



## MZmine 3

---

**Documentation**

*The MZmine Community*

*None*

## Table of contents

---

<b>1. Welcome to the MZmine 3 wiki!</b>	4
<b>1.1 What's new compared to MZmine 2?</b>	4
<b>1.2 About this documentation</b>	4
<b>1.3 How to contribute</b>	4
<b>2. Getting Started</b>	5
<b>3. Main window overview</b>	6
<b>3.1 MS data files and feature lists tab</b>	6
<b>3.2 Main content pane</b>	6
<b>3.3 Main menu</b>	6
<b>3.4 Task overview</b>	6
<b>3.5 Page Contributors</b>	7
<b>4. Processing modules</b>	8
<b>4.1 Data import</b>	8
<b>4.2 Mass detection</b>	10
<b>4.3 Mobility scan merging</b>	14
<b>4.4 ADAP chromatogram builder</b>	15
<b>4.5 IMS Expander</b>	17
<b>4.6 Ion mobility trace builder</b>	18
<b>4.7 Smoothing</b>	20
<b>4.8 Local Minimum Resolver</b>	22
<b>4.9 CCS Calibration and calculation</b>	27
<b>4.10 Gap-filling</b>	30
<b>4.11 MS2 Scan Pairing</b>	31
<b>4.12 <math>^{13}\text{C}</math> isotope filter (Isotope grouper)</b>	33
<b>4.13 Isotope pattern finder</b>	35
<b>5. Workflows</b>	37
<b>5.1 LC-MS Workflow</b>	37
<b>5.2 LC-IMS-MS Workflow Overview</b>	39
<b>5.3 Batch processing</b>	42
<b>5.4 Processing wizard</b>	43
<b>6. Visualisation</b>	45
<b>6.1 Raw data visualisation</b>	45
<b>6.2 Ion mobility raw data overview (LC-IMS-MS)</b>	46
<b>6.3 MS/MS plot</b>	48

7. Additional resources	53
7.1 General terminology	53
7.2 MZmine-specific terminology	55
7.3 Ion mobility spectrometry terminology	57
7.4 Graphical comparison of LC-MS and LC-IMS-MS data	59
8. Performance options	60
8.1 Preferences	60
8.2 Logs	61
9. Command-line arguments	63
10. How to contribute	64
10.1 Contribute to the MZmine documentation	64
10.2 Creating a new page	66
10.3 Page Contributors	67
11. Acknowledgements	68
11.1 Related projects	68
11.2 Libraries we use in MZmine	68

# 1. Welcome to the MZmine 3 wiki!

---

MZmine 3 is an open-source and platform-independent software for mass spectrometry (MS) data processing and visualization. It enables large-scale metabolomics and lipidomics research by spectral preprocessing, feature detection, and various options for compound identification, including spectral library querying and creation.

Since the introduction of MZmine 2 in 2010, the project has matured into a community-driven, highly collaborative platform and its functions continue to expand based on the users' needs and feedbacks. This has also enabled the tight integration of the MZmine ecosystem with popular third-party software for MS data analysis, such as the [SIRIUS](#) suite for *in silico* metabolite annotation, the [GNPS](#) platform with Ion Identity Molecular Networking, the [MetaboAnalyst](#) web app for univariate and multivariate statistical analysis, *etc.*

Such a great progress was made possible by the invaluable contribution of many [developers](#) from research labs distributed all over the world!

Want to get started with MZmine 3? Check out our [getting started](#) page!

---

## 1.1 What's new compared to MZmine 2?

---

MZmine 3 comes with a redesigned and fully customizable [GUI](#) based on the JavaFX technology that allow an interactive visualization and validation of results from every processing step.

A completely new data structure provides the flexibility to process any type of mass spectrometry, including LC-MS, GC-MS and MS-imaging. Moreover, MZmine 3 now supports ion mobility, with a dedicated [LC-IM-MS data visualization](#) module and [feature detection](#) algorithms.

Finally, significant effort was devoted to trace memory issues and bottlenecks, resulting in an unprecedent processing performance and scalability.

**COMING SOON!** We are implementing the [Mass Spec Query Language](#) (MassQL) to explore your MS data with human-readable, succinct queries! The project is supported by the [Google Summer of Code](#) program.

---

## 1.2 About this documentation

---

Here you can find documentation for both processing and visualization modules in MZmine 3. Moreover, data processing pipelines for untargeted [LC-MS](#) and [LC-IMS-MS](#) feature detection are described and general recommendations are given.

**COMING SOON!** We are currently working on a series of short videotutorials to help get you started with the main features of MZmine 3!

---

## 1.3 How to contribute

---

The MZmine community is always welcoming new developers and contributions! You can contribute by improving existing modules or even adding new features in MZmine 3! Please, check out our brief [tutorial](#).

You can also contribute to this wiki and help new users to get started with MZmine 3! See [here](#) how to contribute to the documentation.

---

## 2. Getting Started

---

Before creating your first project, we recommend setting up some things.

1. Set a temporary file directory. Go to *Project* → *Set preferences* → *Temporary file directory*. This requires a restart to take effect.
  - a. We recommend setting the directory to an SSD with enough space for fast processing and visualizations.
  - b. On Windows, old temporary files are deleted when a new session is started.
2. MZmine 2 projects cannot be imported due to changes in the data structure.
3. MZmine 2 batch files cannot be imported due to parameter optimizations.

You can get familiar with the new GUI here: [Main window overview](#)

A quick insight to data processing workflows can be found here: [LC-MS workflow](#) or [LC-IMS-IMS workflow](#)

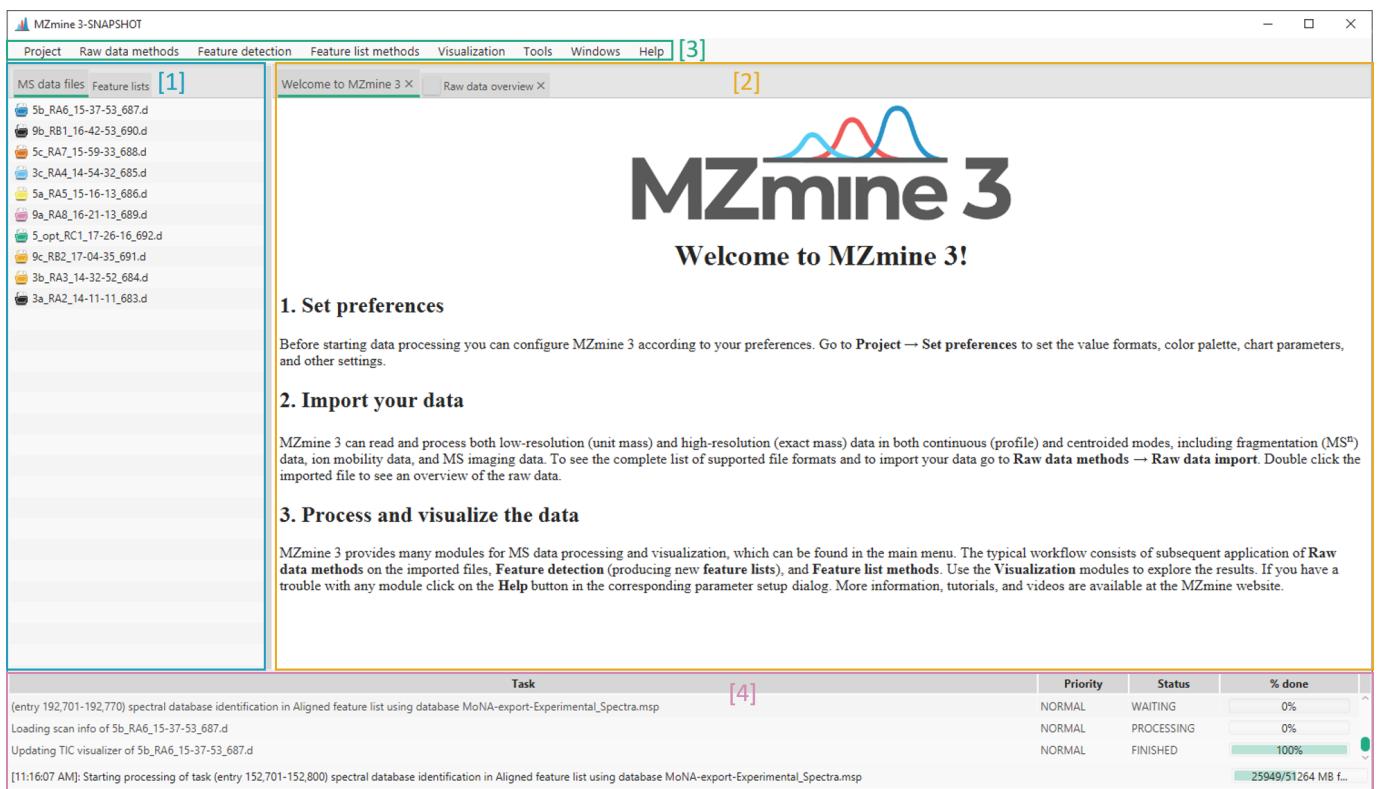
You can also check out the new processing wizard under *Processing wizard* in the main menu.

---

Last update: April 5, 2022 13:31:49

### 3. Main window overview

The MZmine 3 main window is made up of mainly four important building blocks.



#### 3.1 MS data files and feature lists tab

[1]: The (raw) ms data and feature list tabs. Here you can find your imported data files and processed feature lists. *Hint: you can also import files by dragging & dropping them to the ms data tab.*

#### 3.2 Main content pane

[2]: The main content pane. Visualisations such as a raw data overview or a feature list can be viewed here. This pane can also contain multiple tabs. Every tab can also be opened in a new separate window by right-clicking on the header.

#### 3.3 Main menu

[3]: The main menu. Here you can find methods to import and process your data files and feature lists and visualise the results. Furthermore, projects can be saved and preferences can be set.

#### 3.4 Task overview

[4]: The task overview. Current tasks are displayed and their status and progress are indicated. Tasks can also be canceled by right clicking on a task.

## 3.5 Page Contributors

---

**SteffenHeu**

---

Last update: April 5, 2022 13:22:07

## 4. Processing modules

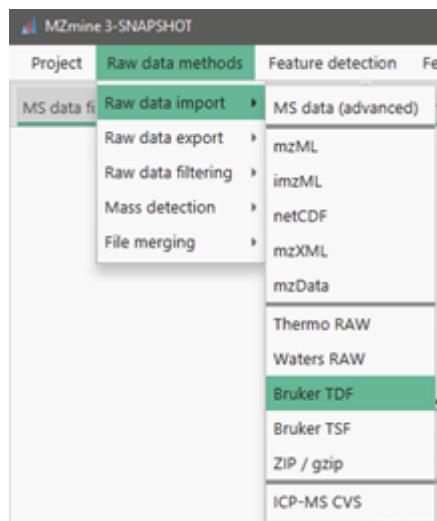
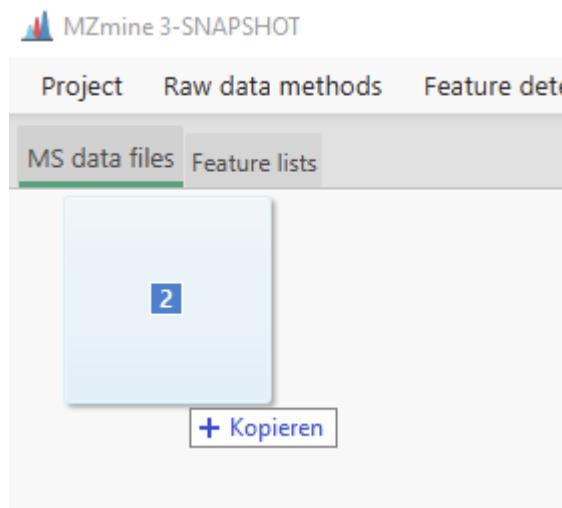
---

### 4.1 Data import

#### 4.1.1 LC-MS data

Raw data can be imported via the main menu **Raw data methods → Raw data import**. Note that multiple data files/folders can be dropped into the **MS data (advanced)** dialog. If individual modules are used, folder based formats can only be imported as one folder at a time. When using the **MS data (advanced)** dialog, inexperienced users should deactivate the direct mass detection steps, since they alter the raw data on the import. Mass detection is then performed, when the scans are loaded and only peaks above the noise level are imported.

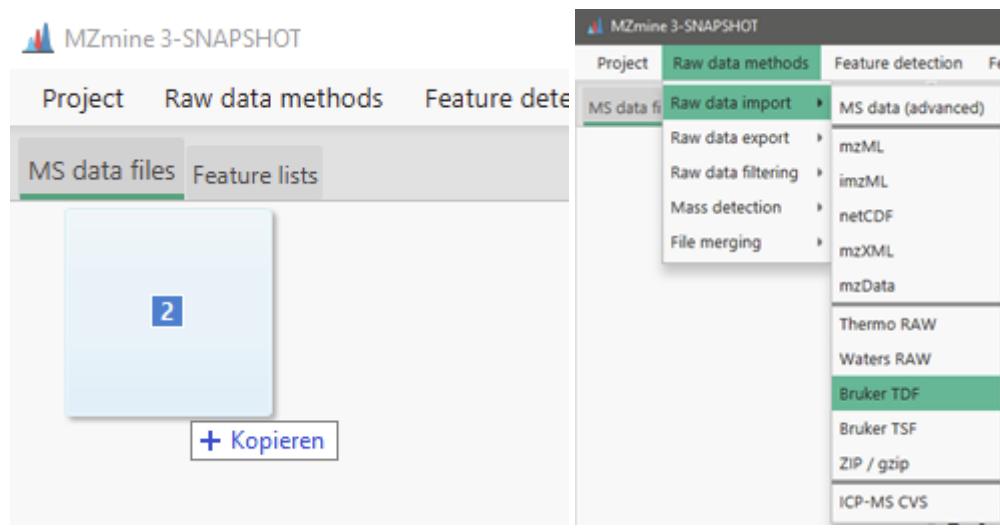
Alternatively, you can simply drag & drop the raw data into the raw data list of the main window.



#### 4.1.2 LC-IMS-MS data

As any other data format, ion mobility data can be imported via the main menu **Raw data methods → Raw data import**. Note that multiple .tdf data folders can be dropped into the **MS data (advanced)** dialog. The Bruker TDF import can only select a single folder. When using the **MS data (advanced)** dialog, inexperienced users should deactivate the direct mass detection steps, since they alter the raw data on the import. Mass detection is then performed, when the scans are loaded and only peaks above the noise level are imported.

Alternatively, you can simply drag & drop the raw data into the raw data list of the main window.



Last update: March 10, 2022 15:53:40

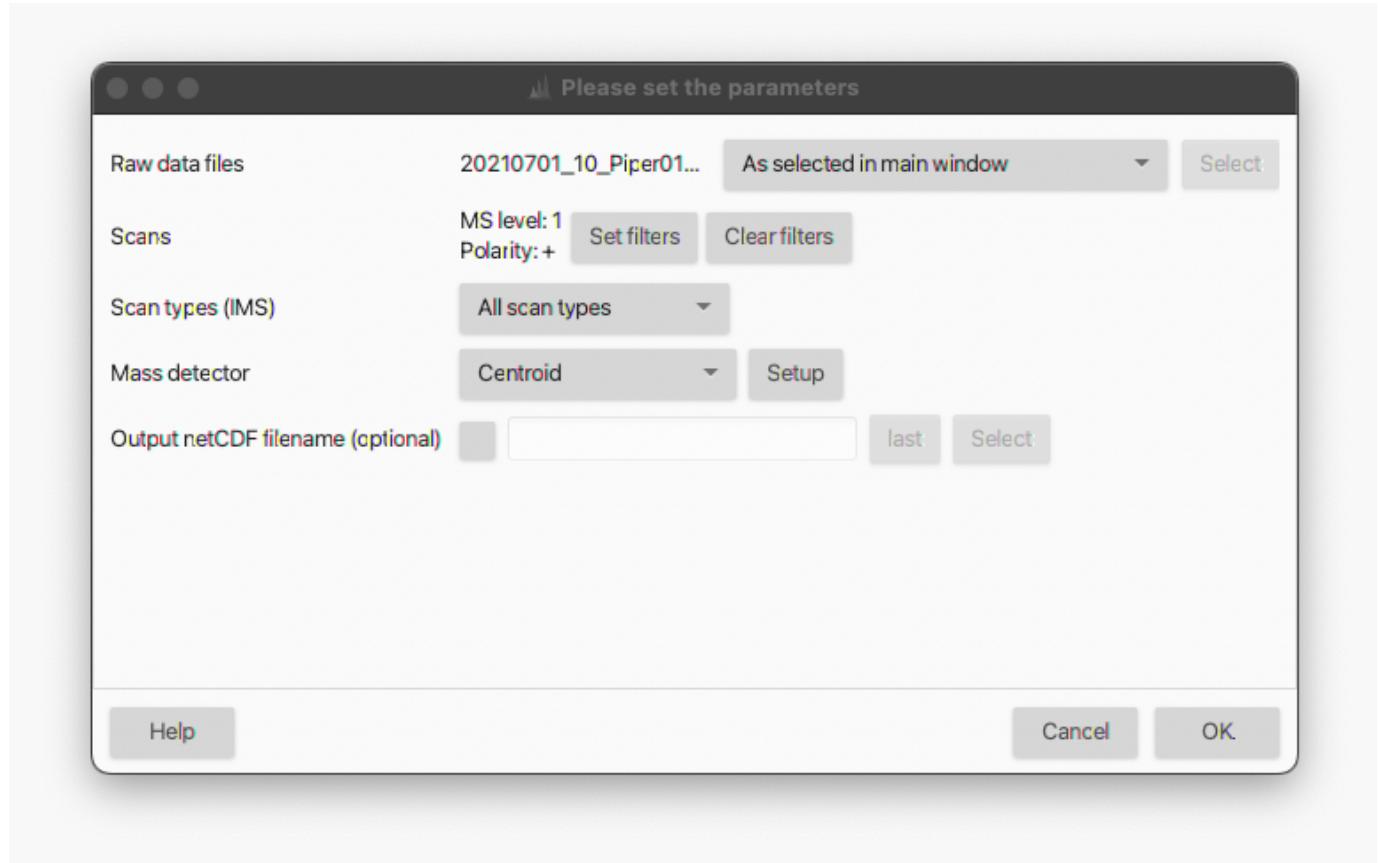
## 4.2 Mass detection

### 4.2.1 Mass detection

The mass detection module generates a [mass list](#) (*i.e.* list of m/z values and corresponding signal intensities) for each scan, in each raw data file. During the mass detection, profile raw data are centroided and a noise filtering is performed based on a user-defined threshold (see [Setting the noise level](#)).

#### Parameters settings

:material-menu-open: Raw data methods → Mass detection → Mass detection



#### Raw data files

Select the input raw data files for the mass detection. All the imported data files can be processed in bulk (*i.e.* *All raw data files*), or a subset can be selected directly from the *MS data files* panel (*i.e.* *As selected in the main window*) or based on the filename (*i.e.* *File name pattern*). As an alternative, the files' directory can be also specified (*i.e.* *Specific raw data files*). Finally, if the *Those created by previous batch step* option is selected, MZmine takes the output of the last processing step as input. This option is only available for [batch processing](#).

#### Scans

Select (or filter out) the MS scans to be processed. Several filters are available (*Select filters* button). A scan number, RT and mobility range can be set (*i.e.* *Scan number*, *Retention time* and *Mobility* options); only the scans belonging to the defined range(s) will be processed. The *Base Filtering Integer* option allows to process one every-N scans. The *Scan definition* field can be used to filter scans based on the scan's description normally included in the raw file's metadata (*e.g.* FTMS). Scans can also be filtered by *MS level* (*i.e.* 1, 2, ..., N), polarity and spectrum type (*i.e.* Centroided, Profile and Thresholded).

### Scan types (IMS)

This parameter applies only to IM data and determines if *mobility scans*, *frame scans* or both (*i.e.* All scan types) are processed. For more details about *mobility* and *frame scans*, see [here](#).

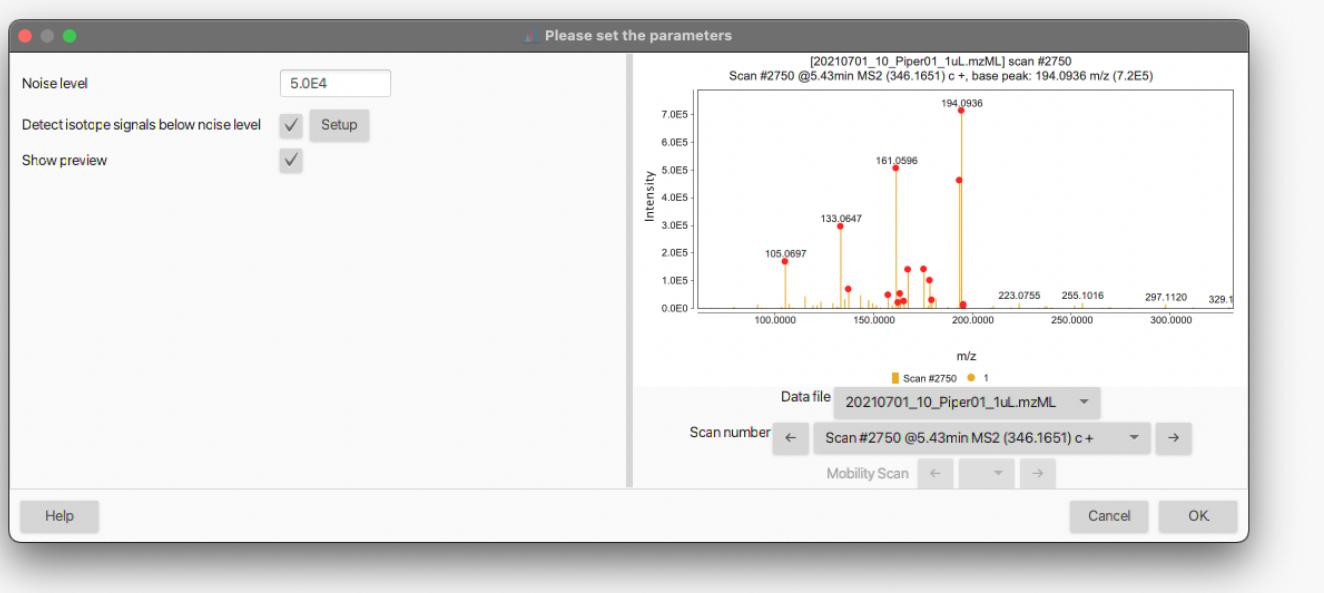
:octicons-light-bulb-16: **Tip.** Since *frame scans* are obtained by merging multiple *mobility scans*, the noise thresholds will likely be different. However, only one noise level can be set per processing step. Therefore, if one wants to run the mass detection for *mobility* and *frame scans* using different noise cutoffs, two module calls are required. As an alternative, mass detection can be performed only on the *mobility scans* by selecting the appropriate noise level. *Mobility scans* can then be merged into *frame scans* with a [dedicated module](#).

