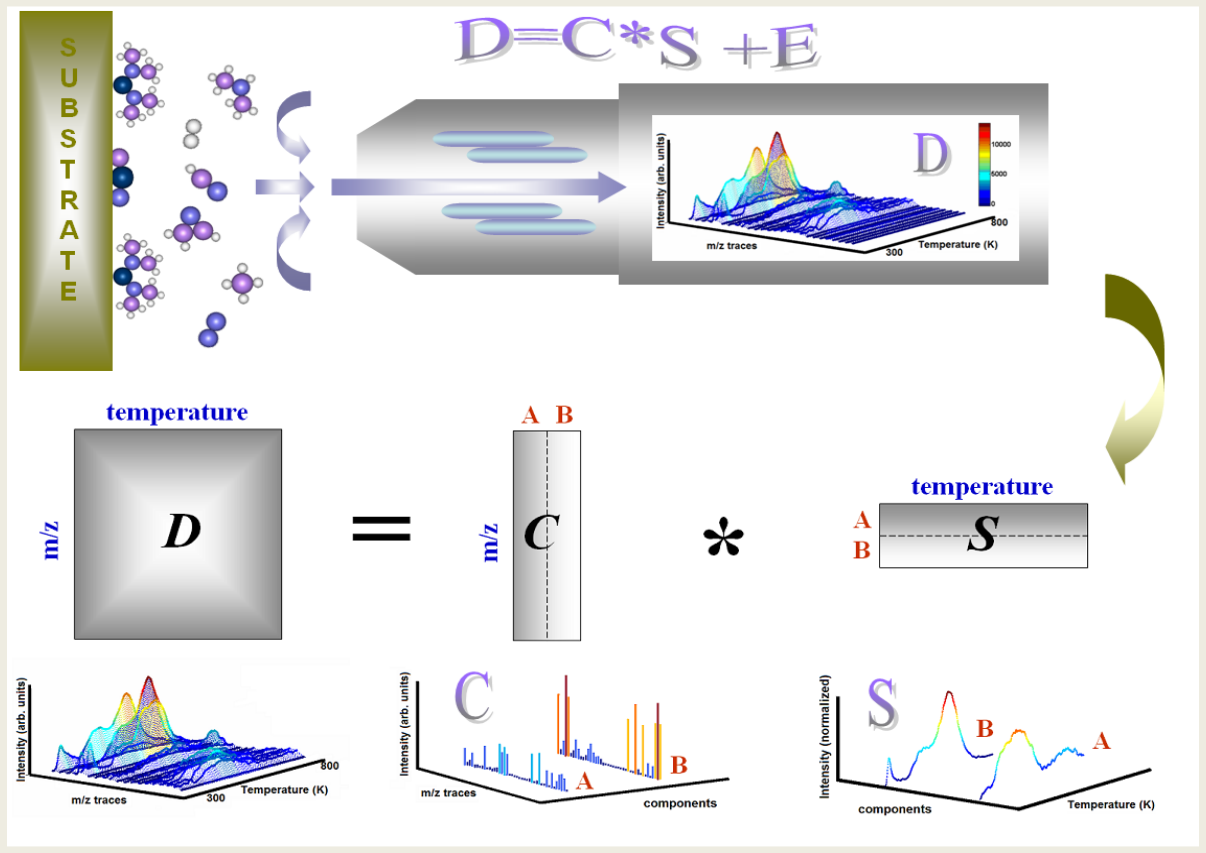
Guide for performing data analysis of mass spectrometric data from temperature-programmed techniques using multivariate analysis



Jing Zhao, Jia-Ming Lin and Andrew V. Teplyakov

Department of Chemistry and Biochemistry, University of Delaware, Newark DE 19713

Juan Carlos F. Rodriguez-Reyes

Department of Industrial Chemical Engineering, Universidad de Ingeniería y Tecnología UTEC, Avenida Cascanueces 2221, Lima 43, Perú

INTRODUCTION

Let´s consider a temperature-programmed desorption/reaction experiment where a Si(100) surface was saturated with ethylchloride-d5 (CD_3CD_2-Cl) at 300 K. The data is shown in Figure A. Usually the information carried by the data is interpreted based on previous knowledge or through comparison with patterns of pristine compounds (to determine the identity of a desorbing compound) or by isolation of individual patterns by subtracting scaled desorption spectra. This manual shows a mathematical procedure for decomposing the data (a ***m x t*** matrix, where ***m*** spectrometric traces are followed over ***t*** temperature points) into two subsets of data: a matrix ***m x c*** and a matrix ***c x t***. The dimension ***c*** represents no less than the number of compounds desorbing from the surface; information regarding the cracking pattern of each compound is carried in the ***m x c*** matrix and the desorption pattern of each compound is carried in the ***c x t*** matrix. The data in Figure A was separated into four desorbing compounds (c=4) and the results are shown in Figure B. As indicated above, for each of these desorbing compounds we obtain their desorption pattern (left) and their cracking pattern (right).

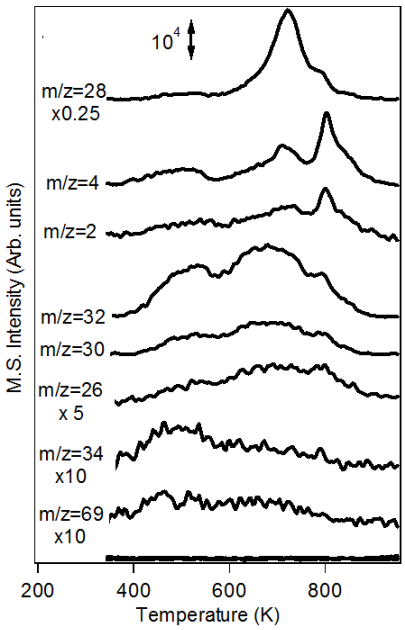


Figure A: Temperature-programmed desorption spectra of ethylchloride-d5 on Si(100) surface. Traces with m/z=2,4,26,28,30,32,34,69 were found with significant peaks from 300K to 900K.

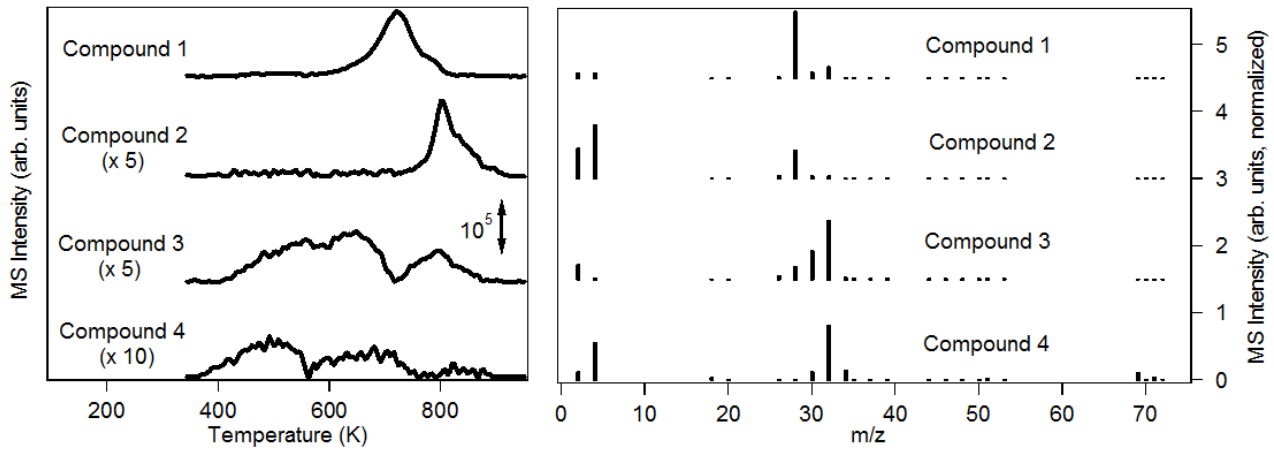


Figure B: Multivariate curve resolution (MCR) of the data presented in Figure A. Assuming the desorption of four compounds, the original data is decomposed into two matrices: the desorption pattern (temperature x 4) and the cracking pattern (4 x m/z).

This guide provides a step-by-step procedure to obtain these results. In addition, a thorough discussion of the application of this technique can be found in Ref [1].

The files used in this guide can be accessed at http://sites.udel.edu/teplyakov/current-research/

Step-by-step procedure

1. Arrangement of the data in excel

Open the file 2014-02-13_dataset_EtCl_on_Si (available at www.teplyakov-data). Notice that the file 2014-02-13_dataset_EtCl_on_Si has several tabs:

- The tab **data** shows the data as it is regularly organized to yield plots (see tab graph_data).
- The tab **matrix** carries the information that will be imported to MATLAB and the PLS_Toolbox. There are no labels for columns and the temperature column has been deleted. The matrix must contain only spectrometric data.
- The tabs **temperature** and **mass** carry the temperature points and the values for m/z that were followed during the experiment. They are going to be inserted as labels in the MATLAB file.

NOTE: When MATLAB imports the excel file it will read the first spreadsheet of the document. For this reason the tab **matrix** should be located first in the excel file.

2. Starting the PLS_toolbox in MATLAB

From the MATLAB window, we need to start the toolbox by typing the command **browse**

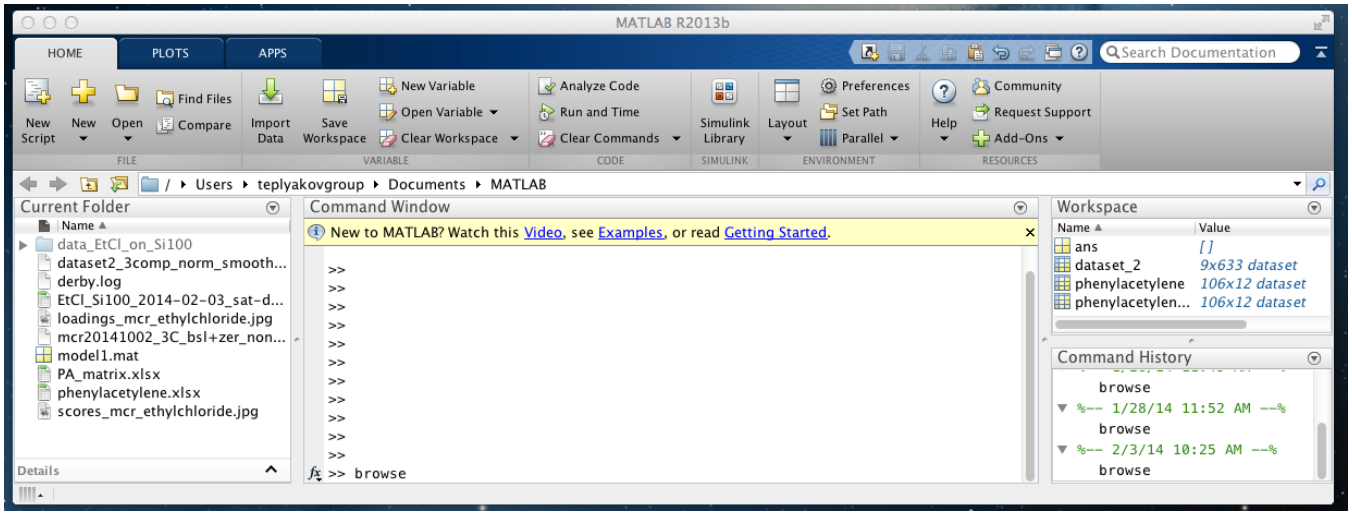


Figure 1

This will open the main window of the PLS_toolbox for MATLAB:

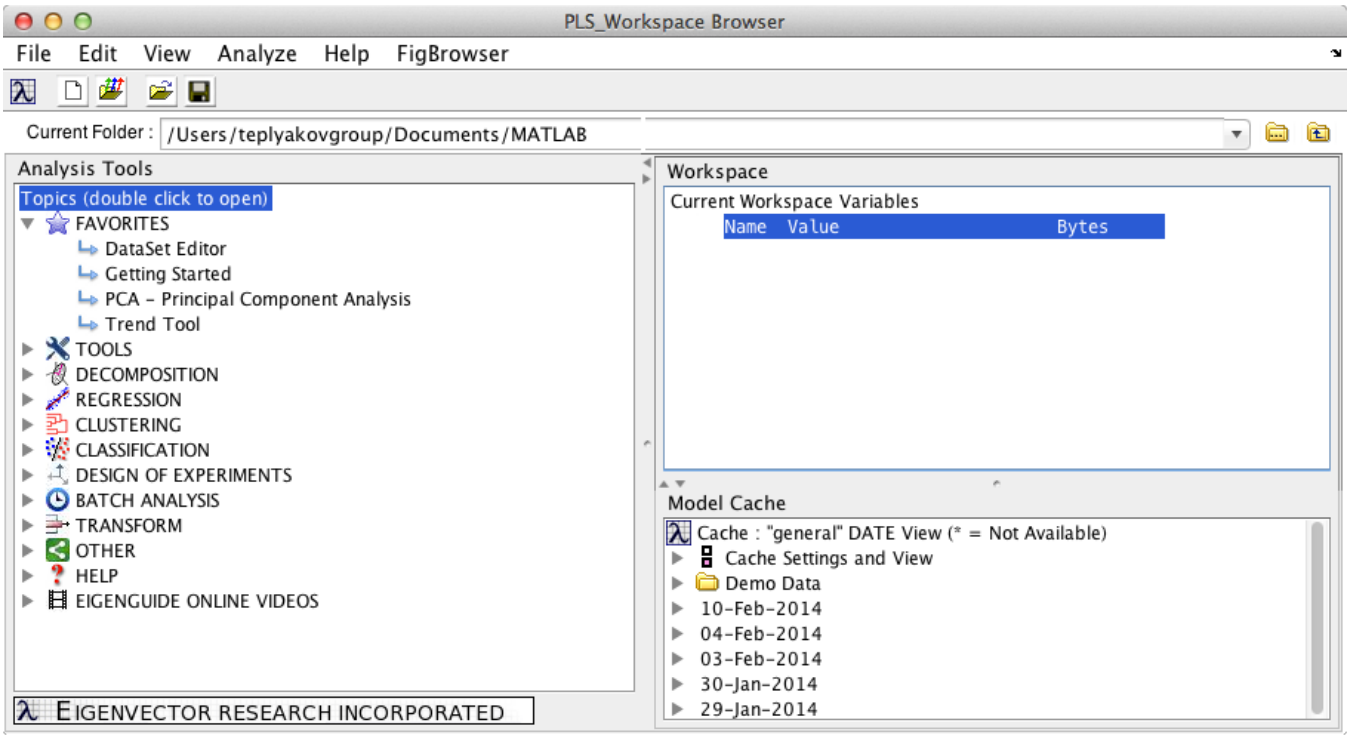


Figure 2

From this window we can select, under analysis tools, decomposition and then MCR,

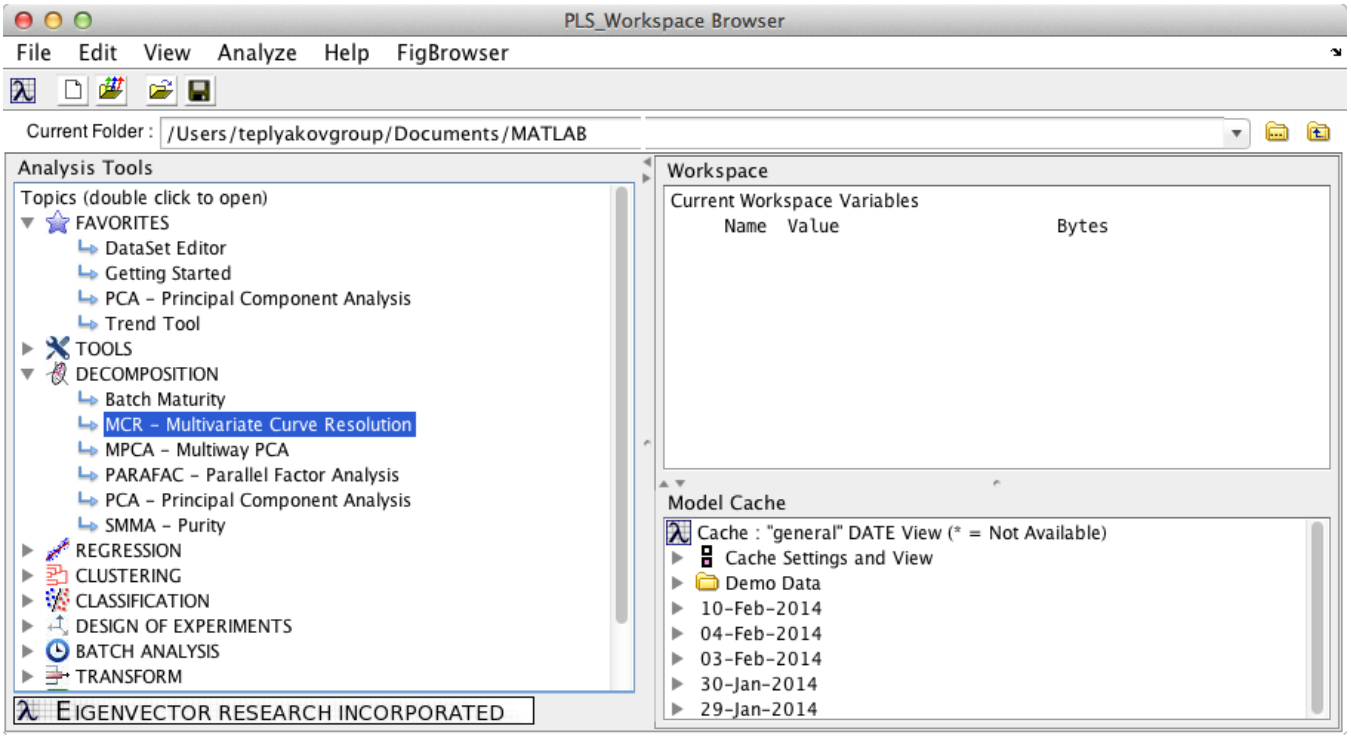


Figure 3

Which will pop-up the window for MCR analysis:

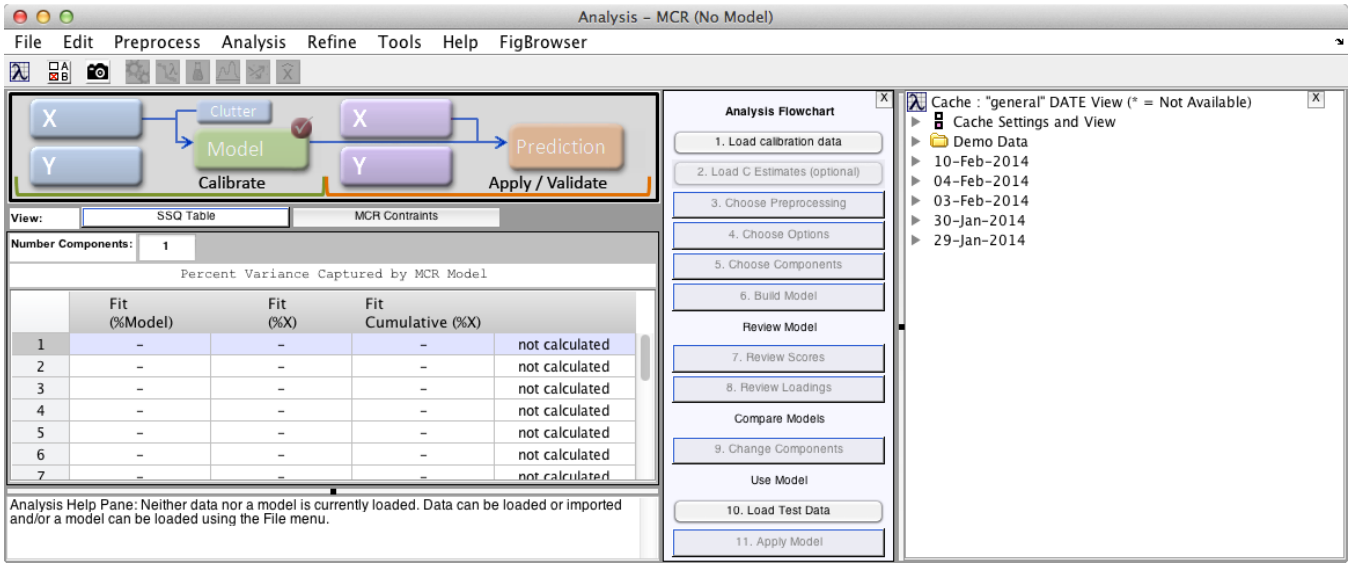


Figure 4

Alternatively, we can type also in the command window mcr

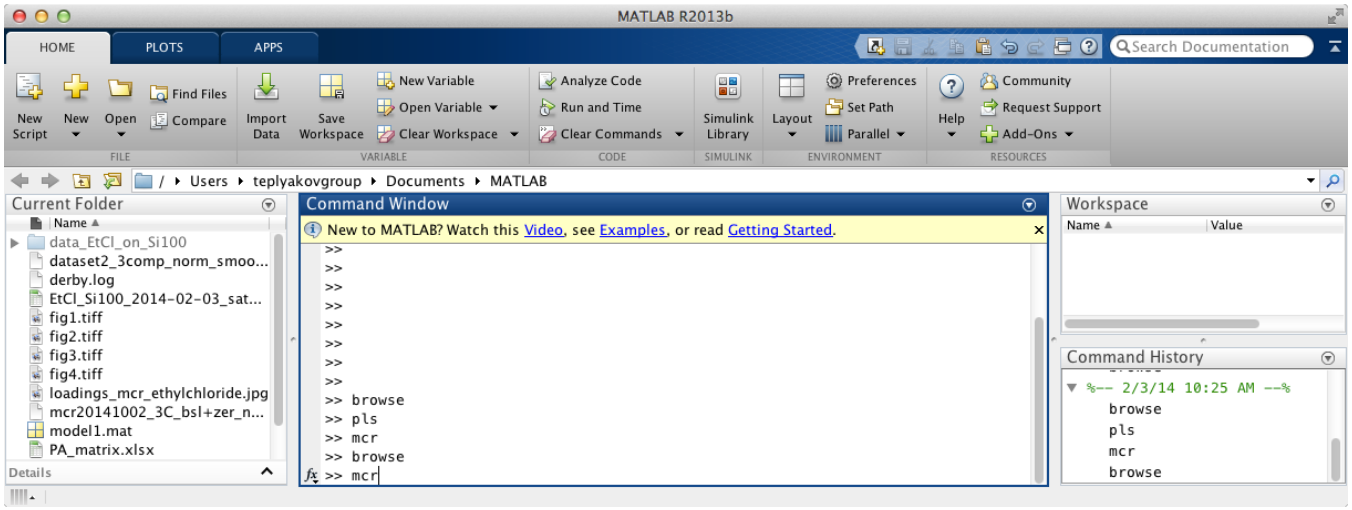


Figure 5

And you will have directly the window shown in Fig. 4.

