**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**

Brendon A. Parsons,^a David K. Pinkerton,^a Bob W. Wright,^b and Robert E. Synovec^{a,*}

(a) Department of Chemistry, University of Washington, Box 351700, Seattle, WA 98198,
U.S.A.

(b) Pacific Northwest National Laboratory, Battelle Boulevard, P.O. Box 999, Richland, WA
99352, U.S.A.

Revised version of JCA-15-2344, submitted for consideration to publish

in *Journal of Chromatography A*

February 18, 2016

* CORRESPONDING AUTHOR: Tel.: +1-206-685-2328; fax: +1-206-685-8665

EMAIL: synovec@chem.washington.edu

Abstract

The illicit chemical alteration of petroleum fuels is of keen interest, particularly to regulatory agencies that set fuel specifications, or taxes/credits based on those specifications. One type of alteration is the reaction of diesel fuel with concentrated sulfuric acid. Such reactions are known to subtly alter the chemical composition of the fuel, particularly the aromatic species native to the fuel. Comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry ($\text{GC} \times \text{GC}$ -TOFMS) is well suited for the analysis of diesel fuel, but may provide the analyst with an overwhelming amount of data, particularly in sample-class comparison experiments comprised of many samples. Tile-based Fisher-ratio (F-ratio) analysis reduces the abundance of data in a $\text{GC} \times \text{GC}$ -TOFMS experiment to only the peaks which significantly distinguish the unaltered and acid altered sample classes. Three samples of diesel fuel from differently branded filling stations were each altered to discover chemical features, i.e., analyte peaks, which were consistently changed by the acid reaction. Using different fuels prioritizes the discovery of features likely to be robust to the variation present between fuel samples and may consequently be useful in determining whether an unknown sample has been acid altered. The subsequent analysis confirmed that aromatic species are removed by the acid alteration, with the degree of removal consistent with predicted reactivity toward electrophilic aromatic sulfonation. Additionally, we observed that alkenes and alkynes were also removed from the fuel, and that sulfur dioxide or compounds that degrade to sulfur dioxide are generated by the acid alteration. In addition to applying the previously reported tile-based F-ratio method, this report also expands null distribution analysis to algorithmically determine an F-ratio threshold to confidently select only the features which are sufficiently class-distinguishing. When applied to the acid alteration of diesel fuel, the suggested per-hit F-ratio threshold was 12.4, which is predicted to maintain the false discovery rate (FDR) below 0.1%. Using this F-ratio threshold, 107 of the 3362 preliminary hits were deemed significantly changing due to the acid alteration, with the number of false positives estimated to be about 3. Validation of the F-ratio analysis was performed using an additional three fuels.

Keywords: tile-based F-ratio, $\text{GC} \times \text{GC}$ -TOFMS, null distribution analysis, diesel fuel, acid alteration, false discovery rate

1. Introduction

Petrochemical fuels are refined by a variety of methods that allow for inherently variable feedstocks to be better controlled in finished products to meet a range of quality metrics and other specifications, such as engine performance and longevity, storage stability, and emissions. Such processes and specifications ensure that the fuels are acceptable for distribution and end use by consumers. However, between the refinery and end user, intermediate parties may further alter the chemical composition of finished fuels, for the purpose of altering the fuel color [1]. Hence, there is substantial interest in the forensic detection of such alteration by law enforcement and regulatory agencies. An alteration that has been observed in the field is the reaction of petroleum fuels with concentrated sulfuric acid (H_2SO_4). Sulfuric acid is a prevalent and inexpensive reagent used in a wide variety of industrial, commercial, and household applications. Based on the chemical composition of petroleum fuels, one would expect reaction with sulfuric acid to cause a variety of chemical changes in fuel due to reactions between the sulfuric acid and native compounds in the fuel [2–4]. Gasoline fuel exhibits substantial chemical changes due to the presence of *tert*-butylating compounds, such as methyl *tert*-butyl ether (MTBE), while the changes to diesel fuel are more subtle, limited to the aromatic moiety of the fuel [2,4]. For the purposes of forensic classification of acid altered diesel fuel, it is necessary to better understand the chemical compositional changes that occur and which may differentiate unaltered from altered fuels.

Reactions between sulfuric acid and aromatic compounds are well represented in the literature, particularly in the fields of surfactant and dye chemistry, in which electrophilic species in the sulfuric acid react with the aromatics to add sulfonic acid, altering the aqueous solubility of the aromatics [5]. Literature on sulfonation reactions using sulfuric acid indicate that various

aromatic species have an intrinsically different reactivity with the electrophilic species in the concentrated acid, and thus will be chemically modified to different extents as the reaction proceeds [6–8]. An interesting aspect of aromatic sulfonation by sulfuric acid addition is that the reaction produces water, leading to a quenching of the reaction as concentrations of the electrophilic species decrease. Most studies on the reaction rates and extent of reactions for aromatic sulfonation using sulfuric acid involve simple reagent systems (i.e., neat reagent or reagent dissolved in a suitable solvent) and an excess of sulfuric acid (typically 3–4 molar equivalents). In the forensic study of acid altered diesel fuel, these ideal conditions are absent: first, the speciation of aromatics in diesel is highly complex, often ranging from benzene to pyrene with various degrees of alkylation and unsaturate substitution, as well as lesser amounts of aromatic heterocycles. Additionally, the sulfuric acid is probably not added stoichiometrically, and factors such as the water content of the diesel fuel may also impact the rates and extent of reaction. Further, industrial syntheses typically include temperature control and removal of water from the sulfuric acid to assist in driving the aromatic sulfonation to completion; such controls may not be present in illicit alterations.

Due to the substantial differences expected between the literature reaction conditions for synthesis of sulfonated aromatics versus the less-ideal conditions for illicit alteration of diesel fuel, a non-targeted “discovery-based” analysis would be beneficial for characterizing the chemical changes that result from the alteration. In non-targeted analysis, the chemical features (i.e., analyte peaks) of interest are not presumed to be known prior to the experiment. Non-targeted approaches may be performed using supervised or unsupervised methods, where supervision refers to the prior classification of samples as they relate to the experimental design [9–11]. Overall, non-targeted methods aim to discover the key chemical features, i.e., the analyte

peaks indicative of chemical marker compounds that robustly distinguish samples from each other. For the purpose of analyzing the chemical composition of diesel fuel, comprehensive two-dimensional (2D) gas chromatography coupled with time-of-flight mass spectrometry (GC $\times$ GC–TOFMS) provides the necessary analytical resolution, sensitivity, and sample throughput [12–19]. Recently, a comprehensive and informative review was reported, dealing with the application of GC $\times$ GC in forensics [20]. Indeed, GC $\times$ GC is increasingly being utilized in forensic studies, but it is acknowledged that the complexity of the data is a major factor in slowing the technology’s implementation. To improve the utility of GC $\times$ GC for all users, and to increase its utilization in forensic applications, it is necessary to develop feature selection software to reduce the complex output to only the most significant chemical features.

We previously reported the development of tile-based Fisher ratio (F-ratio) analysis for the comprehensive discovery-based analysis of GC $\times$ GC–TOFMS data [21,22]. This methodology, herein referred to as tile-based F-ratio analysis, is a supervised non-targeted method that discovers features in a sample-class comparison (e.g., unaltered fuel versus acid altered fuel) that are class-distinguishing, i.e., peaks that are statistically different between sample classes. The tiling portion of the software utilizes a novel binning scheme which improves the discovery of class-distinguishing chemical features (“hits”) while minimizing the discovery of features which are not significantly different between sample classes (“non-hits”). Tile-based F-ratio analysis coupled with GC $\times$ GC–TOFMS is highly effective for the experimental determination of chemical changes in the highly complex matrix of diesel fuel.

