

REVIEW

Open Access

Cosinor-based rhythmometry

Germaine Cornelissen

Correspondence:
corne001@umn.edu
Halberg Chronobiology Center,
University of Minnesota,
420 Delaware Street SE, 55455
Minneapolis, MN, USA

Abstract

A brief overview is provided of cosinor-based techniques for the analysis of time series in chronobiology. Conceived as a regression problem, the method is applicable to non-equidistant data, a major advantage. Another dividend is the feasibility of deriving confidence intervals for parameters of rhythmic components of known periods, readily drawn from the least squares procedure, stressing the importance of prior (external) information. Originally developed for the analysis of short and sparse data series, the extended cosinor has been further developed for the analysis of long time series, focusing both on rhythm detection and parameter estimation. Attention is given to the assumptions underlying the use of the cosinor and ways to determine whether they are satisfied. In particular, ways of dealing with non-stationary data are presented. Examples illustrate the use of the different cosinor-based methods, extending their application from the study of circadian rhythms to the mapping of broad time structures (chronomes).

Keywords: Chronobiology, Chronome, Circadian, Cosinor, External information, Regression, Rhythm parameters, Stationarity

Introduction

Non-random variations are found as a function of time at the cellular level, in tissue culture, as well as in multi-cellular organisms at different levels of physiologic organization [1]. Multi-frequency rhythms usually account for a sizeable portion of the variability [2]. While there is presently much interest in studying circadian rhythms, the biological time structure covers many different ranges of periods beyond the 24-hour day, from fractions of seconds in single neurons to seconds in the cardiac and respiratory cycles, and a few hours in certain endocrine functions. Cycles with periods of about a week, about a month, and about a year are also ubiquitous, as are some other newly discovered cycles with periods of about 5 and 16 months, and much longer periods [3].

The partly built-in nature of circadian rhythms [4,5] is now widely accepted, as is the fact that they are amenable to synchronization by cycles in the environment (e.g., lighting and feeding schedules) [6]. More generally, environmental geophysical cycles such as the day-light cycle, the tides, the phases of the moon, the seasons, as well as a host of other cycles shared between living organisms and the environment in which they evolved, all serve as synchronizers for partly endogenous rhythms [7,8].

The application of chronobiology and its concepts to biology and medicine depends upon the quantitative evaluation of data collected as a function of time. The inclusion of time as a primordial factor in chronobiological investigations broadens the scope of methods for data analysis. The methods presented herein serve the purposes of rhythm

detection and parameter estimation, with applications in the early diagnosis of altered rhythm characteristics indicative of a heightened risk, the optimization of treatment by timing, and a wider understanding of how our physiology is influenced by our environment.

Data collection and study design

Before turning to the methodology itself, it is important to consider aspects of data collection and study design [9] bearing on the choice of analytical tools used for data analysis. Biological data (Y_i , $i = 1, 2, \dots, N$) are typically obtained by having a clock trigger the system (instrument, sensor) to measure a biological variable, yielding a set of data at discrete sampling times (t_i , $i = 1, 2, \dots, N$). Whether the variable examined is discrete (e.g., mitotic counts) or continuous (e.g., oral temperature), the numerical values attributed to Y_i are limited in accuracy and precision by the instrumentation used. Any finite variation of a biological system takes place during a non-zero time interval rather than instantaneously. In terms of data analysis, this means that there is a cut-off frequency f_S beyond which the spectrum of the biological variations is practically zero [10]. Whether the transducer used to measure a given biological variable is analog or digital, it takes a certain time for it to respond and deliver a reading, so that rhythms with periods shorter than this response time cannot be assessed, and rhythms with a period close to it will be distorted [10]. In other words, a cut-off frequency f_T can be defined as the minimal frequency such that for $f > f_T$ the output signal remains practically constant (no variation in the data can be assessed). This means that too dense measurements are redundant and do not bring additional information. In the case of equidistant data obtained at Δt intervals, it has been recommended to choose $\Delta t \leq 1/4f_T$ to assure a good approximation of $Y(t)$ [11].

For chronobiological applications, this sampling requirement (to be able to reconstruct changes as a function of time in the context of the theory of signal processing) is often difficult to meet. Instead, sampling is used in its statistical meaning, where it refers to the selection of a few items from a population to draw inferences for that population. In terms of a data series, the selection of a sampling interval $\Delta t > 1/f_T$ allows only some features of the biological variations to be assessed. In the absence of external information, for data collected over an observation span T , only oscillations with periods in the range of T up to $2\Delta t$ can be assessed. The resolution with which a signal's period can be determined also depends on T : in frequency terms, the smallest difference in frequency between two distinct signals is $1/T$. The highest (no longer assessable) frequency, $1/2\Delta t$, is called the Nyquist frequency, f_N . Within the field of information theory, this is known as the Nyquist-Shannon theorem, which states that if a function $Y(t)$ contains no frequencies higher than f_N , it is completely determined by sampling $Y(t)$ at intervals of $1/2f_N$.

Ostle [12] defines the design of an experiment as the complete sequence of steps taken ahead of time to ensure that the appropriate data are obtained in a way which allows for an objective analysis leading to valid inferences with respect to the stated problem. These steps include the statement of objectives, the formulation of hypotheses, and the choice of design and experimental procedure and of the statistical methods to be used. Principles underlying experimental designs rely on replication, randomization and control. Replication relates to repeated measurements to obtain an

estimate of uncertainty (experimental error or noise) used to derive statistical significance (P-values) and confidence intervals. Noise originates from variations in the biological system considered not to be part of the deterministic portion of the signal, from errors external to the system (errors of experimentation, of observation, and/or of measurement), and from the transducer and sampler (instrumentation error). Reducing the experimental error increases the precision of experiments. Randomization is an important aspect of study design that allows researchers to proceed as though the assumption of independence of the observation errors is true, which is critical in applying a test of significance. Although randomization cannot guarantee independence, it reduces the correlation that tend to characterize errors associated with experimental material (experimental unit or data) adjacent in space or time, while also improving accuracy. Control relates to the amount of balancing, blocking and grouping of the experimental units [12].

The number of replications needed for a given probability of detecting a given difference with statistical significance depends on the standard error per experimental unit [13]. This means that small sample sizes can easily detect large differences, whereas small differences require larger sample sizes. When dealing with rhythmic variables, a sizeable portion of the variance stems from the rhythmic variation. Assessing the rhythmic behavior is thus important to reduce the error term. One important feature of chronobiological study designs is that rhythm stage is often the primary factor, as when assessing the relative efficacy or toxicity of a given treatment administered at different stages of the circadian rhythm. The power of testing for a time effect is usually only slightly affected by the number of timepoints considered when results are analyzed by cosinor, but not when performing an analysis of variance. This difference in approach accounts in part for the controversy between classical designs advocating fewer test groups [14] and chronobiological designs recommending at least 6 timepoints per cycle [15-17].

In the framework of chronobiological study designs, three kinds of data can be distinguished, which determine the choice of method for their analysis and how the results can be interpreted. Longitudinal sampling corresponds to obtaining data on the same individual (experimental unit) as a function of time. One example is the around-the-clock monitoring of blood pressure at about 30-minute intervals for 7 days. Results apply to this particular individual. Transverse (cross-sectional) sampling consists of obtaining only one value per individual (experimental unit), different individuals providing data at the same or different sampling times. Time series of survival times are one example of transverse data. When individuals represent a random sample of a given population, results can be generalized for that population. Hybrid (linked cross-sectional) sampling consists of taking a few serial measurements from several individuals (experimental units). For instance, circulating prolactin is determined at 20-minute intervals for 24 hours in women at low or high familial risk of developing breast cancer later in life. The circadian rhythm can be determined for each woman and summarized across all women in each group for assessing any difference as a function of breast cancer risk [18]. When individuals represent a random sample of their respective populations, results can be generalized to these populations.

Quite generally, but particularly when sampling is performed on more than a single individual, it is important that they are synchronized. Synchronizers (environmental

periodicities determining the temporal placement of biological rhythms) serve this purpose. The rest-activity schedule or the light–dark regimen are effective synchronizers and can be used to determine a reference time (such as time of awakening or light onset in preference to clock hours such as local midnight). Staggered lighting regimens have been used to facilitate data collection in the experimental laboratory [19], making it possible to collect data over several days [20]. Marker rhythms [21] are a useful check of whether synchronization has been achieved, further providing an internal reference. Activity, temperature, heart rate and blood pressure are some useful marker variables that can easily be monitored longitudinally. For instance, blood pressure has been used to guide the timing of administration of anti-hypertensive medication while also providing information regarding the patient's response to treatment [22].

Summary statistics

Before proceeding with any data analysis, it is recommended to first plot the data as a function of time. Such a chronogram can be informative in several ways. The presence of obvious rhythmicity may be recognized and its relative prominence as compared to the noise may be qualitatively (macroscopically) assessed. When sampling covers several cycles, some measure of the cycle-to-cycle variability can be gained. The presence of any increasing or decreasing trends can be observed, as is the existence of any outliers. After curve fitting, a chronogram of residuals can also provide valuable information regarding the adequacy of the model, and the need for data transformation.

A histogram should also be prepared to obtain an estimate of the mean value with its standard deviation, and to check on the assumption of normality. For instance, a long-tailed distribution is indicative of the need for data transformation. Alternatively, the use of robust methods (such as those based on ranks; [23]) may be indicated.

When prior information suggests the presence of a rhythm with known period, stacking the data over a single cycle reduces the noise and reveals the rhythm's waveform. Historically, this approach was used by Franz Halberg to resolve confusing variability in blood eosinophil counts [24-26]. It was also instrumental in showing that the circadian rhythm in core temperature of Fischer rats persisted after the bilateral lesioning of the suprachiasmatic nuclei, albeit with a reduced amplitude and a phase advance [27,28]. It should be noted, however, that stacking the data over an assumed period may yield spurious results if the signal's period differs from its assumed value. For this reason, it is highly recommended to analyze the original data first before proceeding with any stacking. For instance, it is not uncommon to present data by calendar month, even when the data have been collected over several years. This procedure limits the ability to resolve any periodic signal with a period different from precisely 1 year or its harmonic terms (6, 4, 3 months, ...). Stacking the data over a period that has been validated can be complemented by an analysis of variance testing for a time effect when data are binned into a given number of classes of equal duration covering the full cycle (e.g., six 4-hourly classes covering 24 hours). An F-test then serves for testing the equality of class means. While this procedure remains applicable for non-equidistant data, the result depends on the choice of the number of classes used for binning and on the choice of the reference time [29].

Single cosinor

Historically, the single cosinor was developed to analyze short and sparse data series [2,30-32]. Periodograms and classical spectra originally used in chronobiology [33,34] required the data to be equidistant and to cover more than a single cycle. Whereas some spectral analysis techniques are now available to analyze non-equidistant data [35-37], algorithms available in most software packages remain limited to the case of equidistant data.

Least squares procedures do not have this limitation. They are thus useful in curve-fitting problems, where it is desirable to obtain a functional form that best fits a given set of measurements. Although periodic regression presents its own limitations, being sensitive to outliers and not having any constraint to conserve the variance in the data, it possesses two important features: first, when data are equidistant, results at Fourier frequencies are identical to those of the discrete Fourier transform [38]; and second, it advantageously uses prior information. Thus, after the existence of ubiquitous circadian rhythms was demonstrated, it was possible to apply the single cosinor method in many experiments aimed at determining the times of highest efficacy and lowest toxicity in response to a variety of drugs and other stimuli by fitting a 24-hour cosine curve to 6 values, 4 hours apart, each value representing the number of experimental animals that survived a given intervention applied at one of the 6 timepoints when overall about 50% of the animals had died. These results led to the fields of chronopharmacy and chronotherapy [39-42].