3. Importing, editing and saving the data to MATLAB. From the toolbar in the MCR window select File/ import data / Calibration X-block / Excel file

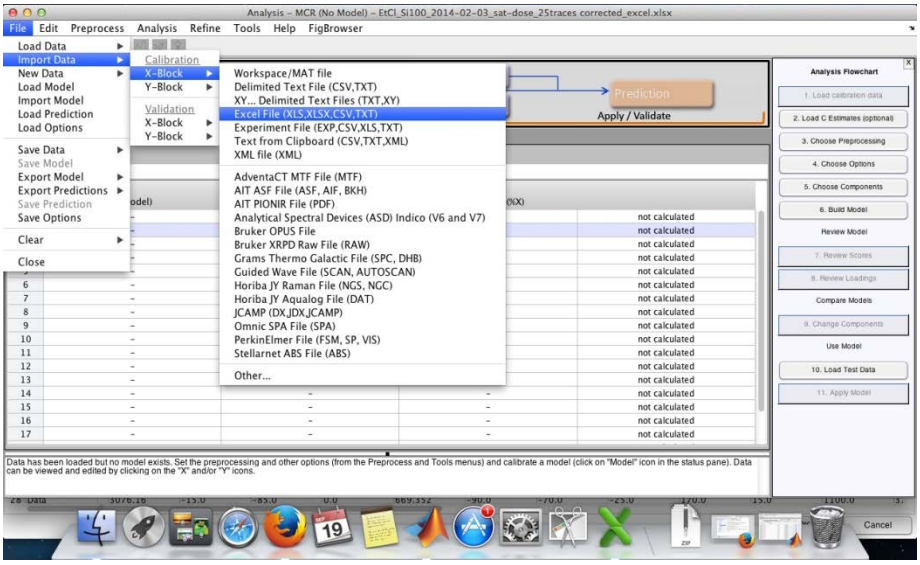


Figure 6

* IMPORTANT: Remember that the spreadsheet with the data matrix should be the first one in the excel file, otherwise it will not upload the right data.

Once the file is selected the data will pop-up in a new window:

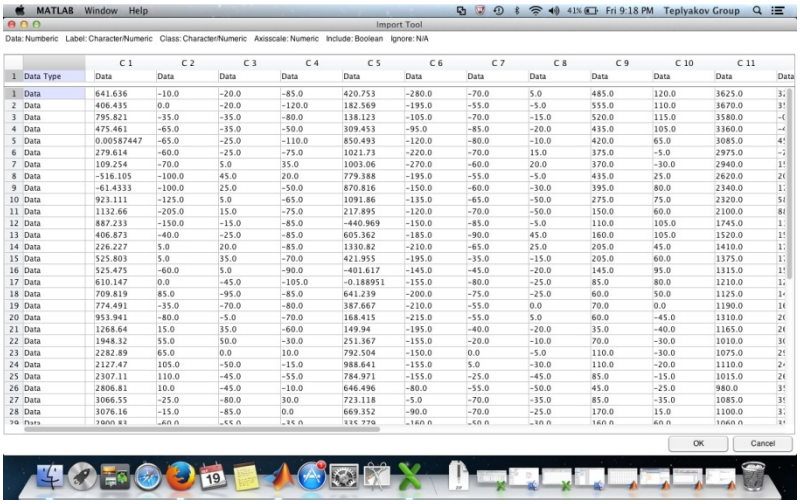


Figure 7

This is a good opportunity to confirm that the uploaded data represents the one in the matrix tab and that it does not have any cells with labels or temperature points. Select in the toolbar of the PLS_toolbox edit / calibration x-block data. We will find our data in a window that allows introducing changes:

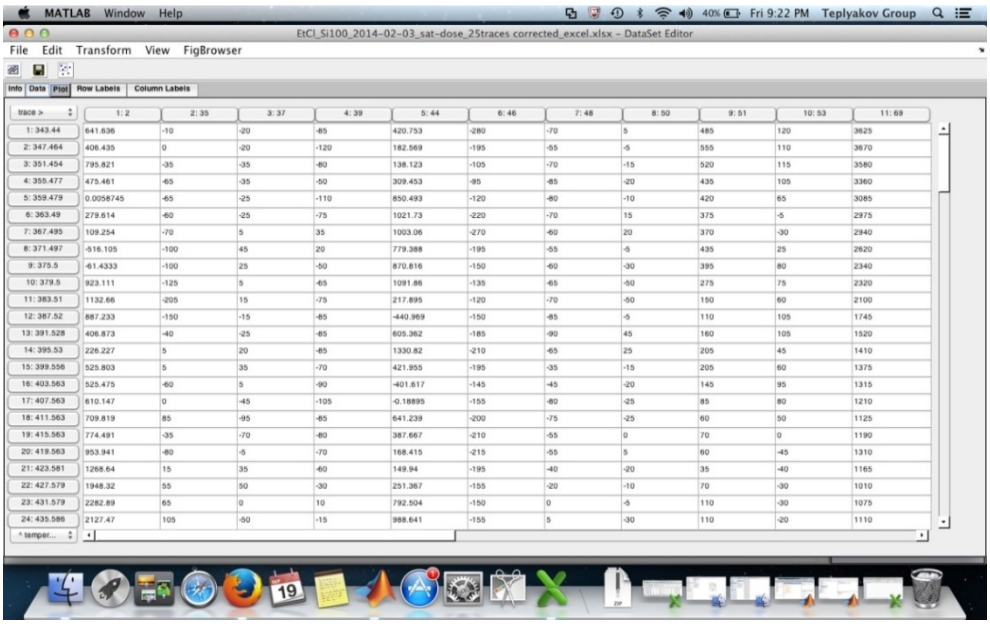


Figure 8

The info tab allows changing the name of the file and brings the info about the dimensions of the matrix. Confirm you have a 157 x 25 matrix (25 masses followed over 157 temperature points).

It is in this window that we need to insert the coordinates for temperature and m/z values. We can open the original excel file and copy the column of temperature points and paste it in the axis scale column and in the label column Copy the T points columns in the tab temperature in the excel file and after clicking in the button(Edit-paste) axis scale paste:

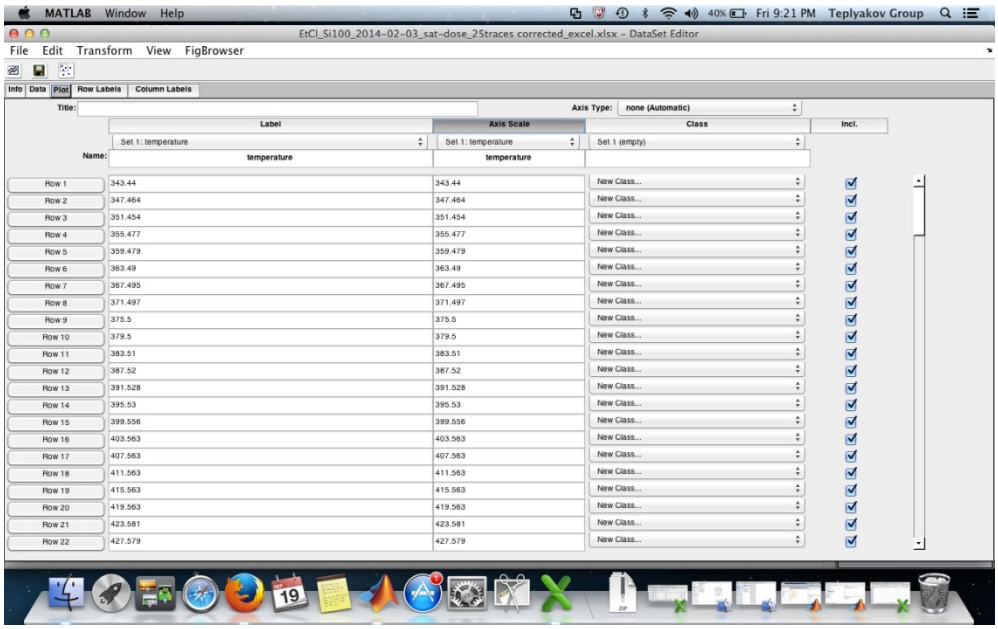


Figure 9

Now we go to the tab column labels and copy the values of m/Z that are in a column in our excel file:

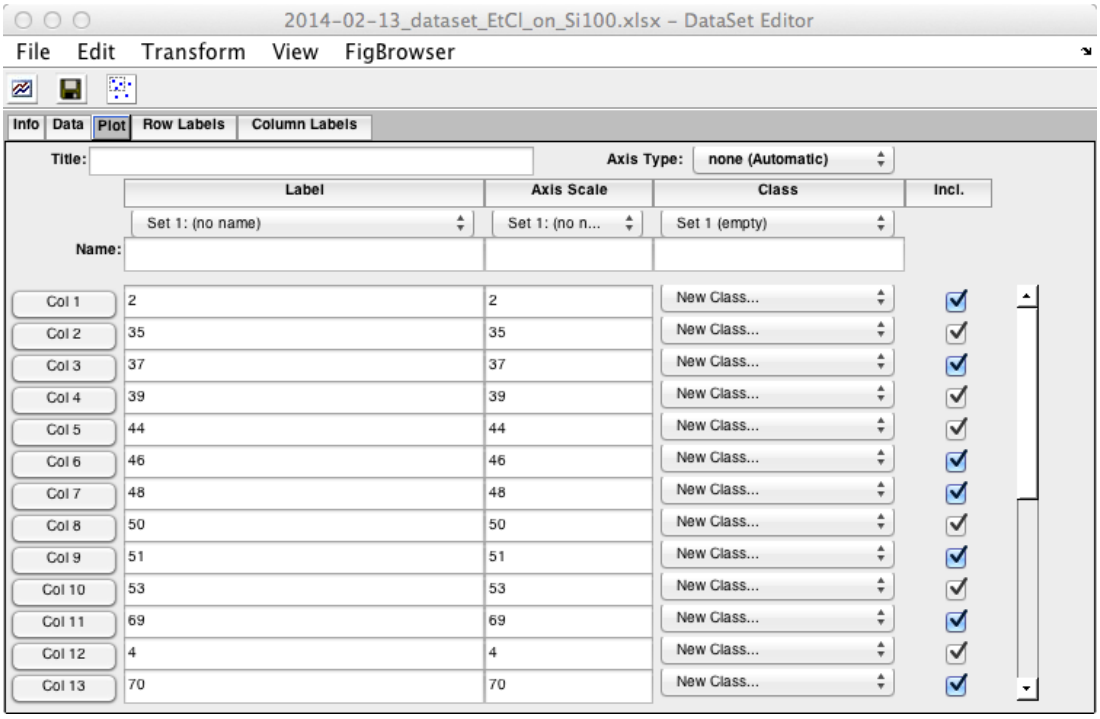


Figure 10

Once we have put all labels and coordinates, the data is ready to be saved in MATLAB format. For this, you can save it in file data_01 and as item data_01.

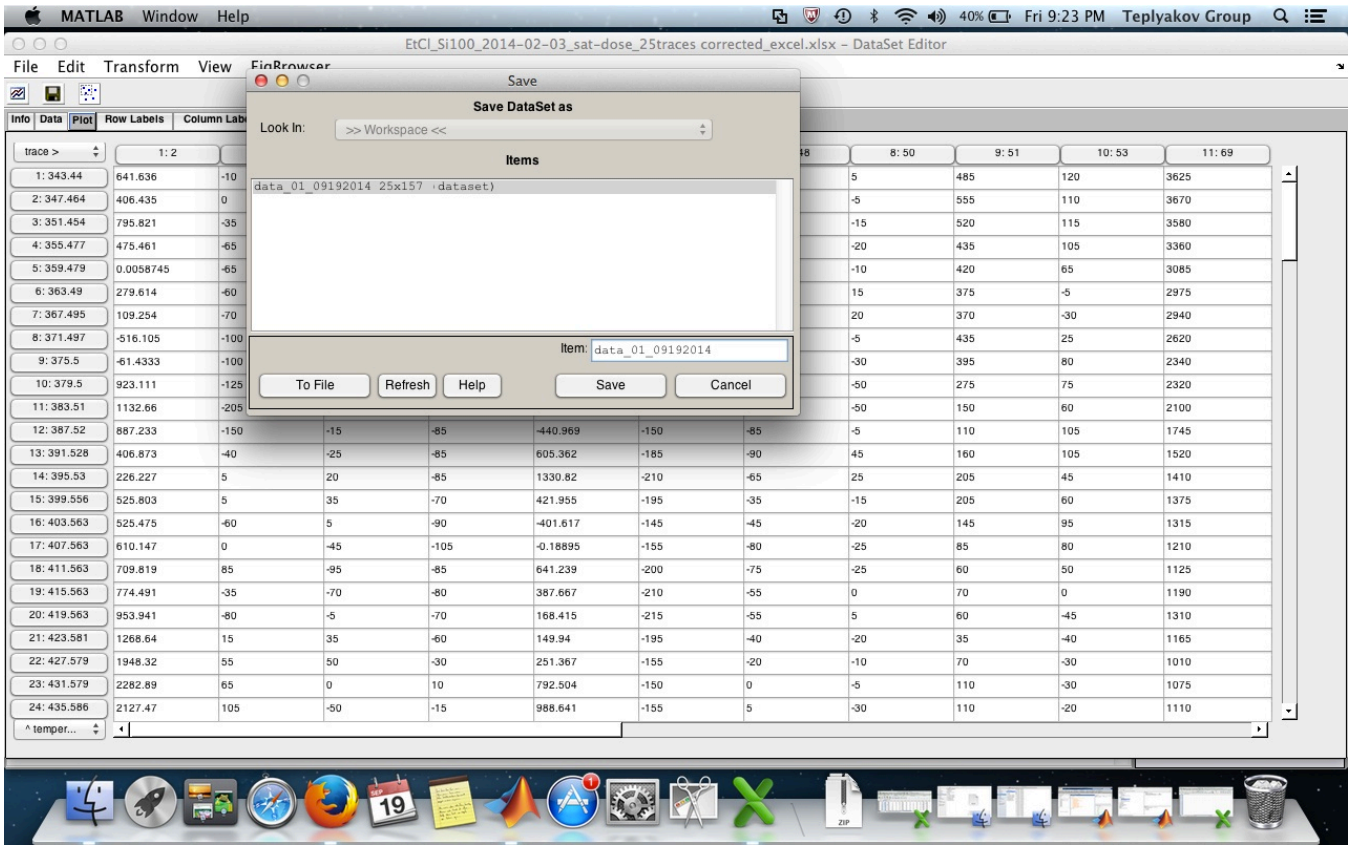

MATLAB Window Help
Fri 9:23 PM Teplyakov Group
EtCl_Si100_2014-02-03_sat-dose_25traces corrected_excel.xlsx - DataSet Editor
File Edit Transform View FigBrowser
Save
Save DataSet as
Look In: >> Workspace <<
Items
data_01_09192014 25x157 (dataset)
Item: data_01_09192014
To File
Refresh
Help
Save
Cancel

Figure 11

4. Performing multivariate analysis on data

Choose **preprocessing / x-block / custom** from the toolbar. The following window will appear:

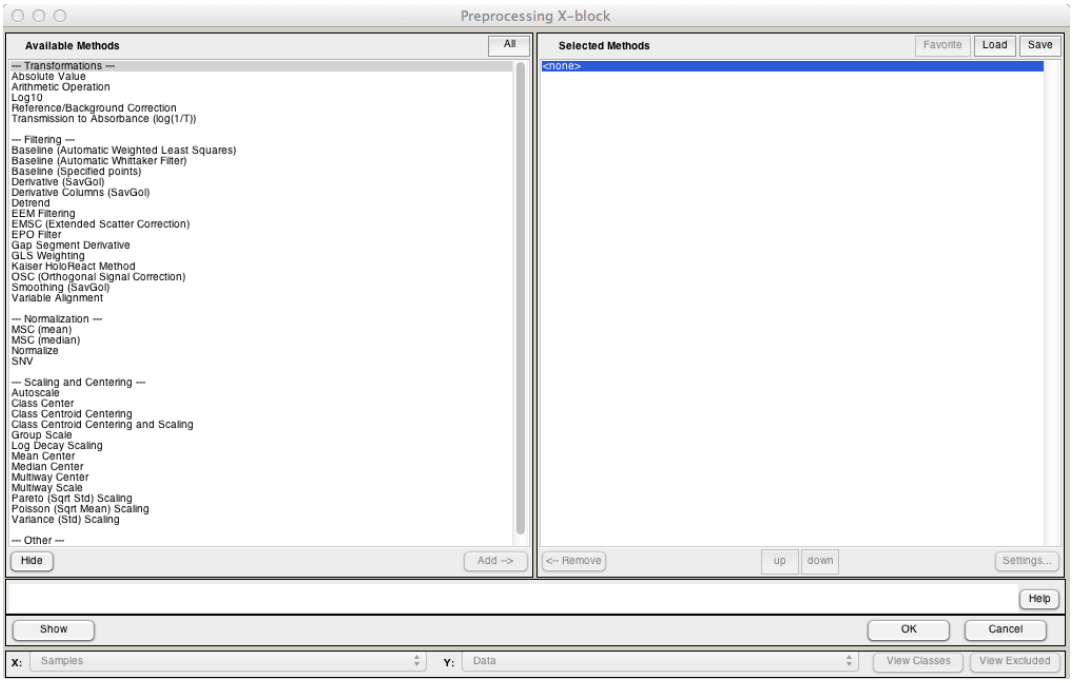


Figure 12

Click the button **show** at the bottom of the window. You will see your data before and after preprocessing. Notice that in MATLAB notation “variables” are the coordinates along rows (m/z values) and “samples” are the coordinates along columns. By default we see the plot variables vs. data:

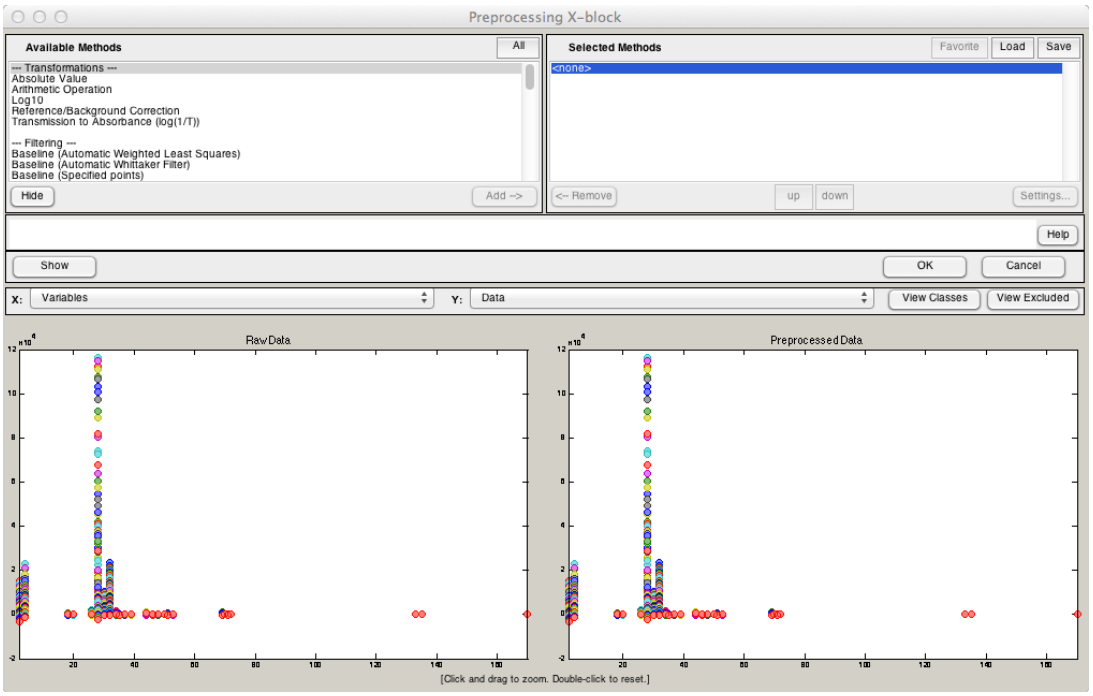


Figure 13

If we select “samples vs. data” we will see the plot of the original data. It is important to make sure we can see the data before we use any preprocessing because it can change terribly the way the data matrix looks. There are several interesting options (baseline correction, smoothing, etc.) but it is necessary to make sure the data does not change significantly. To avoid spurious effects, the data in this example will not have any preprocessing .

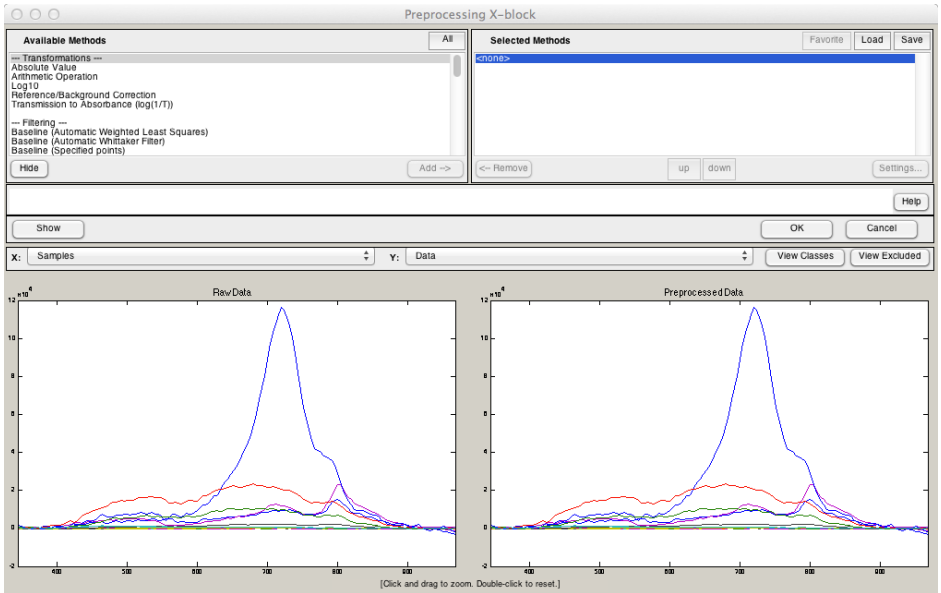


Figure 14

Now, we need to click the “options” button and select the options

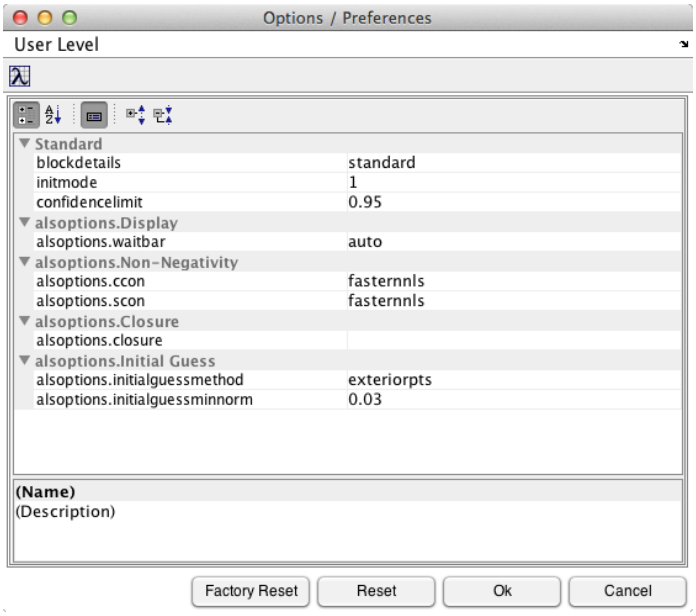


Figure 15

A description of each variable in this window is available when highlighted. The predetermined values are standard, so unless there is a specific reason to change these parameters, we do not change anything here. We have to choose the number of components that will fit our data the best way. The idea is simple: we need to make sure three points are covered:

- The residuals (error matrix) do not carry any significant information and reflect only instrumental noise. This happens when you calculate MCR with an insufficient amount of components.
- We need to avoid two components that are clearly the same but that have been separated. This happens when MCR Is performed with an excess of components.
- The variance of the smallest compound should be significant (e.g. more than 0.5%).

