Statistical inference on the periodicity of irregularly sampled light curves

A THESIS SUBMITTED TO
THE UNIVERSITY OF KENT AT CANTERBURY
IN THE SUBJECT OF STATISTICS
FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY
143 pages

Efthymia Derezea

School of Mathematics, Statistics and Actuarial Science
University of Kent

September 2022

I get out of my car
Step into the night
And look up at the sky
And there's something bright
Traveling fast
Look at it go...

Kate Bush

Abstract

This thesis deals with the problem of period estimation of irregularly sampled time series, and specifically light curves of variable young stars. Knowing the period of these objects can provide important information about the stars' formation and other characteristics. The light curves are measurements of brightness over time conducted in multiple astronomical filters. We examine this problem from three different points of view. First, we need to obtain an accurate period estimate. For that purpose we introduce a weighted t-process regression model for period estimation as a flexible alternative to Gaussian process regression, since it is common for such data to exhibit a fat-tail behaviour, and we extend these models in order to include measurements from multiple astronomical filters. Secondly, we need to accompany our estimates with some credibility as to whether they represent a real periodic signal. This is usually addressed through hypothesis testing. To that end, we introduce a flexible testing scheme using saddlepoint approximation, that can be applied on a range of periodic models including Gaussian process regression. These tests are also extended for data contaminated with red noise, a type of correlated noise that usually appears in such data. Finally, we explore the asymptotic properties of simple harmonic models with additional red noise and show that this estimates are consistent and asymptotically normal. We test our results through extensive simulation studies which are reported along with an application on some real light curves from the Hunting Outbursting Young Stars citizen science project.

Acknowledgements

I would like to thank my supervisor Fred Kume for his guidance and insightful comments, and my second supervisor Dirk Froebrich for the interesting astronomical discusssions.

Also thanks to all those people including my friends, partner and familly, who with their unconditional support throughout this time brightened my days in Canterbury.

Fenia, March 2023.

Contents

Chapter 1

Introduction and background

In this thesis we explore and propose methods around the topic of period estimation of light curves of variable young stars. These data are a complex type of irregularly sampled time series. The problem of estimating the underlying periodicity of time series in the presence of irregularly sampled data appears in many disciplines including Economics (see Baltagi and Wu (1999) [6]), Climatology (see Schulz and Stattegger (1997) [55]), Biology (see Heerah et al. (2020) [31]), or Astronomy (for example for rotation period searches of stars as in Herbst et al. (2007) [32]; Bouvier et al. (2014) [13]). In the following Chapters we discuss methods for obtaining appropriate period estimators, we see how we can associate these estimates with some credibility on their validity through a hypothesis testing setting and we examine the statistical properties of some of these methods.

The study of this problem is important both for the fields of Astronomy and Statistics. Current and future large astronomical surveys potentially generate millions of such light curves, see for example details for the LSST project in Malz et al. (2019) [48]. Thus, there is a need of accurate, automated, and fast methodologies to determine periods reliably. Depending on the type of star being able to estimate the period of light curves can provide important information about the star itself (rotation period), its formation, its internal structure and the environment it is in (e.g. structures in accretion disks around young stars). More details on the importance of the study of light curves can be found in Feigelson et al. (2021)

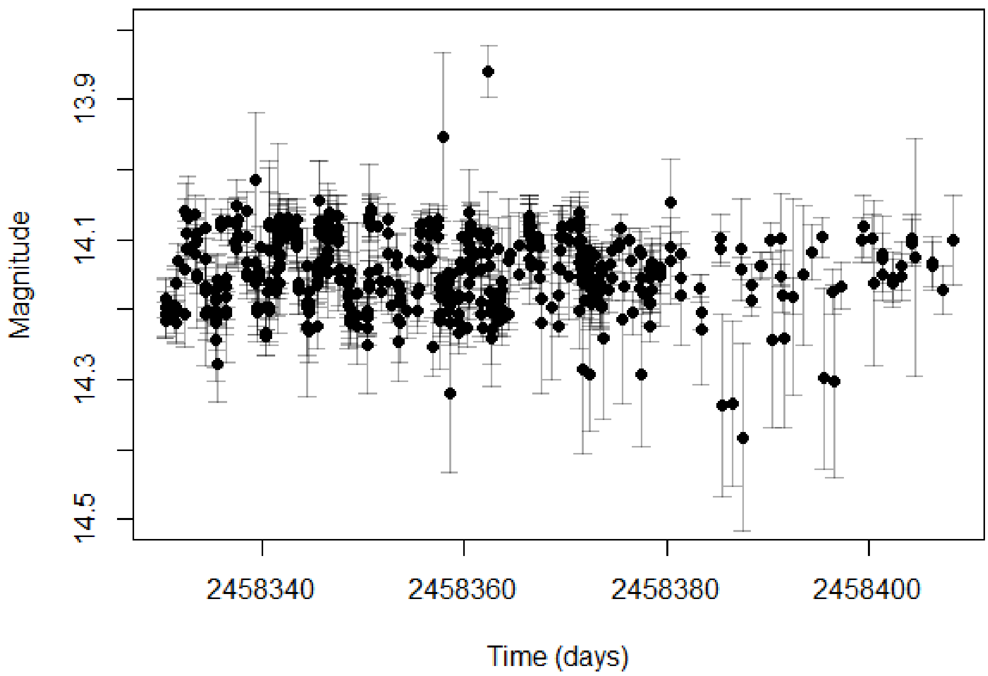


FIGURE 1.1: Object V1598Cyg

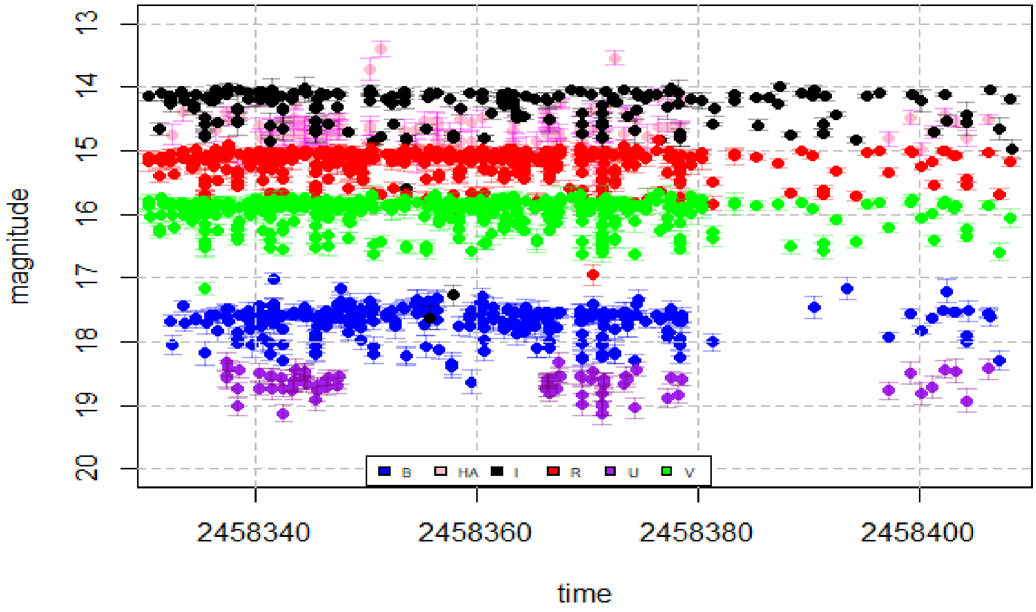
[22]. From the statistical point of view the problem of irregularly sampled time series is under explored. Both for the time and frequency domain analysis of series like these there have been proposed basic methods, but they have not been fully studied for real life problems. This means that there is a lot of room for new statistical methods and theory to be developed, see for example Erdogan et al. (2005) [21].

The data that we work with here represents the brightness of stars, which are usually quoted in magnitudes y. These are determined from the physical flux measurements f as $y = -2.5 \log(f/f_0)$, where f_0 is the flux zero point and it is a constant. For more details on astronomical light curves see Allam et al. (2018) [3]. Thus, numerically large magnitude values indicate fainter brightness and vice versa. The times for the data points in the light curves are usually registered as a Julian date. This is a count of days since noon January 1st, 4713 BC. A typical example of a light curve can be seen in Figure 1.1.

In statistics we are used to studying regularly sampled times series, in other words the data observation occurs in a specific pattern e.g. hourly, daily etc. The sampling of the light curves we study here on the other hand occurs at irregular time points. This happens because the data collection which is carried out by ground telescopes can be affected by many factors such as weather, visibility or the telescopes' schedule. For example we have more observations during summer and fewer in the winter and of course we do not have observations during day time. The factors that affect the frequency of the sampling, thus, are independent of the response variable.

Another interesting feature of the data is that it is accompanied by some additional information about the credibility of the measurement. Specifically, each data point has an associated individual measurement accuracy (see the grey bars centred at the observation points in Figure 1.1), small values of which reflect small measurement uncertainty. These uncertainties represent one sigma errors and are determined during the data processing and calibration of the astronomical images (see Allam et al. (2018) [3]). Here these quantities are treated as known weights. Finally, each star is measured in different astronomical filters. There are 6 potential such filters, namely the Visual (V), Red (R), Infrared (I), Blue (B), Hydrogen-alpha (HA) and Ultraviolet (U). In astronomy these are referred to as passbands and they correspond to different wavelength intervals. In practice we do not have measurements of each star in all six filters, the most common ones are the visual, red and infrared. See for example Figure 1.2 (Top) where object 5686 from Froebrich et al. (2021) [24] is measured all available filters and Figure 1.2 (Bottom) with object 5715 from Froebrich et al. (2021) [24] in Infrared and Red filters. This means that the number of available filters varies for each object. Note that the time observations are also different for each filter since they are usually measured independently.

Throughout this thesis we use data from the Hunting Outbursting Young Stars (HOYS) citizen science project (Froebrich et al. (2018) [23]). These data contain measurements collected by amateur and professional astronomers, as well as universities, and combined in order to create long term light curves. The HOYS



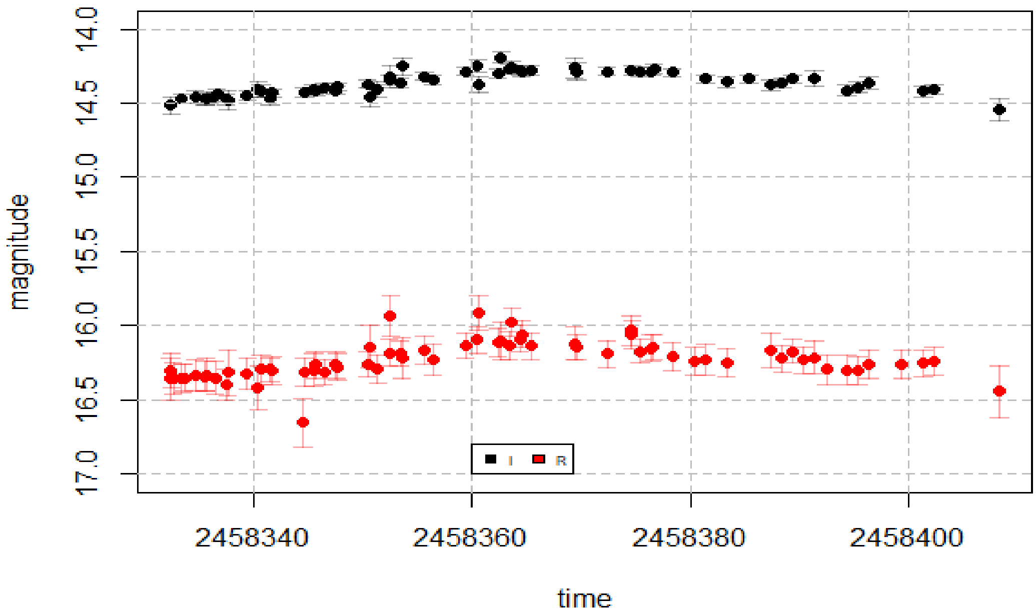


FIGURE 1.2: (Top) Object 5686 from Froebrich et al. (2021) [24] measured in 6 astronomical filters, with known period at 1.67 days. (Bottom) Object 5715 measured in the red and infrared filter, with a known period at 1.39 days.

FIGURE 1.3: The Pelican nebula where many of the objects we study here are located.

database contains measurements of objects from various star forming regions, but here we investigate objects located in the Pelican nebula, see Figure 1.3. Periodic light curves of young stars can be used to measure rotation periods of the stars, as well as to track the evolution of properties of surface features on them. This dataset contains a big range of different light curves. There are some objects like V1598Cyg in Figure 1.1 that are obviously periodic. Others, on the other hand, do not exhibit eminent periodic patterns. See for example object 5686 in Figure 1.2 (Top) where no obvious periodic pattern seems to exist. This object, however, has a period at 1.67 days and this pattern becomes apparent if we look at it folded in period 1.67 days (Figure 2.4 (Right)), we formally discuss folded data at the end of Chapter 2. There are also objects which are variable but not periodic and finally light curves that do not vary at all to name a few.

The problem of period estimation of astronomical light curves is well studied for data that is it reasonable to assume a sinusoidal shape for (Vanderplas (2018) [66]), as we will see in Chapter 2. This assumption however is not always sufficient to describe our data and thus flexible non parametric methods are needed. Fitting a non-parametric model to complex time series like these is a challenging problem. First, obtaining a period estimate is numerically a non trivial task, due to the nature of the periodic functions we need to optimize, as they exhibit thousands of local maxima or minima. The strategy used to deal with this is through evaluating these functions an a grid of periods, or in other words, via fine discretization of the period parameter space (see Section 3.8.2). This is thus a computationally demanding problem even for one light curve and the problem becomes more challenging when thousands of such light curves need to be analysed. In general, the more complex a model is, the larger the computation time and that is the case for all non-parametric approaches discussed here. A way to overcome this difficulty is by implementing our computations in parallel, a process available from most standard programs like R, which reduces significantly the required time to perform calculations. Moreover, these data as described before can exhibit complex features such as the existence of additional measurement accuracies, or multiple filter measurements. These attributes turn the non-parametric modelling of these curves into a difficult task. Our contributions towards this end include the use of a weighted t-process regression in order to model periodic light curves with outliers and additional measurement accuracies. We show how we can model light curves with correlated residuals through a Gaussian process regression setting and finally we extend the Gaussian and t-process models to include measurements from multiple filters.

Another important part of the period estimation problem is adding some credibility to the period estimates we obtained and being able to decide whether our estimate represents true periodic signal in our data or it is just the product of random oscillations. This is usually done through some hypothesis testing schemes. When it comes to non-parametric methodologies there does not exist a test for that purpose and current inference methods are based on time consuming Monte Carlo techniques (e.g. Do et al. (2009) [18]). Here, we introduce a general Hypothesis test setting for non-parametric models and specifically Gaussian process regression, providing thus a solution to the problem of period detection for these flexible models. Furthermore, we show how these tests can be adjusted in the presence of correlated noise (red noise), obtaining more accurate results while drastically reducing the number of periods which could be falsely identified as valid. With this approach we manage to perform calculations in 2% of the time needed when using simulations instead.

Finally, these data are a type of irregularly sampled time series as explained earlier. For example for many of the period estimates we examine in this thesis there do not exist results about their statistical properties, such as their bias or convergence rate to the true parameter. Because of the irregular nature of the time points, usually stochastic models are assumed for the time observations, thus, standard maximum likelihood estimator (mle) results for the asymptotic properties no longer hold. Some results exist for the case of a simple sinusoidal model with white noise (Reimann (1994) [53]). Here we extend these results for sinusoidal data with correlated residuals and show that the parameter estimate under this setting is consistent and asymptotically normal. This result is stated in Chapter 5, see Theorem 5.1.

Below see the outline of this thesis. In Chapter 2 we discuss related work and methods around the problem of period estimation of light curves. The problem of period detection is discussed in Chapter 3 along with our proposed tests under non-parametric settings. Specifically, our contributions to that end can be summarized by the following three points.

- Introduction of a saddlepoint approximation test for non-parametric models, providing a computationally faster alternative to simulation based methods.
- Introduction of a leave-one-out cross-validation error test for Gaussian process regression.
- Extension of the above tests for data with correlated residuals (red noise).

In Chapter 4 we address the problem of period estimation under the Gaussian process regression setting. Our contribution to this subject is three-fold as seen in the next points.

- Introduction of a weighted t-process regression period estimation model as a flexible alternative to Gaussian process regression for data with fat tails.

- Extension of the Gaussian and weighted t-process regression models to incorporate measurements from multiple filters.

- Use of the saddlepoint approximation tests as Kernel selection tools for quasi periodic data.

The asymptotic properties of period estimates for sinusoidal models where the time observations follow a uniform distribution model are discussed in Chapter 5, where following Reimann (1994) [53] we show that:

- The frequency estimate of a sinusoidal model with red noise is consistent and asymptotically normal. The same result is derived for data with AR(1) noise instead.

- Reimann's result holds for sinusoidal models that include the intercept term.

- The same asymptotic properties hold when we take the measurement accuracies into account.

Finally, in Chapter 6 we show our conclusions and discuss some future work paths.

Chapter 2

Period estimation

As we saw in the previous Chapter the focus of this thesis is analysing and estimating the period of light curves, which are a specific type of time series. When dealing with time series there are two common types of modelling strategies, the time domain and the frequency domain analysis (Chatfield (2003) [15]). In this particular problem the latter approach is more appropriate since we are interested in exploring and analysing the cyclic behaviour of the series or more generally the possible repeating patterns that may be present. A useful and commonly used tool for period estimation is the periodogram (Schuster (1898) [56]), which is a measure distinguishing excessive periods relative to the rest. We should mention that throughout this thesis we focus on the case where only one true underlying period exists in the data, although some of the methods discussed could be or have been applied for the study multiple periodicities.

In the following Sections we will discuss the classic periodogram definition and its extensions that will help us account for the irregularities in our data. Specifically, In Section 2.1 we discuss period estimation methods for the simple case of regularly sampled data. A discussion as to why these classical approaches are not ideal for the irregular data we study here is given in Section 2.2 along with the methods designed particularly for irregularly sampled data, both based on linear and non-parametric models. Finally, an extensive discussion on estimating the period based on time versus the phase is provided in Section 2.3.

2.1 Regularly sampled time series

The problem of period estimation of time series is well studied in the literature, for the case of data arriving at equal intervals e.g. hourly, daily etc. In this Section we will discuss the classic approach for spectral analysis. First, let $\mathbf{y}$ denote the vector of n signal observations such that each entry y_j is the data arriving at times t_j, $j = 1, ..., n$. We assume that the data are generated as

$$y_j = g(t_j; p) + \epsilon_j \tag{2.1}$$

where g is the assumed true function with period p while $\epsilon_j \sim N(0, \sigma^2)$ are the measurement errors. For this Section we will assume that our time points are observed regularly, e.g. $t_j = j$. If we take the measurement accuracies, which we will denote as s_j, into account, (2.1) via averaging argument changes to

$$y_j/s_j = g(t_j; p)/s_j + \epsilon*_j, \quad \epsilon*_j \sim N(0, s_j^2) \tag{2.2}$$

taking the form of a weighted regression model.

The simplest way to describe the cyclic behaviour of our data is to assume a single cosine wave for the form of g as seen below.

$$y_j = A\cos(2\pi\omega t_j + \phi) + \epsilon_j \tag{2.3}$$

where A is the amplitude, ϕ is the phase and ω is the frequency (it holds that $\omega = 1/p$). Using the trigonometric identity,

$$A\cos(2\pi\omega t + \phi) = A\cos(\phi)\cos(2\pi\omega t) - A\sin(\phi)\sin(2\pi\omega t)$$

model (2.3) can equivalently be written as,

$$y_j = \beta_1\cos(2\pi\omega t_j) + \beta_2\sin(2\pi\omega t_j) + \epsilon_j \tag{2.4}$$

with

$$\beta_1 = A\cos(\phi), \quad \beta_2 = -A\sin(\phi)$$

Because of the good mathematical properties that sinusoids and cosines have, this model has been used extensively in the literature, see for example Hannan (1973) [30] or Walker (1971) [71]. Given a period (or frequency) we can obtain the estimates for parameters β_1 and β_2 by minimizing the residual sum of squares,

$$S = \sum_{i=1}^{n} (y_i - \beta_1 \cos(2\pi\omega) - \beta_2 \sin(2\pi\omega))^2$$

By calculating the first derivative of S with respect to β_1 and β_2 we obtain the desired estimates. These estimators can be written approximately as (Bloomfield (2004) [12]):

$$\beta_1(\omega) = \frac{2}{n}\sum_{j=1}^{n} y_j \cos(2\pi\omega t_j), \quad \beta_2(\omega) = \frac{2}{n}\sum_{j=1}^{n} y_j \sin(2\pi\omega t_j) \tag{2.5}$$

Looking carefully at equations (2.5), we notice that they are the components of the discrete Fourier Transform, which is given by:

$$d(u) = \frac{1}{n}\sum_{j=1}^{n} g(t_j;\omega)e^{-2\pi i u t_j} \tag{2.6}$$

where, $\exp(i2\omega t_j) = \cos(2\omega t_j) + i\sin(2\omega t_j)$, if g is replaced with out observations y_j. For a general discussion of this subject see Terras (1985) [62] and for the analytic derivation of these equations see Bloomfield (2004) [12]. For the general case of equally spaced time observations the most popular method for estimating periodicities is based on (2.6) and it is the classical periodogram introduced by Schuster (1898) [56] and defined below for the observed data y_j.

$$P(\omega) = |d(\omega)|^2 \tag{2.7}$$

The periodogram is a function that obtains its maximum for the correct period/frequency for noise-free data, we infer ω as $\hat{\omega} = \text{argmax}P(\omega)$. Basically $P(\omega)$ is a measure of how much each period helps in explaining the variance of the data.

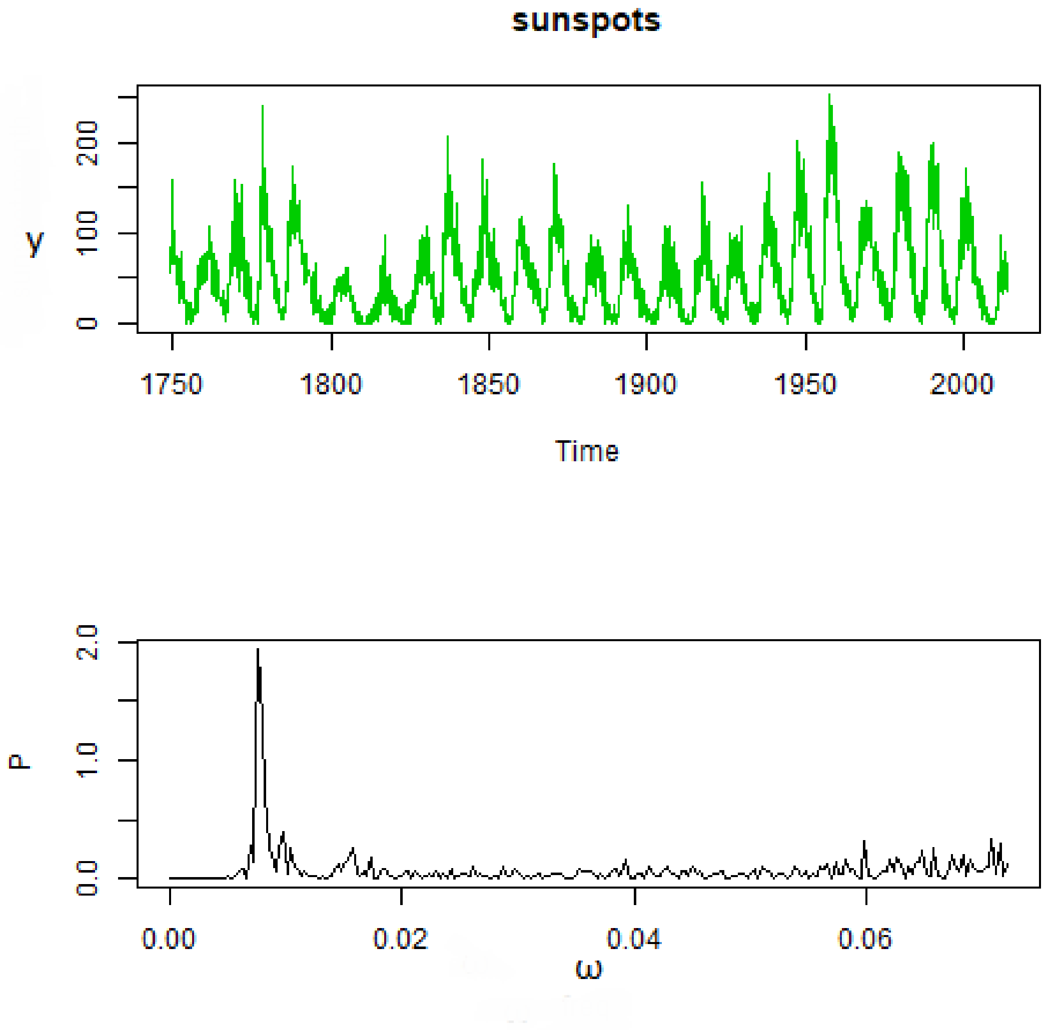


FIGURE 2.1: An example of monthly sunspots time series and periodogram. The data are available as sunspot.month requiring stats in R. In the first plot we see how the actual time series looks, on the second we see the periodogram versus the frequencies. Here, the highest peak is the most significant frequency/period.

In order to interpret the periodogram visually we could plot its values for a series of trial periods and we would expect to see a distinct peak for the period explaining best the data. For example in the bottom part of Figure 2.1 we have plotted the periodogram of the classic sunspot dataset "sunspot.month" from Astronomy, available in R, which is a characteristic example of this type of data, and has been used multiple times for the illustration of periodogram methods.

2.2 Irregularly sampled time series

The problem of frequency estimation of equally spaced data as described in Section 2.1 is well studied (Schuster (1898) [56], Chatfield (2003) [15], Priestley (1981) [52]). These methods, however, fail when our observations do not arrive at regular intervals, as is the case for the data studied here. This happens because the discrete Fourier Transform of unequally spaced data has an inherited "noisy" behaviour by definition, as we explain below. An illustrative discussion regarding this can be found in Vanderplas (2018) [66], where the author shows how an irregular sampling can lead to a noisy Fourier Transform and thus to noisy estimations of the Spectral density.

In order to understand this better, we step back to the Discrete Fourier Transform (2.6) which is an approximation of the Continuous Fourier Transform,

$$\mathcal{F}(u) = \int_{-\infty}^{\infty} g(t;\omega) e^{-i2\pi ut} dt \tag{2.8}$$

which is zero for every value of u except for $\mid \omega \mid$. This describes an infinite signal and our time variable is assumed to be continuous. In real life however we only observe a finite part of the signal and our time observations are discrete.

We can move from the unrealistic continuous signal to the observed one using a spectral Window function. Our observing window behaviour can be captured by a finite Dirac Comb function, which we will denote as $\delta(u)$ and its representation as a Fourier Series,

$$\delta(u) = \frac{1}{n} \sum_{j=1}^{n} e^{-i2\pi ut_j} \tag{2.9}$$

The discrete Fourier Transform thus can be obtained as a point-wise product of the spectral window function and the continuous Fourier transform (Reimann (1994) [53]):

$$d(u) = \mathcal{F}(u) \cdot \delta(u) = \int_{-\infty}^{\infty} \mathcal{F}(u-v)\delta(v) dv \tag{2.10}$$

If our observations are equally spaced $\delta(\omega)$ is 1 at frequencies 1,2,.. etc. See for example Figure 2.2 (Top) where $|\delta(\omega)|$ (Mod of $\delta(\omega)$) is plotted for 315 equally spaced data. The phenomenon of high peaks at integer frequencies is called aliasing (see more details in Reimann (1994) [53]). When time points are irregular $\delta(\omega)$ has the noisy behaviour seen in Figure 2.2 (Bottom). The peaks at integer frequencies are getting smaller and smaller as the frequency grows, and the intermediate frequencies are no longer zero. Thus, looking at (2.10), $d(\omega)$ by default will adopt a similar behaviour. In the next Sections we describe popular approaches designed to deal with similar data.

2.2.1 Linear regression periodograms

Many methods have been developed in the literature to obtain accurate period estimates of irregularly sampled data like the light curves we study here. The most popular approach dealing with this problem is that based on the Lomb-Scargle (LS) periodogram (Scargle (1982) [54]) which is a generalized form of the classical periodogram (2.7) but is not affected by irregular time spacing. It can be shown that this method produces identical results to that obtained by fitting a single sinusoidal wave (2.4) using the ordinary least squares (OLS) regression while assessing the fit at each period using the squared error, (Lomb (1976) [45]; Vanderplas (2018) [66]). The model fitted is basically the same as that of (2.11) below but without the intercept term. The latter method based on fitting a linear model gained more attention because of its connection with the Fourier transform, but most importantly because of its potential natural extensions, for example the introduction of the intercept term as seen below, (see also Cumming et al. (1999) [17] or Vanderplas and Ivezic (2015) [67]).

$$y_j = \beta_0 + \beta_1 \sin(2\pi t_j/p) + \beta_2 \cos(2\pi t_j/p) + \epsilon_j \tag{2.11}$$

In simple words the idea behind the OLS approach is the following: First, we select a series of trial periods. Secondary, given a trial period p estimate the parameters of the model (e.g. β_0, β_1, β_2) by minimizing the sum of squared errors

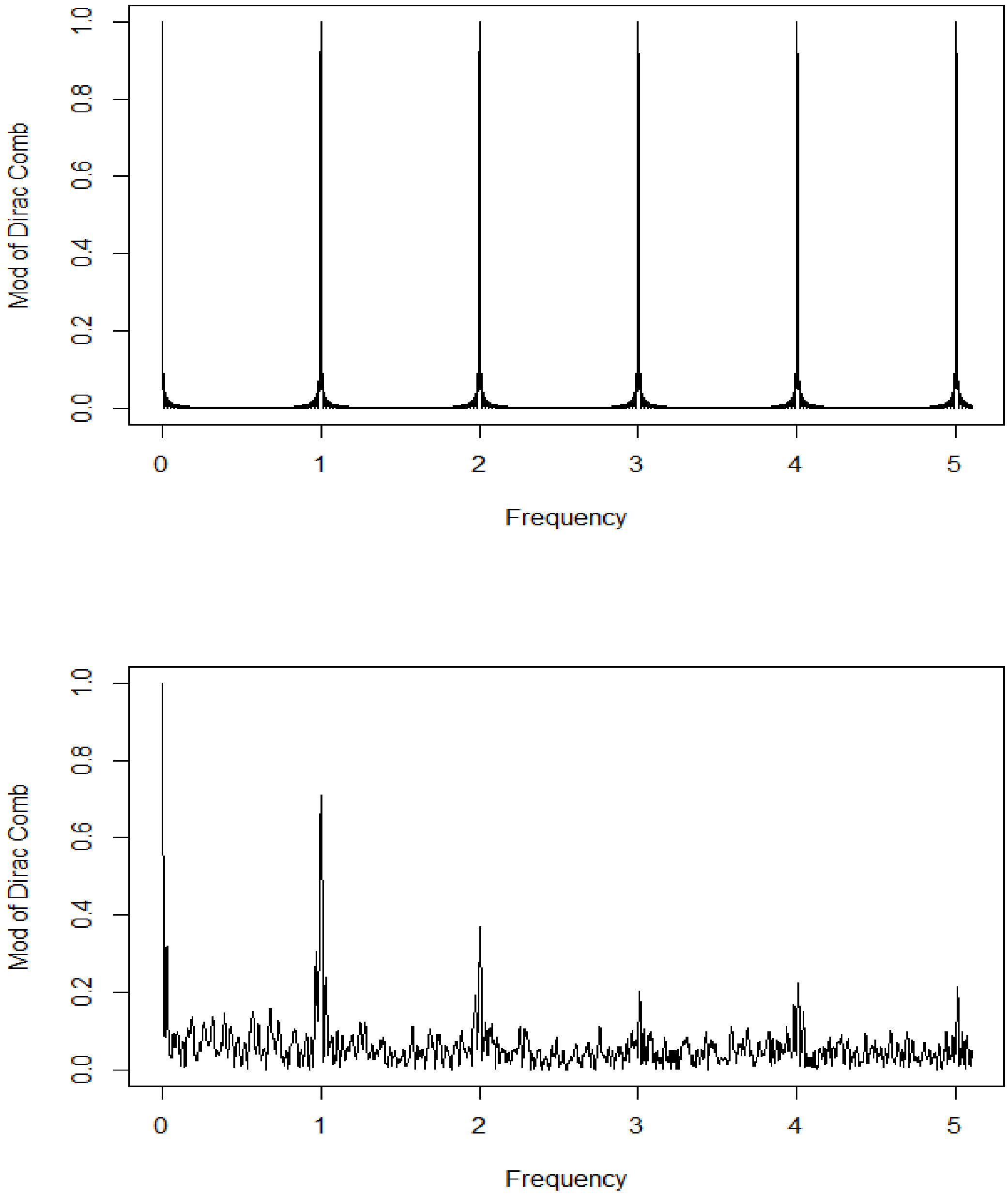


FIGURE 2.2: (Top) The Mod of Dirac Comb defined in (2.9) for 315 equally spaced time points. (Bottom) The Mod of Dirac Comb but this time using 315 real time points from light curve 6149 measured at the visual filter (Froebrich et al. (2021) [24]).

and calculate a corresponding goodness of fit measure as

$$\text{Per}(p) = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.12}$$

Finally, repeat this procedure for all trial periods. The function $\text{Per}(p)$, which in this case is the F-statistic, is our periodogram function, showing the relative importance of the periods. In this case large values indicate valid periods and thus our estimate will be $\hat{p} = \text{argmaxPer}(p)$. In order to understand this behaviour clearly see Figure 2.3 which displays a periodogram obtained with the least squared errors method for a real light curve with period at 0.8246 days, as seen in Froebrich et al. (2020) [25]. As we can see the periodogram has its highest peak at the correct period. For an explanation about the other two lower peaks see Froebrich et al. (2020) [25]. The analytical steps of this approach can be found in Thieler et al. (2013) [63].

These models are clearly very special cases of (2.1), while there exist many other approaches based on various forms for the periodic function g. In fact the biggest advantage of viewing the periodogram from the linear regression point of view is the flexibility it offers. For example we could use a whole range of different goodness of fit measures and, depending on the choice, our estimate $\hat{p}$ would be either the maximizer or minimizer of our periodogram function. In the next Sections we discuss such possible extensions.

2.2.2 Methods based on Robust Regression

In the previous Section we saw how we can obtain our period estimates by using OLS regression. There are many methods developed that explore alternative ways of parameter estimation falling under the umbrella of Robust Regression. The simplest example is the use of the absolute deviations error:

$$\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

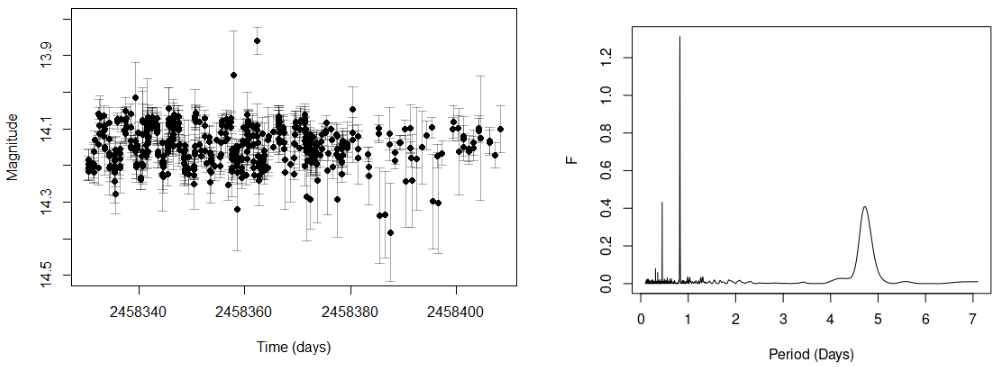


Figure 2.3: (Left) An example of a light curve as measurement of the brightness over time along with the measurement accuracies (error bars). This is object V1598Cyg measured in the visual filter. (Right) An example of a periodogram, for the light curve shown in the left panel, with F-statistic, based on fitting the sinusoidal model (2.11). High peaks indicate important periods. This object has a period of 0.8246 days.

instead of the sum of squared errors as used in OLS. There are many more regression methods available, some examples for our particular problem include that of Katkovnik et al. (1998) [38], where the coefficients of a sinusoidal wave are estimated using Huber's M-regression, or Liang et al. (2009) [44], where the absolute deviations (L1 norm) regression is used to construct a periodogram in order to estimate the period of gene expression data, showing that these approaches perform better than the L2 norm based methods when the data have outliers.

