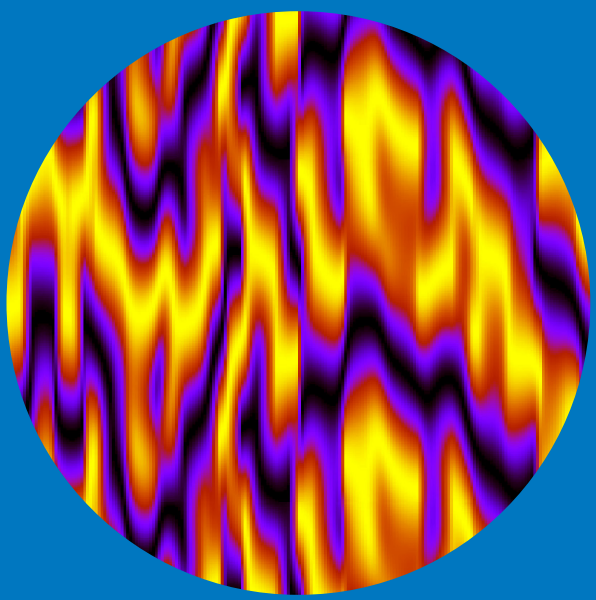
Department of Computer Science

From periodic to cyclic processes in stellar magnetic activity research: time series analysis methods and their applications

Nigul Olspert



Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 224/2018

From periodic to cyclic processes in stellar magnetic activity research: time series analysis methods and their applications

Nigul Olspert

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall AS1 in Maarintie 8, Aalto University on 16th of November 2018 at 12 noon.

Aalto University
School of Science
Department of Computer Science

Supervising professor
Professor Aki Vehtari, Aalto University, Finland

Thesis advisors
Professor Maarit Käpylä, Max Planck Institute for Solar System Research, Germany
Professor Jaan Pelt, Tartu Observatory, Estonia

Preliminary examiners
Professor Ilya Usoskin, Sodankylä Geophysical Observatory, Finland
Professor Katalin Oláh, Konkoly Observatory of the Hungarian Academy of Sciences, Hungary

Opponent
Professor Ivan Andronov, Odessa National Maritime University, Ukraine

Aalto University publication series
DOCTORAL DISSERTATIONS 224/2018

ISBN 978-952-60-8292-9 (printed)
ISBN 978-952-60-8293-6 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-8293-6

Unigrafia Oy
Helsinki 2018

Finland

441 697
Printed matter

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Nigul Olspert

Name of the doctoral dissertation
From periodic to cyclic processes in stellar magnetic activity research: time series analysis methods and their applications

Publisher School of Science

Unit Department of Computer Science

Series Aalto University publication series DOCTORAL DISSERTATIONS 224/2018

Field of research Computer Science

Manuscript submitted 26 August 2018 **Date of the defence** 16 November 2018

Permission to publish granted (date) 22 October 2018 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

One of the unanswered questions in stellar activity research is how the rotation period and the magnetic cycle period of a star are related. A prerequisite to answering this question is being able to estimate both of these quantities as reliably as possible. Throughout the years, the prevailing methods have mostly been based on the well-known Lomb-Scargle periodogram. However, such a periodogram and its analogues are hard to interpret, when the input signal is not fully periodic. Observations of the solar cycle properties through factors, such as, the sunspot number over time, and non-linear dynamo models both clearly indicate that the stellar dynamo process is indeed quasi-periodic and non-stationary. Hence, a more correct approach is to relax the assumption of periodicity. The development and application of such methods is the main aim of this thesis.

To investigate stellar cycles theoretically, the most advanced approach is to use global 3D magnetoconvection models solving the full MHD equations. These have only recently started to show similar quasi-periodic behaviour as the observed datasets. Real and simulated data pose completely different requirements for the analysis methods. While the former are unevenly sampled and sparse, the latter contain vast amounts of multi-dimensional data. For the estimation of magnetic cycles, an additional problem with observational data is their relative shortness. Throughout the thesis I will thoroughly address the above aspects.

In this work, several methods have been developed for analysing time series of active stars. Carrier fit (CF) method is a simple and efficient way for fitting a continuous model into the time series of active stars. Side by side with this method a visualisation technique is used, which allows deviations from strict periodicity at different times to be easily detected, revealing the quasi-periodic and non-stationary effects. Another method, called D^2 phase dispersion statistic is a robust tool for estimating periods of a quasi-periodic time series. It allows a simple generalisation to multiple dimensions, which is useful when analysing datasets of 3D magnetoconvection simulations. We also use probabilistic models for period estimation. For short datasets, the period estimates can become sensitive to the ways the linear trend in the data is handled. We show that for proper treatment one needs to include the trend component in the model, while using prior distributions for regularisation. Other probabilistic models, which have been used in the study include Gaussian processes (GPs) with periodic and quasi-periodic covariance functions. From the toolbox of methods suitable for non-stationary data, we have used ensemble empirical mode decomposition (EEMD).

Our applications involve a young solar analogue LQ Hya, 3D magnetoconvection simulation called PENCIL-Millennium and a Mount Wilson (MW) stellar chromospheric activity dataset. For LQ Hya, we estimated the mean rotation period, surface differential rotation coefficient and fitted a continuous light curve model using the CF method. In the case of PENCIL-Millennium simulation data, we used both EEMD and the D^2 statistic to extract the different dynamo modes with their locations in the convection zone. These modes include a five-year cycle, which is an analogue of the 22-year magnetic cycle of the Sun, and two much longer cycles. Furthermore, with the help of the D^2 statistic, we were able to find a very incoherent short cycle with a period around half a year, which resembles the quasi-biennial oscillations of the Sun. In the analysis of the MW dataset, the main aim was to repeat the cycle length estimation with a simple harmonic model while properly handling trends, but also trying out periodic and quasi-periodic GP models. All three methods led to quite similar results, however, the reliability of the quasi-periodic model remained questionable due to the shortness of the datasets. We confirmed the existence of two different star populations in the activity diagram. However, as opposed to the formerly known positive correlations within both of these branches, we confirmed only a positive correlation within the inactive branch. The results were also compared to the recent 3D magnetoconvection simulations.

Keywords time series analysis methods, quasi-periodicity, stellar magnetic activity

ISBN (printed) 978-952-60-8292-9 **ISBN (pdf)** 978-952-60-8293-6

ISSN (printed) 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2018

Pages 240 **urn** http://urn.fi/URN:ISBN:978-952-60-8293-6

Preface

This thesis has been carried out as a part of the Academy of Finland ReSoLVE Centre of Excellence project, in the DYNAMO team operating at the Department of Computer Science, Aalto University. My modest contribution to this grand project has been related to statistical time series analysis with primary focus on period and cycle search. The path how I got involved into this field deserves a little explanation. I have always had an interest in astronomy, so back in 1993, as a school kid, I met Peep Kalv, an astronomer from Tallinn Observatory, who gave me an exercise to digitalise the photometry of a certain variable star from old photographic plates. The idea would have been to later find its light curve and estimate its rotation period. Unfortunately this task remained unfinished, partially because the instrument broke down and partially because my interests in computers and programming shifted the focus to other things. Much later, during my bachelor studies I was introduced to Jaan Pelt who has been working on time series analysis of astronomical datasets for several decades. His interesting ideas strongly influenced me to choose this field for my further studies. Now, during the last couple of years I have become more and more interested in machine learning and its big potential, not only in time series analysis, but in astrophysics in general. As the cross-disciplinary domains often open up new interesting approaches for solving unanswered problems, I'm looking forward to work in that direction.

It is hard to overestimate how interesting the years working together with the experts in the dynamo theory and stellar astrophysics have been, but there is still a lot to learn. For getting this opportunity, first of all, I am very grateful to my supervisor Maarit Käpylä, who accepted me as a doc-

toral student, never stopped motivating me to continue with the work and most importantly, who is a major contributor in all of the papers included in the theses. She also substantially helped with the thesis manuscript on topics related to dynamo theory. Next I want to thank Jaan Pelt, who has been my long-time supervisor and now also a co-author. The discussions with him have always inspired me and boosted my interest in science. Moreover, several of his methods are central to the current thesis. From the other co-authors I want to thank Jyri Lehtinen and Thomas Hackman for the insights in stellar magnetic activity and Alexander Grigorievskiy in machine learning. I also want to thank Aki Vehtari for giving me the opportunity to join the probabilistic machine learning group seminars, which have significantly broadened my understanding and without the doubt helped me in writing the last two papers included in the thesis. I also thank the pre-examiners of the thesis, Katalin Oláh and Ilya Usoskin, for their valuable comments on the manuscript and for pointing out some mistakes.

Last, but not least, I want to thank my co-workers and team members from Aalto and Helsinki universities with whom there has never been lack of sarcastic, political, scientific and philosophical discussions (but also about music and mushrooms). And to wrap it up, I am glad that I had the possibility to enjoy the beautiful Finnish nature, including the memorable trip to Lapland, and the superb possibilities for outdoor sports around Helsinki, especially skiing and running.

Espoo, October 30, 2018,

Nigul Olspert

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I J. Pelt, N. Olspert, M. J. Mantere, I. Tuominen. Multiperiodicity, modulations and flip-flops in variable star light curves. I. Carrier fit method. *Astronomy & Astrophysics*, Volume 535, id.A23, 12 pp., 11 2011.

II N. Olspert, M. J. Käpylä, J. Pelt, E. M. Cole, T. Hackman, J. Lehtinen, G. W. Henry. Multiperiodicity, modulations, and flip-flops in variable star light curves. III. Carrier fit analysis of LQ Hydrae photometry for 1982-2014. *Astronomy & Astrophysics*, Volume 577, id.A120, 13 pp., 05 2015.

III M. J. Käpylä, P. J. Käpylä, N. Olspert, A. Brandenburg, J. Warnecke, B. B. Karak, J. Pelt. Multiple dynamo modes as a mechanism for long-term solar activity variations. *Astronomy & Astrophysics*, Volume 589, id.A56, 24 pp., 05 2016.

IV N. Olspert, M. J. Käpylä, J. Pelt. Method for estimating cycle lengths from multidimensional time series: Test cases and application to a massive "in silico" dataset. In *2016 IEEE International Conference on Big Data*, Washington, DC, USA, 12 2016.

V N. Olspert, J. Pelt, M. J. Käpylä, J. Lehtinen. Estimating activity cycles

with probabilistic methods: I. Bayesian generalised Lomb-Scargle periodogram with trend. *Astronomy & Astrophysics*, Volume 615, id.A111, 12 pp., 07 2018.

VI N. Olspert, J. Lehtinen, M. J. Käpylä, J. Pelt, A. Grigorievskiy. Estimating activity cycles with probabilistic methods II. The Mount Wilson Ca H&K data. *Astronomy & Astrophysics*, Volume 619, id.A111, 20 pp., 11 2018.

Author's Contribution

Publication I: "Multiperiodicity, modulations and flip-flops in variable star light curves. I. Carrier fit method"

We introduce the carrier fit (CF) method useful for modelling time series of quasi-periodic processes, which are described by a series of band-limited power spectra. We propose two different families of functions suitable for low frequency modulators and discuss the notion of instantaneous amplitude and frequency. Furthermore, we suggest local phase and frequency visualisation techniques, which allow to detect behaviours that deviate from pure harmonic oscillation, with a certain mean period. We also illustrate the benefit of the method on several artificial as well as some real examples. The author contributed to this paper by implementing different types of modulators to the CF model, developing the visualisation tools and by participating in the test computations.

Publication II: "Multiperiodicity, modulations, and flip-flops in variable star light curves. III. Carrier fit analysis of LQ Hydrae photometry for 1982-2014"

This paper is a continuation of Paper I, dedicated to analysing the photometry of a young solar analogue, LQ Hya. Majority of the data analysis presented in this paper was carried out by the author, including: Estimation of the mean rotation period, followed by finding the optimal local and global carrier fit models; visualisation of the results via phase diagrams to detect transient and intermittent behaviour on the surface of the star;

finding the epochs of local minima of the model and comparing them to the results from other studies; and calculating the estimates for the differential rotation coefficient and Carrington latitude for the star.

Publication III: "Multiple dynamo modes as a mechanism for long-term solar activity variations"

Here we analyse the multidimensional time series from the global 3D magnetoconvection simulation that covers more than 80 solar-like magnetic cycles, hence the nickname 'PENCIL-Millennium'. The author performed all the statistical analyses presented in the paper, including: Using an ensemble empirical mode decomposition to establish the significant cyclic modes of the magnetic field and velocity components as well as kinetic helicity. We also established the locations of the dynamo modes in the simulated convection zone. D^2 statistic was used as an independent check for the cycle length estimates. The time dependence of the parities of the modes was calculated, which helped to interpret the irregular epochs at the surface as a result of the interplay of the different parity states of the different dynamo modes. We also analysed the time evolution of the individual magnetic cycle lengths, magnetic 'butterfly' wing inclinations and widths, and whether these quantities are related to magnetic parity and energy states.

Publication IV: "Method for estimating cycle lengths from multidimensional time series: Test cases and application to a massive "in silico" dataset"

We start with defining the D^2 phase dispersion statistic in a form suitable for multidimensional data and continue by investigating its properties via examples involving varying coherence length, S/N as well as adding non-stationary effects (only temporarily existing oscillations). For a real case study, we apply the multidimensional D^2 statistic to the same, but longer PENCIL-Millennium simulation data, now covering nearly 150 magnetic cycles, as in the previous paper. We establish the presence of four magnetic cycles with different periods, coherence lengths and locations in the simulated convection zone. This study confirms some of the results from

the previous study, but also contributes a new finding about the existence of a short and extremely incoherent cycle, that we interpret to be an analogue of solar quasi-biennial oscillation. All the experiments and data analysis in this paper were performed by the author.

Publication V: "Estimating activity cycles with probabilistic methods: I. Bayesian generalised Lomb-Scargle periodogram with trend"

This study introduces a Bayesian period estimation method that accounts for a possible linear trend in the data. We discuss the importance of such a model in the context of time series, which are short in respect to the period of the periodic or cyclic phenomena in the data. We show through examples that using this model is in general more optimal to either de-trending the data or leaving the data unmodified before the period analysis. We also provide an example of a case where using prior distributions for the parameters becomes crucial to gaining meaningful results. As in many astronomical datasets, the amount of observational noise changes over the seasons, we investigate the ways in which the handling of the noise in the model affects the results. In particular, we compare two cases: the constant noise variance vs. the noise variance estimated empirically from the data using small segments. The method development, experiments and the data analysis were all performed by the author.

Publication VI: "Estimating activity cycles with probabilistic methods II. The Mount Wilson Ca H&K data"

The main aim of this paper is to test the previously postulated existence of different populations of magnetically active late type stars, when their magnetic cycle periods are examined as function of their rotation period. For that purpose, we first estimate the cycles from the Mount Wilson Ca H&K dataset using the method developed in the previous paper as well as two Gaussian process models. Using the results, we plot the stellar activity diagrams and perform Gaussian mixture model clustering to show the existence of two different populations, the so-called inactive, and active stars. We confirm the previously known positive correlation in the

inactive population, but find no significant correlation in the active population. We also compare the results to 3D magnetoconvection models to help to understand the results in the light of dynamo theory. The author's contribution includes all the method development and statistical analysis in this paper.

List of abbreviations

BIC	Bayesian information criterion
CF	Carrier fit
CPS	Continuous period search
CV	Cross-validation
CZ	Convection zone
DFT	Discrete Fourier transform
DI	Doppler imaging
EEMD	Ensemble empirical mode decomposition
EMD	Empirical mode decomposition
FAP	False alarm probability
FT	Fourier transform
GP	Gaussian process
LS	Lomb-Scargle
PD	Phase dispersion
PDM	Phase dispersion minimisation
S/N	Signal-to-noise ratio
SN	Sunspot number
STFT	Short time Fourier transform
WSS	Wide sense stationary
WT	Wavelet transform

List of notations

$\arg\max f(x)$	The value of the argument x corresponding to the maximum of the function f
x^*	Complex conjugate of x
$x * y$	Convolution of x and y. $(x * y)(t) = \int_{-\infty}^{\infty} x(\tau)y(t-\tau)d\tau$
$\delta(x)$	Dirac delta function
$\mathbb{E}(x)$	Expected value of random variable x
$\lfloor x \rfloor$	Floor function
$\text{frac}(x)$	Fractional value. $\text{frac}(x) = x - \lfloor x \rfloor$
f	Frequency of the process
$\mathcal{F}$	Fourier transform operator
$\mathbf{A} \circ \mathbf{B}$	Hadamard product of two matrices. $(\mathbf{A} \circ \mathbf{B})_{ij} = A_{ij}B_{ij}$
$\mathbf{I}$	Identity matrix
$\mathbb{I}(\text{condition})$	Indicator function. $\mathbb{I}(\text{true}) = 1, \mathbb{I}(\text{false}) = 0$
$\langle x, y \rangle$	Inner product of x and y. $\langle x, y \rangle = \int_{-\infty}^{\infty} x(t)y^*(t)dt$
δ_{ij}	Kronecker delta. $\delta_{ij} = 1$, if $i = j$, 0 otherwise
$\min(x, y)$	Minimum of x and y
$\|\|\mathbf{a}\|\|$	Vector magnitude or Euclidean norm. $\|\|\mathbf{a}\|\| = \sqrt{\mathbf{a}^\mathsf{T}\mathbf{a}}$
ω	Angular frequency. $\omega = 2\pi f$
$\mathbf{1}$	Vector of all ones
P	Period of the process. $P = 1/f$
R'_{HK}	Stellar chromospheric activity index

1. Introduction

The central topic of the current work is the period estimation from time series of quasi-periodic processes. The motivation for this study arose in the context of solar and stellar activity, where it is common that observed quantities follow not purely periodic, but quasi-periodic behaviour. The best known example of a quasi-periodic process is the solar magnetic activity cycle, observationally studied over hundreds of years. The length of the cycle varies in the range of 9-12 years. The second well known quasi-periodic process is related to the motion of starspots on stellar surfaces. The non-uniform rotation, meaning that different latitudes rotate with different speeds, translates to changing rotation speeds of starspots if they occur at different latitudes. Furthermore, these two phenomena, the magnetic cycle and rotation, are intimately linked in stars[1], major unresolved problems being related to the process linking them. This process is called as the hydromagnetic dynamo, and it is thought to operate so that kinetic energy of rotation and turbulence are processed into magnetic form. If such a process really took place in stars, one would expect that the properties of the magnetic cycles would depend deterministically on the rotation rate. However, the interpretations of observational data, often relying on rather old and basic, although well-established methodologies, have not yet converged; hence, the field of 'cycle science' could benefit from more modern and versatile methods. The development and application of such tools are the leading goals of this thesis.

The observed time series of astronomical datasets are usually unevenly (irregularly) sampled and gapped, which, on one hand, narrows down

[1]Throughout the text when referring to a 'star', the meaning is limited to a late-type active star, even if not exactly specified. We might also use other similar expressions, such as Sun-like star or solar-type star.

the range of suitable statistical methods for period detection and, on the other hand, introduces complications when interpreting the results. The groundbreaking work in period estimation for unevenly sampled time series was carried out several decades ago by Vaníček [146], Deeming [36], Lomb [91], Ferraz-Mello [43], Scargle [125] and others. Since then, a large number of sophisticated models and algorithms have been proposed. The primary focus of these studies has been on detection of strict periodicities, while the domain of quasi-periodicity has not been studied as extensively. One of the earlier steps towards bridging this gap was made by Pelt [104]. More recently with the increase in computational power, Bayesian methods have become popular, where analogues of some of the standard methods in the period analysis have been derived. Furthermore, Gaussian processes, due to their flexibility, are extremely useful in the analysis of quasi-periodic time series.

In addition to real-world data, there is a plethora of available data from numerical simulations. During the past three decades, 3D global magnetohydrodynamical simulations have been carried out with ever-increasing grids and complexity. However, only very recently have these models reached parameter regimes where oscillatory solar-like dynamo modes are found [52, 28, 127, 71, 41]. Some of these simulations exhibit remarkably similar behaviour as the Sun with a quasi-regular cycle, multitude of them also requiring special period search tools to be analysed [70]. The primary difference to the observations is the massive amount of spatially resolved data. Due to these circumstances only well-scalable methods for period estimation can be used.

The structure of the thesis is as follows. Chapter 2 reviews the aspects of solar and stellar magnetic activity and discusses some of the open questions which form the main motivation behind our study. In Chapter 3, we review the widely used period search methods suitable for strictly periodic time series. Chapter 4 introduces the notions of quasi-periodicity (or cyclicity) with tools for the mean period estimation for these kinds of time series. In Chapter 5, the aspects of non-stationarity are discussed as well as the most widely used methods in this domain known as multiresolution methods. Finally, in Chapter 6 we summarise the applications in our research. In these applications, we use different astronomical datasets from real observations as well as simulations.

2. Solar and stellar magnetic activity

2.1 Stellar magnetic cycles

The Sun is a magnetic star and most notably, its magnetic field, being cyclic, varies in time. The primary manifestation of the solar magnetic activity appears in the form of sunspots, which are dark and cool areas seen in the locations of very strong magnetic fields called active regions. The solar cycle is usually defined as the average interval between the times of two consecutive minimum spot coverages. This corresponds to the 11-year Schwabe cycle (see the bottom panel of Fig. 2.1); however, the magnetic cycle is twice as long, thus including two reversals of the magnetic field polarities during the 22 year interval. The latter is known as the Hale cycle. The notion of cycle period is also ambiguous because of the coexisting cycles of different time scales. There is evidence of the centennial time scale Gleissberg cycle, which manifests itself as a modulation of the sunspot cycle amplitudes. Furthermore, there is evidence of non-stationary processes leading to ceased or pronounced surface activity. These seemingly irregular epochs include Maunder minimum, Dalton minimum, and modern grand maximum occurring during 1645-1715, in the beginning of 19th century and in the second half of the 20th century, respectively. There also exist so-called quasi-biennial variations around the time-scale of one or two years seen in some solar activity tracers [15]. These are not the only cyclic phenomena reported in the literature. [for a review see 59].

The 11-year cycle is well visible in the time series of the Sunspot Number (SN), the historical records of which span the last four centuries. Dur-

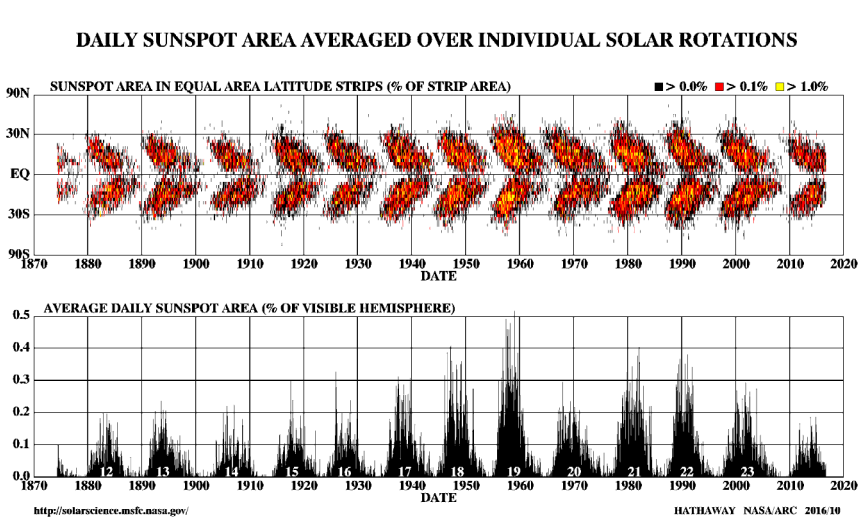


Figure 2.1. Solar cycle. On top: solar butterfly diagram. On bottom: sunspot areas. Image credit: D. H. Hathaway, NASA/ARC.

ing the solar maxima, there is a high number of visible sunspots, and during the minima very low number or zero sunspots. It is also known that the two consecutive solar cycles overlap – the spots of the old cycle start appearing at higher latitudes, while the old cycle spots still exist at lower latitudes. When plotted, this time-latitude dependence forms the so-called butterfly diagram (see the top panel of Fig. 2.1). The solar cycle is usually separated into four phases: minimum, ascending, maximum, and declining (or descending), the latter phase being generally the longest. The time series of SN is not periodic, but the length of the cycle and its amplitude varies. It is worth noting two interesting facts related to this: the integral strengths of the odd cycles are known to be larger than the even cycles (Gnevyshev-Ohl rule), and the length of the ascending phase is inversely proportional to the cycle amplitude (Waldmeier effect) [59].

In addition to SN data, there are other proxies reflecting the spottedness of the Sun, namely, group sunspot number and sunspot areas. These datasets highly correlate with the SN data. The same 11-year cyclicity is also visible in other observed proxies, for instance, total solar irradiation (TSI), which is correlating with SN on longer and anticorrelating on shorter time scales. During the solar maximum TSI is increased, despite the higher number of sunspots compared to the minimum. This is because the bright, hotter magnetic features, such as, faculae and bright elements in the photosphere overcompensate the effect of the dark spots. However,

the short-lived[1] sunspots produce sharp dips in the TSI signal (so-called blanketing effect). In addition to sunspots, there are other phenomena related to the magnetic activity, such as the flares, coronal mass ejections and coronal emission from near ultraviolet to X-ray wavelengths.

There exist other Sun-like stars which are also magnetic and frequently exhibit cyclic behaviour. This fact has been established by long-term monitoring of the active stars through photometry and spectrometry. The first indirect evidence of starspots was reported by Kron [83] via observing the photometric light variations, however, true detections of starspots came only very recently using interferometry. One of the still ongoing surveys of stellar activity cycles is the Mount Wilson Ca II H&K project, which was started by Wilson [156].

The magnetic cycles show up as modulations in the photometry. Simply speaking, one can assume that the photometric mean curve represents the average spottedness level of the star, low brightness corresponding to high spot coverage and vice versa. On the other hand, the amplitude of the light curve can correspond to the axisymmetricity of the spot distribution. Higher amplitudes mean that the spots are grouped closer to each other over longitudes, and lower amplitudes that the spots are more scattered over a wider longitudinal range. Of course, this is affected by the inclination of the star. For instance, no rotational modulation is seen at all when the star is visible to us pole-on.

More phenomena have been discovered from the observations of solar-type stars. One interesting phenomenon is the existence of active longitudes [16, 81, 88], one explanation for which can be provided by the existence of azimuthal dynamo waves [93, 33]. The active longitude is a longitudinally extended and coherent structure which rotates in a rigid body manner independent of the rotation of the gaseous plasma. Also in the Sun, active longitudes have been detected, but their persistence over time is still debated [17, 106]. There is another more hypothetical phenomenon called a 'flip-flop', which refers to the switch in the activity level of longitudes separated from each other by $180°$. In this process, the stronger active region becomes the weaker one and vice versa [55, 92].

[1] From few hours to days or in case of spot groups for months

2.2 Stellar rotation

The Sun as a whole does not rotate as a rigid body, but there is a phenomenon called differential rotation. Both radial and latitudinal differential rotation occur in its outer gaseous plasma layer called the convection zone (CZ). As most of the observable light originates from the lowest layer of the solar atmosphere, called photosphere, right above the top of the CZ, we can directly measure only the surface rotation (due to the presence of the sunspots). The solar internal differential rotation pattern has been determined through helioseismology, which has revealed the presence of shear layers[2] in the boundary regions of the CZ. The rotation speed of the sunspots on the surface is highly dependent on their latitude and/or anchoring depths. Therefore, we can either talk about their individual rotation periods or their statistical average (the so-called Carrington rotation period). Separately, there are notions of equatorial and polar rotation periods, which refer to the rotation periods of the surface layer around latitude of 0° and 90° with respect to the rotational axis. The solar latitudinal differential rotation can be empirically expressed by an equation

$$\Omega(\theta) = A + B\sin^2(\theta) + C\sin^4(\theta), \tag{2.1}$$

where Ω is the angular velocity in degrees per day, θ is the solar latitude and $A = 14.713\ ^\circ/\mathrm{d}$, $B = -2.396\ ^\circ/\mathrm{d}$ and $C = -1.787\ ^\circ/\mathrm{d}$ are the constants, which means fast equator and slow poles. The known sidereal values for equatorial, polar and Carrington rotation periods for the Sun are correspondingly 24.47, ≈ 38 and 25.38 days. The Carrington latitude being around 26°. For the Sun, rotational and cycle periods are quite accurately established, however, this is not the case for distant active late-type stars, for which the spots are not directly resolved (with some recent exception of interferometry [103]). Thus, the primary means of obtaining information about the rotation period and magnetic cycle is through measurements that integrate the emitted light over the whole disk, such as photometry or spectrometry. Many of the observed solar analogues are much younger and more rapidly rotating stars compared to the Sun. The faster rotation has enabled use of the Doppler imaging (DI) and Zeeman Doppler imaging

[2]The shear layer between the radiative and convective zones of the Sun is called tachocline.

methods to track the phases of the spots on the surface of the stars and build surface temperature maps [54, 34, 101, 73] as well as magnetograms [76, 120, 56]. From previous studies, it has become clear that the spot patterns of these stars significantly differ from the patterns of the Sun. Whereas on the Sun, the spots are relatively small and located in the mid and low latitude regions, on the young solar-type stars, the spots tend to be large and located on high latitudes, including polar regions. However, large spots in the latter context can mean an active regions with smaller spots plus faculae. While these giant spots might exist, there is currently no direct proof from observations. Another observed fact is the existence of differential rotation in faster rotators [57, 115, 114].

