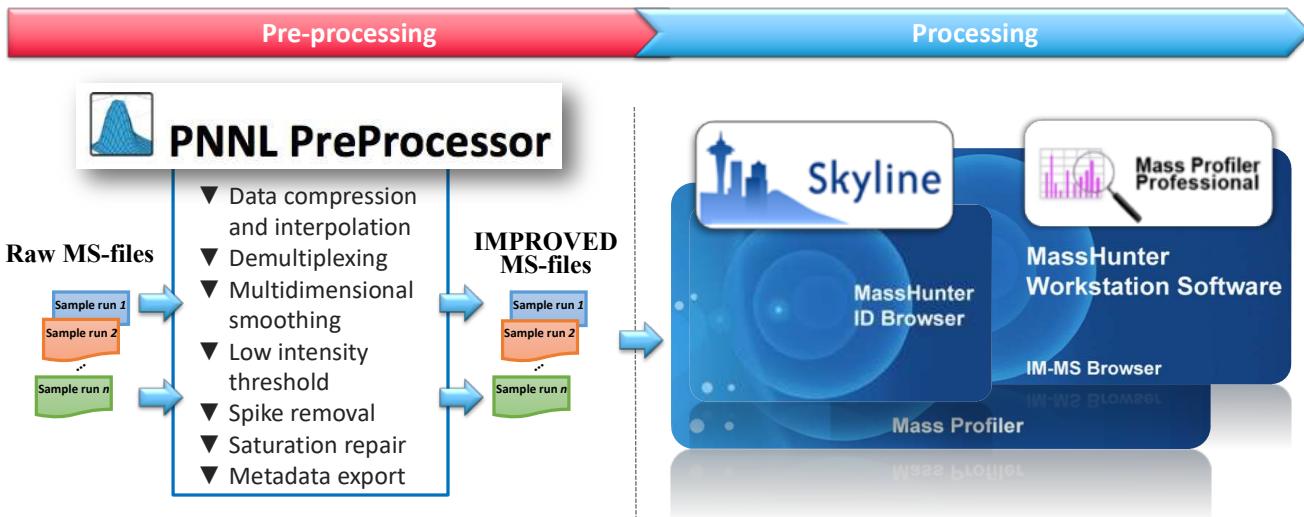


PNNL PreProcessor | Ion mobility MS

USER GUIDE – version 6.0 (2025.10.24)

In close collaboration between Pacific Northwest National Laboratory and Agilent Technologies, we have developed this user-friendly tool for Agilent MassHunter (.d) and UIMF mass spectrometry data files (MS-files) from drift tube (DT) and structure for lossless ion manipulations (SLIM) IM-MS platforms, to generate new MS-files in the same instrument data format with enhanced signal quality.



Available algorithms and utilities in the PreProcessor include: data compression and interpolation, ion mobility demultiplexing, multidimensional smoothing, noise filtering by low intensity threshold and spike removal, saturation repair and metadata export.

If you use the PreProcessor, please cite: Bilbao et al. *A Preprocessing Tool for Enhanced Ion Mobility-Mass Spectrometry-Based Omics Workflows*. Journal of Proteome Research 2021 <https://doi.org/10.1021/acs.jproteome.1c00425>.

Contacts:

- aivett.bilbao@pnnl.gov
- sarah.stow@agilent.com

Do you have a problem to report? Please let us know any feedback!



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965



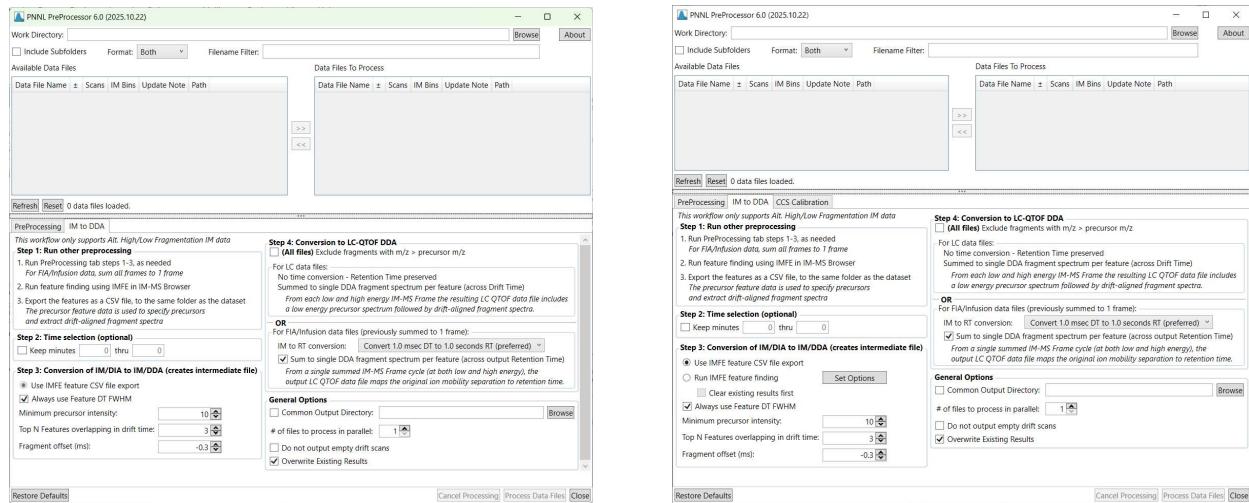
Agilent Technologies

1. What are the benefits?

- Remove artifacts in jagged peaks (i.e. low ion statistics) and enhance low-abundance real signals.
- Improve signal-to-noise and reproducibility with more features consistently detected across your multiple runs and reduce variations in abundance and collision-cross section (CCS).
- Reduce the file size and time for downstream processing.
- Recover the intensity, isotopic ratios and mass accuracy from saturated high-abundance analytes of different molecular ions (i.e., peptides, metabolites, lipids, glycans, and inorganic polymers).
- Computationally extend the dynamic range of your measurements and enable faster and memory-efficient downstream processing!

2. Requirements

- A computer running Windows 7 (64-bit) or later with at least 8GB of RAM (32GB recommended).
- The .NET Framework 4.7.2 (or later). Download available at <https://dotnet.microsoft.com/download/dotnet-framework/thank-you/net472-web-installer>.
- MS-files in UIMF or Agilent MassHunter (.d) format.
- New IM-MS Browser dlls are required for certain features to be visible in the UI (CCS Calibration and Run IMFE feature finding called from within the IM-to-DDA workflow). These dlls support IMFE from the command line and can be obtained by contacting Agilent (sarah.stow@agilent.com). If the dlls are not copied into the Bin folder of IM-MS Browser you will see the UI on the left below (note no CCS Calibration tab and no option to Run IMFE feature finding). If the dlls are copied into the Bin folder then you will see the UI on the right side below.



3. Software features

- Command-line and graphical (workflow style) user interfaces.
- Single-click batch processing of multiple raw MS-files.
- Data compression (by frame and mobility) and filtering by retention time range.
- Data interpolation of the ion mobility dimension to improve the results of the [HRdm](#) demultiplexing and peak deconvolution strategy.
- Multidimensional smoothing of data and repair of saturated peaks.
- [PNNL demultiplexing and artifact removal algorithm](#) integrated. A new selectable pulse coverage percentage to increase sensitivity for low level signals.
- An algorithm to remove noise in form of ‘spikes’.
- Ion mobility MS with/without any separation: LC-IM-MS, solid phase extraction (e.g., RapidFire) IM-MS and direct infusion IM-MS.
- All Ions MS-files (data-independent acquisition) with alternating high/low collision energy fragmentation.
- Exporting metadata information of ion mobility frames (e.g., field, pressure, temperature) and MS actuals to text files.
- Multidimensional smoothing for non-ion mobility TOF-MS data files (e.g., produced by 6530, 6540, 6545 and 6550 Agilent instruments). Conversion of arrival time vs. m/z to CCS in the raw data for SLIM and DT
- Conversion of arrival time to retention time: single frame “IMMS” to “LCMS” format
- IM to DDA conversion of All Ions IM/MS mobility aligned fragmentation data into QTOF Auto MS/MS data file format. Supports both maintaining LC separations as well as single frame “IMMS” to “LCMS” format (new ‘IM to DDA’ tab).
- Command line options

4. Latest features

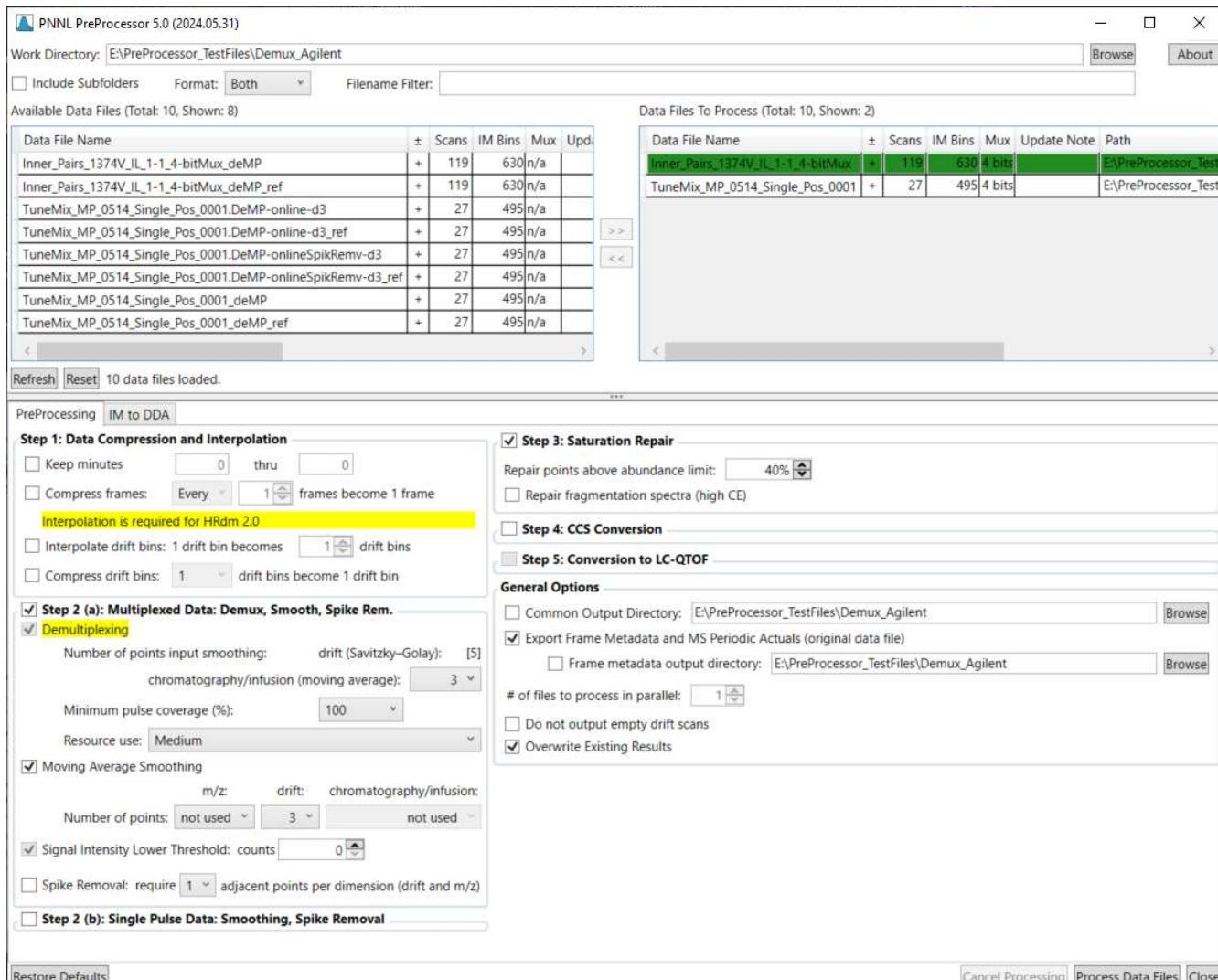
- Polygon extraction is available for batch processing in Step 4a. This requires a .m method file from IM-MS Browser where the polygon has been defined. It can be used alone to generate either an LC-IM-MS data file or in combination with Step 5 to generate a LC-MS data file.
- The IM to DDA conversion of All Ions IM/MS mobility aligned fragmentation data can now call IMFE from within the PNNL PreProcessor to generate a precursor feature list.
- CCS Calibration which is currently only supported in IM-MS Browser is now available in the PNNL PreProcessor. Tune mix data files are processed to generate the calibration and then it is saved to specified data files. Information on the calibration is included in the log files.



