuel 64 (1985) 509–519. https://doi.org/10.1016/0016-2361(85)90086-9.

D.J. Cookson, P. Iliopoulos, B.E. Smith, Composition-property relations for jet and diesel fuels of variable boiling range, Fuel 74 (1995) 70–78. https://doi.org/10.1016/0016-2361(94)P4333-W.

D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity deviation ratio: A new metric for chemometric analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data, Anal. Chim. Acta 871 (2015) 66–76. https://doi.org/10.1016/J.ACA.2015.02.040.

D.R. Deans, Use of heart cutting in gas chromatography: A review, J. Chromatogr. A 203 (1981) 19–28. https://doi.org/10.1016/S0021-9673(00)80278-2.

D.T. Bowman, K.J. Jobst, X. Ortiz, E.J. Reiner, L.A. Warren, B.E. McCarry, G.F. Slater, Improved coverage of naphthenic acid fraction compounds by comprehensive two-dimensional gas chromatography coupled with high resolution mass spectrometry, J. Chromatogr. A 1536 (2018) 88–95. https://doi.org/10.1016/J.CHROMA.2017.07.017.

E. Alborzi, P. Gadsby, M.S. Ismail, A. Sheikhansari, M.R. Dwyer, A.J.H.M. Meijer, S.G. Blakey, M. Pourkashanian, Comparative Study of the Effect of Fuel Deoxygenation and Polar Species Removal on Jet Fuel Surface Deposition, Energy and Fuels 33 (2019) 1825–1836. https://doi.org/10.1021/acs.energyfuels.8b03468.

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

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

E.J. Mercer, F. Halaweish, Determination of free glycerol in biodiesel via solid-phase extraction and spectrophotometric analysis, JAOCS, J. Am. Oil Chem. Soc. 88 (2011) 655–659. https://doi.org/10.1007/S11746-010-1707-6.

F. Dieterle, A. Ross, tz Schlotterbeck, H. Senn, F. Hoffman-, Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1 H NMR Metabonomics, (2006). https://doi.org/10.1021/ac051632c.

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

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

F. Rey-Stolle, D. Dudzik, C. Gonzalez-Riano, M. Fernández-García, V. Alonso-Herranz, D. Rojo, C. Barbas, A. García, Low and high resolution gas chromatography-mass spectrometry for untargeted metabolomics: A tutorial, Anal. Chim. Acta 1210 (2022) 339043. https://doi.org/10.1016/J.ACA.2021.339043.

F.D. Kopinke, G. Zimmermann, G.C. Reyniers, G.F. Froment, Relative Rates of Coke Formation from Hydrocarbons in Steam Cracking of Naphtha. 3. Aromatic Hydrocarbons, Ind. Eng. Chem. Res. 32 (1993) 2620–2625. https://doi.org/10.1021/ie00023a027.

F.D. Kopinke, G. Zimmermann, S. Nowak, On the mechanism of coke formation in steam cracking-conclusions from results obtained by tracer experiments, Carbon N. Y. 26 (1988) 117–124. https://doi.org/10.1016/0008-6223(88)90027-9.

F.W. Plankey, T.H. Glenn, L.P. Hart, J.D. Winefordner, Hadamard Spectrometer for Ultraviolet-Visible Spectrometry, Anal. Chem. 46 (1974) 1000–1005.

G. Datschefski, Role of the jftot in aviation fuel stability research, Fuel Sci. Technol. Int. 6 (1988) 609–631. https://doi.org/10.1080/08843758808915906.

G. Liu, L. Wang, H. Qu, H. Shen, X. Zhang, S. Zhang, Z. Mi, Artificial neural network approaches on composition-property relationships of jet fuels based on GC-MS, Fuel 86 (2007) 2551–2559. https://doi.org/10.1016/j.fuel.2007.02.023.

G. Malmquist, R. Danielsson, Alignment of chromatographic profiles for principal component analysis: a prerequisite for fingerprinting methods, J. Chromatogr. A 687 (1994) 71–88. https://doi.org/10.1016/0021-9673(94)00726-8.

G. Milman, A.J. Barnes, R.H. Lowe, M.A. Huestis, Simultaneous quantification of cannabinoids and metabolites in oral fluid by two-dimensional gas chromatography mass spectrometry, J. Chromatogr. A 1217 (2010) 1513–1521. https://doi.org/10.1016/J.CHROMA.2009.12.053.

G. Tomasi, F. Van Den Berg, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, J. Chemom. 18 (2004) 231–241. https://doi.org/10.1002/CEM.859.

G.L. Alexandrino, J. Malmborg, F. Augusto, J.H. Christensen, Investigating weathering in light diesel oils using comprehensive two-dimensional gas chromatography–High resolution mass spectrometry and pixel-based analysis: Possibilities and limitations, J. Chromatogr. A 1591 (2019) 155–161. https://doi.org/10.1016/J.CHROMA.2019.01.042.

G.S. Ochoa, M.C. Billingsley, R.E. Synovec, Using solid-phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: comparative analysis of aerospace fuel composition, Anal. Bioanal. Chem. (2022) 1–13. https://doi.org/10.1007/s00216-022-04348-1.

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 (2021) 122844. https://doi.org/10.1016/J.TALANTA.2021.122844.

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. https://doi.org/10.1016/J.TALANTA.2021.122844.

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. https://doi.org/10.1016/j.chroma.2020.461401.

G.T. Ventura, G.J. Hall, R.K. Nelson, G.S. Frysinger, B. Raghuraman, A.E. Pomerantz, O.C. Mullins, C.M. Reddy, Analysis of petroleum compositional similarity using multiway principal components analysis (MPCA) with comprehensive two-dimensional gas chromatographic data, J. Chromatogr. A 1218 (2011) 2584–2592. https://doi.org/10.1016/J.CHROMA.2011.03.004.

G.Z. Fang, J. Tan, X.P. Yan, An ion-imprinted functionalized silica gel sorbent prepared by a surface imprinting technique combined with a sol-gel process for selective solid-phase extraction of

cadmium(II), Anal. Chem. 77 (2005) 1734–1739. https://doi.org/10.1021/AC048570V/ASSET/IMAGES/LARGE/AC048570VF00002.JPEG.

H. Parastar, J.R. Radovi, M. Jalali-Heravi, S. Diez, J. Maria Bayona, R. Tauler, Resolution and Quantification of Complex Mixtures of Polycyclic Aromatic Hydrocarbons in Heavy Fuel Oil Sample by Means of GC Â GC-TOFMS Combined to Multivariate Curve Resolution, Anal. Chem. 83 (2011) 9289–9297. https://doi.org/10.1021/ac201799r.

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

I.C. Lee, H.C. Ubanyionwu, Determination of sulfur contaminants in military jet fuels, Fuel 87 (2008) 312–318. https://doi.org/10.1016/J.FUEL.2007.05.010.

I.H.M. van Stokkum, K.M. Mullen, V. V. Mihaleva, Global analysis of multiple gas chromatography-mass spectrometry (GC/MS) data sets: A method for resolution of co-eluting components with comparison to MCR-ALS, Chemom. Intell. Lab. Syst. 95 (2009) 150–163. https://doi.org/10.1016/j.chemolab.2008.10.004.

J. Fleming, Biden's 'Sustainable Aviation Fuel' Goals Mired in Myth: Report Diversified Efforts Needed to Decarbonize Aviation, (2022). https://biologicaldiversity.org/w/news/press-releases/bidens-sustainable-aviation-fuel-goals-mired-in-myth-report-2022-08-24/.

J. Heyne, D. Bell, J. Feldhausen, Z. Yang, R. Boehm, Towards fuel composition and properties from Two-dimensional gas chromatography with flame ionization and vacuum ultraviolet spectroscopy, Fuel 312 (2022) 122709. https://doi.org/10.1016/J.FUEL.2021.122709.

J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemom. Intell. Lab. Syst. 140 (2015) 1–12. https://doi.org/10.1016/j.chemolab.2014.10.003.

J. Ma, G. Wu, S. Li, W. Tan, X. Wang, J. Li, L. Chen, Magnetic solid-phase extraction of heterocyclic pesticides in environmental water samples using metal-organic frameworks coupled to high performance liquid chromatography determination, J. Chromatogr. A 1553 (2018) 57–66. https://doi.org/10.1016/J.CHROMA.2018.04.034.

J. Oh, A. Oldani, T. Lee, L. Shafer, Deep Learning Algorithms for Assessing Sus-tainable Jet Fuels from Two-Dimensional Gas Chromatography, (2022). https://doi.org/10.2514/6.2022-0228.

J. Orzel, B. Krakowska, I. Stanimirova, M. Daszykowski, Detecting chemical markers to uncover counterfeit rebated excise duty diesel oil, Talanta 204 (2019) 229–237. https://doi.org/10.1016/J.TALANTA.2019.05.113.

J. V. Seeley, S.K. Seeley, E.K. Libby, J.D. McCurry, Analysis of Biodiesel/Petroleum Diesel Blends with Comprehensive Two-Dimensional Gas Chromatography, J. Chromatogr. Sci. 45 (2007) 650–656. https://doi.org/10.1093/CHROMSCI/45.10.650.

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

J. Wang, M.F. Reyniers, G.B. Marin, Influence of dimethyl disulfide on coke formation during steam cracking of hydrocarbons, Ind. Eng. Chem. Res. 46 (2007) 4134–4148. https://doi.org/10.1021/ie061096u.

J.A. Murray, Qualitative and quantitative approaches in comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1261 (2012) 58–68. https://doi.org/10.1016/J.CHROMA.2012.05.012.

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

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

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. https://doi.org/10.1021/ac061710b.

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

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

J.M. Amigo, M.J. Popielarz, R.M. Callejón, M.L. Morales, A.M. Troncoso, M.A. Petersen, T.B. Toldam-Andersen, Comprehensive analysis of chromatographic data by using PARAFAC2 and principal components analysis, J. Chromatogr. A 1217 (2010) 4422–4429. https://doi.org/10.1016/J.CHROMA.2010.04.042.

J.M. Baerncopf, V.L. McGuffin, R.W. Smith, Association of Ignitable Liquid Residues to Neat Ignitable Liquids in the Presence of Matrix Interferences Using Chemometric Procedures*,†, J. Forensic Sci. 56 (2011) 70–81. https://doi.org/10.1111/J.1556-4029.2010.01563.X.

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

J.M. Davis, Statistical Theory of Spot Overlap in Two-Dimensional Separations, Anal. Chem. 63 (1991) 2141–2152. https://doi.org/10.1021/ac00019a014.

J.M. Hogan, R.A. Engel, H.F. Stevenson, A Versatile Internal Standard Technique for the Gas Chromatographic Determination of Water in Liquids, Anal. Chem. 42 (1970) 249–252. https://doi.org/10.1021/AC60284A033/ASSET/AC60284A033.FP.PNG_V03.

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

J.W. McIlroy, R.W. Smith, V.L. McGuffin, Assessing the effect of data pretreatment procedures for principal components analysis of chromatographic data, Forensic Sci. Int. 257 (2015) 1–12. https://doi.org/10.1016/J.FORSCIINT.2015.07.038.

K. Mainali, M. Garcia-Perez, Identification and quantification of trace oxygenated compounds in alternative jet fuels: Fluorescence methods for fast detection of phenolic compounds in operational field conditions, Fuel 271 (2020) 117652. https://doi.org/10.1016/j.fuel.2020.117652.

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. https://doi.org/10.1016/J.TALANTA.2022.123301.

K. Pyrzynska, A. Kubiak, I. Wysocka, Application of solid phase extraction procedures for rare earth elements determination in environmental samples, Talanta 154 (2016) 15–22. https://doi.org/10.1016/J.TALANTA.2016.03.022.

K. Robards, P.R. Haddad, P.E. Jackson, Gas Chromatography, in: Princ. Pract. Mod. Chromatgor. Methods, Academic Press, San Diego, 1994.

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

K.J. Johnson, B.W. Wright, K.H. Jarman, R.E. Synovec, High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis, J. Chromatogr. A 996 (2003) 141–155. https://doi.org/10.1016/S0021-9673(03)00616-2.

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. https://doi.org/10.1016/S0169-7439(01)00198-8.

K.J. Johnson, S.L. Rose-Pehrsson, R.E. Morris, Characterization of fuel blends by GC-MS and multi-way chemometric tools, Pet. Sci. Technol. 24 (2006) 1175–1186. https://doi.org/10.1081/LFT-200048192.

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. https://doi.org/10.1021/acs.energyfuels.9b04108.

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

K.M. Pierce, B.W. Wright, R.E. Synovec, Unsupervised parameter optimization for automated retention time alignment of severely shifted gas chromatographic data using the piecewise alignment algorithm, J. Chromatogr. A 1141 (2007) 106–116. https://doi.org/10.1016/J.CHROMA.2006.11.101.

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

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

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. https://doi.org/10.1016/J.TALANTA.2006.01.038.

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

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

K.M. Sharif, S.T. Chin, C. Kulsing, P.J. Marriott, The microfluidic Deans switch: 50 years of progress, innovation and application, TrAC Trends Anal. Chem. 82 (2016) 35–54. https://doi.org/10.1016/J.TRAC.2016.05.005.

L. Fernandez-Lopez, A. Luna-Maldonado, M. Falcon, L. Mastrobattista, J. Navarro-Zaragoza, R. Mancini, Development and validation of a gas chromatography–mass spectrometry method for opiates and cocaine in human bone, J. Pharm. Biomed. Anal. 164 (2019) 636–641. https://doi.org/10.1016/J.JPBA.2018.11.015.

L. Guozhu, H. Yongjin, W. Li, Z. Xiangwen, M. Zhentao, Solid deposits from thermal stressing of n - Dodecane and chinese RP-3 jet fuel in the presence of several initiators, Energy and Fuels 23 (2009) 356–365. https://doi.org/10.1021/ef800657z.

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

L.M. Balster, S. Zabarnick, R.C. Striebich, L.M. Shafer, Z.J. West, Analysis of polar species in jet fuel and determination of their role in autoxidative deposit formation, Energy and Fuels 20 (2006) 2564–2571. https://doi.org/10.1021/ef060275l.

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

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

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. https://doi.org/10.1016/j.forc.2018.01.001.

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

L.W. Hantao, H.G. Aleme, M.M. Passador, E.L. Furtado, F.A. de L. Ribeiro, R.J. Poppi, F. Augusto, Determination of disease biomarkers in Eucalyptus by comprehensive two-dimensional gas chromatography and multivariate data analysis, J. Chromatogr. A 1279 (2013) 86–91. https://doi.org/10.1016/j.chroma.2013.01.013.

M. Barroso, S. Costa, M. Dias, D.N. Vieira, J.A. Queiroz, M. López-Rivadulla, Analysis of phenylpiperazine-like stimulants in human hair as trimethylsilyl derivatives by gas chromatography–mass spectrometry, J. Chromatogr. A 1217 (2010) 6274–6280. https://doi.org/10.1016/J.CHROMA.2010.08.001.

M. Brokl, L. Bishop, C.G. Wright, C. Liu, K. McAdam, J.F. Focant, Multivariate analysis of mainstream tobacco smoke particulate phase by headspace solid-phase micro extraction coupled with comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, J. Chromatogr. A 1370 (2014) 216–229. https://doi.org/10.1016/j.chroma.2014.10.057.

M. Fasciotti, T.V.C. Monteiro, A.A. Ferreira, M.N. Eberlin, L.A. Neves, Two-point normalization using internal and external standards for a traceable determination of δ13C values of fatty acid methyl esters by gas chromatography/combustion/isotope ratio mass spectrometry, Int. J. Mass Spectrom. 418 (2017) 41–50. https://doi.org/10.1016/J.IJMS.2016.12.002.

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. https://doi.org/10.1016/J.FUEL.2018.07.081.

M. Maeder, A. Zilian, Evolving factor analysis, a new multivariate technique in chromatography, Chemom. Intell. Lab. Syst. 3 (1988) 205–213. https://doi.org/10.1016/0169-7439(88)80051-0.

