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

Supporting Information

Brendon A. Parsons, Luke C. Marney, W. Christopher Siegler, Jamin C. Hoggard, Bob W. Wright, and Robert E. Synovec

Abstract

This document provides additional experimental details, a tutorial on implementing tile based F-ratio analysis, additional results and discussion of hit lists from the experiment, and a comparison to pixel and peak table based methods.

Contents

S1. Additional Experimental Details

S1.1 Concentrations of Analytes in Spiked Diesel Fuel

Nominal Spike Level (ppm)	3-Octanone (ppm)	5-Decyne (ppm)	Bromobenzene (ppm)	1-Chlorohexane (ppm)
100	111	121	133	110
50	55.5	60.5	66.6	54.9
25	27.8	30.3	33.4	27.5
12.5	14.3	15.6	17.2	14.2
6.2	6.89	7.51	8.27	6.81
3.2	3.67	4.00	4.40	3.63
1.6	1.93	2.11	2.32	1.91

Table S-1. The actual concentrations are shown in ppm for each nominal spike concentration (first column) for the four spiked analytes in a diesel fuel matrix. The spiked analytes are not initially present in the non-spiked diesel fuel (matrix blank).

Nominal Spike Level (ppm)	3-Octanone (pg)	5-Decyne (pg)	Bromobenzene (pg)	1-Chlorohexane (pg)
100	461.8	503.4	553.3	457.6
50	230.9	251.7	277.1	228.4
25	115.6	126.0	138.9	114.4
12.5	59.5	64.9	71.6	59.1
6.2	28.7	31.2	34.4	28.3
3.2	15.3	16.6	18.3	15.1
1.6	8.0	8.8	9.7	7.9

Table S-2. The mass quantity injected on column is shown in picograms (pg) for each nominal spike concentration (first column) for all spiked analytes. A 1 µL injection of each diesel fuel sample was made in split mode with a split ratio of 200:1.

S1.2 Instrument Parameters

The GC×GC–TOFMS instrumental platform 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×GC (column 1) was a 20 m x 250 μm i.d. x 0.5 μm RTX-5MS film (Restek, Bellefonte, PA) and the secondary column (column 2) was a 2 m x 180 μm i.d. x 0.2 μm RTX-200 film (Restek, Bellefonte, PA). The GC instrument inlet was set at 275 °C and the transfer line was set at 305 °C. Column 1 was held at 50 °C for 0.25 min and then increased at 5 °C/min to 300 °C, where it was held for 5 min. Column 2 was initially set at 55 °C and followed the same temperature program as column 1 giving a total run time of 55.25 min. The modulator was kept 20 °C higher than column 1, and the modulation period was 1 s. The GC instrument was set to maintain a constant (ambient temperature and pressure corrected) flow rate of 2 mL/min at the outlet of column 2, with helium used as the carrier gas. The ion source was set to 300 °C and the detector voltage was set to 1600 V. Mass channels, *m/z* 41-340, were collected at 100 spectra/s after a 6 s solvent delay. A 1 μL injection of each diesel sample was made in split mode with a split ratio of 200:1. Each diesel sample was injected in quadruplicate, however a total of eight injection replicates were collected for the 0 ppm diesel sample for null distribution analysis.

Figure S-1 is a representative GC×GC–TIC chromatogram indicating the locations of the spiked analytes. The separation conditions were selected to allow moderate wraparound to more fully utilize the 2D peak capacity. However, compounds in a given second dimension separation

were not allowed to wraparound into compounds eluting in a subsequent second dimension separation. While the 2D separation may be reregistered for aesthetic considerations, reregistration was not performed since it has no consequence in this study.

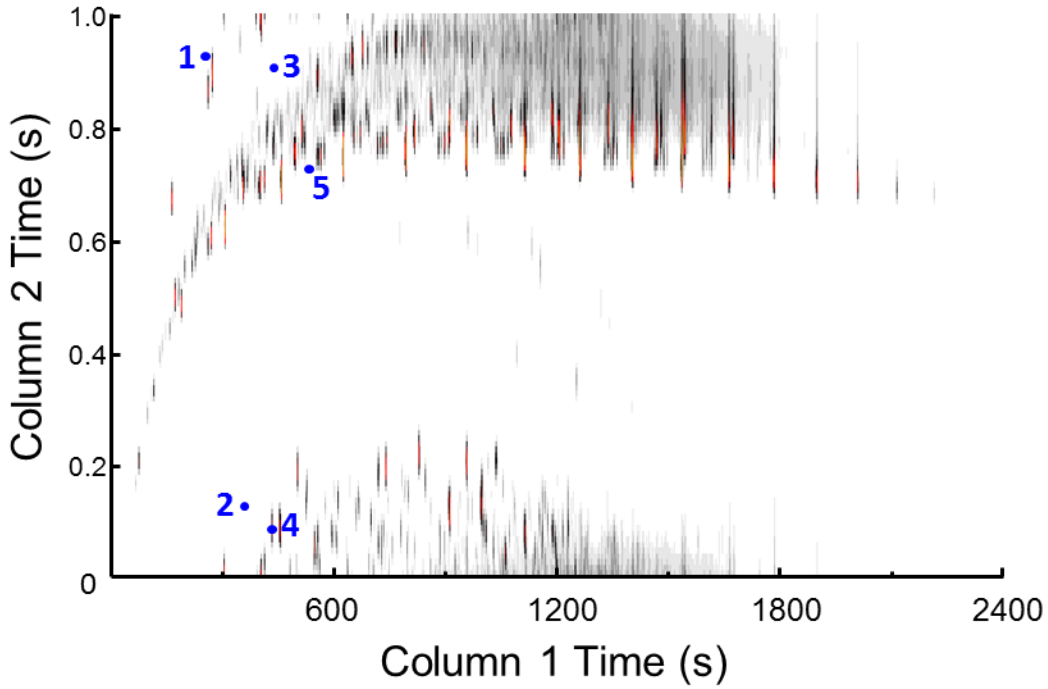

Figure S-1. GC×GC–TOFMS $\log_{10}$ plot of the total ion current (TIC) chromatogram of a 12.5 ppm spiked diesel sample. Locations of all four spiked analytes and the internal standard are marked. The column 1 (first separation dimension) elution order is as follows: (1) 1-chlorohexane; (2) bromobenzene; (3) 3-octanone; (4) 1-bromoheptane (internal standard); (5) 5-decyne.

S1.3 F-ratio Tiling and Redundant Hit Removal Parameters

The GC×GC pixel level data was summed using a 2D grid of 2D tiles, as a function of mass channel (*m/z*), which provides data reduction along both separation dimensions. The tile size chosen for the chromatograms was 6 data points (pixels) in the column 1 dimension (6

modulations) by 10 data points (pixels) in the column 2 dimension (10 mass spectra, or 100 ms). Each tile thus contains an area of 60 pixels. All tile based F-ratio analyses were conducted with the entire collected *m/z* range (41-340). Prior to tile based F-ratio analysis, the data were baseline corrected and normalized to the internal standard signal for each sample injection. A signal-to-noise (*S/N*) threshold was applied to computationally exclude *m/z* having low *S/N* from the F-ratio calculation. This *S/N* threshold was set to a signal equal to three times the standard deviation (3σ) of a tiled noise region, which was taken as the first 10 s of detected signal, during which no peaks eluted, from a representative chromatographic run. The F-ratio was then calculated for each 2D tile for each *m/z*. The average F-ratio, which is used to rank the tiles by significance, was calculated by averaging the F-ratios for each *m/z* which passed the *S/N* threshold for a given 2D tile, with the requirement that the tile have at least three *m/z* above the *S/N* threshold for being included in the analysis. Redundant hits were removed using a novel "pin and cluster" algorithm, as detailed in Section S2.4. The 2D chromatographic parameters for removing redundant hits were ±2 data points (modulations) in the column 1 dimension and ±5 data points (mass spectra) in the column 2 dimension.

