GNAT manual

Version

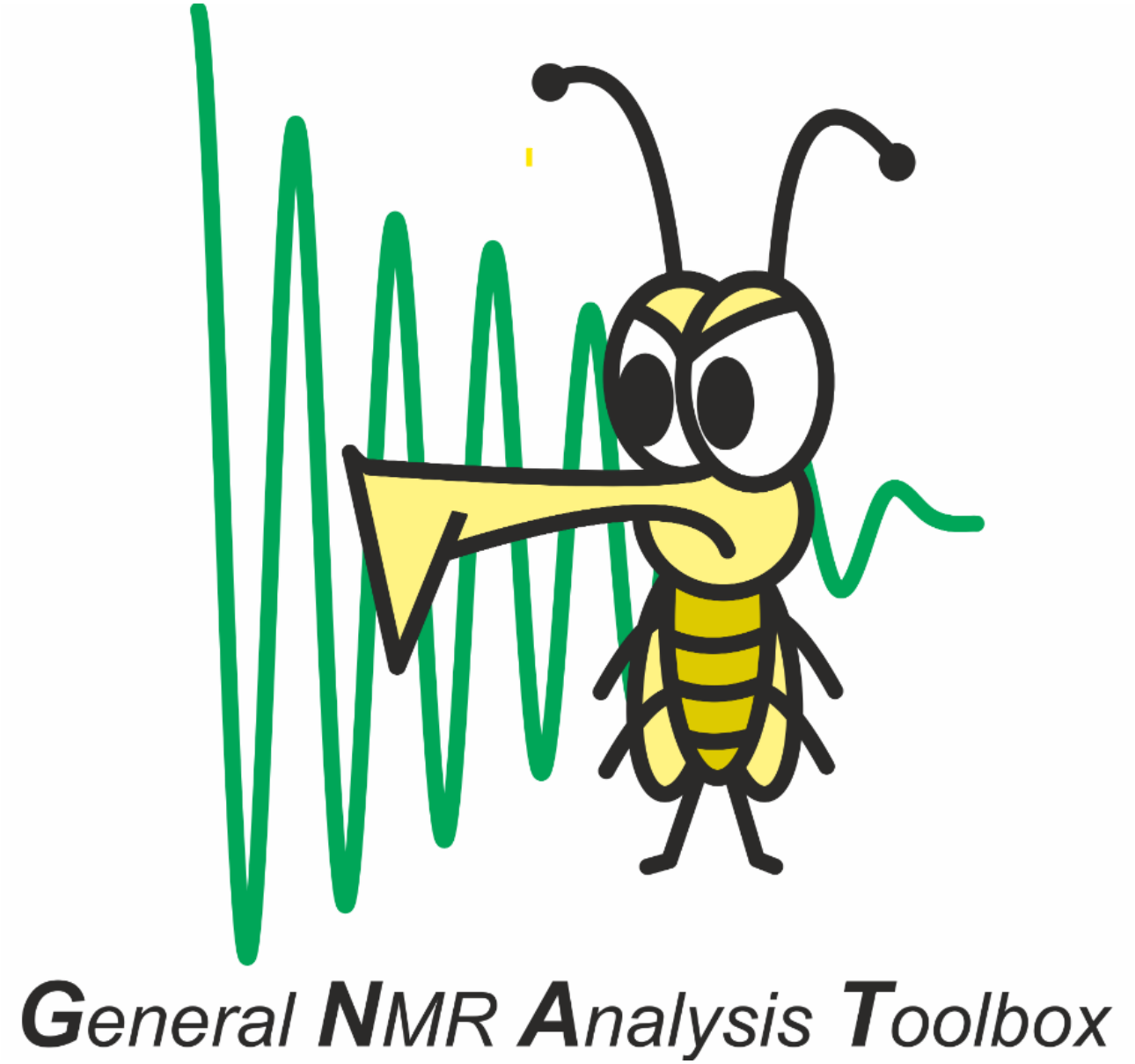


Table of Contents

GNAT manual documentation

Introduction

The GNAT (General NMR Analysis Toolbox) is a free and open-source software package for processing, visualising, and analysing NMR data. It supersedes the popular DOSY Toolbox, which has a narrower focus on diffusion NMR. Data import of most common formats from the major NMR platforms is supported, as well as a GNAT generic format. Key basic processing of NMR data (e.g., Fourier transformation, baseline correction, and phasing) is catered for within the program, as well as more advanced techniques (e.g., reference deconvolution and pure shift FID reconstruction). Analysis tools include DOSY and SCORE for diffusion data, ROSY T1/T2 estimation for relaxation data, and PARAFAC for multilinear analysis. The GNAT is written for the MATLAB® language and comes with a user-friendly graphical user interface. The standard version is intended to run with a MATLAB installation, but completely free-standing compiled versions for Windows, Mac, and Linux are also freely available.

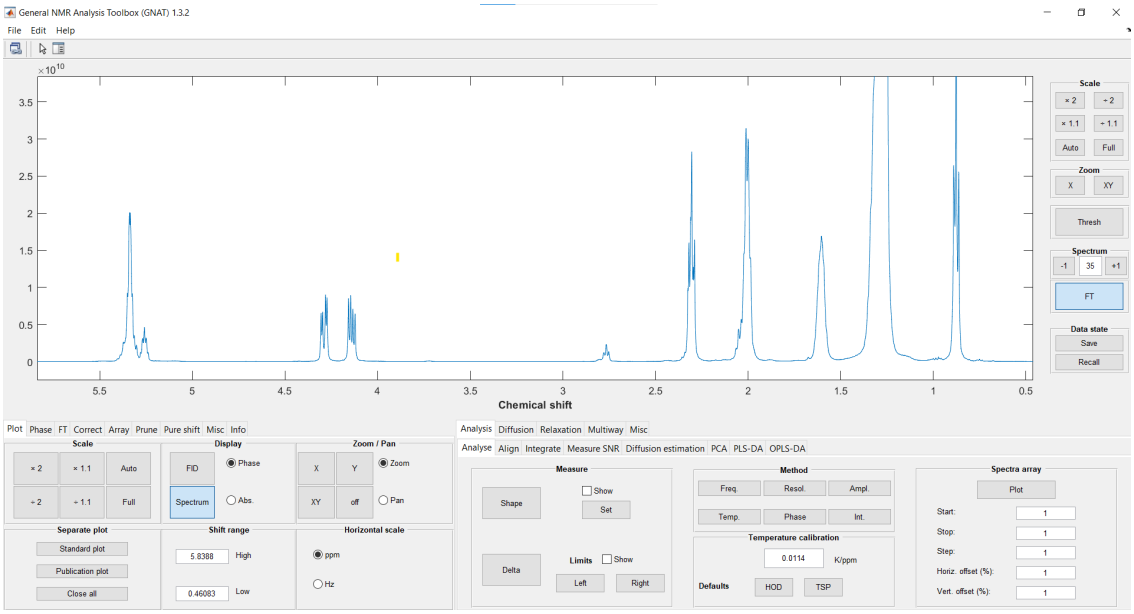
Citations

If you are using GNAT (or the older DOSY Toolbox) please site the following papers:

1. Castanar, L.; Dal Poggetto, G.; Colbourne, A. A.; Morris, G. A.; Nilsson, M. The GNAT: A new tool for processing NMR data. Magnetic Resonance in Chemistry 2018, 56 (6), 546.

2. Nilsson, M. The DOSY Toolbox: A new tool for processing PFG NMR diffusion data. Journal of Magnetic Resonance 2009, 200 (2), 296.

Structure



GNAT symbol.

Functions

File

In the File menu the user can Import and Export files of different format as well as Save and Open files in GNAT format. .. figure:: ./File/Fig1_file_menu.png

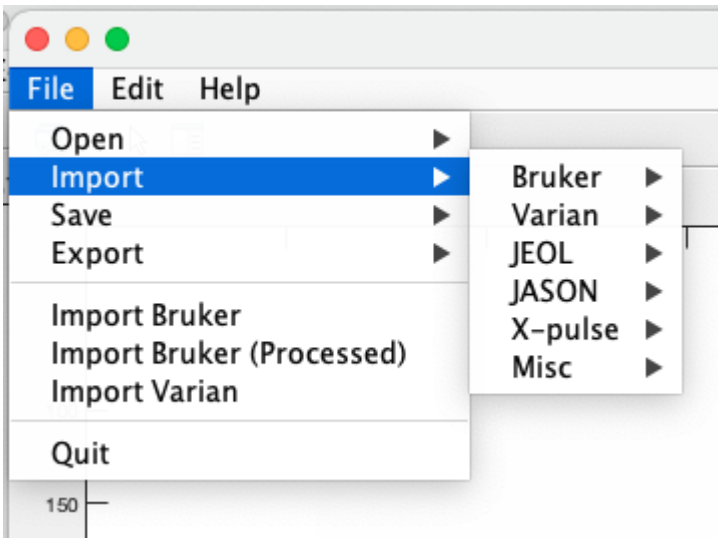
Functionalities

Open

Data in Matlab or GNAT format can be opened here. A description of these formats can be found in the Save section.

Import

Here the user can import data from various external formats. Feedback about the import is given in the Matlab window (or console window for compiled versions)



Bruker

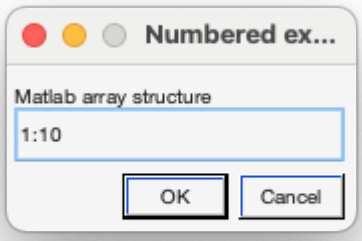
Data from Bruker can be imported in various ways. GNAT will do its best to determine which type of data it is (e.g. diffusion or relaxation encoded) and import the relevant parameters.

Bruker (standard)

Here raw FID data will be imported from a *fid* or *ser* file. For a ser file the data will be imported as an array.

Bruker array

Here a series of 1D spectra from different experiments will be imported. The experiment needs to be in a single folder where each experiment is given a number, which is very common for Bruker data. The order or import is given by providing the import dialog with a Matlab array (see Matlab online documentation for details).



By specifying 1:10 in the dialog the experiments below will be imported (in that order)

1 2 3 4 5 6 7 8 9 10

Here are some other examples

10:5:30 gives 10 15 20 25 30

10,4,5,8 gives 10 4 5 8

10:5:20,32,21 gives 10 15 20 32 21

Bruker 2D array

This option is to import an array of *ser* files. Each *ser* file will be imported as an array (as a standard Bruker import) and the different experiments will form a second array. This is for example of interest f monitoring a chemical reaction with DOSY experiments and using the 3D structure analyse with PARAFAC. The same is true for SCALPEL experiments.

The import procedure is the same as for Bruker Array.

References

1. Khajeh, M.; Botana, A.; Bernstein, M. A.; Nilsson, M.; Morris, G. A. Reaction Kinetics Studied Using Diffusion-Ordered Spectroscopy and Multiway Chemometrics. Analytical Chemistry 2010, 82 (5), 2102.

2. Nilsson, M.; Khajeh, M.; Botana, A.; Bernstein, M. A.; Morris, G. A. Diffusion NMR and trilinear analysis in the study of reaction kinetics. Chem Commun (Camb) 2009, (10), 1252.

3. Dal Poggetto, G.; Castanar, L.; Adams, R. W.; Morris, G. A.; Nilsson, M. Dissect and Divide: Putting NMR Spectra of Mixtures under the Knife. Journal of the American Chemical Society 2019, 141 (14), 5766.

Bruker (Processed)

Here the processed data are the Bruker processed ones. These are resident in the *pdata/* directory. This will allow data that has already been phases, baseline corrected etc to be imported to GNAT. The complex spectrum will be inversely Fourier transformed to a FID. These data can then be further processed in GNAT, just as if it were raw experimental data.

Bruker array (Processed)

This is just like “Bruker array”, but with processed data. However, only processed data in *pdata/1* is used.

Bruker 2D array (Processed)

This is just like “Bruker 2D array”, but with processed data. However, only processed data in *pdata/1* is used.

Bruker pure acquisition order

Imports raw FID data in the order it was acquired, so a 3D experiment is imported a sigle array. This can be useful for looking at increments in arrayed or nD data.

Varian

Varian/Agilent data import is supported here.

Varian

Imports standard Varian data. This is mainly for diffusion (DOSY) and relaxation data, but the standard Varian array structure is also supported.

Varian array

A series of 1D spectra can be imported similar to Bruker array. The experiments need to have numbers as names.

JEOL

JEOL data import is supported here.

JEOL generic

Data import of JEOL generic format is supported by GNAT. A help file for converting to JEOL generic can be found in the *Documentation* folder in the Matlab version of GNAT, or downloaded here:

JEOL export

JASON

TBA

X-pulse

TBA

Misc

Here GNAT supports some miscellaneous data formats.

Matlab structure

Here you can import any data has the format of a Matlab structure saved as a / * .mat/ file. The structure needs contain the fields: fid, sw, sfrq, ppmAxis and nucleus, with the following structure

fid: matlab array with dimensions [np dim2 dim3 dim4] (dim3 and dim4 can be left out). e.g a 1D spectrum will have the dimensions [np 1] and an array of 12 spectra (say a DOSY data set) will have [np 12], while for sets of DOSY data would have [np 12 4]

sw: spectral width in Hz

sfrq: spectrometer frequency in Hz

ppmAxis: chemical shift axis in ppm

nucleus: used isotope e..g 1H or 31P

Save

Data can be saved in different formats as below

Matlab format

The GNAT data are saved as a Matlab data structure (which is the internal format GNAT uses) as *.m

The data can be opened again in GNAT via the Open menu, or loaded directly into Matlab with the load command , which will give the user the NmrData structure. This structure contains the data that GNAT is using - e.g. NmrData.FID contains the original FID.

GNAT format

The data are saved in a GNAT specific format either as raw FID data, inverse Fourier transform of the complex spectrum, or as inverse Fourier transform of the real spectrum. In all cases the data will be saved as a FID, but for the two latter any processing, such as apodisation and baseline correction will be included.

Note: for all original data points to be used for the inverse Fourier transform of the real spectrum at least one zerofilling is needed. Only the raw FID option is automatically completely faithful to the original data. The GNAT data

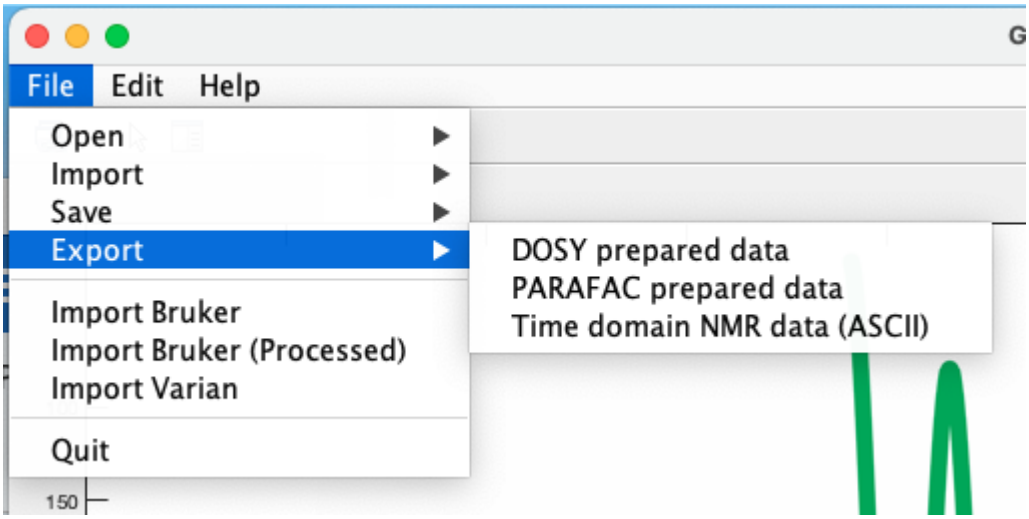
Data can be saved either as ASCII or binary. In both cases there will be a text header-file that describes all the relevant parameters. The ACSII version naturally takes up more disk space and is slower to save and load.

GNAT file format

The file format description can be found in the *Documentation* folder in the Matlab version of GNAT, or downloaded here: GNAT file format

Export

Here data can be exported in various formats



DOSY prepared data

Export of Matlab data prepared for DOSY processing by the m-file dosy_mn.m. This can be useful to do command line processing of DOSY data.

PARAFAC prepared data

Export of Matlab data prepared for PARAFAC processing by the N-way toolbox.

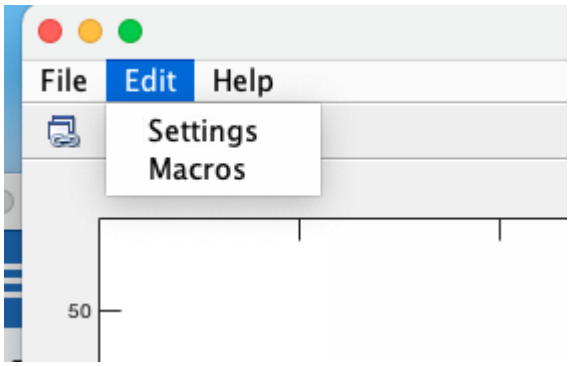
Reference (1) Andersson, C. A.; Bro, R. The N-way Toolbox for MATLAB. Chemometrics and Intelligent Laboratory Systems 2000, 52 (1), 1.

Time domain data

The raw FID data is exported in a ASCII format (related to GNAT format decribed in the Save section)

Edit

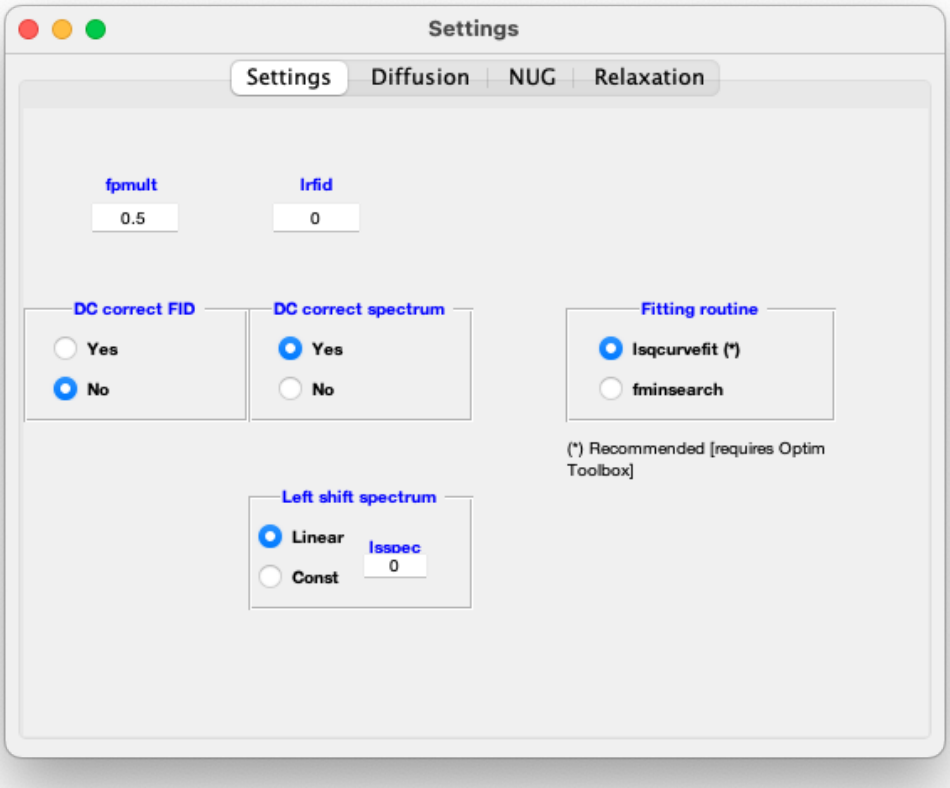
In the Edit menu the user gets access to different setting as well as macros



Functionalities

Settings

Choosing *Settings* from the *Edit* menu open up a GUI :



Settings

In this tab various choices for data handling are made

fpmult

Multiplication factor for the first point in the FID. (0.5 is default)

lrfid

Number of data points to left rotate the FID (o is default)

DC correct fid

Correction for a constant offset of the FID, buy subtracting the average of the last 5% of the FID. (Default is no)

DC correct spectrum

Correction for a constant offset of the spectrum,buy subtracting the average of the edge 5% of the spectrum. (Default is yes)

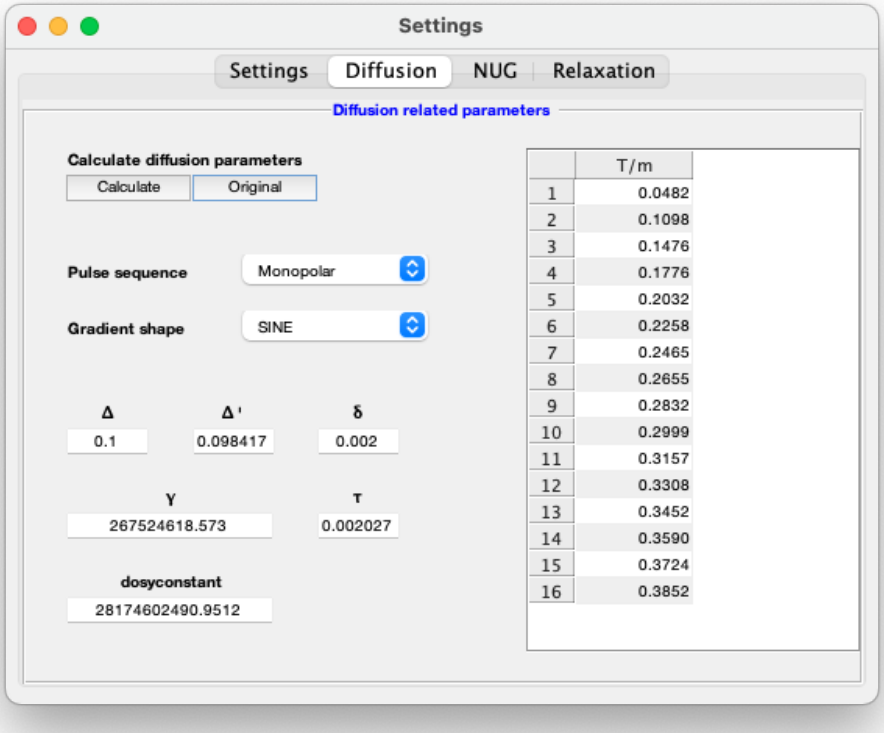
Fitting routine

Choosing *fminsearch* allows some fitting to be done if the Optimization Toolbox is not on the path. The default *lsqcurvefit* is more efficient.

Left shift spectrum* Shift (rotate) the spectrum by a certain amount of data points (deafult 0). If the data is arrayed each spectrum will be shifted by the same amount of data points if *Const* is selected and by linearly increasing amounts for each array element if *Linear* is chosen.

Diffusion

In this tab various parameters for diffusion NMR is accessible:



Diffusion parameters are normally imported directly in GNAT, so in most cases there is no need to make any changes of these. However, when the import has not been successful this can be amended here. The most critical parameter is *dosyconstant* which is calculated as:

\({\delta}^2*{\gamma}^2*{\Delta}^\prime\)

where \({\Delta}^\prime\) is the diffusion time \({\Delta}\) corrected for diffusion during the gradient pulses and is pulse sequence specific, and \({\delta}\) is the diffusion encoding time, \({\gamma}\) is the gyromagnetic ratio. The parameter \({\tau}\) is the difference between gradient pulses in a bipolar pulse pair (bpp) and is used in the calculation of \({\Delta}^\prime\)

A good description of this can be found in this paper:

1. Sinnaeve, D. The Stejskal-Tanner equation generalized for any gradient shape-an overview of most pulse sequences measuring free diffusion. Concepts in Magnetic Resonance Part A 2012, 40A (2), 39.

Calculate diffusion parameters

Pressing the *Calculate* button will calculate *dosyconstant* and \({\Delta}^\prime\) with the given parameters.

Pressing the *Original* button will revert to the originally imported values

Pulse sequence

Choose the type of diffusion pulse sequence used in the drop down list

Gradient shape

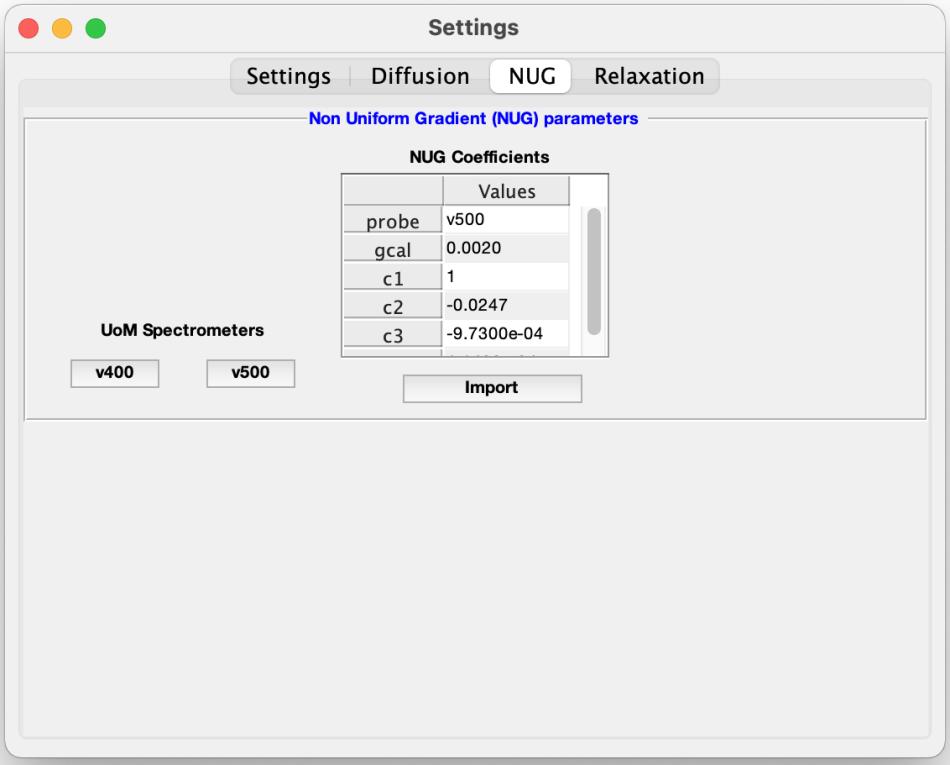
Choose the gradient shape used

Gradient amplitudes

The table to the right shows the gradient amplitudes used (in Tesla per meter)

NUG

In this tab various parameters related to the (non) uniformity of the diffusion encoding gradients (NUG) is accessible:



The NUG coefficients are used to characterize the decay signal in a diffusion NMR experiment, and is specific to each probe. This can be very important for obtaining accurate diffusion coefficients, and is described in detail in this paper:

1. Connell, M. A.; Bowyer, P. J.; Bone, P. A.; Davis, A. L.; Swanson, A. G.; Nilsson, M.; Morris, G. A. Improving the accuracy of pulsed field gradient NMR diffusion experiments: Correction for gradient non-uniformity. Journal of Magnetic Resonance 2009, 198 (1), 121.

with some applications shown here:

1. Nilsson, M.; Connell, M. A.; Davis, A. L.; Morris, G. A. Biexponential fitting of diffusion-ordered NMR data: Practicalities and limitations. Analytical Chemistry 2006, 78 (9), 3040.

2. Nilsson, M.; Morris, G. A. Correction of systematic errors in CORE processing of DOSY data. Magnetic Resonance in Chemistry 2006, 44 (7), 655.

3. Nilsson, M.; Morris, G. A. Improved DECRA processing of DOSY data: correcting for non-uniform field gradients. Magnetic Resonance in Chemistry 2007, 45 (8), 656.

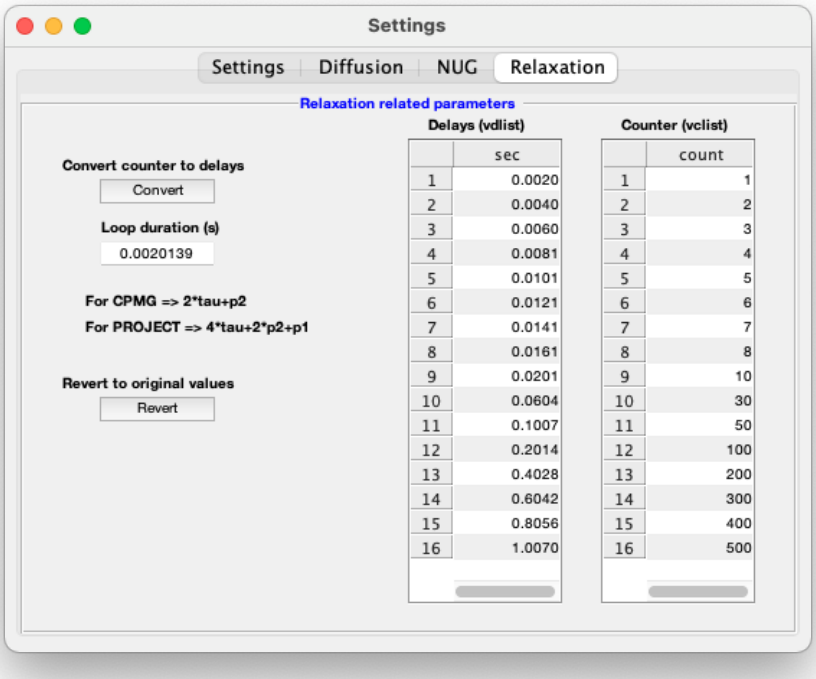
The choice of using NUG coefficients is given in the Diffusion processing module of GNAT.

The table shows the relevant values, which can be typed in by hand or imported from a text file.

There are also a couple of shortcut buttons for some spectrometers in the Manchester laboratory.

Relaxation

In this tab various parameters for relaxation NMR is accessible:



GNAT will attempt to import the correct delays for a relaxation (e.g. T1 or T2) experiment, and provides this interface to correct any mistakes.

If there is direct delay information e.g. in a *vdlist* (variable delay list) this will be used for the *tau* delays. If there is information about counters e.g. in a *vclist* this will be used to produce a vdlist using default parameters by multiplying the number of counters with the associated loop counter time. The recipe for automatically calculating the counters for CPMG and PROJECT is given in the GUI window.

The user can put in an arbitrary loop counter time and recalculate the vdlist pu pressing the *Convert* button; the *Revert* button will restore the original parameters.

Macros

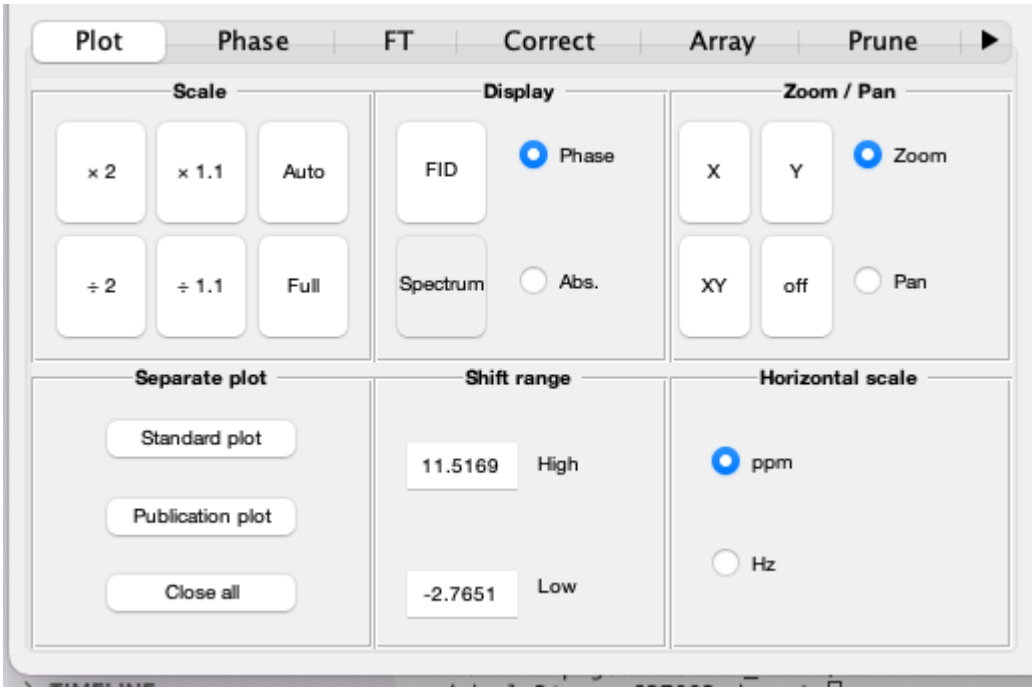
Help

Functionalities

About

Plot

This is the tab for general plot control.



