GNAT manual

Release v2.2

MENUS

1 Introduction **1**

2 Start here **3**

3 Documentation status **5**

4 Citations **7**

5 GNAT interface overview **9**

6 Reference guide **11**

6.1 File . . . 12
6.1.1 Functionalities . . . 12
6.2 Edit . . . 17
6.2.1 Functionalities . . . 17
6.3 Help . . . 24
6.3.1 Functionalities . . . 24
6.4 Plot . . . 24
6.4.1 Scale . . . 24
6.4.2 Display . . . 24
6.4.3 Zoom/Pan . . . 25
6.4.4 Separate plot . . . 25
6.4.5 Displayed range . . . 27
6.4.6 Horizontal scale . . . 27
6.5 Phase . . . 27
6.6 FT . . . 28
6.6.1 Fourier transform section . . . 28
6.6.2 Window function section . . . 29
6.6.3 Reference section . . . 29
6.7 Correct . . . 29
6.7.1 Baseline correction . . . 29
6.7.2 Reference deconvolution . . . 33
6.8 Array . . . 35
6.9 Prune . . . 36
6.9.1 Exclude regions . . . 36
6.9.2 Prune arrays . . . 37
6.10 Pure Shift . . . 37
6.10.1 Process section . . . 37
6.10.2 Parameters section . . . 37

6.10.3 More information . . . 37
6.11 Linear prediction . . . 39
6.12 Misc . . . 39
6.12.1 Play FID section . . . 39
6.12.2 Data state section . . . 39
6.13 Info . . . 39
6.14 Analysis . . . 40
6.14.1 Functionalities . . . 40
6.15 Diffusion . . . 56
6.15.1 Functionalities . . . 56
6.16 Relaxation . . . 86
6.16.1 Functionalities . . . 86
6.17 Multiway . . . 87
6.17.1 Functionalities . . . 87
6.18 Misc . . . 90
6.18.1 Functionalities . . . 90
6.19 Chemometrics . . . 92
6.19.1 Functionalities . . . 92
6.20 Neural Network . . . 119
6.20.1 Functionalities . . . 119

7 Indices and tables **125**

CHAPTER

ONE

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. GNAT supports import from major NMR platforms, standard 1D processing, and a range of analysis tools including DOSY, SCORE, ROSY, PARAFAC, and chemometric methods.

CHAPTER

TWO

START HERE

This manual is currently organised mainly by GUI tab. For a quick start, the most useful path is usually:

1. **File**: import your data.
2. **Phase**, **FT**, and **Correct**: perform basic preprocessing.
3. **Prune** and **Integrate** if needed.
4. Go to **Diffusion**, **Relaxation**, **Multiway**, or **Chemometrics** for the main analysis.

CHAPTER

THREE

DOCUMENTATION STATUS

This is a working manual. Much of the core functionality is documented, but a number of sections are still brief, incomplete, or platform-specific. Where relevant, pages are marked as **Not yet documented**, **Experimental**, **Platform-specific**, or **Work in progress**.

CHAPTER

FOUR

CITATIONS

If you are using GNAT, or the older DOSY Toolbox, please cite 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.

CHAPTER

FIVE

GNAT INTERFACE OVERVIEW

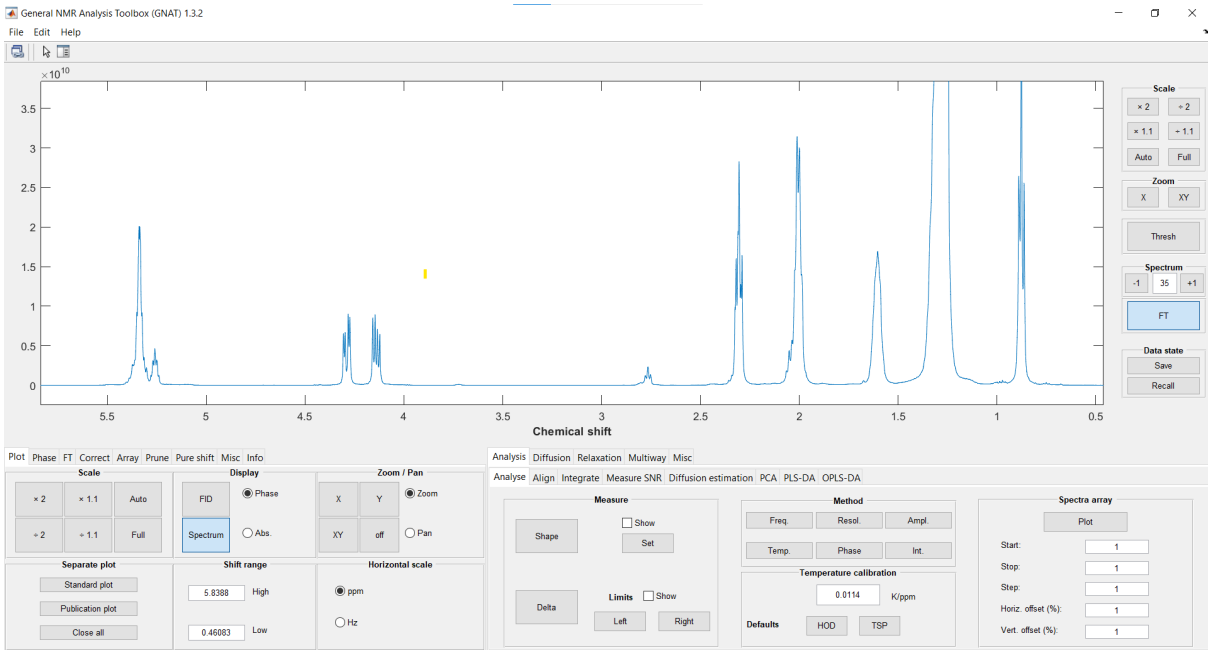


Fig. 1: GNAT main GUI.

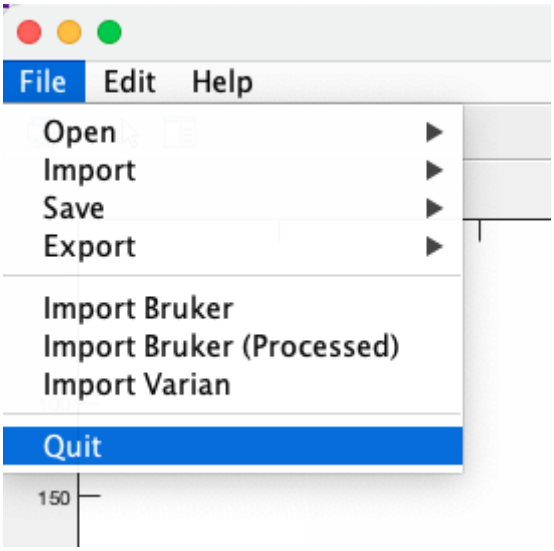
CHAPTER

SIX

REFERENCE GUIDE

6.1 File

The *File* menu is used to open, import, save, and export GNAT data.



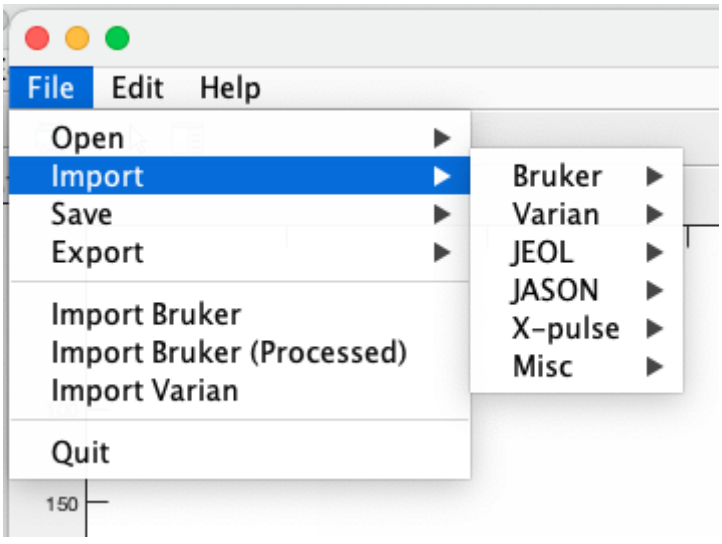
6.1.1 Functionalities

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

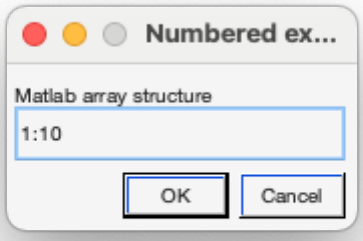
Data from Bruker can be imported in various ways. GNAT will do its best to determine which type of data it is (e.g. diffusion or relaxation encoded) and import the relevant parameters.

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, useful for monitoring a chemical reaction with DOSY experiments and analysing the resulting 3D structure 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 single 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

> **Note**
>
> JASON import is not yet documented in this manual.

X-pulse

> **Note**
>
> X-pulse import is not yet documented in this manual.

Misc

Here GNAT supports some miscellaneous data formats.

MATLAB structure

Here you can import any data that has the format of a MATLAB structure saved as a `*.mat` file. The structure must contain the fields `fid`, `sw`, `sfrq`, `ppmAxis`, and `nucleus`:

- `fid`: MATLAB array with dimensions `[np dim2 dim3 dim4]`. `dim3` and `dim4` may be omitted. A 1D spectrum will typically have dimensions `[np 1]`. An array of 12 spectra, such as a DOSY data set, will have `[np 12]`. A set of DOSY data sets could have `[np 12 4]`.
- `sw`: spectral width in Hz
- `sfrq`: spectrometer frequency in Hz
- `ppmAxis`: chemical shift axis in ppm
- `nucleus`: nucleus, for example `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

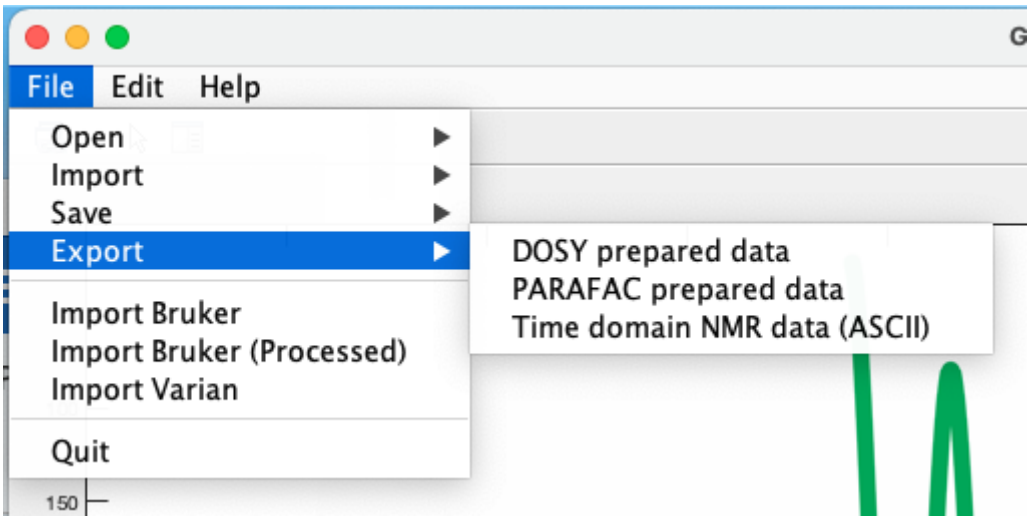
Data can be saved either as ASCII or binary. In both cases there will be a text header-file that describes all the relevant parameters. The ACSII version naturally takes up more disk space and is slower to save and load.

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.

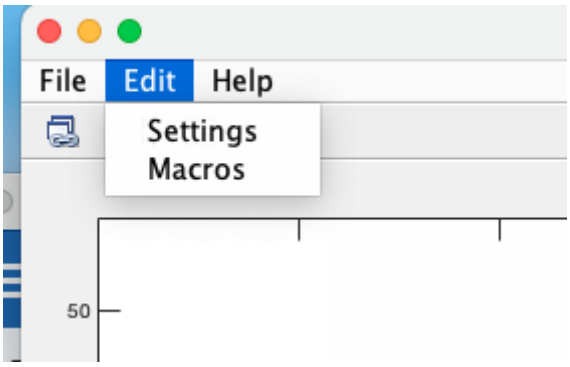
Reference (1) Andersson, C. A.; Bro, R. The N-way Toolbox for MATLAB. Chemometrics and Intelligent Laboratory Systems 2000, 52 (1), 1.

Time domain data

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

6.2 Edit

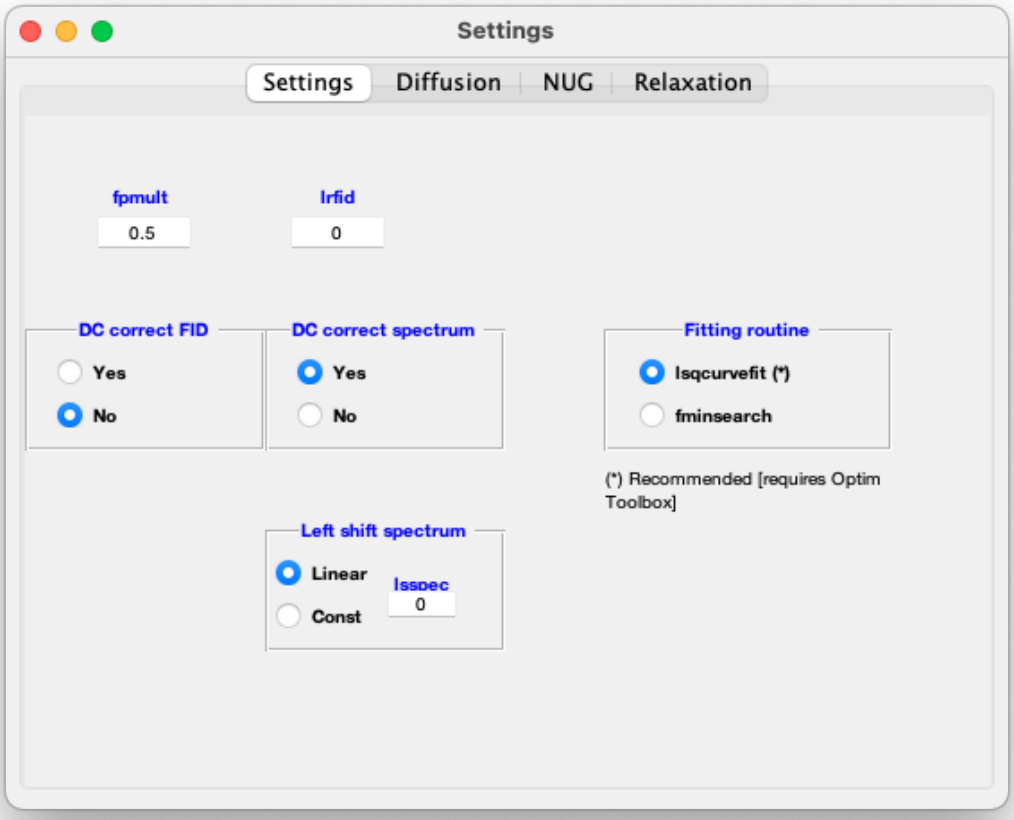
The *Edit* menu provides access to settings and macros.



6.2.1 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 (default 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:

$\delta^2 * \gamma^2 * \Delta'$

where Δ' is the diffusion time Δ corrected for diffusion during the gradient pulses and is pulse sequence specific, and δ is the diffusion encoding time, γ is the gyromagnetic ratio. The parameter τ is the difference between gradient pulses in a bipolar pulse pair (bpp) and is used in the calculation of Δ'

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

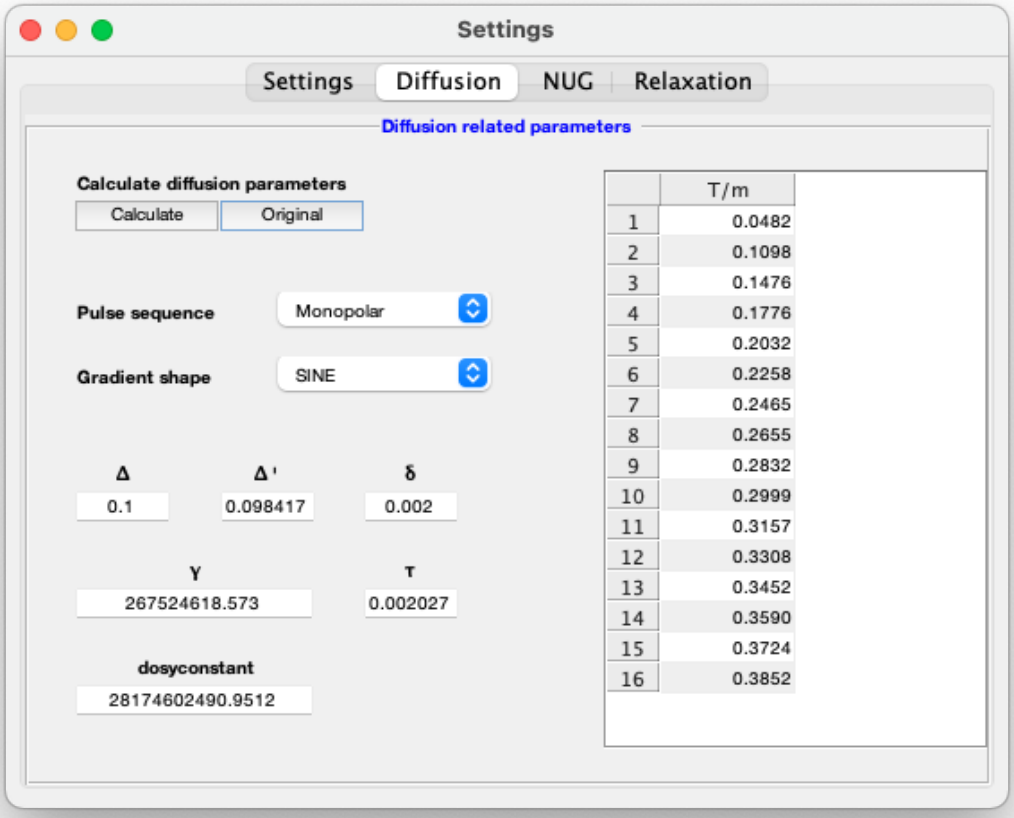
Pressing the *Calculate* button will calculate *dosyconstant* and Δ' with the given parameters.

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


Settings
Settings
Diffusion
NUG
Relaxation
Diffusion related parameters
Calculate diffusion parameters
Calculate
Original
Pulse sequence
Monopolar
Gradient shape
SINE
Δ
Δ'
δ
0.1
0.098417
0.002
γ
τ
267524618.573
0.002027
dosyconstant
28174602490.9512
T/m
1 0.0482
2 0.1098
3 0.1476
4 0.1776
5 0.2032
6 0.2258
7 0.2465
8 0.2655
9 0.2832
10 0.2999
11 0.3157
12 0.3308
13 0.3452
14 0.3590
15 0.3724
16 0.3852

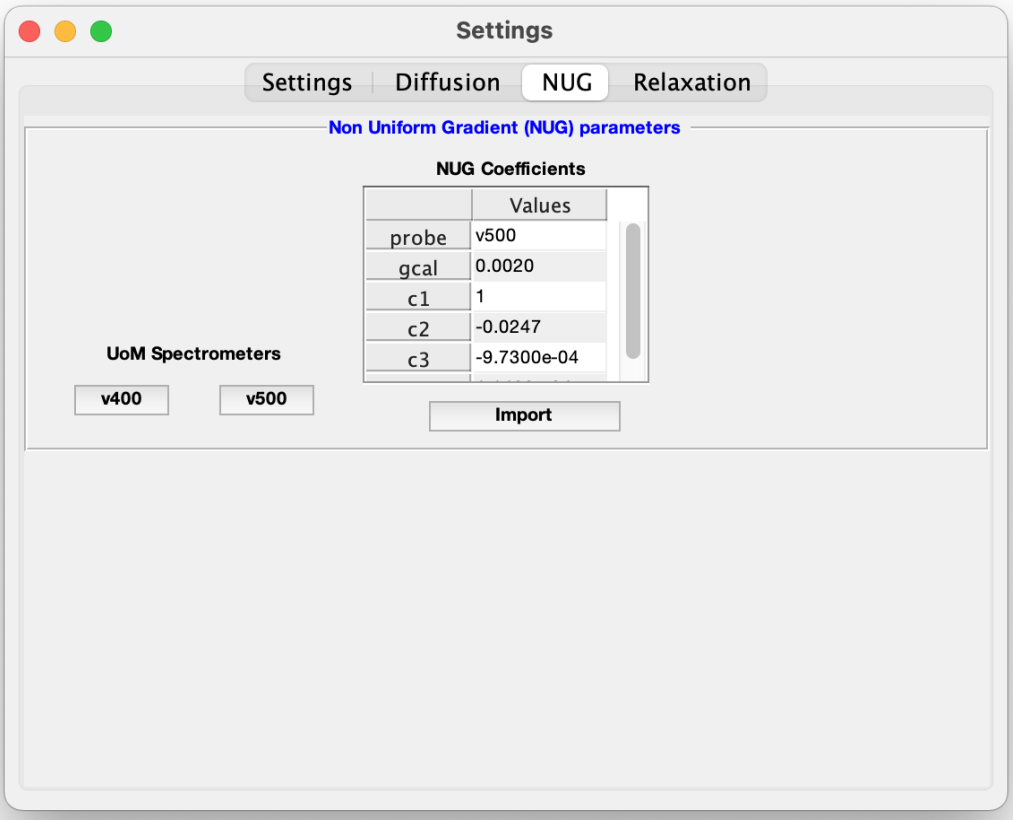
Choose the gradient shape used

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.

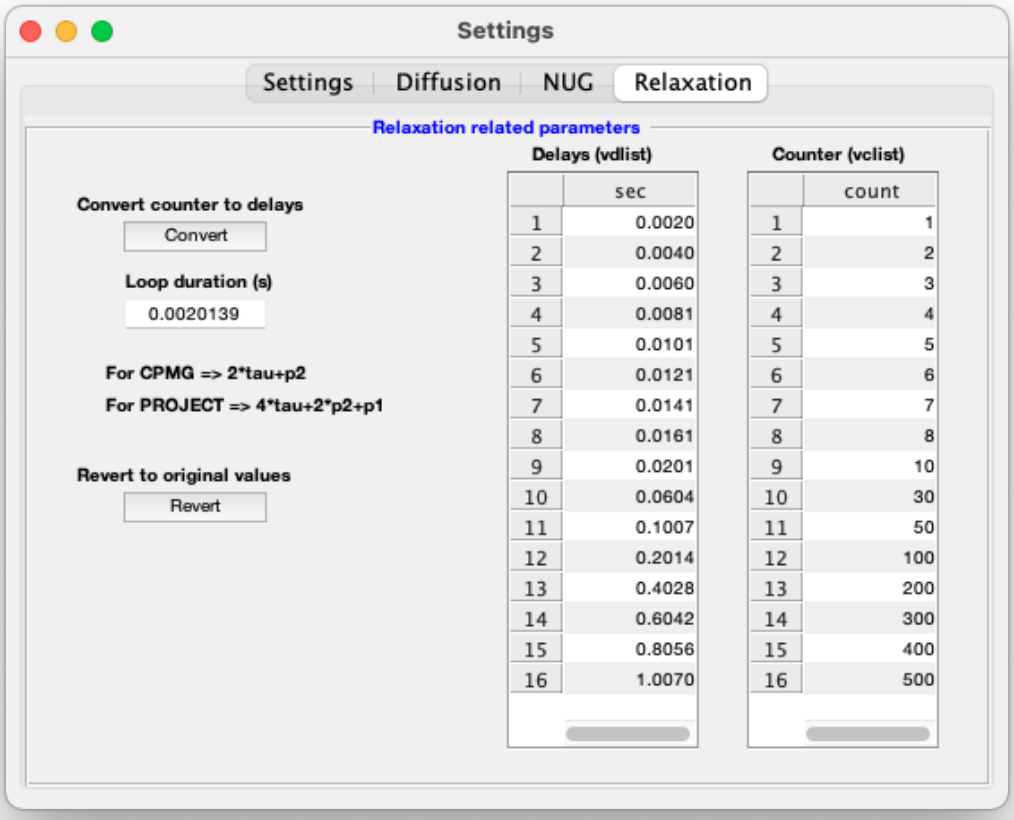
The choice of using NUG coefficients is given in the Diffusion processing module of GNAT.

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:



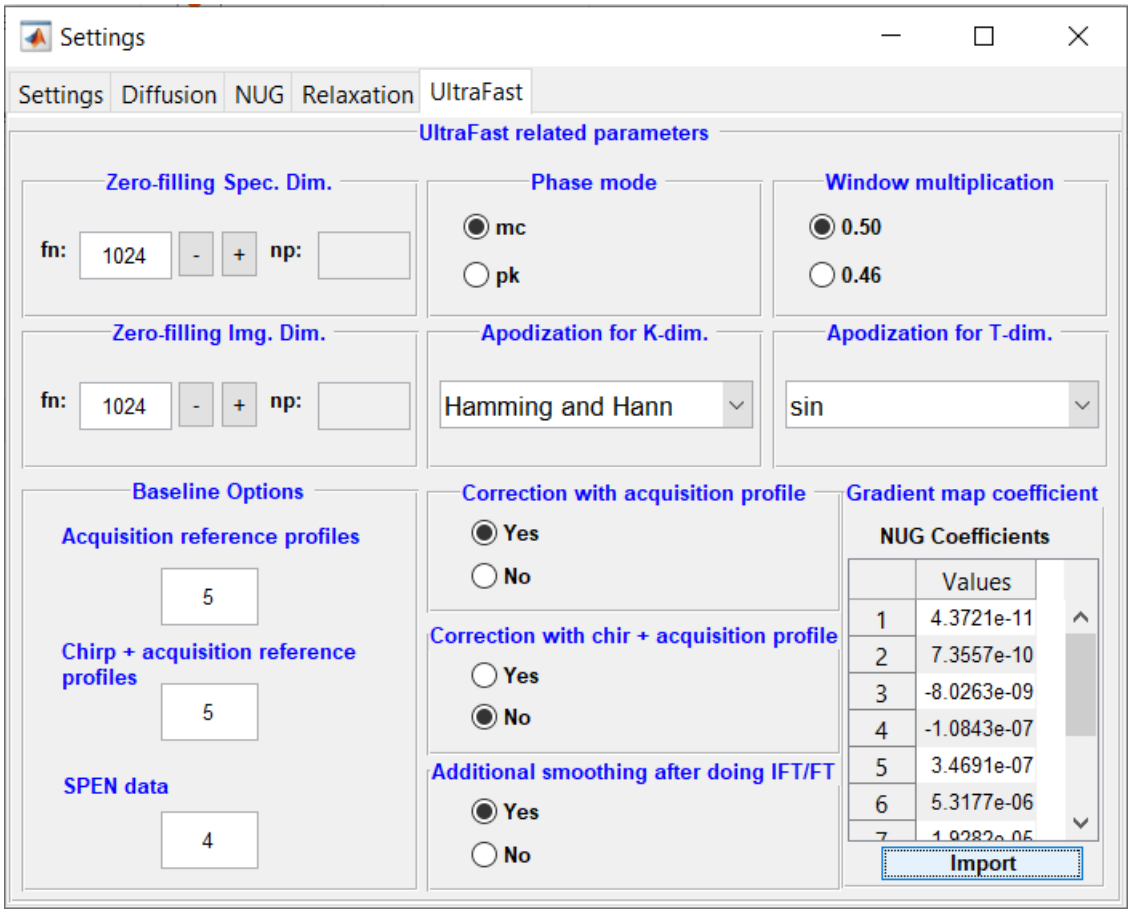
GNAT will attempt to import the correct delays for a relaxation (e.g. T1 or T2) experiment, and provides this interface to correct any mistakes.

If there is direct delay information e.g. in a *vdlist* (variable delay list) this will be usedd for the *tau* delays. If there is information about counters e.g. in a *vclist* this will be usedd 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.

Ultrafast

In this tab various parameters for import ultrafast NMR data is accessible:



Due to the need for special pre-processing, ultrafast data can have its import parameters modified in the settings section. Some of these parameters (e.g. zero filling, phase correction, and baseline correction) can only be modified during the import section.

When any ultrafast data is imported, the question "Do you want to modify the default parameters for importing UF data?" will appear. If the user wants to keep the default parameters, it is enough to press 'no' in the options.

- **Zero-filling Spec. Dim.**
  Zero filling in spectroscopic dimension
- **Zero-filling Img. Dim.**
  Zero filling in imaging dimension
- **Baseline options**
  - **Acquisition reference profile**
    Baseline correction parameter for acquisition reference profiles
  - **Chirp + acquisition reference profile**
    Baseline correction parameter for chirp+acquisition reference profiles
  - **SPEN data**
    Base line correction parameter for SPEN data

- **Phase mode**
  Phase mode of the data. For magnitude mode data: 'mc', and for phase mode data: 'pk'
- **Window multiplication**
  Window multiplication constant
- **Apodization for K-dim.**
  Apodization function for k-dimension

  For more information see: the file apodization_k_t.m and http://spindynamics.org/wiki/index.php?title=apodization.m
- **Apodization for T-dim.**
  Apodization function for t-dimension
- **Correction with acquisition profile**
  correction with acquisition profile
- **Correction with acquisition profile**
  correction with acquisition profile
- **Additional smoothing after doing IFT/FT**
  additional smoothing after doing IFT/FT during reference profile correction to remove unwanted wiggles from the final spectrum
- **The choice of using gradient map coefficients is made by importing these coefficients in the from a text file.**
  Here is an example of a text file: `Gradient map file format`

Macros

GNAT includes support for macros from the *Edit* menu.

> **Note**
>
> Macro functionality is not yet documented in detail in this manual.

6.3 Help

The *Help* menu contains brief information about GNAT.

6.3.1 Functionalities

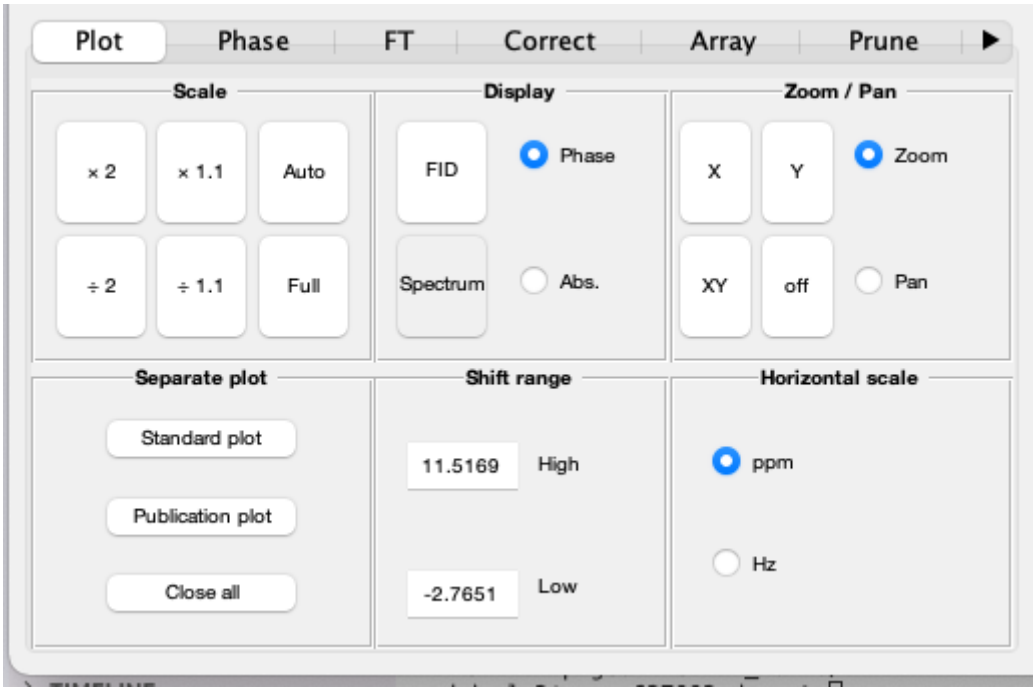
About

This page is a placeholder for general information about GNAT, versioning, and contact details.

> **Note**
>
> This section has not yet been fully documented.

6.4 Plot

This is the tab for general plot control.



6.4.1 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 autoscales the spectrum and the **Full** button plots the full spectrum.

6.4.2 Display

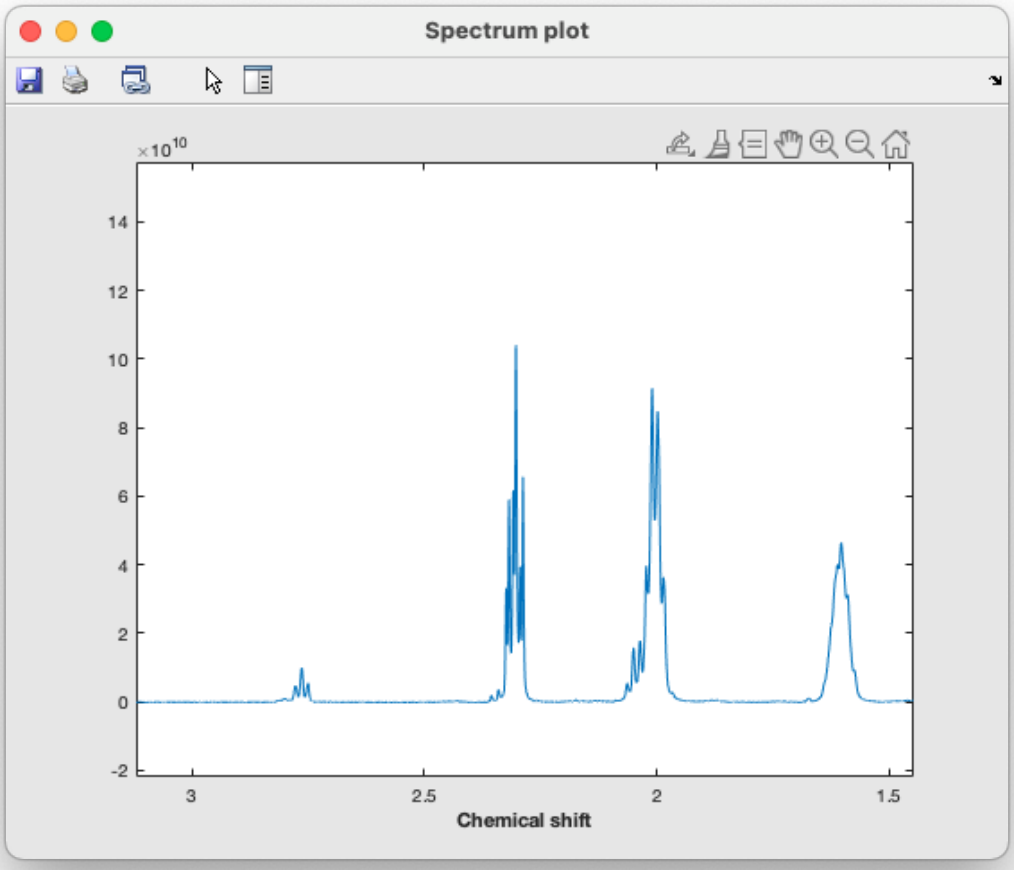
Controls for how the display the data. The user can display the spectrum or FID in either phase sensitive (**Phase**) or absolute value (**Abs**) mode.

6.4.3 Zoom/Pan

Controls for zooming or panning the display.

6.4.4 Separate plot

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

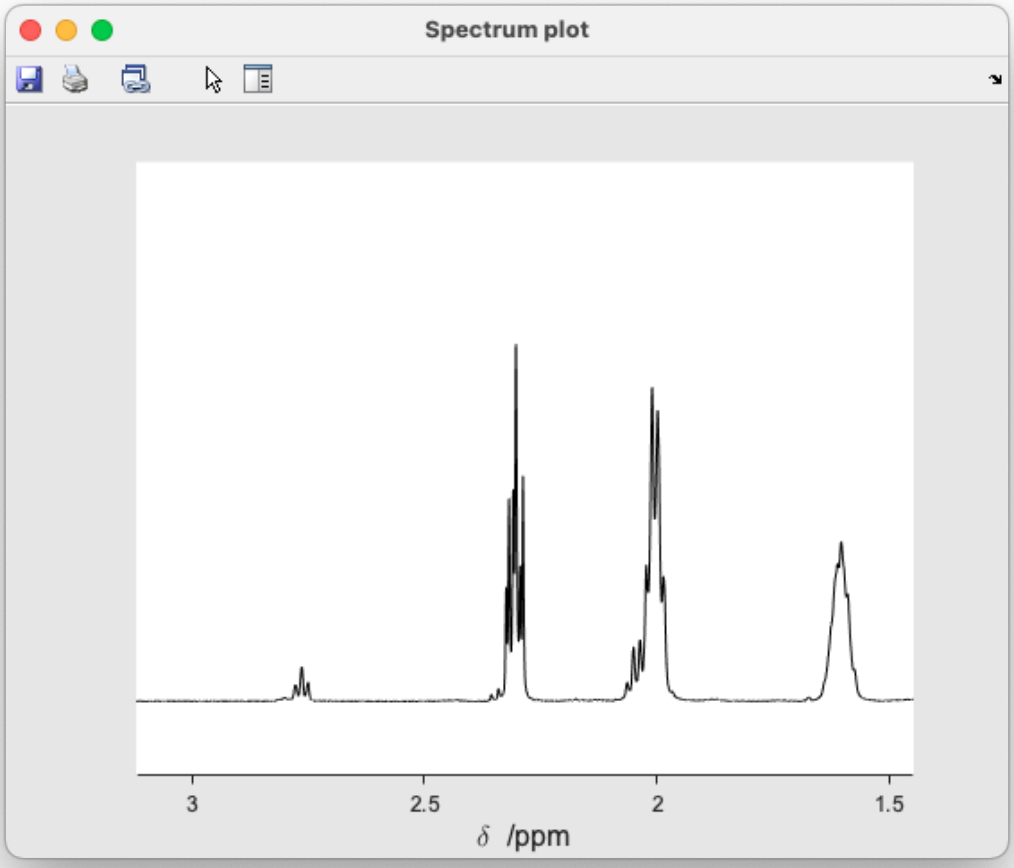


The **Publication plot** button plots the spectrum in a format more suitable for publication.

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.


Spectrum plot
3
2.5
2
1.5
δ /ppm

6.4.5 Displayed range

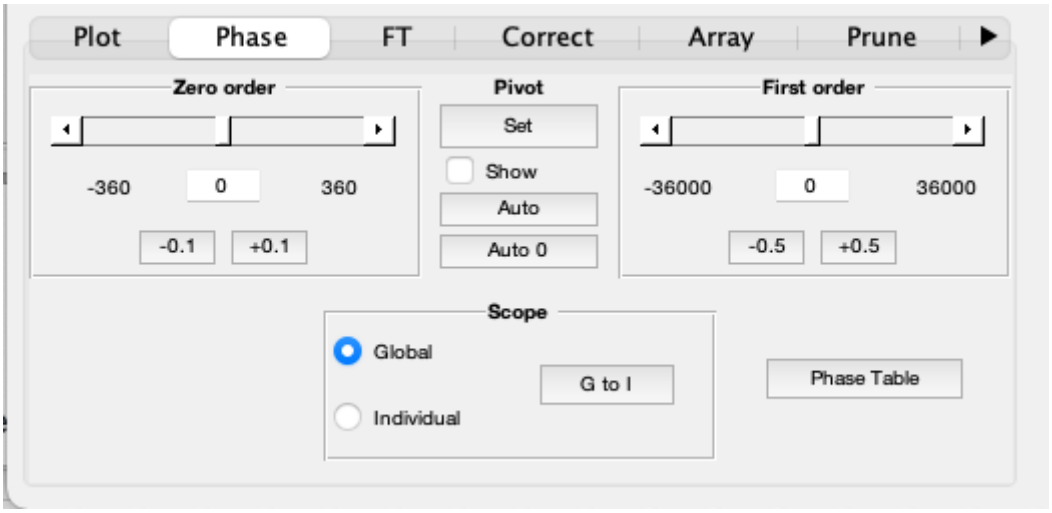
Here the user can set the range of chemical shifts displayed. These values also update when the spectrum is zoomed.

6.4.6 Horizontal scale

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

6.5 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.

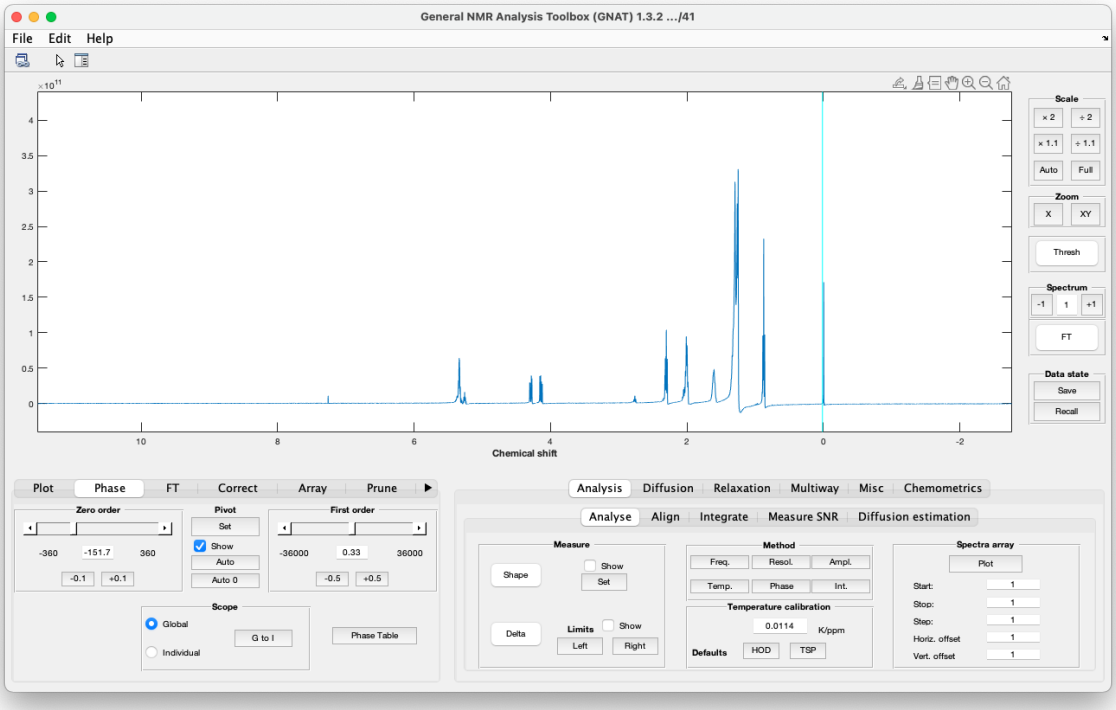
Typical use is:

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.

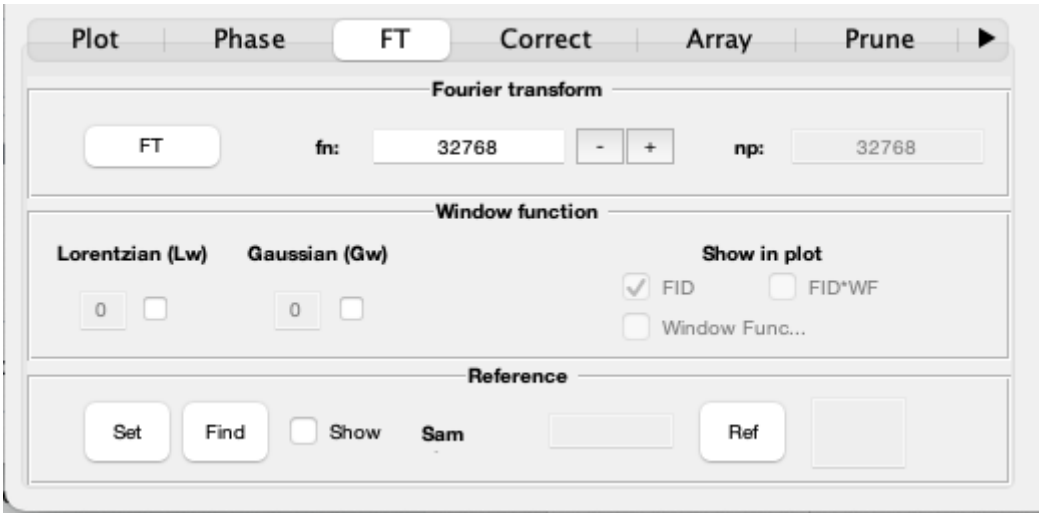
GNAT has the facility to do a separate phase correction for each spectrum in an array. This can e.g. be useful for DOSY data that sometimes show a gradient dependent phase. This is accessed with the **Scope** controls. *Global* uses the same phase for all the spectra in the array, while *Individual* 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.



6.6 FT

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



6.6.1 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.

6.6.2 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.

$e^{-\pi Lwt}$

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.

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

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.

6.6.3 Reference section

Here the user can reference the spectra to get a correct scale in the time and frequency domain.

In the frequency domain:

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.

In the time domain:

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 *FID* button to do the correction in the time domain
4. Use the option *Auto* to do the correction in all spectra imported into GNAT
5. Press the *Ref* button and type in the correct value.