M. Moors, D.L. Massart, R.D. McDowall, Analyte isolation by solid phase extraction (SPE) on silica-bonded phases: Classification and recommended practices, Pure Appl. Chem. 66 (1994) 277–304. https://doi.org/10.1351/pac199466020277.

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

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

M.A. Sharaf, B.R. Kowalski, Extraction of Individual Mass Spectra from Gas Chromatography-Mass Spectrometry Data of Unseparated Mixtures, Anal. Chem. 53 (1981) 518–522. https://doi.org/10.1021/ac00226a031.

M.D. King, G.S. King, A Numerical Method for Extracting Mass Spectra from Gas Chromatography/Mass Spectrometry Data Arrays, Anal. Chem. 57 (1985) 1049–1056. https://doi.org/10.1021/ac00283a020.

M.-F.F.S.G. Reyniers, G.F. Froment, Influence of Metal Surface and Sulfur Addition on Coke Deposition in the Thermal Cracking of Hydrocarbons, Ind. Eng. Chem. Res. 34 (1995) 773–785. https://doi.org/10.1021/ie00042a009.

M.J. DeWitt, T. Edwards, L. Shafer, D. Brooks, R. Striebich, S.P. Bagley, M.J. Wornat, Effect of Aviation Fuel Type on Pyrolytic Reactivity and Deposition Propensity under Supercritical Conditions, Ind. Eng. Chem. Res. 50 (2011) 10434–10451. https://doi.org/10.1021/IE200257B.

M.K. Das, S.C. Bishwal, A. Das, D. Dabral, A. Varshney, V.K. Badireddy, R. Nanda, Investigation of gender-specific exhaled breath volatome in humans by GCxGC-TOF-MS, Anal. Chem. 86 (2014) 1229–1237. https://doi.org/10.1021/ac403541a.

M.K. Jennerwein, A.C. Sutherland, M. Eschner, T. Gröger, T. Wilharm, R. Zimmermann, Quantitative analysis of modern fuels derived from middle distillates – The impact of diverse compositions on standard methods evaluated by an offline hyphenation of HPLC-refractive index detection with GC×GC-TOFMS, Fuel 187 (2017) 16–25. https://doi.org/10.1016/j.fuel.2016.09.033.

M.K. Jennerwein, M. Eschner, T. Gröger, T. Wilharm, R. Zimmermann, Complete group-type quantification of petroleum middle distillates based on comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC-TOFMS) and visual basic scripting, Energy and Fuels 28 (2014) 5670–5681. https://doi.org/10.1021/ef501247h.

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

M.R. Webb, C.N. LaFratta, D.R. Walt, Chromatically Resolved Optical Microscope (CROMoscope): A Grating-Based Instrument for Spectral Imaging, Anal. Chem. 81 (2009) 7309–7313.

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

M.S. Shahriman, M.R. Ramachandran, N.N.M. Zain, S. Mohamad, N.S.A. Manan, S.M. Yaman, Polyaniline-dicationic ionic liquid coated with magnetic nanoparticles composite for magnetic solid phase extraction of polycyclic aromatic hydrocarbons in environmental samples, Talanta 178 (2018) 211–221. https://doi.org/10.1016/J.TALANTA.2017.09.023.

N. Di Giovanni, M.A. Meuwis, E. Louis, J.-F.F. Focant, Untargeted Serum Metabolic Profiling by Comprehensive Two-Dimensional Gas Chromatography-High-Resolution Time-of-Flight Mass Spectrometry, J. Proteome Res. 19 (2020) 1013–1028. https://doi.org/10.1021/acs.jproteome.9b00535.

N. Etxebarria, O. Zuloaga, M. Olivares, L.J. Bartolomé, P. Navarro, Retention-time locked methods in gas chromatography, J. Chromatogr. A 1216 (2009) 1624–1629. https://doi.org/10.1016/J.CHROMA.2008.12.038.

N.D. Spadafora, S. Mascrez, L. McGregor, G. Purcaro, Exploring multiple-cumulative trapping solid-phase microextraction coupled to gas chromatography–mass spectrometry for quality and authenticity assessment of olive oil, Food Chem. 383 (2022) 132438. https://doi.org/10.1016/J.FOODCHEM.2022.132438.

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

N.G.S. Mogollón, F.A.L. Ribeiro, R.J. Poppi, A.L. Quintana, J.A.G. Chávez, D.A.P. Agualongo, H.G. Aleme, F. Augusto, Exploratory Analysis of Biodiesel by Combining Comprehensive Two-Dimensional Gas Chromatography and Multiway Principal Component Analysis, Artic. J. Braz. Chem. Soc. 28 (2017) 740–746. https://doi.org/10.21577/0103-5053.20160222.

N.J. Begue, J.A. Cramer, C. Von Bargen, K.M. Myers, K.J. Johnson, R.E. Morris, Automated method for determining hydrocarbon distributions in mobility fuels, Energy and Fuels 25 (2011) 1617–1623. https://doi.org/10.1021/ef101635a.

N.V. Hung, C. Mohabeer, M. Vaccaro, S. Marcotte, V. Agasse-Peulon, L. Abdelouahed, B. Taouk, P. Cardinael, Development of two-dimensional gas chromatography (GC×GC) coupled with Orbitrap-technology-based mass spectrometry: Interest in the identification of biofuel composition, J. Mass Spectrom. 55 (2020) e4495. https://doi.org/10.1002/JMS.4495.

O. Altin, S. Eser, Analysis of Carboneceous Deposits from Thermal Stressing of a JP-8 Fuel on Superalloy Foils in a Flow Reactor, Ind. Eng. Chem. Res. 40 (2001) 589–595. https://doi.org/10.1021/ie0004489.

O. Altin, S. Eser, Carbon deposition from thermal stressing of petroleum fuels, in: Am. Chem. Soc. Div. Fuel Chem., 2004: pp. 764–766. https://www.researchgate.net/publication/267420915 (accessed June 28, 2021).

O. Coskun, Separation techniques: Chromatography, North. Clin. Istanbul. 3 (2016) 156. https://doi.org/10.14744/NCI.2016.32757.

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

P. Filzmoser, B. Walczak, What can go wrong at the data normalization step for identification of biomarkers?, J. Chromatogr. A 1362 (2014) 194–205. https://doi.org/10.1016/J.CHROMA.2014.08.050.

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. https://doi.org/10.1016/J.TALANTA.2019.120239.

P.E. Sudol, G.S. Ochoa, C.N. Cain, R.E. Synovec, Tile-based variance rank initiated-unsupervised sample indexing for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, Anal. Chim. Acta 1209 (2022) 339847. https://doi.org/10.1016/j.aca.2022.339847.

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. https://doi.org/10.1016/j.chroma.2021.462092.

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 1132 (2020) 157–186. https://doi.org/10.1016/j.aca.2020.07.027.

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. https://doi.org/10.1016/J.CHROMA.2021.462735.

P.H. Stefanuto, K.A. Perrault, L.M. Dubois, B. L'Homme, C. Allen, C. Loughnane, N. Ochiai, J.F. Focant, Advanced method optimization for volatile aroma profiling of beer using two-

dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A 1507 (2017) 45–52. https://doi.org/10.1016/j.chroma.2017.05.064.

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

P.H.C. Eilers, Parametric Time Warping, Anal. Chem. 76 (2004) 404–411. https://doi.org/10.1021/AC034800E/ASSET/IMAGES/MEDIUM/AC034800EE00007.GIF.

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

P.M. Paniv, S. V. Pysh'ev, V.I. Gaivanovich, O.I. Lazorko, Noncatalytic oxidation desulfurization of the kerosene cut, Chem. Technol. Fuels Oils. 42 (2006) 159–166. https://doi.org/10.1007/s10553-006-0049-4.

P.Y. Hsieh, K.R. Abel, T.J. Bruno, Analysis of marine diesel fuel with the advanced distillation curve method, Energy and Fuels 27 (2013) 804–810. https://doi.org/10.1021/ef3020525.

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. https://doi.org/10.1002/(SICI)1099-128X(199905/08)13:3/4<295::AID-CEM547>3.0.CO;2-Y.

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

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

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. https://doi.org/10.1016/j.jchromb.2019.01.029.

R. V. Gough, T.J. Bruno, Composition-explicit distillation curves of alternative turbine fuels, Energy and Fuels 27 (2013) 294–302. https://doi.org/10.1021/ef3016848.

R. van der Westhuizen, M. Ajam, P. De Coning, J. Beens, A. de Villiers, P. Sandra, Comprehensive two-dimensional gas chromatography for the analysis of synthetic and crude-derived jet fuels, J. Chromatogr. A 1218 (2011) 4478–4486. https://doi.org/10.1016/j.chroma.2011.05.009.

R. Venkataraman, S. Eser, Characterization of solid deposits formed from jet fuel degradation under pyrolytic conditions: Metal sulfides, Ind. Eng. Chem. Res. 47 (2008) 9351–9360. https://doi.org/10.1021/ie801007r.

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

R.B. Pellegrino Vidal, A.C. Olivieri, R. Tauler, Quantifying the Prediction Error in Analytical Multivariate Curve Resolution Studies of Multicomponent Systems, Anal. Chem. 90 (2018) 7040–7047. https://doi.org/10.1021/acs.analchem.8b01431.

R.C. Striebich, J. Contreras, L.M. Balster, Z. West, L.M. Shafer, S. Zabarnick, Identification of polar species in aviation fuels using multidimensional gas chromatography-time of flight mass spectrometry, Energy and Fuels 23 (2009) 5474–5482. https://doi.org/10.1021/ef900386x.

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

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. https://doi.org/10.1021/AC052106O/ASSET/IMAGES/LARGE/AC052106OF00007.JPEG.

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. https://doi.org/10.1021/ac971184b.

R.E. Synovec, C.E. Freye, B.A. Parsons, M.C. Billingsley, N. Keim, B. Hill-Lam, J.C. Wilhelm, Recent Advances in the Development of Chemical Analysis Tools to Relate Compositional Variation to Thermal Integrity Data for RP-1, RP-2, and Related Fuels, in: JANNAF 8th Liq. Propuls. Meet., Nashville, TN, 2015.

R.E. Synovec, C.E. Freye, M.C. Billingsley, N. Keim, B. Hill-Lam, Recent Advances in Relating Chemical Compositional Variation in RP-1, RP-2, and Similar Fuels to Thermal Integrity Data, in: JANNAF 10th Liq. Propuls. Meet. Long Beach, CA, 2018: p. Abstract 2018-0001CV.

R.E. Synovec, C.E. Freye, M.C. Billingsley, N. Keim, B. Hill-Lam, Recent Advances in the Development of Chemical Analysis Tools to Relate Compositional Variation to Thermal Integrity Data for RP-1, RP-2, and Related Fuels, in: JANNAF 9th Liq. Propuls. Meet. Phoenix, AZ, 2016.

R.E. Synovec, K.L. Berrier, C.E. Freye, M.C. Billingsley, N. Keim, B. Hill-Lam, A. Bishop, Improvements to GC×GC-TOFMS Composition-Based Models for Hydrocarbon Fuel Thermal Integrity, in: JANNAF 11th Liq. Propuls. Meet., Tampa, FL, 2019.

R.H. Liu, G. Foster, E.J. Cone, S.D. Kumar, Selecting an appropriate isotopic internal standard for gas chromatography/mass spectrometry analysis of drugs of abuse -- pentobarbital example, J. Forensic Sci. 40 (1995) 983–989.

R.J. Strife, J.R. Simms, M.P. Lacey, Combined capillary gas chromatography/ion trap mass spectrometry quantitative methods using labeled or unlabeled internal standards, J. Am. Soc. Mass Spectrom. 1 (1990) 265–271. https://doi.org/10.1016/1044-0305(90)85044-M.

R.R. Mallepally, B.A. Bamgbade, M.A. McHugh, H.O. Baled, R.M. Enick, M.C. Billingsley, Measurements and modeling of the density of rocket propellant RP-2 at temperatures to 573 K and pressures to 100 MPa, Fuel 253 (2019) 1193–1203. https://doi.org/10.1016/J.FUEL.2019.05.089.

S. Eser, R. Venkataraman, O. Altin, Deposition of carbonaceous solids on different substrates from thermal stressing of JP-8 and jet A fuels, Ind. Eng. Chem. Res. 45 (2006) 8946–8955. https://doi.org/10.1021/ie060968p.

S. Ma, M. He, B. Chen, W. Deng, Q. Zheng, B. Hu, Magnetic solid phase extraction coupled with inductively coupled plasma mass spectrometry for the speciation of mercury in environmental

water and human hair samples, Talanta 146 (2016) 93–99. https://doi.org/10.1016/J.TALANTA.2015.08.036.

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

S. Tang, N. Shi, J. Wang, A. Tang, Comparison of the anti-coking performance of CVD TiN, TiO2 and TiC coatings for hydrocarbon fuel pyrolysis, Ceram. Int. 43 (2017) 3818–3823. https://doi.org/10.1016/j.ceramint.2016.12.036.

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

S. Wold, M. Sjöström, L. Eriksson, PLS-regression: A basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.

S. Yuan, H. Li, Z. Liu, Y. Wang, L. Wang, X. Zhang, G. Liu, Measurement of non-hindered and hindered phenolic species in aviation fuels via tandem-SPE with comprehensive GC × GC–MS/FID, Fuel 287 (2021) 119561. https://doi.org/10.1016/j.fuel.2020.119561.

S.A. Dyer, Hadamard transform spectrometry, Chemom. Intell. Lab. Syst. 12 (1991) 101–115.

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 1605 (2019) 460368. https://doi.org/10.1016/J.CHROMA.2019.460368.

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. https://doi.org/10.1021/acs.analchem.0c03456.

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

S.E. Reichenbach, C.A. Zini, K.P. Nicolli, J.E. Welke, C. Cordero, Q. Tao, Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition, J. Chromatogr. A 1595 (2019) 158–167. https://doi.org/10.1016/j.chroma.2019.02.027.

S.K. Jeon, D. Kwon, S. Lee, Identification of weathered multiple petroleum products in contaminated soils by characterizing unresolved complex mixture hump in gas chromatograph data, Sci. Total Environ. 607–608 (2017) 42–52. https://doi.org/10.1016/J.SCITOTENV.2017.06.251.

S.L. Anjos, J.S. Almeida, L.S.G. Teixeira, A.C.M. da Silva, A.P. Santos, A.F.S. Queiroz, S.L.C. Ferreira, S. Mattedi, Determination of Cu, Ni, Mn and Zn in diesel oil samples using energy dispersive X-ray fluorescence spectrometry after solid phase extraction using sisal fiber, Talanta 225 (2021). https://doi.org/10.1016/J.TALANTA.2020.121910.

S.M. Rowland, W.K. Robbins, Y.E. Corilo, A.G. Marshall, R.P. Rodgers, Solid-phase extraction fractionation to extend the characterization of naphthenic acids in crude oil by electrospray ionization fourier transform ion cyclotron resonance mass spectrometry, Energy and Fuels 28

(2014) 5043–5048. https://doi.org/10.1021/EF5015023/ASSET/IMAGES/LARGE/EF-2014-015023_0008.JPEG.

T. Dymerski, J. Namieśnik, K. Vearasilp, P. Arancibia-Avila, F. Toledo, M. Weisz, E. Katrich, S. Gorinstein, Comprehensive two-dimensional gas chromatography and three-dimensional fluorometry for detection of volatile and bioactive substances in some berries, Talanta 134 (2015) 460–467. https://doi.org/10.1016/j.talanta.2014.11.061.

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

T. Miyazaki, K. Okada, T. Yamashita, M. Miyazaki, Two-dimensional gas chromatography time-of-flight mass spectrometry-based serum metabolic fingerprints of neonatal calves before and after first colostrum ingestion, J. Dairy Sci. 100 (2017) 4354–4364. https://doi.org/10.3168/jds.2017-12557.

T. Murakami, Y. Iwamuro, R. Ishimaru, S. Chinaka, H. Hasegawa, Molecularly imprinted polymer solid-phase extraction of synthetic cathinones from urine and whole blood samples, J. Sep. Sci. 41 (2018) 4506–4514. https://doi.org/10.1002/JSSC.201800874.

