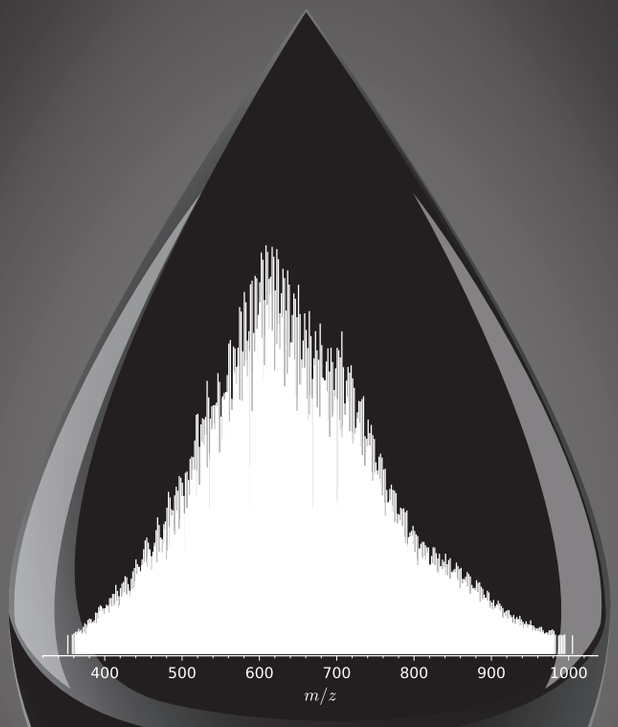
Petroleomics

FTMS Data Processing for complex mixture analysis:

crude oil fractions, biofuels,
dissolved organic matter, environmental



Spectroswiss

Advanced Petroleomics Workflow

The Spectroswiss software solutions (Peak-by-Peak and AutoVectis), alone or empowered by our high-performance data acquisition system (FTMS Booster X1), provide a quantum boost in FTMS data quality and processing sophistication, helping you to take on the most complex and challenging applications.

FTMS instruments

FTMS Booster X1

Mass spectra (.raw/.baf)
Transients (.dat/ser)

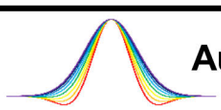
HD-transients (.h5)

1. Visualization:
- TIC & SIC chromatograms;
- mass spectra;
- transients;

2. Scans summation:
- transient or spectral;
- conditional co-adding;
- practically unlimited data sizes.

peakbypeak FTMS DATA ANALYSIS by spectroswiss

Data co-processing from separate experiments is possible



AutoVectis — **absorption mode FT**

3. Baseline correction:
- median of noise extrema BL;
- asymmetric least squares BL;

4. Noise thresholding;
- N-sigma thresholding;
- data-dependent thresholding;

5. Peak picking:
- N-points interpolation;
- targeted/untargeted;

6. Re-calibration:
- iterative re-calibration;
- piecewise binomial functions;

peakbypeak FTMS DATA ANALYSIS by spectroswiss

7. Results validation from external software:
- mass accuracy plot; - spectral slices; - Kendrick plot; C_{12}/C_{13} ratio distribution.

centroided or profile mass spectra
.mzxml, .mzml, .xls, .txt, .csv

External software for petroleomics data analysis
(Composer™, PetroOrg™, *etc.*)

Targeted Petroleomics Workflow for .RAW

Extended processing and analysis of .RAW files using Peak-by-Peak software.
Practically unlimited .RAW file size, on-demand capabilities, multiplexed data processing.

Mass spectra (.RAW): full profile, reduced profile, centroided

1. **Visualization**
2. **Spectral summation**
3. **Non-linear baseline subtraction**
4. **Noise thresholding**
5. **Peak picking**
6. **Re-calibration**
7. **External results re-processing**

peakbypeak
FTMS DATA ANALYSIS
by spectroswiss

External software for petroleomics data analysis:
Composer™, PetroOrg™, *etc.*

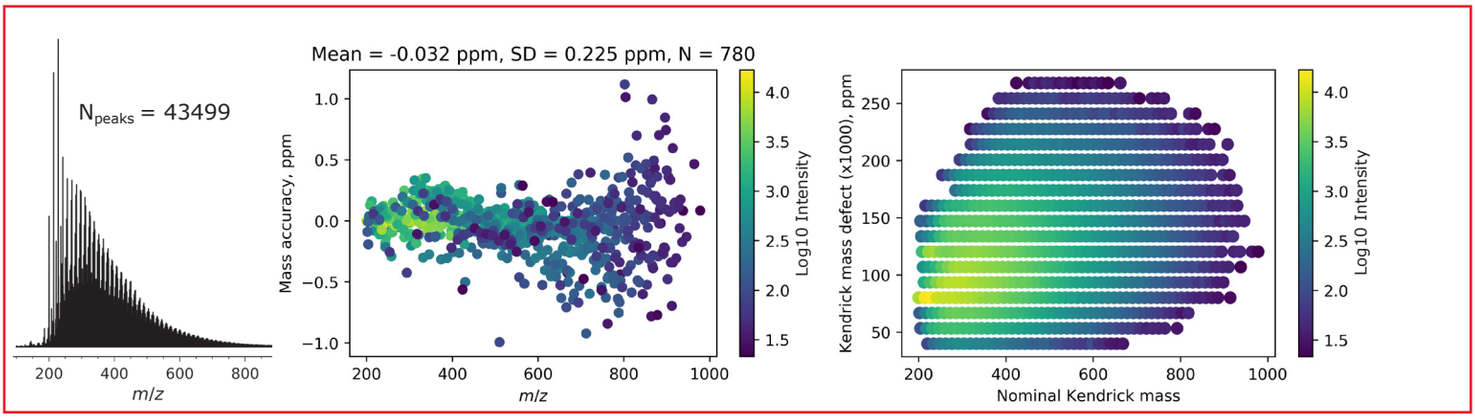


Figure 1. Positive ESI Orbitrap™ FTMS of a heavy crude oil sample (eFT, R=1'000'000 @ 400 *m/z*). The .RAW mass spectrum (Xcalibur™, Thermo Scientific) was acquired with an Orbitrap Elite™ and further processed using Peak-by-Peak software. Panels show recalibrated results: (left) a summed mass spectrum, (middle) mass accuracy distribution, and (right) a Kendrick plot for the N1 compound class. The presented data processing is according to: Zhurov, Kozhinov, and Tsybin; Evaluation of high-field Orbitrap Fourier transform mass spectrometer for petroleomics. *Energy Fuels* (2013) 27, 2974–2983.

