Biostatistics (2012), **0**, 0, *pp.* 1–23
doi:10.1093/biostatistics/MariaCosta˙SM˙Biostatistics

Supplementary Material to 'Inference on periodicity of circadian time series'

Maria J. Costa*[,1], Bärbel F. Finkenstädt*[,2], Véronique Roche[3], Francis Lévi[3], Peter D. Gould[4], Julia Foreman[5], Karen Halliday[5], Anthony Hall[4], David. A. Rand[6]

[1]*Centre for Cancer Prevention, Queen Mary, University of London, London EC1M 6BQ*
[2]*Department of Statistics, University of Warwick, Coventry CV4 7AL*
[3]*Rythmes Biologiques et Cancers, INSERM U776 et Université Paris Sud, 94807 Villejuif Cedex*
[4]*School of Biological Sciences, University of Liverpool, Liverpool, L69 3BX*
[5]*Institute of Molecular Plant Science, University of Edinburgh, Edinburgh EH9 3JD*

[6]Warwick Systems Biology Centre, University of Warwick, Coventry CV4 7AL

maria.costa@qmul.ac.uk, b.f.finkenstadt@warwick.ac.uk

Summary

In this supplementary file we give details of the spectrum function, periodogram, and the problem of spectral kernel smoothing. We describe a case study illustrating the strengths of the SR method. Details on the design of the simulation study in the main paper, including choice of parameter values used, are provided. In addition, the bootstrap algorithm for estimation of p-values for the nonparametric hypothesis test is also described. A brief summary of the smoothing, detrending and tapering approaches applied here to the time series analysed is included. Additional figures that could not be placed in the main paper due to space limitations are included here. References to these figures and corresponding explanations can be found in the main paper.

Key words: Circadian rhythms; Hypotheses tests; Period estimation; Resampling; Spectrum.

*To whom correspondence should be addressed.

1. The Spectrum Function and Kernel Smoothing

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., $f(\omega) = (1/2\pi) \sum_{u=-\infty}^{\infty} \gamma(u)\, e^{-i\omega u}$, $\omega \in [0, \pi]$. As it is an even function of ω, periodic with period 2π, it suffices to consider the interval $[0, \pi]$. The spectrum function $f(\omega)$ decomposes the variance of the time series into its frequency components, and thus a peak in the spectrum indicates an important contribution to the variance of the process 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 Δ. 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 can be written as

$$I(\omega_k) = \frac{\Delta}{2\pi n} \left| \sum_{t=1}^{n} x_t\, e^{-i\omega_k t \Delta} \right|^2 . \tag{1.1}$$

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 periodogram (1.1) 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). We consider the following kernel estimator for $f(\omega)$

$$\hat{f}_b(\omega_k) = \frac{\sum_{j=-\tilde{n}}^{n-1} K_b(\omega_j - \omega_k)\, I(\omega_j)}{\sum_{l=-\tilde{n}}^{n-1} K_b(\omega_l - \omega_k)} , \tag{1.2}$$

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 (1.2) 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 Davison 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

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

with $W_b = K_b(0)/\sum_{j=-\tilde{n}}^{n-1} K_b(\omega_j)$. $\hat{R}(b)$ is an asymptotically unbiased estimator of the integrated mean square error $R(b) = \mathbb{E}\left[\sum_{k=0}^{\tilde{n}-1} \left\{ f(\omega_k) - \hat{f}_b(\omega_k) \right\}^2\right]$. 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 we have used the grid $\left\{1, 2, \ldots, 10^3\right\} \times 10^{-3}$. The kernel function used was $K(x) = (1/\sqrt{2\pi})e^{-x^2/2}$, $x \in (-\infty, \infty)$.

Section 2 in the main paper refers the use of three smoothing parameters, b, $b^\dagger$ and $b^\ddagger$, for the bootstrap resampling of the kernel spectrum estimates. The two 'extra' parameters, $b^\dagger$ and $b^\ddagger$, are necessary for control of the variance and bias of the final spectral estimate. Essentially, in order to control the variance of the kernel spectrum estimate, the residuals ϵ_k in Section 2 of the main paper are generated using an under-smoothed kernel spectrum estimate based on $b^\dagger$. To control for the bias thus introduced, the bootstrap periodogram values are obtained using an over-smoothed kernel spectrum estimate using $b^\ddagger$. The final boostrap spectral estimate, $\hat{f}^*_b(\omega)$ in the main paper, is obtained by kernel smoothing of the bootstrap periodogram values using the optimal smoothing parameter, b.

2. The Fast Fourier Transform Nonlinear Least Squares Method

The Fast Fourier Transform Nonlinear Least Squares (FFT-NLLS) method is a multicomponent cosine estimation algorithm (see, e.g., Straume *and others*, 1991; Plautz *and others*, 1997). Time series are first detrended using simple linear regression and then FFT is used to estimate the spectrum function. The period, phase and amplitude of the largest spectral peak are used to

initialise a one-component cosine fit to the data via NLLS. Upon convergence, approximate nonlinear joint asymmetric confidence intervals are derived for all parameters. If the estimated amplitude is significantly different from zero, then the two largest peaks in the estimated spectrum are used to initialise a two-component cosine fit to the data. This procedure is repeated until at least one cosine component in the fit has amplitude which is not statistically significantly different from zero, as determined by the corresponding confidence interval. At this stage the algorithm stops and the results from the penultimate iteration are used as the most statistically accurate characterization of the data. The retrieved period length is that of the cosine component in the fit with the largest amplitude.

3. Motivating Example

In this section we analyse in detail a case study which highlights the robustness of the SR method while demonstrating the reason behind the occasional erratic behaviour of the FFT-NLLS routine observed in the simulation studies described in the main paper.

Synthetic mRNA data is generated according to the established mathematical model described in Section 4 of this Supplementary Material. The parameters in the model are chosen so as to produce asymmetric cycles matching those often encountered in practice, and such that the period is within the circadian range. The data is displayed in Figure 1(a). The true period length is approximately 22h and data points are taken at an hourly rate. The data is non-sinusoidal, with the level of mRNA rising sharply, reaching its peak within approximately four hours, to then decrease slowly over a time span of between six to nine hours. The plot in Figure 1(b) shows the autocorrelation function (ACF). This has a clear oscillatory behaviour, with the first positive peak around the true value of the period, 22h. Figure 1(c) displays the output of the SR method. The top panel represents the distribution of the spectrum estimator $\hat{f}_b(\omega)$ in Section 2 of the main paper. Two clear peaks can be seen, one at approximately 22h, the largest, and

