

Manual

ROI-MCR-ALS for Metabolomics with MatLab

1. Installation of required software

1.1. Installation of MatLab

To download MatLab with a campus license go to

<https://uk.mathworks.com/academia/tah-support-program/eligibility.html>

and download the newest Version of MatLab (R2021a Update 5 or newer required). If needed use this guidance manual:

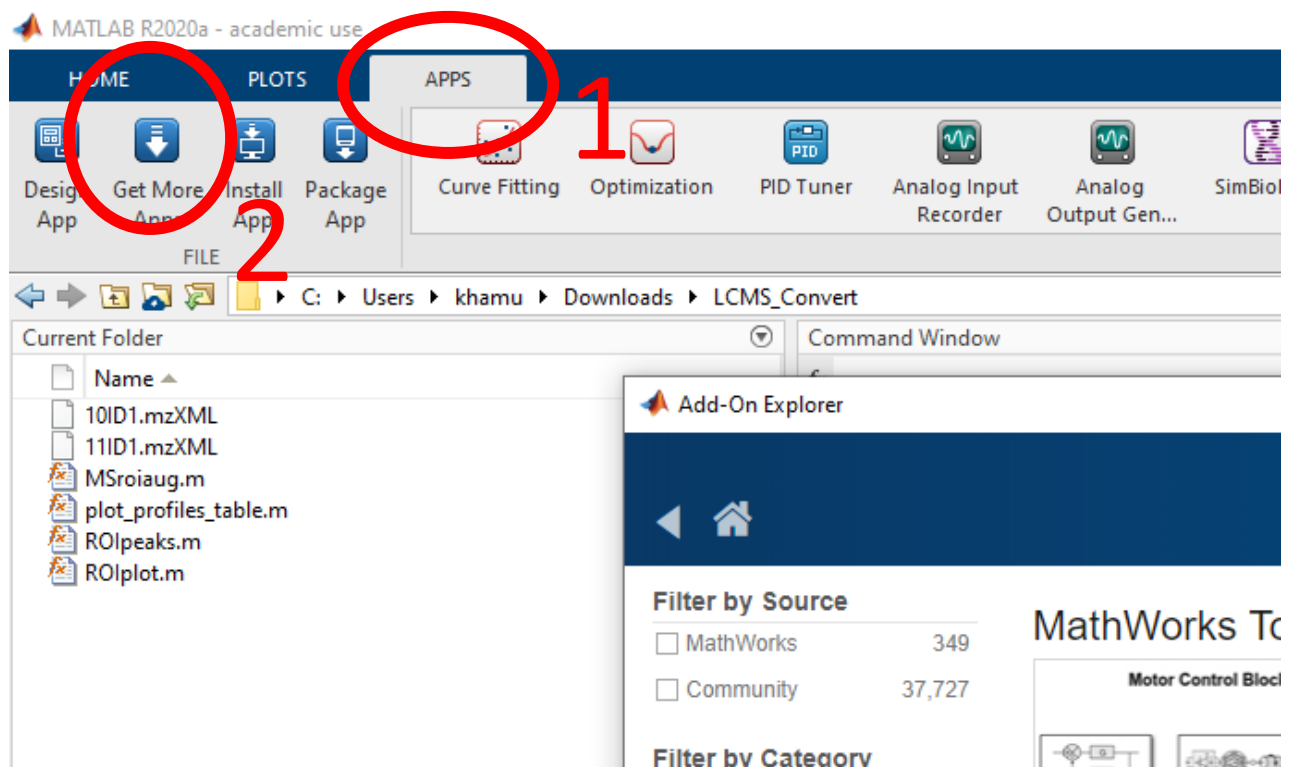
<https://www.hs-aalen.de/uploads/mediapool/media/file/5814/MatLab.pdf>

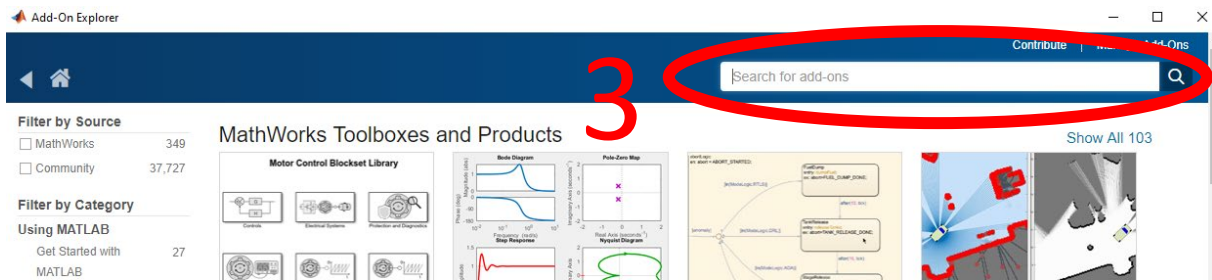
1.2. Installation of required MatLab Add-Ons

Go to the Apps tab, click on "Get more Apps" and type in the search field:

- Statistics and Machine Learning Toolbox
- Bioinformatics Toolbox
- Parallel Computing Toolbox

Install all three Toolboxes





1.3. Add ROIprocess Functions to MatLab Path

Download ROIprocess here:

<https://github.com/AdrianHaun/ROIprocess>

Download MCR-ALS Toolbox 2.0 here:

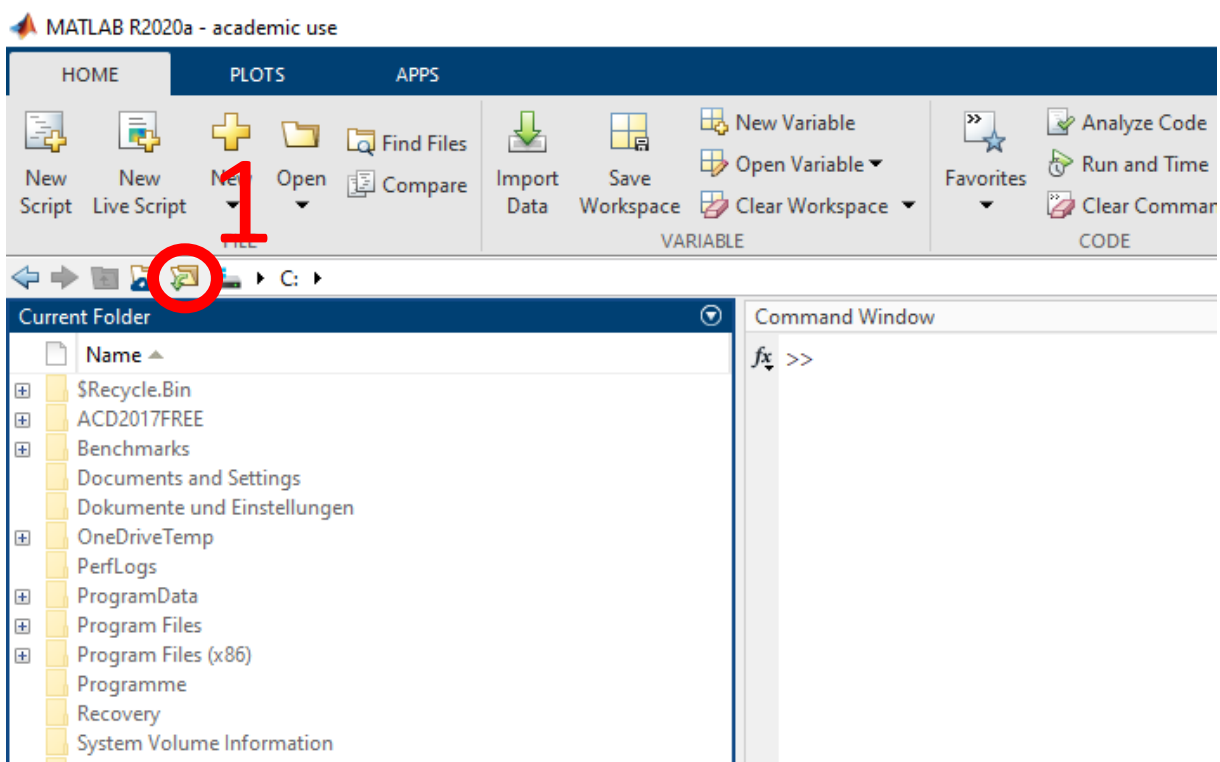
<https://mcrals.wordpress.com/download/mcr-als-2-0-toolbox/>

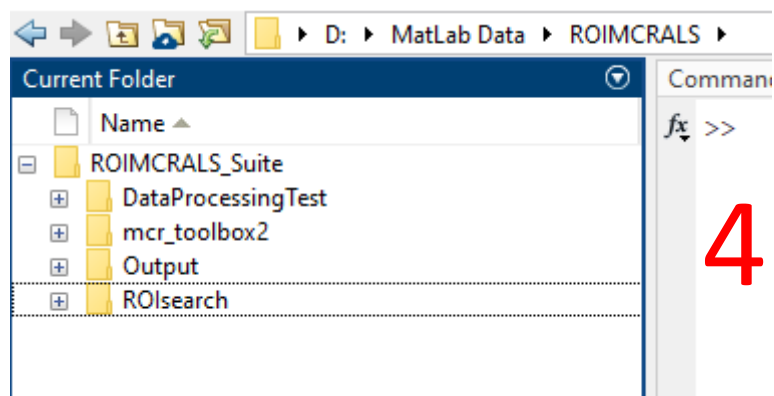
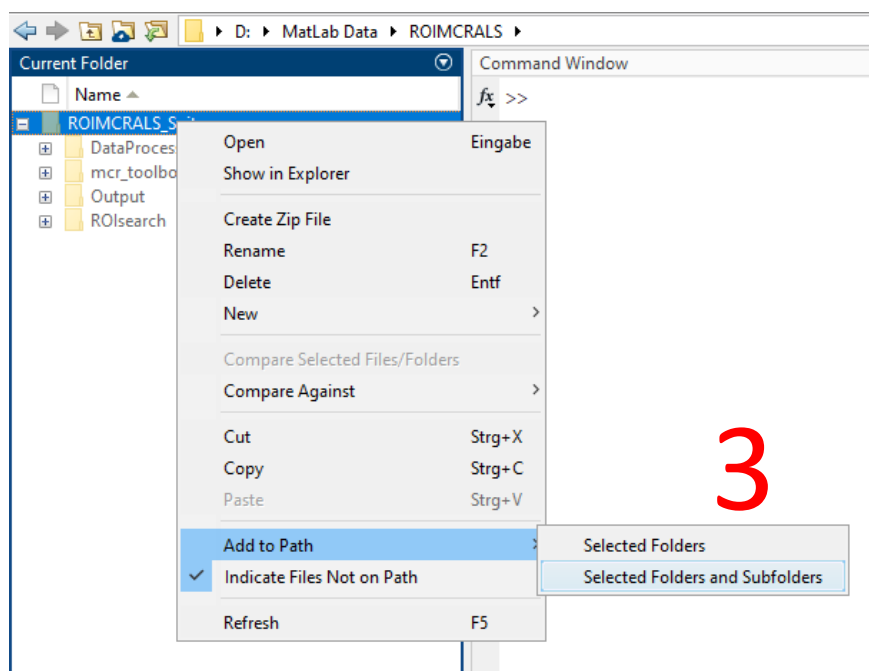
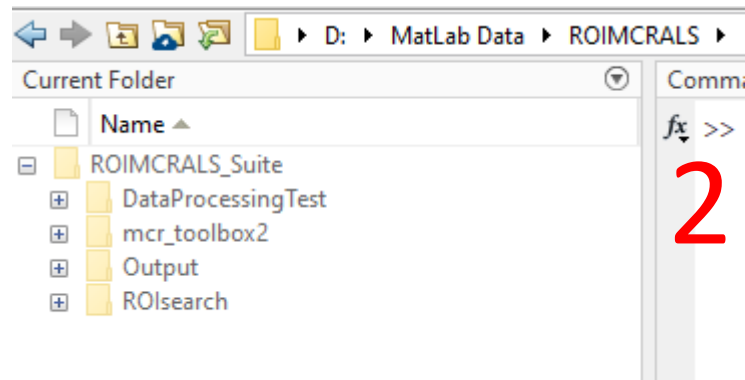
Use a file unpacking software (eg. WinRAR or 7zip) to unpack both .zip files to a folder on your PC and remember the location.

Start MatLab, on the left side click "Browse for folder" and open the folder containing the previously unpacked files.

The sub folders **DataProcessingTest**, **mcr_toolbox2**, **ROIsearch** and **Output** must be in the current folder tab.

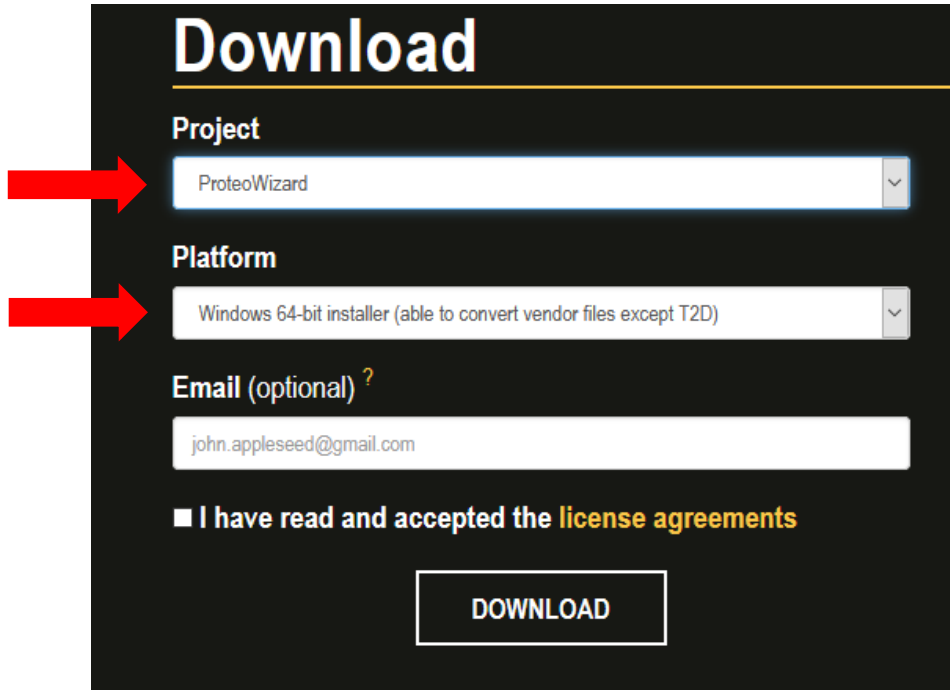
If they are grayed out right click on the parent folder und select Add to path -> Selected Folders and Subfolders





1.4. Download ProteoWizard

Go to <http://proteowizard.sourceforge.net/download.html> and download the ProteoWizard Project



The screenshot shows a download form titled "Download" on a dark background. It contains three main sections: "Project", "Platform", and "Email (optional)". The "Project" dropdown is set to "ProteoWizard" and the "Platform" dropdown is set to "Windows 64-bit installer (able to convert vendor files except T2D)". Two red arrows point to these dropdowns. The "Email" field contains "john.appleseed@gmail.com". Below the email field is a checkbox labeled "I have read and accepted the license agreements", which is checked. At the bottom is a "DOWNLOAD" button.

