ANALYSIS OF RHYTHMS USING R: CHRONOMICS ANALYSIS TOOLKIT (CAT)

A CAPSTONE PROJECT
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF UNIVERSITY OF MINNESOTA
BY

Cathy Lee Gierke

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF BIOLOGICAL SCIENCE

JULY 2013
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Germaine Cornélissen-Guillaume, my thesis advisor, for generously sharing her valuable time with me, and especially for sharing the many amazing stories of chronobiometric research over decades. And for her very precise late night email threads on resolution and P!

I would also like to thank Dr. Deborah Powell for kindly allowing me to use her ambulatory blood pressure monitors so that I could learn using my own blood pressure data. And my thanks for her time and effort on my review committee.

I am indebted to Dr. Alessandro Bartolomucci for introducing me to the problem of chronomic analysis, and for his guidance on the initial stages of this work. And my thanks for his time and effort on my review committee.

I would like to thank Dr. Brad Fruen, MBS advisor extraordinaire, who juggled and coordinated and advised, and made the whole process go smoothly!

There are no words to express my appreciation to my friend John Lindgren for tirelessly and selflessly supporting my interests by studying with me, and listening to my long and complicated explanations of all the fascinating things I was learning. And for always being there to help sort out the mathematics. And for clearly and concisely channeling the needed revisions for this paper.

My deepest appreciation to my loving husband, Chuck, for being understanding throughout, as I took one class a semester for over ten years, and worked full time.... and for agreeing to wear a blood pressure monitor 24x7 to obtain data for this project!
DEDICATION

To my nieces, whom I hope will pursue life with spirit, and knowledge for the power it confers to improve the human condition.

And to Dr. Franz Halberg (7/5/1919 – 6/9/2013), in memory of his amazing journey into biological rhythms, and his foresight in understanding how fundamental their complex inter-weavings are to our health, and to every system in the body. I was very honored to meet him and discuss chronobiology with him, and to have him take an interest in my study, even offering to review this thesis from his hospital bed! He continued doing the work he saw as so important until the very end, sharing his zeal for chronotherapy, and chronomics in general, with the world. What an impact he has made on the world!

Non est vivere sed valere vita est (Martial).
ABSTRACT

A network of biological oscillators modulates genetic, molecular, physiological, and behavioral rhythms. A better understanding of the role rhythms play in organizing these complex systems is dependent on ready access to flexible, robust computational tools, standardized to allow comparison of rhythms across a range of studies. This paper introduces a powerful new toolkit, the Chronomics Analysis Toolkit (CAT), for analysis of biological time series based on well-established methods of digital signal analysis. Although this paper focuses on applications in biology, the routines in CAT are appropriate for assessment of rhythm in any type of data. Because satisfactory results with these specialized statistical and analytical procedures are attendant on certain mathematical presumptions related to data-collection methodology and preparation of the datastream, these related topics are introduced with references for additional guidance. The CAT toolkit includes a smoothing function and an actogram, which permit visual inspection of data and initial (macroscopic) visualization of rhythm. Autocorrelation and crosscorrelation can reveal the presence of rhythms. Although limited, and sometimes difficult to interpret, they are important aids in understanding behavior and properties of a time series. The periodogram is another well-understood and reliable tool to assess rhythmicity and estimate period, in CAT. Another tool presented is the cosinor, comprising a set of parametric methods based on regression techniques for rhythm assessment that have several advantages over the non-parametric methods in CAT. Cosinor, because it is parametric, does not require equidistant data, often unattainable in monitoring of living systems. And it lends itself well to hypothesis testing, providing rigorous, easily interpreted quantitative assessments of mean, amplitude and phase at an assumed period, with a measure of uncertainty for each parameter. For rhythm detection and period estimation, I show that a cosinor-based periodogram is identical to the Fourier periodogram when data are equidistant. Of particular interest is the capability to assess non-stationary data using CAT, since many techniques assume stationarity, and biological data are rarely stationary. A progressive analysis assesses subsections of the data stream, in successive increments through the record, identifying changing rhythm dynamics over time. The functions included in this toolkit are flexible and can be extended or adapted to suit individual needs, owing in part to the open source nature of the code, a library of easily configurable batch files, and flexible analysis parameters. Detailed instructions and illustrative examples are provided in an accompanying website.
# TABLE OF CONTENTS

INTRODUCTION......................................................................................................................... 1

DEFINITION OF RHYTHM ........................................................................................................... 6

PREPARATION AND SIGNAL ACQUISITION .............................................................................. 9
  Discrete or Continuous Sampling: ............................................................................................. 9
  Equidistant or Non-equidistant Samples: .................................................................................. 10
  Missing Data and Interpolation: ............................................................................................... 11
  Assessing Non-stationary Data: ............................................................................................... 12
  Non-sinusoidal Signals: ............................................................................................................ 14
  Resolution: .................................................................................................................................. 15
  Fourier Frequencies: .................................................................................................................. 15
  Range of Frequencies: .............................................................................................................. 16
  Statistical Significance: ............................................................................................................. 17
  Aliasing: ...................................................................................................................................... 18
  Period Discovery vs. Period Analysis: ....................................................................................... 19
  Environment and Synchronization: ......................................................................................... 20
  Phase Reference Time: ............................................................................................................. 21

SIGNAL INSPECTION AND CONDITIONING OF THE RAW DATA ........................................... 23
  Underlying Assumptions: ......................................................................................................... 23
  Visualization of the Signal (Smoothing and Actogram): .......................................................... 24
  Removing High Frequencies: .................................................................................................... 26
  Binning: ...................................................................................................................................... 27
  Normalization: ........................................................................................................................... 28

ESTIMATING PERIODICITY ........................................................................................................ 29
  Autocorrelation: ....................................................................................................................... 29
  Periodogram: ............................................................................................................................. 33
  Crosscorrelation: ..................................................................................................................... 36
  Cosinor: .................................................................................................................................... 37

ADVANCED APPLICATIONS OF COSINOR ................................................................................ 45

SUMMARY ..................................................................................................................................... 49

FUTURE CONSIDERATIONS AND IMPROVEMENTS ............................................................... 50

SUPPLEMENTAL MATERIALS ..................................................................................................... 50

BIBLIOGRAPHY .......................................................................................................................... 51
Introduction

A remarkable range of behavioral, physiological or molecular workings exhibit rhythmic fluctuation at the behest of a network of biological clocks – with a nexus in the suprachiasmatic nuclei, clocks are found in heart, liver, adipose, all major organ tissues, and may be present in every cell in the body (Cornélissen, 2012; Froy, 2010). Implications in medicine, and biology in general, are far reaching. Altered circadian rhythms, in particular, are associated with disease states and increased disease risk. Treatment efficacy has been shown to be dependent on circadian timing, as have other interventions (Halberg, 1969; Halberg et al., 2009). The impacts of biological rhythms encompass treatment of hypertension (Cornélissen and Halberg, 1994; Hermida et al., 2010, 2011), treatment of cancer (Bernard et al., 2010; Halberg et al., 2003b; Scully et al., 2011), sleep-wake phenotypes and disorders (Dijk and Archer, 2010; Von Schantz, 2008), insulin disregulation and increased risk of metabolic syndrome, obesity, and type 2 diabetes (Cornélissen, 2012; Gupta et al., 2008; Halberg et al., 2003b; Kovac et al., 2009; Shi et al., 2013), energy metabolism disorders (Shostak et al., 2013), and major depression, bipolar disorder, and other mood disorders (McClung, 2007, 2013).

In one study looking at effects of treatment timing, the 2-year disease-free survival rate of perioral cancer patients doubled when treatment was delivered at the time of highest tumor temperature relative to other circadian times (Halberg et al., 2003b).

Moreover, disruption of a circadian element, such as the rest-activity cycle, can in turn disrupt gene transcriptomes, with cascading effects on biological processes.
including chromatin modification, gene-expression regulation, macromolecular metabolism, and inflammatory, immune and stress responses (Delezie and Challet, 2011; Golombek et al., 2013; McClung, 2013; Möller-Levet et al., 2013; Sehgal and Mignot, 2011), highlighting the inter-relatedness of circadian rhythmicity and bodily systems.

Circadian rhythms have long been recognized. The finding that these rhythms were partly endogenous, and not just culturally imposed, led to a growing importance of the field, with the term “circadian” being coined around 1950 by Franz Halberg (Halberg, 1959; Halberg et al., 2003a). Today, the study of biological rhythms is a rapidly expanding field in biomedical research. As of June 2013, a keyword search of PubMed (the US National Library of Medicine’s biomedical database) for “circadian” alone, returned over 68,500 journal entries – a number that has increased by over one third in only 6 years (Refinetti et al., 2007). And yet we find few fields of medicine or biology where application of biological rhythms has found its way into current practice. Despite its long history and recent explosive growth, the field of biological rhythms is yet in its infancy. Biological rhythms underlie potentially all bodily functions and have a great deal to tell us about the interdependence of physiological systems that have been, largely, studied independently.