6.7 Correct

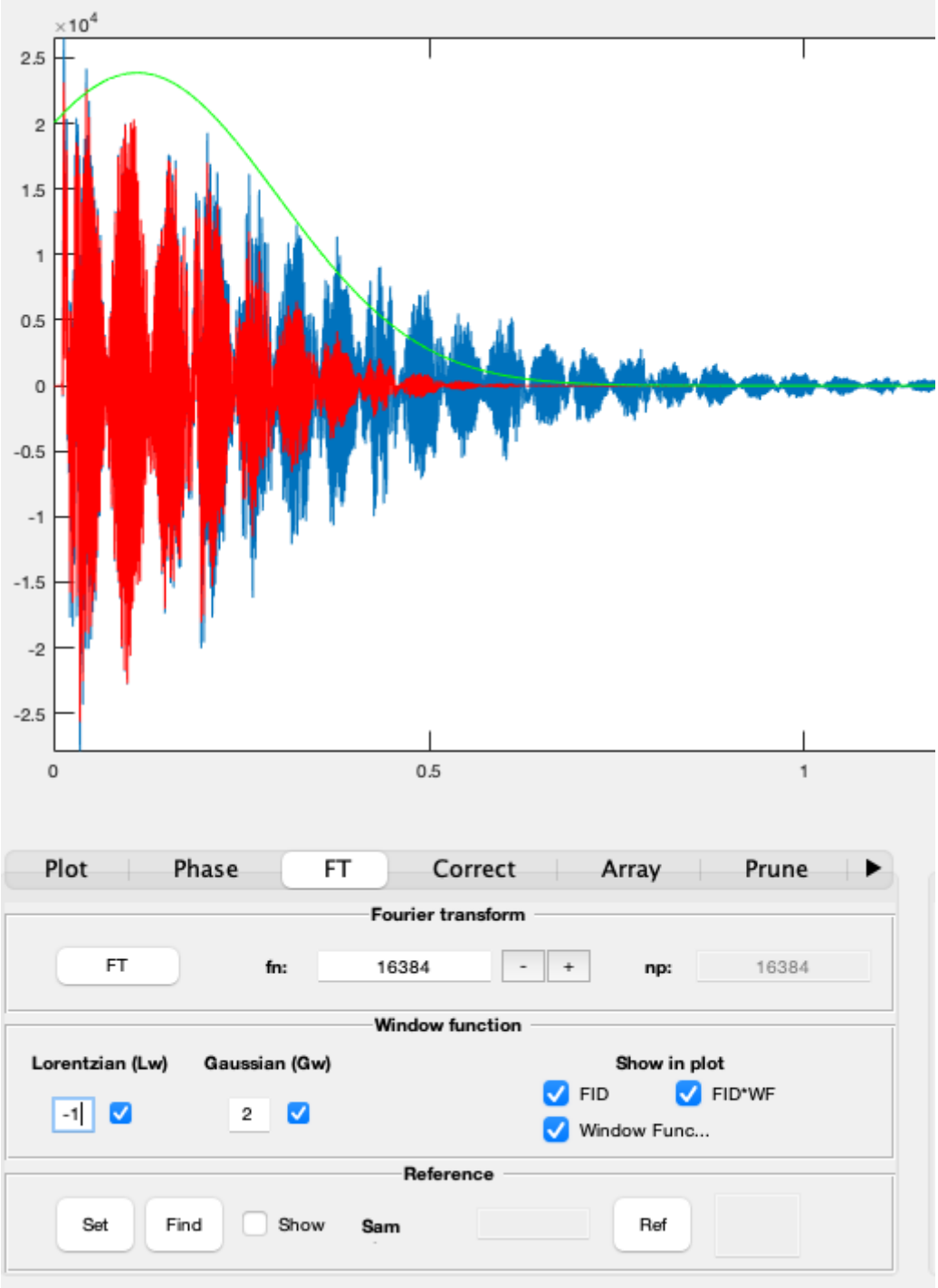
This is the tab for baseline correction and reference deconvolution control.

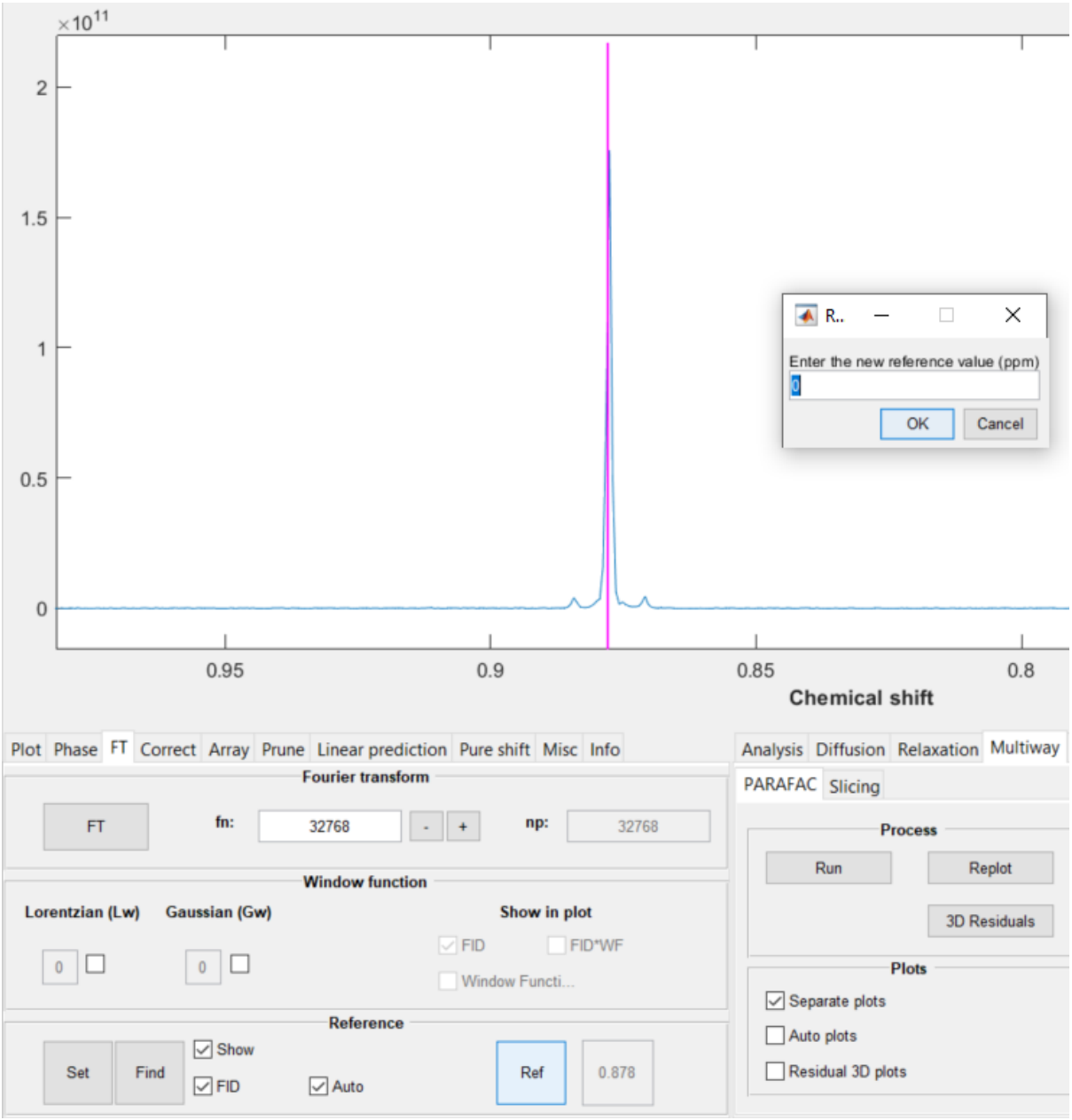
6.7.1 Baseline correction

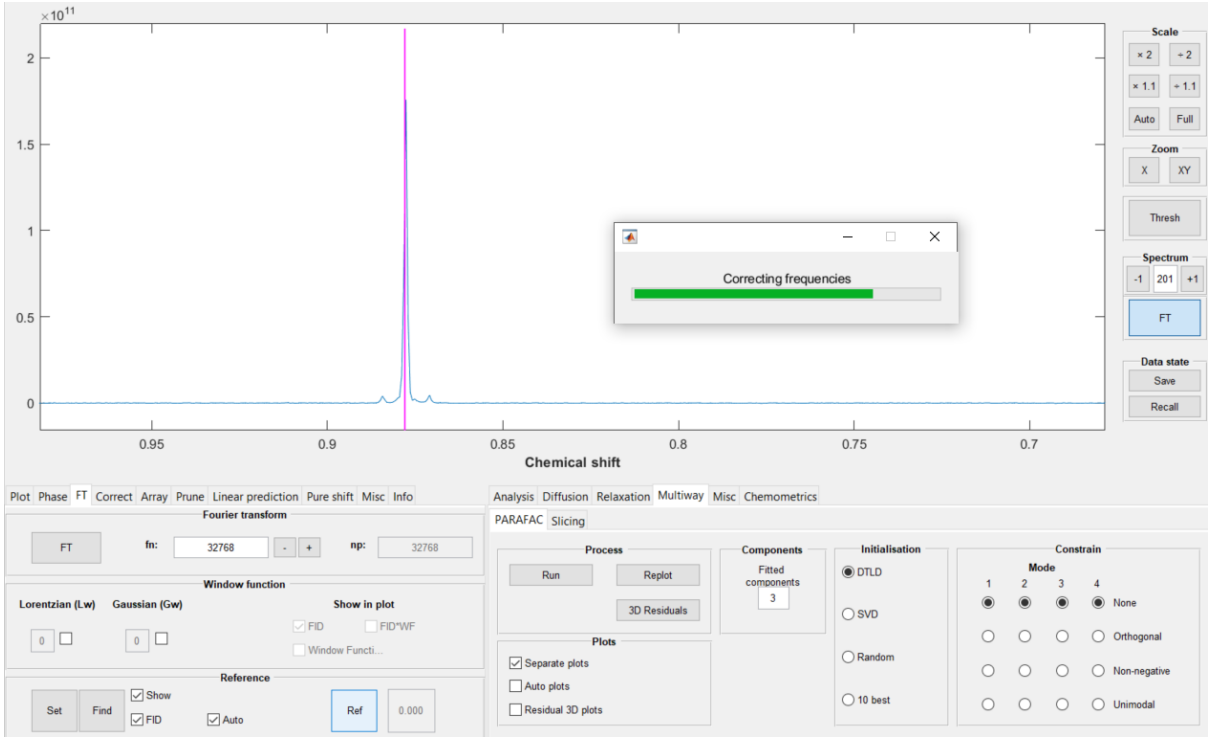
Automatic baseline correction is done by pressing the *Auto* button. The algorithm is taken from:

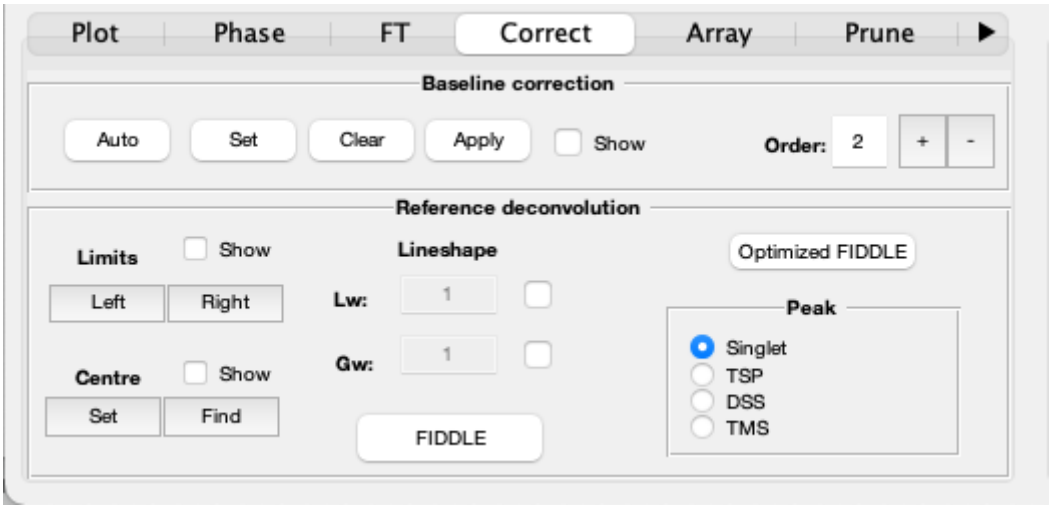
(1) Pearson, G. A. GENERAL BASELINE-RECOGNITION AND BASELINE-FLATTENING ALGORITHM. Journal of Magnetic Resonance 1977, 27 (2), 265.

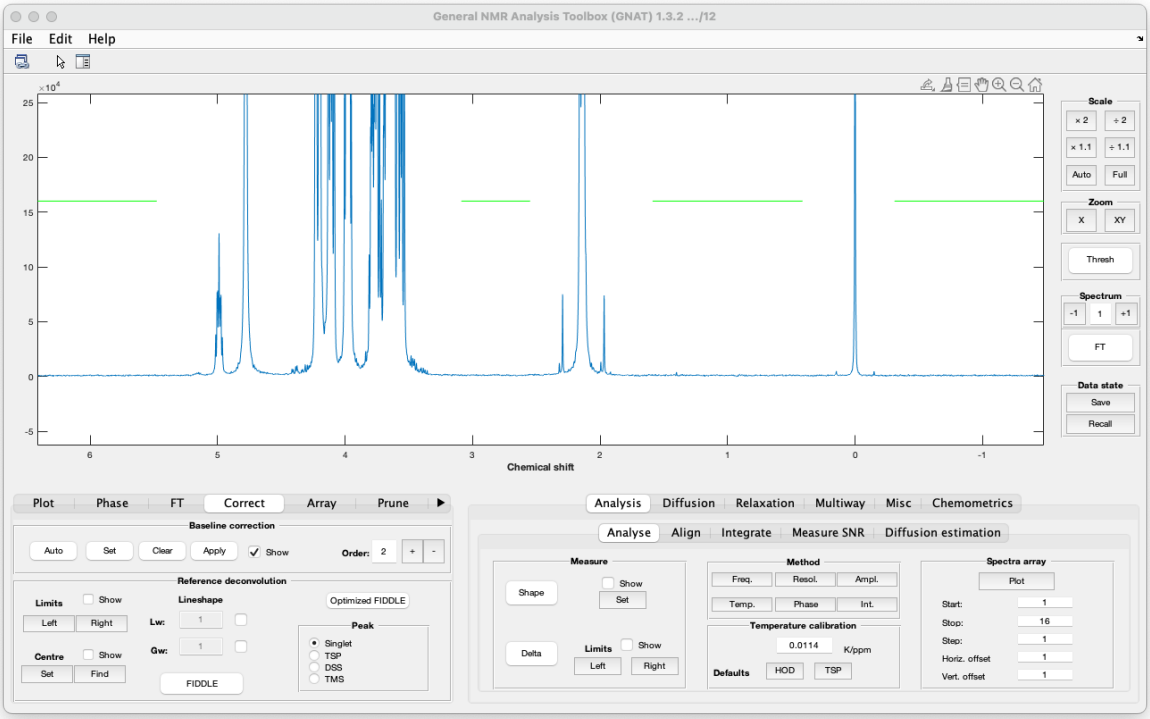
Manual baseline correction is done by manually identifying regions of empty baseline and then fitting a polynomial which is subtracted from the whole spectrum. The order of the polynomial is under user control in the *Order* parameter. The baseline regions are set by pressing the *Set* button and clicking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


×10^4
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
0.5
1
Plot
Phase
FT
Correct
Array
Prune
Fourier transform
FT
fn:
16384
-
+
np:
16384
Window function
Lorentzian (Lw)
Gaussian (Gw)
Show in plot
-1
2
FID
FID*WF
Window Func...
Reference
Set
Find
Show
Sam
Ref


×10^11
2
1.5
1
0.5
0
0.95
0.9
0.85
0.8
Chemical shift
R..
Enter the new reference value (ppm)
0
OK
Cancel
Plot
Phase
FT
Correct
Array
Prune
Linear prediction
Pure shift
Misc
Info
Analysis
Diffusion
Relaxation
Multiway
PARAFAC
Slicing
Fourier transform
FT
fn:
32768
-
+
np:
32768
Process
Run
Replot
3D Residuals
Window function
Lorentzian (Lw)
Gaussian (Gw)
Show in plot
0
0
FID
FID*WF
Window Functi...
Plots
Separate plots
Auto plots
Residual 3D plots
Reference
Set
Find
Show
FID
Auto
Ref
0.878


Correcting frequencies
Chemical shift
Plot Phase FT Correct Array Prune Linear prediction Pure shift Misc Info
Analysis Diffusion Relaxation Multiway Misc Chemometrics
PARAFAC Slicing
Fourier transform
Window function
Reference
Process
Components
Initialisation
Constrain
Plots


Plot
Phase
FT
Correct
Array
Prune
Baseline correction
Auto
Set
Clear
Apply
Show
Order:
2
Reference deconvolution
Limits
Show
Left
Right
Centre
Set
Find
Lineshape
Lw:
Gw:
FIDDLE
Optimized FIDDLE
Peak
Singlet
TSP
DSS
TMS



6.7.2 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).

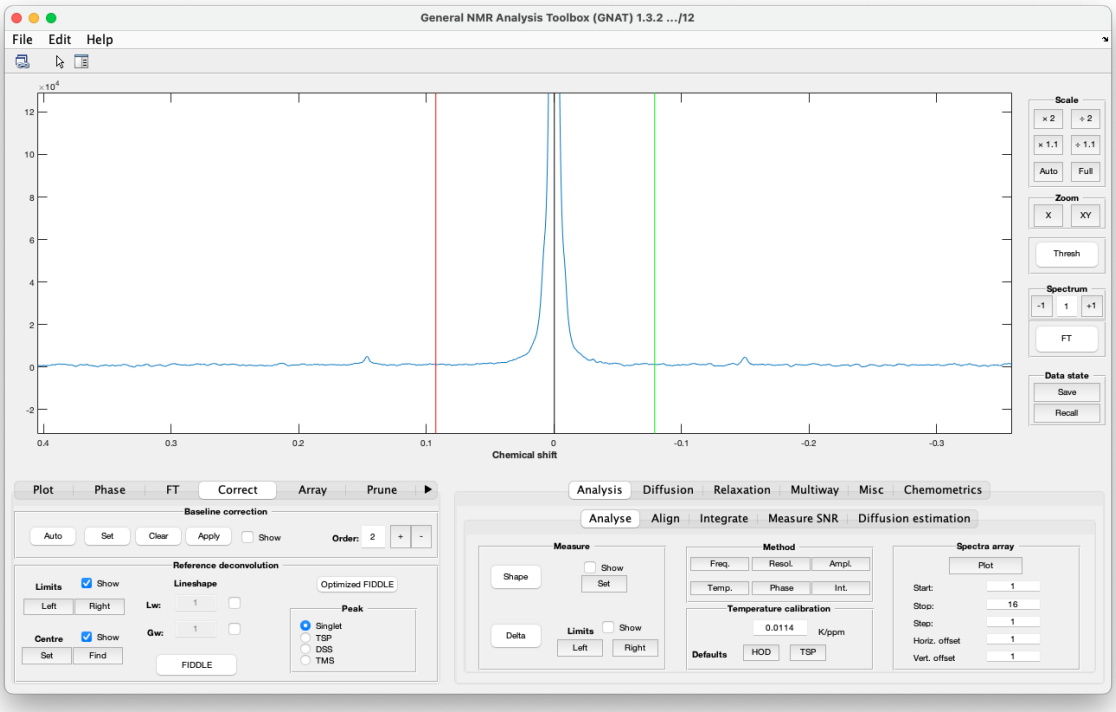
(1) Morris, G. A.; Barjat, H.; Home, T. J. Reference deconvolution methods. Progress in Nuclear Magnetic Resonance Spectroscopy 1997, 31 (2-3), 197.

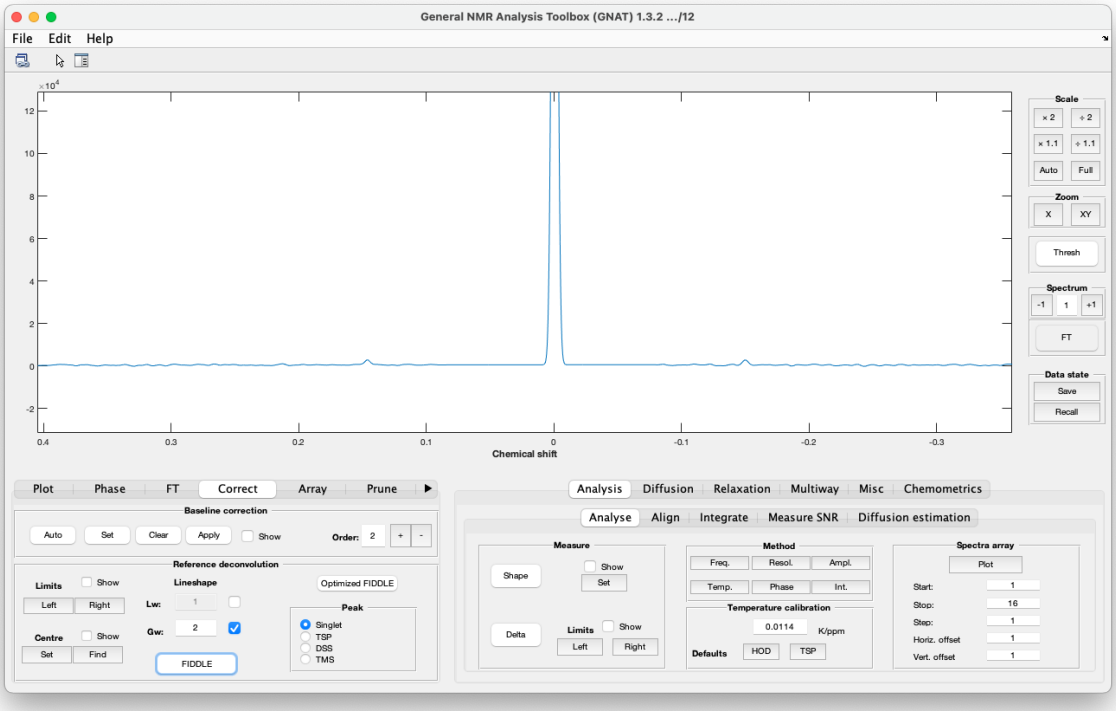
(2) Ebrahimi, P.; Nilsson, M.; Morris, G. A.; Jensen, H. M.; Engelsen, S. B. Cleaning up NMR spectra with reference deconvolution for improving multivariate analysis of complex mixture spectra. Journal of Chemometrics 2014, 28 (8), 656.

The reference peak is ideally a simple singlet, 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 frequency 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.

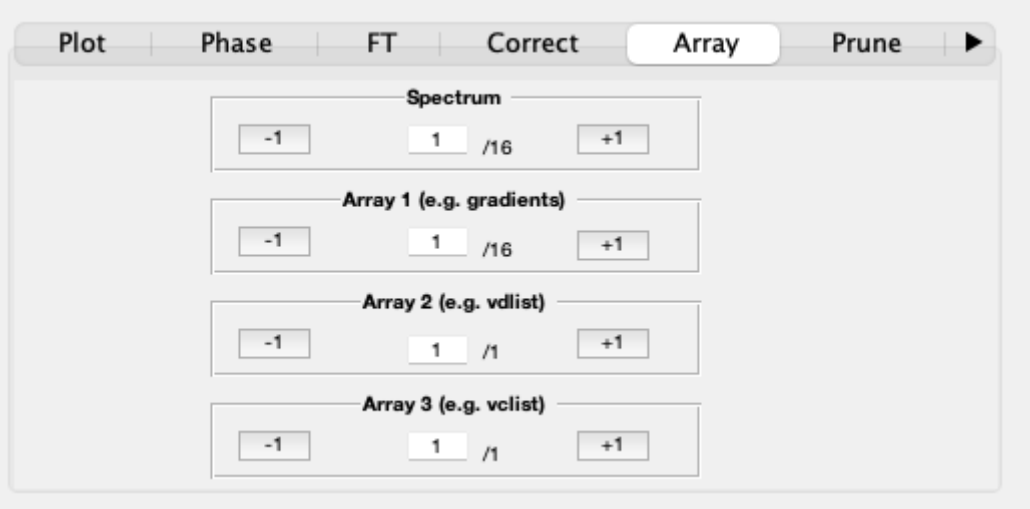
Refrence deconvolution is performed by pressing the *FIDDLE* button. The experimental line shape will be replaced by a "perfect" lineshape as decided by the *Lineshape* parameters. There are the same as in the *FT* tab. Here the lineshape was replaced by a 2 Hz Gaussian shape. The lineshape can be a combination of Lorentzian and Gaussian.


General NMR Analysis Toolbox (GNAT) 1.3.2 .../12
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Plot Phase FT Correct Array Prune
Baseline correction
Auto Set Clear Apply Show Order: 2
Reference deconvolution
Limits Show Left Right
Centre Show Set Find
Lineshape
Lw: Gw:
FIDDLE
Optimized FIDDLE
Peak
Singlet TSP DSS TMS
Analysis Diffusion Relaxation Multiway Misc Chemometrics
Analyse Align Integrate Measure SNR Diffusion estimation
Measure
Shape Delta
Show Set
Limits Show Left Right
Method
Freq. Resol. Ampl. Temp. Phase Int.
Temperature calibration
0.0114 K/ppm
Defaults HOD TSP
Spectra array
Plot
Start: Stop: Step: Horiz. offset Vert. offset


General NMR Analysis Toolbox (GNAT) 1.3.2 .../12
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Plot Phase FT Correct Array Prune
Baseline correction
Auto Set Clear Apply Show Order: 2
Reference deconvolution
Limits Show Left Right
Centre Show Set Find
Lineshape
Lw: Gw: 2
FIDDLE
Optimized FIDDLE
Peak
Singlet TSP DSS TMS
Analysis Diffusion Relaxation Multiway Misc Chemometrics
Analyse Align Integrate Measure SNR Diffusion estimation
Measure
Shape Delta
Show Set
Limits Show Left Right
Method
Freq. Resol. Ampl. Temp. Phase Int.
Temperature calibration
0.0114 K/ppm
Defaults HOD TSP
Spectra array
Plot
Start: Stop: Step: Horiz. offset Vert. offset

6.8 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 time 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 through the spectra in the first array dimension. Here the array is a diffusion experiment so there is only 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 through 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.

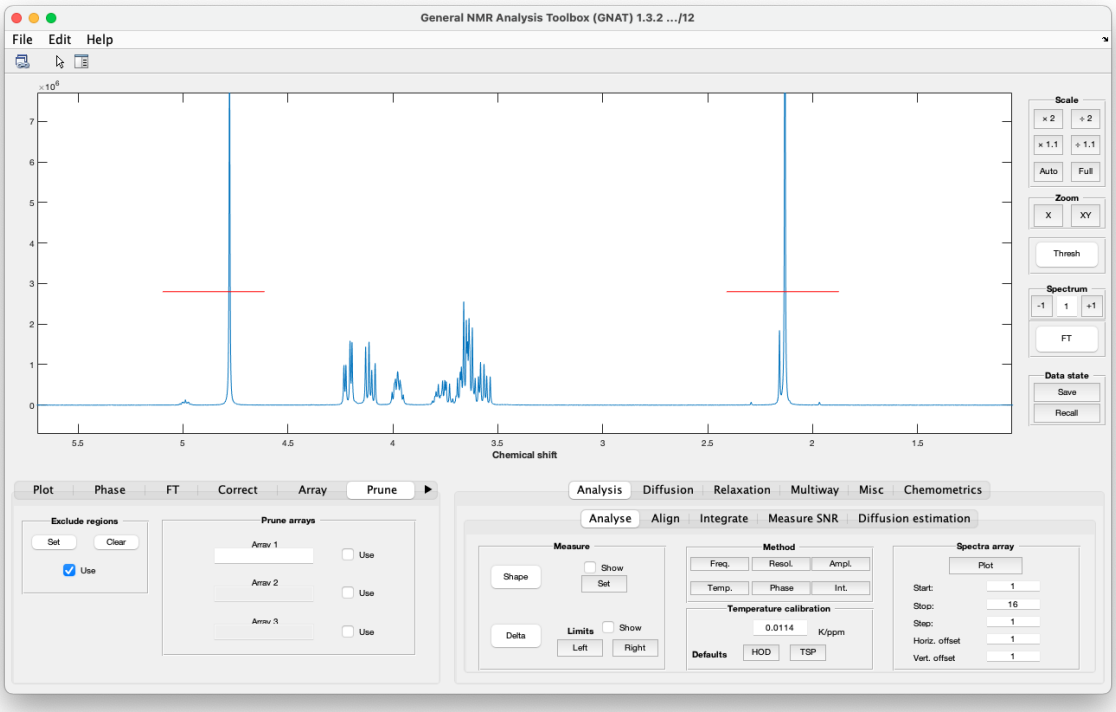
(2) Khajeh, M.; Botana, A.; Bernstein, M. A.; Nilsson, M.; Morris, G. A. Reaction Kinetics Studied Using Diffusion-Ordered Spectroscopy and Multiway Chemometrics. Analytical Chemistry 2010, 82 (5), 2102.

6.9 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.

6.9.1 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.

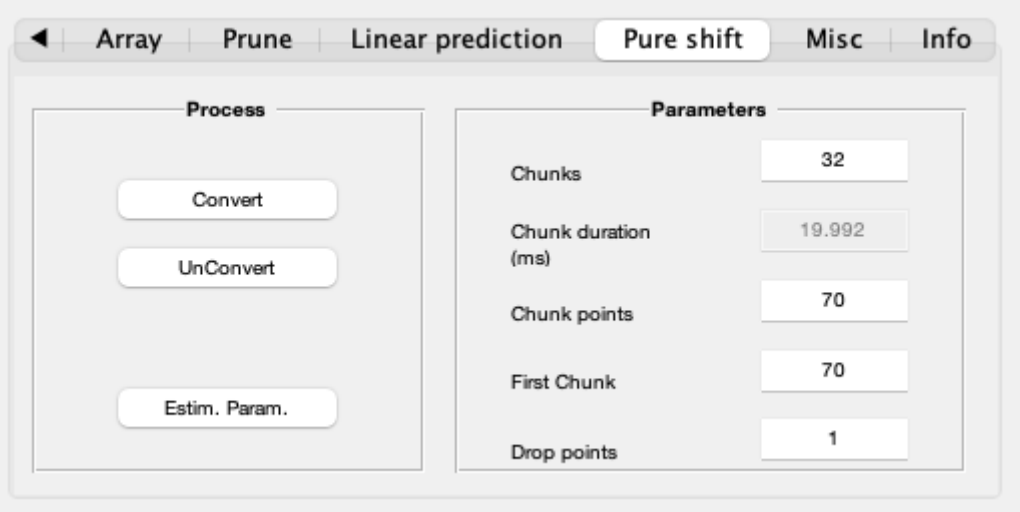


6.9.2 Prune arrays

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*

6.10 Pure Shift

Here the user 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.



6.10.1 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.

6.10.2 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*.

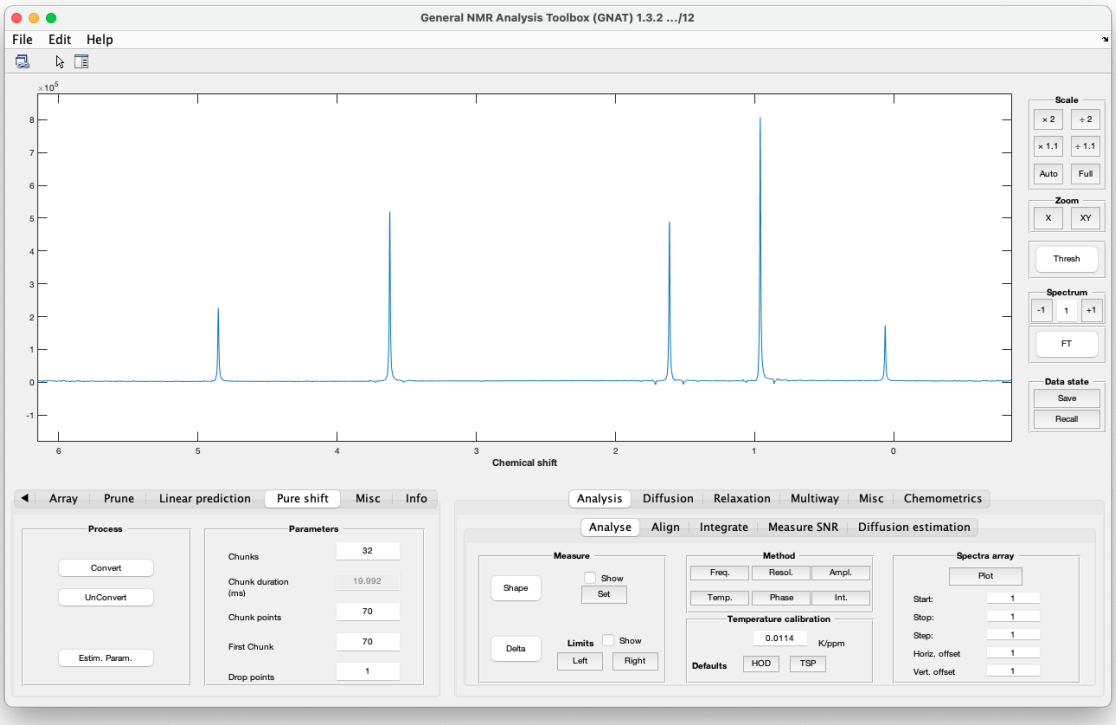
Drop points is the number of complex data points that is discarded in the beginning of each chunk.

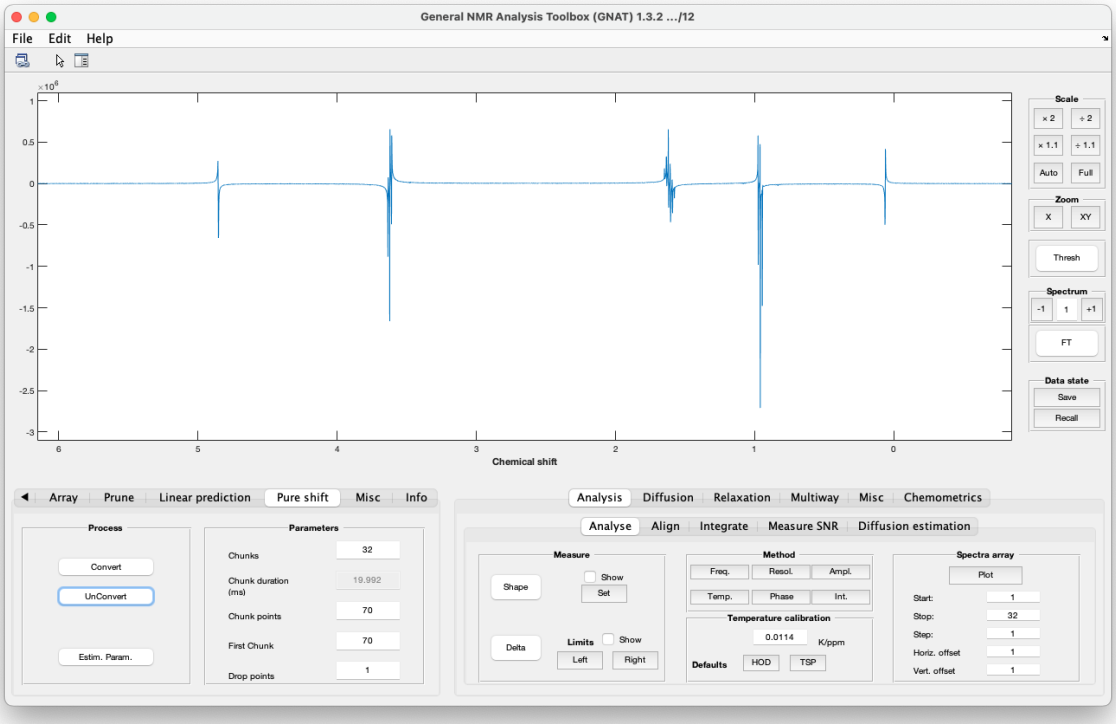
6.10.3 More information

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.


General NMR Analysis Toolbox (GNAT) 1.3.2 .../12
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Array Prune Linear prediction Pure shift Misc Info
Process
Convert
UnConvert
Estim. Param.
Parameters
Chunks 32
Chunk duration (ms) 19.992
Chunk points 70
First Chunk 70
Drop points 1
Analysis Diffusion Relaxation Multiway Misc Chemometrics
Analyse Align Integrate Measure SNR Diffusion estimation
Measure
Shape
Show
Set
Delta
Limits
Left
Right
Method
Freq. Resol. Ampl.
Temp. Phase Int.
Temperature calibration
0.0114
K/ppm
Defaults
HOD
TSP
Spectra array
Plot
Start: 1
Stop: 1
Step: 1
Horiz. offset 1
Vert. offset 1


General NMR Analysis Toolbox (GNAT) 1.3.2 .../12
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Array Prune Linear prediction Pure shift Misc Info
Process
Convert
UnConvert
Estim. Param.
Parameters
Chunks 32
Chunk duration (ms) 19.992
Chunk points 70
First Chunk 70
Drop points 1
Analysis Diffusion Relaxation Multiway Misc Chemometrics
Analyse Align Integrate Measure SNR Diffusion estimation
Measure
Shape
Show
Set
Delta
Limits
Left
Right
Method
Freq. Resol. Ampl.
Temp. Phase Int.
Temperature calibration
0.0114
K/ppm
Defaults
HOD
TSP
Spectra array
Plot
Start: 1
Stop: 32
Step: 1
Horiz. offset 1
Vert. offset 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.

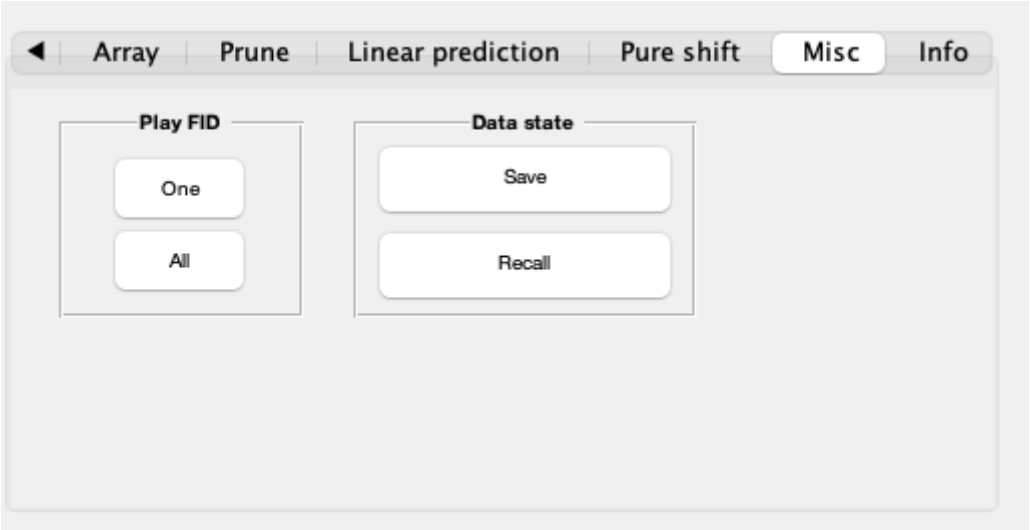
6.11 Linear prediction

Note

This section is reserved for future documentation of linear prediction functionality.

6.12 Misc

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



6.12.1 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)

6.12.2 Data state section

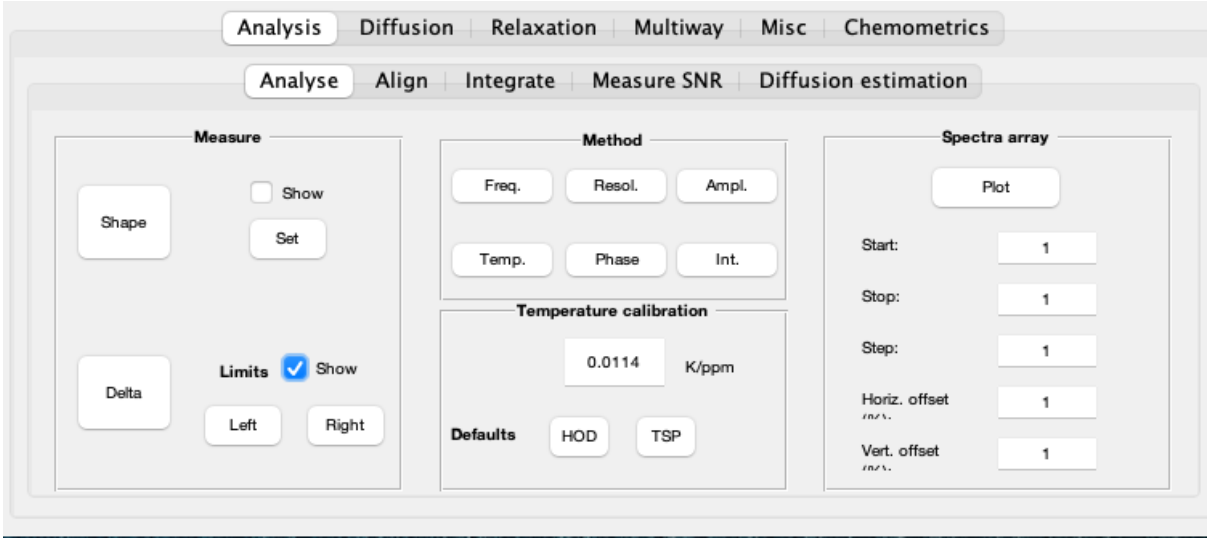
Here the user can save the data in a particular state, with the *Save* button. For examples with some specific processing paramaters, and then recall that state with the *Recall* button.

6.13 Info

This page displays information about import and processing in GNAT. In current versions, much of this feedback is written directly to the MATLAB window, or to the terminal in compiled versions.

6.14 Analysis

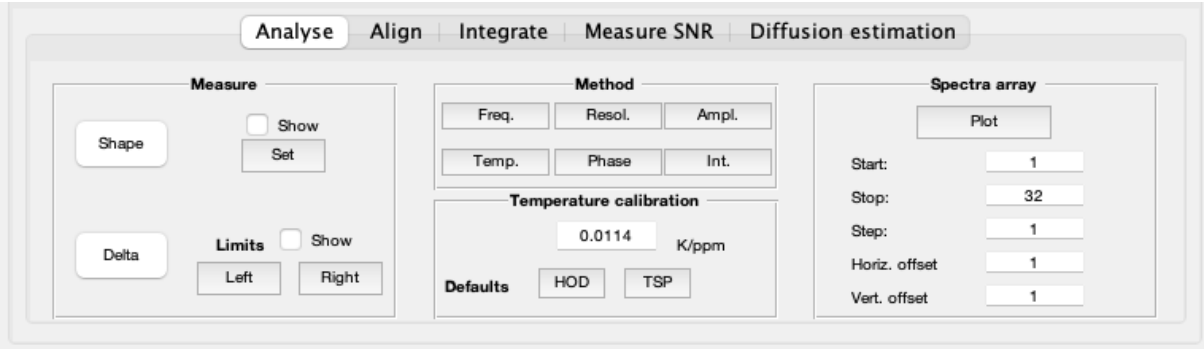
This tab contains general analysis tools, with particular emphasis on arrayed NMR data.



6.14.1 Functionalities

Analyse

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



Measure section

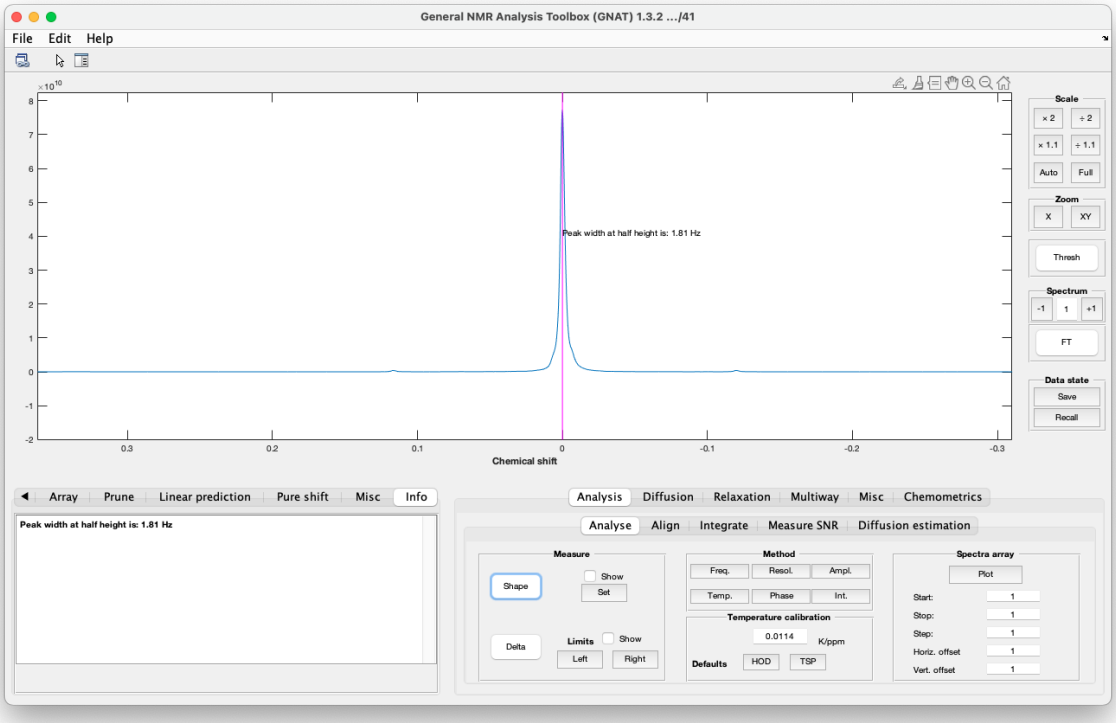
The *Shape* button will display the peak width at half height for the selected peaks - selected with the *Set* button

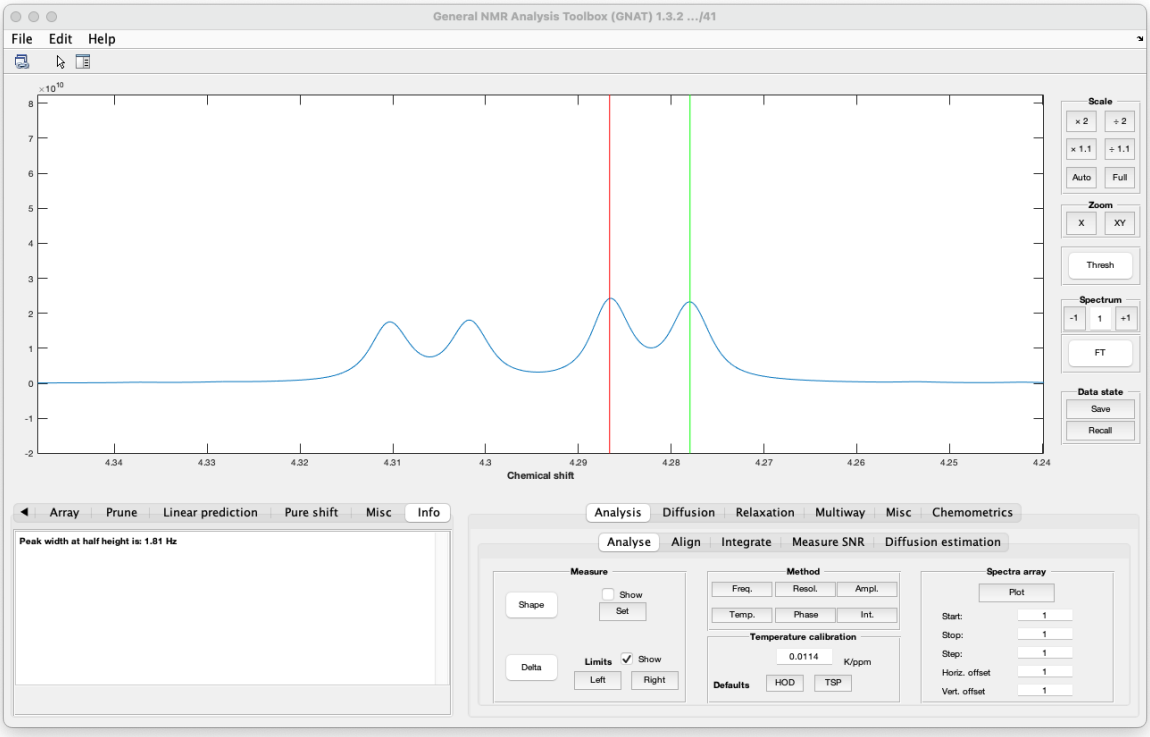
The *Delta* button will display the peak width at half height for the selected peaks - selected with the *Left* and *Right* buttons.