Mass Spectra : Full Profile vs. Reduced Profile

The full profile mass spectra contain all data points including the full noise component, whereas the reduced profile mass spectra are produced by baseline subtraction followed by noise rejection (all data points below a certain threshold are removed) of a full profile data. Peak-by-Peak offers advanced capabilities of data processing for mass spectra presented in either reduced or full profile mode, Figures 2 and 3.

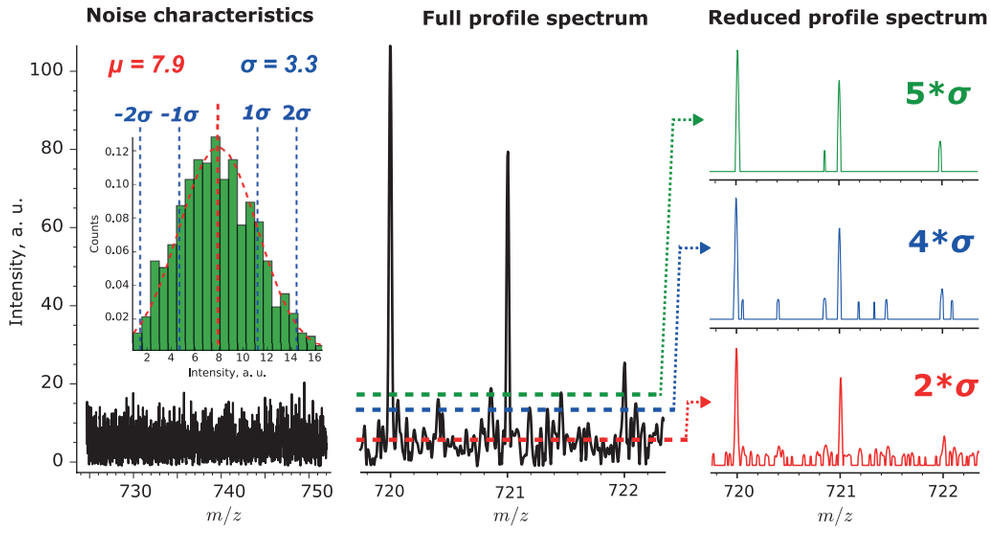


Figure 2. Application of a noise thresholding procedure to a positive APPI LTQ FT mass spectrum of C_{60}. (Left) determination of mean, μ, and standard deviation, σ, of noise for analyte-free mass range.
By definition : $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$; $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$, where *N* is a total number of data points. (Middle) the full profile mass spectrum and (right) the reduced profile mass spectra with different noise thresholding levels - standard deviation of noise, σ, multiplied by the user defined factor, 2, 4 or 5. The thresholding procedure is described in : Zhurov, Kozhinov, Fornelli, and Tsybin. Distinguishing Analyte from Noise Components in Mass Spectra of Complex Samples : Where to Cut the Noise ? *Analytical Chemistry* (2014) 86, 3308-3316.

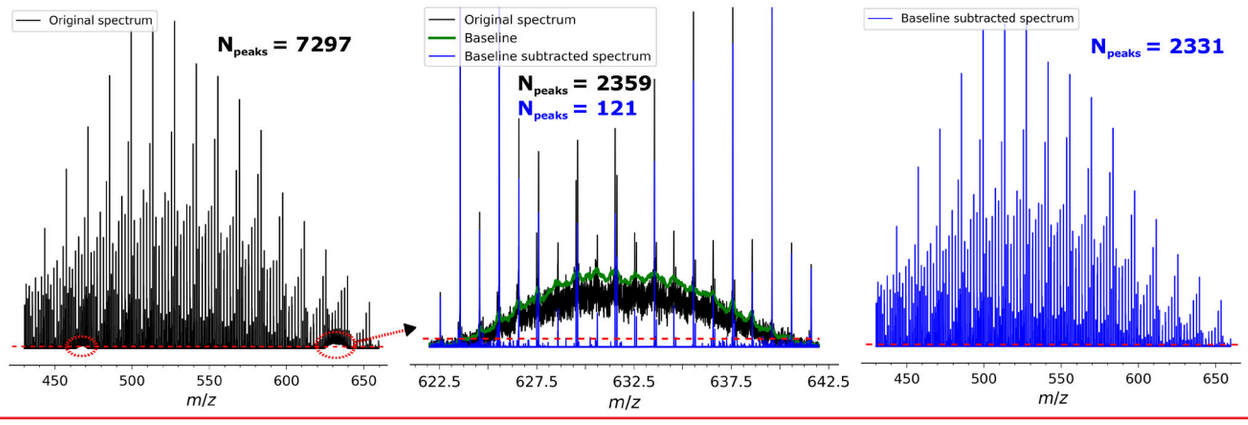
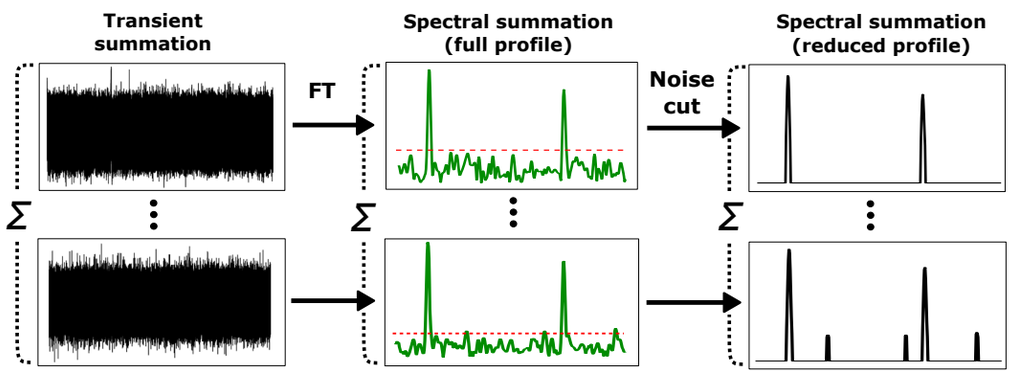


