MatNMR

GUI for statistical analysis of 2D NMR spectra in Matlab

Mattias Hedenström

Department of Chemistry

Swedish NMR Centre

Umeå University

1 INSTALLATION AND SYSTEM REQUIREMENTS

If you received MatNMR as a Matlab application, go to Apps toolstrip in Matlab , click on *Install App* and locate the MatNMR installation file.

If you received MatNMR as a zip-file containing the source code, simply unpack it to a folder and make sure this folder is included in the Matlab path. To define a new path, go to *File -> Set Path -> Add with subfolders*.

The scripts have been tested with Windows 7, Windows 10 and MATLAB versions R2014b and 2015b.

2 FILES

2.1 INCLUDED FILES (ONLY AVAILABLE IF YOU HAVE THE SOURCE CODE, NOT THE STAND-ALONE APPLICATION)

- *MatNMR.m and MatNMR.fig* - Main script for the GUI
- AnalysisPanel.m and AnalysisPanel.fig – GUI for certain function within MatNMR
- *do_log .m* - Writes a log file (.txt) with miscellaneous information about the processed data in the project.
- *file_check.m* – Checks which project files are present in current project folder
- *Ppm_to_dp.m* - Converts ppm values to data point indexes. Used in various processing steps.
- *Progressbar.m*
- *NMR_align2D.m* - Alignment routine
- *ROI_to_vector.m* - Reshapes a spectral region to a row vector
- *Excl_noiseDB.m* - Removes all data points below a defined threshold

2.2 CREATED FILES

- *SpectralData.mat* – Contains both the originally imported spectral data and the spectral data processed in MatNMR.
- *MatNMRspectra.mat and MatNMRspectra.xls* - The final processed data in Matlab and excel format, respectively. The excel file also contain the titles of the spectra. Note: Because of size restrictions in Excel, the spectral data will be saved as column vectors instead of row vectors if the number of data-points exceeds 16384.
- *Settings.mat* - Keeps track of crucial flags and information.

- *Logfile.txt* - Contains information about processing flags and included and excluded spectral regions in an easily accessible format.
- corr.mat – Contains the pearson correlation coefficients if a correlation analysis has been done.

3   WORKING WITH MATNMR

This chapter will be focused on the practical aspects on how to use MatNMR. A more general description of this GUI can be found here in the following publications:

1. Hedenström M, Wiklund S, Sundberg B and Edlund U. Visualization and interpretation of OPLS models based on 2D NMR data. Chemometrics and Intelligent Laboratory Systems 2008, 92:110-117.
2. Hedenström M, Wiklund-Lindström S, Öman T, Lu F, Gerber L, Schatz P, Sundberg B and Ralph J. Identification of Lignin and Polysaccharide Modifications in Populus Wood by Chemometric Analysis of 2D NMR Spectra from Dissolved Cell Walls. Molecular Plant 2009 5:933-942.
3. Öman, T.; Tessem, MB.; Bathen, T. F.; Bertilsson, H.; Angelsen, A.; Hedenström, M.; Andreassen, T., Identification of metabolites from 2D 1H-13C HSQC NMR using peak correlation plots. BMC Bioinformatics 2014,15, 413.