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

T.A. Bolotnik, Y. V. Timchenko, I. V. Plyushchenko, V. V. Levkina, A. V. Pirogov, A.D. Smolenkov, M. V. Popik, O.A. Shpigun, Use of Chemometric Methods of Data Analysis for the Identification and Typification of Petroleum and Petroleum Products, J. Anal. Chem. 2019 7413. 74 (2019) 1336–1340. https://doi.org/10.1134/S1061934819130045.

T.J. Bruno, M.L. Huber, E.W. Lemmon, Effect of RP-1 compositional variability on thermophysical properties, Energy and Fuels 23 (2009) 5550–5555. https://doi.org/10.1021/ef900597q.

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. https://doi.org/10.1021/acs.analchem.2c01549.

T.M. Lovestead, B.C. Windom, J.R. Riggs, C. Nickell, T.J. Bruno, Assessment of the compositional variability of RP-1 and RP-2 with the advanced distillation curve approach, Energy and Fuels 24 (2010) 5611–5623. https://doi.org/10.1021/ef100994w.

U. Tilstam, Sulfolane: A versatile dipolar aprotic solvent, Org. Process Res. Dev. 16 (2012) 1273–1278. https://doi.org/10.1021/op300108w.

V. Abrahamsson, N. Ristic, K. Franz, K. Van Geem, Comprehensive two-dimensional gas chromatography in combination with pixel-based analysis for fouling tendency prediction, J. Chromatogr. A 1501 (2017) 89–98. https://doi.org/10.1016/j.chroma.2017.04.021.

V. V. Lobodin, W.K. Robbins, J. Lu, R.P. Rodgers, Separation and Characterization of Reactive and Non-Reactive Sulfur in Petroleum and Its Fractions, Energy and Fuels 29 (2015) 6177–6186. https://doi.org/10.1021/acs.energyfuels.5b00780.

X. Domingo-Almenara, A. Perera, N. Ramírez, N. Cañellas, X. Correig, J. Brezmes, Compound identification in gas chromatography/mass spectrometry-based metabolomics by blind source separation, J. Chromatogr. A 1409 (2015) 226–233. https://doi.org/10.1016/j.chroma.2015.07.044.

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

Y. Li, Q. Ruan, Y. Li, G. Ye, X. Lu, X. Lin, G. Xu, A novel approach to transforming a non-targeted metabolic profiling method to a pseudo-targeted method using the retention time locking gas chromatography/mass spectrometry-selected ions monitoring, J. Chromatogr. A 1255 (2012) 228–236. https://doi.org/10.1016/J.CHROMA.2012.01.076.

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

Z. Liu, S. Tang, Z. Li, Z. Qin, S. Yuan, L. Wang, L. Wang, X. Zhang, G. Liu, An improved kinetic model for deposition by thermal oxidation of aviation hydrocarbon fuels, Fuel 258 (2019) 116139. https://doi.org/10.1016/J.FUEL.2019.116139.

Z. Yang, B.P. Hollebone, Z. Wang, C. Yang, M. Landriault, Determination of polar impurities in biodiesels using solid-phase extraction and gas chromatography-mass spectrometry, J. Sep. Sci. 34 (2011) 409–421. https://doi.org/10.1002/JSSC.201000533.

Z. Yang, S. Kosir, R. Stachler, L. Shafer, C. Anderson, J.S. Heyne, A GC × GC Tier α combustor operability prescreening method for sustainable aviation fuel candidates, Fuel 292 (2021) 120345. https://doi.org/10.1016/J.FUEL.2021.120345.

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

Z.Z. Kapadia, D. V Spracklen, S.R. Arnold, D.J. Borman, G.W. Mann, K.J. Pringle, S.A. Monks, C.L. Reddington, F. Benduhn, A. Rap, C.E. Scott, E.W. Butt, M. Yoshioka, Impacts of aviation fuel sulfur content on climate and human health, Atmos. Chem. Phys. 16 (2016) 10521–10541. https://doi.org/10.5194/acp-16-10521-2016.

Appendix A

This Appendix is reproduced from the Electronic Supplementary Material of Grant S. Ochoa, Sarah E. Prebihalo, Brooke C. Reaser, Luke C. Marney, Robert E. Synovec, "Statistical Inference of Mass Channel Purity from Fisher Ratio Analysis using Comprehensive Two-Dimensinoal Gas Chromatography with Time-of-Flight Mass Spectrometry Data" *J. Chromatogr. A* 1627 (2020) 461401

Flow Chart Depicting the Workflow for the S-ratio Algorithm

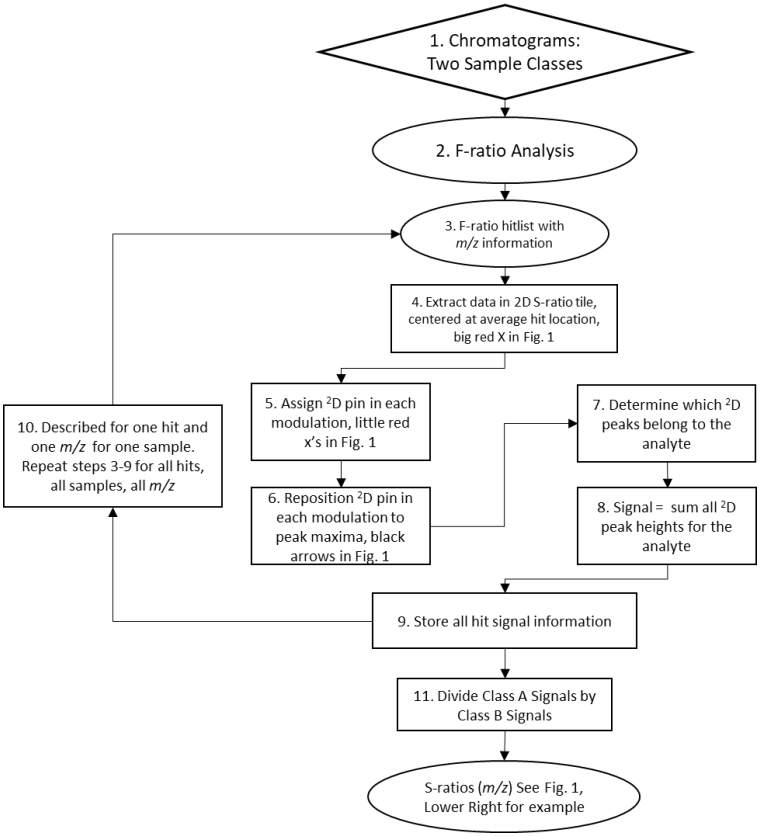


Figure A1. Flow chart depicting the workflow for the S-ratio algorithm. First, F-ratio analysis is performed on two classes of chromatograms yielding a hitlist ordered by decreasing F-ratio with hit locations along with a list of *m/z* that passed the *S/N* threshold for each hit (steps 1-3). Utilizing this information, the chromatographic data in the 2D S-ratio tile is extracted for a given hit around the average $^{1}D \times {}^{2}D$ pin represented by the large red X seen in Fig. 1 for the signal at one *m/z* (step 4). Initial ^{2}D pins are then assigned to each modulation based off the average $^{1}D \times {}^{2}D$ pin for the hit which serve as initial estimates to the peak maxima for the ^{2}D peak in each modulation, represented by the small red x's in Fig. 1 (step 5). These initial pins are then repositioned to the maxima of their respective ^{2}D peaks, black arrows (step 6). In step 7 we determine which ^{2}D peaks belong to the analyte by scanning for locations where consecutive peak maxima switch from decreasing to increasing in signal, indicating the presence of an interfering compound. Peak maxima that start to increase in signal are discarded and the signals

of the ^{2}D peaks determined to belong to the analyte are summed, yielding a single signal value for the hit at the selected *m/z* (step 8). The signal information is stored in an array (step 9) and the process from steps 3-9 is repeated (step 10) for each *m/z* for each selected hit from the F-ratio hitlist for each sample. At the end of the process we have now obtained the signal values for each software selected *m/z* for each hit for each sample. The signal values for class A samples are then divided by class B sample signals yielding six S-ratio values given a six versus six two class experimental design (step 11). The S-ratios can then be averaged yielding a mean S-ratio along with the standard deviation information.

Flow Chart Depicting the Workflow for the *LOF* Algorithm

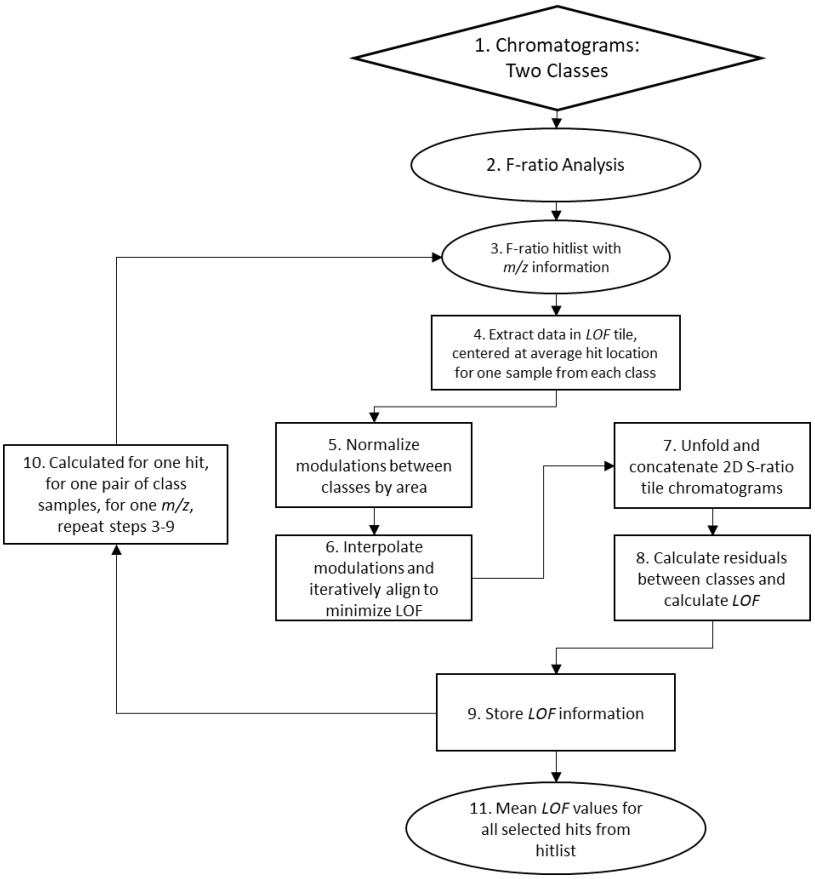


Figure A2. Flow chart depicting the workflow for the lack-of-fit (*LOF*) algorithm. Steps 1-3 are identical to the S-ratio algorithm, using complementary outputs. Utilizing the output information, we extract the chromatographic data in the *LOF* tile for a given hit around the average $^{1}D \times {}^{2}D$ pin represented by the large red X seen in Fig. 1 for one *m/z* for one sample from each class (step 4). The modulations of the extracted data are then normalized such that each modulation between has the same area between classes so that the peak shape can be evaluated with concentration effects removed (step 5). Next the data in each modulation is interpolated such that more points are added in between the existing points without altering the peak shape. These modulations are then iteratively aligned between classes calculating a *LOF* value for each iteration in an attempt to minimize this value so that the resulting *LOF* value is due to peak shape differences and not retention time shifting (step 6). The aligned modulations are then unfolded and concatenated, laid end to end, and the residuals are calculated between classes which are in turn used to calculate a single *LOF* value for the chromatogram pair (steps 7-8). This value is then stored in an array (step 9) and the process from steps 3-9 is repeated for each sample pair for each software selected *m/z* for each hit (step 10). At the end of this process mean *LOF*s and standard deviations can be calculated from the six sample pairs from each hit's mass channels (step 11).

Combinatorial Null Distribution Analysis (CNDA)

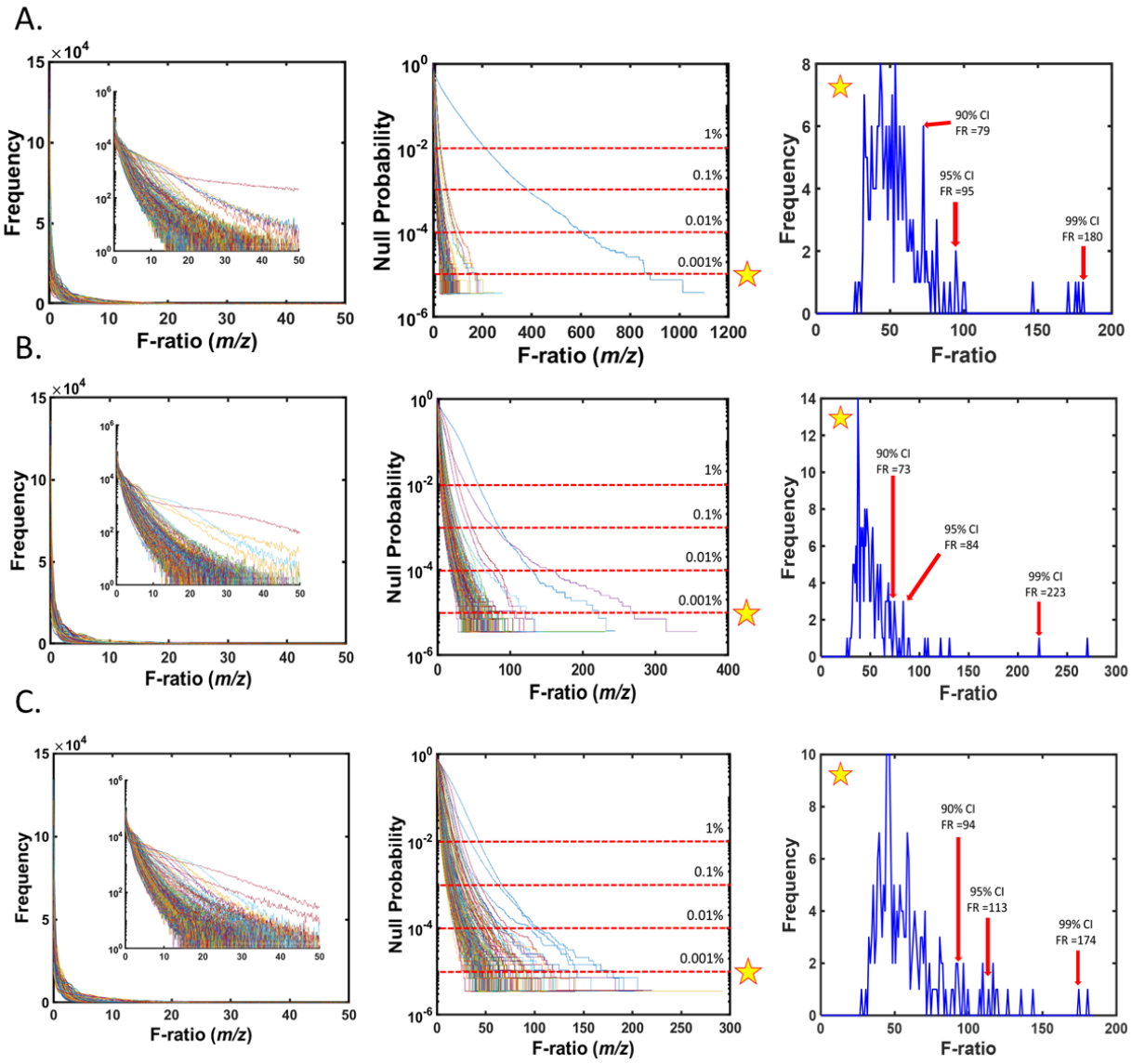


Figure A3. Null distributions using CNDA for each of the concentration comparisons: (A) 80/40 ppm, (B) 40/20 ppm, and (C) 20/10 ppm, and the inset displays a zoom in of distributions (Left). Null probability plots for each of the 200 combinatorial null class comparisons (Middle). Probability levels marked off by dashed red lines indicate the respective percentage of features lying below that line. For instance, for the probability curve passing the starred 0.001% line, the F-ratio value where it crossed represents the threshold that only 0.001% of features in that respective distribution lie above. By finding this value for each probability curve, we can create a distribution of 0.001% null probability F-ratio values (Right). The 80/40 comparison 0.001% probability distribution contains one F-ratio that is off scale at 881, excluded for distribution clarity in (A). The 90, 95, and 99% confidence levels for the 0.001% probability level are marked. Using the 95% level as an example, we can be 95% confident that an F-ratio in the true class comparison that exceeds 63 has a 1 in 100,000 (0.001%) chance of being a true positive.

