# Estimating Periodicity of Oscillatory Time Series Through Resampling Techniques

Maria J. Costa[1*] Bärbel F. Finkenstädt[2] Peter D. Gould[3]
Julia Foreman[4] Karen Halliday[4] Anthony Hall[3]
David. A. Rand[1]

[1]Warwick Systems Biology Centre, University of Warwick, UK
[2]Department of Statistics, University of Warwick, UK
[3]School of Biological Sciences, University of Liverpool, UK
[4]Institute of Molecular Plant Science, University of Edinburgh, UK

**Abstract**

Accurate estimation of the period length of time-course data from cyclical biological processes, such as those driven by the endogenous circadian pacemaker, is crucial for making inferences about the properties of the biological clock found in many living organisms. In this paper we propose a methodology that combines spectral analysis with resampling techniques termed spectrum resampling (SR). Extensive numerical studies show that SR is superior and considerably more robust to non-sinusoidal patterns than currently available methods based on Fourier approximations, namely the FFT-NLLS method by Plautz et al. (1997, *Journal of Biological Rhythms* **12**, 204-217). We also develop a nonparametric test for testing for changes in period length. The test uses resampling techniques and allows for period estimates with different variances. Simulation studies show that it attains correct nominal size and has good power properties when compared to parametric alternatives. The proposed SR method and statistical test are illustrated with real data examples.

*Keywords and phrases:* Circadian rhythms; Hypotheses tests; Period estimation; Resampling; Spectrum.

## 1 Introduction

The identification of periodic patterns is crucial to the understanding of cyclical biological processes such as the circadian rhythms that are found in, for example, humans, plants, animals, fungi and cyanobacteria. These are oscillators which are entrained to a 24 hour period by physiological forcing such as daily light-dark cycles. Recent work has

*Corresponding author: m.j.costa@warwick.ac.uk

demonstrated that the mammalian, fly, fungal and plant circadian clocks have at their heart a network of interacting genes that are regulating each other and recently developed experimental techniques allow one to probe the oscillations of the corresponding molecular components. In particular, high throughput fluorescent imaging techniques allow for the levels of mRNA and reporter protein activity to be measured at a relatively detailed time resolution usually covering around two to six circadian cycles (see, e.g., Hall et al., 2003; James et al., 2008).

Many experiments on circadian clocks are done in constant physiological conditions where there is no forcing and where the period differs from 24 hours. Although circadian, the exact period length is often unknown and may vary under different experimental conditions. The aims of this study are twofold: firstly, to provide an estimator for the period length along with a confidence interval and, secondly, to introduce a hypothesis test for equality of the period under two different experimental conditions for replicate time series data. The underlying statistical methodology is nonparametric, combining spectral analysis with resampling techniques which are straightforward to implement.

Most circadian clock related studies currently estimate period by approximating the oscillatory gene expression profiles by a parsimonious sum of sine and cosine functions within a Fourier approximation context (Straume et al., 1991). Software available for imaging data analysis, such as Lumicycle (Actimetrics, 2010), attempts to find the period by looking for the largest sinusoidal component in such a representation, but provides no measure of its accuracy. The Fast Fourier Transform Nonlinear Least Squares method (FFT-NLLS) by Plautz et al. (1997) is currently the most widely used. It applies a nonlinear least-squares minimization algorithm to estimate the Fourier coefficients and associated confidence intervals. Hence we compare our proposed methodology to the FFT-NLLS routine.

Various circadian data have been found to exhibit non-sinusoidal patterns (see, e.g., Edwards et al., 2006). Figure 1 shows the presence of asymmetric cycles and double peaks in the imaging data of clock genes in the *Arabidopsis thaliana* plant. The period length of non-sinusoidal cycles is more difficult to estimate within a Fourier representation approach because an increased number of sinusoidal components are required. The choice of these can be difficult and poses a burden to the stability of the nonlinear fitting algorithm. Here, we propose to base our period estimation on the spectral density function or spectrum (e.g., Jenkins and Watts, 1968) which is equal to the Fourier transform of the autocovariance function. The spectrum decomposes the variance of the time series process into its frequency components. Time series with a strong periodic component, such as those encountered in circadian experiments, typically produce a spectrum with a clear dominant spectral peak at the frequency corresponding to the period, that provides an estimate of the period length. Combining resampling techniques, such as bootstrap (Efron, 1979), with the spectrum further allows us to refine the estimate of the period and to obtain a measure of its precision, i.e., confidence bands. Our study shows that the resulting estimator is superior and considerably more robust to non-sinusoidal patterns than currently available methods.

The use of both bootstrap and spectral analysis for the study of time series has at-

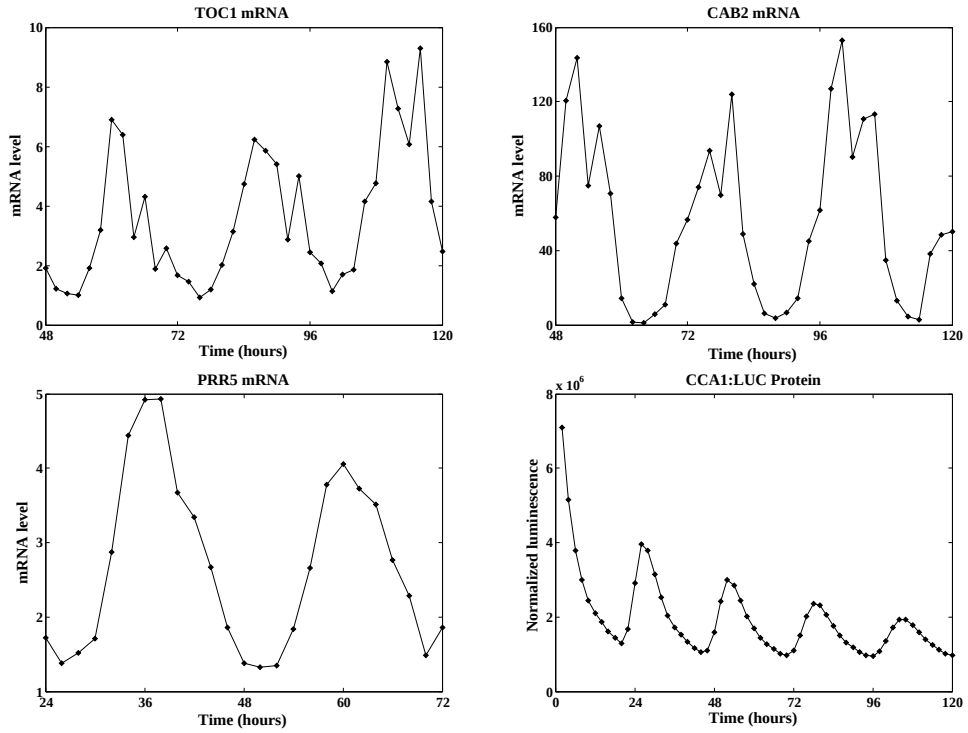


Figure 1. Experimental imaging data. Plants were first entrained in 12 hours light/ 12 hours dark cycles at 22 °C before being transferred into the experimental conditions of interest. Markers represent actual measurements taken every two hours. TOC1 mRNA levels in constant white light and constant temperature of 17 °C, sampled ZT 48-120 (top left). CAB2 mRNA levels in constant white light and constant temperature of 17 °C, sampled ZT 48-120 (top right). PRR5 mRNA levels in constant red light and constant temperature of 27 °C, sampled ZT 24-72 (bottom left). Normalized luminescence of CCA1:LUC protein under constant red light and constant temperature of 17 °C, averaged over eight replicates and sampled ZT 2-120 (bottom right). ZT stands for Zeitgeber time.

tracted considerable attention in recent years (see, e.g., Sergides and Paparoditis, 2007; Zoubir, 2010). Resampling in the frequency domain is appealing because an observed series can be transformed into a set of approximately independent statistics, the ordinates of the so-called periodogram, which is the Fourier transform of the empirical autocovariance function. Moreover, bootstrapping methods developed within the nonparametric regression framework (Härdle and Bowman, 1988) can be adapted to the periodogram. Franke and Härdle (1992) use the fact that the relationship between the theoretical spectrum and the empirical periodogram can be approximately described by means of a multiplicative regression model to propose a nonparametric, residual-based bootstrap. They also establish the asymptotic properties of their method for kernel spectral estimates. Dahlhaus and Janas (1996) extend the validity of this approach to the class of ratio statistics. Paparoditis and Politis (2003) locally resample the periodogram

ordinates. A semiparametric approach is developed by Kreiss and Paparoditis (2003), who first fit an autoregressive model to obtain a set of residuals to which they apply the bootstrap, but define the periodogram through a nonparametric estimator. An overview of bootstrap methods in spectral analysis can be found in Paparoditis (2002). In our approach the bootstrap sample of spectrum estimates gives rise to a sample of period estimates from which, following the bootstrap principle, we obtain an estimate of the period together with confidence bands.