3.1   GETTING STARTED

Start MatNMR with the command *MatNMR* in the Matlab command window or by clicking the MatNMR app icon. This will bring up the main GUI (see Figure 1).
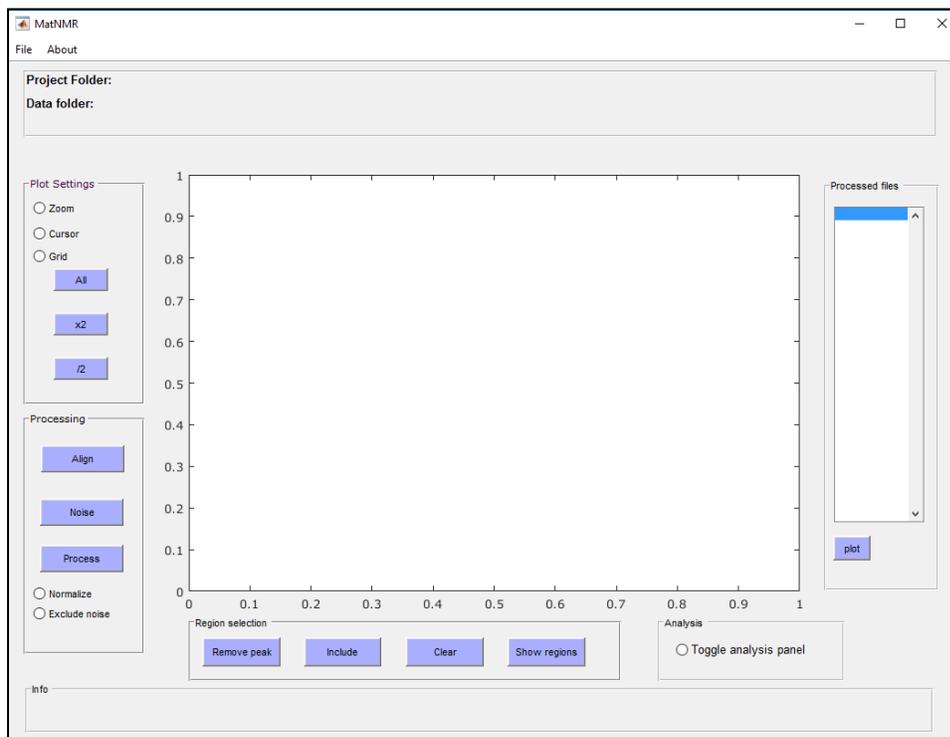
Figure 1. GUI overview

3.2 Creating projects and importing NMR data

First, the project folder needs to be defined by clicking **New Project** in the file menu. All the files generated by MatNMR will be stored in this folder. Next step is to import the spectral data and this is done by clicking **Import NMR Data** in the file menu. The import function currently only works for spectral data processed in Topspin and the spectra need to be stored in the standard Bruker file structure. For example, if 10 different 2D spectra are stored in the folder Test as <exp_nr> 1-10 then the 2rr file from /pdata/1 for each <exp_nr> will be imported.

C:\NMRdata\Test\ <exp nr>\pdata\1**2rr**

Important: All spectra need to contain the same nr of data-points, i.e make sure the processing parameters SI and SI2 in Topspin are the same for all spectra. Also, don't set SI and SI2 too high, 1024 data-points in F2 and 512 in F1 is usually sufficient. Higher values could lead to memory problems in Matlab.

You will now see the imported spectra in the Processed Files list on the right hand of the GUI.

3.3 Open an existing project

You can also open an existing project with **Open Project** in the file menu and select a folder with an existing project.

3.4 Plotting spectra

Figure 2 shows the GUI after data import and plotting a spectrum. After importing NMR data, the path to the Project folder and Data folder will be shown at the top (1).
The plot button (2) will plot the spectrum marked in the list. Plot parameters can be controlled with the plot settings on the left-hand side of the GUI. The title of the spectrum (as defined in Topspin) will be shown in the top right corner of the plot window. **Note: Only the first line of the title that is set in Topspin will be imported so make it as informative as possible. Note: If an empty window is shown after pressing plot, try clicking the All-button on the left.** Plot tools can be found in the menu on the left (3).

