Sleep and Circadian Analysis MATLAB Program (S.C.A.M.P.) Instructions

Written by Christopher G. Vecsey, software developed from the previously published version in:
Donelson N, Kim EZ, Slawson JB, Vecsey CG, Huber R, and Griffith, LC (2012) High-Resolution Positional Tracking for Long-Term Analysis of *Drosophila* Sleep and Locomotion Using the ''Tracker'' Program. PLoS ONE 7(5): e37250. doi:10.1371/journal.pone.0037250.

We have produced a unified MATLAB script suite for rapid and user-friendly analysis of sleep and circadian data from either standard Drosophila *Activity Monitoring (DAM) or using our video-based Tracker system. The following steps describe the general acquisition and processing of raw data, and operation of this script. Even if you're a grizzled, accomplished user of DAM hardware and software,* ***please follow the instructions carefully****. There are some idiosyncrasies that may differ from the way you normally handle your data. When trying out this program for the first time, start with raw TriKinetics data, not previously processed files.*

If you discover issues with the software, or have suggestions for improvement or additional analyses that could be dovetailed in, please let me know. I try to update parts of the code as new operating systems or versions of MATLAB force me to stay compatible with them. Full disclosure: I am a neurobiologist, not a coder.

Contact Info: *Christopher G. Vecsey, Neuroscience Program, Skidmore College, 815 N Broadway, Saratoga Springs, NY 12866. cvecsey@skidmore.edu.*

Getting DAM Data off the Acquisition Computer

The first step is to collect the raw data from the relevant monitors from your experiment into a new folder that has a specific identifiable name and then transfer it to another computer for analysis.

1. Select the files for the monitors you have been using, including the environmental monitor that was in your incubator. For SCAMP analysis, you do not need the "LCdaylight" file.

    a. You can simply drag and drop the files into a separate folder. This will essentially erase those files – the next time data is collected (typically every minute), new files will be created to replace the ones you removed.

    b. If your experiment is still ongoing or if you want to be cautious, you can copy and paste the files into your folder instead. This will allow data to continue to be added to the raw data files.

    c. You do not want to always copy and paste, because the data files will get large and unwieldy.

Setting up DAM Data Set for MATLAB

The second step is to trim each of the large data files (1 per monitor) to the appropriate start and end times for your experiment. Typically what is desired is to start the file at the beginning of the day and end at the very end of the day, so that you have an exact number of days with no extra minutes of recording.

This step will also break the files apart so there is a separate file for each tube in the monitor (32 per monitor), so each fly's data is separate.

2. Open the 'DAM File Scan' program. If you want additional information about this program, it is available through a TriKinetics pamphlet available on the company's website. At the time of writing this, I use the DAMFileScan110X version, and I would recommend you do the same.

3. Click "Select Input Data Folder".

    a. Select folder containing monitors data.

4. Select monitor range (typically all of them).

5. Click "Scan" – this finds the earliest and latest points within any of the files to create an overall time range.

6. Choose bin length to be one minute in length.

    a. Set output file type to "Channel Files" or "Channel Counts".

    b. Set "Extra Readings" to "Sum into Bin".

    c. Name the run (we've found it easiest to use the current date, i.e. "**20150701**")

NOTE: There are a few ways that the names of data files can disrupt analysis. Shown below are standard folder names for processed data (with the raw monitor output text files below) based on the experiment date. The files inside each folder would be called "062811M052C01.txt", "062811M052C02", etc. The "M" and "C" in the file names become placeholders for subsequent processing, so make sure not to include an "M" or "C" in your data folder prefix. Also omit any non-alphanumeric keystrokes.

7. Choose the first and last bins to match the timing of the experiment.

    a. First bin should be the first minute of the experiment (ie. 10:01 if the experiment started at 10:00). **NOTE:** If you loaded flies on 6/29/15 in the middle of the day, most likely the first day you would want to use is 6/30/15. If lights turned on at 10am, then 10:01 would be the first recorded minute of that day.

    b. Last bin should be the last minute of data from the last day of the experiment (ie. 10:00 if the experiment ended at 10:00).

    c. NOTE: Choosing the exact minute can be tricky depending on your computer's type and operating system. You may need to scroll in different ways to move in small increments.

8. Click "Save". This will create a new folder with the name you chose in #9 above. It will be located in the parent folder containing your original data folder.

a. Rename the folder by adding "-1" to the end of the name, i.e."**20150701-1**", and move it inside the original data folder.

9. Repeat steps 9-11 using a bin length of 30 min this time. **NOTE**: You will need to reset both the start and end times for the experiment after you switch to the 30-minute bin length.

10. Keep the run name the same, and click "Save" again. This again creates a new folder. Rename the folder "20150701-30". Again, you can move this folder inside the original data folder.

11. During each of these saves, the program will alert you if there are any issues with the data (skipped or extra bins, for example). There will be extra bins when you switch to the 30-minute bin length – this is normal.

12. You're now ready to run the program "scamp" in MATLAB! Note that in some of these instructions, the words "monitor" and "board" are used interchangeably to refer to the piece of TriKinetics hardware that collects data from 32 individually housed flies at a time.

Sleep and Circadian Analysis MATLAB Program (S.C.A.M.P.)

The third stage of analysis is to run the SCAMP analysis itself. This involves choosing the correct folders of scripts in MATLAB, selecting the FileScanned data for your experiment, naming your groups and de-selecting dead flies, and finally running your sleep and circadian analyses!

Step 1 – Initial Setup

These first steps need only be done one time after downloading the script.

Open MATLAB. At the time of writing this, I am using MATLAB_R2015a. Under "File", choose "Set Path." Click "Add with Subfolders" and choose the folder containing all of the relevant scripts ("Griffith Sleep Analysis").

NOTE: If you already had variants of some of these scripts set under your MATLAB path (from Hall, Rosbash, Griffith, or others), **make sure to remove them**. Otherwise, MATLAB may use incorrect versions of the scripts. Our analysis package will use many built-in MATLAB scripts that should be available automatically. It also uses a few scripts that are only present in the Signal Processing toolbox – make sure that this toolbox is installed in your version of MATLAB. If it is not, circadian analysis may fail.

Step 2 – Running SCAMP

See previous pages about preparing raw DAM data for use in SCAMP. Drag and drop the folder with the data you want to analyze into the command window in MATLAB. The DAM files that you use in SCAMP must be saved in the "Channel Files" style of the DAMfilescan system. Currently, SCAMP does not read "Monitor Files".

