

02-TriDAMP-ThreeGroups

This notebook demonstrates the usage of the *TriDAMP* package for three-way comparison of three groups of replicate datasets. The notebook is based on *MathDAMP*'s 06-MathDAMP-MultipleGroups.nb notebook. The *MathDAMP*'s functionality is used to normalize the datasets and calculate the F ratio (one-way ANOVA) result. The averages of the three groups's datasets are used for the three-way comparison. The F ratio result is then used to filter out statistically less significant differences from the three-way comparison. The *MathDAMP*'s functionality will not be commented in this notebook, for a brief explanation please refer to *MathDAMP*'s 06-MathDAMP-MultipleGroups.nb notebook. For a more detailed explanation of *TriDAMP*'s functionality, please refer to *TriDAMP*.nb (available upon request) and 01-TriDAMP-Basics.nb notebooks.

Step 1: Loading the Data

First, the *MathDAMP* and *TriDAMP* packages have to be loaded. Please adjust the paths to package files according to their location on your system. Due to the size of the datasets and results the global variable `$HistoryLength` is set to 1 to save memory. 1.5 GB of physical memory may be necessary to execute this notebook.

```
$HistoryLength = 1;
<< "/home/baran/math/ms/MathDAMP.1.0.0/MathDAMP.m"
<< "/home/baran/math/ms/MathDAMP.1.0.0/TriDAMP.m"

MathDAMP version 1.0.0 loaded (2006/04/26)

This program is distributed in the hope that it will be useful, but WITHOUT
ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

TriDAMP version 1.0.0 loaded (2006/05/09)

