
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Olsper, N.; Pelt, J.; Käpylä, M. J.; Lehtinen, J. J.

Estimating activity cycles with probabilistic methods I. Bayesian Generalised Lomb-Scargle Periodogram with Trend

Published in:
Astronomy and Astrophysics

DOI:
[10.1051/0004-6361/201732524](https://doi.org/10.1051/0004-6361/201732524)

Published: 01/07/2018

Please cite the original version:

Olsper, N., Pelt, J., Käpylä, M. J., & Lehtinen, J. J. (2018). Estimating activity cycles with probabilistic methods I. Bayesian Generalised Lomb-Scargle Periodogram with Trend. *Astronomy and Astrophysics*, 615, 1-13. [A111]. <https://doi.org/10.1051/0004-6361/201732524>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Estimating activity cycles with probabilistic methods I. Bayesian Generalised Lomb-Scargle Periodogram with Trend

N. Olsper¹, J. Pelt², M. J. Käpylä^{3,1}, and J. J. Lehtinen^{3,1}

¹ ReSoLVE Centre of Excellence, Aalto University, Department of Computer Science, PO Box 15400, FI-00076 Aalto, Finland

² Tartu Observatory, 61602 Tõravere, Estonia

³ Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, D-37077 Göttingen, Germany

Received / Accepted

ABSTRACT

Context. Period estimation is one of the central topics in astronomical time series analysis, where data is often unevenly sampled. Especially challenging are studies of stellar magnetic cycles, as there the periods looked for are of the order of the same length than the datasets themselves. The datasets often contain trends, the origin of which is either a real long-term cycle or an instrumental effect, but these effects cannot be reliably separated, while they can lead to erroneous period determinations if not properly handled.

Aims. In this study we aim at developing a method that can handle the trends properly, and by performing extensive set of testing, we show that this is the optimal procedure when contrasted with methods that do not include the trend directly to the model. The effect of the noise model on the results is also investigated.

Methods. We introduce a Bayesian Generalised Lomb-Scargle Periodogram with Trend (BGLST), which is a probabilistic linear regression model using Gaussian priors for the coefficients and uniform prior for the frequency parameter.

Results. We show, using synthetic data, that when there is no prior information on whether and to what extent the true model of the data contains a linear trend, the introduced BGLST method is preferable to the methods which either detrend the data or leave the data untrended before fitting the periodic model. Whether to use different from constant noise model depends on the density of the data sampling as well as on the true noise model of the process.

Key words. methods: statistical, methods: numerical, stars: activity

1. Introduction

In the domain of astronomical data analysis the task of period estimation from unevenly spaced time series has been relevant topic for many decades. Depending on the context, the term “period” can refer to e.g. rotational period of the star, orbiting period of the star or exoplanet, or the period of the activity cycle of the star. Knowing the precise value of the period is often very important as many other physical quantities are dependent on it. For instance, in the case of active late-type stars with outer convection zones, the ratio between rotational period and cycle period can be interpreted as a measure of the dynamo efficiency. Therefore, measuring this ratio gives us crucial information of the generation of magnetic fields in stars with varying activity levels. In practice, both periods are usually estimated through photometry and/or spectrometry of the star.

For the purpose of analysing unevenly spaced time series, many different methods have been developed over the years. Historically one of the most used ones is the Lomb-Scargle (LS) periodogram (Lomb 1976; Scargle 1982). Statistical properties of the LS and other alternative periodograms have been extensively studied and it has become a well-known fact that the interpretation of the results of any spectral analysis method takes a lot of effort.

The sampling patterns in the data together with the finiteness of the time span of observations lead to multitude of difficulties in the period estimation. One of the most pronounced difficulties is the emergence of aliased peaks (spectral leakage) (Deeming 1975; Lomb 1976; Scargle 1982; Horne & Baliunas 1986). In other words one can only see a distorted spectrum as the observations are made at discrete (uneven) time moments and during a finite time. However, algorithms for eliminating spurious peaks from spectrum have been developed (Roberts et al. 1987). Clearly, different period estimation methods tend to perform differently depending on the dataset. For comparison of some of the popular methods see Carbonell et al. (1992).

One of the other issues in spectral estimation arises when the true mean of the measured signal is not known. The LS method assumes a zero-mean harmonic model with constant noise variance, so that the data needs to be centered before doing the analysis. While every dataset can be turned into a dataset with a zero mean, in some cases, due to pathological sampling (if the empirical mean¹ and the true mean differ significantly) this procedure can lead to incorrect period estimates. In the literature this problem has been addressed extensively and several generalizations of the periodogram, which are invariant to the shifts in the

¹ We use the terms “empirical mean” and “empirical trend” when referring to the values obtained by correspondingly fitting a constant or a line to the data using linear regression.

Send offprint requests to: N. Olsper
e-mail: nigul.olsper@aalto.fi

target (output) value, have been proposed. Such include the method of Ferraz-Mello (1981), who used Gram-Schmidt orthogonalization of the constant, cosine and sine functions in the sample domain. The power spectrum is then defined as the square norm of the data projections to these functions. In Zechmeister & Kürster (2009) the harmonic model of the LS periodogram is directly extended with the addition of a constant term, which has become known as the Generalised Lomb-Scargle (GLS) periodogram. Moreover, both of these studies give the formulations allowing non-constant noise variance. Later, using a Bayesian approach for a model with harmonic plus a constant, it has been shown that the posterior probability of the frequency, when using uniform priors, is very similar to the GLS spectrum (Mortier et al. 2015). The benefit of the latter method is that the relative probabilities of any two frequencies can be easily calculated. Usually in the models, likewise in the current study, the noise is assumed to be Gaussian and uncorrelated. Other than white noise models are discussed in Vio et al. (2010). For more thorough overview of important aspects in spectral analysis please refer to VanderPlas (2017).

The focus of the current paper is on another yet unaddressed issue, namely the effect of linear trend in the data to period estimation. The motivation to tackle this question arose in the context of analysing the Mount Wilson time series of chromospheric activity (hereafter MW) in a quest to look for long-term activity cycles (Olsper et al. 2017), the length of which is of the same order of magnitude than the data set length itself. In a previous study by (Baliunas et al. 1995) detrending was occasionally, but not systematically, used before the LS periodogram was calculated. The cycle estimates from this study have been later used extensively by many other studies, for example, to show the presence of different stellar activity branches (e.g. Saar & Brandenburg 1999).

We now raise the following question – whether it is more optimal to remove the linear trend before the period search or leave the data untrended? Likewise with centering, detrending the data before fitting the harmonic model may or may not lead to biased period estimates depending on the structure of the data. The problems arise either due to sampling effects or due to presence of very long periods in the data. Based on the empirical arguments given in Sect. 3, we show that it is more preferable to include the trend component directly into the regression model instead of detrending the data *a priori* or leaving the data untrended altogether. In the same section we also discuss the effects of noise model of the data to the period estimate. We note that the question that we address in the current study is primarily relevant to the cases where the number of cycles in the dataset is small. Otherwise, the long coherence time (if the underlying process is truly periodic) allows to nail down periods, even if there are uneven errors or small a trend. High number of cycles allows to get exact periods even e.g. by cycle counting.

The generalised least-squares spectrum allowing arbitrary components (including linear trend) was first discussed in Vaníček (1971) and more recently a Bayesian approach allowing several harmonics as well as trend components was introduced in Ford et al. (2011).

