PHASE User Manual (7.0)

Pearsons J.L.1, Lopatkin A.J.2-4, Abhilash L.5, Roelofs A.6, Bell E.V.7, Fernandez M.P.7, Shafer, O.T.5

1- Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109

2- Department of Biology, Barnard College. New York, NY 10027

3- Data Science Institute, Columbia University. New York, NY 10027

4- Department of Ecology, Evolution, and Environmental Biology, Columbia University. New York, NY 10027

5- Advanced Science Research Center at The Graduate Center of The City University of New York, New York, NY

6- Technology Services, College of Literature, Science and the Arts, University of Michigan, Ann Arbor, MI 48109

7- Department of Neuroscience and Behavior, Barnard College. New York, NY 10027

Contents

[Section 1. Introduction 4](#_gjdgxs)

[Section 2. PHASE Software 5](#_30j0zll)

[2.1. Software access 5](#_1fob9te)

[2.2. Installing the standalone version 5](#_3znysh7)

[2.3. Installing MATLAB version 6](#_2et92p0)

[2.4. User-interface 6](#_tyjcwt)

[Section 3. Input and raw data settings 8](#_3dy6vkm)

[3.1. Data processing protocol 8](#_1t3h5sf)

[3.2. Data settings 8](#_4d34og8)

[Section 4. Analysis: data used, and input and output settings 10](#_2s8eyo1)

[4.1. Common export settings 10](#_17dp8vu)

[4.2. Profile analysis and visualization 11](#_3rdcrjn)

[4.2.1. Input 11](#_26in1rg)

[4.2.2. Output 12](#_lnxbz9)

[*4.2.2.1. Sleep 12*](#_35nkun2)

[*4.2.2.2. Normalized Activity 14*](#_1ksv4uv)

[*4.2.2.3. Averaged Activity 16*](#_44sinio)

[4.2.3. Graphs 18](#_2jxsxqh)

[4.3. Phase analysis 18](#_z337ya)

[4.3.1. Input 18](#_3j2qqm3)

[4.3.2. Output 19](#_1y810tw)

[*4.3.2.1. Sleep 19*](#_4i7ojhp)

[*4.3.2.2. Normalized Activity 20*](#_1ci93xb)

[*4.3.2.3. Raw Activity 20*](#_3whwml4)

[4.3.3. Graphs 20](#_2bn6wsx)

[4.4. Anticipation or Latency analysis 21](#_qsh70q)

[4.4.1. Input 21](#_3as4poj)

[4.4.2. Output 21](#_1pxezwc)

[*4.4.2.1. Sleep 21*](#_49x2ik5)

[*4.4.2.2. Normalized Activity 22*](#_2p2csry)

[*4.4.2.1. Raw Activity 23*](#_147n2zr)

[4.4.3. Graphs 23](#_3o7alnk)

[4.4.3.1. Index and Duration 23](#_23ckvvd)

[4.4.3.2. Slope 24](#_ihv636)

[Section 5. Graph tools 25](#_32hioqz)

[Section 6. Savitzky-Golay Filter 26](#_1hmsyys)

[Section 7. Troubleshoot 28](#_2grqrue)

[7.1. Troubleshoot saving and installing standalone installer on Mac OS 28](#_vx1227)

[7.2. Occasional Excel errors while saving PHASE output 28](#_3fwokq0)

[References 29](#_1v1yuxt)

# Section 1. Introduction

This manual is designed to guide users of PHASE in the analysis of *Drosophila* activity, sleep, and circadian entrainment behavior, using data collected from the *Drosophila* Activity Monitor (DAM) system (TriKinetics Inc, Waltham, MA). Section 1 describes how to download and install the PHASE software, and accepted data formats. Section 2 describes all possible analysis that the software can do and the input settings for the analysis. Section 3 describes the Savitzky-Golay filtering that will be used for several of the analysis on PHASE and Section 4 describes all the output files generated and saved by PHASE. We think that PHASE may also be a useful application outside of the *Drosophila* circadian field, provided primary behavior data is in a format read and processed by DAMFileScan.

# Section 2. PHASE Software

## 2.1. Software access

PHASE functions on both Mac (10.13 or later) and Windows (Windows 7 or later). There are two ways to run the PHASE software. Users may run it either as a standalone version or as an App within MATLAB (The MathWorks, Inc, Natick, MA). For the former, users will not require a license for MATLAB and only require MATLAB Runtime R2020 or later (runtime installer is included in the PHASE installers). However, to run PHASE as an App within MATLAB, users will require to acquire a MATLAB license and the MATLAB Signal Processing Toolbox, which contains the Savitzky-Golay filtering function, and the MATLAB Financial Toolbox, which contains the UI Calendar that is used in Data Settings of PHASE. The folders containing the standalone versions of PHASE for Mac OS and Windows may be downloaded from https://github.com/ajlopatkin/PHASE. See the example screenshot below to know which button to use for downloading the installers. For further details users can see the description in the README file or the instructions given below.

