03−MathDAMP−TwoDatasets

The preceding notebook 02−MathDAMP−Elements.nb introduced the basic functionality of the *Math*DAMP package. The core functionality was demonstrated on a comparison of two datasets. This notebook provides a template for such comparison, the core functions are however wrapped into the `DAMPTwoDatasets` function for a more convenient use. The same two datasets used in the previous notebook will be used here as well.
Additional notebooks from the *Math*DAMP package (04−MathDAMP−Outliers.nb, 05−MathDAMP−TwoGroups, and 06−MathDAMP−MultipleGroups.nb) provide templates for locating outliers within a group of datasets, for the comparison of two groups of replicate datasets, and for the comparison of multiple groups of replicate datasets.

Step 1 : Loading the Data

First, the *Math*DAMP package has to be loaded. Please assign the path leading to the *Math*DAMP files to the `MathDAMPPath` variable.

```
MathDAMPPath = "/home/baran/math/ms/MathDAMP.1.0.0/";
<< (MathDAMPPath <> "MathDAMP.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.
```

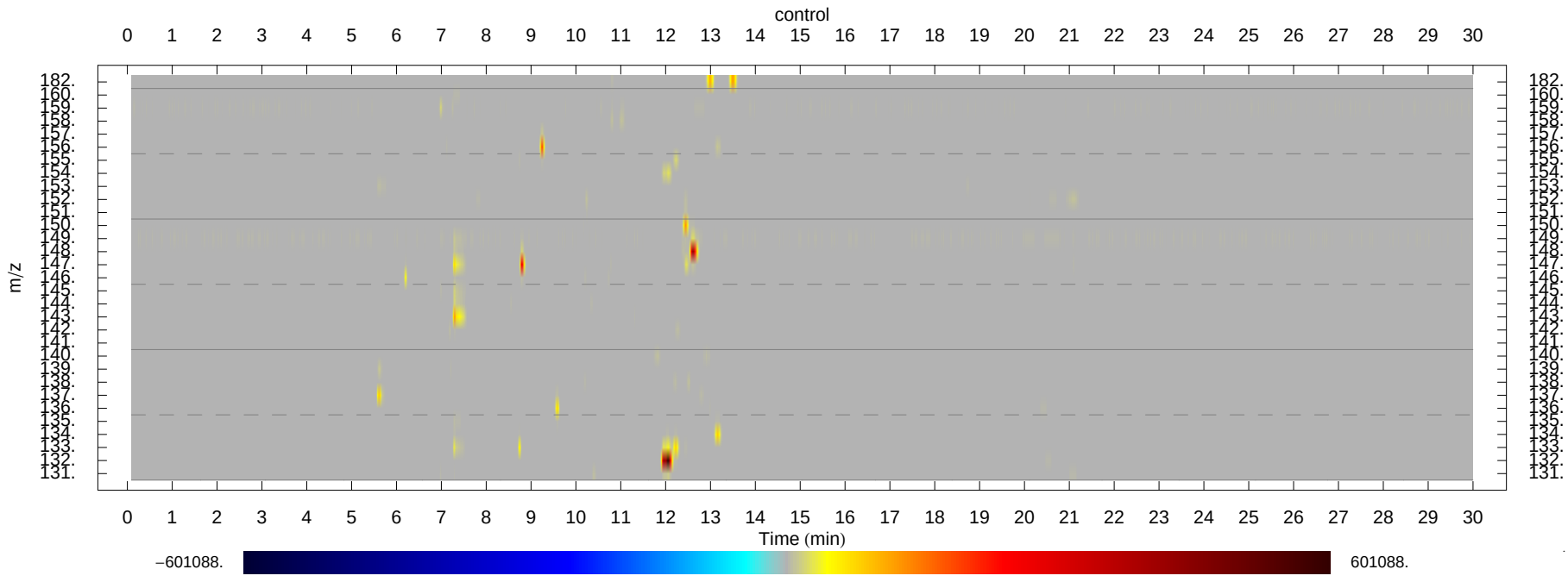
Two datasets acquired by capillary electrophoresis coupled to a quadrupole mass spectrometer (CE−QMS) operated in selected ion monitoring mode (SIM) are used for the demonstration in this notebook. The datafiles are part of the *Math*DAMP package.

```
{ctrl, smpl} = DAMPImportMS[MathDAMPPath <> "/data/" <> #] & /@ {"control.ms", "sample.ms"};
```

■ Optional : Exploring the data, locating the peak of the internal standard in the reference dataset

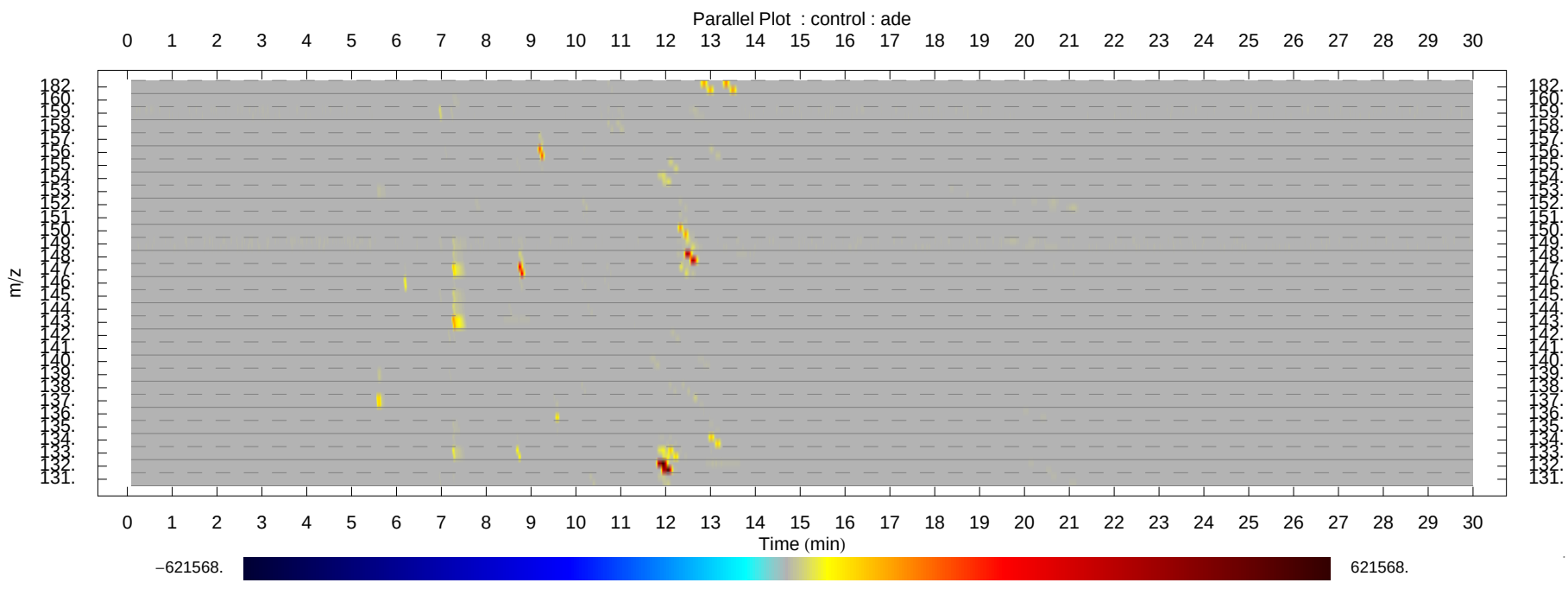
For a preliminary exploration, the loaded data may be visualized on density plots either one by one...

```
DAMPDensityPlot[ctrl];
```



