

PAPER • OPEN ACCESS

A computationally efficient algorithm to obtain an accurate and interpretable model of the effect of circadian rhythm on resting heart rate

To cite this article: Davide Morelli *et al* 2019 *Physiol. Meas.* **40** 095001

View the [article online](#) for updates and enhancements.

You may also like

- [Identification of scalp EEG circadian variation using a novel correlation sum measure](#)
Ali Shahidi Zandi, Philippe Boudreau, Diane B Boivin et al.
- [Expression of circadian gens in different rat tissues is sensitive marker of in vivo silver nanoparticles action](#)
D O Minchenko, O P Yavorovsky, T O Zinchenko et al.
- [Calculating activation energies for temperature compensation in circadian rhythms](#)
C Bodenstein, I Heiland and S Schuster



FREE

The Breath Biopsy® Guide

Fourth edition

DOWNLOAD THE FREE E-BOOK

BREATH
BIOPSY

OPEN ACCESS



CrossMark

RECEIVED
14 May 2019REVISED
9 August 2019ACCEPTED FOR PUBLICATION
22 August 2019PUBLISHED
26 September 2019

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



PAPER

A computationally efficient algorithm to obtain an accurate and interpretable model of the effect of circadian rhythm on resting heart rate

Davide Morelli^{1,2} , Leonardo Bartoloni², Alessio Rossi³  and David A Clifton¹¹ Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, OX2 6DP, United Kingdom² Biobeats Group LTD, 3 Fitzhardinge Street, London, W1H 6EF, United Kingdom³ Computer Science Department, University of Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, ItalyE-mail: davide.morelli@eng.ox.ac.uk**Keywords:** heart rate, chronobiology, circadian rhythm, cosinor, resting heart rateSupplementary material for this article is available [online](#)

Abstract

Objective: Wrist-worn wearable devices equipped with heart rate (HR) sensors have become increasingly popular. The ability to correctly interpret the collected data is fundamental to analyse user's well-being and perform early detection of abnormal physiological data. Circadian rhythm is a strong factor of variability in HR, yet few models attempt to accurately model its effect on HR. **Approach:** In this paper we present a mathematical derivation of the single-component cosinor model with multiple components that fits user data to a predetermined arbitrary function (the expected shape of the circadian effect on resting HR (RHR)), thus permitting us to predict the user's circadian rhythm component (i.e. MESOR, Acrophase and Amplitude) with a high accuracy. **Main results:** We show that our model improves the accuracy of HR prediction compared to the single component cosinor model (10% lower RMSE), while retaining the readability of the fitted model of the single component cosinor. We also show that the model parameters can be used to detect sleep disruption in a qualitative experiment. The model is computationally cheap, depending linearly on the size of the data. The computation of the model does not need the full dataset, but only two surrogates, where the data is accumulated. This implies that the model can be implemented in a streaming approach, with important consequences for security and privacy of the data, that never leaves the user devices. **Significance:** The multiple component model provided in this paper can be used to approximate a user's RHR with higher accuracy than single component model, providing traditional parameters easy to interpret (i.e. the same produced by the single component cosinor model). The model we developed goes beyond fitting circadian activity on RHR, and it can be used to fit arbitrary periodic real valued time series, vectorial data, or complex data.

1. Introduction

In the last decade, wearable activity trackers have become more and more popular to monitor heart rate (HR). However, HR can not be easily interpreted without the proper context. A measured HR value of 70 bpm could be considered high for an athlete while asleep, or low for someone that does not usually train, after a long walk. Several factors are known to affect HR: circadian rhythm will make the HR higher during the day and lower during the night; during physical activity (e.g. a long walk) the muscles will require more blood, causing the heart to beat faster (Stirling *et al* 2008); a high temperature, consequence of an infection or a flu, will cause the heart to beat faster; the need for mental performances, e.g. a student during an exam, also causes the heart to beat faster (Castaldo *et al* 2015); age, sex, and BMI also change the expected heart activity of the user (Shaffer and Ginsberg 2017). To be able to interpret HR measures and understand their physiological meaning, is therefore necessary to build a model of the subject, and to estimate the resting HR (RHR) for that particular subject at that particular

time of the day. Once the expected RHR is calculated, the difference between the measures HR and the RHR can provide valuable insight about the user physiological health and well-being.

RHR is widely investigated in the literature because it is an independent predictor of cardiovascular and all-cause mortality in men and women (Kristal-Boneh *et al* 2000, Fox *et al* 2007, Larsson *et al* 2019, Li *et al* 2019, Sarpti *et al* 2019). However, these studies are focused on a single point measurements without evaluating the variation of RHR through the day that could provide more information about health status of the population. In Faria and Drummond (1982) detected a physiological response fluctuation during the day of RHR and resting temperature, demonstrating a relationship with physical activity tasks. All the studies previously cited evaluated the cardiovascular disease risk and all-cause mortality from RHR and not from HR because of RHR reflects the basal activation of the autonomic nervous system that is found to be a predictor of several pathologies. The variation of HR among days could be only due to the different physical activity performed in a specific day (Fossion *et al* 2018) and not from a real change in health status, that can be detected analyzing the daily fluctuation of RHR.

Physiological measures such as HR and RHR are continuous variables that are subject to the influence of external and internal stimuli, changing over a daily cycle of about 24 h with patterns that have been defined as circadian rhythms (Faria and Drummond 1982, Koukkari and Sothorn 2006). The time series obtained from wrist IoT wearable device are often noisy (i.e. variations in the biological system that are not part of the deterministic portion of the signal and that are derived from external errors such as instrument inaccuracy), short (to save battery) and sparse (i.e. unequal time intervals between observations, missing datapoints) (Fernández *et al* 2003). Despite these time series problems and the fact that the real signal is not a simple sinusoidal, HR circadian rhythms analysis is usually performed using the cosinor method (Cornelissen 2014) with a single component because the resulting parameters can be easily interpreted. In particular, the cosinor curve with single component provides three parameters that describe the circadian rhythm: (i) mid-line estimating statistic of rhythm (MESOR) is an estimation of central tendency of the distribution of values across the cycles of the circadian rhythm computed using a cosine function; (ii) *amplitude* is the difference between the peak and the mean value of a wave; (iii) *acrophase* is the time of the day at which the peak of the circadian rhythm occurs.