## 2.2. Installing the standalone version

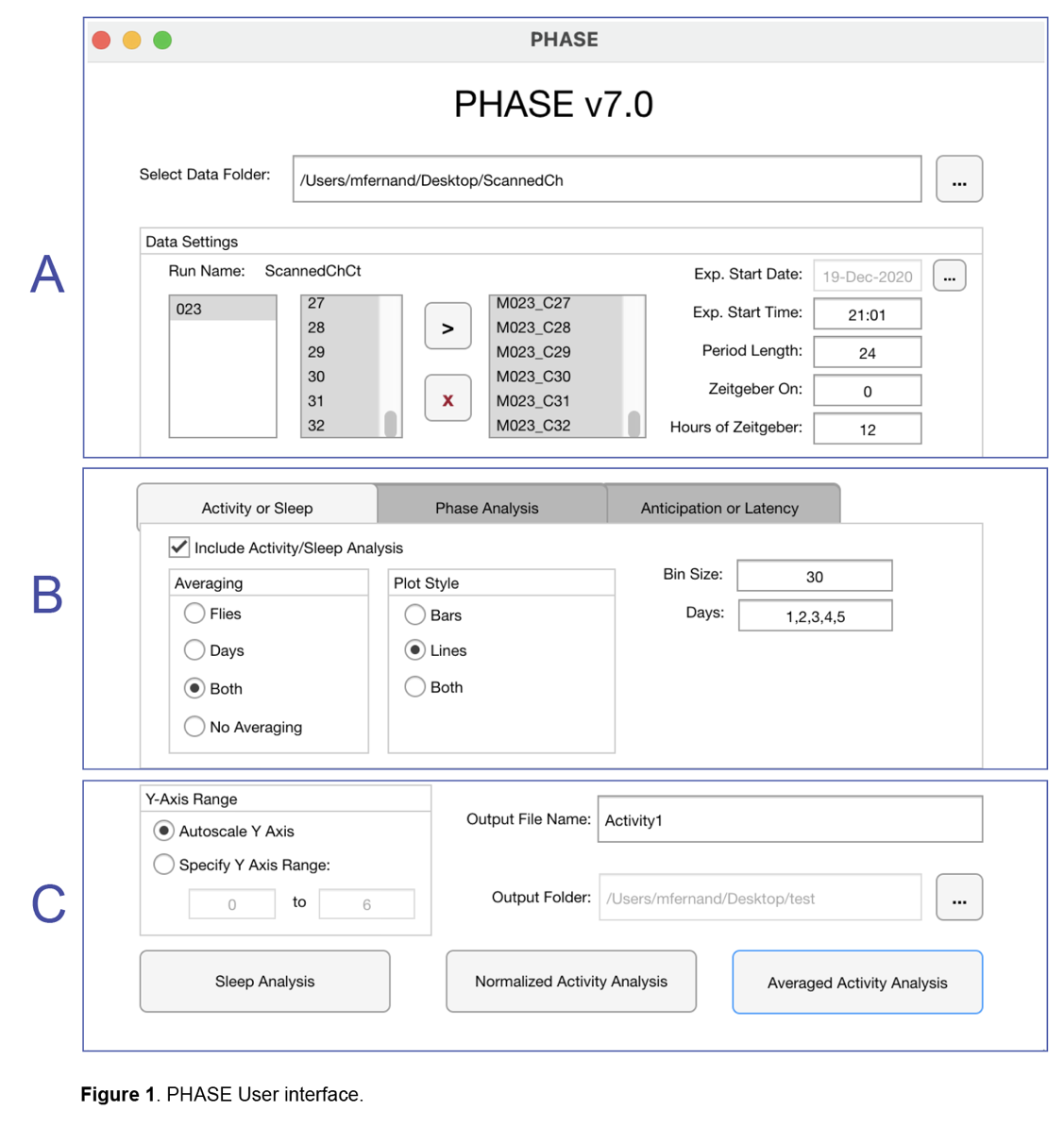
1. Navigate to the Installer Downloads folder and download the appropriate installer. Run it to install PHASE.
2. Ensure that computer security settings allow downloads from the internet. MacOS Gatekeeper may disrupt PHASE file structure and successful installation.

## 2.3. Installing MATLAB version

Acquire or update a MATLAB software license that is less than one year old.

1. Install the MATLAB Signal Processing Toolbox and the MATLAB Financial Toolbox.
2. Clone or download the entire PHASE repo from <https://github.com/ajlopatkin/PHASE>
3. Click to open PHASE.mlappinstall. MATLAB should automatically open and suggest that PHASE be installed into APPS.
4. If MATLAB does not automatically open, open MATLAB independently and drag PHASE.mlappinstall out of Downloads (or wherever the downloaded file was saved) and into the MATLAB window.
5. Confirm PHASE installation by clicking “Install”. PHASE is ready to use and may be found in the APPS tab in MATLAB.

## 2.4. User-interface

The user-interface of PHASE can be deconstructed to three parts as in Figure 1. Part A is the raw data input and data settings section. Part B is the analysis parameters input section, and Part C is the output section of PHASE. Each will be described in detail in the subsequent sections.

# Section 3. Input and raw data settings

## 3.1. Data processing protocol

1. Use [DAMFileScan](https://trikinetics.com/) (Trikinetics Inc, Waltham, MA) to process DAM monitor data into 1-minute bins, in channel (individual fly) counts for the duration of the experiment before trying to upload to PHASE.

**Note:** The 1-minute bin criterion is essential only for analyzing sleep.

1. Place all data from a single experiment into a folder.
   1. All DAM-processed data must have the same file name, and same experiment start and end bins.
   2. Experiments originating from different dates or with different file names must be processed by PHASE from separate data folders.
2. Ensure that no other files are in this folder as PHASE will not be able to read the first experiment start bin.
   1. This includes all folders generated by PHASE.