S2. Tutorial on Implementing Tile Based F-ratio Analysis

S2.1 F-ratio Calculation

The class-to-class variation is calculated as

$$\sigma_{\text{cl}}^2 = \frac{\sum(\bar{x}_i - \bar{x})^2 n_i}{(k - 1)} \quad (1)$$

where n_i is the number of measurements in the ith class, $\bar{x}_i$ is the mean of the ith class, $\bar{x}$ is the overall mean, and k is the number of classes. The within-class variation is calculated as

$$\sigma_{\text{err}}^2 = \frac{\sum \left(\sum\left(\bar{x}_{ij} - \bar{x}\right)^2\right) - \left(\sum(\bar{x}_i - \bar{x})^2 n_i\right)}{(N - k)} \quad (2)$$

where $\bar{x}_{ij}$ is the ith measurement of the jth class, and N is the total number of sample profiles. A F-ratio is then calculated as the ratio between the two variances,

$$\text{Fisher ratio} = \frac{\sigma_{\text{cl}}^2}{\sigma_{\text{err}}^2} \quad (3)$$

S2.2 Overview of Preprocessing

Preprocessing steps reduce the inherent instrumental variation in the data set that contributes to the within-class variation (Eq. 2, and the denominator in Eq. 3). Baseline correction corrects low frequency noise resulting from fluctuations in the mass spectrometer, column flows, and GC oven temperature. Baseline correction can also remove background resulting from column bleed at higher oven temperatures. Baseline correction techniques have

been thoroughly reviewed elsewhere and have not been included here for brevity.[9] The baseline correction technique applied in the tile based F-ratio software is based on a rolling minimum method which relies on the assumption that each column 2 separation (i.e., each modulation) will have a region in which no peaks elute. As long as the separation conditions are designed such that the most retained peaks on column 2 do not wrap around onto less retained peaks on column 2, this assumption holds. In cases where peaks substantially tail on column 2, such as due to chromatographic overloading of the stationary phase, this method of baseline correction may cause a negative bias on the affected *m/z*. In the chromatograms studied herein, the data were appropriate for this method of baseline correction.

Due to the inherent variation in volumetric injection by microsyringe using a GC autosampler, normalization is necessary to avoid excessive variation in signal caused by variation in the amount of sample loaded onto the analytical column. Such variation may be corrected by use of an internal standard. The integrated signal of the internal standard, which is at the same concentration in each sample, is indicative of the amount of sample injected on column. Normalization was performed by multiplying every point in each chromatogram by a scalar determined by the relative abundance of the internal standard in its respective chromatogram. The data reported herein were normalized to a non-native internal standard, 1-bromoheptane.

S2.3 Tile Approach for Binning GC×GC–TOFMS Data

The 2D misalignment of GC×GC–TOFMS chromatographic data across different samples occurs due to minor fluctuations in the mobile phase flow or oven temperature from run to run, or slight changes in the analytical column due to fouling or column maintenance (e.g., column clipping, bakeouts, etc.). Such instrumental fluctuations may cause minor shifts in peak retention times on column 1 and/or column 2. As previously reported,[18] when the data is processed in a pixel based method, these slight shifts may increase the within-class variance for a peak, which commonly diminish the rank of true positives. Further, if the retention time variation coincides with the sample classes, it may lead to the observation of false positives. For pixel based methods, it is possible to align the chromatograms to reduce these occurrences; however, aligning the column 1 dimension can be especially problematic (to the point of being futile) due to low data density in this dimension. It is advisable to have ~ 15 or more data points (i.e., mass spectra) across a peak for reliable alignment.[27] However, most GC×GC–TOFMS analyses are performed with reduced data density on the column 1 separation to optimize the column 2 separation. Even when the optimization is balanced for both of the two separation dimensions, it is common to have only ~ 2 to 4 data points (i.e., modulations) per peak on column 1 (at the $\pm 2\sigma$ width), as was observed in this data set, which had a typical modulation ratio of 3 to 4, depending on the peak width on column 1.

The tile based approach avoids the need for explicit alignment of the peaks by summation of the peak window prior to the calculation of the F-ratio. The size of the tile is a balance of two competing objectives: one, capturing most of the 2D peak signal (using the $\pm 2\sigma$ width) plus

additional space to allow for minor retention time shift in each dimension, and two, keeping the tile size sufficiently small to maintain selectivity for individual peaks, that is, avoiding sampling neighboring peaks. The column 1 tile size for this data set was 6 modulations (6 s), which allowed for shifts of one modulation in either direction. At the modulation ratio applied for this study, the peaklets farthest from the column 1 peak apex contain relatively small amounts of the total peak area, therefore, we anticipate that the performance of the software would not be affected even by a two-modulation (2 s) shift in the column 1 retention time from sample to sample. However, more substantial shifts in column 1 retention time, which were not encountered in this study, would be likely to decrease the ability of the tile based method to discover true positives, and may also lead to the observance of false positives if the retention time shifts spuriously co-vary with the sample classes. The ability of the tile based method to mitigate retention time shifts is an aspect deserving of further study, particularly in context of modulation ratio. In addition to mitigating the effects of minor column 1 and/or column 2 misalignment, this method also improves the *S/N* by summing the signal within each tile.

The tile based approach assists in the discovery of changing analytes and reduces the number of false positives due to retention time misalignment and covariance of detector noise with the sample classes. However, a single grid of tiles is not sufficient to ensure discovery of all changing analytes regardless of their chromatographic location; instead, it is necessary to use four grids. Figure S-2 shows how four grids are applied to a 2D section of the GC×GC chromatogram. A single grid results in the splitting of analyte peaks into multiple tiles, which initially seems to diminish the advantages of tiling (until redundant hit removal is applied as is

described in Section S2.4). By applying four overlapping tile grids, we ensure that each peak is optimally sampled by a tile in one of the tile grids.

As a consequence of the four grids, each analyte is sampled via the tile based F-ratio software multiple times, which can lead to multiple tile hits for an analyte which is significantly changing between sample classes (Table S-3). In complicated sample matrices, such as diesel, the four grids may also sample neighboring peaks, precluding the possibility of simply using only mass spectral matching to eliminate the redundant hits for a given analyte. To effectively remove redundant hits, it is necessary to transform the tile based F-ratio results from tile grid-based space back to pixel based space, taking advantage of the resolution originally present in the 2D chromatographic data.