It was found that the shape of the HR as a function of the hour of the day significantly differs from a simple cosine: during the night the HR becomes slower for approximately 6 h, while during the day it rises for approximately 18 h, with three peaks corresponding to meals (Malik *et al* 1990, Nakagawa *et al* 1998). However, only a few studies in the literature propose a cosinor method with multiple-component to analyze periodic components in non-sinusoidal longitudinal time series (Fernández *et al* 2003, Cornelissen 2014). The result of the multiple-component cosinor analysis includes several parameters which are not easy to interpret in comparison to the three parameters provided by single-component cosinor. Despite the multiple-component model being found to be more suitable to approximate the signals waveform when it deviates from sinusoidal, the single-component cosinor model is still widely used to investigate physiological circadian rhythm because of its simplicity and readability (Cornelissen 2014).

Some models have been proposed to fit the data more accurately. Leise proposed the use of wavelets (Leise and Harrington 2011, Leise 2017). Fossion proposed the use of multiscale adaptive analysis (Fossion *et al* 2017). Those approaches improve the goodness of fit, but fitting parameters are difficult to interpret, compared to the single component cosinor. An additional limitation of those methods is that they are only concerned with the frequencies present in the data, without requiring the signal to follow a specified shape.

With the wide adoption of fitness devices, concerns about the security and privacy of the data collected by these devices have been raised (Aktypi *et al* 2017, Fereidooni *et al* 2017). Among the principles of *Privacy by Design* are *data minimisation*, which states that only the data strictly necessary to perform the intended activities should be collected, and *full functionality*, which states that privacy should not impair the functionalities of the product (Cavoukian *et al* 2009). However, the standard approach of fitness companies is to upload all the user data to their servers for processing, and performing all the data analysis on the cloud, requiring access to users raw HR data. In practice, *data minimisation* is usually sacrificed in favour of *full functionality*.

1.1. Contributions of this paper

In this paper we present a mathematical derivation of the single-component cosinor model with multiple components that fits user data onto an arbitrary function (i.e. the model which describes the circadian rhythm of the population), thus enabling the prediction of the user's circadian rhythm effect on RHR.

Unlike the cosinor model with multiple components provided by Cornelissen (2014), the model proposed in our study provides the same three easy-to-interpret parameters which can be derived from the cosinor model with single component (i.e. *MESOR*, *amplitude*, *acrophase*).

Our model allows to assign weights to every data point used for fitting the expected shape to the user data. This enables advanced analyses with arbitrary error models, like assigning less importance to data points collected while the user was moving (expecting motion artifacts, as explained in Morelli *et al* (2017)). Using weights also it makes possible to implement real-time streaming analysis of the data, performing a new analysis as new

data arrives, assigning less importance to older data. In this study we show how to implement such a real-time strategy assigning exponentially decaying weights to old data, obtaining a circadian model that continuously changes with new data. We show how the three fitted parameters (i.e. *MESOR*, *amplitude*, *acrophase*) can be used to interpret behavioural and physiological changes.

The model presented in this paper is computationally light and, for this reason, it is suitable to be executed on a wearable device. The complexity of the accumulation phase is linear on the number of data points, while the complexity resolution phase depends on the number of frequencies used and not on the number of data points.

To the best of our knowledge, this is the first study that provides an algorithm that fits the circadian rhythm by using a multiple component cosinor, forcing the data to fit a particular shape, i.e. the population average effect of RHR circadian rhythm, only leaving three free parameters (i.e. the same used by single one). Moreover, our algorithm allows to perform all the data sensitive calculations on the wearable device. As a consequence users raw HR data could be kept on the wearable device, while allowing companies to obtain the output of the analysis, implementing both the *full functionality* and the *data minimisation Privacy by Design* principles.

Predicting a user's RHR is an important step to interpret the measured HR data: the residual between measured HR and estimated RHR can then be correlated with physical activity, and inference of valuable insight on this user's well-being and health becomes possible.

1.2. Background

Short and sparse time series were historically analyzed by the single-component cosinor model. However, this model is not suitable for unevenly sampled, noisy time series (i.e. the one showing several outlier values), with missing data such as the data recorded by wearable HR trackers.

Cornelissen (2014) extend the single-component cosinor to a multiple-component model in order to analyze the time series in chronobiology. Instead of solving a system of three equations in three unknowns to define the parameters of the circadian rhythm curve (i.e. *MESOR*, *amplitude*, *acrophase*) there are $2c + 1$ parameters to estimate (i.e. c refers to the number of harmonics selected to describe the circadian). This approach better approximates the signals waveform of the circadian rhythm compared to single-component cosinor model thanks to the higher number of cosinor components used to fit the model on the circadian rhythm curve. For example, a 2-component cosinor model that describes periods of 24 and 12 h has been found to be well-suited to approximate the nightly drop of blood pressure (Halberg et al 1981). However, the parameters provided by this model result hard to interpret and to compare with the results provided from single- or other multiple- component cosinor methods (Fernández et al 2003), e.g. a small change in the phase of an harmonic will result in drastic changes in the resulting shape of the function, and parameters like *amplitude* will be distributed among several components.

Fitting unevenly sampled data to sinusoidal functions can also be performed by least-squares spectral analysis techniques, such as the Lomb–Scargle periodogram (Vander Plas 2018), or the Vaniček method (Taylor and Hamilton 1972). The setting of our problem is different, as in we are interested in fitting functions with a particular shape, that is not a simple sinusoidal, but an arbitrary combination of cosines, with fixed amplitudes and locked phases.

