BRASS version 3

Citation for published version:
Millar, A, Brown, PE & Rand, DA, BRASS version 3: Biological Rhythm Analysis Software Suite, 2010, Software.

Link:
[Link to publication record in Edinburgh Research Explorer](#)

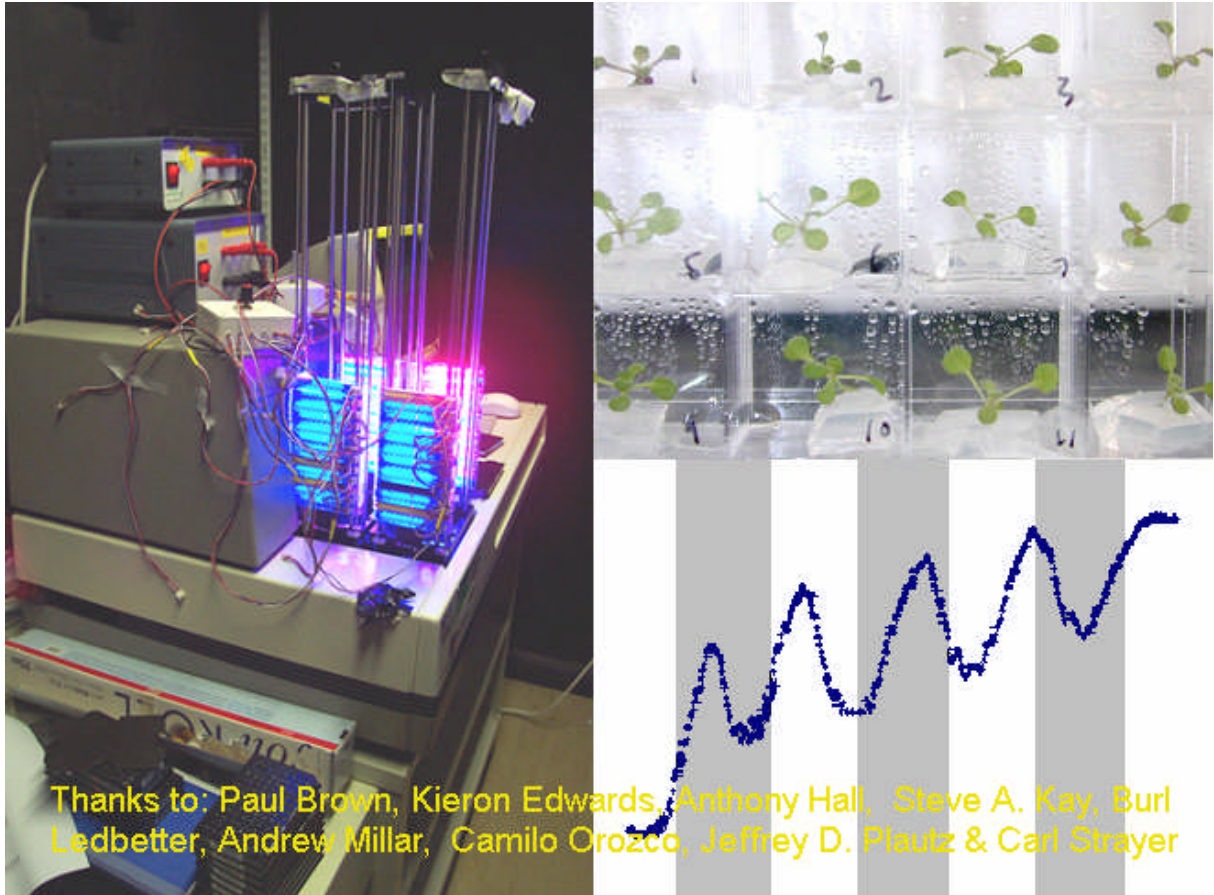**Publisher Rights Statement:**
Copyright Andrew J. Millar, 2010

Biological Rhythms Analysis Software System

Version 3.0

April 2007

By

Paul Brown

Contents

Acknowledgements

Introduction

What is BRASS?

Biological Rhythms Analysis Software System, or BRASS, is a Microsoft Excel workbook which provides a user interface for importing, sorting and analyzing rhythmic time series data and presenting the results. Experimental data can be imported from as wide a variety of sources as possible.

In summary, to completely analyse a set of experiments a user would:

1. Import raw experimental data into BRASS from one or more experiments, producing one Excel workbook for each experiment. Each of these will be in a common format, whatever the source of the original data.

2. Combine the workbooks if data from more than one experiment should be analysed as a single block.

3. Assign genotypes to each data series and de-randomize the data accordingly.

4. Perform Fourier analysis to find oscillations in the data.

5. Optionally proceed to peak time (phase) analysis.

6. A single results workbook is produced at the end containing a range of statistics on period, amplitude, phase, and shape of the waveform for each genotype.

7. This workbook contains functions to allow the automated generation of different types of plots of this data.

BRASS also contains additional functions to:

- Normalise data

- Detrend data

- Produce averages of selected data series

- Plot selected data series

- Plot mean data series for genotypes

Installation Issues

Brass version 3.0 contains two alternative algorithms for performing Fourier analysis. FFT NLLS is a suite of programs which must be obtained separately as described in Appendix F. Mfourfit is an Exel add-in developed by Paul Brown, Alex Morton and David Rand at the Mathematics Institute, University of Warwick and comes supplied with Brass. Please make note of the following points when installing these algorithms

- **If you are using the FFT NLLS set of programs the path from the root directory to BRASS should not contain any spaces**. Spaces cause problems when passing file paths to the command prompt when executing the FFT NLLS set of programs. It is probably best to create a new folder, such as "C:\BRASS"

- **The FFT NLLS set of programs, `four-anl.exe, fls-plot.exe,` and `condense.exe` must be installed in the same folder as BRASS.** See Appendix F for an explanation of these programs and how to obtain them.

- **To install Mfourfit you must have the Matlab run time libraries installed on your computer.** To do this you should download MCRInstaller.exe from the software download page at www.amillar.org. Run this program and follow the onscreen instructions to install the required libraries. However, if you have Matlab 7.1 or later installed on your PC then you should not need to do this.

- **To activate Mfourfit place Brass_Utilities_21.exe in the same folder as Brass and run this program by double clicking on it.** You should get a message looking something like
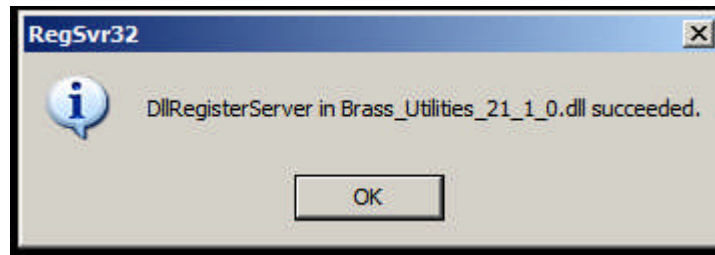
Two new files will have been created, Brass_Utilities_21.ctf and Brass_Utilities_21_1_0.dll. Click the button on the Visual Basic toolbar in order to open the Visual Basic editor. (Show this toolbar if it is not visible by going to **View->ToolBars**). Go to the **Tools->References** menu item to open the References dialog box.
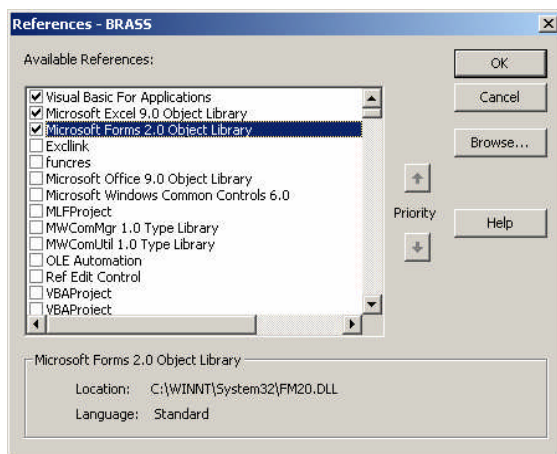
A reference called Brass_Utilities_21 1.0 Type Library should appear in the list. Check this reference if it is not already checked.

Other Issues

- BRASS was developed using mainly Excel 2000 and Windows 2000, but there should not be any problems when using other combinations. We have tested the following additional combinations and found no problems:

  - Windows 98 and Excel 97
  - Windows 2000 and Excel 97
  - Windows XP and Excel 2000
  - Windows XP and Excel XP

- It is necessary to ensure that all the required Visual Basic libraries are available to the

workbook in order for BRASS to function correctly. This should always be the case automatically, but if any error occurs which is caused by a missing library, such as an unknown function or data type, open the references dialog box as described above. The Visual Basic, Excel Object Library, and the Office Object Library are all required and should be checked automatically. Any other checked files marked as MISSING may cause problems. Unchecking them usually solves this.

- Once installed, BRASS should be made read only. This prevents accidental changes to the code. If BRASS is saved under another name all macros should still work.

- Please read the section on Limitations, on page 73, before beginning to use BRASS.

Using BRASS

Most BRASS functions are accessed via the menu item of the same name, which should appear on loading the workbook.

Importing Data

Experimental data is imported into an Excel spreadsheet, which is organised so that readings for each plant / leaf appear in columns. Column headings provide information on raw data file name, location in the experimental set up (region / well name), genotype, and plate number, as shown below. The key heading is region, which must be present. All others are optional. ZT values are displayed as hours with decimal minutes. Worksheets in this book are named after the corresponding worksheets in the raw data book.

The **Import** menu item reveals a sub-menu containing different sources of input data. Each of these represents a source workbook with a different layout. See Appendix A for examples of the different layouts that BRASS will process.
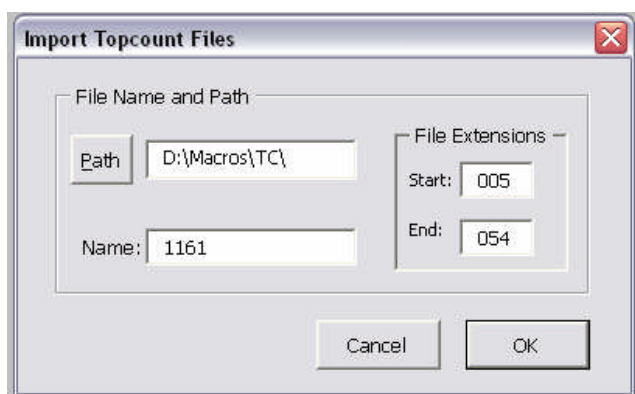
Should any time points be missing from a data series, as is quite common in leaf movement experiments, then the value "-1" is added to the cell. Having an unbroken column of data makes many of the routines much simpler, as it is always clear exactly where the data ends. However, when exporting to FFT NLLS, these values are temporarily removed to prevent then from interfering with the analysis.

After importing data, the user is asked to save the newly created workbook, and then asked if they wish to automatically proceed to assigning genotypes to the data (page 23).

TopCount Data

The **TopCount Data** menu item contains 2 sub-menu items

- **Separate Files** imports a series of Topcount files from one plate, with each representing 96 well readings from one timepoint. It is essential that these files conform to the format show in Appendix A. Import begins by displaying the form below

Clicking Path will allow the user to select one of the required files. This will result in the folder and base file name appearing the appropriate boxes. Note that the base file name should have no more than 8 characters. The file extension boxes indicate which files should be imported. BRASS fills these automatically, depending on which are present, but the user can edit these values, and it does not matter if some are missing. Next, the user is asked to indicate in which order they wish the data series to appear in the imported workbook. The choices are

A1 B1 C1…..F12 G12 H12, or

A1 A2 A3…...H10 H11 H12

The user may have a preference, depending on how different genotypes were arranged in the experiment. The user is then asked to specify at what time dawn (ZT0) occurred on the 24 hour clock. This will be used to convert the time values in hh:mm:ss to ZT time, in hours with decimal minutes. Note that for each file, the time value taken is that for well A1 in all cases. Column headers show the name of the original files, and the plate number, which are assumed to be the same. Genotypes must be added by the user (see later).

Note that as each row on the Import Worksheet consists of the contents of one file, file extensions appear not in the column headings, but in column "C", adjacent to the ZT values.

Users who are used to using the "Packard Import" functions of "I&A" for importing Topcount data in a more general format should refer to the section below on "Custom Import".

- **Combined Files** imports a single file that has been produced in MS-DOS by joining together a series of files, each representing a single timepoint. The format must match that specified in Appendix A. As above, the user is asked to indicate in which order they wish the data series to appear in the imported workbook. The user is then asked to specify at what time dawn (ZT0) occurred on the 24 hour clock. This will be used to convert the time values in hh:mm:ss to ZT time, in hours with decimal minutes. Note that for each file, the time value taken is that for well A1 in all cases. Column headers show the name of the original file, and the plate number, which are assumed to be the same. Genotypes must be added by the user (see later).

Night Owl Data

The **Night Owl Data** menu item imports a single worksheet produced by the Night Owl system, which must be in the format shown in Appendix A. The routines used by "I&A" are used unaltered. The user is asked to specify the starting ZT time, and the sample time interval, both in seconds. Wells are copied over, but genotypes must now be filled in the user (see **Exporting Data** section). Plate number does not apply here.

Note that the worksheet containing the data should be the first in the workbook.

Kujata Y Files

The **Kujata Y Files** menu item imports data from *.y** files, each consisting of a complete set of readings for one leaf. Y files are always named after the camera number (1 to 32) which produced them, with the extension signifying leaf number (1 to 24), ie Cam_1.y01 to Cam_32.y24. Firstly, the form below is displayed, which detects files from which cameras are present in the specified directory and asks the user to select one or more of the available cameras.
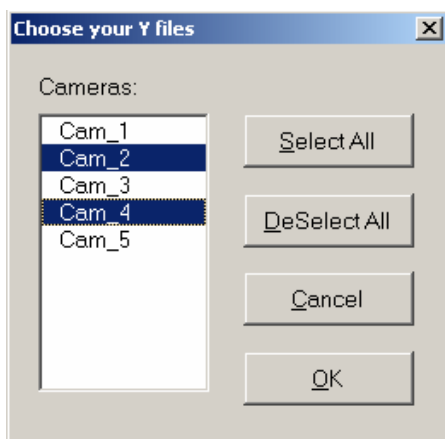
Files with extensions y01 to y24 are then searched for, for each specified camera, and any found are imported. Each camera is assigned its own worksheet in the imported workbook, with sheets being named after the camera. The user is then asked to specify the ZT time at the start of the experiment, in hours and minutes. This is used as an offset and is added to the times read from the y files. The actual times written to each worksheet are those from the y file with the lowest extension number found for that camera. The column headers of the import sheets now show the name of the original y file for each data column (leaf), and the region name, which will be the y file extension number. Genotypes must be added by the user (see later). Plate numbers do not apply here.

Metamorph Data

The **Metamorph Data** menu item contains two sub-menu items

- **Luminescence** imports a worksheet produced by a metamorph luminescence experiment. All the data is expected in single block, one column for each region, with the first region being found in column 5, the region name in row 1 and the data beginning in row 2. The name of the experiment is expected in column 1. Image times

are expected to be in column 3. These are time of day in fractions of a day (0 to 1), although the cells may be formatted to display the values in hh:mm:s format. These are changed to time elapsed since start of experiment, in hours. This format is shown in Appendix A. The user is asked to specify the time of dawn, in hours and minutes. This value is subtracted from each time in order to convert them into ZT times. Data is imported to a worksheet named after the Metamorph workbook. The column headers of the import sheets now show the name of the original Metamorph file, and the region name. Genotypes must be added by the user (see later). Plate numbers do not apply here.

Note that the worksheet containing the data should be the first in the workbook.

- **Movement** imports a worksheet produced by a metamorph leaf movement experiment. This worksheet is expected to have the following format:

Column 1   Image Name
Column 2   Image Plane
Column 3   Image Date and Time, time of day in fractions of a day (0 to 1), although the cells may be formatted to display the values in hh:mm:s format.
Column 4   Region name
Column 5   Centroid X
Column 6   Centroid Y

Headings should be found in row 1, with the data beginning in row2. This format is shown in Appendix A.

Time values are changed to time elapsed since start of experiment, in hours. The user is asked to specify the time of dawn, in hours and minutes. This value is subtracted from each time in order to convert them into ZT times. The X and Y values for each region are then plotted and the user is asked to choose which, if any, of the two to import, as shown below.
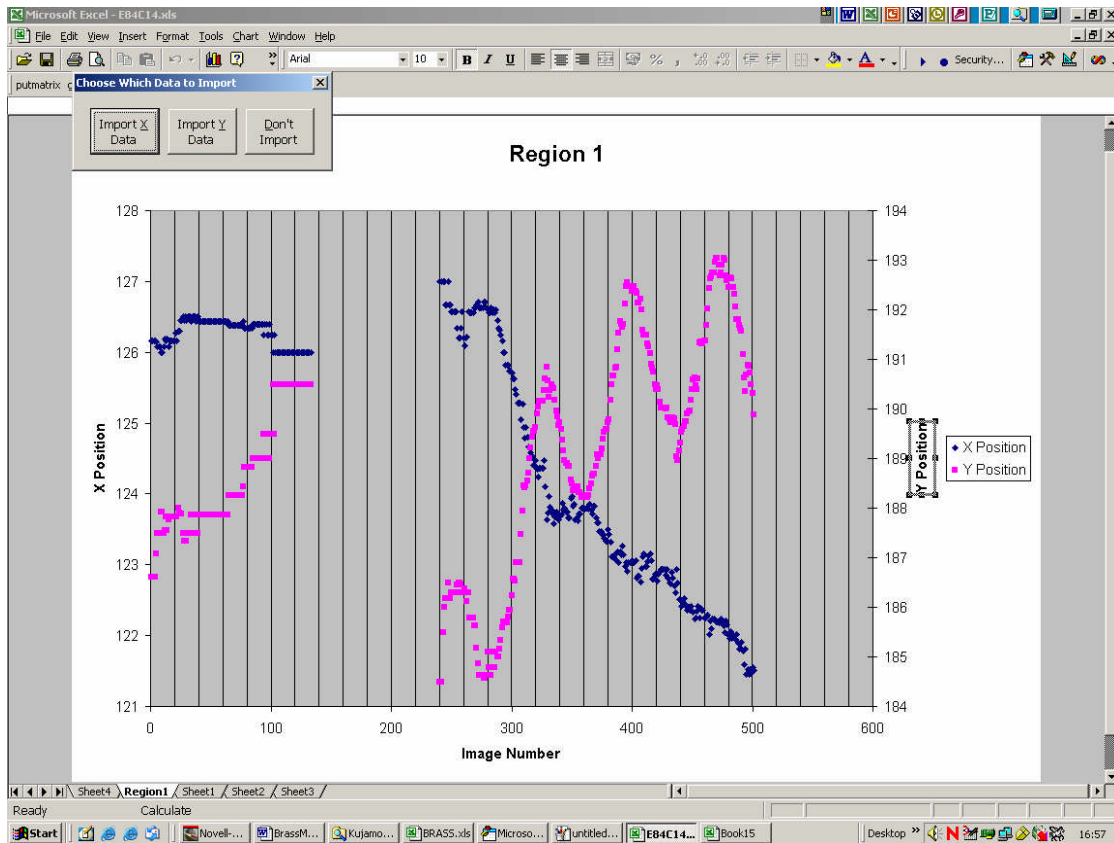
Data is imported into a worksheet named after the Metamorph workbook. The column headers show the name of the original Metamorph file, and the region name, with an X or Y appended. Genotypes must be added by the user (see later). Plate numbers do not apply here. Image numbers from the Metamorph file are added next to their corresponding ZT values in column three of the import worksheet.

Note that the worksheet containing the data should be the first in the workbook.

Custom Import

The **Custom Import** menu item contains two sub-menu items
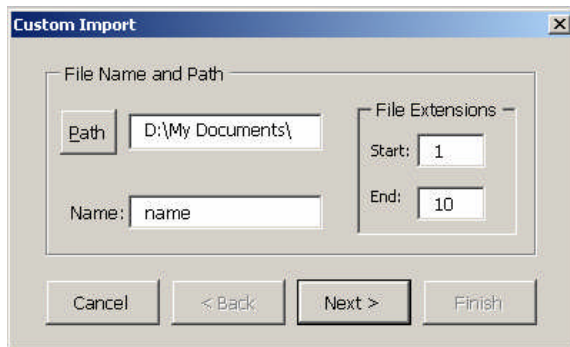
- **Multiple Files** imports multiple delimited text files, each representing all the data for one region, or all the data for one time point, and combines them onto a single worksheet. It is based on the "Packard Import" functions of "I&A", so each file must have the same name, and a numeric extension. Before beginning, the user should first

open an example file in Excel and note the following things

- o The character(s) which act as delimiters between data items. These can be determined by comparing the file open in Excel with the file opened in a text editor.
- o The column/row number where the data is located
- o The column, row, or individual cell where any region name(s) or time values(s) are located.
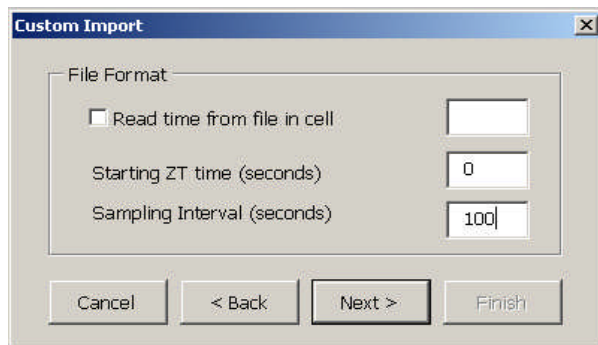
As in "I&A", the first thing the user must do is select the directory where the files are located, enter the file name, and the range of extension values to import. This is done using the form below.

Next, the user must indicate whether each file consists of all the timepoints for a single plant/region, or a single time value for a number of plants/regions. Then, whether the data series appears in a row or a column must be specified.

Importing data for a single timepoint. Each file represents all the data from one sample time, so the user must first specify the time value for each file using the form below.

If a value is present in each file then checking the box and entering the cell reference in

16

the form "A1", for example, will result in these values being read and transferred to the import workbook. Alternatively, unchecking the box and entering values, in seconds for the starting ZT and the sample interval will cause BRASS to calculate ZT values instead. This is done by taking the file with the lowest extension number found within the specified range as being equal to the Starting ZT time. Subsequent files are then assigned a value equal the starting value plus Sampling Interval multiplied by their extension number minus the lowest number. Thus, if a file (timepoint) is missing, the next file will still have the same time values assigned as it would have had if the file were present. This is the same as in "I&A".

If the user elected to read timepoint from file, it is now necessary to specify the format that is to be read. This is so that BRASS can correctly convert the values to ZT values. The options are:

- Time of day in fractions of a day, ie a value between 0 and < 1. These are converted to time since start of experiment in hours. The user is asked to enter the time of dawn. This is subtracted from each value to convert them into ZT values.
- Time of day in hours, a value between 0 and < 24. These are converted to time since start of experiment in hours. The user is asked to enter the time of dawn. This is subtracted from each value to convert them into ZT values.
- Time since start in seconds. These are converted to hours. The user is asked to enter the ZT time at the start of the experiment, in hours and minutes. This is added to each value to produce correct ZT times.
- Time since start in minutes. These are converted to hours. The user is asked to enter the ZT time at the start of the experiment, in hours and minutes. This is added to each value to produce correct ZT times.
- Time since start in hours. The user is asked to enter the ZT time at the start of the experiment, in hours and minutes. This is added to each value to produce correct ZT times.
- Time since start in days. This assumes that hours and minutes are represented as decimal fractions of a day, eg 36 hours would equal 1.5 days. These are converted to hours. The user is asked to enter the ZT time at the start of the experiment, in hours and minutes. This is added to each value to produce correct ZT times.