- x2 and /2 - controls the intensity
- All - shows the whole spectrum
- Zoom - enables click-and-drag zoom mode
- Cursor - get chemical shift information by clicking in the spectrum
- Grid – toggle grid on/off
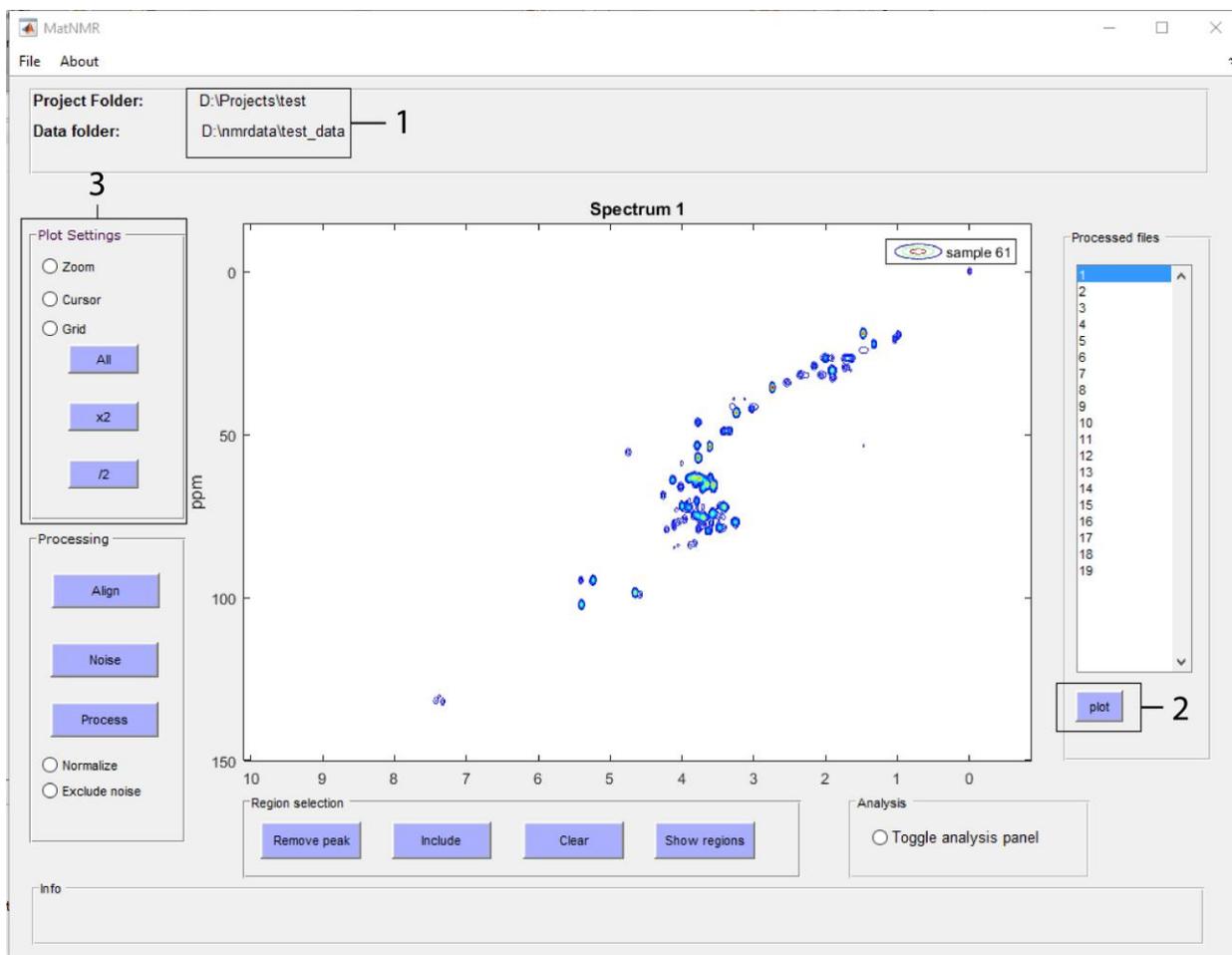
Figure 2. Plotting spectra.

3.5 EXPORTING PLOTS

The spectrum currently plotted in the main window can at any time be exported as a Bruker 2rr file for further visualization in Topspin. All types of plot can be exported in this fashion, regardless of if it is the original spectrum, the processed spectrum, a correlation plot or a loading plot.

Remember that the 2rr file in itself is not enough in order to open it in Topspin, it needs to be placed in a directory where the other files found in, as an example, C:\NMRdata\Test\ <exp nr>\pdata\1\, also exist, like acqu, proc and so on... What I normally do is to create a new folder called "2" in the /pdata folder for one of the imported spectra and then copy all files (except the 2rr) from /pdata/1/ to /pdata/2/.

