Detailed Remote Processing Documentation

VERSION: V1.0 - DATE: 22.NOVEMBER.2021

AUTHOR: LC. BAPTISTA FROM CASC4DE IT DEPARTMENT

CONTACT EMAIL: LUIS.BAPTISTA@CASC4DE.EU

Process Tool program

This program allows to process and analyze **MS FTICR** data-sets, i.e., raw transients as obtained from the FTICR machine.

The program allows to process transients, detect peaks, interact with the spectrum, and store the final result. A signed audit travail of the processing is maintained, in order to insure a complete reproducibility of the process.

There is no theoretical limit on the size of the data-set to process and visualize.

The result of the processing is stored in a `*.msh5` file, with the same filename than the Bruker directory. These files use the standard HDF5 format, and can be read with any program able to access this open format.

Only files in the seafile deposit can be handled.

Simply close the window to exit the program

Standard Operations:

Choose a file

Only files in the SeaDrive deposit can be handled. Use the selector to choose an experiment. Bruker `experiment .d` contains the raw transients, and `.msh5` files are previously processed stored data.

Load

The Load button will get the transient of the selected experiment and display it. Any previous processing will be lost.

Process

Process computes the Spectrum, according to the parameters define in the `Processing Parameters` panel

Peak Pick

`Peak Pick` computes the Peak list, according to the parameters define in the `Processing Parameters` panel. A peaklist file is created in the initial `experiment .d` directory, in `html` and in `csv` formats. An excerpt of the peak list is displayed in the `Peak list` panel.

Peaks are then displayed with the `Show Peaks` button, choosing the color and the label in the Spectrum panel. The number of displayed peaks is limited by the parameter set in the Processing Parameters panel.

Peaks are drawn only in the current displayed zoom. You can redraw several times, allowing to build specific figures.

Save

`Save` stores a `.msh5` file in HDF5 format into the initial `experiment .d` directory

Exit

Simply close the window to exist the program

Tab Panels

- `raw fid`: the transient, if loaded
- `Spectrum`: the processed spectrum, if computed
- `Peak list`: the peak list, if computed - also exported in csv format into the dataset directory
- `Calibration`: to change the calibration values
- `Simulation`: a spectral simulator, able to compute (fine) isotopic patterns at a given Resolving Power
- `Processing Parameters`: all the parameters used for the processing
- `Info`: details on the experiment and Processing audit trails

Processing Parameters

A few processing parameters can be adapted:

- **center fid**: default: Yes
  because of an eventual offset in the electronic, the experimental 'FID' might not be centered on the null value, but on some artefactual value. Removing this artefact before Fourier Transform may improve the spectral quality
- **apod todo**: default: hamming
  The choice of the apodisation (windowing) function:
  - hamming: provides good resolution and SNR at the price of wiggles at the feet of large peaks
  - hanning: provides low wiggles at the price of lower resolution and SNR
  - kaiser: a family of parametric windows of optimized apodisation,
    - 3.5: best possible resolution and SNR at the price of wiggles at the feet of large peaks
    - 5: Similar to Hamming

    - 6: Similar to Hanning
    - 8: very low wiggles, for resolution and SNR close to Hanning
- **baseline todo**: default: offset
  removes the offset on the final spectrum - center the noise - but create unrealistic negative values
- **grass noise todo**: default: "Only when storing file"
  Set points below a certain level (see grass noise level below) to zero. Allows to efficiently compress saved dataset, but loses information
- **grass noise level**: default: 3.0
  the level for "grass noise" removal is taken as the standard deviation of the baseline signal (σ) multiplied by this coefficient. So default is 3σ.
- **peakpicking**: default: Manually
  when to do peak picking, on demand or automatically after FT
- **peak picking noise level**: default: 10.0
  when doing peak picking, all peaks above a threshold expressed as a ratio above the noise level σ is detected. So the default value is set at 10σ (which is quite low).
- **centroid**: default: No
  after peak picking, a centroid is computed, it permits to improve the value of the position, and estimates its width.
- **max peak disp**: default: 200
  when many peaks are detected, the display of all the peaks becomes very slow, so only this many peaks are shown, showing only the highest peaks. Will be redisplayed for each new zoom window. To adapt to your local set-up.

Display

Figures can be interactively explored with the jupyter tools displayed on the side of the dataset.