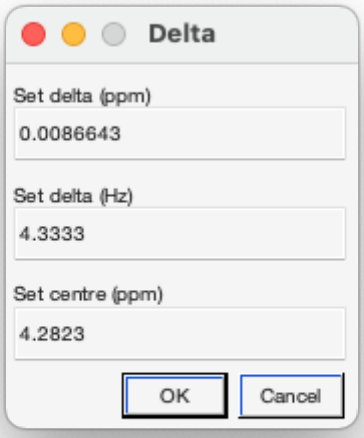
It will open a dialog where the user can set the difference in frequency as well as the centre between Left and Right.

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.


General NMR Analysis Toolbox (GNAT) 1.3.2 .../41
File Edit Help
Peak width at half height is: 1.81 Hz
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Array Prune Linear prediction Pure shift Misc Info
Peak width at half height is: 1.81 Hz
Analysis Diffusion Relaxation Multiway Misc Chemometrics
Analyse Align Integrate Measure SNR Diffusion estimation
Measure
Shape
Show
Set
Delta
Limits
Left
Right
Method
Freq. Resol. Ampl.
Temp. Phase Int.
Temperature calibration
0.0114
K/ppm
Defaults
HOD
TSP
Spectra array
Plot
Start:
Stop:
Step:
Horiz. offset
Vert. offset


General NMR Analysis Toolbox (GNAT) 1.3.2 .../41
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Array Prune Linear prediction Pure shift Misc Info
Peak width at half height is: 1.81 Hz
Analysis Diffusion Relaxation Multiway Misc Chemometrics
Analyse Align Integrate Measure SNR Diffusion estimation
Measure
Shape
Show
Set
Delta
Limits
Left
Right
Method
Freq. Resol. Ampl.
Temp. Phase Int.
Temperature calibration
0.0114
K/ppm
Defaults
HOD
TSP
Spectra array
Plot
Start:
Stop:
Step:
Horiz. offset
Vert. offset



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.

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 function is currently marked as work in progress and should not be relied on for production use.

The alignment interface is intended for manual alignment of spectra in an array.

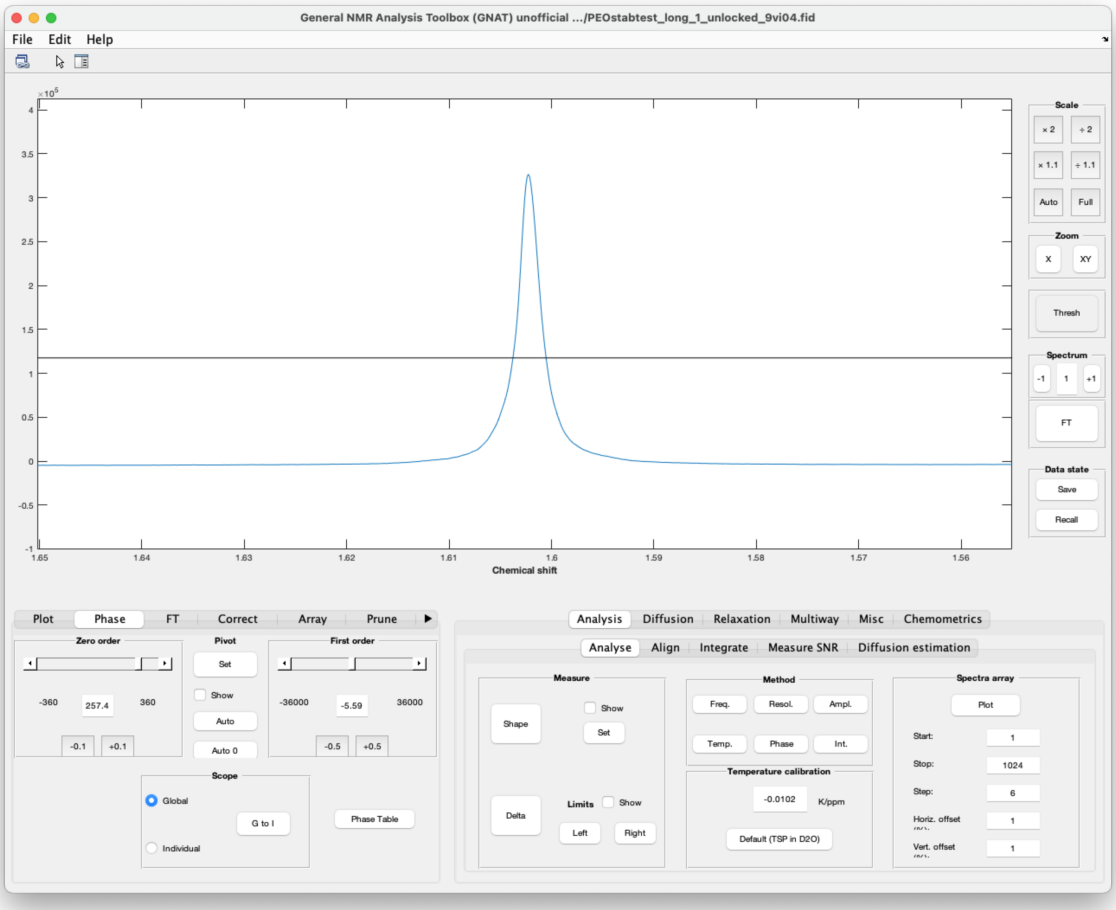
Integrate

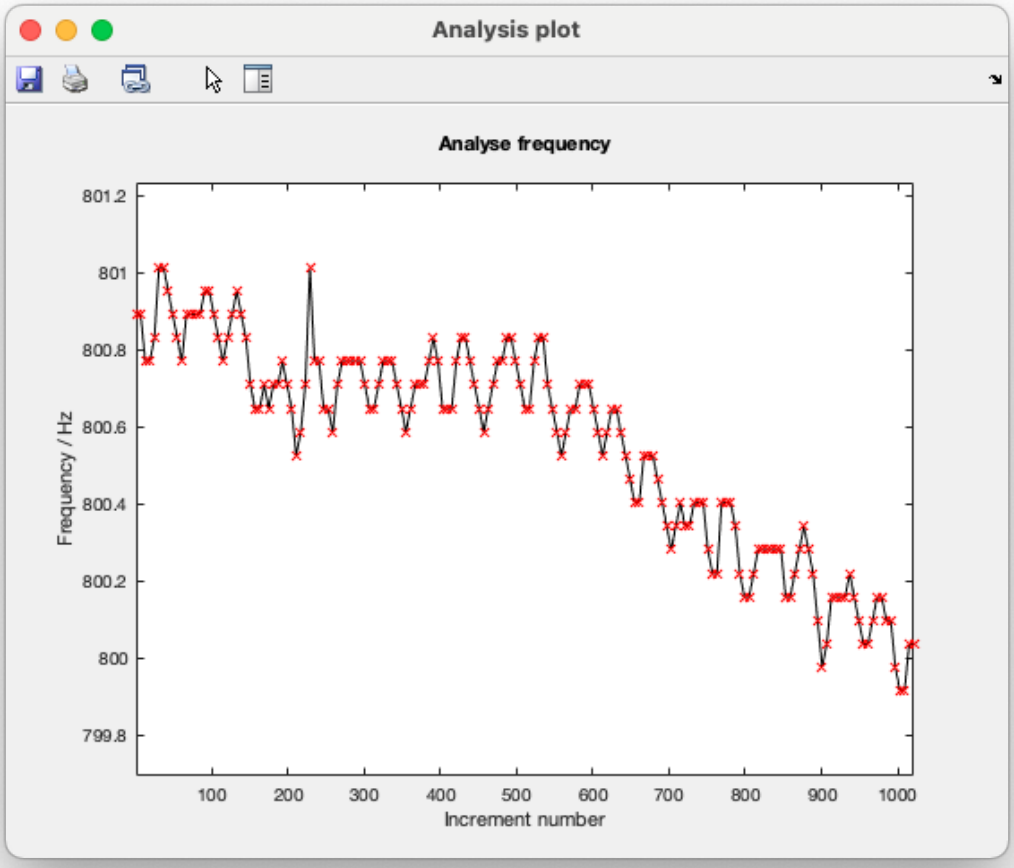
Here the user can find control for integration of spectra.

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.


General NMR Analysis Toolbox (GNAT) unofficial .../PEOstabtest_long_1_unlocked_9vi04.fid
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Plot Phase FT Correct Array Prune
Zero order
Pivot
First order
Set
Show
Auto
Auto 0
Scope
Global
Individual
G to I
Phase Table
Analysis Diffusion Relaxation Multiway Misc Chemometrics
Analyse Align Integrate Measure SNR Diffusion estimation
Measure
Shape
Delta
Limits
Left
Right
Method
Freq.
Resol.
Ampl.
Temp.
Phase
Int.
Temperature calibration
-0.0102
K/ppm
Default (TSP in D2O)
Spectra array
Plot
Start:
Stop:
Step:
Horiz. offset
Vert. offset


Analysis plot
Analyse frequency
801.2
801
800.8
800.6
800.4
800.2
800
799.8
Frequency / Hz
100
200
300
400
500
600
700
800
900
1000
Increment number

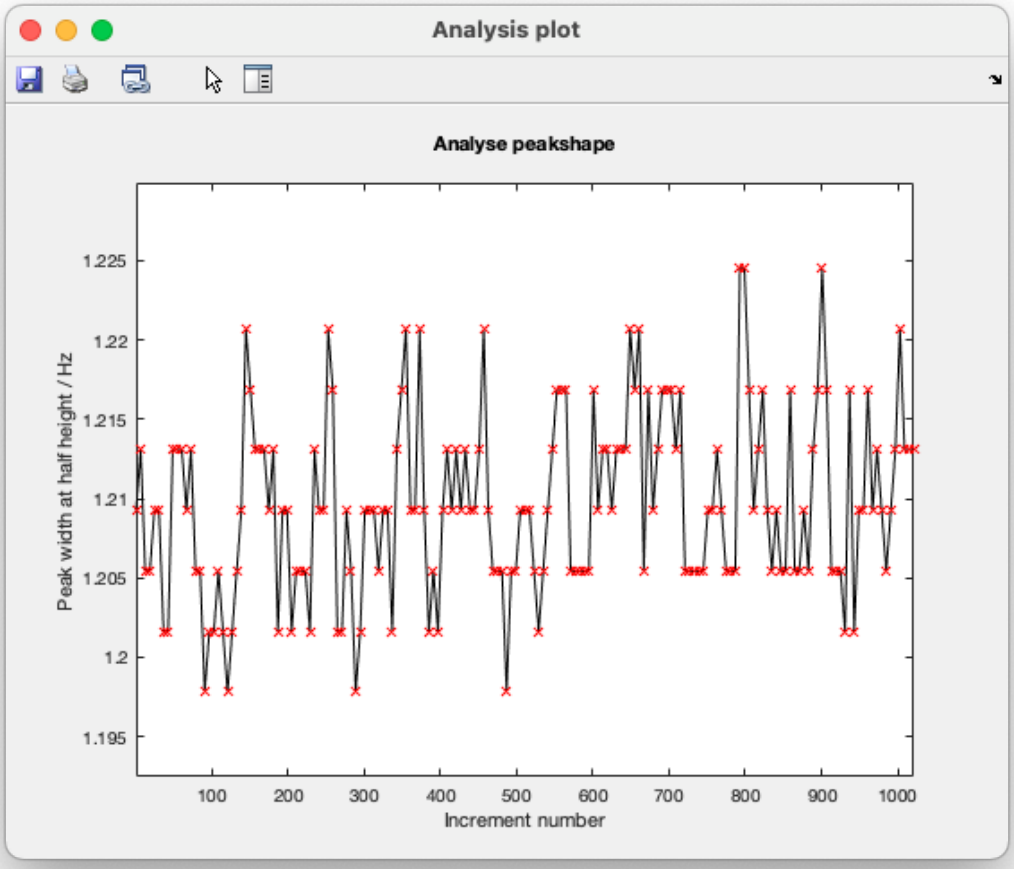


Fig. 1: Pressing the *Ampl.* button will plot peak amplitude as a function of spectrum.

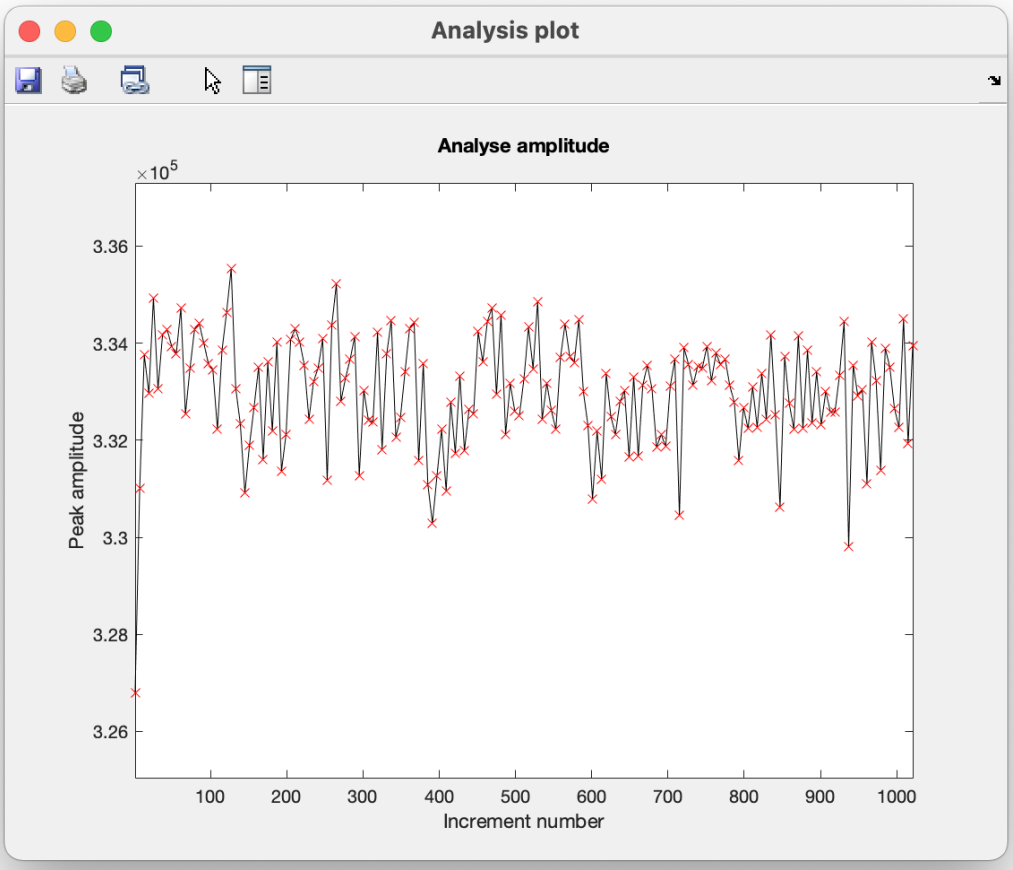


Fig. 2: Pressing the *Temp.* button will plot temperature change as a function of spectrum. The default value is from measuring the changed for the HOD peak relative to TSP in an aqueous sample.

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

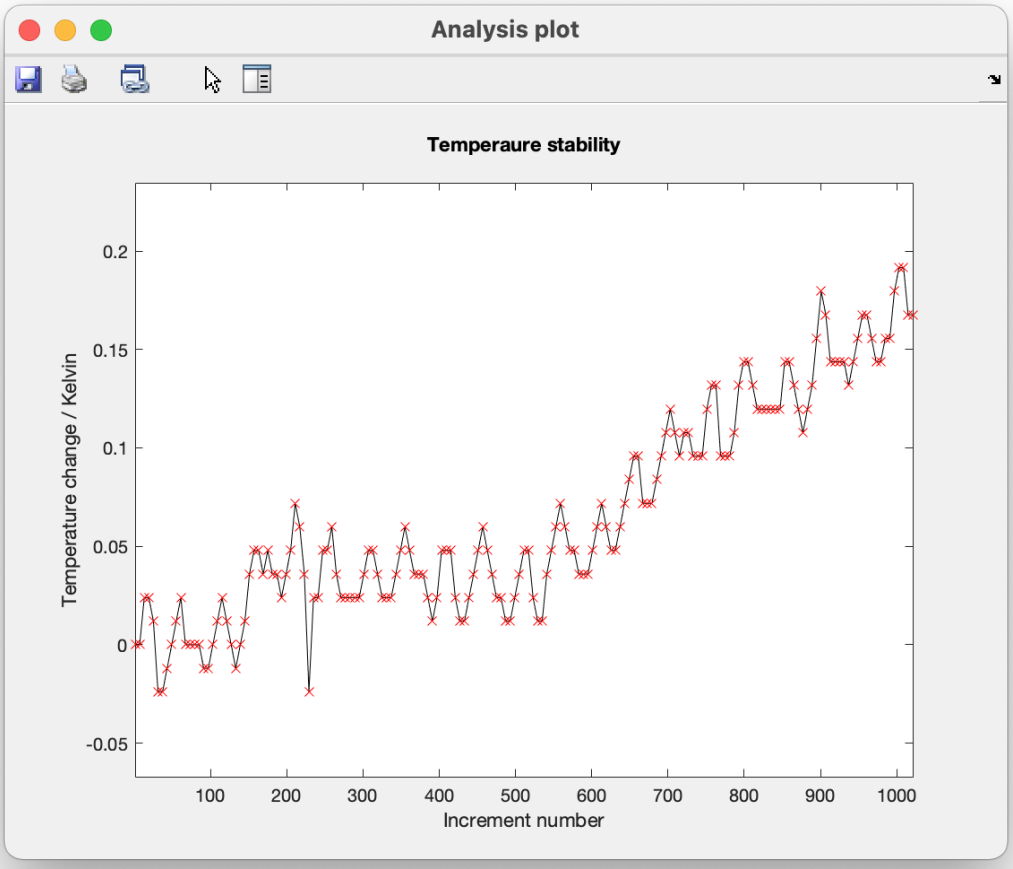


Fig. 3: Pressing the *Phase.* button will plot peak phase as a function of spectrum.

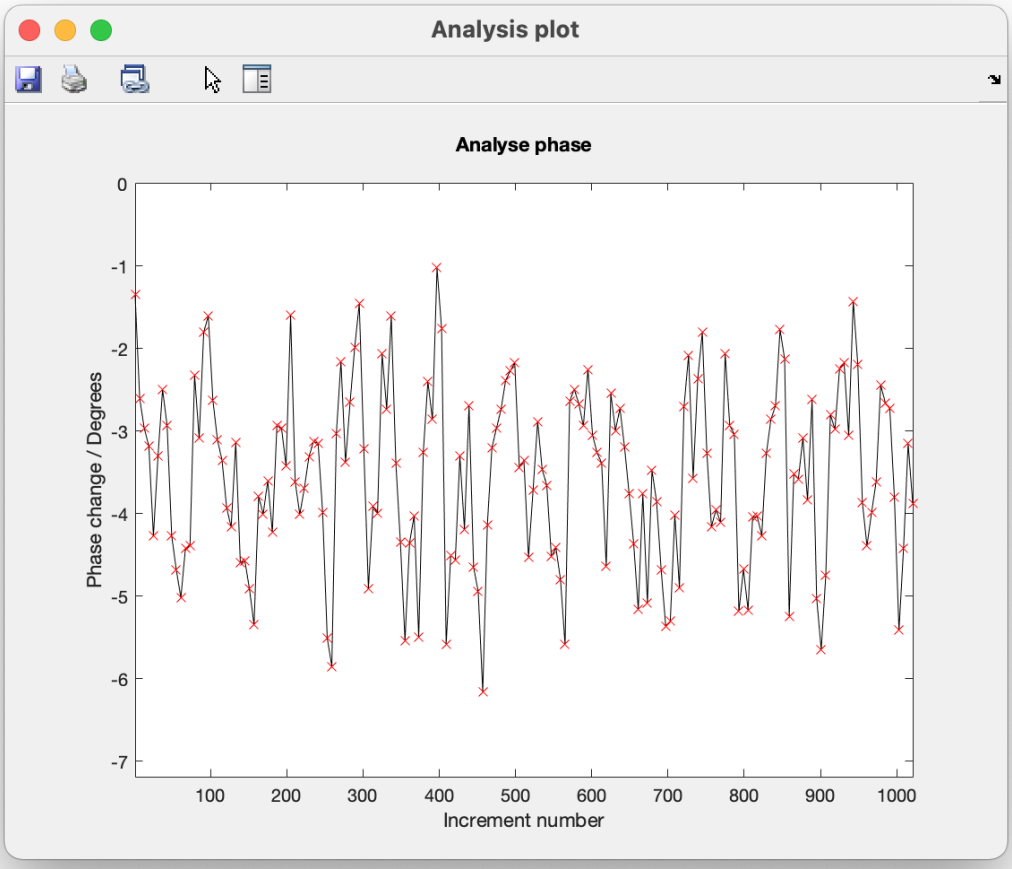
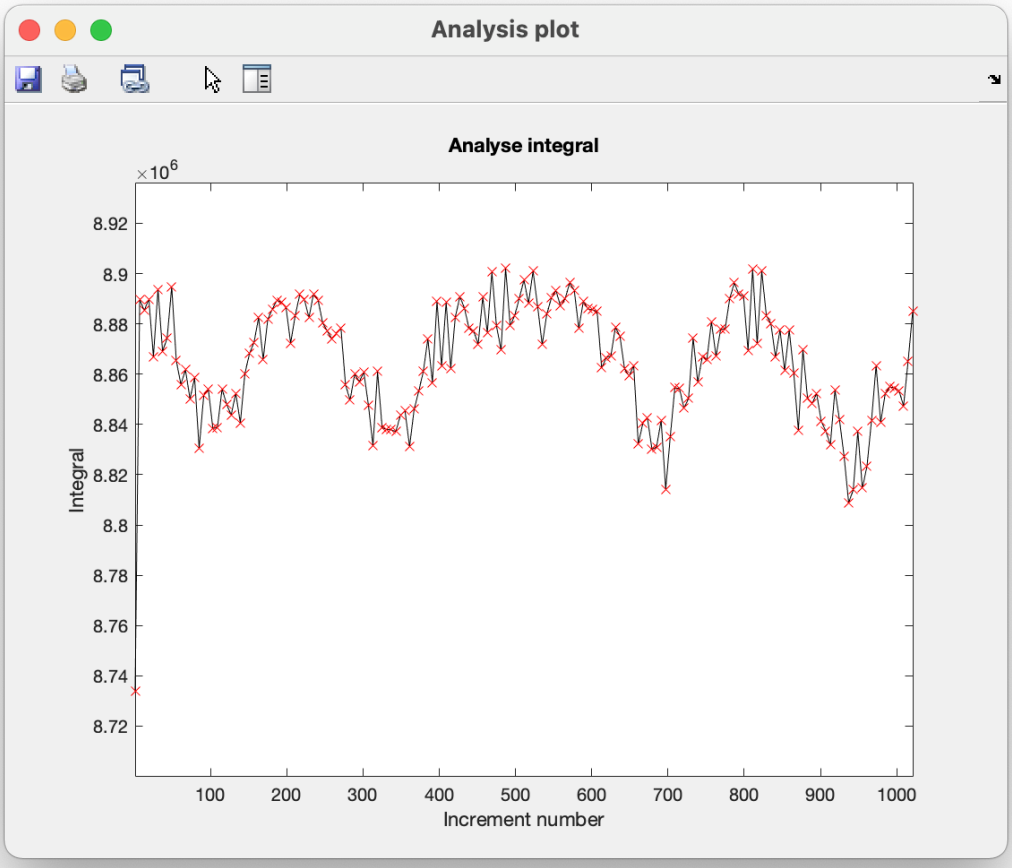
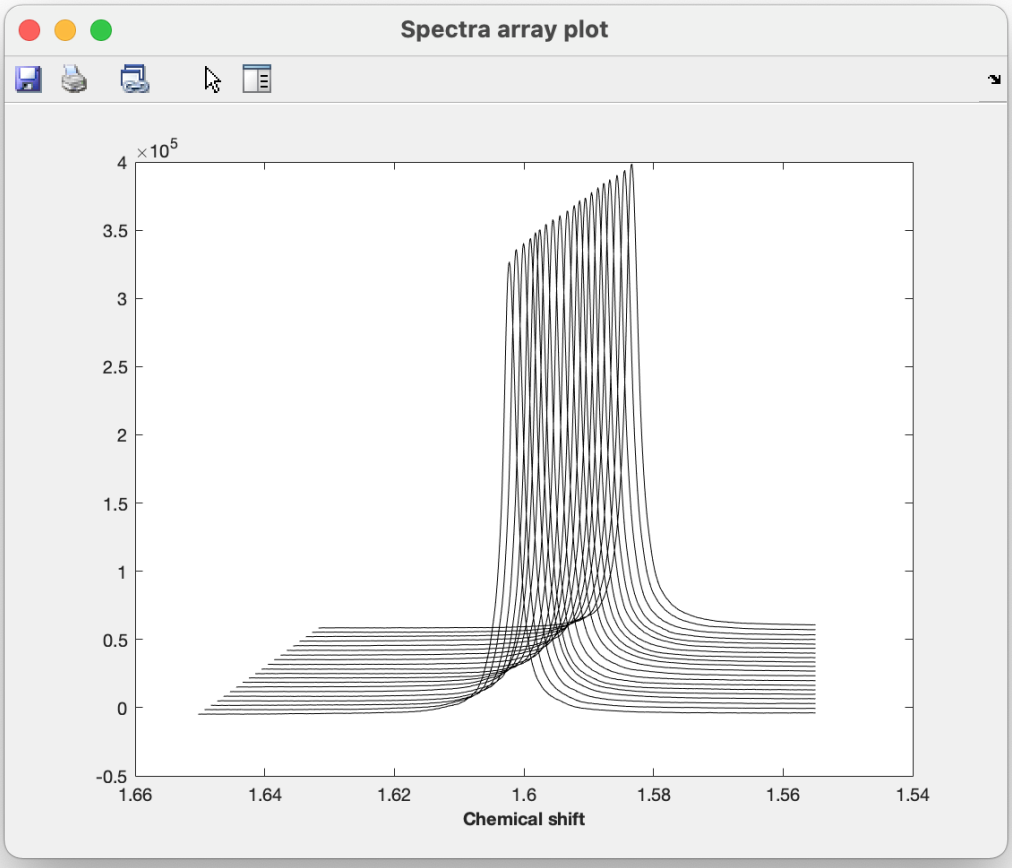
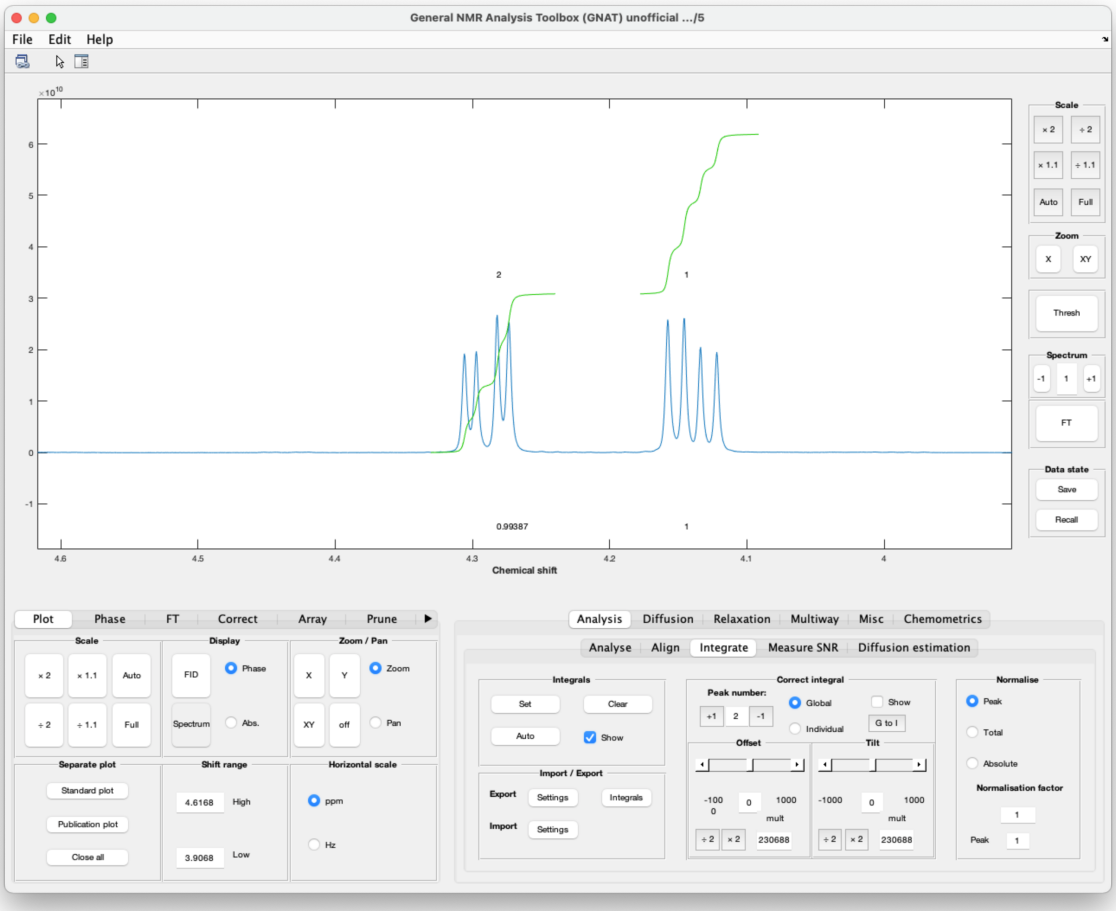


Fig. 4: Pressing the *Int.* button will plot peak integral as a function of spectrum. Integrals are set in the *Integrate* tab.


Analysis plot
Analyse integral
×10^6
8.92
8.9
8.88
8.86
8.84
8.82
8.8
8.78
8.76
8.74
8.72
Integral
100
200
300
400
500
600
700
800
900
1000
Increment number


Spectra array plot
×10^5
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
1.66
1.64
1.62
1.6
1.58
1.56
1.54
Chemical shift


General NMR Analysis Toolbox (GNAT) unofficial .../5
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Plot Phase FT Correct Array Prune
Analysis Diffusion Relaxation Multiway Misc Chemometrics
Analyse Align Integrate Measure SNR Diffusion estimation
Integrals
Import / Export
Correct integral
Peak number:
Offset
Tilt
Normalise
Normalisation factor
Separate plot
Standard plot
Publication plot
Close all
Shift range
Horizontal scale
ppm
Hz

> **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 usedd.

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

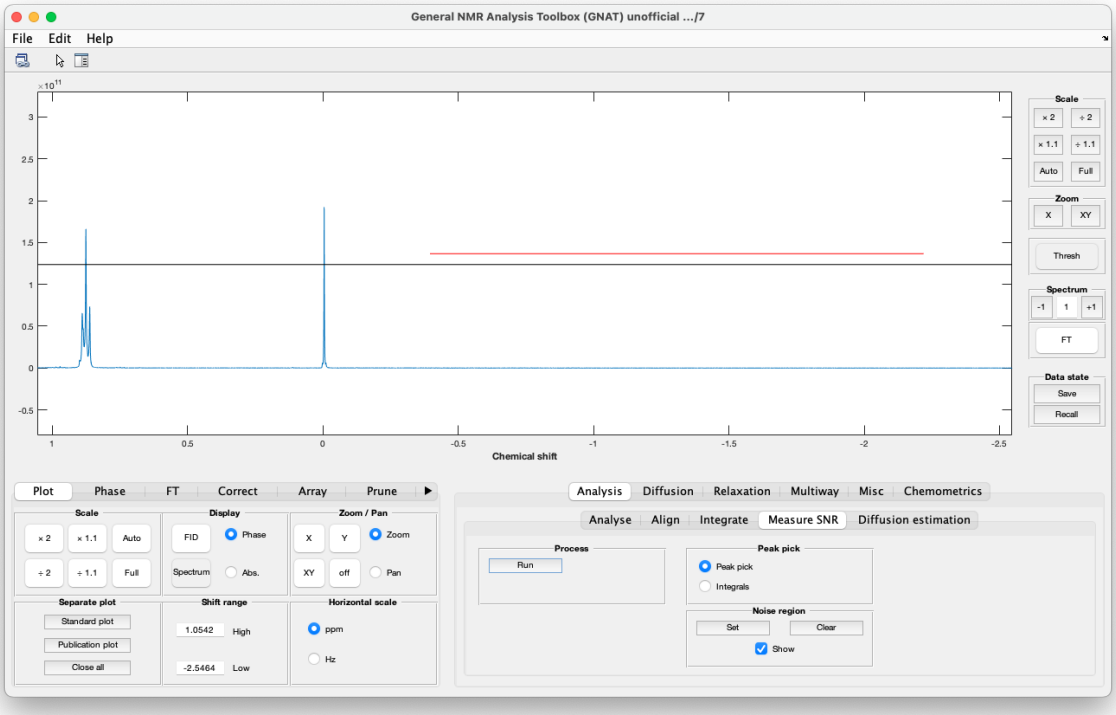
Measure SNR

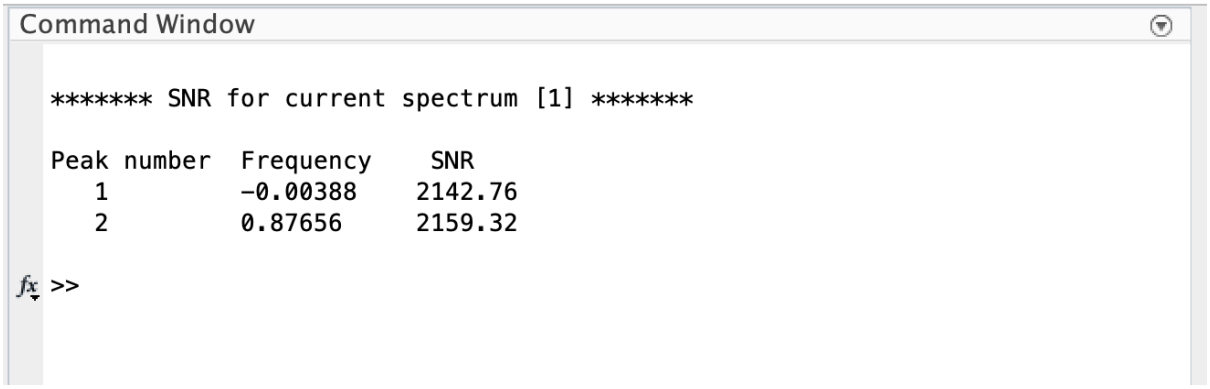
Here the user can find measure the signal-to-noise ratio (SNR) of spectra.

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.

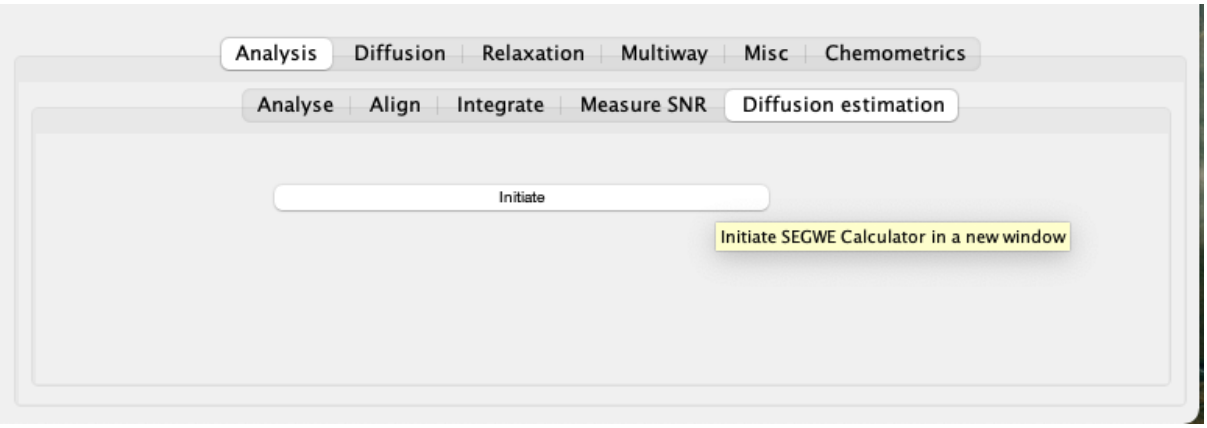
Pressing the *Run* button will display the result in the MATLAB window (or termnal window for compiled versions)


General NMR Analysis Toolbox (GNAT) unofficial .../7
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Plot Phase FT Correct Array Prune
Analysis Diffusion Relaxation Multiway Misc Chemometrics
Analyse Align Integrate Measure SNR Diffusion estimation
Process
Run
Peak pick
Peak pick
Integrals
Noise region
Set
Clear
Show
Separate plot
Standard plot
Publication plot
Close all
Shift range
1.0542
High
-2.5464
Low
Horizontal scale
ppm
Hz
Display
FID
Phase
Spectrum
Abs.
Zoom / Pan
Zoom
Pan

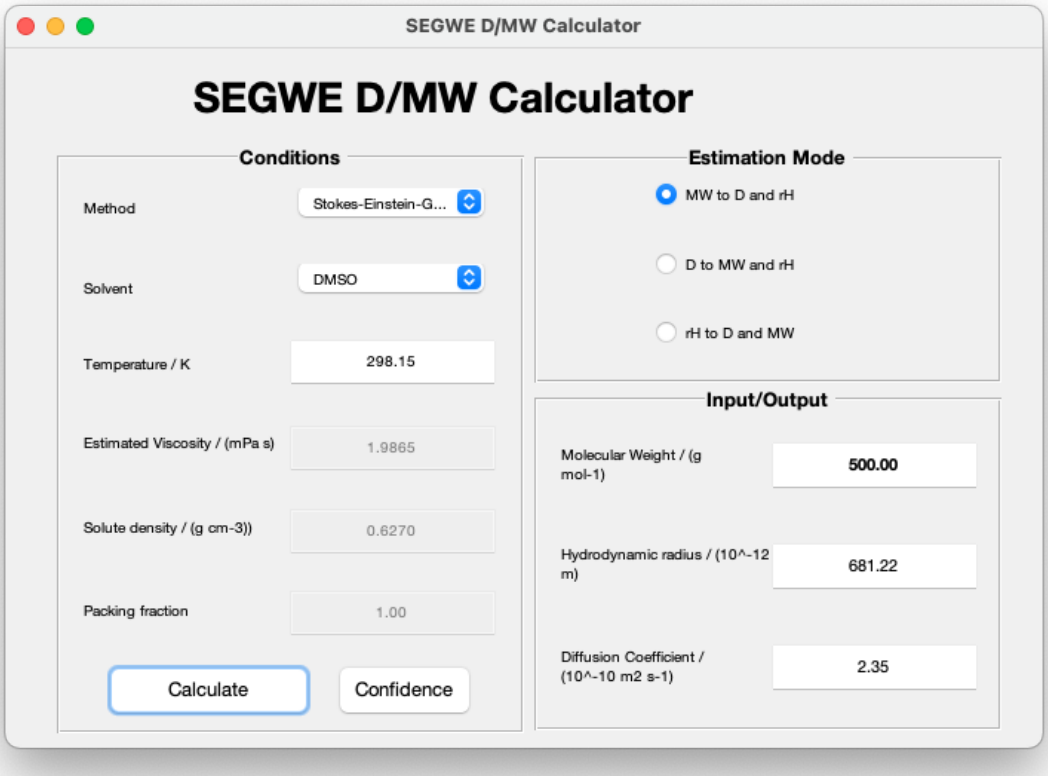

Command Window
******* SNR for current spectrum [1] *******
Peak number Frequency SNR
1 -0.00388 2142.76
2 0.87656 2159.32
>>

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.



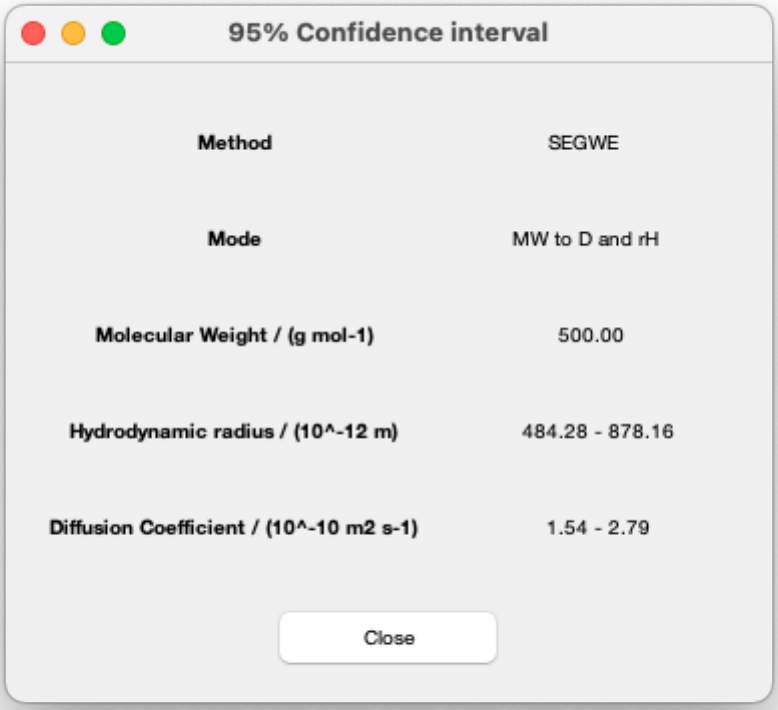
The theory and practicalities behind these calculation is given in the following papers:

(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

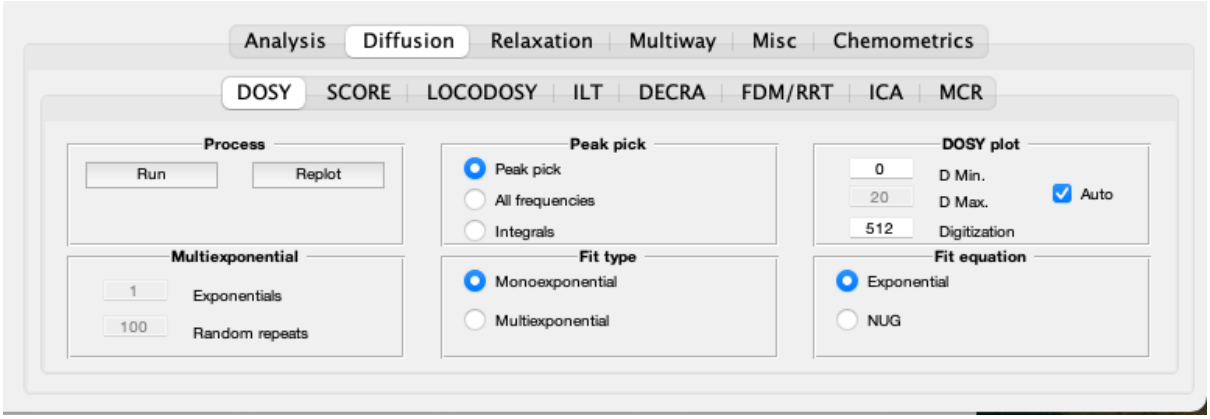
In this section the mode of estimation. One can choose either molecular weight (MW) , the diffusion coefficient (D), or the hydrodynamic radius (rH) of the solute to estimate the other two.

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.

6.15 Diffusion

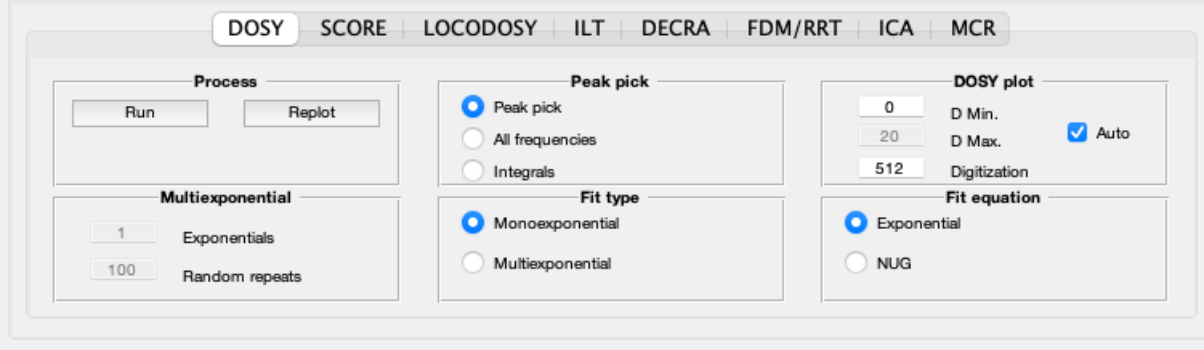
This tab contains the main tools for analysing diffusion NMR data in GNAT.



6.15.1 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.)