## 3.2. Data settings

1. Select Data Folder containing DAMFileScan processed 1-minute binned data using “…” on the top right side (Figure 1, A).
2. Run Name, Exp. Start Date and Exp. Start Time should autofill. The Exp. Start Date and Exp. Start Time should be the exact same date and time of the first bin in the saved channel files.

Note: Data should be analyzed starting on the first bin, not on the start time. For example, for an experiment that starts 9PM (21:00) and is saved in 1-minute bins, the start time should be 21:01.

1. The left most box displays all the Monitors in a given run. The box on the right next to the left most box displays all 32 channels of each monitor. At any given time only channel files from a single monitor must be loaded for analysis.
2. Highlight the channels/individual flies from the second box to be processed from the auto-loaded monitors. To select multiple channels, hold control or command + shift and click desired channels.
   1. Use the green > arrow to push the selected channel files to the analysis box (third box). Make sure that all relevant channels are selected before performing any analysis (as shown in Figure 1), otherwise only the first channel will be analyzed.
   2. Use the red “X” to delete all selected channel files from the analysis box.
   3. Individual channels may also be deleted after loading all 32 channels by selecting the ones that are not of interest and pressing the delete button on the keyboard, not the red “X”. All desired channels must be selected in the top right panel.

**Note:** Failure to autofill results will occur if there are (i) other files in the Data Folder or (ii) if there are monitor files with multiple start bin dates/times, or (iii) multiple file names. Ensure that only monitor files processed by DAMFileScan with the same set of continuous bins, and same name occupy the Data Folder.

1. Change the auto-filled “Exp. Start Date” using the “…” button from the right side of Section A of Figure 1, to select the date at which analysis must begin.
2. Change auto-filled “Exp. Start Time” to the local time at which Zeitgeber comes on (in other words ZT 00) in the HH:MM format. For instance, if lights turn on in an LD cycle at 10 AM, then the Exp. Start Time must be 10:00.
3. Change “Period Length”, “Zeitgeber ON”, and “Hours of Zeitgeber” as needed. “Period Length” is the duration of one cycle in hours. “Zeitgeber ON” refers to the time at which Zeitgeber turns on, in the Zeitgeber Time scale. For instance, if lights turn on at 10 AM, then “Zeitgeber ON” must be 0. “Hours of Zeitgeber” refers to the duration of the entire cycle that the zeitgeber was in the ON state, in hours, e.g., if flies are recorded under LD 12:12 then “Hours of Zeitgeber” will be 12.
4. Auto-fills for “Period Length”, “Zeitgeber ON” and “Hours of Zeitgeber” are “24”, “0” and “12”, respectively. Note that “Day” statistics are derived from the hours between “Zeitgeber ON” and OFF (OFF times are estimated using the “Hours of Zeitgeber”). “Night” statistics are derived from the remaining hours.

Adjust these parameters to quantify sleep or activity during a particular window. For example, to quantify behavior only between ZT3-8, change the “Zeitgeber ON” to “3” and enter “5” for “Hours of Zeitgeber”. All “Day” sleep and activity statistics in Excel outputs generated by PHASE will result from the ZT3-8 window specified.

# Section 4. Analysis: data used, and input and output settings

PHASE can analyze and visualize profiles, objectively compute phases and estimate latency and anticipation of either sleep data, normalized activity data, simply raw activity data, some combination of these or all of these. In order to analyze sleep data users must click the “Sleep Analysis” on the bottom left side (Figure 1, C). In order to analyze normalized activity data or simply raw activity data, users must click the “Normalized Activity Analysis” or “Averaged Activity Analysis” buttons in the middle and right side, respectively (Figure 1, C).

In each of the above described case, PHASE bins the raw data into desired intervals, averages activity/sleep across cycles, individuals, and both cycles and individuals, exports all this data as MS-Excel files, and plots these data and exports these figures as MATLAB compatible figure files and a PDF file. Further, PHASE also uses a Savitzky-Golay filter to smooth raw data and objectively predicts phases, and computes measures of sleep latency and activity anticipation. All these results are exported as described above. Each of these analysis methods requires specific analysis inputs and these will be described in the following sub-sections.

## 4.1. Common export settings

Before delving into details of each of the analysis methods mentioned above, we will introduce the users to Part C of Figure 1.

1. Users must choose *Y*-axis range to appear in all the output plots of PHASE. Users can either choose to allow PHASE to automatically adjust *Y*-axis or can specify *Y*-axis range (Figure 1, C, left).
2. Users must also select an “Output Folder” where all MS-Excel sheets and plots will be saved.
3. Users can also choose to provide an “Output File Name” that uniquely identifies the set of analysis run by PHASE, but this is not necessary to provide.

**Note:** Create a new folder for PHASE’s MS-Excel and graph outputs.

## 4.2. Profile analysis and visualization

A single bout of “sleep” is defined as 5 consecutive 1-minute bins with zero activity (as measured using IR beam crosses) (Hendricks and Sehgal, 2004; Huber et al., 2004). PHASE quantifies sleep this way, and to perform analyses, the “Activity or Sleep” tab requires the following input. To perform this analysis, check the checkbox “Include Activity/Sleep Analysis” under the “Activity or Sleep” tab (Figure 1, B).