Other spectral analysis methods for period estimation have been developed in the literature. The MESA algorithm of Burg (1972) has been implemented in the context of circadian rhythm estimation (see, e.g., Dowse and Ringo, 1989, 1991; Krishnan et al., 2001). The period is estimated using the spectrum of an autoregressive model fitted to the data. This method is, however, sensitive to the number of autoregressive terms in the model (Marple, 1980). Beyond Fourier approximations and spectral analysis, software such as WAVECLOCK (Price et al., 2008) uses wavelet analysis to estimate the period of oscillatory circadian data as a smooth function over time but provides no routine for confidence interval estimation.

Circadian data often consist of several replicate measurements of the same experiment. Figure 2 shows eight time series replicates of TOC1:LUC protein activity in *Arabidopsis thaliana* under a temperature of 17 °C and 27 °C. An important biological question is whether the precision of the clock is indeed conserved under such a change in temperature. We address this question by proposing a test of the hypothesis of equal periodicity between the two groups of replicates. The estimated period lengths constitute a sample from a population whose unknown mean value is the true period under the corresponding experimental condition. The null hypothesis that the two means are the same is tested allowing for the possibility that period estimates may have different variances. Although it is based on our spectral estimator for the period length, the testing procedure is general enough to be based on any period estimator (and its associated variance) including the one provided by the FFT-NLLS method.

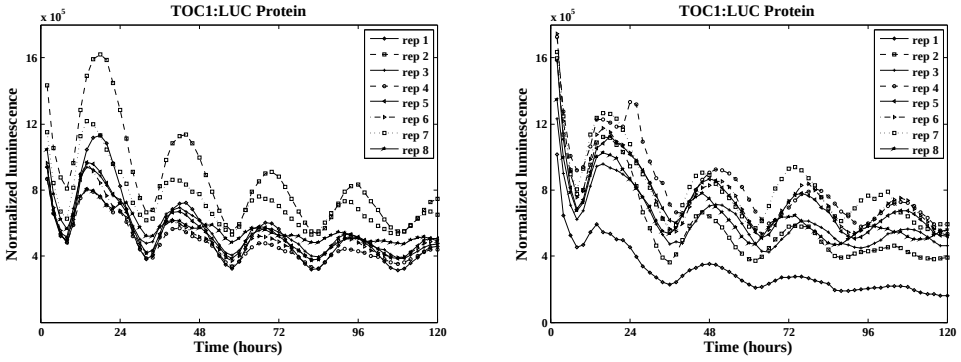


Figure 2. Replicate time series of normalized luminescence of TOC1:LUC protein under constant red light and constant temperature of 17 °C (left) or 27 °C (right), sampled ZT 2-120. Replicate j is denoted by 'rep j', $j = 1, \ldots, 8$. Markers represent actual measurements taken every two hours.

The structure of the paper is as follows. In Section 2 we introduce the proposed period estimation methodology. In Section 3 the results of extensive simulation studies examining the performance of the proposed estimator in the context of non-sinusoidal oscillations are presented. Section 4 looks at the problem of testing for conserved periodicity across experimental conditions by comparing two nonparametric generalizations of the Welch's t-test in terms of size and power. The application of the methods to experimental data is demonstrated in Section 5. We conclude our work in Section 6.

## 2 The Spectrum Resampling Method for Period Estimation

Consider a discrete-time, real-valued stationary time series $\{x_t\}_t$ measured at unit intervals of time. Assume without loss of generality that $\{x_t\}_t$ has mean zero and autocovariance function $\gamma(u)$, $u \in \mathbb{Z} = \{0, \pm 1 \pm 2, \dots\}$. The spectrum $f(\omega)$ of $\{x_t\}_t$ is defined as the Fourier transform of the autocovariance function $\gamma(u)$, i.e.,

$$f(\omega) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma(u)\, e^{-i\omega u}, \qquad \omega \in [0, \pi]\,.$$

As it is an even function of $\omega$, periodic with period $2\pi$, it suffices to consider the interval $[0, \pi]$. The total area under the spectrum curve is equal to the variance of the process. A peak in the spectrum indicates an important contribution to the variance at the frequency corresponding to the peak (see, e.g., Chatfield, 2003, for an introduction to spectral analysis). Consider realizations $x_1, \dots, x_n$ of the process $\{x_t\}_t$ taken at equal time intervals of length $\Delta$. For simplicity assume that $n$ is even and define $\tilde{n} = n/2$. The periodogram is the Fourier transform of the empirical autocovariance function and is given by

$$I(\omega_k) = \frac{\Delta}{2\pi n} \left| \sum_{t=1}^{n} x_t\, e^{-i\omega_k t\Delta} \right|^2 . \tag{1}$$

The frequencies $\omega_k = 2\pi k/n\Delta$, $k = 1, \dots, \tilde{n}$, are called Fourier frequencies (or harmonics) and are multiples of the fundamental frequency $2\pi/n\Delta$. The latter is the lowest frequency at which periods can be resolved whilst the highest detectable frequency is given by the Nyquist frequency, $\pi/\Delta$. The periodogram (1) is an asymptotically unbiased but inconsistent estimator of the spectrum. A consistent estimator is obtained through smoothing techniques such as kernel density smoothing (e.g., Parzen, 1962; Franke and Härdle, 1992). We consider the following kernel estimator for $f(\omega)$

$$\hat{f}_b(\omega_k) = \frac{\sum_{j=-\tilde{n}}^{n-1} K_b(\omega_j - \omega_k)\, I(\omega_j)}{\sum_{l=-\tilde{n}}^{n-1} K_b(\omega_l - \omega_k)}, \tag{2}$$

where $I(0) = 0$, and $I(\omega_{-k}) = I(\omega_k)$, i.e., possible boundary effects are accounted for by periodic smoothing (Lee, 1997). The function $K_b$ in (2) is such that $K_b(\cdot) = b^{-1}K(\cdot/b)$,

where the kernel $K$ is a symmetric probability density function with mean zero and unit variance, and the bandwidth $b$ is a nonnegative parameter that controls the amount of smoothing imposed over $I(\omega)$. The choice of $b$ is crucial to obtaining a good estimator $\hat{f}_b(\omega)$ of $f(\omega)$. Härdle and Bowman (1988) and Davidson and Hinkley (1997) have shown that the optimal value is such that $b = c\, n^{-1/5}$ for some positive constant $c$. Lee (1997) proposes to choose $b$ so as to minimize the risk function

$$\hat{R}(b) = \sum_{k=0}^{\tilde{n}-1} \left\{ I(\omega_k) - \hat{f}_b(\omega_k) \right\}^2 - \frac{1 - 2\,W_b}{2} \sum_{k=0}^{\tilde{n}-1} I(\omega_k)^2 \,, \tag{3}$$

which is an asymptotically unbiased estimator of the integrated mean square error

$$R(b) = \mathbb{E}\left[ \sum_{k=0}^{\tilde{n}-1} \left\{ f(\omega_k) - \hat{f}_b(\omega_k) \right\}^2 \right] ,$$

with $W_b = K_b(0) / \sum_{j=-\tilde{n}}^{n-1} K_b(\omega_j)$. Lee (1997) and Stoica and Sundin (1999) report a very good performance of this choice of $b$ for peridogram smoothing. In practice, the optimal value of $b$ is found numerically by evaluating $\hat{R}(b)$ over a grid of values for $c$ and choosing the one that minimizes $\hat{R}(b)$. In all applications considered here we have used the grid $\left\{1, 2, \ldots, 10^3\right\} \times 10^{-3}$. The kernel function used was $K(x) = (1/\sqrt{2\pi}) e^{-x^2/2}$, $x \in (-\infty, \infty)$.