The problem of not being able to resolve the surface of the stars renders estimation of the rotation periods a complicated task. We do not know the number of spots and their locations on the surface of the star at any given time moment, neither do we know for certain the type of latitudinal differential rotation[3]. This is because theoretically it is expected that stars change their rotation profile from solar- to anti-solar, with slow equator and faster poles, around the solar rotation rate. However, from DI studies, anti-solar differential rotation is found mainly for giant stars [74]. We also cannot distinguish the latitudinal differential rotation from the possible radial one. Nevertheless, using Bayesian inference, fitting physical models to observed data has been achieved [see e.g. 75]. The primary difficulty with these so-called inversion methods is the high number of free parameters. To enable the inference, these models require several prior assumptions (e.g. the number of starspots and their lifetimes). One of the simplest ways to model the light curves of active stars was introduced in Budding [29]. We have implemented this algorithm to provide an online tool for demonstration purposes[4].

To recap, when estimating the rotation periods of active stars, one cannot assume strictly periodic behaviour. Differential rotation leads to quasi-periodicity in the observed light curves [see e.g. 38]. Moreover, the emergence and decay of active regions, possible existence of active longitudes and their flip-flops can introduce non-stationary behaviour.

[3]For high quality observations, DI can distinguish solar- from anti-solar differential rotation.
[4]https://research.cs.aalto.fi/astroinformatics/starspots/

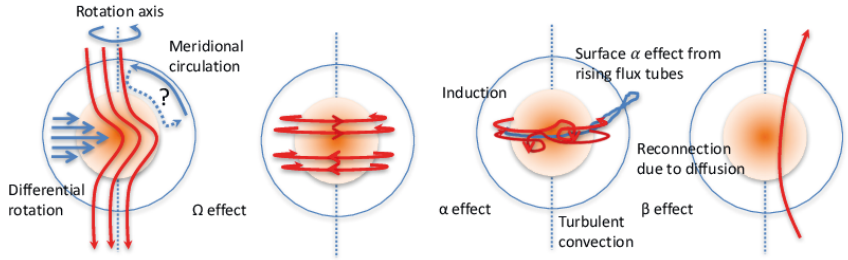


Figure 2.2. Schematic description of the dynamo mechanism. The distributed dynamo relies on the collective inductive action of differential rotation (Ω-effect), generating the toroidal field from the poloidal one, and convective turbulence distributed throughout the CZ (α-effect) generating the poloidal field from the toroidal one (red field lines). Vigorous turbulent diffusion (β-effect) results from the turbulent stirring, and enhances the diffusion coefficients by orders of magnitude above the molecular values. As indicated by helioseismology, there is also a meridional circulation pattern of unknown depth, that affects the dynamo (leftmost panel). Active region formation near the surface is observed, but the mechanism is unknown. One scenario is through thin rising flux tubes (third panel) that may cause α-effect like induction, called as the Babcock-Leighton mechanism. Figure courtesy of M. J. Käpylä.

2.3 Dynamo mechanism – linking rotation and magnetism

The physical mechanism maintaining the large-scale magnetic field is a nonlinear dynamo operating in the CZ of the star. Very generally speaking, the dynamo is a net result of both differential rotation and rotationally affected, in other words, anisotropic, turbulent convection. This is illustrated in Fig. 2.2. Both the relevant induction effects are sensitive to the rotation rate of the object. The α-effect should increase strongly, while Ω-effect should increase more mildly according to theoretical expectations. Hence the vigour of the dynamo mechanism depends on rotation, but also the properties of the excited dynamo modes do so. For example, in some parameter regimes the equatorial symmetry, dynamo period, and axial symmetry can be systematically different. For an introduction and overview of solar dynamo and magnetism, see, for instance, Stix [136], Solanki et al. [133].

To better understand the solar and stellar dynamos, fully 3D global magnetohydrodynamical simulations can nowadays be carried out. This is a more advanced approach to the so-called mean-field dynamo models, where only the evolution of the mean magnetic fields is solved for, most commonly decoupled from the dynamical equations. Moreover, all turbulent effects are unresolved, hence they are parametrised. These models are computationally cheap, and easily allow for solar-like oscillatory solutions. However, oscillatory dynamo modes have only recently been

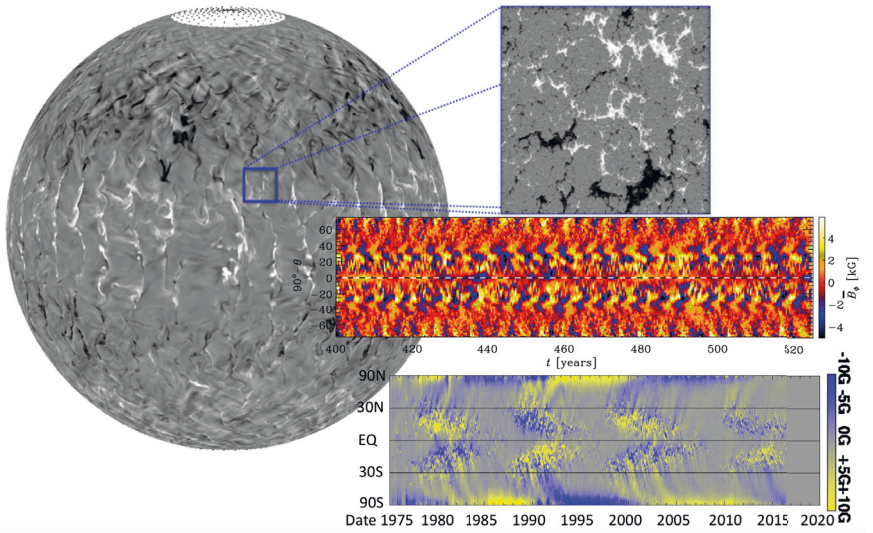


Figure 2.3. Left and upper right panels: Local and global direct numerical simulations of turbulent convection producing a solar-like dynamo solution (black is negative, white is positive radial magnetic field) [72]. Middle panel: The obtained butterfly (time-latitude) diagram from the global model [70]. The time-scales in the simulation do not yet match those in the real Sun. The blue rectangle shows the size of the local computational domain on the solar surface. Image credit: M. J. Käpylä and P. J. Käpylä. Lower right panel: A butterfly diagram from observed magnetograms. Image credit: D. H. Hathaway, NASA/ARC 2016/2017.

found from the much more challenging global convection models. To exemplify these solutions, we have shown excerpts from some of the simulations with the comparison to the magnetic butterfly diagram of the Sun in Fig. 2.3.

Thus far we have discussed two types of quantities associated with active stars: the rotation period and the cycle period[5]. In the following we clarify, why it is important to accurately know both of these quantities, motivating at the same time our work. The crucial aspect is that the stellar magnetic activity is linked to the rotation speed, if the dynamo really operates in the way we expect. Instead of randomly distributed cycle periods one expects to find systematic dependencies on rotation rate.

To observationally study this link, one needs to take into account other well-known dependencies. It is important to note, that the magnetic field strength is related to the intensity of certain spectral lines of a star. For instance, there is a linear correlation between the Ca II H&K line core excess flux density and the absolute value of the magnetic flux density [126]. For this reason, stellar magnetic activity is usually related to the follow-

[5]We use the term 'cycle length' interchangeably with 'cycle period'.

ing quantity: $R'_{\rm HK}$, which is the chomospheric Ca II H&K flux normalised to the bolometric flux. Another quantity empirically known to correlate with magnetic activity is the Rossby number $\rm Ro$, which is defined as the ratio of the rotation period $P_{\rm rot}$ of the star to the convective turnover time τ_c (the latter being an empirical function of the colour index B-V of the star). The relation between $R'_{\rm HK}$ and $\rm Ro$ is, however, not fully established. In addition, there is a dependence between the fractional X-ray luminosity and the rotation period [158, 157]. Furthermore, the spottedness of the late-type stars is found to occur when the Rossby number is lower than about $2/3$. The differential rotation is also known to correlate with the rotation period [57].

Previous studies have postulated the existence of activity populations (clusters) on different diagrams reflecting the above dependencies. The diagrams show the relation usually between (a) $\log P_{\rm rot}/P_{\rm cyc}$ and $\log\langle R'_{\rm HK}\rangle$, (b) $\log P_{\rm rot}/P_{\rm cyc}$ and $\log {\rm Ro}^{-1}$, (c) $\log P_{\rm cyc}$ and $\log P_{\rm rot}$[6], where $P_{\rm rot}$ and $P_{\rm cyc}$ are the rotation and cycle period and $\langle R'_{\rm HK}\rangle$ is the average stellar chromospheric activity index. Two separate populations with positive correlation and later a third population with negative correlation were reported in Brandenburg et al. [24] and Saar and Brandenburg [121] investigating diagrams (a) and (b) respectively. The first two populations have been named 'inactive' and 'active', and the last one 'super-active'. Similarly, studies involving diagrams of type (c) have led to findings of distinct populations [98, 20, 94, 21]. However, the debate over the existence of the populations still continues as there have also been studies reporting non-detections of them [see e.g. 37, 116]. The advantage of diagrams (a) over the other ones is due to the elimination of additional dependence of the quantities on the poorly known τ_c.

Explanations of the above mentioned empirical findings are being sought from the dynamo theory. Böhm-Vitense [20] proposed that the dynamo type would abruptly change from an interface dynamo working near the surface of the CZs for the stars in the active population to the one working at the bottom of the CZs with deeper mixing for the stars in the inactive populations. However, this hypothesis, hinting at the cycle length dependence on the CZ depth, has failed to be confirmed [25]. Another plausible

[6] In Chapter 6 we refer to diagrams (a) and (c) as RCRA and CR diagrams, respectively.

explanation could be the change from nonaxisymmetric to axisymmetric dynamo mode [88], which we set out to study in Paper VI.

2.4 Recapitulation of the aims

In the chain of analysis, different factors contribute to the final result, such as the observational errors in the measurements, statistical methods used, as well as different assumptions about the physical model. Over the years, methods suitable for harmonic time series have been used to estimate the cycle length. However, due to the nonlinearity of the system of partial differential equations governing the stellar dynamo, the solutions are not fully harmonic, but either periodic [162] or more likely quasi-periodic [68, 70]. Both, observations and more recent global magnetohydrodynamical simulations have confirmed this behaviour (see Fig. 2.3). These facts stress that methods primarily suitable for quasi-periodic time series can only be considered to be accurate and reliable for stellar cycle estimation. Due to the differential rotation and finite lifetime of the starspots, similar statement holds for the methods used for the rotation period estimation. This brings us to our main aim in this dissertation, which is to develop and test the methods for period estimation of the quasi-periodic processes. Before focusing on this, we first discuss the basics of the period estimation of harmonic and periodic processes, as this is an important step in understanding many of the aspects which are also relevant in the more general case.

3. Methods for periodic time series

Now we turn to the topic of period estimation from a time series of a periodic process, which is a special case of a stochastic process. We will denote a stochastic process by $y(t)$, where t is time and y is a time dependent random variable. We will mostly discuss the case where y is a real valued scalar.

Useful quantities for describing the stochastic process are its mean and autocovariance[1] functions:

$$\mu(t) = \mathbb{E}\big(y(t)\big), \tag{3.1}$$

$$\operatorname{cov}(t_1, t_2) = \mathbb{E}\big(y(t_1) - \mu(t_1)(y(t_2) - \mu(t_2)\big), \tag{3.2}$$

where $\mathbb{E}$ denotes the expected value (for a definition see Eq. (3.63) later on). The variance of the process at time t is $\sigma^2(t) = \operatorname{cov}(t, t)$. We often use a following uncentred version of the covariance

$$r(t_1, t_2) = \mathbb{E}\big(y(t_1)y(t_2)\big), \tag{3.3}$$

which coincides with the covariance for a zero mean process. In signal processing r is usually also called a covariance. If the process is Gaussian, then the mean and covariance functions fully define the process, otherwise higher-order covariance functions need to be defined.

In this and the next chapter we assume stochastic processes to be wide (or weak) sense stationary (WSS), meaning that the mean is independent of time and the covariance is not depending on exact time moments, but only on the time difference between two time moments $\tau = t_2 - t_1$, which

[1] For convenience we will mostly write just 'covariance' instead of 'autocovariance'.

is usually called a time lag. Then

$$\begin{aligned} \mathrm{cov}(\tau) &= \mathbb{E}\left[\big(y(t)-\mu\big)\big(y(t+\tau)-\mu\big)\right], \\ r(\tau) &= \mathbb{E}\big(y(t)y(t+\tau)\big). \end{aligned} \tag{3.4}$$

$\mathrm{cov}(\tau)$ is an even function, so that $\mathrm{cov}(\tau) = \mathrm{cov}(-\tau)$. The same holds for $r(\tau)$. So, for a WWS process $\mu(t) = \mu$ and $\sigma^2(t) = \sigma^2$.

In a more general case, for non-Gaussian processes one needs to define stationarity through similar properties of higher-order covariance functions. In a more general case the meaning of stationarity can be different. We will turn to it in Chapter 5.

Besides stationarity, in practice often ergodicity is assumed, meaning that the statistical properties of the process can be deduced from a single, sufficiently long, realization of the process. Likewise for the stationarity there are different possible definitions for the ergodicity. For a 2nd order ergodic WSS process

$$\begin{aligned} \mu &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} y(t)dt \\ \mathrm{cov}(\tau) &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \big(y(t)-\mu\big)\big(y(t+\tau)-\mu\big)dt, \\ r(\tau) &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} y(t)y(t+\tau)dt, \end{aligned} \tag{3.5}$$

which means that the expectations can be replaced by the time averages. Similar identities hold for the discrete processes.

Depending on the context, in the following we use the term 'signal' either when referring to a concrete realization of the process or when referring to a component of the process which does not include the observational noise. We do not make differences in the notions of a signal and a process, thus writing $y(t)$ for both. When using the term 'time series', we mean a finite and discrete set of N samples from the realization of the process. We refer to the sequence of measurements as $\{y_n\}$, where $y_n = y(t_n)$ is the measurement at time $t_n, n = 1, \ldots, N$. Often it is more compact to use vector notation, so we also define the column vector of measurements as $\mathbf{y} = [y_1 \ldots y_N]^\mathsf{T}$. Sometimes we refer to time series just as data or dataset.

3.1 Fourier transform and the power spectrum

Under suitable conditions the signal $y(t)$ can be expanded in terms of harmonics of different frequencies f:

$$y(t) = \int_{-\infty}^{\infty} Y(f) e^{2\pi i f t} df, \tag{3.6}$$

where $Y(f)$ is called the Fourier spectrum of the signal $y(t)$. We see that $y(t)$ is a linear superposition of simple waveforms $e^{2\pi i f t}$ with frequencies f and relative contributions given by coefficients $Y(f)$. The latter are obtained from the signal using

$$Y(f) = \int_{-\infty}^{\infty} y(t) e^{-2\pi i f t} dt, \tag{3.7}$$

which is called the Fourier transform (FT) of the signal. The operation of Fourier transform is often denoted by $Y = \mathcal{F}(y)$.

Rayleigh's theorem states that for signals with finite energy

$$E = \int_{-\infty}^{\infty} |y(t)|^2 dt = \int_{-\infty}^{\infty} |Y(f)|^2 df, \tag{3.8}$$

where E is called the total energy of the signal. In the latter equation $|Y(f)|^2$ is sometimes called the energy spectral density of the signal.

In case of stationary processes the signals have infinite energy and the integrals in Eq. (3.8) do not exist. Neither may exist the FT in Eq. (3.7). For that purpose it is practical to consider the truncated Fourier transform

$$Y_T(f) = \int_{-T}^{T} y(t) e^{-2\pi i f t} dt \tag{3.9}$$

and define the so-called power spectrum (also known as spectral density) as

$$P(f) = \lim_{T\to\infty} \frac{1}{2T} \mathbb{E}(|Y_T(f)|^2) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \int_{-T}^{T} r(t,t') e^{2\pi i (t-t')} dt dt'. \tag{3.10}$$

$P(f)$ is useful for assessing the spectral content of stationary processes. In particular, for WSS process, Wiener-Khinchin theorem states, that the

power spectrum and the covariance function form a Fourier pair:

$$\begin{aligned} P(f) &= \int_{-\infty}^{\infty} r(\tau) e^{-2\pi i f \tau} d\tau, \\ r(\tau) &= \int_{-\infty}^{\infty} P(f) e^{2\pi i \tau f} df. \end{aligned} \tag{3.11}$$

where $r(\tau)$ is defined in Eq. (3.3) and f is the frequency. $P(f)$ is a non-negative real and even function of f. For a zero mean process one can replace $r(\tau)$ with $\mathrm{cov}(\tau)$.

If we assume the WSS process to be ergodic, then from the following informal proof we see that for a reasonably long realization of the process, we can calculate an approximation to $P(f)$ using the modulus squared of the truncated Fourier transform $Y_T(f)$:

$$\begin{aligned} &\lim_{T\to\infty} \frac{|Y_T(f)|^2}{2T} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} y(t_1) e^{2\pi i f t_1} dt_1 \int_{-T}^{T} y(t_2) e^{-2\pi i f t_2} dt_2 \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \int_{-T}^{T} y(t_1) y(t_2) e^{2\pi i f (t_1 - t_2)} dt_1 dt_2 \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \int_{-(T-|\tau|)}^{T-|\tau|} y(t) y(t+\tau) e^{-2\pi i f \tau} dt d\tau \\ &= \int_{-\infty}^{\infty} \mathbb{E}\left[y(t) y(t+\tau)\right] e^{-2\pi i f \tau} d\tau = \int_{-\infty}^{\infty} r(\tau) e^{-2\pi i f \tau} d\tau = P(f). \end{aligned} \tag{3.12}$$

3.2 Periodic processes

For the periodic WSS process one can assume that $\mathrm{cov}(t+P) = \mathrm{cov}(t)$, where P is the period of the process. This is equivalent with the statement that for each realization of the process $y(t+P) = y(t)$. In a more general case, one calls the process as mean square periodic, if $\mathbb{E}[(y(t+P) - y(t))^2] = 0$, which, for the WSS process coincides with the previous statements.

In practice the observations are usually contaminated by noise, so that the true signal can not be directly measured. The resulting process then consists of two components, the signal $s(t)$ and noise $\epsilon(t)$. In many situations one assumes that the noise is an independent and identically distributed random variable. Thus we may drop t from the arguments of ϵ. Furthermore, usually ϵ is assumed to be Gaussian with zero mean.

One possible example of a covariance function of a periodic process with

additive Gaussian white noise is:

$$\operatorname{cov}(t_1, t_2) = \sigma_{\mathrm{s}}^2 \cos\big(2\pi(t_1 - t_2)\big) + \sigma_{\mathrm{n}}^2 \mathbb{I}(t_1 = t_2), \tag{3.13}$$

where σ_{s}^2 and σ_{n}^2 are the variances of the periodic signal and the noise components respectively, and $\mathbb{I}$ is the indicator function:

$$\mathbb{I}(\text{condition}) = \begin{cases} 1, \text{if condition is true} \\ 0 \text{ otherwise.} \end{cases} \tag{3.14}$$

3.3 Estimating the power spectrum

Spectral analysis of the WSS periodic processes is motivated by the fact that all periodic functions can be decomposed into the Fourier series. As $\operatorname{cov}(\tau)$ is periodic we can write:

$$\begin{aligned} \operatorname{cov}(\tau) &= \sum_{n=-\infty}^{n=\infty} \gamma_n e^{2\pi i n f t}, \\ \gamma_n &= \frac{1}{P} \int_0^P \operatorname{cov}(\tau) e^{-2\pi i n f t} d\tau, \end{aligned} \tag{3.15}$$

where $P = 1/f$ is the period of the process. This means that the power spectrum of the periodic process consists of a set of decaying delta function-like peaks which are integer multiples of the base frequency f. Similarly, if one decomposes the process $y(t)$ into the Fourier series

$$\begin{aligned} \hat{y}(t) &= \sum_{-\infty}^{\infty} c_n e^{2\pi i n f t}, \\ c_n &= \frac{1}{P} \int_0^P y(t) e^{-2\pi i n f t} dt, \end{aligned} \tag{3.16}$$

it can be shown that the coefficients c_n are uncorrelated random variables, $\mathbb{E}(c_n c_m^*) = \gamma_n \delta_{nm}$. Moreover $\mathbb{E}\big(|y(t) - \hat{y}(t)|^2\big) = 0$ [102].

Now we turn to the description of practical methods developed for estimating the power spectrum. As we saw above, for periodic processes this directly leads to the estimation of the period as one is expecting the highest peak in the spectrum to correspond to the true period. In particular, for the harmonic process with frequency f the power spectrum consists of

two delta function-like peaks one at $-f$ and other at f.

The spectral estimation methods can be generally speaking divided into two categories, parametric and non-parametric ones. The parametric methods assume the spectral density, or equivalently a covariance function in a case of WSS process, to have a certain functional form with respect to period and other parameters. The task is then to find an optimal solution for these parameters. Examples of parametric methods include, for instance, autoregressive, moving average, autoregressive moving average models, as well as maximum entropy spectral estimation. Non-parametric methods on the other hand try to estimate the power spectrum directly from data without postulating the form of covariance explicitly. However, the distinction between parametric and non-parametric methods is somewhat arbitrary. For instance, many non-parametric estimation methods are based on linear least squares regression of harmonic functions. This, however, directly leads to an assumption that the covariance function is also harmonic. Nevertheless, subsequently we primarily consider the methods mostly known as non-parametric methods, with some exceptions later on, like Gaussian processes.

3.3.1 Evenly sampled time series

So far we have been dealing with continuous infinite signals, but in practice, all the measured signals are finite and discrete. It is important to note, that one of the most relevant questions in time series analysis is how to estimate the spectrum from a finite set of observations, in such a way that the estimate would not be biased, it would be consistent and also statistically meaningful. Moreover, the key is to guarantee that the estimation method maintains these properties in the presence of minor variations of assumptions [139]. We do not, however, go into the details of these theoretical questions, but continue with the discussion of several methods that have been developed for the spectral estimation purposes in the discrete case.

First, suppose that we have an evenly sampled set of observations, so that n-th observation at time moment $t_n = n\Delta t$ is $y_n = y(t_n)$, where Δt is the time step and $n = 0 \dots N-1$. The discrete Fourier transform (DFT) is

defined as

$$Y_k = \sum_{n=0}^{N-1} y_n e^{-2\pi i k n/N} \tag{3.17}$$

with the k-th frequency given by $f_k = k/(\Delta t N)$. The data can be recovered via the inverse transform:

$$y_n = \frac{1}{N}\sum_{k=0}^{N-1} Y_k e^{2\pi i k n/N} = \frac{1}{N}\sum_{k=-N/2+1}^{N/2} Y_k e^{2\pi i k n/N}. \tag{3.18}$$

The latter formulas are equivalent as Y_k is periodic with the period N, thus, the summation over $k = 0 \dots N-1$ gives the same answer as over $-N/2 < k \leq N/2$. For real valued data, the values of Y_k for negative k are determined by the values for $0 \leq k \leq N/2$. [19, Ch. 4.2]. Analogue of the Rayleigh's theorem in Eq. (3.8) for an evenly sampled case is the Parseval's theorem, which states:

$$\sum_{n=0}^{N-1} |y_n|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |Y_k|^2. \tag{3.19}$$

According to Nyquist-Shannon theorem, for any function, for which the FT is zero for frequencies above f_N, it is fully determined by the values at equally spaced series of points with step $\Delta t = 1/2f_\mathrm{N}$. Conversely

$$f_\mathrm{N} = \frac{1}{2\Delta t}, \tag{3.20}$$

where f_N is called the Nyquist frequency for the given sampling interval Δt.

In the discrete case Eq. (3.11) has the following analogue:

$$\begin{aligned} P(f) &= \sum_{n=-\infty}^{\infty} r_n e^{-2\pi i f \tau_n}, \quad -f_\mathrm{N} \leq f \leq f_\mathrm{N}, \\ r_n &= \frac{1}{2f_\mathrm{N}} \int_{-f_\mathrm{N}}^{f_\mathrm{N}} P(f) e^{2\pi i f \tau_n} df, \end{aligned} \tag{3.21}$$

where $r_n = \mathbb{E}[y(t_k)y(t_k+\tau_n)]$, $\tau_n = n\Delta t$ and f_N is the Nyquist frequency from Eq. (3.20).

Led by the analogy of the continuous case, the simplest idea to estimate the power spectrum in the discrete case would be to calculate the modulus squared of the DFT and normalise it by the number of data points. This

is called the periodogram:

$$\hat{P}(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} y_n e^{-2\pi i t_n f}\right|^2, \qquad -f_{\rm N} < f \leq f_{\rm N}. \tag{3.22}$$

This equation is equivalent to

$$\hat{P}(f) = \sum_{n=-(N-1)}^{N-1} \hat{r}_n e^{-2\pi i f \tau_n}, \tag{3.23}$$

where $\hat{r}_n$ is the biased sample covariance defined as

$$\hat{r}_k = \frac{1}{N}\sum_{l=0}^{N-k-1}(y_{k+l}-\hat{\mu})(y_l-\hat{\mu}), \quad 0 \leq |k| \leq N-1, \tag{3.24}$$

where $\hat{\mu} = \frac{1}{N}\sum_{n=0}^{N-1} y_n$ is the sample mean [109, Ch. 4.2]. Unbiased sample covariance is defined as $\tilde{r}_k = \frac{N}{N-k}\hat{r}_k$. The sample covariances are both symmetric, that is, $\tilde{r}_k = \tilde{r}_{-k}$ and $\hat{r}_k = \hat{r}_{-k}$. These are the discrete analogues of identities given in Eq. (3.5). Eq. (3.23) can be seen as an empirical way of calculating $P(f)$ in Eq. (3.21) for a finite evenly sampled time series. The name 'periodogram' was first used by Arthur Schuster [128][2].

When the frequencies are selected as $f_k = k/(\Delta t N) = 2k f_{\rm N}/N$, then the periodogram matches the modulus squared of the DFT: $\hat{P}_k = \hat{P}(f_k) = |Y_k|^2/N$ from Eq. (3.17). When $\{y_n\}$ are the realizations of the zero mean Gaussian white noise stationary process with standard deviation $\sigma_{\rm n}$, then $\hat{P}(f_k)/\sigma_{\rm n}^2$ is a χ^2 random variable. Moreover, when $k \neq k'$ ($k, k' = 0, 1 \ldots N/2$), then $\hat{P}_k$ and $\hat{P}_{k'}$ are independent [153]. When $N_\perp$ is the number of these independent frequencies then the probability that at least one of the $\hat{P}_k$ is expected to exceed level z_0 is $p_0 = 1 - (1 - e^{-\hat{P}_k/\sigma_{\rm n}^2})^{N_\perp}$. Conversely

$$z_0 = -\sigma_{\rm n}^2 \ln\left(1 - (1-p_0)^{1/N_\perp}\right). \tag{3.25}$$

z_0 is called the level of false alarm and p_0 the false alarm probability (FAP) [153].

There is a one-to-one correspondence between DFT coefficients Y_k and the amplitudes a_k and b_k of the sine and cosine components with frequencies f_k fit to the data, that is, $\hat{P}_k = a_k^2 + b_k^2$ ($k = 0 \ldots N-1$). This corre-

[2]From now on we will refer to the periodogram defined in the current section as classical periodogram. Otherwise we will use the term 'periodogram' more loosely when referring to any kind of proxy for the power spectrum.

spondence does not hold in the case of uneven sampling [152].