2. Methods

2.1. The mathematical model

2.1.1. Prior from general population

We model the prior of RHR over time $A(h)$ with the sum of c cosines, as defined in the multiple-component cosinor in Cornelissen (2014). This equation can be rewritten using Euler's complex exponential formula as equation (1). The number c of components to use is a free parameter:

$$\begin{aligned} A(h) &= \beta + \sum_{k=1}^c \alpha_k \frac{e^{i2\pi k(h+\phi_k)} + e^{-i2\pi k(h+\phi_k)}}{2} \\ &= \beta + \sum_{k=1}^c \frac{\alpha_{|k|}}{2} e^{i2\pi k(h+\phi_k)} + \sum_{k=1}^c \frac{\alpha_{|-k|}}{2} e^{i2\pi(-k)(h+\phi_{|-k|})} \\ &= \beta + \sum_{k=1}^c \frac{\alpha_{|k|}}{2} e^{i2\pi k(h+\phi_k)} + \sum_{k=-c}^{-1} \frac{\alpha_{|k|}}{2} e^{i2\pi k(h+\phi_{|k|})}. \end{aligned} \quad (1)$$

The parameters β , α , and ϕ correspond to the *MESOR*, *amplitude*, and *acrophase*, and can be found as explained in Cornelissen (2014).

If we define $\psi_0 = 0$, $\psi_{k \neq 0} = \frac{\alpha|k|}{2} e^{i2\pi k\phi|k|}$ we can rewrite equations (1) as (2)

$$A(h) = \beta + \sum_{k=-c}^c \psi_k e^{i2\pi kh}. \quad (2)$$

Note that we have defined ψ so that ψ_{-k} is equal to the complex conjugate of ψ_k . Therefore, the number of real parameters has not changed.

Equations (1) and (2) are equivalent to cosinor analysis (Cornelissen 2014) and the Vaníček method (Taylor and Hamilton 1972). From the next section we introduce novel contributions.

2.1.2. Fitting user data with locked parameters

Single component cosinor can only accurately fit data modulated by a simple sinusoidal factor. Circadian effect on RHR is unlikely to be fully captured by such a simple function. Multiple component cosinor can fit data more accurately (Cornelissen 2014), but the resulting set of fitting parameters, (a *MESOR*; an *acrophase* and an *amplitude* for each component) can not be easily interpreted, making it difficult to interpret the parameters of each component. Moreover, fitting a model for each user with all the free parameters will fit shapes that might not be what circadian is supposed to be, and it will be difficult to get an idea of what is the phase and amplitude.

We use the $A(h)$ model provided in the previous section as the prototype (and prior) for users RHR throughout the day (circadian rhythm). However, every user has a different RHR at night, during the day, and might go to sleep and wake up at different times than most of the users (that generated the model). For this reason we define a refined version of the model where some parameters are let free to adapt to users data. In particular, equation (3) shows the formula that locks parameters of the c components only allowing three degrees of freedom (the same used in single component cosinor analysis: *MESOR*, *amplitude*, and *acrophase*). This is equivalent to single component cosinor analysis, using a different function than a cosine:

$$U(\theta, h) = \theta_0 + \theta_1(A(h - \theta_2) - \beta) = \theta_0 + \theta_1 \sum_{-c \leq k \leq c} \psi_k e^{i2\pi k(h - \theta_2)}. \quad (3)$$

The free parameters in equation (3) are θ_0 , the average RHR of the user; θ_1 , the amplitude of the oscillation of the user RHR with respect to the prior. Values of θ_1 close to 1 indicate that the range of the RHR is similar to the prior, less than 1 indicate that the user RHR has a smaller range compared to the prior, etc; and θ_2 , the phase of the user circadian rhythm with respect to the phase of a cosine with period equal to one day. θ_2 ranges between -0.5 and 0.5 . Values of θ_2 close to 0 indicate that the user is sleeping at midday.

The θ parameters allow to express the user RHR, $U(h)$, as a vertical shift (θ_0), horizontal shift (θ_2), and a change in amplitude (θ_1) of the population prior $A(h)$.

2.1.3. Loss function

To find θ that best approximates $U(h)$ for a particular user, we fit $U(h)$ to the RHR data collected for that user, minimising the distance L between the data and $U(h)$ as shown in equation (4), where y_j is the j th measure of RHR of the user, x_j is the time of the measure, and w_j is the weight assigned to the j th measure:

$$L(\theta) = \sum_j w_j \|y_j - U(\theta, x_j)\|^2. \quad (4)$$

Finding the θ that minimises L is not trivial, because we use periodic functions that are not simple cosines, but a particular combination of cosines, defined by the α and ϕ parameters, that are not allowed to change. The literature of cosinor based rhythmometry does not offer any solutions, neither analytic nor numeric. Therefore, we developed an analytic solution to finding the θ that minimises L .

First, we introduce the following notation to further simplify formulae:

$$\zeta = e^{-i2\pi\theta_2} \quad (5)$$

$$F_d = \sum_j w_j y_j e^{i2\pi dx_j} \quad (6)$$

$$S_d = \sum_j w_j e^{i2\pi dx_j}. \quad (7)$$

We can now develop equation (4) using the definition of U in equation (3) into the following:

$$L(\theta) = \sum_j w_j y_j^2 + \theta_0^2 \sum_j w_j + \theta_1^2 \sum_{k,m} \phi_m \phi_k \zeta^{k+m} S_{k+m} - 2\theta_0 F_0 + \theta_1 \sum_m \phi_m \zeta^m (2\theta_0 S_m - 2F_m) \quad (8)$$

where the indices k and m range between $-c$ and c .