Advances in automated recording of biostatistics now permit regular, ongoing observations of behavioral, physiological and molecular rhythms over many cycles, with
<table>
<thead>
<tr>
<th>Software</th>
<th>Specifications</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MacAnova</strong></td>
<td>Free, cross-platform R-like platform</td>
<td>G. Oehlert and C. Bingham, University of MN</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.stat.umn.edu/macanova/">http://www.stat.umn.edu/macanova/</a></td>
</tr>
<tr>
<td><strong>Circadian Rhythm Lab</strong></td>
<td>Windows, Free</td>
<td>R. Refinetti</td>
</tr>
<tr>
<td></td>
<td>Actogram, Cosinor, more</td>
<td><a href="http://www.circadian.org/softwar.html">http://www.circadian.org/softwar.html</a></td>
</tr>
<tr>
<td><strong>Time Series Analysis - Cosinor 6.3</strong></td>
<td>Windows, Several types of Cosinor, extensive suite of tools</td>
<td>(Gouthière et al., 2005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expert Soft Technologie Inc BogMedical Computing and Applied Stats Lab</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.euroestech.net/labactivityuk.php">http://www.euroestech.net/labactivityuk.php</a></td>
</tr>
<tr>
<td><strong>Chronobiology Kit Analysis</strong></td>
<td>Free Demo, Windows, Actogram, Periodogram</td>
<td>Stanford Software Systems, Santa Cruz, CA</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://query.com/chronokit/">http://query.com/chronokit/</a></td>
</tr>
<tr>
<td><strong>ClockLab Analysis</strong></td>
<td>Matlab-based</td>
<td>Actimetrics Software, Wilmette, IL</td>
</tr>
<tr>
<td></td>
<td>Actogram, Periodogram</td>
<td><a href="http://www.actimetrics.com/ClockLab/">http://www.actimetrics.com/ClockLab/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td>School of Biology and Ecology and Department of Mathematics and Statistics, University of Maine, Orono, Maine</td>
</tr>
<tr>
<td><strong>Sphygmochron</strong></td>
<td>Windows</td>
<td>Germaine Cornélissen-Guillaume</td>
</tr>
<tr>
<td></td>
<td>Cosinor: specialized blood pressure analysis</td>
<td><a href="http://www.phoenix.ieee.org/018_Sphymochron_Spreadsheet/">http://www.phoenix.ieee.org/018_Sphymochron_Spreadsheet/</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sphygmochron_Spreadsheet.htm</td>
</tr>
<tr>
<td><strong>DiscreteTFDs (Time-Frequency Analysis Software)</strong></td>
<td>Free, Matlab-based fractional Fourier, TFD, others</td>
<td><a href="http://tfd.sourceforge.net">http://tfd.sourceforge.net</a></td>
</tr>
<tr>
<td><strong>R packages: Seasons, Psych, CircStats, Stats</strong></td>
<td>Free, open source, platform independent, individual functions</td>
<td>R Foundation for Statistical Computing</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://cran.r-project.org">http://cran.r-project.org</a></td>
</tr>
<tr>
<td><strong>Past software: Chronomova, Chronolab</strong></td>
<td>Free, Windows, cosinor Free, Mac, several programs</td>
<td>(Andres at al., 1995; Revilla, 1995)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Mojón et al., 1992)</td>
</tr>
<tr>
<td><strong>Chronomics Analysis Software (CAT)</strong></td>
<td>Free, open source, platform independent Full featured application: visualization, autocorrelation, multiple cosinor techniques, progressive analysis</td>
<td>Cathy Lee Gierke, Germaine Cornélissen-Guillaume University of Minnesota</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://564394709114639785.weebly.com">http://564394709114639785.weebly.com</a></td>
</tr>
</tbody>
</table>

Table 1: Rhythm analysis software is rare, and may be specialized, costly, or difficult to understand. Above is a sample only.
the consequence that copious amounts of data can be collected. Examples include:
ambulatory blood pressure monitors (ABPM); lasers that record animal activity with
each break in a laser beam; video analyzers to process and track movements, and even
specific behaviors; transgenic and optical reporters that visualize changes in molecular
activity and rhythms. Other variables continue to be difficult to obtain, such as
hormones, which must be assayed from blood, saliva or urine. Such short or sparse data
present additional challenges when assessing rhythm characteristics and led to the
development of the cosinor by Franz Halberg (Halberg et al., 1967).

Data collection has become less labor intensive, in many ways, yet careful
planning and review of data collection remains crucial to achieving satisfactory
subsequent statistical treatment. Moreover, there is a pressing need for the specialized
statistical and analytical software capable of performing a range of rigorous analyses of
rhythms. Currently available tools are not directly targeted to chronobiologic
applications, and the few available packages may be costly or not easily adaptable (see
Table 1). A tool is needed that provides a suite of analysis options, has flexible
parameterization to adapt the computation to varying circumstances, accepts a range of
data formats, and has documentation of features.

This paper presents the Chronomics Analysis Toolkit (CAT), a flexible, integrated
suite of tools for analyses of periodic data (see Table 2 for a list of techniques included
in CAT), freely available at 564394709114639785.weebly.com. While this paper focuses
on chronobiological applications, the tools in CAT are applicable to a wide range of
rhythmic assessments, beyond biology. Each technique in the toolbox is discussed, focusing on usage and interpretation of the output, with consideration of the data preparation required for the technique. The intent is to provide an introductory survey of concepts behind the techniques used in CAT, with some salient pros and cons, allowing the researcher to judge if the methods used are appropriate for their purposes.

<table>
<thead>
<tr>
<th>Function:</th>
<th>Purpose:</th>
<th>Method:</th>
<th>Notes:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Smoothing</strong></td>
<td>Used to visually inspect data quality, checking for missing data, trends, etc.</td>
<td>A moving average, over a user-selected number of points, smooths the data.</td>
<td>Considered a type of filter; removes noise</td>
</tr>
<tr>
<td><strong>Actogram</strong></td>
<td>Used to visually inspect data for approximate period, by manual selection of a period</td>
<td>Plot of data, one cycle per row, with consecutive cycles stacked one below the other, such that edges will align when the correct period is selected</td>
<td></td>
</tr>
<tr>
<td><strong>Auto-correlation</strong></td>
<td>Rhythm detection and quantitative estimate of period in conjunction with periodogram</td>
<td>Point-by-point comparison of a time series to itself, with a progressive lag, to find correlations that indicate a repeating pattern</td>
<td></td>
</tr>
<tr>
<td><strong>Cross-correlation</strong></td>
<td>Quantitative assessment of phase difference between two time series</td>
<td>Point-by-point comparison of two time series, with a progressive lag, to find correlation that indicate the difference in phase (usually of a similar periodicity)</td>
<td></td>
</tr>
<tr>
<td><strong>Periodogram</strong></td>
<td>Rhythm detection and quantitative estimate of period and amplitude</td>
<td>Discrete Fourier transform</td>
<td></td>
</tr>
<tr>
<td><strong>Cosinor</strong></td>
<td>Rhythm detection (hypothesis testing) and quantitative assessment of MESOR, amplitude, and acrophase of each anticipated period, with corresponding standard errors</td>
<td>A parametric modeling technique that fits a cosine curve to the data, using the method of least squares to minimize the residual sum of squares; a cosinor-based periodogram (i.e., a least squares spectrum) is identical to the Fourier periodogram, when data are equidistant</td>
<td>Only a few of the variations of cosinor-based technique are implemented in CAT</td>
</tr>
</tbody>
</table>

Table 2: Each component of the CAT toolbox provides a different perspective into any existing rhythms, producing an easy to interpret result. All require equidistant data, except Cosinor.

This paper is organized as follows: (1) preparation for sampling, and signal acquisition; (2) inspection of the raw data and signal conditioning -- preparation of the data for further analysis by imputation of missing data, binning data points, and by
removal of high frequency noise, or long-term trends; (3) assessment of rhythmicity and characterization of the rhythm; and (4) discussion of applications. Examples, throughout, frequently refer to analysis of blood pressure obtained by ambulatory blood pressure monitors (ABPM); or to comparisons of activity data between stressed vs. non-stressed mice, where activity is measured by counting breaks in a laser beam.

The next sections of this paper, **Definition of Rhythm** and **Preparation and Signal Acquisition**, introduce concepts and considerations in time series analysis that underlie successful usage of CAT, the Chronomics Analysis Toolkit. Proper usage of any software depends on understanding the application. Discussion of the CAT rhythm analysis techniques begins with the section **Signal Inspection and Conditioning**.

**Definition of Rhythm**

Let us begin by defining the mathematical attributes that characterize a biological rhythm. There are five important characteristics: mean, period, amplitude, phase, and waveform. A perfect sine can be fully characterized by the four attributes diagrammed in Figure 1: mean, period, amplitude, and phase (Refinetti et al., 2007). *Mean level* is the midpoint around which the cycle oscillates. *Period* is the time between two consecutive peaks, or the full length of

![Figure 1](attachment:image.png)

**Figure 1**: A perfect sinusoid can be fully characterized by 4 attributes. (Refinetti et al., 2007)
a cycle. If the period is 2 seconds, the frequency of the rhythm would be described as one cycle in 2 seconds, or ½ Hz (.5 Hz), since one half a cycle occurs in one second. 

Amplitude is half the range of oscillation (the full range of oscillation is sometimes called the double amplitude). Phase is the displacement between a specific point in the cycle, usually the peak, and a reference point. (The term acrophase is used when referring to the peak of the cycle of a cosine fitted model.) The importance of the selected reference time is discussed later. For circadian rhythms, the reference time is often chosen in relation to the sleep-wake cycle of the organism, but it varies depending on what is being studied.

The full characterization of a rhythm requires an additional element, namely its waveform, which describes the shape of the rhythm (Halberg et al., 1977). Biological data often contain a great deal of high frequency oscillation, sometimes characterized as noise, as in Figure 2a. Because noise can make the underlying waveform harder to see, smoothing methods are sometimes used to eliminate these high frequency oscillations. See Figures 2b and 2c. Figure 3 shows examples of biological data with different waveforms. Body temperature data are plotted for four different species on the left (only 2 of 7 days are shown), next to a
periodogram assessing the periods present for each set. The same strong 24-hr period is identified in each species, with only weak indications of other periods, although the waveforms are clearly different for the different species.

