PeakDecon Manual

Version 0.1

MATLAB-based Peak deconvolution software

Minkyu Park, Ph.D.

Preface

The development of PeakDecon was intended to deconvolute overlapped chromatograms of size exclusion chromatography. However, this software can be applied to any two-dimensional data that have a non-negative nature.

This work was submitted to a journal and it is currently under review. Please acknowledge it if accepted for publication. Citation information will be added once available.

Published on 2/21/2020

Table of Contents

1. Introduction 4
2. Installation 5
3. Data pre-processing using *guiPre* 6
4. Deconvolution using *guiPD* 9
5. Post-processing using *guiPost* 13
6. Tutorial 15

1. Introduction

What is PeakDecon?

- PeakDecon is a MATLAB-based software to deconvolute overlapped spectral or chromatographic peaks. For the convenience of use, a graphical user interface (GUI) is supported.

Components

- ***guiPre***: Data pre-processing GUI for an efficient data processing. Its functions include the separation of an XLSX, the reduction of data size, the cut off of data, and baseline subtraction. The detailed procedure of use is described in **Data pre-processing using *guiPre***.
- ***guiPD***: The main body of PeakDecon. See Section **Deconvolution using *guiPD***.
- ***guiPost***: Data post-processing GUI. This component was intended to help visualize the processed data. Various plots such as correlation plot, matrix plot, and box plot can be displayed. See **Post-processing using *guiPost***.

Prerequisites

- **MATLAB**: MATLAB version R2019a and R2019b was tested. The compatibility with older versions was not tested.
- **Nonnegative matrix factorization (NMF) algorithm**
  : It can be downloaded from either https://sites.google.com/site/jingukim/ or https://github.com/kimjingu/nonnegfac-matlab. Its incorporation into PeakDecon is introduced in **Installation**.

2. Installation

Concept of installation

Installation of PeakDecon means adding the path of the PeakDecon and NMF packages in the MATLAB search path so that the algorithms can be run in any folder location.

Two installation method

1. Install by running 'runMeFirst.m'
2. Install manually.

1. Install by running 'runMeFirst.m'

(1) Copy all the files in the PeakDecon packages where you want to place them.

(2) Download the NMF packge either from https://sites.google.com/site/jingukim/ or https://github.com/kimjingu/nonnegfac-matlab. Place the MATLAB m files of the NMF package in the same folder as where the PeakDecon package is located.

(3) Enter 'runMeFirst' in the MATLAB command window. Ensure that the current path is the location where 'runMeFirst.m' is placed (i.e., where the software package is located.)

(4) Or, double click 'runMeFirst.m' on the file window of MATLAB. This will open the m file in the editor window. Click the 'run' button of the editor window or press the F5 key.

(5) Or, drag 'runMeFirst.m' file to MATLAB command window

(6) Once the installation is successful, the message 'PeakDecon path is added to the MATLAB search path. Enjoy PeakDecon!!' in the MATLAB command window.

2. Install manually.

(1) You can add the PeakDecon folder location as well as the NMF package to the MATLAB search path using either MATLAB commands of 'pathtool' or 'addpath'. Please check MATLAB online help.

(2) Or, you can copy the package files to a folder that is already added to the MATLAB search path.

3. Data pre-processing using *guiPre*

Introduction

PeakDecon requires a specific data format for data processing. ***guiPre*** was designed to facilitate the subsequent data process in ***PeakDecon***.

How to run *guiPre*?

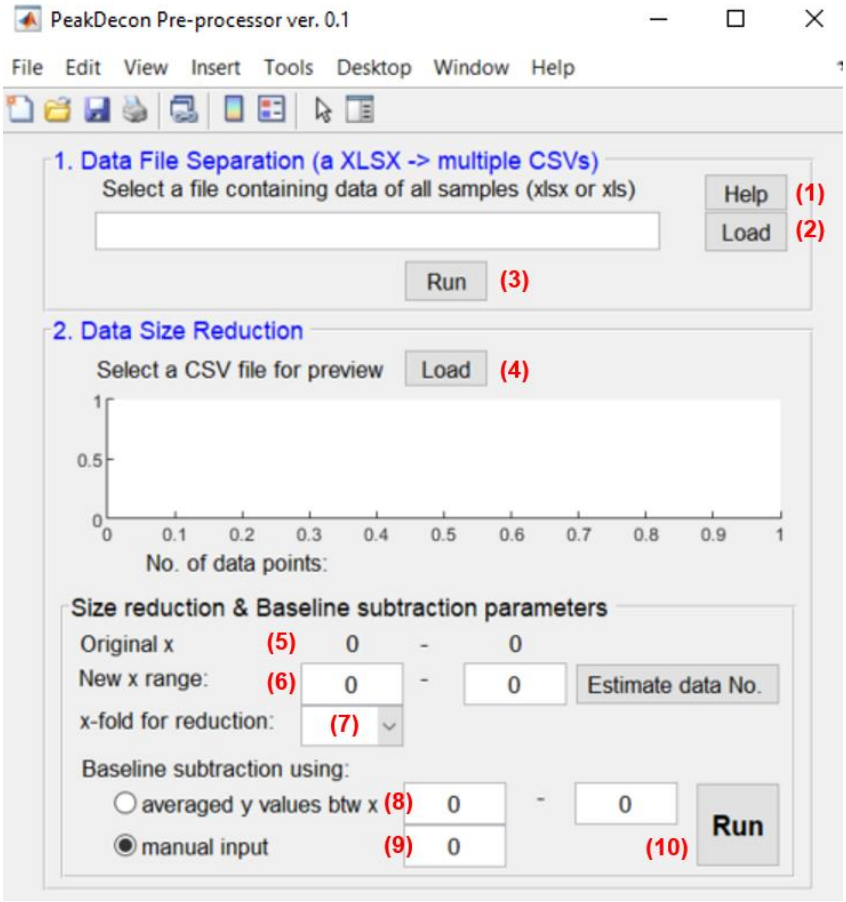
In MATLAB command line, type *guiPre*.

```
>> guiPre
```

If the folder containing the PeakDecon software is not registered in the MATLAB search path, the above command does not work. See **Installation**.

Guide for *guiPre*

guiPre consists of two compartments: 1. Data File Separation, and 2. Data Size Reduction.

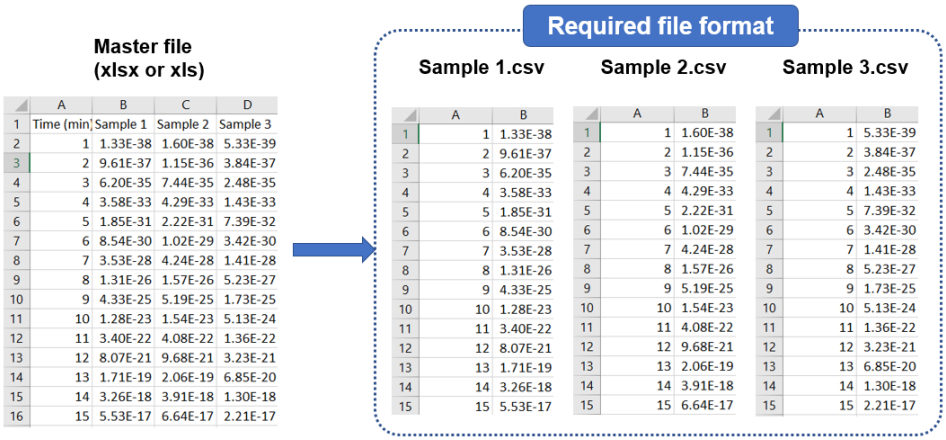


< ***guiPre*** window >

1. Data File Separation

- PeakDecon main program (***guiPD***) requires CSV data files containing two columns: x-axis and y-axis data (only numeric data). If you have a master file containing all the data, ***guiPre*** can separate it into individual CSV data files so that the data format becomes compatible with ***guiPD***.

- Data format:
  If you have a master file in xlsx or xls, the file should have x-axis data in the first column and y-axis data of your samples from the second column. For instance, as shown in the following figure, if you have chromatogram data consisting of time as x-axis and signal as y-axis, and have three samples whose names are Sample 1, Sample 2, and Sample 3; the master file then has four columns. Noted that the first row should contain headers that are used as the file name for CSV. You can also read the instruction of data format by clicking the '**Help**' button **(1)**.



	A	B	C	D
1	Time (min)	Sample 1	Sample 2	Sample 3
2	1	1.33E-38	1.60E-38	5.33E-39
3	2	9.61E-37	1.15E-36	3.84E-37
4	3	6.20E-35	7.44E-35	2.48E-35
5	4	3.58E-33	4.29E-33	1.43E-33
6	5	1.85E-31	2.22E-31	7.39E-32
7	6	8.54E-30	1.02E-29	3.42E-30
8	7	3.53E-28	4.24E-28	1.41E-28
9	8	1.31E-26	1.57E-26	5.23E-27
10	9	4.33E-25	5.19E-25	1.73E-25
11	10	1.28E-23	1.54E-23	5.13E-24
12	11	3.40E-22	4.08E-22	1.36E-22
13	12	8.07E-21	9.68E-21	3.23E-21
14	13	1.71E-19	2.06E-19	6.85E-20
15	14	3.26E-18	3.91E-18	1.30E-18
16	15	5.53E-17	6.64E-17	2.21E-17

	A	B
1	1	1.33E-38
2	2	9.61E-37
3	3	6.20E-35
4	4	3.58E-33
5	5	1.85E-31
6	6	8.54E-30
7	7	3.53E-28
8	8	1.31E-26
9	9	4.33E-25
10	10	1.28E-23
11	11	3.40E-22
12	12	8.07E-21
13	13	1.71E-19
14	14	3.26E-18
15	15	5.53E-17

	A	B
1	1	1.60E-38
2	2	1.15E-36
3	3	7.44E-35
4	4	4.29E-33
5	5	2.22E-31
6	6	1.02E-29
7	7	4.24E-28
8	8	1.57E-26
9	9	5.19E-25
10	10	1.54E-23
11	11	4.08E-22
12	12	9.68E-21
13	13	2.06E-19
14	14	3.91E-18
15	15	6.64E-17

	A	B
1	1	5.33E-39
2	2	3.84E-37
3	3	2.48E-35
4	4	1.43E-33
5	5	7.39E-32
6	6	3.42E-30
7	7	1.41E-28
8	8	5.23E-27
9	9	1.73E-25
10	10	5.13E-24
11	11	1.36E-22
12	12	3.23E-21
13	13	6.85E-20
14	14	1.30E-18
15	15	2.21E-17

  guiPre separates the master file (xlsx) into three CSV files whose file names are the same as the texts in the headers of the master file (i.e., Sample 1, Sample 2, and Sample 3 of the above example).

- In order to load the master file, click the '**Load**' button **(2)**. It pops up a window to help load files.

- Once the data file is loaded, click the '**Run**' button **(3)** to separate the data file. The generated CSV files are located in the subfolder with the same name as the master file name.

- A detailed example is provided in the **Tutorial**.

- If your data files are already compatible with ***guiPD***, you can skip this step.

2. Data Size Reduction

- This compartment loads a data file (CSV) to show the preview of data by clicking the '**Load**' button **(4)**. Subsequently, it processes all the data in the same folder where the loaded file is located.
- A specific range of x-axis can be selected while deleting the other ranges. One can select x range by entering values to the text boxes of '**New x range**' **(6)**. You can check the range of x-axis of data in '**Original x**' **(5)**.
- The data size can be reduced with *N* fold. *N* ranges from 2 to 1000. The preset of fold values can be selected in '**x-fold for reduction**' **(7)**.
- Due to the nature of the NMF algorithm, positive baselines may interfere analysis unless subtracted. Baseline subtraction can be carried out in two different ways. First, the average y-axis value at a range of x-axis can be subtracted from data. This option can be employed by selecting the radio button of '**averaged y values btw x**' and entering values its text boxes **(8)**. The other way is to subtract a manual value from data. To this end, select the radio button of '**manual input**' and enter a value to the corresponding text box **(9)**.
- When all the parameters are set, click the '**Run**' button **(10)**.

4. Deconvolution using *guiPD*

Introduction

guiPD (i.e., PeakDecon) is the main body of the software. This software embedded six NMF algorithms: the alternating nonnegative least squares (ANLS) with block principal pivoting method (ANLS-BPP), ANLS with active set method and givens updating (ANLS-ASGIVENS), ANLS with active set method and column grouping (ANLS-ASGROUP), alternating least squares method (ALS), hierarchical alternating least squares method (HALS), and multiplicative updating method (MU).

It is also feasible to run the sequence of HALS followed by any algorithm among the six embedded ones. This mode was included since the HALS with Gaussian kernel principal component analysis (PCA) initialization followed by ANLS-BPP performs well in some cases.

How to run *guiPD*?