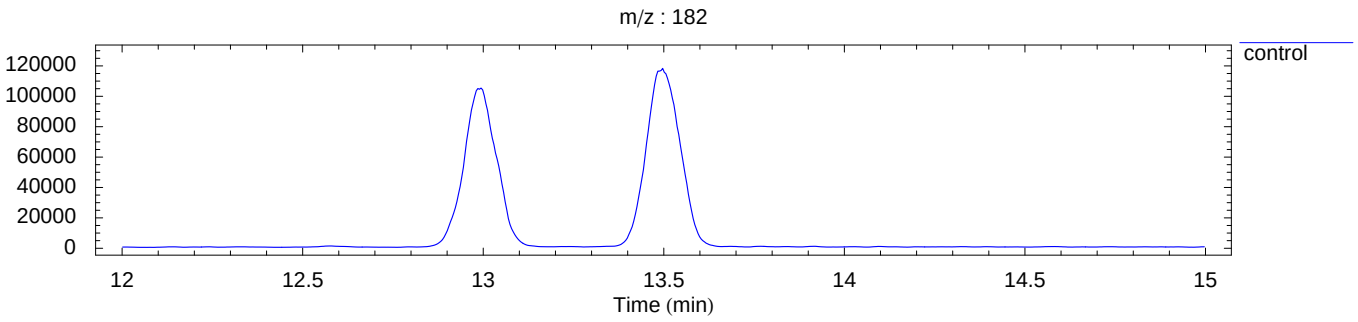
...or using a parallel plot. The parallel plot may provide a useful look at the data even if the datasets are misaligned. The time axis refers to the first dataset in the list submitted to the `DAMPParallelPlot` function as discussed in more detail in notebooks 02−MathDAMP−Elements.nb and MathDAMP.nb.

```
DAMPParallelPlot[{ctrl, smpl}];
```



If the location of the peak of the internal standard will be specified explicitly, it is necessary to locate it in the reference dataset (dataset to which the remaining datasets will be aligned and normalized).

```
DAMPPlotChromatogram[{ctrl}, 182, PlotOptions → {PlotRange → {{12, 15}, All}}];
```



Step 2 : Performing the Differential Analysis

The function `DAMPTwoDatasets` compares two datasets and returns the aligned and normalized datasets, the absolute, relative, and absolute×relative differences along with the aligned annotation tables (as a list of rules). `DAMPNormalizeGroup` function is used internally to align and normalize the datasets along with the annotation tables. This function is also used internally by the functions `DAMPOutliers`, `DAMPTwoGroups`, and `DAMPMultiGroups`. The usage of these functions is demonstrated in the notebooks 04−MathDAMP−Outliers.nb, 05−MathDAMP−TwoGroups, and 06−MathDAMP−MultipleGroups.nb. Please refer to the MathDAMP.nb notebook for more details about the implementation of the functions `DAMPNormalizeGroup` and `DAMPTwoDatasets`. Execute ?*FunctionName* to list a brief description of the respective function's available options.

```
?DAMPNormalizeGroup
```

```
DAMPNormalizeGroup[msdatas,options] aligns msdatas (a list of datasets) and normalizes them according to the areas of the peaks of the internal standard and
  external normalization coefficients (optional). The results are returned as a list of rules: {NormalizedDatasets->...,AlignedAnnotationTables->...}
Options:
Reference - position of the reference dataset within msdatas to which the remaining msdatas will be normalized (default: 1)
AlignmentTimeRange - peak picks from the reference dataset falling within
  this timerange only (specified as {starttime,endtime} in minutes) will be used for alignment (default: All)
RepresentativePeakOptions - options to be passed to the DAMPSelectRepresentativePeaks function to filter
  the initial peak picks (default: {PeaksPerChromatogram->5,PeaksPerInterval->8,IntervalSize->.5})
PeakPickingOptions - options to be passed to the DAMPPickPeaks function (default: {Threshold->5000})
PeakLayoutPlotOptions - options to be passed to the DAMPPlotPeakLayout function (default: {})
FitShiftFunctionOptions - options to be passed to the DAMPFitShiftFunction function (default: {})
AnnotationTables - a list of annotation tables to be aligned to the reference msdata (default: None)
OutputTimeRange - time range to which the resulting normalized datasets should be cropped (default: All)
ExternalNormalizationCoefficients - list of coefficients by which the signal intensities in msdatas
  will be multiplied. The number of coefficients in the list must equal the number of datasets in msdatas (default: None)
Resolution - resolution to which the datasets were binned along the m/z dimension. The annotation
  tables passed through the AnnotationTables option will be binned the same way (default: 1)
InternalStandard - the internal standard for signal intensity normalization may be specified in one of two ways: 1) short name (
  3rd column) from the first annotation table in the list passed via the AnnotationTables option. In this case the position of
  the internal standard will be extrapolated from the aligned annotation table and the vicinity blindly integrated 2) specifying
  the m/z and integration time range (as in the reference dataset) explicitly: {mz,{starttime,endtime}}. (default: None)
AutoISIntegrationVicinity - if the location of the internal standard is extrapolated from the aligned annotation table, this option
  determines the vicinity (in minutes) of the predicted retention/migration time to be blindly integrated (default: {-.25,.25})
SaveMemory - if set to true, signal intensities are rounded to integers in internal calculations and results (default: True)
```

```
?DAMPTwoDatasets
```

