Package 'batman'

Installation and Testing

Table of Contents

1. INSTALLATION INSTRUCTIONS **1**

2. TESTING **3**

Test 1: Single spectrum from designed mixture data *3*

Test 2: Multiple spectra from designed mixture data *4*

Test 3: Multiple spectra from designed mixture data (adjusted relative intensities) *7*

Test 4: Multiple spectra from bacterial supernatants data *8*

3. CHEMICAL SHIFT SORTING **9**

1. Installation instructions

For all platforms:

Install R 2.12.1, or higher (note restriction for Mac binary below).

Install R packages doSNOW and plotrix using the following command in R:

```
install.packages("doSNOW")

install.packages("plotrix")
```

User can specify where to install the package in argument lib. For detailed help of install.packages(), type in R:

```
?install.packages
```

Windows & Linux: Install using online repository:

R install command:

```
install.packages("batman", repos="http://R-Forge.R-project.org")
```

Mac: Download and install from local copy:

There are two options:

1. Download & install from Mac binary

Note that, due to restrictions on R-Forge, the current Mac binary is not compatible with R 3.0.x.

Download the Mac binary file "**batman_1.2.1.tgz**" from (the size of downloaded file should be around 2.7MB):

https://r-forge.r-project.org/scm/viewvc.php/pkg/batman_1.2.1.tgz?root=batman&view=log

And in R type the following command:

```
install.packages("/path/of/batman_1.2.1.tgz", repos = NULL)
```

The fortan compiler (http://cran.r-project.org/bin/macosx/tools/) may be required in order to run BATMAN in some case.

2. Download & install from source

Download the source code file "**batman_1.2.1.tar.gz**" from (the size of downloaded file should be around 2.1MB):

https://r-forge.r-project.org/scm/viewvc.php/pkg/batman_1.2.1.tar.gz?root=batman&view=log

And in R type the following command (make sure you have a C++ compiler is installed before installing from source code and possibly fortan compiler as well which can be downloaded from http://cran.r-project.org/bin/macosx/tools/):

```
install.packages("/path/of/batman_1.2.1.tar.gz", repos = NULL, type="source")
```

(**Note**: in this guide, users should replace the *italic* characters in directory name with the directory in their cases.)

If you receive error messages on installation, check if the required packages doSNOW and plotrix are installed. Also check size of downloaded files. If no error messages are produced, proceed to testing below.

2. Testing

Load batman package in R:

```
library(batman)
```

There are 4 Tests included in this guide using, a) designed mixture data [1] and b) bacterial supernatant data [2]:

- Test 1 is a quick run of batman to check if it is installed properly.
- Test 2 shows how to run batman on the six designed mixture spectra to estimate levels of 6 metabolites.
- Test 3 uses the designed mixture data to explain how to modify the properties of the spectral templates, specifically the relative intensities of each multiplet, thus improving the fit.
- Test 4 is a real biological sample, a bacterial supernatant, with larger number of spectra and possibly more metabolites to fit, which may take longer time to run.

The run times stated in this document were obtained on a Mac Pro with a dual 3.0 GHz quad core processors, 13 GB RAM.

Test 1: Single spectrum from designed mixture data

This is a simple test checking if the package has been installed correctly and can run. The number of iterations is deliberately set very low and one should not expect a particularly good fit. It should take about half minute to run.

Run batman with included test dataset (designed mixture data) in user specified directory:

```
bm<-batman(runBATMANDir = '/user/specified/directory')
```

Or just use the current directory:

```
bm<-batman()
```

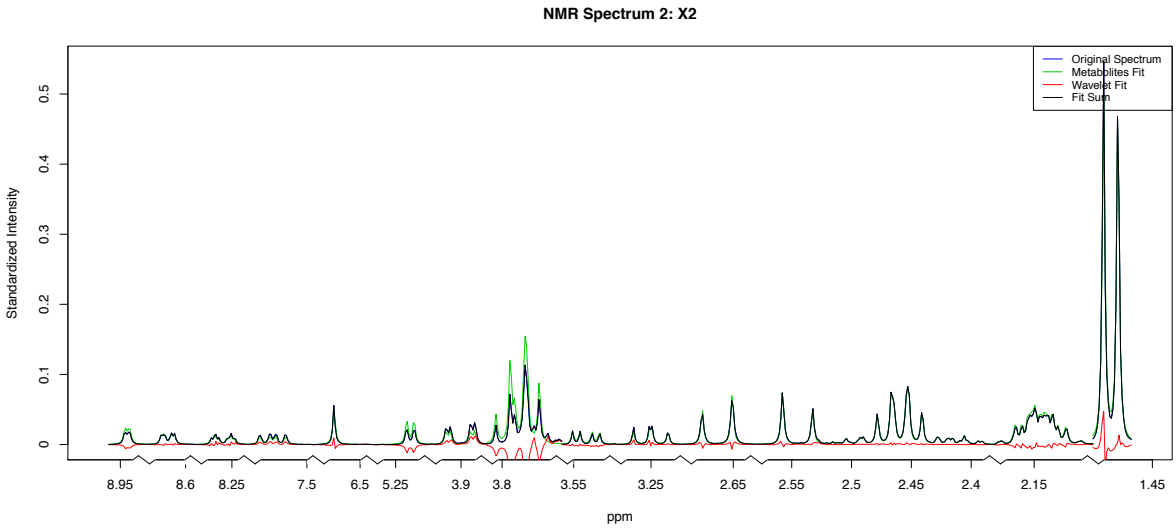
Output from R:

```
Running batman...
Number of burn-in iterations:  200
Number of post-burn-in iterations:  100

The template file used is
 2: The user input template of multiplets in multi_data_user.csv file.

Loading multi_data_user.csv...
  |                                                                      |   0%
Size of each spectrum is 682.
Size of metabolite list is 6.
of which 6 have resonances in/near the specified region and will be fit.
Constructing chain data structure...
time used is 1 seconds.
Running MCMC...
|=============================================================         |  67%
```

```
time used for burnin is 16 seconds.
|=========================================================================================| 100%  |
time used is 23 seconds.
saving posteriors...

time elapsed
 24.205
 second.
Reading in saved data in folder
/user/specified/directory/runBATMAN/BatmanOutput/31_Oct_16_03_38
Completed.
```

