Quantification of irregular circadian cycles using time-series methods

Cite as: AIP Conference Proceedings 2090, 050009 (2019); https://doi.org/10.1063/1.5095924
Published Online: 10 April 2019

Lorena Marisol García-Iglesias, Ana Leonor Rivera, Juan Claudio Toledo-Roy, and Ruben Fossion

ARTICLES YOU MAY BE INTERESTED IN

Heart rate variability of homeless young men
AIP Conference Proceedings 2090, 050004 (2019); https://doi.org/10.1063/1.5095919

On the role of continuous physiological monitoring and time-series analysis in medical prognosis
AIP Conference Proceedings 2090, 050007 (2019); https://doi.org/10.1063/1.5095922

Time-series based evaluation of physiological alterations in age-associated frailty during walking at self-selected velocity
AIP Conference Proceedings 2090, 050002 (2019); https://doi.org/10.1063/1.5095917
Quantification of irregular circadian cycles using time-series methods

Lorena Marisol García-Iglesias\textsuperscript{1,2,a)}, Ana Leonor Rivera\textsuperscript{1,3}, Juan Claudio Toledo-Roy\textsuperscript{1,3} and Ruben Fossion\textsuperscript{1,3,b)}

\begin{itemize}
  \item \textsuperscript{1}Instituto de Ciencias Nucleares, Universidad Autónoma de México, 04510 Mexico City, Mexico.
  \item \textsuperscript{2}Universidad Autónoma del Estado de Morelos, Avenida Universidad 1001, Chamilpa, 62209 Cuernavaca, Morelos.
  \item \textsuperscript{3}Centro de Ciencias de la Complejidad (C3), Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico.
\end{itemize}

\textsuperscript{a)}lorena.garcia@correo.nucleares.unam.mx
\textsuperscript{b)}ruben.fossion@nucleares.mx

\textbf{Abstract.} The original focus of the field of chronobiology was to demonstrate that most – if not all – biological parameters are modulated in a regular way by the periodic alternation of day and night, called circadian cycles. Recently, it has become clear that adverse conditions such as ageing and/or disease induce a reduced or more irregular modulation, such that it has become of interest to quantify these irregularities as a proxy to understand underlying pathology. In the present contribution, we explore how recently developed time-series decomposition techniques such as SSA, EMD, EEMD and CEEMDAN may be applied to chronobiology to describe irregularities in circadian cycles.

\section*{INTRODUCTION}

Circadian cycles are the regular modulation of biological, physiological, behavioral or mental variables by the periodic day-night cycle. The initial objective of the field of chronobiology was to demonstrate that most – if not all – lifeforms show these approx. 24h cycles \cite{1}. More recently, it has become clear that adverse conditions such as aging and/or disease are responsible for a reduced or more irregular modulation, such that the focus of the field has shifted to a quantification of these irregularities to track the underlying pathologies, e.g., age-associated frailty \cite{2} or insomnia \cite{3}.

The most basic tool of chronobiology is cosinor analysis, where a simple periodic function, such as a sine or a cosine function or a superposition of several of these functions, is fitted to the data by linear regression \cite{4}. Cosinor analysis is \textit{model-based} in the sense that it is the user who determines how many different frequencies to include to describe the data. Whereas cosinor analysis may offer the most straightforward description of periodicities in the data, it is not the most adequate method to describe irregularities in the circadian cycle. Wavelet analysis, either continuous (CWT) or discrete (DWT), has been used as an alternative \cite{5,6}. Here, a user-defined filter (mother wavelet) is applied to the data at different scales to assess how a specific mode such as the circadian cycle evolves over time. One of the drawbacks of wavelet analysis is that many different motherwavelets are available, that it is not always a priori clear what wavelet type is the most adequate to describe a particular variable, and that the quality of the analysis may crucially depend on the specific motherwavelet chosen \cite{7}.