The user is now asked to enter how the region names are to be created for each data value. The options, show below, are to read values from a particular row/column, or to create names automatically if they are not present in the file.
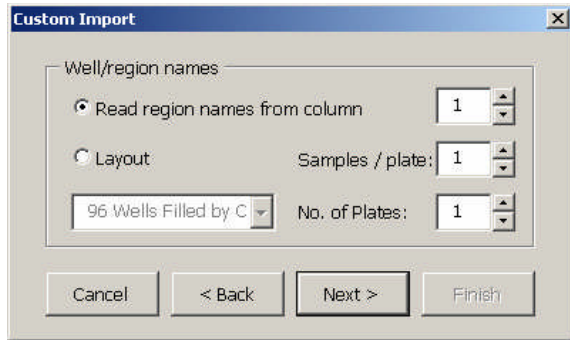
If regions are read from file, then they are located in the first data file only, in the specified row by starting at the column number at which the data row begins (described below), or the specified column at the row number at which the data column begins. As in "I&A", the user can chose to split the regions across several worksheets. BRASS will add worksheets to the Import Workbook called "Plate1", "Plate2" etc, up to the plate number specified above. Region names will then be read in blocks of the size specified in Samples / plate, starting with the top/leftmost block. The first block is copied to "Plate1", and so on.

Alternatively, checking the Layout button allows the user to elect to generate region names using one of the schemes from the list box. These are designed for Topcount plates and are:

- 96 Wells filled by column. This produces the layout

    A1 B1 C1 D1 E1 F1 G1 H1 A2 B2 C2 D2 E2 etc..

- 96 Wells filled by row. This produces the layout

    A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 B1 B2 B3 B4 etc…

- 48 Wells, skip wells produces

    A1 C1 E1 G1 B2 D2 F2 H2 A3 C3 E3 G3 B4 D4 F4 etc…

In this case, sample number is defined so the box is disabled, but the user can still enter plate number for multiple worksheets.

Note that the user is prevented from entering impossible sample sizes, eg if data is in rows then sample number times plate number cannot exceed 256, as Excel does not allow more than 256 columns in a worksheet so this is the maximum amount of data which could be present.

Finally, the form below asks the user to specify the row/column in which the actual data values are to be found, and one or more characters which are to be considered delimiters between individual data items.
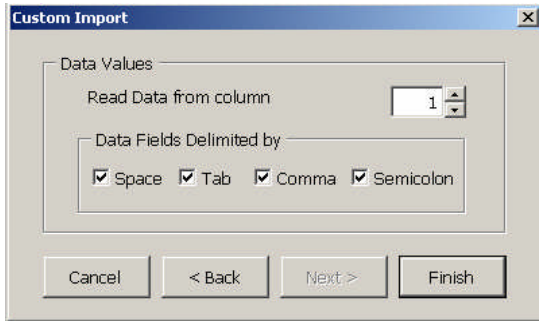
In the case of columns, data is found by starting at row one, and then moving down until the first non-empty cell is found. This is then considered to represent the start of the data. In the case of rows, data is found by starting in the specified row at column one, and moving right until the first empty cell is found. Data can be split into blocks across several worksheets as described above for regions.

Note that as each row on the Import Worksheet consists of the contents of one file, file names appear not in the column headings, but in column "C", adjacent to the ZT values.

Importing data for a single region. Each file represents a complete time series for a single region/plant. Firstly, the user must specify how the region name for each file is to be obtained. Names can be read from a specified cell in each file, or created by simply taking the extension number of the file.
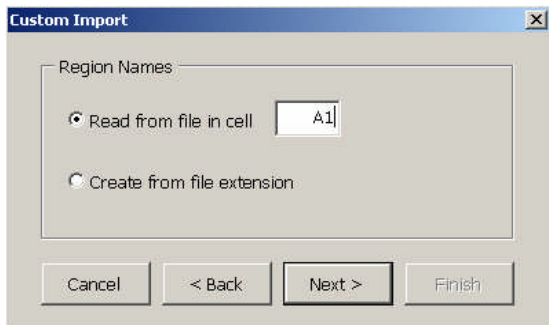
Now the user must specify how the time values are to be obtained.
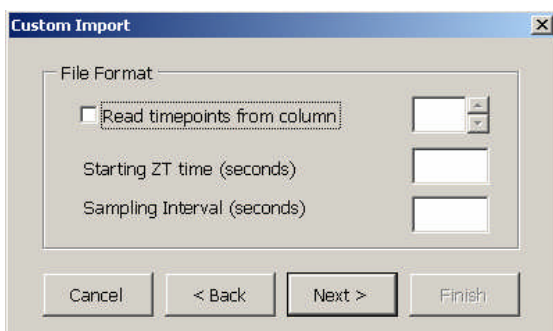
The series can be read from a specified row/column in the first data file. They are located in the specified row by starting at the column number at which the data row begins (described below), or the specified column at the row number at which the data column begins. Alternatively, the user can enter values, in seconds for the starting ZT and the sample interval, which will cause BRASS to calculate ZT values for each data point instead.

If the user elected to read timepoint from file, it is now necessary to specify the format that is to be read. This is so that BRASS can correctly convert the values in ZT values. The options are as described in the previous section.

Finally, the user must specify the row/column in which the data readings are to be found, and one or more characters which are to be considered delimiters between individual data items.
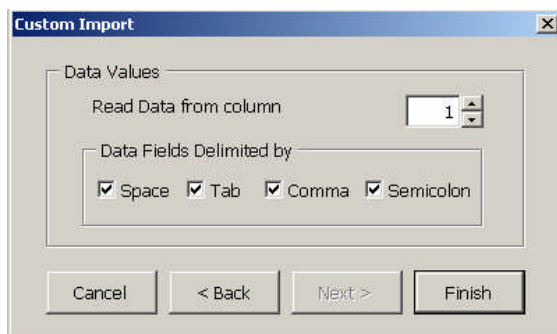
In the case of columns, data is found by starting at row one, and then moving down until the first non-empty cell is found. This is then considered to represent the start of the data. In the case of rows, data is found by starting in the specified row at column one, and moving right until the first empty cell is found.

- **Single Worksheet** imports a single worksheet, which can take any format as long as all the data is in one continuous block. Also, the sheet to be imported must be the first in a workbook. The user is first asked to specify the top left cell of this block, and whether the data series occur in rows or columns. This will determine whether the block must be transposed or not when it is imported.
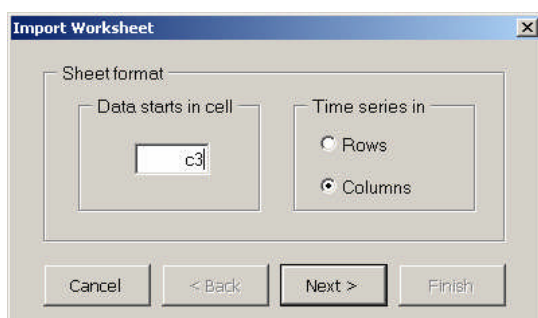
The user must now specify whether any region names or time values can be read from the workbook.
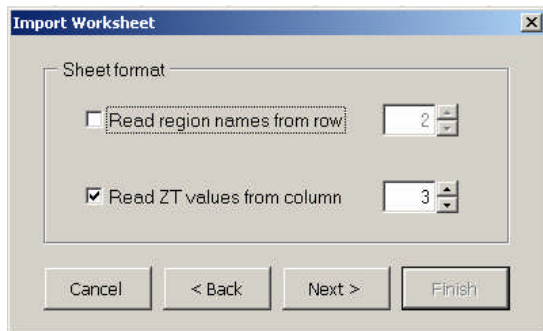
If no region names are read then they must be created, as it is essential that every data set has a region name. In this case they are simply numbered from left to right, starting at one. If no time values are read then they must also be created. The user is asked to specify a starting ZT time, and a sampling interval, both in seconds. Otherwise, the user is now asked to specify the format of the time values which will be read from the workbook. The options consist of:

- Time of day in fractions of a day, ie a value between 0 and < 1. These are converted to time since start of experiment in hours. The user is asked to enter the time of dawn. This is subtracted from each value to convert them into ZT values.
- Time of day in hours, a value between 0 and < 24. These are converted to time since start of experiment in hours. The user is asked to enter the time of dawn. This is subtracted from each value to convert them into ZT values.
- Time since start in seconds. These are converted to hours. The user is asked to enter the ZT time at the start of the experiment, in hours and minutes. This is added to each value to produce correct ZT times.
- Time since start in minutes. These are converted to hours. The user is asked to enter the ZT time at the start of the experiment, in hours and minutes. This is added to each value to produce correct ZT times.
- Time since start in hours. The user is asked to enter the ZT time at the start of the experiment, in hours and minutes. This is added to each value to produce correct ZT times.
- Time since start in days. This assumes that hours and minutes are represented as decimal fractions of a day, eg 36 hours would equal 1.5 days. These are converted to hours. The user is asked to enter the ZT time at the start of the experiment, in hours and minutes. This is added to each value to produce correct ZT times.

Data is imported to a worksheet named after the data workbook. The column headers of the import sheets now show the name of the original workbook, and the region names. Genotypes must be added by the user (see later). Plate numbers do not apply here.

Assigning Genotypes

This can be done automatically after importing data, or by making the imported data sheet you wish to apply genotypes to the active worksheet and selecting the **Analysis->Assign Genotypes** menu item. This loads the form show below.
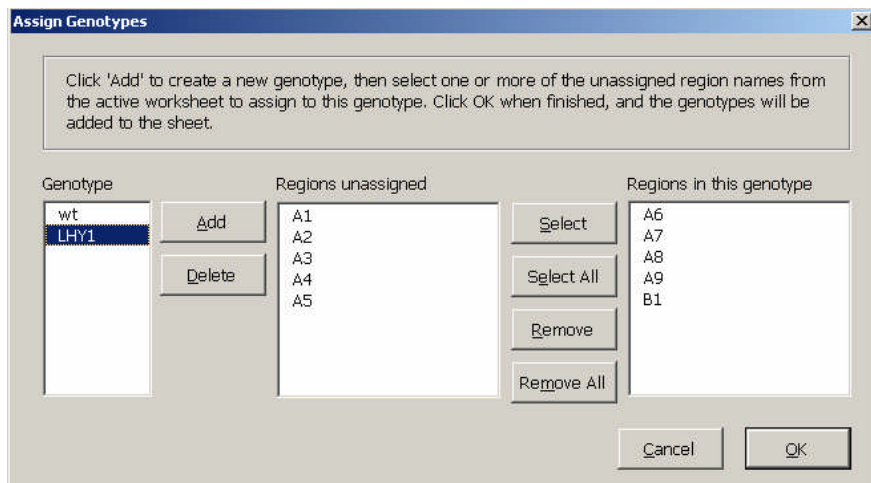
This form reads the region names from row 8 of the active sheet. Any which have a corresponding empty genotype cell appear in the list "Regions unassigned". To create a new genotype click "Add" and enter the name. This will appear in the "Genotype" list. When a particular genotype is selected in this list, any regions which have this name in their corresponding genotype cell will appear in this "Regions in this genotype" list. These can be removed by highlighting them and clicking the "Remove" button. New regions from the unassigned list can be added to the genotype by clicking the "Select" button. A genotype can be removed by selecting it and clicking "Delete". This results in any regions assigned to it appearing in the unassigned list. Clicking "OK" results in the worksheet being updated, with the appropriate genotype being added to each region. If the workbook contains multiple sheets of data, then this process must be repeated for each one.

Note that when region names are moved in and out of the unassigned list, they do not necessarily always appear in their original order, which is the order they are found on the worksheet

Each time "OK" is clicked, the user is asked if they wish to derandomize the workbook. This should only be done when genotypes have been assigned to all the required worksheets.

Note that the workbook is not automatically re-saved after genotypes are assigned.

De-Randomizing Data

This can be done automatically after assigning genotypes, or by making the workbook you wish to de-randomize the active workbook and selecting the **Analysis->DeRandomize workbook** menu item. The de-randomization process examines each sheet in the selected book and creates a new workbook containing a separate sheet for each genotype found on each sheet in the selected book. These sheets are named "Genotype_sheet name". This means that if the same genotype appears on more than one sheet in the selected book, then more than one sheet will be created for this genotype in the de-randomized book. This is in case the ZT times do not match. This is not a problem though as all the sheets for one genotype can be exported to FFT NLLS together, as described in the next section.

After de-randomization, the user is asked to save the newly created workbook, and is then asked if they wish to proceed to export this workbook to FFT NLLS.

Processing Data Series Before Rhythm Analysis

Normalising Data Series

BRASS contains functions to create a new workbook which is a normalised version of an existing imported workbook. The user can select the **Analysis->Normalise Data** menu item, and then select the workbook they wish to normalise. Normalisation works by calculating the average value for each column of data, ignoring any empty cells, or any with a "-1" in them. Each value is then divided by the average value for that column. The user should then save the normalised data under another name to avoid overwriting the original data.

Detrending Data Series

BRASS contains functions to create a new workbook which is a detrended version of an existing imported workbook. The user can select the **Analysis ->Detrend Data** menu item, and then select the workbook they wish to detrend. Now select whether to remove trends in rhythm amplitude, baseline value or both. Click OK and a new workbook will be created. The user should then save the detrended data under another name to avoid overwriting the original data.

Warning: See Appendix F, page 82 for a description of a limitation of the detrending process

Exporting Data for Anlaysis

BRASS will export data to FFT NLLS or Mfourfit automatically and produce results for specified groups of datasets, thus allowing means to be calculated for these groups. Datasets are grouped by creating one or more worksheets containing all the data of interest and no others, and specifying export groups containing one or more worksheets. It is anticipated that user will mainly want to produce means for each genotype. A de-randomized workbook may contain several sheets for the one genotype. These can easily be grouped together and exported as one block.

FFT NLLS

Exporting is done automatically after de-randomizing a workbook, or by selecting the **Analysis->Create Files and Run FFT NLLS** menu item and selecting the required workbook.

Three options for selecting the time windows to be exported are now presented:

1. "All Over Same ZT". If this option is selected then the user will be asked to specify a fixed time window to apply to every plant in the analysis.
2. "Select Best ZT Window for each". If this option is selected then the entire time series will be plotted for each plant and the user will be able to select a range from the graphs individually.
3. "All Possible Time Windows". If this option is selected then user is asked to specify the length of a floating time window, and the increment between windows. Multiple windows are therefore analysed for each plant. BRASS automatically picks out those with the lowest Rel Amp values for each plant.

Whichever of these is selected, export groups must first be defined. The form below is displayed.

This works by the user specifying a group name, usually a genotype, under which to analyse one or more worksheets. This is done by clicking "Add". Then select the genotype from the "Groups" list. The "Worksheets unassigned" list contains the names of all the sheets in the book to be exported, which have not yet been assigned to any group. The "Worksheets in this group" are those which have been assigned to the currently selected group. Worksheets can be added to and removed from the currently selected group using the "Select" and "Remove" buttons. The significance of creating groups is that the periods etc… fitted for all the individual plants in that group, from one or more worksheets, will appear in the same condense file, so allowing BRASS to easily calculate mean values for that group. In this way, if there are multiple sheets for one genotype, then they can all be exported in one block to produce mean values for that genotype.

Very often, however, no genotype will appear on more than one sheet in a derandomized workbook. In this case the "Auto" button allows the user to save time by automatically generating a group name for each worksheet and assigning the worksheet to it. The user can then verify the groups created by clicking on the group names, and edit any as required.

Note that when using this function group names are generated by taking the first 8 characters of each sheet name. It is possible that 8 characters may not uniquely distinguish every worksheet. If this is the case then the user is warned that some worksheets have not been assigned to groups. These remain in the 'Worksheets unassigned' list box, and the user must manually create unique group names for them.

What happens next depends on the selected ZT window option. If a fixed window was selected then the user is simply asked to specify the time window. If the selected start time is before the start of the data then the window will start at the first time point. If all possible time windows are selected then the user must supply details of these using the form below. Using the values shown, the program will start by analysing a time window beginning at time zero, or the earliest ZT time found if values do not start at zero.

The window selected will be 48 hours long. Before analysis the number of blank cells will be counted. These are common in leaf movement experiments when a particular leaf has moved out of the selected area at a particular timepoint, and are signified by containing the value '−1'. If the percentage of cells in the range which are blank is below the maximum permitted, then this range is passed to FFT NLLS. Otherwise, it is not analysed. The next range is then selected by incrementing the start time by the amount specified, in this case 12 hours. This will continue until the start time has advanced so far that the end point of the range exceeds the end of the data. When this happens, then this last range will not be analysed.

Whatever time window is selected, the next stage is to specify the parameters for FFT NLLS using the form below.

The upper and lower limits are the range of periods which `condense` should pick out. "Timeout" is the amount of time that any of the three programs should be allowed to run before BRASS considers that they have crashed, notes this and shuts then down, and then moves on to the next data window. This is useful if the user sets an analysis going and them leaves it to run, as it prevents one bad data set from holding up the entire analysis. As `four-anl` is never run on more than one data file at a time, this can probably be a very short value.

Note that Windows is sometimes unable to shut down `four-anl` after a crash, or to delete the .in file associated with it. This can lead to a number of DOS windows accumulating, but should not interfere with the analysis.

The auto delete check box is useful for keeping the directory tidy, as it removes files which will probably never be used. These are the .sum and .out files produced by every time window, and also the .cnd files when multiple time windows are analysed, as each will have been copied to a master cnd file for that plant (see below). This allows an easy comparison to be made of the effect on period of altering the time window.

Next the user must specify a folder to put the results in, in which new sub-folders will be created, as described below. Finally, the user must enter a confidence limit of statistical certainty for `four-anl` to use. 95% is the value which is commonly used.

The analysis is now ready to begin.

Selecting separate time windows for every plant

Each data set is graphed, and the user must select an appropriate time window, as show below. Remember that any "-1" values have been added to create a continuous data series for graphing and will not be passed to FFT NLLS. The selected time window will be highlighted in red, and the user will be asked to confirm the selection.

The "Back" button allows the user to return to a previously assigned plant and select a different time window.

After all plants have been plotted, a progress bar will appear and FFT NLLS analysis will begin. Upon completion, the user will be asked if they wish to proceed to peak time analysis for the same data, using the same export groups. This is optional and is described on page 51.

After completion, the following folders and files will have been produced in the selected directory.

- A folder named after each worksheet group (genotype) defined by the user. Within each of these is all the data relating to that genotype.

  o Folders named after each worksheet in that genotype group. These contain the individual .out, .sum and .cnd files for each plant. As usual, appearing in the cnd files is the name of the data file that was actually passed to `four-anl`. These file

names were generated by BRASS by taking the name of the column in which that plant appeared in the original worksheet that was exported, flowed by an underscore, followed by the name of the genotype read from the original worksheet, up to a maximum of 8 characters. Using the column name is the only way to uniquely identify a plant on a particular worksheet. Creating separate folders ensures that no files will be overwritten, as could happen if two plants in corresponding columns on separate worksheets generated the same file name. If the auto delete file check box was selected, then the out and sum files will have been deleted.

- A cnd file named after the genotype. To this has been copied the contents of all the cnd files in all the worksheet folders above. Headings indicate which worksheet each data set of values came from. Thus it provides complete information about one genotype, and every plant is uniquely identifiable.

- An Excel workbook named after the genotype. This contains a worksheet corresponding to each worksheet in this export group. To these have been copied the actual data that was passed to `four-anl`. Data is written to the same cells that it was read from in the original export sheet. This allows easy comparison. Column headings are also copied over. In this case the "File" field shows the name of the data file that this data was passed to `four-anl` as. This can therefore be compared to the names in the cnd files, to easily see which data produced which period estimate. Below each column is the corresponding theoretical data. Values are listed first, and then time points below. As in "I&A", every other point has been deleted to reduce the number of values. In the case of data sets to which a period could not be fitted, the message "No Period" is added to the top of that column to remind the user. Alternatively, if FFT NLLS did not execute successfully, and was timed out, then the data is not copied but the message "FFT Failed" is added.
  Note that any "-1" values appearing on the original imported worksheets will still appear on this worksheet even though they were not passed to FFT NLLS. This creates an unbroken block of data to make the graphing functions simpler (see later).

- A results workbook for the entire export will have been created, named after the original imported data workbook, with "_Results" appended. This will contain

  - A master cnd file for each export group, with statistical analysis (page 60) performed.
  - Worksheets named "Waveforms for *group name*" for each group. These worksheets provide a quantitative measure of the shape of the oscillating data series by

applying two of Excel's built in statistical functions, Skewness and Kurtosis. Both these are intended to be applied to a frequency distribution, but can be applied to time series data by picking out consecutive troughs as the limits of one wave, and treating this peak as a distribution. Details of how these values are calculated and what they mean are described on page 56. Values are calculated for both the real and theoretical data, using as many waveforms as can be extracted from the data. Plants are listed in column A, by region name, under heading for each worksheet in the group. Time ranges used appear in columns B (real data) and E (theoretical data). A range of statistical parameters will have been calculated on the distribution of skewness and kurtosis values for all plants in this export group. Note that "n" represents the actual number of waveforms analysed, not the number of plants.

o If peak time analysis was performed (page 51), then a sheet named "*group name*.phs" will be present for each export group. This will be an imported phs file. Plants in the group are listed in column A, by region name, under heading for each worksheet in the group. The time window from which the peak was selected appears in column C, and the peak time in column D. Numeric values, ie excluding "first", "last" and "skipped", will have been copied to columns F and G, and a range of statistical parameters will have been calculated on the distribution of these values for all plants in this export group. Here, "n" represents the number of peak times analysed, and "Total data sets", the total number of plants, including those for which a peak time was not specified.

o A sheet named "Summary", which contains a table of values extracted from all the other sheets. Column B lists the statistical parameters. The other columns represent export groups, and are named after the cnd file for that group. "Number of plants" is the total number of data sets on all worksheets in that export group, including any rejected for analysis and any for which a period could not be fitted. "Number of rhythmic plants" is copied from the cnd page, and represents the number of plants which returned at least one period value. The statistics listed under "Period", "Amplitude" and "RAE" are copied from the relevant cnd pages, and skewness and kurtosis values are copied from the waveform pages. If peak time analysis was performed, then values listed under "Peak Times" will have been copied from the relevant phs page.

Note that the summary page cells are linked to the cells on other sheets rather than simply being copied values. For this reason, of any of the cnd sheets are altered, by repeating the Variance or Rel Amp weighted mean functions (page 60), then the summary sheet will be updated to reflect the changes. *For this reason, it is essential that none of the sheets in this workbook are ever re-named*.

o Sheets named "Histogram", "Plot Means" and "Scatter plot". These are copied over from BRASS itself and are for generating plots of the distribution of values found in this workbook. Their use is described on pages 61 to 63

Analysing a fixed time window

This methods works as described above, except that no graphs are displayed as ZT windows are automatically selected for each plant. This is done by locating the first ZT value equal to or greater than the start time specified by the user, and the first value equal to or greater than the end value specified, or the final ZT value if this is sooner.

Analysing all possible time windows

This is a more complicated method of analysis. Multiple time windows are analysed for each plant. The first window begins at the first ZT value equal to or greater than the specified starting point, and ends at the first ZT value equal to or greater than the sum of the specified starting point and the specified length. Successive time windows are produced by incrementing the starting point of the previous until it is equal to or greater than the sum of the previous starting point and the specified time advance. This continues until the end of a time window exceeds the end of the data.

After all the required parameters have been entered, a progress bar will appear and FFT NLLS analysis will begin. Upon completion, the user will be asked if they wish to proceed to peak time analysis for the same data, using the same export groups. This is optional and is described on page 51.