A detailed discussion to that end can be found in Thieler et al. (2013) [63] where the authors study, under the same framework, periodograms obtained by combinations of models (such as splines or sinusoidal basis) and different types of regression. In a comprehensive simulation study they compare the performance of the methods for light curves both with and without the presence of red noise and show, for example, that robust regression methods are more reliable for data with outliers. This is also supported by some of our results of a period search survey on real light curves (Froebrich et al. (2021) [24]) which showed that for those data the best performing method was that based on the L1 norm. The performance of the methods is usually assessed on their ability to identify the true period, while keeping small the number of falsely identified ones. These concepts are referred to

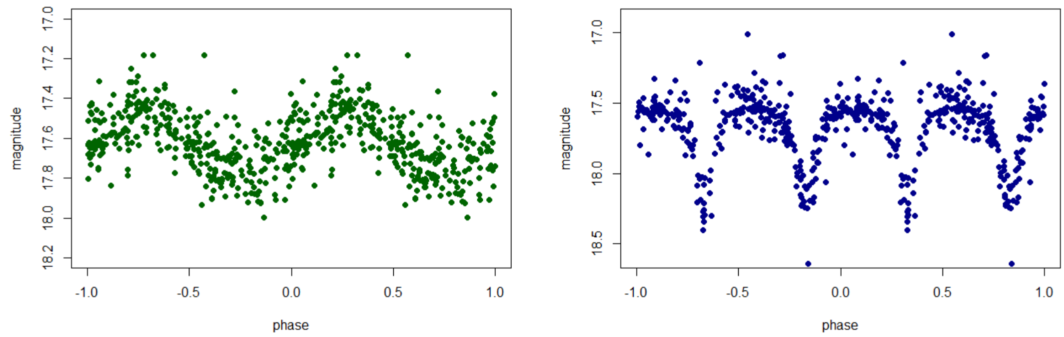


FIGURE 2.4: (Left) Object 6149 from Froebrich et al. (2021) [24] measured on the visual filter, folded in period 2.1763 days. (Right) Object 5686 from the same survey measured in the Blue filter and folded in 1.176 days period.

in Astronomy as completeness and contamination.

2.2.3 Non parametric periodograms

The methods described so far are based on assuming a sinusoidal shape for our data. Although sinusoidal like light curves appear often, see for example Figure 2.4 (Left), there are cases that the data deviate from this behaviour, e.g. Figure 2.4 (Right). There have been methods developed assuming different periodic shapes. See for example Stellingwer et al. (1978) [61] where the authors use step functions to describe the data, or Palmer (2009) [50] for a Fourier series fitting.

These are still however specific cases of a range of possible periodic shapes and their correct choice requires a good prior understanding of the data. Thus, in many cases the sinusoidal or any other specific shape is a very strong assumption while other more flexible non parametric methods are preferred. In Figure 2.5 we see the typical sinusoidal shape compared to various periodic curves. In the next Sections of this Chapter we will describe non-parametric periodograms that address exactly this problem.

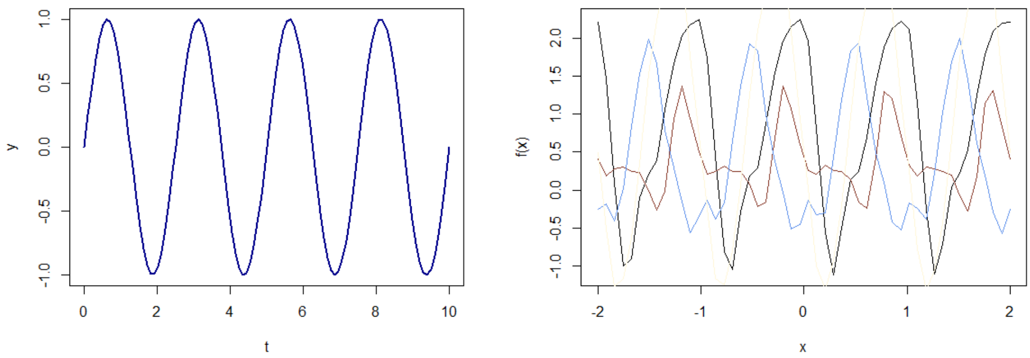


FIGURE 2.5: (Left)Typical example of a sinusoidal shaped curve. (Right) Example of non-sinusoidal periodic curves.

Spline regression

The first method we describe is based on fitting splines. For example, Akerlof et al. (1994) [1] fit cubic b-splines using least squares regression in order to estimate the periods of various stars. Another example is the use of periodic splines as seen in Oh et al. (2004) [49], where the authors use cubic splines with a periodic restriction to fit the model by performing robust regression (i.e. Huber's regression) and finally the fit is assessed based on a cross validation approximation. Although this is not typically considered as a non-parametric method we add it here because it makes fewer assumptions about the shape of the data compared to a sinusoidal wave for example.

A simple way to view splines is as fitting piece-wise polynomials to our data, taking into account constraints for the continuity etc. It could be seen as a combination of fitting polynomials and a step function that splits the data (t here) into separate regions, thus allowing the capture of more complex shapes. The points at which the piecewise polynomials meet are called knots (K). This is a more general model with fewer assumptions on the shape of the data.

A simple cubic spline model with $K = 2$ knots at ξ_k has the form:

$$y_j = \sum_{q=1}^{K+4} \beta_q h_q(t_j) + \epsilon_j \tag{2.13}$$

Where,

$$h_k(t) = t^{k-1}, \quad k = 1, \ldots, 4$$
$$h_{4+k}(t) = (t - \xi_k)^3_+, \quad k = 1, \ldots, K$$

Or simply for ξ_1 and ξ_2:

$$y_j = \beta_1 + \beta_2 t_j + \beta_3 t_j^2 + \beta_4 t_j^3 + \beta_5 (t_j - \xi_1)^3_+ + \beta_6 (t_j - \xi_2)^3_+ + \epsilon_j$$

Where $(t - \xi_k)_+ = t - \xi_k$ if $t > \xi_k$ and zero otherwise. For our problem the t's can be either assumed to represent our time observation or the phased time points, which we formally discuss in Section 2.3, depending on the data and the method used. The simplest way to introduce periodicity is by replacing time with phase. We applied this method for our data, and for that example it performed well but was inferior to methods based on sinusoidal models, the results analytically can be found in Froebrich et al. (2021) [24].

Kernel smoothing regression

A different period estimation approach with fewer assumptions regarding the shape of the data is based on using kernel smoothing regression as studied in great depth in Hall et al. (2000) [27], where the estimates for $\hat{g}$ are obtained using the Nadaraya-Watson estimator (defined below) for a given period and assuming some kernel function. The parameters for g and for a range of candidate periods p are estimated in separate steps, and they propose as $\hat{p}$ the period which corresponds to the minimal sum of squared errors. One could think of the corresponding periodogram function here comprising of values of sums of squared errors for various p.

Analytically, Kernel regression is a non parametric modelling approach where we estimate the fitted values allocating weights to our observations according to a kernel $K((t_j - t_k)/h)$, where h is a constant determining how smooth our estimate of g will be. There is a wide range of kernel choices, the most common one is the normal density (also called Squared exponential kernel). The idea is that when it comes to estimating $g(t_j)$ the estimation is based on values close to t_j since the further away we are from it the smaller the value of the kernel becomes.

A commonly used kernel regression estimator is the Nadaraya-Watson and the fitted values are calculated according to:

$$\hat{g}(t_j) = \frac{\sum_k y_j K(t_j - t_k)}{\sum_k K(t_j - t_k)} \tag{2.14}$$

Notice that the fitted values are a linear combination of our observed values. As we mentioned before the most commonly used kernel is the Normal density. In our problem however it would make more sense to use a periodic kernel instead that we formally define and discuss in the following Section. If a non periodic kernel is used, then we can estimate the period using the phased data as discussed in Section 2.3.

Gaussian process regression

The last non parametric approach we will discuss in this Chapter is based on Gaussian process regression as seen in Wang et al. (2012) [72]. A prior over functions g is assumed to follow a Multivariate Normal distribution $\mathbf{g} \sim N(\mathbf{0}, \mathbf{K})$, where $\mathbf{K}$ is an $n \times n$ covariance matrix in our case, with entries calculated as

$$\mathbf{K}_{jk} = K(t_j; t_k) = A \exp\left(-2\sin^2\left(\frac{\pi}{p}(t_j - t_k)\right)/h^2\right) \quad j,k = 1,2,\cdots,n \tag{2.15}$$

and A denoting the amplitude, p the period and h a smoothing parameter. This particular choice of covariance function given by (2.15) is made in order to introduce periodicity in the model. For each candidate period p the corresponding posterior distribution of the fitted function $\boldsymbol{g}$ at the time points t_j is Multivariate Normal, with mean and covariance:

$$E(\mathbf{g}) = \mathbf{W}\mathbf{y} \quad \mathrm{Cov}(\mathbf{g}) = \mathbf{K} - \mathbf{W}\mathbf{K} \tag{2.16}$$

where $\mathbf{W} = \mathbf{K}[\mathbf{K}+\sigma^2\mathbf{I}]^{-1}$ and with σ^2 denoting the variance of the errors (Williams and Rasmussen (2006) [73], Wang et al. (2012) [72]). It is worth mentioning that the method of splines regression is a special case of GPR when a specific covariance function is selected instead of (2.15) as shown in Kimeldorf and Wahba (1970) [40], see also Williams and Rasmussen (2006) [73] for an analytic discussion to that end. The parameters are estimated from the marginal likelihood (2.17) for a given period p and similar to pseudo-likelihood methodology the optimal p is chosen to be the maximizer of the resulting marginal likelihood:

$$\log P(\mathbf{y}|A, h, \sigma^2) = -\frac{1}{2}\mathbf{y}^\top(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K} + \sigma^2\mathbf{I}| - \frac{n}{2}\log 2\pi \tag{2.17}$$

We discuss this in more detail in Section 3.5. The authors in Wang et al. (2012) [72] claim that their method outperforms the LS periodogram for data that deviate from the sinusoidal assumption, in terms of estimating the correct period. They also emphasize that a formal statistical test is needed to determine whether the estimated/proposed period can be considered valid. We address this in Chapter 3 where we propose such a testing method.

Note that if individual measurement accuracies s_j are to be taken into account, a simple update is needed for using a weighted Gaussian process regression model (See for example Shenf et al. (2017) [59]) with, $\mathbf{W} = \mathbf{K}[\mathbf{K} + \mathbf{Q}^{-2}\sigma_n^2\mathbf{I}]^{-1}$ where $\mathbf{Q}$ is a diagonal matrix with the weights $q_j = 1/s_j$ in its diagonal as,

$$\mathbf{Q} = \begin{bmatrix} q_1 & 0 & 0 & \dots & 0 \\ 0 & q_2 & 0 & \vdots & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & & \dots & q_n \end{bmatrix} \tag{2.18}$$

Under this setting we can assume that our observations are generated by the following Normal distribution:

$$y \sim N(0, \mathbf{K} + \mathbf{Q}^{-2}\sigma_n^2\mathbf{I}) \tag{2.19}$$

We can then estimate our parameters by maximizing the marginal likelihood as discussed before which is in this case will be of the form:

$$\log P(\mathbf{y}|A,h,\sigma^2) = -\frac{1}{2}\mathbf{y}^\top(\mathbf{K}+\mathbf{Q}^{-2}\sigma^2\mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}+\mathbf{Q}^{-2}\sigma^2\mathbf{I}| - \frac{n}{2}\log 2\pi \quad (2.20)$$

Note that the setting of the Gaussian process regression method as described above and the derivation of the corresponding results (e.g Eq. (2.16)) are based on Bayes' theorem and developed from a Bayesian perspective. The posterior over functions $\mathbf{g}$ and thus the marginal likelihood (3.12), however, are available in a closed form. It is common then, in that case, to perform inference for the parameters of interest through the frequentists point of view by maximizing the marginal log-likelihood or minimizing the leave-one-out cross-validation error (more details in Section 2.2.3), instead of assigning hyper-priors and performing Markov chain Monte Carlo (MCMC) simulations. In order to ease computations, we adopt the former strategy throughout this thesis.

2.3 A discussion on folded time series

The methods we discussed in this Chapter involved fitting a statistical model in order to estimate p. In the linear regression method for example we fit model (2.11) and with t, as we already stated, we denote the actual time points that our observations occurred. It is common however to use the phased time points instead, which can be obtained by implementing the following transformation for a time point t_i and for a chosen period p_j

$$f_{ij} = \frac{t_i}{p_j} - \lfloor\frac{t_i}{p_j}\rfloor \in [0,1), \quad i = 1,...,n \quad j = 1,\dots,J \quad (2.21)$$

or,

$$f_{ij} = t_i - p_j\lfloor\frac{t_i}{p_j}\rfloor \in [0,p_j), \quad i = 1,...,n \quad j = 1,...,J \quad (2.22)$$

Where with $\lfloor a \rfloor$ we denote the first integer smaller than or equal to a and with J we denote the number of periods in our period search grid. In fact, in all methods

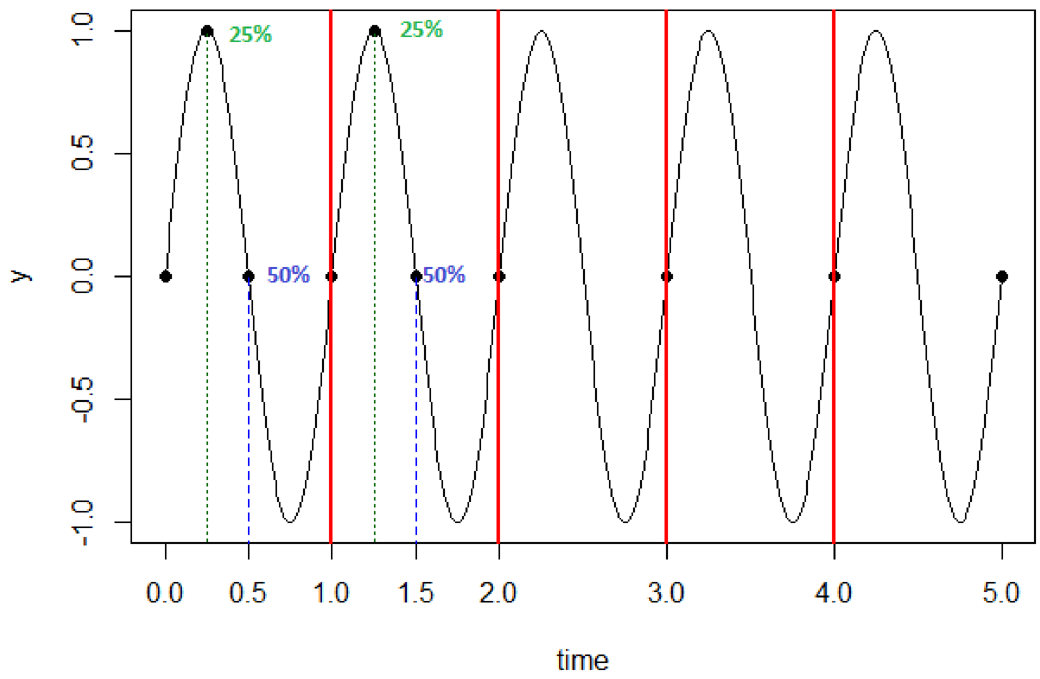


FIGURE 2.6: Understanding phase on a simple sine wave with period = 1.

described above we could use either t or f. Note that there is no practical difference between the two equations, since Eq. (2.22) is obtained by multiplying Eq.(2.21) by p_j.

These phased data points are essentially percentages of the data's cycle, and it can be thought of as folding our curve into its period p. To understand better how it works see Figure 2.6. This image shows a sine wave with period $p = 1$. The vertical lines show the start of each cycle (which is of length 1). The small vertical red lines show the points at which half (50%) of the cycle has been completed, and the blue lines show 25% of the cycle. These are our phased points. For example using equation (2.21) we can calculate the phase of time point $t_i = 0.5$, as $f_i = \frac{0.5}{1} - \lfloor\frac{0.5}{1}\rfloor = 0.5$ or for $t_k = 1.5$ where $f_k = \frac{1.5}{1} - \lfloor\frac{1.5}{1}\rfloor = 0.5$. So both data points t_i, t_k correspond to the 50% of the cycle.

This view of the data can be useful in many cases since for example, it can make

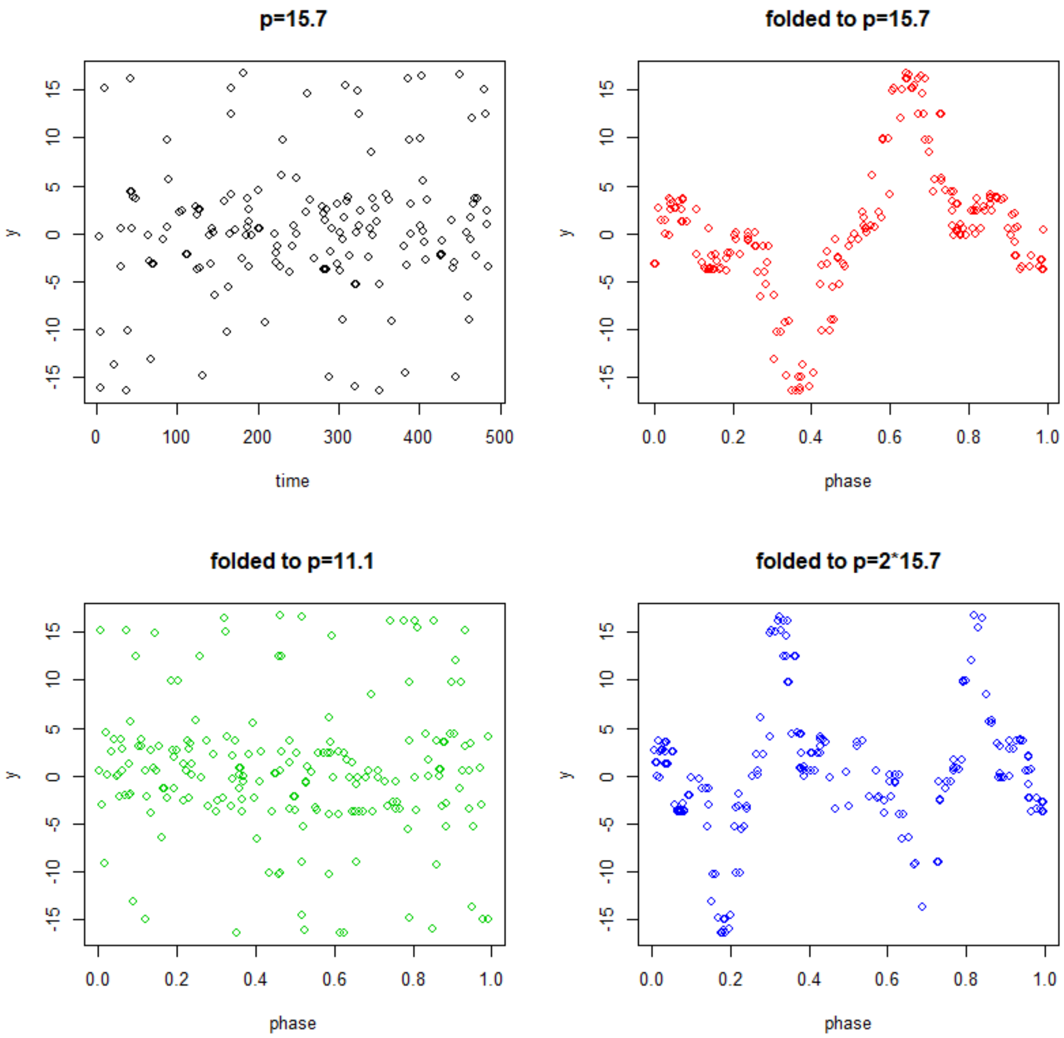


FIGURE 2.7: A simulated light curve folded to different trial periods. We notice that in the first plot the periodic behaviour of the curve is not clear. Once we fold into the correct period the periodic shape becomes apparent. Folding the data to the wrong period results to no pattern. Folding into multiples of the periodicity can also make the periodic shape apparent, making it sometimes difficult to recognize the correct period (see bottom right).

more dense our sample size. Furthermore it is an easy way to bring in the periodicity term in models that do not naturally contain it. The idea is that if we fold the data into the correct period, our phased light curves will "reveal" its periodic shape as in Figure 2.4. If we fold into the wrong period however no pattern should emerge and the data should look like noise. There can be of course intermediate cases, for more details see Figure 2.7.

While using the folded data can be useful in some cases, it can also bring in

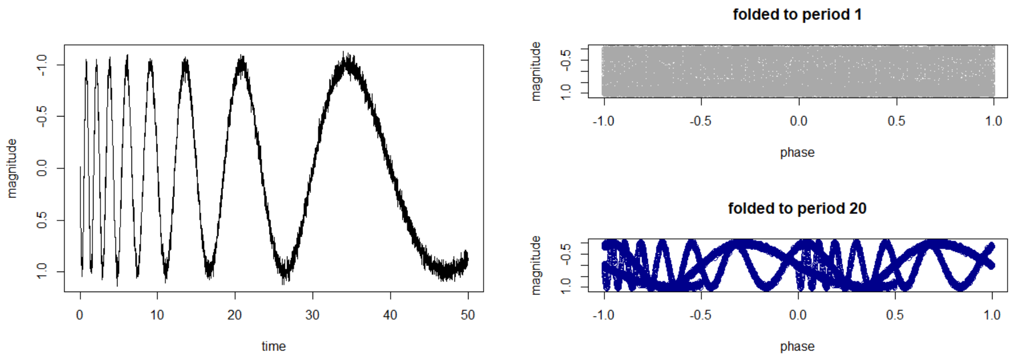


FIGURE 2.8: (Left)A sinusoidal light curve with increasing period. (Right) The light curve folded into period 1 and period 20.

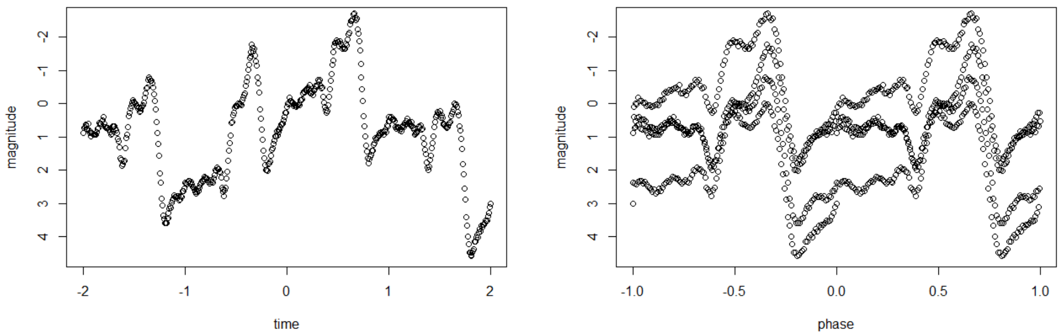


FIGURE 2.9: (Left)A sinusoidal light curve with period 1 and amplitude change. (Right) The light curve folded into period 1.

additional noise in some others and therefore reduce the accuracy of our results. This might be the case if the data are not strictly periodic, for example the period could change over time. In this case by folding our data we miss this information. See Figure 2.8 (Left) where we simulate a light curve with period p that gradually gets larger. Folding this curve in period p for example results into the noisy image seen in Figure 2.8 (Right). Another example is when the period stays the same but our data's shape changes over time. Folding again into the correct period will result into a non informative picture as seen in Figure 2.9 and Figure 2.10. In cases like these using the actual time observations t can be proven more beneficial. In Chapter 4 we explore some possible ways of estimating the period in such cases.

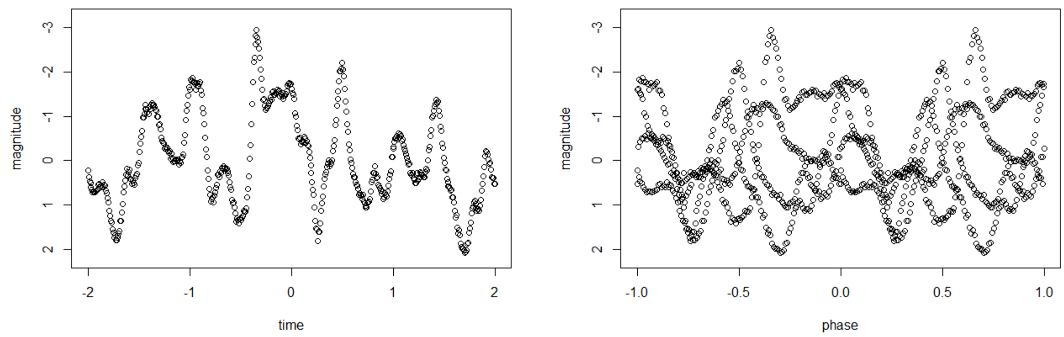


FIGURE 2.10: (Left)A sinusoidal light curve with period 1 and shape change. (Right) The light curve folded into period 1.

2.4 Multiple testing

In Section 2.2.1 we saw how we can obtain a periodogram based on a goodness of fit measure which we evaluate for different periods p. As discussed earlier, a common measure of that form is the well-known F-statistic (2.12). These periodograms always have a maximum peak regardless of the existence or not of signal in the data. We need, thus, to be able to decide if a periodogram peak represents a true periodic signal or it is the product of purely noisy data.

A standard approach to answer that question is based on statistical hypothesis testing. The basic idea is to quantify how likely it is to observe a particular periodogram value at period p if the data consists of only noise. This can be achieved by testing the null hypothesis of purely noise data, against the alternative that a periodic signal exist at period p. In order to perform inference we need to know the distribution of the statistic used under the null hypothesis. In the case of the standard F-statistic (2.12) for example, we know that under the null hypothesis it follows an F distribution, see for more details Section 3.3. This hypothesis testing approach has been also used in the Astronomy literature for this exact problem, see for example Schwarzenberg (1998) [57] where the author derives the theoretical distribution of periodograms, based on this type of goodness of fit statistics, or Thieler et al. (2013) [63] where the authors introduce an empirical test based on the Beta distribution. These methods are discussed analytically in Chapter 3 along with our proposed tests under non-parametric regression settings.

At this point it is important to note that in order to decide which periods are valid or not, we conduct the tests mentioned above multiple times (e.g the number of trial periods tested). This is something we should keep in mind when deciding the significance level α that is the probability of false rejection, or in other words, the probability to identify a period as true when it is not. If for example we fix $a = 0.05$ then that means that the chance to falsely detect a trial period as true is only 5%. But, if we repeat this test 50 times (because this is the number of trial periods we test for example) then the probability of getting at least one false rejection is 92.3%. This probability is also known as the family wise error, which for the case of J independent tests is calculated as:

$$1 - (1 - \alpha)^J$$

There are many methods for correcting the significance level for multiple comparisons, the most well known is probably the Bonferroni correction where the corrected significance level is now calculated as $\alpha' = \alpha/J$, this approach has been criticized as being over conservative and many other corrections have been proposed for example in Sidak (1967) [60] or Good (1958) [26]. The above mentioned methods depend on J the number of tests but in many cases this can be proven problematic, especially when the tests are not independent from each other. This is the case for the period estimation problem too when we try to discretize the frequency spectrum. An analytic discussion regarding this problem can be found in Algeri et al. (2016) [2]. Here we will not pursue this any further and instead we will adopt the approach described in Theiler et al. (2013) [63] where we look at the local maxima or minima of the periodogram. We will also assume that the significance level α is corrected for multiple comparisons.

2.5 Summary

In this Chapter we discussed different methods proposed in the literature for estimating the period of irregularly sampled times series. First, we saw how methods developed for equally-spaced data, such as the classic periodogram which is based on the Fourier Transform, fail under irregularly sampled data schemes. The most popular method designed for this purpose is the Lomb-Scargle periodogram which is connected to the Fourier transform but does not get affected by the sampling. In fact, it is shown that it returns identical results to fitting a simple sinusoidal model with least squares regression and using a chi-squared type goodness of fit. This connection to least squares regression led to the development of a series of empirical periodograms based on different choices of model, regression, goodness of fit statistic etc.

Specifically, the choice of model reflects our assumptions about the shape of the periodic signal, some of the methods developed assumed a sinusoidal, triangular or splines model for instance. The motivation behind the use of different regression methods, such as absolute deviations regression, or Huber's regression, was to obtain robust results under the presence of outliers or data with fat tails. More flexible methods that make fewer assumptions about the data have also been proposed, like the use of kernel smoothing regression or Gaussian process regression (GPR). We mostly focus on the latter model here and provide more details about GPR model and their extensions in Chapter 4. Finally, we briefly mentioned different choices of goodness of fit measures (e.g. F-statistic, Cross-validation, R^2 etc.) which are connected to hypothesis testing methods for deciding if an estimated period represents a true signal or not. This subject is covered analytically in Chapter 3 along with our suggested tests under non-parametric model settings.

We should note at this point that there is no obvious choice as to which method is better. Different approaches perform differently depending on each situation. For example, if outliers are present in the data, robust regression methods return more reliable results compared to least squares but they are compositionality more expensive and have less convenient mathematical properties. Non parametric methods are more flexible, but are again computationally demanding and provide less accurate results than using a linear model if, for some reason, the shape of the data is known. Finally, the choice of goodness of fit depends heavily on the choices of regression and model. In Chapter 6 we discuss how to choose between the different methods depending on the data available each time.

Chapter 3

Hypothesis Testing for period detection

3.1 Motivation

In the previous Chapter we discussed some popular approaches for period estimation by obtaining a periodogram. Extreme periodogram peaks are an indication of potential periods, we need however a quantifiable measure as to what is considered extreme. In other words, every periodogram will always have a maximum (or minimum) peak even if the data is just noise. See for example Figure 3.1, where we generated three light curves consisting of just noise and produced their periodograms as described in Section 2.2.1. The first light curve was generated with white noise and we see a clear peak at 7.2 days. For the next two light curves we used red noise (a correlated type of noise that we will discuss formally in Section 3.5), for these cases, it is also important to notice the maximum value that the periodogram reaches, which is calculated to be 0.3. This relatively high peak becomes even more interesting if we have a look at the last periodogram of a light curve generated with signal this time and a known period at 1.9 days. We see that the value of the highest peak of the periodogram in this case is equal to 0.06, much smaller than some of the cases with no signal.

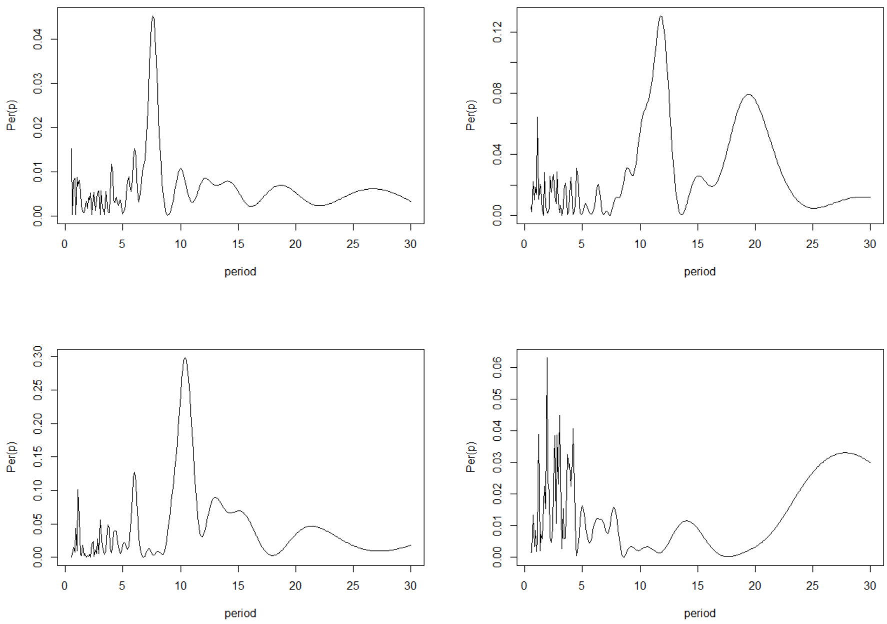


FIGURE 3.1: (Top Left) Periodogram of a light curve consisting of only white noise ($N(0,1)$). (Top Right) The periodogram of a light curve consisting of only red noise ($\rho = 0.5$) (Bottom Left).The periodogram of another light curve with no signal and just red noise ($\rho = 0.9$) (Bottom Right)The periodogram of a light curve generated with a sinusoidal signal and period at 1.9 days. The sampling for all the examples is borrowed from a randomly chosen real light curve.

Thus, it is important to be able to distinguish between real periodic signals and noise fluctuations and the problem is not trivial. The focus of this Chapter is determining whether a periodogram peak represents a valid period or not and in the next Sections we will see how we can answer this question using hypothesis testing.

3.2 The general framework

The common approaches to this problem are based on running hypothesis tests on a sequence of trial periods and relating them to some periodogram entries. In particular, each periodogram entry could be the value of a chosen goodness of fit test statistic in our case, and we use the corresponding p-value to construct the hypothesis testing whether the given statistics are generated from just noise or period bearing data. In other words, the peaks in the corresponding periodogram point to those candidate periods p which are the most likely to generate the data and their corresponding magnitude of the p-values of the goodness of fit statistics indicate their associated credibility. The idea of sequential hypothesis testing for each candidate period p is summarized as follows:

$$\begin{aligned} &H_0 : \mathcal{M}_0 \text{ model, no signal in the data} \\ &\quad \text{vs} \\ &H_1 : \mathcal{M}_1 \text{ model with signal of period } p \end{aligned} \tag{3.1}$$

This strategy is adopted by the astronomy literature as in Schwarzenberg-Czerny (1998) [57], where the distribution under the assumption of model $\mathcal{M}_0$ is derived for empirical periodograms like LS. In many papers the sequential hypothesis testing is based on simulating data according to H_0 as seen in Do et al. (2009) [18], or Halpern et al. (2003) [28] and then comparing their empirical distribution to the observed values of the periodogram peaks.

Note that these tests belong to the general family of pseudo-likelihood ratio tests (using the residual sum of squares) that are based on the relative performance of both models $\mathcal{M}_0$ and $\mathcal{M}_1$. This approach is in fact similar to that applied to the standard model selection in linear regression for nested models. Similar use is seen in time series analysis as shown in Berenblut and Webb (1973) [9] where the authors compare a simple linear model to that with additional AR(1) errors and the resulting statistic is related to that of Durbin and Watson (1950) [19]. More specifically, they show that in this context the one sided hypothesis testing is approximately uniformly most powerful (UMP), see also Paoella (2018) [51] for an analytic discussion to that end. In non-parametric regression as seen in Azzalini and Bowman (1993) [5] a similar test is used for testing the linearity assumption in the data versus non linear models $\mathcal{M}_1$ produced by kernel smoothing regression or spline function families.

Note that similar to the standard F-tests used in the linear regression, the associated test statistics for the above mentioned tests are in fact constructed as ratios of residual sums of squares which correspond to quadratic forms of zero mean normal components. The added difficulty of including our models of interest is that for each candidate p, the terms in the numerator and denominator of those ratios are not necessarily independent. Hence, the standard F-test statistics can not be generally applied. A typical example is that in time series for nested ARMA models where standard results for the F-test no longer apply. This dependence is not fully addressed in the context of this particular problem. Typically simulation based approaches are implemented to overcome such difficulties.

In our context, we will perform hypothesis testing as in (4.10) such that the corresponding model for $\mathcal{M}_0$ would be just a constant, namely no periodic behaviour in the data i.e. the light curve is just noise and function g from (2.1) is the mean. This is equivalent to the model (2.11) reduced to the intercept term β_0. On the other hand, if there was a periodic signal, the test will suggest as appropriate the alternative fitted models $\mathcal{M}_1$ that will contain the periodic terms as in (2.11) (or a more general form as in non-parametric regression). As a result, the stronger the evidence of departure from $\mathcal{M}_0$ towards $\mathcal{M}_1$ for some period p, the stronger the claim that this period is present in the signal.

Note that a natural goodness of fit measure among various regression models is based on comparing the residual sums of squares RSS_0 and RSS_1 for the null and alternative models as:

$$F = \frac{(RSS_0 - RSS_1)}{RSS_1} = \frac{\mathbf{y}^\top \mathbf{M}_0 \mathbf{y} - \mathbf{y}^\top \mathbf{M}_1 \mathbf{y}}{\mathbf{y}^\top \mathbf{M}_1 \mathbf{y}} \tag{3.2}$$

with matrices $\mathbf{M}_0$ and $\mathbf{M}_1$ depending on the models of interest. Note that any linear smoother could be used for the alternative model as long as periodicity can be introduced. In order to perform hypothesis testing, we need to evaluate the distribution of F, namely $P(F \leq f)$ under the null hypothesis. In Azzalini and Bowman (1993) [5] we see that if $\mathbf{M}_0 = \mathbf{I} - \mathbf{1}\mathbf{1}^\top/n$ is some centring matrix with $\mathbf{1}$ denoting the column vector of ones, then

$$P(F > f) = P(\hat{\boldsymbol{\epsilon}}^\top (\mathbf{I} - (1+f)\mathbf{M}_1)\hat{\boldsymbol{\epsilon}} > 0) \tag{3.3}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix and $\hat{\boldsymbol{\epsilon}} = \mathbf{M}_0\mathbf{y}$. If we replace f in (3.3) with a specific observed value of our statistic, e.g t_{obs}, the probability $P(F > t_{obs})$ is the p-value for our generalised F-test.