One other important feature is the robustness of a rhythm. It refers to the strength of a rhythm and corresponds to the proportion of the overall variance accounted for by the signal (Refinetti et al., 2007). Signals may vary in period, phase or other characteristics over time. Notice how the signal for the dog is more variable over time. When signals vary with time, the data are said to be non-stationary. Biological data, due to their variability, are rarely, if ever fully stationary -- mean, period, amplitude and/or phase often change over time. Most time series analysis techniques, including those in CAT, listed in Table 2, have an assumption of stationarity. In the presence of strong non-stationarities, these methods are no longer applicable. However, CAT can perform progressive analyses, a technique where a span of time is broken into multiple overlapping spans and each is assessed separately, thus allowing identification and
quantification of changing characteristics over time in non-stationary data, provided stationarity can be assumed in each separate span of data.

**Preparation and Signal Acquisition**

Several considerations affect data collection. Each analysis technique has underlying assumptions about the data (stationarity; independence; homogeneity of variance; normality; and in the case of model fitting, goodness of fit) that must be satisfied, thereby impacting the choice of data collection or preparation. Additionally, when the CAT toolbox is used for analysis, coding methods employed in CAT functions may also have a bearing on data collection choices (Chatfield, 2003). The sections below discuss these considerations related to planning and signal acquisition. A full assessment depends on the context and objectives of any specific experiment, and the reader is referred to more complete texts on each concept (Bloomfield, 2000; Chatfield, 2003; Percival and Walden, 1993).

**Discrete or Continuous Sampling:** A variable can either assume a value continuously in time (such as temperature) or only at discrete time points (such as daily rainfall). For the purpose of data analysis, both kinds of variables need to be sampled at discrete time points. The variables measured are, themselves, also categorized as discrete or continuous. Measurements such temperature, have values in a continuous range, to the highest precision measurable by instrumentation. These values often change slowly. By contrast, discrete data are usually counts of recurring events, such as
mouse activity, measured by counting the number of times a laser beam is broken during a chosen sampling interval, and can only assume integer values.

The choice of sampling interval used in analysis can be very important. Choosing a small sampling interval (alternately referred to, herein, as bin size), allows the flexibility to re-bin into larger lots in the future, if that becomes desirable. Surprisingly, the sampling interval, or bin size, has a complex relationship with signal strength: beyond a certain point, decreasing the bin size may cause the signal of interest to disappear into noise (Dowse and Ringo, 1994). (See Binning in the Signal Processing section for further discussion.)

For binning purposes, discrete data must sometimes be processed differently than continuous. Both types of data can use averaging of samples over a selected interval to bin into larger lots. Discrete data can also be summed over a selected interval when larger bin sizes are needed. CAT uses summation of samples across a bin size during interpolation, and binning. Interpolation is not needed in cosinor. Averaging or binning can be performed as an initial step, before running CAT Cosinor, when warranted.

**Equidistant or Non-equidistant Samples:** Although automated recordings are normally equidistant, in reality, recording may be halted from time-to-time, or anomalous readings may occur, leaving gaps in the data. Manual recordings are even more irregular than automated ones. This kind of non-equidistant data is problematic, as many rhythm analysis techniques assume equidistant data. Techniques to analyze non-equidistant data are rare, and include the cosinor (included in CAT) and the Lomb-
Scargle Periodogram, which is a well-regarded method, but subject to artifacts, (Ruf, 1999; Schimmel 2001; Van Dongen et al. 1999), sometimes detecting harmonics that do not exist in the data. All of the CAT functions, with the exception of cosinor, expect equidistant observations. Even small intervals of missing data, common in biological data, cause a data set to be non-equidistant. Interpolation is therefore standardly performed by the CAT functions to fill in these intervals for all functions other than cosinor.

**Missing Data and Interpolation:** As noted above, nearly all data collection has intervals of missing data. CAT uses simple nearest-neighbor linear interpolation to fill in missing data. For example, if a block of missing mouse activity data is bracketed with two activity counts, [420 ... 110], with 40 missing intermediate points representing 30-sec intervals, for a total of 20 min of missing data, data are interpolated linearly: 420, 412.5, 405, 397.5, 390, 382.5, etc., through 110, where each data point increments by 7.5 [(420-120)/40], creating a smoothly changing range of data points.

A limit should be established for the amount of missing data allowable, since linear interpolation is meant to handle small gaps (Horton and Kleinman, 2007). An acceptable limit is needed for both the overall data set and the largest allowed single block of missing data. CAT reports the total percentage of missing data points in a file, and the largest single range of missing points. By setting the parameter $R_{maxGap}$, the program will halt execution and report an error if a single block of missing data is found to exceed $R_{maxGap}$ points in length. $R_{maxGap}$ is recommended to be less than 10% of
the target period to be identified, such that, for the detection of a 24-hr period, 

*RmaxGap* should be less than 2.4 hours. Ideally it will be closer to 4% (1 hour out of 24) of a target period. A data set that covers many cycles will be less sensitive to the occasional gap in data than a shorter set.

Since cosinor is tolerant of missing data, no programmatic restriction of percent of missing data is done in the CAT Cosinor function. Information about minimum, maximum and average sampling interval will be added. For any technique, excessive missing data (or interpolated data) will have an impact on the resulting calculation.

Linear interpolation is adequate for filling in small gaps in the data. In conditions where imputed data must closely match the characteristics of the data set, advanced imputation techniques exist (Enders, 2010; Little and Rubin, 2002). Such pre-processed data can be used as input to CAT.

**Assessing Non-stationary Data:** As mentioned earlier, most biological data are non-stationary – and yet most methods available to analyze them assume stationarity. Methods exist to deal with non-stationary data, such as wavelet analysis (Daubechies 1992; Dowse, 2009; Levine et al., 2002; Liese and

![Figure 4: Progressive analysis of blood pressure for an individual newborn exemplifies the 1) raw data in the top row, 2) MESOR and amplitude, and 3) acrophase over the first 16 days of life. (Cornéllissen et al., 2002)](image-url)
Harrington, 2011), and short-time Fourier transform (Allen, 1977). While these are not addressed in this paper, the progressive analysis method employed by CAT is similar to the short-term Fourier transform in its ability to deal with non-stationary data.

Progressive analysis narrows in on the actual rhythms present over time, and their corresponding characteristics. This method breaks a set of data into subsections, or overlapping subsections, and assesses each subsection successively. By analyzing subsections of a time series, there is less variation over time, allowing an assumption of stationarity locally for each interval of time, even if stationarity cannot be assumed overall. This analysis indicates if the rhythm characteristics remain the same or change over time (non-stationary). If any one of them is changing over time, the result from the analysis of the full data set may not represent the actual periodicities present in the time series. In this case, the subsection analyses yield more accurate results. The trade-off here is that, as a result of fewer samples and fewer repeated cycles, the range of frequencies that can be investigated is reduced, and there is a corresponding loss of resolution and broadened confidence intervals (discussed in the next sections).

In a study of 49 newborns, progressive analysis was key to discerning a circadian blood pressure rhythm. Newborns were monitored from shortly after birth for up to 3 weeks, at 20-min intervals (Cornélissen et al., 2002). Individual blood pressure records were very noisy and visual inspection did not reveal a pattern (Figure 4, row 1). Similarly when all data were pooled, analysis by cosinor over the three-week time span did not detect a circadian rhythm. Data collected during days 1-4 showed a statistically
significant circadian variation, but the subsequent 2 days did not. Cosinor analysis of data collected over days 6-9 again reveals a rhythm. So what is going on? Looking at the data in more detail, it can be seen that during the first days of life, the circadian rhythm is peaking in the early morning; over the intervening days, it is gradually shifting to a peak in the afternoon -- the usual human rhythm. A plot of the cycle peaks (acrophase) for the three-week time span (Figure 4, row 3), constructed from the progressive analysis, affords a view of this shifting change, which occurs at different rates across infants. Only by breaking the time series into shorter spans, can these rhythms be observed as they shift over time, notably when multiple neonatal records are stacked separately for different subsections (not shown).

Each of the CAT analysis functions can be parameterized to perform this type of analysis by subsection. A progressive analysis of a data set of 2048 observations, for example, with an interval (subsection length) of 512 data points specified, and an increment of 256, will begin analysis on the first span of 512 data points, from 1 to 512, then increment the starting analysis point by 256, and perform the next analysis on the next span of 512 data points, from (1+256)=257 to (512+256)=768, etc., progressing through the full dataset, ending at 2048. CAT performs a progressive analysis over subsections of any specified length, with any increment, for each span.

**Non-sinusoidal Signals:** According to Fourier, any signal can be broken down into its constituent sine and cosine components. Thus any non-sinusoidal signal (any signal that is not one single sine or cosine wave) would be expected to be composed of
multiple sine and cosine waves, as shown in Figure 5 (Bloomfield, 2000). At the top is the composite signal, and below are the constituent components, adding up to the composite.

FIGURE 5: The three lower constituent components have 3-, 6- and 12-hr periods, and add up to the top 24-hr composite signal.

Certain signals, such as sawtooths or square waves, have a predictable spectral composition. Hence the periodogram will show more than a single line. For example, a square wave having a period of 12 hours, will show up on a spectral analysis with periods of 12 (the fundamental period), 4 (the 3rd harmonic, 12/3), and 2.4 (the 5th harmonic, 12/5),... and every odd harmonic.

Resolution: Resolution refers to the precision with which a rhythm’s period can be estimated from a set of data. Accordingly, it also holds that, “Resolution in digital signal analysis is the capacity of a given system to separate two arbitrarily close frequencies into distinct peaks in the spectrum“ (Levine et al., 2002). Resolution in frequency depends on the length of the time series (Rhoades, 2008). Intuitively, this can be understood by the fact that higher-frequency components in the periodogram are represented by a larger number of cycles.

