



Implementation of Fisher ratio analysis to improve classification of sulfur-contaminated jet fuels using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry data

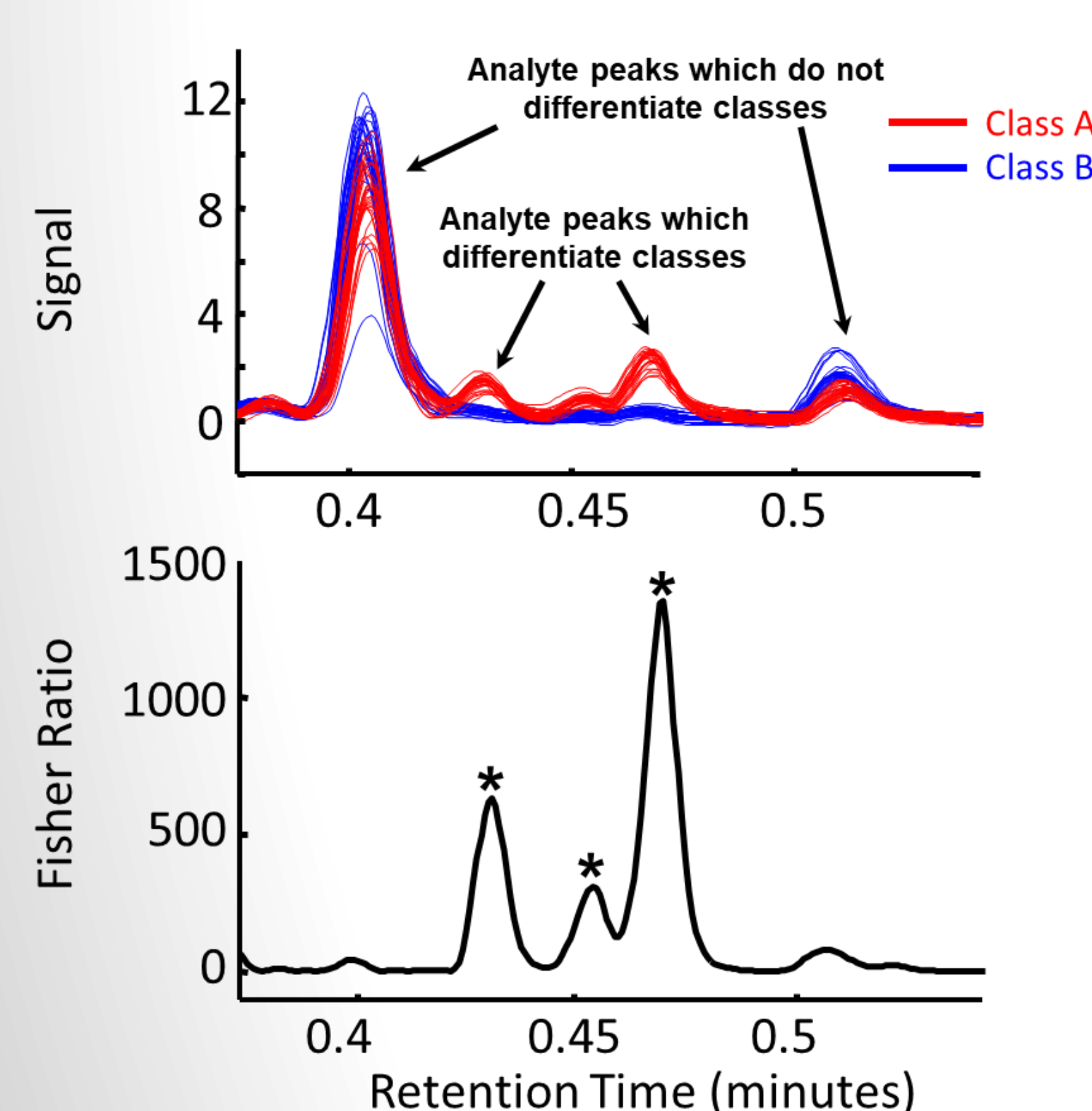


Paige E. Sudol, Grant S. Ochoa, Robert E. Synovec
Department of Chemistry, University of Washington, Seattle WA