5. Usage

Installation is not required (users having issues with anti-virus software can alternatively use the installer). Download and unzip the file in your computer.

- Double click the ‘PNNL-PreProcessor.exe’ file to start it.
- Click the text box to search and select a single MS-file (.d folder) or a directory which contains several MS-files (.d folders).
- Select files and click ‘>>’ to move to the right the files to be processed.
- Adjust the parameters and click the ‘Process Data Files’ button.



For each MS-file, the selected algorithms will be sequentially applied in the same order as they appear in the GUI from top to bottom. A progress bar will indicate the processing status and a new MS-file will be generated within the same folder (or selected Output directory if checked).



Suffix in the MS-file names indicates the preprocessing algorithms applied, for example:

- **_1.00-3.00.d** time range (minutes) kept.
 - **_FC3.d** frame compression with every 3 frames summed into 1.
 - **_DC3.d** drift bin (TOF transient or MS scan) compression with every 3 drift bins summed into 1.
 - **_DI3.d** drift bin interpolation with addition of 3 interpolated drift bins per every original drift bin.
 - **_MA-d3-c3.d** moving average smoothing with 3 points in drift and 3 points in chromatography.
 - **_Min20.d** signal intensity lower threshold of 20.
 - **_Spk.d** spike removal.
 - **_SR.d** saturation repair.
 - **_MA-d3-c3-Min20-Spk_SR.d** moving average smoothing with 3 points in drift and 3 points in chromatography, signal intensity lower threshold of 20, spike removal and saturation repair.
 - **DeMP.d** ion mobility demultiplexed combined or not with other algorithms. Only the final result file will be preserved and intermediate result files will be automatically deleted. To avoid file overwriting or loosing results, manually rename result files (after preprocessing is complete) to compare multiple combinations of preprocessing parameters with demultiplexing. Check the log file inside each result to verify which preprocessing algorithms were applied.
 - **.CCSz.d** drift time (or arrival time) vs. m/z space is mapped to CCS/z vs m/z space based on a CCS calibration file saved to the data file.
 - **.3D.d** the IM dimension is mapped to the LC dimension of the data file
 - **.DDA_3D.d** All Ions IM/MS mobility aligned fragment spectra are written as Auto MS/MS DDA events in an LC-MS data file
 - **_PEx.d** polygon extraction kept in IM format
 - **_PEx_3D.d** polygon extraction and conversion to LC-MS data

PreProcessor workflows with specific algorithms and parameters for different experimental approaches. Empty cells indicate that the module is not used. The output MS-files generated are named as the corresponding input MS-file plus a suffix that indicates the modules applied (due to compatibility with the HRdm software, the suffix “DeMP” is used for the demultiplexed MS-files regardless of additional post-demultiplexing modules applied).

			EXAMPLES OF EXPERIMENTS AND PARAMETERS				
ALGORITHM	SLIM IM-MS	SLIM IM-MS	LC-DT IM-MS multiplexed	LC-DT IM-MS multiplexed + HRdm	LC-DT IM-MS default	LC-DT IM-MS, saturated ions	DT IM-MS (direct infusion)
Data compression	15 drift bins become 1	All frames become 1					All frames become 1
Data interpolation				1 drift bin becomes 5			
Ion mobility demultiplexing			5 points LC	5 points LC			
Multidimensional smoothing			3 points IM	3 points IM	3 points LC, 3 points IM	3 points LC, 3 points IM	3 points IM
Low intensity threshold	1 count		20 counts	20 counts	20 counts	20 counts	50 counts
Spike removal					1 adjacent point per dimension		
Saturation repair						40% (MS1 only)	
CCS conversion		Selected					
Conversion to 'LCMS'		Selected					
<i>Suffix in output MS-file name</i>	DC15-Min1	FCsum CCSz 3D	_DeMP	_DI5.d.DeMP	MA-c3-d3-Min20-Spk	MA-c3-d3-min20-Spk-SR	FCsum-d3-min50

CAUTION with temporary files: temporary files with suffix such as ‘_temp1.d’ will be generated with intermediate results. Do not delete the temp files while any preprocessing is in progress and verify that all temp files were deleted after all preprocessing is finished. For example, an error has occurred if saturation repair was performed and only the file ‘SR_temp1.d’ was generated, please report it.

6. Data Compression and Interpolation

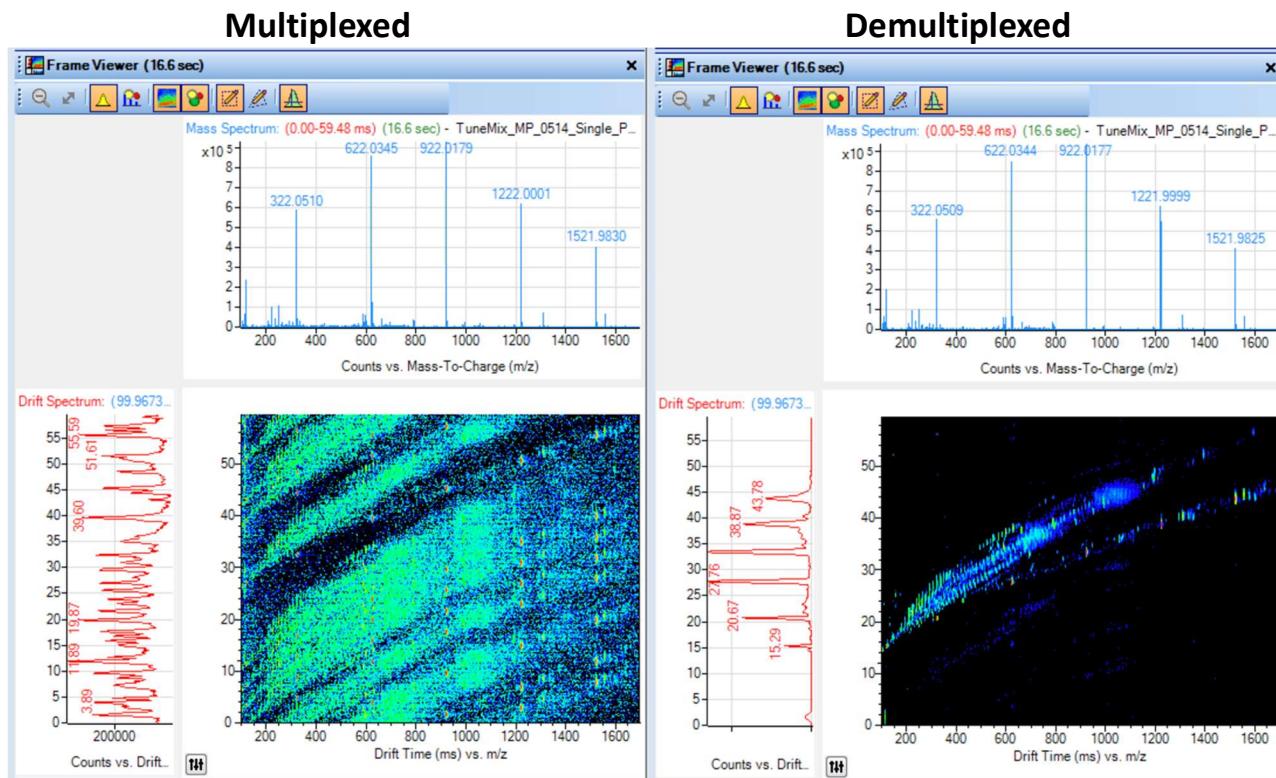
New to version 3 are options to select and alter the sampling rate of the data that will be treated.

- It is now possible to select a time range from the original input MS-file for treatment. The final output MS-file will only contain data from the frames within the time range selected.
- When configuring the acquisition of IM data, the user selects either the time interval or the specific number of frames acquired and summed before being written to disk. For long separations which may have wide chromatographic peak widths this number could be low resulting in many points acquired per LC peak. By using the data compression post-acquisition, additional in-frame sensitivity can be realized by further combining (i.e., summing or “compressing”) adjacent frames. For data that has been acquired as part of an infusion experiment, options are also available to sum all frames into one.
- To increase the effective sampling rate of the ion mobility separation the user can now add additional TOF Transients by means of drift bin interpolation. Specifying 1 bin to become 3 (a good default) results in an effective ion mobility sampling period of 40 μ sec as opposed to 120 μ sec without interpolation when operating in the 1700 m/z range mode. This operation should always be used for multiplexed data that will be subsequently processed with HRdm 2.0. It can also be used for single pulse data.
- To better support SLIM data it is now possible to effectively decrease the TOF sampling rate by combining (i.e., summing or “compressing”) the data in adjacent TOF transients. This is very helpful because SLIM separations occur over a much longer time scale (approximately 1 second) resulting in TOF transient oversampling. Typical drift bin compression values range from 3 to 10.



7. Demultiplexing

The previously developed [PNNL demultiplexing and artifact removal algorithm](#) was integrated into the PNNL PreProcessor. Multiplexing is a data acquisition technique applied during ion mobility separation which can significantly improve the duty cycle and signal-to-noise as well as extend the dynamic range.



