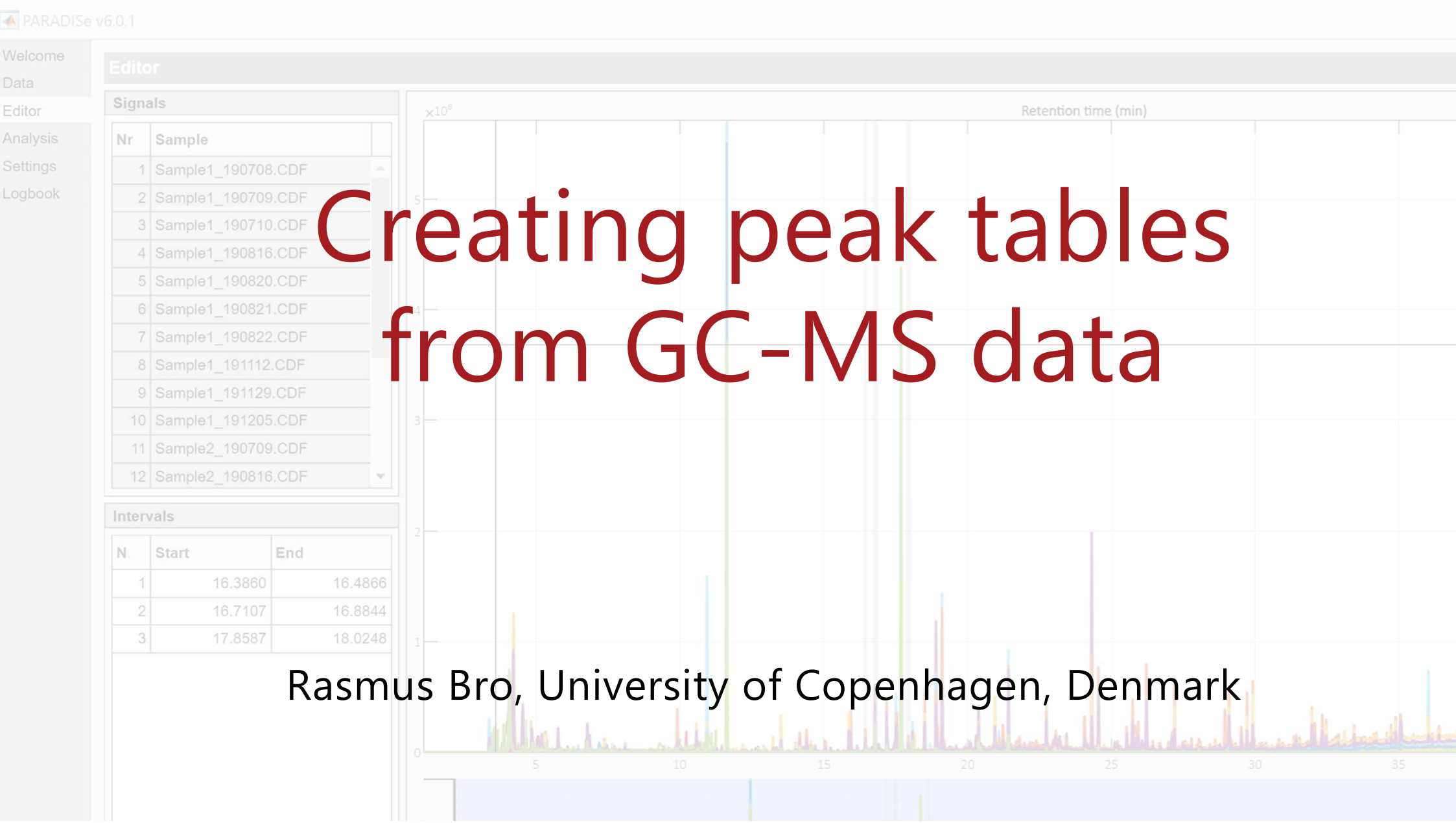
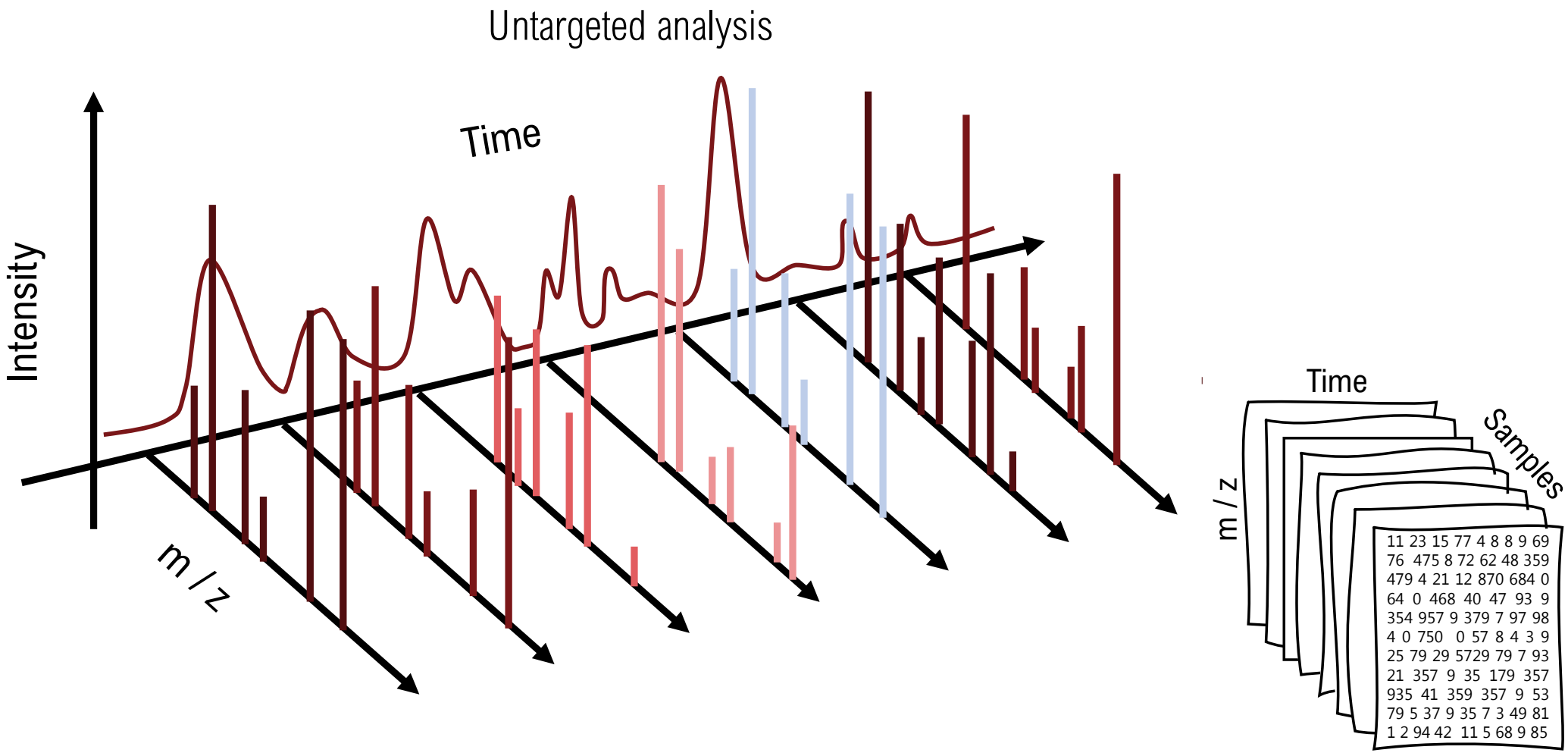
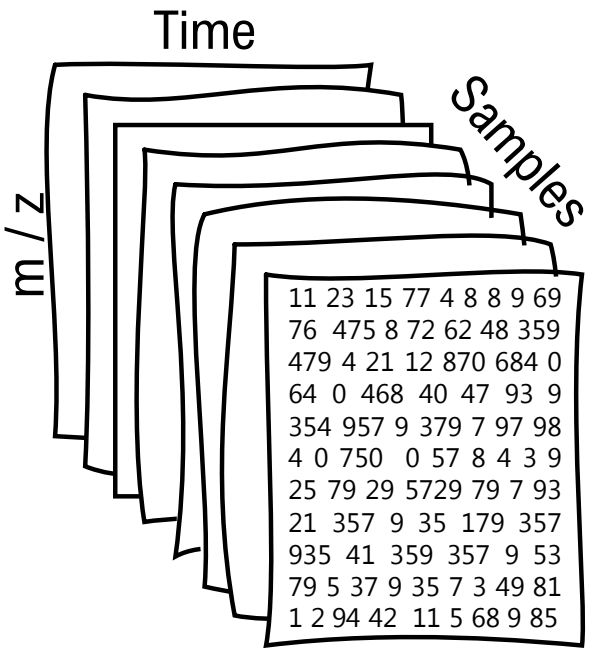


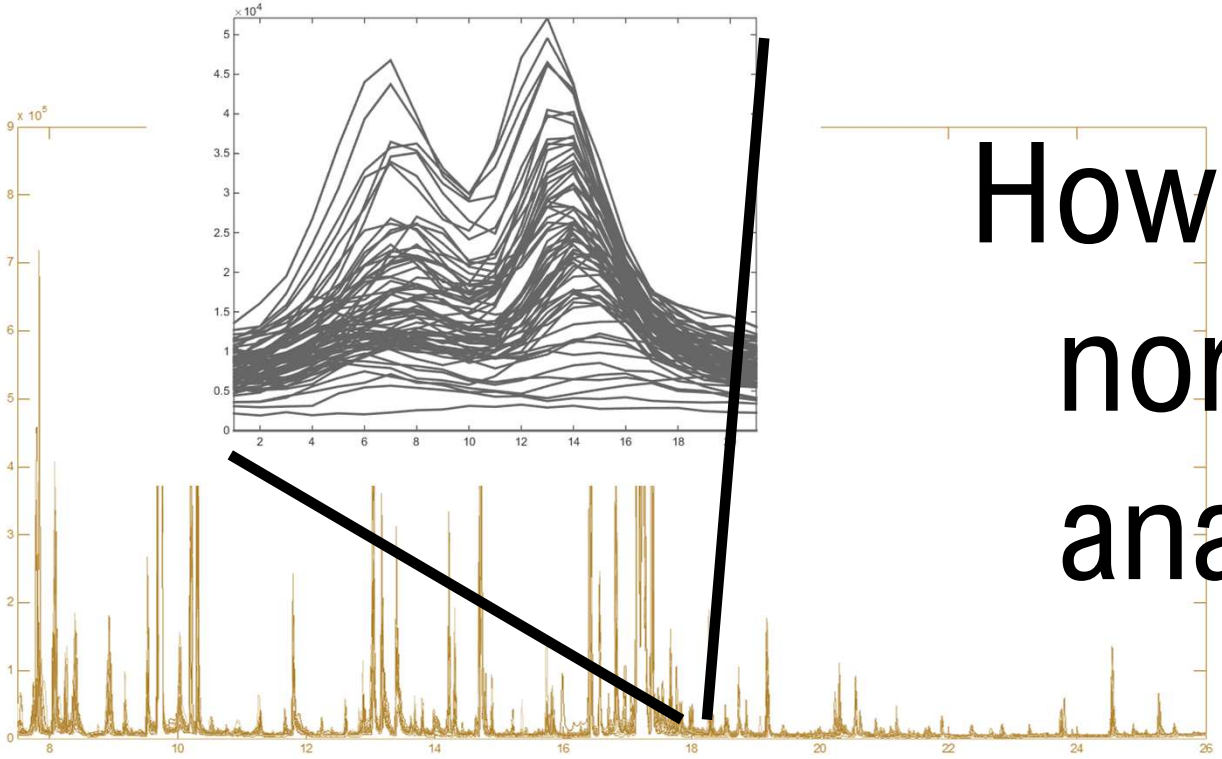
Creating peak tables from GC-MS data

Rasmus Bro, University of Copenhagen, Denmark


Untargeted analysis
Time
Intensity
m / z
Time
Samples
m / z
11 23 15 77 4 8 8 9 69
76 475 8 72 62 48 359
479 4 21 12 870 684 0
64 0 468 40 47 93 9
354 957 9 379 7 97 98
4 0 750 0 57 8 4 3 9
25 79 29 5729 79 7 93
21 357 9 35 179 357
935 41 359 357 9 53
79 5 37 9 35 7 3 49 81
1 2 94 42 11 5 68 9 85



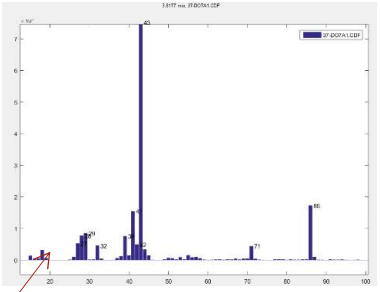
[Comp.] / Samples	Pyridine, 2-ethenyl-	Oxazole	Thiourea, N,N'-dimethyl-	1,3-Butanediamine	Thiourea, N,N'-dimethyl-
ARG-BNS1	14537.68	15004.91	724.1559	55245.34	2339.035
ARG-DDA1	10928.24	18660.14	4266.916	69260.46	7127.093
ARG-FFL1	17169.24	23240.5	790.8323	74449.28	2014.114
ARG-FLM1	11476.84	20224	716.1224	87980.61	2088.36
ARG-ICR1	16027.01	21736.32	18744.36	78769.53	26130.93
ARG-SAL1	7377.091	24200.11	11756.38	81672.69	16879.96
AUS-CAV1	12412.66	14577.94	1695.526	75978.83	3884.537
AUS-EAG1	15684.45	24415.18	0	116384.9	208.576
AUS-HAR1	7010.431	20726.64	0	96737.83	1861.137
AUS-IB41	12200.63	21896.72	2312.168	97213	4686.197
AUS-KIL1	565.6317	10579.45	18156.02	70915.36	25742.15
AUS-KIR1	2929.144	15214.05	32191.06	92669	45012.6
AUS-NUG1	6255.24	30922.8	19791.65	115512	27742.7
AUS-SOC1	14024.34	18639.6	31046.44	84807.11	41654.19
AUS-TGH1	9288.885	20323.24	0	69295	1072.47
AUS-VAF1	5553.512	17021.14	26073.45	85262.24	36196.02

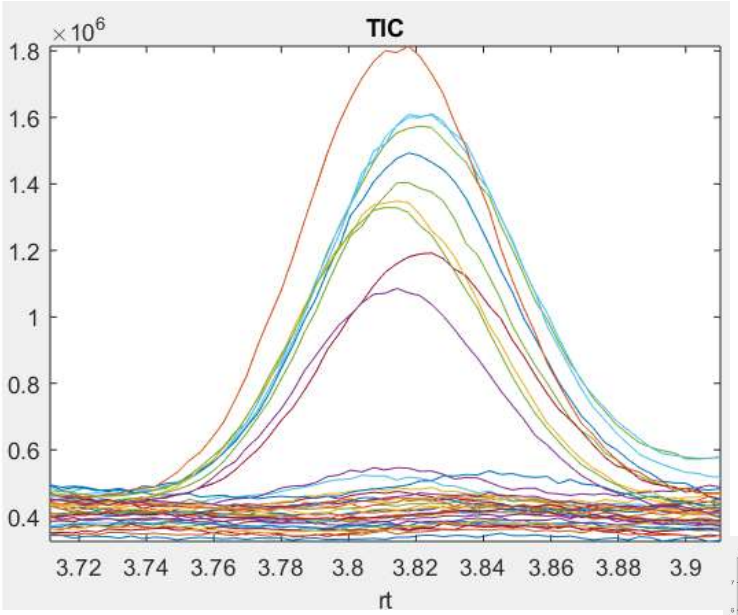


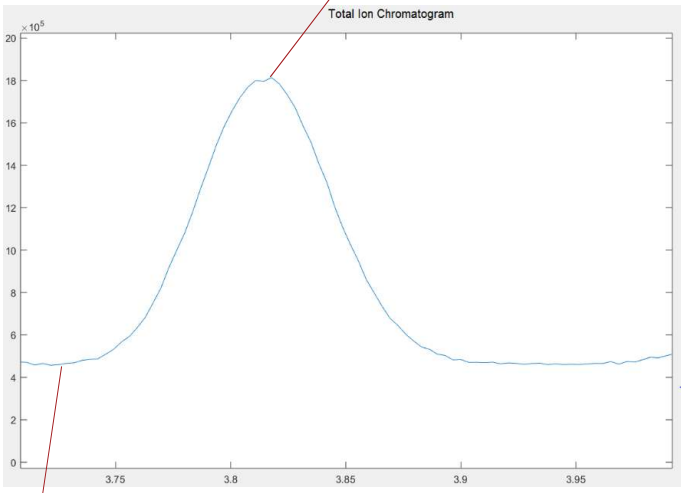
How data is normally analyzed

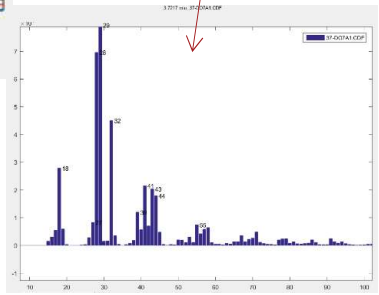
Amdis, Chromatof, Chemstation, Masshunter, XCMS, MZmine

