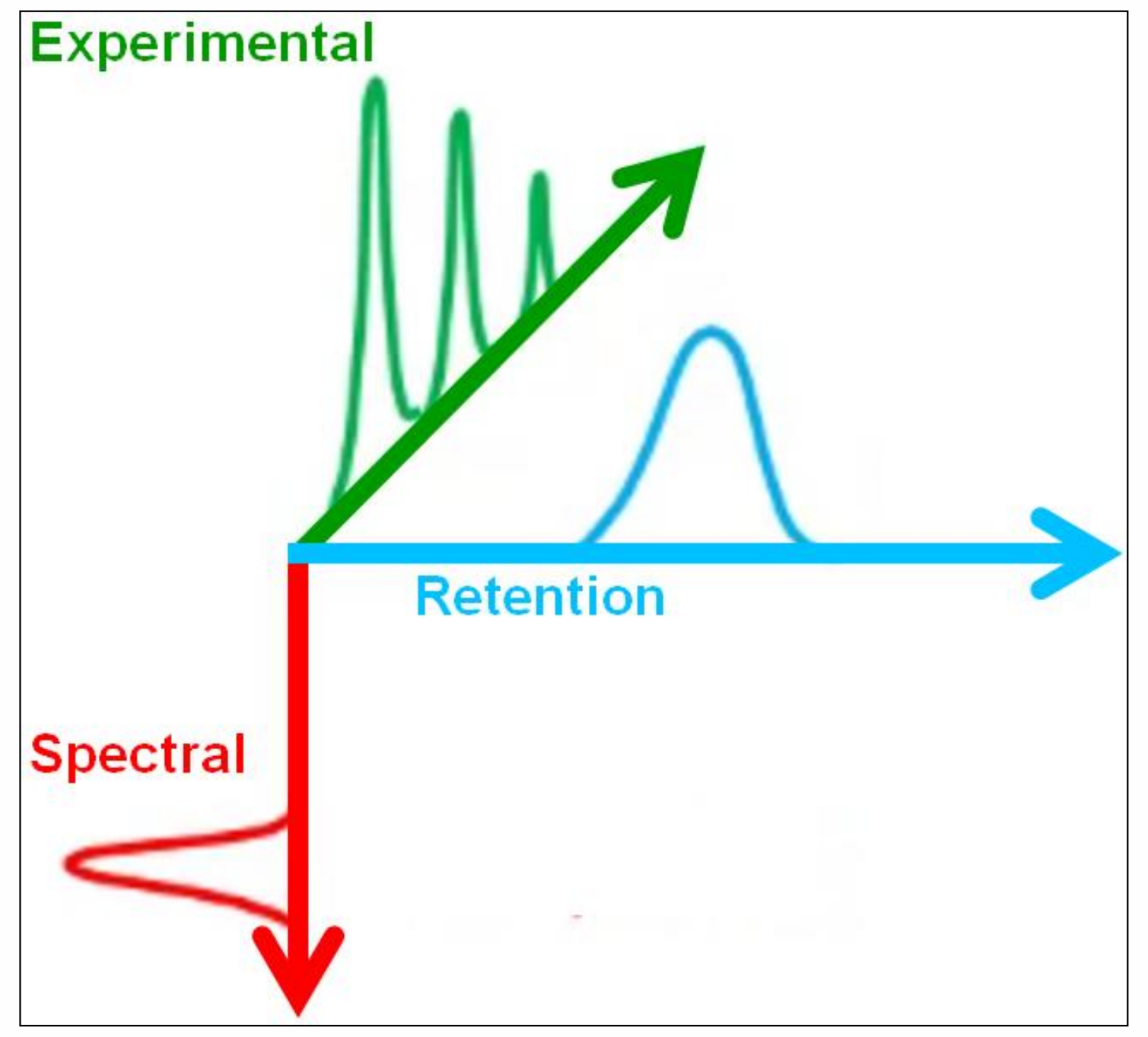
Accurate Deconvolution of Perfectly Coeluting Analytes by Exploiting Differential Expression Across Samples

Kevin Siek; David La Fleur; Jihong Wang; Peter Willis | LECO Corporation, Saint Joseph, Michigan

Introduction



A set of chromatography-mass spectrometry data for multiple related samples may be considered as a tensor with spectral, retention, and experimental dimensions. For example, a series of consecutive time-of-flight spectra comprise an ion chromatogram. A series of consecutive ion chromatograms comprise an experiment. Data from higher-performance techniques (e.g. GCxGC-MS or GC-MS-MS) could introduce additional higher dimensions.