The bootstrap is a resampling technique developed with the aim of gaining information about the distribution of an estimator. The main idea is to treat the original sample of values as the population and to resample from it repeatedly, with replacement, computing the desired estimate each time. This produces a sample of estimates from which a point estimate and confidence intervals can be derived (see, e.g., Davidson and Hinkley, 1997, for details on the bootstrap methodology). Bootstrap typically relies on the ability to identify independent components that can be simulated. These can be either the original sample itself, or the residuals from some suitable model that describes the data. Franke and Härdle (1992) point out that spectrum estimation can be cast in a multiplicative regression framework

$$I(\omega_k) = f(\omega_k)\, \epsilon_k \,, \qquad k = 1, \ldots, \tilde{n} \,, \tag{4}$$

where the residuals $\{\epsilon_k\}_{k=1}^{\tilde{n}}$ are approximately i.i.d. standard exponential random variables. The quality of this approximation can be improved by tapering and padding the data (Dahlhaus and Janas, 1996; Lee, 1997), where a fraction of the data at each end is first smoothed down to zero with a cosine tapering (or Tukey) window (Harris, 1997), and then padded with zeros to the right of the last observed time point until the desired sample size is met. The use of a tapering window avoids discontinuities at the boundaries. If a taper is applied to the data, the definition of the periodogram changes slightly. For simplicity, let $n$ denote the length of the padded series. The periodogram in (1) becomes

$$I(\omega_k) = \frac{\Delta}{2\pi \sum_t w(\Delta t)^2} \left| \sum_{t=1}^{n} w(\Delta t)\, x_t\, e^{-i\omega_k t \Delta} \right|^2 , \tag{5}$$

where $w(t)$ is the cosine tapering window (see Harris, 1997; Brillinger, 2001, for details). Henceforth we refer to (5) as periodogram. In order to resample from the residuals in (4) we need an initial estimate of the spectrum $f(\omega)$. This can be obtained through a kernel spectrum estimate $\hat{f}^{\dagger}_{b^{\dagger}}(\omega)$ of the form (2) with smoothing parameter $b^{\dagger}$. Bootstrap periodogram values are then generated using another kernel estimate $\hat{f}^{\ddagger}_{b^{\ddagger}}(\omega)$ of the form (2), and smoothing parameter $b^{\ddagger}$. Finally, let $b$ be the smoothing parameter from which the final bootstrap estimate of $f(\omega)$ is obtained. The three smoothing parameters are set to $b = c\, n^{-1/5}$ $b^{\dagger} = c\, n^{-1/4}$, $b^{\ddagger} = c\, n^{-1/6}$, with $c$ chosen to minimize the risk function (3) (Franke and Härdle, 1992; Davidson and Hinkley, 1997). A bootstrap estimate of $f(\omega)$ is obtained as follows:

1. Estimate the residuals $\epsilon_k$, $k = 1, \ldots, \tilde{n}$, in (4) by

$$\hat{\epsilon}_k = \frac{I(\omega_k)}{\hat{f}^{\dagger}_{b^{\dagger}}(\omega_k)}, \qquad k = 1, \ldots, \tilde{n}\,.$$

2. Sample, with replacement, independent bootstrap residuals $\epsilon^*_1, \ldots, \epsilon^*_{\tilde{n}}$ from the empirical distribution of the rescaled residuals $\hat{\epsilon}_1/\bar{\epsilon}, \ldots, \hat{\epsilon}_{\tilde{n}}/\bar{\epsilon}$, where

$$\bar{\epsilon} = \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \hat{\epsilon}_k.$$

   The rescaling ensures that the $\epsilon^*_k$'s have mean one. Bootstrap periodogram values are defined using (4) as

$$I^*(\omega_k) = I^*(\omega_{-k}) = \hat{f}^{\ddagger}_{b^{\ddagger}}(\omega_k)\, \epsilon^*_k\,, \qquad k = 1, \ldots, \tilde{n}\,,$$

   with $I^*(0) = 0$.

3. The bootstrap estimate of the spectrum $f(\omega)$ is then defined as

$$\hat{f}^*_b(\omega) = \frac{\sum_{j=-\tilde{n}}^{n-1} K_b(\omega_j - \omega)\, I^*(\omega_j)}{\sum_{l=-\tilde{n}}^{n-1} K_b(\omega_l - \omega)}\,.$$

Repeating steps 2 and 3 $R$ number of times produces a sample of kernel spectrum estimates $\{\hat{f}^*_{b,1}(\cdot), \ldots, \hat{f}^*_{b,R}(\cdot)\}$, where $R$ typically varies between 1,000 and 2,000 (see Franke and Härdle, 1992). Circadian time series data with a strong periodic component produce a spectrum with a sharp peak at the corresponding frequency. Hence, given a bootstrap spectral estimate $\hat{f}^*_{b,r}(\cdot)$, a bootstrap period estimate $\hat{p}^*_r$ of $p$ is easily found for $r = 1, \ldots, R$ as $\hat{p}^*_r = \arg\max_{\omega} \hat{f}^*_b(\omega)$. From the resulting set of $R$ bootstrap period estimates $\{\hat{p}^*_1, \ldots, \hat{p}^*_R\}$ one can derive a point estimate and confidence interval for $p$. We refer to the proposed period estimation methodology as Spectrum Resampling, abbreviated hereafter by SR.

# 3 Simulation Study

The aim of this simulation study is to evaluate the performance of the proposed SR methodology for period estimation and compare it to the FFT-NLLS procedure using synthetic clock data from a mathematical clock model.

## 3.1 General Design

From a theoretical point of view, the framework for understanding the molecular underpinnings governing circadian rhythms is based on a negative transcriptional feedback loop that generates an oscillator with a stable period of around 24 hours (see Roenneberg et al., 2008, for an overview on clock models). Synthetic clock data are thus simulated from a stochastic dynamic model where circadian oscillations are generated by a delayed negative feedback loop which constitutes a generic model for molecular clocks (Jensen, Sneppen, and Tiana, 2003; Monk, 2003). Let $M$ denote the abundance of mRNA molecules and $P$ the abundance of the corresponding protein. The Itô stochastic differential equations (SDEs) for the distributed delay model are given by (Monk, 2003; Heron, Finkenstädt, and Rand, 2007)

$$\begin{aligned} dM &= \zeta_M(t)\,dt + \sigma_M(t)\,dW_M \\ dP &= \zeta_P(t)\,dt + \sigma_P(t)\,dW_P\,, \end{aligned} \tag{6}$$

where

$$\begin{aligned} \zeta_M(t) &= \frac{\nu_1 k_1}{(k_1 + P(t))^{hc}} - \frac{\nu_2 M(t)}{k_2 + M(t)} \\ \zeta_P(t) &= \alpha\, g\,(M(t)) - \delta_P P(t) \\ \sigma_M(t) &= \left[\frac{\nu_1 k_1}{(k_1 + P(t))^{hc}} + \frac{\nu_2 M(t)}{k_2 + M(t)}\right]^{1/2} \\ \sigma_P(t) &= [\alpha\, g\,(M(t)) + \delta_P P(t)]^{1/2}\,, \end{aligned}$$

are drift and volatility functions, respectively, and $W_M$ and $W_P$ are independent Wiener processes (see Heron et al., 2007, for a detailed derivation of the drift and volatility functions in (6)). Here $g\,(M(t)) = \int_0^\infty M(t-s)\phi(s)\,ds$, where $\phi(\cdot)$ is some probability density function defined on the set of positive real numbers. To determine the true period of the oscillations we consider the ordinary differential equation (ODE) model counterpart. This is obtained by setting $\sigma_M(t) = \sigma_P(t) = 0$ in (6) above. The model resembles the clock model developed by Goldbeter (Goldbeter, 1991) with the difference that the intermediate steps between mRNA translation and synthesis of nuclear protein (represented by $P$ in (6)) are replaced by the delay function $g\,(M(t))$ in (6). Our initial choice of parameters is similar to the setting in Goldbeter (1991).