Identification







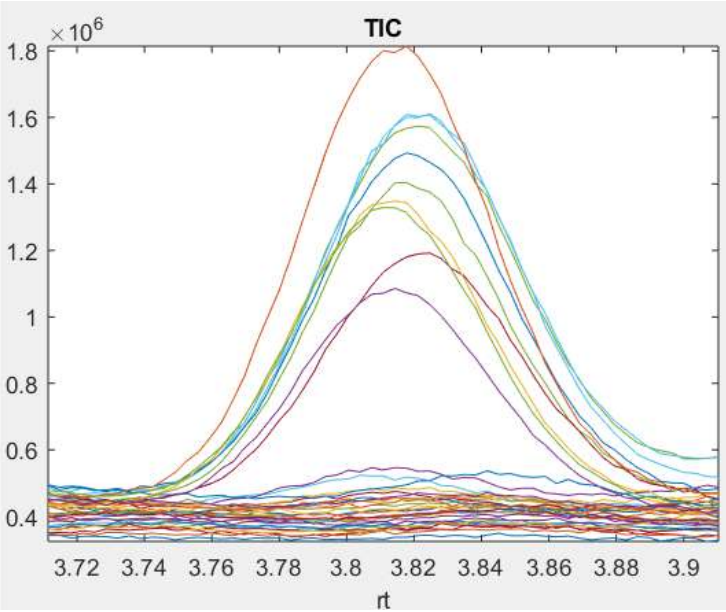


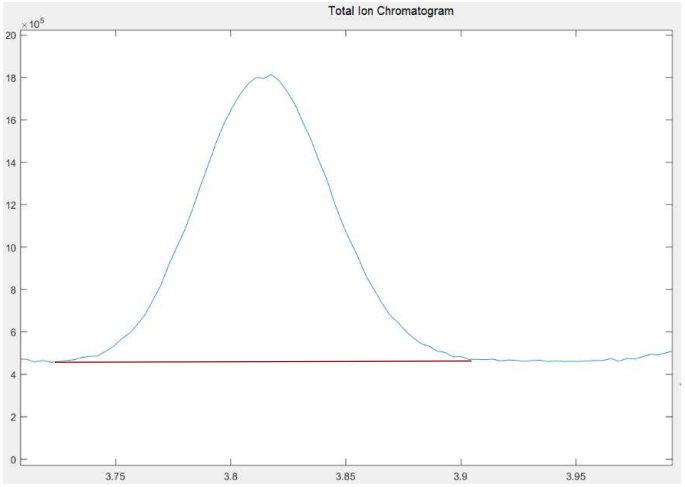
Subtract mass spectra and search:

3-Methyl-2-butanone

The ‘traditional approach’

Quantification

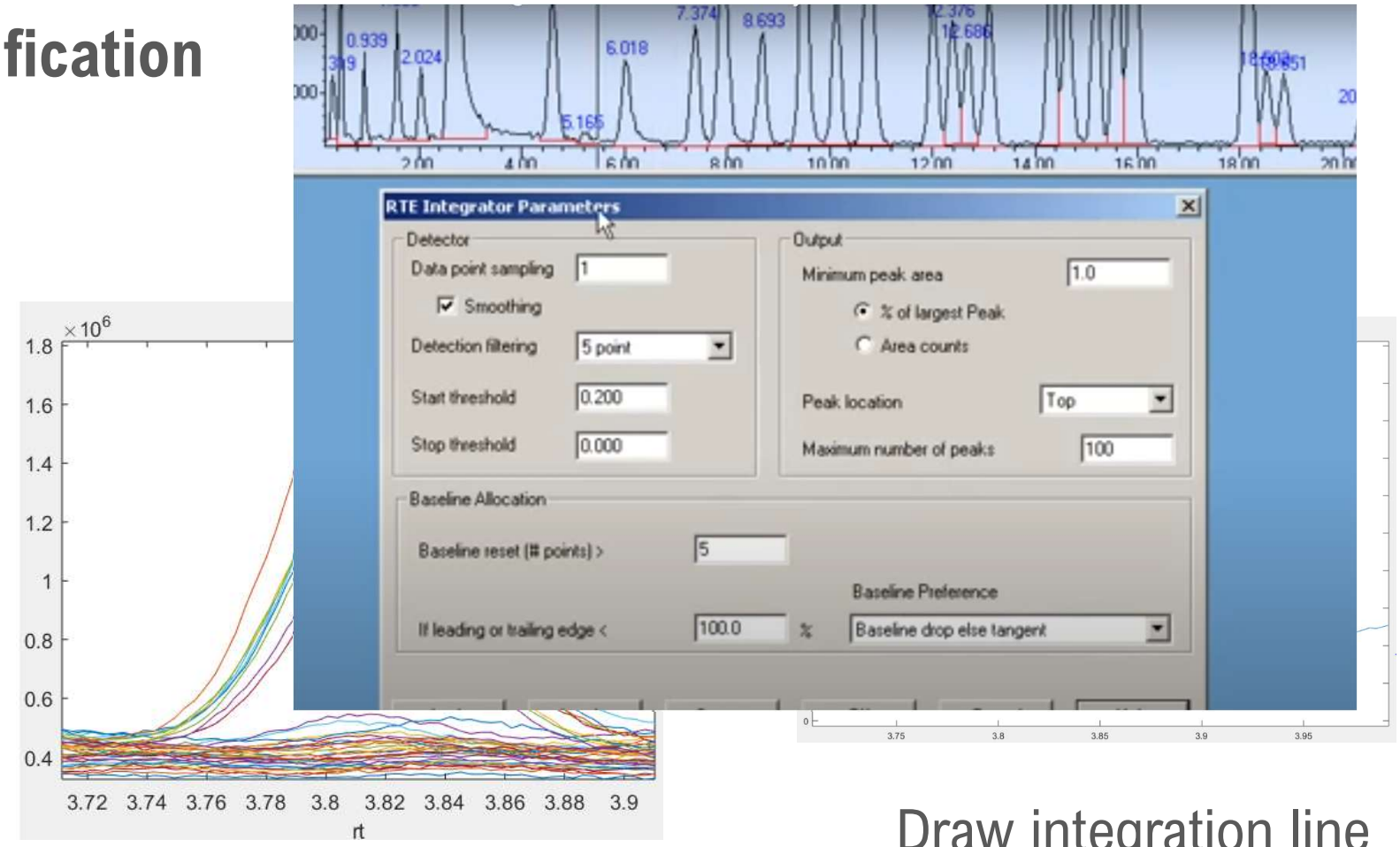




Draw integration line
Calculate peak area

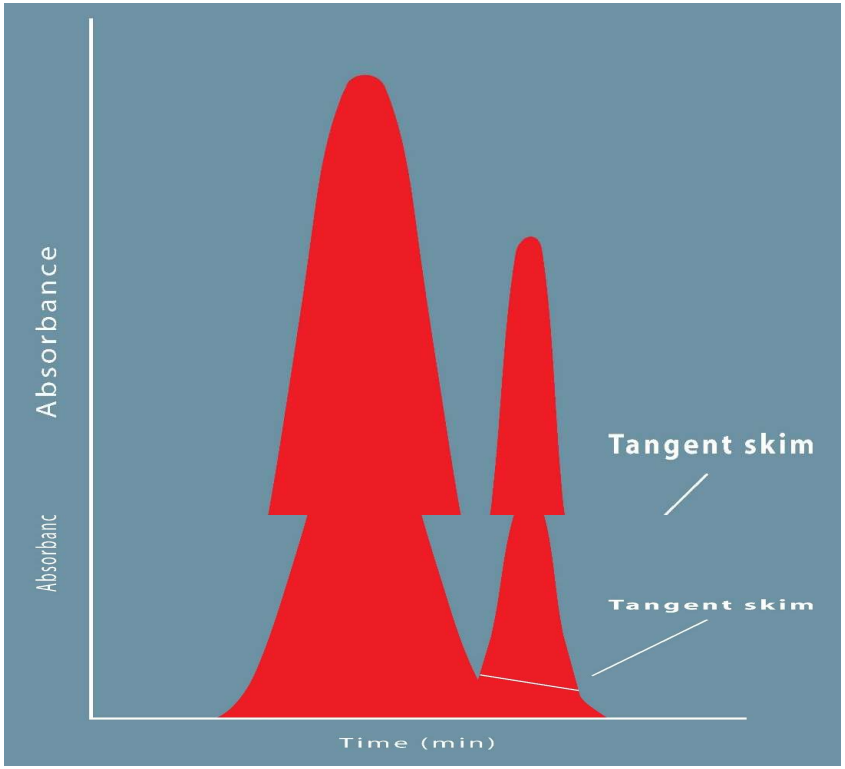
The ‘traditional approach’

Quantification

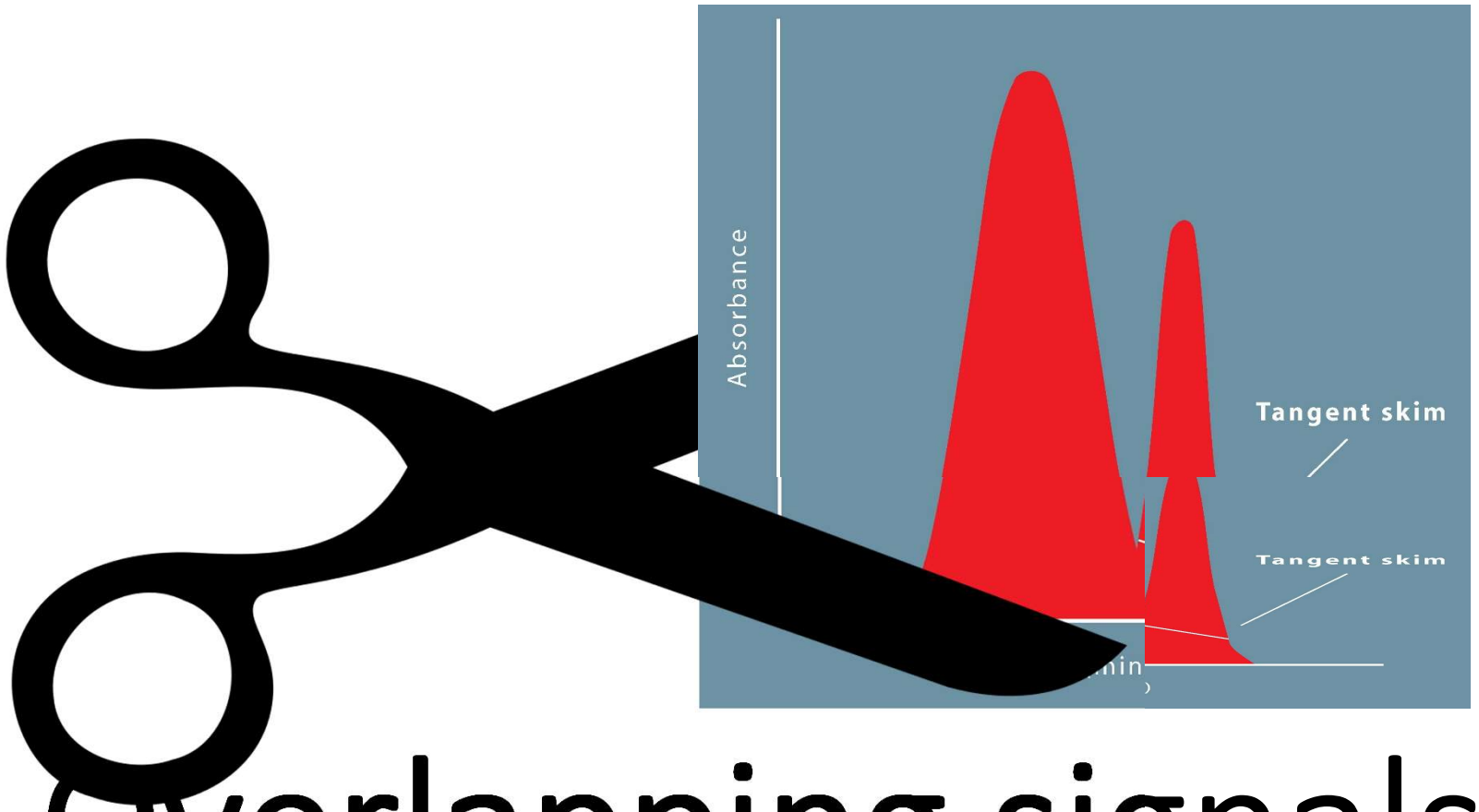


Draw integration line
Calculate peak area

The ‘traditional approach’

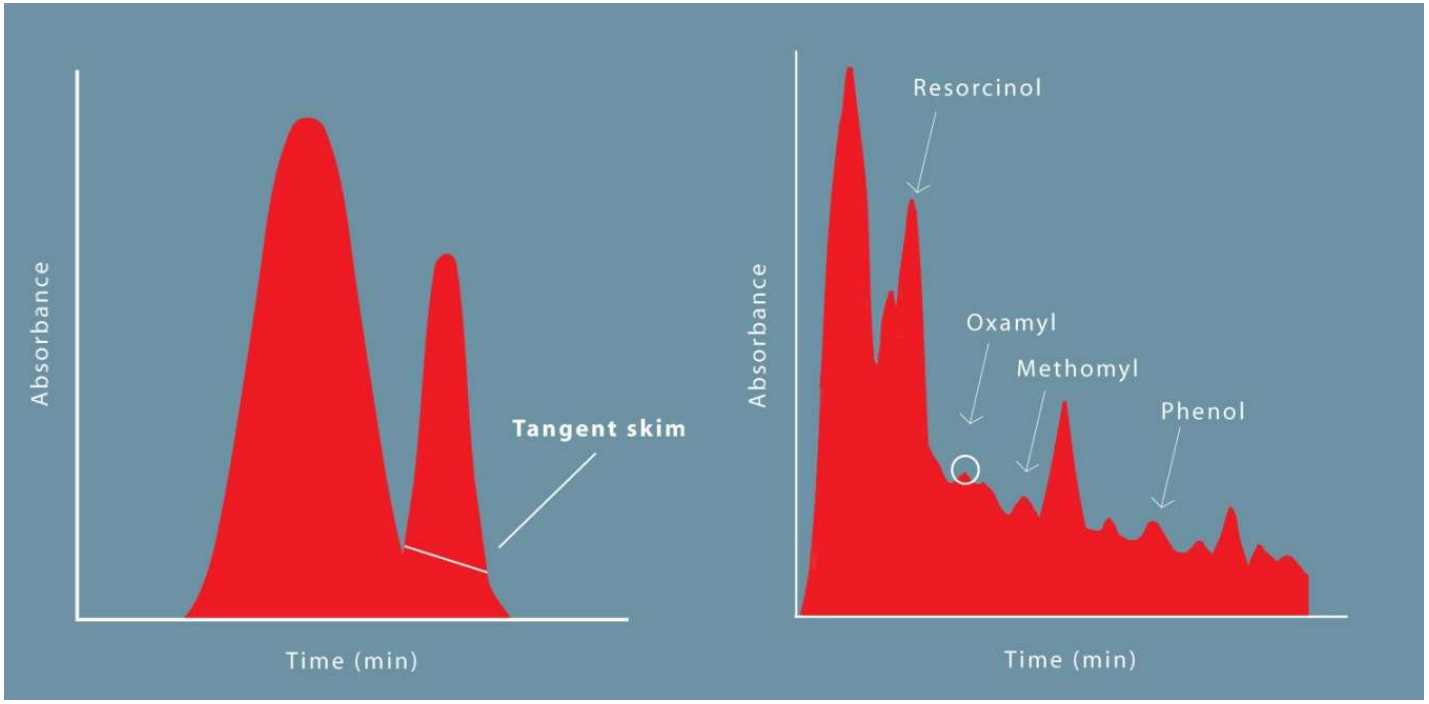


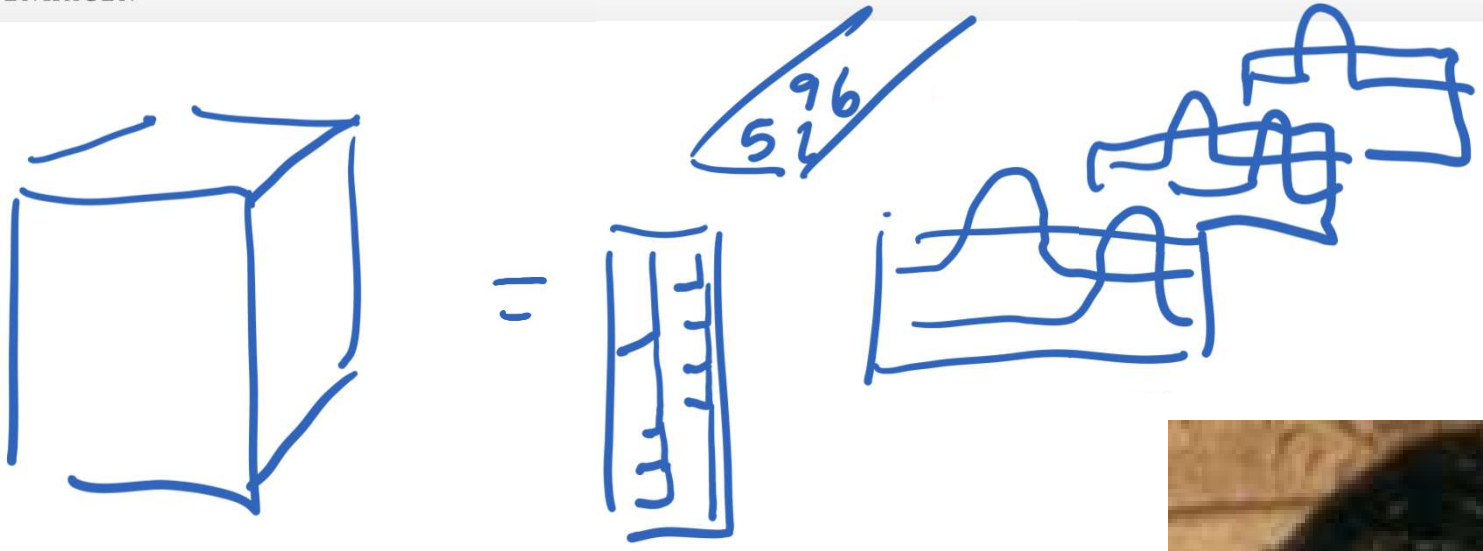
Overlapping signals



Overlapping signals

But what if the peaks are a little more complex?