Despite the simplicity of the periodogram, the method suffers from severe deficiencies. It is an inconsistent estimator, meaning that it does not converge to a true spectral density as $N \to \infty$. While it is asymptotically unbiased, its variance does not tend to zero as $N \to \infty$ [60, Ch. 8.2.2]. It also exhibits very high spectral leakage, although this can be reduced by multiplying $\{y_n\}$ by a window function. Leakage means that power is transferred from the frequencies that concentrate most of the power, to the frequencies that contain less or no power.

Different methods have been developed to overcome the problems of the classical periodogram. For instance in Bartlett's method, the spectra is calculated for smaller windows of data and later averaged [14]. In multitaper method, the signal is first multiplied by a prolate spherical sequence (or Slepian sequences) of different order. The FTs of these windowed signals are then calculated, which are called prolate eigenspectra. Finally the estimate for the power spectrum is calculated as the weighted average over the absolute squares of these eigenspectra. Since the prolate spherical sequences are orthonormal, the eigenspectra are statistically independent estimates from the same sample and the estimation bias of the power spectrum is thus reduced [139].

3.3.2 Unevenly sampled time series

The effect of data sampling on the spectrum

The observational time series are often unevenly sampled – there may be missing data, seasonal gaps and so forth. The general question is, how to estimate $Y(f)$, which depends on an integral of $y(t)$ over the range $(-\infty, \infty)$, from observations in a limited section $(0, T)$ at discrete time moments $\{t_k\}$. It can be shown that the observed spectrum in this case is equal to the true spectrum convolved with a spectral window (or a window function). Following Deeming [36], if one defines an analogue of DFT for irregular sampling

$$Y_N(f) = \sum_{k=1}^{N} y(t_k) e^{-2\pi i f t_k}, \tag{3.26}$$

then

$$Y_N(f) = Y(f) * W_N(f) = \int_{-\infty}^{\infty} Y(f - f') W_N(f') df', \tag{3.27}$$

where $*$ denotes convolution, $Y(f)$ is the actual FT and $W_N(f)$ is the spectral window defined by

$$W_N(f) = \sum_{k=1}^{N} e^{-2\pi i f t_k} = \sum_{k=1}^{N} \mathcal{F}\big(\delta(t - t_k)\big). \tag{3.28}$$

The spectral window is a Fourier transform of the so-called data window, which depends only on the data sampling and interval.

For the power spectrum of the WSS process similar relation holds:

$$P_N(f) = \mathbb{E}\big(Y_N(f) Y_N^*(f)\big) = P(f) * V_N(f), \tag{3.29}$$

where

$$V_N(f) = \sum_{j,k} e^{2\pi i f (t_j - t_k)} = W_N(f) W_N^*(f) \tag{3.30}$$

is the power spectral window and x^* denotes the complex conjugate of x. The effect of the spectral window is usually called aliasing (or spectral leakage, also referred to as spurious periods in observational astronomy). For instance, if the spectral window has large amplitude at some frequency f_0 far away from zero, then the empirical spectra $Y_N(f)$ and $P_N(f)$ have significant contributions coming from the true spectra $Y(f - f_0)$ and $P(f - f_0)$, respectively.

In the case of even sampling the spectral window can be expressed analytically in a form $W_n(f) = \frac{\sin \pi f N \Delta t}{N \sin \pi f \Delta t}$, which is approximately an infinite row of sinc functions spaced $\Delta t^{-1} = 2 f_{\mathrm{N}}$ apart. If $N \to \infty$, then $W_N(f)$ becomes a Dirac comb. As in this case $W_N(f) = W_N(f + n \Delta t^{-1})$, two frequencies f_1 and f_2 are indistinguishable from each other if $f_2 = f_1 \pm n \Delta t^{-1}$. In other words the aliasing is complete [36].

Lomb-Scargle periodogram

One could generalize the Eq. (3.22) to unevenly sampled data, but it turns out that the statistical properties of the estimator would not stay the same. Neither would the periodogram correspond any more to the least squares fit of harmonic functions [153].

However, when the input is a Gaussian white noise, a slightly modified periodogram has the same exponential probability distribution as for even

spacing:

$$P_{\mathrm{LS}}(f) = \frac{1}{2}\left[\frac{\left(\sum_{i=1}^{N} y(t_i)\cos 2\pi f(t_i-\tau)\right)^2}{\sum_{i=1}^{N}\cos^2 2\pi f(t_i-\tau)} + \frac{\left(\sum_{i=1}^{N} y(t_i)\sin 2\pi f(t_i-\tau)\right)^2}{\sum_{i=1}^{N}\sin^2 2\pi f(t_i-\tau)}\right], \tag{3.31}$$

where τ is defined by

$$\tan(4\pi f\tau) = \frac{\sum_{i=1}^{N}\sin 4\pi f t_i}{\sum_{i=1}^{N}\cos 4\pi f t_i}. \tag{3.32}$$

An estimate for the power spectrum formulated in Eq. (3.31) is known as the Lomb-Scargle (LS) periodogram and it is equivalent to the reduction in the sum of squares (RSS) of the least squares fit of harmonic functions to the data [91, 125] (more on that in Sect. 3.3.3). Introduction of τ in Eq. (3.32) makes the LS periodogram invariant under time translations as well as both sine and cosine components of the model orthogonal on the domain of input time moments $\{t_i\}$ [125]. Furthermore, in the case of even sampling, $P(f_n)$ are independent random variables for frequencies

$$f_n = n/T = 2nf_{\mathrm{N}}/N, \quad n = -N/2, \ldots, N/2, \tag{3.33}$$

where $T = N\Delta t$ is the total time interval of the data. f_n are called natural (or standard) frequencies since the DFT evaluated at these frequencies contains just enough information to recover the original data. $P(f)$ at intermediate frequencies, however, are dependent variables [125]. In the unevenly sampled case, when the window function has evenly spaced nulls at frequencies $f_n = nf_1$, those $P(f_n)$ are uncorrelated (for Gaussian processes this yields also independence). With wide variety of sampling schemes the window function has nulls, or relatively small minima, that are approximately evenly sampled. Such nulls comprise a set of natural frequencies at which to evaluate the periodogram [125].

Calculating the LS periodogram has computational complexity $\mathcal{O}(10^2N^2)$, which can be reduced to $\mathcal{O}(10^2N\log N)$ by mapping the data onto an evenly sampled grid and employing the fast Fourier transform [110]. It is also worth noting that for large N the difference from the classical periodogram becomes marginal [152] and LS periodogram is still an inconsistent statis-

tic (for asymptotic properties of sample spectral density see e.g. Anderson [1, Ch. 8.4]).

Generalizations of the LS periodogram

The LS periodogram is valid for a model with zero mean. Commonly, this condition is satisfied by subtracting the mean of the data before calculating the periodogram. However, if the empirical mean of the data significantly differs from the true mean (e.g. in the case of 'pathological' sampling), this can lead to biased period estimates. Secondly, the LS periodogram assumes that the noise in the data is constant. The generalized LS periodogram (GLS) solves both of these problems, by explicitly solving for the mean [163], while the weights are allowed to be dependent on time moments.

In fact a more complex model allowing to fit secular trends together with harmonics to data was introduced much earlier by Vaníček [146]. This method is more loose from LS periodogram also from another perspective, namely that the components of the model are not assumed to be orthogonal, but linearly independent on $\{t_n\}$. In this approach the optimal spectrum is shown to be invariant to any transform of the signal $y(t) \to y(t) + P_m(t)$, where P_m is the sum of the known components of the model. This allows to suppress systematic noise of the form P_m as well as already known peaks in the spectrum [146].

Ferraz-Mello [43] used a Gram-Schmidt orthogonalization procedure on sine, cosine and constant function, before fitting them to the data. Likewise in GLS method, here the weights are not restricted to be constant. One common assumption to all of these methods is that the noise is Gaussian and white (uncorrelated).

3.3.3 Power spectrum estimation as a projection

The LS periodogram can be viewed from two different angles: either as a direct way of approximating the power spectrum or as a least squares fit of the harmonic functions to the data. However, there is also a third angle, originating from the fact that the least squares fit can be understood as a projection.

The N-dimensional space defined by observational time moments $\{t_n\}$ is usually called a sampling space, where the vector of time moments of

the observations and the vector of observations are $\mathbf{t} = [t_1 \dots t_N]^\mathsf{T}$ and $\mathbf{y} = [y_1 \dots y_N]^\mathsf{T}$, respectively. The scalar product of two functions g and h in the sampling space is defined as:

$$\mathbf{g}^\mathsf{T}\mathbf{h} = \sum_{i=1}^{N} g(t_i)h(t_i), \tag{3.34}$$

where $\mathbf{g} = g(\mathbf{t}) = [g(t_1) \dots g(t_N)]^\mathsf{T}$ and similarly for $\mathbf{h}$. Two functions are considered to be orthogonal, if $\mathbf{g}^\mathsf{T}\mathbf{h} = 0$. This notation is useful, as when using quadratic loss function of the fit of K ($K \leq N$) orthogonal functions to the data, one has:

$$L(\mathbf{w}) = \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{K} w_j g_j(t_i) \right)^2 = \mathbf{y}^\mathsf{T}\mathbf{y} + \sum_{j=1}^{K} (w_j^2 \mathbf{g}_j^\mathsf{T}\mathbf{g}_j - 2w_j \mathbf{g}_j^\mathsf{T}\mathbf{y}), \tag{3.35}$$

where $\mathbf{w} = [w_1 \dots w_K]^\mathsf{T}$ is the weight vector. By differentiating and equating to zero one finds:

$$2w_j \mathbf{g}_j^\mathsf{T}\mathbf{g}_j - 2\mathbf{g}_j^\mathsf{T}\mathbf{y} = 0 \Rightarrow w_j = \frac{\mathbf{g}_j^\mathsf{T}\mathbf{y}}{\mathbf{g}_j^\mathsf{T}\mathbf{g}_j}. \tag{3.36}$$

If we write vector $\mathbf{y}$ consisting of two components, one laying in the subspace spanned by vectors $\mathbf{g}_j$[3] and the other perpendicular to it, that is,

$$\mathbf{y} = \sum_{j=1}^{K} a_j \mathbf{g}_j + \mathbf{r}, \tag{3.37}$$

by substituting $\mathbf{y}$ to Eq. (3.36) we see that $w_j = a_j$, meaning that w_j are just the coefficients of projections of $\mathbf{y}$ on $\mathbf{g}_j$. Substituting $\mathbf{w}$ to Eq. (3.35) we obtain:

$$\mathrm{RSS} = \mathbf{y}^\mathsf{T}\mathbf{y} - L = \sum_{j=1}^{K} \frac{(\mathbf{g}_j^\mathsf{T}\mathbf{y})^2}{\mathbf{g}_j^\mathsf{T}\mathbf{g}_j}. \tag{3.38}$$

From Eqs. (3.37) and (3.36) we see that $\mathbf{y}^\mathsf{T}\mathbf{y} - \mathbf{r}^\mathsf{T}\mathbf{r} = \sum_{j=1}^{K} a_j^2 \mathbf{g}_j^\mathsf{T}\mathbf{g}_j = \mathrm{RSS}$. This means that the sum of the squared projections of the data on the basis functions is identical to the RSS from the least squares fit. For the period estimation, the idea is to choose the harmonic basis functions. In a special case of orthogonalised cosine and sine functions: $\mathbf{f}_1 = \mathbf{c}$ and

[3]This K-dimensional subspace of the sampling space is usually called a model space.

$\mathbf{f}_2 = \mathbf{s}$, whose elements are $c_i = \cos(2\pi f(t_i - \tau))$ and $s_i = \sin(2\pi f(t_i - \tau))$, respectively. We see, that

$$\mathrm{RSS(f)} = \frac{(\mathbf{c}^\mathsf{T}\mathbf{y})^2}{\mathbf{c}^\mathsf{T}\mathbf{c}} + \frac{(\mathbf{s}^\mathsf{T}\mathbf{y})^2}{\mathbf{s}^\mathsf{T}\mathbf{s}} = 2P_{\mathrm{LS}}(f), \tag{3.39}$$

where $P_{\mathrm{LS}}(f)$ corresponds to the LS periodogram defined in Eq. (3.31).

Here we gave a short description of power spectrum estimation from the point of view of projections in a very special case when the basis functions are orthogonal and the weights of the data points are equal. Detailed discussions on this topic can be found in [146, 43, 84, 46, 47, 131].

3.4 Phase dispersion minimisation (PDM) methods

An alternative to period estimation via the power spectrum is a group of methods called phase dispersion (PD) minimisation methods. The idea behind these methods is to first fold the time series with a given trial period to construct a so-called phase diagram and then measure the scatter of the data points on that diagram. In contrast to power spectra, where one looks for maxima to represent the true period, in the case of PD spectra one looks for minima. The second difference between the approaches is that while for the periodic signal the power spectra have one peak for the fundamental frequency f_0 and additional peaks for each overtone $f_n = nf_0$ ($n \in \mathbb{N}$), then PD spectra have one peak for f_0, while additional peaks for each sub-harmonic $f_n = f_0/n$.

Lafler and Kinman [85] introduced a statistic, which measures the squared differences of the adjacent points on the diagram as follows:

$$\Theta = \frac{\sum_{i=1}^{N}(m_i - m_{(i+1) \mod N})^2}{\sum_{i=1}^{N}(m_i - \overline{M})^2}, \tag{3.40}$$

where N is the number of data points, $\overline{M} = \sum_{i=1}^{N} m_i/N$ is the mean of the data points and m_i are the measurements ordered in phase with respect to the trial period P, such that $\varphi_{i+1} > \varphi_i$, where $\varphi_i = \mathrm{frac(t_i/P)}$.

Similar idea was used by Stellingwerf [135], where the statistic measures the squared differences of the data points inside the phase bins to the bin means:

$$\Theta = \frac{s^2}{\sigma^2}, \tag{3.41}$$

where

$$s^2 = \frac{\sum_{j=1}^{M}(n_j - 1)s_j^2}{\sum_{j=1}^{M} n_j - M} \quad \text{and} \quad \sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}. \tag{3.42}$$

Here M is the number of bins, n_j is the number of data points in j-th bin, σ^2 is the total variance, $\overline{x} = \sum_{i=1}^{N} x_i$ is the mean of the data points, $\overline{x}_j = \sum_{i=1}^{n_j} x_{ji}$ is the mean of the data points belonging to bin j, x_{ji} is the i-th data point in the j-th bin and $s_j^2 = \big(\sum_{i=1}^{n_j}(x_{ji} - \overline{x}_j)^2\big)/(n_j - 1)$ is the variance of data in the j-th bin.

A different way to view the PDM methods is to notice that they corresponds to the least squares fit of the orthogonal step functions of the form:

$$\Psi^l(t) = \begin{cases} \frac{1}{n_t}, \text{if } l - 1 \leq r\frac{t}{P} \mod r < l \\ 0 \text{ otherwise}, \end{cases} \tag{3.43}$$

where n_t is the number of data points in the l-th phase bin and r is depending on the method [131]. For instance, in the case of Lafler-Kinman method, $r = n/2$.

3.5 D^2 statistic

Statistic which we call a D^2 statistic essentially belongs to the group of PDM methods, but as it is more general than the others, in the following we dedicate several subsections to describe it. The D^2 method was used in several of applications in our work. In this section we will give a definition of the special case of the statistic suitable for strictly periodic time series and in the next chapter where we discuss quasi-periodicity, we will generalize it.

The D^2 statistic was first introduced by Pelt [104] in the following form:

$$D^2(f) = \frac{N}{2}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} g(t_i, t_j, f)\big(y(t_i) - y(t_j)\big)^2}{\sum_{i=1}^{N}\sum_{j=1}^{N} g(t_i, t_j, f)}, \tag{3.44}$$

where N is the number of data points, $f = 1/P$ is the trial frequency and $g(t_i, t_j, f)$ is the phase selection function, which is symmetric, depending only on $\tau = |t_i - t_j|$. The name D^2 comes from the fact, that the statistic measures the sum of squared differences of the observations weighted by the selection function g. Other functional forms instead of squared differ-

ence could be considered, but we will not discuss these options here. From Eq. (3.44) we see that the D^2 statistic is invariant under both time translations and translations in the signal values. The latter property allows us to assume that the signal has a zero mean without loss of generality. Using shorthands $y_i = y(t_i)$ and $g_{ij} = g(t_i, t_j, f)$, and defining the data vector $\mathbf{y} = [y_1 \dots y_N]^\mathsf{T}$ and selection matrix $\mathbf{G} \in \mathbb{R}^{N\times N}$, whose elements are g_{ij}, one can write (by using the fact that $g_{ij} = g_{ji}$) Eq. (3.44) in a vector form:

$$D^2(f) = N\frac{\mathbf{1}^\mathsf{T}\mathbf{G}(\mathbf{y}\circ\mathbf{y}) - \mathbf{y}^\mathsf{T}\mathbf{G}\mathbf{y}}{\mathbf{1}^\mathsf{T}\mathbf{G}\mathbf{1}} = \frac{N\mu_{\mathbf{G}(\boldsymbol{y}\circ\mathbf{y})} - \boldsymbol{y}^\mathsf{T}\mathbf{G}\boldsymbol{y}}{\mu_{\mathbf{G}}}, \tag{3.45}$$

where $\mathbf{1} \in \mathbb{R}^N$ is the vector of all ones, $\circ$ denotes the Hadamard (or element-wise) product, $\mu_{\mathbf{x}} = \mathbf{1}^\mathsf{T}\mathbf{x}/N$ is the mean of the elements of vector $\mathbf{x}$ and $\mu_{\mathbf{X}} = \mathbf{1}^\mathsf{T}\mathbf{X}\mathbf{1}/N^2$ is the mean of the elements of matrix $\mathbf{X}$.

Now we consider two special cases of the selection function, to reveal better the idea of D^2 statistic.

3.5.1 Cosine selection function

When one uses the following form for the selection function $g(t_i, t_j, f) = 1 + 2\cos\big(2\pi f(t_i - t_j)\big)$, it can be shown that in the case of evenly sampled data the D^2 statistic becomes identical to the LS periodogram [104]. We will repeat this procedure in a slightly different manner here.

In this special case the selection matrix has the form:

$$\mathbf{G} = \mathbf{1}\mathbf{1}^\mathsf{T} + 2\mathbf{c}\mathbf{c}^\mathsf{T} + 2\mathbf{s}\mathbf{s}^\mathsf{T}, \tag{3.46}$$

where $\mathbf{c} \in \mathbb{R}^N$ and $\mathbf{s} \in \mathbb{R}^N$ represent the vectors whose elements are correspondingly $c_i = \cos(2\pi f t_i)$ and $s_i = \sin(2\pi f t_i)$. Now Eq. (3.45) becomes

$$\begin{aligned} D^2(f) =& N\frac{\mathbf{1}^\mathsf{T}(\mathbf{1}\mathbf{1}^\mathsf{T} + 2\mathbf{c}\mathbf{c}^\mathsf{T} + 2\mathbf{s}\mathbf{s}^\mathsf{T})(\mathbf{y}\circ\mathbf{y}) - \mathbf{y}^\mathsf{T}(\mathbf{1}\mathbf{1}^\mathsf{T} + 2\mathbf{c}\mathbf{c}^\mathsf{T} + 2\mathbf{s}\mathbf{s}^\mathsf{T})\mathbf{y}}{\mathbf{1}^\mathsf{T}(\mathbf{1}\mathbf{1}^\mathsf{T} + 2\mathbf{c}\mathbf{c}^\mathsf{T} + 2\mathbf{s}\mathbf{s}^\mathsf{T})\mathbf{1}} \\ =& N\frac{N\mathbf{y}^\mathsf{T}\mathbf{y} - (\mathbf{1}^\mathsf{T}\mathbf{y})^2 + 2\big(\mathbf{1}^\mathsf{T}\mathbf{c}\mathbf{c}^\mathsf{T}(\mathbf{y}\circ\mathbf{y}) + \mathbf{1}^\mathsf{T}\mathbf{s}\mathbf{s}^\mathsf{T}(\mathbf{y}\circ\mathbf{y}) - \mathbf{y}^\mathsf{T}\mathbf{c}\mathbf{c}^\mathsf{T}\mathbf{y} - \mathbf{y}^\mathsf{T}\mathbf{s}\mathbf{s}^\mathsf{T}\mathbf{y}\big)}{N^2 + 2N^2\mu_{\mathbf{c}}^2 + 2N^2\mu_{\mathbf{s}}^2} \\ =& \frac{\mathbf{y}^\mathsf{T}\mathbf{y} - N\mu_{\mathbf{y}}^2 + 2\big(\mu_{\mathbf{c}}\mathbf{c}^\mathsf{T}(\mathbf{y}\circ\mathbf{y}) + \mu_{\mathbf{s}}\mathbf{s}^\mathsf{T}(\mathbf{y}\circ\mathbf{y}) - 1/N(\mathbf{c}^\mathsf{T}\mathbf{y})^2 - 1/N(\mathbf{s}^\mathsf{T}\mathbf{y})^2\big)}{1 + 2\mu_{\mathbf{c}}^2 + 2\mu_{\mathbf{s}}^2}. \end{aligned} \tag{3.47}$$

If our vector of observations $\mathbf{y}$ has zero mean then $\mu_{\mathbf{y}} = 0$. Also, if the

number of data points is large ($N \gg 1$) and we have high enough sampling rate (or at least non-pathological sampling), the terms with $\mu_{\mathbf{c}}$ and $\mu_{\mathbf{s}}$ can be considered small (because $\mu_{\mathbf{s}} = \frac{1}{N}\sum_{i=1}^{N}\sin(2\pi f t_i) \ll 1$ and $\mathbf{s}^{\mathsf{T}}(\mathbf{y}\circ\mathbf{y}) < \mathbf{y}^{\mathsf{T}}\mathbf{y}$, thus these terms are only minor corrections to the first term $\mathbf{y}^{\mathsf{T}}\mathbf{y}$). With these assumptions formula Eq. (3.47) becomes

$$D^2(f) = \mathbf{y}^{\mathsf{T}}\mathbf{y} - \frac{2}{N}(\mathbf{c}^{\mathsf{T}}\mathbf{y})^2 - \frac{2}{N}(\mathbf{s}^{\mathsf{T}}\mathbf{y})^2, \tag{3.48}$$

which is identical to Eq. (3.39) for evenly sampled time series as in that case $\mathbf{c}^{\mathsf{T}}\mathbf{c} \approxeq \mathbf{s}^{\mathsf{T}}\mathbf{s} \approxeq N/2$ (assuming $N \gg 1$ and f is not close to the sampling rate).

3.5.2 Boxcar selection function

Using boxcar function as a selection function, the D^2 statistic coincides more closely with other PDM statistics introduced in Sect. 3.4. In this case

$$g_{ij} = \mathbb{I}\left(\min\big(\Delta\varphi(\mathrm{i},\mathrm{j}), 1 - \Delta\varphi(\mathrm{i},\mathrm{j})\big) < \epsilon\right), \tag{3.49}$$

where $\mathbb{I}$ is the indicator function, $\Delta\varphi(i,j) = \mathrm{frac}(f|t_i - t_j|)$, $f = 1/P$, and ϵ is the phase separation. The intuition behind using the boxcar function is simple: to only allow those data points to enter the statistic, which are closer than ϵ to each other in phase.

Now we want to give another perspective to the statistic from the point of view of autocovariance. Let us assume that the time moments are ordered such, that $t_i > t_j$ if $i > j$ and the sampling is even. Then the matrix $\mathbf{G}$ is a symmetric Toeplitz matrix with diagonals consisting of either fully ones or zeros. If the total number of points is N and the total time range is T then the parameter ϵ defines the width $\lfloor 2W \rfloor$ of the band of diagonals (how many consecutive diagonals with ones there are), where $W = \epsilon NP/T$. Please note that W is not an integer. We further denote with $N_{\mathrm{per}} = \lfloor T/P \rfloor$, the number of full periods in the dataset, and $I = NP/T$, the index difference corresponding to one period (also not an integer). As the approximate number of ones in each row is $2N\epsilon$ and the rows are cyclically shifted versions of each other, the first term in the numerator and

the denominator of Eq. (3.45) become

$$\begin{aligned} \mathbf{1}^\mathsf{T}\mathbf{G}(\mathbf{y} \circ \mathbf{y}) &\approxeq 2N\epsilon\mathbf{y}^\mathsf{T}\mathbf{y}, \\ \mathbf{1}^\mathsf{T}\mathbf{G}\mathbf{1} &\approxeq 2N^2\epsilon. \end{aligned} \tag{3.50}$$

For the second term, we notice that for a diagonal of ones starting at column i in the first row we get the contribution $\sum_j^{N-i} y_j y_{i+j}$. There are on average N_per chunks of such columns, such that on average the k-th chunk starts at column $\lfloor kI \rfloor$ and ends at column $\lfloor kI + 2W \rfloor$. This covers the upper triangle (above the main diagonal) of the matrix. Due to the symmetry, the contribution from the lower triangle is exactly identical, so that the full second term becomes

$$\mathbf{y}^\mathsf{T}\mathbf{G}\mathbf{y} \approxeq 2\left(\sum_{k=0}^{N_\mathrm{per}} \sum_{i=\lfloor kI \rfloor}^{\lfloor kI+2W \rfloor} \sum_{j=1}^{N-i} y_j y_{i+j}\right) - \mathbf{y}^\mathsf{T}\mathbf{y}. \tag{3.51}$$

Subtracting $\mathbf{y}^\mathsf{T}\mathbf{y}$ is needed, because the main diagonal has been counted twice.

In the following we again assume that the signal has a zero mean. Then we see that $\mathbf{y}^\mathsf{T}\mathbf{y} = N\hat{\sigma}^2$ and $\sum_{j=1}^{N-i} y_j y_{i+j} = N\hat{r}_i$, where $\hat{\sigma}^2$ is the sample variance and $\hat{r}_i$ is the biased sample covariance defined in Eq. (3.24). Finally, combining equations (3.45), (3.50) and (3.50) we get

$$D^2(f) \approxeq \left(N + \frac{1}{2\epsilon}\right)\hat{\sigma}^2 - \frac{1}{\epsilon}\sum_{k=0}^{N_\mathrm{per}} \sum_{i=\lfloor kI \rfloor}^{\lfloor kI+2W \rfloor} \hat{r}_i. \tag{3.52}$$

The phase separation ϵ is usually taken as small number, for instance 0.1. If we have a large number of data points, then the first term in the equation becomes approximately equal to $N\hat{\sigma}^2$. However, only the second term is interesting to us from the perspective of period estimation. It basically represents a filtering of the biased sample covariance near the time lags which are integer multiples of the period.

3.5.3 Generalization to multiple dimensions

A nice property of the D^2 statistic is that it is straightforward to generalize it to multidimensional data. Let us consider instead of scalar observations $\{y_i\}$ a multivariate set of observations $\{\mathbf{y}_i\}$ with vector el-

ements $\mathbf{y}_i = [y_1(t_i) \dots y_M(t_i)]^\mathsf{T}$, where M is the number of covariates. Then we can change the squared difference in Eq. (3.44) with the L^2 norm $||\mathbf{a}||^2 = \sum_{i=1}^{M} a_i^2$. D^2 statistic then becomes

$$D^2(f) = \frac{N}{2} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} g(t_i, t_j, f) ||\mathbf{y}(t_i) - \mathbf{y}(t_j)||^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} g(t_i, t_j, f)} \tag{3.53}$$

and the analogue of Eq. (3.45) is

$$D^2(f) = N \frac{\mathbf{1}_N^\mathsf{T} \big(\mathbf{G}(\mathbf{Y} \circ \mathbf{Y}) - \mathbf{Y} \circ (\mathbf{G}\mathbf{Y})\big) \mathbf{1}_M}{\mathbf{1}_N^\mathsf{T} \mathbf{G} \mathbf{1}_N}, \tag{3.54}$$

where the i-th row of $\mathbf{Y} \in \mathbb{R}^{N \times M}$ is now containing vector $\mathbf{y}_i^\mathsf{T}$. Here $\mathbf{1}_N$ and $\mathbf{1}_M$ are all-one vectors of dimensions N and M, respectively. As $\mathbf{1}_N^\mathsf{T} \mathbf{X} \mathbf{1}_M = NM\mu_\mathbf{X}$, where $\mathbf{X} \in \mathbb{R}^{N \times M}$ and $\mu_\mathbf{X}$ is the mean of all elements of the matrix $\mathbf{X}$, then the latter equation can also be written as

$$D^2(f) = N \frac{\mu_{\mathbf{G}(\mathbf{Y} \circ \mathbf{Y})} - \mu_{\mathbf{Y} \circ (\mathbf{G}\mathbf{Y})}}{\mu_\mathbf{G}}. \tag{3.55}$$

Furthermore, when we now denote by $\mathbf{y}_j$ the j-th column of $\mathbf{Y}$, then

$$D^2(f) = \frac{\sum_{j=1}^{M} \left(\mathbf{1}_N^\mathsf{T} \mathbf{G} (\mathbf{y}_j \circ \mathbf{y}_j) - \mathbf{y}_j^\mathsf{T} \mathbf{G} \mathbf{y}_j\right)}{\mu_\mathbf{G}} = \sum_{j=1}^{M} D_j^2(f). \tag{3.56}$$

We see that the multidimensional statistic is identical to summing the statistic calculated for the individual components of the measured vector variable. This is similar to ensemble averaging technique used to reduce the noise in the estimate.