### **4.2.1. Input**

1. Tell PHASE whether it must calculate sleep by averaging over either “Flies” or “Days” or “Both” or “No Averaging”. Outputs that each of these will produce will be described in the next section.
   1. Averaging by “Flies” averages behavior of all selected flies over indicated “Days” and provides a single graph with each day of data displayed consecutively (in the form of a time-series).
   2. Averaging by “Days” averages over indicated “Days” for each selected fly. Results in a single day-averaged plot.
   3. “Both” results in a single plot that first averages all flies’ behavior for each fly, then by days.

**Note:** When “Averaging” by “Both” days and flies is selected, averaging is done first by flies and then (in a separate step) by days. The plots generated, therefore, represent SEM across cycles, and may be used for preliminary visualization. Final figures must be made such that error bars represent variation across individuals, as they represent true replication.

* 1. “No Averaging” creates one time-series for every fly.

1. Indicate the desired nature of profile plots. Available choices, as users can see, are bar graphs, line plots and both.
   1. All styles will center “Zeitgeber ON” hours between Zeitgeber OFF hours (shaded bars or regions) at the left and right of the graph.
   2. The exceptions are (i) “No Averaging” and (ii) “Flies”, where multiple days are selected, which will display each day consecutively beginning with the first day’s Zeitgeber ON to last day’s Zeitgeber OFF.
   3. “Bars” creates a traditional bar graph displaying binned activity or sleep data with standard error of mean (SEM) error bars.
   4. “Lines” creates a line graph with SEM shaded above and below the mean line.
   5. “Both” creates plots with both graph styles.

**Note:** The SEM is calculated on the second step only, that is, it's the standard error across flies.

1. Indicate the desired binning interval in the output by entering value in the “Bin Size” box.

**Note:** The bin size must be in minutes. For instance, bin size of 1-h must be input as 60.

1. Indicate the days of analysis that PHASE must include for analysis of sleep. Multiple days can be input here, each day separated by a comma (for instance, 1,2,3). The days don’t necessarily have to be consecutive. For example, 1,2,3,5 is a valid input.

**Note:** Day “1” begins at the “Exp. Start Time” for the indicated “Day Length” with “Day” beginning at “Zeitgeber ON” and “Night” beginning after the given “Hours of Zeitgeber” in “Data Settings”. Day “2” begins the indicated “Day Length” after day “1”.

### **4.2.2. Output**

#### 4.2.2.1. Sleep

Sleep Analysis quantifications are summarized in the Excel file:

“SleepAnalysis\_OutputFileName\_YYYMMDD\_HHMM”. The following is a description of the sheets that are created by PHASE in the MS-Excel file, in the order described.

1. SummedStatistics (Raw)

This MS-Excel sheet contains the following columns of data, values of which are summed over all indicated days in the “Days” input. This sheet is the same irrespective of the averaging option chosen in the analysis input settings.

* 1. *Board:* Monitor number (1-120)
  2. *Fly:* Channel number (1-32)
  3. *TotalMinutes:* Sum of all data intervals with non-zero sleep values multiplied by the length of that interval (in minutes).
  4. *AverageMinutes:* The “TotalMinutes” divided by the total number of data collection bins.
  5. *TotalBouts:* The total number of bouts of sleep. A bout is defined as a series of non-zero values of any length.

**Note:** Because each data bin has been marked as sleep or not sleep and since only bins within a minimum 5-minute series are marked sleep, there's an effective minimum of 5 minutes for a bout.

* 1. *AverageBoutDuration:* The sum of the bout durations in minutes divided by the total number of bouts.
  2. *DayTotal:* Total minutes of sleep during “Zeitgeber ON” hours.
  3. *AverageDayMinutes:* The “DayTotal” divided by the total number of data collection bins in duration between “Zeitgeber ON” and Zeitgeber OFF.
  4. *DayBouts:* The total number of bouts beginning during “Zeitgeber ON” hours.
  5. *AverageDayBoutDuration:* The total minutes of sleep in bouts that began during “Zeitgeber ON” hours divided by the total number of bouts that during the same time.
  6. *NightTotal:* Total minutes of sleep outside the “Zeitgeber ON” hours.
  7. *AverageNightMinutes:* The “NightTotal” divided by the total number of data collection bins outside of the “Zeitgeber ON” hours.
  8. *NightBouts:* The total number of bouts beginning outside the “Zeitgeber ON” hours.
  9. *AverageNightBoutDuration:* The total minutes of sleep in bouts that began after “Zeitgeber ON” hours divided by the total number of bouts that during the same time.

1. DayAvgStatistics (Raw)

This sheet contains all the column-wise information as described above, but averaged over cycles, instead of sum. This sheet is the same irrespective of the averaging option chosen in the analysis input settings.

1. BinnedData (Raw)

Contents of this sheet will depend on what averaging option was chosen.

* 1. When averaged over “Flies”, data are presented at time-series averaged over all individuals. The first time point in the far-left column will be ZT0.
  2. When averaging over “Days”, data are centered around “Zeitgeber ON” hours. Therefore, the first, far-left bin in Column A is not ZT0. Each subsequent column has individual fly sleep data averaged over all the days. When PHASE reports individual fly data, it stores the identity by using the monitor number and channel number. For instance, channel 2 of monitor 13, will be represented as M013C02.
  3. When averaging over “Both”, flies and days, a single column of data is presented, centered around ZT0.
  4. When “No Averaging” is selected, raw time series through the selected days for every fly is reported.

