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:

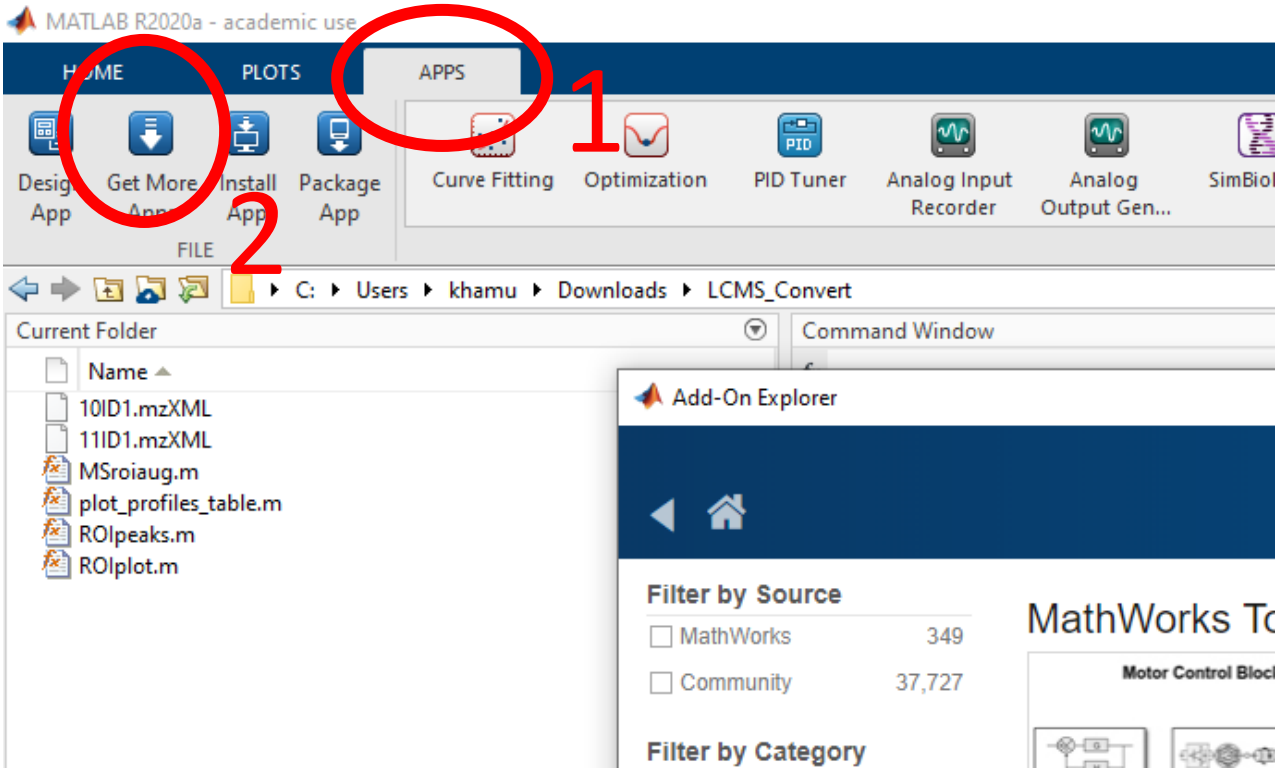
https://www.hs-aalen.de/uploads/mediapool/media/file/5814/MatLab.pdf

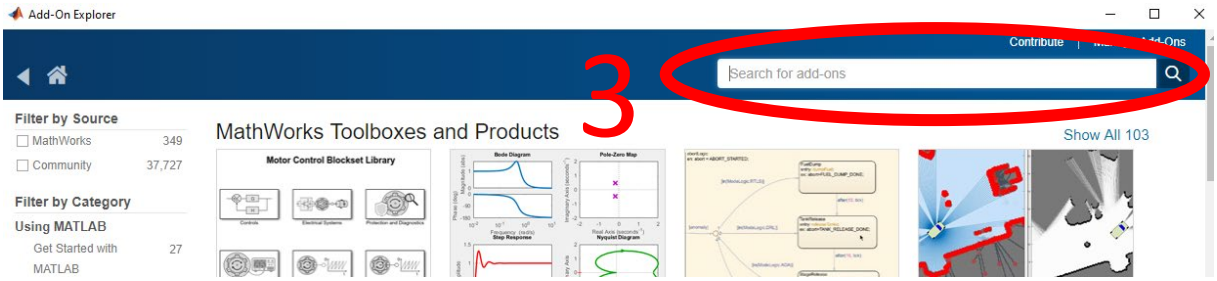
1.2. Installation of required MatLab Add-Ons

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 ROIprocess2 here:
https://github.com/AdrianHaun/ROIprocess

Download MCR-ALS Toolbox 2.0 here:
https://mcrals.wordpress.com/download/mcr-als-2-0-toolbox/

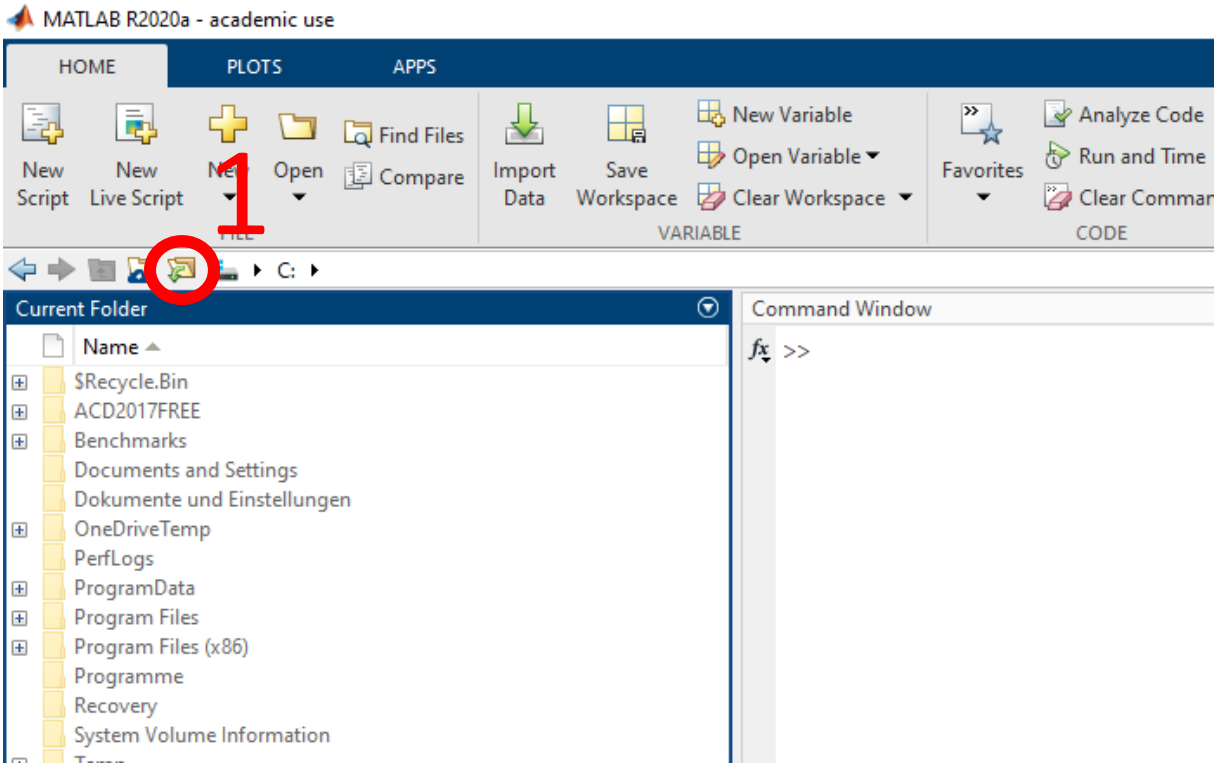
Use a file unpacking software (eg. WinRAR or 7zip) to unpack both .zip files to a folder on your PC and remember the location.

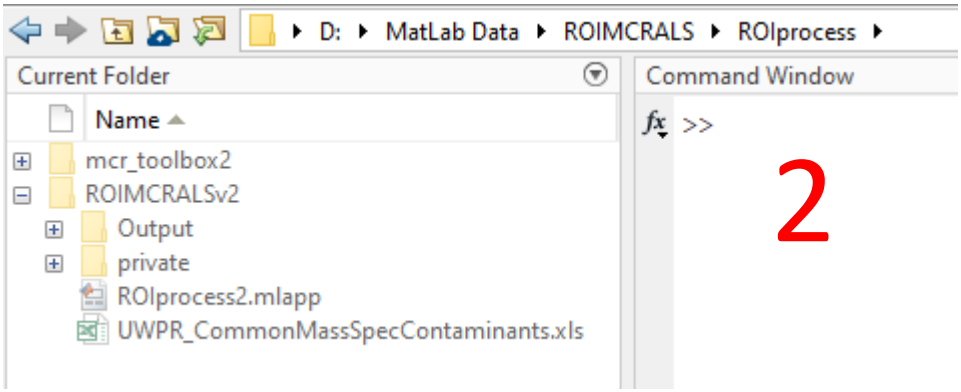
Start MatLab, on the left side click "Browse for folder" and open the folder containing the previously unpacked files.

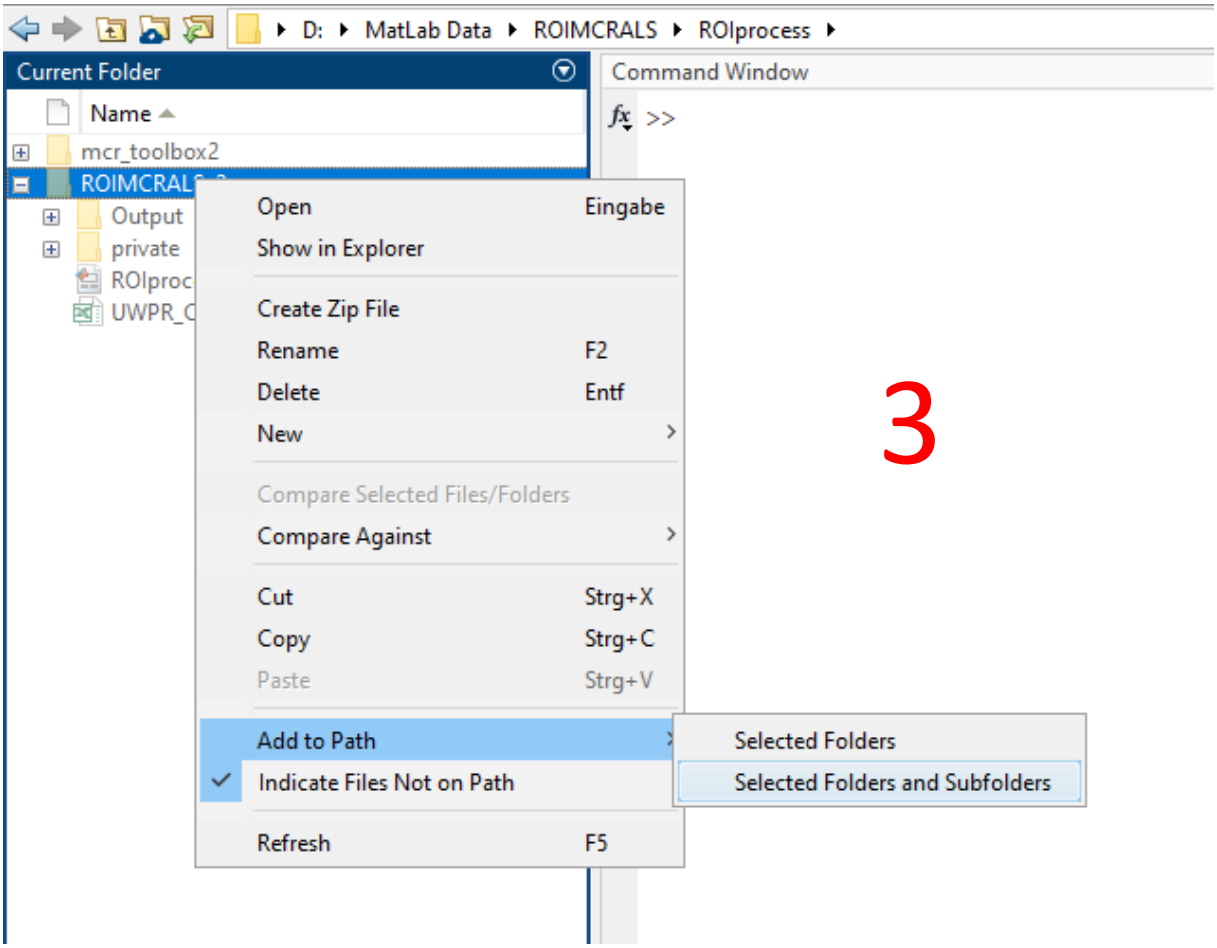
The folders mcr_toolbox2 and ROIMCRALSv2 must be in the current folder tab.