Figure 3. Application of a baseline correction procedure to a positive ESI FT-ICR MS reduced profile mass spectrum (.RAW) of a biofuel. (Left) the original mass spectrum, (middle) an expanded view of the mass range containing a significant baseline deviation, and (right) mass spectrum after baseline subtraction. Note the change in a number of peaks picked (noise threshold at 5*σ). Data processed via Peak-by-Peak software. The baseline calculation procedure is described in : Friedrichs ; A model-free algorithm for the removal of baseline artifacts.
Journal of Biomolecular NMR (1995) 5(2), 147-153.

FTMS Data Processing : Summation

Summation of mass spectral data from a number of measurements improves sensitivity and spectral dynamic range, Figures 4 and 5. Peak-by-Peak supports summation for transients (e.g., .DAT or .H5) and mass spectra (reduced and full profile mode, e.g., .RAW mass spectra).



Dependence of noise characteristics	*N* scans summation	
	Transient	Spectral
Mean of noise	$\downarrow\sqrt{N}$	*const*
Standard deviation of noise	$\downarrow\sqrt{N}$	$\downarrow\sqrt{N}$

Figure 4. The three approaches to scan summation, exemplified for a positive APPI FT-ICR MS data of C_{60}. Figure demonstrates (left) transient summation, (middle) full profile and (right) reduced profile (5*σ) spectral summation. Table summarizes reduction of noise mean (related to baseline value) and noise standard deviation (SD) as a number of scans *N* summed. Transient summation provides the highest value for performance improvement via data summation.

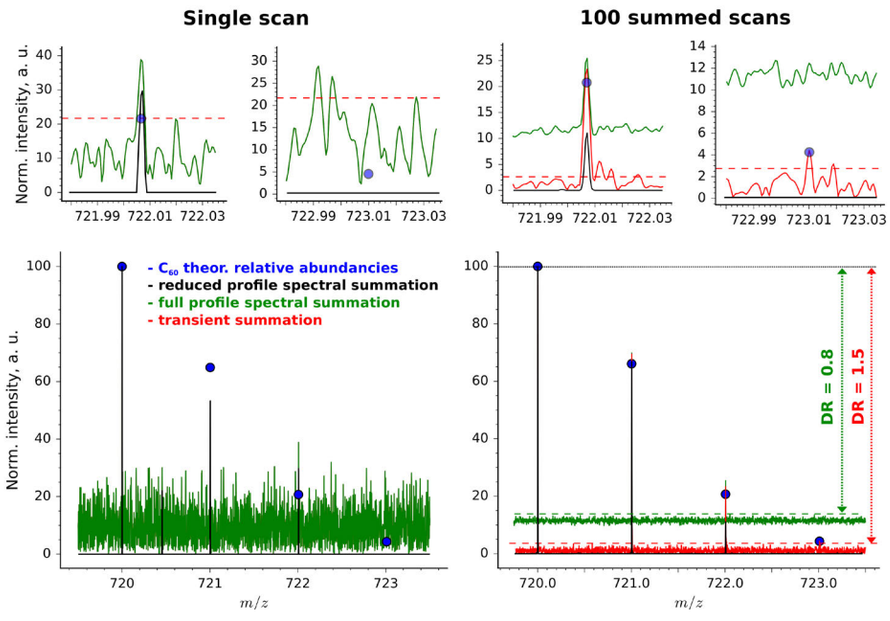
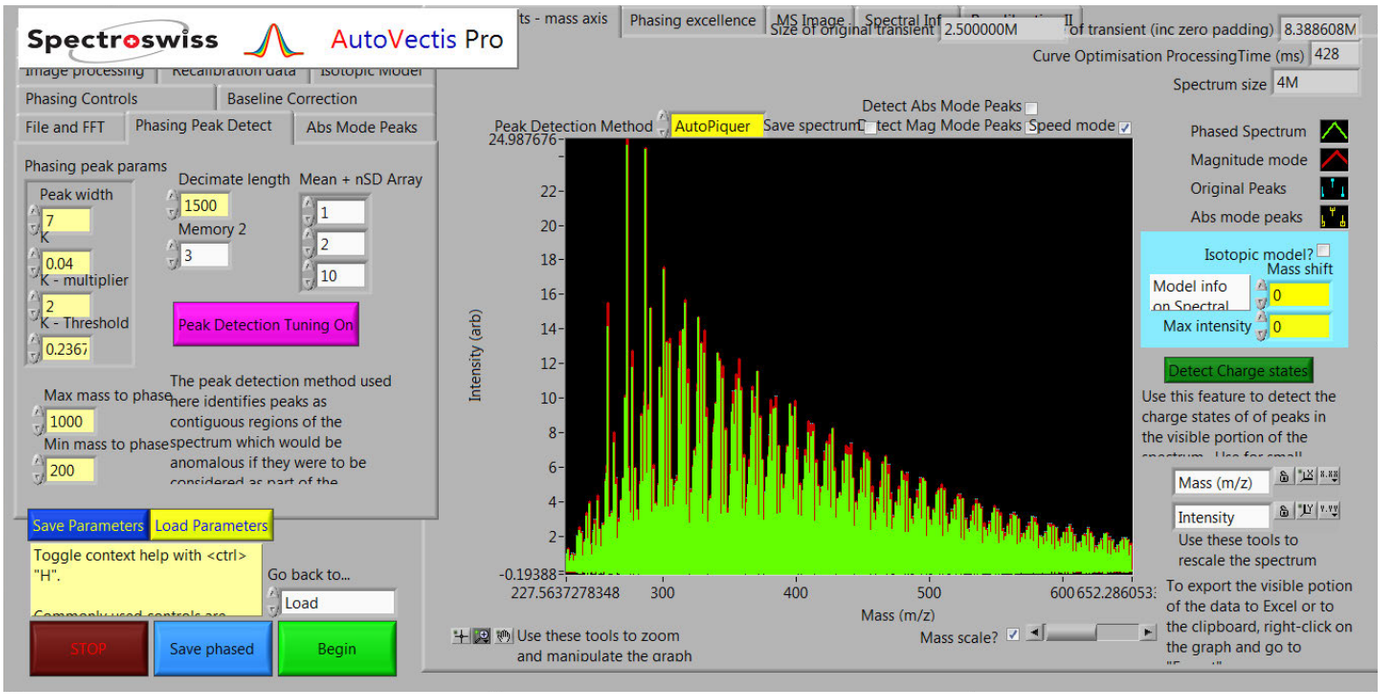