In the process of feature selection using tile-based F-ratio analysis, the chromatograms for the sample-class comparison experiment are comprehensively analyzed, with only the most essential application of signal thresholds, enabling the discovery of class-distinguishing analytes,

even at the trace concentration level. The F-ratio, which is the calculated value for the one-way analysis of variance (ANOVA) F -test [23] is used to rank chromatographic features in a hit list according to their class-distinguishing abilities, with a large F-ratio indicating that a given feature is highly class-distinguishing, and a small F-ratio indicating that a given feature is not class-distinguishing. In order to avoid false positives (type I errors) and to utilize analyst time most efficiently, only features with a sufficiently high F-ratio should be pursued in subsequent analyte deconvolution, identification and quantification steps; this may be accomplished by setting an F-ratio threshold for the hit list. However, determination of such a threshold is often a subjective matter that requires manual analyst intervention to decide the point at which the selected features become unreliable, and often leads to excessive, time-consuming manual data analysis steps [24,25].

As an alternative to manual threshold determination, statistics-based approaches should be considered and applied. One approach is to utilize the critical value for the F -test (F_{crit}) to determine a threshold for significance [26,27]. Another statistics-based approach is using a z-scores method [28]. An alternative approach for F-ratio threshold determination is to analyze the within-class variation for each sample class within the experiment, and utilize the maxima of these values as the cutoff, under the basis that any significant between-class variation will exceed that of the within-class variation [29]. We recently reported a new approach, null distribution analysis, which is a permutation-based test [30] to establish the distribution of spurious F-ratio values for a given GC $\times$ GC-TOFMS experimental design analyzed using tile-based F-ratio analysis [22], which used a frequentist strategy to infer a threshold to differentiate hits from non-hits. Herein, we describe improvements to null distribution analysis to determine a statistically-appropriate F-ratio threshold aimed at minimizing false discoveries while simultaneously

maximizing true discoveries. The improved null distribution analysis methodology is then applied to find class-distinguishing analytes, i.e., chemical marker compounds which discriminate acid altered from unaltered diesel fuel.

2. Theory

For robust determination of class-distinguishing analytes in non-targeted GC $\times$ GC–TOFMS F-ratio experiments, it is necessary to choose an appropriate threshold to cut off the analysis of the resulting hit list, in order to save analyst time and to avoid false discoveries. Most traditional statistical approaches would suggest simply using a critical value, in this case F_{crit} , which is determined for the degrees of freedom in the experiment and significance level desired, and which is provided in tables found in most quantitative analysis or statistics textbooks. However, the F_{crit} approach makes the assumption that the F -distribution arises from within-class and between-class variates that each follow a chi-squared distribution, and which are independent; the validity of the assumptions may vary. A more-pressing consideration is that the use of critical values is based on single hypothesis testing, and does not consider the overall error rate when testing multiple hypotheses. When testing multiple hypotheses, it is necessary to consider and control the false discovery rate (FDR) [31], the basis of which is described next.

In statistical hypothesis testing, the significance level, α , predicts the probability of a type I error (rejecting the null hypothesis when it is actually true, resulting in a false positive); in a sample-class comparison experiment, this would be “discovering” a feature that is not class-distinguishing and is instead caused by noise, baseline, or other spurious covariance. The significance level, α , provides a statistic for the type I error rate for testing a single hypothesis. However, in the application of the F-ratio method to a GC $\times$ GC–TOFMS sample-class

comparison experiment, every feature in the chromatogram represents an individual hypothesis, i.e., each individual feature is tested to determine whether it is statistically class-distinguishing. In the traditional approach, hypothesis testing proceeds as follows: a significance level is chosen for a test statistic to calculate a critical value, for example, an analyst may choose $\alpha = 0.001$, which corresponds to a 1:1000 chance of a type I error. Statistics are calculated for the feature being tested and the generated p -value for the test is then compared against α to decide whether to reject the null hypothesis at that chosen significance level. Generally, if the p -value is less than α , the analyst considers the feature to be statistically different between classes. When this procedure is applied with $\alpha = 0.001$, an individual hypothesis may be expected to have a 0.1% chance of falsely failing, that is, producing a false positive. However, when testing multiple hypotheses, such as testing each peak in a chromatogram, the overall chance of at least one hypothesis test producing a false positive is actually greater than 1:1000, and is proportional to the number of hypotheses tested. This phenomenon is known as the multiplicity problem, which requires adjustment of the statistics according to the number of hypotheses tested [31,32]. Indeed, methods to control the FDR are an active area of contemporary statistics, particularly in the field of genetic microarray analysis and other biological analyses [33–35]. Herein, we now introduce null distribution analysis as a method to control the FDR, while avoiding the assumptions of F_{crit} methods, concurrent with inherently addressing the multiplicity problem.

Null distribution analysis is a technique that leverages the large number of simultaneous hypotheses that may be tested in a GC $\times$ GC–TOFMS dataset to experimentally determine the distribution of potential false positives [22]. By rearranging the sample classes to test the null hypothesis (that there are no differences between the classes), the effects of non-meaningful variation in the dataset may be estimated for subsequent sample-class comparisons. The null

distribution method is utilized for determining an F-ratio threshold which holds no assumptions about the underlying sources of variation that lead to non-significant F-ratios. Further, the method is applied to the results of the testing of multiple hypotheses, such that the results are inherently applicable to the multiple hypotheses testing in the sample-class comparison, and directly address the probabilities of false positives in context of the FDR. In our initial report for a spiked diesel study [22], the “null” sample classes were constructed by manually arranging six simple pairwise null combinations for a four versus four comparison (two sample classes, each comprised of four GC $\times$ GC–TOFMS chromatograms). Six pairwise arrangements were chosen to demonstrate proof-of-principle of the methodology, as the null arrangements for this case were simple to explain, and provided adequate sampling of all the possible null comparisons.

However, to increase the extensibility of null distribution analysis, it is desirable to algorithmically determine all possible null comparisons in a rigorous and automated fashion.

Null distribution determination can be rigorously expanded, from a manual subset of arrangements, to include all possible null class arrangements through the use of combinatorial theory. In order to calculate the total number of unique null arrangements that can be generated from a given sample-class comparison experiment, the number of classes in the null arrangements must be equal to the number of classes in the original data, and there must be an equal number of samples from each class in the null class; this means that the number of samples in each class of data must be a multiple of the number of classes. For example, in a 2×2 sample-class comparison (two samples in Class 1 and two samples in Class 2), a null arrangement is generated by creating two new classes: A and B, where Class A consists of one sample from Class 1 and one sample from Class 2, and Class B would then consist of the remaining samples from Classes 1 and 2. Because the order of the samples within a class does not affect the F-ratio

calculation, we can determine the number of null arrangements by “counting” the number of unique combinations of a subset of k samples from a class of size n , a combinatorial method of sampling where order does not matter, often referred to as “ n choose k ”[36].

$${}_nC_k = \frac{n!}{k!(n-k)!} \quad (1)$$

The value ${}_nC_k$ in Eq. (1) is used to calculate the number of ways to select k samples from the set of n samples where n is the size of Class 1 and k is the number of samples from Class 1 to be included in each null class. For a two class system, $k = n/2$. This calculation must be performed for each class, and the total number of combinations is the product of these two values. However, there is no mathematical difference between Class A and Class B (i.e., Class A consisting of samples 1 and 2, Class B consisting of samples 3 and 4 is the same as Class A consisting of samples 3 and 4, Class B consisting of samples 1 and 2). Therefore, the number of unique null arrangements will be $\frac{1}{2}$ the total number of combinations from Classes 1 and 2, utilizing Eq. (1),

$$\# \text{ Unique Null Arrangements} = \frac{1}{2} * {}_{n_1}C_{k_1} * {}_{n_2}C_{k_2} = \frac{1}{2} * \frac{n_1!}{k_1!(n_1-k_1)!} * \frac{n_2!}{k_2!(n_2-k_2)!} \quad (2)$$

With the above combinatorial method for creating null arrangements, we are now able to study the potential benefits of using more null distributions than previously reported, in which we explored the use of a small number of randomly selected null arrangements (i.e., 6 arrangements) [22]. Using Eq. (2), the number of possible unique null arrangements may be calculated; for common sizes of sample-class comparison experiments, 4×4 , 6×6 , and 8×8 , there are 18, 200, and 2450 unique null arrangements, respectively. In the previous report, 6 null arrangements were selected from the 4×4 comparison that was performed. From a practical perspective a 6×6 sample-class size comparison, producing 200 unique null arrangements, is a sensible compromise between possibly having insufficient statistical coverage of the null distribution behavior (i.e., with only 18 unique null arrangements using a 4×4 comparison), and

producing an overabundance of null arrangements which may begin to impact computational performance (i.e., using the 8×8 comparison that produces 2450 null arrangements). Thus, the 6×6 sample-class size comparison is utilized in the current study.