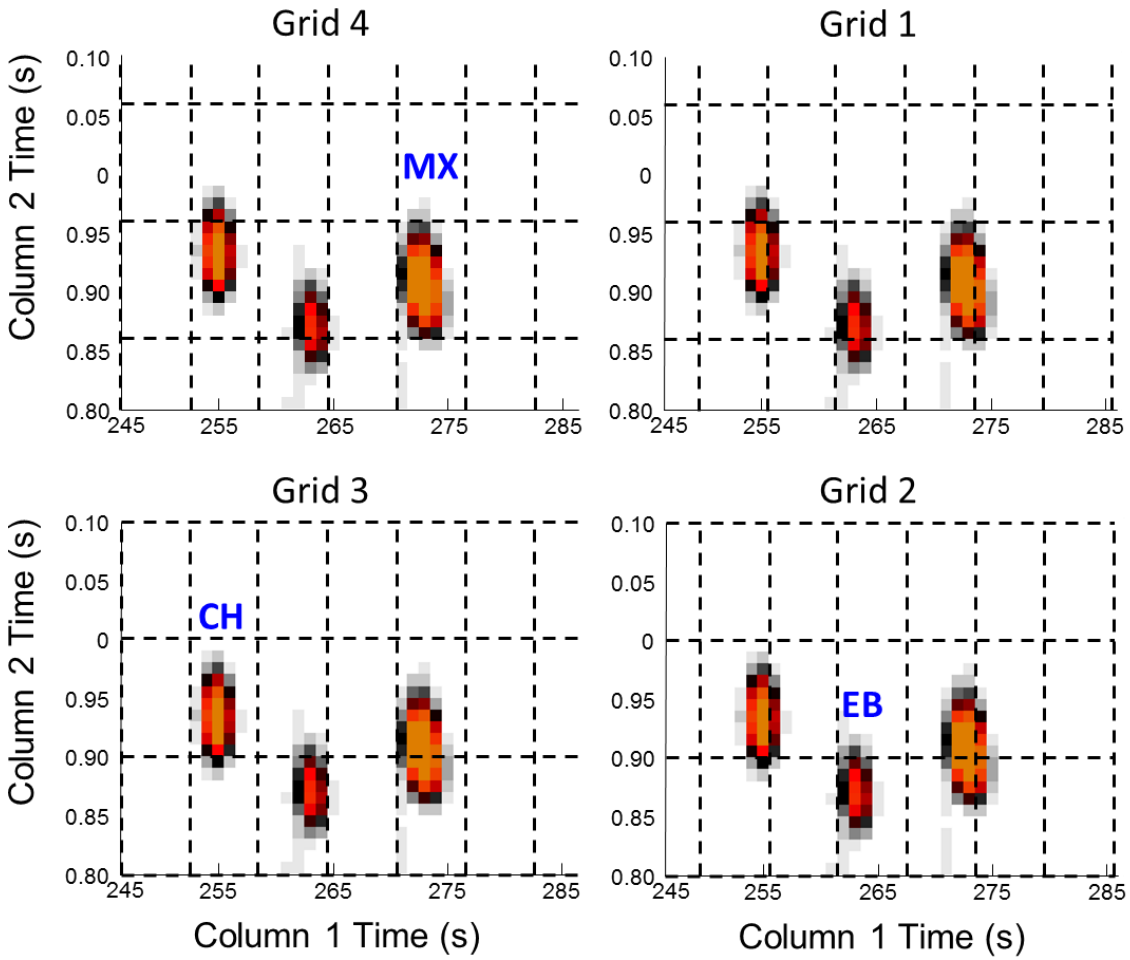

Figure S-2. An analytical ion chromatogram (AIC) comprised of m/z 41:43+55:57+93 for a 25 ppm spiked diesel fuel is shown for a portion of the 2D separation with the four grids overlaid. Each complete tile in the four grids samples 6 data points (6 modulations or 6 s) on the column 1 dimension and 10 data points (0.1 s) on the column 2 dimension. The four grids ensure that each peak is more optimally sampled by one of the tiles. The peaks are labeled in the particular grid containing the tile that best samples each peak. 1-Chlorohexane (**CH**) is best sampled in Grid 3; ethylbenzene (**EB**) is best sampled in Grid 2; and m-xylene (**MX**) is best sampled in Grid 4. The peaks were identified by mass spectral matching to the NIST11 mass spectral library and reported retention indices. While the use of four grids allows each peak to be more optimally sampled by one tile, it also results in the occurrence of redundant hits when peaks are split by non-optimal tiles. In the case of 1-chlorohexane (**CH**), seven tiles capture sufficient peak signal to generate hits.

F-ratio hit no.	Average F-ratio	Tile no., ^1D	Tile no., ^2D	Grid	Compound
1	320.3	60	2	2	**bromobenzene**
2	260.5	43	10	3	**1-chlorohexane**
3	209.1	60	1	1	**bromobenzene**
4	68.5	43	9	4	**1-chlorohexane**
5	42.5	89	7	1	**5-decyne**
6	40.2	455	10	1	false positive
7	37.4	42	10	2	**1-chlorohexane**
8	33.5	42	9	1	**1-chlorohexane**
9	31.8	74	10	3	**3-octanone**
10	21.9	60	2	1	**bromobenzene**
11	20.2	60	1	4	**bromobenzene**
12	19.0	42	10	1	**1-chlorohexane**
13	17.2	60	2	3	**bromobenzene**
14	16.0	120	6	2	false positive
15	15.8	43	10	4	**1-chlorohexane**
16	13.8	208	4	4	false positive
17	12.9	176	5	4	false positive
18	12.2	74	9	4	**3-octanone**
19	11.8	74	10	4	**3-octanone**
20	11.6	89	8	2	**5-decyne**
21-14070	…	…	…	…	false positives

Table S-3. The top twenty entries in the initial hit list from the tile based F-ratio software (prior to redundant hit removal and null classification) applied to the comparison of the nominal 25 ppm spike level versus the 12.5 ppm spike level. There are many redundant hits associated with each spiked analyte that will be removed by pinning and clustering. There are a total of 14070 entries in the list. The spiked analytes were identified by matching mass spectra and retention times to analyte standards. The tile size was 6 s on column 1 by 0.10 s on column 2.

S2.4 Redundant Hit Removal by "Pinning and Clustering"

The first tile based F-ratio report[18] demonstrated that the tile based approach provided a computationally-fast way to improve the sensitivity contrast between true positives and false positives for discovery-based analyses. However, the final analysis was somewhat complicated by the presence of multiple hits per "discovered" analyte: the minor hits had to be removed by hand by the analyst in order to obtain a hit list with a single entry for each class distinguishing analyte. While there should ideally be one tile hit per class distinguishing analyte, the four grids used to bin the data result in multiple samplings of a given peak (as shown in Figure S-2), which leads to multiple hits for class distinguishing analytes. These multiple features per class distinguishing analyte are referred to as redundant hits. For an efficient analysis, redundant hits must be removed automatically. Herein, we introduce and describe an algorithmic method to remove redundant hits by focusing the multiple 2D tile locations back to the original high-resolution 2D chromatographic data.

Since redundant hits are due to the same analyte being sampled multiple times by the tile based approach, redundant hits have very similar, if not the same 2D chromatographic peak location. Briefly, the pinning algorithm analyzes each hit found by the initial tile based F-ratio approach and locates the maximum signal difference (between the sample classes) observed at the *m/z* with the highest F-ratio associated with each tile. The *m/z* with the highest F-ratio is used because it is the most selective ion for the class-distinguishing peak; chromatographic interferents which are not changing between sample classes have *m/z* with lower F-ratio values, and are not selective for the peak of interest. The locations of the maximum signal differences for the top F-ratio *m/z* for each tile are then denoted in the 2D chromatographic space, analogous

to how locations may be pinned on a map. The corresponding information from the each tile hit, including the hit's array of F-ratio values at each *m/z* for that location (which comprises an F-ratio "spectrum"), the average F-ratio value, and the original tile and grid locations, is indexed to its respective pin. Since redundant hits are attributable to the same analyte peak eluting at the same, or similar, 2D retention times, multiple pins for the same analyte peak are consolidated into small regions.

Next, with the use of a cluster algorithm, we remove the pin locations with redundant F-ratio information in an automated fashion. The cluster algorithm ranks the pins by their associated average F-ratios, and then removes those that are within a user-specified 2D chromatographic distance from one another. This approach is based on the observation that the pin having the highest average F-ratio optimally locates the peak maximum of the class distinguishing analyte. The highest F-ratio pin in a given cluster is preserved and assigned as a hit in the final hit list, while the lesser pins (i.e., lesser redundant hits) are removed. The window locations that are indexed with the pins can be easily used for further deconvolution or identification, and the F-ratio spectra for that 2D window location provides the particular *m/z* that are the most important for the comparison of the two sample classes and are the most chemically selective in the chromatographic separation for the peak of interest.