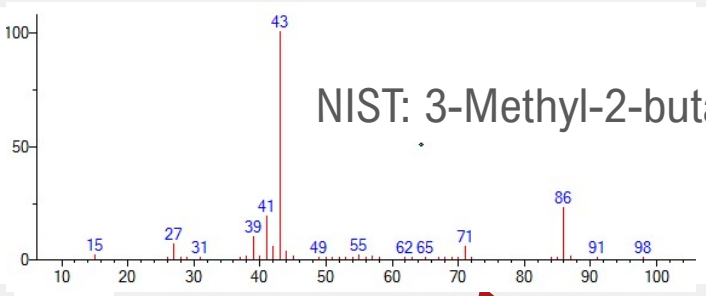




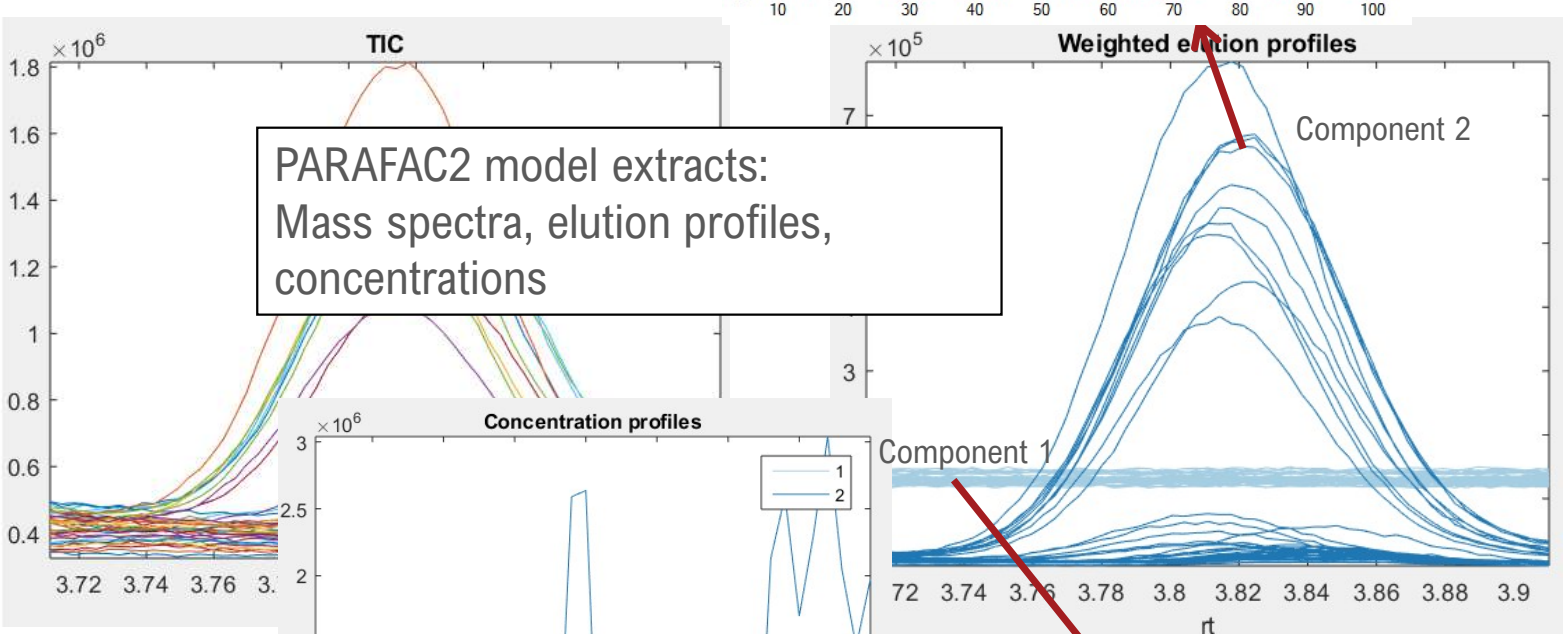
PARAFAC2

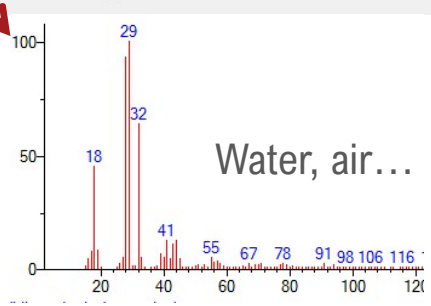
Harshman, R. A. (1972). PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics*, *22*, 30-44.

PARAFAC2 with two components



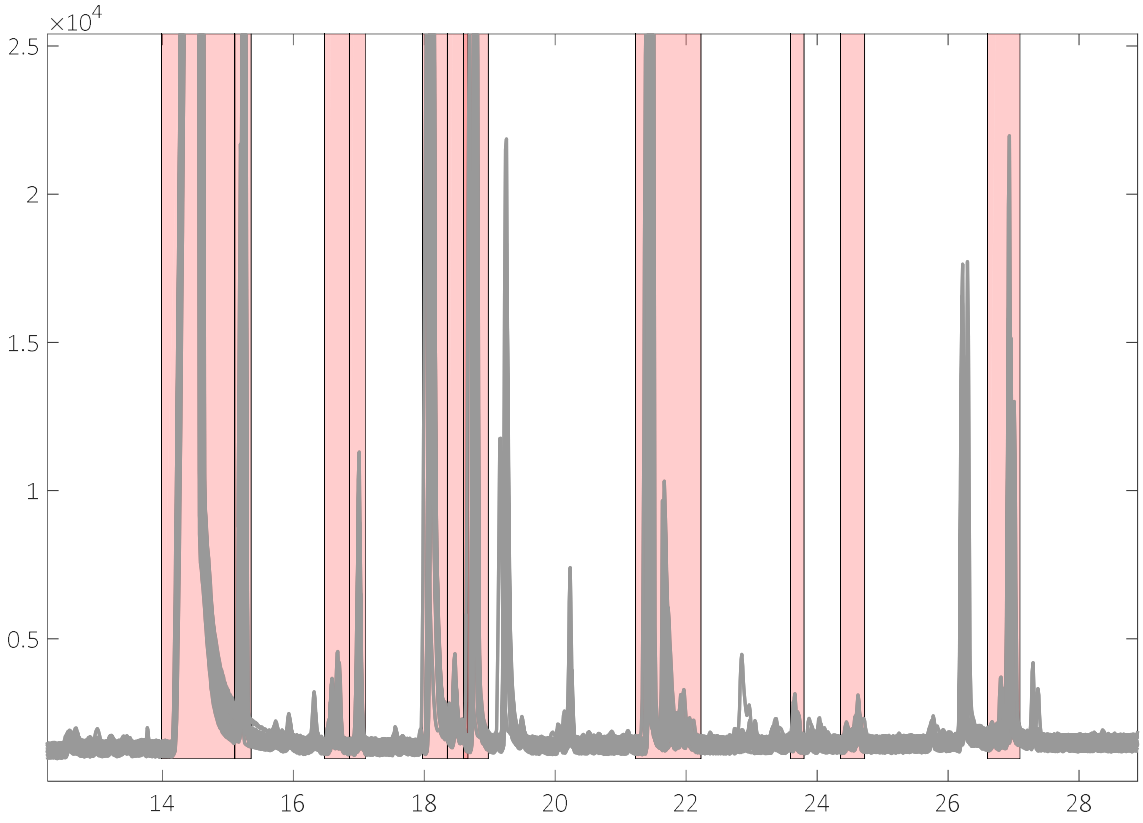
PARAFAC2 model extracts:
Mass spectra, elution profiles, concentrations

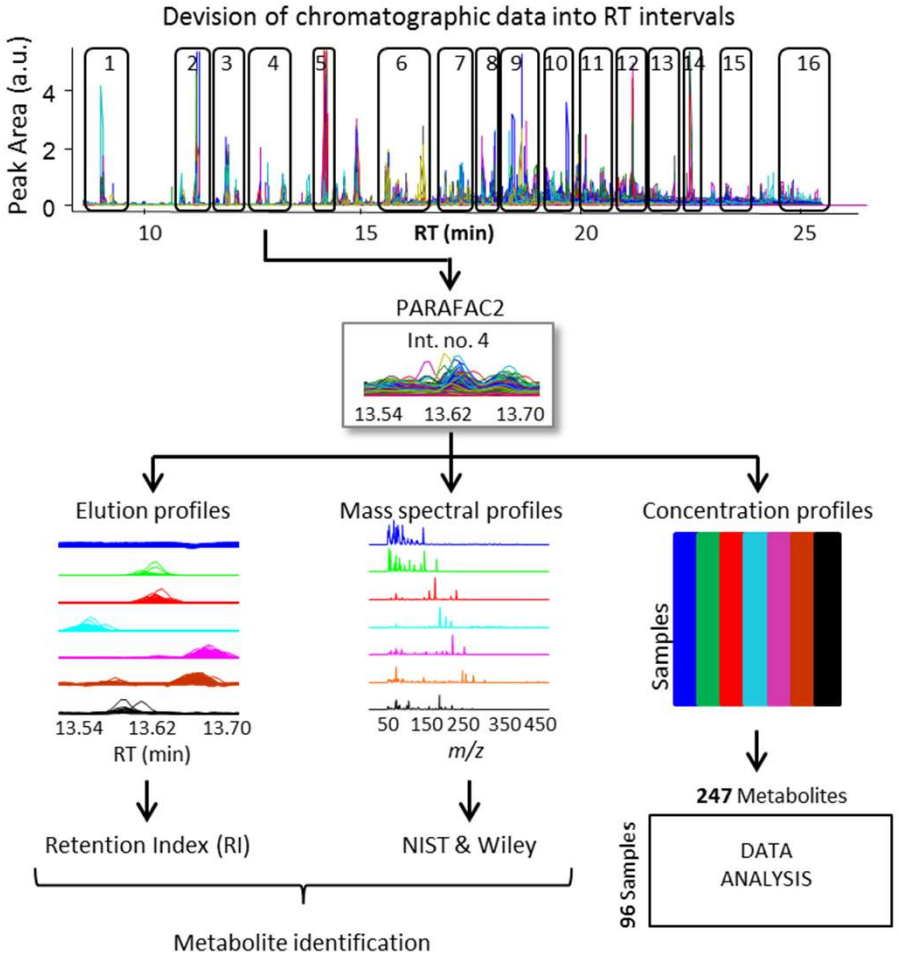




How to do it 1: Select intervals

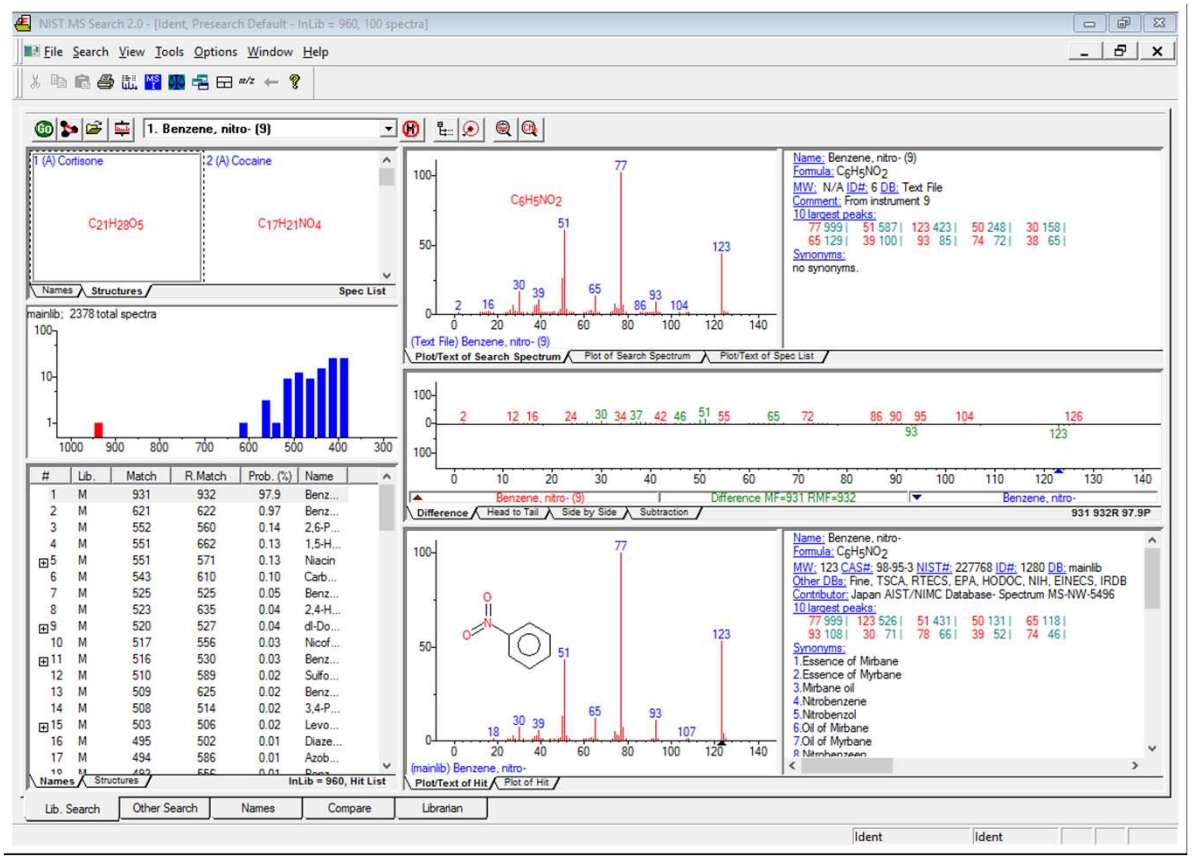
Don’t worry – it is not that critical





How to do it 2: Fit PARAFAC2 in each interval and identify the chemical components