3.3 Some remarks on linear regression

In the case of ordinary linear regression (2.11) a periodic structure is assumed and the statistic (3.2), if appropriately scaled, coincides with the ordinary F-statistics. The rescaling constant $c = \frac{n-m_1}{m_1-m_0}$ depends on the respective numbers of parameters m_0 and m_1 for models $\mathcal{M}_0$ and $\mathcal{M}_1$. Namely, under the null hypothesis assumption of white noise, the statistics

$$F \cdot c \sim F(m_1 - m_0, n - m_1) \tag{3.4}$$

with the numerator and denominator terms of (3.2) being independent random variables following chi-square distributions with $m_1 - m_0$ and $n - m_1$ degrees of freedom. Here, $\mathbf{M}_0 = \mathbf{I} - \mathbf{1}\mathbf{1}^T/n$ is a fixed centring matrix, $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ with $\mathbf{X}$ being the design matrix of our model depending on observation time points. Note that from the standard calculations of the projection matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ in regression we can easily see that $\mathbf{M}_1 = (\mathbf{I} - \mathbf{P})^\top(\mathbf{I} - \mathbf{P}) = \mathbf{I} - \mathbf{P}$ and in particular the vector $\mathbf{1}$ is in the null space of $\mathbf{M}_1$. This implies that $\mathbf{y}^\top\mathbf{M}_1\mathbf{y}$ is invariant of the intercept in the model $\mathcal{M}_0$ and therefore the statistic (3.2) is not affected if the data points $\mathbf{y}$ are centralised, namely replaced by residual vector $\hat{\boldsymbol{\epsilon}} = \mathbf{M}_0\mathbf{y}$. Note that under this setting, the eigenvalues of $\mathbf{M}_1$ can be either zero or one and that is why (3.4) holds. The corresponding p-values are immediately available from standard packages like R.

It is worth mentioning that the popular R^2 statistic is a linear combination of the standard F statistic and thus can be written in a similar form as a ratio of quadratic forms as seen below,

$$R^2 = \frac{RSS_0 - RSS_1}{RSS_0} = \frac{F \cdot (m_1 - m_0)}{F \cdot (m_1 - m_0) + (n - m_1)} = \frac{\mathbf{y}^T\mathbf{M}_0\mathbf{y} - \mathbf{y}^T\mathbf{M}_1\mathbf{y}}{\mathbf{y}^T\mathbf{M}_0\mathbf{y}}$$

It is known that if $X \sim F(d_1, d_2)$ then $\frac{d_1 X/d2}{1+d1X/d2} \sim \text{Beta}(\frac{d_1}{2}, \frac{d_2}{2})$. So R^2 under the null hypothesis follows a $\text{Beta}(\frac{m_1-m_0}{2}, \frac{n-m_1}{2})$ (Jonson and Kotz (1970) [37]).

3.4 Non OLS regression

In situations where more general alternatives are explored for $\mathcal{M}_1$, the structure of $\mathbf{M}_1$ does not lead to the standard F-distribution for the statistic (3.2) and hence it is not straightforward to obtain its p-values. Naive Monte Carlo methods are adopted for these cases, and have been extensively used in the relevant literature such as Do et al. (2009) [18], Halpern et al. (2003) [28], Ivezic et al. (2019) [35] where for each trial period say p, many noise curves are simulated and the corresponding p-value is estimated based on the corresponding empirical distribution of the F-statistic (or something equivalent). In fact, we can show that for these cases there is no need to run simulations as the p-values are easily evaluated using the saddlepoint approximation to the distribution of F, which is formally introduced in Section 3.6. A computationally efficient alternative is important given the large amount of data available. Our method improves drastically the calculation time when non parametric models are used, for Gaussian process regression for example, by performing the calculations in 2% of the time needed for doing the same analysis using Monte Carlo. For more details on time comparisons see Table 4 in Section 3.8.1.

3.4.1 Generalized F-test for non-parametric periodograms

In this Section we consider more complicated models $\mathcal{M}_1$ to describe our alternative hypothesis and see how the F-statistic (3.2) can be used for non-parametric settings such as kernel or Gaussian process regression (and in general for any linear smoother). The statistic (3.2) has been used in Azzalini and Bowman (1993) [5] as a linearity test, where a comparison was conducted between a linear and a non-parametric model.

By following the same idea, we compare a constant (as the null model) with a non-parametric model for a given period. Under this setting, while the centring matrix $\mathbf{M}_0 = \mathbf{I} - \mathbf{1}\mathbf{1}^\top/n$ remains the same, $\mathbf{M}_1 = (\mathbf{I} - \mathbf{W})^\top(\mathbf{I} - \mathbf{W})$ is not a projection matrix, since $\mathbf{W}$ is a $n \times n$ matrix depending on the observation times and additional model parameters as seen for example in (2.16) and (2.15).

The appropriate period detection test of the optimal periodogram entries requires the evaluation of the distribution of F under the null hypothesis, which due to $\mathbf{M}_1$ can no longer be assumed to follow a standard F distribution. We should note again that this test is also depending on the alternative model through $\mathbf{M}_1$. The probability of interest $P(F > t_{\text{obs}})$, is given by (3.3) and thus the problem reduces to that of evaluating these values for the distribution of such Quadratic form of some normally distributed terms.

3.4.2 CVF-test for Gaussian process regression

In the previous Section we discussed how the generalised F-test can be used to create the corresponding periodogram for detecting valid periods when $\mathcal{M}_1$ is some non-parametric model and the resulting matrix $\mathbf{M}_1$ is not a projection matrix any more. Here we will look more specifically at the explicit expressions for $\mathbf{M}_1$ when the non-parametric model used is the Gaussian process regression (GPR) and the natural goodness of fit measure based on cross-validation score is used.

As seen in Section (2.2.3), in order to estimate the parameters for the GPR model (including the period), we can either maximize the marginal likelihood or minimize the cross-validation error. We can build a statistic/periodogram based on the leave-one-out cross-validation error

$$CVE = \sum_{j=1}^{n} \left(y_j - g_{-j}(t_j)\right)^2 \tag{3.5}$$

where g_{-j} is function g estimated excluding the jth data point. We can use this measure for comparing the predictive ability of the GPR model at period p with that of the mean $\mathcal{M}_0$. As seen in Williams and Rasmussen (2006) [73], each term in (3.5) can be simplified to

$$y_j - g_{-j}(t_j) = \frac{[(\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} y]_j}{[(\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1}]_{jj}} \tag{3.6}$$

and in the vectorized form these entries are shown to be $\mathbf{y}^\top(\mathbf{K} + \sigma_n^2 I)^{-1}\mathbf{B}_2$ where $\mathbf{B}_2$ is a diagonal matrix,

$$\mathbf{B}_2 = \begin{bmatrix} 1/d_1 & 0 & \dots & 0 \\ 0 & 1/d_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \dots & & 1/d_n \end{bmatrix} \tag{3.7}$$

such that d_j is the jth diagonal element of the inverse of the matrix $\mathbf{K} + \sigma_n^2 \mathbf{I}$, namely, $d_j = diag((\mathbf{K} + \sigma_n^2)^{-1})_j$. Therefore, by denoting as $\mathbf{B} = (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1}\mathbf{B}_2$, we can write CVE as a quadratic form in $\mathbf{y}$ as

$$CVE = \sum_{j=1}^{n} \left(\left(\frac{[(\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1}]_j}{d_j} \right) \cdot y_j \right)^2 = \mathbf{y}^\top \mathbf{B}\mathbf{B}^\top \mathbf{y} \tag{3.8}$$

In the same spirit as (3.2), a similar goodness of fit statistic, which we will refer to as cross validation F (CVF), based on the cross validation errors for the null model and that for a particular choice of g using GPR is given below as:

$$CVF = \frac{CVE_0 - CVE_1}{CVE_1} = \frac{\sum_{j=1}^{n}(y_j - \bar{y}_{-j})^2 - \sum_{i=j}^{n}(y_j - g_{-j}(t_j))^2}{\sum_{j=1}^{n}(y_j - g_{-j}(t_j))^2} \tag{3.9}$$

where $\bar{y}_{-j}$ denotes the j^{th} leave-one-out mean. It can be easily shown that the CVF-statistic can be written in the form of (3.2) where $\mathbf{M}_1 = \mathbf{B}\mathbf{B}^\top$ and $\mathbf{M}_0 = \mathbf{M}\mathbf{M}^\top$, is a matrix used for the leave one out mean terms:

$$\mathbf{M} = \begin{bmatrix} 1 & -\frac{1}{n-1} & \dots & -\frac{1}{n-1} \\ -\frac{1}{n-1} & 1 & & \vdots \\ \vdots & & \ddots & \\ -\frac{1}{n-1} & \dots & & 1 \end{bmatrix} \tag{3.10}$$

The problem of evaluating the distribution of the CVF-statistic reduces again to that of estimating a simple Normal Quadratic form, and Saddlepoint approximation can be implemented in order to obtain fast and accurate results.

3.5 Departure from white noise

For all the methods described so far we assumed white noise residuals. In many problems however the residuals might not be i.i.d. and additional structures should be assumed in order to capture their behaviour. In astronomy for example it is common for light curves to be contaminated with red noise as seen in Vaughan (2003) [69], where additive systematic errors are assumed to be generated by a power-law spectrum, see also Timmer and Konig (1995) [65] on a method on generating such data. An example of different types of noises can be found in Figure 3.2. Specific interest is shown in period detection under the presence of red noise. The assumed correlation structure of red noise coincides with that of an AR(1) process if the observations are collected at equally spaced time observations, see e.g. Von Storch and Zwiers (2001) [70]. In these cases of AR(1) errors, (not of our primary interest here), Benlloch et al. (2001) [8] identified valid periodogram peaks against red background noise using Monte Carlo methods. Later some exact tests were developed, see for example Vaughan (2005) [68]. In the case of unequally spaced time series however most of the corresponding methods used are based on Monte Carlo approaches in order to identify the valid periods in the presence of a red background noise, for example Zhou and Sornette (2002) [74].

3.5.1 The generalised F-test for red noise

In this Section we consider the situation when a particular correlated noise (red noise) structure is assumed for the residual terms ϵ. Let us assume for example that the correlation structure between any pair of residuals observed at a single time unit apart is $\mathrm{Corr}(\epsilon_j, \epsilon_{j-1}) = \rho$ for any j and some $\rho \in (-1, 1)$. This in turn

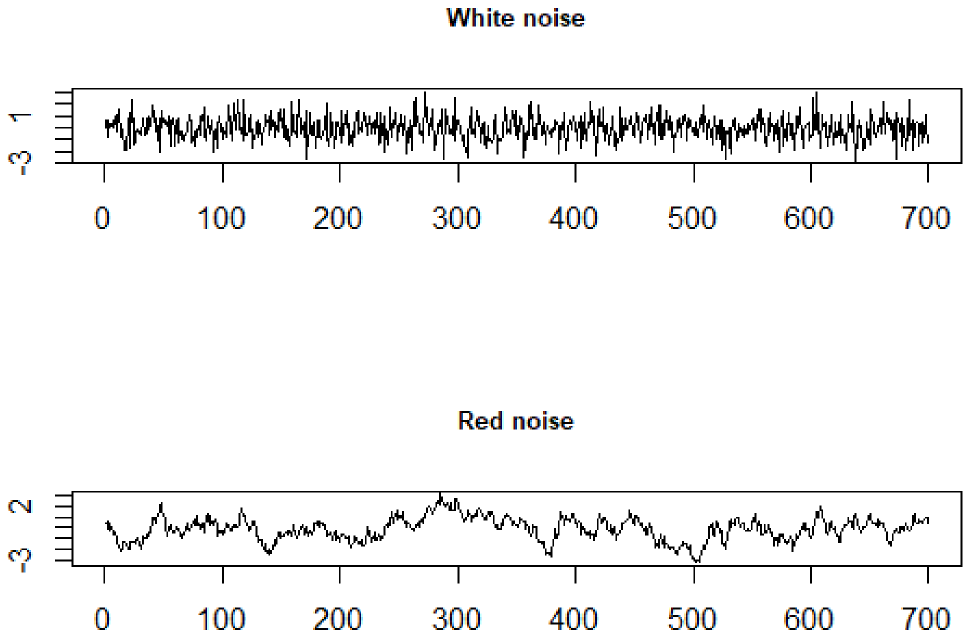


Figure 3.2: Examples of different types of noise.

implies that $Corr(\epsilon_j, \epsilon_{j-2}) = \rho^2$ and in general for any pair of residuals observed at time units t_j and t_k, $\mathrm{Corr}(\epsilon_{t_j}, \epsilon_{t_k}) = \rho^{|t_j - t_k|}$. The red noise terms are distributed as:

$$\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{C}_\rho) \quad \text{where} \quad \mathbf{C}_\rho = \begin{bmatrix} 1 & \rho^{|t_1-t_2|} & \rho^{|t_1-t_3|} & \dots & \rho^{|t_1-t_n|} \\ & 1 & \rho^{|t_1-t_2|} & \vdots & \rho^{|t_2-t_n|} \\ \vdots & & \ddots & & \\ & & & & \\ & & & \dots & 1 \end{bmatrix} \tag{3.11}$$

Note that the matrix $\mathbf{C}_\rho$ takes into account the irregularly sampled nature of the data, such that the correlation between the observations depends on their time distance, the further away two points are, the less the correlation assumed for the corresponding residuals. If $\rho = 0$, $\mathbf{C}_\rho = \mathbf{I}$ then the model reduces to that of the white noise. In order to calculate (3.2) for various p, we need to estimate the model parameters of $\mathcal{M}_1$. This, for the case of Gaussian process regression, is achieved by maximizing the marginal likelihood or alternatively by minimizing the leave-one-out cross-validation error. Under the presence of red noise, the marginal likelihood for GPR is:

$$\log P(\mathbf{y}|A, h, \sigma^2, \rho) = -\frac{1}{2}\mathbf{y}^\top(\mathbf{K} + \sigma^2\mathbf{C}_\rho)^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K} + \sigma^2\mathbf{C}_\rho| - \frac{n}{2}\log 2\pi \quad (3.12)$$

For constructing (3.2) for the non-parametric models we take $\mathbf{M}_0 = \mathbf{C}_\rho^{1/2}(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\mathbf{C}_\rho^{1/2}$ and $\mathbf{M}_1 = \mathbf{C}_\rho^{1/2}(\mathbf{I} - \mathbf{W})^\top(\mathbf{I} - \mathbf{W})\mathbf{C}_\rho^{1/2}$. Note that the matrix $\mathbf{M}_1$ can be similarly adjusted accordingly for linear models too. In the case of equally spaced data $\mathbf{C}_\rho$ reduces to the standard AR(1) correlation structure. In a similar manner as in Section 3.4.1 and Section 3.4.2, the p-value $P(F > t_{\text{obs}})$ can be calculated as seen in (3.3). The same principles can be applied in order to adjust the CVF-statistic for the correlated noise assumption. See also Section 3.8.2 for a comparison between red noise and white noise Gaussian process regression for simulated light curves with additive red noise.

3.6 P-value evaluation for generalised F-test

In this Section we will refer to more technical details regarding the implementation of the above mentioned tests. Note that the corresponding statistics whose p-values need to be generated, are of the general form

$$F = \frac{\hat{\boldsymbol{\epsilon}}^\top \mathcal{A} \hat{\boldsymbol{\epsilon}}}{\hat{\boldsymbol{\epsilon}}^\top \mathcal{B} \hat{\boldsymbol{\epsilon}}}$$

where $\mathcal{A}$ and $\mathcal{B}$ are some symmetric matrices of the same dimension n depending on the model and $\hat{\boldsymbol{\epsilon}}$ represent i.i.d normal error components. For example, in the case of the generalized F-statistic (3.2), $\mathcal{A} = \mathbf{I} - \mathbf{M}_1$ and $\mathcal{B} = \mathbf{M}_1$. The corresponding p-values for some observed value of statistics t_{obs} is therefore

$$P\left(\frac{\hat{\boldsymbol{\epsilon}}^\top \mathcal{A} \hat{\boldsymbol{\epsilon}}}{\hat{\boldsymbol{\epsilon}}^\top \mathcal{B} \hat{\boldsymbol{\epsilon}}} > t_{obs}\right) = P(\hat{\boldsymbol{\epsilon}}^\top (\mathcal{A} - t_{obs}\mathcal{B})\hat{\boldsymbol{\epsilon}} > 0)$$

Clearly, $\mathcal{A} - t_{obs}\mathcal{B}$ is symmetric but not necessarily positive definite, and as shown in Kuonen (1999) [43]

$$P(\hat{\boldsymbol{\epsilon}}^\top(\mathcal{A} - t_{obs}\mathcal{B})\hat{\boldsymbol{\epsilon}} > 0) = P(X = \sum_{i=1}^{l} \lambda_i \chi_1^2 > 0)$$

Where, λ_i are the $l \leq n$ non-zero eigenvalues, not necessarily positive of $(\mathcal{A} - t_{obs}\mathcal{B})$ and χ_1^2 stands for independent χ^2 random variables with 1 degree of freedom. Note that the matrices $\mathcal{A}$ and $\mathcal{B}$, when standard linear regression is used, have only two possible eigenvalues: either 0 or 1. This leads to the ratio of two independent chi-square components with $n - m_1$ and $m_1 - m_0$ degrees of freedom respectively which reduces to the standard F distribution (see Butler (2007) [14], p. 378). As the distribution of X is not generally known in closed form, see Johnson et al. (1995) [36], we need to evaluate these probabilities numerically. There are many methods in the literature for approximating such probabilities, for example numerical integration Imhof (1961) [34] or using the method of matching moments. The moment generating function of the corresponding convolution X is known however and rather than performing the corresponding numerical Laplace inversion, a practically convenient method that we adopt here is that based on saddlepoint approximation as it is numerically efficient for the accuracy that we need to operate. Moreover, saddlepoint approximation has been successfully implemented in many time series applications e.g. Paolella (2018) [51] and more recently in non Euclidean statistics data like shapes and directions e.g. Kume et al. (2013) [41] or Kume et al. (2005) [42]. We adopt the methodology that is suggested in Kuonen (1999) [43] as our hypothesis testing methodology is also similar in nature.

3.6.1 Saddlepoint approximation

Saddlepoint approximation is an approximation method for evaluating a density or a cumulative distribution function given the analytical expression of its moment generating function or the cumulant generating function $\mathcal{K}$ and its derivatives $\mathcal{K}'$ and $\mathcal{K}''$, see Butler (2007) [14] for more details. In fact, our focus here is on the

CDFs, namely $F(x) = P(X > x)$, or $P(F > f)$ specifically for the F-statistic case. Two popular approximations are available for these survival probabilities, as suggested in Lugannani and Rice (1980) [46] and Barndorff-Nielsen (1990) [7]. In particular, the approximation of Barndorff-Nielsen (1990) [7] for $\hat{F}(x)$ is derived as

$$\hat{F}(x) = 1 - \Phi\{w + \frac{1}{w}\log(\frac{u}{w})\} \tag{3.13}$$

where Φ is the standard normal distribution, and w and u are given by:

$$w = \mathrm{sgn}(\hat{s})[2(\hat{s} - \mathcal{K}(\hat{s}))]^{1/2}, \qquad u = \hat{s}(\mathcal{K}''(\hat{s}))^{1/2}$$

The saddlepoint solution is $\hat{s}$ such that

$$\mathcal{K}'(\hat{s}) = x \tag{3.14}$$

In Johnson and Kotz (1995) [36] the authors provide details about normal quadratic forms along with their cumulant function. In our case, we consider the quadratic forms of zero mean normal variables. It can be shown that for some independent random variables, $Z_i \sim N(0,1)$, the corresponding cumulant generating function of the quadratic form $\sum \lambda_i Z_i^2$ is

$$\mathcal{K}(s) = -\frac{1}{2}\sum \log(1 - 2s\lambda_i) \tag{3.15}$$

and

$$\mathcal{K}'(s) = \sum \frac{\lambda_i}{1 - 2s\lambda_i} \quad \text{and} \quad \mathcal{K}''(s) = \sum \frac{2\lambda_i^2}{(1 - 2s\lambda_i)}$$

As can be seen from the expression for $\mathcal{K}'$, at $s = 1/2\lambda_i$ this function is unbounded and therefore the solution of $\hat{s}$ should be carefully chosen. In particular, we have found that the best strategy here is to look for the solution which is the nearest to 0. This implies that the solution of the saddlepoint equation (3.14) can be also negative. See Figure 3.3 as an example of the general behaviour of $\mathcal{K}'$. Some examples regarding the accuracy of saddlepoint approximation for our particular problem can be found in Section 3.8.1.

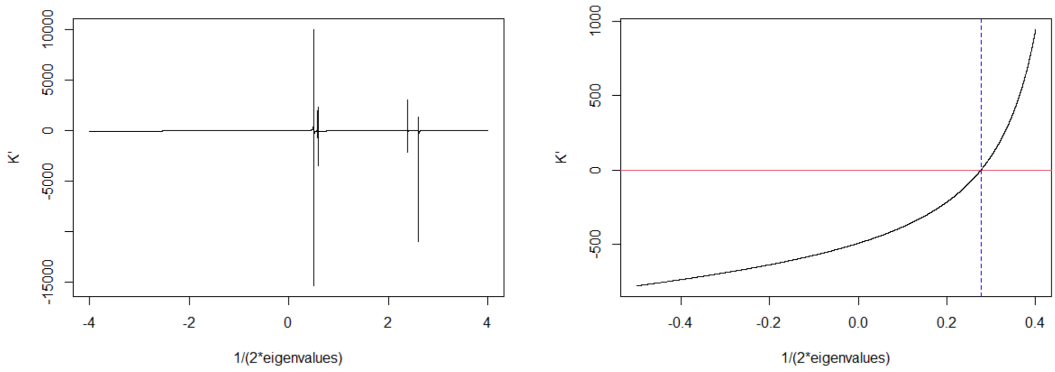


FIGURE 3.3: (Left) An example of the $\mathcal{K}'$ function, we see that its behaviour can be quite problematic making the choice of a suitable range to search for the solution important. (Right) For the same example only the range from the biggest negative to the smallest positive. The blue dashed line shows the solution of the Saddlepoint equation which in this case is 0.2763643.

3.7 Power analysis

In this Section we will examine the performance of the proposed tests. In general we are interested in minimizing type I and II errors. Type I error is the probability to falsely identify a period as correct when no periodicity exists, or in other words $P(\text{reject } H_0|H_0 \text{ true})$. It is usually predetermined in our case, and it is the level of significance α. Type II error is the probability to not identify a correct period , or $P(\text{Do not reject } H_0|H_1 \text{ true}) = \beta$ and $1-\beta$ is the power of the test, that is the probability to identify a period as valid when the period is correct, and it is usually denoted as γ. The higher the value of the power the better the performance of the test.

3.7.1 Example 1: Weighted linear regression

We will estimate γ using simulated light curves with known periodicity at 2.4 days. We mentioned in Section 3.2 that in many cases the distribution of the F-statistic (3.2) no longer follows a standard F distribution. One of these situations is when weighted regression is used. As a first example here we will compare the power for the generalized F-test to that of the standard F-test and to the approach described in Thieler et al. (2013) [63] when weighted linear regression is used. For this example α will be set to $1 - 0.998$. We will simulate 1000 sinusoidal light curves with known period consisting of 200 data points for different signal to noise ratios (SNR), SNR $= \text{var}(y)/\text{var}(\epsilon)$, starting from $SNR = 0.1$. The model used to generate the data is given by (2.2) with g described by (2.11). These light curves will be generated using the package *RobPer* (Thieler et al. (2016) [64]). The results are shown in Figure 3.4 (Left) for the significance level $\alpha = 1 - 0.998$, note that this level is chosen as the multiple tests correction for $\alpha = 0.05$ and 30 repetitions according to Sidak (1967) [60]. We notice that the generalised F-test outperforms the other two while closely followed by the standard F-test. Of course the larger the SNR, the larger the power, meaning that the probability of correct inference, which depends on the estimated p-values, gets larger as the SNR increases.

In order to be able to include the testing approach introduced in Thieler et al. (2013) [63], which we denote here as "RThieler", for each generated light curve we calculated the relevant periodgram for 196 trial periods ranging from 0.5 to 20 and used the *RobPer* package to calculate the relevant critical value. To obtain a clearer view on the test's performance we further calculate the average number of periods identified as correct, for different signal to noise ratios. In this example we are searching for periods from 0.5 to 10 days with 1 decimal accuracy. Note that the simulated light curves were generated with only one actual period, and thus extra periods detected are false positives. The results can be seen in Figure 3.4 (Right), in this plot the number of periods detected should be 1 and thus deviations from that are an indication of poor performance. We also estimated the probability of not rejecting the null hypothesis when it is true based on 1000 iterations of simulating purely noise data. For the chosen significance level this probability was 0 for all three models. We can see that the generalized F-test clearly outperforms the other two, with the standard F-test overestimating the number of correct periods and the RThieler test underestimating it. Overall, from this example we can conclude that using the generalized F-test when weighted

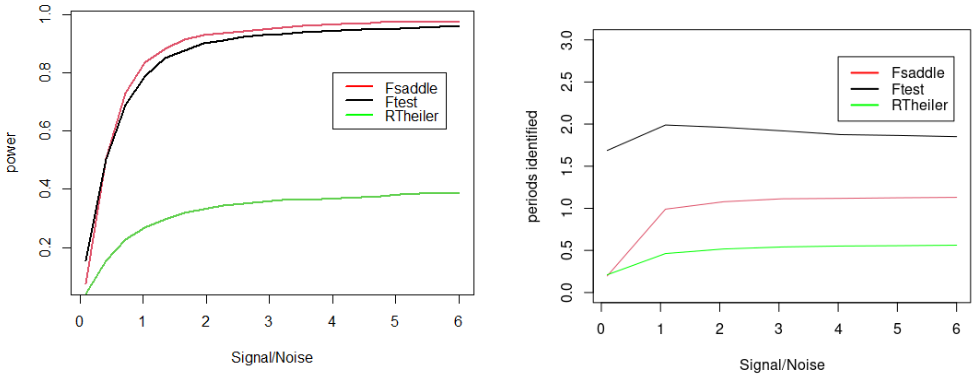


FIGURE 3.4: (Left) Comparison of the power between the F-test with Saddle-point approximation (Fsaddle), standard F-test (Ftest) and Theiler's approach (RThieler), for different signal to noise ratios. (Right) Comparison between the average number of periods identified as correct for the three tests mentioned. For each value of SNR 1000 curves with 200 data points were generated. The significance level was set to $\alpha = 1 - 0.998$.

regression is used can improve the quality of our results compared to the other two methods.

3.7.2 Example 2: Gaussian process regression

In Section 3.4.1 and Section 3.4.2 we discussed that the generalized F and CVF-tests can be readily used for more complex models such as Gaussian process regression. In this Section we simulate artificial light curves from a GPR model, fit a Gaussian process regression and perform our tests in order to estimate their power. For this example we borrow the sampling of a randomly selected real light curve (object 3314_I from Froebrich et al. (2021) [24]). We generate our periodic signal from a Gaussian prior using a periodic kernel as in (2.15) with period at 5.2 days, parameters h and A are set equal to 1. A typical example of the shape of our simulated data can be seen in Figure 3.5 (Left). We generate light curves for different signal to noise ratios, ranging from 0.01 to 8. The noise is generated from a zero-mean Normal distribution. We estimate the power for each different value of SNR based on 1000 repetitions. In Figure 3.5 (Right) we see the power of the generalized F-test at a $1 - \alpha$ =99% significance level. We notice that the test performs very well with its power being estimated larger than 0.7 for even SNRs as small as 0.5. For a SNR larger or equal to 1 the power is constantly estimated close to 1.

To our knowledge there does not exist another test under the GPR setting to pose as a comparison. We include in our plot the estimated power of the standard F-test using a sinusoidal model, for the same simulated data, as a reference, acknowledging that the comparison of these tests is not fair. The generalized F-test performs a lot better than the standard F-test in this example as expected. We performed the same analysis for the CVF-test too and obtained very similar results to those obtained from the generalized F-test. In Table 3.1 we can see a comparison of the power of generalized F and CVF-tests for the same simulated data and SNR fixed to 0.65. The power is calculated for different significance levels. Both tests perform very well with their power estimated to be larger than 0.84 for a significance level of $1 - \alpha = 99.5\%$ for example. Of course the larger the $1 - \alpha$ significance level the smaller the estimated power. We should note that the significance levels were chosen as multiple testing corrections according to Sidak (1967) [60] and correspond to different numbers of tests (e.g. 1, 10, 100, 1000) that could be conducted. We notice that the tests behave similarly with the CVF-test having a slightly larger power in this particular example. Finally we also calculated the probability of not rejecting the null hypothesis when the null hypothesis is true by simulating purely noise data. In all cases the probability was larger than or equal to 0.98 for our proposed tests and the same holds for when a linear model and the standard F-test is applied.

Example 2

Here we show another example similar to the previous one, but this time we simulate data from a GPR model that are close to a sinusoidal shape. For this example we borrow again the sampling of the same real light curve (object 3314_I from Froebrich et al. (2021) [24]) and follow exactly the same procedure as described before to generate a periodic signal at 5.2 days. A typical example of the shape

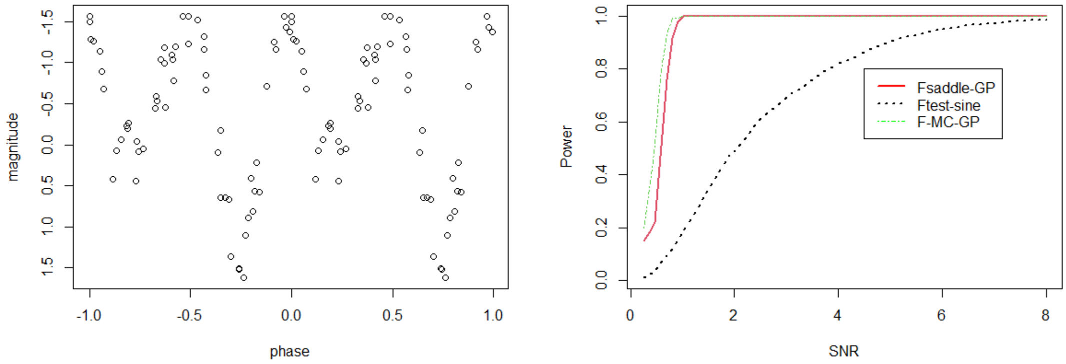


FIGURE 3.5: (Left) A phased light curve generated from the GPR model with period 5.2 days with SNR=6. This is a typical example of the shape of the light curves generated. (Right) The estimated power of the GPR generalized F-test of light curves generated for different SNR (red line). the green line shows the power when the same analysis is perform using Monte Carlo and the dashed line shows the power of the standard F-test for a sinusoidal based fitting.

TABLE 3.1: This table shows the estimated power($\hat{\gamma}$) and the estimated probability of not rejecting the null hypothesis when it is correct ($1-\hat{\alpha}$) of the generalized F and CVF-tests for different significance levels. The alternative hypothesis here is a signal generated form the GPR model with period at 5.2 days.

	$\hat{\gamma}$		$1-\hat{\alpha}$	
1-sig. level	**F GPR**	**CVF GPR**	**F GPR**	**CVF GPR**
$0.95^{1/1} = 0.95$	0.968	0.970	0.899	0.895
$0.95^{1/10} = 0.9948838$	0.842	0.848	0.980	0.979
$0.95^{1/100} = 0.9994872$	0.616	0.612	0.997	0.995
$0.95^{1/1000} = 0.9999487$	0.352	0.354	1.000	0.999

of this simulated data can be seen in Figure 3.6 (Left). Similarly as before we estimate the power and the results can be seen in Figure 3.6 (Right). We see that the test performs very well with its power being estimated larger than 0.7 for even SNRs as small as 0.5.

We include again the estimated power of the standard F-test and a sinusoidal model. The tests in this example seem to perform similarly in terms of power with the generalized F-test being slightly better. In Table 3.2 we can see a comparison of the power of generalized F and CVF-tests similar to the previous example. Again both tests perform well with their estimated power being larger than 0.92 for a significance level of $1-\alpha = 99.5\%$ for example. We notice again that the

Table 3.2: This table shows the estimated power ($\hat{\gamma}$) of the generalized F and CVF-tests for different significance levels.

sig. level	F GPR	CVF GPR
$0.95^{1/1} = 0.95$	0.987	0.987
$0.95^{1/10} = 0.9948838$	0.922	0.924
$0.95^{1/100} = 0.9994872$	0.714	0.715
$0.95^{1/1000} = 0.9999487$	0.439	0.446

tests behave similarly with the CVF-test having a slightly larger power in this example too.

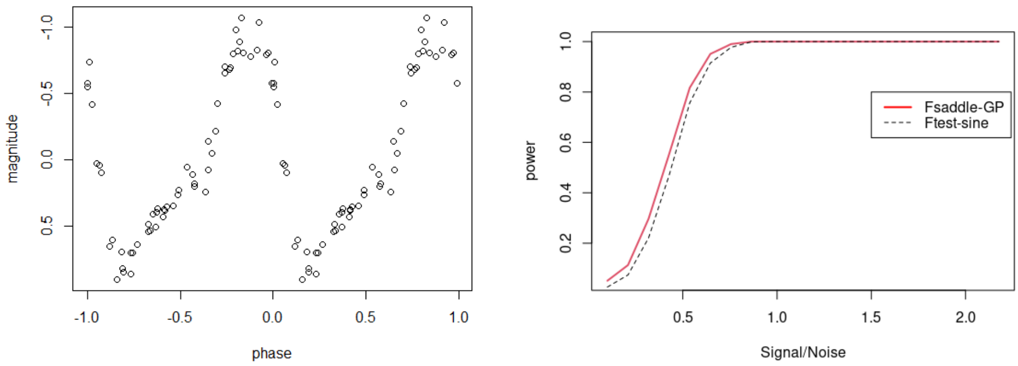


Figure 3.6: (Left) A phased light curve generated from the GPR model with period 5.2 days with SNR=6. This is a typical example of the shape of the light curves generated. (Right) The estimated power of the GPR generalized F-test of light curves generated for different SNR (red line). The dashed line shows the power of the standard F-test for a sinusoidal based fitting.

3.8 Further Numerical examples

3.8.1 Exploring the Saddlepoint approximation accuracy

In this subsection we will see some numerical examples in order to understand further the behaviour of Saddlepoint approximation for the statistics proposed in this paper. We will start by approximating the cumulative distribution function (CDF) of the generalized F-statistic under the null hypothesis when the alternative is a periodic Gaussian process model. For this simulation the same kernel parameters as those in Section 3.7.2 are used.

Table 3.3: Time in seconds for 50 repetitions of the different methods.

	Saddlepoint	Imhof	Monte Carlo	Exact
GPR	16.30	17.56	774.90	–
OLS	0.79	1.24	4.10	0.02

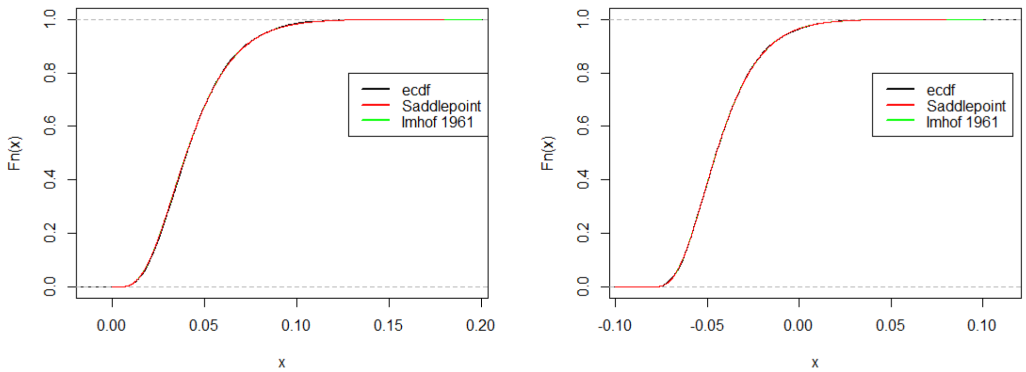


Figure 3.7: (Left) An example of the ECDF and the saddlepoint approximation of the corresponding generalized F-statistic based on the RSS of White Noise and GPR models. The green line that agrees with the saddlepoint approximation is the approximation based on the numerical integration method. (Right) The same comparison of methods for the CVF-statistic of GPR.

