

# ***ProMetab***

**MATLAB™ code for *Processing Metabolite* profiles derived from NMR spectra**

**Release 1.1 (December 2004)**

**Mark R. Viant  
University of Birmingham, UK**

# Table of Contents

Table of Contents	1
Acknowledgements	2
1. Introduction	
1.1 About this software	3
1.2 Conditions of use	3
1.3 User support and contact details	3
1.4 Disclaimer	4
1.5 New features in Release 1.1	4
2. How to use the software	
2.1 Filenames and location	5
2.2 Format for NMR data	5
2.3 The <i>PM_input_parameters.m</i> file	5
2.4 <i>ProMetab</i> tutorial	5
2.5 Multivariate data analysis	9

## **Acknowledgements**

Several agencies have provided support for the development of this software, particularly via personnel funding. These include:

- The National and California Sea Grant College Programs of the U.S. Department of Commerce's National Oceanic and Atmospheric Administration.
- University of California Toxic Substances Research and Teaching Program, U.S.A.
- Natural Environment Research Council, U.K.

# 1. Introduction

## 1.1 About this software

*ProMetab* software is a metabolomics data processing tool that converts raw Bruker NMR spectra into a format for multivariate chemometric analysis. It is written in MATLAB<sup>TM</sup> (The MathWorks, Inc.; [www.mathworks.com](http://www.mathworks.com)), which provides a technical computing environment for high-performance numeric computation and visualization.<sup>1</sup> *ProMetab* can interpret both 1D and projections of 2D NMR spectra, and then segments the data into chemical shift bins of user-defined width. Following removal of unwanted spectral features, the User can compress specific groups of bins into single segments to minimize the effects of pH-induced shifting of the NMR peaks. Various normalisation strategies and data transformations are available to the User. Multivariate analyses of the processed data can subsequently be performed using a MATLAB<sup>TM</sup> toolbox such as PLS\_Toolbox (Eigenvector Research, Inc.; [www.eigenvector.com](http://www.eigenvector.com)), or exported and analysed using other statistical packages.

## 1.2 Conditions of use

This software is provided free of charge to the scientific community. However, I request that interested parties contact me directly should they want a copy (see Section 1.3). The maintenance of a User list will enable immediate notification of any bugs or improvements in the code. I encourage Users who improve or add to the existing code to send me their updates so that these can be implemented and then re-distributed to other Users (with appropriate acknowledgements). Furthermore, to facilitate notification of this resource, I request that Users reference the following paper in their publications:

M. R. Viant, Improved Methods for the Acquisition and Interpretation of NMR Metabolomic Data, *Biochem. Biophys. Res. Comm.* **310**, 943-948 (2003).

Finally, no part of this software may be used, or integrated into other products, for commercial benefit.

## 1.3 User support and contact details

Should you encounter problems with the software, have suggestions for improving the code, or would like an updated copy of the program, please contact me at:

Dr. Mark R. Viant  
School of Biosciences  
The University of Birmingham  
Birmingham B15 2TT  
United Kingdom

Phone: +44-(0)121-414-2219

FAX: +44-(0)121-414-5925

Email: [M.Viant@bham.ac.uk](mailto:M.Viant@bham.ac.uk)

URL: <http://www.biosciences.bham.ac.uk/staff/staff.htm?ID=81>

---

<sup>1</sup> MATLAB is a registered trademark of The MathWorks, Inc.

## ***1.4 Disclaimer***

While this software has been thoroughly tested using many NMR datasets, no guarantee is given for accuracy of results.

## ***1.5 New features in Release 1.1***

This version of the software includes one additional algorithm over Release 1.0. The original log transformation of the binned data has been replaced with a generalized log transformation [P. V. Purohit, D. M. Rocke, M. R. Viant and D. L. Woodruff, Discrimination Models using Variance Stabilizing Transformation of Metabolomic NMR Data, *OMICS*, **8**, 118-130 (2004)]. This transformation stabilizes the variance across the bins within the dataset and has the effect of increasing the weightings of the less intense peaks. This is particularly valuable when using principal components analysis since it circumvents the problem of only a few peaks dominating the analysis. Other changes to *ProMetab* are planned during 2005. These include a completely new graphical user interface, and increased flexibility to handle both Bruker and Varian file formats.