Scale

Controls to scale the spectrum in the plot spectrum/FID. You can multiply or divide by a factor 2 or 1.1. The **Auto** button autsoscales the spectrum and the **Full** button plots the full spectrum.

Display

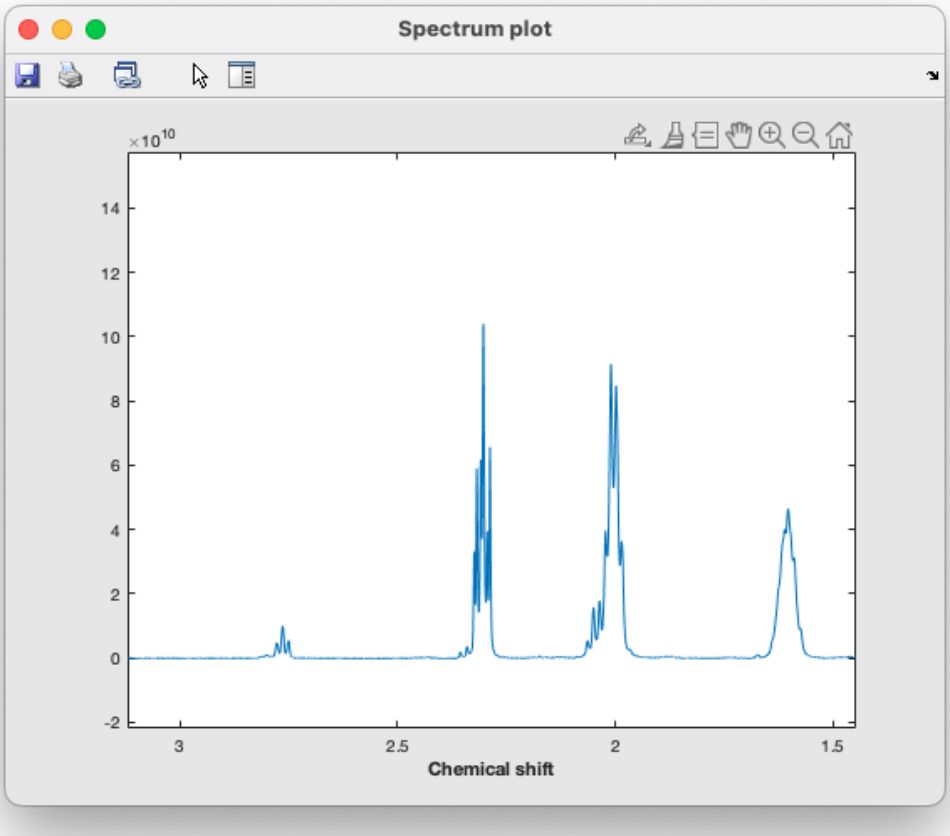
Controls for how the display the data. The user can display the spectrum or FID in either phase sensitive (**Phase**) or absolute value (**Abs**) mode.

Zoom/Pan

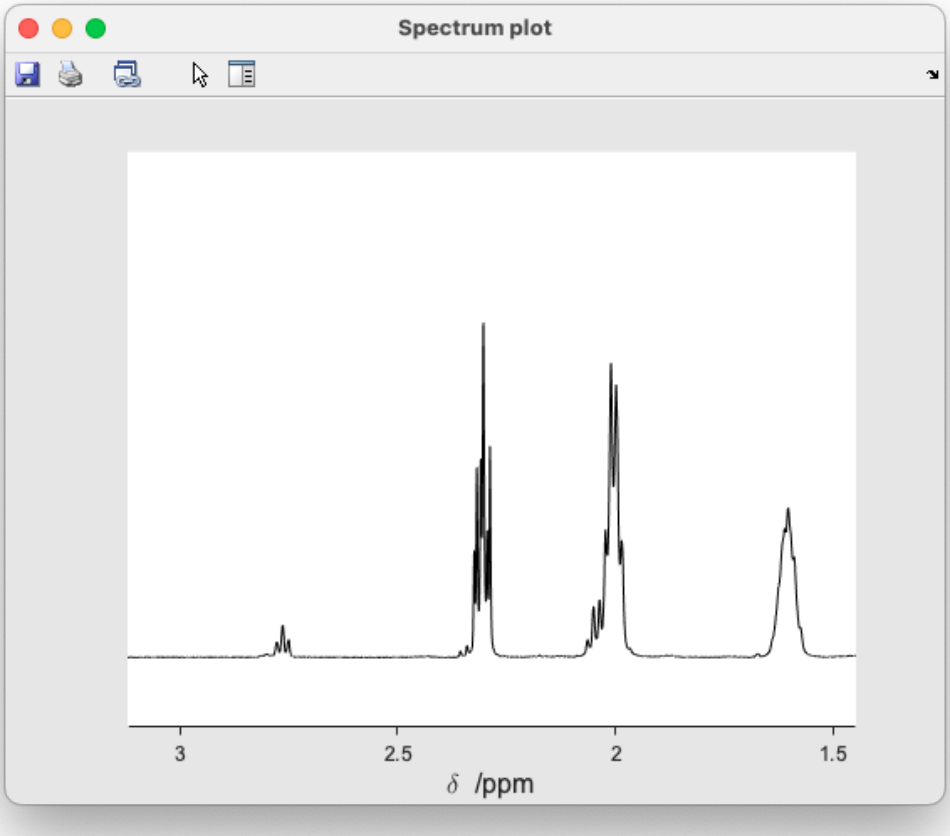
Controls for zooming or panning the display.

Separate plot

The **Standard plot** button plots the spectrum as seen in the main window.



The **Publication plot** button plots the spectrum in a format more suitable for publication.



The plots can be saved in the available Matlab formats (e.g. fig, svg, eps, png, jpg, pdf)

The **Close all** button closes all Matlab windows expect the main window.

Warning

This can be handy when there are too many open windows but it really means all Matlab windows, so if you have another, non GNAT, Matlab window open, or another GNAT instance they will all be closed.

Separate plot

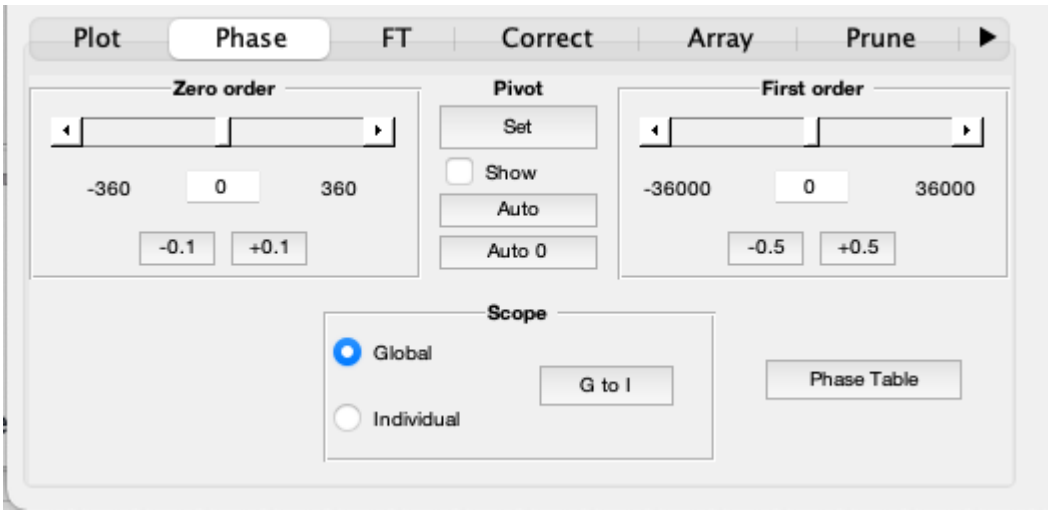
Here the user can set the range of chemical shifts displayed. These will also update if you e.g. zoom the spectrum.

Horizontal scale

Here the user can set the unit of the chemical shifts displayed to either **ppm** or **Hz** .

Phase

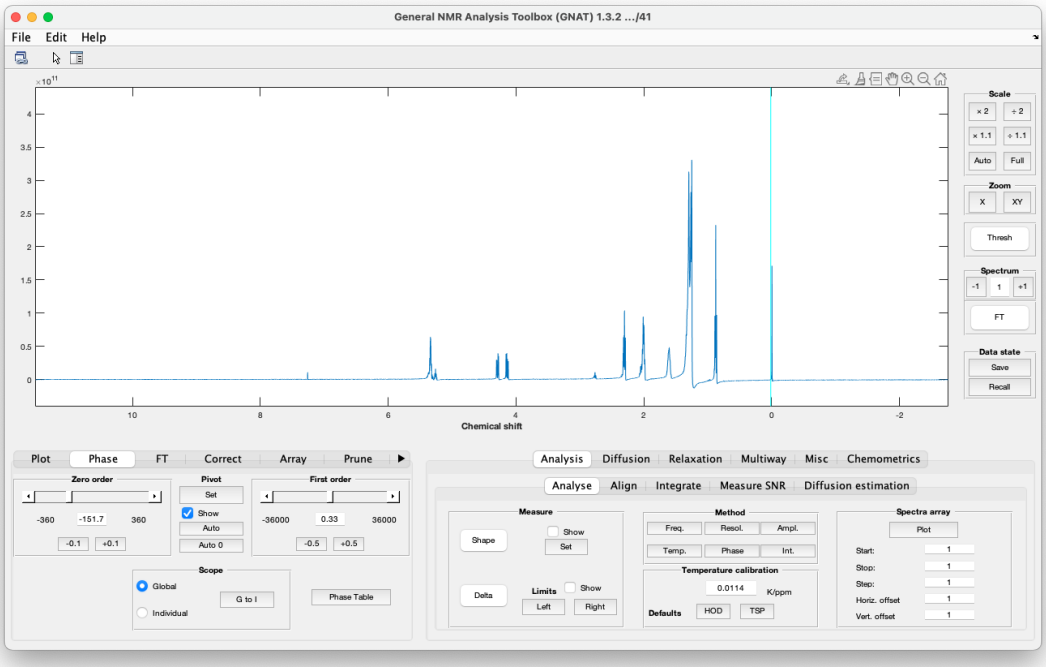
This is the tab for phasing the spectra



Phasing is done using a zero and first order phase correction, using the **Zero order** and **First order** controls.

Typical use is:

1. Set a pivot (light blue) using the **Pivot** controls.
2. Adjust the zero order phase so that it is correct at the pivot.
3. Adjust the first order so that the whole spectrum is correct. Changing the first order phase will keep the phase constant at the *pivot* .



The buttons *Auto* and *Auto 0* will attempt an automatic phase correction, where *Auto 0* is only optimising the zero order phase.

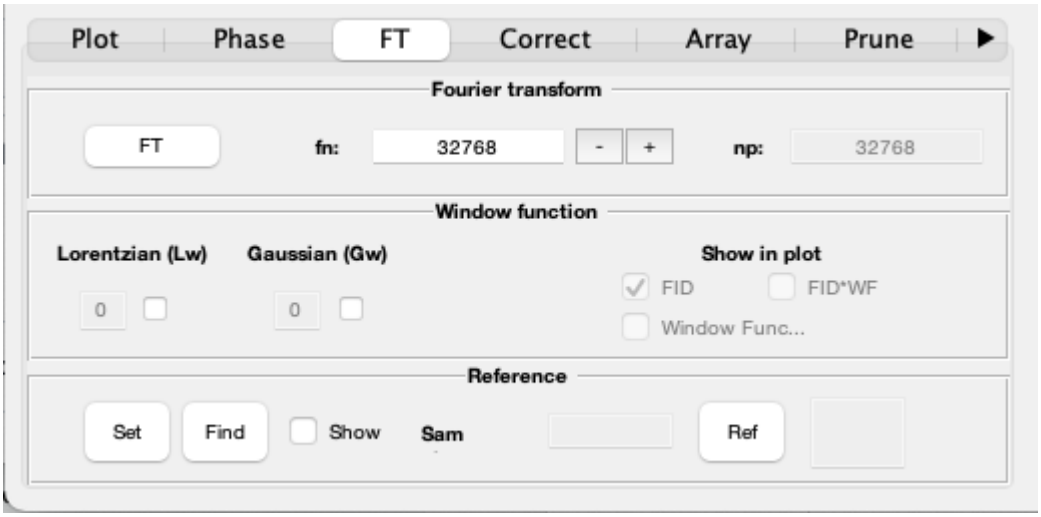
GNAT has the facility to do a separate phase correction for each spectrum in an array. This can e.g. be useful for DOSY data that sometimes show a gradient dependent phase. This is accessed with the **Scope** controls. *Global* uses the same phase for all the spectra in the array, while *Individual* uses separate correction for all array elements. The *G to I* button will copy the global phase parameters to all the Individual ones. It is often useful to first do a *Global* correction and then copy that t the *Individual* array elements before doing fine adjustments to some or all of the array elements.

Note

Switching between array elements can be done in the Array tab or in the **Spectrum** control in the shortcuts to the left of the main window.

FT

This is the tab for controlling the Fourier transform parameters, and to reference the spectra.



Fourier transform section

The actual Fourier transform is performed when the *FT* button is pressed. This button can also be found in the shortcuts to the left in the main window.

np stands for “”number of points” and corresponds to the number of complex data points in the FID.

fn stands for “Fourier number” and is the actual number of complex points that is used in the Fourier transform. This can be lower than **np** (and only the first **np** points of the FID is used), but more typically it is larger. When **fn** > **np** the FID data will be extended with zeros (zerofilling). The + and - buttons will change the **fn** by to te nearest power of two.

Window function section

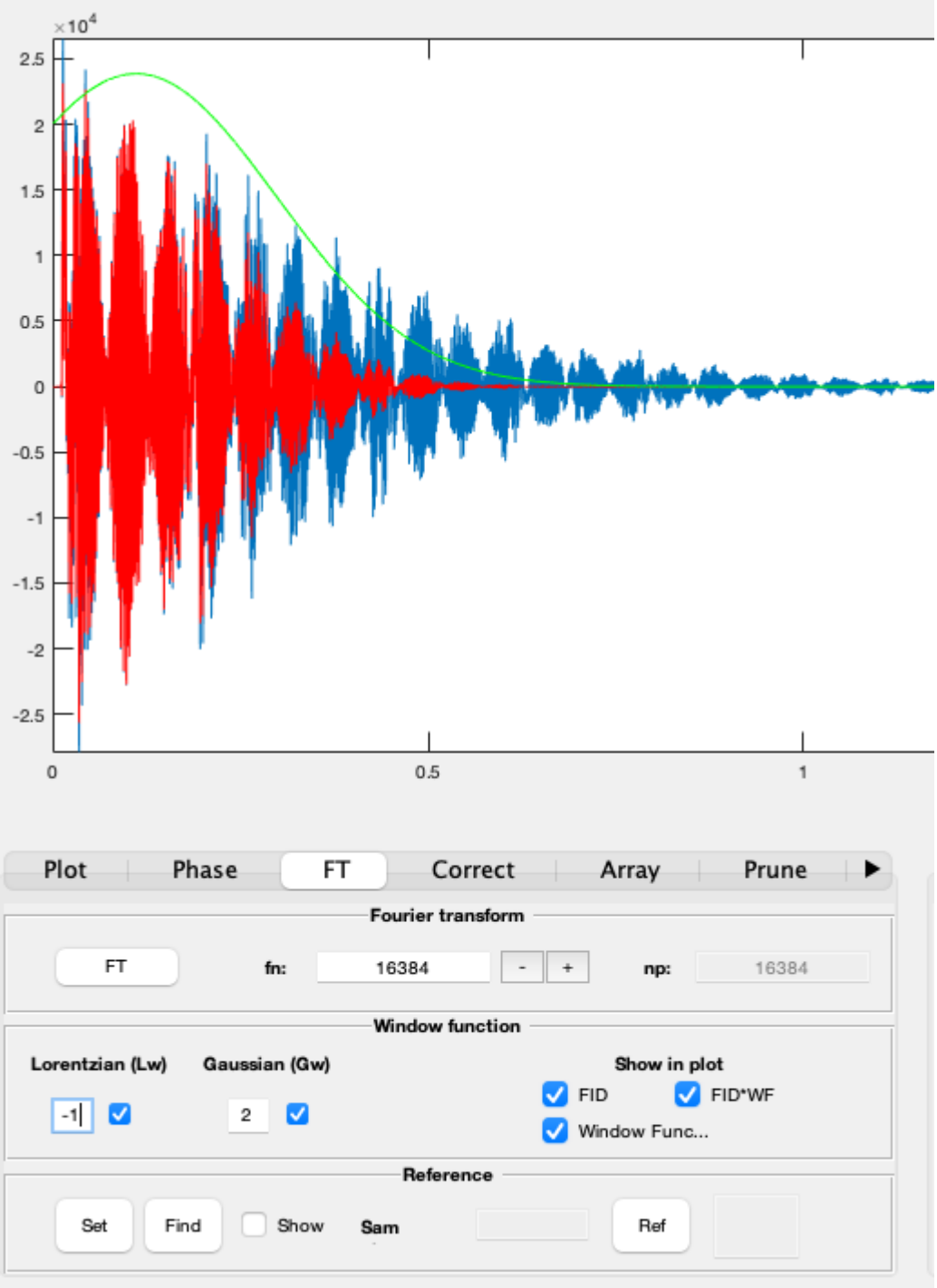
Here the user can apply window functions to the FID. The *Lorentzian* function will apply a line broadening of *Lw* Hz by multiplying the FID with a suitable exponential function. *Lw* can be positive or negative.

$e^{-\pi Lw t}$

The *Gaussian* function ill apply a line broadening of *Gw* Hz by multiplying the FID with a suitable exponential function. A as it is squared, negative *Gw* gives the same result as a positive.

$e^{-{\frac{(\pi Gw t)^2}{4 ln2}}}$

A graphical display of the window functions can be accessed by the **Show in plot** control. This requires that displaying the FID is chosen in the Plot tab. The window function is shown in green and the resulting FID in red.



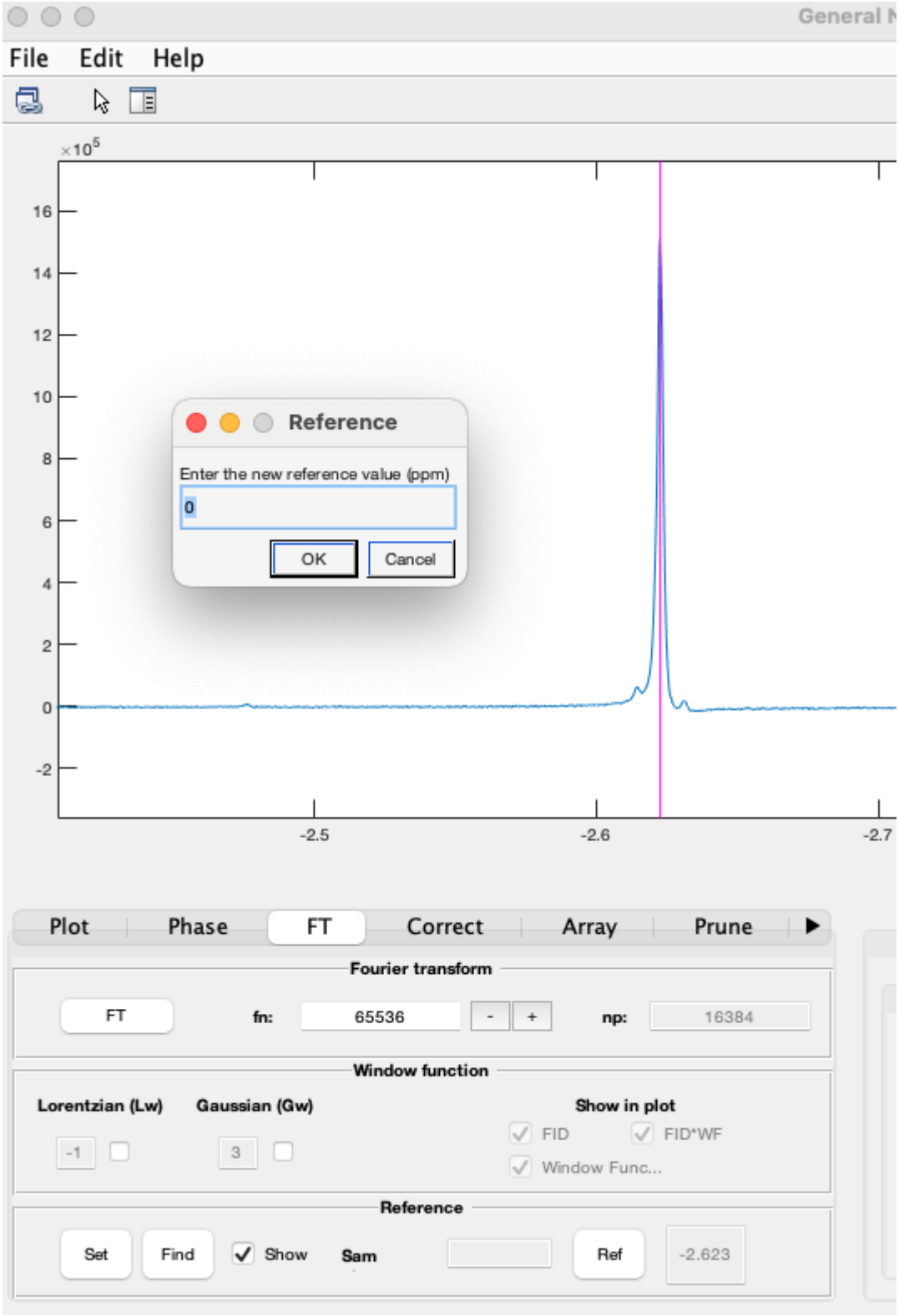
Reference section

Here the user can reference the spectra to get a correct frequency scale.

1. Set the reference line (purple) using the *Set* button.

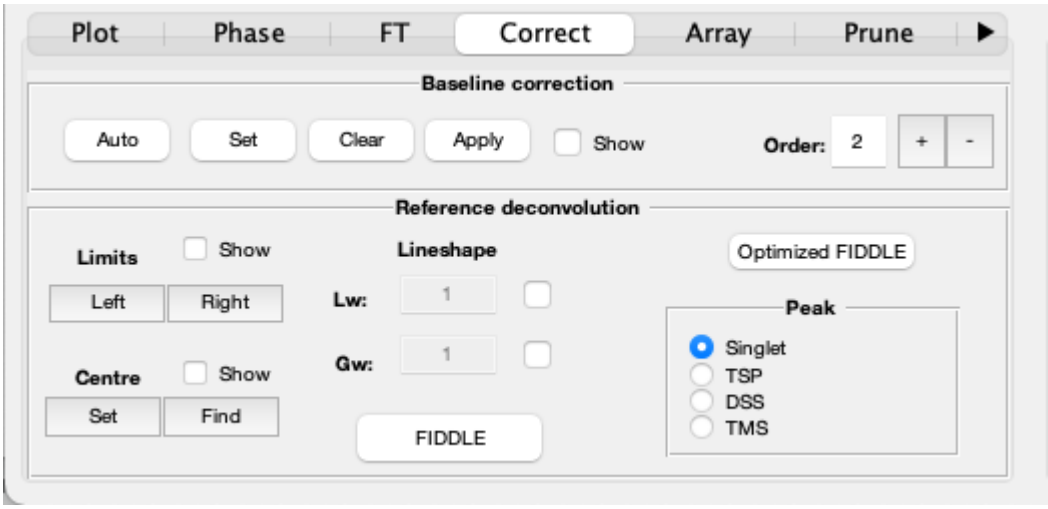
2. Press the *Find* button to find the peak maximum (if desired).

3. Press the *Ref* button and type in the correct value.



Correct

This is the tab for baseline correction and reference deconvolution control.

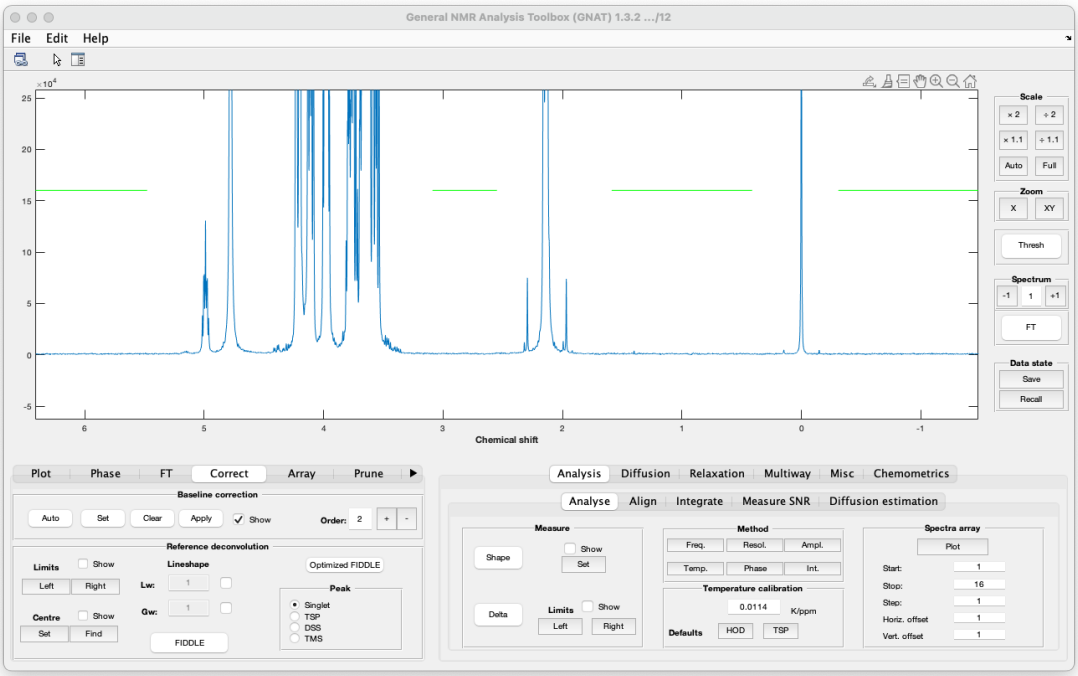


Baseline correction

Automatic baseline correction is done by pressing the *Auto* button. The algorithm is taken from:

1. Pearson, G. A. GENERAL BASELINE-RECOGNITION AND BASELINE-FLATTENING ALGORITHM. Journal of Magnetic Resonance 1977, 27 (2), 265.

Manual baseline correction is done by manually identifying regions of empty baseline and then fitting a polynomial which is subtracted from the whole spectrum. The order of the polynomial is under user control in the *Order* parameter. The baseline regions are set by pressing the *Set* button and clicking with the mouse in the spectrum window. *Clear* will clear current settings and *Apply* will apply the baseline correction.



Reference deconvolution

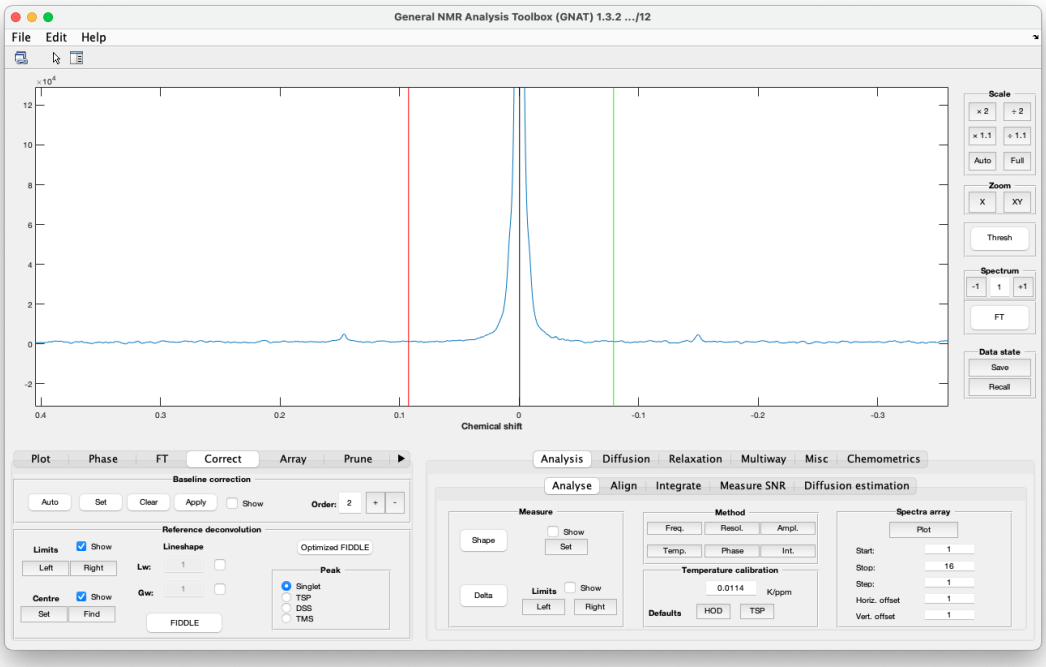
Reference deconvolution uses information from a reference peak to correct errors (e.g. shimming, phase, and frequency) in the whole spectrum. A thorough explanation can be found in the following papers (and references therein).

1. Morris, G. A.; Barjat, H.; Home, T. J. Reference deconvolution methods. Progress in Nuclear Magnetic Resonance Spectroscopy 1997, 31 (2-3), 197.

2. Ebrahimi, P.; Nilsson, M.; Morris, G. A.; Jensen, H. M.; Engelsen, S. B. Cleaning up NMR spectra with reference deconvolution for improving multivariate analysis of complex mixture spectra. Journal of Chemometrics 2014, 28 (8), 656.

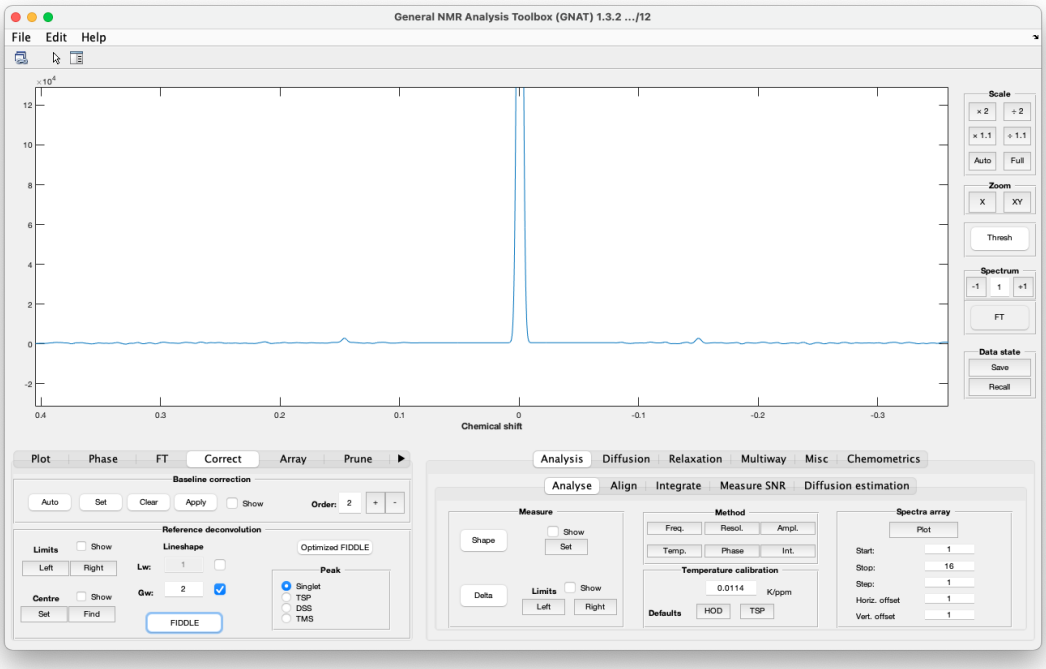
The reference peak is ideally a simple singled, but some common reference materials such as TSP, DSS, and TMS can also be used if their Si satellite signals are taken into account. The type of signal can be chosen in the *Peak* box.