How to do it 3: Use mass spectra to identify



How to do it 4: Make peak table

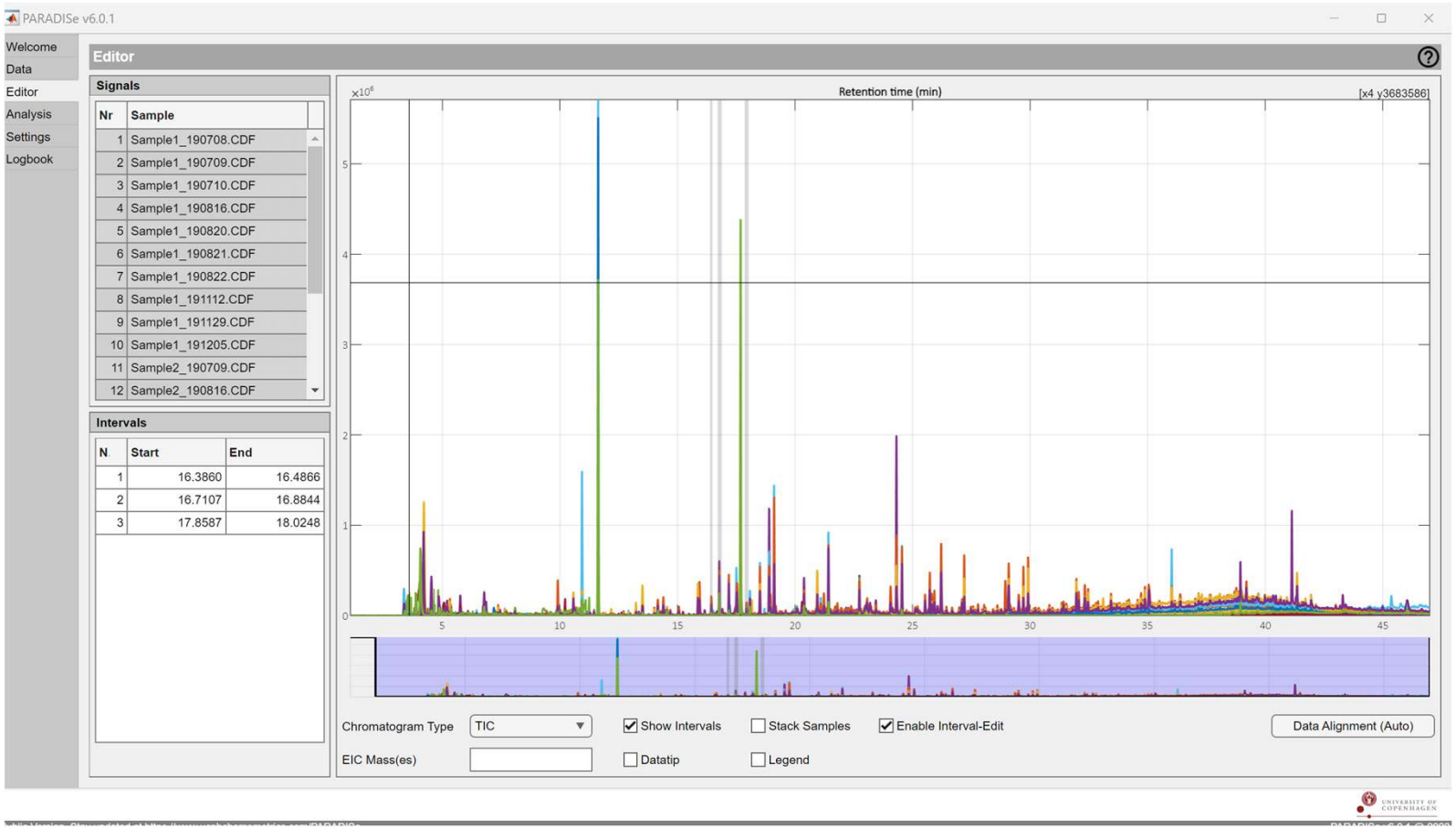
File Edit Transform View FigBrowser

Info Data Plot Row Labels Column Labels

Label Set 1 >	1: Ethylacetate	2: Ethanol	3: EthylPropionate	4: ButanoicAcid, et...	5: isobutylAlcohol	6: 1-butanol 3-meth..	7: Myrcene	8: Iso amylalcohol	9: Etyl Caproate
1: BR.1	7.5961571e+08	41393129	17902681	37568396	27017320	5.3985409e+08	2811695.7	1.2094407e+09	1.437568e+08
2: BR.2	7.6252091e+08	28659911	27683927	1.073743e+08	61947111	7.0724089e+08	8040094.3	1.2214659e+09	3.5925737e+08
3: CA.1	1.0576947e+09	33483838	32461276	73689514	45219636	1.1907209e+09	18046390	1.0858745e+09	1.9190012e+08
4: CA.2	1.051996e+09	36670850	23284583	46895827	77923900	1.1342623e+09	3110083.2	1.489822e+09	1.6581216e+08
5: CA.3	1.0278748e+09	96647909	45277493	44013683	34695044	1.1634978e+09	2818270.4	1.2098891e+09	1.3822257e+08
6: DE.1	1.0525303e+09	53632202	47613151	1.059843e+08	48190103	1.1621429e+09	1662815.3	1.4810101e+09	2.9515687e+08
7: DE.2	1.2978792e+09	46254285	66235674	90251136	42743930	1.4240906e+09	1182217.1	1.3728263e+09	2.4437158e+08
8: EU.1	1.1069234e+09	45146774	28800999	74840253	66039075	1.0523956e+09	1696614.1	1.7475233e+09	2.9582716e+08
9: EU.2	1.1659407e+09	20184890	27885482	80407334	71403759	1.1041941e+09	2165266.1	1.9329068e+09	3.1322591e+08
10: EU.3	1.0449123e+09	2.3932967e+08	37114051	95704743	53410764	1.3086851e+09	1540020.5	1.7368e+09	3.6812392e+08
11: EU.4	1.112918e+09	2.1928651e+08	44327512	84528779	62132751	1.049772e+09	1408825.1	1.6643148e+09	2.0859253e+08
12: FB.1	9.8910734e+08	4.2996183e+08	28900611	89368452	89661847	1.5782691e+09	774399.14	1.6515615e+09	2.7175456e+08
13: FB.2	1.0262077e+09	4.4422959e+08	41100911	1.7588415e+08	96386755	1.6972058e+09	271312.32	1.8981702e+09	3.5244115e+08
14: FB.3	1.0220532e+09	4.6739437e+08	40504422	1.8752781e+08	79501078	1.9316177e+09	465541.48	2.0558383e+09	4.4367049e+08
15: FO.1	8.5450441e+08	24062258	58930442	58412255	61071244	1.5999178e+08	3563329.6	1.5649239e+09	2.2100563e+08
16: FO.2	9.6885104e+08	10576361	26914535	57919276	50610603	3.649403e+08	268966.49	1.3867549e+09	2.5318616e+08
17: FU.1	1.0089355e+09	4.302324e+08	40874578	1.4144724e+08	56622318	1.3384891e+09	1794446.4	1.6706078e+09	3.8742713e+08
18: FU.2	1.0983342e+09	4.120951e+08	31108465	1.4026276e+08	79792925	1.5239737e+09	1341524.8	1.5380575e+09	3.5770928e+08
19: HB.1	7.292056e+08	18512413	92866284	68935462	1.2538492e+08	21649058	58869916	1.8814001e+09	3.1825371e+08
20: HB.2	6.7687583e+08	38296666	43028962	38971009	71832146	1.1469062e+08	24077124	2.0679402e+09	1.9368067e+08
21: HI.4	1.0513647e+09	35807140	19116411	65591409	68537900	1.7751781e+09	873919.82	1.4576126e+09	2.3590077e+08
22: HI.7	1.1105261e+09	58784921	51892761	81622287	67445642	1.9495565e+09	2150585.2	2.0408539e+09	2.3829981e+08
23: HI.8	1.0357279e+09	19498403	17351190	57924942	61732584	1.4776179e+09	331916.25	1.5419158e+09	2.5820037e+08
24: HI.9	1.0826754e+09	28855505	17873584	58939662	73254702	1.4954376e+09	812806.29	1.5753531e+09	2.639788e+08
25: HI.10	1.1380725e+09	9442046.5	16318172	89260927	56525071	1.5881895e+09	1278046.2	1.2068533e+09	4.009337e+08
26: HI.13	1.3584484e+09	26094994	30514192	78480582	88738497	1.4457027e+09	2849069.4	1.7220419e+09	3.3298972e+08
27: JA.1	1.1674513e+09	43722274	39899883	1.0380964e+08	69354570	1.7622099e+09	1112214.4	1.4039243e+09	2.7280185e+08
28: JA.2	8.4272529e+08	32747089	63537401	1.3139738e+08	47754696	6.9636803e+08	541527.29	2.2289966e+09	3.2954727e+08
29: KR.1	9.4208219e+08	18480222	27316805	69914569	34052283	1.0067676e+09	1345485.1	1.3537962e+09	4.9214685e+08

^ Beer Sample...

All handled in the software PARADISe

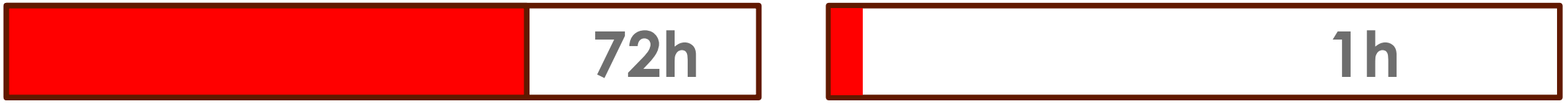


Traditional

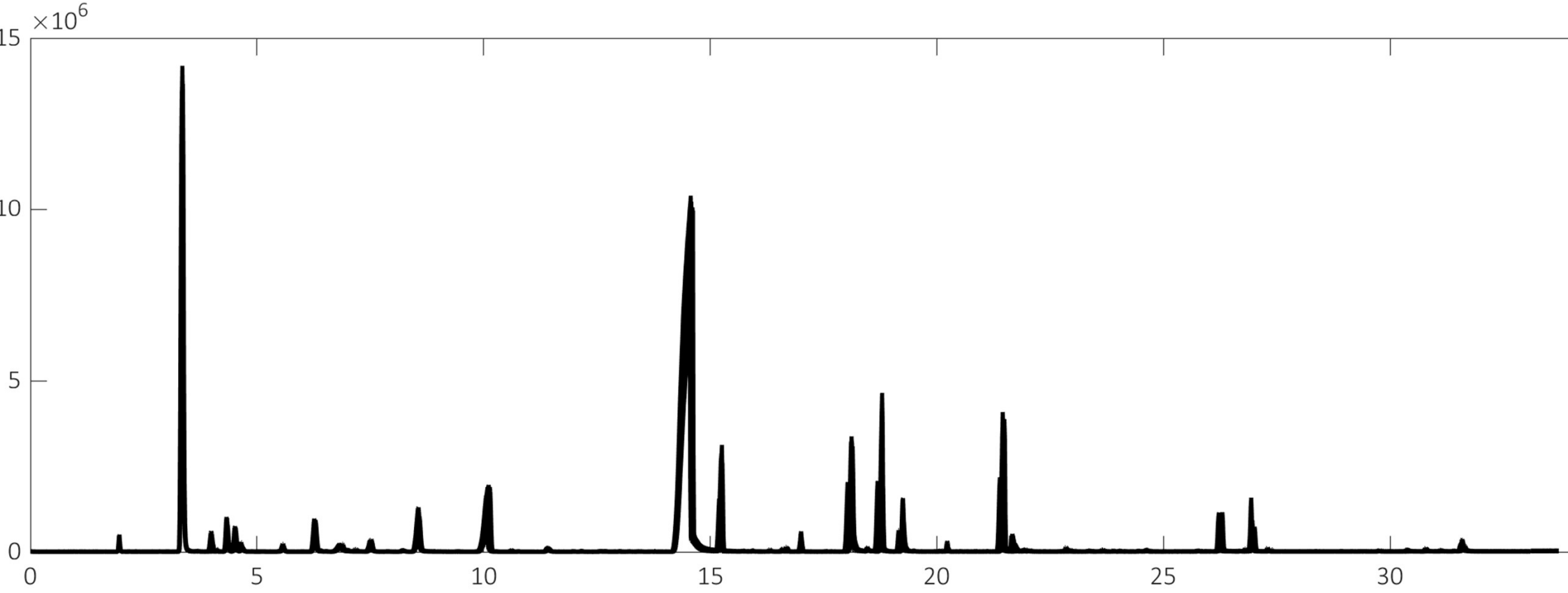
57 aroma compounds

44 Wine Samples

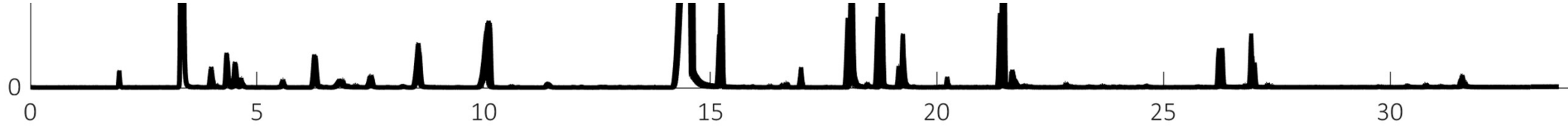
Traditional	PARADISe
57 aroma compounds	120 aroma compounds