### Mass detector

Select the algorithm to be used for the mass detection. Several mass detection algorithms are available and can be selected in the drop-down menu. The choice depends on the raw data characteristics (profile/centroded, mass resolution, etc.). The *Centroid* algorithm must be used for already-centroded data. A step-by-step guide to convert profile into centroded data is provided in the [GNPS documentation](#). Other algorithms are available for profile raw data and are described in more details [here](#). The *Exact mass* algorithm is highly recommended for profile HRMS data. When *Auto* is selected, the *Centroid* and *Exact mass* algorithms are used by default for centroded and profile data, respectively.

### SETTING THE NOISE LEVEL

All the mass detection algorithms allow to set a threshold for the noise filtering (*i.e.* *Noise level*) by hitting the *Setup* button next to the *Mass detector* field. A dialog box like the following will open up:



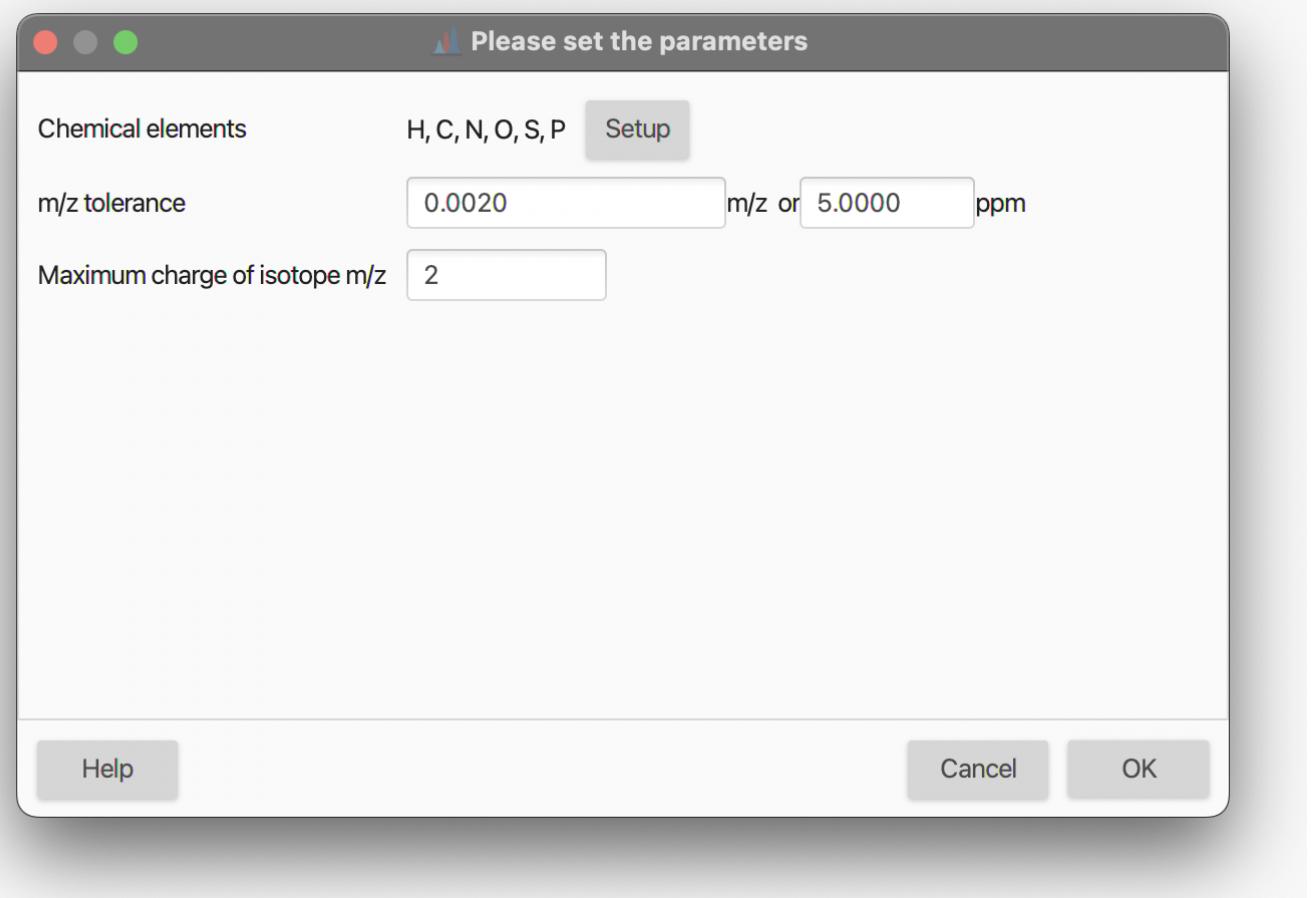
The noise threshold can be entered either in standard or scientific notation. By checking the *Show preview* box, an interactive visualization panel will open to help the user to adjust the noise level (see also [How do I determine the noise level in my data?](#)). The red dots denote the mass signals retained in the mass list according to the set threshold. Different data files and scan numbers can be visualized using the corresponding drop-down menus.

### DETECT ISOTOPE SIGNALS BELOW NOISE LEVEL

The *Centroid* and *Exact mass* algorithms provide the option to retain signals that are below the noise level (and would be otherwise discarded), but correspond to isotopes of the detected masses. Theoretical isotopic distributions are calculated for each mass detected in the *mass list* based on the specified chemical elements. If a signal below the noise threshold that matches a theoretical isotopic mass is found in the raw data, it will be included in the final mass list.

:octicons-light-bulb-16: **Tip.** In the case of LC-MS data processing, the low-intensity isotope signals included in the final mass list will undergo the whole feature detection workflow (see, for example, [LC-MS data processing workflow](#)). Due to the low intensity, these masses often produce LC peaks with poor peak shape during the chromatogram building step and might be discarded if they do not meet the user-defined parameters (*e.g.* minimum number of data points and intensity, see [ADAP chromatogram builder](#) for more details). Therefore, it might be advisable not to use this option during the mass detection, but rather use the Isotope finder module (CREATE DOC).

By ticking the corresponding checkbox and hitting the *Setup* button, the following dialog box opens up:



#### Chemical elements

Elements considered when generating the isotopic distributions. Select the elements from the periodic table by hitting the *Setup* button.

#### m/z tolerance

Maximum allowed difference between measured and theoretical isotope  $m/z$ . It is an [intra-scan m/z tolerance](#). The tolerance can be set in  $m/z$ , ppm or both. Since mass deviations expressed in ppm are dependent on the  $m/z$  (*e.g.* higher at low  $m/z$  and lower at high  $m/z$ ), MZmine automatically uses the largest tolerance.

#### Maximum charge of isotope m/z

Maximum allowed charge state of the isotope to be retained in the mass list. Default value is 1.

#### HOW DO I DETERMINE THE INSTRUMENTAL NOISE LEVEL IN MY DATA?

To-do. The detector noise is usually determined by a lot of signals of the same intensity.

Last update: April 5, 2022 19:53:57

## 4.2.2 Mass detection algorithms

---

To-do

---

Last update: March 26, 2022 22:05:35

## 4.3 Mobility scan merging

---

If a .mzML file is imported, the merged frame spectrum must be created via the **File merging → Mobility scan merging** module. This is required to gain access to MZmine's regular LC-MS functionality. This step uses the centroided and thresholded data produced by the [mass detection](#) step.

**This step is not required when importing native Bruker .tdf or .tsf data from .d folders.** When importing native Bruker data, a merged spectrum for the frame is created automatically by the vendor library.

---

Last update: April 5, 2022 11:25:19

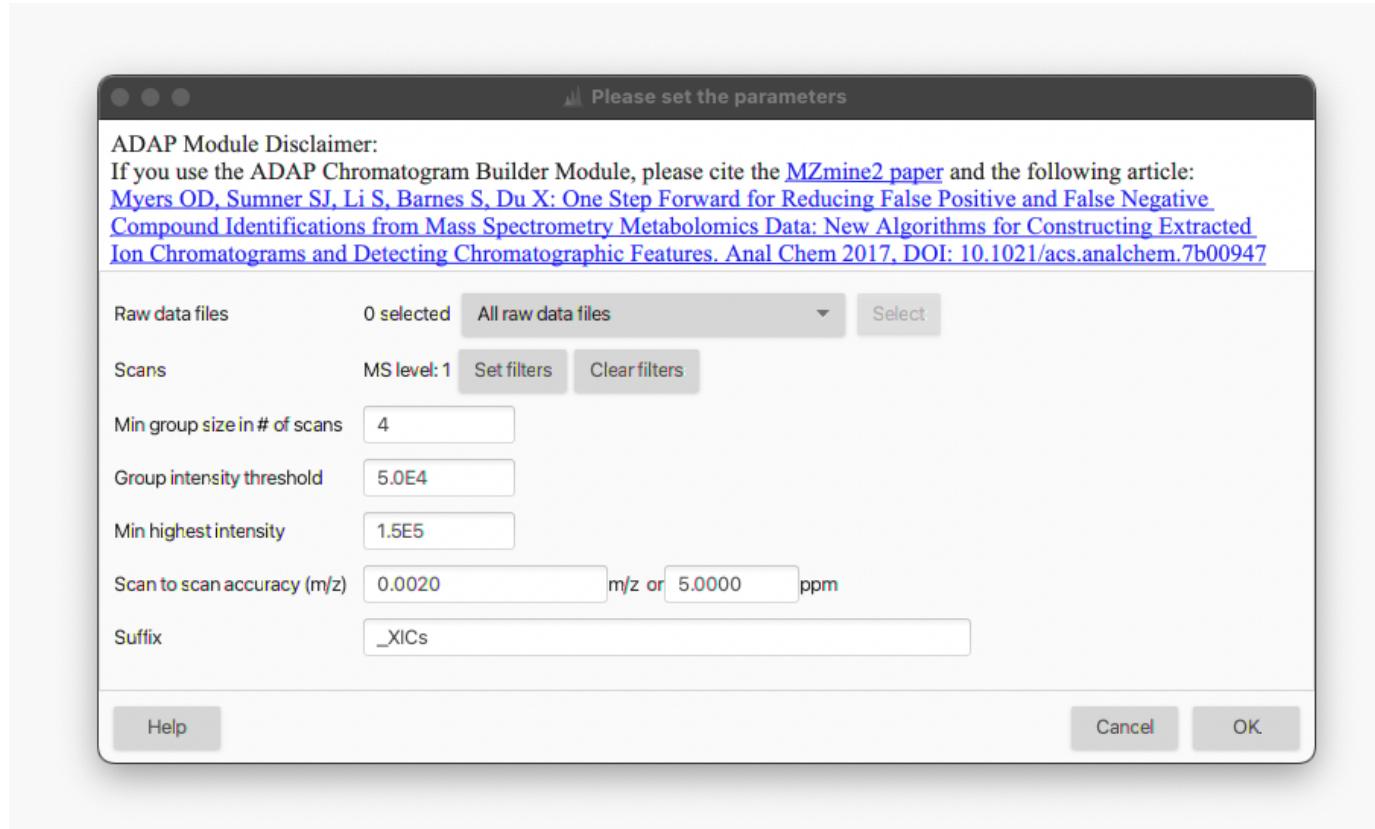
## 4.4 ADAP chromatogram builder

The *ADAP chromatogram builder* module is one of the LC-MS feature detection algorithms provided by MZmine 3. The module essentially builds an EIC for each  $m/z$  value that was detected over a minimum number of consecutive scans in the LC-MS run. Each data file is processed individually. The [mass list](#) associated to each MS1 scan in a data file (see [Mass detection](#) module) are taken as input and a [feature list](#) is returned as output. Since a mass list must be available, the *Mass detection* module must be run first.

The *ADAP chromatogram builder* algorithm operates as follows. Only MS1 scans are processed. All the data points are extracted from all the MS1 scans in a data file and sorted in order of decreasing intensity. The processing starts from the most intense data point and, since no EICs have yet been created, a new EIC is initialized and associated to the corresponding  $m/z$  value. The processing proceeds with the second-highest data point and the corresponding  $m/z$  is checked to determine if it "belongs" to the existing EIC based on the user-defined tolerance (*i.e.* "Scan to scan accuracy ( $m/z$ )" parameter). If so, the data point is added to the EIC and the EIC-associated  $m/z$  is updated. Otherwise, a new EIC is initialized. The process is iterated until all the data points have been processed and a set of EICs has been created. Finally, the EICs are checked according to the user-defined parameters (*i.e.* minimum number of data points and intensity). The EICs matching the requirements are retained in the [feature list](#), whereas the rest are discarded. The so-built EICs can then be resolved into individual features by one of the deconvolution algorithms provided by MZmine 3 (*e.g.* [Local minimum resolver](#) module).

### 4.4.1 Parameters settings

:material-menu-open: Feature detection → LC-MS → ADAP chromatogram builder



#### RAW DATA FILES

Select the input raw data files for chromatogram building. Mass lists associated with the data files will be automatically selected. See option descriptions in [Mass detection](#) module.

**SCANS**

Select (or filter out) the MS scans to be processed. Although setting the *MS level = 1* is usually sufficient for this module, several filters are available (see option descriptions in [Mass detection](#) module). For example, specific RT ranges (*e.g.* dead volume, equilibration time, calibration segments, *etc.*) can be excluded from the processing by setting the corresponding filter.

**MIN GROUP SIZE IN # OF SCANS**

Minimum number of consecutive MS1 scans where a *m/z* must be detected with a non-zero intensity in order for the corresponding EICs to be considered valid and retained in the feature list. :octicons-light-bulb-16: **Tip**. This parameter largely depends on the chromatographic system setup (*e.g.* HPLC vs UHPLC) and the acquisition rate (*a.k.a.* MS scan speed) of the mass spectrometer. The best way to optimize this setting is by manually inspecting the raw data and determining the typical minimum number of data points of the LC peaks. Usually, no less than 4-5 should be used.

**GROUP INTENSITY THRESHOLD**

Minimum signal intensity that the group scans (see previous parameter) must exceed in order for the corresponding EICs to be considered valid and retained in the feature list. :octicons-light-bulb-16: **Tip**. A good starting point for this parameter is 3 times the noise level used in the [Mass detection](#), if the instrumental noise is used as cutoff (see [here](#) for more details).

**MIN HIGHEST INTENSITY**

Minimum intensity that the highest point in the EIC must exceed in order for the corresponding trace to be considered valid and retained in the feature list. :octicons-light-bulb-16: **Tip**. A good starting point for this parameter is 7-10 times the noise level used in the [Mass detection](#), if the instrumental noise is used as cutoff.

**SCAN TO SCAN ACCURACY (M/Z)**

Maximum allowed difference between an EIC-associated *m/z* and a new data point to be added to the existing EIC trace. It is essentially the maximum allowed mass accuracy deviation between consecutive data points in the EICs. The tolerance can be set in *m/z*, ppm or both. It is an [inter-scan m/z tolerance](#) and it depends on the mass accuracy, resolution and stability of the instrument. :octicons-light-bulb-16: **Tip**. The best way to optimize this parameter is by manually inspecting the raw data and determining the typical fluctuation of the accurate mass measurement over consecutive scans. A good starting point is 0.002-0.005 *m/z* and 5-10 ppm for Orbitrap instruments, while 0.005 *m/z* and 10-15 ppm can be used for TOF devices.

**SUFFIX**

String added to the filename as suffix when creating the corresponding feature list.

---

Last update: April 6, 2022 09:25:41

## 4.5 IMS Expander

The IMS expander will search for data points in mobility scans for existing features. This requires prior chromatogram building (see [ADAP Chromatogram builder](#) and resolving in retention time dimension (see [Resolving](#)).

### Parameter settings

#### M/Z TOLERANCE

If selected, a tolerance will be applied to the feature's detected m/z while searching for data points in mobility dimension. Otherwise, the accepted m/z range is determined by the feature's m/z distribution in accumulated frame spectra.

**Recommended setting: selected, 0.003 m/z and 15 ppm**

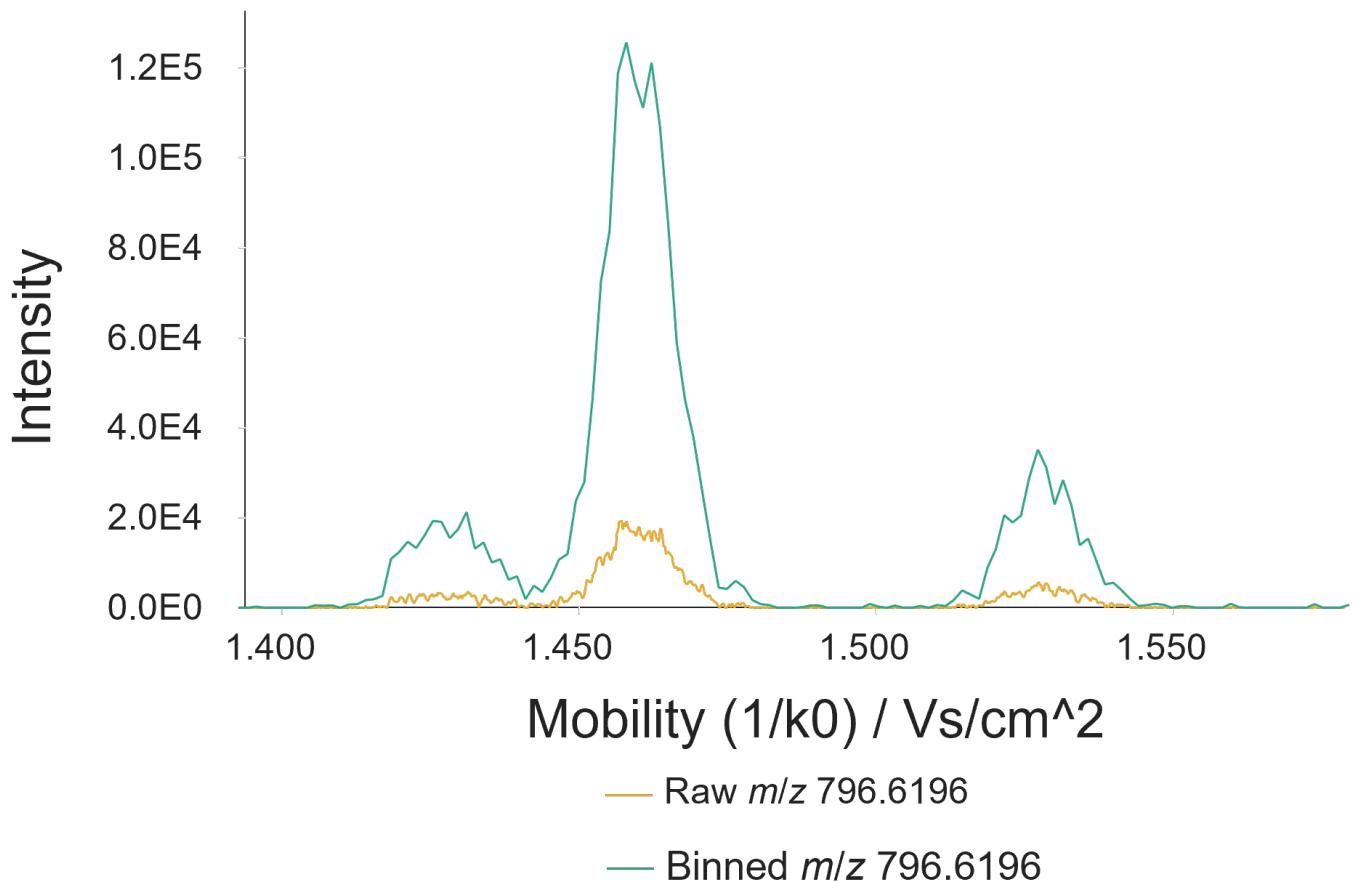
#### RAW DATA INSTEAD OF THRESHOLDED

Enables searching in mobility scan raw data instead of the thresholded (=mass detected) data. Only possible for centroid raw data files.

#### OVERRIDE DEFAULT MOBILITY BIN WIDTH (SCANS)

If selected, the default number of binned mobility scans can be overridden. Useful for data with high mobility resolution.

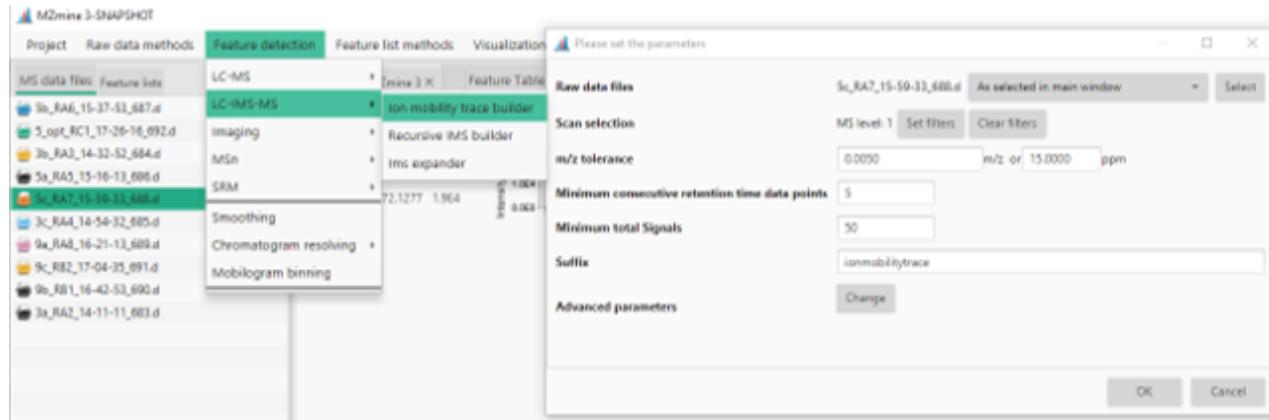
### Binned mobilogram example



## 4.6 Ion mobility trace builder

The **Ion mobility trace builder** will build ion mobility traces from the raw data. Alternatively, the **Recursive IMS builder** can be used, which requires less ram but takes longer.

The **Ion mobility trace builder** is found under **Feature detection → LC-IMS-MS → Ion mobility trace builder**.



### Scan selection

The scan selection parameter specifies the scans that shall be processed for feature detection. Usually, setting the ms level to 1 is sufficient. If a calibration segment is present, it can be cut out via the retention time filter in the scan selection.

### m/z tolerance

The **m/z tolerance** specifies the scan-to-scan tolerance for ion mobility traces. This tolerance window may need to be set higher than for classic LC-MS feature detection (e.g. to 0.005 m/z and 15-20 ppm instead of 10 ppm) due to lower intensities therefore less accuracy in individual mobility scans compared to LC-MS scans. Note that the overall accuracy is achieved via LC-IMS-MS is the same due to the higher number of scans.

### Minimum consecutive retention time data points

This parameter specifies the number of consecutive detections of the same m/z value in a chromatographic peak (rt dimension only). This means that a single m/z has to be detected in, e.g. 5 frames with an intensity higher than zero. This parameter helps to filter noise. Consecutive detections in the mobility dimension do not affect this parameter. Usually no less than 5 should be set here if the MS1 acquisition rate is sufficient.