After completion, the following folders and files will have been produced in the selected directory

- A folder named after each export group (genotype) defined by the user. Within each of these is all the data relating to that genotype.

  o Folders named after each worksheet in that genotype group. These contain the individual .out, .sum and .cnd files for each time window. These are named in the format "*Column_Genotype_StartTime_StopTime.*cnd", where column is the name of

the column where the data was read from in the original worksheet. As usual, appearing in the cnd files is the name of the data file that was actually passed to `four-anl`. These are named (max 8 chars without extension), as follows "*Column_Genotype.xx*", where xx represents the start ZT time of this window. Corresponding .sum and .out files are also present. This way there will never be any uncertainty over which period refers to which plant and which time window. In addition to these files, master cnd files are created for each plant on the worksheet. These are names in the format "*Column_Region_Genotype*.cnd". They are made by copying the contents of each cnd file for each time window found for that plant. The data file names within these files therefore identify which time window each period value refers to, as shown below.

Creating separate folders ensures that no files will be overwritten, as could happen if two plants in corresponding columns on separate worksheets generated the same file name. If the auto delete file check box was selected, then the individual cnd, out and sum files will have been deleted, leaving just the master files for each plant.

o   A master condense file named after the export group. This was produced by copying the contents of the cnd file which produced the best Rel Amp value for each plant. The names of the data files in this file will therefore identify which was the best ZT window for each plant in the genotype. Headings indicate which worksheet each data set of values came from. Thus it provides complete information about one genotype, and every plant is uniquely identifiable. Underneath each data file name has been added the ZT range it refers to, as shown below

o An Excel workbook named after the export group, as show below.

This contains a worksheet corresponding to each worksheet in this export group (genotype). To these have been copied the data which produced the best Rel Amp values for each plant. Data is written to the same cells that it was read from in the original export sheet. This allows easy comparison. Column headings are also copied over. In this case the "File" field shows the name of the data file that this data was passed to `four-anl` as. This can therefore be compared to the names in the cnd files, to easily see which data produced which period estimate. Below each column is the corresponding theoretical data. Values are listed first, and then time points below. As in "I&A", every other point has been deleted to reduce the number of values. In the case of plants to which a period could not be fitted for any time window, the message "No Period" is added to the top of that column to remind the

user. Alternatively, if every time window for that plant failed the validation criteria, ie too many empty cells, then the message "Invalid Data" is added. Finally, if FFT NLLS did not execute successfully for any time window, then the message "FFT Failed" is added.

Note that any "-1" values appearing on the original imported worksheets will still appear on this worksheet even though they were not passed to FFT NLLS. This creates an unbroken block of data to make the graphing functions simpler (see later).

In addition to the data sheets, this workbook contains an additional sheet called "Report". This lists each ZT window for each plant that caused any of these three problems, as shown below

- A results workbook for the entire export will have been created, named after the original imported data workbook, with "_Results" appended. This will contain

  o The master cnd file for each export group, with statistical analysis (page 60) performed.
  o Worksheets named "Waveforms for *group name*" for each group. These worksheets provide a quantitative measure of the shape of the oscillating data series by applying two of Excel's built in statistical functions, Skewness and Kurtosis. Both these are intended to be applied to a frequency distribution, but can be applied to time series data by picking out consecutive troughs as the limits of one wave, and treating this peak as a distribution. Details of how these values are calculated and what they mean are described on pages 56. Values are calculated for both the real and theoretical data, using as many waveforms as can be extracted from the data. Plants are listed in column A, by region name, under heading for each worksheet in

the group. Time ranges used appear in columns B (real data) and E (theoretical data). A range of statistical parameters will have been calculated on the distribution of skewness and kurtosis values for all plants in this export group. Note that "n" represents the actual number of waveforms analysed, not the number of plants.

o If peak time analysis was performed (page 51), then a sheet named "*group name*.phs" will be present for each export group. This will be the imported .phs file. Plants in the group are listed in column A, by region name, under heading for each worksheet in the group. The time window from which the peak was selected appears in column C, and the peak time in column D. Numeric values, ie excluding "first", "last" and "skipped", will have been copied to columns F and G, and a range of statistical parameters will have been calculated on the distribution of these values for all plants in this export group. Here, "n" represents the number of peak times analysed, and "Total data sets", the total number of plants, including those for which a peak time was not specified.

o A sheet named "Summary", which contains a table of values extracted from all the other sheets. Column B lists the statistical parameters. The other columns represent export groups, and are named after the cnd file for that group. "Number of plants" is the total number of data sets on all worksheets in that export group, including any for which a period could not be fitted. "Number of rhythmic plants" is copied from the cnd page, and represents the number of plants which returned at least one period value. The statistics listed under "Period", "Amplitude" and "RAE" are copied from the relevant cnd pages, and skewness and kurtosis values are copied from the waveform pages. If peak time analysis was performed, then values listed under "Peak Times" will have been copied from the relevant phs page.

Note that the summary page cells are linked to the cells on other sheets rather than simply being copied values. For this reason, of any of the cnd sheets are altered, by repeating the Variance or Rel Amp weighted mean functions (page 60), then the summary sheet will be updated to reflect the changes. *For this reason, it is essential that none of the sheets in this workbook are ever re-named*.

o Sheets named "Histogram", "Plot Means" and "Scatter plot". These are copied over from BRASS itself and are for generating plots of the distribution of values found in this workbook. Their use is described on pages 61 to 63.

Other Ways of Running FFT NLLS

The **Analysis** menu item contains three other sub items that refer to FFT NLLS

1. **Run FFT NLLS** is designed to allow the suite of programs to be used to analyse a pre-existing data file, .sum file or .out file. It launches the form below.
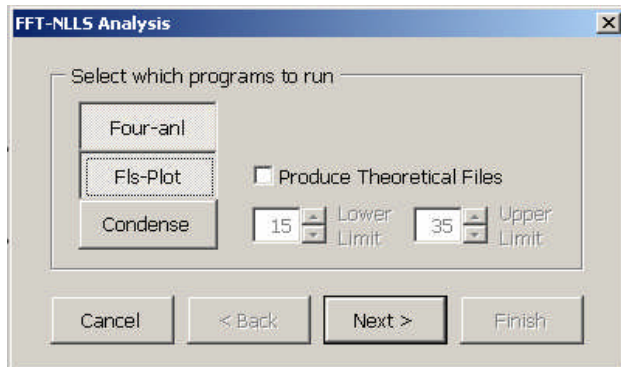
The user selects which of the programs to run, and supplies the inputs each requires. The form below appears asking the user to select a file of the type appropriate to the first program to be run. In this case, as four-anl was selected, a .in file is requested.
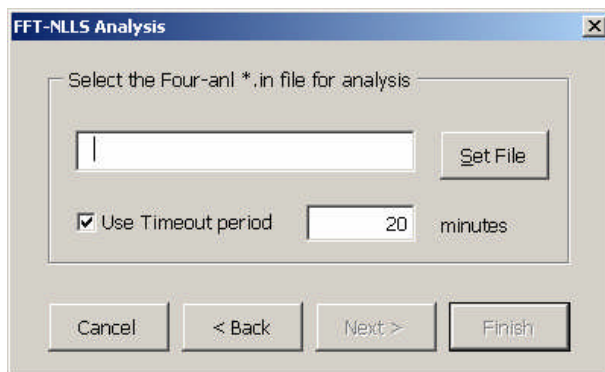
BRASS times the execution of each program. The timeout period is the time after which a program will be shut down if it has not finished executing, ie it is assumed to have crashed. If the box is not checked them programs will be allowed to run for an infinite time.

As the programs execute, the form below appears and provides information about what is happening, and messages if anything goes wrong.
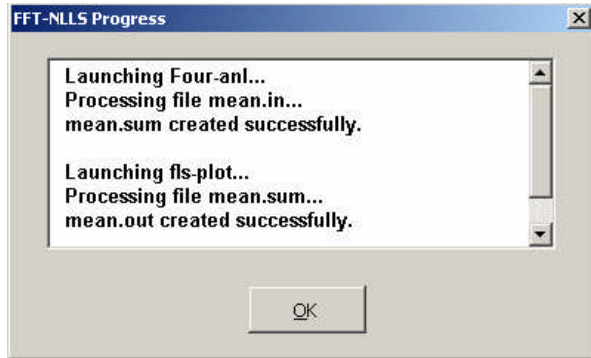
The new files created can them be found in the specified directory.

2. **Range to FFT NLLS** is for producing an instant period value for a selected data range. The user should highlight the exact cells they wish to analyse, and then click this menu item. After selecting the directory to put the resulting files in, this range is then passed to FFT NLLS, and the resulting condense file imported into BRASS. This is useful for verfying period values (Appendix B) , but note that if the selected data range contains "-1" values, then it will return a different value to that in the cnd file, as these values will have been temporarily removed during the export process.

3. **Launch DOS Prompt** is simply used to open `four-anl` or `condense`, so they may be run in the conventional way. Before opening them, the user is asked to set the current directory. This will be the directory that the programs look for input files in, so this should be set to the directory containing the in or .out file that is to be analysed. This way it will not be necessary to enter the full file path.

Note that `fls-plot` cannot be launched in this way, as it requires the use of command line arguments.

Mfourfit

Mfourfit is a new program that is an alternative to FFT NLLS. A more detailed description of the algorithm can be found in the appendix. In outline though, it attempts to fit a Fourier series to the data by varying the period of the series and looking for which period gives the best fit. Goodness of fit is measured by a term called the Akaike information criterion (AIC), which is a way measuring how well the theoretical data fits the real data, with a penalty for complexity (lots of Fourier coefficients) in the theoretical data. The smaller this value the better the fit. More information can be found on Wikipedia.

As well as period, Mfourit will return for each time series an amplitude value (maximum / minimum value for the first cycle of the theoretical fit) and a time for each peak found. Each peak time also has a corresponding sharpness (the second derivative of the slope at the peak, ie a measure of the rate of change of the slope at the peak so the higher the value the sharper the peak) and size value.

Warning: Unlike FFT NLLS, Mfourfit almost always returns a period value. This is because it returns the best fit it finds over the specified range. For non-rhythmic input data this is often the lowest or the highest period specified by the user.

Exporting to Mfourfit is done automatically after de-randomizing a workbook, or by selecting the **Analysis ->Run Mfourfit** menu item and selecting the required workbook.

Three options for selecting the time windows to be exported are now presented:

1. "All Over Same ZT". If this option is selected then the user will be asked to specify a fixed time window to apply to every plant in the analysis.
2. "Select Best ZT Window for each". If this option is selected then the entire time series will be plotted for each plant and the user will be able to select a range from the graphs individually.
3. "All Possible Time Windows". If this option is selected then user is asked to specify the length of a floating time window, and the increment between windows. Multiple windows are therefore analysed for each plant. BRASS automatically picks out those with the lowest Rel Amp values for each plant.

Whichever of these is selected, export groups must first be defined. The form below is displayed.

This works by the user specifying a group name, usually a genotype, under which to analyse one or more worksheets. This is done by clicking "Add". Then select the genotype from the "Groups" list. The "Worksheets unassigned" list contains the names of all the sheets in the book to be exported, which have not yet been assigned to any group. The "Worksheets in this group" are those which have been assigned to the currently selected group. Worksheets can be added to and removed from the currently selected group using the "Select" and "Remove" buttons. The significance of creating groups is that the periods etc… fitted for all the individual plants in that group, from one or more worksheets, will appear in the same condense file, so allowing BRASS to easily calculate mean values for that group. In this way, if there are multiple sheets for one genotype, then they can all be exported in one block to produce mean values for that genotype.

Very often, however, no genotype will appear on more than one sheet in a derandomized workbook. In this case the "Auto" button allows the user to save time by automatically generating a group name for each worksheet and assigning the worksheet to it. The user can then verify the groups created by clicking on the group names, and edit any as required.

Note that when using this function group names are generated by taking the first 8 characters of each sheet name. It is possible that 8 characters may not uniquely distinguish every worksheet. If this is the case then the user is warned that some worksheets have not been assigned to groups. These remain in the 'Worksheets unassigned' list box, and the user must manually create unique group names for them.

What happens next depends on the selected ZT window option. If a fixed window was

selected then the user is simply asked to specify the time window. If the selected start time is before the start of the data then the window will start at the first time point. If all possible time windows are selected then the user must supply details of these using the form below. Using the values shown, the program will start by analysing a time window beginning at time zero, or the earliest ZT time found if values do not start at zero.

The window selected will be 48 hours long. Before analysis the number of blank cells will be counted. These are common in leaf movement experiments when a particular leaf has moved out of the selected area at a particular timepoint, and are signified by containing the value '−1'. (These values are not used in the rhythm analysis.) If the percentage of cells in the range which are blank is below the maximum permitted, then this range is passed to Mfourfit. Otherwise, it is not analysed. The next range is then selected by incrementing the start time by the amount specified, in this case 12 hours. This will continue until the start time has advanced so far that the end point of the range exceeds the end of the data. When this happens, then this last range will not be analysed. If 'keep results for best window' is selected then only values relating to the time window that produced the best periodic fit is saved. If 'All windows' is selected then the period fitted to each time window will be saved. See results section below.

Whatever time window is selected, the next stage is to specify the upper and lower limits of the period which Mfourfit should attempt to fit to the data.

Next the user has the option to remove any trend in the experimental data before attempting to fit. This can be done independently for trends in amplitude of rhythm and for baseline trend.
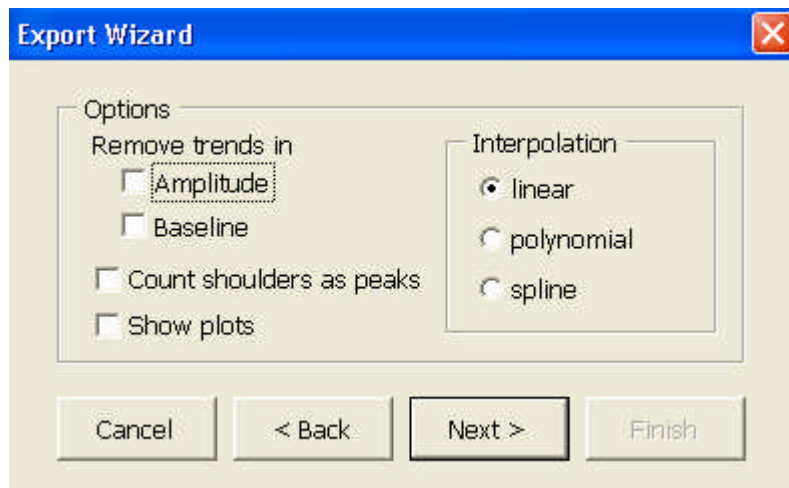
Warning. We have found that detrending data is essential to get an accurate measure of period. We therefore recommend that users always select these boxes. See Appendix F, page 82, for an explanation of detrending and a limitation of the process.

If 'Count shoulders as peaks' is selected, then times where the gradient of the curve is temporarily reduced, but without passing through zero, are counted as peaks. Two examples are shown below.
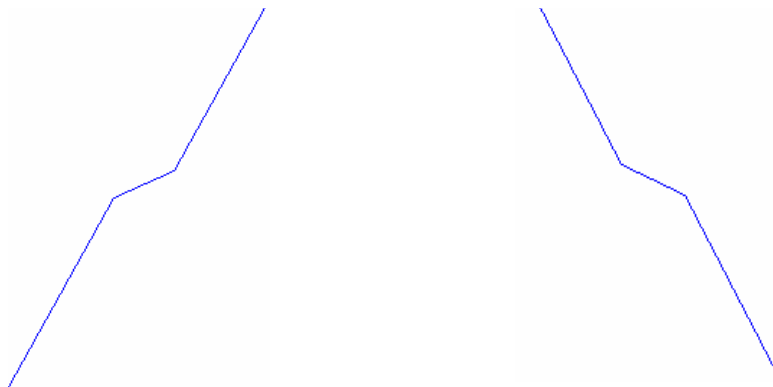
Checking 'Show Plots' will cause popup windows to appear during analysis showing each series of real data analyzed and the corresponding theoretical fit.

Mfourfit produces a theoretical time series with points corresponding to the real data. The user must also choose how to interpolate between these points to create a smooth curve. The simplest method is linear interpolation. For noisy data polynomial or spline fits give a better fit. Comparing the two, spline produces a smoother result for smooth data, and polynomial has no overshoots and less oscillation if the data are not smooth.

Warning. If you select linear interpolation then the theoretical data points corresponding in time to a real data points will be joined by straight lines, and therefore all peak phase

values calculated will correspond to real data points, which may not be the genuine phase. A spline fit is probably best if you are interested in phase.

Finally the user must specify a folder to put the results in, in which new sub-folders will be created, as described below.

The analysis is now ready to begin.

Selecting separate time windows for every plant

Each data set is graphed, and the user must select an appropriate time window, as show below. Remember that any "-1" values have been added to create a continuous data series for graphing and will not be passed to Mfourfit. The selected time window will be highlighted in red, and the user will be asked to confirm the selection.
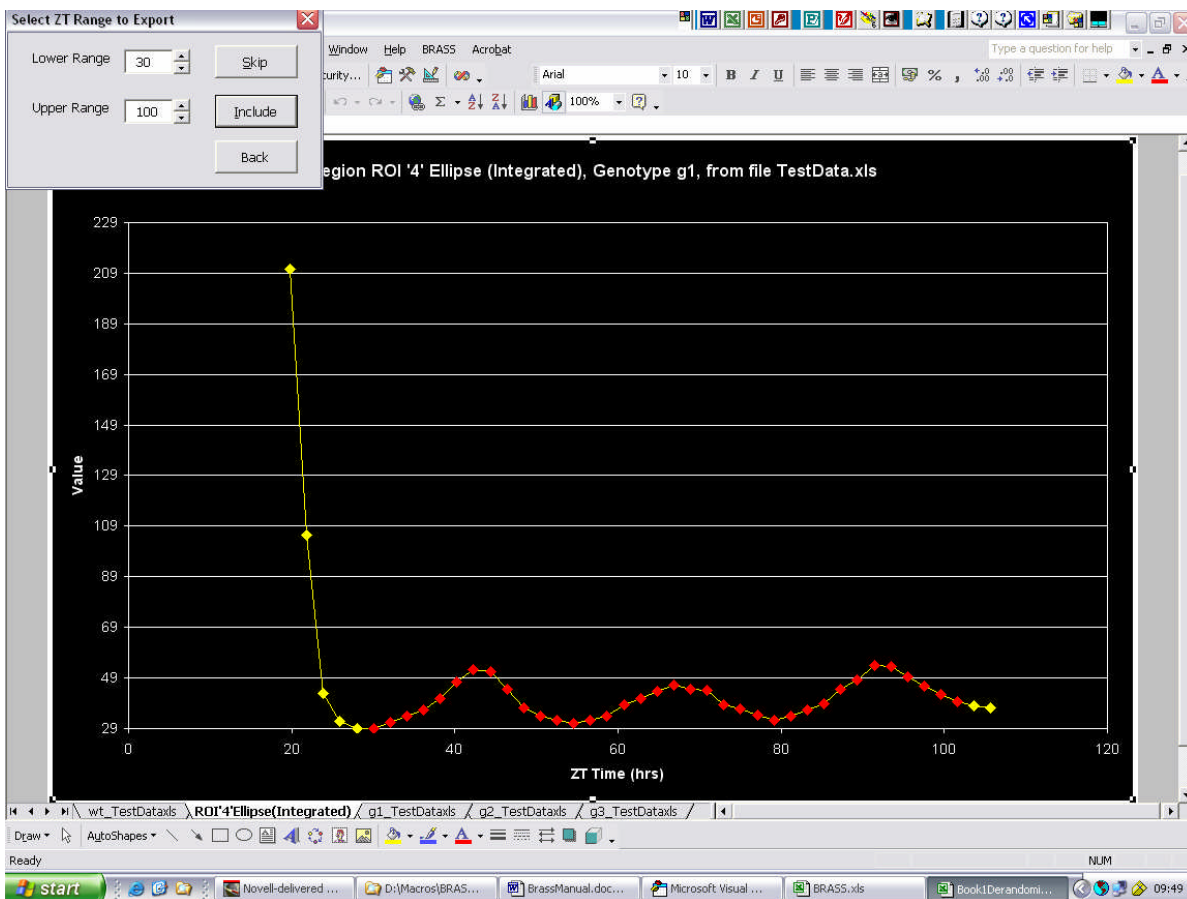
The "Back" button allows the user to return to a previously assigned plant and select a

different time window.

After all plants have been plotted, a progress bar will appear and analysis will begin. Upon completion, the user will be asked if they wish to proceed to peak time analysis for the same data, using the same export groups. This is optional and is described on page 51.

After completion, the following folders and files will have been produced in the selected directory.

- A folder named after each worksheet group (genotype) defined by the user. Within each of these is an Excel workbook named after the genotype. This contains the following sheets

  o Worksheets of time series data corresponding to each worksheet in this export group. To these have been copied the actual data that was used in the analysis. Data is written to the same cells that it was read from in the original export sheet. This allows easy comparison. Column headings are also copied over. Below each column is the corresponding theoretical data. Values are listed first, and then time points below. If Mfourfit did not execute successfully for a particular dataset then the message "No Period" is added to the top of that column.

    Note that any "-1" values appearing on the original imported worksheets will still appear on this worksheet even though they were not passed to FFT NLLS. This creates an unbroken block of data to make the graphing functions simpler (see later).

  o A worksheet named *genotype*.cnd. This contains, for each time series in this group, the period, amplitude, 'AIC' (confidence of the fit) and the phase, size and 'sharpness' of each peak. Statistical analysis has been performed on the data. See the section on statistics for an explanation of this.

  o A worksheet named "Waveforms for *genotype*". This provides a quantitative measure of the shape of the oscillating data series by applying two statistical measures, skewness and kurtosis. Both these are intended to be applied to a frequency distribution, but can be applied to time series data by picking out consecutive troughs as the limits of one wave, and treating this peak as a distribution. Details of how these values are calculated and what they mean are described on page 56. Values are calculated for both the real and theoretical data,

using as many waveforms as can be extracted from the data. Data series are listed in column A, by region name, under heading for each worksheet in the group. Time ranges used appear in columns B (real data) and E (theoretical data). A range of statistical parameters* will have been calculated on the distribution of skewness and kurtosis values for all plants in this export group. Note that "n" represents the actual number of waveforms analysed, not the number of time series.

Note the descriptive statistics performed on waveform data include skewness and kurtosis. These refer to the built in Excel functions and are used here to analyse the frequency distribution on the waveforms found, ie skewness of skewness values, skewness of kurtosis values etc…

- o If peak time analysis was performed (page 51), then a sheet named *genotype*.phs will be present. Plants in the group are listed in column A, by region name, under heading for each worksheet in the group. The time window from which the peak was selected appears in column C, and the peak time in column D. Numeric values, ie excluding "first", "last" and "skipped", will have been copied to columns F and G, and a range of statistical parameters will have been calculated on the distribution of these values for all plants in this export group. Here, "n" represents the number of peak times analysed, and "Total data sets", the total number of plants, including those for which a peak time was not specified.

- A results workbook for the entire export will have been created, named after the original imported data workbook, with "_Results" appended. This will contain

  - o A copy of the *genotype*.cnd, *genotype*.phs and waveforms worksheets described above for each export group, with statistical analysis (page 60) performed.
  - o A sheet named "Summary", which contains a table of values extracted from all the other sheets in this workbook. "Number of plants" is the total number of data sets on all worksheets in that export group, including any rejected for analysis and any for which a period could not be fitted. "Number of rhythmic plants" is copied from the cnd page, and represents the number of plants which returned at least one period value.

    Note that the summary page cells are linked to the cells on other sheets rather than simply being copied values. For this reason, of any of the cnd sheets are altered, by repeating the AIC weighted mean functions (page 60), then the summary sheet will be updated to reflect the changes. *For this reason, it is essential that none of the*

sheets in this workbook are ever re-named.

o Sheets named "Histogram", "Plot Means" and "Scatter plot". These are copied over from BRASS itself and are for generating plots of the distribution of values found in this workbook. Their use is described on pages 61 to 63.

Analysing a fixed time window

This methods works as described above, except that no graphs are displayed as ZT windows are automatically selected for each plant. This is done by locating the first ZT value equal to or greater than the start time specified by the user, and the first value equal to or greater than the end value specified, or the final ZT value if this is sooner.