Independently from the field of chronobiology, another field of time-series analysis has evolved with applications in many different areas of knowledge such as climate studies, medicine and economics. Time-series analysis is not necessarily most interested in one specific dominant mode, and often allow to decompose a time series as a sum of separate time-series components of different scales. Fourier spectral analysis is perhaps the best well known decomposition technique where a time-series is decomposed exactly as a sum of sines and cosines of different frequencies \cite{7}. A drawback of Fourier analysis is that the data cannot always be described using such a predefined set of periodic
basis functions, such as in the case of a dominant trend or a time series where the frequency content changes in time such as a chirp signal, and the result of the analysis may be contaminated with artefacts. Other, more data-driven techniques have been developed where the basis functions are generated by the data themselves, such as nonlinear mode decomposition (NMD) [8], singular spectrum analysis (SSA) [9, 10] and empirical model decomposition (EMD) and its variants [11, 12, 13, 14, 15]. Curiously, with a few rare exceptions [3, 16, 17], these more recent methods have not been applied in chronobiology.

The purpose of the present contribution is to explore if and how these new time-series techniques can be applied to describe circadian cycles, and whether they perhaps offer an improved description, such that patient populations can be diagnosed more easily.

METHODS

In the present contribution, we use an actigraphy time series of 4 successive weeks. Actigraphy offers an objective measure of the level of physical activity using movement counts per sample interval, called "epochs", which was here arbitrarily chosen to be 1 min. In principle, resting and waking intervals can be distinguished as absence or low levels of activity vs. high levels of movement, respectively. The data was recorded with an Actigraph wGT3X-BT, worn at all times throughout day and night. The subject monitored was a healthy young female university student of 18yo, during 3 weeks of lectures and 1 week of holidays (Easter).

We will analyze the data using cosinor analysis, nonlinear mode decomposition (NMD) [8], singular spectrum analysis (SSA) [9, 10], empirical model decomposition (EMD) [11, 12], ensemble empirical mode decomposition (EEMD) [13, 15] and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [14]. A detailed description of the various methods is beyond the scope of this contribution and we refer to the literature. We will only describe the cosinor method because it allows to define the different parameters of the circadian cycle [4].

The cosinor approach is based on regression techniques and is also applicable to equidistant or non-equidistant time series \( x(n) \) of \( N \) discrete data points,

\[
x(n) = \{x_1, x_2, \ldots, x_N\}.
\]

Given a specific value for period \( T \), the procedure consists of fitting a continuous cosine function \( y(t) \) to time series \( x(n) \),

\[
y(t) = M + A \cos(2\pi t/T + \phi).
\]

When exposed to the normal day-and-night cycle, this period can be expected to be \( T \approx 24h \). Minimizing the summed square residual errors \( e_n^2 = (x_n - y_n)^2 \) for all data points \( n = 1, 2, \ldots, N \), allows to find values for the so-called circadian parameters: the rhythm-adjusted mean or mesor \( M \), the amplitude \( A \) and the phase \( \phi \). Here, \( \phi \) indicates the height of the cosine wave at the start of the monitoring; because each monitoring session can start at an arbitrary time of the day, the phase \( \phi \) does not give any physiological information on the monitored individual. Instead, a more interesting variable is the acrophase \( \phi_0 \) which can be defined as the time of day where the circadian cycle obtains its maximum, with respect to a fixed moment in time which is the same for all subjects, e.g. taking midnight as a reference, and which can be expressed as hours and minutes (hh:mm), or alternatively, as an angle (taking into account the relation \( 360^\circ = 24hrs \)), relative to this reference time.