This program is distributed in the hope that it will be useful, but WITHOUT
ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```

Datasets acquired by capillary electrophoresis coupled to a time-of-flight mass spectrometer (CE-TOFMS) will be used for the demonstration in this notebook. The *.bdt datafiles were generated using a separate in house software from *.csv datafiles exported by the Analyst QS software. The original data were binned to a 0.02 m/z units resolution, baselines were subtracted from the individual electropherograms (as with the `DAMPSubtractBaselines` function with default options), noise was removed (as with the `DAMPRemoveNoise` function with default options), the data were binned to 1 m/z units resolution along the m/z axis and saved in a binary format as *.bdt datafiles.

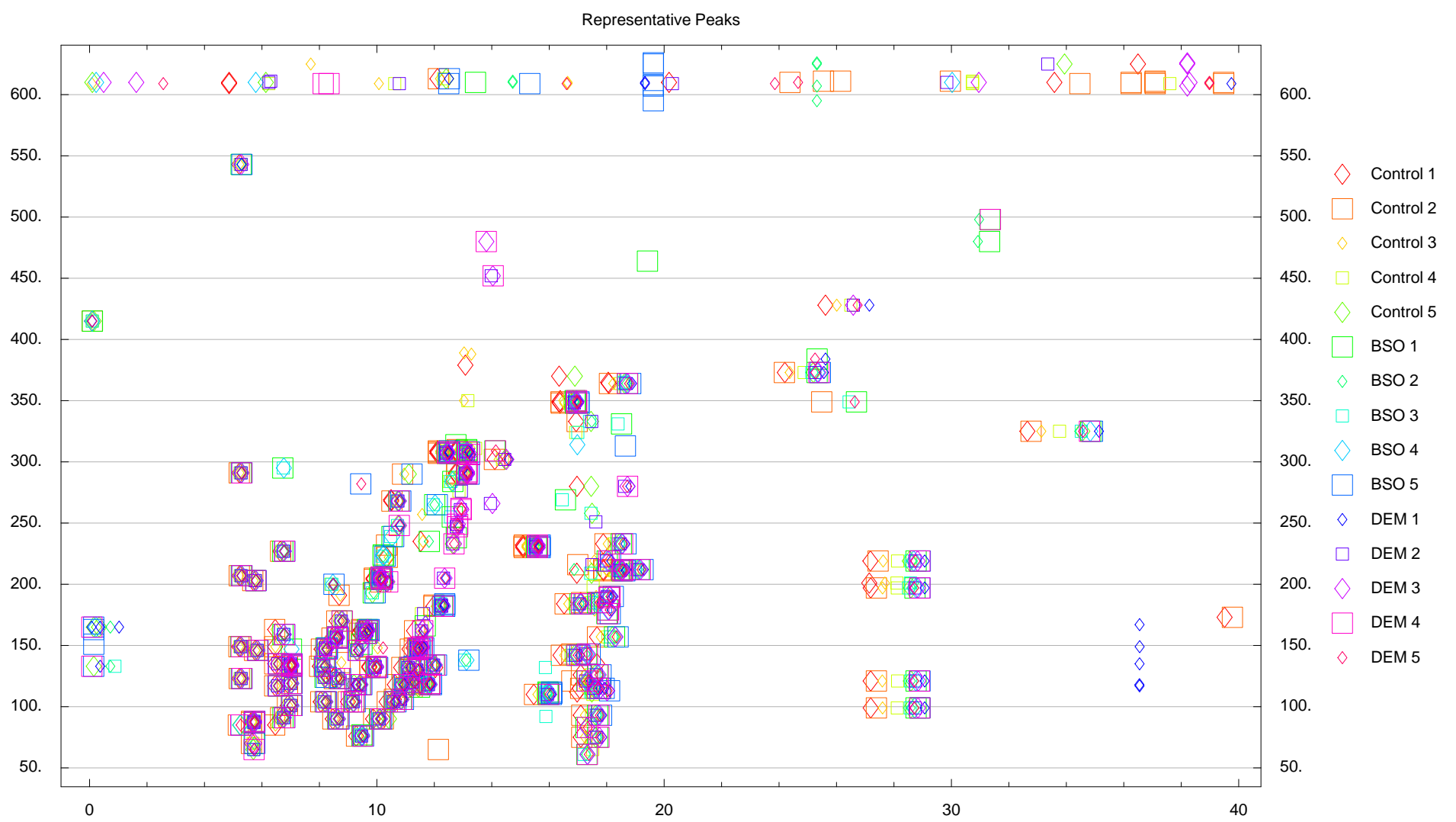
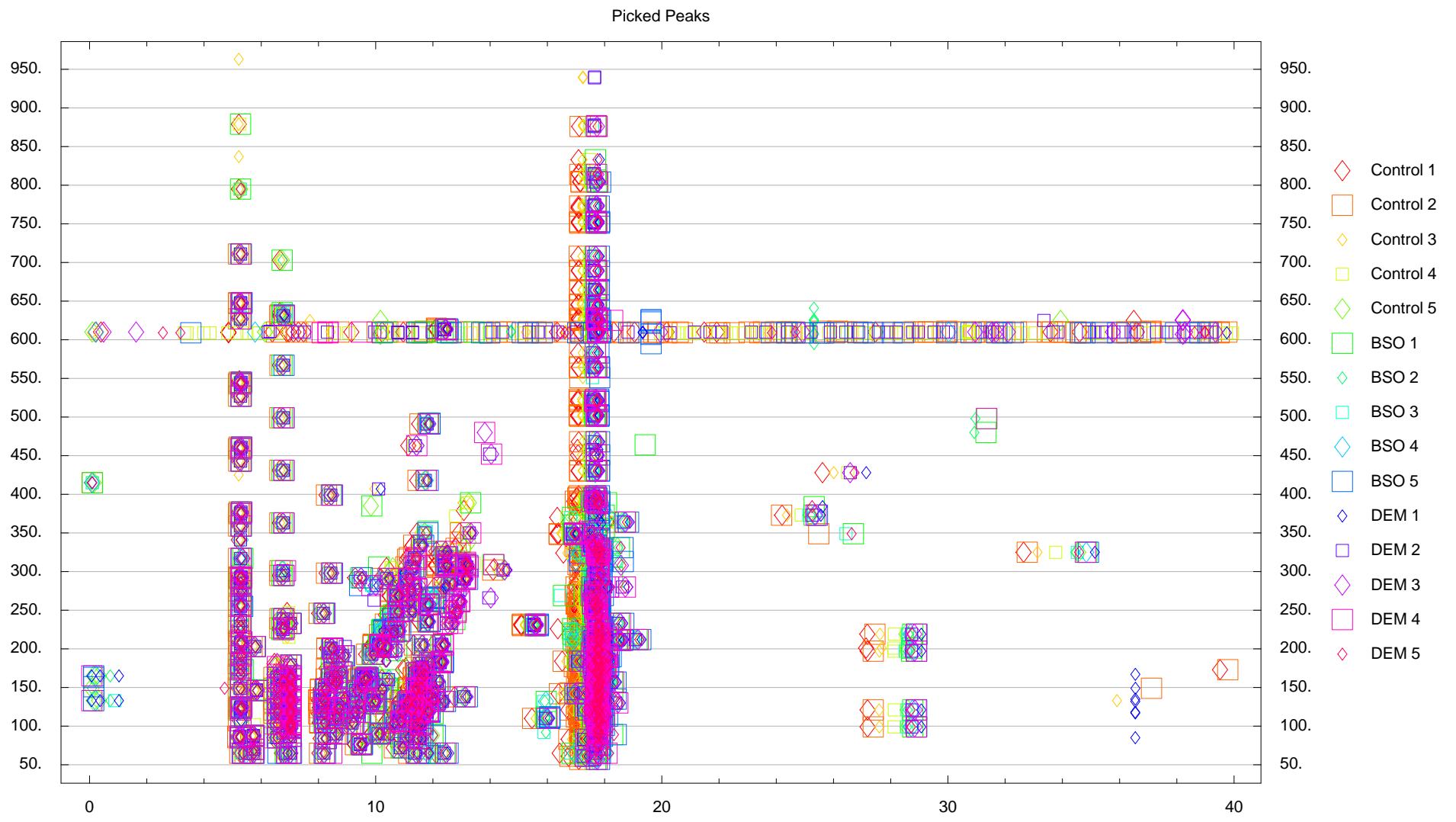
```
fnames = Partition[FileNames["/home2/baran/data/liver - ext/*.bdt"], 5][[2, 1, 3]];
replicates = If[Equal@@(Length[#] & /@ fnames), Length[fnames[[1]],
  Print[StyleForm["Error: Number of replicates not identical within groups!", FontColor -> Hue[0]]]; 0];
data = DAMPImportBDT[#[[1]], #[[2]]] & /@ Transpose[{Join@@fnames,
  Join@@(Table[# <> " " <> ToString[i], {i, replicates}] & /@ {"Control", "BSO", "DEM"})]};
NumberForm[MemoryInUse[], DigitBlock ->
  3]

450,845,672
```

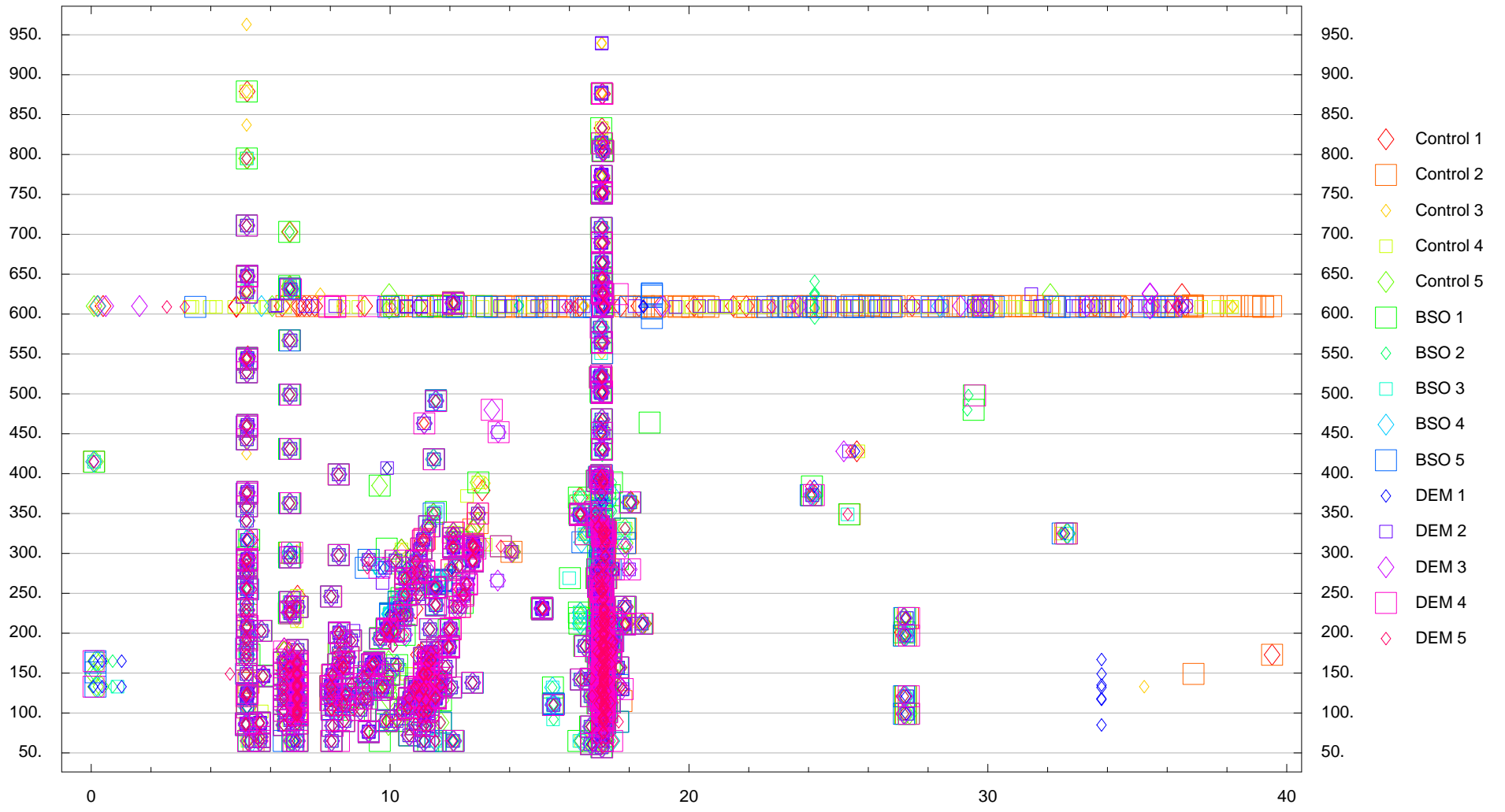
Step 2: Performing the Differential Analysis

The normalization of datasets is done in the same way as in *MathDAMP*'s comparison of multiple groups of replicate datasets (please refer to the 06-MathDAMP-MultipleGroups.nb notebook for description).