To minimize the loss we put to zero the partial derivatives with respect to $\theta_0, \theta_1, \theta_2$. We obtain the system in equation (9):

$$\begin{cases} \theta_0 S_0 + \theta_1 \sum_k \phi_l \zeta^k S_k - F_0 & = 0 \\ \sum_m \phi_m \zeta^m [\theta_0 S_m + \theta_1 \sum_k \phi_k \zeta^k S_{m+k} - F_m] & = 0 \\ \sum_m m \phi_m \zeta^m [\theta_0 S_m + \theta_1 \sum_k \phi_k \zeta^k S_{m+k} - F_m] & = 0. \end{cases} \quad (9)$$

Substituting θ_0 and then θ_1 in the equations of the system (9) we find a polynomial in ζ :

$$0 = \sum_{m,l,k} m \phi_m \phi_l \phi_k \zeta^{m+l+k} \Psi_{m,l,k} \quad (10)$$

where

$$\begin{aligned} \Psi_{m,l,k} &= F_k S_l S_m + F_k S_0 S_{m+l} - F_0 S_k S_{m+l} \\ &\quad - F_m S_l S_k - F_m S_0 S_{k+l} + F_0 S_m S_{k+l}. \end{aligned} \quad (11)$$

Since the indices m, l , and k range between $-c$ and c , and $\forall m. \Psi_{m,m,m} = 0$, the polynomial in equation (10) has degree $6c - 2$.

The roots of the polynomial in ζ can be found with spectral methods, i.e. using the eigenvalues of the companion matrix (Edelman and Murakami 1995) to find an approximation, and then the solutions can be refined using the Newton–Raphson method.

From the roots of the polynomial we can determinate the values of θ_2 , and subsequently of θ_1 and θ_0 , where the total derivative of $L(\theta)$ is zero. We compute $L(\theta)$ using equation (8) on these points to determine the absolute minimum.

Our model is based on the weighted least squares approach. Data do not need to respect the assumption of homoscedasticity, and weights can be designed to focus the analysis on certain areas of the data, e.g. give less weight to data collected in movement.

2.1.4. Engineering considerations

The system in equation (9) depends on data indirectly through the values of F and S . Moreover, also the square loss function can be expressed in term of these values (and the total sum of y_i^2). This means the whole algorithm does not require access to the full dataset, but only a limited set of values (namely the values of S, F and the sum of squares of y values if we want to output an absolute value for the loss). All these values are moreover an additive function of data sets, that is for two disjoint set of values $A = \{x_i, y_i\}$ and $B = \{x_j, y_j\}$ $S_{k,A \cup B} = S_{k,A} + S_{k,B}$ and $F_{k,A \cup B} = F_{k,A} + F_{k,B}$. This implies that it can be computed separately for each subset of samples, and then added together to obtain the total value: for example it can be accumulated directly in the user's device, with big advantages in computational complexity, data transfer and privacy. In particular the asymptotic analysis of the algorithm complexity is the following (where N is the number of data points, c the number of components):

- The accumulation of data into the surrogate values S and F is $\mathcal{O}(N \cdot c)$.
- To solve the polynomial we need to find eigenvalues of a $6c - 2$ sided matrix which requires $\mathcal{O}(c^3)$ operations.
- To compute the loss for a given set of fit parameters we required $\mathcal{O}(c^2)$ operations, which has to be repeated for each candidate solution, for a total of $\mathcal{O}(c^3)$ operations.

The total complexity is thus $\mathcal{O}(c^3 + Nc)$, in particular it only depends linearly on the size of data.

2.2. Experimental setting

To test the accuracy of our model of circadian effect on RHR against the traditionally used single component cosinor, we used the data collected by Biobeats, a corporate well-being company⁴, through their product Biobase, a stress management app that uses a wrist worn wearable device (Biobeam) to passively collect activity data (number of steps performed every 20 s), sleep phases by a triaxial accelerometer with a sample rate of 100 Hz, and heart beats by photoplethysmography sensors (i.e. Interbeat Interval durations are measured for 2 min every

⁴ www.biobeats.com/

30 min, with a sample rate of 50 Hz; and the average HR of 1 min of heart activity is measured every 10 min, with a sample rate of 25 Hz). To test our model we used a dataset collected by Biobeats from two pilots with financial sector organizations. Forty-seven individuals ($F = 52\%$, $M = 47\%$), with an age range of 20–63 years ($M = 29.29$, $SD = 11.38$), agreed to participate to the experiment and wore a wrist worn wearable device that collected HR every 10 min, activity data (steps) every 20 s, and sleep phases every night, for four weeks. The number of HR measurements per user was in average 3471 ($SD 2964$).

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

2.2.1. Building a prior by fitting β and ψ on the general population

To build a prior for the RHR we proceeded as follows:

- for each user we kept only time series with at least 100 HR samples. After this preprocessing step dataset had a total of 163 179 HR measures;
- we split the dataset using a leave-one-out strategy, training the prior on a dataset that did not include the user for which we wanted to calculate the θ parameters;
- for each HR measure, we calculated the amount of steps taken by the user in the previous 5 min;
- we removed from the dataset the HR points where the number of steps in the previous 5 min was larger than 10. This step was carried out to ensure we estimate RHR, instead of HR that is influenced by the physiological request of physical activity. After this preprocessing step the number of HR datapoints decreased from 163 179 to 132 498;
- on each user we first performed cosinor analysis on a single component to find the *Acrophase* (every user will generally have different sleep habits);
- we then aligned the data of all users, forcing all users to have the same phase as a cosine with period equal to 1 d; and
- finally, we used multiple-components cosinor on the phase aligned data from all users, using $c = 4$ cosines, with period of 1 d, half a day, a third of a day, and a quarter of a day.

Figure 1 shows one of the priors fitted using the leave-out-out approach. In that case the values for β and ψ , from equation (2), are the following:

- $\beta = 68.75$;
- $\psi_1 = 1.63$;
- $\psi_2 = 0.69e^{-i0.97\pi}$;
- $\psi_3 = 0.13e^{-i0.06\pi}$;
- $\psi_4 = 0.19e^{i0.49\pi}$.