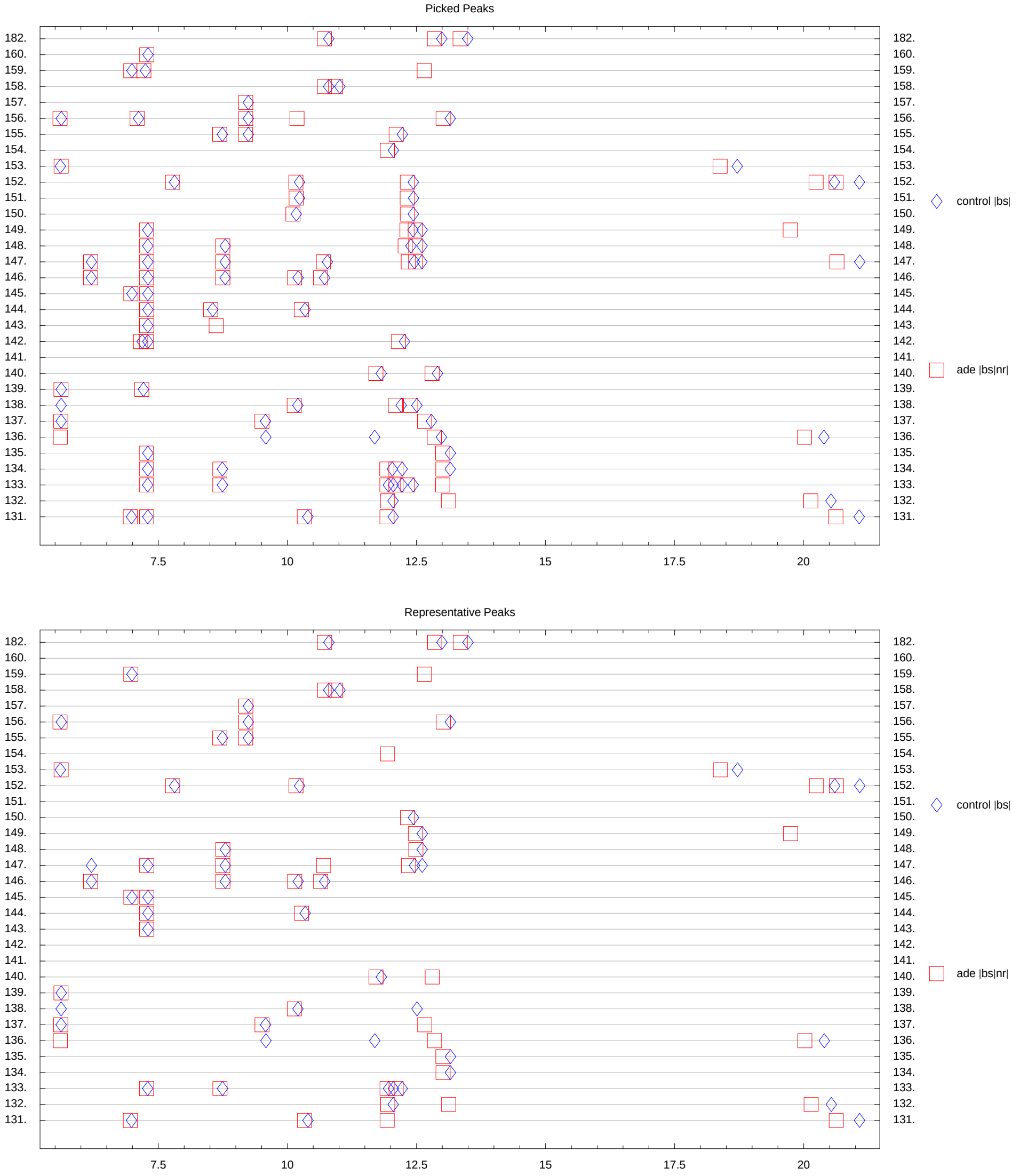
```
DAMPTwoDatasets[msdata1,msdata2,options] generates datasets representing the absolute, relative, and absolute×relative
  differences between msdata2 and msdata1 and returns them along with the normalized datasets and aligned annotation tables
  as a list of rules: {NormalizedDatasets->...,Absolute->...,Relative->...,AbsoluteRelative->...,AlignedAnnotationTables->...}
Options:
NormalizeGroupOptions - a list of options for the DAMPNormalizeGroup function which is used internally to normalize the datasets (default: {})
ThresholdForRelative - relative difference in the relative result will be set to 0 if neither of the
  two corresponding signal intensities from msdata1 and msdata2 are equal to or greater than this threshold (default: 0)
```

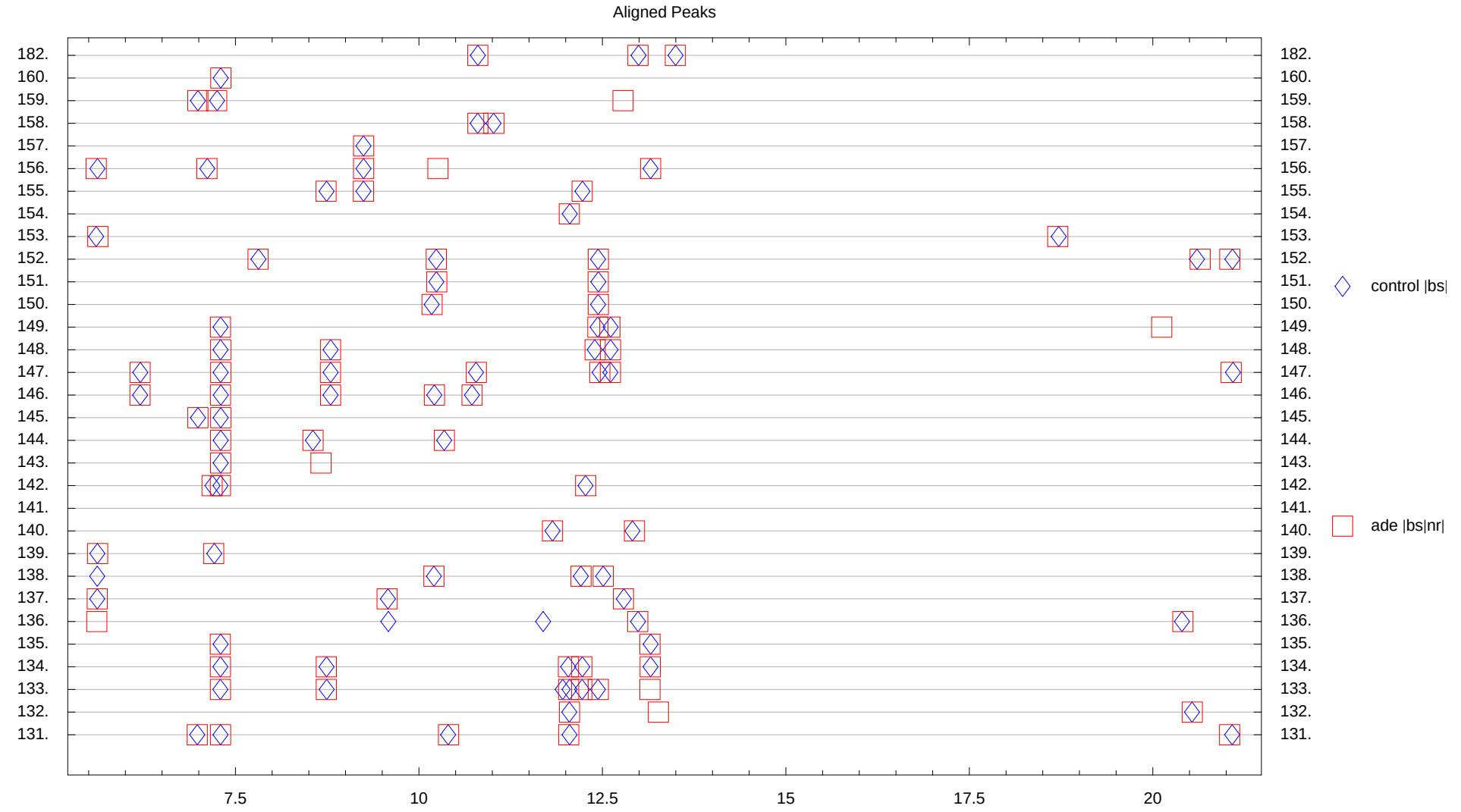
The loaded datasets are preprocessed prior to differential analysis in the same way as in the 02−MathDAMP−Elements.nb notebook (baseline subtraction and noise removal).