The only way we can make sure we have acquired data with a correct number of components is by performing the analysis starting with 2 components. For example, in the analysis window we need to select the second row in the spreadsheet:

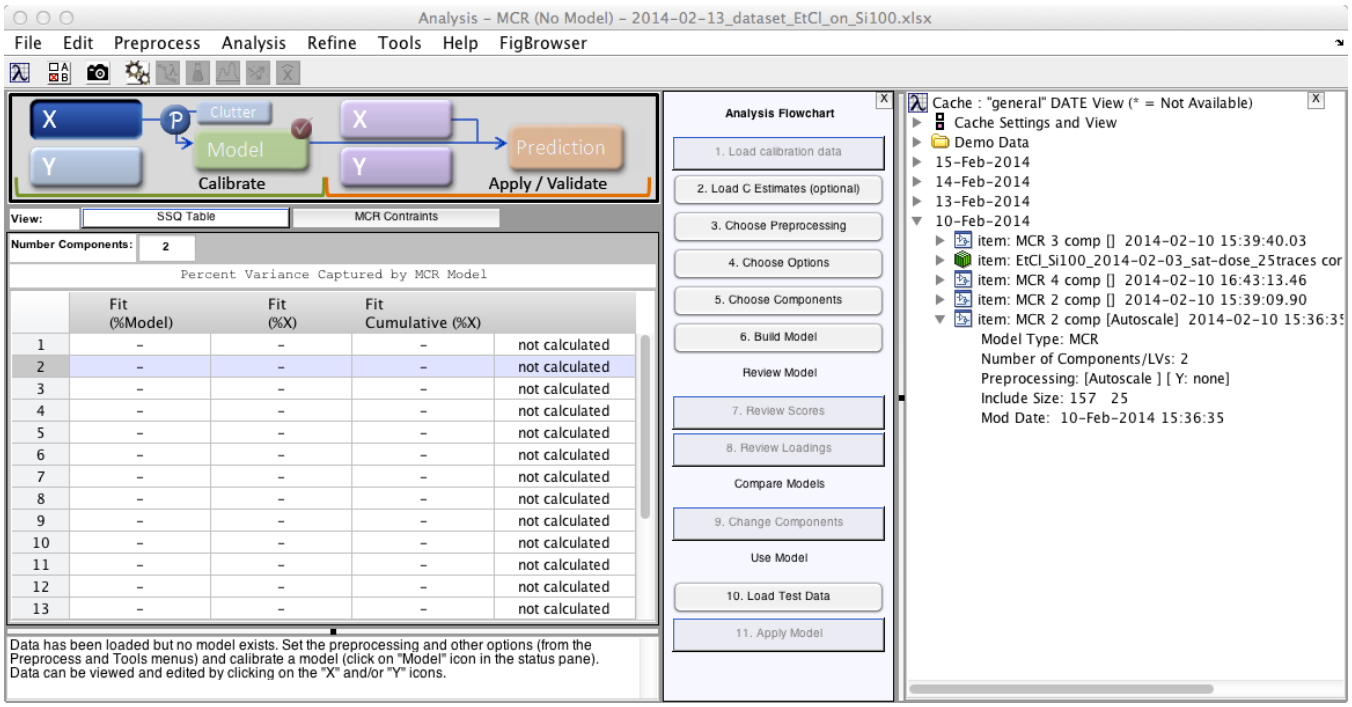


Figure 16

Then, we will click on the button **build model**. After a few seconds you will have the MCR analysis done, and in the table that was originally empty we will have the values of % of data captured by the model:

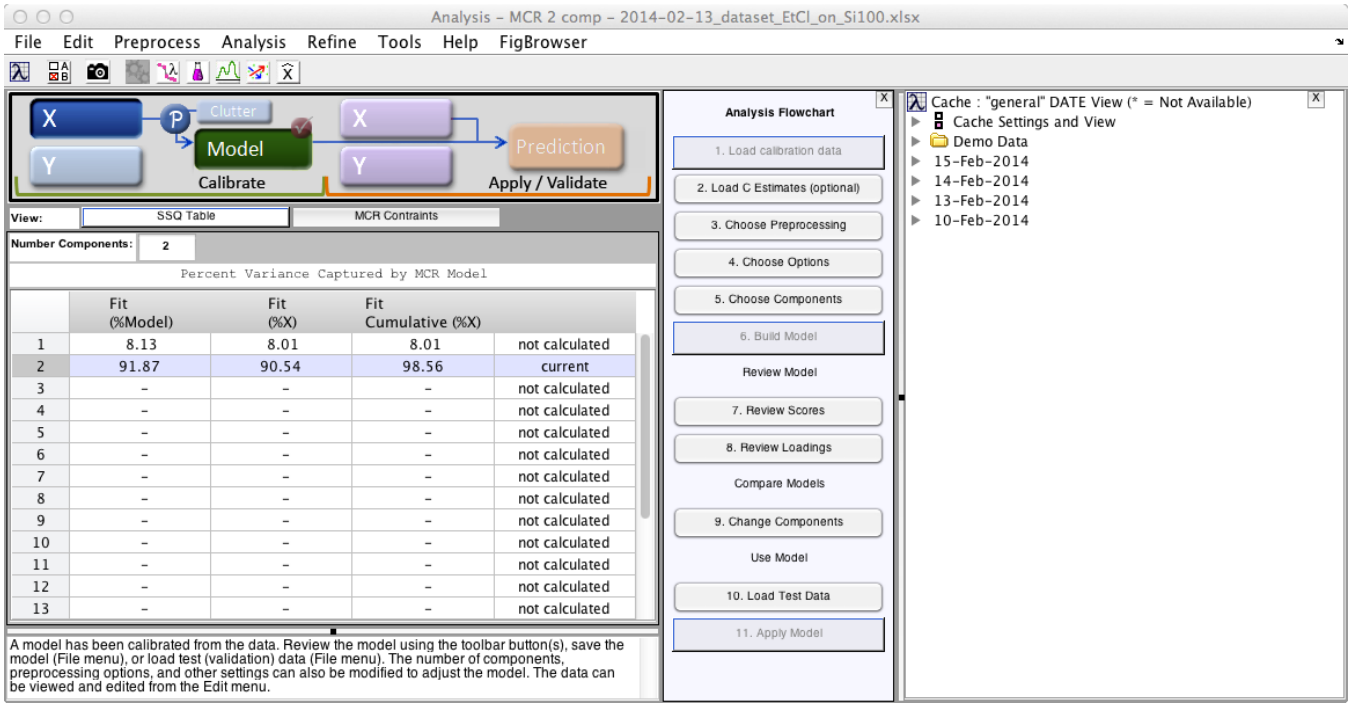


Figure 17

If we review the scores (or matrix 157 x 2) we will see the desorption patterns for the two components, together with the amount of variance captured by each component.

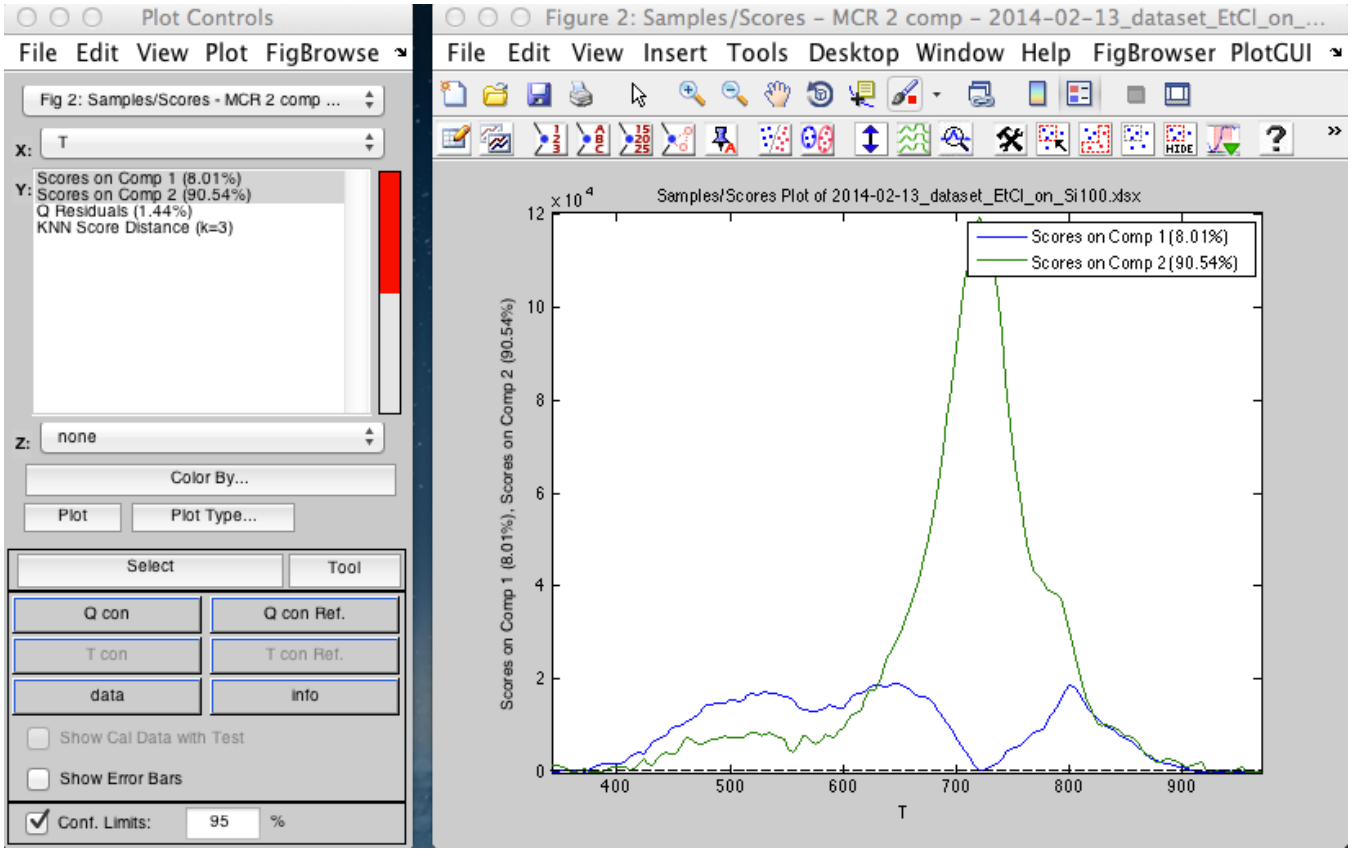


Figure 18

We can also plot the residuals of the scores. This will show what has not been included in the model

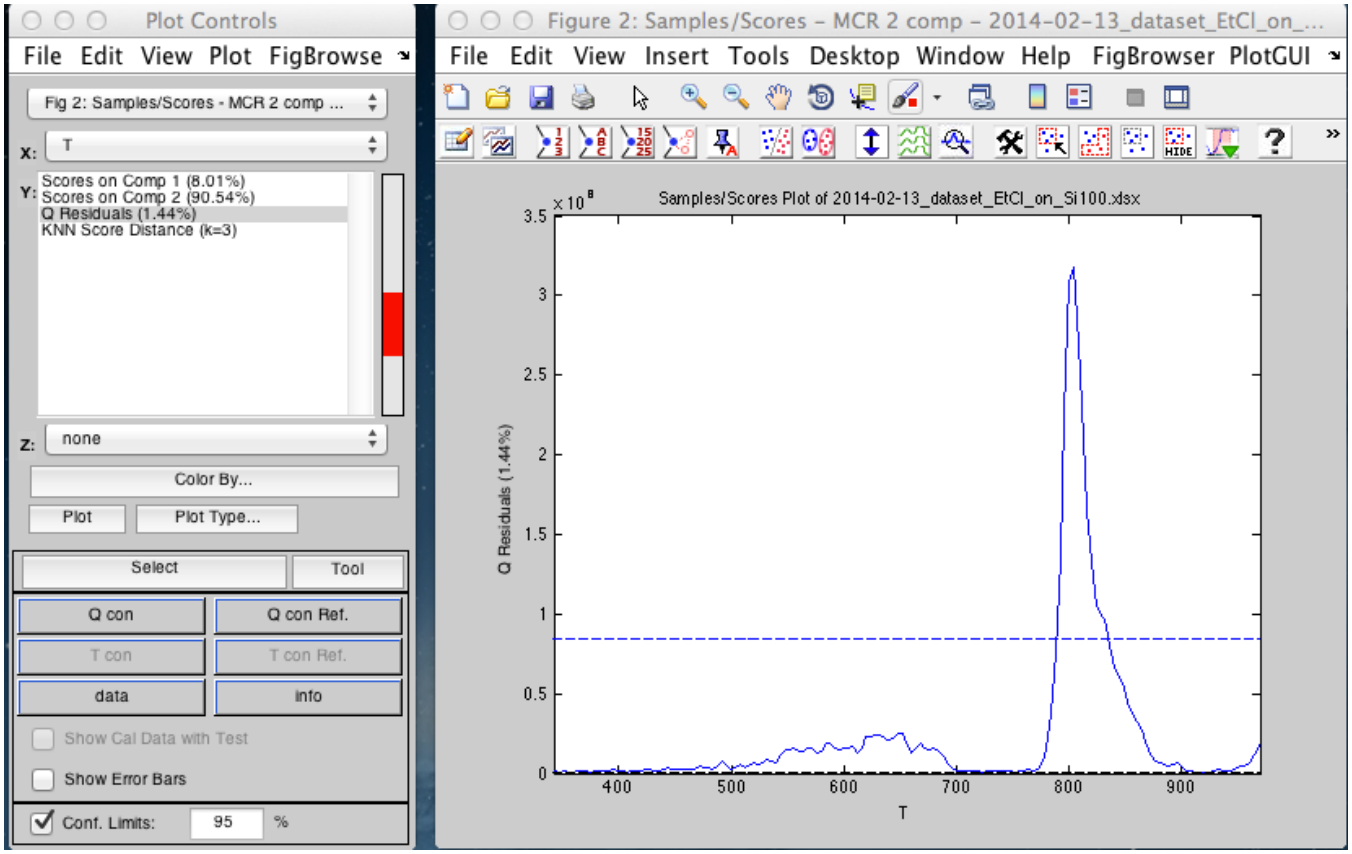


Figure 19

It is clear that we have a peak that has not been included in the graph. This is a strong indication that 2 components are not sufficient.

We can also obtain information from the loadings (2 x 25 matrix). The mass spectra of the two components and the residuals are shown in Figures 20 and 21. (click “view loadings”)

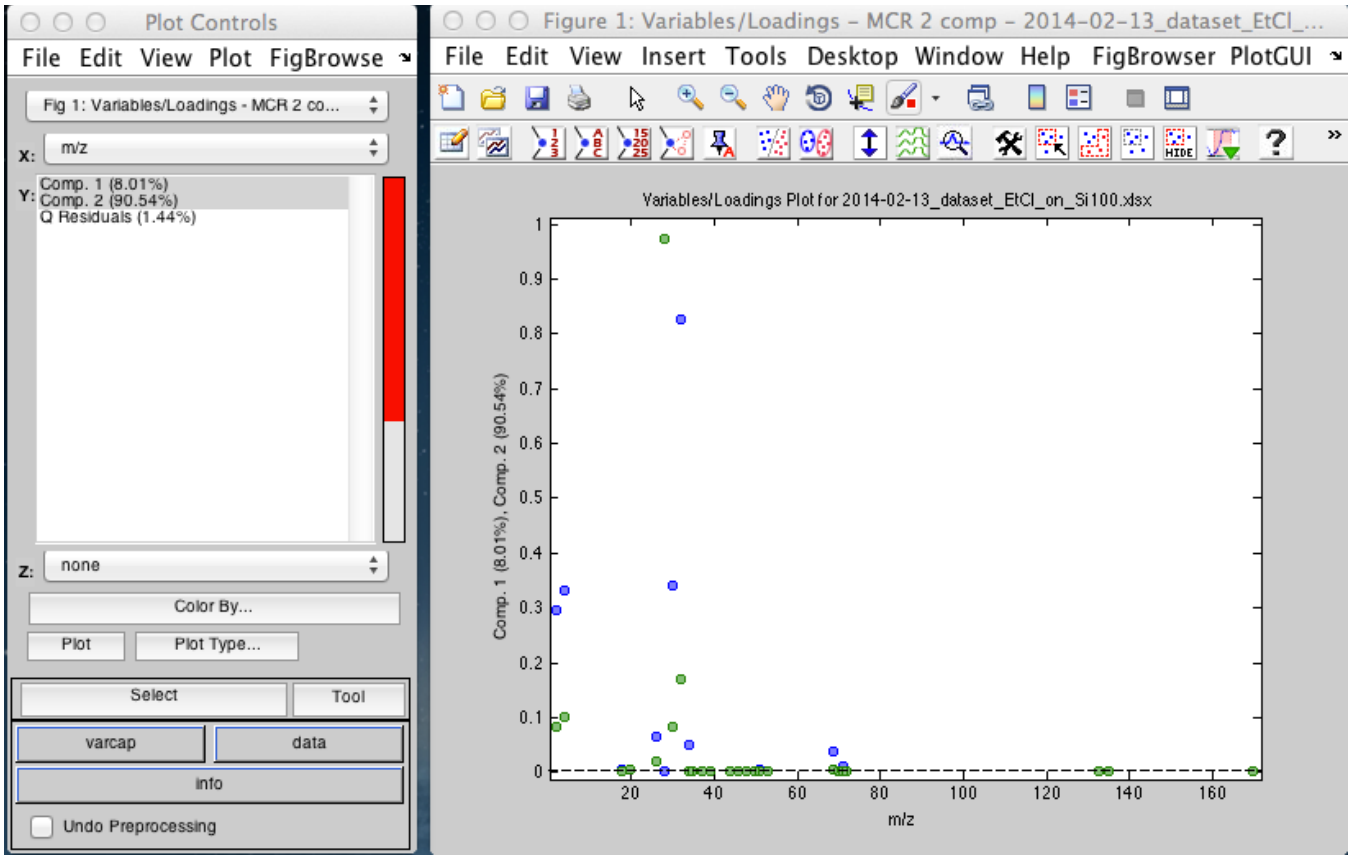

Plot Controls
File Edit View Plot FigBrowse
Fig 1: Variables/Loadings - MCR 2 co...
X:
m/z
Y:
Comp. 1 (8.01%)
Comp. 2 (90.54%)
Q Residuals (1.44%)
Z:
none
Color By...
Plot
Plot Type...
Select
Tool
varcap
data
info
Undo Preprocessing
Figure 1: Variables/Loadings - MCR 2 comp - 2014-02-13_dataset_EtCl_...
File Edit View Insert Tools Desktop Window Help FigBrowser PlotGUI
Variables/Loadings Plot for 2014-02-13_dataset_EtCl_on_Si100.xlsx
Comp. 1 (8.01%), Comp. 2 (90.54%)
m/z
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
20
40
60
80
100
120
140
160

Figure 20

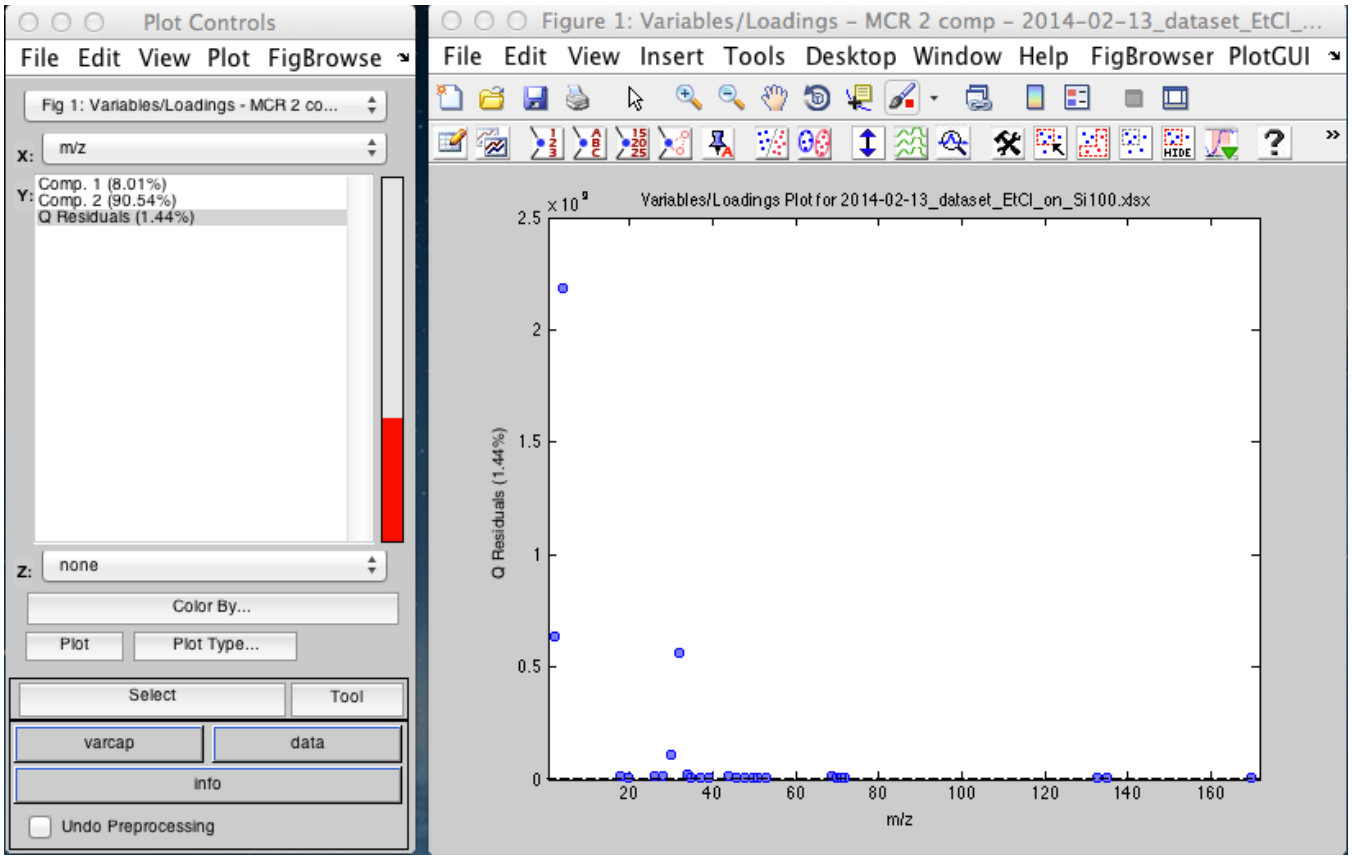


Figure 21

Notice that the residuals show a significant amount of mass 4.

Let´s try 3 components. Doing the same analysis, we have Figures 22 and 23 for scores and 24 and 25 for loadings. It seems the data is better represented. The residuals are more randomly distributed, which means we are taking all the information out.