Introduction

Tile-based Fisher ratio (F-ratio) analysis is a supervised chemometric method for the “discovery” of analytes in comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry (GC×GC-TOFMS) datasets. Herein, F-ratio analysis is used as a feature selection tool in conjunction with principal component analysis (PCA), an unsupervised chemometric method, to improve sample classification. However, the limits of this data analysis strategy to distinguish small concentration differences in the context of large sample-to-sample variation have not been explored. Thus, we examine the suitability of applying F-ratio analysis for identifying low level concentrations of sulfur-containing compounds spiked into jet fuels with large differences in native composition across all fuel samples. This research provides insight into uncovering subtle sample-to-sample differences of interest, without obstruction by a diverse, unwanted chemical background. A mixture of 14 sulfur-containing compounds of approximately equal mass was prepared. Three jet fuels (JP-1800A, JP-4, and JP-8) were spiked with the sulfur mix at nominal concentrations of 10 ppm and 20 ppm. The neat and spiked jet fuels were run in quadruplicate on the Pegasus BT 4D GC×GC-TOFMS instrument (LECO). The data was exported into MATLAB and subjected to PCA and F-ratio analysis. Two F-ratio comparison analyses were performed: 20 ppm versus neat, and 20 ppm versus 10 ppm. The top 15 “hits” for each F-ratio comparison were analyzed, identified using NIST databases, and quantified using an in-house developed signal algorithm. Signals for the 3 top F-ratio mass channels (m/z) were then analyzed via PCA.

Fisher Ratio



$$F - ratio = \frac{\text{Between Class Variance}}{\sum(\text{Within Class Variance})}$$

F-ratio method emphasizes class-distinguishing ability over absolute signal

- L.C. Marney, W.C. Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Talanta. 115 (2013) 887–895.
- B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Anal. Chem. 2015, 87 (7), 3812–3819.
- Parsons, B.A., et al. J. Chromatogr. A. 1440 (April 1, 2016): 179–90.
- N.E. Watson, B.A. Parsons, R.E. Synovec, J. Chromatogr. A. 1459 (2016) 101–111

Experimental Design

Class 1

JP-1800A Neat x4
JP-4 Neat x4
JP-8 Neat x4

Class 2

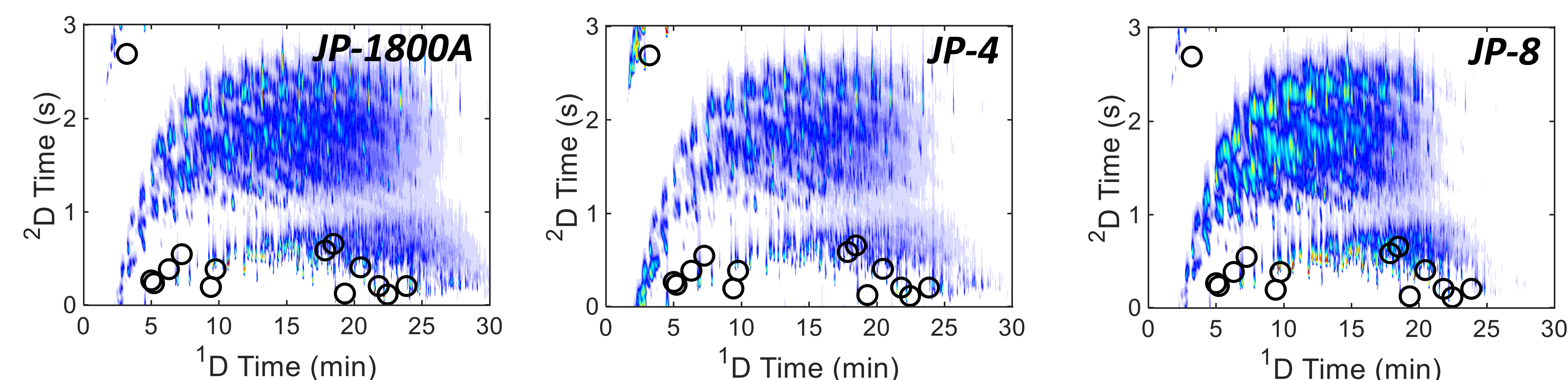
JP-1800A Spiked x4
JP-4 Spiked x4
JP-8 Spiked x4