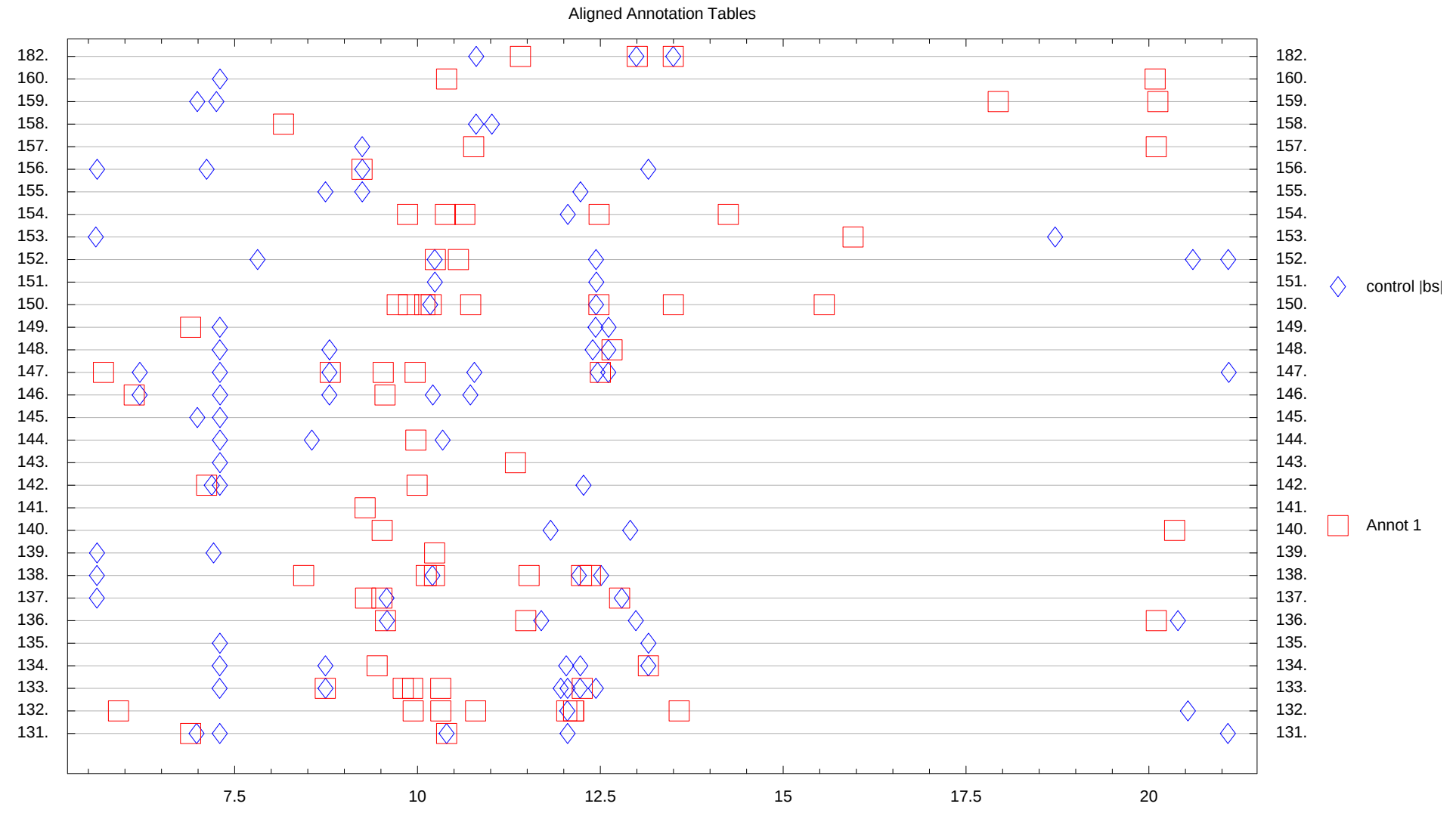
```
{ppctrl, ppsmpl} = DAMPRemoveNoise[DAMPSubtractBaselines[#]] & /@ {ctrl, smpl};
```