Additional Spiked Analyte Data

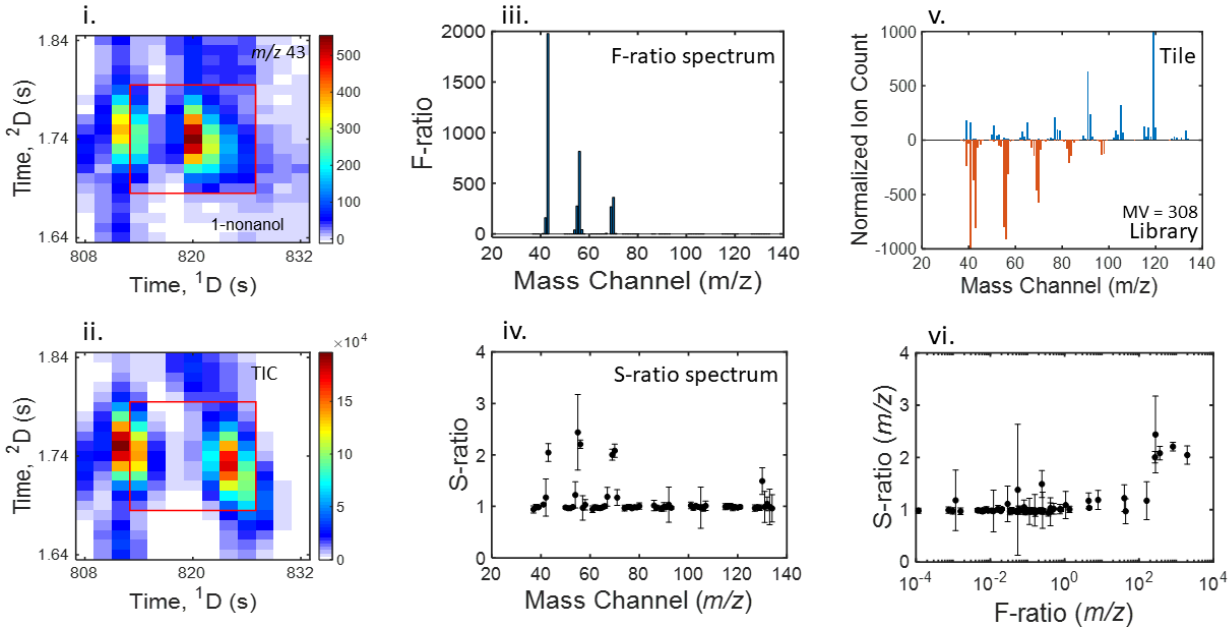


Figure A4. Illustration of the F-ratio(*m/z*) and S-ratio(*m/z*) spectra from the 80/40 ppm comparison for a severely overlapped analyte, 1-nonanol. Two-dimensional data for the region around each analyte for the top F-ratio *m/z* (i), inferred to be relatively pure, and the total ion current (TIC) (ii) from an average 80 ppm chromatogram. The red box outlines the tile space used by the S-ratio algorithm. The F-ratio(m/z) and S-ratio(m/z) spectrum is provided in (iii) and (iv), respectively. The mass spectrum from the red box region is also provided (v) and compared to the library spectrum. Finally, a plot of S-ratio(*m/z*) as a function of F-ratio(*m/z*) in (vi) provides insight into the correlation between high F-ratio(*m/z*) and having the S-ratio(*m/z*) correspond to the true concentration ratio of 2. Since very few F-ratio(*m/z*) correspond to results with an S-ratio(*m/z*) ≈ 2, this demonstrates a very challenging case.

Additional Comparison Statistical Plots

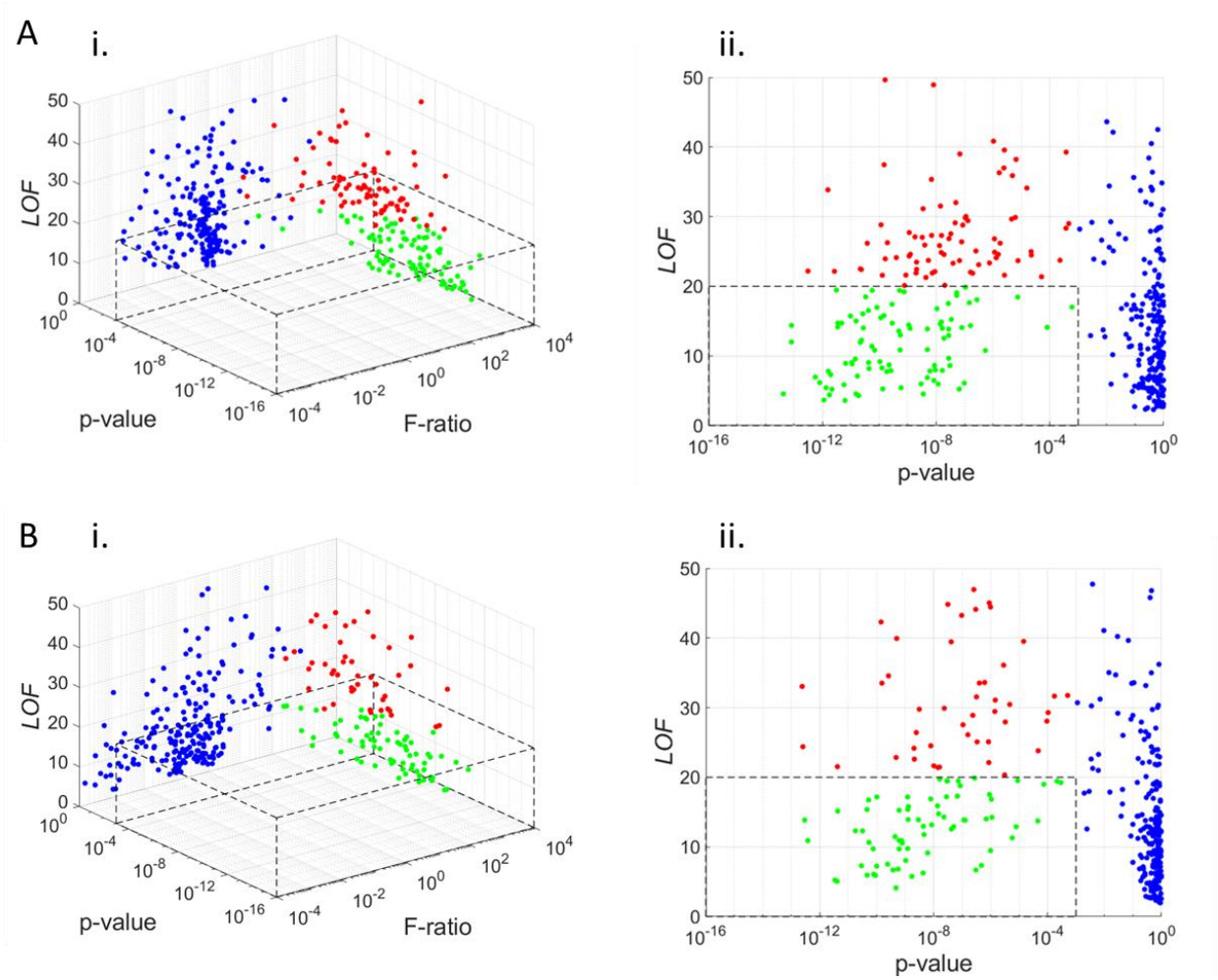


Figure A5. F-ratio(m/z), p-value(m/z), and *LOF(m/z)* for all *m/z* of all 15 spiked analytes for the for the 40/20 and 20/10 ppm comparisons, (A) and (B), respectively The dashed lines represent the regions defined by the p-value ≤ 0.001 and *LOF* ≤ 20 % thresholds (A) Blue *m/z* have not passed either threshold, red have only passed the p-value threshold, and green *m/z* have passed both thresholds. (B) A 2D projection of the *LOF* and p-value dimensions illustrates these two metrics independent of the F-ratio dimension. The blue *m/z* carry insignificant *m/*z belonging to interfering compounds. Red *m/z* have statistically significant signal differences between classes however they are interfered to a large extent with overlapping compounds. Lastly the green *m/z* are considered both statistically significant and sufficiently pure to be used for signal quantification, while not relying upon the F-ratio(*m/*z). We see that as the average concentration of the comparisons decreases there are fewer *m/z* that pass the F-ratio *S/N* filter and/or have higher *LOF* values.

Additional S-ratio versus F-ratio Threshold Plots

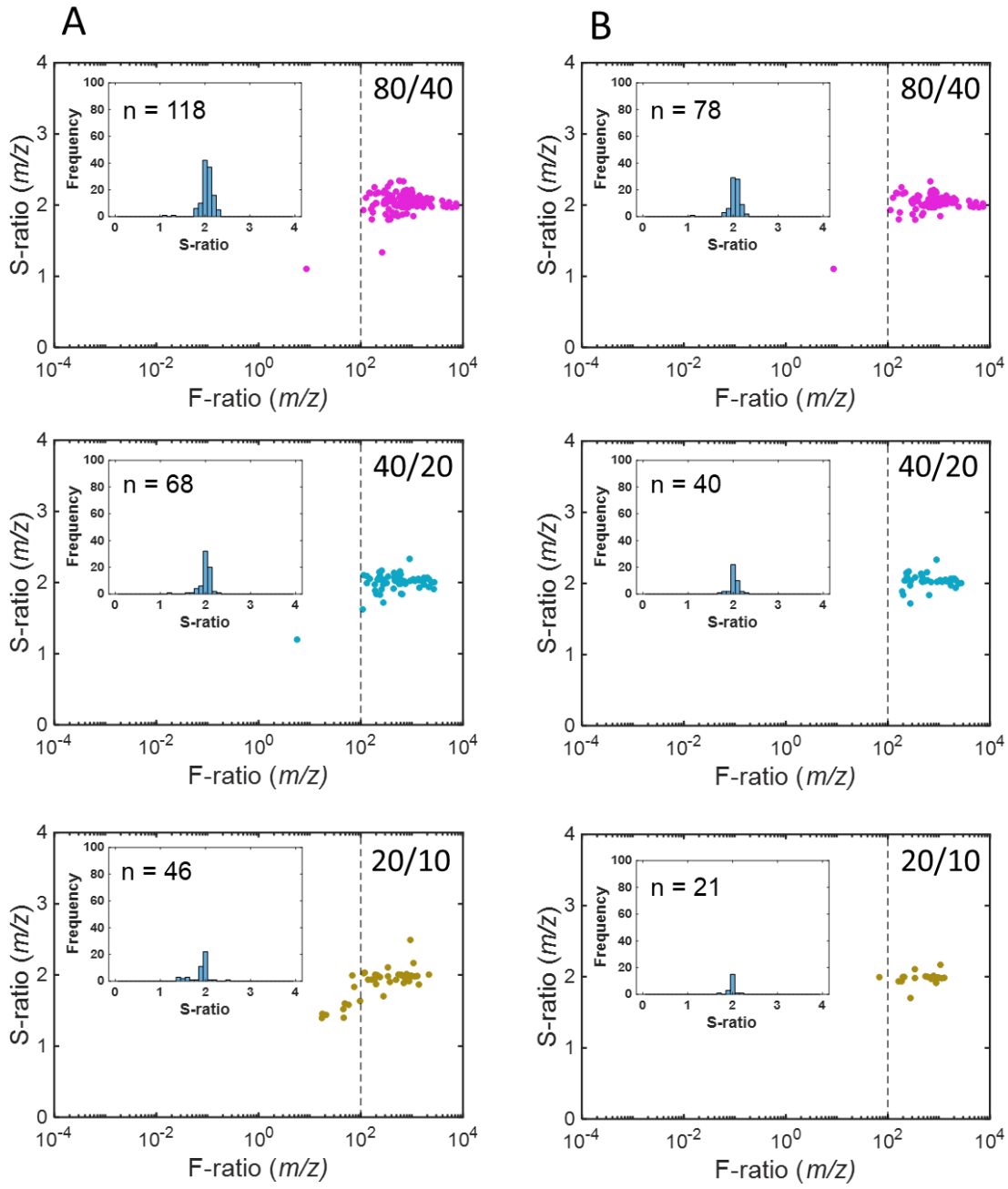


Figure A6. Plots of S-ratio vs F-ratio for each concentration comparison of the *m/z* that passed a p-value threshold of < 0.001, but with a *LOF* threshold of <15% (A) and <10% (B). The CNDA F-ratio threshold = 97 is marked for context. Inset in each plot are histograms of the *m/z* that pass these two metrics along with a count. Compared to the *LOF* threshold of < 20 in Figure 7 there are much fewer *m/z* in each comparison, reflecting the tightening of the parameter.

S-ratio Results for 40/20 ppm Comparison

Table A1. S-ratio results for each of the 15 spiked analytes in the 40/20 ppm comparison, arranged by decreasing F-ratio. The S-ratio for the top *m/z* which passed the p-value and *LOF* thresholds is provided and compared against the actual concentration ratio. Cyclohexyl isothiocyanate did not have any *m/z* passing these thresholds. Average concentration ratio deviation was 1.9% with a standard deviation of 2.1%.

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

S-ratio Results for 20/10 ppm Comparison

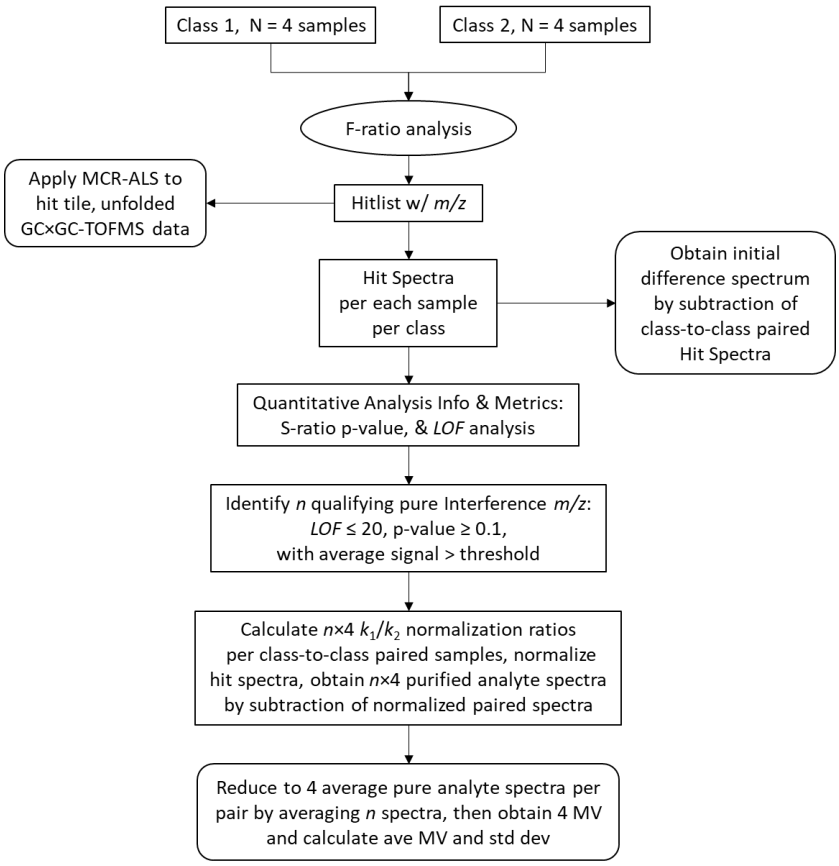
Table A2. S-ratio results for each of the 15 spiked analytes in the 20/10 ppm comparison, arranged by decreasing F-ratio. The S-ratio for the top *m/z* which passed the p-value and *LOF* thresholds is provided and compared against the actual concentration ratio. Limonene had no *m/z* pass both of these filters. 1-Nonanol had 1 passing *m/z* which appears to belong to an interferent. Average concentration ratio deviation was 6.1% with a standard deviation of 11%. If 1-nonanol is excluded, the resulting average deviation was 2.8% with a standard deviation of 2.4%.