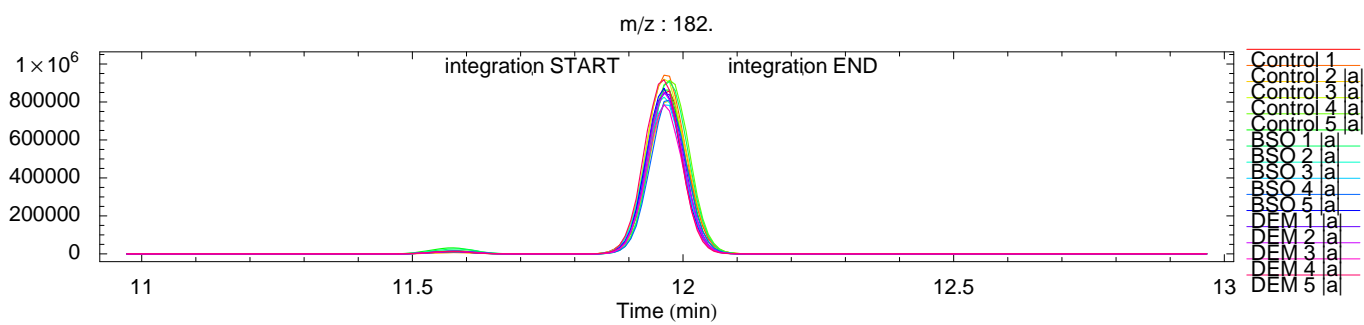
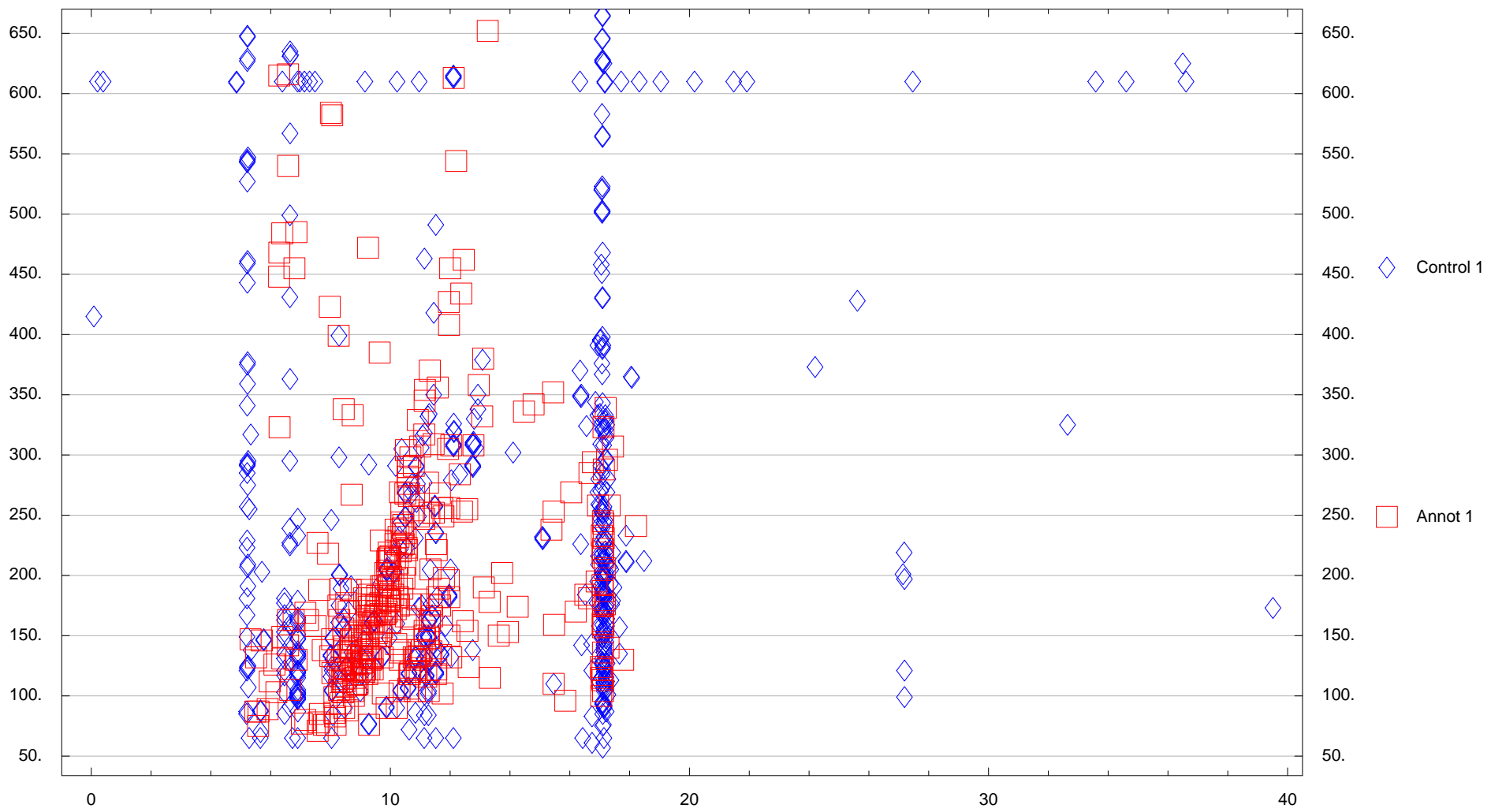
```
Clear[rslt];
rslt =
  DAMPMultiGroups[data, replicates, NormalizeGroupOptions -> {Reference -> 1, InternalStandard -> 363, AutoISIntegrationVicinity -> {-.25, .25},
  PeakPickingOptions -> {Threshold -> 4000}, RepresentativePeakOptions -> {PeaksPerChromatogram -> 5, PeaksPerInterval -> 5, IntervalSize -> .5},
  AlignmentTimeRange -> All, FitShiftFunctionOptions -> {GapPenalty -> {3, .5}}, PeakLayoutPlotOptions -> DAMPCETOFMSPeakLayoutOptions,
  AnnotationTables -> {DAMPLoadAnnotationTable["/home/baran/math/ms/MathDAMP.1.0.0/iab_cems_cation.csv"]},
  OutputTimeRange -> {4, 24}, ExternalNormalizationCoefficients -> None, GroupNames -> {"Control", "BSO", "DEM"}];
```



Aligned Peaks



Aligned Annotation Tables



IS normalization coefficients : {1., 0.931947, 1.07215, 0.970092, 0.972815, 1.06909, 1.01094, 1.09095, 1.14835, 1.17723, 1.11306, 1.08559, 1.12947, 1.23681, 1.03703}

Step 3: Three-way comparison, Exploring the Results, Listing the Candidates

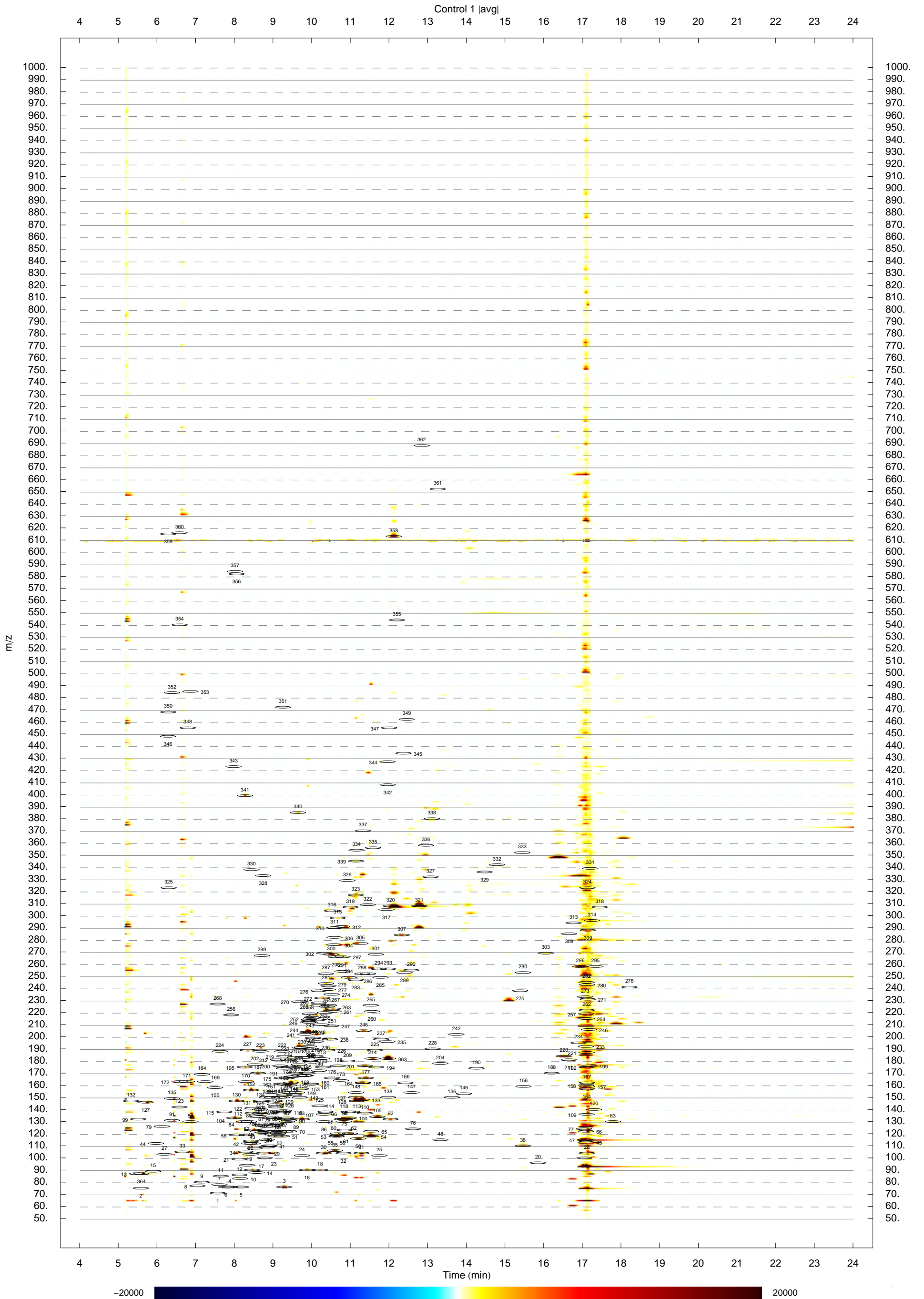
Averaged datasets are calculated for all three groups of replicate datasets which were normalized in the previous step. The averaged datasets are used to calculate the absolute, relative and absolute×relative three-way comparison results. The F ratio result dataset from previous section is used to filter the absolute×relative three-way comparison dataset (the F ratio threshold 3.9 corresponds to $P=0.05$ for comparisons of 3 groups of 5 replicates).