Download

Project

ProteoWizard

Platform

Windows 64-bit installer (able to convert vendor files except T2D)

Email (optional) ?

john.appleseed@gmail.com

☒ I have read and accepted the **license agreements**

DOWNLOAD

To check whether you're using a 32-bit or 64-bit version of Windows follow the steps on the following website:

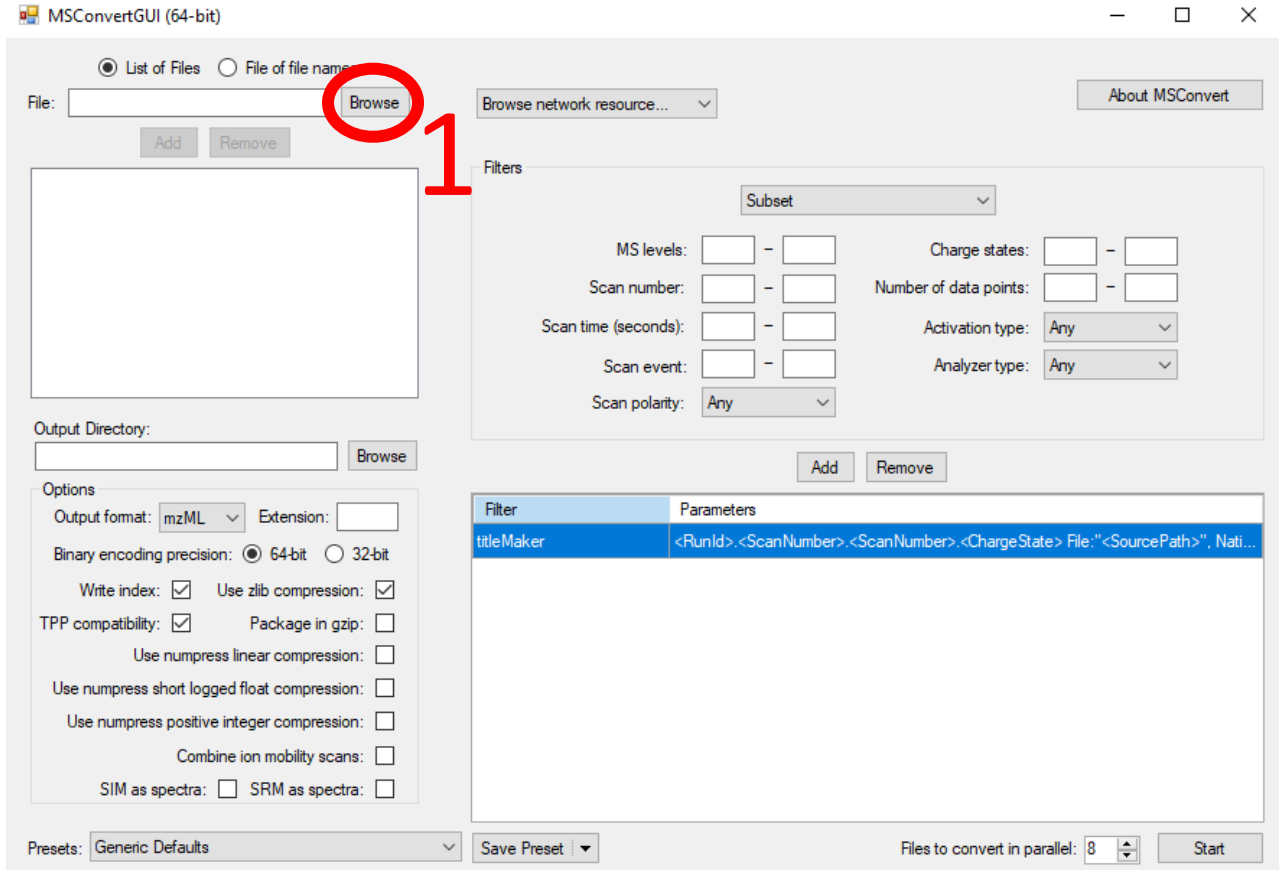
<https://www.howtogeek.com/howto/21726/how-do-i-know-if-im-running-32-bit-or-64-bit-windows-answers/>

After download run the installer to install MSConvert.

2. Data Conversion

2.1. Data Conversion

Launch MSConvert and open the MS files you want to convert. Click Browse, a new window opens, then navigate to the MS files and select them. Click open.



MSConvertGUI (64-bit)

File: **Browse**

Output Directory:

Options

Output format: Extension:

Binary encoding precision: ☒ 64-bit ☐ 32-bit

Write index: ☒ Use zlib compression: ☒

TPP compatibility: ☒ Package in gzip: ☐

Use numpress linear compression: ☐

Use numpress short logged float compression: ☐

Use numpress positive integer compression: ☐

Combine ion mobility scans: ☐

SIM as spectra: ☐ SRM as spectra: ☐

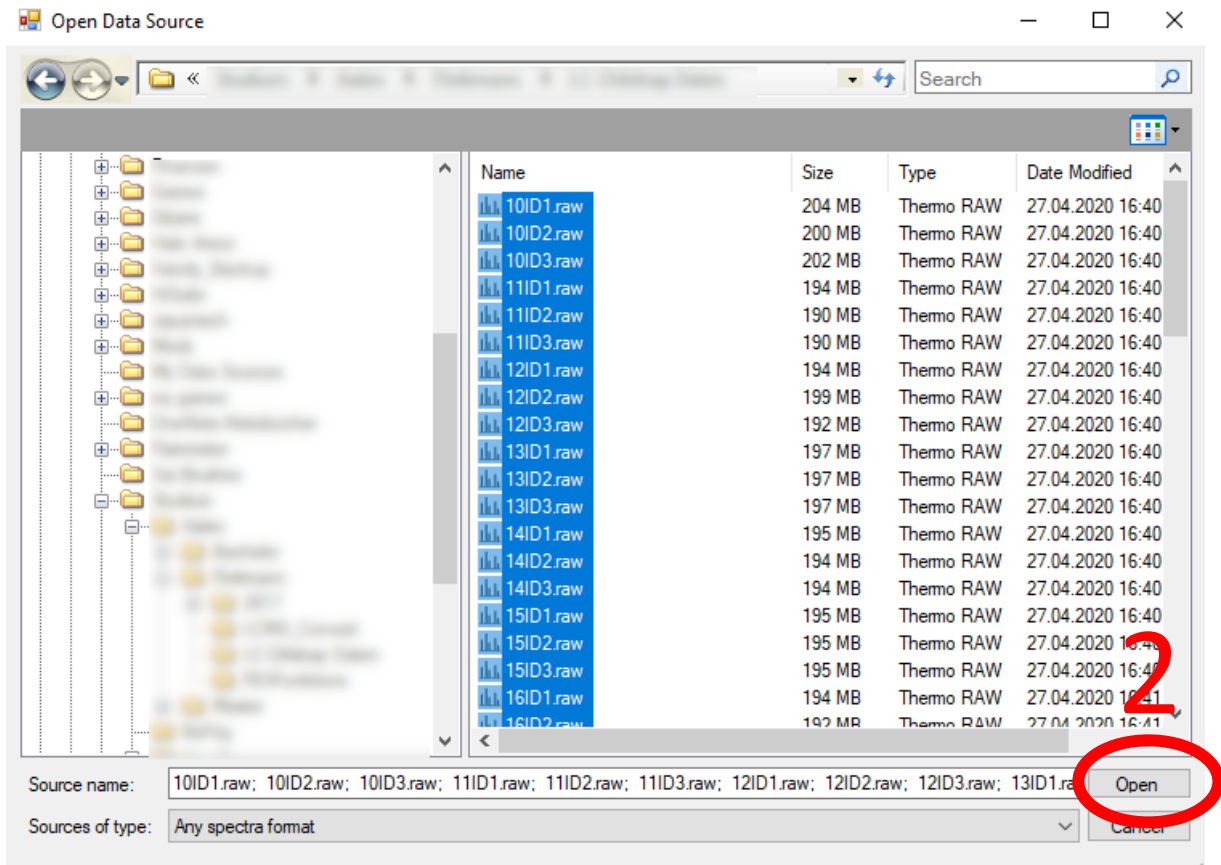
Presets:

Files to convert in parallel:

Filters

Subset

| Filter | Parameters |
|------------|--|
| titleMaker | <RunId>.<ScanNumber>.<ScanNumber>.<ChargeState> File:"<SourcePath>", Nati... |



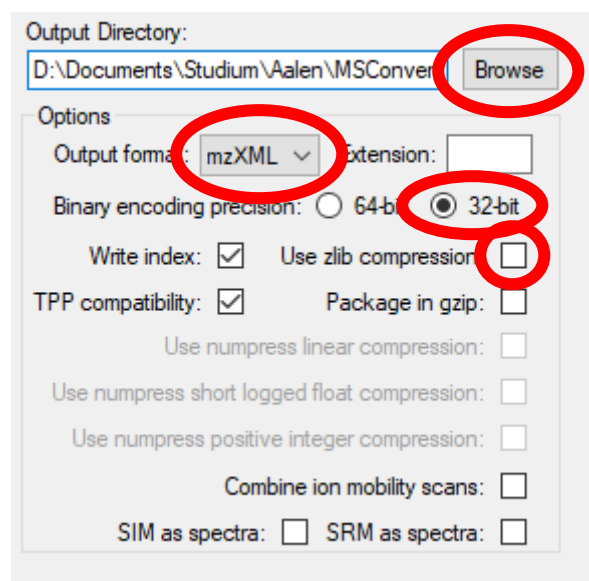
Select an output directory by clicking the Browse button and selecting a folder.

Change following conversion options:

Output format: mzXML

Binary encoding precision: 32-bit

Use zlib compression: untick



Click Start, a new window opens, wait until data conversion is complete.