The dimensions of the cluster window are based on the observed distribution of pins for a typical class distinguishing feature, which is smaller than the tile, so the cluster window is sized to capture the majority of the peak signal, as well as to allow for minor retention time variance. Further, depending on chromatographic interference from sample matrix peaks, the best tile may

not be centered on the class distinguishing feature. During the process of redundant hit removal, the peak maximum is located and indexed, allowing the cluster window to be properly centered on the peak, and simplifying further analysis of the data. The cluster window dimensions for this study were ±2 s from the center in the column 1 dimension (total of 5 modulations) and ±0.05 s in the column 2 dimension (total of 11 mass spectra).

Figure S-3 illustrates the process of redundant hit removal for the 1-chlorohexane peak in the 25 ppm versus 12.5 ppm comparison, using the pinning and clustering algorithms. The 2D peak is the average of the 25 ppm injection replicates minus the average of the 12.5 ppm replicates using *m/z* 55, which is the highest F-ratio *m/z* for 1-chlorohexane. This "average difference" peak is used to determine the location of the pin (i.e., the 2D chromatographic location that has the greatest difference between classes). The maximum of the average difference peak corresponds to the peak maxima for the 25 ppm class and the 12.5 ppm class. The box with a dashed black line is the tile that best sampled the 1-chlorohexane peak (see Figure S-2). The multiple tiles that sampled the 1-chlorohexane peak are each assigned ("pinned") to the location within the respective tiled 2D chromatographic window, as indicated by a white star. A small star indicates that a single tile was pinned to that location, while a large star indicates that two tiles were pinned to that location; there are a total of seven pins (i.e., a total of seven tiles sampled the 1-chlorohexane peak, capturing sufficient class distinguishing signal to generate a hit). The cluster window, represented by the box with a solid red line, is centered around the pin with the highest F-ratio (in this instance, at 254, 0.94). As shown, the centered cluster window captures all of the pins for the 1-chlorohexane peak. The pin with the

highest F-ratio in the cluster window is retained in the hit list, and the other pins are removed as redundant hits. This process is repeated for all tile hits until there is only one pin remaining per discovered feature (i.e., no redundant hits within the defined cluster boundaries).

Table S-4 is the hit list for the 25 ppm versus 12.5 ppm comparison after redundant hits were removed by the automated pinning and clustering algorithms. Whereas prior to redundant hit removal (Table S-3) there were many hits for each class distinguishing analytes, after redundant hit removal (Table S-4) there is only one hit per analyte. Additionally, the redundant hit removal focuses the tiled results back to the pixel level so that the hit list report provides an accurate determination of the 2D retention times of the discovered class distinguishing analyte features (Table S-4). Table S-5 and Table 1 (primary manuscript) provide another example of redundant hit removal using the 12.5 ppm versus 6.2 ppm comparison, with hit lists prior to and following redundant hit removal, respectively. The complete tile based F-ratio analysis procedure is summarized by a flowchart in Figure S-4.

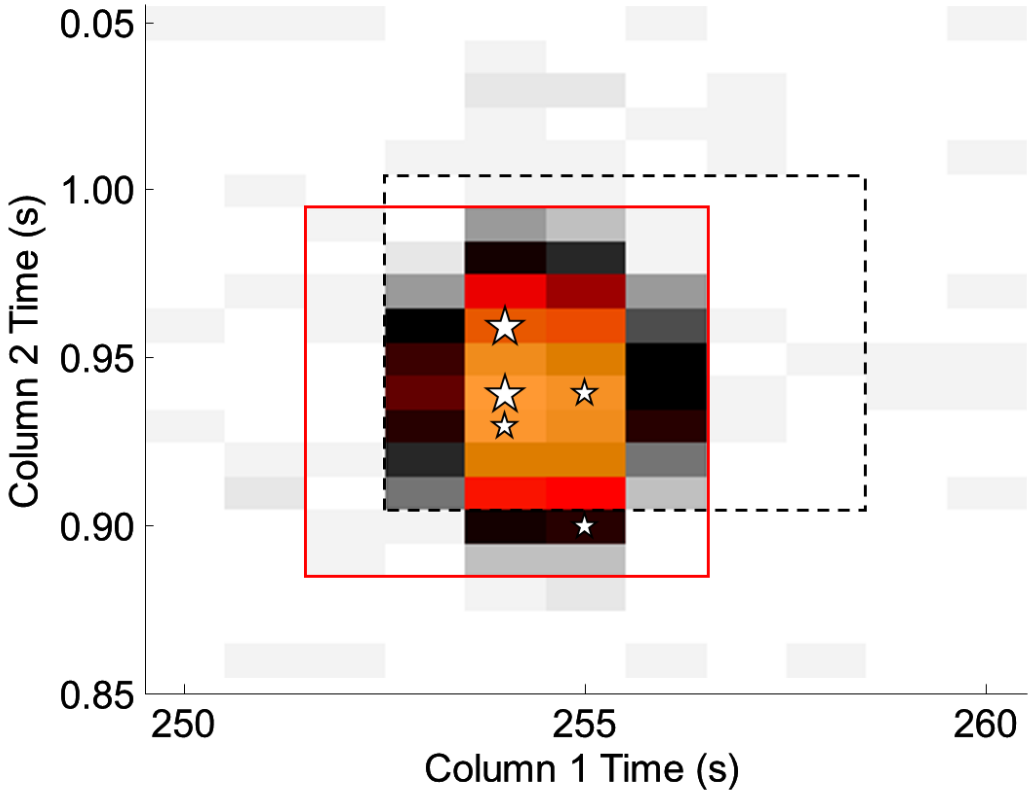

Figure S-3. An example of redundant hit removal applied to the 1-chlorohexane peak in the 25 ppm versus 12.5 ppm comparison. The peak is the average of the 25 ppm injection replicates minus the average of the 12.5 ppm replicates using *m/z* 55, which is the highest F-ratio *m/z* for 1-chlorohexane.

F-ratio hit no.	Average F-ratio	$^{1}t_R$ (s)	$^{2}t_R$ (s)	Null classification	Compound
1	320.3	360	0.14	hit	**bromobenzene**
2	260.5	254	0.94	hit	**1-chlorohexane**
3	42.5	534	0.73	hit	**5-decyne**
4	40.2	2733	0.99	hit	false positive
5	31.8	441	0.92	hit	**3-octanone**
6	16.0	720	0.56	potential hit	
7-4793	…	…	…	potential hits and non-hits	

Table S-4. The hit list for the 25 ppm versus 12.5 ppm comparison, after redundant hit removal by the pinning and clustering algorithms. The four spiked analytes are found within the first five entries in the hit list. Compare to Table S-3, for results prior to redundant hit removal.

F-ratio hit no.	Average F-ratio	Tile no., ^{1}D	Tile no., ^{2}D	Grid	Compound
1	214.4	60	1	1	**bromobenzene**
2	114.5	43	9	4	**1-chlorohexane**
3	107.2	43	10	3	**1-chlorohexane**
4	104.9	60	2	2	**bromobenzene**
5	97.5	42	9	1	**1-chlorohexane**
6	68.0	42	10	2	**1-chlorohexane**
7	60.1	89	7	1	**5-decyne**
8	21.0	61	1	4	**bromobenzene**
9	20.2	61	2	3	**bromobenzene**
10	16.9	43	10	4	**1-chlorohexane**
11	15.5	60	1	2	**bromobenzene**
12	14.8	157	5	1	false positive
13	14.0	42	10	1	**1-chlorohexane**
14	13.0	60	1	4	**bromobenzene**
15	12.5	60	2	3	**bromobenzene**
16	12.3	89	7	4	**5-decyne**
17	11.6	60	2	1	**bromobenzene**
18	11.6	200	6	1	false positive
19	11.2	74	10	3	**3-octanone**
20	10.0	89	8	2	**5-decyne**
21-14253	…	…	…	…	false positives

Table S-5. The hit list for the 12.5 ppm versus 6.2 ppm comparison, prior to redundant hit removal and null classification. The hit list following redundant hit removal is included in Table 1 (primary manuscript).

