SHIFTS-4

Every SHIFTS module uses configuration files (INI), but some of the parameters it contains can also be specified through the command line. Run each module with the help option, ***-h***, (e.g. `python DMcalibrator.py -h`) for a list of parameters that are available through the command line and their description.

1. SHIFTSadapter

This module adapts a Comet-PTM or Recom file so that it can be analyzed with SHIFTS.

- Input:
  - A tab-separated file from Comet-PTM or Recom output (with header).
  - A configuration file (INI). There is a default INI in the "config" folder.
- Output:
  - A tab-separated file (without header).
  - A log file containing the header information.

2. DMcalibrator

This module calculates the calibrated values for experimental masses and deltamasses. An 'alpha' value is calculated as the median of the absolute error column divided by the MZ column, using the dataset filtered by minimum score and maximum ppm:

alpha = (df_filtered[abs_error]/df_filtered[mzcolumn]).median()

This value is used to calibrate the experimental MZ values as follows:

*cal_exp_mz = exp_mz * (1 - alpha)*

- Input:
  - A file from SHIFTSadapter output.
  - A configuration file (INI). There is a default INI in the "config" folder.
  - A mass file (INI). There is a default INI in the "config" folder.
- Output:
  - The same file with additional columns for theoretical MH and MZ, calibrated MH and MZ, calibrated deltamass MH and MZ, absolute error, ppm error, sequence with calibrated deltamass.
  - A log file containing systematic error before and after calibration, alpha, StdDevMAD_ppm, number of PSMs before and after filtering.

3. PeakModeller

This module concatenates a group of files, generates a histogram grouped by deltamass bins of a user-specified width, and calculates the frequency and slopes (first and second derivatives) for each bin, using linear regression.

- Input:
  - A text file containing a list of calibrated files to be modelled together.
  - A configuration file (INI). There is a default INI in the "config" folder.
- Output:

- A tab-separated file containing all the input files together (DMTable)
- A tab-separated file containing the histogram (DMHistogram)
- A log file.

4. PeakSelector

This module filters a histogram of deltamasses according to user-specified thresholds for slope and frequency, and calculates a list of apexes using the given number of points. Using the *x* (bin midpoint) and *y* (first derivative) values for those points it will identify the apex as the *x* value where *y* equals 0, using interpolation.

- Input:
  - A DMHistogram to be filtered.
  - A configuration file (INI). There is a default INI in the "config" folder.
- Output:
  - A text file containing the apex list.
  - A log file containing the number of apexes that were calculated.

5. PeakAssignator

This module will assign every PSM to the closest peak from the provided apex list and identify it as either belonging to that peak, or an orphan. To do so, it will calculate the absolute distance between the deltamass value and the assigned peak in ppm:

$$distance = abs(assigned\_peak - delta\_MH)$$

$$distance\_ppm = (distance / (theoretical\_mass + assigned\_peak)) * 1e6$$

If that value is larger than the user-defined threshold, it will be considered an orphan.

- Input:
  - A DMtable.
  - An apex list.
  - A configuration file (INI). There is a default INI in the "config" folder.
- Output:
  - A DMtable with additional columns for the closest peak, assignation as peak or orphan, assigned deltamass, ppm error, sequence with assigned deltamass.
  - A log file.

6. PeakFDRer

This module will calculate global, local, and peak FDR values for a DMtable subdivided by experiment. For the global FDR, rather than taking an entire experiment, this module will separate it in two deltamass regions (defined by the parameter ***dm_region_limit***, default value -56) and calculate a global FDR for each region.

- Input:
  - A DMtable.
  - A tab-separated file containing a column with the experiment name and another with the file path (must match the file paths in the ***Filename*** column of your DMtable).
  - A configuration file (INI). There is a default INI in the "config" folder.
- Output:
  - A DMtable with additional columns for the global, local and peak FDR ranks and values.
  - A log file.