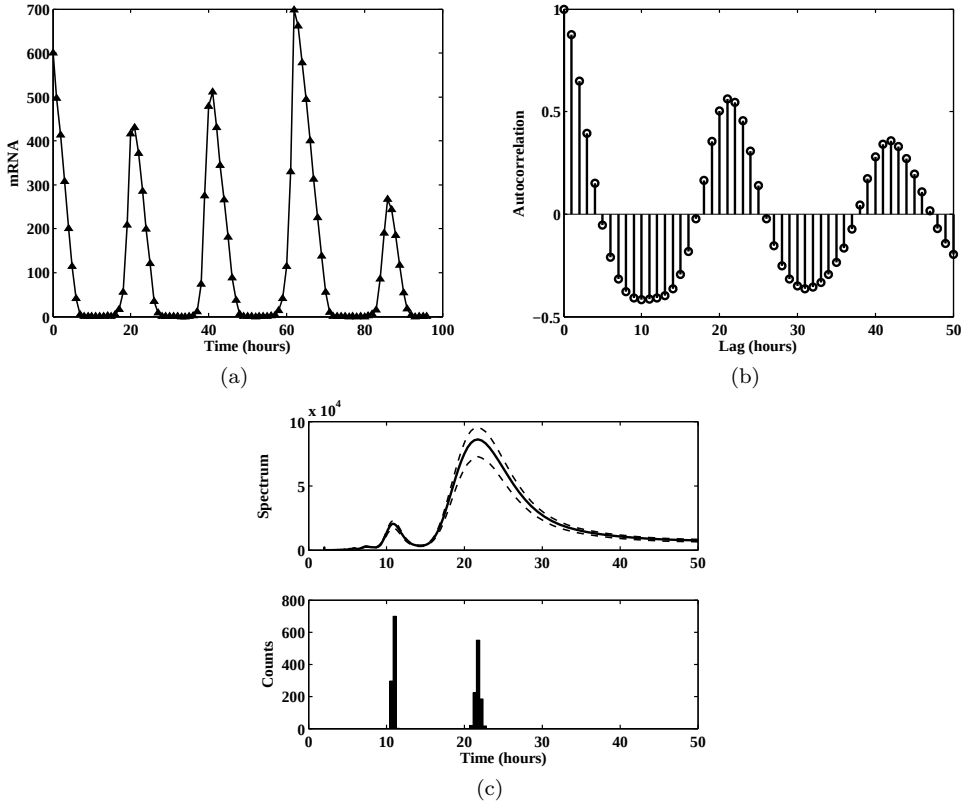

Fig. 1. Case Study. (a) Dynamics of synthetic mRNA data. Markers represent synthetic hourly observations. (b) ACF for synthetic mRNA data. (c) SR method output. Top panel displays the estimated distribution of the spectrum function estimator. Solid line represents the mean of the bootstrap sample of spectrum functions, dashed lines represent the 95% pointwise confidence intervals. Bottom panel shows the distribution of the boostrap sample of period length estimates corresponding to the two largest peaks in the spectrum.

one at approximately 11h, corresponding to the peak/trough pattern. The bottom panel shows the distribution of the estimators of the frequencies corresponding to the two largest peaks in the spectrum as determined by the SR method. These are tightly concentrated around the corresponding peaks in the spectrum function. The SR period length estimator, $\hat{p} = \arg\max_\omega \hat{f}_b(\omega)$, retrieves an estimated period length of 21.72h, with 95% confidence interval [21.28h, 22.22h].

The FFT-NLLS outputs a period length estimate of 13.79h, with corresponding 95% confidence interval [13.33h, 14.25h], clearly far from the true value of the period. Table 1 presents the details of the Fourier representation in the FFT-NLLS output. The results are ordered by

decreasing value of the amplitude estimate of each sinusoidal component in the fit. In addition, the actual order in which each component is added to the fit is also included. It shows that, at the first iteration, when a single cosine curve is fit, the correct period length is found. The fit with two cosine terms then captures the second peak in the spectrum function. However, as the algorithm tries to accommodate the non-sinusoidal shape of the oscillation, additional, spurious terms are also included. In this case, the result is the inclusion of two sinusoidal components with large amplitudes, with period lengths 13.79h and 13.88h, but whose effect in the overall Fourier fit is essentially null as a result of the opposing phases.

Table 1. *Case Study. Elements of FFT-NLLS output and the order in which they are added to the fit, by decreasing order of amplitude.*

Order	Period	Amplitude	Phase
5	13.79	470.33	2.63
6	13.88	438.90	-4.10
1	21.59	209.37	0.75
2	10.77	99.69	0.91
3	33.46	71.72	0.73
4	65.60	58.97	5.72

4. Design of Simulation Studies

In this section we describe the details of the simulation studies comparing the SR and FFT-NNLS methods and whose results are presented in Section 3 of the main paper.

4.1 *The Delayed Negative Feedback Loop Model*

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 negative feedback loop (DNFL) model are given by (Monk, 2003; Heron, Finkenstädt and Rand, 2007)

$$\begin{aligned} dM &= \zeta_M(t)\,dt + \sigma_M(t)\,dW_M \\ dP &= \zeta_P(t)\,dt + \sigma_P(t)\,dW_P\,, \end{aligned} \tag{4.3}$$

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, Finkenstädt and Rand, 2007, for a detailed derivation of the drift and volatility functions in (4.3)). 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 (4.3) 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 (4.3)) are replaced by the delay function $g\,(M(t))$ in (4.3). Our initial choice of parameters is similar to the setting in Goldbeter (1991).

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. We first use the iterative Euler method to obtain discrete realisations of the stochastic process at small time intervals of length 1/10h. The synthetic data is then generated by sampling values at a given level of coarseness or sampling frequency, s_f. The latter is a function of the sample size, n, and the number of cycles, n_c, as follows: for periodic expression levels with period p, $s_f = (p \times n_c)/n$.

4.2 *Outliers*

The definition of outlier used in this paper is motivated by biological rather than distributional arguments. Hence, the boxplots in the main paper are constructed in a slightly different way. The three central lines composing the box correspond to the usual 25th, 50th and 75th percentiles of the simulated distribution of $\log_{10}(\text{SqE})$ for both the SR method and the FFT-NLLS routine excluding all the so-called outliers. However, the whiskers now correspond to the observed minimum and maximum of $\log_{10}(\text{SqE})$ associated with period estimates within the circadian range.

4.3 *Sample Size Requirements and Consistency*

For all the simulations in this section we use model (4.3) with ϕ being the gamma density function with mean $\mu = 8$ and variance $\sigma^2 = 8$, and set $h_c = 4$, $\nu_1 = 1.5$, $\nu_2 = 1.3$, $k_1 = 0.2^4$, $k_2 = 0.2$, $\alpha = 2$, and $\delta_P = 0.5$, which results in sinusoidal shaped cycles with a true period p of approximately 24h.