Fourier Frequencies: The periodogram consists of – and is limited to -- discrete Fourier frequencies, determined by the record length, resulting in practical limitations.
A data set containing N=2400 data points collected in 6-min bins for 10 days, for example, can be analyzed at periods of 2400/1, 2400/2, 2400/3 bins, etc., such that, in the circadian range, only periods of 2400/12, 2400/11, 2400/10, and 2400/9 bins (or 240/12, 240/11, 240/10, and 240/9 hr) are estimated, corresponding to 20.0, 21.8, 24.0, and 26.7-hr (Dowse and Ringo, 1994; Refinetti, 2004, 2006; Refinetti et al., 2007). This corresponds to a resolving power of 2 to 3-hr for a 24-hr period. Of note, however, is that methods exist to calculate intermediate periods between actual Fourier frequencies, thereby allowing the period to be determined more precisely (Rabiner, 1969; Welch, 1967), but the uncertainty of the period estimate remains the same, depending on the record length. Those fractional Fourier techniques are not used in CAT, although the CAT cosinor has the capability to calculate intermediate Fourier frequencies.

**Range of Frequencies:** The range of frequencies (or periods) that it is possible to investigate depends on the length of the time series and the sampling rate (or bin size). The longest period that can be assessed is the length of the data set. The shortest period that can be assessed is determined by the sampling rate.

Series length is the first relevant factor for determining the range of periods that can be estimated, and defines the longest period that can be estimated from a series. Where a period of a certain length is known to exist, meaningful results can be obtained on the basis of data covering a single cycle, but whenever possible it is desirable to record multiple cycles. In 1993, Refinetti studied accuracy and noise tolerance of six
rhythm analysis techniques, using simulated running wheel activity with a known period. He reported common practice in studies of circadian rhythms was to use 10 cycles. A time series of 10 cycles minimizes the effect of potential non-stationarities that may be encountered in longer data streams. Less than 10 days may allow short-term random variations to have undue influence on the determination of period and rhythm characteristics (Refinetti, 1993).

Common practice depends on study goals, however. In the case of blood pressure diagnosis and treatment, 7 days of automated readings is often recommended at the outset to obtain a more reliable estimation of the circadian rhythm as well as a rough estimate of the weekly variation (Halberg et al., 2010b).

The second of these two factors, sampling rate, determines the shortest period that can be detected. Accordingly, as sampling rate increases, rhythms with shorter periods can be detected. For instance, if the sampling rate for a mouse activity experiment is every half hour, then it would be possible to evaluate periodicities longer than an hour (but not equal to or shorter than one hour). Put in frequency terms, this means the sampling rate must be more than twice that of the highest frequency to be analyzed; this limit is the Nyquist frequency (Levine et al., 2002; Pierce, 1961). Note this is a theoretical minimum, and it is usually advocated to sample at a frequency at least three times that of the highest frequency component of interest to the investigator.

**Statistical Significance:** Because the number of samples \((N)\) in a series determines the number of degrees of freedom, the \(P\)-value and estimation of
Confidence intervals of rhythm parameters depend on the number of samples, $N$. If more than one test is carried out on a given time series (testing for the presence of more than a single rhythm), the probability level used for hypothesis testing and computation of a 95% confidence intervals needs to be adjusted for the number of tests performed using special methods, for instance using a Bonferroni correction (Repinetti et al., 2007).

**Aliasing:** In signal processing, aliasing refers to an effect that causes different signals to become indistinguishable. The effect may be seen in an old movie where a wheel on a forward moving wagon appears to be moving backward. The movie frames move at a speed too slow to see the wheel’s forward motion, but at a speed that captures a sequence of wheel positions appearing to be moving backward! Stated another way, given a theoretical signal containing two sinusoids that cross at multiple points, a dataset consisting of only those crossing points will be insufficient to allow correct determination of the sinusoid actually present – instead, a sinusoid with a longer period will be detected (Figure 6). Aliasing is caused by a sampling frequency that is too low to detect a high frequency signal, mis-identifying it as a lower-frequency signal.
Specifically, it occurs when the sampling interval is at or longer than one half the period being recorded (Chatfield, 2003; Dowse, 2007; Hamming, 1983).

**Period Discovery vs. Period Analysis:** Analysis of rhythm is not only done to discover a rhythm *de novo*; it may also be done to study the details of a rhythm that is already partially characterized. When the period is known or can be anticipated, as for the circadian rhythm in blood pressure, the observation of one cycle may be sufficient to detect the rhythm with statistical significance. Whereas, when the period is not known, data need to be collected over more than one cycle. Even though three points fully characterize a rhythm with assumed period over one cycle, to be able to test for statistical significance and to provide confidence intervals for the rhythm parameters, additional data are needed, for either known or unknown periods. In the latter case, data must cover more than one cycle (Refinetti et al., 2007).

The cosinor method of analysis was designed for studying circadian rhythms, which are ubiquitous and often prominent, even when sampling is sparse. The CAT Cosinor function is the most highly quantitative of the CAT functions, reporting an array of metrics, including the rhythm-adjusted mean (MESOR), amplitude and phase for each period assessed, and the standard errors for each metric. The specifics of cosinor-based methods will be elaborated below, but cosinor has important distinction from the other CAT techniques: it does not require data to be equidistant. This can be critical since there are many experiments in which data collection at equal intervals is not possible. DNA microarray gene expression experiments, for example, commonly acquire time
series with different sampling intervals (Liew et al., 2007).

The cosinor may also be used for discovery by assessing a spectrum of Fourier frequencies, as is done with a periodogram. The strongest period(s) identified can then be used to build a cosinor model, of interest in various applications, such as disease prediction. In this case, the more closely the raw data approaches a cosine curve, or multiple component cosine curve, the better this model performs. All of the other functions in CAT (Table 2) assume the data are equidistant, and analyze the raw data directly using non-parametric techniques, rather than a regression-based model.

**Environment and Synchronization:** In living organisms, many biological functions have at least partly endogenous circadian rhythms coordinated with one or several external environmental factors (Halberg et al., 1953, 1977). These synchronizers (also called entraining agents) include light/dark cycles, cycles of temperatures, feeding times, social interaction or stressor events that occur with regularity, etc. In the absence of environmental synchronizers, circadian rhythms persist, albeit with a period slightly, but statistically significantly, different from 24-hr, as shown in isolation experiments (Siffre et al., 1966; Wever, 1979). When the endogenous circadian period is close to rhythms in the environment, it usually becomes synchronized and assumes the same period as the environment. This has been shown in several experiments where the period length was changed from 24-hr to either a slightly shorter or longer period. Environmental synchronizers also influence the phase. Careful attention is
therefore needed to control or account for potential impacts of entraining agents and other environmental factors (Asher et al., 2010).

Heart rate, for instance, is a rhythmic system that is impacted by both endogenous and environmental factors. There is a distinct circadian component, usually characterized by lower values during sleep or rest, a small increase around mid-sleep followed by a larger increase after waking, a postprandial dip that is more pronounced with age, and a slow decline in the evening (see Figures 22, 23 & 24 for sample plots of weekly, and stacked daily blood pressure) (Cornélissen et al., 2002; Otsuka and Cornélissen, 2012). Blood pressure follows a very similar pattern. Both blood pressure and heart rate vary with a variety of factors including emotions (Halberg et al., 2010a), mental health (McClung, 2013; Ehlers, 1988), exercise (Levine et al., 1977; Rowland, 2011; Singh et al., 2012), and smoking (Scarpelli et al., 1989).

A pair of studies on newborn heart rate demonstrates the potential impact of environmental factors – feeding time, in this case. In one hospital where babies were fed every 3 hours, except for one time point at night, automatic 24-hr monitoring of 164 babies during the first week of life revealed an about 3-hr cycle in heart rate rhythm that was stronger than the circadian rhythm. At a different hospital, where babies where fed on a 4-hr schedule, a heart rate rhythm of 4-hr, not 3, was found (Cornélissen et al., 2002; Otsuka and Cornélissen, 2012; Schuh et al., 1989).

**Phase Reference Time:** *Phase* is the displacement between a specific point, usually the peak of the cycle, in relation to a reference point. Changing the reference
time thus leads to a change in the phase estimation. As such, it is important to use a consistent and meaningful reference point when comparing results between different analyses.

The importance of the reference point can be seen in a study of 62 newborns monitored at 30-min intervals for 48 hours during the first week of life, where blood pressure was analyzed for circaseptan (about-weekly) rhythms. When analyzed as a function of the day of the week (i.e., with a reference date of Sunday), using ANOVA, single cosinor and population-mean cosinor, only slight support is found for a weekly rhythm \( (P = .255, .030, .656 \text{ respectively}) \). When analyzed, however, as a function of days since birth, using these same techniques, there is significant support for a weekly rhythm \( (P = .011, .001, .001 \text{ respectively}) \), suggesting that developmental age, rather than societal schedule, was important in the synchronization of the circaseptan rhythm of blood pressure in the first days after birth (Cornélissen, 2012).

The reference phase chosen will depend on the purpose of the study. If the purpose is to evaluate the impact of an environmental factor or stressor on a rhythm, the reference point would be chosen relative to the environmental factor. Or it may be chosen according to the period expected to be impacted -- if a circadian rhythm is being studied, a specific point of day, perhaps dark onset, would be chosen; if a circaseptan (weekly) period is suspected, a point in the week would be chosen. A record of the timing of potentially entraining factors or events that occur over the span of data collection can aid in later analysis of possible impacts to endogenous rhythms.
Signal Inspection and Conditioning of the Raw Data

This section discusses data concerns that can impact accurate, successful analysis of rhythm and, briefly, conditioning methods sometimes used to address them. The CAT package, and rhythm analysis techniques in general, rely on certain assumptions about the data being analyzed. CAT provides visualization tools to support investigation of how well data meet these underlying assumptions, reviewed below.