In the same window, type "scamp" to begin analysis – **this needs to be lower case**. A window will open (see **Figure 1** below). On PCs, it will prompt you to choose the folder containing your 1-minute data. On some Macs, this prompt is absent but the action should be the same.

Once you have made your choice, a new folder will appear. Again, in some Macs there will be no prompt, but on PCs, you will be prompted to choose the folder containing your 30-minute data.

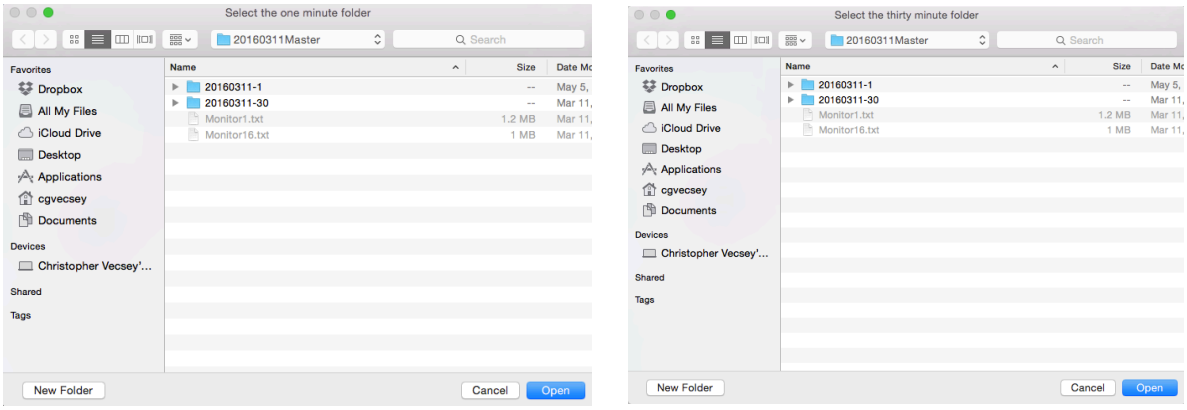

Figure 1. Windows for Data Folder Selection. These particular examples indicate whether 1-minute or 30-minute data should be selected. If on your computer the windows do not have informative titles, the first time a window pops up, select the 1-minute data folder, and the second time select the 30-minute data folder.

STEP 3 – Formatting groups and exclusion of dead flies

Once you've selected the folders with your 1 and 30-minute files, a new window will open (see **Figure 2** below). All of the monitors of flies being analyzed will be listed at the left. Clicking on a particular monitor will bring up an array of 32 channels on the right corresponding to the flies on the monitor that is selected at the left. Editable text boxes (**Group**) allow the user to specify group names for individual flies based on genotype, treatment, etc.

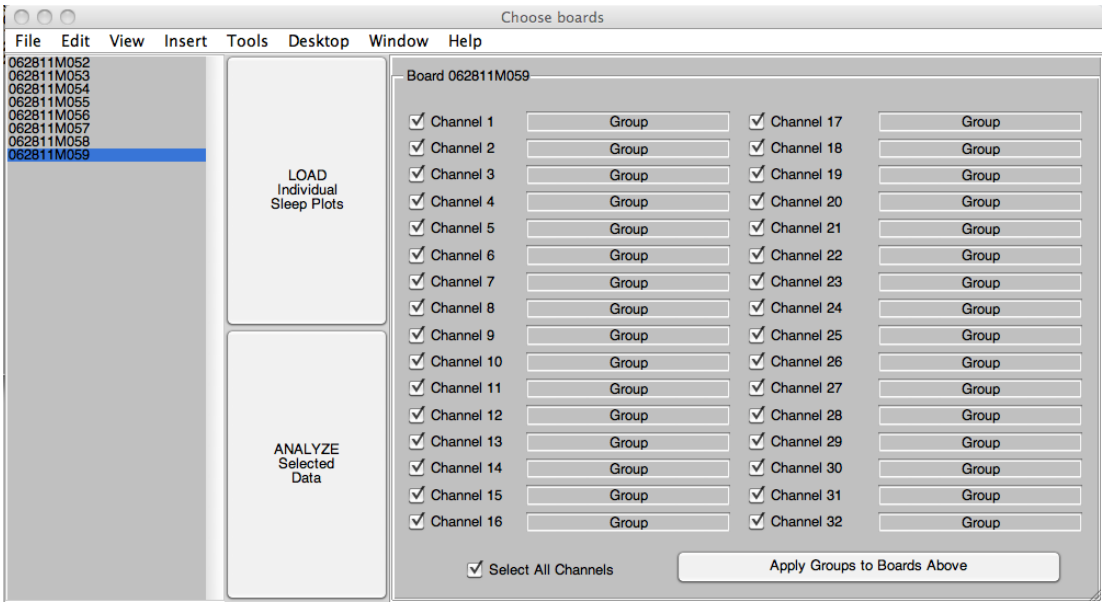

Figure 2. The Fly Selection Window.

NOTE: If the group patterns of flies are similar across monitors, it is recommended to set the group names for the last board in the list, then to click the "**Apply Groups to Boards Above**" button. That way,

you'll only have to enter group names one time. If group patterns are not identical in all boards, then proceed to the next board up the list, make whatever group changes are necessary, and so on.

While selecting a board (or group of boards), clicking the "**LOAD Individual Sleep Plots**" button will bring up a figure (see **Figure 3** below) showing double-plotted activity plots for all of the channels on the selected board. Blue indicates inactivity. This is used to determine which flies are dead or alive. The flies are arranged horizontally from 1 to 32. Flies that the user wishes not to use for subsequent analysis can be un-checked.

When all groups have been specified, and all unwanted flies have been un-checked, select all of the boards you wish to analyze, then click the **"ANALYZE Selected Data"** button.

NOTE: It's advisable to save the **fly selection** window at this time as a .fig MATLAB file. That way, you can re-open this file at a later date, and reselect different boards, rename groups, or run other analyses, if necessary.