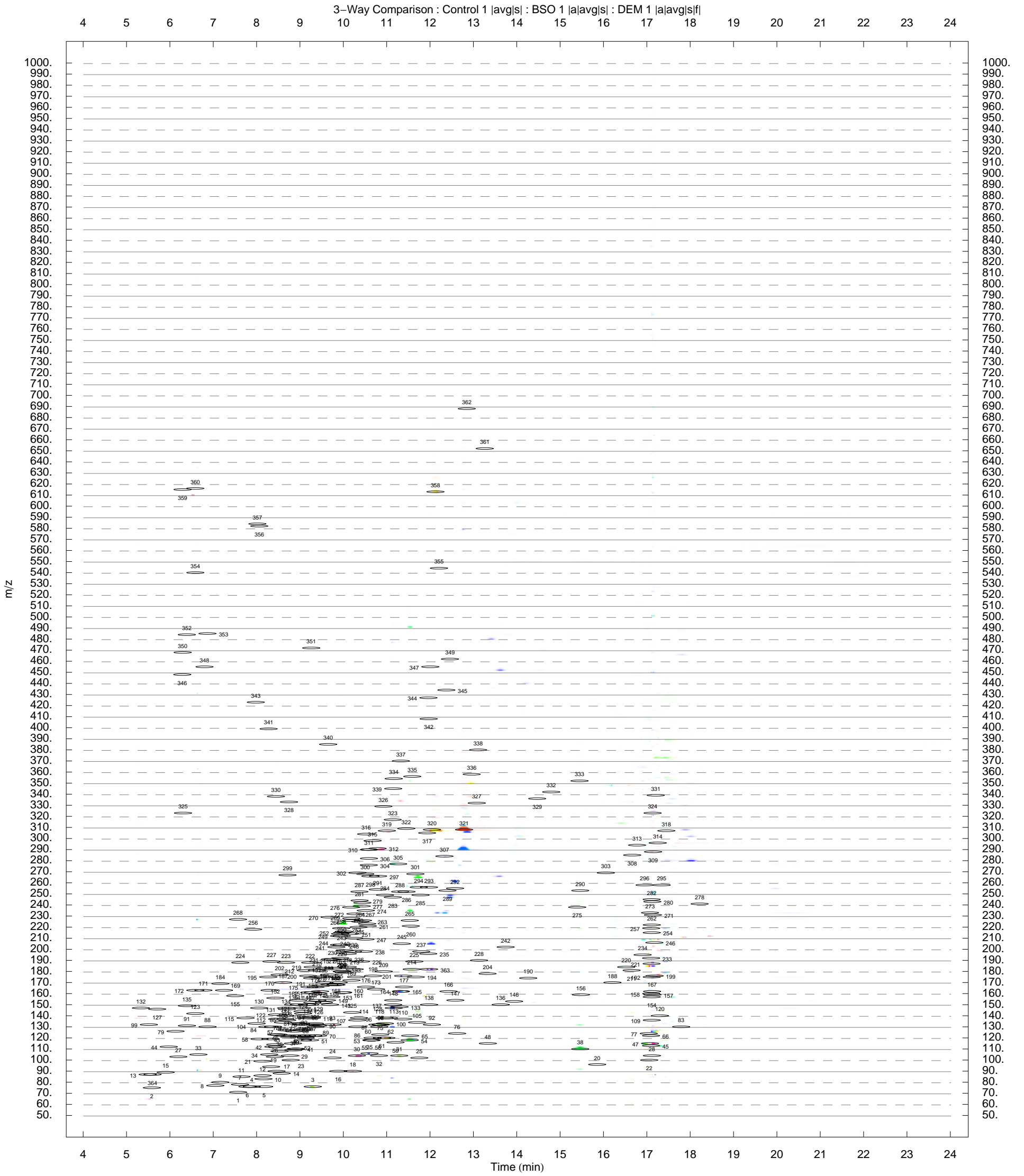
```
Clear[data]
avgdata = DAMPApplyFunctionToGroup[#, Mean, SampleNameSuffix -> "avg"] & /@ Partition[NormalizedDatasets /. rslt, replicates];
abs3way = TriDAMPCompare[avgdata];
rel3way = TriDAMPCompare[DAMPSmooth[#] & /@ avgdata, Relative -> True];
absrel3way = rel3way;
absrel3way[[1, All, All, 2]] *= abs3way[[1, All, All, 2]];
filtabsrel3way = TriDAMPFilter[absrel3way, DAMPSmooth[FRatios /. rslt], 3.9];
NumberForm[MemoryInUse[], DigitBlock -> 3]

986,624,248
```

Below is a visualizations for one of the averaged datasets and a visualization of the filtered absolute×relative three-way comparison.

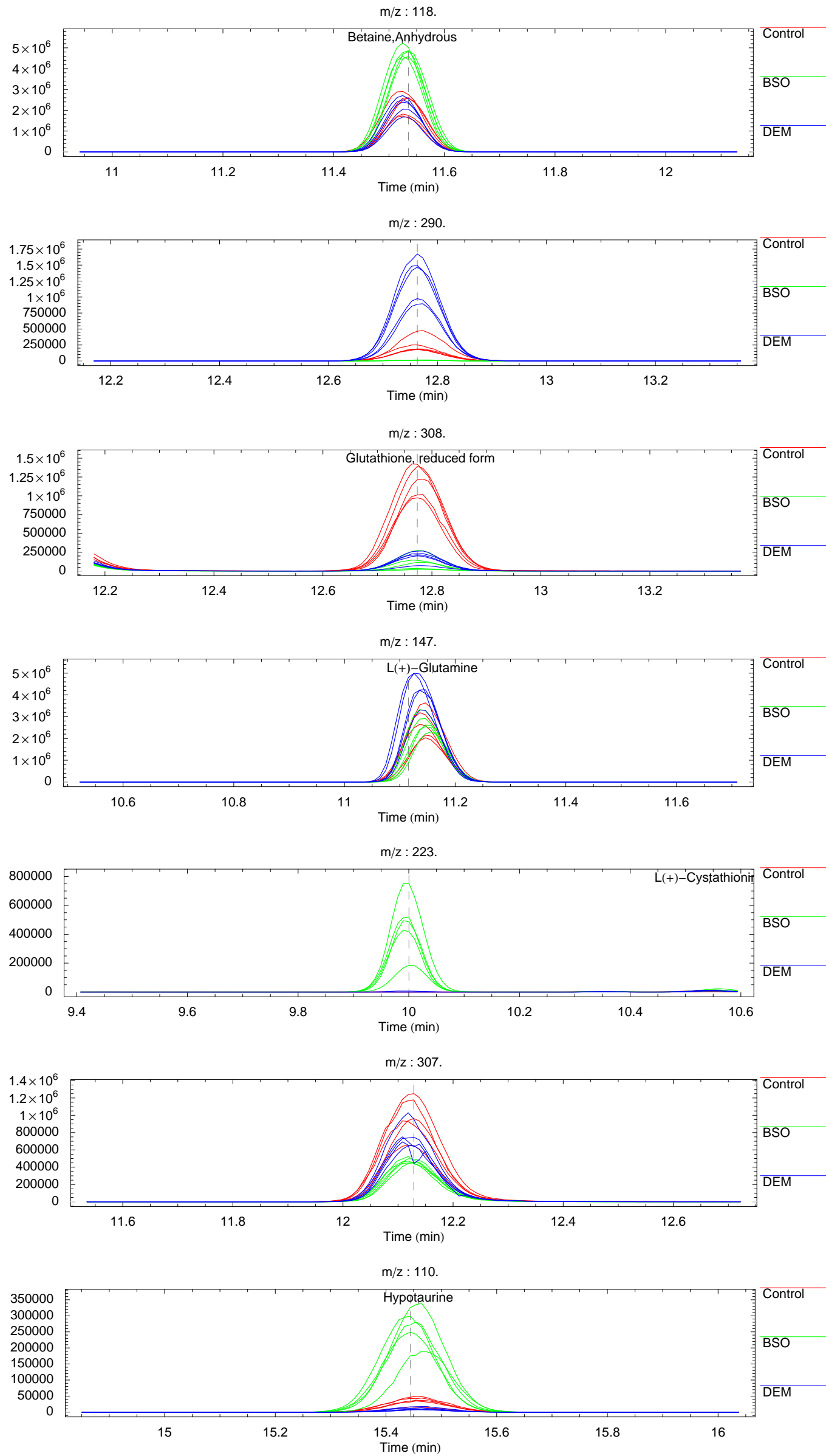
```