### Minimum total signals

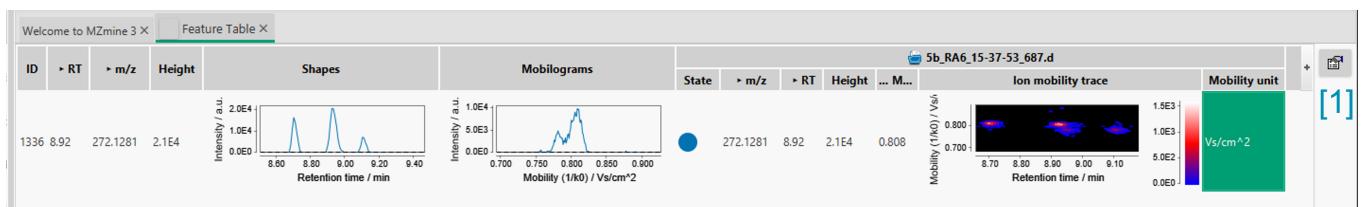
Specifies the total number of peaks in the mobility dimension in all mobility scans. Every "dot" in an ion mobility trace represents a single datapoint. ( see [ion mobility traces](#) )

### Advanced parameters

For most applications, these parameters do not need to be set/changed. For high mobility resolved data the mobilograms might become noisy due to less ions reaching the detector at the same time. By default, the number of binned scans is set to cover about 0.0008 Vs/cm<sup>2</sup> per bin. The effect of binning can be seen [here](#). If you are unsure about the nature of your data, you can perform trace building with the standard parameters and apply/preview the binning afterwards via the **Feature detection → Mobilogram binning** module.

### Processing result

After performing ion mobility trace detection, a feature list is created in the feature list tab (see [feature lists tab](#)). In the feature table, multiple columns are created. The displayed columns can be set via the button on the right of the feature table ([1]).



The **shapes** displays an EIC of the ion mobility trace (intensities summed in rt dimension). The **mobilograms** column shows a mobilogram for the ion mobility trace (intensities summed in mobility dimension). The shapes and projections can be smoothed and resolved. However, the ion mobility trace is always represented by the raw data and remains unaltered. After resolving, the shapes and mobilograms have to be recalculated from the raw data, which is why the smoothing is lost after resolving.

Last update: April 5, 2022 13:31:49

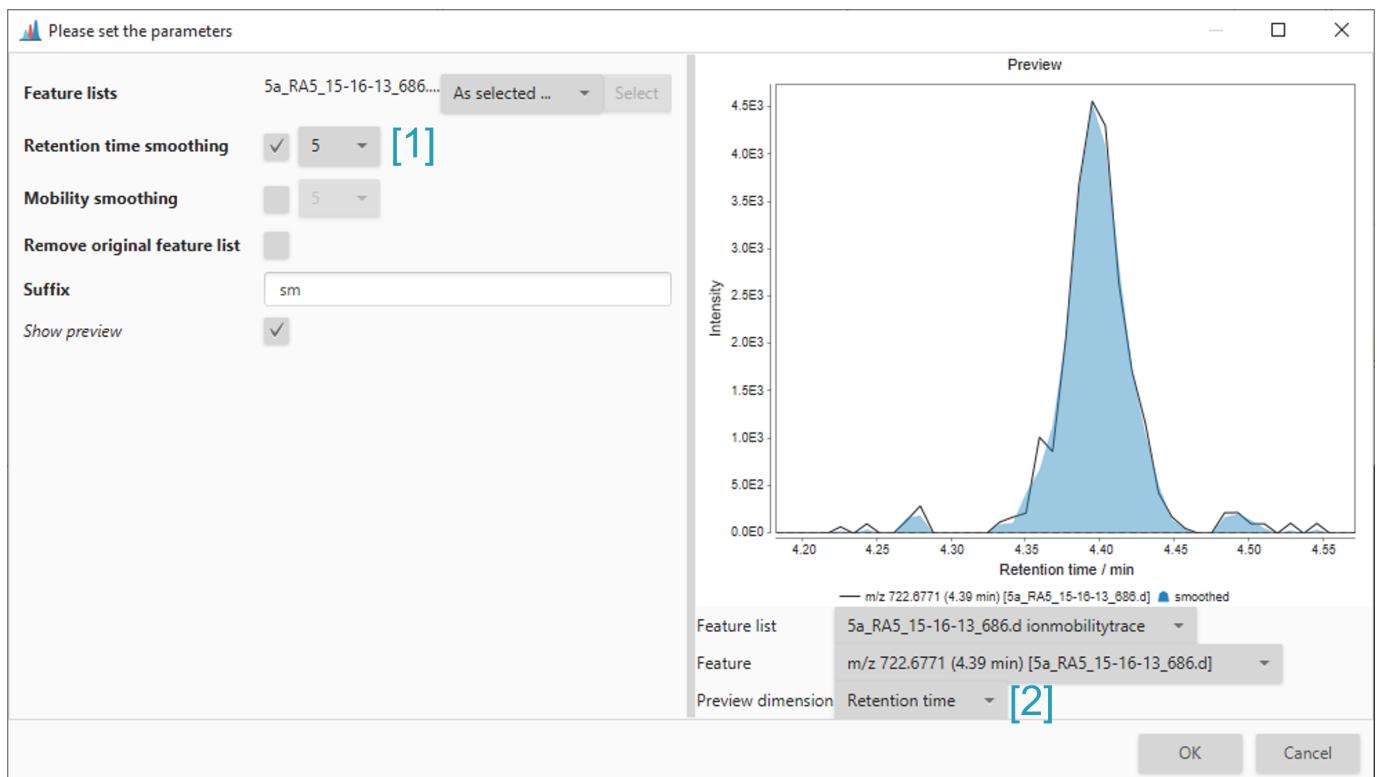
## 4.7 Smoothing

### 4.7.1 Retention time dimension

Smoothing chromatograms is optional. The necessity of smoothing in RT dimension is determined by the noisiness of chromatographic peaks. These can be influenced by the overall spray stability, instrument accumulation times, transfer efficiency and many more.

The number of data points to be smoothed in rt dimension can be set at [1]. Note that the correct preview dimension is selected at [2].

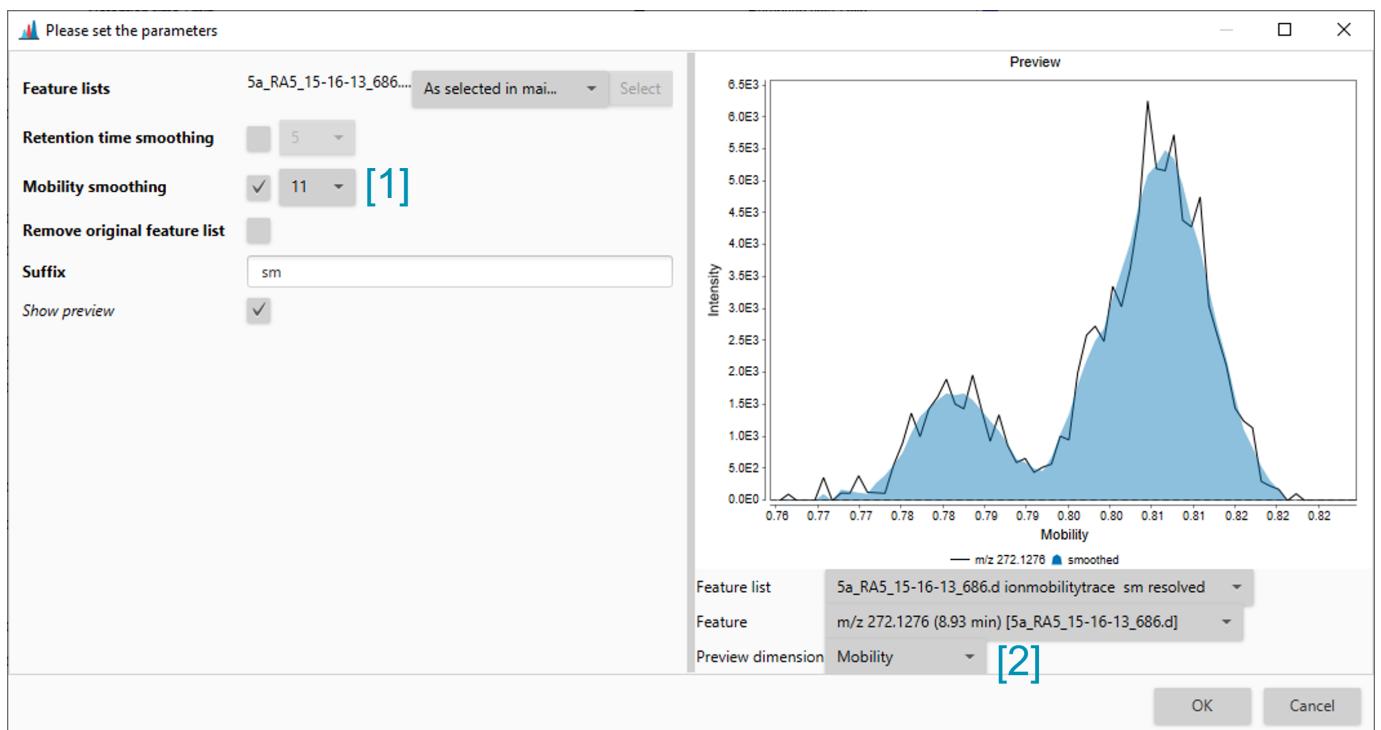
For large batch modes, the **Remove original feature list** parameter should be selected. While parameters are being optimised, this is not recommended, because removing a feature list cannot be undone.



### 4.7.2 Mobility dimension

After resolving a feature in RT dimension, the mobilograms will be recalculated from the raw data (the resolved ion mobility trace). Therefore, a smoothing step is necessary if the data requires it. The smoothing dialog is opened via **Feature detection → Smoothing**

Select to smooth the mobility dimension [1] and select it as preview dimension [2]. The filter width depends on the number of spectra acquired in the observed mobility range. Usually, a value between 5 and 15 should be appropriate.



Last update: March 10, 2022 15:42:38

## 4.8 Local Minimum Resolver

---

During the EICs building, overlapping and partially co-eluting peaks are retained as single features in the feature list (see, for example, [ADAP chromatogram builder](#)). The *local minimum resolver* module aims at splitting such "shoulder" LC peaks into individual features (*i.e.* [chromatographic deconvolution](#)) based on local minima. In fact, a local minimum in the EIC trace might correspond to the valley between two adjacent, partially-resolved peaks.

The algorithm examines all the data points in the EIC trace starting from the earliest RT. A scan window is set (see *Minimum search range RT/Mobility* parameter) and centred around the examined data point. A data point is considered a local minimum if it is the lowest intense point within the scan window. Moreover, it has to be X times lower than the highest point in the window (see *Min ratio of peak top/edge* parameter). When a local minimum is found, a set of user-defined intensity and peak duration requirements is checked. If they are fulfilled, the original overlapping peaks are split into new, distinct features.

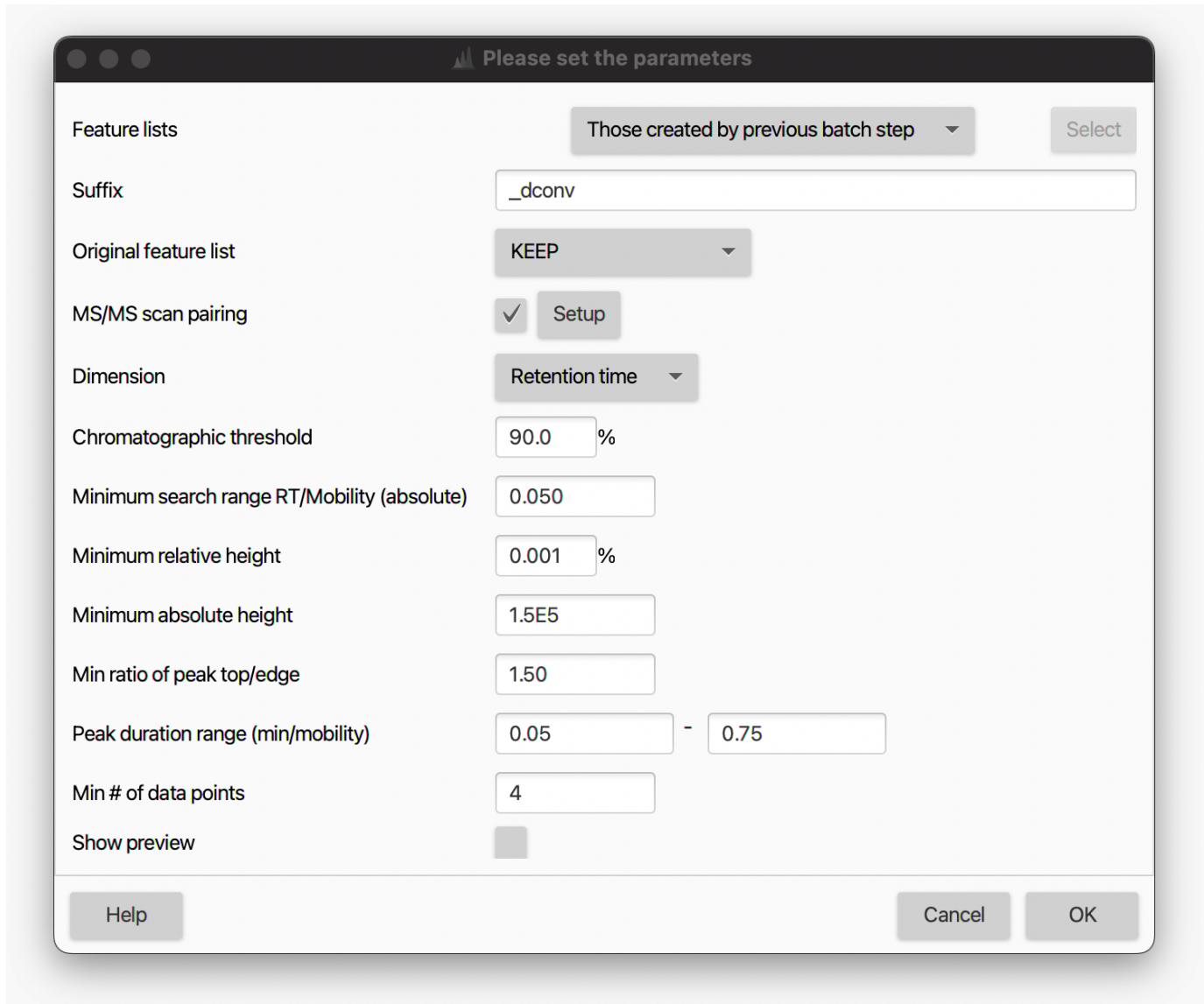
With the implementation of IM support in MZmine 3, this module was expanded and can now be applied to both the RT and IM dimension (see [Resolving the ion mobility dimension](#)).

:octicons-light-bulb-16: **Tip.** The LMR is particularly suitable for LC-MS data with little noise and nice peak shapes.

---

### 4.8.1 Parameters settings

:material-menu-open: Feature detection → Chromatogram resolving → Local minimum resolver

**SUFFIX**

String added as suffix when creating the corresponding feature list.

**ORIGINAL FEATURE LIST**

Keep or remove the processed feature list(s).

**MS/MS SCAN PAIRING**

Pair MS/MS fragmentation spectra collected in [DDA](#) mode to the resolved features. This is optional at this stage as the same can be done with a separate [module](#). See [MS2 scan pairing](#) documentation for more details.

**DIMENSION**

Dimension to be resolved. Select *Retention time* or *Mobility* to run the module over the RT or ion mobility dimension, respectively.

**CHROMATOGRAPHIC THRESHOLD**

Percentage of data points in the EIC removed before local minima search. The algorithm finds the intensity value (threshold) such that the specified percentage of EIC's data points are below the threshold. All such data points are removed. For example, a *Chromatographic threshold* = 50% will discard the lowest-intense 50% data points. This represents an important filter for noisy chromatogram and significantly reduces the precessing time.

:octicons-alert-16: Since the algorithms examines the EICs throughout the entire RT range (*i.e.* also the zero data points are considered), we recommend to set this value rather high (*e.g.* 95%) and lower it only if necessary. :octicons-alert-16: ION MOBILITY 80%

#### MINIMUM SEARCH RANGE RT/MOBILITY (ABSOLUTE)

Size of the moving window examined for local minima search. Overly low values can cause peak edges to be cut off. On the other hand, too high might lead to incomplete separation of narrowly eluting compounds.

#### MINIMUM RELATIVE HEIGHT

Minimum relative intensity (respect to the highest data point in the EIC) a peak need to reach to be retained as a feature. Overly high thresholds may lead to low-intensity features being discarded. :octicons-light-bulb-16: **Tip.** Modern HRMS instruments can show linear dynamic ranges up to five orders of magnitude. If we take an Orbitrap device with a detector saturation level around 1.0E10 intensity, a *Minimum relative height* = 0.001 would correspond to 1.0E5.

#### MINIMUM ABSOLUTE HEIGHT

Minimum absolute intensity a peak needs to reach to be retained as a feature. This parameter depends on what you want to detect, the instrument and detector type. Usually, Orbitrap instruments report higher intensities than TOF instruments. However, the noise level is also higher for Orbitrap than for TOF instruments. For TOF instruments 1E3 or even 5E2 can be appropriate, whilst Orbitraps can require 1E5 or 5E4. :octicons-light-bulb-16: **Tip.** The same value used as *Min highest intensity* in the EICs building step (*e.g.* [ADAP chromatogram builder](#)) can normally be used here.

**Min ratio of peak top/edge:** Describes the minimum ratio between the highest and lowest point of a peak. This mostly affects detection of low intensity and not-baseline-resolved signals and should be optimised using such a signal as an example.

:octicons-light-bulb-16: **Tip.** We recommend values between 1.7 (not baseline separated) and 2 to start the optimisation.

**Peak duration range (min/mobility):** Describes valid peak lengths. Can be used to filter out very short or long noise signals.

**Min # of data points:** Can be used in addition to **Peak duration range** to filter out noise. Should be set no lower than 5 in most cases.

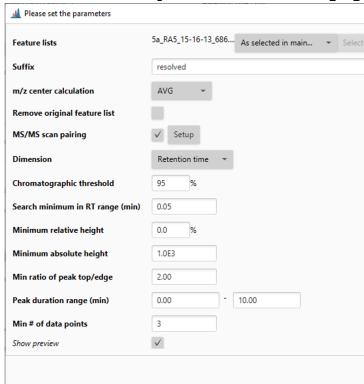
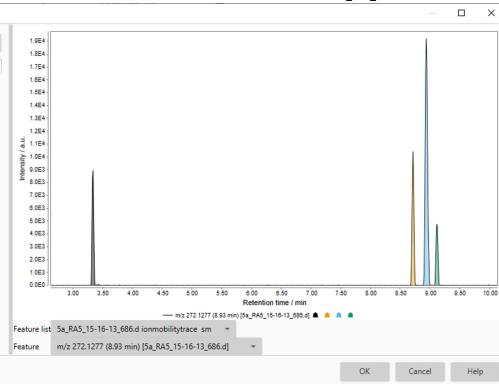
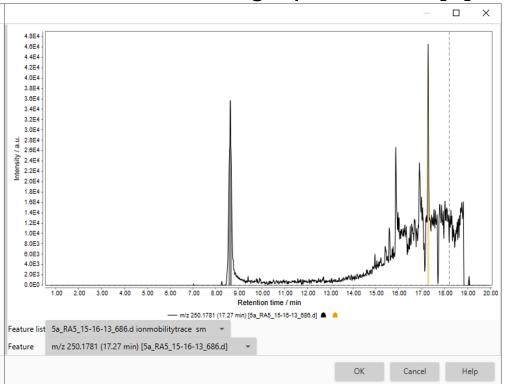
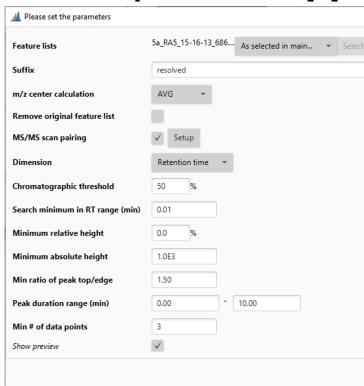
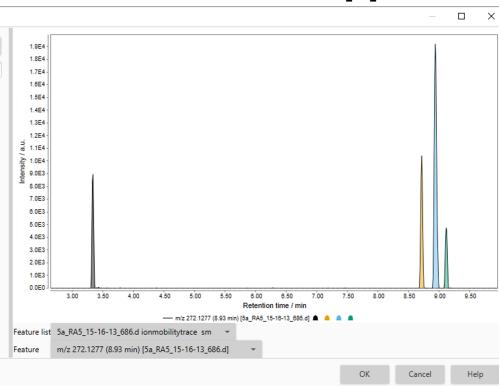
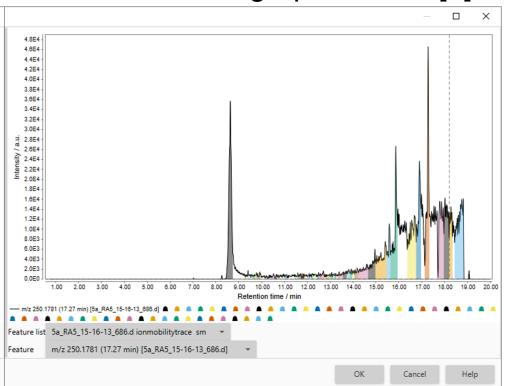
**Show preview:** To-do

## 4.8.2 Example

---

Optimal algorithm settings [1] can lead to recognition of all good features in a "good" EICs [2] and to few noise recognised as feature in EIC that also contains chromatographic noise [3]. On the other hand, poorly set parameters [4] will likely recognize all features in a good EIC [5], but also retain a lot of noise as feature in a noisy EIC [6].