2. Method

Now we turn to the description of the present method, which is a generalization of the method proposed in Mortier et al. (2015). We introduce a simple Bayesian linear regression model where besides the harmonic component we have an offset plus a trend². This is summarised in the following equation:

$$y(t_i) = A \cos(2\pi f t_i - \phi) + B \sin(2\pi f t_i - \phi) + \alpha t_i + \beta + \epsilon(t_i), \quad (1)$$

where $y(t_i)$ and $\epsilon(t_i)$ are the observation and noise at time moment t_i , $f = 1/P$ is the frequency of the cycle, A , B , α , β and ϕ are free parameters. Specifically, α is the slope and β the offset. We assume that the noise is Gaussian and independent between any two time moments, but we do not assume the equality of variances. For parameter inference we use a Bayesian model, where the posterior probability is given by

$$p(f, \theta | D) \propto p(D | f, \theta) p(f, \theta), \quad (2)$$

where $p(D | f, \theta)$ is the likelihood of the data, $p(f, \theta)$ is the prior probability of the parameters and $\theta = [A, B, \alpha, \beta]^T$ is the vector of nuisance parameters. Parameter ϕ is not optimized, but set to a frequency dependent value simplifying the inference such that cross terms with cos and sin components will vanish (see Sect. 2.1). The likelihood is given by

$$p(D | f, \theta) = \left(\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_i}} \right) \exp \left(-\frac{1}{2} \sum_{i=1}^N \frac{\epsilon_i^2}{\sigma_i^2} \right), \quad (3)$$

where $\epsilon_i = \epsilon(t_i)$ and σ_i^2 is the noise variance at time moment t_i . To make the derivation of the spectrum analytically tractable we take independent Gaussian priors for A, B, α, β and a flat prior for the frequency f . This leads to the prior probability given by

$$p(f, \theta) = \mathcal{N}(\theta | \mu_\theta, \Sigma_\theta), \quad (4)$$

where $\mu_\theta = [\mu_A, \mu_B, \mu_\alpha, \mu_\beta]^T$ is the vector of prior means and $\Sigma_\theta = \text{diag}(\sigma_A^2, \sigma_B^2, \sigma_\alpha^2, \sigma_\beta^2)$ is the diagonal matrix of prior variances.

The larger the prior variances the less information is assumed to be known about the parameters. Based on what is intuitively meaningful, in all the calculations we have used the following values for the prior means and variances:

$$\begin{aligned} \mu_A = 0, \mu_B = 0, \mu_\alpha = \alpha_{\text{slope}}, \mu_\beta = \beta_{\text{intercept}}, \\ \sigma_A^2 = 0.5\sigma_y^2, \sigma_B^2 = 0.5\sigma_y^2, \sigma_\alpha^2 = \frac{\sigma_y^2}{\Delta T^2}, \sigma_\beta^2 = \sigma_y^2 + \beta_{\text{intercept}}^2, \end{aligned} \quad (5)$$

where α_{slope} and $\beta_{\text{intercept}}$ are the slope and intercept estimated from linear regression, σ_y^2 is the sample variance of the data and ΔT is duration of the data. If one does not have any prior information about the parameters one could set the variances to infinity and drop the corresponding terms from the equations, but in practice to avoid meaningless results with unreasonably large parameter values some regularization would be required.

² The implementation of the method introduced in this paper can be found at <https://github.com/olsper/BGLST>

2.1. Derivation of the spectrum

Our derivation of the model closely follows the key points presented in Mortier et al. (2015). Consequently, we use as identical notion of the variables as possible. Using the likelihood and prior defined by Eqs. (3) and (4), the posterior probability of the Bayesian model given by Eq. (2) can be explicitly written as

$$\begin{aligned} p(f, \boldsymbol{\theta}|D) &\propto p(D|f, \boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &= p(D|f, \boldsymbol{\theta})\mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta) \\ &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_i} \prod_{i=1}^k \frac{1}{\sqrt{2\pi}\sigma_{\theta_i}} e^{-\frac{1}{2}E}, \end{aligned} \quad (6)$$

where

$$E = \sum_{i=1}^N \frac{\epsilon_i^2}{\sigma_i^2} + \sum_{i=1}^k \frac{(\theta_i - \mu_{\theta_i})^2}{\sigma_{\theta_i}^2}. \quad (7)$$

Here θ_i , $i = 1, \dots, k$, $k = 4$, denotes the i th element of $\boldsymbol{\theta}$, i.e. either A , B , α , or β . From Eqs. (1) and (3) we see that the likelihood and therefore posterior probability for every fixed frequency f is multivariate Gaussian w.r.t. the parameters A , B , α and β . In principle we are interested finding the optimum for the full joint posterior probability density $p(f, \boldsymbol{\theta}|D)$, but as for every fixed frequency the latter distribution is a multivariate Gaussian we can first marginalise over all nuisance parameters, find the optimum for the frequency f e.g. by doing grid search and later analytically find the posterior means and covariances of the other parameters. The marginal distribution for the frequency parameter is expressed as

$$p(f|D) \propto \int p(D|f, \boldsymbol{\theta})\mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)d\boldsymbol{\theta}. \quad (8)$$

Here the integrals are assumed to be taken over the whole range of parameter values, i.e. from $-\infty$ to ∞ . We introduce the following notations:

$$w_i = \frac{1}{\sigma_i^2}, w_A = \frac{1}{\sigma_A^2}, w_B = \frac{1}{\sigma_B^2}, w_\alpha = \frac{1}{\sigma_\alpha^2}, w_\beta = \frac{1}{\sigma_\beta^2}, \quad (9)$$

$$W = \sum_{i=1}^N w_i + w_\beta, Y = \sum_{i=1}^N w_i y_i + w_\beta \mu_\beta, \quad (10)$$

$$C = \sum_{i=1}^N w_i \cos(2\pi f t_i - \phi), \quad (11)$$

$$S = \sum_{i=1}^N w_i \sin(2\pi f t_i - \phi), \quad (12)$$

$$T = \sum_{i=1}^N w_i t_i, \quad (13)$$

$$\widehat{YC} = \sum_{i=1}^N w_i y_i \cos(2\pi f t_i - \phi) + w_A \mu_A, \quad (14)$$

$$\widehat{YS} = \sum_{i=1}^N w_i y_i \sin(2\pi f t_i - \phi) + w_B \mu_B, \quad (15)$$

$$\widehat{CC} = \sum_{i=1}^N w_i \cos^2(2\pi f t_i - \phi) + w_A, \quad (16)$$

$$\widehat{SS} = \sum_{i=1}^N w_i \sin^2(2\pi f t_i - \phi) + w_B, \quad (17)$$

$$\widehat{TT} = \sum_{i=1}^N w_i t_i^2 + w_\alpha, \quad (18)$$

$$\widehat{YT} = \sum_{i=1}^N w_i y_i t_i + w_\alpha \mu_\alpha, \quad (19)$$

$$\widehat{TC} = \sum_{i=1}^N w_i t_i \cos(2\pi f t_i - \phi), \quad (20)$$

$$\widehat{TS} = \sum_{i=1}^N w_i t_i \sin(2\pi f t_i - \phi), \quad (21)$$

$$\widehat{YY} = \sum_{i=1}^N w_i y_i^2 + w_A \mu_A^2 + w_B \mu_B^2 + w_\alpha \mu_\alpha^2 + w_\beta \mu_\beta^2. \quad (22)$$

If the value of ϕ is defined such that the sine and cosine functions are orthogonal (for the proof see Mortier et al. 2015), namely

$$\phi = \frac{1}{2} \arctan \left[\frac{\sum_{i=1}^N w_i \sin(4\pi f t_i)}{\sum_{i=1}^N w_i \cos(4\pi f t_i)} \right], \quad (23)$$

then we have

$$\begin{aligned} E &= \widehat{CC}A^2 - 2\widehat{YCA} + 2\alpha\widehat{TCA} + 2\beta CA \\ &\quad + \widehat{SS}B^2 - 2\widehat{YSB} + 2\alpha\widehat{TSB} + 2\beta SB \\ &\quad + \widehat{TT}\alpha^2 - 2\widehat{YT}\alpha + 2T\beta\alpha + W\beta^2 - 2Y\beta + \widehat{YY}, \end{aligned} \quad (24)$$

where we have grouped the terms with A on the first line, terms with B on the second line and the rest of the terms on the third line.