Single-component cosinor

Notably in studies of circadian rhythms, it is indeed possible to assume that the period is known, being synchronized to the external 24-hour cycle. The regression model for a single component can be written as

$$Y(t) = M + A\cos(2\pi t/\tau + \phi) + e(t) \quad (1)$$

where M is the MESOR (Midline Statistic Of Rhythm, a rhythm-adjusted mean), A is the amplitude (a measure of half the extent of predictable variation within a cycle), ϕ is the acrophase (a measure of the time of overall high values recurring in each cycle), τ is the period (duration of one cycle), and $e(t)$ is the error term (Figure 1).

When τ can be assumed known, using well-known trigonometric angle sum identity, the model can be rewritten as

$$Y(t) = M + \beta x + \gamma z + e(t) \quad (2)$$

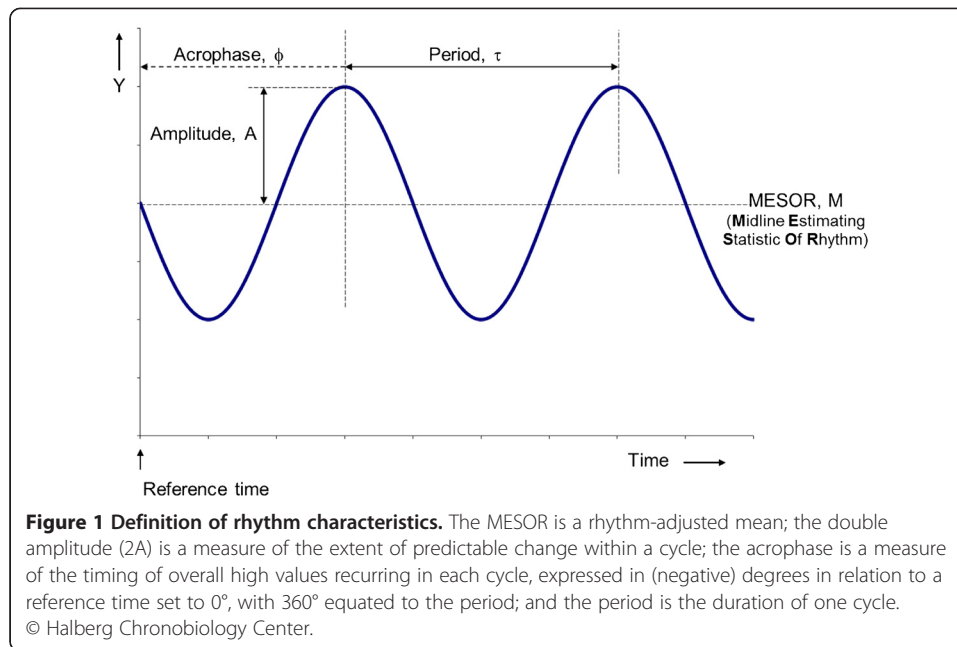
where

$$\beta = A\cos\phi; \gamma = -A\sin\phi; x = \cos(2\pi t/\tau); z = \sin(2\pi t/\tau).$$

The principle underlying the least squares method is the minimization of the residual sum of squares (RSS), that is the sum of squared differences between measurements Y_i (obtained at times t_i , $i = 1, 2, \dots, N$) and the values estimated from the model at corresponding times

$$\text{RSS} = \sum_i [Y_i - (\hat{M} + \hat{\beta}x_i + \hat{\gamma}z_i)]^2 \quad (3)$$

This approach is valid when all individual standard deviations are equal, as is often the case.



Estimates for M , β , and γ are obtained by solving the normal equations, obtained by expressing that RSS is minimal when its first-order derivatives with respect to each parameter are zero.

The normal equations are

$$\begin{aligned} \sum Y_i &= MN + \beta \sum x_i + \gamma \sum z_i \\ \sum Y_i x_i &= M \sum x_i + \beta \sum x_i^2 + \gamma \sum x_i z_i \\ \sum Y_i z_i &= M \sum z_i + \beta \sum x_i z_i + \gamma \sum z_i^2 \end{aligned} \quad (4)$$

or in matrix form

$$\begin{pmatrix} \sum Y_i \\ \sum Y_i x_i \\ \sum Y_i z_i \end{pmatrix} = \begin{pmatrix} N & \sum x_i & \sum z_i \\ \sum x_i & \sum x_i^2 & \sum x_i z_i \\ \sum z_i & \sum x_i z_i & \sum z_i^2 \end{pmatrix} \begin{pmatrix} M \\ \beta \\ \gamma \end{pmatrix} \text{ or } \mathbf{d} = \mathbf{S}\mathbf{u} \quad (5)$$

Estimates of M , β and γ (or vectorially, \mathbf{u}) are thus obtained as

$$\hat{\mathbf{u}} = \mathbf{S}^{-1} \mathbf{d} \quad (6)$$

Estimates for the amplitude and acrophase can be derived from the estimates of β and γ by the following relations

$$\begin{aligned} \hat{A} &= (\hat{\beta}^2 + \hat{\gamma}^2)^{1/2} \\ \hat{\phi} &= \arctan(-\hat{\gamma}/\hat{\beta}) + K\pi \quad \text{where } K \text{ is an integer} \end{aligned} \quad (7)$$

The correct value of $\hat{\phi}$ is determined by taking into account the signs of $\hat{\beta}$ and $\hat{\gamma}$.

For rhythm detection, the total sum of squares (TSS) is partitioned into the sum of squares due to the regression model (MSS) and the residual sum of squares (RSS). TSS is the sum of squared differences between the data and the arithmetic mean. MSS is

the sum of squared differences between the estimated values based on the fitted model and the arithmetic mean. As noted above, RSS is the sum of squared differences between the data and the estimated values from the fitted model.

$$TSS = MSS + RSS \text{ or } \sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2 \quad (8)$$

The model is statistically significant when the model sum of squares is large relative to the residual sum of squares, as determined by the F test

$$F = (MSS/2)/(RSS/(N-3)) \quad (9)$$

where 2 and N-3 are the numbers of degrees of freedom attributed to the model ($k = 3$ parameters - 1) and to the error term ($N - k$). The null hypothesis (H_0) that there is no rhythm (the amplitude is zero) is rejected when $F > F_{1-\alpha}(2, N-3)$, where α relates to the chosen probability level for testing H_0 .

For parameter estimation, it seems reasonable to consider the MESOR (M) separately and (β, γ) together. The $1-\alpha$ confidence interval for \hat{M} is then given by

$$\hat{M} \pm t_{1-\alpha/2}(N-3) \hat{\sigma} \sqrt{s_{11}^{-1}} \quad (10)$$

where s_{ij}^{-1} are the elements of S^{-1} ,

$$\hat{\sigma} = [RSS/(N-3)]^{1/2} \quad (11)$$

and $t_p(f)$ denotes the p th probability point of Student's t on f degrees of freedom. The covariance matrix for $(\hat{\beta}, \hat{\gamma})$ is given by

$$\hat{\sigma}^2 \begin{pmatrix} s_{22}^{-1} & s_{23}^{-1} \\ s_{32}^{-1} & s_{33}^{-1} \end{pmatrix},$$

from which a $1-\alpha$ confidence region for $(\hat{\beta}, \hat{\gamma})$, or equivalently for $(\hat{A}, \hat{\phi})$ can be derived. It is based on the F-statistic used for rhythm detection, being evaluated at the estimated value of \hat{A} instead of at $A = 0$. The resulting equation is that of an ellipse (Figure 2):

$$\sum (x_i - \bar{x})^2 (\beta - \hat{\beta})^2 + 2 \sum (x_i - \bar{x})(z_i - \bar{z}) (\beta - \hat{\beta})(\gamma - \hat{\gamma}) + \sum (z_i - \bar{z})^2 (\gamma - \hat{\gamma})^2 \leq 2\hat{\sigma}^2 F_{1-\alpha}(2, N-3) \quad (12)$$

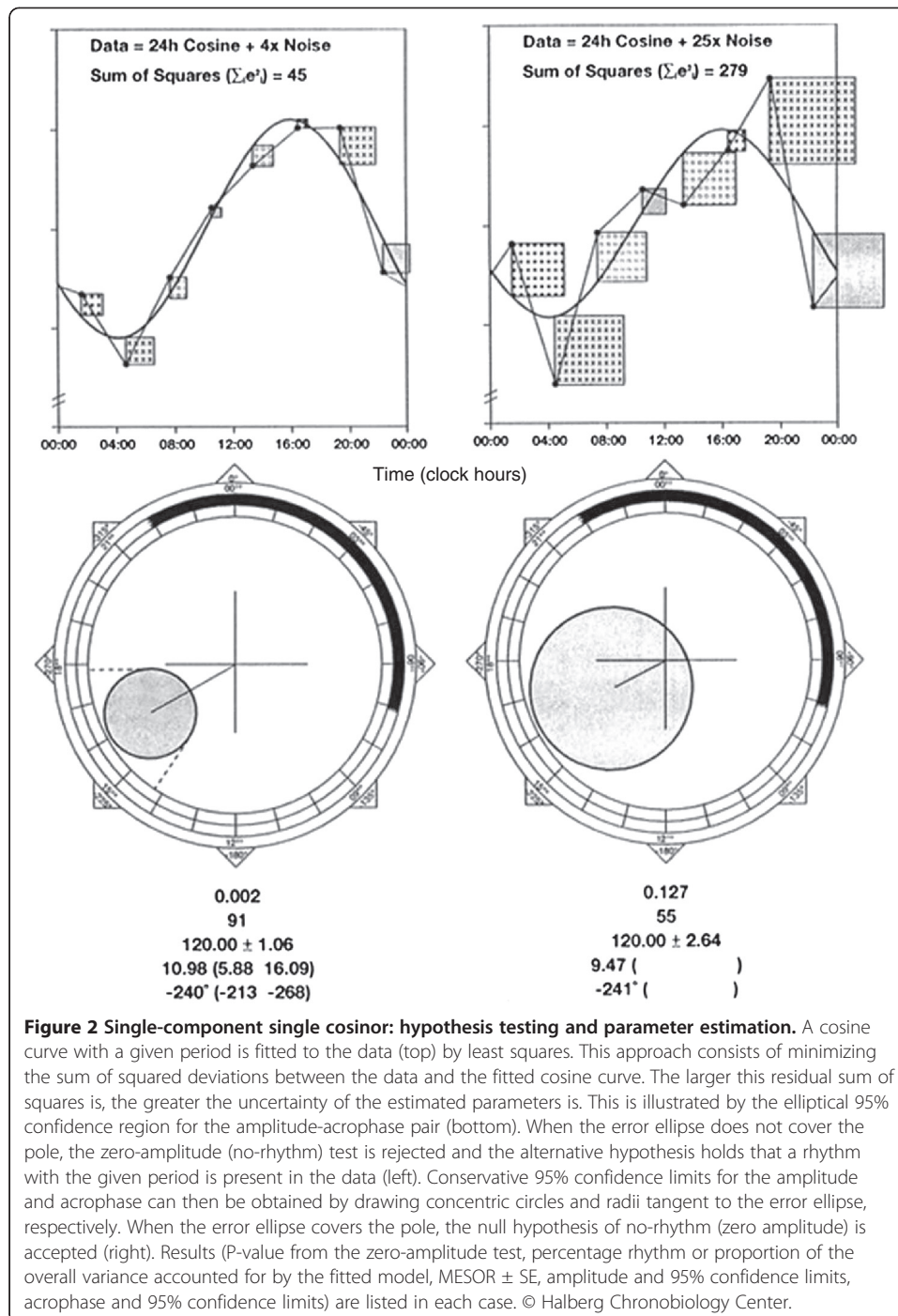
where

$$\bar{x} = \left(\sum_i X_i \right) / N$$

and

$$\bar{z} = \left(\sum_i Z_i \right) / N$$

The region delineated by this ellipse represents the confidence region for the rhythm parameters. Conservative confidence intervals for \hat{A} and $\hat{\phi}$ are obtained by computing the minimal and maximal distances from the pole (zero) to the error ellipse and by drawing tangents from the pole to the error ellipse, respectively [31]. These confidence limits are conservative in the sense that the area they delineate is larger than the confidence ellipse. Limits closer to those corresponding to the α -level chosen for testing can also be computed [43].