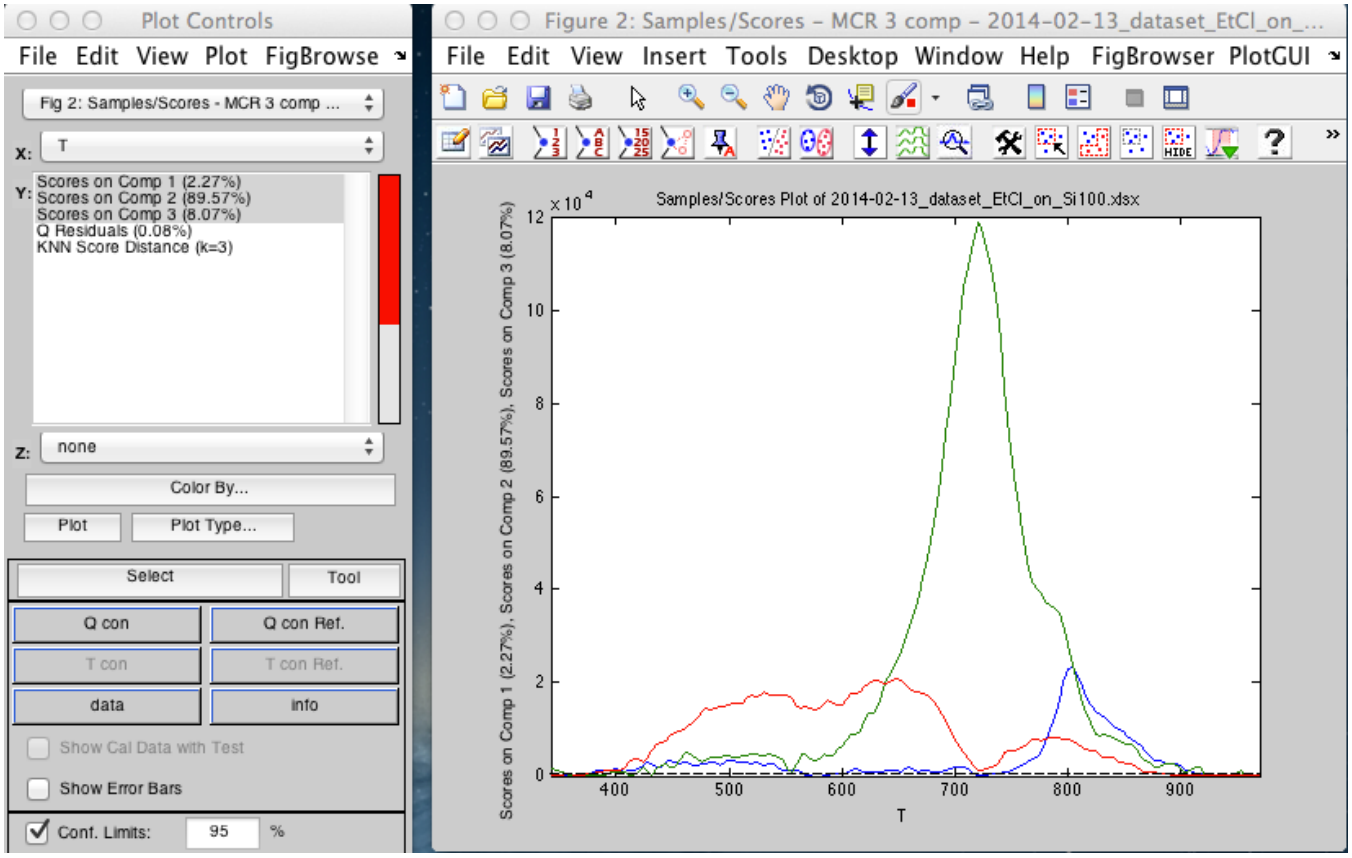

Plot Controls
File Edit View Plot FigBrowse
Fig 2: Samples/Scores - MCR 3 comp ...
X: T
Y: Scores on Comp 1 (2.27%)
Scores on Comp 2 (89.57%)
Scores on Comp 3 (8.07%)
Q Residuals (0.08%)
KNN Score Distance (k=3)
Z: none
Color By...
Plot
Plot Type...
Select
Tool
Q con
Q con Ref.
T con
T con Ref.
data
info
Show Cal Data with Test
Show Error Bars
Conf. Limits: 95 %
Figure 2: Samples/Scores - MCR 3 comp - 2014-02-13_dataset_EtCl_on_...
File Edit View Insert Tools Desktop Window Help FigBrowser PlotGUI
Samples/Scores Plot of 2014-02-13_dataset_EtCl_on_Si100.xlsx
x 10^4
Scores on Comp 1 (2.27%), Scores on Comp 2 (89.57%), Scores on Comp 3 (8.07%)
T
0
2
4
6
8
10
12
400
500
600
700
800
900

Figure 22

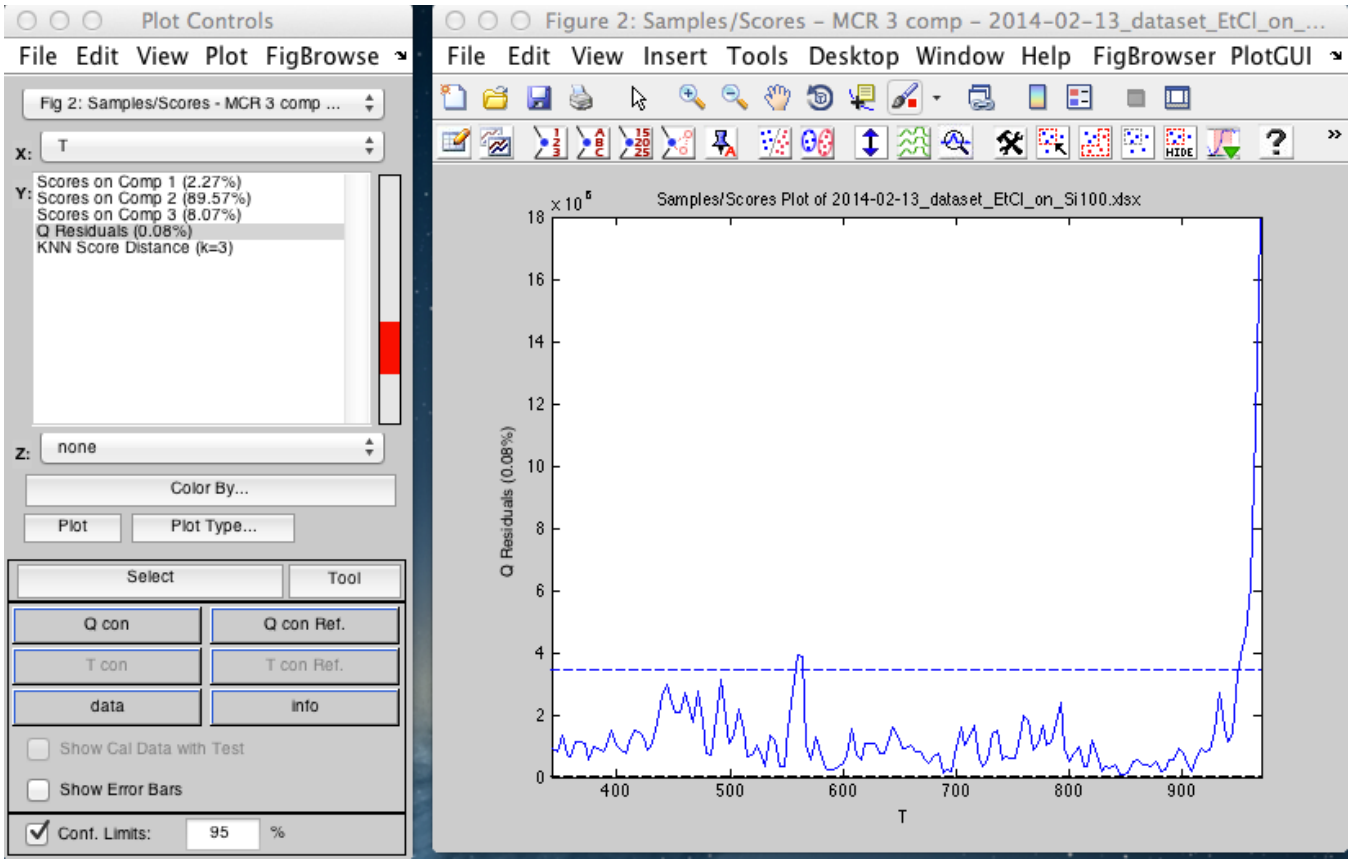

Plot Controls
File Edit View Plot FigBrowse
Fig 2: Samples/Scores - MCR 3 comp ...
X: T
Y: Scores on Comp 1 (2.27%)
Scores on Comp 2 (89.57%)
Scores on Comp 3 (8.07%)
Q Residuals (0.08%)
KNN Score Distance (k=3)
Z: none
Color By...
Plot
Plot Type...
Select
Tool
Q con
Q con Ref.
T con
T con Ref.
data
info
Show Cal Data with Test
Show Error Bars
Conf. Limits: 95 %
Figure 2: Samples/Scores - MCR 3 comp - 2014-02-13_dataset_EtCl_on_...
File Edit View Insert Tools Desktop Window Help FigBrowser PlotGUI
Samples/Scores Plot of 2014-02-13_dataset_EtCl_on_Si100.xlsx
x 10^5
Q Residuals (0.08%)
T
0
2
4
6
8
10
12
14
16
18
400
500
600
700
800
900

Figure 23

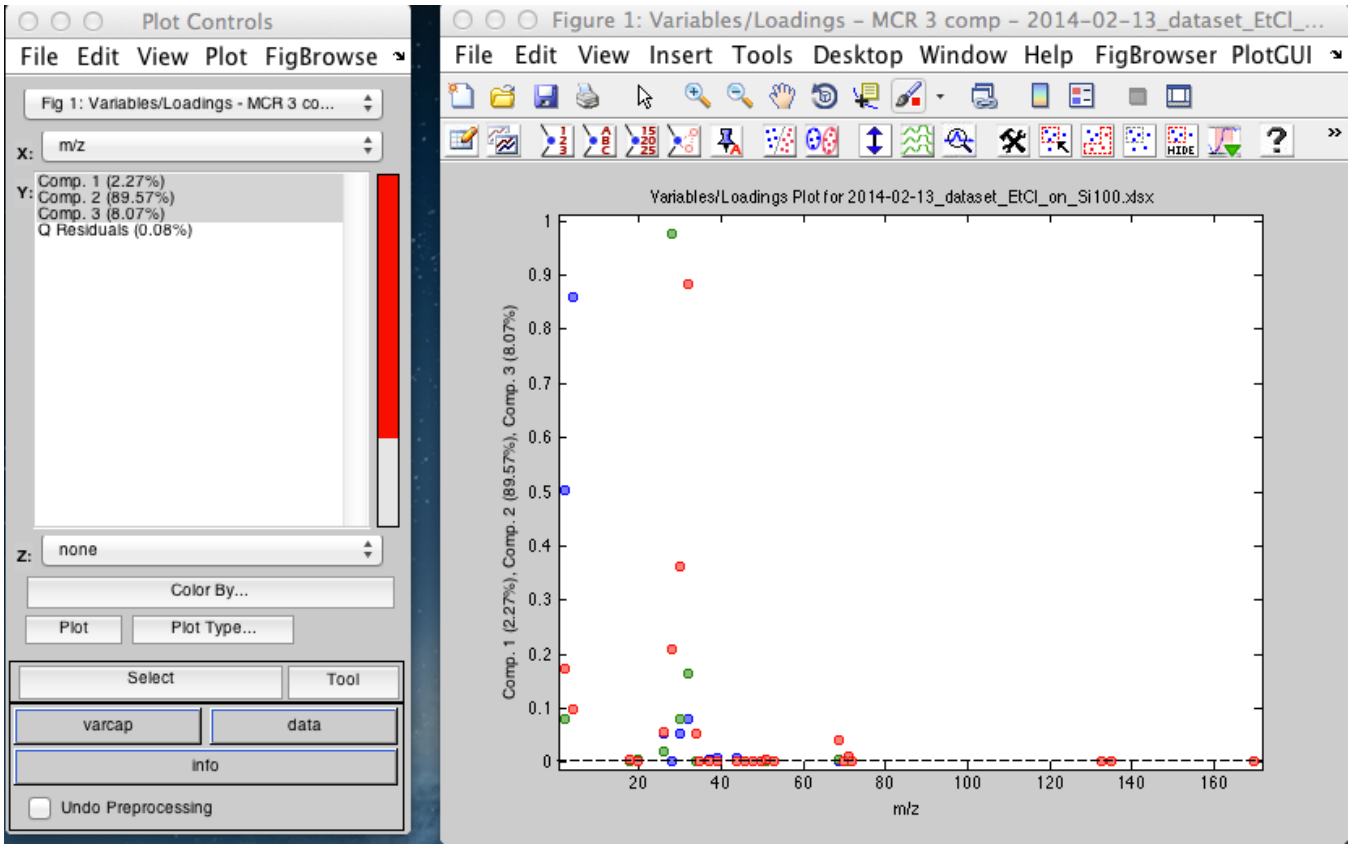

Plot Controls
File Edit View Plot FigBrowse
Fig 1: Variables/Loadings - MCR 3 co...
X:
m/z
Y:
Comp. 1 (2.27%)
Comp. 2 (89.57%)
Comp. 3 (8.07%)
Q Residuals (0.08%)
Z:
none
Color By...
Plot
Plot Type...
Select
Tool
varcap
data
info
Undo Preprocessing
Figure 1: Variables/Loadings - MCR 3 comp - 2014-02-13_dataset_EtCl_...
File Edit View Insert Tools Desktop Window Help FigBrowser PlotGUI
Variables/Loadings Plot for 2014-02-13_dataset_EtCl_on_Si100.xlsx
Comp. 1 (2.27%), Comp. 2 (89.57%), Comp. 3 (8.07%)
m/z
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
20
40
60
80
100
120
140
160

Figure 24

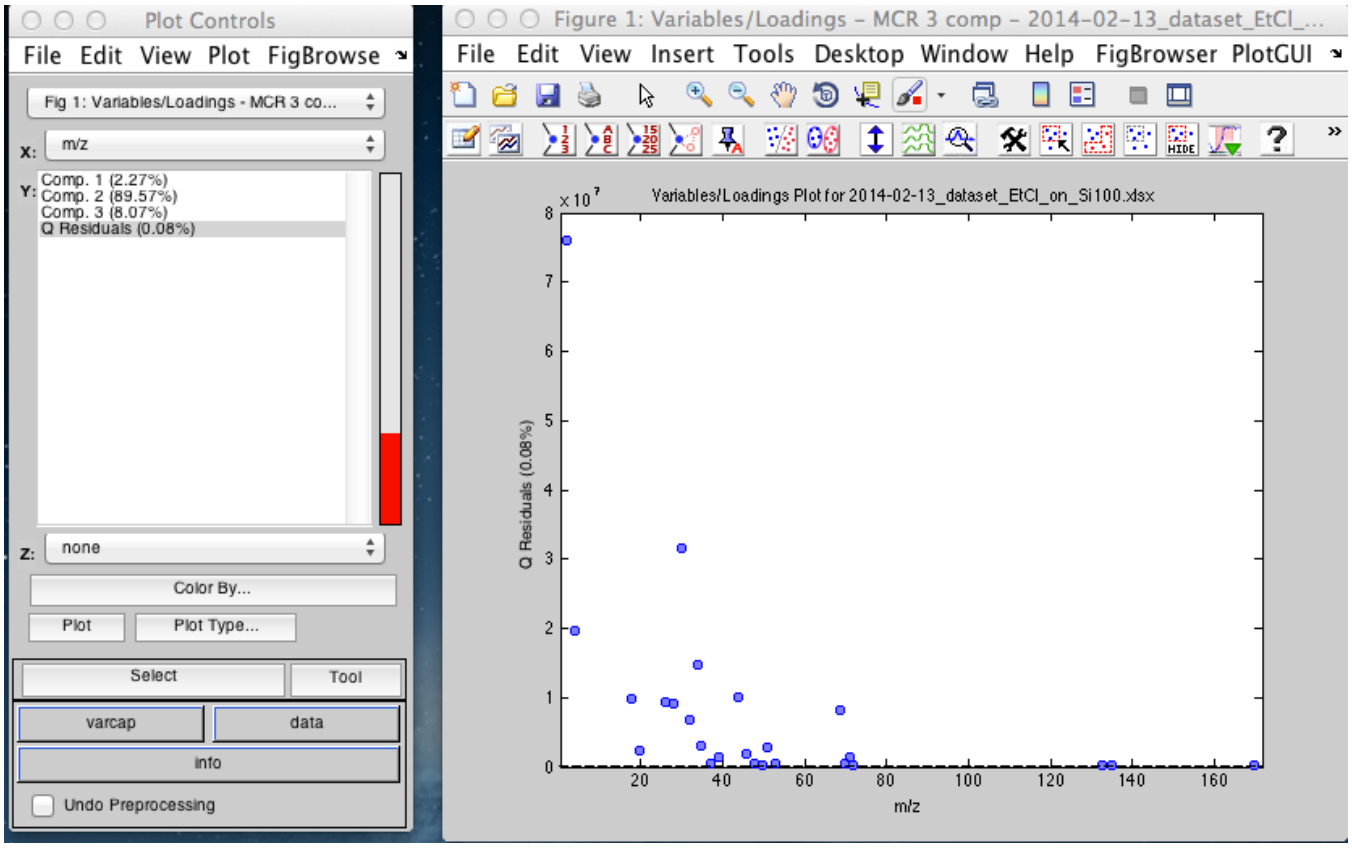

Plot Controls
File Edit View Plot FigBrowse
Fig 1: Variables/Loadings - MCR 3 co...
X: m/z
Y: Comp. 1 (2.27%)
Comp. 2 (89.57%)
Comp. 3 (8.07%)
Q Residuals (0.08%)
Z: none
Color By...
Plot
Plot Type...
Select
Tool
varcap
data
info
Undo Preprocessing
Figure 1: Variables/Loadings - MCR 3 comp - 2014-02-13_dataset_EtCl_...
File Edit View Insert Tools Desktop Window Help FigBrowser PlotGUI
x 10^7
Variables/Loadings Plot for 2014-02-13_dataset_EtCl_on_Si100.xlsx
Q Residuals (0.08%)
m/z
0
1
2
3
4
5
6
7
8
20
40
60
80
100
120
140
160

Figure 25

Let´s try 4 components. Figures 26, 27 (scores) and 28, 29 (loadings):

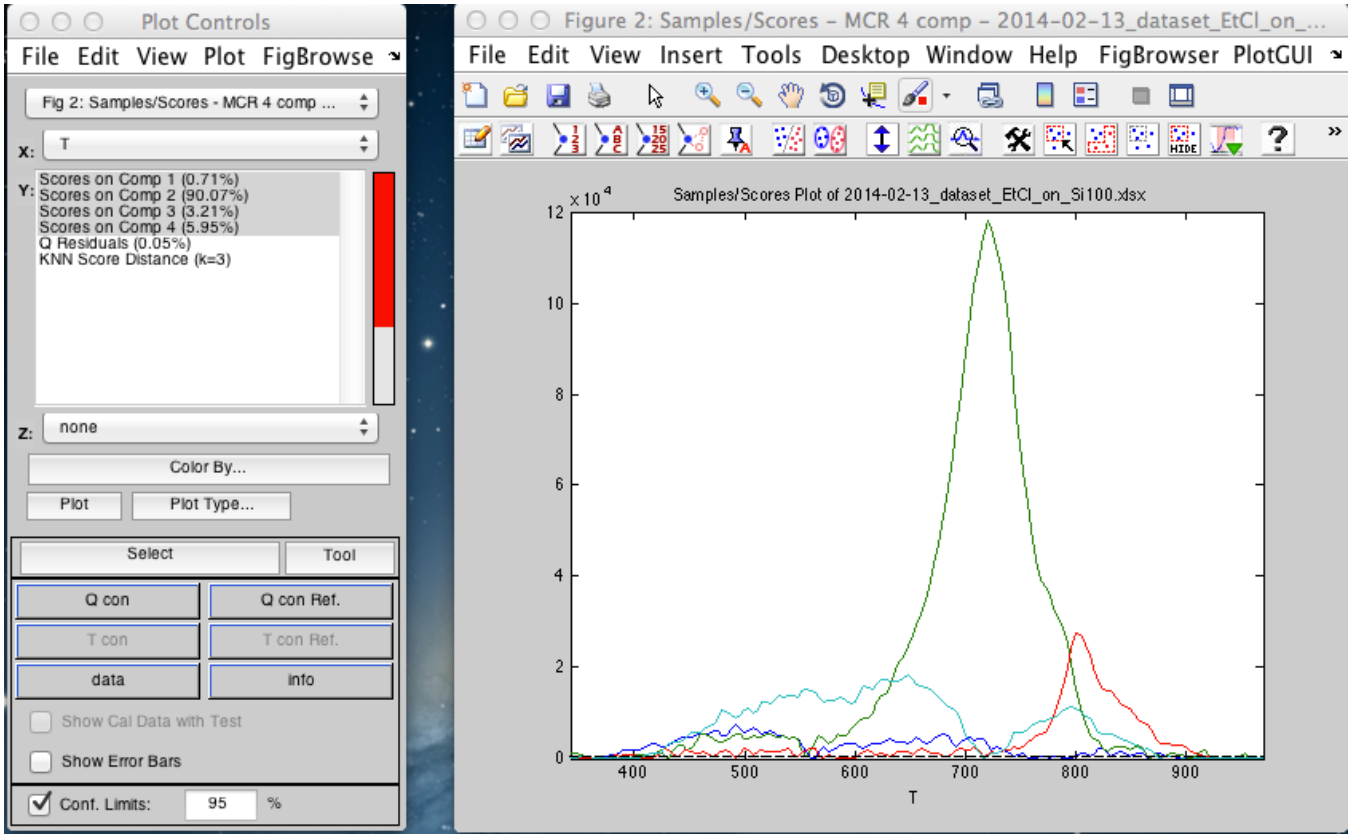


Figure 26

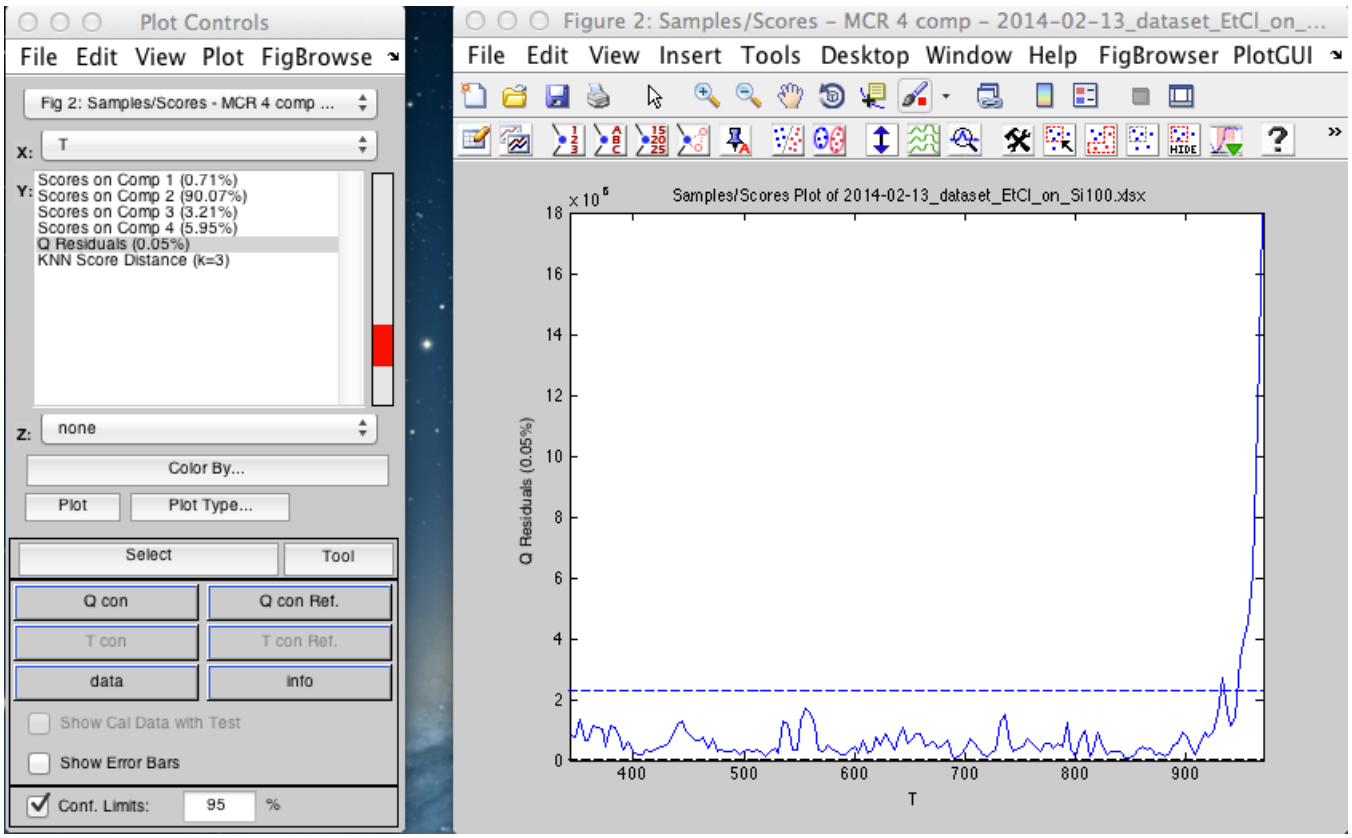

Plot Controls
File Edit View Plot FigBrowse
Fig 2: Samples/Scores - MCR 4 comp ...
X: T
Y: Scores on Comp 1 (0.71%)
Scores on Comp 2 (90.07%)
Scores on Comp 3 (3.21%)
Scores on Comp 4 (5.95%)
Q Residuals (0.05%)
KNN Score Distance (k=3)
Z: none
Color By...
Plot
Plot Type...
Select
Tool
Q con
Q con Ref.
T con
T con Ref.
data
info
Show Cal Data with Test
Show Error Bars
Conf. Limits: 95 %
Figure 2: Samples/Scores - MCR 4 comp - 2014-02-13_dataset_EtCl_on_...
File Edit View Insert Tools Desktop Window Help FigBrowser PlotGUI
Samples/Scores Plot of 2014-02-13_dataset_EtCl_on_Si100.xlsx
x 10^6
Q Residuals (0.05%)
T
0
2
4
6
8
10
12
14
16
18
400
500
600
700
800
900