The green lines shown above may only show when using R from terminal (in Mac and Linux) or CMD (in Windows). Time elapsed will vary depending on the machine power.

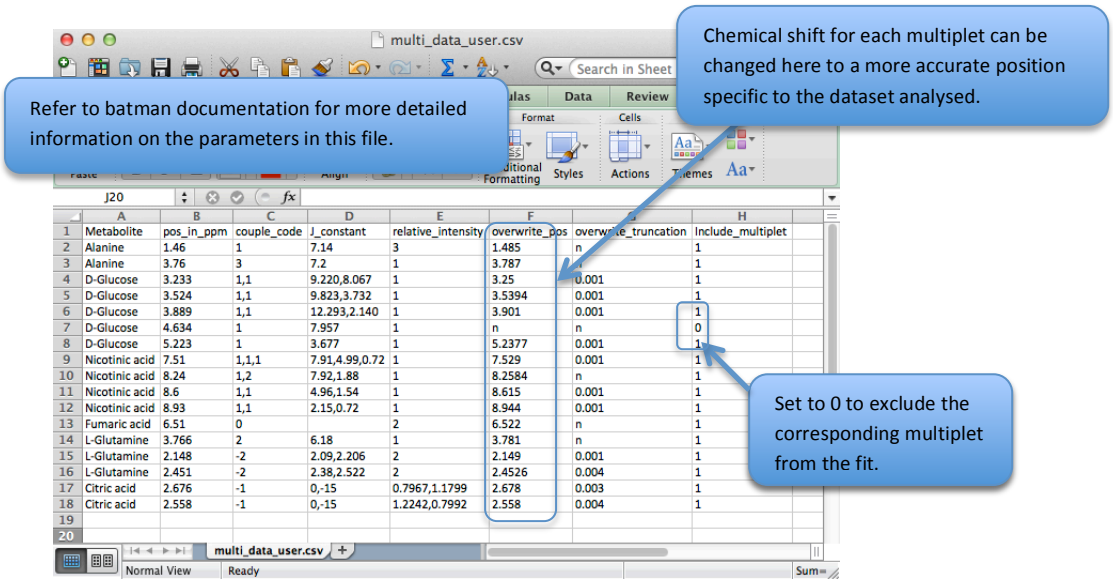
The mean posterior fit result similar to the one below will be shown in a figure window. A PDF file of this plot with the name "specFit_2_.pdf" can be found in folder "*/user/specified/directory/*runBATMAN/BatmanOutput/*date_time*".



Test 2: Multiple spectra from designed mixture data

This test fits the six designed mixture spectra and uses the parallel processes function to process a separate spectrum on each core. The final example output has been processed with a large number of iterations, which takes about 11 minutes to run (using six cores, thus about the same time needed to run one spectrum), but the user can try first with less iterations.

Below is a screen shot of the multiplet template file "*/user/specified/directory/*runBATMAN/BatmanInput/multi_data_user.csv" used for this test. The user does not need to change anything in this file at the moment. In the next test, details will be given on how to change the relative intensities of multiplets to improve the fit.



	A	B	C	D	E	F	G	H
1	Metabolite	pos_in_ppm	couple_code	J_constant	relative_intensity	overwrite_pos	overwrite_truncation	Include_multiplet
2	Alanine	1.46	1	7.14	3	1.485	n	1
3	Alanine	3.76	3	7.2	1	3.787		1
4	D-Glucose	3.233	1,1	9.220,8.067	1	3.25	0.001	1
5	D-Glucose	3.524	1,1	9.823,3.732	1	3.5394	0.001	1
6	D-Glucose	3.889	1,1	12.293,2.140	1	3.901	0.001	1
7	D-Glucose	4.634	1	7.957	1	n	n	0
8	D-Glucose	5.223	1	3.677	1	5.2377	0.001	1
9	Nicotinic acid	7.51	1,1,1	7.91,4.99,0.72	1	7.529	0.001	1
10	Nicotinic acid	8.24	1,2	7.92,1.88	1	8.2584	n	1
11	Nicotinic acid	8.6	1,1	4.96,1.54	1	8.615	0.001	1
12	Nicotinic acid	8.93	1,1	2.15,0.72	1	8.944	0.001	1
13	Fumaric acid	6.51	0		2	6.522	n	1
14	L-Glutamine	3.766	2	6.18	1	3.781	n	1
15	L-Glutamine	2.148	-2	2.09,2.206	2	2.149	0.001	1
16	L-Glutamine	2.451	-2	2.38,2.522	2	2.4526	0.004	1
17	Citric acid	2.676	-1	0,-15	0.7967,1.1799	2.678	0.003	1
18	Citric acid	2.558	-1	0,-15	1.2242,0.7992	2.558	0.004	1
19								
20								