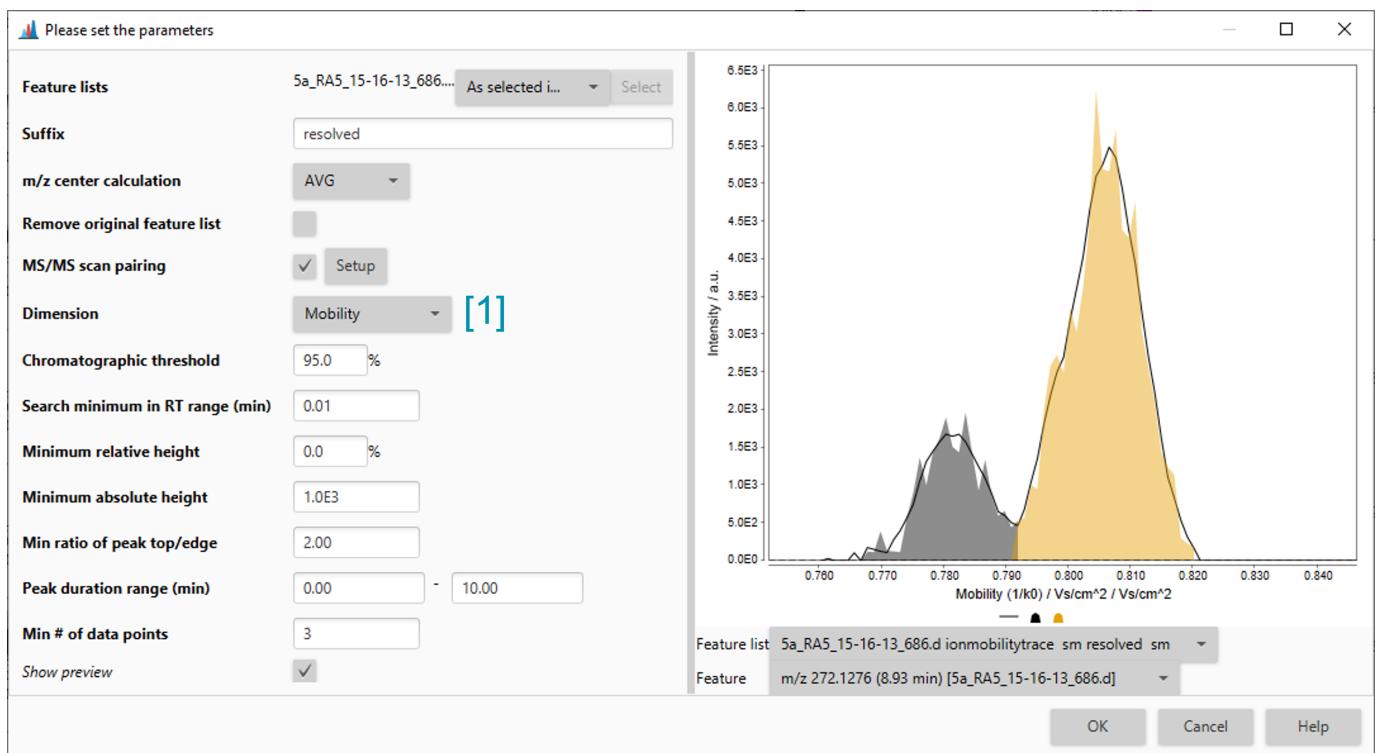
Therefore, we recommend optimising the parameters on good EICs and checking the results of these parameters with a noisy EIC. Most of the time, a noisy EIC can be found by sorting the feature table by decreasing area.

**Good parameters [1]****Good EIC [2]****Chromatographic noise [3]****Bad parameters [4]****Good EIC [5]****Chromatographic noise [6]**

#### 4.8.3 Resolving the ion mobility dimension

To-finish

In general, the same principles apply as in the retention time resolving step. However, a few differences shall be noted. In the screenshot you can also see, that the resolved mobilograms are recalculated from the raw data and previously applied smoothing steps are therefore lost and must be reapplied if necessary.



#### Dimension

Mobility has to be selected as a dimension to resolve mobilograms [1].

#### Chromatographic threshold

Since there are less scans in mobility dimension (e.g., 400 - >1000 per frame, depending on the instrument type and setting) than in rt dimension (e.g. 5000 for LC-MS depending on acquisition rate), the threshold should be lowered to 80 or less.

#### Search minimum range

This value determines the search range in mobility dimension. Therefore, this value has to be set lower when resolving a TIMS (Bruker data) mobilogram, because the absolute numerical values are smaller (e.g., 0.01). When resolving mobilograms from Waters or Agilent data (mobility as drift time in ms), the values are higher and must therefore be increased.

Other parameters such as minimum intensities or minimum number of data points should be adjusted depending on what you want to detect.

---

Last update: April 6, 2022 09:25:41

## 4.9 CCS Calibration and calculation

Accurate determination of CCS values requires a valid CCS calibration and molecule charge states to be detected.

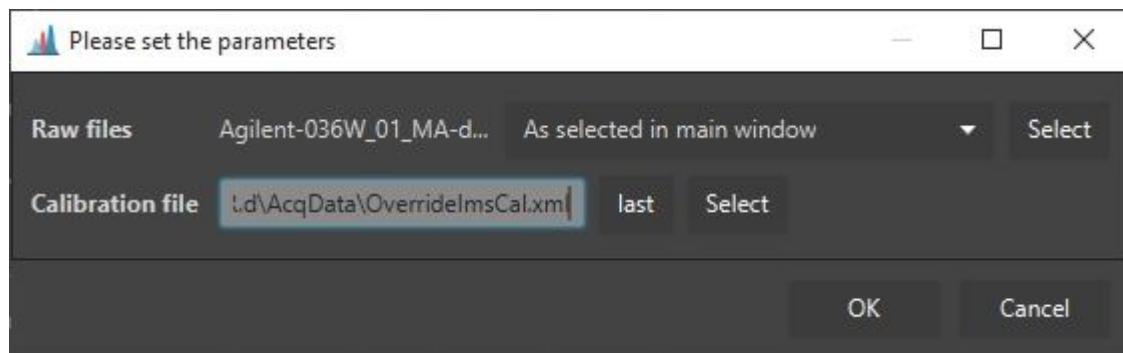
- **timsTOF** raw data can be recalibrated using data analysis and imported in MZmine. The recalibrated data will be used by default. (see [Calculating CCS values](#))
- **mzML** raw data requires the determination of a calibration function from the raw data (e.g. as detected features) or as import from an external file. (see [Creating or importing a CCS calibration](#))

### 4.9.1 Creating or importing a CCS calibration

#### Importing a CCS calibration

Agilent calibration data can be imported from the "OverrideImsCal.xml" file in the Agilent raw data folder. Waters calibration data can be imported from the "mob\_cal.csv" file in the Waters raw data folder. The "\_extern.inf" file is also required, but will be read automatically when the "mob\_cal.csv" is selected.

The calibration import is accessed via **Feature list methods -> Processing -> External CCS Calibration**. Then select the calibration "OverrideImsCal.xml"/"mob\_cal.csv" from the raw data folder, and select the raw data files the calibration should be applied to.

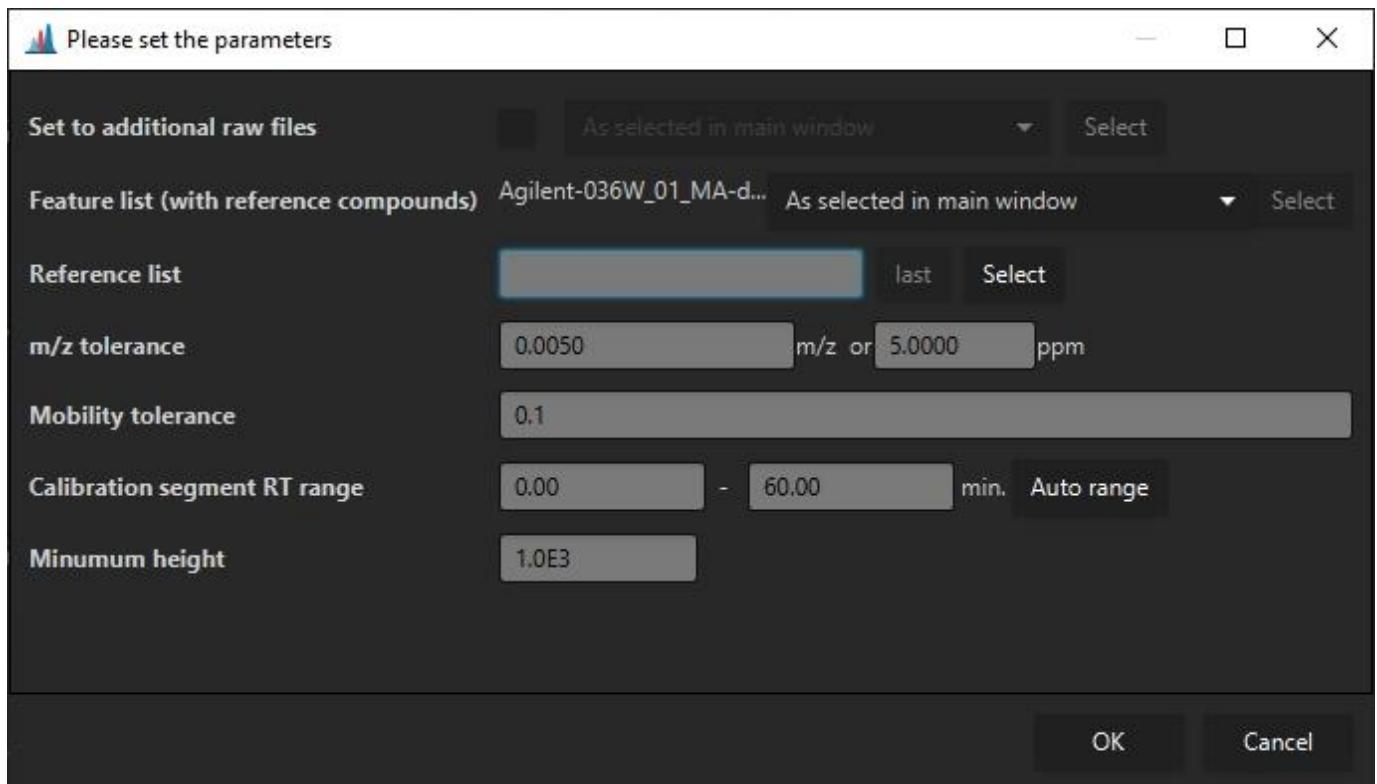


#### Reference CSS calibration

If a mobility calibrant is infused during an HPLC run of every sample, a CCS calibration can be calculated on a per-run basis. (Common procedure on Bruker devices) Otherwise, a single run can be used to calibrate multiple files.

Please note that this is currently only supported for TIMS and DTIMS data.

The calibration module can be accessed via **Feature list methods -> Processing -> Internal reference calibration**.



**Set to additional raw files** If a calibration calculated from a single feature list shall be applied to multiple other raw files, the raw files can be selected here. This requires only a single raw file to be selected.

**Feature list (with reference compounds)** Specifies (a) feature list(s) that contains the reference compounds. If multiple feature lists are selected, every feature list will be searched for reference compounds, and the calibration will be used for the raw data files in the particular feature list. This means that no raw data file may be selected. (Cannot set multiple calibrations to a single raw file.)

If a single feature list is selected, the calibration may be applied to additional raw data files via the **Set to additional raw files** parameter.

**Reference list** Specifies a ".csv" reference list of for CCS calibrant ions. Must contain the columns "mz", "mobility", "ccs", "charge". Columns must be separated by ";". The ion mode may be specified via the charge of the ion, e.g., as 1 or -1. Only the correct polarity will be used to calculate the calibration.

**m/z tolerance** The m/z tolerance for the reference compounds.

**Mobility tolerance** the mobility tolerance to detect the reference compounds.

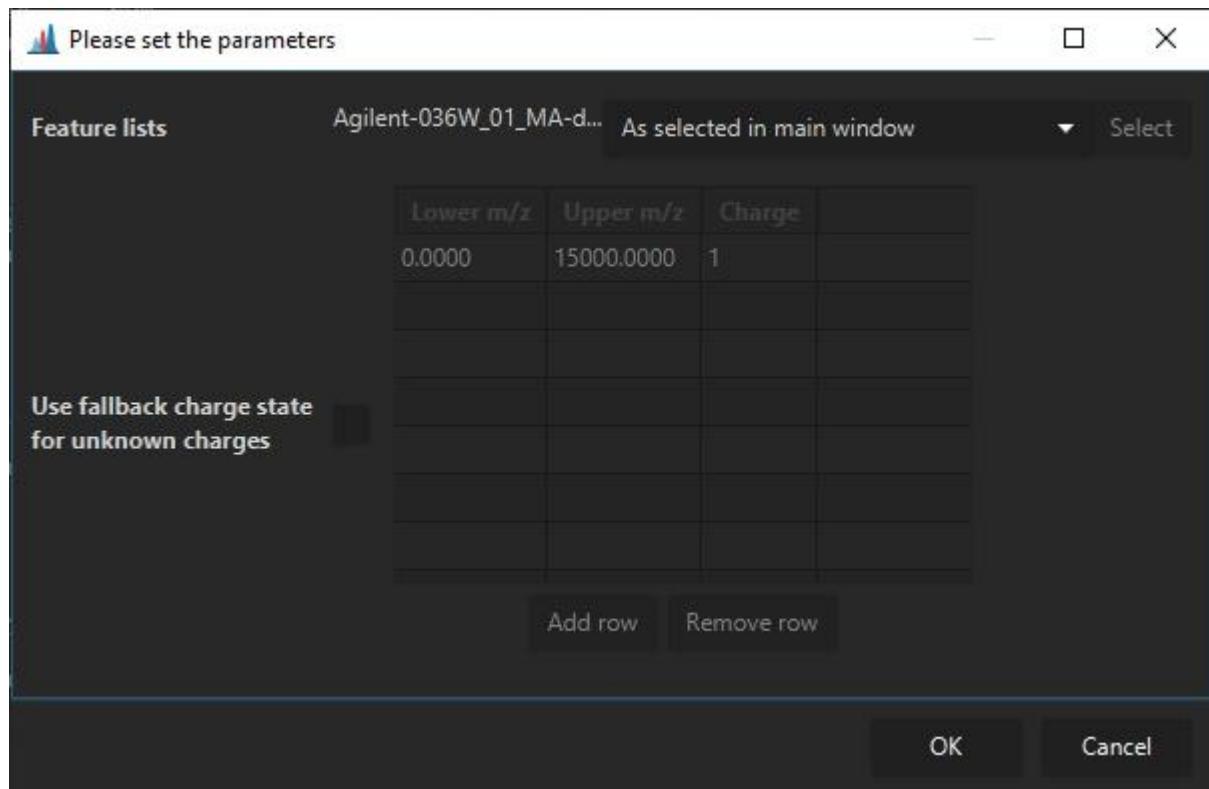
**Calibration segment RT Range** Specifies the rt range that shall be searched for calibrant ions. Usually either the beginning or end of a HPLC run.

**Minimum height** A minimum intensity for reference compounds to be used as calibrant signals for determination of the calibration.

#### 4.9.2 Calculating CCS values

After a calibration as been set (Agilent/Waters/Bruker mzML) (Bruker tdf works out-of-the-box) CCS values can be calculated via **Feature list methods -> Processing -> Calculate CCS values**.

Here, a default charge state may be set, in case it could not be determined. Otherwise, the charge state determined via the isotope pattern will be used.



Last update: March 10, 2022 15:42:38

## 4.10 Gap-filling

---

Some chromatographic peaks may not be detected in every sample for several reason: - Reason 1 - Reason 2 - Reason 3

- This causes undesirable gaps (missing values) in the aligned feature table.
  - To tackle this issue, a value for the peak needs to be imputed
  - A simple gap-filling approach is to integrate the area where the peak is expected but not detected
  - These areas usually correspond to spectral noise
  - By doing so, no bias is introduced
  - The gap-filled feature table can now be used in downstream data analysis
- 

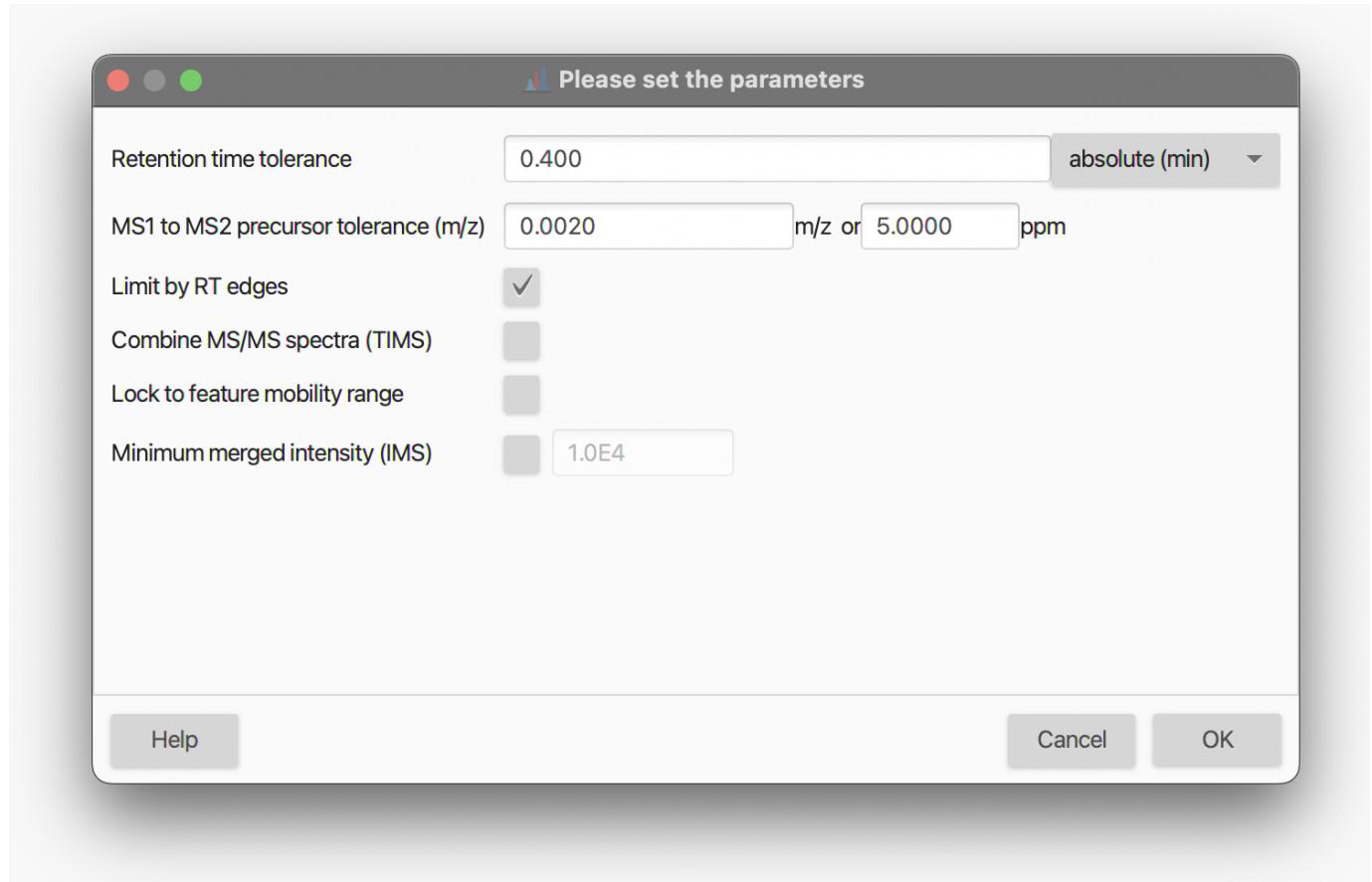
Last update: March 10, 2022 15:42:38

## 4.11 MS2 Scan Pairing

Description

### 4.11.1 Parameters settings

:material-menu-open: Feature list methods → Processing → Assign MS2 to feature



#### RETENTION TIME TOLERANCE

To-do

#### MS1 TO MS2 PRECURSOR TOLERANCE (M/Z)

To-do

#### LIMIT BY RT EDGES

To-do

#### COMBINE MS/MS SPECTRA (TIMS)

To-do

#### LOCK TO FEATURE MOBILITY RANGE

To-do

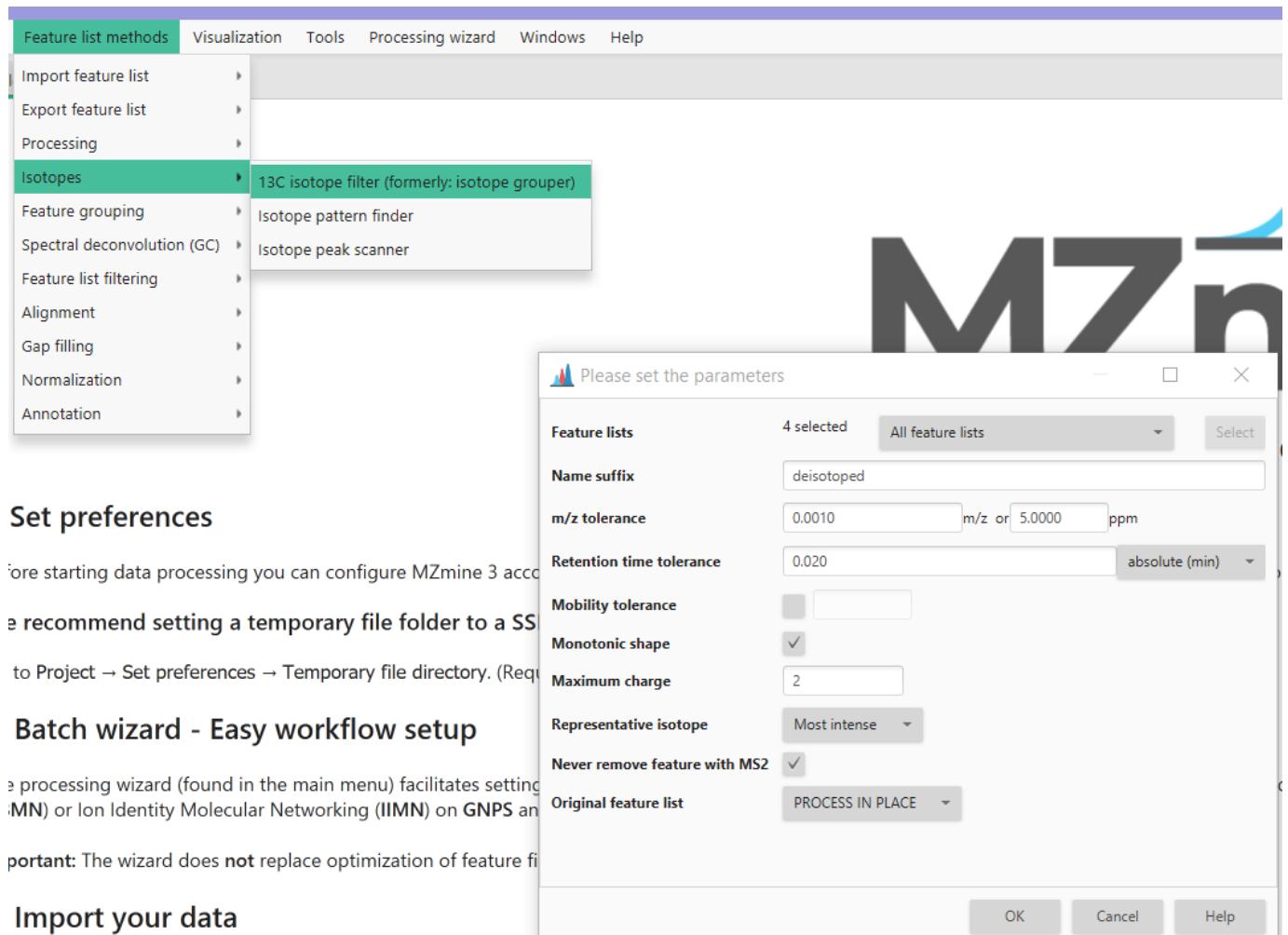
#### MINIMUM MERGED INTENSITY

To-do

Last update: April 5, 2022 13:22:51

## 4.12 $^{13}\text{C}$ isotope filter (Isotope grouper)

The  **$^{13}\text{C}$  isotope filter** is found under **Feature list methods → Isotopes →  $^{13}\text{C}$  isotope filter (formerly: isotope grouper)**.



### Set preferences

Before starting data processing you can configure MZmine 3 according to your needs.

We recommend setting a temporary file folder to a suitable location.

To do this, go to Project → Set preferences → Temporary file directory. (Requires administrator rights)

### Batch wizard - Easy workflow setup

The processing wizard (found in the main menu) facilitates setting up workflows for Ion Mobility Mass Spectrometry (IMMS) or Ion Identity Molecular Networking (IIMN) on GNPS and other platforms.