The user needs to define the peak to be used. The centre of the peak is set using the *Set* and *Find* buttons, where the latter finds the frequence of the peak maximum. The edges of the peak is set by the *Left* and *Right* buttons and should include a small amount of base line on each side. The limits can be set either inside or outside of the 13C satellites, depending on how narrow the peak is. Here they are set inside the satellites.



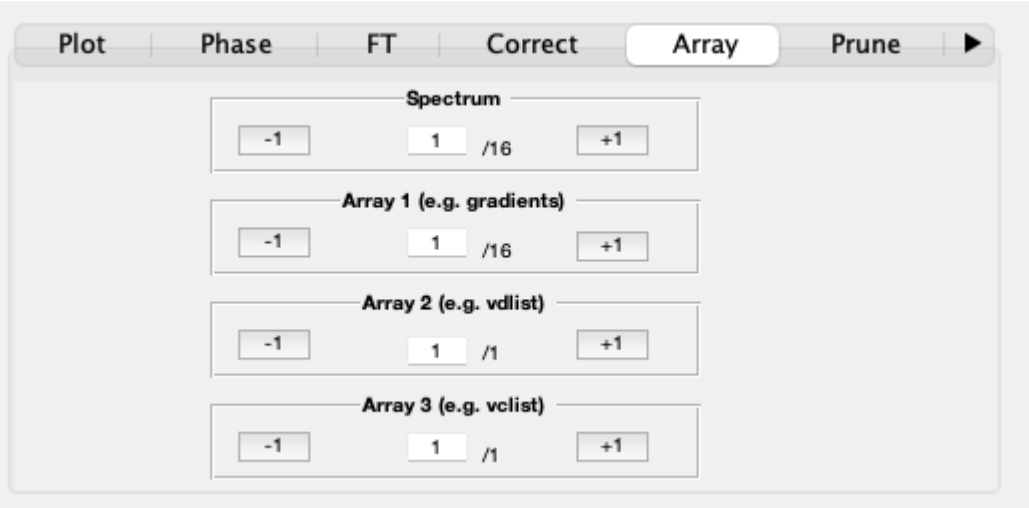
Refrence deconvolution is performed by pressing the *FIDDLE* button. The experimental line shape will be replaced by a “perfect” lineshape as decided by the *Lineshape* parameters. There are the

same as in the FT tab. Here the lineshape was replaced by a 2 Hz Gaussian shape. The lineshape can be a combination of Lorentzian and Gaussian.



Array

GNAT is made to deal with arrayed data. This is the tab for controlling the array element to display. The most common arrayed data sets are diffusion and relaxation measurements (the predecessor of GNAT, the DOSY Toolbox, was written for processing diffusion data). However many types of arrayed data can be invesigated; examples include reaction timr course, or just the different t1 increments in a classic 2D NMR experiments such as COSY.



GNAT can currently handle arrays that are arrayed in a maximum of 3 dimensions. An example of such data would be a combined diffusion, T2 relaxation, and TOCSY t1 SCALPEL experiment.

1. Dal Poggetto, G.; Castanar, L.; Adams, R. W.; Morris, G. A.; Nilsson, M. Dissect and Divide: Putting NMR Spectra of Mixtures under the Knife. Journal of the American Chemical Society 2019, 141 (14), 5766.

In the *Spectrum* box you can decide which spectrum to display from the total number of spectra in the arrayed data. This can also be accessed in the shortcuts in to the left in the main window.

In the *Array 1* box you can go though the spectra in the first array dimension. Here the array is a diffusion exprimnet so there is ony one array (or increasing gradient strength) and the *Spectrum* and *Array 1* boxes will do the same thing.

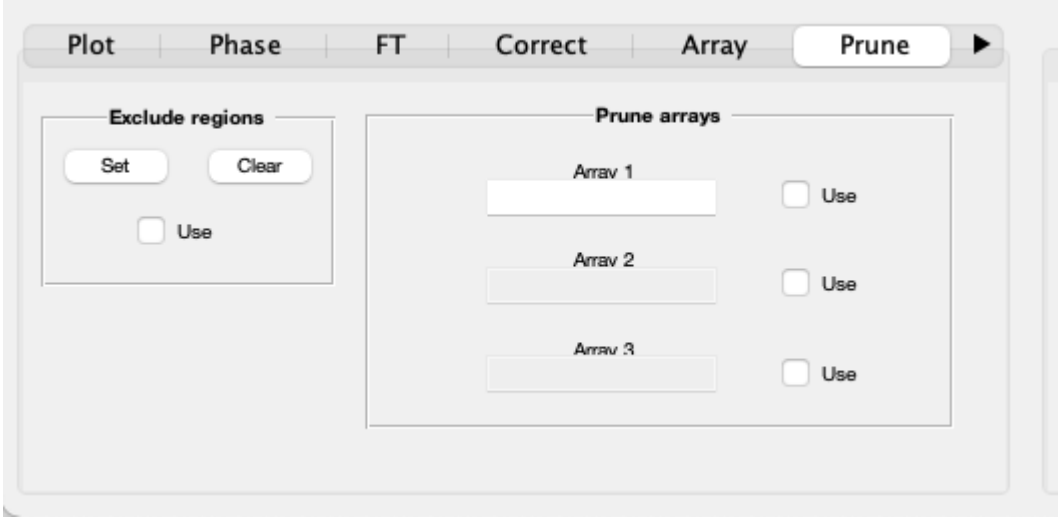
In the *Array 2* box you can go though the spectra in the second array dimension. If the first array is a diffusion experiment (as in our example) and the second is a time course during a chemical reaction (i.e. acquiring a diffusion experiment for each time point in a chemical reaction) then if Array 1 (gradient strength) is set at 2 changing the Array 2 display will change time points for the second gradient levels. If Array 2 is set to 8, then changing Array 1 will change gradient levels for the 8th time point.

1. Nilsson, M.; Khajeh, M.; Botana, A.; Bernstein, M. A.; Morris, G. A. Diffusion NMR and trilinear analysis in the study of reaction kinetics. Chem Commun (Camb) 2009, (10), 1252.

2. Khajeh, M.; Botana, A.; Bernstein, M. A.; Nilsson, M.; Morris, G. A. Reaction Kinetics Studied Using Diffusion-Ordered Spectroscopy and Multiway Chemometrics. Analytical Chemistry 2010, 82 (5), 2102.

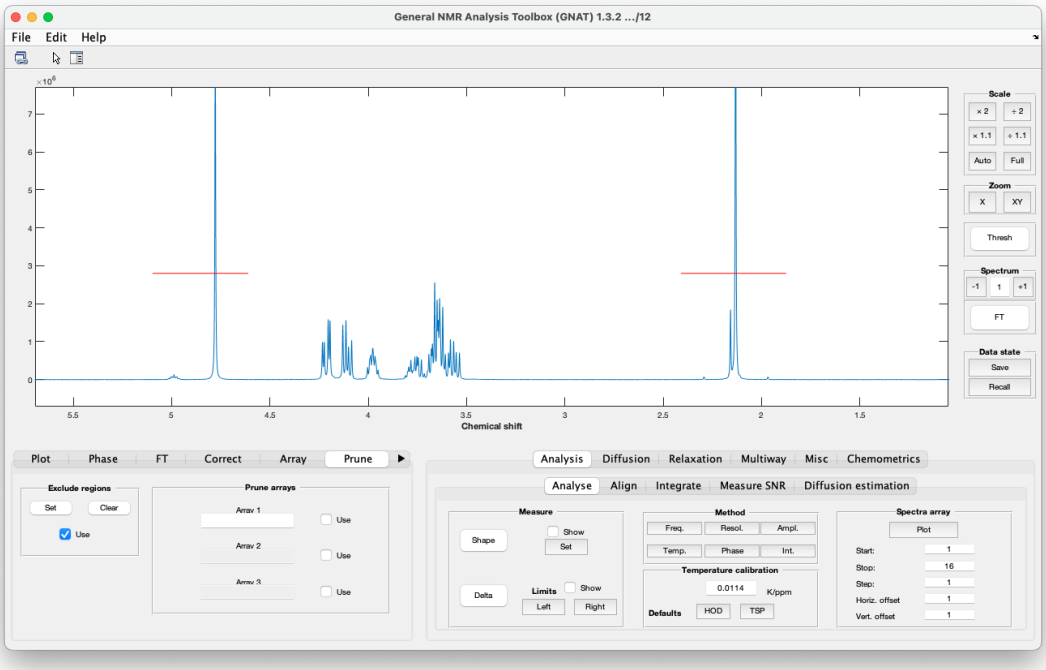
Prune

In this tab the user can access controls to exclude (prune) parts of the data for analysis. For example a certain region of the spectrum, like a solvent peak, may not be helpful to include in a DOSY spectrum.



Exclude regions

In a way similar to baseline correction (see Prune tab.) the user can select regions from the spectrum to be excluded. In the example here the regions around 2.1 and 4.9 ppm (red line) will be removed from analysis.



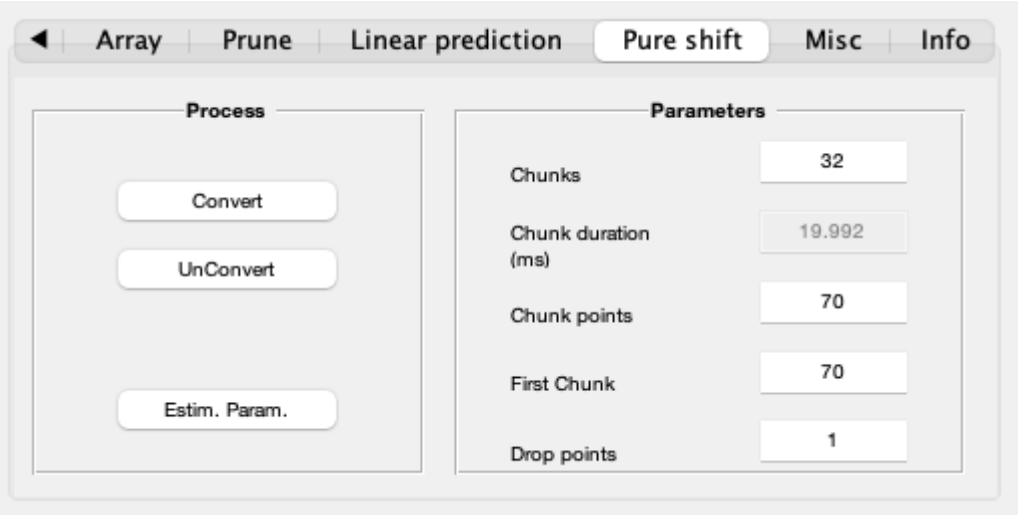
Prune arrays

Here one or several array elements can be removed. This could for example be useful if a certain gradient level in diffusion experiment, or delay time in a relaxation experiment is corrupted. The

array element to be excluded is determined by a Matlab array. Some examples of this is given for importing arrayed data Import

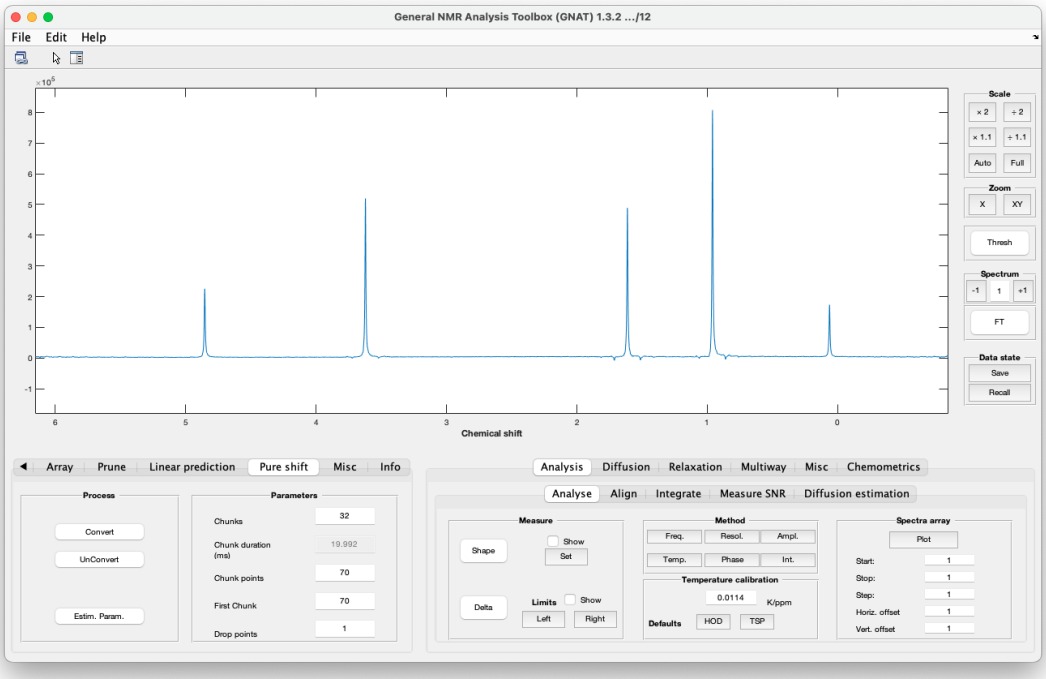
Pure Shift

Here the use can process interferogram style pure shift data. The raw data where each experiment contains a “chunk” of the FID is assembled to a single pure shift FID.

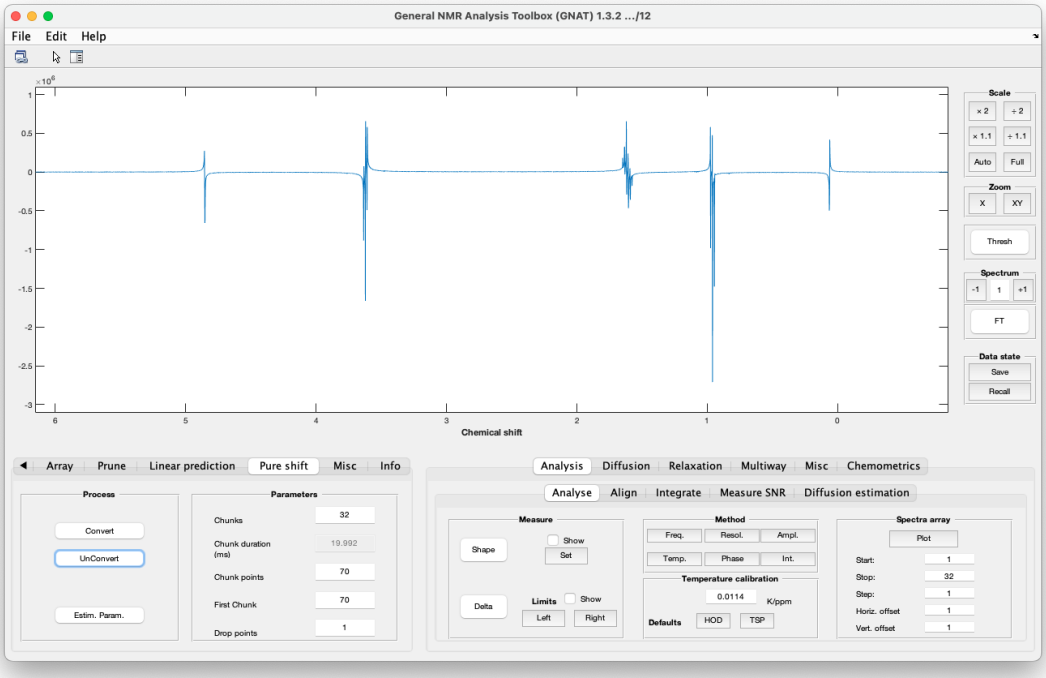


Process section

Pressing the *Convert* button will convert the raw data to pure shift data using the parameters in the **Parameters** section.



Pressing the *UnConvert* button will revert to raw data.



The *Estim. Paramaters* will try to guess the pure shift conversion parameters from the raw data set.

Parameters section

The *Chunks* parameter decides how many chunks the assembled pure shift FID consists of (and *Chunk duration* is the duration on each chunk in milliseconds)

The *Chunk points* parameter decides how many complex data points each chunk consists of. Sometimes this is different for the *First Chunk* .

Drop points is the number of comlex data points that is discarded in the beginning of each chunk.

More information

Pure shift NMR is a big topic and cannot be covered in this manual. The user is referred to some of the excellent reviews available (references below.)

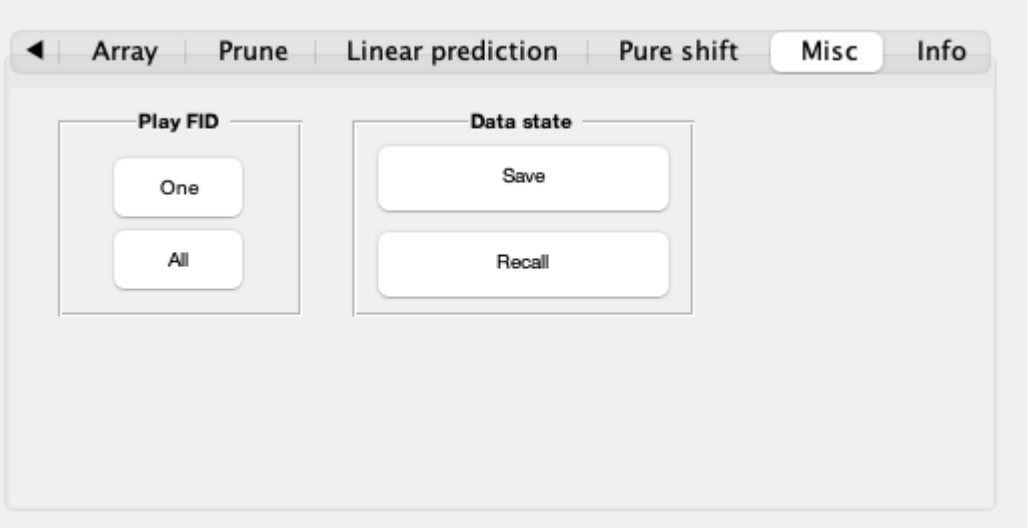
There is also a lot of information from a workshop held in Manchester: (Pure shift workshop)

1. Zangger, K. Pure shift NMR. Progress in Nuclear Magnetic Resonance Spectroscopy 2015, 86-87, 1.
2. Adams, R. W. In eMagRes; John Wiley & Sons, Ltd, 2014. https://doi.org/10.1002/9780470034590.emrstm1362
3. Foroozandeh, M.; Morris, G. A.; Nilsson, M. PSYCHE Pure Shift NMR Spectroscopy. Chemistry-a European Journal 2018, 24 (53), 13988.
4. Castañar, L. Pure shift 1H NMR: what is next? Magnetic Resonance in Chemistry 2017, 55 (1), 47.

Info

Misc

Here the user can find some miscellaneous functionality that did not fit in well anywhere else.



Play FID section

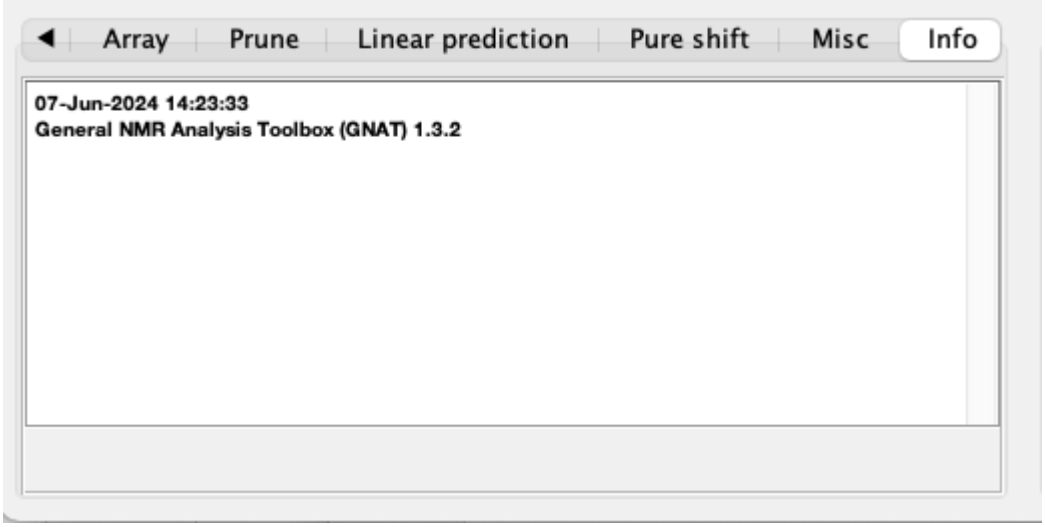
Here the user can listen to their data. The *One* button will play the current FID while the *All* button will play all FID in arrayed data (can take a lot of time for a large array)

Data state section

Here the user can save the data in a particular state, with the *Save* button. For examples with some specific processing paramaters, and then recall that state with the *Recall* button.

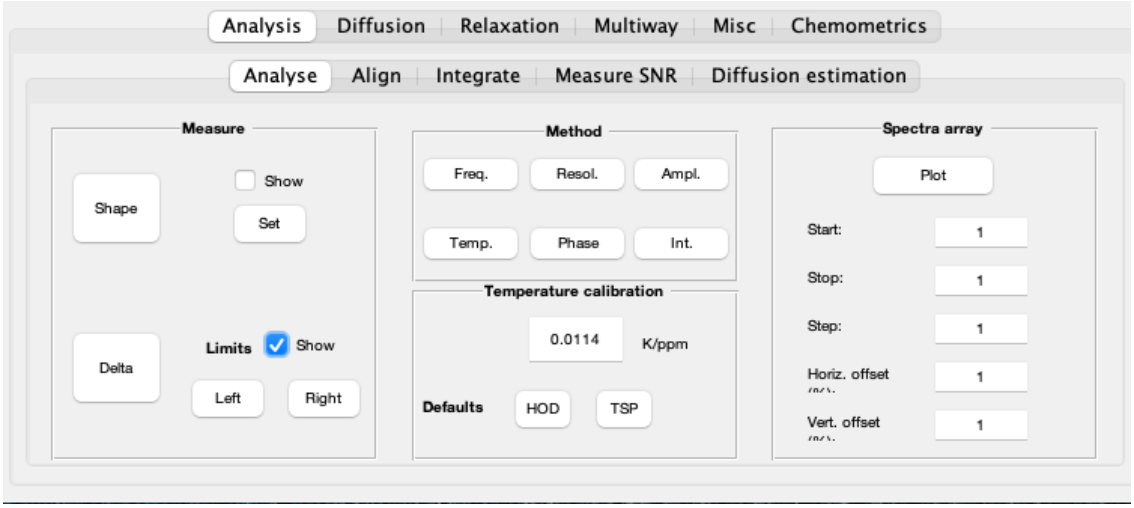
Info

Here information about e.g. teh import and processing in GNAT is displayed. This was more heavily used in early versions of GNAT, but as it tended to slow things down this is now mostly done directly to the Matlab window, or the Terminal window for compiled versions.



Analysis

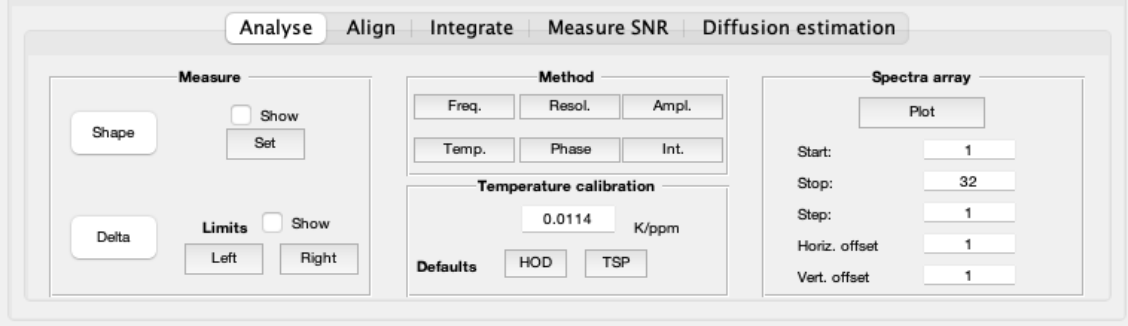
This is the head tab for various way to analyse your NMR data, with particular emphasis on arrayed data.



Functionalities

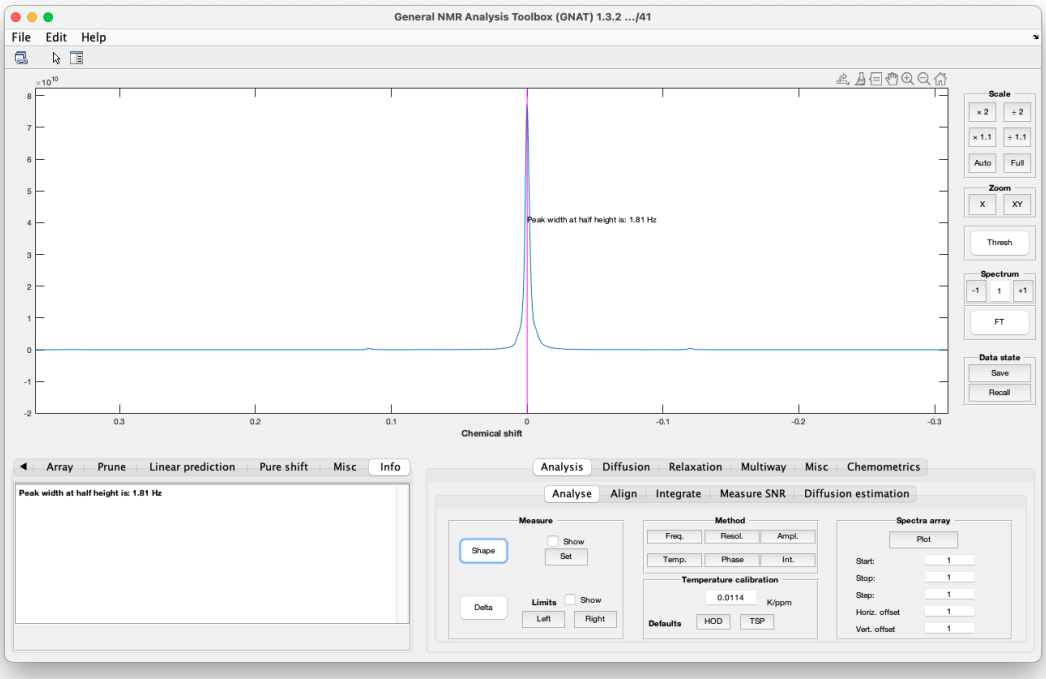
Analyse

Here the user can find various way to analyse their NMR data, with particular emphasis on arrayed data.

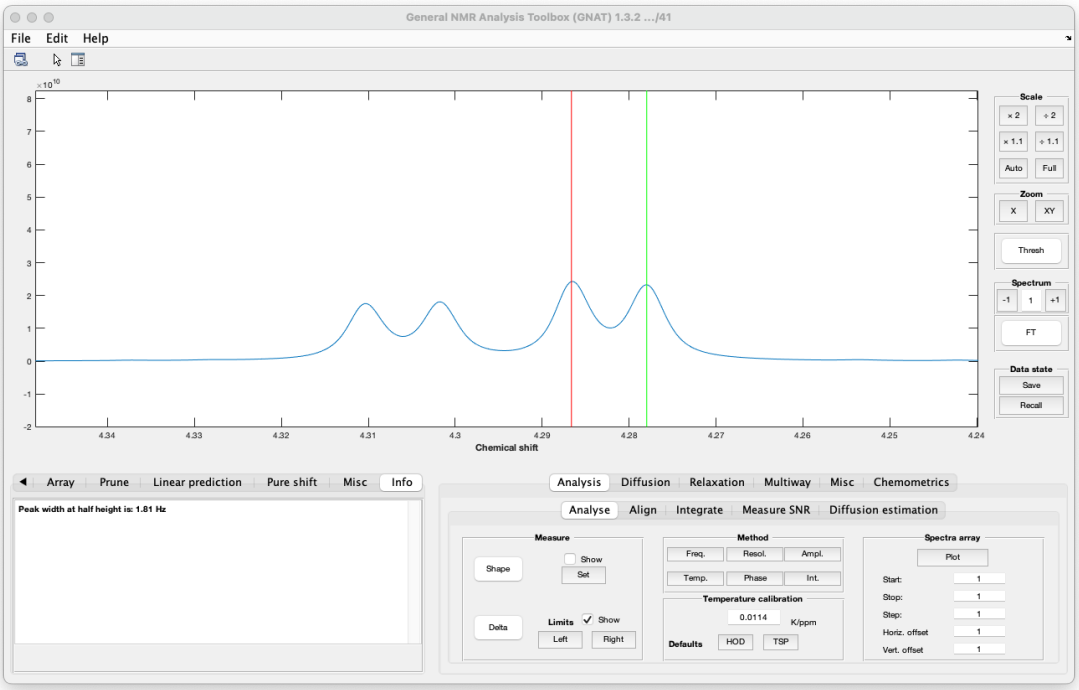


Measure section

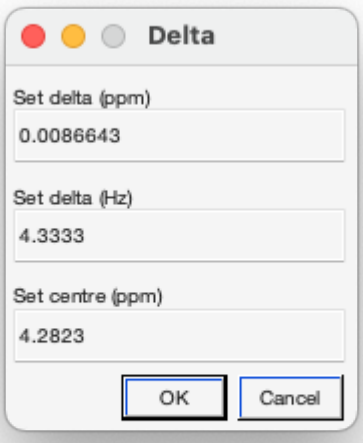
The *Shape* button will display the peak width at half height for the selected peaks - selected with the *Set* button



The *Delta* button will display the peak width at half height for the selected peaks - selected with the *Left* and *Right* buttons.



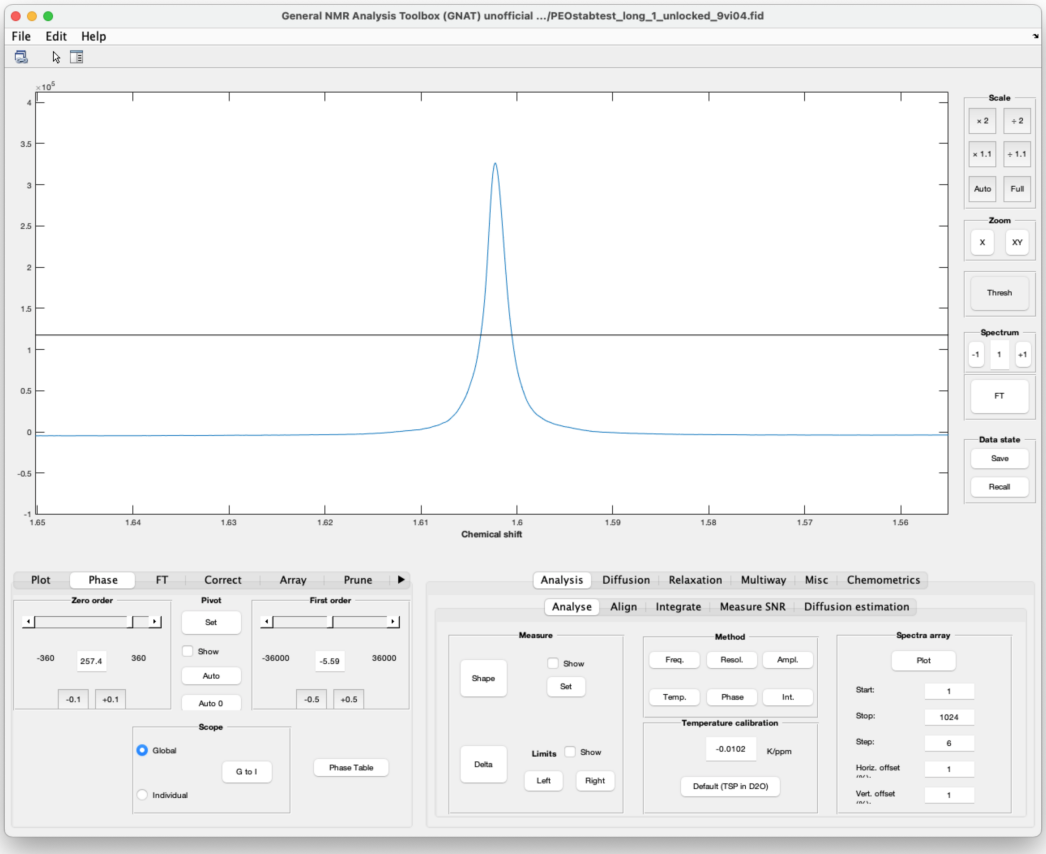
It will open a dialog where the used can set the difference in frequency as well as the centre between Left and Right.