**Underlying Assumptions:** Major assumptions are stationarity, independence of data, homogeneity of variance and normally distributed noise. For data modeling by cosinor, goodness of fit is also an underlying assumption. While it is not always possible to satisfy all underlying assumptions, some data conditioning can bring data closer into alignment with the requirements of the analytical methods (De Prins and Cornélissen,
Pre-processing, or conditioning, techniques include detrending, data transformations (log, square root...) and binning or averaging. CAT does not perform pre-processing except for binning. It is noted here that certain types of data, such as mouse activity or some circulating hormones such as melatonin are not normally distributed and have heterogeneous variance (Lewy and Sack, 1997). Also noted, a slowly varying variable such as body temperature, if sampled too often, violates the assumption of independence.

**Visualization of the Signal:** Although quantitative analysis is required to characterize rhythms, visual inspection is important to assess the quality of the data, and consider if modifications are needed. Missing data, trend lines, or unexpected behaviors might be observed, prompting further inspection or data conditioning.

Inspecting Figure 7, for example, a long sloped line spanning two days can be seen where there was a large gap of missing data, filled by linear interpolation.

For visualization purposes, the CAT Smoothing function plots a moving average, which cleans up the signal, allowing the rhythm to be

![Actogram: Each row is a day: bars are activity levels during each interval.](image)

*Figure 8: CAT Actogram: A distinct vertical alignment gives an approximate estimate of the period length: \( \text{modulus}=1440 \text{ hours.} \)*
observed more clearly. CAT
smooths the data by
averaging the K nearest data
points on either side of every
data point (such that 2K+1
data points are averaged),
and that average becomes
the new point. The default is
K=6 for the CAT smoothing
function, but K is a
parameter specified at run
time. More examples of CAT
output can be seen in the
Supplemental Materials.

In addition, CAT
includes an Actogram
function, a standard method
used to visualize periodicity
at a macroscopic level (Figure
8). An actogram comprises multiple rows, each row displaying data over a duration
specified by the user, corresponding to the anticipated period. Each successive row

Figure 9: CAT Actogram: A diagonal line indicates a period longer than the period selected for display. Parameter mod=1340 min.

Figure 10: CAT Actogram: The same data as Figs 8 & 9, but the width obscures the period. Parameter mod=1000 min (16.67 hr).
displays the data over successive cycles. The alignment of data at a selected cycle length gives a rough assessment of the period at a glance. A distinct vertical edge, as seen in Figure 8, shows activity beginning at the same time in each cycle, indicating the selected width of the Actogram (or folding period) is roughly equivalent to the period in the data. A diagonal from upper left to the lower right, as in Figure 9, means the period in the data is longer than the period specified for the actogram (onset activity is later in each cycle). Even when there is no apparent rhythm, as in Figure 10, the data may still be periodic. Figures 8, 9 and 10 show actograms using the same data set, with different modulus parameters. The Actogram function in CAT can be configured to assess for any period length by changing this modulus parameter.

Removing High Frequencies: The jagged, highly variable data points common to biological data (see Figure 2) can hinder analysis if the techniques used are too sensitive to such variability (Levine et al., 2002). These high frequency oscillations can be due to artifacts of data acquisition (quality of instruments, influences from external sources...), or they may be inherent in the data. Mouse activity data, for example, are highly variable from moment to moment, and that high variability is characterized by high frequency oscillations. Such high frequency oscillations, regardless of the source, are often referred to as noise, and are therefore sometimes removed from the data stream using filters, such as the moving average used for smoothing in Figure 7. Noisy signals are more likely to register the presence of weaker periodicities that do not actually exist in the data (Levine et al., 2002). Before applying a filter, however, one may want to
consider whether this variability is true noise, or whether it may be an as-yet-not-understood rhythm. Another consideration with filters is that certain filters, such as a Butterworth filter, can introduce a phase shift in the data (Dowse, 2007; Levine et al., 2002); or alternately, may cause the assumption of independence of the data to be violated, as a moving average will do (Bloomfield, 2000).

The methods used in CAT, because they are tolerant of white noise, do not usually require filtering. The moving average employed by the CAT Smoothing function described above is one method sometimes used to reduce high frequency oscillations. CAT does not use the resulting smoothed data for further processing, but the display gives an idea how the data would be changed by smoothing or filtering, and is an important visual aid in assessing data.

**Binning:** As discussed earlier, both discrete and continuous data can be averaged over a selected interval, or bin. An additional option for discrete data is to sum them over a selected interval instead of averaging, and this is the method currently used in CAT. When deciding on the binning size, the specific application and constraints are important, as well as considerations of desired resolution, range of frequencies to be detected, statistical power, possible artifacts due to binning, among other things, as discussed above. Note that binning can serve as a type of filter, as it can reduce high frequency oscillation (Chatfield, 2003). Sample size and desired bin size are specified on the call to CAT. After missing data are interpolated, the data are binned.

Published studies in the circadian field have used widely-varying sampling
intervals, from 6 minutes, to 1 or 2 hr (Palmer et al., 1994; Refinetti, 1993, 2004), depending on characteristics of the subject under study. Dowse recommends a minimum of 30-min intervals for locomotor activity data having periods in the range of 6 to 40 hr, based on empirical tests of bin sizes ranging from 5 min to 4 hr, in 1-min increments (Dowse, 2007, 2009; Dowse and Ringo, 1994). His testing modeled variations in the period of oscillation, amount of noise and bin size. Testing employed autocorrelation, Maximum Entropy Spectral Analysis (MESA) and transfer functions at each bin size to determine the effects of varying bin size on the results. He reports that bin sizes much smaller than 10 minutes may cause artifact, or the period may be obscured in noise from so many samples. The resulting general “rule of thumb” is that bin size be selected in the range of 1% to 10% of the period of greatest interest (Dowse and Ringo, 1994). Because individual experimental conditions vary widely, however, any general recommendations must be weighed in the context of the study’s aims, the effort and cost of data collection, and the amount of noise affecting measurements.

Normalization: Normalization of the data allows for direct comparison of results between different data sets, and ultimately facilitates the comparison among behavioral, molecular, genetic and other signals, contributing to understanding the complex relationships between these systems (Levine et al., 2002). In CAT, the AutoCorrelation, Periodogram, and CrossCorrelation calculations automatically normalize amplitude. For cosinor, normalization on the Y-axis can be achieved by expressing data as a percentage of each series’ arithmetic mean value prior to analysis.
Estimating Periodicity

CAT uses several techniques to detect rhythms and get an approximate estimate of their period: Autocorrelation; Periodogram; Cosinor. Each method has its own strengths and weaknesses and, used together, allow conditional understanding of the rhythm. Visualizations give an intuitive sense of the rhythm present; significance levels are shown. All methods assume equidistant samples, except cosinor.

Autocorrelation is a fundamental statistical method for identifying periodicities. It has been tested empirically on varying types of biological data sets, and found to be reliable and accurate in the circumstances tested (Dowse, 2009, 2007; Dowse and Ringo, 1994; Levine et al., 2002; Palmer et al., 1994; Refinetti, 1993, 2004, 2006; Refinetti et al., 2007). Autocorrelation begins by comparing a data set to itself, point by point, from start to end, using a standard correlation analysis. Since each data point is compared against itself, the correlation is perfect, and the resulting correlation coefficient, \( r \), is 1. The two identical sets are then shifted, or lagged, by one point, and the comparison is repeated. The computed \( r \) will not be as perfect for this position. This shifting, or lagging, is continued, one point at a time, until it has moved about a third of the way through the data, (N/3) in CAT. The resulting \( r \) values are plotted in sequence as a function of the lag, in what is called a correlogram. If the series is rhythmic, \( r \) decreases and increases regularly with a period equivalent to that present in the data series. The output is normalized at each step by dividing by the data series’ variance, yielding a correlation coefficient, \( r \), that assumes values between -1 and 1.
Each time the data set is lagged, the values on the two ends have no pair for the correlation calculation and are discarded; hence the power of the test is gradually diminished. For this reason, the usual limit of the autocorrelation computation is about N/3. The 95% confidence interval and hence significance of a peak is given as $2/\sqrt{N}$, where N is the number of data points (Chatfield, 2003; Dowse, 2009).

Interpretation: Chatfield, a recognized expert in time-series analysis, calls the correlogram "probably the most useful tool in time-series analysis after the time plot." He also says that interpreting a correlogram "is one of the hardest tasks in time-series analysis" (Chatfield, 2003). For this reason, a period estimate is not given on the autocorrelation. In the presence of a strong single periodicity, the lag of the first peak gives an approximation of the period, as does the difference between successive peaks. A very rough rule of thumb for interpreting a correlogram is that repeated peaks

Figure 11: CAT AutoCorrelation: One third of the 3024 bins (20 days) are used. All peaks exceed the dashed 95% confidence band. Interval length ($\Delta t$): 2min; 5 samples per bin; $2^*5 = 10$-min bins. Amplitude = .5 The first peak at 23.83 hrs gives an approximation of the period. An accompanying periodogram (Figure 15) of the plotted $r$ values determines the actual estimate of the period.
exceeding the confidence level detect periodicity (see Figure 11-13). It has been suggested, in studies of drosophila, to use the third peak to judge statistical significance— the rhythm is statistically significant if the third peak is above the dotted significance line on the correlogram (Dowse, 2007, 2009; Levine et al., 2002). A great deal of variation between cycles in the data, or a decrease in amplitude, will cause the correlogram to decay more rapidly than in a regular series.

