



NMRProcFlow

Macro-command Reference Guide

This document is the reference guide of the macro-commands

Daniel Jacob

UMR 1332 BFP, Bordeaux Metabolome - 2022



NMRProcFlow - Macro-command Reference Guide

Content

Pre-processing	3
airpls	5
align	6
bucket	7
clupa	10
gbaseline	11
normalisation	12
shift	14
zero	15



Pre-processing

Description

The spectral preprocessing for 1D NMR should be applied in case the input raw data are FID, either from a Bruker or Varian spectrometer. The term pre-processing designates here the transformation of the NMR spectrum from time domain to frequency domain, including the phase correction and the fast Fourier-transform (FFT).

Syntax

```
##% Vendor=string; Type=string; LB=numeric; ZF=0|2|4;  
    PHC1=TRUE|FALSE; ZNEG=TRUE|FALSE; TSP=TRUE|FALSE;
```

##%	This character sequence plays the role of a shebang at the beginning of the preprocessing command line. The different parameters must be written in the same line, and all of them based on the same model <i>variable=value</i> , and each of them separated by a semicolon.
Vendor	Vendor or nmrML - value must be taken in the set: bruker, varian, nmrm1.
Type	Type of the raw data - value must be taken in the set: fid, 1r. In case of the type is '1r', the preprocessing is not applied, and the others parameters are ignored.
LB	Line Broadening – Typical value is 0.3 (only if type=fid)
ZF	Zero Filling factor (i.e the factor of multiplication of the FID)
PHC1	indicates if the first order phase correction is applied
ZNEG	Zeroing Negative Values
TSP	indicates whether a 'zero reference signal' is present or not (i.e. TSP, TMS, DSS, ...)



Example

```
Vendor=nmrml; Type=fid; LB=0.3; ZF=4; PHC1=TRUE; ZNEG=FALSE; TSP=TRUE;
```

References

James Keeler (2011) Understanding NMR Spectroscopy, 2nd Edition, Ed John Wiley & Sons Ltd, [ISBN: 978-0-470-74608-0](#)



airpls

Description

The airPLS (adaptive iteratively reweighted penalized least squares) algorithm based on [Zhang et al., 2010] is a baseline correction algorithm that works completely on its own, and that only requires a “detail” parameter for the algorithm, called Lambda. Because this Lambda parameter can vary within a very large range (from 10 up to 1.e+06), we converted this parameter within a more convenient scale for the user, called 'level correction factor' chosen by the user from '1' (soft) up to '6' (high). The lower this level correction factor is set, the smoother baseline will be. Conversely, the higher this level correction factor is set, the more baseline will be corrected in details.

Syntax

```
airpls <ppm range> <lambda>
```

<ppm range>	The ppm zone to apply the baseline correction.
<lambda>	The detail parameter – the true detail parameter used in the algorithm is rescaled according to the formula : $10^{(6-\lambda)}$

Example

```
airpls 0.674 1.601 3
```

References

Zhang Z, Chen S, and Liang Y-Z (2010) Baseline correction using adaptive iteratively reweighted penalized least squares, Analyst, 2010,135, 1138-1146. [doi:10.1039/B922045C](https://doi.org/10.1039/B922045C)



align

Description

A peak alignment algorithm based on Least-squares. Align spectra either based on a particular spectrum chosen within the spectra set, or based on the average spectrum. In this latter case, the re-alignment procedure is executed three times, the average spectrum being recalculated at each time.

In order to limit the relative ppm shift between the spectra to be realigned and the reference spectrum, the user can set this limit by adjusting the parameter 'Relative shift max.' that sets the maximum shift between spectra and the reference. The range goes from 0 (no ppm shift allowed) up to 1 (maximum ppm shift equal to 100% of the selected ppm range).

Syntax

```
align <ppm range> <relative shift max> <groupid>
```

<ppm range>	The ppm zone to apply the baseline correction.
< relative shift max >	sets the maximum shift between spectra and the reference
<groupid>	The group id of the spectra to be aligned. In batch mode, only the 0 value is considered (i.e all values)

Example

```
align 4.241 4.687 0.11 0
```



bucket

Description

An NMR spectrum may contain several thousands of points, and therefore of variables. In order to reduce the data dimensionality binning or bucketing is commonly used. In binning, the spectra are divided into bins (so called buckets) and the total area (sum of each resonance intensity) within each bin is calculated to represent the original spectrum. The more simple approach (**Uniform bucketing**) consists in dividing all spectra with uniform spectral width (typically 0.01 to 0.04 ppm).