Figure 5. Application of Peak-by-Peak summation procedures to a positive APPI FT-ICR MS data of C_{60}. (Left) a single scan mass spectrum and (right) a summed mass spectrum (100 scans) obtained using different representation and summation approaches : **red** – transient summation, **green** – full profile and black – reduced profile spectral summation. (**Blue** dots) theoretical relative abundancies for $C_{60}^{+\cdot}$.

AutoVectis : Absorption Mode FT

FTMS data representation in **absorption mode FT** instead of a magnitude mode FT allows for 2-fold increase in resolution throughout a complete broadband mass spectrum of any mass range. The input data for absorption mode FT signal processing are FTMS transients and the output data are the mass spectra. The AutoVectis software package automates the steps required to produce absorption mode FT mass spectra. A genetic algorithm is used to optimize the phase correction function that relates ion frequency to the appropriate corrective phase angle to apply.



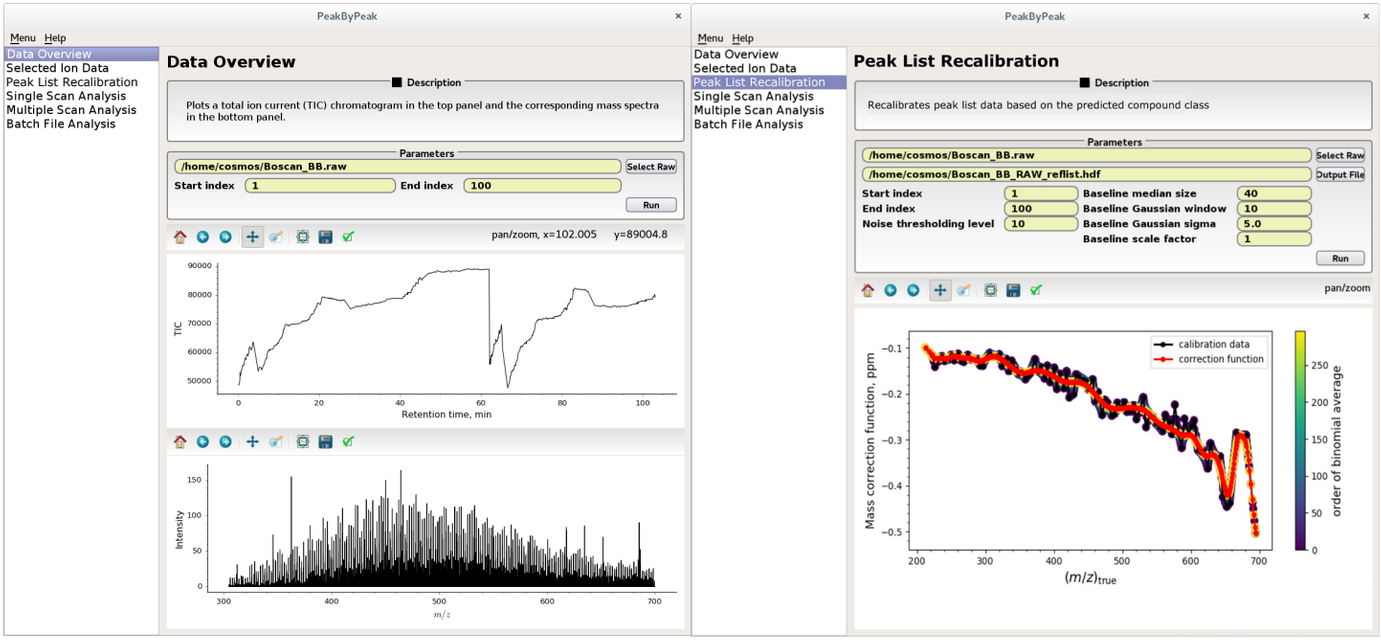
Key features of AutoVectis :

- absorption mode Fourier transform signal processing for FTMS ;
- quadratic and non-quadratic phase correction functions ;
- custom apodization functions to minimize undesirable baseline deviation ;
- support for data formats of common FTMS instruments and software ;
- configured for petroleomics applications, compatible with Peak-by-Peak ;
- developed in collaboration with Dr. David Kilgour.