Standard errors (SEs) for \hat{A} and $\hat{\phi}$ can also be derived from the covariance matrix for $(\hat{\beta}, \hat{\gamma})$, by using Taylor's series expansion:

$$\begin{aligned} SE(\hat{A}) &= \hat{\sigma} [s_{22}^{-1} \cos^2 \hat{\phi} - 2s_{23}^{-1} \sin \hat{\phi} \cos \hat{\phi} + s_{33}^{-1} \sin^2 \hat{\phi}]^{1/2} \\ SE(\hat{\phi}) &= \hat{\sigma} [s_{22}^{-1} \sin^2 \hat{\phi} + 2s_{23}^{-1} \sin \hat{\phi} \cos \hat{\phi} + s_{33}^{-1} \cos^2 \hat{\phi}]^{1/2} / \hat{A} \end{aligned} \quad (13)$$

Regression diagnostic tests

It should be noted that the P-value obtained from the F-test and the corresponding confidence limits derived for \hat{M} , $\hat{\beta}$ and \hat{y} are valid only if assumptions underlying the use of the least squares procedure are satisfied. These assumptions are (1) the model fits the data well, (2) the residuals are normally distributed, (3) the variance is homogeneous, (4) the residuals are independent, and (5) the parameters do not change over time.

Model adequacy:

Goodness of fit can be examined when replicates are available, either from multiple data collection at different timepoints or from data covering multiple cycles. RSS can then be further partitioned into the “pure error” and the “lack of fit”. An F-test comparing the pure error and lack of fit sums of squares provides a test of the model adequacy [44]. The pure error sum of squares (SSPE) is defined as the sum of squared differences (across all timepoints) between the data collected at a given timepoint and their respective arithmetic mean, whereas the sum of squares ascribed to lack of fit (SSLOF) is obtained by subtracting SSPE from RSS

$$\begin{aligned} \text{SSLOF} &= \text{RSS} - \text{SSPE} \\ \text{SSPE} &= \sum_i \sum_l (Y_{il} - \bar{Y}_i)^2 \end{aligned} \quad (14)$$

with $\bar{Y}_i = (\sum_l Y_{il})/n_i$ where n_i is the number of data collected at time t_i .

The appropriateness of the model is rejected if

$$F = [\text{SSLOF}/(m-1-2p)]/[\text{SSPE}/(N-m)] > F_{1-\alpha}(m-1-2p, N-m) \quad (15)$$

where m is the number of timepoints and p is the number of (cosine) components in the model ($p = 1$ for the single-component cosinor).

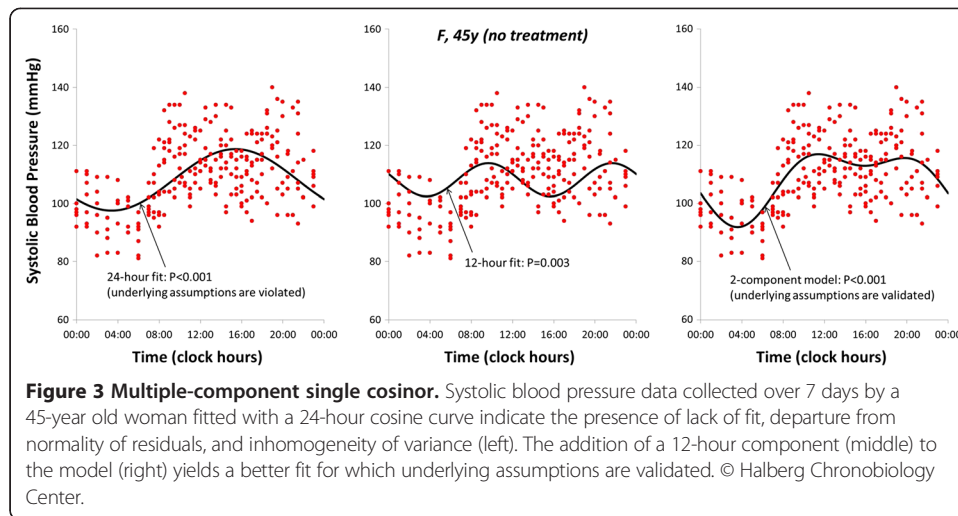
In the presence of lack of fit, adding components in the model may be considered (Figure 3).

Normality of residuals:

The rankit plot provides an attractive visual technique to test normality [44,45]. In this test, the errors (e_i) are sorted by increasing order and expected values of a normal sample of size N with zero mean (rankits, z_i) are calculated. If the residuals are normally distributed, the regression of e_i on z_i is a straight line. The Shapiro-Wilk test of normality can be applied for small sample sizes ($N \leq 50$) [46]. For larger sample sizes, a chi-square test of goodness of fit can be used [23] by comparing expected and observed frequencies of residuals grouped in classes.

Homogeneity of variance:

The variance of the error term sometimes depends on the expected level of the variable examined. This can be the case of hormonal data such as melatonin that assumes large values by night but only very low values during the day. All day-time values being very small, the variance is also small. But as night-time data can vary greatly, their variance is also inflated. Deviation from variance homogeneity can be revealed by a plot of residuals as a function of the fitted values [47]. A horizontal band around zero in such a plot indicates that the assumption is valid. If it is violated, data can be transformed, for instance by taking their square root or their logarithm. A numerical test can also be



performed by fitting the model to the square of the estimated values instead of the data in order to obtain residuals re_i . If

$$F = (N-2p-2)r^2 / (1-r^2) > F_{1-\alpha}(1, N-2p-2) \quad (16)$$

where r denotes the regression coefficient of re_i on e_i , the assumption of homogeneity of variance is rejected.

Independence of residuals:

While violation of independence usually does not affect the estimate of the parameter themselves, their confidence intervals tend to be under-estimated [48]. When residuals are positively correlated, they tend to assume the same sign for long sequences. The runs test is a non-parametric test that allows to test whether sequences (runs) of positive and negative residuals occur randomly. Specifically, if successive errors are independent, there cannot be any regular sequences, either too long or too short. In other words, the number of runs cannot be too small or too large, respectively. For a given sample size, tables list limits for acceptable numbers of runs compatible with the assumption of independence [23,49,50].

When residuals are correlated, the data can either be low-passed filtered by averaging or decimation (using only one every k values, thereby lengthening the sampling interval from Δt to $k\Delta t$). A slightly different model can also be considered wherein the error term is replaced by an autoregressive error term [10,51,52].

Stationarity:

The problem of stationarity arises primarily in long time series, when the MESOR, amplitude, acrophase and/or period can change as a function of time. This may occur, for instance, when a person being monitored travels across time zones (e.g., from the USA to Europe). A head cold or pain can also bring about transient changes in circadian rhythm characteristics of variables such as body temperature, blood pressure or heart rate. Changes in period can be anticipated when time clues (environmental synchronizers) are removed. They are also present in the case of components with no strong environmental synchronizer. This concerns primarily non-photic cycles such as the about 1.3-year transyear and the about 5-month cis-half-year found in solar wind speed and solar flares, respectively. These components are wobbly by nature, even in

the environment. Counterparts in biology have been found, as discussed elsewhere [3,53-56]. There are several approaches available to analyze non-stationary time series, such as wavelets, short-term Fourier transforms, and gliding spectral windows complemented by chronobiologic serial sections, as discussed below.

For short and sparse time series, for which the cosinor method was originally developed, underlying assumptions are usually valid (or at least statistical power is not sufficient to invalidate them). With longer and denser data, the likelihood of violating one or several underlying assumptions increases. Violation of one assumption can also result in violating one or several other assumptions. For instance, when there is lack of fit, residuals tend not to be independent and may not follow a normal distribution. Confidence intervals and P-values tend to be affected more than the estimation of parameters. Even when one or several underlying assumptions are violated, the information from the cosinor analysis may be of value as long as results are properly qualified. Many of the conventional methods of data analysis depend on similar assumptions, with the exception of robust non-parametric techniques [10,57-61].

Multiple-component cosinor

The single-component cosinor is easily extended to a multiple-component model (Figure 3)

$$Y(t) = M + \sum_j A_j \cos(2\pi t/\tau_j + \phi_j) + e(t), j = 1, 2, \dots, p \quad (17)$$

Instead of solving a system of 3 equations in 3 unknowns, there are $2p + 1$ normal equations to estimate M and p pairs of (β_j, γ_j) or (A_j, ϕ_j) when τ_j are assumed known. Generally, in the normal equations $d = Su$

$$d = \sum_i x_{iv} Y_i \text{ and } Su = \sum_j u_j \sum_i x_{iv} x_{ij} \text{ for } i = 1, 2, \dots, N; \quad (18)$$

$$j = 1, 2, \dots, 2p + 1; \text{ and } v = 1, 2, \dots, 2p + 1$$

where $\{x_{ij}\}$ are the $\cos(2\pi t_i/\tau_j)$ and $\sin(2\pi t_i/\tau_j)$.

Estimates of u ($M, \beta_1, \gamma_1, \beta_2, \gamma_2, \dots, \beta_p, \gamma_p$) are obtained as

$$\hat{u} = S^{-1} X'Y \text{ where } S = X'X = \left| \sum_i x_{ij} x_{ik} \right|.$$

A confidence ellipsoid can be determined [43] from which approximate confidence intervals can be derived for each component's amplitude and acrophase, as outlined above. Computations are greatly simplified in the case of equidistant data covering an integer number of cycles [45].

A multiple-component model is useful to obtain a better approximation of the signal's waveform when it deviates from sinusoidality. For instance, a 2-component model consisting of cosine curves with periods of 24 and 12 hours has been extensively used to analyze ambulatory blood pressure data (Figure 3) [62-64]. On the average, these two components account for the larger overall variance in this case [65]. This model is usually well-suited to approximate the nightly drop in blood pressure that reaches a nadir around mid-sleep [66], the slight increase thereafter and a sharper increase after awakening, the post-prandial dip seen more prominently

in the elderly [67], and a slow decline in the evening. Whereas better-fitting models can be obtained for each individual patient, the choice of a given model used as a reference standard makes it possible to derive reference values (such as 90% prediction limits) for the model's parameters for specified populations, usually clinically healthy men or women in several age groups [62]. Deviations from these norms can then be viewed as indicative of rhythm alteration. In addition to the well-known cardiovascular disease risk associated with an elevated blood pressure MESOR, outcome studies [68-72] have determined that certain other rhythm alterations affecting the circadian amplitude and acrophase are also associated with an increase in cardiovascular disease risk [64].