3. Experimental

3.1. Sample Preparation

Six diesel fuel samples were collected from separate fueling stations in the Seattle area, from various fuel brands, labeled herein as samples AR, FH, MB, SG, SS, and RC. Only samples SS and RC contained significant levels of biodiesel in the form of fatty acid methyl esters (FAMES). The fuel samples were collected and stored in amber glass bottles. A spike solution was prepared to deliver two internal standards, bromoheptane (Aldrich, 99%) and bromobenzene (Aldrich, 99%), to each fuel at final nominal concentrations of 800 ppm for each of the internal standards. After thorough mixing, two 25.0 g aliquots of each of the six spiked fuels were added to individual glass sample jars (4 fl. oz. capacity with polypropylene-lined lids) along with PTFE-coated magnetic stir bars. For each fuel, 5.0 g of concentrated sulfuric acid (Macron Fine Chemicals, ACS reagent grade) was added to one jar, while the other was a control (no acid added); control samples were handled similarly to avoid false discoveries due to sample handling. The samples were stirred using a magnetic stir plate at 1000 rpm for 10 min, which was sufficient time to produce the desired acid alteration chemical changes for the diesel samples in which acid was added. Upon completion of stirring, 1.5 ml of each sample was transferred to polypropylene 1.5 ml microcentrifuge tubes and centrifuged at 10,000 rpm ($5,000 \times g$) for 5 min to separate the diesel and acid phases. The diesel phase of each centrifuged sample was then transferred to 1.5 ml autosampler vials for analysis by GC $\times$ GC–TOFMS.

3.2. Instrumental

The GC $\times$ GC–TOFMS instrument consisted of an Agilent 6890N gas chromatograph equipped with an Agilent 7683 autoinjector (Agilent Technologies, Palo Alto, CA) coupled with a LECO Pegasus III TOFMS equipped with a 4D thermal modulator upgrade (LECO, St. Joseph, MI). The primary column of the GC $\times$ GC (column 1) was a 20 m $\times$ 250 μ m i.d. $\times$ 0.5 μ m Rtx-5MS film (Restek, Bellefonte, PA) and the secondary column (column 2) was a 2 m $\times$ 180 μ m i.d. $\times$ 0.2 μ m Rtx-200 film (Restek, Bellefonte, PA). The GC instrument inlet was set to 275 $^{\circ}$ C and the transfer line was set to 290 $^{\circ}$ C. Column 1 was held at 40 $^{\circ}$ C for 1 min and then increased at 5 $^{\circ}$ C/min to 285 $^{\circ}$ C, where it was held for 5 min, for a total run time of 55 min. The column 2 oven and modulator block followed the same temperature program as column 1, with +5 $^{\circ}$ C and +20 $^{\circ}$ C offsets, respectively. The modulation period was 1 s and the hot and cold pulse times for each stage were 0.4 s and 0.1 s, respectively. A constant (ambient temperature and pressure corrected) flow rate of 2 ml/min at the outlet of column 2, with helium (Praxair, Grade 5.0) as the carrier gas. The ion source was set to 250 $^{\circ}$ C, the electron impact energy was 70 eV, and the detector voltage was set to 1600 V. Mass channels, m/z 33–250, were collected at 100 spectra/s after a 10 s acquisition delay. A 1 μ l injection of each diesel sample was made in split mode with a split ratio of 200:1. Each diesel sample was injected in duplicate.

3.3. Computational

To discover class-distinguishing chemical features for the acid alteration, a sample-class comparison experiment was constructed using three different diesel fuels collected from separate filling stations in the Seattle, WA area: **samples AR, FH, and MB**; the three fuels were analyzed prior to and following acid alteration. Two chromatograms from of each of the three fuels prior to acid alteration comprised the unaltered class, and two chromatograms of each of the three

fuels after acid alteration comprised the altered class. Findings from the sample-class comparison experiment were validated using the remaining three samples: SG, SS, and RC. All chromatograms were normalized for injection variation using the sum of signals method (often referred to as TIC normalization) [9], which was validated against the bromoheptane and bromobenzene internal standards. The normalization values were used in the tile-based F-ratio analysis, as well as in plotting and quantification of selected class-distinguishing features.

GC $\times$ GC-TOFMS data from the 12 runs for the sample-class comparison (3 fuels $\times$ 2 replicates $\times$ 2 classes) were imported from the LECO ChromaTOF software v 3.32 (LECO, St. Joseph, MI) to Matlab R2014b (8.4.0.150421) (MathWorks, Inc., Natick, MA) via an in-house developed data converter [37]. The imported data were analyzed with in-house developed tile-based F-ratio software, performed on a mid-level personal computer, having an Intel Core i7-4770 processor (3.4 GHz), a 250 GB Samsung 840 solid state hard drive, and 16 GB dual-channel DDR3 RAM, with Windows 7 SP1 as the operating system. The tile-based F-ratio software was used to discover chemical features in the fuels that differentiated the unaltered and altered classes [21,22]. The tile and cluster sizes, given as column 1 $\times$ column 2 (number of modulations $\times$ number of spectra) were 6 $\times$ 10 and 4 $\times$ 10, respectively. Selection of the tile size is a balance of two competing objectives. First, the tile aims to capture most of the 2D peak signal (using the $\pm 2\sigma$ width) plus additional space to allow for minor retention time shifts in each dimension. Second, the tile size is kept sufficiently small to maintain selectivity for individual peaks (i.e., to minimize sampling neighboring peaks) and to minimize inclusion of noise [22]. The per- m/z S/N threshold was 3σ with ≥ 3 m/z required above the threshold for inclusion of the tile; all m/z collected were included in the analysis. The null probability limit was 0.1%.

Selected analyte features of interest were quantified using ChromaTOF. Briefly, all column 2 peaks (i.e., peaklets) that result from the modulation of the column 1 peak for selected analytes are individually detected and deconvoluted; unique m/z are then found for the chromatographic profiles present in the peaklets. Next, peaklets are combined to form a single 2D peak for each selected analyte, based on the mass spectra and column 2 retention times of the peaklets, as well as the expected peak widths in both dimensions. Finally, the combined 2D peak for each analyte was integrated using the unique m/z to calculate the quantified peak sum. The deconvoluted mass spectra, along with analyte retention times, were compared to the NIST11 mass spectral library and used to preliminarily identify discovered features.

Null distribution analysis was performed similarly to that described previously [22], with the distinction that in this report, we perform a 6×6 sample-class comparison, including the complete set of all 200 unique null arrangements. Furthermore, the 200 null distributions were combined by averaging and used to prepare a comprehensive null probability curve, which was used to determine the F-ratio threshold. The individual null distributions were then used to infer the statistical confidence for the F-ratio threshold. Figure 1 is a flowchart summarizing the workflow performed in this study.

4. Results and Discussion

4.1. Tile-based F-ratio analysis

In this study, the analytical challenge was to discover changing analytes in a highly-complex sample with maximal discovery of significantly-changing analytes, even at very low relative concentrations, concurrent with minimal discovery of analytes that do not significantly change between sample classes. Further, we aimed to discover said changes without explicit

prior knowledge of which chemicals, or even chemical classes, would be substantially modified by the alteration. GC $\times$ GC–TOFMS allows for the comprehensive analysis of diesel fuel with a far greater degree of chromatographic resolution than one-dimensional GC and additionally improves detection limits through the zone-focusing effects of the thermal modulator [16–19]. These benefits are of substantial value for the study of the acid alteration of diesel fuel, in which the affected compounds are at low abundances compared to the predominantly aliphatic content of diesel fuel (which is not affected by the alteration) and which may only be partially removed by the alteration. Figure 2 demonstrates the challenge: under visual comparison of the 2D total ion current (TIC) chromatograms, the visible differences between the fuels is more apparent than those prior to and following the acid alteration.