For each choice of parameter values in the model we simulated 200 replications. For each replicate, the SDE model is used to generate random data with intrinsic noise for which a period estimate is obtained. To determine the true period, $p$, a long ODE is run, and the mean peak distance is computed after convergence is achieved. Let $\hat{p}_i$ represent

the mean period estimate for replicate $i$ and define $\mathrm{SqE}_i = (p - \hat{p}_i)^2$ as the associated squared error. The overall performance of the period estimators is then measured by the mean squared error (MSE)

$$\mathrm{MSE} = \overline{\mathrm{SqE}} = \frac{1}{200}\sum_{i=1}^{200}(p - \hat{p}_i)^2 .$$

For each method a period estimate $\hat{p}$ is obtained by averaging across the 200 replications. From these an estimate of the standard deviation of $\hat{p}$ is also extracted. For all simulations we take $R$, the number of bootstrap samples, to be 1,000. From a biological point of view, a period estimate is discarded as an "outlier" if it falls outside the circadian range, say $[15\mathrm{h}, 35\mathrm{h}]$. Since we sometimes obtain a large number of so-called outliers with the FFT-NLLS estimator we also include the results disregarding all outliers, denoted by otl-FFT-NLLS.

### 3.2 Consistency

Our first simulation study focuses on the consistency properties of the SR estimator. A sufficient requirement is that it be consistent in quadratic mean and therefore we want to find evidence that

$$\lim_{N\to\infty} \mathrm{MSE} = 0 ,$$

where $N$ is either the number of cycles, the sample size, or both. For the simulations we use model (6) with $\phi$ being the gamma density function with mean $\mu = 8$ and variance $\sigma^2 = 8$, and set $hc = 4$, $\nu_1 = 1.5$, $\nu_2 = 1.3$, $k_1 = 0.2^4$, $k_2 = 0.2$, $\alpha = 2$, and $\delta_P = 0.5$, which results in sinusoidal shaped cycles with a true period of approximately 24 hours. We first use the iterative Euler method to obtain discrete realisations of the stochastic process at small time intervals of length 0.1h. The synthetic data is then generated by sampling values at a given level of coarseness or sampling frequency, SF. The latter is a function of the sample size, $n$, and the number of cycles, NC, as follows: if periodic expression levels with period $p$ are generated every fraction $\xi$ of an hour, then

$$\mathrm{SF} = \frac{\mathrm{NC}\,(p/\xi)}{n} .$$

Table 1 gives the MSE estimates for our chosen combinations of $n$ and NC. These reflect a range of different situations in terms of the amount of information available. For example, $\mathrm{NC} = 12$ and $n = 60$ results in one measurement taken every 4.8 hours, i.e., only five observations to cover a 24 hours cycle, while setting $\mathrm{NC} = 6$ and $n = 80$ yields one measurement every 1.8 hours, i.e., 14 observations per 24 hours cycle. From Table 1 we can draw the following conclusions:

- Both the SR and the FFT-NLLS estimators perform poorly for data where observations are taken at very short time intervals but covering only two complete cycles.

Table 1. Results For Simulation Study on Consistency

| | | MSE | | |
|---|---|---|---|---|
| $n$ | NC | SR | FFT-NLLS | otl-FFT-NLLS |
| 30 | 2 | 4.271 | 9.370 | – |
| | 4 | 0.694 | 1.942 | – |
| | 6 | 2.858 | 49.940 | 1.231 |
| 40 | 3 | 2.174 | 5.856 | 4.547 |
| | 6 | 0.155 | 0.290 | – |
| | 9 | 0.068 | 2.771 | 1.896 |
| 60 | 2 | 4.828 | 10.091 | – |
| | 4 | 0.477 | 0.894 | – |
| | 8 | 0.083 | 0.135 | – |
| | 12 | 0.969 | 21.319 | 0.141 |
| 80 | 3 | 2.167 | 14.197 | 4.225 |
| | 6 | 0.118 | 0.180 | – |
| | 9 | 0.062 | 0.079 | – |
| | 12 | 0.046 | 1.526 | 0.058 |
| 120 | 2 | 5.997 | 1.154 e+4 | 8.421 |
| | 4 | 0.390 | 0.476 e+4 | 0.714 |
| | 8 | 0.079 | 0.136 | – |
| | 12 | 0.049 | 0.778 | 0.057 |
| 160 | 3 | 1.415 | 7.248 e+4 | 2.396 |
| | 6 | 0.125 | 5.665 e+4 | 0.702 |
| | 9 | 0.063 | 2.285 e+4 | 0.092 |
| | 12 | 0.051 | 1.519 | 0.066 |
| 240 | 2 | 9.158 | 0.449 e+5 | 8.023 |
| | 4 | 0.384 | 2.074 e+5 | 0.876 |
| | 8 | 0.090 | 1.494 | 0.778 |
| | 12 | 0.049 | 1.571 | 0.134 |

– , if no outliers.

- The estimated MSE of the SR estimator generally decreases for larger sample size. The only instance where a rise in the value of MSE was observed was for very low sampling frequency, obtained when $n = 30$ or $n = 60$, and NC $= 6$ or NC $= 12$ respectively, which results in observations being taken approximately every five hours.

- There is a tendency for a better period estimate if one samples more cycles rather

than more frequently. However the sampling frequency should not be too small (i.e., more frequent than 5 hourly intervals).

- The SR estimator always outperforms the FFT-NLLS one, even when outlier estimates are ignored, apart from the two combinations $\mathtt{n}=60$ and $\mathrm{NC}=12$, and $\mathtt{n}=240$ and $\mathrm{NC}=2$.

These considerations are confirmed by boxplots of $\log_{10}(\mathrm{SqE})$ for SR and otl-FFT-NLLS across all replicates for different values of $\mathtt{n}$ and NC (see Supplementary Figure 1).

### 3.3 Asymmetric Cycles

In our next simulation study we investigate the performance of the estimators in terms of MSE in the presence of asymmetry. Circadian data can be observed to exhibit asymmetric cycles characterised by a short rise in expression followed by a longer, more gradual, decline as shown in Figure 1. In model (6) the level of asymmetry can be controlled through the set of parameters $\nu_1$, $\mathtt{k}_1$, $\nu_2$, and $\mu$, the mean of the gamma density $\phi$ in (6). We define three levels of asymmetry, mild ($\nu_1=0.5, \mathtt{k}_1=0.2^{\mathtt{hc}}, \nu_2=0.3, \mu=7$), moderate ($\nu_1=1.5, \mathtt{k}_1=0.2^{\mathtt{hc}}, \nu_2=0.3, \mu=6$), and severe ($\nu_1=34,\ \mathtt{k}_1=0.003^{\mathtt{hc}},\ \nu_2=0.1,\ \mu=6$). The other parameters are set to $\mathtt{hc}=5$, $\mathtt{k}_2=0.2$, $\alpha=2$, and $\delta_{\mathrm{P}}=0.5$ so that the period is approximately 24 hours. We fix the sample size to $\mathtt{n}=120$ which covers four complete cycles if samples are taken hourly (see Supplementary Figure 2). As an indicator of the level of asymmetry in the cycle we define the parameter $\eta_{\mathrm{AL}}=(\mathtt{l}-\mathtt{r})/(\mathtt{l}+\mathtt{r})$, where $\mathtt{l}$ and $\mathtt{r}$ are, respectively, the distance between the peak of the oscillation and its left and right extremities. The value of $\eta_{\mathrm{AL}}$ varies between $-1$ and 1, with positive (negative) values corresponding to left (right)-hand side asymmetry. A summary of the simulation results can be found in Table 2 (see also Supplementary Figure 3 for boxplots of $\log_{10}(\mathrm{SqE})$ for all three levels of asymmetry). For each level of asymmetry the true value of the period, $\mathtt{p}$, as well as the level of asymmetry as defined by $\eta_{\mathrm{AL}}$ are also included. The results of this simulation study can be summarised as follows:

- As can be expected both estimators show increasing MSE with increasing level of asymmetry.
- The SR methodology clearly outperforms the FFT-NLLS in terms of lower MSE at all three levels of asymmetry. For a moderate level of asymmetry, the difference in MSE between SR and FFT-NLLS is around 2-fold, increasing substantially for a severe level of asymmetry.
- For a severe level of asymmetry the FFT-NLLS estimator fails to detect circadian periodicity, while the SR method still obtains reasonable estimates at a markedly higher precision than FFT-NLLS.
- When non-circadian period estimates are removed the SR method still outperforms otl-FFT-NLLS with MSE $= 11.6371$ ($\hat{\mathtt{p}} = 27.3414$).