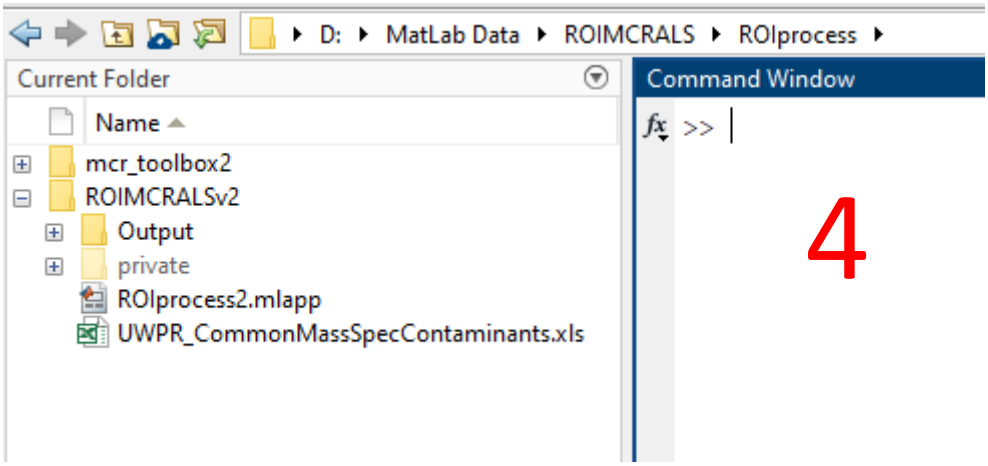
If they are grayed out right click on the parent folder und select Add to path -> Selected Folders and Subfolders

Note: sub folders called private contain support functions that can only be called by the main function. These folders are always grayed out.


MATLAB R2020a - academic use
HOME
PLOTS
APPS
New Script
New Live Script
Open
Find Files
Compare
Import Data
Save Workspace
New Variable
Open Variable
Clear Workspace
Favorites
Analyze Code
Run and Time
VARIABLE
CODE
1
C:
Current Folder
Name
$Recycle.Bin
ACD2017FREE
Benchmarks
Documents and Settings
Dokumente und Einstellungen
OneDriveTemp
PerfLogs
ProgramData
Program Files
Program Files (x86)
Programme
Recovery
System Volume Information
Command Window
>>

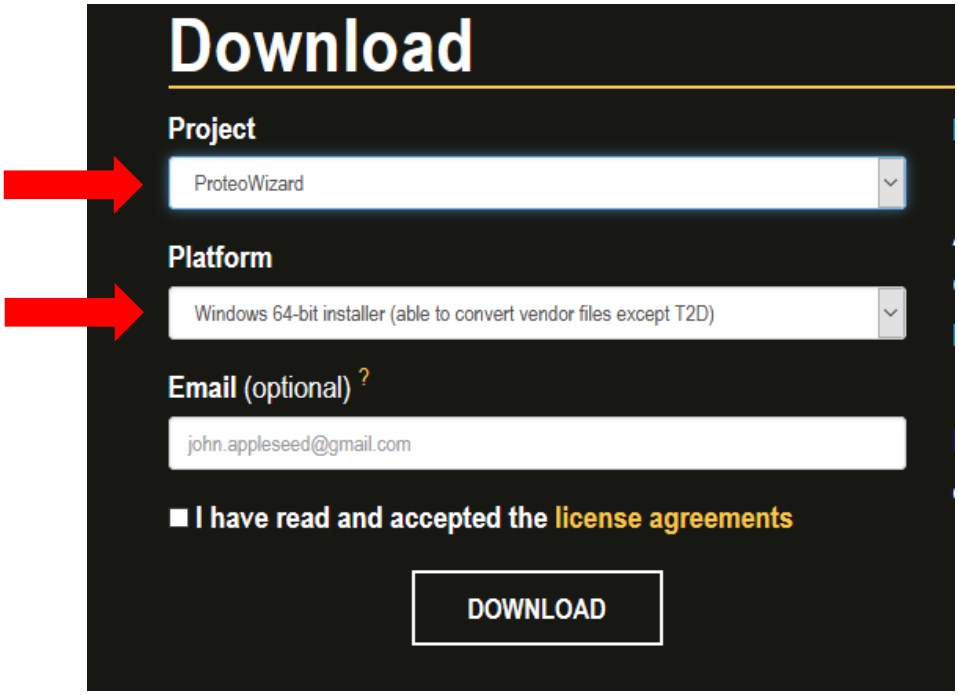

D:
MatLab Data
ROIMCRALS
ROIprocess
Current Folder
Name
mcr_toolbox2
ROIMCRALSv2
Output
private
ROIprocess2.mlapp
UWPR_CommonMassSpecContaminants.xls
Command Window
>>
2


D: ▸ MatLab Data ▸ ROIMCRALS ▸ ROIprocess ▸
Current Folder
Name
mcr_toolbox2
ROIMCRALS
Output
private
ROIproc
UWPR_C
Command Window
>>
Open Eingabe
Show in Explorer
Create Zip File
Rename F2
Delete Entf
New
Compare Selected Files/Folders
Compare Against
Cut Strg+X
Copy Strg+C
Paste Strg+V
Add to Path
Indicate Files Not on Path
Refresh F5
Selected Folders
Selected Folders and Subfolders
3


D: ▸ MatLab Data ▸ ROIMCRALS ▸ ROIprocess ▸
Current Folder
Name
mcr_toolbox2
ROIMCRALSv2
Output
private
ROIprocess2.mlapp
UWPR_CommonMassSpecContaminants.xls
Command Window
>>
4

1.4. Download ProteoWizard

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



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

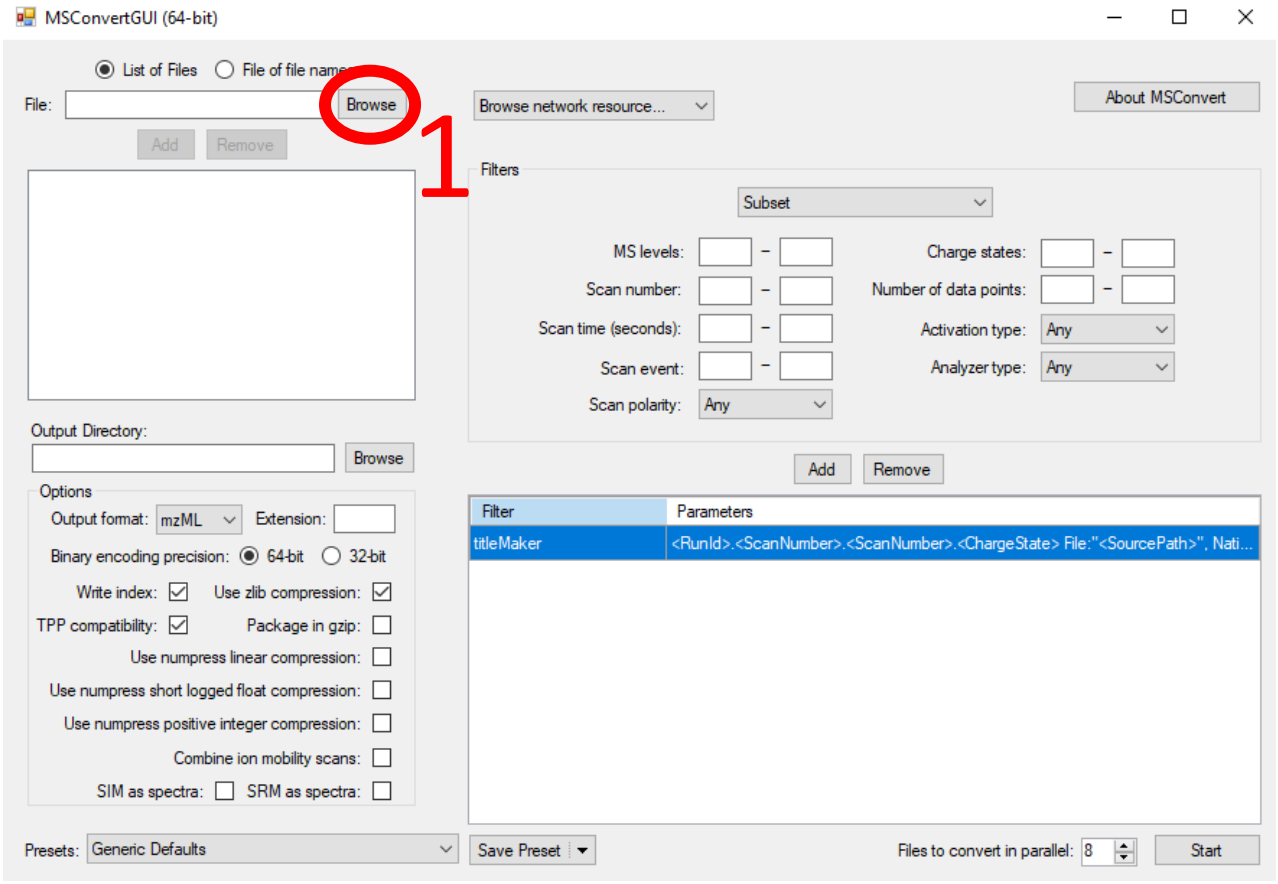
https://www.howtogeek.com/howto/21726/how-do-i-know-if-im-running-32-bit-or-64-bit-windows-answers/