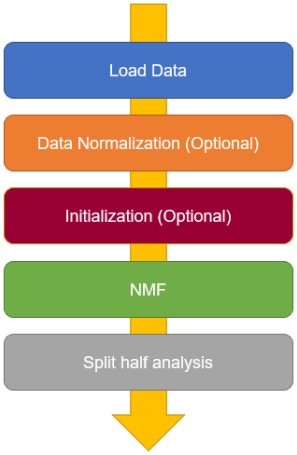
In MATLAB command line, type *guiPD*.

```
>> guiPD
```

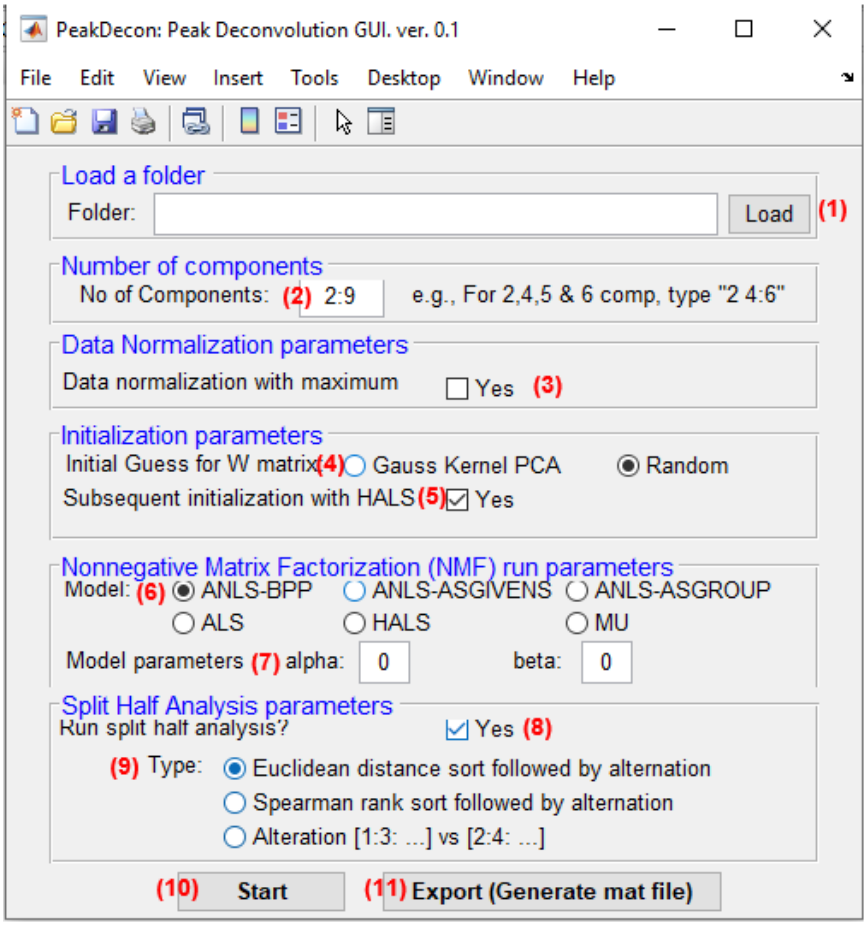
If the folder containing the PeakDecon software is not registered in the MATLAB search path, the above command does not work. See **Installation**.

Guide for *guiPD*

guiPD consists of six compartments: ‘Load a folder’, ‘Number of components’, ‘Data Normalization parameters’, ‘Initialization Parameters’, ‘’NMF Run Parameters’, and ‘Split half analysis’. The order of the compartments is agreed with the workflow of data processing as shown below.



< Workflow of data processing >



< ***guiPD*** window >

- You can load data files by selecting a folder containing either CSV files or subfolders containing CSV files **(1)**. The latter case is necessary if you want to visualize deconvolution results with respect to different groups during post-processing (i.e., ***guiPost***). For example, ***guiPost*** allows box plots to compare components among different groups if processed data contains grouping information. Files in each subfolder are treated to be in the same group. It should be borne in mind that if the latter is your case, no CSV file can be present under the selected folder, which violates the hierarchy of data structure in the given codes. **Tutorial** provides an example of using data with groups.

- The number of components for NMF modeling should be pre-defined before running the algorithm. To this end, you can enter a scalar or vector value in the text box of '**No of Components**' **(2)**. The notation follows MATLAB's vector notation. (e.g., '2 4:6' means a vector of [2 4 5 6]'.

- Normalization is helpful to have reliable modeling when data have reasonably greater values than the noise. Normalization can be performed by checking the checkbox of '**Data normalization with maximum**' **(3)**.

- Initialization can be manually setup. The default value is to employ the HALS with random initialization. The developer suggests the use of the HALS with random initialization (i.e., default).

- If one chooses Gauss Kernel PCA rather than Random, the Gaussian kernel PCA is run first to calculate the *W* matrix. The computed *W* and a randomized *H* are used as the initial values for the HALS. Finally, the *W* and *H* calculated from the HALS are used for the initial values of the main NMF algorithm selected in **(6)**. If it is not necessary to a manual initialization, choose the '**Random**' radio button in **(4)** and uncheck the checkbox in **(5)**.
  In case of running Gaussian Kernel PCA without the HALS, check '**Gauss Kernel PCA**' radio button in **(4)** and uncheck the checkbox in **(5)**.

- The main NMF algorithm can be selected in **(6)**. α and β values of the NMF algorithms can be selected **(7)**. The meaning of the parameters can be found in the articles referred to in https://github.com/kimjingu/nonnegfac-matlab. The default values are zero.

- The split-half analysis is available as a means of cross-validation. If checking the checkbox of '**Run split half analysis**?' **(8)**, it enables split-half analysis. Otherwise, split half analysis can be skipped. In this software, three options are provided **(9)**. First, data are sorted based on the Euclidean distance of the *H* matrix of simulation using all data. Subsequently, the sorted data are assigned alternatively to two splits. The similarity of the components among three simulations (i.e., one with all data, and the other two with each split data) can be measured by average cosine similarity that is the average of cosine similarity values of all possible combinations of splits. There are three combinations possible (all data vs. split 1, all data vs. split 2, and split 1 vs split 2).
  The second option is to use the Spearman rank as a criterion for sorting. This option has the same underlying principle as the Euclidean distance sorting option. The last option is to assign data alternately to two splits without any sorting algorithm. The default is to use the Euclidean distance.

- When all the parameters are set, click the '**Run**' button **(10)**.

- Once the model run is completed, export the results to a mat file by clicking '**Export (Generate mat file)**' **(11)**. The generated mat file can be visualized using ***guiPost***.

5. Post-processing using *guiPost*

Introduction

guiPost is designed to help visualization of processed data. This function provides plotting of components, split-half analysis, and cosine similarity index results. In addition, various visualization methods such as correlation matrix plot, matrix plot by group, box plot by group, PCA, and clustergram.

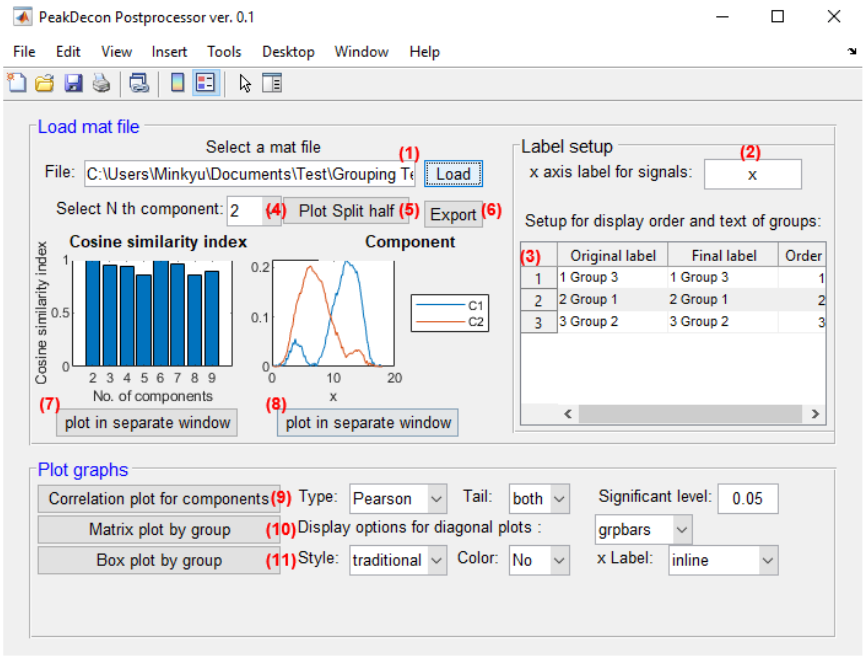
How to run *guiPost*?

In MATLAB command line, type *guiPost*.

```
>> guiPost
```

If the folder containing the PeakDecon software is not registered in the MATLAB search path, the above command does not work. See **Installation**.

Guide for *guiPost*



< ***guiPost*** window >

- The exported mat file in ***guiPD*** can be loaded by clicking '**Load**' **(1)**.
- Since input data have no information about the x-axis label, it needs to be manually entered in the '**x axis label for signals**' **(2)**. The entered value is applied to all the subsequent plots.

- If grouping information is included in the mat file, the display order can be manually adjusted in **(3)**. In addition, labels for groups can also be changed in **(3)**.
- The number of components can be selected for visualization in **(4)**.
- Split-half analysis can be shown if conducted by clicking '**Plot Split half**' **(5)**.
- Results including *W*, *H*, and the area of peak components can be exported by pressing '**Export'** button **(6)**
- While cosine similarity result is shown in a subfigure of the main window, it can be plotted in a separate window. Click '**plot in separate window**' **(7)**.
- In a similar vein, the plot for components can also be displayed in a separate window by clicking '**plot in separate window**' **(8)**.
- '**Plot graphs**' can visualize the results with various visualization methods.
- '**Correlation plot for components**' plots a correlation plots among components **(9)**. Please refer to MALAB's manual for the command 'corrplot' for the related parameter setup. This plot method is helpful to check the collinearity among components.
- If grouping information is available, '**Matrix plot by group**' can be performed **(10)**. Please refer to MALAB's manual for the command 'gplotmatrix' for the related parameter setup. This plot is similar to the correlation plot but useful to compare components among different groups.
- If grouping information is available, '**Box plot by group**' can be performed **(11)**. Please refer to MALAB's manual for the command 'boxplot' for the related parameter setup. This plot is useful to assess the trend of each component change among different groups.

6. Tutorial

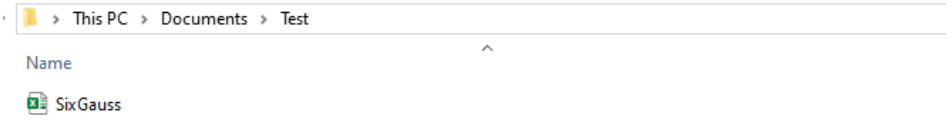
This tutorial provides an easy-to-follow example. Users may easily get an idea of how to use PeakDecon by following the **Scroll-down tutorial** step by step.

Data set

- A master excel file (xlsx) containing 100 samples is provided along with the software.
- The data is synthetic and composed of six unimodal Gaussian peaks (i.e., 6 components).

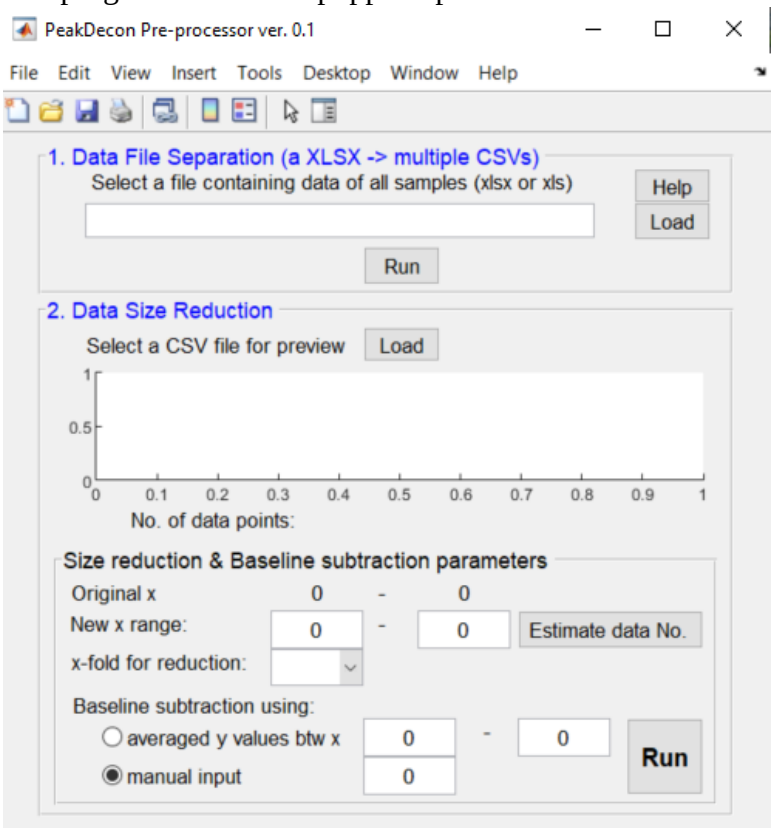
Scroll-down tutorial

- Locate ‘SixGauss.xlsx’ data file to the desired folder. The file was located in the ‘Test’ folder in this tutorial.