Table 2. Simulation results for synthetic data with asymmetric cycles

| Asymmetry Level | Parameter | SR | FFT-NLLS | otl-FFT-NLLS |
|---|---|---|---|---|
| Mild | # Outliers | 0 | 0 | – |
| | Mean ($\hat{p}$) | 24.650 | 24.502 | – |
| ($p = 24.670$) | STD | 0.043 | 0.052 | – |
| ($\eta_{AL} = 0.041$) | MSE | 0.002 | 0.031 | – |
| Moderate | # Outliers | 0 | 0 | – |
| | Mean ($\hat{p}$) | 25.539 | 25.449 | – |
| ($p = 25.760$) | STD | 0.075 | 0.081 | – |
| ($\eta_{AL} = 0.230$) | MSE | 0.055 | 0.103 | – |
| Severe | # Outliers | 11 | 80 | – |
| | Mean ($\hat{p}$) | 28.595 | 47.872 | 29.650 |
| ($p = 24.554$) | STD | 6.390 | 27.508 | 2.487 |
| ($\eta_{AL} = 0.565$) | MSE | 56.954 | 1.297 e+3 | 32.099 |

STD - standard deviation; – , if no outliers; $p$, $\hat{p}$ in hours.

We also investigated the coverage probability of the confidence intervals produced by the SR and the FFT-NLLS methods. For the SR methodology, confidence intervals are estimated by computing the respective quantiles of the bootstrap sample of period estimates (Carpenter and Bithell, 2000). For mild and moderate asymmetry levels, coverage for the SR estimator was 100% in both cases for a nominal level of 95% which indicates that the SR method is conservative. The FFT-NLLS estimator resulted in only 54% and 16% coverage, respectively, which is critically below the nominal level, i.e., the estimated confidence intervals tend to be much too narrow. For a severe level of asymmetry, both the SR and the FFT-NLLS are below the nominal level of 95%, with 35% and 0% coverage, respectively. This confirms the above conclusion about the failure of the FFT-NLLS estimator in the severe asymmetry case.

### 3.4 Cycles with Shoulder Pattern

Various core clock genes have mRNA profiles with bimodal cycles or shoulder patterns as shown in Figure 1. In the limit, these shoulders become almost double peaks and thus mask the true circadian periodicity, significantly hindering period estimation. To accomodate for a shoulder pattern in synthetic data we reformulate the mRNA dynamics in model (6) as follows

$$dM = \zeta_M(t)\,dt + \sigma_M(t)\,dW_M\,, \tag{7}$$

where $\zeta_M(t) = \tau(t) - \delta_M M(t)$, and $\sigma_M(t) = [\tau(t) + \delta_M M(t)]^{1/2}$. The function $\tau(t)$ in (7) regulating transcription of mRNA is given by the following step function for $t \in [0, p]$

Table 3. Simulation results for synthetic data with shoulder pattern cycles

| Shoulder Level | Parameter | SR | FFT-NLLS | otl-FFT-NLLS |
|---|---|---|---|---|
| Mild | # Outliers | 0 | 27 | – |
| | Mean ($\hat{p}$) | 24.423 | 28.074 | 24.464 |
| | STD | 0.097 | 12.002 | 0.100 |
| | MSE | 0.188 | 159.923 | 0.225 |
| Moderate | # Outliers | 0 | 35 | – |
| | Mean ($\hat{p}$) | 24.387 | 31.722 | 24.469 |
| | STD | 0.097 | 15.983 | 0.078 |
| | MSE | 0.159 | 313.792 | 0.226 |
| Severe | # Outliers | 0 | 39 | – |
| | Mean ($\hat{p}$) | 24.343 | 33.310 | 24.434 |
| | STD | 0.100 | 18.849 | 0.192 |
| | MSE | 0.128 | 440.187 | 0.225 |

STD - standard deviation; – , if no outliers; $\hat{p}$ in hours.

(where $p$ is the desired period)

$$\tau(t) = \begin{cases} \tau_1(t) = a \sin(\theta t) + a, & 0 \leqslant t \leqslant p/4 \\ \tau_2(t) = b \sin(\theta t + \psi) + b, & p/4 < t \leqslant p/2 \\ 0, & p/2 < t \leqslant p \end{cases} \tag{8}$$

and $\tau(t) = \tau(t - kp)$, $t \in (kp, (k+1)p]$, $k = 1, \ldots, N$, where $N$ is such that $(N+1)p$ is the largest observed time point. The period of the functions $\tau_1(t)$ and $\tau_2(t)$ is defined by $\theta$ and fixed to be lower than $p/2$. The latter ensures that a time span of length $p/4$ covers the peak in the sine wave that generates the peaks in the mRNA level oscillations. Here we set $p = 24$ hours. We assume that mRNA is degraded at a constant rate, $\delta_M$. The level of the shoulder pattern can be controlled through the ratio $a/b$. How far the two peaks in a cycle are apart is controlled by $\psi$. Again we consider three different levels: mild, moderate and severe, in increasing order of complexity, and for each we investigate and compare the performance of the SR methodology with that of the FFT-NLLS. The sample size is set to $n = 96$ with hourly observations. A mild shoulder level is obtained by setting the parameters in (7)-(8) to $a = 0.3$, $b = 0.1$, $\psi = 10$ and $\delta_M = 0.3$. Choosing $\theta = \pi/5$ yields a sine curve with a period of 10 hours. To increase the height of the shoulder we simply increase the value of $b$ in (8). Hence, a moderate shoulder level is defined by $b = 0.15$ and a severe shoulder level corresponds to $b = 0.2$ (see Supplementary Figure 4).

Table 3 summarizes the simulation results for all three shoulder levels (see Supplementary Figure 5 for boxplots of $\log_{10}(\text{SqE})$). From the results in Table 3 we can draw the following conclusions:

- The FFT-NLLS produces a large number of outliers. Here between 13% and 20% of the results had to be discarded or be examined for periodicity otherwise.
- The SR estimator outperforms the FFT-NLLS estimator in terms of lower MSE for all levels of shoulder pattern and also with the benefit given to FFT-NLLS by discarding all outliers. The difference between the two estimators is most apparent for moderate and severe shoulder levels (see Supplementary Figure 5).
- The SR estimator is fairly robust in that no outliers are observed, even for a severe shoulder level.

The coverage probabilities of the resulting confidence intervals for a nominal level of 95% were as follows: the average coverage was 100% for the SR estimator for all levels of shoulder behaviour, whereas for the FFT-NLLS it varied between only 12%, for severe shoulder level, and 22% for mild shoulder level. This confirms the findings of the previous section that the confidence intervals obtained by the SR approach are conservative, while those obtained by the FFT-NLLS method here are too narrow resulting in unacceptably low coverage probabilities.

# 4 A Two-Sample Bootstrap Test for the Comparison of Periods

Motivated by the experimental setup described in Figure 2 we now consider the problem of testing whether sets of replicate measurements from two different experimental conditions have the same underlying periodicity. Hence we focus on the comparison between the means of the two groups. We assume that the two experiments have size $n_1$ and $n_2$ given by the number of replicate time series in each experimental group.