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

Appendix B

This Appendix is reproduced from the Electronic Supplementary Material of Grant S. Ochoa, Paige E. Sudol, Timothy J. Trinklein, Robert E. Synovec, "Class comparison enabled mass spectrum purification for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry" *Talanta* 236 (2022) 461401

Figure B1. Flow chart depicting the workflow for the CCE-MSP algorithm. First, F-ratio analysis is performed on the 4 versus 4 class comparison of the 30 ppm versus 15 ppm sulfur compound spiked JP-8 jet fuel, from which a hitlist is output alongside the *m/z* passing the *S/N* threshold for each hit. Following the center of the flow chart, for each hit, based upon the hit spectra obtained, the S-ratio, p-value and *LOF* are calculated to determine *m/z* purity. These metrics are then used to determine the *m/z* which qualify as pure interference *m/z*, resulting in *n* *m/z* per hit. Using the hit spectra, k_1/k_2 ratios are calculated for each of the pure interference *m/z*. These ratios are utilized to normalize the class 1 hit spectra to their paired class 2 hit spectra. Then, the normalized class 1 spectra are subtracted from their paired class 2 spectra resulting in *n*×4 purified analyte spectra. These spectra are averaged per pair, resulting in 4 purified analyte spectra, from which MVs are obtained and an average MV calculated. Additionally, for each hit an initial analyte spectrum is obtained from the pin location which are then paired between classes and subtracted to obtain the initial difference spectrum without normalization. Finally, from each hit tile the data is extracted and unfolded for each sample, then submitted to MCR-ALS to attempt to isolate each analyte spectrum for comparison to the CCE-MSP results.



In-depth experimental description

Analysis of the neat and spiked jet fuels was performed (4 replicates each) using a Pegasus BT 4D GC×GC-TOFMS instrument (LECO Corporation, St. Joseph, MI) with a quad-jet thermal modulator, an Agilent 7890 gas chromatograph (Agilent Technologies, Palo Alto, CA) and an L-PAL3 GC auto sampler. The auto injector was programmed to inject 0.5 μl at a 200:1 split ratio at an inlet temperature of 275 °C. High-performance liquid chromatography grade acetone and hexane (Fisher Scientific, Waltham, MA) were used as solvent rinses between sample injections. The ^{1}D column was an Rxi-17Sil MS (29.5 m × 250 μm inner diameter (i.d.) × 0.25 μm film thickness), and the ^{2}D column was an Rxi-1MS (1.9 m × 180 μm i.d. × 0.18 μm film thickness). Ultrahigh purity helium (grade 5, 99.999%, Praxair, Seattle, WA) was used as the carrier gas at a constant flow rate of 2.0 ml/min. The temperature of the ^{1}D oven was held at 40 °C for 1 min, then ramped at a rate of 5 °C/min to 200 °C and held there for 1.5 min. The ^{2}D oven was held at a +12 °C offset relative to the ^{1}D oven, the modulator block was held at +18 °C offset relative to the ^{1}D oven. The modulation period, P_M, was 3 s with 0.75 s for both the hot and cold pulse times. The transfer line temperature was set to 285 °C and the TOFMS ion source temperature was set to 225 °C. Mass channels (*m/z*) 45-300 were collected at unit resolution with an ionization voltage of 70 eV and a collection rate of 100 spectra/s after a 10 s acquisition delay.

Figure B2. Two-dimensional chromatograms for 2-hexylthiophene. This collection of GC×GC chromatographic data shows the 2D location for 2-hexylthiophene (hit 8) and the complexity of the surrounding chemical environment. The TIC for both the (A) 30 ppm spike level and the (B) 15 ppm spike level are shown indicating there is no discernable change at a macro level between classes. (C) The same location at *m/z* 97, the most intense selective channel for the analyte, showing the analyte peak roughly in the center of the plot. (D) Lastly, this location at *m/z* 117, an intense, selective channel for the closest interference peak. Notice the difference in scale between (A) and (B) relative to (C) and (D).

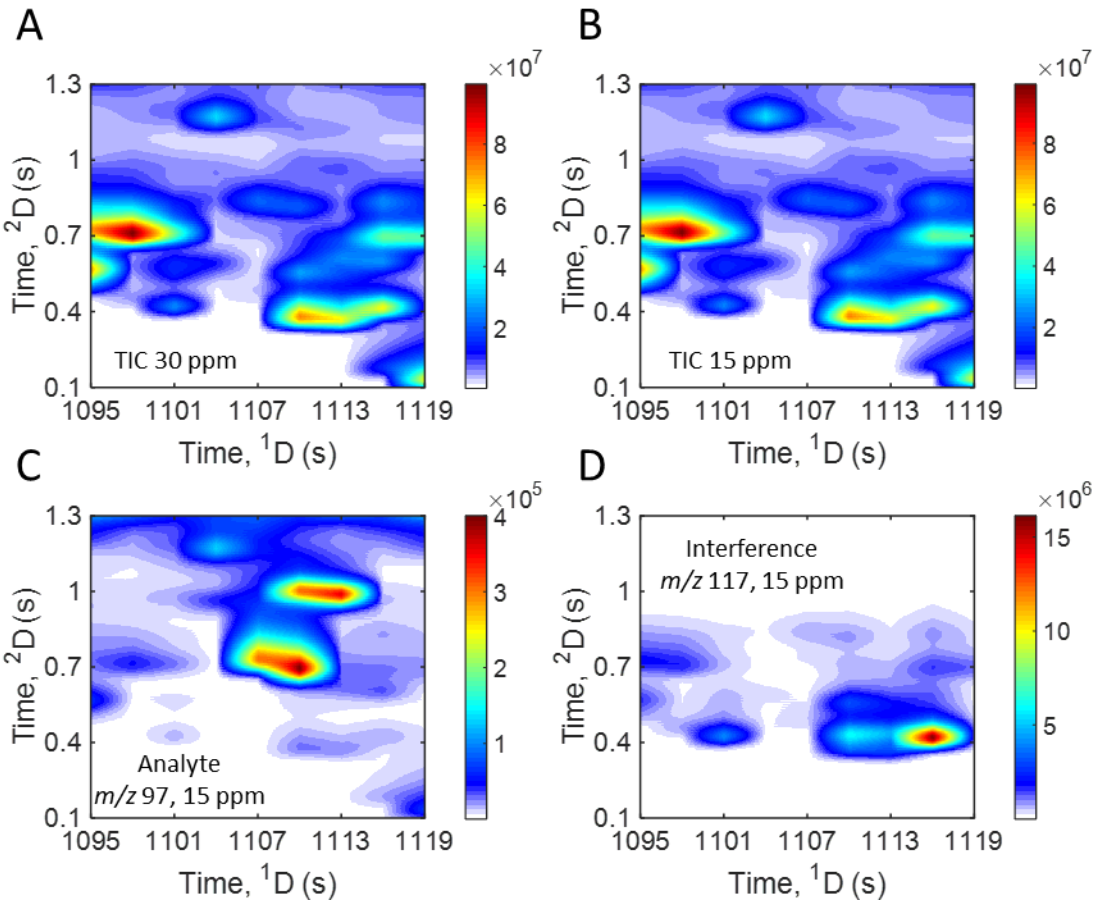


Figure B3. Two-dimensional chromatograms for 2-propylthiophene. This collection of GC×GC chromatographic data shows the 2D location for 2-propylthiophene (hit 10) and the complexity of the surrounding chemical environment. The TIC for both the (A) 30 ppm spike level and the (B) 15 ppm spike level are shown indicating there is no discernable change at a macro level between classes. (C) The same location at *m/z* 45, the most intense selective channel for the analyte, showing the analyte peak roughly in the center of the plot. (D) Lastly, this location at *m/z* 105, an intense, selective channel for the closest interference peak. Notice the difference in scale between (A), (B) and (D) relative to (C).

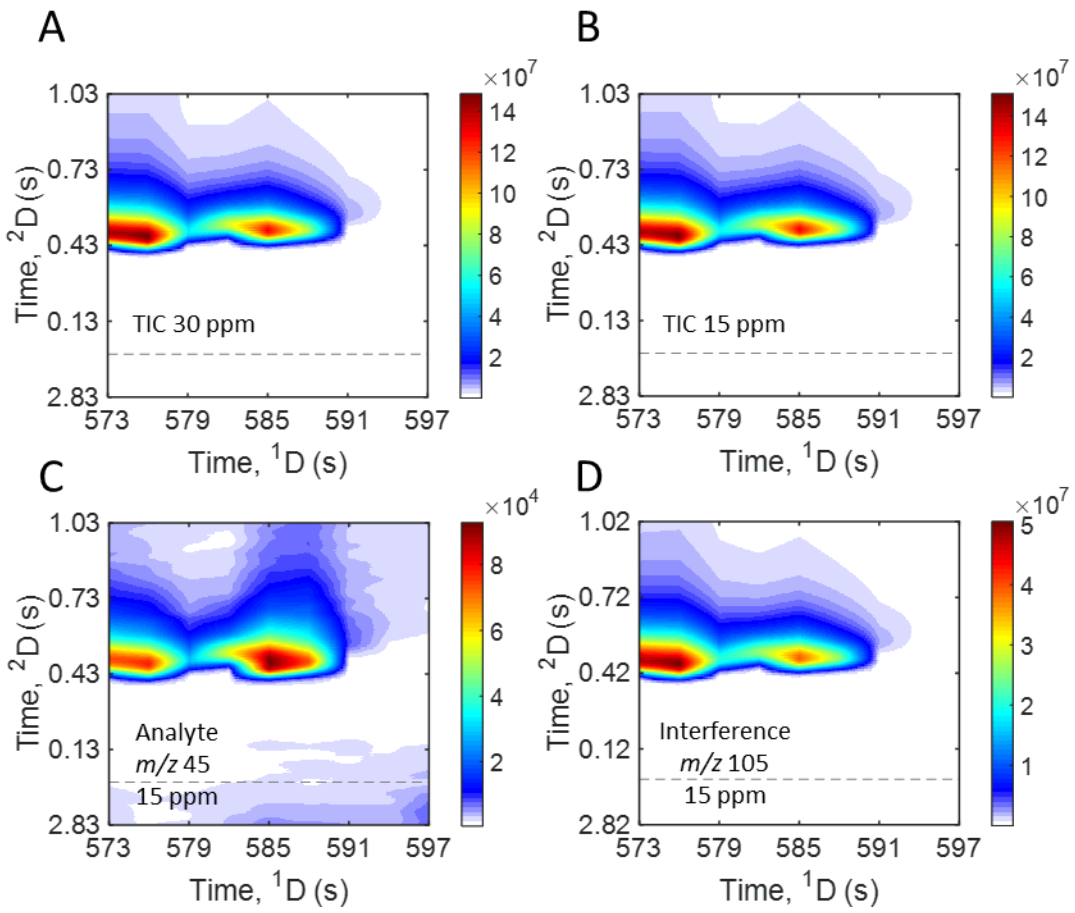


Figure B4. Two-dimensional chromatograms for 1,4-thioxane**.** This collection of GC×GC chromatographic data shows the 2D location for 1,4-thioxane (hit 12) and the complexity of the surrounding chemical environment. The TIC for both the (A) 30 ppm spike level and the (B) 15 ppm spike level are shown indicating there is no discernable change at a macro level between classes. (C) The same location at m/z 104, the most intense selective channel for the analyte, showing the analyte peak roughly in the center of the plot. (D) Lastly, this location at m/z 82, an intense, selective channel for the closest interference peak. Notice the difference in scale between (A) and (B) relative to (C) and (D).

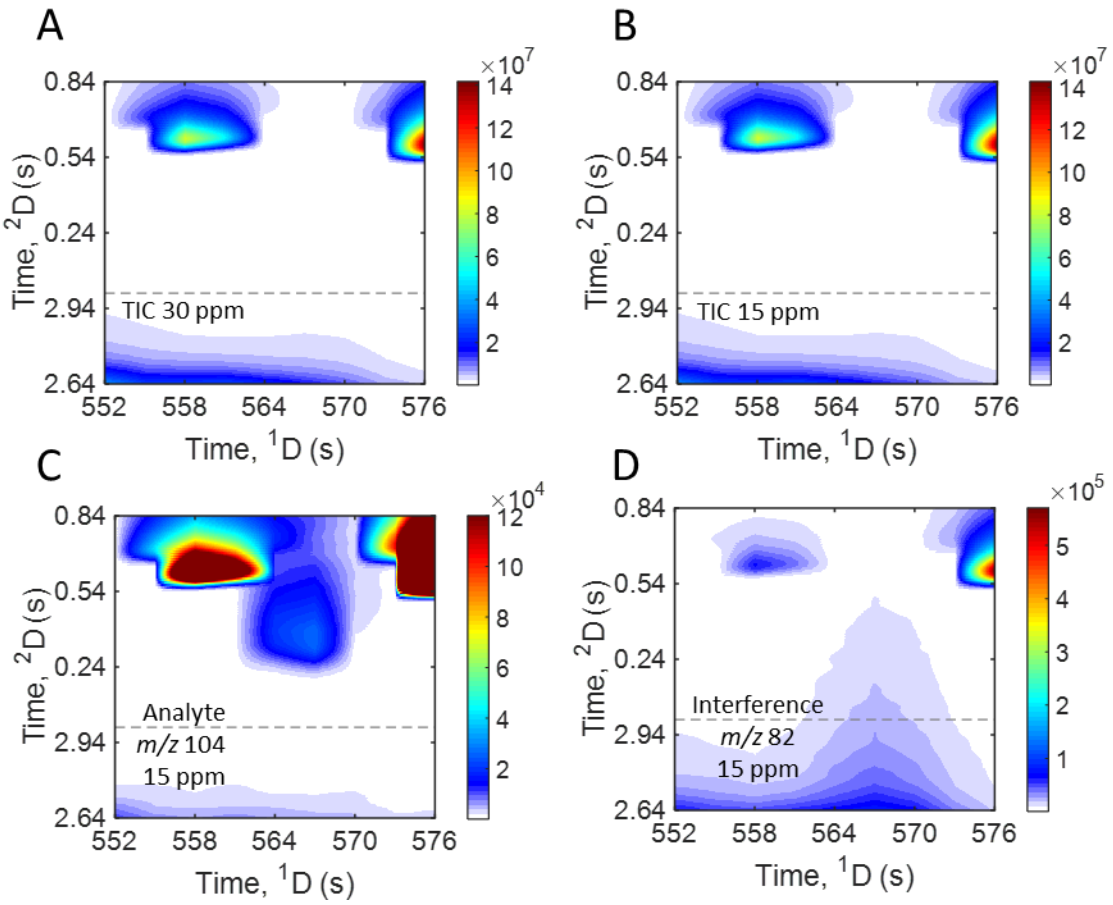


Figure B5. Alternate example of spectral normalization for 2-butyl-5-ethylthiophene. Implementation of CCE-MSP using an alternate sample pair of data for 2-butyl-5-ethylthiophene, where the $k_1/k_2 = 0.7196$, a normalization ratio well away from 1. When $S(m/z)_1$ is subtracted from $S(m/z)_2$ without normalization in such a case the residual interference spectrum significantly obscures the analyte spectrum, seen in the lower left plot, yielding a MV of 690. Applying the k_1/k_2 normalization to equalize $S(m/z)_1$ to $S(m/z)_2$ before subtraction results in the lower right plot, a purified analyte spectrum with a much higher MV of 819. This demonstrates the usefulness CCE-MSP presents when dealing with larger differences in signal due to run-to-run variation.