## 2. How to use the software

### 2.1 Filenames and location

The following files are supplied with *ProMetab*:

<i>ProMetab</i> User Manual v1.1	- this manual
<i>ProMetab_v1_1.m</i>	- main data processing script
<i>PM_input_parameters.m</i>	- contains several input parameters for <i>ProMetab</i>
<i>PM_read_1Ddata.m</i>	- function to read in Bruker NMR spectral data
<i>PM_read_procs.m</i>	- function to read in Bruker processing parameters

All these files should be copied to your MATLAB™ working directory. If a new working directory is created then make sure to point MATLAB™ to this location using *File -> Set Path*. Add the new working directory to the top of the search path.

### 2.2 Format for NMR data

NMR spectra must have been recorded on a Bruker spectrometer. *ProMetab* is compatible with data collected on both UNIX and Windows systems. All spectra must have the same file name and be numbered sequentially (although can start at any value). The NMR spectra should be copied to the MATLAB™ working directory, retaining the *same* file and directory structure as created by Bruker.

### 2.3 The *PM\_input\_parameters.m* file

Once optimized, many of the input parameters for *ProMetab* remain fixed for a given analysis. To avoid re-entering these parameters for each analysis, they have been stored in an M-file called *PM\_input\_parameters.m*, which is automatically read by *ProMetab*. The User is required to edit these parameters from within the MATLAB™ Command Window by typing:

```
edit pm_input_parameters
```

The file is heavily commented in order to explain each of the parameters to the User. In general, the parameters should initially be set to default values and then as the analysis proceeds, and the various parameters are optimized by the User, the *PM\_input\_parameters.m* file will require updating.

### 2.4 *ProMetab* tutorial

Run the software from the MATLAB™ Command Window by typing:

```
prometab_v1_1
```

You will be prompted for various details concerning the NMR files:

**Process 1-D or 2-D spectra? [default = 1]:**  
**Enter filename of NMR data:**  
**How many spectra require processing?**  
**Enter file number of first file [default = 1]:**

Currently the '2-D' option only includes the 1-D projections of 2-D J-resolved spectra.

Then you will be prompted for various 'binning' options.

**Enter start of binning [default = 10.0 ppm]:**  
**Enter end of binning [default = 0.2 ppm]:**  
**Enter bin stepsize [default = 0.005 ppm]:**

Note that the minimum value for the 'end of binning' is forced to 0.2 ppm. The program will then read in the Bruker XWINNMR data, and report...

### **Reading Bruker file *n***

...for each file *n*.

You will then be asked if you want the raw unnormalized NMR data presented. The default option is 'no' since viewing unnormalized data is often unnecessary.

**Display raw unnormalized 1-D spectra (yes=1)? [default = no]:**

If bin compression was requested in file *PM\_input\_parameters.m*, then the following message appears:

**Compressing bins now...**

Next you enter the normalization phase, and are presented with several options:

**Enter normalization procedure?**  
**(1) TSA, (2) TMSP, (3) TMSP w/ sample scaling, (4) none [default = 1]:**

These include:

1. Normalize each spectrum to the Total Spectral Area (TSA) of that binned spectrum, effectively yielding a total area of one.
2. Normalize each spectrum to the area of the TMSP NMR standard.
3. Normalize each spectrum to the area of the TMSP NMR standard, and additionally multiply by a unique sample scaling factor listed in file *PM\_input\_parameters.m*. This is useful if you know a particular sample was re-constituted at, for example, one-half the concentration of all other samples.
4. Apply no normalization procedure at all.

The results of normalizing the data will be displayed (for example):