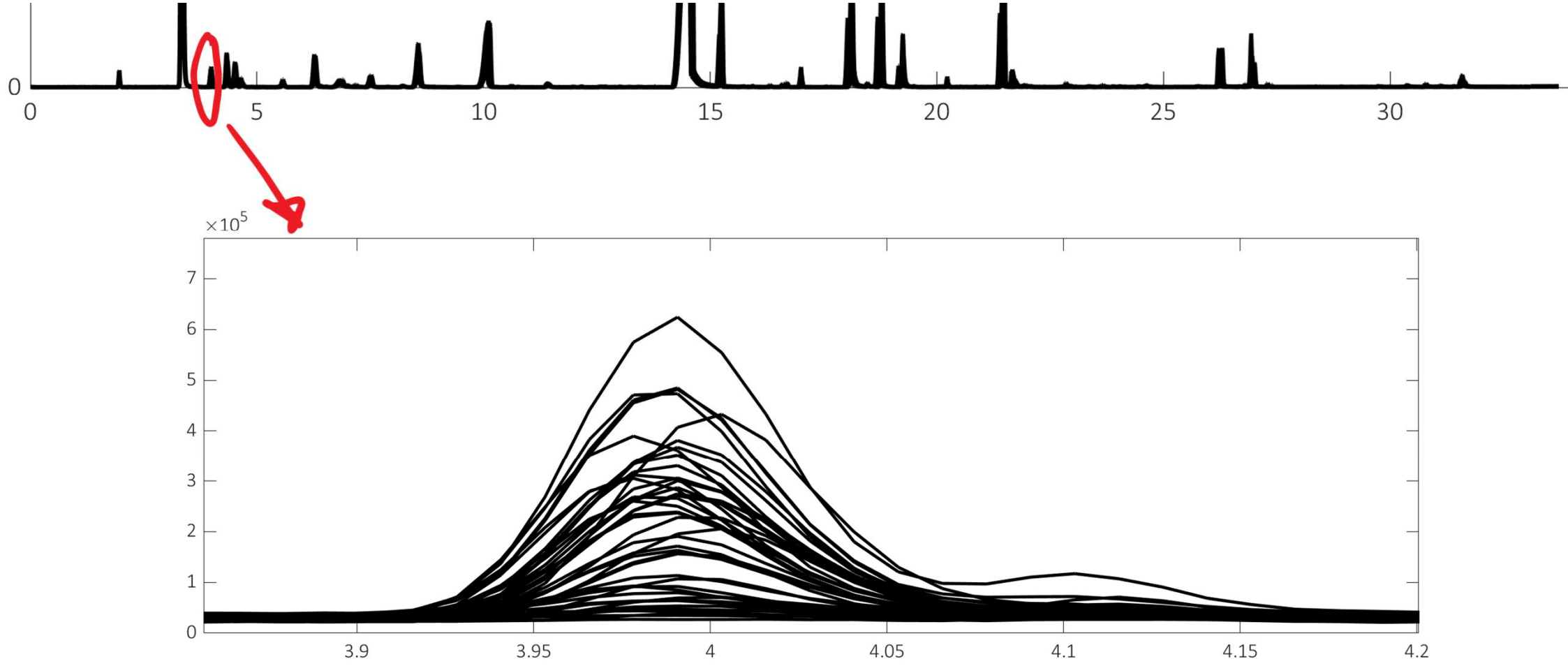
44 Wine Samples



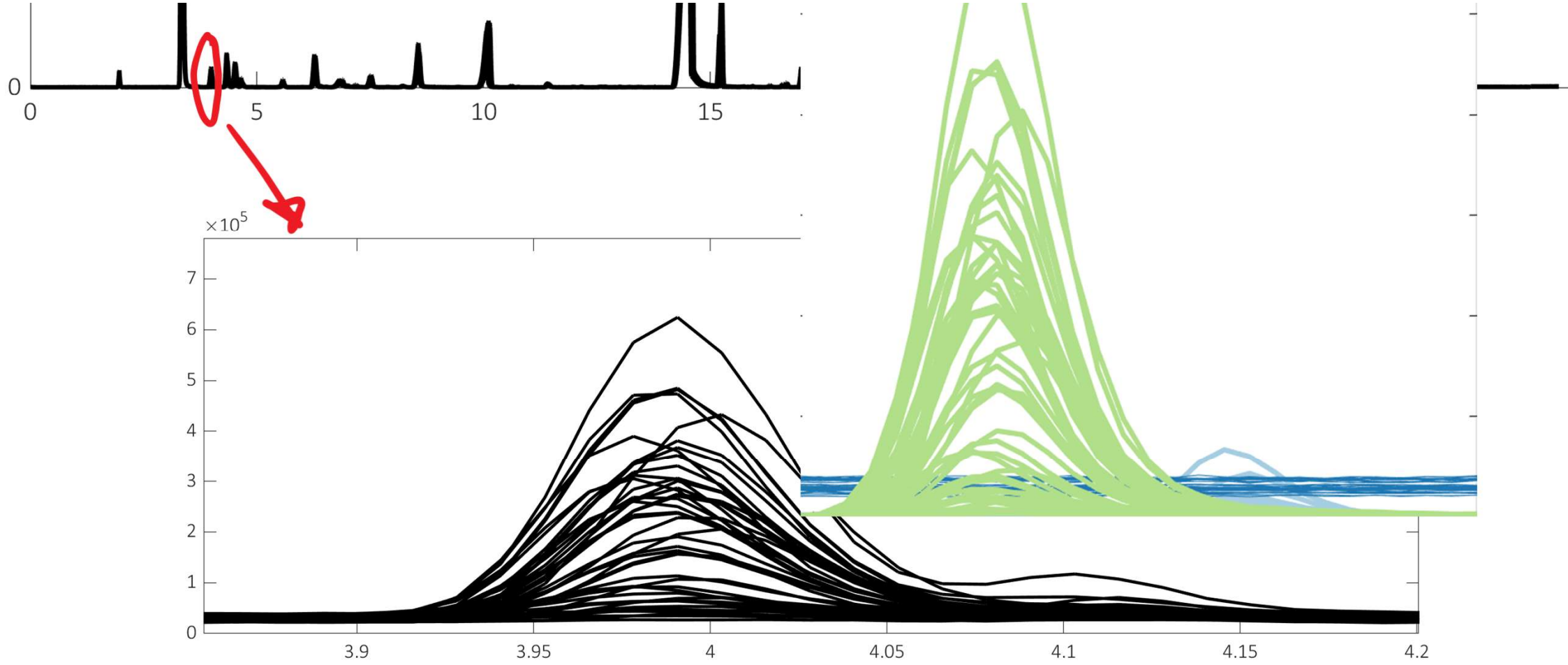
Cabernet Sauvignon



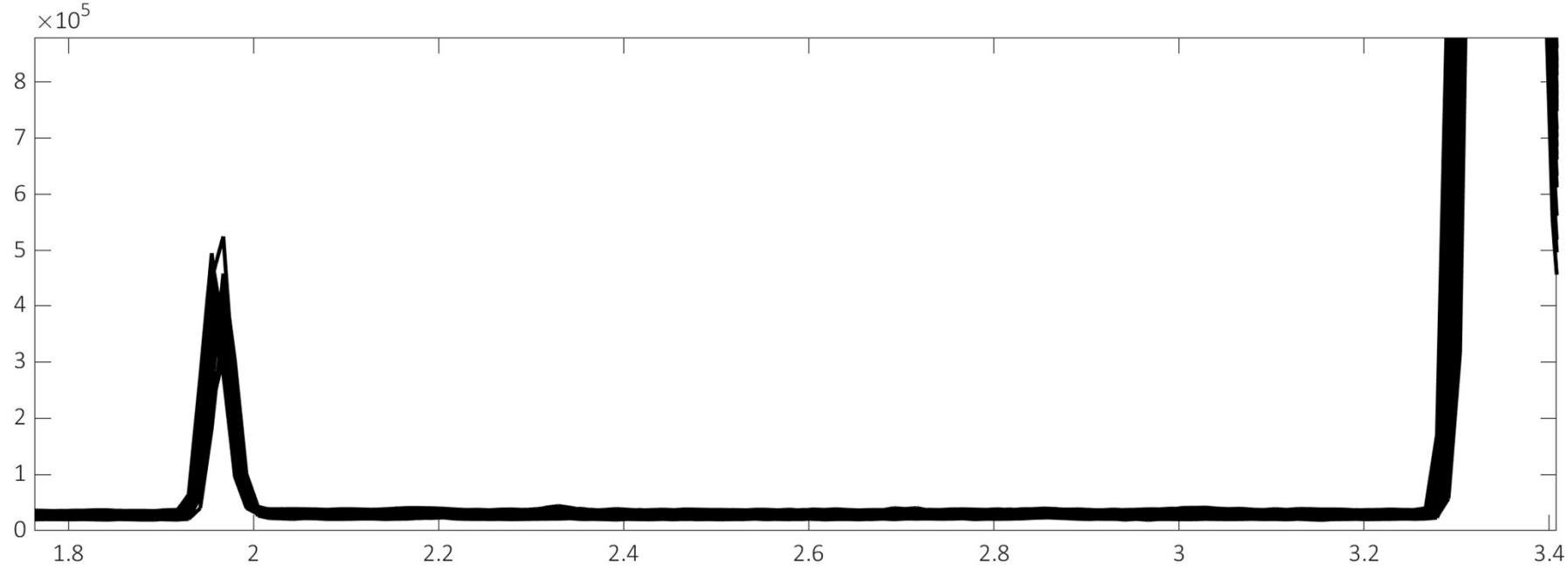
Cabernet Sauvignon



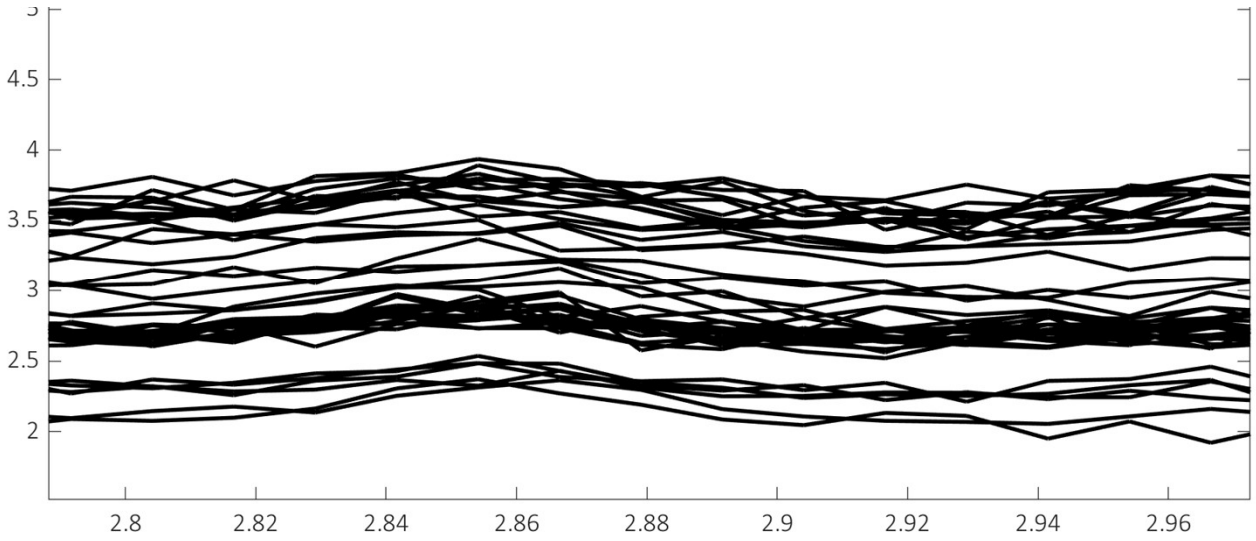
Cabernet Sauvignon

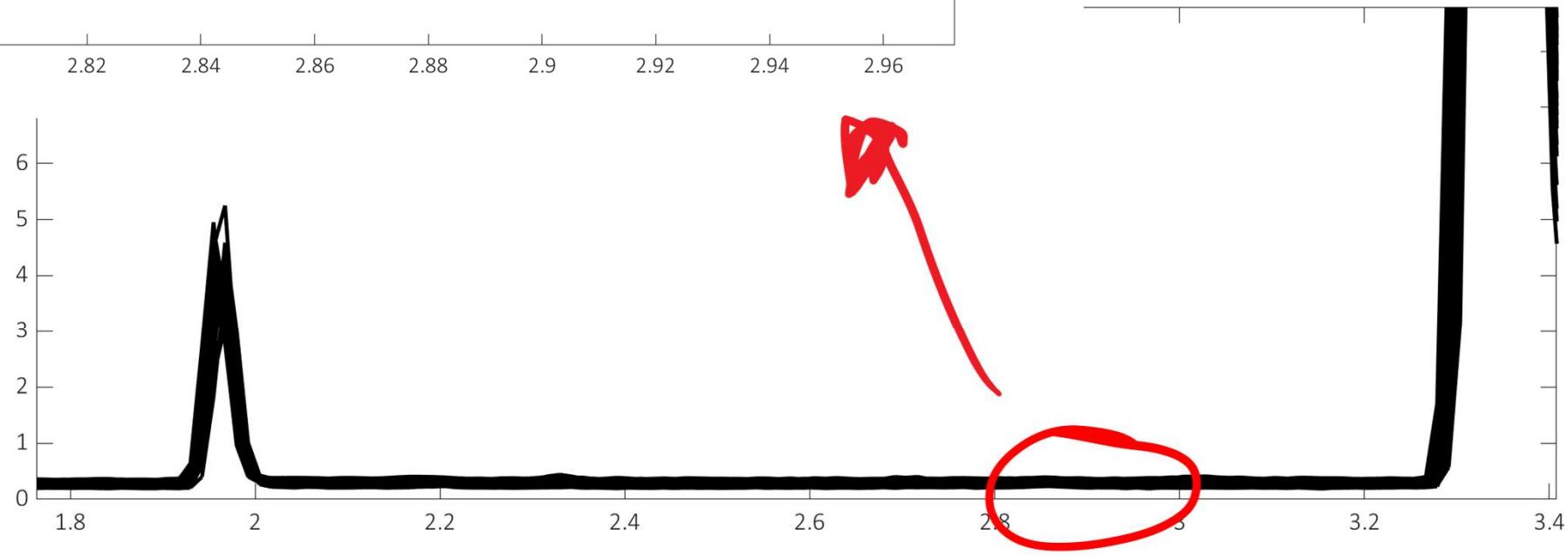


Cabernet Sauvignon

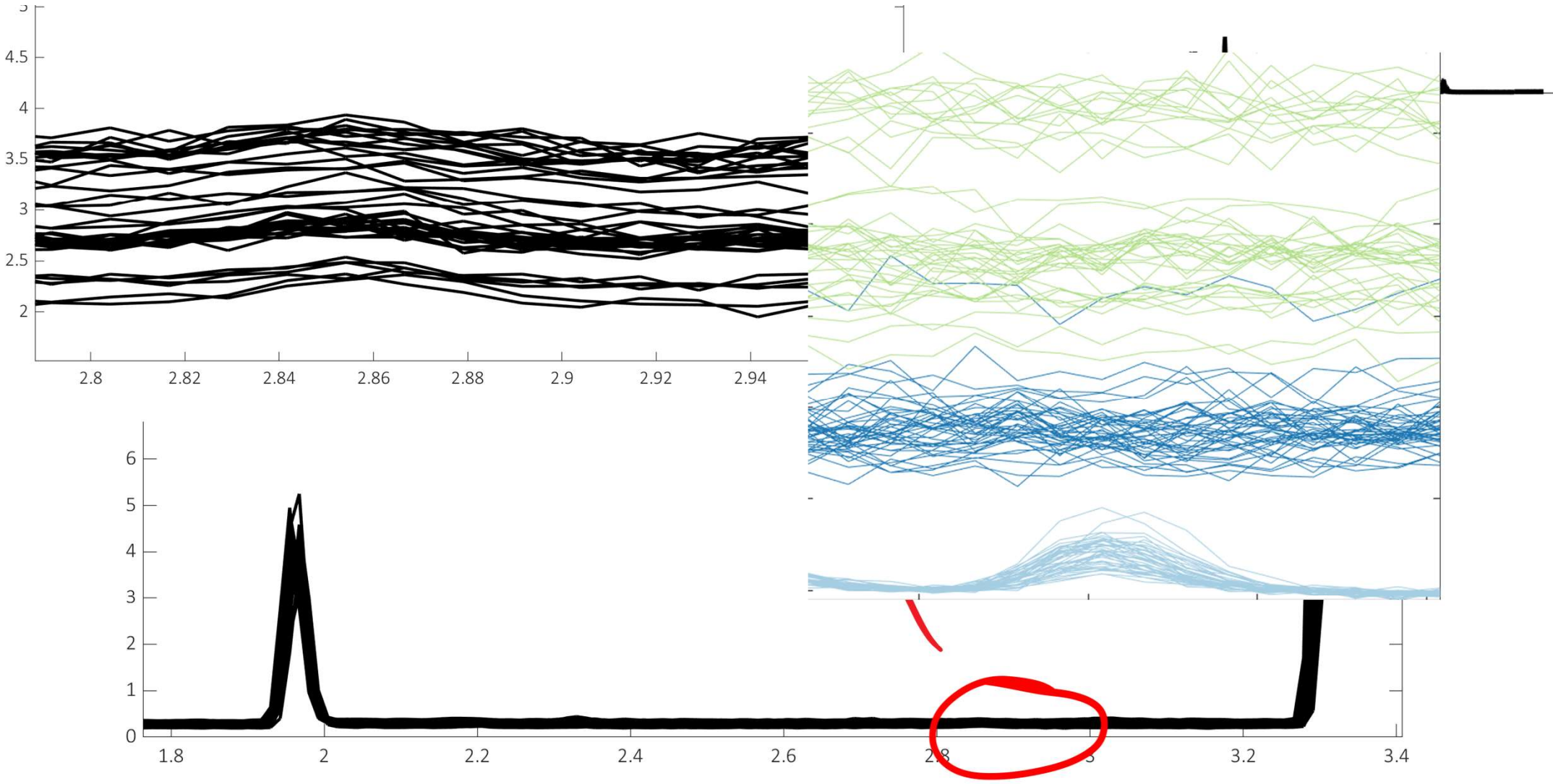


Cabernet Sauvignon

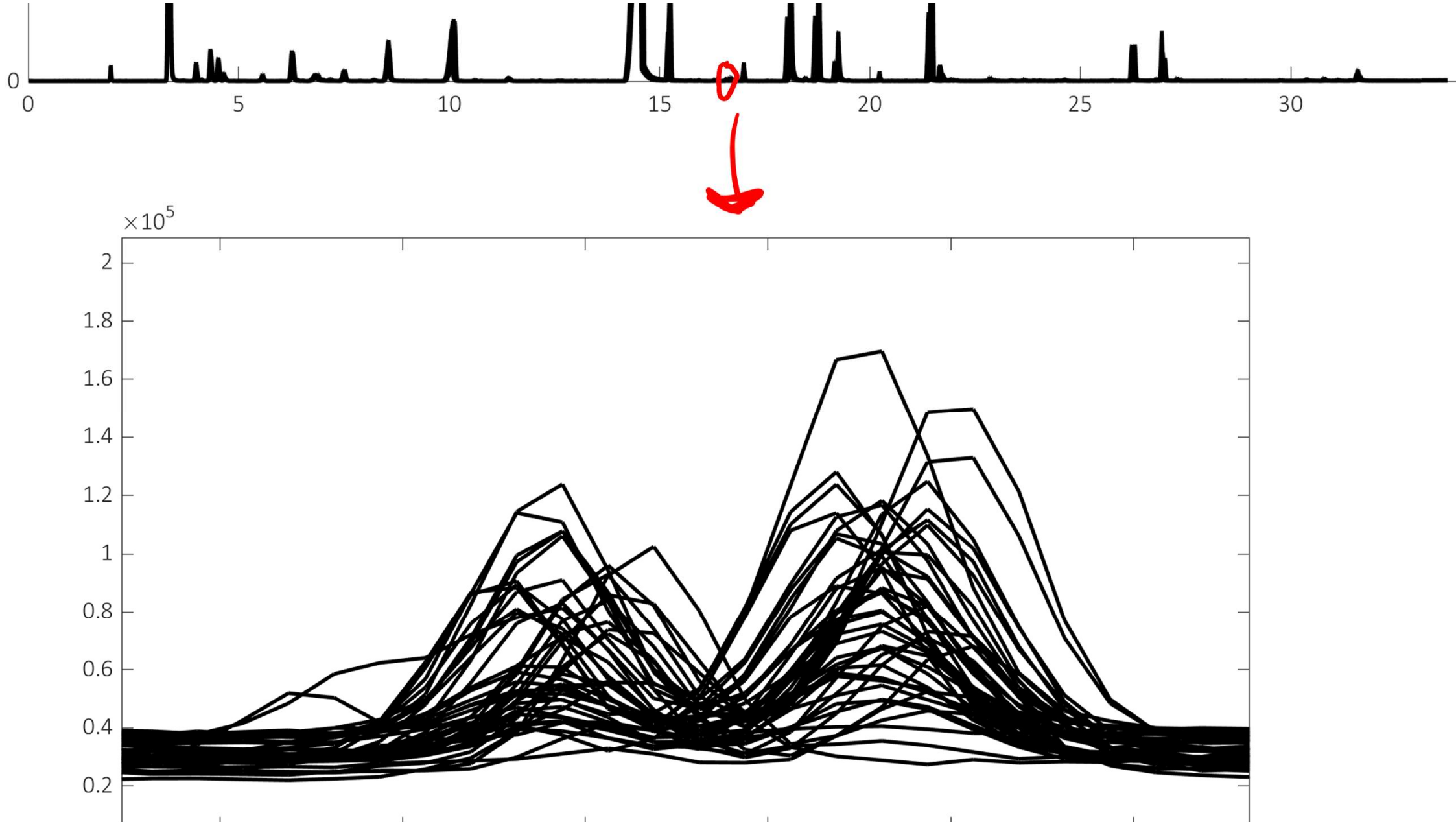


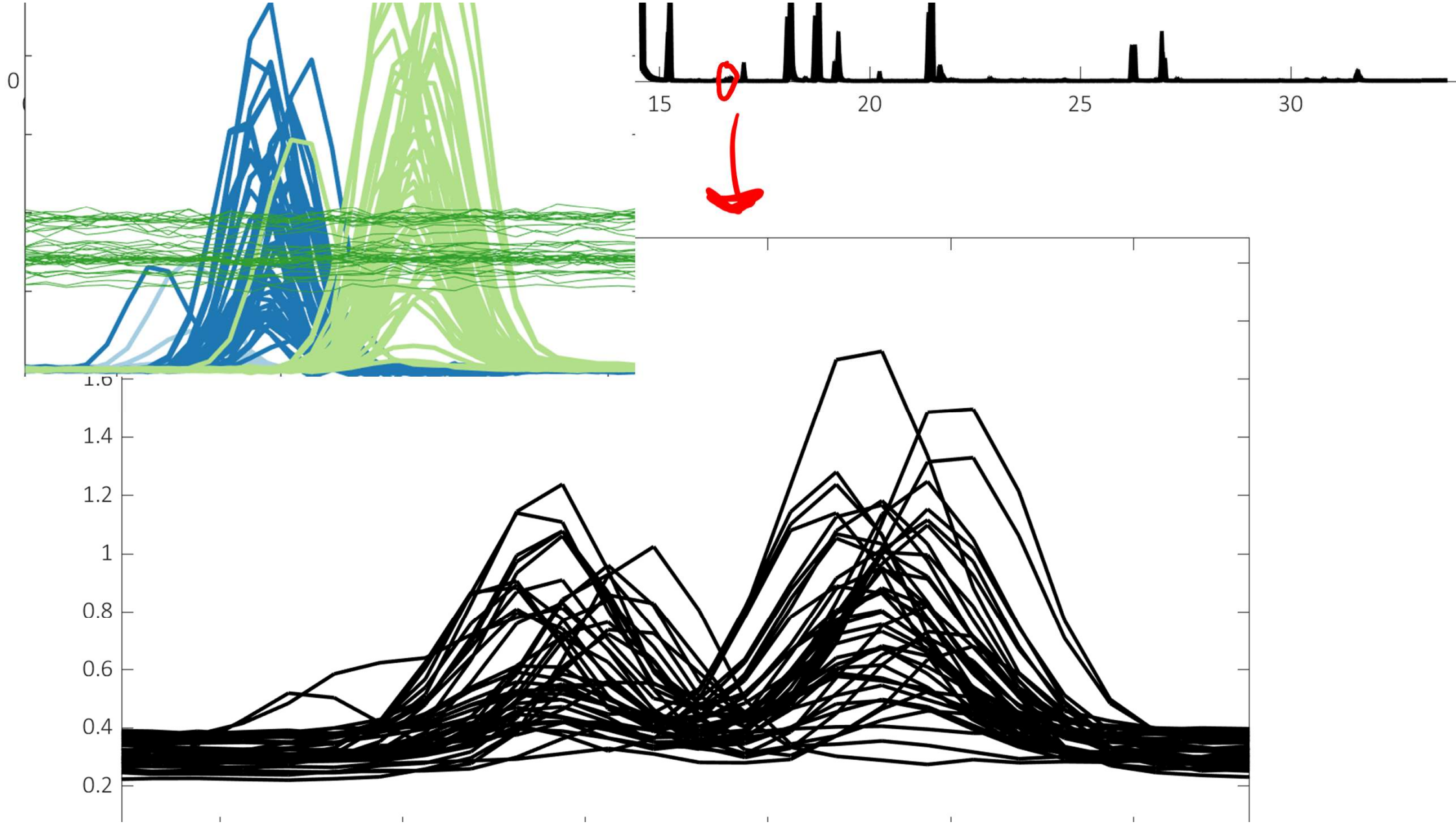


Cabernet Sauvignon



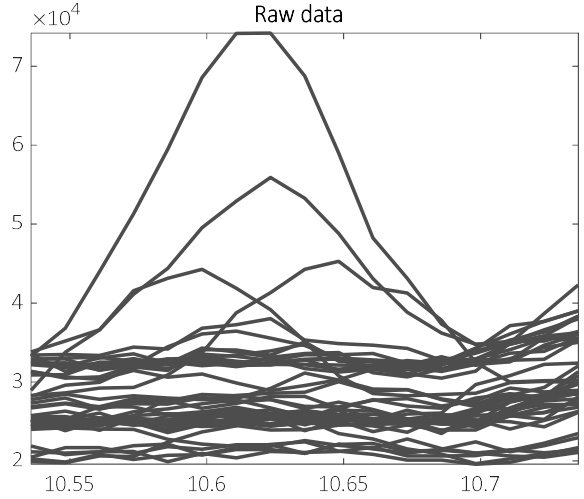
Cabernet Sauvignon

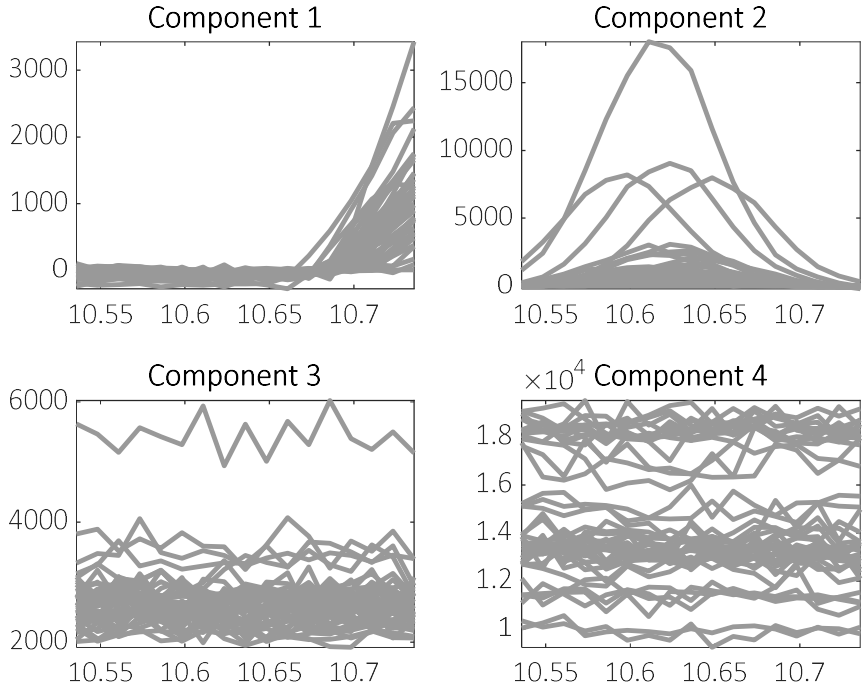

0
0
5
10
15
20
25
30
×10^5
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2