¹D Column: Rxi-17 Sil MS, 26 m, 250 μ m i.d., 0.25 μ m d.f.
²D Column: Rxi-1 MS, 1.9 m, 180 μ m i.d., 0.18 μ m d.f.
Temp Program: 40°C, hold 1.5 min, 5°C/min to 200°C, hold 1 min
Flow Rate: 2 ml/min
Mass Channels Collected: m/z 45–334

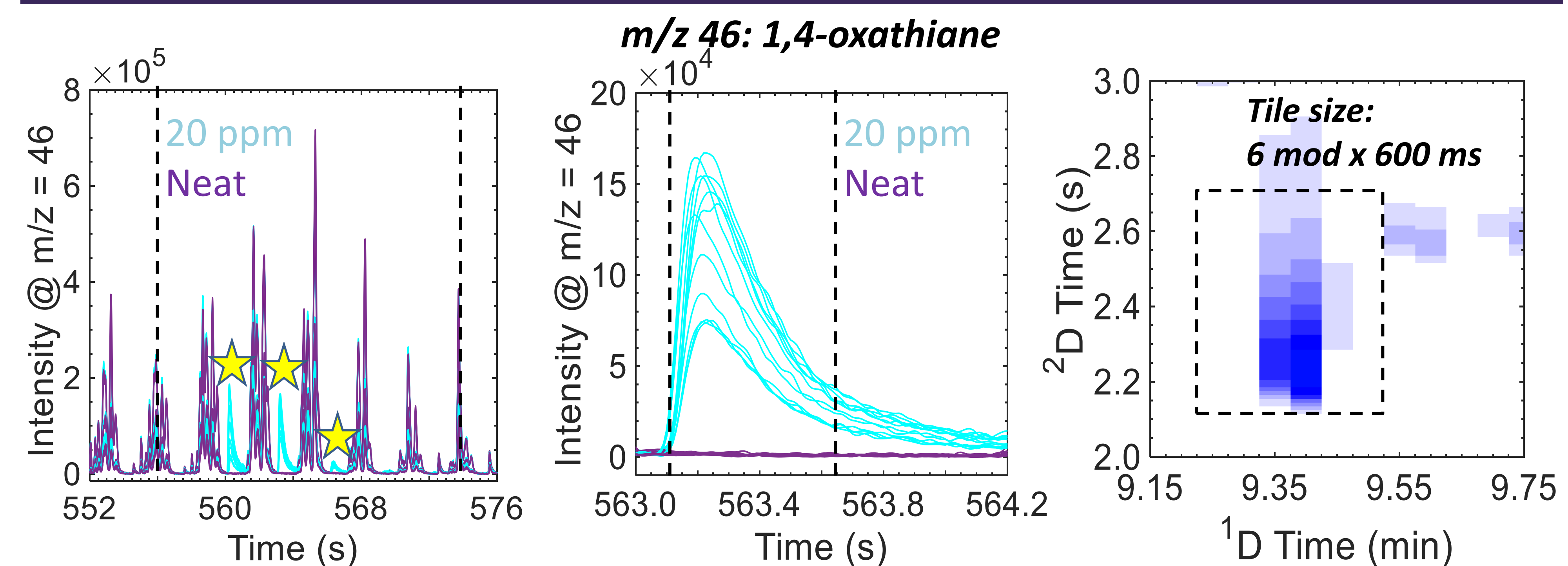


Above: LECO Pegasus BT 4D GC×GC-TOFMS used to collect data for the neat and spiked jet fuels.