Reference on original implementation for LTQ FT :
Kilgour David, Nagornov Konstantin, Kozhinov Anton, Zhurov Konstantin, and Tsybin Yury ; Producing absorption mode FT-ICR mass spectra with non-quadratic phase correction functions.
Rapid Commun. Mass Spectrom., 2015, 29, 1087-1093

Peak-by-Peak : FTMS Data Processing

Powerful and flexible software for advanced processing of FTMS data
Tailored workflows (graphical user interfaces) for petroleomics applications



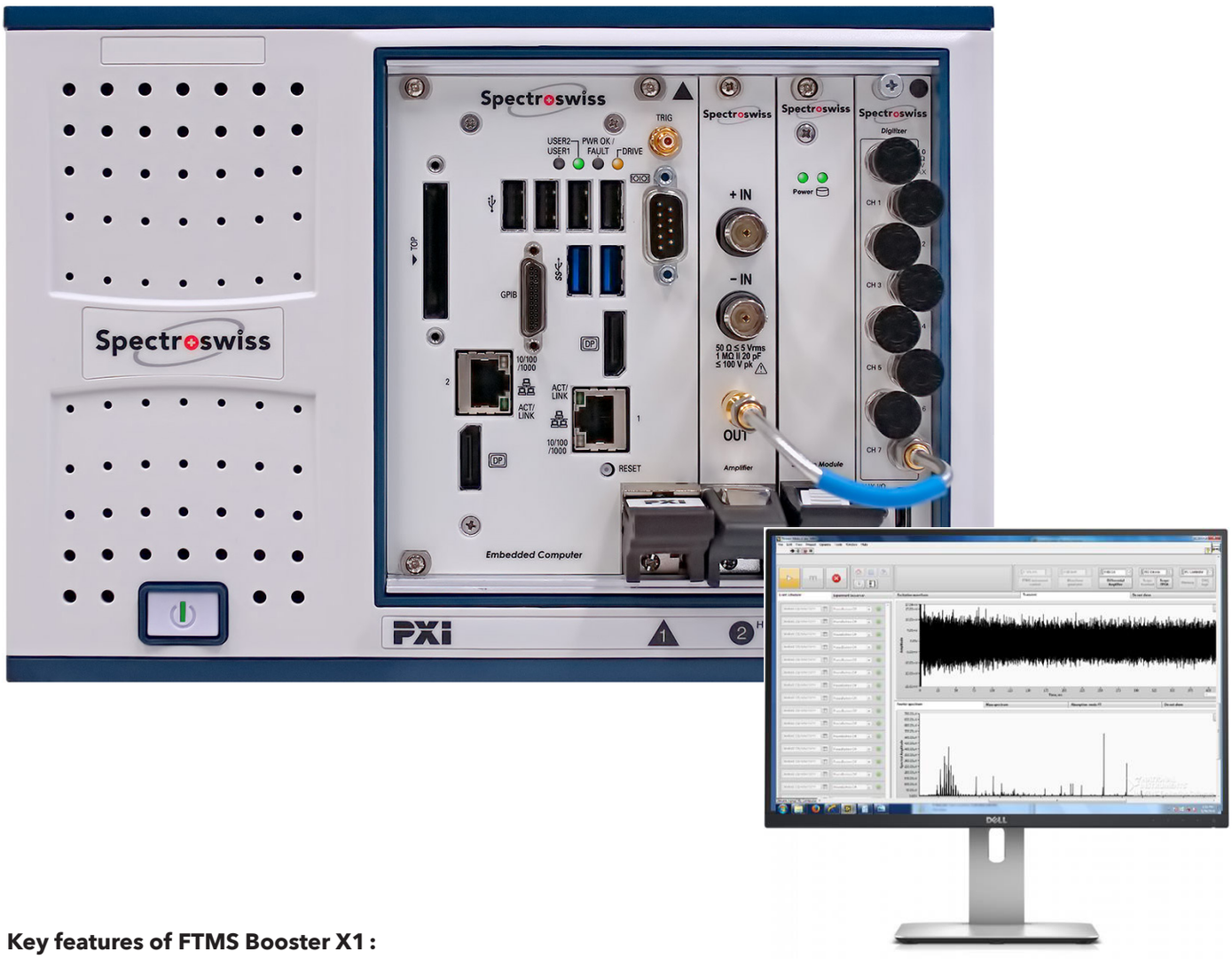
Key features of Peak-by-Peak:

- integration and seamless processing of .RAW files, .d folders and .HDF5 files;
- on-demand file formats for transient & mass spectra, including: HDF5, MGF, mzXML, mzML, imzXML;
- visualization of transients and mass spectra, visual comparison from separate data sets;
- spectral and transient summation for high sensitivity and dynamic range: data averaging across multiple LC-MS runs and across very long data sets, e.g., 10'000 scans;
- basic transient signal processing: discrete FT with parallel, multi-CPU computing;
- advanced signal processing: FT with pre- and post-processing as required in FTMS, including signal apodization, zero-fillings, magnitude mode FT spectral representation;
- noise thresholding to distinguish between analyte and noise, reduced-profile mode;
- peak picking with N-points interpolation and efficient baseline correction;
- mass scale calibration equations: (i) custom and standard calibration laws, (ii) pre-sets of calibration equations for FT-ICRs with given magnetic field strengths; (iii) file input-output operations with mass calibration parameters (supported file formats are: .DAT, FID/SER); (iv) iterative re-calibration procedure for petroleomics;
- handling of (very) large amounts of data;
- parallel computing on a given number of CPU-cores for fast data processing;
- advanced technical capabilities: cross-platform code supporting Windows, Linux, Mac OS;
- rapid development of custom software applications by the user.

peakbypeak
FTMS DATA ANALYSIS
by spectroswiss

FTMS Booster X1 : Data Acquisition System

Provides access to transients from any FTMS instrument
Improves performance and productivity of your FTMS instrument



Key features of FTMS Booster X1 :

- hardware interface to FTMS instruments, for example LTQ FT series, LTQ Orbitrap™ series, Orbitrap™ Fusion™ series and Q Exactive™ Orbitrap™ series instruments ;
- works in parallel with standard commercial data acquisition software, e.g., XCalibur™ ;
- real-time visualization of transients and mass spectra with or without data summation ;
- high-resolution and sensitivity (dynamic range) data acquisition due to low-noise RF front-end electronics and pre-programmed high-throughput FPGA chip with on-the-fly digital signal processing enabling full transient detection and pure absorption-mode FT on Orbitrap™ platforms or magnitude-mode FT on FT-ICR MS platforms ;
- rapid data transfer for measurements of large data sets (e.g., mass spectra in full profile mode), high duty cycles (e.g., fast FTMS instruments), or continuous recording of long transients (e.g.,> 50 s);
- file format of output data (transients or mass spectra): HDF5 ;
- high-capacity (2 TB), fast SSD buffer for transient signals and mass spectra, with on-line interface for streaming data to the customer's file server(s).