In practical situations it might be that only some of the dimensions carry the periodic signal while the others carry only noise. In that case we could select only those dimensions for which the integral of the D^2 dispersion spectrum over certain frequency range is less than the integral of the average spectrum. This way we can both fine-tune the domain of the oscillations and improve the period estimate.

3.5.4 Generalization to multiple periods

Let us consider a function $h_M(x_1, \dots, x_M)$ of M arguments which is periodic with respect to each argument, that is, $h_M(x_1, \dots, x_m + P_m, \dots, x_M) = h_M(x_1, \dots, x_m, \dots, x_M)$, $m = 1, \dots, M$. Then the multiperiodic function

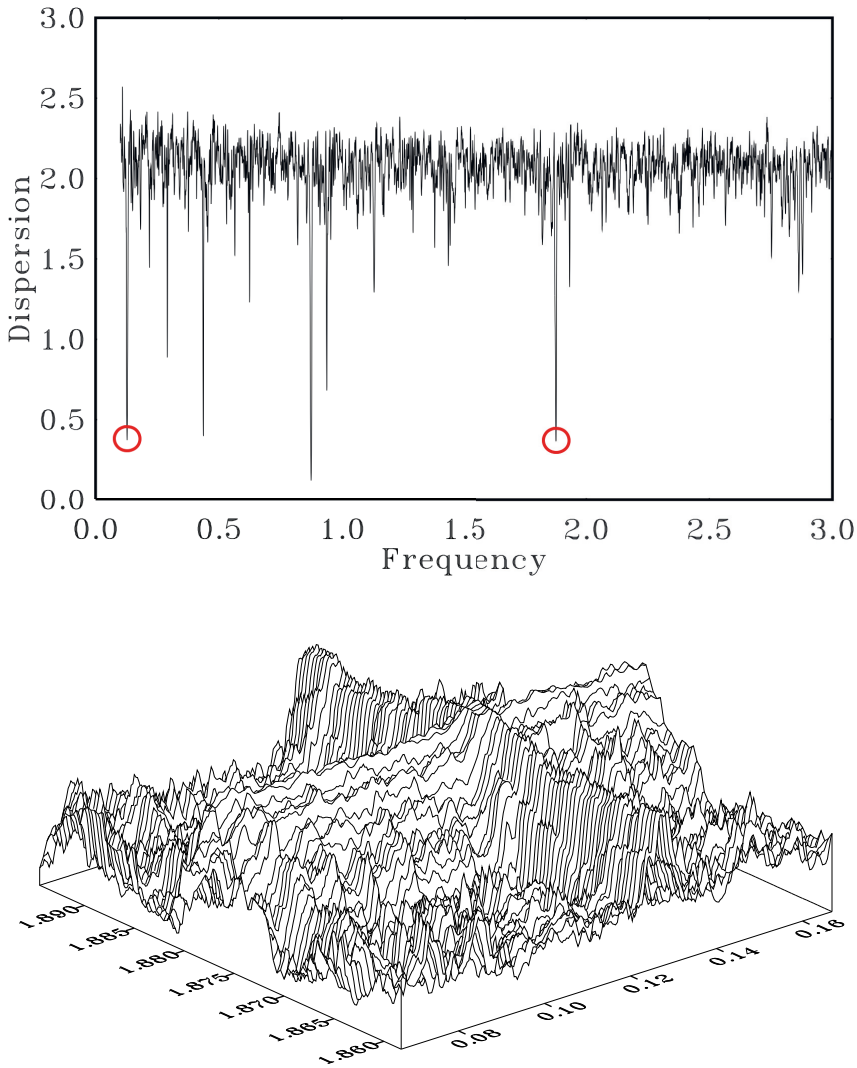


Figure 3.1. Example spectra of a multiperiodic model with frequencies $f_1 = 1.87655$ and $f_2 = 0.12764$. On top: one-dimensional D^2 spectrum. The true frequencies are marked with red circles. On bottom: two-dimensional D^2 spectrum. For better visualisation the two-dimensional spectrum has been inverted. Figure with kind permission from J. Pelt [105].

can be defined as $h(t) = h_M(t, \ldots, t)$. A special case of a multiperiodic function is the sum of M periodic functions. If we include the observational noise, we have the model $y(t) = h(t) + \epsilon(t)$ [104].

To estimate the periods from a time series of such a process, one can generalize Eq. (3.44) by introducing a phase selection term for each frequency $f_m = 1/P_m$:

$$D^2(f_1, \ldots, f_m) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \prod_{m=1}^{M} g(t_i, t_j, f_m) \big(y(t_i) - y(t_j)\big)^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} \prod_{m=1}^{M} g(t_i, t_j, f_m)}. \qquad (3.57)$$

The idea behind this equation is to take into account the phase proximity of the data with respect to each argument. However, as the computational complexity of calculating the spectrum for n frequency points for each component is proportional to n^M, then in practice considering the cases with $M > 2$ become infeasible [104]. This generalization is different than those of multicomponent Fourier based models, where the frequencies form an arithmetic progression [130, 11]. For a multifrequency periodogram see, for instance, Baluev [9, 10].

In figure Fig. 3.1 we show one-dimensional as well as two-dimensional D^2 spectra for the model $y(t) = A_1 \cos(2\pi f_1 t + \phi_1) + A_2 \cos(2\pi f_2 t + \phi_2) + \epsilon$, where $A_1 = 0.55073, A_2 = 0.58717, f_1 = 1.87656, f_2 = 0.12764, \phi_1 = 0.09111, \phi_2 = 0.091564$. It is evident that from the one-dimensional spectrum neither of the true frequencies can be easily inferred, while both of the frequencies are clearly visible on the two-dimensional spectrum.

3.6 Significance and error estimation

There have been many studies over the years where the statistical properties of different formulations of the empirical power spectra have been analysed. Primary attention has been paid to the models of the data such as Gaussian white noise and pure harmonic signal. Discussions of this topic is far outside the scope of the current thesis, however, in this section we briefly touch some relevant parts of it.

Estimating the power spectrum is obviously worthless without assessing the significance of the peaks, as it is possible that the peaks in the spectrum are just outcomes of pure chance. Answering this question is

part of statistical inference called hypothesis testing. Hypothesis testing requires constructing a statistical model of the data which assumes that the chance or random processes alone are responsible for the results. This is usually called the null hypothesis. In spectral estimation most commonly used null hypothesis is a Gaussian white noise. The significance under this hypothesis is sometimes possible to calculate analytically. For instance in the case of LS periodogram for the given significance (or FAP) level p_0 the peaks in spectrum reach

$$z_0 = -\ln\left(1 - (1 - p_0)^{1/N_{\perp\!\!\!\perp}}\right), \tag{3.58}$$

where $N_{\perp\!\!\!\perp}$ is the number of statistically independent frequencies searched for the maximum [125]. Note the close similarity to Eq. (3.25). In other words the periodogram ordinates follow the χ^2-distribution, with an assumption that the true noise variance is known. In practice, however, the variance is never exactly known, especially when the dataset is small and one estimates it empirically from the data. Depending on the different formulations of the LS periodogram, the peaks follow either the Fisher-Snedecor F-distribution or the beta-distribution [77, 131, 7].

Difficulties in assessing the significance in practice arise for three reasons: the dependence of the peaks in the estimated spectra, the closed form solutions for the distributions of the periodogram peaks might not exist or are hard to derive, and finally, the true noise model might be different from white noise or is totally unknown. Now we turn to the former of the two questions and in the end of this section we will shortly discuss other than white noise models

In the equations like Eq. (3.58) one should know the number of independent frequencies $N_{\perp\!\!\!\perp}$, however, this value is strongly dependent on the data sampling. Scargle [125] indicated that if the spectral window has a set of evenly spaced nulls at $f_n = nf_1$, then the peaks at f_n are uncorrelated. In practice often the nulls are approximately evenly spaced, thus it is a good estimate for $N_{\perp\!\!\!\perp}$. Horne and Baliunas [62] gave an empirical formula for the number of independent frequencies, based on simulations with different spacings of data:

$$N_{\perp\!\!\!\perp} = -6.362 + 1.193N_0 + 0.00098N_0^2, \tag{3.59}$$

where N_0 is the number of trial frequencies. Vio et al. [152] proposed even a simpler formula to approximate the number of independent frequencies, namely $N_{\perp} = N/2$, where N is the number of data points.

The cumulative distribution function (CDF) for the periodogram peaks can be also calculated empirically using Monte Carlo simulations. Linnell Nemec and Nemec [90] described a procedure for estimating the p-value of the PDM statistic Θ for a single period via doing m random permutations of the data and calculating $\sum_{i=1}^{m} \mathbb{I}(\Theta_i \leq \Theta_{\rm r})/m$, where Θ_i is Θ calculated for the i-th permutation and $\Theta_{\rm r}$ is the same for the real data. For power spectrum estimates one should use $\geq$ in the formula.

To overcome the need to know the number of independent peaks, Frescura et al. [50] proposed the following procedure for producing the empirical CDF of the periodogram peaks:

1. Using the sampling times of the actual data set to be analysed, construct a large number of pseudo-Gaussian random time series.

2. Select a convenient grid of frequencies that covers the frequency range in the periodogram that is to be inspected.

3. Construct the periodogram for each pseudo-random time series, sampling it at each of the selected frequencies.

4. In each periodogram, identify the highest periodogram power that occurs at the preselected frequencies only, and use these highest values to construct the CDF of the highest power values.

In some situations the white noise is not a suitable null hypothesis. It was pointed out by Gilman et al. [53] that the power spectra for many time series of atmospheric and solar indices are reminiscent of the spectrum of red noise, which is not constant, but decaying with increasing frequency according to a power law. Therefore, one can substantially overestimate the significance of the low-frequency peaks and conversely, underestimate it for high-frequency peaks. It is known that both photometric observations and radial velocity measurements of the stars contain red noise. It has been shown that the stellar p-mode oscillations have a noisy impact

on the precision Doppler velocities [100]. This jitter is dependent on the length of the time spent on observing a target at any given epoch. It was noted that the correlation of the photometric data at the millimagnitude level cannot be neglected [108]. The correlated noise in the data significantly affects the detection limit of exoplanet transit events. For instance, in Baluev [8], the nondiagonal covariance matrix was taken into account in the likelihood function of the data, and in Feng et al. [42], Bayesian linear regression model accounting for correlated noise was introduced and applied to the Kepler data.

Red noise-like light curves are also typical of those from active galactic nuclei (e.g. Seyfert galaxies and blazars). Tests for periodicities in the time series of these types of objects are discussed in Vaughan [148, 149], where the continuum of the spectrum is parametrically modelled.

Now we turn to the discussion of error estimation. While analytical formulas have been derived for calculating the uncertainty of the frequency estimate [129, 144], often the easiest way for error estimation is to use bootstrapping. If the number of data points is N, then a bootstrap sample is obtained by randomly sampling with replacement from the original dataset N times [39]. Then for each bootstrap sample one calculates the statistic of interest and obtains the empirical distribution for it.

For the methods such as LS, which provide the regression model, one can use this idea to first sample randomly with replacement from the residuals and add them back to the regression curve[4]. This way a new sample of the time series is generated, for which the model is fit again and new frequency estimate is obtained. The procedure is repeated many times, so that one obtains an empirical distribution for the frequency. The latter can in turn be used to calculate the standard deviation or confidence intervals for the frequency estimate. Sometimes the test for Gaussianity of the distribution is recommended to eliminate the potential errors or systematic biases introduced by the computation.

For the model free methods like PDM we do not have a regression curve, so that a different approach is needed. In the following we describe the procedure suitable for the D^2 statistic. From Eq. (3.44) it is evident that under the stationarity assumption the actual time moments of the measurements do not play a role in the statistic, but only the differences. We

[4]Here we assume that the noise is constant in time.

can thus convert the original time series $\{(y_i, t_i)\}$, $i = 0, \ldots, N-1$ to a time lag series $\{(\Delta y_k, \tau_k)\}$, where $\Delta y_k = y(t_i) - y(t_j)$, $\tau_k = |t_i - t_j|$, $k = iN + j$. Next, we can use the bootstrap technique to sample N^2 points from the new time series and calculate D^2 statistic for each bootstrap sample. Another way to look at the procedure just described is to notice that it is equivalent to forming a new time series by sampling N points (y_i, t_i) with replacement from the original time series. For the bootstrap method to yield realistic error estimates, sufficiently long time series are needed.

3.7 Probabilistic methods

Earlier we saw that one way to estimate the power spectrum is to use a least squares fit of harmonic functions to the data. From probabilistic viewpoint the latter is equivalent to a maximum likelihood estimation with Gaussian likelihood. More generally, we can postulate any kind of probabilistic model explaining the data, which depends on different parameters, including the frequency. Together with the use of prior information about the parameters one infers the posterior probability distribution of the parameters. The meaning of the prior distributions is similar to the regularisation in the case of frequentists statistics. The parameter values at the mode of the posterior distribution are sometimes taken as the best estimate, which is called a maximum a posteriori (MAP) estimate. However, instead of one point estimate, Bayesian inference is based on computing expectations over the posterior distribution and thus considering all possible parameter values weighted by their posterior density or probability. Considering the distribution provides a natural way to get also uncertainty estimates for the interesting quantities.

In the domain of period estimation, the probabilistic methods have only recently become popular. Hence, let us first, before moving to the description of Bayesian period estimation, summarise the relevant rules of probability.

3.7.1 Rules of probability

Following Barber [12, Ch. 1.1], consider the case of discrete random variables, and denote the probability of an event $x = \mathrm{a}$ by $p(x = \mathrm{a})$ or just by

$p(x)$. The probabilities are always positive, so that $p(x) \geq 0$. Normalisation condition requires that $\sum_{x \in \mathrm{dom}(x)} p(x) = 1$, where $\mathrm{dom}(x)$ denotes the domain of x (the set of all possible events). The probability of two events $x = \mathrm{a}$ and $y = \mathrm{b}$ happening simultaneously is written as $p(x = \mathrm{a}, \mathrm{y} = \mathrm{b})$, or simply by $p(x, y)$. This is called a joint distribution. The distribution of a single variable is obtained from the joint distribution by marginalisation:

$$p(x) = \sum_{y} p(x, y). \tag{3.60}$$

The probability of an event x conditioned on knowing an event y (the probability of x given y) is defined as

$$p(x|y) \equiv \frac{p(x, y)}{p(y)}. \tag{3.61}$$

From this formula, using $p(x, y) = p(y, x)$, one can derive the Bayes rule, which has a central role in Bayesian inference:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}. \tag{3.62}$$

If two random variables x and y are independent, then their joint distribution factorises, so that $p(x, y) = p(x)p(y)$.

These formulas are also applicable to continuous random variables, but then the meaning of p is the probability density and the sums are replaced by integrals. Normalisation condition then takes the form $\int_{-\infty}^{\infty} p(x) = 1$ and the probability of x falling into interval $[a, b]$ is given by $\int_a^b p(x)dx$.

Previously we have frequently used the notion of an expected value. Now we give its formal definition. For a continuous random variable case the expected value of a function f of random variable x is defined by

$$\mathbb{E}_{p(x)} f(x) = \int_{-\infty}^{\infty} f(x)p(x)dx. \tag{3.63}$$

For a single variable case as shown above, or when the distribution over which the expectation is taken is evident, the subscript $p(x)$ is usually omitted. For a discrete random variable case the formula is similar, only the integral is replaced by the sum.

3.7.2 Bayesian linear regression

Considering a set of model parameters θ and data $\mathcal{D}$, the Bayes rule in Eq. (3.62) becomes

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int_\Theta p(\mathcal{D}|\theta)p(\theta)d\theta}, \tag{3.64}$$

where $\Theta = \mathrm{dom}(\theta)$. In this equation $p(\theta|\mathcal{D})$ is called the posterior, $p(\mathcal{D}|\theta)$ the likelihood, $p(\theta)$ the prior and $p(\mathcal{D})$ the evidence. The probability distribution of the data point x_*, measured in the future, is given by the so-called posterior predictive equation

$$p(x_*|\mathcal{D}) = \mathbb{E}_{p(\theta|\mathcal{D})}p(x_*|\theta) = \int_\Theta p(x_*|\theta)p(\theta|\mathcal{D})d\theta. \tag{3.65}$$

In the simplest case, assuming independent Gaussian noise with known constant variance, the Bayesian linear regression involves the parameters $\theta = \{\boldsymbol{\theta}, \mu, \sigma^2\}$ and data $\mathcal{D} = \{\mathbf{y}, \mathbf{X}\}$, where $\boldsymbol{\theta} \in \mathbb{R}^K$ is the vector of regression weights, μ is the offset, σ^2 is the noise variance, $\mathbf{y} \in \mathbb{R}^N$ is the vector of output variable, $\mathbf{X} \in \mathbb{R}^{N\times K}$ is the matrix of input variables, whose N rows represent the vector of input $\mathbf{x}_i \in \mathbb{R}^K$ corresponding to output y_i.

The likelihood is given by

$$\begin{aligned} p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}, \mu, \sigma^2) &= \mathcal{N}(\mathbf{y}|\mu + \mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I}_N) \\ \propto \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mu\mathbf{1}_N - \mathbf{X}\boldsymbol{\theta})^\mathsf{T}(\mathbf{y} - \mu\mathbf{1}_N - \mathbf{X}\boldsymbol{\theta})\right), \end{aligned} \tag{3.66}$$

where $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ is the unit matrix and $\mathbf{1}_N \in \mathbb{R}^N$ is the vector of all ones [97, Ch. 7.6].

One can take the mean μ to be zero without the loss of generality, as one can introduce a unit pseudo variate for each input vector $\mathbf{x}_i$, and include μ into the vector $\boldsymbol{\theta}$. So, assuming zero mean the above formulation corresponds to the model

$$y_i = \boldsymbol{\theta}^\mathsf{T}\mathbf{x}_i + \epsilon \quad i = 1, \ldots, N, \tag{3.67}$$

where $\epsilon \sim \mathcal{N}(\mu, \sigma^2)$.

If we use a Gaussian prior for the parameters $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathbf{V}_0)$, then the

posterior becomes

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathbf{V}_0)\mathcal{N}(\mathbf{y}|\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I}_N) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\theta}_N, \mathbf{V}_N), \tag{3.68}$$

where

$$\begin{aligned} \boldsymbol{\theta}_N &= \mathbf{V}_N\mathbf{V}_0^{-1}\boldsymbol{\theta}_0 + \frac{1}{\sigma^2}\mathbf{V}_N\mathbf{X}^\mathsf{T}\mathbf{y}, \\ \mathbf{V}_N^{-1} &= \mathbf{V}_0^{-1} + \frac{1}{\sigma^2}\mathbf{X}^\mathsf{T}\mathbf{X}, \\ \mathbf{V}_N &= \sigma^2(\sigma^2\mathbf{V}_0^{-1} + \mathbf{X}^\mathsf{T}\mathbf{X})^{-1}. \end{aligned} \tag{3.69}$$

Using Eq. (3.65), the posterior predictive is given by

$$\begin{aligned} p(y|\mathbf{x}, \mathcal{D}, \sigma^2) &= \int \mathcal{N}(y|\mathbf{x}^\mathsf{T}\boldsymbol{\theta}, \sigma^2)\mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\theta}_N, \mathbf{V}_N)d\boldsymbol{\theta} \\ &= \mathcal{N}(y|\boldsymbol{\theta}_N^\mathsf{T}\mathbf{x}, \sigma^2 + \mathbf{x}^\mathsf{T}\mathbf{V}_N\mathbf{x}). \end{aligned} \tag{3.70}$$

If the noise cannot be assumed to be constant, $\sigma^2\mathbf{I}_N$ has to be replaced with $\mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)$. In a more general case when the independency of the noise cannot be any more assumed, one needs to use a full covariance matrix $\boldsymbol{\Sigma}$.

Above we assumed $\boldsymbol{\Sigma}$ to be known. For Bayesian linear regression with unknown noise variance see, for instance, Murphy [97, Ch. 7.6].

3.7.3 Bayesian period estimation

One can map the input vector $\mathbf{x}$ using an arbitrary (nonlinear) function $g : \mathbb{R}^K \to \mathbb{R}^{K'}$ to get a new input vector $\mathbf{g} = g(\mathbf{x})$. Then all the previous equations are valid, but instead of K and $\mathbf{x}$ we have K' and $\mathbf{g}$, respectively. In Bayesian period estimation natural choice for g is the combination of harmonic functions. From now on in this section we assume $\mathbf{x}$ to be single dimensional time and we denote it with t. Let us select $g(t)$ such that

$$\mathbf{g}_i = g(t_i) = [\cos(2\pi f t_i - \phi), \sin(2\pi f t_i - \phi), 1]^\mathsf{T}, \tag{3.71}$$

where ϕ has a fixed value for each frequency f and its meaning will become clear later. We can write our regression model as

$$y_i = \boldsymbol{\theta}^\mathsf{T}\mathbf{g}_i + \epsilon = \theta_1\cos(2\pi f t_i - \phi) + \theta_2\sin(2\pi f t_i - \phi) + \theta_3 + \epsilon \quad i = 1, \ldots, N. \tag{3.72}$$

Including f into the list of parameters of the probabilistic model and using uniform prior for it, the posterior can be written as

$$p(\boldsymbol{\theta}, f|\mathcal{D}) \propto \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathbf{V}_0)\mathcal{N}(y|\mathbf{G}\boldsymbol{\theta}, \sigma^2\mathbf{I}_N), \tag{3.73}$$

where $\mathbf{G} \in \mathbb{R}^{N\times 3}$ is a matrix whose i-th row is $\mathbf{g}_i^\mathsf{T}$. This equation is otherwise identical to Eq. (3.68), but instead of $\mathbf{X}$ we have $\mathbf{G}$. To find the optimal period one could in principle sample from $p(\boldsymbol{\theta}, f|\mathcal{D})$ using, for instance, Markov Chain Monte Carlo methods. However as with respect to $\boldsymbol{\theta}$ this distribution is multivariate Gaussian, equivalent approach is to marginalise over $\boldsymbol{\theta}$ and find the optimal frequency from $p(f|\mathcal{D})$. The latter quantity is usually called a probabilistic spectrum or Bayesian spectrum.

Assuming independence of the noise and using uniform priors for both $\boldsymbol{\theta}$ and f, a formula for probabilistic spectrum was derived in Mortier et al. [96]:

$$p(f|\mathcal{D}) \propto \frac{1}{\sqrt{|K|(\mathbf{c}^\mathsf{T}\mathbf{c})(\mathbf{s}^\mathsf{T}\mathbf{s})}} \exp\left(M - \frac{L^2}{4K}\right), \tag{3.74}$$

where

$$\begin{aligned}
K &= \frac{1}{2}\left(\frac{(\mathbf{w}^\mathsf{T}\mathbf{c})^2}{\mathbf{c}^\mathsf{T}\mathbf{c}_w} + \frac{(\mathbf{w}^\mathsf{T}\mathbf{s})^2}{\mathbf{s}^\mathsf{T}\mathbf{s}_w} - \mathbf{1}^\mathsf{T}\mathbf{w}\right),\\
L &= \mathbf{w}^\mathsf{T}\mathbf{y} - \frac{(\mathbf{w}^\mathsf{T}\mathbf{c})(\mathbf{c}^\mathsf{T}\mathbf{y}_w)}{\mathbf{c}^\mathsf{T}\mathbf{c}_w} - \frac{(\mathbf{w}^\mathsf{T}\mathbf{s})(\mathbf{s}^\mathsf{T}\mathbf{y}_w)}{\mathbf{s}^\mathsf{T}\mathbf{s}_w},\\
M &= \frac{1}{2}\left(\frac{(\mathbf{c}^\mathsf{T}\mathbf{y}_w)^2}{\mathbf{c}^\mathsf{T}\mathbf{c}_w} + \frac{(\mathbf{s}^\mathsf{T}\mathbf{y}_w)^2}{\mathbf{s}^\mathsf{T}\mathbf{s}_w}\right),\\
\mathbf{w} &= [1/\sigma_1^2 \ldots 1/\sigma_N^2]^\mathsf{T},\\
\mathbf{y} &= [y_1 \ldots y_N]^\mathsf{T},\\
\mathbf{c} &= [\cos(2\pi f t_1 - \phi) \ldots \cos(2\pi f t_N - \phi)]^\mathsf{T},\\
\mathbf{s} &= [\sin(2\pi f t_1 - \phi) \ldots \sin(2\pi f t_N - \phi)]^\mathsf{T},\\
\mathbf{y}_w &= \mathbf{w} \circ \mathbf{y} = [w_1 y_1 \ldots w_N y_N]^\mathsf{T},\\
\mathbf{c}_w &= \mathbf{w} \circ \mathbf{c} = [w_1 c_1 \ldots w_N c_N]^\mathsf{T},\\
\mathbf{s}_w &= \mathbf{w} \circ \mathbf{s} = [w_1 s_1 \ldots w_N s_N]^\mathsf{T},\\
\phi &= \frac{1}{2}\arctan\left(\frac{\mathbf{w}^\mathsf{T}\mathbf{s}_2}{\mathbf{w}^\mathsf{T}\mathbf{c}_2}\right),\\
\mathbf{c}_2 &= [\cos(4\pi f t_1) \ldots \cos(4\pi f t_N)]^\mathsf{T},\\
\mathbf{s}_2 &= [\sin(4\pi f t_1) \ldots \sin(4\pi f t_N)]^\mathsf{T}.
\end{aligned} \tag{3.75}$$

It is important to note that $\phi = 2\pi f\tau$, where τ is defined in Eq. (3.32). Moreover, the term M is proportional to the power calculated in GLS [96,

163]. Thus, Eq. (3.74) can be considered as a probabilistic analogue of GLS. In Paper V we generalised the above model by adding a linear trend component to g_i. We will discuss it more in Sect. 6.3.

When the noise variance is unknown, but constant in time, one can, similarly to other parameters, integrate it out to obtain $p(f|\mathcal{D})$. Most convenient in that case is to use flat priors for $\boldsymbol{\theta}$ and f, and Jeffreys' prior for σ^2. This approach was introduced in Bretthorst [26], where the spectrum was derived for a model consisting of complex signal allowing arbitrary amplitude modulation plus noise. For a special case without the modulation, the spectrum is given by

$$p(f|\mathcal{D}) \propto \frac{1}{\sqrt{(\mathbf{c}^\mathsf{T}\mathbf{c})(\mathbf{s}^\mathsf{T}\mathbf{s})}} \left(\mathbf{y}^\mathsf{T}\mathbf{y} - \frac{(\mathbf{y}^\mathsf{T}\mathbf{c})^2}{\mathbf{c}^\mathsf{T}\mathbf{c}} - \frac{(\mathbf{y}^\mathsf{T}\mathbf{s})^2}{\mathbf{s}^\mathsf{T}\mathbf{s}} \right)^{(2-N)/2}, \tag{3.76}$$

where in the brackets one can recognise the LS periodogram.

The Bayesian approach in period estimation has two major benefits. First, the probabilistic spectrum has a straightforward interpretation as the ratio of any two peaks corresponds to the ratio of probabilities of those frequencies being true [96]. Second, by approximating the spectral line with a Gaussian, one can easily obtain an uncertainty estimate for the frequency [27, Ch. 2.4]. As mentioned above, for estimating the uncertainty in the case of LS periodogram one needs to use bootstrapping or analytic approximations.