In order to examine the performance of Saddlepoint approximation, we simulate 10000 values of the generalized F-statistic under the null hypothesis assumption and calculate its empirical cumulative distribution function (ECDF), which we compare with the CDF obtained by using Saddlepoint, see Figure 3.7 (Left). Saddlepoint approximation gives results that are very close to the empirical CDF of the statistic. In addition, we plot the CDF approximated using numerical integration according to Imhof (1961) [34] getting almost the same results with the other approaches. The only difference is that Saddlepoint approximation is faster, see more details at Table 3.3 for a time comparison between the methods. Note that whenever we report ECDF values below we mean the empirical CDF of 10000 replications. Similar results have been produced for the saddlepoint approximation of the CVF-statistic. In Figure 3.7 (Right) we see that the saddlepoint approximation of CVF under the null hypothesis when GPR is used as the alternative also works well.

The hypothesis testing proposed can be also used to model correlated background

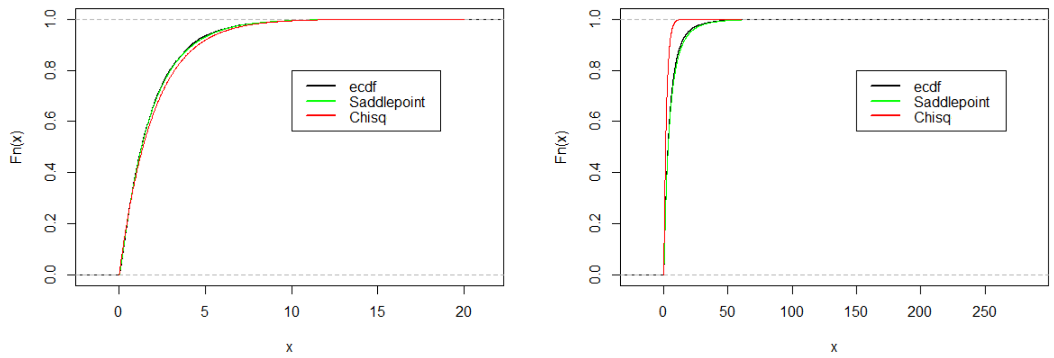


FIGURE 3.8: (Left) A comparison of the CDF of the F-statistic under the null hypothesis of correlated noise for a light curve of 200 observations. The black line shows the empirical CDF, the green line corresponds to Saddlepoint approximation and the red line is the asymptotic chi-square distribution. (Right) A comparison of the CDF of F-statistic under the null hypothesis of correlated noise for a light curve of 200 observations. Note that when the sample size is very small the asymptotic result no longer holds but the saddlepoint method approximated the CDF quite accurately.

noise and in many cases can perform better than standard asymptotic tests. In this example we will compare the approximations of the CDF of the F-statistic under the presence of AR(1) type noise and when the alternative is fitting a sinusoidal model. Under this scenario the standard F-statistic can no longer be assumed to follow an $F(m_1 - m_0, n - m_1)$ distribution, but there is an asymptotic result (Hamilton (2020) [29]) based on chi-square distribution with $m_1 - m_0$ degrees of freedom. In Figure 3.8 we compare the approximations of the CDF of the F-statistic for two different scenarios based on the sample size. For a relatively big sample size of 200 data points both the Saddlepoint approximation and the asymptotic result are close to the empirical CDF. When the sample size gets small however, as in Figure 3.8 (Right) with 20 data points, the asymptotic result based on chi square no longer works well. The Saddlepoint on the other hand still approximates the distribution quite accurately.

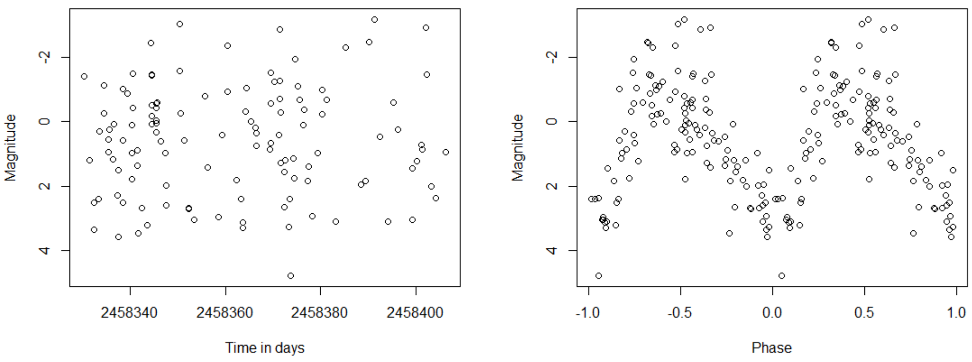


FIGURE 3.9: (Left) Simulated light curve from GPR prior with period at 5.1 days. The sampling is borrowed from a real light curve and this shape is used throughout the example of Table 1. (Right)he same simulated light curve folded into period of 5.1 days.

3.8.2 Red noise vs White noise GPR model

In Section 3.5 we saw how our test statistics F and CVF can be adjusted when our data are contaminated with red noise. Adjusting the models and tests to the correct noise assumption can improve drastically our results. More specifically, assuming a white noise model when the residuals are correlated can lead to periods failing to be identified and also to an increased number of falsely detected periods. As an example we will generate 100 light curves for 3 different correlated noise scenarios ($\rho = 0.1, 0.5, 0.9$) with variance 1. The curves will be generated from a Gaussian process prior with a periodic kernel with period at 5.1 days, $A = 1$ and $h = 1$. The mean behaviour of the curves will be fixed for all curves in all scenarios and it will be a realisation from the Gaussian process prior. We will borrow the sampling times from a real light curve (object 6785_I from Froebrich et al. (2021) [24]). A typical example of the simulated data used here can be seen in Figure 3.9. Note that the performance of the methods is not affected by the magnitude's scale.

For these simulated light curves we will fit a Gaussian process regression model with a correlation structure $\mathbf{C}_\rho$ for AR(1) errors (e.g. (3.11) but for equally spaced time). For this example, we estimate the parameters by minimizing the equivalent squared leave-one-out cross-validation error (3.5). We will search for potential

Table 3.4: A comparison between the red and white noise Gaussian process regression models fitted to red noise data. The columns labelled “correct peak” show the percentage of the periodograms that had the maximum peak at correct period $p = 5.1$. The columns labelled “false periods” show the average number of periods falsely identified as valid.

	GPR red noise		GPR white noise	
ρ	**correct peak**	**false periods**	**correct peak**	**false periods**
0.1	90%	1.34	33%	3.56
0.5	79%	0.88	25%	3.62
0.9	84%	1.20	44%	2.90

periods between 1 and 10 days with 1 decimal accuracy. Note that this particular period search grid is chosen for computational convenience and it is selected in order to contain the exact true period. For real applications it is preferable to use the two-stage grid search method as seen in Wang et al. (2012) [72] and Reimann (1994) [53], where some N periods are initially chosen corresponding to the top N periodogram peaks based on a rough period grid, and then a finer period search grid is applied around each of the N chosen periods from the previous step, see also Section 2.5. The results obtained by the red noise model will also be compared to those from the standard GPR model. We notice that for all three scenarios the red noise GPR model most of the time has the maximum peak of the F-statistic periodogram at the correct period, on the other hand, the white noise model (standard GPR) most of the time fails to do so. These results are summarized in Table 3.4 under the columns labelled as “correct peak”.

Furthermore, we apply the generalized F-test as described in Section (3.5) to all simulated curves for a significance level $1 - \alpha$ set to $0.95^{1/91} = 0.9994365$. This is the correction for multiple tests as seen in Sidak (1967) [60], 91 is the number of tests conducted which is the number of trial periods in our case. Almost always the correct period (at 5.1 days) was identified as significant (more than 99% of the time). When the wrong model (white noise) was used however, we had an increased number of false period discoveries compared to that of using the red noise GPR model. Specifically, after performing our period detection test, in many cases other periods, except for the true one, appeared to be falsely significant. These results are summarized in Table 3.4 under the columns labelled as “false periods”, where we show the average number of periods falsely appearing to be significant, for each simulation scenario.

3.9 General implementation notes

For the examples in this Section we used a very rough period grid of one decimal accuracy. This grid is chosen for computational simplicity and it also shows us how methods perform under relatively crude search schemes. In general it is advised to follow the two stage period search method seen in Reimann (1994) [53] and described in Section 3.8.2.

Another thing to note is that we selected our period grid homogeneously in the range between 0.5 and 30 days. For this application we had prior knowledge as to where the potential periods were, since they were studied in Froebrich et al. (2021) [24], and so this approach is sensible as the results also showed. In general applications however, when no prior information is available, it is preferable to select periods in-homogeneously, by building a grid based on the equivalent frequency range instead and ensure that short periods will not be missed.

3.10 Application to real light curves

In this Section we apply the methodology from the previous Sections to real light curves. They are obtained as part of the Hunting Outbursting Young Stars Citizen Science Project [23]. This project combines observational data from professional, university and amateur observatories to construct long-term light curves of young stars. By its very nature the project hence creates in-homogeneously sampled light curves. The example objects investigated here are all situated in the Pelican Nebula, a vast star forming region in Cygnus.

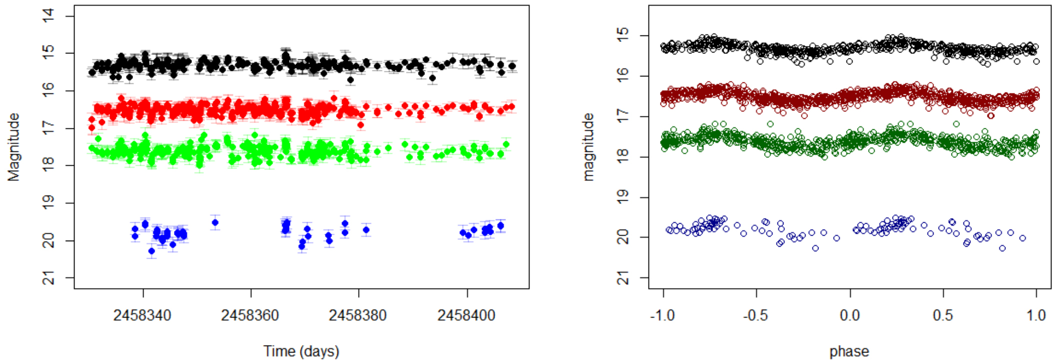


FIGURE 3.10: (Left) Light curve number 6149. Each colour represents a different filter. Black is for Infrared, red for Red filter, blue for Blue filter and green for Visual. The error bars represent the measurement accuracies. (Right) Light curve number 6149 folded in period 2.1763 days for different filters.

3.10.1 Example 1

We first investigate object number 6149 from Froebrich (2021) [24]. This object belongs to the population of young stars in the Pelican Nebula and shows a clear periodic behaviour in the four filters Blue (B), Visual (V), Red (R), and Infrared (I). The period of the variations has been determined as 2.1763 d, from the median of the periods in the individual filters (Froebrich et al. (2021) [24]). See Figure 3.10 where the light curve is plotted for the different filters against time and folded in period 2.1763 days.

We first analyse the light curve independently for each filter, by fitting a weighted Gaussian process model using the periodic kernel and assuming independent residuals. In principle, light curves measured in different filters for the same star should exhibit the same periodic behaviour, so if a method identifies the same period for all filters it is a good sign that the correct period is detected.

The periodogram is obtained using F and CVF statistics, see Figure 3.11 (Left). For this example we search for periods between 0.5 and 30 days with a very rough grid of 1 decimal accuracy. We then apply sequentially the generalized F and CVF-tests and identify in most cases the period at 2.2 days as the most important real period. This is the period that corresponds to the maximum F-statistic value and a

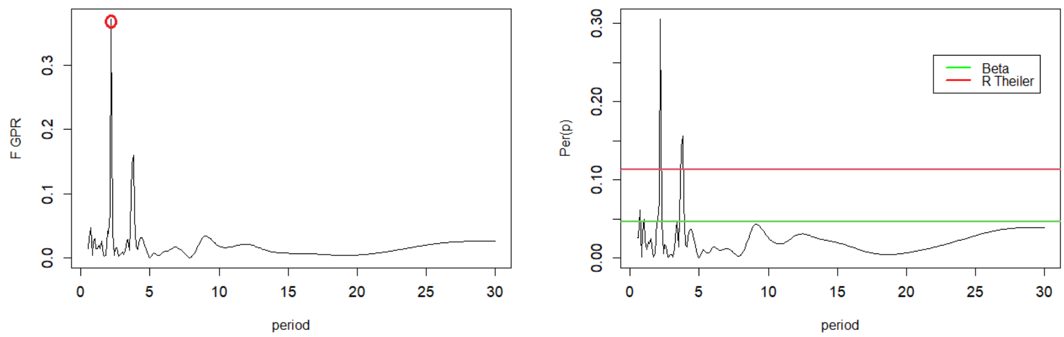


FIGURE 3.11: (Left) The weighted GPR F periodogram for light curve 6149 measured in the visual filter. With the red circle we mark the period identified as important according to the generalized F-test. (Right) The weighted least squares sine periodogram for the same light curve. The lines represent the critical values (same for all trial periods) according to the Standard F-test and the "RTheiler" approach, peaks above the lines are considered as important periods.

significant p-value. We notice however, that for some filters the tests identify more than one period as valid and in some cases as in filter R the tests fail to identify any periods. We will further run the same analysis but this time assuming some correlation structure for the residuals as described in Section 3.5. We see that with these adjustments the generalized F and CVF-tests return as valid only one period at 2.2 days. It is worth noticing that for filter R the tests identify also the period at 2.2 days.

Next, for comparison, we analyse the light curve independently for each filter again, by fitting a sinusoidal wave using weighted least squares regression (this is the model described in Section 3.7.1). We obtain the periodogram based on the R^2 statistic seen in Figure 3.11 (Right). We then proceed by performing the standard F-test and Beta distribution based test as seen in Thieler et al. (2013) [63]. The results for all methods and filters are summarized in Table 3.5. We can see that all methods in most cases identify the most important period as 2.2. The standard F-test seems to be overestimating the number of valid periods. All methods seem to find as valid other periods than 2.2, with some exceptions for the GPR based tests.

TABLE 3.5: The period identified as most important from the various tests, for different filters. The numbers in the parenthesis denote the number of other periods identified as important from the tests. The dash represents the case that a test did not identify any period as valid.

Filter	F GPR	CVF GPR	F GPR red	CVF GPR red	F sine	Thieler sine
B	2.2 (2 extra)	2.2 (2 extra)	2.2 (0 extra)	2.2 (0 extra)	2.2 (2 extra)	-
R	-	-	2.2 (0 extra)	2.2 (0 extra)	2.2 (3 extra)	2.2 (1 extra)
I	2.2 (1 extra)	2.2 (1 extra)	2.2 (1 extra)	2.2 (1 extra)	2.2 (1 extra)	2.2 (1 extra)
V	2.2 (0 extra)	2.2 (0 extra)	4.3 (1 extra)	4.3 (1 extra)	2.2 (3 extra)	2.2 (1 extra)

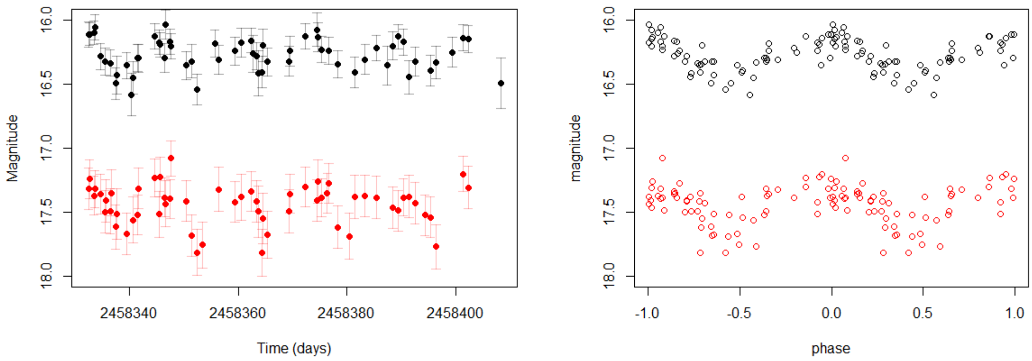


FIGURE 3.12: (Left) Light curve number 3314. Black is for Infrared filter and red for Red filter. The error bars represent the measurement accuracies. (Right) Light curve number 3314 folded in period 13.8783 days for different filters.

3.10.2 Example 2

As a second example, we analyse object number 3314, measured in filter I and R with a studied period at 13.8783 days, as seen in Froebrich (2021) [24]. See Figure 3.12 (Left) where the light curve is plotted for both filters and Figure 3.12 (Right) where the light curve is folded in a 13.8783 day period.

TABLE 3.6: The period identified as most important from the tests in Example 2, for different R and I. In this case no other period was identified as important from the tests (0 extra periods).

Filter	F GPR	CVF GPR	F sine	Thieler sine
R	13.9 (0 extra)	13.9 (0 extra)	13.9 (0 extra)	13.9 (0 extra)
I	13.9 (0 extra)	13.9 (0 extra)	13.9 (0 extra)	13.9 (0 extra)

Similarly to example 1 we apply all previous methods to both filters. The results are summarized in Table 3. We see that for this particular example all methods behaved in the same way. They all identified only one important period at 13.9

days. In this example the red noise model was applied too returning the same results and was excluded from the table for simplicity.

3.11 Summary

This Chapter dealt with the topic of period detection through hypothesis testing. This approach is fairly standard when linear models and least squares regression is used, as the distribution of common statistics such as F is known under the null hypothesis of no periodic signal in the data. This problem gets more difficult when more complex modelling schemes are implemented and the distribution of the statistic used is not known. Usually this problem is addressed through computationally expensive simulations. The generalized F-statistic, however, introduced in this Chapter is readily used for a series of models including weighted regression, non iid noise settings such as AR(1) errors and non-parametric linear smoothers, and is a faster and accurate alternative to Monte Carlo methods. Furthermore, a test specifically designed for Gaussian process regression is introduced based on the leave-one out cross-validation error, which could be also applied under red noise settings.

Our simulations show that both tests perform better than the alternatives for the weighted linear regression model when the sample size is relatively large. Furthermore, we see how both tests exhibit a large power under the Gaussian process regression setting for data with signal to noise ration larger than 0.6. Examining the behaviour of the tests under the presence of red noise showed that taking this aspect of correlated errors into account, when present, can significantly reduce the number of falsely detected periods. Finally, our tests were applied to two real light curves and returned similar results to existing approaches.

Chapter 4

Extensions to Gaussian process regression period estimation methodology

4.1 Motivation

In Chapters 2 and 3 we discussed extensively the use of Gaussian process regression (GPR) as a tool for period estimation. Although GPR is a very flexible model that requires minimal assumptions, there are still attributes in the light curves that may not be adequately captured. For example, as discussed in Section 2.2.2 and as suggested by some of the results in the period search study in Froebrich et al. (2021) [24], it is common for our data to be contaminated with outliers, or that the errors exhibit fatter tails than those assumed by a Normal distribution. This behaviour could be attributed to many factors such as observational errors or sudden changes in the star's brightness unrelated to its overall underlying periodic signal. There have been several methods developed in the literature that adapt the concept of robust regression to the problem of period estimation, see for example Oh et al. (2004) [49], where the authors obtain their period estimates through a Huber's regression setting or Thieler et al. (2013) [63] where a combined methodology for different types of robust regression is provided. To our knowledge however there has not been an equivalent method under the GPR setting.

Moreover as discussed in Chapter 1 light curves are usually measured in multiple astronomical filters. There exist some methods in the literature for multiple filters along the line of ordinary least squares but again no work has been done in order to incorporate multiple filter measurements for Gaussian process models. Another example is that of quasi periodic light curves. Data like these have periods or shapes that change over time, see for example Figure 2.9 and 2.10 which display such examples of simulated light curves. In order to effectively capture these effects and get accurate period estimations there is a need for these GPR models to be extended. This is the purpose of this Chapter and its outline is as follows.

In Section 4.2 we extend the flexibility of the GPR models by considering t-process regression models instead in order to amend for possible outliers. Moreover, we show the extension of this method in order to take measurement accuracies into account in a form of a weighted t-process regression. The multiple filter extension of these models is discussed in Section 4.3. The problem of quasi periodic data is discussed in Section 4.4 where we build different spherical kernels capable of capturing such behaviours. In Section 4.5 we adapt the hypothesis testing ideas from Chapter 3 for these models and finally in Section 4.6 we apply these models to two real light curves.

4.2 t-process regression period estimation

The aim of this Section is to provide tools for period estimation under the GPR setting when the available data contain outliers. For that purpose we show here how we can model our light curves using the t-process as a more robust approach against data with fat tails. In Shah et al. (2014) [58] we see how the t-process can be used as an alternative to the Gaussian process. According to the authors the t-process works as well as the Gaussian process, only it is more flexible. There are many non-parametric approaches for estimating the period of unequally spaced

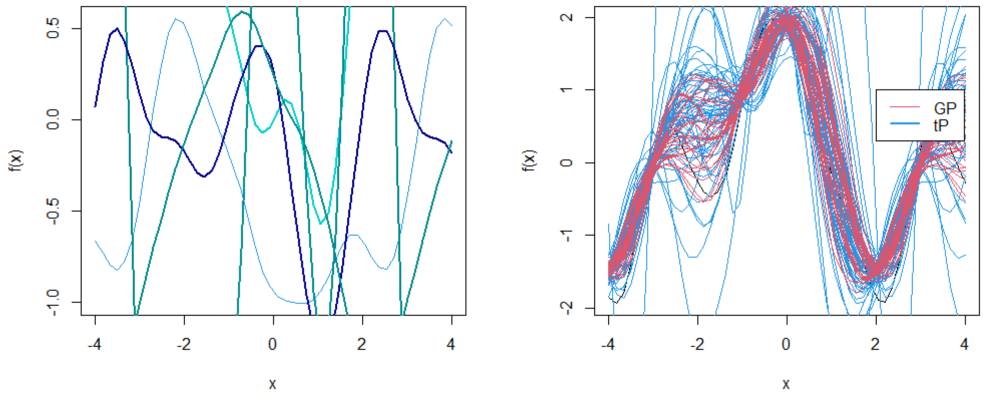


FIGURE 4.1: (Left) A sample from the distribution over priors from a t process with a periodic kernel. (Right) Samples from the distribution over posteriors. With the blue colour we denote samples from the t-process, with the red we denote samples from the Gaussian process. We notice the wider range that the former cover compared to the latter.

data as we saw in the previous Chapters. The most relevant work can be found in Wang et al. (2012) [72] where the authors estimate the periods of light curves using Gaussian process regression. Here we discuss the use of t-process regression instead for this particular period estimation problem.

In order to get a first idea of the difference between Gaussian and t-process we can look at Figure 4.1 (Right), where we can see samples from the posterior over functions **g**. With blue colour we denote samples from the t-process and with red samples from the Gaussian process. We notice how much wider the range of the functions is in the first case. This is something we would expect because of the fat tail behaviour of the t-distribution. In Figure 4.1 (Left) we see samples from a t-process with the periodic kernel (2.15). An equivalent plot for the Gaussian process can be seen in Section 2.2.3 Figure 2.5 for example.

4.2.1 Definition of the t-process

In Shah et al. (2014) [58] we see how we can derive the student t-process from a Gaussian process by placing an inverse Wishart process prior to the covariance function K. We can write for a process with mean zero:

$$\mathbf{y}|\sigma\mathbf{I} \sim GP(0, (\nu-2)\sigma\mathbf{I})$$

$$\sigma\mathbf{I} \sim IWP(\nu, \mathbf{K})$$

where with $\nu > 2$ we denote the degrees of freedom. The inverse Wishart distribution is a conjugate prior for the covariance matrix of a Gaussian distribution. So σ can be marginalized and thus we can assume that our observations follow a Multivariate t distribution, $\mathbf{y} \sim MTV(0, \nu, \mathbf{K})$ with the pdf,

$$P(\mathbf{y}) = \frac{\Gamma(\frac{\nu+n}{2})}{((\nu-2)\pi)^{n/2}}|\mathbf{K}|^{-1/2}\left(1 + \frac{\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y}}{\nu-2}\right)^{-\frac{\nu+n}{2}}$$

Note that the multivariate t distribution is consistent under marginalization, meaning that conditioning on a subset of components of an n dimensional t will result in a t distribution. This property is useful in performing t-process regression and deriving the distribution of our fitted values or predicted values.

In most real life examples our observations will be contaminated with some noise with variance σ^2. For a given data set with n observations and a matrix $\mathbf{X}$ with the independent variables, we can assume our observations $\mathbf{y}$ follow a multivariate t and we can write,

$$\mathbf{y} \sim MVT(0, \nu, \mathbf{K} + \sigma^2\mathbf{I})$$

Now if we want to predict the behaviour of g for some n_1 predictive values $\mathbf{X}^*$ we can find the predictive distribution for the predicted values $\mathbf{g}^*$, which is still a multivariate t because it is consistent under marginalization. The conditional distribution of our predictive values is

$$\mathbf{g}^*|\mathbf{X}, \mathbf{y}, \mathbf{X}^* \sim MVT(\bar{\mathbf{g}^*}, \nu + n_1, \mathrm{cov}(g^*))$$

$$\bar{\mathbf{g}^*} = K(\mathbf{X}^*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I}]^{-1}\mathbf{y}$$

$$\mathrm{cov}(g^*) = [K(\mathbf{X}^*, \mathbf{X}^*) - K(\mathbf{X}^*, \mathbf{X})(K(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I})^{-1}K(\mathbf{X}, \mathbf{X}^*)]B$$

$$B = \frac{\nu - 2 + \mathbf{y}^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{y}}{\nu + n_1 - 2}$$

It is straightforward to obtain the distribution of the fitted values as well. We just need to replace $\mathbf{X}^*$ with $\mathbf{X}$ and of course n_1 is n which is the number of observations in our data set. For more details on t-process regression see Chen et al. (2020) [16]. Note that the mean is of the same form as that of the Gaussian process regression fitted distribution. The actual values however will be different since we obtain the parameter estimates by maximizing different marginal likelihoods. The marginal log likelihood has the closed form seen below:

$$\log P(\mathbf{y}|\nu, \mathbf{K}) = -\frac{1}{2}\log\left(|\mathbf{K} + \sigma^2\mathbf{I}|\right) - \frac{\nu + n}{2}\log(1 + \frac{\mathbf{y}^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{y}}{2}) + S \quad (4.1)$$

$$S = -\frac{n}{2}\log((\nu - 2)\pi) + \log(\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\nu/2)})$$

4.2.2 Period estimation

In the previous Section we saw that we can obtain estimations for our parameters by maximizing the marginal likelihood. Here our primary interest is on the parameter p, the period. Similar to the GPR setting we can introduce periodicity in our model through the choice of the Kernel which can be, for example, the common periodic kernel (2.15). Maximizing the marginal likelihood of the t-process with respect to p is a non trivial task as also seen for the other models, because it exhibits numerous local maxima. In Figure 4.2 (Left) we see an example of how the marginal likelihood (4.3) looks for different values of p.

For this reason we obtain our period estimate by following the two grid procedure described in Section 3.8.2. Given a trial period of the grid, we estimate the rest of the parameters which in our case include the kernel parameters A and h, the variance σ^2 and the degrees of freedom ν. For a given period the likelihood is a smooth function of parameters A, h and σ, see for example Figure 4.2 (Right). In this case $\hat{p} = \operatorname{argmax} \log P(y|\nu, K)$, but we can also build a t-process regression periodogram using the generalized F statistic introduced in Section 3.2 by following

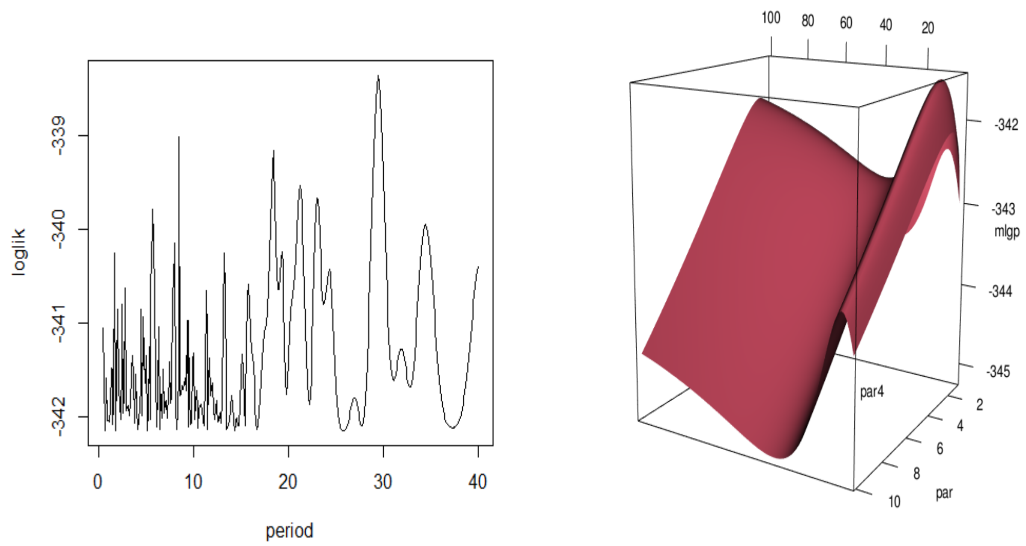


FIGURE 4.2: (Left) Example of the t process marginal likelihood with only the period varying. Notice the numerous local maxima. (Right) Example of the t process marginal likelihood with the parameters for the degrees of freedom and smoothness varying. Given the period the form of the likelihood becomes smoother.

exactly the same steps seen there, the only difference is that matrix $\mathbf{M}_1$ will be calculated using the parameters obtained from the t-process likelihood.

4.2.3 Taking the measurement accuracies into account

It is very common in data like light curves to have also measurement errors $s_i = 1/w_i$ for each observation. If this type of information is reliable, it can improve significantly our results. If that is the case we should like to be able to take them into account in the setting of t-process regression too. This can be done with the use of a weighted t-process regression. Under the general weighted regression setting our basic model from (2.1) can be written as $w_i y_i = g(t_i)w_i + \epsilon_i$ where with w_i we denote the weights. This can be further written as $y_i = g(t_i) + \epsilon_i/w_i$. The error component is usually assumed to be a white noise and in most settings follows a Gaussian distribution of the form $\epsilon_i \sim N(0, \sigma^n \mathbf{I})$. In our case however the errors follow a t distribution as seen below.

We can start by defining the simplest case for a matrix $\mathbf{Q}$ containing the weights as seen in (2.18). We observe that since $\epsilon \sim MVT(0, \nu, \sigma^n \mathbf{I})$, then $\mathbf{Q}^{-1}\epsilon \sim MVT(0, \nu, \mathbf{Q}^{-2}\sigma^2\mathbf{I})$. Given this observation we can assume the following model for our observed values $\mathbf{y}$,

$$\mathbf{y} \sim MVT(0, \nu, \mathbf{K} + \mathbf{Q}^{-2}\sigma^2\mathbf{I})$$

Under this setting the predictive distribution (and similarly the fitted distribution) is of the following form

$$g^*|\mathbf{X}, \mathbf{y}, \mathbf{X}^* \sim MVT(\bar{g*}, \nu + n_1, \mathrm{cov}(g^*)) \tag{4.2}$$

$$\bar{g*} = K(\mathbf{X}^*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \mathbf{Q}^{-2}\sigma^2\mathbf{I}]^{-1}\mathbf{y}$$

$$\mathrm{cov}(g^*) = [K(\mathbf{X}^*, \mathbf{X}^*) - K(\mathbf{X}^*, \mathbf{X})(K(\mathbf{X}, \mathbf{X}) + \mathbf{Q}^{-2}\sigma^2\mathbf{I})^{-1}K(\mathbf{X}, \mathbf{X}^*)]B$$

$$B = \frac{\nu - 2 + \mathbf{y}^T(\mathbf{K} + \mathbf{Q}^{-2}\sigma^2\mathbf{I})^{-1}\mathbf{y}}{\nu + n_1 - 2}$$

Similarly as in the t-process case we can obtain our periodogram only this time we maximize the marginal likelihood of the weighted t-process, as seen below, in order to obtain the estimates of the parameters and consequently the periodogram entries.

$$\log P(\mathbf{y}|\nu, \mathbf{K}) = -\frac{1}{2}\log(|\mathbf{K} + \mathbf{Q}^{-2}\sigma^2\mathbf{I}|) - \frac{\nu + n}{2}\log(1 + \frac{\mathbf{y}^T(\mathbf{K} + \mathbf{Q}^{-2}\sigma^2\mathbf{I})^{-1}\mathbf{y}}{2}) + S \tag{4.3}$$

$$S = -\frac{n}{2}\log((\nu - 2)\pi) + \log\left(\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\nu/2)}\right)$$

The period is estimated in exactly the same way as described in Section 4.2.2.

4.3 Multiple filter extension

In Chapter 1 we saw that each star can be measured in multiple astronomical filters or bands, as they are sometimes referred to, see for example Figure 1.2 (Top) where object 5686 from Froebrich et al. (2021) [24] is measured in six different filters. In general, as we discussed in previous Sections, a star should have the same period in all filters it has been measured in. Thus, it makes sense to try to model the magnitude in all filters together instead of modelling them independently as seen in this thesis so far. Little work has been done in the literature towards that end. One example can be found in Vanderplas and Ivezic (2015) [67], where the authors introduce a generalized version of the Lomb-Scargle periodogram extended to incorporate measurements in multiple filters. This method lies within the OLS framework, where a harmonic model as the one seen in (2.11) is assumed with additional dummy variables corresponding to each filter. To our knowledge, no work has been done in order to extend the Gaussian process model for multiple filters. In the next Sections we show a multiple-filter extension of the GPR model and a more generalized version of that based on weighted t-process regression.

4.3.1 Multiband GPR model

Here we show how the GPR model (2.16) can be extended to include multiple filters measurements. We will follow the same idea of using dummy variables for each filter as seen in Vanderplas and Ivezic (2015) [67]. Specifically we augment our dataset with as many dummy variables as the available filters. We denote these new variables as d_m, $m = 1, 2, ..., M$ where M is the number of available filters. By definition $d_{m_j} = 1$ if our jth observation was measured in the mth filter and 0 otherwise. We can thus represent our dependent variables by a matrix $\mathbf{X}$ of $N \times M + 1$ dimensionality:

$$\mathbf{X} = \begin{bmatrix} t_1 & d_{1_1} & \dots & d_{M_1} \\ t_2 & d_{1_2} & \dots & d_{M_2} \\ \vdots & \vdots & & \vdots \\ t_N & d_{1_N} & \dots & d_{M_N} \end{bmatrix} \tag{4.4}$$

where N is the number of all observations combined from all filters.

We can write our multiple-filter GPR model along the same lines as in Section 2.2.3, and we will refer to this model as the multiband GPR model. In this case the estimated values of $\mathbf{g}$, which is of length N now, will be assumed to follow a multivariate Normal distribution,

$$\hat{\mathbf{g}} \sim N(\mathbf{W}\mathbf{y}, \mathbf{K} - \mathbf{W}\mathbf{K}) \tag{4.5}$$

Here $\mathbf{K}$ is an $N \times N$ covariance matrix whose entries are obtained as

$$\mathbf{K}_{jk} = A \exp\left(-2\sin^2(\frac{\pi}{p}|\mathbf{X}_{[j,]} - \mathbf{X}_{[k,]}|)/h^2\right) \quad j,k = 1,2,\ldots,N \tag{4.6}$$

with $\mathbf{X}_{[j,]}$ we denote the jth row of the standardized matrix $\mathbf{X}$. The mean and variance of 4.5 are given by equations 2.16 in Section 2.2.3. Note that in this case $\mathbf{y}$ will also be of length N as it will include the magnitudes observed in all filters.

We can estimate the multiband GPR model's parameters including the period p, which is of primal interest, by maximizing the marginal likelihood (3.12) or by minimizing the leave-one-out cross validation error (3.5) following the double grid procedure described in Section 3.8.2. Furthermore, we can also build a multiband GPR periodogram using the generalized F or CVF statistics given by equations (3.2) and (3.9) respectively which allow us to perform the hypothesis testing scheme we described in the previous Chapter. For more details on the Hypothesis test under this setting see Section 3.4.1.