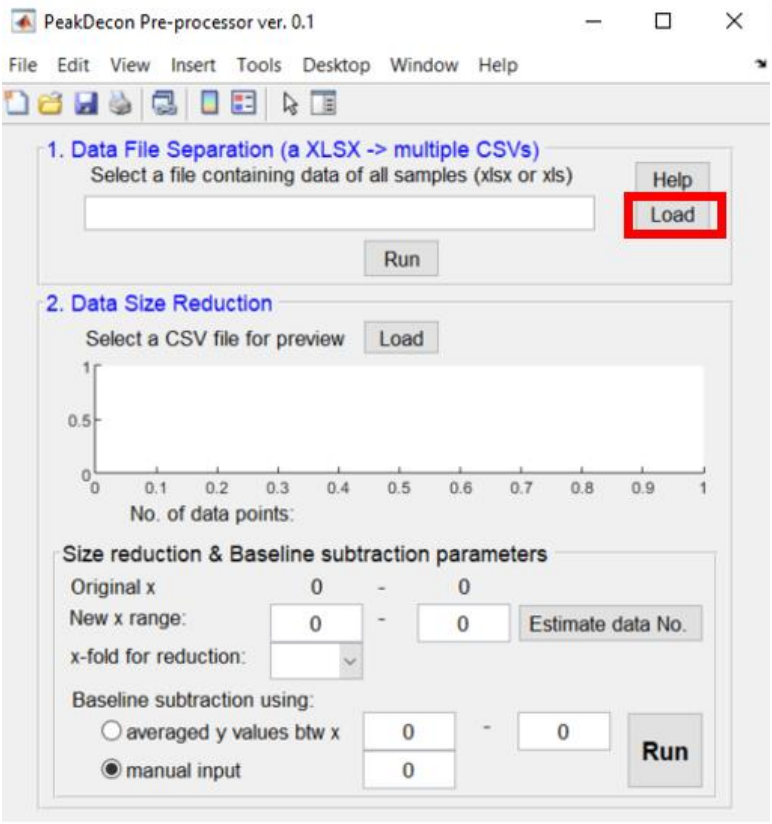


- Type ‘guiPre’ to run ***guiPre***. Make sure to save all the data and figures in MATLAB before running ***guiPre*** because ***guiPre*** will close all the MATLAB windows and clear all variables.

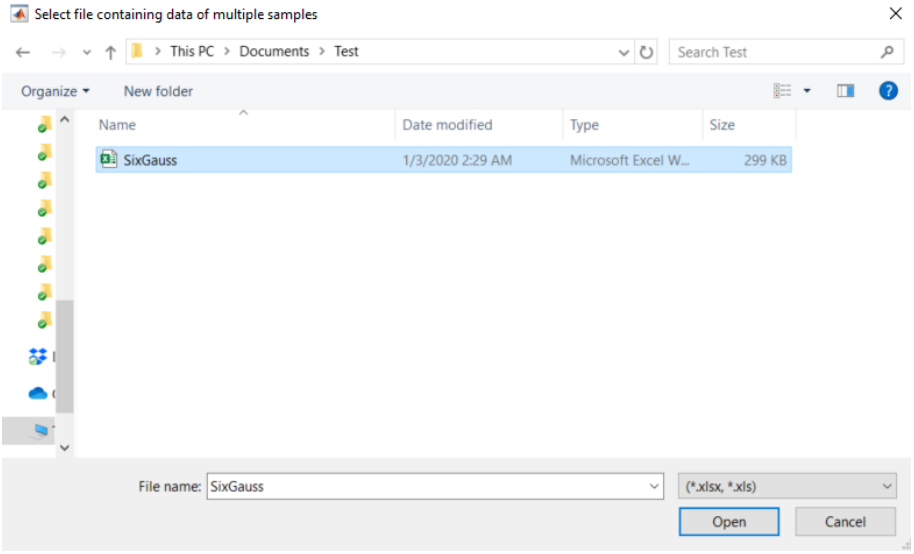
- The program window is popped up as shown below.



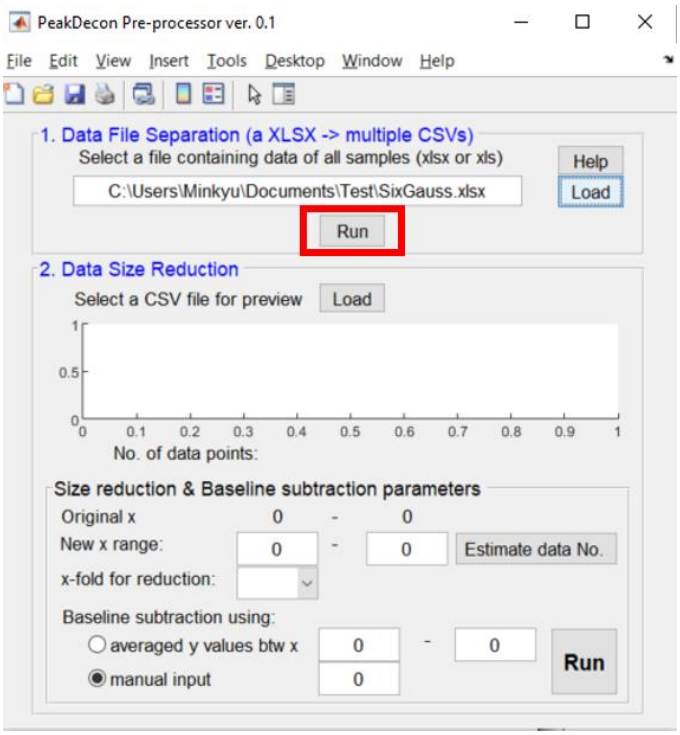
- Click the ‘Load’ button to load the data file.



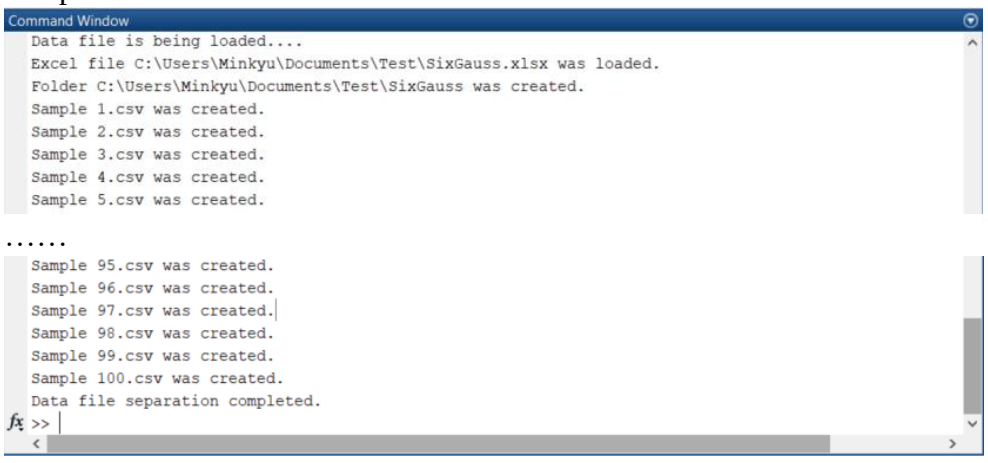
- Select the data file (SixGauss.xlsx) and click ‘Open’.



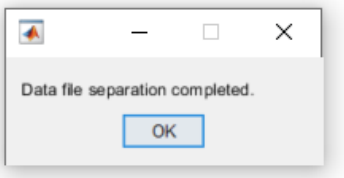
- Then, the data file location is shown in a text box. Click ‘Run’ to separate the xlsx file into CSV files.



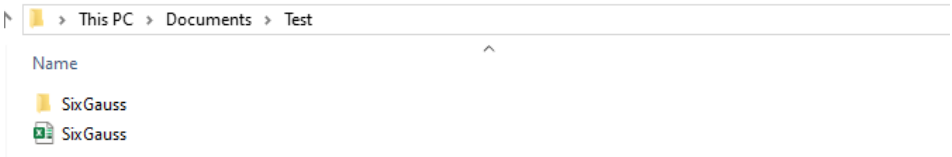
- The text log is displayed on the command window of MATLAB. Once the separation process is completed, you can see ‘Data file separation completed’ in the command window.



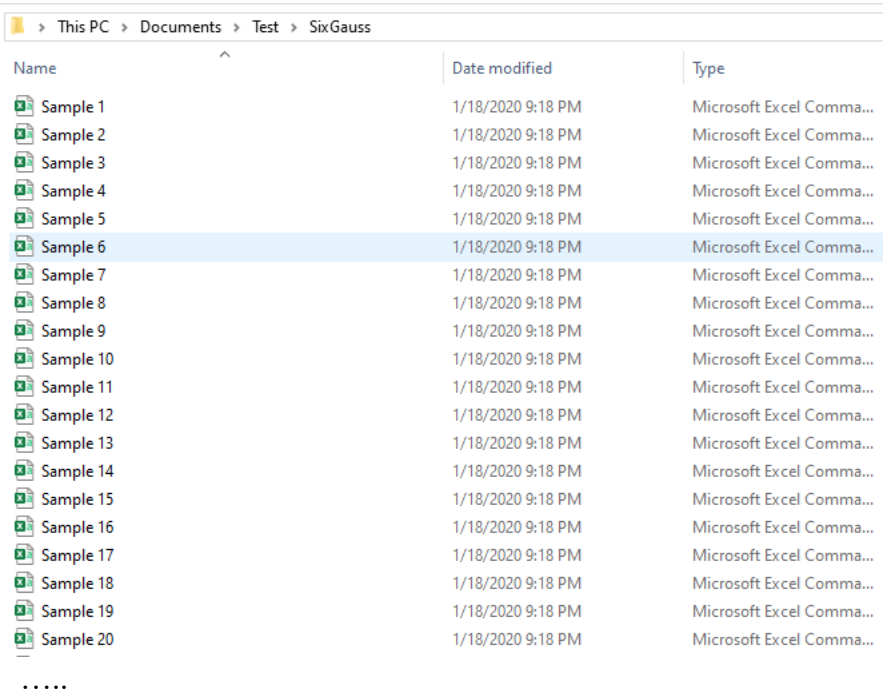
- Also, a message box notifies the completion.



- The separated files are stored in the folder with the same name as the master xlsx file.

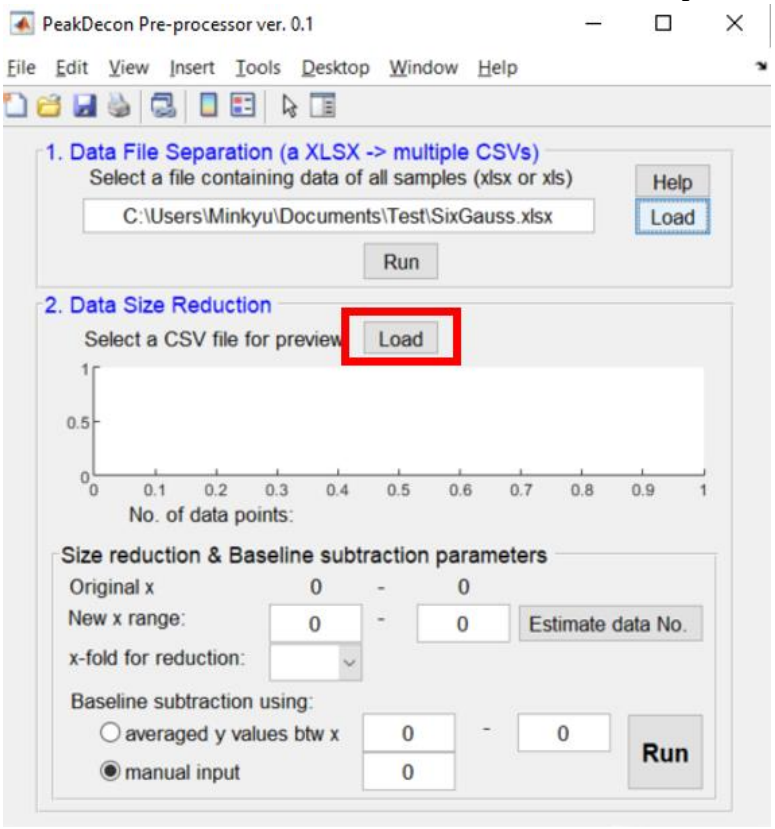


- As shown below, new CSV files are created in the ‘SixGauss’ folder. It is noted that the names of CSV files are assigned from the header of the xlsx file.

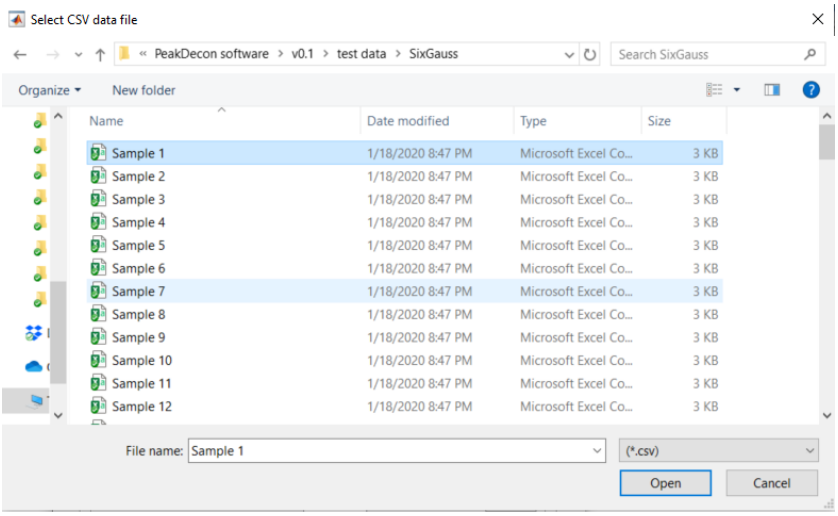


…..