3.6   DATA PROCESSING

Once a set of NMR spectra has been imported, it is time to process the data within MatNMR. This will produce a matrix containing the spectral data that can be used for further analysis. There are several optional preparation steps, found in the Processing and Region Selection tool groups in the GUI that can be performed before constructing the final data matrix. These are described below.

3.6.1   Alignment

Click **Alignment** and then zoom in on a reference peak and press Enter. Make sure that no other peaks with higher intensity then the desired reference peak is present in the zoomed region for any of the imported spectra. This alignment routine will search for the maximum intensity in the defined region for all spectra and align the peak maximum to the chemical shift of this peak in the first spectrum. If you have calibrated your spectra in Topspin, the alignment routine here is not strictly necessary.

3.6.2   Noise exclusion

The purpose of this step is to reduce the size of the final data matrix by removing data-points below a manually defined noise threshold. Clicking **Noise** will plot the first spectrum in the "processed files" list as a line plot. Zoom in on a small peak and switch to cursor mode in the Plot Settings menu, then left-click on a peak with slightly higher intensity than the noise level and press Enter, see Figure 3. This will store the noise threshold in the file settings.mat.
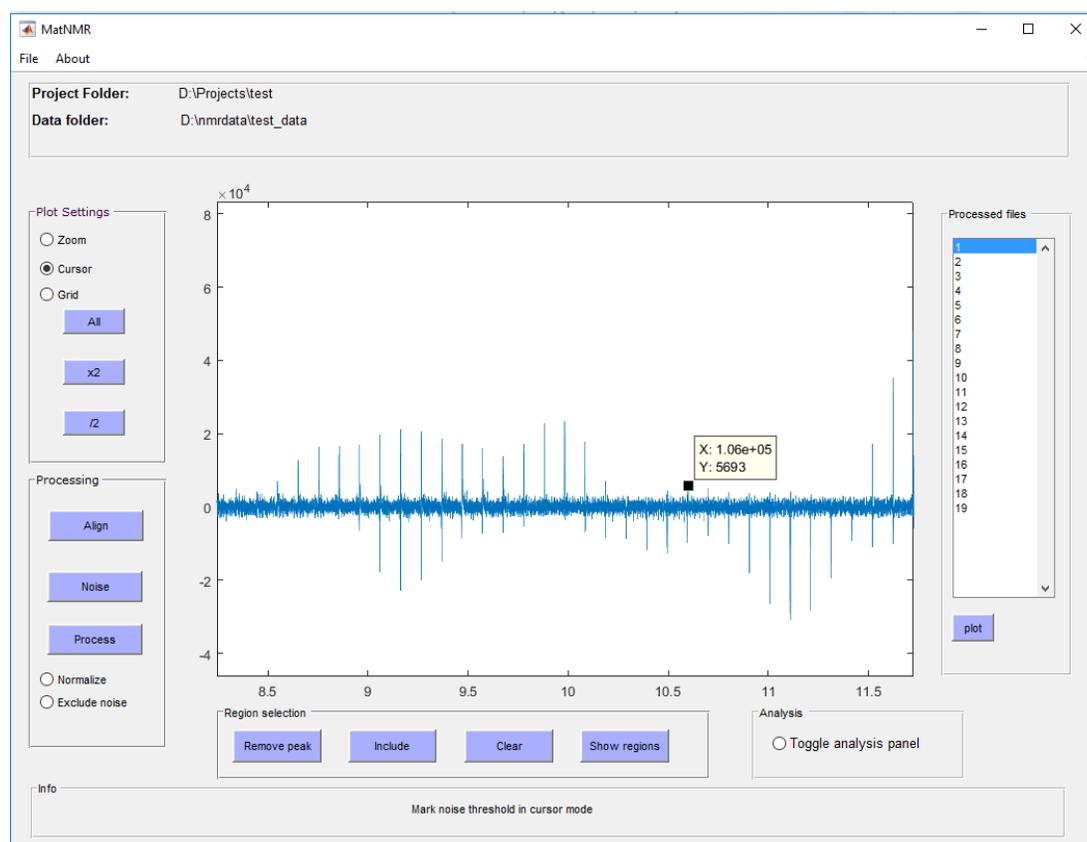

Figure 3. Noise exclusion

3.6.3   Region selection

With the menu below the spectrum window, it is possible to select a sub-region that you want to include in the final data matrix and also to exclude regions/peaks that are of no interest such as solvent peaks. If no region is selected to be included then the whole spectrum will be used.