Due to the arbitrary division of spectrum, one bin may contain pieces from two or more resonances and one resonance may appear in two bins which may affect the data analysis. We have chosen to implement the Adaptive, Intelligent Binning method [De Meyer et al. 2008] that attempts to split the spectra so that each area common to all spectra contains the same single resonance, i.e. belonging to the same metabolite (**Intelligent bucketing**). In such methods, the width of each bucket is then determined by the maximum difference of chemical shifts among all spectra.

Such as AIBIN method, a method called ERVA (Extraction of Relevant Variables for Analysis) (Jacob et al, 2013) attempts to split the spectra so that each area common to all spectra contains the same single resonance, i.e. belonging to the same metabolite. This mathematical method is based on a convolution product between a spectrum (S) and the second-order derivative of the Lorentzian function (SDL). This products buckets that are exactly centered on each resonance having their width equal to twice the half-width of the resonance.

Finally, the bucketing type called **Variable-Size Bucketing** allows to simply provide a list of static bucket zones without any parameter.

Syntax

```
bucket <binning type> <ppm noise> <resolution> <snr>  
<ppm range 1>  
<ppm rang 2>  
...  
EOF
```

Or

```
bucket vsb  
<ppm range 1>
```



<ppm rang 2>

...

EOF

<binning type>	The binning method can be either ' unif ' for uniform bucketing, ' aibin ' for intelligent bucketing, ' erva ' for Extraction of Relevant Variables for Analysis, or ' vsb ' for variable-size bucketing
<ppm noise>	Only for ' aibin ', ' erva ' & ' unif ': The ppm range considered as noisy zone. Indeed the algorithm needs to estimate the noise level to be more efficient.
<resolution>	Only for ' aibin ', ' erva ' & ' unif ': The resolution parameter: typically from 0.01 to 0.04 ppm for the uniform bucketing, from 0.1 to 0.6 for the intelligent bucketing, and from 0.001 to 0.0001 for ERVA.
<snr>	Only for ' aibin ', ' erva ' & ' unif ': The Signal-Noise Ratio defines the minimal intensity levels to be taken into account for the binning
<ppm range n>	the ppm zones to apply the binning

Example

```
bucket aibin 10.2 10.5 0.5 5
0.6 4.72
4.865 5.5
6.3 9.5
EOL

bucket vsb
5.431 5.406
5.249 5.225
4.669 4.631
EOL
```

References



De Meyer, T., Sinnaeve, D., Van Gasse, B., Tsiorkova, E., Rietzschel, E. R., De Buyzere, M. L., et al. (2008). NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Analytical Chemistry*, 80(10), 3783–3790. [doi: 10.1021/ac7025964](https://doi.org/10.1021/ac7025964)

Jacob D., Deborde C. and Moing A. (2013). An efficient spectra processing method for metabolite identification from ¹H-NMR metabolomics data. *Analytical and Bioanalytical Chemistry* 405(15) 5049–5061 [doi: 10.1007/s00216-013-6852-y](https://doi.org/10.1007/s00216-013-6852-y)



clupa

Description

A peak alignment algorithm, called hierarchical Cluster-based Peak Alignment (CluPA), aligns a target spectrum to the reference spectrum in a top-down fashion by building a hierarchical cluster tree from peak lists of reference and target spectra and then dividing the spectra into smaller segments based on the most distant clusters of the tree.