As an example we simulate a light curve measured in three filters from the GPR model with the same periodicity $p = 5.2$ across filters, but assuming a different shape for each filter. We borrow the sampling (i.e. the observation times) of a real light curved measured in the Infrared, Red and Visual filter (object 7159 from Froebrich et al. (2021) [24]). This simulated light curve can be seen in Figure 4.3 (Top left), note that for illustration purposes we used here a signal to noise ratio equal to 2. The multiband GPR periodogram based on the generalized F statistic can be seen in Figure 4.3 (Top right). For comparison reasons we also show the equivalent GPR periodogram for each filter individually (see Figure 4.3 (Bottom line)). Note that in this example we obtained the same results from the

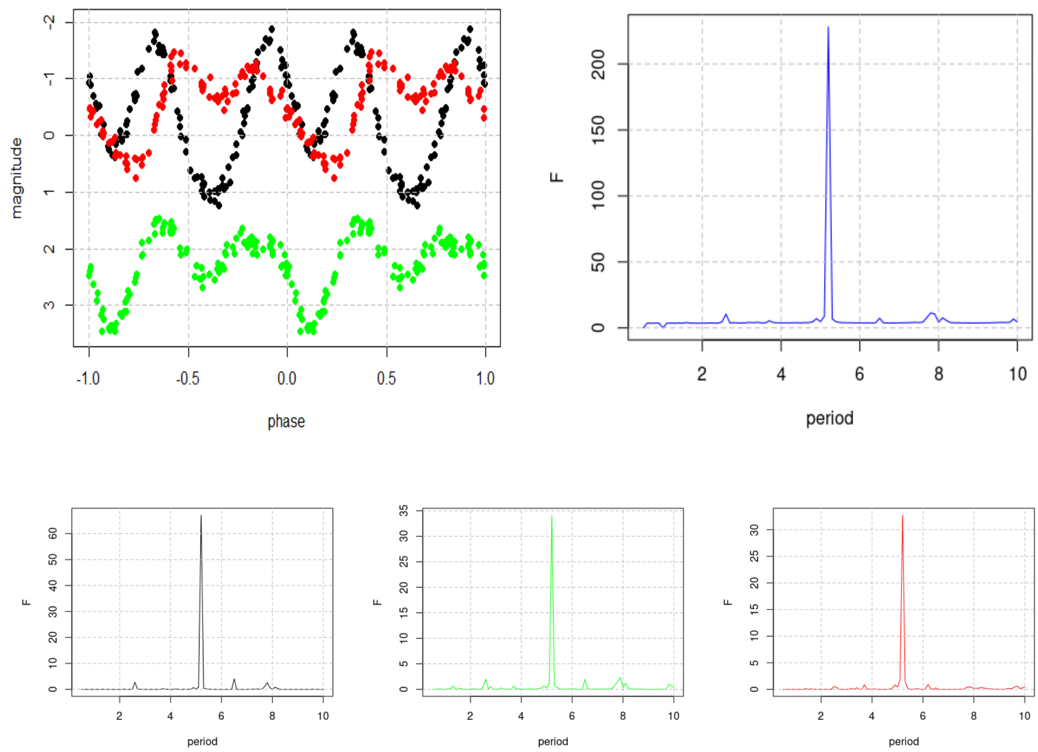


Figure 4.3: (Top left) Simulated light curve from a GPR model for three filters. (Top right) The multiband GPR periodogram for this simulated light curve. (Bottom line) The GPR periodogram for each filter individually.

multi-filter method as those for individual filters which is expected since we had the same amount of data for each filter and the SNR was quite high.

4.3.2 Multiband weighted tPR model

In Section 4.2.2 we discussed how the period of fat-tailed data could be modelled by moving from the Gaussian Process to the t-Process regression model. Furthermore, an extension of the GPR model for data from multiple filters was provided in Section 4.3.1. Here these two approaches are combined and we show a more generalized version of the Multiband GPR model (4.5). The new model we discuss here is the multiple filter extension of the weighted t process regression (4.2) we saw in Section 4.2.3, and we will refer to it as the Multiband weighted tPR model. We follow the same idea as seen in the previous Section for the multiband GPR model and augment our dataset by M dummy variables. In this case the estimated

$\hat{\mathbf{g}}$ values will be given by the following multivariate t distribution,

$$\hat{\mathbf{g}} \sim MVT(\mathbf{W}\mathbf{y}, \nu + N, (\mathbf{K} - \mathbf{W}\mathbf{K})B) \tag{4.7}$$

where,

$$\mathbf{W} = \mathbf{K}[\mathbf{K} + \mathbf{Q}^{-2}\sigma^2\mathbf{I}]^{-1}, \quad B = \frac{\nu - 2 + \mathbf{y}^T(\mathbf{K} + \mathbf{Q}^{-2}\sigma^2\mathbf{I})^{-1}\mathbf{y}}{\nu + N - 2}$$

$\mathbf{K}$ is calculated as seen in the previous Section as (4.6) and $\mathbf{Q}$ is given by (2.18). Note that both these two matrices are of $N \times N$ dimension in this case. We can estimate the parameters by maximizing the marginal likelihood (4.3) as seen in Section 4.2.3.

Again we can use the generalized F statistic to build a multiband weighted tPR periodogram as discussed in the previous Section, we can see an example of it in Figure 4.4 (Top right). This periodogram was calculated for a simulated light curve with period at 5.2 days measured in the Red and Infrared filter. The simulated data can be seen in Figure 4.4 (Top left), in order to generate them we borrowed the sampling from object 3314 form Froebrich et al. (2021) [24], the signal was generated from a weighted t-process and the weights were assumed to follow a Gamma distribution, $\Gamma(3, 10)$. For comparison reasons we also calculated the weighted tPR periodograms for each filter individually, see Figure 4.4 (Bottom line). In this example we notice that although the multiband tPR periodogram has its maximum peak at the correct period, it also has some relatively high peaks for other periods too that are not correct. The individual periodograms however do not exhibit the same behaviour. This is an example where modelling all the filters together acts like adding extra noise to the data. This problem could be avoided however if instead of the kernel (4.6), we used the summation of two separate kernels, one for the time points and one for the filter dummy variables. By doing so we enforce the penalization of data coming from different filters. This extension is not explored further here however and it is left as part of future work.

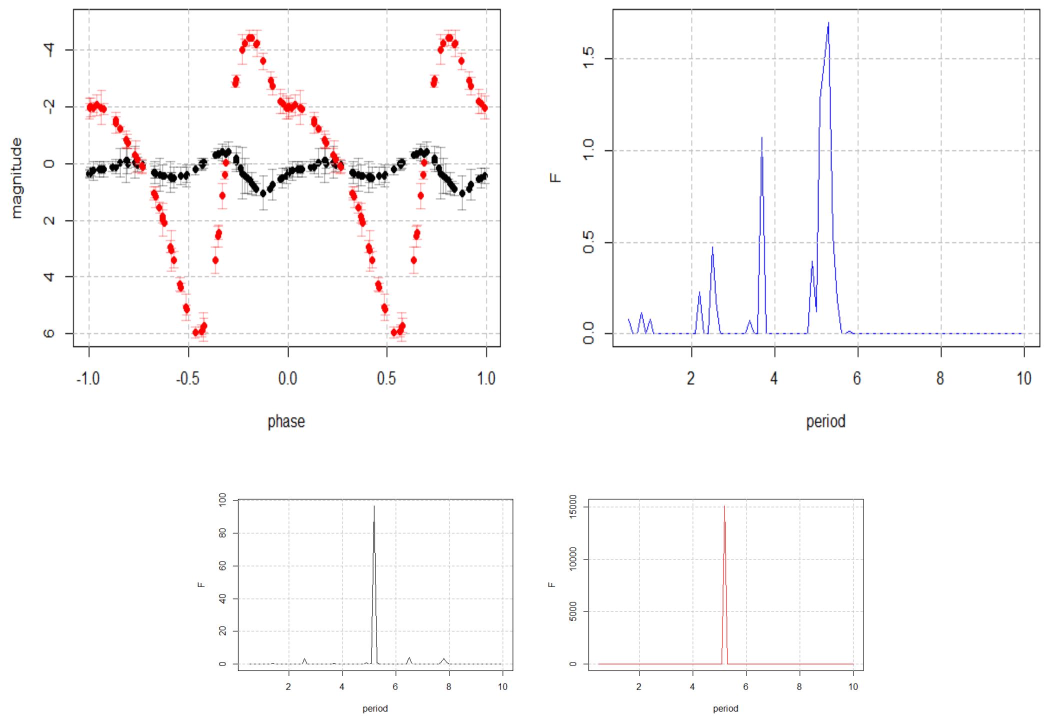


FIGURE 4.4: (Top left) A simulated light curve from the weighted tPR model for two filters. The error bars correspond to the measurement accuracies. (Top right) The weighted tPR periodogram for this light curve. (Bottom) Individual weighted tPR periodograms for each filter.

4.4 Periodic and quasi periodic kernels

We are aiming to estimate the periods of light curves. So far we have worked solely with the periodic kernel (2.15) which first appeared in Mackay (1998) [47]. When using this kernel we make specific assumptions regarding the expected behaviour of the curve that we are investigating. Apart from smoothness we also assume that our curve is *strictly* periodic, meaning that the same shape is repeated at period p. This assumption however might not be always realistic. There are many examples of data exhibiting quasi periodic behaviours; maybe the shape changes every period, or the period changes with time, ect. We would like to be able to take that into account when estimating the periods of light curves. In this Section we see examples of kernels that can incorporate quasi periodic attributes like these.

4.4.1 The periodic kernel

In this Section we will examine closely the periodic kernel of (2.15). Specifically we show how we can obtain this particular kernel. In Mackay (1998) [47] and Williams (2006) [39] we see how the periodic kernel is obtained by projecting our data points to a circle using the stationary squared exponential kernel. This procedure just described is referred to in the literature as Warping or Embedding. It is usually used to introduce non-stationarity by introducing an arbitrary non linear mapping (or warping) $u(\mathbf{x})$ of the input $\mathbf{x}$ (that is the time t inside the kernel function in our case) and then use a stationary covariance function in u-space, $\mathbf{x}$ and $\mathbf{u}$ need not have the same dimensionality.

In general, given any kernel $K(u_j, u_k)$, we can introduce an arbitrary non-linear mapping $\mathbf{x} \to \mathbf{u}(\mathbf{x})$ and define a new kernel

$$D(x_j, x_k) = K(\mathbf{u}(x_j), \mathbf{u}(x_k)), \quad j, k = 1, ..., n$$

To begin with given the SE kernel

$$K(x_j, x_k) = \exp(-\frac{1}{2}\frac{|x_j - x_k|^2}{h^2}) \tag{4.8}$$

where $|\cdot|^2$ denotes the L2 norm, and by introducing the mapping to the circle $\mathbf{u}(\mathbf{x}) = (\cos(\mathbf{x}), \sin(\mathbf{x}))$ we obtain the desired kernel (2.15). We define $\cos(\mathbf{x})$ here as a vector with elements given by $\cos(x_j)$ (similarly for sine). As we see in this case the one-dimensional $\mathbf{x}$ is mapped to a higher dimensional space (2D). See Figure 4.5 for more details regarding the mapping and the behaviour of the resulting covariance function. We notice that points integer times a period apart will return the same kernel value.

Calculations

In order to obtain the periodic kernel we use the SE kernel and project the $\mathbf{x}$ inputs to the circle. We start with the SE kernel given by (4.8). Using the projection

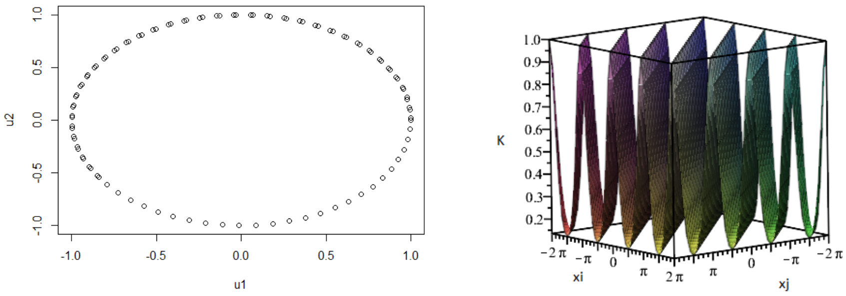


FIGURE 4.5: (Left) Projecting $\mathbf{x}$ to $\mathbf{u} = (\cos(\mathbf{x}), \sin(\mathbf{x}))$ resulting in a circle. (Right) The periodic kernel for different values of $\mathbf{x}$ and period π.

$\mathbf{u} = (\cos(\mathbf{x}), \sin(\mathbf{x}))$ we write the SE as

$$D(x_j, x_k) = \exp\left(-\frac{1}{2}\frac{|\mathbf{u}(x_j) - \mathbf{u}(x_k)|^2}{h^2}\right)$$

Taking the norm we get,

$$|\mathbf{u}(x_j) - \mathbf{u}(x_k)|^2 = (\cos(x_j) - \cos(x_k))^2 + (\sin(x_j) - \sin(x_k))^2$$

$$= 1 + 1 - 2(\cos(x_j)\cos(x_k) + \sin(x_j)\sin(x_k)) =$$

$$= 2 - 2\cos(x_j - x_k)$$

Taking into account that $\cos(2\theta) = 1 - 2\sin^2\theta, \quad \theta = \frac{x_j - x_k}{2}$

$$|\mathbf{u}(x_j) - \mathbf{u}(x_k)|^2 = 4\sin^2(\frac{x_j - x_k}{2})$$

Replacing this into equation (4.8) this results in the desired form of the periodic kernel

$$D(x_j, x_k) = \exp\left(-\frac{1}{2}\frac{4\sin^2((x_j - x_k)/2)}{h^2}\right)$$

Note that for simplicity in the above calculations we assumed the period to be equal to 2π. The kernel in its general form can be seen below:

$$D(x_j, x_k) = \exp\left(\frac{-2\sin^2((\pi/p)(x_j - x_k))}{h^2}\right)$$

4.4.2 The cylinder kernel

As described at the beginning of this Chapter, sometimes the data can deviate from the strict periodic pattern captured by the periodic kernel (2.15). In this Section we describe a kernel that models periodic data whose shape changes with time. We can obtain many kernels with periodic behaviour and different characteristics by following the procedure described above but projecting to different spaces. In this Section we will study the resulting kernel when we warp our data around the cylinder as seen in Figure 4.6. The idea is similar as before, we use the SE kernel as seen at (4.8) and project our input to the cylinder $\mathbf{u} = (\cos(\mathbf{x}), \sin(\mathbf{x}), \mathbf{x})$ and this results in the following kernel.

$$K(x_j, x_k) = \exp\left(\frac{-2\sin^2((\pi/p)(x_j - x_k))}{h^2}\right) \cdot \exp\left(-\frac{1}{2}\frac{|x_j - x_k|^2}{h^2}\right) \tag{4.9}$$

Notice that this is exactly the same kernel we would obtain if we multiply the SE with the periodic one (since multiplication of two kernels results in a kernel, see details in Williams and Rasmussen (2006) [73]). The calculations can be found at the end of this Section. In Duvenaud et al. (2013) [20] the authors refer to this particular kernel as "local periodic" in the sense that we still fit periodic functions but we allow their shape to change. This can be quite useful in real life applications where our data is not perfect and although the main periodic behaviour remains the same the actual shape of the curve can change.

From the geometric point of view, in the previous example where we projected to the circle, for every period a circle is completed, meaning that points passing through the same phase will return identical values no matter how many periods apart they are. When warping around the cylinder, on the other hand, we allow points periods apart to have different values. So when it comes to samples drawn from a prior using this kernel the periodic behaviour will be stable meaning that peaks and dips will appear with period p but the shape of the functions will exhibit different behaviour with time. For more details see Figure 4.7.

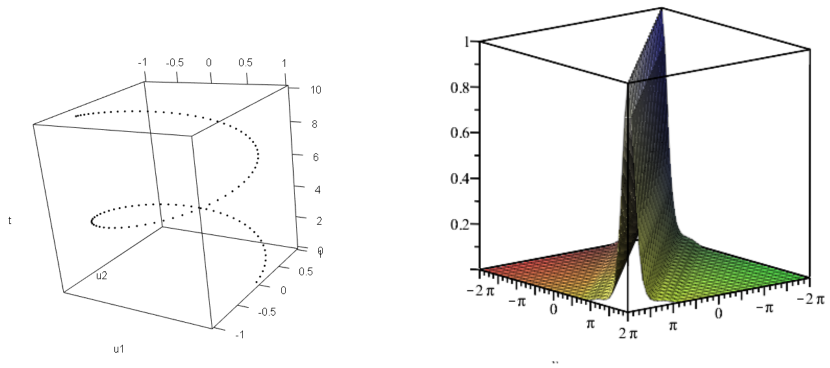


FIGURE 4.6: (Left) Projecting $\mathbf{x}$ to $\mathbf{u}(\mathbf{x}) = (\cos(\mathbf{x}), \sin(\mathbf{x}), \mathbf{x})$ resulting in a cylinder. (Right) The periodic kernel for different values of $\mathbf{x}$ and period π.

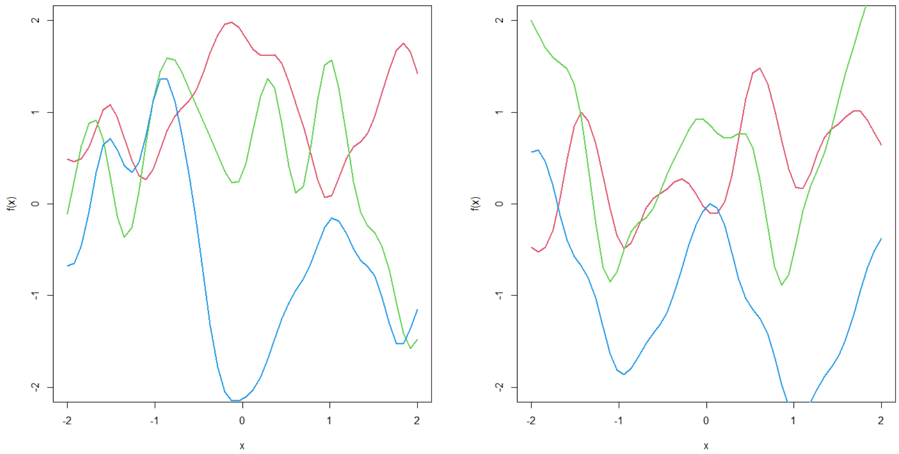


FIGURE 4.7: (Left) Samples from the GP prior with the cylinder kernel. (Right) Samples from the GP prior with the SE+PER kernel.

This should not be confused with the kernel obtained by adding the SE and the periodic one. The difference between this and the SE+PER, is that the latter can be viewed as the noisy version of the periodic one allowing for departures from the strictly periodic kernel, resulting in a more flexible fitting. The shape of the curve under this setting does not change over time, we just allow small deviations from it. A discussion regarding the use of this kernel for the problem of period estimation in light curves can be found in Angus et al. (2018) [4], the authors refer to it as the “quasi periodic” kernel.

Calculations

In order to obtain the cylinder kernel we use the SE kernel and project the $\mathbf{x}$ inputs to the cylinder. We start with the SE as seen in 4.8, and use the projection $\mathbf{u}(\mathbf{x}) = (\cos(\mathbf{x}), \sin(\mathbf{x}), \mathbf{x})$. Taking the norm,

$$|\mathbf{u}(x_j) - \mathbf{u}(x_k)|^2 = (\cos(x_j) - \cos(x_k))^2 + (\sin(x_j) - \sin(x_k))^2 + (x_j - x_k)^2 =$$

$$= 2 - 2\cos(x_j - x_k) + (x_j - x_k)^2 = 4\sin^2(\frac{x_j - x_k}{2}) + (x_j - x_k)^2$$

Replacing this into the SE we get,

$$D(x_j, x_k) = \exp\left(-\frac{1}{2}(\frac{4\sin^2(\frac{x_j - x_k}{2})}{h^2} + \frac{|x_j - x_k|^2}{h^2})\right)$$

And for the general periodic case we can write,

$$D(x_j, x_k) = \exp\left(\frac{-2\sin^2((\pi/p)(x_j - x_k))}{h^2}\right) \cdot \exp\left(-\frac{1}{2}\frac{|x_j - x_k|^2}{h^2}\right)$$

4.4.3 The spiral kernel

Following the same idea and adapting a different warping we will show how a kernel that allows the period to change over time can be obtained. For that purpose we will wrap our inputs around the spiral as seen in Figure 4.8 (Left). As previously we will use as base the SE kernel and use the projection $\mathbf{u}(\mathbf{x}) = (\cos(\mathbf{x})\mathbf{x}, \sin(\mathbf{x})\mathbf{x})$ in order to obtain the following kernel

$$K(x_j, x_k) = \exp\left(\frac{-2x_j x_k \sin^2((\pi/p)(x_j - x_k))}{h^2}\right) \cdot \exp\left(-\frac{1}{2}\frac{|x_j - x_k|^2}{h^2}\right)$$

We notice that the resulting kernel is very similar to the cylinder one except for the periodic part being multiplied by the term $x_j x_k$. The calculations can be seen at the end of this Section. The $x_j x_k$ term changes according to the actual location in time scale of the two points so data integer multiples of phases apart will not have the same correlation (as opposed to the periodic kernel for example). In

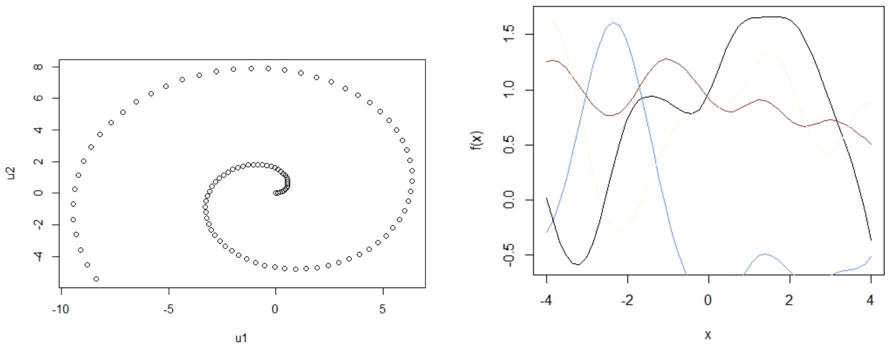


FIGURE 4.8: (Left) Projecting $\mathbf{x}$ to $\mathbf{u}(\mathbf{x}) = (\cos(\mathbf{x})\mathbf{x}, \sin(\mathbf{x})\mathbf{x})$ resulting in a spiral. (Right) Samples from the GP prior with the Spiral kernel.

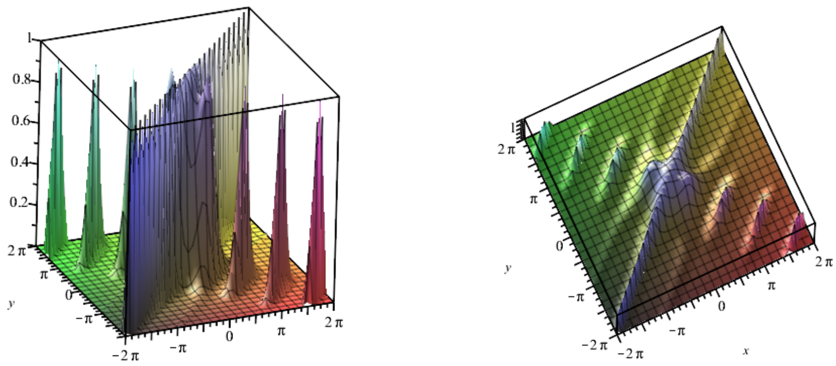


FIGURE 4.9: (Left) Spiral kernel from point of view 1. (Right) Spiral kernel from point of view 2.

Figure 4.8 (Right) we see examples of priors drawn from the spiral kernel. This kernel allows both the period and the shape to change with time. This results in a quite complicated shape which might not be ideal for period estimation but might be useful for modelling other types of data. We won't pursue the exploration of applications of this kernel further in this thesis and we will leave it as part of a future work.

Calculations

In order to obtain the spiral kernel we use the SE kernel and project the $\mathbf{x}$ inputs to the spiral. We start with the SE as seen in 4.8, and use the projection $\mathbf{u}(\mathbf{x}) =$

$(\cos(\mathbf{x})\mathbf{x}, \sin(\mathbf{x})\mathbf{x})$. Taking the norm we get,

$$|\mathbf{u}(x_j) - \mathbf{u}(x_k)|^2 = (\cos(x_j)x_j - \cos(x_k)x_k)^2 + (\sin(x_j)x_j - \sin(x_k)x_k)^2 =$$

$$= x_j^2(\cos^2(x_j)+\sin^2(x_j))+x_k^2(\cos^2(x_k)+\sin^2(x_k))-2x_jx_k(\cos(x_j)\cos(x_k)+\sin(x_j)\sin(x_k)) =$$

$$(x_j-x_k)^2+2x_jx_k(1-(\cos(x_j)\cos(x_k)+\sin(x_j)\sin(x_k))) = (x_j-x_k)^2+4x_jx_k\sin^2(\frac{x_j - x_k}{2})$$

Replacing this into the SE we get,

$$D(x_j, x_k) = \exp\left(-\frac{1}{2}(\frac{4x_jx_k\sin^2(\frac{x_j-x_k}{2})}{h^2} + \frac{(x_j - x_k)^2}{h^2})\right)$$

Which for the general periodic case has the form,

$$D(x_j, x_k) = \exp\left(\frac{-2x_jx_k\sin^2((\pi/p)(x_j - x_k))}{h^2}\right) \cdot \exp\left(-\frac{1}{2}\frac{|x_j - x_k|^2}{h^2}\right)$$

4.5 Applications of generalized F and CVF-tests

In Chapter 3 we discussed the importance of having a reliable statistical testing order to decide if the period estimated is true or if there is no periodic signal in the data. Furthermore, we introduced the generalized F-test (3.2) and CVF-test (3.9) in order to address the problem of period detection when Gaussian process regression is used to model the data. In this Section we show how these tests can be used in exactly the same way for period detection when t-process regression is used, and also a different application of these tests as tools for appropriate kernel selection.

4.5.1 Period detection

In Section 3.4.1 we saw how the generalized F-test can be used under the GPR setting. The matrix $\mathbf{M}_1$ represented the alternative hypothesis, that of a periodic signal from a Gaussian process, and depends on $\mathbf{W}$ (2.16). This holds exactly in the same way when tPR is used instead, as the mean for the two models is given by the same equation (2.16). Numerically however, the matrix $\mathbf{W}$ will be different for the two models, since the parameters were estimated by maximizing a different likelihood. Here we perform the same hypothesis testing analysis for the tPR model as described in the previous Chapter. Note however that in order for the methodology to be applied in Chapter 3 the generalized F statistic was assumed to be a Normal Quadratic form under the null hypothesis. In order to make use of the same test here, we make the same assumption for the errors under the null hypothesis although in reality they follow a student t distribution. We acknowledge that this can lead to less accurate results especially for small degrees of freedom ν and we will investigate the true distribution of the generalized F-statistic under the student t setting as part of a future work. With that in mind we report this hypothesis testing analysis as described in the previous Chapter, and see that we obtain reasonable results even for the hardest case of data generated with $\nu = 3$ degrees of freedom. Note that as $\nu \to \infty$ the t-process is equivalent to the Gaussian process and this is the reason we refer to the case of three degrees of freedom as being the "hardest" scenario for this example.

In order to assess the tests' performance under the t-process setting we conduct an extensive power analysis and draw comparisons between the GPR and tPR models. To begin with, we simulate artificial light curves as seen in Section 3.7.2 with a known period at 5.2 days and their shape generated by a t-process with 3 degrees of freedom and the periodic kernel (2.15). We estimate the power for a series of different signal to noise ratios based on 100 light curves for each SNR value.

The results are reported in Figure 4.10 (Left). For each set of light curves we fit both a tPR (red line) and a GPR (blue line) model, and the power is calculated at a 1-0.999 significance level. The generalized F-test seems to perform well under both settings with the tPR model seeming to perform slightly better for SNR < 0.3. In Figure 4.10 (Right) we perform exactly the same analysis only the simulated data are generated from a GPR model this time. We notice the same behaviour with the models being almost identical in this case.

Table 4.1: Percentage of times the Gaussian process or t process marginal likelihood had their highest peak at the correct period. The columns show the model used and the rows the type of simulated data.

	Fit	
Data	**Gaussian process**	**t process**
Gaussian	91%	91%
t	73%	81%

Having such similar performance is to be expected since the GPR model is a special case of the tPR and the errors under the null hypothesis were assumed to be normal in both cases, that is why they were almost identical for the Gaussian process data. The fact that the generalized F-test performed very well for both settings does not mean that there is no practical use for the tPR model. The tPR model is more flexible and it can provide us in some cases with more accurate results compared to GPR, for example. As shown in Table 4.1 we notice exactly that. For this example we simulate 100 light curves from a t-process with 3 degrees of freedom and 100 light curves from a Gaussian process. We fit to the two data sets both a GPR and a tPR model and we obtain the parameter estimates by maximizing the corresponding marginal likelihood. In principle the likelihood should obtain its maximum value for a period close to the true one. In this example we use a very crude period grid that contains the true value, while the rest of the grid periods are quite away from the correct one, forcing the likelihood to obtain its maximum value for the period at 5.2 days. In this table we show the percentage of light curves whose corresponding likelihoods had their maximum value at the correct period. We notice that when the data come from a Gaussian process we have the same amount of periods correctly identified for both models. When the data come from a t-process however we see that the number of correctly identified periods is higher when the tPR model is used instead of the GPR.

4.5.2 Period detection under the multiband setting

As discussed in Section 4.3 we can use the generalized F and CVF statistics in order to build a periodogram. In a similar manner as described in the previous

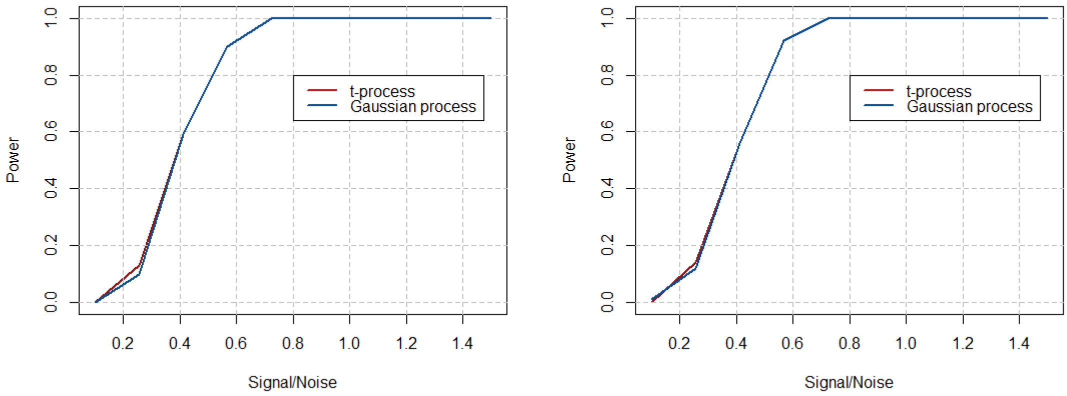


FIGURE 4.10: (Left) The estimated power of the generalized F-test for simulated light curves generated by a t-process with 3 degrees of freedom. The red line shows the power when fitting a t-process, while the blue line shows the power when fitting a Gaussian process. (Right) The same analysis, only the simulated data are produced by a Gaussian process.

Section we can perform hypothesis tests to decide whether our periodogram peaks represent a valid period or not under this multiple filter setting. The generalized F statistic (3.2) is readily used exactly as seen in Section 3.4.1. Under this setting the matrix $\mathbf{M}_1$ representing the multiband GPR model will be calculated as $\mathbf{M}_1 = (\mathbf{I} - \mathbf{W})^\top(\mathbf{I} - \mathbf{W})$ with $\mathbf{W}$ being an $N \times N$ matrix and calculated as described in Section 2.2.3. For the null hypothesis we need a model that assumes a different mean for each filter. This model will be represented by the matrix $\mathbf{M}_0 = \mathbf{I} - \mathbf{X_0}(\mathbf{X_0}^\top\mathbf{X_0})^{-1}\mathbf{X_0}^\top$, where $\mathbf{X_0}$ is an $N \times M+1$ matrix:

$$\mathbf{X_0} = \begin{bmatrix} 1 & d_{1_1} & \dots & d_{M_1} \\ 1 & d_{1_2} & \dots & d_{M_2} \\ \vdots & \vdots & & \vdots \\ 1 & d_{1_N} & \dots & d_{M_N} \end{bmatrix}$$

After calculating the matrices $\mathbf{M}_0$ and $\mathbf{M}_1$ we can perform our hypothesis test in exactly the same way as described in Chapter 3.

Here we assess the performance of the generalized F-test under the multiband GPR setting through a detailed power analysis. We simulate artificial light curves measured in three filters for different signal to noise ratios and we borrow the sampling from the real light curves 7159, measured in the Visual, Red and Infrared filters from Froebrich et al. (2021) [24]. For each filter we generate our signal from a GP with common period at 5.2 days. Each filter is assumed to have a different shape. We estimate the power of the test by generating 100 light curves for each SNR and we fit to each data set the multiband GPR model. The results can be seen in Figure 4.11 (Left), the power of the multiband GPR test is denoted by the blue line. For comparison reasons we include in this plot the power of fitting a GPR model individually for each filter (red, green and black lines). Note that in this Figure the starting value of SNR is 0.1. We notice that the power of the multiband GPR model is higher than the power obtained individually for each filter for SNRs smaller than 0.6. After that the performance of the tests is the same and consistently equal to 1.

To complete our analysis we estimate also the probability of the opposite correct inference, meaning the probability of not rejecting H_0 when H_0 is correct. We calculate first this probability by generating purely noise data and fitting the multiband GPR model to them for different variances and testing against different periods, in all cases it was estimated to be larger than 0.98. Note that this probability is different if there is a periodic signal in our data but we test against the wrong period. In this case the estimated $1-\alpha$ probability is high for periods away from the correct one or integer multiples of it. These results are summarized in Figure 4.11 (Right) where we see the estimated $1-\alpha$ probability for the multiband GPR model (blue line) and the equivalent probabilities for each individual filter. In this case we notice that the results for the multiband GPR model are close to the other lines and perhaps in most cases we get values somewhere in the middle. Thus, from this power analysis we see that the multiband GPR model works well and it can potentially generate better results from those obtained individually for each filter. Note that this analysis is conducted for a significance level $\alpha = 1-0.99$.

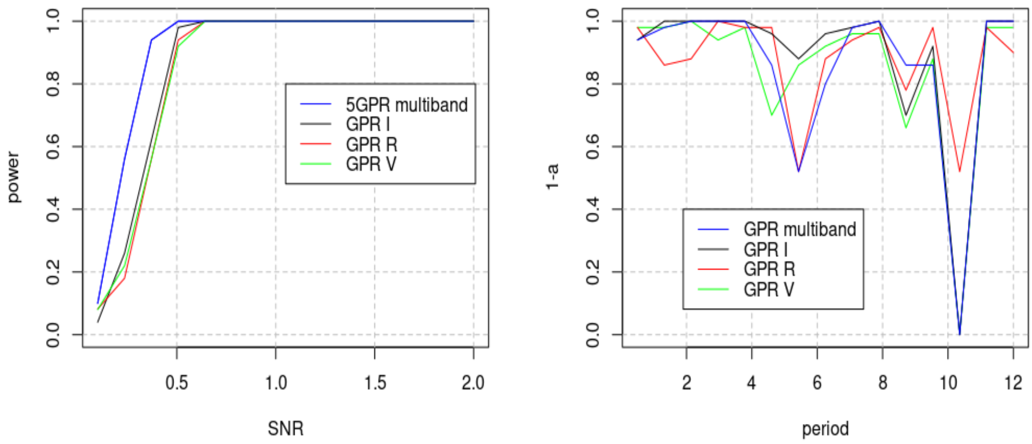


FIGURE 4.11: (Left) The estimated power of the multiband GPR generalized F-test (blue line) for different signal to noise ratios. The red line corresponds to the power of the red filter individually using a GPR model. Similarly the black line is for the infrared and the green for the visual. (Right) The estimated probability of the opposite correct inference when testing against different periods.

4.5.3 Kernel selection

In Section 4.4 we saw that by assuming different periodic kernels we can model different attributes that our data potentially exhibit. It would be useful thus to be able to decide which kernel is more suitable to describe our light curves. In this Section we show how the generalized F and CVF-tests can be used in order to choose between kernels. The general idea here is that we want to compare two different periodic models. In other words,

$$\begin{aligned} &H_0 : \mathcal{M}_0 \text{ model with strict periodic signal at } p \\ &\quad \text{vs} \\ &H_1 : \mathcal{M}_1 \text{ model with quasi periodic signal at } p \end{aligned} \tag{4.10}$$

Under this setting the generalized F statistic is given by (3.2). The matrix $\mathbf{M}_0$ which represents the null hypothesis can be derived here from the GPR model with the periodic kernel (2.15) and it would be calculated as $(\mathbf{I}-\mathbf{W}_0)^\top(\mathbf{I}-\mathbf{W}_0)$, where $\mathbf{W}_0$ is given by (2.16). For the alternative hypothesis we will have a similar matrix but it will dependent on a different kernel, the cylinder one (4.9) for example. We can write $\mathbf{M}_1 = (\mathbf{I} - \mathbf{W}_1)^\top(\mathbf{I} - \mathbf{W}_1)$, where $\mathbf{W}_1$ is the same as $\mathbf{W}$ from (2.16) but for a different kernel. The rest of the hypothesis testing procedure is exactly the same as described in Chapter 3.