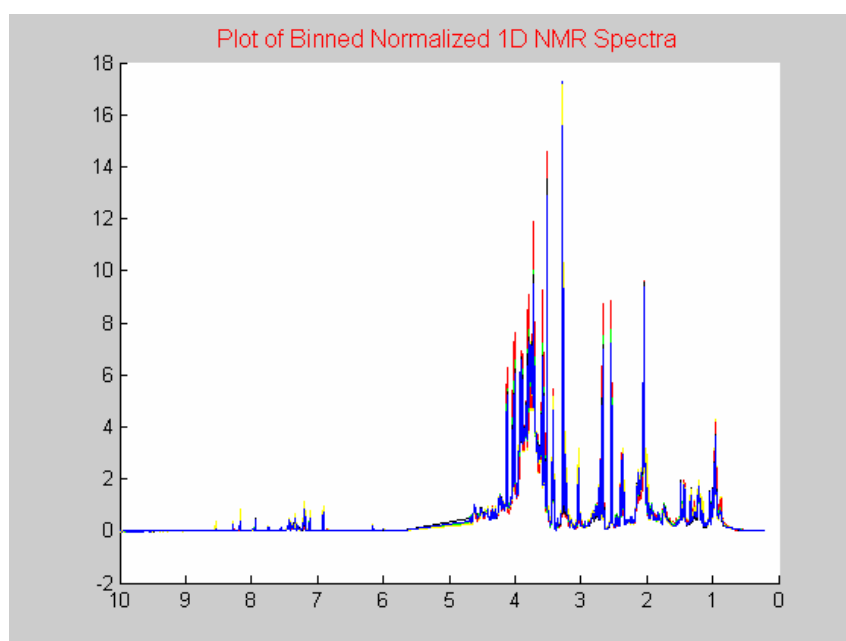
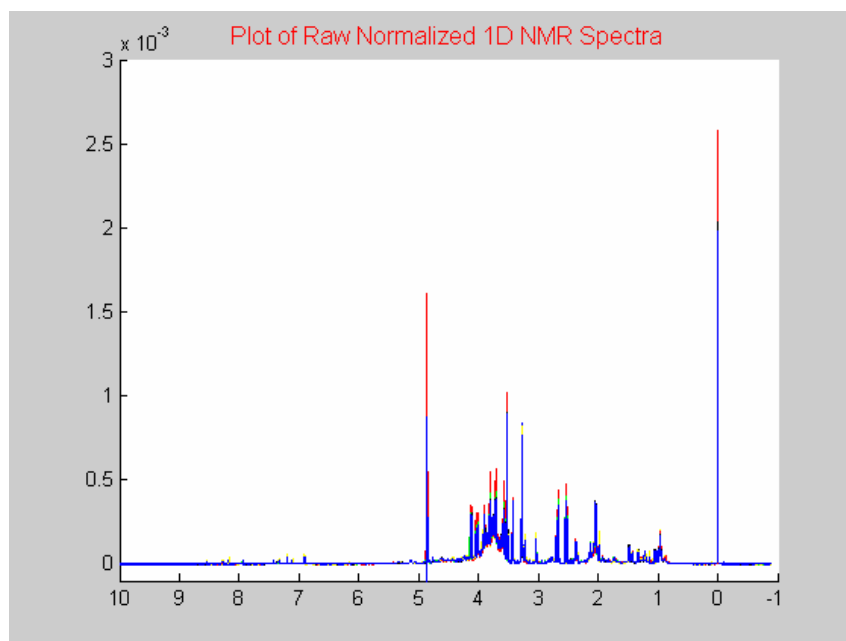
<b>Original TSA</b>	<b>TMSP area</b>	<b>Normalized TSA</b>
<b>3.4187e+010</b>	<b>2.1778e+009</b>	<b>1.0000e+003</b>

**3.1399e+010 2.0825e+009 1.0000e+003**  
**3.0522e+010 2.1590e+009 1.0000e+003**  
**2.7355e+010 2.1571e+009 1.0000e+003**  
**3.1290e+010 2.1778e+009 1.0000e+003**

You will then be asked if you want the raw normalized NMR data presented, and the binned normalized data presented:

**Display raw normalized 1-D spectra (yes=1)? [default = yes]:**  
**Display binned normalized 1-D spectra (yes=1)? [default = yes]:**

These figures will appear as (for example):





Note that in the binned spectra the TMS standard at 0.0 ppm has been cropped, and the unwanted water resonance at ca. 4.7 ppm has been removed.

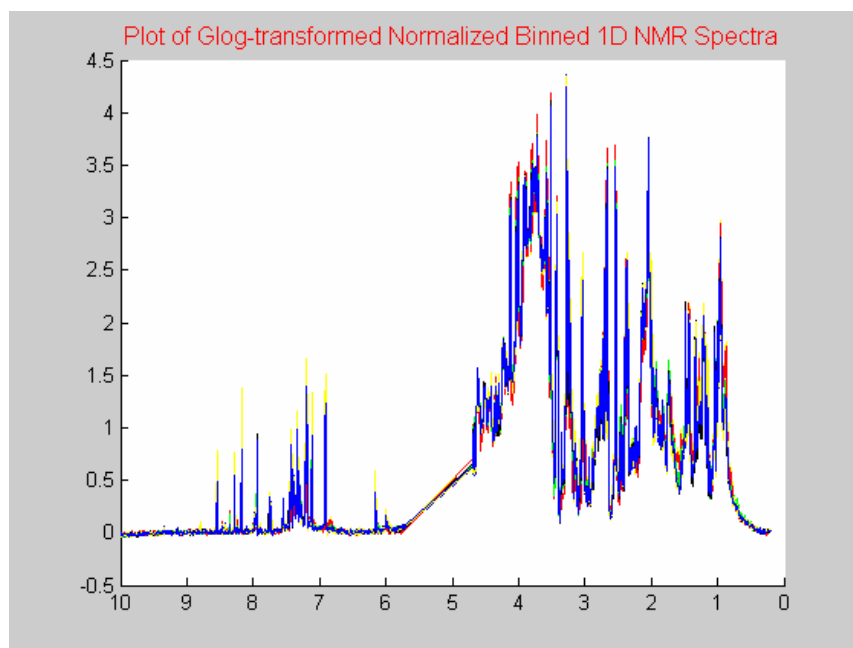
Finally you will be asked if you wish to transform the data using the generalized log transformation, and the value of the transformation parameter,  $\lambda$ , is echoed to the screen:

**Enter transformation procedure (1) generalized log, (2) none [default = 1]:**  
**lambda = 0.2**

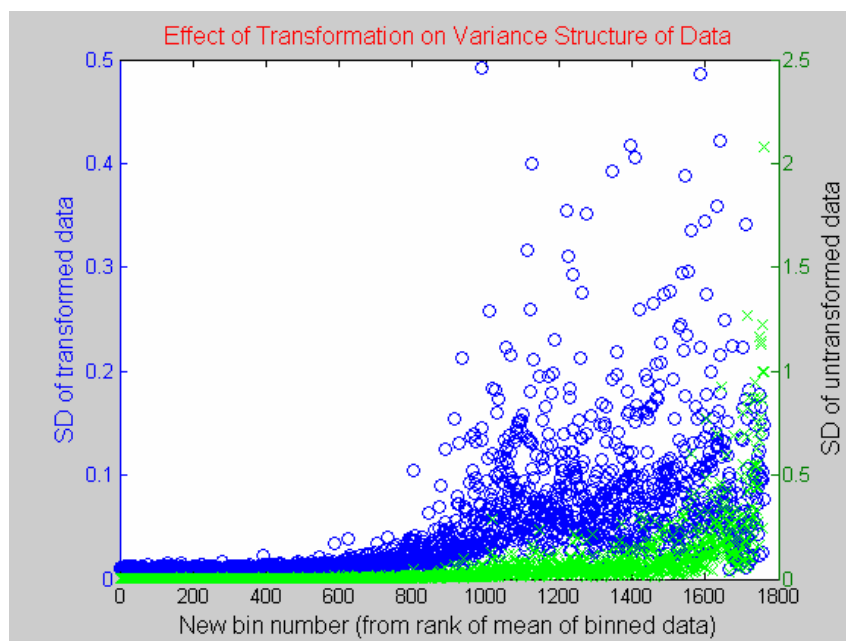
Note that optimisation of the transformation parameter is described in the file *PM\_input\_parameters.m*. Then you are prompted to plot the transformed, binned, normalized NMR data, and also to plot the relationship between the bin number (re-ordered according to the ranked means of the binned data) and standard deviation of the pre- and post-transformed data, which facilitates the optimisation of  $\lambda$ :

**Display transformed binned 1-D spectra (yes=1)? [default = yes]:**  
**Display bin vs. SD of pre- and post-transformed data (yes=1)? [default = no]:**

These figures will appear as (for example):



The beneficial effect of the generalized log transformation is readily apparent. The weaker spectral features have been increased in intensity relative to the stronger peaks, yet the noise (e.g. between 9-10 ppm) is still very small in comparison to the peaks.



The second beneficial effect of the generalized log transformation is the stabilization of the variance across the bins, avoiding the problem of a very limited number of bins having significantly larger variances than the majority of bins (pre-transformation = green crosses, and post-transformation = blue circles).

## 2.5 Multivariate data analysis

The processed NMR data will now be in a format for multivariate data analysis, such as principal components analysis. At this point the User has a number of choices. The first (and simpler) option is to conduct these analyses within MATLAB™. This can be done, for example, using the PLS\_Toolbox available from Eigenvector Research, Inc. ([www.eigenvector.com](http://www.eigenvector.com)), which contains a collection of chemometric routines. Both principal components analysis and partial least squares regression can be conducted via convenient graphic user interfaces.

The second option is to export the processed NMR data in ASCII format, and subsequently load into another multivariate data analysis package. To save the generalized log-transformed, binned and normalized NMR data as space-delimited ASCII (where each row corresponds to a sample, and each column corresponds to a bin), type the following in the MATLAB™ Command Window:

```
save <filename> spec_array_glognbin -ascii
```

To save the data without transformation, replace *spec\_array\_glognbin* with *spec\_array\_nbin*. The file will be saved in the MATLAB™ working directory. Then import and process this data in your own multivariate analysis software package.

Finally, to save the MATLAB™ workspace, use *File -> Save Workspace As...*