S2.5 Flowchart of Tile Based F-ratio Software Functions

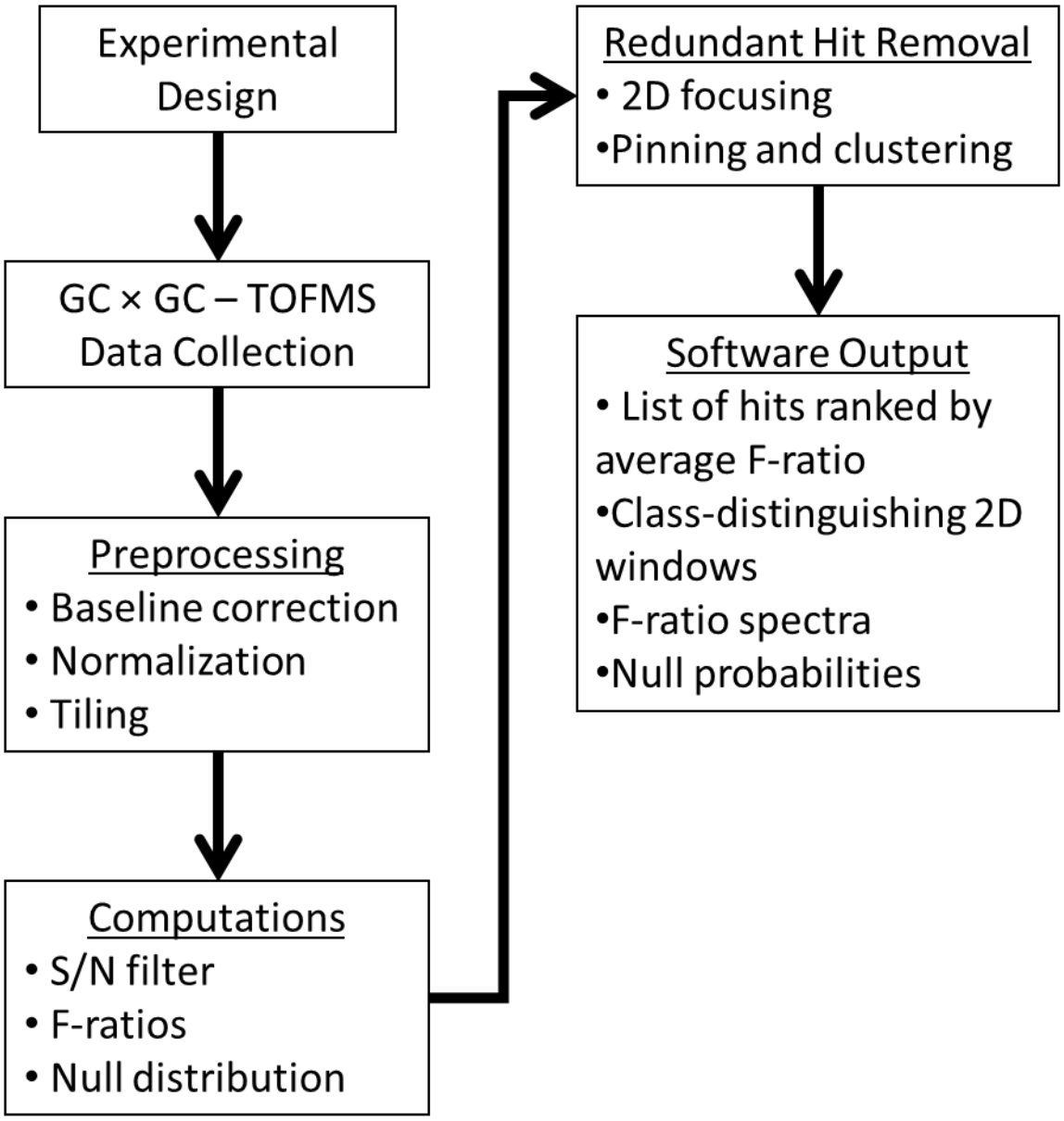

Figure S-4. A summary of steps comprising the tile based F-ratio software.

S3. Additional Discussion of Hit Lists

Tables S-4, S-6, and S-7 are hit lists for additional interesting concentration comparisons from the range of concentration levels studied for this publication. Table S-4 is the hit list for the 25 ppm versus 12.5 ppm comparison, in which all four spiked analytes are found with average F-ratios that exceed the upper limit of the 0.1% null probability range (i.e., greater than 20, as explained in the primary manuscript and illustrated in Figure 1). There is a single false positive hit interspersed with the four analytes. Tables S-6 and S-7 are selected hit lists from the spike versus matrix blank comparisons. Table S-6 is the hit list for the 12.5 ppm versus 0 ppm comparison, in which all four spiked analytes are found with average F-ratios above the 0.1% null probability range (see Figure 3). Due to greater between-class variation with the spike versus blank comparisons, possibly due to sample handling, there are more false positive hits observed compared to the concentration ratio of 2 comparisons. The 12.5 ppm versus 0 ppm comparison has a total of nine false positives with average F-ratios above 20. Table S-7 is the hit list for the lowest absolute concentration studied, the 1.6 versus 0 ppm comparison. At this level, three of the four spiked analytes are found with average F-ratios above the 0.1% null probability range, and 3-octanone is found in the null probability range of 10 to 20. As with the 12.5 ppm versus 0 ppm comparison, there are a larger number of false positives with average F-ratios above the 0.1% null probability limit, in this case, 24 false positive hits.

F-ratio hit no.	Average F-ratio	^1t_R (s)	^2t_R (s)	Null classification	Compound
1	2436.1	360	0.13	hit	**bromobenzene**
2	1333.5	255	0.93	hit	**1-chlorohexane**
3	518.4	534	0.73	hit	**5-decyne**
4	67.4	441	0.91	hit	**3-octanone**
5	43.5	109	0.30	hit	false positive
6	32.2	49	0.62	hit	false positive
7-4895	…	…	…	potential hits and non-hits	

Table S-6. The hit list for the 12.5 ppm versus 0 ppm comparison. The four spiked analytes are found as the first four entries in the hit list.

F-ratio hit no.	Average F-ratio	^1t_R (s)	^2t_R (s)	Null classification	Compound
1	315.0	255	0.93	hit	**1-chlorohexane**
2	257.0	360	0.13	hit	**bromobenzene**
3	36.9	185	0.5	hit	false positive
4	33.6	534	0.73	hit	**5-decyne**
5-37	…	…	…	Hits and potential hits	
38	16.8	441	0.91	potential hit	**3-octanone**
39-4913	…	…	…	non-hits	false positives

Table S-7. The hit list for the 1.6 ppm versus 0 ppm comparison. The four spiked analytes are found within the first 38 entries in the hit list. There are 24 false positive hits interspersed with the hits for the spiked analytes. The distribution of false positives for the 1.6 ppm versus 0 ppm comparison has a substantial tail, which manifests as a prevalence of more false positives.

