MS Compare – BioMarker Discovery Case Study

LC-MS Data:

Data files were downloaded from Listgarten et al. The data set consists of 14 serum samples; 7 control samples and 7 samples spiked with a Peptide Calibration Standard. Each LC-MS run contains 1205 scans (10 – 54 minutes) and a mass range between 400 and 1600 Dalton. The samples were measured on a LCQ quadrupole ion trap tandem mass spectrometer (Thermo Finnigan). To import the data into MS Xelerator an ASCII conversion routine was written.

Data Complexity:

It was stated in the article that the amount of spike-in was not so large as to rise above the baseline serum peaks and thus not trivially easy to find. To asses the data complexity in an independent way, **MPeaks** was run on one of the spike-in samples. MPeaks peak-picking using default settings returned more than 27000 peaks for this sample. In the MPeaks overview plot (Figure 1) one of the mass chromatograms (m/z 744) is plotted, this masschromatogram alone contains about 24 peaks. The peaks shown range between 0.2 % and 2.3 % (compared to the largest peak in the data file) and appear to be quite narrow.

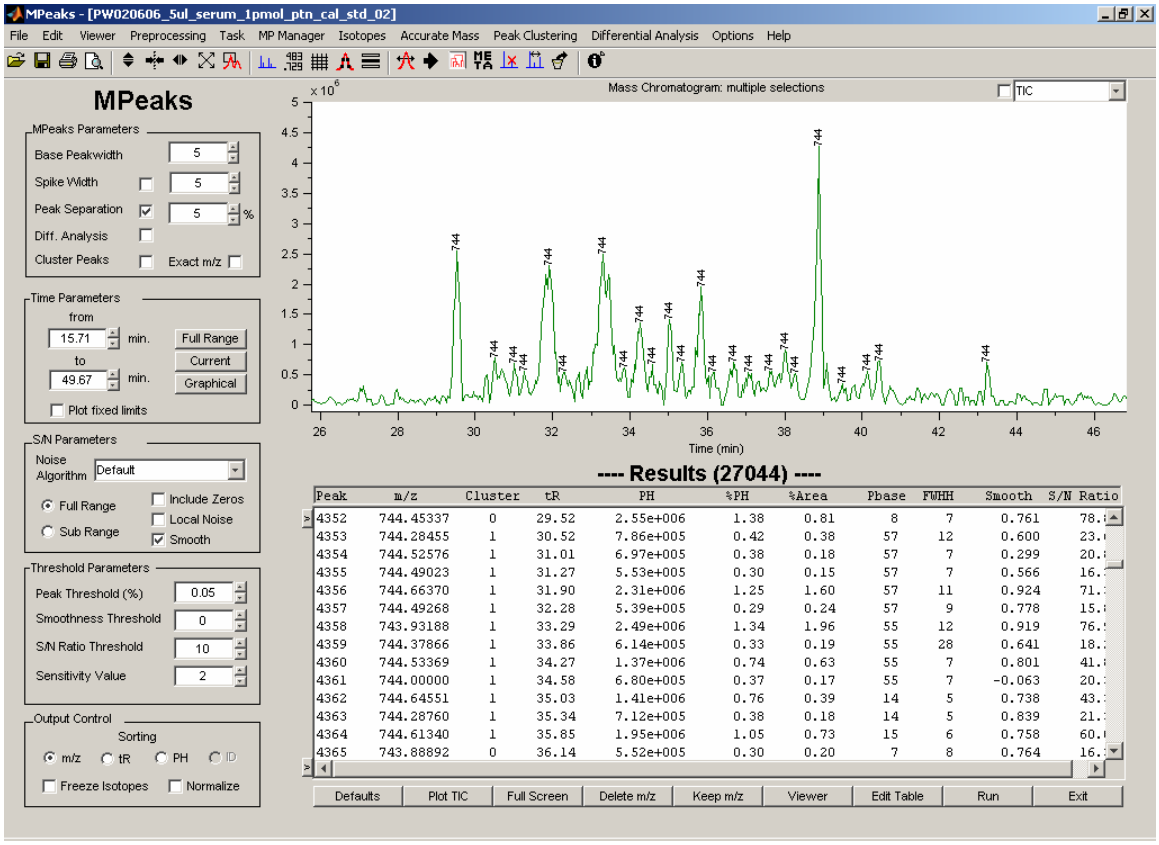


Figure 1: MPeaks Overview Screen

MS Compare will be used to detect the differences between the two groups of samples and to demonstrate the utilities and algorithms available. Figure 2 displays the MS Compare window with the 14 samples loaded. The bottom window displays the TIC overlay of all samples, one of the larger peaks was selected from the TIC. MS compare automatically identifies the selected sample and plots the extracted ion currents to the top window (or the mass spectrum). As can be seen from the Extracted Ion Current m/z 951 a severe alignment problem exists, the difference between the minimum and maximum retention times for this peak is about 2 minutes.

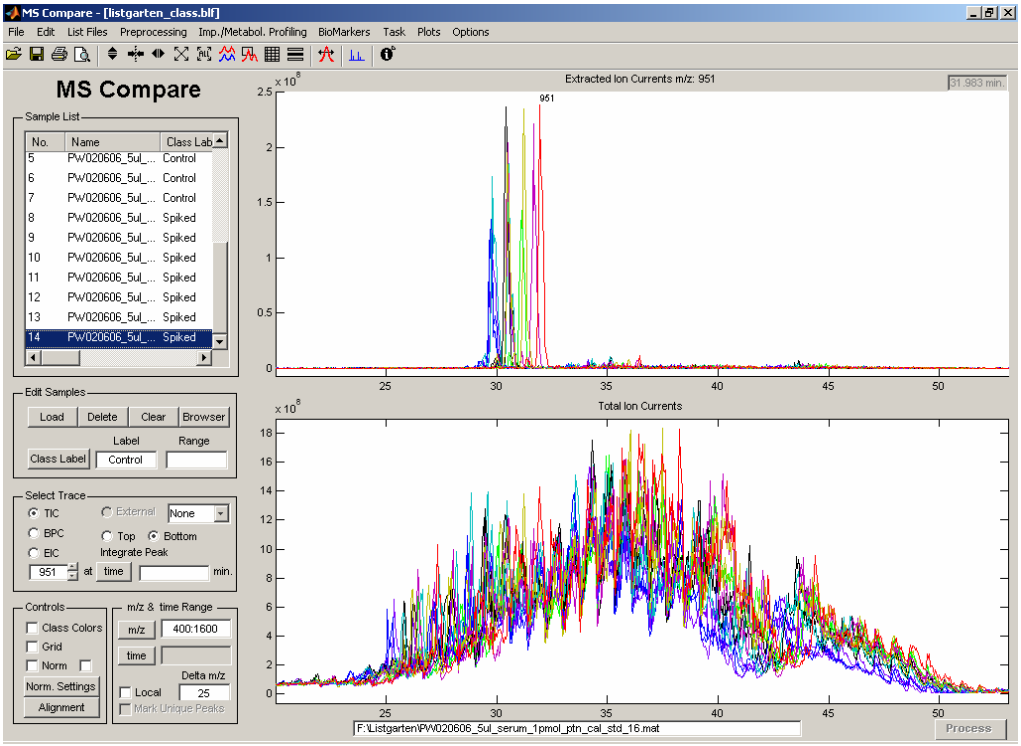


Figure 2: MS Compare – Overlay of TIC's (bottom) and extracted ion currents (m/z 951, top)

Data Alignment:

The MS Compare module contains 4 alignment algorithms (offset shifting, cross correlation, peak based warping and correlation optimized warping). The data were aligned based on correlation optimized warping using the base peak chromatogram. Results before and after alignment are shown in Figure 3.