Remove peaks by clicking **Remove peak** then click and drag a box around the peaks you want to remove. Once all peaks are marked, press Enter to exit the selection mode.

Include a region by clicking **Include** and drag a box around the area you want to include. If you selected the wrong region by mistake you can remove all selected region by clicking **Clear**. This will delete all excluded regions as well as the included region.

Information about removed and included regions will be stored in settings.mat and also in logfile.txt

Show region will display the selected and excluded regions. Figure 4 shows an example where one peak has been excluded (red) and a sub-region has been selected (green).
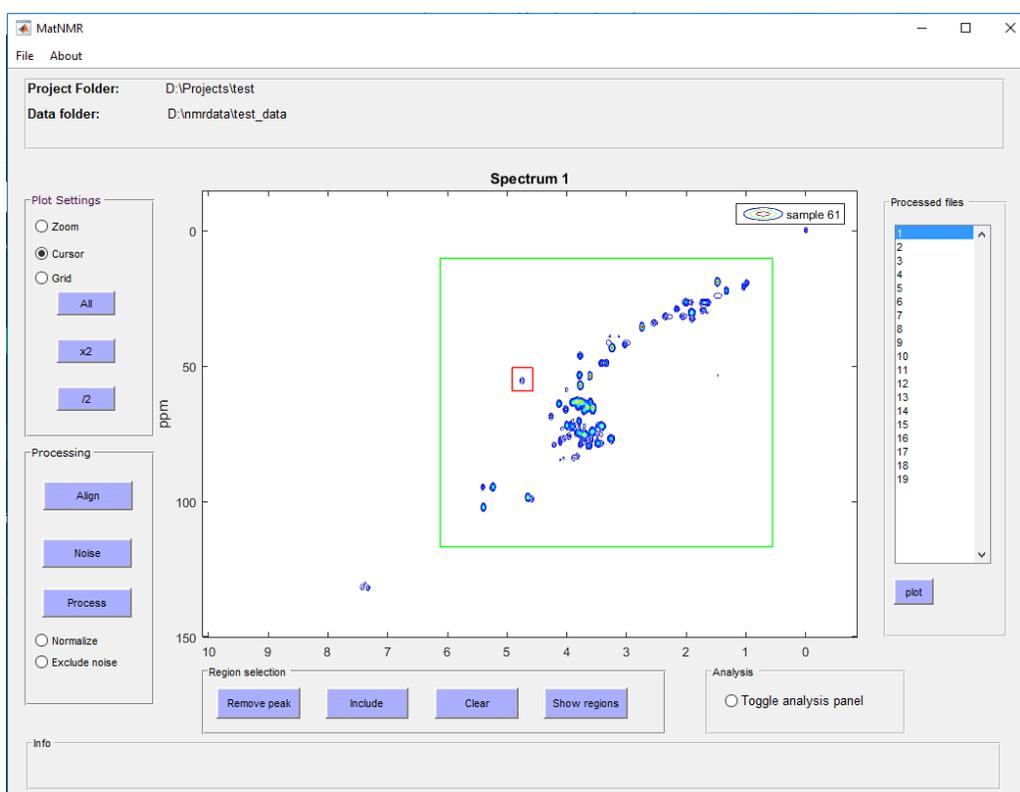
Figure 4. Include and exclude spectral regions.