Figure 27

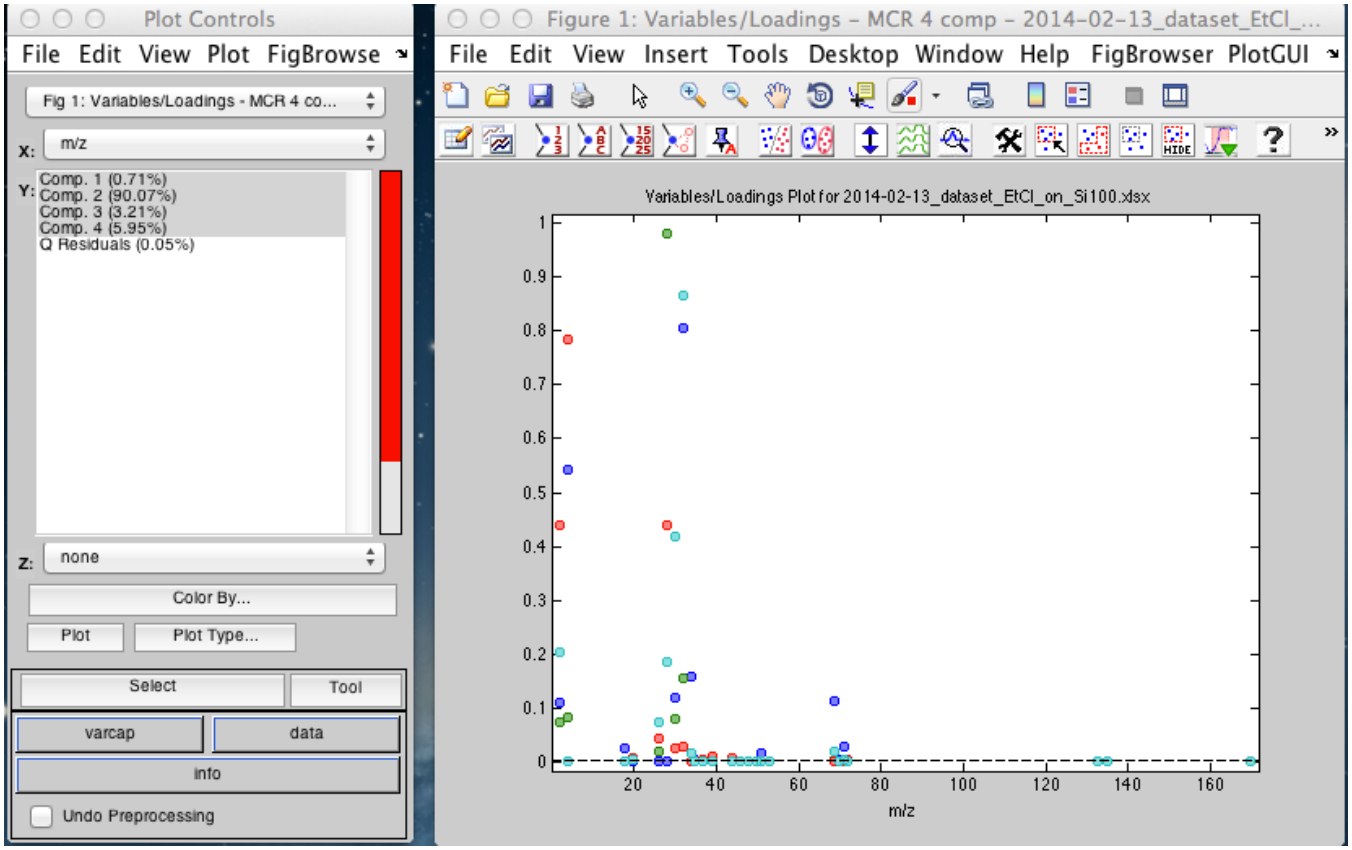

Plot Controls
File Edit View Plot FigBrowse
Fig 1: Variables/Loadings - MCR 4 co...
X:
m/z
Y:
Comp. 1 (0.71%)
Comp. 2 (90.07%)
Comp. 3 (3.21%)
Comp. 4 (5.95%)
Q Residuals (0.05%)
Z:
none
Color By...
Plot
Plot Type...
Select
Tool
varcap
data
info
Undo Preprocessing
Figure 1: Variables/Loadings - MCR 4 comp - 2014-02-13_dataset_EtCl_...
File Edit View Insert Tools Desktop Window Help FigBrowser PlotGUI
Variables/Loadings Plot for 2014-02-13_dataset_EtCl_on_Si100.xlsx
m/z

Figure 28

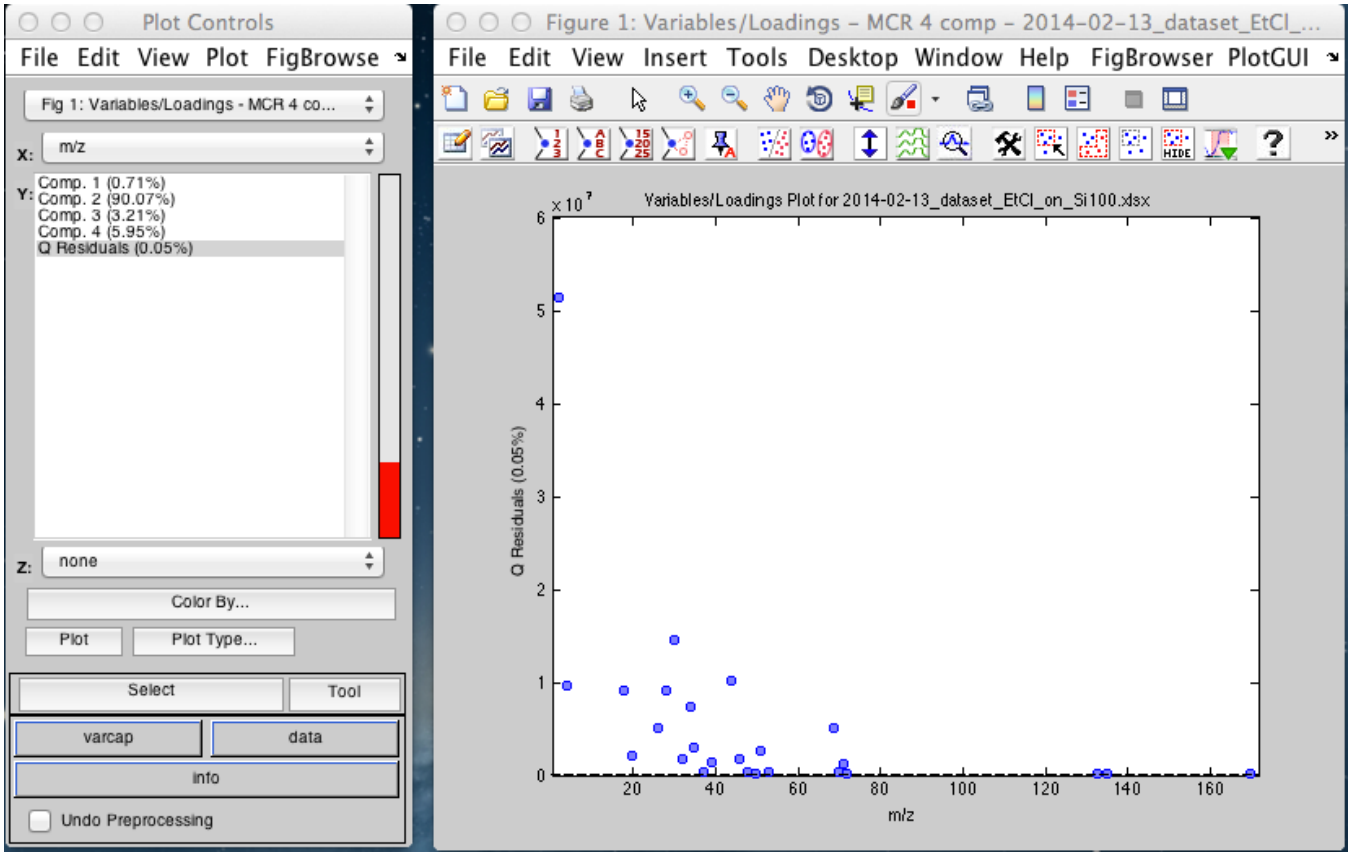


Figure 29

NOTE: It may be important for visualization purposes, to plot the loadings as bars instead of points. This is a regular MATLAB procedure that will produce a graph like this:

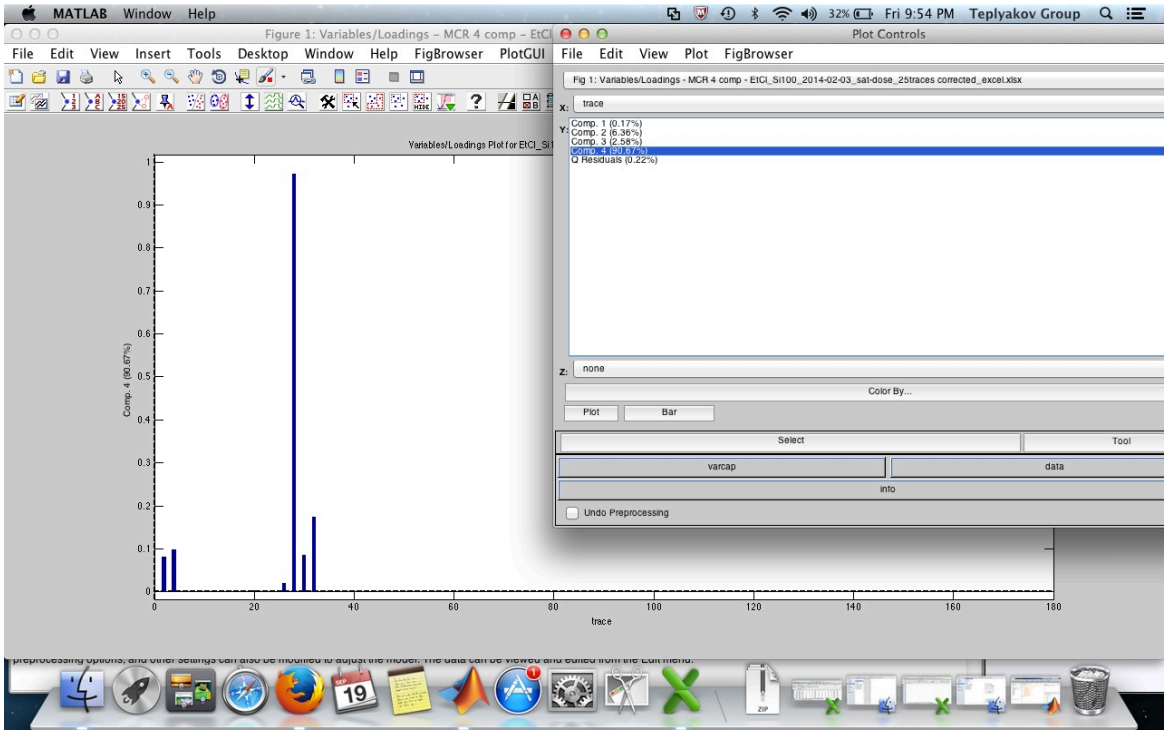


Figure 30

The residuals do not show any information, so we should be close to the real number of components. To confirm this, we perform MCR considering 5 components. The results are shown below. It is notable that in this case there is a compound carrying only 0.25% of the variance, which indicates that even though this is an accurate description mathematically, it does not add anything new in physical sense. Figures 31 and 32 show the desorption spectra and the residuals:

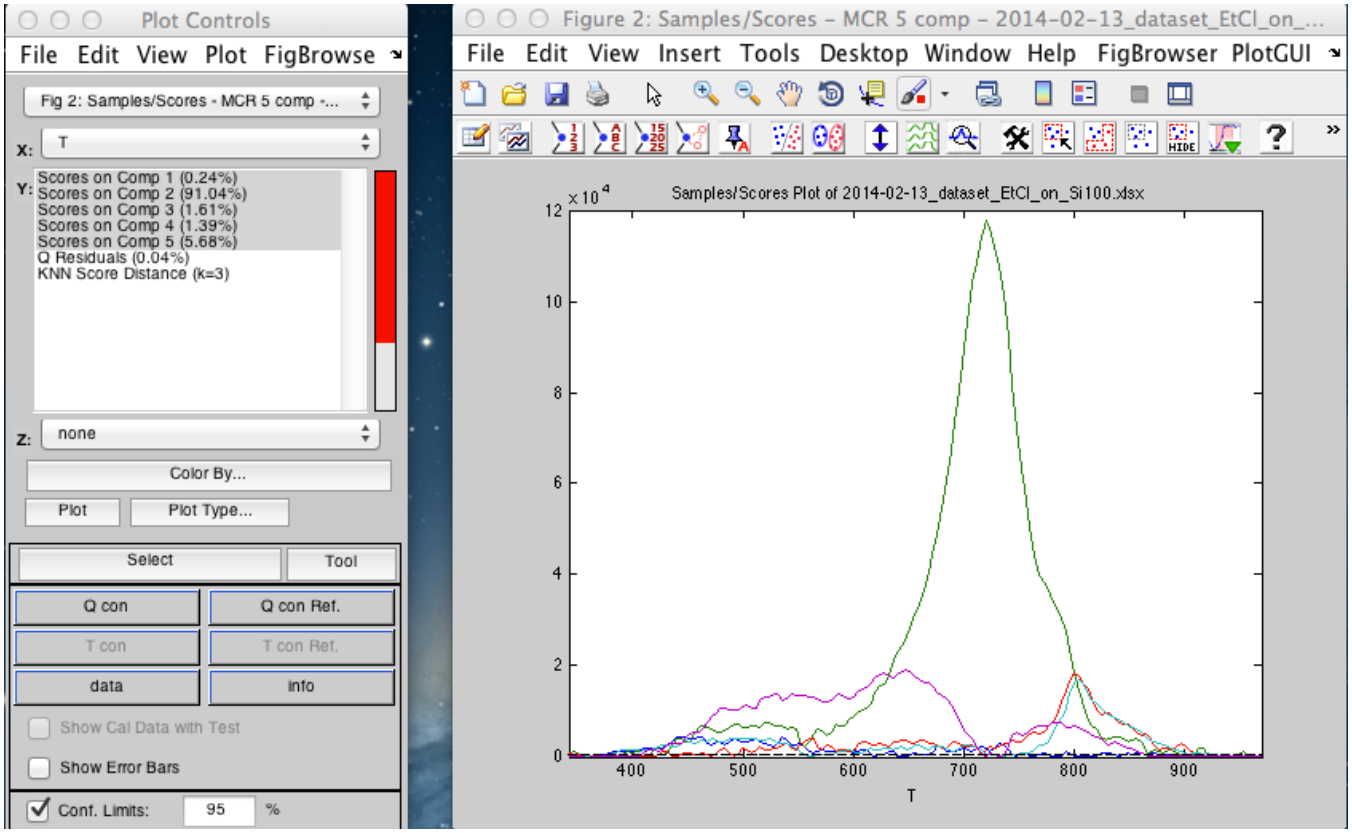

Plot Controls
File Edit View Plot FigBrowse
Fig 2: Samples/Scores - MCR 5 comp -...
X: T
Y: Scores on Comp 1 (0.24%)
Scores on Comp 2 (91.04%)
Scores on Comp 3 (1.61%)
Scores on Comp 4 (1.39%)
Scores on Comp 5 (5.68%)
Q Residuals (0.04%)
KNN Score Distance (k=3)
Z: none
Color By...
Plot
Plot Type...
Select
Tool
Q con
Q con Ref.
T con
T con Ref.
data
info
Show Cal Data with Test
Show Error Bars
Conf. Limits: 95 %
Figure 2: Samples/Scores - MCR 5 comp - 2014-02-13_dataset_EtCl_on_...
File Edit View Insert Tools Desktop Window Help FigBrowser PlotGUI
Samples/Scores Plot of 2014-02-13_dataset_EtCl_on_Si100.xlsx
x 10^4
12
10
8
6
4
2
0
400
500
600
700
800
900
T

Figure 31

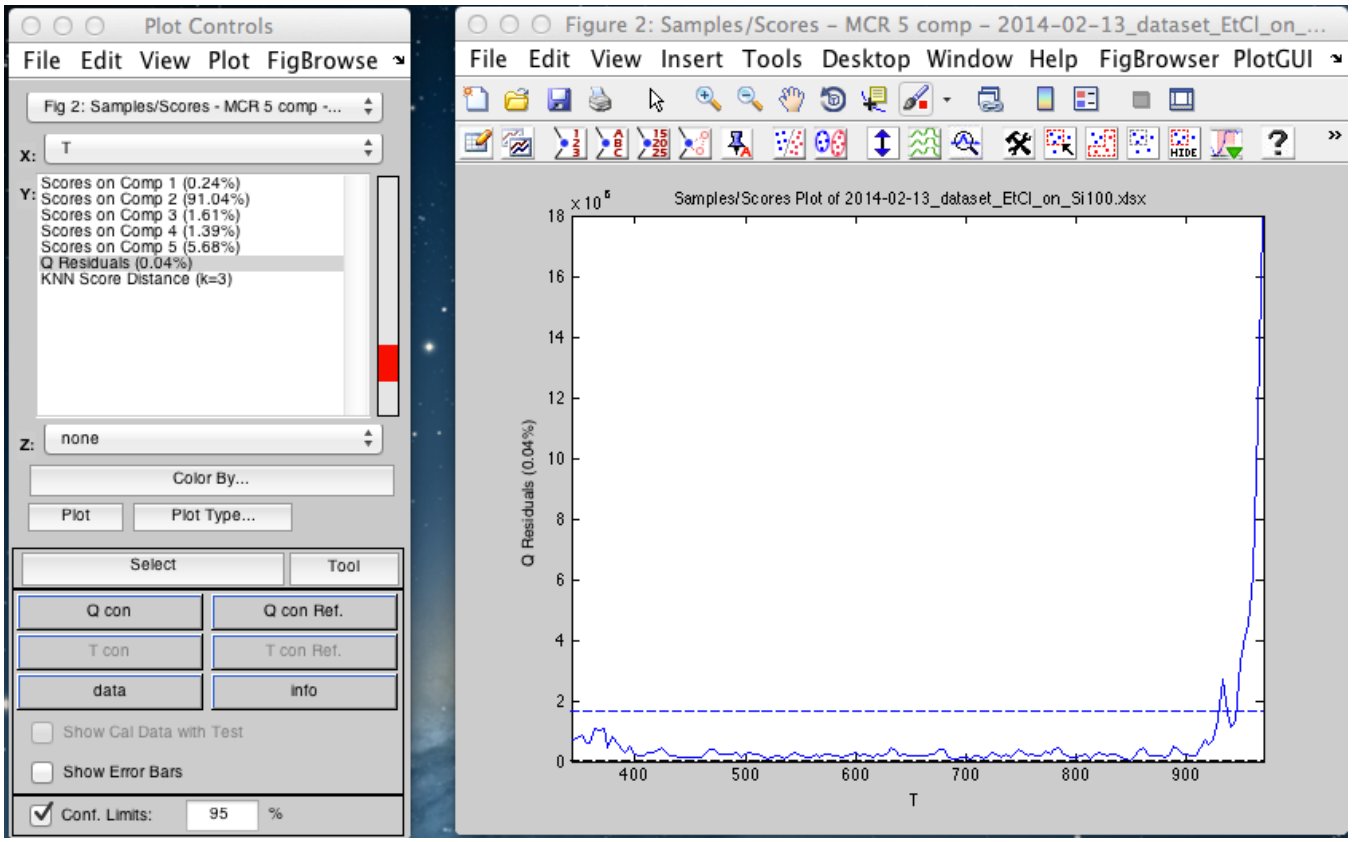


Figure 32

A close inspection of the scores shows that components C3 and C4 are essentially the same (Fig. 33), which suggests that the consideration of a fifth component is forcing the separation of one component into two.