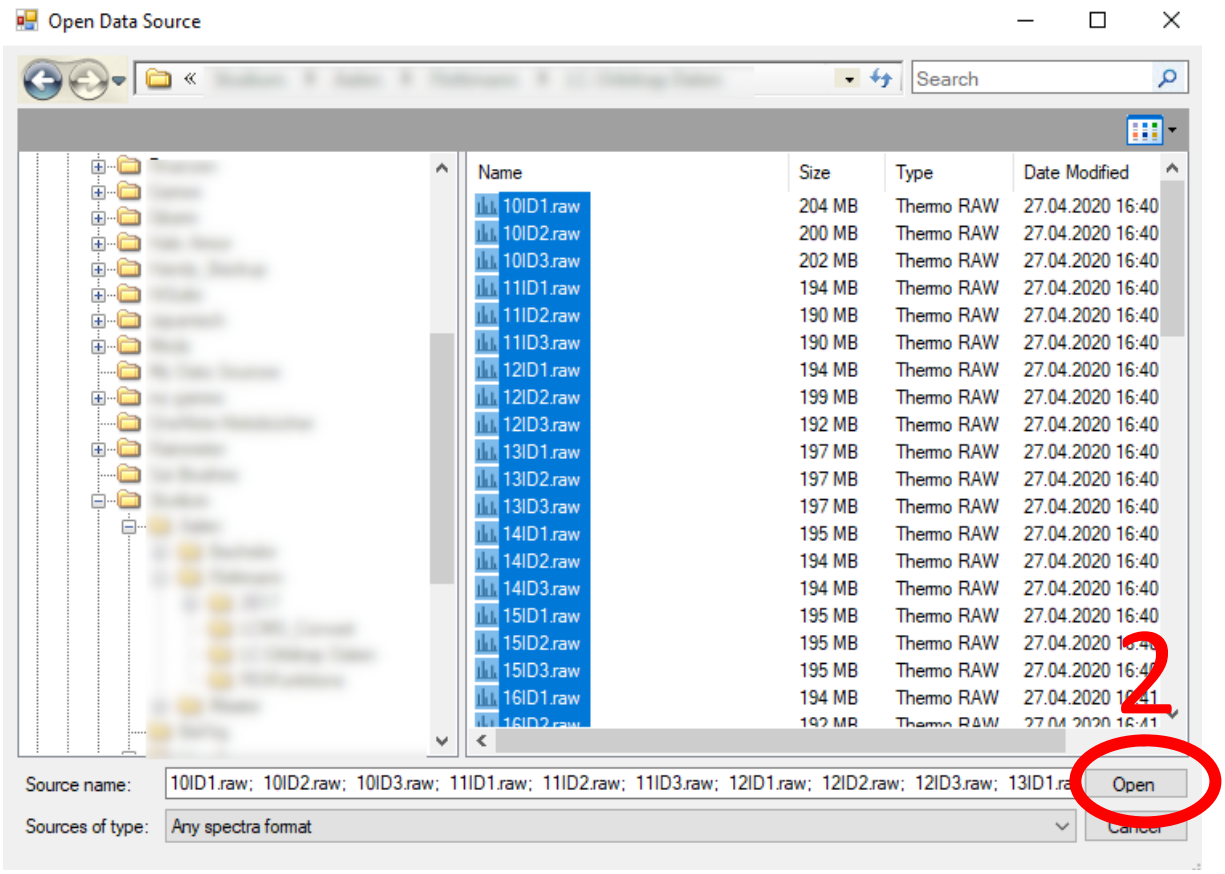
After download run the installer to install MSConvert.

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.

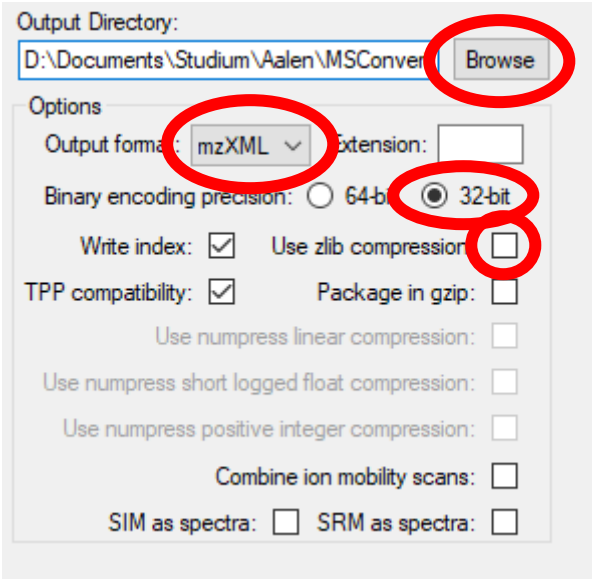




Select an output directory by clicking the Browse bottom and selecting a folder.

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.

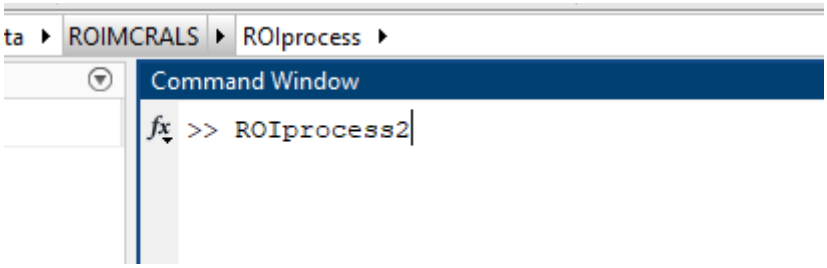
3. ROI analysis

3.1. Start ROIprocess2 and select test file

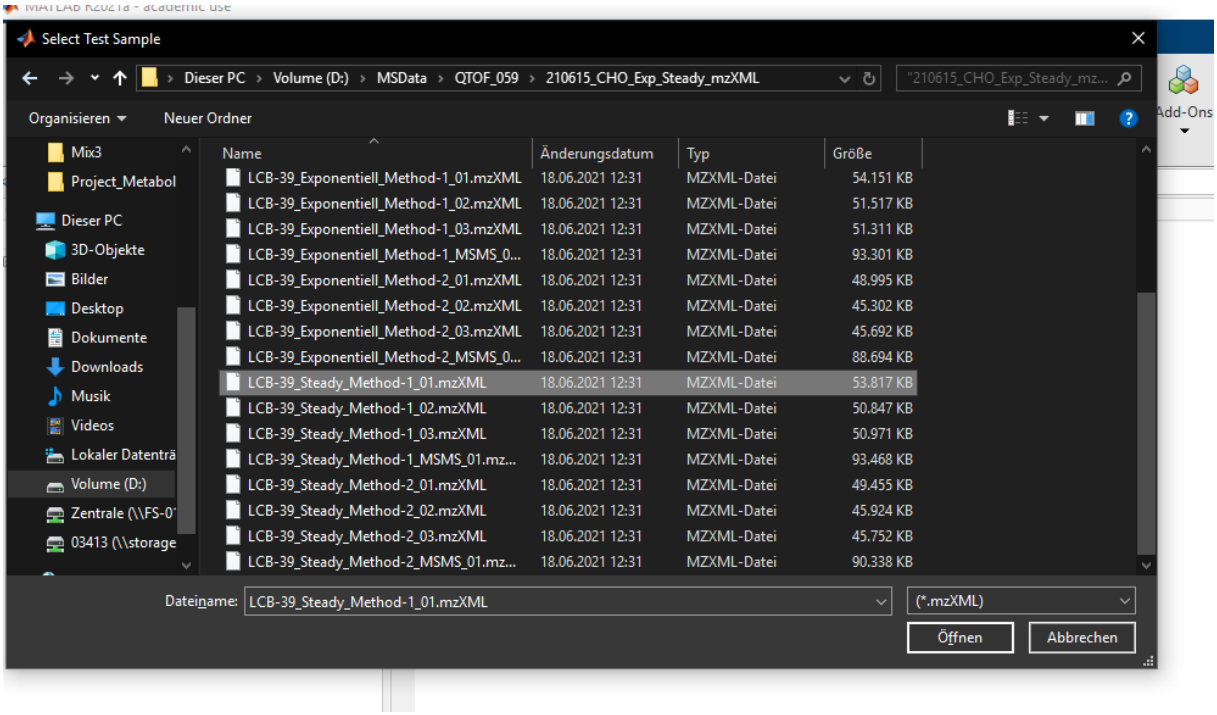
Type or copy the command

ROIprocess2

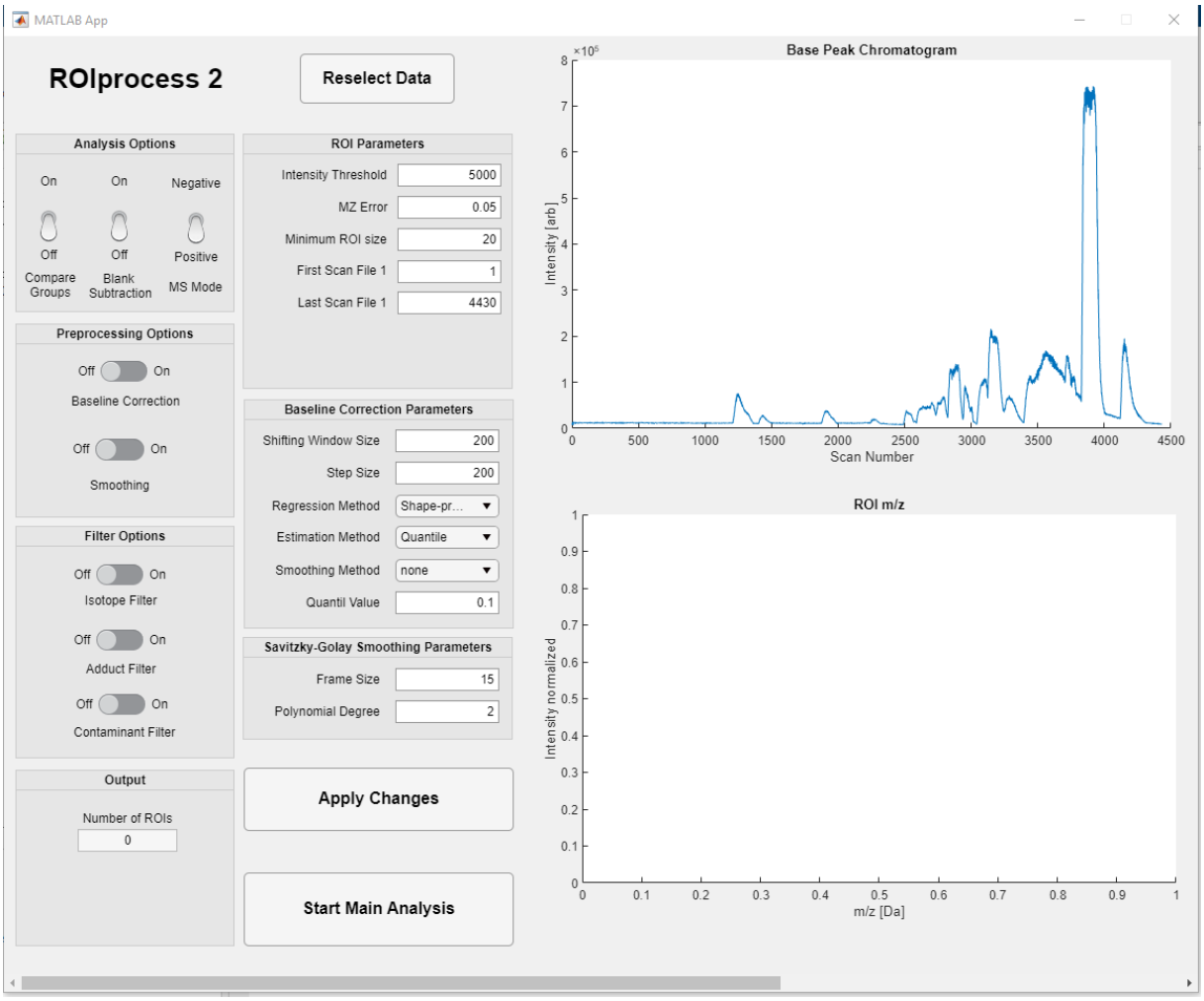
in the command window and press Enter.