In cases where the differences between sample classes are subtle, algorithmic methods are critical for rapid and comprehensive discovery of class-distinguishing analytes. The tile-based F-ratio software enables analysis of GC $\times$ GC–TOFMS data with minimal user intervention, and is able to discover analytes even at low absolute concentration and a low concentration ratio. Further, the tile-based algorithm mitigates recognized challenges in the analysis of GC $\times$ GC–TOFMS data, including the impacts of minor chromatographic retention time shifts and ensuring all of the column 2 peaklets for a given modulated analyte are accurately combined for 2D peak quantification [38–40]. By improving the signal-to-noise ratio (S/N) and reducing the impact of spurious chromatographic and detector variations prior to the F-ratio calculation, the tile-based algorithm maximizes the discovery of significant features while minimizing features of lesser value [21,22].

The analysis initially produces a hit list of features ranked in order of decreasing F-ratio values. The F-ratio is calculated for each feature in the sample-class comparison chromatograms

that has sufficient signal, here, having three or more m/z above the S/N threshold of 3σ . Based on the number of pixels (data points) in the 2D chromatogram, there are 328,900 possible hits (3289 column 1 modulations $\times$ 100 column 2 mass spectra) for the sample-class comparison experiment. Tiling the 2D chromatogram reduces the data to 21,880 possible hits; applying the per-tile requirements for S/N further reduces this to 16,813 hits. Finally, redundant hit removal reduced the hit list to 3362 features prior to applying the F-ratio threshold determined by the combinatorial null distribution analysis.

In a sample-class comparison experiment such as the acid alteration of diesel fuel, only a relatively small fraction of the total number of chromatographic features are expected to be class-distinguishing. As a result, the preliminary hit list is dominated by small F-ratio values (features that are likely not class-distinguishing) with a lesser number of large F-ratio values (features that are likely class-distinguishing). Figure 3 is a histogram of all 3362 F-ratio values in the preliminary hit list for the sample-class comparison; the distribution is approximately centered on 1, with the vast majority of values below 10. The inset portion of Figure 3 is the tail of the distribution from 10 to 100; the tail of the described distribution contains the chromatographic features that are the most class distinguishing for the acid alteration. However, Figure 3 shows that the distribution of F-ratio values is continuous, with no clear distinction between the lesser-value features in the main part of the distribution, and the greater-value features in the tail of the distribution. To distinguish class-distinguishing features from among all of the chromatographic features in the sample-class comparison, it is necessary to statistically define an F-ratio threshold using null distribution analysis.

4.2. Combinatorial null distribution and threshold determination

Combinatorial null distribution analysis comprehensively evaluates the possible null distributions that may occur in a given experiment. Performing multiple null distribution combinations allows us to predict the variability in the F-ratio distributions of non-hits for the sample-class comparison, allowing the inference of an F-ratio threshold that is robust to reasonable experimental variations in the sample-class comparison. Ultimately, our goal is to ensure that the chosen threshold applied to the preliminary hit list for the sample-class comparison provides the correct balance of avoiding false positives while also finding the lower-ranked features that may still be informative for the experimental study.

As described in the theory section, a 6×6 sample-class comparison allows for the computation of 200 unique null distributions. As noted, increasing the number of samples within each sample class rapidly increases the number of possible unique null combinations, e.g., 2450 for an 8×8 sample-class comparison. While increasing the number of samples within the sample classes may provide a more-comprehensive sampling of the possible ranges of diesel fuel composition, the calculation of the comprehensive null distributions quickly becomes computationally burdensome. Statistical subsampling methods such as Monte-Carlo sampling may be an avenue to maintain a reasonable computational burden while realizing potential advantages to expanded sample sizes, and may be a fruitful area for future work.

For this study, sample injection replicates were treated as unique samples for the combinatorial null classes, as retention time and injection variation between injections of the same fuel may be similar in magnitude to differences between similar fuels, as was performed in previous studies [21,22]. Further, sample replicates reflect the nature of fuel station sampling: due to centralized refining and distribution, it is possible for different stations to dispense

essentially the same fuel. In discovery-based analysis, there is also the risk of introducing too much sample variation, for example, two fuels (SS and RC) out of the six that were collected contained substantial amounts of FAME-based biodiesel and had lower contents of aromatics. Including these fuel samples in the sample-class comparison experiment would substantially increase the within-class variation and reduce the ability to find class-distinguishing features due to the acid alteration. Utilizing sample replicates strikes a balance between sampling the reasonable variation between fuels that may be encountered in the field, and introducing excessive within-class variation that may preclude the success of the F-ratio analysis.

Figure 4 illustrates a single null arrangement where the null classes are comprised of the switching of the replicates between the sample classes, such that Null Class A contains the first chromatogram (of the two replicates) of each fuel prior to (Class 1) and following acid alteration (Class 2), and Null Class B contains the remaining chromatograms. The null rearrangement eliminates the true source of variation in the sample-class comparison experiment (i.e., the reaction with sulfuric acid), such that the only sources of variation are those which are not germane to the class comparison, such as chromatographic and detector noise and variation. When the tile-based F-ratio software is applied to a null comparison, the resulting null hit list represents the distribution of F-ratio values that one would expect to encounter had there been no meaningful class differences in the sample-class comparison experiment.

Figure 5A includes the null distributions from the 200 unique null combinations for the 6×6 sample-class comparison, per Eq. (2). Overall, nearly all of the null F-ratios are below 10, and most of the null distributions are centered on F-ratio values from about 1 to 2. The null distributions are transformed into null probability curves, which represent the proportion of the null features that would be included in the analysis at given F-ratio cutoffs. The null probabilities

are calculated by dividing the number of null features above a given F-ratio by the total number of features in the null distribution, multiplied by 100 to yield a percentage. Figure 5B shows the resulting null probability curves; the dashed red line represents the 0.1% null probability limit, which is defined as the F-ratio that is exceeded by only 1 in 1000 values in the respective null distribution [22]. Most of the null distributions suggest a null probability limit of an F-ratio of approximately 6 to 10; however, there are curves that lie further to the right, e.g., the right-most null probability curve which has a 0.1% null probability limit with an F-ratio of 44. The suggested 0.1% null probability limits are more easily visualized in histogram form. Figure 5C shows the histogram (black trace) of the 0.1% null probability limits from all 200 null distributions overlaid with the average null distribution (blue) and the distribution of hits from the sample-class comparison from Fig. 3 (red). Figure 5D is a zoomed view of the overlapping distributions from F-ratios in the range of 10 to 100.

The distribution of 0.1% null probability limits provides insight to guide the analyst in selecting a suitable F-ratio threshold for the purpose of focusing on the true positive hits while minimizing analytical effort on hits that end up being false positives. For example, if the F-ratio threshold is set to 13.6, 90% of the 0.1% null probability limits are included. That is, in the application of a threshold of 13.6, we would expect only a 10% chance of the FDR exceeding 1 in 1000 features (0.1%). At a threshold of 27.6, 99% of the 0.1% null probability limits are included, and we would expect only a 1% chance of the FDR exceeding 0.1%. Here, we chose the threshold by averaging all 200 null distributions and then determining the 0.1% null probability limit; this method suggests a threshold of 12.4 (indicated by the blue arrow in Figure 5C). In applying the threshold of 12.4 to the preliminary sample-class comparison hit list (Fig. 3), we cover 87% of the 0.1% null probability limits, such that there is only a 13% chance that

the FDR will exceed 0.1%; thus we expect to encounter approximately 3 false positives above the threshold. At the threshold of 12.4, 107 out of the initial 3362 potential features were found to be class-distinguishing, the significance of which is discussed in the following section.