If samples in the experiment represent discrete timepoints in a chemical process, distinct biological classes, or other such experimental designs, chemical feature amounts are expected to exhibit different profiles across the sample set. Significant differential expression of features across a sample set should permit multi-way factor analysis (including multilinear and unfolded approaches) to accurately report pure spectra and concentration profiles, even if some features coelute perfectly in one or more samples.

Algorithm Description

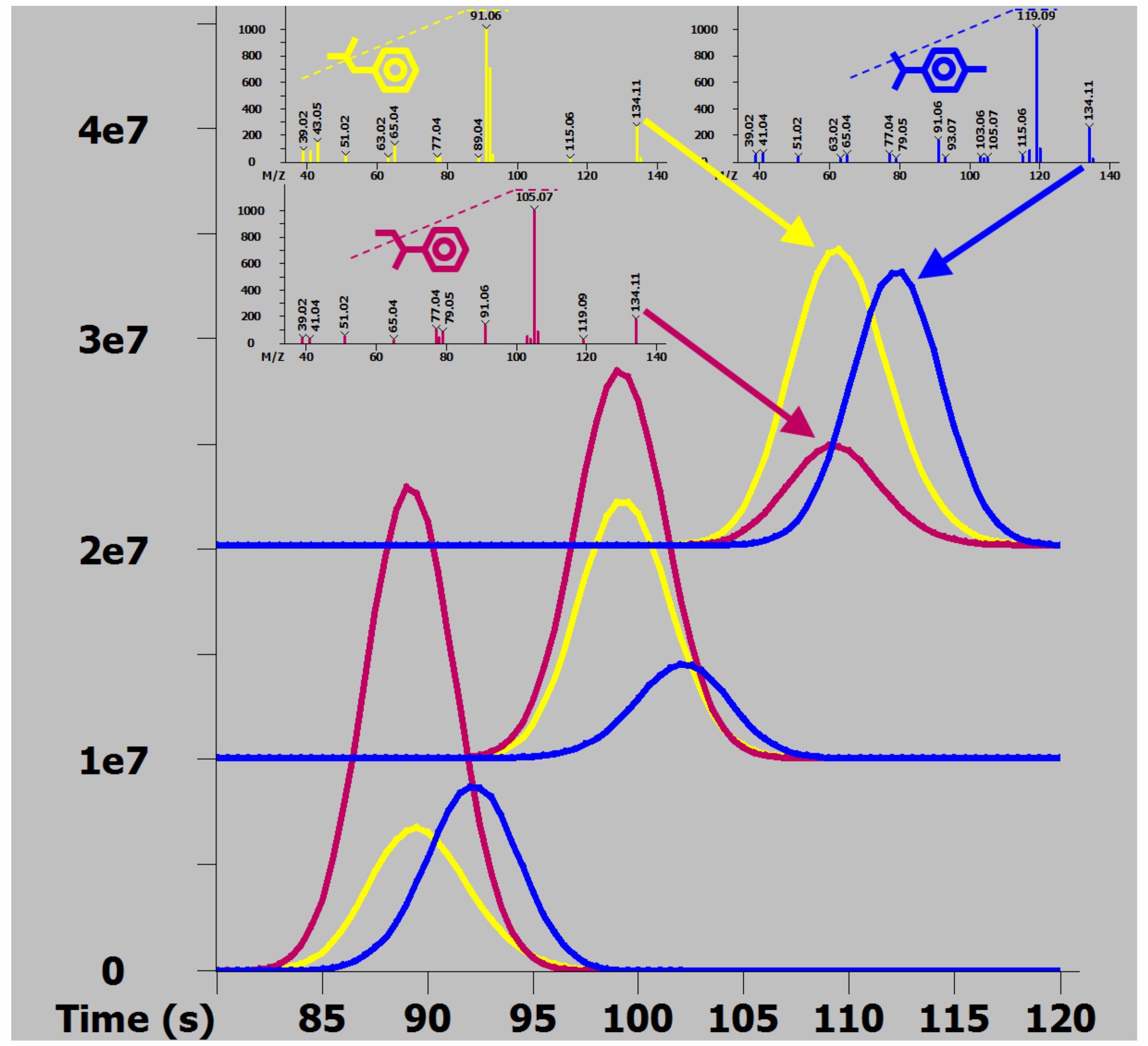
The algorithm used to generate results reported herein uses factor analysis to decompose the sample set tensor into constituent factors and loadings that are rotated to yield chemically meaningful spectra and concentration profiles. Since spectra are expected to remain consistent for each chemical feature, but minor chromatographic peak shifting cannot be precluded, the algorithm operates with spectra folded into chromatograms, but with chromatograms unfolded across the experiment, to not impose trilinearity where it should not be expected.