For ion mobility multiplexing, the drift time dimension is encoded with a pseudo random sequence by allowing multiple ion packets to be concurrently resident in the drift tube. As the multiple packets traverse the drift cell, fast moving ions from one packet overlap with slow moving ions from a previous packet and the information for all packets is recorded by the detector in one MS-file. The multiplexed raw data is demultiplexed post-acquisition using a Hadamard transform-based decoding algorithm which is based on the gating of the ions and the pseudo random sequence used for multiplexing during acquisition and a second algorithm developed at PNNL to detect and remove artifacts.

Version 3.0 provides the following new capabilities:

- The user can now select the Minimum pulse coverage (%). Following demultiplexing the software performs a noise filtering or artifact removal by applying a reverse confirmation for all output results to check that corresponding signal is present. This is one important contributor to why the PNNL PreProcessor produces results with a very low level of artifacts, compared to other demultiplexing methods. Reducing this value from 100% allows preserving signals with relatively low ion statistics while still providing very low artifact presence.



- Three options for resource use are available for demultiplexing workflows. As demultiplexing is a computationally intensive operation, the user can now select either low, medium or high. The selection is not absolute, but rather relative to the computer's number of cores.

CAUTION: The demultiplexing algorithm has been optimized to use multiple threads and supports all post-demultiplexing operations selected, except chromatographic smoothing which occurs as a single-thread operation in a second step. To avoid very slow processing *do not use the parameter 'Moving algorithm smoothing: chromatography/infusion'*.

8. Multidimensional smoothing

Raw intensity values in each frame (an IM separation) are smoothed first in the drift dimension followed by smoothing of the chromatographic dimension considering neighboring frames. The parameters are described in the following subsection.

8.1. Number of points

The number of points used to compute the average or new smoothed value should be specified for each dimension. For instance, 5 points for drift mean that each new point is computed as the average considering also the 2 drift spectra before and the 2 drift spectra after (same m/z and within the same frame).

- As a general rule, the number of points should be less than the number of sampling points covering the peak width at half height. Smoothing with a large number of points cause a detrimental loss in resolution (see section below on Over Smoothing).
- Smoothing can be applied to all m/z , drift and chromatographic dimensions or any subset combination.
- Smoothing across the m/z dimension is typically not necessary for Agilent IM-MS systems.
- Larger chromatographic number of points requires more memory and processing time.

8.2. Over Smoothing

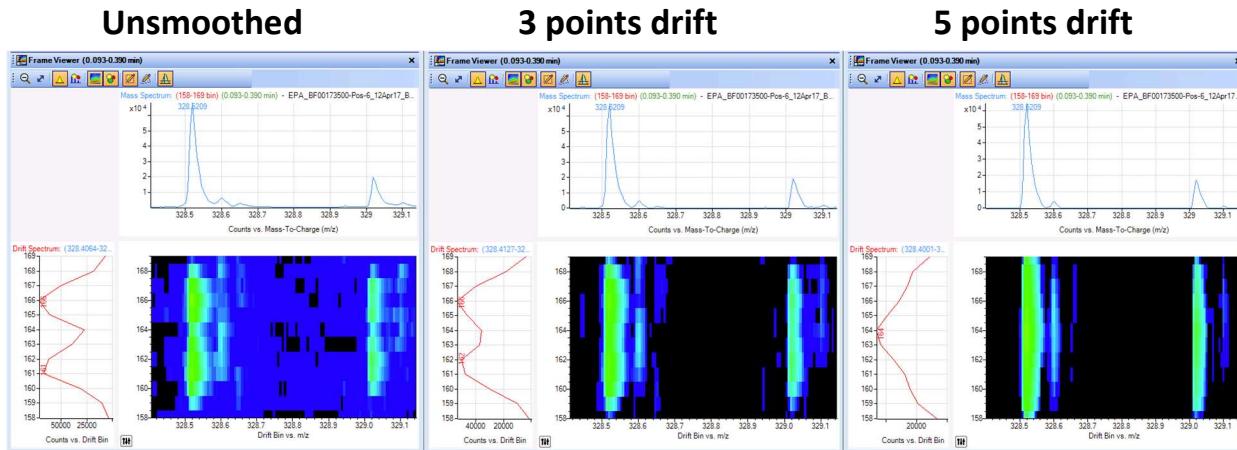
When first using the software or processing data from a new sample type, you should check your raw data to inspect the results. One of the possible problems that you may run into is over smoothing or blurring two adjacent peaks into one. This happens when smoothing is applied with a number of points that is too large (i.e. smooth too many points, relative to the width of a peak).

To avoid over smoothing, please do the following: check one of your MS-files and see how many points you have across the early drift peaks that you are interested in, the low m/z range has usually narrower drift peaks and earlier drift times.

In the IM-MS Browser, select: Configuration -> Set Drift Spectrum units to Drift Bins.



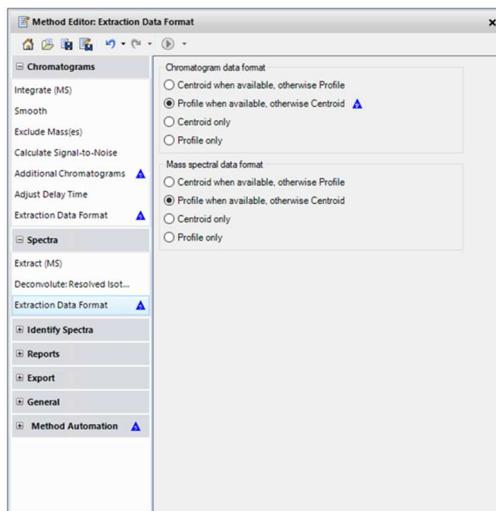
In the following example, the MS-file in the left is the original or unsmoothed one, those drift peaks have less than ~6 points per peak (the first one from 158-164 Drift Bin), those are likely isomers. To preserve them, use only 3 points for smoothing (as in the middle MS-file), since using 5 or 7 points will merge those drift peaks (as in the right MS-file).



8.3. Smoothing for non-ion mobility TOF-MS data files and using them in QUIL software
 Smoothing is also available for MS-files produced by 6530, 6540, 6545 and 6550 Agilent TOF instruments. However, the PNNL PreProcessor does not generate centroid peak data. *Use the Agilent 'IM-MS Reprocessor' to recentroid the result file prior to opening it in Agilent QUIL software.*

Alternatively, change the settings in QUIL to show the raw data instead of the centroid peak data:

- Go to ‘Method –> Method Editor’...
- Select ‘Spectra –> Extraction Data Format’
- Select ‘Chromatogram data format –> Profile when available, otherwise Centroid’
- Select ‘Mass spectral data format –> Profile when available, otherwise Centroid’



9. Signal intensity lower threshold

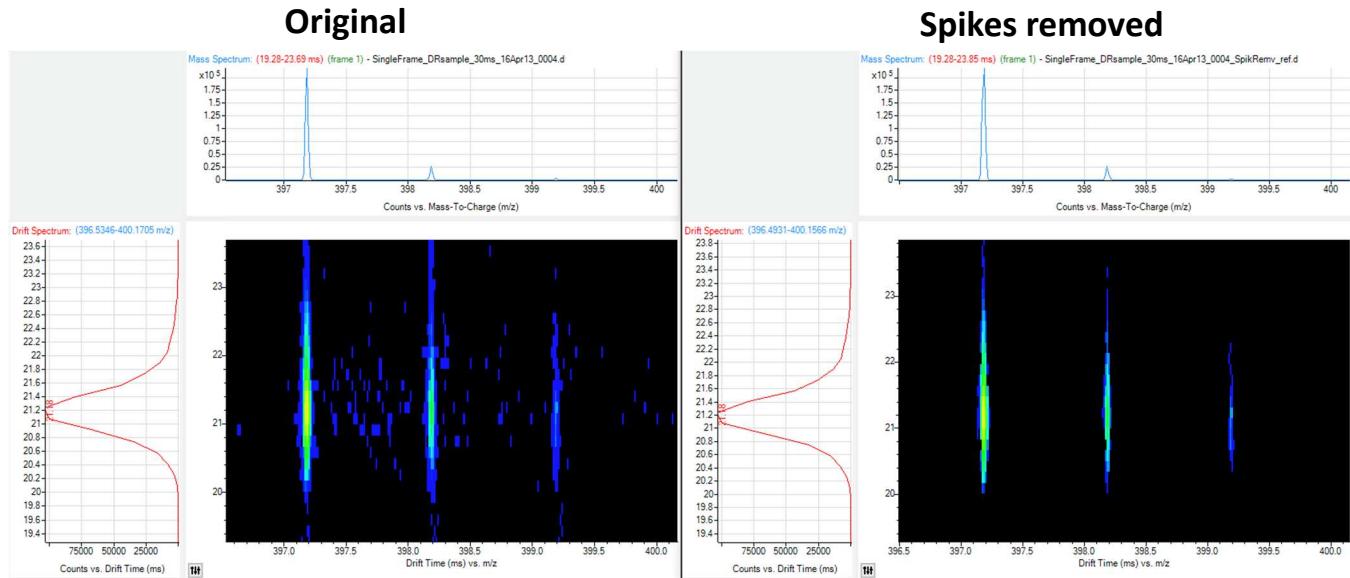
After applying the smoothing, intensity values that do not exceed the threshold will be removed.

CAUTION: a threshold higher than the recommended default value (20) in the default settings may cause non-detection or decreased quality of features. However, a higher value could be useful to reduce required memory and increase processing speed for very complex MS-files.

10. Spike removal

We refer to as spikes to the noisy peaks that have inconsistent and very few raw data points. Since the spikes may have intensity values above the lower intensity threshold, we detect them by comparing each point against the neighboring or adjacent points in drift and m/z . Each point must have either the 1 or 2 user-specified neighbors in each dimension in order to pass the filter. Points that do not pass the filter will be removed.

Note: spike removal is *not recommended* for All Ions MS-files because some relevant low-mass and low-abundance fragment ions could be removed from the fragmentation spectra. Similarly, spike removal is *not recommended* for studies where low-abundance/low concentration ions are relevant.



11. Saturation repair

Complete raw MS-files from untargeted LC-IM-MS analyses are automatically preprocessed to find ions exceeding a defined intensity level (a percentage of the abundance limit given the MS instrumentation).

The undistorted isotopic distribution is recovered from spectra at unsaturated time from the peak tail (intensity below the selected saturation level). In contrast to our [previous algorithm](#), this new method does not require theoretical isotopic distributions, works at the raw or profile level (i.e., non-centroid spectra) and produces MS-files in instrument format (instead of text tables) as output.

The most intense unsaturated isotopic peak is utilized as reference to successively reconstruct the time profile of the saturated ones.