Results from the CAT Autocorrelation functions can be seen in Figures 11 through 13. In Figure 11, a time series of 20 days of activity data from stressed mouse #4: X8, very similar to Figures 8 & 9, was correlated with itself. The repeated peaks, well exceeding the dashed confidence lines, with very little attenuation, indicate a strong rhythm. An accompanying periodogram, performed on the autocorrelation results (Figure 15 & 16), provides the actual estimate of the period. The maximum lag used for

![Figure 12: CAT AutoCorrelation: Two days of data. Less than one cycle can be seen on the plot, resulting in an inaccurate identification of the peak. Multiple cycles, preferably 5 or more, should be analyzed.](image)

Animal - 4 :  X8
2007-10-18 02:00:20 -- 2007-10-20 01:58:20

Max Lag= 95.04 ; 1st Peak: 93 bins ( 15.5 hrs)
Total bins: 288 ; pts/bin: 5 ( 10 mins)
Span: 0 Inc: 0  activity stress c57--basIn.txt May 25, 2013 - 18:06:51
Figure 13: CAT AutoCorrelation: Top plot has 1 sample/bin, for a 2 min bin size. Bottom plot uses 6 samples/bin, for a 12 min bin. Peak location is given in intervals of bin size. Approximation of the period (peak) can get closer to the actual period with smaller bin size, but period uncertainty remains the same, depending solely on record length.

The dashed line is significance level, and the solid line is for reference in the figures herein (not in CAT), estimating the (normalized) height and attenuation of subsequent peaks.

Figure 12 is data from the same animal (#4: X8), but contains only 2 days, and demonstrates the problems resulting from short time-series. The Autocorrelation, performed on N/3 of the data bins results in only 48/3 (16) hr of data being used. CAT reports an error and does not accept an input file containing less than 3 days of data.

The plots in Figure 13 use the same data as in Figure 11, but all three plots have
different bin sizes, which result in different period estimates. Using autocorrelation, point estimates of the period can get closer to the actual period with smaller bin size, but period uncertainty remains the same, depending solely on record length. Other factors such as noise and waveform can also impact this relationship. Note the power and confidence intervals also vary between the correlograms.

For comparison, Figure 14 shows a plot of random data. Chatfield advises familiarity with correlogams, using model data from known sources as well as real data, as the best way to learn to interpret them. For detailed specifications on the autocorrelation function used in CAT see (Venebles and Ripley, 2002).

**Periodogram:** The periodogram is another long-time standard method for finding periodicity. Based on Fourier’s insight that a signal can be reproduced by a series of sines and cosines, Fourier analysis is used in nearly every field of engineering and physical sciences, from MRI and image processing to X-ray crystallography and geophysics. In the periodogram, Fourier analysis is used to calculate the amplitude (strength) at specific periods: N/1, N/2, N/3, up to N/(N/2)=2, where N is the number of
equidistant samples in the time
series (Rhoades, 2009). Each
sequence begins with N, and
continues to 2 (the Nyquist limit).
Mathematically speaking, an
assessment ends just before 2, not
at 2, as there must be more than 2
points per cycle. The periodogram
only includes these specific
periods, so N should be chosen
with this in mind. See Table 3 for
sample periods using differing
series lengths. Unlike other CAT

techniques that allow finer control
of the periods, the periods (or
frequencies) included in the
periodogram depend directly on
the number of data points
analyzed.

Interpretation:

Periodograms in Figures 15 and 16

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N/1</td>
<td>48.00</td>
<td>56.00</td>
<td>108.00</td>
<td>2000.00</td>
</tr>
<tr>
<td>N/2</td>
<td>24.00</td>
<td>28.00</td>
<td>54.00</td>
<td>1000.00</td>
</tr>
<tr>
<td>N/3</td>
<td>16.00</td>
<td>18.67</td>
<td>36.00</td>
<td>666.67</td>
</tr>
<tr>
<td>N/4</td>
<td>12.00</td>
<td>14.00</td>
<td>27.00</td>
<td>500.00</td>
</tr>
<tr>
<td>N/5</td>
<td>9.60</td>
<td>11.20</td>
<td>21.60</td>
<td>400.00</td>
</tr>
<tr>
<td>N/6</td>
<td>8.00</td>
<td>9.33</td>
<td>18.00</td>
<td>333.33</td>
</tr>
<tr>
<td>N/7</td>
<td>6.86</td>
<td>8.00</td>
<td>15.43</td>
<td>285.71</td>
</tr>
<tr>
<td>N/8</td>
<td>6.00</td>
<td>7.00</td>
<td>13.50</td>
<td>250.00</td>
</tr>
<tr>
<td>N/9</td>
<td>5.33</td>
<td>6.22</td>
<td>12.00</td>
<td>222.22</td>
</tr>
<tr>
<td>N/10</td>
<td>4.80</td>
<td>5.60</td>
<td>10.80</td>
<td>200.00</td>
</tr>
<tr>
<td>N/11</td>
<td>4.36</td>
<td>5.09</td>
<td>9.82</td>
<td>181.82</td>
</tr>
<tr>
<td>N/12</td>
<td>4.00</td>
<td>4.67</td>
<td>9.00</td>
<td>166.67</td>
</tr>
<tr>
<td>N/13</td>
<td>3.69</td>
<td>4.31</td>
<td>8.31</td>
<td>153.85</td>
</tr>
<tr>
<td>N/14</td>
<td>3.43</td>
<td>4.00</td>
<td>7.71</td>
<td>142.86</td>
</tr>
<tr>
<td>N/15</td>
<td>3.20</td>
<td>3.73</td>
<td>7.20</td>
<td>133.33</td>
</tr>
<tr>
<td>N/16</td>
<td>3.00</td>
<td>3.50</td>
<td>6.75</td>
<td>125.00</td>
</tr>
<tr>
<td>N/17</td>
<td>2.82</td>
<td>3.29</td>
<td>6.35</td>
<td>117.65</td>
</tr>
<tr>
<td>N/18</td>
<td>2.67</td>
<td>3.11</td>
<td>6.00</td>
<td>111.11</td>
</tr>
<tr>
<td>N/19</td>
<td>2.53</td>
<td>2.95</td>
<td>5.68</td>
<td>105.26</td>
</tr>
<tr>
<td>N/20</td>
<td>2.40</td>
<td>2.80</td>
<td>5.40</td>
<td>100.00</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>N/25</td>
<td>2.24</td>
<td>4.32</td>
<td>80.00</td>
<td></td>
</tr>
<tr>
<td>N/26</td>
<td>2.15</td>
<td>4.15</td>
<td>76.92</td>
<td></td>
</tr>
<tr>
<td>N/27</td>
<td>2.07</td>
<td>4.00</td>
<td>74.07</td>
<td></td>
</tr>
<tr>
<td>N/28</td>
<td>2.00</td>
<td>3.86</td>
<td>71.43</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>N/54</td>
<td></td>
<td></td>
<td></td>
<td>37.04</td>
</tr>
<tr>
<td>N/55</td>
<td></td>
<td></td>
<td></td>
<td>36.36</td>
</tr>
<tr>
<td>N/56</td>
<td></td>
<td></td>
<td></td>
<td>35.71</td>
</tr>
<tr>
<td>N/57</td>
<td></td>
<td></td>
<td></td>
<td>35.09</td>
</tr>
<tr>
<td>N/58</td>
<td></td>
<td></td>
<td></td>
<td>34.48</td>
</tr>
<tr>
<td>N/59</td>
<td></td>
<td></td>
<td></td>
<td>33.90</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/80</td>
<td></td>
<td></td>
<td></td>
<td>25.00</td>
</tr>
<tr>
<td>N/81</td>
<td></td>
<td></td>
<td></td>
<td>24.69</td>
</tr>
<tr>
<td>N/82</td>
<td></td>
<td></td>
<td></td>
<td>24.39</td>
</tr>
<tr>
<td>N/83</td>
<td></td>
<td></td>
<td></td>
<td>24.10</td>
</tr>
<tr>
<td>N/84</td>
<td></td>
<td></td>
<td></td>
<td>23.81</td>
</tr>
<tr>
<td>N/85</td>
<td></td>
<td></td>
<td></td>
<td>23.53</td>
</tr>
</tbody>
</table>
Figure 15: CAT Periodogram: Bins/cycle, using the same experimental data set as Figures 11 & 13; 24-hr is clearly the strongest signal. Period in hours listed by amplitude.

use the same data as the correlograms in Figures 11 and 12. Vertical lines, called spectral lines, convey the amplitude of the sine (or cosine) curves at each frequency; they indicate the prominence of signals present in the data. Figure 15 is very clean with 1 strong spectral peak indicated at 24-hr, and 2 weaker ones at 12 and 8 hr. By contrast, Figure 16, based on only 2 days of data, is less definitive, having the same three most prominent spectral peaks, but with less differentiation between the periods identified.

In general, a non-sinusoidal waveform (a rhythm not composed of one pure sine or cosine) will be represented in a periodogram by a spectral line at the fundamental frequency, with additional (usually smaller) spectral peaks at the additional harmonics that define the waveform. Very rarely a signal can be represented by harmonic terms in

Figure 16: CAT Periodogram: Using the same data as Figure 12, two days of data, results in less clarity than Figure 15. 24-, 12- & 8-hr lines are largest. Period in hours is listed by amplitude.
the absence of a spectral peak at the fundamental period.