Further details are given in the batman documentation which can be found at https://r-forge.r-project.org/scm/viewvc.php/*checkout*/documentation%20and%20test/batman.pdf?root=batman.

Open file "batmanOptions.txt" in folder "*/user/specified/directory*/runBATMAN/BatmanInput", find the following parameter lines:

```
specNo - Ranges of spectra number to be included (e.g. 1,3-4 etc.):  2
paraProc - No of parallel processes (multicores) (only 1 core will be used for single spectrum): 1
nItBurnin - Number of burn-in iterations: 200
nItPostBurnin - Number of post-burn-in iterations: 100
```

And change their parameters to:

```
specNo - Ranges of spectra number to be included (e.g. 1,3-4 etc.):  1-6
paraProc - No of parallel processes (multicores) (only 1 core will be used for single spectrum): 6
nItBurnin - Number of burn-in iterations: 7000
nItPostBurnin - Number of post-burn-in iterations: 1000
```

Save the options file.

This will change the input from spectrum number 2 to all the 6 (1 to 6) spectra in test dataset.

Note:
paraProc parameter: This depends on the number of cores on the user's machine.
nItBurnin & nItPostBurnin parameters: user can test with less iterations which will need less time to run.

Now run the same command used in Test 1:

Run batman with included test dataset (designed mixture data) in user specified directory:

```
bm<-batman(runBATMANDir = '/user/specified/directory')
```

Or just use the current directory:

```
bm<-batman()
```

Output from R:

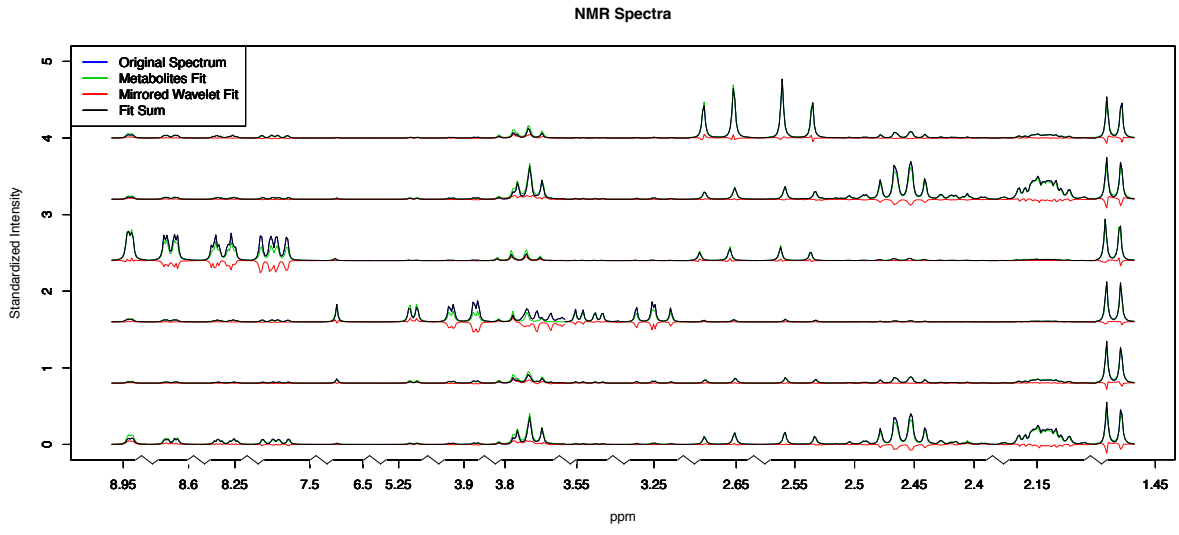
```
Running batman...
Number of burn-in iterations:  7000
Number of post-burn-in iterations:  1000

The template file used is
 2: The user input template of multiplets in multi_data_user.csv file.

Loading multi_data_user.csv...

Number of parallel processes (multicores) used to run the multi-spectra analysis:  6

 Parallel processing of multi spectra currently cannot display the progress
 bar (or any words), please be patient for the results :)

time elapsed
 643.854
 second.
Reading in saved data in folder
/user/specified/directory/runBATMAN/BatmanOutput/date_time
Completed.
```

The 3 figures with the results for all 6 spectra (not shown here) will be generated.