0
15
20
25
30
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2

Automating further

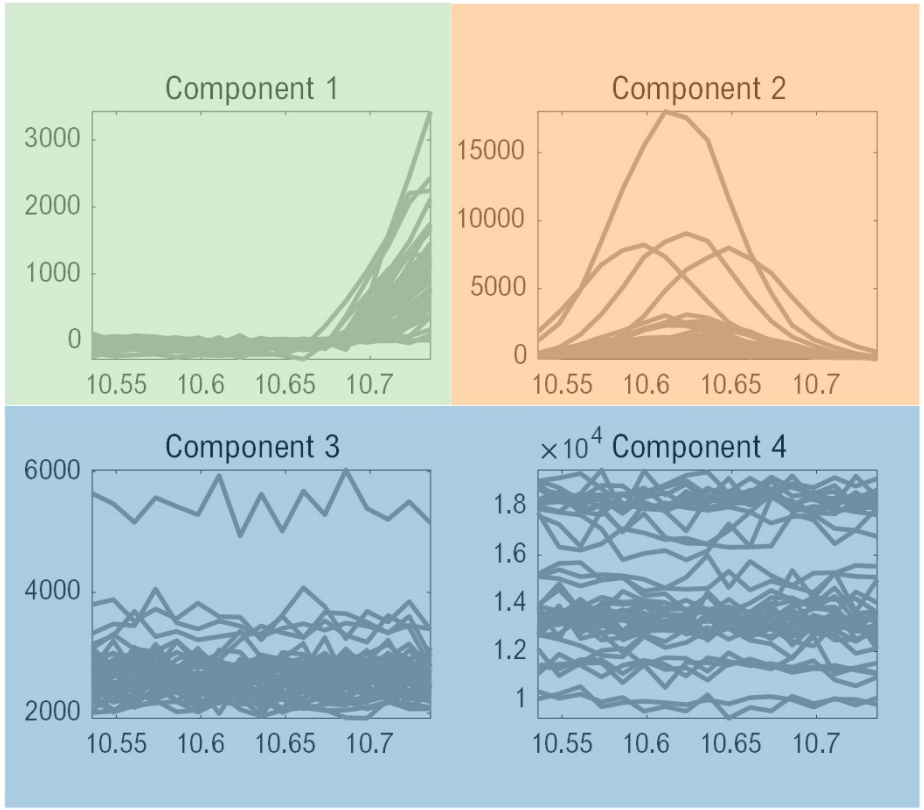
Which features are chemical?



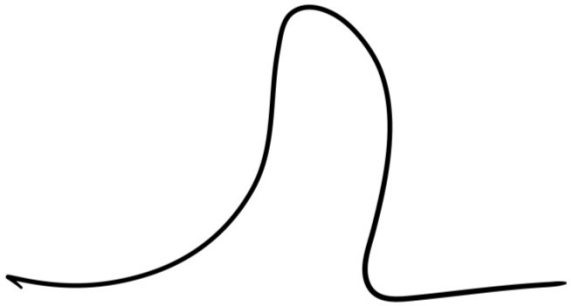


What we still have to do select

Baseline
Cutoff
Peak
Other

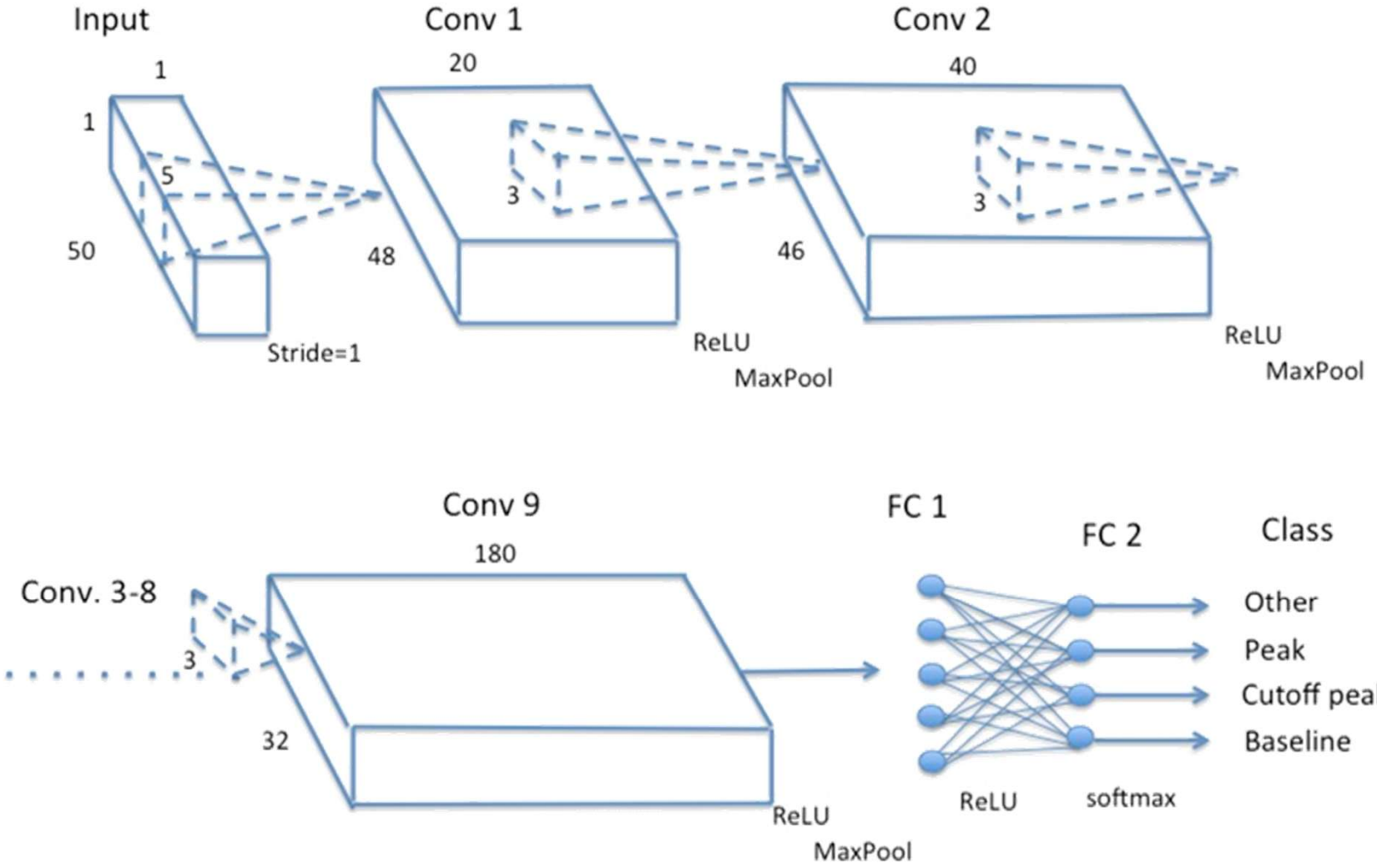


How to assess a peak?