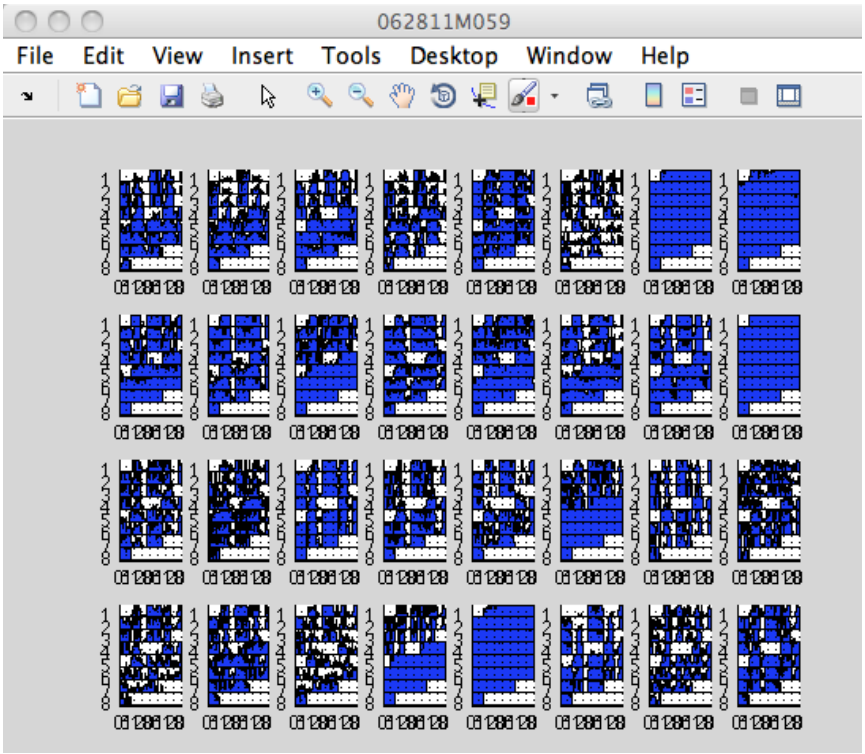

Figure 3. An Individual Sleep Plot Window. Dead flies are evident in slots 7, 8, 9, 11, 16, 22, 28, and 29. This experiment included a period of sleep deprivation, which is why there is a white rectangle in the center of the data for live flies.

Step 4 – Selection of Bin Length for Analysis

The **"Select Analysis"** window (see **Figure 4** below) will open (note that the Choose Boards window remains open). This is the main center of operations for the SCAMP program.

To the left are the available sleep and circadian analyses. A standard subset of each will be checked by default.

To the right are lists of the groups that you defined in the Choose Boards window, and a list of experimental days based on the length of the files you inputted.

Below those lists are several buttons that allow the user to specify what type of analyses they wish to perform. In the center of the Choose Analysis window are the action buttons that will output graphs and data export.

The first step is to choose the length of "bin" that you want to analyze across. The preset value for for sleep analysis (**Default (12/24)**) will analyze across 12-hour bins (separating day and night, e.g.) and also across the full 24-hour day.

If you wish to analyze across smaller bins, you may enter the # of hours corresponding to the desired bin size in the editable text box. For example, writing "3" will break the data into 3-hour bins. If, after analyzing with a custom bin length, you wish to re-analyze using the standard 12/24 system, write "12/24" or "Default (12/24hr)" into the text box and click the "**Analyze for Chosen Bin**" button.

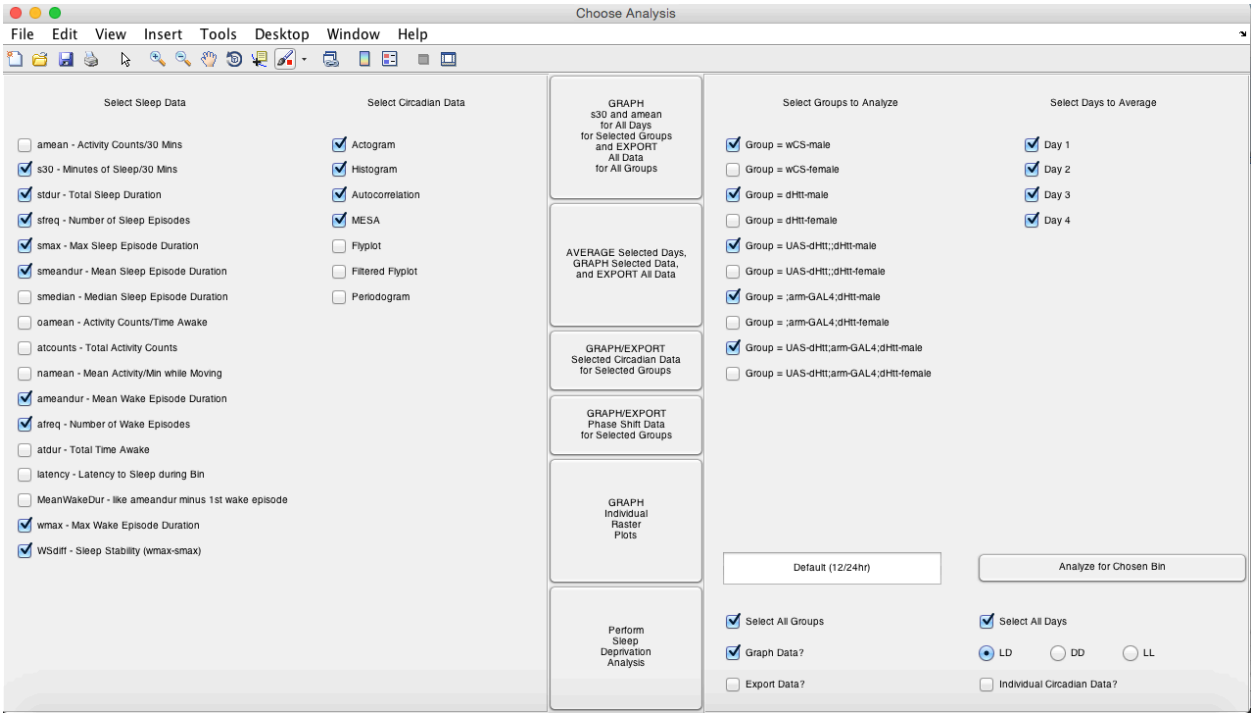

Figure 4. The Main Analysis Window.

You must click the "**Analyze for Chosen Bin**" button before performing any other graphing or export.

A message saying "Bin Analysis Complete" will appear in the MATLAB command window when this analysis has finished. With large numbers of flies or slow computers, this may take a few minutes.

Step 5 – The Analysis Buttons

A. The "**GRAPH s30 and amean for All Days for Selected Groups and EXPORT All Data for All Groups**" button will do <u>one of 3 things</u>:

1. If neither the "**Graph Data?**" nor the "**Export Data?**" checkboxes are selected, it will do nothing.
2. If the "**Graph Data?**" checkbox is selected, clicking this analysis button will output two graphs (see **Figure 5** below) depicting "**s30**" and "**amean**" data in 30-minute bins across all days of the experiment for the groups that you have selected. Which days are selected at the right has no effect on the output from this button.
3. If you check the "**Export Data?**" checkbox instead, this analysis button will output .csv files (see **Figure 6** below) corresponding to each of the sleep analyses at the left of the analysis window.