4.3. Acid alteration

Application of F-ratio analysis with combinatorial threshold determination found a total of 107 features that distinguish the altered fuels from the unaltered fuels. For purposes of concisely describing the chemical changes that occur from acid alteration of the diesel fuel, we will focus primarily on the top twenty hits found, which are summarized in an abbreviated hit list in Table 1. The hit list provides the F-ratio and 2D retention times for each discovered feature; additionally, the software determines whether the feature increases or decreases following the acid alteration. Features that were removed by the acid alteration are indicated by a minus sign (–), while features that are generated by the acid alteration are indicated by a plus sign (+). The null probability column indicates the overlap of each hit with the combined null distribution. All of the top twenty hits had an F-ratio exceeding that of the null distribution, and thus had a null probability of 0.00%. Indeed, the top 55 out of 107 hits above the threshold had a null probability of 0.00%. Using the outputs of the software, analytes were tentatively identified by matching to the NIST11 mass spectral library.

The primary types of class-distinguishing features in the hit list were aromatic species, alkenes/alkynes, and compounds with both aromatic and alkene/alkyne moieties. The majority of discovered features were removed partially or entirely by the acid alteration. Only a small number of features were found to be generated by the acid alteration. The only feature which was generated by the acid alteration with adequate *S/N* and chromatographic resolution to be identified was sulfur dioxide, which was not present prior to acid alteration, but was consistently

found following the acid alteration. To demonstrate the variety of chemical functional groups that were changing in response to the acid alteration, Figure 6A plots the top twenty hits as an overlay on a representative TIC chromatogram from the unaltered class; the blue circles are centered on the retention time for the class-distinguishing feature, and are sized according to their relative F-ratios (i.e., larger hits have bigger circles). The blue numbers next to the circles correspond to the hit numbers in Table 1. The presence of multiple hits for sulfur dioxide (Hits 2 and 14) was due to breakthrough during thermal modulation, resulting in double modulation and non-Gaussian column 2 chromatographic profiles for the volatile gas. Though we focus on the top twenty hits found for brevity, there were a total of 107 hits found above the F-ratio threshold of 12.4. Figure 6B shows all 107 hits, demonstrating that further class-distinguishing features show a similar pattern to the first twenty.

Overall, in agreement with literature on electrophilic aromatic sulfonation, the aromatic fractions of the diesel fuels were chemically altered by reaction with sulfuric acid resulting in substantial reduction of these compounds in the acid altered fuels. Compounds such as pyrene and alkyl-substituted pyrenes were removed by the acid alteration, most likely via electrophilic aromatic sulfonation, and subsequent partitioning to the aqueous acid phase of the reaction mixture [2]. Overall, the relative changes in concentration agree with literature on reactivity of various aromatic species to sulfonation [8]. Literature on aromatic nitration, which also follows an electrophilic aromatic substitution mechanism, further support the relative reactivities of the aromatics [41]. Figure 7 summarizes the relative peak sums for a series of aromatic species that were present in the diesel fuel samples, including benzene, naphthalene, phenanthrene, and pyrene. For the discovery of robust markers for the acid alteration of diesel fuel, it is preferable to utilize diverse fuels to comprise the sample classes; here we used fuels from three different

sources to sample the variability of fuels. As shown in Figure 7, the amounts of aromatic species present in the diesel fuels prior to acid alteration were variable between samples; this is consistent with literature findings that the aromatic fraction of diesel fuel is the most important for chemometric discrimination of samples [42]. In comparing each fuel to itself prior to and following acid alteration, it is apparent that the degree of removal by the acid alteration for the aromatics is consistent with the expected reactivity. Benzene is less reactive and is removed to a small degree by acid alteration, while pyrene is more reactive and is removed to a much larger extent. Notably, though benzene exhibits a consistent pattern of slightly decreasing signal abundance after acid alteration, the amounts of benzene natively present in the unaltered fuels is relatively variable compared to the amount of benzene removed by the acid alteration. Though the alteration induces meaningful class-to-class differences in the relative amounts of benzene, the within-class variation is substantially larger, leading to a small F-ratio. As a result, benzene is not found as a hit in the analysis. In contrast, pyrene is extensively removed by the acid alteration and has substantial between-class variation, exceeding that of the within-class variation. As a result, pyrene has a substantial F-ratio of 53.7 and is found as Hit 22 in the final hit list. Naphthalene is intermediate to benzene and pyrene; it is removed to a smaller degree than pyrene, but is relatively consistent prior to alteration, and is found as Hit 82 with an F-ratio of 15.5. We additionally examined other analytes present in the fuel that were expected to not change, such as decane; these analytes were quantified and confirmed to be unaffected by the alteration, and are not shown for brevity.

In addition to finding that aromatics with greater degrees of aromaticity were removed to a greater extent, the hit list revealed other aromatics with alkene and alkyne substituents were also removed by reaction with sulfuric acid. The top hit (Hit 1) was tentatively identified as 1-

propynylbenzene; as shown in Figure 8, 1-propynylbenzene has relatively low within-class variation in the unaltered class, and was likely entirely removed by the acid alteration, as there was no detectable peak in the acid altered class. The peak's relatively low within-class variance coupled with its high between-class variance leads to a significant F-ratio of 410, placing it at the top of the hit list. Figure 8 includes similar plots of the next three top analytes, two of which are similar to Hit 1 in that they were detected only in the unaltered class, and in that they had both aromatic and alkene/alkyne moieties. As noted in the abbreviated hit list in Table 1, we also observed the unexpected removal of non-aromatic species from the fuel, particularly, alkenes; a subsequent review of literature suggests that the mechanism may be alkene sulfonation or sulfation [5]. A further interesting and unexpected observation was the presence of sulfur dioxide (SO₂) in the acid altered fuels. Sulfur dioxide was not found in any of the unaltered-fuels, but was found at relatively high abundances in all of the acid altered fuels, as shown in Figure 8. It is unknown whether the sulfur dioxide was present as a pure compound in the acid altered fuels, as a product of residual sulfuric acid in the altered fuels, or whether it was produced by breakdown of other species during analysis in the hot (275 °C) inlet. Potential sources of sulfur dioxide include sulfones that may be byproducts of aromatic sulfonation [5], which potentially could remain in the diesel phase. However, more study is needed to confirm this suggestion.

Six fuels were collected from separate fuel stations, with three fuels included in the F-ratio sample-class comparison, while the remaining three fuels were utilized as validation samples in this initial study. To demonstrate that the features discovered by the class-comparison experiment were useful for other acid altered fuels, the signals for the top four features from the analysis, as listed in Table 1, were quantified for all six fuels. Table 2 provides the mean integrated signals, with *N.D.* indicating that the analyte was not measured above the limit of

detection. Overall, the class-distinguishing features discovered in the sample-class comparison using samples AR, FH, and MB could also be used to distinguish the acid altered validation samples SG, SS, and RC. The validation samples consistently exhibited the same pattern of hits 1, 3, and 4 being removed by the acid alteration, and hit 2 being created by the acid alteration. Compared to the samples used in the sample-class comparison, the validation samples had smaller signal intensities in the unaltered fuels for 1-propynylbenzene, and larger intensities for sulfur dioxide. For the aromatic species comprising hits 1, 3, and 4, sample SS had substantially lower signal intensities in the unaltered fuels, but was still distinguished from the altered class by the removal of these analytes by the acid alteration. While the results of this study are promising, further validation in future studies is warranted, with a larger set of samples.

The tile-based F-ratio software performed well in finding significant differences between the unaltered and altered classes, even with the variation between fuels within each class. Figure 9 demonstrates a key analytical challenge addressed by the software: many of the class-distinguishing features that were discovered were only present at very limited absolute signal abundance. Figure 9A-B are the TIC chromatograms for the 1D chromatographic region corresponding to Hit 1 (1-propynylbenzene) and Hit 2 (sulfur dioxide), respectively. The blue traces correspond to the signals from the unaltered fuels, while the red traces correspond to the altered fuels. The black arrows indicate the 2D retention times of the discovered features. Notably, there is no apparent peak at the retention time of Hit 1 in the TIC chromatogram plotted in 9A. However, the *S/N* improvement via the tiling process and utilization of selective *m/z* made it possible to discover the feature, even with its relatively low signal abundance. Figures 9C-D are the extracted ion current (XIC) chromatograms for Hit 1 (*m/z* 116) and Hit 2 (*m/z* 64), respectively. Hit 1 is removed by the acid alteration, while Hit 2 is created by the acid alteration.