We examine the performance of the generalized F-test as a kernel selection tool by performing the following simulation study. First we generate light curves from a GPR model and the periodic kernel (2.15) for different signal to noise ratios. We estimate the power of the F-test against the GPR model with the cylinder (4.9) and quasi periodic kernel. For each SNR we simulate 100 light curves, the period in all cases is fixed and equal to 5.2. The results are summarized in Figure 4.12 (Left). We notice that the power is larger than 0.8 in both cases for SNR > 0.75 which means that the test indeed allows us to select the correct kernel. Moreover, we notice that the power when testing against the cylinder kernel is slightly higher than when testing against the quasi periodic kernel.

Finally, as a second example we generate data from the GPR model but with the cylinder kernel. We perform the same analysis as described in the previous paragraph but we test against the periodic and the quasi periodic kernel in this case. The results can be found in Figure 4.12 (Right). In this case the power for both kernels is slightly lower than that seen in the Left panel of the figure. The estimated power is still high however and larger than 0.8 for SNRs larger than 1. Again we can see that the test allows us to select the correct kernel in most cases. Note that for these examples we borrowed the sampling from the real light curve 7159 measured in the Infrared filter from Froebrich et al. (2021) [24] and the tests were performed for a significance level equal to 0.01.

4.6 Application to real light curves

In this Section we apply the models described in Chapter 4 to two real light curves. For comparison reasons we select the same light curves as the ones used in Chapter 3 and analysed in Section 3.10.

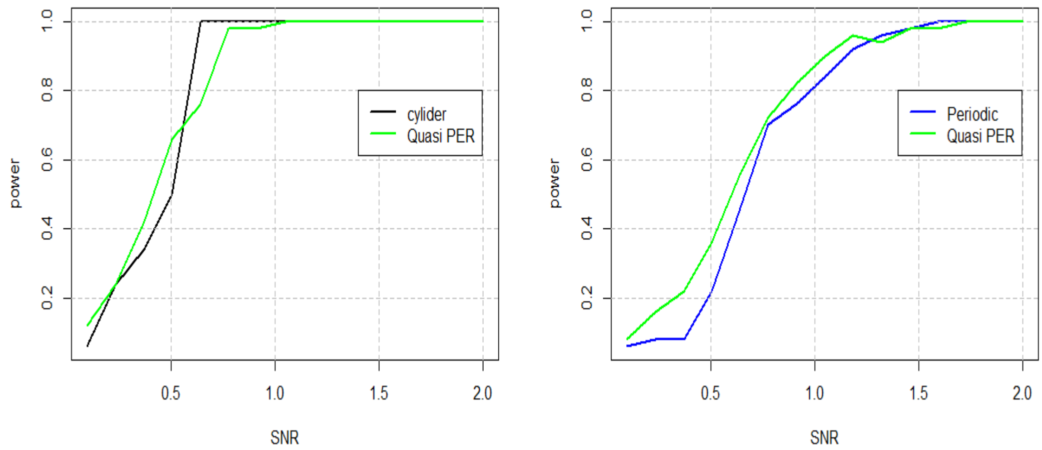


FIGURE 4.12: (Left) The estimated power of the generalized F-test as a kernel selection tool for light curves generated from a periodic kernel. The black line shows the power when we fit the cylinder kernel and the green is for the quasi periodic kernel. (Right) The same analysis, only the simulated data are generated from the cylinder kernel. The blue line corresponds to the power when fitting the periodic kernel.

4.6.1 Example 1

We begin with light curve 6149 from Froebrich et al. (2021) [24] with a studied period at 2.1763 days. More details on this dataset can be found in Section 3.10 along with its plot, see Figure 3.10. Here we calculate the weighted t-process periodogram based on the F-statistic individually for each filter and for three different kernels, namely the Periodic, cylinder and quasi periodic. We choose again the same rough grid as in the previous Chapter searching for periods between 0.5 to 30 days and 1 decimal accuracy. Moreover, we perform the equivalent hypothesis testing for all trial periods and a significance level given by $\alpha = 1 - 0.95^{1/q}$, where $q = 296$ is the number of trial periods. The results are summarized in Table 4.2. We notice that in almost all cases we found as the most dominant period that at 2.2 days, so the choice of kernel did not affect the results in that sense with the exception of the cylinder kernel which indicated as correct the period at 2.6 days for the Blue filter. Moreover we see that in most cases we obtain less extra detected periods compared to the methods described in the previous Chapter. Finally, the multiband GPR model with no weights and the multiband

Table 4.2: The period identified as most important for light curve 6149. In the parenthesis we see the number of extra periods identified as valid. The dash symbolizes no periods identified.

Filter	F tPR per	F tPR cyl	F tPR quasi	mult GPR	mult tPR
B	2.2 (2 extra)	2.6 (1 extra)	2.2 (2 extra)	10.9 (1 extra)	2.2 (6 extra)
R	–	–	–		
I	2.2 (1 extra)	2.2 (0 extra)	2.2 (1 extra)		
V	2.2 (0 extra)	2.2 (0 extra)	2.2 (1 extra)		

weighted tPR model are calculated for all filters. These results are found under the columns “mult GPR” and “mult tPR” respectively in Table 4.2. We see that for this dataset the multiband periodograms do not perform very well. In the case of the multiband GPR with no weights we fail completely to identify the correct period. Note however that this is the only method where we did not use the measurement accuracies (in order to use the model exactly as introduced in Section 4.3.1) which is part of the reason this method did not perform as well as the rest. In the case of the multiband weighted tPR we have a lot more false positives compared to the other methods, but the correct period is identified as the most important.

Table 4.3: This table shows the period identified as most important for the 3314 light curve. In the parenthesis we see the number of further periods identified as correct.

Filter	F tPR	F tPR cyl	F tPR quas	mult GPR	mult tPR
R	13.9 (0 extra)	12.5 (0 extra)	13.9 (0 extra)	13.8 (2 extra)	13.9 (1 extra)
I	13.9 (0 extra)	16.2 (0 extra)	13.9 (0 extra)		

4.6.2 Example 2

Here we examine our second real light curve, object 3314 from Froebrich et al. (2021) [24] with period at 13.8783 days. More details and a plot of the curve can be found in Section 3.10. We perform exactly the same analysis as described in the previous Section and the results can be found in Table 4.3.

We notice that the weighted t-process method with the periodic and quasi periodic kernel identified as valid only one period at 13.9 days similarly to the results from the previous Chapter. When the cylinder kernel is used however we identify different periods for each filter. Because of its large deviation from the period at 13.8783 days in this example we see that using the cylinder kernel might not be an appropriate choice. In situations, however, where no prior information about the dataset is available, we should apply the hypothesis test described in Section 4.5.3 in order to decide which kernel choice seems to describe the data better. The mutiband GPR and weighted tPR periodograms are also calculated and in this example return similar results to the other methods.

4.7 Summary

The Gaussian process regression model has been used in the literature for estimating the period of light curves making fewer assumptions about the data compared to linear regression models. In this Chapter we described an extension to the GPR method based on the t-process regression, which provided more robust results (compared to GPR) in the case of fat-tailed data. We saw how the generalised test introduced in Chapter 3 can be approximately used under the t-process setting too. Our simulations showed that the estimated power of the test is larger that 0.8 for signal to noise ratios over 0.6 even for a t-process with 3 degrees of freedom. When applied to the two light curves studied in this thesis we obtained identical results to GPR.

An extension to modelling simultaneously observations from multiple filters in also considered both for the Gaussian and t processes regression models. After a small modification the generalized F-test can be also used under this mutiband setting. According to our simulation the test performs satisfactorily with a larger power compared to performing the analysis separately for each test. When applied to the real data however this method overall performed less well compared to the other approaches and possible solutions relating to kernel adjusting were discussed.

Finally, we studied the case of non strictly periodic data and we saw that we can model different periodic patterns by choosing different kernels. The generalised F-test was used under this setting as a kernel selection tool. Our simulation results show that this test successfully selects the correct kernel for signal to noise ratios larger than 0.9. When applied to the real data the quasi periodic kernel returned almost the same results with the periodic. The cylinder kernel however seemed to perform better for the first light curve (by identifying less false positive periods) but returned quite different results for the second one.

Chapter 5

On the asymptotic properties of frequency estimates

5.1 Introduction

In the previous Chapters we discussed how to estimate the frequency parameter ω through different model assumptions and optimization techniques, and ways to decide if our estimate represents a true periodic signal. In this Chapter we will explore the asymptotic properties of some frequency estimates considered in this thesis and explore at a more technical level some approaches for modelling unequally-spaced time series. As we saw in Chapter 2 the most common approach in frequency estimation (and light curve modelling) is by assuming a sinusoidal shape for the underlying signal and so the focus of this Chapter will mostly revolve around this model and its extensions.

We consider initially the simplest sinusoidal model as seen in Section 2.2.1 and represented by (2.11). For this model the time observations arrive at regular intervals. This model has been studied in the literature extensively, and the asymptotic properties have been provided. In Walker (1971) [71] for example the authors show that the estimate obtained by approximating the classical periodogram with least squares is asymptotically normal and converges to the true parameter at an $n^{2/3}$

rate. Note that the convergence rate shows us how fast the difference $\hat{\omega} - \omega$ goes to zero as the sample size increases. Furthermore, its asymptotic variance for normally distributed errors is given by:

$$\frac{24\sigma^2}{a^2 + b^2} \tag{5.1}$$

where, a and b are the coefficients of the simple sinusoidal model (2.4) first introduced in Chapter 2. A key thing to note about this result is that it assumes equally spaced observations. The data considered here however are unequally-spaced, and as explained in Section 2.2, this can affect our estimates. For this reason there is a need to develop models that can capture this behaviour. The first step towards that end is to assume an appropriate model for the time sampling. Here we will mostly discuss the jittered sampling model introduced in Beutler (1970) [10] and given by $t_j = j + \delta_j$, where δ_j is a random variable following a known $h(\cdot)$ distribution with zero mean. This model assumes that the time observations come at equal intervals, jittered with additional small disturbances δ_j. Alternative models include the cumulative sum of some positive continuous variables or an ordered uniform sample on $[0, s]$, these are models M1 and M2 introduced in Hall et al. (2000) [27].

For the problem of frequency estimation, the model given by

$$y_j = a\cos(\omega t_j) + b\sin(\omega t_j) + \epsilon_j, \quad \epsilon_j \sim g(0, \sigma^2) \tag{5.2}$$

$$t_j = j + \delta_j \quad j = 1, \ldots, n$$

has been considered in Reimann (1994) [53] and it was studied in great depth. For this model the distribution $g(\cdot)$ of the residuals is considered known. In that work, the author first studies the periodogram estimate of frequency which is the maximizer of (2.7), and shows that periodogram estimates α, β, ω are obtained by minimizing the following modified residual sum of squares:

$$U_n(a, b, \omega) = \sum_{j=1}^{n} y_j^2 - 2\sum_{j=1}^{n} y_j\{a\cos(\omega t_j) + b\sin(\omega t_j)\} + \frac{1}{2}n(a^2 + b^2)$$

Furthermore, they prove that the periodogram estimate is asymptotically equivalent to the least squares one, for ω bounded away from 0 and π, and equally-spaced data, by showing that the difference between U_n and RSS is uniformly $O(1)$. Moreover, the author shows that under certain regularity conditions the asymptotic variance of the periodogram frequency estimate for unequally-spaced data is larger than that for equally-spaced ones since it is affected by the sampling as seen below:

$$n^{3/2}\hat{\omega} \rightarrow N(\omega, \frac{24\sigma^2}{a^2+b^2} + 6[1 - |\phi(2\omega)|^2]) \tag{5.3}$$

where ϕ is the characteristic function of h defined as:

$$\phi(x) = E[e^{ix\delta}]$$

and thus the variance depends on the time point sampling distribution. The maximum likelihood estimate on the other hand does not get affected by the time distribution and the asymptotic variance is the same as that of equally spaced data, given by (5.1). Note that in model 5.2 there are two types of errors introduced. The first type is denoted by ϵ_j and it is the classical residual error, which we will refer to as "observational errors" throughout this thesis. The second type is denoted by δ_j and it corresponds to the deviations from the equally-spaced sampling, in order to avoid confusion we will refer to that as "measurement errors".

In the following Sections we consider extensions of the jittered sinusoidal model (5.2). In Section 5.2 we extend this model assuming residuals with red noise and show the relevant asymptotic results. Moreover, we discuss additional technical details of our results in Section 5.3 and consider additional useful extensions to model (5.2). We assess our results through an extensive simulation study in Section 5.4.In Section 5.5 we demonstrate a simple application of the generalized F-test for linear models with correlated residuals. Finally, we apply our models to two real light curves in Section 5.6.

5.2 Frequency estimation of light curves with red noise

As discussed in previous Chapters, it is common in Astronomy (and other fields) for the data to be contaminated with red noise. In Section 3.8.2 we saw an example of how failing to account for the red noise can lead to false inference. In this Section we extend the jittered sinusoidal model (5.2) to include correlated residuals and we explore the asymptotic properties of the corresponding frequency MLE.

In Section 3.5 we mentioned that red noise is the equivalent of fitting an AR(1) model to the errors when the data are sampled regularly. Since our observations are irregularly sampled, assuming a correlation structure for our errors that takes into account this irregularity in sampling might be a more realistic approach. With this in mind we consider the following model,

$$y_j = \alpha\cos(\omega t_j) + \beta\sin(\omega t_j) + \epsilon_{t_j}, \quad j = 1, \dots, n \tag{5.4}$$

$$\epsilon_{t_j} = \rho^{|t_j - t_{j-1}|}\epsilon_{t_{j-1}} + \tilde{u}_j, \quad \tilde{u}_j \sim g(0, \sigma^2 \frac{1-\rho^{2|t_j-t_{j-1}|}}{1-\rho^2})$$

$$t_j = j + \delta_j, \quad \delta_j \sim h(\cdot)$$

where ρ corresponds to the correlation between two data points and it takes values in the interval $(-1, 1)$. In this model the error dependence is seen as a decreasing function of time difference, and we will refer to it as the red noise jittered sinusoidal model. We can view the second line of the above model as an AR(1) process where not all entries have been observed. In order to understand this better, assume a hypothetical series of observational errors ϵ_j consisting of $J > n$ entries containing the values observed at equal spaces spanning from our first to our last actual observation. These hypothetical errors can be written as an AR(1) process as seen below.

$$\epsilon_j = \rho\epsilon_{j-1} + u_j, \quad u_j \sim g(0, \sigma^2)$$

and for $k > 0$, the observational errors, as seen in Paolella (2018) [51], can be written as:

$$\epsilon_j = \sum_{i=0}^{\infty} \rho^i u_{j-i} = u_j + \rho u_{j-1} + ... + \rho^k u_{j-k} + ...$$

which we can further write as,

$$= u_j + ... + \rho^{k-1} u_{j-k-1} + \rho^k \sum_{i=0}^{\infty} \rho^i u_{j-k-i} = \sum_{i=0}^{k-1} \rho^i u_{j-i} + \rho^k \sum_{i=0}^{\infty} \rho^i u_{j-k-i}$$

If we denote $\tilde{u}_j = \sum_{i=0}^{k-1} \rho^i u_{j-i}$ then we can write

$$\epsilon_j = \rho^k \epsilon_{j-k} + \tilde{u}_j$$

which gives us the relation between two errors k times apart. This for our problem translates as the distance between two consecutively observed time points, e.g. $k = |t_j - t_{j-1}|$. Thus, we can write the relation between two consecutive observed errors as $\epsilon_{t_{j-1}} = \rho^{|t_j - t_{j-1}|} \epsilon_{t_{j-1}} + \tilde{u}_j$ as seen in (5.4). If we further, denote $q_j = \frac{1-\rho^{2|t_j - t_{j-1}|}}{1-\rho^2}$ then, $u_j = q_j^{-1/2} \tilde{u}_j \sim N(0, \sigma^2)$. The pdf of our data will be of the following form:

$$f_j(x_j, \theta) = g(q_j^{-1/2} \tilde{u}_j) h(t_j - j) = \tag{5.5}$$

$$g(q_j^{-1/2} \{y_j - \alpha \cos(\omega t_j) - \beta \sin(\omega t_j) - \rho^{|t_j - t_{j-1}|} (y_{j-1} - \alpha \cos(\omega t_{j-1}) - \beta \sin(\omega t_{j-1}))\}) h(t_j - j)$$

In order to perform inference we need to estimate parameters $\theta = (\alpha, \beta, \omega)$, this can be done by maximizing the log-likelihood of the model (5.4) for our data $x_j = (y_j, t_j)$ which is given by

$$\log \prod_{j=1}^{n} f_j(x_j, \theta)$$

Note that because of the imposed structure on our observation times, our observations depend on each other via the ϵ_j correlations, and of course time is a randomly distributed variable, we cannot apply the standard MLE asymptotic results. Under certain regularity conditions, it can be shown that the ML estimate of the frequency parameter $\hat{\omega}$, for this noise setting, is consistent and asymptotically normal and converges to the true parameter at an $O(n^{-2/3})$ rate. This is the same convergence rate as for the iid model (5.2) as shown in Reimann (1994) [53]. The main result is stated formally below in the form of a theorem.

Theorem 5.1. *For the red noise jittered sinusoidal model* (5.4) *it can be shown that under the following regularity conditions, and for $a^2+b^2 \neq 0$, $\omega \neq \kappa\pi$ where $\kappa \in \mathbb{Z}$, $\hat{\theta}$ is consistent and asymptotically normal with mean $\theta = (\alpha, \beta, \omega)$ and covariance matrix*

$$\mathrm{Cov}(\mathbf{N}^{1/2}(\hat{\theta}-\theta)) = (1-\rho^2)^{-1} k_p^{-1} 2 I^{-1} \mathbf{\Gamma}^{-1}$$

where $\mathbf{N} = \mathrm{diag}[n, n, n^3]$, $I = \int g'(u)^2 \frac{1}{g(u)} du$,

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & 0 & \frac{b}{2} \\ 0 & 1 & -\frac{a}{2} \\ \frac{b}{2} & -\frac{a}{2} & \frac{a^2+b^2}{3} \end{bmatrix}$$

and,

$$k_p = E\left(\frac{1}{1-\rho^{2|\delta_1-\delta_2+1|}}\right) + E\left(\frac{\rho^{2|\delta_1-\delta_2+1|}}{1-\rho^{2|\delta_1-\delta_2+1|}}\right) - 2\cos(\omega) E\left(\frac{\rho^{|\delta_1-\delta_2+1|}}{1-\rho^{2|\delta_1-\delta_2+1|}}\right) \quad (5.6)$$

***Conditions**:*

***(1)** for $r > 0$:*

$$\int |\frac{g'}{g}(u)|^{2+r} g(u) du < \infty \quad , \quad \int \left(\frac{\partial^2}{\partial u^2} g^{1/2}(u)\right)^2 du < \infty$$

(2)

$$\forall \eta > 0, \quad \sup_{|t|>\eta} |\phi(t)| < 1$$

Proof. In order to prove Theorem 5.1 we will use Theorem III 4.1 in Ibragimov and Has'Minski (2013) [33], p.191. which states that: The maximum likelihood estimator $\hat{\theta}$ of some observations x_j with pdf $f_j(x, \theta)$ depending on j, is asymptotically normal with variance Ψ_n^{-2} when the four conditions (A)-(D) in Reimann (1994) [53] page 42 are satisfied. Note that conditions (A)-(D) are also formally stated in Appendix A.1. Condition (C) is related to condition (1) from Theorem 5.1, which is a restriction where we only consider smooth likelihoods, and conditions (B) and (D) are connected with condition (2) which prevents the RSS from being as small as for the true frequency at multiples of ω.

Here $\Psi_n^2 = \sum_{j=1}^n I_j(\theta)$, is the information of the first n observations, and it must be positive definite. We will show that these conditions are met, following closely Reimann (1994) [53], but for the red noise setting. (Note that we only show in detail the steps that are different to those in Reimann's proof.)

Fisher's information matrix $I_j(\theta)$ is given by:

$$I_j(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log f_j(x_j;\theta)\right)^2 |_\theta\right] = E\left[\left(\frac{\frac{\partial}{\partial\theta} f_j(x_j;\theta)}{f_j(x_j;\theta)}\right)^2 |_\theta\right] =$$

$$\int \left[\frac{\partial}{\partial\theta} f_j(x_j;\theta)\right]\left[\frac{\partial}{\partial\theta} f_j(x_j;\theta)\right]^T \frac{dx_j}{f_j(x_j;\theta)} =$$

$$= \int \mathbf{C}\frac{dx_j}{f_j(x_j;\theta)}$$

In order to obtain the entries of matrix $\mathbf{C}$ we calculate the partial derivatives of f_j as seen below:

$$A = \frac{\partial}{\partial a} f_j(x_j,\theta) = g'(u)(-\cos(\omega t_j) + \rho^{|t_j - t_{j-1}|}\cos(\omega t_{j-1}))h(t_j - j)\frac{1-\rho^2}{1-\rho^{2|t_j - t_{j-1}|}}$$

$$B = \frac{\partial}{\partial b} f_j(x_j,\theta) = g'(u)(-\sin(\omega t_j) + \rho^{|t_j - t_{j-1}|}\sin(\omega t_{j-1}))h(t_j - j)\frac{1-\rho^2}{1-\rho^{2|t_j - t_{j-1}|}}$$

$$C = \frac{\partial}{\partial \omega} f_j(x_j,\theta) = (a\sin(\omega t_j)t_j - b\cos(\omega t_j)t_j - a\rho^{|t_j - t_{j-1}|}\sin(\omega t_{j-1})t_{j-1} + b\rho\cos(\omega t_{j-1})t_{j-1})\times$$
$$g'(u)h(t_j - j)\frac{1-\rho^2}{1-\rho^{2|t_j - t_{j-1}|}}$$

Thus $\mathbf{C}$ is given by:

$$\mathbf{C} = g'(u)^2 h(t_j - j)^2 \begin{bmatrix} A^2 & AB & AC \\ * & B^2 & BC \\ * & * & C^2 \end{bmatrix}$$

Fisher's information then can be written as

$$I_j(\theta) = \int g'(u)^2 h^2(t_j-j)\tilde{C}\frac{dx_j}{g(u)h(t_j-j)} = \int\int \frac{g'(u)^2}{g(u)}\tilde{C}h(t_j-j)dudt = \int \frac{g'(u)^2}{g(u)}duE[\tilde{C}] =$$

$$= IE[\tilde{C}]$$

Finally, the matrix Ψ^2 will be of the following form:

$$\Psi_n^2 = k_p(1-\rho^2)\frac{I}{2}\begin{bmatrix} n+O(1) & O(1) & \frac{b}{2}n^2+O(n) \\ * & n+O(1) & -\frac{a}{2}n^2+O(n) \\ * & * & \frac{a^2+b^2}{3}n^3+O(n^2) \end{bmatrix}$$

$$= k_p(1-\rho^2)\frac{I}{2}\mathbf{N}^{1/2}\left\{\mathbf{\Gamma}+O(n^{-1})\right\}\mathbf{N}^{1/2}$$

The first entry of matrix Ψ_n^2 is calculated as:

$$\Psi^2_{[1,1]} = I(1-\rho^2)\sum_{j=1}^{n}\{E(\frac{\rho^{2|t_j-t_{j-1}|}\cos^2(\omega t_{j-1})}{1-\rho^{2|t_j-t_{j-1}|}}) + E(\frac{\cos^2(\omega t_j)}{1-\rho^{2|t_j-t_{j-1}|}})-$$

$$-E(\frac{2\rho^{|t_j-t_{j-1}|}\cos(\omega t_j)\cos(\omega t_{j-1})}{1-\rho^{2|t_j-t_{j-1}|}})\}$$

$$= \frac{I}{2}(1-\rho^2)\sum_{j=1}^{n}\{E(\frac{\rho^{2|\delta_1-\delta_2+1|}}{1-\rho^{2|\delta_1-\delta_2+1|}}) + \cos(2\omega(j-1))E(\frac{\rho^{2|\delta_1-\delta_2+1|}}{1-\rho^{2|\delta_1-\delta_2+1|}}\cos(2\omega\delta))+$$

$$+E(\frac{1}{1-\rho^{2|\delta_1-\delta_2+1|}}) + \cos(2\omega j)E(\frac{1}{1-\rho^{2|\delta_1-\delta_2+1|}}\cos(2\omega\delta))-$$

$$-2\cos(\omega)E(\frac{\rho^{|\delta_1-\delta_2+1|}}{1-\rho^{2|\delta_1-\delta_2+1|}}) - \cos(2\omega(j-1))E(\frac{\rho^{|\delta_1-\delta_2+1|}}{1-\rho^{2|\delta_1-\delta_2+1|}}\cos(\delta_1+\delta_2))\}$$

$$= \frac{I}{2}(1-\rho^2)nk_p + O(1)$$

We calculate in a similar manner the rest of the entries. For $a^2+b^2 \neq 0$ and $\omega \neq \kappa\pi$, Ψ_n is positive definite, so we can write:

$$\Psi_n^{-2} = k_p^{-1}(1-\rho^2)^{-1}2I^{-1}\mathbf{N}^{-1/2}\left\{\mathbf{\Gamma}^{-1}+O(n^{-1})\right\}\mathbf{N}^{-1/2}$$

The square roots of the above matrices are of the following form:

$$\Psi_n = \frac{\sqrt{I}}{\sqrt{2}} k_p^{1/2}(1-\rho^2)^{1/2}\mathbf{N}^{1/4}\{\mathbf{\Gamma}^{1/2} + O(n^{-1})\}\mathbf{N}^{1/4}$$

$$\Psi_n^{-1} = \frac{\sqrt{2}}{\sqrt{I}} k_p^{-1/2}(1-\rho^2)^{-1/2}\mathbf{N}^{-1/4}\{\mathbf{\Gamma}^{-1/2} + O(n^{-1})\}\mathbf{N}^{-1/4}$$

After obtaining these four matrices it can be shown by following Reimann (1994) [53] that conditions (A)-(D) are satisfied. Thus Theorem III from Ibragimov and Has'Minskii (2013) [33] can be applied and that concludes the proof. Conditions (A)-(D) are formally stated in Appendix A.1.

□

Discussion

Note that the bounds on α, β and ω are needed in order to ensure uniformity of convergence. At this stage it is important to note that Theorem 5.1 does not make specific assumptions about the distributions g and h of the observational and measurement errors respectively. Regarding the latter, the only requirement is that h is a zero mean distribution. While this gives us some flexibility regarding the distribution of the δ_j's, it results in an asymptotic variance depending on k_p which is not in a closed form.

A common choice for the distribution h is to use Uniform. In Reimann (1994) [53] for example, the author generates artificial light curve sampling using the jittered model and assumed that $\delta \sim U(-1/6, 1/6)$. This particular choice of the parameters of the Uniform was made in order to reflect the fact that observations are usually made at night time. We can show that if we specify the distribution of the measurement errors δ_j's to be a $U(-d, d), \quad d < 1/2$, we can write k_p in a closed form as seen below.

$$k_p = 1 + \frac{\ln(\frac{1-\rho^{-4d+2}}{1-\rho^{4d+2}})}{8d\ln(\rho)} - 2\cos(\omega)\frac{\ln(\frac{(1+\rho^{2d+1})(1-\rho^{-2d+1})}{(1-\rho^{2d+1})(1+\rho^{-2d+1})})}{8d\ln(\rho)} \tag{5.7}$$

We obtain this result by conducting the following calculations.

$$E(\frac{1}{1-\rho^{2|\delta_1-\delta_2+1|}}) = \frac{1}{4d}\int_{-2d+1}^{2d+1}\frac{1}{1-\rho^{2x}}dx = \frac{1}{8d\ln\rho}\int_{\rho^{-4d+2}}^{\rho^{4d+2}}\frac{1}{u(1-u)}du =, \quad u=\rho^{2x}$$

$$= 1 + \frac{\ln(\frac{1-\rho^{-4d+2}}{1-\rho^{4d+2}})}{8d\ln\rho} \tag{5.8}$$

$$E(\frac{\rho^{2|\delta_1-\delta_2+1|}}{1-\rho^{2|\delta_1-\delta_2+1|}}) = E(\frac{1}{1-\rho^{2|\delta_1-\delta_2+1|}}) - 1 \tag{5.9}$$

$$E(\frac{\rho^{|\delta_1-\delta_2+1|}}{1-\rho^{2|\delta_1-\delta_2+1|}}) = \frac{1}{4d}\int_{-2d+1}^{2d+1}\frac{\rho^x}{1-\rho^{2x}}dx = \frac{1}{4d\ln\rho}\int_{\rho^{-2d+1}}^{2d+1}\frac{1}{1-v^2}dv =, \quad v=\rho^x$$

$$= \frac{\ln(\frac{(1+\rho^{2d+1})(1-\rho^{-2d+1})}{(1-\rho^{2d+1})(1+\rho^{-2d+1})})}{8d\ln(\rho)} \tag{5.10}$$

By combining equations (5.8), (5.9) and (5.10) we obtain the closed form for k_p as seen in (5.7).

Moreover, if we select the distribution of observational errors g to be a Normal with zero mean and variance σ^2 we can obtained a more simplified result. Specifically for the frequency estimate, when the errors are normally distributed, we have:

$$n^{3/2}\hat{\omega} \to N(n^{3/2}\omega, \frac{24\sigma^2}{\alpha^2+\beta^2}(1-\rho^2)^{-1}k_p^{-1}) \tag{5.11}$$

This is the same variance as for the jittered model in (5.2) shown in Reimann (1994) [53] re-scaled by $1/k_p(1-\rho^2)$, depending on the correlation parameter, frequency and also on the sampling of the data through parameter d.

5.2.1 Sinusoidal model with AR(1) residuals

In this Section we consider a simpler version of the red noise jittered sinusoidal model that could be useful in some applications. Specifically, we assume that although the time points are unequally-spaced the corresponding observational errors are not and can be sufficiently modelled by an autoregressive process at lag 1. In other words, under this setting we assume that the correlations between the errors are the same regardless of the time spacings. This model is given below and we will refer to it as the AR(1) jittered sinusoidal model.

$$y_j = \alpha\cos(\omega t_j) + \beta\sin(\omega t_j) + \epsilon_j, \quad j = 1,\ldots,n \tag{5.12}$$

$$\epsilon_j = \rho\epsilon_{j-1} + u_j, \quad u_j \sim g(0,\sigma^2)$$

$$t_j = j + \delta_j, \quad \delta_j \sim h(\cdot)$$

It can be easily shown that this model is a special case of the red noise jittered sinusoidal model (5.4). If we assume that the difference in sampling does not affect the relation between the observational errors then the second line of model (5.4) collapses to the second line of model (5.12). Under this setting we can estimate our parameters in a similar way as seen previously by maximizing the likelihood of our model. The pdf of our data x_j in this case will be of the following form,

$$f_j(x_j,\theta) = g(\epsilon_j - \rho\epsilon_{j-1})h(t_j - j) = \tag{5.13}$$

$$= g(y_j - \alpha\cos(\omega t_j) - \beta\sin(\omega t_j) - \rho(y_{j-1} - \alpha\cos(\omega t_{j-1}) - \beta\sin(\omega t_{j-1})))h(t_j - j)$$

which is clearly a special case of the jittered red noise pdf (5.5) when $\rho^{|t_j - t_{j-1}|} = \rho^1$. It can be easily shown that a similar asymptotic result to Theorem 5.1 holds for the estimated parameters under this setting. The details are formally stated in the following corollary.

Corollary 5.2. *For the AR(1) jittered sinusoidal model described in* (5.12) *it can be shown that under the regularity conditions (1)-(2) stated in Theorem 5.1, and for the same parameter restrictions, $\hat{\theta}$ is consistent and asymptotically normal with mean $\theta = (\alpha,\beta,\omega)$ and covariance matrix*

$$\mathrm{Cov}(\mathbf{N}^{1/2}(\hat{\theta} - \theta)) = k^{-1}2I^{-1}\mathbf{\Gamma}^{-1}$$

where $k = 1 + \rho^2 - 2\rho\cos(\omega)$, $\mathbf{N} = \mathrm{diag}[n, n, n^3]$, $I = \int g'(u)^2 \frac{1}{g(u)} du$ *and*

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & 0 & \frac{b}{2} \\ 0 & 1 & -\frac{a}{2} \\ \frac{b}{2} & -\frac{a}{2} & \frac{a^2+b^2}{3} \end{bmatrix}$$

Discussion

This result is very similar to the one for the general case only simpler since the asymptotic variance now does not depend on the sampling. Details on the proof can be found in Appendix A.2. Note that again we do not specify completely the distributions h and g and we get a closed form for the variance without the additional assumption that the measurement errors follow a Uniform distribution. Similarly to before, if we choose g to be a normal we get the following result for the frequency parameter which is of primal interest.

$$n^{3/2}\hat{\omega} \to N(n^{3/2}\omega, \frac{24\sigma^2}{\alpha^2 + \beta^2}(1 + \rho^2 - 2\rho\cos(\omega))^{-1}) \tag{5.14}$$

The asymptotic variance of the frequency estimate depends on both the frequency and the correlation of the residuals. Note that if our errors are iid, ρ is zero and thus the variance from (5.14) reduces to the variance for the iid model as seen in Reimann (1994) [53].

The effect of frequency is more dramatic and values of ω close to $\kappa\pi$, $\kappa \in \mathbb{Z}$ result to extreme variances. See for example Figure 5.1 (Left). In order to understand better how changes in ρ and ω affect the asymptotic variance see Figure 5.1 (Right). In this picture we show how $\sqrt{k} = \sqrt{1 + \rho^2 - 2\rho\cos(\omega)}$ changes when varying the angle ω and radius ρ. For example $\sqrt{k}$ gets larger when changing the angle from $\pi/4$ (black colour) to $3\pi/4$ (blue colour), and thus in this example the variance decreases. Similarly if we change the radius from 0.5 (black colour) to 0.8 (red) the variance again gets smaller.

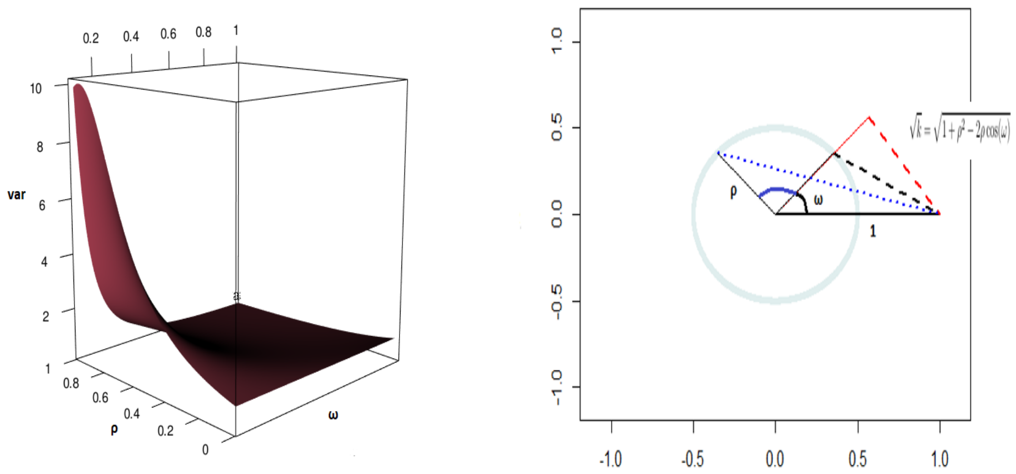


FIGURE 5.1: (Left) The variance of the jittered AR1 model for varying values of ρ and ω. (Right) A visual example as to how changes in frequency and correlation can affect our variance.

5.3 General remarks

5.3.1 About the asymptotic term O(1) in the proofs

From Theorem 5.1 we saw that $\hat{\omega}$ converges to the true parameter at an $O(n^{-3/2})$ rate. A key part of the equivalent proof was to show that,

$$\sum_{j=1}^{n} \left(E[\cos(2\omega\delta)]\cos(2\omega j) - E[\sin(2\omega\delta)]\sin(2\omega j)\right) = O(1) \tag{5.15}$$

In this Section we provide a brief explanation as to why equation (5.15) holds. In simple words this means that the summation above is bound by a constant regardless of how big n is. We start by examining the expectations $E[\cos(2\omega\delta)]$ and $E[\sin(2\omega\delta)]$ that do not depend on the sample size anyway. They are both expectations of transformations of variable δ which as we saw before is assumed to have zero mean. Since δ is centred around zero $2\omega\delta$ will be also centred around zero. This means that the expectation of $\sin(2\omega\delta)$ will be zero by symmetry, or in other words $E[\sin(2\omega\delta)] = 0$.