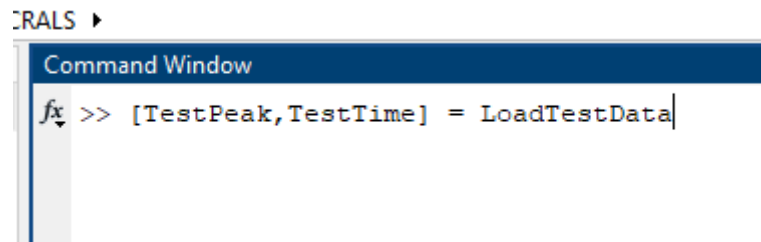
3. Load and Review Test File

3.1. Load Test Data file

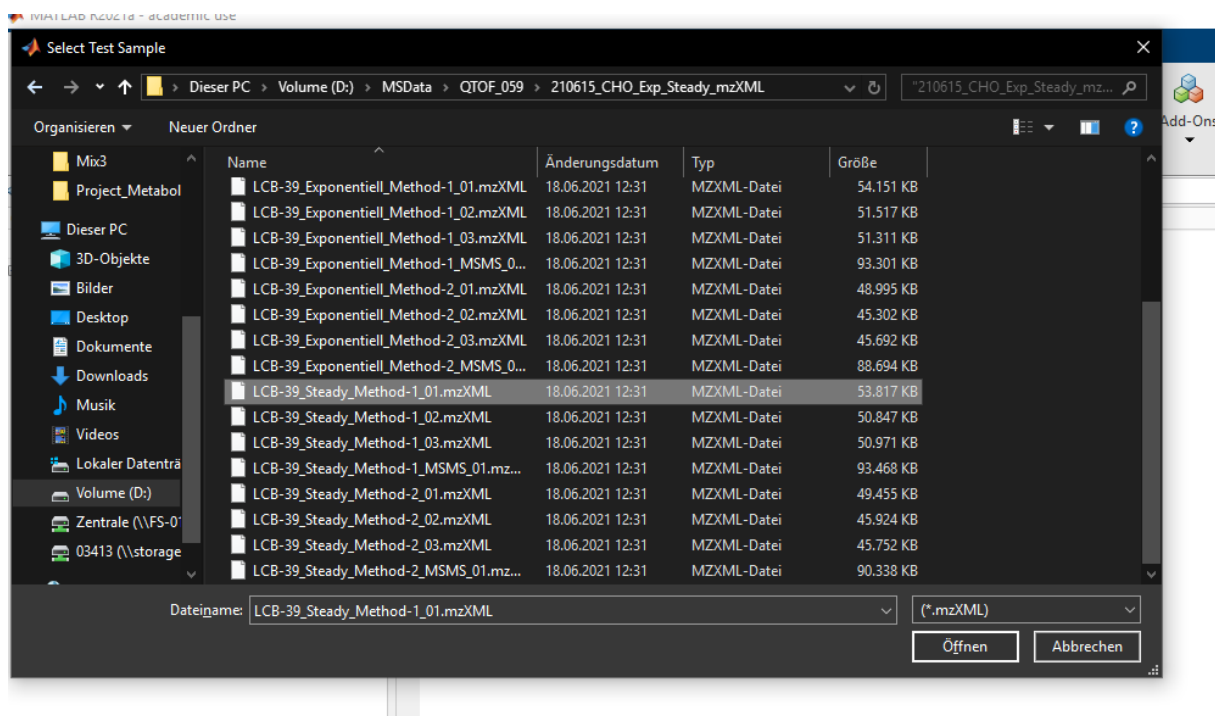
Type or copy the command

`[TestPeak,TestTime] = LoadTestData`

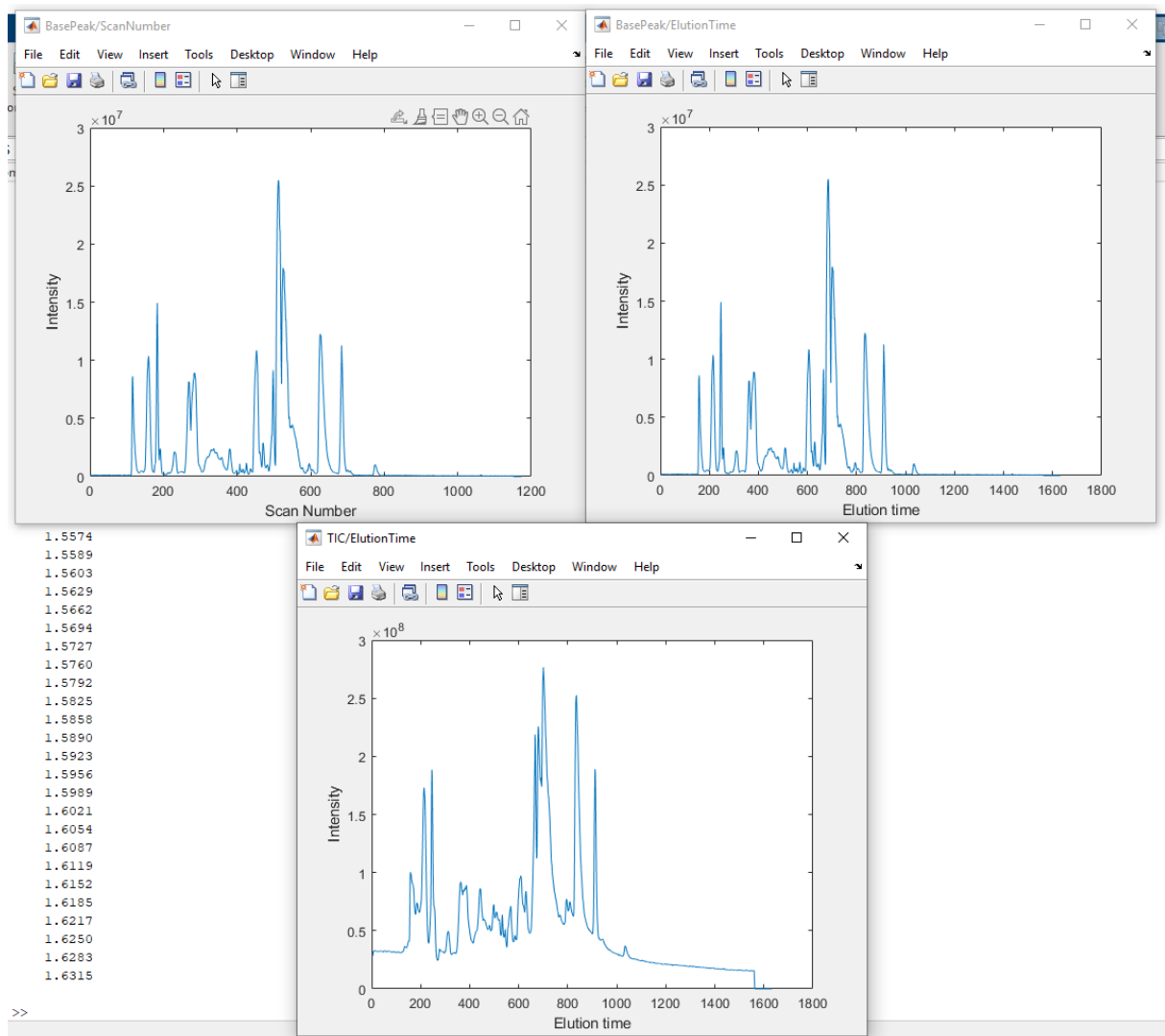
in the command window and press Enter.



A window will appear, select the test file you want to review and click Open.



After file loading three plots are generated.



Review the Basepeak/Scan number plot to determine the starting scan number and last scan number which contain relevant information as well as noise intensity. These numbers are important parameters in the ROI search.

Scan numbers and intensity values can be read out by hovering the mouse cursor over the plot. In this example scan 80 – 800 contain peaks and background noise is around 120000.

The figures can be closed after that.

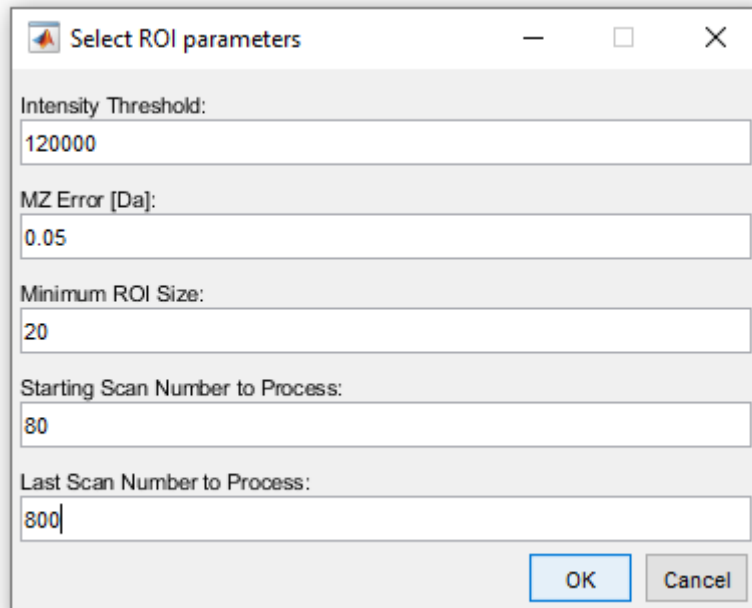
3.2. Test ROI parameters

Type or copy the command

```
[MSroi_test,mzroi_test,Times] = TestROI(TestPeak,TestTime)
```

in the command window and press Enter.

A window will appear, type in the ROI search parameters you want to test and click OK.

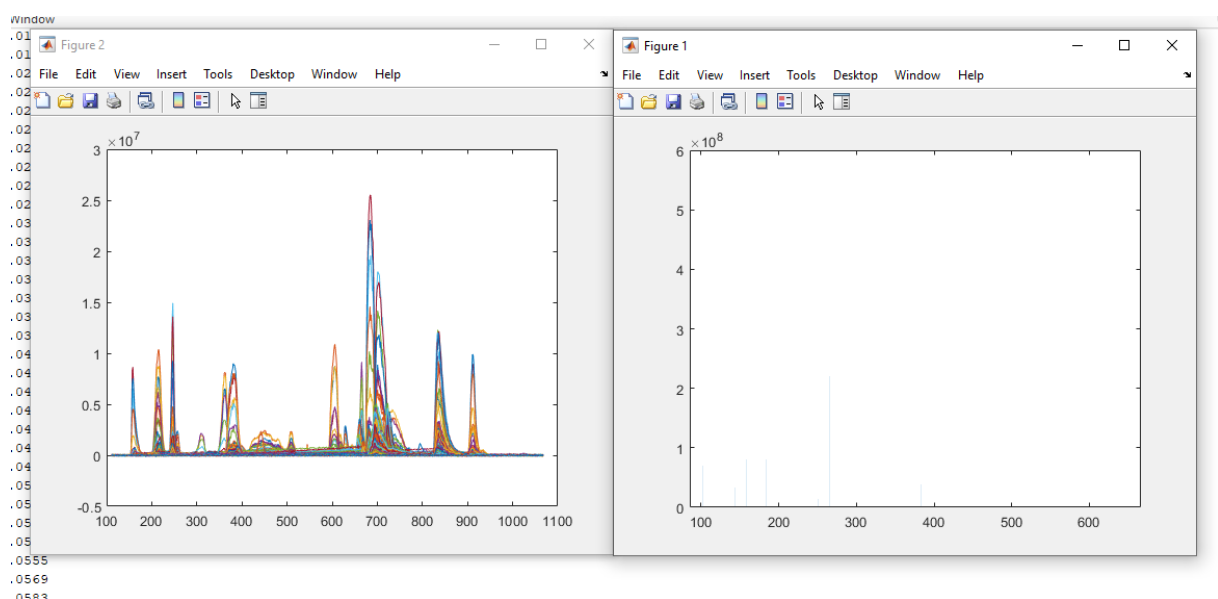


The dialog box titled "Select ROI parameters" contains the following fields and values:

- Intensity Threshold: 120000
- MZ Error [Da]: 0.05
- Minimum ROI Size: 20
- Starting Scan Number to Process: 80
- Last Scan Number to Process: 800

Buttons: OK, Cancel

The ROI search is performed and two figures appear. Figure 1 plots the intensity of every found ROI mz value and figure 2 shows the chromatogram for every ROI.



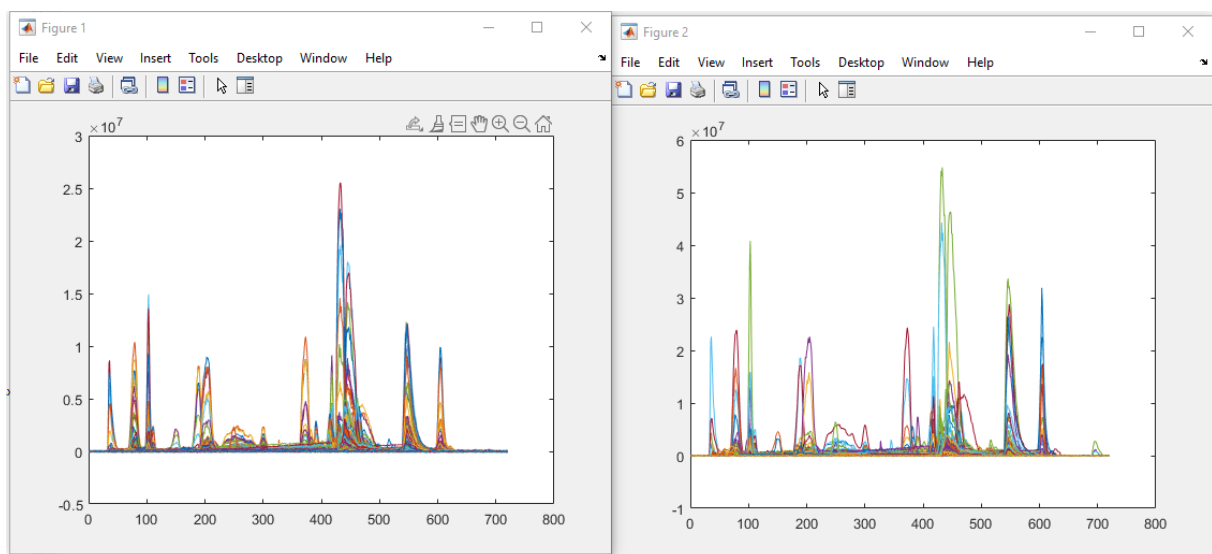
To test different parameters rerun the command.

Rename the output variables or the previously generated workspace variables are overridden, e.g.:

```
[MSroi_test2,mzroi_test2,Times2] = TestROI(TestPeak,TestTime)
```

To compare the two results visually use the commands:

```
figure
plot(MSroi_test)
figure
plot(MSroi_test2)
```



Repeat until satisfactory parameters are found.

3.3. Test Baseline Correction (optional)

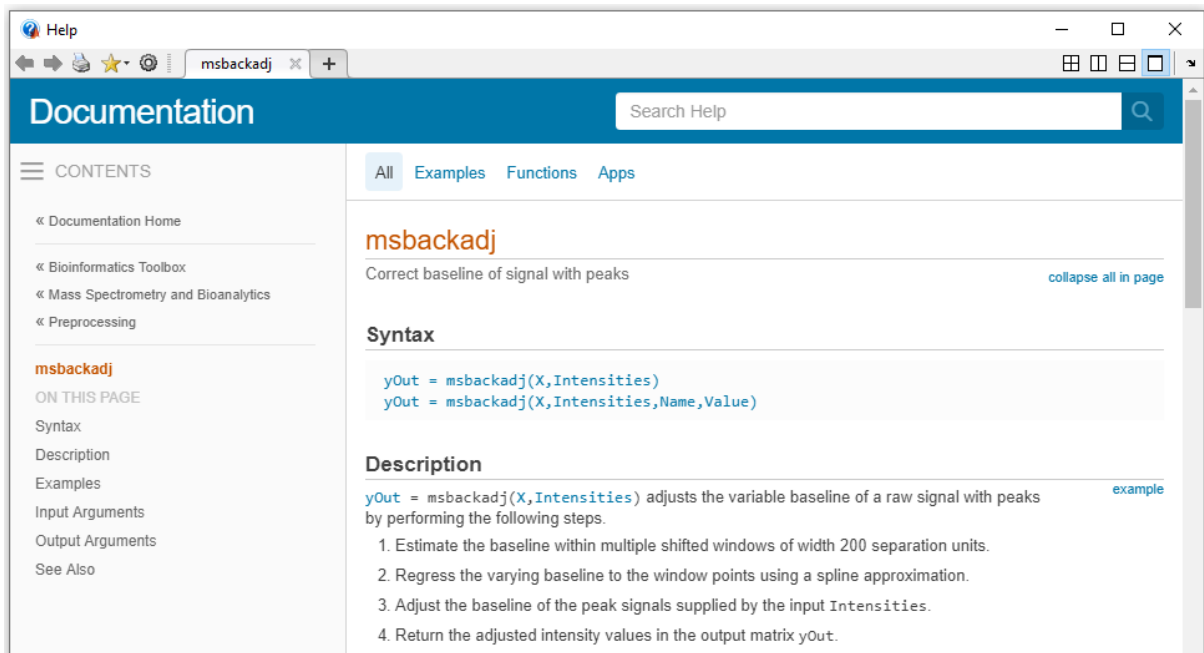
If baseline correction is desired use the command:

```
MSroiBaseCorr=TestBaseCorrect(MSroi_test,Times)
```

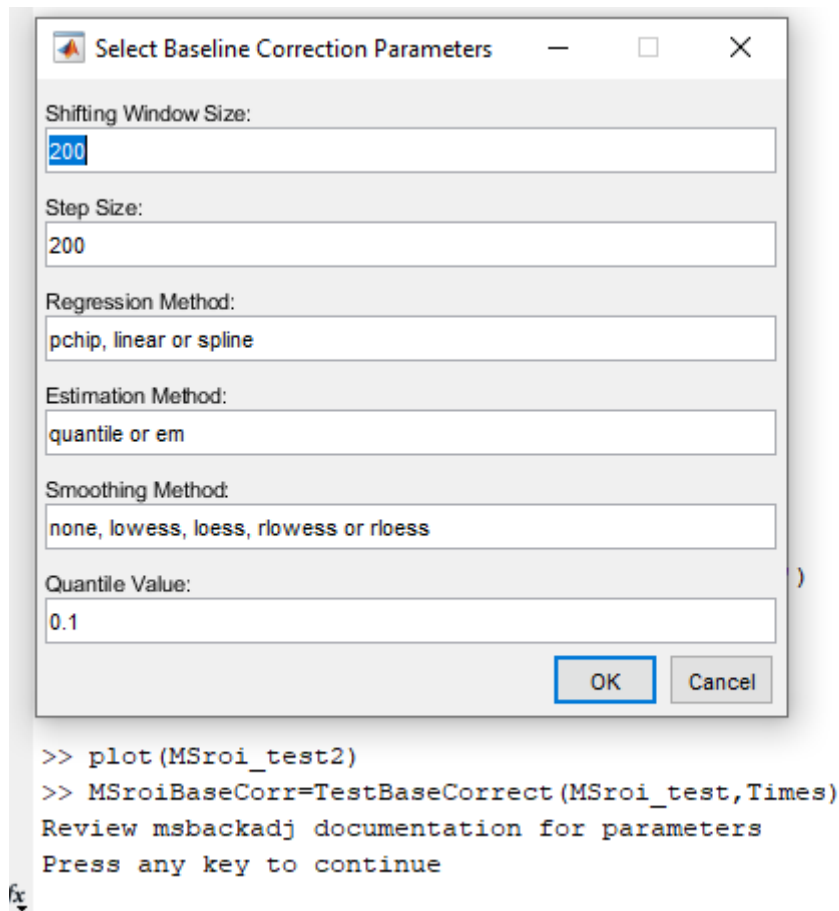
to test baseline correction parameters.