11.1. IMPORTANT NOTES FOR CURRENT VERSION

- Chromatographic separation required: saturated analytes must have a well-defined chromatographic peak (even if several isotopic peaks have a ‘flat top’, peaks should have a clear start and end time points and at least one isotopic peak below the saturated threshold defined).
- Interferences: the software ignores saturated ions if interferences are detected.
- Ion charge states from 1-5 supported.
- All Ions MS-files supported (high fragmentation energy frames are ignored by default).

11.2. Output MS-files

The tool works in a serial way, two MS-files are generated if both smoothing and saturation repair options are selected: the first MS-file is only smoothed, and the second MS-file is both smoothed and saturation repaired. For example, with default settings, the suffix in the MS-file names indicates the preprocessing applied:

- `_MA-d3-c3.d`: smoothed only
- `_MA-d3-c3_SR.d`: both smoothed and saturation repaired

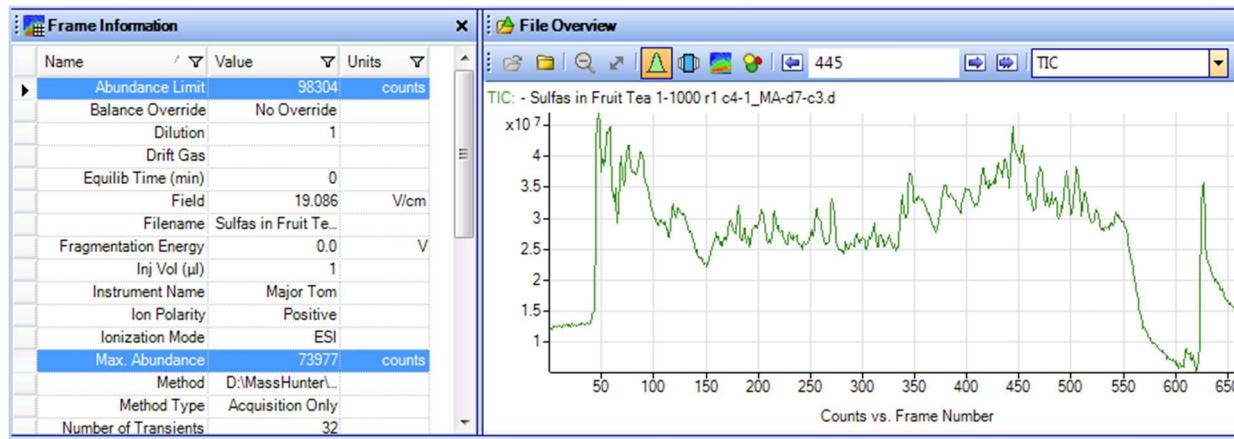
Why both ‘_MA-d3-c3.d’ and ‘_MA-d3-c3_SR.d’ MS-files are kept? Because the saturation repair results may not be comparable across multiple runs, it is recommended to verify the results and delete the unused files after. For example, when the sample is very complex and there are interferences for a saturated peak in some runs but not in the others, the peak will not be repaired across all runs and therefore quantitation results will not be comparable across runs. Unused files can be manually deleted by filtering them either typing ‘SR.d’ or similar in the Explorer windows.

11.3. Abundance limit

The percent saturation limit default is 40% and is the approximate onset of non-linearity, peak broadening and an increasing mass error due to saturation of the detector. The following is an example of a threshold at 70% and shows where the associated Abundance Limit can be found. It should be noted that saturation effects worsen as a peak’s percentage abundance increases and that 100% is a theoretical and not realized maximum.



You can see the Abundance values in your MS-file using the IM-MS Browser by selecting 'View -> Frame Information':



The Abundance Limit is 98304 (given by the detection system and acquisition settings) and the Max. Abundance of the detected ions (for the actual signals recorded in that frame) is 73977.

Therefore, the software will attempt to repair all ions in that frame that have raw intensity between 68812 (70% of the Abundance Limit 98304) and 73977.

The software reads the actual value `FrameSpecAbundLimit` from the `IMSFrameMeth.xml` file in the AcqData subfolder.

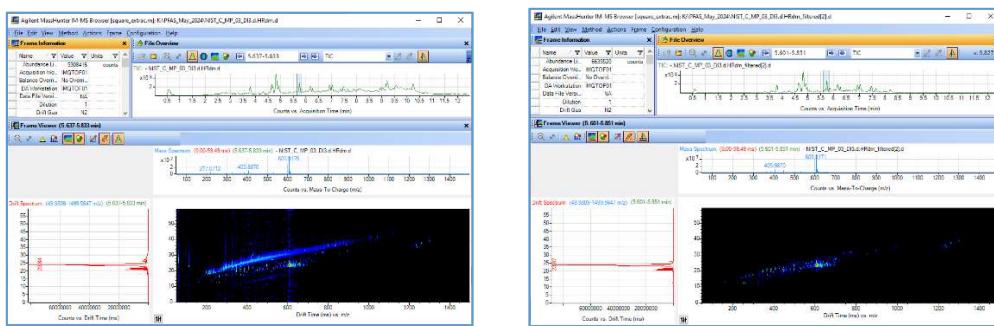
```
<!--1=Raw; 2=HiLow; 4=DualGain; 8=TLPP; etc-->
<DetectorGainMode>4</DetectorGainMode>
<DetectorGainRatio>12</DetectorGainRatio>
<!--Volts-->
<FragEnergy>0</FragEnergy>
<!--1=Fixed; 2=MultiSegment-->
<FragEnergyMode>1</FragEnergyMode>
<!--1=None; 2>Selective; 4=NonSelective; 8=HiLoFra...
<FragOpMode>1</FragOpMode>
<!--milliseconds-->
<FrameDtPeriod>0.120224</FrameDtPeriod>
<!--nanoseconds-->
<FrameMsXPeriod>0.5</FrameMsXPeriod>
<FrameSpecAbundLimit>98304</FrameSpecAbundLimit>
<FrameSpecFmtId>1</FrameSpecFmtId>
```



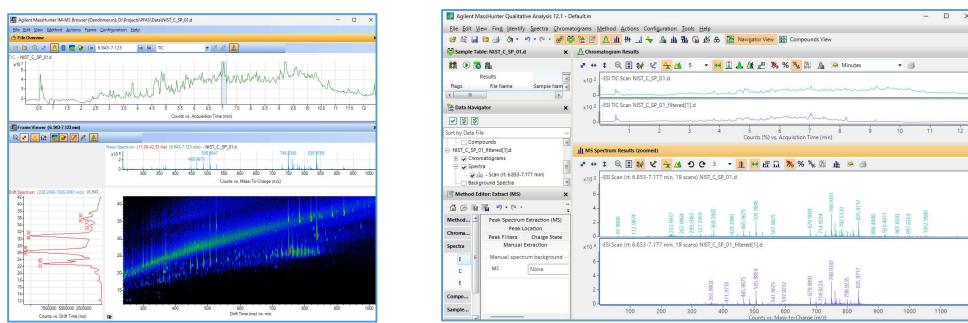
12. Polygon Extraction

The polygon extraction feature in the PNNL PreProcessor is an extension of the same functionality present in IM-MS Browser with the ability to support batch processing. The polygon region (m/z vs. drift time) is defined in IM-MS Browser and saved as a method file (*.m) (Draw polygon > Right Click > Use Selection as Method Bookmark > Method > Save). This file is then loaded into the PNNL PreProcessor to indicate the region of the data file to be present in the resulting data file.

There are a few different use cases for the polygon extraction. First, to save processing time the polygon region can be extracted from an LC-IM-MS data file and an LC-IM-MS data file will be generated with only the polygon region of interest in each LC frame. For example in the images below, PFAS ions which are present in a complex matrix generally elute in a region of IM space below the background matrix signal. By extracting the polygon region of PFAS ions, downstream feature finding algorithms will not waste time mining the background signal that is not of interest.



A second use case is to extract the region of interest from an LC-IM-MS data file and export an LC-MS data file without the IM dimension (using both Steps 4a and 5 in the PNNL PreProcessor). In this example, traditional LC-MS data analysis workflows can be used, and IM is simply used to remove co-eluting signal which cleans up the mass spectra in the resulting LC-MS data file.



We do not recommend using this with All Ions fragmentation data as it could remove important precursor or fragment ions which could leave to false positive or negative identification in downstream analysis. Additionally, this has been added as Step 4a because it cannot be used in combination with CCS conversion in Step 4b.



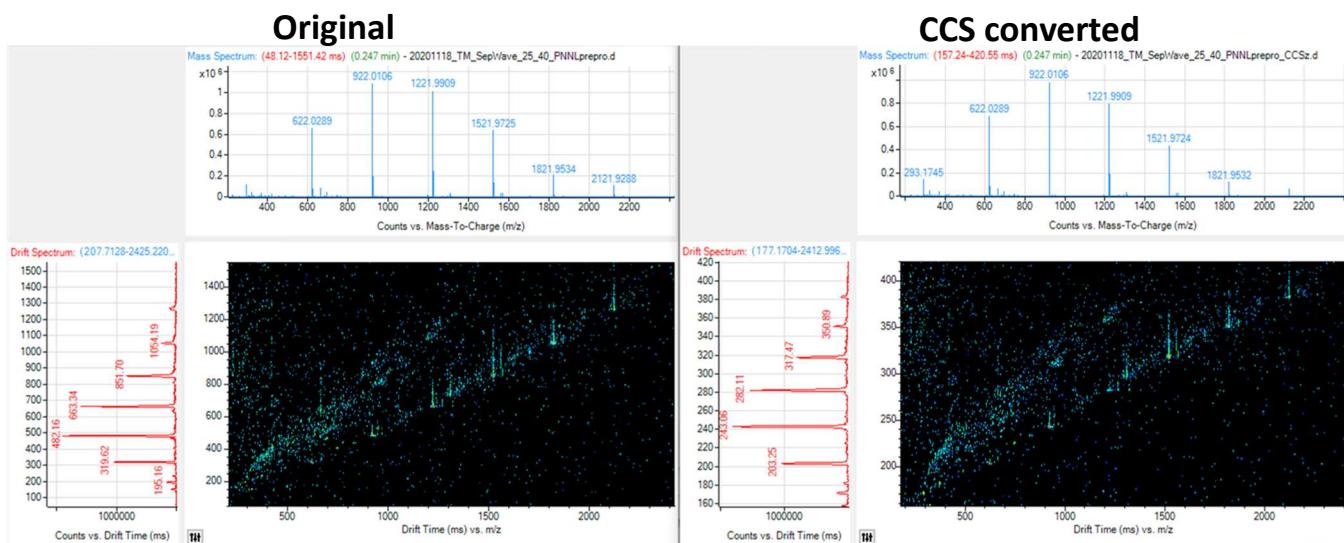
13. CCS Conversion

Conversion of arrival time to CCS and generation of new MS-files with raw data where the mobility dimension is fully converted into CCS space. The calibration coefficients for the function, linear for DT and 3rd order polynomial for SLIM, are automatically detected from the raw MS-file (found in file OverridelmsCal.xml or SLIMlmsCal.xml). The calibration must be performed using the Agilent IM-MS Browser prior conversion to CCS.