Population-mean cosinor

When data are collected as a function of time on 3 or more individuals, the population-mean cosinor procedure renders it possible to make inferences concerning a population rhythm, provided the k individuals represent a random sample from that population. Each individual series is analyzed by the single- or multiple-component single cosinor to yield estimates of $\left\{ \hat{u} = \hat{M}_i, \hat{\beta}_{1i}, \hat{\gamma}_{1i}, \hat{\beta}_{2i}, \hat{\gamma}_{2i}, \dots, \hat{\beta}_{pi}, \hat{\gamma}_{pi}, i = 1, 2, \dots, k \right\}$. The goal is to make inferences concerning the population averages of the parameters, u^* . The "*" indicates that the expectations are population averages and not averages over the k individuals sampled. Individual vectors u_i are assumed to represent a random sample from a $(2p + 1)$ -variate, normal population with mean u^* . The within-individual variances are also assumed to be equal, so that the pooled estimate of variance can be estimated as

$$\hat{\sigma}^2 = \sum_j (n_j - (2p + 1)) \hat{\sigma}_j^2 / (N - (2p + 1)k) \quad (19)$$

When the sample sizes for all individuals are the same or almost the same, as is often the case in hybrid designs, the population estimates are unweighted averages of the individual parameters

$$\hat{u}^* = \sum_j \hat{u}_{jn} / k \text{ for } j = 1, 2, \dots, k \text{ and} \quad (20)$$

$$n = 1, 2, \dots, 2p + 1 \left(\hat{u}_n = \left\{ \hat{M}, \hat{\beta}_1, \hat{\gamma}_1, \hat{\beta}_2, \hat{\gamma}_2, \dots, \hat{\beta}_p, \hat{\gamma}_p \right\} \right)$$

and the population amplitudes and acrophases can be estimated using the relations

$$\hat{\beta}^* = \hat{A}^* \cos \hat{\phi}^*; \hat{\gamma}^* = -\hat{A}^* \sin \hat{\phi}^*$$

In the above conditions and assuming normality of errors and individual parameters, sample variances can be computed as

$$\hat{\sigma}_M^2 = \sum_j (\hat{M}_j - \hat{M}^*)^2 / (k-1); \hat{\sigma}_{\beta_n}^2 = \sum_j (\hat{\beta}_{nj} - \hat{\beta}^*)^2 / (k-1);$$

$$\hat{\sigma}^{\gamma n^2} = \sum_j (\hat{\gamma}_{nj} - \hat{\gamma}^*)^2 / (k-1); \hat{\sigma}_{M\beta_1}^2 = \sum_j (\hat{M}_j - \hat{M}^*)(\hat{\beta}_{1j} - \hat{\beta}_1^*) / (k-1); \quad (21)$$

and similarly for the other cross-products

where

$$j = 1, 2, \dots, k \text{ and } n = 1, 2, \dots, p$$

In the case when the population-mean cosinor can be applied separately for each trial period ($p = 1$), a confidence interval for M^* is given by

$$\hat{M}^* \pm t_{1-\alpha/2}(k-1)\hat{\sigma}_M/k^{1/2} \quad (22)$$

and a joint $1-\alpha$ confidence ellipse for $(\hat{\beta}^*, \hat{\gamma}^*)$ consists of all points (β_z^*, γ_z^*) satisfying

$$\frac{(\beta_z^* - \hat{\beta}^*)^2}{\hat{\sigma}_\beta^2} - 2r(\beta_z^* - \hat{\beta}^*)(\gamma_z^* - \hat{\gamma}^*)/\hat{\sigma}_\beta\hat{\sigma}_\gamma + \frac{(\gamma_z^* - \hat{\gamma}^*)^2}{\hat{\sigma}_\gamma^2} \leq \frac{2(1-r^2)(k-1)F_{1-\alpha}(2, k-2)}{(k(k-2))} \quad (23)$$

where

$$r = \hat{\sigma}_{\beta\gamma}/(\hat{\sigma}_\beta\hat{\sigma}_\gamma)$$

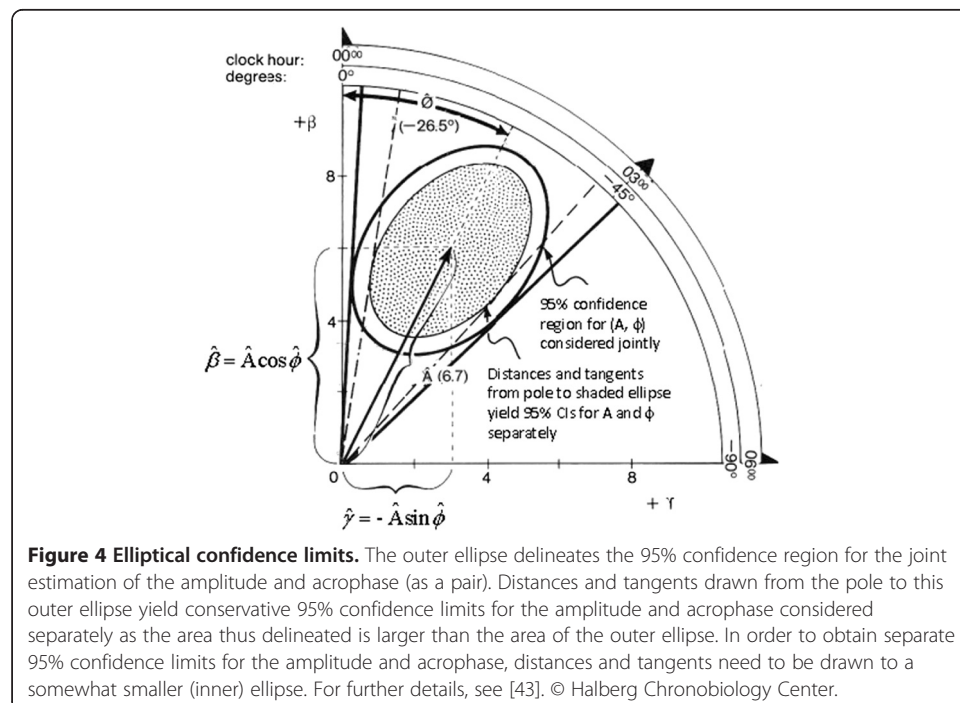
The null hypothesis of $A^* = 0$ is rejected if

$$\left[\frac{k(k-2)}{2(k-1)} \right] \left[\frac{1}{(1-r^2)} \right] \left[\frac{\hat{\beta}^{*2}}{\hat{\sigma}_\beta^2} - 2r\hat{\beta}^*\hat{\gamma}^*/\hat{\sigma}_\beta\hat{\sigma}_\gamma + \frac{\hat{\gamma}^{*2}}{\hat{\sigma}_\gamma^2} \right] > F_{1-\alpha}(2, k-2) \quad (24)$$

and approximate confidence intervals for \hat{A}^* and $\hat{\phi}^*$ can be obtained by computing the minimal and maximal distances from the pole (zero) to the error ellipse and by drawing tangents from the pole to the error ellipse, respectively (Figure 4). As for the single cosinor, closer approximate limits can also be computed [43].

Parameter tests

Test statistics have been developed to test the equality of MESORs, amplitudes and acrophases considered jointly or separately for the case of the single cosinor and the population-mean cosinor [43]. These tests can allow for a clearer interpretation of the results, for instance in a circadian experiment involving 6 timepoints 4 hours apart: Student t-tests are sometimes applied at each separate timepoint without adjustment of



the P-values for multiple testing; when differences in opposite directions are found, parameter tests may reveal a difference in the circadian amplitude in the absence of a difference in MESOR or in the circadian acrophase [73].

Least squares spectra and population-mean cosinor spectra

The circadian rhythm is often prominent. It is also ubiquitous. These features enabled the single cosinor procedure to be applied to many short data series covering no more than a single cycle in order to yield valuable information regarding the organization of the circadian system in different species. Computationally, estimates of the MESOR, amplitude and acrophase can be obtained for any trial period. This procedure, however, is valid only if there is sufficient evidence for considering this particular trial period. In the absence of such evidence, results can no longer be taken at their face value.

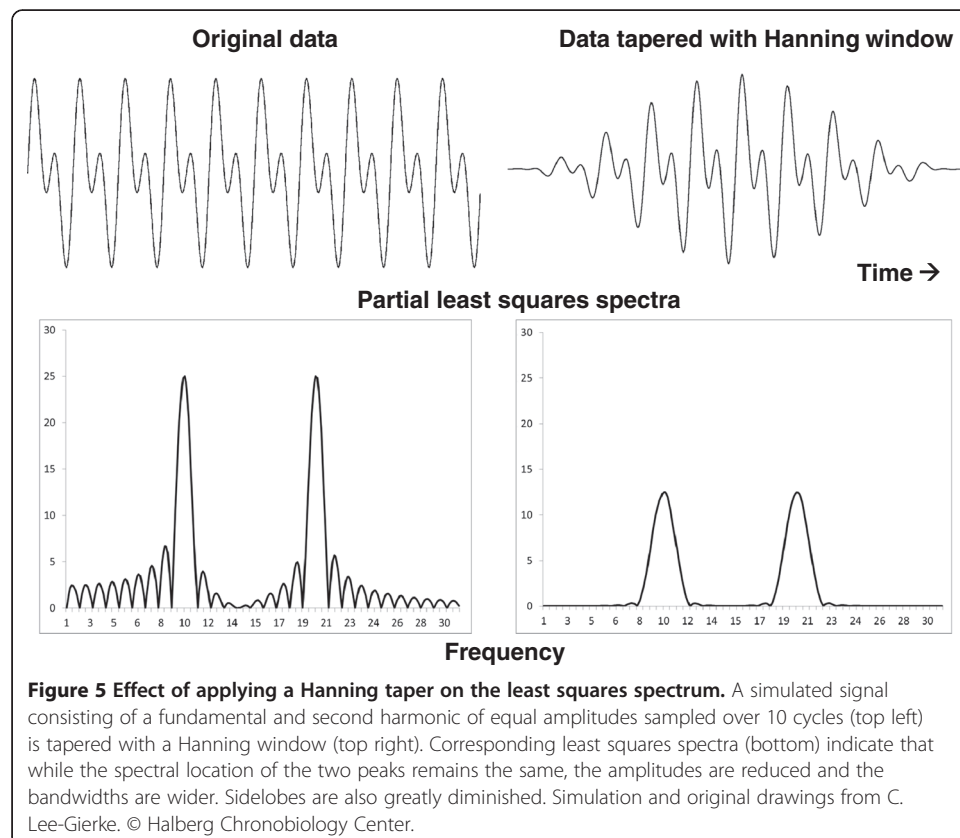
It has become much easier for chronobiologists to collect data over much longer spans and/or at much shorter intervals, but it has been more difficult to obtain series of equidistant data. Even for variables that are obtained with automated instrumentation (such as telemetry or ambulatory blood pressure monitors), it is not uncommon to have missing data or to have additional data collected manually at times different from the scheduled times. Investigations have also extended outside the circadian realm. For these reasons, a least squares approach to time series analysis remains attractive, as long as caution is properly taken in interpreting the results.

Just as a chronogram provides useful information prior to quantitative data analysis, a view of the time structure of the data in the frequency domain can also be informative. For this purpose, using the cosinor at Fourier frequencies in the range of $1/T$ (where T is the length of the data series) up to $1/2\Delta t$ (where Δt is the sampling interval) can be viewed as no more than another macroscopic view of the data. A plot of amplitudes as a function of frequency (least squares spectrum) is equivalent to a discrete Fourier transform when data are equidistant [38].

- Large spectral peaks indicate the presence of signals and provide an approximate estimate of their periods. This information can be used to validate anticipated components while also revealing the presence of other cycles. For rhythms that are anticipated, rhythm detection and parameter estimation can proceed as outlined above as long as P-values are adjusted for multiple testing [74]. Caution needs to be taken regarding non-anticipated cycles. The information thus gained can be used to design the next study or to examine other similar data series that could serve as replications. Additional analyses can be performed to determine the extent of stability of the unanticipated component, for instance by means of applying a chronobiologic serial section [21] or a gliding spectral window [75].
- Plotting log-amplitudes versus log-frequency provides useful information regarding the noise structure [65]. White noise corresponds to about the same background amplitudes across the entire frequency range. Larger background amplitudes at lower than at higher frequencies represents colored (or correlated) noise, indicating that underlying assumptions are not met, resulting in under-estimated P-values and too-liberal confidence intervals of rhythm parameters. The noise structure can in itself be valuable. It is used for instance in assessing the $1/f$ behavior of heart rate variability [76].