Data Processing for LTQ FT

Data summation using transients compared to the reduced profile spectral summation improves sensitivity and provides up to 3-fold more analyte peaks per mass unit in a wider mass range, Figures 6 and 7. Data acquisition was performed using FTMS Booster X1 and data processing via Peak-by-Peak.

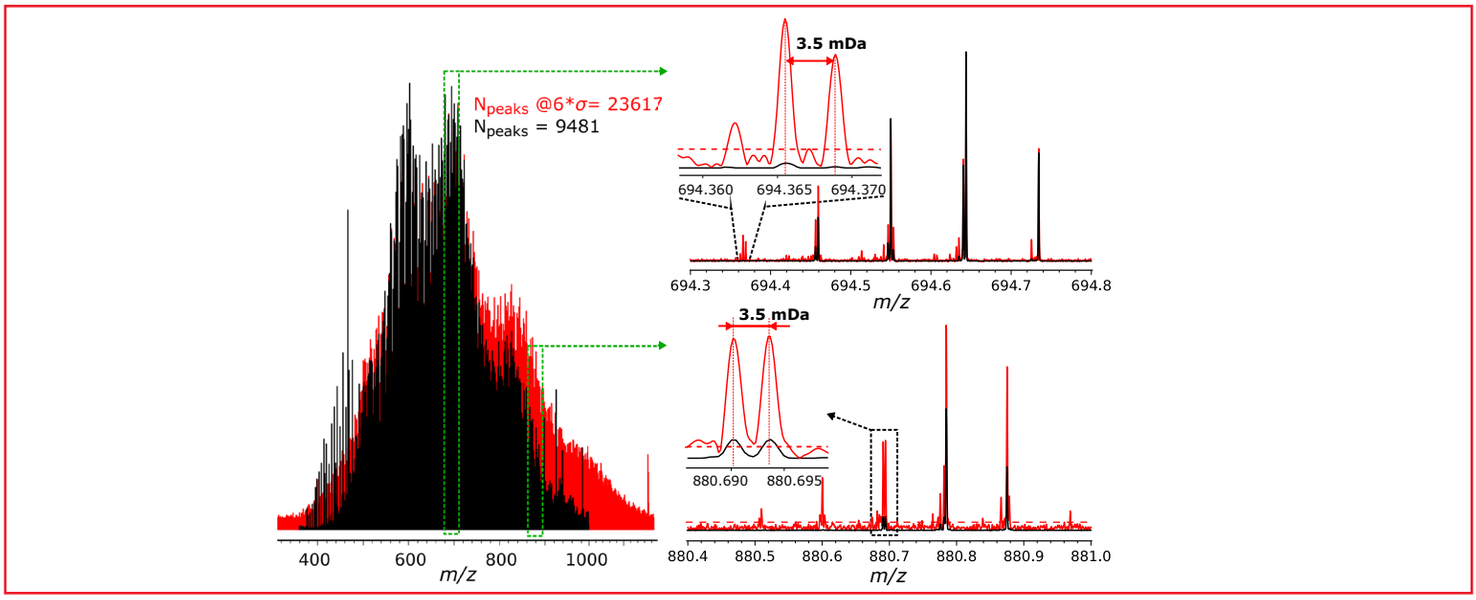


Figure 6. Positive APPI FT-ICR MS (10 T LTQ FT™, Thermo Scientific) of a crude oil fraction. **Black** : reduced profile spectral summation (.RAW, Xcalibur™). **Red** : transient summation (.H5, FTMS Booster X1). Both mass spectra are represented in a magnitude mode FT and normalized to the same base peak. Preset resolution R=750,000 @ 400 *m/z*, AGC=off, IT=100 ms, 10 microscans per scan.