In the following we will repeatedly use the formula of the following definite integral:

$$\int_{-\infty}^{\infty} e^{-ax^2 - 2bx} dx = \sqrt{\frac{\pi}{a}} e^{\frac{b^2}{a}}, \quad a > 0. \quad (25)$$

To calculate the integral in Eq. (8), we start by integrating first over A and B . Assuming that \widehat{CC} and \widehat{SS} are greater than zero and applying Eq. (25), we will get the solution for the integral containing terms with A :

$$\begin{aligned} &\int_{-\infty}^{\infty} \exp \left(-\frac{\widehat{CC}A^2 - 2\widehat{YCA} + 2\alpha\widehat{TCA} + 2\beta CA}{2} \right) dA \\ &= \sqrt{\frac{2\pi}{\widehat{CC}}} \exp \left(\frac{(\widehat{YC} - \alpha\widehat{TC} - \beta C)^2}{2\widehat{CC}} \right) \\ &= \exp \left(\frac{-\alpha\widehat{YCTC}}{\widehat{CC}} + \frac{\alpha^2\widehat{TC}^2}{2\widehat{CC}} + \frac{\alpha\beta\widehat{TCC}}{\widehat{CC}} \right) \\ &\quad \cdot \exp \left(\frac{-\beta\widehat{YCC}}{\widehat{CC}} + \frac{\beta^2 C^2}{2\widehat{CC}} \right) \\ &\quad \cdot \sqrt{\frac{2\pi}{\widehat{CC}}} \exp \left(\frac{\widehat{YC}^2}{2\widehat{CC}} \right). \end{aligned}$$

(26) We similarly introduce new variable Q for the coefficients involving all terms with β from the same equations:

$$Q = -\frac{\widehat{YCC}}{\widehat{CC}} - \frac{\widehat{YSS}}{\widehat{SS}} + Y - \frac{2MN}{4K}. \quad (34)$$

With these substitutions we are ready to integrate over β using again Eq. (25) while assuming $P < 0$:

$$\int_{-\infty}^{\infty} \exp(P\beta^2 + Q\beta) = \sqrt{\frac{\pi}{-P}} \exp\left(\frac{Q^2}{-4P}\right). \quad (35)$$

After gathering all remaining constant terms from Eqs. (24), (26), (27), (30), we finally obtain:

$$p(f|D) \propto \frac{2\pi^2}{\sqrt{(\widehat{CC}\widehat{SS}KP)}} \cdot \exp\left(\frac{\widehat{YC}^2}{2\widehat{CC}} + \frac{\widehat{YS}^2}{2\widehat{SS}} - \frac{M^2}{4K} - \frac{Q^2}{4P} - \frac{\widehat{YY}}{2}\right). \quad (36)$$

For the purpose of fitting the regression curve to the data one would be interested in obtaining the optimal values for the nuisance parameters A, B, α, β . This can be easily done after fixing the frequency to its optimal value using Eq. (36) with grid search, and noticing that $p(\boldsymbol{\theta}|D, f_{\text{opt}})$ is a multivariate Gaussian distribution. In the following we list the corresponding posterior means of the parameters³:

$$\mu_{\beta} = -\frac{Q}{2P}, \quad (37)$$

$$\mu_{\alpha} = -\frac{N\mu_{\beta} + M}{2K}, \quad (38)$$

$$\mu_A = \frac{\widehat{YC} - \widehat{TC}\mu_{\alpha} - C\mu_{\beta}}{\widehat{CC}}, \quad (39)$$

$$\mu_B = \frac{\widehat{YS} - \widehat{TS}\mu_{\alpha} - S\mu_{\beta}}{\widehat{SS}}. \quad (40)$$

Formula for the full covariance matrix for the parameters, assuming a model with constant noise variance, can be found in Murphy (2012, Chapter 7.6.1).

If we now consider the case when $\widehat{SS} = 0$ (for more details see Mortier et al. 2015), we see that also $S = 0$, $\widehat{YS} = 0$ and $\widehat{TS} = 0$. Consequently the integral for B is proportional to a constant. We can define the analogues of

³ This is similar to Empirical Bayes approach as we use the point estimate for f .

In the last expression we have grouped onto separate lines the terms with α , terms with β not simultaneously containing α , and constant terms. Similarly for the integral containing terms with B :

$$\begin{aligned} \int_{-\infty}^{\infty} \exp\left(-\frac{\widehat{SSB}^2 - 2\widehat{YSB} + 2\alpha\widehat{TSB} + 2\beta SB}{2}\right) dB \\ = \sqrt{\frac{2\pi}{\widehat{SS}}} \exp\left(\frac{(\widehat{YS} - \alpha\widehat{TS} - \beta S)^2}{2\widehat{SS}}\right) \\ = \exp\left(\frac{-\alpha\widehat{YSTS}}{\widehat{SS}} + \frac{\alpha^2\widehat{TS}^2}{2\widehat{SS}} + \frac{\alpha\beta\widehat{TSS}}{\widehat{SS}}\right) \\ \cdot \exp\left(\frac{-\beta\widehat{YSS}}{\widehat{SS}} + \frac{\beta^2 S^2}{2\widehat{SS}}\right) \\ \cdot \sqrt{\frac{2\pi}{\widehat{SS}}} \exp\left(\frac{\widehat{YS}^2}{2\widehat{SS}}\right). \end{aligned} \quad (27)$$

Now we gather the coefficients for all the terms with α^2 from the last line of Eq. (24) (keeping in mind the factor $-1/2$) as well as from Eqs. (26), (27) into new variable K :

$$K = \frac{1}{2} \left(-\widehat{TT} + \frac{\widehat{TC}^2}{\widehat{CC}} + \frac{\widehat{TS}^2}{\widehat{SS}} \right). \quad (28)$$

We similarly introduce new variable L for the coefficients involving all terms with α from the same equations:

$$L = \left(\widehat{YT} - \beta T + \frac{-\widehat{YCTC} + \beta\widehat{TC}C}{\widehat{CC}} + \frac{-\widehat{YSTS} + \beta\widehat{TSS}}{\widehat{SS}} \right) \quad (29)$$

After these substitutions, integrating over α can again be accomplished with the help of Eq. (25) and assuming $K < 0$:

$$\begin{aligned} \int_{-\infty}^{\infty} \exp(K\alpha^2 + L\alpha) = \sqrt{\frac{\pi}{-K}} \exp\left(\frac{L^2}{-4K}\right) \\ = \sqrt{\frac{\pi}{-K}} \exp\left(\frac{(M + N\beta)^2}{-4K}\right) \\ = \sqrt{\frac{\pi}{-K}} \exp\left(\frac{N^2\beta^2}{-4K}\right) \exp\left(\frac{2MN\beta}{-4K}\right) \exp\left(\frac{M^2}{-4K}\right), \end{aligned} \quad (30)$$

where

$$M = \widehat{YT} - \frac{\widehat{YCTC}}{\widehat{CC}} - \frac{\widehat{YSTS}}{\widehat{SS}} \quad (31)$$

and

$$N = \frac{\widehat{TC}C}{\widehat{CC}} + \frac{\widehat{TSS}}{\widehat{SS}} - T. \quad (32)$$

At this point what is left to do is the integration over β . To simplify things once more, we gather the coefficients for all the terms with β^2 from Eqs. (24), (26), (27), (30) into new variable P :

$$P = \frac{C^2}{2\widehat{CC}} + \frac{S^2}{2\widehat{SS}} - \frac{W}{2} - \frac{N^2}{4K}. \quad (33)$$

the constants K through Q for this special case as

$$K_C = \frac{1}{2} \left(-\widehat{TT} + \frac{\widehat{TC}^2}{\widehat{CC}} \right), \quad (41)$$

$$L_C = \left(\widehat{YT} - \beta T + \frac{-\widehat{YCTC} + \beta \widehat{TC}^2}{\widehat{CC}} \right), \quad (42)$$

$$M_C = \widehat{YT} - \frac{\widehat{YCTC}}{\widehat{CC}}, \quad (43)$$

$$N_C = \frac{\widehat{TC}^2}{\widehat{CC}} - T, \quad (44)$$

$$P_C = \frac{C^2}{2\widehat{CC}} - \frac{W}{2} - \frac{N_C^2}{4K_C}, \quad (45)$$

$$Q_C = -\frac{\widehat{YCC}}{\widehat{CC}} + Y - \frac{2M_C N_C}{4K_C}. \quad (46)$$

Finally, we arrive at the expression for the data likelihood

$$p(f|D) \propto \frac{2\pi^2}{\sqrt{(\widehat{CC}K_C P_C)}} \cdot \exp \left(\frac{\widehat{YC}^2}{2\widehat{CC}} - \frac{M_C^2}{4K_C} - \frac{Q_C^2}{4P_C} - \frac{\widehat{YY}}{2} \right). \quad (47)$$

Similarly we can handle the situation when $\widehat{CC} = 0$. In the derivation we also assumed that $K < 0$ and $P < 0$. Our experiments with test data showed that the condition for $K < 0$ was always satisfied, but occasionally P obtained values zero and probably due to numerical rounding errors also very low positive values. These special cases we handled by dropping the corresponding terms from Eqs. (36) and (47). Theoretically confirming or disproving that the conditions $K < 0$ and $P \leq 0$ always hold is however out of the scope of current study.