The algorithm we developed is similar to a recently described PARAFAC2-based approach[1] which requires two user-specified parameters: retention interval and number of factors expected per interval. The algorithm we developed converts data to absolute number of ions and takes the square root. Features are then removed until the residual can be fully accounted for by Poisson ion statistics, down to ≈3σ level. This approach easily and accurately predicts the correct number of factors to extract, independent of user input. Thus, one of two parameters required by the previously reported approach is eliminated.

The algorithm was composed and tested in MATLAB R2014b 8.4.0.150421.

Algorithm Application to Synthetic Data

Compendial (NIST 2014 mainlib) or real spectra acquired in-house of small hydrocarbon molecules and derivatized metabolites were simplified by truncating the upper m/z to 300 and eliminating spectral signals <2.5% abundance. Giving these spectra Gaussian concentration profiles made simulated chromatograms. Partial and perfect coelutions were created in the synthetic data set, according to semi-standard non-polar retention indexes listed in NIST 2014. The example below shows a perfect coelution of sec-butylbenzene and isobutylbenzene partially coeluting with p-cymene.



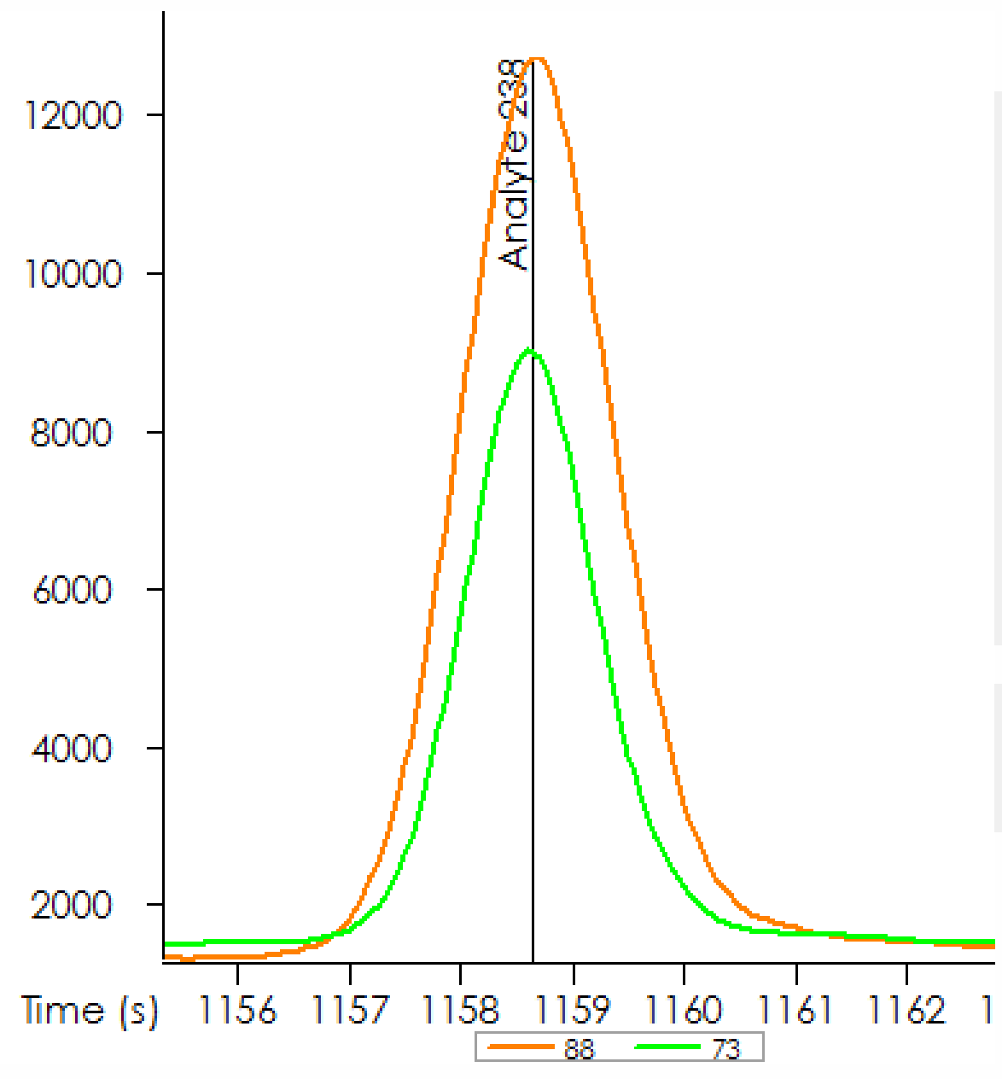
A simulated dead coelution of trimethylsilyl-2-furoate + bis(trimethylsilyl) oxalate was accurately reported as two independent components when covariance of these analytes across a set of 15 samples was r <0.9. The correct hit for each analyte was ranked #1 from a NIST 2014 library search. Forward similarity scores vs. mainlib and replib ranged from 930 to 946 for the oxalic acid 2TMS derivative and from 769 to 949 for the 2-furoic acid TMS derivative. Dead coelutions where analyte covariance exceeded r = 0.9 across the sample set were reported as single components with chimera spectra.