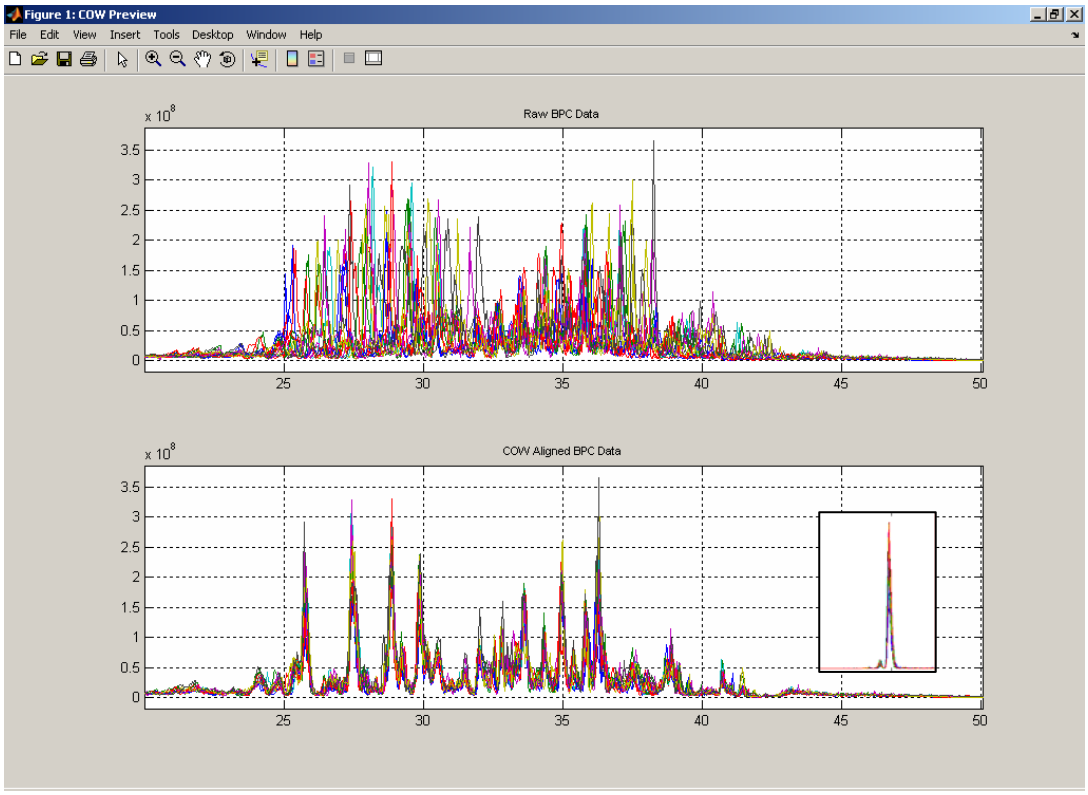


Figure 3: Unaligned (top) and aligned (bottom) BPC's of the 14 LC-MS runs. Inset: m/z 951, compare with figure 2 The alignment was performed in 22 seconds, using default settings.

Finding Differential Peaks using MS Compare:

MS Compare contains several algorithms to detect differential peaks between two classes/groups of samples. The first two methods are more or less visually based. A third method performs an extensive search for significant differences between both groups by scanning all 2D-LC/MS surfaces simultaneously.

Method 1: Comparing TCM/ BCM traces

The TCM (Total Chromatogram Mass spectrum) is basically the average mass spectrum of a LC/MS data set (comparable with the Total Ion Current). The BCM (Base Chromatogram Mass spectrum) can be compared to the Base Peak Chromatogram.

The advantage of using "mass spectra" is the absence of alignment problems. The disadvantage however, is that small differential peaks will not be easy to detect in the presence of larger peaks having the same m/z value.

Scanning these "mass spectra" for differences can be done as shown in Figure 4. In this case the BCM mass spectra are plotted in the bottom window between m/z 400 and m/z 1600 (blue: controls, red: spiked samples). Finding differences is basically a visual exercise. MS Xelerator will automatically identify m/z values for which a 100% separation between both groups can be found (green vertical lines). Using this screening method we find 10 unique m/z values. Clicking on one of the unique m/z values will extract the accompanying masschromatograms (top window, m/z 649).