We will compare the results of the different methods. An important parameter to evaluate the quality of the circadian cycle as determined by the various methods is the coefficient of determination \( R^2 \), which compares the variance of the residual errors \( e_n \) around the fitted model \( y \) to the variance of the time series \( x(n) \) around its average value \( \langle x \rangle \),

\[
R^2 = 1 - \frac{\text{Var}(e)}{\text{Var}(x)} = 1 - \frac{\sum_{n=1}^{N}(x_n - y_n)^2}{\sum_{n=1}^{N}(x_n - \langle x \rangle)^2},
\]

such that \( R^2 \) is a measure for the fraction of the variance of the time series that can be explained by the model \( y(t) \).
RESULTS

DWT analysis requires to determine the particular motherwavelet to be used. Previous studies indicated that Daubechies-4 is the most appropriate motherwavelet for the study of actigraphy data [6, 7] and this is the type of wavelet that is used in the following. EMD and its variants have “hidden” variables and these had to be adjusted by trial and error while performing the analyses. SSA has 1 obvious parameter which is the size of the phase space $L$ but the recommendation of the literature is to chose it as a multiple of the dominant periodicity of the data, such that here it was chosen as $L = 1440\text{min}=24\text{hrs}$ [3].

Using the various methods mentioned before, we extracted the trend or mesor of the time series, i.e., the mode which represents the average behaviour of the time series and which does not oscillate, and the circadian cycle which
was defined as the mode with an average period of 24h. The NMD method only extracts dominant modes of oscillation and no mesor, therefore the mesor was represented by the constant average value of the time series. The extracted modes are compared to the original actigraphy data in Figs. 1 and 2.

The circadian parameters of mesor $M$, amplitude $A$, period $T$ and acrophase $\phi_0$ can be derived from the mode of the circadian cycle. Cosinor analysis only derives average values for the circadian parameters for the whole duration of the data, whereas time-series based methods including DWT allow to calculate a value for the circadian parameters for every individual day. Therefore, time-series methods allow to study not only the average value of the circadian parameters but also their day-to-day variability [3]. Considering the distribution of the values of every circadian parameter for each individual day, this variability may be quantified using the moments of the distribution, such as the standard deviation (SD), coefficient of variation ($CV=SD/mean$), skewness (Skew) and kurtosis (Kurt), see Table 1.

It can be seen that SSA, DWT and EEMD describe the circadian cycle best (highest values of $R^2$). Results for the average values of the circadian parameters agree rather well between the different methods, see also Fig. 3. The least day-to-day variability for the circadian parameters as quantified by SD and CV is obtained for NDM, whereas EMD tends to maximize the day-to-day variability. An exception is the variability of mesor $M$ which for most methods is a rather constant mode but which results rather variable within the SSA approach.

Finally, we calculated the correlation between the circadian cycles as extracted for the different methods, using Pearson’s correlation coefficient $r$ which in principle can only be used for normally distributed data and Spearman’s rank correlation coefficient $\rho$ which is more generally valid, see Table 2. The correlation coefficients are always quite high, $r \approx \rho > 7.0$, which is not surprising because it is always the same circadian cycle that is described although by different methods. When comparing the individual correlation coefficients to the average correlation coefficient (either the global average over all methods, or a local average within the same method), we find indications for 3 groups of methods. On the one hand, circadian cycles as derived by cosinor and NMD correlate well between each other, on the

![Figure 2](image_url)

FIGURE 2. (Continuation of Fig. 1). The trend component (black dashed curve) and circadian cycle (black continuous curve) were extracted using the EMD, EEMD and CEEMDAN methods (from top to bottom).
FIGURE 3. Graphical representation of the results of Table 1 for the circadian parameters mesor $M$, amplitude $A$, acrophase $\phi_0$ and period $T$ with average value and day-to-day variability as represented by SD. In the case of period $T$, also the expected value of 24h=1440min is indicated (horizontal gridline).

other hand EMD and CEEMDAN also correlate well, and finally SSA, DWT and EEMD seem to be closely related between each other but do also correlate above average with the other methods.

DISCUSSION

Fig. 3 suggests that the average values of the circadian parameters mesor $M$, amplitude $A$, period $T$ and acrophase $\phi_0$ as derived by the different methods agree rather well and all average values lie within the error bars of the day-to-day variability of the other methods.