Alternatively to inspect the results, the user can plot all 6 spectra in a single stack plot as follows:

```
plotBatmanFitStack(bm,offset = 0.8,placeLegend = 'topleft',yto = 5)
```

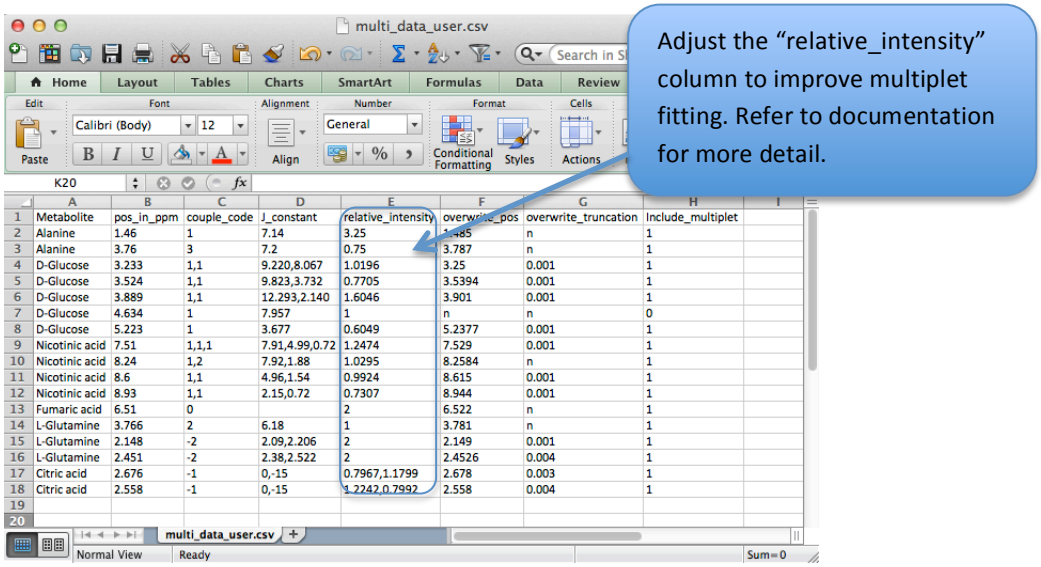
A stack plot of the fit result is shown below: (pdf file with the name "specFit_stack_.pdf" can be found in folder "*/user/specified/directory/*runBATMAN/BatmanOutput*/date_time*")



Test 3: Multiple spectra from designed mixture data (adjusted relative intensities)

Often, the relative intensities of resonances do not reflect the theoretical ratios expected from the molecular structure. This could be due to, for example, incomplete relaxation of different spin systems, and varies according to the pulse sequence and sample type. The effect may lead to poor fitting. This test uses the same spectra as in Test 2 but with some adjustment of relative intensities of the multiplet template used to attempt to improve the fit. It should take about the same time as of Test 2.

Below is a screen shot of multiplet template file "*/user/specified/directory*/runBATMAN/BatmanInput/multi_data_user.csv" used for this test. The user can either change the indicated column to the same value as shown, or rename the file "multi_data_user.csv" used in Test 2 to a different one, and then change the file name from "*/user/specified/directory*/runBATMAN/BatmanInput/multi_data_user_adj.csv" to "*/user/specified/directory*/runBATMAN/BatmanInput/multi_data_user.csv", so that batman will use the adjusted template parameters.