β represents the population average. ψ_1 describes the first of the four components cosinor fitting. The modulus expresses how important is this component in the fitting, and the argument expresses the phase of this component. ψ_1 has argument equal to zero, because we explicitly aligned all users to a cosine. We can see that the first component captures 62% of the overall information, as the modulus of ψ_1 is equal to 0.62 multiplied by the sum of the moduli of all ψ_1, ψ_2, ψ_3 , and ψ_4 .

2.2.2. Building the models of a user resting heart rate by fitting θ on the data of that user

To obtain the analysis of the user of interest, we proceeded as follows:

- for every HR measure, we calculated the number of steps in the 5 min preceding the timestamp of the HR measure;
- we removed all the HR measures with a corresponding number of steps larger than 10. Therefore we removed all the HR measures that happened during, or immediately after, some physical activity;
- with a rolling window approach, for every HR measure
 - we applied exponential decay to the previous HR measures. The decay function is applied to the weights w_i used in the loss function L . $w_i = e^{-\Delta t}$, where Δt is the time, in days, between the datapoint i and the latest datapoint, that triggered the update of the model. Therefore, the weight approximately halves every day;
 - we then calculated the surrogates S and F , solved the polynomial in ζ , finding the θ parameters of equation (3) that minimise the loss function.

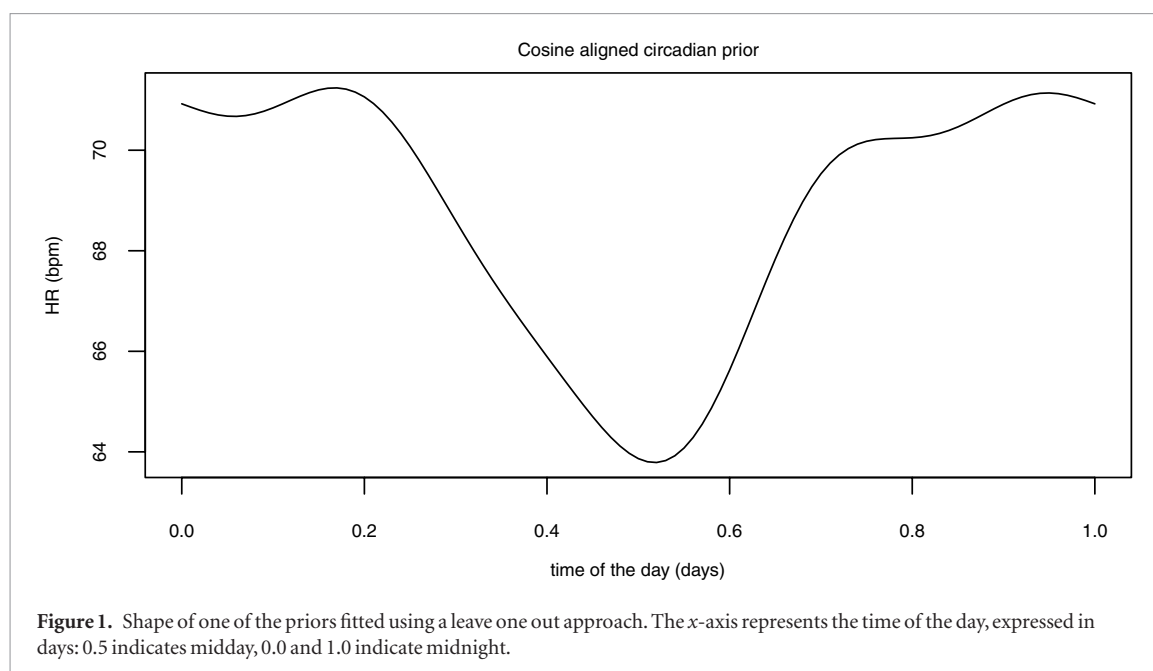


Figure 2 shows θ_0 , θ_1 , and θ_2 , calculated at each HR measure.