Analysing all possible time windows

This is a more complicated method of analysis. Multiple time windows are analysed for each plant. The first window begins at the first ZT value equal to or greater than the specified starting point, and ends at the first ZT value equal to or greater than the sum of the specified starting point and the specified length. Successive time windows are produced by incrementing the starting point of the previous until it is equal to or greater than the sum of the previous starting point and the specified time advance. This continues until the end of a time window exceeds the end of the data.

After all the required parameters have been entered, a progress bar will appear and mfourfit analysis will begin. Upon completion, the user will be asked if they wish to proceed to peak time analysis for the same data, using the same export groups. This is optional and is described on page 51.

After completion, the following folders and files will have been produced in the selected directory

• A folder named after each export group (genotype) defined by the user. Within each of these is an Excel workbook named after the genotype. This contains the following sheets

o Time series worksheets corresponding to each worksheet in this export group

(genotype). Data is written to the same cells that it was read from in the original export sheet. This allows easy comparison. Column headings are also copied over. Below each column is the corresponding theoretical data for the time window which produced the best fit. Values are listed first, and then time points below. If Mfourfit did not execute successfully for any time window, the message "No Period" is added to the top of that column to remind the user. Alternatively, if every time window for that plant failed the validation criteria, ie too many empty cells, then the message "Invalid Data" is added.

Note that any "-1" values appearing on the original imported worksheets will still appear on this worksheet even though they were not used in the analysis. This creates an unbroken block of data to make the graphing functions simpler (see later).

In addition to the data sheets, this workbook contains an additional sheet called "Report". This lists each ZT window for each plant that caused either of the problems above.
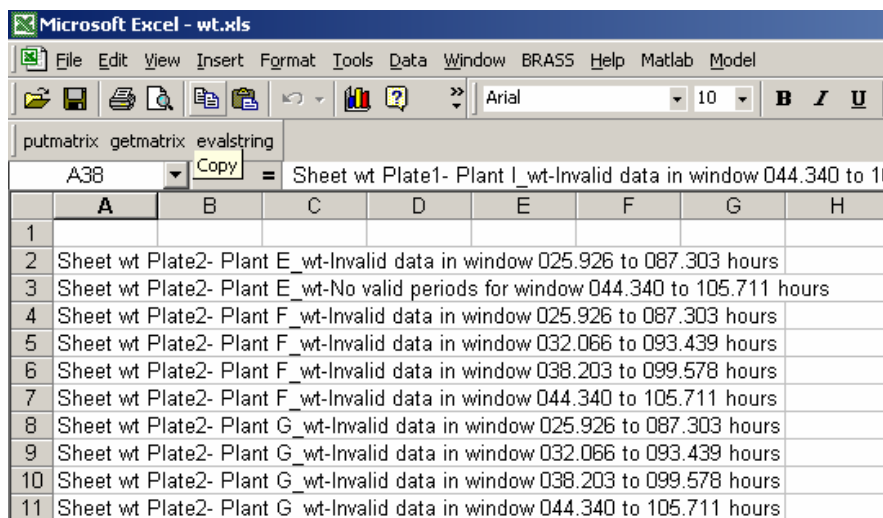
o A worksheet named *genotype*.cnd. This contains, for each valid time window analysed for each time series in this group, the period, amplitude, 'AIC' (confidence of the fit) and the phase, size and 'sharpness' of each peak. Statistical analysis has been performed on the data. See the section on statistics for an explanation of this. If the user opted to keep results for best window only, then only the time window with the best AIC score will be present for each time series.

o A worksheet named "Waveforms for *genotype*". This provides a quantitative

measure of the shape of the oscillating data series by applying two statistical measures, skewness and kurtosis. Both these are intended to be applied to a frequency distribution, but can be applied to time series data by picking out consecutive troughs as the limits of one wave, and treating this peak as a distribution. Details of how these values are calculated and what they mean are described on page 56. Values are calculated for both the real and theoretical data, using as many waveforms as can be extracted from the data. Data series are listed in column A, by region name, under heading for each worksheet in the group. Time ranges used appear in columns B (real data) and E (theoretical data). A range of statistical parameters* will have been calculated on the distribution of skewness and kurtosis values for all plants in this export group. Note that "n" represents the actual number of waveforms analysed, not the number of time series.

*Note the descriptive statistics performed on waveform data include skewness and kurtosis. These refer to the built in Excel functions and are used here to analyse the frequency distribution on the waveforms found, ie skewness of skewness values, skewness of kurtosis values etc…

- o If peak time analysis was performed (page 51), then a sheet named "*genotype*.phs" will be present. Plants in the group are listed in column A, by region name, under heading for each worksheet in the group. The time window from which the peak was selected appears in column C, and the peak time in column D. Numeric values, ie excluding "first", "last" and "skipped", will have been copied to columns F and G, and a range of statistical parameters will have been calculated on the distribution of these values for all plants in this export group. Here, "n" represents the number of peak times analysed, and "Total data sets", the total number of plants, including those for which a peak time was not specified.

- A results workbook for the entire export will have been created, named after the original imported data workbook, with "_Results" appended. This will contain

  - o A copy of the *genotype*.cnd, *genotype*.phs and waveforms worksheets described above for each export group, with statistical analysis (page 60) performed.

  - o A sheet named "Summary", which contains a table of values extracted from all the other sheets in this workbook. "Number of plants" is the total number of data sets on all worksheets in that export group, including any rejected for analysis and any for which a period could not be fitted. "Number of rhythmic plants" is copied from the

cnd page, and represents the number of plants which returned at least one period value.

Note that the summary page cells are linked to the cells on other sheets rather than simply being copied values. For this reason, of any of the cnd sheets are altered, by repeating the AIC weighted mean functions (page 60), then the summary sheet will be updated to reflect the changes. *For this reason, it is essential that none of the sheets in this workbook are ever re-named.*

o Sheets named "Histogram", "Plot Means" and "Scatter plot". These are copied over from BRASS itself and are for generating plots of the distribution of values found in this workbook. Their use is described on pages 61 to 63.

Measuring Peak Time Values

It appears that both the FFT NLLS suite of programs and Mfourfit do not always provide the best estimate of phase if the period of a data set varies with time. For this reason we have included in BRASS a manual way of measuring the time of selected peaks. This can be done automatically following export of data to FFT NLLS or Mfourfit, using the same export groups. Alternatively, the user can click the menu item **Analysis->Phase Values->Manual Selection** and select the workbook they wish to analyse. This will probably be book which has been de-randomized (page 24), if there is more than one genotype present. If this is the case then the form below is then show to allow the user to define the groups.
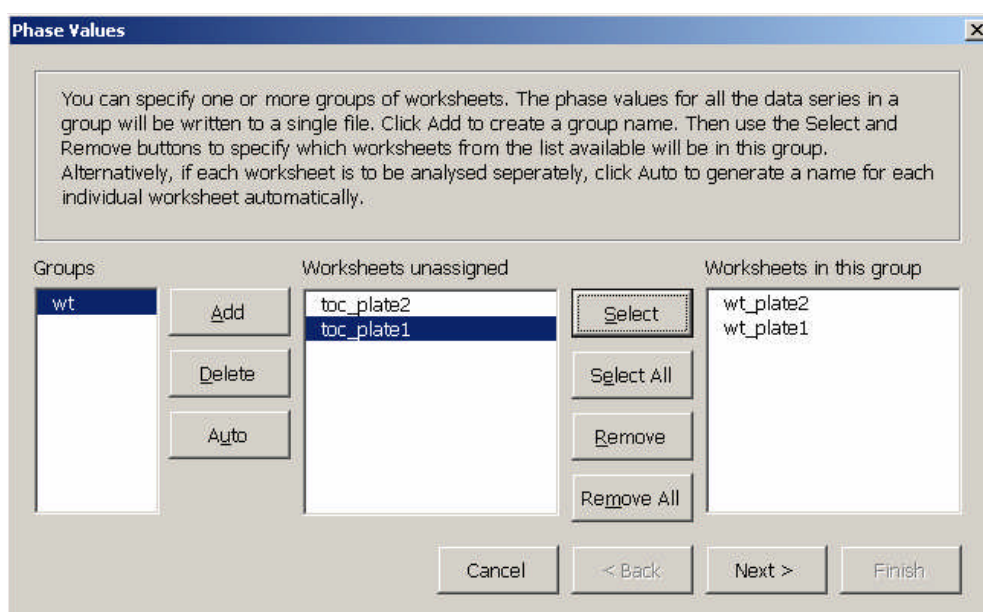
This works by the user specifying a group name, usually a genotype, under which to analyse one or more worksheets. This is done by clicking "Add". Then select the genotype from the "Groups" list. The "Worksheets unassigned" list contains the names of all the sheets in the book to be analysed, which have not yet been assigned to any group. The "Worksheets in this group" are those which have been assigned to the currently selected group. Worksheets can be added to and removed from the currently selected group using the "Select" and "Remove" buttons. At the end of the analysis, one file will have been produced for each group created, allowing mean peak times to be calculated for all the data series in this group, across several worksheets if required.

Next, the user is asked to specify a directory in which to put the results files. Analysis then begins.

For each data series, a three point average value for each timepoint is calculated by taking the mean of the value for that timepoint, and the preceding and succeeding points. This smoothes the data out and makes finding peak values more reliable. The user should be aware, however, that they are lot looking at the raw data. Cells containing the value "-1" are omitted. The averaged data is then plotted, as shown on the next page.

Users will probably wish to confine their search for a peak to a specified time window. By default, only the first 24 hours of data is plotted, but the user can easily change this by altering the drop down lists on the form displayed. As can be seen, the highest value in the time window is located and coloured red. This is the value that has been selected as representing the time of the peak. This value is displayed above the upper group of buttons. The user then has the option to accept this value by clicking "Accept". However, the selected point may not always be the true peak, but an anomalous point. If this is the case, the user can click "Lower". This will result in the next highest point being located. This can be done repeatedly until the user is satisfied with the selected time. Clicking "Higher" allows the user to move back to the previously selected point.
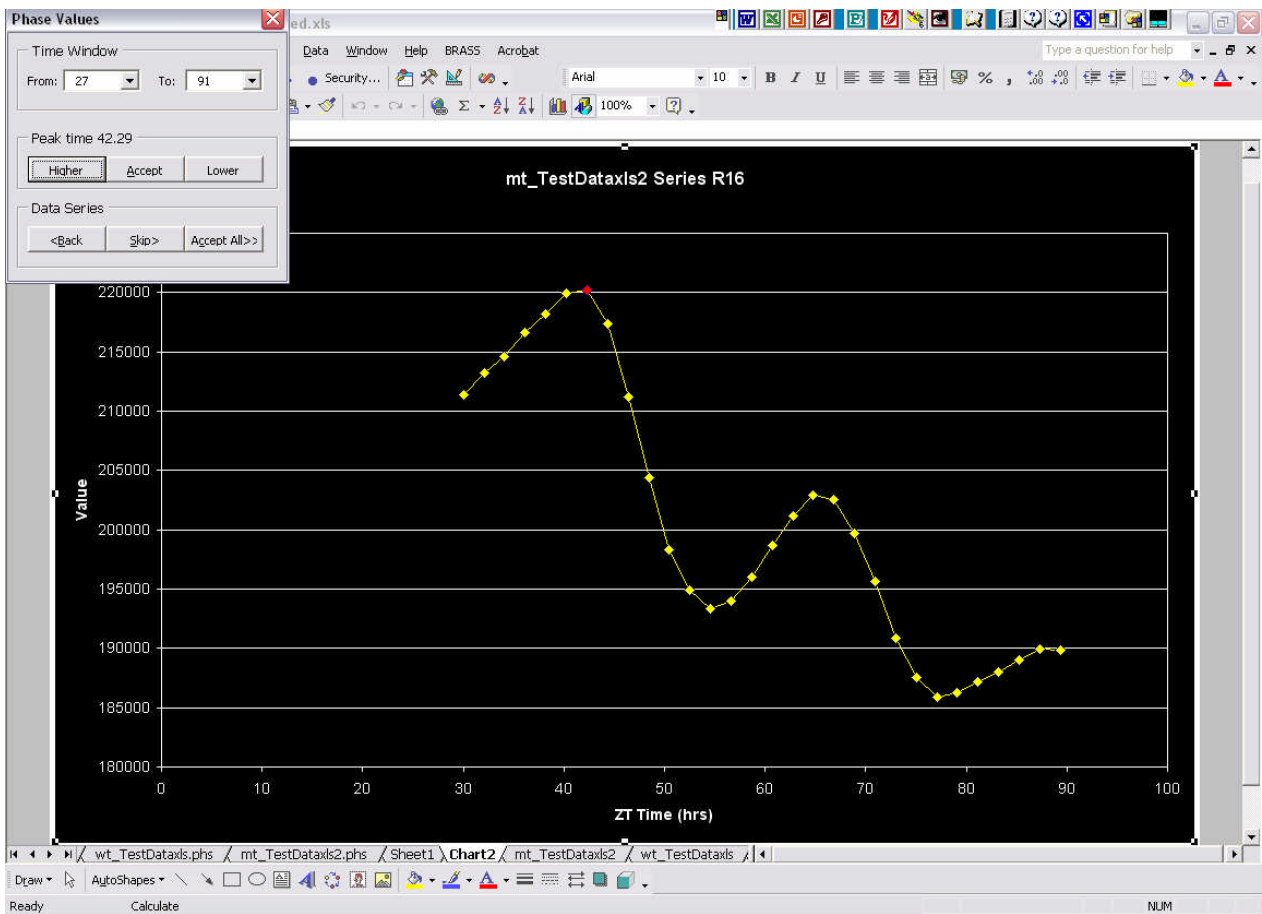
The lower group of buttons allow the user to "Skip" a data series. The label "Skipped" is

then written to the results file. The button "Accept All", allows the user to save time by not displaying the graphs at all for the remainder of this genotype. In this case the highest point in the specified time window is always accepted as peak time, unless it happens to be the first or last point in the time window. When this is the case, it is assumed that the user would want to reject this point as not being a true peak time, and so the label "First" or "Last" is written to the results file. The "Back" button allows the user to return to the previous data series and change the point selected.

Results will appear as a new sheet in the data workbook. If peak time analysis was carried out automatically after exporting to FFT NLLS or Mfourfit, then these sheets will have been copied into the results workbooks as well.

The format of these files when imported is shown below.

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3	**Manual Estimates of Peak times for wt**										
4											
5											
6											
7											
8	wt					ROI '1' Ell	48.43		**Descriptive Statistics**		
9						ROI '2' Ell	66.84		Mean=	62.94	
10	ROI '1' Ellipse	46 to 91 hours		48.43		ROI '3' Ell	66.84		SD=	7.23	
11	ROI '2' Ellipse	46 to 91 hours		66.84		ROI '4' Ell	66.84		SEM=	2.18	
12	ROI '3' Ellipse	46 to 91 hours		66.84		ROI '5' Ell	48.43		Median=	66.84	
13	ROI '4' Ellipse	46 to 91 hours		66.84		ROI '6' Ell	66.84	Sample Variance=		52.27	
14	ROI '5' Ellipse	46 to 91 hours		48.43		ROI '7' Ell	66.84		Kurtosis=	1.87	
15	ROI '6' Ellipse	46 to 91 hours		66.84		ROI '8' Ell	64.80		Skewness=	-1.85	
16	ROI '7' Ellipse	46 to 91 hours		66.84		ROI '11' El	64.80		Minimum=	48.43	
17	ROI '8' Ellipse	46 to 91 hours		64.80		ROI '12' El	64.80		Maximum=	66.84	
18	ROI '9' Ellipse	46 to 91 hours		First		ROI '14' El	66.84		Range=	18.41	
19	ROI '10' Ellipse	46 to 91 hours		First					Sum=	692.30	
20	ROI '11' Ellipse	46 to 91 hours		64.80					n=	11	
21	ROI '12' Ellipse	46 to 91 hours		64.80				Total data sets=		14	
22	ROI '13' Ellipse	46 to 91 hours		First							
23	ROI '14' Ellipse	46 to 91 hours		66.84							
24											
25											

The title at the top of the worksheet indicates the group name and the imported workbook from which the data came. The region name and time window from which the peak was selected, and the peak value for each data series is listed under the name of the worksheet it came from, allowing each to be uniquely identified. For each region, the column number the data was taken from in the original imported workbook is also shown. This aids the user in matching values in this file with values in the cnd file for the same export group. Valid peak time times are then copied into column "G", and a range of

Excel's built in statistical functions are performed on the distribution of these values for all plants in this export group. Here, "n" represents the number of peak times analysed, and "Total data sets", the total number of plants, including those for which a peak time was not specified ("skipped", "first", or "last"), of which in this case there are none.

Correcting Phase Values Before Statistical Analysis

Very often the phase values generated by analysis or manually need further processing before they are suitable for statistical analysis.

Extrapolating Phases

Peak times may have been measured from different time windows for different data series, for example a cnd sheet may have a number of values clustering around 10 hours, and another set may cluster around 34 as these were measured in the following cycle period, assuming a period of around 24 hours. This would give a biphasic distribution with a mean phase of somewhere in between the two peaks. To prevent this go to **Analysis->Phase Values->Correct Values** and when prompted enter a time value to extrapolate the phase values towards. BRASS then takes each phase value on the current sheet and extrapolates by adding or subtracting a whole number of corresponding period values to get as close to this value as possible. So in the above example, if you enter a value of 10, all phase values clustering around 34 will have around 24 subtracted, so all values will now cluster around 10 hours.

Sorting Phase Values

When statistical analysis is performed, only the peak time listed first for each time series is used. By default this is the first chronologically. This may not always be desirable though. Open the cnd sheet of interest and go to **Analysis->Phase Values->Sort Phases** to sort multiple peak times by another criteria. The other options are size, sharpness and peak closest in time to a value entered by the user.

When phases are altered by either of the above the statistics are automatically re-calculated.

Analysing Waveforms

During the export of data to FFT NLLS, the waveform of any oscillations found is analysed by treating all the values between 2 appropriately separated troughs as a frequency distribution and performing two statistical functions which describe the shape of a distribution on these values. These are skewness and kurtosis. This "miss-use" of these functions provides a useful quantification of the shape of the waveform. Excel's built in functions are not appropriate for this interpretation of time series data, so functions written by Alex Morton, Department of Statistics, University of Warwick are used instead. Analysis is done by talking the time window selected for export and picking out every possible pair of troughs whose distance of separation is within the range of time periods of interest, as specified by the user during export to FFT NLLS or Mfourfit. Each of these data ranges is then considered to represent one waveform.

Skewness measures the degree of asymmetry of a distribution around its mean. A positive value indicates more values greater than the mean, and a negative value indicates more values less than the mean. In terms of time series data, waveforms would look like this:
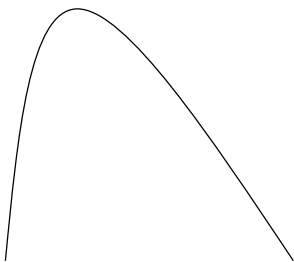
Positive skewness   Negative skewness
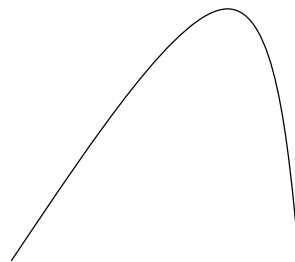
Kurtosis measures the relative peakedness or flatness of a distribution compared with the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution. In terms of time series data, waveforms would look like this:

Positive kurtosis   Negative kurtosis

The end result of this presence in the results summary workbook of a sheet named "Waveforms for *export group name*" for each export group.

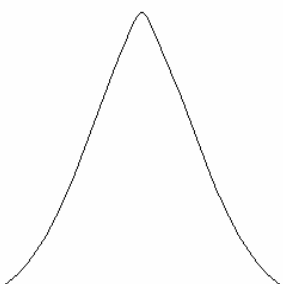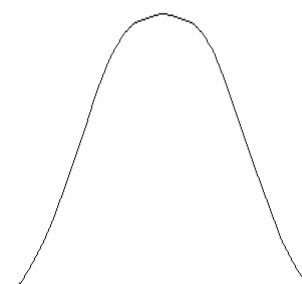|	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Waveform analysis								Descriptive Statistics				
2		Real Data			Theoretical Data					Real Data		Theoretical Data	
3	Data	Times	Skewness	Kurtosis	Times	Skewness	Kurtosis			Skewness	Kurtosis	Skewness	Kurtosis
4									Mean=	-0.09898	2.323531	0.008703	2.269214
5	wt_TestDataxls								SD=	0.121826	0.145058	0.130461	0.088651
6									SEM=	0.040609	0.048353	0.033685	0.02289
7	R1, col E	54.6 to 79.	-0.18805	2.178715	24.7 to 53.	-0.01135	2.308047		Median=	-0.16789	2.326781	-0.01251	2.250706
8					53.8 to 82.	0.228162	2.220711		Sample Variance=	0.014841	0.021042	0.01702	0.007859
9	R2, col F	32.1 to 54.	0.090795	2.179377	24.6 to 52.	-0.0257	2.307294		Kurtosis=	-0.16451	-1.33491	2.567598	1.452561
10		54.6 to 79.	-0.19092	2.487163	52.2 to 79.	-0.02549	2.307197		Skewness=	1.24174	-0.2239	1.303716	0.779071
11									Minimum=	-0.19092	2.103172	-0.19274	2.136378
12	R3, col G	54.6 to 79.	-0.15053	2.513012	21.7 to 50.	-0.1376	2.213919		Maximum=	0.115041	2.513012	0.341135	2.482725
13					50.7 to 79.	0.025625	2.304354		Range=	0.305961	0.40984	0.533877	0.346347
14	R4, col H	32.1 to 54.	0.115041	2.103172	20.1 to 39.	-0.0389	2.136857		Sum=	-0.89079	20.91178	0.13055	34.03821
15		54.6 to 79.	-0.17416	2.403418	39.1 to 55.	0.341135	2.244905		n=	9	9	15	15
16					55.4 to 78.	-0.07506	2.389433						
17					78.2 to 97.	-0.04438	2.136378						
18	R5, col I	54.6 to 79.	-0.05295	2.326781	24.4 to 53.	0.026537	2.250706						
19					53.0 to 81.	0.027042	2.250666						
20	R6, col J	54.6 to 79.	-0.16789	2.304021	54.1 to 78.	0.045803	2.482725						
21													
22	R7, col K	54.6 to 79.	-0.17212	2.416119	21.4 to 50.	-0.19274	2.219339						
23					50.6 to 79.	-0.01251	2.265676						
24													

Column A lists all the worksheets in the export group, and all the regions (data sets) on each sheet. For each region, the column number the data was taken from in the original imported workbook is also shown. This aids the user in matching values in this file with values in the cnd file for the same export group. Columns B, C and D describe the actual experimental data, and columns E, F and G the theoretical data generated by the FFT NLLS suite of programs. The time ranges picked out for analysis are shown in columns B and E. A range of Excel's built in statistical functions are then performed on the distribution of values for skewness and kurtosis (columns I to M).

Graphing Time Series Data

These functions are for plotting time series data from the active worksheet. The Graph menu item contains 3 sub-menu items

1. **Graph Selected**. This function launches a form which allows the user to select one or more data series from a list of all the data series found on the active worksheet. If the "Plot Separately" box is checked then each selected series is plotted on its own graph. Otherwise, all selected series are plotted on one graph. If theoretical data is found on the worksheet, then the "Add Theoretical Data" box appears. This becomes enabled only if the "Plot Separately" box is checked. If this second box is checked then corresponding theoretical data appears on the same axes as the experimental data.

2. **Graph All**. This function plots every data series found on the sheet. Series are plotted on graphs in groups of not more than 8 for clarity. Theoretical data is not plotted.