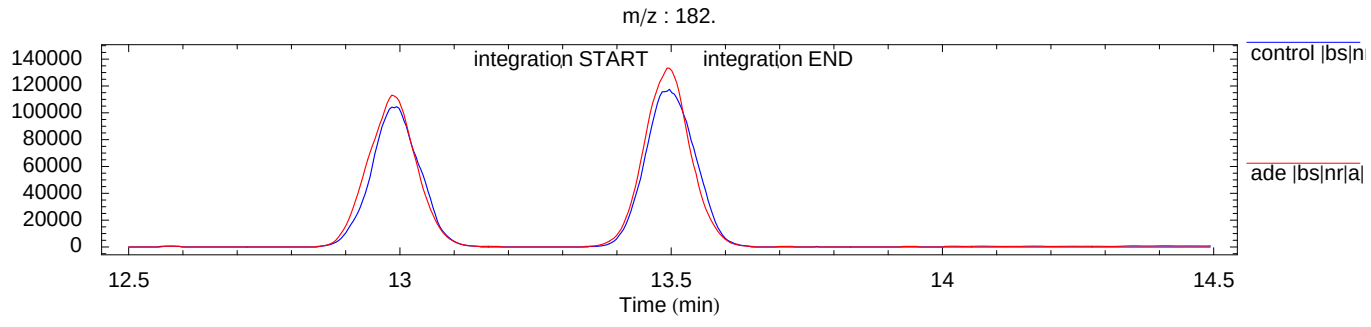
Most of the options for the `DAMPTwoDatasets` and `DAMPNormalizeGroup` are specified explicitly in the following command to allow easy editing of the options. The annotation table for the cation mode CE−MS analysis is used. This table was assembled according to a CE−TOFMS analysis of a mixture of standard compounds. Methioninesulfone is used as an internal standard. Its short name (in the annotation table) 363 is passed to the `DAMPNormalizeGroup` function via the `InternalStandard` option. The location of the peak of the internal standard will be extrapolated from the aligned annotation table. Overlaid electropherograms of the vicinities of the expected peaks of the internal standard are plotted along with indicators of the beginning and the end of blindly integrated regions for visual confirmation. To specify the location of the peak explicitly, use the notation {mz,{starttime,endtime}} instead of the short name. In this case it would be {182,{13.3,13.7}} (according to the electropherogram at the end of the optional section).
The `ppctrl` dataset will be used as the reference dataset. To use the `ppsmpl` dataset as the reference dataset, set the option `Reference` to 2.
The annotation table is reduced to contain only items with m/z values relevant to the analyzed datasets.

```
rslt = DAMPTwoDatasets[ppctrl, ppsmpl,
   NormalizeGroupOptions → {Reference → 1, AlignmentTimeRange → All, InternalStandard → 363, AutoISIntegrationVicinity → {-.2, .2},
     PeakPickingOptions → {Threshold → 2000}, RepresentativePeakOptions → {PeaksPerChromatogram → 5, PeaksPerInterval → 5, IntervalSize → .5},
     FitShiftFunctionOptions → {GapPenalty → {3, .5}},
     AnnotationTables → {Select[DAMPLoadAnnotationTable[MathDAMPPath <> "/iab_cems_cation.csv"], MemberQ[ppctrl[[2]], 1. Round[#[[1]]] &]},
     OutputTimeRange → All, ExternalNormalizationCoefficients → None}, ThresholdForRelative → 0];
```


Picked Peaks
182.
160.
159.
158.
157.
156.
155.
154.
153.
152.
151.
150.
149.
148.
147.
146.
145.
144.
143.
142.
141.
140.
139.
138.
137.
136.
135.
134.
133.
132.
131.
7.5
10
12.5
15
17.5
20
control |bs|
ade |bs|nr|
Representative Peaks
182.
160.
159.
158.
157.
156.
155.
154.
153.
152.
151.
150.
149.
148.
147.
146.
145.
144.
143.
142.
141.
140.
139.
138.
137.
136.
135.
134.
133.
132.
131.
7.5
10
12.5
15
17.5
20
control |bs|
ade |bs|nr|