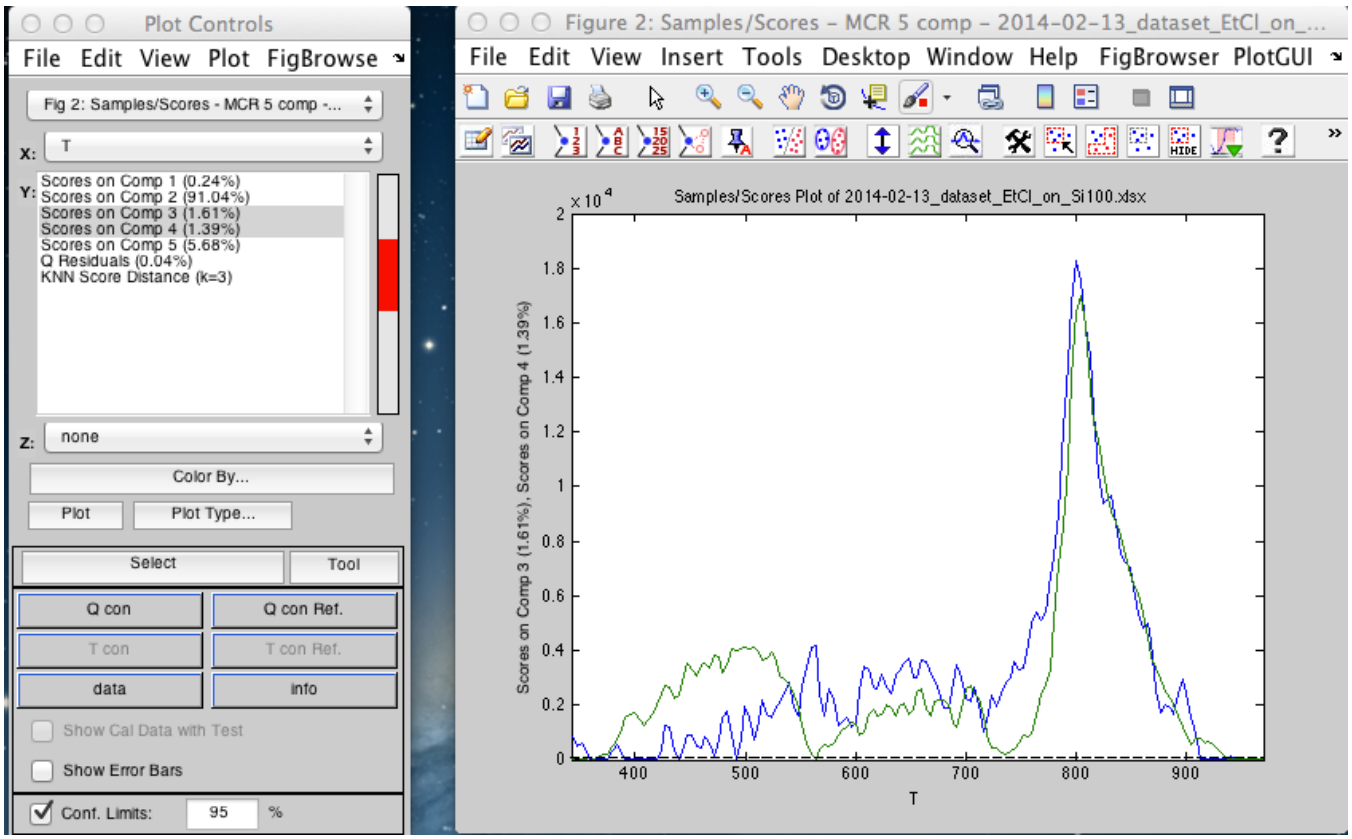

Plot Controls
File Edit View Plot FigBrowse
Fig 2: Samples/Scores - MCR 5 comp -...
X: T
Y: Scores on Comp 1 (0.24%)
Scores on Comp 2 (91.04%)
Scores on Comp 3 (1.61%)
Scores on Comp 4 (1.39%)
Scores on Comp 5 (5.68%)
Q Residuals (0.04%)
KNN Score Distance (k=3)
Z: none
Color By...
Plot
Plot Type...
Select
Tool
Q con
Q con Ref.
T con
T con Ref.
data
info
Show Cal Data with Test
Show Error Bars
Conf. Limits: 95 %
Figure 2: Samples/Scores - MCR 5 comp - 2014-02-13_dataset_EtCl_on_...
File Edit View Insert Tools Desktop Window Help FigBrowser PlotGUI
Samples/Scores Plot of 2014-02-13_dataset_EtCl_on_Si100.xlsx
x 10^4
Scores on Comp 3 (1.61%), Scores on Comp 4 (1.39%)
T
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
400
500
600
700
800
900

Figure 33

Loadings (data and residuals) are shown in fig. 34 and 35.

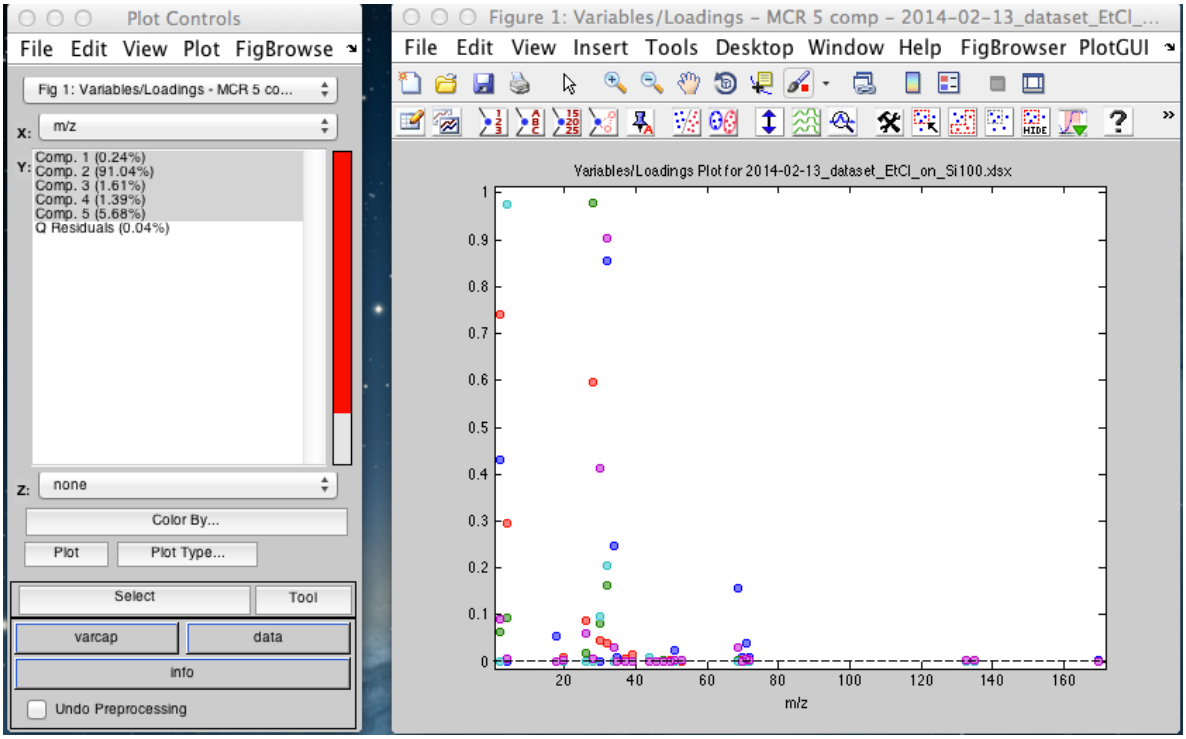


Figure 34

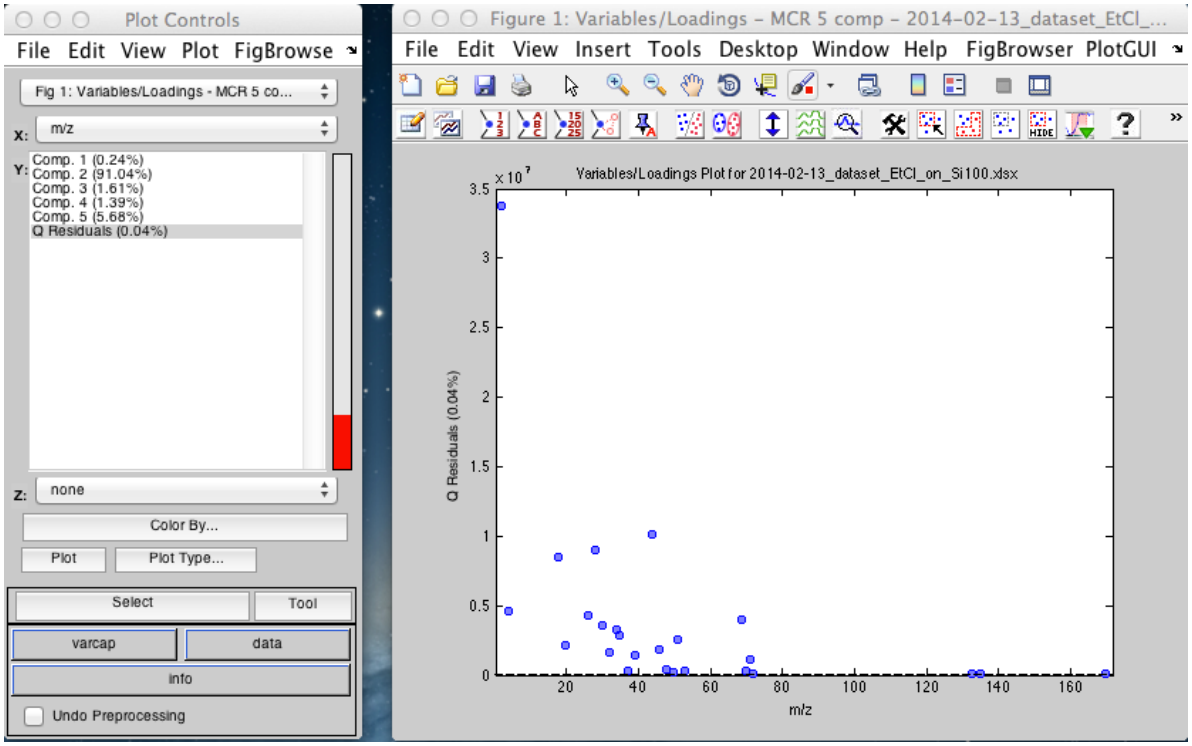


Figure 35

From our different analyses, we have found that four components seem to be the optimal number of desorption compounds. Choose again four components and build a model

Saving the model/exporting the data

Go to file / save model. Save model as mcr_4_component.

Open the saved model from MATLAB main window

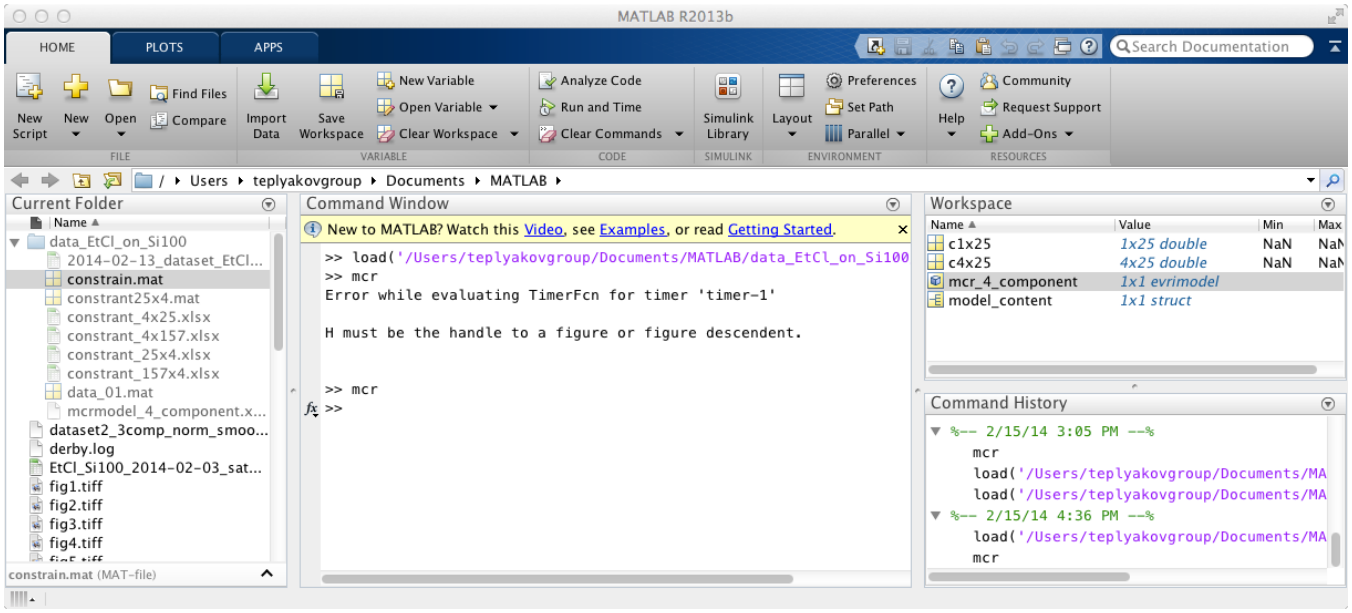


Figure 36

Double-Click on the model saved and it will appear in a window inside MATLAB. A window asking how you want to visualize the data will appear. Select **view raw contents** and you will see the data.

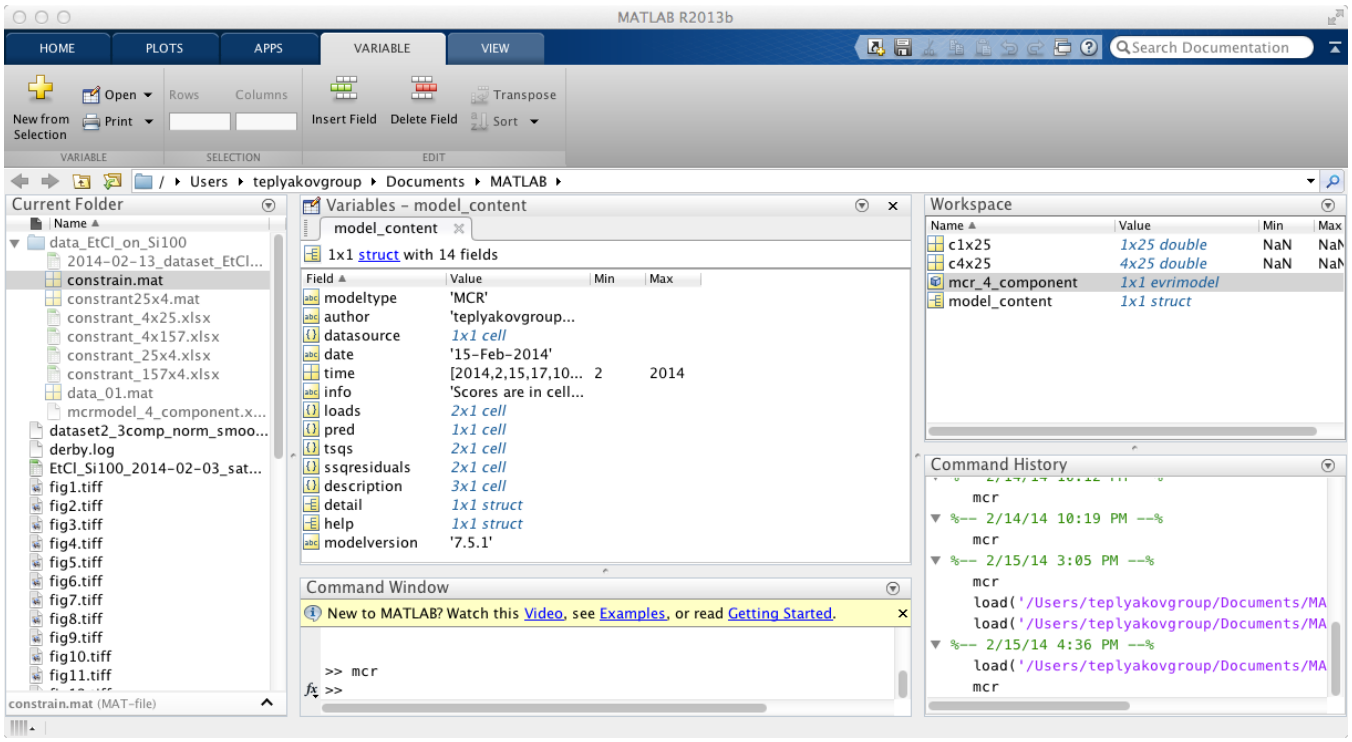


Figure 37

Select the cell named loads. This carries the two matrices C and S.

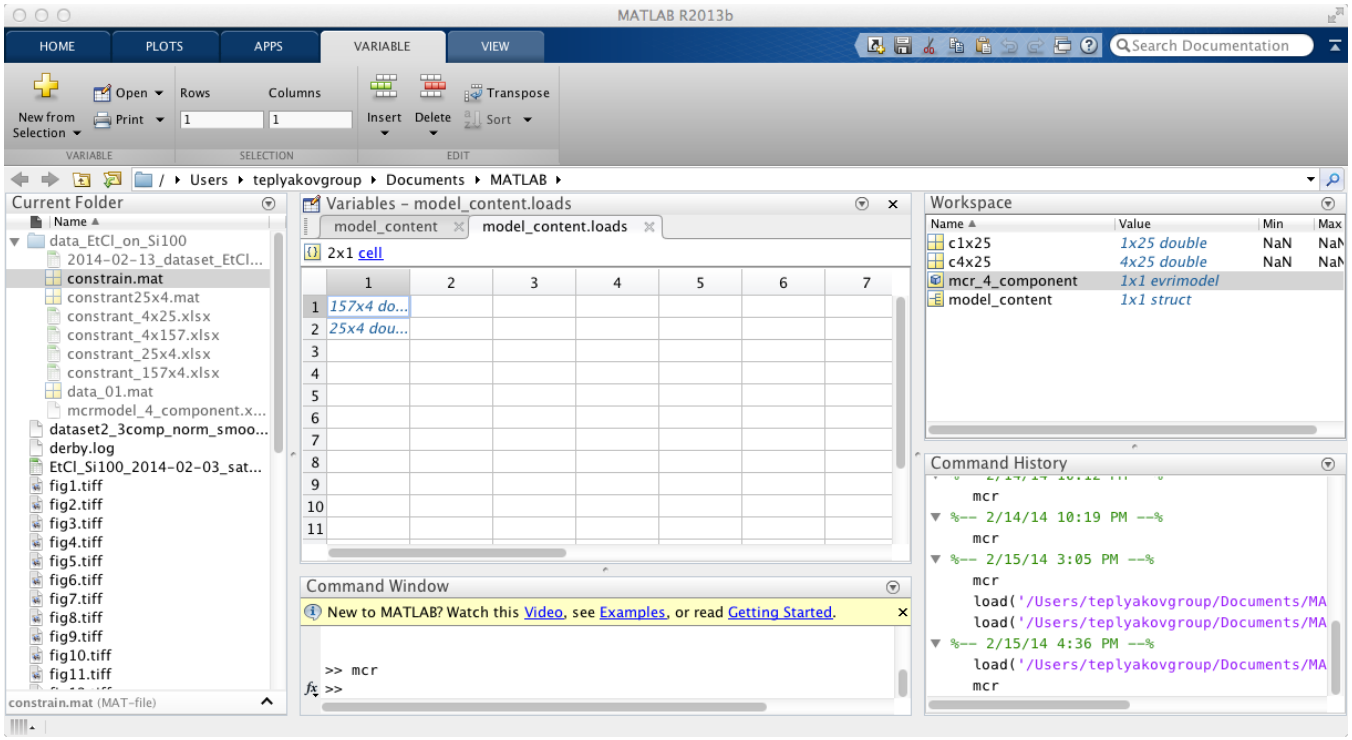


Figure 38

The cells will show the dimensions 157x4 and 25x4. These are C and S. As we double-click on these cells the data will appear.

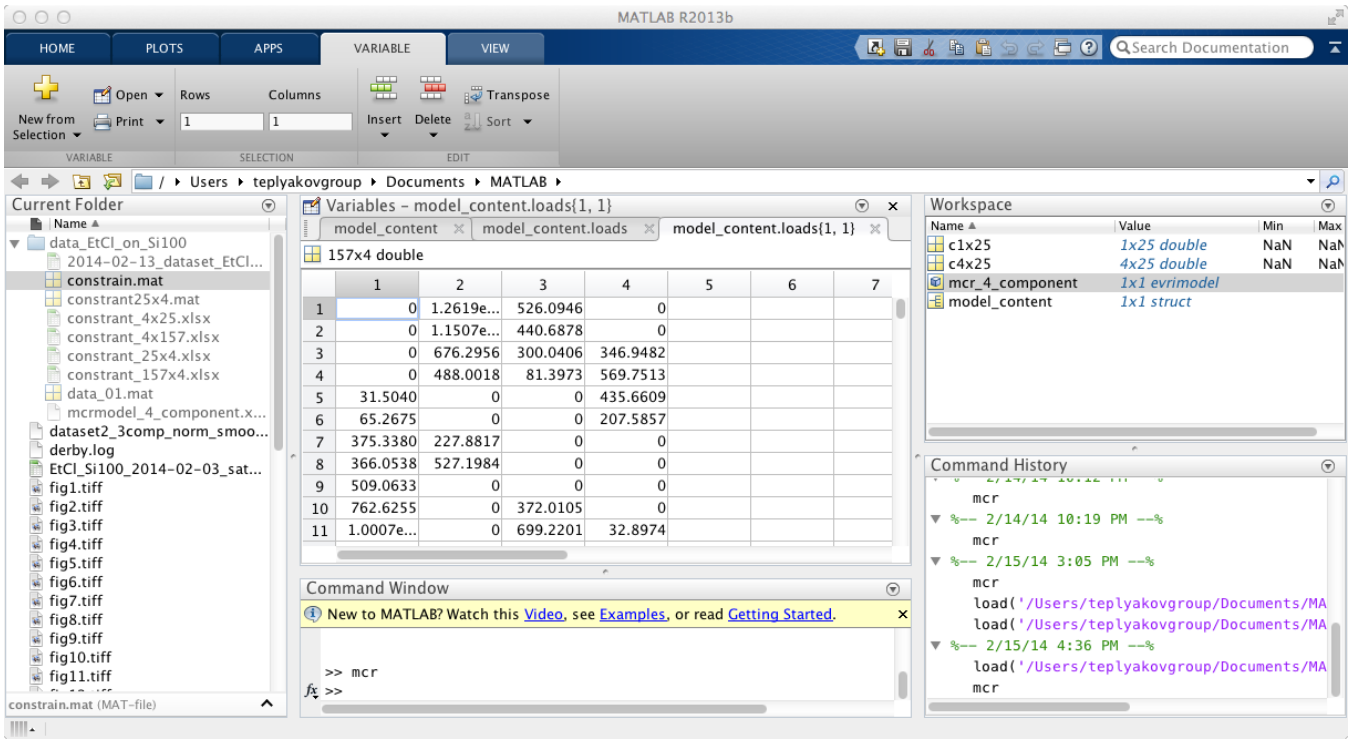


Figure 39

We can copy it and paste it into an excel file. Notice that the 157 x 4 matrix does not carry any information of temperature, so if we want to have the complete data in excel we have to insert the column with temperatures. Excel file mcr_4_component have the data successfully exported and S and C are plotted. Figures 40 and 41 show the data for C and S plotted in excel.