3.6.4   Process

By clicking **Process**, the final data matrix containing the selected region will be generated. This matrix will be stored as MatNMRSpectra.mat where each row will contain the data points (from the selected region or the whole spectrum) from one spectrum. Check the radio buttons *Norm regions and Exclude noise* if you want to perform those steps in the processing. The file MatNMRSpectra.xlsx will also be created after the processing, see chapter 2.2 for a description of the excel file.

If the *Norm regions* option is enabled, the selected region will be normalized to a constant sum. *Exclude noise* will delete the data points that are below the set threshold in ALL spectra, i.e. as long as the intensity for a specific data point is above the threshold in any spectrum, this data point will not be excluded. Strongly recommended because it reduces the size of the final matrix considerably! Technical note: The original position of the data points that are left after this step is stored in settings.mat as setting.Index. This index is needed to correctly visualize the processed spectral data.

4   ANALYSIS PANEL

Open the analysis panel by clicking the *Toggle analysis panel* in the main GUI. This will also switch the main GUI into "analysis mode" where you no longer can use the processing or region selection functions and plotting spectra will plot the data in the MatNMRSpectra.mat, not the spectral data originally imported. For example, if you included a sub-region in the processing step, only this region will be plotted. The analysis panel is divided into 2 parts, one for correlation analysis and one for visualization of loadings from a multivariate model, see Figure 5.

When the Toggle analysis panel is toggled off, the analysis panel will close and the main GUI will operate as described in chapter 3.
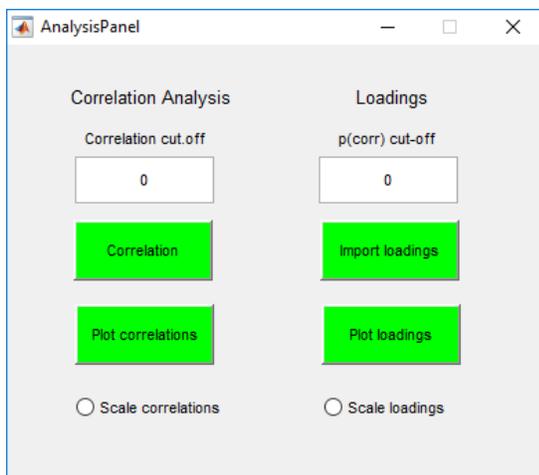

Figure 5. Analysis panel

4.1   CORRELATION ANALYSIS

The aim of this feature is to determine which peaks correlate with each other in a set of 2D spectra and to visualize the correlations.

First step is to plot a spectrum processed in MatNMR in the main GUI. This is done by simply chosing a spectrum in the list and press **plot** (the plot button should be green when the analysis panel is open). You then need to define a peak that you want to calculate correlations to, called the seed peak. This is done with the **Correlation** button in the analysis panel. When you click the Correlation button, a crosshair will show up, place it on the peak of interest and click left mouse button.

The pearson correlation coefficients to all other datapoints in the processed spectrum will be calculated and stored in corr.mat.