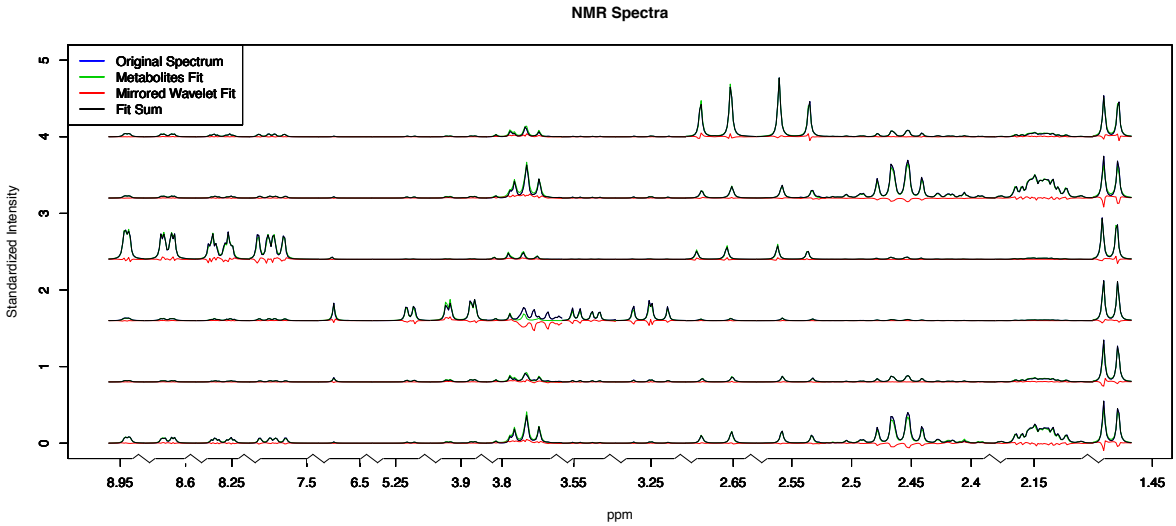


	A	B	C	D	E	F	G	H	I
1	Metabolite	pos_in_ppm	couple_code	J_constant	relative_intensity	overwrite_pos	overwrite_truncation	Include_multiplet	
2	Alanine	1.46	1	7.14	3.25	1.485	n	1	
3	Alanine	3.76	3	7.2	0.75	3.787	n	1	
4	D-Glucose	3.233	1,1	9.220,8.067	1.0196	3.25	0.001	1	
5	D-Glucose	3.524	1,1	9.823,3.732	0.7705	3.5394	0.001	1	
6	D-Glucose	3.889	1,1	12.293,2.140	1.6046	3.901	0.001	1	
7	D-Glucose	4.634	1	7.957	1	n	n	0	
8	D-Glucose	5.223	1	3.677	0.6049	5.2377	0.001	1	
9	Nicotinic acid	7.51	1,1,1	7.91,4.99,0.72	1.2474	7.529	0.001	1	
10	Nicotinic acid	8.24	1,2	7.92,1.88	1.0295	8.2584	n	1	
11	Nicotinic acid	8.6	1,1	4.96,1.54	0.9924	8.615	0.001	1	
12	Nicotinic acid	8.93	1,1	2.15,0.72	0.7307	8.944	0.001	1	
13	Fumaric acid	6.51	0		2	6.522	n	1	
14	L-Glutamine	3.766	2	6.18	1	3.781	n	1	
15	L-Glutamine	2.148	-2	2.09,2.206	2	2.149	0.001	1	
16	L-Glutamine	2.451	-2	2.38,2.522	2	2.4526	0.004	1	
17	Citric acid	2.676	-1	0,-15	0.7967,1.1799	2.678	0.003	1	
18	Citric acid	2.558	-1	0,-15	1.2242,0.7992	2.558	0.004	1	
19									
20									

Plot all 6 spectra in a single stack plot use the same command as in Test 2:

```
plotBatmanFitStack(bm,offset = 0.8,placeLegend = 'topleft',yto = 5)
```

A stack plot of the fit result is shown below: (pdf file with the name "specFit_stack_.pdf" can be found in folder "*/user/specified/directory/*runBATMAN/BatmanOutput*/date_time"*)



It can be observed that the fit for some metabolites is improved (particularly nicotinic acid resonances around 8ppm).

If you are able to generate the above results, you have correctly and successfully installed BATMAN!

Test 4: Multiple spectra from bacterial supernatants data

This test uses 31 spectra from bacterial supernatants. To reduce run time we fit a small ppm range corresponding to resonances from the metabolites glycine, valine, threonine and glucose. It will take about 1 minute per spectrum. The total run time depends on the number of cores (parallel processes) batman can use. The chemical shift sorting information is also provided with this test, details of how to get this information are explained in section 3: **Chemical Shift Sorting**.

The test dataset and input files for batman are zipped in folder “Test4.zip”, which can be downloaded from

https://r-forge.r-project.org/scm/viewvc.php/documentation%20and%20test/?root=batman

Unzip this file, then:

1. Backup the 4 files: batmanOptions.txt, chemShiftPerSpec.csv, metabolitesList.csv, and multi_data_user.csv in folder “*/user/specified/directory*/runBATMAN/BatmanInput”, for example by giving them another name. (chemShiftPerSpec.csv may not exist if you haven’t run this option before).
2. Move all the 5 files from the unzipped folder “Test4” (include: batmanOptions.txt, chemShiftPerSpec.csv, metabolitesList.csv, multi_data_user.csv, NMRdata_Test4.txt) into the input folder “*/user/specified/directory*/runBATMAN/BatmanInput”.

The user should check if the input for paraProc parameter in “batmanOptions.txt” is appropriate for your machine, which is set to 7 in the default file. In this case the 31 spectra take around 4 minutes to process.

```
paraProc - No of parallel processes (multicores) (only 1 core will be used for single spectrum): 7
```

Now run the same command used in previous tests with two additional input arguments:

```
bm<-batman(runBATMANDir = '/user/specified/directory', txtFile =
'/user/specified/directory/runBATMAN/BatmanInput/NMRdata_Test4.txt', figBatmanFit =
FALSE)
```

Or

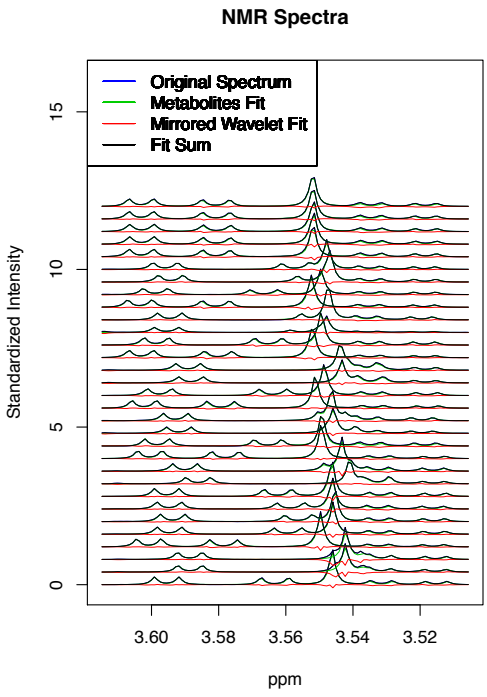
```
bm<-batman(txtFile = './runBATMAN/BatmanInput/NMRdata_Test4.txt',figBatmanFit = FALSE)
```

The user can leave the setting of "`figBatmanFit`" to its default (`TRUE`) by removing that argument from input, in which case 16 figures will be produced when batman fitting finishes.