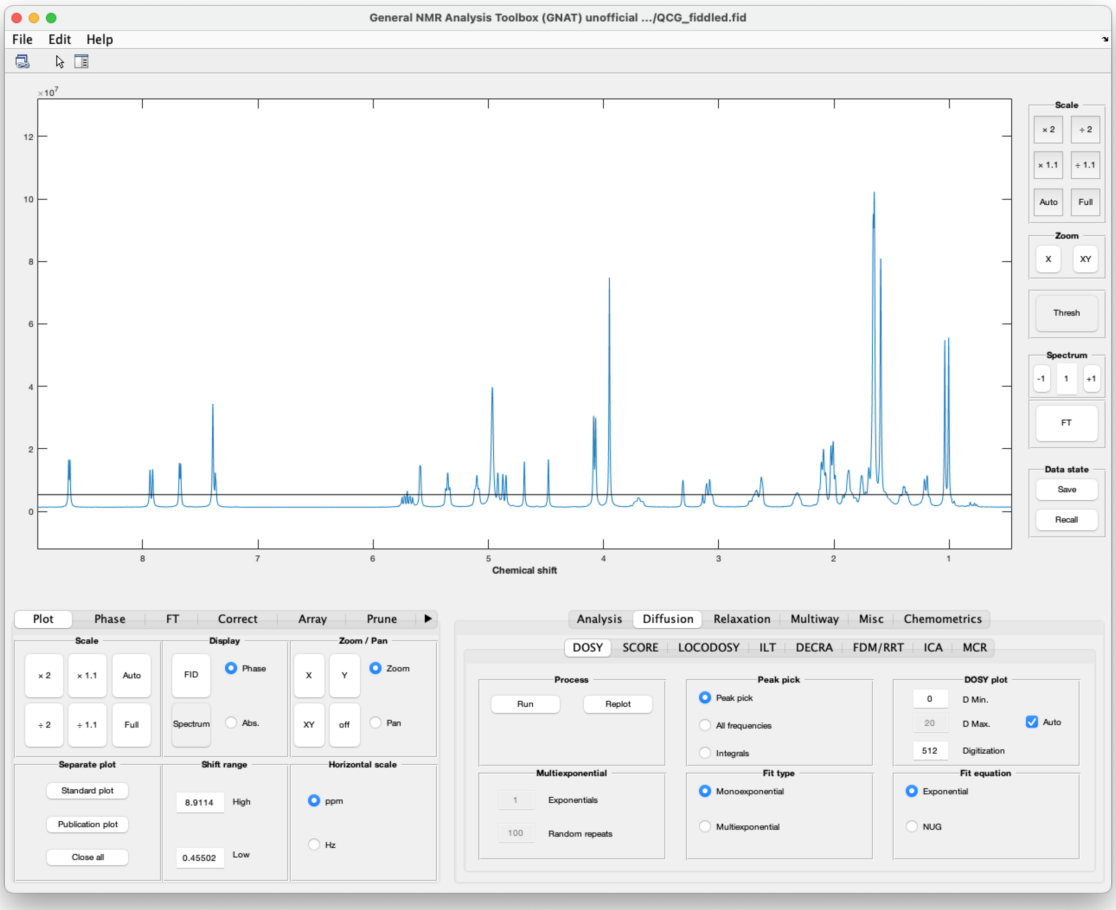
> **Note**
>
> **Quickstart**: zoom into the part of the spectrum you are interested in, set a threshold with the *Thresh* button (right part of the GUI) and press the *Run* button

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.


General NMR Analysis Toolbox (GNAT) unofficial .../QCG_fiddled.fid
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Plot Phase FT Correct Array Prune
Analysis Diffusion Relaxation Multiway Misc Chemometrics
DOSY SCORE LOCODOSY ILT DECRA FDM/RRT ICA MCR
Process
Run
Replot
Peak pick
Multiexponential
Fit type
Fit equation
DOSY plot
D Min.
D Max.
Digitization
Auto
Exponentials
Random repeats
Monoexponential
Multiexponential
Exponential
NUG
Separate plot
Standard plot
Publication plot
Close all
Shift range
High
Low
Horizontal scale
ppm
Hz
Display
Zoom / Pan
Phase
Abs.
Zoom
Pan

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 usedd 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 $10^{-10}\ m^2 s^{-1}$)

D max sets the highest diffusion coefficient to be displayed (in $10^{-10}\ m^2 s^{-1}$) 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 available when *Multiexponential* is selected in the *Fit type* section, is for parameters related to multi exponential fitting of DOSY data. That means 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:

(2) 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)

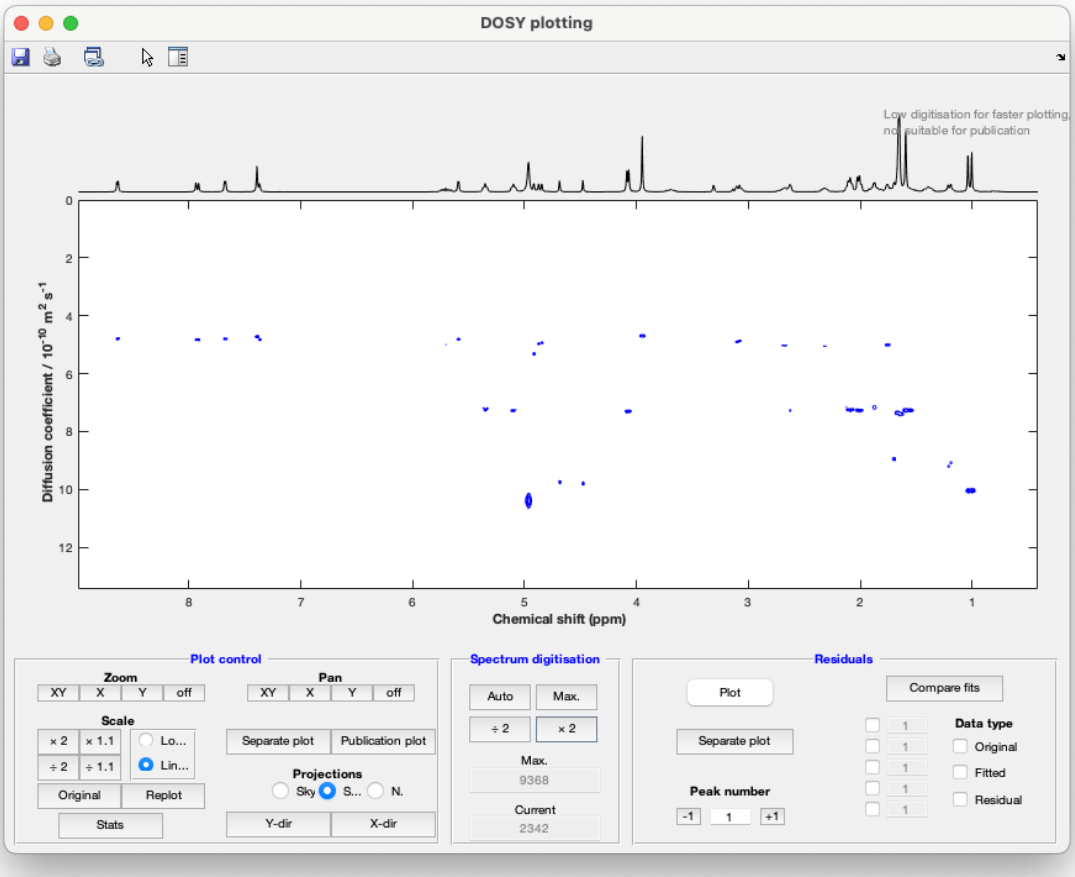
Fit equation

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

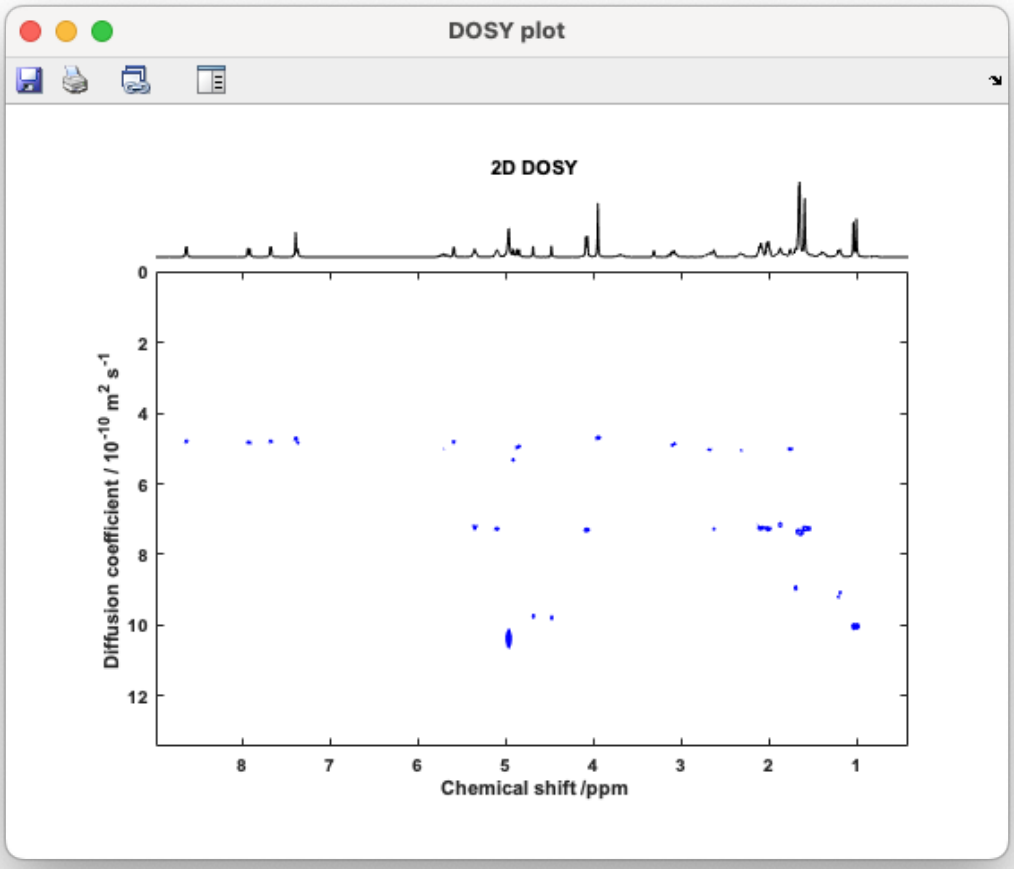
Controls to scale the contours in the 2D DOSY plot . You can multiply or divide by a factor 2 or 1.1. The **Original** button resets to the initial scale and the **Replot** button plots the whole initial spectrum.

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.

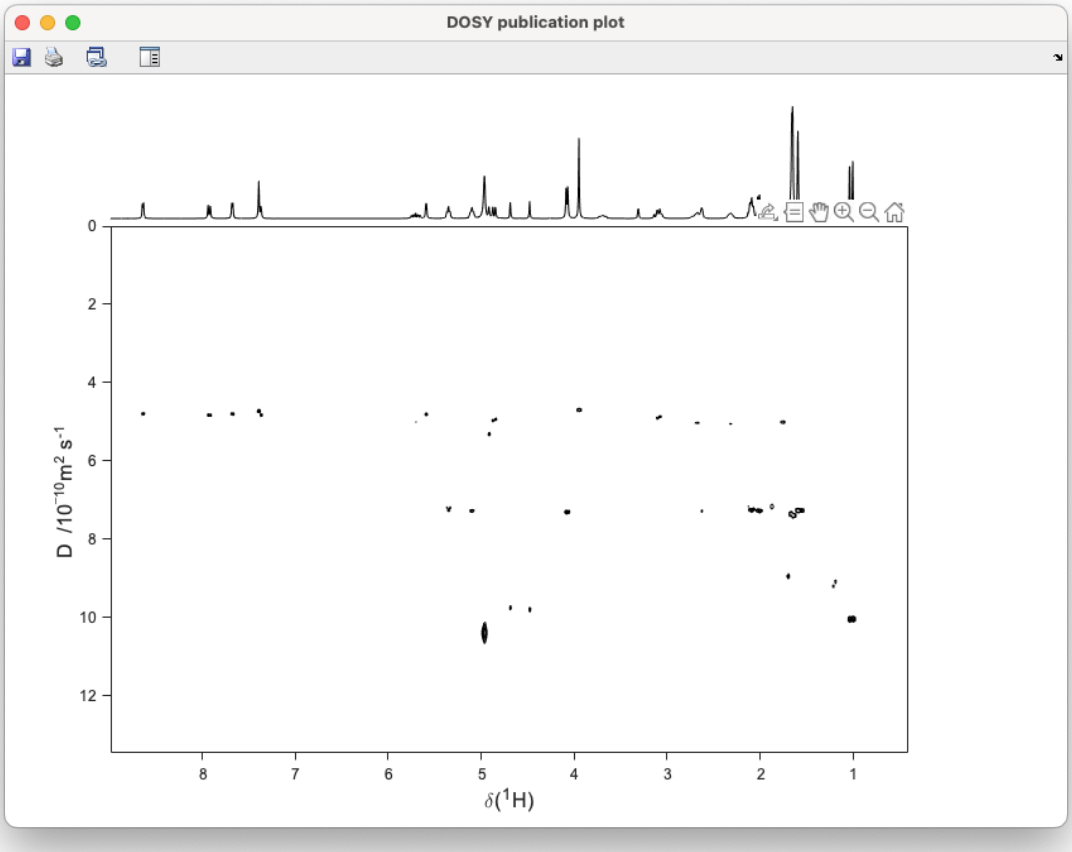
Separate plot

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.

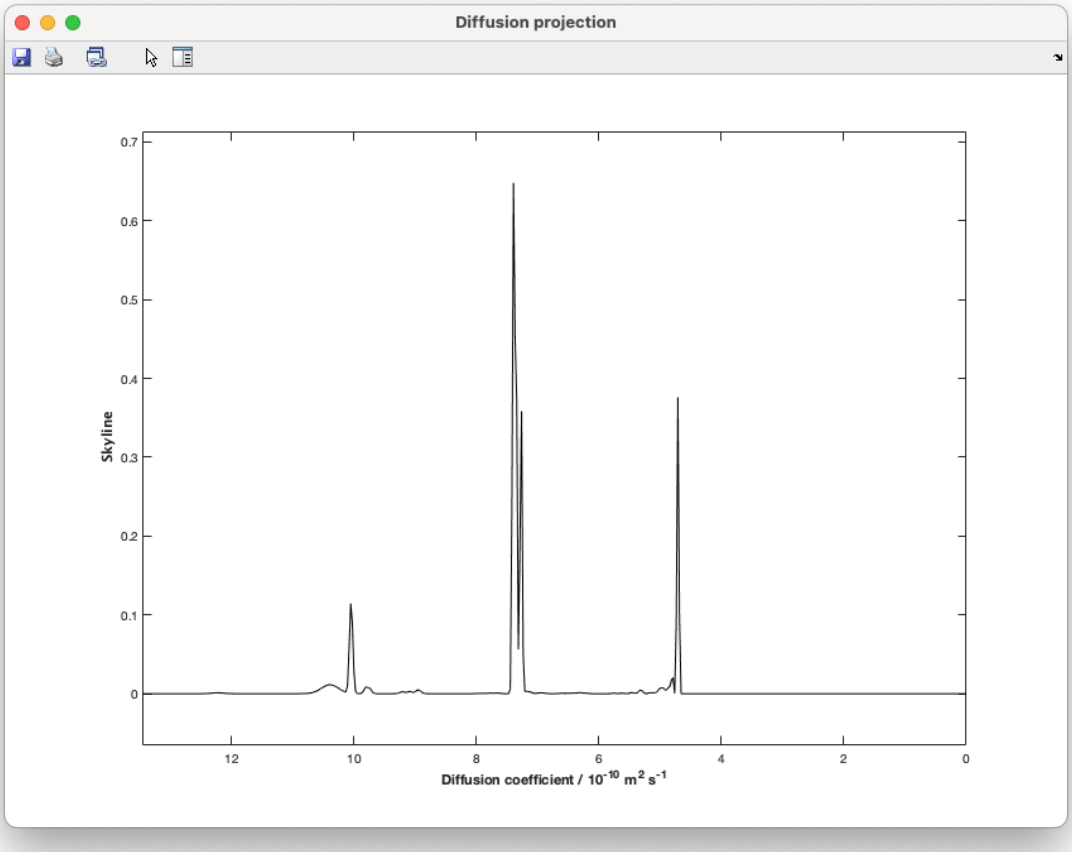
The plots can be saved in the available MATLAB formats (e.g. fig, svg, eps, png, jpg, pdf)


DOSY publication plot
0
2
4
6
8
10
12
D /10⁻¹⁰m² s⁻¹
8
7
6
5
4
3
2
1
δ(¹H)

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)

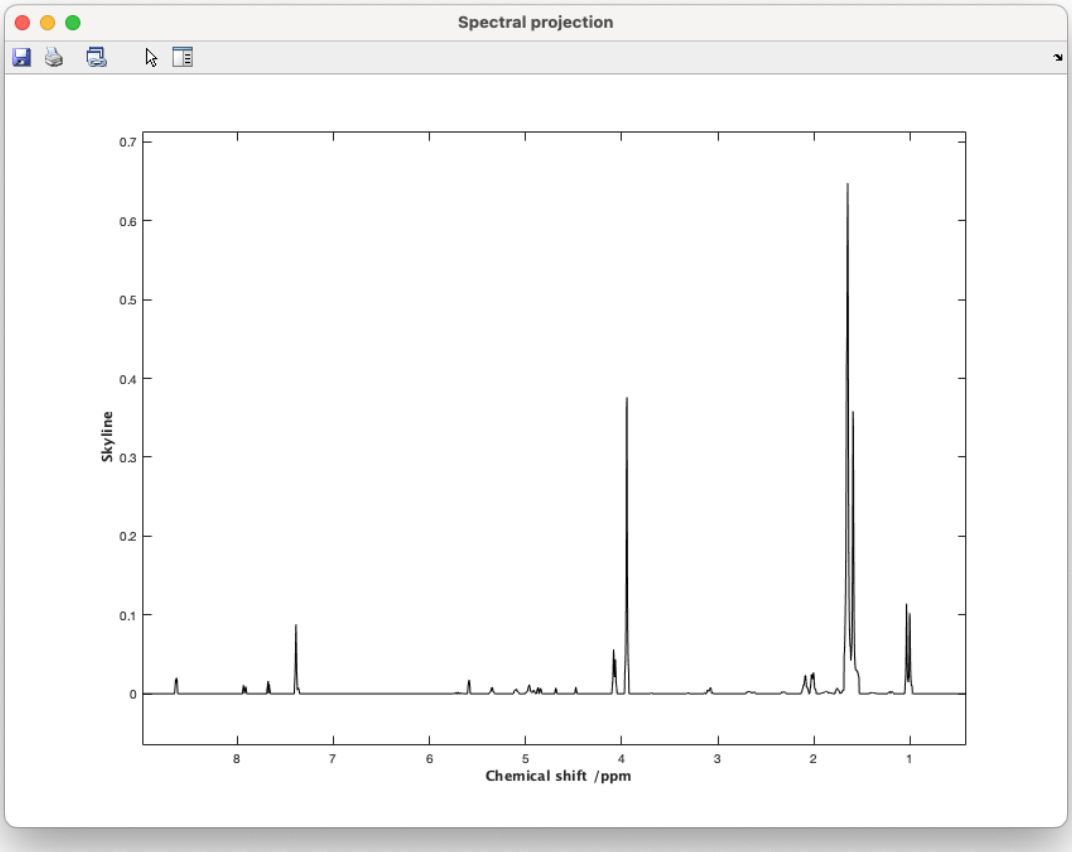
Spectrum digitization section

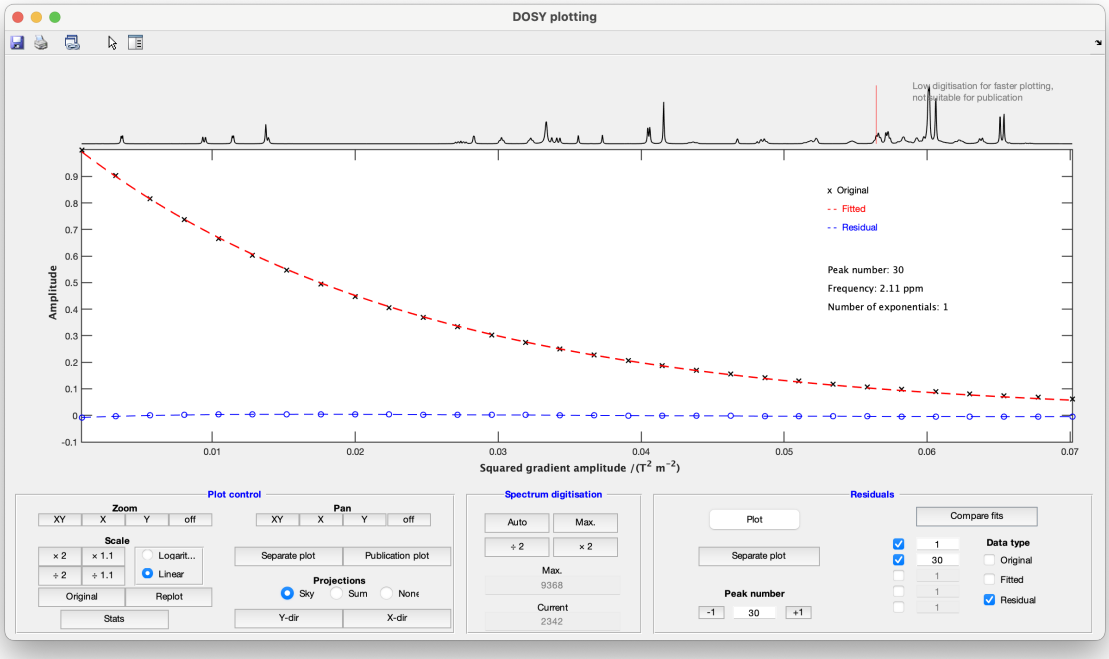
Matlab is unfortunately very slow at plotting 2D data. Therefore GNAT has an automatic downsampling to something closer to the screen resolution. This speeds up the plotting, but can also sometime cause distortions in the plot. If downsampling is present a grey warning text is displayed in the top right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

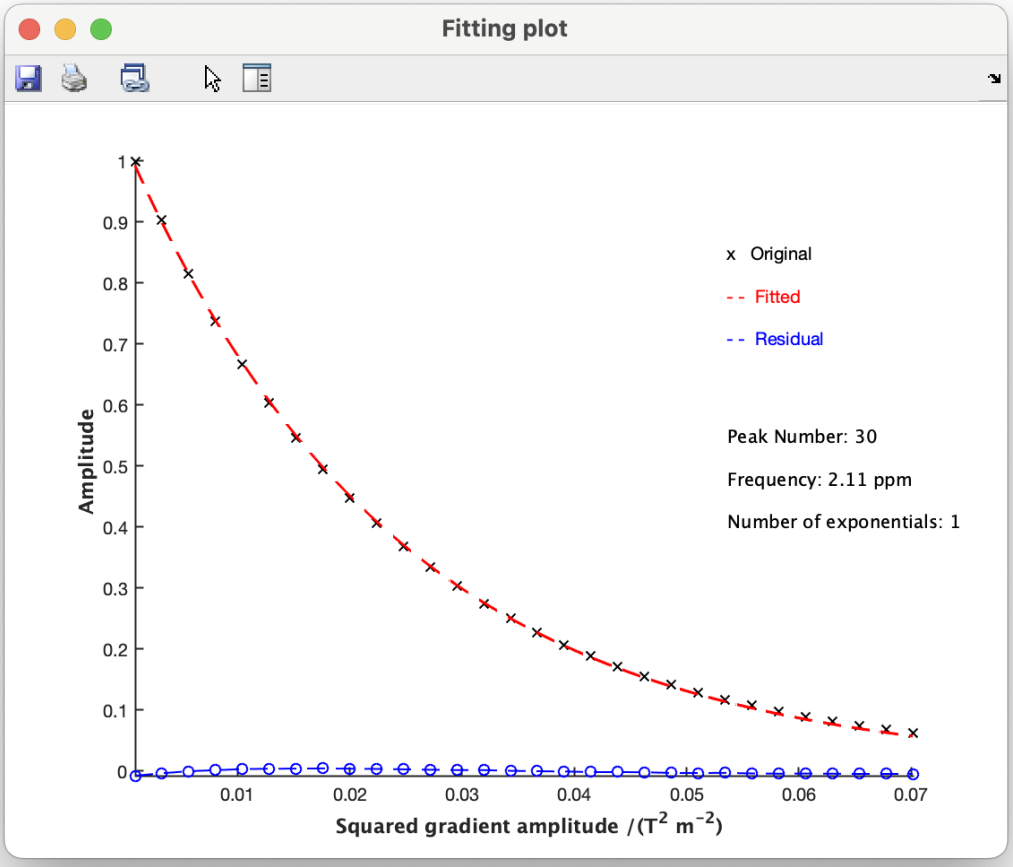
Residuals section

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


Spectral projection
Skyline
Chemical shift /ppm


DOSY plotting
Low digitisation for faster plotting, not suitable for publication
x Original
Fitted
Residual
Peak number: 30
Frequency: 2.11 ppm
Number of exponentials: 1
Amplitude
Squared gradient amplitude /(T² m⁻²)
Plot control
Zoom
Pan
Scale
Projections
Spectrum digitisation
Residuals
Data type
Original
Fitted
Residual
Peak number

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.



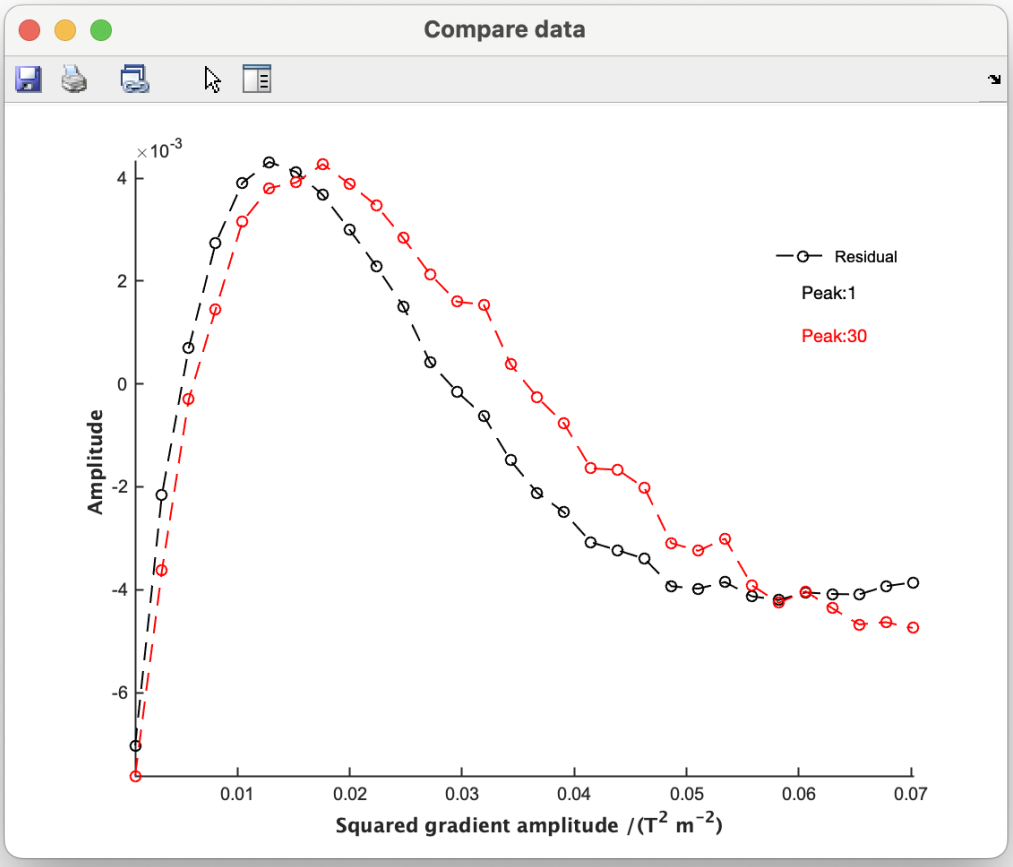
Fits and residuals can also be compared for up to 5 peaks, and the user can select any combination of raw data (original), fit and residuals. The comparison is displayed using the **Compare fits** button. Below is a comparison of the residuals for peak 1 and peak 30.

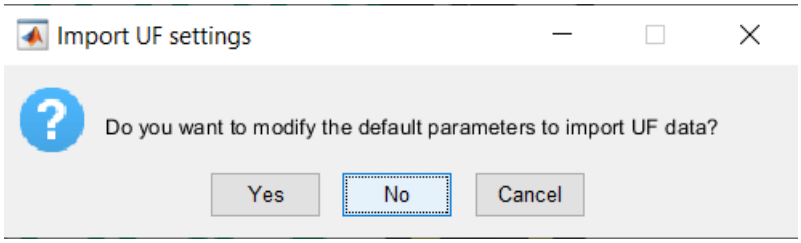
SPEN DOSY

> **Warning**
>
> For the current version, the functions listed here work for Windows operating systems. If you encounter any problems on other systems, please report them to mathias.nilsson@manchester.ac.uk

Loading procedure

To import ultrafast experiments, the data must be loaded using the menu: file > Import > Bruker ultrafast. The user is initially directed to import SPEN DOSY data. It is possible to change the default parameters for importing ultrafast data in the settings section. To do this, the user must press Yes in the Import UF settings window (see the *Settings* section.):


Compare data
×10^-3
4
2
0
-2
-4
-6
Amplitude
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Squared gradient amplitude /(T^2 m^-2)
Residual
Peak:1
Peak:30


Import UF settings
Do you want to modify the default parameters to import UF data?
Yes
No
Cancel

The user must then choose the directory containing the reference NMR experiment. If the data has not been sheared, a window should pop up, where the user needs to shear data in kt space (Figure 1, left) in case the acquisition gradient were not exactly equals. Otherwise it leads to shifting peaks.

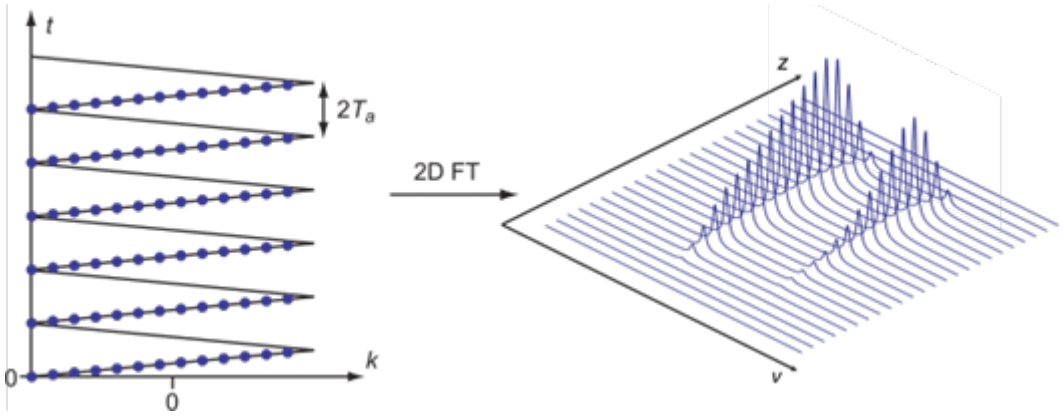


Fig. 5: Figure 1. On the left is the acquisition trajectory in a two-dimensional (k, t) space, obtained with echo-planar spectroscopic imaging. Applying a double Fourier transform results in the spectroscopic image shown on the right.

More information about these plots can be found here:

1. J.-N. Dumez, Prog. Nucl. Magn. Reson. Spectrosc. 2018, 109, 101–134, DOI: 10.1016/j.pnmrs.2018.08.001.

The first window to pop up aims at align the echoes from the SPEN DOSY experiment's data that were just loaded (Figure 2). Then, if the user chose to apply an acquisition or chirp + profile correction, a second window will pop up. It will be asking again the user to shear these data. When the user validates the echoes alignment, a "shearval.txt" file is created in both the SPEN DOSY expno folder and the reference experiment. This file contains a unique numerical value. When the user wants to processed again its data, as a shearval.txt file is already existing, is it asked whether if they want to reuse the same shear value, or if they want to perform the data shearing again.

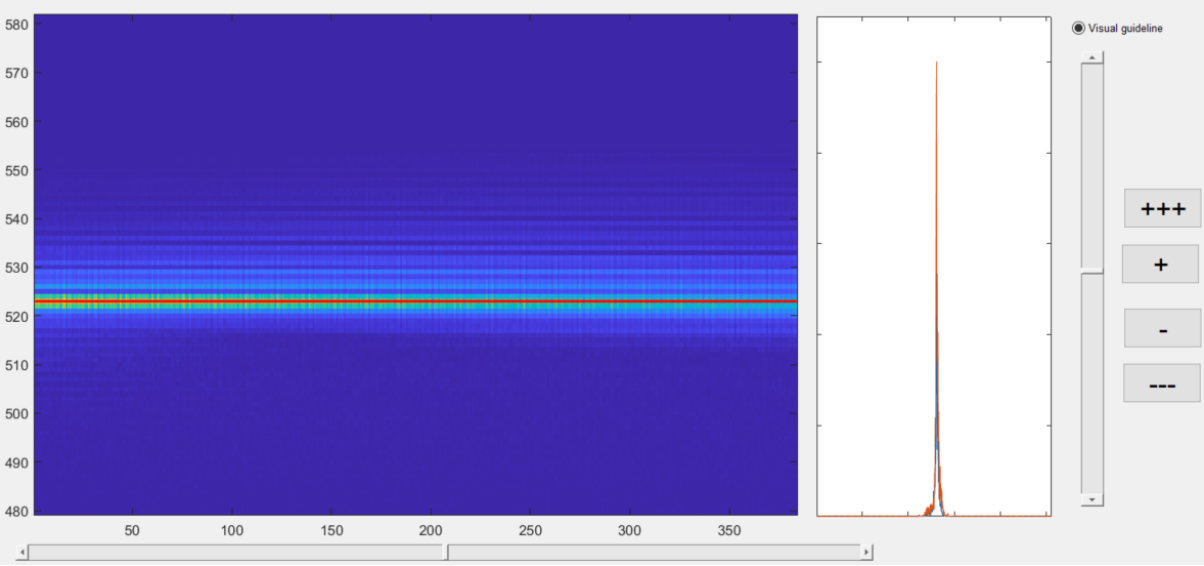


Fig. 6: Figure 2: SPEN DOSY shearing GUI to align the echoes with the ideal format (red line).

SPEN DOSY experiments are designed for magnitude processing (Figure 3a) that usually goes with a sine apodisation window in the t dimension. In the GNAT, the preferred phase mode of the data should

be “mc” in that case. In that case, no phasing correction is asked and the next step for the user is to select the length of the diffusion profile..

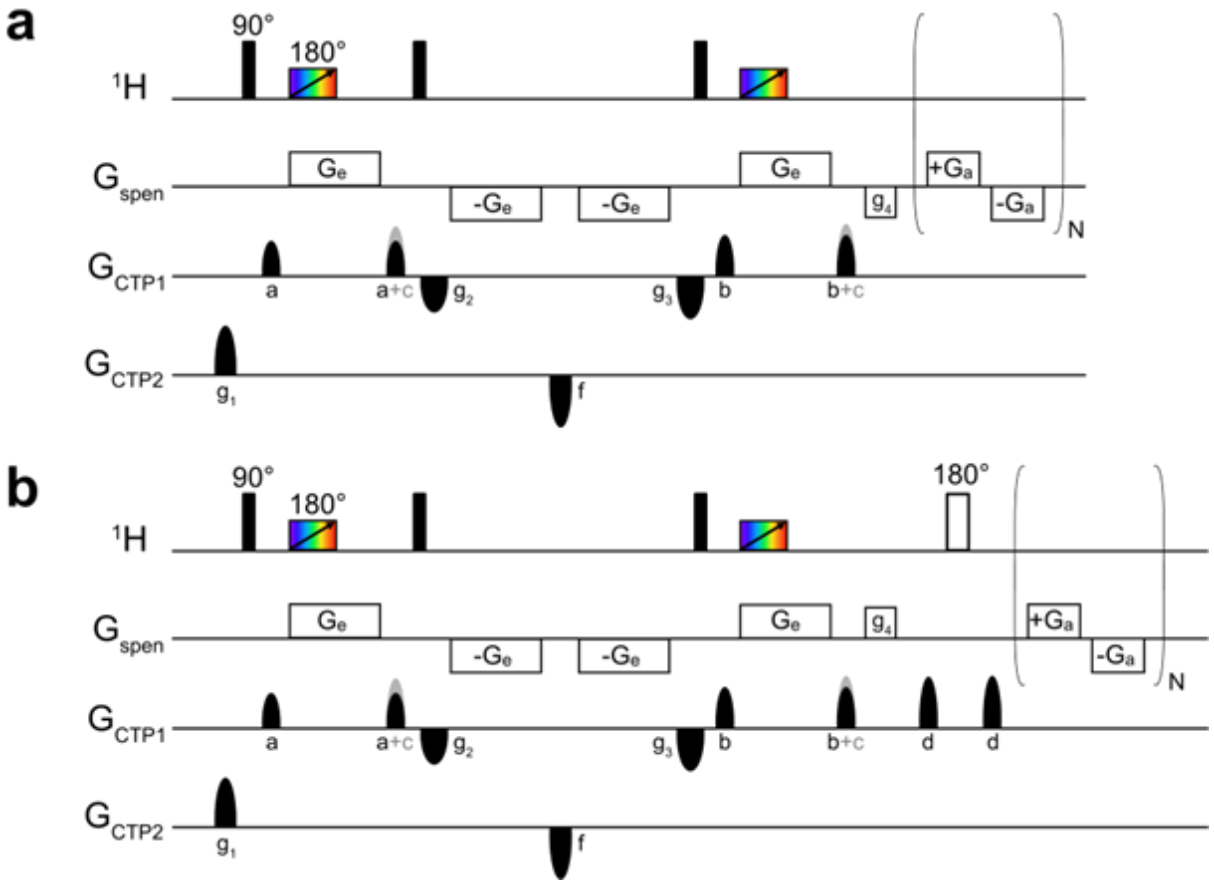


Fig. 7: Figure 3: SPEN DOSY pulse sequences, (a) for magnitude processing and (b) for pure-absorption processing. The black filled half-ellipses are CTP gradient pulses. Narrow black rectangles correspond to 90° pulses. The gradient coloured rectangles represent chirp pulses. White-filled rectangles represent are rectangular magnetic-field gradient pulse.

However, in order to gain in resolution, it is possible to design a SPEN DOSY experiment for pure-absorption processing (Figure 3b). Experimentally, a hard 180° pulse is added between the end of the second encoding gradient and the start of the EPSI. Thus, the chemical shift evolution is fully refocused at the moment of the first gradient echo of the train of echoes collected by the EPSI. With this type of SPEN DOSY experiment, the “pk” processing option must be chosen and a phasing correction is preferred. The apodisation usually consists in a cosine window in that case. The following section explains how this phasing step could be done in GNAT:

Select the pivot point by clicking, first, on the `Select pivot` button, and then on the spectrum (Figure 4). Then, the phase correction can be performed with the two sliders under the spectrum. The top slider corresponds to the 0-order phase correction, while the bottom one is the first-order phase correction. When the phase correction is satisfactory, click and `Save and close`. The phasing parameters will be saved, and could be reused for the same data. Finally, the phase correction of the reference experiment will also be asked.

Then, the user needs to select the length of the diffusion profile they want to process during the ST fit. At this time of the data processing, a 2D Fourier transform is applied on the data in the kt space, so that an image of the data could be obtained. Such image should pop up, with scaled colors. The user needs to select, by clicking 1 time each resonance they want to extract the diffusion and the reference profile (if

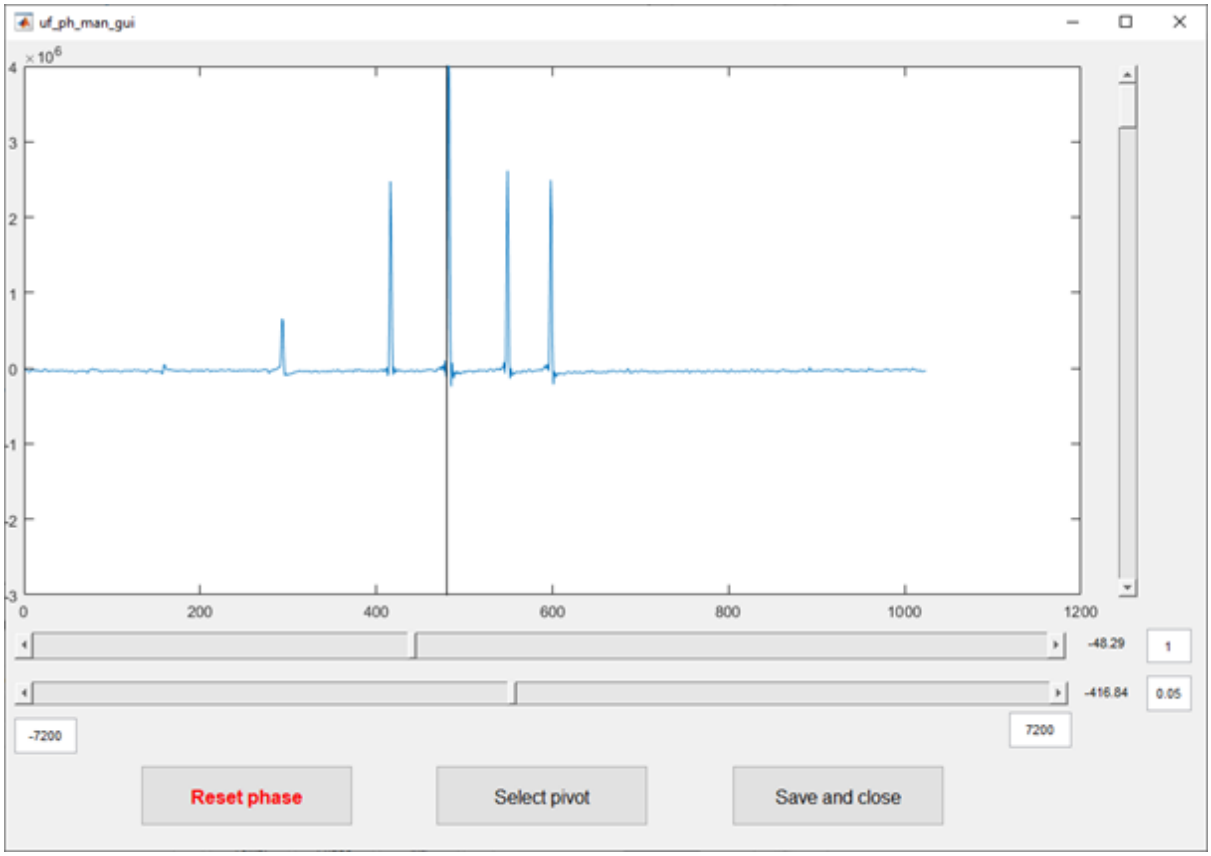


Fig. 8: Figure 4: Phase correction GUI for pure-absorption SPEN DOSY data.

activated) from. It could be one or several resonances.

Then, the diffusion and the acquisition (if activated) profiles that come from the selected resonances should be visualised by the user. This time, they need to select the ends of the diffusion and reference profiles, so that these selected limits are valid for both the diffusion profile and the reference profile. Be careful about always perform the selection in increasing order of positions. After this step, a spectrum should appear in this main GNAT window.

Experimental details

The reference spatial profiles ideally need to be recorded with a 10% H2O and 90% D2O sample. Two reference profiles are relevant to be recorded: - Sref1, that reflects the probe's sensitivity, it is obtained by the pulse sequence from Figure 3a. Note that the acquisition block may also consist of a single readout gradient instead of echo-planar spectroscopic imaging. - Sref2, that corresponds to the effect of the pair of the chirp pulses on the magnetisation's magnitude. Sref2 is obtained by dividing the outcome of this pulse sequence by Sref1. When a linear sweep is used, Sref2 was found to be close to 1. But for a quadratic sweep, such correction is important.

The SPEN DOSY data collected with pulse sequences for magnitude processing (Figure 3a), or for pure-absorption processing (Figure 3b) need to be processed, respectively, in magnitude mode or phase mode.

Once the field gradient is mapped experimentally, the retrieved map is fitted to a polynomial function, that yields polynomial coefficients. The SPEN DOSY gradient strength list, as a function of position, is then be multiplied in GNAT by those coefficients to account for the gradient non-uniformity. The value of G is instead expressed as a function of position (z-position for this example):

K(z): effective gradient area

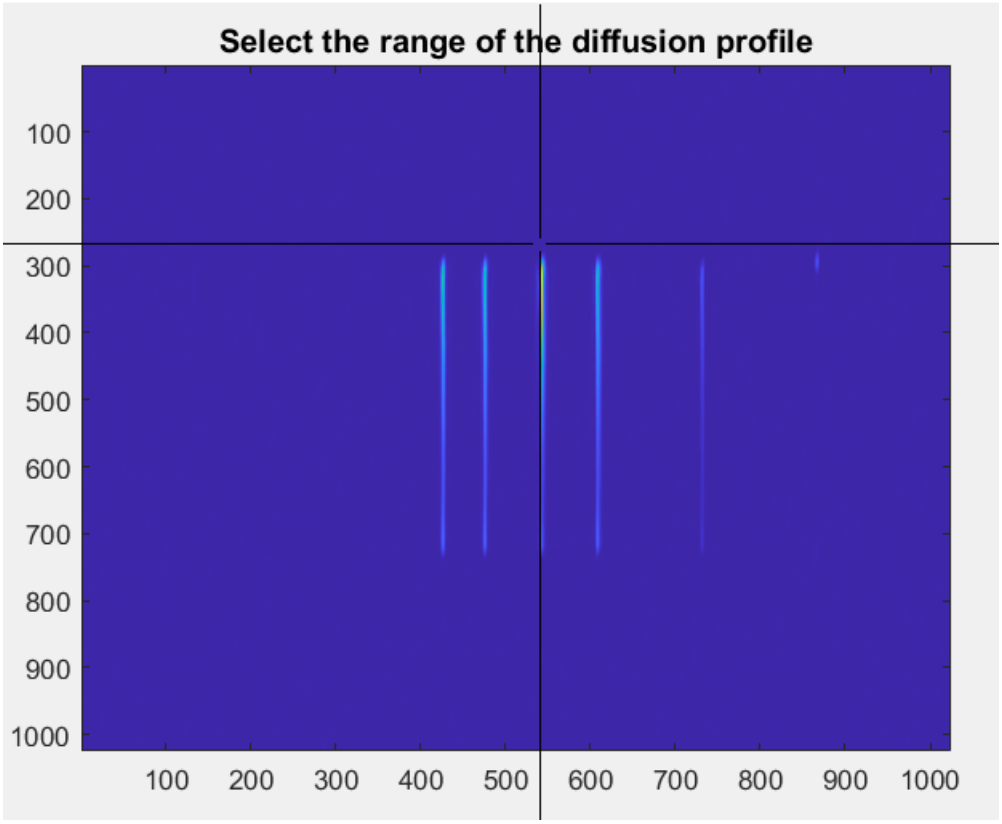


Fig. 9: Figure 5. GUI for selection of the diffusion profile. After selecting the profiles, the user can press Enter to continue.

