Introduction

Finnee is a Matlab toolbox that has been designed to help processing data that has been acquired using a separation technique hyphenated with high-resolution mass spectrometry. Finnee is aimed specifically for MS1 data: the input format is mzML with MS profile scans. Finnee is a computer assisted MS analysis tool (CAMSAT) that use a different approach than all other existing tools (MZmine, KPIC, xcms…). The core principle of Finnee is to start with MS scans recorded as profile and perform various correction (axis normalisation, baseline drift correction, noise reduction…) before running the centroid algorithms, this allows to drastically decrease the size of the file, improve the accuracy of the accurate masses and facilitate the extraction of the (pure) ion profiles from the data. Finnee used object oriented programming (OOP). Within the same Finnee object, each transformation will generate a new dataset, where a dataset will contain all the scans related to the run. This sheet contains the optimised parameters to process the urine dataset (https://data.mendeley.com/datasets/6rn82jdv8d/1). For more information see https://finneeblog.wordpress.com or https://github.com/glerny/Finnee2016/wiki .

Datasets and methods associated

Step 1: Creations of the Finnee object

Command: > myFinnee = Finnee('overwrite')

If not enter as optional parameters (see help Finnee), the address to the mzML files, the generic name and output folder will be asked.

Step 2: Normalisation of every MS scan to a common (master) mz axis

Command: > myFinnee = myFinnee.align2newMZ(1, [], 'masterAxis', 'mz1:mz2:Id')

Example: > myFinnee = myFinnee.align2newMZ(1, [], 'masterAxis', 50:1000:max')

This command will create a new mz axis based on the parameters 'mz1:mz2:max' and interpolate the intensity of every MS scans to this new axis. mz1 and mz2 are the starting and final mz values respectively, and Id is the index of the scan that will be used as model (used max to use the most intense scan)

Step 3: Correction for baseline drift

Command: > myFinnee = myFinnee.BaselineCorrection(2)

This command allows correcting chromatographic profiles for baseline drift, where each chromatographic profile is the variation of the intensity as a function of time at a given m/z interval. The command makes use of two graphic user interfaces (GUI) where the first one allows selecting the profiles that need correction (typically you should select a value between 40 and 60%), and the second allows selecting the baseline model and optimizing the parameters (fastest function arPLS2).

Step 4: **Noise removal**

Command: > myFinnee = myFinnee.filterDataset(3, 'RemoveNoise:5:3:10:f');

This command allows to remove background noise as defined by the string of character after '…:5:3:10f'.

Step 5: **Centroid scans**

Command: > myFinnee = myFinnee.doCentroid(4, 'LocalMax:2:0');

This command allows to transform profile MS scans to centroid MS scans using the LocalMax algorithm.

Step 6: **Peak list**

Command: > myPeakList = PeakList(myFinnee.Datasets{5}, 0, 0.002, 10);

This command allows extracting pure ion profiles from the datasets 5. In this case pure ion profiles are recognised as any series of 10 or more points that does not differ by more than 0.002 in their mz values.

Traces and methods associated

As previously explained, every transformation will create a new dataset. Information in those datasets can be visualised using classical 2D representation (denominated as a trace) such as the base peak profile (BPP) or total ion profile (TIP). For example the command

- myFinnee.Datasets{x}.TIP.plot;

will plot the TIP linked with the x dataset (from previously 1: original data; 2: mz normalized data, 3: baseline corrected data; 4: noise reduced; 5: centroid dataset). It is also possible to obtain spectra between different time interval

- myFinnee.Datasets{x}.getSpectra([t_i t_f]).plot;

or a profile calculated using only the mz values between a given interval:

- myFinnee.Datasets{x}.getProfile([mz_i mz_f]).plot;

PeakList and methods associated

The PeakList is the final result of Finnee it contains the list of profiles that were automatically extracted from a single dataset. All figures of merits are accessible using

- myPeakList.FOM{1}.Headings
- myPeakList.FOM{1}.Data

each profile can be plotted using

- myPeakList.LstPIP{1}{Ix}.plot

where Ix is the index to the profile (first column in myPeakList.FOM{1}.Data). The centergram representation can be obtained using

- myPeakList.Centergram

and the cluster plot representation using

- myPeakList.ClusterPlot(0.95)