Plot all 31 spectra in a single stack plot as follows:

```
plotBatmanFitStack(bm, offset = 0.4, yto = 15, overwrite = T, placeLegend = 'topleft',
orientation = "P")
```

A stack plot of the fit result is shown below: (pdf file with the name "specFit_stack_.pdf" can be found in folder "*/user/specified/directory/*runBATMAN/BatmanOutput*/date_time*")



3. Chemical Shift Sorting

The "SplineFitBATMAN" is a MATLAB tool developed to work with BATMAN. The idea is to plot a stack of spectra, where the order is controlled by the chemical shift of a shifting peak. This often

shows a coherent shift across many peaks in the spectra allowing their positions to be estimated more clearly. Then a spline fitting tool allows the user to quickly interactively specify the position of each multiplet across a large number of spectra. This gives BATMAN accurate prior knowledge on the position of each resonance in each spectrum and can be extremely beneficial with peaks that shift a lot. Note that, when BATMAN is used with this information, peaks are still allowed to shift, according to the parameter rdelta (which can be controlled for individual multiplets in mulit_data.csv). The tool gets input from "chemShiftPerSpec.csv" and, "NMRdata.txt" (generated by BATMAN). The codes can be found in in SplineFitBATMAN_MATLABcode folder from:

https://r-forge.r-project.org/scm/viewvc.php/SplineFitBATMAN_MATLABcode/?root=batman

The "csFlag" in batman options file should be set to "1" in order to use this tool.

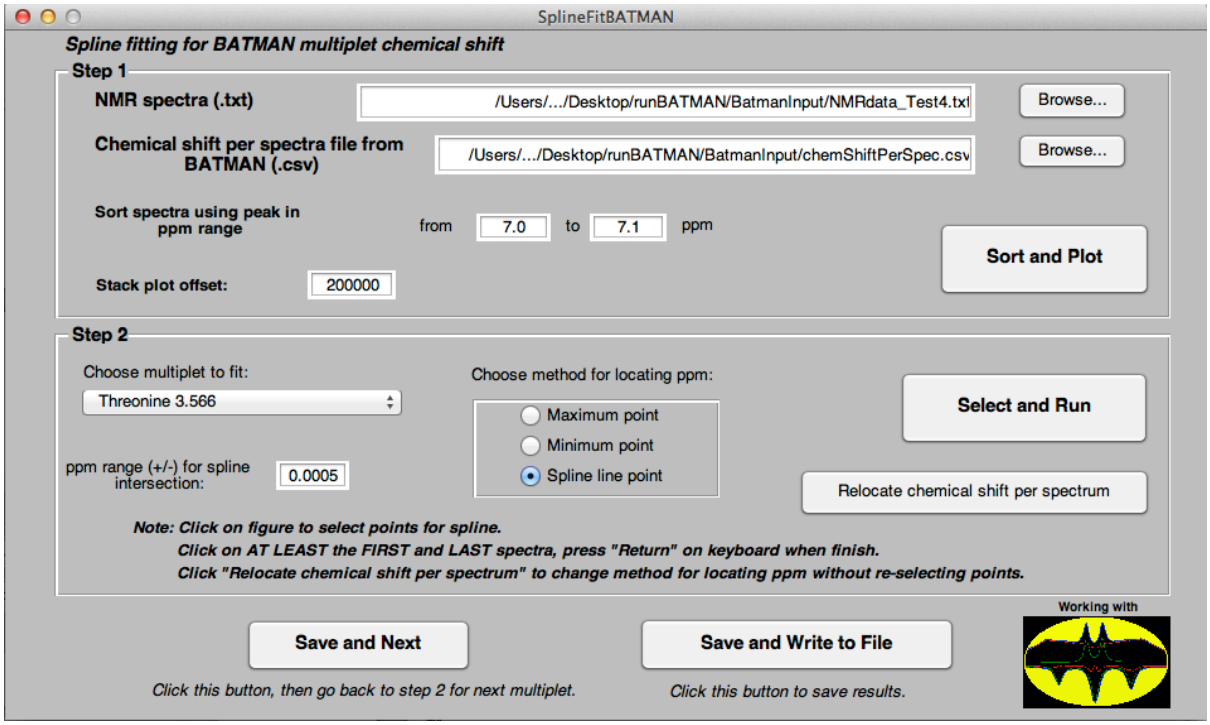
```
csFlag - Specify chemical shift for each multiplet in each spectrum? (chemShiftperSpectra.csv file) (Yes - 1 / No - 0): 1
```

At this stage, you should have finished **Test 4,** which uses the provided file "chemShiftPerSpec.csv". Rename that file to a different name, and run batman again. The following error message (in red) will show.