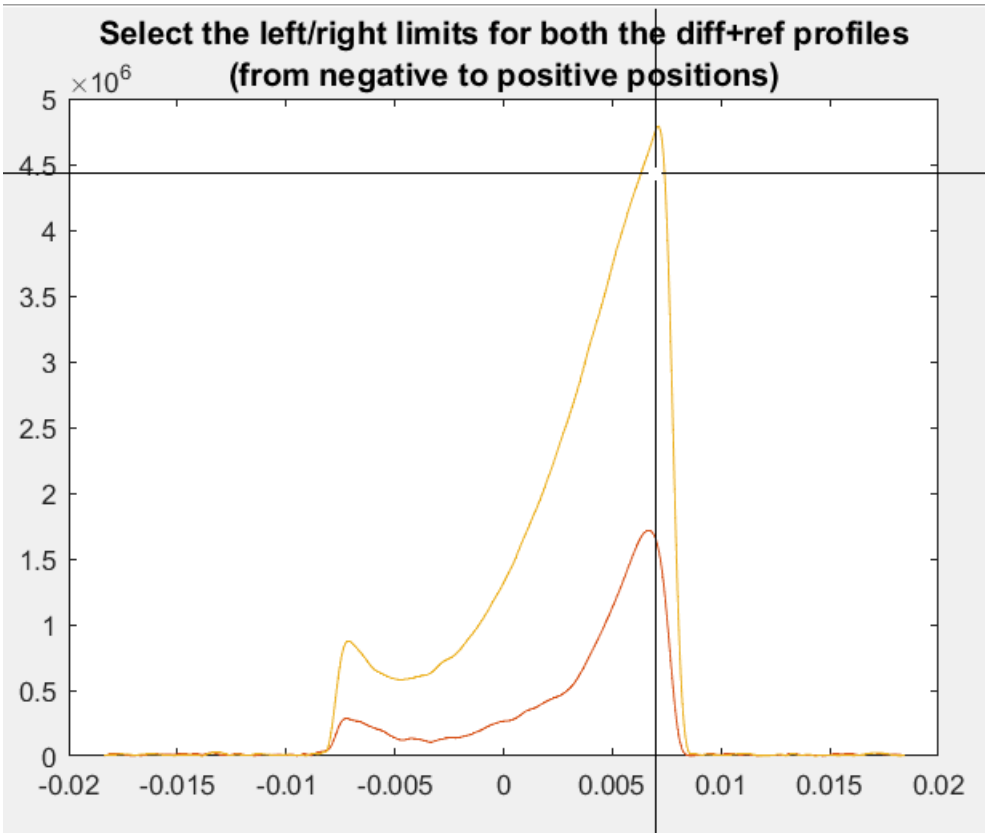


Fig. 10: Figure 6. GUI for selection of the limits for the acquisition and diffusion profile

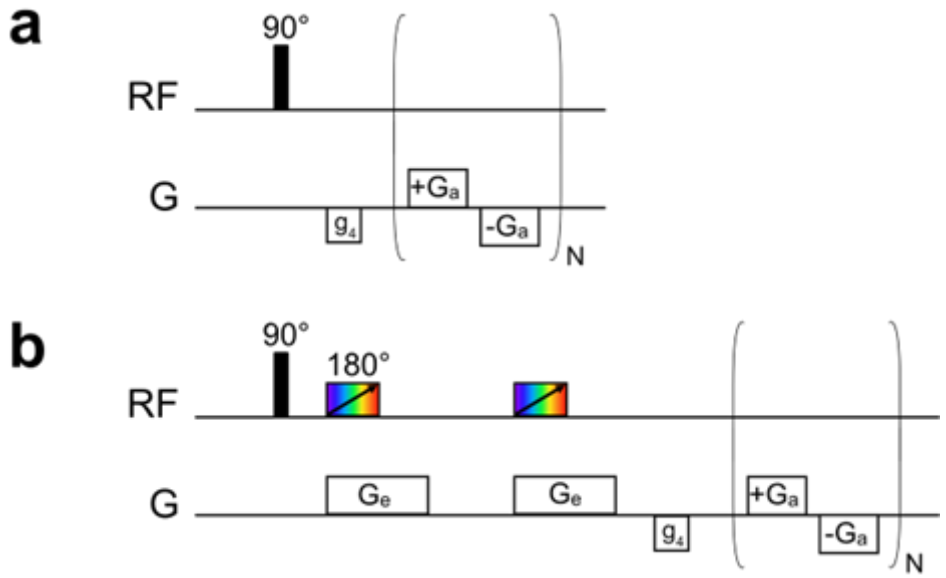


Fig. 11: Figure 7: Pulse sequences to obtain reference spatial profiles. The pulse sequence (a) gives an estimate of the probe's sensitivity Sref1, while (b) is used to retrieve a combination of Sref1 and Sref2, where Sref2 represents the effect of the pair of frequency-swept pulses on the magnetisation's magnitude.

$$K(z) = \frac{\partial \Phi}{\partial z} = -\gamma \boldsymbol{G_e}(\boldsymbol{z}) T_e \left(\frac{2z}{L_s} - 1 \right)$$

:math:\gamma``: gyromagnetic ratio

Ge: encoding gradient strength

Te: encoding gradient duration

Ls: length of the encoded area

See also:

2. M. A. Connell, P. J. Bowyer, P. Adam Bone, A. L. Davis, A. G. Swanson, M. Nilsson, G. A. Morris, J. Magn. Reson. 2009, 198 (1), 121–131, DOI: 10.1016/j.jmr.2009.01.025.

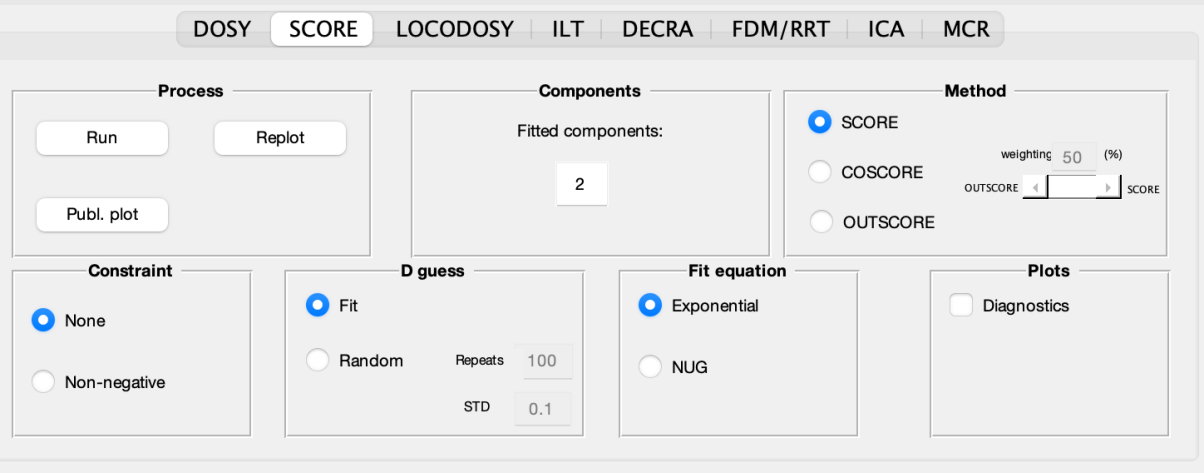
3. B. Lorandel, R. Mishra, O. Cazimajou, A. Marchand, A. Bernard, J.-N. Dumez, J. Magn. Reson. 2023, 107543, DOI: 10.1016/j.jmr.2023.107543.

Miscellaneous

If SPEN DOSY data processing uses non-uniform gradient correction, there can be identical values at different positions for the effective gradient area. As a result, looking at the fitting curves in that case might show experimental values "that are going backwards". This is just a visual issue; the ST fit is still done correctly for each values of G.

SCORE

This is the tab for SCORE processing if diffusion data.



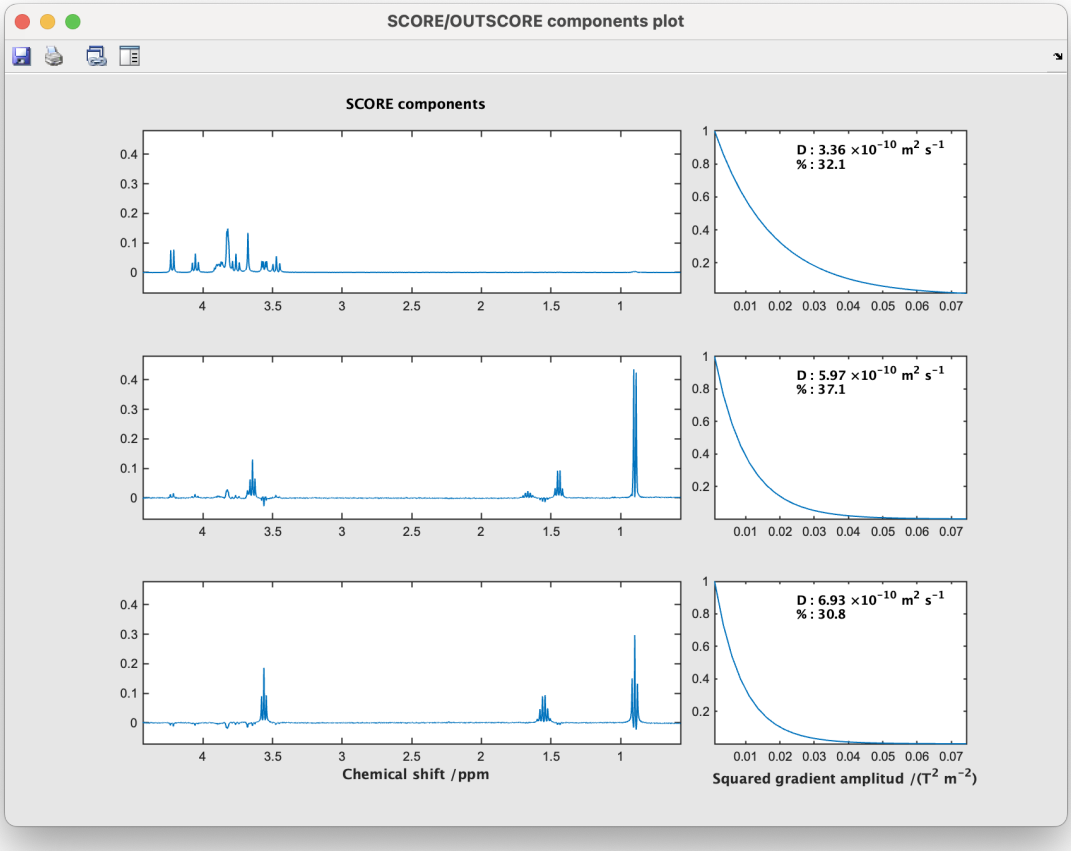
Before processing the data in this tab, make sure that the spectra have been properly preprocessed (phase, baseline correction etc), and that the diffusion parameters are correct (see the *Settings* section.)

Process section

Here the user can access various processing functions. The button *Run* will start SCORE processing with the parameters set in the other sections in this tab, but also e.g., the settings in the *Prune* tab, and in the Diffusion tab

The **Replot** button will plot again the last processed data (e.g. in case it was closed by mistake) The **Publ. Plot** will make a SCORE plot more suitable for publication in reports etc

A three component fit of a mixture of 1-propanol, 1-pentanol and sucrose.


SCORE/OUTSCORE components plot
SCORE components
D : 3.36 ×10−10 m2 s−1
% : 32.1
D : 5.97 ×10−10 m2 s−1
% : 37.1
D : 6.93 ×10−10 m2 s−1
% : 30.8
Chemical shift /ppm
Squared gradient amplitud /(T2 m−2)

Components section

Here the user can determine the number of SCORE components to fit. For unknown mixtures this can often be explored by trial and error.

Methods section

Here the user can choose between different minimization criteria for SCORE analysis. SCORE [2] minimizes the residuals, OUTSCORE minimizes cross-talk [1] and COSCORE [unpublished] uses a combination of the two with the weighting factor det with the slider.

Constraint section

Here the user can choose to constrain the spectrum to be non-negative. In most cases this makes physical sense, but it also slows down the algorithm very significantly. Commonly non-negative constraints is not needed, or even produced unwanted results, but sometimes it can be really helpful.

D guess section

The SCORE algorithm requires initial guesses for the diffusion coefficients. The default (*Fit*) is automatic based on the D value obtained for a monoexponential fit of the whole spectrum. This works well most of the time, but sometimes the algorithm can end up in a local minimum. To test that the user can use the *Random* function where a number of random repeats are done and the values that gives the best fit (lowest residual) is chosen. The repeats are a Gaussian distribution centred on the monoexponential fit value. The number of repeats is set in the *Repeats* box and the and the width of the Gaussian is set by the *STD* value.

Fit equation section

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.

Plots section

Here the user can opt in for additional plots. By selecting the tick box *Diagnostics* several plots for investigating the fit are provided.

The *Residuals* plot shows the total residual in the spectrum and decay.

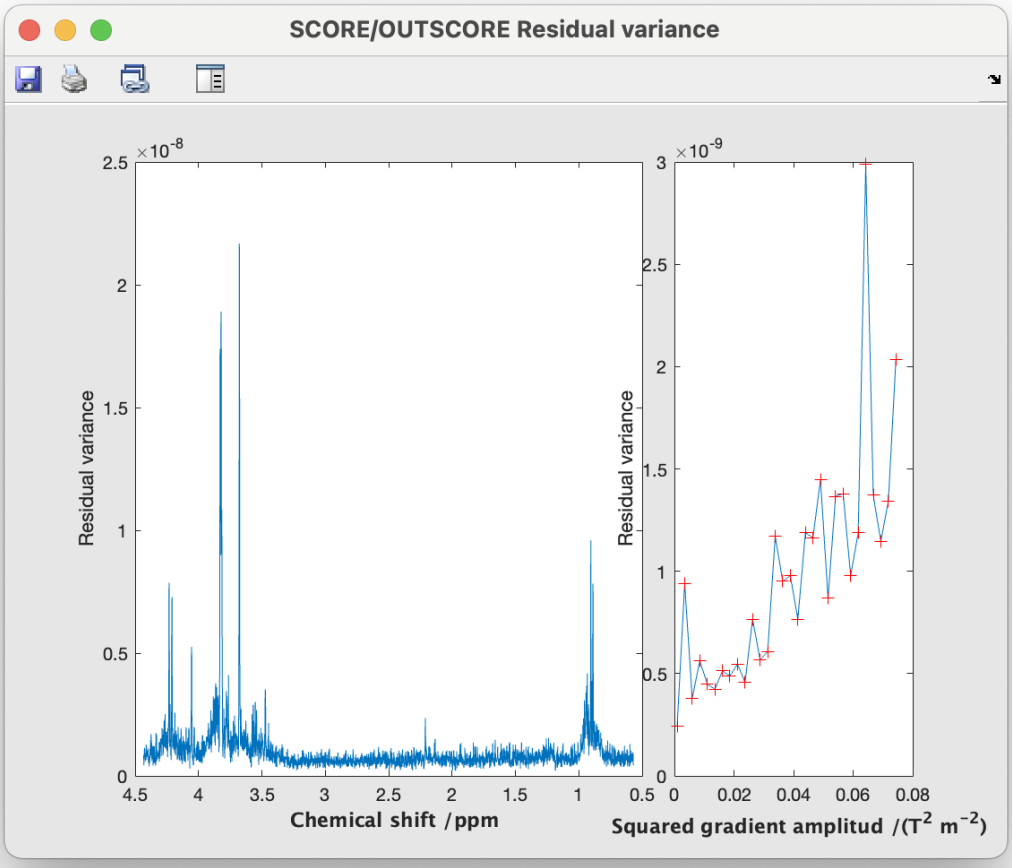
The *Leverage* plot shows the leverage (i.e. how much influence it has on the model) of each data point in the fit.

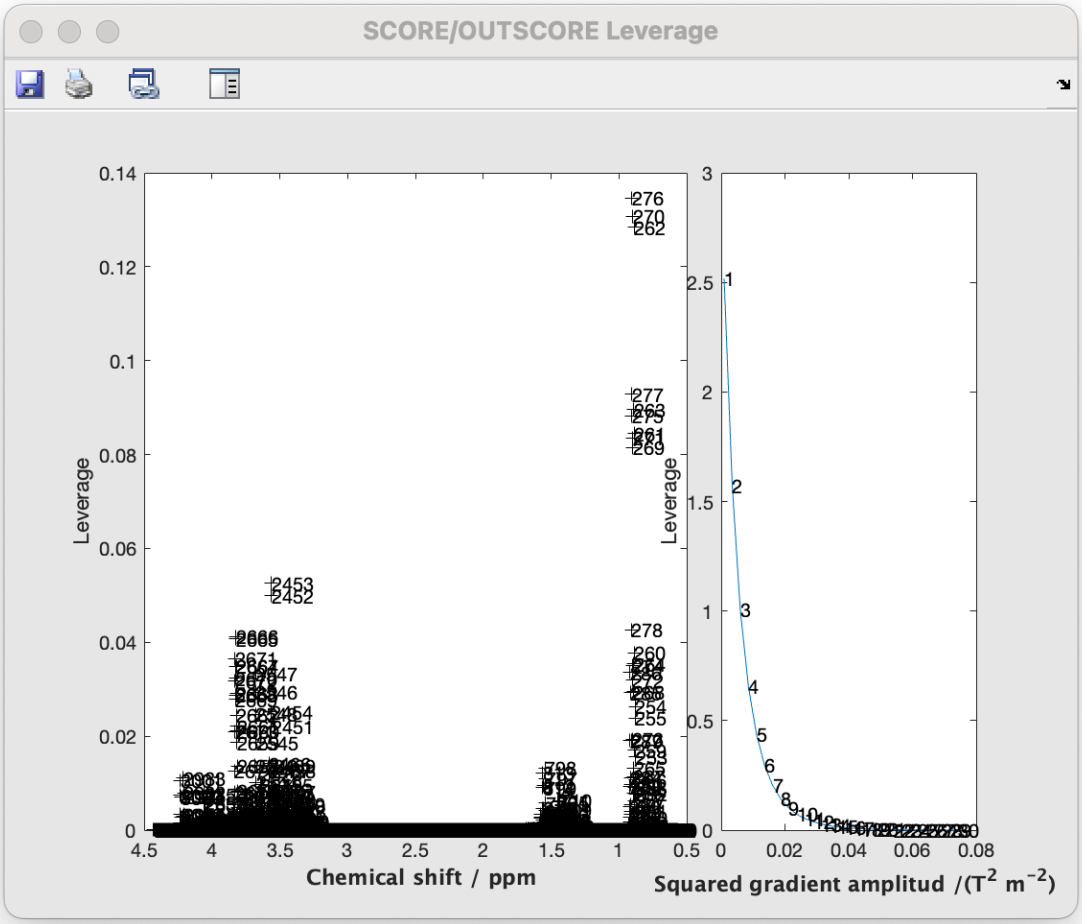
The *Residuals vs Leverage* shows the correlation between residuals and leverage. A data point that had both high residual and leverage is a suspected outlier. (e.g. it could be a gradient level that suffers from unwanted refocussing)

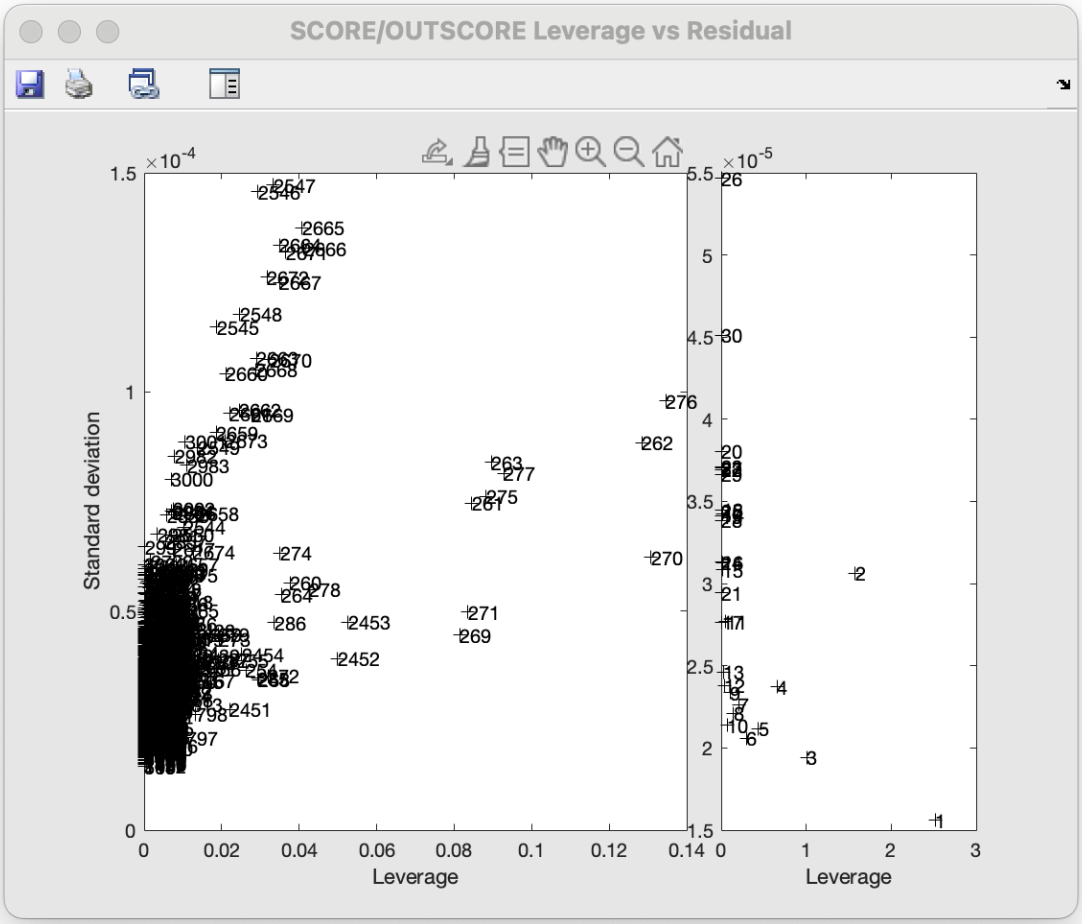
References

Here are some of the papers related to SCORE processing.

(1) Colbourne, A. A.; Meier, S.; Morris, G. A.; Nilsson, M. Unmixing the NMR spectra of similar species - vive la difference. Chemical Communications 2013, 49 (89), 10510-10512. DOI: 10.1039/c3cc46228e.


SCORE/OUTSCORE Residual variance
×10^-8
Residual variance
Chemical shift /ppm
×10^-9
Residual variance
Squared gradient amplitud /(T^2 m^-2)


SCORE/OUTSCORE Leverage
Leverage
Chemical shift / ppm
Leverage
Squared gradient amplitud /(T2 m-2)


SCORE/OUTSCORE Leverage vs Residual
Standard deviation
Leverage
Leverage

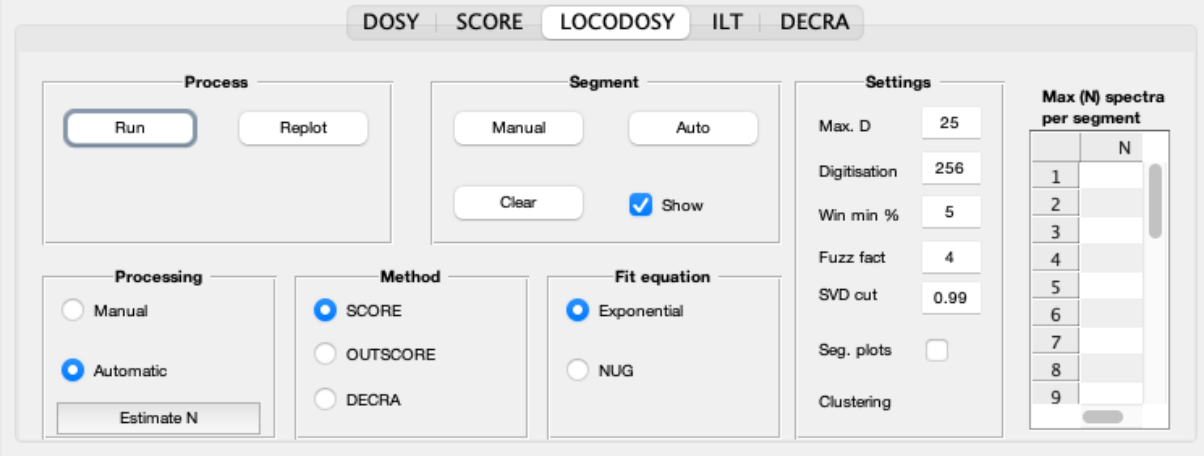
(2) Nilsson, M.; Morris, G. A. Speedy component resolution: An improved tool for processing diffusion-ordered spectroscopy data. Analytical Chemistry 2008, 80 (10), 3777-3782. DOI: 10.1021/ac7025833.

(3) Nilsson, M.; Morris, G. A. Correction of systematic errors in CORE processing of DOSY data. Magnetic Resonance in Chemistry 2006, 44 (7), 655-660. DOI: 10.1002/mrc.1805.

(4) Lorandel, B.; Rocha, H.; Cazimajou, O.; Mishra, R.; Bernard, A.; Bowyer, P.; Nilsson, M.; Dumez, J. Speedy Component Resolution Using Spatially Encoded Diffusion NMR Data. MAGNETIC RESONANCE IN CHEMISTRY 2025, 63 (1), 49-61. DOI: 10.1002/mrc.5488.

LOCODOSY

This is the tab for LOCODOSY 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), click *Auto* in the **Segment** section and press the *Run* button

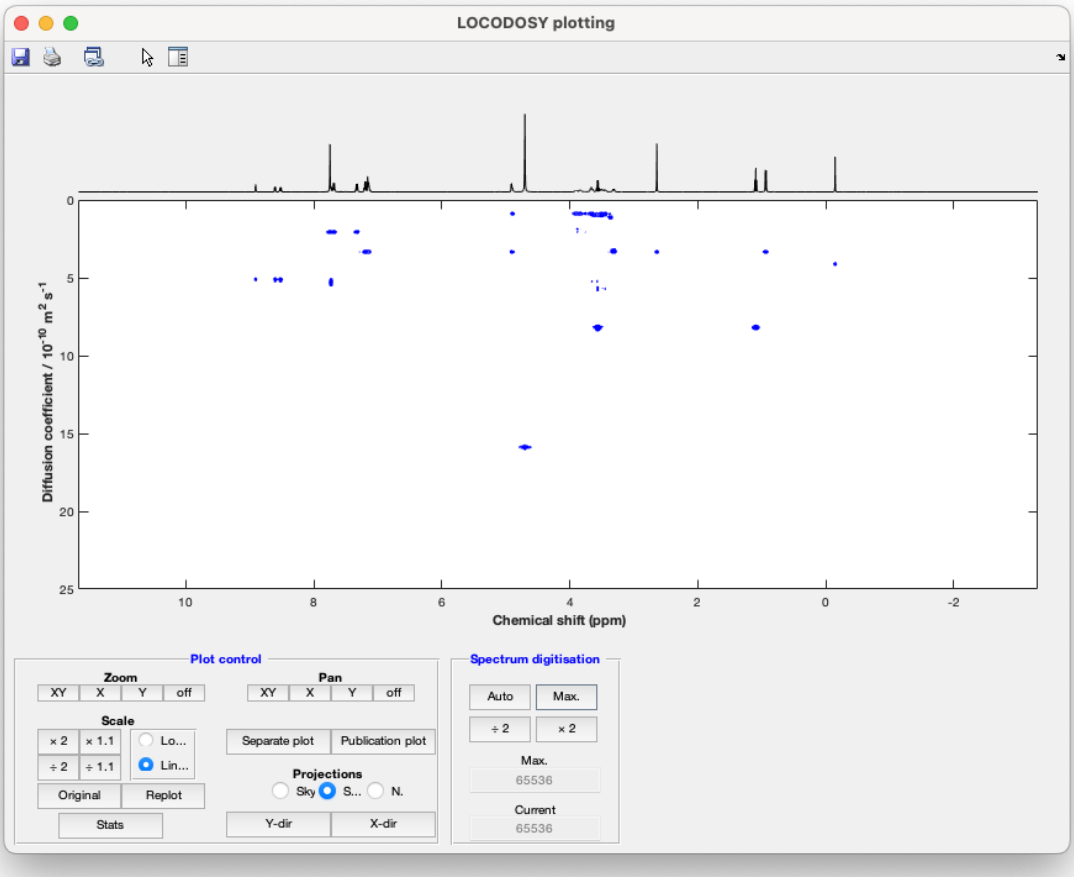
The idea behind LOCODOSY is described in the below article.

(1) Colbourne, A. A.; Morris, G. A.; Nilsson, M. Local Covariance Order Diffusion-Ordered Spectroscopy: A Powerful Tool for Mixture Analysis Journal of the American Chemical Society 2011, 133, 7640.

Process section

Here the user can access various processing functions. The button *Run* will start LOCODOSY 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)


LOCODOSY plotting
Diffusion coefficient / 10^-10 m^2 s^-1
0
5
10
15
20
25
10
8
6
4
2
0
-2
Chemical shift (ppm)
Plot control
Zoom
XY
X
Y
off
Pan
XY
X
Y
off
Scale
× 2
× 1.1
÷ 2
÷ 1.1
Lo...
Lin...
Original
Replot
Stats
Separate plot
Publication plot
Projections
Sky
S...
N.
Y-dir
X-dir
Spectrum digitisation
Auto
Max.
÷ 2
× 2
Max.
65536
Current
65536

Segment section

Here the user can choose how to segment the spectrum for LOCODOSY processing.

The *Auto* butting will do an automatic segmentation, the *Manual* button will let the used decide the segments, and the *Clear* button will clear any previous segments. To show the segment tick the *Show* box. Sometimes the threshold line will cover the segment views, you it may be a good idea to remove that by clicking the *Thresh* button again.

Settings section

Here some parameters for the LOCODOSY plot in the *DOSY Plotting* GUI are set, together with some parameters for the automatic segmentation are accessed.

Max. D sets the highest diffusion coefficient to be displayed (in $10^{-10}\ m^2 s^{-1}$)

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)

Win min % sets the limit for diffusion values larger that Max. D. (It is recommended to leave this at the default)

Fuzz % sets the limit for the fit statistics of diffusion values. (It is recommended to leave this at the default)

SVD cut % determined the cutoff for the automatic determination of number of components per segment. (It is recommended to leave this at the default)

Seg. plots, if ticked all SCORE/OUTSCORE/DECRA plots for each segment are displayed in separate windows.

Processing section

Here the maximum number of components (N) for each section is set. Click the button *Estimate N* to get an automatic determination of N for each segment is performed. This can be manually entered in the *Max (N) spectra per segment* table.

Method

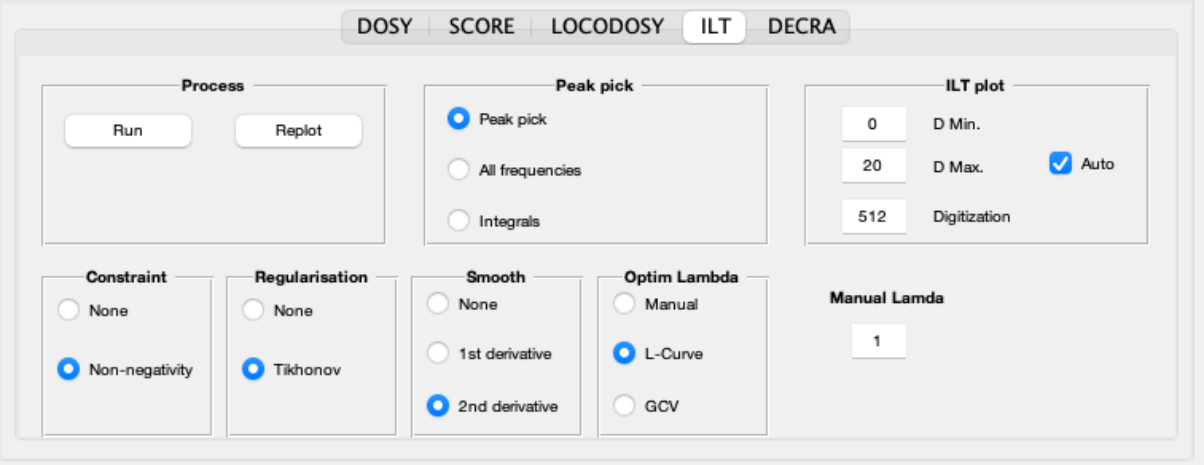
Here the user can change between with multivariate method to use for processing each segment. More information about these can be found in the *SCORE* and *DECRA* pages

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.

ILT

This is the tab for inverse Laplace transform (ILT) 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 description of this version of ILT to diffusion data is given in the below article, and references therein.

(1) Day, I. J. On the inversion of diffusion NMR data: Tikhonov regularization and optimal choice of the regularization parameter Journal of Magnetic Resonance 2011, 211, 178.

Process section

Here the user can access various processing functions. The button *Run* will start ILT 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)

An ILT fit a mixture of 1-propanol, 1-pentanol.

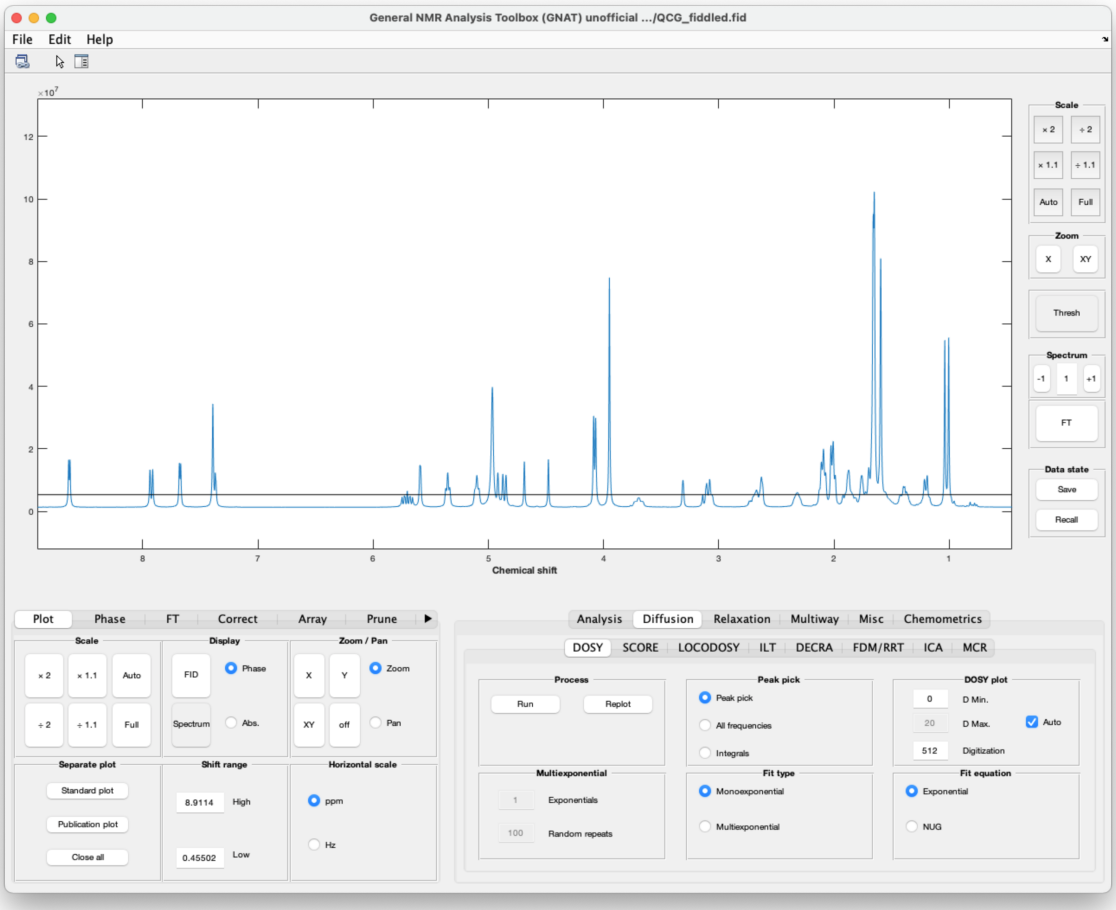
Peak pick section

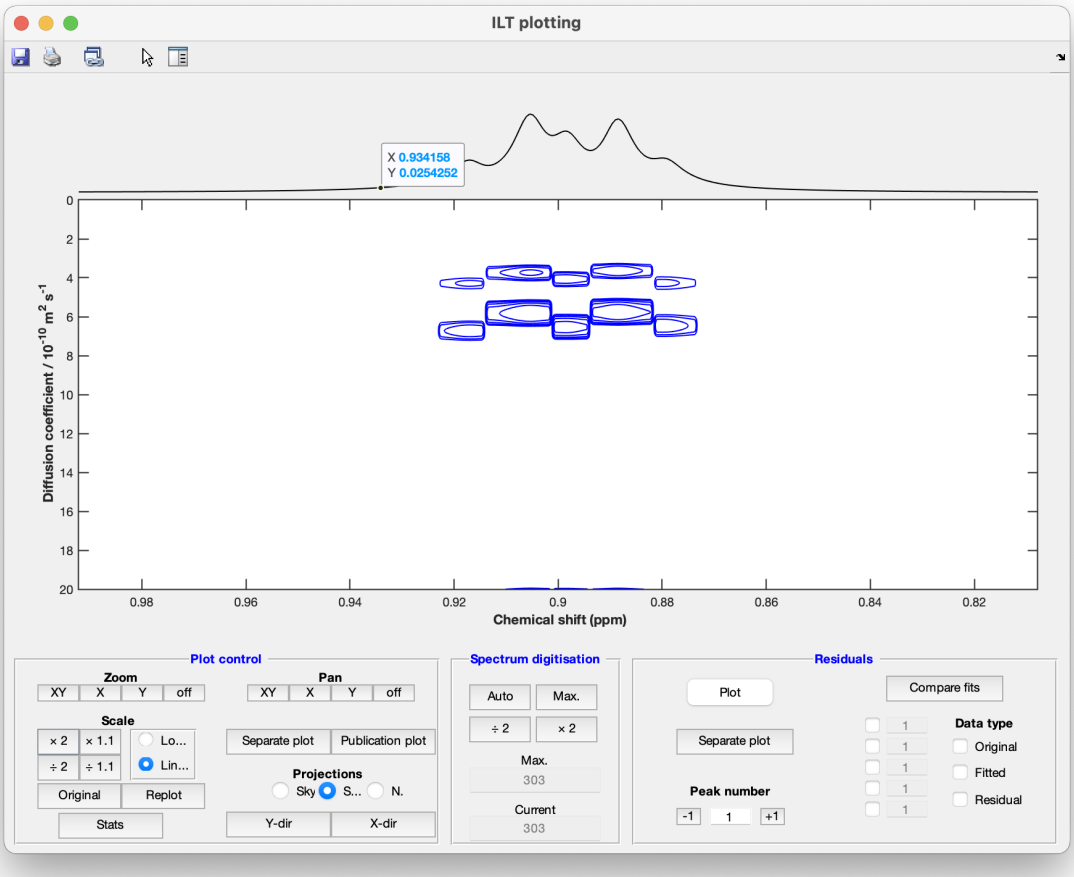
Here the user can choose which peaks that will be usedd for ILT processing.

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.


General NMR Analysis Toolbox (GNAT) unofficial .../QCG_fiddled.fid
File
Edit
Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Plot
Phase
FT
Correct
Array
Prune
Analysis
Diffusion
Relaxation
Multiway
Misc
Chemometrics
DOSY
SCORE
LOCODOSY
ILT
DECRA
FDM/RRT
ICA
MCR
Process
Run
Replot
Peak pick
Peak pick
All frequencies
Integrals
DOSY plot
D Min.
D Max.
Auto
Digitization
Multiexponential
Exponentials
Random repeats
Fit type
Monoexponential
Multiexponential
Fit equation
Exponential
NUG
Scale
Display
Zoom / Pan
FID
Phase
Spectrum
Abs.
Zoom
Pan
Separate plot
Standard plot
Publication plot
Close all
Shift range
High
Low
Horizontal scale
ppm
Hz


ILT plotting
X 0.934158
Y 0.0254252
Diffusion coefficient / 10^-10 m^2 s^-1
0
2
4
6
8
10
12
14
16
18
20
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0.82
Chemical shift (ppm)
Plot control
Zoom
XY
X
Y
off
Pan
XY
X
Y
off
Scale
× 2
× 1.1
÷ 2
÷ 1.1
Lo...
Lin...
Original
Replot
Stats
Separate plot
Publication plot
Projections
Sky
S...
N.
Y-dir
X-dir
Spectrum digitisation
Auto
Max.
÷ 2
× 2
Max.
303
Current
303
Residuals
Plot
Separate plot
Peak number
-1
1
+1
Compare fits
Data type
Original
Fitted
Residual

ILT 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 $10^{-10}\ m^2 s^{-1}$)

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

Constraint section

Here the user can opt to impose Non-negativity constraints (default) on the diffusion coefficient.

FRegularization

Here the user can decide to use Tikhonov regularisation (default) or not.

Smooth

Here the user can decide to smooth the result using second (default) or first derivatives, or not.

Optim. Lambda

Here the user can decide to use L-curve (default) or GCV optimisation of the lambda parameter for Tikhonov regularisation. There is also an option to set it manually.

References

(1) Hansen, P. C. Regularization tools version 4.0 for matlab 7.3 Numerical Algorithms 2007, 46, 189.

(2) Hansen, P. C. REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems Numerical Algorithms 1994, 6, 1.

(3) Day, I. J. On the inversion of diffusion NMR data: Tikhonov regularization and optimal choice of the regularization parameter Journal of Magnetic Resonance 2011, 211, 178.

DECRA

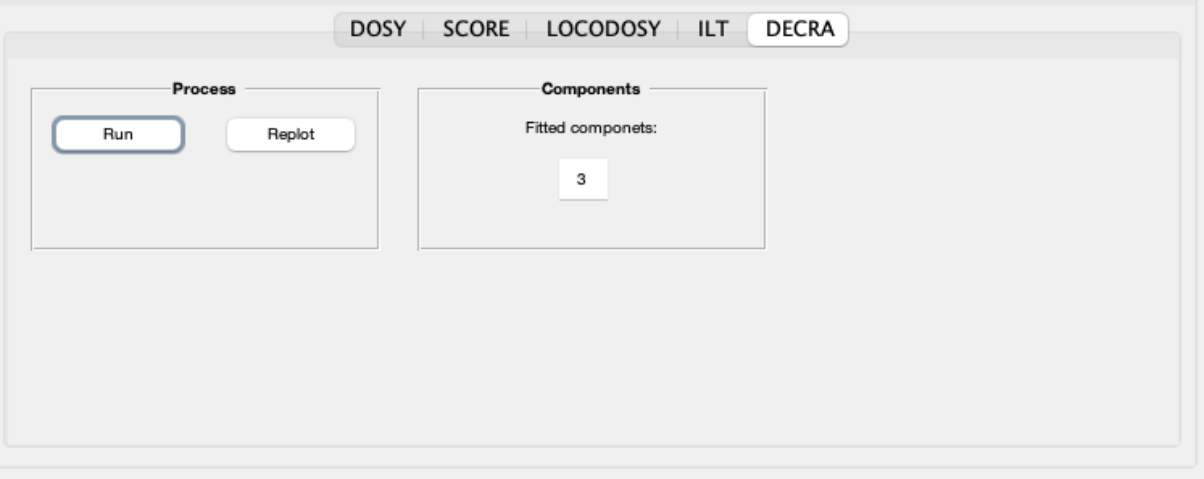
This is the tab for DECRA processing if diffusion data.

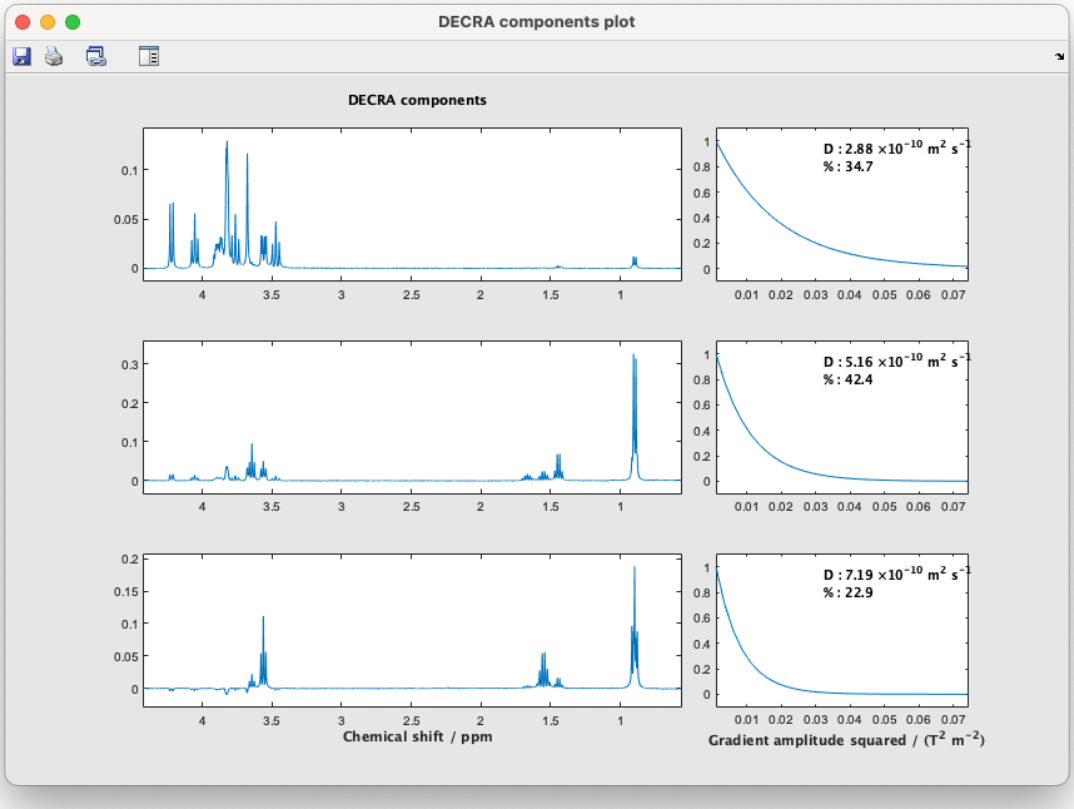
Before processing the data in this tab, make sure that the spectra have been properly preprocessed (phase, baseline correction etc), and that the diffusion parameters are correct (see the *Settings* section.)