Syntax

```
clupa <ppm noise> <ppm range> <resolution> <snr> <groupid>
```

<ppm noise>	The ppm range considered as noisy zone. Indeed the algorithm needs to estimate the noise level to be more efficient.
<ppm range>	The ppm zone to apply the baseline correction.
<resolution>	The resolution parameter given a ppm interval. It defines the smallest ppm interval to be aligned.
<snr>	The Signal-Noise Ratio defines the minimal intensity levels to be taken into account during the peak finding step
<groupid>	The group id of the spectra to be aligned. In batch mode, only the 0 value is considered (i.e all values)

Example

```
clupa 10.2 10.5 4.241 4.687 0.01 5 0
```

References

Vu T.N., Valkenborg D., Smets K., Verwaest K.A., Dommissie R., Lemièrre F., Verschoren A., Goethals B., Laukens K. (2011) An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. BMC Bioinformatics.;12:405 [doi: 10.1186/1471-2105-12-405](https://doi.org/10.1186/1471-2105-12-405)



gbaseline

Description

The global baseline correction was based on [Bao et al., 2012], but only two phases were implemented: i) Continuous Wavelet Transform (CWT), and ii) the sliding window algorithm. The user must choose the correction level, from 'soft' up to 'high'.

Syntax

```
gbaseline <ppm noise> <ppm range> <window size> <smoothing size>
```

<ppm noise>	The ppm range considered as noisy zone. Indeed the algorithm needs to estimate the noise level to be more efficient.
<ppm range>	The ppm zone to apply the baseline correction.
<window size>	The width (number points) of the convolution window used for the baseline estimation.
<smoothing size>	The width (number points) of the convolution window used for the smoothed spectrum calculation

Example

```
gbaseline 10.2 10.5 -0.5 11 75 50
```

References

Bao, Q., Feng, J., Chen, F., Mao, W., Liu, Z., Liu, K., et al. (2012). A new automatic baseline correction method based on iterative method. *Journal of Magnetic Resonance*, 218, 35–43. [doi: 10.1016/j.jmr.2012.03.010](https://doi.org/10.1016/j.jmr.2012.03.010).



normalisation

Description

In order to make all spectra comparable with each other, the variations of the overall concentrations of samples have to be taken into account. We propose two normalization methods.

In NMR metabolomics, the total intensity normalization (called the Constant Sum Normalization) is often used so that all spectra correspond to the same overall concentration. It simply consists in normalizing the total intensity of each individual spectrum to a same value.

The Probabilistic Quotient Normalization [Dieterle et al. 2006] assume that biologically interesting concentration changes influence only parts of the NMR spectrum, while dilution effects will affect all metabolite signals. Probabilistic Quotient Normalization (PQN) starts by the calculation of a reference spectrum based on the median spectrum. Next, for each variable of interest the quotient of a given test spectrum and reference spectrum is calculated and the median of all quotients is estimated. Finally, all variables of the test spectrum are divided by the median quotient.

Syntax

```
normalisation <method>  
<ppm range 1>  
<ppm rang 2>  
...  
EOF
```

<method>	The normalization method can be either ' CSN ' for Constant Sum Normalization or ' PQN ' for Probabilistic Quotient Normalization.
----------	--

<ppm range <i>n</i> >	the ppm zones to apply the normalization
-----------------------	--



Example

```
normalisation PQN  
5.089 9.547  
0.651 4.68  
EOL
```

References

Dieterle F., Ross A., Schlotterbeck G. and Senn H. (2006). Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. Analytical Chemistry, 78:4281-4290. [doi:10.1021/ac051632c](https://doi.org/10.1021/ac051632c)

Kohl SM, Klein MS, Hochrein J, Oefner PJ, Spang R, Gronwald W. (2012) State-of-the art data normalization methods improve NMR-based metabolomic analysis, Metabolomics 146-160, [doi: 10.1007/s11306-011-0350-z](https://doi.org/10.1007/s11306-011-0350-z)



shift

Description

To effectively align certain delicate ppm zones, it is possible to shift beforehand these zones on one or more spectra (corresponding to the same factor level) in order to align afterwards these ppm zones including all spectra.

Syntax

```
shift <ppm range> <ppm shift value> <groupid>
```

<ppm range>	The ppm zone to apply the shift.
<ppm shift>	The ppm shift value applied on the ppm zone
<groupid>	The group id of the spectra to be aligned. In batch mode, only the 0 value is considered (i.e all values)

Example

```
shift 4.241 4.687 0.01 2 3
```



zero

Description

Put to zero all the ppm ranges given within a list

Syntax

```
zero  
<ppm range 1>  
<ppm rang 2>  
...  
EOF
```

<ppm range <i>n</i> >	The ppm zones to apply the zeroing.
-----------------------	-------------------------------------

Example

```
zero  
4.705 4.881  
5.484 6.333  
EOL
```