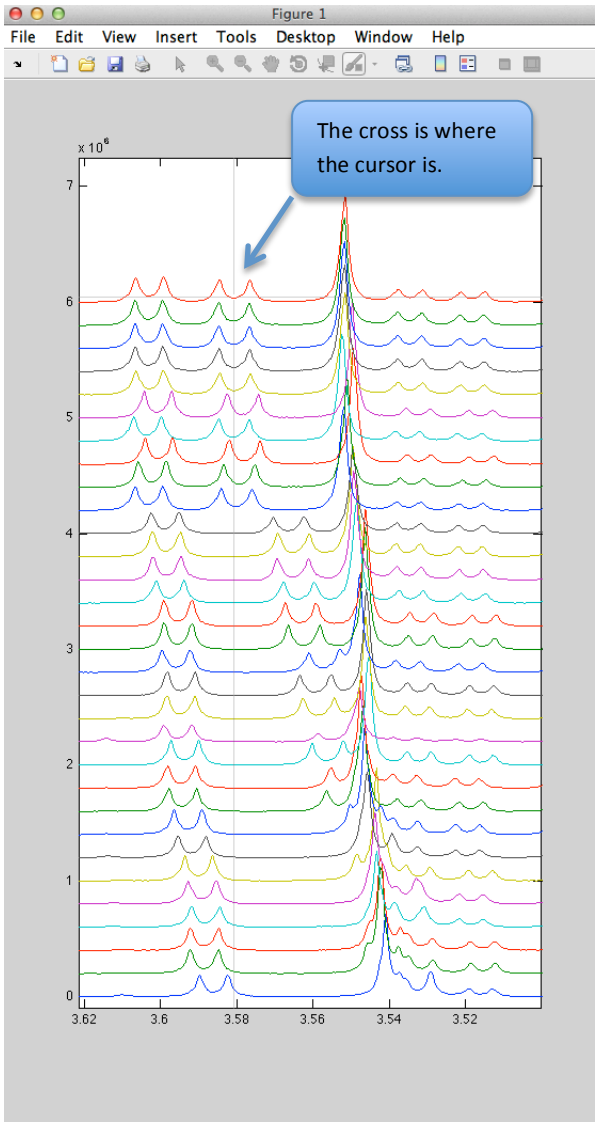
```
> bm<-batman(txtFile = './runBATMAN/BatmanInput/NMRdata_Test4.txt',figBatmanFit = FALSE)

Running batman...
Number of burn-in iterations:  6000
Number of post-burn-in iterations:  1000

The template file used is
 2: The user input template of multiplets in multi_data_user.csv file.

Loading multi_data_user.csv...
Copying multiplet list from multi_data_user.csv to chemShiftPerSpec.csv...
Error in batman(txtFile = "./runBATMAN/BatmanInput/NMRdata_Test4.txt",  :
  No chemShiftPerSpec.csv file found in BatmanInput folder.
     Creating one now, please modify the values.
```

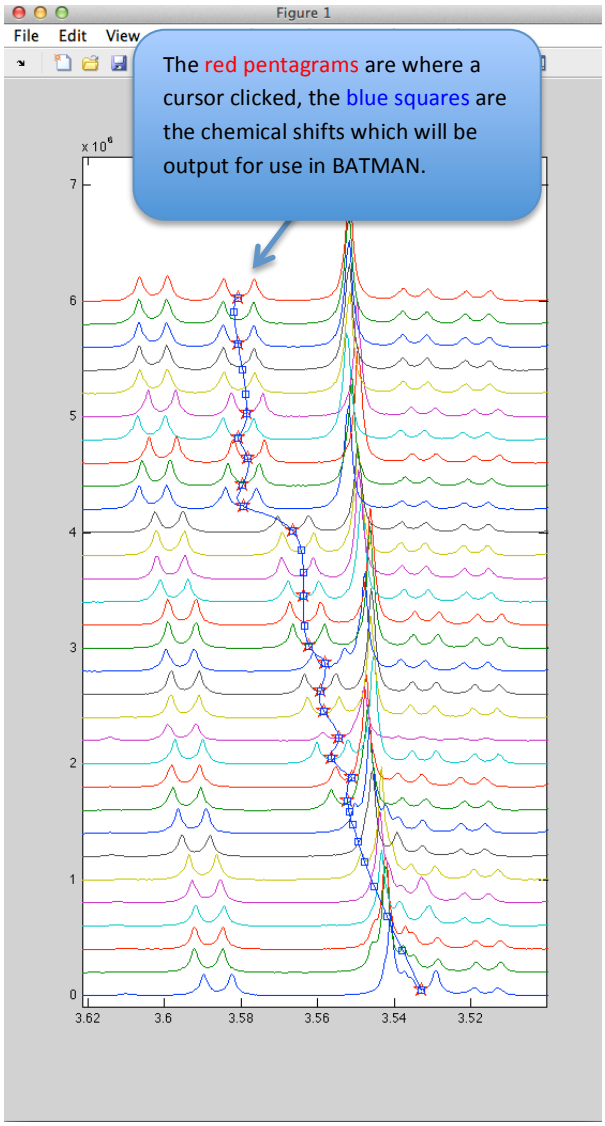
A new "chemShiftPerSpec.csv" file should be created in batman input folder. Keep this file closed when using "SplineFitBATMAN" to write in it.