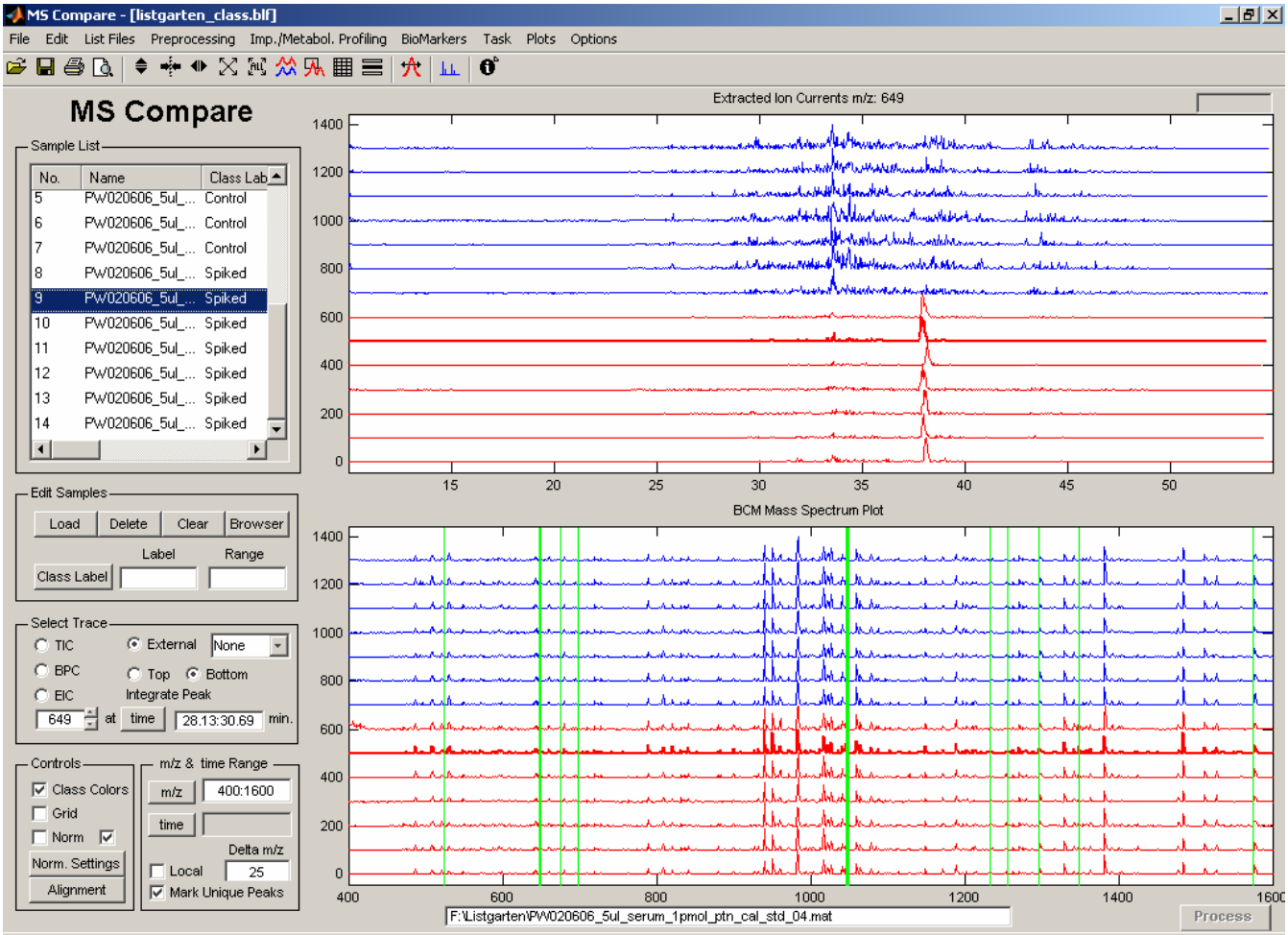


Figure 4: MS Compare: Differential analysis by comparison of BCM mass spectra. Unique m/z values are marked.

Method 2: Comparing TIC/BPC traces

A visually similar method exists for comparing TIC/BPC traces. However, TIC's and BPC's can be created on any mass range the user wants. As a default TIC's will be build using a mass window of 25 Dalton. A fast screening of the above data set can be done in less than a minute. A total of 7 differential chromatographic peaks were found using this method.

Method 3: Scanning and comparing 2D-surfaces

To automatically scan and compare all 14 data sets for unique peaks, MS Compare creates a so-called two-dimensional surface map as shown in figure 5. On the x-axis retention time is plotted, on the y-axis the m/z value and peaks fulfilling all

criteria are marked using a dark color. From the binary map the unique peaks are easily identified. Maps can be created using a number of criteria. In this case full selectivity was requested. Alternatives are: absolute or relative differences, t-statistics, Fischer Discriminant value, etc. Plots are interactive, clicking on a unique position will plot the extracted ion currents of all 14 samples in overlay in the top window. A default run finds 18 unique peaks.