Using this functionality, MS-files that were acquired with different IM method parameters can be analyzed together and, rather than CCS values from centroids, the full CCS distributions in the raw data can be visualized directly and easily compared.

If the data does not contain LC separation, a transformation can be applied to convert the IM dimension to an LC dimension and generate a non-IM file (standard QTOF .d or 3D, see next section) which can be processed by existing LC-MS analysis tools that do not support the IM dimension.

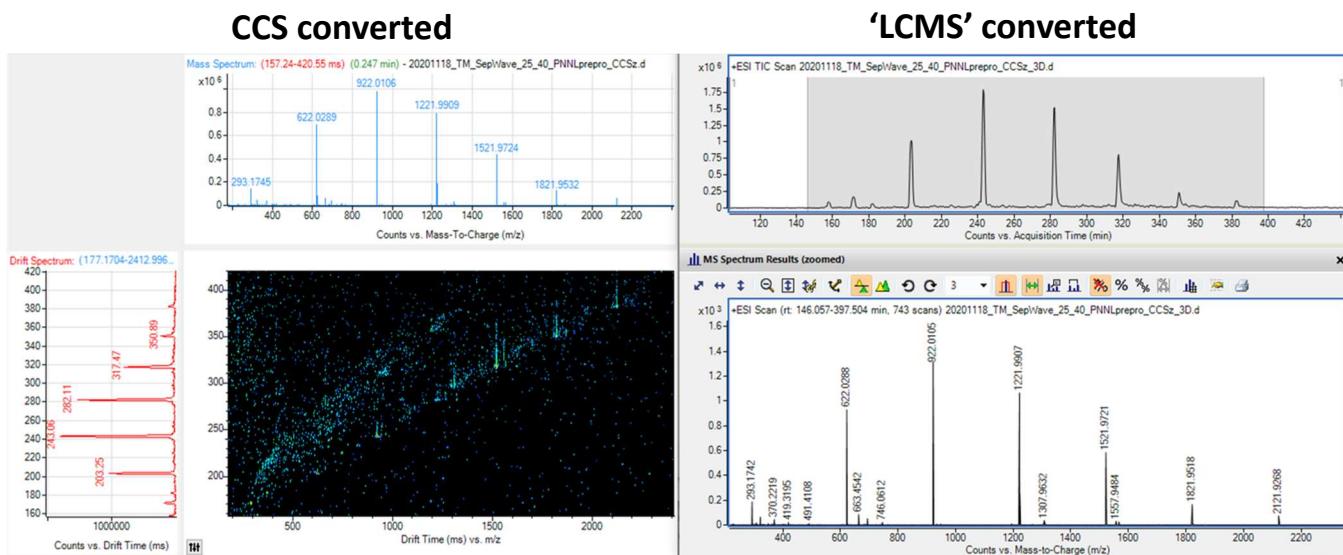
The figure below shows an example for the Agilent tune mix ions. In the CCS converted file on the right, the Drift Time axis (red trace) represents the ion's CCS directly in the raw data.



14. Conversion to 3D/LCMS'

Conversion of arrival time to retention time: single frame “IMMS” data to “LCMS” format. The time axis with arrival times (or converted CCS) are parsed to retention time (RT) in a new MS-file in the MassHunter QTOF/3D format which can be processed by existing LC-MS software tools (e.g., Agilent Qualitative Analysis or Skyline). In the case of CCS converted MS-files, 1 sq Angstrom is equivalent to 1 min RT in the LCMS data file. For unconverted CCS MS-files, several options are available to scale the units from milliseconds in arrival time to seconds or minutes in RT.

The figure below shows an example for the Agilent tune mix ions. After conversion to ‘LCMS’, the MS-file with suffix “_3D” can be visualized using the Agilent Qualitative Analysis software (IM-MS Browser will not work for this format). While the m/z dimension remains the same, the Drift Time axis (red trace on the left) becomes the Acquisition Time or RT axis on the right (top trace).



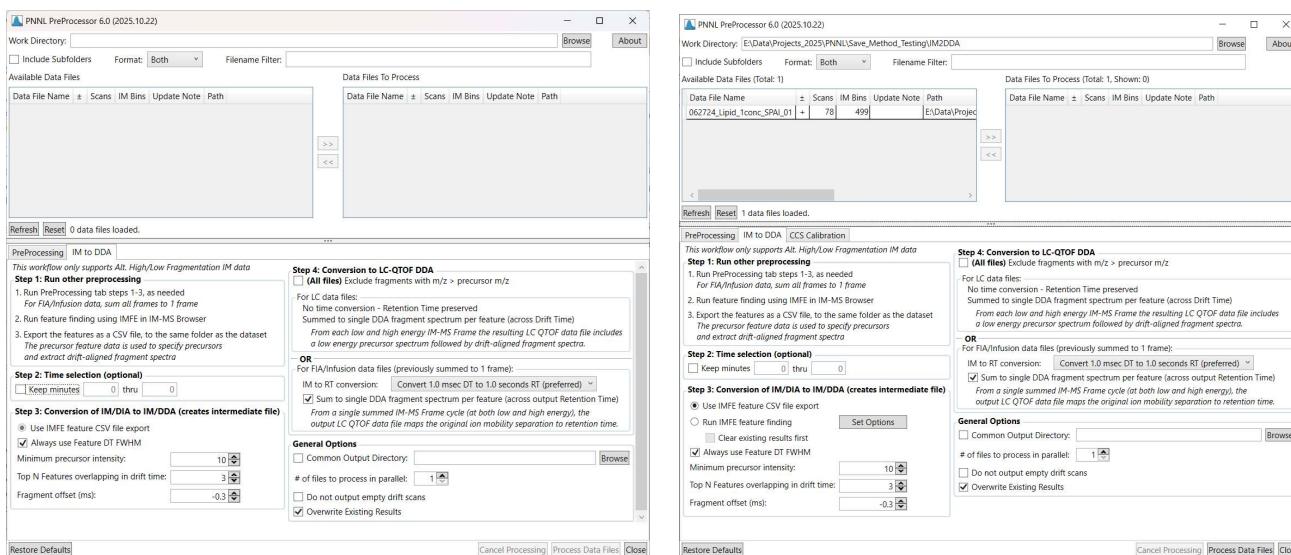
Note: If the MS-file contains fragmentation spectra it will be ignored and only the MS1 will be included in the new MS-file.



15. IM to DDA Conversion

The ability to convert All Ions IM/MS mobility aligned fragmentation data to LC-MS Auto MS/MS DDA format is now available on a new tab in the PNNL PreProcessor. Before running the steps in this tab, consider the following requirements:

1. Any previous preprocessing steps on the first tab (i.e. demultiplexing, smoothing, frame and drift bin compression) must be run as a prior separated step (i.e., the input of this tab would be the output data file of the “PreProcessing” tab).
2. A feature list is needed to indicate precursor ions for the IM to DDA conversion. This list can either be generated in IM-MS Browser and exported as a .csv file prior to running the IM to DDA conversion or IMFE can be called from within the PNNL PreProcessor via command line to run IMFE as part of the conversion. The image on the left below shows the UI when the dlls to run IMFE via command line are missing from the IM-MS Browser Bin folder and the image on the right shows the UI when the dlls are present. Note that if the feature list is generated in IM-MS Browser it can be manually edited or generated using an IM-MS Browser feature list as a template. Only the m/z, DT, and RT have to be accurate for the conversion. Both the Z and CCS should be accurate if you want them to appear that way in the scan header of the new DDA file. The other values such as abundance, height, Q-score, and RT start and stop can all be best estimates as the algorithm does not use this info in the conversion. Each feature you add should have a new feature ID number.



While abundant text guidance is provided on the ‘IM to DDA’ tab, default settings do not need to be changed for most workflows. Once the conversion is started, the PreProcessor uses the feature list as a set of precursor ions and creates MS/MS spectra based on the mobility aligned fragment ions for each



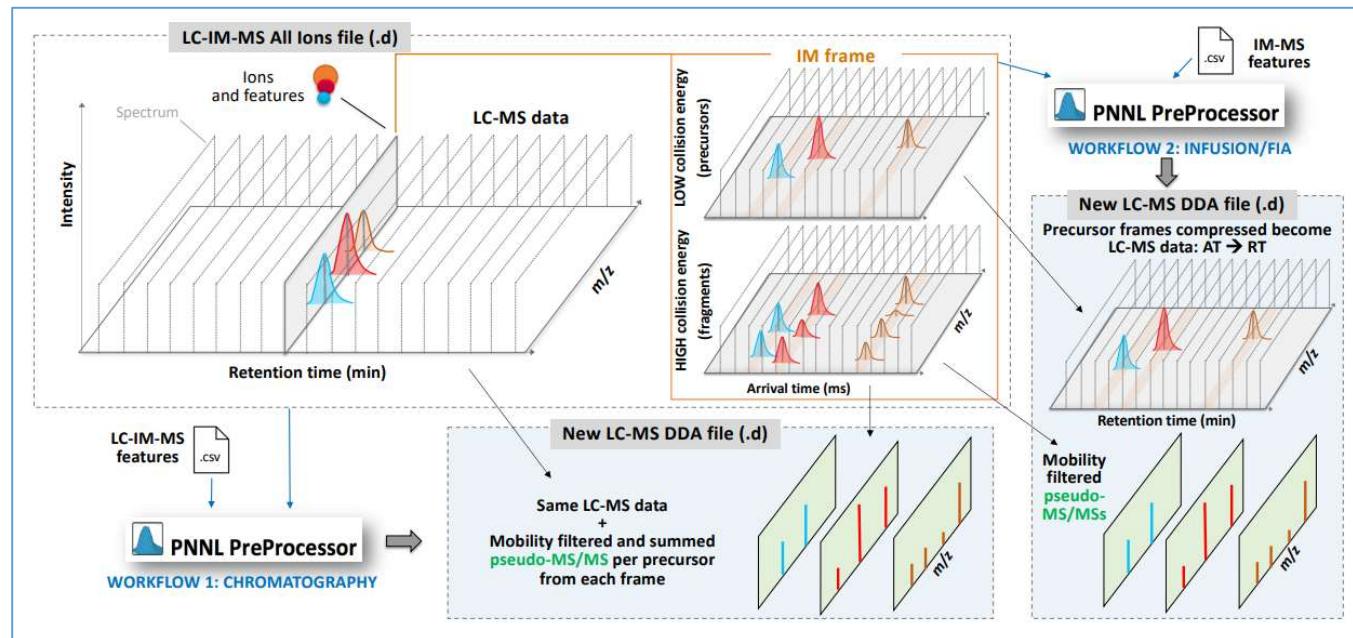
precursor. This is then stored as a new output file in the format of a LC-MS Auto MS/MS data file. Two data files are created in this conversion and are both appended with DDA at the end of the file name:

...DDA.d – This file is still in 4D IM format. Current IM data analysis software cannot process this data file. It is a placeholder for future investigations.