GC×GC Chromatograms of Jet Fuels



Selection of tile size for Fisher ratio analysis

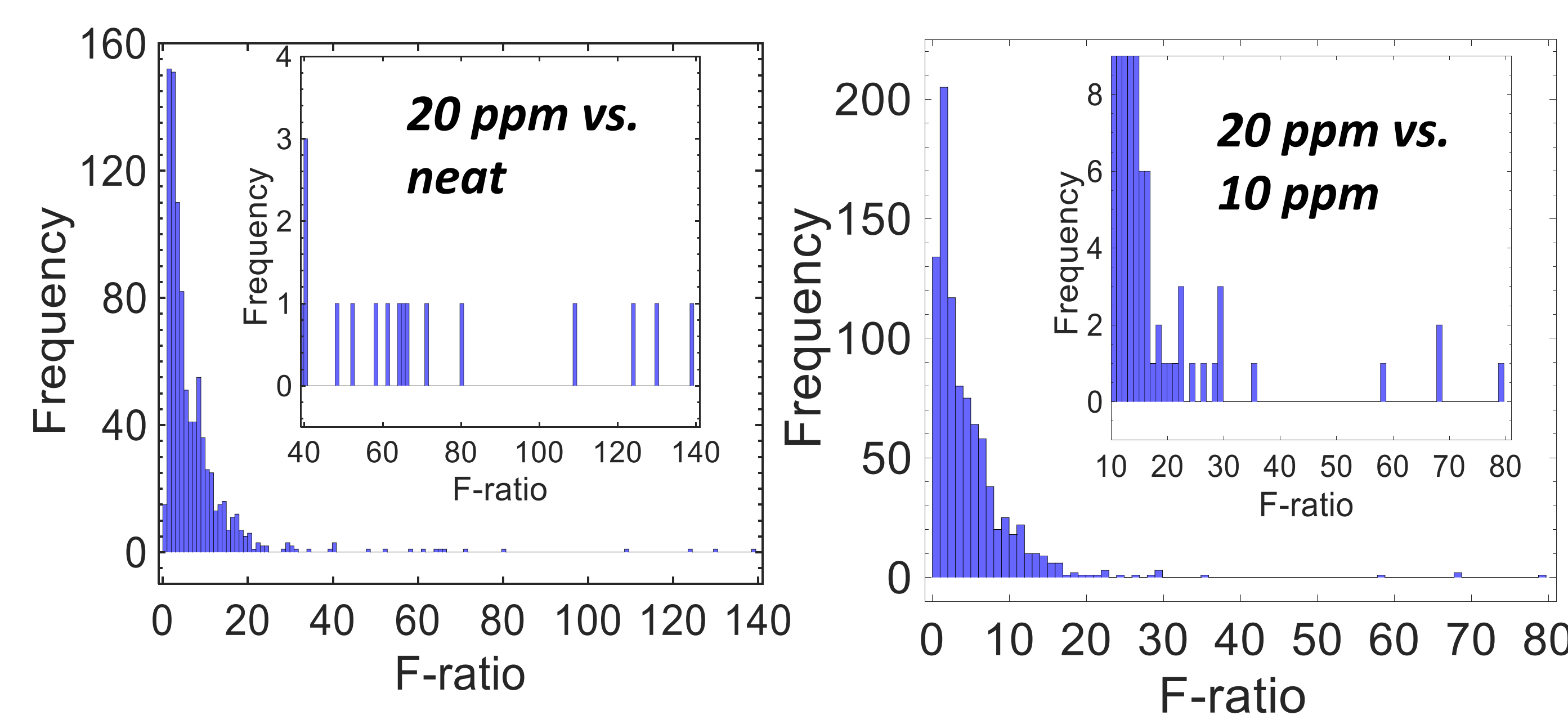


Left: Selection of ¹D tile size based on a selective analyte, 1,4-oxathiane. An overlay of the unfolded GC×GC chromatograms at m/z 46 for all 24 samples (12 spiked and 12 neat) is provided, with the modulations for 1,4-oxathiane in the spiked samples labeled with yellow stars. A ¹D tile size of 6 modulations (black dashed lines) covers about twice the ¹ W_b for 1,4-oxathiane (3 modulations), which is ideal for tile-based F-ratio analysis to account for retention time shifting without capturing multiple analytes of interest. Note that one modulation is equal to 3 s which is the separation run time on ²D.