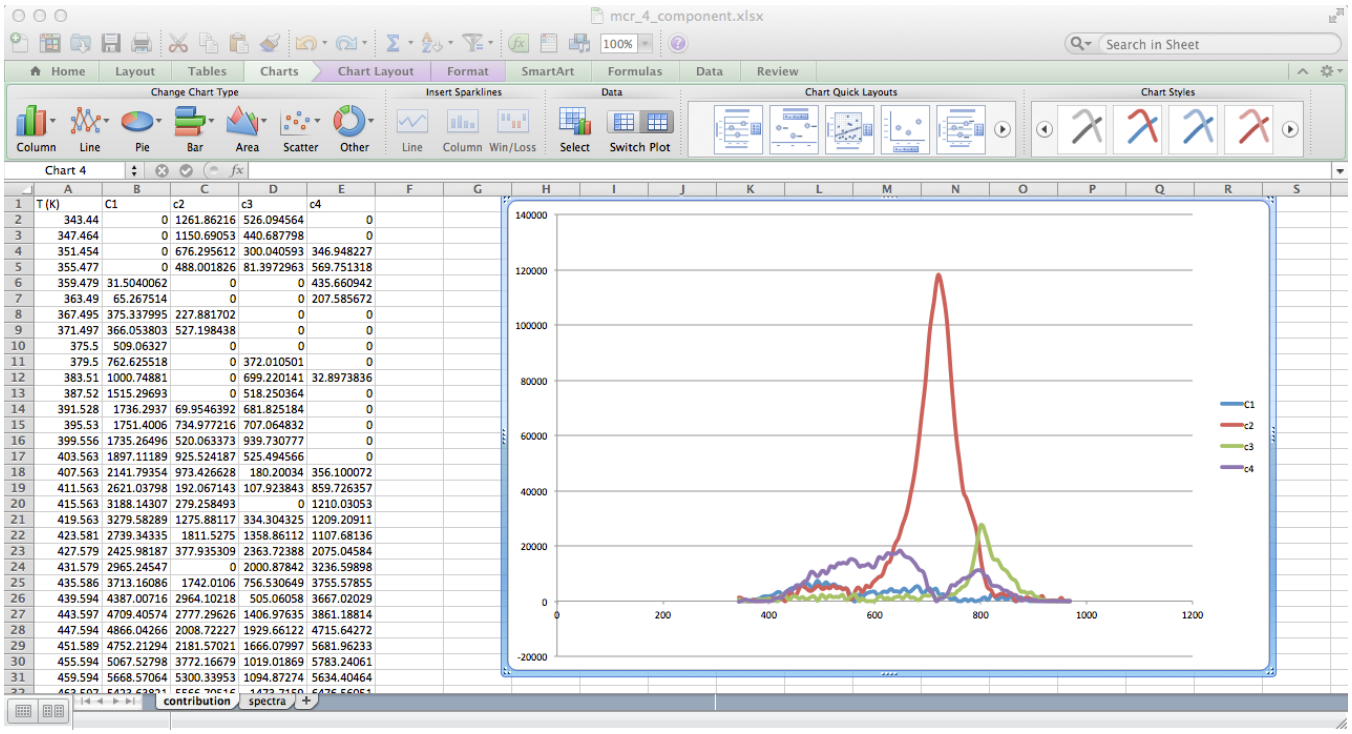


Figure 40

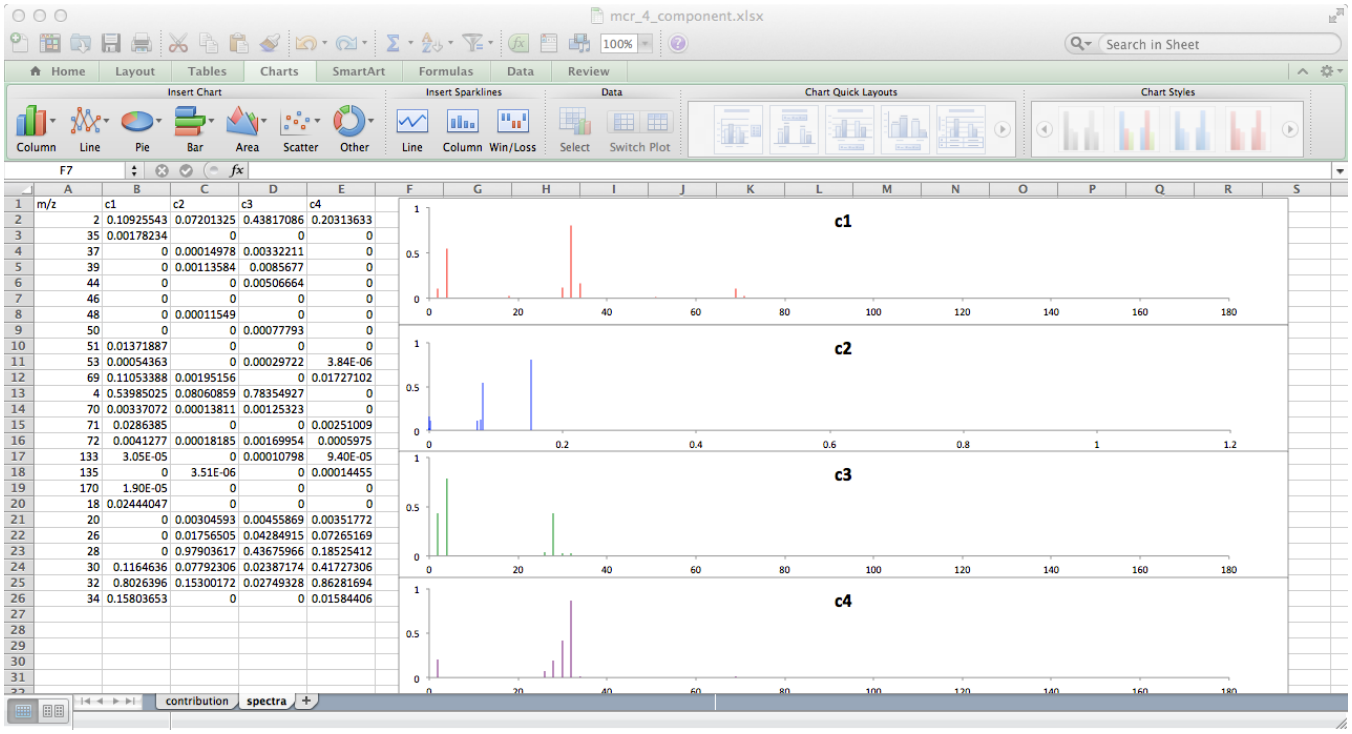


Figure 41

We have proven that 4 components are sufficient.

5. Using <u>prior knowledge</u> to further optimize the result

Now, even though the results are interesting, they have some issues. For example, In this case, it is well known that H_2 (D_2) desorbs from a silicon surface at around 800 K. Therefore, component C3 should feature only masses 2 and 4 in its cracking pattern. What we have instead is a cracking pattern governed by m/z= 2, 4, 26, 28 and 32:

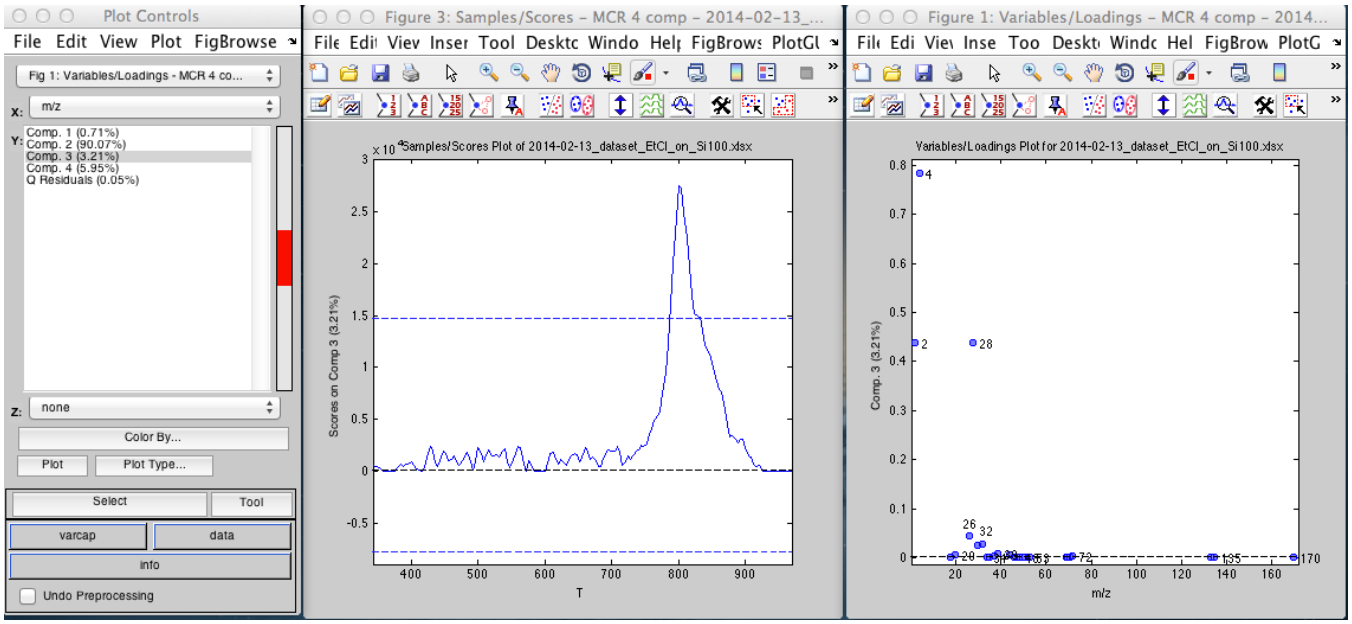


Figure 42

It is at this point (or in cases like this) we can use our chemical knowledge to improve the results. We do this by introducing this knowledge in the form of constraints, as it will be shown below. Please make sure that since we have accepted 4 components, we need to obtain again the model for four components (choose number of components and later built model). Once this is done, we can click on the mcr constraints tab:

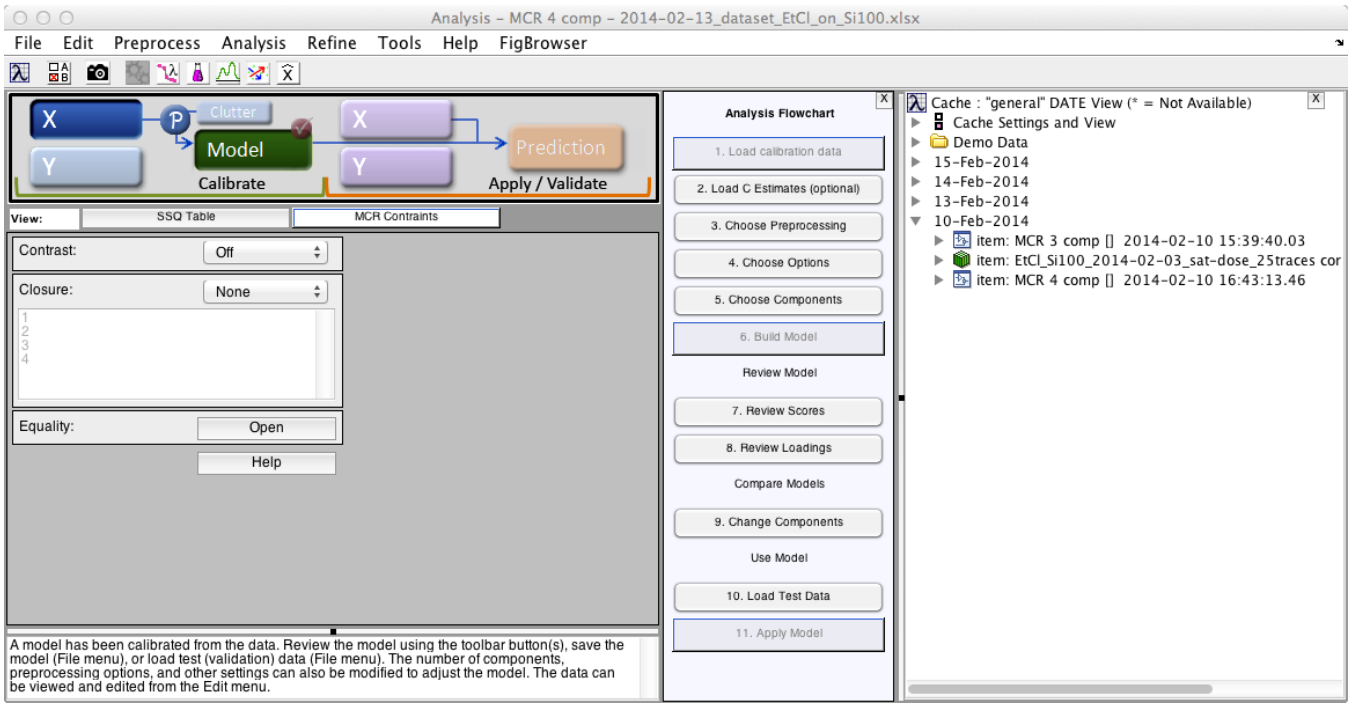


Figure 43

In the equality portion we select open. A window called “equality constrains” will pop up

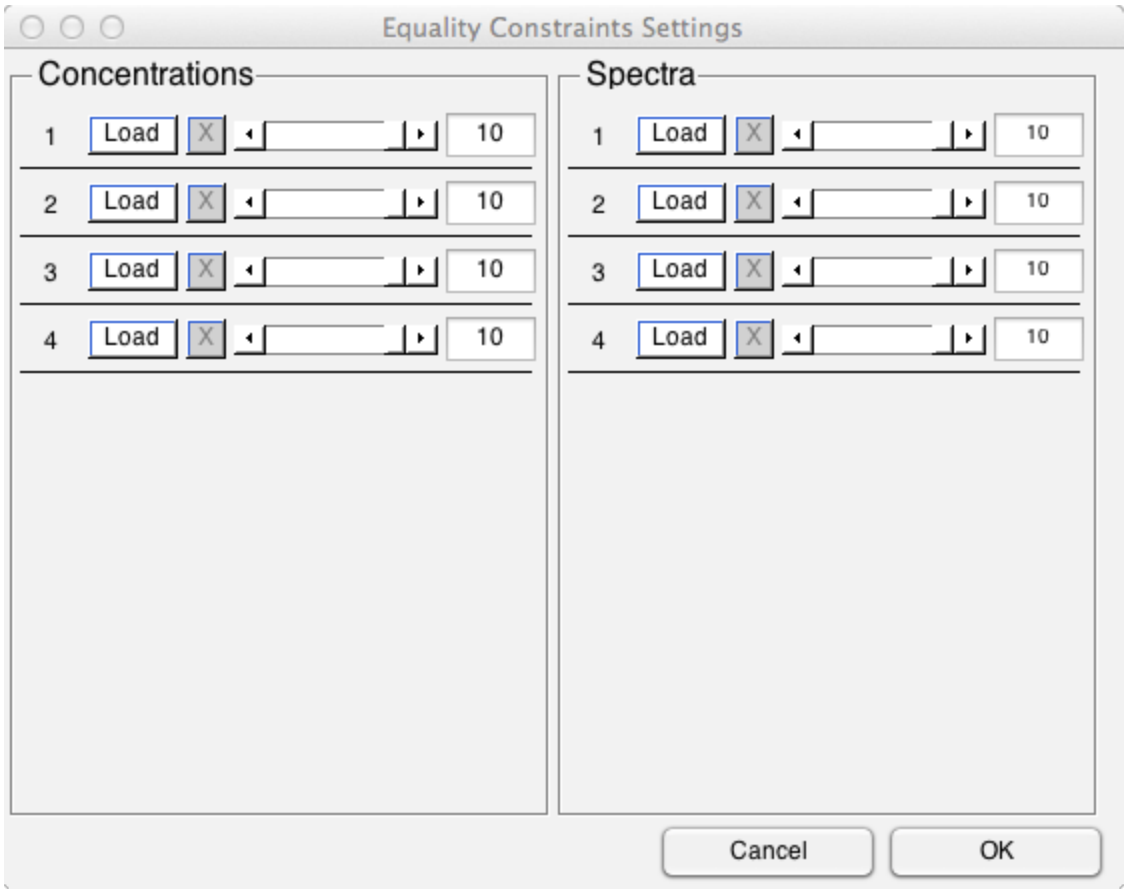
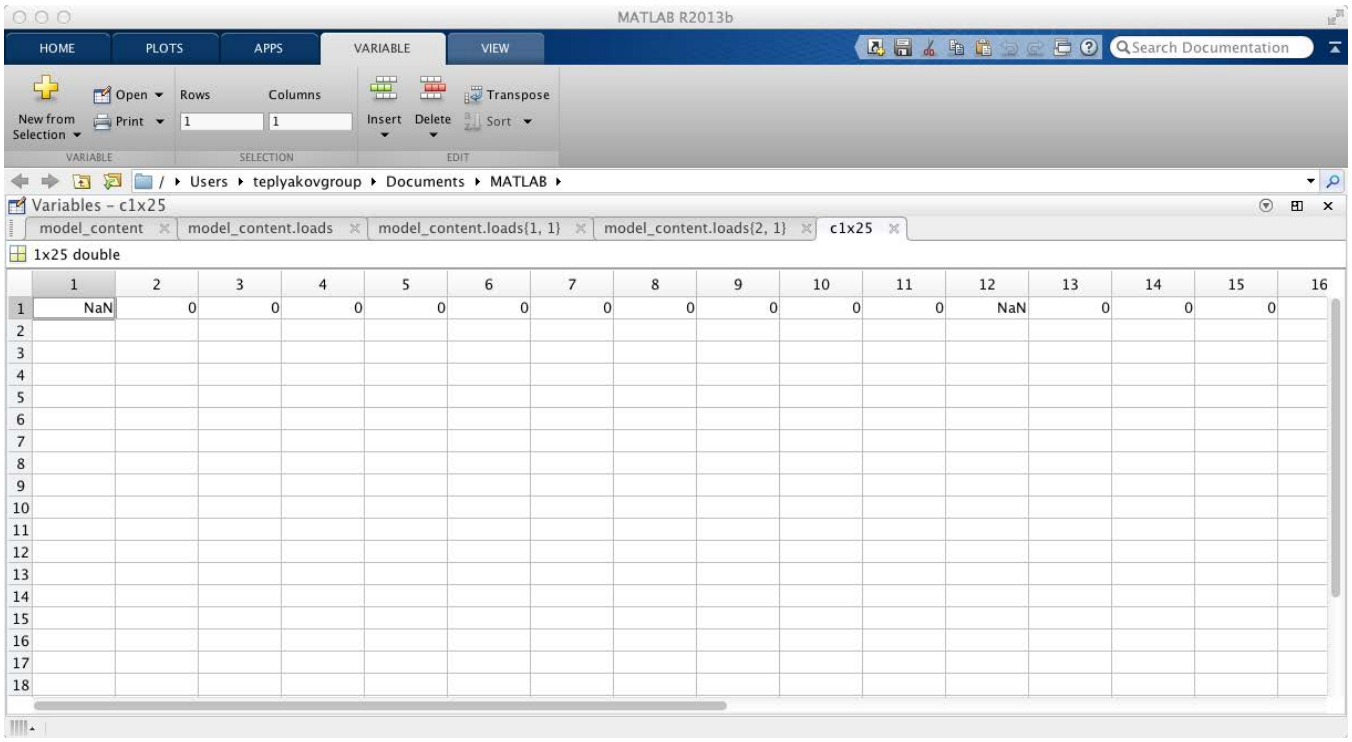


Figure 44

Here we can insert constraints in C and S matrices component by component. Constrains come from the previous knowledge about a system. Since the component that represents hydrogen evolution is the component number 3, we can insert a constraint that will allow only masses 2 and 4 to be part of this component. Since we want to put constrains in the mass values, we need to do this in the spectra section of the window. We select load in number 3 and we need to import a 1 x 25 dataset. This data set should look like this:



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	NaN	0	0	0	0	0	0	0	0	0	0	NaN	0	0	0	
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																
16																
17																
18																

Figure 45

We have set to zero all values for m/z except for masses 2 and 4, which are allowed to vary with the denomination NaN. These values of NaN appear in columns 1 and 12 because they correspond to masses 2 and 4 according to our original data .The 1 x 25 constraint row can be made in excel and be imported or it can be prepared directly in a MATLAB spreadsheet.

Once we load the data we will have the component uploaded with constrains in grey.

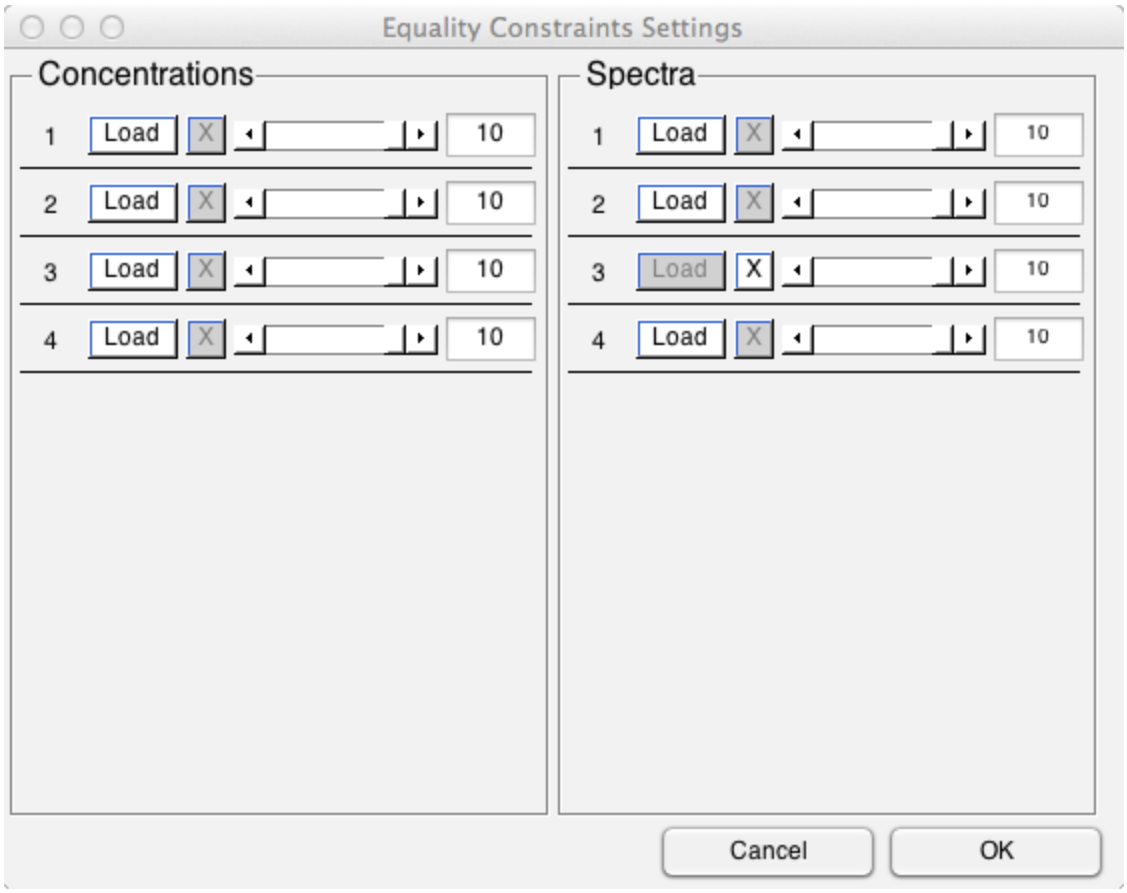


Figure 46

We click OK. Then, we build the model again. If we look at the scores and loadings for component 3, we will see that it does not carry other masses besides 2 and 4, which is what we wanted.