Since the behaviour of the SR estimator depends both on the sampling frequency and the number of cycles covered, as described in Section 3 in the main paper, we design our simulation study on the consistency properties of the SR estimator so that it covers several scenarios corresponding to different values of sampling frequency. The sample size, n, is set to be $n = (p \times n_c)/s_f$, where as before s_f is the sampling frequency, in hours, and n_c the number of cycles. We vary $s_f \in \{1h, 2h, 4h, 5h\}$, and $n_c \in \{2, 4, 8, 16, 32\}$. For each value of n the MSE of the SR estimator is obtained as described in Section 3 in the main paper. The results can be found in Figure 2, where boxplots of $\log_{10}(\text{SqE})$ of period estimates obtained for synthetic mRNA time series using the SR method are displayed.

4.4 *Non-sinusoidal Cycles and Noise*

In model (4.3) the level of asymmetry can be controlled through the set of parameters ν_1, k_1, ν_2, and μ, the mean of the gamma density ϕ in (4.3). For mild asymmetry levels we set

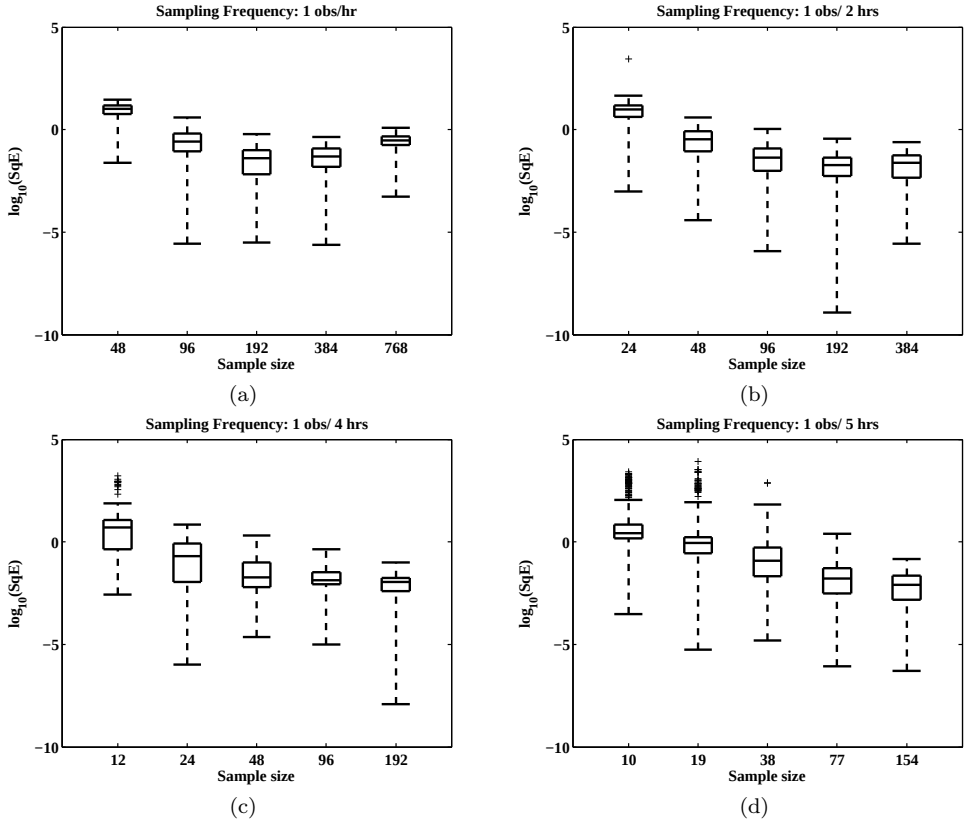

Fig. 2. Consistency. Boxplots of $\log_{10}(\mathrm{SqE})$ for period estimates obtained from synthetic mRNA dynamics using the SR method. (a) $\mathrm{s_f} = 1\mathrm{h}$. (b) $\mathrm{s_f} = 2\mathrm{h}$. (c) $\mathrm{s_f} = 4\mathrm{h}$. (d) $\mathrm{s_f} = 5\mathrm{h}$.

$\nu_1 = 0.5, \mathrm{k}_1 = 0.2^{\mathrm{hc}}, \nu_2 = 0.3, \mu = 7$, a moderate level corresponds to $\nu_1 = 1.5$, $\mathrm{k}_1 = 0.2^{\mathrm{hc}}$, $\nu_2 = 0.3$, $\mu = 6$, and severe level to $\nu_1 = 34$, $\mathrm{k}_1 = 0.003^{\mathrm{hc}}$, $\nu_2 = 0.1$, $\mu = 6$. The other parameters are set to $\sigma^2 = 8$, $\mathrm{hc} = 5$, $\mathrm{k}_2 = 0.2$, $\alpha = 2$, and $\delta_{\mathrm{P}} = 0.5$ so that the period is approximately 24h. The dynamics of synthetic mRNA and corresponding transcription functions for the different asymmetry levels are represented in Figure 3.

To accomodate for a shoulder pattern in synthetic data we reformulate the mRNA dynamics in model (4.3) as follows,

$$\mathrm{d}M = \zeta_M(t)\,\mathrm{d}t + \sigma_M(t)\,\mathrm{d}W_M\,, \tag{4.4}$$

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 (4.4)

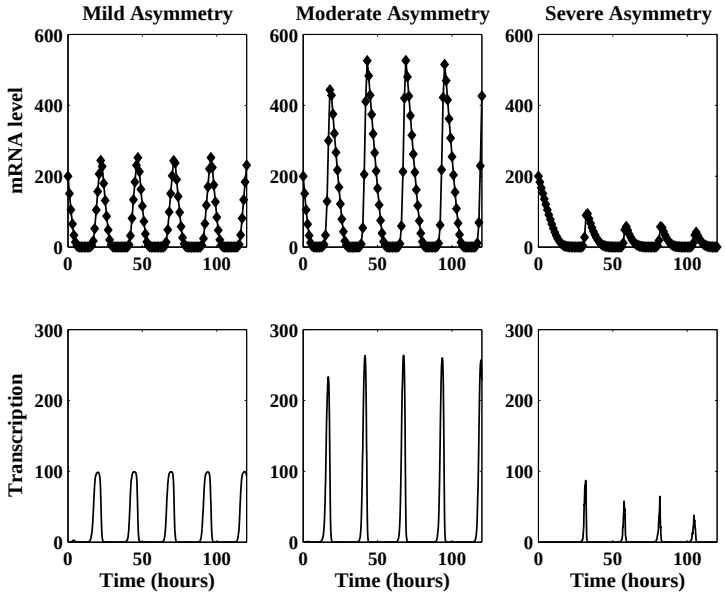

Fig. 3. Non-sinusoidal Cycles. Dynamics of synthetic mRNA (top panels) and corresponding transcription functions (bottom panels) for the different asymmetry levels. From left to right, mild asymmetry level, moderate asymmetry level, and severe asymmetry level. Markers represent synthetic hourly mRNA observations.