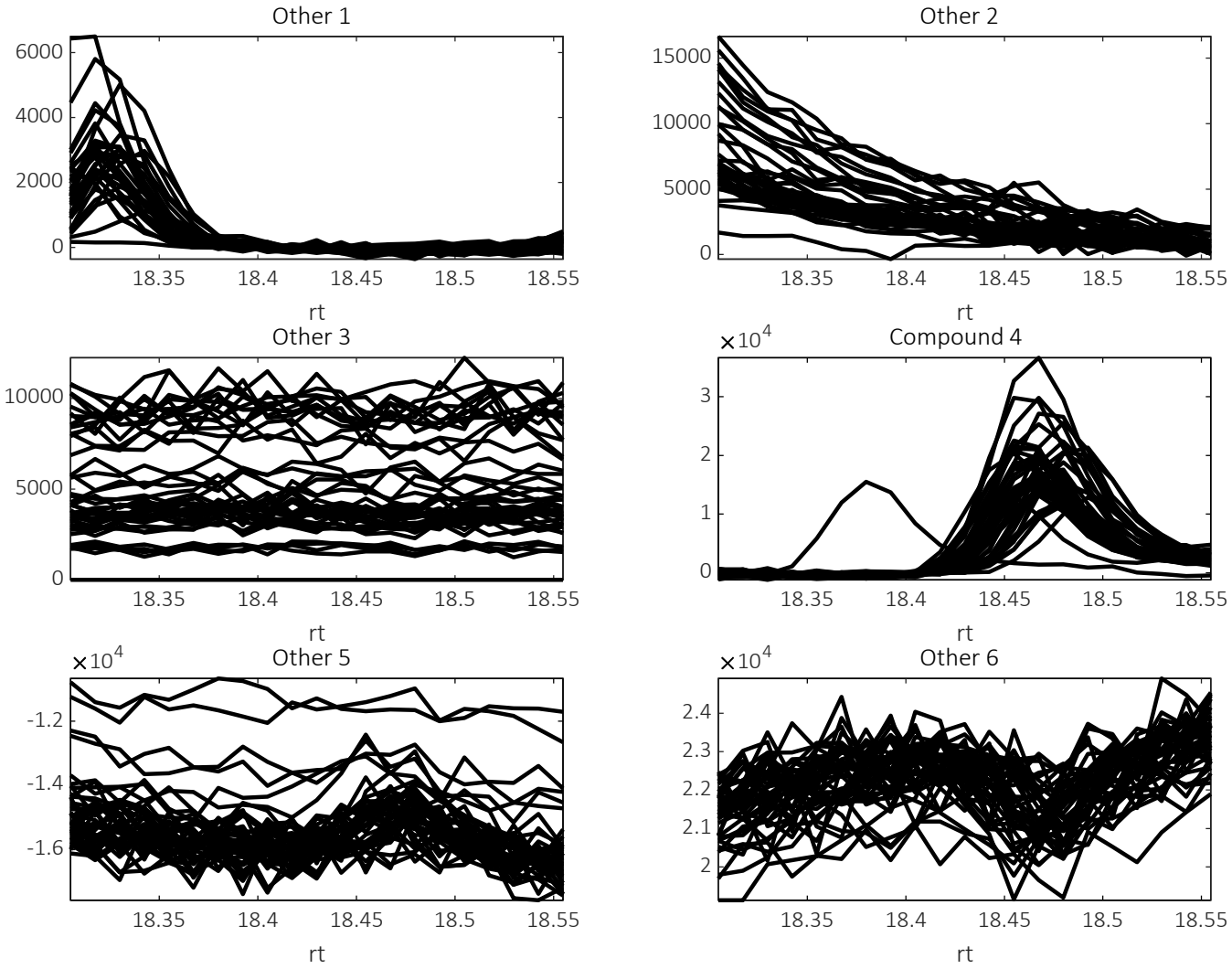
YES

NO



Neural net with 42 hidden layers

Profiles can now be individually qualified

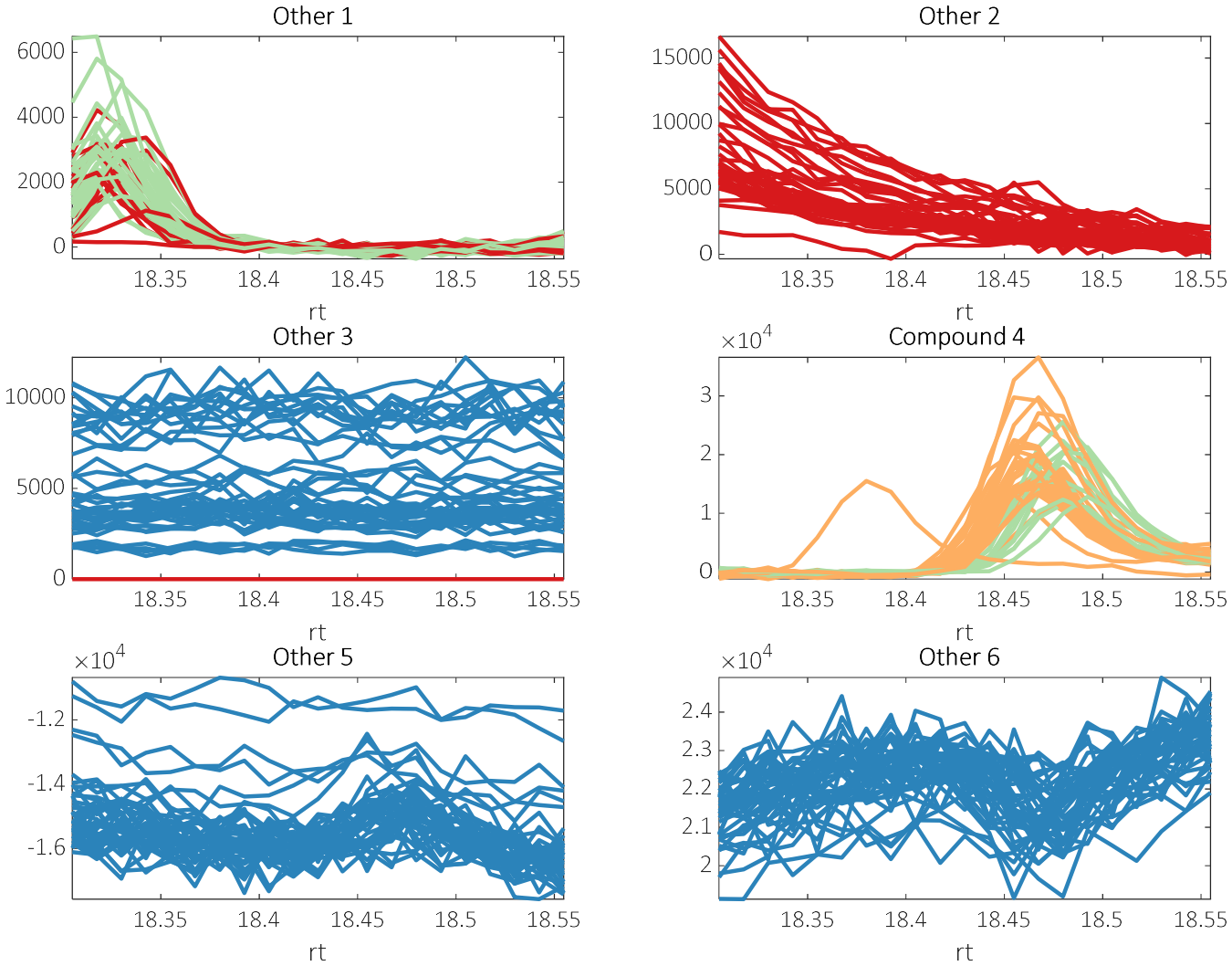


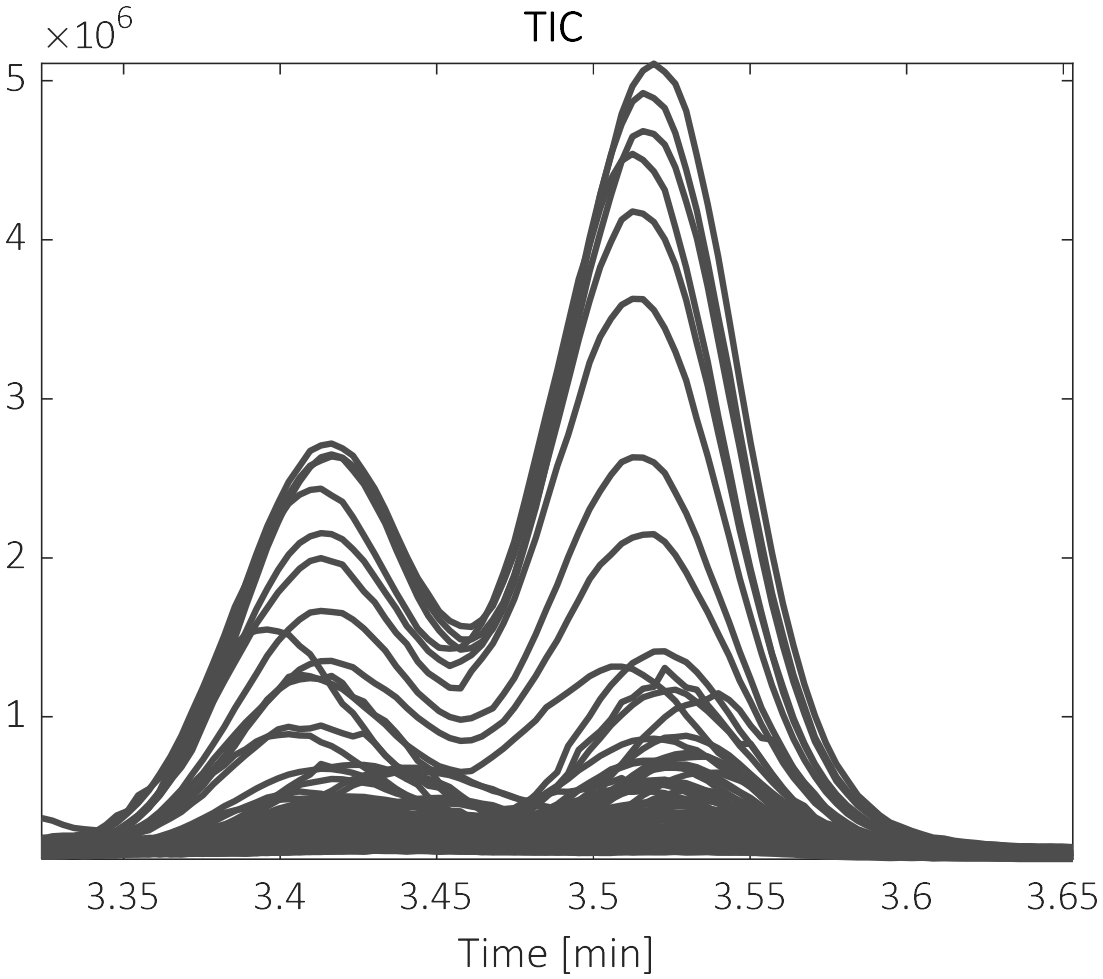
Other

Peak

Cutoff

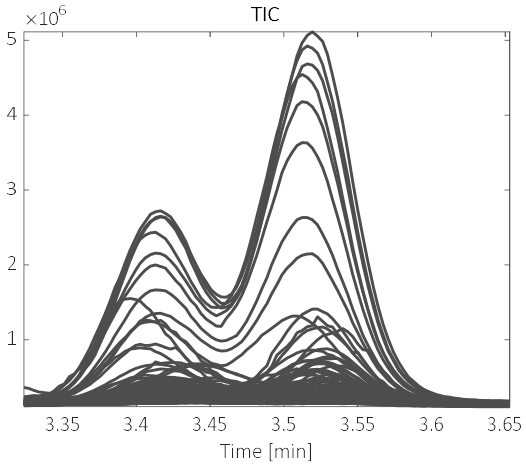
Baseline

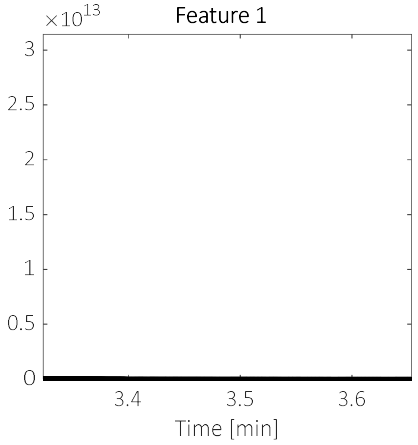


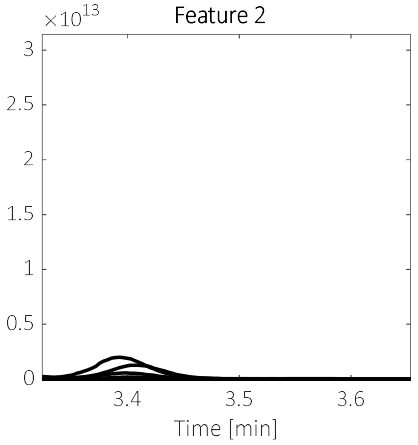


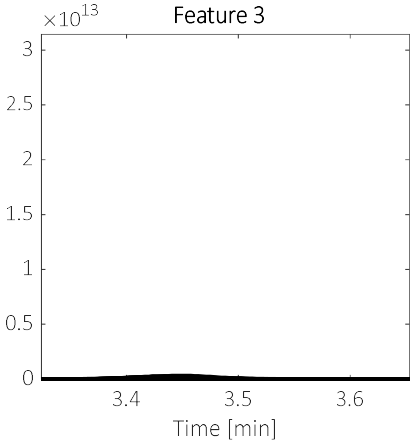
Well-known problem
2- and 3-methylbutanal

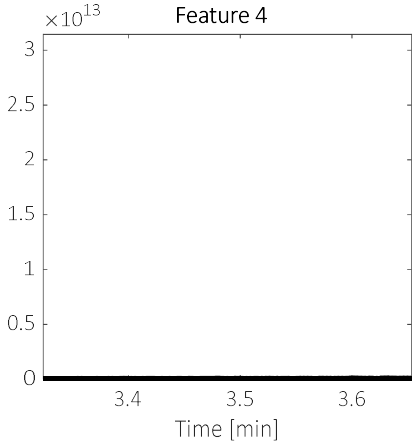
Example apple wine

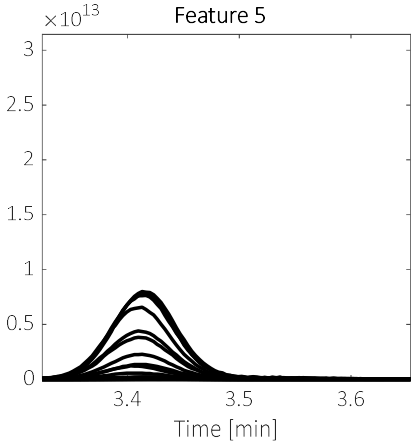


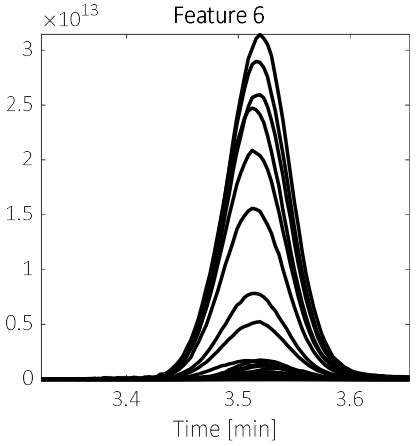




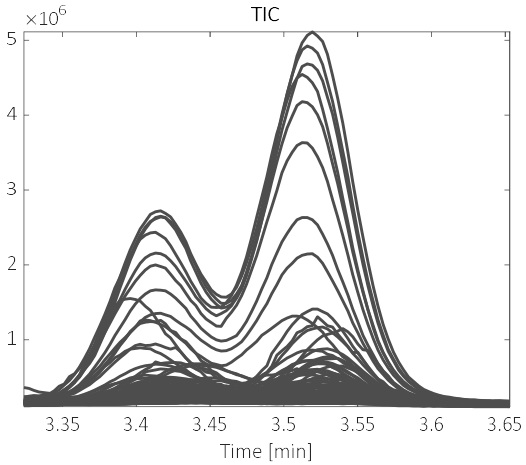


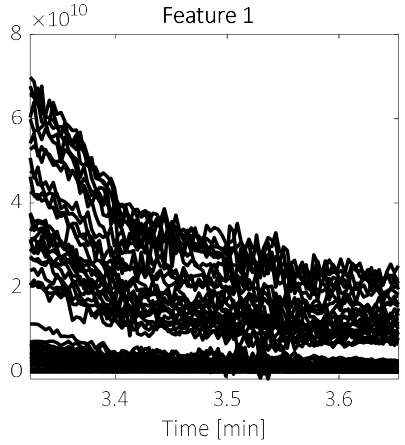


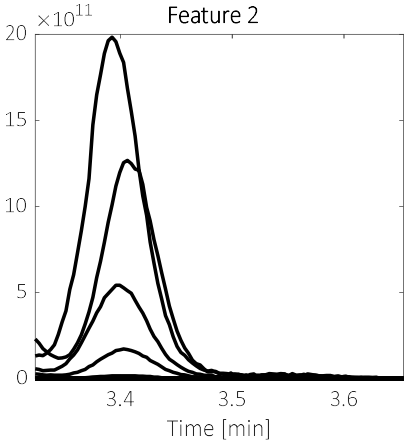


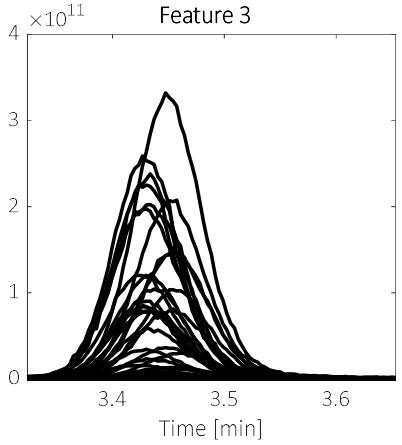


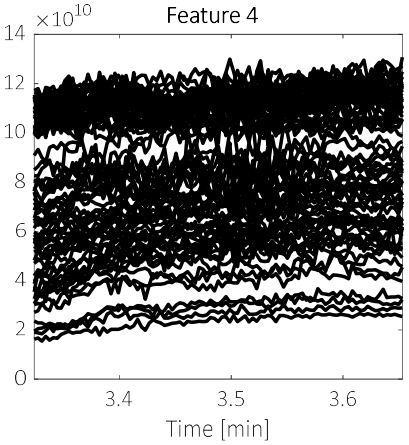
Scaled by concentration

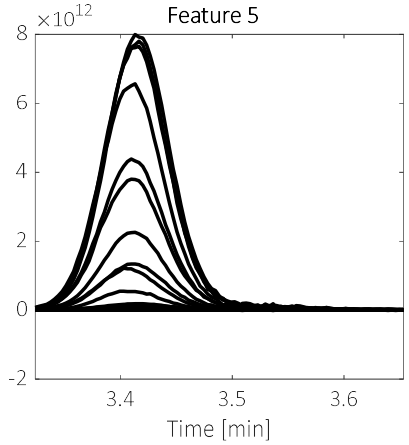

TIC
×10^6
5
4
3
2
1
3.35
3.4
3.45
3.5
3.55
3.6
3.65
Time [min]

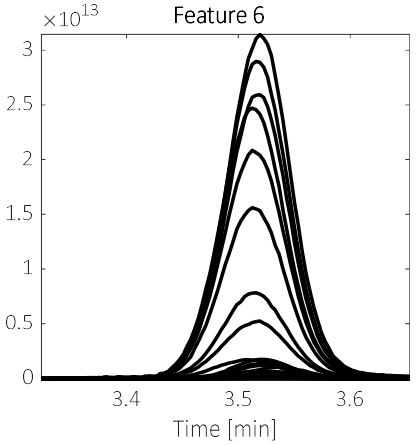

Feature 1
×10^10
8
6
4
2
0
3.4
3.5
3.6
Time [min]


Feature 2
×10^11
20
15
10
5
0
3.4
3.5
3.6
Time [min]

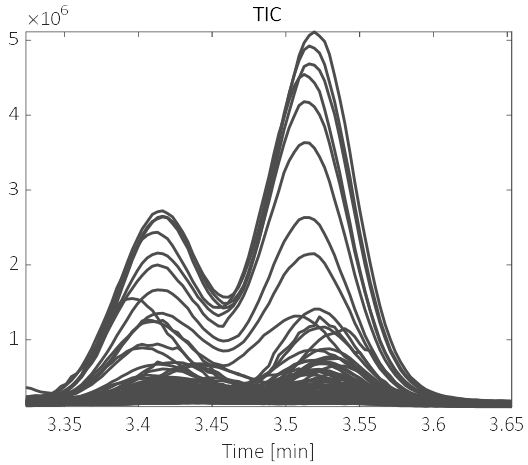

Feature 3
×10^11
4
3
2
1
0
3.4
3.5
3.6
Time [min]

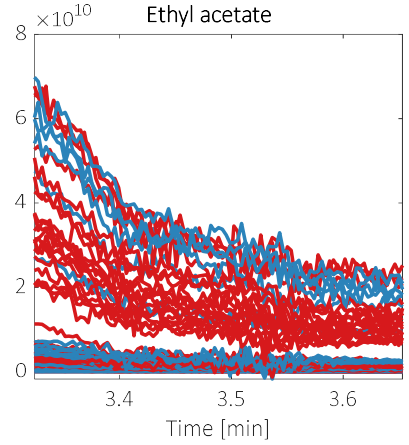

Feature 4
×10^10
14
12
10
8
6
4
2
0
3.4
3.5
3.6
Time [min]

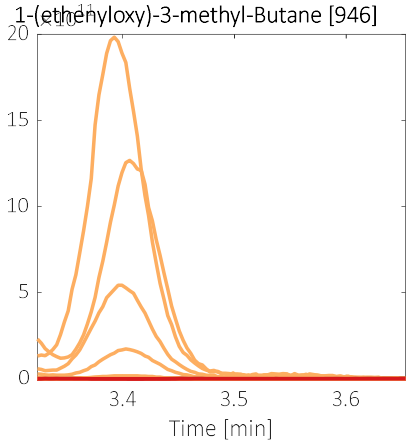

Feature 5
×10^12
8
6
4
2
0
-2
3.4
3.5
3.6
Time [min]

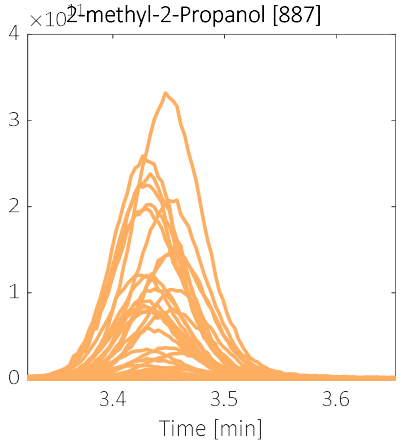

Feature 6
×10^13
3
2.5
2
1.5
1
0.5
0
3.4
3.5
3.6
Time [min]

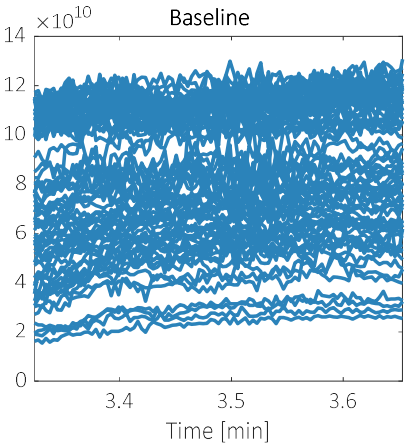
Non-scaled

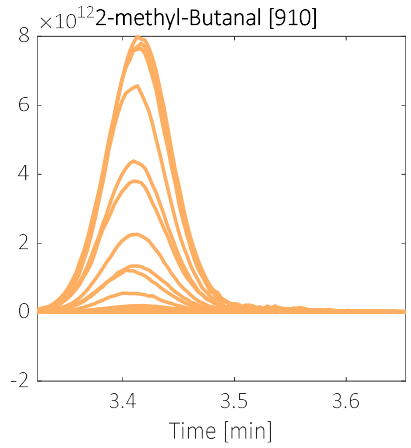

TIC
×10^6
Time [min]

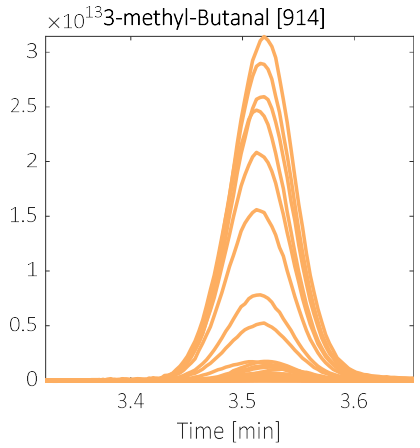

Ethyl acetate
×10^10
Time [min]


1-(ethenyloxy)-3-methyl-Butane [946]
Time [min]

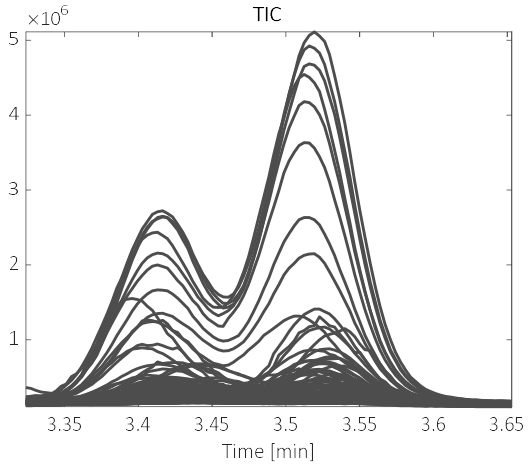

2-methyl-2-Propanol [887]
×10^21
Time [min]

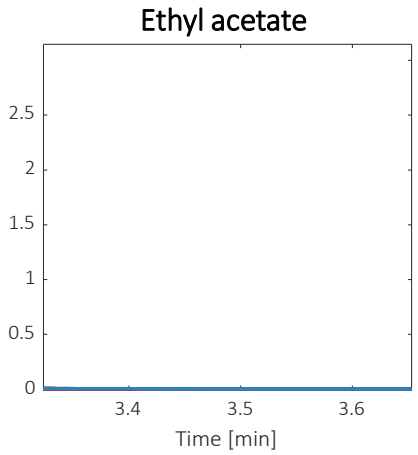

Baseline
×10^10
Time [min]

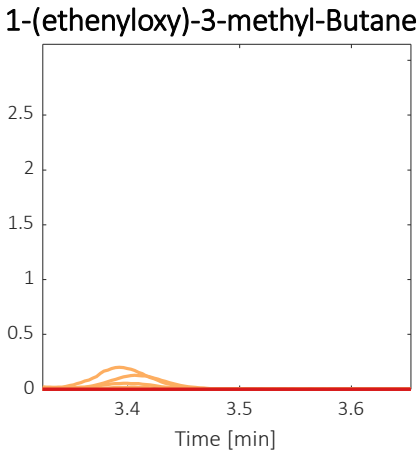

2-methyl-Butanal [910]
×10^12
Time [min]

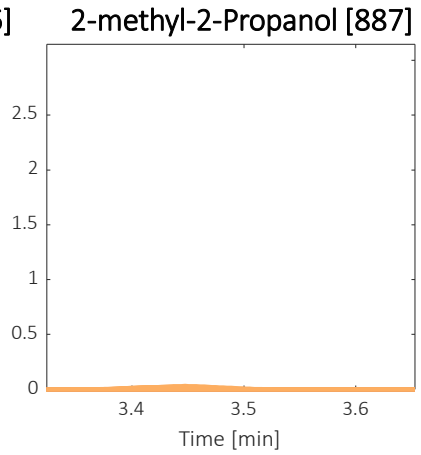

3-methyl-Butanal [914]
×10^13
Time [min]