...DDA_3D.d – New 3D LC-MS DDA data file that can be processed in current DDA data analysis software

The data file with the additional 3D is the new LC-MS data file to use in downstream workflows.

The image below is a graphical representation of the two workflows that are mentioned in Step 4.



Workflow 1: If the LC dimension in the original LC-IM-MS file is to be maintained then the DDA fragment spectrum are summed per feature across drift time. No time conversion is needed as original LC time is written in the output data file.

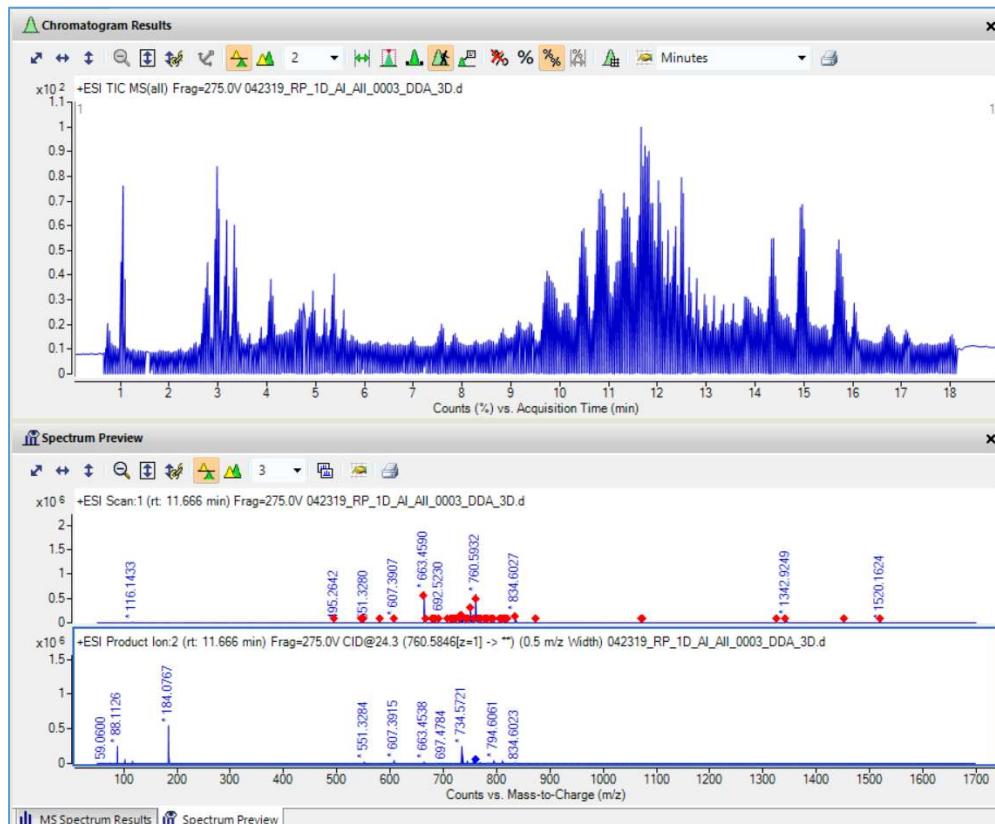
Note: For this workflow, features that have the same RT and m/z (e.g., isomers) but different IM drift times will each be written in the output data file with separated MS/MS, but some downstream LC-MS data analysis software will likely group these into the same compound. An IMPrecursorMap.csv is written to the AcqData folder of the converted 3D file and can be compared to the input CSV feature file to track isomers in the conversion.

If you have infusion/FIA data or if you want to process individual peaks of your LC run you can maintain IM isomer separation into different compounds in the resulting 3D DDA file with the second workflow in the image above.

Workflow 2: This workflow requires that all the frames be compressed down to a singular frame prior to performing feature finding in IM-MS Browser. Step 4 on the 'IM to DDA' tab automatically picks the first or second option based on the number of frames in the data file. So, if they are summed it uses the

second workflow where the IM dimension is then mapped to the LC dimension (the original LC dimension no longer exists because the frames were summed to 1). This allows IM isomers to be separated in the resulting data 3D data file.

A logical first place to examine the new 3D data files is MassHunter Qual software as these new LC-MS data files can no longer be opened in IM-MS Browser.

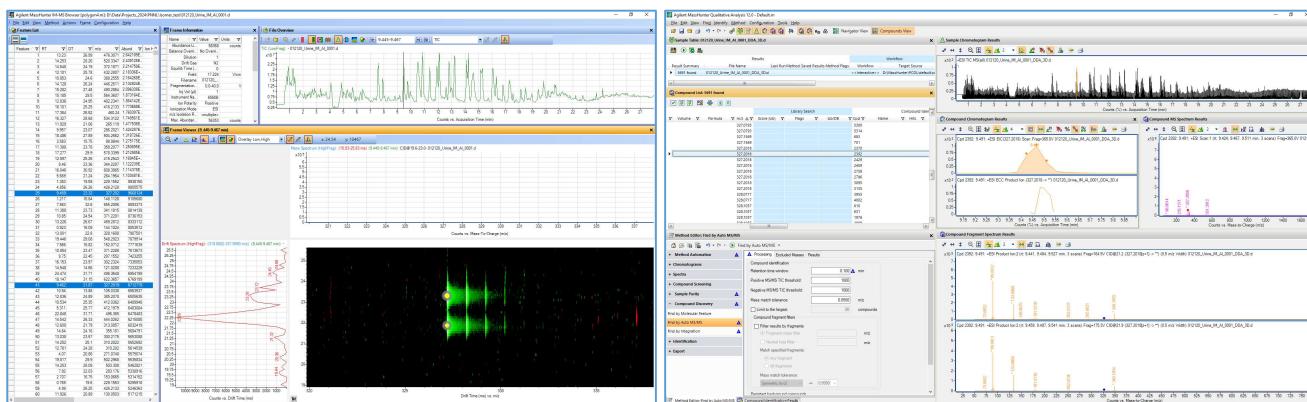


If you have worked with LC-MS Auto MS/MS data files before the image above will look familiar. As you use the “walking man” option to move through the chromatogram the MS1 and MS2 spectra will appear in the Spectrum Preview. Red diamonds on the MS1 scan indicate the precursors that have corresponding MS/MS spectra. The MS/MS spectra associated with the precursors are cycled through on the bottom spectra with a blue diamond now indicating the location of the precursor ion. If you CCS calibrated your data file the value stored in the “Frag=XXX” location in the header is the CCS value of the precursor ion.

Next, we will walk through a few examples to illustrate how IM isomers are handled with Find by Auto MS/MS in Qual. While the converted 3D file does contain MS/MS spectra for all the isomers that were found as unique features in IM-MS Browser, downstream applications may group these isomers in the same compound based on similarity in RT and m/z.

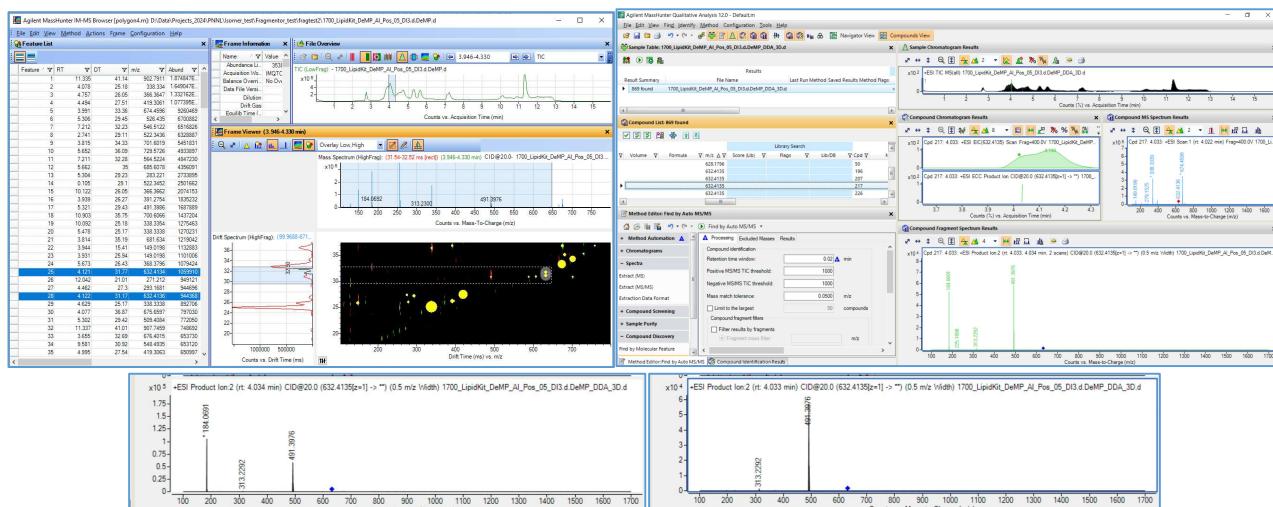


Example 1 – Workflow 1, ramped CE: In the first example, the LC separation is maintained in the conversion and a ramped CE is used in the All Ions IM/MS experiment. This results in the two isomers which have different IM arrival times having different CE values and therefore they appear as two CE experiments in the same compound.



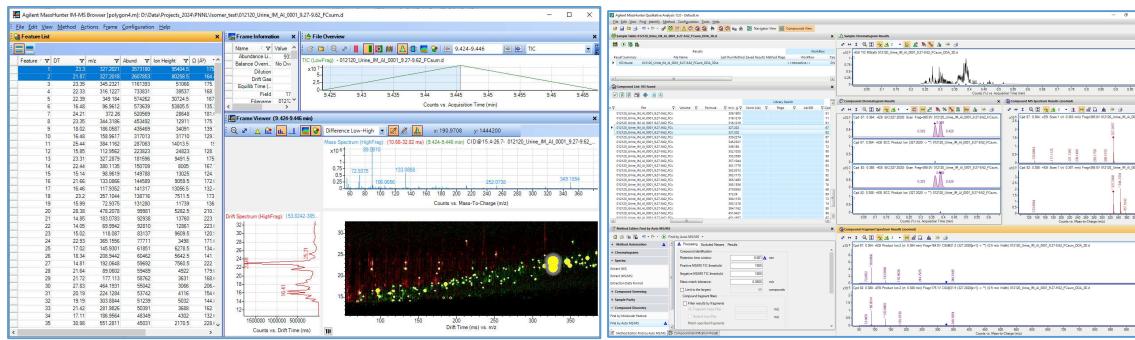
The Find by Auto MS/MS algorithm also has an option (Results tab in Method Editor for Find by Auto MS/MS) to average the MS/MS spectra across multiple Ces. This is the behavior of Find by Auto MS/MS and other data analysis software may have hard coded or other ways of handling MS/MS spectra that you will need to familiarize yourself with.