## 4.1 Nonparametric Boostrap Tests

Let $T$ denote the test statistic of interest with observed value $t$, and let $p_\nu = \Pr(T \geqslant t \mid H_0)$ be the corresponding p-value for some null hypothesis $H_0$. In the bootstrap setting the value of $p_\nu$ is typically approximated by means of a Monte Carlo test (see, e.g., Davidson and Hinkley, 1997, and the references therein). The latter compares the observed statistic $t$ to $R$ independent values of $T$ which are obtained from corresponding samples independently simulated under the null hypothesis model. Let these simulated values be denoted by $t_1^*, \ldots, t_R^*$, then, under $H_0$, all $R+1$ values $t, t_1^*, \ldots, t_R^*$ are equally likely values of $T$, and so

$$p_\nu \approx \frac{1 + \#\{t_r^* \geqslant t\}}{R+1}. \tag{9}$$

In this setting the value of $R$ typically varies between 99 and 999 (Davidson and Hinkley, 1997). Nonparametric bootstrap tests in general compute the p-value of a test with minimal assumptions of how the data are distributed. Therefore there can be many candidates for a null model in the nonparametric case, each corresponding to different

restrictions imposed in addition to $H_0$. The choice of test statistic is usually based on the physical context of the problem. Suppose that each observed periodicity for each replicate time series, $y_{ij}$ say, carries a known positive weight $w_{ij}$, $j = 1, \ldots, n_i$, $i = 1, 2$, and that $\sum_j w_{ij} = 1$, $i = 1, 2$. These weights correspond to the inverse of the relative error of the period estimate, defined as the ratio between half the width of the estimate's confidence interval and the period estimate itself. Theoretically, the relative error takes values in $[0, 1]$. The closer its value is to zero the more precise the period estimate is. This definition of relative error is close to that of the relative amplitude error of the FFT-NLLS period estimator (Plautz et al., 1997). A test of equality of the mean periodicity for both experimental groups is formulated by the model

$$y_{ij} = \mu_i + \sigma_{ij}\, \epsilon_{ij}\,, \quad j = 1, \ldots, n_i\,, \quad i = 1, 2\,, \tag{10}$$

where the $\epsilon_{ij}$'s have zero mean and variance one, and are i.i.d. over $j$ given $i$ (but not necessarily Gaussian). The variances $\sigma^2_{ij}$ are of the form

$$\sigma^2_{ij} = \frac{\nu_i}{w_{ij}}\,, \tag{11}$$

for some $\nu_i > 0$. The null hypothesis to be tested is $H_0 : \mu_1 = \mu_2$, against the alternative $H_A : \mu_1 \neq \mu_2$. We consider two nonparametric tests that differ in their assumptions about the $\sigma_{ij}$'s.

**Test 1 ($T_1$): Homogenous variance within each group, $\sigma^2_{ij} = \sigma^2_i$, $\forall\, j$**

Let $\hat{\sigma}^2_i$ and $\bar{y}_i$ be the usual sample variance and mean for the $i$th sample, and let $h_i = n_i/\hat{\sigma}^2_i$. The test statistic is defined as (Davidson and Hinkley, 1997)

$$t = h_1(\bar{y}_1 - \hat{\mu}_0)^2 + h_2(\bar{y}_2 - \hat{\mu}_0)^2\,, \tag{12}$$

where $\hat{\mu}_0 = (h_1\, \bar{y}_1 + h_2\, \bar{y}_2)/(h_1 + h_2)$ is the estimate of the common mean under the null hypothesis. Note that the sample with higher sample variance, or lower sample size, will contribute less to the pooled mean estimate. The estimates of the variances under the null are

$$\hat{\sigma}^2_{i0} = \frac{1}{n_i - 1} \sum_j (y_{ij} - \hat{\mu}_0)^2\,,$$

and the null model studentized residuals are given by

$$e_{ij} = \frac{y_{ij} - \hat{\mu}_0}{\sqrt{\hat{\sigma}^2_{i0}}}\,.$$

Data sets are simulated under the null model

$$y^*_{ij} = \hat{\mu}_0 + \hat{\sigma}_{i0}\, \epsilon^*_{ij}\,, \tag{13}$$

with the $\epsilon^*_{ij}$'s randomly sampled from the pooled residuals $\{e_{ij}, j = 1, \ldots, n_i\,, i = 1, 2\}$. The p-value for test $T_1$ can be estimated as follows:

1. Compute the value of the test statistic, $t$, the pooled mean $\hat{\mu}_0$, and the variances $\hat{\sigma}^2_{i0}$, $i = 1, 2$.

2. Sample, with replacement, $n_1 + n_2$ values $\epsilon^*_{ij}$ from $\{e_{ij}, j = 1, \ldots, n_i\,, i = 1, 2\}$.

3. Compute the simulated data sets $y^*_{ij}$, $j = 1, \ldots, n_i\,, i = 1, 2$, using (13).

4. For each data set compute sample averages and variances, the $h_i$'s, the pooled mean, and finally the test statistic $t^*$.

5. Check whether $(t^*)^2 \geqslant t^2$ (because we are considering the general alternative $H_A : \mu_1 \neq \mu_2$).

Steps 2-5 are repeated $R$ number of times. The p-value is then estimated by (9).

**Test 2 ($T_2$): Heteroscedasticity, $\sigma^2_{ij} = \nu_i / w_{ij}$, with $w_{ij}$ known**

In order to account for the possibility that each element within a sample may have a different variance of the form (11) we consider the following weighted estimates of the sample mean and variance

$$\bar{y}_i = \sum_j w_{ij} y_{ij}\,, \quad \hat{\sigma}^2_i = \sum_j w_{ij}(y_{ij} - \bar{y}_i)^2\,, \quad i = 1, 2\,. \tag{14}$$

The pooled mean $\hat{\mu}_0$ is defined as before, but now with $\bar{y}_i$ and $\hat{\sigma}^2_i$ replaced by the corresponding estimates in (14). The main difference between $T_1$ and $T_2$ is in the data generating process under the null hypothesis. Let

$$\hat{\nu}_i = \frac{\hat{\sigma}^2_i}{n_i - 1}\,. \tag{15}$$

It can be shown that, if the error terms in (10) are normally distributed, $\hat{\nu}_i$ in (15) is the uniformally minimum variance unbiased estimator of $\nu_i$ for the model in (10) (Goldberg, Kercheval, and Lee, 2005). The test statistic is now

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\hat{\nu}_1 + \hat{\nu}_2}}\,.$$

Define $\hat{\nu}_{i0}$ as the estimate of $\nu_i$ under the null hypothesis, i.e., $\hat{\nu}_{i0} = \hat{\sigma}^2_{i0}/(n_i - 1)$, where $\hat{\sigma}^2_{i0} = \sum_j w_{ij}(y_{ij} - \hat{\mu}_0)^2$. Let

$$e_{ij} = \frac{y_{ij} - \hat{\mu}_0}{\sqrt{\hat{\nu}_{i0}/w_{ij}}} \tag{16}$$

be the studentized residuals. Data sets satisfying the null hypothesis are generated as

$$y^*_{ij} = \hat{\mu}_0 + \sqrt{\hat{\nu}_{i0}/w_{ij}}\,\epsilon^*_{ij}\,,$$

with $\epsilon^*_{ij}$ obtained from $e_{ij}$ in (16) as before. Estimation of the p-value for test $T_2$ proceeds as described above for test $T_1$ with the appropriate modifications.

For completeness we also consider the two-sample Welch's t-test (Welch, 1947), represented here by $T_0$. It assumes within sample homogeneity as in $T_1$, and that the error terms in (10) follow a normal distribution. In this case, the test statistic $t$ in (12) is shown to follow a t-Student distribution with degrees of freedom given by the Welch-Satterthwaite equation (Satterthwaite, 1946; Welch, 1947). Although here we built the statistical test on the SR period estimator proposed above, it should be noted that the test can be built on other estimators of the period of a time series.

### 4.2 Simulation Study on Size and Power

We compare the three different tests $T_0$, $T_1$ and $T_2$ based on size and power properties using Monte Carlo studies. Due to the fact that only a finite number of bootstrap samples are used to estimate p-values some power loss is expected for tests $T_1$ and $T_2$ (Davidson and MacKinnon, 2000). We consider a simple data generating process that follows the variance heterogeneity assumption of $T_2$. This is a realistic approach as can be seen from the plots in Figure 3, where period estimates and corresponding relative errors obtained with the SR method are represented for the experimental data of Figure 2. They show how period estimates from different replicates within the same experiment have different levels of precision. Hence, given sample sizes $n_1$ and $n_2$ we first generate a set of relative errors

$$r_{ij} = r_i + \kappa_{ij}\,, \quad j = 1, \ldots, n_i\,, \quad i = 1, 2\,,$$

where the $\kappa_{ij}$'s follow a normal distribution with mean zero and variance $\sigma^2_{r_i}$, for given $r_i$ and $\sigma^2_{r_i}$. The set of weights $w_{ij}$ is then defined as $w_{ij} = 1/r_{ij}$, and normalized so that $\sum_j w_{ij} = 1$. We generate the $y_{ij}$'s using a distribution with mean $p_i$, the true value of the period, and variance given by $\nu_i/w_{ij}$, where again the $\nu_i$'s are chosen beforehand. We chose a gamma distribution with mean and variance such that the conditions above for the distribution generating the $y_{ij}$'s are met. In all simulations the number of bootstrap replicates, $R$, is 499, and the number of Monte Carlo replicates, $M$, is 5,000.