The map can be converted to a table as shown. Based on this table a number of Multivariate Analysis techniques can be performed, like Hierarchical Clustering and Principal Component Analysis (PCA).

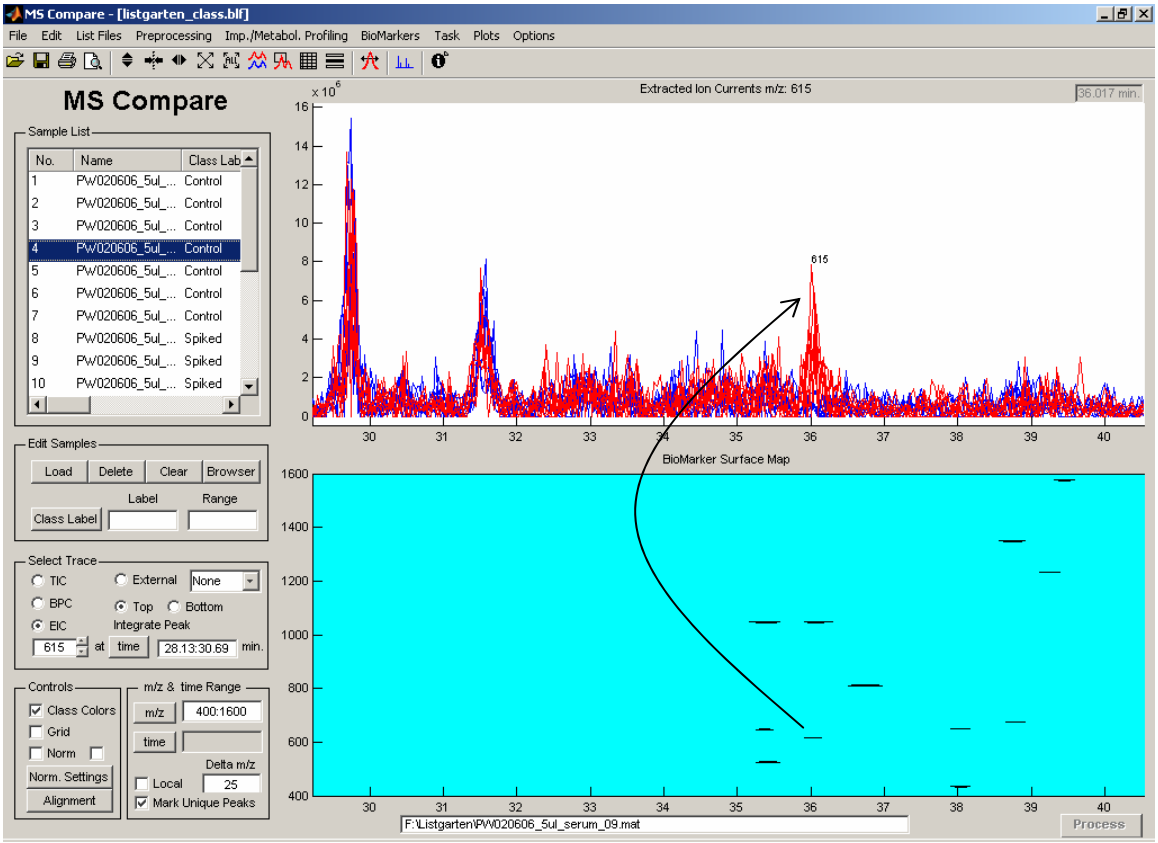
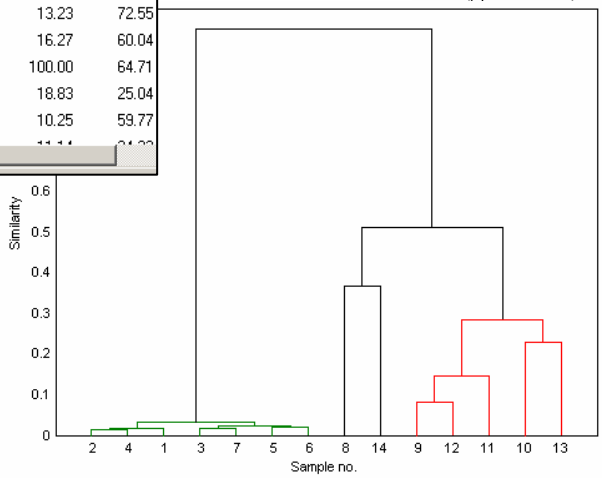