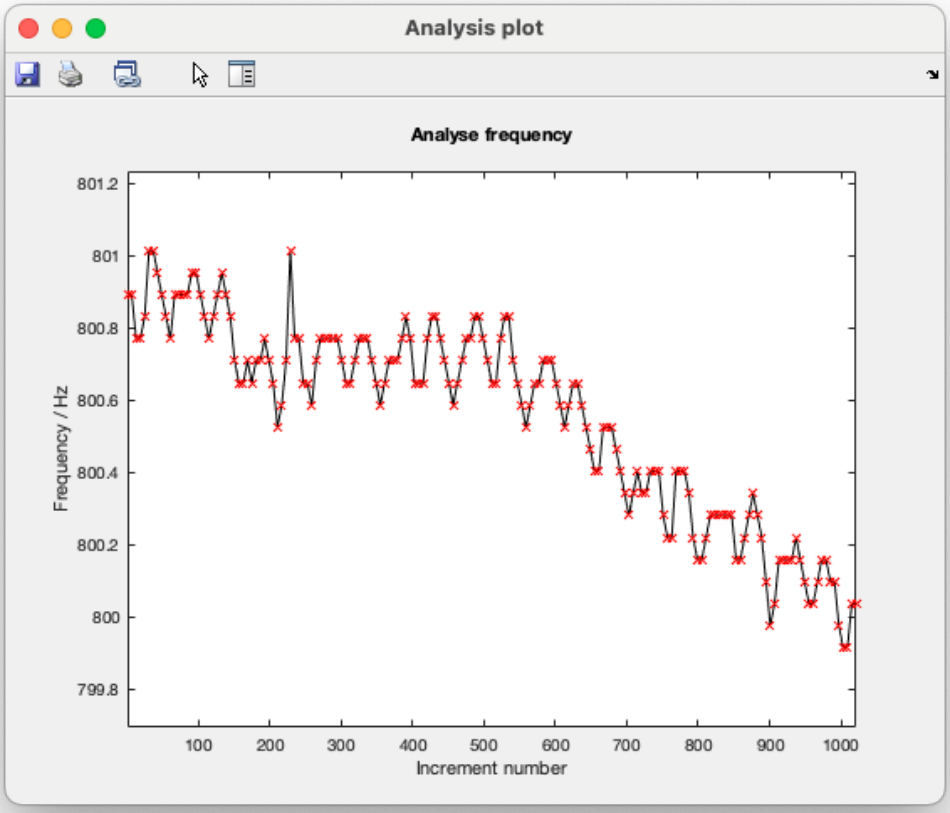
Method section

This section is intended for investigating changes in an array of spectra. To demonstrate this data from a time course to investigate spectrometer stability. A simple 1H pulse acquire experiment was recorded from a sample of PEO in D2O as 1 min intervals (plots show every sixth spectrum i.e. 1 per minute) . The spectrometer was in an room with A/C but the VT control was turned off.

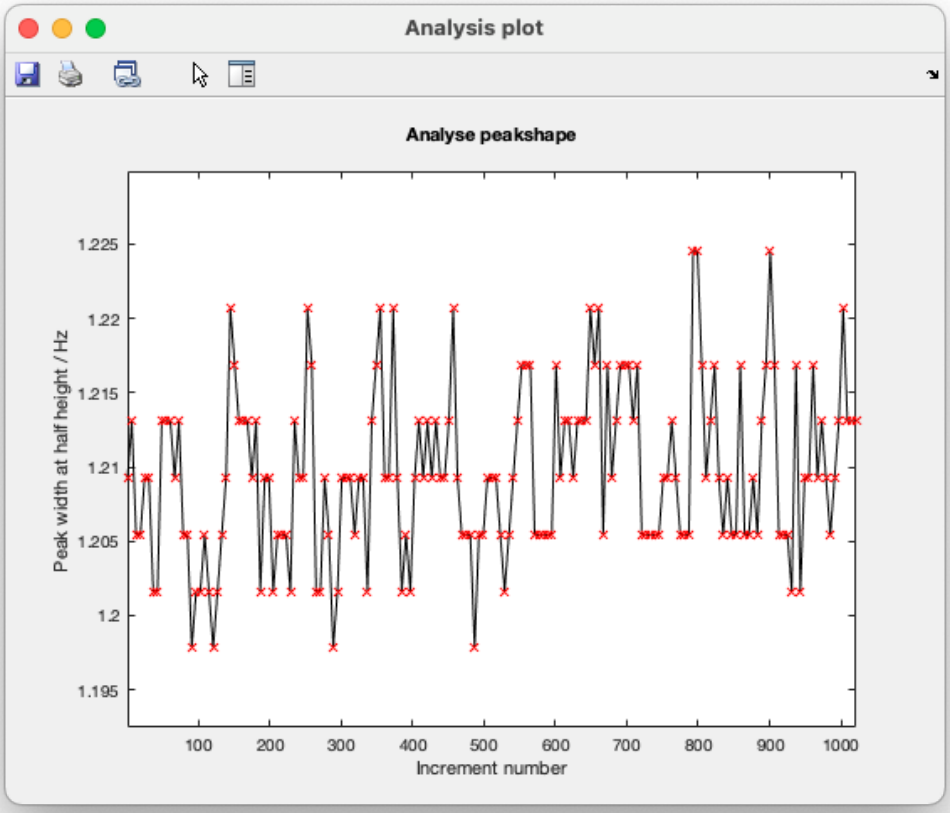
Typical use is to zoom in on a single peak and set the Threshold (Thresh button on the right side of the main window) and then press one of the buttons.



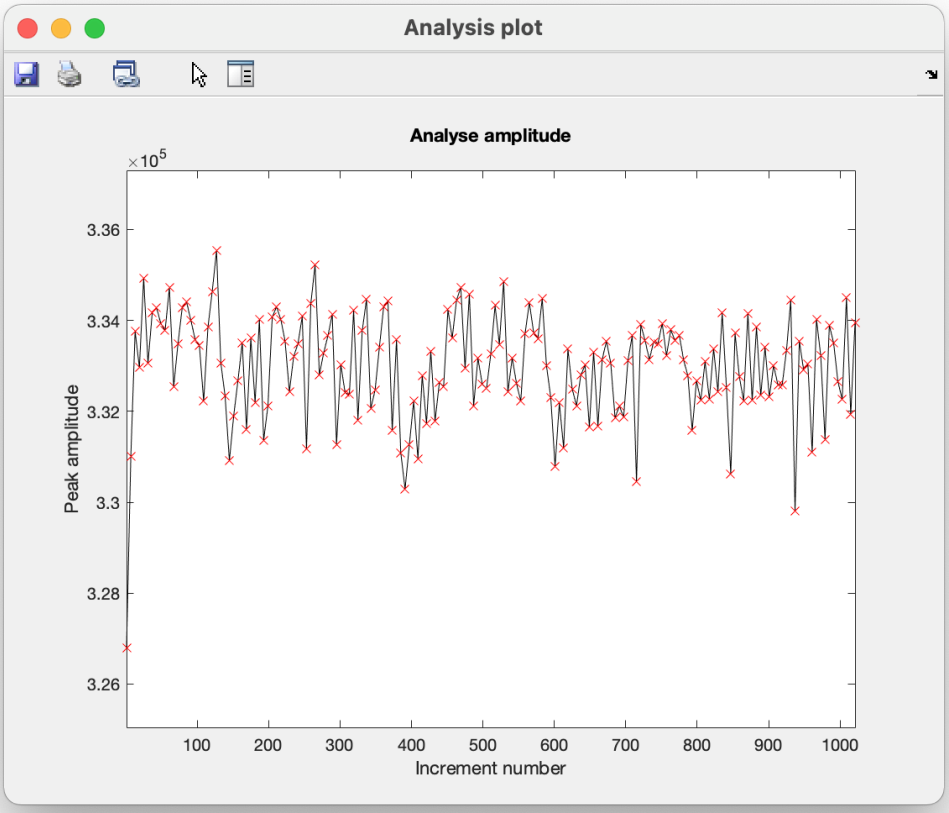
Pressing the *Freq.* button will plot peak frequency as a function of spectrum.



Pressing the *Resol.* button will plot peak width at half height as a function of spectrum.

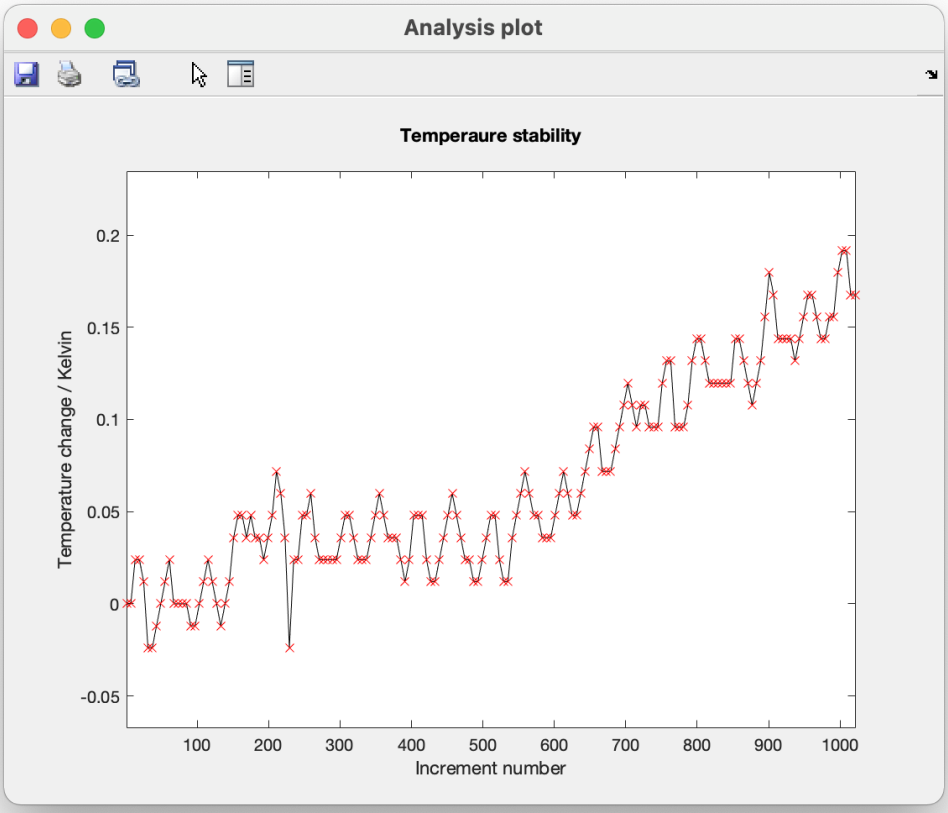


Pressing the *Ampl.* button will plot peak amplitude as a function of spectrum.

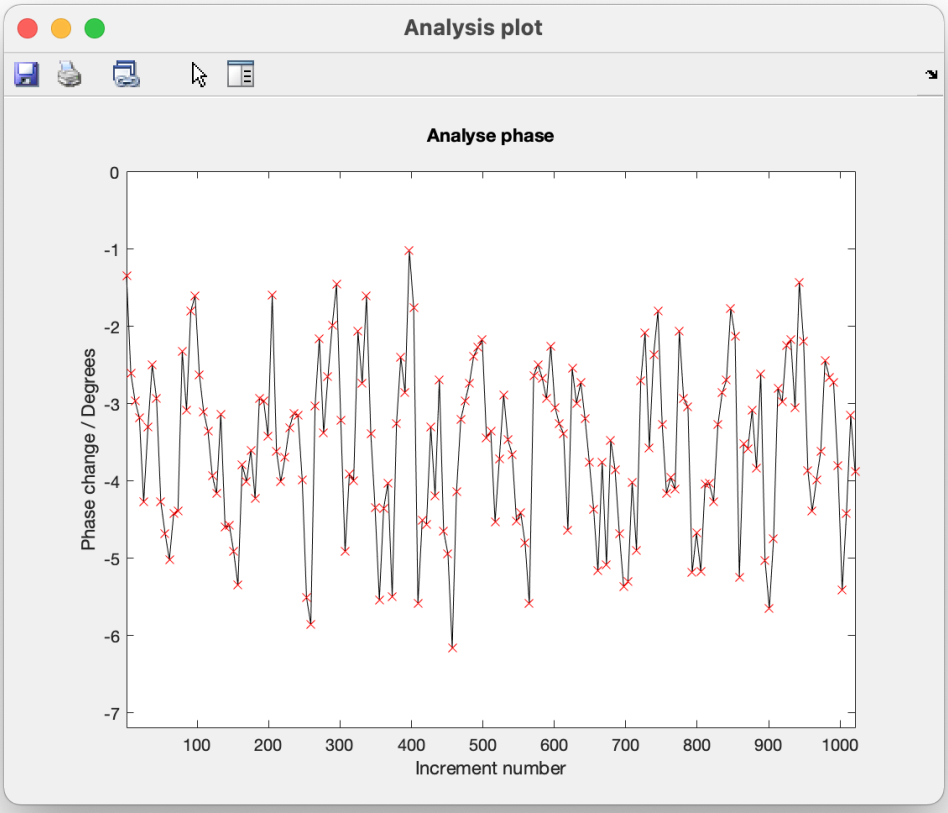


Pressing the *Temp.* button will plot temperature change as a function of spectrum. The default value is from measuring the changed for the HOD peak relative to TSP in an aqueous sample.

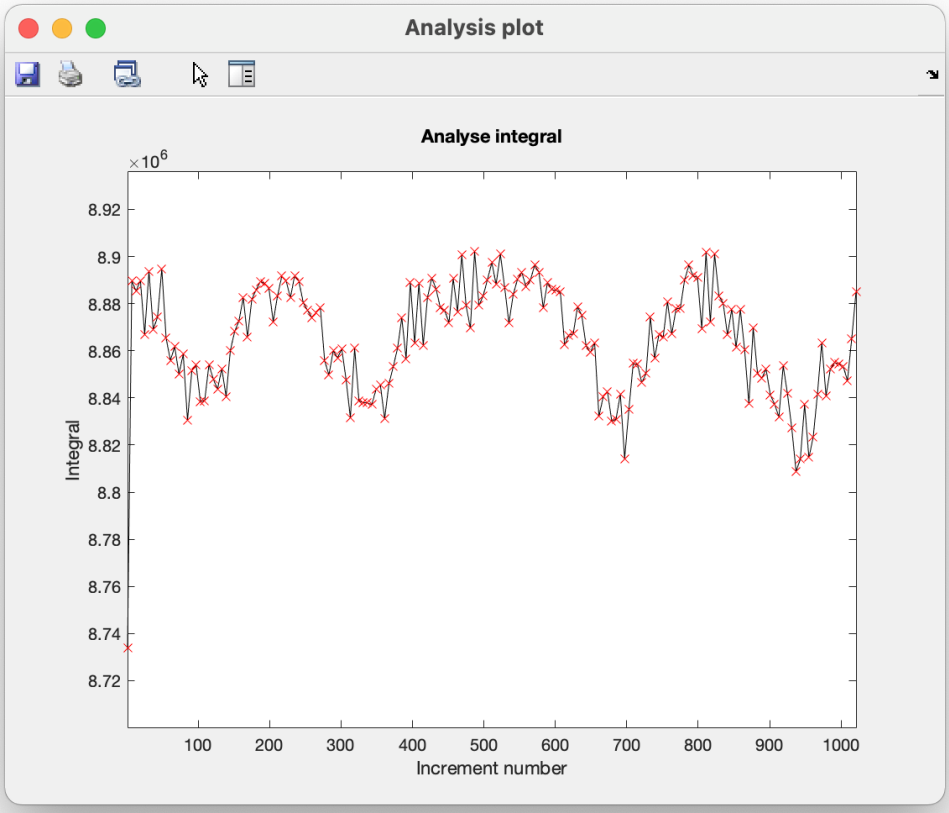
1. HOFFMAN, R.; DAVIES, D. In MAGNETIC RESONANCE IN CHEMISTRY, 1988; Vol. 26.



Pressing the *Phase.* button will plot peak phase as a function of spectrum.

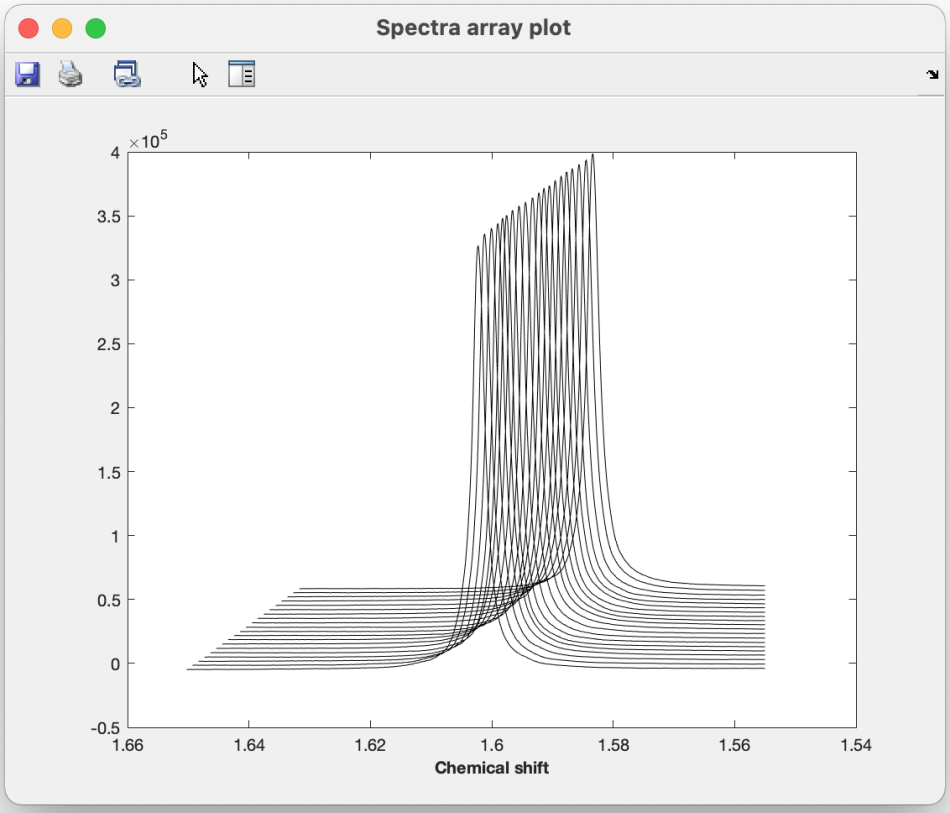


Pressing the *Int.* button will plot peak integral as a function of spectrum. Integrals are set in the Integrate tab.



Spectra array section

Here the user can plot the spectra from the array in a separate window by pressing the *Plot* button. The displayed region is plotted. The array elements to be plotted can be chosen using the *Start* , *Step* and *Stop* parameters, where Start is the first spectrum, Stop is the last, and Step determines how bit steps to take in the array. The horizontal and vertical offset can be also be controlled.



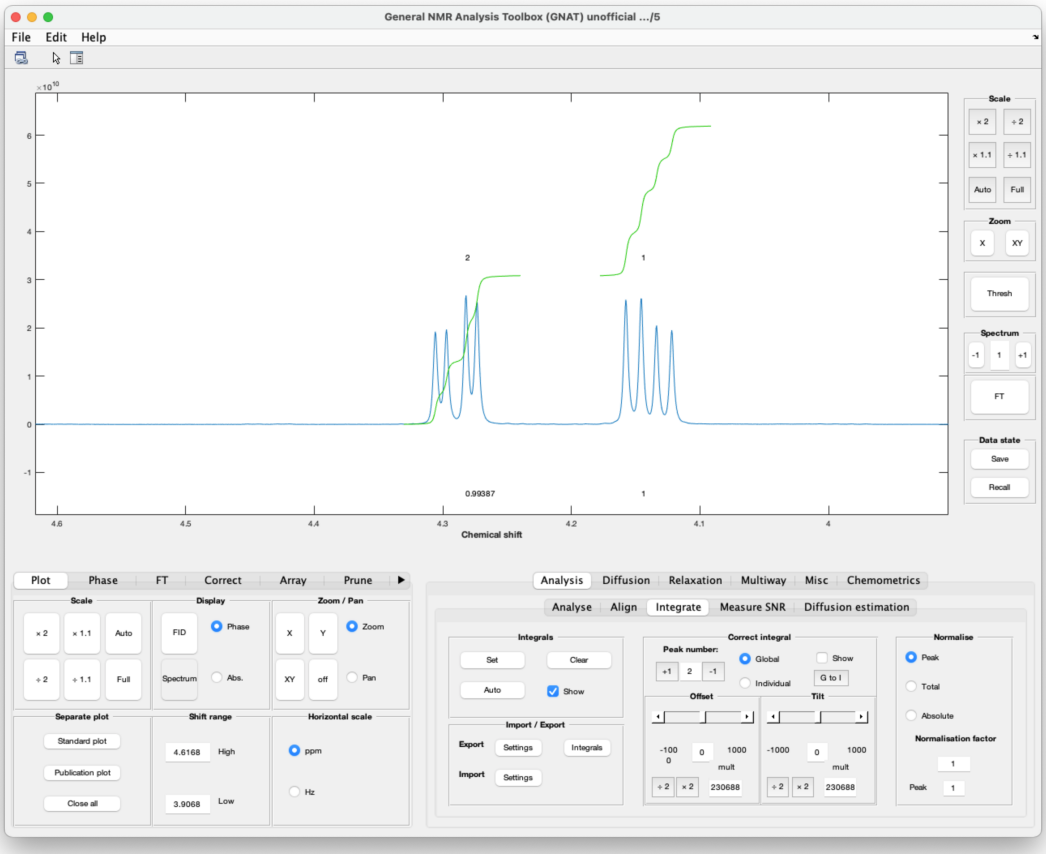
Align

Warning

This is not working at the moment. Work in progress.

Integrate

Here the user can find control for integration of spectra.



Integrals section

The *Set* buttons allows the user to set the integral regions by clicking the mouse for the position set by the pointer. The *Clear* button clears all integrals and the *Show* tick-box decides whether the regions are displayed in the spectrum.

The *Auto* button sets integral regions automatically.

> **Warning**
>
> The *Auto* button algorithm is not working well at the moment. Work in progress.

Import/Export section

Here integral values and region limits can be exported or imported. The integral regions can be imported/exported with the *Settings* button. These are currently only in a GNAT specific text format.

The integral values (and region settings) can be exported using the *Integrals* button. These can be exported in either the GNAT specific text format, or as a *.xlsx (for e.g Excel and other spreadsheet

programmes). The integral values are determined by the type of normalisation chosen (see below).

Correct integral section

Here the user can correct errors in offset of tilt of the integral regions. Which peak to correct the offset/tilt is selected in the *Peak number* box. Selecting the *Show* tick box will show a red line in the current peak.

For arrayed data the offset/tilt settings can be the same for all array elements, by selecting the *Global* radio button, or separate for all array elements by selecting the *Individual* radio button. To copy the global parameters to all of the array elements press the *G to I* button. (This is like the system for phase parameters in the Phase tab).

The *Offset* and *Tilt* parameters are adjusted in the Offset and Tilt boxes, respectively. The value can be adjusted by using the sliders or typing the value directly in the box. The value under the *mult* test is a data set specific multiplication factor and depends on the total integral of the raw spectrum. It is automatically set, but can be adjusted by the user as needed, either by typing a value directly in the box or by using the buttons to double of half the current value.

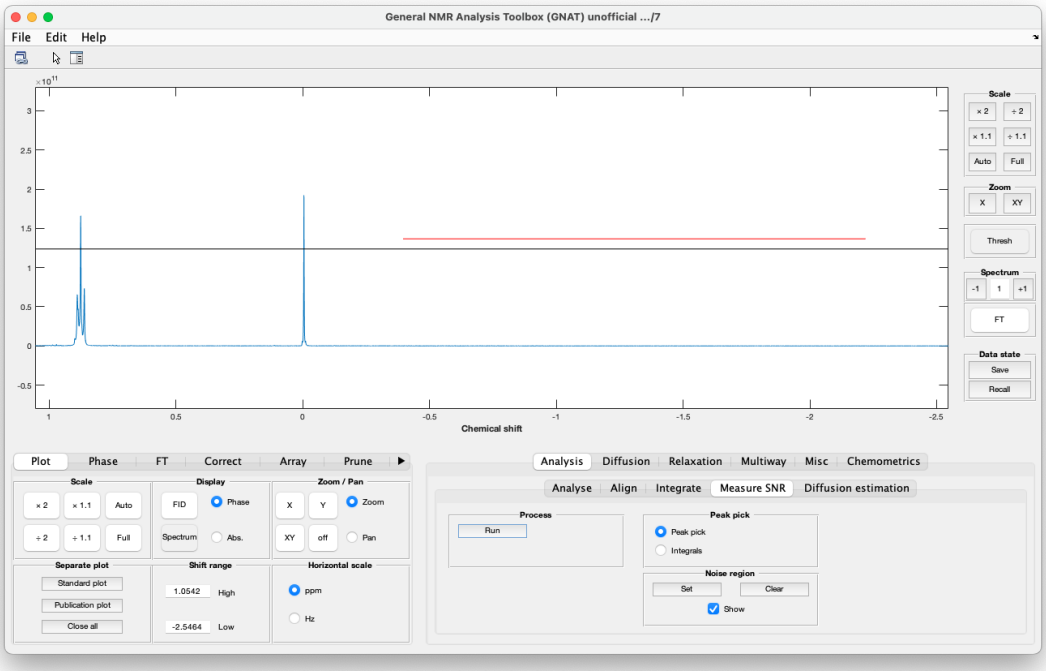
Normalise section

Here the user choses the sort of normalisation used for the integration. If the *Absolute* radio button is selected there is no normalisation and the raw integrals will be used.

Normalisation can be done to a *Total* value for all the integrals; the value is set in the *Normalisation factor* box. Normlisation to a specific *Peak* can also be chosen, in which case the peak number is selected in the *Peak* box.

Measure SNR

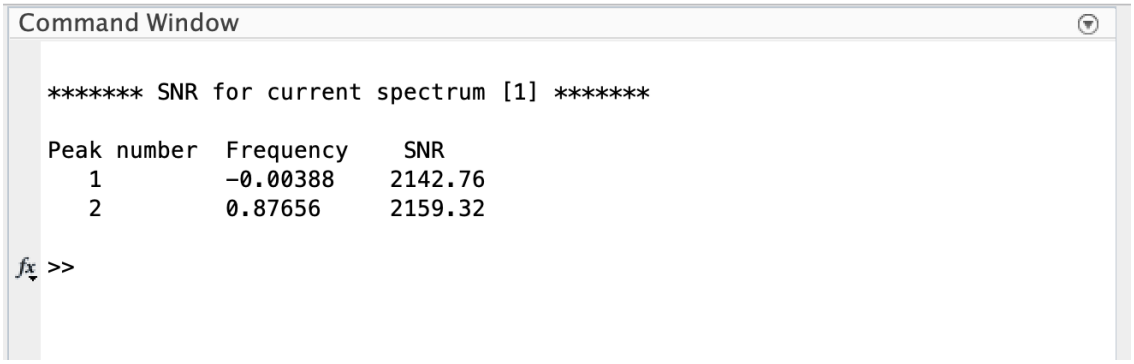
Here the user can find measure the signal-to-noise ratio (SNR) of spectra.



You can use peak picking over a threshold (*Thresh* button in the shortcuts on the right of the main window - black line shown in figure), or define the peaks in the Integrate tab. The noise region (i.e. a piece of baseline with only noise) is defined in the *Noise region* section (red line in the figure).

The SNR for peak picking is defined a the max value of the peak divided by 2 times the root mean square amplitude of the noise, and for integrated peaks it is the sum of the integral values divided by 2 times the root mean square amplitude of the noise.

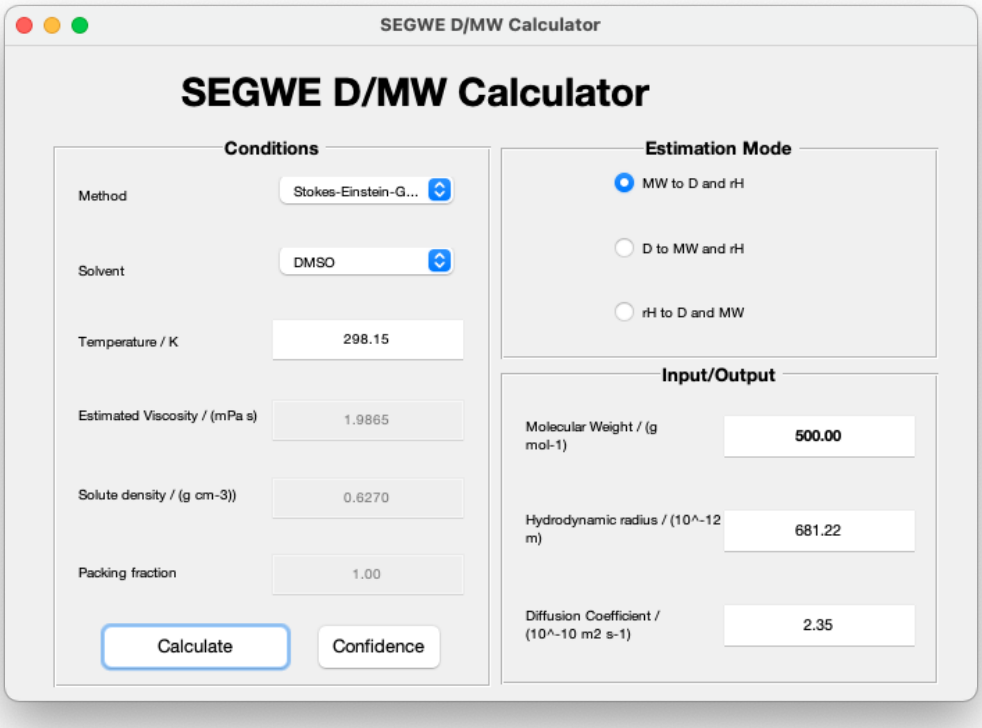
Pressing the *Run* button will display the result in the Matlab window (or termnal window for compiled versions)



Diffusion Estimation

Here the user can estimate diffusion coefficients, molecular weights and hydrodynamic radii.

Pressing the *Initiate* button will open a separate GUI for these estimations.



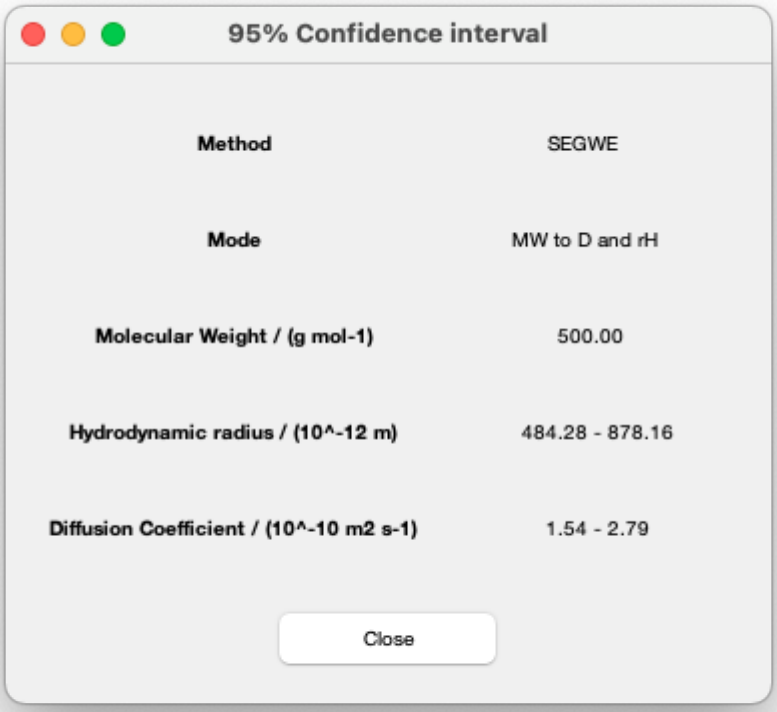
The theory and practicalities behind these calculation is given in the following papers:

1. Evans, R.; Deng, Z.; Rogerson, A. K.; McLachlan, A. S.; Richards, J. J.; Nilsson, M.; Morris, G. A. Quantitative Interpretation of Diffusion-Ordered NMR Spectra: Can We Rationalize Small Molecule Diffusion Coefficients? Angewandte Chemie-International Edition 2013, 52 (11), 3199.