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



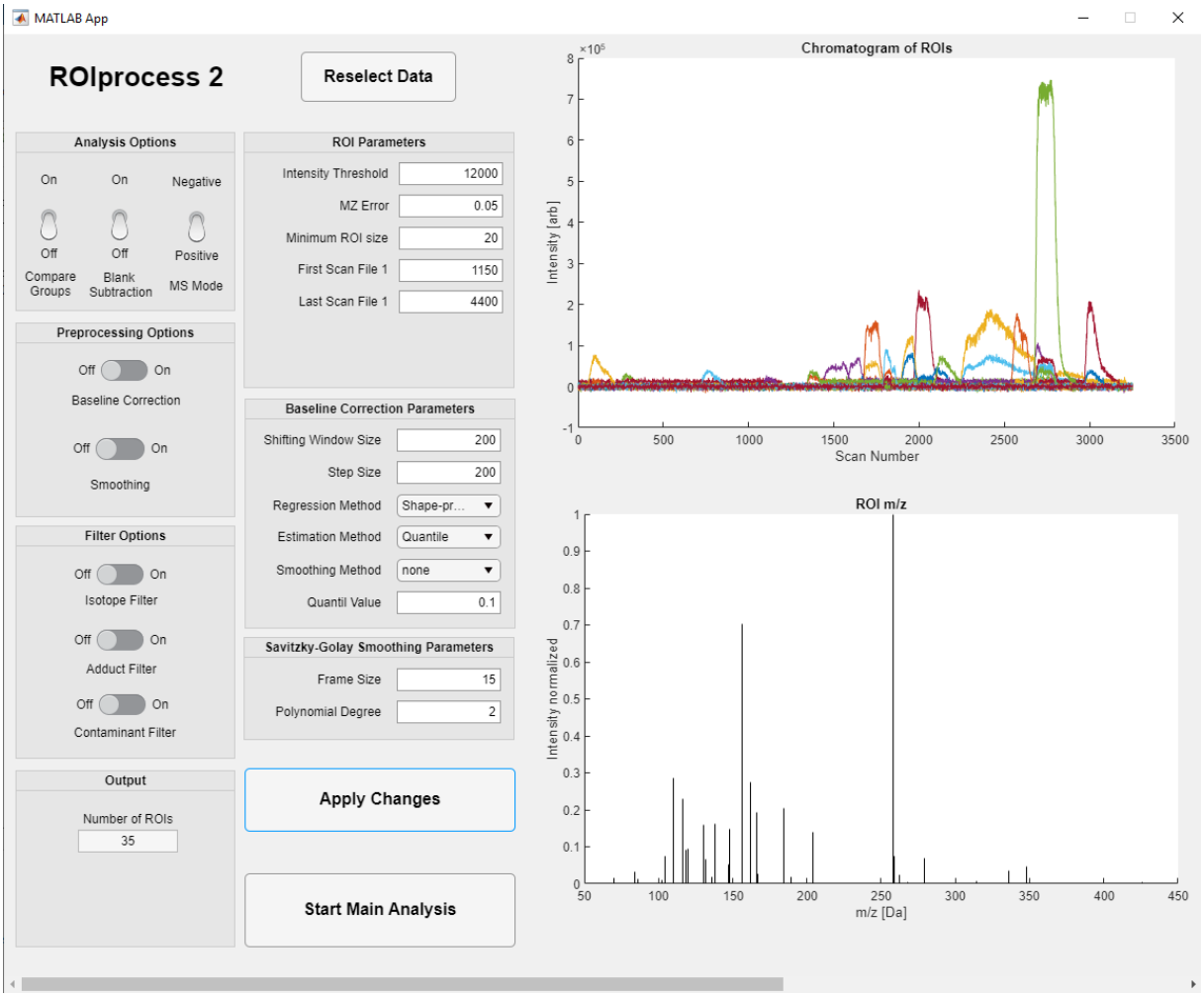
The main window opens and the base peak chromatogram is potted after the file is loaded.



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 red out by hovering the mouse cursor over the plot. In this example scan 1150 – 4400 contain peaks and background noise is around 12000.

Update the numbers in the ROI Parameters panel. Then update the graph by clicking the Apply changes button. Now the ROI m/z values will be plotted and the number of found ROIs is shown in the Output panel.



You can add preprocessing and filter options by using the switches on the left. Corresponding parameters can be changed in the parameter panels.

When the Analysis Options Compare Groups is switched on, a window opens and you are asked to select a test file for Group 2. After loading the file, the basepeak chromatogram for file 2 is plotted in a new graph.

If the Option Blank Subtraction is selected, you have to select a blank file for each group.

The button Reselect Data opens windows to reselect test files. The number of files to select depends if Pairwise Analysis and Blank Subtraction are enabled. After loading, the basepeak chromatograms are plotted again.

3.2. ROI and data pretreatment Batch processing

When settings are finalized click the Start Main Analysis button to start the main analysis using the activated options and specified settings.

Depending on the Pair-wise analysis and Blank subtraction options up to four windows will open asking to specify data files.

The options in order of appearance are:

Input	Options	Description	Note
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
Select Blanks (for Dataset 1)	File explorer	Select Blank files (for group 1)	Multiple files allowed. If multiple blanks are selected, the average blank file will be used.
Select Blanks for Dataset 2	File explorer	Select data files for group 2	Only for pairwise analysis. Different number of blank files per group are allowed

Caution: ROIprocess2 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 when more memory is needed than available.

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

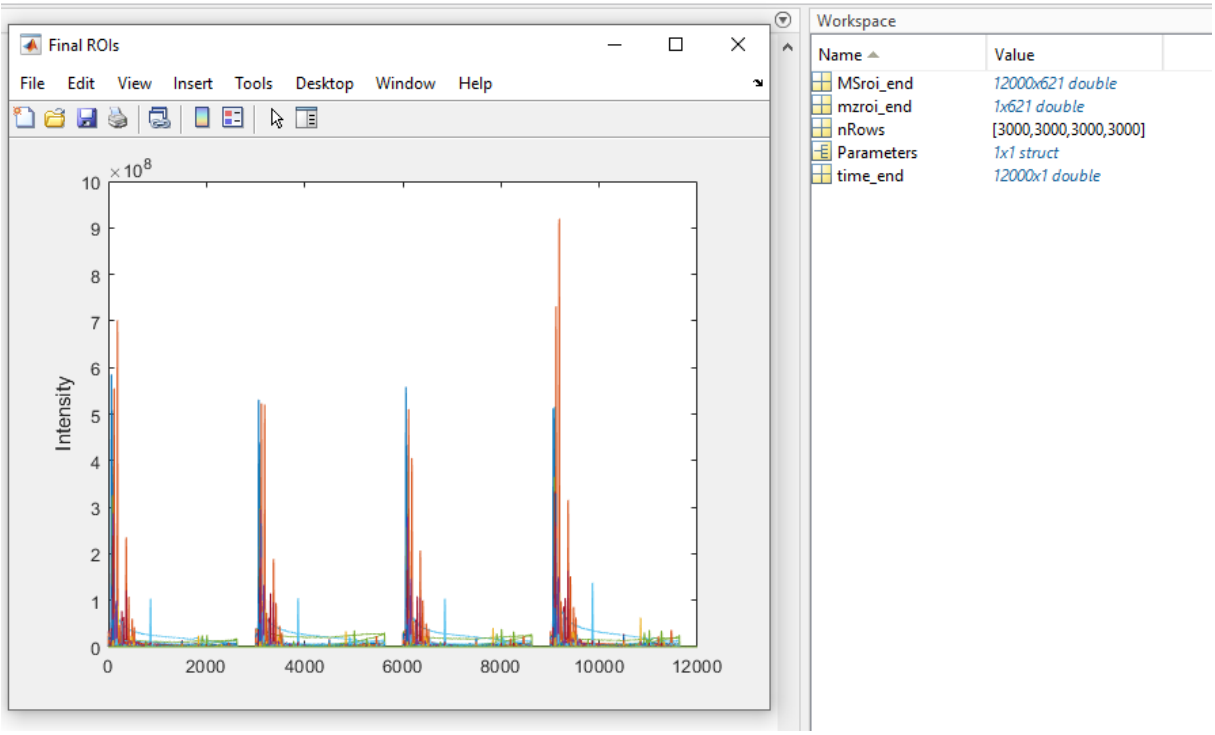
You can change the number of workers on the Home tab in the Environment section, by selecting Parallel > Parallel Preferences.

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

The main analysis 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 and the ROIprocess2 app can now be closed.

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

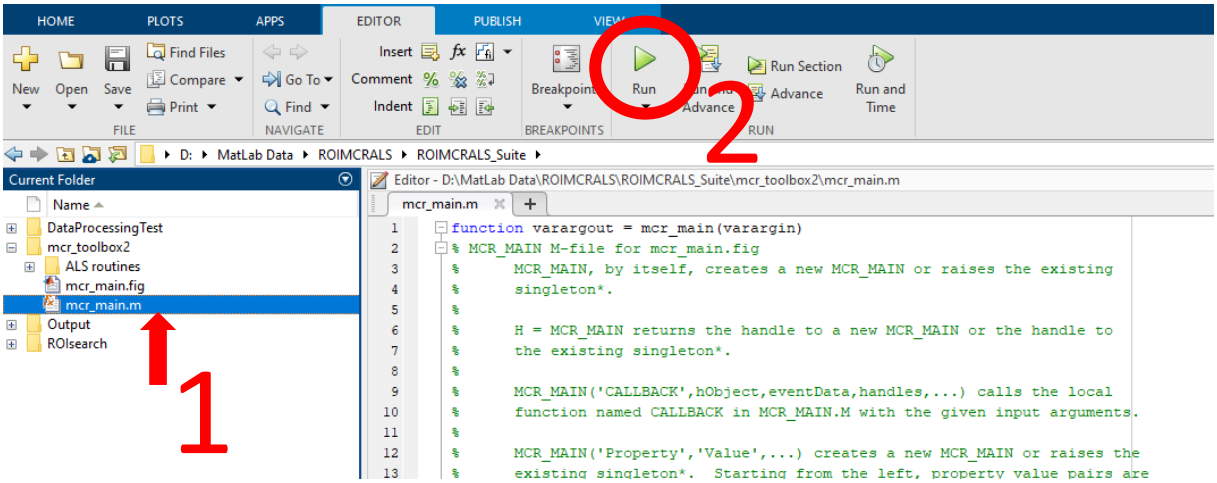
4.1. Running MCR-ALS

Type

mcr_main

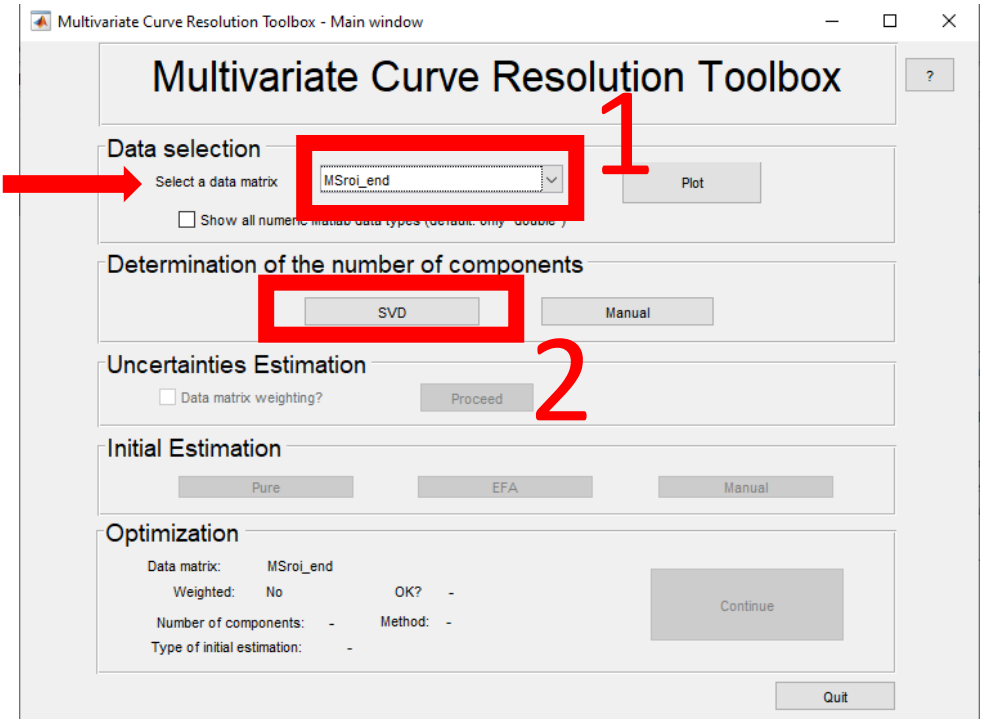
in the command window and press enter

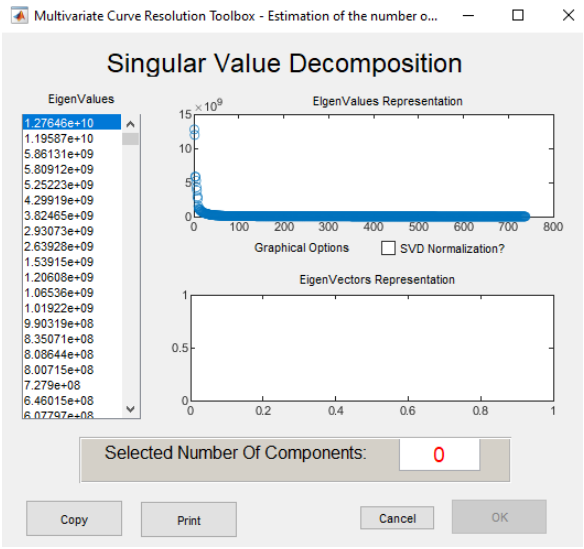
or go to the folder mcr_toolbox2, double click mcr_main.m to open it, then click Run in the Editor Tab.