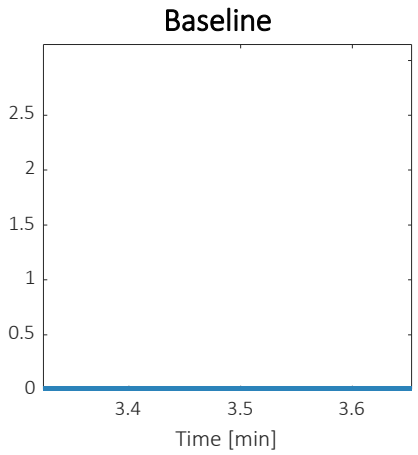
Non-scaled

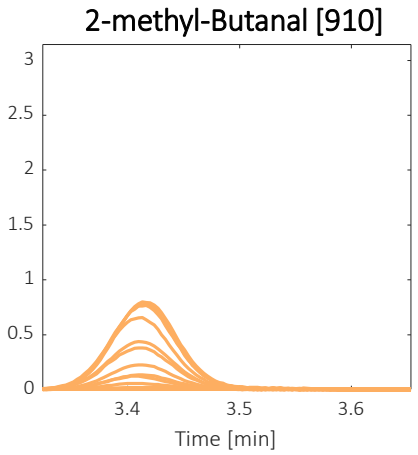


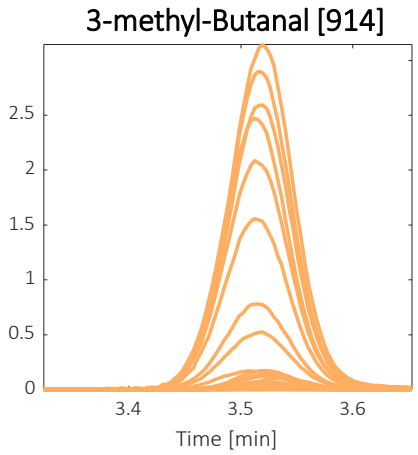












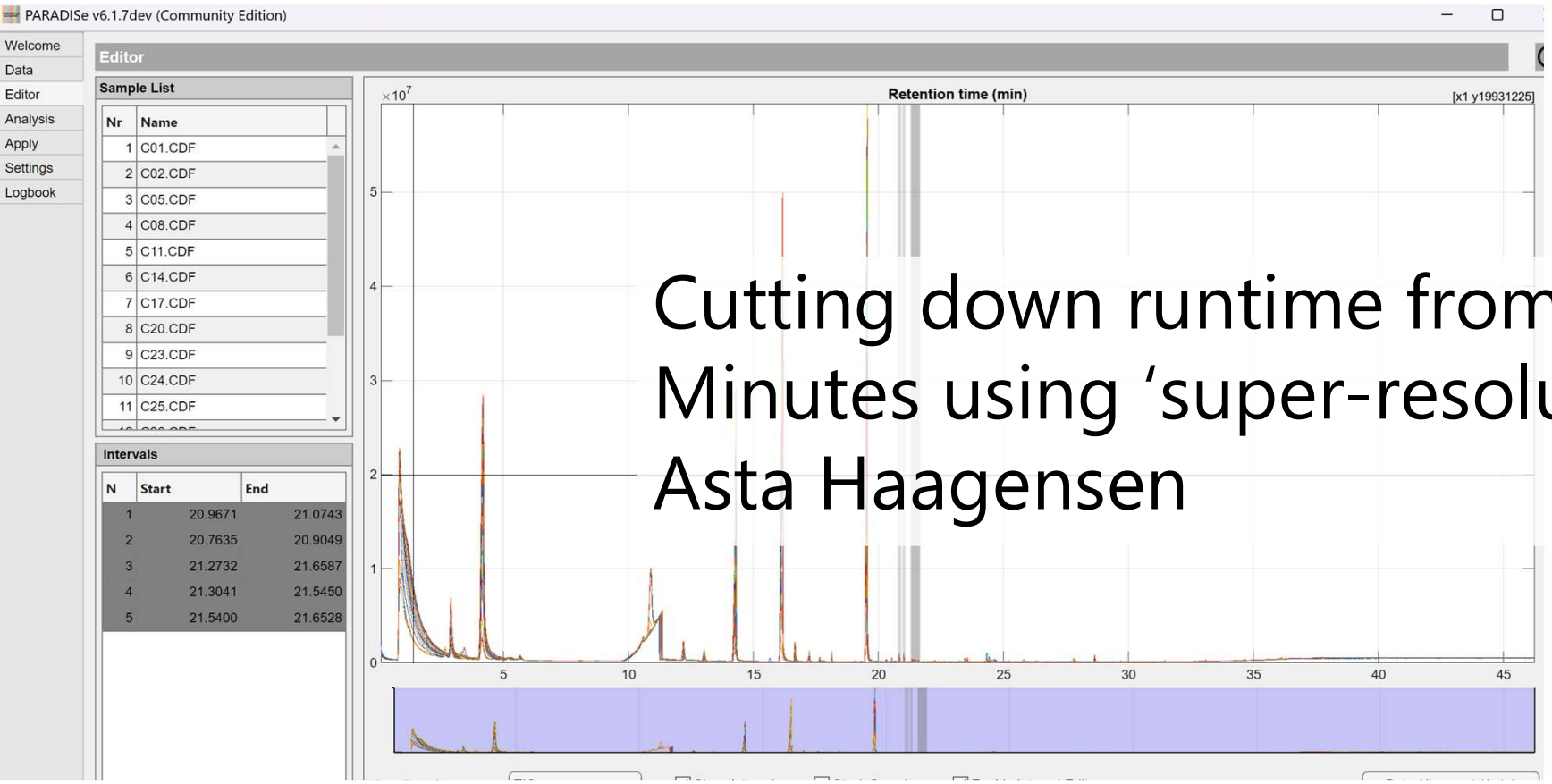
Scaled by concentration

We can handle

- retention time shifts
- baseline variation
- better estimated spectra (idenfication)
- severe co-elution
- no LOD
- speed up analysis
- improve robustness (less user-influence)
- thousands of samples

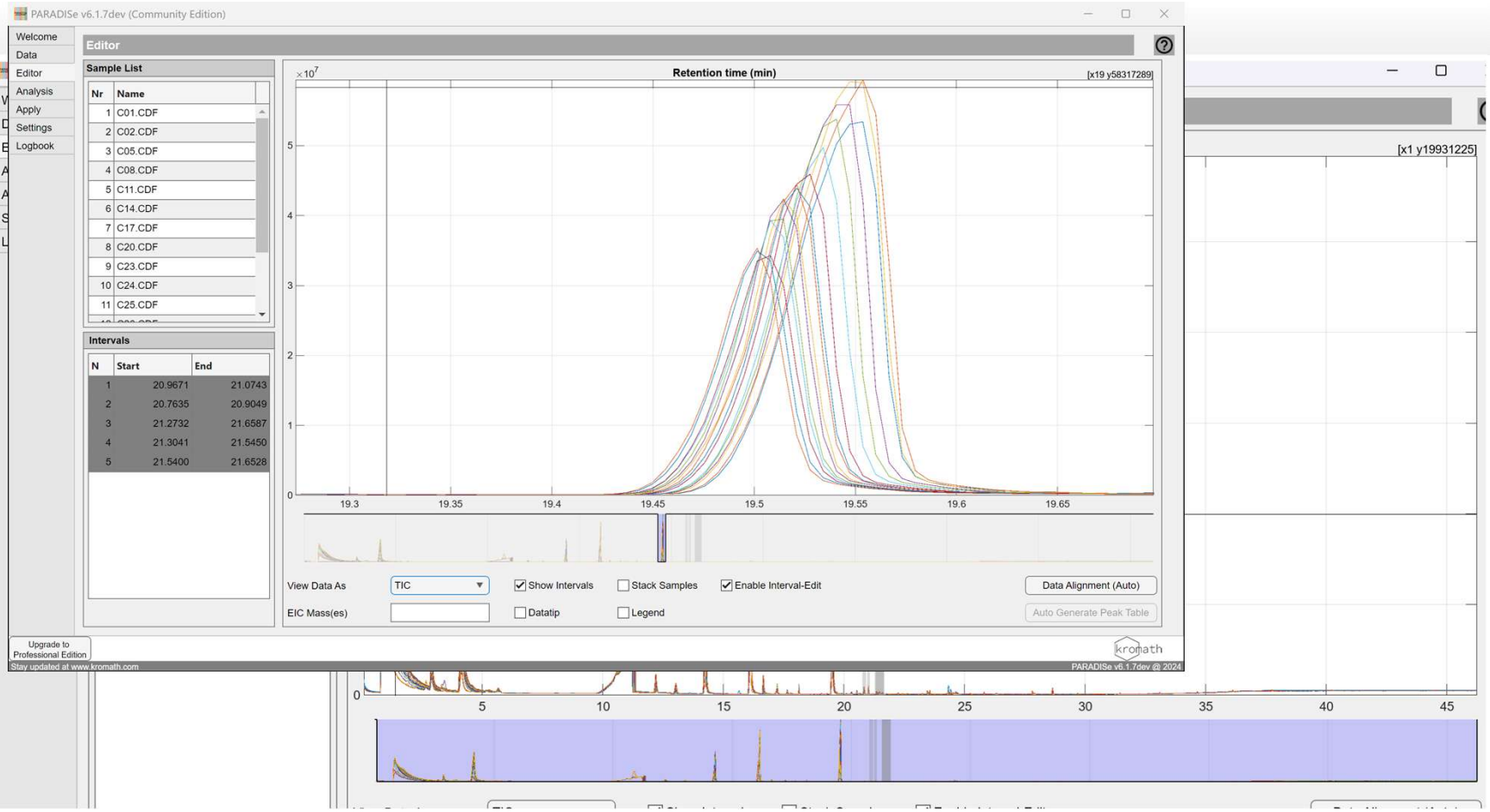
In our department we have gone from spending three man-year per year on GC-MS to 0.1!!

What we can do now



Cutting down runtime from 40 to 10 Minutes using ‘super-resolution’
Asta Haagensen

And more ….


PARADISe v6.1.7dev (Community Edition)
Welcome
Data
Editor
Analysis
Apply
Settings
Logbook
Editor
Sample List
Nr Name
1 C01.CDF
2 C02.CDF
3 C05.CDF
4 C08.CDF
5 C11.CDF
6 C14.CDF
7 C17.CDF
8 C20.CDF
9 C23.CDF
10 C24.CDF
11 C25.CDF
Intervals
N Start End
1 20.9671 21.0743
2 20.7635 20.9049
3 21.2732 21.6587
4 21.3041 21.5450
5 21.5400 21.6528
Retention time (min)
[x19 y58317289]
View Data As
TIC
Show Intervals
Stack Samples
Enable Interval-Edit
Data Alignment (Auto)
EIC Mass(es)
Datatip
Legend
Auto Generate Peak Table
Upgrade to Professional Edition
Stay updated at www.kromath.com
PARADISe v6.1.7dev @ 2024
[x1 y19931225]

And more ....

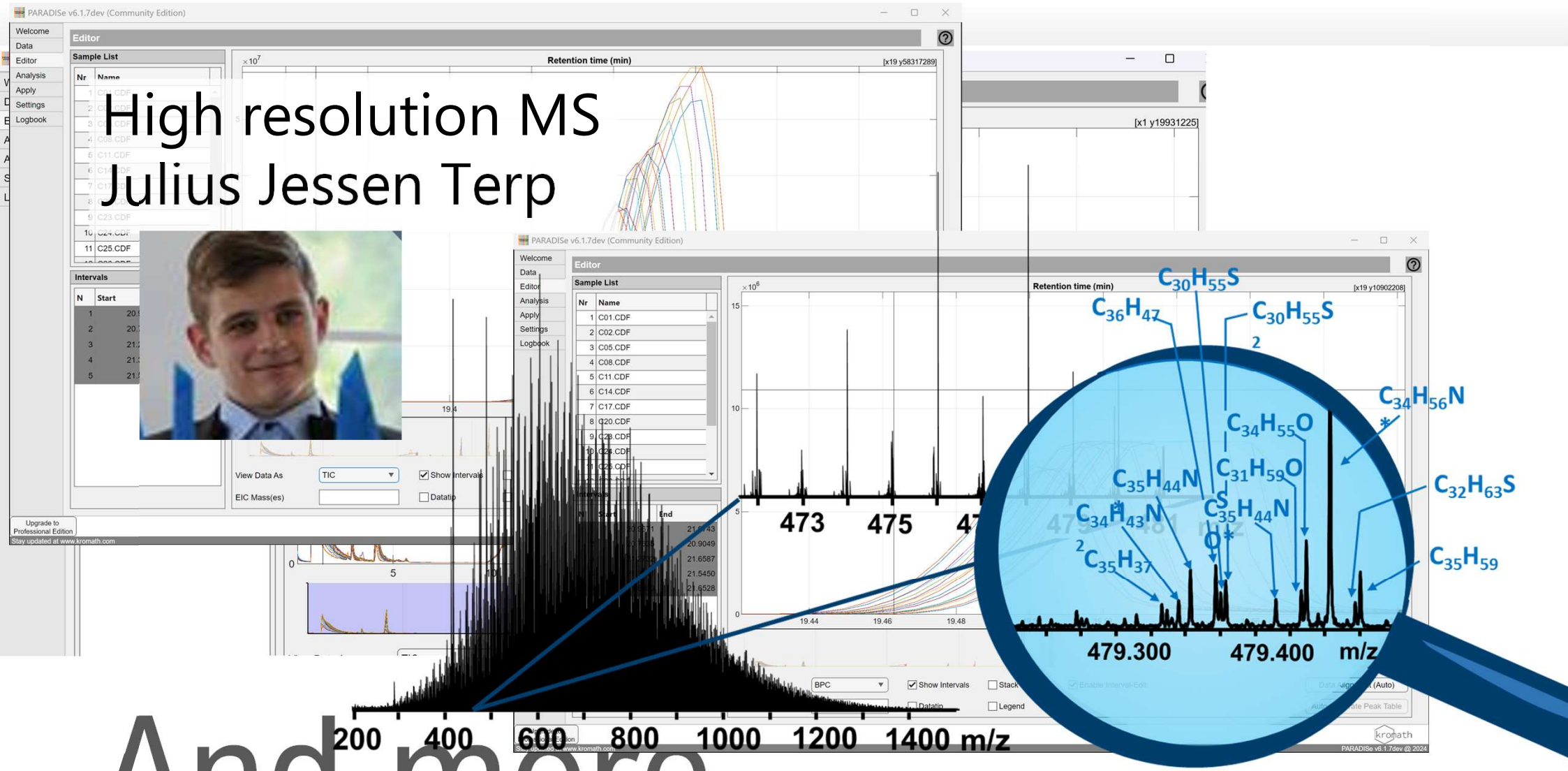
Handling saturation

Beatriz Quintanilla

And more ....

High resolution MS

Julius Jessen Terp



And more ....

Schneide, Gallagher, Bro, Shift invariant soft trilinearity: Modelling shifts and shape changes in gas-chromatography coupled mass spectrometry, Chemometrics Intelligent Laboratory Syst, Volume 251, 2024,

Baccolo G, Yu H, Valsecchi C, Ballabio D, Bro R, Comparison of machine learning approaches for the classification of elution profiles, Chemometrics & Intelligent Laboratory Systems, 243, 2023, 10500

Huiwen Y, Bro R, Gallagher N, PARASIAS: A new method for analyzing higher-order tensors with shifting profiles, Analytica Chimica Acta, 2023, https://doi.org/10.1016/j.aca.2022.339848

Marie Roald, Schenker, Calhoun, Adali, Bro, Cohen, and Evrim Acar, et al. (2022). "An AO-ADMM Approach to Constraining PARAFAC2 on All Modes." SIAM Journal on Mathematics of Data Science 4(3): 1191

Huiwen Y, Bro R, Gallagher N, PARASIAS: A new method for analyzing higher-order tensors with shifting profiles, Analytica Chimica Acta, 2022

Baccolo, Quintanilla-Casas, Vichi, Augustijn, Bro (2021). "From untargeted chemical profiling to peak tables – A fully automated AI driven approach to untargeted GC-MS." Trends in Analytical C

Yu H, Bro R, PARAFAC2 and local mínima, Chemometrics and Intelligent Laboratory Systems 219 (2021) 104446

Yu, H., Augustijn, D., Bro, R., Accelerating PARAFAC2 algorithms for non-negative complex tensor decomposition, Chemometrics and Intelligent Laboratory Systems, 2021, 214, 104312.

Risum A. B., Bro R., Using deep learning to evaluate peaks in chromatographic data, Talanta, (2019), 204, 255-260

K. Tiana, L. Wu, S. Min, R. Bro, Geometric search: A new approach for fitting PARAFAC2 models on GC-MS data, Talanta 185 (2018) 378–386

Cohen, Bro (2018) Nonnegative PARAFAC2: A Flexible Coupling Approach. Latent Variable Analysis and Signal Separation. LVA/ICA 2018. Springer, Cham.

Petersen, M. A. and R. Bro. PARADISe – a ground-breaking tool to treat complex GC-MS datasets. Flavour Science. Proceedings of the XV Weurman Flavour Research Symposium (2017) 421-426

L. G. Johnsen, P. B. Skou, B. Khakimov, and R. Bro. Gas chromatography – mass spectrometry data processing made easy. Journal of Chromatography A 1503:57-64, 2017.

L. G. Johnsen, J. M. Amigo, T. Skov, and Rasmus Bro. Automated resolution of overlapping peaks in chromatographic data. J.Chemom. 28:71-82, 2014.

J. M. Amigo, T. Skov, and R. Bro. ChroMATHography: Solving Chromatographic Issues with Mathematical Models and Intuitive Graphics. Chemical Reviews 110:4582-4605, 2010.

J. M. Amigo, T. Skov, J. Coello, S. Maspoch, and R. Bro. Solving GC-MS problems with PARAFAC2. Trends in Analytical Chemistry 27 (8):714-725, 2008.

Thomas Skov and R. Bro. Solving fundamental problems in chromatographic analysis. Analytical and Bioanalytical Chemistry 390:281-285, 2008.

Thomas Skov, F. van den Berg, G. Tomasi, and R. Bro. Automated alignment of chromatographic data. J.Chemom. 20:484-497, 2007.

H. A. L. Kiers, J. M. F. ten Berge, R. Bro, PARAFAC2 – Part I. A direct fitting algorithm for the PARAFAC2 model, J. Chemom., 13, 275-294, 1999

R. Bro, H. A. L. Kiers, C. A. Andersson, PARAFAC2 – Part II. Modeling chromatographic data with retention time shifts, J. Chemom., 13, 295-309, 1999

R. Bro, S. de Jong, A fast non-negativity constrained least squares algorithm, J. Chemom., 1997, 11, 393-401

Download: www.kromath.com

More software: ucphchemometrics.com

Where to get it