Center: Selection of ²D tile size based on a selective analyte, 1,4-oxathiane. An overlay of a single 1,4-oxathiane modulation for all 24 sample chromatograms (12 spiked and 12 neat) is provided. Minimal retention time shifting on ²D can be observed. The average ² W_b for this peak is 760 ms, which is largely encompassed by a 2D tile size of 600 ms (black dashed lines). The other sulfur-containing analytes have narrower ² W_b (not pictured here), so a 2D tile size of 600 ms represents a good compromise to ensure individual analytes are captured within a single tile.

Right: Folded GC×GC chromatogram at m/z 46 for average spiked samples, as determined by tile-based F-ratio software. The ²D peak for 1,4-oxathiane is enclosed by dashed black lines representing the optimum tile size chosen (6 modulations \times 600 ms).

Fisher ratio distributions



Left: F-ratio distribution for the 20 ppm versus neat comparison. The top hits (highest F-ratios) are highlighted in the inset. The 14 spiked sulfur-containing compounds are easily found in the top 15 hits for this comparison.

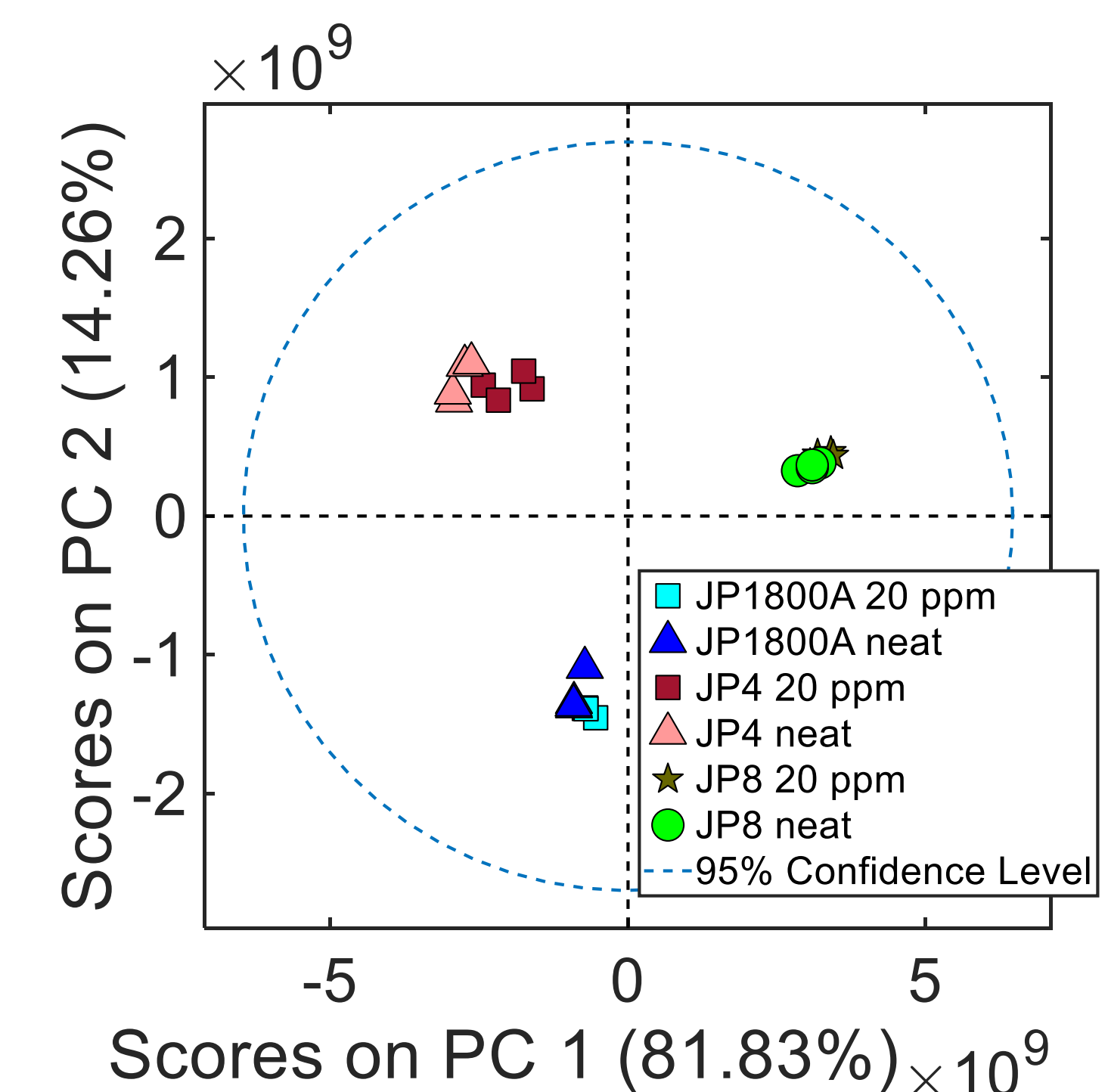
Right: F-ratio distribution for the 20 ppm versus 10 ppm comparison. The F-ratios are noticeably smaller here due to the smaller concentration difference between sample classes. Only 10 of the 14 spiked sulfur-containing compounds can be identified in the top 15 hits. As shown in the inset, the four remaining sulfur-containing compounds are obscured by false positives at lower F-ratios.