A graphical interface opens. In the dropdown Menu select the MSroi_end matrix to analyze.

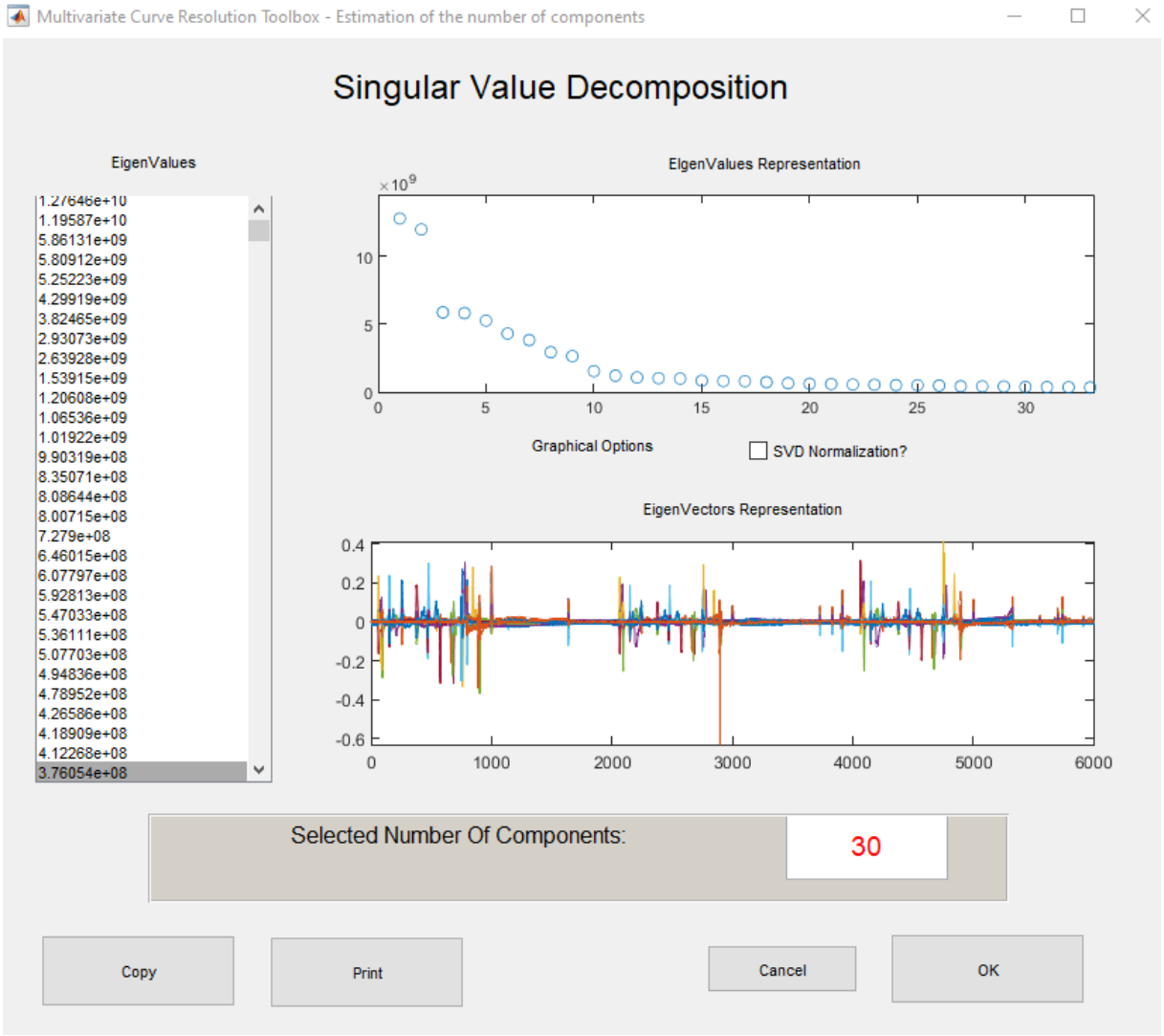
Proceed by running SVD.





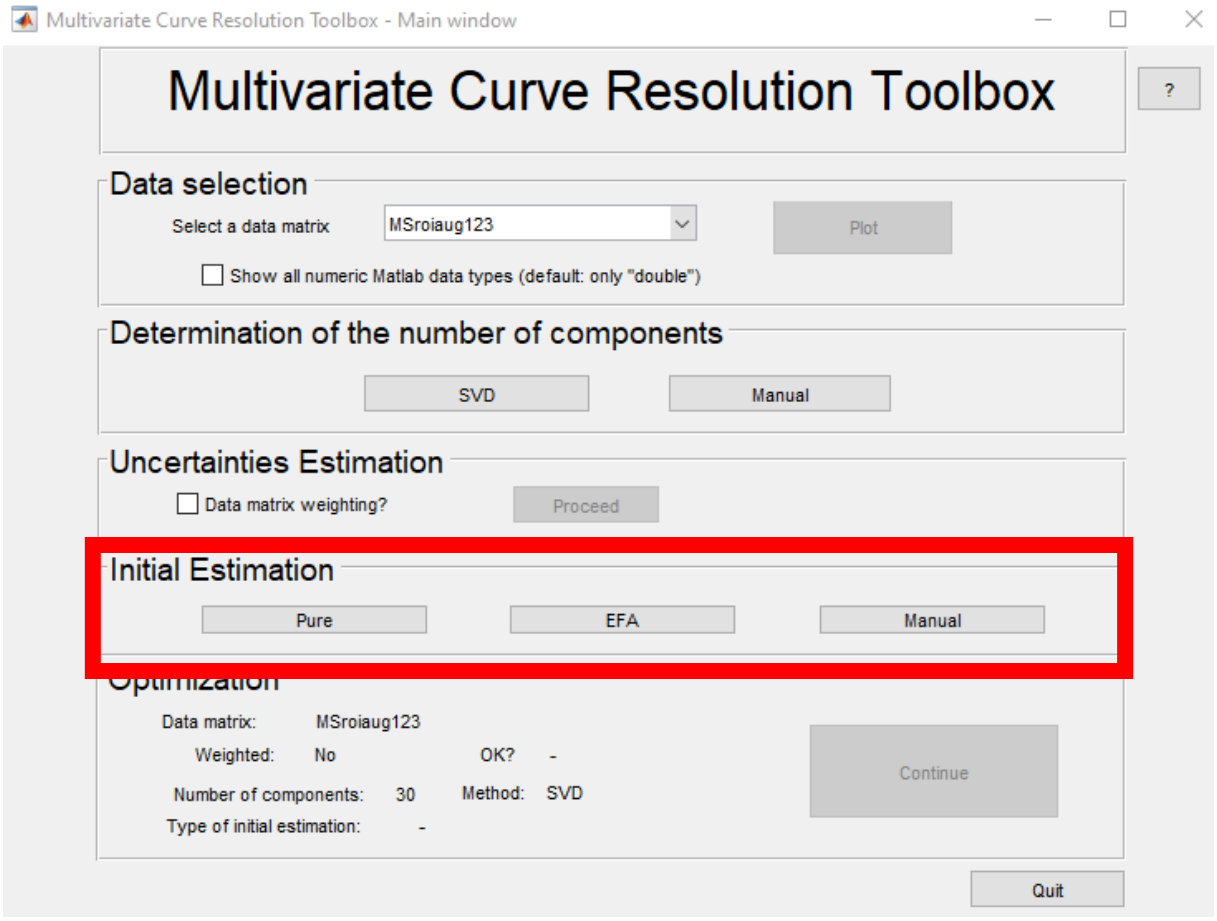
Select the number of components by reviewing the EigenValues Representation. The window can be resized for easier visual inspection. By click and dragging in the graph you can zoom in.

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.


Multivariate Curve Resolution Toolbox - Estimation of the number of components
Singular Value Decomposition
EigenValues
1.27646e+10
1.19587e+10
5.86131e+09
5.80912e+09
5.25223e+09
4.29919e+09
3.82465e+09
2.93073e+09
2.63928e+09
1.53915e+09
1.20608e+09
1.06536e+09
1.01922e+09
9.90319e+08
8.35071e+08
8.08644e+08
8.00715e+08
7.279e+08
6.46015e+08
6.07797e+08
5.92813e+08
5.47033e+08
5.36111e+08
5.07703e+08
4.94836e+08
4.78952e+08
4.26586e+08
4.18909e+08
4.12268e+08
3.76054e+08
EIgenValues Representation
×10^9
10
5
0
0
5
10
15
20
25
30
Graphical Options
SVD Normalization?
EigenVectors Representation
0.4
0.2
0
-0.2
-0.4
-0.6
0
1000
2000
3000
4000
5000
6000
Selected Number Of Components:
30
Copy
Print
Cancel
OK

4.2. Initial Estimates

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



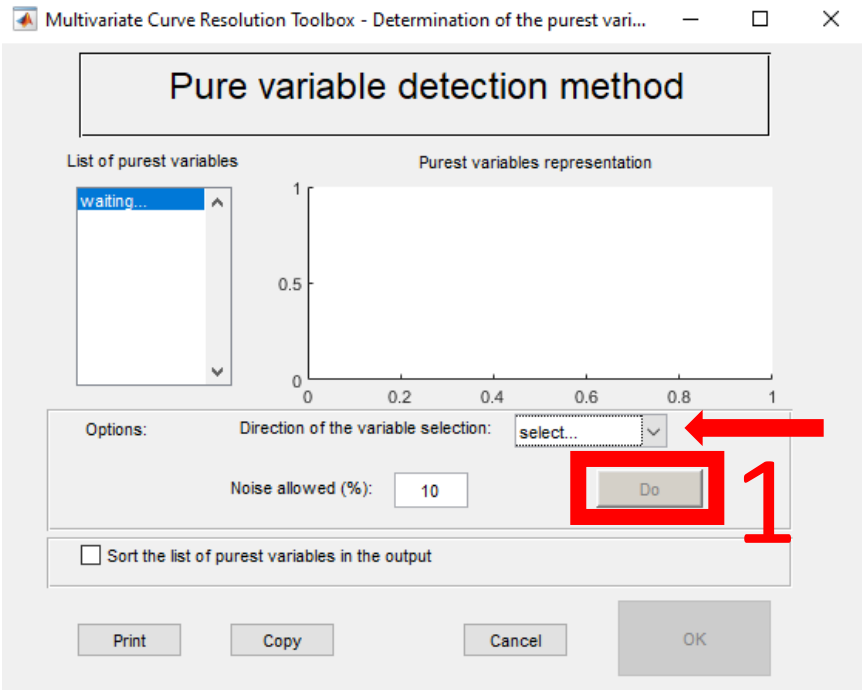
4.2.1.Pure

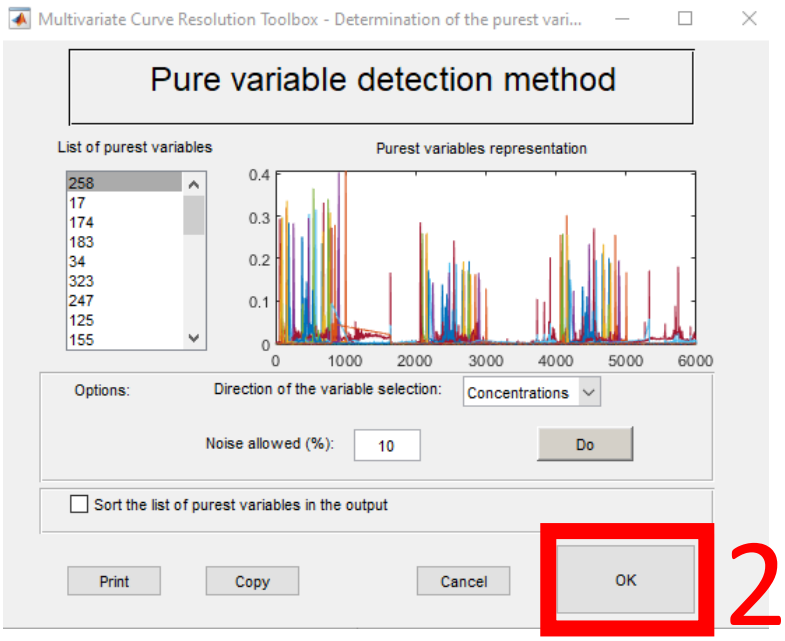
When selecting Pure estimates, the following Window opens

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.