2. Evans, R.; Dal Poggetto, G.; Nilsson, M.; Morris, G. A. Improving the Interpretation of Small Molecule Diffusion Coefficients. Analytical Chemistry 2018, 90 (6), 3987.

Conditions

In the *Conditions* section the user can select the underlying method for the calculations. The default is the SEGWE (Stokes-Einstein-Gierer-Wirtz) described in the above papers, and the user can also choose the conventional Stokes-Einstein method (see above papers for more information). The user also need to choose a solvent and a temperature. The most common NMR solvents are available and there is also a user defined version where the user provides the solvent viscosity and molecular weight. For the SEGWE method the solute density and packing fraction is fixed at predetermined values, but for the SE method these are under user control. Pressing the *Calculate* button will display the result depending on the input values selected in the *Estimation Mode* and *Input/Output* sections. Pressing the *Confidence* button will give the confidence values for the estimation (only works for the SEGWE method)



Estimation Mode

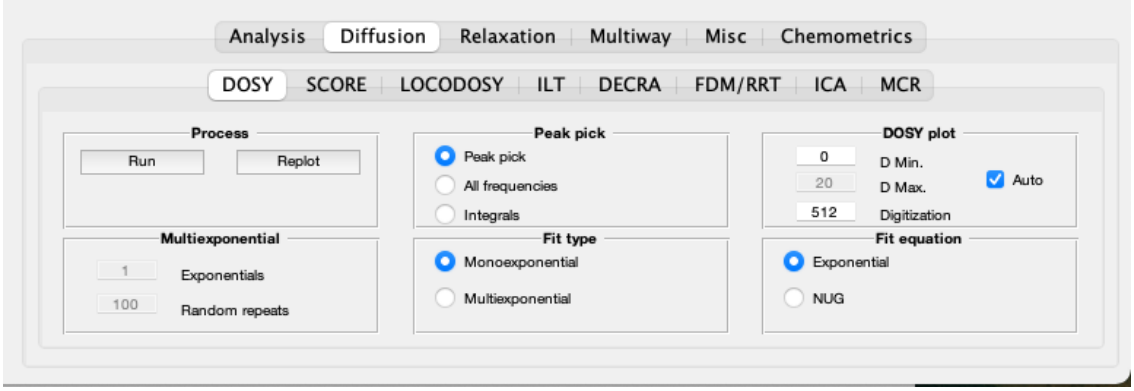
In this section the mode of estimation. One can choose either molecular weight (MW) , the diffusion coefficient (D), or the hydrodynamic radius (rH) of the solute to estimate the other two.

Input/Output

Here the the input value of the chosen parameter (MW, D, or rH) is set by the user, and the resulting estimation of the other two is displayed after the calculation.

Diffusion

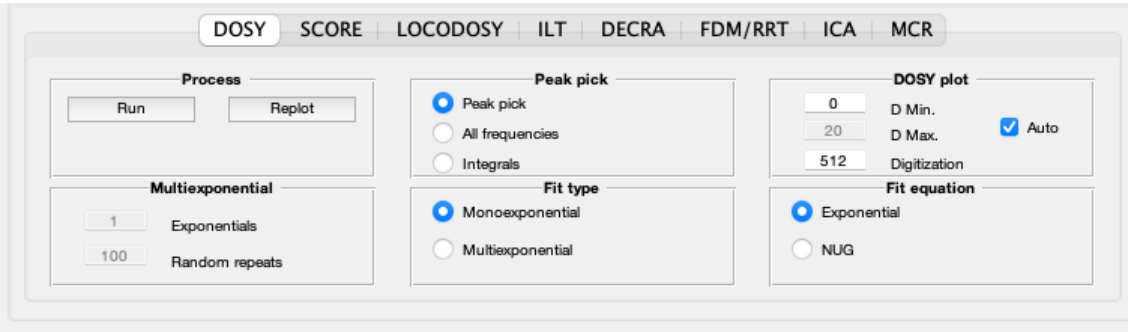
This is the head tab for various way to analyse your diffusion NMR data.



Functionalities

DOSY

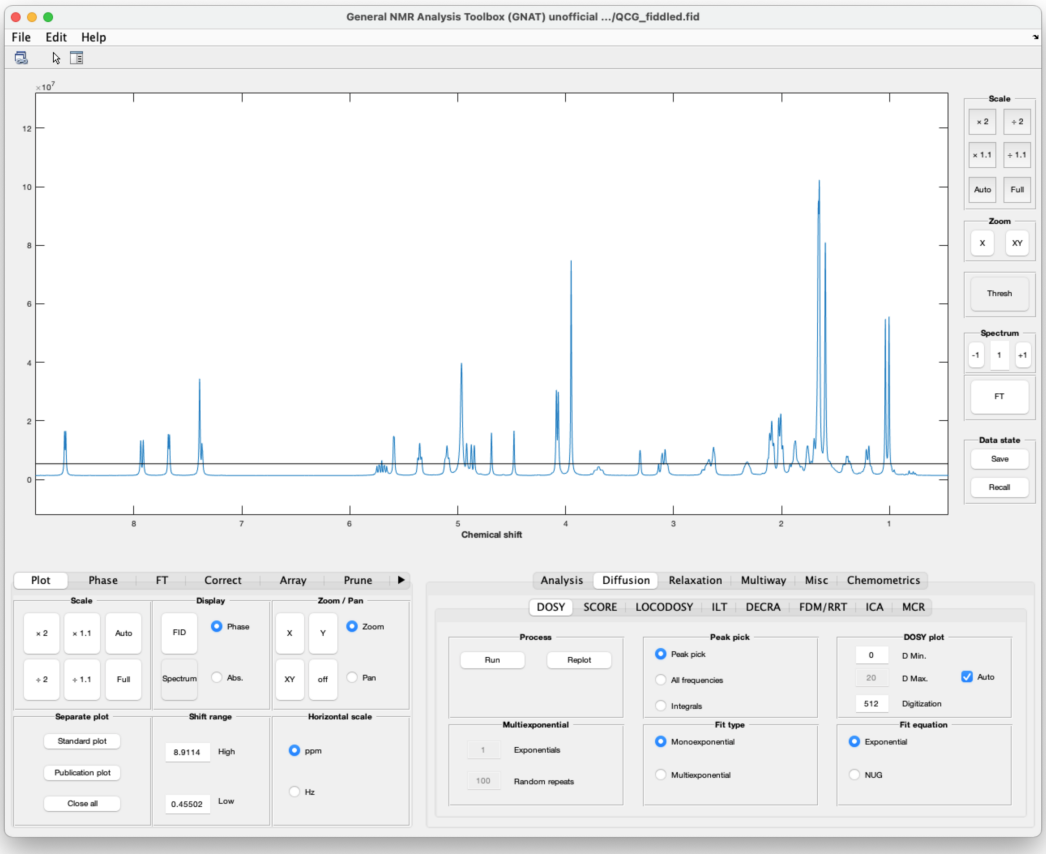
This is the tab for DOSY processing if diffusion data.



Before processing the data in this tab, make sure that the spectra have been properly preprocessed (phase, baseline correction etc), and that the diffusion parameters are correct (see the Settings section.)

Note

Quickstart : zoom into the part of the spectrum you are interested in, set a threshold with the *Thresh* button (right part of the GUI) and press the *Run* button



A good introduction to DOSY is given in the below article, and references therein.

1. Nilsson, M. The DOSY Toolbox: A new tool for processing PFG NMR diffusion data. Journal of Magnetic Resonance 2009, 200 (2), 296.

Process section

Here the user can access various processing functions. The button *Run* will start DOSY processing with the parameters set in the other sections in this tab, but also e.g. the threshold with the *Thresh* button (right part of the GUI), the settings in the Prune tab, and in the Diffusion tab in the Settings . This will open up the DOSY Plotting GUI.

The *Replot* button will open up the DOSY Plotting GUI with the last processed data (e.g. in case it was closed by mistake)

Peak pick section

Here the user can choose which peaks that will be used for DOSY processing.

The default is *Peak pick* which automatically picks all the peaks over the threshold set with the *Thresh* button (right part of the GUI).

The option *All frequencies* will use all data points over the threshold.

The option *Integral* will use all integral regions set in the Integrate tab.

DOSY plot section

Here some parameters for the DOSY plot in the DOSY Plotting GUI are set.

D min sets the lowest diffusion coefficient to be displayed (in $10^{-10} \: m^2s^{-1}$)

D max sets the highest diffusion coefficient to be displayed (in $10^{-10} \: m^2s^{-1}$) if the *Auto* box is ticked (default) the *D max* will be set depending on the highest fitted diffusion coefficient.

Digitization sets the number of data points in the diffusion dimension. The number of points in the spectral dimension is the same as the number of spectral points displayed in the main window. The plotting routines in Matlab can be quite slow so a high number may make plotting glacial. If this becomes a real problem (e.g. on older hardware) it is advisable to plot a limited spectral and/or diffusion range. (more about the digitization on the DOSY Plotting GUI page)

Multiexponential section

This section, which only becomes avalibale when *Multiexponential* is selected in the *Fit type* section, is for parameters relates to multi exponential fitting of DOSY data. That mean that for each peak (as selected in the peak pick section) The programme will try to fit two, or more, components. The algorithm tries to fit that maximum number of components and if that fails it will revert to a lower number. The criterion for a successful fit is that the standard error for the diffusion coefficient is less than 20%.

The *Exponentials* box sets the max number of exponentials to try (integer value)

The *Random repeats* box sets the maximum number of random starting values that are tried for each peak. The random values are taken from a Gaussian distribution around the fitted values for a monoexponential

Some more information about multiexponential fits of DOSY data can be found in:

1. Nilsson, M.; Connell, M. A.; Davis, A. L.; Morris, G. A. Biexponential fitting of diffusion-ordered NMR data: Practicalities and limitations. Analytical Chemistry 2006, 78 (9), 3040.

Fit type

Here there is the option to set either a monoexponetial fit (default) or multiexponential fit. (see Multiexponential section above for more information)

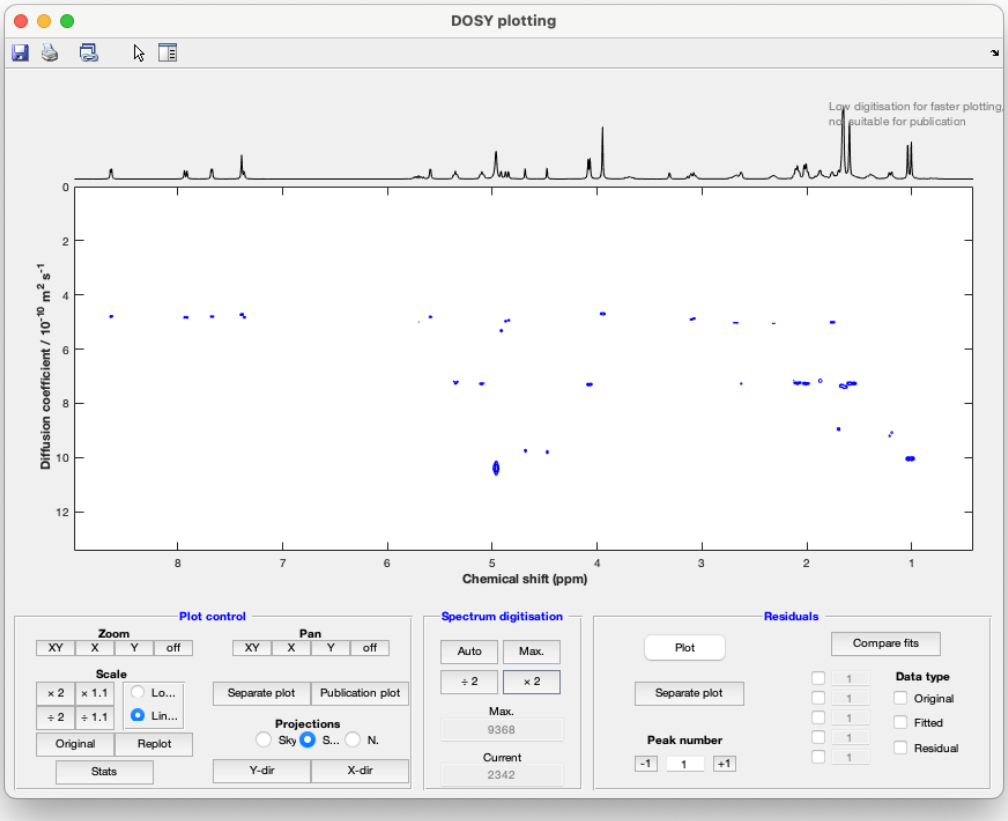
Fit equation

Here the equation describing the diffusional signal decay is set. The default is the Stejskal-Tanner equation (i.e. a pure exponential). The NUG (non-uniform gradient) is probe specific and can provide more accurate results. More information can be found in the Settings page.

Functionalities

DOSY Plotting

This is the main GUI for plotting DOSY spectra, and inspecting the data.



It is also use for relaxation data (see the ROSY section.)

> **Note**
>
> The grey text in the top corner mean that low digitization is used to speed up the plotting. To produce spectra with higher quality see the section on *Spectrum digitization*

Plot control section

Zoom/Pan

Controls for zooming or panning the display.

Scale

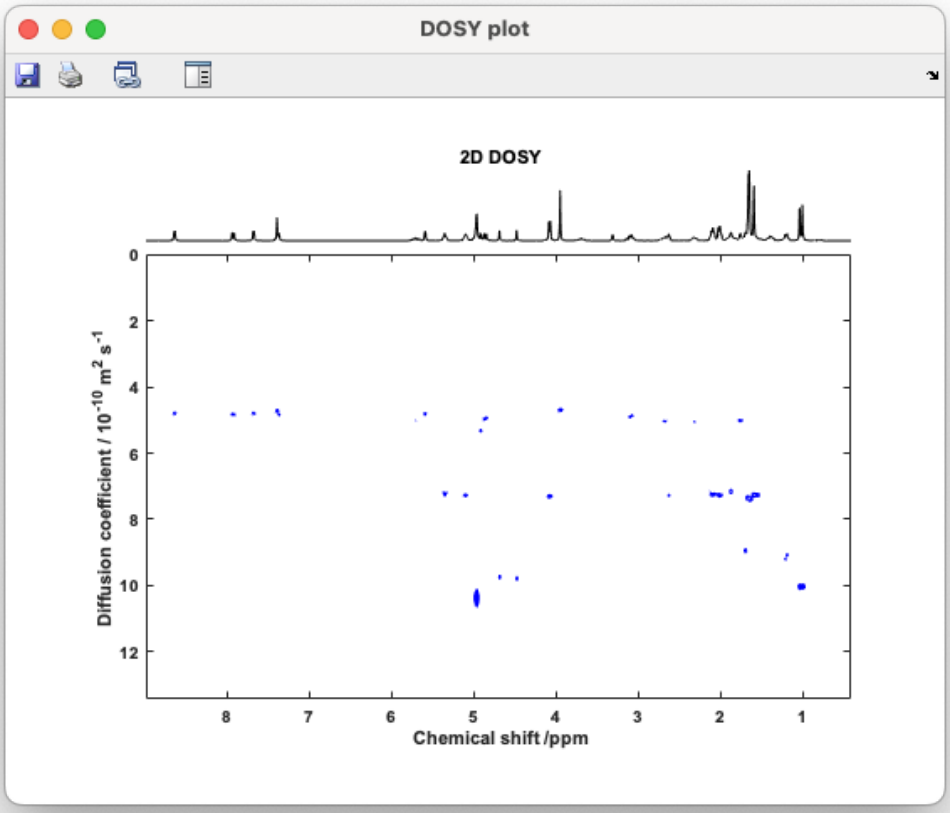
Controls to scale the contours in the 2D DOSY plot . You can multiply or divide by a factor 2 or 1.1. The **Original** button resets to the initial scale and the **Replot** button plots the whole initial spectrum.

The radio buttons switches between linear and logarithmic scale for the diffusion dimension.

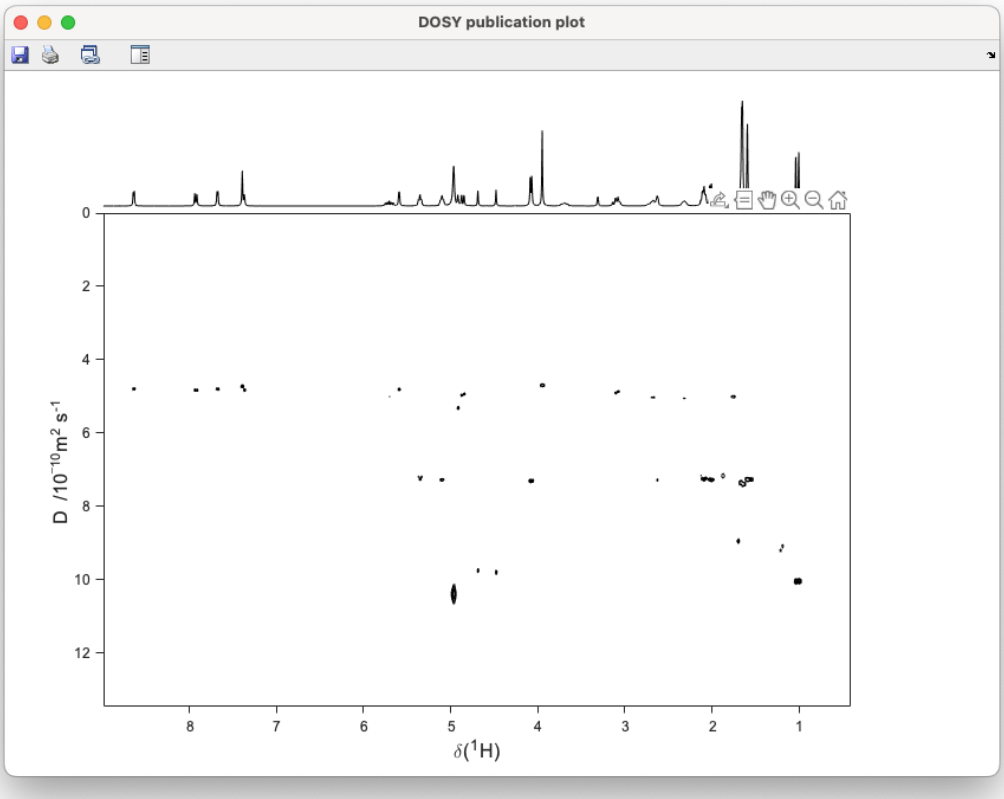
The **Stats** button allows the user to save a text file with the fitting statistics: the dosystats.txt file.

Separate plot

The **Separate plot** button plots the spectrum as seen in the main window.



The **Publication plot** button plots the spectrum in a format more suitable for publication.

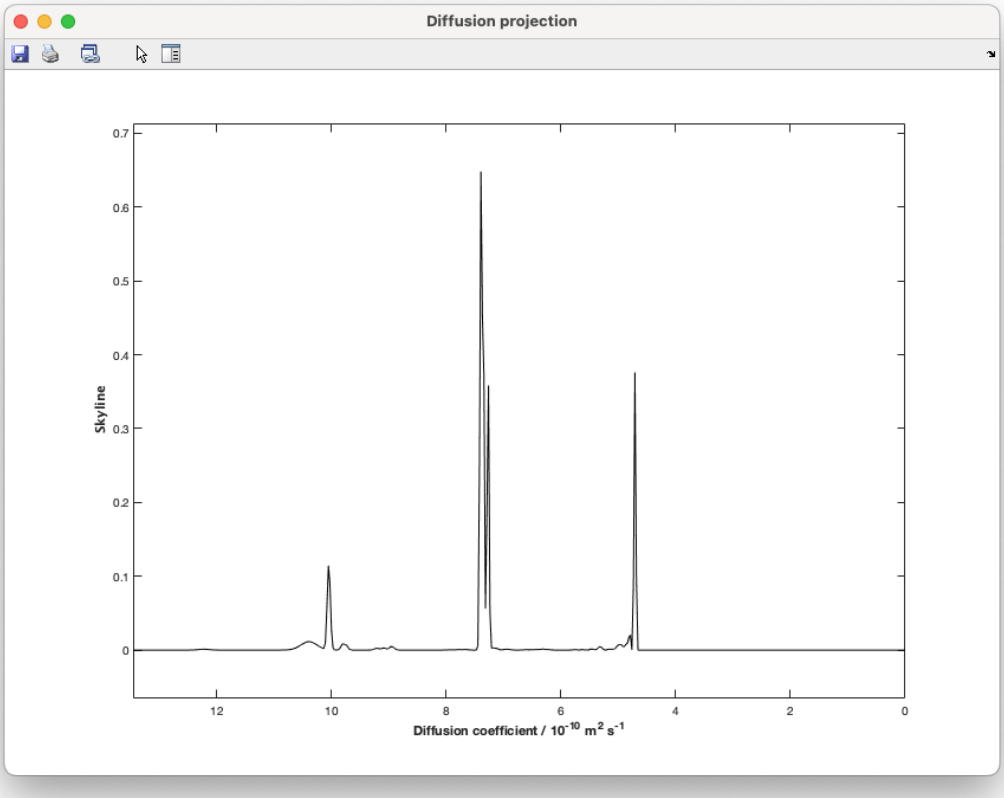


The plots can be saved in the available Matlab formats (e.g. fig, svg, eps, png, jpg, pdf)

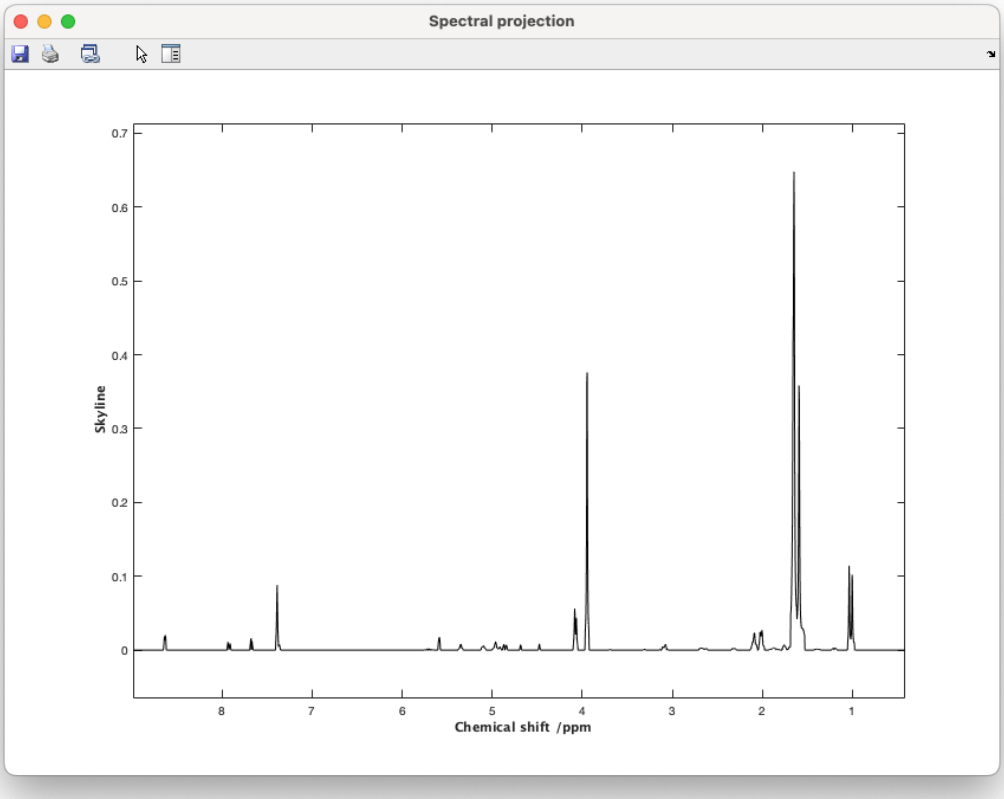
Projections

Here the user can plot the Y (spectral) or X (diffusion) projections of the displayed DOSY plot.

The **X projection** (skyline)



The **Y projection** (skyline)



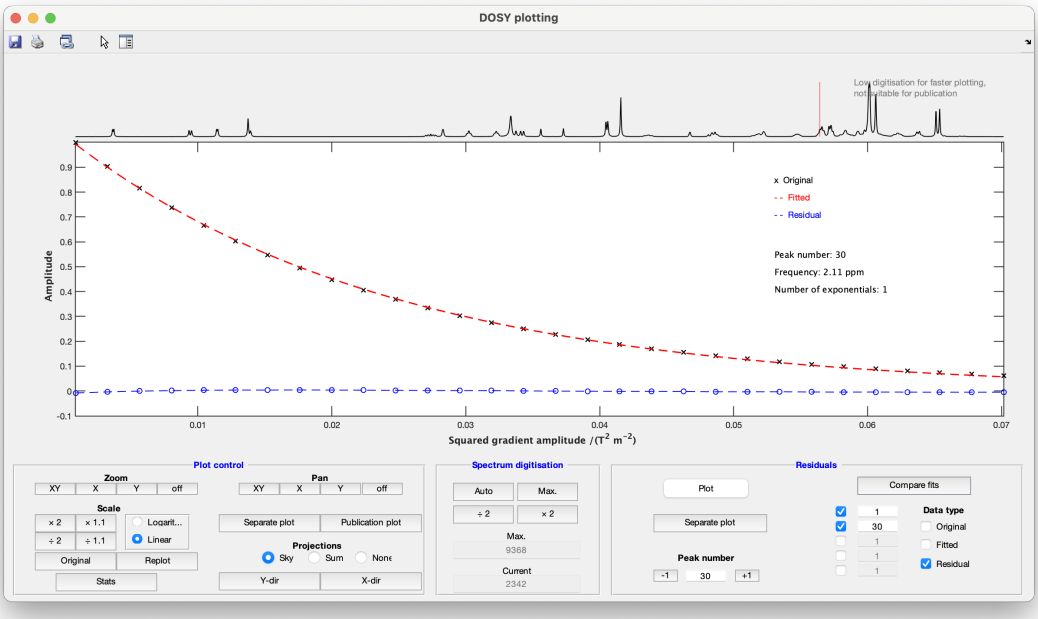
Spectrum digitization section

Matlab is unfortunately very slow at plotting 2D data. Therefore GNAT has an automatic downsampling to something closer to the screen resolution. This speeds up the plotting, but can also sometime cause distortions in the plot. If downsampling is present a grey warning text is displayed in the top right corner of the spectrum.

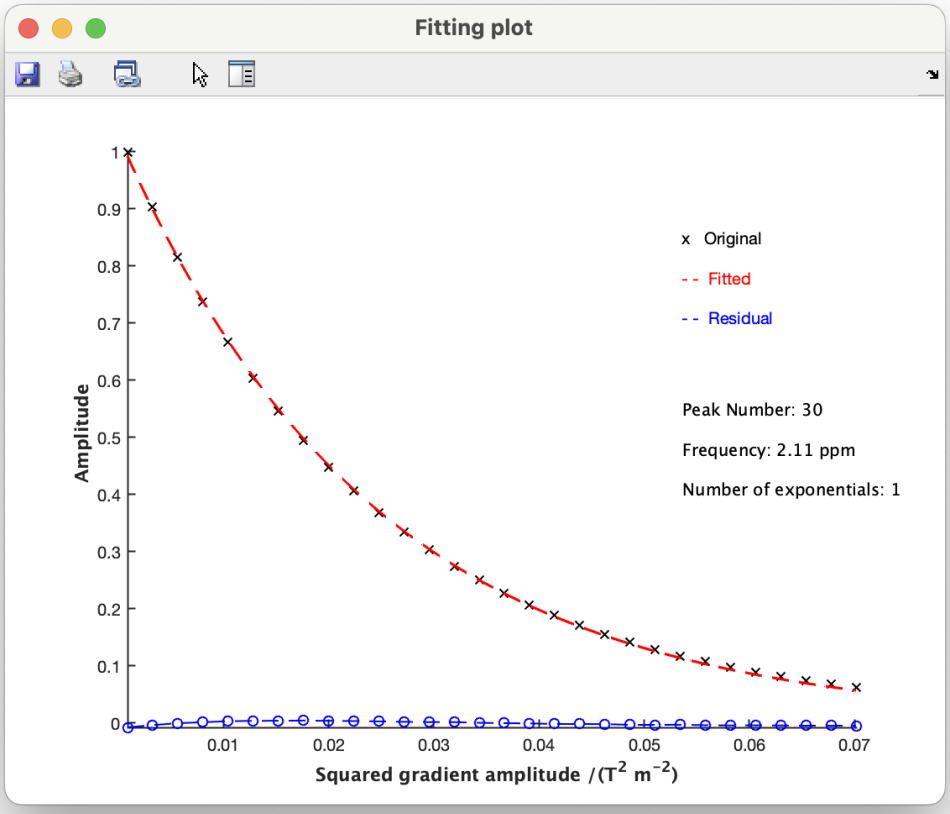
The sampling is under user control where the **Auto** button chooses the parameters automatically, trying to usefully match the screen resolution. The **Max** button sets the resolution to max, which can sometimes be very slow. There are also buttons to have or double the digitization, and the values are current and max values are shown below.

Residuals section

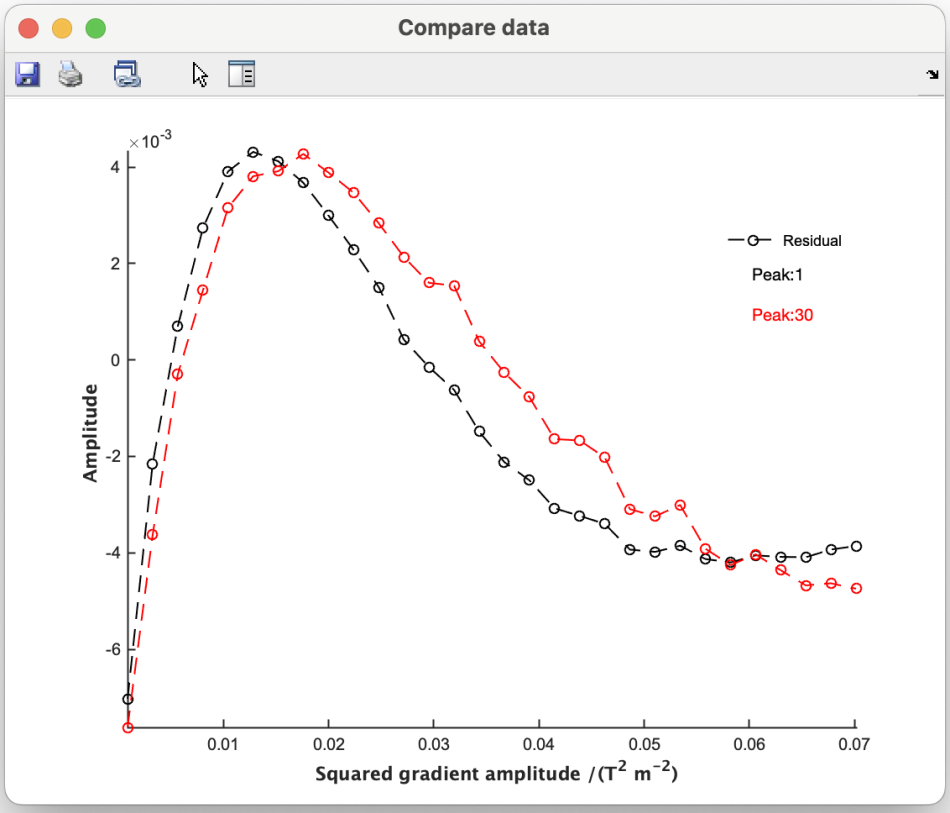
This section is for plotting and comparing individual fits and residuals. The **Plot** button plots fit and residual in the main DOSY GUI window.



The current peak indicated with a red line in the spectrum on the top of the plot, and the peaks number can be selected in below. The **Separate plot** button produces a separate plot for the selected peak.



Fits and residuals can also be compared for up to 5 peaks, and the user can select any combination of raw data (original), fit and residuals. The comparison is displayed using the **Compare fits** button. Below is a comparison of the residuals for peak 1 and peak 30.