DAMPDensityPlot[avgdata[[1]], MaxScale -> 20000, AnnotationTables -> (AlignedAnnotationTables /. rslt), Sequence @@ DAMPCETOFMSDensityPlotOptions];
TriDAMPDensityPlot[filtabsrel3way, IntensityScale -> {0, 10000},
  ScaleModifier -> (#1/3 &), LegendPosition -> "Bottom", LegendOptions -> {Identifiers -> (GroupNames /. rslt)},
  AnnotationTables -> (AlignedAnnotationTables /. rslt), Sequence @@ DAMPCETOFMSDensityPlotOptions];
```

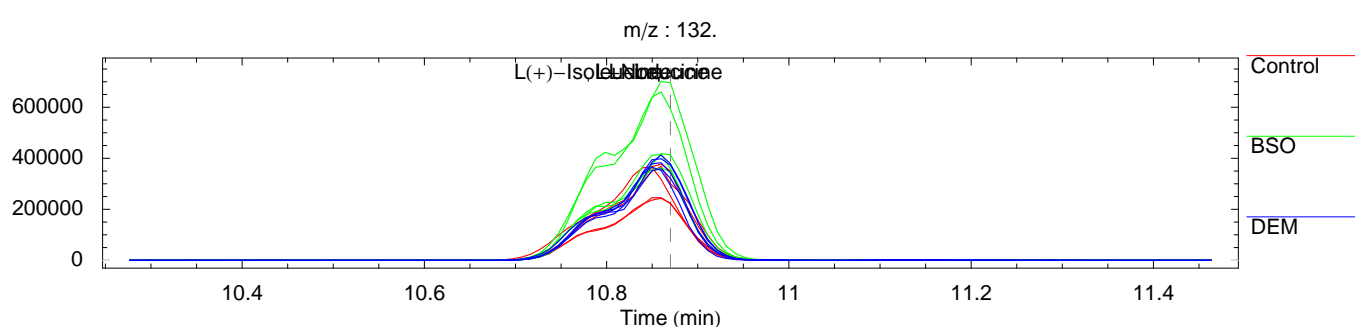
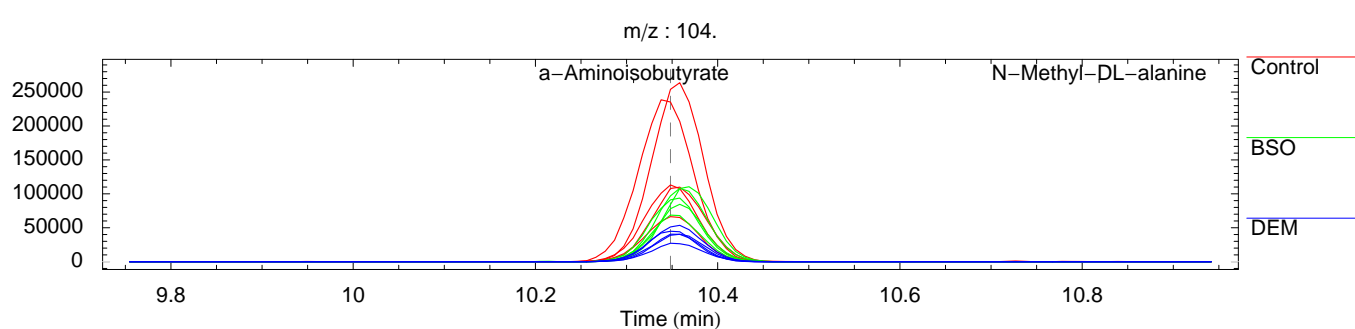
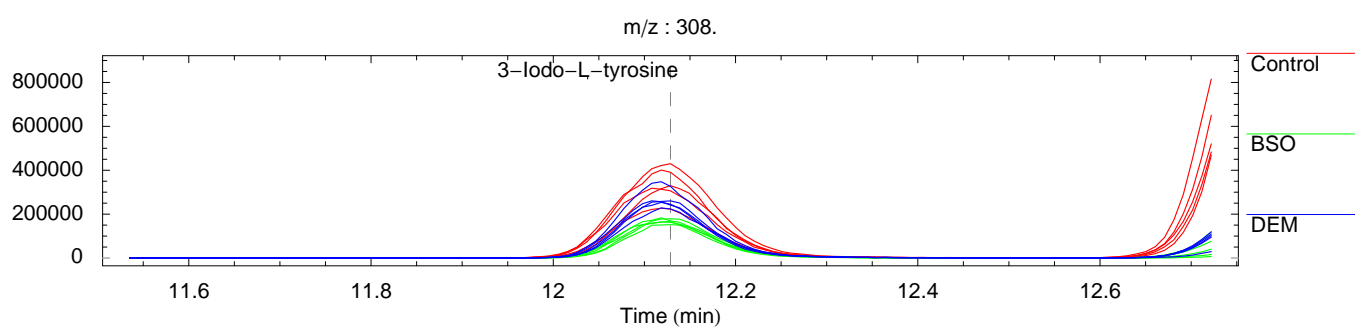
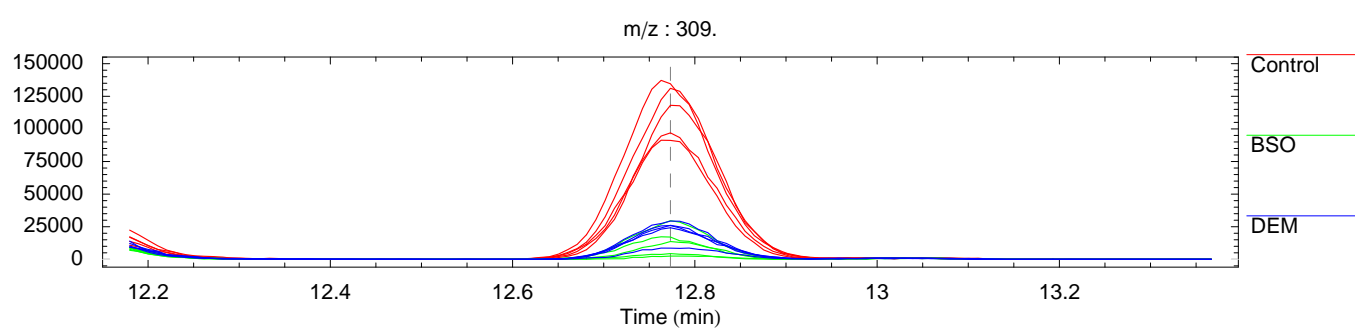
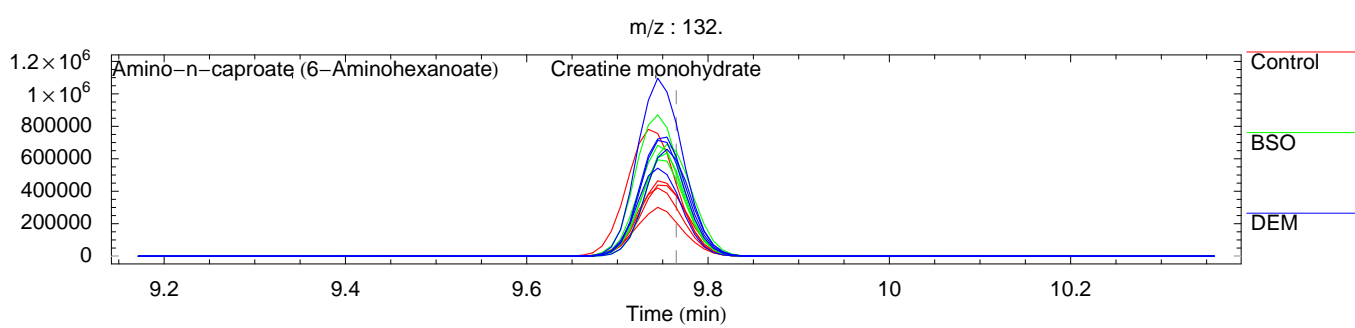
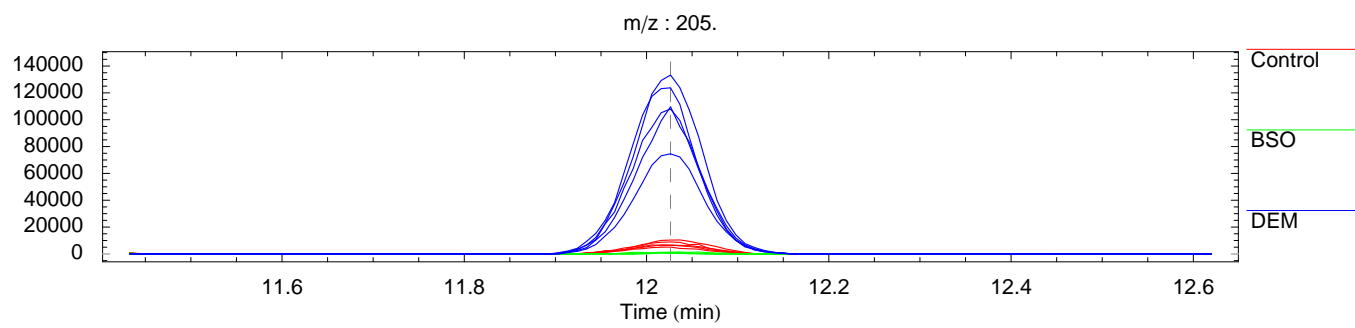
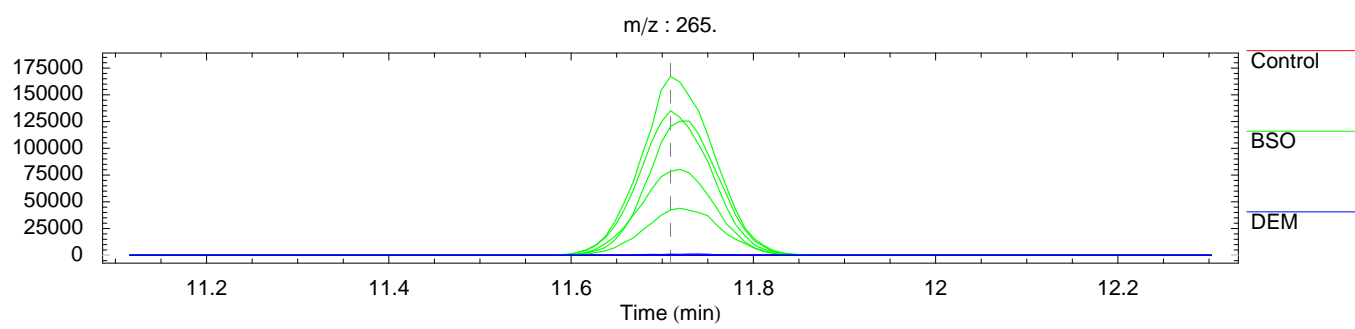
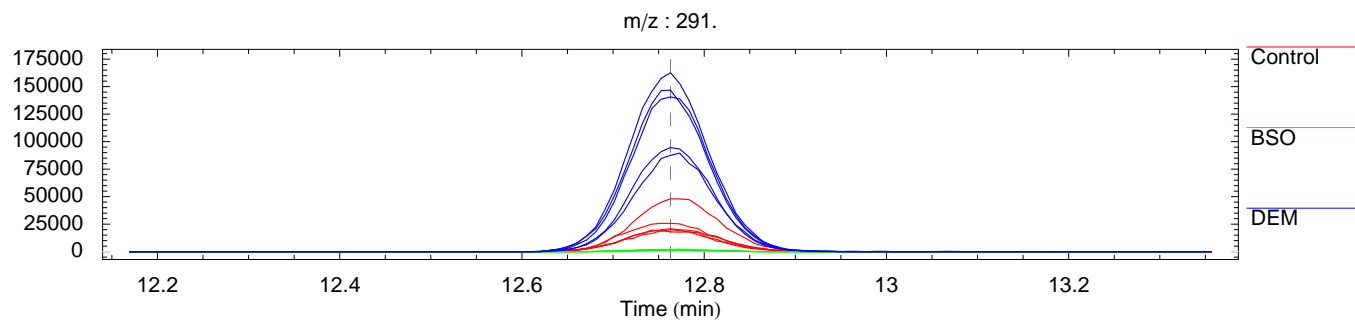