Important: The wizard does **not** replace optimization of feature lists.

### Import your data

#### Parameters

##### NAME SUFFIX

Suffix to be added to feature list name.

##### M/Z TOLERANCE

Maximum allowed difference between two features' m/z values in order for them to be considered the same. The value is specified both as absolute tolerance (in m/z) and relative tolerance (in ppm). The tolerance range is calculated using maximum of the absolute and relative tolerances.

##### RETENTION TIME TOLERANCE

Maximum allowed difference between the retention time values of two features that will be taken into account when performing the  $^{13}\text{C}$  isotope filtering.

##### MOBILITY TOLERANCE

If enabled (and mobility dimension was recorded), isotopic peaks will only be grouped if they fit the given tolerance.

**MONOTONIC SHAPE**

If true, then monotonically decreasing height of isotope pattern is required. This is usually the case for  $^{13}\text{C}$  isotope pattern of small molecules.

**MAXIMUM CHARGE**

Maximum charge to consider for detecting the isotope patterns. The charge state with the maximum number of detected isotope features will be used.

**NEVER REMOVE FEATURE WITH MS2**

If checked, all features with MS2 are retained without applying any further filters on them.

**ORIGINAL FEATURE LISTS**

The input feature list can either be kept, removed, or directly filtered (PROCESS IN PLACE). This allows for more control over memory usage. Where available, in place processing is the most performant.

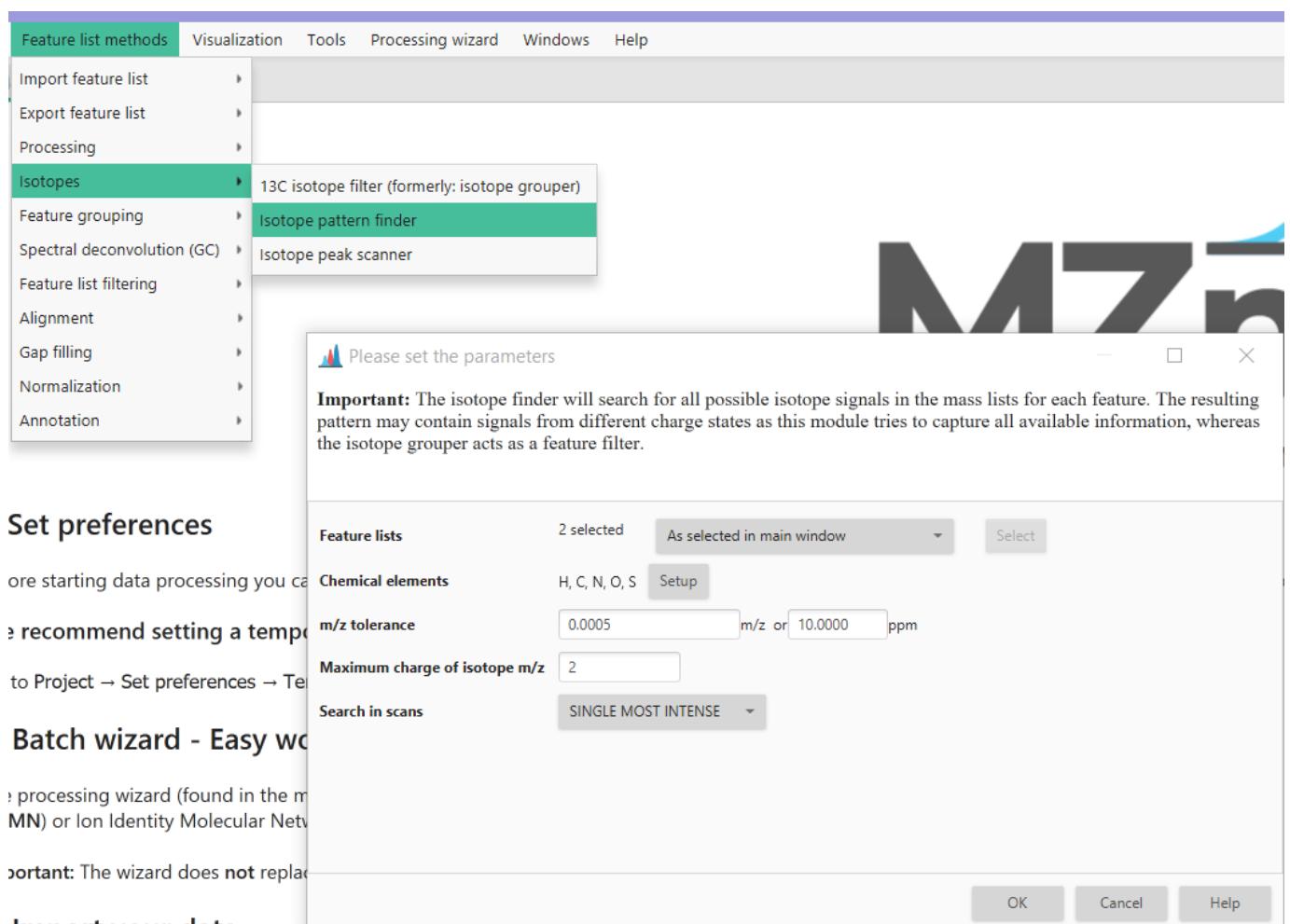
---

Last update: April 1, 2022 09:23:04

## 4.13 Isotope pattern finder

The **Isotope pattern finder** is found under **Feature list methods → Isotopes → Isotope pattern finder**.

The module searches isotope patterns for each feature in selected feature lists by going back to the mass spectra. Starting from the feature m/z the algorithm will first backtrack any possible preceding isotope signals using a list of delta masses created from elements, their stable isotopes, and an m/z tolerance. For example, a -2 signal might be detected when searching for Br isotopes. In a second step, all picked potential isotope m/z values are used to search next isotope (with higher m/z). This is done for each charge state. I



### Set preferences

ore starting data processing you can

re recommend setting a tempe

to Project → Set preferences → Te

**Batch wizard - Easy way**

processing wizard (found in the m

important: The wizard does not replace

### Import your data

#### Parameters

##### CHEMICAL ELEMENTS

All stable isotopes of the chosen elements are used to create a list of mass differences to search. Signals with this mass difference (m/z difference with different charge states) are then considered as potential isotope signals.

##### M/Z TOLERANCE

Maximum allowed difference between two features' m/z values in order for them to be considered the same. The value is specified both as absolute tolerance (in m/z) and relative tolerance (in ppm). The tolerance range is calculated using maximum of the absolute and relative tolerances.

**MAXIMUM CHARGE OF ISOTOPE M/Z**

Maximum possible charge of isotope m/z distributions. All present m/z values obtained by dividing isotope masses with 1,2 ...,maxCharge values will be considered. The default value is 1, but insert an integer greater than 1 if you want to consider ions of higher charge states.

**SEARCH IN SCANS**

Currently, the supported option is "Single most intense", which means the search for isotopes will happen in the single most intense MS scan of each feature.

---

Last update: April 11, 2022 20:20:26

## 5. Workflows

---

### 5.1 LC-MS Workflow

The workflow proposed herein is intended as a general pipeline for untargeted LC-MS (or LC-MS/MS) data preprocessing. The main goal is essentially to turn the highly-complex LC-MS raw data into a list of features, and corresponding signal intensity, detected across the analysed samples. Such feature lists can then be exported for further downstream analysis (e.g., identification, search against spectral libraries, statistical analysis, etc.). A schematic representation of the workflow is shown below:



References: - 10.1039/9781788019880-00232 - 10.1007/978-1-0716-0239-3\_3 - 10.1016/bs.coac.2018.08.003

#### 5.1.1 Raw data processing

The raw data processing consists of essentially two steps: [Data import](#) and [Mass detection](#)

##### **Raw data import**

Either open (e.g. mzML) and native vendor (e.g. Thermo, Bruker) data formats can be imported in MZmine 3. All the supported formats are listed here ([LINK to Doc](#)). For more details see the [Data import](#) module.

##### **Mass detection**

This step produces a list (referred to as "mass list") of the m/z values found in each MS scan across the LC run that exceed a user-defined threshold (i.e. noise level). For more details see the [Mass detection](#) module.

#### 5.1.2 Feature processing

The goal of the "Feature processing" is to obtain a list of all the detected features (characterized by a RT and m/z value) from the raw LC-MS data.

##### **Chromatogram building**

The first step in the "Feature processing" is to build the so-called extracted ion chromatograms (EICs) for each detected mass (see "Mass detection"). There are two modules in MZmine 3 that can fulfil this task: [ADAP chromatogram builder](#) (widely used) and [Grid mass](#) (create docs).

The "detected" features in each file are listed in the so-called "feature lists", which are then further processed.

(e.g. to ) and aligned to connect corresponding features across all samples.

##### **Smoothing in retention time dimension (optional)**

- Optional, depends on the LC peak shape
- For more details see the [Mass detection](#) module.

##### [Smoothing](#)

## Feature resolving

### Local minimum resolver

#### [13C isotope filter \(isotope grouper\)](#)

- Removes  $^{13}\text{C}$  isotope features from the feature list
- Has limited chance to detect isotope patterns for all features as only detected features (with all their filters and constraints are considered)
- Use the isotope finder for more sensitive detection of possible isotope signals

#### [13C isotope filter \(isotope grouper\)](#)

## Isotope pattern finder

- Searches for the isotope signals of selected chemical elements in the mass list of each feature.
- The isotope pattern detected by the **isotope finder** module has priority over the one detected by the **isotope filter (grouper)** module, if both are available.

### Isotope pattern finder

## Gap-filling

Gap-filling can be performed on the aligned feature lists to cope with missing features that might be artifacts of the feature-detection process. - For more details see the Gap-filling module.

---

## 5.1.3 Page Contributors

[SteffenHeu](#) (47.69%), [tdamiani](#) (27.69%), [Robin Schmid](#) (9.23%), [lalalana5](#) (9.23%), [lalalana5](#) (6.15%)

---

Last update: April 11, 2022 20:27:30

## 5.2 LC-IMS-MS Workflow Overview

---

Compared to regular LC-MS, LC-IM-MS data is more complex due to the additional separation dimension. Since some terms might not be straightforward for new users, a basic explanation of IM separation principles and the terminology used within this documentation is provided [here](#).

### 5.2.1 Supported formats

---

- Vendor formats:
  - .tdf (Native Bruker LC-IMS-MS and MALDI-IMS-MSI format)
  - .tsf (Native Bruker MALDI-IMS-MS (single shot) format)
  - .mzML
  - Created via [MSConvert](#) from native Bruker data
  - Created via [MSConvert](#) from native Waters/Agilent data
- 

### 5.2.2 Feature detection workflows

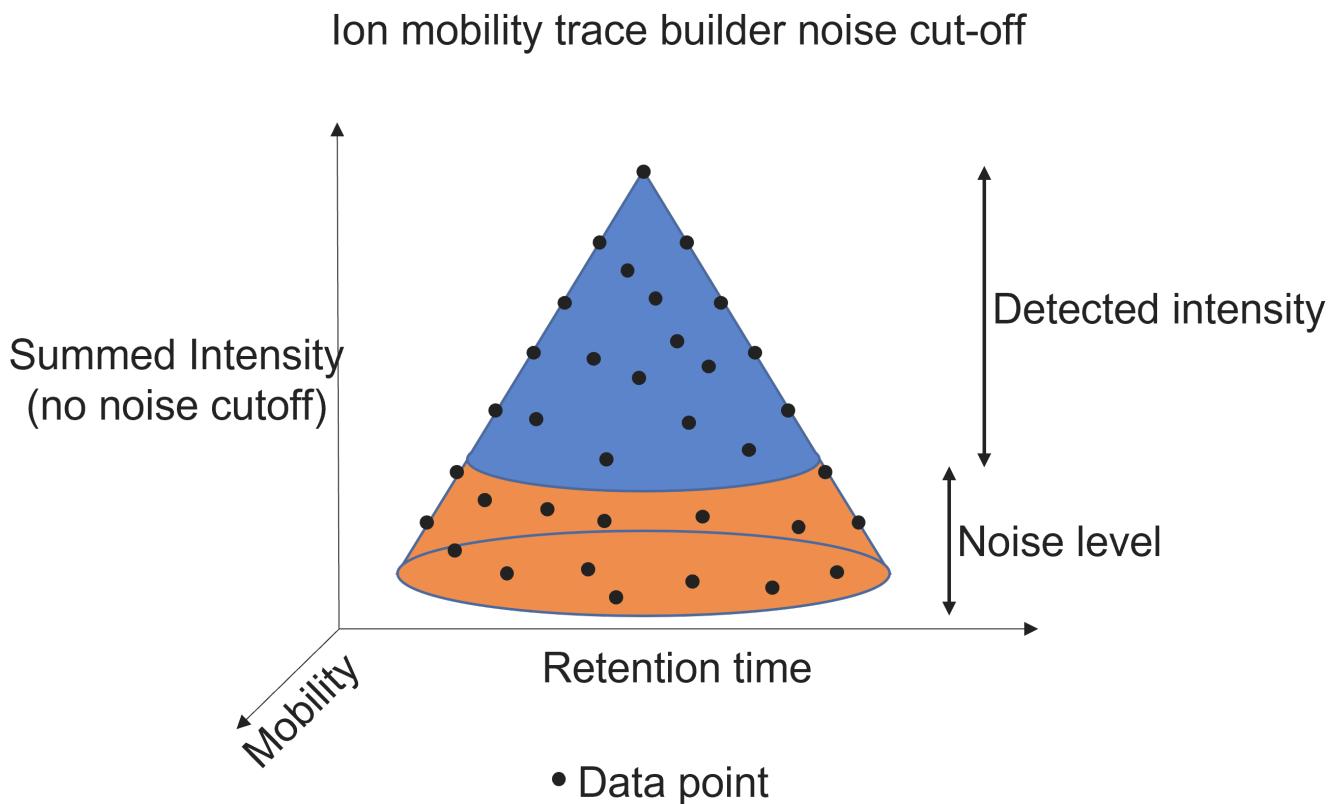
---

Ion mobility data can be processed in MZmine 3 in two ways. The first few steps are different for the two workflows (see below).

1. [LC-IMS-MS workflow via ADAP Chromatogram builder and IMS expander \( recommended\)](#)
2. [LC-IMS-MS workflow via Ion mobility trace builder / Recursive IMS builder](#)

While these lists might seem fairly similar, there are some differences in the processing approach. The LC-IMS-MS workflow builds ion mobility traces from the data in the mobility scans, whilst the LC-MS workflow builds EICs from the summed frames. For ion mobility data imported from .mzML files, accumulated frame spectra have to be built from the individual mobility scans after [mass detection](#). Since the mass detection impacts the computation of accumulated frame spectra in the same way it would impact the [ion mobility trace builder](#), the differences from this workflow and the [ADAP workflow](#) will be negligible. However, frame spectra for native Bruker .tdf raw data are summed by the vendor library during file import. Here, the frame spectra are generated from the raw data and thus result in higher intensities, since the low abundant data points on the edges of

the mobility and retention time peaks are not cut-off by the mass detection step. (see below)



Therefore, the more low abundant compounds might be detected, if the LC-MS workflow is recommended.

#### LC-MS workflow (recommended)

LC-IMS-MS data can also be processed via the regular LC-MS modules. If necessary, detected features can be expanded into the mobility dimension.

For this workflow, generation of summed frame spectra via the [Mobility scan merging](#) module is a mandatory step, if the data was imported from an .mzML file (automatically generated via native Bruker import).

- Data import
- Mass detection
- [Mobility scan merging](#) (mzML data)
- ADAP Chromatogram builder
- Smoothing in retention time dimension (optional)
- Resolving in retention time dimension
- Expanding EICs in mobility dimension
- Smoothing in mobility dimension (optional)
- Resolving in mobility dimension
- [Smoothing in rt and mobility dimension \(optional\)](#)
- Some recognised features might have rather noisy signals (in rt and mobility dimension) after the mobility resolving step. If smoother shapes are required, the smoothing can be reapplied afterwards. In that case, smoothing can be applied to both dimensions at once.

### LC-IMS-MS workflow

The LC-IMS-MS workflow will directly build [ion mobility traces](#) from the raw data in the mobility scans. This workflow does not necessarily require summed frame spectra. However, if extracted ion chromatograms shall be visualized via the [Chromatogram visualizer](#), the frame intensities are used. In case these are not present, the chromatograms will be blank. Note that feature intensities from the LC-IMS-MS workflow might not exactly match the frame chromatograms due to summing being executed prior to thresholding (for native Bruker data). Furthermore, multiple isomers might hide behind a single chromatographic peak.

- [Data import](#)
- [Mass detection](#)
- [Ion mobility trace builder](#)
- [Smoothing in retention time dimension \(optional\)](#)
- [Resolving in retention time dimension](#)
- [Smoothing in mobility dimension \(optional\)](#)
- [Resolving in mobility dimension](#)
- [Smoothing in rt and mobility dimension \(optional\)](#)
- Some recognised features might have rather noisy signals (in rt and mobility dimension) after the mobility resolving step. If smoother shapes are required, the smoothing can be reapplied afterwards. In that case, smoothing can be applied to both dimensions at once.

### 5.2.3 Graphical comparison of LC-MS and LC-IMS-MS data

---

#### Data comparison

### 5.2.4 Page Contributors

---

[SteffenHeu](#) (94.12%), [tdamiani](#) (5.88%)

---

Last update: April 5, 2022 14:10:02

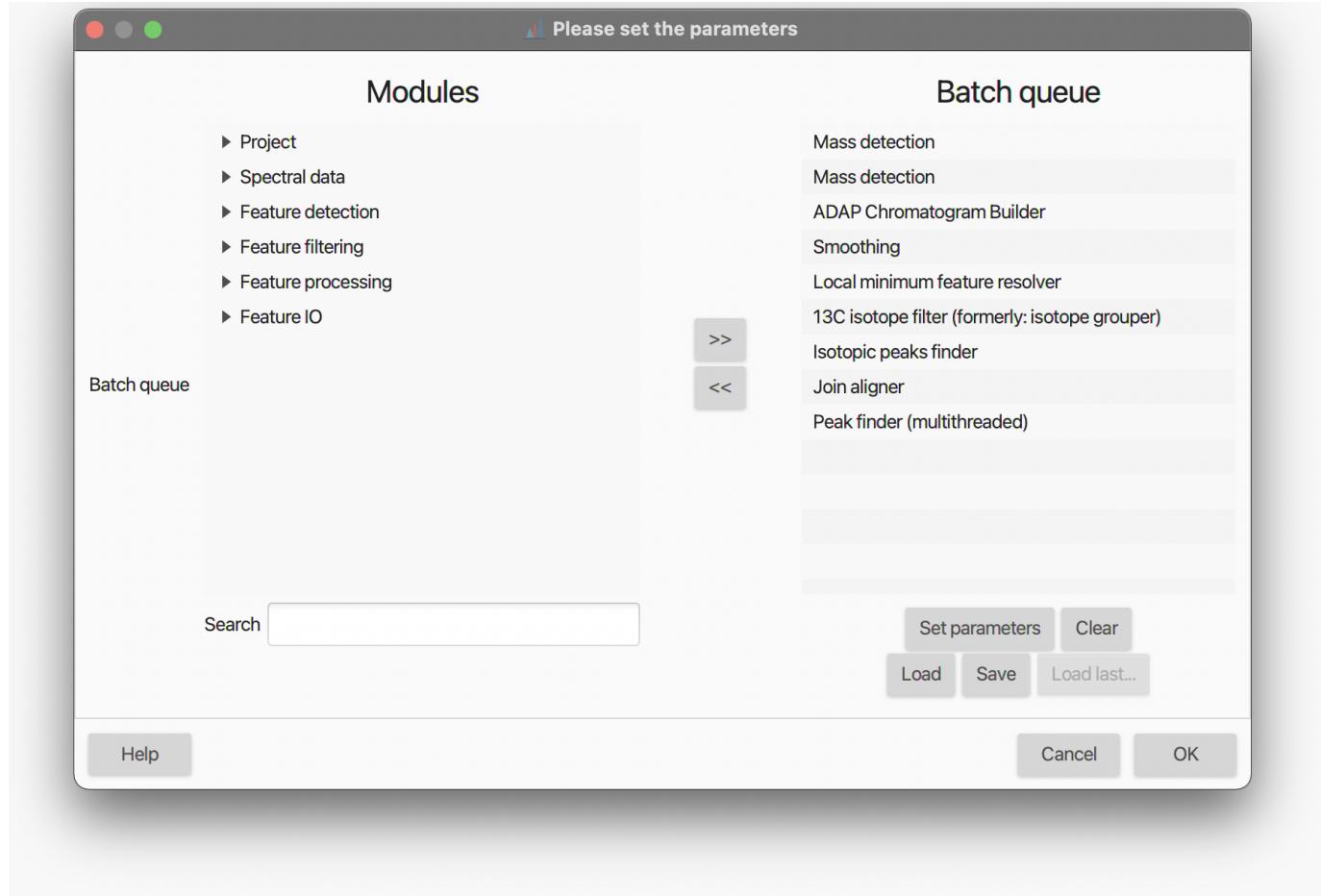
## 5.3 Batch processing

Besides the interactive [GUI](#), MZmine allows the user to run processing workflows in an automated manner using the "batch mode". Entire processing pipelines (including data import/export) can be run with few clicks, or even through the command-line application. This makes MZmine suitable to be integrated into automated data analysis pipelines (e.g. QC systems).

Batch files (XML format) are essentially lists of tasks run by MZmine one after another. Any of the methods available in MZmine 3 can be included in the batch file.

### 5.3.1 How to run batch processing

#### Project :material-arrow-right-thin: Batch mode



When a new step is added to the queue its parameter setup dialog is shown. The "Set parameters" button allows the user to modify a step's parameter settings. The "Clear" button removes all steps. The "Load" and "Save" buttons make it possible to read and write batch steps to XML files.