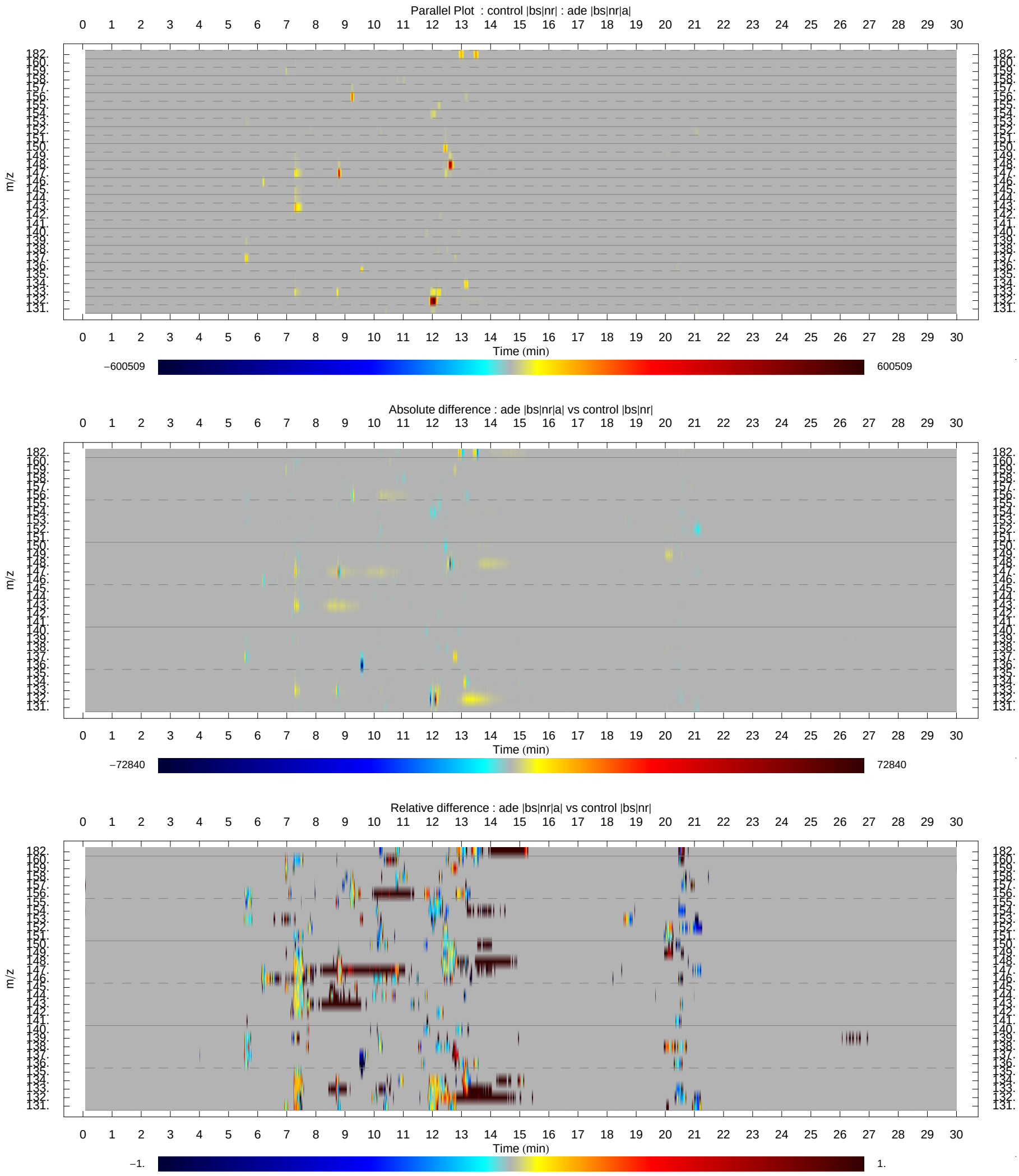


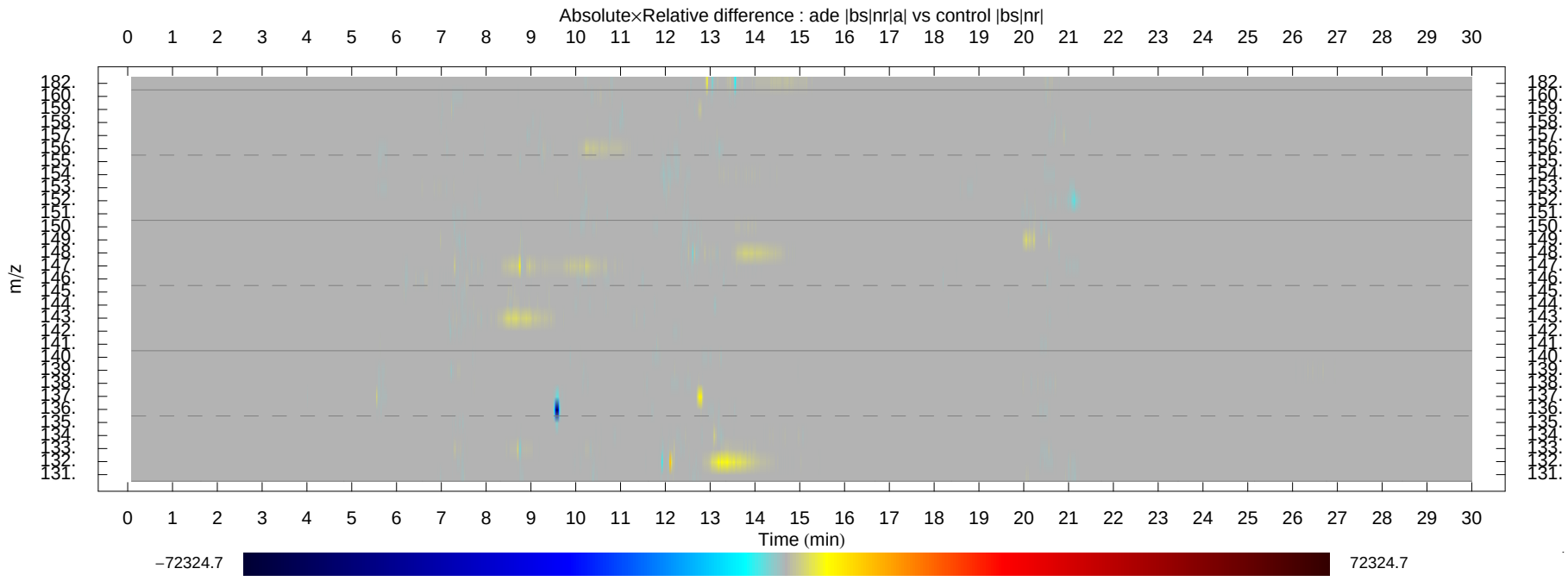

```
IS normalization coefficients : {1., 0.952856}
```

Step 3 : Exploring the Results, Listing the Candidates

The visualization of the results returned by the `DAMPTwoDatasets` is shown below (the parallel plot, the absolute, relative, and absolute×relative differences). The signal intensity threshold for calculating the relative difference was set to 0 this time (in contrast to the example in the 02−MathDAMP−Elements.nb notebook). Although the relative difference dataset contains a higher number of signals and smears, the absolute×relative difference dataset does not seem to be significantly affected.
Annotation is not shown on the plots below. On how to show the annotation or to alter the appearance of the plots, please refer to the 02−MathDAMP−Elements.nb notebook.

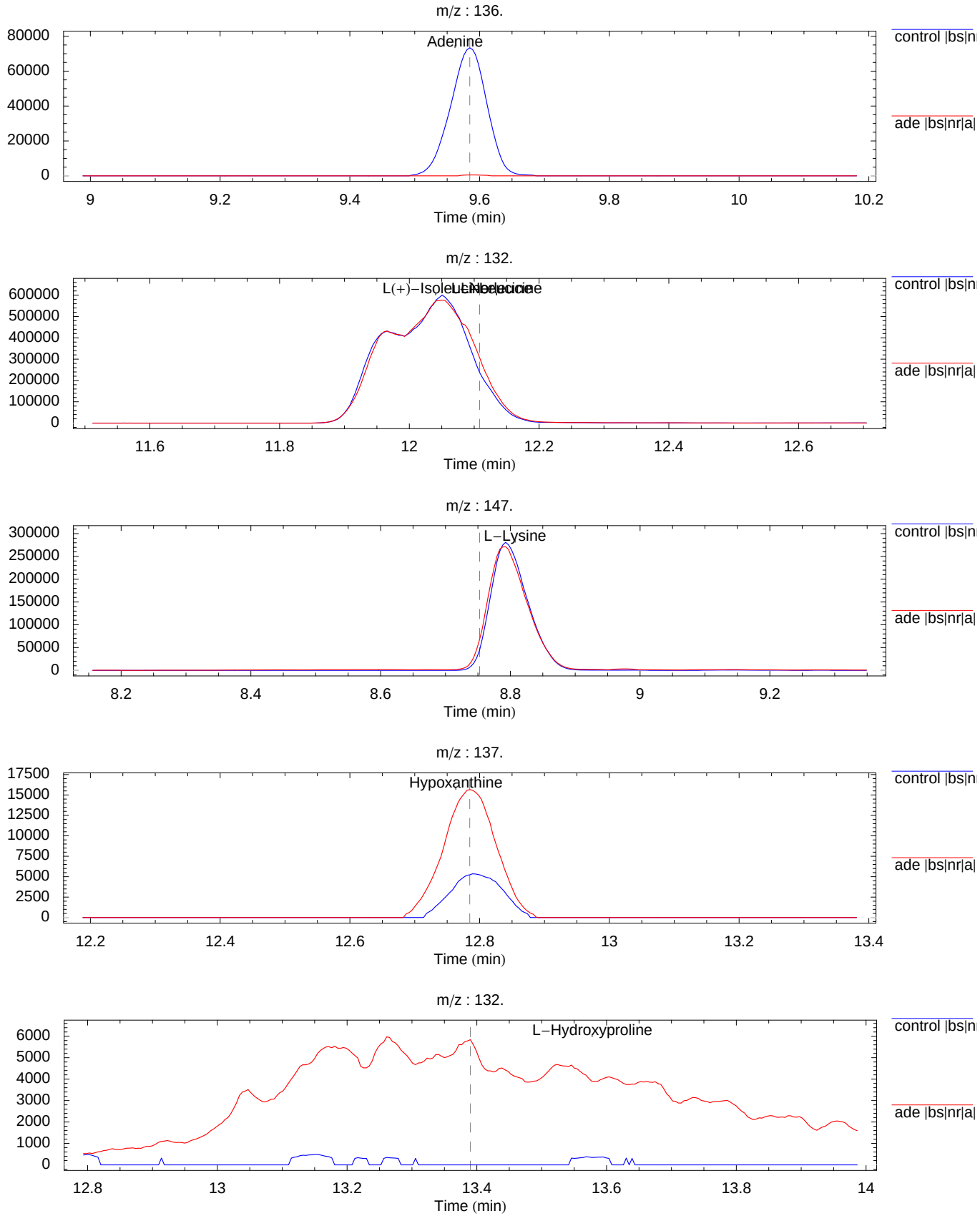
```
DAMPParallelPlot[NormalizedDatasets /. rslt];
DAMPDensityPlot[# /. rslt] & /@ {Absolute, Relative, AbsoluteRelative};
```