**CrossCorrelation**: Crosscorrelation can be used to identify the phase difference between two time series sharing a common frequency. The calculation is the same as autocorrelation, except that two data sets are used instead of one. The two datasets are compared point-for-point to calculate $r$, then lagged and $r$ is calculated again. The resulting cross-correlogram will show a peak at lag zero if the data are in phase. Displacement of the first peak from lag zero in either direction is a direct measure of the phase difference between the two rhythms (Dowse, 2007). A reference data set may be compared to the data set of interest, or two data sets may be compared directly to find the difference in phase between them.

![Figure 17: CAT Crosscorrelation](image)

A short 6-day control group with scant periodicity is compared to the robust 20-day experimental data set. Depicts a phase difference of 7.5 hr ($45/155)*24$, where the control average lags the experimental average – control peaks later than experimental.
For the purposes of a study, a reference point may be different things. In the case of a study of the relationship between an entraining agent (light/dark), and the rhythm it entrains (e.g., blood pressure, activity), an appropriate reference time may be a stage of the entrainer, such as dark onset (assuming a synchronized relationship has been established). To investigate potential entraining agents for a rhythm, it may be necessary to record multiple potential reference points, and compare each to a reference point, to see which has a consistent phase relationship with the rhythm under study, possibly using a phase response curve (PRC), which plots changes of phase as a function of when a stimulus (e.g., light) occurs (Levine et al., 2002).

**Interpretation:** Two mouse activity data sets with similar periods are crosscorrelated using CAT in Figure 17. A 6-day data set with weak approximate 24-hr periodicity is compared against a 20-day data set with a similar, but robust rhythm, and in this case, the average of 5 animals from each data set was compared. The peak in crosscorrelation occurs at the lag corresponding to 45 bins (144 bins in a day), revealing a 7.5-hr phase difference between the two data sets, where the reference (first) datastream lags (peaks after) the experimental (second) datastream. Note the largest $r$ is below 0.4 (on a scale from 0 to 1), and the peaks attenuate, both of which are indicators of the degree of robustness of correlation. The periods present in each data set should be similar for a clean analysis of phase using crosscorrelation.

**Cosinor:** The single and population-mean cosinor techniques were first developed and extensively applied to analysis of biological rhythms by Franz Halberg, at
<table>
<thead>
<tr>
<th>Cosinor extensions:</th>
<th>Multiple techniques provide different information</th>
<th>In CAT:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Cosinor, Single-component model</td>
<td>One or more cosine curves are fitted to the data, separately. This is the base cosine method.</td>
<td>Yes</td>
</tr>
<tr>
<td>Individual Cosinor, Multiple-component model</td>
<td>Multiple cosine curves are fitted to the data as one model.</td>
<td></td>
</tr>
<tr>
<td>Population Mean Cosinor</td>
<td>Results from single- or multiple-component individual cosinor are summarized for a population using vectorial averaging.</td>
<td></td>
</tr>
<tr>
<td>Least Squares Spectrum</td>
<td>This is a configuration of the single-component individual cosinor applied to all Fourier frequencies, yielding a spectrum identical to a periodogram, for the special case of equidistant data. (Can also be applied to frequencies intermediate to the Fourier frequencies, which the periodogram cannot calculate.)</td>
<td>Yes</td>
</tr>
<tr>
<td>Progressive Analysis, Single-component model</td>
<td>Single-component individual model cosinor is applied to subsections of the data record and moved progressively along at specified increments. Part or all of the data record can be assessed.</td>
<td>Yes</td>
</tr>
<tr>
<td>Progressive Analysis, Multiple-component model</td>
<td>Multiple-component individual model cosinor is applied to subsections of the data record and moved progressively along at specified increments. Part or all of the data record can be assessed.</td>
<td></td>
</tr>
<tr>
<td>Progressive Analysis, Least Squares Spectrum</td>
<td>This is a progressive analysis performed at the frequencies of a periodogram, using single-component cosinor model. Yields a heatmap for visualizing changes through time. Frequencies intermediate to Fourier frequencies can be included. Also not limited to 1/T and 1/Δt (part or all of the data record can be assessed).</td>
<td>Yes</td>
</tr>
<tr>
<td>Nonlinear Cosinor</td>
<td>Can estimate the period, together with MESOR, amplitude and phase of one or more components. Includes capabilities for all of the above-described extensions, as well.</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Cosinor can be performed in a variety of ways, depending on goals of the assessment.

The University of Minnesota, to handle short time-series and sparse data when prior information is available (Halberg et al., 1967). Its ability to handle non-equidistant and missing data is a powerful feature. Cosinor is a regression technique that fits one or more cosine curves to the data, separately or concomitantly, minimizing the sum of squares of the differences between the actual measurements and the fitted model (the residuals), for the specified period (Halberg, 1980). From this model, one obtains, for the period considered, an estimate of (i) the rhythm-adjusted mean or midline estimating statistic of rhythm (MESOR), defined as the average value of the curve fitted
Figure 18: CAT: Var 1=Analysis of 56 equidistant blood pressure readings at 3-hr intervals, was done over the course of a week. a) Periodogram: Summary of the periodicities corresponding to the four largest amplitudes b) Cosinor yields equivalent periods and amplitudes to periodogram when Fourier frequencies are assessed.

to the data, (ii) amplitude (A), defined as half the height of oscillation in a cycle

approximated by the fitted cosine curve (difference between the maximum and the
MESOR), and (iii) acrophase (\(\phi\), a measure of phase), the lag from a defined reference
time point (e.g. local midnight, or other significant point) to the crest time in the fitted
curve (see Figure 1) (Halberg et al., 1977). Statistical significance is determined for the
period tested by an F-test with respect to the null hypothesis (zero amplitude or no-
rhythm). Cosinor also reports an estimate of the percentage rhythm, or proportion of
variance accounted for by the model (Cornéllissen and Halberg, 2005).

The single cosinor is the core of a set of cosinor-based methods, that calculates
the best fit of a cosine model made up of either one, or multiple cosine components, at
specified periods. A population-mean cosinor (to be added to CAT) can summarize
single cosinor results (either single- or multiple-component models) from a set of
individuals at a common time period by vectorial averaging of individual results from the
single cosinor, estimating the extent of similarity among individuals (Cornélissen and Halberg, 2005). Additional extensions of the cosinor are listed in Table 4. Those performed by CAT are described further in the coming section.

The single cosinor, when used at the Fourier frequencies, yields results identical to a periodogram (Bloomfield, 2000). Figure 18 shows a periodogram calculated with CAT, from equidistant blood pressure readings, with a list of the strongest four periodicities (largest amplitudes). Frequencies corresponding to those on the periodogram are also calculated with cosinor, yielding the same four strongest periodicities, with identical amplitudes, listed below the periodogram. A complete listing of the amplitudes calculated for each period is given in Figure 19. They are identical to 6 decimal places. Bloomfield (2000) proved the equivalence between the discrete Fourier Transform (periodogram) and the regression approach (cosinor).

**Periodogram: Fourier Periods (in hrs, top 3 rows) and Amplitudes (last 4 rows; first value is arithmetic mean)**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>11</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000000</td>
<td>18.666667</td>
<td>14.000000</td>
<td>9.333333</td>
</tr>
<tr>
<td>2.68826e-14</td>
<td>2.557129e-00</td>
<td>5.090192e-00</td>
<td>8.40636e-01</td>
</tr>
<tr>
<td>1.089843e+00</td>
<td>3.11665e+00</td>
<td>4.459975e+00</td>
<td>2.468414e+00</td>
</tr>
<tr>
<td>2.737095e-14</td>
<td>2.204135e-00</td>
<td>1.638726e-00</td>
<td>1.40714e-00</td>
</tr>
<tr>
<td>1.089843e+00</td>
<td>3.11665e+00</td>
<td>4.459975e+00</td>
<td>2.468414e+00</td>
</tr>
<tr>
<td>0.843315e-01</td>
<td>1.90636e-01</td>
<td>1.842991e-00</td>
<td>3.011311e-00</td>
</tr>
</tbody>
</table>

**Cosinor: Amplitudes corresponding to periodogram periods, above**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.557129</td>
<td>5.090192</td>
<td>0.884063</td>
<td>4.358985</td>
</tr>
<tr>
<td>1.638726</td>
<td>1.40714</td>
<td>1.40714</td>
<td>2.923299</td>
</tr>
<tr>
<td>5.266700</td>
<td>8.266700</td>
<td>2.013514</td>
<td>1.086559</td>
</tr>
<tr>
<td>3.071582</td>
<td>0.843315</td>
<td>1.906360</td>
<td>1.842991</td>
</tr>
</tbody>
</table>

**Figure 19: Cosinor: Comparison of amplitudes at Fourier periods demonstrates that the CAT Cosinor (bottom) and Periodogram raw result (top) are equivalent to at least 6 decimal places.**

CAT has implemented the single cosinor, with a multiple-components cosinor near completion. The multiple-components model fits a combination of multiple cosine
Curves of selected periods, and can model more complex curves (Figure 20). Cosinor, like all CAT functions, can estimate circadian or other periodicities, from years to seconds, etc. Calculations can be performed at any frequency, not limited to the Fourier frequencies. The single-component cosinor in CAT allows specific frequencies/periods to be selected for any analysis. Any subsection of the time series can be designated for analysis. CAT Cosinor produces both graphical and numeric listings of results.

Reference Time selection: As we have seen, the reference time is important in estimating rhythms. CAT Cosinor allows the user to specify this as a parameter.