regulating transcription of mRNA is given by the following step function for $t \in [0, p]$ (where p is the desired period)

$$\tau(t) = \begin{cases} \tau_1(t) = a \sin(\theta t) + a, & 0 \leqslant t \leqslant p/4 \\ \tau_2(t) = b \sin(\theta t + \psi) + b, & p/4 < t \leqslant p/2 \\ 0, & p/2 < t \leqslant p \end{cases} \tag{4.5}$$

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 θ 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 $p = 24$h. We assume that mRNA is degraded at a constant rate, δ_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 ψ. We consider three different levels of shoulder pattern: mild, moderate and severe, in increasing order of complexity. The sample size is set to $n = 96$ with hourly observations. A mild shoulder

level is obtained by setting the parameters in (4.4)-(4.5) 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 10h. To increase the height of the shoulder we simply increase the value of b in (4.5). Hence, a moderate shoulder level is defined by $b = 0.15$ and a severe shoulder level corresponds to $b = 0.2$. Figure 4 contains plots of the synthetic mRNA data and the corresponding transcription functions.

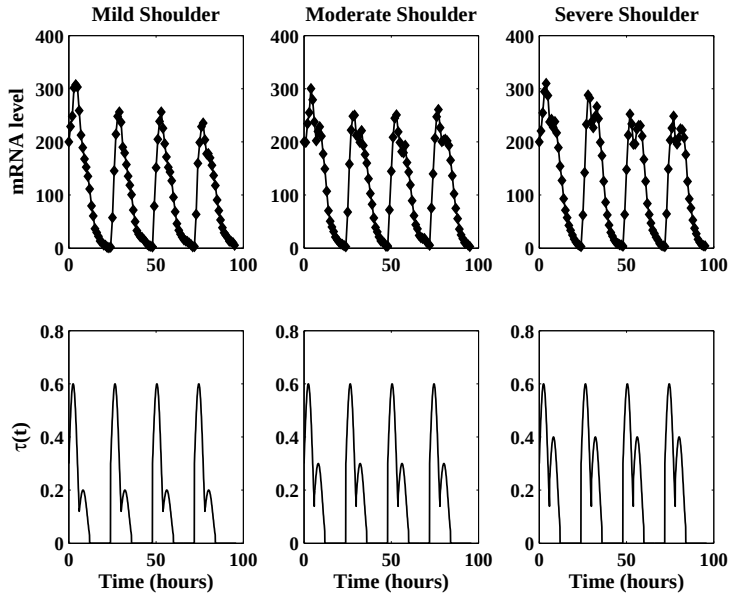

Fig. 4. Non-sinusoidal Cycles. Dynamics of synthetic mRNA (top panels) and corresponding transcription rates $\tau(t)$ (bottom panels) for the different shoulder models. From left to right, mild shoulder level, moderate shoulder level, and severe shoulder level. Markers represent synthetic hourly mRNA observations.

To simulate data with measurement error we add a simple Gaussian term to the mRNA dynamics of the ODE counterpart of the model in (4.3). The parameters in the model are the same as in Section 4.3 of this Supplementary Material. We fix the sample size to $n = 96$ with samples taken hourly. For each level of signal-to-noise ratio (SNR) the variance of the Gaussian distribution, σ^2 say, is determined as follows,

$$\sigma^2 = \frac{(1/n)\sum_{i=1}^{n} M(t_i)}{\mathrm{SNR}}.$$

The dynamics of synthetic mRNA for the different levels of SNR are represented in Figure 5.

Increasing the value of SNR results in well defined, smooth oscillations. The values of SNR chosen reflect those encountered in practice.

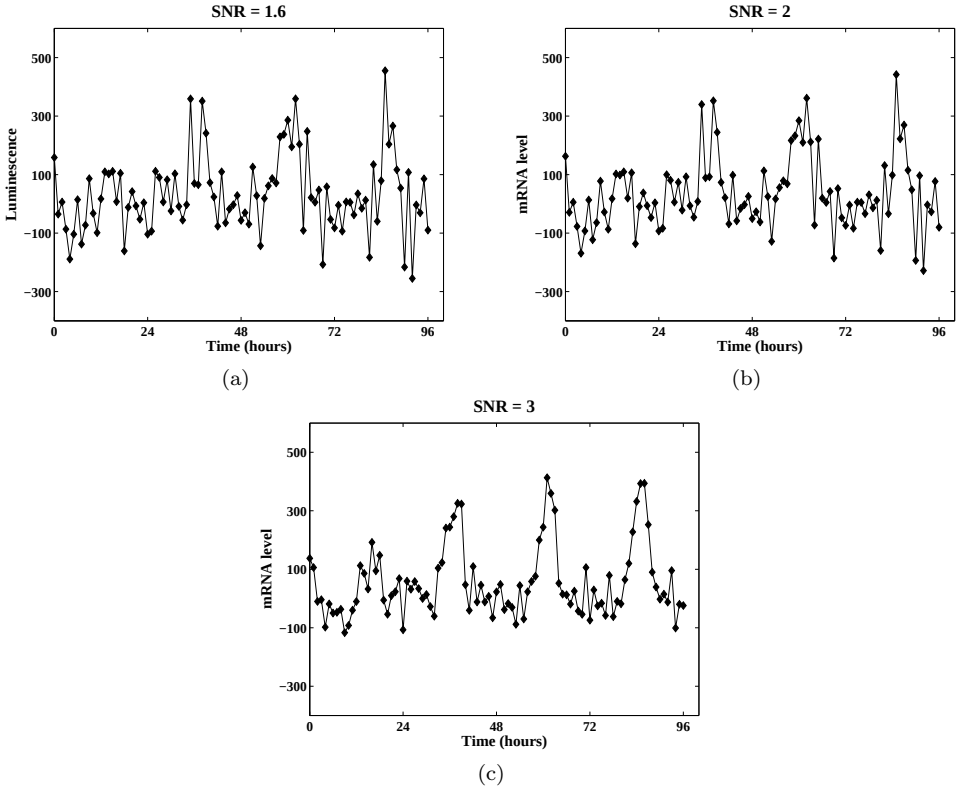

Fig. 5. Non-sinusoidal Cycles. Dynamics of synthetic mRNA for different levels of SNR. (a) SNR = 1.6; (b) SNR = 2; (c) SNR = 3. Markers represent synthetic hourly mRNA observations.

5. Bootstrap Estimate of p-value

The p-value of the nonparametric hypothesis test T_2 (and also T_1) described in Section 5 in the main paper can be estimated using a simple bootstrap algorithm as follows:

1. Compute the value of the test statistic, s, 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 ϵ^*_{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 appropriate model.