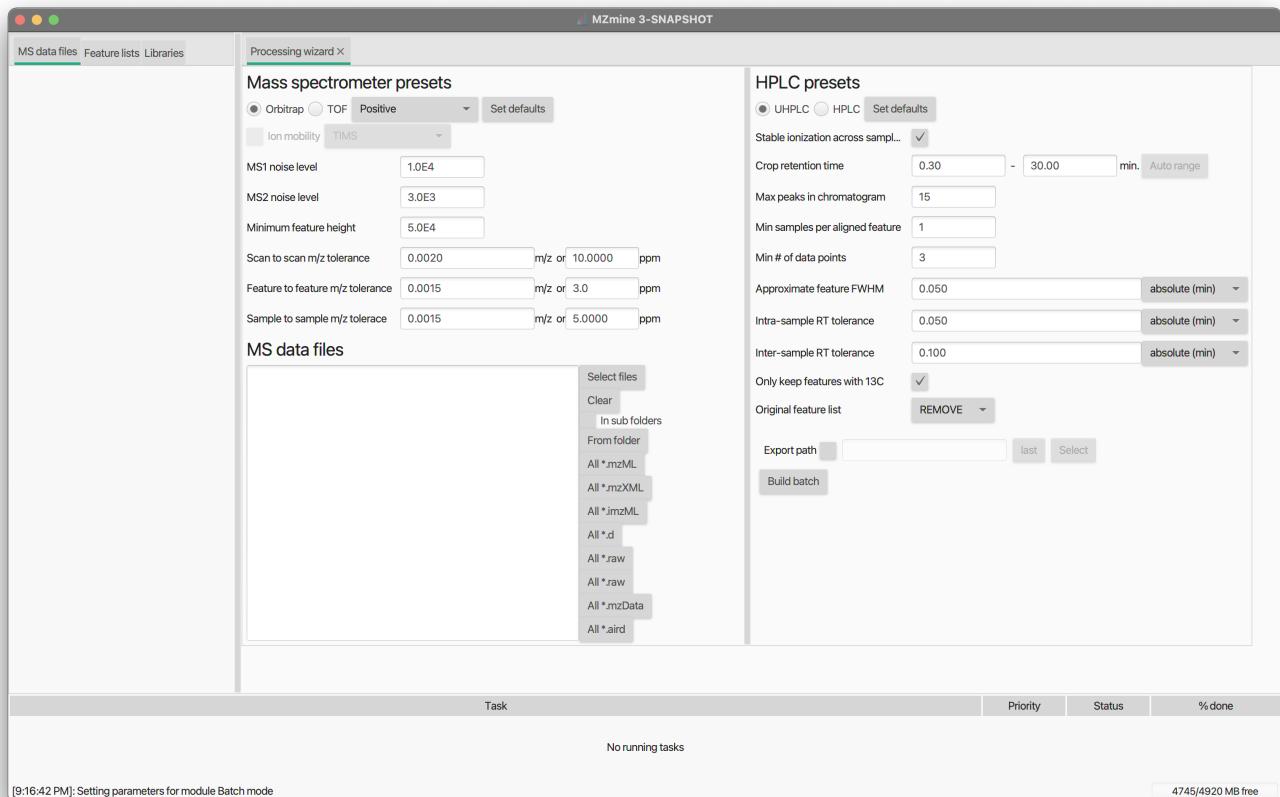
The first step of a batch queue is performed on those raw data files and/or peak lists selected by the user. The remaining steps are performed on the results produced by each preceding step (File/Feature list selection must be set to *Those created by previous batch step*). For example, if the first step of the batch queue is the [ADAP chromatogram builder](#), it will produce peak lists as a result. If the following step were Peak list deconvolution then it will be performed on the peak lists produced by the preceding Chromatogram builder step.

:octicons-alert-16: **Tip** MZmine "remembers" the last settings used.

## 5.4 Processing wizard

The processing wizard is intended to quickly set up a general workflow for the processing of untargeted LC-MS and LC-IM-MS data. By clicking the "Set default" button, default settings for mass and feature detection are also provided according to the selected MS type (Orbitrap or TOF) and LC system (UHPLC or HPLC). Once the desired parameters have been set, hit the "Build batch" button and a pre-populated batch window will open up.

### Tools :material-arrow-right-thin: Processing wizard



#### Mass spectrometers presets

MS type: When TOF is selected, the "Ion mobility" can be enabled

MS1 and MS2 noise level:

Minimum feature height:

Scan to scan m/z tolerance:

Feature to feature m/z tolerance:

#### HPLC presets

Stable ionization across samples:

Crop retention time:

Max peaks in chromatogram:

Min samples per aligned feature:

Min # of data points:

Approximate feature FWHM:

Intra-sample RT tolerance:

Inter-sample RT tolerance:

Only keep feature with 13C:

Original feature list:

Export path:



### Warning

The default settings were optimized on sample datasets used during the MZmine 3 development. Although probably suitable for many applications, it is strongly recommended not to blindly rely on them. Rather, optimal processing parameters should be chosen based on the LC-(IM)-MS system performance and data acquisition settings.

---

Last update: April 5, 2022 19:53:57

## 6. Visualisation

---

### 6.1 Raw data visualisation

---

todo

#### 6.1.1 Raw data overview (LC-MS)

---

todo

#### 6.1.2 Ion mobility raw data overview (LC-IMS-MS)

---

todo

#### 6.1.3 Chromatogram plot

---

todo

#### 6.1.4 Spectrum plot

---

todo

#### 6.1.5 Page Contributors

---

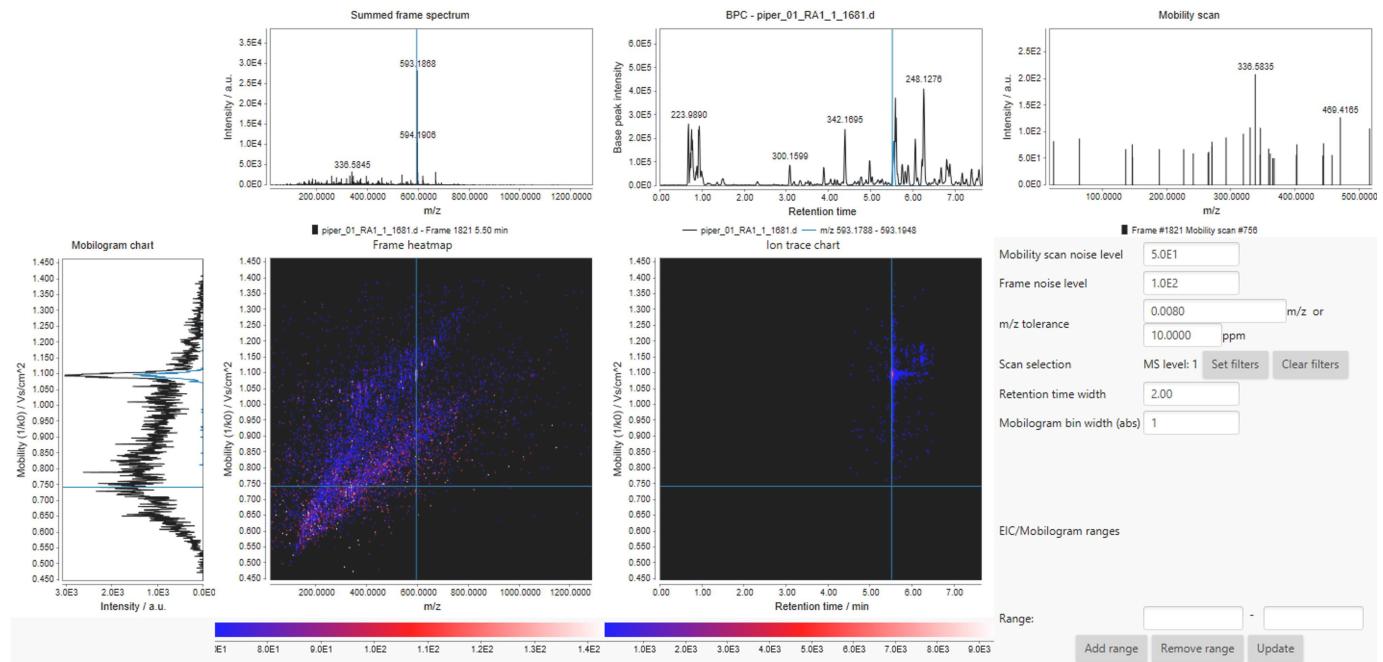
[SteffenHeu](#)

---

Last update: April 5, 2022 13:22:07

## 6.2 Ion mobility raw data overview (LC-IMS-MS)

The "Ion mobility raw data visualization" module allow a comprehensive navigation of the complex LC-IM-MS raw data. The screenshot below shows an example of LC-IM-MS data acquired with a Bruker timsTOF instrument:



The main window consists of 5 panels and a set of displaying parameters. All the panels are interconnected, which means that moving the cursor in one panel, automatically updates the others. Cursors are displayed as light-blue solid lines in the panels.

### 6.2.1 Summed frame spectrum panel [1]

The MS spectrum corresponding to each **frame** is shown in this panel. The displayed MS spectrum is the sum of all the **mobility scans** acquired over that frame (see [Ion mobility spectrometry terminology](#)).

### 6.2.2 BPC panel [2]

In this panel, the **base peak chromatogram** is displayed. Each data point corresponds to an individual **frame**. Moving the cursor frame-by-frame automatically updates the 'frame heatmap' and 'summed frame spectrum' panels. Moving the cursor frame-by-frame automatically updates the 'summed frame spectrum' panels as changing data point in regular LC-MS data would display a different MS scan. Since each frame is made of several **mobility scans**, the 'mobilogram chart' and 'frame heatmap' panels automatically updates too. *Note*. It is currently not possible to display the [TIC chromatogram](#)

### 6.2.3 Mobility scan [3]

Todo Note that this is the only panel that does not possess a cursor as [...].

### 6.2.4 Mobilogram chart [4]

Todo The signal intensity is displayed as a continuous colour scale.

### 6.2.5 Frame heatmap [5]

Todo The signal intensity is displayed as a continuous colour scale.

---

## 6.2.6 Ion trace chart [6]

Todo

---

## 6.2.7 Displaying parameters [6]

**Mobility scan noise level:** This parameter controls the signals shown in the XXX panels (panel n°X). For example, a noise level of 5.0E1 will show only the signals above this value (see below) SCREENSHOT

**Frame noise level:** This parameter sets a threshold for the signals shown in the "Summed frame spectrum panel" (panel n°X). Signals from MS spectra acquired over the same frame are summed and shown

**m/z tolerance:** Todo

**Scan selection:** Todo

**Retention time width:** Todo

**Mobilogram bin width (abs):** Todo

**EIC/mobilogram ranges:** Todo

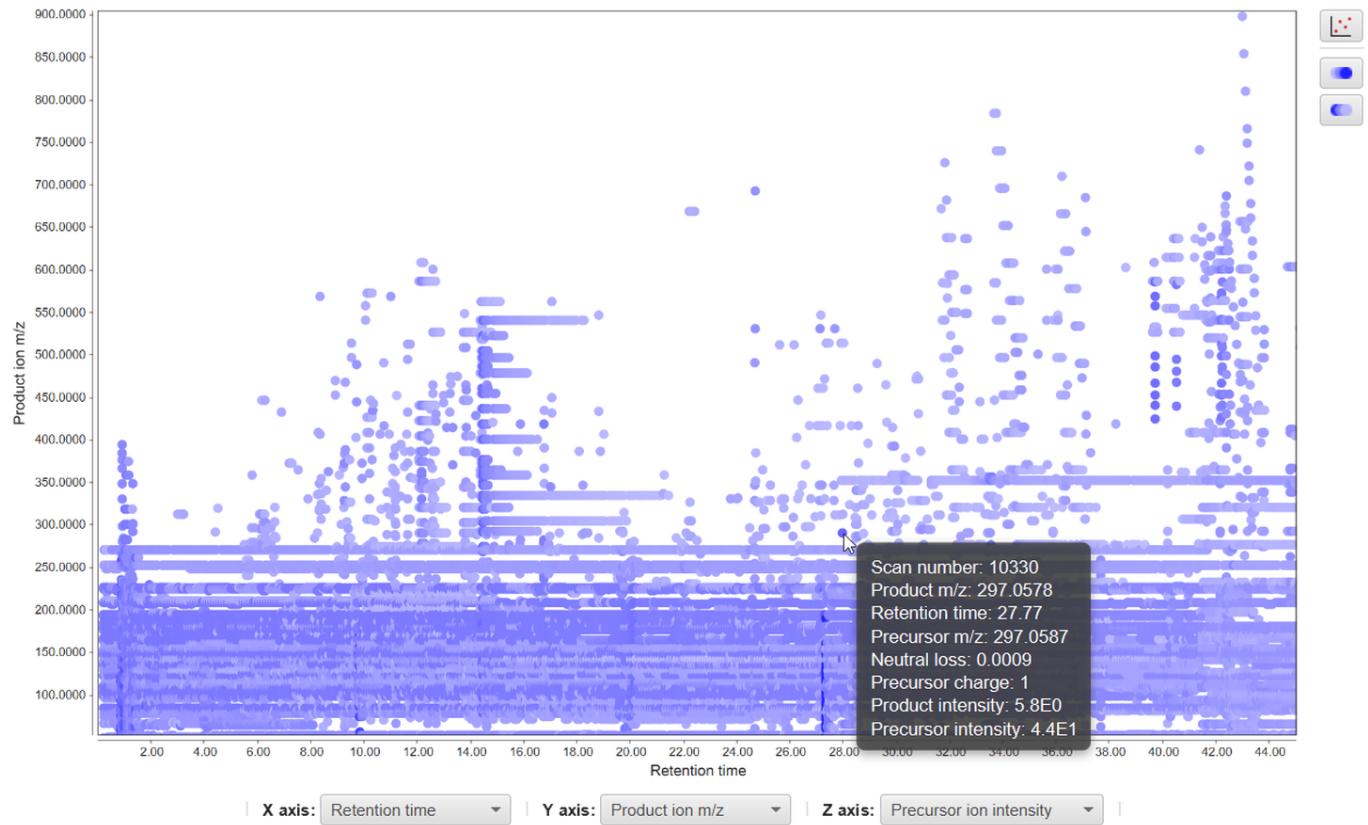
To-do list: - Explain EIC and EIC in mobilogram chart

---

Last update: March 11, 2022 09:00:25

## 6.3 MS/MS plot

This module provides a colored scatter plot of the MS/MS data. There are 4 options for X and Y axes: retention time, precursor ion m/z, product ion m/z, neutral loss and 3 options for Z axis (color): precursor ion intensity, product ion intensity, retention time. The module additionally allows you to filter ions by their intensities and to perform diagnostic fragmentation filtering. In order to focus on the values of interest you can highlight specific data points and sort them by color axis. This tool can be very useful to get an overview of large amounts of MS/MS data by tuning parameters and filters.



### 6.3.1 Parameters

#### Raw data file

Selection of the raw data file to visualize. Only one file can be selected.

#### X axis

Selection of the values for X axis. There are 4 options available: Retention time, Precursor ion m/z, Product ion m/z, Neutral loss.

#### Y axis

Selection of the values for Y axis. Options are the same as for X axis.

#### Z axis

Selection of the values for Z axis. There are 3 options available: Precursor ion intensity, Product ion intensity, Retention time.

#### MS level

MS level of the scans to be plotted.

#### Retention time

Retention time range.

**m/z range**

Range of m/z values for precursor ions in MSn scans.

**m/z tolerance**

Maximum allowed difference between two m/z values to be considered same.

**Intensities filtering**

Optional parameter to filter ions by intensity. There are 3 different ways of filtering:

- Number of best fragments - Number of ions with highest intensities from each scan to be visualized.  
For example 5(for each scan 5 ions with highest intensities will be plotted).
- Base peak percent, % - Ions with intensity values lower than the given percent of base peak intensity will be plotted.  
For example 95(ions with intensity values lower than 0.95 multiplied by base peak intensity will not be plotted).
- Intensity threshold - Ions having intensities lower than the given value will not be plotted.  
For example 6.0E6(ions with intensity values lower than 6.0E6 will not be plotted).

**Diagnostic fragmentation filtering**

Optional parameter for diagnostic fragmentation filtering described below. It has 2 subparameters: diagnostic product ions and diagnostic neutral loss values. Scans not containing any ion satisfying each input criterion will not be considered for the visualization.

### 6.3.2 Diagnostic fragmentation filtering

---

Due to common structural features, compounds within the same class undergo similar MS/MS fragmentation and as a result of many identical product ions and/or neutral losses. Diagnostic fragmentation filter (product ion filter) is a post-acquisition approach to screen LC-MS/MS datasets for entire classes of both known and unknown natural products. This tool searches all MS/MS spectra for product ions and/or neutral losses that has defined as being diagnostic for the entire class of compounds. In other words it screens LC-MS/MS datasets for MS/MS spectra containing production ions and/or neutral losses that are specific to that class of compounds. The user defines the diagnostic product ions and/or the diagnostic neutral loss values (Da) to use in the filtering.

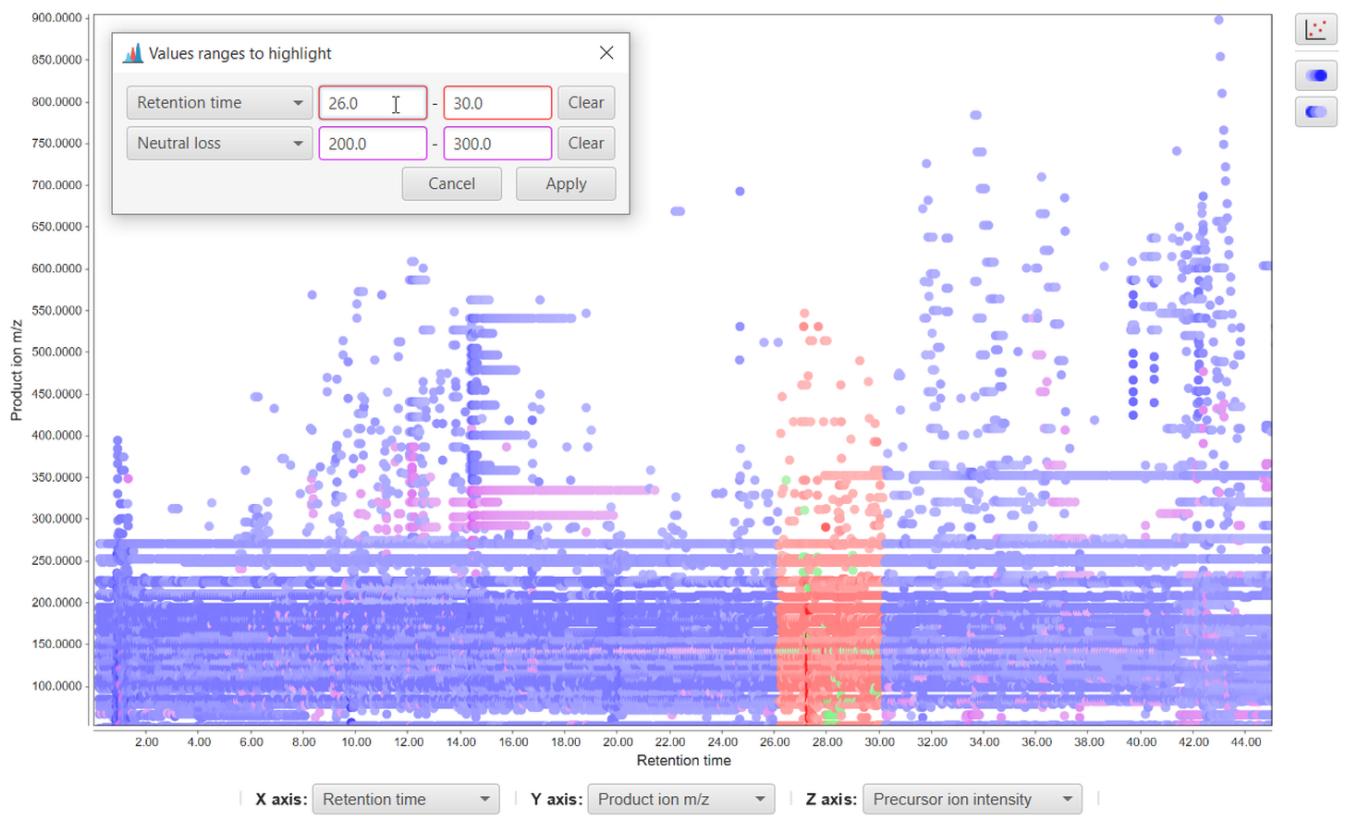
The user can also define the minimum diagnostic ion intensity (% base peak) to use in the filtering. If a recurrent neutral loss occurs, a line pattern in the plot can be observed. If compounds carrying those diagnostic product ions and/or the neutral loss values are detected the resulting plot will show their product ion m/z and precursor ion m/z. Additionally, an output file may be specified that will output the results of the filtering. For a detailed view of diagnostic fragmentation filtering: [Walsh, Jacob P., et al. "Diagnostic Fragmentation Filtering for the Discovery of New Chaetoglobosins and Cytochalasins." Rapid Communications in Mass Spectrometry \(2018\).](#)

### 6.3.3 Functionality

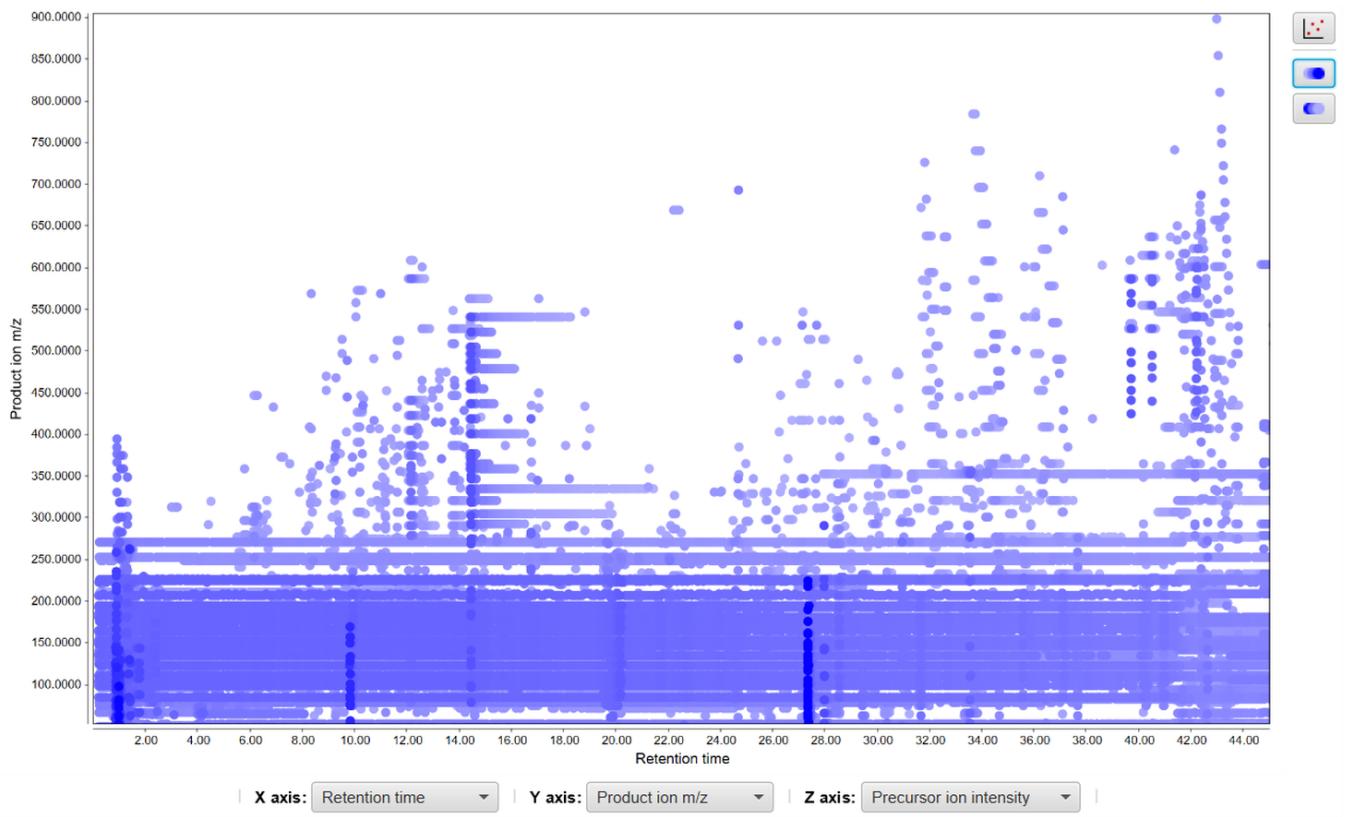
---

This plot is using the third part library JfreeChart for its basic functionality.