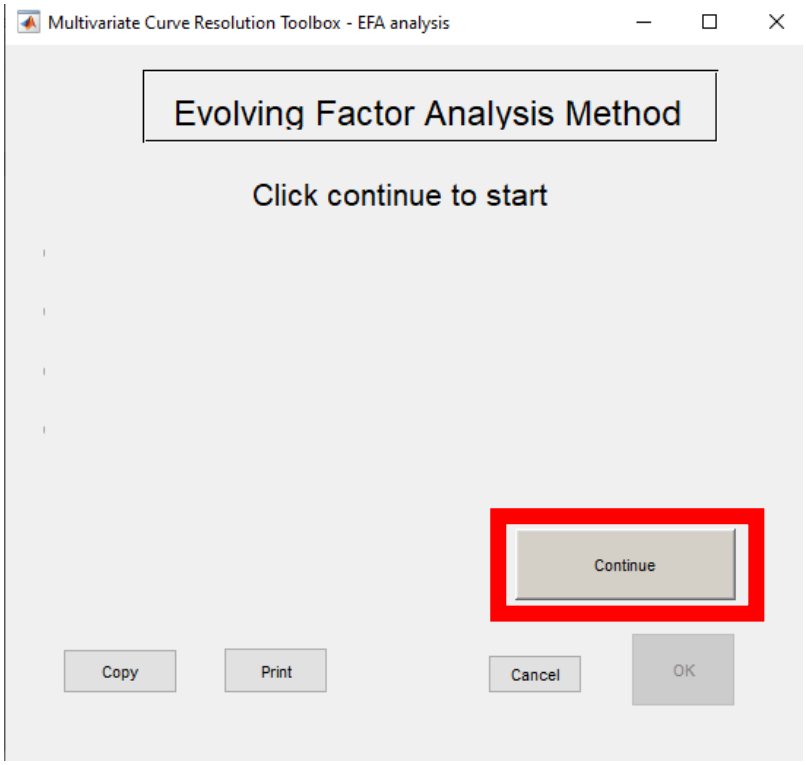




4.2.2.EFA

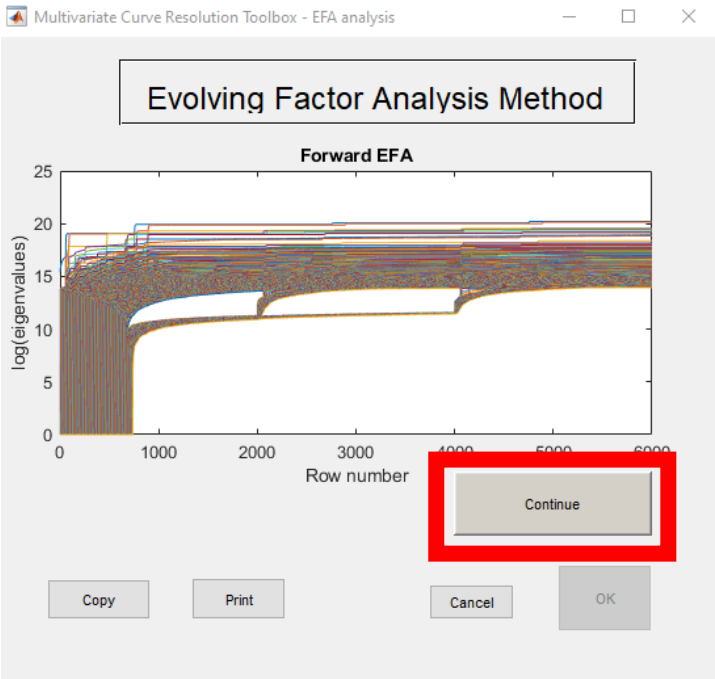
When selecting EFA the following window opens, click Continue to start forward EFA.

This step takes several minutes to calculate.

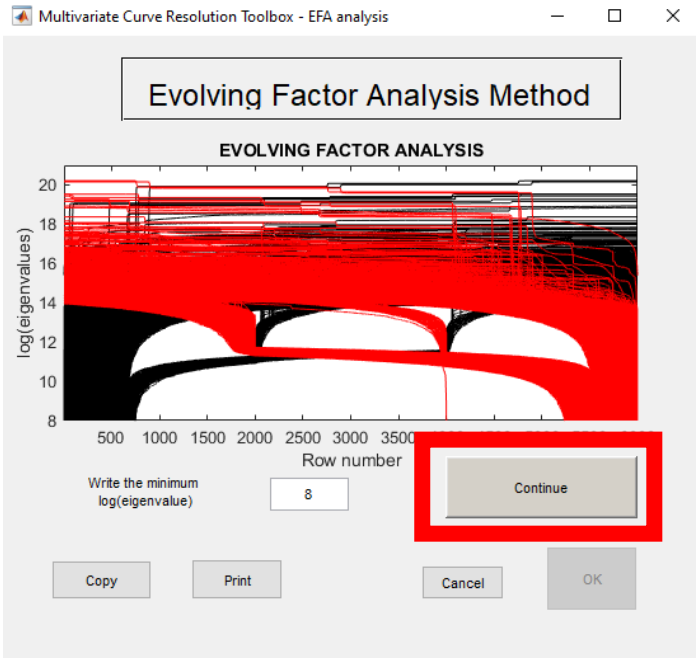


Click Continue Forward EFA is finished, to start Backward EFA.

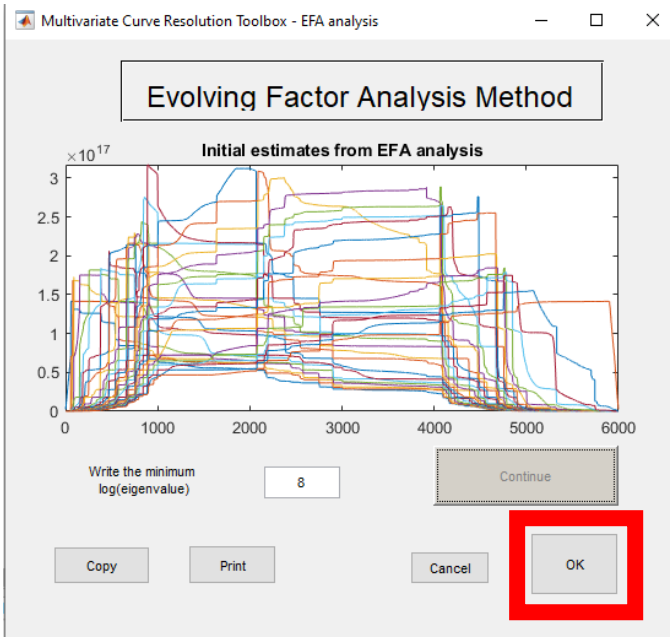
This step takes several minutes to calculate.



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.

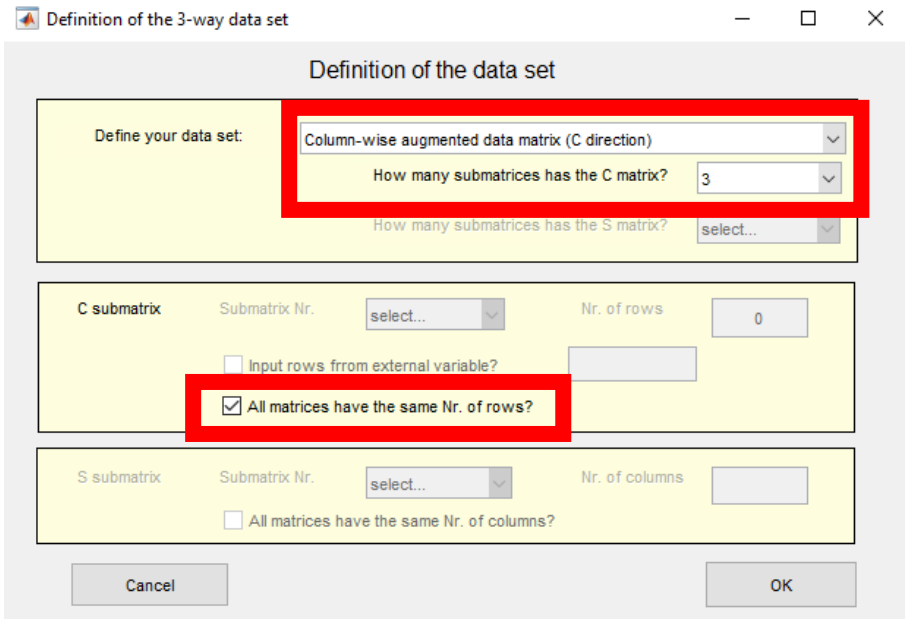
A new window opens, and several plots will be drawn. This step takes some time.

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

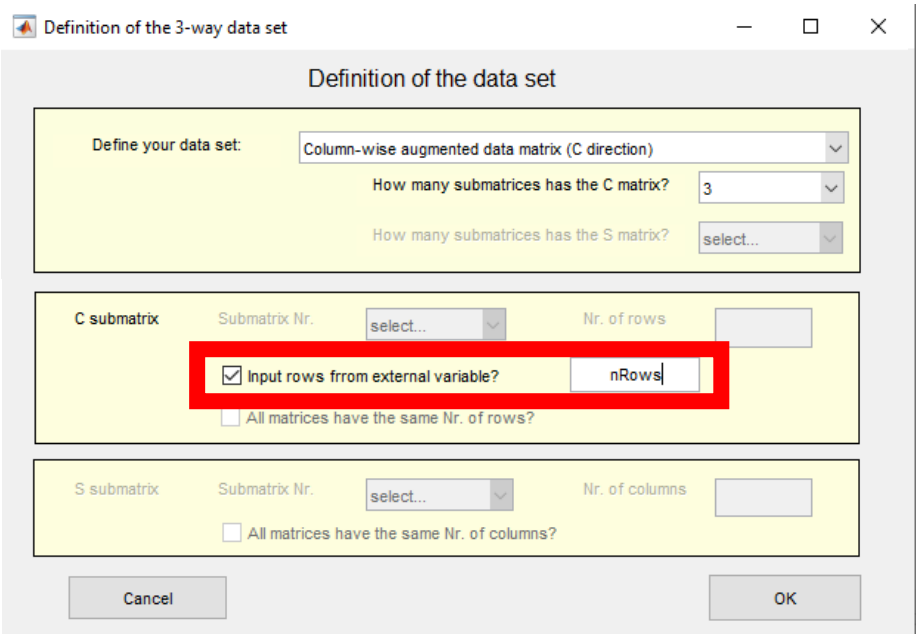
Select Column-wise augmented data matrix (C direction) in the drop down menu and check if the number of submatrices is set correctly.

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?".