2.2. Dealing with multiple harmonics

The formula for the spectrum is given by Eq. (36), which represents the posterior probability of the frequency f , given data D and our harmonic model M_H i.e. $p(f|D, M_H)$. Although being a probability density, it is still convenient to call this frequency-dependent quantity a spectrum. We point out, that if the true model matches the given model, M_H , then the interpretation of the spectrum is straightforward, namely being the probability distribution of the frequency. This will give us a direct way for error estimation, e.g. by fitting a Gaussian to the spectral line (Bretthorst 1988, Chapter 2). When the true model has more than one harmonic, the interpretation of the spectrum is not anymore direct due to mixing of the probabilities from different harmonics. The correct way to address this issue would be to use more complex model with at least as many harmonics that there are expected to be in the underlying process (or even better, to infer the number of components from data). However, as a simpler workaround, the ideas of cleaning the spectrum introduced in Roberts et al. (1987) can be used to iteratively extract significant frequencies from the spectrum calculated using a model with single harmonic.

2.3. Significance estimation

To estimate the significance of the peaks in the spectrum we perform a model comparison between the given model and a model without harmonics, i.e. only with offset and trend. One practical way to do this is to calculate

$$\Delta \text{BIC} = \text{BIC}_{M_{\text{null}}} - \text{BIC}_{M_H}, \quad (48)$$

where M_{null} is the linear model without harmonic, $\text{BIC} = \ln(n)k - 2\ln(\hat{L})$ is the Bayesian Information Criterion (BIC), n is the number of data points, k the number of model parameters and $\hat{L} = p(x|\hat{\theta}, M)$ is the likelihood of data for model M using the parameter values that maximize the likelihood. This formula is an approximation to the logarithmic Bayes factor, more precisely $\Delta \text{BIC} \approx 2 \ln K$, where K is the Bayes factor. Strength of evidence for models with $2 \ln K > 10$ are considered very strong, with $6 \leq 2 \ln K < 10$, strong and with $2 \leq 2 \ln K < 6$ positive. For a harmonic model M_H , θ include optimal frequency, and coefficients of the harmonic component, trend and offset, for M_{null} only the last two coefficients.

In the derivation of the spectrum we assumed that the noise variance of the data points is known, which is usually not the case. In probabilistic models, this parameter should also be optimized, however in practice often sample variance is taken as the estimate for it. This is also the case with LS periodogram. However, it has been shown that when normalizing the periodogram with the sample variance the statistical distribution slightly differs from the theoretically expected one (Schwarzenberg-Czerny 1998). More realistic approach, especially in the case when the noise cannot be assumed stationary is to use e.g. subsample variances in a small sliding window around the data points.

As a final remark in this section, we want to emphasize that the probabilistic approach does not remove the burden of dealing with spectral aliases due to data sampling. There is always a chance of false detection, thus the interpretation of the spectrum must be done with care.

3. Experiments

In this section we undertake some experiments to compare the performance of the introduced method with LS and/or GLS periodograms. As the effect of an offset in the randomly sampled data to the period estimate has been well described in Mortier et al. (2015), then we skip the discussion of similar examples. Instead, we will discuss the effects of a linear trend in the data and a nonconstant noise variance to the period estimate, as we find them to be less addressed topics in the literature.

3.1. Effect of a linear trend

Let us start with the question of how much the presence of a linear trend in the data affects the period estimate. To measure the performance of the BGLST method introduced in Sect. 2, we do the comparison to plain GLS and GLS with preceding linear detrending (GLS-T). Throughout this section we assume a constant noise variance. We draw the data with time span $\Delta T \approx 30$ time units randomly from the process with following harmonic covariance function:

$$\text{cov}(t_i, t_j) = \sigma_S \cos(2\pi f(t_i - t_j)) + \sigma_T t_i t_j + \delta_{ij} \sigma_N^2, \quad (49)$$

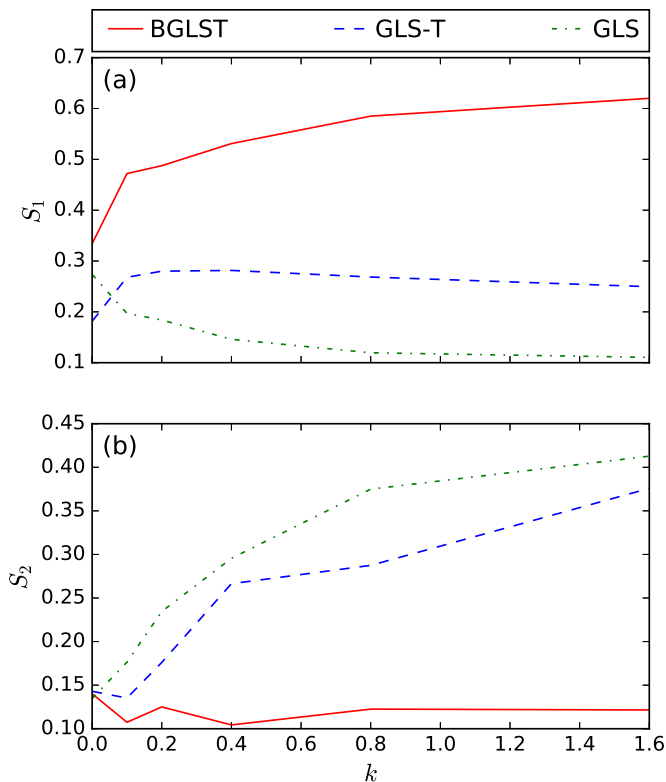


Fig. 1. Results of the experiments with varying linear trend using uniform sampling. (a) Performance measure S_1 as function of k . (b) The same for performance measure S_2 .

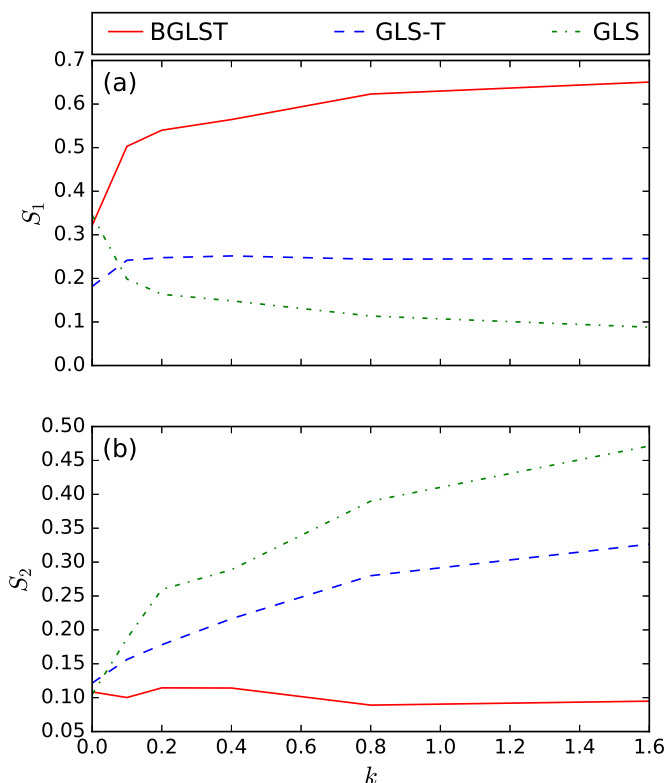


Fig. 2. Results of the experiments with varying linear trend using sampling similar to MW datasets. (a) Performance measure S_1 as function of k . (b) The same for performance measure S_2 .

where σ_S , σ_T , and σ_N are the signal, trend, and noise variances, respectively, and δ_{ij} is a Kronecker delta. We varied the trend variance $\sigma_T = k(\sigma_S + \sigma_N)/\Delta T$, such that $k \in [0, 0.1, 0.2, 0.4, 0.8, 1.6]$. For each k we generated $N = 2000$ time series, where each time the mean of the signal was drawn uniformly from interval $[-1, 1]$, S/N ratio from $[0.2, 0.8]$ and period $P = 1/f$ from $[2, 2/3\Delta T]$. We repeated the experiments with two forms of sampling: uniform and the one based on the samplings of MW datasets. The prior means and variances for our model have been chosen according to Eq. (5).