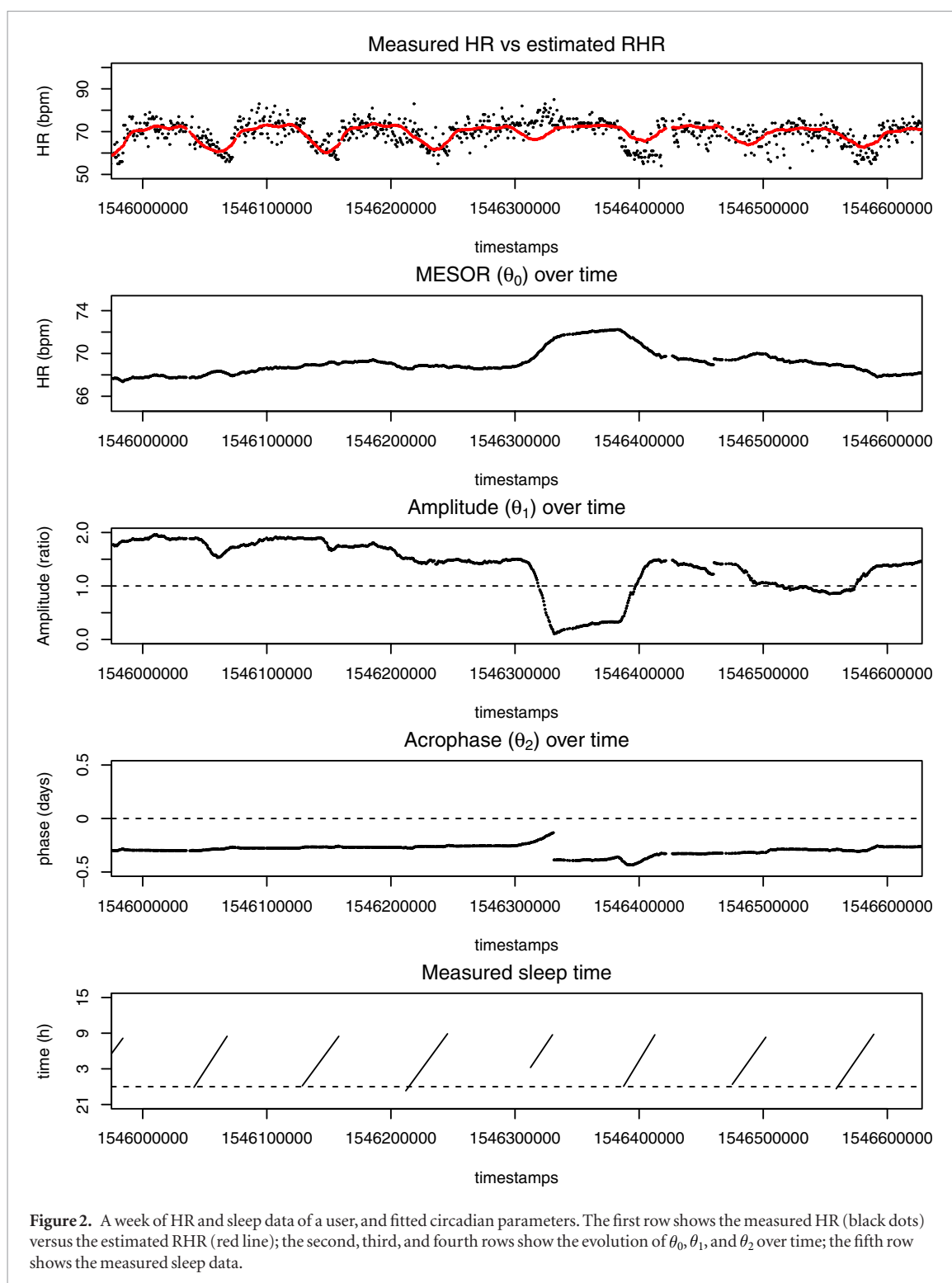
The choice of using an exponential decay function over time ensures that recent data is weighted more than old data. This approach allows the model to evolve over time, reacting to changes in sleep habits and physiological status. An alternative approach could have been to consider measures in finite time moving windows. However, this latter approach could yield unpredictable results in presence of wide sections of missing data, a frequent situation in data from field experiments, caused by users forgetting to put the band back on after recharging it. An exponential decay approach ensures that even in presence of wide sections with missing data, a usable model is always available.

Figure 2 shows seven days of HR and sleep data of one of the users, and the estimated circadian parameters:

- the first graph shows the HR measures of the user as black dots, and the RHR as estimated by the circadian model as red dots. The x-axis is time expressed as UNIX-timestamps, the y-axis is HR in beats per minute;
- the second graph shows the estimated *MESOR*, θ_0 , evolving over time, changing for every HR measure. The x-axis is time expressed as UNIX-timestamps, the y-axis is HR in beats per minute;
- the third graph shows the estimated *amplitude*, θ_1 , evolving over time, changing for every HR measure. The x-axis is time expressed as UNIX-timestamps, the y-axis is the ratio between the amplitude of the user model and the amplitude of the population model. A value around 1 (the dotted horizontal line) indicates that the circadian amplitude is in line with the population average, values above 1 indicate that the amplitude is larger than the population average, values below 1 that the amplitude is below the population average;
- the fourth graph shows the estimated *acrophase*, θ_2 , evolving over time, changing for every HR measure. The x-axis is time expressed as UNIX-timestamps, the y-axis is phase in days with respect to a cosine with period 1 d, with range between -0.5 and 0.5 d. The horizontal dotted line shows a phase of zero, indicating that the user sleeps at midday; and
- the fifth graph shows the sleep data measures by the wearable, in the graph we show at what time the user went to bed and woke up, as a line. The horizontal dotted line indicates midnight. We did not use this data in our model, and is shown here as a reference for the sleep behaviour of the user.

Looking at the estimated circadian parameters, it can be seen that on the fourth night, slightly after timestamp = 1546 300 000, the *MESOR* increases and the *amplitude* decreases, indicating non optimal physical conditions of the user, and that the *acrophase* suddenly changes, increasing from the stable average. Observing the sleep data from the band it can be seen that on that same night the user suddenly changed sleep habits (from a healthy stable time to bed of about 11 pm, to 3 am). The lack of proper sleep can be seen in the HR data and is reflected in the changes in θ_0 and θ_1 . The sudden change in the user sleep habit is reflected in the sudden change in θ_2 .

Figure 3 shows the data of one day of a user, with the fitted single component cosinor model and our circadian model. We can see that our model better captures the sudden change in HR when the user awakens, and the fact that the time awake is longer than the time asleep. More examples with the fitted circadian model of multiple users can be found in the supplementary materials (stacks.iop.org/PM/40/095001/mmedia).