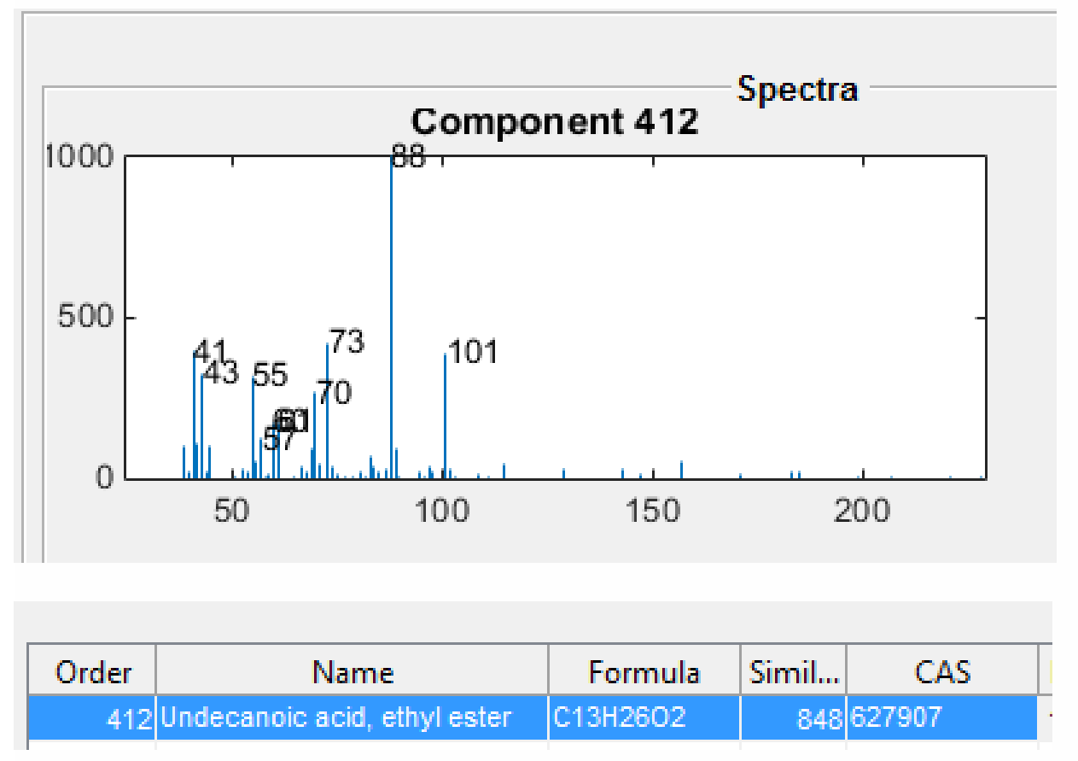
Delivering the Right Results

Algorithm Application to Real Data

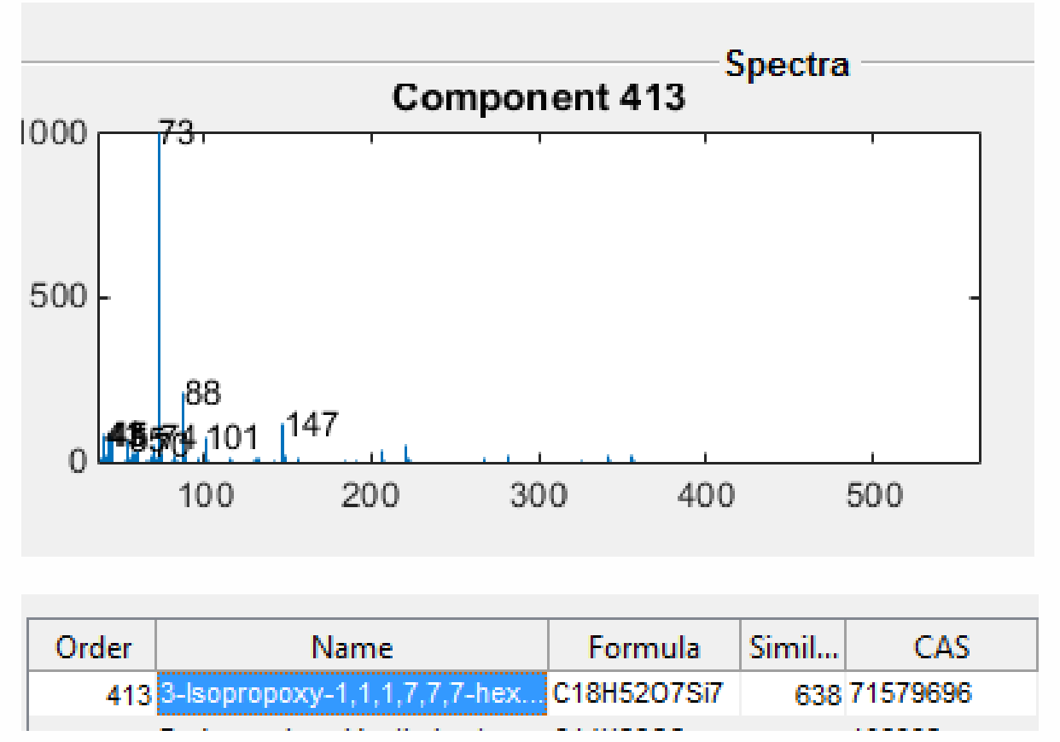
The algorithm was applied to real GC-TOFMS data previously reported[2]. Accurately deconvolved close to perfect coelutions are shown below.

Ethyl dodecanoate and siloxane.