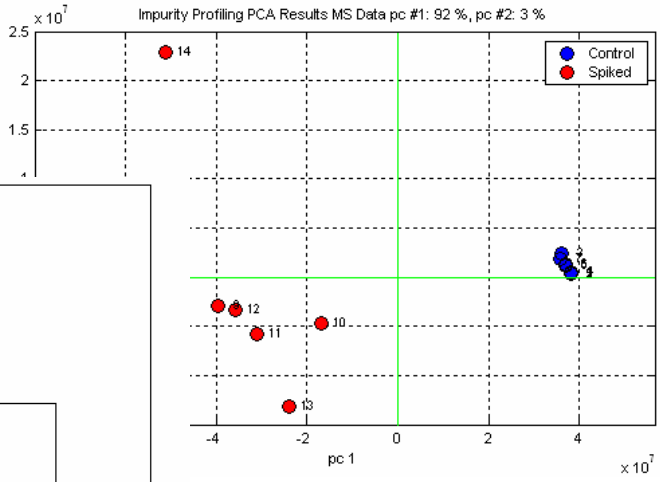
Figure 5: MS Compare showing unique peaks in surface map (retention time versus m/z value).

Peak	tR(min.)	m/z	s1: PW...	s2: PW...	s3: PW...	s4: PW...
1	38.036	434	14.99	19.23	7.01	4.89
2	35.410	524	45.19	9.69	16.16	55.28
3	35.410	525	50.64	32.70	21.79	76.30
4	36.025	615	31.92	18.63	14.97	65.61
5	35.367	647	24.53	25.84	26.85	65.52
6	38.036	649	51.05	13.89	22.98	31.17
7	38.036	650	28.27	43.30	11.62	76.45
8	38.786	675	0.00	55.54	13.23	72.55
9	36.738	811	33.90	25.01	16.27	60.04
10	35.367	1046	41.84	28.29	100.00	64.71
11	35.410	1047	16.16	29.75	18.83	25.04
12	36.064	1047	22.01	32.26	10.25	59.77

Table of Unique Peaks



Clustering



PCA

The MS Compare module offers different methods to present the results. Two plots are shown in Figure 6. In the bottom window the peak area is plotted as a function of the (unique) peak number. From this plot the difference in peak area between both groups and the spread inside samples from the spiked group is easily observed. For the control samples most peaks are virtually absent, as was expected for this case study.

The top window shows the concentration profiles for 4 selected peaks in overlay mode . This plot is very useful in cases where the samples were measured as a function of time or some other kind of continuous parameter.