SCORE

LOCODOSY

ILT

DECRA

FDM

ICA

MCR

Relaxation

Functionalities

ROSY
T

RSCORE

Multiway

Functionalities

PARAFAC

Slicing

Misc

Functionalities

Sim DOSY

Macros

Chemometrics

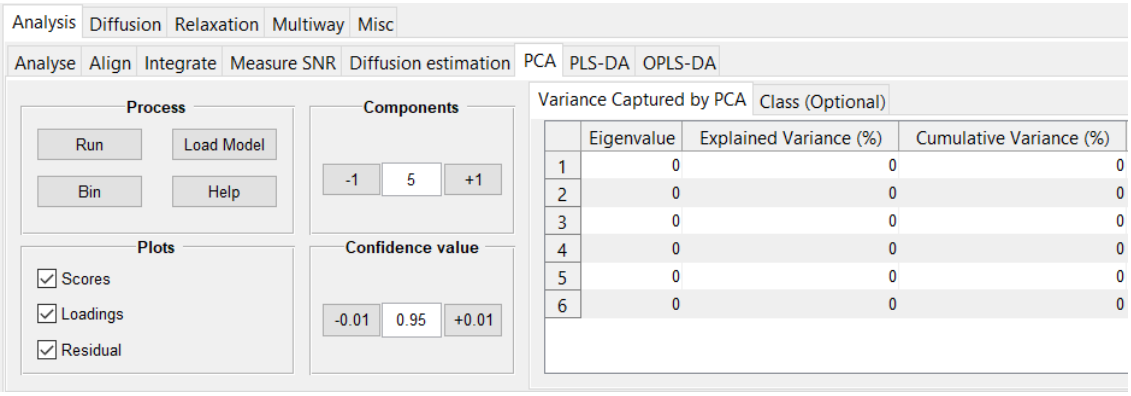
Functionalities

PCA

PCA is an effective method for extracting information from massive data sets. It aims to reduce a larger set of predictor variables to a smaller set with minimal information loss by linearly combining the original variables to form new variables known as principal components (PC's), which maximize the explained variance for a given number of components.

Structure of PCA Tab

It's possible to perform a PCA analysis in GNAT (General NMR Analysis Toolbox) by accessing the Analysis Functionalities on GNAT (right functions) on the tab Chemometrics>PCA. The steps below describe how to create a new PCA transformation on your data:



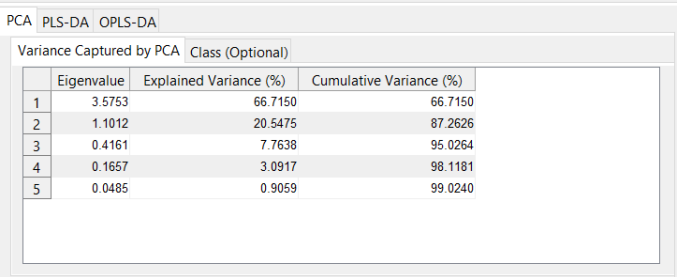
The major controls for PCA plots are determined by the panels **Components** , **Confidence Value** , and **Plots** :

1. Select the number of components to be visualized.

2. Select the confidence value for the limits to detect possible outliers.

3. Determine the plot to be visualized - Scores, Loadings and Residual.

The table **Variance Captured by PCA** shows the explained variance After pressing the button `Run` in the **Process** panel. The number of components presented on this table is dictated by the value imputed on the `Components` edit box.

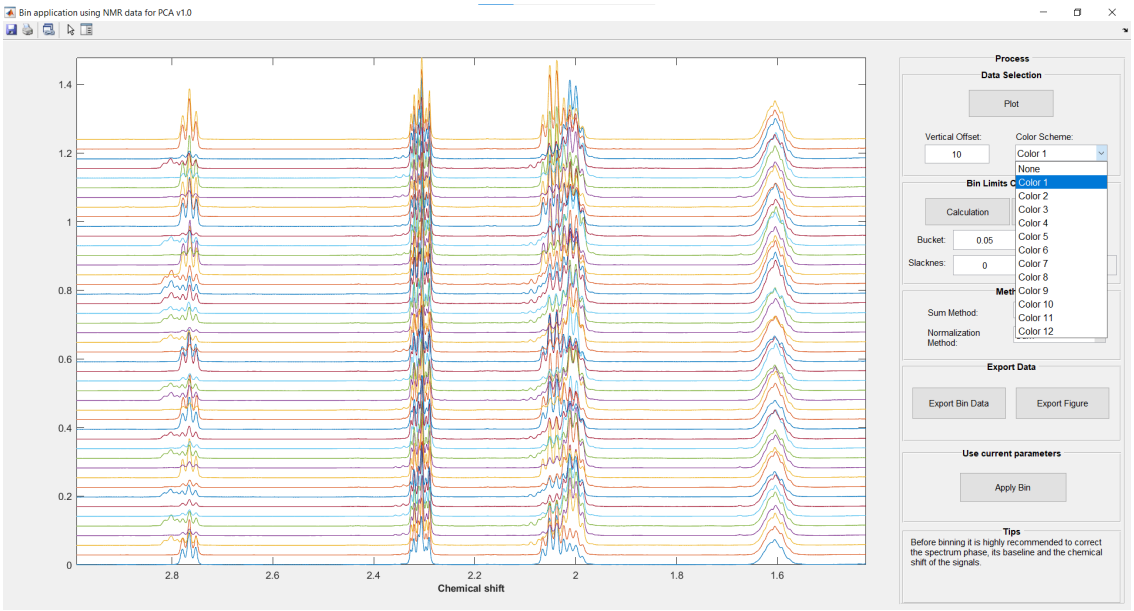


	Eigenvalue	Explained Variance (%)	Cumulative Variance (%)
1	3.5753	66.7150	66.7150
2	1.1012	20.5475	87.2626
3	0.4161	7.7638	95.0264
4	0.1657	3.0917	98.1181
5	0.0485	0.9059	99.0240

Binning GUI

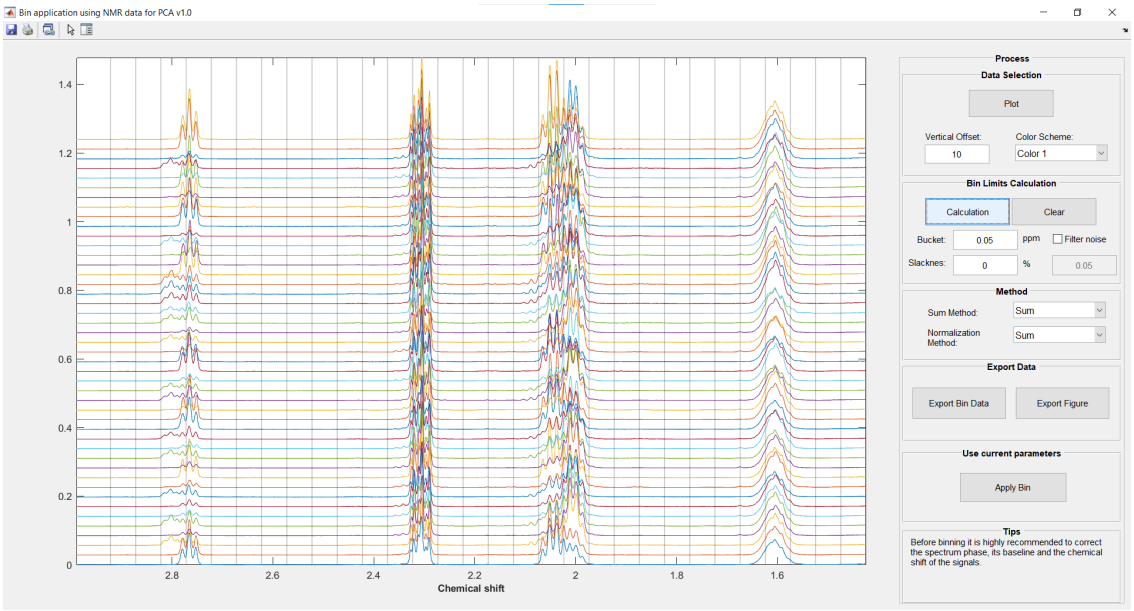
The pre-processing method **Binning** is available by pressing the button `Bin` in the the panel **Process** . The 'Binning' GUI will open. The active windows on GNAT (spectrum display on the main axis) will determine the limits of the spectrum to apply the binning.

Press `Full` on GNAT (panel on the left) to apply Binning method on the full spectra. The user can also select the width of each integral region.

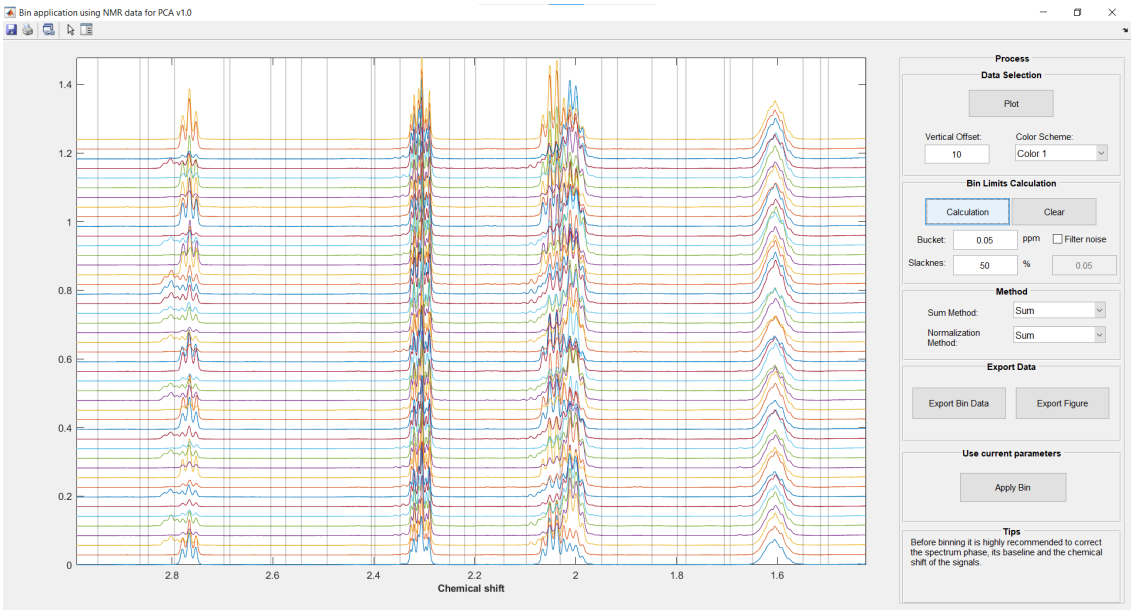


Binning is the process of integrating spectral data into areas of similar length in order to reduce the impact of differences in peak locations induced by physicochemical influences in the samples.

After defining a value for `Bucket` in the **Bin Limits Calculation** , the spectra is separated into non-overlapping regions/bins of predetermined size in the traditional technique, with widths ranging between 0.01 and 0.05 ppm. A typical 64k point NMR spectrum would be reduced using bin widths of 0.04 ppm, resulting in ~250 bin integral values.



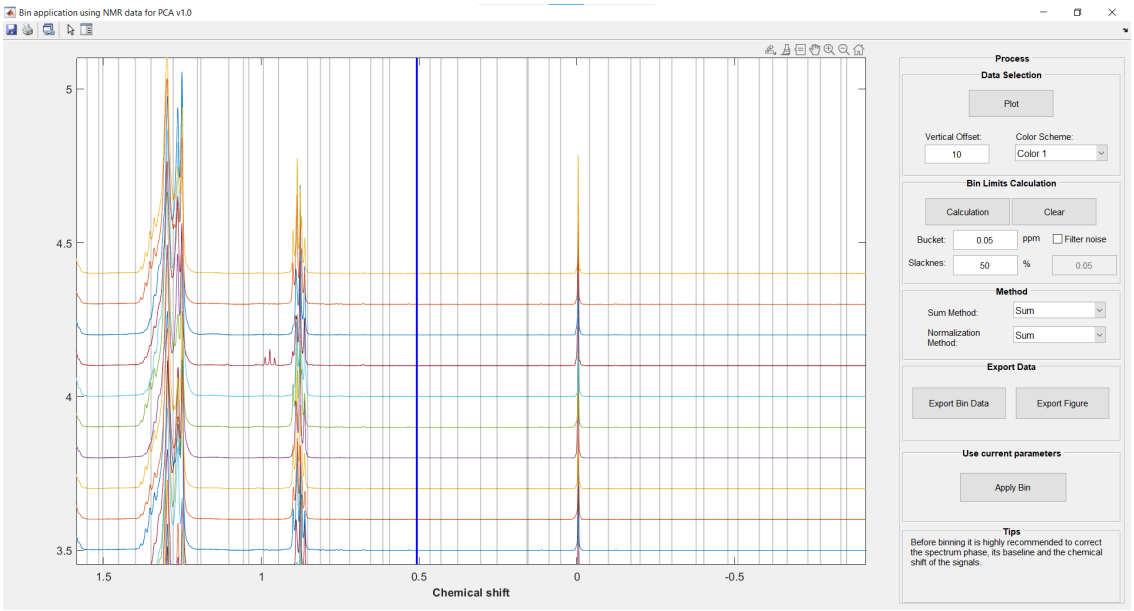
When defining a value for `Slackness` (a value between 0 and 100) the optimized bin boundary will be calculated. Sleckness is a threshold can vary while looking for local minima in the mean spectrum, in % of the `Bucket` value.



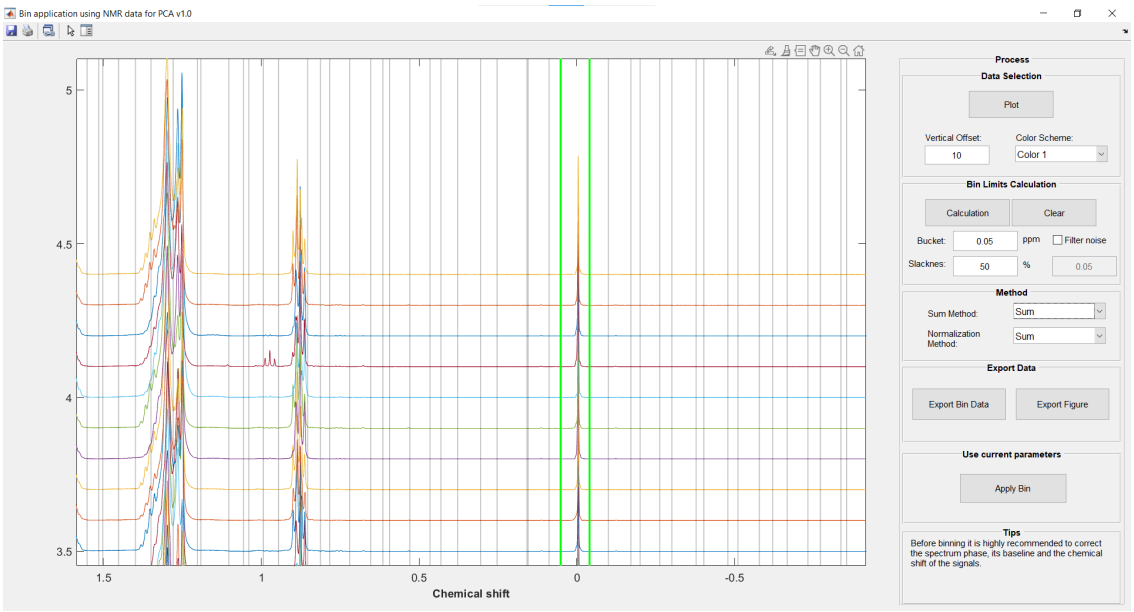
Warning

The function `Filter Noise` is used to define bin limits close to the spectrum signals. However, it is not fully optimized at the moment

All the limits can be moved after left-clicking in a existing limit



If the method **Reference** is selected on the `Normalization method` pop-menu, the user can right click the right/left limits of the NMR region that will be use to normalize each spectrum



After finishing the calculation, it is necessary to press the button `Apply` so save the modification.
.. figure:: PCA/Fig12_BIN_GUI_apply.png

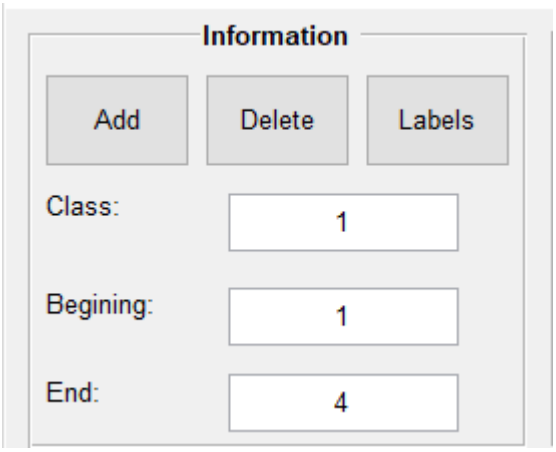
Note

It is important to note that, to apply the modification made in the Bin GUI, is necessary to maintain the GUI open before creating the PCA , PLS-DA or OPLS-DA models

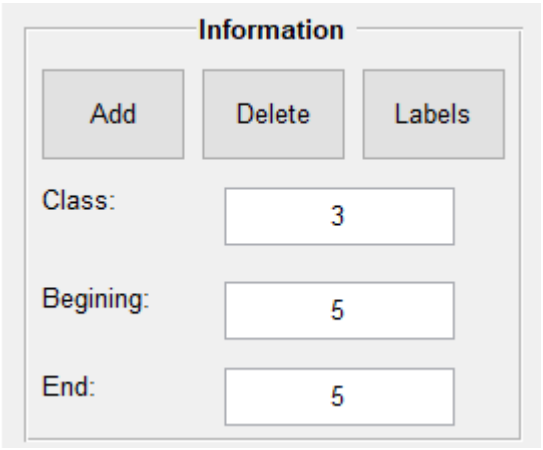
Class GUI

PCA is a non-supervised approach, hence is not necessary to determine one class for each sample in the dataset loaded into GNAT. However, the user can utilize the **Class** tab to build this array. The user can use the button `Add` after defining the **Class** , **Beginning** and **End** for each sample.

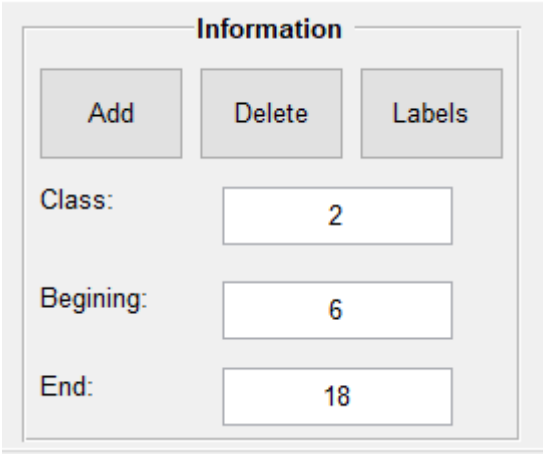
Case 1 : The first four samples belong to the class 1, so the following parameters should be used in the tab:



Case 2 : Sample 5 belongs to class 3:



Case 3 : Samples between 6 and 18 belong to class 2:



All the samples need to be associated to a numerical class. Labels to each class can be defined later using the Class GUI. All samples that are not associated to a label with have the number of their class as their label. It's also possible to exclude sample of the dataset imported by using the include menu.

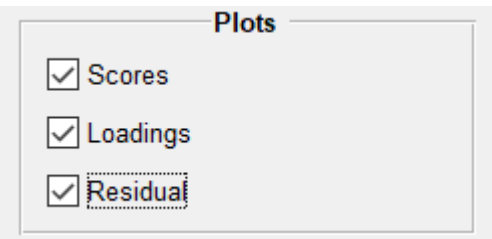
Warning

All the inputs for each edit box need to be a number. Different error messages will appear for other type of inputs. But, when the last class added need to be deleted, the user can erase the value of the **Class** editbox and press delete to perform this

After defining the classes, the user can divide the imported dataset into a calibration and validation set of samples using the Split Cal/Val panel. There are three algorithms for this division:

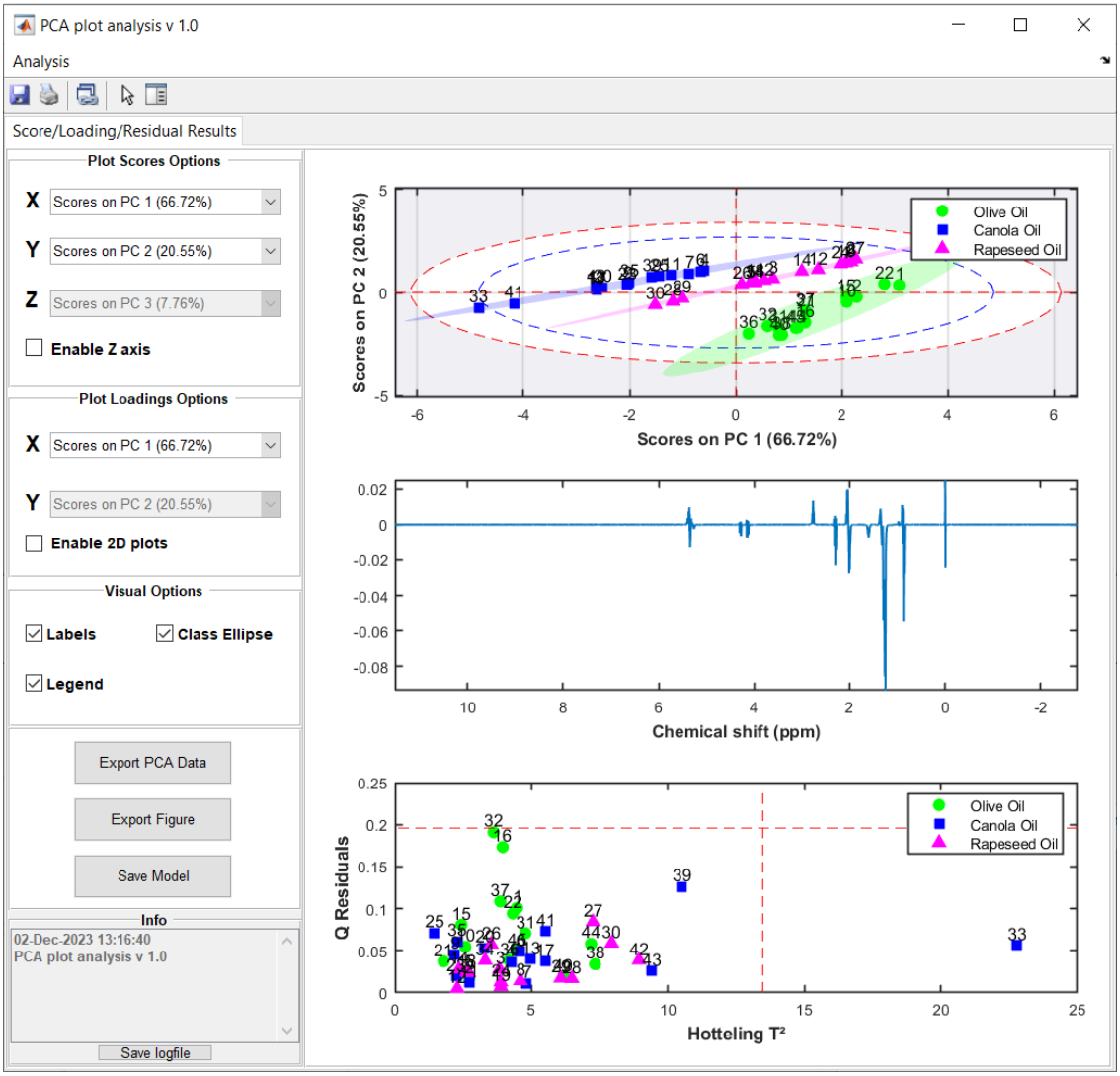
Plot GUI

Before pressing `Run` to calculate the PCA model, the user need to define which plots will be show in the Plot GUI. There are three option of plots: **Scores** , **Loadings** and **Residual**



The tab **Score/Loadings/Residuals Result** displays a visualization of the PCA model's calculated scores. The user may modify the scores on the X and Y axes, as well as plot the 3D graph of these scores, under the **Plot scores options** panel.

The **Processing panel** on the left allows the user to choose the preprocessing technique for the dataset's columns (e.g., Meancenter, Autoscale, or Pareto), as well as the number of latent variables, confidence value, and variable selection method. It is also able to toggle on and off the score plot features (i.e., Labels, Legend, and Class Ellipse) in the **Visual Options panel** .



Outliers are commonly identified using Hotelling's Residuals Q and T 2. The T 2 statistic is a measure of variation in the PCA model, but the Q statistic is a measure of the amount of variation that the PCA model does not capture, as seen in its residual matrix E (MUJICA et al., 2011). The Mahalanobis distance defines the T 2 statistic, while the Euclidean distance defines the Q statistic (KOURTI; MACGREGOR, 1995; QIN, 2003).

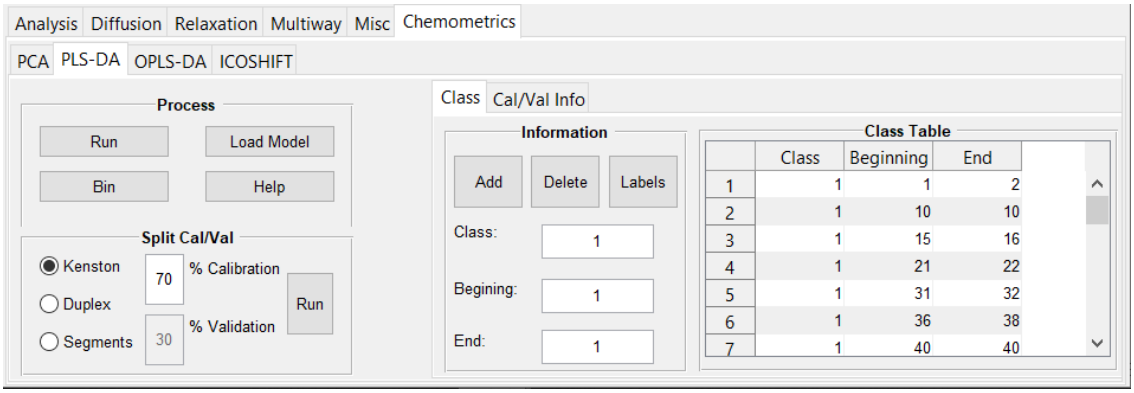
The Q statistic quantifies a sample's orthogonal projection to the space provided by the PCA model. 2015; HARROU et al. In other words, the output matrix from this calculation may be viewed as a measure of how effectively the sample is described.

Hotelling's T2 may be defined in the PCA model space by the Mahalanobis distance. The Mahalanobis distance describes the variance in the sample distribution for distinct data projection planes, taking into account their relevance for the model. This allows us to confirm that the sample distribution distance in some directions is greater than the distance in others.

PLS-DA

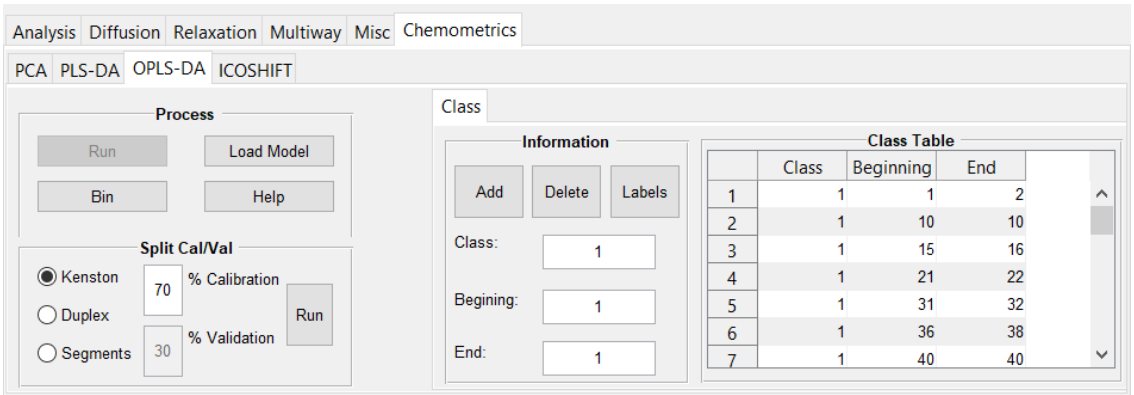
PLS-DA is a regression method that uses a matrix \(X_{(ij)}\) as a predictor and a matrix \(Y_{(ik)}\) with dummy variables as the answer. The dummy matrix \(Y\) contains categorical variables (i.e. 0 or 1). The model generated produces a \(\hat{Y}\) matrix for discriminating purposes.The discriminating rule compares anticipated response values from Y^ to a predefined threshold (e.g., 0.5) or calculated using the relationship between the sensitivity and specificity calculated for the model.

The main tab for PLS-DA computation within GNAT is shown bellow:



Note

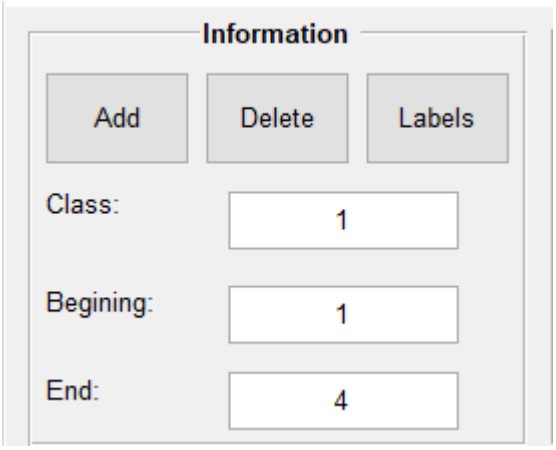
It is important to note that all of the methods used for creating a PLS-DA model may also be used for OPLS-DA models, as shown bellow.



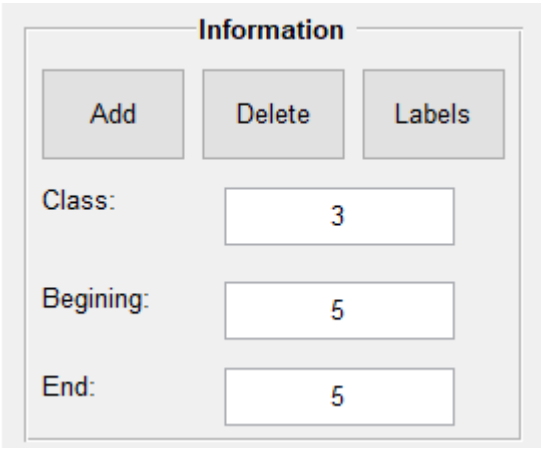
Class GUI

PLS-DA is a supervised approach, hence one class must be established for each sample in the dataset loaded into GNAT. The user can utilize the **Class** tab to build this array. The user can use the button `Add` after defining the **Class** , **Beginning** and **End** for each sample.

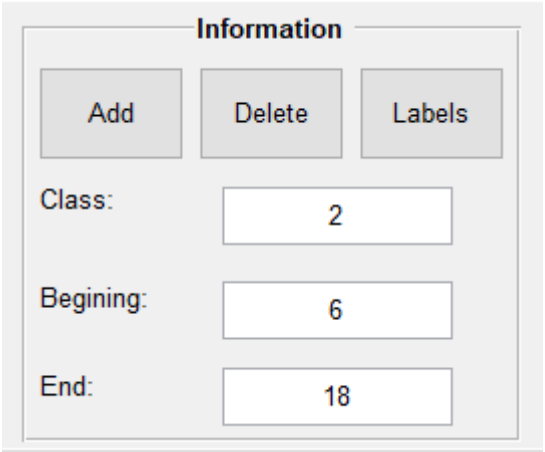
Case 1 : The first four samples belong to the class 1, so the following parameters should be used in the tab:



Case 2 : Sample 5 belongs to class 3:



Case 3 : Samples between 6 and 18 belong to class 2:



All the samples need to be associated to a numerical class. Labels to each class can be defined later using the Class GUI. All samples that are not associated to a label with have the number of their class as their label. It's also possible to exclude sample of the dataset imported by using the include menu.