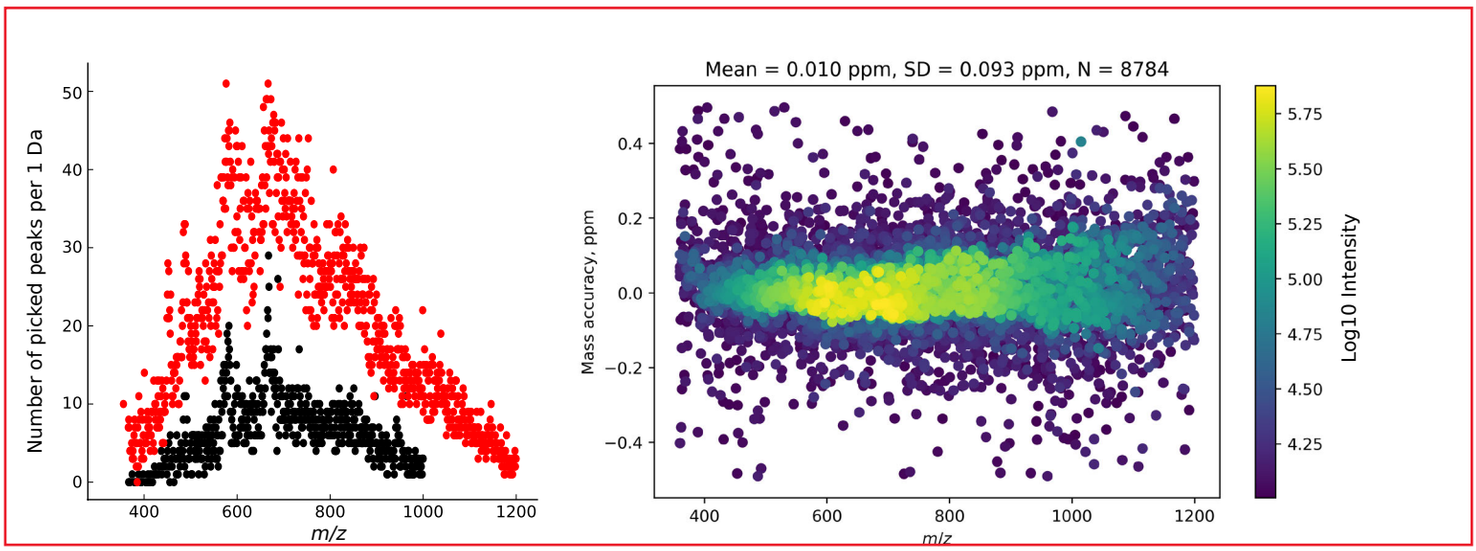


Figure 7. Results of a data analysis of mass spectra shown in Figure 6 : (left) number of analyte peaks of a crude oil fraction with the abundance higher than $6*\sigma_{noise}$ for (**red**) transient and (**black**) reduced profile spectral summation picked in the 1 Da (1 *m/z*) mass windows spreading the mass range of 300-1200 *m/z*. (Right) mass error distribution calculated for 8784 monoisotopic peaks identified in the mass spectrum represented in absorption mode FT (via AutoVectis) and obtained with transient summation.

Data Processing for FT-ICR MS

Absorption mode FT spectral representation of a 3-fold shorter transient compared to a magnitude mode FT of the original transient provides similar resolution performance at the improved sensitivity (more peaks detected) and higher throughput, Figure 8. Data processing via Peak-by-Peak and AutoVectis.

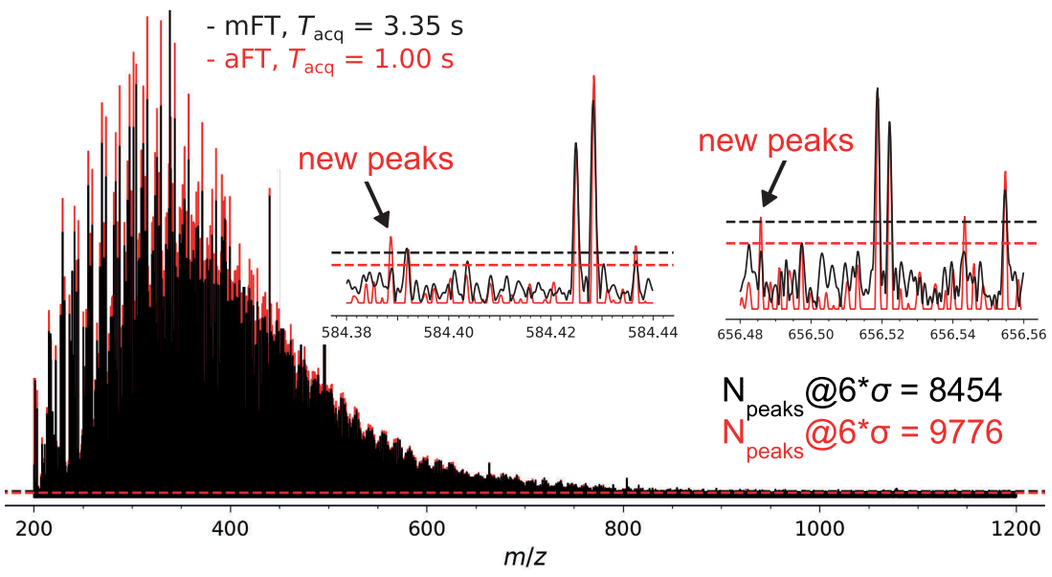


Figure 8. Positive ESI FT-ICR MS of a heavy oil sample. A summed transient containing 100 scans (fid) was acquired with a 12 T SolariX™ FT-ICR MS (Bruker Daltonics) instrument and processed using Peak-by-Peak and AutoVectis. (**Black**) the transient with an original length of 3.3 s was processed in a magnitude mode FT and (**red**) the transient was first cropped down to 1.0 s and represented in absorption mode FT.