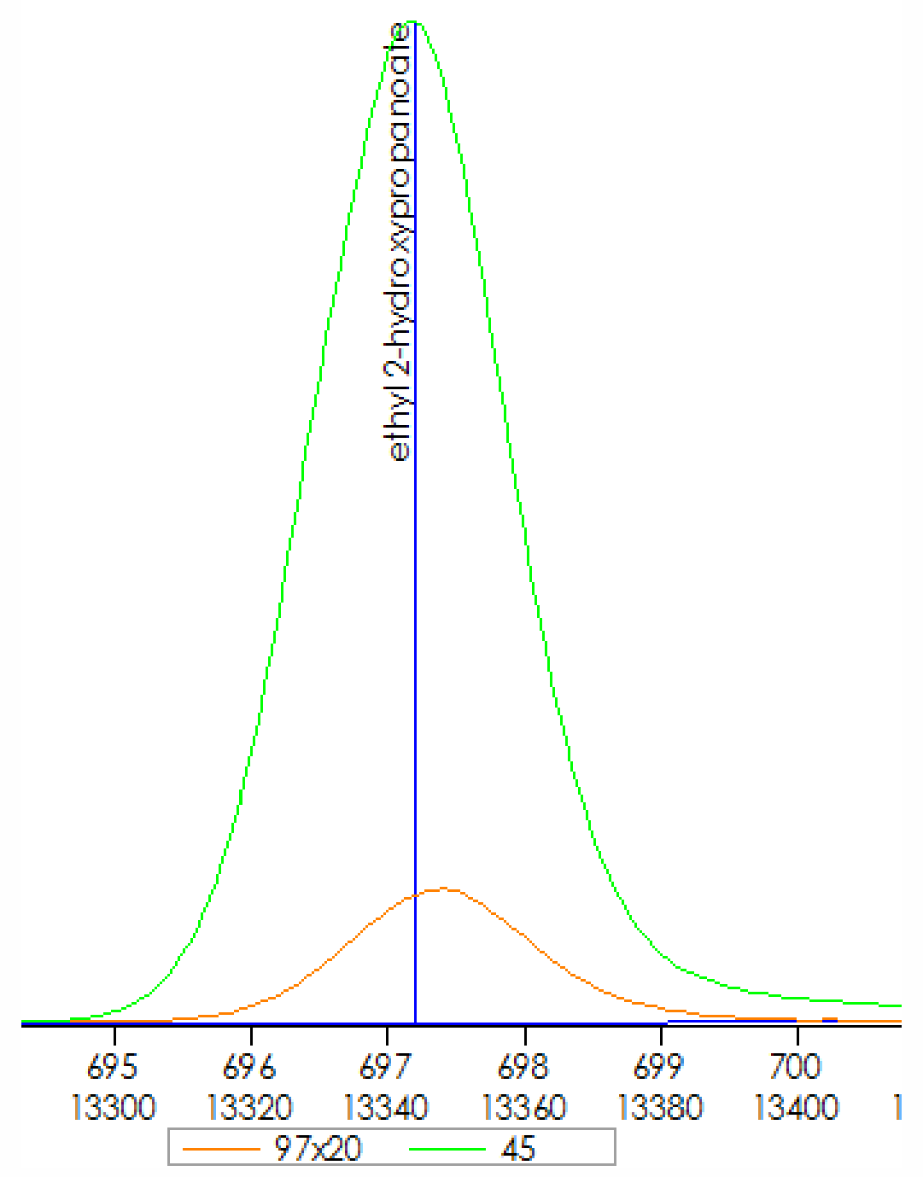


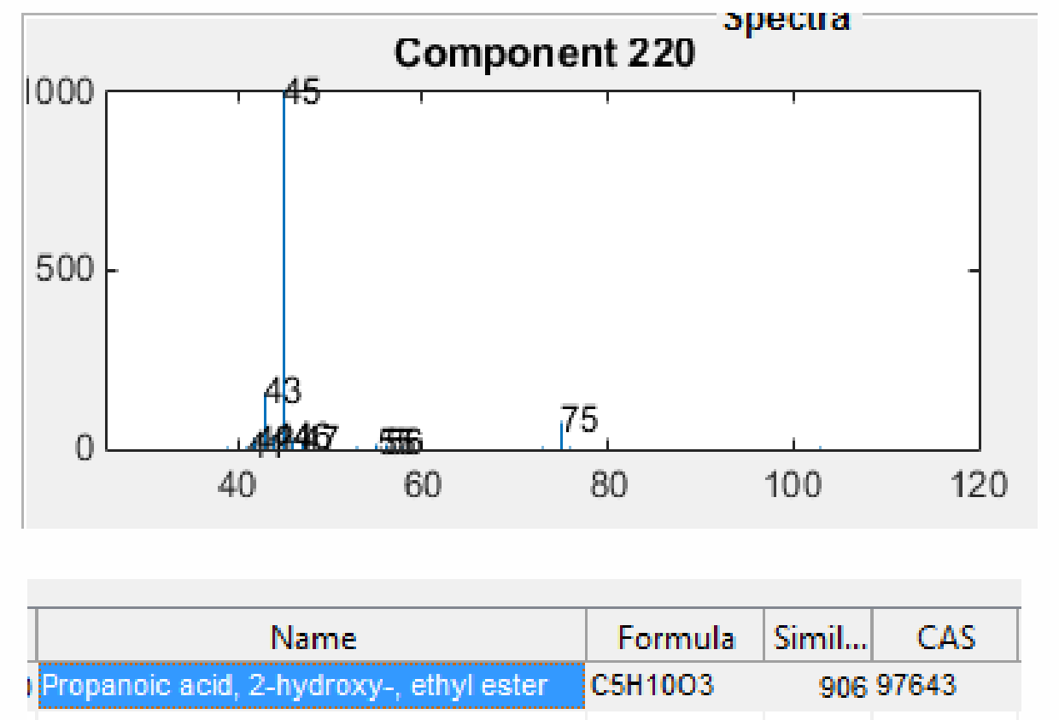
Order	Name	Formula	Simil...	CAS
412	Undecanoic acid, ethyl ester	C13H26O2	848	627907