- Single spectral peaks are found only if the data cover an integer number of cycles. If this is not the case, the signal spreads over several spectral lines [10]. When this happens and the underlying signal was anticipated, it is possible to determine the period (frequency) corresponding to the maximal amplitude by applying the single cosinor procedure not only at the Fourier frequencies but at additional intermediary frequencies as well. Whereas this may provide a clearer picture of the signal, it should be realized that the resolution in frequency ($1/T$) remains the same, being determined by the series length, T . Tapers such as a Hanning window [77] can be used to reduce sidelobes associated with the finite observation span, but this procedure also affects the estimation of the rhythm parameters. While a Hanning taper does not affect the location of spectral peaks in a spectrum, the width of the peak is wider and the amplitude is reduced (Figure 5). It remains useful, however, for a macroscopic view of the time structure of the data.

Least squares spectra can be very helpful in exploratory analyses, but it should be realized that assumptions underlying the use of the single cosinor (notably independence and normality) are violated more often than not. Population-mean cosinor spectra are a useful complementary approach not prone to this limitation. This method is similar to the power spectrum obtained by smoothing the periodogram, which is more reliable for testing unknown periodicities [78], with the important difference, however, of retaining the phase information. The averaging (smoothing) can be done either in the frequency domain by averaging across consecutive Fourier frequencies, or in the time domain. The



population-mean cosinor spectrum uses averaging in the time domain. The method consists of subdividing the observation span T into several (e.g., k) subsections (or intervals, I) of equal length ($I = T/k$). A least squares spectrum is computed for each interval, using the same common reference time. The population-mean cosinor procedure is then applied at each trial frequency to summarize results across the k intervals. The procedure can be repeated by using different values of k . Unknown signals consistently detected by this approach may thus be viewed with added confidence.

Extended linear-nonlinear cosinor

When the period is unknown, the single cosinor model (Equations 1 and 12) can no longer be linearized in its parameters as the period is in the argument of the cosine function. Starting from an initial (guess) estimate for the period, all parameters can be estimated using iterations aimed at minimizing the residual sum of squares. Marquardt [79] developed an algorithm which performs an optimum interpolation between the Taylor series and gradient methods. He also derived a way to approximate confidence intervals for all parameters, including the period [80]. For the particular case of single-component models, Bingham offers an easily understood approach [81].

For low-frequency signals, simulated annealing [82] is another suitable method that has the advantage of not requiring the specification of initial values for the periods. This approach does not perform well, however, for very sharp signals in the higher frequency range of the spectrum. Both simulated annealing and Marquardt's nonlinear approach performed best in distinguishing two signals with close periods sampled over less than a beat cycle, when compared to other approaches [83].

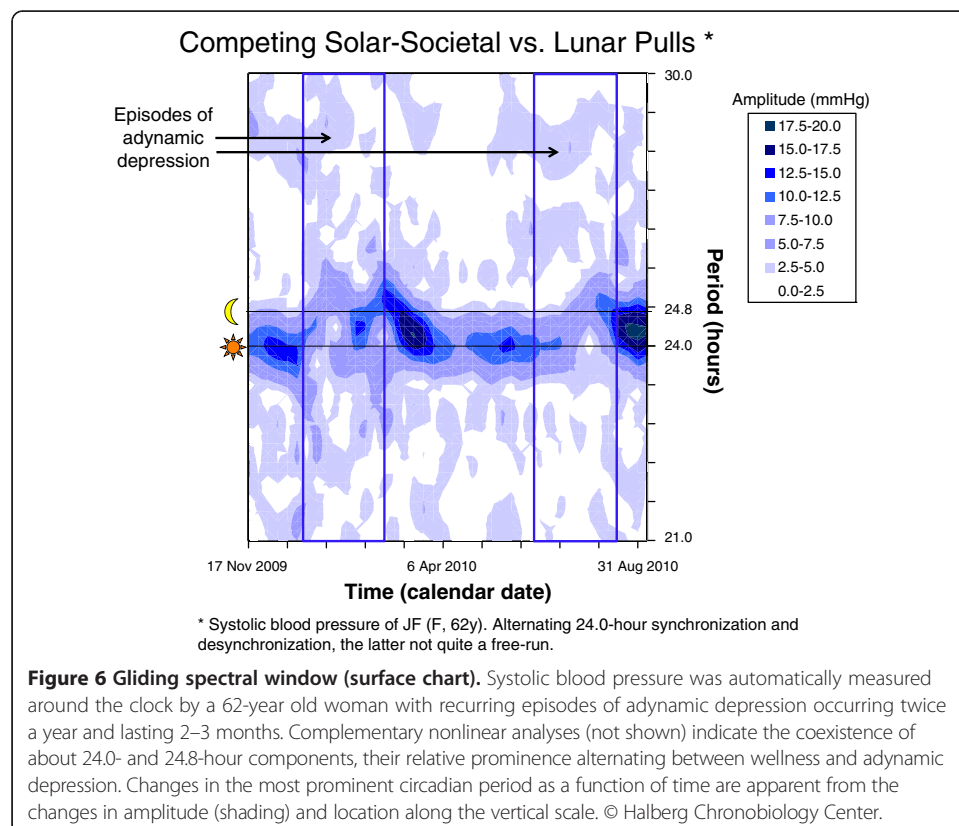
For signals with a symmetrical waveform, the nonlinear procedure can yield an acceptable estimate of the fundamental period on the basis of very short records not even covering a full cycle [84]. This is not the case, however, when the waveform is asymmetrical. Simulations indicate that about 5 cycles are needed to obtain a reasonable estimate of the period in this case, when the model fitted includes only the fundamental component. Including additional harmonic terms in the model allows the nonlinear procedure to correctly estimate the fundamental period with data covering no more than 2 cycles [84].

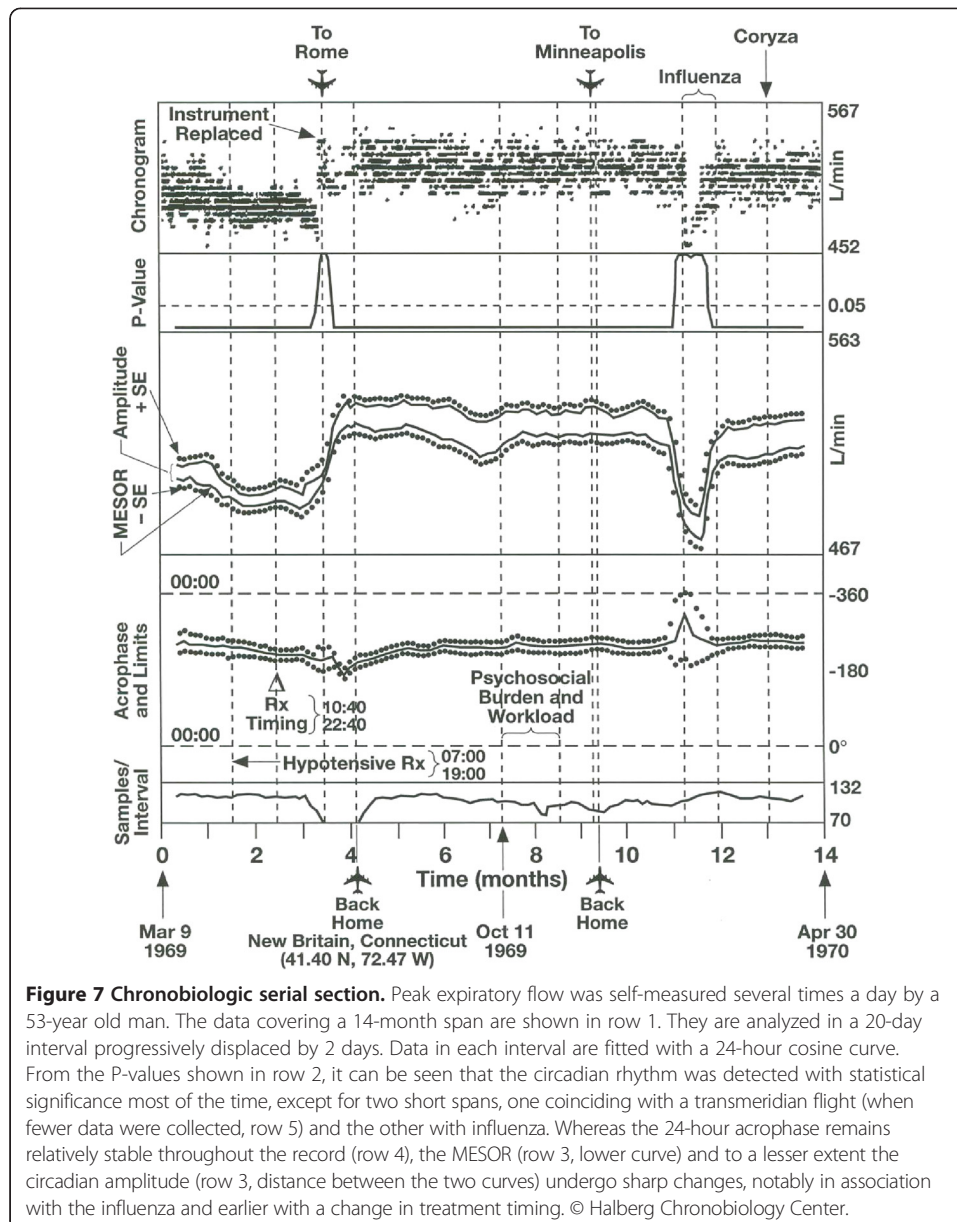
Analysis of non-stationary data

When data are equidistant or rendered equidistant by averaging and filling data gaps by interpolation, wavelets can be performed [85]. This approach has been useful to uncover components not detected earlier [86]. Short-term Fourier transforms can be used to visualize changes in the spectral structure of the data as a function of time [87]. Alternatively, gliding spectral windows [75] can be computed. The method consists of defining an interval (I) that is progressively displaced by a given increment (δt) throughout the data series. A least squares spectrum is computed over each interval over a specified frequency range. Both a fractional harmonic increment ($\delta h < 1$) and overlapping intervals ($\delta t < I$) are chosen to help visualize the time course of changes in frequency and/or amplitude occurring as a function of time in a 3D graph and/or a surface chart. In such a display, time and frequency are the two horizontal axes and amplitude is shown on the vertical axis in a 3D plot or as different shadings in a surface