Remember to change the input variable **MSroi_test** and **Times** to those workspace variables generated by the ROI parameters you want to use further.

The documentation for **msbackadj** opens, review it for the different parameters that can be selected.

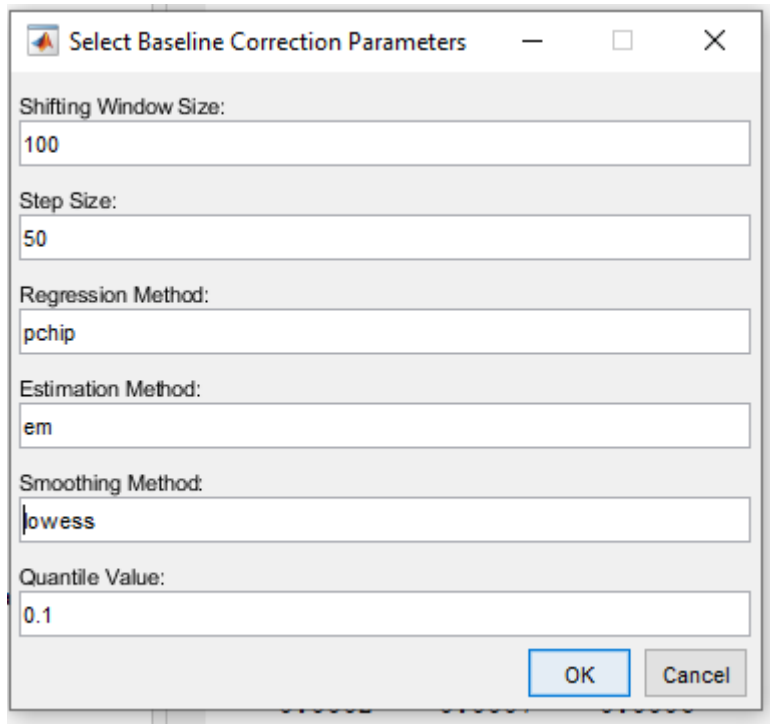


After that return to the command window and press a key to continue.



Define all parameters and click OK.

Caution: Regression Method, Estimation Method and Smoothing Method must be typed in as text, and only one per line. If not, an error is produced.

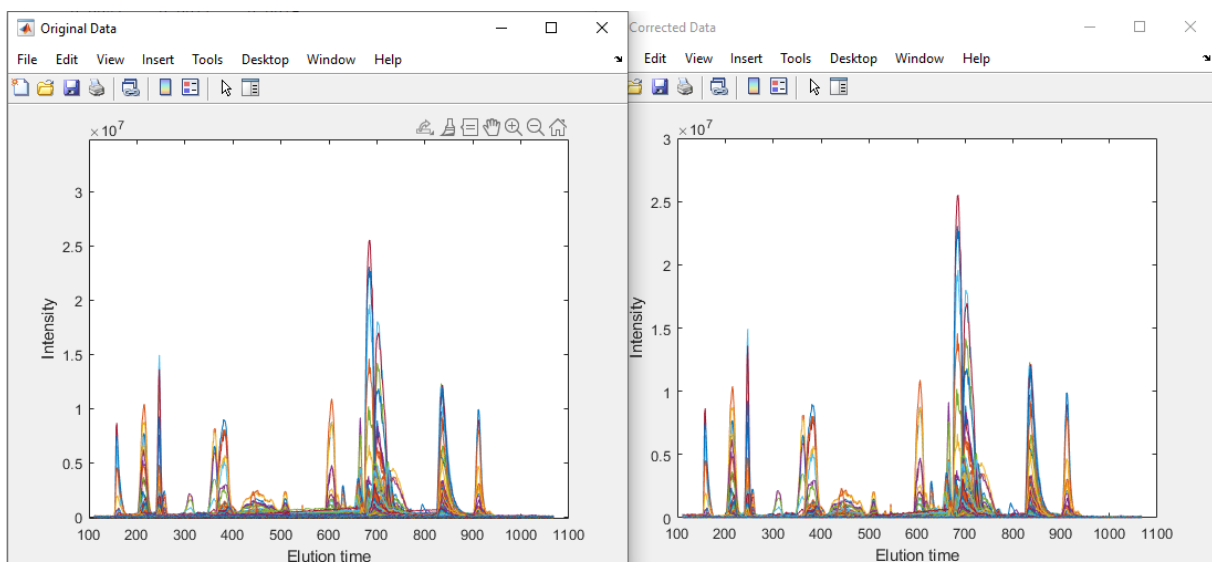


The dialog box titled "Select Baseline Correction Parameters" contains the following fields:

- Shifting Window Size: 100
- Step Size: 50
- Regression Method: pchip
- Estimation Method: em
- Smoothing Method: lowess
- Quantile Value: 0.1

Buttons: OK, Cancel

This step can take several minutes.



Two plots are generated, the original data and the corrected data. Review the plots and redo this step if necessary, until satisfactory parameters are found.

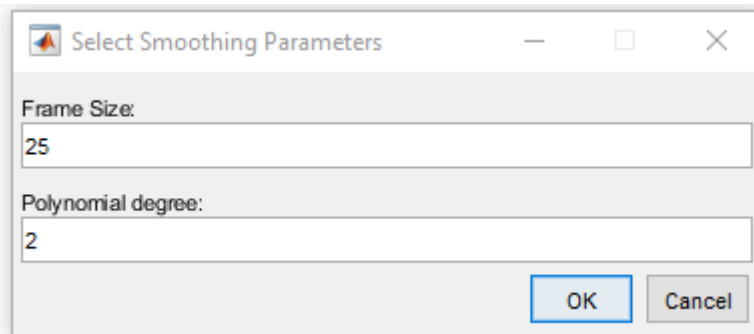
3.4. Test Savitzky-Golay-Filter (optional)

If Savitzky-Golay smoothing is desired use the command:

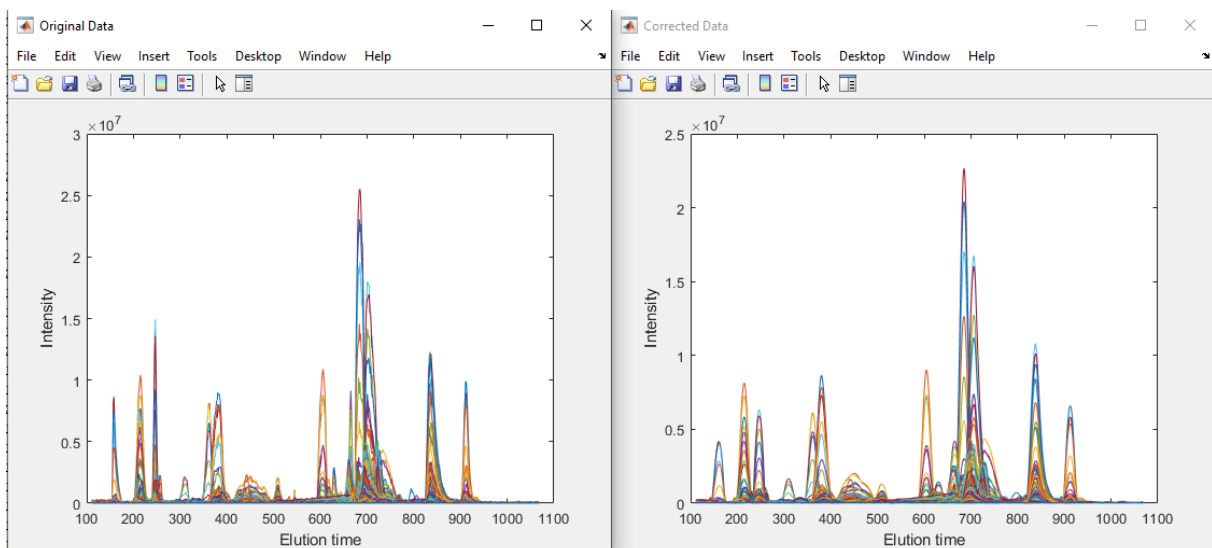
```
MSroiSmooth=TestGolaySmooth(MSroiBaseCorr,Times)
```

to test smoothing parameters.

A window appears, define smoothing parameters to test and hit enter.



Two plots are generated, the data from the previous step and the corrected data. Review the plots and redo this step if necessary, until satisfactory parameters are found.



Generated workspace variables are not needed for batch processing. Save them if desired and remove them by using

clear all

in the command window.

3.5. ROI and data pretreatment Batch processing

Run the command

```
[MSroi_end,mzroi_end,time_end,nRows,Parameters] = ROIprocess
```

in the command window.

All further parameter inputs and file selections are handled by GUIs.

The options in order of appearance are:

| Input | Options | Description | Note |
|-------------------------------|-------------------|---|---|
| Pairwise or single analysis? | Pairwise / Single | Pairwise for comparison of two groups or same sample with different measurement techniques. Single for e.g. repetition measurements | Pairwise adds fold change and significance test to evaluation output. |
| Select Dataset (1) | File explorer | Select data files (for group 1) | |
| Select Dataset 2 | File explorer | Select data files for group 2 | Only for pairwise analysis. Different number of data files per group are allowed |
| Subtract Blank? | Yes / No | Enables subtraction of blank files | |
| Select Blanks (for Dataset 1) | File explorer | | Multiple files allowed. If multiple blanks are selected, the average blank file will be used. |
| Select Blanks for Dataset 2 | File explorer | | Only for pairwise analysis. Different number of blank files per group are allowed |

| Input | Options | Description | Note |
|---------------------------------------|--|--|---|
| Datasets with different Scan Numbers? | Yes / No | Enables processing of data sets measured with different separation techniques. E.g. Group 1 LC-MS, Group 2 CE-MS | Only for pairwise analysis. |
| ROI parameters | Intensity threshold MZ error [Da] Minimum ROI Size Starting Scan Number to process (Dataset 1) Last Scan Number to Process (Dataset 1) Starting Scan Number to process Dataset 2 Last Scan Number to Process Dataset 2 | Minimum ROI Intensity Mass window considered to be the same ROI Minimum number of occurrences necessary to be an ROI Can be used to limit ROI search to relevant chromatographic region Can be used to limit ROI search to relevant chromatographic region for Group 2 | Inputs for Dataset 2 only for pairwise analysis when enabled by previous Input. Last scan number must be less or equal to number of scans in data files of that group, or an error is produced. Use [TestPeak,TestTime] = LoadTestData To check number of scans. |
| Perform Baseline Correction? | Yes / No | Enables Baseline Correction processing | |

| Input | Options | Description | Note |
|--------------------------------|---|--|--|
| Baseline Correction Parameters | Shifting Window Size Step Size Regression Method Estimation Method Smoothing Method Quantile Value | See https://uk.mathworks.com/help/bioinfo/ref/msbackadj.html for detailed parameter description | Only when enabled by previous Input. |
| Apply Savitzky-Golay Filter? | Yes/No | Enables Smoothing processing | |
| Select Smoothing Parameters | Frame Size Polynomial degree | | Only when enabled by previous Input. |
| Remove Isotopes? | Yes/No | Uses Isotope masses in the file UWPR_CommonMassSpec Contaminants.xls and removes isotopes from output | |
| Remove Adducts? | Yes/No | Uses Adduct masses in the file UWPR_CommonMassSpec Contaminants.xls and removes Adducts from output | |
| Remove common Contaminants? | Yes/No | Uses common contaminant masses in the file UWPR_CommonMassSpec Contaminants.xls and removes them from output | |
| Choose MS polarity | Positive / Negative | Sets correct contaminant list for common contaminant removal | Only when common contaminant removal is chosen |

Caution: ROIprocess uses parallelization to process as many data files at the same time as possible. This uses high amounts of system memory, especially when working with high resolution MS files and setting MZerror very low. Causing out of memory errors.

By lowering the number of workers less memory is needed, but process times are increased.

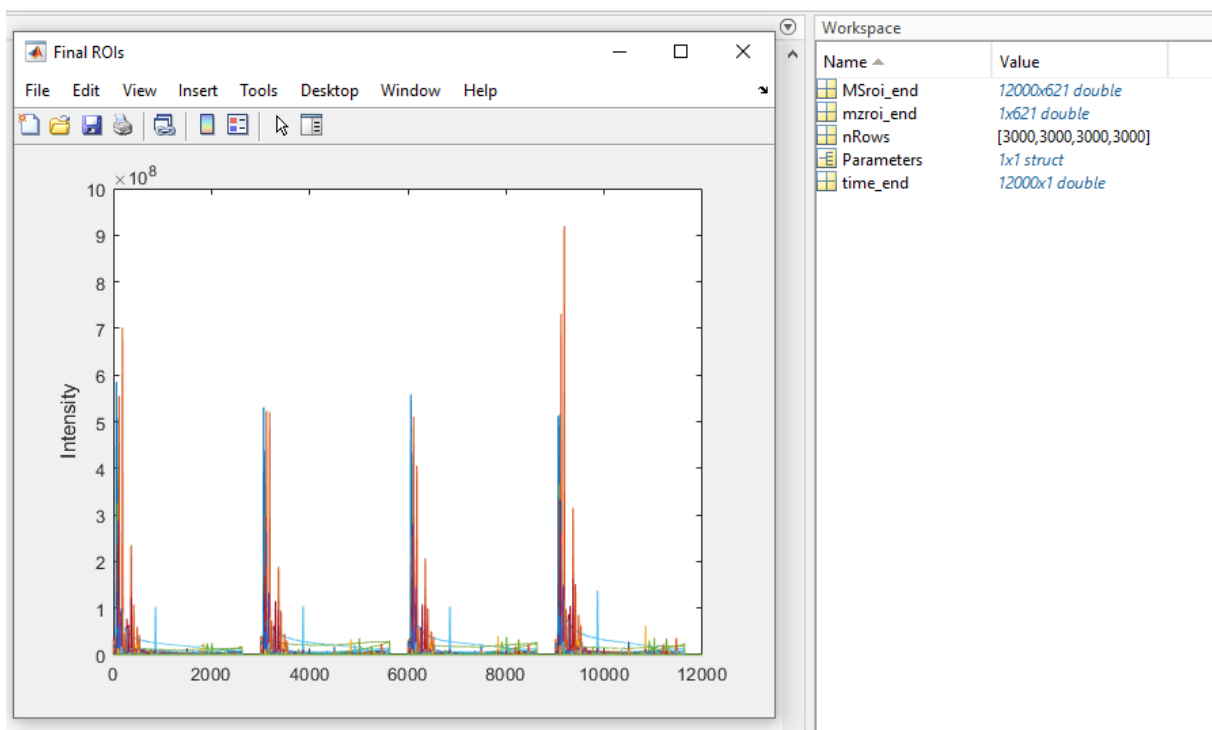
You can change the number of workers on the **Home tab** in the **Environment section**, by selecting **Parallel > Parallel Preferences**.