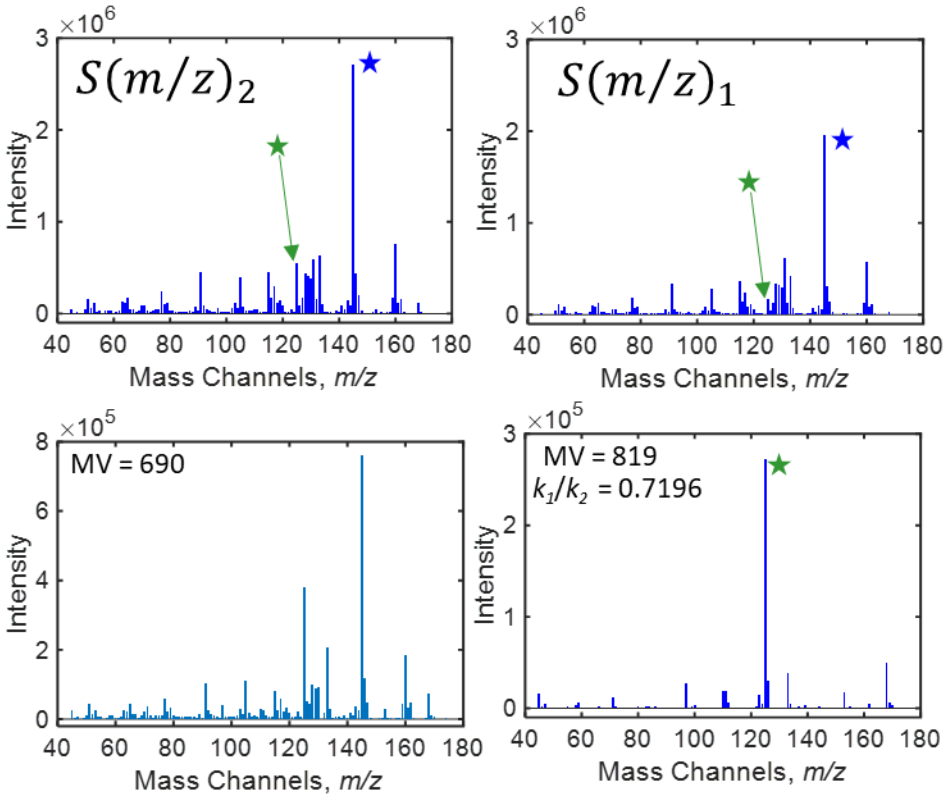
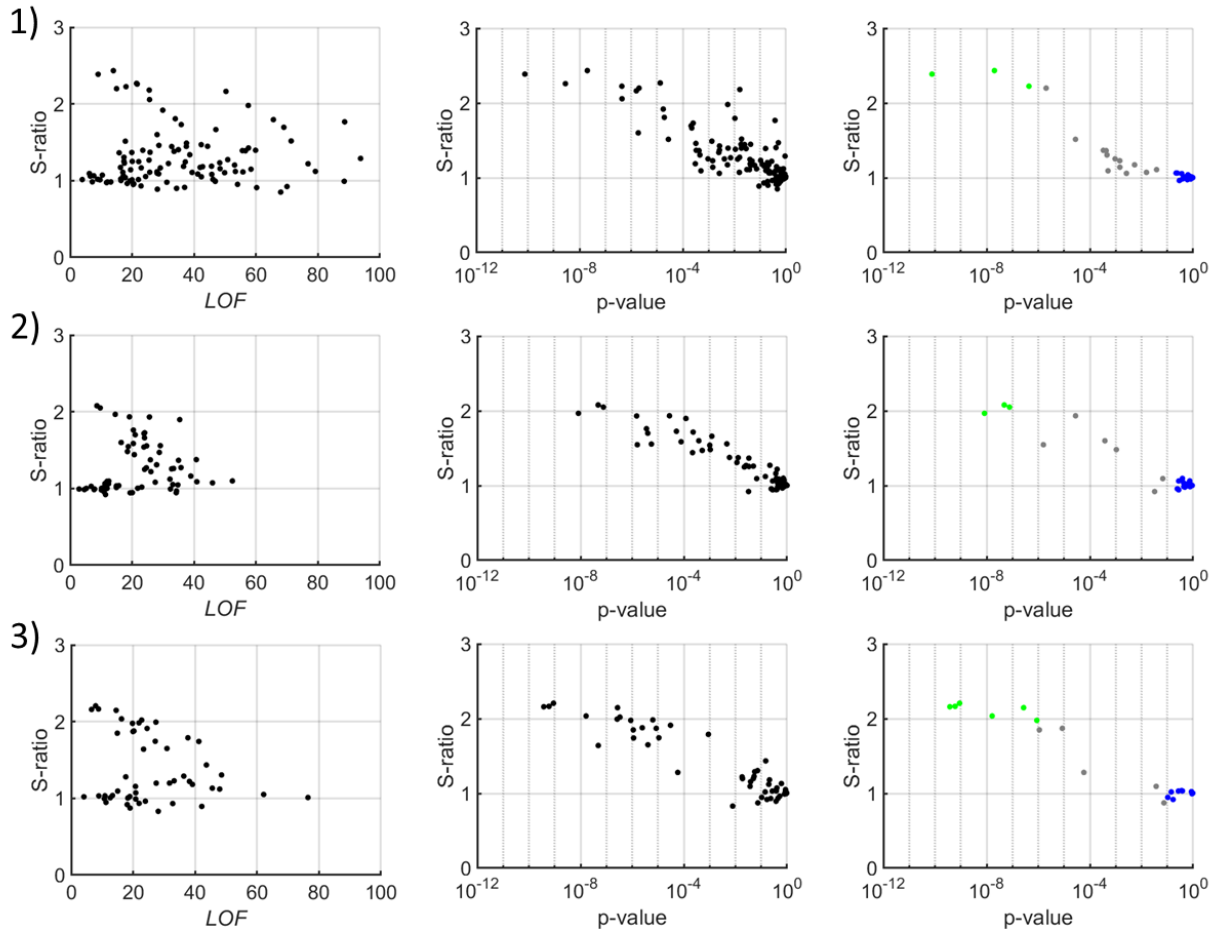
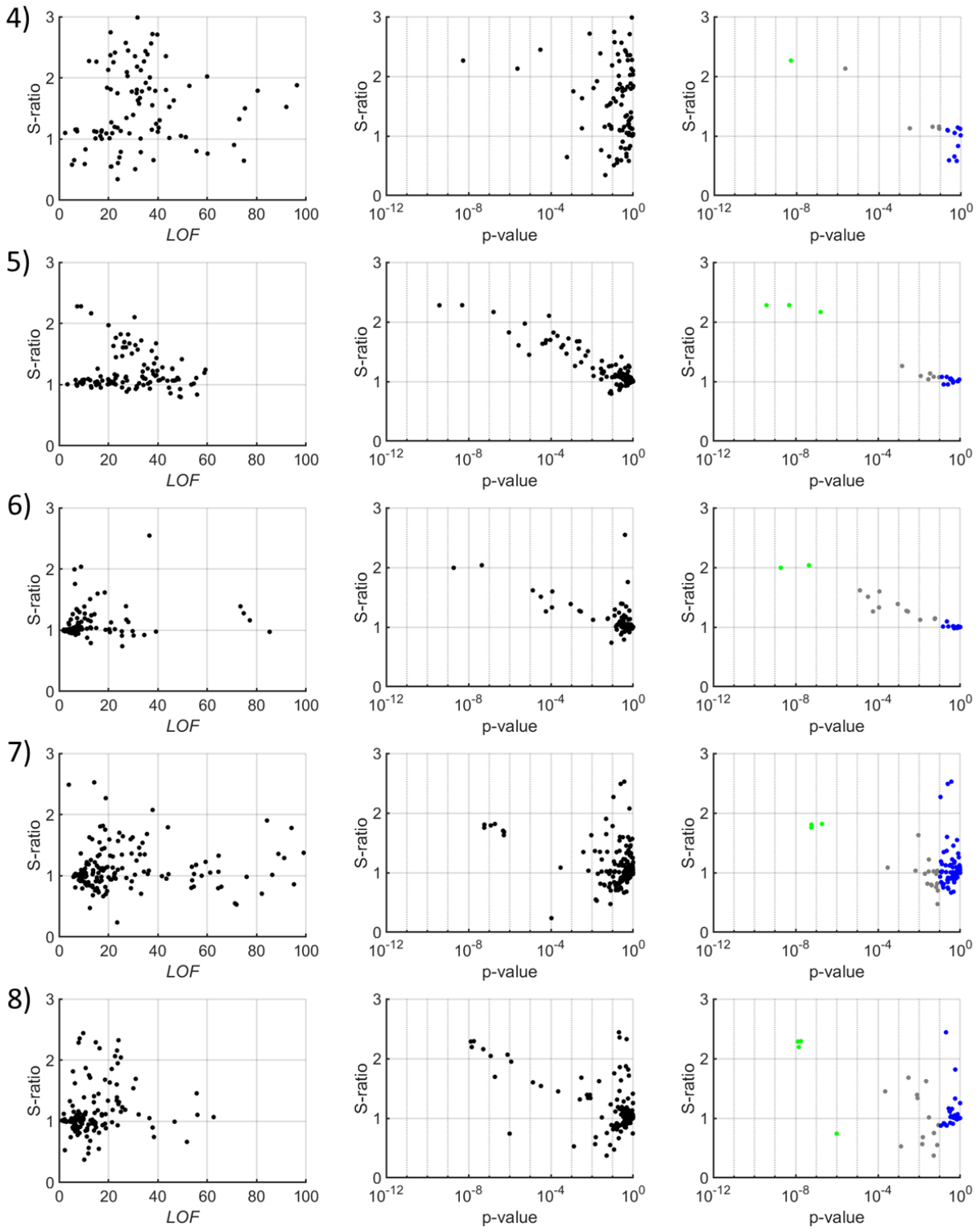


Figure B6. Statistical metrics for all of the analyte hits: (1) 2,5-dimethylthiophene, (2) 2-methylthiophene, (3) thiophene, (4) 2-chloroethyl phenyl sulfide, (5) 3-methylthiophene, (6) 2-butyl-5-ethylthiphene, (7) 2-methylbenzothiophene, (8) 2-hexylthiophene, (9) tetrahydrothiophene, (10) 2-propylthiophene, (11) benzothiophene, (12) 1,4-thioxane, (13) 3-methylbenzothiophene, and (14) 3-acetyl-2,5-dimethylthiophene. Plots display the statistical metrics for determining pure analyte and pure interference *m/z* for each hit. Dots on the rightmost plot are color coded: blue *m/z* represent pure interference *m/z*, green represent pure analyte *m/z*, and grey represent analyte *m/z* with varying contribution from interference peaks. Pure analyte *m/z* approach an S-ratio equivalent to the true concentration ratio of 2, except for two analytes which had a native concentration level in the neat fuel (hits 7 and 13) with an S-ratio < 2. There are three analyte hits which do not have any pure analyte *m/z* (hits 9, 12, 14). The transition from interference to analyte *m/z* is also seen in the S-ratio versus *LOF* plots, where transitionary *m/z* have high *LOF* due to the interference contribution.




4)
5)
6)
7)
8)
S-ratio
LOF
p-value

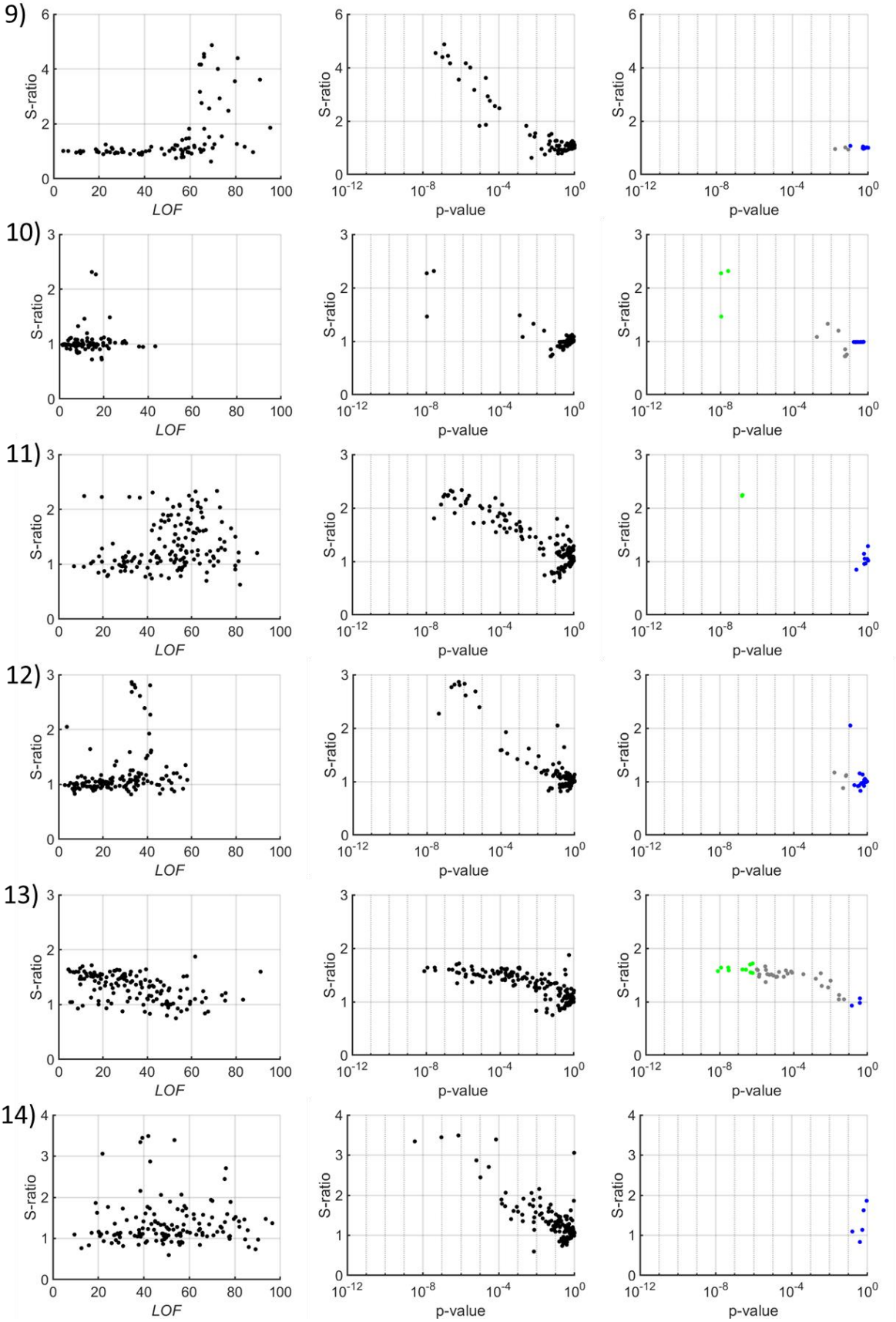
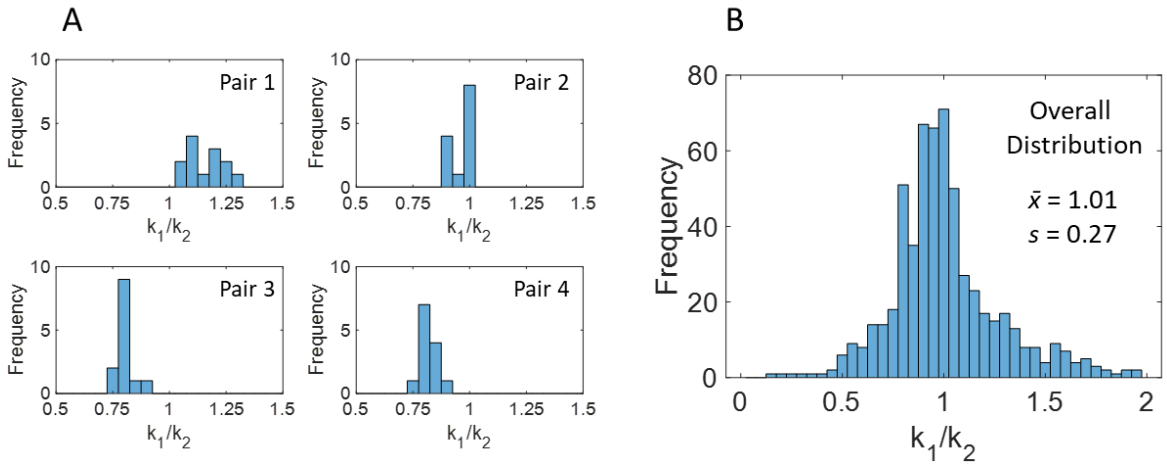

9)
10)
11)
12)
13)
14)
S-ratio
LOF
p-value

Figure B7. In-depth analysis of k_1/k_2 values

(A) Histograms displaying the distribution of k_1/k_2 ratios calculated for the 13 pure interference *m/z* for the four sample pairs for 2-butyl-5-ethylthiophene (hit 6). Notice that the average normalization varies between the pairs, indicating that selection of the 'best' k_1/k_2 is unique to a given pair of samples. The normalization step is critical to analyte spectrum purification, but on the other hand, use of the average k_1/k_2 in Eq. (10) for the S-ratio determination is not needed if a modest error in concentration ratio can be tolerated. This apparent contradiction is clarified when one considers that it is the impact of the s_{Int}/s_A in Eq. (6) coupled with the 2D chromatographic resolution between the analyte and the adjacent interference peak(s) that renders k_1/k_2 normalization critical for analyte spectrum purification. (B) The overall histogram displays the distribution of all k_1/k_2 ratios for qualifying interference *m/z* for all 14 analyte hits. With an average k_1/k_2 = 1.01 and s.d. = 0.27, the majority of interference *m/z* need only a modest normalization while others require larger corrections, although the recovery of a purified analyte spectrum is improved even for small corrections, eg., Fig. 3D.