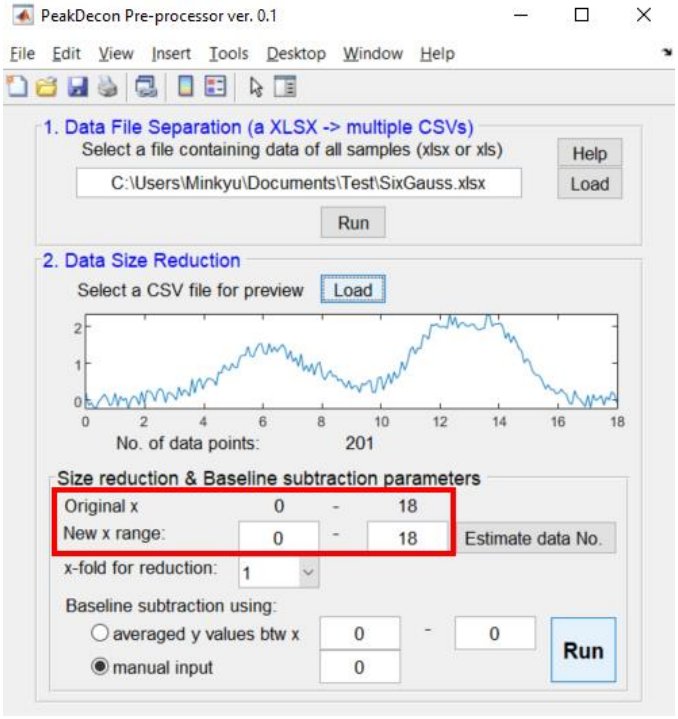
- As the next step, the size of data will be reduced to half. To this end, click ‘Load’ to load a CSV file. It is noteworthy that all the data files in the same folder where the selected file is located will be processed together.



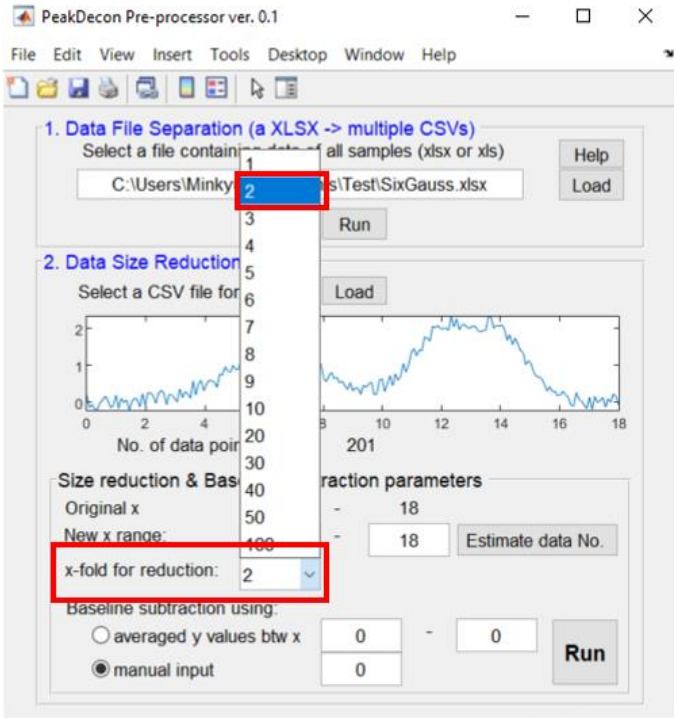
- Select one of the data files. Any file is okay. In this tutorial, ‘Sample 1.csv’ file is selected.



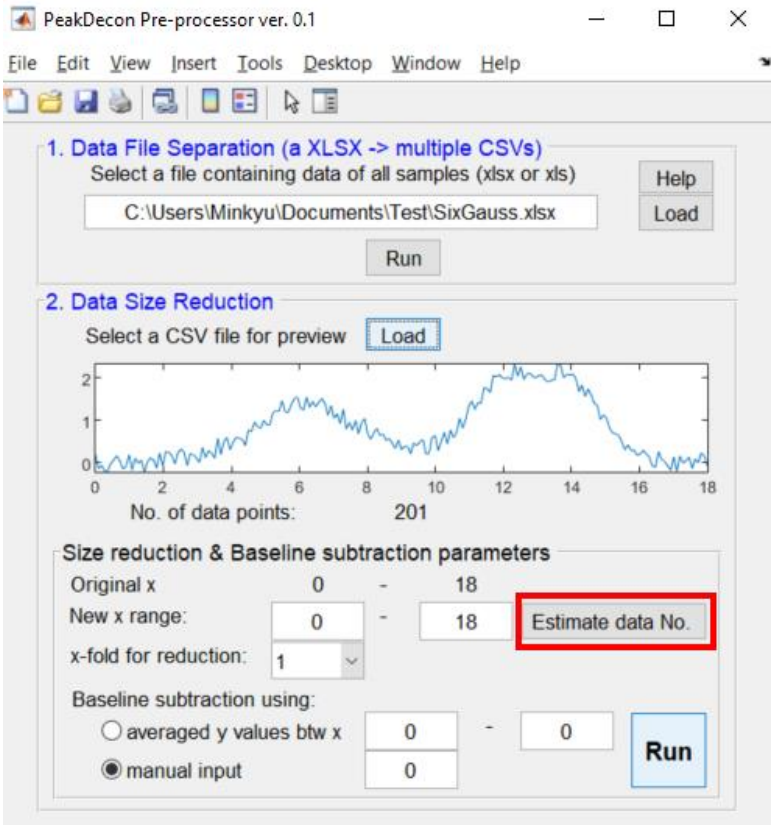
- Once a data file is selected, its data is visualized in the subfigure. Also, the range of the x-axis is automatically displayed. A specific range of x-axis can be selected by entering values for ‘New x range’, but this is not applied in this tutorial.



- In order to reduce the data size to half, select ‘2’ in ‘x-fold for reduction’.

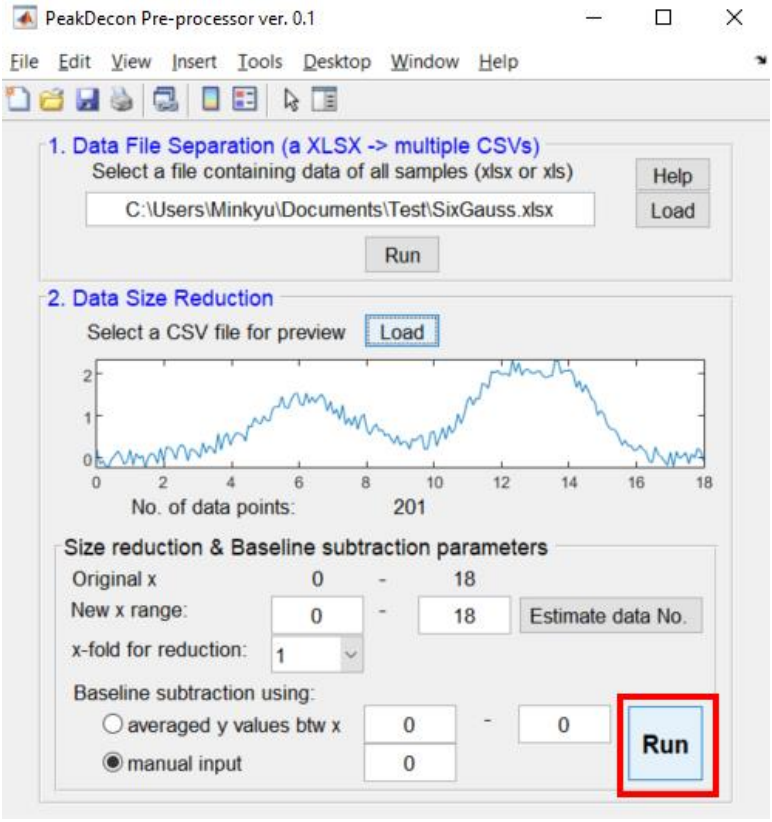


- Click ‘Estimate data No.’ button to see the estimated number of data points if the fold change is applied.

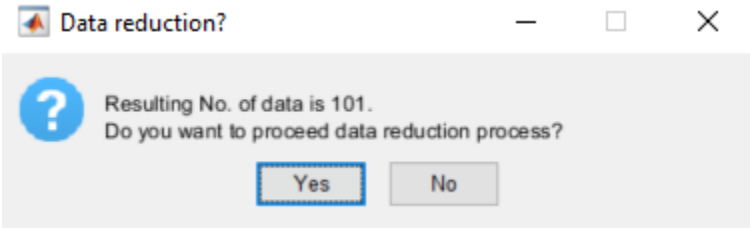


The following message box shows the estimated number.

- Click ‘Run’ to reduce the size of the data.



- Click ‘Yes’ to proceed with the reduction process.



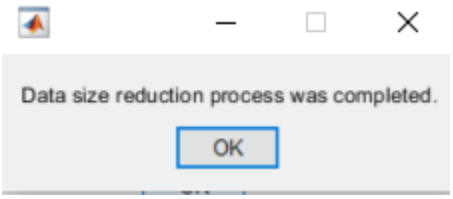
- Progress is displayed in the MATLAB command window and print ‘Data size reduction process was completed’ once the reduction process is done.

```
------------------------------------------------------------
D:\Dropbox\My_PhD\Paper_work\4_SEC FLD NMF\PeakDecon software\v0.1\test data\SixGauss_0_18_N101 is created.
------------------------------------------------------------
Sample 1.csv was created.
Sample 10.csv was created.
Sample 100.csv was created.
Sample 11.csv was created.
Sample 12.csv was created.
```

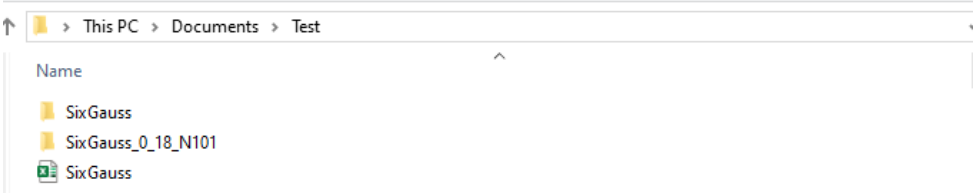
.....

```
Sample 95.csv was created.
Sample 96.csv was created.
Sample 97.csv was created.
Sample 98.csv was created.
Sample 99.csv was created.
------------------------------------------------------------
Data size reduction process was completed.
fx >>
```

Also, a message box notifies the completion too.

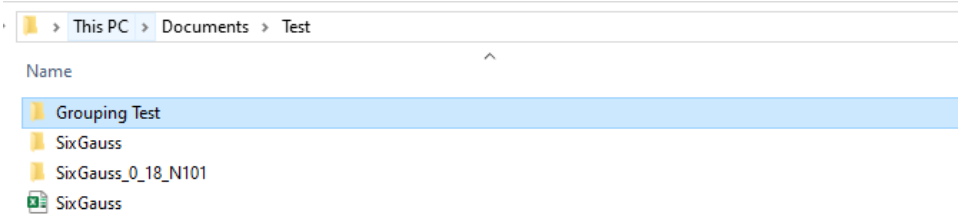


- The software automatically generates a folder whose name is composed of the original folder location and additional information about the reduction process. For example, “0_18_N101” means that the range of x-axis is from 0 to 18 and the number of data points is 101.

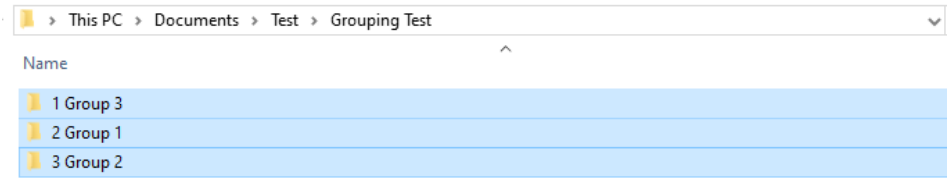


- Before moving onto ***guiPD*** (i.e., the main deconvolution process), the rearrangement of data is needed to include grouping information. Grouping can be done by dividing files into subfolders. As explained in **Guide for *guiPD***, files can be grouped by placing them in a subfolder. Let’s follow the subsequent example.

- New data files with reduced sizes were placed in the ‘SixGauss_0_18_N101’ folder. We will separate them into new folders. To this end, make a folder ‘Grouping Test’ in the same folder location where the folder ‘SixGauss_0_18_N101’ is located.

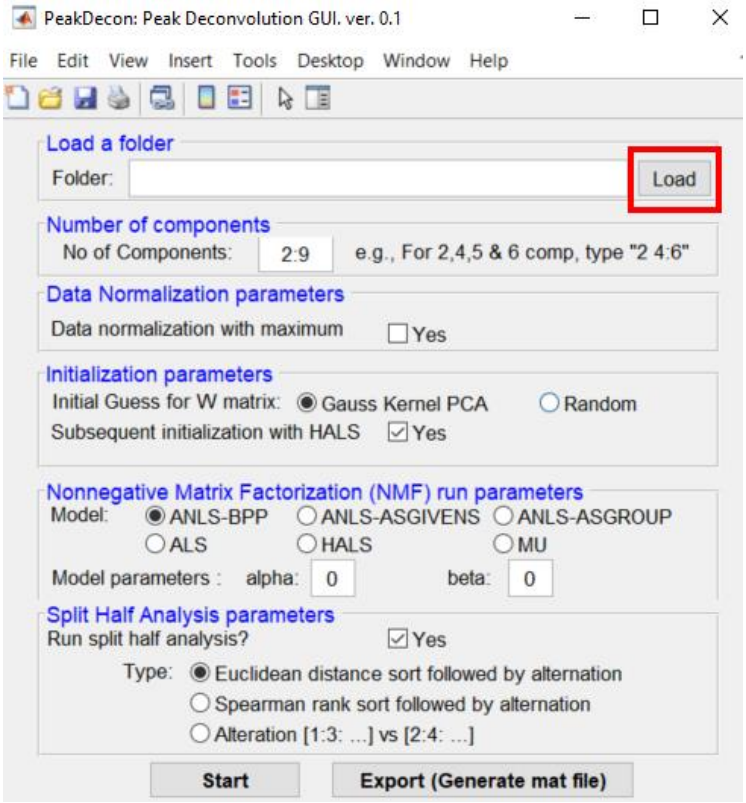


- Enter the ‘Grouping Test’ folder and make three new folders: ‘1 Group 3’, ‘2 Group 1’, and ‘3 Group 2’. The reason that the disagreed numbers at the beginning of the folder names with the actual group numbers is that MATLAB reads files and folders in the ascending order of letters and numbers. Therefore, the disagreed folder names are intended to load Group 3 first, followed by Group 1 and Group 2. This causes the unintended orientation of visualization when using ***guiPost*** if one wants to display results in the order of Group 1, 2, and 3. Hence, this tutorial simulates a scenario that does not have the order of groups as intended. Later, we will change the order of groups when visualizing them using ***guiPost***.