4. For each data set compute sample averages and variances, the h_i's, the pooled mean, and finally the test statistic s^*.

5. Check whether $(s^*)^2 \geqslant s^2$ (if considering the general alternative $H_A : \mu_1 \neq \mu_2$).

Steps 2-5 are repeated R times. The p-value is then estimated by $p_\nu \approx (1 + \#\{s^*_r \geqslant s\})/(R + 1)$.

6. Size and Power Study

We compare the three different tests T_0, T_1 and T_2 based on size and power properties using Monte Carlo studies. We consider a simple data generating process that follows the variance heterogeneity assumption of T_2. This is a realistic approach, from our experience, as it is often the case that period estimates from different replicates within the same experiment have different levels of precision.

Given sample sizes n_1 and n_2 we first generate a set of relative errors $r_{ij} = r_i + \kappa_{ij}$, $j = 1, \ldots, n_i$, $i = 1, 2$, where the κ_{ij}'s follow a normal distribution with mean zero and variance $\sigma^2_{r_i}$, for given r_i and $\sigma^2_{r_i}$. A 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 ν_i/w_{ij}, where again the ν_i's are chosen beforehand. We chose a gamma distribution with mean and variance such that the conditions set above for the distribution generating the y_{ij}'s are met. In all simulations the number of bootstrap replicates R for p-value estimation is 499, and the number of Monte Carlo replicates M is 5,000.

To study the true size of each of the tests proposed in the main paper 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, σ_{r_i} and ν_i, $i = 1, 2$, two samples of synthetic data are generated as described in Section 5 in the main paper, with $p_i = 24$h, $i = 1, 2$. 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, $\{p_v^1, \ldots, p_v^M\}$, from which the EDF is estimated as follows,

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

Let $\hat{F}_i$ be the EDF associated with test T_i, $i = 0, 1, 2$, and let α be the nominal significance level. If the distribution of the test statistic is correct, the p-value should be distributed uniformly in $[0, 1]$ (U[0,1]). Thus, a plot of $\hat{F}_i(\alpha)$ against α 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 α, 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 at the 5% significance level. We also compare each $\hat{F}_i$ with a U[0,1] distribution using the Anderson-Darling test statistic, A^2_M (Anderson and Darling, 1952). If its value is greater than the 5% critical value of 2.492 (Stephens, 1974) we reject the null hypothesis that $\hat{F}_i$ is U[0,1], i.e., that the test T_i attains the correct nominal size. For the simulations we take $n_i \in \{8, 16\}$, $\nu_i \in \{0.05, 0.08\}$, and $\sigma_{r_i} \in \{0.1, 0.2\}$, while fixing $r_1 = r_2 = 0.5$ and $p_1 = p_2 = 24$ hours. The chosen values reflect real experimental conditions. We consider the following grid for α, $\{0.002, 0.004, \ldots, 1\}$. Results can be found in Figures 6(a)-6(d) and Table 2.

Now let $\delta = p_2 - p_1$, i.e., the difference in the mean period between the two groups. For each value of δ 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 α 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. The grid chosen for α is the same as for the size study. The plots in Figure 6(e) and 6(f) show estimated power curves corrected for true size for different values of δ whilst fixing $\nu_i = 0.08$, $r_i = 0.5$, and $\sigma_{r_i} = 0.1$, $i = 1, 2$.

Table 2. *Size and Power simulation study. Results from Anderson-Darling test, with values of* A^2_M *for each test and each scenario.*

Test	Scenario (ν_i, σ_{r_i})			
	$(0.05, 0.1)$		$(0.08, 0.2)$	
	$n_i = 8$	$n_i = 16$	$n_i = 8$	$n_i = 16$
T_0	0.723	1.012	0.690	0.866
T_1	0.881	1.112	1.296	1.133
T_2	2.092	1.414	0.655	0.293

7. Tapering and Padding Time Series Data

The quality of the approximation for the distribution of the residuals in the multiplicative regression model described in Section 2 in the main paper can be improved by tapering and padding the data (Dahlhaus and Janas, 1996; Lee, 1997). First, a fraction of the data at each end is smoothed down to zero with, for example, a cosine tapering (or Tukey) window (Harris, 1997). The 'new' data is 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 becomes

$$I(\omega_k) = \frac{\Delta}{2\pi \sum_t w(\Delta t)^2} \left| \sum_{t=1}^{n} w(\Delta t)\, x_t\, e^{-i\omega_k t \Delta} \right|^2, \tag{7.6}$$

where $w(t)$ is the cosine tapering window (see Harris, 1997; Brillinger, 2001, for details). In the main paper the term periodogram refers to the definition in (7.6).