Warning

All the inputs for each edit box need to be a number. Different error messages will appear for other type of imputs. But, when the last class added need to be deleted, the user can erase the value of the **Class** editbox and press delete to perform this

After defining the classes, the user can divide the imported dataset into a calibration and validation set of samples using the Split Cal/Val panel. There are three algorithms for this division:

Split methods

Kennard-Stone

The Kennard-Stone method selects a subset of samples from x which provide uniform coverage over the data set and includes samples on the boundary of the data set.

It begins by identifying the two samples with the greatest Euclidean distance (i.e the two samples farthest apart), then rating them as the most representative. In each subsequent phase, the remaining samples with the largest distance from the previously selected samples are picked and

appended to the bottom of the previous rank list. This technique is continued until a set number of samples have been selected and rated.

As defined in GNAT, this division is made by 70 % of samples been selected to the calibration set and 30 % in the validation set.

Reference 1. R. W. Kennard & L. A. Stone (1969): Computer Aided Design of Experiments, Technometrics, 11:1, 137-148.

Duplex

The Duplex method is similar to the Kennard-Stone algorithm, but it allows for the selection of separate calibration and validation sites.

The method begins by picking the pair of points that are the farthest apart. They are assigned to calibration sets and deleted from the list of points. The same procedure is repeated to find a pair of samples for the test set. Then, the algorithm iterates over the remaining samples to locate the sample farthest from the samples in the calibration set, followed by the sample farthest from the test set and assigning it to their respective sets. This is repeated until the desired number of samples in the calibration set is met.

Reference 1. R.D. Snee, Validation of regression models: methods and examples, Technometrics 19 (1977) 415-428 2. M. Daszykowski, B. Walczak, D.L. Massart, Representative subset selection, Analytica Chimica Acta 468 (2002) 91-103

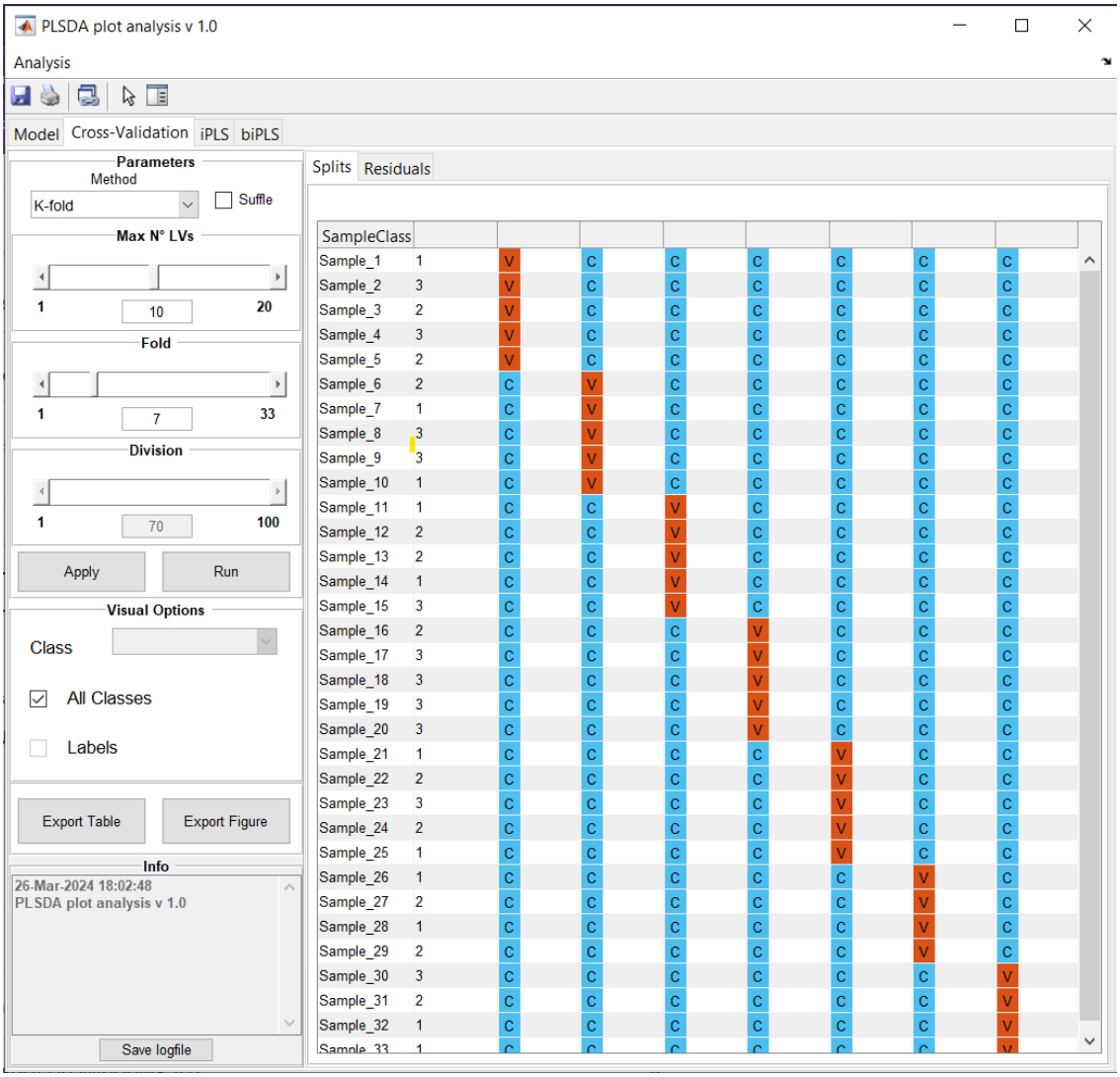
Segments

The dataset is divided continuously between calibration and validation based on the percentages specified in the panel. This method is only recommended for cases when the dataset imported into GNAT presents the samples in a random order based on the classes of each sample. Otherwise, it is recommended to use the Duplex or Kennard-Stone method.

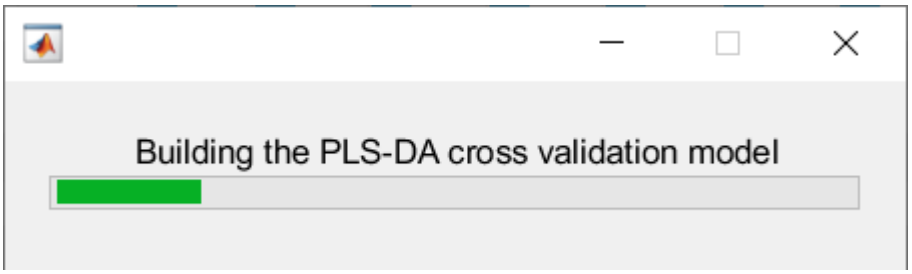
Cross-Validation

After dividing the dataset into calibration and validation, the user may hit the run button to begin calculating the PLS-DA model. The Cross-Validation tab has a settings panel on the right to define the “Method”, number of “Latent Variable” and number of “Folds” for the CV calculation. The “Division” parameter is available when the method “Mont-Carlo” is selected

After pressing `Apply` is presented a visualization of the split in calibration and validation set for each fold.



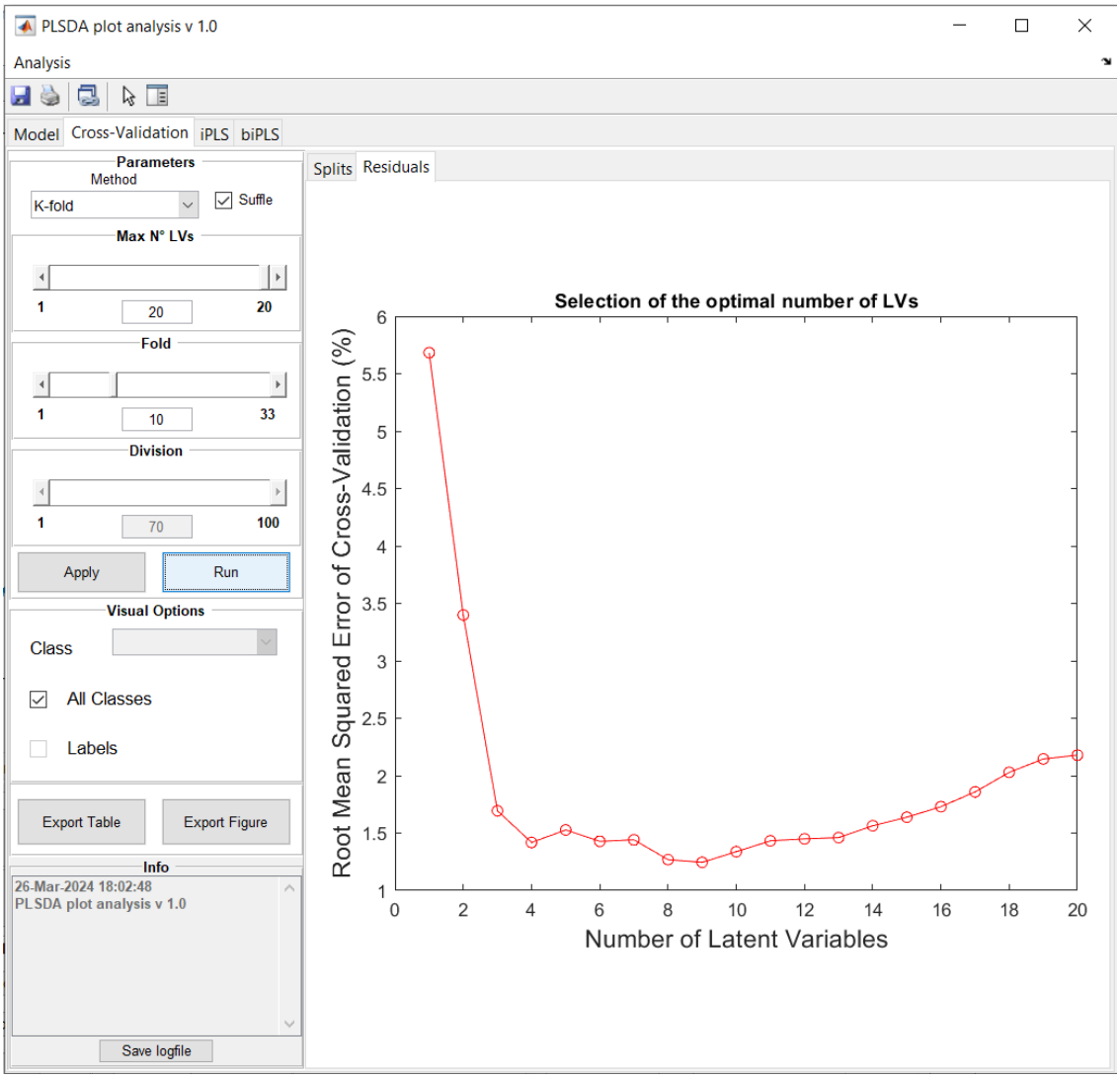
The `Shuffle` checkbox will select at random the samples for the validation set in each folder. This method is recommended when the samples are ordered. After pressing `Run` the Cross-Validation method is calculated:



RMSECV

A typical cross-validation strategy frequently includes many sub-validation trials that each involve choosing different subsets of data for model creation and testing. The ideal number of PLS components can be visualized pn the tab “residual”, in which is possible to visualize the Root Mean Square Error of Cross Validation (RMSECV).

The RMSECV calculation is used graphically to show how many latent variables are needed for your PLS-DA model. The number of latent variables retained in the model should, in principle, grow as the the residuals of the calibration data decreases.



iPLS

The methods for variable selection are presented in the tabs iPLS and biPLS. The implementation of the iPLS algorithm was based of the work developed for the iToolbox for Nørgaard L, 2001. In a nutshell, the iPLS method divides the spectrum into multiple equidistant regions and analyzes the value of the calibration prediction error for that interval for different values of Latent Variables, demonstrating to the user what the ideal value of latent variables is for the specific interval.

To calculate iPLS, the user must specify two parameters: the number of `intervals` and the number of `segments` .

1. `Intervals` : The number of subdivisions in the NMR spectra. The division is made up of many bins of identical size. If you have previously used this preprocessing, you may import these restrictions into the Binning GUI.

2. `Segments` : The division mechanism used to calculate crossvalidation for the iPLS model. It operates similarly to the Segments method which divides the dataset continually between calibration and validation based on the number of samples given in the panel.

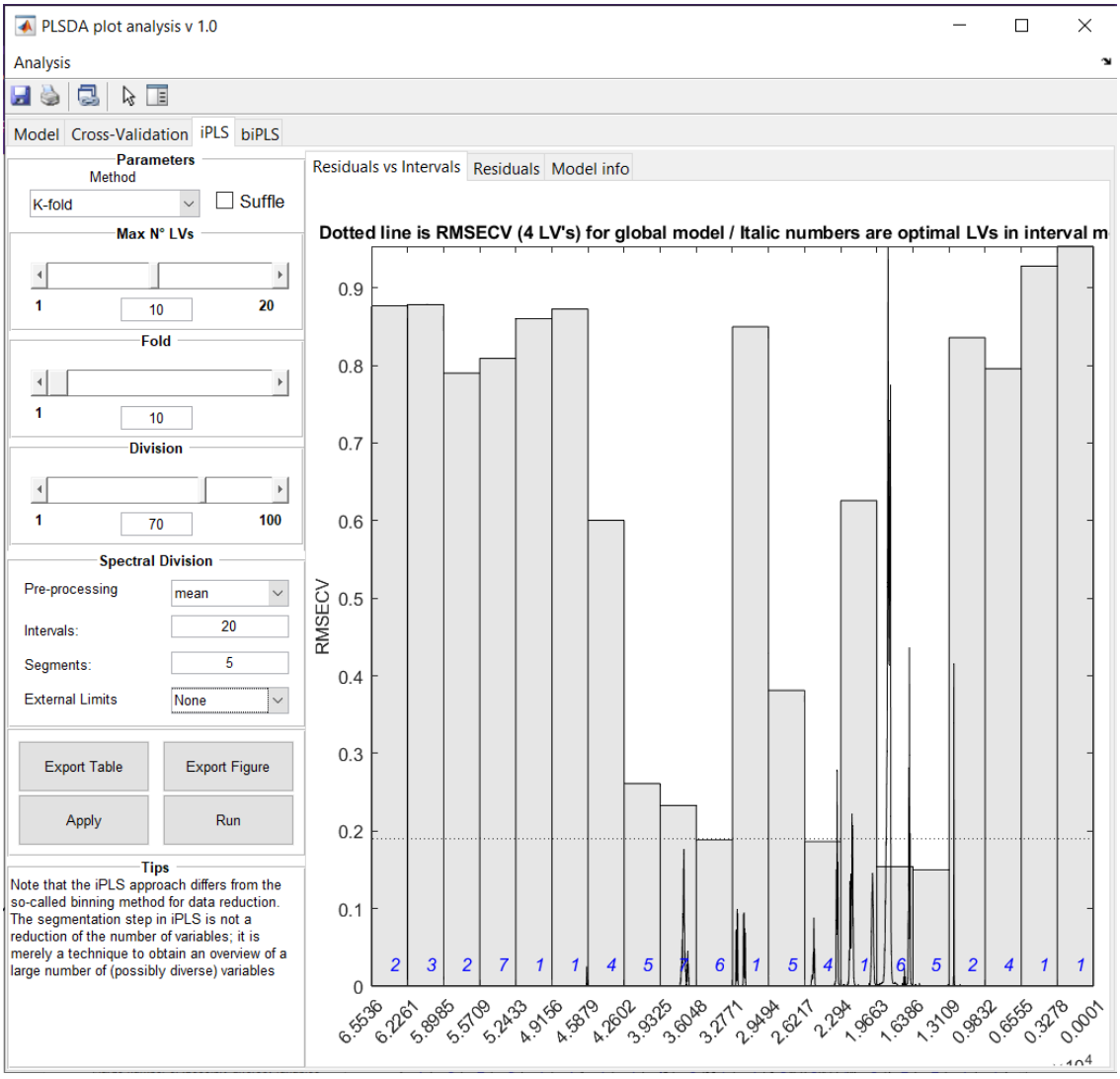
The forward approach works as follows:

1. Split spectral data into N intervals

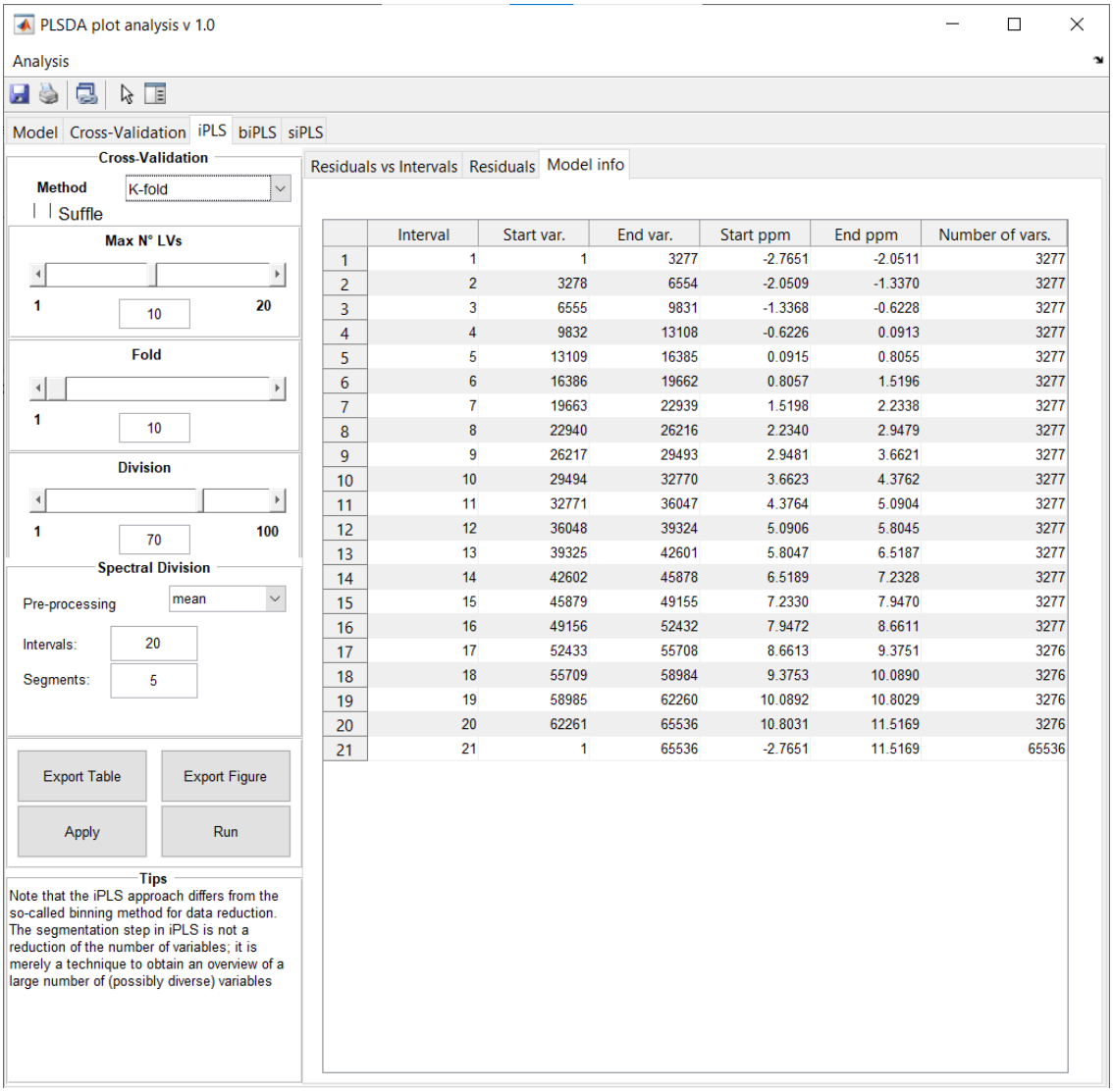
2. Create an empty vector with selected intervals

3. Create a model where intervals in the vector (already selected) are combined with one of the rest. If combination improves the model, add this new interval to the vector. 4. Repeat previous step until there is no improvements.

The iPLS model will select only a single variable for the construction of the PLS-DA model. The variable selected is the one with the lowest RMSECV value for this purpose, in this case, for the example presented in the next figure, variable number 5.



The limits for each interval, as well as the number of points, are displayed on the **model info** tab.



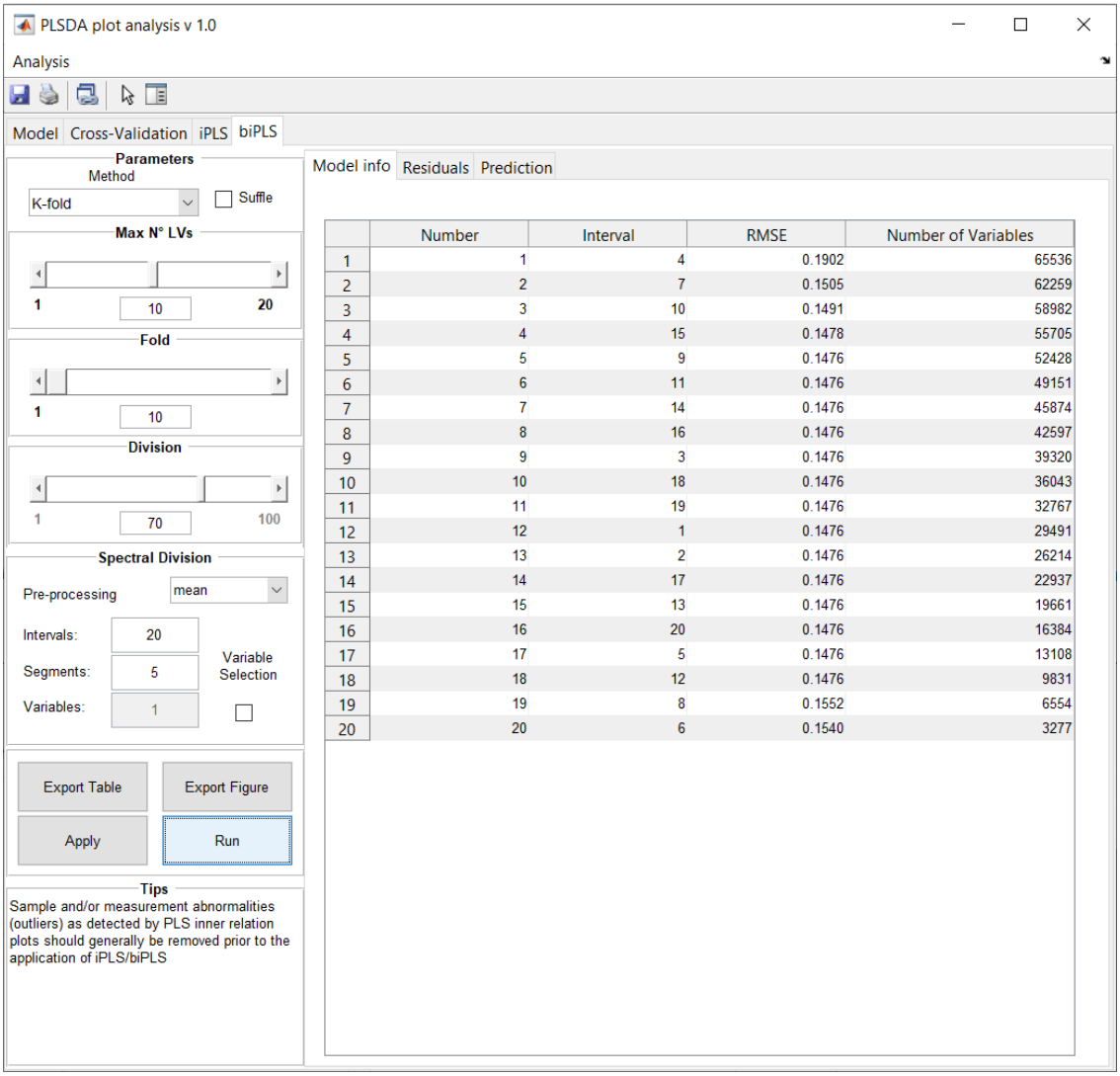
	Interval	Start var.	End var.	Start ppm	End ppm	Number of vars.
1	1	1	3277	-2.7651	-2.0511	3277
2	2	3278	6554	-2.0509	-1.3370	3277
3	3	6555	9831	-1.3368	-0.6228	3277
4	4	9832	13108	-0.6226	0.0913	3277
5	5	13109	16385	0.0915	0.8055	3277
6	6	16386	19662	0.8057	1.5196	3277
7	7	19663	22939	1.5198	2.2338	3277
8	8	22940	26216	2.2340	2.9479	3277
9	9	26217	29493	2.9481	3.6621	3277
10	10	29494	32770	3.6623	4.3762	3277
11	11	32771	36047	4.3764	5.0904	3277
12	12	36048	39324	5.0906	5.8045	3277
13	13	39325	42601	5.8047	6.5187	3277
14	14	42602	45878	6.5189	7.2328	3277
15	15	45879	49155	7.2330	7.9470	3277
16	16	49156	52432	7.9472	8.6611	3277
17	17	52433	55708	8.6613	9.3751	3276
18	18	55709	58984	9.3753	10.0890	3276
19	19	58985	62260	10.0892	10.8029	3276
20	20	62261	65536	10.8031	11.5169	3276
21	21	1	65536	-2.7651	11.5169	65536

biPLS

Backward interval partial least squares (biPLS) is a variable selection approach that is primarily used to decrease the PLS model's variables and reduce the number of sub-intervals by analysis RMSECV of multiple intervals every new run. When we build the model using biPLS functions, it is possible to determine multiple relevant variables for better class separation by PLS-DA models.

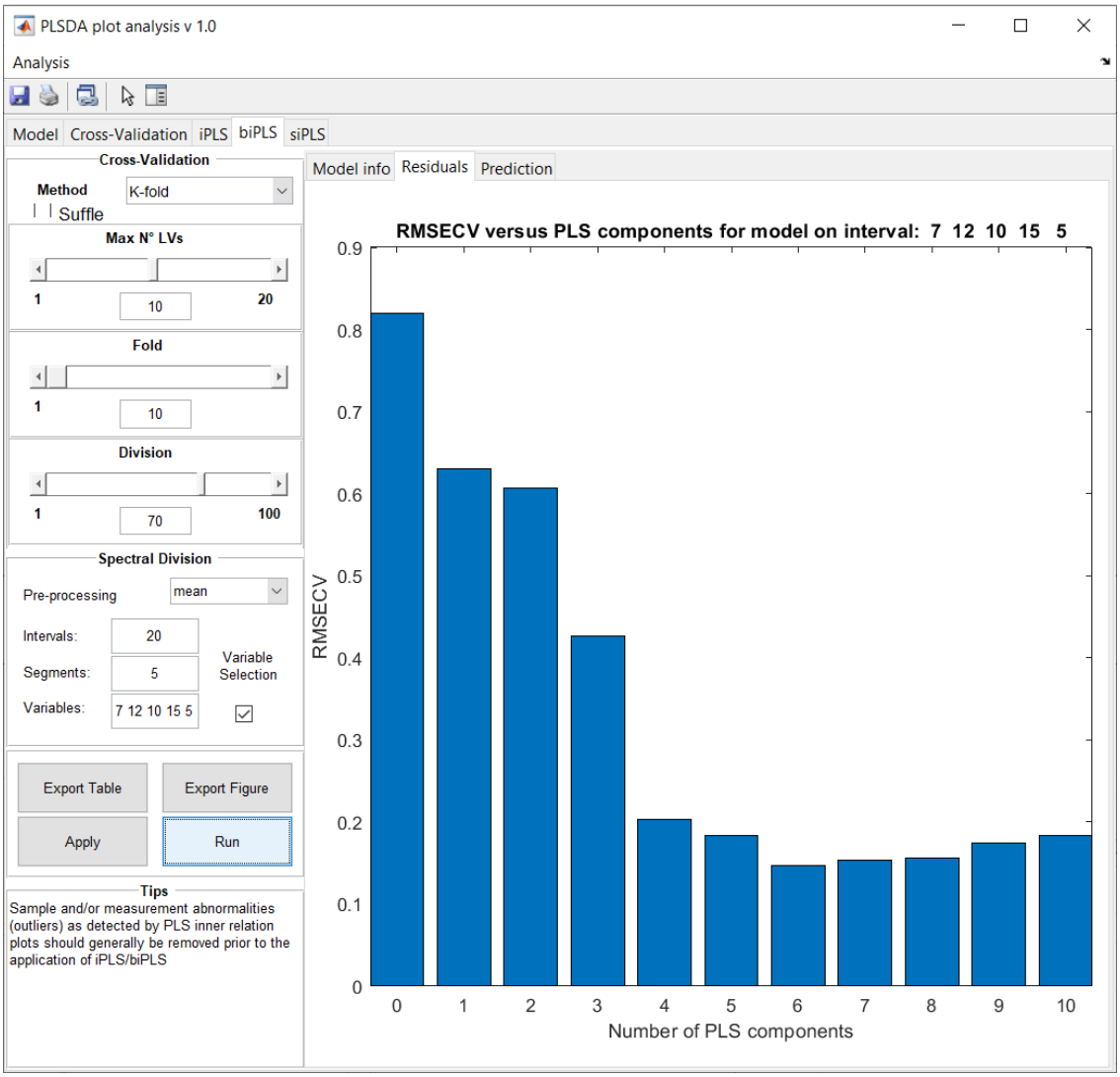
The method is calculated using the same parameters used for the "iPLS" model. After defining the intervals and segments, the user can press the button `Apply` and `Run` to start the analysis.

In the "Model info" tab table, the colum RMSE dictate which variables should be selected. The error is reduced until the interval 8, so the ideal variables should be 8 and 6. The model is recalculated using these variables after checking the edit box "Variable selection" and adding these variables in the edit box.

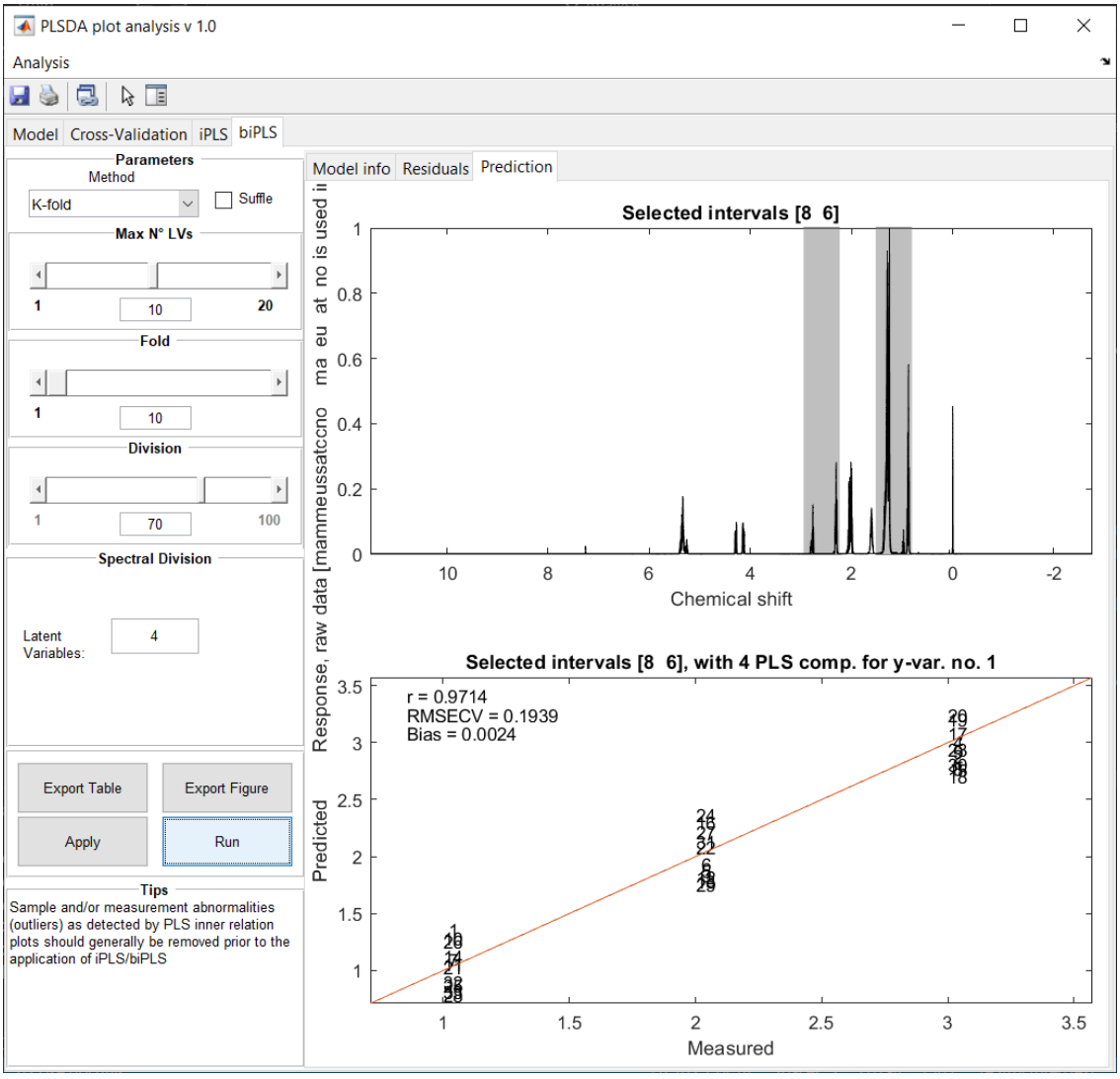


	Number	Interval	RMSE	Number of Variables
1	1	4	0.1902	65536
2	2	7	0.1505	62259
3	3	10	0.1491	58982
4	4	15	0.1478	55705
5	5	9	0.1476	52428
6	6	11	0.1476	49151
7	7	14	0.1476	45874
8	8	16	0.1476	42597
9	9	3	0.1476	39320
10	10	18	0.1476	36043
11	11	19	0.1476	32767
12	12	1	0.1476	29491
13	13	2	0.1476	26214
14	14	17	0.1476	22937
15	15	13	0.1476	19661
16	16	20	0.1476	16384
17	17	5	0.1476	13108
18	18	12	0.1476	9831
19	19	8	0.1552	6554
20	20	6	0.1540	3277

To continue the analysis, the user need to go to the tab “Residuals” and press `Apply` and `Run` again. The RMSECV graph will be plot, showing the ideal number of LVs for this model. In the example presented, this value of LV can be between 4 and 6.