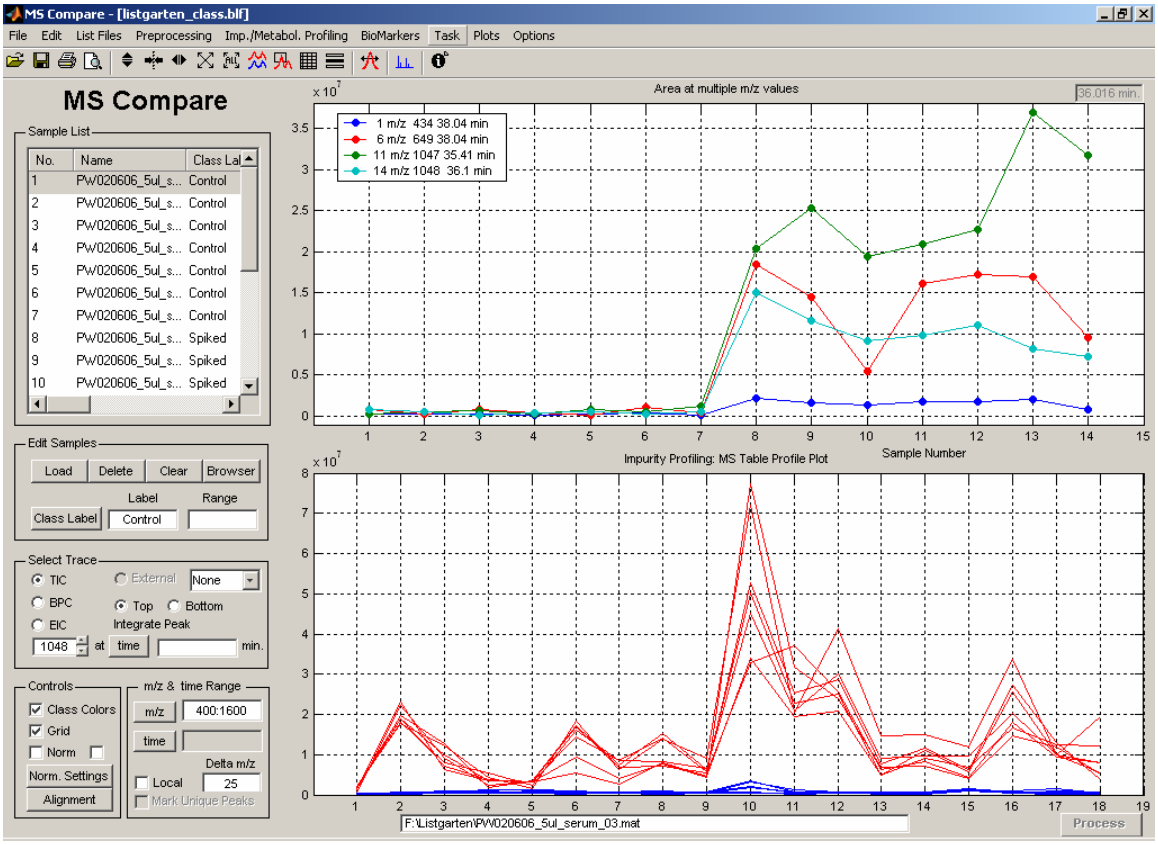
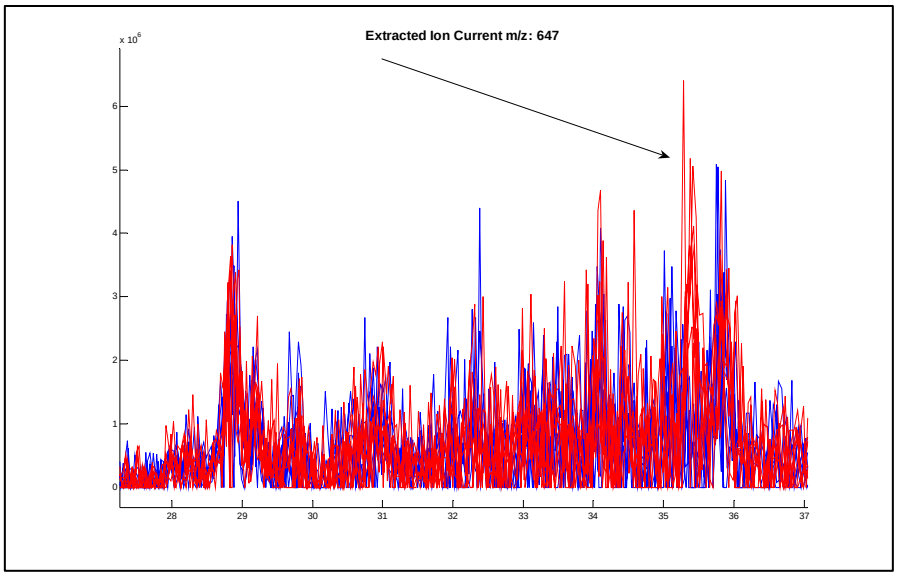


Figure 6: MS Compare – Table profile plots

As an example of the sensitivity of the method, the masschromatograms for unique peak number 5 (m/z 647) are plotted in Figure 7. The differential peaks eluting at 35.5 minutes are hardly visible due to the high noise levels.



In this example the surface map was created without any preprocessing. Preprocessing methods available in MS Compare are: de-spiking, smoothing, normalization and baseline correction. If a slight smoothing would have been applied, the number of unique peaks raises to 47.
In this example, smoothing has a very positive effect on the data quality.

MS compare is very interactive and links directly to the Browser and other modules. At any moment the user will be able to compare a large number of samples based on mass chromatograms or mass spectra, either manually or by using algorithmic techniques.