NOTE: Running this analysis multiple times will overwrite these files unless you move each set of export files into separate folders before the next export.

NOTE: This analysis will not work if there are exactly 17 experimental groups.

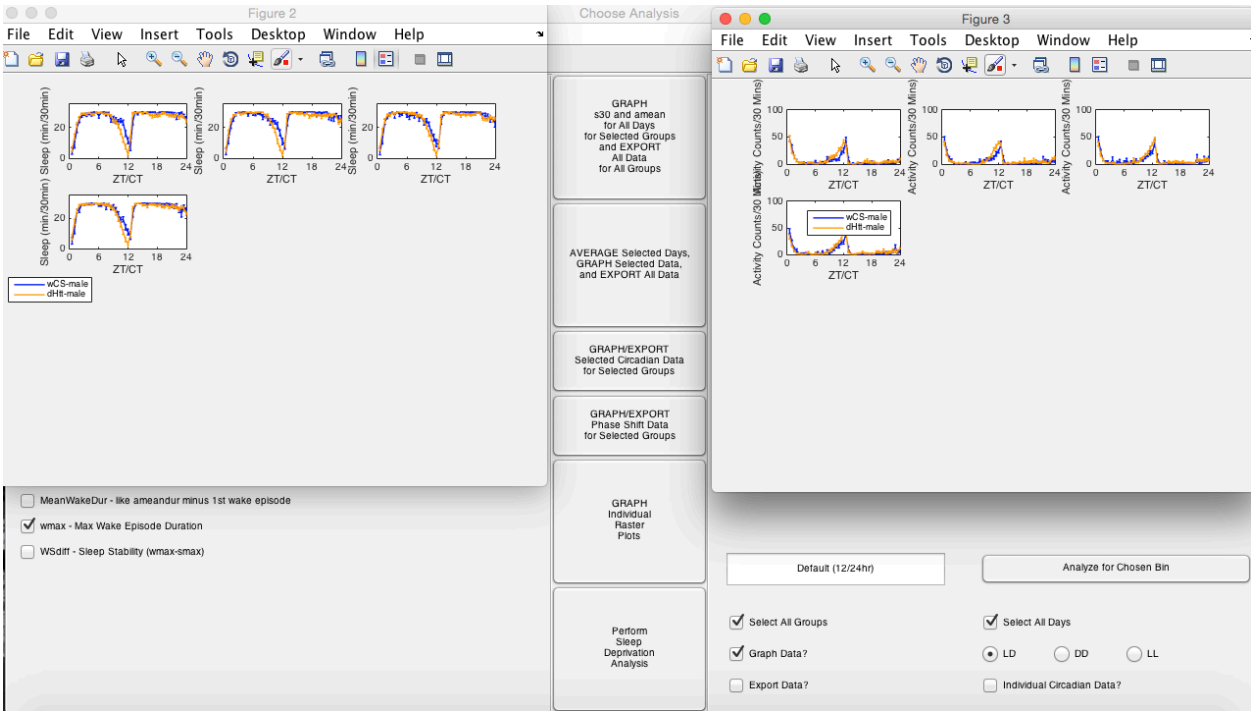

Figure 5. s30 and amean Analysis Graphs.

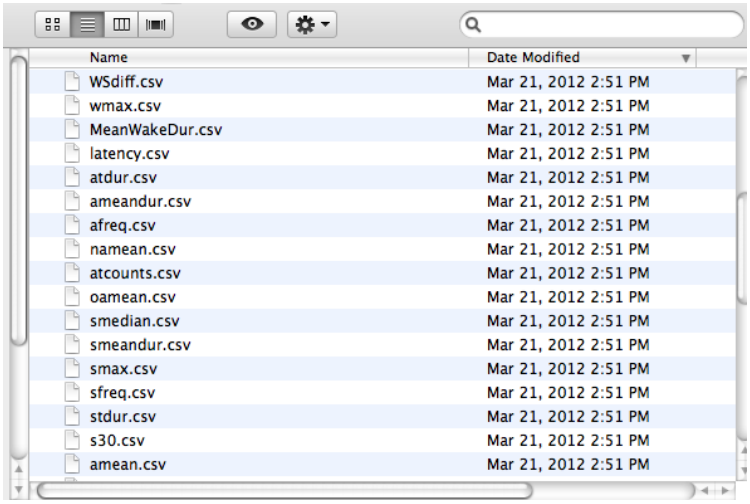

Figure 6. The 17 Data Files Created when the "Export All Data" button is Selected for the "Graph All Days" Analysis Function.

B. The "**AVERAGE Selected Days, GRAPH Selected Data, and EXPORT All Data**" button functions the same as above, but the function of this button depends on whether "**Graph Data?**" or "**Export Data?**" are checked.

This button's output depends additionally on which days are selected. Having no days selected will result in an error message.

This analysis will average across all of the days that are selected. This analysis also alters the labeling of figures based on whether the user has selected either LD or DD/LL, indicating either a normal Light/Dark cycle or a Dark/Dark or Light/Light schedule.

Checking the "**Graph Data?**" checkbox, then clicking this analysis button will output a figure (see **Figure 7** below) depicting averaged data for whichever sleep analyses are selected at the left of the analysis window.

> **NOTE:** Selecting more than 9 analyses at once will result in a graphing error. If you want to graph all analyses, perform this analysis twice, with different analyses selected each time.