**Note:** Time is referred to as t\_Xmin. For instance, ZT0 is referred to as t\_0min, and so on.

1. Settings

This sheet records all the settings utilized for the specific analysis session.

1. Daily data sheets

All subsequent sheets record day-wise sleep statistics (Day *X* Statistics (Raw)) as described above and sleep profiles (Day *X* BinnedData (Raw)).

**Note:** If the user chooses to analyze more than 6 days of data, then a different MS-Excel file is generated to store all the data for the 7th cycle onwards.

#### 4.2.2.2. Normalized Activity

Exports from “Normalized Activity Analysis” are summarized in the MS-Excel file “NormalizedActivityAnalysis\_OutputFileName\_YYYMMDD\_HHMM”. The following is a description of the sheets that are created by PHASE in the MS-Excel file, in the order described.

1. SummedStatistics (Raw)

This MS-Excel sheet contains the following columns of data, values of which are summed over all indicated days in the “Days” input. This sheet is the same irrespective of the averaging option chosen in the analysis input settings.

* 1. *Board:* Monitor number (1-120)
  2. *Fly:* Channel number (1-32)
  3. *TotalCounts:* The sum of all counts of activity in the data series for the selected days.
  4. *AverageCounts:* The “TotalCounts” divided by the total number of user defined bins. For example, a 30-minute “Bin Size” over a 24-hour day would divide all “TotalCounts” values by 48.
  5. *DayTotalCounts:* The total IR beam crosses that took place during “Zeitgeber ON” hours.
  6. *AverageDayCounts:* The “DayTotalCounts” divided by the total number of data collection bins in “Zeitgeber ON” hours.
  7. *NightTotalCounts:* The total IR beam crosses that took place outside of the “Zeitgeber ON” hours.
  8. *AverageNightCounts:* The “NightTotalCounts” divided by the total number of data collection bins outside of “Zeitgeber ON” hours.

1. DayAvgStatistics (Raw)

This sheet contains all the column-wise information as described above, but averaged over cycles, instead of sum.

1. BinnedData (Raw)

Contents of this sheet will depend on what averaging option was chosen.

* 1. When averaged over “Flies”, data are presented at time-series averaged over all individuals. The first time point in the far-left column will be ZT0.
  2. When averaging over “Days”, data are centered around “Zeitgeber ON” hours. Therefore, the first, far-left bin in Column A is not ZT0. Each subsequent column has individual fly activity data averaged over all the days. When PHASE reports individual fly data, it stores the identity by using the monitor number and channel number. For instance, channel 2 of monitor 13, will be represented as M013C02.
  3. When averaging over “Both”, flies and days, a single column of data is presented, centered around ZT0.
  4. When “No Averaging” is selected, raw time series through the selected days for every fly is reported.

**Note:** Time is referred to as t\_Xmin. For instance, ZT0 is referred to as t\_0min, and so on.

1. SummedStatistics

SUM of activity for each fly over the time window chosen (number of days).

1. DayAvgStatistics

AVERAGE of activity for each fly over the time window chosen (number of days).

1. BinnedData

Contents of this sheet are in the same format as that in the sheet BinnedData (Raw). This contains activity counts normalized to total counts of activity within that day.

1. Settings

This sheet records all the settings utilized for the specific analysis session.

1. Daily data sheets

All subsequent sheets record day-wise activity statistics (Day *X* Statistics (Raw)) as described above and activity profiles (Day *X* BinnedData (Raw)). Additionally, daily data sheets also include Day *X* Statistics and Day *X* BinnedData. These report data in the same format as described above.

**Note:** If the user chooses to analyze more than 6 days of data, then a different MS-Excel file is generated to store all the data for the 7th cycle onwards.

#### 4.2.2.3. Averaged Activity

Exports from “Averaged Activity Analysis” are summarized in the MS-Excel file “AveragedActivityAnalysis\_OutputFileName\_YYYMMDD\_HHMM”. The following is a description of the sheets that are created by PHASE in the MS-Excel file, in the order described.

1. SummedStatistics (Raw)

Contents of this sheet are the exact same as described under the Normalized Activity section.

1. DayAvgStatistics (Raw)

Contents of this sheet are the exact same as described under the Normalized Activity section.

1. BinnedData (Raw)

Contents of this sheet are the exact same as described under the Normalized Activity section.

1. SummedStatistics
2. DayAvgStatistics
3. This sheet records all the settings utilized for the specific analysis session.
4. BinnedData

Contents of this sheet are in the same format as that in the sheet BinnedData (Raw). This contains total activity counts in the user-defined bin divided by the user-defined bin size. In other words, values represent the average activity in a given bin.

**Note:** It is not advisable to use this measure for downstream analysis, because these measurements do not preserve the shape of the waveform. They are likely to flattens/broaden peaks, and therefore, may be best utilized for first-pass visualization and not statistical analysis.

1. Settings

This sheet records all the settings utilized for the specific analysis session.

1. Daily data sheets

All subsequent sheets record day-wise activity statistics (Day *X* Statistics (Raw)) as described above and activity profiles (Day *X* BinnedData (Raw)). Additionally, daily data sheets also include Day *X* Statistics and Day *X* BinnedData. These report data in the same format as described above.

**Note:** If the user chooses to analyze more than 6 days of data, then a different MS-Excel file is generated to store all the data for the 7th cycle onwards.