In these experiments we compare three different methods – the BGLST, GLS-T and GLS. We measure the performance of each method using the following two statistics:

$$S_1 = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\Delta_i < \Delta_i^A) \mathbb{I}(\Delta_i < \Delta_i^B), \quad (50)$$

where \mathbb{I} is the indicator function, $\Delta_i = |f_i - f_i^{\text{true}}|$ is the absolute error of the period estimate in the i -th experiment using given method (either BGLST, GLS-T or GLS) and Δ_i^A and Δ_i^B are the corresponding values for the other two methods. The second performance statistic we measure is the average relative error of the estimates:

$$S_2 = \frac{1}{N} \sum_{i=1}^N \delta_i, \quad (51)$$

where $\delta_i = \Delta_i / f^{\text{true}}$.

The results for the experiments with uniform sampling are shown in Fig. 1 and for sampling taken from MW dataset in Fig. 2. From panels (a) of the figures we see that the fraction of number of experiments where the estimate of BGLST method is better than the estimate of the other two methods is increasing with trend variance. When there is no trend ($k = 0$) our method coincides with the GLS method as expected, but as we see, detrending has negative effect in that case. When the true trend increases then GLS-T starts to outperform the GLS, but both are far from the BGLST method. From panels (b) we see that when there is no trend in the dataset at all, all the three methods have the same average relative error and while for our method it stays the same or slightly decreases with increasing k , for both other methods the errors start to increase rapidly.

3.1.1. Special cases

Next we illustrate the benefit of using BGLST model with two examples, where the differences to the other models are the well emphasized. For that purpose we first generated such a dataset where the empirical slope significantly differs from the true slope. The test contained only one harmonic with a frequency of 0.014038 and a slope 0.01. We again fit three models – BGLST, GLS and GLS-T. The comparison of the results are shown in Fig. 3. As is evident from Fig. 3(a), the empirical trend 0.0053 recovered by the GLS-T method (blue dashed line) differs significantly from the true trend (black dotted line). Both the GLS and GLS-T methods perform very poorly in fitting the data, while only the BGLST model produces a fit that represents the data points adequately. Moreover, BGLST recovers very close to the true trend value 0.009781. From Fig. 3(b) it is evident that the BGLST model retrieves a frequency closest to the

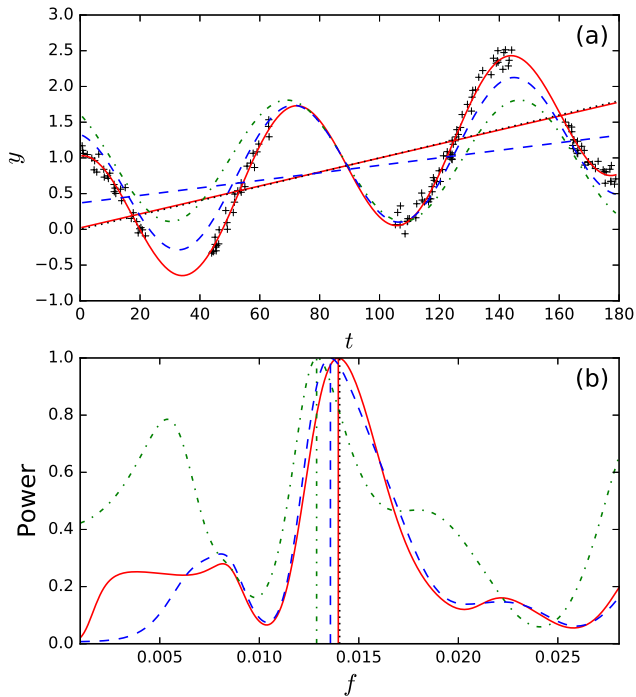


Fig. 3. Comparison of the results using different models. The true model of the data contains one harmonic, trend and additive white Gaussian noise. (a) Data (black crosses), BGLST model (red continuous curve), GLS-T model (blue dashed curve), GLS model (green dash-dotted curve), true trend (black dotted line), trend from BGLST model (red continuous line), empirical trend (blue dashed line). (b) Spectra of the corresponding models with vertical lines marking the locations of maxima. The black dotted line shows the position of the true frequency.

true one, while the other two methods return too low values for it. The performance of the GLS model is the worst, as expected. The frequency estimates for BGLST, GLS-T and GLS are correspondingly: 0.013969, 0.013573 and 0.01288. This simple example shows that when there is a real trend in the data, due to the sampling patterns, detrending can be erroneous, but it still leads to better results than not doing the detrending at all.

Finally we would like to show a counterexample where detrending is not the preferred option. This happens when there is a harmonic signal in the data with a very long period, in this test case 0.003596. The exact situation is shown in Fig. 4(a), wherefrom it is evident that the GLS-T method can determine the harmonic variation itself as a trend component, and lead to a completely erroneous fit. Fig. 4(b) we show the corresponding spectra. As expected, the GLS model, coinciding with the true model, gives the best estimate, while the BGLST model is not far off. The detrending, however, leads to significantly worse estimate. The estimates in the same order are the following: 0.003574, 0.003673 and 0.005653. The value of the trend learned by BGLST method was -0.00041, which is very close to zero.

The last example clearly shows that even in the simplest case of pure harmonic, if the dataset does not contain exact number of periods the spurious trend component arises. If pretrended, then this introduces bias into period estimation. In the case of BGLST method the sinusoid can be fitted to its fragments and zero trend can be recovered.

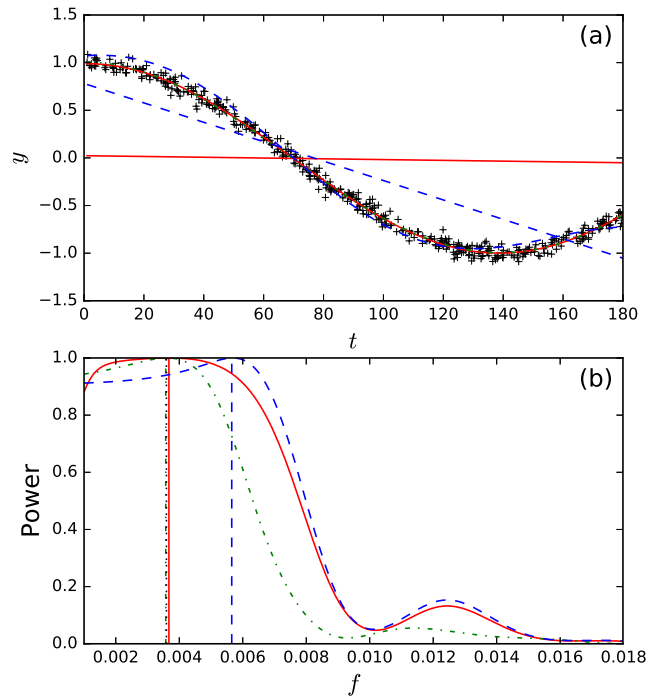


Fig. 4. Comparison of the results using different models. The true model contains one long harmonic with additive white Gaussian noise. (a) Data (black crosses), BGLST model (red continuous curve) and its trend component (red line), GLS-T model with trend added back (blue dashed curve), empirical trend (blue dashed line), GLS model (green dash-dotted curve). (b) Spectra of the corresponding models with vertical lines marking the locations of maxima. The black dotted line shows the position of the true frequency.