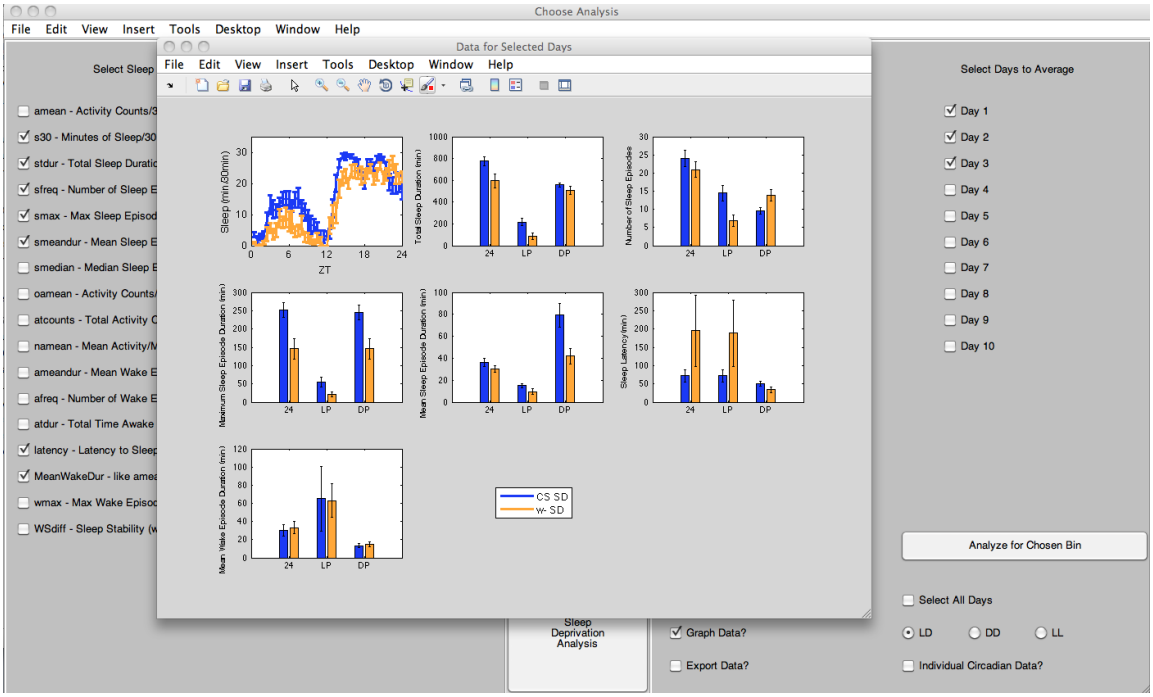

Figure 7. Example of a Graph of Averaged Selected Days.

Checking the "Export Data?" checkbox, then clicking this analysis button will output .csv files (see **Figure 8** below) corresponding to each of the sleep analyses at the left of the analysis window. As above, be careful about overwriting by running export multiple times.

NOTE: This analysis will not work if there are exactly 17 experimental groups.

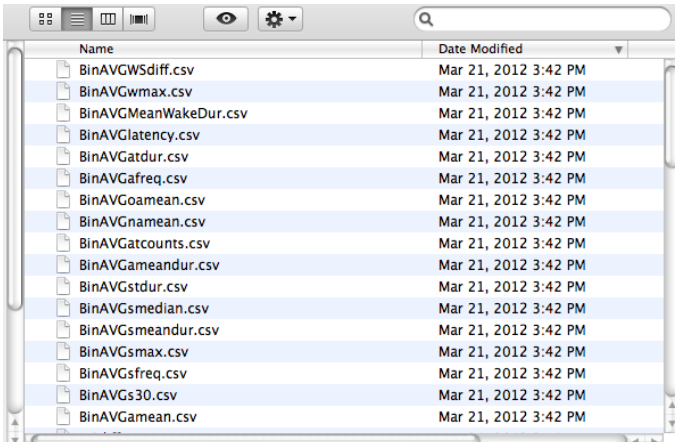

Figure 8. The 17 Data Files Created when the "Export All Data" button is Selected for the "Average Selected Days" Analysis Function.

C. The "**GRAPH/EXPORT Selected Circadian Data for Selected Groups**" button requires you to click the "**Analyze for Chosen Bin**" button, but what the user defines as the bin length will not alter the output.

Unlike the top two previous buttons, this one does not require you to select the "**Graph Data?**" checkbox for graphing to occur.

Four default output graphs have been selected from the list of circadian analyses (see top graph in **Figure 9** below). The user can change the choices via the checkboxes. If "**Individual Circadian Data?**" has been selected, clicking this button will also output the same output graphs for individual flies. Their board/channel #s will be labeled at the left of each set of graphs. For legibility, a maximum of 8 animals will be graphed in a single window, so with large groups this function will output many windows. They will be labeled "Group **X**, Figure 1", "Group **X**, Figure 2", etc. (see bottom graph in Figure 9 below).

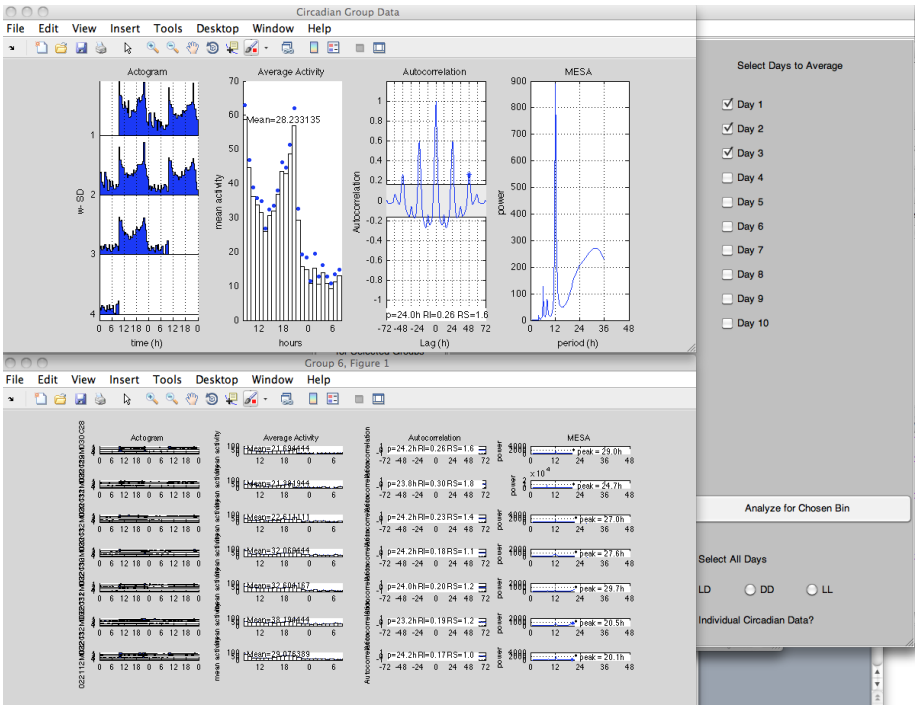

Figure 9. Group and Individual Circadian Graphs.

This button is also unique because it will only export data from the groups that have been selected. In order to export, the "**Individual Circadian Data?**" checkbox must be selected. The exported data ("Circadian Data Export.csv") consists of the 3 data values shown at the bottom of the autocorrelation graph (px: period, ri: rhythmicity index, rs: rhythmicity strength).