See <https://uk.mathworks.com/help/parallel-computing/parallel-preferences.html> for more info.

The script ROIprocess will extract, ROI search, and combine all .mzXML files selected in the previous step and applies selected data processing steps.

This can take several minutes, depending on file size and selected parameters.

Processing is done when the output **MSroi_end**, **mzroi_end**, **time_end**, **nRows** and **Parameters** appear in the workspace. **MSroi_end** is also plotted and shows the final ROIs.



The Plot can be closed after reviewing.

MSroi_end contains the intensities of the ROIs, sorted by time (columns) and ROI (row) of the combined matrix

mzroi_end contains the m/z information of each ROI in the combined matrix

time_end is the combined timetable

nRows is a row vector containing the number of scans processed in each data file

Parameters contains all selected processing parameters and data files for easier traceability.

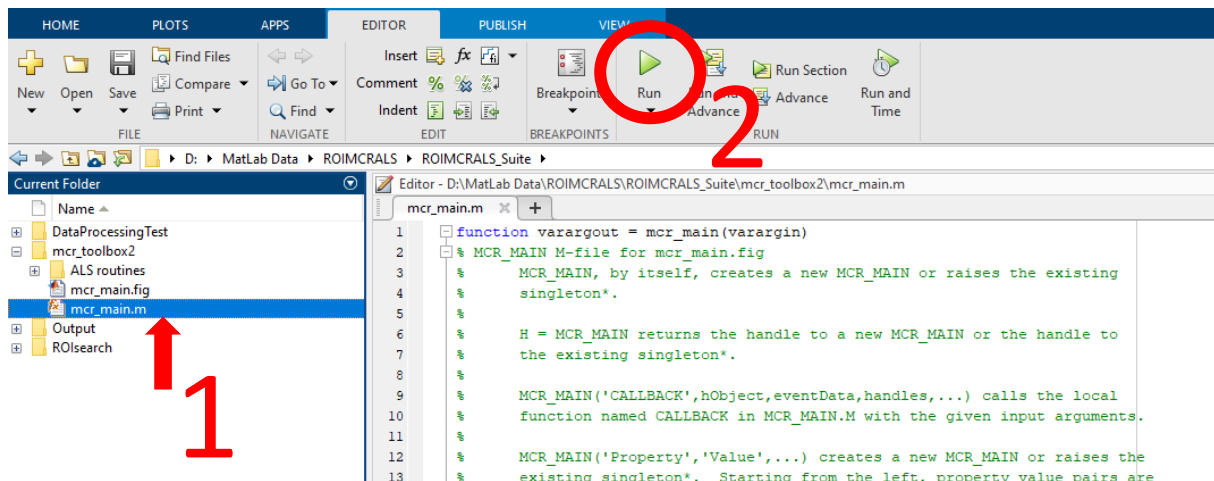
Caution: All output variables are required in further steps!
Renaming is possible if code input variable names are changed accordingly.

Save your progress by clicking "Save Workspace" in the home tab (optional, but recommended).

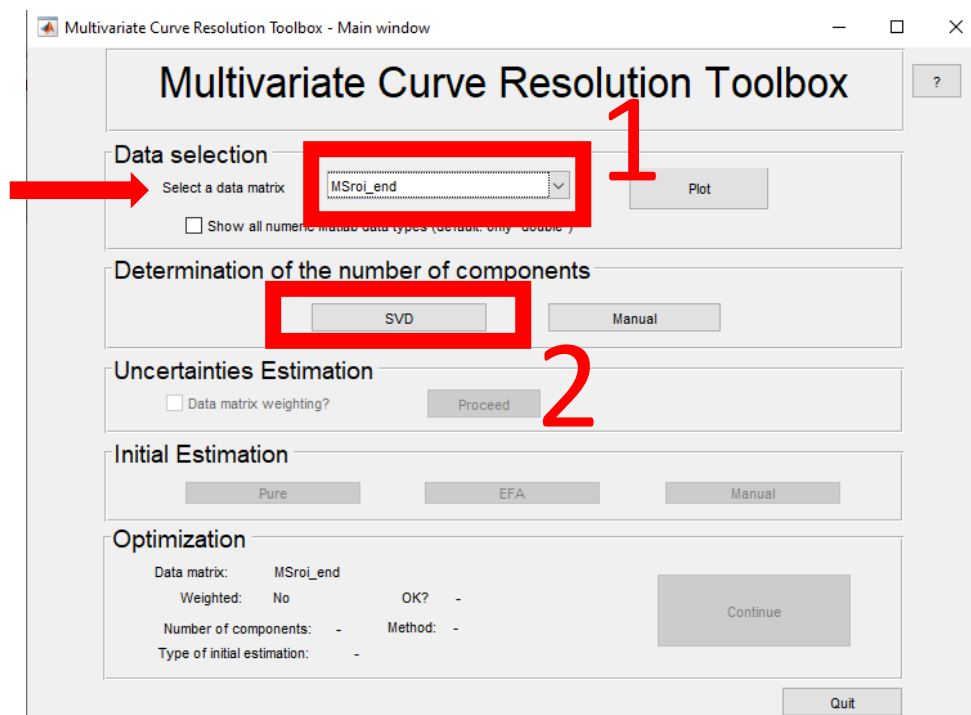
4. MCR-ALS

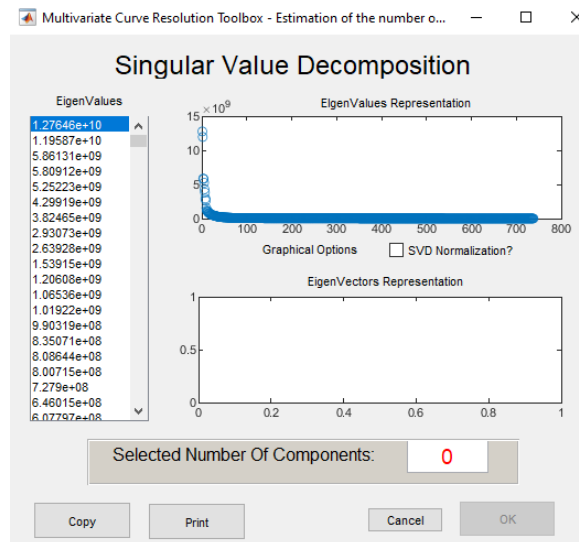
4.1. Running MCR-ALS

Go to the folder `mcr_toolbox2`, double click `mcr_main.m` to open it, then click Run in the Editor Tab.



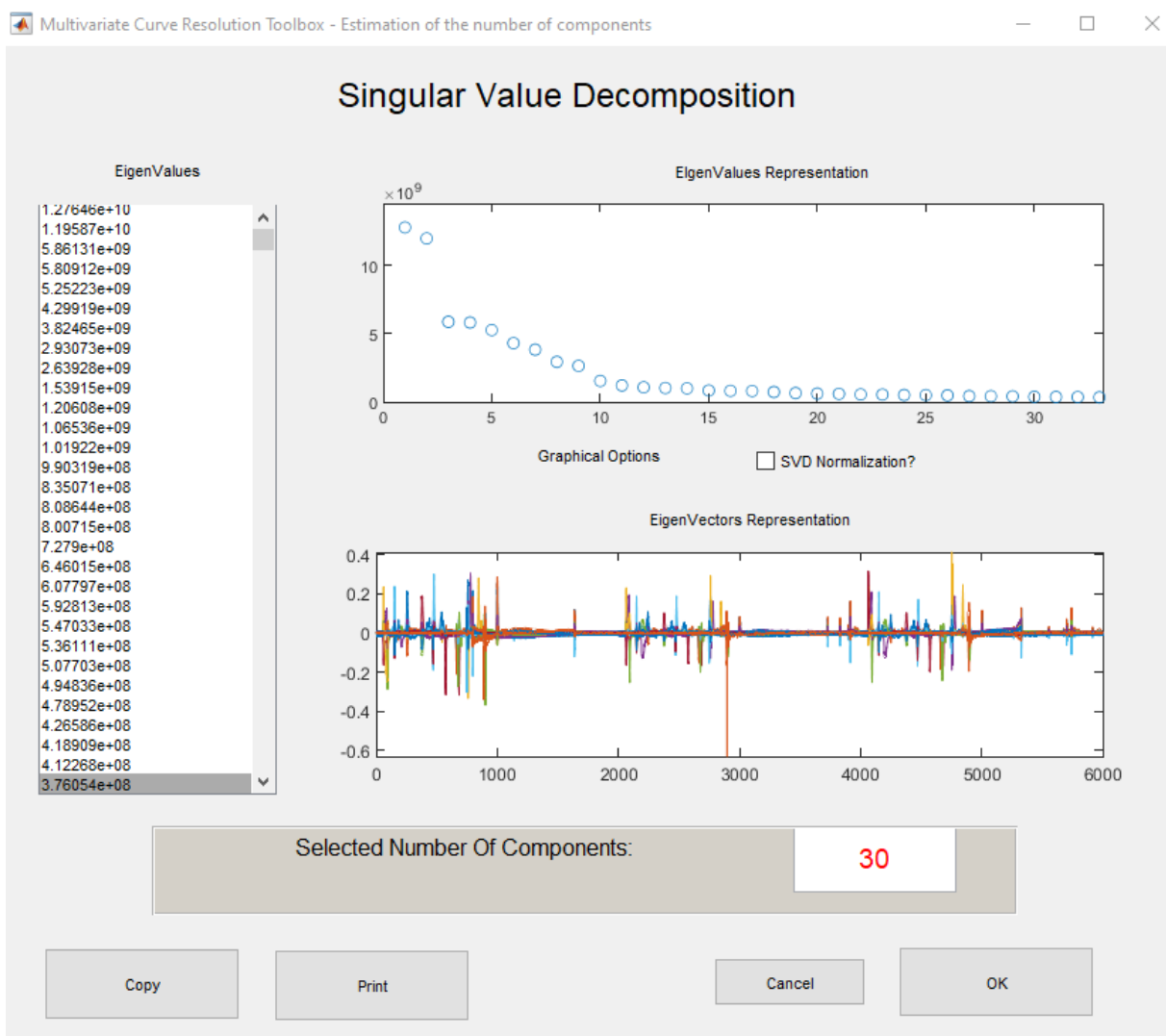
A graphical interface opens. In the dropdown Menu select the `MSroi_end` matrix to analyze. Proceed by running SVD.





Select the number of components by reviewing the EigenValues Representation. The window can be resized for easier visual inspection. By click and dragging in the graph you can zoom in.

When you have decided how many components you want to include in MCR-ALS, choose it by selecting the EigenValue in the list on the left side.



4.2. Initial Estimates

ALS requires initial estimates to run. Choose either Pure or EFA for initial estimation.

Multivariate Curve Resolution Toolbox - Main window

Multivariate Curve Resolution Toolbox

Data selection

Select a data matrix: MSroiaug123

☐ Show all numeric Matlab data types (default: only "double")

Determination of the number of components

Uncertainties Estimation

☐ Data matrix weighting?

Initial Estimation

Optimization

Data matrix: MSroiaug123
Weighted: No OK? -
Number of components: 30 Method: SVD
Type of initial estimation: -

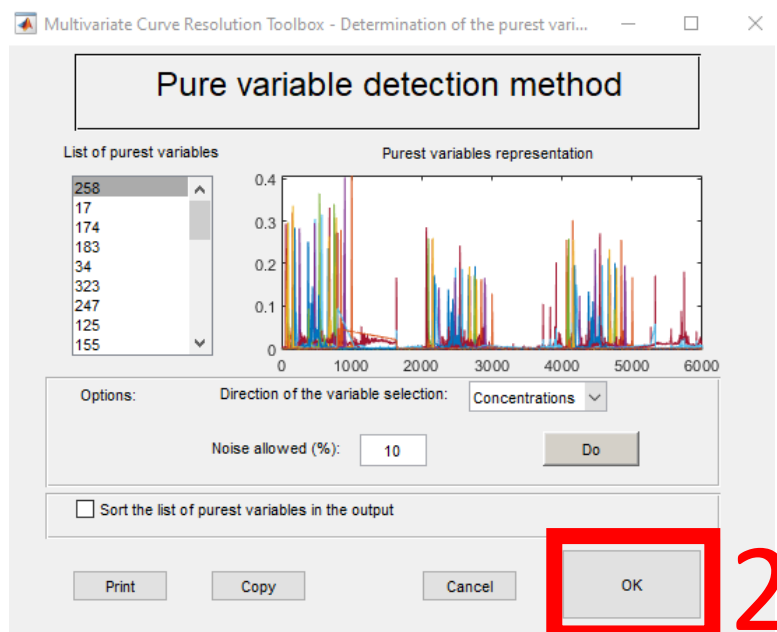
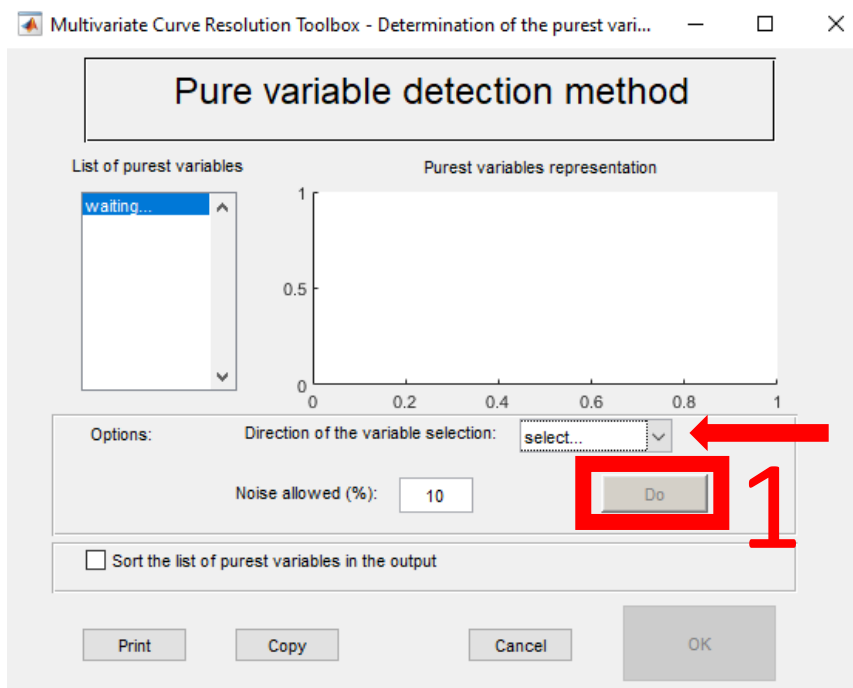
4.2.1.Pure

When selecting Pure estimates, the following Window opens

To proceed selecting concentration or spectra as initial estimate in from the drop-down menu.