S3.1 Graphical Summary of 3-octanone for the 6.2 ppm versus 0 ppm Comparison

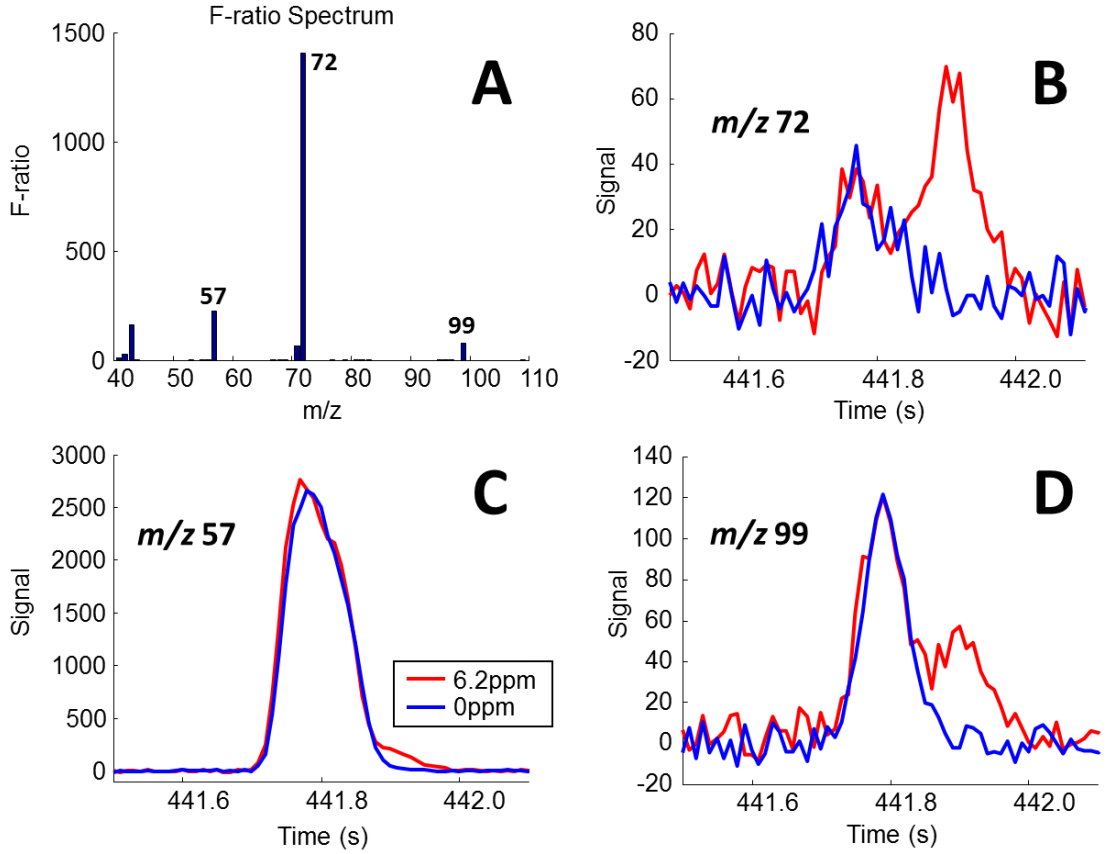

Figure S-5. 3-octanone in the 6.2 ppm versus 0 ppm comparison. (A) The F-ratio spectrum. (B-D) The three most selective *m/z* are individually plotted for an injection from each the 6.2 ppm spiked diesel sample (red trace) and the 0 ppm diesel (blue trace). The plots are extracted ion chromatograms (XIC) from the column 2 separation for the most intense modulation of the 3-octanone peak at the 6.2 ppm spike level. The rightmost peak is 3-octanone, while the leftmost peak is comprised of several interfering compounds.

Figure S-5 provides an illustration of the exquisite sensitivity and selectivity provided by the tile based F-ratio software, in this case for the "discovery" of 3-octanone above the 0.1% null probability range. At this level, 3-octanone has an average F-ratio of 86.4, substantially above the 0.1% null probability range of 10-20. Figure S-5(A) is the F-ratio spectrum (the F-ratio for each *m/z* that was present in the tile above the *S/N* threshold) for the hit corresponding to 3-

octanone. The F-ratio spectrum is plotted to *m/z* 110, as there were no significant F-ratios above this *m/z*. The *m/z* 57, 72, and 99 are the most selective versus the significant matrix peaks, which overlap the 3-octanone peak at low chromatographic resolution in both separation dimensions. Figures S-5(B-D) includes segments of the column 2 separation from the most abundant modulation of 3-octanone, displayed as extracted ion chromatograms (XIC) at *m/z* 57, 72, and 99. These chromatograms illustrate both the low absolute abundance of 3-octanone at this challenging concentration level, as well as the limited chromatographic resolution and mass spectral selectivity versus the interfering matrix peak. Despite low *S/N* and low mass spectral selectivity, 3-octanone has a significant average F-ratio value and is easily found in the output of the tile based F-ratio software as the fourth feature in the hit list, above the 0.1% null probability limit. Figure S-6 is the mass spectrum for 3-octanone, as obtained from the 2011 NIST Mass Spectral Library. Comparing the F-ratio spectrum in Figure S-5(A) to the mass spectrum (Figure S-6), we see that as expected the two spectra have several *m/z* in common, though the relative intensities of the *m/z* are altered by the mass spectral selectivity in the F-ratio calculation for the various chromatographic interferents.

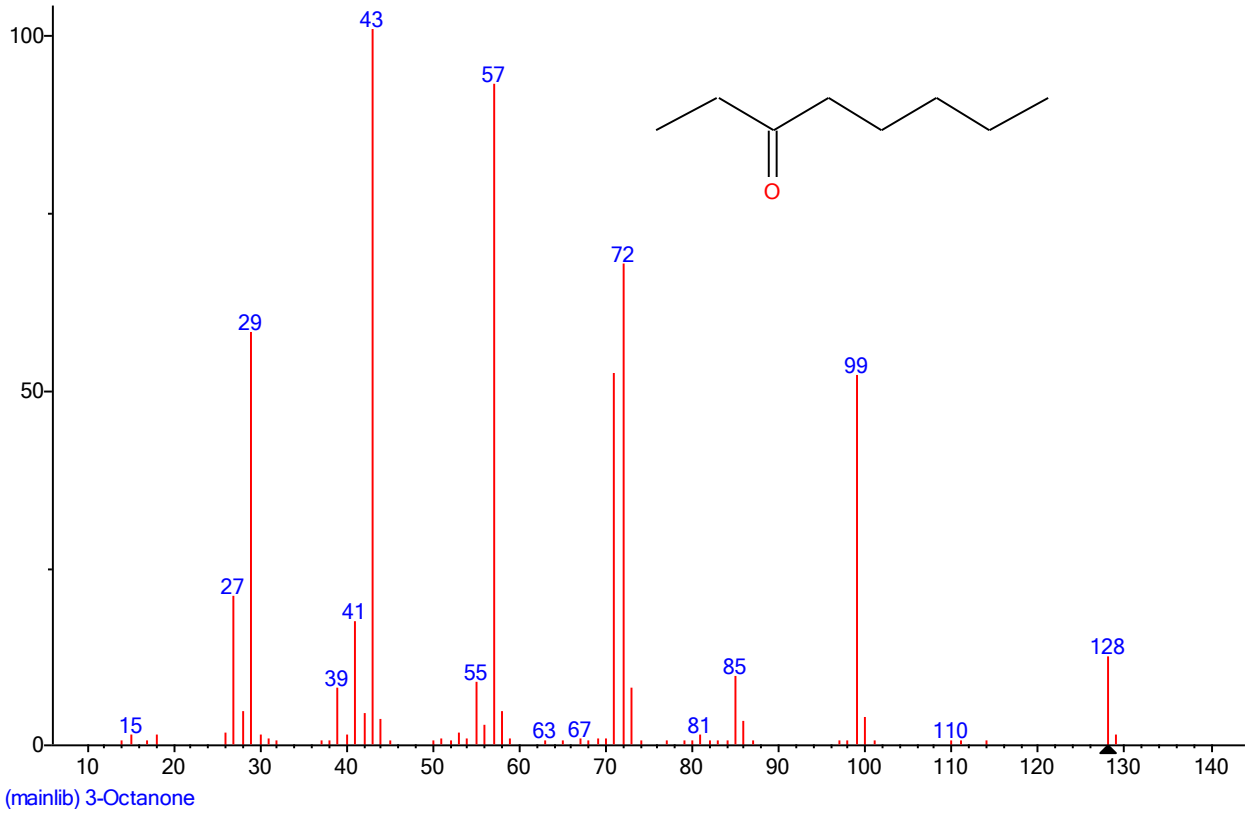

Figure S-6. The mass spectrum for 3-octanone, as obtained from the 2011 NIST Mass Spectral Library.