NOTE: If, when you run circadian analysis, you get an error saying that the function "butter.m" could not be found, this indicates that the version of MATLAB you are using does not have the Signal Processing toolbox installed. Making sure that the toolbox is installed should solve the problem.

D. Phase Shift Analysis. This is specialized analysis for experiments in which a phase shift is induced in some groups but not others. For example, following baseline days in light/dark (LD) conditions, a brief light pulse could be given to some groups during the night. Then the animals would transition into an extended period of constant darkness to assess their internal rhythm. This analysis will find peaks in the shifted and non-shifted groups and compare the timing of those peaks to find a difference (phase shift).

The analysis requires that you choose groups to analyze in pairs – you will need to run the analysis multiple times to examine multiple pair comparisons. For a single genotype, choose the groups that were phase-shifted or were not shifted. Once you have selected exactly 2 groups, click the "GRAPH/EXPORT Phase Shift Data for Selected Groups"

Two windows will open, typically one on top of the other. In **Figure 10** below, these windows have been separated so both are visible. The first window allows you to manually de-select activity peaks so that there are matching peaks between the two groups. This is a tricky process in some data sets, especially during the LD period when there can be a few peaks per day. The important thing is to end up with matched sets of peaks between the two groups.

The analysis can't continue correctly until there are the same # of selected peaks in each set. If this is not true, the graph will display the message "**WRONG! Click on peaks to remove them until peak sets match**" (see **Figure 11**) Note that additional peaks cannot be added in this graph to correct this issue – they can only be removed by navigating the cross-hairs over the marker for each peak you want to remove and clicking on it.

Once there are the same # of peaks in each group, the graph's message will change to "**MAYBE RIGHT**". This is because, even though the total #s of peaks are the same, some peaks could still be mismatched between groups. It is up to the user to determine when the peaks are correctly selected. At this point, as the title of the figure indicates, **"When satisfied, click in gray area outside of plot".** This will tell the script to calculate and graph the phase differences based on the peaks you selected.

The current window will close, and two graphs will be plotted in the second window (see **Figure 12**). This is the graphical output for the phase shift analysis function.

> **NOTE:** One potential source of confusion with the peak selection window is that if you click on a different graph or elsewhere on the screen, when you eventually do click in the gray area of the window to finalize the analysis, it may not graph the results correctly.

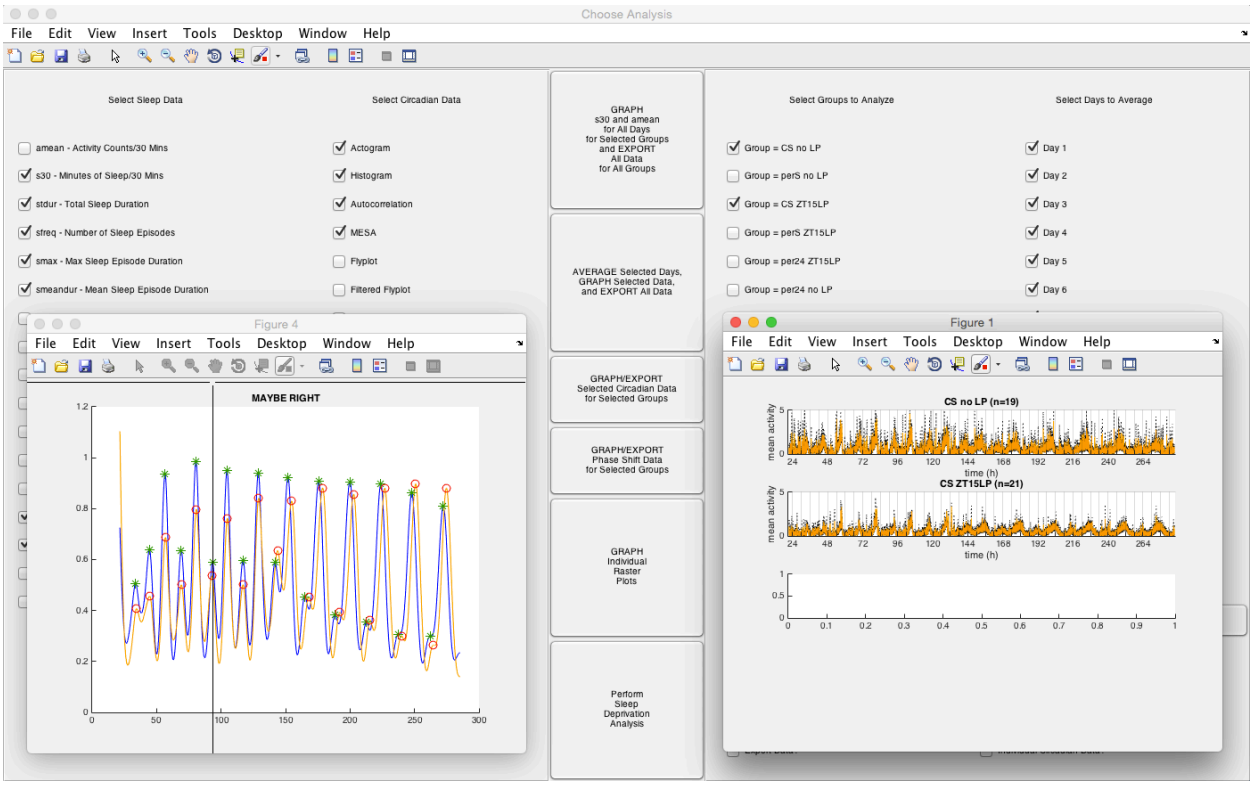

Figure 10. The Two Windows That Open During Phase Shift Analyis. Choices of peaks will be carried out using the left window.