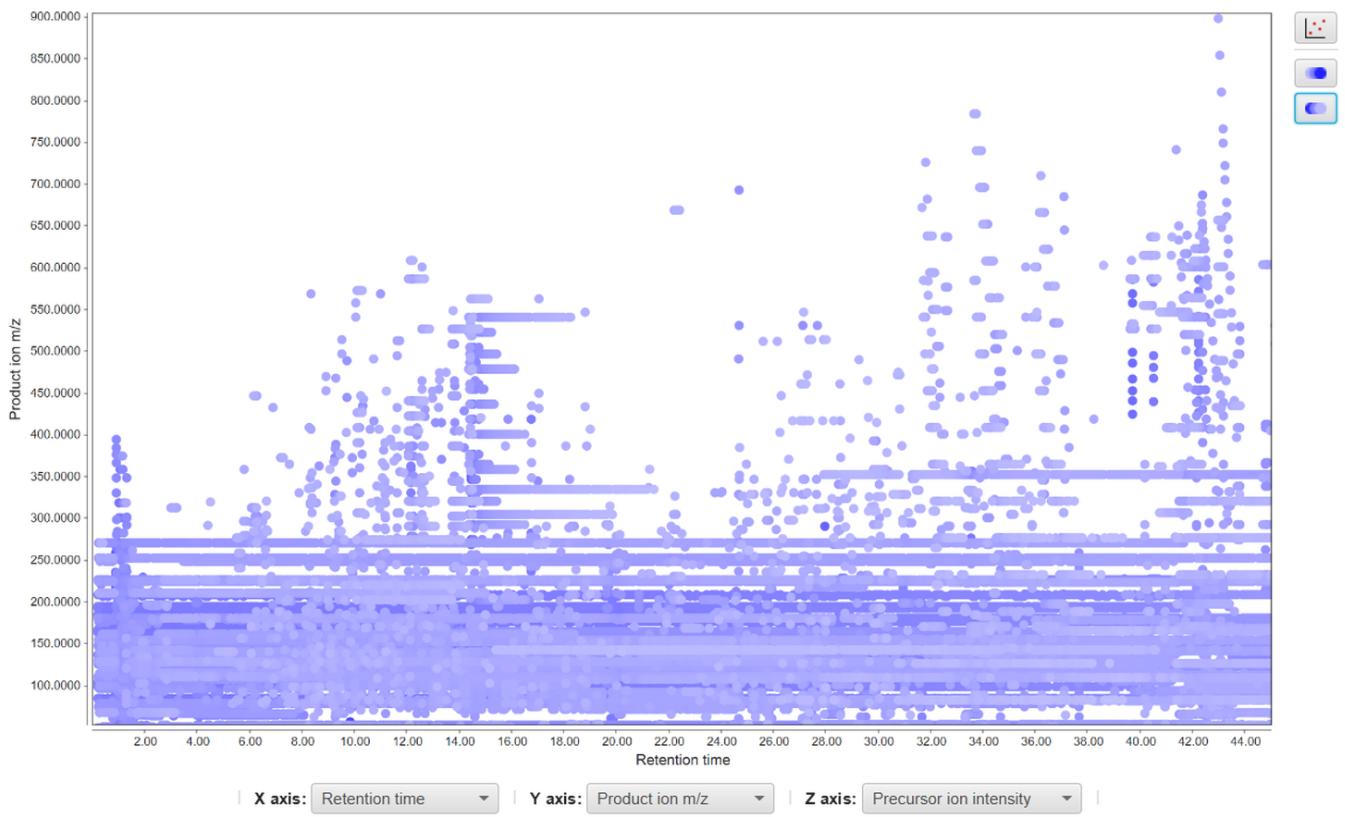
- Drag the mouse from left to right - selecting the area to zoom
- Drag the mouse from right to left - zoom out
- Select combo boxes below - change axes types
- Hold the mouse on data point - show detailed information in a tooltip
- Double click on data point - show spectrum plot
-  - highlight points representing ions with specific values given by input ranges (Note: colors of range input boxes determine the highlighting color, green color denotes ions satisfying both ranges)



-  - show intense points in front



- show pale points in front



Last update: April 1, 2022 14:27:24

## 7. Additional resources

---

### 7.1 General terminology

---

#### 7.1.1 MS

---

##### Precursor and fragment ions

The precursor ion (a.k.a. "parent ion") is the ion that dissociates to a smaller fragment ions in a MS/MS experiment. A fragment ion (a.k.a. "daughter ion" or "product ion") is the charged product of an ion dissociation. A fragment ion may be stable itself or may dissociate further to form other charged fragment ions and neutral species of successively lower mass.

##### Accurate mass, exact mass and mass accuracy

The accurate mass is the experimentally-determined mass of an ion measured with an high-resolution mass spectrometer. The exact mass is the calculated mass of an ion based on its elemental formula, isotopic composition and charge state. While the accurate mass is an experimentally-measured quantity, the exact mass is a theoretically-calculated quantity. The mass accuracy is defined as the difference between the measured value (accurate mass) and the true value (exact mass). It can be expressed either in absolute (mDa) or relative (ppm) terms.

##### Monoisotopic mass

Exact mass of an ion calculated using the mass of the lightest isotope of each element.

##### Mass resolution

Todo. Often called/expressed as mass resolution power

##### Full scan acquisition mode

Todo

##### Data-dependent acquisition mode

In data-dependent acquisition (DDA) schemes, the mass spectrometer detects 'suitable' precursor ions in each MS scan and selects them for fragmentation in consecutive MS2 scans.

Todo:'Cycle time' and 'topN' acquisition schemes

##### Data-independent acquisition mode

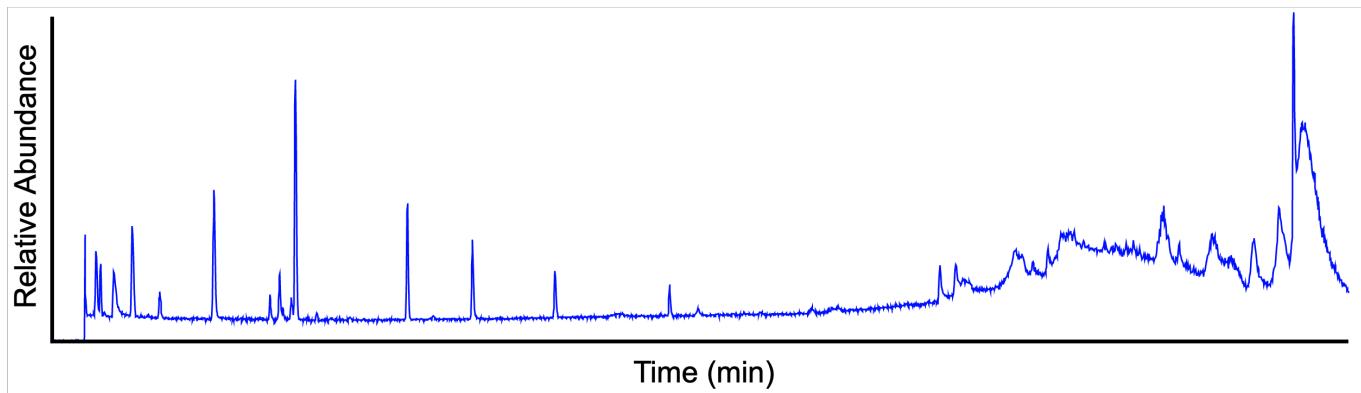
Todo

#### 7.1.2 LC-MS

---

##### Total ion current chromatogram

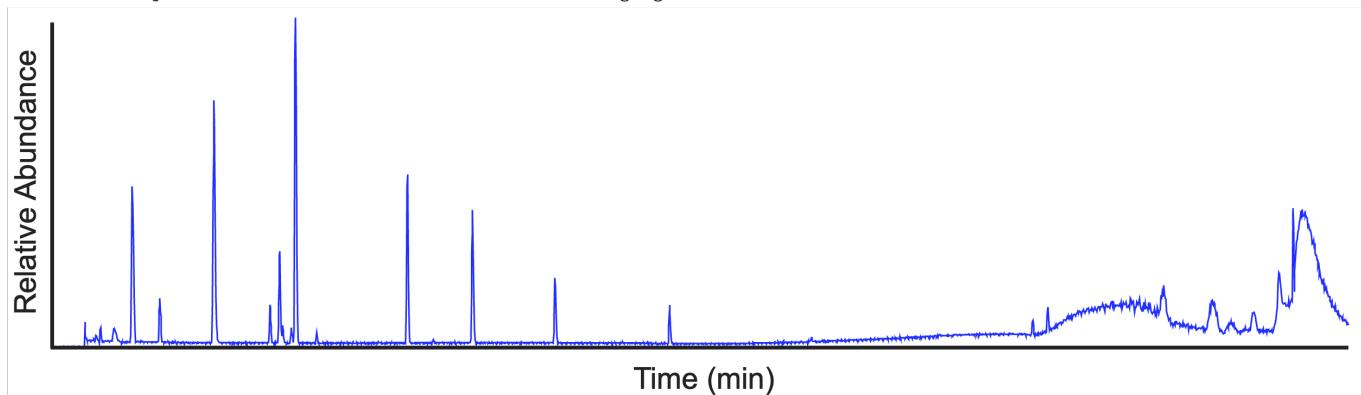
The total ion current (TIC) chromatogram displays the summed signal intensity (y-axis) over the entire m/z range at any one retention time point (x-axis) in the LC-MS run. The following figure shows a TIC chromatogram of a 9-compounds mixture analysed on LC-MS system.



**Note.** In complex samples, the TIC chromatogram often provides limited information as multiple analytes elute simultaneously obscuring individual species.

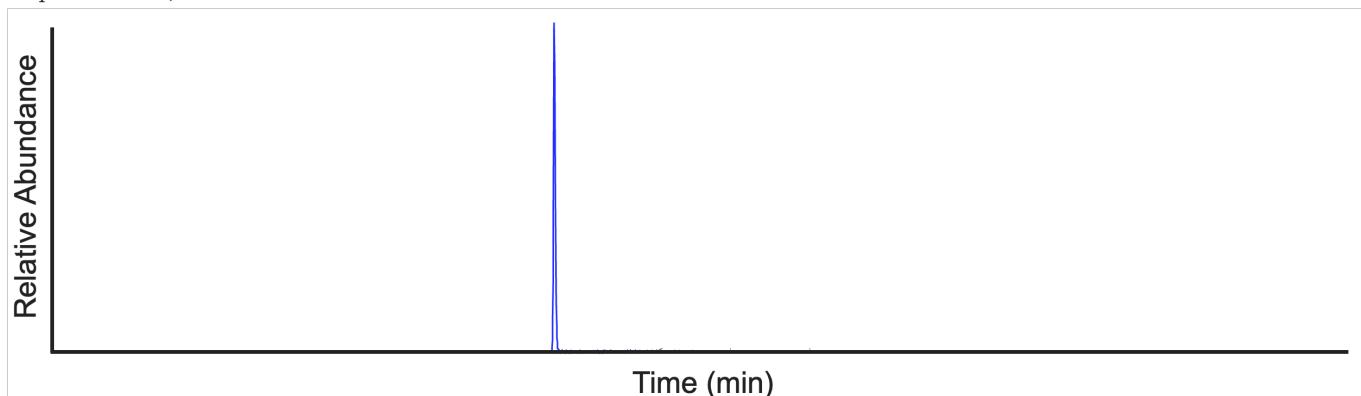
#### Base peak chromatogram

The base peak chromatogram (BPC) displays the signal intensity of the most intense mass peak in the MS spectra at any one retention time point (x-axis) in the LC-MS run. The following figure shows the same data as above, visualized in BPI mode.



#### Extracted ion chromatogram

The extracted ion chromatogram (EIC) displays the signal intensity of a specific  $m/z$  value, within a defined tolerance (e.g.  $\pm 5$  ppm), at any one retention time point in the LC-MS run. The following figure shows the EIC of  $m/z$   $455.2945 \pm 5$  ppm (same sample as above).



#### Chromatographic resolving

Peak overlapping, or co-elution, is a common problem in any chromatographic separation technique. In the case of LC-MS (especially untargeted *omics* analysis), it is virtually impossible to obtain a full baseline separation for the hundreds (or

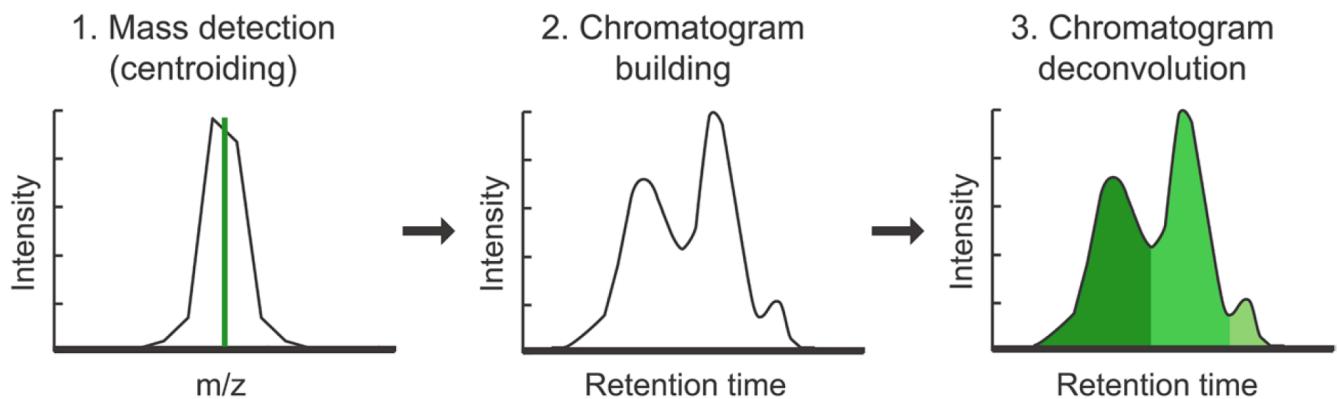
thousands) of analytes eluted through the column. The split of partially-overlapping and shoulder peaks into individual features is generally referred to as *chromatographic resolving* and is one of the most crucial steps of data processing. TO FINISH.

## 7.2 MZmine-specific terminology

### Masses and Features

In MS data processing, the term *mass* is normally used to refer to an individual signal in a mass spectrum, which corresponds to an ion detected by the mass spectrometer (see [Mass detection](#)).

In LC-MS, a *feature* is defined as a bounded, two-dimensional ( $m/z$  and RT dimensions) signal characterized by a pair of  $m/z$  and RT values and associated with the detected signal intensity. In the case of LC-IM-MS data, a feature is also characterized by the ion mobility value recorded for the ion (see [LC-MS and LC-IMS-MS data comparison](#)). MZmine 3 provides a selection of different algorithms for LC-(IM)-MS feature detection, depending on the nature of the MS data (e.g. mass accuracy and resolution). All the algorithms follow the same logic: EICs are constructed starting from each  $m/z$  value in the mass lists and subsequently deconvoluted into individual features (see figure). Further, additional information, such as isotope pattern, adduct type, etc. can be associated to the individual features.



### Mass list

In MZmine we call *mass list* the output of the [mass detection](#) module. A *mass list* is a list of  $m/z$  values (and corresponding signal intensities) found in each mass spectrum (MS or MSn) of each processed raw data file. Every mass spectrum contained in the raw file is processed individually and the signals exceeding the set noise threshold are included in the mass list. See [Mass detection](#) module.

### Feature list

In MZmine, *feature lists* are the output of the feature detection process (see [Masses and features](#)). The set of detected features in each LC-MS run is stored as a list, hence the name "feature list" (see, for example, [ADAP chromatogram builder](#) and [Local minimum resolver](#) for more details). Multiple feature lists can undergo further processing (e.g. feature alignment) which results in a table (often referred to as *feature table*) where samples are arranged in columns, features in rows and each entry contains the signal intensity detected for the corresponding feature, in the corresponding sample.

### Intra and inter-scan tolerances

To-do

#### 7.2.1 References

- Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* (2010). DOI: 10.1186/1471-2105-11-395

- Pluskal, T. et al. Processing Metabolomics and Proteomics Data with Open Software: A Practical Guide, Chapter 7: Metabolomics Data Analysis Using MZmine (2020). DOI: 10.1039/9781788019880-00232
  - Smoluch M., Piechura K. Mass Spectrometry: An Applied Approach, Chapter 3: Basic Definitions (2019). DOI: 10.1002/9781119377368.ch3
- 

Last update: April 5, 2022 19:53:57

## 7.3 Ion mobility spectrometry terminology

---

### 7.3.1 Background

Ion-mobility mass-spectrometry, here simply referred to as ion-mobility (IM), is an analytical technique where ions are separated through a gas-filled mobility cell prior to the MS acquisition. Ions drift through the IM cell with different velocity based on their interaction with the buffer gas, which allows for the separation of different shaped molecules. Modern devices are able to perform IM separation on a millisecond timescale, typically within 10 to 100 ms. Thus, IM nicely fits in-between LC separation (~seconds timescale) and MS detection of TOF instruments (~microseconds timescale). This allows LC-IM-MS instruments to acquire several MS spectra during each [accumulation](#), without incurring sensitivity loss. For example, assuming a typical 100  $\mu$ s MS-acquisition time of TOF analyzers, around 1000 spectra can be recorded within 100 ms of IM separation. Therefore, as opposed to LC-MS, multiple MS (or MS<sub>2</sub>) spectra are associated to each RT in LC-IM-MS data. A more detailed explanation of LC-MS and LC-IMS-MS raw data structure is provided [here](#).

#### Trapped ion mobility spectrometry (TIMS) - Todo

Trapped ion mobility spectrometry (TIMS) reverses the concept of traditional drift tube IM. Rather than moving ions through a stationary gas, TIMS holds ions stationary against a moving gas and then releases them according to their mobility. Video: <https://www.youtube.com/watch?v=cWjz32wky2A>

#### Time-dispersive ion mobility spectrometry (DTIMS and TWIMS) - Todo

Time-dispersive IM devices include "traditional" drift tube (DTIMS) and travelling-wave (TWIMS) devices. In classic DTIM, ions migrate through an inert buffer gas under the influence of a weak electric field, whereas collisions with buffer gas molecules retard the progress of the ions. As larger ions have more collisions with the gas, they are more strongly retarded than their smaller counterparts. Thus, smaller ions, having a smaller cross section, arrive earlier at the detector than ions with a larger collisional cross section (CCS). The ion mobility K is then defined as the ratio of the analyte's steady-state net drift velocity to the applied electric field, and it is convention to calculate the reduced ion mobility K<sub>0</sub> at standard pressure and standard temperature, often reported as the inverse reduced ion mobility 1/K<sub>0</sub>.

### 7.3.2 Terminology

---

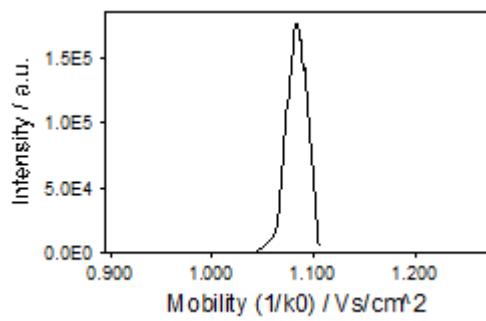
#### Accumulations, Mobility Scans and Frames

Although mainly used for TIMS, the term "*accumulation*" refers to the pack of ions gathered at the head of the IM device prior to the release and separation in the IM cell. As explained [above](#), since the accumulation-separation cycle typically last ~100 ms, multiple MS spectra (referred to as "*mobility scans*" in MZmine) are acquired during each cycle. The set of *mobility scans* collected during each IM separation constitutes a "*frame*". A *frame* can be seen as the IM separation of a single *accumulation*, along which multiple MS spectra are collected. Several *frames* are contained within one LC peak. Thus, the frame number are a natural unit to measure chromatographic RT. See [here](#) for more details.

--FIGURE HERE--

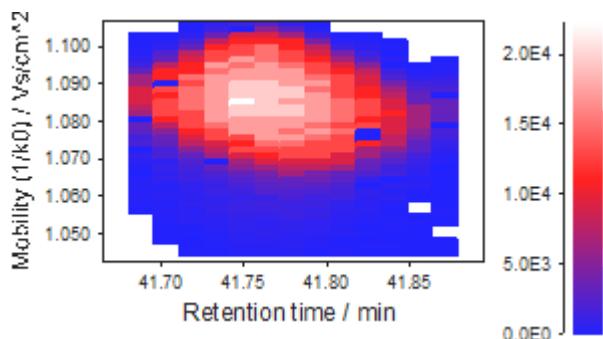
#### Mobilograms

A "*mobilogram*" represents the intensity of an m/z or m/z range along the mobility axis. A *mobilogram* may be build from multiple frames and summed or built from a single frame.



#### Ion mobility trace

An "ion mobility trace" basically represents a mobility resolved extracted ion chromatogram (EIC).



#### Cross Collisional Section

IM-derived CCS values can be used as an additional molecular descriptor to support the compound unknown identification process.(Paglia et al. 2014) However, the number of acquired spectra per run increases from several thousand to several million, requiring memory-efficient software and new processing algorithms.