3. **Graph Mean Level**. This function plots a histogram of the average values of each data series, ie the average reading over the entire ZT time window. Empty cells, and cells with "-1" in are ignored. If more than 32 data series are found on the sheet then series are split over 2 or more histograms for clarity. This function is useful for luciferase imaging experiments, for detecting data series which fail to luminesce, or which are excessively and constitutively bright.

In all cases graphs are added to the data workbook.

Averaging Data Series

BRASS contains functions to create an average data set for each worksheet in a workbook. The user should first make the workbook they wish to average the active workbook. This should be a valid imported data workbook. Then select the **Averaging->All** menu item. A column is then added to each worksheet to the right of the last data series which represents a mean data series for that worksheet. A chart is plotted of the mean data series from each worksheet.

Note that the empty column between the final actual data series and the average series prevents the average series being processed if the worksheet is subsequently exported to FFT NLLS or Mfourfit.

Alternatively, to average selected time series on a particular worksheet, select **Averaging->Selected**. You can then select columns from the current sheet. The average of these is again added to the right of the data.

Statistical Analysis of Results

Statistical analysis can be performed on cnd sheets by using the **Statistics** menu item.

Cnd Sheets Produced by FFT NLLS

Enter an 'x' in the 'Periods to be omitted from the mean' column next to any data to be omitted from the analysis. Select the **VarWtdMean** or **RAEWtdMean** to weight individual values for period, amplitude, phase and rel-amp by variance or rel-amp values.

Cnd Sheets Produced by Mfourfit

Enter an 'x' in the 'Omit?' column next to any data to be omitted from the analysis. Select **AICWtdMean** to weight individual values for period, amplitude and phase by AIC.

Another type of mean, we have called "Partially weighted" is also calculated. This is done by calculating the mean of all the weighting factors (variance or rel-amp or AIC) for a set of period, amplitude, phase or rel-amp values, and then adding the mean to each weighting factor in the set. The purpose of this is to prevent single values with very low errors from distorting the mean too much. These new weighting factors are written in a column to the right of the original values, entitled "Partially Weighted C.I. x 0.5", "Partially Weighted RAE", or "Partially weighted AIC".

As well as these two types of mean, the full range of built in Excel descriptive statistics are calculated for each set of period, amplitude, phase or rel-amp values. These are unweighted mean, SEM, standard deviation, median, variance, kurtosis, skewness, minimum value, maximum value, range and sum.

The parameter appearing last, "Total rhythmic data sets" is the total number of data sets that returned at least one valid period.

Plotting Results Parameters

The results workbooks produced by BRASS contain three worksheets to enable rapid plotting of the results data in three different formats, to allow easy comparison between different genotypes. These sheets act upon any cnd, phs or waveform sheets present in the workbook.

**Note** that in order for the routines to find the data, none of the data sheets should ever be re-named.

1. **Histograms**. The "Histogram" sheet contains a drop down list box. The user should select the parameter to plot, period, amplitude or phase from any cnd sheets, peak times from any phs sheets, or skewness or kurtosis values from any waveform sheets, and then click "Get Data". This causes the appropriate values to be copied to the "Histogram" sheet. They are divided into bins according to the values specified for "Upper limit", "Lower limit", and "Bin". Values are suggested automatically for these but the user can enter their own choice, which will cause the bin-frequency values to be re-calculated. When ready, clicking "Plot Histogram" causes the chart to be created as a separate sheet, with different files (genotypes) plotted as different series.
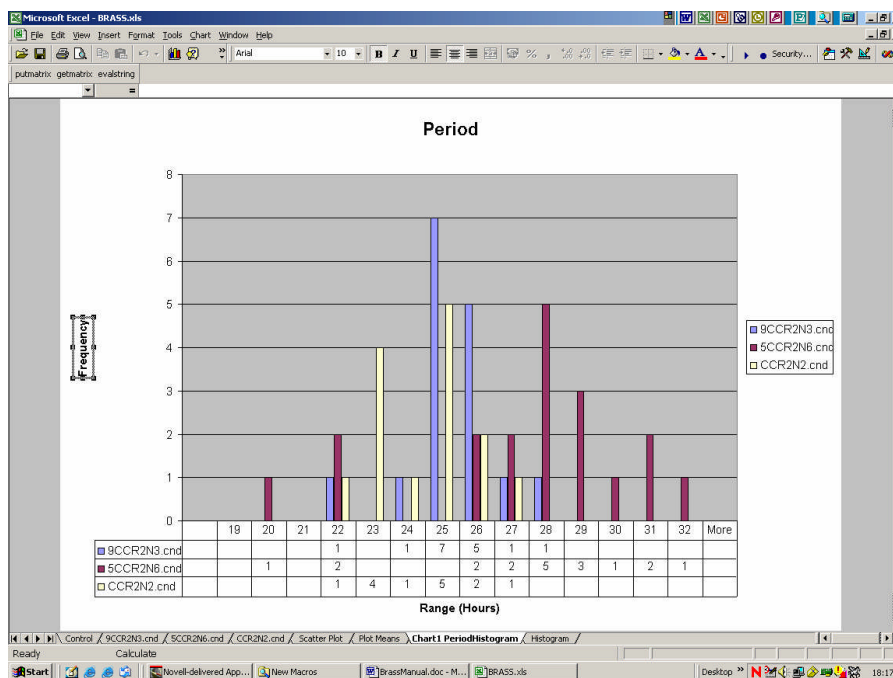
2. **Mean Plots**. The "Plot Means" sheet is designed to plot a single parameter mean from each worksheet, with appropriate error bars. Drop down list boxes allow the user to specify the parameter (period, amplitude or phase from any cnd sheets, peak times

from any phs sheets, or skewness or kurtosis values from any waveform sheets), type of mean (unweighted, partially weighted, weighted), and error bars (SEM, SD).

Note that the two types of weighted mean are only available for those parameters which come from the cnd sheets
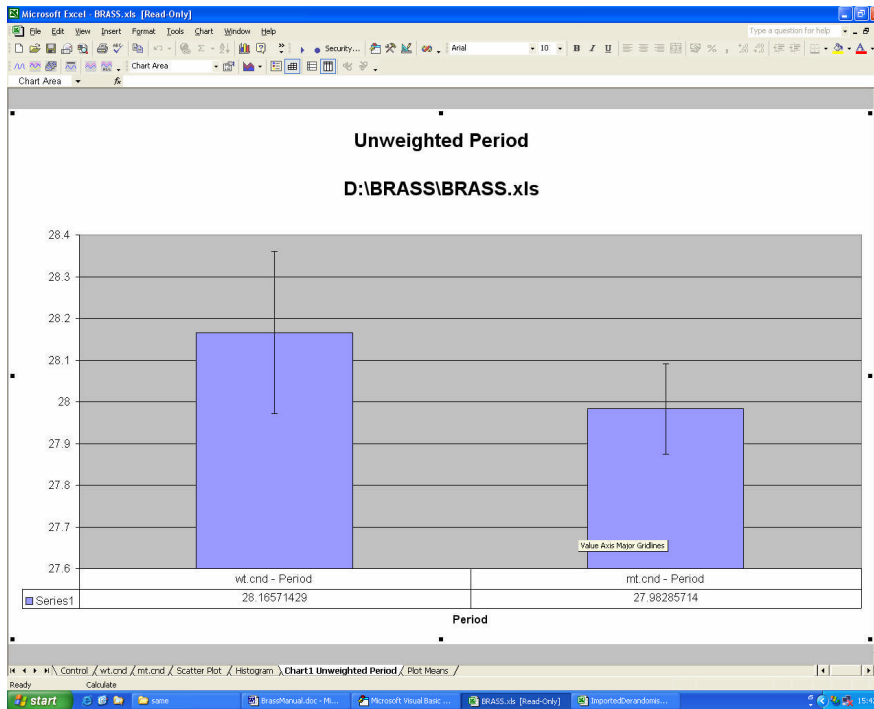
3. **Scatter Plots**. The "Scatter Plot" sheet contains drop down list boxes to enable the user to specify parameters for the X and Y axes. These can be any of the parameters from the cnd sheets: Period, Amplitude, Phase, Rel Amp, Rel Per, Rel Phase or AIC. Different condense files (genotypes) are plotted as different series.
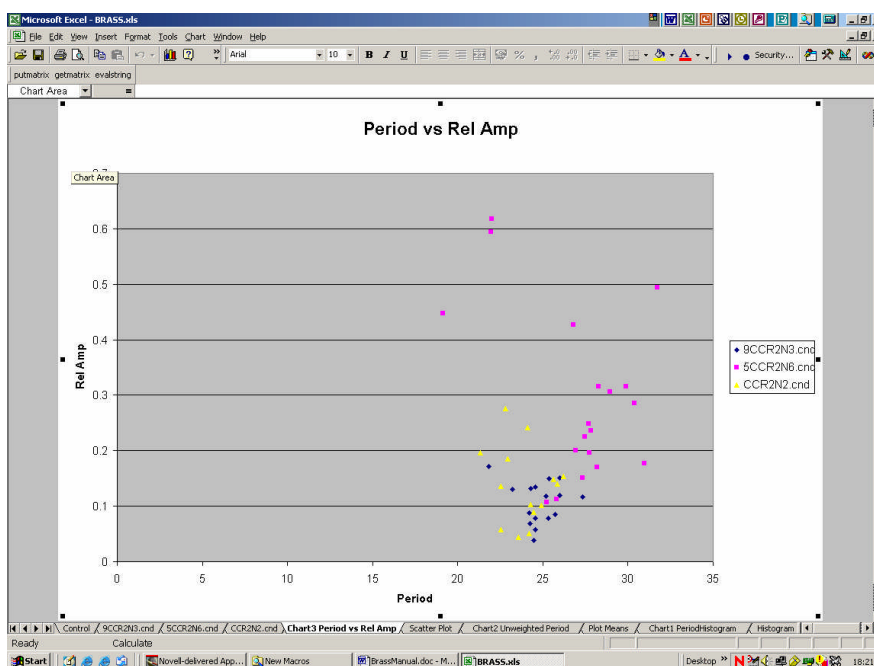
Note that some data series will have more than one period fitted by FFT NLLS. In this case only the primary period is plotted for each data series, ie that which appears first in the cnd sheet. Parameters from phs and waveform sheets are not available for scatter plots.

Note that for all of these charts the data used to create them is copied to an invisible worksheet. This ensures that they will not be affected when the sheet that created them is updated by the user when drawing another chart, for example re-generating statistics or performing a phase sort (page 55).

.

Exporting Results

Brass contains a function to export all the period, amplitude and phase values from time series on each cnd sheet in a single worksheet. Open the desired book and select **Export cnd Results** from the BRASS menu. This will create a new worksheet in the book called "Exported Values" containing all the data in a single table. This can them be exported as a text file as a convenient form to be read into another program.

Appendix A – Format of Import Data Sheets

This appendix contains examples of each of the raw data sheets recognised by BRASS. If sheets do not conform to these standards then they may not be imported correctly. In the case of Metamorph user can define the precise layout of the worksheets created. Alternatively it should be possible to use the **Custom Import** functions.

TopCount Files

These are tab delimited MS-DOS files. When a single file representing one timepoint is opened in Excel, using the default import options, it is essential that the data should appear in the following cells

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		1	2	3	4	5	6	7	8	9	10	11	12
2	A	3	8	5	3	3	5	0	5	3	5	3	0
3	B	3	5	3	3	3	5	3	3	3	5	3	5
4	C	5	5	3	3	3	3	0	3	5	5	0	0
5	D	3	5	5	5	3	5	3	5	3	5	3	8
6	E	0	3	3	5	3	5	5	3	3	5	3	3
7	F	5	5	5	5	5	3	0	3	3	3	3	3
8	G	3	5	5	3	0	3	3	3	3	3	3	3
9	H	3	3	5	5	5	5	0	3	3	5	3	3
10													
11													
12		1	2	3	4	5	6	7	8	9	10	11	12
13	A	10:26:10 PM	10:26:20 PM	10:26:10 PM	10:26:20 PM	10:26:10 PM	10:26:20 PM	10:26:10 PM	10:26:20 PM	10:26:10 PM	10:26:20 PM	10:26:10 PM	10:26:20 PM
14	B	10:26:15 PM	10:26:25 PM	10:26:15 PM	10:26:25 PM	10:26:15 PM	10:26:25 PM	10:26:15 PM	10:26:25 PM	10:26:15 PM	10:26:25 PM	10:26:15 PM	10:26:25 PM
15	C	10:26:20 PM	10:26:30 PM	10:26:20 PM	10:26:30 PM	10:26:20 PM	10:26:30 PM	10:26:20 PM	10:26:30 PM	10:26:20 PM	10:26:30 PM	10:26:20 PM	10:26:30 PM
16	D	10:26:25 PM	10:26:35 PM	10:26:25 PM	10:26:35 PM	10:26:25 PM	10:26:35 PM	10:26:25 PM	10:26:35 PM	10:26:25 PM	10:26:35 PM	10:26:25 PM	10:26:35 PM
17	E	10:26:30 PM	10:26:40 PM	10:26:30 PM	10:26:40 PM	10:26:30 PM	10:26:40 PM	10:26:30 PM	10:26:40 PM	10:26:30 PM	10:26:40 PM	10:26:30 PM	10:26:40 PM
18	F	10:26:35 PM	10:26:45 PM	10:26:35 PM	10:26:45 PM	10:26:35 PM	10:26:45 PM	10:26:35 PM	10:26:45 PM	10:26:35 PM	10:26:45 PM	10:26:35 PM	10:26:45 PM
19	G	10:26:40 PM	10:26:50 PM	10:26:40 PM	10:26:50 PM	10:26:40 PM	10:26:50 PM	10:26:40 PM	10:26:50 PM	10:26:40 PM	10:26:50 PM	10:26:40 PM	10:26:50 PM
20	H	10:26:45 PM	10:26:55 PM	10:26:45 PM	10:26:55 PM	10:26:45 PM	10:26:55 PM	10:26:45 PM	10:26:55 PM	10:26:45 PM	10:26:55 PM	10:26:45 PM	10:26:55 PM
21													

If the files are combined in MS-DOS, then they should appear as below, with each time point below the previous one. Note that there is only one blank row, in this case row 21, between consecutive data sets.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		1	2	3	4	5	6	7	8	9	10	11	12
2	A	218	2896	230	1931	1701	308	479	144	296	109	4	4
3	B	2173	9961	628	785	2678	1627	195	82	768	1380	4	4
4	C	1125	3048	460	1647	3923	10807	222	51	2214	456	8	4
5	D	1737	6745	423	1290	2970	3779	8	8	2499	518	4	4
6	E	3011	5286	727	1343	937	838	226	8	82	339	8	4
7	F	1035	337	226	127	415	850	172	97	74	172	8	4
8	G	5377	5069	431	2547	752	797	55	164	164	27	12	12
9	H	2165	596	337	242	1844	3914	8	261	444	136	4	4
10													
11													
12		1	2	3	4	5	6	7	8	9	10	11	12
13	A	12:09:57 PM	12:11:00 PM	12:11:04 PM	12:12:07 PM	12:12:12 PM	12:13:15 PM	12:09:57 PM	12:11:00 PM	12:11:04 PM	12:12:07 PM	12:12:12 PM	12:13:15 PM
14	B	12:10:01 PM	12:10:55 PM	12:11:08 PM	12:12:03 PM	12:12:16 PM	12:13:10 PM	12:10:01 PM	12:10:55 PM	12:11:08 PM	12:12:03 PM	12:12:16 PM	12:13:10 PM
15	C	12:10:05 PM	12:10:51 PM	12:11:12 PM	12:11:59 PM	12:12:20 PM	12:13:06 PM	12:10:05 PM	12:10:51 PM	12:11:12 PM	12:11:59 PM	12:12:20 PM	12:13:06 PM
16	D	12:10:09 PM	12:10:47 PM	12:11:16 PM	12:11:54 PM	12:12:24 PM	12:13:02 PM	12:10:09 PM	12:10:47 PM	12:11:16 PM	12:11:54 PM	12:12:24 PM	12:13:02 PM
17	E	12:10:13 PM	12:10:43 PM	12:11:21 PM	12:11:50 PM	12:12:28 PM	12:12:58 PM	12:10:13 PM	12:10:43 PM	12:11:21 PM	12:11:50 PM	12:12:28 PM	12:12:58 PM
18	F	12:10:17 PM	12:10:39 PM	12:11:25 PM	12:11:46 PM	12:12:32 PM	12:12:53 PM	12:10:17 PM	12:10:39 PM	12:11:25 PM	12:11:46 PM	12:12:32 PM	12:12:53 PM
19	G	12:10:21 PM	12:10:34 PM	12:11:29 PM	12:11:42 PM	12:12:36 PM	12:12:49 PM	12:10:21 PM	12:10:34 PM	12:11:29 PM	12:11:42 PM	12:12:36 PM	12:12:49 PM
20	H	12:10:25 PM	12:10:30 PM	12:11:33 PM	12:11:38 PM	12:12:40 PM	12:12:45 PM	12:10:25 PM	12:10:30 PM	12:11:33 PM	12:11:38 PM	12:12:40 PM	12:12:45 PM
21													
22		1	2	3	4	5	6	7	8	9	10	11	12
23	A	230	2534	230	1754	1380	296	448	109	257	109	4	8
24	B	2144	10910	596	743	2280	1446	183	70	557	1165	8	4
25	C	723	2530	390	1368	3993	10819	203	66	1789	366	12	4
26	D	1705	6638	357	1171	2888	3762	12	4	2043	413	8	4
27	E	2966	5052	665	1294	945	653	222	8	78	292	8	4
28	F	969	341	205	123	370	727	144	78	78	144	4	4
29	G	4847	5167	440	2444	686	731	47	156	148	35	8	8
30	H	1951	600	288	205	1746	3849	8	210	398	97	4	4
31													

Night Owl Data Sheet

A Night Owl workbook contains a single worksheet, which contains data for 90 regions from a 10 * 9 experimental setup, in the following format

Time values are not read from these file, but, as in "I&A" are generated by the import routines.

Kujata Y Files

Y files are space delimited text files. Column one consists of time values, in seconds from start of experiment, and column two consists of values for pixel position of the leaf. Values of '-1' appear for timepoints when the leaf could not be located.
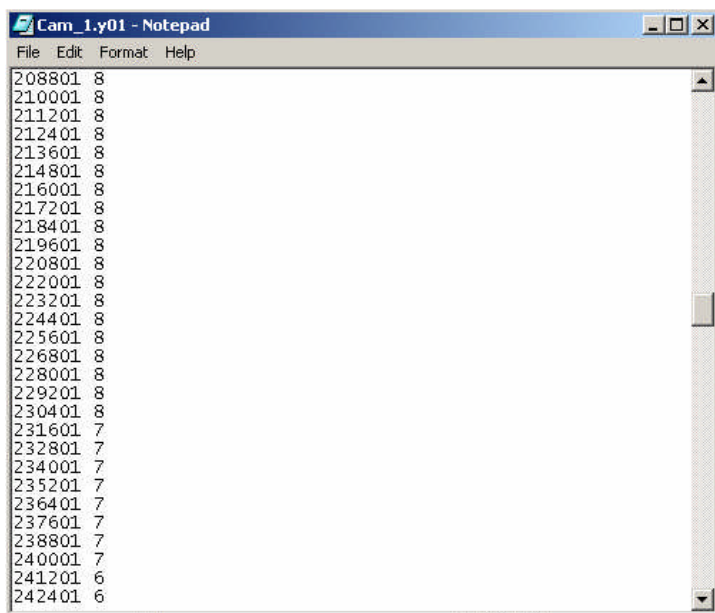
Metamorph Luminescence Data Sheet

Metamorph allows the user to specify the format of the workbook created. This is done as follows (specifically in Metamorph 4.5).

Build a stack and draw regions as usual, and save your regions. Next, select the **Log** menu and open **Initiate dynamic data exchange with Excel**. Then select **Open data log**. Tick the "Dynamic data exchange" box and click OK. By the "Application" tab, select "Microsoft Excel 97", and again hit OK. Next, select the **Apps** menu and the **Option graph intensities**. Make sure in the "Measure from" frame, "Stack" is selected, in the "Measure regions over" frame "Plane number" is selected, and in the "Region measurements", "Integrated" is selected. Then click on the "Configure log" button, and make sure only "Image name", "Image plane", "Image date and time", "Region name" and "Integrated" are selected. Click OK, check the "Log data" box is ticked, and click the "Begin" button. This should generate an Excel sheet with the following format.

Experiment name

Time of day, in fractions of a day. Although here the cells are formatted to display time values

Continuous block of data begins in this cell begins in this cell. Each column is one region

Region names in row one

Column two, 'Image Plane', and column 4, 'Region Name', are not read by BRASS.

Metamorph Leaf Movement Data Sheet

These are specified in a similar way to the process described above, and should consist of the following format.

Image Plane

Time of day, in fractions of a day. Although here the cells are formatted to display time

Region names

X values   Y values

Image Name

Appendix B - Verifying Results

As with all new software, BRASS should be tested thoroughly. I have done this myself with test data, but the most demanding test will always be when users process real experimental data. I have designed the format of the output workbooks and condense files to make verification as easy as possible. The following refers to analysis performed with FFT NLLS.

Single ZT Windows

1. Running FFT NLLS produces a workbook for each group (genotype) of worksheets that was exported, containing a worksheet for each worksheet in the export group. This will be called *Genotype*.xls. The selected data ranges on the newly created sheets should be identical to the corresponding cells on the corresponding worksheets of the original workbook. This includes the column headings identifying the genotypes and regions of each plant. Checking this indicates that the correct figures were analysed under the correct genotype name.

2. At the top of column headings of these new sheets will be a file name indicating the name of the file that this data was passed to `four-anl` as. This file name will include the name of the genotype of the plant, and the column the data was read from in the original sheet, which should be the same as the column the data now appears in. The individual cnd files for each of these plants can be opened (*Genotype\Sheetname\Column_Genotyp*.cnd) and the period corresponding to this file name can be found. Those cnd files with no periods found should have "No Period" or "FFT Failed" in their corresponding columns.

3. Opening the master cnd file for this genotype (*Genotype\Genotype*.cnd) will show that the contents of all the cnd files for all the plants in this group have been copied over. Headings indicate which *Sheetname* folder they came from.

Multiple ZT Windows

1. Running FFT NLLS produces a workbook for each group (genotype) of worksheets that was exported, containing a worksheet for each worksheet in the export group. This will be called *Genotype*.xls. The selected data ranges on the newly created sheets should be identical to the corresponding cells on the corresponding worksheets of the original workbook. This includes the column headings identifying

the genotypes and regions of each plant. Checking this indicates that the correct figures were analysed under the correct genotype name.

2. At the top of column headings of these new sheets will be a file name indicating the name of the file that this data was passed to `four-anl` as. This file name will include the name of the genotype of the plant, and the extension will be the starting ZT time of the data in the column below. The name will begin with the column the data was read from in the original sheet, which should be the same as the column the data now appears in. If they have not been automatically deleted, the individual cnd files for each of these plants can be opened (*Genotype\Sheetname\Column_Genotype_StartTime_StopTime*.cnd) and the period corresponding to this file name can be found. There will also be a series of files for this plant for each time window.

3. Opening the master cnd file for this plant (*Genotype\Sheetname\Column_Region_Genotype*.cnd) shows that all these files have been copied. The value with the lowest Rel Amp can be noted. It should correspond to the ZT window of the data on the new worksheet.

4. Opening the master cnd file for this genotype (*Genotype\Genotype*.cnd) will show that it is this period value which has been picked out for this plant.

5. The "Report" sheet of the workbook *Genotype*.xls will indicate which time windows produced valid periods for each plant, and which did not. These can be checked against the master cnd files for each plant (*Genotype\Sheetname\Column_Region_Genotype*.cnd). Those time windows with no period values in the cnd file should be listed in the "Report" sheet with an explanation. Plants which have "Invalid Data" or "No Period" or "FFT Failed" should have no period values in their cnd files for any time window.