Figure 9C shows that 1-propynylbenzene is not detected above the baseline in the acid altered samples. Conversely, Figure 9D shows that sulfur dioxide is not detected above the baseline in the unaltered samples. Figure 9D also illustrates that even non-Gaussian peak profiles can be successfully examined; as mentioned above, sulfur dioxide, a gas under standard conditions, is incompletely trapped during thermal modulation, resulting in non-standard peak shapes. These figures demonstrate the effectiveness with which the software finds class-distinguishing features, even for features with low *S/N* or non-Gaussian chromatographic peak profiles.

Conclusions

Overall, the tile-based F-ratio analysis of the acid altered diesel fuels revealed that the chemical changes induced by reaction with sulfuric acid are subtle yet of potential forensic significance, especially under the non-stoichiometric and short reaction time conditions utilized in this report. Further, the F-ratio analysis utilized three different fuels, each individually acid altered, adding to the challenge since differences between fuels may exceed the differences between classes. However, analytes were discovered that were both sufficiently consistent in the unaltered class, and significantly changed in relative concentration following the acid alteration. The hypothesis that the relative reactivity of aromatic compounds toward electrophilic aromatic sulfonation was highly supported, as suggested in other literature. Furthermore, removal of alkenes and alkynes in the acid altered fuels was observed, suggesting additional mechanisms of reaction with sulfuric acid which may alter the chemical composition of the fuel. Finally, consistent generation of sulfur dioxide in acid altered fuels was observed. Further investigations are warranted to determine the origin, identities and analytical concentrations of the key species

altered by reaction with sulfuric acid, all of which may potentially serve as chemical marker compounds for the forensic determination of diesel alteration.

In addition to the application of previously reported tile-based F-ratio software and methods to the acid alteration of diesel fuel, this report introduced an improved method for null distribution analysis for the determination of defensible thresholds specific to individual sample-class comparison experiments. Through evaluating the comprehensive set of 200 null distributions for a 6×6 sample-class comparison, we were able to observe the variability of the null distributions, and thus infer the reliability of a given null probability limit for use as a threshold to avoid (or certainly to minimize) false positives in the hit list. In this study, the combined null distribution suggested a 0.1% null probability threshold of 12.4; at this threshold value, 87% of the null probability limits are included, suggesting that there is only a 13% chance of the FDR exceeding 1 in 1000 for the sample-class comparison. Applying the F-ratio threshold of 12.4 to the acid alteration experiment, we found 107 features that should be class-distinguishing. Since the true positives in the experiment are not known *a priori*, it is not possible to explicitly determine the FDR, though we would expect it to be less than 1:1000 (about 3 false positives in this experiment), based on the combined 0.1% null probability limit of 12.4 having 87% coverage in the comprehensive null combinations. In the top 20 hits studied in depth, no false positives were encountered. The major advantage of the combinatorial method presented here for F-ratio threshold determination is that it allows for a hit list cutoff to be selected that is statistically suitable for the sample-class comparison being studied. The cutoff is determined by the variability of the data, not by an *a priori* decision of how many hits should be chosen. If the analyst were to arbitrarily choose a number of hits to examine, many class-distinguishing features may be missed or an excessive number of non-hit features would be

597 analyzed [22]. However, future work is underway to further characterize the robustness of the F-
598 ratio cutoff determination in the context of the FDR in a controlled spike experiment in which
599 true positives will be known *a priori* and with which the FDR predicted by null distribution
600 analysis may be experimentally validated.

601 **Acknowledgements**

602 This work was supported by the Internal Revenue Service (IRS) under an Interagency
603 Agreement with the US Department of Energy (DOE) under Contract DE-AC05-76RLO 1830
604 with the Pacific Northwest National Laboratory. We thank Dr. Luke Marney and Dr. Jamin
605 Hoggard for their contributions to the initial development of the tile-based Fisher ratio software.

606 References

- 607 [1] B. Krakowska, I. Stanimirova, J. Orzel, M. Daszykowski, I. Grabowski, G. Zaleszczyk, et
608 al., Detection of discoloration in diesel fuel based on gas chromatographic fingerprints, *Anal.*
609 *Bioanal. Chem.* 407 (2014) 1159–1170. doi:10.1007/s00216-014-8332-4.
- 610 [2] C. Martín-Alberca, C. García-Ruiz, O. Delémont, Study of chemical modifications in
611 acidified ignitable liquids analysed by GC–MS, *Sci. Justice.* 55 (2015) 446–455.
612 doi:10.1016/j.scijus.2015.06.006.
- 613 [3] C. Martín-Alberca, C. García-Ruiz, O. Delémont, Study of acidified ignitable liquid residues
614 in fire debris by solid-phase microextraction with gas chromatography and mass
615 spectrometry, *J. Sep. Sci.* 38 (2015) 3218–3227. doi:10.1002/jssc.201500337.
- 616 [4] C. Martín-Alberca, M. López-López, C. García-Ruiz, Analysis of pre-ignited improvised
617 incendiary devices using portable Raman, *Talanta.* 144 (2015) 612–618.
618 doi:10.1016/j.talanta.2015.06.072.
- 619 [5] G.P. Dado, E.A. Knaggs, M.J. Nepras, Sulfonation and Sulfation, in: *Kirk-Othmer Encycl.*
620 *Chem. Technol.*, John Wiley & Sons, Inc., 2000.
621 [http://onlinelibrary.wiley.com/doi/10.1002/0471238961.1921120611140107.a01.pub2/abstra](http://onlinelibrary.wiley.com/doi/10.1002/0471238961.1921120611140107.a01.pub2/abstract)
622 [ct.](http://onlinelibrary.wiley.com/doi/10.1002/0471238961.1921120611140107.a01.pub2/abstract)
- 623 [6] H. Cerfontain, H.J.A. Lambrechts, Z.R.H. Schaasberg-Nienhuis, R.G. Coombes, P.
624 Hadjigeorgiou, G.P. Tucker, Aromatic sulphonation. Part 91. The sulphonation of anisole,
625 phenol, phenyl methanesulphonate, potassium phenyl sulphate, and a series of methyl-,
626 bromo-, and chloro-substituted anisoles and phenols in concentrated aqueous sulphuric acid,
627 *J. Chem. Soc. Perkin Trans. 2.* (1985) 659–667. doi:10.1039/P29850000659.
- 628 [7] R.A. Cox, Organic reactions in sulfuric acid. The excess acidity method, *Acc. Chem. Res.* 20
629 (1987) 27–31. doi:10.1021/ar00133a004.
- 630 [8] A.R. Katritzky, M.S. Kim, D. Fedoseyenko, K. Widyan, M. Siskin, M. Francisco, The
631 sulfonation of aromatic and heteroaromatic polycyclic compounds, *Tetrahedron.* 65 (2009)
632 1111–1114. doi:10.1016/j.tet.2008.11.023.
- 633 [9] K.M. Pierce, B. Kehimkar, L.C. Marney, J.C. Hoggard, R.E. Synovec, Review of
634 chemometric analysis techniques for comprehensive two dimensional separations data, *J.*
635 *Chromatogr. A.* 1255 (2012) 3–11. doi:10.1016/j.chroma.2012.05.050.
- 636 [10] K.M. Pierce, J.S. Nadeau, R.E. Synovec, Chapter 17 - Data Analysis Methods, in: C.F.
637 Poole (Ed.), *Gas Chromatogr.*, Elsevier, Amsterdam, 2012: pp. 415–434.
638 <http://www.sciencedirect.com/science/article/pii/B9780123855404000171>.
- 639 [11] S. Yang, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, Gas Chromatography and
640 Comprehensive Two-Dimensional Gas Chromatography Hyphenated with Mass
641 Spectrometry for Targeted and Nontargeted Metabolomics, in: M. Lämmerhofer, W.
642 Weckwerth (Eds.), *Metabolomics Pract.*, Wiley-VCH Verlag GmbH & Co. KGaA, 2013: pp.
643 69–92. <http://onlinelibrary.wiley.com/doi/10.1002/9783527655861.ch4/summary>.
- 644 [12] K.M. Pierce, S.P. Schale, Predicting percent composition of blends of biodiesel and
645 conventional diesel using gas chromatography–mass spectrometry, comprehensive two-
646 dimensional gas chromatography–mass spectrometry, and partial least squares analysis,
647 *Talanta.* 83 (2011) 1254–1259. doi:10.1016/j.talanta.2010.07.084.