S4. Comparison of Tile Based Software to Pixel Based Software and Peak Table Based Methods

The tile based F-ratio software comprehensively analyzes GC×GC–TOFMS data while simultaneously maximizing discovery of true positives and minimizing the discovery of false positives. The approach taken by the tile based software is most similar to that of pixel based analysis,[28-33] in that the tile based software aims to perform minimal processing prior to statistical analysis of the data to find class distinguishing features. Specifically, tile based[18] and pixel based[28-33] feature selection analyses avoid peak finding, explicit 2D alignment of the chromatograms, deconvolution, and mass spectral matching to libraries. Rather, the workflow of tile based and pixel based analyses is to first discover the class distinguishing analytes in an experimental comparison, and then apply alignment, deconvolution, and mass spectral matching only as needed on the features which have been selected as having statistical discriminatory power in the analysis. The preceding approaches are in marked contrast to peak table methods, which apply statistical analysis only after generating lists of all of the "analyte" peaks found in the chromatograms, which follows the peak finding, deconvolution, matching, and peaklet combination (for GC×GC–TOFMS) steps.[34-37]

To demonstrate the potential of the tile based software amid other methods applied in the GC×GC–TOFMS field, we have performed preliminary comparisons to both pixel based F-ratio analysis, as well as the peak table based data processing that would occur prior to F-ratio analysis. The analysis which follows is based on the software we have available in our laboratory. At this time we are not able to compare to peak region methods (such as those used in

GC Image), but such a comparison would also be beneficial in future work. The tables presented in the following sections were used to compare to those shown for the tile based results.

S4.1 Pixel Based Analysis Comparison

A pixel based analysis was performed with the in-house developed pixel based F-ratio software[28-31], using parameters matched to those of the tile based software (signal normalization to the bromoheptane internal standard, *S/N* threshold of 3σ, all *m/z* analyzed, etc.). Tables S-8 and S-9 summarize the results of the pixel based F-ratio analysis for the 12.5 ppm versus 0 ppm comparison, and the 1.6 ppm versus 0 ppm comparison, respectively. These pixel based tables correspond to the same concentration comparisons as (and should be compared to) Tables S-6 and S-7 for the tile based analysis. Compared to the tile based software, the pixel based software performs well, finding most of the analytes near the top of the hit list. However, as expected, more false positives are found interspersed between the hits, as compared to the tile based analysis. At the 12.5 versus 0 ppm comparison level, we see that the pixel based analysis (Table S-8) exhibits a false positive at the top of the hit list, outranking even bromobenzene. The tile based analysis (Table S-6) finds bromobenzene as the top hit, and has no false positives between the four spiked analytes. In the significantly more challenging 1.6 versus 0 ppm comparison, the tile based and pixel based analyses each find two of the four analytes at the top of their respective hit lists; however, the pixel based analysis (Table S-9) has approximately twice as many non-discriminating features (76) interspersed between the four spiked analytes, versus the tile based analysis (Table S-7) (34).

F-ratio hit number	Average F-ratio	^1t_R (s)	^2t_R (s)	Compound
1	896.7	1622	0.26	false positive
2	853.2	360	0.11	**bromobenzene**
3	715.5	534	0.73	**5-decyne**
4	568.7	441	0.91	**3-octanone**
5	322.4	255	0.94	**1-chlorohexane**
6	103.8	165	0.07	false positive
7-2428	…	…	…	false positives

Table S-8. The hit list for the pixel based analysis of the 12.5 ppm versus 0 ppm comparison. The four spiked analytes are found as the first five entries in the hit list. The first hit of the analysis is a false positive.

F-ratio hit number	Average F-ratio	^1t_R (s)	^2t_R (s)	Compound
1	376.6	534	0.73	**5-decyne**
2	367.1	360	0.13	**bromobenzene**
3	133.4	2708	0.90	false positive
4	108.1	2415	0.68	false positive
5	102.1	255	0.93	**1-chlorohexane**
6-79	…	…	…	false positives
80	21.0	441	0.91	**3-octanone**
81-2459	…	…	…	false positives

Table S-9. The hit list for the pixel based analysis of the 1.6 ppm versus 0 ppm comparison. There are false positives interspersed with the true positives.

As was observed in table S-6 through S-9, false positives were observed between true positives in both the pixel based and tile based analyses, though to a lesser degree in the latter. Tile based analysis consistently finds the four spiked analytes at the top of the hit list at all spike versus blank comparisons including and above the 6.2 ppm versus 0 ppm comparison. Conversely, pixel based analysis encounters false positives interspersed between true positives even at higher concentration comparisons. Tables S-10 and S-11 compare the results of the pixel

based and tile based methods for the 25 ppm versus 0 ppm comparison. At this concentration level comparison, pixel based analysis finds two false positives between the hits for 3-octanone and 1-chlorohexane. Analyzing the same data, tile based analysis finds no false positives between the four spiked analytes.

F-ratio hit number	Average F-ratio	^1t_R (s)	^2t_R (s)	Compound
1	1102.7	360	0.14	**bromobenzene**
2	765.6	534	0.73	**5-decyne**
3	343.1	441	0.92	**3-octanone**
4	157.4	1003	0.50	false positive
5	143.6	2337	0.95	false positive
6	134.8	255	0.93	**1-chlorohexane**
7-2672	…	…	…	false positives

Table S-10. The hit list for the pixel based analysis of the 25 ppm versus 0 ppm comparison. Two false positives are interspersed with the true positives.

F-ratio hit number	Average F-ratio	^1t_R (s)	^2t_R (s)	Null classification	Compound
1	2167.2	360	0.13	hit	**bromobenzene**
2	1869.7	255	0.94	hit	**1-chlorohexane**
3	271.6	534	0.73	hit	**5-decyne**
4	131.2	441	0.91	hit	**3-octanone**
5-5123	…	…	…	potential hits and non-hits	

Table S-11. The hit list for the tile based analysis of the 25 ppm versus 0 ppm comparison. There are no false positives interspersed with the true positives.

Overall, we found in our comparison of the pixel based and tile based methods that both successfully find analytes spiked into a complex matrix (diesel fuel) at low concentrations (substantially less than 100 ppm) relative to the matrix. Compared to pixel based analysis, tile based analysis demonstrated similar discovery rates for the true positives, with substantially

lower rates for false positives. Whereas pixel based analysis demonstrated false positives interspersed with the four spiked analytes at spike versus blank comparisons of up to 25 ppm versus 0 ppm, tile based analysis applied to the same data did not encounter interspersed false positives until the 3.2 ppm versus 0 ppm concentration comparison.

S4.2 Peak Table Based Comparison

To evaluate the feasibility of peak table based F-ratio analysis for the data analyzed by the tile based F-ratio software, we compiled peak tables using the instrument software (LECO ChromaTOF v 3.32). While we were unable to perform F-ratio analysis, which requires the LECO Statistical Compare add-on, we critically evaluated the peak table results to demonstrate for which analyte concentrations the F-ratio analysis of the peak tables would likely be viable. This critical evaluation allowed us to study both the advantages and disadvantages of peak table analysis and to predict the ultimate performance of a non-targeted analysis of the results. The peak table analysis software that is provided with the LECO Pegasus 4D is a powerful utility which integrates peak finding, deconvolution, mass spectral matching, and peaklet combination (for GC×GC–TOFMS) to generate a list of all peaks found in a chromatogram, along with metrics such as peak height, area, deconvolution purity, etc. Peak tables can be viewed in ChromaTOF or exported as .csv tables to other applications.