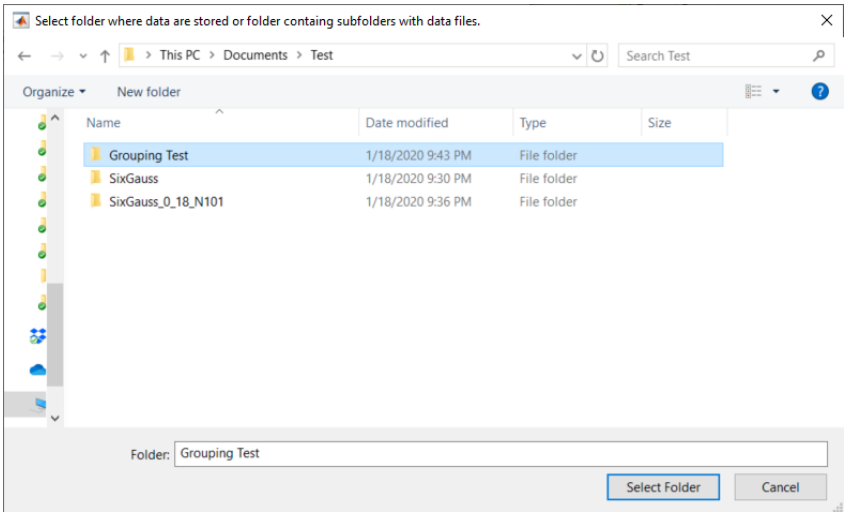


- Run ***guiPD*** by entering ‘guiPD’ in the MATLAB command window. It is noted that all the MATLAB windows will be closed and variables will be cleared when running ***guiPD***.

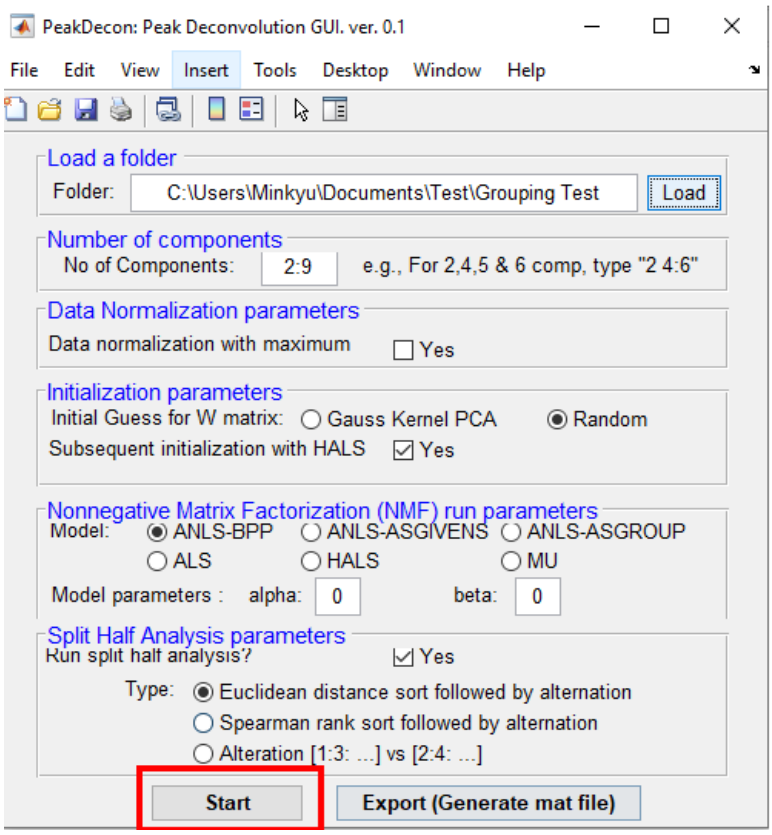
- Load a folder containing data files. Click ‘Load’ on the GUI of PeakDecon,



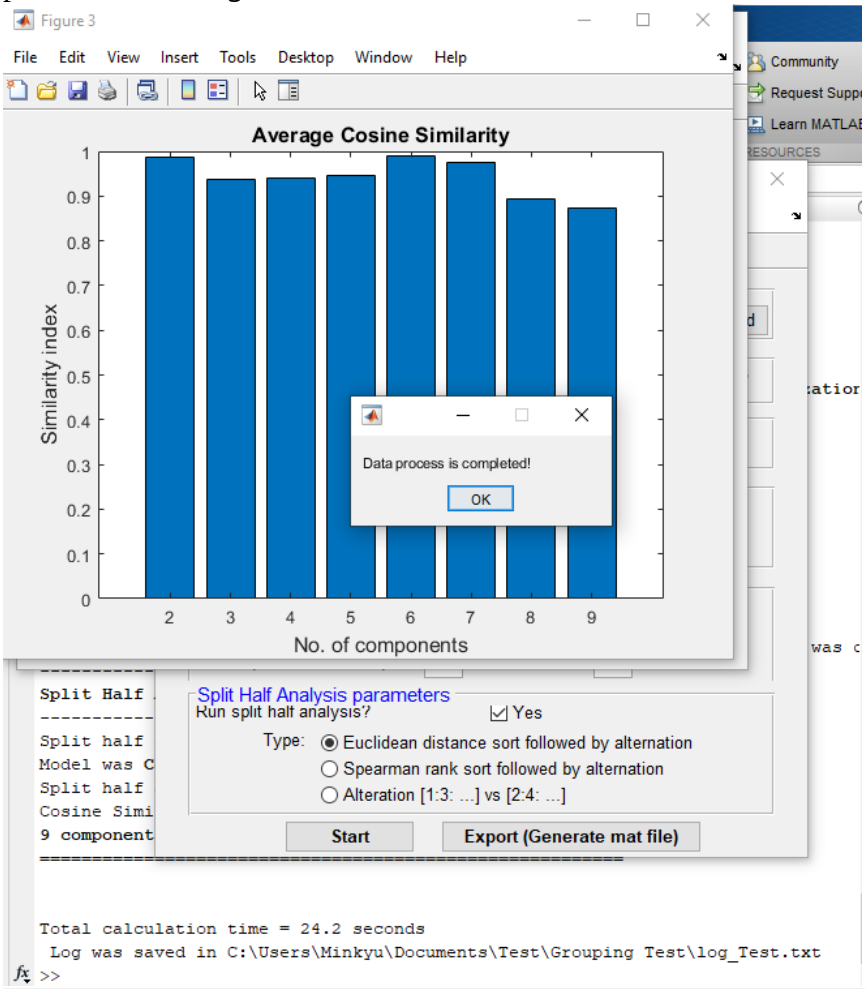
- Select the folder ‘Grouping Test’ that contains the three folders generated earlier. Click ‘Select Folder.’

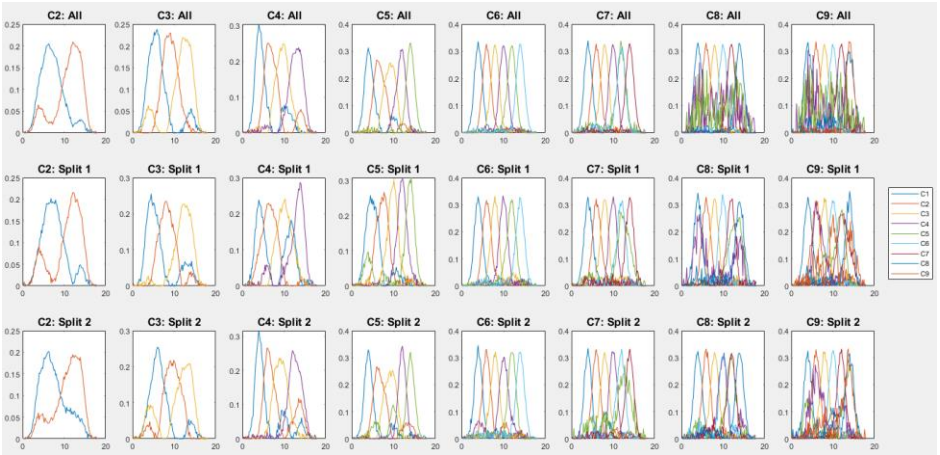


- This tutorial runs the deconvolution algorithm with the default setup. That is, the number of components ranges from 2 to 9. The normalization process is not performed. The ANLS-BPP is used as the main deconvolution algorithm with the initial values of *W* and *H* matrices calculated with the HALS. Hence, the order of running is the HALS with random initialization followed bu the ANLS-BPP. The split half analysis with the alternation of data sorted based on the Euclidean distance is conducted. Without changing parameters, click 'Start' with the default setup.

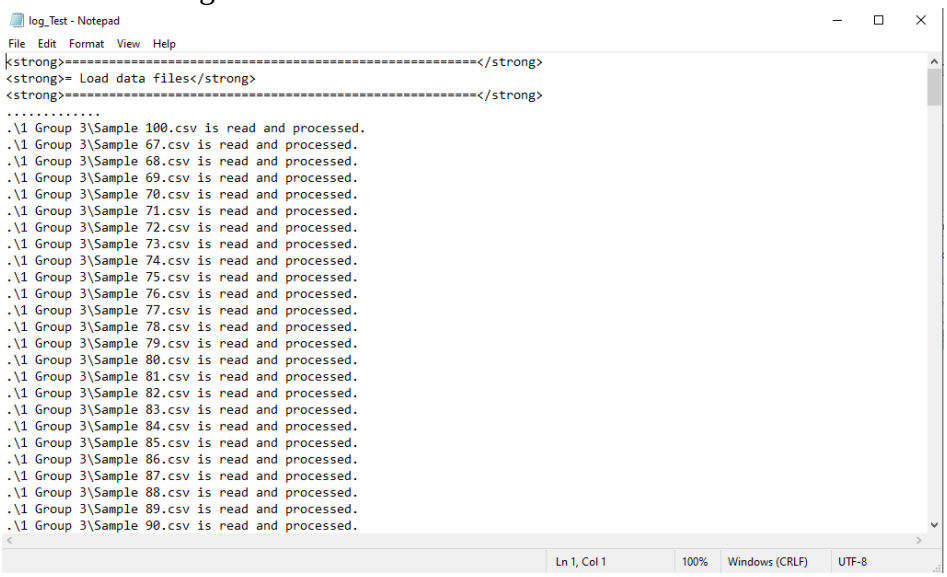


- Once the model run is completed, the average cosine similarity values of 2-9 component models are displayed on a figure. Also, the split half analysis is shown as a graph. The log of the run is displayed on the MATLAB command window. You can close the graphs since they can be plotted later with ***guiPost***.

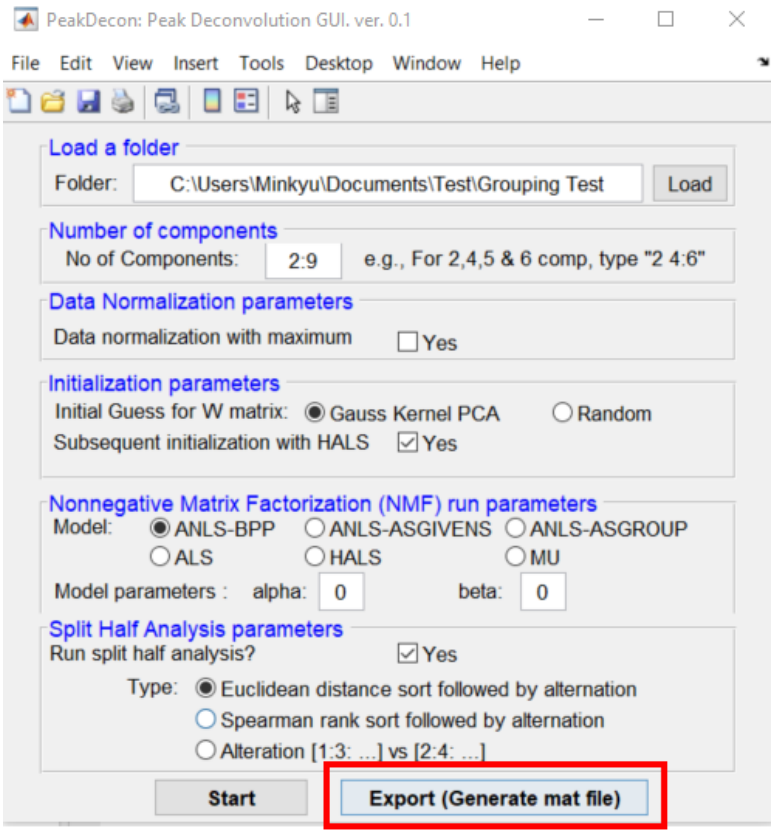




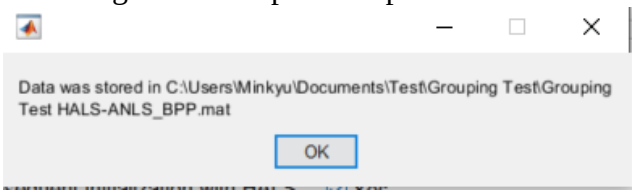
- A text file of log in the same folder selected.