- [13] R. van der Westhuizen, H. Potgieter, N. Prinsloo, A. de Villiers, P. Sandra, Fractionation by liquid chromatography combined with comprehensive two-dimensional gas chromatography–mass spectrometry for analysis of cyclics in oligomerisation products of Fischer–Tropsch derived light alkenes, *J. Chromatogr. A.* 1218 (2011) 3173–3179. doi:10.1016/j.chroma.2010.10.009.
- [14] J.V. Seeley, S.K. Seeley, Multidimensional Gas Chromatography: Fundamental Advances and New Applications, *Anal. Chem.* 85 (2013) 557–578. doi:10.1021/ac303195u.
- [15] 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 (GC×GC-TOFMS) and Visual Basic Scripting, *Energy Fuels.* 28 (2014) 5670–5681. doi:10.1021/ef501247h.
- [16] A.L. Lee, K.D. Bartle, A.C. Lewis, A Model of Peak Amplitude Enhancement in Orthogonal Two-Dimensional Gas Chromatography, *Anal. Chem.* 73 (2001) 1330–1335. doi:10.1021/ac001120s.
- [17] P. Marriott, R. Shellie, Principles and applications of comprehensive two-dimensional gas chromatography, *TrAC Trends Anal. Chem.* 21 (2002) 573–583. doi:10.1016/S0165-9936(02)00814-2.
- [18] C. Vendevure, F. Bertoncini, L. Duval, J.-L. Duplan, D. Thiébaut, M.-C. Hennion, Comparison of conventional gas chromatography and comprehensive two-dimensional gas chromatography for the detailed analysis of petrochemical samples, *J. Chromatogr. A.* 1056 (2004) 155–162. doi:10.1016/j.chroma.2004.05.071.
- [19] J. Krupčík, P. Májek, R. Gorovenko, J. Blaško, R. Kubínek, P. Sandra, Considerations on the determination of the limit of detection and the limit of quantification in one-dimensional and comprehensive two-dimensional gas chromatography, *J. Chromatogr. A.* 1396 (2015) 117–130. doi:10.1016/j.chroma.2015.03.084.
- [20] A. Sampat, M. Lopatka, M. Sjerps, G. Vivo-Truyols, P. Schoenmakers, A. van Asten, The forensic potential of comprehensive two-dimensional gas chromatography, *TrAC Trends Anal. Chem.* (n.d.). doi:10.1016/j.trac.2015.10.011.
- [21] L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data, *Talanta.* 115 (2013) 887–895. doi:10.1016/j.talanta.2013.06.038.
- [22] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach, *Anal. Chem.* 87 (2015) 3812–3819. doi:10.1021/ac504472s.
- [23] K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: comprehensive GC×GC with ANOVA-based feature selection and principal component analysis, *Chemom. Intell. Lab. Syst.* 60 (2002) 225–237. doi:10.1016/S0169-7439(01)00198-8.
- [24] R.E. Mohler, K.M. Dombek, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, Comprehensive analysis of yeast metabolite GC×GC–TOFMS data: combining discovery-mode and deconvolution chemometric software, *Analyst.* 132 (2007) 756–767. doi:10.1039/B700061H.
- [25] J.E. Welke, V. Manfroi, M. Zanus, M. Lazzarotto, C. Alcaraz Zini, Differentiation of wines according to grape variety using multivariate analysis of comprehensive two-

- dimensional gas chromatography with time-of-flight mass spectrometric detection data, *Food Chem.* 141 (2013) 3897–3905. doi:10.1016/j.foodchem.2013.06.100.
- [26] M. Brokl, L. Bishop, C.G. Wright, C. Liu, K. McAdam, J.-F. Focant, Multivariate analysis of mainstream tobacco smoke particulate phase by headspace solid-phase micro extraction coupled with comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry, *J. Chromatogr. A.* 1370 (2014) 216–229. doi:10.1016/j.chroma.2014.10.057.
- [27] P.-H. Stefanuto, K.A. Perrault, R.M. Lloyd, B. Stuart, T. Rai, S.L. Forbes, et al., Exploring new dimensions in cadaveric decomposition odour analysis, *Anal. Methods.* 7 (2015) 2287–2294. doi:10.1039/C5AY00371G.
- [28] P.-H. Stefanuto, K.A. Perrault, S. Stadler, R. Pesesse, H.N. LeBlanc, S.L. Forbes, et al., GC × GC–TOFMS and supervised multivariate approaches to study human cadaveric decomposition olfactive signatures, *Anal. Bioanal. Chem.* 407 (2015) 4767–4778. doi:10.1007/s00216-015-8683-5.
- [29] H.P. Bailey, S.C. Rutan, Comparison of chemometric methods for the screening of comprehensive two-dimensional liquid chromatographic analysis of wine, *Anal. Chim. Acta.* 770 (2013) 18–28. doi:10.1016/j.aca.2013.01.062.
- [30] D.J. Vis, J.A. Westerhuis, A.K. Smilde, J. van der Greef, Statistical validation of megavariate effects in ASCA, *BMC Bioinformatics.* 8 (2007) 322. doi:10.1186/1471-2105-8-322.
- [31] Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *J. R. Stat. Soc. Ser. B Methodol.* 57 (1995) 289–300. doi:10.2307/2346101.
- [32] Y. Benjamini, D. Yekutieli, The Control of the False Discovery Rate in Multiple Testing under Dependency, *Ann. Stat.* 29 (2001) 1165–1188. doi:10.2307/2674075.
- [33] B. Efron, R. Tibshirani, Empirical bayes methods and false discovery rates for microarrays, *Genet. Epidemiol.* 23 (2002) 70–86. doi:10.1002/gepi.1124.
- [34] B. Efron, Bayesians, Frequentists, and Scientists, *J. Am. Stat. Assoc.* 100 (2005) 1–5. doi:10.1198/0162145050000000033.
- [35] S. Woo, J.T. Leek, J.D. Storey, A computationally efficient modular optimal discovery procedure, *Bioinformatics.* 27 (2011) 509–515. doi:10.1093/bioinformatics/btq701.
- [36] H.J. RYSER, *Combinatorial Mathematics*, 1st ed., Mathematical Association of America, 1963. <http://www.jstor.org/stable/10.4169/j.ctt5hh8v6> (accessed October 5, 2015).
- [37] J.C. Hoggard, *peg2mat3p8*; <http://depts.washington.edu/synlab/software/>, (2011).
- [38] G. Vivó-Truyols, H.-G. Janssen, Probability of failure of the watershed algorithm for peak detection in comprehensive two-dimensional chromatography, *J. Chromatogr. A.* 1217 (2010) 1375–1385. doi:10.1016/j.chroma.2009.12.063.
- [39] S. Samanipour, P. Dimitriou-Christidis, J. Gros, A. Grange, J. Samuel Arey, Analyte quantification with comprehensive two-dimensional gas chromatography: Assessment of methods for baseline correction, peak delineation, and matrix effect elimination for real samples, *J. Chromatogr. A.* 1375 (2015) 123–139. doi:10.1016/j.chroma.2014.11.049.
- [40] H.D. Bean, J.E. Hill, J.-M.D. Dimandja, Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography–mass spectrometry data, *J. Chromatogr. A.* 1394 (2015) 111–117. doi:10.1016/j.chroma.2015.03.001.