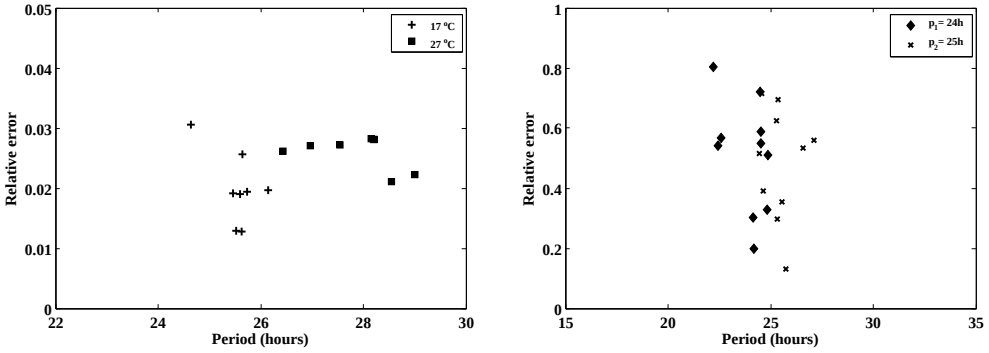


Figure 3. Relative error plots. Period estimates obtained with the SR methodology from the replicate time series experiment measuring TOC1:LUC protein activity in *Arabidopsis thaliana* (left). Example of synthetic replicate data with $\delta = 1$h and $n_1 = n_2 = 10$ (right).

To study the true size of each of the tests we estimate the empirical distribution function (EDF) of the corresponding p-values under the null using a Monte Carlo experiment. Given parameters $n_i$, $r_i$, $\sigma_{r_i}$ and $\nu_i$, $i = 1, 2$, two samples of synthetic data are generated as described above with $p_i = 24h$, $i = 1, 2$. We then calculate the p-value of the test as described in the previous section. This process is repeated $M$ times, generating a sample of $M$ p-values, $\{p_\nu^1, \ldots, p_\nu^M\}$, from which the EDF is estimated as follows,

$$\hat{F}(x) = \frac{1}{M} \sum_{m=1}^{M} I\left(p_\nu^m \leqslant x\right), \quad x \in (0, 1).$$

Let $\hat{F}_i$ be the EDF associated with test $T_i$, $i = 0, 1, 2$, and let $\alpha$ be the nominal significance level. If the distribution of the test statistic is correct, the p-value should be distributed uniformly in $[0, 1]$. Thus, a plot of $\hat{F}_i(\alpha)$ against $\alpha$ should yield a $45^o$ line. It turns out that these plots are not very informative here as all three tests seem to perform equally well for all the different parameter settings we considered. Davidson and MacKinnon (1998) propose the use of P-value Discrepancy Plots, which plot $\hat{F}(\alpha) - \alpha$ against $\alpha$, to assess and compare the performance of hypotheses tests. To test whether the observed p-value discrepancies are the result of experimental error we use the Kolmogorov-Smirnov test. Let $r_i = 0.5$, $i = 1, 2$. We consider the following grid for $\alpha$, $\{0.002, 0.004, \ldots, 1\}$. In general, all tests attain the correct nominal size, where all variation is due to experimental randomness (see Supplementary Figure 6). The only exception is the case where $n_i = 8$, $\nu_i = 0.05$ and $\sigma_{r_i} = 0.1$, when $T_0$ and $T_1$ tend to under-reject for test sizes around 0.4. It is worth noticing that the discrepancy plots of $T_0$ and $T_1$ are almost indistinguishable.

Now let $\delta = p_2 - p_1$, i.e., the difference in the mean period between the two groups. For each value of $\delta$ the power of the test is estimated via a Monte Carlo study similar to that used to estimate the EDFs under $H_0$. After a sample of $M$ p-values is obtained, the power of the test at level $\alpha$ is the proportion of times that p-value $\leqslant \alpha$, i.e., the proportion of times the test rejects the null hypothesis. Following Davidson and MacKinnon (1998) we compare power against true size rather than nominal size by plotting the EDFs of the p-values under both the null and alternative hypotheses. At each Monte Carlo replication, two sets of synthetic data, one satisfying the null hypothesis and the other not, are generated using the same sequence of random numbers to reduce experimental error. We fixed $\nu_i = 0.08$, $r_i = 0.5$, $\sigma_{r_i} = 0.1$, $i = 1, 2$, and varied the period difference $\delta$ and the sample sizes $n_i$. The grid chosen for $\alpha$ is the same as before. Figure 3 displays an example of two such generated synthetic samples with true periods set to $p_1 = 24h$ and $p_2 = 25h$. As expected, the power performance of all tests improves when both $\delta$ and $n_i$ increase (see Supplementary Figure 7). The $T_2$ test seems to have a slight advantage, especially for larger values of $\delta$. This is not surprising given that in order to generate plausible synthetic data these have to fulfil the assumptions of $T_2$. As in the size study, $T_0$ and $T_1$ perform equally well with regards to their power.

Table 4. Estimated periods and confidence intervals for circadian experimental data using the SR method

| Gene | Period Estimate (C.I.) |
|---|---|
| TOC1 mRNA | 25.452 (24.390,26.316) |
| CAB2 mRNA | 24.236 (23.256,25.317) |
| PRR5 mRNA | 24.211 (22.989,25.317) |
| CCA1:LUC Protein | 26.245 (25.641,27.027) |

C.I. - 95% confidence interval.

# 5 Application

Recall the experimental data on circadian oscillations introduced in Section 1. For each time series the period of the oscillation is estimated using the proposed SR methodology. The results for the data presented in Figure 1 can be found in Table 4. As expected, all estimated periodicities are close to 24 hours. Also, the width of the 95% confidence intervals changes slightly with the number of cycles available for estimation: they are wider for gene PRR5, for which only two complete cycles are observed, and shorter for the averaged CCA1:LUC profile which consists of four complete cycles.

Where replicate time series are available, as is the case for the data on TOC1:LUC presented in Figure 2, the tests $\mathsf{T}_0$, $\mathsf{T}_1$ and $\mathsf{T}_2$ introduced in Section 4 are used to find evidence of whether or not the period is equal with changing experimental conditions. At the 5% significance level the estimated p-values are 3.292e-5, 0.002 and 0.001 for $\mathsf{T}_0$, $\mathsf{T}_1$ and $\mathsf{T}_2$ respectively. Hence, all three tests reject the null hypothesis of period conservation when the experimental conditions change. Therefore, an increase in temperature leads to a significant change in the period of the oscillations for the reporter gene TOC1:LUC under constant red light.

# 6 Summary and Discussion

In this study we propose an improved estimator for the period of an oscillatory time series using nonparametric methods. The method relies on bootstrapping spectral estimates and is termed spectrum resampling, or SR method. In a comparison based on simulated data from circadian models we find that the SR method outperforms the currently used FFT-NLLS routine based on Fourier series approximations. It is substantially more robust to non-sinusoidal patterns of the oscillations and confidence intervals are readily available. Although the latter are conservative they are usually found to be more realistic than the confidence intervals provided by the FFT-NLLS method. The SR methodology is simple to implement and is currently developed as a piece of software freely available. It should be noted that a key assumption to any period estimation technique, including the SR method, is that the time series are stationary. In practise this can often be

achieved through detrending of the data, by fitting and subtracting a cubic polynomial. We found that the latter provides enough flexibility to accommodate all of the trends encountered in practice. In addition, the logarithmic transformation should be applied to the data, in particular if the oscillations are found to dampen with time.