3.2. Effect of a nonconstant noise variance

We continue with investigating the effect of a nonconstant noise variance to the period estimate. In these experiments the data is generated from a purely harmonic model with a zero mean and no linear trend. We compare the results from the BGLST to the LS method. As there is no trend and offset in the model, we set the prior means for these two parameters to zero and variances to very low values. This essentially means that we are approaching the GLS method with a zero mean. In all the experiments the time range of the observations is $t_i \in [0, T]$, where $T = 30$ units, the period is drawn from $P \sim \text{Uniform}(5, 15)$ and the signal variance is $\sigma_S^2 = 1$. For each experiment we draw two values for noise variance $\sigma_1^2 \sim \text{Uniform}(2\sigma_S^2, 10\sigma_S^2)$ and $\sigma_2^2 = \sigma_1^2/k$, where $k \in [1, 2, 4, 8, 16, 32]$ which are used in different setups as indicated in the second column of Table 1. For each k we repeat the experiments $N = 2000$ times.

Let $\Delta_i = |f_i - f_i^{\text{true}}|$ denote the absolute error of the BGLST period estimate in the i -th experiment and $\Delta_i^{\text{LS}} = |f_i^{\text{LS}} - f_i^{\text{true}}|$ the same for the LS method. Then we measure a performance statistic

$$S_3 = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\Delta_i < \Delta_i^{\text{LS}}), \quad (52)$$

where \mathbb{I} is the indicator function. Let $\delta_i = \Delta_i/f^{\text{true}}$ and $\delta_i^{\text{LS}} = \Delta_i^{\text{LS}}/f^{\text{true}}$ be the relative errors of the period deter-

Table 1. Description of experimental setups with nonconstant noise variance. The first column indicates the number of the setup, the second column shows how the variance σ_i^2 for i -th data point was selected and third column the criteria of drawing the time moment t_i for i -th data point.

No.	Form of noise	Type of sampling
1	$\sigma_i^2 = \sigma_1^2 + \frac{t_i}{T}(\sigma_2^2 - \sigma_1^2)$	$t_i \sim \text{Uniform}(0, T)$
2	$\sigma_i^2 = \begin{cases} \sigma_2^2, & \text{if } t_i < T/2 \\ \sigma_1^2, & \text{if } t_i \geq T/2 \end{cases}$	$t_i \sim \text{Uniform}(0, T)$
3	$\sigma_i^2 = \sigma_2^2 + \frac{\sigma_i^{2(\text{emp})} - \sigma_{\min}^{2(\text{emp})}}{\sigma_{\max}^{2(\text{emp})} - \sigma_{\min}^{2(\text{emp})}}(\sigma_1^2 - \sigma_2^2)$	Based on MW dataset

Notes. In the 3rd row the $\sigma_i^{2(\text{emp})}$ denotes the empirical variance of the i -th datapoint and $\sigma_{\min}^{2(\text{emp})}$ and $\sigma_{\max}^{2(\text{emp})}$ the minimum and maximum intra-seasonal variances. In other words we have renormalized the intra-seasonal variances to the interval from σ_1^2 and σ_2^2 .

mination. The second performance statistic we measure is the relative difference in the average relative errors between the methods i.e.

$$S_4 = 1 - \frac{\sum_{i=1}^N \delta_i}{\sum_{i=1}^N \delta_i^{\text{LS}}}. \quad (53)$$

The list of experimental setups with the descriptions of the models are shown in Table. 1. In the first experiment the noise variance is linearly increasing from σ_1^2 to σ_2^2 . In practice this could correspond to decaying measurement accuracy over time. In the second experiment the noise variance abruptly jumps from σ_2^2 to σ_1^2 in the middle of the time series. This kind of situation could be interpreted as a change of one instrument to another, more accurate, one. In both of these experiments the sampling is uniform. In the third experiment we use samplings patterns of randomly chosen stars from MW dataset and the noise variance is calculated based on empirical intra-seasonal variances. In the first two experiments the number of data points was $n = 200$ and in the third between 200 and 400 depending on the randomly chosen dataset, which we downsampled.

In Fig. 5 the results of different experimental setups are shown, where the performance statistics S_3 and S_4 are plotted as function of $\sqrt{k} = \sigma_1/\sigma_2$. We see that when the true noise is constant in the data ($k = 1$), the BGLST method performs identically to the LS method while for greater values of k the difference grows bigger between the methods. For the 2nd experiment, where the noise level abruptly changes, the advantage of using a model with nonconstant noise seems to be the best, while for the other experiments the advantage is slightly smaller, but still substantial for larger k -s.

In Fig. 7 we have shown the models with maximally differing period estimates from the 2nd experiment for $\sqrt{k} = 5.66$. On the left column of the figures there is shown the situation where our method outperforms the most the LS method and on the right column vice versa. For colour and symbol coding, please see to the caption of Fig. 7. This plot is a clear illustration of the fact that even though on average the method with nonconstant noise variance is better than LS method, for each particular dataset this might not be the case. Nevertheless it is also apparent from the figure that when the LS method is winning over BGLST method,

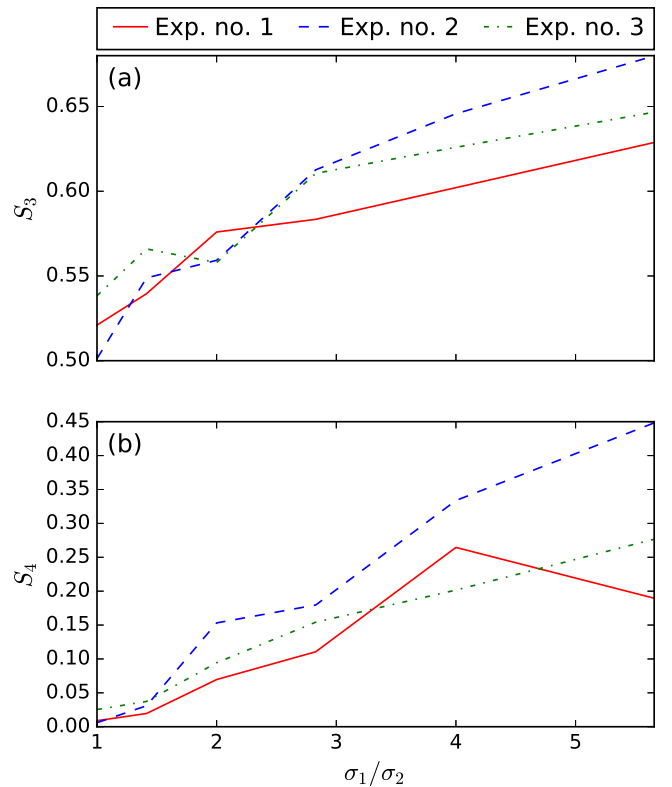


Fig. 5. Results of the experiments with known nonconstant noise variance. (a) Performance measures S_3 as function of σ_1/σ_2 . (b) The same for performance measure S_4 .

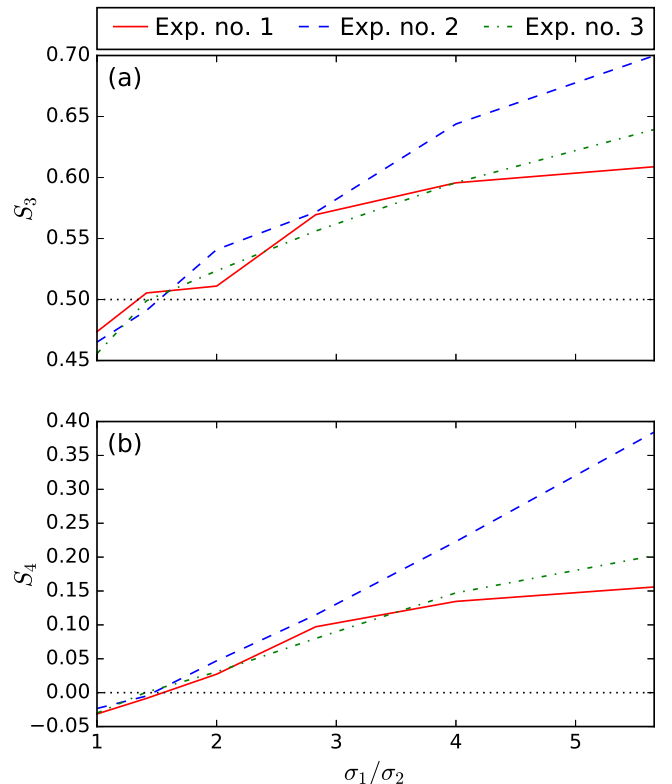


Fig. 6. Results of the experiments with unknown nonconstant noise variance. The noise variance is empirically estimated from the data. (a) Performance measures S_3 as function of σ_1/σ_2 . (b) The same for performance measure S_4 . The black dotted horizontal line shows the break even point between the two methods.