- Zoom with 
- shift and resize  (with left and right click)
-  and  allow to navigate in the zoom history

The drawing zone can be resized using the little gray triangle on the lower-right corner

Figures can also be saved as a `png` graphic file with 

Calibration

The calibration used by SPIKE is based on a 2 or 3 parameters equation:

$$freq = A / (m/z) - B + C / (m/z)^2$$

where A B and C are imported from the Bruker `ML1` `ML2` `ML3` parameters.

Be careful Bruker uses a sign inversion on `ML2` depending on the value of `ML3`

- this is not used, and the equation used for calibration is always the equation above, even if C is 0.

Nevertheless, the equation above is not optimal, and unstable in certain conditions. We are working on a better definition, which will allow more robust calibrations and this set-up will be changed in the future for a more flexible and robust set-up

LCMS Tool program

This program allows to analyse **LC-MS FTICR** data-sets, i.e., a series of MS spectra acquired during a chromatography run. The information is bi-dimensional, with one chromatographic axis and one MS axis.

The program allows to look at the whole data at once, as well as extracting MS spectra at a given retention time, or a chromatogram extracted at a given *m/z*.

The raw data `ser` *(series of transients)* has to be processed before being handled here. The processing is performed in background on the deposit system, and a few hours should be expected for the processing to be performed. There is no theoretical limit on the size of the data-set to process and visualize. The result of the processing is stored in a `*.msh5` file with the same filename than the Bruker directory. These files use the standard HDF5 format, and can be read with any program able to access this open format.

Only files in the Seafile deposit can be handled.

Tab Panels

There several tab panels that play different roles

1D Extraction

Slices from the complete experiment can be computed and displayed here,

- On the **MS** line click on get to get a MS spectra extracted at the retention time given by the slider - labelled in minutes *(you can also type directly the value)*.
- On the **LC** line click on get to get a chromatogram of the *m/z* peak location given by the slider *(you can also type directly the value)*. If needed the chromatogram can be smoothed for better looking using a Savitsky-Golay method - 0 means no smoothing.
- both extractions can be summed other a small region around the given location - which width is given by the second slider.

for both type of dataset, the Peak Pick button computes the position of peaks of the dataset currently displayed, and stores it as a CSV file in the project folder.

The Save button stores the content of the window, as a `.msh5` file for MS spectra or in CSV format for chromatograms.

2D spectrum:

A 2D view of the LC-MS experiment, displayed as a contour map.

To speed-up the display, a low resolution of the spectrum is displayed when a large zone of the experiment is displayed, resolution is optimized after zooming-in. The resolution being used is displayed on the top left of the 2D map, (see in the `Info` Panel for the different resolution of the dataset.) and R is an estimate of the maximum resolving power available in the center of the zoom window (computed from the sampling of the *m/z* axis rather than from the actual peak width!).

The zoom box on the top shows the current zoom limits, which can be modified at will. The Apply button activates the entry.

The vertical slider chooses the values at which the levels are drawn, and the Redraw and Reset buttons will recompute the display, `Redraw` with the current zoom and scale parameters, and `Reset` with the default parameters.

Peak list

Display the last computed peak list from the 1D panel.

Info

Display details on the experiment. In particular the different resolutions layered in the document are presented.

Standard Operation:

Choose a file

LC-MS raw data are stored in the `ser` file in Bruker directories. They have to be processed before being visualized with this utility.

The selector will present only those `.msh5` files which are available for analysis.

There is no limit on the size of the dataset to be explored. *During the development, tests where performed on a 1800 × 4096k experiment (1800 spectra of 4096k length) and the program was reasonably swift.*

However, the very first time you read an experiment, it may take some time to load the cache (1-2 minutes) - be patient -

Explore the experiment in 2D mode

Using the interactive tools available in the 2D panel (see above)

The TIC profile on the right and the MS total spectrum on the top allow to precisely locate the signals of interest.

Extract spectra and chromatogram

Using the 1D panel, you can look at a MS spectrum for a given retention time; or extract a chromatographic profile associated to a *m/z* value, or a range of values.

You can realize a pick-peaking as well as store the displayed dataset.

`.msh5` files created from the MS spectra stored with this tool can then loaded back into the program `Process_Tool`.

Exit

Simply close the window to exit the program

Display

Figures can be interactively explored with the jupyter tools displayed on the side of the dataset.

- Zoom with 
- shift and resize  (with left and right click)
-  and  allow to navigate in the zoom history

The drawing zone can be resized using the little gray triangle on the lower-right corner

Figures can also be saved as a `png` graphic file with 

Calibration

The calibration used by SPIKE is based on a 2 or 3 parameters equation:

$freq = A / (m/z) - B + C / (m/z)^2$

where *A* *B* and *C* are imported from the Bruker `ML1` `ML2` `ML3` parameters.