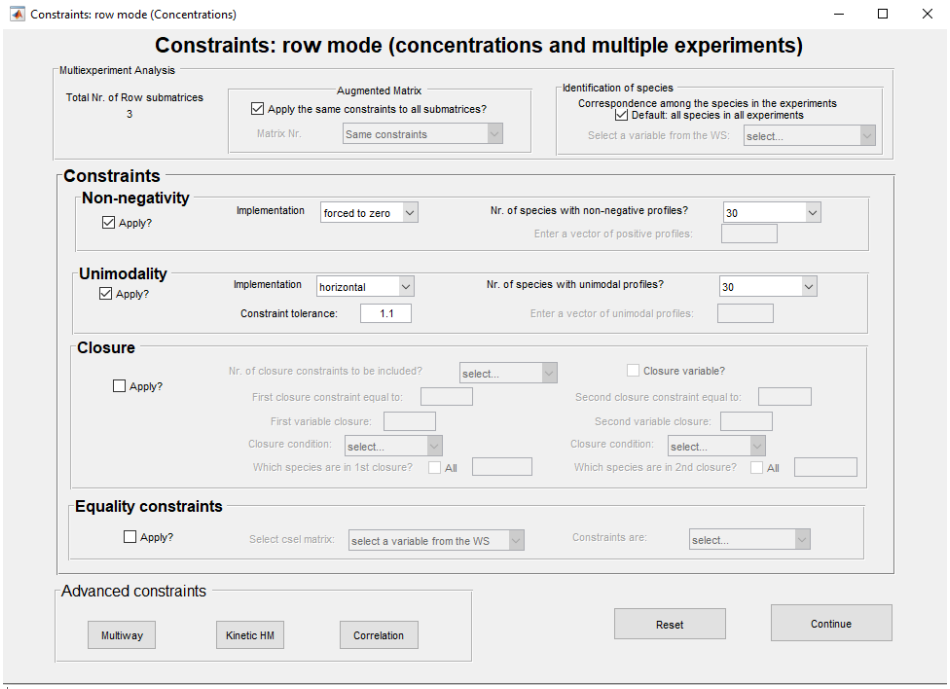
If you use pairwise analysis with different number of scans for each group check "Input from external variable?" and type nRows in the field next to it.

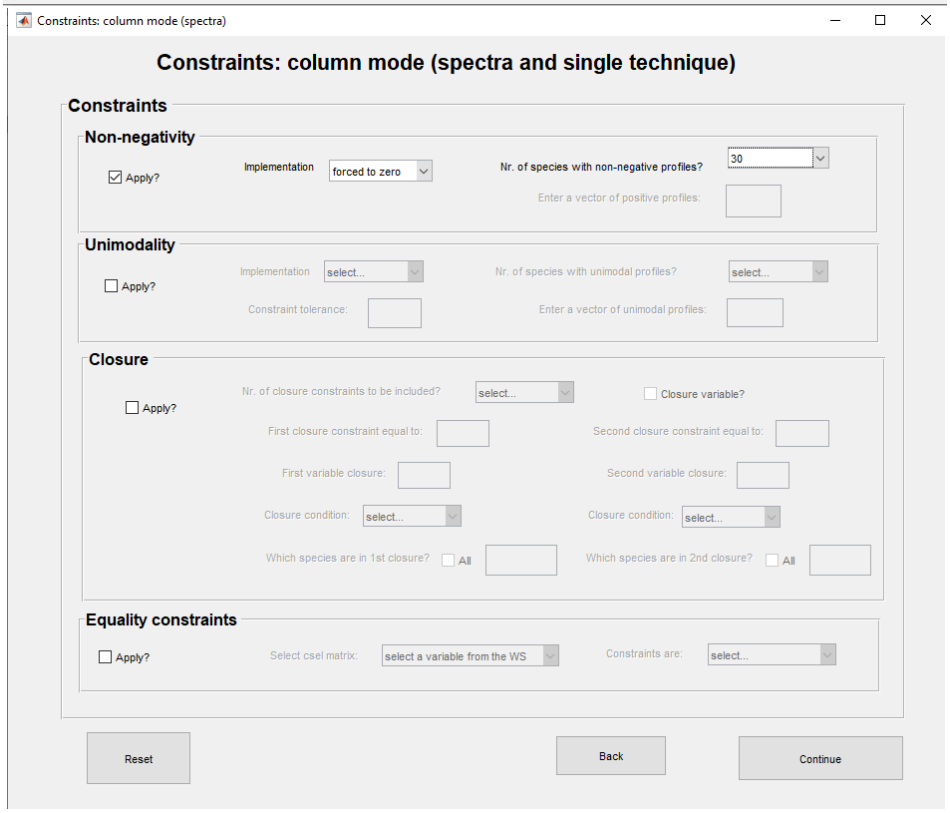


Click OK to continue.

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

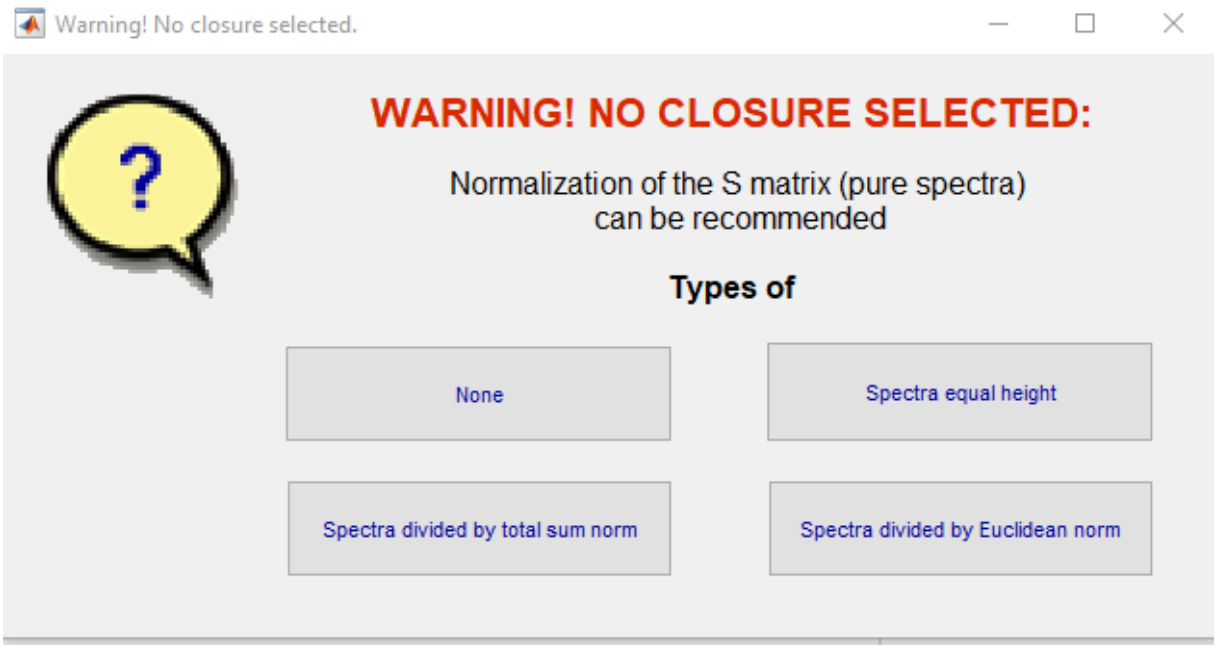
Click Continue.





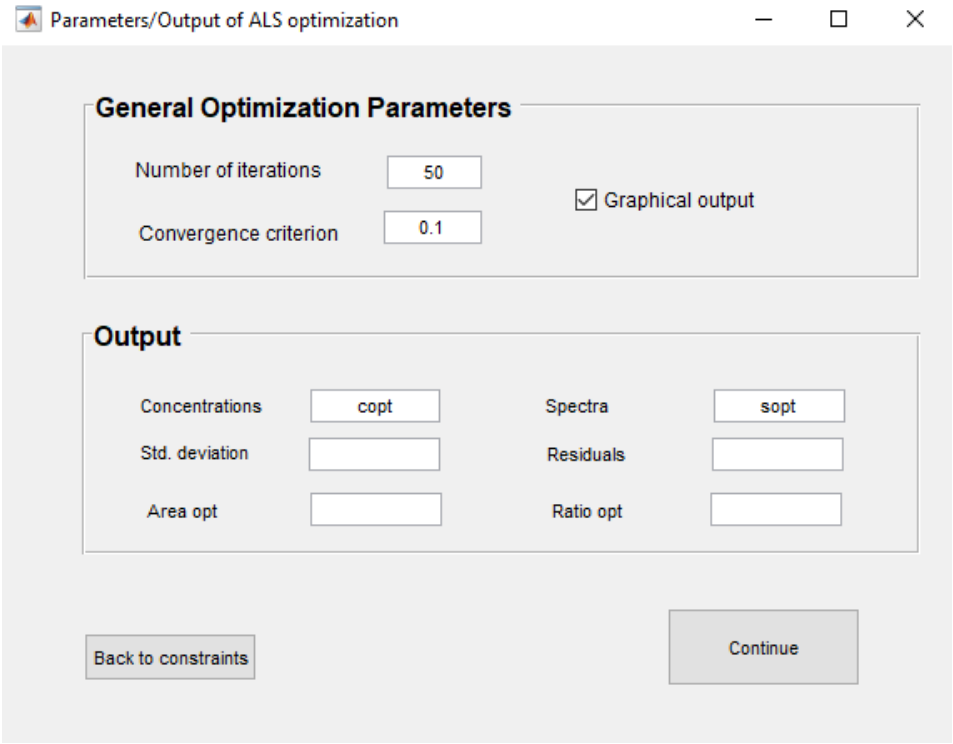
When no Closure constraint was selected a warning appears where you can choose to normalize the S matrix.

Pick one of the four options.

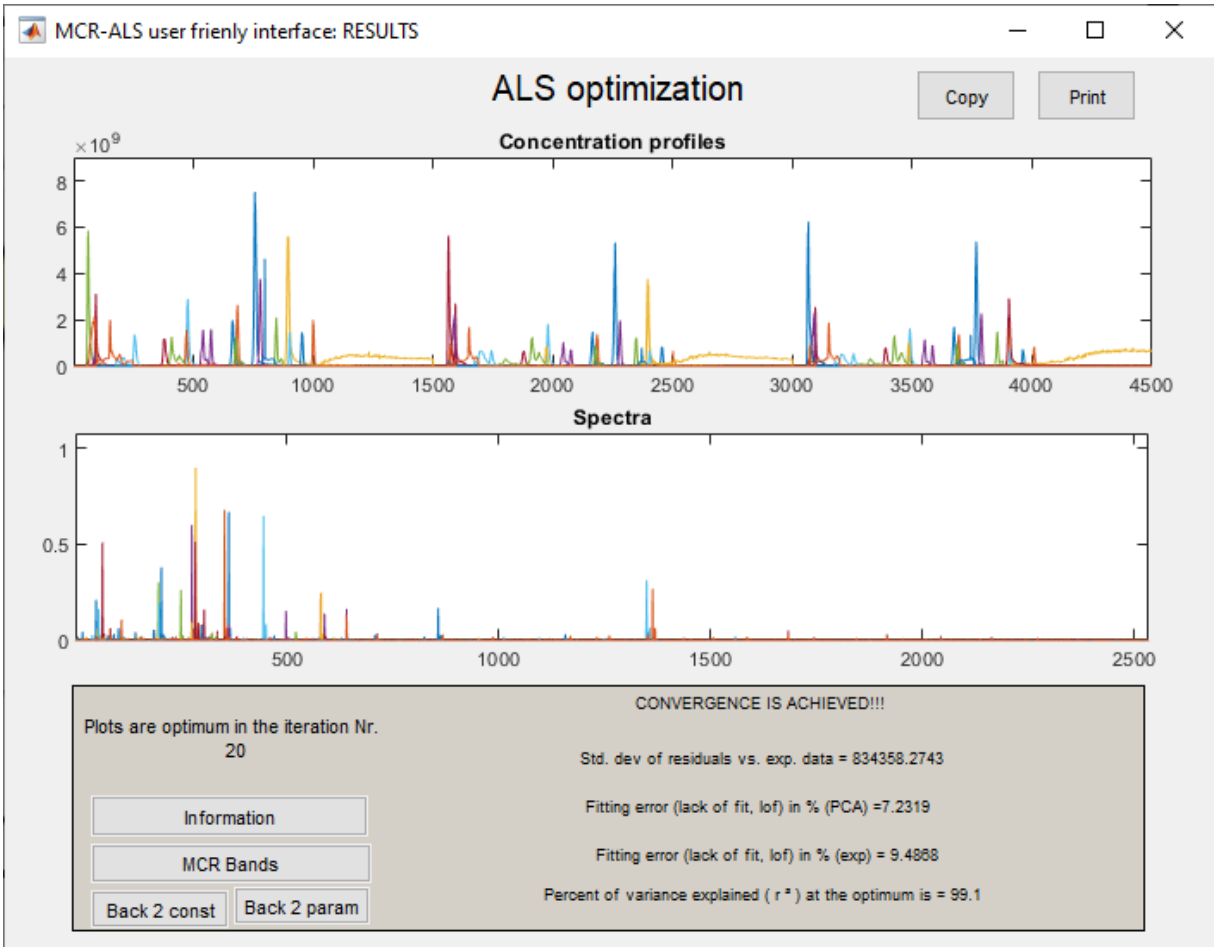


Lastly set general optimization settings and select the Output of ALS optimization by typing the name you want the matrices to have, into the corresponding field.