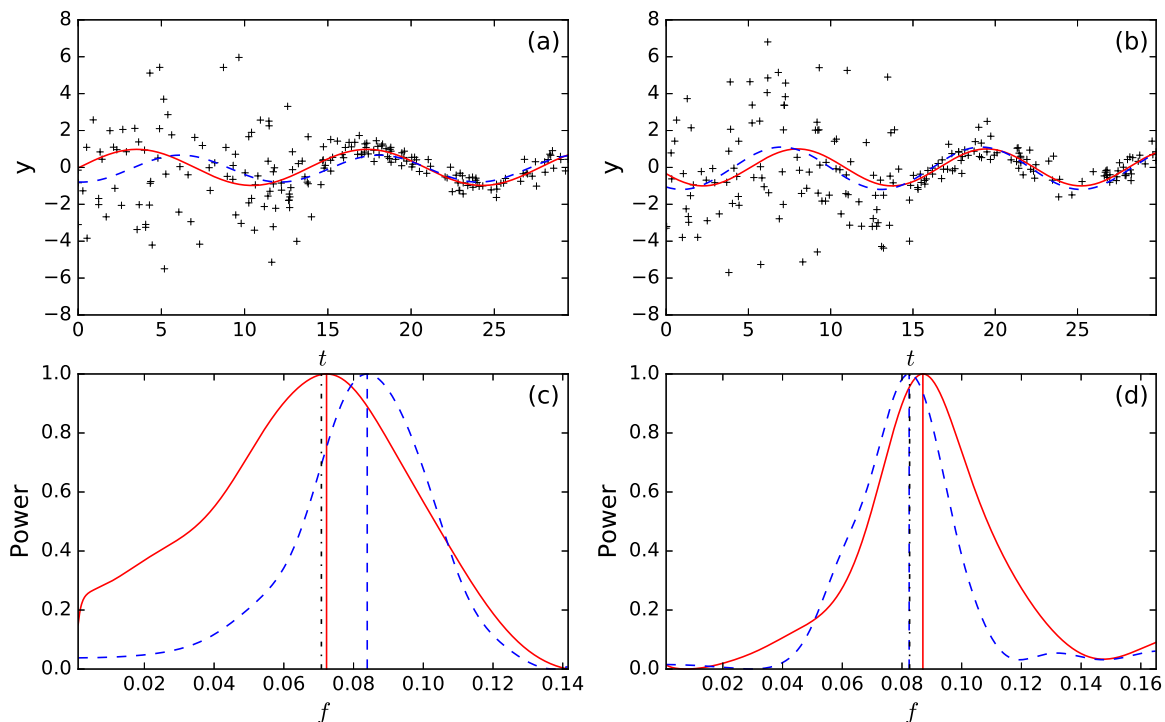


Fig. 7. Comparison of the methods for experiment with setup 2. (a) and (b) Data points with black crosses, red continuous curve – BGLST model, blue dashed line – LS model fit. (c) and (d) Spectra of the models with same colours and line styles. The vertical lines correspond to the optimal frequencies found and the black dotted line shows the true frequency. The plots on two columns correspond to the biggest difference in the period estimates from both methods w.r.t. each other. On the left our method outperforms LS method, on the right vice versa.

the gap tends to be slightly smaller than in the opposite situation.

In the previous experiments the true noise variance was known, but this rarely happens in practice. However, when the data sampling is sufficiently dense, we can estimate the noise variance empirically using e.g. narrow sliding window around each data point. We now repeat the experiments using such an approach. In experiment setups 1 and 2 we used windows with the length 1 unit and in the setup 3 the intra-seasonal variances. In the first two cases we increased the number of data points to 1000 and in the latter case we did not downsample the data. The performance statistics for these experiments are shown in Fig. 6. We see that when the true noise variance is constant ($k = 1$) then using empirically estimated noise in the model leads to slightly worse results than with constant noise model, however roughly starting from the value of $k = 1.5$ in all the experiments the former approach starts to outperform the latter. The location of the break-even point obviously depends on how precisely the true variance can be estimated from the data. This, however depends on how dense is the data sampling and how short is the expected period in the data, because we want to avoid counting signal variance as a part of the noise variance estimate.

3.3. Real datasets

As the last examples we consider time series of two stars from the MW dataset. In Figs. 8 and 9 we show the differences between period estimates for the stars HD37394 and HD3651 correspondingly. For both stars we see that the

linear trends fitted directly to the data (blue dash-dotted line) significantly differ from the trend component in the harmonic model (red dash-dotted line) with both types of noise models (compare the left and right columns of the plots), and consequently, also the period estimates in between the methods always tend to vary somewhat.

Especially in the case of HD37394, the retrieved period estimates differ significantly. Assuming a constant noise model for this star results in the BGLST and GLS/GLST to deviate strongly, the BGLST model producing a significantly larger period estimate. With intra-seasonal noise model, all three estimates agree more to each other, however, BGLST still gives the longest period estimate. In the previous study by Baliunas et al. (1995) the cycle period for this star has been reported to be 3.6 yrs which exactly matches the period estimate of GLS-T estimate with constant noise model⁴. The period estimate of BGLST method, being less sensitive to the noise model, is 5.8 yrs.

In the case of HD3651 the period estimates of the models are not that sensitive to the chosen noise model, but the dependence on how the trend component is handled, is significant. Interestingly none of these period estimates match the estimate from Baliunas et al. (1995) (13.8 yrs), where the GLS-T was used. All three estimates shown here are above 16 yrs. As can be seen from the model fits in Fig. 8(a) and (b), they significantly deviate from the data, so we must conclude that the true model is not likely to be harmonic. Conceptually the generalization of the model proposed in the current paper to the nonharmonic case is

⁴ In their study they have dropped the data prior 1980, but this has negligible effect on the results