Process section

Here the user can access various processing functions. The button *Run* will start DECRA processing with the parameters set in the other sections in this tab, but also e.g., the settings in the *Prune* tab, and in the Diffusion tab

The **Replot** button will plot again the last processed data (e.g. in case it was closed by mistake)


DOSY
SCORE
LOCODOSY
ILT
DECRA
Process
Run
Replot
Components
Fitted componets:
3


DECRA components plot
DECRA components
D : 2.88 ×10⁻¹⁰ m² s⁻¹
% : 34.7
D : 5.16 ×10⁻¹⁰ m² s⁻¹
% : 42.4
D : 7.19 ×10⁻¹⁰ m² s⁻¹
% : 22.9
Chemical shift / ppm
Gradient amplitude squared / (T² m⁻²)

A three component fit of a mixture of 1-propanol, 1-pentanol and sucrose.

Components section

Here the user can determine the number of DECRA components to fit. For unknown mixtures this can often be explored by trial and error.

References

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

(2) Pedersen, H. T.; Bro, R.; Engelsen, S. B. Towards rapid and unique curve resolution of low-field NMR relaxation data: Trilinear SLICING versus two-dimensional curve fitting Journal of Magnetic Resonance 2002, 157, 141.

(3) Antalek, B.; Hewitt, J. M.; Windig, W.; Yacobucci, P. D.; Mourey, T.; Le, K. The use of PGSE NMR and DECRA for determining polymer composition Magnetic Resonance in Chemistry 2002, 40, S60.

(4) Windig, W.; Antalek, B. Direct exponential curve resolution algorithm (DECRA): A novel application of the generalized rank annihilation method for a single spectral mixture data set with exponentially decaying contribution profiles Chemometrics and Intelligent Laboratory Systems 1997, 37, 241.

FDM

FDM-based processing is available for diffusion data.

> **Note**
>
> This page is not yet documented in detail.

ICA

ICA-based processing is available for diffusion data.

> **Note**
>
> This page is not yet documented in detail.

MCR

MCR-based processing is available for diffusion data.

> **Note**
>
> This page is not yet documented in detail.

6.16 Relaxation

This tab contains the main tools for analysing relaxation NMR data in GNAT.

6.16.1 Functionalities

ROSY

ROSY is the main plotting and analysis route for relaxation-ordered data in GNAT.

> **Note**
>
> This page is currently only lightly documented and will need expansion in a later pass.

RSCORE

RSCORE provides SCORE-style processing for relaxation data.

> **Note**
>
> This page is not yet documented in detail.

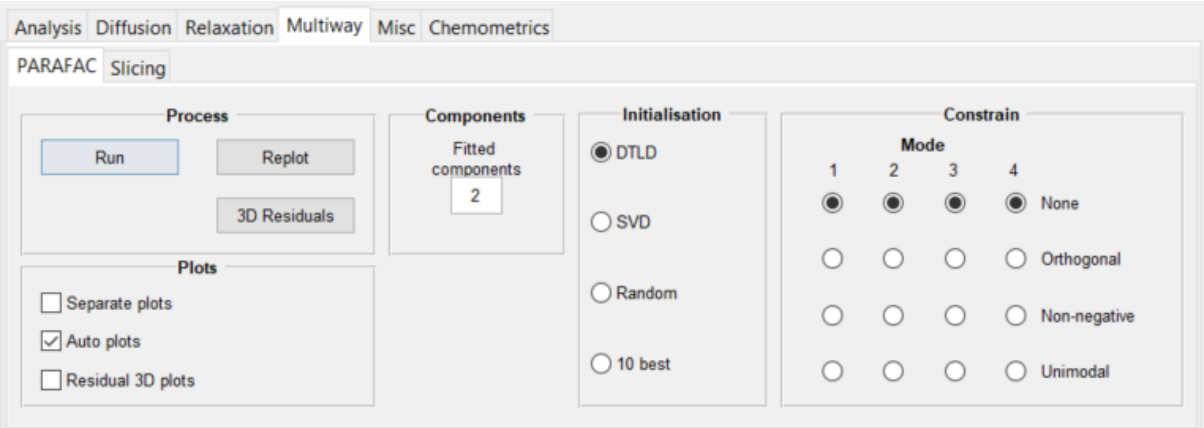
6.17 Multiway

This tab contains tools for multiway analysis.

6.17.1 Functionalities

PARAFAC

This is the tab for PARAFAC processing.



For PARAFAC (PARAllel FACtor analysis) analysis make sure that the spectra have been preprocessed (phase, baseline correction etc), and the active area of the spectrum that will we analysed is presented in the main axis of GNAT

> **Note**
>
> **Quickstart**: zoom into the part of the spectrum you are interested in, set the number of components, the initialisation and the constrains for all dimensions in your data in with PARAFAC tab (right part of the GUI) and press the Run button.

A introduction to PARAFAC analysis is the contex of mixtute analysis is given in the below article.

(1) Nilsson, M.; Khajeh, M.; Botana, A.; Bernstein, M. A.; Morris, G.A. Diffusion NMR and trilinear analysis in the study of reactionkinetics. Chem. Commun. 2009, No. 10, 1252-1254.

(2) Khajeh, M.; Botana, A.; Bernstein, M. A.; Nilsson, M.; Morris, G.A. Reaction Kinetics Studied Using Diffusion-Ordered Spectroscopyand Multiway Chemometrics. Anal. Chem. 2010, 82, 2102-2108

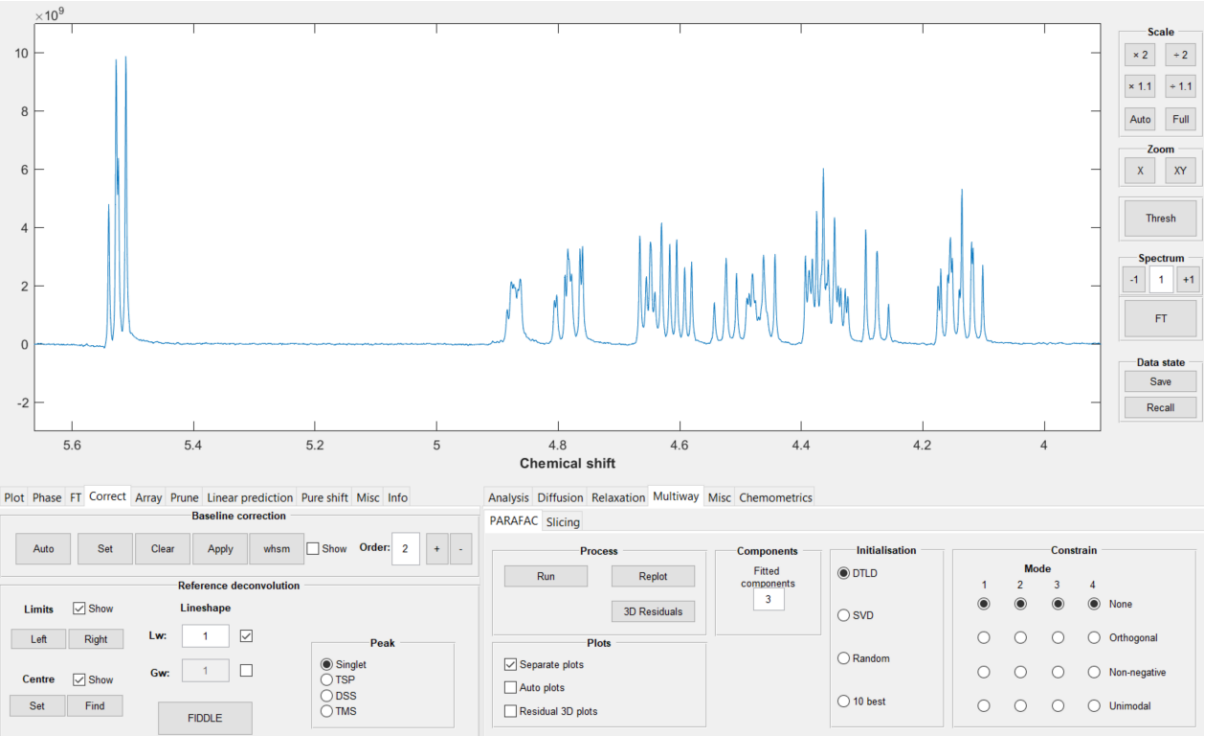
Description

In NMR applications, PARAFAC can decompose a multidimensional data, in this example NMR data array, of order N (where $N \geq 3$) into a sum of outer products involving N vectors (a low-rank model). For instance, when N=3, the array might represent a 3D NMR spectrum with dimensions of chemical shift (^{1}H), relaxation time (T1/T2), diffusion, NOE, etc. This decomposition is particularly useful for analyzing complex NMR datasets, such as those from relaxation experiments or metabolomics studies, where it can help separate overlapping signals and identify individual components in mixtures. An example application would be the analysis of 3D NMR spectra from the SCALPEL experiment as shown in the reference below.

(3) Dal Poggetto, G., Castañar, L., Adams, R.W., Morris, G.A. and Nilsson, M., 2019. Dissect and divide: Putting NMR spectra of mixtures under the knife. Journal of the American Chemical Society, 141(14), pp.5766-5771.

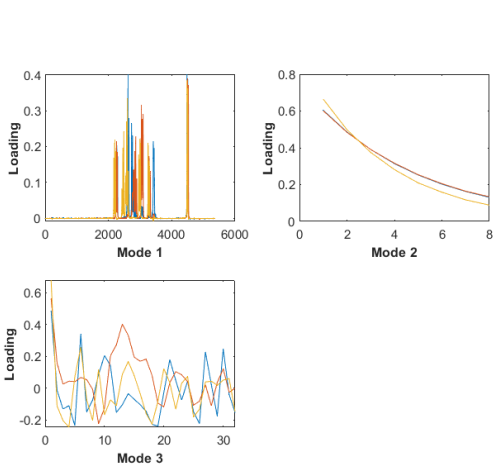
Plot section

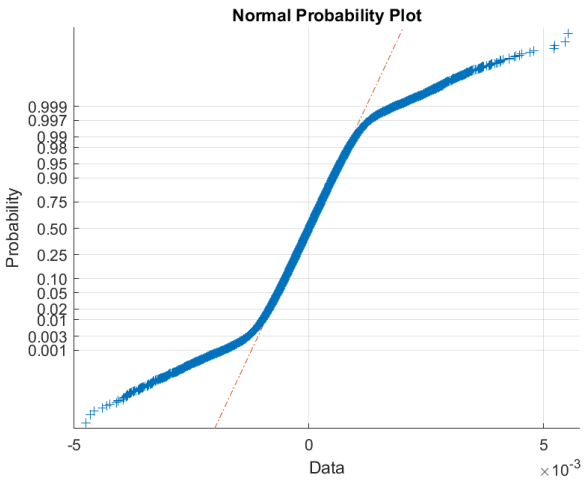
Here, the mixture of glucose, maltose and glucose-6-phosphate was imported and preprocessed in GNAT.

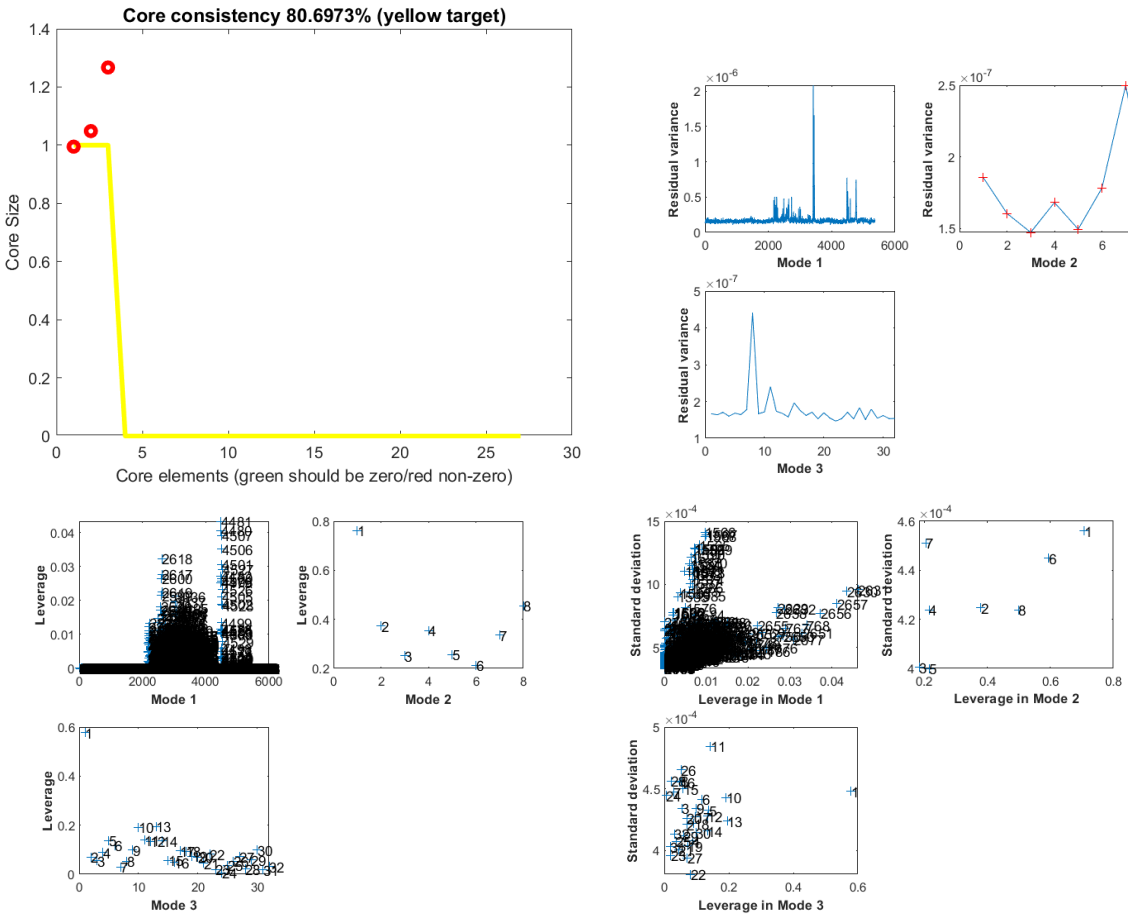


Auto plots

The results for PARAFAC calculations are shown in the next figure. The data in the same presented in the previous reference. This result is imported and processed in GNAT, but it is crucial to state that the data was processed with reference deconvolution, without this method the separation of components fail. More information of this method of processing can be found in *Correct*. The user can analyse the separation of components and its statistical plots when the option `Auto plots` is selected in the tab of PARAFAC. For the mixture of glucose, maltose and glucose-6-phosphate the initialization method DTLD was selected



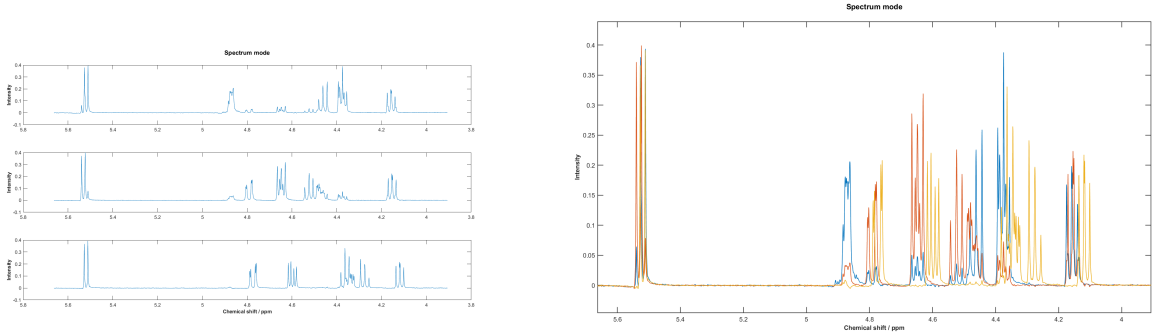


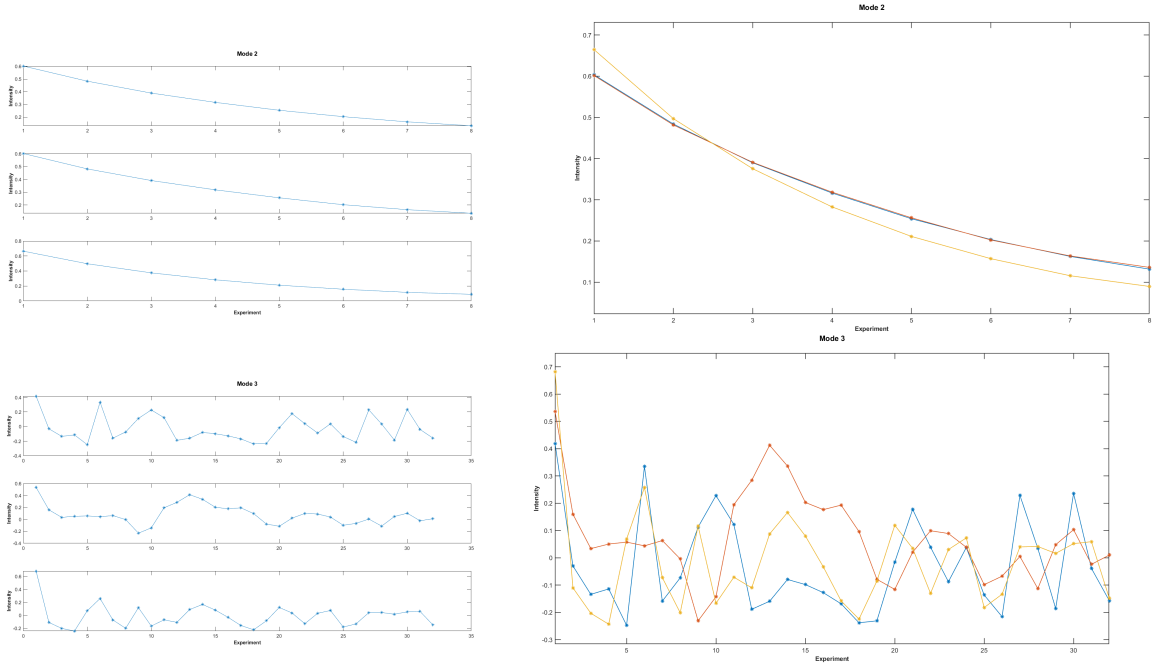


> **Warning**
>
> These plots are suitable for 3D analyses. The plots and figure of merit for higher dimension in GNAT are been developed.

Separated plots

A more clear visualisation of the components can be achieved by the `Separated plots` option. Here, the focus are only in the modes (Loadings) for the separated components. An example of these plots for the same dataset is presented below:





> **Note**
>
> Just click in the desired image to open it in a higher resolution.

Slicing

Slicing utilities are available for multiway analysis workflows.

> **Note**
>
> This page is not yet documented in detail.

6.18 Misc

Miscellaneous analysis utilities are collected here.

6.18.1 Functionalities

Sim DOSY

This section covers simulation of DOSY data.

> **Note**
>
> This page is not yet documented in detail.

Macros

This section is reserved for analysis-related macro functionality.

> **Note**
>
> This page is not yet documented.

6.19 Chemometrics

This section contains the current chemometric tools available in GNAT.

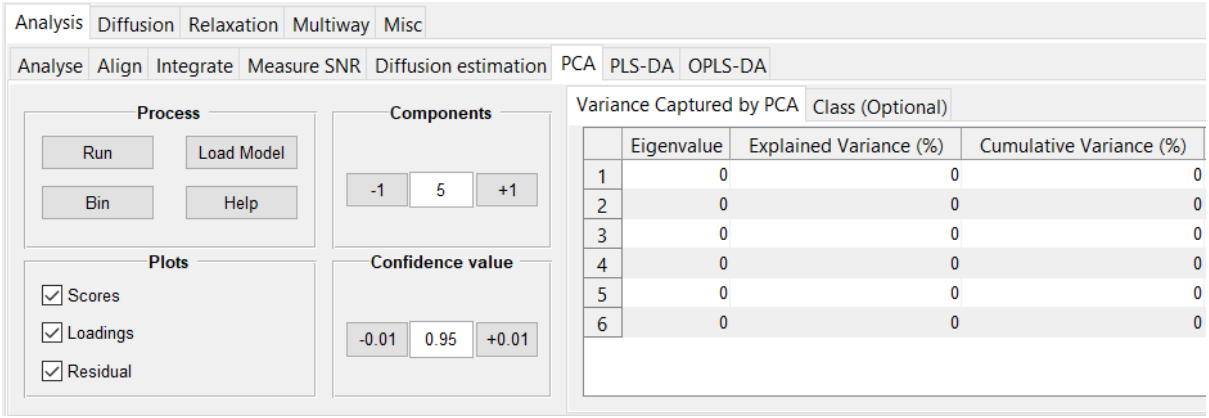
6.19.1 Functionalities

PCA

PCA is an effective method for extracting information from large 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

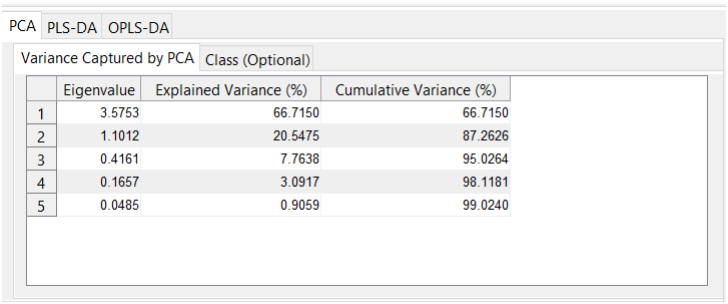
It's possible to perform a PCA analysis in GNAT (General NMR Analysis Toolbox) by accessing the Analysis Functionalities on GNAT (right functions) on the tab Chemometrics>PCA. The steps below describe how to create a new PCA transformation on your data:



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.

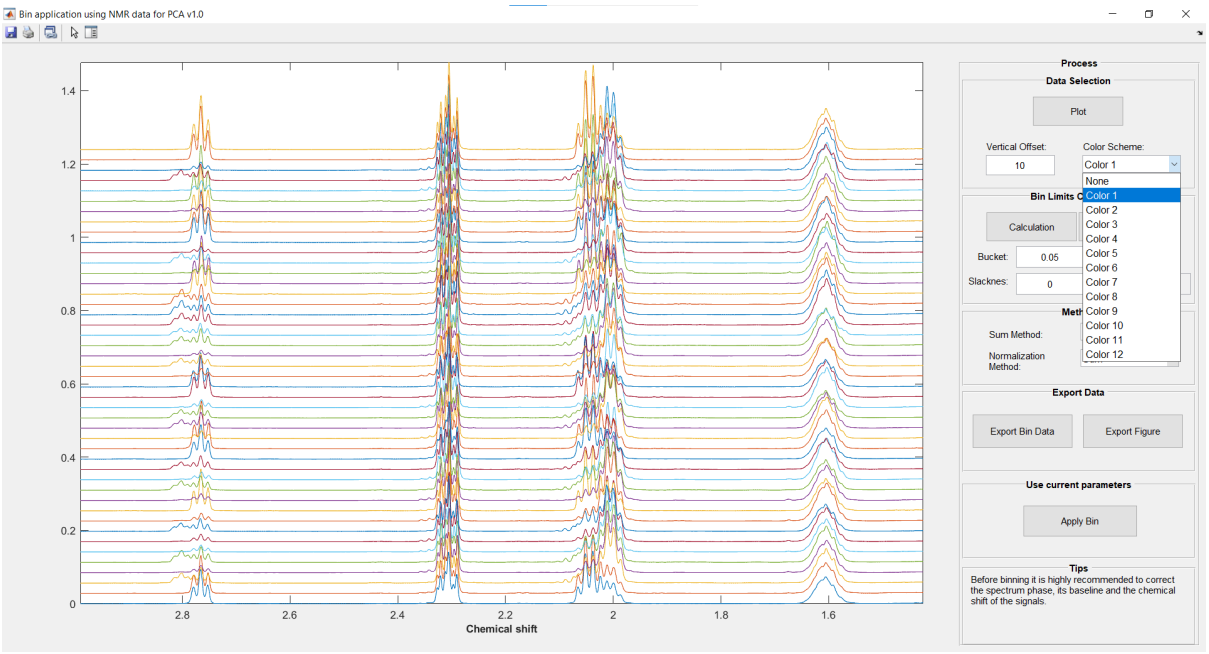
The table **Variance Captured by PCA** shows the explained variance After pressing the button `Run` in the **Process** panel. The number of components presented on this table is dictated by the value entered in the `Components` edit box.



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.

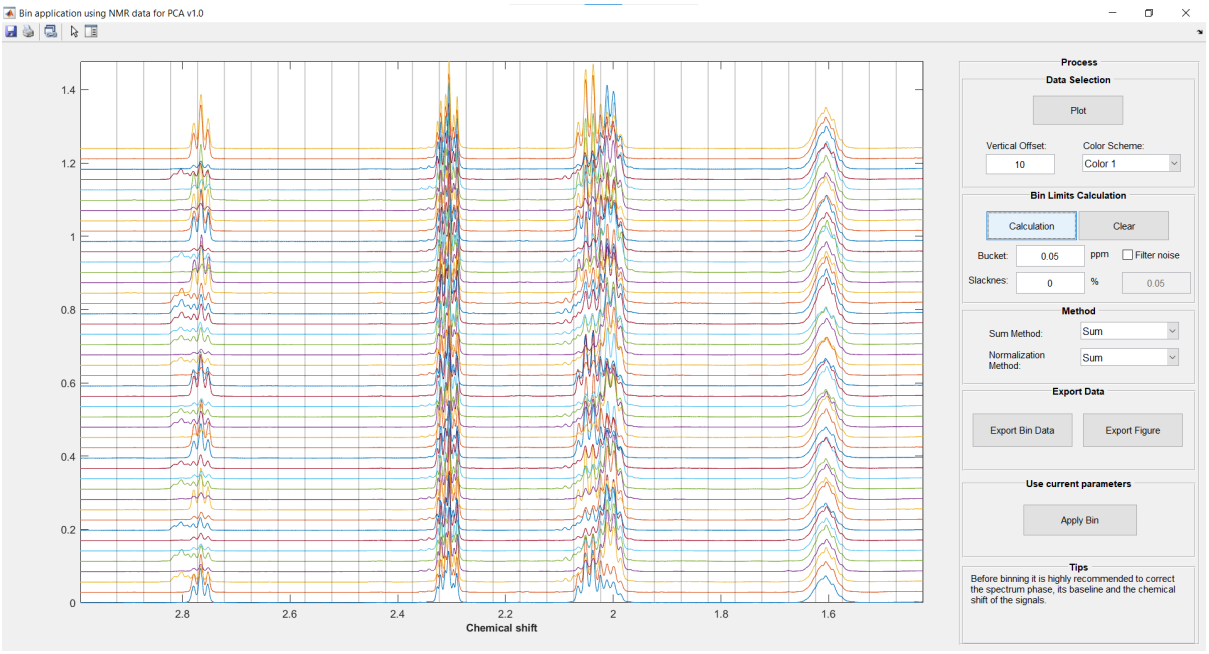
> **Warning**
>
> The function `Filter Noise` is used to define bin limits close to the spectrum signals. However, it is not fully optimized at the moment

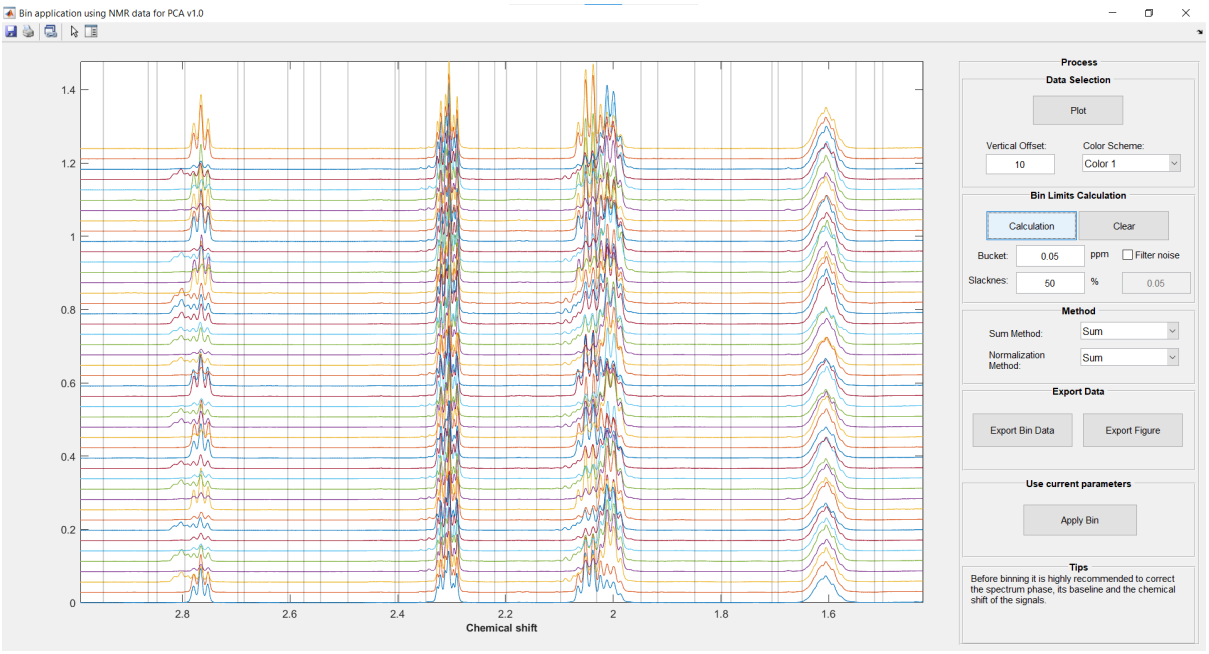
All the limits can be moved after left-clicking in a existing limit

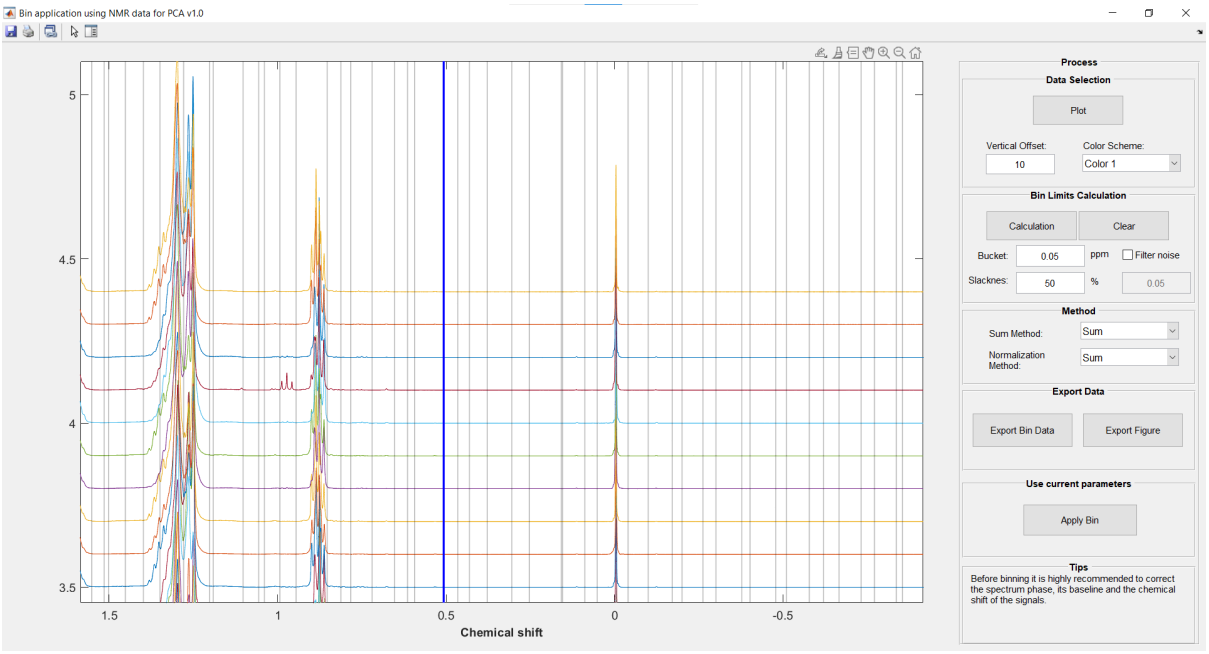
If the method **Reference** is selected on the `Normalization method` pop-menu, the user can right click the right/left limits of the NMR region that will be used to normalize each spectrum

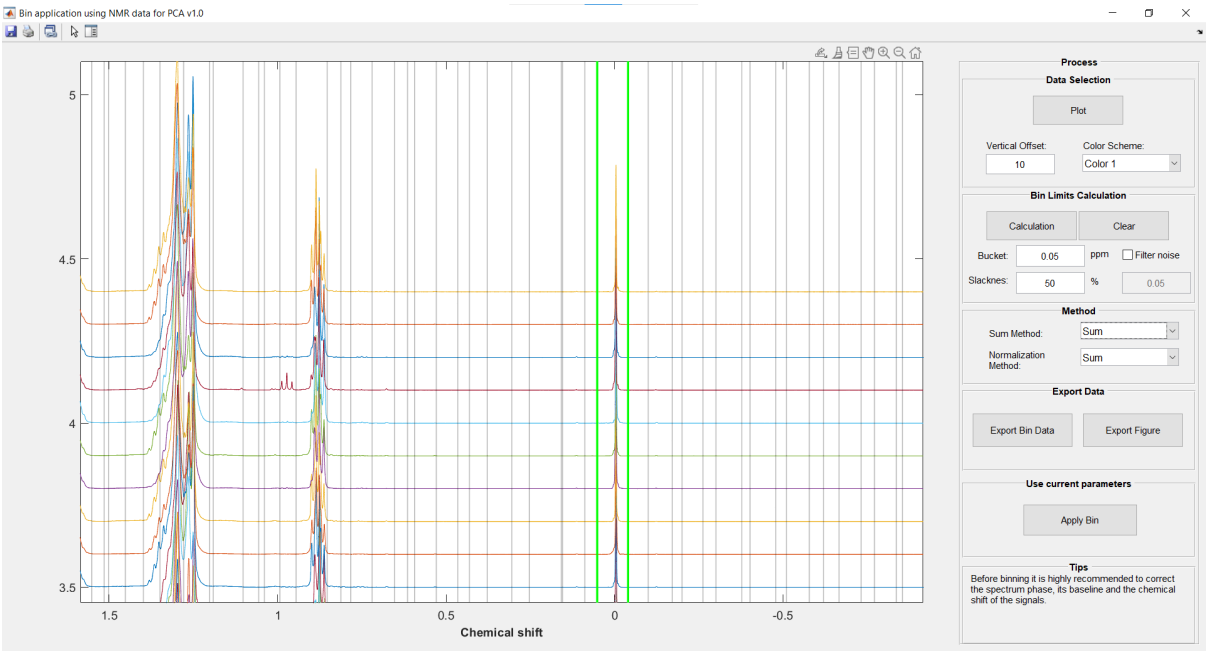
After finishing the calculation, it is necessary to press the button `Apply` so save the modification. .. figure:: PCA/Fig12_BIN_GUI_apply.png

Optionally, any user can export the dataset into .csv and .txt into formats suitable to be imported into


Bin application using NMR data for PCA v1.0
Chemical shift
Process
Data Selection
Plot
Vertical Offset:
10
Color Scheme:
Color 1
Bin Limits Calculation
Calculation
Clear
Bucket:
0.05
ppm
Filter noise
Slacknes:
0
%
0.05
Method
Sum Method:
Sum
Normalization Method:
Sum
Export Data
Export Bin Data
Export Figure
Use current parameters
Apply Bin
Tips
Before binning it is highly recommended to correct the spectrum phase, its baseline and the chemical shift of the signals.


Bin application using NMR data for PCA v1.0
Chemical shift
Process
Data Selection
Plot
Vertical Offset:
10
Color Scheme:
Color 1
Bin Limits Calculation
Calculation
Clear
Bucket:
0.05
ppm
Filter noise
Slacknes:
50
%
0.05
Method
Sum Method:
Sum
Normalization Method:
Sum
Export Data
Export Bin Data
Export Figure
Use current parameters
Apply Bin
Tips
Before binning it is highly recommended to correct the spectrum phase, its baseline and the chemical shift of the signals.


Bin application using NMR data for PCA v1.0
5
4.5
4
3.5
1.5
1
0.5
0
-0.5
Chemical shift
Process
Data Selection
Plot
Vertical Offset:
10
Color Scheme:
Color 1
Bin Limits Calculation
Calculation
Clear
Bucket:
0.05
ppm
Filter noise
Slacknes:
50
%
0.05
Method
Sum Method:
Sum
Normalization Method:
Sum
Export Data
Export Bin Data
Export Figure
Use current parameters
Apply Bin
Tips
Before binning it is highly recommended to correct the spectrum phase, its baseline and the chemical shift of the signals.


Bin application using NMR data for PCA v1.0
5
4.5
4
3.5
1.5
1
0.5
0
-0.5
Chemical shift
Process
Data Selection
Plot
Vertical Offset:
10
Color Scheme:
Color 1
Bin Limits Calculation
Calculation
Clear
Bucket:
0.05
ppm
Filter noise
Slacknes:
50
%
0.05
Method
Sum Method:
Sum
Normalization Method:
Sum
Export Data
Export Bin Data
Export Figure
Use current parameters
Apply Bin
Tips
Before binning it is highly recommended to correct the spectrum phase, its baseline and the chemical shift of the signals.

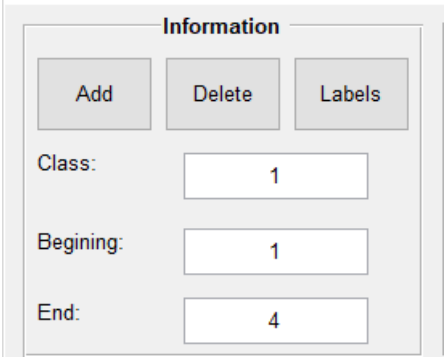
Bruker, Metaboanalyst and R .. figure:: PCA/Fig10_BIN_GUI_export.png

> **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

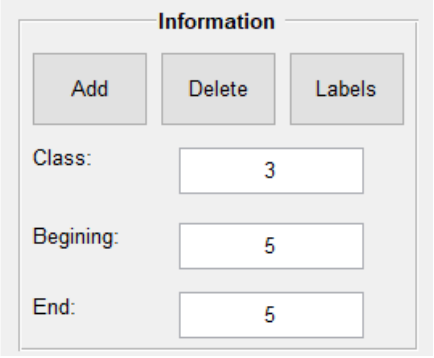
Class GUI

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.

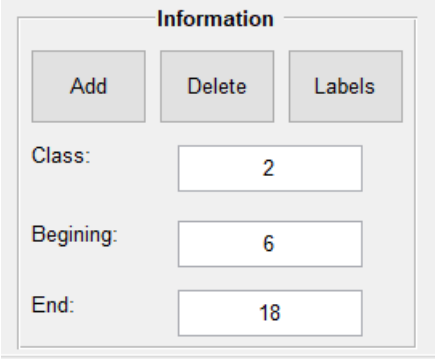
Case 1: The first four samples belong to the class 1, so the following parameters should be used in the tab:



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.

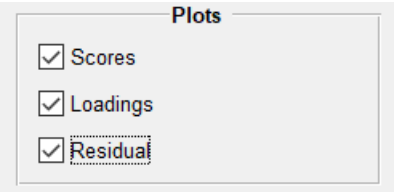
Warning

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

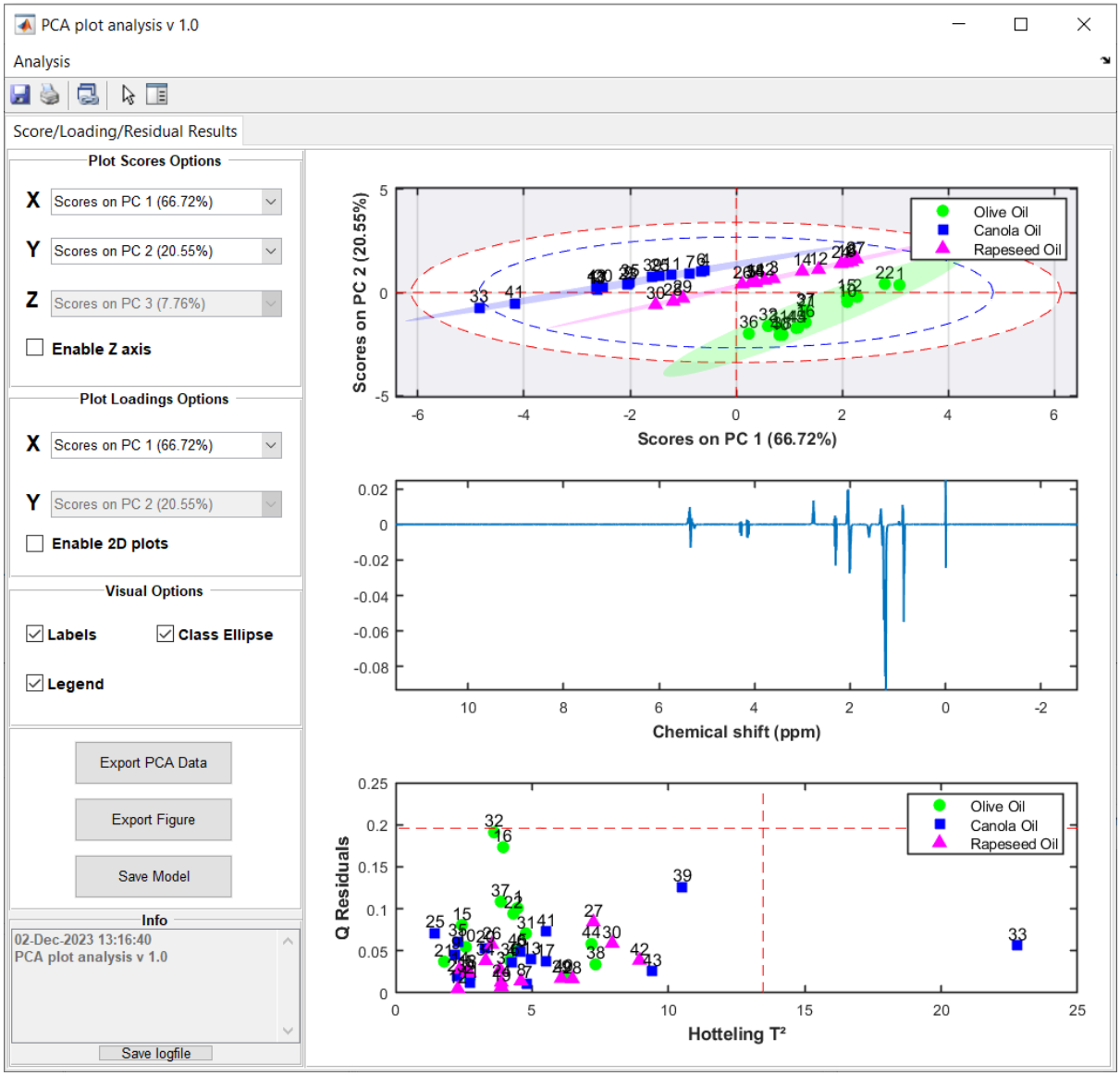
Before pressing Run to calculate the PCA model, the user need to define which plots will be shown in the Plot GUI. There are three option of plots: **Scores**, **Loadings** and **Residual**



The tab **Score/Loadings/Residuals Result** displays a visualisation 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


PCA plot analysis v 1.0
Analysis
Score/Loading/Residual Results
Plot Scores Options
X
Scores on PC 1 (66.72%)
Y
Scores on PC 2 (20.55%)
Z
Scores on PC 3 (7.76%)
Enable Z axis
Plot Loadings Options
X
Scores on PC 1 (66.72%)
Y
Scores on PC 2 (20.55%)
Enable 2D plots
Visual Options
Labels
Class Ellipse
Legend
Export PCA Data
Export Figure
Save Model
Info
02-Dec-2023 13:16:40
PCA plot analysis v 1.0
Save logfile
Olive Oil
Canola Oil
Rapeseed Oil
Scores on PC 2 (20.55%)
Scores on PC 1 (66.72%)
Chemical shift (ppm)
Q Residuals
Hotteling T²

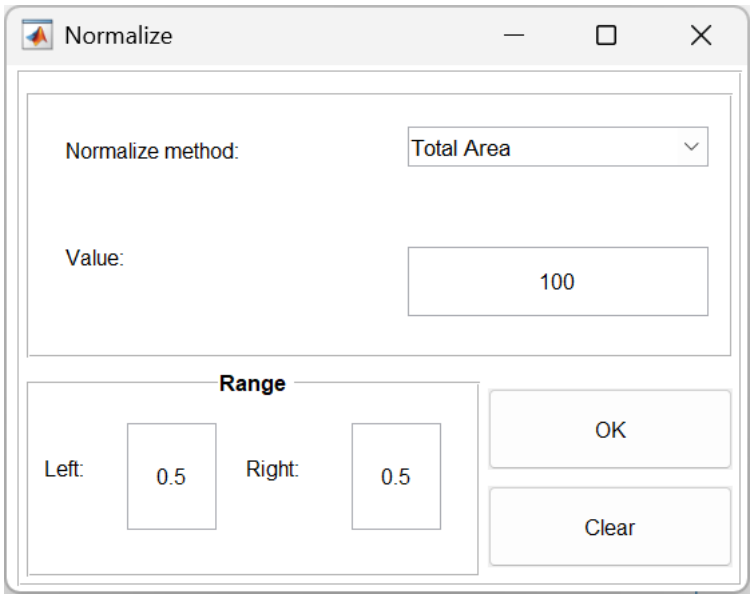
of variation in the PCA model, but the Q statistic is a measure of the amount of variation that the PCA model does not capture, as seen in its residual matrix E (MUJICA et al., 2011). The Mahalanobis distance defines the T 2 statistic, while the Euclidean distance defines the Q statistic (KOURTI; MACGREGOR, 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.