Run the "SplineFitBATMAN" tool in Matlab, the following GUI should appear.


SplineFitBATMAN
Spline fitting for BATMAN multiplet chemical shift
Step 1
NMR spectra (.txt)
/Users/.../Desktop/runBATMAN/BatmanInput/NMRdata_Test4.txt
Browse...
Chemical shift per spectra file from BATMAN (.csv)
/Users/.../Desktop/runBATMAN/BatmanInput/chemShiftPerSpec.csv
Browse...
Sort spectra using peak in ppm range
from
7.0
to
7.1
ppm
Sort and Plot
Stack plot offset:
200000
Step 2
Choose multiplet to fit:
Threonine 3.566
Choose method for locating ppm:
Maximum point
Minimum point
Spline line point
Select and Run
ppm range (+/-) for spline intersection:
0.0005
Relocate chemical shift per spectrum
Note: Click on figure to select points for spline.
Click on AT LEAST the FIRST and LAST spectra, press "Return" on keyboard when finish.
Click "Relocate chemical shift per spectrum" to change method for locating ppm without re-selecting points.
Working with
Save and Next
Click this button, then go back to step 2 for next multiplet.
Save and Write to File
Click this button to save results.

The figures below show how the tool can be used to fit a spline curve through the sorted stack plot to estimate the positions of shifting multiplets.





Step 1:

1.1 The input for NMR spectra file should be .txt file generated by BATMAN. Calling "batman" once in R should generate NMRdata.txt in ../runBATMAN/BatmanInput folder which can be used here.

1.2 The input for chemical shift per spectra file, chemShiftPerSpec.csv can also be found in folder ../runBATMAN/BatmanInput. Keep this file closed when writing to file.

1.3 "**Sort spectra using peak in ppm range**": choose a peak to sort on, suggest an isolated but shifting peak. In this example (test 4 data) we suggest sorting on the histidine peak in the region 7.0-7.1ppm.

1.4 "**Stack plot offset**": sets the offset for stack plotting NMR spectra. Adjust the offset so that the multiplets of interest do not intersect with each other between spectra. This parameter can be modified for different multiplets.

1.5 Click "**Sort and Plot**" to plot the stack of spectra sorted according to chemical shift of the selected peak. Repeat steps 1.3-1.5 until satisfied with the stack plot.

Step 2:

2.1 Choose a multiplet from the drop down menu, and then zoom in to the range containing that multiplet on the figure. This does not have to be the multiplet used in step 1.3 above.
2.2 The algorithm can output ppm positions found by
a) simply fitting a spline between points clicked on the figure,
b) finding the maximum intensity in the vicinity of the spline position (e.g. to find position of a singlet or triplet) or
c) finding the minimum intensity in the vicinity of the spline position (e.g. to find position of a doublet).
The parameter "**ppm range for spline intersection**" sets the search range on the left and right (+/-) of the intersection between spline and each spectrum.
2.3 Click "**Select and Run**".
2.4 Click on the points at which you wish to place the spline. You do not need to click on every spectrum in the stack. For example, a peak, which shifts smoothly across the stack, may only require 2 or 3 clicks to define the spline. Make sure to click on at least the FIRST and LAST spectra in stack plot each time. At the end, press '**return**' on the keyboard and the tool will draw the spline and the chemical shift positions to be output for use in BATMAN (blue squares).
2.5 If you are not satisfied with results, try changing the parameters including "**ppm range for spline intersection**" and "**method for locating ppm**", and then click "**Relocate chemical shift per spectrum**" (no need to re-select points), or "**Select and Run**" to re-select points.
2.6 If satisfied with estimated positions, click "**Save and Next**", to save and go to next multiplet.
2.7 When finished, click "**Save and Write to File**" to save results to file chemShiftPerSpec.csv.

Note that you do not need to do this for all multiplets or all metabolites. If no spectrum-specific chemical shifts are available for a given multiplet, BATMAN will use the default parameters specified in multi_data.csv (or multi_data_user.csv, as selected by the user).

Troubleshooting:

If not satisfied with results try:

- checking if the FIRST and LAST spectra were clicked, or
- choosing different method for locating ppm, or
- clicking on figure at more accurate points, or
- clicking more spectra on figure.

Sometimes MATLAB does not register all clicks on the plot. Try leaving more time between clicks.

If satisfied with fitting results, return to **Test 4** to use the newly estimated chemical shifts in batman.

References:

1. Viant, M.R., et al., *International NMR-based environmental metabolomics intercomparison exercise.* Environ Sci Technol, 2009. **43**(1): p. 219-25.
2. Behrends, V., et al., *Metabolite profiling to characterize disease-related bacteria: gluconate excretion by Pseudomonas aeruginosa mutants and clinical isolates from cystic fibrosis patients.* J Biol Chem, 2013. **288**(21): p. 15098-109.