To understand this better we can look at Figure 5.2 (Left), that shows the unit circle defined by $\cos(2\omega\delta)$ (x-axis) and $\sin(2\omega\delta)$ (y-axis). For this example we

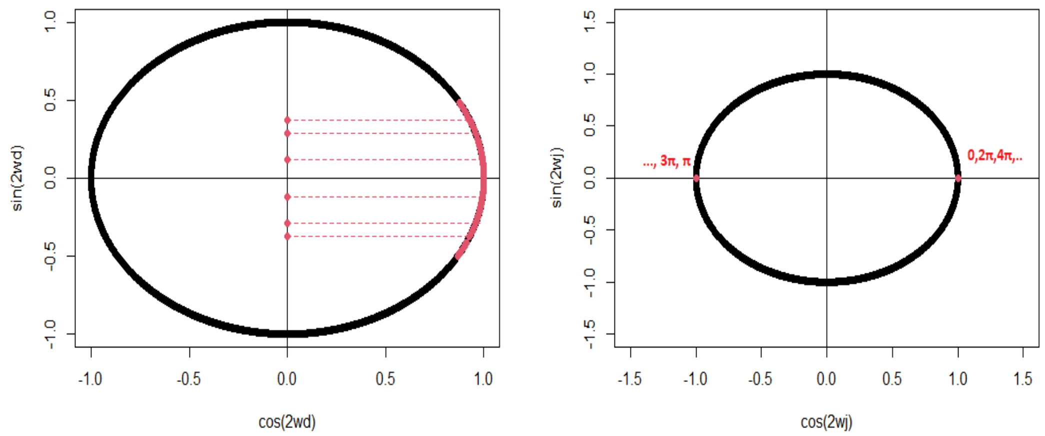


FIGURE 5.2: (Left) Example using the unit circle for the expectation of $\sin(2\omega\delta)$. (Right) An example of the unit circle. The red dots show the values of $\cos(\omega)$ for $\omega = 0, \pi, 2\pi, ...$

assume δ to follow a Uniform distribution, specifically $\delta \sim U(-1/6, 1/6)$. We generate 100 points and the frequency is taken equal to 0.5. The red dots in the plot on top of the unit circle correspond to the data points $(\cos(2 \cdot 0.5\delta_j), \sin(2 \cdot 0.5\delta_j))$, $\quad j = 1, ..., 100$. The dashed lines show the projections of some of these points on the y-sine axis. We can see that the data points are equally spread around zero, that means that their mean will be zero.

Since $E[\sin(2\omega\delta)] = 0$, that means we only have to focus on the first part of the summation. $E[\cos(2\omega\delta)]$ is a constant that does not depend on n, so we just have to show that $\sum_{j=1}^{n} \cos(2\omega j)$ does not depend on the sample size. We use

$$\sum_{j=1}^{n} \cos(2\omega j) = \frac{\sin(n\omega)}{\sin(\omega)} \cos(\omega(n+1))$$

so that,

$$\sum_{j=1}^{n} E[\cos(2\omega\delta)] \cos(2\omega j) = E[\cos(2\omega\delta)] \sum_{j=1}^{n} \cos(2\omega j)$$

$$= E[\cos(2\omega\delta)] \frac{\sin(n\omega)}{\sin(\omega)} \cos(\omega(n+1)) = c \sin(n\omega) \cos(\omega(n+1))$$

where $c = E[\cos(2\omega\delta)]/\sin(\omega)$ is a constant and since the quantity $\sin(n\omega)\cos(\omega(n+1))$ is at most 1 we can write,

$$\sum_{j=1}^{n} E[\cos(2\omega\delta)]\cos(2\omega j) \leq c \quad \text{or,} \quad \sum_{j=1}^{n} E[\cos(2\omega\delta)]\cos(2\omega j) = O(1)$$

.

5.3.2 Further extensions

The main structure of the jittered sinusoidal model is quite simple and that could be restrictive in real applications when other models are preferred. For example as discussed in Section 2.2.1 a simple harmonic model with an additional intercept term is widely used. Moreover, in astronomy is common to have additional information regarding the accuracy of the measurements, which is usually taken into account in the form of a weighted regression. Of course these are only but a few examples of potential alterations for model (5.2) that one could apply. In this Section we will show that an equivalent asymptotic result to the one discussed in Section 5.2 holds for these two examples too.

Model with intercept

We consider the jittered sinusoidal model seen in Reimann (1994) [53] but with an additional intercept term.

$$y_j = c + \alpha\cos(\omega t_j) + \beta\sin(\omega t_j) + \epsilon_j, \quad j = 1, \ldots, n \tag{5.16}$$

$$t_j = j + \delta_j$$

The pdf for our x_j data can be written similarly in the following way,

$$f_j(x_j, \theta) = g(y_j - c - \alpha\cos(\omega t_j) - \beta\sin(\omega t_j))h(t_j - j) \tag{5.17}$$

and we can show that exactly the same asymptotic result as seen in Reimann (1994) [53] holds under this setting too. Here we have an extra parameter to estimate and we can denote our new parameter vector as $\theta = (a, b, c, \omega)$. The result is summarized in the following Theorem.

Theorem 5.3. *For the jittered sinusoidal model with the intercept term* (5.16) *it can be shown that under the regularity conditions (1)-(2) and the same parameter restrictions as seen in Theorem 5.1, $\hat{\theta}$ is consistent and asymptotically normal with mean θ and covariance matrix*

$$\mathrm{Cov}(\mathbf{N}^{1/2}(\hat{\theta} - \theta)) = 2I^{-1}\mathbf{\Gamma}_c^{-1}$$

where $\mathbf{N} = \mathrm{diag}[n, n, n^3]$, $I = \int g'(u)^2 \frac{1}{g(u)} du$ *and*

$$\mathbf{\Gamma}_c = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{b}{2} \\ 0 & 0 & 1 & -\frac{a}{2} \\ 0 & \frac{b}{2} & -\frac{a}{2} & \frac{a^2+b^2}{3} \end{bmatrix}$$

The specifics of the proof can be found in Appendix A.3. Note that the matrix $\mathbf{\Gamma}_c$ includes the matrix $\mathbf{\Gamma}$ as seen in Reimann (1994) [53]. Specifically for the frequency estimate, when the errors are normally distributed, we have :

$$n^{3/2}\hat{\omega} \to N(n^{3/2}\omega, \frac{24\sigma^2}{\alpha^2 + \beta^2})$$

Which is the same variance we obtained for the frequency for the model without the intercept term. Note, that from these calculations it is easy to see that Theorem 5.1 holds even if we include the intercept term in the red noise jittered sinusoidal model (5.4).

Model with weights

We can take the measurement accuracies s_j into account in the form of some known weights, which we will denote as $q_j = 1/s_j$. Under this setting the jittered model (5.2) can be written as:

$$q_j y_j = q_j \alpha \cos(\omega t_j) + q_j \beta \sin(\omega t_j) + \epsilon_j, \quad j = 1, \ldots, n \tag{5.18}$$

$$t_j = j + \delta_j$$

Hence we have the following pdf for our x_j data

$$f_j(x_j, \theta) = g(q_j y_j - q_j \alpha \cos(\omega t_j) - q_j \beta \sin(\omega t_j)) h(t_j - j) \tag{5.19}$$

Theorem 5.4. *For the weighted jittered model described in (5.18) it can be shown that under the same regularity conditions (1)-(2) and the same parameter restrictions, $\hat{\theta}$ is consistent and asymptotically normal with mean θ and covariance matrix*

$$\mathrm{Cov}(\mathbf{N}^{1/2}(\hat{\theta} - \theta)) = k^{-1} 2 I^{-1} \mathbf{\Gamma}^{-1}$$

where, $k = \sum_{j=1}^{n} g_j^2$

The proof can be found in Appendix A.4. Again, for the frequency estimate, when the measurement errors are normally distributed, we have:

$$n^{3/2} \hat{\omega} \to N(n^{3/2} \omega, \frac{24\sigma^2}{\alpha^2 + \beta^2} k^{-1})$$

This is the same variance as for the jittered model (5.2) multiplied by the sum of squared weights. Larger weights are an indication of data that we trust more and a smaller variance, or the other way around.

5.4 Simulations

In this Section we provide numerical evidence that the main asymptotic results stated in (5.14) and (5.11) for the red noise models indeed hold. For that purpose we perform an extensive simulation study for both models and report our findings in the following Sections.

5.4.1 Light curves with AR(1) residuals

We begin with the result for the jittered sinusoidal model with AR(1) errors. We generate light curves from model (5.12) for different combinations of the frequency ω and the residuals' correlation ρ. Specifically, we choose for ω the values 2.4π and 1.1π, and for ρ the values 0.1, 0.5, 0.9. In addition to that, we allow the sample size n to vary too and we show the results for light curves of length 50, 100 and 200 data points. For each combination of the above mentioned parameters we generate 100 light curves. We keep parameters α, β and σ^2 fixed and equal to: $\alpha^2 + \beta^2 = 1$ and $\sigma^2 = 0.04$. We do this for simplicity but also for comparison reasons, as these are the same values reported in Reimann's (Reimann (1994) [53]) simulation study for the simple jittered sinusoidal model (5.2).

The estimation of ω as described in the previous Chapters is non trivial. For this reason we discretize the parameter space of ω in two steps (see also Section 3.8.2). We let ω range from 0 to 50, and for the first estimation step we use a 0.001 accuracy grid. We then select the 5 frequency values for which the likelihood obtained its 5 highest points and discretized again around these 5 frequencies using a grid of $\frac{1}{20000n}$ accuracy. This is the same grid used in Reimann (1994) [53] and is selected here for comparison reasons.

Our results are summarized in Table 5.1 where we compare the theoretical variance according to (5.14) to the estimated one from the simulations. The theoretical variances are shown in bold and the first entry is obtained as:

$$\frac{24\sigma^2}{a^2+b^2}(1+\rho^2-2\rho\cos(\omega)) = \frac{24\cdot 0.04}{1}(1+0.9^2-2\cdot 0.9\cdot\cos(2.4\pi)) = 0.77$$

We notice that the estimated variances are close to the ones predicted from theory and the values get closer as the sample size grows. Furthermore, we see that the estimated biases are quite small in all cases and can be found under the columns named "Bias/se".

Finally, as an extra layer of proof we perform the following hypothesis testing for the sample variances

$$H_0 : \tau^2 = \tau_0^2$$

$$H_1 : \tau^2 \neq \tau_0^2$$

where τ^2 is the unknown true variance of our variance sample and τ_0^2 is the theoretical value being tested. Under the null hypothesis, the sample variance follows a $\tau^2\chi^2_{(99)}$ distribution since our sample size is 100 as described previously. In the square brackets in Table 5.1 under the columns "Theoretical var", we report the 95% confidence intervals corresponding to the hypothesis test described above. Note that the confidence interval reported in not affected by n (the length of the series), it is affected by the number of simulations performed in order to calculate the estimated variance, which here is 100 for all parameter combinations. We notice that the estimated and asymptotic variances are close to each other for all parameter choices, and that all values fall within the intervals, thus, equation (5.14) and the asymptotic rate of $n^{-3/2}$ are confirmed.

For comparison reasons, we perform the same analysis by simulating light curves with AR(1) noise but instead we fit the white noise model. The results can be found in Table 5.2. For the simple jittered sinusoidal model the theoretical variance is equal to 0.96 for the same choice of parameters and the corresponding 95% confidence interval is [0.71, 1.25]. We see that the estimated variances are quite far away from 0.96 that the white noise model predicts. We notice that the effect

TABLE 5.1: Variances from artificial sampling. Data with AR(1) noise

			$\omega = 2.4\pi$			$\omega = 1.1\pi$	
ρ	**n**	**Theoretic Var**	**Estimated Var**	**Bias/se**	**Theoretic Var**	**Estimated Var**	**Bias/se**
0.9	50	**0.77** [0.57,1]	0.58	-0.15	**0.27** [0.20,0.35]	0.14	-0.01
	100		0.87	0.03		0.17	0.05
	200		0.73	-0.16		0.23	0.07
0.5	50	**1.02** [0.76,1.33]	1.09	0.03	**0.44** [0.33,0.57]	0.39	0.03
	100		0.90	0.09		0.41	-0.06
	200		1.02	0.10		0.42	0.10
0.1	50	**1.01** [0.75,1.31]	0.93	-0.02	**0.80** [0.59,1.04]	0.73	-0.06
	100		0.87	0.03		0.92	-0.01
	200		1.15	-0.01		0.99	-0.13

is more dramatic when the residuals' correlation is large, but the values get closer to 0.96 as ρ gets smaller and thus the data resemble more white noise.

Finally, we perform the same analysis as in Table 5.1 but this time we borrow the sampling of three real light curves from Froebrich et al. (2021) [24]. The light curves selected are objects 9324 R, 2287 R, 7159 I with 65, 56, and 107 data points respectively. We notice that for the first two light curves the simulation results are reasonable and close to those obtained for the light curves with artificial sampling. However, for the third light curve we get variances quite larger than expected. This happens because the sampling of the first two light curves resembles that of the jittered model, for the last light curve however this model is insufficient to describe the underlying sampling. See for example Figure 5.3, where the spacings (i.e differences between two consecutive time points) from the jittered model are compared to those of light curve 7159 I, used in the third example. We see that the histogram of the real light curve deviates a lot from what we would expect to see from data generated according to the jittered model.

5.4.2 Light curves with unequally-spaced red noise

We perform again the same simulation analysis but this time in order to verify the result from Theorem 5.1. The way simulations were performed remains exactly the same as described in the previous Section. For the first part of the analysis we simulate light curves with artificial sampling according to the jittered model and frequency fixed at 2.4π. In this example we vary the residuals' correlation ρ,

TABLE 5.2: Variances from artificial sampling. Data with AR(1) noise fitted with white noise model.

		$\omega = 2.4\pi$		$\omega = 1.1\pi$	
ρ	**n**	**Estimated Var**	**Bias/se**	**Estimated Var**	**Bias/se**
0.9	50	0.53	-0.16	0.20	0.06
	100	0.49	-0.03	0.12	-0.19
	200	0.38	0.23	0.21	0.09
0.5	50	0.75	-0.06	0.39	-0.03
	100	0.68	0.04	0.30	-0.14
	200	1.21	0.10	0.56	-0.05
0.1	50	0.92	-0.03	0.72	-0.06
	100	0.87	0.03	0.68	-0.10
	200	1.15	0.08	1.10	-0.14

TABLE 5.3: Variances with sampling from real light curves, with AR(1) noise.

		$\omega = 2.4\pi$			$\omega = 1.1\pi$		
ρ	**lc**	**Theoretic Var**	**Estimated Var**	**Bias/se**	**Theoretic Var**	**Estimated Var**	**Bias/se**
0.9	lc1	**0.77** [0.57,1]	0.54	0.10	**0.27** [0.20,0.35]	0.40	0.07
	lc2		0.66	0.06		0.29	0.09
	lc3		1.93	0.10		0.57	-0.09
0.5	lc1	**1.02** [0.76,1.33]	0.88	-0.03	**0.44** [0.33,0.57]	0.62	0.16
	lc2		0.59	0.19		0.34	-0.13
	lc3		2.69	-0.02		1.45	-0.13
0.1	lc1	**1.01** [0.75,1.31]	0.80	0.06	**0.80** [0.59,1.04]	0.81	-0.06
	lc2		0.58	-0.14		0.56	0.01
	lc3		1.84	-0.08		1.62	0.04

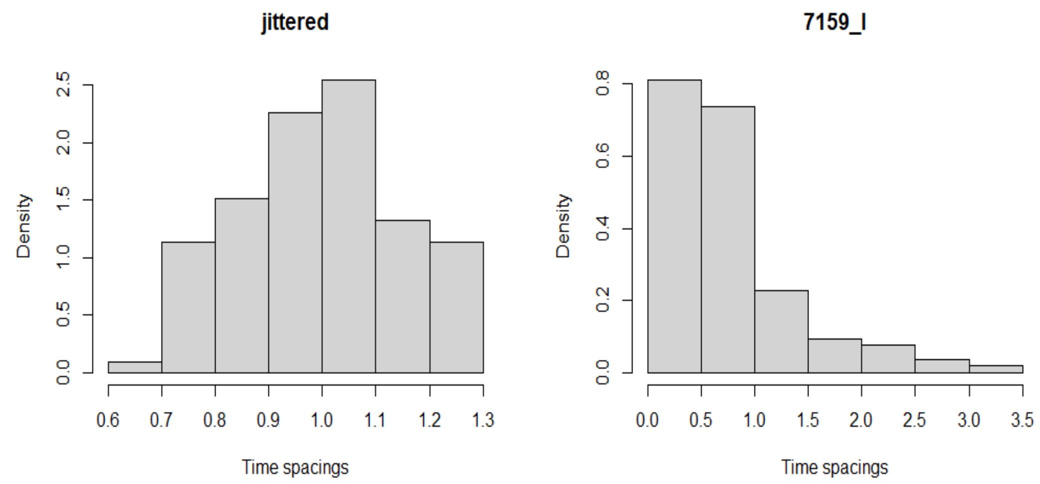


FIGURE 5.3: (Left) Histogram of time spacings generated from the jittered model for d=1/6. (Right) Histogram of the spacings of the time points of light curve 7159_I.

Table 5.4: Variances from artificial sampling. Model with unequally spaced red noise and $\omega = 2.4\pi$.

		d=1/6			d=1/3		
ρ	n	**Theoretic Var**	**Estimated Var**	**Bias/se**	**Theoretic Var**	**Estimated Var**	**Bias/se**
0.9	50	**0.75** [0.56, 0.98]	0.43	-0.13	**0.63** [0.47, 0.82]	1.17	-0.03
	100		0.50	-0.02		1.2	0.05
	200		0.63	0.09		0.99	0.13
0.5	50	**1** [0.74, 1.30]	0.79	-0.05	**0.88** [0.65, 1.14]	1.03	-0.11
	100		0.82	0.01		0.96	0.04
	200		1.11	0.05		0.97	0.10
0.1	50	**1.01** [0.75, 1.31]	0.98	-0.03	**0.99** [0.73, 1.29]	0.85	0.04
	100		0.92	0.02		0.88	0.04
	200		1.21	0.06		1	0.07

the number of data points n and we compare two values of the parameter d which controls the level of disturbances of our time points. The results are summarized in Table 5.4. The first half of the table shows the results for $d = 1/6$ which is a standard value for this parameter as described previously, and the other half shows the results for $d = 1/3$. In the latter case our sampling has larger disturbances and thus is further away of the equally-spaced scheme compared to the sampling generated with $d = 1/6$.

We notice that for $\rho = 0.5$ and $\rho = 0.1$ the estimated variances fall within the confidence intervals in all cases. When the correlation gets larger however (0.9 in our case) we need a larger number of data in order for the estimated variance to fall within the confidence interval. In the case of $d = 1/6$ this is achieved for 200 data points when the disturbances get larger however these sample sizes are still small, as can be seen from the top right part of the table.

Table 5.5 shows exactly the same analysis as seen in Table 5.2, only the data are generated according to the red noise jittered sinusoidal model (5.4). Similarly as before we see that the estimated variances for $\rho > 0.5$ are far away from the 0.96 which is predicted from the white noise model, and thus a model that takes this behaviour into account is needed.

Finally in Table 5.6 we perform the same analysis as described in the first paragraph, only we borrow the sampling from the three real light curves described in the previous Section. Here we choose $d = 1/6$ as seen in Reimann (1994) [53] for similar data, which reflects the idea that observations occur at night time.

TABLE 5.5: Variances from artificial sampling. Data with red noise fitted with white noise model.

		$d = 1/6$		$d = 1/3$	
ρ	**n**	**Estimated Var**	**Bias/se**	**Estimated Var**	**Bias/se**
0.9	50	0.42	-0.13	0.92	0.08
	100	0.56	-0.05	0.93	-0.08
	200	0.63	0.09	0.92	0.08
0.5	50	0.79	-0.05	0.95	-0.07
	100	0.81	0.01	0.87	-0.06
	200	1.13	0.08	0.92	0.05
0.1	50	0.97	-0.03	1	-0.04
	100	0.92	0.02	0.83	-0.06
	200	1.28	0.10	0.89	0.06

TABLE 5.6: Variances with sampling from real light curves, with unequally-spaced red noise.

		$\omega = 2.4\pi$			$\omega = 1.1\pi$		
ρ	**lc**	**Theoretic Var**	**Estimated Var**	**Bias/se**	**Theoretic Var**	**Estimated Var**	**Bias/se**
0.9	lc1		0.35	-0.03		0.37	0.10
	lc2	**0.75**	0.36	-0.15	**0.63**	0.29	0.07
	lc3	[0.56,0.98]	1.56	0.06	[0.47,0.82]	0.67	-0.01
0.5	lc1		0.87	-0.20		0.62	-0.09
	lc2	**1**	0.47	0.14	**0.88**	0.39	0.01
	lc3	[0.74,1.30]	1.9	-0.21	[0.65,1.14]	1.66	-0.07
0.1	lc1		0.71	0.01		0.62	0.11
	lc2	**1.01**	0.61	-0.20	**0.99**	0.47	-0.14
	lc3	[0.75,1.31]	2.74	-0.05	[0.73,1.29]	2.61	0.08

5.5 Adjusting the generalized F-test

In Chapter 3 we discussed the importance of applying a statistical test in order to decide if our frequency (or period) estimate represents a real signal. The same holds for the models we considered in this Chapter. Thus, in order for the methodology to be complete we will show how the generalized F-test seen in Section 3.2 can be applied here too when we use the jittered sinusoidal model with unequally spaced red noise (5.4).

The format of the test will be the same as in Equation (3.2). Matrices $\mathbf{M}_0$ and $\mathbf{M}_1$ representing the model with no signal and the periodic model are calculated as follows:

$$\mathbf{M}_0 = \mathbf{C}_\rho^{1/2}(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\mathbf{C}_\rho^{1/2}$$

$$\mathbf{M}_1 = \mathbf{C}_\rho^{1/2}(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{C}_\rho^{1/2}$$

Where $\mathbf{X}$ is the design matrix of the sinusoidal model (2.11) and $\mathbf{C}_\rho$ is shown in (3.11). Similarly to the previous Chapters we use Saddlepoint approximation to evaluate the p-values (3.3).

In order to assess the performance of our test under this setting, we perform an extensive power analysis and compare the test to an alternative one, namely RTheiler (Thieler et al. (2013) [63]). For all the examples we will discuss below, we estimate the power by simulating 100 light curves for each scenario (e.g each combination of SNR and d) and the significance level used is 1-0.99.

First, we estimate the power of the test for different signal to noise ratios and we also explore how different time disturbances may affect the test's performance. In Figure 5.4 (Left) we see these comparisons. Different types of lines correspond to time observations jittered by different Uniform distributions. We notice that the generalized F-test performs well for reasonable SNRs (e.g. larger than 0.5) and its power is constantly higher compared to the RThieler test. Furthermore, we notice that for both tests the power is lower for time observations disturbed by a $U(-1/2, 1/2)$ distribution.

Finally, we explore how different levels of correlation in our residuals can affect the test's power, see Figure 5.4 (Right). We notice that both tests have their highest power for ρ close to 0.5 and the power drops quickly as the dependence in the residuals gets larger.

5.6 Application to real light curves

In this Section we will apply the methods discussed in this Chapter to two real light curves we analysed in Chapter 3.

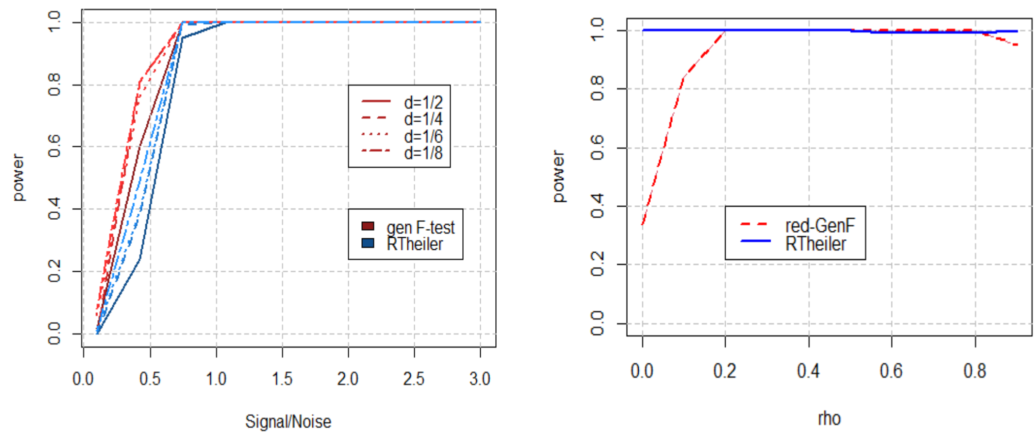


FIGURE 5.4: (Left) The power of the generalized F-test for the unequally spaced red noise model for different signal to noise ratios. For this example $\rho = 0.5$ and different types of lines correspond to different time disturbances. (Right) In this example we see the power of the generalized F-test for different values of ρ. The signal to noise ratio is fixed at 0.5 and the disturbances in time come from $U(-1/6, 1/6)$.

5.6.1 Example 1

This is light curve number 6149 from Froebrich et al. (2021) [24], with a studied period at 2.1763 days. More details can be found in Section 3.10 and the data can be seen plotted in Figure 3.10.

Here we compare the three main models discussed in this Chapter, namely the sinusoidal model with iid (5.2), AR(1) (5.12) and unequally spaced red noise (5.4) respectively. In Table 5.7 we report the estimated period for each method for the four available filters. We notice that all methods agree to one decimal place and are close to 2.1763. Note that these results also agree with those from Section 3.10. We also report the 95% confidence intervals for each estimate. We notice that the intervals are similar, with those for AR(1) being slightly narrower.

After examining the autocorrelation of the residuals after fitting a white noise model we noticed that in some cases there seemed to be some significant correlations between the residuals. The estimated correlation parameter $\hat{\rho}$ after fitting the AR(1) model was between 0.16 and 0.21 in all cases, see also Figure 5.5 (Left) which depicts the acf plot for the visual filter after fitting the white noise model.

TABLE 5.7: The period identified as most important from the three different models discussed in this Chapter assuming iid, AR(1) and unequally spaced AR(1) residuals. In parenthesis we show the 95% confidence interval for each estimation.

Filter	White	AR(1)	Red
B	**2.1895** (2.1781, 2.2011)	**2.1891** (2.1797, 2.1986)	**2.1891** (2.1773, 2.2011)
R	**2.1795** (2.1789, 2.1801)	**2.1799** (2.1795, 2.1803)	**2.1799** (2.1793, 2.1805)
I	**2.1705** (2.1685, 2.1717)	**2.1698** (2.1688, 2.1706)	**2.1698** (2.1685, 2.1711)
V	**2.1695** (2.1685, 2.1705)	**2.1698** (2.1689, 2.1707)	**2.1698** (2.1688, 2.1708)

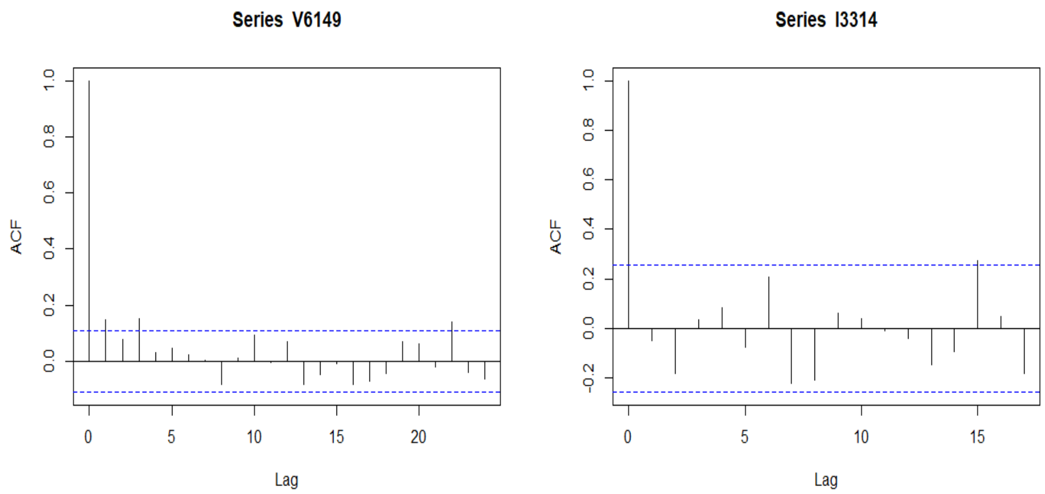


FIGURE 5.5: (Left) The acf plot of the residuals of object 6149 V after fitting the white noise sinusoidal model. (Right) The same but for object 3314 I.

Notice that some small level of autocorrelation seems to exist between the residuals, however $\hat{\rho}$ is estimated very close to zero and the autocorrelation function exhibits a similar behaviour after the red noise model has been fitted. This might be an indication that other dynamics, not explained by the red noise model, exist in this data. When fitting the red noise model however $\hat{\rho}$ was estimated as very close to zero in most cases. Finally, we performed the period detection test described in Section 5.5 and more analytically in Chapter 3, for all three models. For the iid case the test rejected the null hypothesis (of no periodicity) indicating a significant period at an 1- 99% level. The test however did not reject the null hypothesis for the other two models assuming correlated residuals.

5.6.2 Example 2

The second light curve we examine is that corresponding to star number 3314 with a period at 13.8783 days as seen in Froebrich et al. (2021) [24]. More details on this data set, along with its plot can be found in Section 3.10. We perform the same analysis as described for the previous example in Section 5.6.1. Again all models appear to be in close agreement. The results can be seen in Table 5.8.

TABLE 5.8: The estimated period according to the models with iid, AR(1) and unequally spaced AR(1) residuals along with their 95% confidence intervals.

Filter	**White**	**AR(1)**	**Red**
R	**13.8705** (13.3767, 14.4022)	**13.8701** (13.3415, 14.4423)	**13.8701** (13.3759, 14.4022)
I	**13.8605** (13.5023, 14.2383)	**13.8701** (13.5124, 14.2471)	**13.8701** (13.5073, 14.2529)

After applying the period detection test we report that the null hypothesis was rejected for all models at the same 99% significance level. We examined the autocorrelation of the residuals after fitting the white noise model in this case too, with no significant autocorrelations identified, see Figure 5.5 (Right) as an example for filter I. The correlation parameter here was < 0.00001 in all cases for both the AR(1) and the red noise models.

5.7 Summary

The asymptotic properties of the maximum likelihood frequency estimates of sinusoidal models for irregularly sampled data are studied in this Chapter. A result from Reimann (1994) [53] for iid data is extended for the case where red noise is present. We showed that under this setting and for the noise structure known, $\hat{\omega}$ is again consistent and asymptotically normal. The asymptotic variance in this case also depends on the frequency, error correlation and sampling. The convergence rate is the same as for the iid case. Our simulations returned consistent results to those predicted by the proposed theorems for sample sizes of 200 data points or more.

Furthermore, we saw that a special case of the red noise model is the AR(1) error structure. The same results hold under this case too but the asymptotic variance no longer depends on the sampling. The simulation results in this case were in agreement with the proposed theoretic values even for small sample sizes (i.e 50). Both results are unaffected by the use of the intercept term in the sinusoidal model. The application to real data returned results similar to the previous methods, accompanied with a confidence interval for each estimate. For these two examples the results were almost the same between white and red noise settings indicating a non-correlated error structure for these examples.

Chapter 6

Conclusions and Future work

In this thesis we studied the problem of analysing the periodicity of light curves through three different perspectives. First we discussed the problem from the period estimation point of view, secondly we examined thoroughly adding some credibility to those estimates and finally we explored the statistical properties of some of them.

In particular, in Chapter 3 we provided tests for period detection under non-parametric linear smoother settings, especially Gaussian process regression. The generalized F-statistic is easily and readily adjustable for a range of models. For example it works better than alternative tests when Weighted Least Squares regression is used and the sample size is relatively large. It is also easy to use under ARMA model setting providing more accurate results than asymptotic methods for small sample sizes.

The CVF-test statistic based on leave-one-out cross validation for Gaussian process regression behaves similarly to the F. Both statistics can be adjusted for Weighted Gaussian process regression scenarios (when measurement accuracies are available), and also for deviations from white background noise. Our simulation results show that adjusting for red noise (when present) leads to more periods correctly identified and smaller number of falsely identified periods. The power of both tests when Gaussian process regression is used is quite high for reasonable signal to noise ratios (e.g. for data with $SNR > 0.6$). The proposed methods are also applied to real light curves returning similar results to standard approaches.

In Chapter 4, we provided a flexible alternative to the GPR methodology, using a t-process. This approach according to our simulations works as well as the GPR method only it captures better data with outliers and fat tails. Moreover we showed how the measurement accuracies can be taken into account under this setting too.

In addition, we extended the GPR and tPR models for light curves with measurements in multiple filters and showed how the measurement accuracies can be taken into account again for these multiple filter approaches. Our simulations showed that the multiband models work well and in some cases provide better inference than by analysing the light curves individually. This is especially the case if the light curves are not well measured in all filers. If we have perfect measurements in all filters and the shape of the light curves is quite different between the filters then analysing them individually might provide more accurate results.

Furthermore, we showed how the generalized F and CVF-tests can be use under these settings too in order to build periodograms and perform the corresponding tests. Finally, the possibility of modelling quasi periodic data was examined by using different periodic kernels that account for different periodical assumptions and we showed that the generalized F and CVF-tests can be used as kernel selection tools. All these methods were applied to real data returning in most cases similar results to the previously discussed methods.

In Chapter 5, we examined the statistical properties of the period estimates of irregularly sampled time series. As explained earlier not a lot of work has been done in this domain of statistics. We discussed a result from Reimann et al. (1994) [53] for the MLE estimate of period based on the simple sinusoidal model (5.2) and extended it for data with red noise.

Specifically, we showed that under some regularity conditions the MLE of the frequency parameter of the red noise jittred sinusoidal model (5.4) is consistent and asymptotically normal and converges to the true parameter at an $O(n^{-3/2})$ rate. This result was verified through an extensive simulation study. When the residuals' correlation is less than 0.5 the result is verified even for small sample sizes of length of 50 points for example. When this correlation gets closer to 1 however then a sample size of 500 or more is required. Finally we showed that a similar result holds when we include the intercept in model (5.2) and also for their weighted regression equivalents.

Methods Comparison

In this thesis we explored a big range of methods around the subject of finding periodical patterns in time series. Each method has its strengths and weaknesses and from our study no method was identified as overall better than the rest. here we will try ti look at these methods collectively and draw comparisons between them.