The SR method as introduced here does not provide an estimate of the amplitude or phase of the oscillations, which may be of further interest. However, the SR method can be easily extended to allow for this. At each bootstrap replicate, once the Fourier frequency corresponding to the peak in the spectrum is found, the amplitude and phase of the corresponding sinusoidal can be computed using simple Fourier analysis. Hence, in principle, the SR method can provide both point estimates and confidence intervals for the amplitude and phase of the oscillations.

Some limitations remain. For example, the confidence intervals produced by the proposed methodology using the percentile approach tend to be conservative. We could have used other definitions, as proposed in Carpenter and Bithell (2000), but we feel that the percentile method is simple to interpret and inexpensive to implement computationally. Finally, the SR methodology requires a minimum of two complete cycles worth of observed data. However, most data sets resulting from circadian experiments fulfill this requirement.

In addition we have focused on the scenario where groups of replicate time series resulting from two different experimental conditions are available, and the question is whether the periodicity is the same between the two sets of replicates. We introduced two different nonparametric statistical tests, based on bootstrapping, which can be seen as generalizations of the t-test: one allowing for heteroscedasticity within each group and the other for heteroscedasticity between the two groups only. Experimental data suggests that the assumption of heteroscedasticity within a group is the most realistic one. However, the simulation studies presented here indicate that both nonparametric tests, as well as the Welch's t-test, attain correct nominal size. Allowing for within group heteroscedasticity resulted in a slight advantage in terms of the power of the test. All of the tests can be applied in conjunction with any estimator of periodicity that also retrieves a measure of the estimate's variance. The application to circadian data shows that the periodicity is significantly different under a change in temperature although both groups still oscillate with periods that are within a circadian range.

## Acknowledgements

This research was funded by BBSRC and EPSRC under the SABR initiative. MJC is funded by ROBuST (Regulation Of Biological Signalling by Temperature) grant BB/F005261/1. JF and KH are funded by ROBuST grant BB/F005237/1. PDG and AH are funded by ROBuST grant BB/F005318/1. DAR holds an EPSRC Senior Research Fellowship EP/C544587/1 and his work was also funded by the European Union BIOSIM Network Contract 005137.

## Author Contributions

MJC and BFF developed the SR method. MJC conducted the numerical studies in Sections 3, and 4. PDG and JF performed the experiments providing the real data examples in Section 1 under guidance of KH and AH. MJC wrote the paper with assistance from BFF and DAR. DAR initiated the collaboration between the theoretical and experimental groups.

# References


Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory.* Classics in Applied Mathematics. SIAM.

Burg, J. P. (1972). The relationship between maximum entropy spectra and maximum likelihood spectra. *Geophysics* **37,** 375–376.

Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine* **19,** 1141–1164.

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction.* Chapman & Hall/CRC.

Dahlhaus, R. and Janas, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *The Annals of Statistics* **24,** 1934–1963.

Davidson, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press.

Davidson, R. and MacKinnon, J. G. (1998). Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School* **66,** 1–26.

Davidson, R. and MacKinnon, J. G. (2000). Bootstrap tests: how many bootstraps ? *Econometrics Review* **19,** 55–68.

Dowse, H. B. and Ringo, J. M. (1989). The search for hidden periodicities in biological time series revisited. *Journal of Theoretical Biology* **139,** 487–515.

Dowse, H. B. and Ringo, J. M. (1991). Comparisons between periodograms and spectral analysis: apples are apples after all. *Journal of Theoretical Biology* **148,** 139–144.

Edwards, K. D., Anderson, P. E., Hall, A., Salathia, N. S., Locke, J. C., Lynn, J. R., Straume, M., Smith, J. Q., and Millar, A. J. (2006). FLOWERING LOCUS C mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock. *The Plant Cell* **18,** 639–650.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* **7,** 1–26.

Franke, J. and Härdle, W. (1992). On bootstrapping kernel spectral estimates. *The Annals of Statistics* **20,** 121–145.

Goldberg, L. R., Kercheval, A. N., and Lee, K. (2005). t-Statistics for weighted means in credit risk modelling. *Journal of Risk Finance* **6,** 349–365.

Goldbeter, A. (1991). A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase. *Proceedings of the National Academy of Sciences* **88,** 9107–9111.

Hall, A., Bastow, R. M., Davis, S. J., Hanano, S., McWatters, H. G., Hibberd, V., Doyle, M. R., Sung, S., Halliday, K. J., Amasino, R. M., and Millar, A. J. (2003). The TIME FOR COFFEE gene maintains the amplitude and timing of Arabidopsis circadian clocks. *The Plant Cell* **15,** 2719–2729.

Härdle, W. and Bowman, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association* **83,** 102–110.

Harris, F. J. (1997). On the use of windows for harmonic analysis with the discrete Fourier transform. *Biometrika* **84,** 965–969.

Heron, E. A., Finkenstädt, B., and Rand, D. A. (2007). Bayesian inference for dynamic transcriptional regulation: the hes1 system as a case study. *Bioinformatics* **23,** 2589–2595.

James, A. B., Monreal, J. A., Nimmo, G. A., Kelly, C. L., Herzyk, P., Jenkins, G. I., and Nimmo, H. G. (2008). The circadian clock in Arabidopsis roots is a simplified slave version of the clock in shoots. *Science* **322,** 1832–1835.

Jenkins, G. M. and Watts, D. G. (1968). *Spectral Analysis and Its Applications*. Holden-Day.

Jensen, M. H., Sneppen, K., and Tiana, G. (2003). Sustained oscillations and time delays in gene expression of protein Hes1. *FEBS Letters* **541,** 176–177.

Kreiss, J.-P. and Paparoditis, E. (2003). Autoregressive-aided periodogram bootstrap for time series. *The Annals of Statistics* **31,** 1923–1955.

Krishnan, B., Levine, J. D., Lynch, M. K. S., Dowse, H. B., Funes, P., Hall, J. C., Hardin, P. E., and Dryer, S. E. (2001). A new role for cryptochrome in a Drosophila circadian oscillator. *Nature* **411,** 313–317.

Lee, T. C. (1997). A simple span selector for periodogram smoothing. *Biometrika* **84,** 965–969.

Marple, L. (1980). A new autoregressive spectrum analysis algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **28,** 441–454.

Monk, N. A. M. (2003). Oscillatory expression of Hes1, p53, and NF-κB driven by transcriptional time delays. *Current Biology* **13,** 1409–1413.

Paparoditis, E. (2002). *Frequency Domain Bootstrap for Time Series*, chapter VI, pages 365–381. Empirical Process Techniques for Dependent Data. Birkhäuser.

Paparoditis, E. and Politis, D. N. (2003). The local bootstrap for periodogram statistics. *Journal of Time Series Analysis* **20,** 193–222.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* **33,** 1065–1076.

Plautz, J. D., Straume, M., Stanewsky, R., Jamison, C. F., Brandes, C., Dowse, H. B., Hall, J. C., and Kay, S. A. (1997). Quantitative analysis of drosophila period gene transcription in living animals. *Journal of Biological Rhythms* **12,** 204–217.

Price, T. S., Baggs, J. E., Curtis, A. M., FitzGerald, G. A., and Hogenesch, J. B. (2008). WAVECLOCK: wavelet analysis of circadian oscillation. *Bioinformatics* **24,** 2794–2795.

Roenneberg, T., Chua, E. J., Bernardo, R., and Mendoza, E. (2008). Modelling biological rhythms. *Current Biology* **18,** 826–835.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2,** 110–114.

Sergides, M. and Paparoditis, E. (2007). Bootstrapping the local periodogram of locally stationary processes. *Journal of Time Series Analysis* **29,** 264–299.

Stoica, P. and Sundin, T. (1999). Optimally smoothed periodogram. *Signal Processing* **78,** 253–264.

Straume, M., Frasier-Cadoret, S. G., and Johnson, M. L. (1991). *Least Squares Analysis of Fluorescence Data*, chapter 4, pages 117–240. Topics in Fluorescence Spectroscopy, Volume 2: Principles. Plenum, New York.

Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrika* **34,** 28–35.

Zoubir, A. M. (2010). Bootstrapping spectra: Methods, comparisons and application to knock data. *Signal Processing* **90,** 1424–1435.