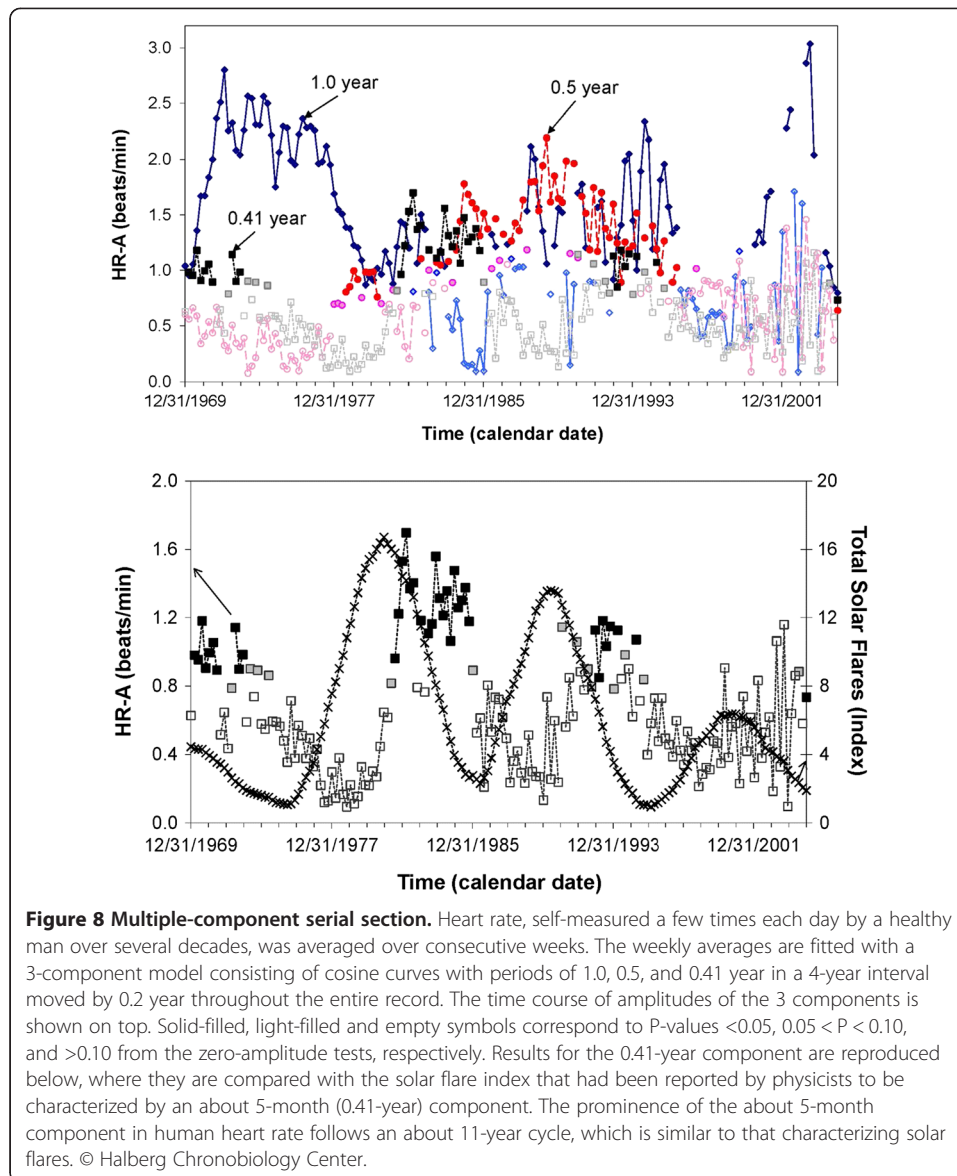
chart. One example relates to competing about 24.0- and 24.8-hour components coexisting in the physiology of an apparently seleno-sensitive woman with adynamic episodes recurring twice a year and lasting 2–3 months, as illustrated for systolic blood pressure in Figure 6. Another example illustrates the changing prominence of the about-weekly and about-daily rhythms in blood pressure and heart rate during the first 40 days of life of a clinically healthy boy [88]. Whereas the procedure can be performed on non-equidistant data, the interpretation of results is greatly helped when data are equidistant, as changes in sampling rate are also associated with changes in spectral structure appearing on the graph. A judicious choice of I and of the frequency range examined is important in order to minimize sidelobes. The use of a Hanning taper [77] is also helpful in this kind of exploratory analysis.

Whenever focus can be placed on a specified component with a given trial period, a chronobiologic serial section [21] can be performed. As for the gliding spectral window, an interval I is selected that is progressively displaced throughout the time series by δt increments. To data in each interval, the single-component single cosinor procedure is applied. To visualize the results, a chronogram is shown on top, followed by the sequence of MESORs, amplitudes and acrophases as they change as a function of time, provided with a measure of uncertainty. Corresponding P-values from the zero-amplitude test and the number of data per interval are also displayed to help interpret any change in the results. This procedure has been extensively used in studies of phase shifts associated with transmeridian flights [89], as illustrated in Figure 7, and in cases when the circadian rhythm is desynchronized from 24 hours [90,91].





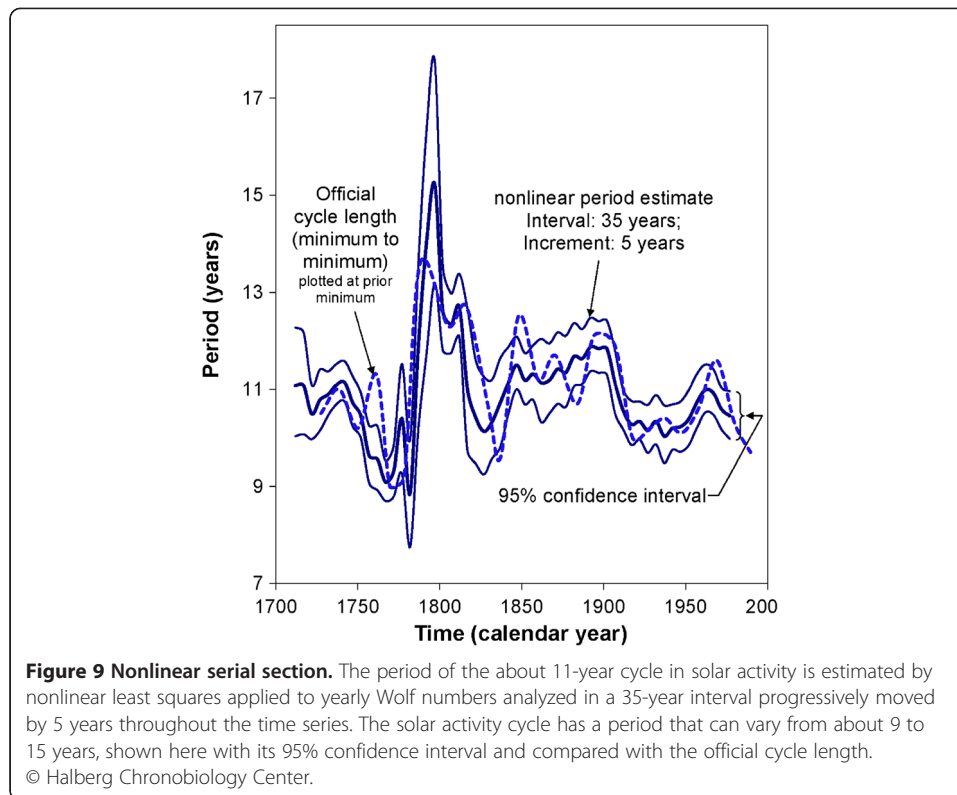
The procedure has been extended in two ways. First, a multiple-component instead of a single-component single cosinor model can be fitted in each interval. This procedure has been used for instance to illustrate that the prominence of an about 5-month component of heart rate self-measured over 4 decades by a clinically healthy man follows the about 11-year change in solar flares in which this component had been documented [92]. In this analysis, the 5-month component was fitted together with 1.0- and 0.5-year components over a 4-year interval displaced by 0.2-year increments [93]. Figure 8 illustrates the changing prominence of the three components as a function of time. An about 11-year cycle in the prominence of the about 5-month component is highlighted. Second, a nonlinear model can be fitted in each interval and the period displayed as a function of time with its 95% confidence interval. This approach was



used to illustrate the great variability in the length of the about 11-year solar activity cycle, Figure 9 [94].

Discussion and conclusion

There are, of course, other important tools for the analysis of time series [95-103]. Most of them, however, require that the data be equidistant. This overview focused specifically on the use of the cosinor and its different extensions. The method is fairly simple and its results lead to meaningful interpretations. Despite its several shortcomings related primarily to the difficulty of satisfying all assumptions underlying the use of regression techniques, its wide-ranging applications have played an important role in the development of chronobiology as a quantitative scientific discipline. Used with caution, results based on a combination of exploratory analyses with the different cosinor routines and other conventional statistical tests, progress has also been made in the



field of chronomics which aims at mapping broad time structures from the high-frequency brain waves to the multi-decadal cycles characterizing space-terrestrial weather influencing human physiology and pathology [3,104].

Despite its simplicity, some reluctance remains for some investigators to use the cosinor for estimating rhythm parameters or for considering more than a single test time in designing experiments. Too many studies still rely on testing only at a fixed time of day (to control for the circadian variation) or at most at two times 12 hours apart, ignoring the possibility that the two selected timepoints may be at the midline crossings rather than at the peak and trough where differences are maximal. As discussed elsewhere, such practice can be misleading in missing an existing difference or even in finding a difference in mean when none exists [15-17]. Computing day-night differences in lieu of an amplitude and acrophase is also widely done to interpret ambulatory blood pressure monitoring records in terms of “dipping” [105], despite the documentation in several outcome studies of the superiority of a chronobiological approach [106,107].

To some extent, this status quo may be accounted for by the lack of dissemination of computer software offering chronobiologists tools for time series analysis applicable to non-equidistant data. This situation is slowly changing, however. Personal computers have become more powerful and statistical packages have become more readily available for relatively easy use by investigators not necessarily versed in all statistical details underlying the programs included in the software packages. While professional statistical software packages remain somewhat expensive for individual users, several open-source packages (such as Octave and R) offer an attractive alternative, notably since

some are platform-independent, running on PCs, Macs or Linux systems [108]. Programmers have taken advantage of the tools available in these packages to write code to perform analytical tasks of interest to chronobiologists. Perhaps the most comprehensive package is that developed by Oehlert and Bingham [109], offering a large array of procedures that can be applied by writing minimal coding instructions to call the different macros. Selected programs used in chronobiology have long been offered on the website of Refinetti [110], with clear instructions on how to run the programs. While not open-source, the Expert Soft Technologie website [111] also offers an array of cosinor-based and other procedures, including techniques for the study of non-stationary signals. These programs have been used in the study of shift-workers [112].

In summary, selected methods for the study of biologic time series have been reviewed and their relative merits have been discussed in the light of underlying assumptions. Some illustrative applications have also been mentioned. When the choice of a model is justified, and it is functional and explicative, quantitative methods of data analysis are extremely valuable to specify the model and obtain estimates of its parameters. Even when underlying assumptions are not fully met, point estimates of the parameters can be very useful. More caution is needed, however, in deciding whether P-values and confidence intervals are trustworthy, since violation of underlying assumptions tends to yield results that are too liberal. Once this limitation is taken into consideration, data analysis methods as described herein constitute extremely valuable tools for research in chronobiology and chronomics.

Competing interests

The author declares that she has no competing interests.

Acknowledgment

Dedicated to the memory of Franz Halberg who developed the cosinor method.

Support

Halberg Chronobiology Fund, University of Minnesota Supercomputing Institute.

Received: 27 January 2014 Accepted: 18 March 2014

Published: 11 April 2014

References

1. Halberg F: **Temporal coordination of physiologic function.** *Cold Spr Harb Symp quant Biol* 1960, **25**:289–310.
2. Halberg F, Tong YL, Johnson EA: **Circadian System Phase – An Aspect of Temporal Morphology; Procedures and Illustrative Examples.** In *Proc. International Congress of Anatomists. The Cellular Aspects of Biorhythms, Symposium on Biorhythms.* Edited by Mayersbach HV. New York: Springer-Verlag; 1967:20–48.
3. Halberg F, Cornelissen G, Katinas GS, Hillman D, Otsuka K, Watanabe Y, Wu J, Halberg F, Halberg J, Sampson M, Schwartzkopff O, Halberg E: **Many rhythms are control information for whatever we do: an autobiography.** *Folia anthropologica* 2012, **12**:5–134.
4. Halberg F, Visscher MB: **Regular diurnal physiological variation in eosinophil levels in five stocks of mice.** *Proc Soc exp Biol (N.Y.)* 1950, **75**:846–847.
5. Halberg F, Visscher MB, Bittner JJ: **Relation of visual factors to eosinophil rhythm in mice.** *Amer J Physiol* 1954, **179**:229–235.
6. Halberg F, Visscher MB, Bittner JJ: **Eosinophil rhythm in mice: Range of occurrence; effects of illumination, feeding and adrenalectomy.** *Amer J Physiol* 1953, **174**:109–122.
7. Haus E, Cornelissen G, Halberg F: **Introduction to chronobiology.** In *Chronobiology: Principles and Applications to Shifts in Schedules.* Edited by Scheving LE, Halberg F. Alphen aan den Rijn, The Netherlands: Sijthoff and Noordhoff; 1980:1–32.
8. Halberg F, Cornelissen G, Gumarova L, Halberg F, Ulmer W, Hillman D, Siegelova J, Watanabe Y, Hong S, Otsuka K, Wu J, Lee JY, Schwartzkopff O, Wendt H: **Integrated and as-one-goes analyzed physical, biospheric and noetic monitoring: Preventing personal disasters by self-surveillance may help understand natural cataclysms: a chronosphere (chrono-noosphere).** London: SWB International Publishing House. in press.
9. Halberg E, Halberg F: **Chronobiologic study design in everyday life, clinic and laboratory.** *Chronobiologia* 1980, **7**:95–120.
10. De Prins J, Cornelissen G, Malbecq W: **Statistical procedures in chronobiology and chronopharmacology.** In *Annual Review of Chronopharmacology, Volume 2.* Edited by Reinberg A, Smolensky M, Labrecque G. Oxford: Pergamon Press; 1986:27–141.