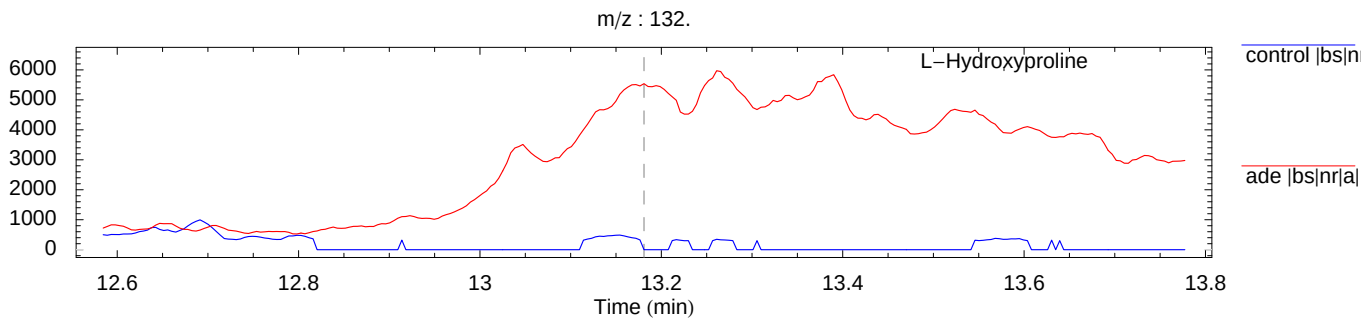


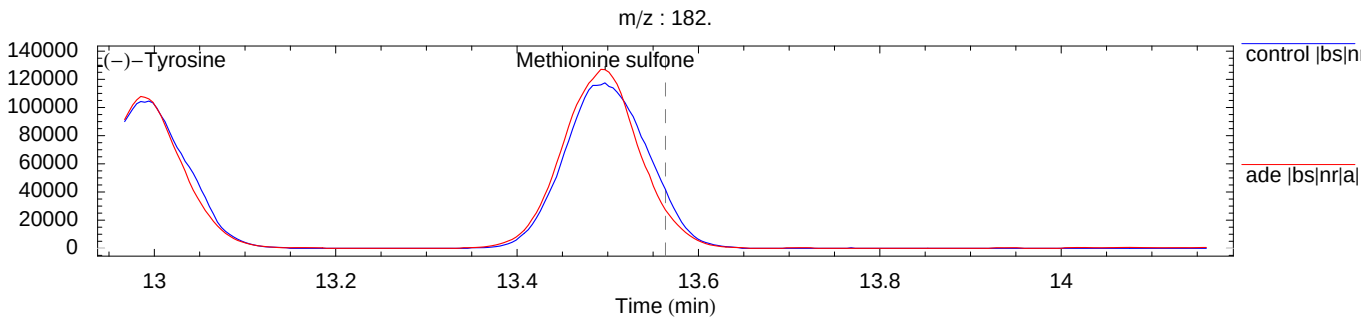


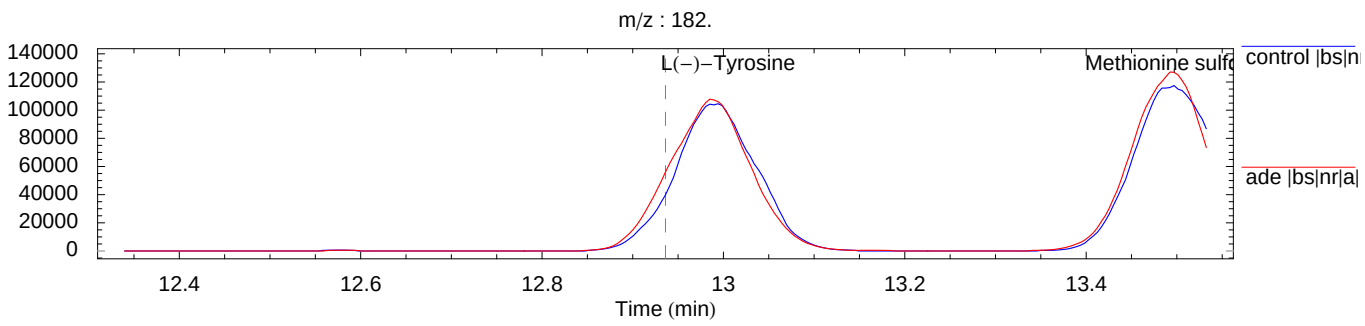
For the visual confirmation of significant differences between the datasets (and for the rejection of false positives), overlaid electropherograms are plotted in descending order of significance. Below are the electropherograms of the top 12 differences from the absolute×relative difference result. The vertical dashed line indicates the position of the most significant difference according to the selected criteria.