Plot correlations will show the correlations in the main window and the title will show which chemical shifts were used as the seed peak. There are 2 parameters that can be used to modify how the correlations are represented. The first one is the *Correlation cut-off* that can be set in the analysis panel. This will take a value 0-1 and only the correlations with a higher absolute value than the cut-off will be plotted. This is useful because in most cases we only want to know what peaks correlate most strongly to our seed peak. The other option is the *Scale correlation* radio button. If this is ticked when the correlations are plotted then the correlations for each data-point will be multiplied with the standard deviation this data-point has in the processed spectra. Instead of being a value between 0 and 1, the peaks in the correlation plot will then correspond to the original peak intensities scaled by the correlations, resulting in a more visually appealing representation of the data. Note that the correlation cut-off will still operate on the true correlations (0-1). Shown in Figure 6 is an example of a correlation plot where only correlation >0.9 are shown and where the correlation coefficients have been scaled.
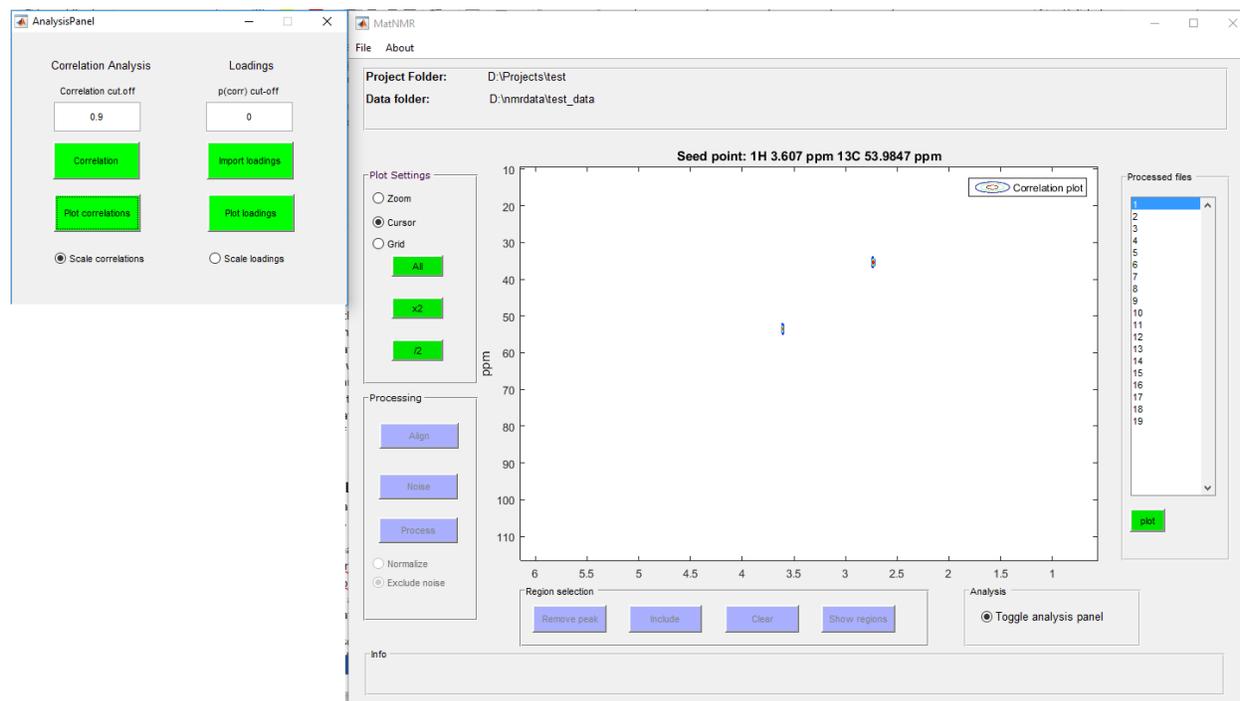

Figure 6. Example of a correlation plot

4.2 LOADINGS

This part of the analysis panel is used for visualization of loading vectors from multivariate models, such as PCA and OPLS.

The analysis panel is streamlined to work with loadings from multivariate models calculated in Simca-P (Umetrics, Sweden) where you make a list of the loading vector, p(corr), you're interested in and save that as a matlab file. **Import loadings** is used to locate this file and to read it. Plot loadings will plot the loading vector as a 2D plot. The *p(corr) cut-off* can be set between 0 and 1. For example, you can set this to 0.8 to only show the loadings most important in the model. *Scale loadings* works just as *Scale correlations* described in chapter 4.1.

Note! The section below has not been tested properly as of yet.

It is also possible to visualize loadings from e.g a PCA calculated in Matlab. Below is shown a simple example on how to do this. X is the data matrix you will get when loading the MatNMRspectra.mat

[t,p] = pca(X,2) %calculate a two-component PCA

DATA = p; %the loading vector must be called DATA for compatibility reasons

Save('loadings.mat','DATA'); %save loadings, the name of the .mat file can be whatever you want

Now you can click **Import loadings** and open loadings.mat and continue as described above. However, the *p(corr) cut-off* does not work exactly the same because Matlab does not seem to scale the loadings to be between 0 and 1.