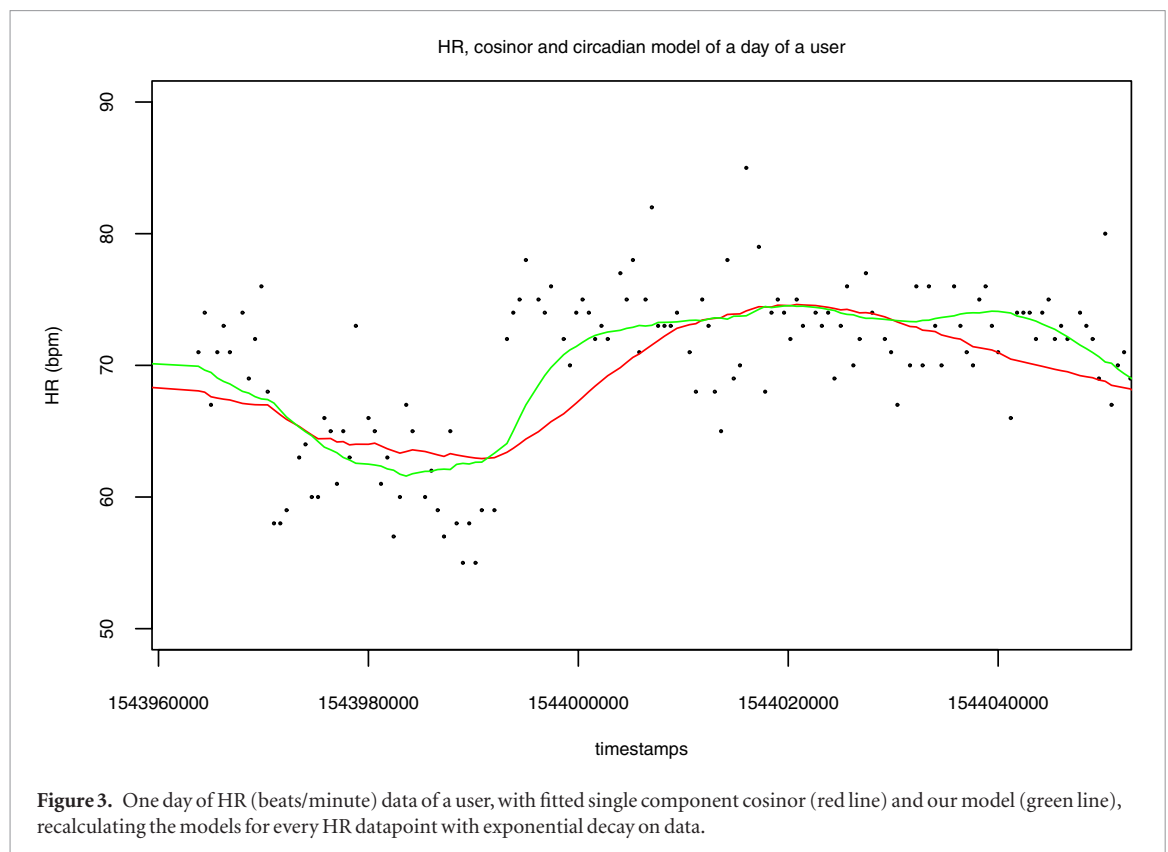


2.2.3. Predict user's resting heart rate

We applied the method described in the previous sections to obtain an evolving model of the users. We used those models to predict the next HR, given all the user's previous HR datapoints. We compared the prediction accuracy of our model against the single component cosinor model.

The cosinor RMSE is 5.73 ± 2.39 , our circadian model RMSE is 5.15 ± 1.07 . Our circadian model RMSE is in average 10% lower than the cosinor RMSE. A paired t-test returns, 0.59 mean of the differences between cosinor RMSE and our circadian RMSE, with a 95% confidence between 0.02 and 1.15 (p-value < 0.05).

We also tested our model against the null hypothesis that simply using the previous HR measure to predict the current HR measure yields better results than modelling circadian effect on HR. The null hypothesis model has RMSE equal to 5.94 ± 1.38 . A paired t-test returns a mean of differences between our circadian model RMSE and the null hypothesis model RMSE of 0.79, with a 95% confidence between 0.57 and 1.01 (p-value < 0.001).



The paired t-test between cosinor RMSE and null hypothesis RMSE returns a non significant difference of means (p-value >0.5).

The loss function $L(\theta)$ can be used to calculate the standard error, useful to have an estimation of the distribution of the predicted data. This information could be used to automatically analyse HR activity, finding anomalous data points that could be caused by factors not modelled by circadian effect on RHR, such as physiological conditions.

3. Conclusion

In this paper we presented a novel mathematical model that can be used to approximate a user's RHR. The model achieves the expressiveness of the multiple component cosinor, i.e. the model fits the data. At the same time the parameters of the model are easy to interpret parameters: *MESOR*, *amplitude* and *acrophase*, the same produced by the single component cosinor model. The model does not require uniformly sampled data, and allows to weight each datapoint independently. Hence, the results showed in this paper suggest that it is possible to accurately predict circadian RHR rhythm by using wrist worn PPG device permitting to make inference about health status of the people at home without the need of expansive devices. Moreover, our model allows the fitting to require a signal with a predetermined shape (a combination of sinusoid functions with fixed amplitudes and phases). We use this characteristic to force the shape of the expected effect of circadian rhythm on RHR, an operation not easily doable with the method used in the literature.

Continuously assessing the evolution of all the parameters of RHR circadian rhythm (i.e. *MESOR*, *amplitude* and *acrophase*) it is possible to make inference about all-cause mortality. In particular, it is possible to evaluate the quality of sleep-wake circadian rhythm from the wrist RHR. For example, low arousability, e.g. circadian rhythm with long sleep time duration, have been found to be a possible risk factor for sudden-infants-death-syndrome (SIDS) (Dvir et al 2018). A home used wrist-worn HR sensors device with an accurate circadian rhythm analysis showing increase in the sleep time duration (low arousability) might be used to prevent SIDS. Moreover, heart diseases could be induced by sleep disorder also in adulthood, i.e. too long (more than 8 h per night) or too short (less than 7 h per night) sleep duration). Hence, an accurate sleep time duration analysis may be used to screen heart diseases using simple home device (Shahar et al 2001).

The model provided in this study is computationally cheap, depending linearly on the size of the data. The computation of the model does not need the full dataset, but only two surrogates (F and S). This implies that the model can be implemented in a streaming approach, where the data is accumulated on the device, with important consequences for security and privacy of the data, that never leaves the user devices. Moreover, this model is,

to the best of our knowledge, the only computational model that can be used to predict users RHR and to analyze the data using the parameters traditionally used to describe the circadian modulation of physiological activity. The model we developed goes beyond fitting circadian activity on RHR, and it can be used to fit arbitrary periodic real valued time series, but also vectorial data (e.g. gesture recognition from the accelerometer), or complex data. With an extension to the mathematical model it could be used to fit non periodic data using a linear combination of an arbitrary set of functions.