Optionally change the allowed Noise.

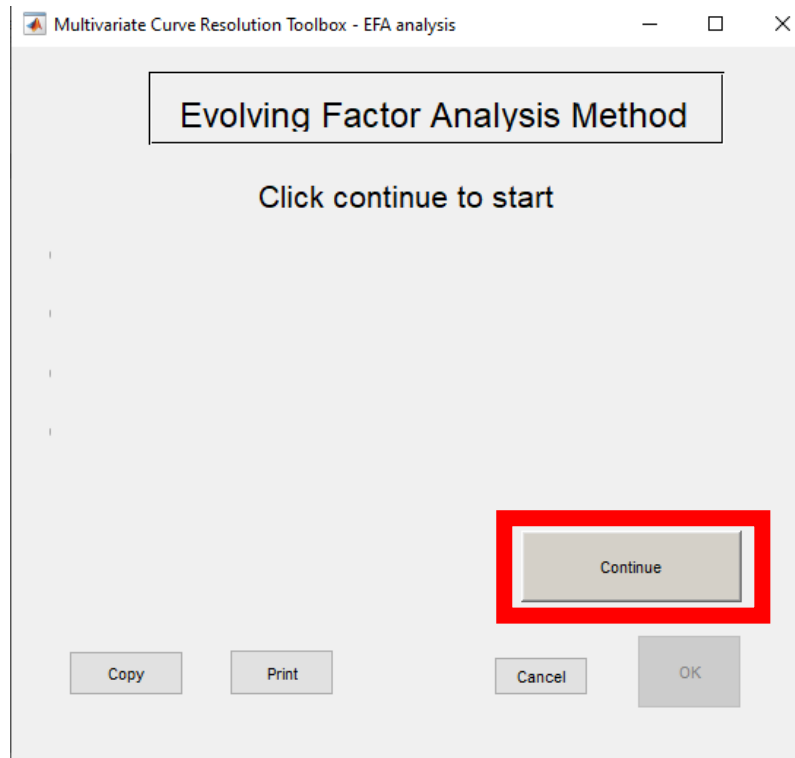
Click "Do", when the calculation is finished click OK.



4.2.2.EFA

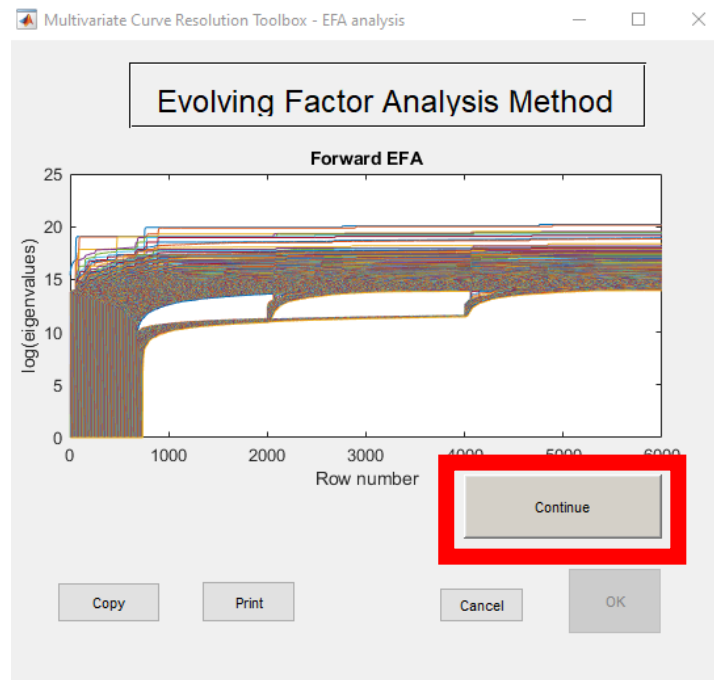
When selecting EFA the following window opens, click Continue to start forward EFA.

This step takes several minutes to calculate.

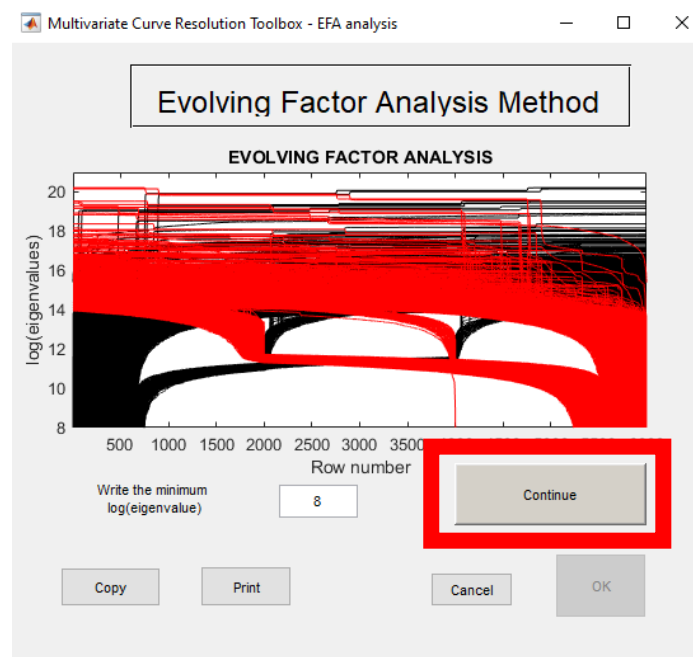


Click Continue Forward EFA is finished, to start Backward EFA.

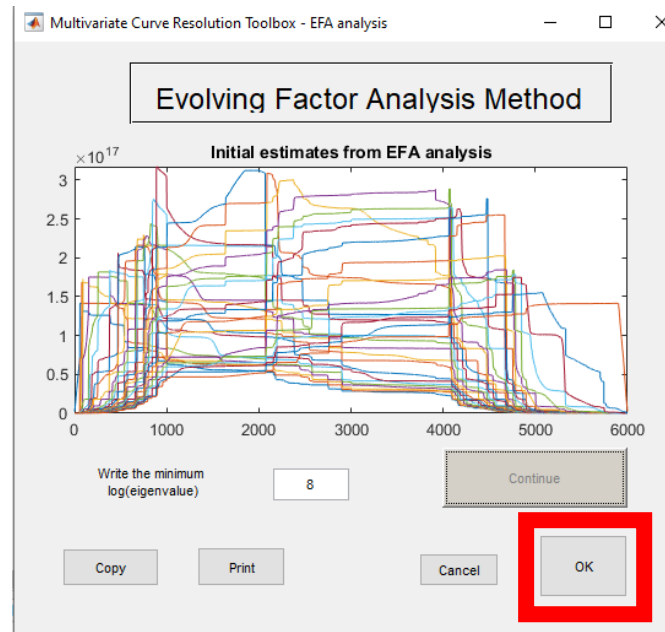
This step takes several minutes to calculate.



When finished select the minimum log(eigenvalue) and click Continue.



A Graph shows the Initial Estimates, then click OK



4.3. Optimization

Start optimization by clicking Continue in the MCR Toolbox main window.

A new window opens, and several plots will be drawn. This step takes some time.

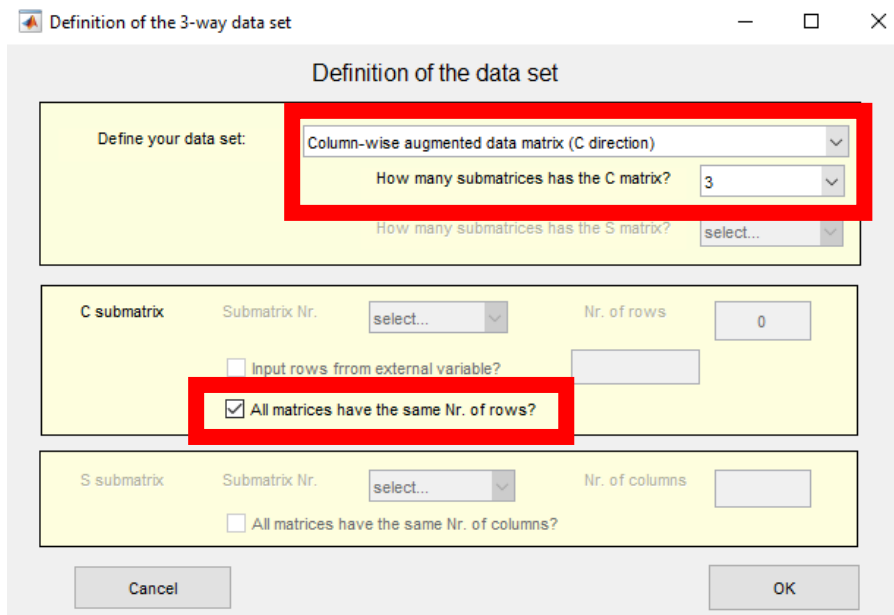


Select the total number of matrices that are combined in the data set, then click continue.

Select Column-wise augmented data matrix (C direction) in the drop down menu and check if the number of submatrices is set correctly.

In the C submatrix section define the Nr. of rows for each submatrix.

If all data files have been processed to use the same number of scans, (always the case for single analysis) you can check “All matrices have same Nr. of rows?”.



Definition of the 3-way data set

Define your data set:

Column-wise augmented data matrix (C direction)

How many submatrices has the C matrix? 3

How many submatrices has the S matrix? select...

C submatrix

Submatrix Nr. select... Nr. of rows 0

☐ Input rows from external variable?

☒ All matrices have the same Nr. of rows?

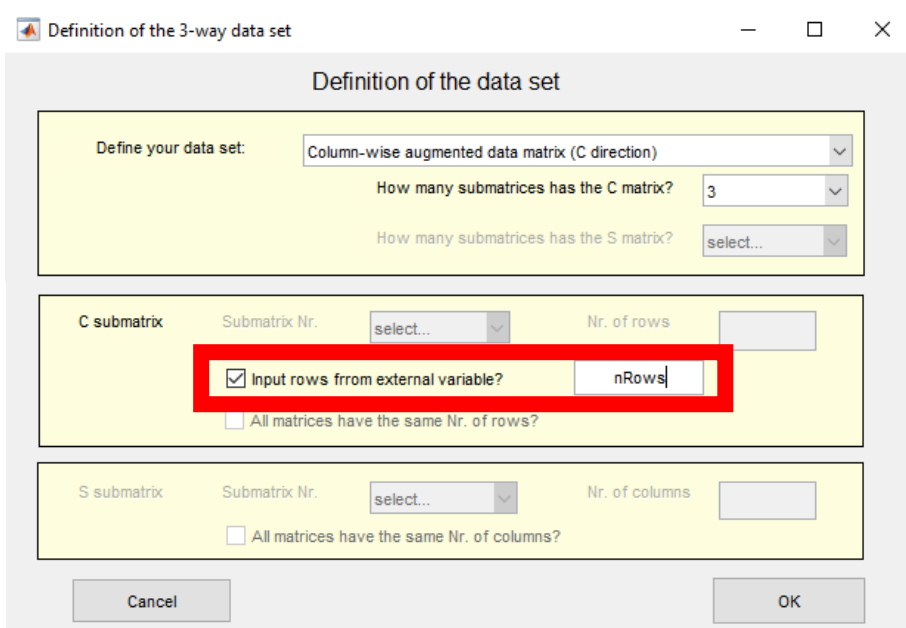
S submatrix

Submatrix Nr. select... Nr. of columns

☐ All matrices have the same Nr. of columns?

Cancel OK

If you use pairwise analysis with different number of scans for each group check “Input from external variable?” and type **nRows** in the field next to it.



Definition of the 3-way data set

Define your data set:

Column-wise augmented data matrix (C direction)

How many submatrices has the C matrix? 3

How many submatrices has the S matrix? select...

C submatrix

Submatrix Nr. select... Nr. of rows

☒ Input rows from external variable? nRows

☐ All matrices have the same Nr. of rows?

S submatrix

Submatrix Nr. select... Nr. of columns

☐ All matrices have the same Nr. of columns?

Cancel OK

Click OK to continue.

On the next two screens set the constraints for row mode (concentrations and multiple experiments) and column mode (spectra and single technique) respectively.

Click Continue.

Constraints: row mode (Concentrations)

Constraints: row mode (concentrations and multiple experiments)

Multiexperiment Analysis
Total Nr. of Row submatrices
3

Augmented Matrix
☒ Apply the same constraints to all submatrices?
Matrix Nr. Same constraints

Identification of species
Correspondence among the species in the experiments
☒ Default: all species in all experiments
Select a variable from the WS: select...

Constraints

Non-negativity

☒ Apply?
Implementation forced to zero
Nr. of species with non-negative profiles? 30
Enter a vector of positive profiles:

Unimodality

☒ Apply?
Implementation horizontal
Nr. of species with unimodal profiles? 30
Constraint tolerance: 1.1
Enter a vector of unimodal profiles:

Closure

☐ Apply?
Nr. of closure constraints to be included? select...
☐ Closure variable?

First closure constraint equal to:
Second closure constraint equal to:

First variable closure:
Second variable closure:

Closure condition: select...
Closure condition: select...

Which species are in 1st closure? ☐ All
Which species are in 2nd closure? ☐ All

Equality constraints

☐ Apply?
Select csel matrix: select a variable from the WS
Constraints are: select...

Advanced constraints

Multiway
Kinetic HM
Correlation

Reset
Continue

Constraints: column mode (spectra)

Constraints: column mode (spectra and single technique)

Constraints

Non-negativity

☒ Apply?
Implementation forced to zero
Nr. of species with non-negative profiles? 30
Enter a vector of positive profiles:

Unimodality

☐ Apply?
Implementation select...
Nr. of species with unimodal profiles? select...
Constraint tolerance:
Enter a vector of unimodal profiles:

Closure

☐ Apply?
Nr. of closure constraints to be included? select...
☐ Closure variable?

First closure constraint equal to:
Second closure constraint equal to:

First variable closure:
Second variable closure:

Closure condition: select...
Closure condition: select...