CCE-MSP when the analyte hit is in only one of the two classes

In terms of the theoretical principles in the main paper, we consider the case in which the analyte is detected in only one of the two classes. Here, we let the analyte be in class 2, with concentration $C_{A,2}$, in order to frame the equations directly from the main paper. Now, the hit spectra $S(m/z)_1$ and $S(m/z)_2$ are given by

$$S(m/z)_1 = k_1 \sum_{i=1}^{n} R(m/z)_{Int} C_{Int,1} \quad \text{(S1)}$$

$$S(m/z)_2 = k_2[R(m/z)_A C_{A,2} + \sum_{i=1}^{n} R(m/z)_{Int} C_{Int,2}] \quad \text{(S2)}$$

for the analyte (A) and the background interference (Int), where $R(m/z)_A$ is the analyte mass spectrum, and $R(m/z)_{Int}$ is the interference spectrum with an overall interference concentration of $C_{Int,1}$ and $C_{Int,2}$, for class 1 and class 2, respectively. Thus, under the conditions stated above, $C_{Int,1} = C_{Int,2}$, with the interference taken as invariant between classes. However, $C_{A,1} = 0$ while $C_{A,2} > 0$ which facilitates analyte discovery by F-ratio analysis. Once again, by assuming $k_1 = k_2$ the "initial difference" of Eq. (1) from Eq (2) produces a rough estimate of the analyte spectrum,

$$S(m/z)_2 - S(m/z)_1 = \textit{Analyte Spectrum + Residual Interference Spectrum} =$$
$$= \{k_2 R(m/z)_A C_{A,2}\} + \{k_2 \sum_{i=1}^{n} R(m/z)_{Int} C_{Int,2} - k_1 \sum_{i=1}^{n} R(m/z)_{Int} C_{Int,1}\} \quad (3)$$

with the analyte spectrum given by,

$$\textit{Analyte Spectrum} = k_2 R(m/z)_A C_{A,2} \quad (4)$$

and the residual interference spectrum given by,

$$\textit{Residual Interference Spectrum} = k_2 \sum_{i=1}^{n} R(m/z)_{Int} C_{Int,2} - k_1 \sum_{i=1}^{n} R(m/z)_{Int} C_{Int,1} \quad (5)$$

As with analyte in both classes, to obtain the analyte spectrum when the analyte is in only one of the two classes, pure *m/z* for the background interference are identified using the p-value and *LOF* information. Then, the ratio k_1/k_2 is calculated using the signal at the pure interference *m/z* from the hit spectra $S(m/z)_1$ and $S(m/z)_2$,

$$Normalization\ Factor = \frac{s(m/z)_1}{s(m/z)_2} = \frac{k_1}{k_2} \quad (6)$$

where s(*m/z*) is a scalar. Normalization factors from Eq. (6) can be applied to eliminate the residual interference spectrum contamination from Eq. (5),

$$Residual\ Int.\ Spectrum = \frac{k_1}{k_2} k_2 \sum_{i=1}^{n} R(m/z)_{\mathrm{Int}} C_{\mathrm{Int},2} - k_1 \sum_{i=1}^{n} R(m/z)_{\mathrm{Int}} C_{\mathrm{Int},1} = 0 \quad (8)$$

Following normalization of the hit spectrum $S(m/z)_2$, subtraction of the hit spectrum $S(m/z)_1$ from the normalized hit spectrum $S(m/z)_2$, reveals the purified analyte spectrum,

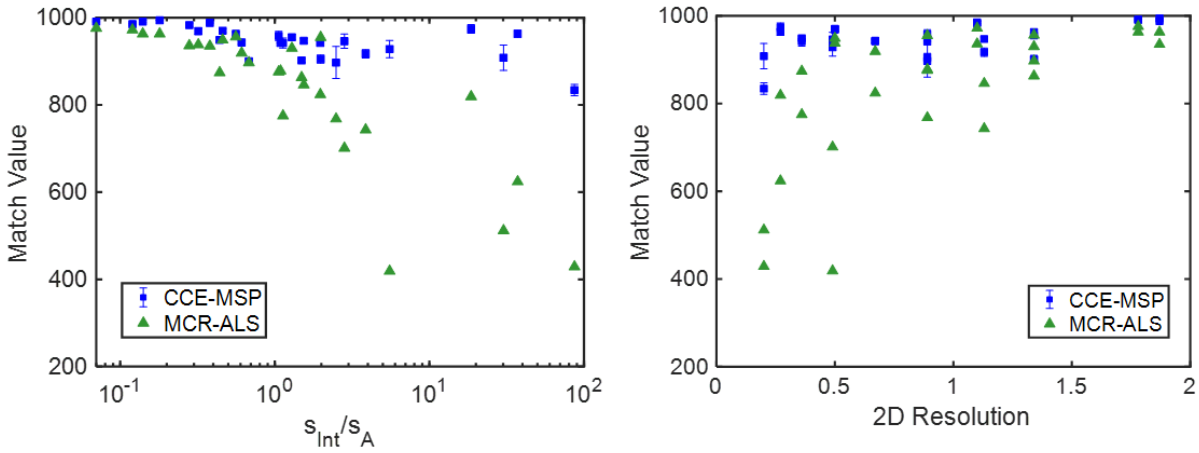
$$(\frac{k_1}{k_2}) S(m/z)_2 - S(m/z)_1 = R(m/z)_{\mathrm{A}} k_1 (C_{\mathrm{A},2}) \quad (9)$$

whereby the analyte spectrum $R(m/z)_{\mathrm{A}}$ is scaled by the quantity $k_1(C_{\mathrm{A},2})$.

Table B1. Match value table for neat fuel comparisons**.** Match values for the spectra obtained from the CCE-MSP results calculated for the 30 ppm versus neat and the 15 ppm versus neat comparisons. The 2D resolution for each analyte is the same for both classes while the s_{Int}/s_A is calculated for each class comparison individually, otherwise this table is constructed like Table 2 in the main text.

			30 ppm				15 ppm			
Hit No	**Analyte Name**	**2D Resolution**	**s_{int}/s_a**	**MV Initial**	**MV CCE-MSP**	**MV MCR-ALS**	**s_{int}/s_a**	**MV Initial**	**MV CCE-MSP**	**MV MCR-ALS**
1	2,5-dimethylthiophene	0.89	1.09	953 ± 3	942 ± 1	956	1.98	916 ± 14	905 ± 9	930
2	2-methylthiophene	1.34	0.56	951 ± 3	963 ± 4	819	1.29	946 ± 4	955 ± 6	624
3	thiophene	1.87	0.18	993 ± 1	994 ± 1	963	0.38	990 ± 1	988 ± 1	935
4	2-chloroethyl phenyl sulfide	0.36	0.44	850 ± 47	949 ± 5	846	1.13	898 ± 12	942 ± 12	743
5	3-methylthiophene	1.34	0.68	885 ± 12	900 ± 5	972	1.49	874 ± 32	902 ± 4	936
6	2-butyl-5-ethylthiophene	0.2	30.17	749 ± 50	908 ± 29	897	86.78	731 ± 88	834 ± 13	863
7	2-methylbenzothiophene	0.5	0.46	976 ± 3	970 ± 5	878	0.32	963 ± 1	969 ± 7	955
8	2-hexylthiophene	0.49	2.82	975 ± 5	946 ± 16	976	5.53	934 ± 21	928 ± 20	963
9	tetrahydrothiophene	0.89	1.06	900 ± 15	958 ± 10	512	2.49	838 ± 56	897 ± 37	429
10	2-propylthiophene	0.27	18.61	708 ± 124	974 ± 10	874	37.22	719 ± 194	963 ± 8	775
11	benzothiophene	1.1	0.12	989 ± 2	985 ± 2	876	0.28	981 ± 1	983 ± 5	768
12	1,4-thioxane	1.13	1.54	947 ± 15	947 ± 5	949	3.87	920 ± 13	917 ± 9	938
13	3-methylbenzothiophene	1.78	0.07	984 ± 1	991 ± 2	701	0.14	981 ± 1	991 ± 3	419
14	3-acetyl-2,5-dimethylthiophene	0.67	0.61	907 ± 9	943 ± 5	919	1.97	862 ± 19	942 ± 4	824

Figure B8. Match value plots for 30 ppm vs neat and 15 ppm vs neat comparisons**.** Plots of the MV data for the 30 ppm vs neat and 15 ppm vs neat F-ratio analysis comparisons presented in Table S1. As with the 30 ppm vs 15 ppm comparison summarized in Fig. 7, the MV for MCR-ALS and the initial difference spectrum decreases as s_{Int}/s_A increases, indicating the complexity from overlapping interferences is directly hampering the effectiveness of MCR-ALS, whereas the MVs provided by CCE-MSP remain consistently high despite the interferences. This effect is mirrored and/or compounded by 2D resolution. While the 30 ppm MVs are generally higher than the 15 ppm MVs, due to higher *S/N* of the analyte, the MV still decreases when 2D resolution is lower.



Description of the GC×GC-TOFMS data simulations and Table S2

Ten GC×GC-TOFMS chromatograms with 5 components were simulated in which two components were changing in signal intensity by a factor of two between classes (2-fold higher in the second sample class) while the other three components did not change in signal intensity between sample classes. Six sample replicates were simulated for both sample classes for each simulation, to provide the data for F-ratio analysis. The 2D size of each chromatogram simulated was 36 s on ^{1}D with a 3 s modulation period on ^{2}D at a TOFMS collection frequency of 100 Hz. Retention times for each of the 5 component peaks were randomly assigned between 14 s and 22 s on ^{1}D and between 140 ms and 170 ms on ^{2}D, in order to generate a significant degree of peak overlap for all 5 components. Peaks were simulated by first generating a Gaussian peak based upon the randomly assigned ^{1}D retention time with a constant $^{1}w_b$ of 9 s. Then, the modulated ^{2}D peaks were created by generating Gaussian peaks for each modulation based upon the randomly assigned ^{2}D retention time, with a constant $^{2}w_b$ of 180 ms, and the amount of area captured in each modulation by the ^{1}D peak profile. Total peak areas for the 5 components of each chromatogram were randomly selected from an exponential function [1]. These peak areas represent the mean peak areas per class for each component. Also, the areas for each of the 6 replicates per sample class were randomly assigned with a relative standard deviation in signal intensity of 10%, which defined the within-class variance in the subsequent F-ratio calculations. The fully modulated peak profiles for each of the 5 components for a given chromatogram were then multiplied across a random mass spectrum selected from a library, creating the GC×GC-TOFMS data for each component [2]. Each component was then added to each other to create one replicate chromatogram. For these simulations there was no retention time shifting introduced thus each component appeared at the same retention time in each replicate. This chromatogram generation process was repeated for all replicates. In total 10 different 6 versus 6 sample class comparisons were generated followed by F-ratio analysis with one example shown in the main text for brevity. Results for all 10 simulations are presented in Table S2. The average analyte-to-analyte 2D resolution = 0.37 with s = 0.17, for the two analytes that change between classes, while the mean 2D resolution = 0.67 with s = 0.29 for all pairs between the 5 components in the separations. The average MV for both analytes across all 10 simulations was 743, with s = 295, for the CCE-MSP outputs (per Figure 7B in the main paper), while after applying ALS to the CCE-MSP outputs the MV improved significantly to an average MV = 972,

s = 48 (per Figure 7C in the main paper). Note that all of the simulations produce a substantial improvement in MV for one or both of the analyte hits. The variability in the MV improvement is due to the randomness in the simulation parameters applied. The low analyte-to-analyte 2D resolution and analyte mass spectrum similarity (the MV between the two analyte hits) all play a role in making each simulation challenging to extract high quality spectra.

Table B2. Chromatographic resolution situations and MV for the 10 simulated chromatograms. Analyte to analyte resolution and total mean resolution (average of all resolutions pairwise) demonstrate the high overlap of all components. Match values for each analyte hit are given for the CCE-MSP spectra results and for the results of the application of ALS to the CCE-MSP spectra.

Rep No.	Analyte to Analyte R_s	Total Mean R_s	Analyte Similarity	MV Analyte 1 CCE-MSP	MV Analyte 2 CCE-MSP	MV Analyte 1 ALS	MV Analyte 2 ALS
1	0.51	1.13	63	932	858	996	897
2	0.57	1.04	27	453	970	984	997
3	0.16	0.96	42	674	984	999	982
4	0.34	0.57	82	561	998	1000	1000
5	0.31	0.48	740	827	1000	999	999
6	0.31	0.64	74	321	1000	840	999
7	0.65	0.62	60	573	1000	873	999
8	0.42	0.59	22	179	1000	940	999
9	0.32	0.42	19	995	320	999	995
10	0.13	0.24	27	265	957	944	999

References

[1] J.M. Davis, Extension of Statistical Overlap Theory to Poorly Resolved Separations, Anal. Chem. 69 (1997) 3796–3805. https://doi.org/10.1021/AC9701391.

[2] J.M. Davis, J.C. Giddings, Statistical Theory of Comopnent Overlap in Multicomponent Chromatograms, Anal. Chem. 55 (1983) 8. https://pubs.acs.org/sharingguidelines (accessed August 13, 2021).

Appendix C

This Appendix is reproduced from the Electronic Supplementary Material of Grant S. Ochoa, Matthew C. Billingsley, Robert E. Synovec, "Using solid phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: Comparative analysis of aerospace fuel composition" *Anal. Bioanal. Chem.* (2022) 1-13.

Instrument Conditions. Samples were analyzed with the Pegasus BT 4D GC×GC-TOFMS (LECO Corporation, St. Joseph, MI) instrument using a quad-jet thermal modulator, an Agilent 7890 gas chromatograph (Agilent Technologies, Palo Alto, CA) and an L-Pal3 GC autosampler. Six replicates were collected for the neat and pass samples of the RP-1 fuel B0112868 and the three RF2 fuel formulations RF-A, RF-B, and RF-C. A 0.5 µL aliquot of each replicate was injected at a split ratio of 200:1 in the GC inlet, which was held constant at 275°C. The ^{1}D column (26.4 m × 250 µm i.d. × 0.25 µm film thickness) had a mid-polarity Rxi-17Sil MS stationary phase of phenyl crosslinked dimethyl polysiloxane (Restek, Bellefonte, PA) and the ^{2}D column (1.9 m × 180 µm i.d. × 0.18 µm film thickness) had a non-polar Rxi-1 MS stationary phase (100% dimethylpolysiloxane). Ultrahigh purity helium (Grade 5, 99.999%, Praxair, Seattle, WA, USA) was the carrier gas at a constant flow of 2.0 mL/min. The ^{1}D column was held at 40°C for 1.5 min, increased to 200°C at 5°C/min, and held at 200°C for 1 min. The ^{2}D column and modulator block utilized the same temperature program with offsets of +12°C and +30°C, respectively. The modulation period, P_M = 3 s, with a hot and cold pulse time of 0.75 s each. After an acquisition delay of 10 s, *m/z* 45-200 were collected at a collection frequency of 100 spectra/s with a detector voltage of 2005 V. The ion source and transfer line were held constant at 225°C and 285°C, respectively.