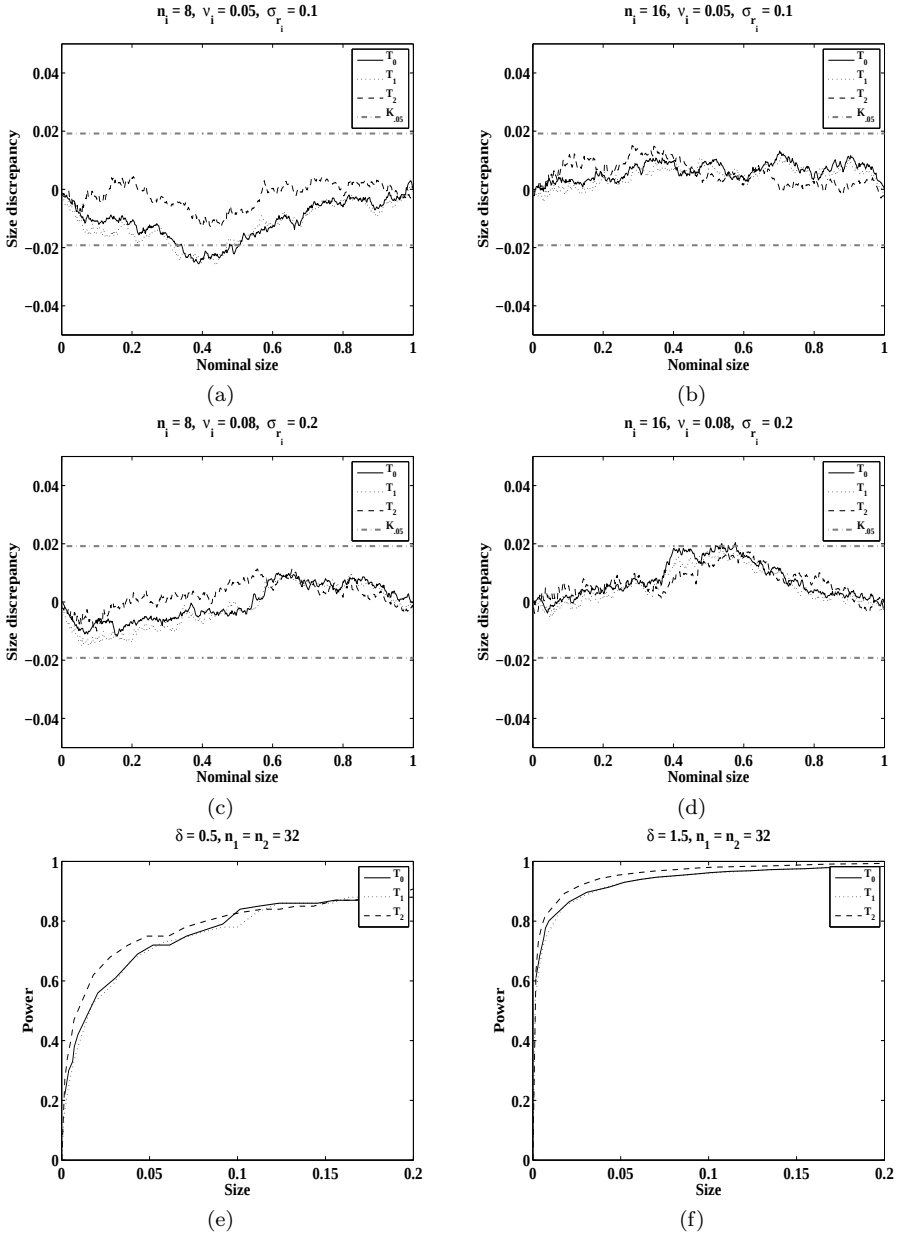

Fig. 6. Size and Power simulation study. (a)-(d) P-value discrepancy plots for T_0, T_1 and T_2 for different parameter combinations. In all cases, $\mathrm{r_i} = 0.5$, $\mathrm{i} = 1, 2$. Horizontal lines represent 5% critical value of the Kolmogorov-Smirnov test, $\mathsf{K}_{.05}$. (e) - (f) Estimated power curves corrected for true size for different values of δ (in hours). Other parameters in the data generating process are fixed to $\nu_\mathrm{i} = 0.08$, $\mathrm{r_i} = 0.5$, $\sigma_{\mathrm{r_i}} = 0.1$, $\mathrm{i} = 1, 2$.

8. Smoothing and Detrending of Time Series Data

Noisy data can be smoothed using a number of techniques. Here we consider non-overlapping smoothing of time series. We first partition the time series into $\mathfrak{b}$ non-overlapping blocks of length $\mathfrak{l}$ such that $\mathfrak{b} \times \mathfrak{l} = \mathfrak{n}$, the length of the original, unsmooth time series. For each block $\mathfrak{b}$ we compute the mean of the $\mathfrak{l}$ observations and set this as the 'new' observation.

Detrending of time series data can be accomplished by a two-step process. We first compute the logarithm of the data. This serves to stabilize the amplitude of the oscillations across the observed range and does not change the periodicity properties of any rhythms that might be present. We then fit and remove a polynomial of degree $\mathfrak{d}$ to the overall trend. We find that for the applications considered here a cubic polynomial suffices.

9. Chronotherapy Study

Figure 7 shows plots of temperature recordings from four different locations on the skin of a patient with metastatic colorectal cancer before, during and after treatment with intensive chronotherapy with chronoIFLO4. Observations are taken every minute for approximately four days. Plots of the ACF and estimated kernel spectrum distribution for the smoothed data are presented in Figure 8. We can see that even for very noisy data the ACF, and consequently the spectrum function, are able to recover the true value of the period length.

10. Human Gene Luminescence Data

Figure 9 shows plots of individual PER2:LUC time series replicates for both treatments, G1 and G2, and for the control group. There is a strong downward trend in the data which can be removed by applying the logarithm transformation combined with polynomial detrending as described in Section 8 of this Supplementary Material. Plots of the ACF and estimated kernel spectrum distribution for both treatments and control are presented in Figure 10.

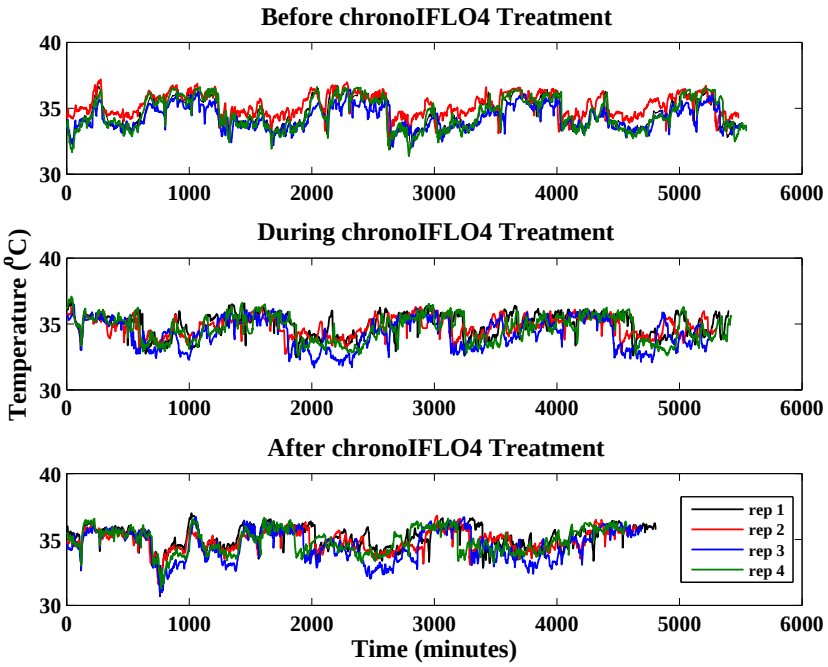

Fig. 7. Chronotherapy Study. Skin temperature data collected over approximately three days (unsmoothed) for a patient suffering from metastatic colorectal cancer before, during and after treatment with intensive chronotherapy with chronoIFLO4. For each treatment stage replicate time series represent data collected from different sites in the skin of the patients body.

11. Plant Luminescence Data

Figure 11 shows eight time series replicates of TOC1:LUC and CCA1:LUC protein activity in *Arabidopsis thaliana* at 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. Plots of the ACF and estimated kernel spectrum distribution for both CCA1:LUC and TOC1:LUC at 17°C are presented in Figure 12.