Acknowledgment

This work is partially supported by the European Community's H2020 Program under the funding scheme INFRAIA-1-2014-2015: Research Infrastructures grant agreement 654024, www.sobigdata.eu, SoBigData. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

ORCID iDs

Davide Morelli  <https://orcid.org/0000-0002-1041-6793>

Alessio Rossi  <https://orcid.org/0000-0002-6400-5914>

References

- Aktypi A, Nurse J R and Goldsmith M 2017 Unwinding Ariadne's identity thread: privacy risks with fitness trackers and online social networks *Proc. 2017 on Multimedia Privacy and Security (ACM)* pp 1–11
- Castaldo R, Melillo P, Bracale U, Caserta M, Triassi M and Pecchia L 2015 Acute mental stress assessment via short term HRV analysis in healthy adults: a systematic review with meta-analysis *Biomed. Signal Process. Control* **18** 370–7
- Cavoukian A et al 2009 Privacy by design: the seven foundational principles *Information & Privacy Commissioner, Toronto, Ontario, Canada Report 5*
- Cornelissen G 2014 Cosinor-based rhythmometry *Theor. Biol. Med. Modelling* **11** 16
- Dvir H, Elbaz I, Havlin S, Appelbaum L, Ivanov P and Bartsch R 2018 Neuronal noise as an origin of sleep arousals and its role in sudden infant death syndrome *Sci. Adv.* **4** eaar62
- Edelman A and Murakami H 1995 Polynomial roots from companion matrix eigenvalues *Math. Comput.* **64** 763–76
- Faria I and Drummond B 1982 Circadian changes in resting heart rate and body temperature, maximal oxygen consumption and perceived exertion *Ergonomics* **25** 381–6
- Fereidooni H, Frassetto T, Miettinen M, Sadeghi A R and Conti M 2017 Fitness trackers: fit for health but unfit for security and privacy *IEEE/ACM Int. Conf. on Connected Health: Applications, Systems and Engineering Technologies (IEEE)*
- Fernández J, Mojn A, Hermida R and Alonso I 2003 Methods for comparison of parameters from longitudinal rhythmometric models with multiple components *Chronobiol. Int.* **20** 495–513
- Fossion R, Rivera A and Estanol B 2018 A physicists view of homeostasis: how time series of continuous monitoring reflect the function of physiological variables in regulatory mechanisms *Physiol. Meas.* **39** 084007
- Fossion R, Rivera A L, Toledo-Roy J C, Ellis J and Angelova M 2017 Multiscale adaptive analysis of circadian rhythms and intradaily variability: application to actigraphy time series in acute insomnia subjects *PLoS One* **12** e0181762
- Fox K et al and (Heart Rate Working Group) 2007 Resting heart rate in cardiovascular disease *J. Am. Coll. Cardiol.* **50** 823–30
- Halberg E, Halberg F and Shankaraiah K 1981 Plexo-serial linear-nonlinear rhythmometry of blood pressure, pulse and motor activity by a couple in their sixties *Chronobiologia* **8** 351–66
- Koukkari W and Sothern R 2006 *Introducing Biological Rhythms: a Primer on the Temporal Organization of Life, with Implications for Health, Society, Reproduction, and the Natural Environment* (Netherlands: Springer)
- Kristal-Boneh E, Silber H, Harari G and Froom P 2000 The association of resting heart rate with cardiovascular, cancer and all-cause mortality. Eight year follow-up of 3527 male israeli employees (the cordis study) *Eur. Heart J.* **21** 116–24
- Larsson S, Drca N, Mason A M and Burgess S 2019 Resting heart rate and cardiovascular disease *Circ.: Genomic Precis. Med.* **12** e002459
- Leise T L 2017 Analysis of nonstationary time series for biological rhythms research *J. Biol. Rhythms* **32** 187–94
- Leise T L and Harrington M E 2011 Wavelet-based time series analysis of circadian rhythms *J. Biol. Rhythms* **26** 454–63
- Li F, Sheng C, Zhang D, An D, Huang J, Li Y and Wang J 2019 Resting heart rate in the supine and sitting positions as predictors of mortality in an elderly chinese population *J. Hypertension* **37** 2024–31
- Malik M, Farrell T and Camm A 1990 Circadian rhythm of heart rate variability after acute myocardial infarction and its influence on the prognostic value of heart rate variability *Am. J. Cardiol.* **66** 1049–54
- Morelli D, Bartoloni L, Colombo M, Plans D and Clifton D A 2017 Profiling the propagation of error from PPG to HRV features in a wearable physiological-monitoring device *Healthcare Technol. Lett.* **5** 59–64
- Nakagawa M, Iwao T, Ishida S, Yonemochi H, Fujino T, Saikawa T and Ito M 1998 Circadian rhythm of the signal averaged electrocardiogram and its relation to heart rate variability in healthy subjects *Heart* **79** 493–6
- Sartipy U, Savarese G, Dahlström U, Fu M and Lund L H 2019 Association of heart rate with mortality in sinus rhythm and atrial fibrillation in heart failure with preserved ejection fraction *Eur. J. Heart Fail.* **21** 471–9
- Shaffer F and Ginsberg J 2017 An overview of heart rate variability metrics and norms *Frontiers Public Health* **5** 258
- Shahar E, Whitney C, Redline S, Lee E, Newman A, Nieto F, O'connor G, Boland L, Schwartz J and Samet J 2001 Sleep-disordered breathing and cardiovascular disease cross-sectional results of the sleep heart health study *Am. J. Respiratory Crit. Care Med.* **163** 19–25
- Stirling J R, Zakythinaki M, Refoyo I and Sampedro J 2008 A model of heart rate kinetics in response to exercise *J. Nonlinear Math. Phys.* **15** 426–36
- Taylor J and Hamilton S 1972 Some tests of the vaníček method of spectral analysis *Astrophys. Space Sci.* **17** 357–67
- Vander Plas J T 2018 Understanding the Lomb–Scargle periodogram *Astrophys. J. Suppl. Ser.* **236** 16