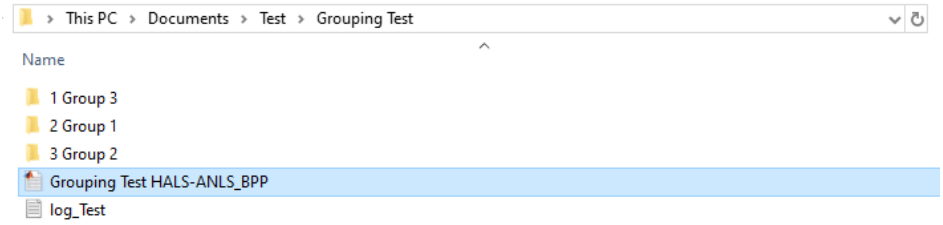
- Click ‘Export (Generate mat file)’ to save the processed data. The created mat file can be used for ***guiPost***.



- A message box of export completion is shown.

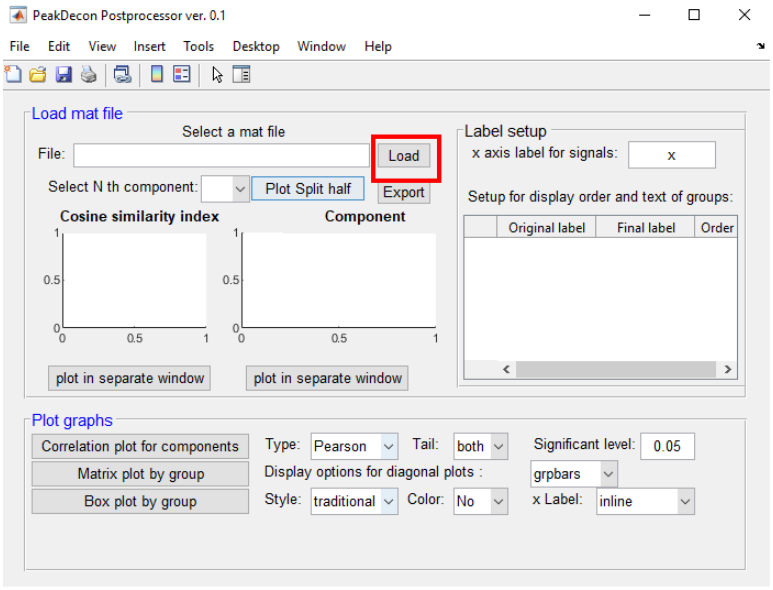


- The created file is located in the same folder selected for the deconvolution.

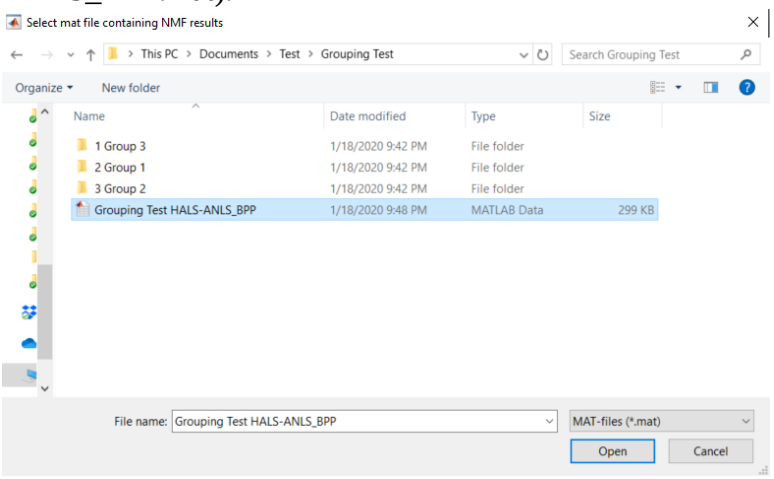


- The next step is the visualization of the run models. To this end, run ***guiPost*** by entering 'guiPost' in the MATLAB command window.

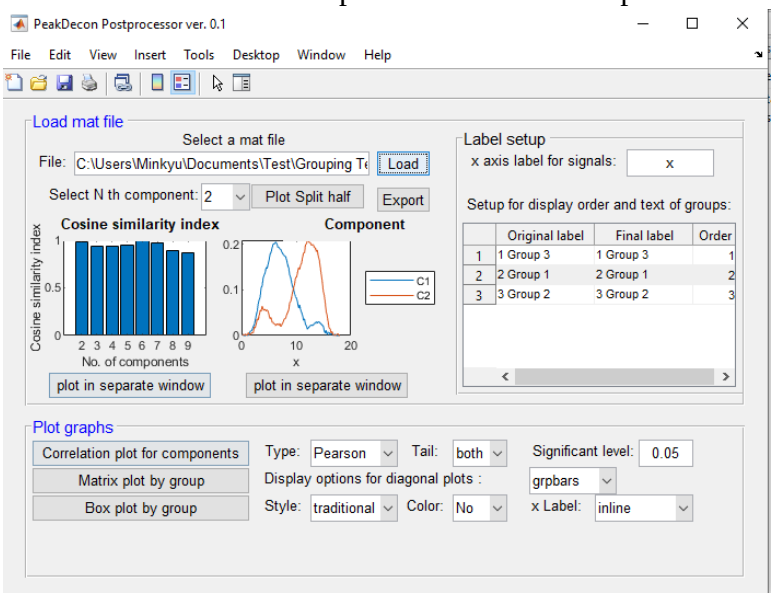
- Click 'Load' on the ***guiPost*** window.



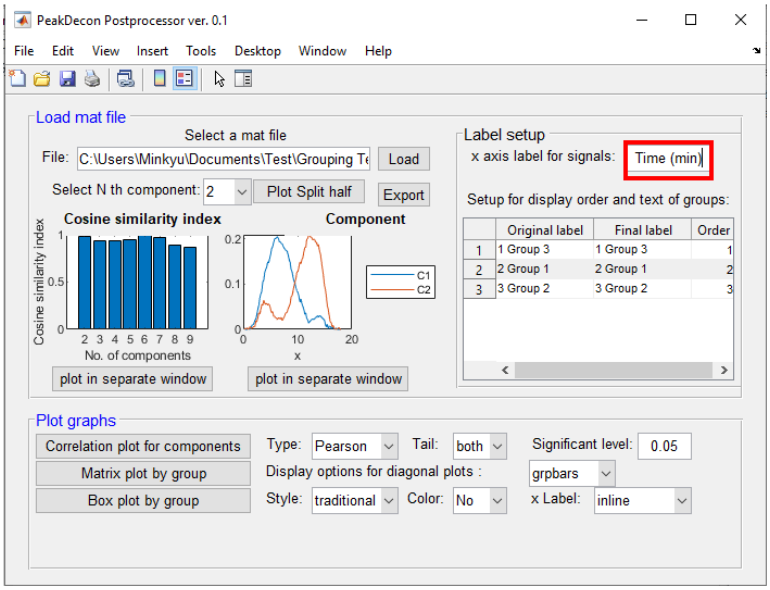
- Select the mat file saved earlier (i.e., Grouping Test HALS-ANLS_BPP.mat).



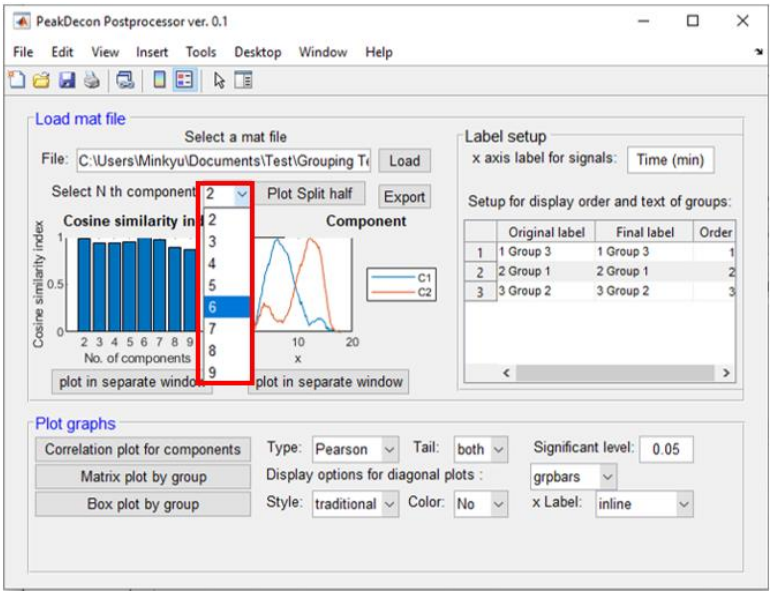
- Loading the data file displays the cosine similarity index values of the model results and the components for the 2-component model.



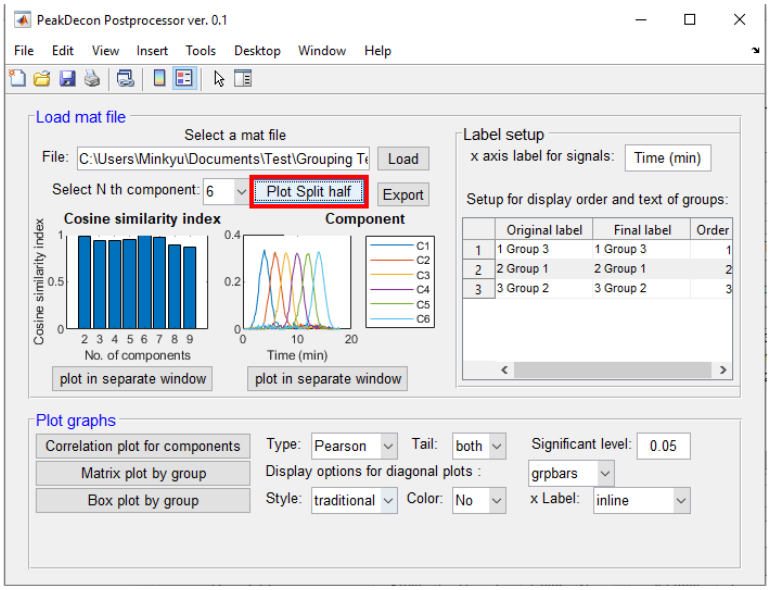
- Since the data files do not have the label for x axis, it is entered manually for visualization in ***guiPost***. Since the data used is assumed to be a chromatographic data, 'Time (min)' is entered in the text box of 'x axis label for signals.'



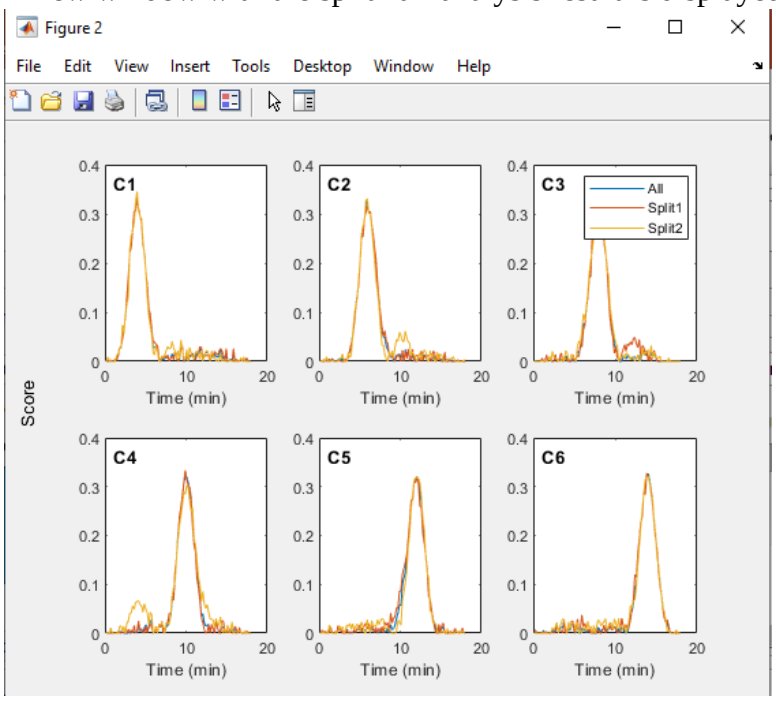
- Here, the cosine similarity index of the 6-component model is the highest while the 2-component model has a similar value. Since the data were synthesized by combing 6 unimodal peaks, the highest cosine similarity value of the 6-component model suggests that the split-half analysis with cosine similarity assessment can be a reliable means of cross-validation. In order to further evaluate the 6-component model, click the list box of 'Select N th component' and select '6'.