### **4.2.3. Graphs**

Graphs generated by PHASE represent activity or sleep data collected in the “Bin Size” specified for each individual fly included in “Data Settings”. *Y*-values on the activity graphs represent either activity data “Averaged” by the “Bin Size” or “Normalized” to total activity. Zeitgeber ON hours (“Day”) are centered between Zeitgeber OFF hours (“Night”) at the left and right of the graph except where multiple days of sleep are displayed consecutively in “Averaging” by “Flies” or “No Averaging”. Bar and line graphs represent Zeitgeber OFF hours with darker bars or shaded background regions, respectively. Bar graphs display SEM as error bars above the mean activity/sleep bar. Line graphs shade the SEM area above and below a darker, mean activity/sleep line. When “Averaging” by “Both” days and flies is selected, averaging is done first by flies and then (in a separate step) by days. The SEM is calculated on the second step only, that is, it's the standard error across days of the averaged fly data (this error must not be used for comparisons across genotypes).

## 4.3. Phase analysis

“Phase Analysis” applies a Savitzky-Golay filter (see Section 6 for details on this filter and how it is used) to either raw IR beam sleep data or activity data or normalized activity data.

**Note:** PHASE only detects phases of peaks, and no other circadian phase markers such as onset, offset and trough.

### **4.3.1. Input**

1. Indicate that PHASE must analyze phases by checking the box labelled “Include Phase Analysis”.
2. Indicate the “Filter Order” to be used (see Section 6 for clarity on “Filter Order”).
3. Indicate the “Filter Frame Length (minutes)” to be used (see Section 6 for clarity on “Filter Frame Length”).

**Note:** Filter frame length must be an odd number.

1. Indicate the ZT around which phases must be called. Multiple ZT values can be used separated by commas.
2. Tell PHASE what the “Minimum Distance Between Peaks” must be. This value must be provided in minutes. PHASE uses the MATLAB function *findpeaks* to find local maxima within the “Minimum Distance Between Peaks” of filtered data. The “Minimum Distance” restricts the *findpeaks* function to an acceptable peak-to-peak distance (in minutes), effectively ignoring any peaks which are close together. Therefore, smaller minimum distances will allow PHASE to find many, small peaks. Larger minimum distances will favor larger, broader peak features.
3. Tell PHASE data from which days must be used to call phases. Multiple days can be used as input, separated by commas. Where multiple days are specified in “Days” averaging takes place prior to filter application.

**Note:** If day wise phases are required, users must analyze one day at a time.

### **4.3.2. Output**

#### 4.3.2.1. Sleep

Phase calls on sleep data are summarized in the Excel file “SleepPhase\_OutputFileName\_YYYMMDD\_HHMM”. Values reported in this Excel file are the same irrespective of the averaging option selected in the analysis input settings.

1. Phase

This sheet contains three columns for every ZT chosen in the phase analysis input section. The values for each ZT are in descending order of the ZT value. For instance, in two ZT values are used in the input, i.e., 0,12. Columns B, C and D will report Peak Time (ZT), Peak Height and Peak Area, respectively around ZT12. Columns E, F and G will report the same for ZT0, and so on. Column A will always be the monitor and channel identity of every fly in the format M*XXX*C*YY*, such that a fly from monitor 1, channel 3 will be labelled as M001C03.

*Peak Time (ZT):* PHASE picks the time (ZT) at which the maximum sleep occurs. ZT is computed relative to provided input in the “Data Settings”.

*Peak Height:* PHASE uses the function *findpeaks* to also determine the peak height/prominence (in minutes of sleep) and the peak width/half-prominence (in minutes)

*Peak Area:* This is calculated as the product of Peak Height and Peak Width.

**Note:** PHASE can find, quantify and visualize all peaks within smoothed/filtered behavior data. However, it is advisable that users search for peaks closest to given ZT, or set of ZT(s).

1. Settings

This sheet records all the settings utilized for the specific analysis session.

#### 4.3.2.2. Normalized Activity

Phase calls on 1-minute binned normalized (by total daily activity) activity data are summarized in the Excel file “NormalizedActivityPhase\_OutputFileName\_YYYMMDD\_HHMM”. Values reported in this Excel file are the same irrespective of the averaging option selected in the analysis input settings.

1. Phase

Contents of this sheet are in the exact same format as described for the output of sleep data as described above.

1. Settings

This sheet records all the settings utilized for the specific analysis session.

#### 4.3.2.3. Raw Activity

Phase calls on raw activity data (IR beam crosses in 1-minute intervals) are summarized in the Excel file “ActivityPhase\_OutputFileName\_YYYMMDD\_HHMM”. Values reported in this Excel file are the same irrespective of the averaging option selected in the analysis input settings.

1. Phase

Contents of this sheet are in the exact same format as described for the output of sleep data as described above.

1. Settings

This sheet records all the settings utilized for the specific analysis session.

### **4.3.3. Graphs**

Graphs generated from this analysis represent 1-minute binned activity, normalized activity, or sleep data for each individual fly included in “Data Settings” for the day or average of days indicated in “Days to Use”. The blue line represents the filtered data curve of the 1-minute data. Graphs shade each peak area in blue and label the peak maximum with a blue carrot and the ZT value.

## 4.4. Anticipation or Latency analysis

“Anticipation or Latency Analysis” applies a Savitzky-Golay filter (see Section 6 for details on this filter and how it is used) to either raw IR beam sleep data or activity data or normalized activity data.