Which species are in 1st closure? ☐ All
Which species are in 2nd closure? ☐ All

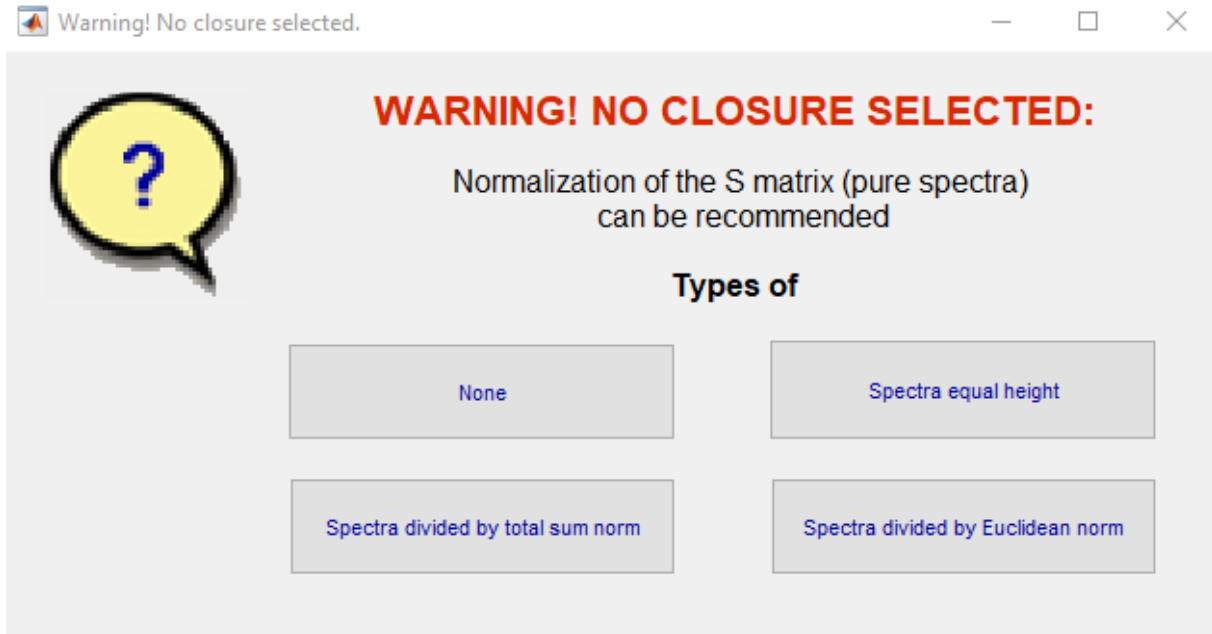
Equality constraints

☐ Apply?
Select csel matrix: select a variable from the WS
Constraints are: select...

Reset
Back
Continue

When no Closure constraint was selected a warning appears where you can choose to normalize the S matrix.

Pick one of the four options.



Warning! No closure selected.

WARNING! NO CLOSURE SELECTED:

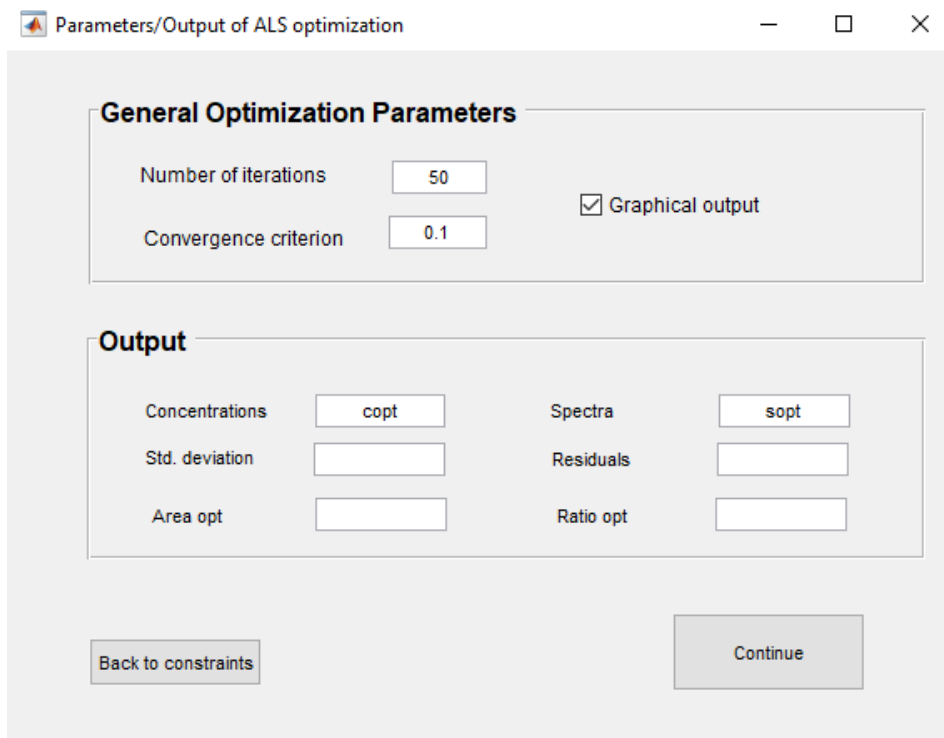
Normalization of the S matrix (pure spectra)
can be recommended

Types of

| | |
|-----------------------------------|-----------------------------------|
| None | Spectra equal height |
| Spectra divided by total sum norm | Spectra divided by Euclidean norm |

Lastly set general optimization settings and select the Output of ALS optimization by typing the name you want the matrices to have, into the corresponding field.

Then click continue.



Parameters/Output of ALS optimization

General Optimization Parameters

Number of iterations: 50

Convergence criterion: 0.1

☒ Graphical output

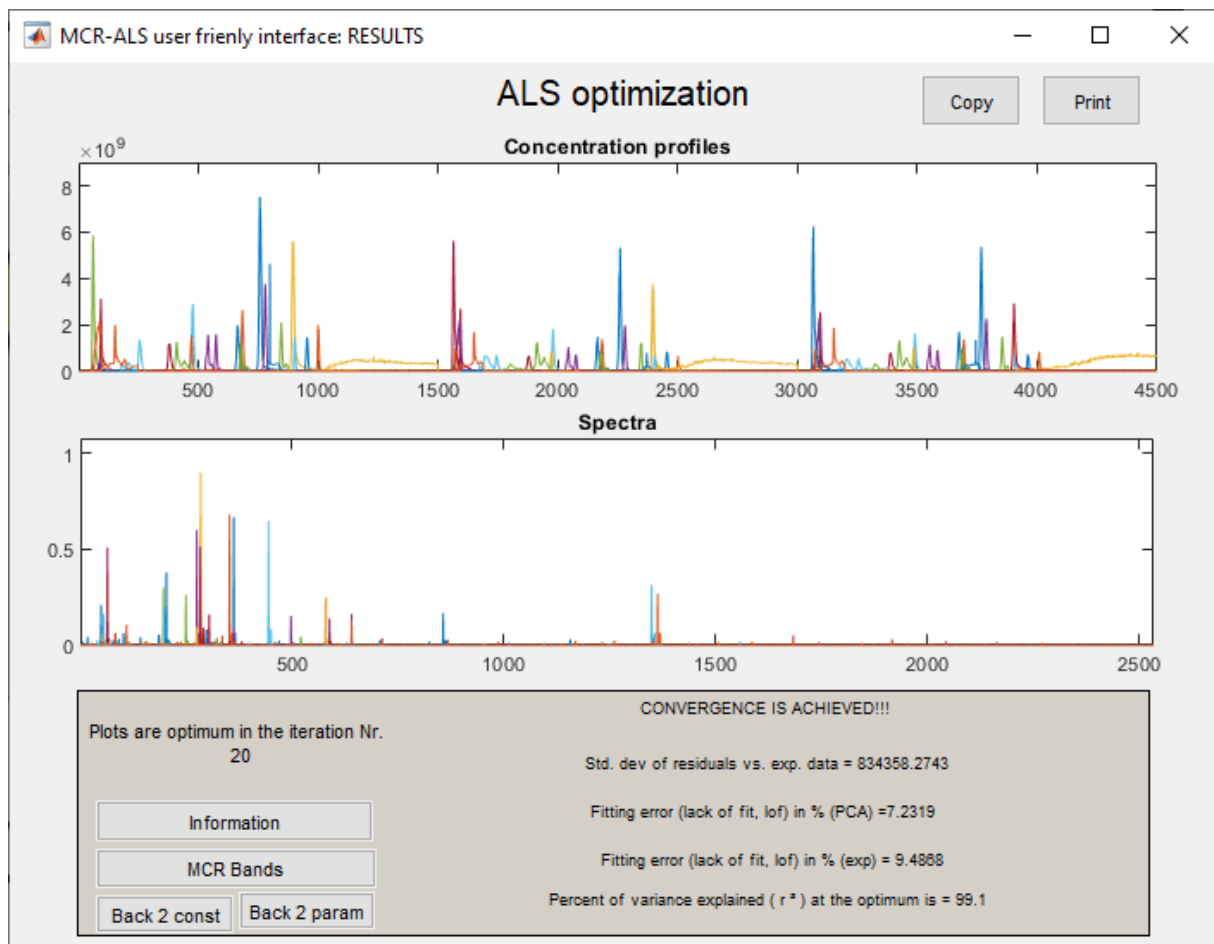
Output

| | | | |
|----------------|------|-----------|------|
| Concentrations | copt | Spectra | sopt |
| Std. deviation | | Residuals | |
| Area opt | | Ratio opt | |

Back to constraints

Continue

After ALS iteration achieved convergence, the results are shown.



All matrices you defined in the previous screen, are generated in the Workspace tab and can be used for further calculations or analysis.

If no convergence is achieved you can go back to constraints [Back to const] or back to optimization parameters [Back to param] to change settings, without redoing the whole process.

Caution: `mcr_main.m` can sometimes produce an error after using the Back 2 const option and changing constraint parameters. In this case delete all workspace variables except `MSroi_end`, `mzroi_end`, `time_end`, `nRows` and `Parameters` and start over.

Your Workspace now contains new additional variables `copt`, `mcr_als` and `sopt`.

| Workspace | | |
|------------|-----------------------|--|
| Name ▲ | Value | |
| copt | 12000x25 double | |
| mcr_als | 1x1 struct | |
| MSroi_end | 12000x621 double | |
| mzroi_end | 1x621 double | |
| nRows | [3000,3000,3000,3000] | |
| Parameters | 1x1 struct | |
| sopt | 25x621 double | |
| time_end | 12000x1 double | |

copt is the MCR-ALS result containing concentration profiles for each compound (column wise)

mcr_als is a data construct containing all parameters, estimates etc. used in the MCR-ALS analysis

sopt is the MCR-ALS result containing spectra data for each compound (row wise)

5. Result Evaluation

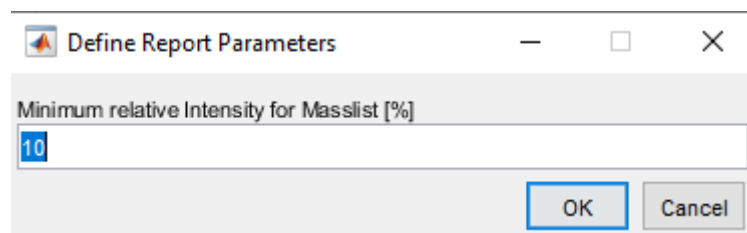
5.1. Report Generation and Component Plots

Use the command

```
ReportTable = MCRout(copt,sopt,mzroi_end,time_end,nRows,Parameters)
```

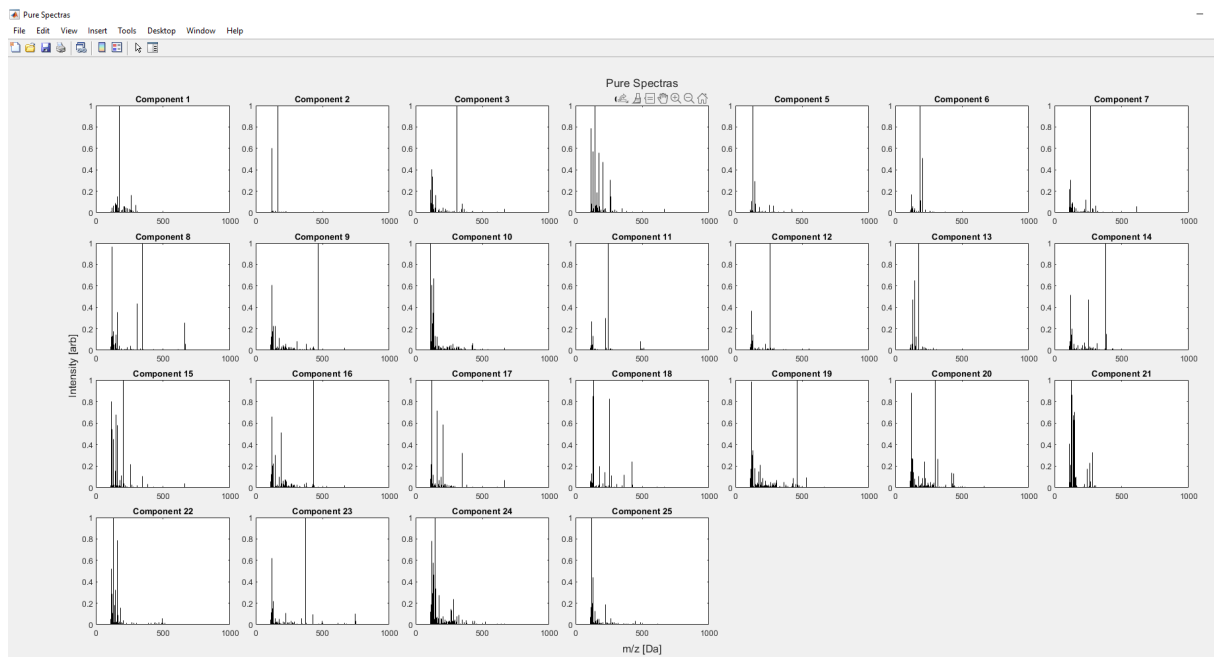
in the command window to generate a Report and plot the average concentration profile and spectra for each pure component.

You can change the minimum relative intensity threshold for m/z values to be included in the component mass list.