Author Contributions

MJC and BF developed the SR method. MJC conducted the numerical studies and wrote the paper with assistance from BF and DAR. VR performed the experiments for the chronotherapy study under guidance of FL. PDG

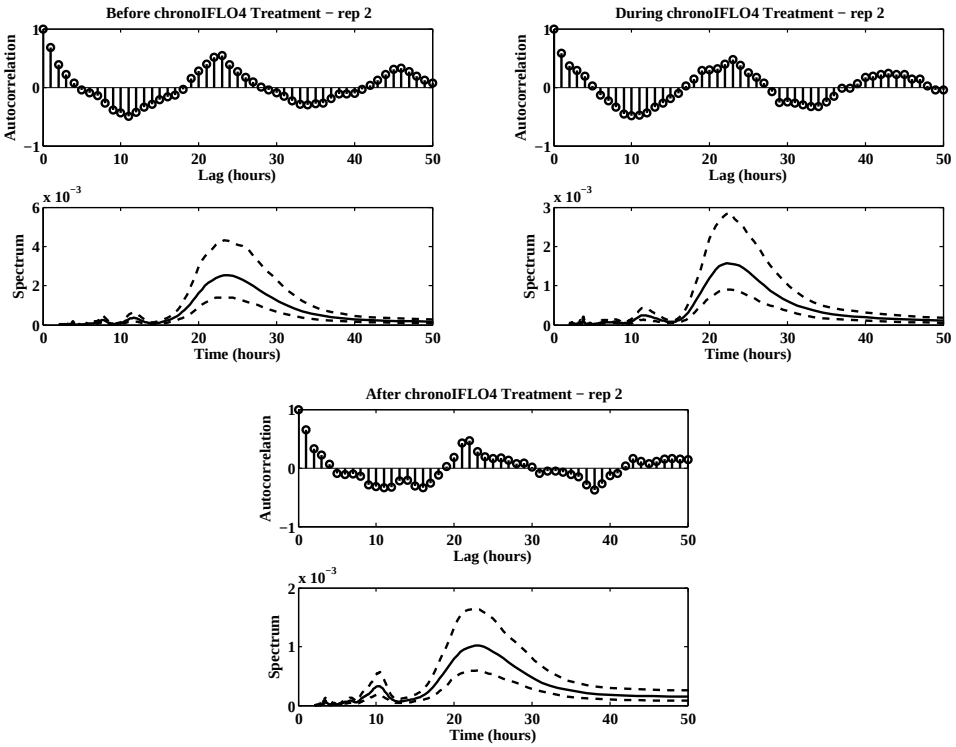

Fig. 8. Chronotherapy Study. Autocorrelation functions and estimated distribution of spectrum functions using the SR method for skin temperature data on replicate 2. In bottom panels, the solid line represents the mean of the bootstrap sample of spectrum functions, and the dashed lines represent 95% pointwise confidence intervals.

performed the experiments for the plant luminescence data under guidance of AH. JF provided data used in the development of the project. KH is principal investigator in ROBuST. DAR initiated the collaboration between the theoretical and experimental groups.

References

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain 'goodness-of-fit'criteria based on stochastic processes. *Annals of Mathematical Statistics* **23**, 193–212.

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

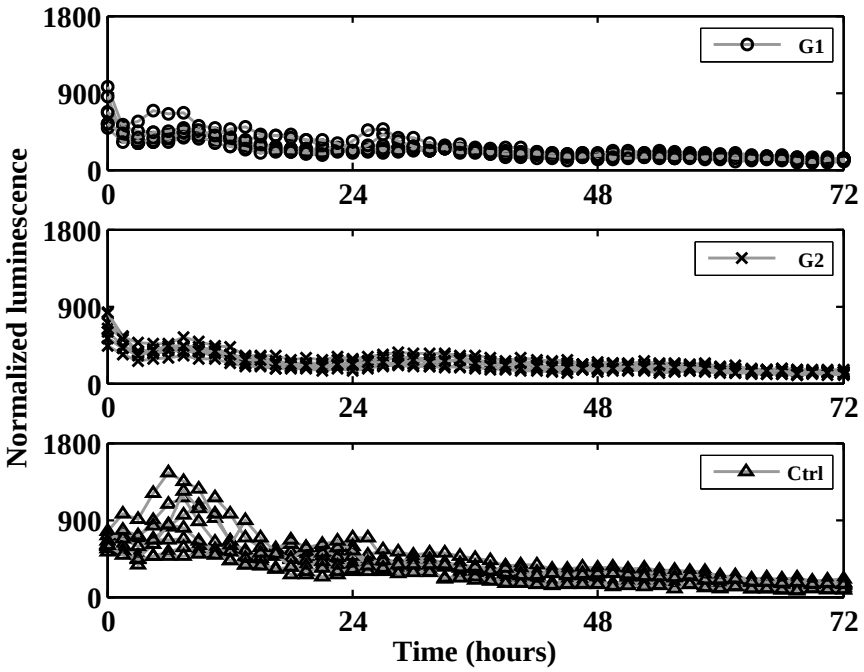

Fig. 9. Human Gene Luminescence Data. Eight individual replicates of PER2:LUC expression for two treatments, G1 and G2, and control (Ctrl). Markers correspond to positions of observed values.

Chatfield, Chris. (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, Russell and MacKinnon, James G. (1998). Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School* **66**, 1–26.

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

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

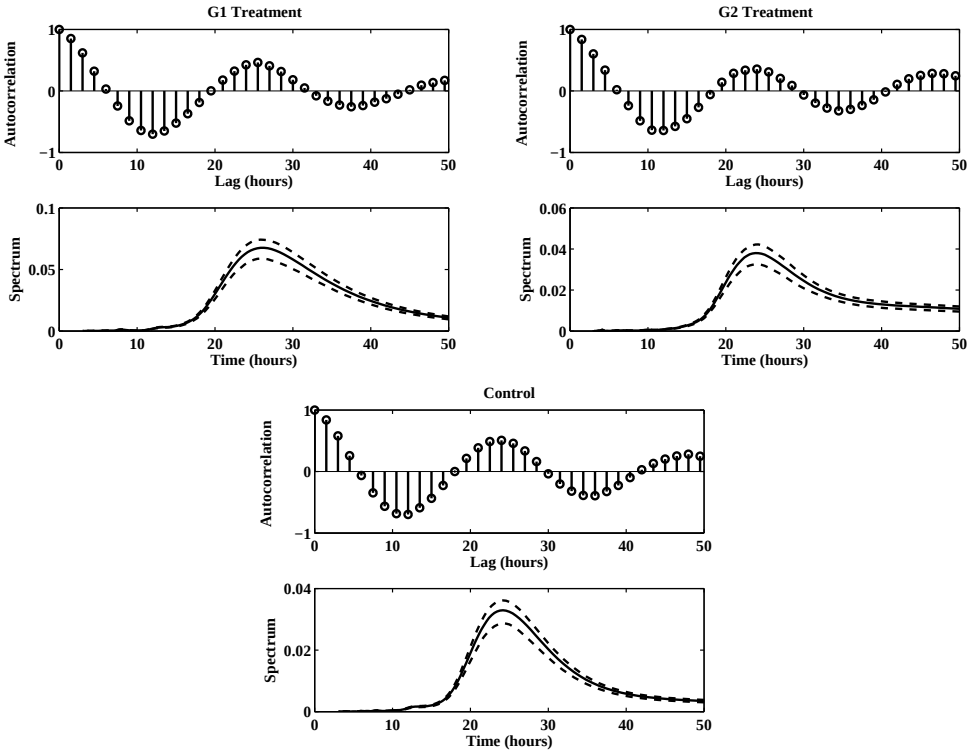

Fig. 10. Human Gene Luminescence Data. Autocorrelation functions and estimated distribution of spectrum functions using the SR method for the average PER2:LUC luminescence profile in different treatment regimes. In bottom panels, the solid line represents the mean of the bootstrap sample of spectrum functions, and the dashed lines represent 95% pointwise confidence intervals.