Appendix C - Error Handling

All logic errors cause a form to be displayed with a cause and suggestions. This is the same form as used by "I&A". Errors will mostly be caused by worksheets being of the wrong format for the function being used. If they persist and the cause is not obvious, email me at p.e.brown@warwick.ac.uk with the error description and the data file you were using.

Code number generated by Excel

Description generated by Excel

Suggestions generated by BRASS
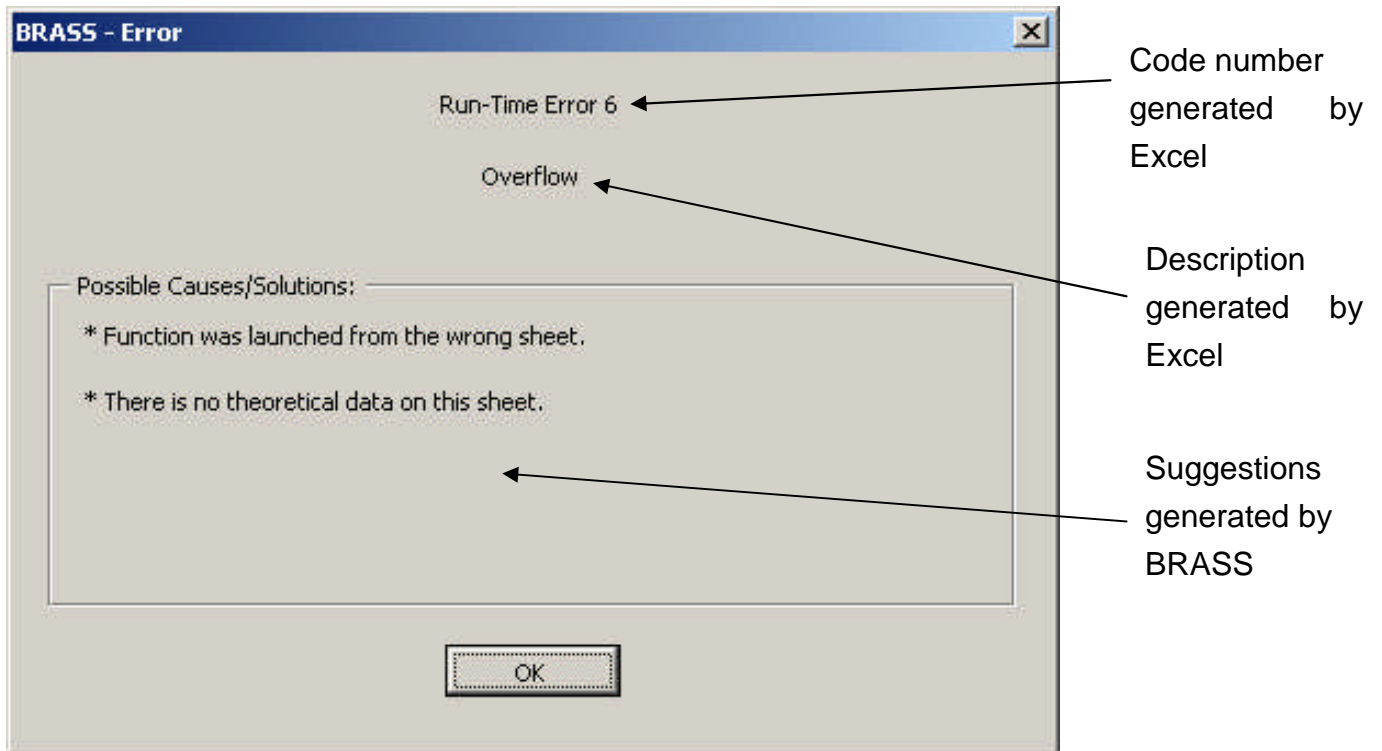
Clicking "OK" causes all macros to be exited, so whatever process was taking place when the error occurred must be begun again from the start.

Appendix D - Limitations of BRASS

Users should be aware of the following points:

- Never re-name any of the worksheets in BRASS, or any of the controls found on them.

- Don't change from Excel to another application while FFT NLLS is running. This may break the communication between it and Excel, and Excel may crash.

- There must be no empty columns in the middle of the data in a worksheet that is to be exported, as when the routines come to an empty column they will think that the data has ended.

- The same applies when using the Assign Genotypes form.

- When BRASS automatically prompts the user to save a workbook, the user clicks OK, then Windows gives a warning that a file with the same name exists, the user may elect not to save. This causes BRASS to produce an error "Method 'SaveAs' of object '_Workbook' failed. This can just be ignored. The book will not have been saved.

- Avoid the characters ;, : . \ / ? * [ ] < > | - & " and the space character when naming worksheets, regions or genotypes. These can cause problems when used in a worksheet, folder or file name. BRASS will remove them.

- The length of a worksheet name is limited to 31 characters by Excel. Similarly, any file processed by FFT NLLS is limited to 8 characters. BRASS will truncate any names which are longer than the permitted length.

- When creating new folders during data export, if a folder with the intended name already exists, then the new folder will have a '0' appended to its name. If this name exists too then '01' , then '012' will be appended and so on. Similarly, if a cnd or phs file that is being imported has the same name as a pre-existing worksheet, the new sheet will have a '1' appended to its name.

- In order for the export functions to work correctly, all spreadsheet cells containing data readings must be in number format rather than text. This should always be the case when using data from Topcount, Metamorph, Kujata or Night Owl. However, it may not be when using **Custom Import** to import data from other sources. In the imported data

sheets created by BRASS numeric data should appear right-aligned, whereas textual values will be left-aligned. The easiest way to fix this is to highlight the cells and go to **Format->Cells**, go to the "Number" tab and select "Number" from the list. I have, however, noticed a problem with Excel 2000 and XP where this sometimes does not work until the user selects every cell individually, clicks the cursor in the formula bar and then presses the <Return> key.

- When copying data into a spreadsheet from a custom text file, ensure that all text and numbers appearing on the same line are properly delimited so that they are imported into separate cells. For example, if your file contains the line

```
Time:10.5
```

this may be imported as a text string into a single cell, making the numerical value for time inaccessible. This can be fixed by inserting a TAB character immediately before the value which will ensure that 'Time:' and '50' will go into separate cells.

Appendix E – mFourfit

mFourfit works by attempting to fit every period from the lowest to the highest specified by the user, in increments of 0.01 hours and returning the one that gives the best fit. This is done as follows:

1. Start by generating a set of 5 sine and 5 cosine waves covering the time window of the experimental data. Formulas used are

$$\cos\left(\frac{2\cdot\pi\cdot t\cdot n}{\tau}\right) \quad \text{and} \quad \sin\left(\frac{2\cdot\pi\cdot t\cdot n}{\tau}\right)$$

where t is time, τ is period, and n is 1 to 5.

This produces the following waves if Period = 24
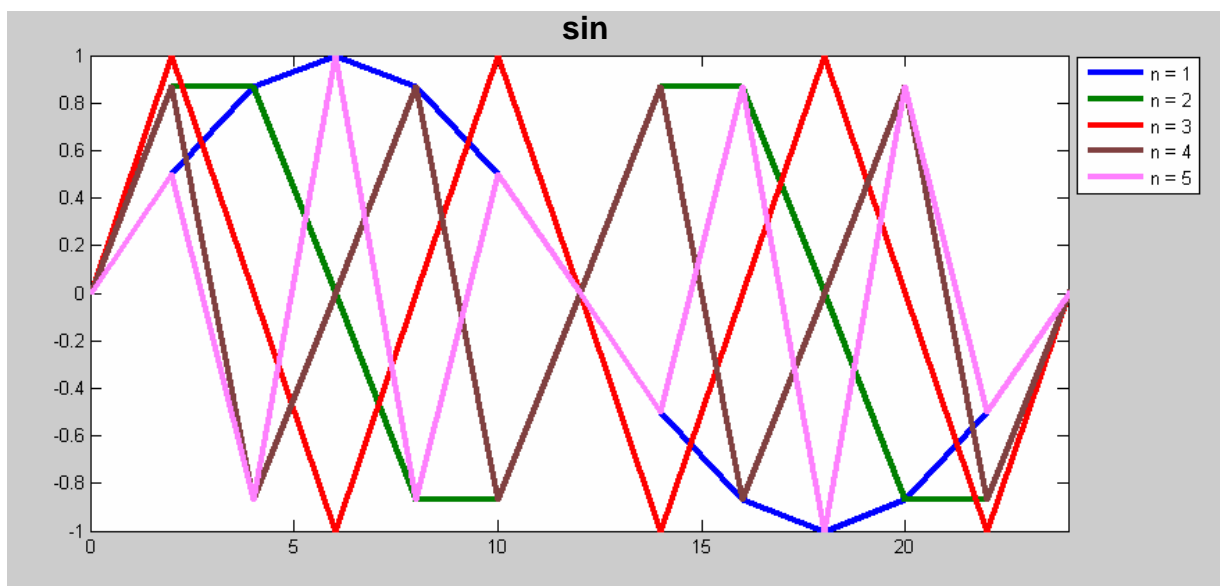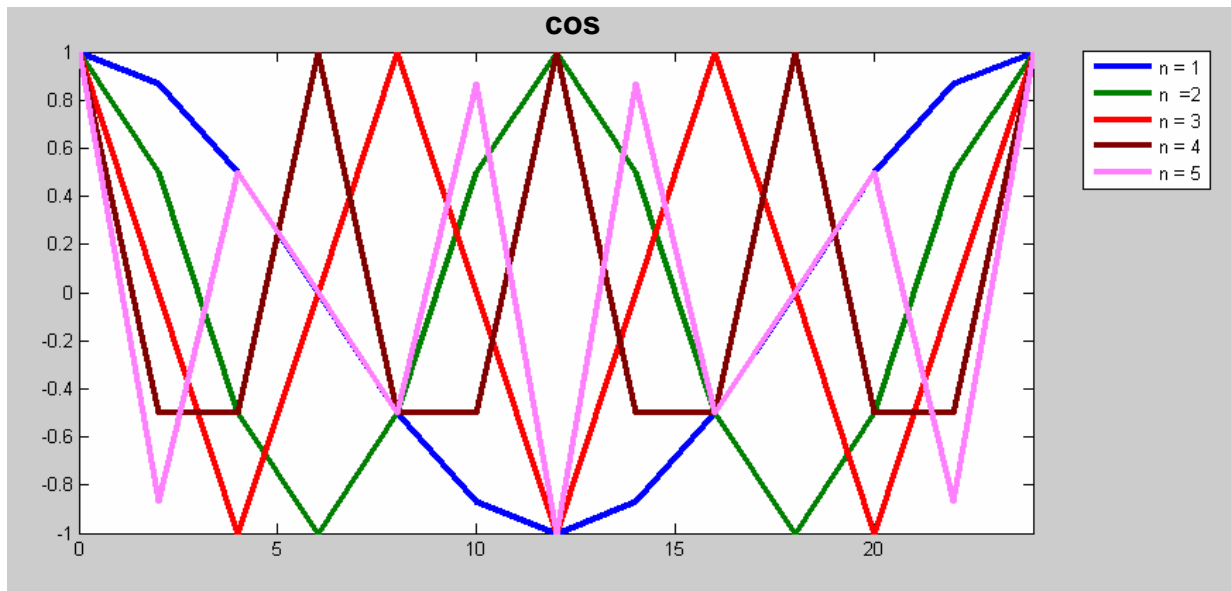
2. These waves, together with a flat line of value 1 for all timepoints (the offset) form the matrix A, which is nt * 11, where nt is the number of timepoints. The experimental data forms the vector b, which is nt * 1. These are passed to the Matlab function `lscov()` (ordinary least squares covariance).

   `[x,stdx,mse,S] = lscov(A,b)` returns the 11 * 1 vector x, such that A*x = b. Put simply, this means that when the sine and cosine components are weighted by the corresponding elements of x at each timepoint, this minimises the difference between them and the real data.

   mse is a scalar quantity, the mean squared error.

   S is an 11 * 11 matrix, the estimated covariance of x.

3. A theoretical time series is then generated by this weighting. The sum of squares difference (SS) is calculated between this and the real data.

4. The next step is to improve the fit by removing components which don't have significant weighting. The covariance (c) of the component n is taken as element (n,n) of matrix S divided by mse and the following value is calculated

$$\sqrt{c \cdot \frac{SS}{nt}}$$

   where nt is the number of time points.

   If this value is greater than or equal to half the value of the weight for this component, then this component is deemed to be not significant and its weight is set to zero, so eliminating this component from the theoretical time series.

5. The sum of squares (SS) is then recalculated from the new theoretical time series, and the likelihood of the fit calculated according to the formula

$$\frac{-nt \cdot \log\left(\frac{SS}{nt}\right)}{2}$$

   A smaller (less negative) value indicates a better fit.

6. The Akaike information criterion (AIC), is then calculated by subtracting from the likelihood the number of significant components. This penalises solutions with lots of components. The smaller this value the better the fit. More information can be found on Wikipedia.

This whole process is repeated for all the periods, and the one with the smallest AIC value is the best fit.

Once the best period has been found, the real data is analysed for trends in baseline and rhythm amplitude. This is done as follows

1. A regression line is fitted though the raw data. This gives the baseline trend

2. The absolute difference between this trend line and the real data is plotted against time, and a regression line fitted through this data. Each data point in this line is doubled, to give the trend in amplitude.

3. The theoretical data is scaled to have a mean of zero and an amplitude of one. This line is then multiplied by the amplitude trend line, and then the baseline trend is added to the result. This produces a proper visual fit to the real data.

At this point, the theoretical fit has data points at time points corresponding to the real data. Before further analysis of phase and amplitude, an interpolation is performed to produce time points every 10 min. This can be done using either of the Matlab functions `spline()` or `interp1()`. For noisy data `spline()` gives a smoother fit.

Preliminary comparison of mFourfit and FFT-NLLS period and phase estimates

The mFourfit procedure was developed to estimate the phase of circadian rhythms under entraining conditions, where output markers can have complex waveforms. Quantitative comparisons have not been made exhaustively: this document gives indicative results to accompany the release of BRASS 3.0 in 2007.

Period estimates:
mFourfit (with spline interpolation) and FFT-NLLS methods showed reasonable agreement in average period estimates from sample data of Arabidopsis expressing a range of LUC reporters in two test experiments under constant light (Figures 1, 2).

Variance and bias. For strongly rhythmic reporters, the average period difference for each trace between methods was in the range 0.4 – 0.7 hours, which is less than ErrPer (i.e. it falls within the error of period estimation for each trace from FFT-NLLS). There is a tendency for mFourfit to return a wider range of period estimates, such that the

unweighted SD of period estimates for mFourfit was either equal to or greater than the SD for FFT-NLLS (line D in figure 2, for example). More significant mean differences were apparent for some reporters (I or J in Figure 1, for example).

These differences did not correlate with the weakest rhythms, though the reporters with highest RAE did show smaller differences in period estimates between methods. Their detailed provenance remains to be determined. We would expect either complexity in or variation of waveform to have a significant effect. FFT-NLLS will account for these non-stationary effects using low-period components that are ignored when reporting the circadian component, whereas mFourfit will include such effects in its single waveform. This might also account for the greater variability in mFourfit period estimates.

Arrhythmic traces. mFourfit is not restricted to fitting cosine waveforms to data and so can fit a broader range of rhythmic traces. However, mFourfit will always fit a waveform unless the programme fails completely, so this means that it will often fit periodic rhythms to arrhythmic data. In one example, rhythms in the circadian range were fitted to 71% of traces from an overtly arrhythmic marker by mFourfit, compared to 21% by FFT-NLLS. Thus, FFT-NLLS is a more useful method for classifying arrhythmic phenotypes, and the RAE *versus* period plots are a useful depiction of rhythmic robustness.

Figure 1: Mean periods in LL
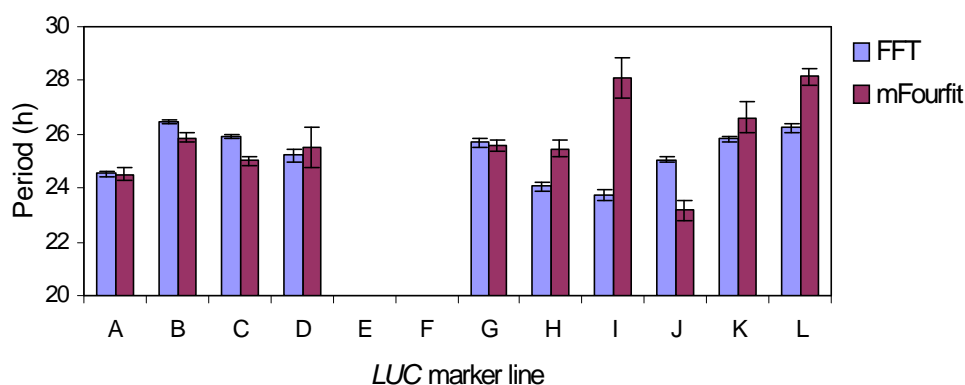
Figure 2: individual plant periods in LL
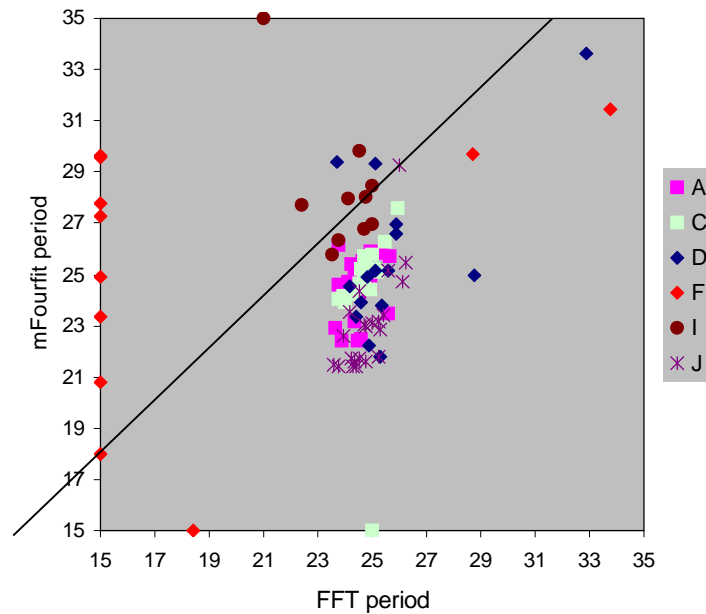
Comparison of FFT and mFourfit period estimates under constant light. Figure 1 shows mean periods for sets of 15 to 22 individual seedlings containing LUC reporter genes A to L. Error bars are one variance-weighted SEM (WtdSEM for FFT-NLLS, AIC WtdSEM for mFourfit). E and F are overtly arrhythmic lines, which are omitted. Figure 2 shows the individual period estimates contributing to Figure 1, though for clarity not all lines are shown. Data points at minimum values (15h period) represent traces that were only scored as rhythmic by one of the methods. This demonstrates mFourfit's ability to fit rhythms to arrhythmic traces (such as line F; Figure 2). The spread in Y tends to be greater than in X.

Phase:

The two methods show good agreement in phase estimates from plants under constant light (Figure 3) and a reasonable correlation for plants under light:dark cycles (Figure 4).

mFourfit is expected to have some benefits for phase estimates under entraining conditions, because the period of the fitted waveforms can be fixed to that of the external cycle. This simplification assumes that the plants are stably entrained. FFT-NLLS, in contrast, can confound phase estimation if the estimated period departs from the period of the entraining cycle. Under entraining conditions, mFourfit phase estimates show less variance than FFT phase estimates (Figure 4, for example marker lines B and I). However, a slightly larger proportion of plants failed to provide any rhythm analysis with mFourfit under LD cycles.

Figure 3: peak phase in LL
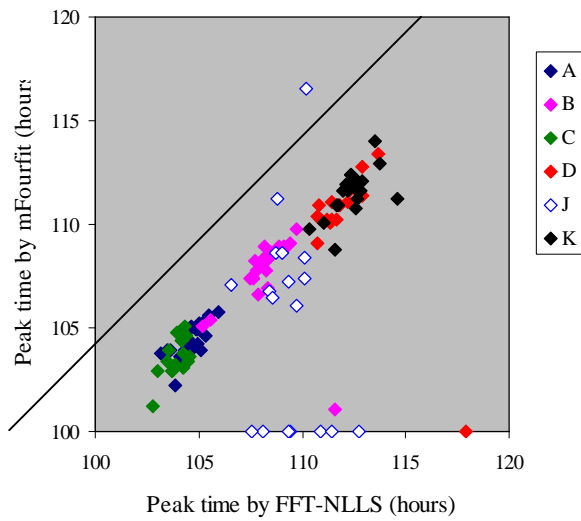
Figure 4: peak phase in LD
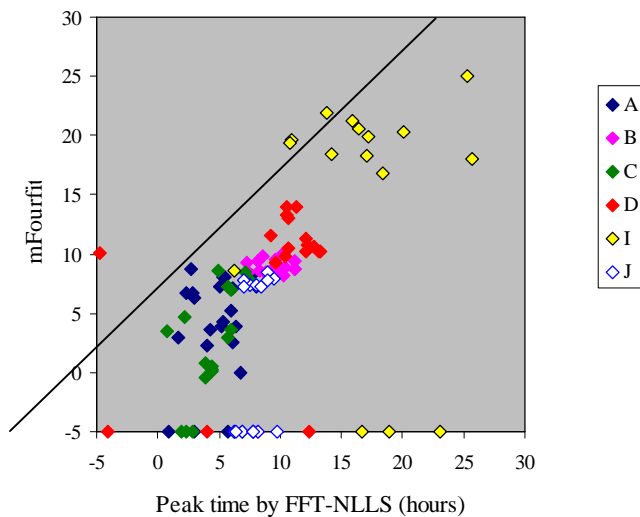
Comparison of FFT and mFourfit phase estimates under constant light (<u>Figure 3</u>) and during light:dark cycles (<u>Figure 4</u>). Samples that yielded no rhythmic fit (hence no phase estimate) by one method have a phase set manually to -5h. For clarity, not all marker lines are shown.

In conclusion, the methods differ in some respects and should therefore be explored in greater detail for specific applications. FFT-NLLS may be most useful for data from free-running conditions, for which it was developed. It has a clear advantage in identifying arrhythmic marker lines, with no apparent disadvantage in period estimation (and reduced

variance). mFourfit was developed for the analysis of rhythms under entraining conditions, where it gives reduced variance in the estimated phase for some marker lines.

Appendix F The Detrending Algorithm

The following is a description of how a typical time series data set is detrended for both amplitude of any oscillation, and baseline value.