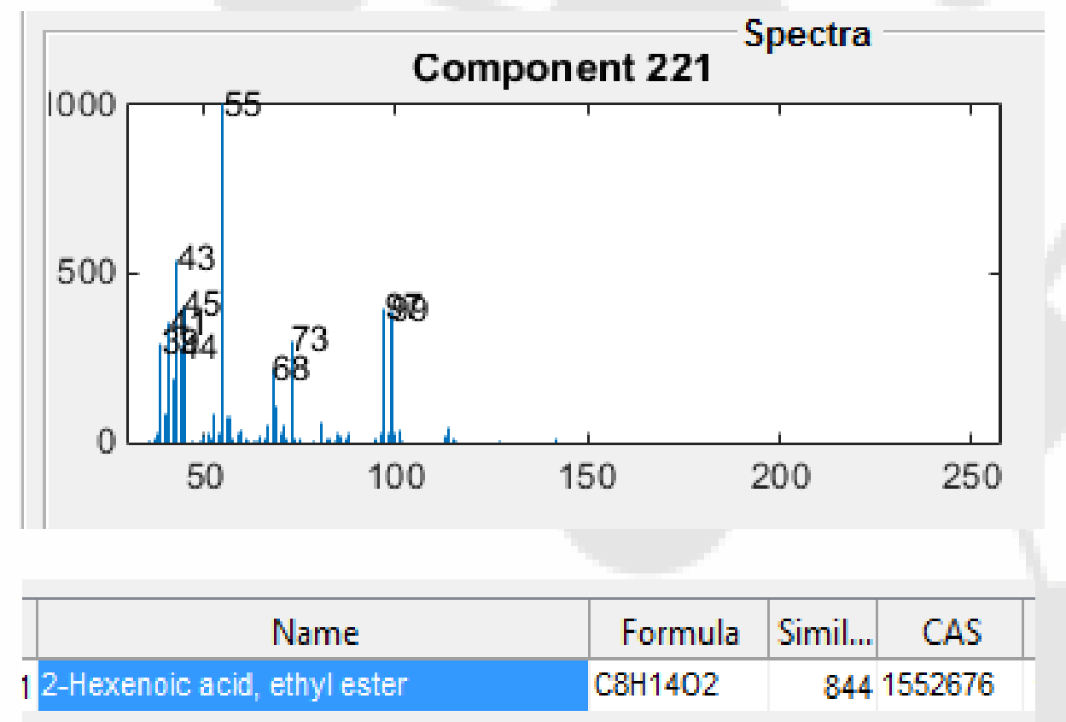
Order	Name	Formula	Simil...	CAS
413	3-Isopropoxy-1,1,1,7,7,7-hex...	C18H52O7Si7	638	71579696

Ethyl lactate (major) and ethyl 2-hexenoate (minor) NOTE: XIC drawn for minor component is multiplied 20X for visibility on the plot.





Name	Formula	Simil...	CAS
Propanoic acid, 2-hydroxy-, ethyl ester	C5H10O3	906	97643



Name	Formula	Simil...	CAS
2-Hexenoic acid, ethyl ester	C8H14O2	844	1552676

Future work includes comparative evaluation using real and synthetic data sets.

References

[1] L.G. Johnsen et al. J. Chromatography A 1503 (2017) 57-64.
[2] E. M. Humston-Fulmer et al. Metabolomics 2015 Proceedings; Poster 344.