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

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

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

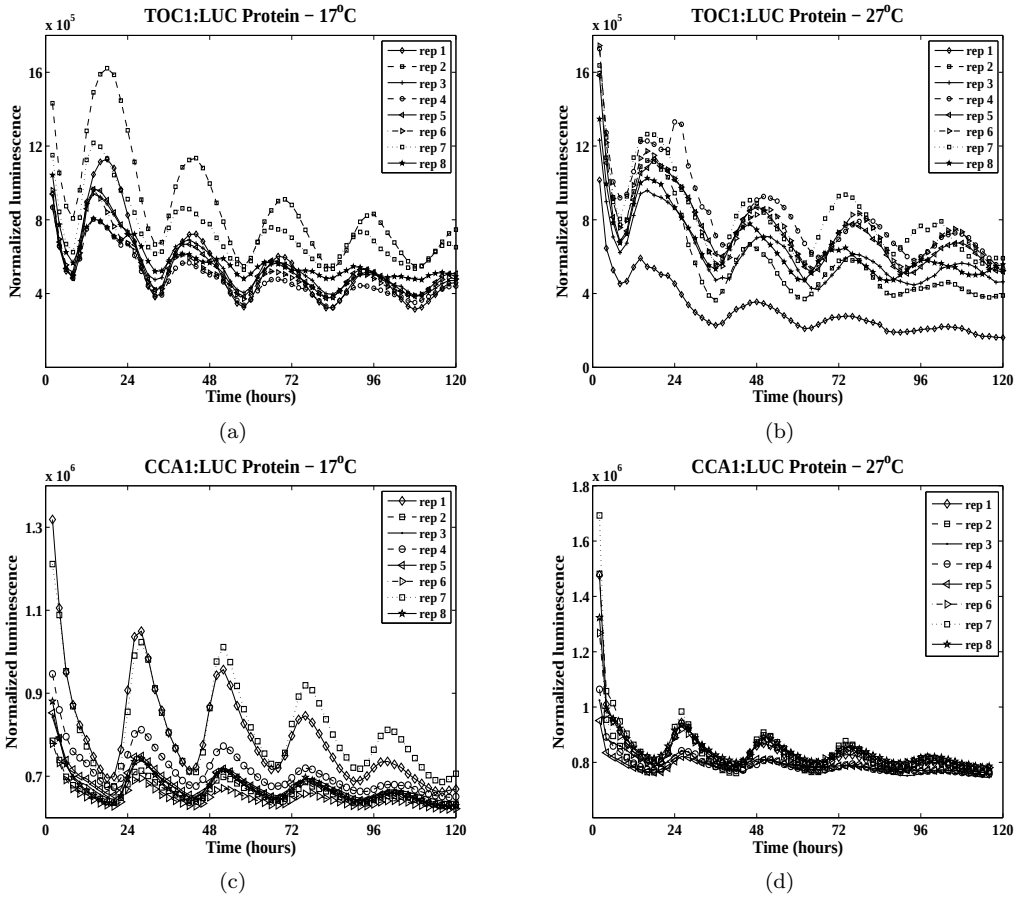

Fig. 11. Plant Luminescence Data. (a) Eight replicate time series of normalized luminescence of TOC1:LUC protein under constant red light and constant temperature of 17°C. (b) Eight replicate time series of normalized luminescence of TOC1:LUC protein under constant red light and constant temperature of 27°C. (c) Eight replicate time series of normalized luminescence of CCA1:LUC protein under constant red light and constant temperature of 17°C. (d) Eight replicate time series of normalized luminescence of CCA1:LUC protein under constant red light and constant temperature of 27°C. Data sampled ZT 2-120, except for CCA1:LUC at 27°C, where ZT 2-114. Replicate j is denoted by 'rep j', $j = 1, \ldots, 8$. Markers represent actual measurements taken every two hours.

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

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

Parzen, Emanuel. (1962). On estimation of a probability density function and mode. *The*

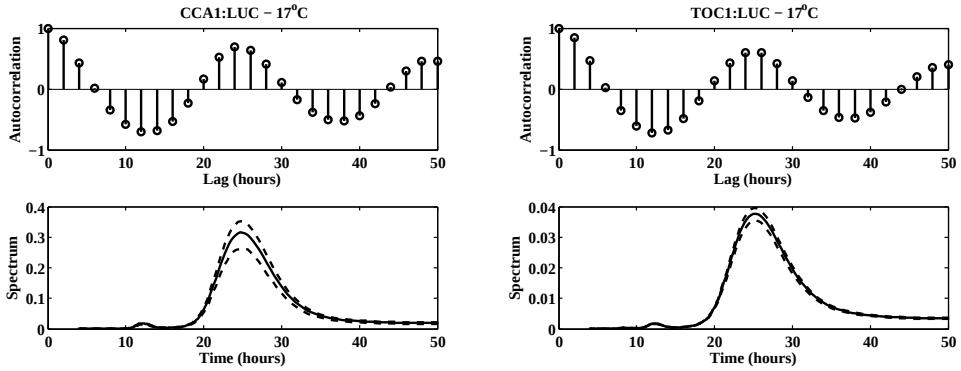

Fig. 12. Plant Luminescence Data. Autocorrelation functions and estimated distribution of spectrum functions using the SR method for the averaged CCA1:LUC and TOC1:LUC luminescence profiles at 17°C. In bottom panels, the solid line represents the mean of the bootstrap sample of spectrum functions, and the dashed lines represent 95% pointwise confidence intervals.

Annals of Mathematical Statistics **33**, 1065–1076.

Plautz, Jeffrey D., Straume, Martin, Stanewsky, Ralf, Jamison, Creston F., Brandes, Christian, Dowse, Harold B., Hall, Jeffrey C. and Kay, Steve A. (1997). Quantitative analysis of drosophila period gene transcription in living animals. *Journal of Biological Rhythms* **12**, 204–217.

Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *JASA* **69**, 730–737.

Stoica, Petre and Sundin, Tomas. (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, Topics in Fluorescence Spectroscopy, Volume 2: Principles. Plenum, New York, pp. 117–240.