To assess the significance of the model, one can calculate the Bayes factor between a given model and an alternative model [97, Ch. 5.3.3]. The latter one is usually taken as the constant model, which assumes that the data points are following the Gaussian white noise. Alternatively, one can use Bayesian information criterion (BIC), which is an approximation to the logarithmic marginal likelihood of the data, given the model [97, Ch. 5.3.2.4].

3.8 Selecting the frequency grid

One crucial aspect when calculating the periodograms is the selection of the frequency grid in a way that its range would be wide enough to cover all detectable frequencies and sampling would be dense enough not to miss a period detection. On the other hand, from the point of view of

computational efficiency, one would want to select the grid as narrow and sparsely sampled as possible. For the equispaced case the highest detectable frequency from the data is defined by the Nyquist frequency f_{N} and the frequency grid can be chosen based on the natural frequencies given by Eq. (3.33).

For irregular sampling the Nyquist limit is usually much higher than defined by the average sampling rate of the data, namely $f_{\mathrm{N}} = 1/2p$, where p is the greatest common divisor between $t_i - t_1$, that is, $t_i = t_1 + n_i p$, $n_i \in \mathbb{N}, i = 1, \ldots, N$ [40].

Another, yet simpler way of finding f_{N}, is to find the lowest frequency f_0 which satisfies

$$\sum_{k=1}^{N-1} \sum_{l=k+1}^{N} \left(\sin 2\pi f_0 (t_k - t_l) \right)^2 = 0. \tag{3.77}$$

If we denote by $\delta_{\min}$ the minimum distance between the consecutive time moments, then one can calculate $f_0 = J/2\delta_{\min}$ for successively increasing values of $J \in \mathbb{N}$ and evaluate Eq. (3.77). The first frequency for which the equality holds is then the Nyquist frequency for the given sampling [78]. Besides the zero frequency, the Nyquist frequency is the point relative to which the spectrum has the mirror symmetry, that is, $P(f_{\mathrm{N}} + f) = P(f_{\mathrm{N}} - f)$ for any f. From this immediately follows that the spectrum repeats itself after interval $2f_{\mathrm{N}}$.

One can estimate f_{N} also using the spectral window $W_N(f)$ from Eq. (3.28). For the evenly sampled case this quantity is periodic with a period $2f_{\mathrm{N}}$ satisfying $W_N(2f_{\mathrm{N}})/N = 1$. Led by this idea, one can approximate the Nyquist frequency in the unevenly sampled case by finding the smallest f, which is greater than zero and satisfying $W_N(2f)/N \approx 1$ [4].

The next thing one should consider is the density of the frequency grid. A good rule of thumb, when choosing the number of frequencies which to evaluate, is

$$N_{\mathrm{eval}} = n_0 T f_{\max}, \tag{3.78}$$

where n_0 is the oversampling factor, usually taken 5, T is the time span of the data and $f_{\max}$ is the upper bound for the frequency search range. If feasible, $f_{\max}$ would be set equal to f_{N}. This equation roots from the simple idea that the data observed through a rectangular window of length T will have sinc-shaped peaks of width $\approx 1/T$ [144].

In the context of PDM methods, a practically identical formula was given by Pelt [104]:

$$N_{\mathrm{eval}} = \frac{T}{\Delta\varphi}\left(f_{\max} - f_{\min}\right), \tag{3.79}$$

where $\Delta\varphi$ is the maximum allowed error in phase, usually taken 0.1, $f_{\min}$ and $f_{\max}$ are the minimum and maximum frequencies evaluated, respectively.

In practice, however, one often cannot choose $f_{\max}$ equal to f_{N}, as it may lead to a too large frequency grid. One then needs to use prior knowledge about the possible placement of the true frequency and set the upper bound accordingly [144].

3.9 Cleaning the spectrum

By spectral cleaning one means the procedure of recovering the true ('clean') spectrum $Y(f)$ from the 'dirty' spectrum $Y_N(f)$ in Eq. (3.27). The direct deconvolution is, however, an ill posed problem, as the window function is mostly zero everywhere. There is no unique solution to the problem [144]. However, most of the period estimation methods do not involve calculating $Y_N(f)$, but rather a proxy to the power spectrum $P(f)$. Therefore, more loosely speaking, one can think of spectral cleaning as a method for filtering out true periodicities from the forest of false peaks for any statistic used. When we consider a noise free periodic signal, then essentially all period estimation methods have three major problems: spectral leakage, appearance of aliases and the presence of peaks for over- or undertones. Leakage and aliases both arise due to the effect of the window function on the true spectrum as discussed in Sect. 3.3.2. Sometimes leakage is attributed to the finite length of the time series and aliases to the discreteness of the observations. The third problem appears because of the non-harmonicity of the signal. The models usually involve a trial frequency through simple periodic functions, so that when the local minimum is reached for a certain frequency f, then the same usually holds either for nf or f/n ($n \in \mathbb{N}$) as well.

If we have equally spaced observations with step Δt in time, for a given

trial period P one can find the aliases (or so-called spurious periods) using

$$\frac{1}{P_{\mathrm{spur}}} = \frac{1}{P} \pm \frac{n}{k\Delta t}, \tag{3.80}$$

where n and k are integers [137]. Even when the sampling is not strictly even, one can still use the formula for estimating the approximate locations of the aliased peaks in the periodogram.

An elaborate technique for eliminating the aliases and filtering out the peaks of the true periodicities from the spectrum called CLEAN was developed by Roberts et al. [119]. This is summarized by the following algorithm:

Algorithm 1: CLEAN algorithm

Input: $\{t_i, y_i\}$

Output: The cleaned spectrum

$i \leftarrow 1$

$Y(f) \leftarrow Y_N(f)/N$, where $Y_N(f)$ is calculated using Eq. (3.26)

while *Significant peaks found in* $Y(f)$ **do**

  Find the frequency of the highest peak in spectrum

  $f_i \leftarrow \arg\max Y(f)$

  Calculate the complex amplitude of the peak using

  $\alpha = \frac{Y(f_i) - Y^*(f_i)W(2f_i)}{1-|W(2f_i)|^2}$, where $W(f) = W_N(f)/N$ and $W_N(f)$ is from Eq. (3.28)

  $c_i \leftarrow g\alpha$, where g is the gain (usually taken between 0.1 and 1)

  $Y(f) \leftarrow Y(f) - \big(c_i W(f - f_i) + (c_i^* W(f - f_i)\big)$

  $i \leftarrow i + 1$

end

$K \leftarrow i - 1$

Fit the Gaussian to the spectral window $W(f)$ to get $B(f)$

Form the final clean spectrum as

$Y(f) \leftarrow \sum_{i=1}^{K} \big(c_i B(f - f_i) + c_i^* B(f - f_i)\big) + Y(f)$

return $Y(f)$

The idea of the algorithm comes from the fact that for a pure sine wave $y(t) = A\cos(2\pi f_0 t + \phi)$ the true spectrum is $Y(f) = a\delta(f - f_0) + a^*\delta(f - f_0)$, where $a = Ae^{i\phi}/2$. The purpose of the gain g is to allow removing only fraction of the amplitude of the peak at each step, thus reducing the errors introduced by mismeasuring the peak amplitudes and compensating for the frequency shift due to mixing of multiple frequencies.

Very close, but somewhat simpler approach compared to the CLEAN algorithm is the following algorithm, which directly works with the dirty power spectrum $P(f)$, instead of $Y_N(f)$:

Algorithm 2: Cleaning the power spectrum

Input: $\{t_i, y_i\}$

Output: The frequencies of the found peaks

$\mathbf{y} \leftarrow [y_1 \dots y_N]$

Calculate the power spectrum $P(f)$ for $\mathbf{y}$

$F \leftarrow \emptyset$

while *Significant peaks found in* $P(f)$ **do**

  Find the frequency of the highest peak in spectrum

  $f_{\max} = \arg\max P(f)$

  $F \leftarrow F \cup \{f_{\max}\}$

  Fit the model $g(t)$ using $f_{\max}$ to the data

  Sample the model forming $\mathbf{g} = [g(t_1) \dots g(t_N)]$

  Subtract the model from the data to form new data: $\mathbf{y} \leftarrow \mathbf{y} - \mathbf{g}$

  Calculate power spectrum $P(f)$ for $\mathbf{y}$

end

return F

This method can be applied to any model based periodogram, such as the LS periodogram. Note that if at each step the highest peak corresponds to the true frequency, then by subtracting the periodic component with the given frequency from the data removes also the aliases of that frequency. However, the highest peak itself can sometimes correspond to an alias, for instance, in the case when the aliases of two true frequencies add up. Cleaning such a spectrum can be tricky and needs more reasoning, especially taking into consideration the average sampling rate and using Eq. (3.80). To avoid selecting false peaks into the result set, one can, at each step, find a set of K highest peaks and repeat the procedure for each of these peaks separately. Some stopping criteria must be introduced. The chain with the least residual error is likely to be the best candidate for the set of true frequencies. This approach was used in Paper VI when estimating the cycle periods of active stars.

In Fig. 3.2(a), we show the LS periodogram of a time series with two harmonics having frequencies $f_1 = 0.62$ and $f_2 = 1.62$. Due to the sampling pattern with approximate time step of two units the aliases of the

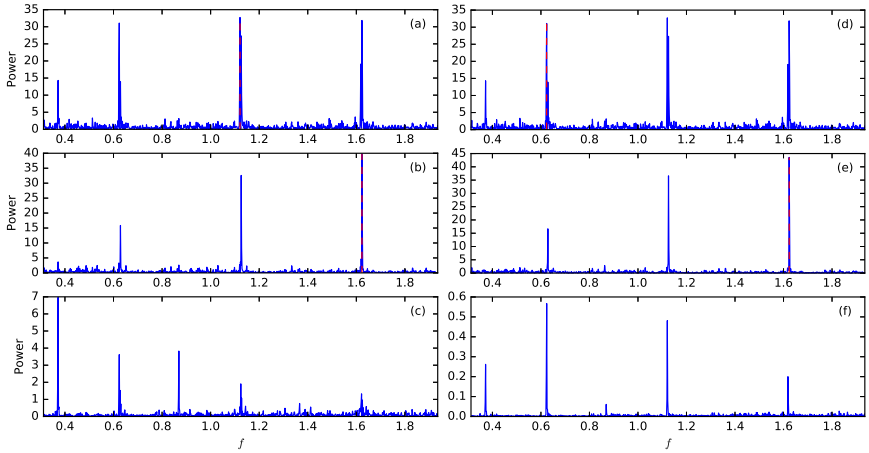


Figure 3.2. Cleaning of the spectrum of a time series with two harmonics. The central peak is an alias due to sampling. Original spectrum with the highest, but incorrect peak marked for removal with red dashed line (a). Residual spectrum after removal of the first peak and showing the new highest peak (b). Final spectrum after both peaks in the previous steps have been removed (c). (d)-(f) show the same procedure, but in the first step one of the true frequencies is removed.

harmonics add up and the highest peak in the spectrum falls on a wrong frequency around $f_{\mathrm{A}} = 1.12$. In panel (b) of the same figure, we show the residue spectrum when the harmonic with the wrong frequency f_{A} has been subtracted from the data. Note that the peak around f_{A} has not significantly dropped, however, now the highest peak is at the true frequency f_2. When the harmonic with this frequency is removed, the total power drops significantly and one would think that the spectrum is finally cleaned (panel (c)). Note, however that some of the power has redistributed to the peaks where in the previous step there was less power (for instance, the peak around 0.38, which is an alias of the peak at $-f_1$). In panels (d), (e) and (f) of the Fig. 3.2, we show an alternative cleaning procedure starting from the second highest, but true peak at f_1. Now after also removing the highest peak at f_2, we see that the reduction in the power has dropped an order of magnitude more than compared to the previous example. Thus, it is not always sufficient to greedily clean the spectrum by selecting the peaks with strongest powers first.

To overcome the problems with the methods described above a method for cleaning the spectrum multiple peaks at a time was proposed by Foster [45]. However, we will not go into more details here.

3.10 Summary

In this chapter we covered some of the well known methods in the domain of nonparametric spectral estimation. For comparison of some of the methods please see Carbonell et al. [30], Taylor and Hamilton [138]. Over the years many more methods have been developed, which we did not mention here. Some of these methods are based, for instance, on the interpolation of the time series or resampling the time series onto an even grid, followed by application of FT. Some of the other methods not covered include a hybrid method combining PDM and LS periodogram [122], and methods developed for multiband time series [145, 66]. For more extensive reviews on spectral analysis of unevenly sampled data, please refer to VanderPlas [144], Babu and Stoica [4].

4. Methods for quasi-periodic time series

In the previous chapter we discussed periodic processes and practical ways of assessing the period. Most of the models or methods discussed so far have an assumption of one single, usually harmonic, signal component in the observed time series. When the signal is not strictly harmonic, but periodic, the methods still work, but the spectra contain extra peaks due to the non-harmonicity. The spectra get even more complicated when there are several periodic signals in the time series. In Sect. 3.9 we discussed ways of filtering out separate periods in the presence of multiperiodicity using tools originally developed for single period analysis. In this chapter we turn to the other direction and discuss the processes which are not any more strictly periodic, that is, the period can vary over time. We refer to such processes as quasi-periodic (or cyclic) processes[1]. By quasi-periodic process we mean such a WSS process which is locally periodic with a constant period. In other words, small amplitude and frequency modulations are allowed, but on average the process has a single well defined period. The emphasis is on a constant period, as, for instance, the chirp signal is also locally harmonic, however, as the period changes over time, it is not a quasi-periodic, but a non-stationary signal. In terms of covariance functions, we can express the cyclicity in the form $\mathrm{cov}(t, t') = g(t, t')p(t, t')$, where g is a damping term and p is a periodic term. Here g is in general more slowly changing function than p, as otherwise the process becomes fully stochastic. We start with an example to show that 'traditional' period estimation can fail when the time series is cyclic.

[1]Throughout this thesis we use both of these terms interchangeably.

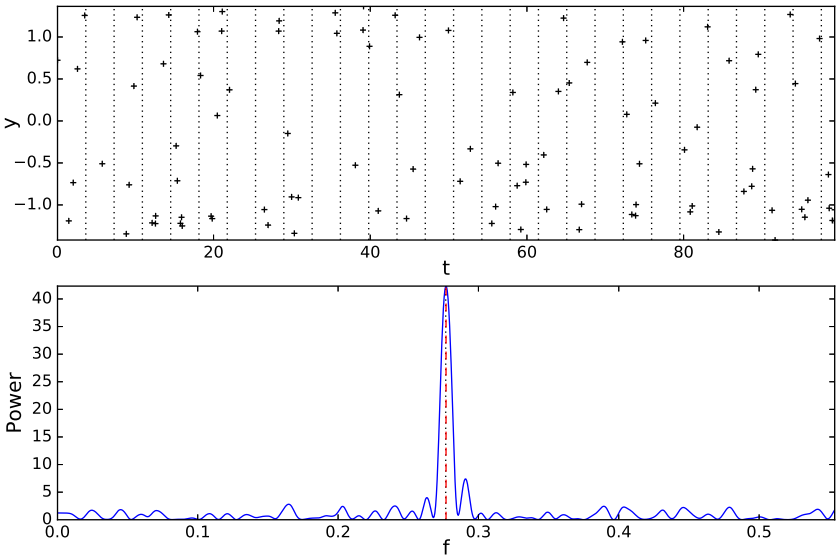


Figure 4.1. Top: Time series of a harmonic process. Vertical dotted lines mark the interval of the period. Bottom: LS periodogram of the time series. Red dashed line shows the position of the true frequency and the black dotted line the estimated frequency.

Consider a covariance function of the form

$$\mathrm{cov}(t,t') = \exp\left(-\frac{(t-t')^2}{\ell^2}\right)\cos\left(2\pi f(t-t')\right). \tag{4.1}$$

This is one of the examples of a locally harmonic (or quasi-harmonic) process. Without the exponential term it would be a fully harmonic process, however, the exponential damping reduces the correlation to zero at timescales $\gtrapprox \ell$. For this reason, ℓ is usually called a time-scale[2]. To illustrate the effect of non-harmonicity on the power spectrum, we have plotted two figures. On Fig.4.1 there is plotted a time series of a purely harmonic process without the noise and where the number of data points is 100. We see that the true frequency of the process is closely matching the estimate from the LS periodogram. On Fig.4.2 we have plotted a time series with a cyclic process with the covariance function given by Eq. (4.1). The true frequency, sampling pattern and the number of data points are identical to the harmonic case. The time-scale of this process was chosen to be five times the period. We note that the LS spectrum has lost its interpretability, as neither the estimated frequency based on the maximum peak, nor any of the other peaks match the true frequency.

[2]In some of the papers we refer to this quantity as coherence time.

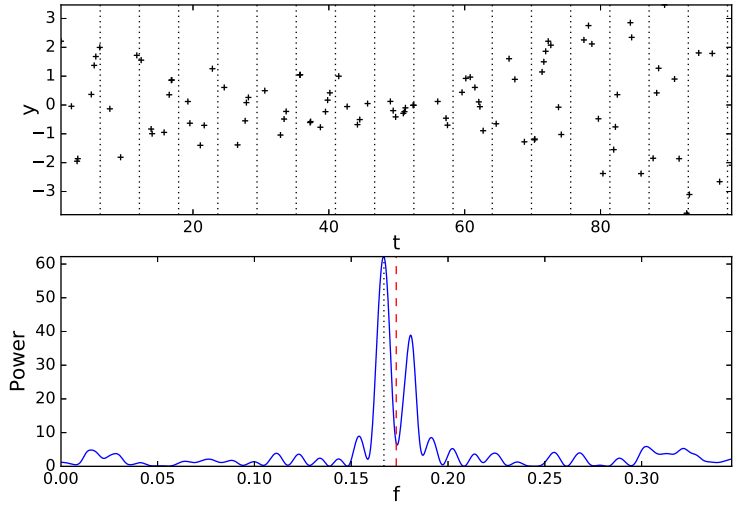


Figure 4.2. Top: Time series of a cyclic process with $\ell = 5P$. Vertical dotted lines mark the interval of the period. Bottom: LS periodogram of the time series. Red dashed line shows the position of the true frequency and the black dotted line the estimated frequency.

4.1 Carrier fit method

One simple way to model a locally harmonic time series is to assume that the amplitudes of the sine and cosine components in the regression model are not constant, but slightly varying over time:

$$y(t) = a(t)\cos(2\pi f t) + b(t)\sin(2\pi f t) + \epsilon(t), \tag{4.2}$$

where $a(t)$ and $b(t)$, the so-called 'modulators', are slowly changing functions compared to the period $P = 1/f$ of the harmonic components, the so-called 'carriers', and $\epsilon(t) \sim \mathcal{N}\big(0, \sigma^2(t)\big)$ is the zero mean Gaussian independent noise. The idea behind this model is that our process is a product of two independent processes, one being harmonic and the other slowly changing in time. Covariance of such a process (without noise and assuming certain restrictions) is of the form $g(t,t')\cos\big(2\pi f(t-t')\big)$, where $g(t, t \pm \tau) \gg 0$ if $\tau \lessapprox P$. We call this model as the Carrier fit (CF) model, hinting that the carrier harmonic with frequency f is slightly modulated in time [107]. More generally, if we want the carrier signal to be not strictly harmonic, but periodic, we can also include the overtones of the

base frequency into the model:

$$y(t) = \sum_{k=0}^{K} \big(a_k(t)\cos(2\pi k f t) + b_k(t)\sin(2\pi k f t)\big) + \epsilon(t), \tag{4.3}$$

where the zeroth component $a_0(t)$ represents the slow trend of the signal. One further needs to select the functional form of $a_k(t)$ and $b_k(t)$. Let the interval of the observation times be $[t_{\min}, t_{\max}]$. If we divide this interval into L subintervals, so that $[t_{\min}, t_{\max}] = [t_0, t_1) \cup [t_1, t_2) \cup \cdots \cup [t_{L-1}, t_L]$ and define polynomial $P(t)$ of degree M on each of these subintervals, we can build $a_k(t)$ and $b_k(t)$ as combinations of these polynomials as follows:

$$\begin{aligned} a_k(t) &= \sum_{l=0}^{L-1} \mathbb{I}(t_l \le t < t_{l+1}) P_{kl}^{(a)}(t) = \sum_{l=0}^{L-1} \mathbb{I}(t_l \le t < t_{l+1}) \sum_{m=0}^{M} c_{klm}^{(a)} t^m, \\ b_k(t) &= \sum_{l=0}^{L-1} \mathbb{I}(t_l \le t < t_{l+1}) P_{kl}^{(b)}(t) = \sum_{l=0}^{L-1} \mathbb{I}(t_l \le t < t_{l+1}) \sum_{m=0}^{M} c_{klm}^{(b)} t^m. \end{aligned} \tag{4.4}$$

In this formulation, functions $a_k(t)$ and $b_k(t)$ are called splines. To guarantee the smoothness of the resulting curves, the polynomials must share common derivatives (usually the first and the second derivative) at the boundaries of the subintervals. In practice it is often sufficient to consider the case of cubic splines ($M = 3$) and usually the subintervals are taken to be of an identical width.

Another option to define the modulators is to use a truncated Fourier series with low frequency base harmonic $f_D \approxeq 1/(t_{\max} - t_{\min})$:

$$\begin{aligned} a_k(t) &= \sum_{l=0}^{L} \left(c_{kl}^{(a)} \cos(2\pi l f_D t) + s_{kl}^{(a)} \sin(2\pi l f_D t)\right), \\ b_k(t) &= \sum_{l=0}^{L} \left(c_{kl}^{(b)} \cos(2\pi l f_D t) + s_{kl}^{(b)} \sin(2\pi l f_D t)\right). \end{aligned} \tag{4.5}$$

In both of the above cases the number of components, L, defines how strong the modulation of the carrier signal is [107]. The goodness of the model fit is heavily dependent on K and L. To achieve optimal values for K and L, avoiding over- or underfitting, one can use Cross-validation (CV) or calculate the Bayesian information criterion (BIC). Although the CF is a method based on a linear least squares regression, due to many components in the model, the easiest way to get the uncertainty for the model

function is to use bootstrapping.

The CF method is suitable for such time series for which the spectrum is contained within a narrow band and its integer multiples. The width of the band is connected to L and the number of bands to K.

4.1.1 Visualisation of the CF model

Being a continuous fit, the CF model gives a good means to visualise the phase behaviour of the signal. We give an example in this section, but the same idea can be used with any quasi-periodic regression model. When the model $g(t)$ with optimal parameter values has been found, one can sample it forming the sequence $\{t_i, g(t_i)\}$, and calculate the full period counts and phases of the model points with respect to the carrier period P and some preselected epoch: $n_i = \lfloor t_i/P \rfloor, \varphi_i = \mathrm{frac}(t_i/P)$. Then one can plot the triplets $(n_i P, \varphi_i, g(t_i))$ as a colour map. When one wants to emphasise the phase behaviour and is not so interested in the actual amplitude of the signal, one can normalise $g(t_i)$ values in each bin. These diagrams become useful as they make it easy to spot phase intermittencies and long term drifts with respect to the carrier period. Here we can notice the close relation to the ideas of instantaneous frequency and amplitude discussed later in Sect. 5.1.

In Fig. 4.3 we show example phase diagrams for an artificial quasi-periodic time series with a covariance function given by Eq. (4.1), where $P = 5$ and $\ell = 100$. On the left panel, we see that the maximum amplitude is reached around $t = 1500$, while on both panels, long frequency drifts are visible from $t = 500$ to $t = 1000$.

For a comparison to synthetic data, in Fig. 4.4, we include a plot from Lindborg et al. [89], where we show the phase diagram of an active star II Peg using the known rotation period $P_{\mathrm{rot}} = 6^{\mathrm{d}}.72433$. Persistent long downward trends in the phase diagram clearly indicate that the active regions on the surface of the star rotate with faster period than P_{rot}. The plausible explanation for this is the existence of an azimuthal dynamo wave.

In Paper I, more examples are given about the visualisation, including plotting the instantaneous frequency of the signal. In Paper II, we used the visualisation technique to illustrate the changes in the light curve of

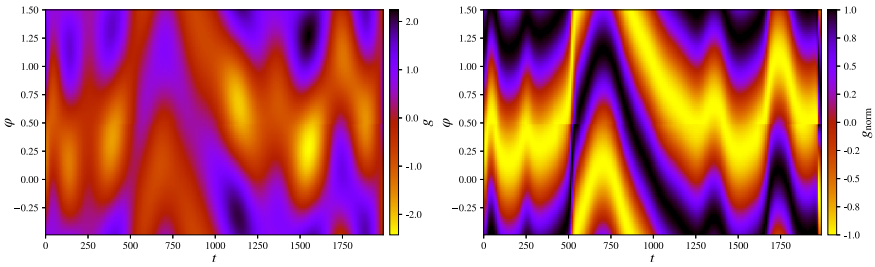


Figure 4.3. Phase diagrams of a quasi-harmonic time series with $P = 5$ and $\ell = 100$. Left: Unnormalised phase diagram. Right: Normalised phase diagram.

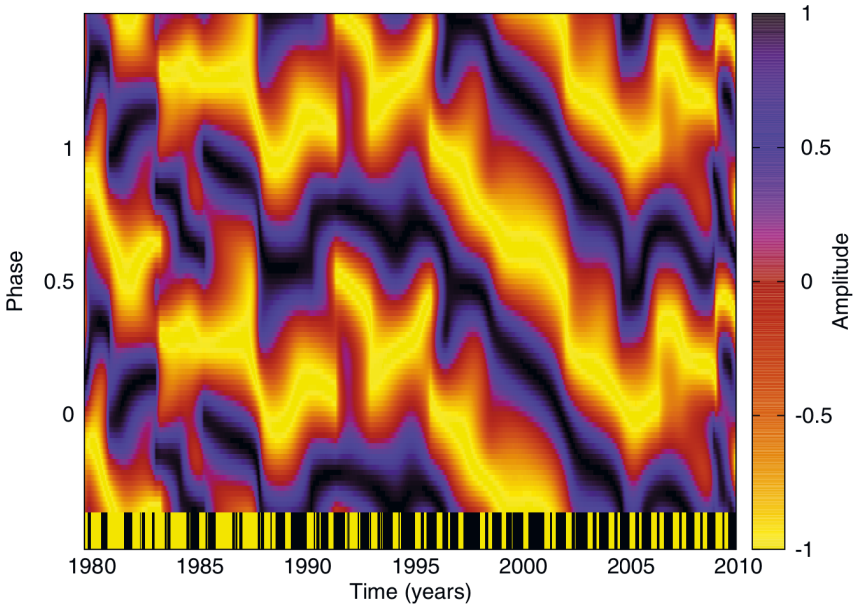


Figure 4.4. Phase diagram of II Peg with rotation period $P = 6^{\mathrm{d}}.72433$. There is a clear linear downward trend visible indicating faster rotation of the active regions on the surface of the star. The 'barcode' in the bottom shows the data sampling.

another active star LQ Hya.

4.2 Continuous period search method

Consider a process which consists of a sum of L locally periodic processes in the subintervals discussed in the previous section. We assume that the period of these local processes is a random variable from some distribution, in the simplest case Gaussian. We would be interested in estimating the mean and variance of this distribution. An empirical method suitable for this purpose, called continuous period search (CPS), was developed by

Lehtinen et al. [86]. This method consists of fitting periodic models to the small segments of data whose widths are set based on a priori estimate P_0 of the mean period:

$$y(t) = M + \sum_{k=1}^{K} \big(a_k \cos(2\pi k f t) + b_k \sin(2\pi k f t)\big) + \epsilon(t), \tag{4.6}$$

where K is the number of overtones in the model, M is the mean of the model, a_k and b_k are the amplitudes of the harmonic components and $\epsilon(t) \sim \mathcal{N}\big(0, \sigma^2(t)\big)$ is the zero mean Gaussian independent noise. The initial period estimate, P_0, can be taken from literature, if available, or obtained by using any of the other period search methods discussed in this chapter. The width of the segment needs to be long enough to cover at least two periods, but short enough for the constant period model to hold. The analysis is repeated while gradually shifting the segment in time until the whole time span of the observations is covered. The grid search is performed for the period interval from $(1-q)P_0 \leq P \leq (1+q)P_0$, where q is a small number regulating the interval of the search. The simplest way to estimate the optimal value for K in each segment is to find the value corresponding to the smallest BIC. The mean period is calculated as the weighted average of the independent period estimates (estimated from nonoverlapping segments). However, the sample variance of the independent period estimates does not directly correspond to the variance of the mean period. The sample variance is highly dependent on the average S/N level of the data and can be nonzero even for the case of constant period, that is, when the true variance is zero [86].