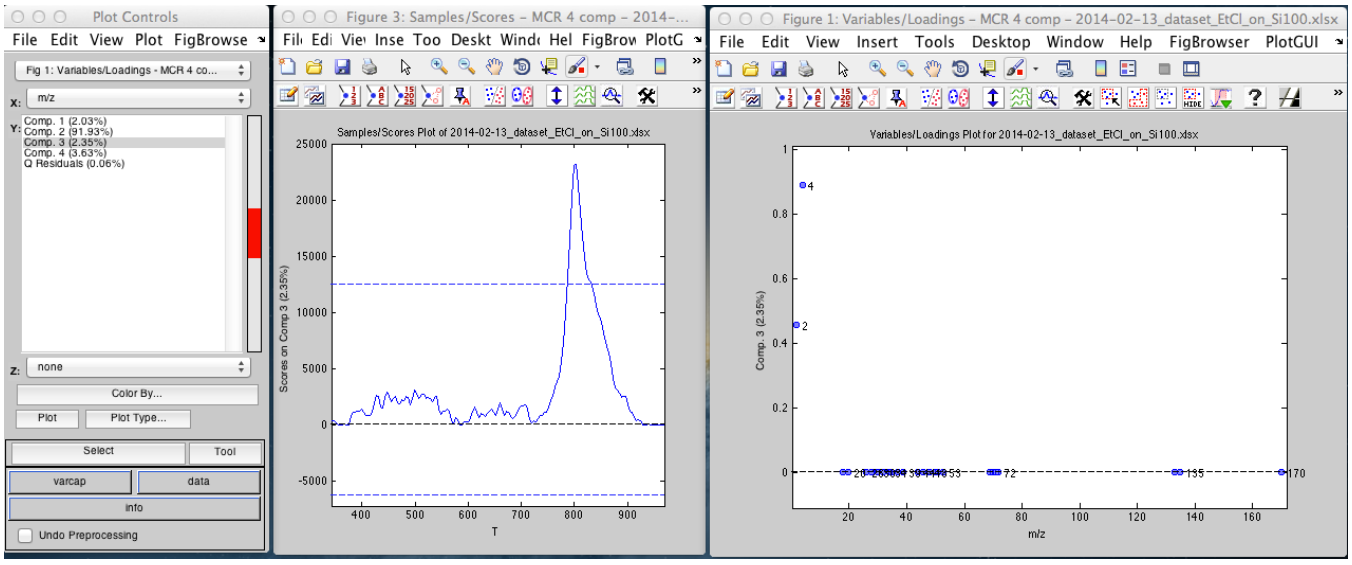


Figure 47

Of course, we can see the other scores and loadings as well:

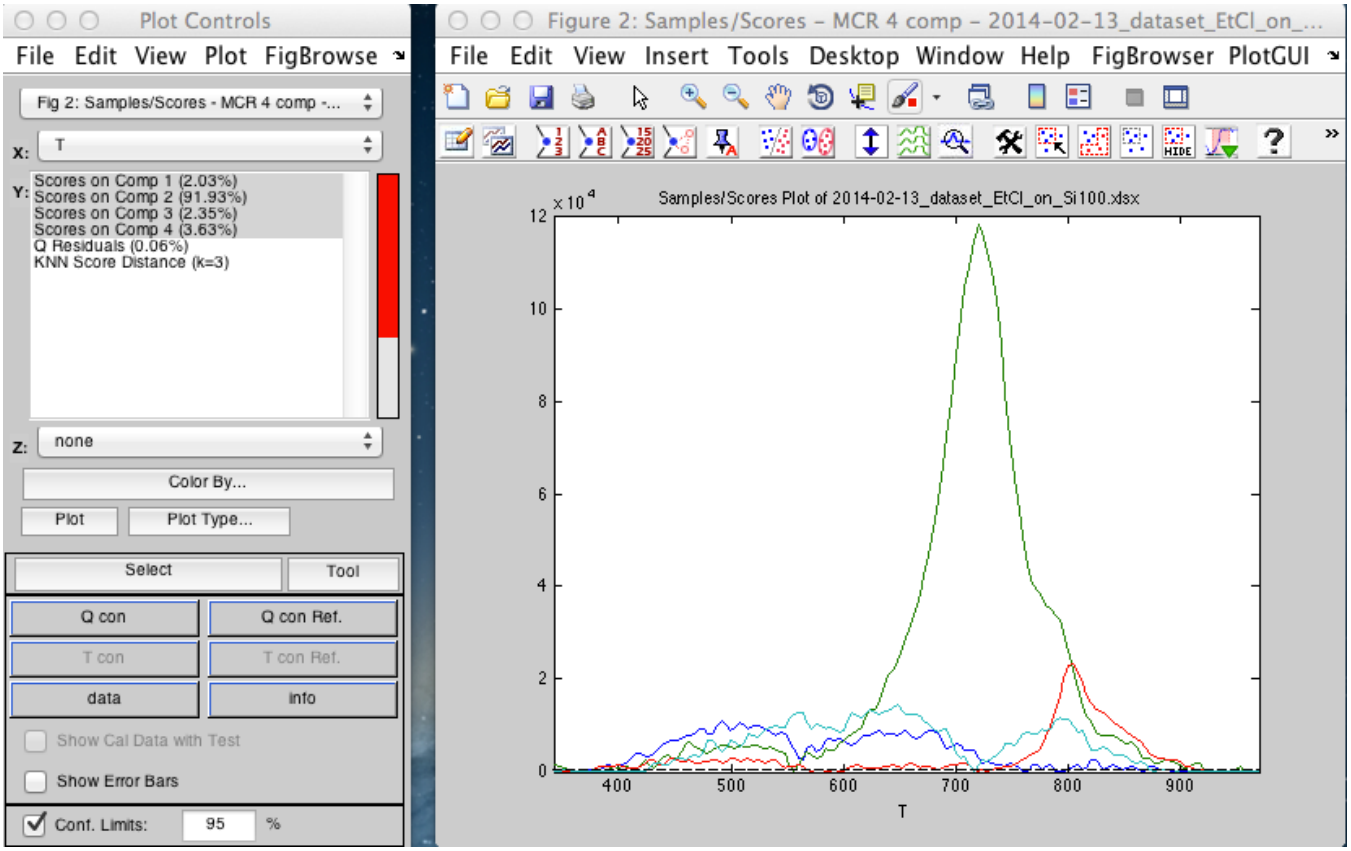


Figure 48

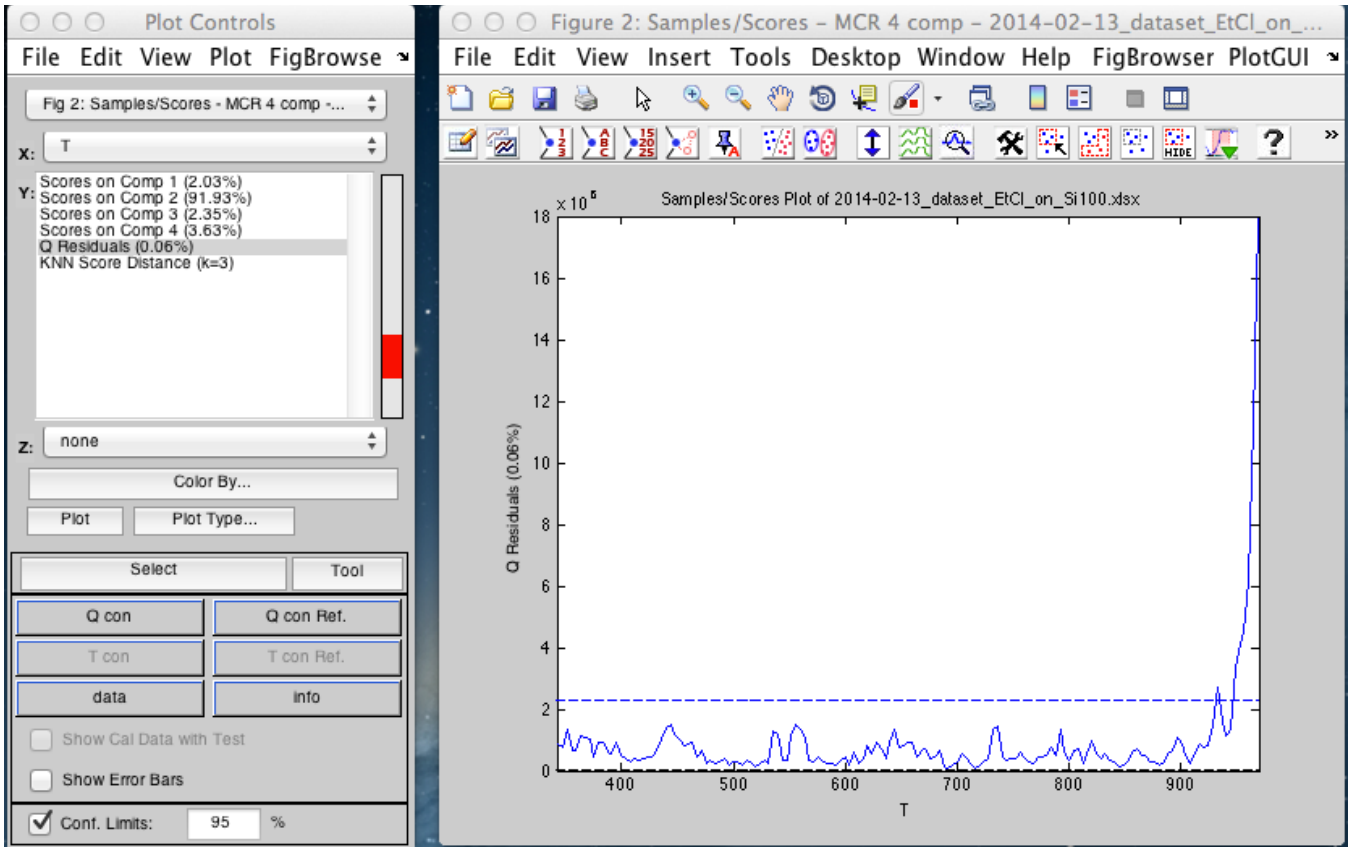

Plot Controls
File Edit View Plot FigBrowse
Fig 2: Samples/Scores - MCR 4 comp -...
X: T
Y: Scores on Comp 1 (2.03%)
Scores on Comp 2 (91.93%)
Scores on Comp 3 (2.35%)
Scores on Comp 4 (3.63%)
Q Residuals (0.06%)
KNN Score Distance (k=3)
Z: none
Color By...
Plot
Plot Type...
Select
Tool
Q con
Q con Ref.
T con
T con Ref.
data
info
Show Cal Data with Test
Show Error Bars
Conf. Limits:
95
%
Figure 2: Samples/Scores - MCR 4 comp - 2014-02-13_dataset_EtCl_on_...
File Edit View Insert Tools Desktop Window Help FigBrowser PlotGUI
Samples/Scores Plot of 2014-02-13_dataset_EtCl_on_Si100.xlsx
x 10^5
Q Residuals (0.06%)
T
0
2
4
6
8
10
12
14
16
18
400
500
600
700
800
900

Figure 49

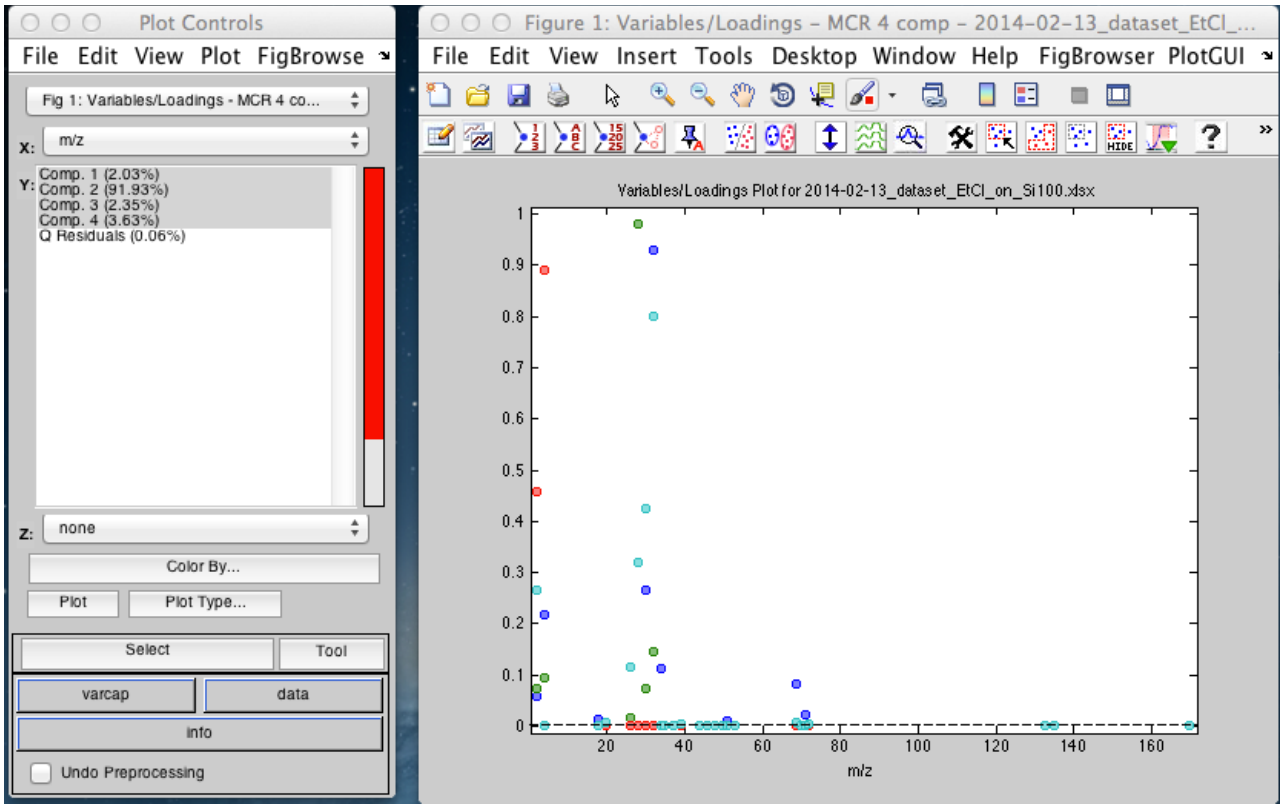


Figure 50

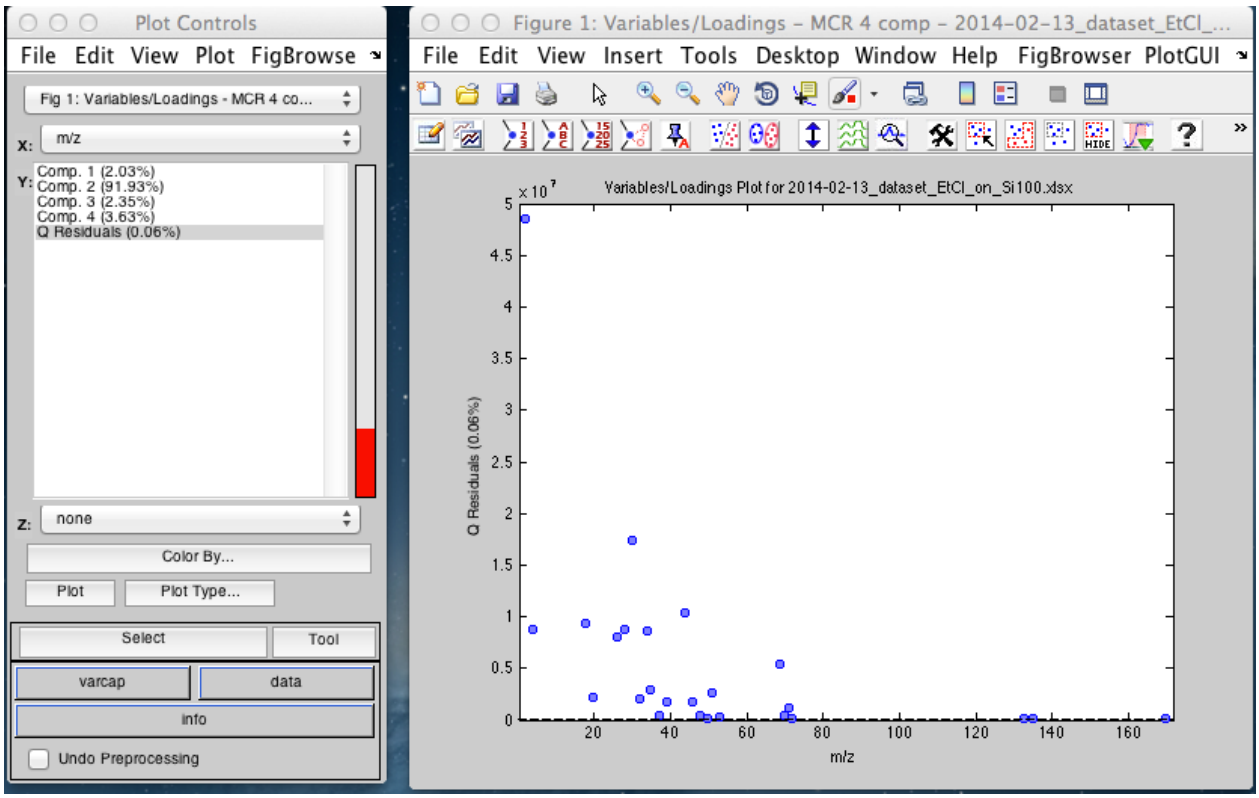


Figure 51

If we want to be more exigent about how a peak should look like, we can also insert constrains in temperature. For example, H2 (D2) is expected to leave the surface around 800 K, with the onset of desorption starting above 700 K. Then, it is possible to insert a constraint in the C matrix. It will have to have the dimensions (157 x 1) (one column) and with values of 0 up to the row 95, which is the one that corresponds to T=700 K. By doing this, now C and S of C3 appear grey in the window

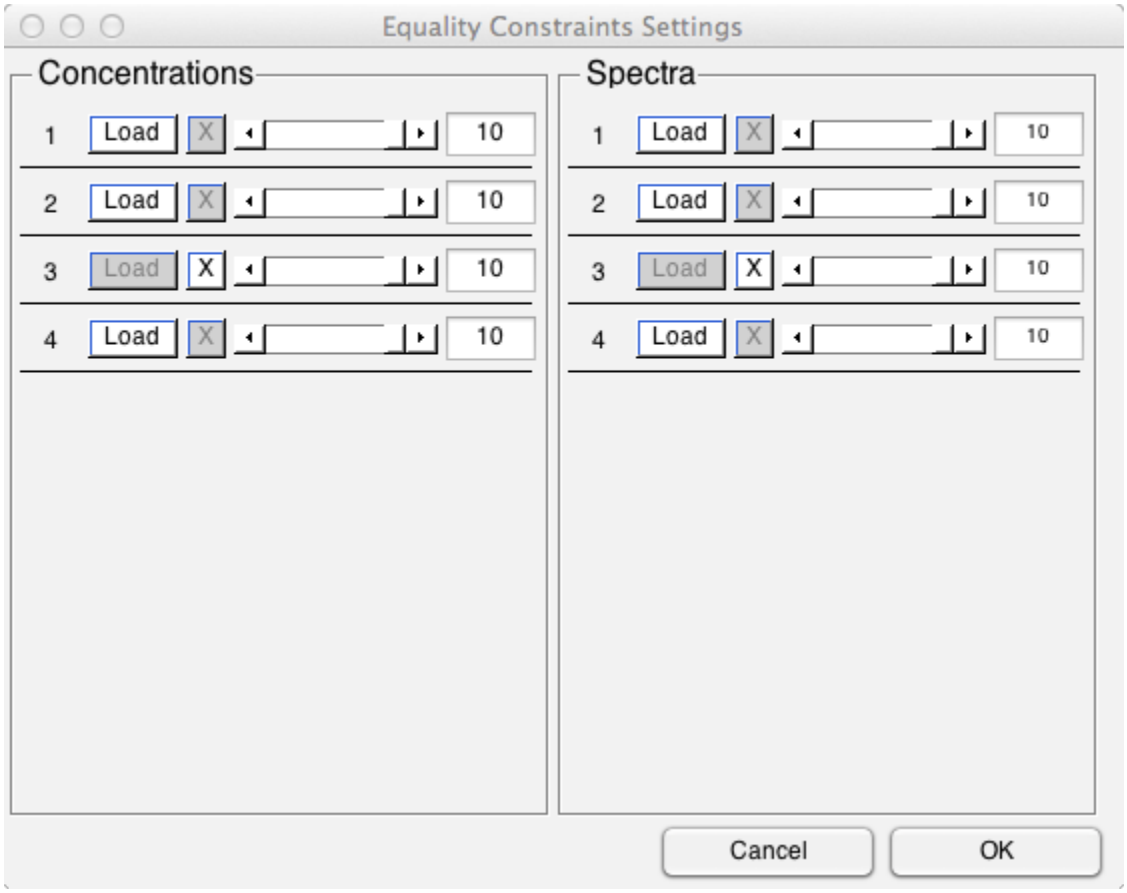


Figure 52

The result for component 3 is highlighted in Figure 53

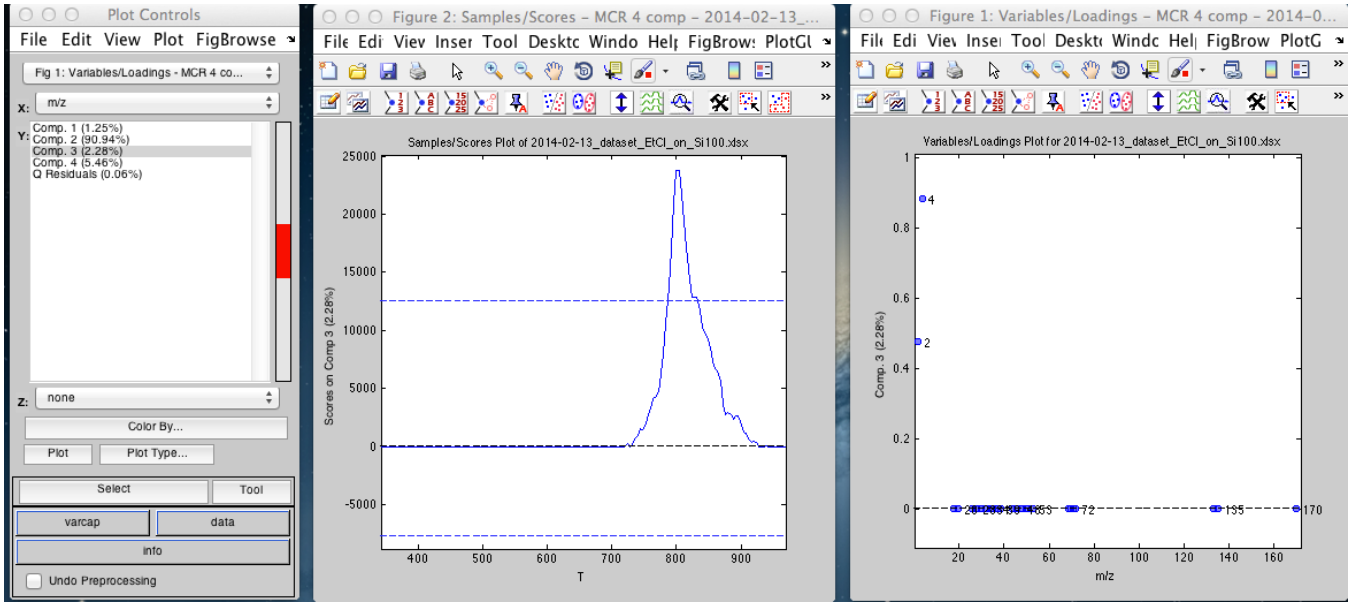


Figure 53

This model has been saved and exported into excel. The data is available as file mcr_4_component_constrains.