Normalization

The spectral intensities may be normalized using the procedure available in the "More Processing" menu. Six normalization approaches are provided: total area, largest peak, largest peak range, specified peak, specified peak range, and manual normalization (see Figure 17).



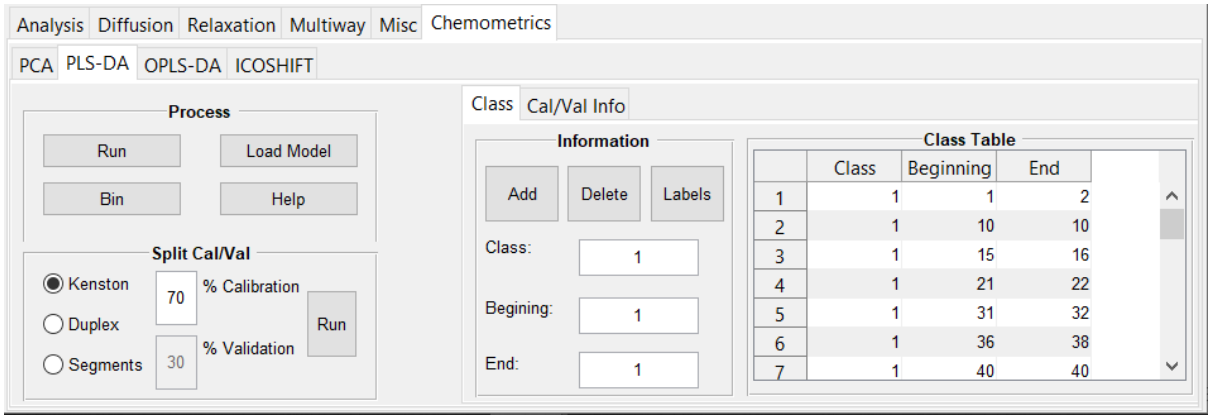
- Total area: Scales each spectrum so that its integrated area equals a prescribed value (see Figure 17). Regions may be omitted from the calculation (for example, residual water signals) by applying the Cut or by zooming in on the zone of interest in the spectrum.
- Largest peak: Scales spectra such that the amplitude of the most intense peak in the active area of GNAT is set to a chosen value.
- Probabilistic quotient: Implements the probabilistic quotient normalization approach introduced by Dieterle et al.; further methodological details are available in *Anal. Chem.* 2006, 78, 4281–4290.
- Integral reference: Scales spectra by setting the intensity at a user-specified chemical shift (ppm) to a chosen value.
- Integral range: Scales spectra so that the integrated signal over a user-specified spectral interval (by default, the current zoom region) equals a designated value.

Interval-dependent methods like Integral range are configured via the range panel within the Normalization graphical user interface.

PLS-DA

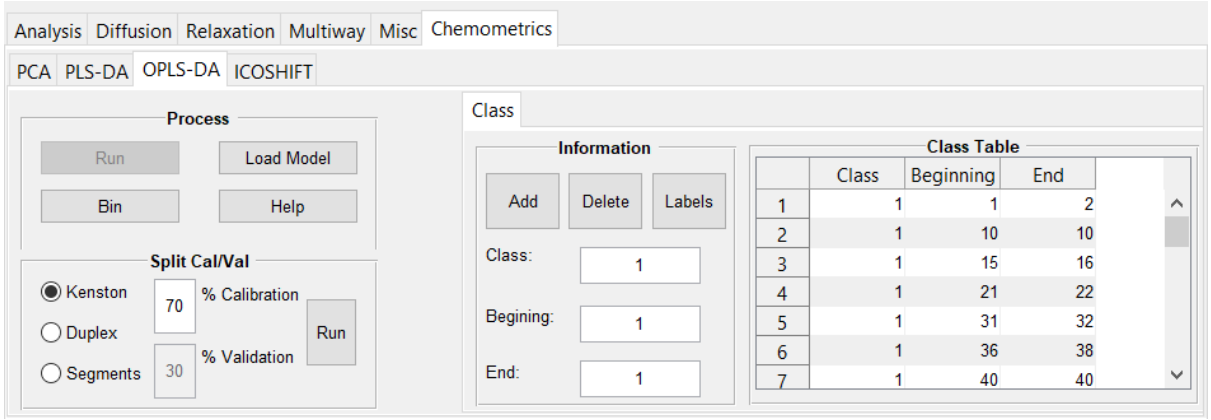
PLS-DA is a regression method that uses a matrix $X_{(ij)}$ as a predictor and a matrix $Y_{(ik)}$ with dummy variables as the answer. The dummy matrix Y contains categorical variables (i.e. 0 or 1). The model generated produces a $\hat{Y}$ matrix for discriminating purposes.The discriminating rule compares 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 below:



Note

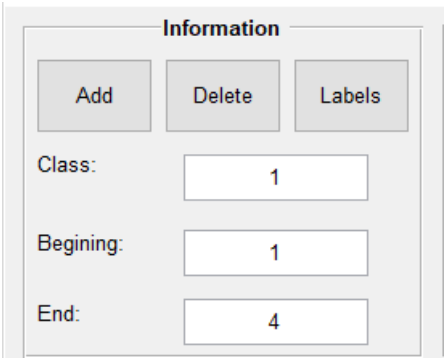
It is important to note that all of the methods used for creating a PLS-DA model may also be used for OPLS-DA models, as shown below.



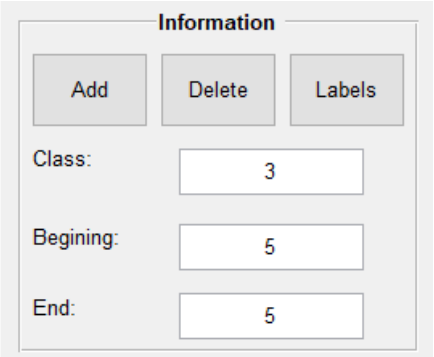
Class GUI

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.

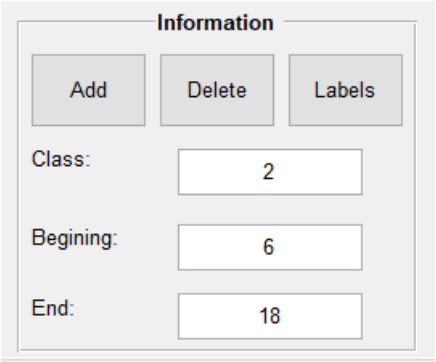
Case 1: The first four samples belong to the class 1, so the following parameters should be used in the tab:



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 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:

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 visualisation 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 visualised in the tab "residual", in which is possible to visualize the Root Mean Square Error of Cross Validation (RMSECV).

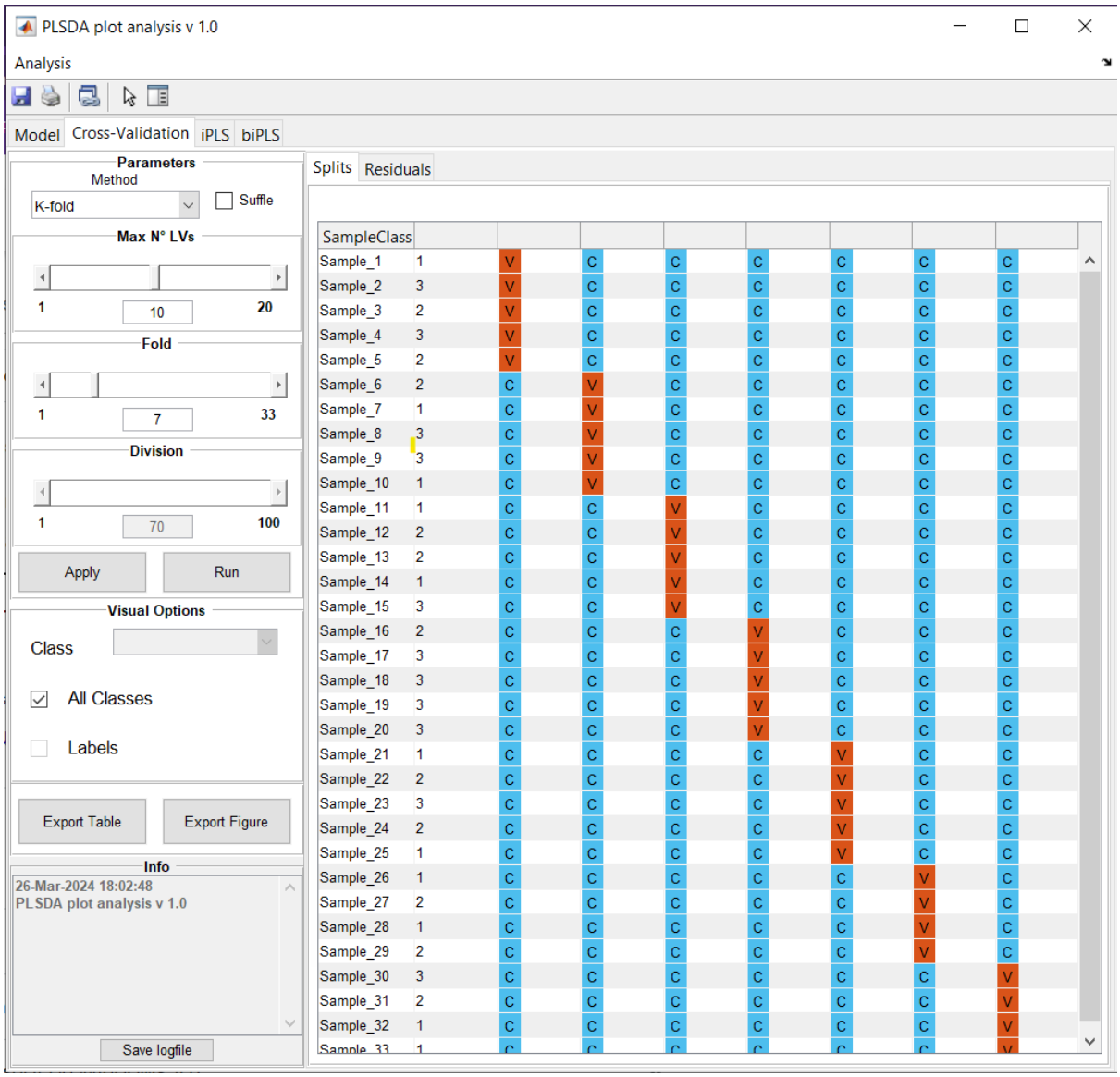
The RMSECV calculation is used graphically to show how many latent variables are needed for your PLS-DA model. The number of latent variables retained in the model should, in principle, grow as the the residuals of the calibration data decreases.

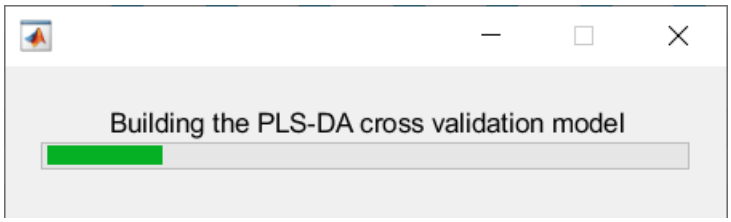
iPLS

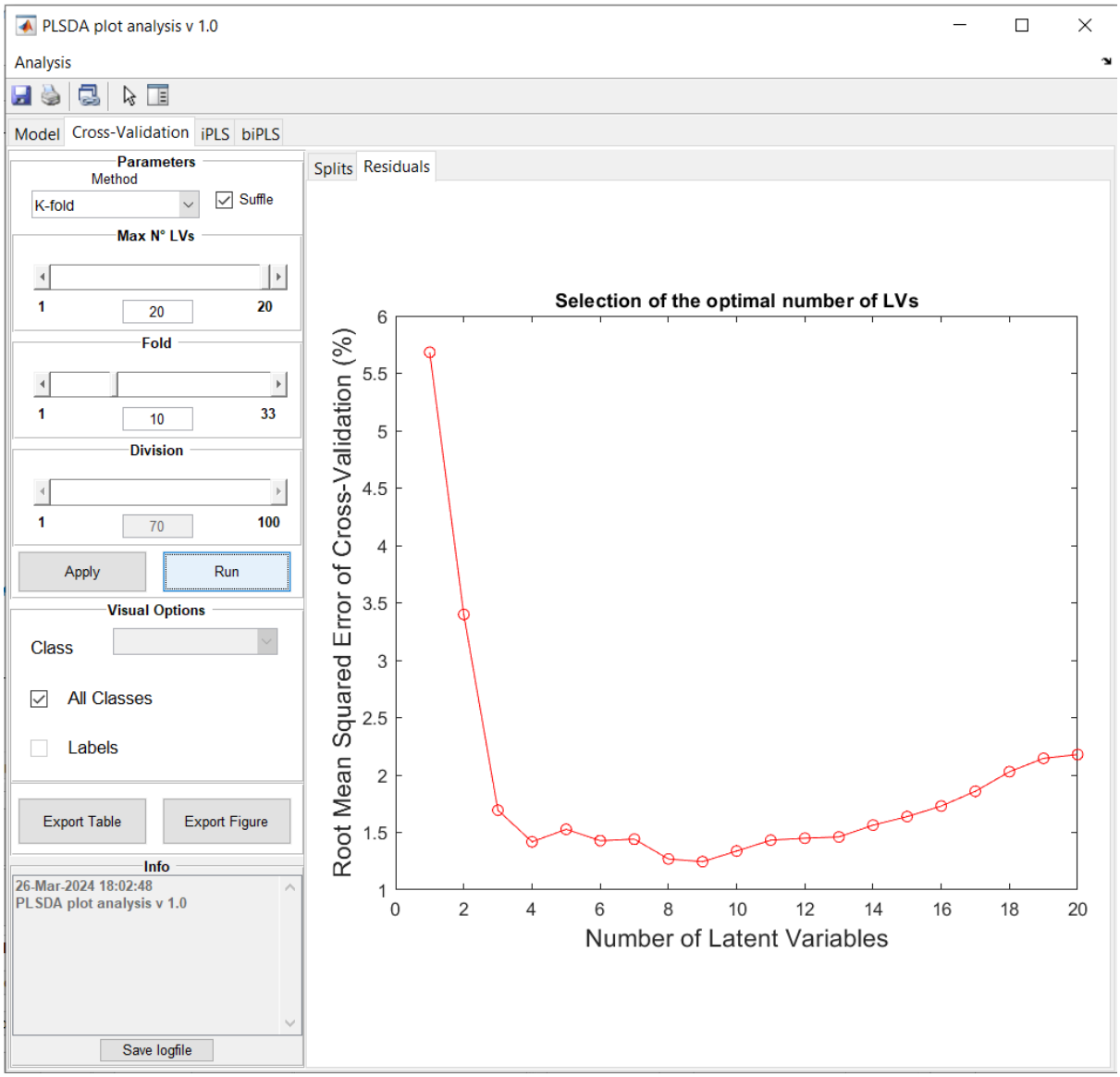
The methods for variable selection are presented in the tabs iPLS and biPLS. The implementation of the iPLS algorithm was based of the work developed for the iToolbox for Nørgaard L, 2001. In a nutshell, the iPLS method divides the spectrum into multiple equidistant regions and analyzes the value of 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.


PLSDA plot analysis v 1.0
Analysis
Model
Cross-Validation
iPLS
biPLS
Parameters
Method
K-fold
Suffle
Max N° LVs
1
10
20
Fold
1
7
33
Division
1
70
100
Apply
Run
Visual Options
Class
All Classes
Labels
Export Table
Export Figure
Info
26-Mar-2024 18:02:48
PLSDA plot analysis v 1.0
Save logfile
Splits
Residuals
SampleClass


Building the PLS-DA cross validation model

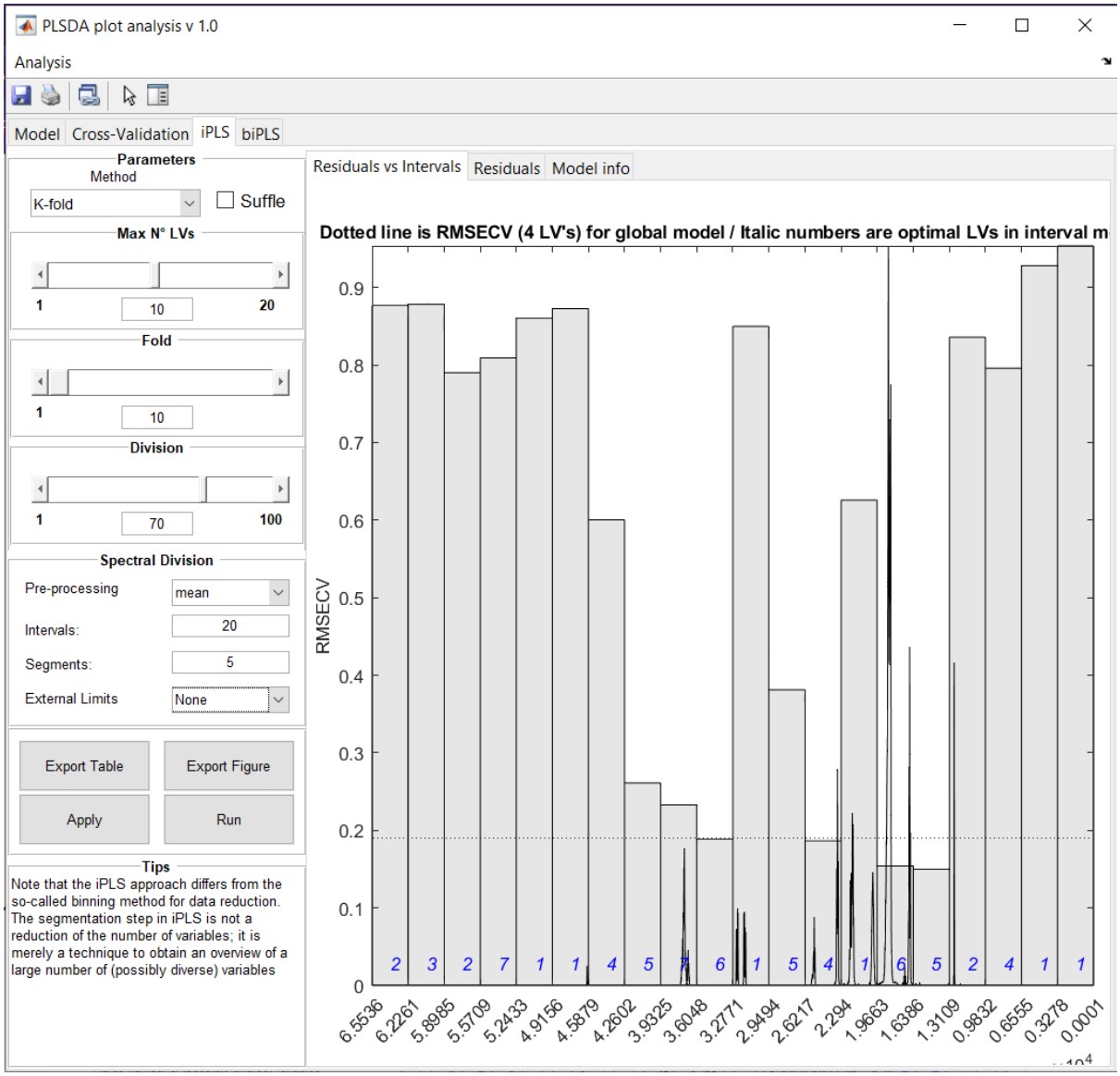

PLSDA plot analysis v 1.0
Analysis
Model
Cross-Validation
iPLS
biPLS
Parameters
Method
K-fold
Suffle
Max N° LVs
1
20
20
Fold
1
10
33
Division
1
70
100
Apply
Run
Visual Options
Class
All Classes
Labels
Export Table
Export Figure
Info
26-Mar-2024 18:02:48
PLSDA plot analysis v 1.0
Save logfile
Splits
Residuals
Selection of the optimal number of LVs
Root Mean Squared Error of Cross-Validation (%)
Number of Latent Variables

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.

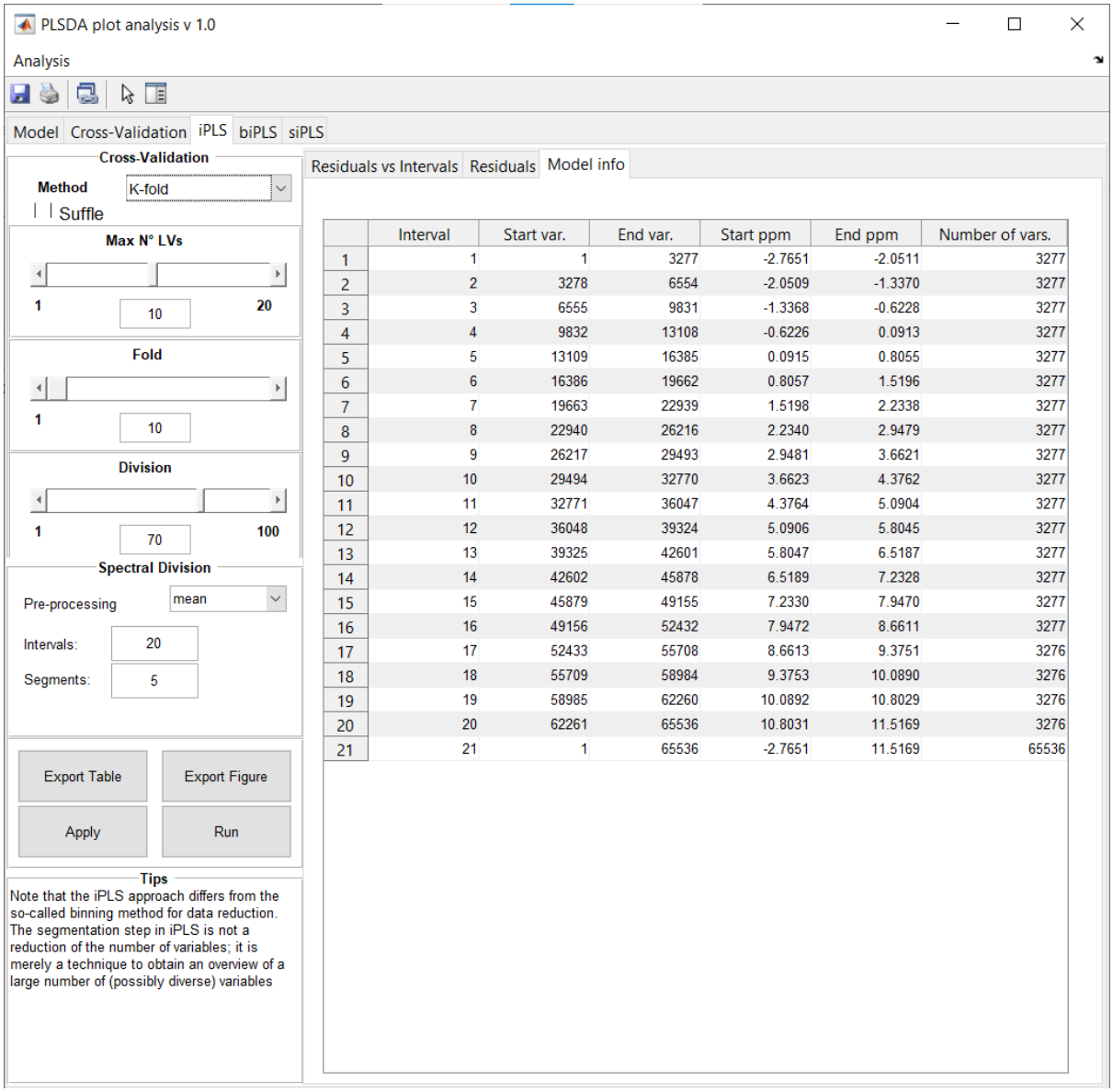
The forward approach works as follows:

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.



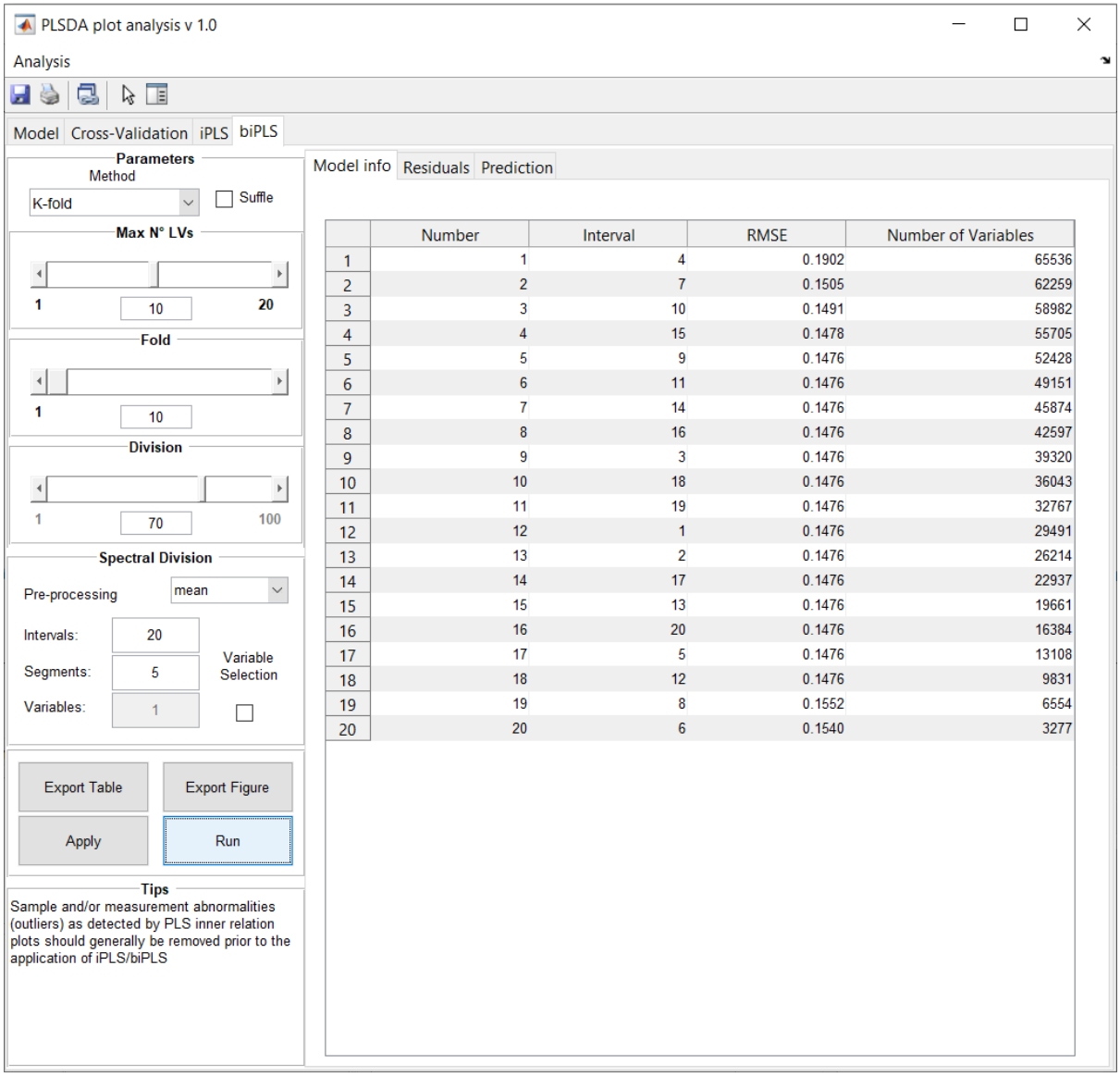
	Interval	Start var.	End var.	Start ppm	End ppm	Number of vars.
1	1	1	3277	-2.7651	-2.0511	3277
2	2	3278	6554	-2.0509	-1.3370	3277
3	3	6555	9831	-1.3368	-0.6228	3277
4	4	9832	13108	-0.6226	0.0913	3277
5	5	13109	16385	0.0915	0.8055	3277
6	6	16386	19662	0.8057	1.5196	3277
7	7	19663	22939	1.5198	2.2338	3277
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.

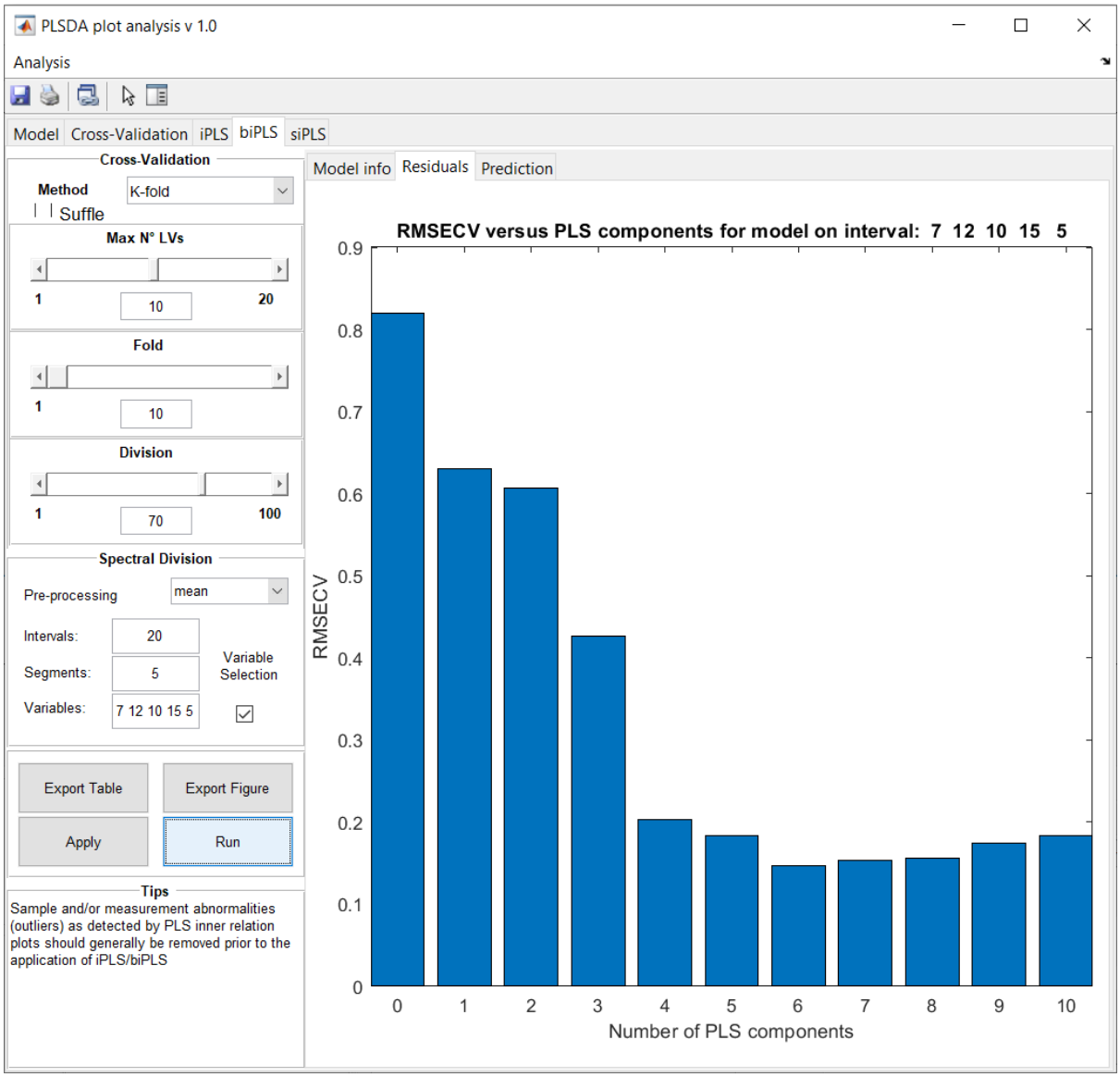
In the "Model info" tab table, the colum RMSE dictate which variables should be selected. The error is reduced until the interval 8, so the ideal variables should be 8 and 6. The model is recalculated using these variables after checking the edit box "Variable selection" and adding these variables in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

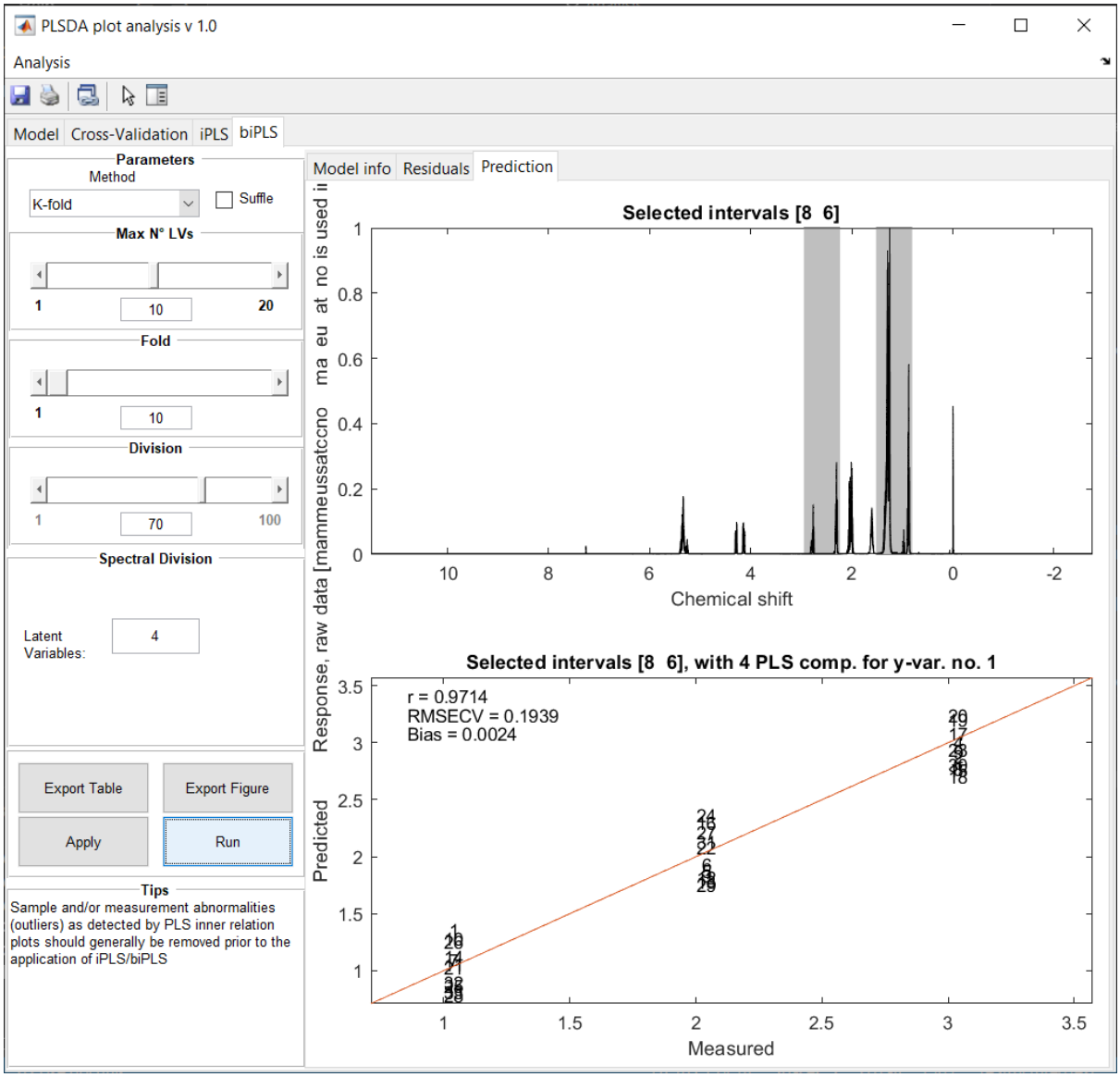


	Number	Interval	RMSE	Number of Variables
1	1	4	0.1902	65536
2	2	7	0.1505	62259
3	3	10	0.1491	58982
4	4	15	0.1478	55705
5	5	9	0.1476	52428
6	6	11	0.1476	49151
7	7	14	0.1476	45874
8	8	16	0.1476	42597
9	9	3	0.1476	39320
10	10	18	0.1476	36043
11	11	19	0.1476	32767
12	12	1	0.1476	29491
13	13	2	0.1476	26214
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 put in the ideal value for the LV. After pressing `Apply` and `Run` the user can see the variables used in the model.

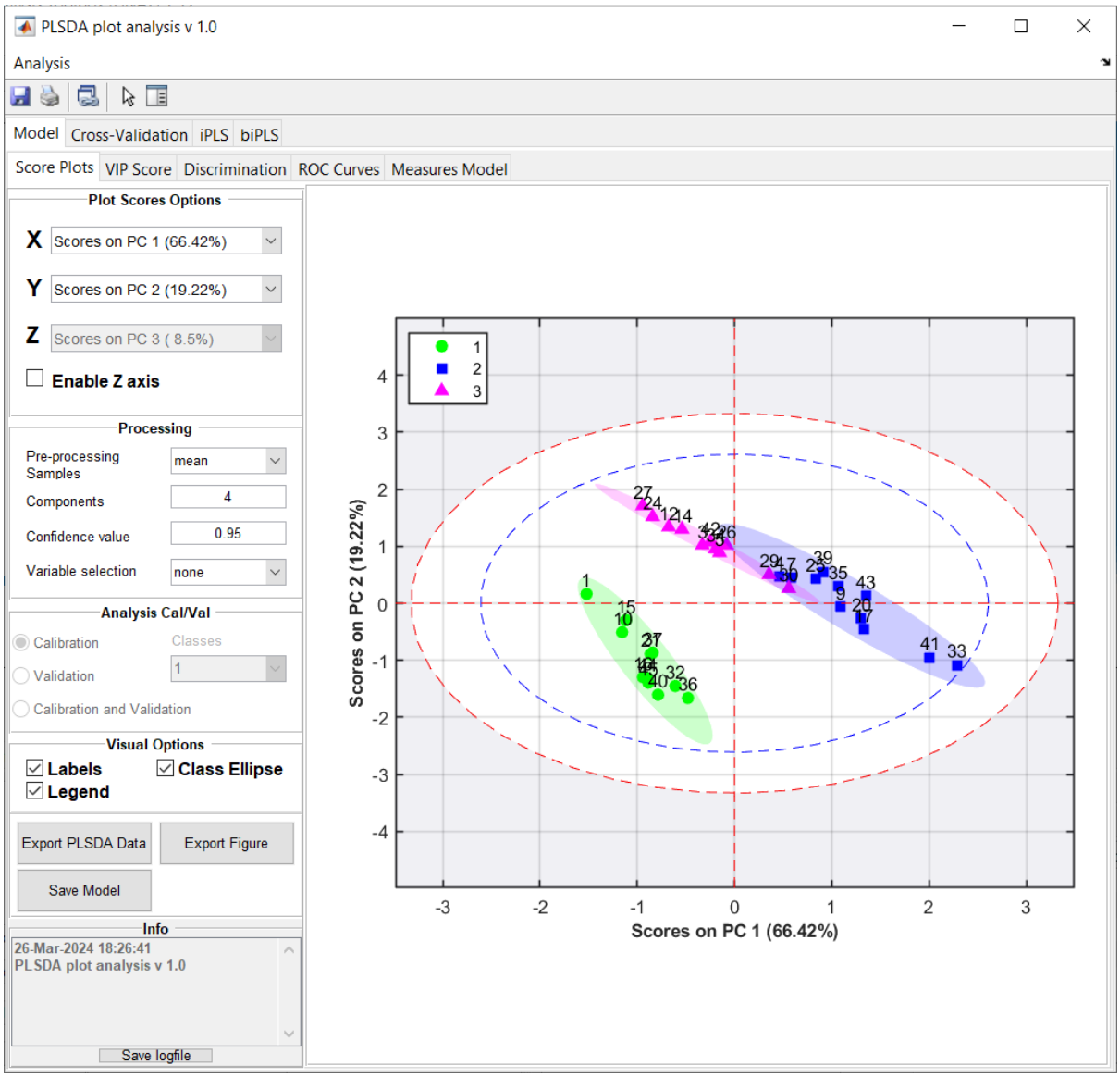

PLSDA plot analysis v 1.0
Analysis
Model
Cross-Validation
iPLS
biPLS
siPLS
Cross-Validation
Method
K-fold
Suffle
Max N° LVs
1
10
20
Fold
1
10
Division
1
70
100
Spectral Division
Pre-processing
mean
Intervals:
20
Segments:
5
Variables:
7 12 10 15 5
Variable Selection
Export Table
Export Figure
Apply
Run
Tips
Sample and/or measurement abnormalities (outliers) as detected by PLS inner relation plots should generally be removed prior to the application of iPLS/biPLS
Model info
Residuals
Prediction
RMSECV versus PLS components for model on interval: 7 12 10 15 5
RMSECV
Number of PLS components
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
1
2
3
4
5
6
7
8
9
10


PLSDA plot analysis v 1.0
Analysis
Model
Cross-Validation
iPLS
biPLS
Parameters
Method
K-fold
Suffle
Max N° LVs
1
10
20
Fold
1
10
Division
1
70
100
Spectral Division
Latent Variables:
4
Export Table
Export Figure
Apply
Run
Tips
Sample and/or measurement abnormalities (outliers) as detected by PLS inner relation plots should generally be removed prior to the application of iPLS/biPLS
Model info
Residuals
Prediction
Selected intervals [8 6]
Chemical shift
Selected intervals [8 6], with 4 PLS comp. for y-var. no. 1
r = 0.9714
RMSECV = 0.1939
Bias = 0.0024
Predicted
Measured

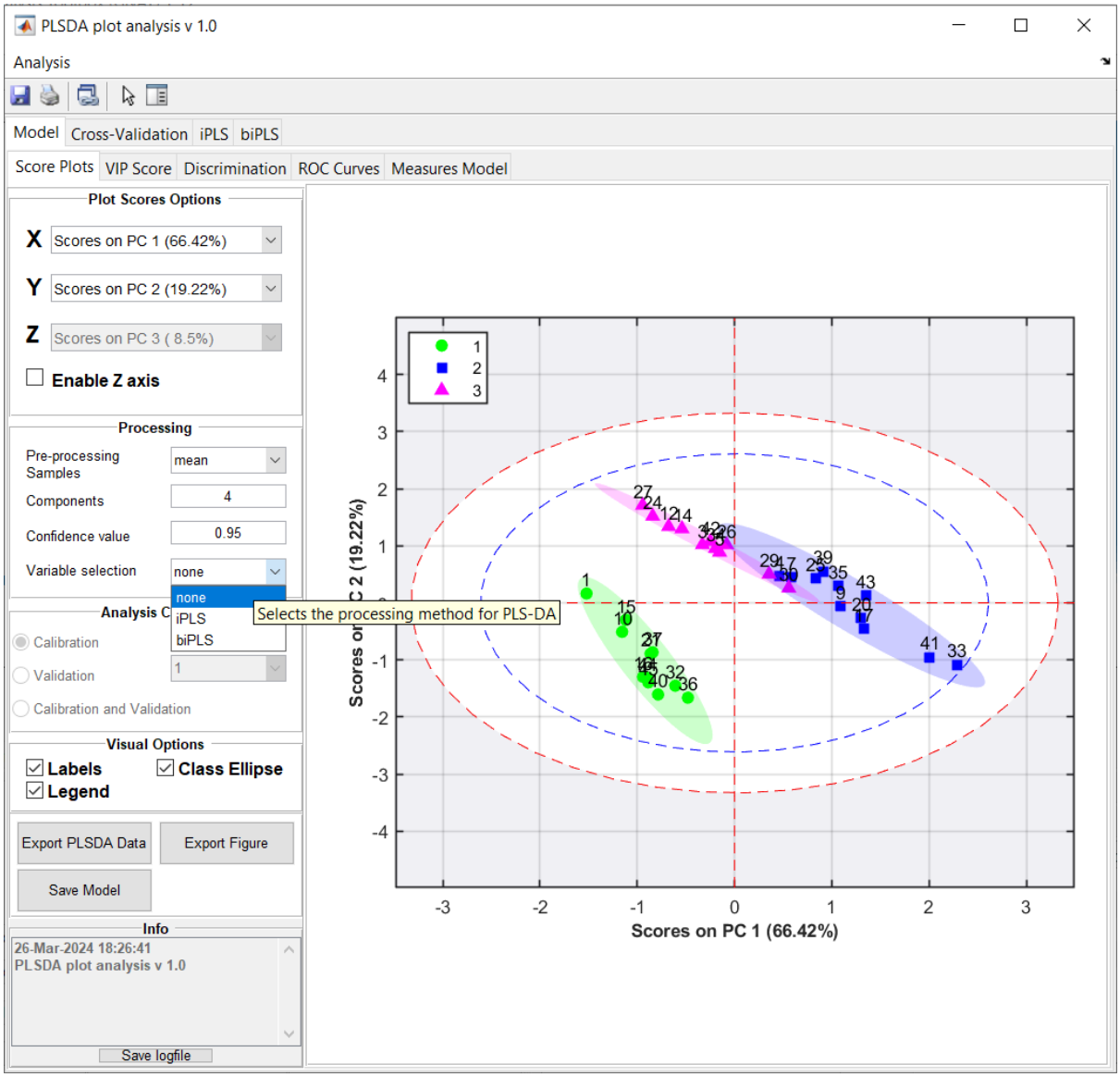
Scores Plot

The tab **Model** displays a visualisation 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 (circles), Rapeseed oil (squares) and Sunflower oil (triangles). The blue ellipse (–) represents the confidence limit for 95 % of confidence and the red ellipse (–) for 99 %.