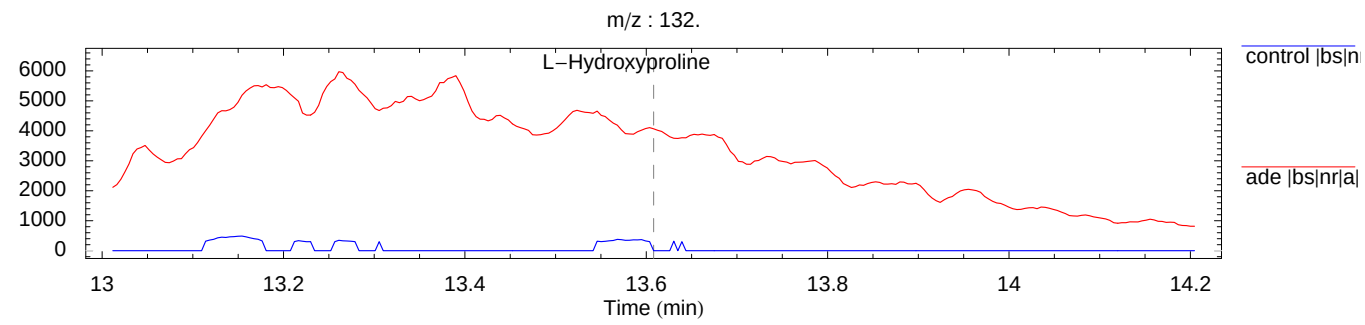
```
DAMPPlotCandidates[NormalizedDatasets /. rslt, AbsoluteRelative /. rslt,
  PlotCount → 12, PlotChromatogramOptions → {AnnotationTable → (AlignedAnnotationTables /. rslt)[[1]]}];
```

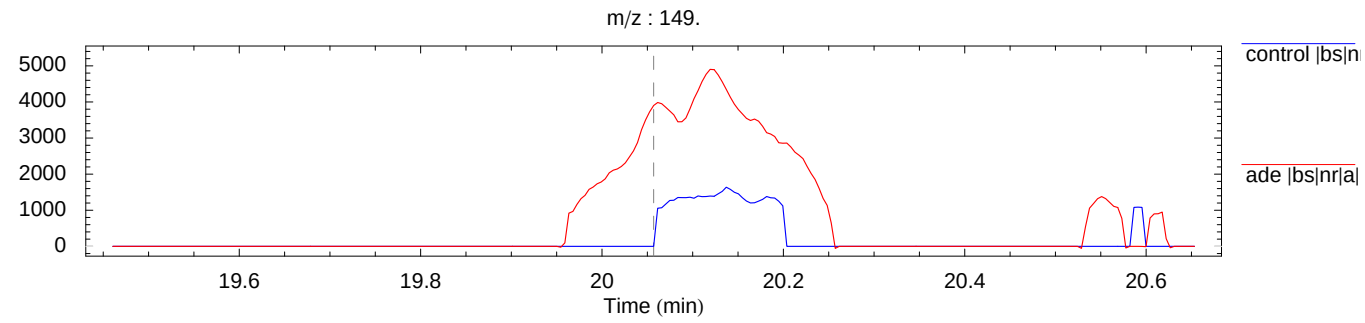


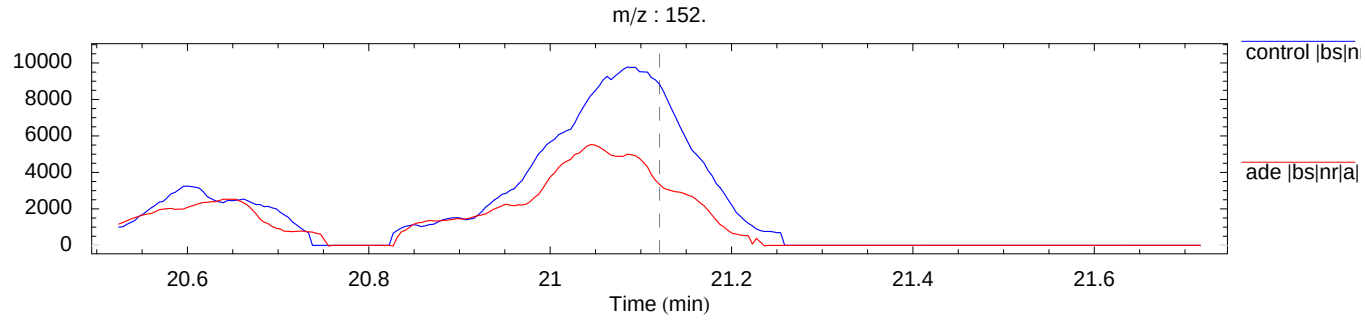

m/z : 132.
L−Hydroxyproline
6000
5000
4000
3000
2000
1000
0
12.6
12.8
13
13.2
13.4
13.6
13.8
Time (min)
control |bs|n
ade |bs|nr|a|

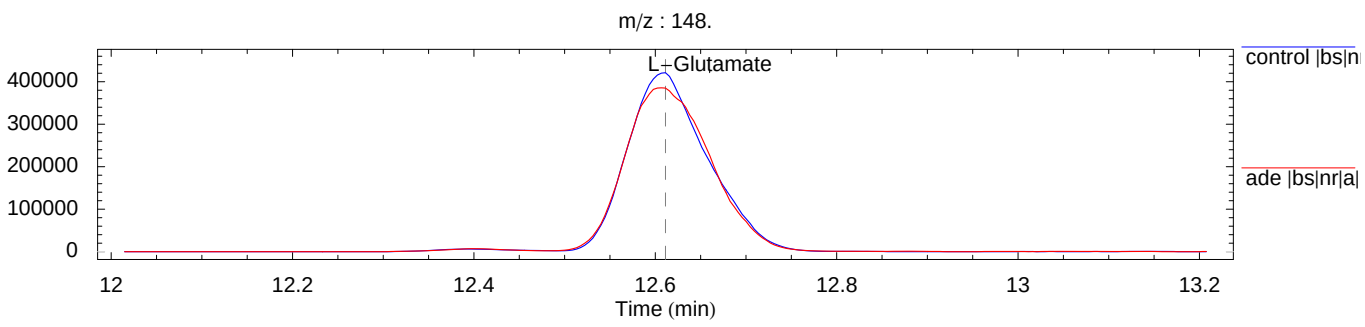

m/z : 182.
(−)−Tyrosine
Methionine sulfone
140000
120000
100000
80000
60000
40000
20000
0
13
13.2
13.4
13.6
13.8
14
Time (min)
control |bs|n
ade |bs|nr|a|


m/z : 182.
L(−)−Tyrosine
Methionine sulfc
140000
120000
100000
80000
60000
40000
20000
0
12.4
12.6
12.8
13
13.2
13.4
Time (min)
control |bs|n
ade |bs|nr|a|


m/z : 132.
L−Hydroxyproline
6000
5000
4000
3000
2000
1000
0
13
13.2
13.4
13.6
13.8
14
14.2
Time (min)
control |bs|n
ade |bs|nr|a|


m/z : 149.
5000
4000
3000
2000
1000
0
19.6
19.8
20
20.2
20.4
20.6
Time (min)
control |bs|n
ade |bs|nr|a|


m/z : 152.
10000
8000
6000
4000
2000
0
20.6
20.8
21
21.2
21.4
21.6
Time (min)
control |bs|n
ade |bs|nr|a|


m/z : 148.
L−Glutamate
400000
300000
200000
100000
0
12
12.2
12.4
12.6
12.8
13
13.2
Time (min)
control |bs|n
ade |bs|nr|a|