Note the difference between the CF and CPS methods. While in the CF method the model is fit globally and the modulators $a_k(t)$ and $b_k(t)$ are time dependent, in the CPS method the model is fit locally and a_k and b_k are constant. Both CF and CPS methods essentially belong to the category of time-frequency analysis tools as they are similar to short-time Fourier transform, which we will discuss in Sect. 5.4. For other similar methods, please refer to [3, 95, 51].

Although the process described in this section is not strictly speaking quasi-periodic by our definition, if we assume that the distribution of periods has a finite support and the jumps between the periods are not too abrupt in time, we can still effectively describe it with a quasi-periodic

covariance function. These assumptions are quite realistic in the case of active stars, as the range of periods is limited in Eq. (2.1) and the spots or spot groups do not emerge and decay instantaneously. Therefore, other methods discussed in the current chapter become similarly useful for modelling these kinds of time series.

4.3 D^2 statistic for cyclic data

In Sect. 3.5 we defined the D^2 statistic with selection function g depending on the period, however, we can introduce an additional dependence on the time-scale, so that

$$g(t, t', f, \ell) = g_1(t, t', f) g_2(t, t', \ell), \tag{4.7}$$

where g_1 is the phase selection term and g_2 is the time selection term, both depending on t and t' only through the time difference $\tau = |t - t'|$. Furthermore, we assume that g_2 is significantly greater than zero only if $\tau \lessapprox \ell$. The intuition behind this kind of a selection function is to allow only those pairs of points to enter the statistic, whose time difference is not significantly greater than ℓ. With these modifications, after dropping the factor N from Eq. (3.44) and normalising by the sample variance σ^2, the D^2 statistic becomes

$$D^2(f) = \frac{1}{\sigma^2} \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} g(t_i, t_j, f, \ell) \big(y(t_i) - y(t_j)\big)^2}{\sum_{i=1}^{N} \sum_{j=i+1}^{N} g(t_i, t_j, f, \ell)}, \tag{4.8}$$

where $g(t, t', f, \ell)$ is defined in Eq. (4.7). In vector form, the Eq. (3.45) still holds, the only difference being that the elements of $\mathbf{G}$ now depend both on f and ℓ.

To open up the idea behind the D^2 statistics further, let us consider the case of evenly sampled data with time step $\tau = t_2 - t_1$, assuming that the data points have been ordered in time. For simplicity, let us also assume that N is odd. Then the matrix $\mathbf{G}$ in Eq. (3.45) is a symmetric Toeplitz matrix and $\mathbf{Gy}$ represents a convolution of $\mathbf{g}$ and $\mathbf{y}$, where $\mathbf{g}$ is the middle column of matrix $\mathbf{G}$ with close to zero elements removed (this is a justified approximation for $\ell \ll t_N - t_1$ because we assume that the selection function g_{ij} is small if $|t_i - t_j| > \ell$). Due to this restriction the

first term in the numerator of Eq.(3.45) is approximately proportional to $\mathbf{y}^\mathsf{T}\mathbf{y}$ thus not carrying any information regarding cyclic features in the data. This information is solely contained in the second term of the same equation. If we define matrix $\mathbf{Y}$ containing shifted and truncated versions of columns $\mathbf{y}$, that is, if $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$, then $\mathbf{Y} = \begin{bmatrix} y_2 & y_1 & 0 \\ y_3 & y_2 & y_1 \\ 0 & y_3 & y_2 \end{bmatrix}$, and due to the symmetry of the convolution, $\mathbf{G}\mathbf{y} = \mathbf{Y}^\mathsf{T}\mathbf{g}$. Consequently, the second term in the numerator of Eq. (3.45) becomes $\mathbf{g}\mathbf{Y}^\mathsf{T}\mathbf{y}$. If we assume zero mean, then $\mathbf{Y}^\mathsf{T}\mathbf{y} = N\hat{\mathbf{r}}$, where $\hat{\mathbf{r}} = [\hat{r}_{-(N-2)} \dots \hat{r}_{N-2}]^\mathsf{T}$ and $\hat{r}_k$ is from Eq. (3.24). We see that in this special case of evenly sampled data, D^2 statistic is proportional to the weighted sum of covariances $\hat{r}(\tau_i)$ at time lags $\tau_i = i\tau$, the weight being the selection function $g(\tau_i)$: $\sum_i g(\tau_i)\hat{r}(\tau_i)$. Such a natural interpretation of the D^2 statistic does not, however, directly hold for the data with uneven sampling.

For g_2 one can consider either a Gaussian or a boxcar function. It is natural to combine the former with a cosine and the latter with a boxcar selection function for the phase, respectively:

$$g = \big(1 + 2\cos(2\pi f(t_i - t_j))\big) e^{-\frac{(t_i - t_j)^2}{2\ell^2}} \tag{4.9}$$

and

$$g = \mathbb{I}\left(\min\big(\Delta\varphi(\mathrm{i},\mathrm{j}), 1 - \Delta\varphi(\mathrm{i},\mathrm{j})\big) < \epsilon\right) \mathbb{I}(|t_i - t_j| < \ell), \tag{4.10}$$

where the meaning of $\Delta\varphi(i,j)$ is the same as in Eq. (3.49).

In the case of continuous stationary process, it can be shown that the D^2 statistic with Gauss-cosine selection function becomes identical to the convolution of the power spectrum with a Gaussian, whose width is inversely proportional to ℓ, while in the case of boxcar selection function the filtering of the covariance function analogy from Sect. 3.5.2 remains the same, only difference being that the covariance function is now truncated at ℓ.

In Fig. 4.5 we show the D^2 spectrum for the same dataset as was used in the example shown in Fig. 4.2. We see that for the longer time-scales used in the statistic, the spectrum becomes closer to the LS spectrum, while for shorter time-scales the spectrum becomes more smooth and the bias between the true frequency and estimated frequency gets smaller.

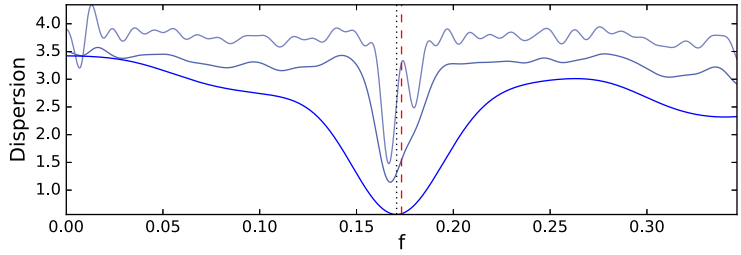


Figure 4.5. D^2 spectrum calculated for the same time series as in Fig. 4.2. Red dashed line shows the position of the true frequency and the black dotted line the estimated frequency. Different curves correspond to the statistics calculated for different time-scales.

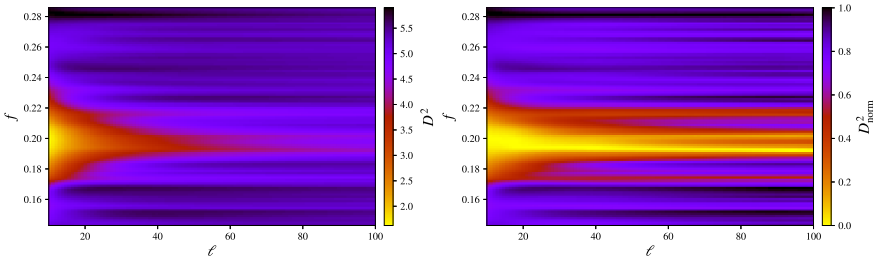


Figure 4.6. D^2 spectra of another quasi-harmonic time series with $f = 0.2$ and $\ell = 20$. Left: Unnormalised spectrum. Right: Normalised spectrum.

Instead of plotting one-dimensional spectra separately for each time-scale, it is more informative to plot the spectra as a single colour map as shown in Fig. 4.6. In this case the true cycle period was chosen five and the time-scale 20. The total time duration of the time series was 2000. On the left panel of this figure the D^2 spectrum is plotted without any normalisation. We see that the spectrum gets weaker towards longer length scales. To make the spectrum equally prominent for all time scales, it is sometimes useful to plot the colour spectrum in a normalised manner. This is shown on the right panel of the same figure. From both plots it is evident, that the spectrum becomes multimodal (i.e. splits) around $\ell \gtrapprox 20$. This is a clear indication of the quasi-periodic nature of the signal.

One relevant question is, whether, and how well one could use D^2 statistic to infer the time-scale of the cyclic process. Obviously there must be a link between the time-scale of the selection function and the time-scale of the process, but the explicit form of this link is not known at the moment. However, our tests have shown that in many cases the rough estimate of the time-scale can be obtained by finding the location on the D^2 spectrum where the unimodality changes to multimodality.

The generalisation of the D^2 statistic introduced here is practical also for a strictly periodic time series. As the time-scale parameter acts as a smoothing factor, one can significantly reduce the number of trial frequencies at the first stage of the period search, and find only the approximate locations of the highest peaks. On the second stage, one performs the finer grid search with any other period estimation method around the local minima found during the first stage. Finally, in the third stage, one can try to optimise all parameters in the model together alongside with the frequency to get the final estimates. This method is known as a three stage period analysis [69].

4.4 Probabilistic methods for cycle detection

In Sect. 3.7.3 we discussed a probabilistic period estimation method, which was based on Bayesian linear regression using harmonic functions. In principle one can use the same analogy and build a probabilistic model for the CF method introduced in Sect. 4.1. A more general and flexible way to accomplish the same task, however, is to use Gaussian processes (GPs), which we will discuss next.

4.4.1 Gaussian processes

GP is a flexible model used in statistics and machine learning for many purposes, including regression. In this section we only focus on the regression from the point of view of period estimation. Shortly put, the idea behind the GP is that the covariance structure of the data points is assumed to be known and the distribution to be Gaussian. The latter condition makes the inference simple. In a GP all the unknown parameters are captured inside the covariance function and it is common to call them hyperparameters.

More formally, GP is a collection of random variables which have a joint Gaussian distribution, and it is completely defined by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ [111]. GP is formally written as

$$y(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \tag{4.11}$$

where

$$\begin{aligned} m(\mathbf{x}) =& \mathbb{E}[y(\mathbf{x})], \\ k(\mathbf{x}, \mathbf{x}') =& \mathbb{E}[(y(\mathbf{x}) - m(\mathbf{x}))(y(\mathbf{x}') - m(\mathbf{x}'))]. \end{aligned} \tag{4.12}$$

Eq. (4.11) can be understood as a probability distribution over functions.

In the following, for convenience, we write t instead of x as we are considering only time series. For a finite set of random variables, the GP forms a multivariate Gaussian given by:

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}), \tag{4.13}$$

where $\mathbf{y} \in \mathbb{R}^N$ is the vector of observables, $\boldsymbol{\mu} \in \mathbb{R}^N$ is the prior mean vector (usually assumed to be zero) and $\mathbf{K} \in \mathbb{R}^{N \times N}$ is the prior covariance matrix whose elements depend only on the pairs of time moments, that is, $K_{ij} = k(t_i, t_j)$, where k is a covariance function of the process. N is the number of data points. The log likelihood of the data can be expressed as:

$$\log p(\mathbf{y}) = -\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}| - \frac{N}{2}\log 2\pi. \tag{4.14}$$

Maximising this quantity leads to the optimal hyperparameters defining the form of the covariance function k. If we assume the covariance function to be quasi-periodic, this can be used to estimate the period of the process.

Quasi-periodic covariance function is usually defined as a product between a periodic covariance function k_{P} and a more slowly changing covariance function k_{Q}, which decays as function of $|t - t'|$. k_{P} is given by

$$k_{\mathrm{P}}(t, t') = \exp\left(-\frac{2\sin^2\left(\pi f(t - t')\right)}{\ell_{\mathrm{P}}^2}\right), \tag{4.15}$$

where ℓ_{P} and $f = 1/P$ are the time-scale and frequency of the periodic component. ℓ_{P} in the equation defines how much the form of the signal can deviate from the sine wave. The longer the ℓ_{P}, the more harmonic the process is and vice versa. One frequently used option for k_{Q} is the squared exponential covariance function

$$k_{\mathrm{Q}}(t, t') = \exp\left(-\frac{(t - t')^2}{2\ell_{\mathrm{Q}}^2}\right), \tag{4.16}$$

where ℓ_{Q} is the time-scale of the squared exponential component. Both of these equations describe the processes with unit variance. Quasi-periodic covariance function can then be written as

$$k_{\mathrm{QP}}(t,t') = \sigma_{\mathrm{S}}^2 k_{\mathrm{Q}}(t,t') k_{\mathrm{P}}(t,t'), \tag{4.17}$$

where σ_{S}^2 is the signal variance. ℓ_{Q} defines how coherent the periodicity of the process is. Longer values of ℓ_{Q} correspond to more strict periodicity and vice versa. For very low values of ℓ_{Q} the process becomes fully stochastic.

Usually we only observe noisy measurements so that the final covariance matrix must take that into account. Assuming Gaussian independent noise

$$k(t,t') = k_{\mathrm{QP}}(t,t') + \mathbb{I}(t=t')\sigma_{\mathrm{N}}^2(t), \tag{4.18}$$

where $\mathbb{I}$ is the indicator function and σ_{N}^2 is the noise variance.

In the beginning of the chapter we gave an example of a quasi-harmonic process with the covariance function

$$k_{\mathrm{QH}}(t,t') = \sigma_{\mathrm{S}}^2 k_{\mathrm{Q}}(t,t') k_{\mathrm{H}}(t,t'), \tag{4.19}$$

where

$$k_{\mathrm{H}}(t,t') = \cos\big(2\pi f(t-t')\big). \tag{4.20}$$

In fact, k_{H} is obtained from k_{P} by taking the limit $\ell_{\mathrm{P}} \to \infty$. This form of covariance function we used in a GP regression in Paper VI.

The GP model scales as $\mathcal{O}(N^3)$ due to the need for inverting the covariance matrix. To make GPs feasible for larger datasets, different approximation methods have been suggested. In the following we mention some of these methods. Downsampling of the data, which may seem the simplest of the options, can be used when the estimated time-scale and the period of the process are much larger than the resulting average time step. In another method, called a Nyström approximation the original covariance matrix $\mathbf{K}$ is approximated by $\tilde{\mathbf{K}} = \mathbf{K}_{NM}\mathbf{K}_{MM}^{-1}\mathbf{K}_{MN}$, where $M < N$, such that $\mathbf{K} = \begin{bmatrix} \mathbf{K}_{MM} & \mathbf{K}_{M(N-M)} \\ \mathbf{K}_{(N-M)M} & \mathbf{K}_{(N-M)(N-M)} \end{bmatrix}$ and $\mathbf{K}_{MN} = \mathbf{K}_{NM}^{\mathsf{T}}$ is the top left $M \times N$ block of $\mathbf{K}$. This approximation reduces the computational cost to $\mathcal{O}(M^2 N)$. The Nyström method is very closely related to

another method called the subset of regressors. In the context of stellar cycle search, the subset of M time moments can be selected, for instance, as the centres of the observational seasons. Instead of directly downsampling, one can greedily select data points into the active set by maximising the differential entropy score at site j before and after the inclusion of the datapoint. This is equivalent to finding a site with the largest variance. This method is known as the subset of datapoints [111, Ch. 8]. A more elaborate approximation method based on variational inference, where the subset of inducing inputs and model parameters are simultaneously optimised was introduced in Titsias [140].

For model selection, leave-one-out cross-validation (LOO-CV) is recommended over the Bayes factor [150]. In the case of LOO-CV, one calculates the sum of predictive log probabilities, leaving out one data point at a time [111, Ch. 5].

While GP can be used for period estimation of the fully harmonic time series, for instance, by using cosine covariance function, due to the scalability issues, other methods discussed in the previous chapter are more preferable in practice.

4.4.2 Probabilistic analogue of D^2 statistic

For simplicity, throughout this section we assume that the mean of the process is zero. Instead of Eq. (4.13), let us use pairwise factor graph model, such that

$$p(\mathbf{y}) = \prod_{\substack{i,j \\ i \neq j}} p_2(y_i, y_j), \tag{4.21}$$

where we assume $p_2 \left(\begin{bmatrix} y_i \\ y_j \end{bmatrix} \right) = \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K_{ii} & K_{ij} \\ K_{ij} & K_{jj} \end{bmatrix} \right)$, where $\sigma_i^2 = K_{ii}$ and K_{ij} is an element of the full covariance matrix $\mathbf{K}$ defined in the previous section. The idea behind Eq. (4.21) is that we sample the Gaussian process $N(N-1)$ times using pairs of time moments $(t_i, t_j), i, j = 1 \dots N, i \neq j$, and average over the log likelihoods. The full log probability of the data is given by

$$\log p(\mathbf{y}) = -\mathbf{1}^\mathsf{T} \mathbf{G}^{(1)} (\mathbf{y} \circ \mathbf{y}) + \mathbf{y}^\mathsf{T} \mathbf{G}^{(2)} \mathbf{y} - \frac{1}{2} G_{\log}, \tag{4.22}$$

where

$$G_{ij}^{(1)} = \begin{cases} \frac{\sigma^2}{\sigma^4 - Kij^2} & \text{if } i \neq j \\ c & \text{if } i = j, \end{cases} \tag{4.23}$$

$$G_{ij}^{(2)} = \begin{cases} \frac{K_{ij}}{\sigma^4 - Kij^2} & \text{if } i \neq j \\ c & \text{if } i = j \end{cases} \tag{4.24}$$

and

$$G_{\log} = \sum_{ij} \log(\sigma^4 - K_{ij}^2), \quad i \neq j, \tag{4.25}$$

where c is any constant. Also, $\mathbf{G}^{(2)} = \mathbf{G}^{(1)} \circ \mathbf{K}/\sigma^2$.

The functional forms of the D^2 statistic given by Eq. (3.45) is quite similar to the one of Eq. (4.22). That is the reason why we call it a probabilistic analogue of D^2 statistic. Now it becomes more evident that in the definition of the D^2 statistic, the purpose of the selection function g is explained by the need to imitate the covariance structure of the data, however, the form of g given, for instance, by Eq. (4.10) does not exactly coincide with k given by Eq. (4.19).

As the evaluation of $\log p(\mathbf{y})$ in Eq. (4.22) is quadratic with respect to N, it can be used as a practical approximation to the full GP solution. However, obviously there is no guarantee for the solution to coincide with the full GP one.

4.4.3 State space methods

Other well known alternatives to GPs are the state space methods. The huge advantage of these methods compared to GPs is that they scale linearly with the number of data points. However, the downside is that not all covariance functions are convertible to the state space form, having finite number of terms. On the other hand, in practical situations, one can use as precise approximations as needed. Here we consider a special case of the state space methods, known as the Kalman filter, which is the closed form solution to the Bayesian filtering equations, where the dynamic and measurement models are linear Gaussian [124, Ch. 4.3]:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{q}_{k-1}, \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{x}_k + \mathbf{r}_k, \quad k = 1 \dots K, \end{aligned} \tag{4.26}$$

where $K+1$ is the number of data points, $\mathbf{x}_k \in \mathbb{R}^n$ is the vector of (hidden) state, $\mathbf{y}_k \in \mathbb{R}^m$ is the vector of measurement (we assume $m = 1$), $\mathbf{q}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1})$ is the process noise, $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ is the measurement noise, $\mathbf{A}_{k-1}$ is the transition matrix of the dynamic model and $\mathbf{H}_k$ the measurement model matrix. The distribution for the initial state is taken as $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0)$. Kalman filter corresponds to a Markov chain, where the state at step k is conditionally dependent only on the previous state at step $k-1$ and the measurement at step k only depends on the state at step k:

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{x}_{k-1}) &= \mathcal{N}(\mathbf{x}_k|\mathbf{A}_{k-1}\mathbf{x}_{k-1}, \mathbf{Q_{k-1}}), \\ p(\mathbf{y}_k|\mathbf{x}_k) &= \mathcal{N}(\mathbf{y}_k|\mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k). \end{aligned} \tag{4.27}$$

As with other methods, we are interested in maximising the likelihood of the data $p(\mathbf{y}_{1:K}|\boldsymbol{\theta})$[3], where $\boldsymbol{\theta}$ is the vector of parameters such as the period and time-scale. The likelihood can be factorized as follows:

$$p(\mathbf{y}_{1:K}|\boldsymbol{\theta}) = \prod_{k=1}^{K} p(\mathbf{y}_k|, \mathbf{y}_{1:k-1}, \boldsymbol{\theta}). \tag{4.28}$$

The equations for the Kalman filter give closed form solution for calculating $p(\mathbf{y}_k|, \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$, namely

$$p(\mathbf{y}_k|, \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{y}_k|\mathbf{H}_k\mathbf{m}_k^-, \mathbf{S}_k), \tag{4.29}$$

where $\mathbf{H}_k$, $\mathbf{m}_k^-$ and $\mathbf{S}_k$ can be calculated iteratively starting from $k = 1$ [124, Ch. 4.3].

To know the state transition matrix $\mathbf{A}_k$ and noise covariance matrix $\mathbf{Q}_k$, one needs first to know the feedback matrix $\mathbf{F}$ and the spectral density $\mathbf{Q}_c$ of the white noise $\mathbf{w}(t)$ in the stochastic differential equation defining the Markov process:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}\mathbf{x}(t) + \mathbf{L}\mathbf{w}(t), \tag{4.30}$$

where $\mathbf{L}$ is the noise effect matrix. Knowing $\mathbf{F}$, $\mathbf{A}_k$ can be calculated as follows

$$\mathbf{A}_k = \Phi(\Delta t_k), \tag{4.31}$$

where $\Phi(\tau) = \exp(\mathbf{F}\tau)$ and $\Delta t_k = t_k - t_{k-1}$. Here $\exp$ means the matrix

[3] $\mathbf{y}_{1:K}$ stands for the set $(\mathbf{y}_1, \ldots, \mathbf{y}_K)$.

exponential. The covariance of the initial state $\mathbf{P}_0$ (usually written as $\mathbf{P}_\infty$ instead) is taken as the stationary covariance of $\mathbf{x}(t)$, which can be calculated from the following Lyapunov equation:

$$\mathbf{F}\mathbf{P}_0 + \mathbf{P}_0\mathbf{F}^\mathsf{T} + \mathbf{L}\mathbf{Q}_\mathrm{c}\mathbf{L}^\mathsf{T} = \mathbf{0}. \tag{4.32}$$

The noise covariance matrix can be calculated from

$$\mathbf{Q}_k = \int_0^{\Delta t_k} \Phi(\Delta t_k - \tau)\mathbf{L}\mathbf{Q}_\mathrm{c}\mathbf{L}^\mathsf{T}\Phi(\Delta t_k - \tau)^\mathsf{T} d\tau, \tag{4.33}$$

which is quite inefficient to use directly. However, when combining the last equation with Eq. (4.32), one gets a more practical formula for calculating $\mathbf{Q}_k$:

$$\mathbf{Q}_k = \mathbf{P}_0 - \mathbf{A}_k\mathbf{P}_0\mathbf{A}_k^\mathsf{T}. \tag{4.34}$$

The mean $\mathbf{m}_0$ of the initial state is usually taken zero.

Now we consider a Kalman filter which corresponds to the Gaussian process with the quasi-periodic covariance function given by Eq. (4.17). To calculate the matrices $\mathbf{F}$, $\mathbf{Q}_\mathrm{c}$, $\mathbf{L}$, $\mathbf{P}_0$ and $\mathbf{H}$, one needs first to know the corresponding matrices for the squared exponential term and periodic term separately. The derivation of these quantities for the squared exponential covariance function was given by Hartikainen and Särkkä [58]. Usually a sufficient precision is achieved when order $N = 6$ is used in the approximation. Both, the derivation of the quantities for the periodic covariance function and the recipe for combining them with the quantities for the squared exponential covariance, to form the final quantities was given by Solin and Särkkä [134]. The level of non-harmonicity controlled by ℓ_P in the GP model is here controlled by the number of harmonics, J.

Now all the quantities, including the covariance $\mathbf{R}_k$ of measurement noise, are known, so we can apply the Kalman filter to the data. What is of primary importance to us, the Eq. (4.28) can now be evaluated for each trial frequency, so that we can form a probabilistic spectrum.

4.5 Summary

Application of the methods developed for quasi-periodic time series is the central part of the given thesis. In particular, the D^2 statistic, which

we have developed and explored, is a very practical and efficient alternative to GPs in many situations. However, the mathematical properties and limitations of D^2 statistic are still not fully known. In the current and previous chapter we opened up the intuition behind the statistic using evenly sampled time series as an example and explored its properties from the probabilistic perspective. Another scalable alternative to GPs is the Kalman filter, which is a method developed many decades ago. Since then it is thoroughly investigated and frequently used in many domains, for instance, in control theory. Interestingly, in period and cycle search related to astronomical datasets the full power of this method is yet to be exploited.

5. Methods for non-stationary time series

By far not all processes in the nature are stationary. More specifically in the context of astrophysical systems, periodic fluctuations often arise intermittently, as transient phenomena, but even for a time series with consistent periodicity usually time evolution of other parameters (e.g. amplitude) are seen [48]. For instance, in the case of the Sun, it is known that during the period around 1645 to 1715 solar activity, measured via counting the sunspots, vanished. This is known as the Maunder minimum[1]. Also many other such grand minima are known in the solar activity indices. Other typical examples of non-stationary time series would be, for instance, the recorded speech or sound signals. More specifically, by non-stationarity in this chapter we primarily mean that the power spectrum of the process is changing over time. This does not, however, mean that the process cannot be WSS at the same time. In other words, for the signal of a stationary process one would expect the power spectra to be similar regardless of whether it is calculated, for instance, for the first half of the time series or for the second half. Moreover, with ergodicity one would expect to get consistent results with increasing number of data points. For non-stationary signals this is not so, as the characteristic properties of the signal, such as the amplitude, frequency range etc. are time dependent [109, Ch. 11.1].

In practice often methods assuming stationarity are in fact applied to non-stationary data. Sometimes it does not matter, given that we know how to interpret the results. For instance, if the first half of the time series contains a harmonic with one frequency and the second half a har-

[1]The cycle is known to exist during the Maunder minimum, but at the level below or comparable to the sunspot formation threshold [143]

monic with another frequency, the power spectrum will yield both frequencies. However, as the phase information is lost, we cannot any more infer whether these two frequencies were persistent throughout the data or only occasionally. To give another example, if we would calculate the D^2 spectrum for the chirp signal, we would get a minimum around the mean frequency. Without knowing that the data is non-stationary, we would be prone to make a conclusion that there is a cycle of given frequency. The pure Fourier spectrum of a non-stationary signal carries of course also the phase information, but it is still virtually useless in helping to make direct conclusions which frequencies are dominant at which times.

To analyse time series of non-stationary processes, many different methods have been developed, which are mainly known as time-frequency or multiresolution methods. These methods usually assume even sampling of the data, which makes them less suitable for ground-based astronomical observations. However, for satellite measurements as well as for simulated data they are still very valuable.

In the following we will consider a signal $y(t)$ with finite energy. The main idea of the time-frequency methods is to seek either for suitable joint spectrum $Y(t, f)$ or spectral density distribution $P(t, f)$. The former are so-called linear methods and the latter quadratic methods. Meaningful spectral density distribution $P(t, f)$ should satisfy the total energy requirement

$$E = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(t, f) df dt, \tag{5.1}$$

where E was defined in Eq. (3.8), as well as the time and frequency marginals:

$$\int_{-\infty}^{\infty} P(t, f) df = |y(t)|^2, \tag{5.2}$$

$$\int_{-\infty}^{\infty} P(t, f) dt = |Y(f)|^2. \tag{5.3}$$

Naturally one would want $P(t, f)$ to satisfy also the so-called finite-time and -frequency criteria, meaning that if $y(t)$ is zero outside the interval $[t_1, t_2]$, so is $P(t, f)$ and similarly for $\mathcal{F}(y)$. As there exist infinite number of joint distributions satisfying the above conditions, the solution to the problem, how to construct for a given signal $y(t)$ the distribution $P(t, f)$, does not have a mathematical origin. Solving it depends on particular goals one wants to achieve [32, Ch. 6.10]. Before we turn to the discussion

of suitable distributions, let us give a definition of the analytic signal.