Example 2 – Workflow 1, fixed CE: In another example the ramped CE is not used and therefore the MS/MS spectra for the isomers automatically get averaged. The two individual MS/MS spectra for scan 4.033 RT and 4.034 RT are averaged in the green spectra in Find by Auto MS/MS (you can tell because the 184 fragment ion is lower in that spectra than the original 4.033 RT spectra shown below).



Note that in these examples if you really wanted to keep the isomers separate you could set the RT window to a really small value (0.001 for these data files) and you would get the isomers as separate compounds, but this results in a very large number of Auto MS/MS events which really slows down Qual and other software.

Example 3 - Workflow 2, ramped CE: Finally, in the last example we look at what happens in the use the FIA/Infusion workflow and you sum the LC peak from the first example down to 1 frame and then map to LC space. Here the two isomers are now separated in what the software thinks is the LC dimension and are therefore separate compounds.



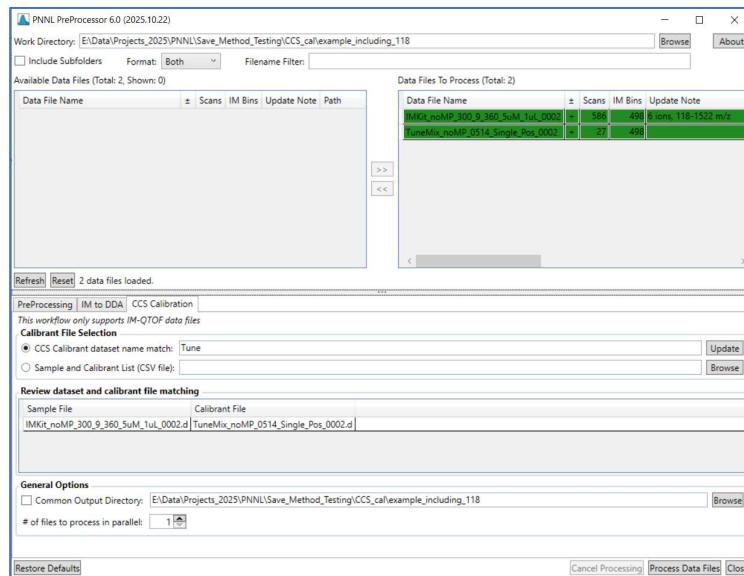
Once you have familiarized yourself with your converted data file you should be able to process it in any other downstream data analysis software packages that support the Agilent .d format. Conversions to open-source formats (i.e. mzML, abf) also work and then open the possibility to using many other third party LC-MS data analysis software packages.



16. CCS Calibration

NOTE: IF YOU DO NOT HAVE THE IM-MS BROWSER dlls THIS TAB WILL NOT APPEAR

CCS Calibration can now be performed in the PNNL PreProcessor in addition to IM-MS Browser. This allows for CCS Calibration to be performed in the logical place of the PNNL PreProcessor along with other preprocessing steps. Before the calibration can be performed the user must specify which file(s) in their folder correspond to the tune mix calibrant (used to create the calibration) and which file(s) correspond to their sample data files (calibration is applied to via copying of an OverrideImsCal file). The user has two methods to classify the data files: 1. Via name match (displayed below) where the user types in “tune” or “tm” or some other portion of the data file name that is unique to the tune mix data file, or 2. Via a csv file that contains the full path for each sample and corresponding tune mix (example is shown below). Once the name or csv file has been uploaded there is a table in the UI that allows the user to review the sample and calibrant data file classification. If positive and negative data files are present in the folder then the preprocessor can assign the correct polarity tune mix to the correct sample data files. It also can assign tune mix files based on time stamps.



The screenshot shows a Microsoft Excel spreadsheet titled "textcmdpn.csv". The first row contains the headers "sample" and "calibrant". The "sample" column lists several lipid calibration files, and the "calibrant" column lists corresponding tune mix files. The data is as follows:

sample	calibrant
1 sample	K:\Lipid_06_2024\06_27_DriftTube6560\test_pnnl_cal\posneg\cmd_line_csv\062724_Lipid_1conc_SP_01.d
2 K:\Lipid_06_2024\06_27_DriftTube6560\test_pnnl_cal\posneg\cmd_line_csv\062724_Lipid_02conc_SP_01.d	K:\Lipid_06_2024\06_27_DriftTube6560\test_pnnl_cal\posneg\cmd_line_csv\062724_TuneMix_Neg_SP.d
3 K:\Lipid_06_2024\06_27_DriftTube6560\test_pnnl_cal\posneg\cmd_line_csv\062724_Neg_Lipid_02conc_01.d	K:\Lipid_06_2024\06_27_DriftTube6560\test_pnnl_cal\posneg\cmd_line_csv\062724_TuneMix_SinglePulse.d
4 K:\Lipid_06_2024\06_27_DriftTube6560\test_pnnl_cal\posneg\cmd_line_csv\062724_Neg_Lipid_1conc_01.d	K:\Lipid_06_2024\06_27_DriftTube6560\test_pnnl_cal\posneg\cmd_line_csv\062724_TuneMix_SinglePulse.d
5 K:\Lipid_06_2024\06_27_DriftTube6560\test_pnnl_cal\posneg\cmd_line_csv\062724_Neg_Lipid_5conc_01.d	K:\Lipid_06_2024\06_27_DriftTube6560\test_pnnl_cal\posneg\cmd_line_csv\062724_TuneMix_SinglePulse.d
6	
7	
8	
9	
10	



While the result of the CCS calibration is the same as that from IM-MS Browser (assuming all tune mix ions are present in sufficient intensity) the implementation is slightly different. IM-MS Browser uses a targeted extraction process to identify the tune mix ions whereas the PNNL PreProcessor uses IMFE which is an untargeted feature extraction process. This is why it is important for the ions to be present at sufficient intensity for IMFE to find them. The number of tune mix ions and m/z range is indicated in the Update Note once the calibration determination and application has been completed (shown in the image above). The tune mix ions that were used for the calibration are also listed in the PNNL-PreProcessorLog file that gets written in the AcqData folder of the tune mix data file. The log file in the sample data files report the tune mix data file name that was used to generate the calibration.

```

PNNL-PreProcessorLog - Notepad
File Edit Format View Help
Results filter: MinCharge: 1
Results filter: MinQuality: 70
IMFE Console log (summarized):
Working on of 1: E:\Data\Projects_2025\PNNL\Save_Method_Testing\CCS_cal\example_including_118\TuneMix_noMP_0514_Single_Pos_0002.d
Using 'E:\Data\Projects_2025\PNNL\Save_Method_Testing\CCS_cal\example_including_118\TuneMix_noMP_0514_Single_Pos_0002.d_MfeParams.xml' as the parameter
0.0% complete: Finding features: initializing
0.0% complete: Finding features (step 1/105)
14.0% complete: Finding features (step 18/105)
27.0% complete: Finding features (step 33/105)
38.0% complete: Finding features (step 46/105)
49.0% complete: Finding features (step 59/105)
60.0% complete: Finding features (step 72/105)
71.0% complete: Finding features (step 85/105)
84.0% complete: Finding features (step 100/105)
90.0% complete: Loading feature results (reading)
81.0% complete: Finding features (step 97/105)
95.0% complete: Loading feature results (parsing)
100.0% complete: Finding features: done;
DONE.
100.0% complete: Total features = 2743; Time taken = 11.0 seconds; Written to: E:\Data\Projects_2025\PNNL\Save_Method_Testing\CCS_cal\example_including_
Single-Field CCS Calibration performed: Tfix: -0.148446980880454, Beta: 0.139065672065178
Calibration score: RMS Dev: 0.02190949157808, SE: 0.0268335374452903
Calibrants used:
m/z      z      CCS      DT      Residual (ms)
118.086255  1      121.3    14.99   -0.0273975041758057
322.048121  1      153.73   20.39   0.0329045668464332
622.02896   1      202.96   27.48   0.0183908904288472
922.009798  1      243.64   33.21   -0.0203630950632814
1221.990637 1      282.2    38.64   -0.0137610329563245
1521.971476 1      316.96   43.54   0.0102261749201702
10/23/2025 5:22:24 PM
Total time 11.7070086 seconds

```

```

PNNL-PreProcessorLog - Notepad
File Edit Format View Help
----- 10/23/2025 5:22:24 PM
PNNL PreProcessor 6.0 (2025.10.22)
Copied CCS calibration file OverrideImscal.xml from E:\Data\Projects_2025\PNNL\Save_Method_Testing\CCS_cal\example_including_118\TuneMix_noMP_0514_Single_Pos_0002.d
10/23/2025 5:22:24 PM
Total time 0.0010006 seconds

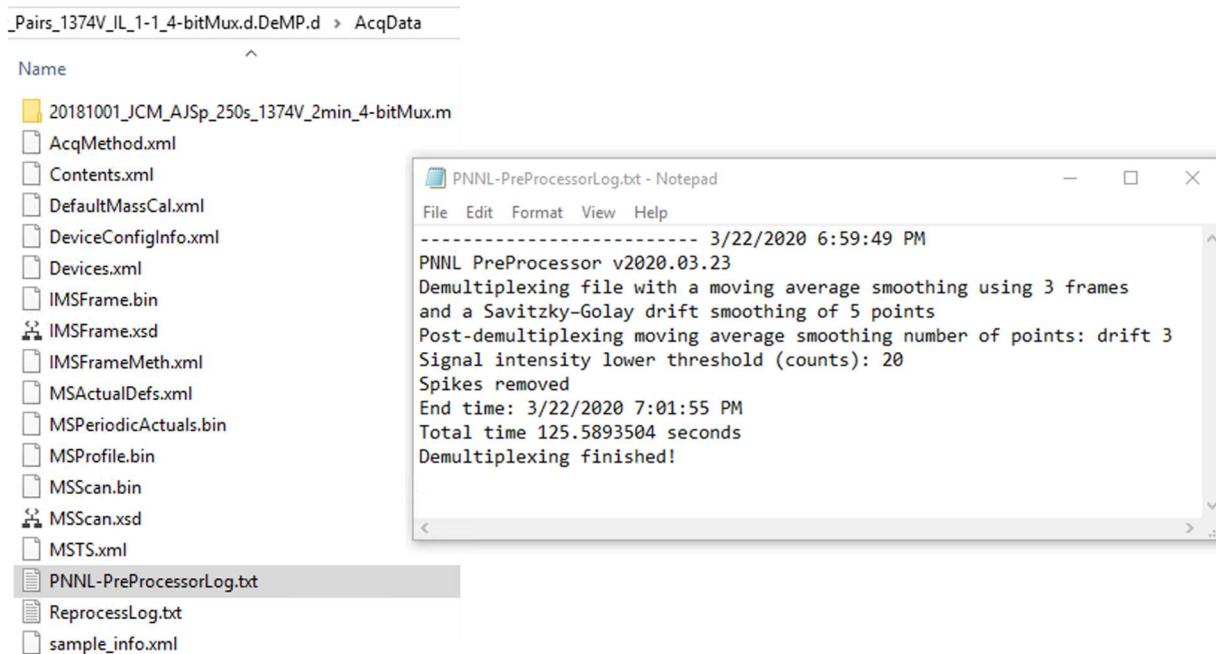
```

The CCS calibration described above in the UI is also available via command line. More details are provided by typing PNNL Preprocessor.exe -ccs -help.