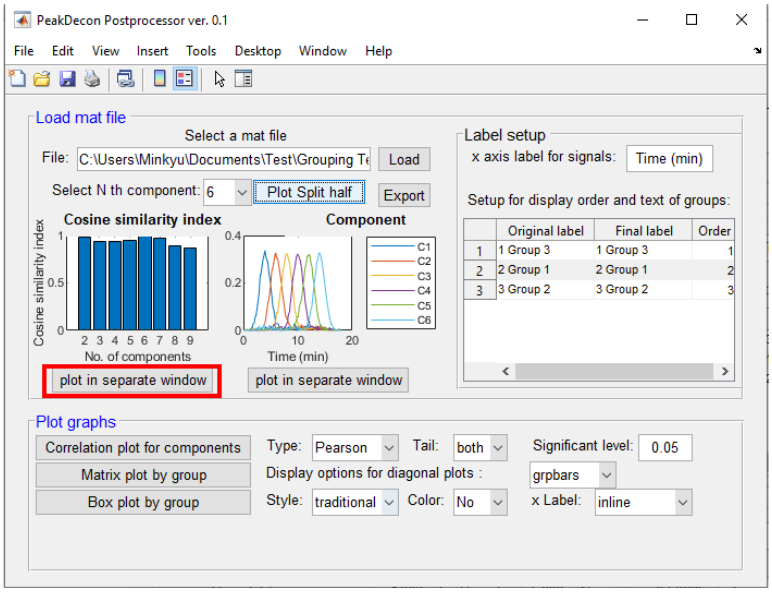
- The selection of the 6-component model automatically changes the subfigures for the cosine similarity and the components. Click ‘Plot Split half result’



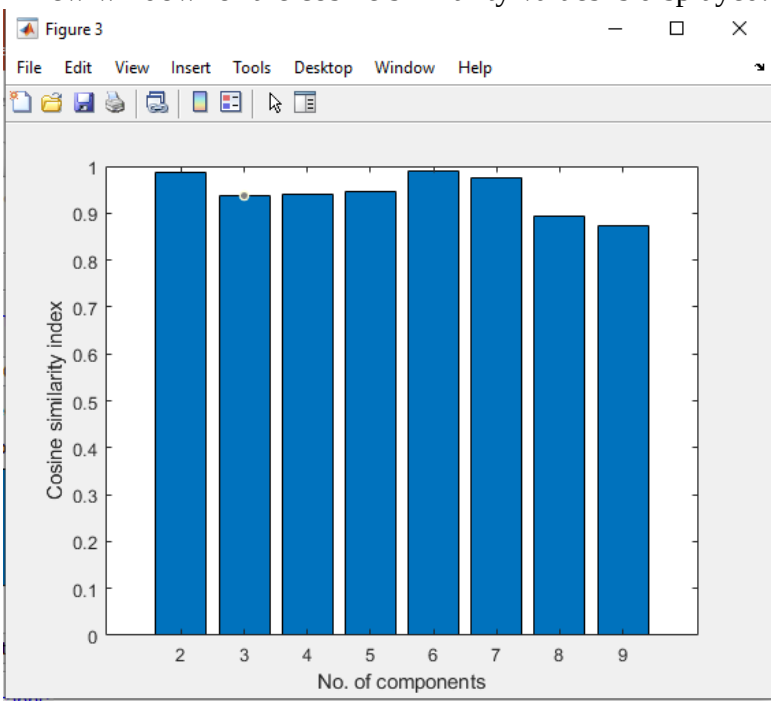
- A new window with the split half analysis result is displayed.



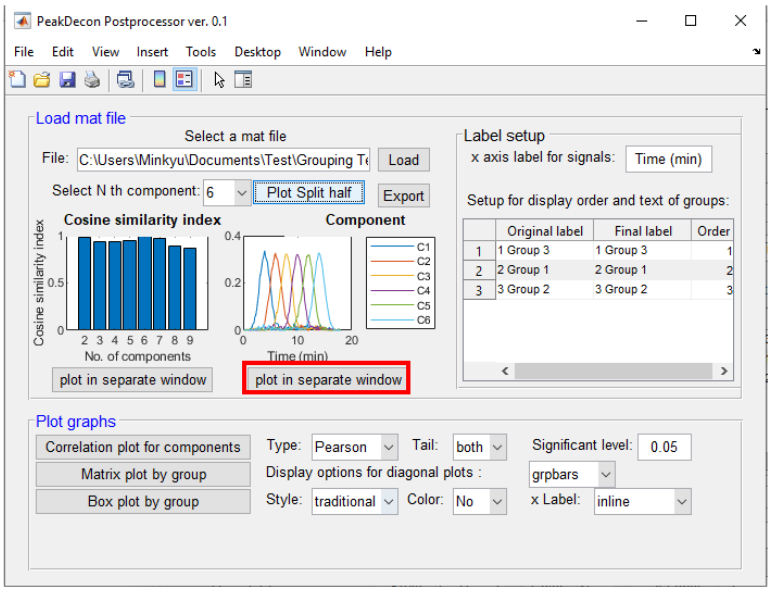
- Click ‘plot in separate window’ for the cosine similarity.



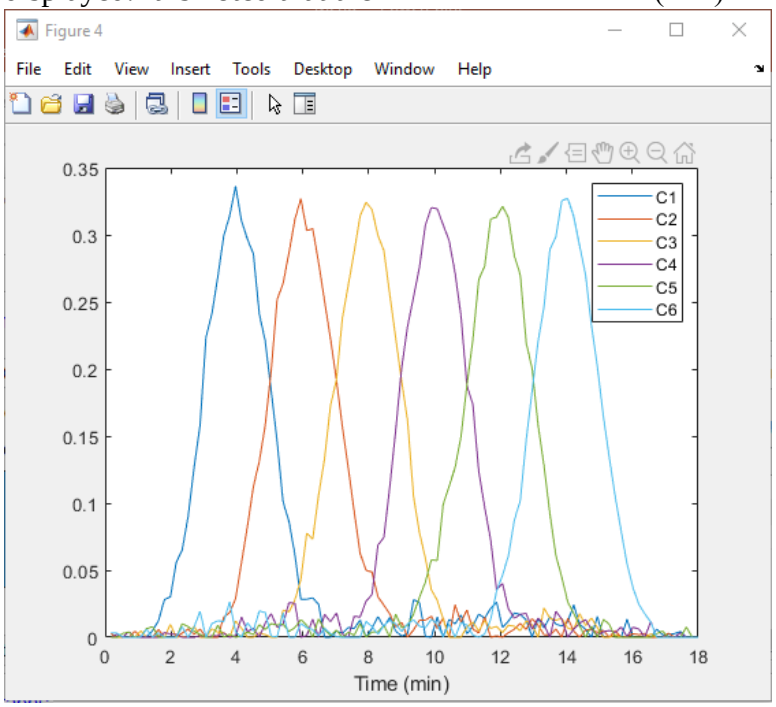
- A new window for the cosine similarity values is displayed.



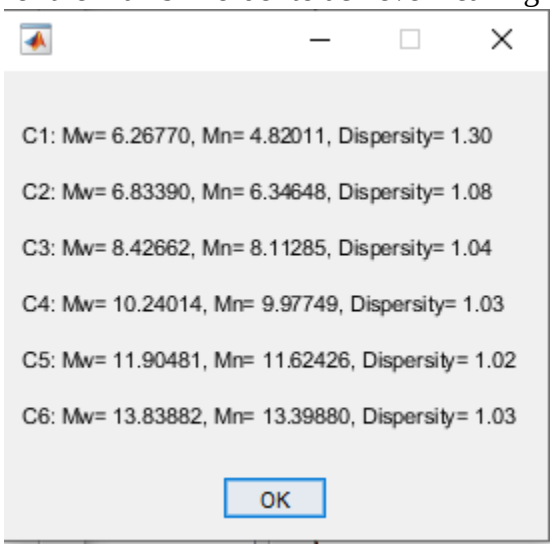
- Click ‘plot in separate window’ for the component.



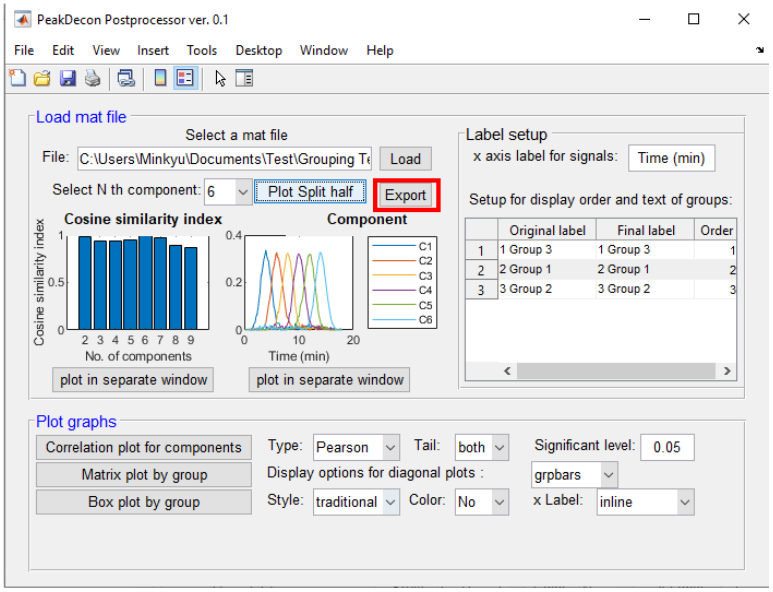
- A new window of the components for the six-component model is displayed. It is noted that the x-axis label is ‘Time (min)’ as set up earlier.



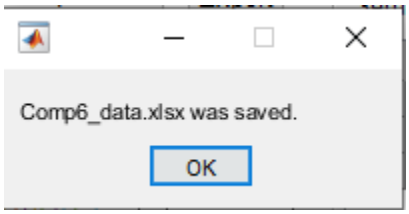
You can also see a message box displaying the values of M_w (weight molecular weight), M_n (number molecular weight), and dispersity index (M_w/M_n). The example data here are synthetic, and time (min) was allocated to the x-axis arbitrarily. Therefore, the values in the message box are not meaningful. Since the calculation of the two molecular weight depends heavily on x-axis, one should use the appranet molecular weight for the x-axis in order to achieve meaningful values of M_w and M_n.



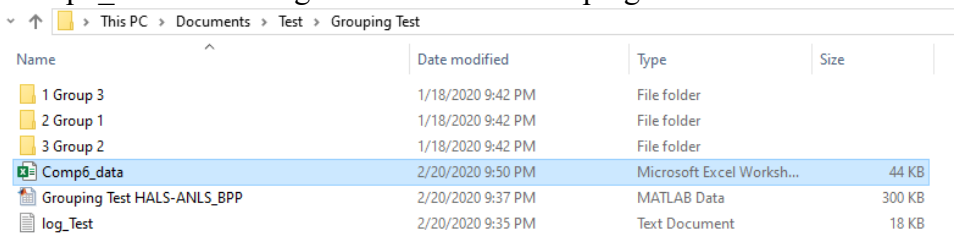
- Click ‘Export’ to export the results of the 6-component model.



- A message box is poped up for showing the completion of the export.



- Comp6_data.xlsx was generated in the ‘Grouping Test’ folder.

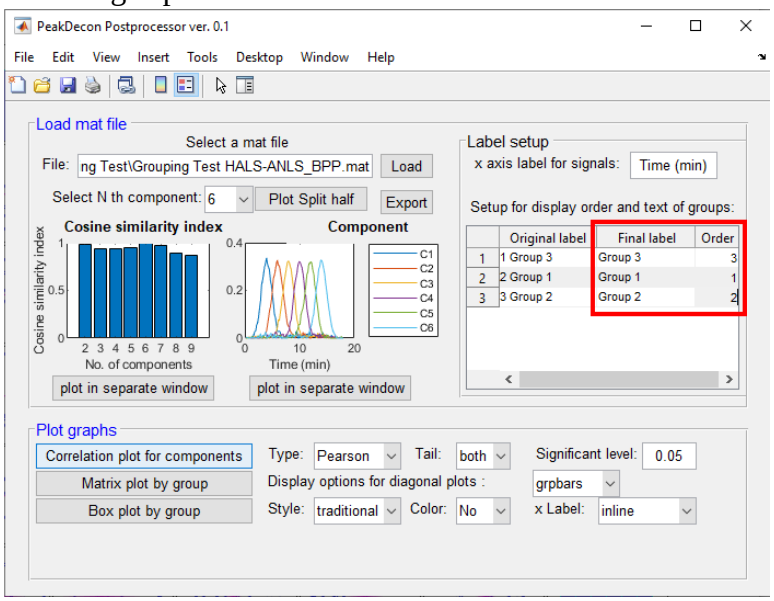


- The saved data contains *W*, *H*, and the area of components for samples.