5.1 Analytic signal

The analytic signal $z(t)$ of a real signal $y(t)$ is a complex signal formed from the positive part of the Fourier spectrum of the real signal, which is multiplied by two:

$$z(t) = 2\int_0^\infty Y(f)e^{2\pi i f t}df, \tag{5.4}$$

where $Y = \mathcal{F}(y)$. In the time domain this is identical to

$$\mathcal{A}(y)(t) = z(t) = y(t) + \frac{i}{\pi}\int_{-\infty}^{\infty}\frac{y(t')}{t-t'}dt' = y(t) + iH(y)(t), \tag{5.5}$$

where H is called the Hilbert transform. The improper integral needs to be understood in the principal value sense[2]. The analytic signal is useful, because its mean frequency and frequency spread can be naturally defined as $\mu_f = \int_{-\infty}^{\infty} fP(f)df$ and $\sigma_f^2 = \int_{-\infty}^{\infty}(f-\mu_f)^2P(f)df$.

If $z(t) = A(t)e^{i\varphi(t)}$ is an analytic signal then the spectrum of $A(t)$ is between $(-\omega_1, \omega_1)$ and the spectrum of $e^{i\varphi(t)}$ is zero for $\omega \leq \omega_1$, that is, the low frequency component is contained in the amplitude term and the high frequency component in the exponential term. The instantaneous frequency of the analytic signal can be defined as $\omega_{\text{inst}}(t) = \frac{d\varphi(t)}{dt}$. However, despite this intuitive formula, the instantaneous frequency does not always have a straightforward physical interpretation. It can have negative values, not match with any true frequencies of the signal, and lie outside the bandwidth of the signal [32, Ch. 2.7]. An exception is a chirp signal, for which it has the correct physical meaning.

5.2 Cohen's class of distributions

A very useful tool in time-frequency analysis is the Cohen distribution (CD), or rather a class of distributions:

$$C_y(t,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(\nu,\tau)y(u+0.5\tau)y^*(u-0.5\tau)e^{i(\nu u-\nu t-\omega\tau)}dud\nu d\tau, \tag{5.6}$$

[2] $\int_{-\infty}^{\infty}\frac{y(t')}{t-t'} = \lim_{\epsilon\to 0^+}\left(\int_{-\infty}^{-\epsilon}\frac{y(t')}{t-t'} + \int_{\epsilon}^{\infty}\frac{y(t')}{t-t'}\right)$.

where y is the input signal and $\phi(\nu, \tau)$ is the kernel function of the distribution. By selecting different kernel functions one can construct different special cases of the CD. For instance by selecting $\phi(\nu, \tau) = 1$ one gets the Wigner-Ville distribution (WVD):

$$W_y(t, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} y(t + 0.5\tau) y^*(t - 0.5\tau) e^{-i\omega\tau} d\tau. \tag{5.7}$$

The kernel function needs to satisfy certain criteria for the C_y to be useful in the analysis. To highlight some of them, then C_y is real only if $\phi(\nu, \tau) = \phi^*(-\nu, -\tau)$. It satisfies the time-marginal property Eq. (5.2) only if $\phi(\nu, 0) = 1$, for all ν, and the frequency-marginal property Eq. (5.3) only if $\phi(0, \tau) = 1$, for all τ. The total energy requirement Eq. (5.1) is satisfied only if $\phi(0, 0) = 1$.

Another property is related to the instantaneous frequency of the signal. It can be shown that $\omega_{\text{inst}}(t)$ can be calculated as $\omega_{\text{inst}}(t) = \frac{\int_{-\infty}^{\infty} \omega C_y(t,\omega) d\omega}{\int_{-\infty}^{\infty} C_y(t,\omega) d\omega}$, given that $\phi(\nu, 0) = 1$ and $\partial\phi(\nu, \tau)/\partial\tau|_{\tau=0} = 0$, for all ν. C_y satisfies the finite-time support criterion if $\int_{-\infty}^{\infty} \phi(\nu, \tau) e^{-i\nu t} d\nu = 0$ for $|t/\tau| > 0.5$, and the finite-frequency support criterion if $\int_{-\infty}^{\infty} \phi(\nu, \tau) e^{-i\omega\tau} d\tau = 0$ for $|\omega/\nu| > 0.5$. WVD satisfies all the above criteria. For more of the criteria and proofs, please refer to Cohen [32, Ch. 9.4] and Porat [109, Ch. 12.5].

There are several issues with CD, first one being that it is not always guaranteed to be nonnegative. The second one is the appearance of the so-called cross terms, that is, if the input signal is a sum of many signals, then due to nonlinearity, the resulting CD is not the sum of the individual CDs, but the cross CDs appear. The cross terms may carry misleading information, for instance, making the spectrum nonzero (within the outer limits) where it is not supposed to be. The primary task of the kernel function is to reduce the effect of the cross terms. For that purpose the kernel needs to have certain properties on top of the ones discussed above. More details can be found from Cohen [31, 32], Porat [109].

5.3 Time-frequency distributions for stochastic processes

An extension of the Wiener-Khinchin theorem in Eq. (3.11) for non-stationary processes can be written as

$$P(t,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} r(t,t+\tau)e^{-i\omega\tau}d\tau, \tag{5.8}$$

where $P(t,\omega)$ is called a time-varying spectral density [123]. The Fourier transform of the latter, which is a function of two frequencies, is known as the generalised spectral density.

For a stochastic process $y(t)$ one can define CD as $\bar{C}_y(t,\omega) = \mathbb{E}[C_y(t,\omega)]$, where the expectation is taken over the realisations of the process (i.e. ensemble averaging). One can then write

$$\begin{aligned}\bar{C}_y(t,\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(\nu,\tau)\mathbb{E}[y(u+0.5\tau)y^*(u-0.5\tau)]e^{i(\nu u-\nu t-\omega\tau)}dud\nu d\tau \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(\nu,\tau)r_{\rm sym}(t,\tau)e^{i(\nu u-\nu t-\omega\tau)}dud\nu d\tau,\end{aligned} \tag{5.9}$$

where $r_{\rm sym}(t,\tau) = \mathbb{E}\big(y(t+0.5\tau)y^*(t-0.5\tau)\big)$ is the symmetric covariance function [32, Ch. 13.11]. In particular, for WVD we get

$$\bar{W}_y(t,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} r_{\rm sym}(t,\tau)e^{-i\omega\tau}d\tau, \tag{5.10}$$

which is close to Eq. (5.8), with the difference between $r_{\rm sym}$ and r [123].

While repeated realisations can be achieved in laboratory environment, in astronomical time series context, it is impossible. Thus, the definition of $\bar{C}_y(t,\omega)$ has more theoretical, than practical interest to us.

The Cohen class of distributions discussed above belong to the so-called quadratic methods. Now we turn to the discussion of some of the linear methods.

5.4 Short-time Fourier transform

For the continuous signal, short-time Fourier transform (STFT), also known as windowed Fourier transform, is defined as

$$Y(t,\omega) = \int_{-\infty}^{\infty} y(t')w^*(t'-t)e^{-i\omega t'}dt', \tag{5.11}$$

where $y(t)$ is the measured signal and $w(t)$ is the (most often real) window function usually satisfying

$$w(t) \propto \begin{cases} 1 & \text{for } t \text{ close to } 0 \\ 0 & \text{for } t \text{ far away from } 0, \end{cases} \tag{5.12}$$

and it should be centred around zero, that is, $\int_{-\infty}^{\infty} t|w(t)|^2 dt = 0$ [109, Ch. 11.2].

The power spectrum at t is

$$P(t, \omega) = |Y(t, \omega)|^2, \tag{5.13}$$

which is called the spectrogram. The spectrogram is another special case of CD, whose kernel is the ambiguity function of the window function: $\phi(\nu, \tau) = \int_{-\infty}^{\infty} w(u + 0.5\tau)w^*(u - 0.5\tau)e^{i\nu u}du$. A desirable property of the spectrogram is its guaranteed nonnegativity. Also, its cross spectra are reduced in comparison to WVD [32, Ch. 8.11].

If the energy of the window $\int_{-\infty}^{\infty} |w(t)|^2 dt = 1$, then the spectrogram satisfies the total energy requirement, however, it does not generally satisfy the marginals, because the energy distribution of the signal is scrambled with the one of the window [32, Ch. 7.3].

The inverse STFT is defined as

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Y(t, \omega) w'(t - t') e^{i\omega t} dt' d\omega, \tag{5.14}$$

where $w'(t - t')$ must satisfy $\int_{-\infty}^{\infty} w'(t)w^*(t)dt = 1$. Thus, to recover the original signal from $Y(t, \omega)$, one can choose $w'(t) = w(t)$.

For the discrete time series $\{y_n\}$, STFT is given by

$$Y(m, \omega) = \sum_{n=-\infty}^{\infty} y_n w^*_{n-m} e^{-i\omega n}, \tag{5.15}$$

where $w_n = w(t_n)$ [109, Ch. 11.7].

For an application of STFT on the SN data, as well as the data of other solar analogues, please see Oláh et al. [99].

5.5 Uncertainty principle

A signal cannot be simultaneously both band-limited and time-limited, because multiplication in time domain transfers to convolution in frequency domain and vice versa: $\mathcal{F}(x \cdot y) = \mathcal{F}(x) * \mathcal{F}(y)$. If we normalize the squared modulus of the signal $|y(t)|^2 = p(t)$ and its Fourier transform $|Y(f)|^2 = P(f)$ such that $\int_{-\infty}^{\infty} p(t)dt = 1$ and $\int_{-\infty}^{\infty} P(f)df = 1$, and assume them to be usual probability densities, then we can define the effective duration and bandwidth of the signal as

$$\sigma_t^2 = \mathbb{E}_{p(t)}\big((t-\mu_t)^2\big) \quad \sigma_f^2 \ = \mathbb{E}_{P(f)}\big((f-\mu_f)^2\big), \tag{5.16}$$

where $\mu_t = \mathbb{E}_{p(t)}t$ and $\mu_f = \mathbb{E}_{P(f)}f$. Then it can be shown that $\sigma_t\sigma_f \geq 1/4\pi$. For the proof see, for instance, Bracewell [22] or Cohen [32, Ch. 3.3]. More specifically, in the case of STFT, the lower bound in this product is achieved when a Gaussian function $e^{-t^2/2}$ is used as $w(t)$. Thus, to achieve the best compromise between frequency and time resolution, one should use the Gaussian window function. The latter is also known as the Gabor transform [109, Ch. 11.2], [123]. The minimum frequency that can be resolved by a finite duration time window is called a Rayleigh frequency.

5.6 Wavelet transform

5.6.1 Continuous signals

The broad topic of wavelet transform (WT) is out of scope of the current thesis. However, as it is one of the most popular methods in time-frequency analysis, for the sake of completeness, we will summarise some of the main ideas of it here. The continuous WT is defined as

$$W(a,b) = \langle y, \psi_{a,b}\rangle = \int_{-\infty}^{\infty} y(t)\psi_{a,b}^*(t)dt, \tag{5.17}$$

where the functions $\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right)$ are called the wavelets, and a and b are the scale and translation (or dilation and position) parameters.

The reconstruction of the input signal is given by

$$y(t) = C_h^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a^{-2} W(a,b) \psi_{a,b}(t) db da. \tag{5.18}$$

For the reconstruction to exist, the admissibility condition requires that $0 < C_h < \infty$, where $C_h = \int_{-\infty}^{\infty} |\omega|^{-1} |\Psi(\omega)|^2 d\omega$, $\Psi = \mathcal{F}(\psi)$. The necessary condition for the latter is $\Psi(0) = 0$ [109, Ch. 11.5].

When one defines a wavelet as

$$\psi_\sigma(t) = c_\sigma \pi^{-1/4} e^{-t^2/2} \left(e^{i\sigma t} - \kappa_\sigma \right), \tag{5.19}$$

where $\kappa_\sigma = e^{-\sigma^2/2}$ and $c_\sigma = \left(1 + e^{-\sigma^2} - 2e^{-3\sigma^2/4}\right)^{-1/2}$, then the WT has a close similarity to the Gabor transform. This wavelet is known as the Morlet wavelet. For period determination of the phenomena in the nature, the Morlet wavelet is one of the most natural choices for the analysis.

The discrete WT (DWT) assumes that the parameters a and b are discretised, such that $a = a_0^{-m}, b = n a_0^{-m} b_0, m, n \in \mathbb{Z}, a_0 > 1, b_0 \neq 0$. In the following we will discuss only the orthonormal wavelet basis, which is possible to construct if one takes $a_0 = 2$ and $b_0 = 1$. In this special case the input signal is projected onto the set of basis functions $\psi_{m,n}(t) = 2^{m/2} \psi(2^m t - n)$, where $\langle \psi_{m,n}, \psi_{m',n'} \rangle = \delta_{mm'} \delta_{nn'}$. The coefficients of DWT and the reconstruction are given by

$$\begin{aligned} W_{m,n} &= \langle y, \psi_{m,n} \rangle = \int_{-\infty}^{\infty} y(t) \psi_{m,n}^*(t) dt, \\ y(t) &= \sum_{m,n=-\infty}^{\infty} W_{m,n} \psi_{m,n}(t). \end{aligned} \tag{5.20}$$

To achieve the orthonormality, the function $\psi(t)$ has to be selected such that

$$\psi(t) = 2 \sum_{n=-\infty}^{\infty} d_n g(2t - n), \tag{5.21}$$

were the values d_n form a square summable sequence and function $g(t)$ itself satisfies an orthonormality in the following sense:

$$\langle g(t-m), g(t-n) \rangle = \delta_{mn}, \quad m, n \in \mathbb{Z}. \tag{5.22}$$

The simplest example of a function satisfying this criteria is

$$g_0(t) = \begin{cases} 1 & \text{if } 0 \le t < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.23}$$

One can construct a so-called Haar wavelet basis taking $\psi(t) = g_0(t) - 2g_0(2t-1)$. To generate other functions $g(t)$ satisfying Eq. (5.22), one can use the following iterative formula: $g_{m+1}(t) = \sum_{n=-\infty}^{\infty} c_n g_m(2t-n)$, and take $g(t) = \lim_{m\to\infty} g_m(t)$. The sequence c_n must also be square summable. Then d_n can be calculated using $d_n = (-1)^n c^*_{1-n}$ and the wavelet is defined via Eq. (5.21) [109, Ch. 11.6], [118]. An algorithm for generating orthonormal wavelets $\psi(t)$ with compact support (involving also the recipe for selecting the sequence c_n) can be found in [67].

The main difference between STFT and WT is that the former is a constant resolution (constant Q) transform, while for the latter frequency resolution increases with a, whereas time resolution decreases. Similarly to spectrogram one can calculate the so-called scalogram $P(a,b) = |W(a,b)|^2$. For the orthonormal wavelet basis, the total energy of the signal is $E = \int\int P(a,b) dadb/a^2$ [118] (i.e. the total energy requirement is satisfied).

5.6.2 Discrete signals

For a discrete evenly sampled signal $y_k = y(k\Delta t)$, DWT is defined as

$$w_{m,n} = 2^{m/2} \sum_{k=-\infty}^{\infty} y(k\Delta t)\psi(2^m k\Delta t - n\Delta t), \tag{5.24}$$

where Δt is the sampling period. This is called a discrete-time wavelet transform. Under certain conditions $w_{m,n} = W_{m,n}$ [132]. For $\Delta t = 1$, $w_{m,n}$ can be calculated iteratively as follows:

$$s_{m,n} = \sqrt{2} \sum_{k=-\infty}^{\infty} c_k s_{m+1,2n+k}, \quad w_{m,n} = \sqrt{2} \sum_{k=-\infty}^{\infty} d_k s_{m+1,2n+k}, \tag{5.25}$$

where c_k and d_k were introduced in the previous section. In practice, of course, the signals have a finite length, so that the sums in the above formula are also truncated. Suppose there are $N = 2^M$ points, then the calculation goes backwards starting from the largest scale, M, while using the initial condition $s_{M,n} = y_n$ and continuing down to the coarsest

scale needed in the analysis. Eq. (5.25) represents a repeated application of convolution and decimation (downsampling by 2) of the signal [109, Ch. 11.7], [151, 118, 132].

WT for unevenly sampled time series has been less studied topic. Foster [48] discusses a DWT using Morlet mother wavelet, giving examples of application to white noise and purely harmonic signals. DWT for unevenly sampled time series can be defined as

$$W(\omega,\tau) = \sqrt{\omega}\sum_{i=1}^{N} y(t_i)\psi^*_{\omega,\tau}(t_i), \tag{5.26}$$

where

$$\psi_{\omega,\tau}(t) = e^{i\omega(t-\tau)-c\omega^2(t-\tau)^2} \tag{5.27}$$

is the abbreviated Morlet wavelet. DWT with the abbreviated Morlet transform is very similar to Gabor transform, however, here the window is frequency dependent. Discrete wavelet power is defined as

$$\mathrm{DWP} = \frac{|W|^2}{s^2\omega n(\sqrt{2}\omega,\tau)}, \tag{5.28}$$

where s^2 is the sample variance, and $n(\omega,\tau) = \sum_{i=1}^{N} e^{-c\omega^2(t_i-\tau)^2}$. The need for the denominator comes from the fact that for white noise $\mathbb{E}(|W|^2) = \sigma^2\omega n(\sqrt{2}\omega,\tau)$. DWP has relatively poor performance and the causes are similar to that of a naive periodogram. The estimates become especially bad in the gaps of the data. Foster [48] argues, that using ideas analogous to GLS, it is better to project the data onto three trial functions $\phi_1(t) = 1, \phi_2(t) = \cos(\omega(t-\tau))$ and $\phi_3(t) = \sin(\omega(t-\tau))$. To not to loose the idea of locality, one uses the weights $w(t) = e^{-c\omega^2(t-\tau)^2}$ in the projection. Note that this is the second factor in Eq. (5.27). Let us define vectors of trial functions, weights and data as $\boldsymbol{\phi}_i = [\phi_i(t_1)\dots\phi_i(t_N)]^\mathsf{T}, i = 1,\dots,3$, $\mathbf{w} = [w(t_1)\dots w(t_N)]^\mathsf{T}/(\sum_{i=1}^{N} w(t_i))$, $\mathbf{y} = [y(t_1)\dots y(t_N)]^\mathsf{T}$. Furthermore, let us define matrix of weights $\mathbf{W} = [\mathbf{w},\mathbf{w},\mathbf{w}]$, matrix of trial functions $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\boldsymbol{\phi}_3]$, and denote the vector and matrix operations with subscript $\mathbf{w}$, such that $\mathbf{a}_\mathbf{w} = \mathbf{a}\circ\mathbf{w}$ and $\mathbf{A}_\mathbf{w} = \mathbf{A}\circ\mathbf{W}$. Then the projection coefficients are described by a vector

$$\mathbf{p} = \mathbf{S}^{-1}(\boldsymbol{\Phi}^\mathsf{T}\mathbf{y}_\mathbf{w}), \tag{5.29}$$

where $\mathbf{S} = \mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}_{\mathbf{w}}$. This equation is a generalisation of Eq. (3.36) for the non-orthogonal basis functions using weighted projections. The signal part (or the model function) is given by $s(t) = \sum_{i=1}^{3} p_i \phi_i(t)$. If we now define the weighted variation of the data $V_y = \mathbf{y}^{\mathsf{T}}\mathbf{y}_{\mathbf{w}} - \left\|\mathbf{1}^{\mathsf{T}}\mathbf{y}_{\mathbf{w}}\right\|^2$ and the signal $V_s = \mathbf{s}^{\mathsf{T}}\mathbf{s}_{\mathbf{w}} - \left\|\mathbf{1}^{\mathsf{T}}\mathbf{s}_{\mathbf{w}}\right\|^2$, where $\mathbf{s} = [s(t_1)\dots s(t_N)]^{\mathsf{T}}$, then finally one can define a weighted wavelet Z-transform (WWZ)

$$Z = \frac{(N_{\text{eff}} - 3)V_s}{2(V_y - V_s)}, \tag{5.30}$$

where $N_{\text{eff}} = n^2(\omega,\tau)/n(\sqrt{2}\omega,\tau)$. The performance of this statistic in determining the frequency of a purely harmonic signals has been shown to be much better than the one of Eq. (5.28) [48].

A method with adaptive window size, depending on the local density of data points, called the least squares wavelet analysis, was developed by Ghaderpour and Pagiatakis [51]. It was shown to outperform WWZ in detecting the components of a certain multicomponent, non-stationary signal. For a single harmonic with constant frequency and amplitude, it showed more leakage in the spectrogram compared to WWZ, however, the latter one showed time dependent amplitude of the peak at true frequency. More discussion on wavelets for irregularly spaced data and how to determine the optimal scale and shift grids can be found in Andronov [2].

5.7 Hilbert-Huang transform and empirical mode decomposition

A slightly different approach to multiresolution analysis is the Hilbert-Huang transform (HHT) [65]. In this approach the input signal is decomposed into a sum of basis functions called intrinsic mode functions (IMFs). In contrast to the other methods, the basis functions for HHT are not fixed a priory, but derived directly from the data and they are not guaranteed to be orthogonal. The decomposition implicitly makes the assumption that, at any given time, data may have many coexisting simple oscillatory modes of significantly different frequencies [64].

HHT is fully performed in the time domain consisting of two steps. In the first step, the input signal is separated into a set of IMFs using an algorithm called empirical mode decomposition (EMD). In the second step,

for each extracted mode, Hilbert spectral analysis is applied, which allows each IMF to be described as an analytical signal.

Let us denote the input signal by $\mathbf{y} = [y(t_1) \dots y(t_N)]^\mathsf{T}$, then

$$\mathbf{y} = \sum_{m=1}^{M} \mathbf{c}^{(m)} + \mathbf{r}, \tag{5.31}$$

where N is the number of data points, M is the number of IMFs, $\mathbf{c}^{(m)}$ is the m-th extracted IMF and $\mathbf{r}$ is the final residue, which is a sample of a monotonic function. The IMFs need to satisfy the following two criteria:

1. The difference of the number of zero crossings and the number of extrema must be less than or equal to one.

2. The mean of the envelope along the maxima and the envelope along the minima must be zero at each time moment.

These criteria are crucial as only then the IMFs can represent zero mean amplitude/frequency modulated signals for which meaningful instantaneous frequency exists.

The IMF extraction process is described by the following algorithm:

Algorithm 3: EMD algorithm

Input: $\mathbf{y} = [y(t_1) \dots y(t_N)]^\mathsf{T}$

Output: IMFs and the residue

$\mathbf{r} \leftarrow \mathbf{y}$

$m \leftarrow 1$

while $\mathbf{r}$ *is not monotonic* **do**

  $\mathbf{h} \leftarrow \mathbf{r}$

  while $\mathbf{h}$ *does not satisfy IMF criteria* **do**

    Detect the local minima $\mathbf{h}_{\min}$ and maxima $\mathbf{h}_{\max}$ of $\mathbf{h}$

    Fit the lower and upper envelope curves $f_{\mathrm{L}}(t)$ and $f_{\mathrm{U}}(t)$ through $\mathbf{h}_{\min}$ and $\mathbf{h}_{\max}$

    Calculate the mean of the envelopes $f_{\mathrm{M}}(t) = \big(f_{\mathrm{L}}(t) + f_{\mathrm{U}}(t)\big)/2$

    Calculate the residues between $\mathbf{h}$ and $\mathbf{f}_{\mathrm{M}} = [f_{\mathrm{M}}(t_1) \dots f_{\mathrm{M}}(t_N)]^\mathsf{T}$

    $\mathbf{h} \leftarrow \mathbf{h} - \mathbf{f}_{\mathrm{M}}$

  end

  $\mathbf{c}^{(m)} \leftarrow \mathbf{h}$

  $m \leftarrow m + 1$

  $\mathbf{r} \leftarrow \mathbf{r} - \mathbf{h}$

end

return $m, \mathbf{c}^{(1)} \dots \mathbf{c}^{(m)}, \mathbf{r}$

The inner loop in the algorithm is usually referred to as the sifting process. The EMD algorithm is sensitive to many things, especially how the extrema are detected and what kind of interpolation and boundary conditions are used. Also the stopping criterion for a found IMF affects the results [112]. The decomposition highly depends on how all of these aspects are handled, as the potential errors accumulate from one step to another. Effects of different boundary conditions are illustrated in Holzinger and Benedikt [61].

Having obtained a decomposition into IMFs, one can construct the corresponding analytical signals and based on the instantaneous amplitude $A(t)$ and frequency $\omega_{\mathrm{inst}}(t)$ form the so-called Hilbert power spectrum $H(\omega, t)$. The latter is defined as the square of the instantaneous amplitude $A^2(t)$ while using $\omega_{\mathrm{inst}}(t)$ for ω.

The orthogonality of the IMFs is not guaranteed theoretically, however, it is usually satisfied in practice. If we take the square of Eq. (5.31) and

denote $\mathbf{c}^{(M+1)} = \mathbf{r}$

$$||\mathbf{y}||^2 = \sum_{i=1}^{M+1} ||\mathbf{c}^{(i)}||^2 + \sum_{i=1}^{M+1} \sum_{j=1}^{M+1} \mathbf{c}^{(i)\mathsf{T}} \mathbf{c}^{(j)}, \tag{5.32}$$

then we can define an index of orthogonality as

$$\mathrm{IO} = \sum_{i=1}^{M+1} \sum_{j=1}^{M+1} \mathbf{c}^{(i)\mathsf{T}} \mathbf{c}^{(j)} / ||\mathbf{y}||^2. \tag{5.33}$$

Similarly, one can define the measure of orthogonality between IMF pairs [65]. Note that the orthogonality defined above is global, but the real meaning applies only locally. It can be that the neighbouring IMFs contain selections of data, sharing the same frequency at different time segments. Locally, however, any two IMFs should be orthogonal. The leakage usually depends on the length (finiteness) of the dataset as well as how good decomposition was achieved [65]. Therefore, as an independent check to the 'quality' of the decomposition one should ensure that IO is close to zero in real applications.

The EMD algorithm has been also generalized to multiple dimensions. This algorithm is known as the multivariate EMD or MEMD. For further information see, for instance, [113, 63].

One of the major drawbacks of the original EMD is the problem known as mode mixing, which is a consequence of signal intermittency. To overcome this problem, a noise-assisted data analysis method called the Ensemble EMD (EEMD) was proposed [160], which is based on the fact that EMD acts as a dyadic filter bank akin to WT when applied to Gaussian white noise [44]. The easiest way to understand why EMD can fail in certain situations is to imagine an example, where a low-frequency harmonic is superimposed with intermittent high-frequency oscillations. In that case, the envelopes generated by the first sifting iteration should go through the extrema of the high-frequency component where it is present, and follow the low-frequency component elsewhere. However, this is not what actually results. Instead, the envelopes are distorted, leading to the first IMF being a combination of both signal components [160]. Adding noise introduces additional extrema to the signal, resolving the problem. Now, of course, the IMFs are contaminated by the very same noise, which one needs to average out by doing repeated trials.

In the EEMD algorithm Gaussian white noise is added to the input data forming $\mathbf{y}^{(k)} = \mathbf{y} + \boldsymbol{\epsilon}^{(k)}$, where $\boldsymbol{\epsilon}^{(k)} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ and $k = 1, \ldots, K$ is the number of trials. Here σ^2 controls the S/N level of the trials. For high-frequency signals it is recommended to use small σ^2, for instance, 10% of the data variance, and vice versa for low-frequency signals [141]. From this point on, the EMD algorithm is applied to each $\mathbf{y}^{(k)}$ and IMFs $\mathbf{c}^{(m,k)}$ are obtained, where $\mathbf{c}^{(m,k)}$ is the m-th IMF of the k-th trial. In the end, all the IMFs of the same order are averaged, to obtain final IMFs $\mathbf{c}^{(m)} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{c}^{(m,k)}$. In [141] an alternative algorithm was proposed, where the noise is not added in the beginning, but to each residue $\mathbf{r}$ in the EMD algorithm. In this case the averaging is done at each IMF level, not in the end. This method allows to choose σ^2 adaptively at each level. Arguably this method requires less sifting iterations and the original signal can be recovered exactly. Moreover, smaller ensemble size K can be used.