Then click continue.



After ALS iteration achieved convergence, the results are shown.

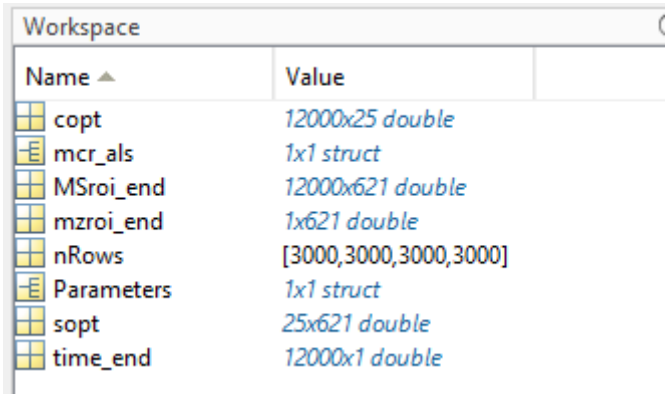


All matrices you defined in the previous screen, are generated in the Workspace tab and can be used for further calculations or analysis.

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.



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

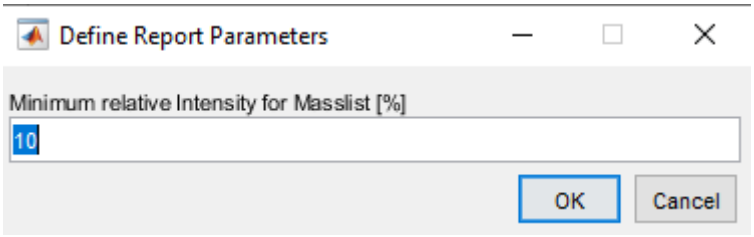
5.1. Report Generation and Component Plots

Use the command

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

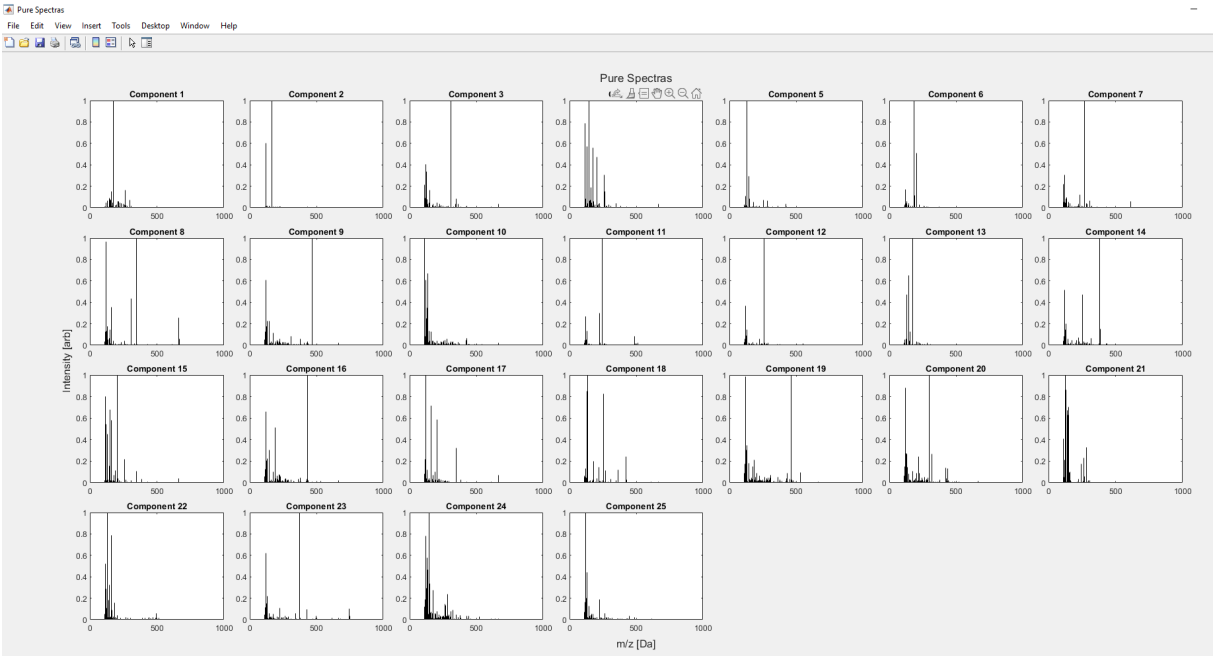
in the command window to generate a Report and plot the average concentration profile and spectra for each pure component.

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

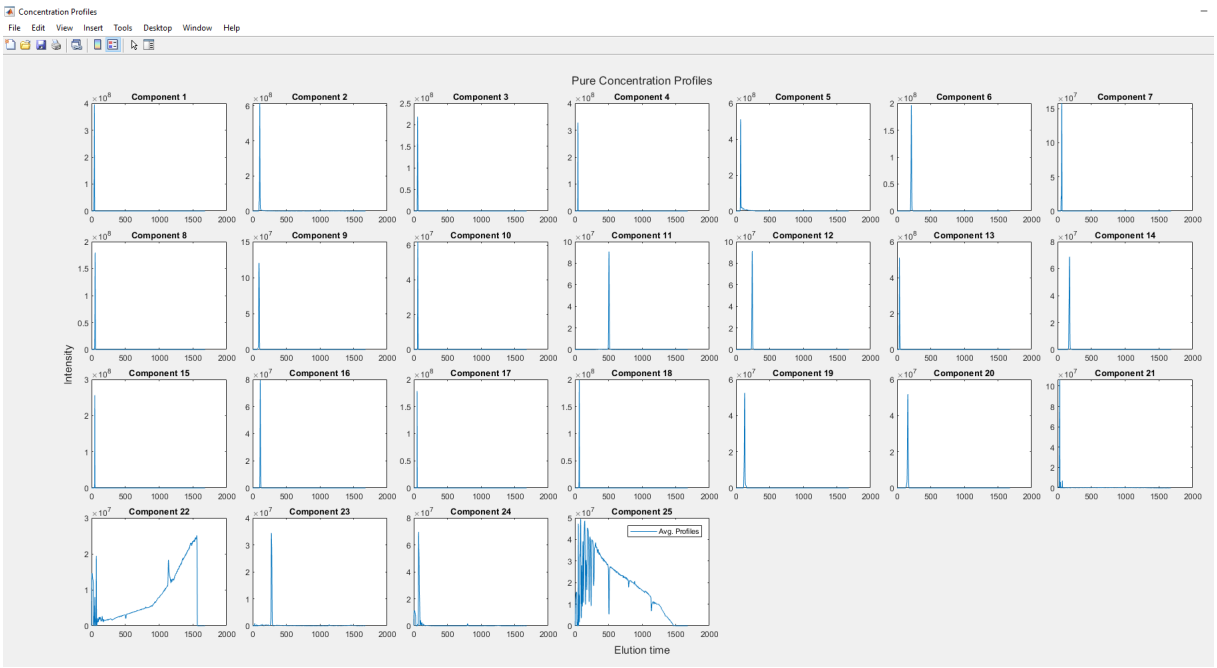


The following output is generated:

MS spectra for every pure component:

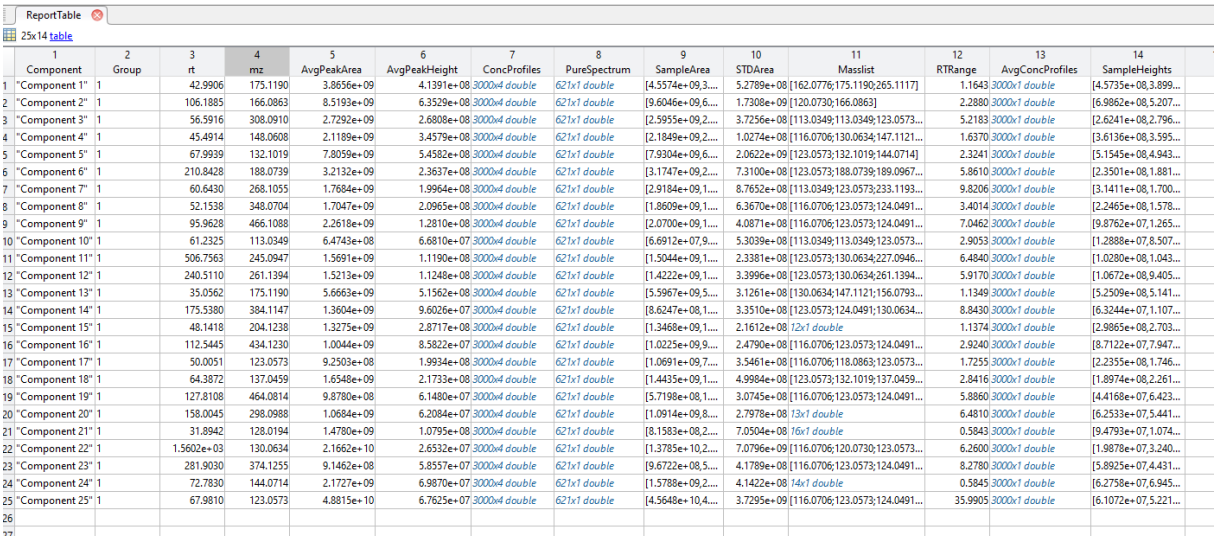


Average Concentration Profile for every pure component:



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 Component	2 Group	3 rt	4 mz	5 AvgPeakArea	6 AvgPeakHeight	7 ConcProfiles	8 PureSpectrum	9 SampleArea	10 STDArea	11 Masslist	12 RTRange	13 AvgConcProfiles	14 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;113.0349;123.0573...	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	5.4582e+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	2.3637e+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;113.0349;123.0573...	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	[2.9865e+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...
26														

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 ≥ Correlation ≥ 0.3

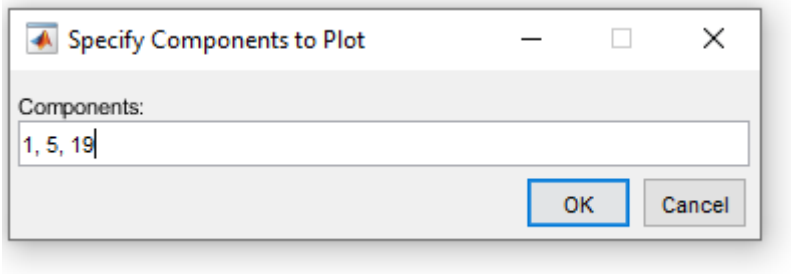
Noise Component: Correlation < 0.3

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.