12 vs. 12 Fisher ratio results

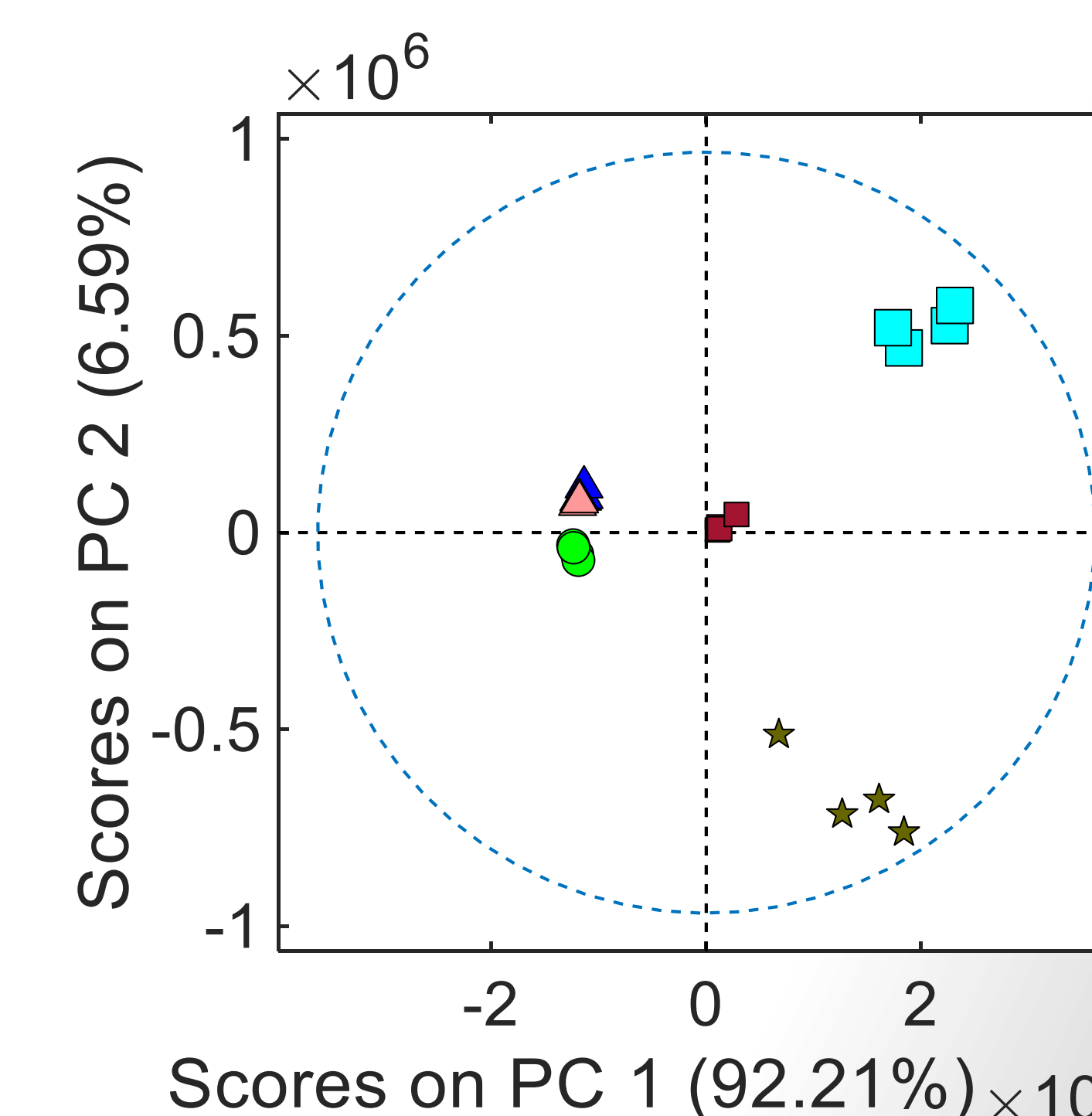
Identity	¹ t_R (min)	² t_R (s)	Hit no. 20 ppm v neat	F-ratio, 20 ppm v neat	Hit no. 20 ppm v 10 ppm	F-ratio, 20 ppm v 10 ppm
Benzothiophene	19.30	0.13	1	139.3	3	68.4
3-methylbenzothiophene	22.45	0.12	2	130.4	1	79.1
2-methylbenzothiophene	21.80	0.21	3	123.7	2	68.6
2-chloroethylphenylsulfide	23.85	0.21	4	109.2	4	58.1
2-hexylthiophene	18.45	0.66	5	79.7	6	29.8
3-acetyl-2,5-dimethylthiophene	20.45	0.41	6	71.6	5	35.2
2-butyl-5-ethylthiophene	17.85	0.59	7	66.3	11	23.9
2-propylthiophene	9.75	0.39	8	64.9	25	16.3
2,5-dimethylthiophene	7.25	0.55	9	64.0	13	22.6
Thiophene	3.2	2.69	10	61.1	21	16.8
Tetrahydrothiophene	6.30	0.39	11	58.3	9	28.0
1,4-oxathiane	9.40	0.20	12	52.2	8	29.4
2-methylthiophene	5.00	0.27	14	40.6	56	12.6
3-methylthiophene	5.20	0.24	15	40.5	51	13.0

PCA: 20 ppm vs. Neat Comparison

Before Feature Selection



After Feature Selection



Conclusion

Prior to feature selection via F-ratio analysis, the spiked and neat samples could not be distinguished as individual classes in the PCA scores plot, as shown for the 20 ppm versus neat comparison above. Tile-based F-ratio analysis of the 20 ppm versus neat class comparison identified all 14 sulfur-containing compounds in the top 15 hits of the hitlist. When these signals were input into PCA, the neat and spiked samples could be clearly distinguished as individual classes along the PC1 axis, as shown above. The spiked samples also separated according to fuel type along the PC2 axis, which can likely be attributed to small differences in added spike concentration. These results highlight the impressive sensitivity of the LECO Pegasus BT 4D GC×GC-TOFMS system. Tile-based F-ratio analysis of the 20 ppm versus 10 ppm class comparison only identified 10 of the 14 sulfur-containing compounds in the top 15 hits. Therefore, PCA of the signals for these top 15 hits showed a less distinct separation between neat and spiked samples, and so this scores plot was excluded for brevity. Nevertheless, the results presented herein demonstrate that tile-based F-ratio analysis performs exceptionally well for identification of low-level contaminants in the presence of a largely varying chemical background.