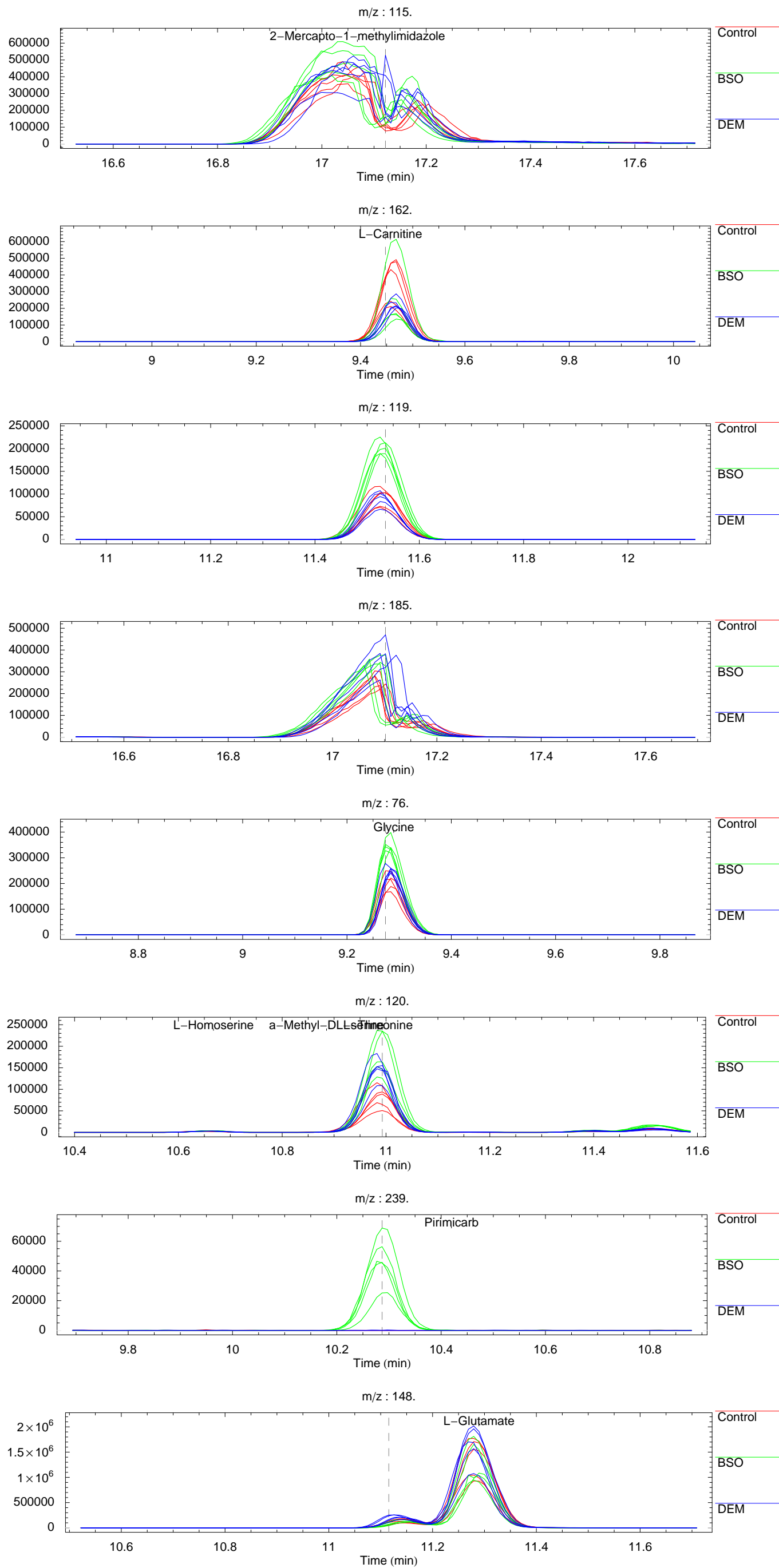


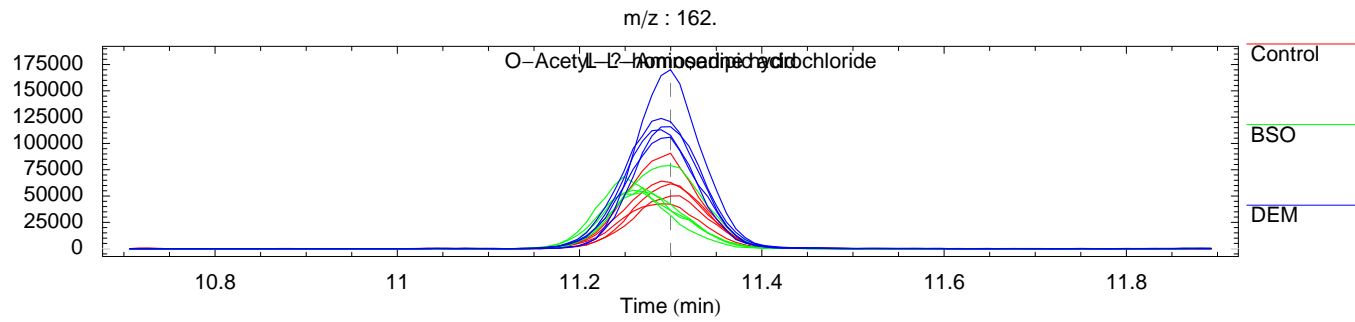
Overlaid electropherograms corresponding to the most significant differences from any result can be generated in descending order of significance. Below are electropherograms corresponding to the 24 most significant differences from the filtered absolute \times relative three-way comparison result.

```
TriDAMPPlotCandidates[(NormalizedDatasets /. rslt), filtabsrel3way, PlotCount -> 24, TimeRange -> All,
PlotChromatogramOptions -> {PlotOptions -> {PlotStyle -> (Join@@ (Table[#, {replicates}] & /@ {Hue[0], Hue[1/3], Hue[2/3]}))},
LegendData -> Transpose[{{Hue[0], Hue[1/3], Hue[2/3]}, (GroupNames /. rslt)}], AnnotationTable -> (AlignedAnnotationTables /. rslt)[[1]]};