Period selection: One or more specific periods can be selected for analysis, or CAT can be instructed to analyze a range of periods. The user can specify a custom range of periods for analysis, or allow CAT by default to do the calculations for the set of all Fourier periods for the data set, from T/1 (1 cycle per T), to 1/2Δt, ending with the Nyquist. This produces the equivalent of a periodogram, although unlike a
periodogram, it can be performed on non-equidistant data. Whenever a range is specified (e.g., 252 - 9.5), an increment can also be considered (e.g., .5), identifying the set of periods to be assessed (e.g., 252/1, 252/1.5, 252/2, 252/2.5... 252/9.5). In this way, the cosinor version of the periodogram produces the equivalent of a periodogram, with the addition of intermediate frequencies!

When the first period estimated is T (record length in time) and the harmonic increment is 1, the periods correspond to Fourier frequencies. Using a fractional harmonic increment allows estimates of rhythm characteristics at intermediate periods. This is advantageous in identifying the period associated with the largest amplitude (or rather percentage rhythm), even if it is not a Fourier period. It should be remembered, however, that this approach may allow a more accurate point estimate of the period, but it does not improve the uncertainty with which the period can be estimated.

**Results:** For each period considered, results show the estimates of the MESOR, amplitude, phase, and a standard error for each rhythm.

Figure 21: CAT Cosinor: Graph of MESOR, Amplitude, Acrophase and P, as they change with the periods estimated, from 6 to 168 hours. Green dots are S.E. on either side of the parameter estimate.
parameter. An F-test is used for rhythm detection, yielding the significance level
associated with the fitted curve and the corresponding percentage rhythm ($R^2$), for the
period assessed. When multiple periods are selected, a graph is produced for MESOR,
amplitude, phase and P-value, showing how each varies for the periods investigated
(Figure 21).

Time span selection: CAT Cosinor can also be configured to calculate a specified
subset of the time series. Any period selection can be combined with any time span
selection. The progressive analysis, discussed next, uses successive spans of time.
The analysis selected can also be performed on multiple variables in a single run (sequentially) by selecting columns from the input file to be processed, where each column is an individual variable. For example, to run an analysis in CAT on systolic blood pressure, diastolic blood pressure and heart rate, where each is a separate data column, select those three input file columns at execution.

Progressive Analyses: This method, used by other CAT functions as well, allows increased insight into how, or whether, the data series may be varying over time.

Progressive analysis breaks a set of data into subsections, or overlapping subsections, and assesses each subsection successively, constructing a heat map of rhythm amplitudes as a function of frequencies over time (Figure 22). Progressive analysis can be performed on single, multiple, or a range of periods. Output, in this case, consists of the table of values as above, in addition to a heat map, in color or black and white, of the

Figure 23: CAT Cosinor: Progressive analysis output in Word. Shown are the first few periods of the first span of 326 data points, from a 1536-point time series. For each span, all Fourier frequencies are calculated. For each frequency, MESOR, Amplitude, and Acrophase (Phi) are calculated, with their respective standard error (SE). Percent rhythm (PR) is also given for each frequency. (Complete output samples can be seen in supplemental materials.)

The analysis selected can also be performed on multiple variables in a single run (sequentially) by selecting columns from the input file to be processed, where each column is an individual variable. For example, to run an analysis in CAT on systolic blood pressure, diastolic blood pressure and heart rate, where each is a separate data column, select those three input file columns at execution.

Progressive Analyses: This method, used by other CAT functions as well, allows increased insight into how, or whether, the data series may be varying over time.

Progressive analysis breaks a set of data into subsections, or overlapping subsections, and assesses each subsection successively, constructing a heat map of rhythm amplitudes as a function of frequencies over time (Figure 22). Progressive analysis can be performed on single, multiple, or a range of periods. Output, in this case, consists of the table of values as above, in addition to a heat map, in color or black and white, of the
spans plotted over time for each period. An additional graph set similar to Figure 4 can be produced as well, showing MESOR, amplitude, phase and P-value, as they vary across the span of time being investigated.

Output formats, and output content can be configured as needed. Options for output formats include: a .txt file, containing a table of rhythm characteristics for each period considered; or a more attractively formatted Word document of the same content (Figure 23). The line graphs (Figure 21) and the heatmap (Figure 22) can be turned on or off as needed. The graphics can be produced in several formats: PDF, JPG, PNG or postscript (PS). The CAT website details usage.

Advanced Applications of Cosinor

Cosinor is a full-featured tool that lends itself to hypothesis testing (Bingham et al., 1982). A specialized cosinor-based application, the sphygmochron, demonstrates the potential for usage of the CAT cosinor, and rhythm analyses more generally. The Halberg Chronobiology Center at the University of Minnesota uses cosinor-based analyses for a wide array of chronobiologic studies on variables such as blood pressure, heart rate and hormonal determinations that display circadian and other rhythms (Cornélissen and Halberg, 1994; Cornélissen et al., 1992; Cornélissen et al., 2007b; Halberg et al., 2013). The sphygmochron is used to analyze blood pressure readings for Vascular Variability Anomalies (VVAs) that have been associated with cardiovascular disease risk in several outcome studies (Cornélissen et al., 2007a, in press; Halberg et al., 2010b; Otsuka et al., 1997; Schaffer et al., 2001). Automated blood pressure monitors
(ABPM), worn around the clock, record the full range of daily and weekly blood pressure and heart rate rhythms, as seen in Figure 24. Using the sphygmochron, blood pressure readings over a week or more are stacked, using techniques not addressed here, to present a graph of an average daily blood pressure rhythm for that individual (Figures 25 and 26). This summary is compared to gender- and age-appropriate reference data for identification of VVAs. When the blood pressure data are fitted to a 2-component cosine model, having periods of 24 and 12-hr, an excessive circadian blood pressure amplitude can be diagnosed, which is a VVA (Figure 27) (Cornélissen et al., 2004; Halberg et al., 2010b; Müller-Bohn et al., 2002). The sphygmochron produces a number of graphs of systolic and diastolic blood pressure, and heart rate, and several pages of
numerical analysis. Output from the sphygmochron can be seen in Supplemental Materials on the CAT website.

During the course of working on this project, another individual and I collected blood pressure data by wearing the ABPMs around the clock. Through analysis using sphygmochron, my own blood pressure could be seen to be in the normal range, relative to my age and gender (Figure 24). The other individual showed signs of Circadian Hyper-Amplitude Tension (CHAT) (Cornélissen et al., 2008) in intermittent weekly collections (Figure 27). The CHAT resulted primarily from high blood pressure values in early afternoon. In conference with his doctor, this individual changed the
timing of his blood pressure medication from evening to noon, in order to lower the afternoon blood pressure peaks. Currently his blood pressure readings are not showing CHAT. A paper discussing this case (Lee Gierke et al., 2013) is available in Supplemental Materials.

The diagnostic potential in the area of blood pressure, as demonstrated by the sphygmochron, exemplifies the type of health impacts that are possible when rhythm analysis is applied to therapeutic concerns (Halberg et al., 2006; Halberg et al., 2010b). It is for these reasons tools like CAT are so important, to move the nascent science of biological rhythm analysis forward. Dr. Franz Halberg believed, “The greatest promise of circadian systems is a better universal health care at less cost and, for science, much new information” (Halberg et al., 2003a).

Figure 27: Sphygmochron: Diagnostic analysis of a vascular variability anomaly, CHAT. Red fill quantifies excessive amplitude.
Summary

A package of complementary rhythm analysis functions has been presented, capable of analyzing periodic data from chronobiologic as well as other areas of study. Used in conjunction, these tools perform robust and rigorous analyses of rhythmic data. Visualization tools aid in assessing needed data adjustments, and a variety of techniques allow for both a visual understanding of the data under analysis, as well as rigorous quantitative analysis. CAT functions are adaptable to accommodate study aims ranging from initial investigations, seeking to establish the presence of a rhythm and its general characteristics, to detailed quantitative studies of individual rhythms, testing a specific hypothesis. The visual analysis techniques of Actogram, Smoothing, Autocorrelation and CrossCorrelation provide a high-level, macroscopic understanding of the rhythm behavior, supporting development of study aims. Whereas, the detailed quantitative cosinor results, with measures of uncertainty, support sophisticated analyses and hypothesis testing. Examples of usage on biological data, and potential impacts in the area of health have been provided.

CAT is written in R (R Project, 2012), and open source, so easily extensible, for the advanced user. Yet it is easy to use and can be executed from a desktop shortcut, calling one of the many batch files from the CAT website library, or run from the programmer-friendly RStudio GUI console. Batch files from the CAT library can be easily edited to configure the parameters for customized scenarios. The CAT website supports usage, including documentation on installing R, CAT parameterization and execution,
sample input and output files, as well as fully functional vignettes users can run to validate their setup and learn more about CAT. Vignettes include sample data files, parameterization batch files and expected output, and can be found on the CAT site: http://564394709114639785.weebly.com

**Future considerations and improvements**

The Chronomics Analysis Toolkit, CAT, will continue to be developed in cooperation with the Halberg Chronobiology Center and the University of MN. Planned future enhancements, among others, include:

- Extend data input format options
- Addition of multiple-component single cosinor, and other extensions
- Graphical user interface for configuring and running CAT
- Enhanced error analysis and reporting
- Submit CAT to the Comprehensive R Archive Network (CRAN) of the R Project for Statistical Computing, making it more easily accessible in a standard format

Data conditioning enhancements:

- Noise reduction using a Butterworth, or other optional filters
- Binning and imputation enhancements

**Supplemental Materials**

Supplemental materials related to applications of cosinor, and data collection performed during the development of this paper can also be found on the CAT site: http://564394709114639785.weebly.com/ under MBS Supplemental Materials.
REFERENCES


Technology for Biological Rhythm-Guided Therapy and Prevention of Diseases, (John Wiley & Sons, Inc.).