### **4.4.1. Input**

1. Indicate that PHASE must analyze anticipation or latency by checking the box labelled “Include Anticipation or Latency Analysis”.
2. Indicate the “Filter Order” to be used (see Section 6 for clarity on “Filter Order”).
3. Indicate the “Filter Frame Length (minutes)” to be used (see Section 6 for clarity on “Filter Frame Length”).

**Note:** Filter frame length must be an odd number.

1. Indicate “Window Length (minutes)” to be used. All anticipation and/or latency analyses are performed within this user-defined window length.
2. Indicate the ZT around which phases must be called. Multiple ZT values can be used separated by commas.
3. Tell PHASE data from which days must be used to analyse anticipation and/or latency. Multiple days can be used as input, separated by commas. Where multiple days are specified in “Days” averaging takes place prior to filter application.

**Note:** If day wise data for anticipation and/or latency are required, users must analyze one day at a time.

### **4.4.2. Output**

#### 4.4.2.1. Sleep

Data are summarized in the Excel file “SleepLatency\_OutputFileName\_YYYMMDD\_HHMM”. Values reported in this Excel file are the same irrespective of the averaging option selected in the analysis input settings. Wherever multiple days are chosen for analysis, averaging is performed before filtering or calculating slopes. Analysis of latency around each user-defined ZT is provided in a separate Excel sheet. The columns in each of these sheets contain the following data in the following order.

1. The left-most column has monitor and channel identity of every fly in the format M*XXX*C*YY*, such that a fly from monitor 1, channel 3 will be labelled as M001C03.
2. *Min Sleep Time (ZT):* PHASE moves positively along the *x*-axis (time) from a user-defined ZT to derive the first minima/least amount of sleep from filtered data. The time (ZT) at which this occurs is reported in this column.
3. *Max Sleep Time (ZT):* PHASE moves positively along the *x*-axis from a user-defined ZT to derive the maxima/greatest amount of sleep from filtered data. The time value (ZT) is reported in this column.
4. *Latency (minutes):* PHASE determines the duration in minutes between defined maxima and minima and reports this latency value for each fly in this column.
5. *Index (AUC):* This column documents area under the curve (AUC) of the filtered data between the minima and maxima defined above, for every individual.
6. *Slope:* PHASE also finds the slope of a linear regression fitted to unsmoothed sleep data for each user-defined ZT within the indicated “Window Length (minutes)”.

#### 4.4.2.2. Normalized Activity

Anticipation analysis performed on Savitzky-Golay filtered 1-minute binned normalized activity data (normalized by total daily activity) are summarized in the Excel file “NormalizedActivityAnticipation\_OutputFileName\_YYYMMDD\_HHMM”. Values reported in this Excel file are the same irrespective of the averaging option selected in the analysis input settings. Wherever multiple days are chosen for analysis, averaging is performed before filtering or calculating slopes. Analysis of anticipation around each user-defined ZT is provided in a separate Excel sheet. The columns in each of these sheets contain the following data in the following order.

1. The left-most column has monitor and channel identity of every fly in the format M*XXX*C*YY*, such that a fly from monitor 1, channel 3 will be labelled as M001C03.
2. *Min Activity Time (ZT):* PHASE moves negatively along the *x*-axis (time) from a user-defined ZT to derive the first minima/least amount of activity from filtered data. The time (ZT) at which this occurs is reported in this column.
3. *Max Activity Time (ZT):* PHASE moves negatively along the *x*-axis from a user-defined ZT to derive the maxima/greatest amount of activity from filtered data. The time value (ZT) is reported in this column.
4. *Anticipation (minutes):* PHASE determines the duration in minutes between defined maxima and minima and reports this anticipation value for each fly in this column.

**Note:** PHASE excludes the “startle effect” by gathering maxima and minima from the first bin after a given ZT.

1. *Index (AUC):* This column documents area under the curve (AUC) of the filtered data between the minima and maxima defined above, for every individual.
2. *Slope:* PHASE also finds the slope of a linear regression fitted to unsmoothed normalized activity data (1-minute binned) for each user-defined ZT within the indicated “Window Length (minutes)”.

#### 4.4.2.1. Raw Activity

Anticipation analysis performed on Savitzky-Golay filtered 1-minute binned raw activity data are summarized in the Excel file “ActivityAnticipation\_OutputFileName\_YYYMMDD\_HHMM”. Values reported in this Excel file are the same irrespective of the averaging option selected in the analysis input settings. Wherever multiple days are chosen for analysis, averaging is performed before filtering or calculating slopes. Analysis of anticipation around each user-defined ZT is provided in a separate Excel sheet. The columns in each of these sheets contain the same contents as those described above for the Normalized Activity section.

### **4.4.3. Graphs**

#### 4.4.3.1. Index and Duration

Graphs represent 1-minute binned raw activity data (when “Averaged Activity Analysis” is selected), normalized activity data (“Normalized Activity Analysis”), or sleep data for each individual fly included in “Data Settings” for the day or average of days indicated in “Days”. Each region included in the analysis is shaded in light grey. The blue line represents the filtered data curve. Blue carrots indicate the filtered activity or sleep maximum and minimum bins determined by PHASE within the “Window Length” from each ZT. The duration (minutes) between each maximum and minimum is annotated for each region above the blue carrot.