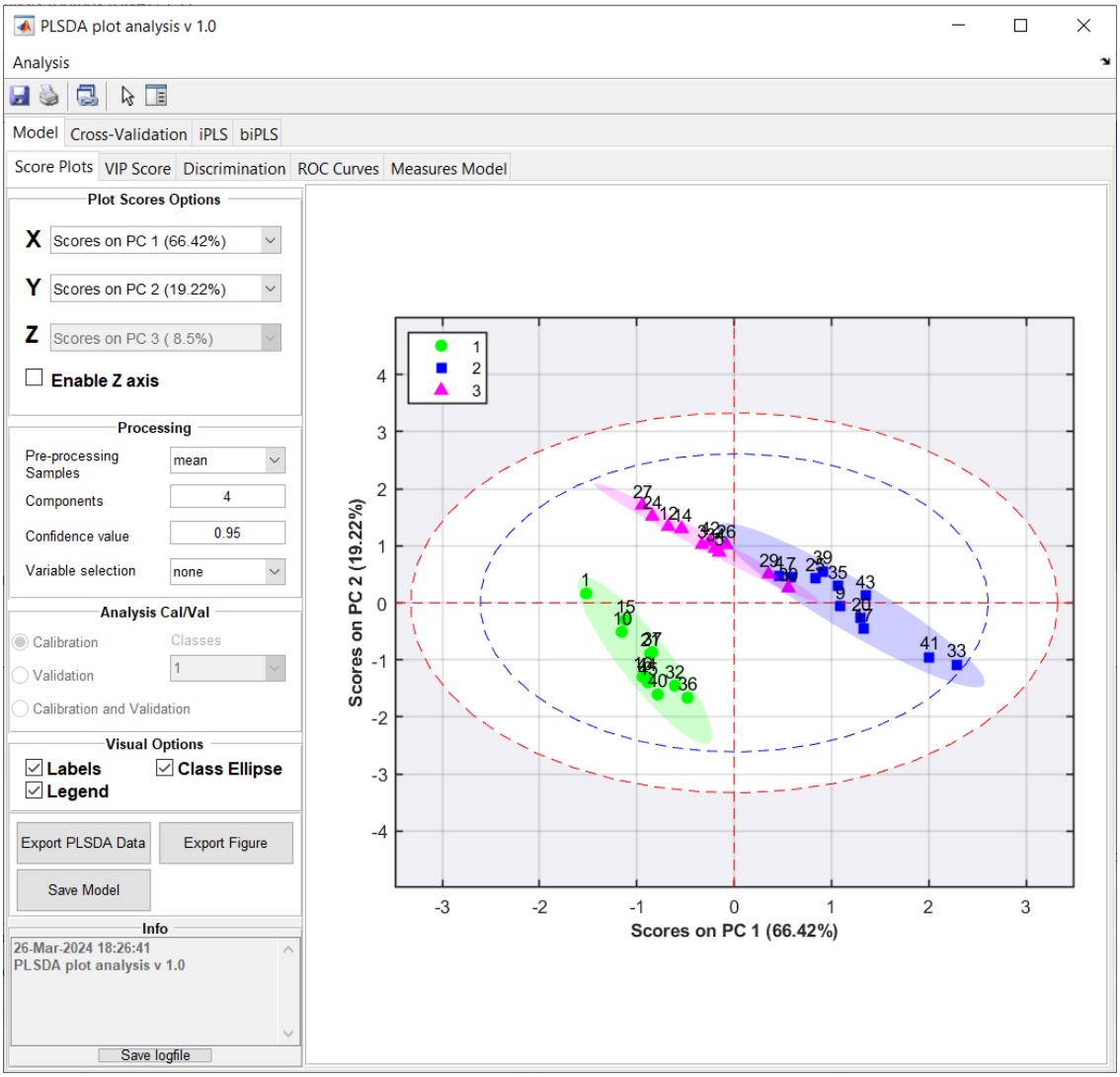
Finally, the user can go to the **Predict tab** , check the Latent variables check box and pu the ideal value for the LV. After pressing `Apply` and `Run` the user can see the variables used in the model.



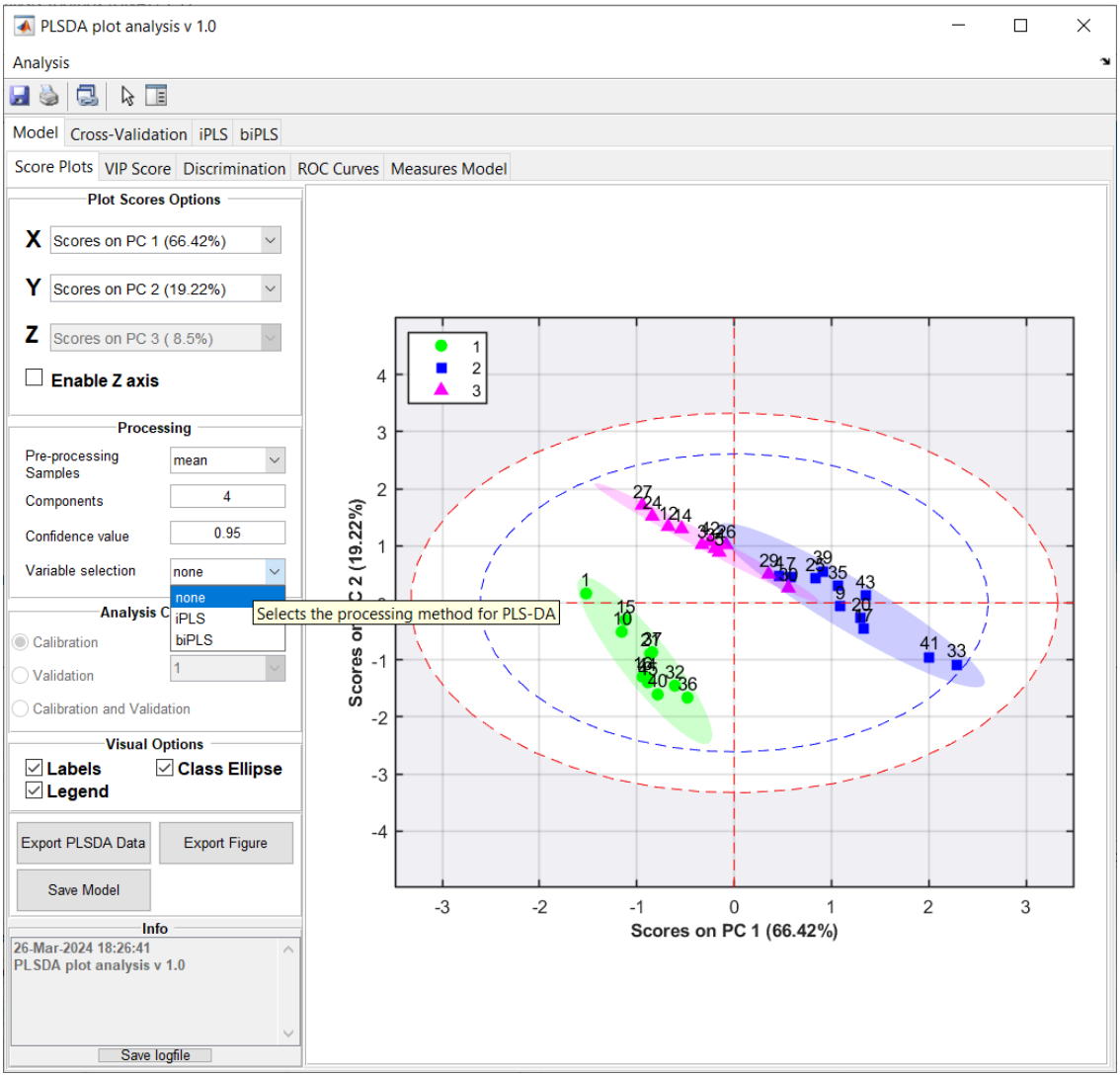
Scores Plot

The tab **Model** displays a visualization of the PLS-DA model's calculated scores. The user may modify the scores on the x and y axes, as well as plot the 3D graph of these scores, under the **Plot scores options** panel.

The figure below presents a PLS-DA calculation results for the 1H NMR spectra dataset of three edible oils – Olive oil (●), Rapeseed oil (■) and Sunflower oil (▲). The blue ellipse (–) represents the confidence limit for 95 % of confidence and the red ellipse (–) for 99 %.

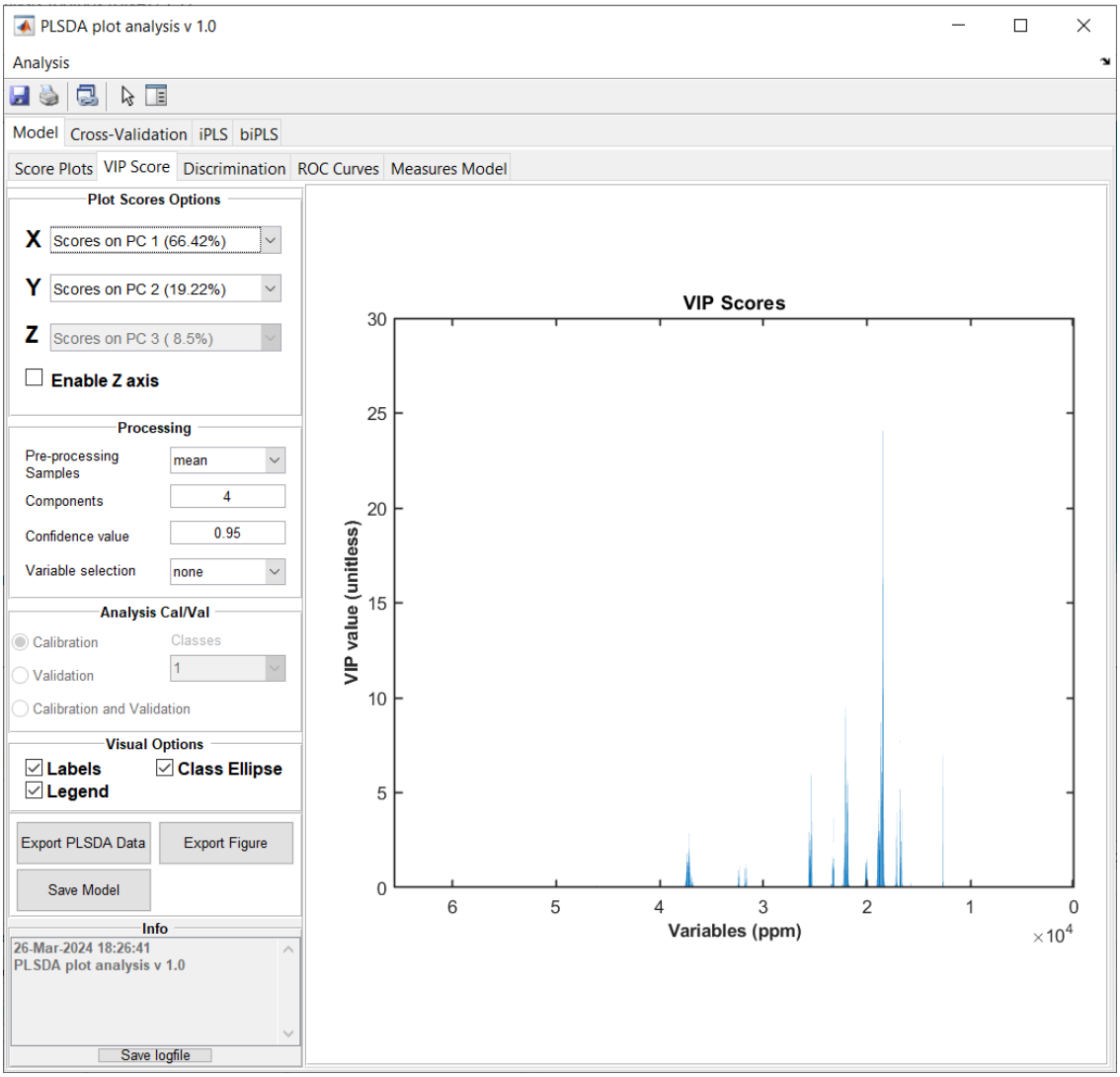


The **Processing panel** on the left allows the user to choose the preprocessing technique for the dataset's columns (e.g., Meancenter, Autoscale, or Pareto), as well as the number of latent variables, confidence value, and variable selection method. It is also able to toggle on and off the score plot features (i.e., Labels, Legend, and Class Ellipse) in the **Visual Options panel** .



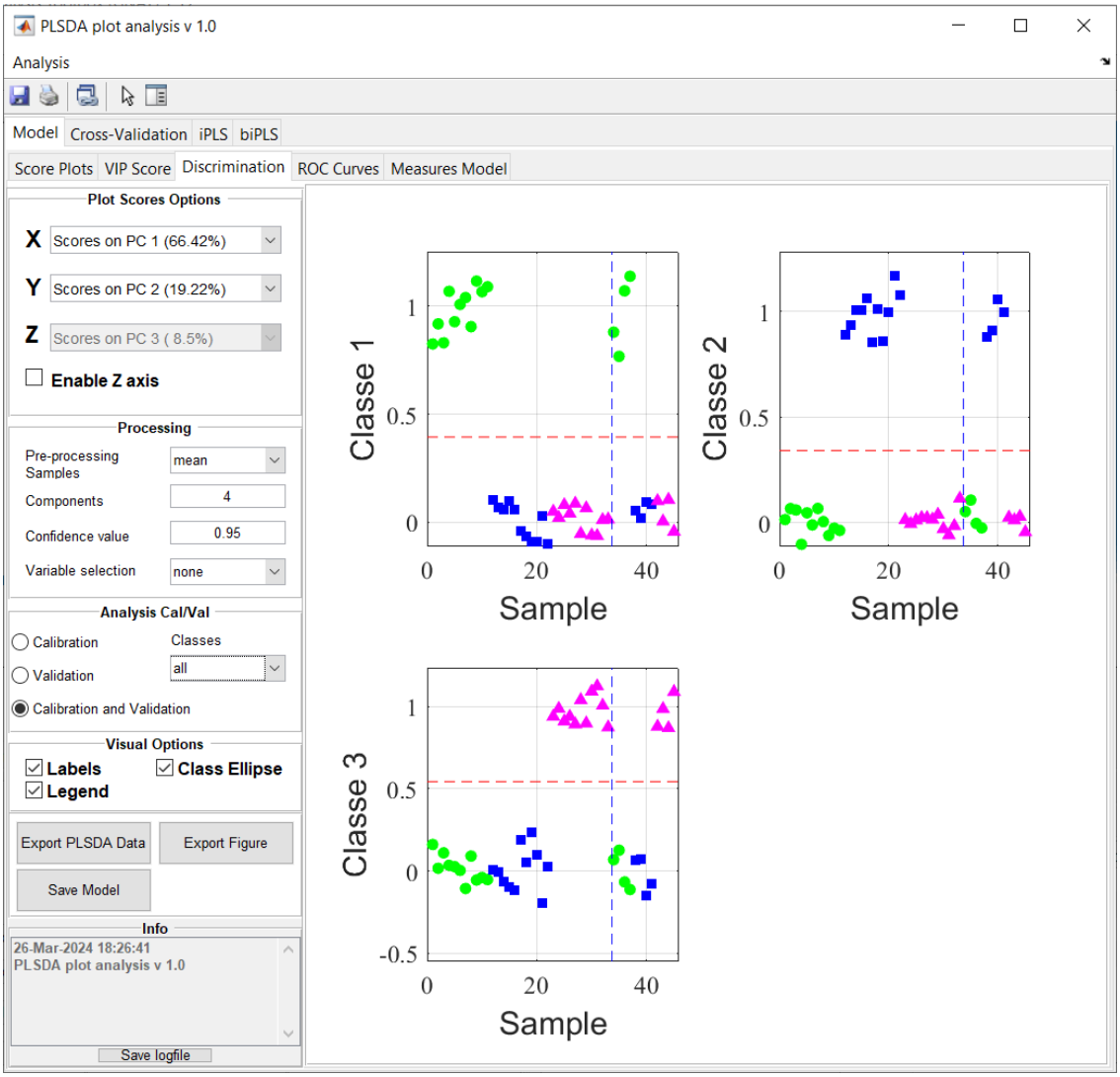
VIP

In PLS-DA and OPLS-DA models, the variable importance in projection (VIP) value is utilized to evaluate the relevance of each variable and choose biomarkers. A variable with a VIP Score near to or more than one (one) might be considered significant in a particular model. The Y-axis shows the VIP scores for each variable on the X-axis



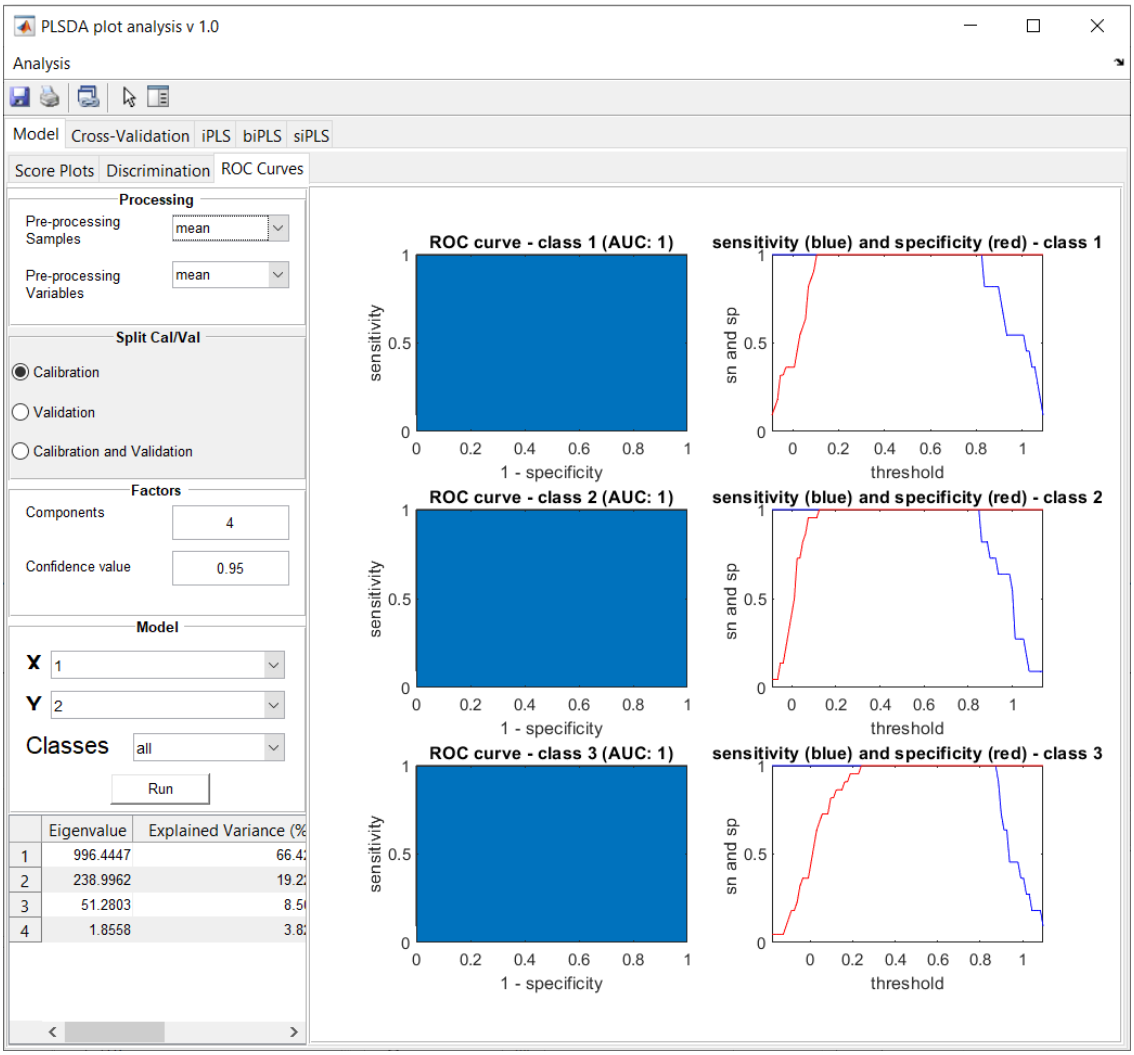
Discrimination plot

Still within the Model tab, it is possible to view the class discrimination mode of the PLS-DA model for all classes (both calibration and validation in the Analysis Cal/Val panel) in the **Discrimination tab** . By default, this model was calculated with 4 LV (Latent Variables) and meancentered.



ROC curves

ROC curves may be used to show the specificities and sensitivities that can be achieved with different projected y-value thresholds in a PLSDA model. The **Analysis Cal/Val** panel menu **Classes** allows the user to choose which ROC they want to visualize.



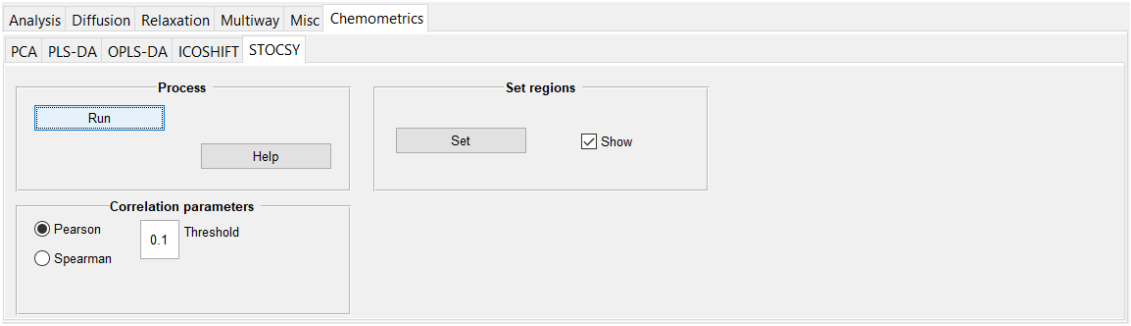
OPLS-DA

All the information to calculate an OPLS-DA model can be found in the PLS-DA page

STOCSY

STOCSY (Statistical Total Correlation Spectroscopy) is a form of homonuclear NMR spectroscopy that reveals correlations among all nuclei in a spin system. This approach uses correlations between the intensity of spectral components in numerous spectra to get a statistically generated spectrum.

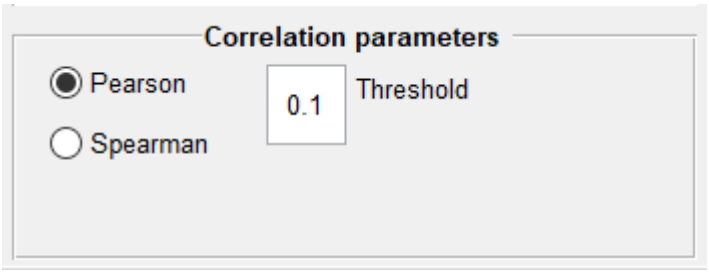
The main tab for STOCSY computation within GNAT is shown bellow:



Note

It is important to note that the method is restricted to 1D analysis

Defining correlation parameters



Threshold Correlation threshold p-value for testing the hypothesis of no correlation (by default 0.1)

Correlation method The selection of correlation coefficient measure {'pearson' or 'spearman'}

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

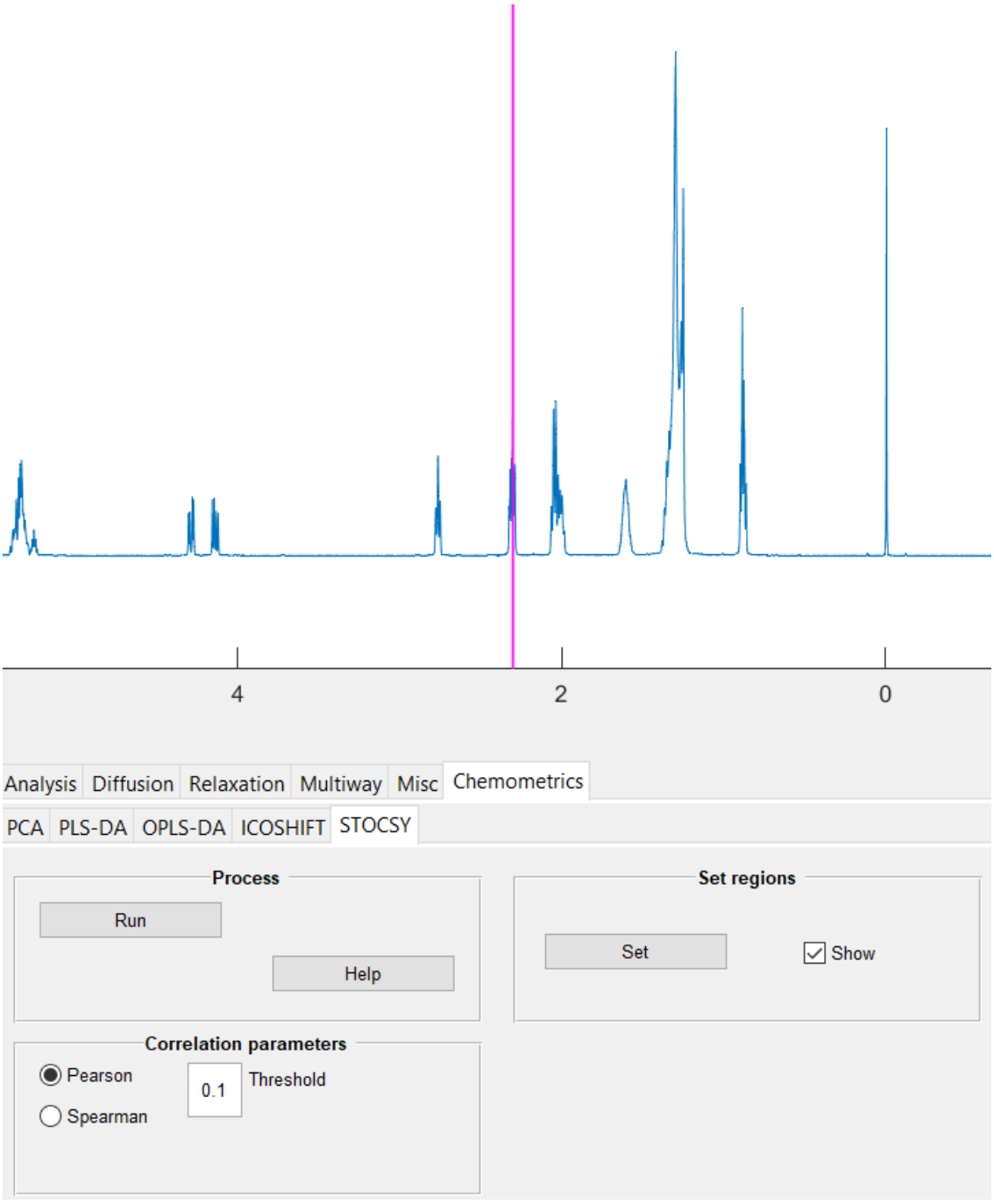
r = correlation coefficient
x_i = values of the x-variable in a sample
$\bar{x}$ = mean of the values of the x-variable
y_i = values of the y-variable in a sample
$\bar{y}$ = mean of the values of the y-variable

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

ρ = Spearman's rank correlation coefficient
d_i = difference between the two ranks of each observation
n = number of observations

Defining ppm region

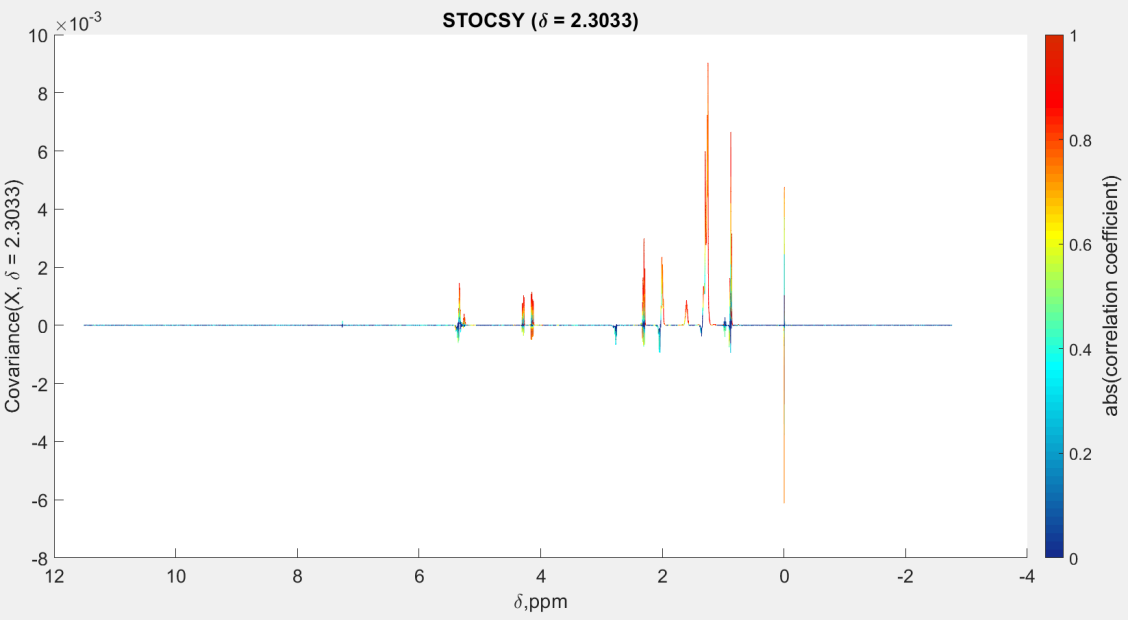
In STOCSY 1D the user can utilize the **Set region** panel to select the signal in the active spectrum in GNAT to be use in the analysis. The user can use the button `Set` after selecting the checkbox `Show`



Reference 1. R. W. Kennard & L. A. Stone (1969): Computer Aided Design of Experiments, Technometrics, 11:1, 137-148.

STOCSY calculation

After selecting the signal, the user may hit the `Run` button to begin calculating the STOCSY model. A new figure with the 1D analysis will show up



Contact & Credits

References

Indices and tables

- Index
- Module Index
- Search Page