17. Log files

The PNNL-PreProcessorLog.txt file in the AcqData subfolder contains information about the software version, preprocessing algorithms applied, parameters used and execution time, e.g.,



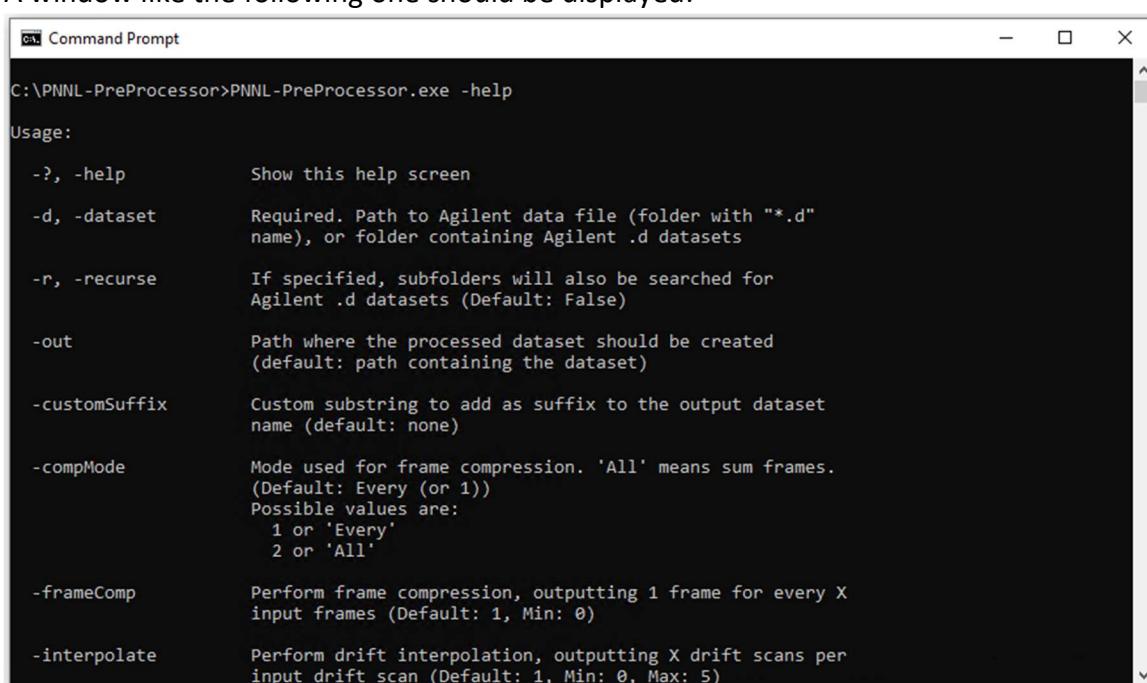
Note: Information about the saturated ions found, repaired points and specific messages for ions that were not repaired are in the text file PNNL-PreProcessorSaturation.csv.

18. Running the tool from the command prompt

Most of the functionality can be run from the Windows Command Prompt. Get started by using the following steps to display a list of available options:

1. Open the Windows Command Prompt: type "command prompt" into the Windows Start menu to search for it and click it.
2. Change your current location to the folder where the PreProcessor is in your computer. For example, if it is in "C:\PNNL-PreProcessor" type:
cd C:\PNNL-PreProcessor
3. Type the help command to see a list of available options from the first tab of the preprocessor:
PNNL-PreProcessor.exe -help

A window like the following one should be displayed:



```
PS C:\PNNL-PreProcessor>PNNL-PreProcessor.exe -help

Usage:
  -?, -help      Show this help screen
  -d, -dataset   Required. Path to Agilent data file (folder with "*.d" name), or folder containing Agilent .d datasets
  -r, -recurse    If specified, subfolders will also be searched for Agilent .d datasets (Default: False)
  -out           Path where the processed dataset should be created (default: path containing the dataset)
  -customSuffix  Custom substring to add as suffix to the output dataset name (default: none)
  -compMode      Mode used for frame compression. 'All' means sum frames. (Default: Every (or 1))
                 Possible values are:
                 1 or 'Every'
                 2 or 'All'
  -frameComp     Perform frame compression, outputting 1 frame for every X input frames (Default: 1, Min: 0)
  -interpolate   Perform drift interpolation, outputting X drift scans per input drift scan (Default: 1, Min: 0, Max: 5)
```

4. The IM-to-DDA tab has limited command line support which is visible via an additional -dda command to see the options:

PNNL PreProcessor.exe -dda -help



5. The new CCS Calibration tab has command line support, but requires an additional -ccs command to see the options:

PNNL PreProcessor.exe -ccs -help

```

E:\Data\Projects_2025\PNNL\Release_6.0_20251022>PNNL-PreProcessor.exe -ccs -help

Usage:

-?, -help      Show this help screen
-d, -dataset   Path to Agilent data file (folder with "*.d" name),
               folder containing Agilent .d datasets, or text with full
               paths to datasets (one per line).
-r, -recurse   If specified, subfolders will also be searched for
               Agilent .d datasets (Default: False)
-out          Path where the processed dataset should be created
               (default: path containing the dataset)
-threads       Max number of threads to use when processing (multiple
               files at once) (Default: 1)
-calMatch     Text to match in a dataset name for CCS calibrant
               reference files; required when '-d'/'-dataset' argument
               is used (Default: "")
-csv          CSV file path for sample and calibrant assignments;
               'sample' and 'calibrant' columns required, with full
               paths to data files; not used when '-d'/'-dataset' is
               supplied (Default: null)
-recal         If supplied, existing CCS calibration results in
               calibrant reference files will be ignored/overwritten
               (Default: False)
-ParamFile    Path to a file containing program parameters. Additional
               arguments on the command line can supplement or override
               the arguments in the param file. Lines starting with '#'
               or ';' will be treated as comments; blank lines are
               ignored. Lines that start with text that does not match a
               parameter will also be ignored.
-CreateParamFile Create an example parameter file. Can supply a path; if
                  path is not supplied, the example parameter file content
                  will output to the console.

E:\Data\Projects_2025\PNNL\Release_6.0_20251022>

```

18.1 Example commands

Apply smoothing in IM and LC to a single MS-file:

PNNL-PreProcessor.exe -d MyDataFile.d -smooth -driftKernel 3 -lcKernel 3 -minIntensity 20

* Note: Kernel refers to number of points used for smoothing.

Apply smoothing in IM and LC to all MS-files inside a folder called “C:\MyData\Experiment1”:

PNNL-PreProcessor.exe -d C:\MyData\Experiment1 -smooth -driftKernel 3 -lcKernel 3 -minIntensity 20

* Note: if your folder path has spaces use quotation marks: “C:\My Data\Experiment 1”.

Apply smoothing in IM and LC to all MS-files in a list in text format called “C:\MyData\DataFiles.txt”, where each line contains an MS-file name including a full path (this allows processing of multiple MS-files from different directories):

PNNL-PreProcessor.exe -d C:\MyData\DataFiles.txt -smooth -driftKernel 3 -lcKernel 3 -minIntensity 20

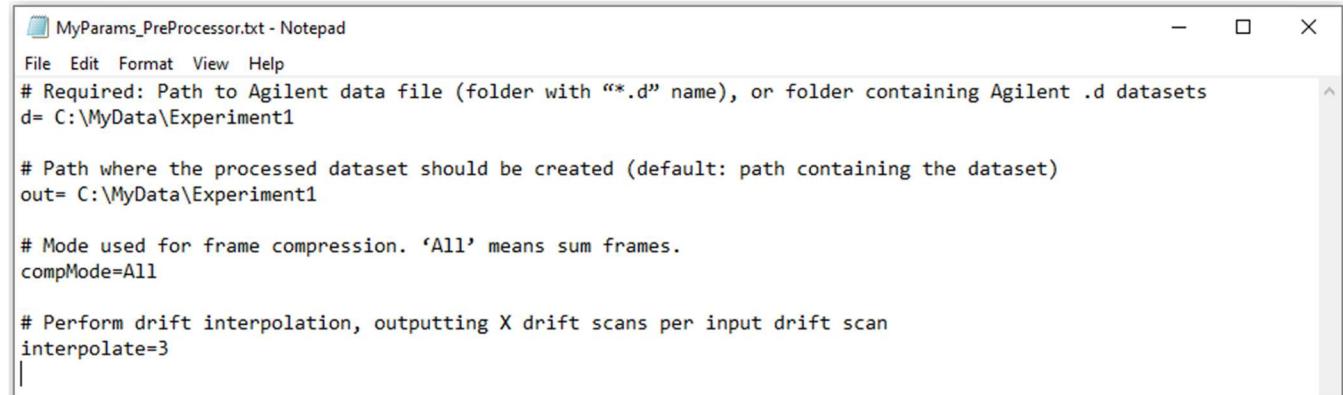


18.2 Using a file with parameters

Generate and use an example file with parameters inside a folder called “C:\MyData\Experiment1”:

PNNL-PreProcessor.exe -CreateParamFile C:\MyData\Experiment1\MyParams_PreProcessor.txt

You can open and edit the parameters file using the Notepad (lines starting with the symbol '#' indicate that it is a comment and will be ignored). Below is an example for compressing all frames into a single frame and performing drift interpolation with 3 drift bins per every drift bin, for all MS-files in a folder:



```
MyParams_PreProcessor.txt - Notepad
File Edit Format View Help
# Required: Path to Agilent data file (folder with "*.d" name), or folder containing Agilent .d datasets
d= C:\MyData\Experiment1

# Path where the processed dataset should be created (default: path containing the dataset)
out= C:\MyData\Experiment1

# Mode used for frame compression. 'All' means sum frames.
compMode=All

# Perform drift interpolation, outputting X drift scans per input drift scan
interpolate=3
|
```

To use the file, save the changes and run:

PNNL-PreProcessor.exe -ParamFile C:\MyData\Experiment1\MyParams_PreProcessor.txt

* Note: The steps are performed in the order as they appear in the GUI, regardless of the order in the command prompt or the parameter file.