11. De Prins J, Lechien JP: **Echantillonnage et analyse spectrale.** *Cahiers du Centre d'Etudes de Recherches Opérationnelle* 1977, **19**:357–364.
12. Ostle B: *Statistics in research.* Iowa: Iowa State University Press; 1963.
13. Cochran WG, Cox GM: *Experimental designs.* New York: J Wiley & Sons, Inc. second edition; 1957.
14. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG: **Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design.** *Brit J Cancer* 1976, **34**:585–612.
15. Bingham C, Cornelissen G, Halberg F: **Power of "Phase 0" chronobiologic trials at different signal-to-noise ratios and sample sizes.** *Chronobiologia* 1993, **20**:179–190.
16. Halberg F, Bingham C, Cornelissen G: **Clinical trials: the larger the better?** *Chronobiologia* 1993, **20**:193–212.
17. Cornelissen G, Halberg F: **Treatment with open eyes: markers-guided chronotheranostics.** In *Chronopharmacetics: Science and Technology for Biological Rhythm-Guided Therapy and Prevention of Diseases.* Edited by Youan BC. Hoboken, NJ: Wiley; 2009:257–323.
18. Halberg F, Cornelissen G, Sothorn RB, Wallach LA, Halberg E, Ahlgren A, Kuzel M, Radke A, Barbosa J, Goetz F, Buckley J, Mandel J, Schuman L, Haus E, Lakatua D, Sackett L, Berg H, Wendt HW, Kawasaki T, Ueno M, Uezono K, Matsuoka M, Omae T, Tarquini B, Cagnoni M, Garcia Sainz M, Perez Vega E, Wilson D, Griffiths K, Donati L, Tatti P, Vasta M, Locatelli I, Camagna A, Lauro R, Tritsch G, Wetterberg L: **International geographic studies of oncological interest on chronobiological variables.** In *Neoplasms—Comparative Pathology of Growth in Animals, Plants and Man.* Edited by Kaiser H. Baltimore: Williams and Wilkins; 1981:553–596.
19. Halberg F, Guillaume F, Sanchez de la Peña S, Cavallini M, Cornelissen G: **Cephalo-adrenal interactions in the broader context of pragmatic and theoretical rhythm models.** *Chronobiologia* 1986, **13**:137–154.
20. Jozsa R, Halberg F, Cornelissen G, Zeman M, Kazsaki J, Csernus V, Katinas GS, Wendt HW, Schwartzkopff O, Stebelova K, Dulkova K, Chibisov SM, Engebretson M, Pan W, Bubenik GA, Nagy G, Herold M, Hardebrand R, Hüther G, Pöggeler B, Tarquini R, Perfetto F, Salti R, Olah A, Csokas N, Delmore P, Otsuka K, Bakken EE, Allen J, Amory-Mazaudier C: **Chronomics, neuroendocrine feedsideways and the recording and consulting of nowcasts forecasts of geomagnetics.** *Biomed Pharmacother* 2005, **59**(Suppl 1):S24–S30.
21. Halberg F, Carandente F, Cornelissen G, Katinas GS: **Glossary of chronobiology.** *Chronobiologia* 1977, **4**(1):189.
22. Watanabe Y, Halberg F, Otsuka K, Cornelissen G: **Toward a personalized chronotherapy of high blood pressure and a circadian overswing.** *Clin Exp Hypertens* 2013, **35**:257–266.
23. Sokal RR, Rohlf FJ: *Biometry.* secondth edition. Freeman & Co.; 1981.
24. Halberg F, Visscher MB, Flink EB, Berge K, Bock F: **Diurnal rhythmic changes in blood eosinophil levels in health and in certain diseases.** *J Lancet* 1951, **71**:312–319.
25. Halberg F, Cornelissen G, Katinas G, Syutkina EV, Sothorn RB, Zaslavskaya R, Halberg F, Watanabe Y, Schwartzkopff O, Otsuka K, Tarquini R, Perfetto P, Siegelova J: **Transdisciplinary unifying implications of circadian findings in the 1950s.** *J Circadian Rhythms* 2003, **1**(2):61.
26. Cornelissen G, Halberg F: **Chronomedicine.** In *Encyclopedia of Biostatistics.* 2nd edition. Edited by Armitage P, Colton T. Chichester, UK: John Wiley & Sons Ltd; 2005:796–812.
27. Powell EW, Pasley JN, Scheving LE, Halberg F: **Amplitude-reduction and acrophase-advance of circadian mitotic rhythm in corneal epithelium of mice with bilaterally lesioned suprachiasmatic nuclei.** *Anat Rec* 1980, **197**:277–281.
28. Cornelissen G, Halberg F: *Introduction to Chronobiology.* Minneapolis, MN: Medtronic Chronobiology Seminar #7; 1994:52. Library of Congress Catalog Card #94-060580.
29. Hillman D, Fernández JR, Cornelissen G, Berry DA, Halberg J, Halberg F: **Bounded limits and statistical inference in chronobiometry.** *Prog Clin Biol Res* 1990, **341B**:417–428.
30. Halberg F: **Chronobiology.** *Annu Rev Physiol* 1969, **31**:675–725.
31. Halberg F, Johnson EA, Nelson W, Runge W, Sothorn R: **Authorhythmometry – procedures for physiologic self-measurements and their analysis.** *Physiol Tchr* 1972, **1**:1–11.
32. Halberg F: **Chronobiology: methodological problems.** *Acta med rom* 1980, **18**:399–440.
33. Halberg F, Panofsky H: **I. Thermo-variance spectra; method and clinical illustrations.** *Exp Med Surg* 1961, **19**:284–309.
34. Panofsky H, Halberg F: **II. Thermo-variance spectra; simplified computational example and other methodology.** *Exp Med Surg* 1961, **19**:323–338.
35. Marquardt DW, Acuff SK: **Direct quadratic spectrum estimation from unequally spaced data.** In *Applied time series analysis.* Edited by Anderson OD, Perryman MR. North-Holland: North-Holland Publ. Co; 1982:199–227.
36. Brillinger D, Fienberg S, Gani J, Hartigan J, Krickeberg K (Eds): *Time Series Analysis of Irregularly Observed Data, vol. 25.* New York, Berlin, Heidelberg, Tokyo: Springer-Verlag; 1984.
37. Brillinger D, Fienberg S, Gani J, Hartigan J, Krickeberg K (Eds): *Robust and Nonlinear Time Series Analysis, vol. 26.* New York, Berlin, Heidelberg, Tokyo: Springer-Verlag; 1984.
38. Bloomfield P: *Fourier Analysis of Time Series: An Introduction.* New York: Wiley; 1976.
39. Reinberg A, Halberg F: **Circadian chronopharmacology.** *Annu Rev Pharmacol* 1971, **2**:455–492.
40. Haus E, Halberg F, Kühl JFW, Lakatua DJ: **Chronopharmacology in animals.** *Chronobiologia* 1974, **1**(Suppl. 1):122–156.
41. Halberg F, Haus E, Nelson W, Sothorn R: **Chronopharmacology, chronodietetics and eventually clinical chronotherapy.** *Nova Acta leopold* 1977, **46**:307–366.
42. Reinberg A, Halberg F: *Chronopharmacology.* Oxford/New York: Pergamon Press; 1979.
43. Bingham C, Arbogast B, Cornelissen Guillaume G, Lee JK, Halberg F: **Inferential statistical methods for estimating and comparing cosinor parameters.** *Chronobiologia* 1982, **9**:397–439.
44. Weisberg S: *Applied Linear Regression.* New York: J Wiley & Sons; 1980.
45. Bliss CI: *Statistics in Biology, Volume 1.* New York: McGraw Hill; 1967.
46. Shapiro SS, Wilk MB: **An analysis of variance test for normality (complete samples).** *Biometrika* 1965, **52**:591–611.
47. Draper NR, Smith H: *Applied Regression Analysis, Second Edition.* New York: Wiley & Sons; 1981.
48. Anderson RL: **The problem of autocorrelation in regression analysis.** *J Am Stat Assoc* 1954, **49**:113–129.
49. Sachs L: *Applied Statistics: A Handbook of Techniques.* New York: Springer-Verlag; 1982.