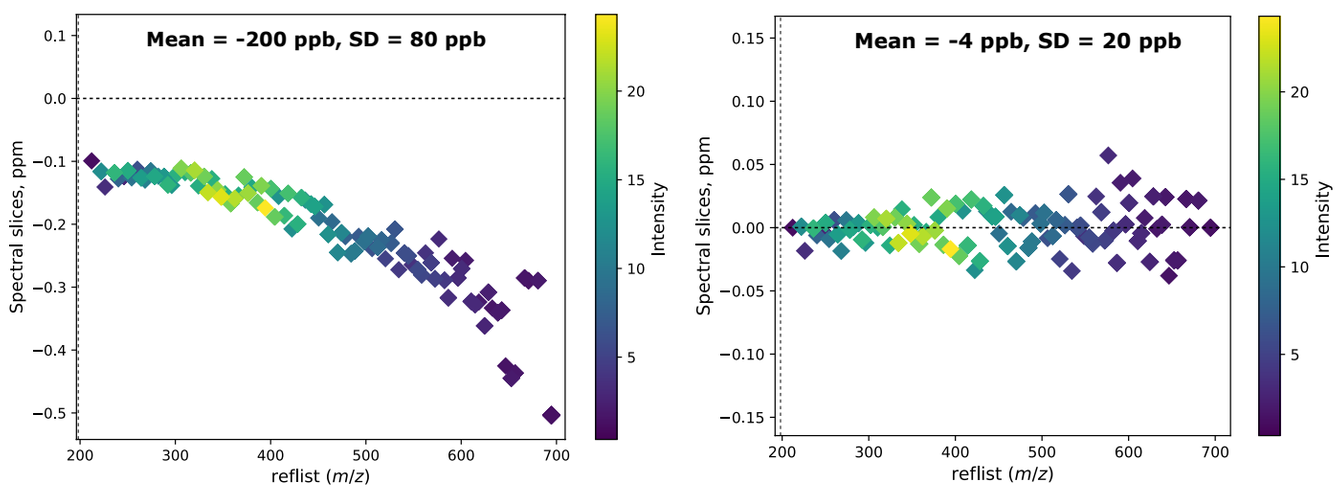


Figure 9. Application of an iterative re-calibration procedure (Peak-by-Peak) to a positive ESI FT-ICR MS mass spectrum of a heavy oil sample presented in Figure 3. The mass error distributions are (left) before and (right) after iterative re-calibration of a mass spectrum obtained from the summed 3.3 s transient (magnitude mode FT). The re-calibration procedure is described in: Kozhinov, Zhurov, and Tsybin; Iterative method for mass spectra recalibration via empirical estimation of mass calibration function for FTMS-based petroleomics. *Analytical Chemistry* (2013) 85, 6437-6445

Multiplexed Data Processing : Sensitivity

Gas chromatography/mass spectrometry (GC/MS) and liquid chromatography/mass spectrometry (LC/MS) are routinely used for identifying volatile and non-polar/polar components in crude oils. However, the sensitivity of GC/MS or LC/MS FTMS experiments applied to a complex mixture analysis may be poor for low-abundant components due to a limited number of scans per elution peak. The conditional scan summation through a number of separate GC/LC MS experiments (repetitions) is an attractive opportunity for improving the measurement sensitivity. Multiplexed transient or spectral summation for a targeted elution peak is enabled in Peak-by-Peak software.

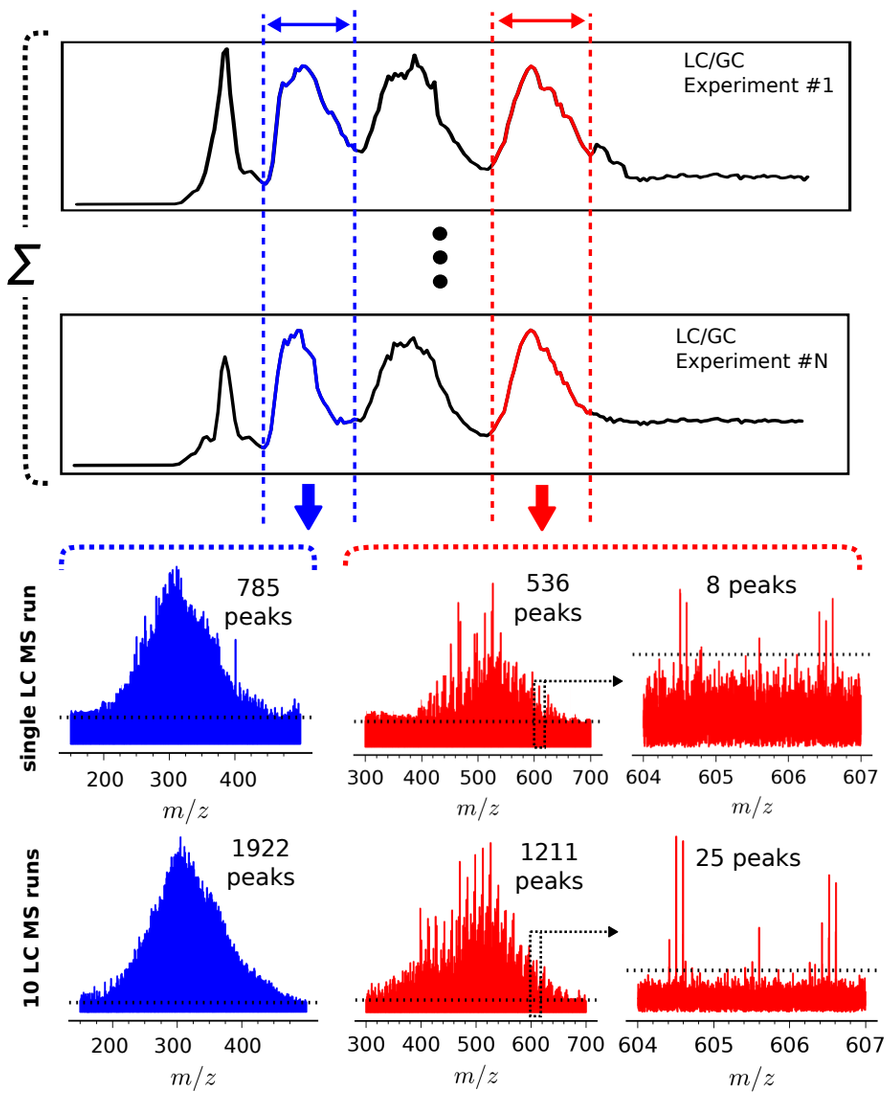


Figure 10. Improving sensitivity in the FTMS measurements via multiplexed summation of data (spectral or transient) across multiple chromatographic experiments (GC or LC hyphenated with MS). Data summed over a single GC/LC MS run is compared to data summed over 10 GC/LC MS runs.

A word from our customers

Dr. Florian Albrieux
French Institute of Petroleum (IFPEN), France

«We have been using Peak-by-Peak software since 2016 for processing and analysis of spectral and transient data acquired on our 7 T LTQ FT from petroleomic and ex-biomass samples. Peak-by-Peak enables summation of a large number of transients to provide an exceptional data quality, specifically in terms of a dynamic range and S/N, both in direct infusion and in chromatographic modes. In addition, it allows to perform conditional co-adding of transients based on frequency analysis, iterative re-calibration for improved mass accuracy, and provides flexible and tailored solutions as a function of a specific workflow employed. All in all, it dramatically increases the method efficiency for our petroleomics-type applications.»

Spectroswiss

EPFL Innovation Park | Building I | 1015 Lausanne | Switzerland
T: +41 21 693 98 06 | info@spectroswiss.ch
www.spectroswiss.ch