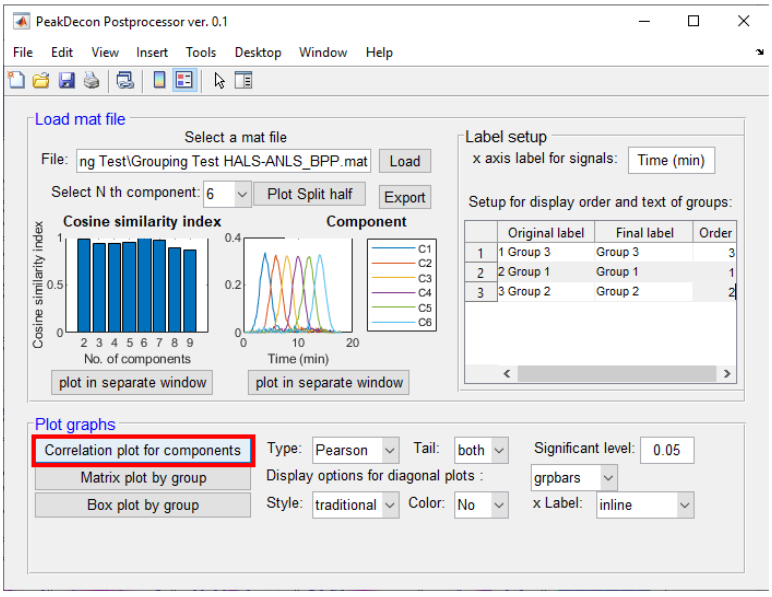
	A	B	C	D	E	F	G
1	Sample	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6
2	Sample 100	4.887877098	2.055293073	2.361442746	4.116906148	4.794527365	3.891212256
3	Sample 67	2.627067281	3.610778229	3.153518587	3.17699708	0.184355706	0.213148622
4	Sample 68	3.109516775	0.39327145	4.503341152	1.136742039	1.447111232	0.870804846
5	Sample 69	2.827318977	3.52409978	0.320654728	3.697326612	4.320225563	1.698547248
6	Sample 70	3.543349129	4.217259047	4.85930024	4.181735628	1.574054073	4.184892753
7	Sample 71	1.10167071	3.056749504	0.232610583	3.671368028	5.275469595	0.839075469
8	Sample 72	4.784042451	0.521954038	0	1.106015803	3.875279972	3.641418416
9	Sample 73	1.810395055	3.886088563	4.260634411	2.968215005	1.729529737	3.248540994
10	Sample 74	1.443408486	3.983638607	0.797639118	2.532367302	2.865186318	3.470857722
11	Sample 75	0.247379658	0.581638479	0.958318868	4.928030993	0.707607099	4.632170058
12	Sample 76	2.097260313	1.345980606	2.262988319	4.02676652	0.532260831	3.998241335
13	Sample 77	2.368213428	0.321770345	0.379139922	3.484078404	3.716503671	3.167654781
14	Sample 78	2.828376069	1.1686116	1.774693365	1.615292462	4.385646745	4.516534689
15	Sample 79	4.486031142	2.674246292	0.337707215	3.798845709	3.074736463	1.459918556
16	Sample 80	2.682231383	2.325887286	1.713431257	2.105495089	3.494486097	1.616915604
17	Sample 81	3.741649307	4.3107342	0.9002365	3.39793828	0.300963969	1.249084872
18	Sample 82	3.385817774	0.313529146	1.584519408	1.602736335	4.97687101	4.391437399
19	Sample 83	2.114980996	2.864399408	4.968345385	1.080645022	1.416470832	1.084868827
20	Sample 84	1.337842367	4.512027238	0.127945992	2.944509032	2.775451137	3.523368806
21	Sample 85	1.837505025	3.342193171	0.655279801	3.605584717	0	0.492542499
22	Sample 86	3.363436327	4.300698693	4.657304555	0.661588839	2.003312615	1.412154268
23	Sample 87	3.969868061	2.833928125	4.760101015	3.279870997	4.708579668	4.052726888
24	Sample 88	0.054731885	0.02897577	3.82213088	2.227286377	4.850530801	0
25	Sample 89	0.939155808	4.661257443	3.679601884	0.62110872	1.029229896	0.177407682
26	Sample 90	0	3.222272049	3.799703343	4.489798889	2.465139984	4.070682977
27	Sample 91	3.091888599	1.245006515	2.648446748	4.751532577	3.901848566	2.211013825
28	Sample 92	4.787609816	2.941359446	4.054739433	0	1.303630849	3.721254638
29	Sample 93	1.482509227	5.080065177	0.489335265	1.5465806	4.868081345	2.599151727
30	Sample 94	3.586472679	1.856394567	1.25786378	2.590689165	4.763630179	4.768067856
31	Sample 95	3.375742718	4.951581609	3.967276533	2.520277478	2.671977027	2.13419196
32	Sample 96	4.443508879	5.079103099	1.46612241	1.474231158	0.217461187	2.311233766
33	Sample 97	1.339924164	1.405008374	1.593461456	2.785068375	1.311591027	1.243899191
34	Sample 98	1.207328802	4.83971112	1.287389507	0.883161853	1.349717775	4.3102954
35	Sample 99	0	4.3889105	5.024553726	1.514235754	3.378849914	2.986073777
36	Sample 1	0.624126907	3.278711089	1.264194147	0.022412503	4.631619691	4.467095309
37	Sample 10	4.020143041	2.219447751	4.537740528	3.222188209	1.347923921	4.514017636
38	Sample 11	1.346967642	1.870082122	1.938001574	1.097388355	3.091036875	2.757736823

Sheet1 | W | H | Area

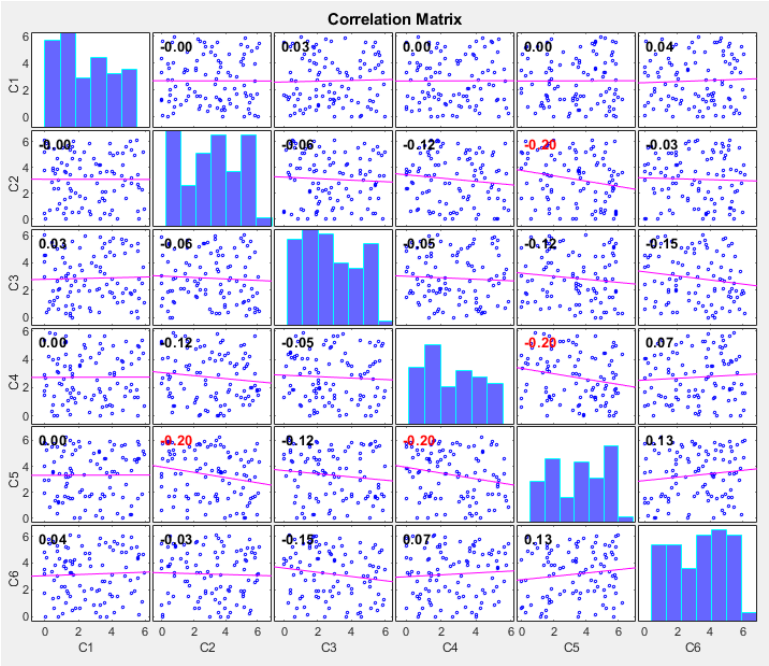
- As the next step, group labels and the display order will be changed. The three folders created earlier are in the order of ‘1 Group 3’, ‘2 Group 1’, and ‘3 Group 2’. Ultimately, we want to display the groups in the order of ‘2 Group 1’, ‘3 Group 2’, and ‘1 Group 3’ with the new names of ‘Group 1’, ‘Group 2’, and ‘Group 3.’ Change the cells of the table as shown in the following capture.



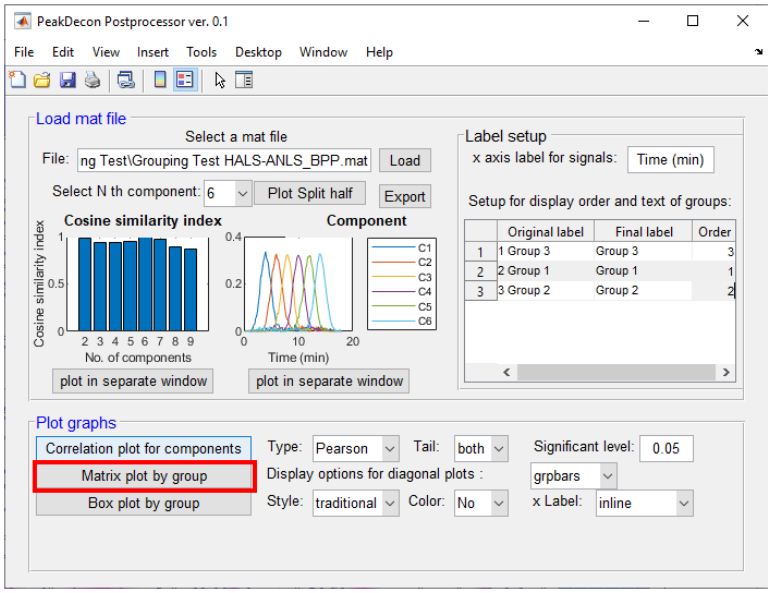
- Subsequently, click ‘Correlation plot for compounds’.



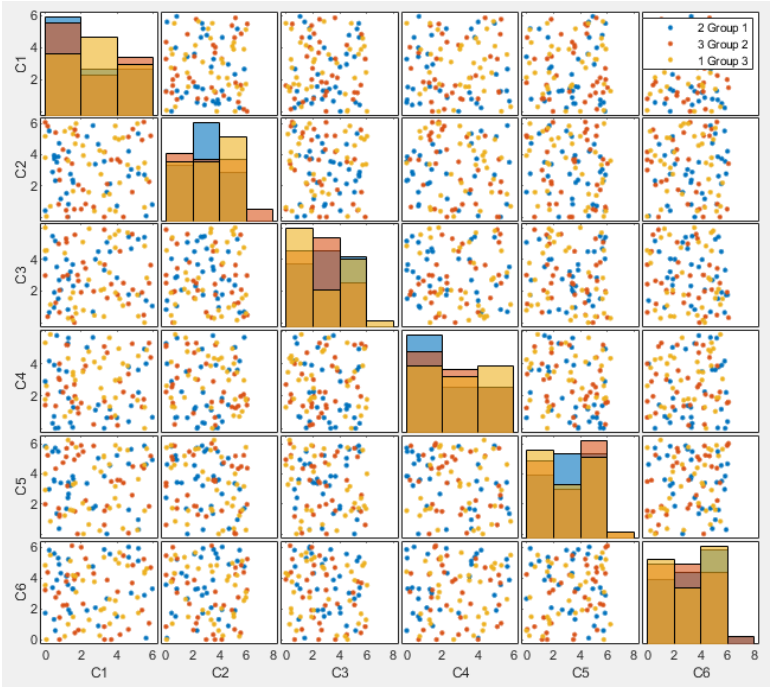
- A correlation plot is created to compare six components. If you want to change the statistical parameters such as the type of correlation and tail, change them accordingly. MATLAB's online manual for ‘corrplot’ is helpful.



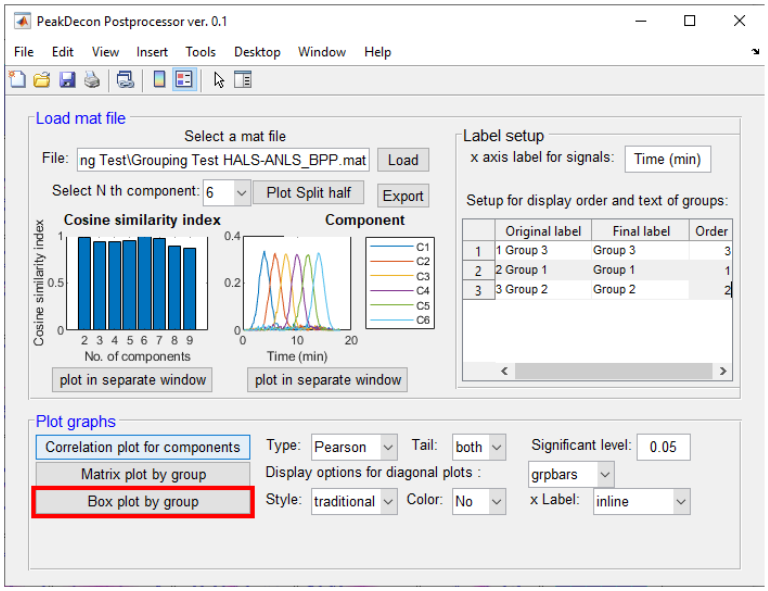
- Click ‘Matrix plot by group’.



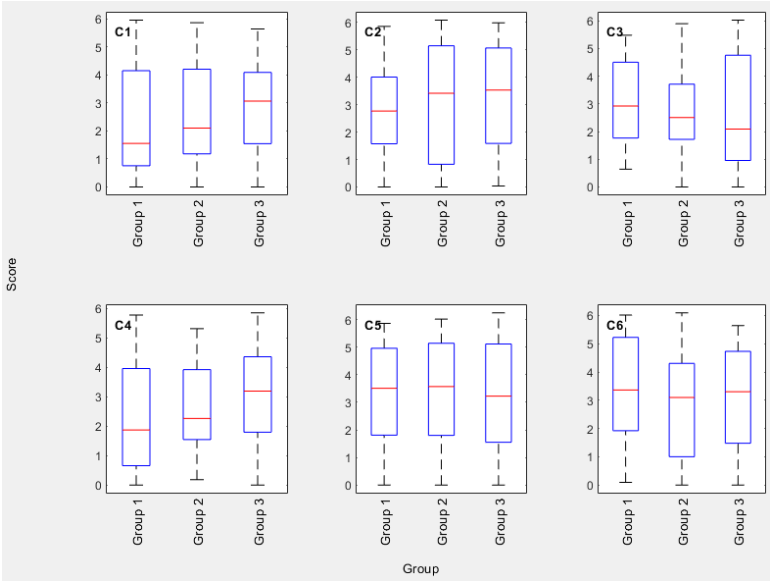
- This plot method is similar to the correlation plot. However, the comparison among groups is more facile. The diagonal plot type can be changed. MATLAB's online manual for ‘gplotmatrix’ is useful.



- Click ‘Box plot by group.’ The style, color, and x label of the box plot can be changed. MATLAB's online manual for ‘boxplot is useful.



- This box plot is helpful to assess the trend of each component change with respect to the group. It is noted that the order of group is 1, 2, and 3 as set previously.



- Now, enjoy PeakDecon with your data!!!