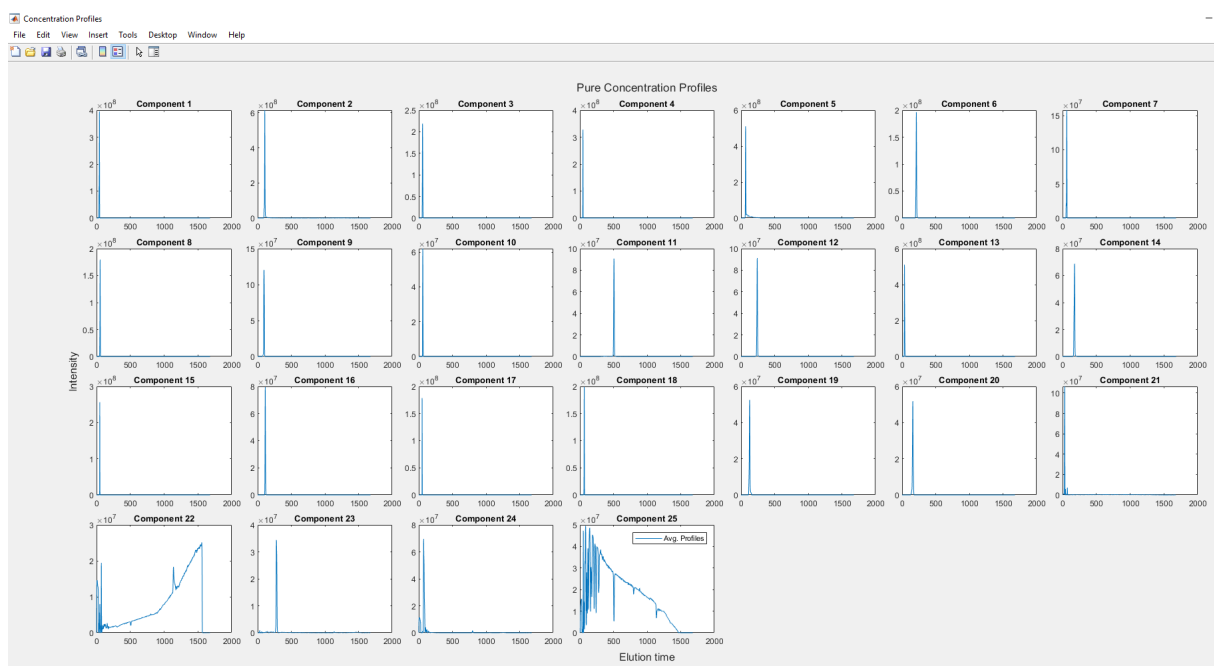


The following output is generated:

MS spectra for every pure component:



Average Concentration Profile for every pure component:



If pair-wise analysis is performed, average concentration profiles for each group will be plotted.

The ReportTable appears in the workspace and can be opened by double-clicking.

| ReportTable | | | | | | | | | | | | | | |
|-------------|----------------|----|------------|-------------|---------------|--------------------------|--------------|------------------|---|----------|---------------|----------------------|---------------|--|
| 25x14 Table | | | | | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | |
| Component | Group | rt | mz | AvgPeakArea | AvgPeakHeight | ConcProfiles | PureSpectrum | SampleArea | STDArea | Masslist | RTRange | AvgConcProfiles | SampleHeights | |
| 1 | "Component 1" | 1 | 42.9906 | 175.1190 | 3.8656e+09 | 4.1391e+08,3000x4 double | 621x1 double | [4.5574e+09,3... | 5.2789e+08 [162.0776;175.1190;265.1117] | 1.1643 | 3000x1 double | [4.5735e+08,3.899... | | |
| 2 | "Component 2" | 1 | 106.1885 | 166.0863 | 8.5193e+09 | 6.3529e+08,3000x4 double | 621x1 double | [9.6046e+09,6... | 1.7308e+09 [120.0730;166.0863] | 2.2880 | 3000x1 double | [6.9862e+08,5.207... | | |
| 3 | "Component 3" | 1 | 56.5916 | 308.0910 | 2.7292e+09 | 2.6808e+08,3000x4 double | 621x1 double | [2.5955e+09,2... | 3.7256e+08 [113.0349;123.0573;233.1193] | 5.2183 | 3000x1 double | [2.6241e+08,2.796... | | |
| 4 | "Component 4" | 1 | 45.4914 | 148.0608 | 2.1189e+09 | 3.4579e+08,3000x4 double | 621x1 double | [2.1849e+09,2... | 1.0274e+08 [116.0706;130.0634;147.1121] | 1.6370 | 3000x1 double | [3.6136e+08,3.595... | | |
| 5 | "Component 5" | 1 | 67.9939 | 132.1019 | 7.8059e+09 | 5.4382e+08,3000x4 double | 621x1 double | [7.9304e+09,6... | 2.0622e+09 [123.0573;132.1019;144.0714] | 2.3241 | 3000x1 double | [5.1545e+08,4.943... | | |
| 6 | "Component 6" | 1 | 210.8428 | 188.0739 | 3.2132e+09 | 2.3837e+08,3000x4 double | 621x1 double | [3.1747e+09,2... | 7.3100e+08 [123.0573;188.0739;189.0967] | 5.8610 | 3000x1 double | [2.3501e+08,1.881... | | |
| 7 | "Component 7" | 1 | 60.6430 | 268.1055 | 1.7684e+09 | 1.9964e+08,3000x4 double | 621x1 double | [2.9184e+09,1... | 8.7652e+08 [113.0349;123.0573;233.1193] | 9.8206 | 3000x1 double | [3.1411e+08,1.700... | | |
| 8 | "Component 8" | 1 | 52.1538 | 348.0704 | 1.7047e+09 | 2.0965e+08,3000x4 double | 621x1 double | [1.8609e+09,1... | 6.3670e+08 [116.0706;123.0573;124.0491] | 3.4014 | 3000x1 double | [2.2465e+08,1.578... | | |
| 9 | "Component 9" | 1 | 95.9628 | 466.1088 | 2.2618e+09 | 1.2810e+08,3000x4 double | 621x1 double | [2.0700e+09,1... | 4.0871e+08 [116.0706;123.0573;124.0491] | 7.0462 | 3000x1 double | [9.8762e+07,1.265... | | |
| 10 | "Component 10" | 1 | 61.2325 | 113.0349 | 6.4743e+08 | 6.6810e+07,3000x4 double | 621x1 double | [6.6912e+07,9... | 5.3039e+08 [113.0349;123.0573;124.0491] | 2.9053 | 3000x1 double | [1.2888e+07,8.507... | | |
| 11 | "Component 11" | 1 | 506.7563 | 245.0947 | 1.5691e+09 | 1.1190e+08,3000x4 double | 621x1 double | [1.5044e+09,1... | 2.3381e+08 [123.0573;130.0634;227.0946] | 6.4840 | 3000x1 double | [1.0280e+08,1.043... | | |
| 12 | "Component 12" | 1 | 240.5110 | 261.1394 | 1.5213e+09 | 1.1248e+08,3000x4 double | 621x1 double | [1.4222e+09,1... | 3.3996e+08 [123.0573;130.0634;261.1394] | 5.9170 | 3000x1 double | [1.0672e+08,9.405... | | |
| 13 | "Component 13" | 1 | 35.0562 | 175.1190 | 5.6663e+09 | 5.1562e+08,3000x4 double | 621x1 double | [5.5967e+09,5... | 3.1261e+08 [130.0634;147.1121;156.0793] | 1.1349 | 3000x1 double | [5.2509e+08,5.141... | | |
| 14 | "Component 14" | 1 | 175.5380 | 384.1147 | 1.3604e+09 | 9.6026e+07,3000x4 double | 621x1 double | [8.6247e+08,1... | 3.3510e+08 [123.0573;124.0491;130.0634] | 8.8430 | 3000x1 double | [6.3244e+07,1.107... | | |
| 15 | "Component 15" | 1 | 48.1418 | 204.1238 | 1.3275e+09 | 2.8717e+08,3000x4 double | 621x1 double | [1.3468e+09,1... | 2.1612e+08 12x1 double | 1.1374 | 3000x1 double | [2.9856e+08,2.703... | | |
| 16 | "Component 16" | 1 | 112.5445 | 434.1230 | 1.0044e+09 | 8.5822e+07,3000x4 double | 621x1 double | [1.0225e+09,9... | 2.4790e+08 [116.0706;123.0573;124.0491] | 2.9240 | 3000x1 double | [8.7122e+07,7.947... | | |
| 17 | "Component 17" | 1 | 50.0051 | 123.0573 | 9.2503e+08 | 1.9934e+08,3000x4 double | 621x1 double | [1.0691e+09,7... | 3.5461e+08 [116.0706;118.0863;123.0573] | 1.7255 | 3000x1 double | [2.2355e+08,1.746... | | |
| 18 | "Component 18" | 1 | 64.3872 | 137.0459 | 1.6548e+09 | 2.1733e+08,3000x4 double | 621x1 double | [1.4435e+09,1... | 4.9984e+08 [123.0573;132.1019;137.0459] | 2.8416 | 3000x1 double | [1.8974e+08,2.261... | | |
| 19 | "Component 19" | 1 | 127.8108 | 464.0814 | 9.8780e+08 | 6.1480e+07,3000x4 double | 621x1 double | [5.7198e+08,1... | 3.0745e+08 [116.0706;123.0573;124.0491] | 5.8860 | 3000x1 double | [4.4168e+07,6.423... | | |
| 20 | "Component 20" | 1 | 158.0045 | 298.0988 | 1.0684e+09 | 6.2084e+07,3000x4 double | 621x1 double | [1.0914e+09,8... | 2.7978e+08 13x1 double | 6.4810 | 3000x1 double | [6.2533e+07,5.441... | | |
| 21 | "Component 21" | 1 | 31.8942 | 128.0194 | 1.4780e+09 | 1.0795e+08,3000x4 double | 621x1 double | [8.1583e+08,2... | 7.0504e+08 16x1 double | 0.5843 | 3000x1 double | [9.4793e+07,1.074... | | |
| 22 | "Component 22" | 1 | 1.5602e+03 | 130.0634 | 2.1662e+10 | 2.6532e+07,3000x4 double | 621x1 double | [1.3785e+10,2... | 7.0796e+09 [116.0706;120.0730;123.0573] | 6.2600 | 3000x1 double | [1.9878e+07,3.240... | | |
| 23 | "Component 23" | 1 | 281.9030 | 374.1255 | 9.1462e+08 | 5.8557e+07,3000x4 double | 621x1 double | [9.6722e+08,5... | 4.1789e+08 [116.0706;123.0573;124.0491] | 8.2780 | 3000x1 double | [5.8925e+07,4.431... | | |
| 24 | "Component 24" | 1 | 72.7830 | 144.0714 | 2.1727e+09 | 6.9870e+07,3000x4 double | 621x1 double | [1.5788e+09,2... | 4.1422e+08 14x1 double | 0.5845 | 3000x1 double | [6.2758e+07,6.945... | | |
| 25 | "Component 25" | 1 | 67.9810 | 123.0573 | 4.8815e+10 | 6.7625e+07,3000x4 double | 621x1 double | [4.5648e+10,4... | 3.7295e+09 [116.0706;123.0573;124.0491] | 35.9905 | 3000x1 double | [6.1072e+07,5.221... | | |

Columns contain the following information:

Component: Component Number

Group: group assignment if pair-wise analysis is performed

rt: average elution time

mz: highest intensity m/z value in component spectra (Main MZ)

AvgPeakArea: Average peak area

AvgPeakHeight: Average peak height

ConcProfiles: Concentration profile for each data file, sorted column wise

PureSpectrum: Pure Spectrum

SampleArea: row vector containing peak area for each data file

STDArea: standard deviation of peak area between samples

Masslist: Row vector containing every m/z value over the specified threshold

RTRange: peak width of the average concentration profile

AvgConcProfiles: Average Concentration Profile

SampleHeights: row vector containing peak height for each data file

FoldChange: Fold change between groups, calculated by AvgArea1/ AvgArea2-1 and vice versa. Only for pair-wise analysis

pValue: p-value calculated by two-sample t-test

SignificantDifference: Yes, if null hypothesis is rejected at 5% significance level

5.2. Compare MCR-ALS result to raw data

Use the command

```
[EICs,ReportTable]=CheckConcProfiles(ReportTable,Parameters)
```

to compare the calculated pure component concentration profiles to raw file EICs.

Select a file in the file browser. If pair-wise analysis is used, you must select representative files for both groups.

The function automatically loads the file and uses the main MZ of every component to extract the corresponding EIC. This step can take several minutes, depending on the selected files.

The average concentration profile and corresponding EIC is plotted for every component and can be reviewed visually.

Additionally, the correlation between average concentration profile and EIC is calculated and stored in a new column **Correlation** in **ReportTable**. The column **ProfileEvaluation** is a quick aid to determine if a calculated component is a real component or noise component.

Classification is done by evaluating the calculated correlation. A high correlation indicates a pure component, low correlation indicates a noise component or wrongly calculated concentration profile.

Pure Component: Correlation > 0.7

Uncertain, Check Profile vs EIC Plot: $0.7 \geq \text{Correlation} \geq 0.3$

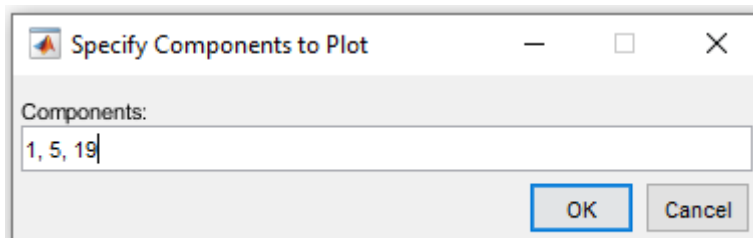
Noise Component: Correlation < 0.3

Those boundaries are arbitrarily set, and should always be manually verified.

To check individual plots, use the command

```
PlotProfileCheck(EICs,ReportTable)
```

Specify which components to plot, by typing in the numbers.



For each specified component a new plot will open.

5.3. Compare Component mass list to YMDB search results

Use the command

```
ReportTable=evalDatabaseQuery(ReportTable)
```

to evaluate a YMDB MS1 search result, stored as delimited text file or spreadsheet.

This function compares all m/z values in the component mass lists to the search result and stores matched results in **ReportTable** in the column **DataBankResult**.

Caution: This function currently works only with YMDB search results or .csv files with the same data structure.
If different data bank searches are desired, contact me at Adrian.Haun@hs-aalen.de

5.4. Compare Component mass list to suspect list

Use the command

```
ReportTable = compare2suspects(ReportTable,Parameters)
```

to compare found m/z values in the component mass lists to a list of suspect targets.

First define the acceptable mass difference in Da, either use the same mass error used in the ROI search or set it manually.

Then select the excel file containing the suspect target mass list.

Caution: This function currently works only with suspect lists saved as .xlsx with the sheet name **Suspect Export**.

5.5. Find MS/MS Spectra

Use the command

```
ReportTable = extractMSMS(ReportTable,Parameters)
```

to search for MS2 spectra for the m/z values in every component mass list.

First define the acceptable mass difference in Da, either use the same mass error used in the ROI search or set it manually.

Then select the data files to search.

The function searches for MS2 spectra for precursor masses inside the mass difference window and only in the retention time range of that component.

The output is stored in **ReportTable** in the column **MSMS**.