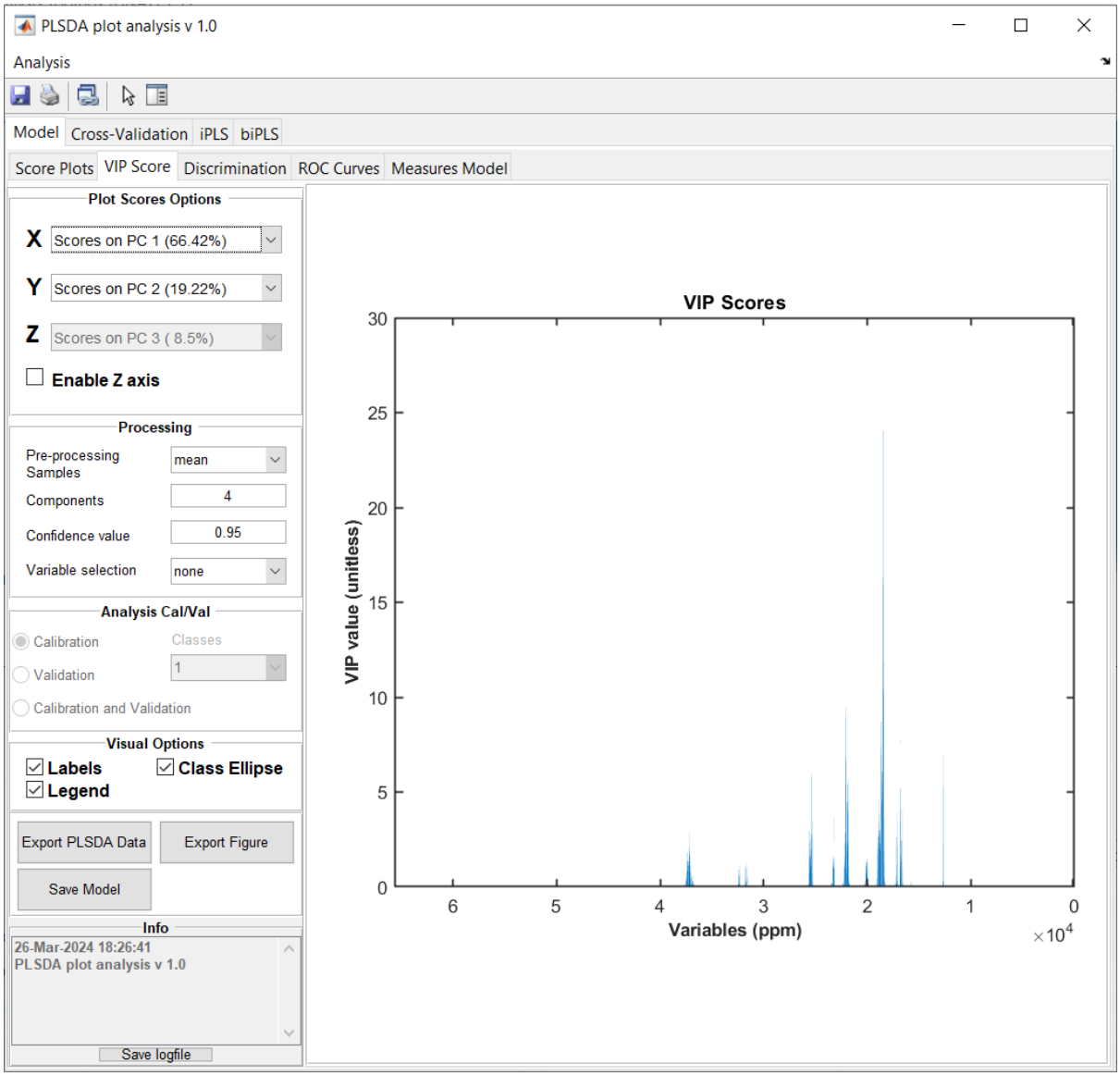


The **Processing panel** on the left allows the user to choose the preprocessing technique for the dataset's columns (e.g., Meancenter, Autoscale, or Pareto), as well as the number of latent variables, confidence value, and variable selection method. It is also able to toggle on and off the score plot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


PLSDA plot analysis v 1.0
Analysis
Model
Cross-Validation
iPLS
biPLS
Score Plots
VIP Score
Discrimination
ROC Curves
Measures Model
Plot Scores Options
X
Scores on PC 1 (66.42%)
Y
Scores on PC 2 (19.22%)
Z
Scores on PC 3 (8.5%)
Enable Z axis
Processing
Pre-processing
Samples
mean
Components
4
Confidence value
0.95
Variable selection
none
none
iPLS
biPLS
Selects the processing method for PLS-DA
Analysis C
Calibration
Validation
1
Calibration and Validation
Visual Options
Labels
Class Ellipse
Legend
Export PLSDA Data
Export Figure
Save Model
Info
26-Mar-2024 18:26:41
PLSDA plot analysis v 1.0
Save logfile
1
2
3
Scores on PC 1 (66.42%)

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

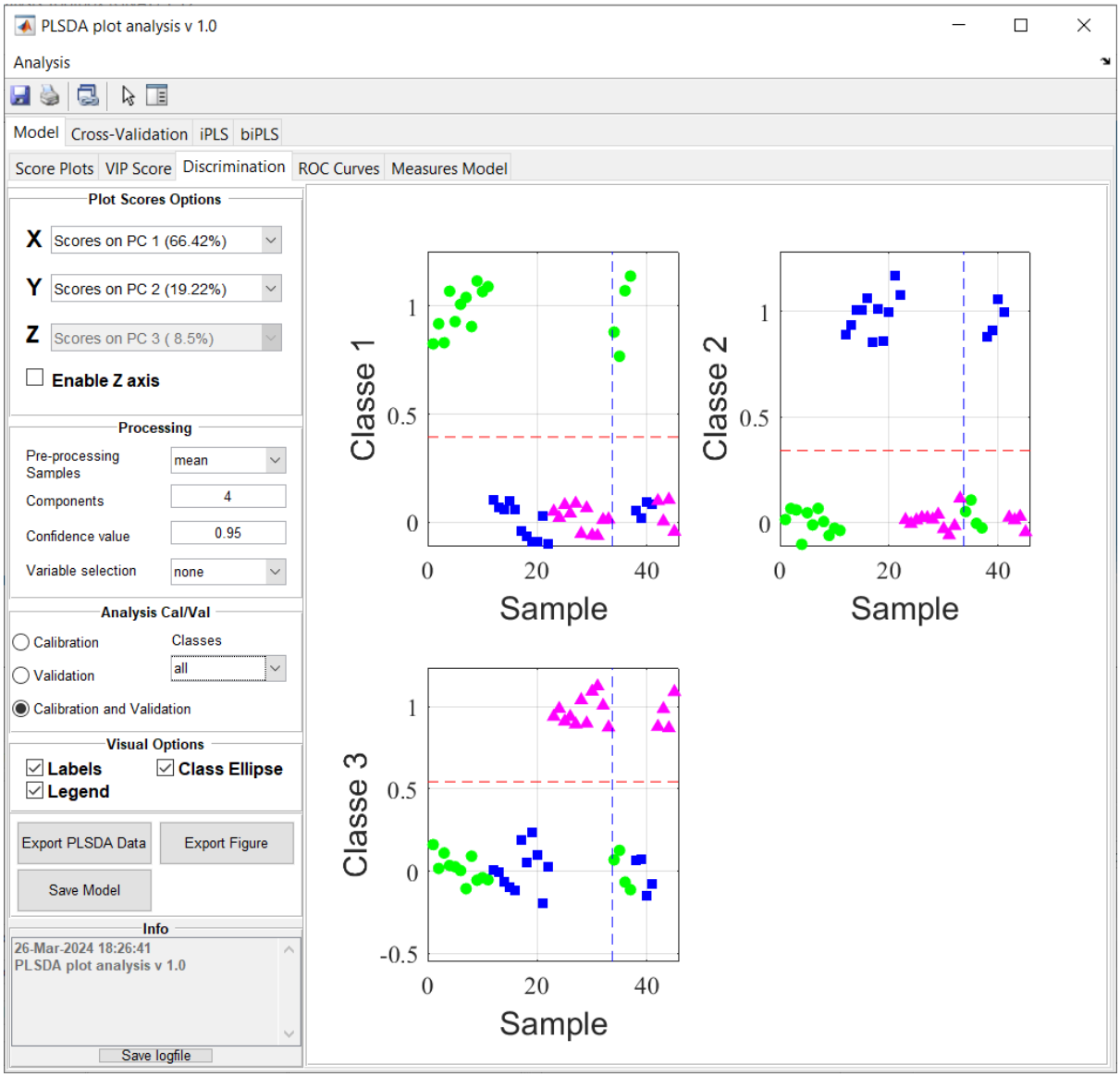


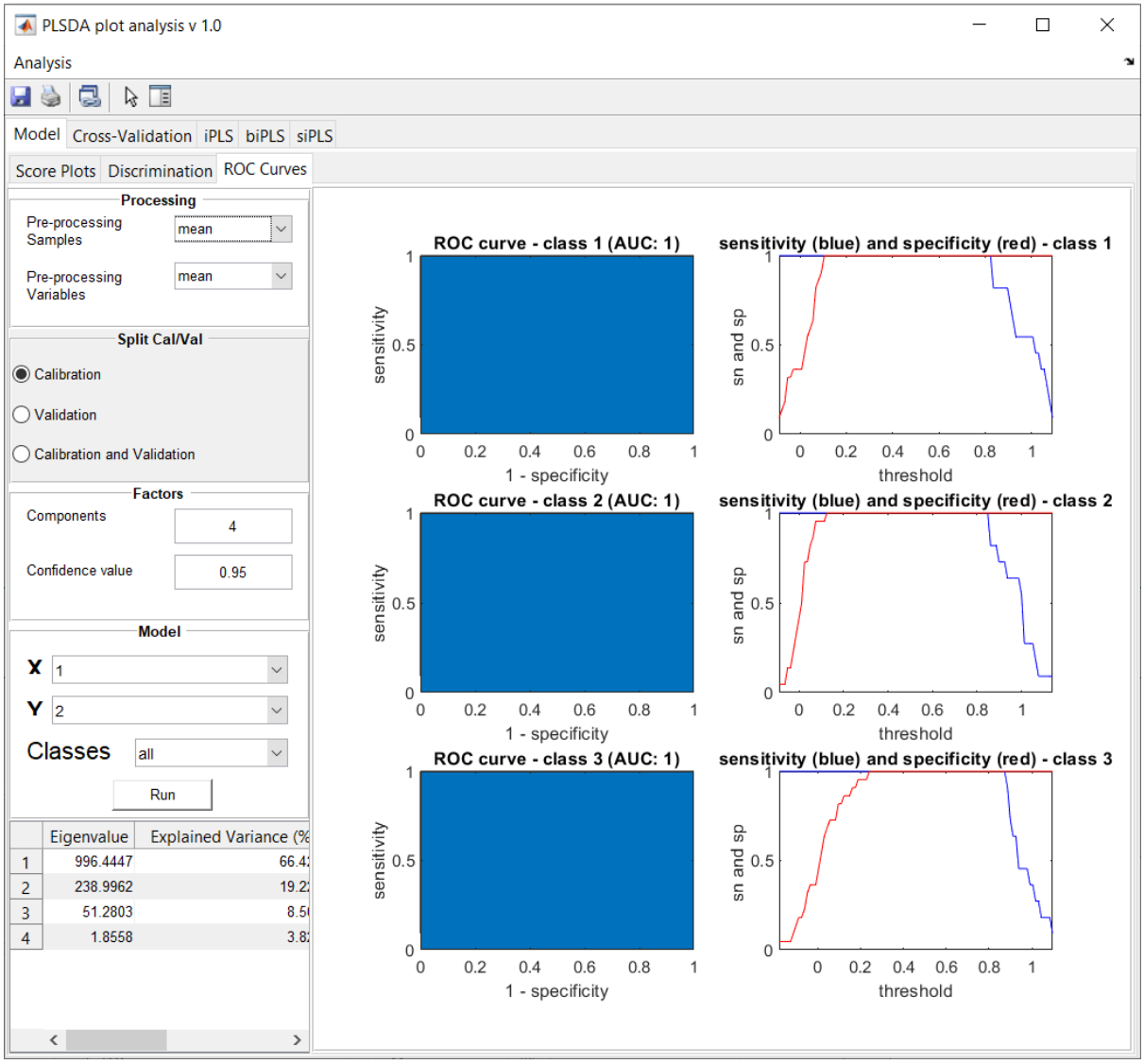
Discrimination plot

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.


PLSDA plot analysis v 1.0
Analysis
Model
Cross-Validation
iPLS
biPLS
Score Plots
VIP Score
Discrimination
ROC Curves
Measures Model
Plot Scores Options
X
Scores on PC 1 (66.42%)
Y
Scores on PC 2 (19.22%)
Z
Scores on PC 3 (8.5%)
Enable Z axis
Processing
Pre-processing
Samples
mean
Components
4
Confidence value
0.95
Variable selection
none
Analysis Cal/Val
Calibration
Validation
Calibration and Validation
Classes
all
Visual Options
Labels
Class Ellipse
Legend
Export PLSDA Data
Export Figure
Save Model
Info
26-Mar-2024 18:26:41
PLSDA plot analysis v 1.0
Save logfile
Classe 1
Classe 2
Classe 3
Sample
Sample
Sample


PLSDA plot analysis v 1.0
Analysis
Model
Cross-Validation
iPLS
biPLS
siPLS
Score Plots
Discrimination
ROC Curves
Processing
Pre-processing Samples
mean
Pre-processing Variables
mean
Split Cal/Val
Calibration
Validation
Calibration and Validation
Factors
Components
4
Confidence value
0.95
Model
X
1
Y
2
Classes
all
Run
Eigenvalue
Explained Variance (%
1
996.4447
66.4
2
238.9962
19.2
3
51.2803
8.5
4
1.8558
3.8
ROC curve - class 1 (AUC: 1)
sensitivity (blue) and specificity (red) - class 1
ROC curve - class 2 (AUC: 1)
sensitivity (blue) and specificity (red) - class 2
ROC curve - class 3 (AUC: 1)
sensitivity (blue) and specificity (red) - class 3
sensitivity
sn and sp
1 - specificity
threshold

OPLS-DA

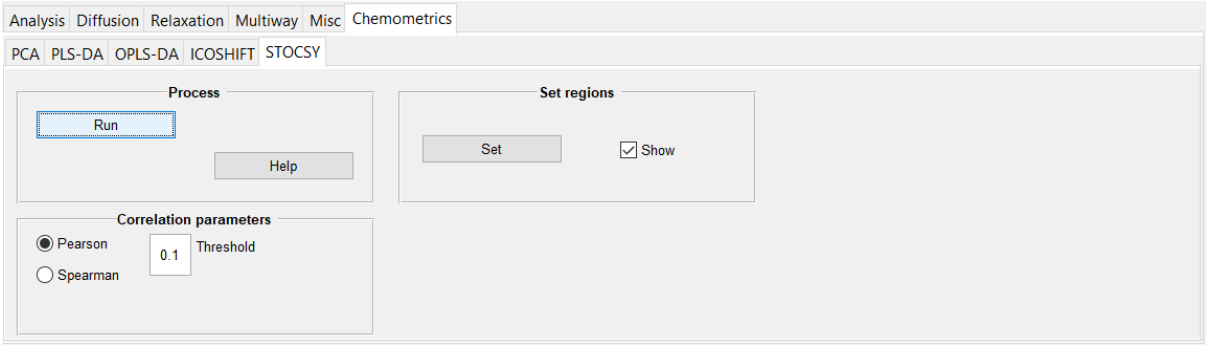
The OPLS-DA workflow in GNAT follows the same general setup as PLS-DA.

See the *PLS-DA* page for class definition, calibration/validation splitting, cross-validation, and model inspection.

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.

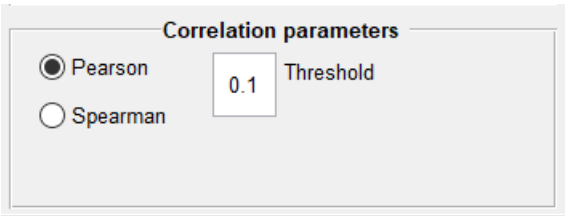
The main tab for STOCSY computation within GNAT is shown below:



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

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

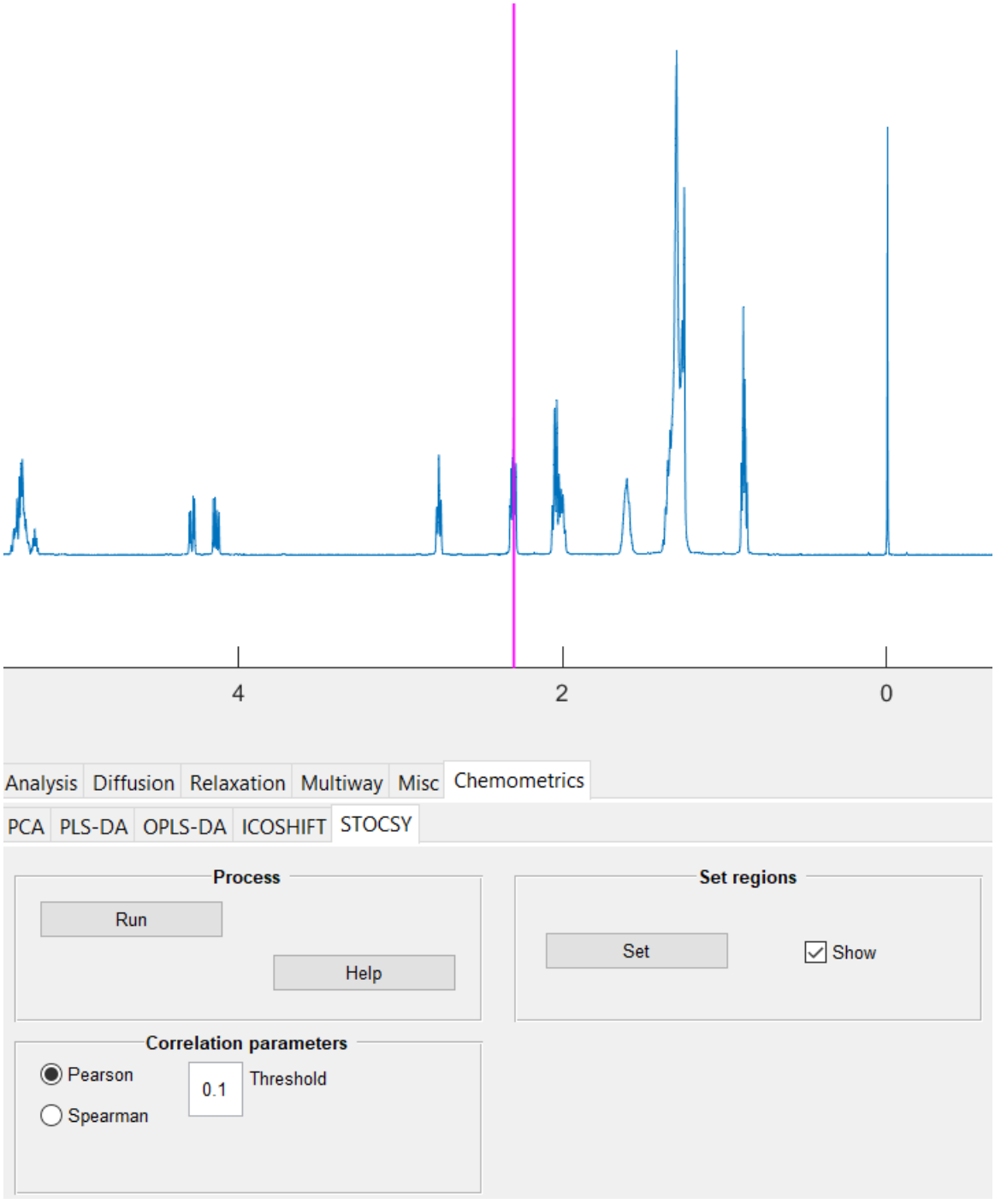
ρ = Spearman's rank correlation coefficient

d_i = difference between the two ranks of each observation

n = number of observations

Defining ppm region

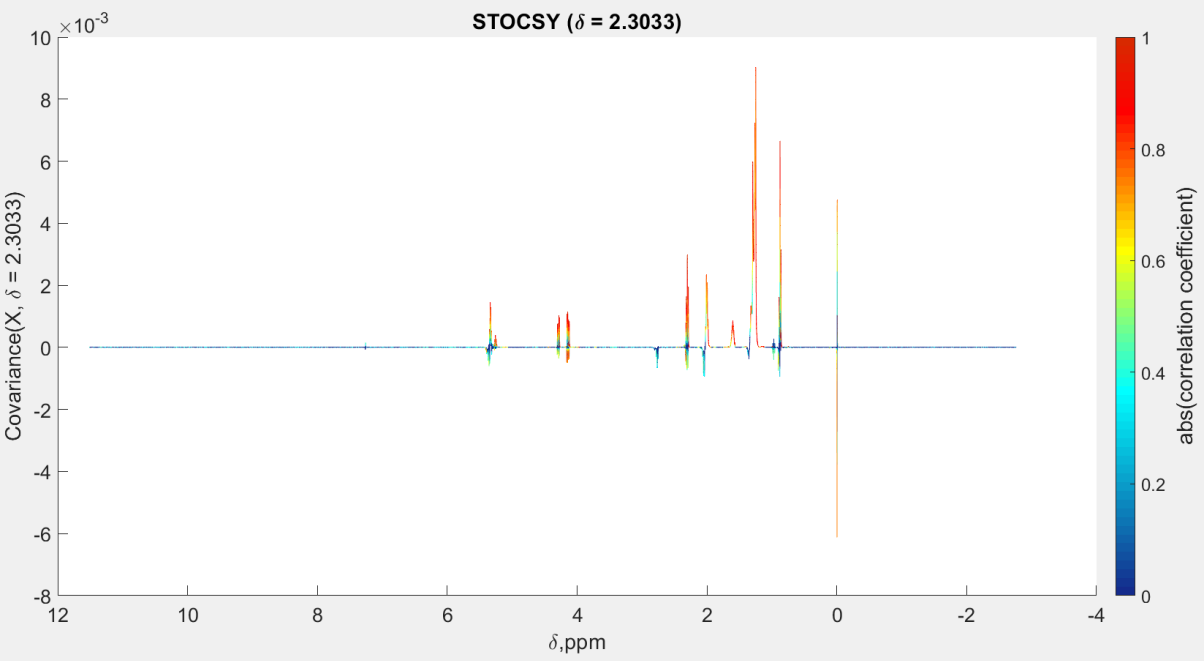
In STOCSY 1D the user can utilize the **Set region** panel to select the signal in the active spectrum in GNAT to be use in the analysis. The user can use the button `Set` after selecting the checkbox `Show`



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

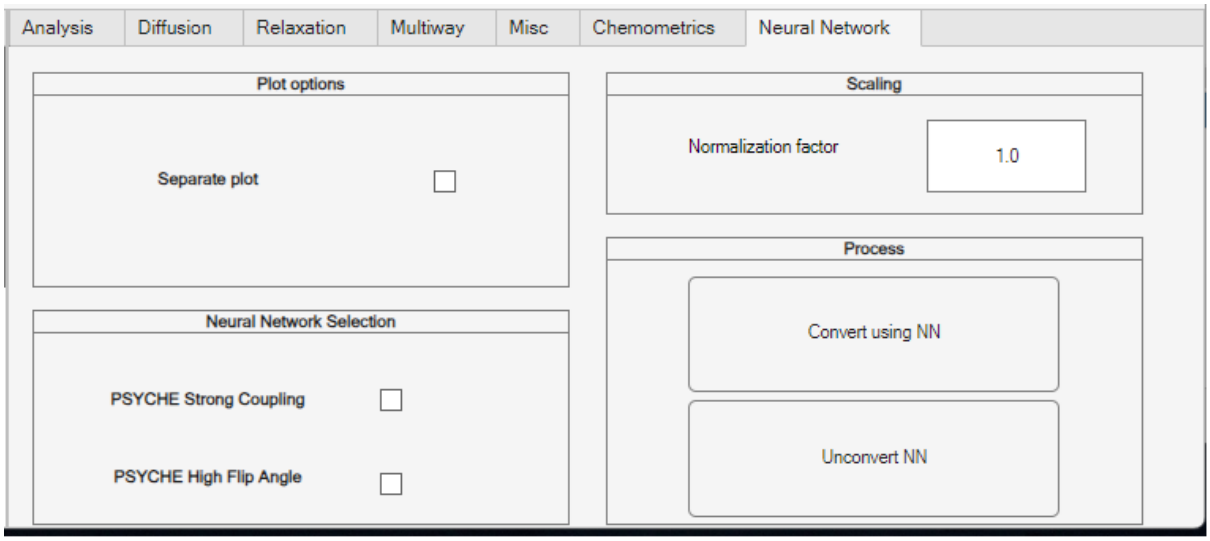
STOCSY calculation

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



6.20 Neural Network

This module provides neural network-based enhancement of NMR spectra within GNAT.



6.20.1 Functionalities

Neural Network

Overview

Neural networks are used to enhance experimental Pure Shift NMR spectra in GNAT.

Models

- EDHR_Net.onnx: Deals with large flip angle artifacts.[1]
- EDHR_Net_SC.onnx: Strong coupling cases.

Workflow

1. Load ONNX model
2. Normalize input
3. Apply prediction
4. Rescale output

MATLAB Implementation

```
net = importNetworkFromONNX(modelPath);
Y = predict(net, X);
```

Scaling

```
X = X./((max(abs(X)))/scaleVal);
```

Pure Shift Processing with Neural Network (Step-by-step guide)

This section describes the workflow for loading, processing, and converting Pure Shift NMR data using the neural network module in GNAT.

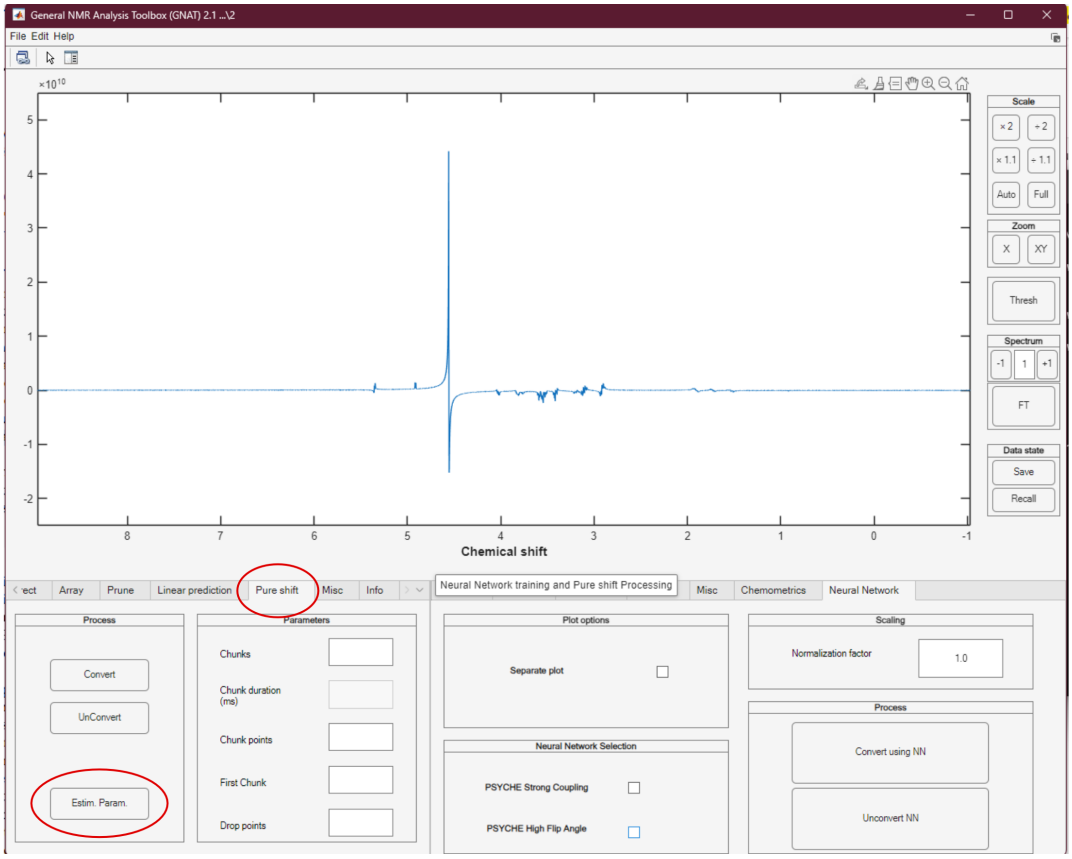
1. Data Loading

The data can be imported using two approaches:

File > Import Bruker > Select the directory containing the fid/ser file

a) Direct processing in GNAT

- Navigate to the *Pure Shift* tab.
- Click **Estim. Param.** to automatically extract acquisition parameters.
- Click **Convert** to generate the Pure Shift spectrum.

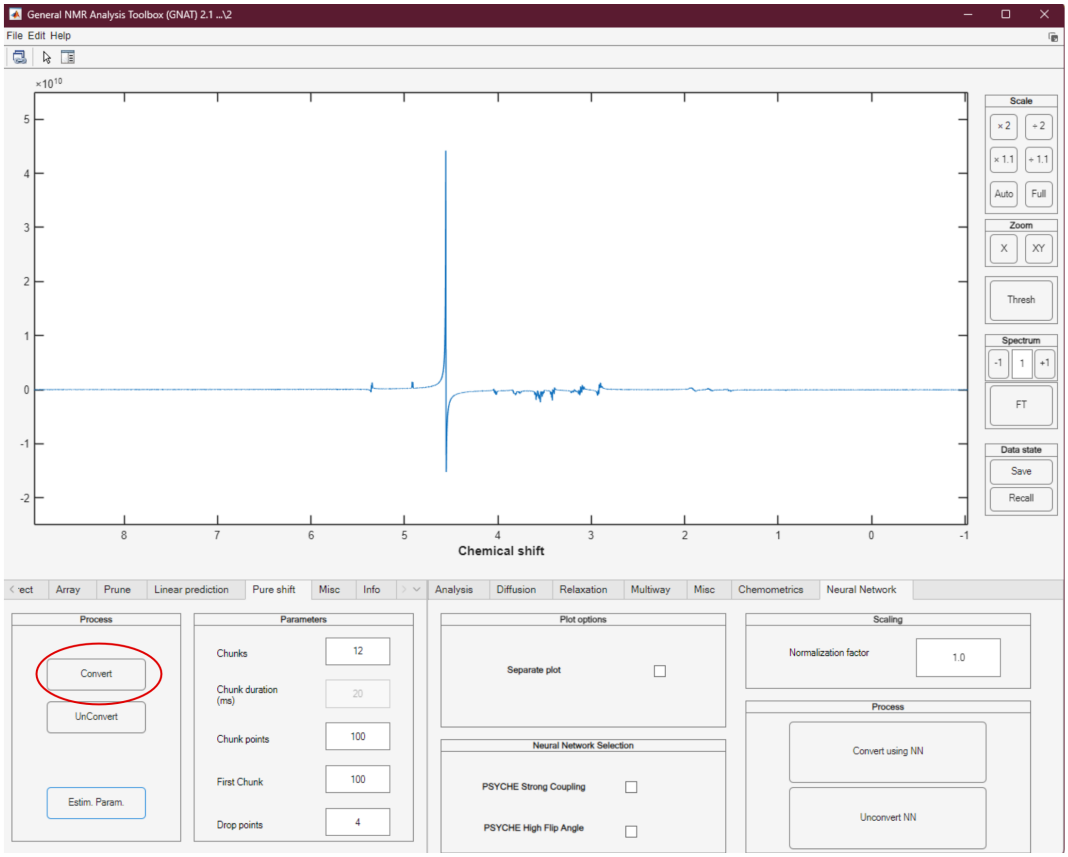


b) Pre-processing in TopSpin

Alternatively, process the dataset in Bruker TopSpin using the `pshift` macro:

- Apply the `pshift` macro in TopSpin.
- Load the processed dataset directly into GNAT.

> **Note**
>
> Phase correction can be applied either in TopSpin or GNAT. Improper phasing may degrade neural network performance.



2. Verification of Data Size

- Open the **FT** tab.
- Verify that the number of points is **4096**.

Important

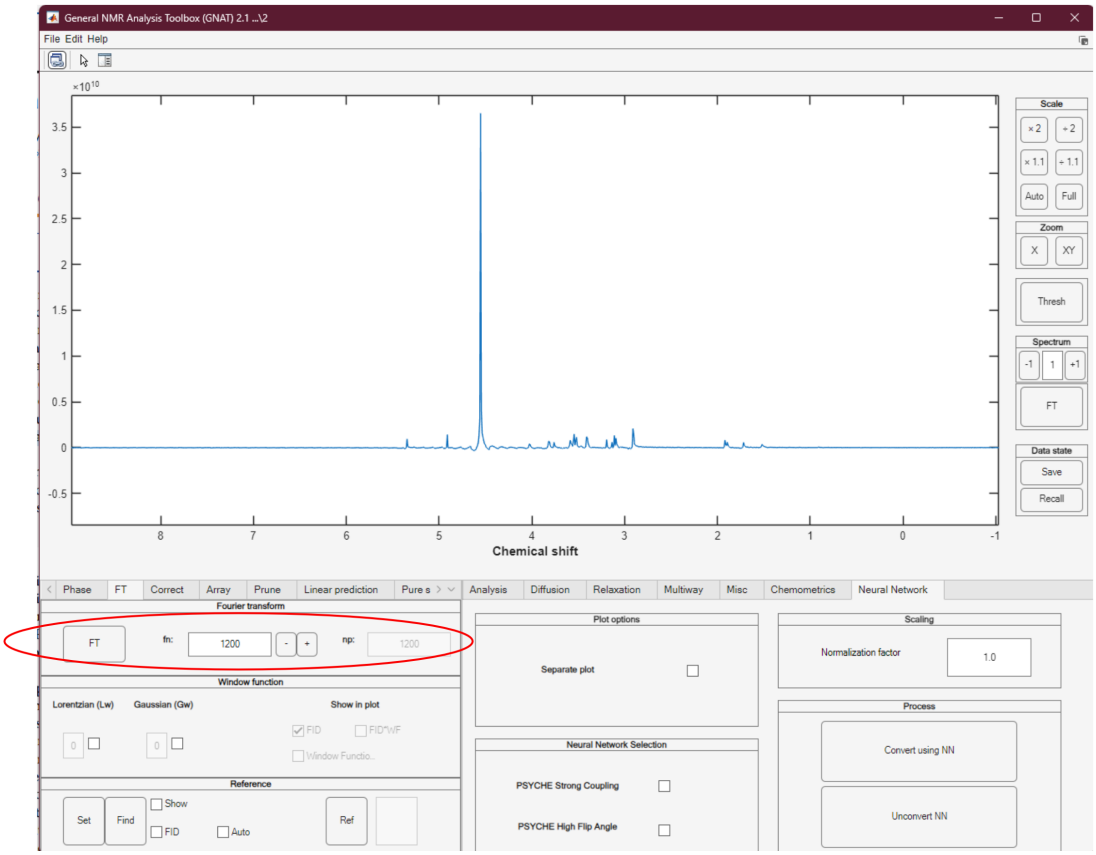
The neural network models require a fixed input size of 4096 points. Apply zero-filling if necessary.

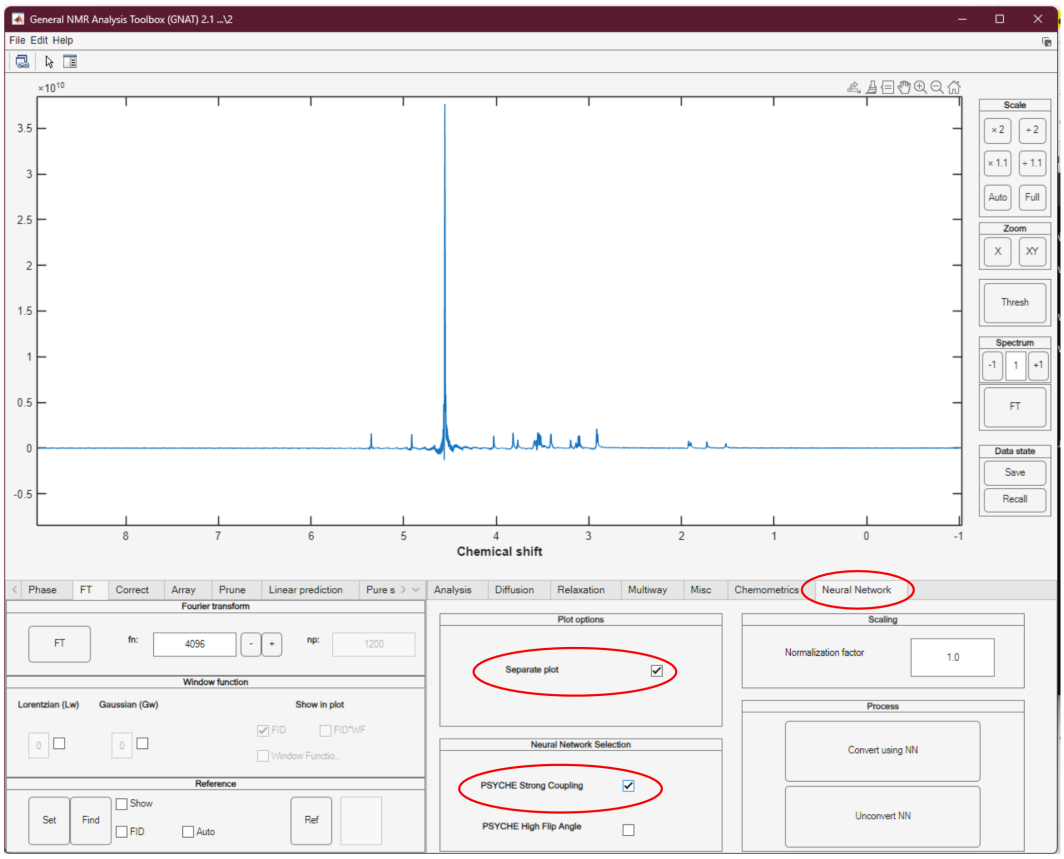
3. Neural Network Selection

- Navigate to the **Neural Network** tab.
- Select the desired model.
- Optionally enable **Separate Plot**.

Tip

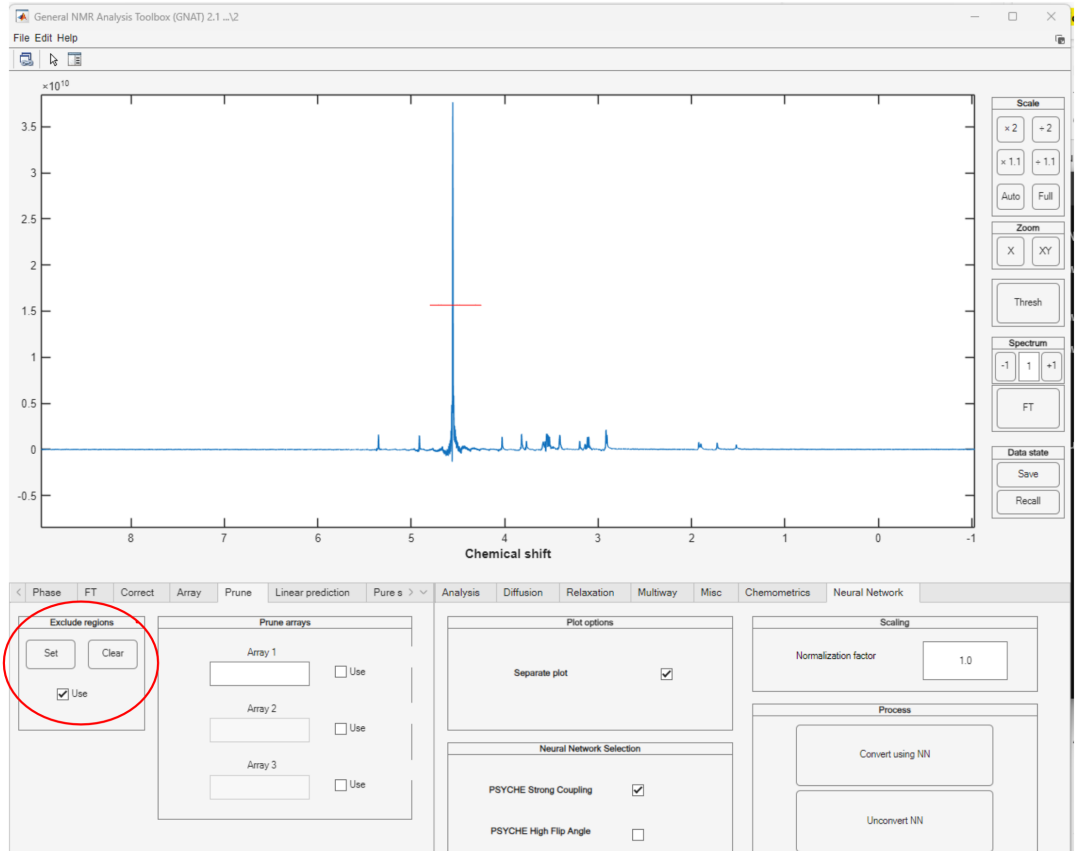
Useful for comparing input and output spectra.


General NMR Analysis Toolbox (GNAT) 2.1 ...\2
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Phase
FT
Correct
Array
Prune
Linear prediction
Analysis
Diffusion
Relaxation
Multiway
Misc
Chemometrics
Neural Network
Fourier transform
fn:
1200
np:
Window function
Lorentzian (Lw)
Gaussian (Gw)
Show in plot
Reference
Set
Find
Show
Auto
Ref
Plot options
Separate plot
Neural Network Selection
PSYCHE Strong Coupling
PSYCHE High Flip Angle
Scaling
Normalization factor
1.0
Process
Convert using NN
Unconvert NN


General NMR Analysis Toolbox (GNAT) 2.1 ...\2
File Edit Help
Chemical shift
Scale
Zoom
Thresh
Spectrum
FT
Data state
Save
Recall
Phase
FT
Correct
Array
Prune
Linear prediction
Analysis
Diffusion
Relaxation
Multiway
Misc
Chemometrics
Neural Network
Fourier transform
fn:
4096
np:
1200
Window function
Lorentzian (Lw)
Gaussian (Gw)
Show in plot
Reference
Set
Find
Show
Auto
Ref
Plot options
Separate plot
Neural Network Selection
PSYCHE Strong Coupling
PSYCHE High Flip Angle
Scaling
Normalization factor
1.0
Process
Convert using NN
Unconvert NN

4. Peak Pruning (Recommended)

- Open the **Prune** tab.
- Use **Exclude Regions** to remove unwanted peaks.



> **Warning**
>
> Strong peaks distort normalization and affect neural network output.

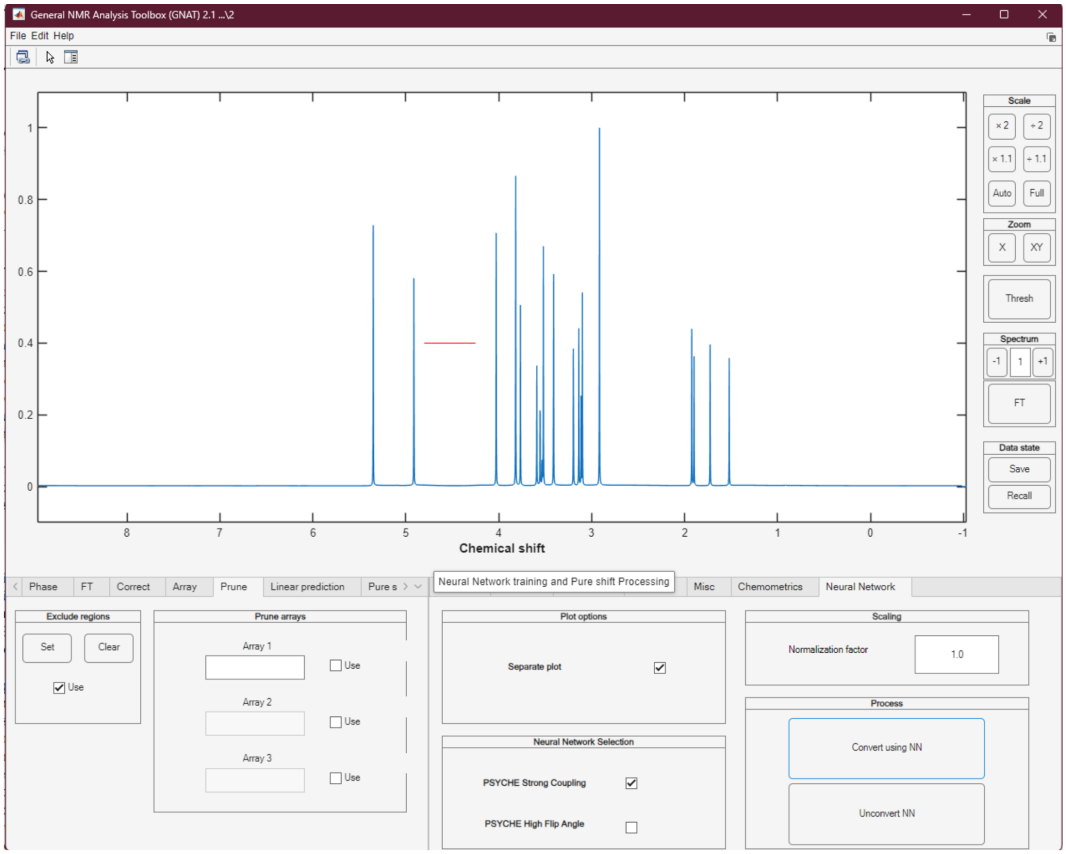
5. Neural Network Conversion

- Click **Convert using NN**.
- Monitor MATLAB command window or terminal for logs.

> **Note**
>
> Errors in model loading or formatting will appear in the terminal.

6. Reverting the Conversion

- Click **Unconvert NN** to restore the original spectrum.



Summary of Workflow

1. Import data
2. Ensure 4096 points
3. Select model
4. Prune peaks
5. Convert
6. Validate or revert

Notes

- Input must be 4096 points
- Scaling affects small peaks if pruning is not applied

References

[1] Sensitivity-Enhanced Pure Shift Spectroscopy Empowered by Deep Learning and PSYCHE, Xiaoxu Zheng, Wen Zhu, Xiaoqi Shi, Xinjing Gao, Yao Luo, Qing Zeng, Jiyang Dong, Zhong Chen, and Yanqin Lin Analytical Chemistry 2025, 97 (25), 13480–13486 DOI: 10.1021/acs.analchem.5c01899

CHAPTER

SEVEN

INDICES AND TABLES

- genindex
- modindex
- search