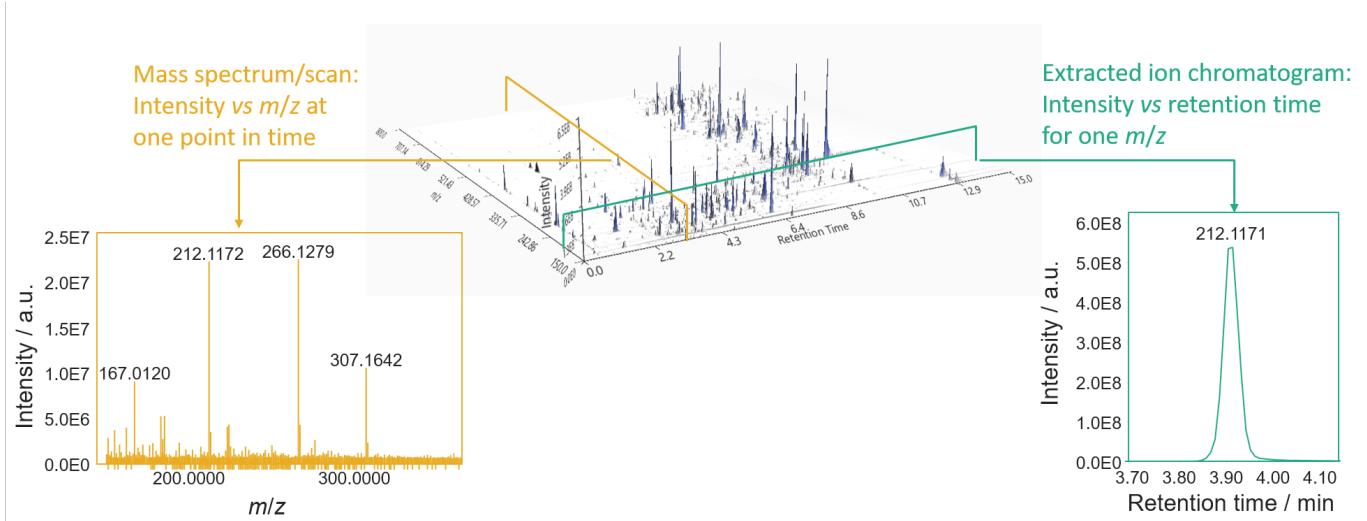
#### 7.3.3 References

<https://doi.org/10.1074/mcp.TIR118.000900>

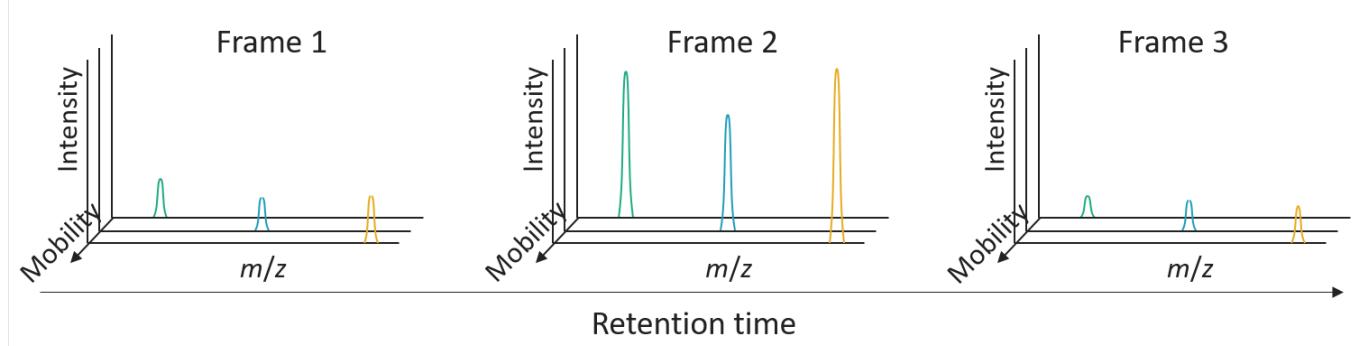
Last update: February 14, 2022 22:31:14

## 7.4 Graphical comparison of LC-MS and LC-IMS-MS data

Classic LC-MS data consists of three dimensions: m/z, intensity, and retention time. At every retention time, a whole mass spectrum is acquired (yellow). Putting all scans together creates a three-dimensional plane. By slicing the three-dimensional data at a single m/z (+- a tolerance), EICs can be created (green).



On the other hand, ion mobility resolved data consists of a three-dimensional data plane at each retention time. The three dimensions being m/z, intensity, and mobility (as drift time (ms) or inverse reduced mobility  $1/k_0$  [Vs/(cm<sup>2</sup>)]). The 3D projection of regular LC-MS data can be created by summing all mobility scans of a frame to create a frame spectrum. (see [Mobility scan merging](#))



Last update: March 10, 2022 15:42:38

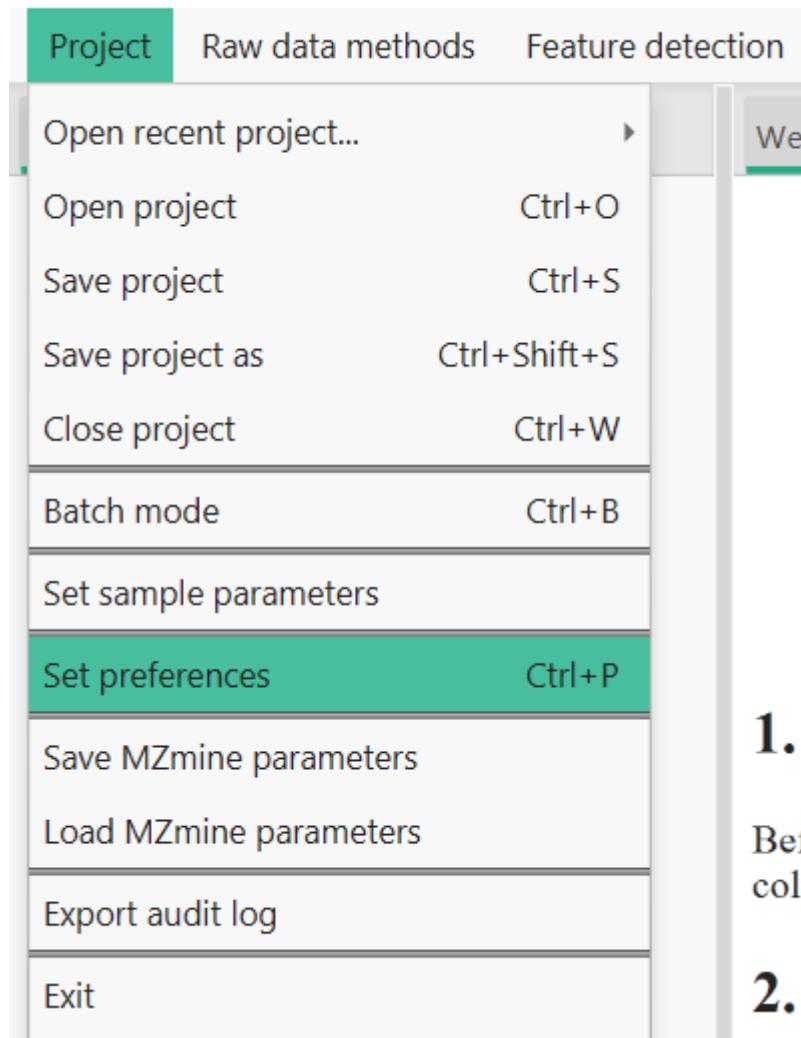
## 8. Performance options

---

This section contains information on how to tune MZmine 3 for different systems.

### 8.1 Preferences

The preferences can be changed in MZmine's graphical user interface by accessing *File/Set preferences* from the menu. The choices will be stored in a (hidden) *.mzmine3.conf* file in the user's home directory (Windows: *C:\Users\USERNAME*) once MZmine is closed.



1.

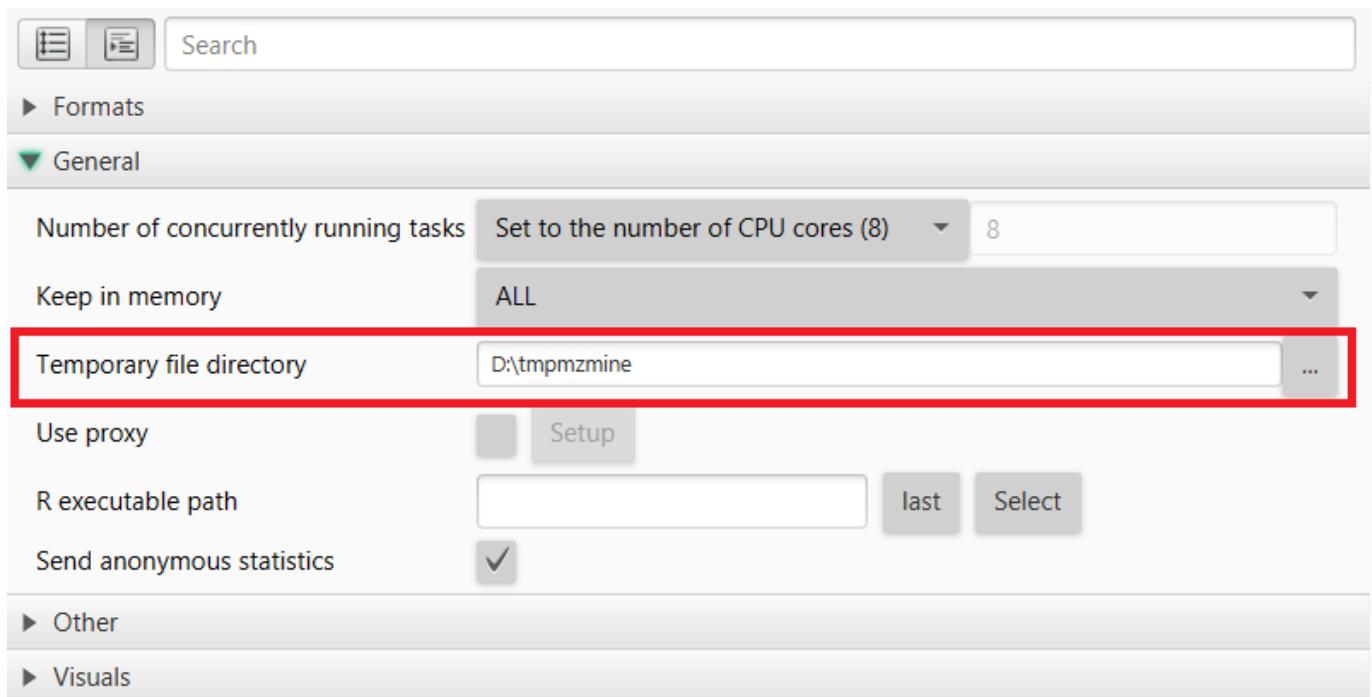
Be  
col

2.

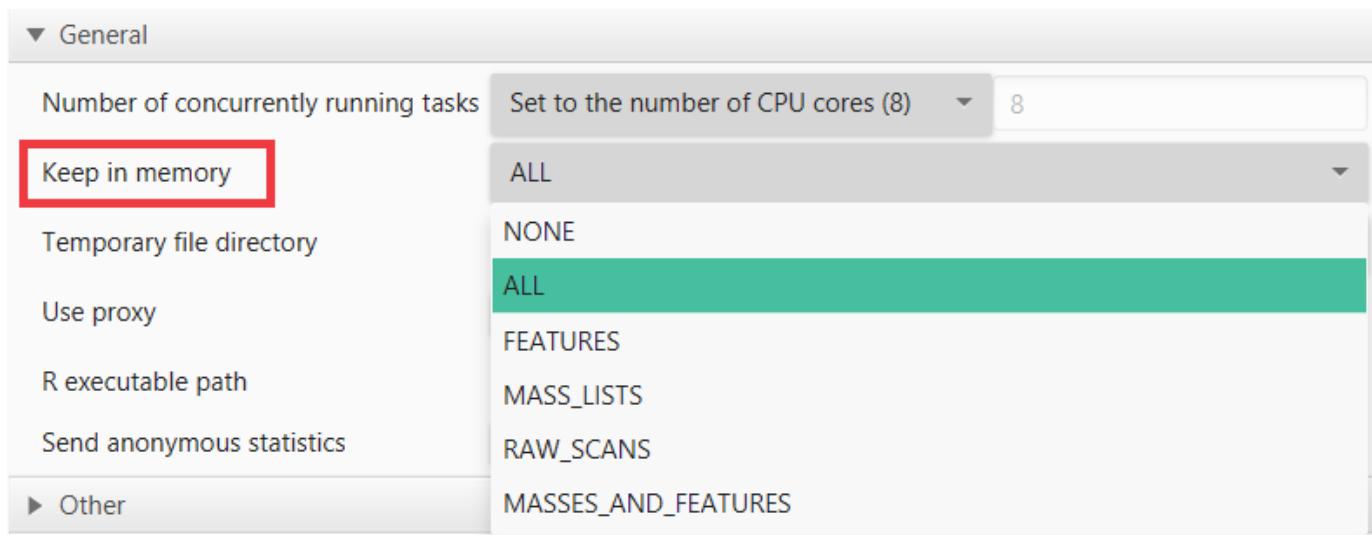
#### 8.1.1 Temporary files

MZmine will create multiple temporary files at various times of the processing stage, e.g., when importing spectral data, running mass detection, or creating feature lists. These files will be stored in a folder that can be specified in the preferences.

We recommend putting this folder on an SSD drive, ideally an M.2 for the best performance. The temporary files will be deleted when MZmine is closed (Mac & Linux) or when a new session is started (Windows).



### 8.1.2 Memory options



The parameter **Keep in Memory** defines what data is kept in memory (RAM) or otherwise memory mapped to the temp directory.

- Generally this setting should be *none (default)*.
- If memory is no issue this option might be set to *all* process all spectral and feature data in memory.
- The option *masses\_features* keeps centroid mass lists and features in memory while memory mapping raw spectral data.
- The option *mass\_lists* will keep only mass lists in RAM, while memory mapping the raw spectral data and features.

## 8.2 Logs

Currently, the logs are written to an *mzmine\_0\_0.log* file in the user's home directory. Please submit your log files together with any issues on [GitHub](#).

Last update: April 8, 2022 08:12:41

## 9. Command-line arguments

---

Command-line arguments offer a variety of options that generally override the corresponding parameters in the preferences.

### Windows

An easy way to start MZmine with arguments is to create a shortcut to the MZmine.exe, right-click, and define the target with additional arguments. This example runs MZMine in batch mode (headless), imports the specified batch file, overrides the memory management to **none** (which is the default), effectively using memory mapping to store and access spectral, centroid, and feature data from temporary files stored in the defined temp directory. By leaving out the *memory* or *temp* arguments, the values stored in the current *preferences* file will be used, or the default values if no *preferences* file was found.

#### Start MZmine batch with memory mapping (DEFAULT)

```
"C:\Program Files\MZmine\MZmine.exe" -batch "D:\batch\my_batch_file.xml" -memory none -temp "D:\tmpmzmine"
```

#### Start MZmine batch on machines with enough memory (RAM) with -memory all

```
"C:\Program Files\MZmine\MZmine.exe" -batch "D:\batch\my_batch_file.xml" -memory all -temp "D:\tmpmzmine"
```

### 9.0.1 Argument table

Argument	Options (default)	Description
-batch	a path, e.g. "D:\batch.xml"	Path to batch file
-memory	<b>none</b> , all, features, centroids, raw, masses_features	Defines what data is kept in memory (RAM) or otherwise memory mapped to the temp directory. Generally this setting should be <i>none</i> . If memory is no issue this option might be set to <i>all</i> process all spectral and feature data in memory. The option <i>masses_features</i> keeps centroid mass lists and features in memory while memory mapping raw spectral data.
-temp	a path, e.g., "-temp "D:\tmpmzmine\"	The defined directory should be on a fast drive (usually SSD > HDD > network drive) with enough free space. Local drives are usually preferred. MZmine uses memory mapping to efficiently store and access spectral and feature data. This can lead to a considerable temporary consumption of disk space. Make sure that the selected drive has enough space (maybe 20 GB + 1 GB/10 files; generously over estimated).

Last update: April 8, 2022 08:12:41

## 10. How to contribute

---

### 10.1 Contribute to the MZmine documentation

#### 1. Make a GitHub Account

You'll need to make a [GitHub Account](#).

#### 2. Click Edit Button on Page You Want to Edit

##### MZmine 3 Documentation

[Home page](#)  
[Main window overview](#)  
[LC-MS workflow](#)  
[LC-IMS-MS workflow](#)  
[Raw data visualisation](#)

##### LC-IMS-MS Workflow

##### Supported formats

- Vendor formats: \*

  - .tdf (Native Bruker LC-IMS-MS and MALDI-IMS-MSI format) \*
  - .tsf (Native Bruker MALDI-IMS-MS (single shot) format)

- .mzML \*

  - Created via [MSConvert](#) from native Bruker data \*
  - Created via [MSConvert](#) from native Waters data

	<a href="#">Table</a>
	<a href="#">Support</a>
	<a href="#">Background</a>
	<a href="#">Terminology</a>
	<a href="#">Motif</a>
	<a href="#">Form</a>
	<a href="#">Français</a>
	<a href="#">Mobile</a>
	<a href="#">Ionization</a>
	<a href="#">Raw data</a>
	<a href="#">Raw results</a>
	<a href="#">Mass spectra</a>
	<a href="#">Sequence analysis</a>

#### 3. Fork the Repository When Prompted (only the first time)



You need to fork this repository to propose changes.

Sorry, you're not able to edit this repository directly—you need to fork it and propose your changes from there instead.

[Fork this repository](#)

[Learn more about forks](#)

#### 4. Make the Edits in MarkDown

mzmine\_documentation / docs / Ion-mobility-data-proc Cancel changes

Spaces 3 Soft wrap

```

1 # LC-IMS-MS Workflow
2 ## Supported formats
3
4 * Vendor formats:
5 *
6     * .tdf (Native Bruker LC-IMS-MS and MALDI-IMS-MSI format)
7 *
8     * .tsf (Native Bruker MALDI-IMS-MS (single shot) format)
9 * .mzML
10 *
11     * Created via [MSConvert](https://proteowizard.sourceforge.io/download.html) from native Bruker
12     data
13 *
14     * Created via [MSConvert](https://proteowizard.sourceforge.io/download.html) from native Waters
15     data
16
17 **Note**: mzML via MSConvert from Agilent raw data can be imported, but certain restrictions might
18 hinder processing workflows due to the nature of the raw data format.
19
20
21 ***
22
23 ## Background information and terminology
24
25 Since ion mobility spectrometry (IMS) resolved data is more complex due to the additional dimension
26 when compared to regular LC-MS data, some terms shall be clarified before going into details of the
27 processing steps.
28
29 ### Mobility separation and data format
30
31 Ion mobility separation usually occurs on the millisecond timescale, fitting nicely in-between
32 liquid chromatography (LC) (few seconds per chromatographic peak) and mass spectra acquisition of
33 TOF instruments (several micro seconds). Therefore, the mobility dimension can be resolved by
34 acquiring multiple spectra during a mobility separation (e.g. 1000 spectra per 100 ms). This leads
35 to multiple mass spectra acquired at one IMS accumulation. Thus, at one retention time, multiple
36 spectra are acquired. A detailed comparison of LC-MS and LC-IMS-MS raw data can be

```

Attach files by dragging & dropping, selecting or pasting them.

## 5. Propose Changes

Please describe the change you are making.

### Commit changes

update mobility resolving step

add msms pairing description in mobility resolving step

steffen.heuckeroth@gmx.de

Choose which email address to associate with this commit

- o Commit directly to the `master` branch.
- Create a new branch for this commit and start a pull request. Learn more about pull requests.

## 6. Create Pull Request

The screenshot shows a GitHub repository page for 'mzmine / mzmine\_documentation'. The 'Pull requests' tab is selected. A yellow banner at the top indicates that 'SteffenHeu-patch-1 had recent pushes 1 minute ago'. A green button labeled 'Compare & pull request' is visible.

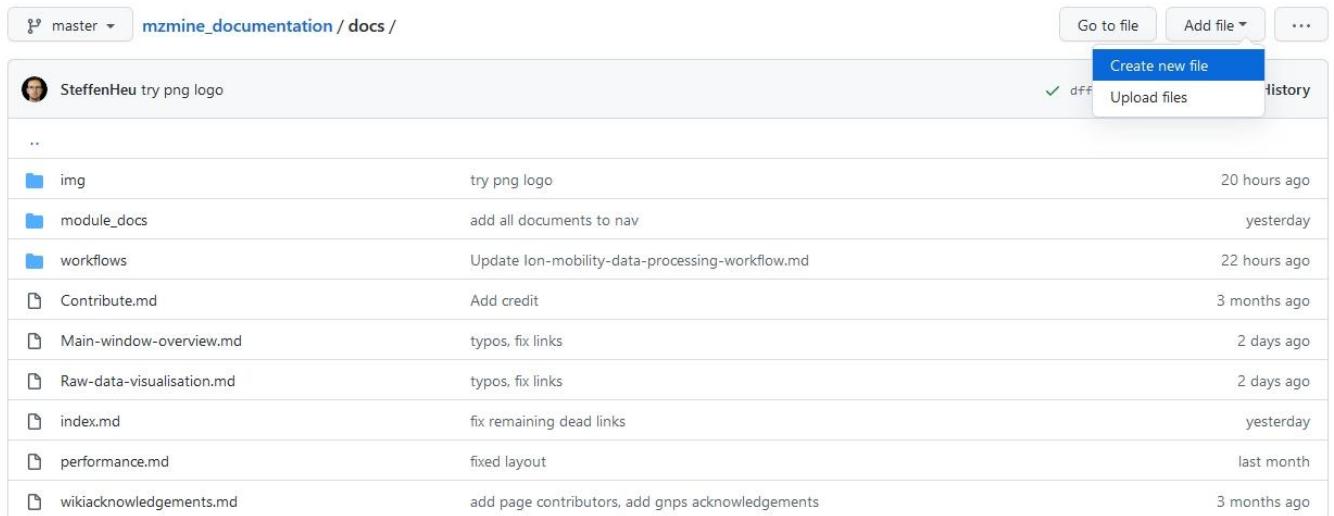
## 7. Finalize Pull Request with Description

The screenshot shows the GitHub pull request creation interface. The base branch is set to 'master'. The compare branch is 'SteffenHeu-patch-1'. A green checkmark indicates that the branches are 'Able to merge'. The pull request description includes the text 'update mobility resolving step' and 'add msms pairing description in mobility resolving step'. There is a note at the bottom encouraging users to follow GitHub Community Guidelines.

## 10.2 Creating a new page

**Follow steps 1 - 3.**

**Navigate to mzmine\_documentation/docs in your fork and create a new file**



mzmine_documentation / docs /		
SteffenHeu try png logo		...
 img	try png logo	20 hours ago
 module_docs	add all documents to nav	yesterday
 workflows	Update Ion-mobility-data-processing-workflow.md	22 hours ago
 Contribute.md	Add credit	3 months ago
 Main-window-overview.md	typos, fix links	2 days ago
 Raw-data-visualisation.md	typos, fix links	2 days ago
 index.md	fix remaining dead links	yesterday
 performance.md	fixed layout	last month
 wikiacknowledgements.md	add page contributors, add gnps acknowledgements	3 months ago

**Follow steps 4 - 7.**

## 10.3 Page Contributors

[SteffenHeu](#)

This page was adapted from the [GNPS documentation](#).

---

Last update: April 5, 2022 13:22:07

## 11. Acknowledgements

---

We would like to point out that this wiki was set up in tight collaboration with the [GNPS](#) staff. We highly appreciate your help!

### 11.1 Related projects

---

- [GNPS](#)
- [SIRIUS](#)

### 11.2 Libraries we use in MZmine

---

- [Apache XML Graphics](#) - EPS image export
- [Chemistry Development Kit](#) - Isotope pattern and molecular calculations
- [Freehep](#) - EMF image export
- [Google Guava](#) - Utility classes
- [JDK Documentation](#)
- [JChemPaint](#) - 2D molecule visualization
- [JFreeChart](#) - TIC, Spectra and 2D visualizers
- [J Mol](#) - 3D molecule visualization
- [jmzml](#) - mzML file import
- [jmzTab](#) - mzTab file import and export
- [NetCDF-Java](#) - NetCDF file import
- [VisAD](#) - 3D visualizer
- [WEKA](#) - Clustering and other machine learning algorithms
- [Bruker TDF SDK](#) - Native tdf/tdf file import (requires VC++ 2017 redist.)
- [Thermo raw file parser](#) - Native Thermo raw import

---

Last update: March 10, 2022 15:35:04