We are however also interested in the day-to-day variability itself as a new parameter to evaluate the health status of subjects and populations. We have previous results where differences in the variability of circadian parameters helps to distinguish between young healthy controls and young adults with acute insomnia [3]. We wish to understand how the different methods quantify the variability of the various circadian parameters. Circadian cycles as described by cosinor and NMD have no or smaller day-to-day variability. Correlation analysis indicates that the circadian cycles described by these methods are rather similar. These methods describe more “rigid” circadian cycles, which can possibly be explained by these models being more model-based. This is clear in the case of cosinor where a user-defined sine or cosine function is fitted to the data. NMD is based on CWT analysis using by construction the lognormal motherwavelet which may not necessarily be the most adequate wavelet type to describe every type of data, such that parts of the day-to-day variability of the data might be overlooked by this method. Circadian cycles as described by EMD and CEEMDAN are also rather similar, they tend to have a large day-to-day variability such that more “volatile” circadian cycles seem to be described. These methods are empirical, purely data-driven, without any model constraints, e.g., EMD suffers from mode-mixing, where the instantaneous frequency of a mode can vary drastically from one time moment to the next, and more elaborate versions of the method such as EEMD and CEEMDAN try to correct for these artefacts. Finally, SSA, DWT and EEMD seem to describe a more moderate day-to-day variability in between both
extremes of cosinor and NMD on the one hand, and EMD and CEEMDAN on the other hand. The modes as derived by SSA are normal modes and by construction have a very stable instantaneous frequency. The differences between methods in their quantification of the day-to-day variability of the different circadian parameters would appear to be larger than the differences in the average values, and it is not a priori clear what fraction of the variability is due to the method of quantification and what fraction is due to the data itself.

There are differences in the user-friendliness of the methods as well. The family of EMD methods has “hidden” parameters that can be adjusted to each specific type of data and results may depend on these parameters. No prescription exists as of how to chose these parameters. EMD decomposes time series in components starting from the higher frequencies. Therefore, one has to carry out the whole calculation to obtain the circadian cycle and the mesor which are the slowest modes, and these calculations can take up a lot of computer time. EMD and the other variants of the method are very susceptible to outliers, missing data, or very irregular behaviour of the data, because there is no underlying model to guide the method how to derive the circadian cycle. NMD on the other hand focusses on the oscillating dominant modes and no mesor.

According to the coefficient of determination $R^2$, SSA, DWT and EEMD describe the circadian cycle best,
see Table 1. We suspect that such improved description might increase the statistical significance when comparing different populations. This is a hypothesis we are planning to investigate in the future.

CONCLUSION

Chronobiology and time-series analysis appear to be 2 fields that have evolved independently from each other. Recent time-series decomposition techniques such as NMD, SSA, EMD and EEMD, CEEMDAN have only exceptionally been used to describe circadian cycles in experimental data. These techniques go beyond the quantification of the average values of circadian parameters as mesor, amplitude, period and acrophase, and allow to describe the day-to-day variability of these parameters as well, which could offer a new parameter to distinguish between different populations. We find that all methods agree rather well in their quantification of the average values of the circadian parameters but there less agreement in the quantification of their day-to-day variability. The SSA method is easy to implement, is fast in the execution of its algorithm, and maximizes the coefficient of determination of the fit to the data.

ACKNOWLEDGMENTS

We acknowledge financial support from the Dirección General de Asuntos del Personal Académico (DGAPA) of the Universidad Nacional Autónoma de México (UNAM) grants IN106215, IV100116 and IA105017, from the Consejo Nacional de Ciencia y Tecnología (CONACYT) grants Fronteras 2015-2-1093, Fronteras 2016-01-2277 and CB-2011-01-167441, and the Newton Advanced Fellowship awarded to R.F. by the Academy of Medical Sciences through the UK Government’s Newton Fund program. We are grateful to Markus Mueller for fruitful discussions.

REFERENCES