Figure C1. SPE Procedure Workflow. Diagram showing the workflow for the SPE procedure for removing polar compounds from the fuel samples. The silica based SPE cartridge is first conditioned with hexane, after which the neat fuel is loaded, concentrating the polar compounds into a band on top of the cartridge (displayed as a pink band resulting from the fuel dye). The polar deficient fuel is collected from the cartridge and the first mL off is isolated as the 'pass' fuel sample. The pass and neat samples are subjected to GC×GC-TOFMS for data collection then F-ratio analysis to discover the extracted compound locations in the original neat fuel. Separately, the SPE cartridge is washed with hexanes to remove residual non-polar compounds after which the polar compounds are eluted with a minimal volume of methanol to obtain the fuel extract sample, nominally the low-level polar species.

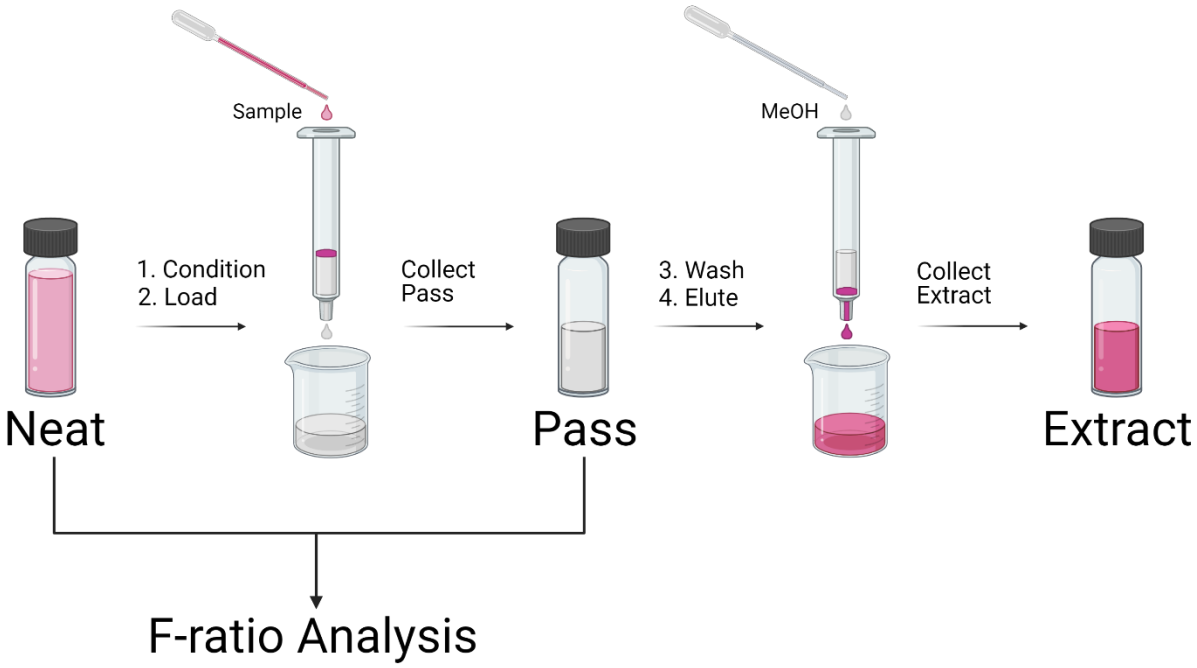


Figure C2. Stitch Chromatograms for RP-1 1v1 analyses. Stitch chromatograms (a-f) of the top 50 hits for the 1v1 analyses for each of the six chromatogram pairs of fuel B0112868 displaying the entirety of the chromatographic space. In addition to the hits in the region of interest the effect of the 1v1 analysis variation can be seen with other peaks appearing in the top 50 hits.

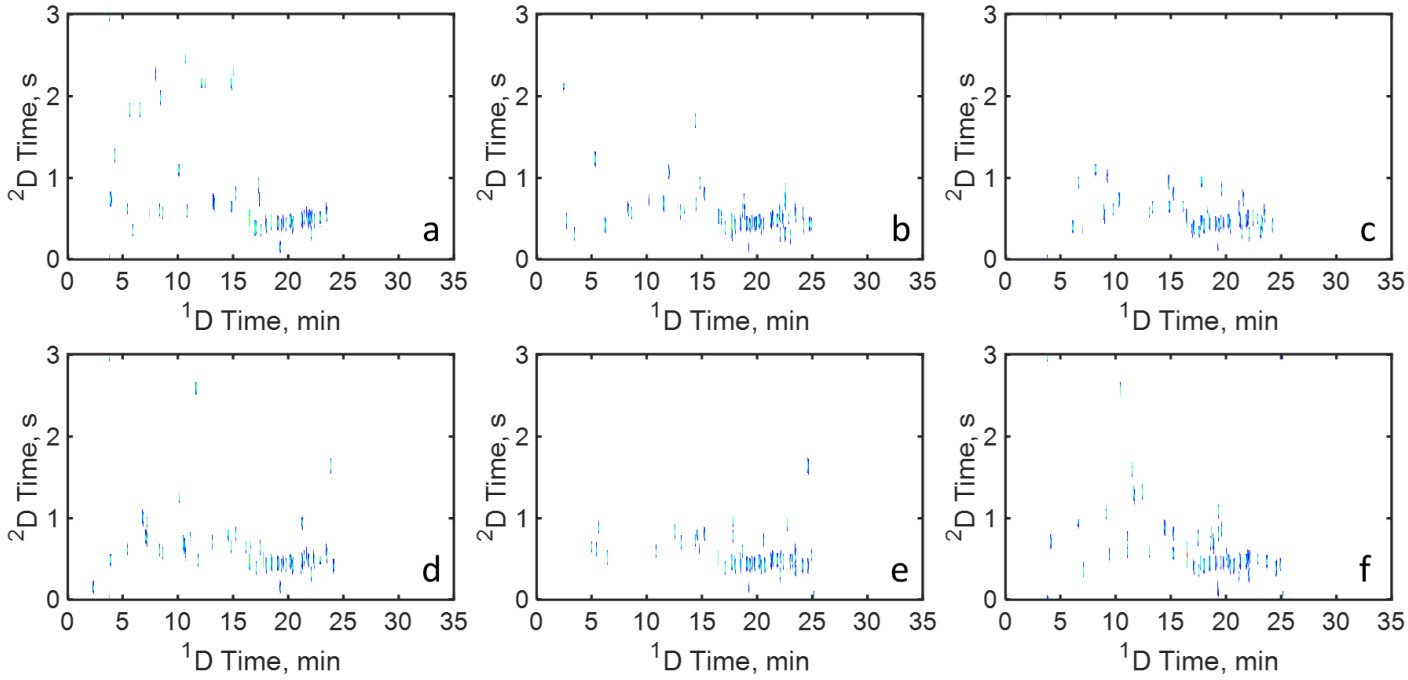


Figure C3. Neat vs Neat Comparisons Between RF-A and RF-B. (a) GC×GC-TOFMS TIC chromatograms of rocket fuel RF-A with the hit markers for the neat versus neat comparison between RF-A and RF-B, and (b) the hit markers for the pass versus pass comparison for the same pair. (c) The hits due to low-level polar species are only present in the neat versus neat comparison but not the pass versus pass comparison and are plotted over the 1:10 diluted RF-A extract chromatogram. Notice that not all the compounds along the polar band are identified as hits, meaning these compounds are present in each fuel to a similar extent. (d) The pass versus pass comparison hits are plotted over the TIC chromatogram of fuel RF-A with a more intense color scale.

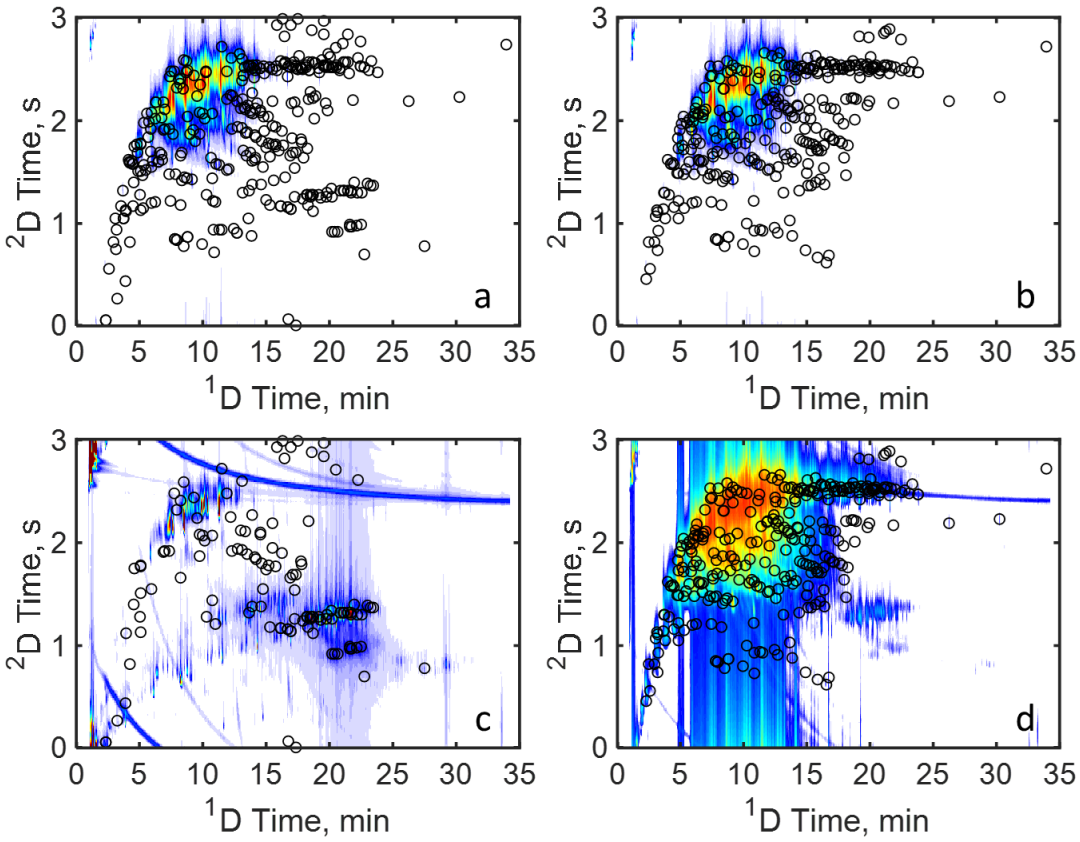


Figure C4. Neat vs Neat Comparisons Between RF-B and RF-C. (a) GC×GC-TOFMS TIC chromatograms of rocket fuel RF-B with the hit markers for the neat versus neat comparison between RF-B and RF-C, and (b) the hit markers for the pass vs pass comparison for the same pair. (c) The hits due to low-level polar species are only present in the neat versus neat comparison but not the pass versus pass comparison and are plotted over the 1:10 diluted RF-B extract chromatogram. Nearly every peak along the polar compound band has a hit marker, indicating one fuel (RF-B) had significantly more low-level polar species from the other fuel (RF-C). (d) The pass versus pass comparison hits are plotted over the TIC chromatogram of fuel RF-B with a more intense color scale.

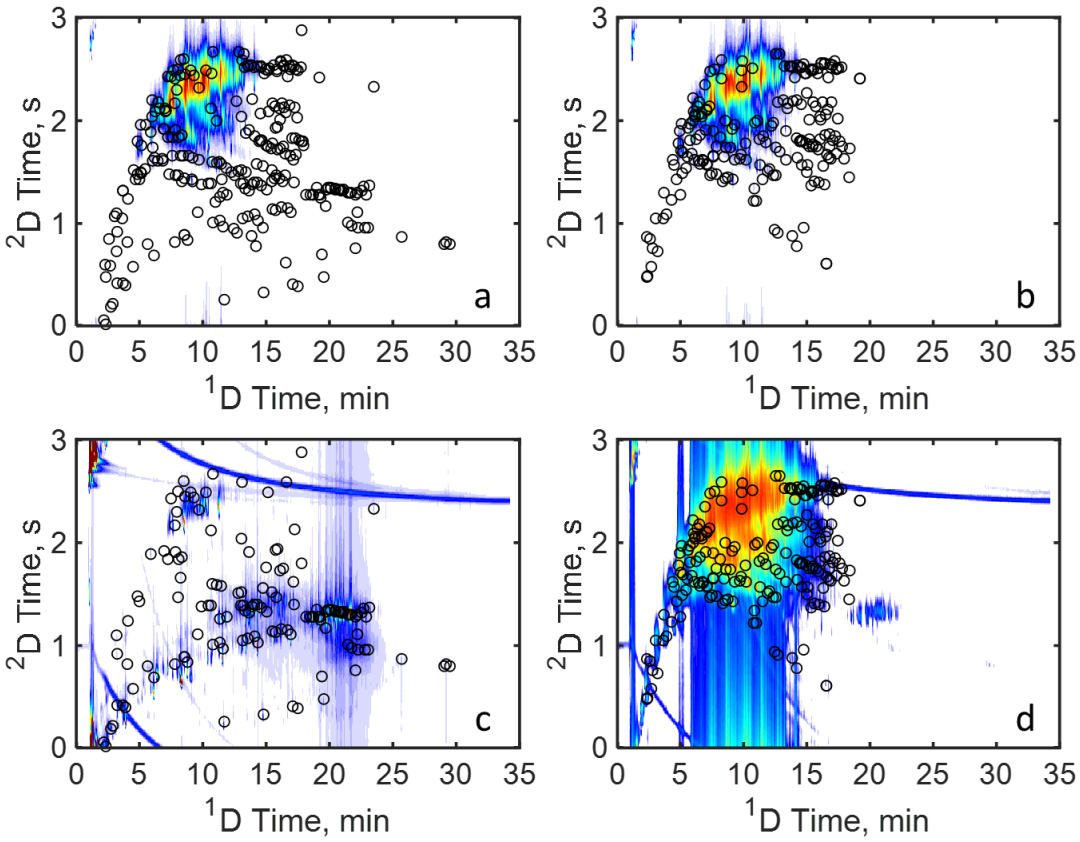


Figure C5. Signal-ratio Stitch Chromatograms for Inter-Fuel Comparisons. The bulk hydrocarbon hits (a-c) and low-level polar species hits (d-f) of the three inter-fuel comparison pairs from left to right, RF-A versus RF-B, RF-A versus RF-C, and RF-B versus RF-C. The chromatograms are generated by first calculating a stitch chromatogram for each fuel in each pair using either the pass versus pass hit list (bulk hydrocarbon) or low-level polars hit list. The two stitch chromatograms are then divided by each other with the first fuel in the pair divided by the second, yielding these signal-ratio (i.e., apparent concentration ratio) stitch chromatograms that detail the relative concentration of differences found between the fuels. This approach allows an analyst to gain a sense of what compounds differ between classes at a glance for the entire chromatogram. For example, fuel RF-A is more concentrated in long chain alkanes than either RF-B or RF-C as seen in (a) and (b). Since the polar compound band is blue for chromatograms (d-f) this means that fuel RF-C, which is always in the denominator for the signal-ratio calculation, has either much less concentration than either of the other fuels or is completely lacking these compounds. In such cases any zero in the stitch chromatograms is replaced with a 1 to ensure an accurate signal-ratio direction. Furthermore, the bulk hydrocarbon stitch chromatograms (a-c) give information to differences in carbon number distribution, which may serve as a benchmark for the efficiency of a production method or provide insight to differences in physical properties.