At the end of each Chapter we applied the methods discussed/proposed to two randomly selected real light curves from Froebrich et al. (2021) [24]. For comparison reasons we chose the same two light curves for all Chapters applications. Object no 6149 has a studied period at 2.1763 days as explained before. Overall for this light curve almost all methods found as most significant the period at 2.2 days. Standard F-test had a larger amount of false positives compared to generalized F, CVF and Theiler's test. The t-process regression method returned exactly the same results as the Gaussian process one. We also applied two further periodic kernels the cylinder and Quasi periodic one and they again returned similar results. The multiple filter version of of the tPR method identified the correct period but had many false positives. The multiband GPR method failed to identify the correct period. Note however that for the latter method we did not include the measurement accuracies which were taken into account in all other methods. This is an indication that accounting for these weights can improve our results. Finally red noise models both under the sinusoidal and GPR settings returned

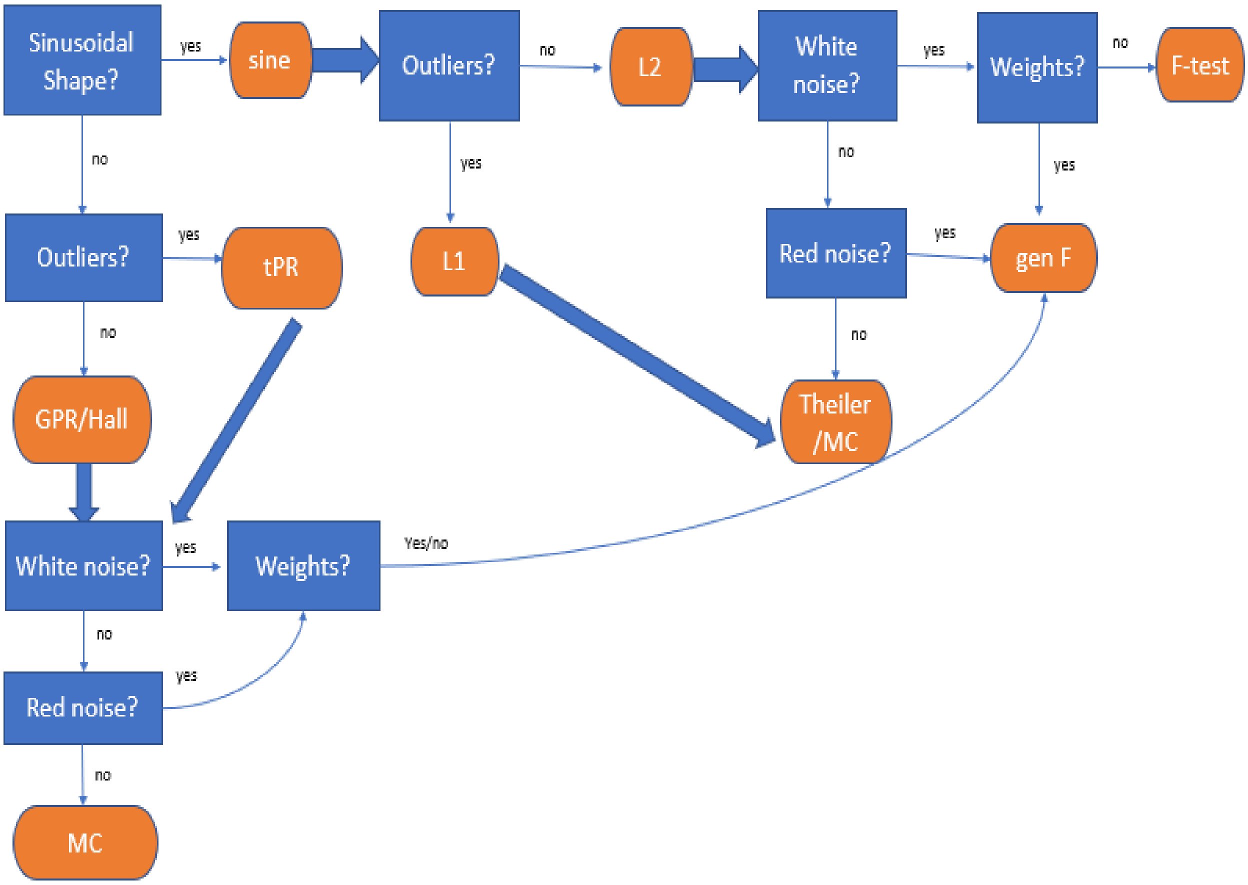


FIGURE 6.1: Flow chart of the methods discussed in this thesis. The starting point is top left. With L2 we denote the sum of least squares. With L1 we denote the sum of absolute deviations but other robust regression techniques can be used instead. With Hall we denote the approach by Hall et al. (2000) [27] based on kernel smoothing regression. Finally MC stands for Monte Carlo.

similar results, which is expected since the data as we discussed earlier did not seem to have correlated residuals. As we can see from Figure 3.10 the shape of this light curve resembles a sinusoidal wave and that's why both non-parametric and non methods behaved similarly.

The second object studied was light curve no 3314 with period at 13.8783 days. In this case almost all methods returned identical results, they identified as most important the period at 13.9 days and zero false positives. The only approach that gave a different period was when the cylinder kernel was used. The multiple filter approaches identified the correct period this time with the multiband tPR having 1 and the multiple GPR 2 false positives. Similarly as before looking at Figure 3.12 we see that the light curve resembles the shape of a sine wave.

In order to get a clearer idea as to when it is better to use each method we built the flow chart seen in Figure 6.1. The blue boxes represent the questions we should answer regarding our data. For example what is their shape or their background noise. The orange boxes show us the recommended method to use in each case. Note that in this chart we did not include the quasi periodic and the multiple filter measurements cases for simplicity, since the idea and thought steps to follow are the same. Note also that at the start of the flow chart we ask the question whether our data is sinusoidal or not because that is the most common shape for the data we study. Generally this question should be posed as whether the shape of the curve is known or not. When the shape is known it is advised to model the data parametrically for more accurate and computationally faster results.

Future Work

In Chapter 3 we introduced a flexible hypothesis testing setting for identifying valid period. These tests are flexible but there are situations where they cannot be used (e.g Robust regression/L1 settings). Furthermore, we showed how these tests can be used under red noise settings, but further work is needed in order to include other potentially useful background noise models. Finally there might be cases where the background noise is not known at all and thus other methods are needed in order to obtain accurate results.

In Chapter 4 we extended the GPR framework in order to include measurements from multiple filters. Although these multiband methods seem to work well they are computationally demanding as the number of filters grow. For that reason more work is needed for these models to be computationally efficient. Moreover we saw how choosing different kernels can help us model light curves that are not strictly periodic and we can choose between them by applying the generalized F-test. This setting however is restrictive regrading the types of data we can model and requires multiple comparisons between kernels in order to select the best. It would be useful thus to have a more general model that incorporates all cases. Another thing to consider is determining the p-values of quadratic forms of t components in order to obtain more accurate results when using the generalized F-test under the t-process setting. Estimating the CDF of Student t Quadratic forms is an interesting and not adequately explored problem with potential applications to time series in general. Some relevant work to that end can be found in Bishop et al. (1978) [11] where the authors estimate percentages of Student t Quadratic forms using simulations or the Pearson curve fitting method.

Finally, in Chapter 5 we discussed the asymptotic properties of frequency estimates obtained under sinusoidal-type models. We did not explore however the asymptotic properties of the non-parametric estimates. To our knowledge, the only relevant work to that end can be found in Hall et al. (2000) [27] where the authors show that period estimates obtained through a kernel smoothing regression scheme are consistent and asymptotically normal. It would be interesting thus to show that a similar result holds and the GPR and tPR setting too. Moreover for our results we assumed a very specific model for the time points, given by (5.2). Other models have been proposed too (see Hall et al. (2000) [27]) and since as we discussed previously this model might not always be representative of our data it would be useful to explore if the same results hold for different time models. Finally, another thing to consider is the situation where the distribution of the time points is not known.

Appendix A

Proofs for asymptotic theorems

A.1 Conditions (A)-(D) for Theorem 5.1

To make use of Theorem III4.1 of Ibragimov and Has'minskii (2013) [33] (p.191), Ψ_n^2 must also satisfy the following four conditions. These are the same conditions stated in Reimann (1994) [53] (p.42) or equations (4.2), (4.4), (4.8) and (4.9) found at pages (191-194) in [33].

(A): For defined $|A| = \sup_{|\lambda|=1} |\lambda^T A \lambda|$,

$$\lim_{n\to\infty} \sup tr \Psi_n^{-2} = 0, \quad \lim_{n\to\infty} |\Psi_n^{-1}\Psi^2(n,\tau)\Psi_n^{-1}| < \infty$$

where τ is an estimate of θ and $\Psi_n = \Psi_{(n,\theta)}$

(B): For $\rho > 0$ and $v \in \Re^3$,

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \sum_{j=1}^{n} E|[\frac{\partial}{\partial\theta} \ln f_j(x_j;\theta)]^T \Psi_n^{-1} v|^{2+\rho} = 0$$

(C): Assume sequence $\lambda(n) \to \infty$ as $n \to \infty$.

$f_j^{1/2}(x_{j,\theta})$ is twice continuously differentiable w.r.t θ, and as $n \to \infty$

$$\sup_{\theta\in\Theta} \sup_{|v|<\lambda(n),\theta+\Psi_n^{-1}v} \sum_{j=1}^{n} \int |\Psi_n^{-1} \frac{\partial^2}{\partial\theta^2} f_i^{1/2}(x_j, \theta + \Psi_n^{-1}v)\Psi_n^{-1}|^2 dx_j = o(\lambda(n)^{-2})$$

(D): For $\zeta > 0$

$$\inf_{\theta\in\Theta} \inf_{|v|<\lambda(n),\theta+\Psi_n^{-1}v} |\Psi_n|^{-\zeta} \sum_{j=1}^{n} \int [f_j^{1/2}(x_j;\theta+\Psi_n^{-1}v) - f_j^{1/2}(x_j;\theta)]^2 dx_j > 0$$

The steps followed to prove these conditions for our model are exactly the same as shown in [53] at pages 44-46.

A.2 Proof for Corollary 5.2

We follow exactly the same steps as seen in the proof for Theorem 5.1. Matrix C in this case will be calculated as:

$$\frac{\partial}{\partial a} f_j(x_j,\theta) = g'(u)(-\cos(\omega t_j) + \rho\cos(\omega t_{j-1}))h(t_j - j)$$

$$\frac{\partial}{\partial b} f_j(x_j,\theta) = g'(u)(-\sin(\omega t_j) + \rho\sin(\omega t_{j-1}))h(t_j - j)$$

$$\frac{\partial}{\partial \omega} f_j(x_j,\theta) = g'(u)(a\sin(\omega t_j)t_j - b\cos(\omega t_j)t_j - a\rho\sin(\omega t_{j-1})t_{j-1} + b\rho\cos(\omega t_{j-1})t_{j-1})h(t_j - j)$$

Thus $\mathbf{C}$ is seen bellow:

$$\mathbf{C} = g'(u)^2 h(t_j - j)^2 \begin{bmatrix} A^2 & AB & AC \\ * & B^2 & BC \\ * & * & C^2 \end{bmatrix}$$

$$A^2 = \rho^2\cos^2(\omega t_{j-1}) + \cos^2(\omega t_j) - 2\rho\cos(\omega t_j)\cos(\omega t_{j-1})$$

$$AB = \cos(\omega t_j)\sin(\omega t_j) - \rho\cos(\omega t_j)\sin(\omega t_{j-1}) - \rho\cos(\omega t_{j-1})\sin(\omega t_j) + \rho^2\cos(\omega t_{j-1})\sin(\omega t_{j-1})$$

$$B^2 = \rho^2\sin^2(\omega t_{j-1}) + \sin^2(\omega t_j) - 2\rho\sin(\omega t_j)\sin(\omega t_{j-1})$$

$$AC = -at_j\cos(\omega t_j)\sin(\omega t_j + bt_j\cos^2(\omega t_j) + a\rho t_{j-1}\cos(\omega t_j)\sin(\omega t_{j-1}) -$$

$$-b\rho t_{j-1}\cos(\omega t_j)\cos(\omega t_{j-1})+a\rho t_j\cos(\omega t_{j-1})\sin(\omega t_j)-b\rho t_j\cos(\omega t_{j-1})\cos(\omega t_j)-$$

$$-a\rho^2 t_{j-1}\cos(\omega t_{j-1})\sin(\omega t_{j-1})+b\rho^2 t_{j-1}\cos^2(\omega t_{j-1}))$$

$$BC=-at_j\sin^2(\omega t_j)+bt_j\sin(\omega t_j)\cos(\omega t_j)+a\rho t_{j-1}\sin(\omega t_j)\sin(\omega t_{j-1})-$$

$$-b\rho t_{j-1}\sin(\omega t_j)\cos(\omega t_{j-1})+a\rho t_j\sin(\omega t_j)\sin(\omega t_{j-1})-b\rho t_j\sin(\omega t_{j-1})\cos(\omega t_j)-$$

$$-a\rho^2 t_{j-1}\sin^2(\omega t_{j-1})+b\rho^2 t_{j-1}\sin(\omega t_{j-1})\cos(\omega t_{j-1})$$

$$C^2=a^2t_j^2\sin^2(\omega t_j)-abt_j^2\sin(\omega t_j)\cos(\omega t_j)-a^2\rho t_j t_{j-1}\sin(\omega t_j)\sin(\omega t_{j-1})+$$

$$+ab\rho t_{j-1}\sin(\omega t_j)\cos(\omega t_{j-1})+b^2t_j^2\cos^2(\omega t_j)+ab\rho t_{j-1}\cos(\omega t_j)\sin(\omega t_{j-1})-$$

$$-b^2\rho t_{j-1}\cos(\omega t_j)\cos(\omega t_{j-1})+a^2\rho^2 t_{j-1}^2\sin^2(\omega t_{j-1})-$$

$$-ab\rho^2 t_{j-1}^2\sin(\omega t_{j-1})\cos(\omega t_{j-1})+b^2\rho^2 t_{j-1}^2\cos^2(\omega t_{j-1})$$

Finally, matrix Ψ^2 will be of the following form:

$$\Psi_n^2=k\frac{I}{2}\mathbf{N}^{1/2}\left\{\mathbf{\Gamma}+O(n^{-1})\right\}\mathbf{N}^{1/2}$$

where $k=1+\rho^2-2\rho\cos(\omega)$. The first entry $[1,1]$ of Ψ_n^2 is obtained similarly by,

$$\Psi^2_{[1,1]}=I\sum_{j=1}^{n}\left\{E(\rho^2\cos^2(\omega t_{j-1}))+E(\cos^2(\omega t_j))+E(-2\rho\cos(\omega t_j)\cos(\omega t_{j-1}))\right\}=$$

$$=\frac{I}{2}\sum_{j=1}^{n}\{1+\cos(2\omega j)E(\cos(2\omega\delta))+\rho^2+\rho^2\cos(2\omega(j-1))E(\cos(2\omega\delta))-$$

$$-2\rho\cos(\omega)E(\cos(\omega(\delta_1-\delta_2)))-2\rho\cos(2\omega(j-1))E(\cos(\omega(\delta_1+\delta_2)))\}=$$

$$=\frac{I}{2}n(1+\rho^2-2\rho\cos(\omega))+O(1)$$

The form of matrices Ψ_n^{-2}, Ψ_n and Ψ_n^{-1} can be seen below:

$$\Psi_n^{-2}=k^{-1}2I^{-1}\mathbf{N}^{-1/2}\left\{\mathbf{\Gamma}^{-1}+O(n^{-1})\right\}\mathbf{N}^{-1/2}$$

$$\Psi_n = \frac{\sqrt{I}}{\sqrt{2}} k^{1/2} \mathbf{N}^{1/4} \{\mathbf{\Gamma}^{1/2} + O(n^{-1})\} \mathbf{N}^{1/4}$$

$$\Psi_n^{-1} = \frac{\sqrt{2}}{\sqrt{I}} k^{-1/2} \mathbf{N}^{-1/4} \{\mathbf{\Gamma}^{-1/2} + O(n^{-1})\} \mathbf{N}^{-1/4}$$

Conditions (A)-(D) are verified as seen in Reimann (1994) [53].

A.3 Proof for Theorem 5.3

We will follow the same steps described in the previous two sections, and we will start by calculating matrix Ψ_n^2 which in this case will be of a 4×4 dimension. Fisher's information matrix for this case can be seen bellow:

$$I_j(\theta) = IE \begin{bmatrix} 1 & \cos(\omega t_j) & \sin(\omega t_j) & t_j(b\cos(\omega t_j) - a\sin(\omega t_j)) \\ * & \cos^2(\omega t_j) & \cos(\omega t_j)\sin(\omega t_j) & t_j\cos(\omega t_j) b\cos(\omega t_j) - a\sin(\omega t_j) \\ * & * & \sin^2(\omega t_j) & t_j\sin(\omega t_j) b\cos(\omega t_j) - a\sin(\omega t_j) \\ * & * & * & t_j^2 b\cos(\omega t_j) - a\sin(\omega t_j)^2 \end{bmatrix}$$

We are ready now to calculate the entries of Ψ_n^2 matrix as follows:

$$[\Psi_n^2]_{11} = I\sum_{j=1}^{n} E[1] = In = \frac{n}{2}I2$$

$$[\Psi_n^2]_{12} = I\sum_{j=1}^{n} E[\cos(\omega t_j)] = I\sum_{j=1}^{n} E[\cos(\omega\delta)\cos(\omega j) - \sin(\omega\delta)\sin(\omega j)] =$$

$$= I\sum_{j=1}^{n}\{Re\phi(\omega)\cos(\omega j) - Im\phi(\omega)\sin(\omega j)\} = O(1)$$

$$[\Psi_n^2]_{14} = I\sum_{j=1}^{n}\{b\cos(\omega j)E[\delta\cos(\omega\delta)] - b\sin(\omega j)E[\delta\sin(\omega\delta)] +$$

$$+ jb\cos(\omega j)E[\cos(\omega\delta)] - jb\sin(\omega j)E[\sin(\omega\delta)] - a\cos(\omega j)E[\delta\sin(\omega\delta)] - a\sin(\omega j)E[\delta\cos(\omega\delta)] -$$

$$- ja\cos(\omega j)E[\sin(\omega\delta)] - ja\sin(\omega j)E[\cos(\omega\delta)]\} = O(n)?$$

$$\Psi_n^2 = \frac{I}{2}\begin{bmatrix} 2n & O(1) & O(1) & O(n) \\ * & n+O(1) & O(1) & \frac{b}{2}n^2+O(n) \\ * & * & n+O(1) & -\frac{a}{2}n^2+O(n) \\ * & * & * & \frac{a^2+b^2}{3}n^3+O(n^2) \end{bmatrix} = \frac{I}{2}\mathbf{N}^{1/2}\{\mathbf{\Gamma}_c+O(n^{-1})\}\mathbf{N}^{1/2}$$

In the same manner as in previous sections we can obtain the rest of the matrices.

$$\Psi_n^{-2} = 2I^{-1}\mathbf{N}^{-1/2}\left\{\mathbf{\Gamma}_c^{-1}+O(n^{-1})\right\}\mathbf{N}^{-1/2}$$

$$\Psi_n = \frac{\sqrt{I}}{\sqrt{2}}\mathbf{N}^{1/4}\{\mathbf{\Gamma}_c^{1/2}+O(n^{-1})\}\mathbf{N}^{1/4}$$

and,

$$\mathbf{\Gamma}_c^{1/2} = \begin{bmatrix} \sqrt{2} & 0 & 0 & 0 \\ * & \frac{2a^2+b^2}{2(a^2+b^2)} & \frac{ab}{2(a^2+b^2)} & \frac{b\sqrt{3a^2+3b^2}}{2(a^2+b^2)} \\ * & * & \frac{a^2+2b^2}{2(a^2+b^2)} & \frac{-a\sqrt{3a^2+3b^2}}{2(a^2+b^2)} \\ * & * & * & \frac{(a^2+b^2)^{3/2}}{\sqrt{3}(a^2+b^2)} \end{bmatrix}$$

$$\Psi_n^{-1} = \frac{\sqrt{2}}{\sqrt{I}}\mathbf{N}^{-1/4}\{\mathbf{\Gamma}_c^{-1/2}+O(n^{-1})\}\mathbf{N}^{-1/4}$$

Conditions (A)-(D) hold in a similar manner.

A.4 Proof for Theorem 5.4

We will show that a similar result to the previous holds under the weighted regression setting. We will start with the calculations for Fisher's information matrix:

$$\frac{\partial}{\partial a}\left[g(q_jy_j - q_j\alpha\cos(\omega t_j) - q_j\beta\sin(\omega t_j))h(t_j-j)\right] = -g'(\epsilon)q_j\cos(\omega t_j)h(t_j-j)$$

$$\frac{\partial}{\partial b}\left[g(q_jy_j - q_j\alpha\cos(\omega t_j) - q_j\beta\sin(\omega t_j))h(t_j-j)\right] = -g'(\epsilon)q_j\sin(\omega t_j)h(t_j-j)$$

$$\frac{\partial}{\partial \omega}\left[g(q_jy_j - q_j\alpha\cos(\omega t_j) - q_j\beta\sin(\omega t_j))h(t_j-j)\right] = g'(\epsilon)(aq_j\sin(\omega t_j)t_j - bq_j\cos(\omega t_j)t_j)h(t_j-j)$$

If we square the vector above for our three parameters we will obtain matrix C described below:

$$C = \begin{bmatrix} g'(\epsilon)^2 q_j^2 \cos^2(\omega t_j) h^2(t_j - j) & g'(\epsilon)^2 q_j^2 h^2(t_j - j)\cos(\omega t_j)\sin(\omega t_j) & g'(\epsilon)^2 q_j^2 h^2(t_j - j) t_j \cos(\omega t_j) c(t_j) \\ * & g'(\epsilon)^2 q_j^2 \sin^2(\omega t_j) h^2(t_j - j) & g'(\epsilon)^2 q_j^2 h^2(t_j - j) t_j \sin(\omega t_j) c(t_j) \\ * & * & g'(\epsilon)^2 q_j^2 h^2(t_j - j) t_j^2 c^2(t_j) \end{bmatrix} \quad \text{(A.1)}$$

where $c(t_j) = b\cos(\omega t_j) - a\sin(\omega t_j)$.

Hence Fisher's information can be calculated following the next steps.

$$I_j(\theta) = \int C \frac{dx_j}{f_j(x_j; \theta)}$$

Since all elements in matrix C are multiplied with $g'(\epsilon)^2 q_j^2 h^2(t_j - j)$ we can take these values out, and obtain matrix $\tilde{C}$ depending on t_j and we can write,

$$I_j(\theta) = \int g'(\epsilon)^2 q_j^2 h^2(t_j - j)\tilde{C}\frac{dx_j}{g(\epsilon)h(t_j - j)} = q_j^2 \int \frac{g'(\epsilon)^2}{g(\epsilon)}\tilde{C}h(t_j - j)dx_j = q_j^2 \int \frac{g'(\epsilon)^2}{g(\epsilon)} E[\tilde{C}] =$$

$$= q_j^2 IE \begin{bmatrix} \cos^2(\omega t_j) & \cos(\omega t_j)\sin(\omega t_j) & t_j \cos(\omega t_j) b\cos(\omega t_j) - a\sin(\omega t_j) \\ * & \sin^2(\omega t_j) & t_j \sin(\omega t_j) b\cos(\omega t_j) - a\sin(\omega t_j) \\ * & * & t_j^2 b\cos(\omega t_j) - a\sin(\omega t_j)^2 \end{bmatrix}$$

In a similar manner as described in the previous section we can calculate matrix Ψ_n^2 which will be of the following form.

$$\Psi_n^2 = \sum_{j=1}^{n} q_j^2 \frac{I}{2} \begin{bmatrix} n + O(1) & O(1) & \frac{b}{2}n^2 + O(n) \\ * & n + O(1) & -\frac{a}{2}n^2 + O(n) \\ * & * & \frac{a^2+b^2}{3}n^3 + O(n^2) \end{bmatrix} = k\frac{I}{2}\mathbf{N}^{1/2}\left\{\mathbf{\Gamma} + O(n^{-1})\right\}\mathbf{N}^{1/2}$$

where $k = \sum_{j=1}^{n} q_j^2$ Similarly as before, Ψ_n is positive definite, so we can write:

$$\Psi_n^{-2} = k^{-1} 2 I^{-1} \mathbf{N}^{-1/2}\left\{\mathbf{\Gamma}^{-1} + O(n^{-1})\right\}\mathbf{N}^{-1/2}$$

where,

$$\mathbf{\Gamma}^{-1} = \begin{bmatrix} \frac{a^2+4b^2}{a^2+b^2} & \frac{-3ab}{a^2+b^2} & \frac{-6b}{a^2+b^2} \\ * & \frac{4a^2+b^2}{a^2+b^2} & \frac{6b}{a^2+b^2} \\ * & * & \frac{12}{a^2+b^2} \end{bmatrix}$$

Finally,matrices Ψ_n and Ψ_n^{-1} are similar to the case with no weights and can be seen below:

$$\Psi_n = \frac{\sqrt{1/8Ikn}}{a^2+b^2} \begin{bmatrix} 2a^2+b^2+O(n^{-1}) & ab+O(n^{-1}) & b\sqrt{3a^2+3b^2}+O(n^{-1}) \\ * & a^2+2b^2+O(n^{-1}) & -a\sqrt{3a^2+3b^2}+O(n^{-1}) \\ * & * & \frac{2n}{3^{1/2}}(a^2+b^2)^{3/2}+O(1) \end{bmatrix}$$

$$\Psi_n^{-1} = \frac{\sqrt{2/(Ikn)}}{a^2+b^2} \begin{bmatrix} a^2+2b^2+O(n^{-1}) & -ab+O(n^{-1}) & -3bn^{-1}+O(n^{-2}) \\ * & 2a^2+b^2+O(n^{-1}) & 3an^{-1}+O(n^{-2}) \\ * & * & n^{-1}\sqrt{(}3a^2+3b^2)^{+}O(n^{-2}) \end{bmatrix}$$

It is straightforward to show that conditions (A)-(D) hold in a similar manner.

Bibliography

[1] C. Akerlof, C. Alcock, R. Allsman, T. Axelrod, D. Bennett, K. Cook, K. Freeman, K. Griest, S. Marshall, H.-S. Park, et al. Application of cubic splines to the spectral analysis of unequally spaced data. *The Astrophysical Journal*, 436:787–794, 1994.

[2] S. Algeri, D. A. van Dyk, J. Conrad, and B. Anderson. On methods for correcting for the look-elsewhere effect in searches for new physics. *Journal of Instrumentation*, 11(12):P12010, 2016.

[3] T. Allam Jr, A. Bahmanyar, R. Biswas, M. Dai, L. Galbany, R. Hložek, E. E. Ishida, S. W. Jha, D. O. Jones, R. Kessler, et al. The photometric lsst astronomical time-series classification challenge (plasticc): Data set. *arXiv preprint arXiv:1810.00001*, 2018.

[4] R. Angus, T. Morton, S. Aigrain, D. Foreman-Mackey, and V. Rajpaul. Inferring probabilistic stellar rotation periods using gaussian processes. *Monthly Notices of the Royal Astronomical Society*, 474(2):2094–2108, 2018.

[5] A. Azzalini and A. Bowman. On the use of nonparametric regression for checking linear relationships. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(2):549–557, 1993.

[6] B. H. Baltagi and P. X. Wu. Unequally spaced panel data regressions with ar (1) disturbances. *Econometric Theory*, pages 814–823, 1999.

[7] O. E. Barndorff-Nielsen. Approximate interval probabilities. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):485–496, 1990.

[8] S. Benlloch, J. Wilms, R. Edelson, T. Yaqoob, and R. Staubert. Quasi-periodic oscillation in seyfert galaxies: Significance levels. the case of markarian 766. *The Astrophysical Journal Letters*, 562(2):L121, 2001.

[9] I. Berenblut and G. Webb. A new test for autocorrelated errors in the linear regression model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 35(1):33–50, 1973.

[10] F. Beutler. Alias-free randomly timed sampling of stochastic processes. *IEEE transactions on information theory*, 16(2):147–152, 1970.

[11] T. A. Bishop, E. J. Dudewicz, J. M. MURITZ, and M. A. Stephens. Percentage points of a quadratic form in student t variates. *Biometrika*, 65(2):435–439, 1978.

[12] P. Bloomfield. *Fourier analysis of time series: an introduction.* John Wiley & Sons, 2004.

[13] J. Bouvier, S. P. Matt, S. Mohanty, A. Scholz, K. G. Stassun, and C. Zanni. Angular Momentum Evolution of Young Low-Mass Stars and Brown Dwarfs: Observations and Theory. In H. Beuther, R. S. Klessen, C. P. Dullemond, and T. Henning, editors, *Protostars and Planets VI*, page 433, Jan. 2014.

[14] R. W. Butler. *Saddlepoint approximations with applications*, volume 22. Cambridge University Press, 2007.

[15] C. Chatfield. *The analysis of time series: an introduction.* Chapman and hall/CRC, 2003.

[16] Z. Chen, B. Wang, and A. N. Gorban. Multivariate gaussian and student-t process regression for multi-output prediction. *Neural Computing and Applications*, 32(8):3005–3028, 2020.

[17] A. Cumming, G. W. Marcy, and R. P. Butler. The lick planet search: detectability and mass thresholds. *The Astrophysical Journal*, 526(2):890, 1999.

[18] T. Do, A. M. Ghez, M. R. Morris, S. Yelda, L. Meyer, J. R. Lu, S. D. Hornstein, and K. Matthews. A near-infrared variability study of the galactic black hole: a red noise source with no detected periodicity. *The Astrophysical Journal*, 691(2):1021, 2009.

[19] J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4):409–428, 1950.

[20] D. Duvenaud, J. Lloyd, R. Grosse, J. Tenenbaum, and G. Zoubin. Structure discovery in nonparametric regression through compositional kernel search. In *International Conference on Machine Learning*, pages 1166–1174. PMLR, 2013.

[21] E. Erdogan, S. Ma, A. Beygelzimer, and I. Rish. Statistical models for unequally spaced time series. In *proceedings of the 2005 SIAM International conference on data mining*, pages 626–630. SIAM, 2005.

[22] E. Feigelson et al. 21st century statistical and computational challenges in astrophysics. 2020. *Annual Review of Statistics and Its Application*, 8.

[23] D. Froebrich, J. Campbell-White, A. Scholz, J. Eislöffel, T. Zegmott, S. J. Billington, J. Donohoe, S. V. Makin, R. Hibbert, R. J. Newport, R. Pickard, N. Quinn, T. Rodda, G. Piehler, M. Shelley, S. Parkinson, K. Wiersema, and I. Walton. A survey for variable young stars with small telescopes: First results from HOYS-CAPS. *Monthly Notices of the Royal Astronomical Society*, 478(4):5091–5103, Aug. 2018.

[24] D. Froebrich, E. Derezea, A. Scholz, J. Eislöffel, S. Vanaverbeke, A. Kume, C. Herbert, J. Campbell-White, N. Miller, B. Stecklum, et al. A survey for variable young stars with small telescopes–iv. rotation periods of ysos in ic 5070. *Monthly Notices of the Royal Astronomical Society*, 506(4):5989–6000, 2021.

[25] D. Froebrich, A. Scholz, J. Eislöffel, and B. Stecklum. A survey for variable young stars with small telescopes–iii. warm spots on the active star v1598

cyg. *Monthly Notices of the Royal Astronomical Society*, 497(4):4602–4613, 2020.

[26] I. J. Good. Significance tests in parallel and in series. *Journal of the American Statistical Association*, 53(284):799–813, 1958.

[27] P. Hall, J. Reimann, and J. Rice. Nonparametric estimation of a periodic function. *Biometrika*, 87(3):545–557, 2000.

[28] J. P. Halpern, K. Leighly, and H. Marshall. An extreme ultraviolet explorer atlas of seyfert galaxy light curves: Search for periodicity. *The Astrophysical Journal*, 585(2):665, 2003.

[29] J. D. Hamilton. *Time series analysis*. Princeton university press, 2020.

[30] E. J. Hannan. The estimation of frequency. *Journal of Applied probability*, 10(3):510–519, 1973.

[31] S. Heerah, R. Molinari, S. Guerrier, and A. Marshall-Colón. Granger-causal testing for irregularly sampled time series with application to nitrogen signaling in arabidopsis. *bioRxiv*, 2020.

[32] W. Herbst, J. Eislöffel, R. Mundt, and A. Scholz. The Rotation of Young Low-Mass Stars and Brown Dwarfs. In B. Reipurth, D. Jewitt, and K. Keil, editors, *Protostars and Planets V*, page 297, Jan. 2007.

[33] I. A. Ibragimov and R. Z. Has' Minskii. *Statistical estimation: asymptotic theory*, volume 16. Springer Science & Business Media, 2013.

[34] J.-P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3/4):419–426, 1961.

[35] Ž. Ivezić, A. J. Connolly, J. T. VanderPlas, and A. Gray. *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data*, volume 8. Princeton University Press, 2019.

[36] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions, volume 2*, volume 289. John wiley & sons, 1995.

[37] N. Jonson, S. Kotz, and N. Balakrishnan. Continuous univariate distributions, vol. 2, 1970.

[38] V. Katkovnik. Robust m-periodogram. *IEEE Transactions on Signal processing*, 46(11):3104–3109, 1998.

[39] C. KI Williams. *Gaussian processes formachine learning.* Taylor & Francis Group, 2006.

[40] G. S. Kimeldorf and G. Wahba. A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, 41(2):495–502, 1970.

[41] A. Kume, S. P. Preston, and A. T. Wood. Saddlepoint approximations for the normalizing constant of fisher–bingham distributions on products of spheres and stiefel manifolds. *Biometrika*, 100(4):971–984, 2013.

[42] A. Kume and A. T. Wood. Saddlepoint approximations for the bingham and fisher–bingham normalising constants. *Biometrika*, 92(2):465–476, 2005.

[43] D. Kuonen. Miscellanea. saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika*, 86(4):929–935, 1999.

[44] K.-c. Liang, X. Wang, and T.-H. Li. Robust discovery of periodically expressed genes using the laplace periodogram. *BMC bioinformatics*, 10(1):15, 2009.

[45] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science*, 39(2):447–462, 1976.

[46] R. Lugannani and S. Rice. Saddle point approximation for the distribution of the sum of independent random variables. *Advances in applied probability*, 12(2):475–490, 1980.

[47] D. J. MacKay. Introduction to gaussian processes. *NATO ASI series F computer and systems sciences*, 168:133–166, 1998.

[48] A. Malz, R. Hložek, T. Allam, A. Bahmanyar, R. Biswas, M. Dai, L. Galbany, E. Ishida, S. Jha, D. Jones, et al. The photometric lsst astronomical time-series classification challenge plasticc: Selection of a performance metric for classification probabilities balancing diverse science goals. *The Astronomical Journal*, 158(5):171, 2019.

[49] H.-S. Oh, D. Nychka, T. Brown, and P. Charbonneau. Period analysis of variable stars by robust smoothing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(1):15–30, 2004.

[50] D. M. Palmer. A fast chi-squared technique for period search of irregularly sampled data. *The Astrophysical Journal*, 695(1):496, 2009.

[51] M. S. Paolella. *Linear models and time-series analysis: regression, ANOVA, ARMA and GARCH*. John Wiley & Sons, 2018.

[52] M. B. Priestley. *Spectral analysis and time series: probability and mathematical statistics*. Number 04; QA280, P7. 1981.

[53] J. D. Reimann. *Frequency estimation using unequally-spaced astronomical data*. PhD thesis, Citeseer, 1994.

[54] J. D. Scargle. Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, 1982.

[55] M. Schulz and K. Stattegger. Spectrum: Spectral analysis of unevenly spaced paleoclimatic time series. *Computers & Geosciences*, 23(9):929–945, 1997.

[56] A. Schuster. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism*, 3(1):13–41, 1898.

[57] A. Schwarzenberg-Czerny. The distribution of empirical periodograms: Lomb-scargle and pdm spectra. *Monthly Notices of the Royal Astronomical Society*, 301(3):831–840, 1998.

[58] A. Shah, A. Wilson, and Z. Ghahramani. Student-t processes as alternatives to gaussian processes. In *Artificial intelligence and statistics*, pages 877–885. PMLR, 2014.

[59] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang. Short-term solar power forecasting based on weighted gaussian process regression. *IEEE Transactions on Industrial Electronics*, 65(1):300–308, 2017.

[60] Z. Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967.

[61] R. F. Stellingwerf. Period determination using phase dispersion minimization. *The Astrophysical Journal*, 224:953–960, 1978.

[62] A. Terras. *Harmonic Analysis on Symmetric Spaces and Applications I*. Springer Science & Business Media, 2012.

[63] A. M. Thieler, M. Backes, R. Fried, and W. Rhode. Periodicity detection in irregularly sampled light curves by robust regression and outlier detection. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 6(1):73–89, 2013.

[64] A. M. Thieler, R. Fried, J. Rathjens, et al. Robper: An r package to calculate periodograms for light curves based on robust regression. *Journal of Statistical Software*, 69(9):1–36, 2016.

[65] J. Timmer and M. Koenig. On generating power law noise. *Astronomy and Astrophysics*, 300:707, 1995.

[66] J. T. VanderPlas. Understanding the lomb–scargle periodogram. *The Astrophysical Journal Supplement Series*, 236(1):16, 2018.

[67] J. T. VanderPlas and Ž. Ivezic. Periodograms for multiband astronomical time series. *The Astrophysical Journal*, 812(1):18, 2015.

[68] S. Vaughan. A simple test for periodic signals in red noise. *Astronomy & Astrophysics*, 431(1):391–403, 2005.

[69] S. Vaughan, R. Edelson, R. Warwick, and P. Uttley. On characterizing the variability properties of x-ray light curves from active galaxies. *Monthly Notices of the Royal Astronomical Society*, 345(4):1271–1284, 2003.

[70] H. Von Storch and F. W. Zwiers. *Statistical analysis in climate research*. Cambridge university press, 2001.

[71] A. Walker. On the estimation of a harmonic component in a time series with stationary independent residuals. *Biometrika*, 58(1):21–36, 1971.

[72] Y. Wang, R. Khardon, and P. Protopapas. Nonparametric bayesian estimation of periodic light curves. *The Astrophysical Journal*, 756(1):67, 2012.

[73] C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[74] W.-X. Zhou and D. Sornette. Statistical significance of periodicity and log-periodicity with heavy-tailed correlated noise. *International Journal of Modern Physics C*, 13(02):137–169, 2002.