Detrending amplitude

First fit a regression line through the data, we will call this R1, shown here in red. The experimental data is in blue.
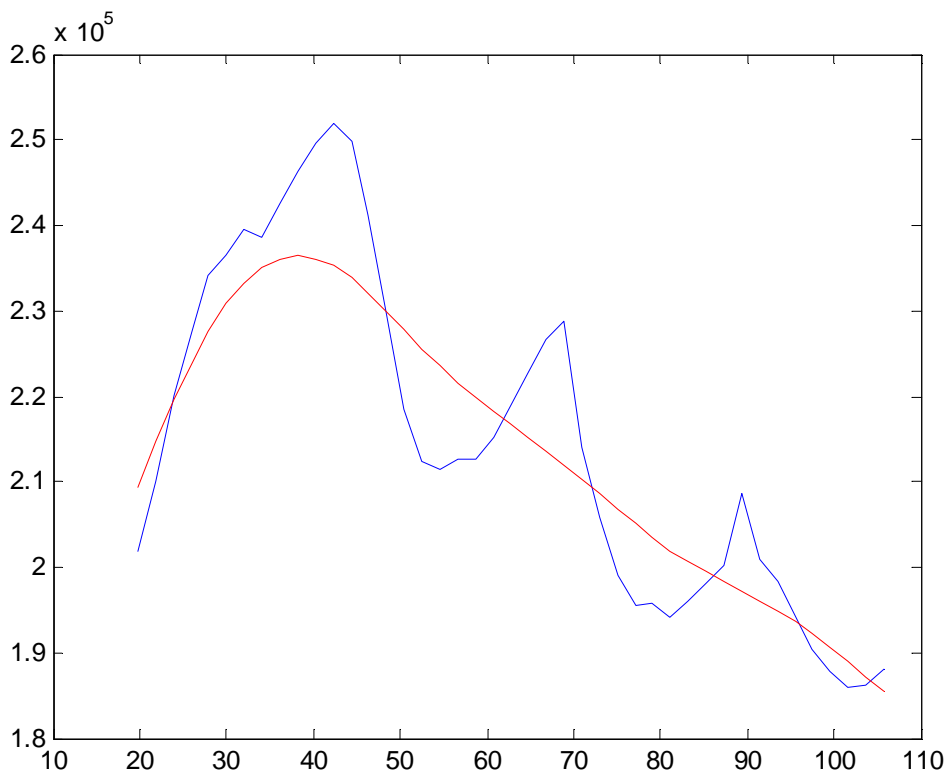
The regression line is a vector which at the ith point is the value of a linear regression put through the data weighted by a gaussian kernel with the kernel centred at the ith time. When this vector is plotted as a function it gives the trend of the time-series.

The regression uses the fact that the value of a and b that minimises
$chisq = sum_i \, w_i{}^*(x_i - a - b{}^*t_i)^2$ satisfies $aS + bSx = Sy$ and $aSx + bSxx = Sxy$,
where $S = sum_i \, w_i$, $Sx = sum_i \, w_i {}^* x_i$, $Sy = sum_i \, w_i{}^*y_i$, $Sxy = sum_i \, w_i {}^* x_i {}^* y_i$.
Here x is times and y is the data.

Subtract R1 from the data to give the curve below. We will call this curve x.
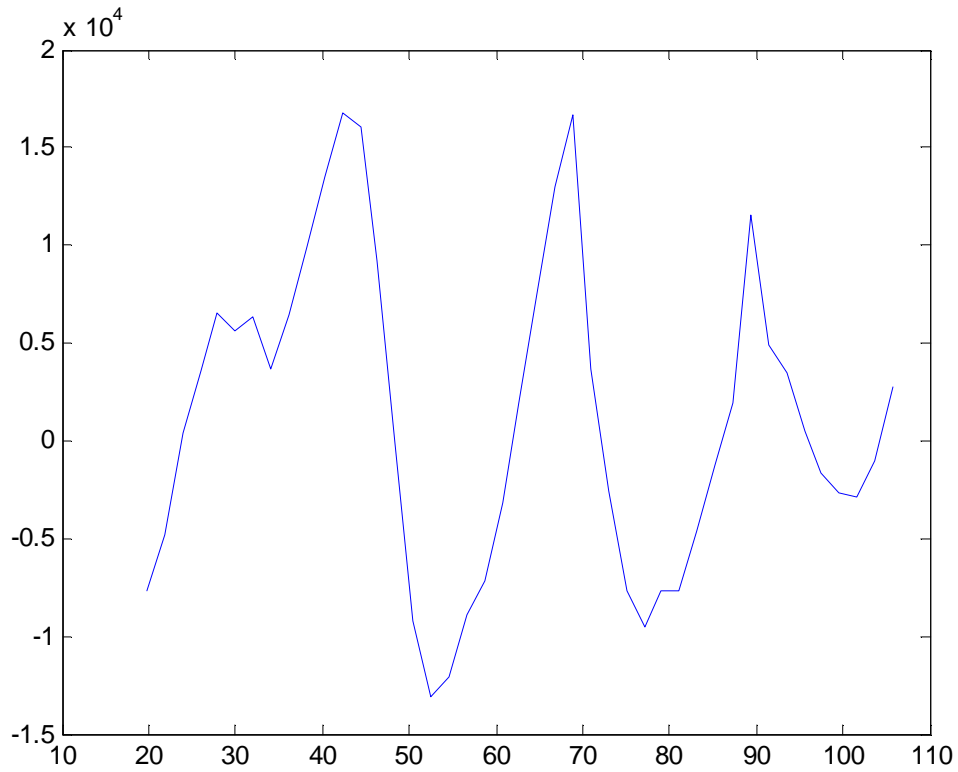
Fit a regression line through the absolute values of x. We will call this R2, shown here in green. The absolute values of x are shown in blue.
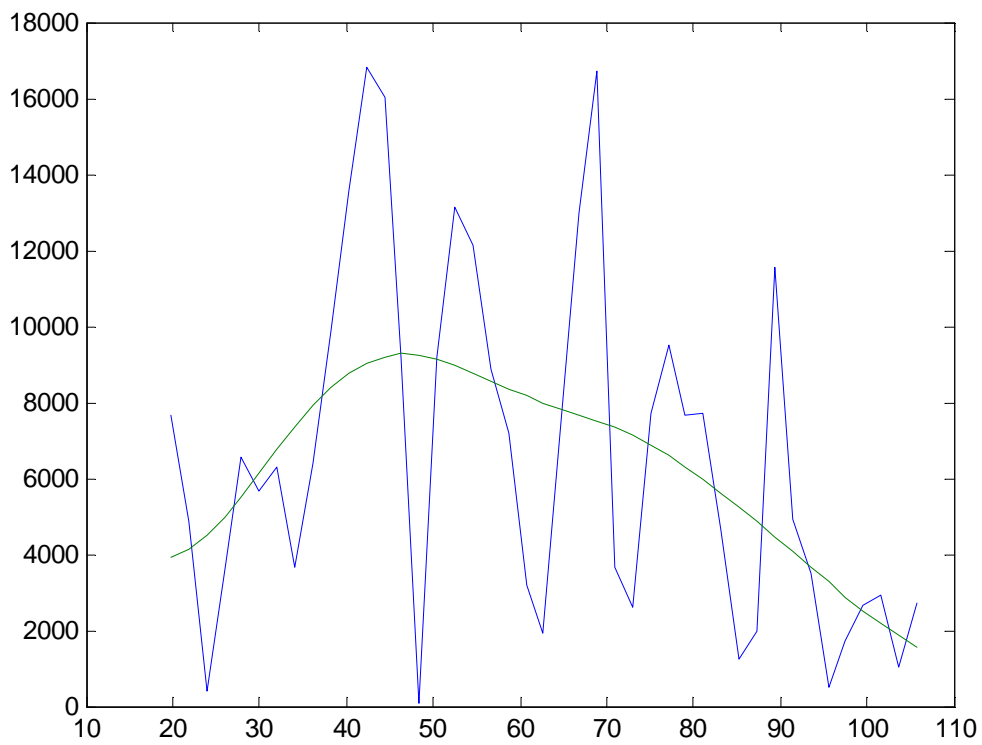
Divide the real values of x by R2, on a point by point basis. The results are shown below.
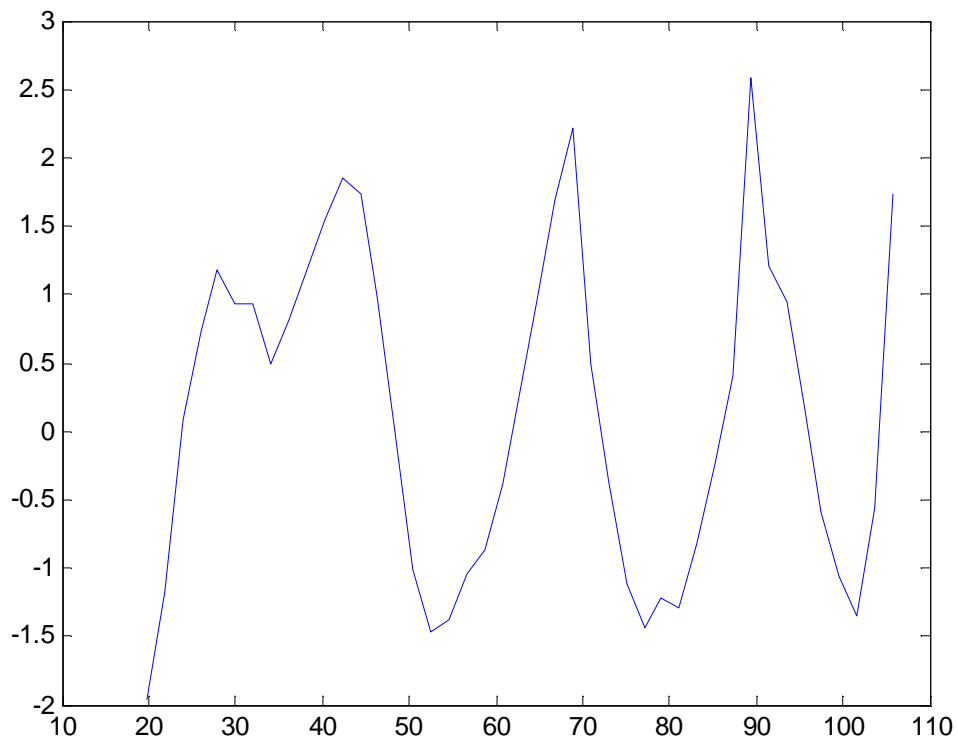
Multiply the result by the mean value of curve R2 to give the curve shown below.
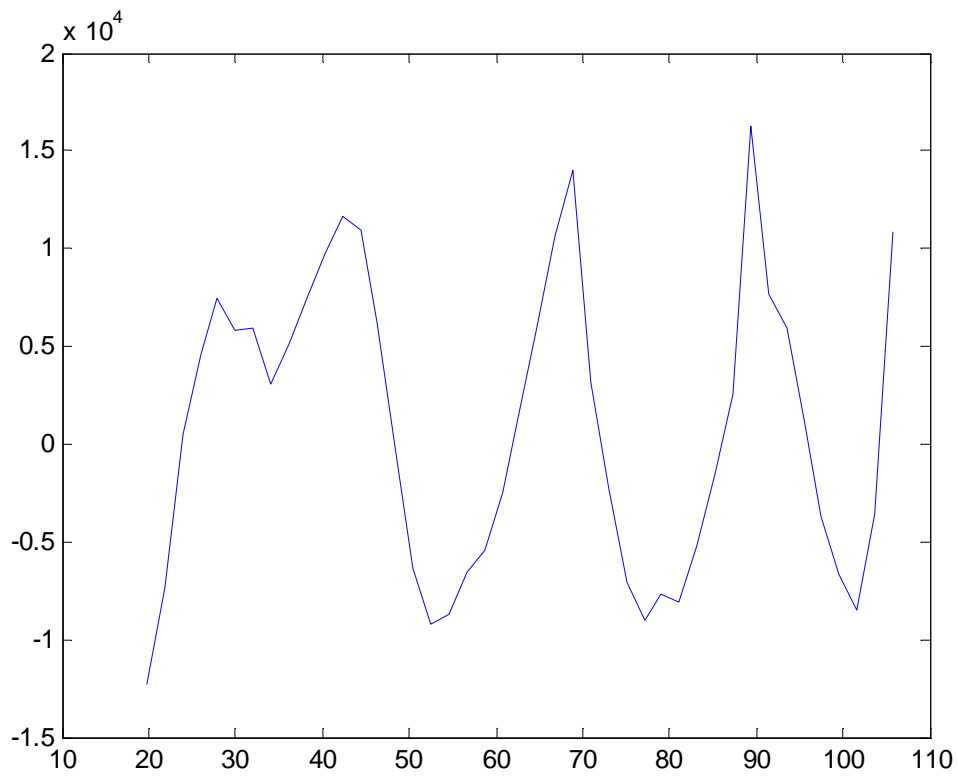
Finally, add the original regression line R1 to this curve to give the solid blue line shown below. This line has the trend in amplitude removed, but the baseline follows the experimental data, shown as the dashed line.
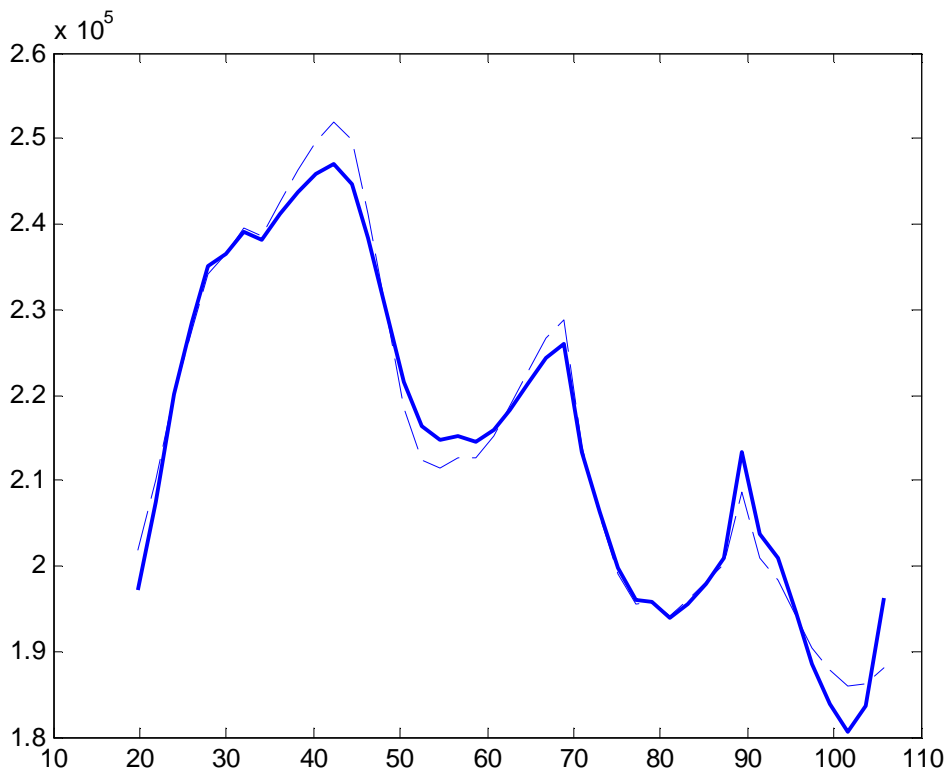
Next, we will remove the baseline trend from this line.

Detrending the Baseline

First, fit a regression line through the data. The regression line is shown in red.
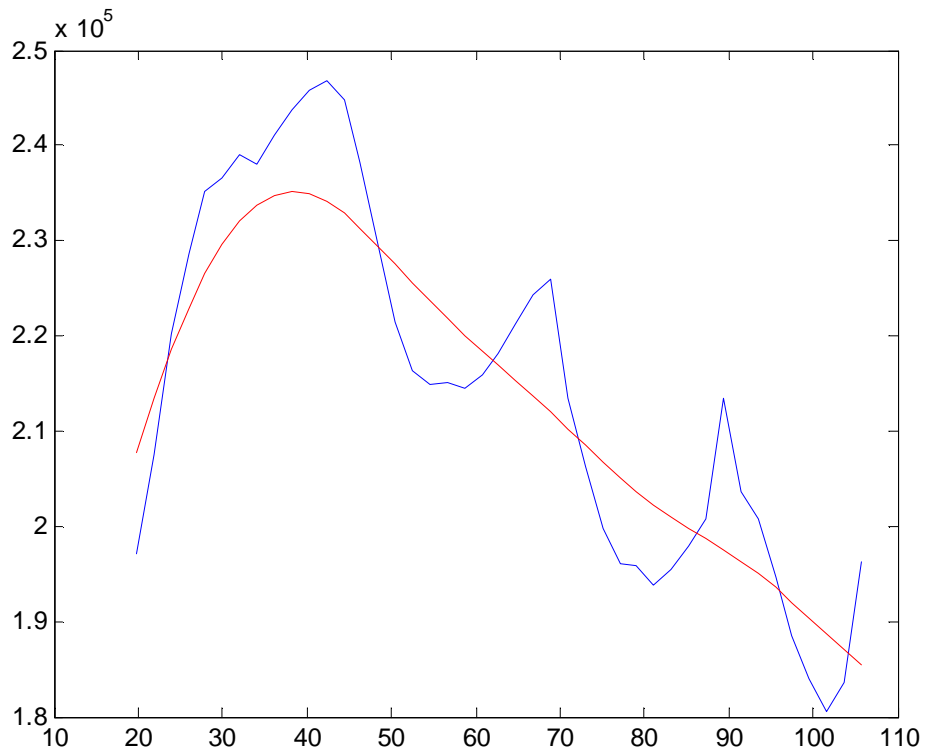
Now subtract the regression line from the data to give the curve below.
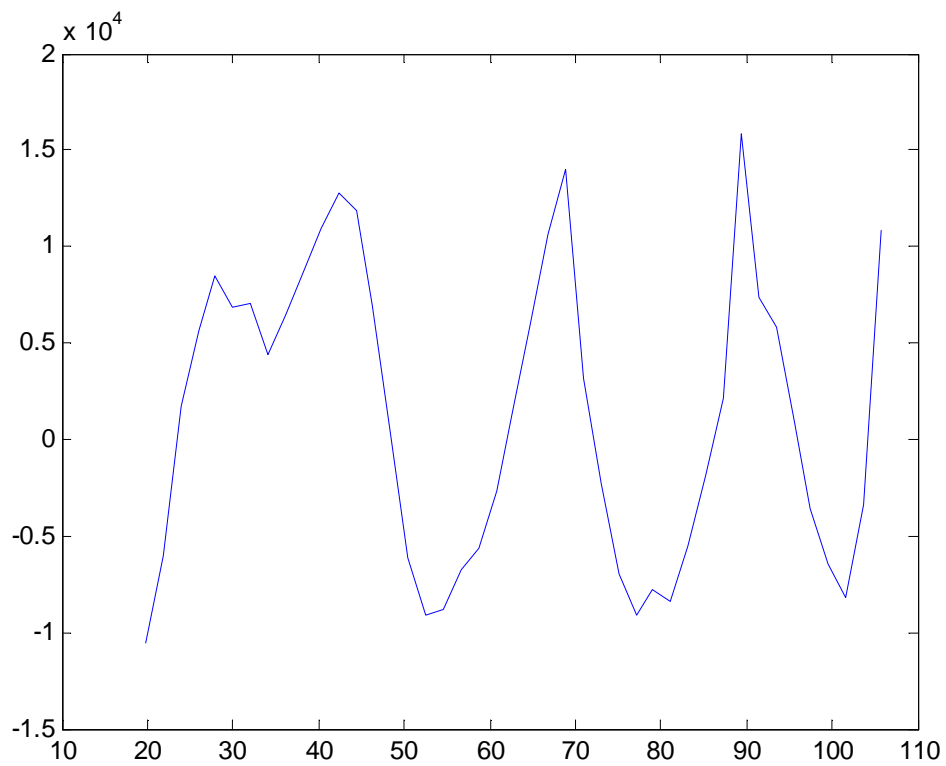
Finally, add the mean of the regression line to this curve to give the solid blue line below. The dashed line is the original data with the trend in amplitude removed.
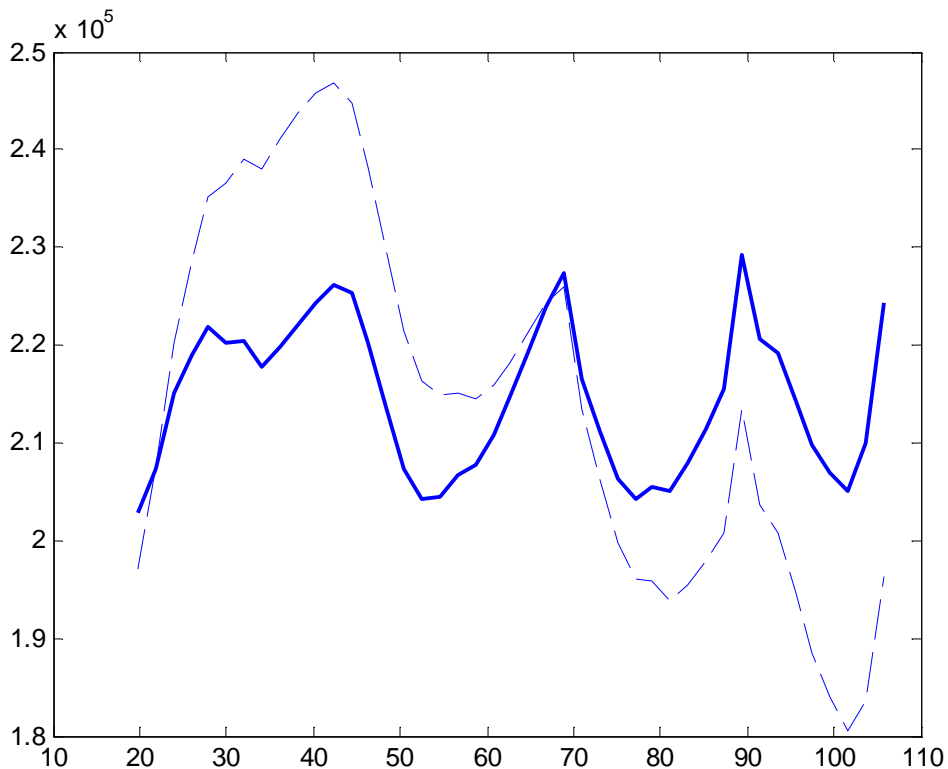
Of course the base line trend could have been removed first as the two processes are independent of one another.

Limitations of Detrending

The regression line fitting process must follow changes in the baseline value of a data series, but must not follow the circadian oscillation. Thus it must follow oscillations with a period much greater than 24 hours, but ignore oscillations with a period approaching this value. This is determined by the 'bandwidth' of the process. This has been optimised for a circadian period of 24 hours, ie the regression line accurately follows the baseline trend, but not any 24 hour oscillation. This may therefore cause inaccuracies when detrending mutant data with a cycle period very different from 24 hours. Future releases of BRASS will allow the user to fine tune the bandwidth to suit any expected cycle period.

Appendix G – FFT NLLS

BRASS works in conjunction with the FFT NLLS suite of programs to fit periods, phase and amplitudes to data. This consists of a set of three DOS programs; four-anl, fls-plot and condense. These were written by Marty Straume, formerly of University of Virginia. The following sections are his explanations of the curve fitting process, interpretation of the results, and user instructions for running these programs without the BRASS front end.

FFT-NLLS (Fast Fourier Transform – Nonlinear Least Squares)

Analyses conducted by FFT-NLLS (Fast Fourier Transform – Nonlinear Least Squares) conform to the above convention. Time series are first linear regression detrended by FFT-NLLS to produce zero-mean, zero-slope data. An FFT power spectrum is then calculated of the linear regression detrended data. The period, phase, and amplitude of the most powerful spectral peak are used to initialize a one-component cosine function (i.e., an order-1 fit, initially) of the form

$$y_{LR}(t) = c + \sum_{i=1}^{n} \alpha_i \cos\left[\frac{2\pi(t + \phi_i)}{\tau_i}\right]$$

which $y_{LR}(t)$ is the linear regression detrended time series to which analysis is being performed, c is a constant offset term, n is the order of fit, t is time, and α_i, ϕ_i, and τ_i are the amplitude, phase, and period, respectively, of the i-th cosine component. The parameters of this function are then estimated by nonlinear least squares minimization by a modified Gauss-Newton NLLS algorithm (Straume et al., 199 1; Johnson & Frasier, 1985). Upon convergence, approximate nonlinear asymmetric joint confidence limits are estimated for all parameters (period, phase, amplitude, and constant offset) at 95% confidence probability (Straume et al., 1 99 1). If the amplitude is significantly different from zero, then the procedure is repeated at the next higher order. The two most powerful FFT spectral peaks are then used to initialize a two-component cosine function (i.e., $n = 2$), which is subsequently NLLS minimized to the linear regression detrended data, and confidence limits are again evaluated. This process is repeated iteratively until at least one cosine component is identified with an amplitude that is not statistically significant. Alternatively, the procedure terminates if an error condition is encountered during computation (often an attempt at inversion of a near-singular matrix) or if the relative frequency difference between any two derived cosine components is within (1/N? t) (the approximate resolution limit supportable by the data, where N represents the number of

time points and Δt represents the sampling interval). Satisfying any of these termination criteria suggests that a model of too high an order is being attempted to fit the data. Penultimate results are then retained as the most statistically accurate quantitative characterization of the time series. In this way, the period, phase, and amplitude of the data are selected objectively and in fully automated manner with no requirement for user interaction (and, hence, no opportunity for influence of user-introduced bias).