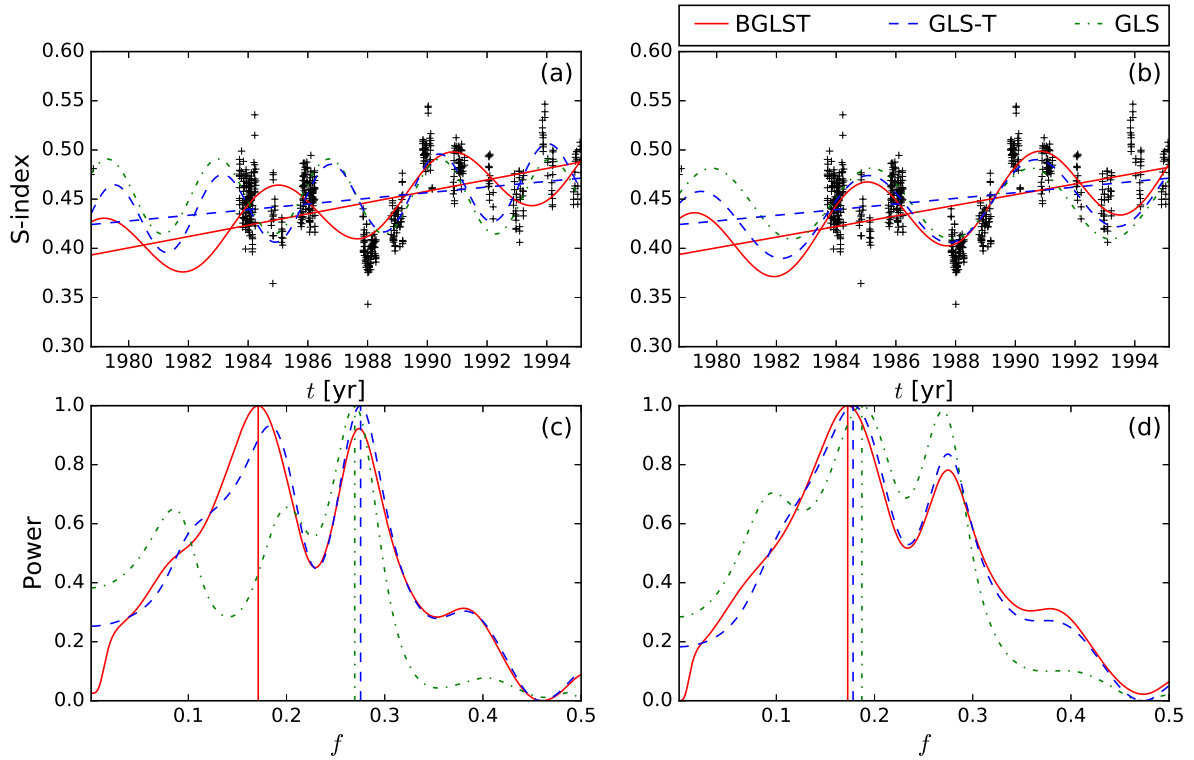


Fig. 8. Comparison of the results for star HD37394 using BGLST, GLS-T and GLS models with constant noise model on the plots on the left column and with intra-seasonal noise model on the right column. (a) and (b) Data (black crosses), BGLST model (red curve), GLS-T model with trend added back (blue dashed curve), GLS model (green dash-dotted curve), the trend component of BGLST model (red line) and the empirical trend (blue dashed line). (c) and (d) Spectra of the models with same colours. The dashed lines mark the locations of corresponding maxima.

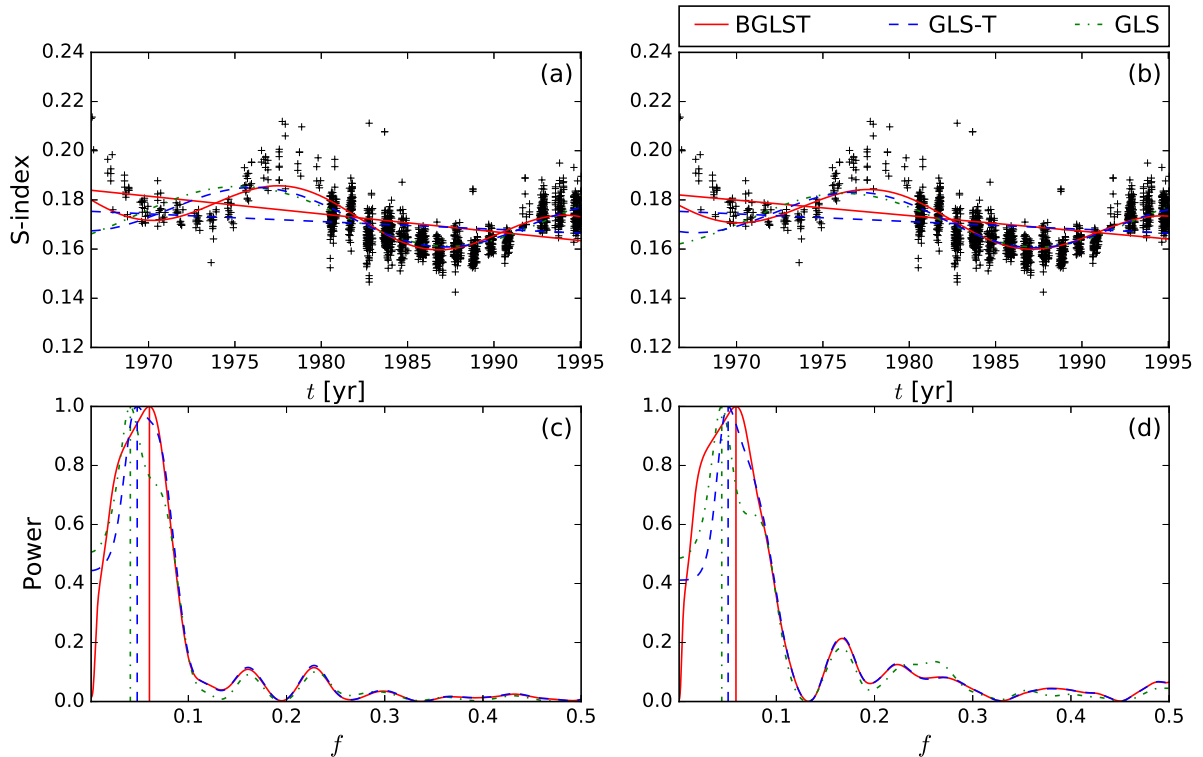


Fig. 9. Comparison of the results for star HD3651 using BGLST, GLS-T and GLS models. The meaning of the panels and colour coding is identical to the in Fig. 8.

straightforward, but becomes analytically and numerically less tractable. Fitting periodic and quasi-periodic models to the MW data, which are based on Gaussian Processes, are discussed in Olsper et al. (2017).

4. Conclusions

In this paper we introduced a Bayesian regression model which involves, besides harmonic and offset components, also a linear trend component. The main focus of this paper was on addressing the effect of linear trends in data to the period estimate. We showed that when there is no prior information on whether and to what extent the true model of the data contains linear trend, it is more preferable to include the trend component directly to the regression model rather than either detrending the data or leaving the data untrended before fitting the periodic model.

We also showed that if the true noise model of the data is far from being constant, then by neglecting this knowledge, period estimates start to deteriorate. Unfortunately in practice the true noise model is almost never known and in many cases impossible to estimate empirically (e.g. sparse sampling rate compared to the true frequency). In this study we, however, focused only on the long period search task, which made estimating the noise variance on narrow local subsamples possible. Based on the experiments we saw that if the true noise is not constant and the extremes of the S/N ratio differ at least two times, then using a model with empirically estimated noise variance is well justified.

Acknowledgements. This work has been supported by the Academy of Finland Centre of Excellence ReSoLVE (NO, MJK, JP), Finnish Cultural Foundation grant no. 00170789 (NO) and Estonian Research Council (Grant IUT40-1; JP). MJK and JL acknowledge the 'SOLSTAR' Independent Max Planck Research Group funding.

The HK Project_v1995 NSO data derive from the Mount Wilson Observatory HK Project, which was supported by both public and private funds through the Carnegie Observatories, the Mount Wilson Institute, and the Harvard-Smithsonian Center for Astrophysics starting in 1966 and continuing for over 36 years. These data are the result of the dedicated work of O. Wilson, A. Vaughan, G. Preston, D. Duncan, S. Baliunas, and many others.

References

- Baliunas, S. L., Donahue, R. A., Soon, W. H., et al. 1995, ApJ, 438, 269
- Bretthorst, L. G. 1988, Bayesian Spectrum Analysis and Parameter Estimation, Vol. 48 (Springer)
- Carbonell, M., Oliver, R., & Ballester, J. L. 1992, A&A, 264, 350
- Deeming, T. J. 1975, Ap&SS, 36, 137
- Ferraz-Mello, S. 1981, AJ, 86, 619
- Ford, E. B., Moorhead, A. V., & Veras, D. 2011, ArXiv e-prints [arXiv:1107.4047]
- Horne, J. H. & Baliunas, S. L. 1986, ApJ, 302, 757
- Lomb, N. R. 1976, Astrophysics and Space Science, 39, 447
- Mortier, A., Faria, J. P., Correia, C. M., Santerne, A., & Santos, N. C. 2015, A&A, 573, A101
- Murphy, K. P. 2012, Machine learning: a probabilistic perspective (Cambridge, MA)
- Olsper, N., Lehtinen, J., Grigorievskiy, A., Käpylä, M. J., & Pelt, J. 2017, ArXiv e-prints [arXiv:1712.XXXX]
- Roberts, D. H., Lehar, J., & Dreher, J. W. 1987, AJ, 93, 968
- Saar, S. H. & Brandenburg, A. 1999, ApJ, 524, 295
- Scargle, J. D. 1982, 263, 835
- Schwarzenberg-Czerny, A. 1998, MNRAS, 301, 831
- VanderPlas, J. T. 2017, ArXiv e-prints [arXiv:1703.09824]
- Vaniček, P. 1971, Ap&SS, 12, 10
- Vio, R., Andreani, P., & Biggs, A. 2010, A&A, 519, A85
- Zechmeister, M. & Kürster, M. 2009, A&A, 496, 577