```



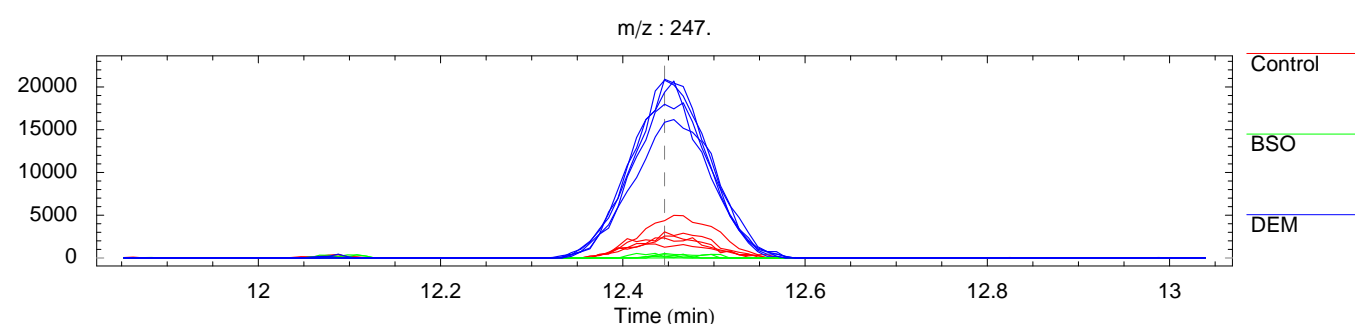
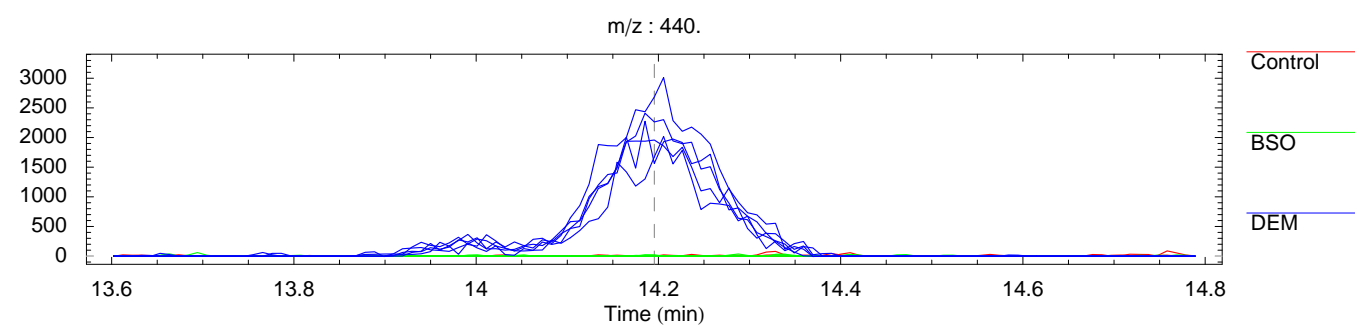
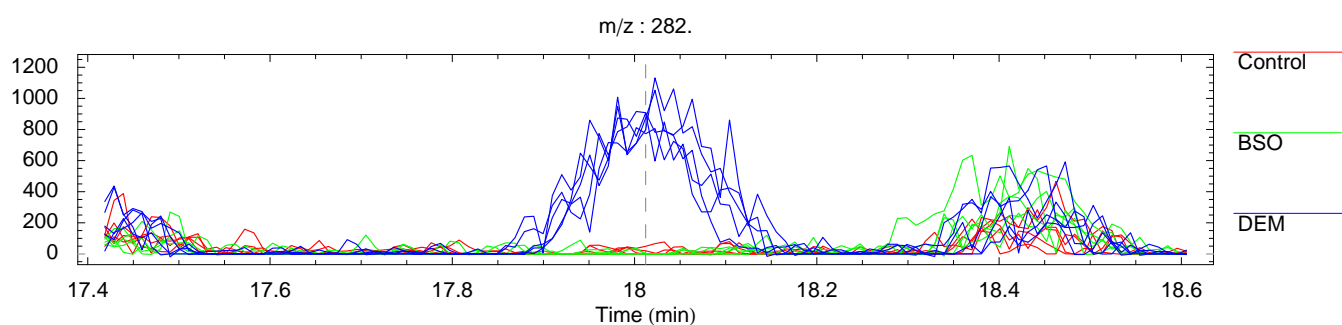
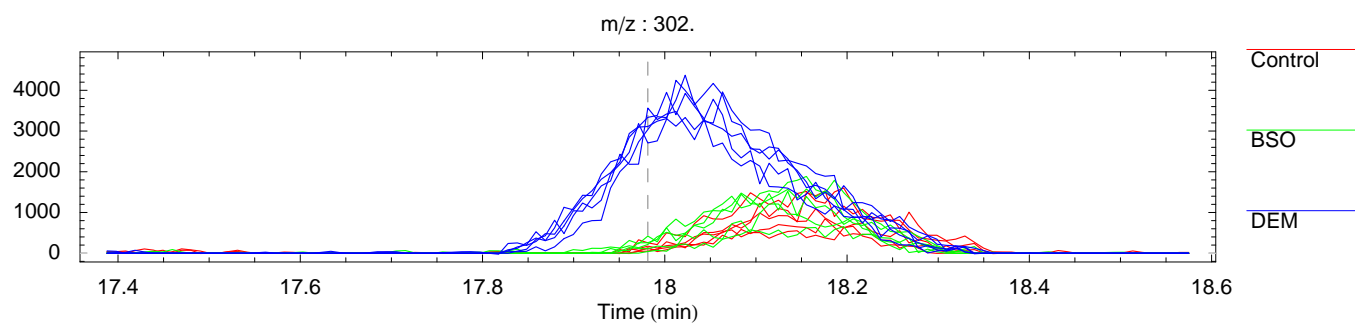
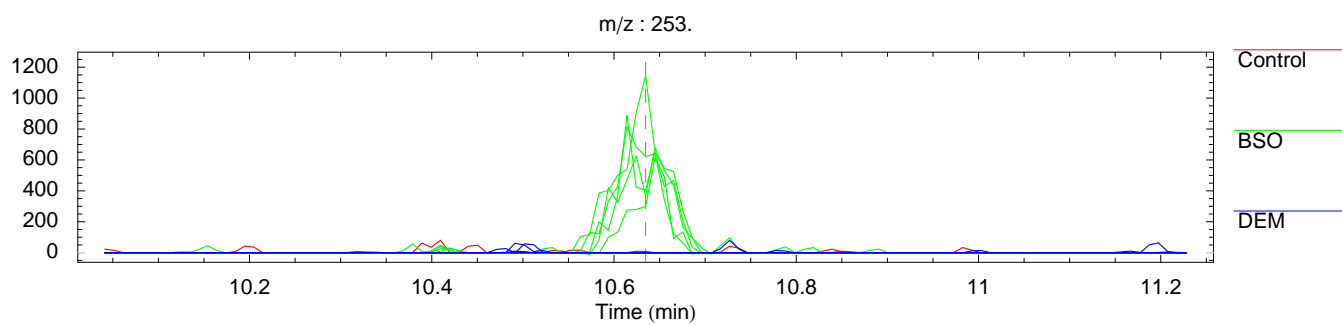
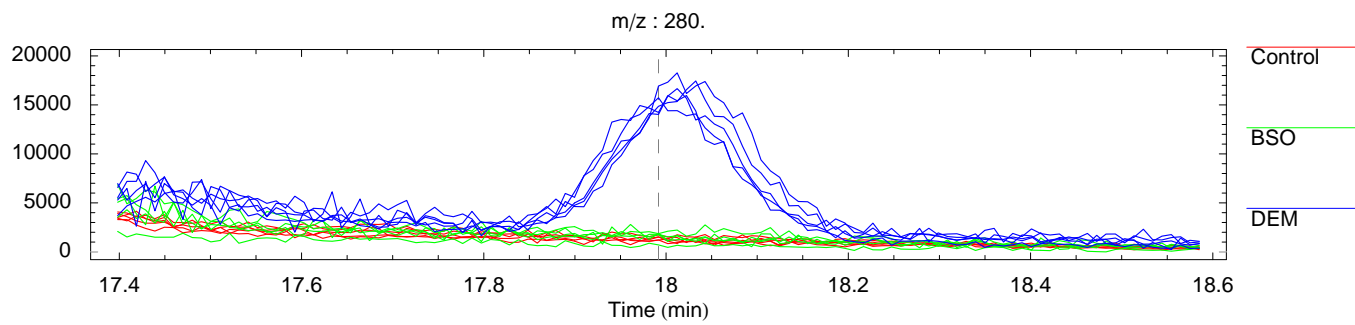


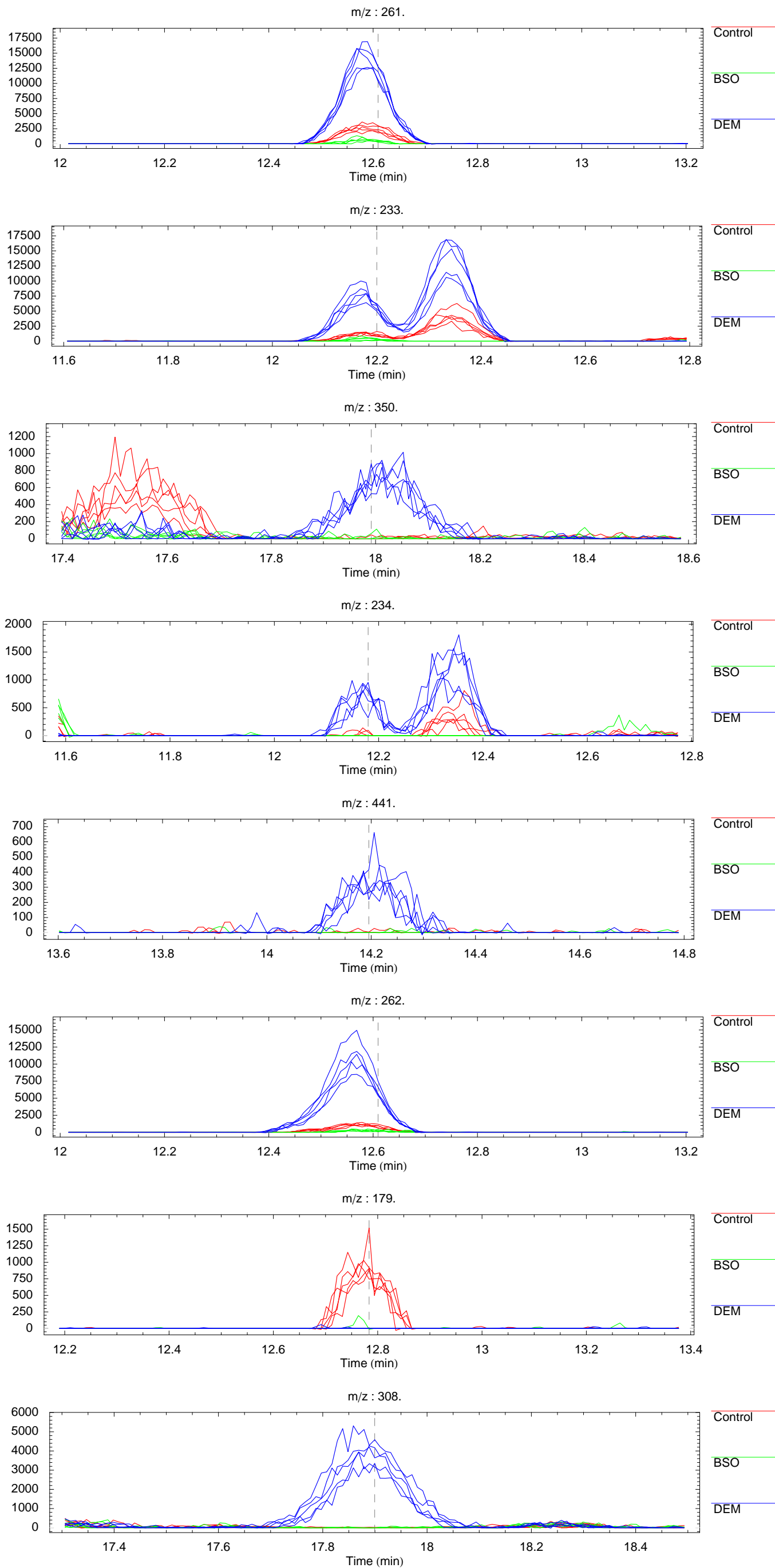


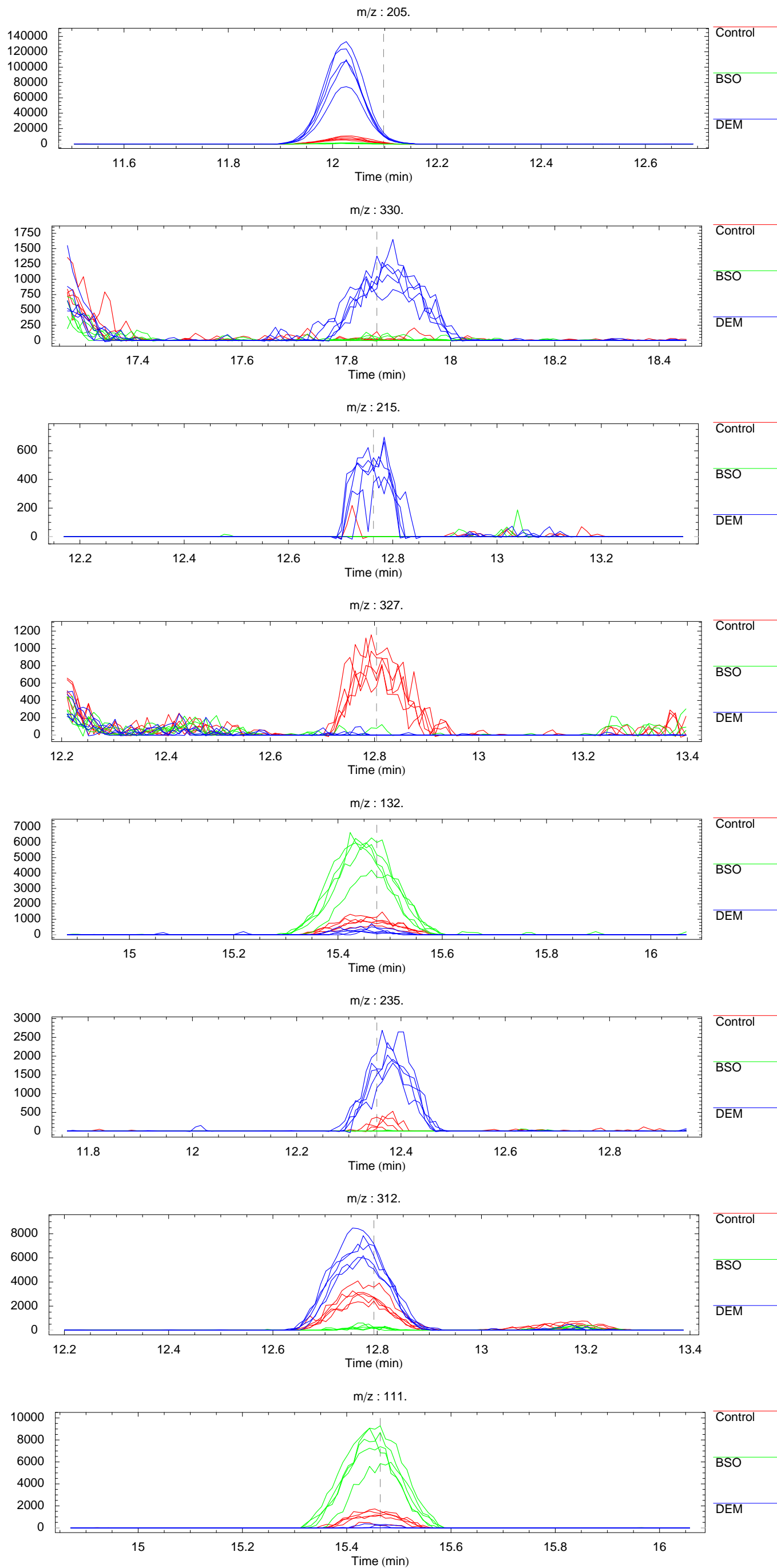


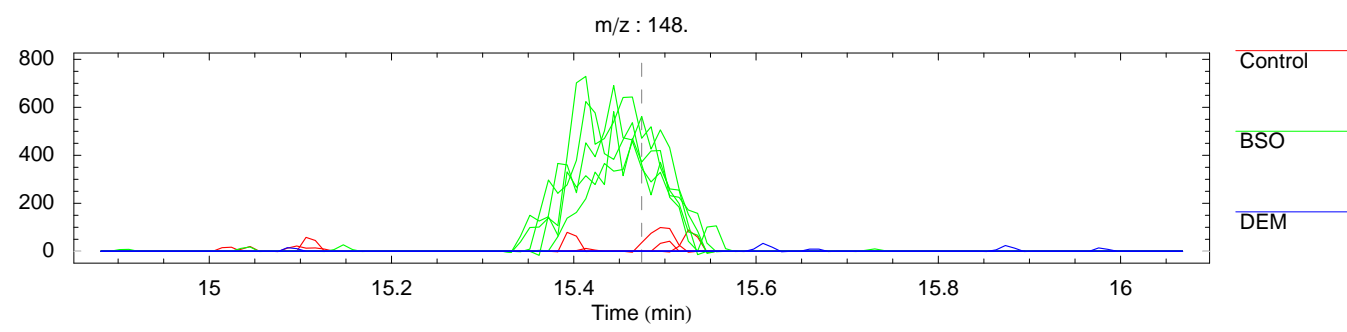
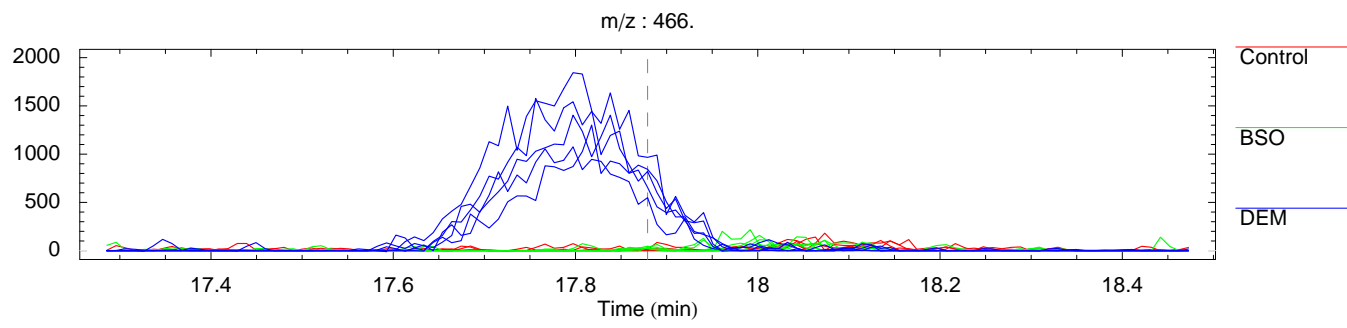
The candidate differences may be ranked according to the F ratio result as in *MathDAMP*'s comparison of multiple groups of replicate datasets. As discussed in the 05-MathDAMP-Two-Groups.nb notebook, different results may have different strength and weaknesses, so it may prove beneficial to generate the lists of candidates according to multiple results to minimize the possibility of missing an important difference if it ranks lower in one particular result.

```
plotcolors = {Hue[0], Hue[1 / 3], Hue[2 / 3]};
DAMPPlotCandidates[NormalizedDatasets /. rslt, DAMPCrop[DAMPsmooth[FRatios /. rslt], mzRange -> {50, 1000}, TimeRange -> {4, 28}],
  PlotCount -> 24, PlotChromatogramOptions -> {PlotOptions -> {PlotStyle -> (Join@@(Table[#, {replicates}] & /@plotcolors))},
  LegendData -> Transpose[{plotcolors, (GroupNames /. rslt)}], AnnotationTable -> (AlignedAnnotationTables /. rslt)[[1]]];
```









`NumberForm[MaxMemoryUsed[], DigitBlock -> 3]`

1,257,515,552