Be careful Bruker uses a sign inversion on `ML2` depending on the value of `ML3`

- this is not used, and the equation used for calibration is always the equation above, even if C is 0.

Nevertheless, the equation above is not optimal, and unstable in certain conditions. We are working on a better definition, which will allow more robust calibrations and this set-up will be changed in the future for a more flexible and robust set-up

2D Tool Program

This program allows to analyse **2D-MS FTICR** data-sets. Only processed 2DMS files in the Seafile deposit can be handled.

Select the file you want look at, and Load, it will show-up as a full width 2D image. *The very first time you read a 2D experiment, it may take some time to load the cache (1-2 minutes) - be patient -*

2D Spectrum

Is shown as a full width 2D image.

- the F2/horizontal axis is the high resolution, direct axis. You find fragment ions along this line.
- the F1/vertical axis indirect axis, usually at lower resolution. You have parent ions along this axis.
- the top and right spectra show the diagonal of the experiment *($m/z_{F1} = m/z_{F2}$)* . It should be equivalent to the complete parent + fragment spectrum.

What you are seeing is the contour plot of the 2D experiment.

> *Think of the experiment like a landscape filled with water (an archipelago of small islands), and you are seeing a map of this continuous landscape. If a dot appears, it means that it is higher than the sea level.*

- with the `scale` slider, you can select the levels at which the contours are drawn.
  - *(You raise or lower the sea "level")*
  - `Reset` gets back to 1.0
  - `Redraw` is sometime needed and recompute the map
- `Side spectra scale` changes the size of the side spectra
- you can zoom by dialing limit values in the `Zoom Box`
- `Highest displayed mass` limit the highest displayed mass in the zoom box

The data has a hierarchical multiresolution structure. When zooming, the `Redraw` button loads the version of the data with the optimal resolution. The smaller the zoom box, the better the resolution. `#1` means you have the full resolution; higher numbers are lower resolution.

Classical problems

- **If you are seeing a nearly empty image**
  - try using the vertical slider on the left to adjust the viewing scale
- **If you are seeing just vertical lines**
  - it probably means your experiment is dominated with scintillation noise,
    - try reprocess with a more SANE iterations and a small SANE rank
    - try acquiring with a spray as stable as possible
- **If you are seeing many closely related diagonals, and nearly empty top and right spectra**
  - this is usually the case when a wrong demodulation was applied, the correct value should be the pulse frequency at the end of the P1 pulse.

Tab Panels

There several tab panels that play different roles

1D Extraction

Slices from the complete experiment can be computed and displayed here,

- a horizontal slice, extracted at the *m/z* of a parent ion will show the spectrum of all the fragments of this parent.
- a vertical slice, extracted at the *m/z* of a fragment ion will show the spectrum of all the parents which may generate fragments at this *m/z*.

Be careful that isotopic patterns are not aligned to the horizontal or vertical lines but are skewed, with an angle *z1/z2* relative to the main diagonal. (Where *z1* - respectively *z2* are the charges of the ion on axis1 - resp. axis2). This will distort the spectra you extract along the main axes.

Peak list

is not active yet.

Info

Display details on the experiment. In particular the different resolutions layered in the document are presented.

Standard Operation:

Choose a file

2D-MS raw data are stored in the `ser` file in Bruker directories. They have to be processed before being visualized with this utility.

The selector will present only those `.msh5` files which are available for analysis.

There is no limit on the size of the dataset to be explored.

Explore the experiment in 2D mode

Using the interactive tools available in the 2D panel (see above)

Exit

Simply close the window to exit the program

Display

Figures can be interactively explored with the jupyter tools displayed on the side of the dataset.

- Zoom with 
- shift and resize  (with left and right click)
-  and  allow to navigate in the zoom history

The drawing zone can be resized using the little gray triangle on the lower-right corner

Figures can also be saved as a `png` graphic file with 

Calibration

The calibration used by SPIKE is based on a 2 or 3 parameters equation:

$freq = A / (m/z) - B + C / (m/z)^2$

where *A* *B* and *C* are imported from the Bruker `ML1` `ML2` `ML3` parameters.

Be careful Bruker uses a sign inversion on `ML2` depending on the value of `ML3`

- this is not used, and the equation used for calibration is always the equation above, even if C is 0.

Nevertheless, the equation above is not optimal, and unstable in certain conditions. We are working on a better definition, which will allow more robust calibrations and this set-up will be changed in the future for a more flexible and robust set-up