- 738 [41] W.A. Pryor, G.J. Gleicher, J.P. Cosgrove, D.F. Church, Reaction of polycyclic aromatic
739 hydrocarbons (PAH) with nitrogen dioxide in solution. Support for an electron-transfer
740 mechanism of aromatic nitration based on correlations using simple molecular orbital theory,
741 J. Org. Chem. 49 (1984) 5189–5194. doi:10.1021/jo00200a035.
- 742 [42] A.M. Hupp, L.J. Marshall, D.I. Campbell, R.W. Smith, V.L. McGuffin, Chemometric
743 analysis of diesel fuel for forensic and environmental applications, Anal. Chim. Acta. 606
744 (2008) 159–171. doi:10.1016/j.aca.2007.11.007.

Figure Captions

Figure 1. Flowchart of the experimental design with three fuels: AR, MB, FH.

Figure 2. Example total ion count (TIC) chromatograms for representative chromatograms from two unaltered fuels (MB_u and AR_u) and those same fuels after acid alteration (MB_a and AR_a). Visible differences in the TIC chromatograms are more obvious between fuel samples than between their unaltered and altered states.

Figure 3. The distribution of average F-ratios for the features in the preliminary hit list for the sample-class comparison of unaltered and acid altered fuels. The distribution is approximately centered on an F-ratio of 1 to 2. The inset plot is the histogram of the tail from 10 to 100, which shows the features which are most likely to be class-distinguishing. There were four features with average F-ratios greater than 100 which are not shown in the inset for clarity.

Figure 4. Illustration of combinatorial rearrangement for 1 of the 200 unique null combinations. The original unaltered class (Sample Class 1) and original altered class (Sample Class 2) are rearranged to create a null distribution from the analysis of the null classes (Null Class A and Null Class B). In this null arrangement, the first injection replicates of all samples are compared against the second injection replicates, for example, $AR_u:1$ is the first injection replicate for unaltered AR fuel, and $AR_u:2$ is the second injection replicate. This rearrangement mitigates the between-class variance that was present in the original sample-class comparison, such that the null hypothesis may instead be tested.

Figure 5. (A) The comprehensive null distributions (200 total) for the 6×6 sample-class comparison of unaltered and altered diesel fuel. Most null F-ratio values were less than 10, with most distributions centered around 1 to 2. (B) All null probability curves; the dashed red line

corresponds to a 0.1% null probability limit. (C) The combined null distribution (blue, long-dashed line), sample-class comparison distribution (red, solid line), from Fig. 3 for comparison, and distribution of 0.1% null probability limits (black, short-dashed line). The 0.1% null probability limit from the combined null distribution is 12.4, and is indicated by the arrow. (D) The tail of the distribution in (C), which shows the relative abundance of higher F-ratio features in the sample-class comparison.

Figure 6. (A) The top twenty hits from the tile-based F-ratio analysis of the acid alteration experiment overlaid on a representative 2D TIC chromatogram from the unaltered fuel class. The blue circles are centered on each hit, and the size of the circle corresponds to the average F-ratio for the hit. The hit numbers for each feature correspond to those in Table 1. (B) The top 107 hits (i.e., all hits above the 0.1% null probability limit of 12.4) from the tile-based F-ratio analysis of the acid alteration experiment.

Figure 7. The quantified peak sums for unique m/z for selected species in the diesel fuel samples. The solid blue bars correspond to the average peak sum of the two injection replicates for the fuel prior to acid alteration, while the patterned red bars represent the peak sum after acid alteration. The numbers in parentheses in the analyte names are the hit numbers if they were found by the tile-based F-ratio analysis.

Figure 8. The quantified peak sums for unique ions for the first four hits from the F-ratio analysis of the diesel fuel samples. The solid blue bars correspond to the average peak sum of the two injection replicates for the fuel prior to acid alteration, while the patterned red bars represent the peak sum after acid alteration. Missing bars indicate that the analytes was not detected in that sample. The numbers in parentheses in the analyte names are the hit numbers from the tile-based F-ratio analysis.

790 **Figure 9.** The top two hits (Hits 1 and 2, tentatively identified as 1-propynylbenzene and sulfur
791 dioxide, respectively) from the tile-based F-ratio analysis of the acid alteration experiment. Blue
792 lines correspond to the unaltered fuel samples while the red lines correspond to the acid altered
793 fuels. (A-B) The total ion current (TIC) chromatograms for the 1D chromatographic regions
794 centered on the analyte features for hits 1 and 2, respectively. (C-D) The extracted ion current
795 (XIC) chromatograms for selective mass channels for the 1D regions for hits 1 (m/z 116) and 2
796 (m/z 64), respectively. Arrows indicate the 2D peak apices as identified by the pinning and
797 clustering step of the software. Quantitative information for the peaks is provided in Figure 8.

Figure1

Figure2

Figure3

Figure4

Figure5

Figure6

Figure7

Figure8

Figure9

Table1

Hit no.	Avg F-ratio	Class Change	Null Prob	¹ t _R (s)	² t _R (s)	Preliminary ID
1	410.4	-	0.00%	733	0.61	1-propynylbenzene
2	153.2	+	0.00%	60	0.10	sulfur dioxide
3	122.4	-	0.00%	1115	0.56	1,1-dimethyl-1H-indene
4	122.4	-	0.00%	933	0.66	1-methyl-1H-indene
5	94.6	-	0.00%	638	0.54	2-propenylbenzene
6	93.5	-	0.00%	1132	0.59	1,2-dimethyl-1H-indene
7	92.1	-	0.00%	2427	0.95	2-methylpyrene
8	86.6	-	0.00%	2402	0.98	1-methylpyrene
9	83.6	-	0.00%	2549	0.93	1,3-dimethylpyrene
10	81.1	-	0.00%	1132	0.49	1,3-dimethyl-1H-indene
11	79.8	-	0.00%	671	0.14	1-decene
12	77.5	-	0.00%	1474	0.67	acenaphthene
13	75.3	-	0.00%	475	0.10	cis-2-nonene
14	74.0	+	0.00%	60	0.61	sulfur dioxide
15	73.4	-	0.00%	644	0.55	(Z)-1-propenylbenzene
16	73.1	-	0.00%	1782	0.61	4,4'-dimethylbiphenyl
17	65.2	-	0.00%	1104	0.51	1,3-dimethyl-1H-indene
18	61.7	-	0.00%	845	0.17	5-methyl-1-decene
19	61.6	-	0.00%	589	0.41	1,2,3-trimethylbenzene
20	60.8	-	0.00%	2434	0.96	3-methylpyrene

Table 1. The top twenty hits from the tile-based F-ratio analysis of the acid alteration experiment. There were a total of 3362 analyte peak features, 107 of which were above the 0.1% null probability limit. 55 hits were found with 0.00% null probability. A 0.00% null probability implies the given hit has an F-ratio that exceeds the largest F-ratio found in all of the 200 null probability distributions. The class change metric indicates whether the analyte is decreasing as a result of acid alteration (-) or increasing as a result of acid alteration (+). The analytes were identified based on 2D retention times and matching to the NIST11 mass spectral library.

Table2

Sample	Class	1-propynylbenzene (Hit 1)	Sulfur dioxide (Hit 2)	1,1-dimethyl-1H-indene (Hit 3)	1-methyl-1H-indene (Hit 4)
AR	U	1700	N.D.	2110	2180
	A	N.D.	47600	N.D.	N.D.
FH	U	1850	N.D.	2220	2230
	A	N.D.	37300	N.D.	N.D.
MB	U	1740	N.D.	2220	2380
	A	N.D.	19500	N.D.	N.D.
SG	U	700	N.D.	2410	2090
	A	N.D.	171000	N.D.	N.D.
SS	U	410	N.D.	1290	1110
	A	N.D.	90900	N.D.	N.D.
RC	U	840	N.D.	2430	2440
	A	N.D.	152000	N.D.	N.D.

Table 2. The mean integrated signals for the top four analytes from the F-ratio analysis, for each of the six fuels that were analyzed, for both the unaltered (U) and altered (A) classes. *N.D.* indicates that the analyte was not detected. Samples AR, FH, and MB were analyzed to discover class-distinguishing analytes. Samples SG, SS, and RC were independent samples that serve to validate the findings of the F-ratio analysis.