The statistical significance of each derived rhythmic component is assessed by way of the relative amplitude error, defined as the ratio of, in the numerator, the amplitude error (one-half the difference between the upper minus the lower 95% amplitude confidence limits) to, in the denominator, the most probable derived amplitude magnitude. Theoretically, this metric will range from 0.0 to 1.0; 0.0 indicating a rhythmic component known to infinite precision (i.e., zero error), 1.0 (or greater) indicating a rhythm that is not statistically significant (i.e., error equal to (or exceeding) the most probable amplitude magnitude), and intermediate values indicative of varying degrees of rhythmic determination.

FFT-NLLS is specifically designed to process data sets that are relatively short and/or noisy and is generally capable of extracting relatively weak rhythms. Additionally, it extracts meaningful periods despite mean and variance nonstationarities that may exist in data. It assigns period, phase, amplitude, and rhythmic strength/level of rhythmic determination (via the relative amplitude error) in a totally automated manner that is free of user-introduced analytical bias.

Please be so kind as to cite the following three references when referring to the FFT-NLLS analysis in presentations to which it has contributed:

Primary reference
Plautz JD, Straume M, Stanewsky R, Jamison CF, Brandes C, Dowse HB, Hall JC, Kay SA (1997) Quantitative analysis of Drosophila period gene transcription in living animals. *J Biol Rhythms* 12: 204-217.

Secondary references
Straume M, Frasier-Cadoret SG, Johnson ML (1991) Least squares analysis of fluorescence data. In Topics in Fluorescence Spectroscopy, Volume 2: Principles, JR Lakowicz, ed, pp. 117-240. Plenum, New York.
Johnson ML, Frasier SG (1985) Nonlinear least squares analysis. *Methods Enzymol* 117: 301- 342.

Interpretation of FFT-NLLS results

The results reported by FFT-NLLS are the best-fit parameter values identified for each data set at the specified level of confidence probability. This last point is important! Reported results will vary depending on the confidence probability level employed during analysis. For valid comparisons of results of analysis of different data files, it is necessary to have identical conditions represented in the results of analysis (i.e., same numbers of data points per data set, same sampling frequency per data set, same confidence probability level used during analysis). The results offered by FFT-NLLS present the amplitude, period, and phase of all statistically significant rhythmic components that the FFT-NLLS algorithm could find at the specified level of confidence probability. Importantly, as such, the parameter error estimates produced by FFT-NLLS are not necessarily estimates of the standard errors of the parameters (unless a confidence probability value of 68.26% was used in analysis). The parameter error estimates produced by FFT-NLLS are estimates of parameter confidence at the user specified level of confidence probability, whether that is 95% or 90% or 99% or 68% or …

Considerations in assessment of phase

For purposes of interpretation, any assessment of an analytical determination of a phase angle reference is dependent on at least (i) exactly how time is represented in the data set being analyzed, (ii) any coordinate relationship between the endogenous rhythm and an exogenous periodicity of possible consequence, and (iii) the sign convention employed to represent phase advances versus delays thereof.

Phase angle reference is most commonly considered in terms of maximum, minimum, half-maximum, or half-minimum of some experimental observable.

Representation of time relates to how time zero is referred, because any analytical assessment of phase is implicitly relative to the absolute reference of time zero.

A coordinate relationship between exogenous and endogenous periodicities of consequence refers to the situation in which (for example, in the context of circadian rhythmicity) one is interested in a phase angle reference on a particular day (i.e., relative to an exactly 24-hour wall-time clock; the exogenous periodicity of the earth's rotation, day/night) of some circadian biological rhythm that has been analytically assessed to possess a period of oscillation that is not exactly 24-hours (i.e., a non-24-hour free-running period, for

example; the endogenous periodicity analytically assessed for the biological circadian rhythm).

The sign convention commonly employed in circadian biology refers to advances as positive (i.e., + or >0) and delays as negative (i.e., – or <0). When considering a relative situation comparing two rhythms, the concepts of advance and delay are straightforward. However, when considering an individual rhythm, the reference becomes absolute time zero; thus, in such a case, a positive phase means in advance of time zero (i.e., negative time) and a negative phase means delayed relative to time zero (i.e., positive time). This latter point can be a source of confusion.

Relative Amplitude Error (RAE) and rhythmic determination

The relative amplitude error, RAE, is the value of the amplitude error estimate divided by the value of the most probable amplitude estimate. RAE can range from a value of zero for an infinitely well-determined rhythmic component (zero error) to a value of one, theoretically, for a minimally-determined rhythmic component (error in the amplitude equals the amplitude value itself). There will be times when an RAE value greater than unity will be reported. Although theoretically not possible within the context of linear, symmetric models, the multi-component cosine model that underlies the FFT-NLLS analysis is neither. As such, RAE values greater than unity do sometimes occur (most often, in my experience, when a very-long-period component is identified during analysis to account for long-term drifting that is occurring in the data set being analyzed). In such cases, it is usually a rhythm that is not of interpretive interest, but rather is probably there to accommodate trend.

Interpretation of phase

The phase value is referenced to whatever time zero is in your data file's time axis. A positive phase value means advanced relative to zero (i.e., negative time) and a negative phase value means delayed relative to zero (i.e., positive time). Please note this relationship in defining phase so as to avoid confusion! The phase marker is cosine acrophase (i.e., peak of cosine wave) and it recurs modulo whatever the period for that rhythmic component is. So, be careful when comparing phase reference points that you do so correctly with respect to some absolute time marker, not just to what may be a quite-arbitrarily assigned zero-time-value in the time axis of the data files being analyzed. Also, please recognize that the phase referred to is that of a rhythmic component. In the event of multiple rhythmic components being identified in an analysis of a data file, the

acrophase of a rhythmic component may not always bear a discernable resemblance to an acrophase of the observed, composite rhythm (i.e., of the composite best-fit curve that may be composed of a linear combination of a number of contributing cosine waves of varying amplitudes, periods, and phase relationships).

Extrapolation of phase estimates

If the time of acrophase of an FFT-NLLS rhythmic component is desired at a time "distant" from absolute time zero of the data set under consideration, it must be calculated in modulo relative to its associated period.
Let me demonstrate by example to hopefully present most clearly what I mean to convey here. Let us assume in the examples that follow that absolute time zero refers to midnight wall-clock time.

1) Suppose an analysis by FFT-NLLS returns a statistically significant period estimate of exactly 24.0 hours, with a corresponding phase estimate reported as – 6.0 hours. The time of acrophase between the times of [(–24.0/2)= –12.0] and [(+24.0/2)=12.0] is at t=6.0 hours [i.e., –(–6.0)=6.0], thus $f\,max,orig$ = 6.0 hours. If one is interested in knowing at what wall-clock time the acrophase occurs on the fourth day after time zero as represented in the data set under consideration, the time range of interest becomes 72 = t = 96. To accommodate this situation, one must add values of the derived period estimate (i.e., 24.0, in this case) to the time of acrophase (i.e., 6.0, in this case) until a time of acrophase is produced that is within the desired time range (i.e., 72 = t = 96, in this case). Since the period estimate in this case is exactly 24.0 hours, the result becomes $fmax,extrap$ = 78.0 hours, which corresponds to $f\,max,extrap$ = 6.0 hours relative to the previous midnight.

2) Suppose an analysis by FFT-NLLS returns a statistically significant period estimate of exactly 23.0 hours, with a corresponding phase estimate reported as –6.0 hours. The time of acrophase between the times of [(–23.0/2)= –11.5] and [(+23.0/2)=1 1.5] is at t=6.0 hours [i.e., –(–6.0)=6.0], thus $f\,max,orig$ = 6.0 hours. If one is interested in knowing at what wall-clock time the acrophase occurs on the fourth day after time zero as represented in the data set under consideration, the time range of interest becomes 72 = t = 96. To accommodate this situation, one must add values of the derived period estimate (i.e., 23.0, in this case) to the time of acrophase (i.e., 6.0, in this case) until a time of acrophase is produced that is within the desired time range (i.e., 72 = t = 96, in this case). Since the period estimate in this case is exactly 23.0 hours, the result becomes $fmax,extrap$ = 75.0 hours, which

corresponds to *f max,extrap* = 3.0 hours relative to the previous midnight.

3) Suppose an analysis by FFT-NLLS returns a statistically significant period estimate of exactly 25.0 hours, with a corresponding phase estimate reported as –6.0 hours. The time of acrophase between the times of [(–25.0/2)= –12.5] and [(+25.0/2)=12.5] is at t=6.0 hours [i.e., –(–6.0)=6.0], thus *f max,orig* = 6.0 hours. If one is interested in knowing at what wall-clock time the acrophase occurs on the fourth day after time zero as represented in the data set under consideration, the time range of interest becomes 72 = t = 96. To accommodate this situation, one must add values of the derived period estimate (i.e., 25.0, in this case) to the time of acrophase (i.e., 6.0, in this case) until a time of acrophase is produced that is within the desired time range (i.e., 72 = t = 96, in this case). Since the period estimate in this case is exactly 25.0 hours, the result becomes *fmax,extrap* = 81.0 hours, which corresponds to *f max,extrap* = 9.0 hours relative to the previous midnight.

Exactly the same process will apply regardless of the sign of the phase. I.e., if it's positive, then its negation becomes negative, indicating that acrophase occurred in the negative time domain between the times of (–t/2) and (t/2) (i.e., *f max,orig* < 0). Then, if a phase optimum is desired on some day other than the day nearest t=0, the relevant range of times must be determined after which exactly the same extrapolation procedure is employed, modulo t. Of course, if the relevant range of times is prior to t=0, instead of after t=0, then the time of phase optimum is arrived at by subtraction, rather than addition, modulo t.

So, the key steps to consider are to (i) negate the phase value reported by FFT-NLLS, (ii) define the range of times between which a phase optimum estimate is desired, and (iii) extrapolate modulo t by period addition (forward in time relative to t=0) or period subtraction (backward in time relative to t=0) until an acceptable phase value is produced (after which it may be again converted into relative wall-clock time, for example).

Newly-developed (ca. 2002) analysis for rhythms (COSOPT, as extension of CORRCOS)

Recently developed is another analytical procedure for rhythms analysis, COSOPT (Ceriani et al., 2002; Panda et al., 2002), an extension of CORRCOS (Harmer et al., 2000; Duffield et al., 2002), that may offer some utility beyond the limitations of FFT-NLLS; most notably in cases of particularly short, sparse, and noisy data sets.

See the following references for applications of the COSOPT and CORRCOS algorithms:

COSOPT references

Ceriani MF, Hogenesch JB, Yanovsky M, Panda S, Straume M, Kay SA (2002) Genome-wide expression analysis in Drosophila revelas genes controlling circadian behavior. *J Neurosci* 22: 9305-9319.

Panda S, Antoch MP, Miller BH, Su A!, Schook AB, Straume M, Schultz PG, Kay SA, Takahashi JS, Hogenesch JB (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* 109: 307-320.

CORRCOS references

Harmer SL, Hogenesch JB, Straume M, Chang H-S, Han B, Zhu T, Wang X, Kreps JA, Kay SA (2000) Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science* 290: 21 10-21 13.

Duffield GE, Best JD, Meurers BH, Bittner A, Loros JJ, Dunlap JC (2002) Circadian programs of transcriptional activation, signaling, and protein turnover revealed by microarray analysis of mammalian cells. *Curr Biol* 12: 551-557.

Using FFT-NLLS without BRASS

Input file format

Data file format is a 3-column ASCII (i.e., text file) format in which each row is a data point. No extra rows for headers or other identifying information of any sort. Column 1 contains the values of your observable (i.e., the y-axis values). Column 3 contains corresponding values of time (i.e., the x-axis values). If you have individual estimates of uncertainty for each of your y-values, they go in column 2 (i.e., the standard errors associated with corresponding y-values). If you have no error estimates for column 2 (which is often the case), just fill column 2 with zeroes for every point and a uniformly-weighted (with zeroes in column 2), as opposed to a variably variance-weighted (with individual uncertainty estimates in column 2), nonlinear least squares analysis will be performed. These files must be stored as text (ASCII) files. You can use any kind of column delimiter you want (space, tab, or comma) because a free-format read is used for data input (so it should accommodate any of the mentioned delimiters without problems). Name the data files as "FILENAME.DAT" where "FILENAME" is any name you choose (not more than 8 characters) and ".DAT" identifies it as a data file.

THE FIRST STEP OF FFT-NLLS ANALYSIS: "FOUR-ANL"

Prior to beginning an analytical session, position yourself into the subdirectory that contains all of the data files you wish to analyze (i.e., at the command prompt, change directory to make yourself resident in the subdirectory containing your data as the current working directory). The first program to execute is "FOUR-ANL". When invoked, "FOURANL" first prompts you to supply a file name to which results are to be written. I recommend you provide a name like "FILENAME.SUM" where "FILENAME" can be any name you want (not more than 8 characters), but append ".SUM" as the extension. (By using my recommended extensions, you and I will always be able to communicate clearly regarding which analysis files we are talking about and, additionally, you will establish a consistent naming convention that will greatly reduce the likelihood of future confusion.) Type in the filename and press <ENTER>. "FOUR-ANL" next prompts you for the name of a data file to be analyzed. Type in the filename and press <ENTER>. "FOUR-ANL" next prompts you for a confidence probability value to be used in the analysis of this data file. A good default choice that I will recommend here is "95." (this value is expected in units of "percent", hence 95. and not 0.95). "FOUR-ANL" then goes ahead and performs the analysis of this data file until it decides that it has completed its work. (We can discuss at some other time, if you are interested, the more detailed workings of exactly how "FOURANL" performs its calculations, on what its decision making processes are based, and on variations for selection of confidence probability level values.) At the conclusion of analysis of the current data file, "FOUR-ANL" again prompts you for the name of the next data file to be analyzed followed again by the confidence probability level at which this data file is to be analyzed. In this way, any number of data files can be analyzed in a given analytical session, the results of which are summarized sequentially in the summary file "FILENAME.SUM".

AUTOMATING FFT-NLLS ANALYSIS BY "FOUR-ANL"

Because of these cyclically repetitive queries for data files and associated confidence probability levels for analysis, this analytical process can conveniently be automated by using a script input from which all program input can be redirected (i.e., instead of having to type the requested information in directly from the keyboard every time "FOUR-ANL" requests it). If you are not familiar with this (command prompt redirection of script files into an executable program), I encourage you to check documentation or another knowledgeable resource about how to redirect program input from a file at a command prompt (this is done in the same manner at both DOS/Win command prompts and at UNIX command prompts, and is of the form "EXECUTABLE < SCRIPTFILE"). This automation feature (coupled with a

controlling batch file, perhaps) can allow very lengthy and/or multiple analytical sessions (that may take hours or days, even) to proceed completely automatically and unattended, with no user interaction required at all, all results summarized in ".SUM" files.

THE SECOND STEP OF FFT-NLLS ANALYSIS: "FLS-PLOT"

When analysis by "FOUR-ANL" of all the data files in a given subdirectory is complete, make sure you again position yourself into the same subdirectory before proceeding with the second step of the FFT-NLLS analytical process for these files. Execute from the command line the program "FLS-PLOT". "FLS-PLOT" will first prompt you for the name of the summary produced by "FOUR-ANL". Type in the appropriate name of the file we referred to as "FILENAME.SUM" above and hit <ENTER>. "FLS-PLOT" will next prompt you for the name of the output file you would like "FLS-PLOT" to create. I suggest you provide "FLS-PLOT" with an output file name like "FILENAME.OUT" where "FILENAME" is again any name you want to use (not more than 8 characters) with the extension ".OUT". The file "FILENAME.OUT" will contain a summary of only the best fits of the analysis by "FOUR-ANL" (as opposed to the complete, verbose summary of the analytical session contained in "FILENAME.SUM").

"FLS-PLOT" SEQUENTIAL GRAPHIC OUTPUT OPTION

"FLS-PLOT" next prompts you for whether you want (enter "1 ") or do not want (enter "0") to see graphs of each data set analyzed along with the corresponding best fit curve. In general, it's probably a very good idea to go ahead and view the graphs, along with a pencil and a pad of paper so you can make appropriate notes, to verify that good fits are indeed okay, and to identify and note fits that may not look very good by eye (it sometimes happens that the best attempt by "FOUR-ANL" to fit the data is not satisfactory, and you need to catch this if/when it happens by viewing the best fits with the data). If you are not concerned with viewing each graph but just want "FLS-PLOT" to hurry up and create "FILENAME.OUT", then select not to view the graphs.

TECHNICAL NOTE REGARDING "FLS-PLOT" GRAPHIC SCREEN OUTPUT

Prior to viewing on-screen the graphic output produced by "FLS-PLOT", please toggle your command prompt session to full-screen mode, not window mode. Toggling between these two display modes in Windows is accomplished by holding down the <ALT> key and then

depressing the <ENTER> key. It is necessary to be in full-screen mode prior to displaying to the screen the graphics that are created by "FLS-PLOT". If you wish to produce hard copy graphs of the screen graphics displays, get in touch with me; I can recommend an inexpensive software product (Snag-It32 by TechSmith; see www.techsmith.com) that I am very pleased with for this purpose. I can go over with you at that time how I set up my environment for manually capturing to a printer the graphics produced on-screen by "FLS-PLOT".

"FLS-PLOT" CREATION OF THEORETICAL BEST-FIT TEXT FILES

"FLS-PLOT" next prompts you for whether or not you want it to create theoretical best-fit files (these are 1001 point files of the theoretical best-fit curve ranging from t(min) to t(max) of the data file analyzed by "FOUR-ANL" and summarized in file "FILENAME.SUM"). Respond with a "0" for no or a "1 " for yes. If no, "FLS-PLOT" will not produce any theoretical output files. If yes, "FLS-PLOT" will produce sequentially numbered files corresponding to each file summarized in "FILENAME.SUM" in the same order they were analyzed. The theoretical best-fit files are named "00000001 .THE", "00000002.THE", "00000003.THE", etc. (These are text (ASCII) files that are useful for producing publication quality graphs, after importing into your favorite graphics software along with the original data file.)

"FLS-PLOT" SEQUENTIAL GRAPHICS VIEWING AND RESCALING OPTION

If you are viewing the graphs, hit <ENTER> to move beyond the graphics screen. You cannot go back or select which file to view; "FLS-PLOT" just moves sequentially forward through the files, one at a time, until finished. You are provided an option to rescale the graph, if you wish, and view it again after rescaling. However, in general, there is little value in doing this and you will probably just choose to move on by hitting <ENTER> again at the query regarding whether you want to rescale or not and, in doing so, going to the next graph. (To proceed from graph to graph without rescaling, you will just need to hit <ENTER> twice in a row, thus clearing the previous graph and drawing the next.) After proceeding through all of the files analyzed and summarized in "FILENAME.SUM", "FLSPLOT" terminates and the file "FILENAME.OUT" will have been written in its entirety (along with the ".THE" files, if theoretical best-fit files were requested).

THE THIRD, AND FINAL, STEP OF FFT-NLLS ANALYSIS: "CONDENSE"

The third, and final, step of the FFT-NLLS analysis procedure involves running the program "CONDENSE" from the command prompt (again, you need to be resident in the same subdirectory in which you have been working, containing your ".DAT" data files, your "FOUR-ANL" ".SUM" summary file, and your "FLS-PLOT" ".OUT" file). "CONDENSE" first prompts you for the name of the "FLS-PLOT" output file to condense, to which you respond with "FILENAME.OUT" followed by <ENTER>. "CONDENSE" then asks you for the name of the CONDENSE output file to create, to which you respond "FILENAME.CND" where "FILENAME" is anything you want (not more than 8 characters) and ".CND" is the extension to use to identify this file as output from "CONDENSE". "CONDENSE" then prompts you for lower and upper limits of period components to search for, extract, and write to the output file. If you want to summarize everything in the ".OUT" file (i.e., excluding no statistically significant identified period components, which is generally what you will want to do), enter "-1 .,1 .e35" at this prompt. By doing so, you are instructing "CONDENSE" to report all period components with periods greater than -1. and less than 10-to-the-35th-power (i.e., our approximation of infinity); in other words, report all period components identified in the analysis. This query exists in case you may ever wish to select some subset range of periods to extract from the analysis (for example, only those between 19 and 29, or something like that, if you were interested in quickly identifying those data files for which a rhythm moderately close to circadian was detected; in such a case, you could name the "CONDENSE" output file to something like "FILENAME.CIR" ("CIR" for circadian) and have "CONDENSE" summarize only circadianly rhythmic files for you).

INTERPRETATION OF FFT-NLLS RESULTS

The results reported in the ".CND" file are the best-fit FFT-NLLS parameter values identified for each data set at the specified level of confidence probability. This last point is important! Reported results will vary depending on the confidence probability level employed during analysis by "FOUR-ANL". Don't get caught trying to compare apples with oranges! For valid comparisons of results of analysis of different data files, it is necessary to have identical conditions represented in the results of analysis (i.e., same numbers of data points per data set, same sampling frequency per data set, same confidence probability level used during analysis). The contents of the ".CND" file present the amplitude, period, and phase of all statistically significant rhythmic components that the FFT-NLLS algorithm

could find at the specified level of confidence probability. The values in parentheses under each value are the approximate joint parameter confidence intervals at the user specified level of confidence probability. Note this last point! The values in parentheses are not necessarily estimates of the standard errors of the parameters (unless a confidence probability value of 68.26% was used in analysis). The values in parentheses are estimates of parameter confidence at the user specified level of confidence probability, whether that is 95% or 90% or 99% or 68% or..

RELATIVE AMPLITUDE ERROR AND RHYTHMIC DETERMINATION

The column labeled RelAmp is the relative amplitude error (the RAE, which is the value of the amplitude error estimate divided by the value of the most probable amplitude estimate). RelPer and RelPhi are not used in any interpretive capacity at this time. RAE can range from a value of zero for an infinitely well determined rhythmic component (zero error) to a value of one, theoretically, for a minimally determined rhythmic component (error in the amplitude equals the amplitude value itself). There will be times when an RAE value greater than unity will be reported. Although theoretically not possible within the context of linear, symmetric models, the multi-component cosine model that underlies the FFT-NLLS analysis is neither. As such, RAE values greater than unity do sometimes occur (most often, in my experience, when a very-long-period component is identified during analysis to account for long-term drifting that is occurring in the data set being analyzed). In such cases, it is usually a rhythm that is not of interpretive interest, but rather is probably there to accommodate trend.

INTERPRETATION OF PHASE

The phase value is referenced to whatever time zero is in your data file's time axis. A positive phase value means advanced relative to zero (i.e., negative time) and a negative phase value means delayed relative to zero (i.e., positive time). Please note this relationship in definig phase so as to avoid confusion! The phase marker is cosine acrophase (i.e., peak of cosine wave) and it recurs modulo whatever the period for that rhythmic component is. So, be careful when comparing phase reference points that you do so correctly with respect to some absolute time marker, not just to what may be a quite arbitrarily assigned zero time value in the time axis of the data files being analyzed. Also, please recognize that the phase referred to is that of a rhythmic component. In the event of multiple rhythmic components being identified in an analysis of a data file, the acrophase of a

rhythmic component may not always bear a discernable resemblance to an acrophase of the observed, composite rhythm (i.e., of the composite best-fit curve that may be composed of a linear combination of a number of contributing cosine waves of varying amplitudes, periods, and phase relationships).