50. Conover WJ: *Practical Nonparametric Statistics, Second Edition*. New York: Wiley & Sons; 1980.
51. Cornelissen Guillaume G, Halberg F, Fanning R, Kanabrocki EL, Scheving LE, Pauly JE, Redmond DP, Carandente F: **Analysis of circadian rhythms in human rectal temperature and motor activity in dense and short series with correlated residuals**. In *Biomedical Thermology*. Edited by Gautherie M, Albert E. New York: Alan R. Liss; 1982:167–184.
52. Dunstan FDJ, Barham S, Kemp KW, Nix ABJ, Rowlands RJ, Wilson DW, Phillips MJ, Griffiths K: **A feasibility study for early detection of breast cancer using breast skin temperature rhythms**. *Statistician* 1982, **31**:37–52.
53. Halberg F, Cornelissen G, Sothorn RB, Hillman D, Watanabe Y, Haus E, Schwartzkopff O, Best WR: **Decadal cycles in the human cardiovascular system**. *World Heart J* 2012, **4**:263–287.
54. Halberg F, Powell D, Otsuka K, Watanabe Y, Beaty LA, Rosch P, Czaplicki J, Hillman D, Schwartzkopff O, Cornelissen G: **Diagnosing vascular variability anomalies, not only MESOR-hypertension**. *Am J Physiol Heart Circ Physiol* 2013, **305**:H279–H294.
55. Cornelissen G, Otsuka K, Halberg F: **Remove and Replace for a Scrutiny of Space Weather and Human Affairs**. In *Int. Conf., Space Weather Effects in Humans: In Space and on Earth*. Edited by Grigoriev AI, Zeleny LM. Moscow, Russia: Space Res Inst; 2013:508–538.
56. Cornelissen G, Watanabe Y, Otsuka K, Halberg F: **Influences of Space and Terrestrial Weather on Human Physiology and Pathology**. In *Bioelectromagnetic and Subtle Energy Medicine*. Edited by Rosch P. Russia: CRC Press. in press.
57. Narula SC, Saldiva PH, Andre CD, Elian SN, Ferreira AF, Capelozzi V: **The minimum sum of absolute errors regression: a robust alternative to the least squares regression**. *Stat Med* 1999, **18**:1401–1417.
58. Ahdesmaki M, Lahdesmaki H, Gracey A, Shmulevich L, Yli-Harja O: **Robust regression for periodicity detection in non-uniformly sampled time-course gene expression data**. *BMC Bioinforma* 2007, **8**:233.
59. Blume JD, Su L, Olveda RM, McGarvey ST: **Statistical evidence for GLM regression parameters: a robust likelihood approach**. *Stat Med* 2007, **26**:2919–2936.
60. Pires AM, Rodrigues IM: **Multiple linear regression with some correlated errors: classical and robust methods**. *Stat Med* 2007, **26**:2901–2918.
61. Anderson R: *Modern Methods for Robust Regression*. Sage Publications, Inc.: University of Toronto; 2008.
62. Cornelissen G, Otsuka K, Halberg F: **Blood pressure and heart rate chronome mapping: a complement to the human genome initiative**. In *Chronocardiology and Chronomedicine: Humans in Time and Cosmos*. Edited by Otsuka K, Cornelissen G, Halberg F. Tokyo: Life Science Publishing; 1993:16–48.
63. Cornelissen G, Halberg F, Bakken EE, Singh RB, Otsuka K, Tomlinson B, Delcourt A, Toussaint G, Bathina S, Schwartzkopff O, Wang ZR, Tarquini R, Peretto F, Pantaleoni GC, Jozsa R, Delmore PA, Nolley E: **100 or 30 years after Janeway or Barter, Healthwatch helps avoid “flying blind”**. *Biomed Pharmacother* 2004, **58**(Suppl 1):S69–S86.
64. Halberg F, Cornelissen G, Otsuka K, Siegelova J, Fiser B, Dusek J, Homolka P, Sanchez DelaPena S, Singh RB, BIOCOS project: **Extended consensus on means and need to detect vascular variability disorders (VVDs) and vascular variability syndromes (VVSs)**. *World Heart J* 2010, **2**:279–305.
65. Cornelissen G: **Instrumentation and data analysis methods needed for blood pressure monitoring in chronobiology**. In *Chronobiotechnology and Chronobiological Engineering*. Edited by Scheving LE, Halberg F, Ehret CF. Dordrecht, The Netherlands: Martinus Nijhoff; 1987:241–261.
66. Halberg E, Halberg F, Shankaraiah K: **Plexo-serial linear-nonlinear rhythmometry of blood pressure, pulse and motor activity by a couple in their sixties**. *Chronobiologia* 1981, **8**:351–366.
67. Cornelissen G, Halberg F, Otsuka K, Singh RB: **Separate cardiovascular disease risks: circadian hyper-amplitude-tension (CHAT) and an elevated pulse pressure**. *World Heart J* 2008, **1**:223–232.
68. Otsuka K, Cornelissen G, Halberg F, Oehlert G: **Excessive circadian amplitude of blood pressure increases risk of ischemic stroke and nephropathy**. *J Med Eng Technol* 1997, **21**:23–30.
69. Chen CH, Cornelissen G, Halberg F, Fiser B: **Left ventricular mass index as “outcome” related to circadian blood pressure characteristics**. *Scr Med (Brno)* 1998, **71**:183–189.
70. Müller-Bohn T, Cornelissen G, Halhuber M, Schwartzkopff O, Halberg F: **CHAT und Schlaganfall**. *Deutsche Apotheker Zeitung* 2002, **142**:366–370.
71. Schaffer E, Cornelissen G, Rhodus N, Halhuber M, Watanabe Y, Halberg F: **Blood pressure outcomes of dental patients screened chronobiologically: a seven-year follow-up**. *JADA* 2001, **132**:891–899.
72. Cornelissen G, Siegelova J, Watanabe Y, Otsuka K, Halberg F: **Chronobiologically-interpreted ABPM reveals another vascular variability anomaly (VVA): excessive pulse pressure product (PPP) – updated conference report**. *World Heart J* 2012, **4**:237–245.
73. Cornelissen G, Halberg F, Burioka N, Peretto F, Tarquini R, Bakken EE: **Do plasma melatonin concentrations decline with age?** *Am J Med* 2000, **109**:343–344.
74. Bonferroni CE: **Teoria statistica delle classi e calcolo delle probabilità**. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 1936, **8**:3–62.
75. Nintcheu-Fata S, Cornelissen G, Katinas G, Halberg F, Fiser B, Siegelova J, Masek M, Dusek J: **Software for contour maps of moving least-squares spectra**. *Scr Med (Brno)* 2003, **76**:279–283.
76. Otsuka K, Cornelissen G, Halberg F: **Age, gender and fractal scaling in heart rate variability**. *Clin Sci* 1997, **93**:299–308.
77. Nuttall AH: **Some Windows with Very Good Sidelobe Behavior**. *IEEE Trans Acoustics Speech Signal Process* 1981, **29**:84–91.
78. Bendat JS, Piersol AG: *Random Data-Analysis and Measurement Procedures*. New York: Wiley Interscience; 1971.
79. Marquardt DW: **An algorithm for least squares estimation of nonlinear parameters**. *J Soc Indust Appl Math* 1963, **11**:431–441.
80. Marquardt DW: *Least Squares Estimation of Nonlinear Parameters*, IBM share library distribution No. 309401; 1966.
81. Bingham C, Cornelissen G, Halberg E, Halberg F: **Testing period for single cosinor: extent of human 24-h cardiovascular “synchronization” on ordinary routine**. *Chronobiologia* 1984, **11**:263–274.
82. Czaplicki J, Cornelissen G, Halberg F: **GOSA, a simulated annealing-based program for global optimization of nonlinear problems, also reveals transyears**. *J Applied Biomedicine* 2006, **4**:87–93.

83. Refinetti R, Cornelissen G, Halberg F: **Procedures for numerical analysis of circadian rhythms.** *Biol Rhythm Res* 2007, **38**:275–325.
84. Cornelissen G, Grambsch P, Halberg F: **Editorial.** *World Heart J* 2011, **3**:123–134.
85. Daubechies I: **The wavelet transform, time-frequency localization and signal analysis.** *IEEE Trans Inf Theory* 1990, **36**:961–1005.
86. Sello S, Halberg F, Cornelissen G: **Human babies: a slow-to-read, sensitive population magnetometer, also read by wavelets.** In *Noninvasive Methods in Cardiology*. Edited by Halberg F, Kenner T, Fiser B, Siegelova J. Brno, Czech Republic: Faculty of Medicine, Masaryk University; 2011:123–140.
87. Czaplicki J, Cornelissen G, Halberg F: **Spectrograms recognize multiple circadian periods by blood pressure and heart rate tensiometry.** *World Heart J* 2012, **4**:9–22.
88. Watanabe Y, Nintcheu-Fata S, Katinas G, Cornelissen G, Otsuka K, Hellbrügge T, Schwartzkopff O, Bakken E, Halberg F: **Methodology: partial moving spectra of postnatal heart rate chronome.** *Neuroendocrinol Lett* 2003, **24**(Suppl 1):139–144.
89. Levine H, Cornelissen G, Halberg F, Bingham C: **Self-measurement, automatic rhythmometry, transmeridian flights and aging.** In *Chronobiology: Principles and Applications to Shifts in Schedules*. Edited by Scheving LE, Halberg F. Alphen aan den Rijn, The Netherlands: Sijthoff and Noordhoff; 1980:371–392.
90. Sanchez de la Peña S, Halberg F, Galvagno A, Montalbini M, Follini S, Wu J, Degioanni J, Kutyna F, Hillman DC, Kawabata Y, Cornelissen G: **Circadian and Circaseptan (about-7-day) Free-Running Physiologic Rhythms of a Woman in Social Isolation.** In *Proc. 2nd Ann. IEEE Symp. on Computer-Based Medical Systems*. Washington DC: Computer Society Press; 1989:273–278.
91. Halberg F, Cornelissen G, Hillman D, Ilyia E, Cegielski N, El-Khoury M, Finley J, Thomas F, Brandes V, Kino T, Papadopoulou A, Chrousos GP, Costella JF, Mikulecky M: **Multiple circadian periods in a lady with recurring episodes of adynamic depression: case report.** In *Noninvasive Methods in Cardiology*. Edited by Halberg F, Kenner T, Fiser B, Siegelova J. Brno, Czech Republic: Faculty of Medicine, Masaryk University; 2011:45–67.
92. Rieger A, Share GH, Forrest DJ, Kanbach G, Reppin C, Chupp EL: **A 154-day periodicity in the occurrence of hard solar flares?** *Nature* 1984, **312**:623–625.
93. Cornelissen G, Halberg F, Sothorn RB, Hillman DC, Siegelova J: **Blood pressure, heart rate and melatonin cycles synchronization with the season, earth magnetism and solar flares.** *Scr Med* 2010, **83**:16–32.
94. Cornelissen G, Halberg F, Gheonjian L, Paatashvili T, Faraone P, Watanabe Y, Otsuka K, Sothorn RB, Breus T, Baevsky R, Engebretson M, Schröder W: **Schwabe's ~10.5- and Hale's ~21-Year Cycles in Human Pathology and Physiology.** In *Long- and Short-Term Variability in Sun's History and Global Change*. Edited by Schröder W. Bremen: Science Edition; 2000:79–88.
95. Box GEP, Jenkins GM: *Time Series Analysis: Forecasting and Control*, Holden-Day; 1970.
96. Mills TC: *Time Series Techniques for Economists*. Cambridge: Cambridge University Press; 1990.
97. Percival DB, Walden AT: *Spectral Analysis for Physical Applications*. Cambridge: Cambridge University Press; 1993.
98. Anderson N: **On the calculation of filter coefficients for maximum entropy spectral analysis.** *Geophysics* 1974, **39**:69–72.
99. Burg JP: *Maximum Entropy Spectral Analysis*. Oklahoma City: Paper presented at the 37th Annual Int. Meeting Soc. Of Explo. Geophy; 1967.
100. Enright JT: **The search for rhythmicity in biological time-series.** *J Theor Biol* 1965, **8**:426–468.
101. Lomb NR: **Least-squares frequency analysis of unequally spaced data.** *Astrophysics Space Sci* 1976, **39**:447–462.
102. Scargle JD: **Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data.** *Astrophysical J* 1982, **263**:835–853.
103. Van Dongen HPA, Olofse E, Van Hartevelt JH, Kruyt EW: **Searching for Biological Rhythms: Peak Detection in the Periodogram of Unequally Spaced Data.** *J Biol Rhythm* 1999, **14**:617–620.
104. Halberg F, Cornelissen G, Katinas G, Hillman D, Schwartzkopff O: **Season's Appreciations 2000: Chronomics complement, among many other fields, genomics and proteomics.** *Neuroendocrinol Lett* 2001, **22**:53–73.
105. Verdecchia P, Schillaci G, Guerrieri M, Gatteschi C, Benemio G, Boldrini F, Porcellati C: **Circadian blood pressure changes and left ventricular hypertrophy in essential hypertension.** *Circulation* 1990, **81**:528–536.
106. Al-Abdulgader AA, Cornelissen G, Halberg F: **Vascular Variability Disorders in the Middle East: case reports.** *World Heart J* 2010, **2**:261–277.
107. Cornelissen G, Halberg F, Otsuka K, Singh RB, Chen CH: **Chronobiology predicts actual and proxy outcomes when dipping fails.** *Hypertension* 2007, **49**:237–239.
108. Lee-Gierke C: *Analysis of Rhythms Using R: Introducing Chronomics Analysis Toolkit (CAT)*. Minnesota: A Master's Thesis Submitted to the Graduate Faculty of the University of Minnesota; 2013.
109. Oehlert GW, Bingham C: *MacAnova. A Program for Statistical Analysis and Matrix Algebra*. <http://www.stat.umn.edu/macanova/>.
110. Refinetti R: *Circadian Software*. <http://www.circadian.org/softwar.html>.
111. Gouthiere L, Mauvieux B: *Etapes essentielles dans l'analyse des rythmes: qualité des données expérimentales, recherche de périodes par analyses spectrales de principes divers, modélisation*. <http://www.euroestech.net/resources/eedadr.pdf>.
112. Gouthiere L, Mauvieux B, Davenne D, Waterhouse J: **Complementary methodology in the analysis of rhythmic data, using examples from a complex situation, the rhythmicity of temperature in night shift workers.** *Biol Rhythm Res* 2005, **36**:177–193.

doi:10.1186/1742-4682-11-16

Cite this article as: Cornelissen: Cosinor-based rhythmometry. *Theoretical Biology and Medical Modelling* 2014 **11**:16.