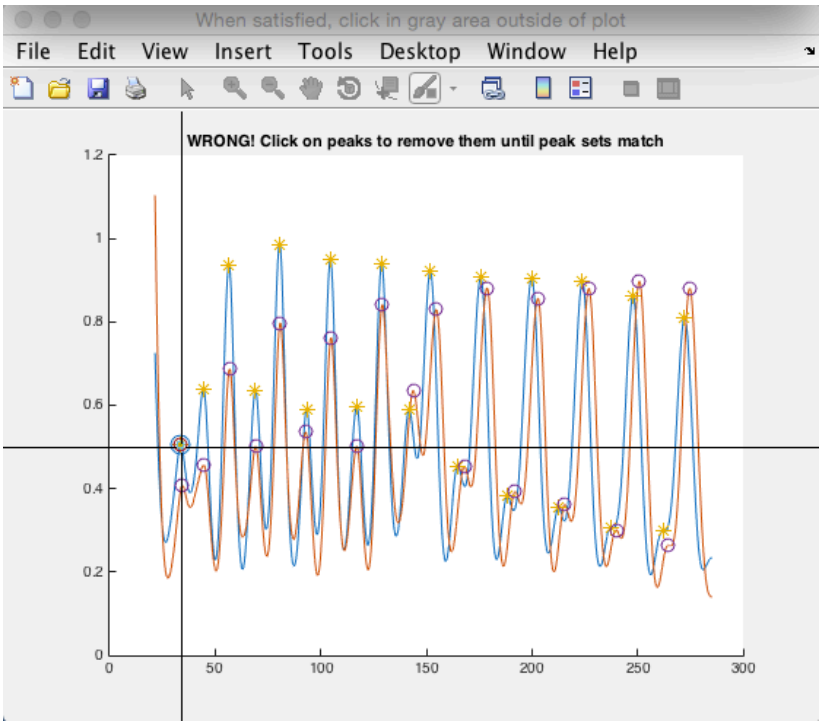

Figure 11. Appearance of Manual Peak Selection Window When Numbers of Selected Peaks Are Unequal Between Groups. After de-selecting the peak under the cross-hairs, the numbers of peaks were not equal between the two groups. This led to the warning message in bold at the top of the plot.

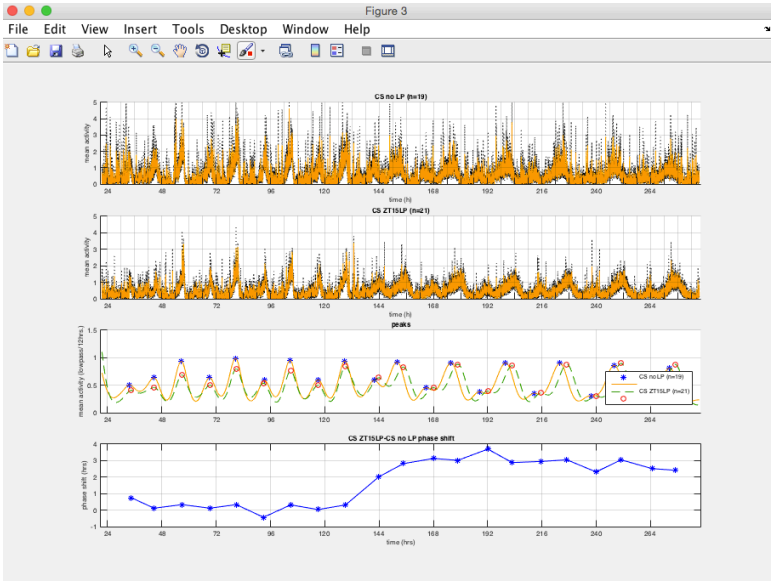

Figure 12. Graphical Output of Phase Shift Analysis. After completing selection of peaks, the bottom two graphs will be plotted.

If the **Export** checkbox is selected at the time when Phase Shift Analysis is run, completing the peak selection process will also generate an Excel-compatible export file of the data, including columns of the peak times for each group and the phase difference calculated by subtracting Group 1 from Group 2.

NOTE: This subtraction is always performed Group 2 – Group 1. If this is not the direction you want, you will have to manually inverse the results. You will also have to determine based on your experimental conditions when the phase shift actually occurred, and choose the appropriate phase shift data from the export file to use for further analysis.

E. The "**GRAPH Individual Raster Plots**" button does as labeled, and creates graphs based on user selections for groups and days (see **Figure 13** below). Each row shows data for each selected day for a given animal. This output can take some time for large numbers of animals, groups, and days. In the raster plots, blue lines denote periods of sleep. This button does not have data export functionality. The "**Graph Data?**" checkbox does not need to be selected for this button to output these graphs.

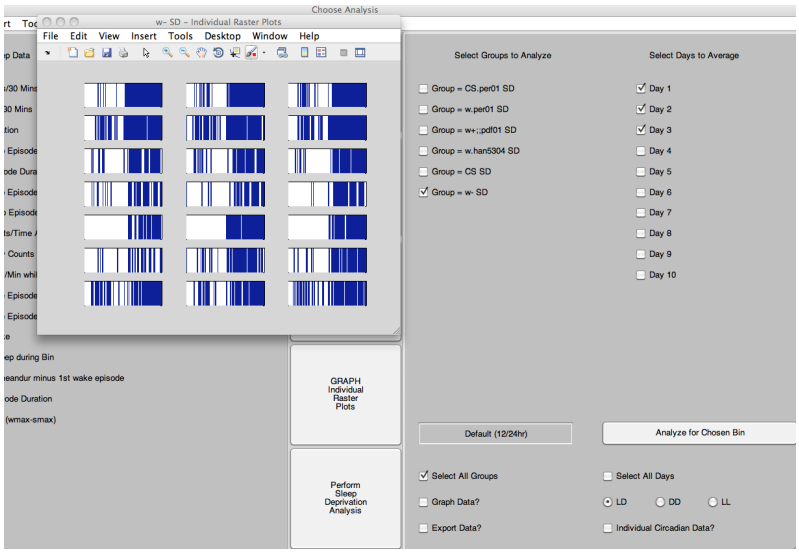

Figure 13. An Example of a Raster Plot Graph. Data from six individual flies over 3 days are shown. This type of graphical output can be useful for visualizing how reproducible flies' sleep patterns are across multiple days.

F. For the "**Perform Sleep Deprivation Analysis**" button to function correctly, the user needs to have performed "**Analyze for Chosen Bin**" first, using the default 12/24 hour binning. All other selections of group and day or sleep/circadian analyses will not be carried over for sleep deprivation analysis.

Clicking the button opens a new window (see left panel of **Figure 14** below). All group names will be listed as in the primary analysis window.

A list of preset values are found in the lower half of the window. These presets are designed for a situation in which the experiment had 3 baseline days, followed by a day when sleep deprivation began 12 hours into the day and lasted for 12 hours. The preset time to follow sleep amount during recovery is 24 hours.

Once the desired groups for graphing have been chosen, select the "**Graph Data?**" checkbox and click the "**Analyze for Choices Above**" button. This will output a graph showing a cumulative sleep loss and recovery plot (see right panel of **Figure 14** below).

For each 30-minute bin during a day, this calculates the average sleep across the baseline days chosen by the user. Then, starting on the day of deprivation, the amount of sleep in each 30-minute bin is subtracted from the equivalent bin's value from the baseline average.