An important aspect of peak table analysis, as implemented in ChromaTOF, is the selection of appropriate parameters for the generation of the peak tables. These parameters include selections such as expected column 1 and column 2 peak widths, maximum number of peaks to find, *S/N* thresholds, matching thresholds, and others. Selection of these parameters

greatly influences the success of peak finding and deconvolution, and consequently the discovery rates for true and false positives.[55] In order to best compare, in principle, peak table based methodology to our tile based software, the parameters were selected to balance the need for low detection limits with the preference for reducing false positives. Table S-12 summarizes the relevant data processing parameters selected for the peak table analysis.

Baseline offset	1
Smoothing	Auto
Column 1 peak width (baseline to baseline)	0.2 s
Maximum number of unknown peaks to find	10000
Signal-to-noise threshold	10
Number of apexing masses	1
Match required to combine peaks	800
Column 2 peak width	4 s
Library search mode	Normal/forward
Masses to library search	All (m/z 41:340)
Maximum molecular weight allowed	300
Mass threshold	10
Minimum similarity for name assignment	800
Mass for area/height calculation	DA (deconvoluted analytical ion chromatogram)

Table S-12. The relevant parameters applied for peak table analysis of the data (ChromaTOF v 3.32).

After processing the individual chromatograms, the peak tables were exported to Excel, and the relevant entries (the four spiked analytes, and the internal standard) were manually selected and combined. The peak areas were normalized to the internal standard, 1-bromoheptane. Table S-13 summarizes the results of the peak table analysis. As found in the application of the pixel based and tile based methods, the peak table approach was sensitive to analyte concentration. Peaks were reliably found above the 12.5 ppm concentration level, more

so for the less-interfered bromobenzene and 1-chlorohexane peaks. As evident by the relative standard deviations, the more-interfered 5-decyne and 3-octanone peaks were quantified with lesser precision. In the most lenient approach to an F-ratio analysis on the peak table data, the "discovery limit" (for this study, the nominal concentration level at which the analyte could feasibly be discovered in the course of F-ratio analysis) would be the 3.2 ppm and 12.5 ppm for 5-decyne and 3-octanone, respectively. Below the discovery limits, no peaks were found at the elution times of the respective analytes, as indicated in Table S-13 by "N.F."

	mean area (relative standard deviation) for spiked analyte			
Nominal concentration level	bromobenzene	1-chlorohexane	5-decyne	3-octanone
50 ppm	224767 (4%)	175228 (3%)	127339 (7%)	57834 (55%)
25 ppm	115272 (5%)	87680 (5%)	44211 (23%)	9991 (60%) [b]
12.5 ppm	56438 (7%)	45466 (5%)	15207 (47%)	2437 (35%) [a,b]
6.2 ppm	26239 (5%)	19356 (11%)	2044 (41%) [a,b]	N.F.
3.2 ppm	12623 (7%) [b]	9484 (14%)	842 (54%) [a,b]	N.F.
1.6 ppm	3814 (28%) [b]	2533 (23%) [b]	N.F.	N.F.

Table S-13. Peak table results from the ChromaTOF processing of the chromatograms spanning the spiked analyte concentration levels from 1.6 ppm to 50 ppm. Given are the average peak areas for the four spiked analytes for which a peak was found, along with the relative standard deviation for the four injection replicates. N.F. indicates that no peak was found for the analyte at the given analyte-concentration pair. ([a]) indicates analyte-concentration pairs having fewer than 3 apexing *m/z.* ([b]) indicates analyte-concentration pairs having unidentified or misidentified analytes for at least one of the injection replicates.

In a more stringent usage of peak table data for F-ratio analyses, additional metrics would be applied to the peak table data prior to an F-ratio calculation. An important parameter for ChromaTOF data processing is the number of apexing *m/z* for a given peak. Apexing *m/z* are important for spectral deconvolution; to reduce false positives, the ChromaTOF documentation

recommends that 3 apexing *m/z* be required for peak inclusion; the default is 2 apexing *m/z*, which allows for detection of more peaks, but at the risk of more false positives and splitting of true positives. If at least 3 apexing *m/z* were required for inclusion in a peak table based F-ratio analysis, we find that the discovery limit would increase for 5-decyne and 3-octanone, to 12.5 ppm and 25 ppm, respectively. Analyte-concentration pairs having fewer than 3 apexing *m/z* are denoted by "[a]" in the Table S-13.

An additional constraint that may be applied in the course of an F-ratio analysis using peak table data would be the requirement of reliable mass spectral matching. Under ideal conditions, all analytes would be high *S/N*, free of chromatographic interference, and included in mass spectral libraries. However, since these ideal conditions are not typical of most—if any—real-world analyses, the quality of mass spectral matching must be taken into account. To reduce false positives, a match value of at least 800 (out of 999) was required for name assignment for a given peak. Under this constraint, peaks with insufficient match value, or incorrect matches, would not be included in the F-ratio analysis. Analyte-concentration pairs having at least one replicate that was unidentified or misidentified are denoted by "[b]" in the Table S-13. Under this more stringent usage of peak tables, the discovery limit increases for all of the analytes, to 6.2 ppm (bromobenzene), 3.2 ppm (1-chlorohexane), 12.5 ppm (5-decyne), and 50 ppm (3-octanone). Because the tile based method is able to discover features even under conditions that are challenging for peak finding and deconvolution, its discovery limit for 3-octanone is much lower compared to peak table methods (found as a highly ranked hit even at the 6.2 versus 0 ppm comparison).

Overall, the peak table based F-ratio analysis performs well at higher *S/N*, and for analytes with little chromatographic interference. However, due to inherent limits in peak finding, deconvolution, and mass spectral matching, the peak table analysis was, for this data set, less reliable at lower *S/N*, especially for analyte peaks which are substantially interfered chromatographically, such as 3-octanone. Tile based F-ratio analysis significantly improved the "discovery limits" for the more challenging analytes and demonstrated that when performing discovery-based analysis on complicated chromatographic data sets, statistical analysis prior to deconvolution and careful quantification is both viable and beneficial. We also note that the tile based analysis workflow for discovery of class distinguishing features is substantially faster than peak table processing. A four chromatogram versus four chromatogram comparison using tile based analysis may be performed in less than 2 minutes (10 minutes including null distribution analysis) on a mid-level desktop PC. Using the same PC to process the same eight GC×GC–TOFMS chromatograms for peak table analysis took approximately 80 minutes, which only includes generation of the peak tables; combination and statistical comparison of the peak tables would further add to the computational burden. The computational savings of the tile based analysis provide substantial advantages for rapid discovery, and additionally allows for rapid iteration of parameters when developing methods for a given experiment.

S4.3 Summary of the Comparisons of the Tile, Pixel, and Peak Table Based Methods

Overall, the tile based analysis provides a reduced propensity for the discovery of false positives. Whereas the pixel based analysis encountered false positives interspersed with the four spiked analytes as high as the 25 ppm versus 0 ppm comparisons, the tile based analysis did not

encounter interspersed false positives until the 3.2 ppm versus 0 ppm comparison. Compared to pixel based and tile based analyses, which perform minimal processing of the data prior to statistical analysis, peak table analysis performs peak finding, deconvolution, quantification, and matching prior to the statistical analysis of the processed data. Under typical use of peak table methods, this requires that the potentially class distinguishing peaks are properly located and identified in order to be included in the statistical analysis. In a stringent peak table F-ratio analysis where the parameters are chosen to reduce false positives, the lowest concentrations at which analytes are reliably discovered was be substantially higher than that of the pixel based or tile based methods. While the limit of discovery 3-octanone was ~56 ppm using the peak table method, this challenging analyte is easily discovered at the ~7 ppm when using the either the tile based or pixel based methods.