#### 4.4.3.2. Slope

Graphs represent slopes calculated from unsmoothed 1-minute raw activity data (“Averaged Activity Analysis”), normalized activity data (“Normalized Activity Analysis”) or sleep data within the analysis window of a given ZT for all flies included in “Data Settings”. As sleep values range from 1 or 0 for a given bin, these values are far more discreet than raw activity values (presented as unsmoothed IR beam crosses). For example, given the 0-or-1 sleep values, for three days of activity there are only four possible average values: 0, 1/3, 2/3, or 1 depending how many of those days the fly was asleep during each interval. Each ZT value included in the “Anticipation or Latency Analysis” is plotted individually. The linear regression/slope for individual flies is in light grey. The average linear regression/slope for all flies included in analysis is in dark grey.

# Section 5. Graph tools

Once a graph has been generated by PHASE, a window containing it will automatically open. Clicking on the graph will display a number of graph tools on the top of the graph. A description of these tools follows.

1. House: Restore original graph, undo any changes.
2. Magnifying glasses: Zoom in and out.
3. Hand: Move the graph.
4. Data tip: Gives exact (*x,y*) coordinates of selected data point.

To remove data tips, right click the data tip box and select ‘remove data tip’.

1. Paintbrush: Highlight specific data, e.g., a bar.
2. Arrow with box: Save or copy graph.

To further edit using MATLAB copy as vector graphic.

For a fixed image of the graph copy as image.

1. Legend: In the toolbar above the graph, select the white box with red and blue squares.

Customize by right clicking.

# Section 6. Savitzky-Golay Filter

PHASE uses the MATLAB function *sgolay* from the Signal Processing Toolbox to filter 1-minute activity or sleep data with a Savitzky-Golay FIR smoothing filter of a given “Filter Order” and “Filter Frame Length”. Savitzky-Golay filtering is optimal for smoothing data with a large frequency span (such as 1-minute activity data); reducing signal-to-noise while preserving the original shape of signals that are often distorted using standard averaging filters (Orfanidis, 1995). *sgolay* replaces successive central points in frames (minutes of behavior within the “Filter Frame Length”) with the weighted average of all measures within the frame. The weighted coefficients are obtained by fitting all frame values to a polynomial of a given degree (“Filter Order”) using the linear least squares method.

PHASE applies a low-order polynomial function specified by the “Filter Order” across an odd-number of minutes of activity or sleep behavior within the “Filter Frame Length”. The “Filter Order” and “Filter Frame Length” is user-specified in “Phase Analysis”, and “Anticipation or Latency Analysis”. Therefore, it is imperative that users find the best representation of activity and sleep data that minimizes noise and signal distortion and maximizes signal. This is an iterative process until the final polynomial order and frame length are defined. To best utilize PHASE’s smoothing functions, users may begin by generating activity or sleep graphs using the autofilled parameters for “Filter Order” and “Filter Frame Length”. After visual inspection, users should apply the following principles of Savitzky-Golay filtering in order to more accurately represent their data and avoid generating erroneous estimates of phase and anticipation and/or latency.

**Notes:**

1. The polynomial degree in “Filter Order” must not exceed the minutes defined by “Filter Frame Length”.
2. The “Filter Frame Length” must be an odd number. Savitzky-Golay methodology assumes that the data within the specified frame are equidistant. Symmetry requires the window to contain an odd number of points in order to replace a central point by the best representation of all values within the frame.
3. Signal-to-noise generally increases as the polynomial degree decreases. A higher “Filter Order” will best preserve narrow feature height and width but perform less well on broad peaks. A lower “Filter Order” conversely, will smooth broad features, at the expense of minimizing narrow features.
4. Signal-to-noise increases as “Filter Frame Length” increases as there are more data points over which to derive the weighted average of the center point. Increased frame length also increases original signal distortion, however.
5. The ratio between the “Filter Order” and “Filter Frame Length” controls original data distortion. Smoothing generally increases as the ratio between the two values increases.

# Section 7. Troubleshoot

## 7.1. Troubleshoot saving and installing standalone installer on Mac OS

1. Complete one of the following steps before downloading the PHASE Installer on MacOS:
   1. Allow downloads from unknown.
   2. Open “System Preferences” by pressing the Apple logo on the top left of the screen.
   3. In “Security & Privacy”, under general, check the box to allow apps to be downloaded from “Anywhere”.
   4. If this option is hidden on MacOS version, open the “Terminal” and type this line of code: sudo spctl –master-disable. Click the return button. Enter the administrator password.
   5. If unsuccessful with step (d), first download the program and then reopen the “System Preferences” window. At the bottom of the window will be the option for “Open Anyways”.

## 7.2. Occasional Excel errors while saving PHASE output

This may be due to running excel processes.

1. Go to Ctrl+Alt+Del and start Task Manager.
2. End all EXCEL processes
3. Try running PHASE again.

# References

1. Hendricks, J. C., & Sehgal, A. (2004). Why a fly? Using Drosophila to understand the genetics of circadian rhythms and sleep. Sleep. https://doi.org/10.1093/sleep/27.2.334
2. Huber, R., Hill, S. L., Holladay, C., Biesiadecki, M., Tononi, G., & Cirelli, C. (2004). Sleep homeostasis in Drosophila melanogaster. Sleep. https://doi.org/10.1093/sleep/27.4.628
3. Orfanidis, S. J. (1995). Introduction to Signal Processing. In Introduction to Signal Processing. https://doi.org/10.1109/TCOM.1972.1091274