This is done cumulatively, so if during the first two 30-minute bins of the sleep deprivation period an animal slept 20 minutes less than during the same bin during the baseline period, then the first value would be -20, and the next would be -40. And so on.

Typically, once sleep deprivation ends, animals begin to regain sleep, sleeping more than they would have during that same bin during the baseline. Therefore, this plot will usually look like a V, with animals gradually accruing sleep debt during the sleep deprivation period and then regaining sleep during recovery.

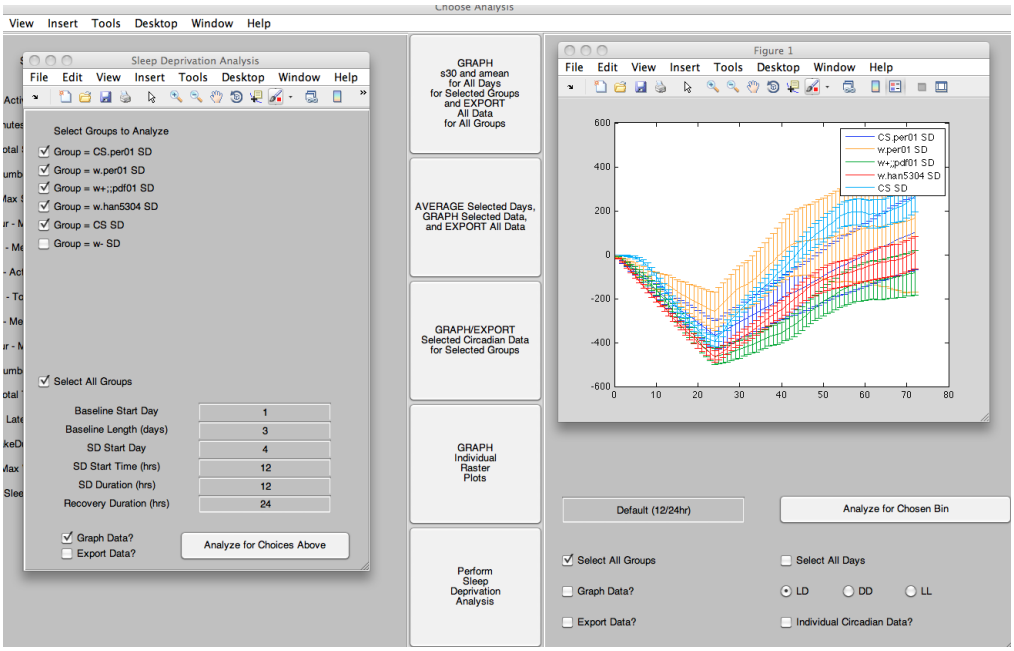

Figure 14. Sleep Deprivation Analysis Windows. Options are selected in the window on the left, and a cumulative sleep loss plot is graphed on the right.

It is important to note that this function could be used to analyze sleep loss or gain resulting from other temporary manipulations to the animal, such as heating, light pulsing, starvation, etc.

Selecting the "**Export Data?**" checkbox will export the same data that is plotted ("CumulativeSleepLossexport.csv").

Environmental Monitor Data Analysis

It is often helpful to check on the environmental conditions throughout your experimental protocol. If you included a TriKinetics Environmental Monitor in the incubator along with your experimental monitors, you can use a separate MATLAB script called EnvMonData.m to generate a graph showing light, temperature, and humidity plots for the duration of your experiment.

1. Type "EnvMonData" into the MATLAB command screen. **NOTE:** The capital letters are required.
2. A dialog box will open (see left panel of **Figure 14** below). Navigate to the folder containing 1-minute data files ("20150701-1", according to the example mentioned above).
3. Find the series of 32 channel files from the environmental monitor. Select the following 3 files simultaneously: …C04, …C09, and …C14, and run the analysis by clicking "Open". These 3 channels contain the average data for each minute of recording for Light, Temp, and Humidity, respectively. See TriKinetics documentation about these readings if further info is needed.
4. This will generate a graph (see right panel of **Figure 14** below) with Light in blue, Temperature in orange, and Humidity in green. The Light units are lux, Temp units are Celsius with one decimal point (i.e., a value of 255 = 25.5ºC), and Humidity is in %. X-axis is in minutes (1440 per day).

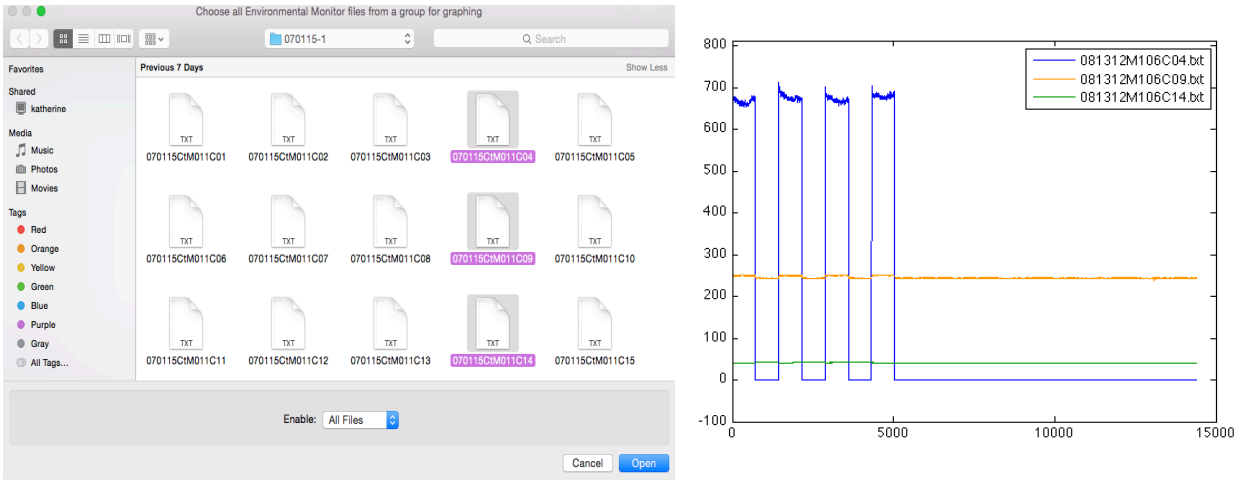

Figure 14. Environmental Analysis – Data Selection Window and Output Graph. Here are shown results from an experiment consisting of 4 days of Light/Dark conditions followed by 6 days of Dark/Dark conditions. Environmental variables are ~ 675 lux (blue), 25ºC (orange), and 40% humidity (green).