Likewise with many other methods, to assess the significance of the IMFs, one can use a Gaussian white noise as a null hypothesis. In [159] it was shown, that in the case of normalised Gaussian white noise $\bar{P}_n\bar{E}_n = 1$, for the n-th IMF with average period $\bar{P}_n$ and energy $\bar{E}_n$. In other words, $\log \bar{E}_n$ is linearly decreasing with increasing n, as $\bar{P}_n$ increases as 2^n. In the same paper E_n was shown to follow a χ^2 distribution. This information can be used to calculate the significance of the IMFs.

Some of the application of EMD and EEMD on solar datasets can be found in [80, 13, 161, 79].

5.8 Summary

The field of analysis of non-stationary processes is still developing. For instance the EMD algorithm and its derivatives contain several heuristic and ad-hoc elements. The limitations of its applicability and guarantees of accuracy are still not fully known. To open the mathematical understanding behind the ideology of EMD, a so-called synchrosqueezed wavelet transform was proposed [35]. In this study a class of well separated (intrinsic-mode-type) functions is defined which can be obtained via mapping the pair of shift and scale parameters (b, a) of the WT to shift and instantaneous frequency $(b, \omega(a, b))$. Discussion of this rather involved pa-

per, however, is out of the scope of the current thesis.

In this chapter we did not discuss at all probabilistic methods. One of the possible ways to tackle the problem in that case is to use GPs with non-stationary covariance functions. Some of the more recent works in that direction can be found in [142, 49, 117].

6. Applications

This chapter discusses the applications of CF, D^2, EEMD and two probabilistic methods, introduced in the previous chapters, on real datasets of stellar activity as well as on global magnetohydrodynamic simulations. We explain the reasons for deciding on those particular methods, the alternatives that could have been considered, and summarise the main findings. As mentioned in Chapter 2, from the point of view of time series analysis, there are primarily two relevant tasks in the domain of stellar activity: estimating the rotation period and the magnetic cycle length of a star. In both cases, there are difficulties of the same as well as of different origin.

As with any observational time series, the complicating factor is the presence of noise, the level of which is usually unknown. For the sake of simplicity, the noise is in most cases assumed to be uncorrelated, but in practice this is not so (for instance, imagine two consecutive observing nights with similar weather conditions when the atmospherical effects on the observations are similar). In addition to that, there might exist distortions of data due to instrumental effects.

Usually the datasets are sufficiently long to cover hundreds or even thousands of rotation periods of the star, however, in most cases, the sampling in the datasets is uneven with regularities and gaps (missing data). The uneven sampling itself is not a big obstacle, if it was according to, for instance, uniform distribution, however, the periodic sampling patterns induce aliases in the spectrum rendering it more challenging to detect the true period. The common patterns usually involve diurnal, monthly, and yearly periodicities. Moreover, there might not be one constant period persistent throughout the whole dataset. As discussed in Chapter 2, due

to differential rotation, even the notion of rotational period is ambiguous, unless explicitly stated what is meant. An exception to this are the close binary stars, which are tidally locked. In this case, the rotation period can be defined through orbital period.

The primary problem when estimating the lengths of activity cycles from observations is the limited size of the datasets. The longest ground-based monitoring programmes currently span a couple of decades, which means that in the best case only a small number of full cycles are observed for these active stars. However, due to the nonlinearities in the stellar dynamo equations, the cycles cannot be expected to be simply harmonic or periodic, but quasi-periodic. This factor greatly impedes or even renders estimation of the true cycle period from short datasets impossible. The sampling patterns in the data are usually minor problem in this context as the cycle lengths are longer than those time scales.

The 3D simulations of global magnetoconvection pose other types of challenges. The solution over the full numerical grid is harvested, resulting in massive amounts of evenly sampled data, which is difficult to analyse due to its complexity and size. The key question is to find out mechanisms that excite cyclic dynamo solutions. Many dynamo modes can be simultaneously excited and their properties need to be quantified (such as the coherence time, exact period and spatial localisation). The non-linear dynamics naturally generate quasi-periodic signals and non-stationary behaviour. Due to these complications, standard methods fail and computationally efficient ones are needed.

6.1 Multiperiodicity, modulations and flip-flops in variable star light curves

In our first two papers, we focused on modelling the light curves of some variable stars. In Paper I, we introduced the CF method with local phase and frequency visualisation technique as described in Sect. 4.1.1. The advantages of the method were illustrated by providing examples including mismatched carrier frequency, abrupt change in period, two beating frequencies, chirp signal, and non-harmonicity. Finally, we applied the method on two targets: MW Lyr and FK Com. In the former case, we compared the phase plots from the CF method to those of bi-periodic model

from an earlier study and, in the latter case we showed the qualitative difference between the spline and harmonic modulators in the CF model.

In Paper II, we selected a young solar analogue LQ Hya as a particular research target. This star is obviously quite a complicated object as inconsistent period estimates had been obtained from previous studies. The challenge was to find ways to establish the possible causes responsible for producing these different periods. Such causes include differential rotation, azimuthal dynamo wave and stellar rotation itself. Transient phenomena can occur as spots rotate with different periods while having a finite lifetime. Azimuthal dynamo wave, on the other hand, manifests itself as a persistent active longitude that migrates in the orbital reference frame. This has been predicted by early dynamo models [82, Ch. 16.3] and later found from 3D magnetoconvection models, such as [33].

The estimates of the periods previously found cluster tightly within a certain range of values, while their 'duty times' (when they are present in the time series) are mostly unknown. To establish the duty times, one can consider multiresolution analysis methods discussed in the previous chapter or the CPS method described in Sect. 4.2. In fact, the latter approach was used in Lehtinen et al. [87]. However, one of the limitations of this approach is that one can obtain the light curve only in the intra-seasonal regions of the data. Nevertheless, our study aimed at finding a global continuous model such that there would also be a light curve estimate in the gaps between the observational seasons. The trade-off between the two approaches is accuracy in the observational seasons versus the predictive power in the gaps. For the model fitting, we decided to use the CF method introduced in Paper I as being one of the fastest and easiest linear regression methods, given that the carrier period is known. As we knew from the previous studies and by looking at the power spectrum, there does not exists a single peak whose frequency would be best suited for that purpose, rather we needed to smooth the spectrum and find the average frequency. This could be achieved in several ways, but we chose to use the D^2 statistic from Sect. 4.3 by varying the time-scale parameter. We took the carrier frequency as optimal frequency for such a time-scale where the D^2 spectrum is still unimodal. One possible interpretation of a carrier frequency found in such a way is the so-called Carrington rotation period defined in the solar context. This is the period corresponding to the mean

latitude around which the starspots emerge on a visible to us hemisphere. From the CF model we extracted the primary and secondary minima and compared them to the surface temperature maps from DI study by Cole et al. [34] as well as with the minima from the study by Lehtinen et al. [87]. The agreement between the studies was reasonably good. The importance of the global model is better understood in this respect, for if the results from different studies are compared, there is no guarantee that the observing seasons overlap.

From the point of view of visualising the intermittent and transient phenomena observed in the photometry, we used the technique described in Sect. 4.1.1. There we detected longer time frames during which the local period differed significantly from the mean period. However, it was not clear, whether these phenomena could correspond to active longitudes or azimuthal dynamo waves. We detected one epoch when the data is the most consistent with a long-lived non-axisymmetric drifting spot structure. In any case, it is very clear that azimuthal dynamo waves are less pronounced in solar-type single stars than in giant binary systems, one of which was also recently analysed by the CF method [89]. We also detected several flip-flop events with apparently irregular timing patterns, hence suggesting that these events are not directly connected with the stellar cycle, as previously suggested by Berdyugina et al. [18].

Using the half-width of the D^2 spectrum around the minimum, we calculated a rough estimate for the relative differential rotation coefficient, which is expressed as $k = (\Omega_{\rm eq} - \Omega_{\rm pol})/\Omega_{\rm eq}$. Consider solar-like differential rotation with fast equator and slow poles with known equatorial period. Dropping the third term in Eq. (2.1), one can then calculate a period corresponding to any given latitude or, vice versa, a latitude corresponding to any given period. We took the estimate for the equatorial period as the highest cut-off frequency in the power spectrum near the base frequency band and calculated the latitude for the mean rotational period found from D^2 analysis (interpreted it as the Carrington latitude). However, the estimate turned out to be unrealistically low (close to equator), even smaller than the value for the Sun.

We want to further remark on the importance of finding the average rotation period as we did. In principle the goodness of the fit for the CF method is not as sensitive to the carrier frequency, given that it is reason-

ably close to the centre of the base frequency band. However, by finding a time-scale of the process, we can extract additional information, potentially related to an average lifetime of a spot or the turnover time for convection. These are, of course, quite speculative suggestions and were left out of the results.

By postulating the form of the covariance of the data, one can estimate the mean rotation period and time-scale of the process also by using, for instance, GP or Kalman filter. As a side result, one also obtains a continuous fit for the data with these methods. Such applications would constitute potentially interesting future studies for the given or other similar stars.

6.2 Finding cycles from PENCIL-Millennium simulation

Our next analysis target was the data from a solar-like semi-global 3D magnetoconvection simulation, representing a very viscous Sun-like star, which we refer to as a PENCIL-Millennium simulation. The higher diffusivities are needed as it is impossible to approach anywhere near the high Rayleigh and Reynolds numbers of real stars. In the given model a roughly 10^6 times higher luminosity is implied and the convective velocity is roughly 100 times greater in comparison to the Sun. To obtain the same rotational influence on the flow, the rotation period is thus increased by the same factor. The progenitor run of this model was the first ever solar-like dynamo solution from global direct numerical simulations obtained [71]. It is also the most extensive in time simulation as it covers over 200 magnetic cycles at the present moment. Full details of the simulation set up can be found in Paper III.

The solution of the simulation is oscillatory, but very complex, the cycle period changing significantly depending on the depth wherefrom it is plotted (even by eye), and non-stationary Maunder minima-like epochs were related to it. For the reasons listed in the previous paragraph, the main goal of this study was not to make direct comparisons with the observed characteristics of the solar magnetic cycle, but to discover the causes for the irregular behaviour. In contrast to previous studies, we analysed the solution in the whole CZ. The following is a list of some of the conclusions

reached based on the solution:

- The magnetic field is hemispherically asymmetric, while the velocity field remains symmetric.

- There is a clear evidence of equatorward migrating dynamo wave in the low latitudes and poleward in the higher latitudes.

- The dominating magnetic field component near the surface is the radial one with roughly twice the strength of the azimuthal component, while the azimuthal field becomes dominant at greater depths, being roughly four times stronger than the radial field at the bottom of the CZ.

- The toroidal (azimuthal) field is roughly three times weaker near the surface than at the bottom of the CZ, however, the latter is gradually decaying in time.

- There are epochs of ceased surface activity, most prominent around $20-40$ years in the south and $35-45$ years in the north.

- The azimuthal velocity grows when the magnetic field is weak, and decays when it is strong.

- The angular velocity variations (torsional oscillations), occur with twice the frequency of the magnetic cycle.

- The epochs of decreased or practically ceased surface activity are not global magnetic energy minima, but maxima as strong magnetic fields are stored in the deeper parts of the CZ.

Now we turn to the magnetic cycle analysis of the solution, which was the most relevant part of the study in the context of the current thesis. We analysed the time series of azimuthally averaged quantities using the EEMD method described in Sect. 5.7. At each time moment the domain of the data forms a 2D meridional disc with unit radius, from where the polar regions having latitudes that satisfy $|\theta| > \pm 75^\circ$ and radii $r < 0.7$ are

cut out. At each location in the domain we have the values of the components of azimuthally averaged, but otherwise fully turbulent, magnetic field vector, velocity vector and the kinetic helicity, which we sampled with 12 points in the radial direction and with 18 points in the latitudinal direction, to render it suitable for the 1D EEMD algorithm. We used ensemble size 100 and noise variance equal to the sample variance of the time series. There were primarily two reasons for favouring the use of EEMD algorithm over others. First, the coexistence of multiple quasi-periodic cycles and second, as mentioned above, the epochs of reduced activity (missed or vanishing cycles). We aimed at adaptively extracting the cycles as a collection of IMFs from the data. Adaptivity is important, as using, for instance, regression models for the same purpose can become infeasible, as the number of cycles is not known beforehand, leading possibly to excessively high number of models to test. Likewise, per each trial model, optimising the parameters can be very costly. The other alternatives to EEMD, such as WT and STFT, could have been considered for this task, however, the benefit of EEMD is that it provides a straightforward way to obtain the decomposition of the input signal into the components directly in the time domain.

To estimate the significance of the IMFs, we used an idea similar to the one discussed in Sect. 5.7; namely, we compared the energy of the IMF to the theoretical energy of the Gaussian white noise. We called this quantity M_i^{reg} as the mode regularity. Significantly higher value of this quantity than $1/M$, where M is the number of modes, is an indication that the mode is also significant.

Our main observation from the analysis was that for the magnetic field, IMF 7 with an average period of five years was the most prominent for all components, however, the locations of the highest amplitude of the cycles differed. For the azimuthal field, the cycle was mainly confined to low latitudes at the surface region where the equatorward migration of the field was observed. The cycle was found mainly near the surface of the CZ close to the latitudinal boundaries for the radial field, and around higher latitudes and middle to lower layers for the latitudinal field. The only other mode that we found significant was IMF 11, forming a cycle of roughly 50 years, for the magnetic field components at the bottom of the CZ. All the other modes seemed to be containing purely noise.

Next, we analysed the velocity components and kinetic helicity to see the potential modulations in the rotation as well as in the α-effect. The total energy of the modes for the azimuthal velocity was spread over IMFs from 6 to 11. For the radial and latitudinal as well as for the kinetic helicity, the strongest mode was IMF 1. However, the only significant mode in all four cases turned out to be the IMF 11. Therefore, it is evident that neither of these quantities is modulated by the dominant magnetic cycle.

For each IMF, we also calculated the equatorial symmetry or the so-called parity which shows whether the quadrupolar or dipolar configuration with respect to the equator is dominating on average (both in time and space). For all the significant magnetic field IMFs, the average parity turned out to be weakly negative, which corresponds to the dipolar (solar-like) mode. IMF 11 was more dominantly antisymmetric than IMF 7, however, the parity of IMF 7 was strongly variable, obtaining values of nearly -1 and $+1$. The irregular epochs at the surface were interpreted to be a result of the interplay of the different parity states. Hence, based on the dynamo mode properties revealed by EEMD, we were able to isolate a totally new mechanism of disturbed epochs in stellar dynamos.

We note that the division into the IMFs was sensitive to the variance of the additive noise in the EEMD algorithm. However, the effects of the noise on the results have not yet been thoroughly investigated in the literature. Another problem is that as the errors in the decomposition cumulatively increase when being passed through the residues to each subsequent mode, there is a high chance that the IMFs with higher indices become less reliable. Additional errors appear due to a limited ensemble size, implying that the noise is not fully cancelled out. Due to these shortcomings and also because of a somewhat simplistic way of determining the significance of the extracted IMFs, we decided to conduct an independent study and repeat the analysis using a multidimensional D^2 statistic introduced in Sect. 3.54. This forms the topic of the Paper IV.

There were two more reasons underlining the need for the second study. Namely, the fact that a secular trend was observed for the long-term cycle in the bottom of the CZ and that we now had almost twice as long simulation data available. This enabled a more reliable analysis for the longer cycle(s). Furthermore, during the first study, we qualitatively observed the presence of a very short cycle, whose period we did not manage to

estimate with the given tools.

Concerning the methodology, it is important to note that for the multicyclic time series, the cycle length estimation with D^2 statistic involves additional complications. The basic assumption when using Eq. (4.8) is that ℓ is greater than the hypothetical cycle period. What happens when this is not satisfied is that the statistic does not 'see' the full variance of the data, in other words, that the base level starts decreasing with ℓ. This effect has unwanted consequences when the data consists, for example, of two cycles, such as one of them being much shorter than the other and the coherence time of the former being simultaneously shorter than the period of the latter. To overcome this problem, it is first better to detect the period of the longer cycle and subtract the corresponding cycle from the data. As the D^2 method does not provide the regression model, then the easiest way to achieve this is to high-pass filter the data, which, in certain cases, can be achieved sufficiently well by subtracting the moving average using a suitable window size from the data[1].

Now we come to one of the main results of this study, namely confirming the existence and establishing the length of a short and very incoherent cycle around half a year. This cycle resembles the quasi-biennial oscillations of the Sun, observed, for instance, in its activity indices [15]. The extreme incoherency and relatively weak amplitude of this cycle rendered it difficult to spot it in the power spectrum, and were probably also the main reasons why this cycle was not detected in Paper III using the EEMD method. The cycle was most likely split between the two separate IMFs whose average amplitudes were below our detection limit. Furthermore, we confirmed the existence of two long cycles, one around 50 years and another around 100. However, none of these cycles seemed to be equally strong on both hemispheres. The former seemed to be weaker in the northern, while the latter in the southern hemisphere. Moreover, for all the cycles, except the shortest one, the periods were slightly different on both hemispheres. Such an 'impairment' in between the rotations of different hemispheres is also known from observations [5], now for the first time found from global simulations.

While the multidimensional aspect of the D^2 statistic does not add any-

[1] In our case, this is straightforward as the PENCIL-Millennium simulation data possessing an even sampling.

thing novel to the analysis, as it is equivalent of averaging the estimates for one-dimensional time series, in some cases it is more natural to think in terms of vector or matrix quantities changing over time, rather than as a set of scalars. To illustrate the idea, some examples were presented in the paper. The same paper also empirically investigated the effects of the cycle coherency and S/N on the D^2 spectrum. Although the D^2 statistic is primarily meant for analysing stationary data, we also investigated the effect of non-stationarity on the spectrum, in particular when the cycle is only temporary. However, it became clear that solely based on the D^2 spectrum, the temporary aspects cannot be separated from the aspects of low coherency (i.e. the case when the time-scale of the quasi-periodic process is relatively short).

An important aspect when working with big datasets, such as that of PENCIL-Millennium, is the optimisation. In the original form, the D^2 statistic scales as N^2 per each dimension, trial frequency and time-scale. To render the algorithm feasible, we used data binning in the time lag domain, such that for the highest trial frequency there would still be a sufficient amount of bins. For each bin, the squared differences of the observed values were pre-computed, ensuring that these values were reusable for each trial frequency and time-scale. For further speed-up, we used parallelisation over the dimensions.

We note that although in Paper III we used EEMD as a multiresolution tool, we did not plot the Hilbert spectra, but rather calculated the average cycle periods based on the zero-crossings of the IMFs. Possible future studies using PENCIL-Millennium simulation data could involve a more detailed analysis involving Hilbert spectra, and comparisons to other time-frequency methods, such as WT and STFT. Furthermore, application of the multidimensional EEMD, or multi-output GPs can be considered, which would also allow studying the spatial coherence between the signals. However, without any reduction, the dataset is most likely overly large for these tasks. Another possibility would be to attempt generalising the D^2 statistic for a similar purpose.

6.3 Finding cycles with probabilistic methods and handling trends in data

Papers V and VI turned our attention to cycle length estimation of the Mount Wilson (MW) stellar chromospheric activity dataset using probabilistic methods. The probabilistic methods have become increasingly feasible over the years as computational power has increased. The primary benefit of these methods is that they are usually easier to interpret compared to the non-probabilistic methods. This is due to most of those methods a priori define a concrete model (via the explicit likelihood function) of the data. Knowing the model facilitates an understanding of the applicability and limitations of the method. In contrast, some of the non-probabilistic methods discussed in this thesis, for instance, the D^2 statistic are sometimes harder to interpret, as we do not know what are the assumptions about the data (unless explicitly shown what is the probabilistic equivalent of the given method). The second benefit of the probabilistic methods is that the uncertainties of the parameter estimates are the direct coproducts of the inference process (unless we are only optimising the parameters). The trade-off is usually such that to gain the error estimates for non-probabilistic methods one needs to use bootstrap techniques, but for probabilistic methods one needs to sample from the posterior distributions, unless the posterior is analytically tractable. Another aspect is the possibility to use prior distributions, which correspond to regularisation in the case of non-probabilistic methods. The primary need for priors arises due to an insufficient amount of data for the given complexity of the data. From the point of view of period estimation, there are not always obvious reasons why one would want to use the priors, as most of the models involve few parameters; nevertheless, there is a clear exception, which we will shortly discuss.

We observed that many of the time series of the S-indices of the stars from MW dataset had apparent linear trends. Most likely, they represent very long cycles whose small fraction of the phase is only covered by the time series. On the other hand, we cannot neglect the slight possibility of instrumental degradation effects either, but because both increasing and decreasing trends are present in the data, this renders it unlikely. From the perspective of cycle period estimation, MW datasets can be considered

relatively short, covering only a couple of full cycles in the best cases. This fact renders the results sensitive to the way in which the trend is handled. In Paper V, through number of examples, we conclude that the optimal way to estimate the cycle lengths in this case is to include the linear trend component into the model. We considered a model which consists of a cosine, sine, slope and intercept components. It was revealed that, in some cases, the model components can start compensating for each other, while it is clear that these solutions correspond to unphysical situations with enormous regression coefficients. An example of one such case is given in Paper V. We concluded that if not all, then at least some of the parameters should be restricted by imposing priors. We called such a model the Bayesian generalised Lomb-Scargle periodogram with trend (BGLST).

In Paper V we also investigate empirically how the noise handling affects the results. In the case of MW dataset we observed a high variation of the intra-seasonal variance, which we could not understand well. However, we wanted to know with the help of simple generated datasets, whether assuming the noise variance to be constant, while it actually changes over time, leads to deterioration in the cycle period estimate. The idea was to empirically estimate the local noise based on intra-seasonal variances. The conclusion, whether this approach is more beneficial than just assuming the constancy of the noise was not very clear, as the performance depends on the degree to which the noise changes over time.

To return to the discussion of the MW dataset and the main motivation behind the study in paper VI, there has been a long debate over the existence of clusters in the period to cycle length ratio – stellar activity (RCRA) diagrams. If such 'clustering' existed, it would imply that the stellar dynamos would not be totally random processes. On the contrary, given a certain rotation rate, there would be a more likely dynamo state, whereto the star would be found. As the whole dynamo process is to convert the kinetic energy of largely turbulent and, hence, chaotic, fluid motions into organised magnetic fields with a cyclic pattern, a chaos in the realised cycle periods would be rather unexpected. To locate the star on this diagram, one needs to know the average chromospheric activity index $\langle R'_{\mathrm{HK}} \rangle$, the rotation period, and the magnetic cycle period of the star. As recently extended MW observations were made public, it became rea-

sonable to repeat the analysis conducted earlier by Baliunas et al. [6]. To estimate the cycle lengths we used the ideas and results from Paper V and first applied the BGLST model to the data. Using the rotation period and $\langle R'_{\rm HK} \rangle$ values from other studies, we plotted the RCRA diagram. Our diagram was significantly sparser than that of Saar and Brandenburg [121], the primary reasons for that were higher significance levels chosen in the analysis. We applied a Gaussian mixture clustering method to the data on the diagram and showed that two clusters explain the data the best. We could confirm the positive correlation within the inactive cluster, which is rather robustly established independent of a method chosen. However, no correlation was found in the active population. The time series of the stars in this population indicated high abundance of trends, which we interpreted as longer cycles, too long to be detected with our methods.

Our findings also show the Sun in a different light as previously thought. Some of the previous studies have positioned the Sun in between the two populations, but, we did not find such an evidence. Instead, the Sun seems to clearly belong to the inactive population. Hence, the solar dynamo should be explicable similarly to the ones generally on the inactive cluster. However, this cluster is currently inexplicable by any dynamo scenario so far, as we discussed in Paper VI.

In Chapter 2 we mentioned that reason behind grouping of stars into different populations could be the change from nonaxisymmetric to axisymmetric dynamo mode. However, the observation made from the current study was, that the estimated locations of this transition from numerical simulations and observational data mismatch. Furthermore, overplotting our results and the results from Lehtinen et al. [88] with recent results from turbulent magnetoconvection simulations [155, 154] do not show agreement with the inactive population, while at least qualitative behaviour in the active population could be reproduced.

Time series of the active stars cannot be assumed purely harmonic, therefore, in principle BGLST and other harmonic models are not well suited for the cycle length estimation task. This would be the case if the datasets were sufficiently long to cover at least tens or, even better, hundreds of cycles. However, as mentioned earlier, the time series in the MW dataset are extremely short in that respect. This finding justifies the usage of the harmonic models, nevertheless, we wanted to go further

and tried two more flexible models. These models were GPs with periodic and quasi-periodic (more precisely quasi-harmonic) covariance functions, respectively. The comments on feasibility and results of applying these methods to the data can be found in Paper VI. Here we only mention that by using the synthetic data, at least in principle, the quasi-periodic model was proven to be more accurate than the harmonic model in cycle length estimation. On the other hand, the real data contained high irregularity, thus making the significance of the models for many stars questionable. From this perspective, the analysis using GP models was partially inconclusive.

One other interesting finding from this study was the bimodal shape of the histogram of overall number of stars, the dip located at the mid-activity index corresponding to Vaughan-Preston gap, which is the decreased abundance of stars in certain range of $\langle R'_{\mathrm{HK}} \rangle$ [147]. However, the histogram of the number of stars with cycles was unimodal with the peak slightly towards higher values of $\langle R'_{\mathrm{HK}} \rangle$, thus, the cyclic stars are over-abundant in a certain region. The Sun was located at the inactive edge of this region and stars even slightly less active than it appear to shut off their cyclic dynamos. Transition from solar- to anti-solar differential rotation has been proposed as a potential mechanism enabling this [23].

7. Conclusions

Methods for period estimation of unevenly spaced time series have been developed and their properties studied over the last several decades. Most of these methods are well-suited to harmonic or periodic time series, but not applicable to quasi-periodic and non-stationary time series, which are prevalent in the domain of stellar magnetic activity. The primary focus in this thesis was to develop and apply some methods dedicated especially to analysing quasi-periodic time series. We gave an overview of the well known methods, such as the LS periodogram, and introduced simple, but more general methods, such as the D^2 statistic and the CF model. The latter methods were used several times in the applications on real-world as well as 3D magnetoconvection simulation data. With the help of the D^2 statistic, we estimated the mean rotation period of an active star as well as the magnetic cycle periods from the PENCIL-Millennium simulation. The CF method was used to build a continuous model for the light curve of the star.

From the toolbox of methods suitable for non-stationary time series we used EEMD as one of the most natural methods for decomposing a complicated signal into a set of significant modes carrying non-overlapping spectral information. EEMD was applied to the same PENCIL-Millennium simulation, thus allowing us to 'decifer' the surface magnetic behaviour by analysing the parities of the separate modes.

In the later stages of the work, we turned to probabilistic methods, which have become increasingly practical as computational speed has increased. We used a simple Bayesian linear regression model with a linear trend, and GPs as tools for cycle length estimation. The proper handling of linear trend becomes important when analysing relatively short stellar

activity time series, while the limitations of the linear regression models are overcome by GPs, which are much more flexible due to the freedom of choice of the covariance function. In the applications we used the probabilistic methods on the MW dataset, where we statistically confirmed the existence of two separate activity populations.

GP models are indeed very useful in the analysis of quasi-periodic time series, however, due to their poor scalability, finding faster alternatives is a relevant topic at present and in the future. Such alternatives include different GP approximations usually exploiting the properties of the covariance function as well as the state space methods. We also proposed one probabilistic method which can be seen as an analogue of the D^2 statistic. These methods carry great potential for future studies involving time series of active stars.

The motivations behind developing new period estimation methods can be found from different domains of astrophysics. For instance, currently, one of the hot topics is exoplanet search. However, our main motivation stems from dynamo theory as well as from the need to uncover the reasons and ways in which the stellar rotation period and cycle length are connected. As it is currently quite unclear the ways in which the dynamo mechanism works in detail, such studies set important constraints for dynamo models. One precondition for helping answer this question is to be able to estimate the rotation and cycle periods as reliably as possible. Our efforts in analysing several datasets of solar-type stars, as well as the dataset from 3D magnetoconvection simulation, provided useful information on the dynamo mechanism. Furthermore, it also provided directions on ways to improve the numerical models. Perhaps the most important result presented in this thesis is the confirmation of a positive slope of the inactive population in RCRA diagram. This result implies that none of the standard dynamo scenarios are applicable to solar-like stars, therefore holding major implications for dynamo theories. In the future, solar dynamo theory must bring in new thoughts and concepts to be able to explain the stellar observations.

It is clear that continuing the analysis of solar and stellar observations as well as the results from numerical simulations is very important to broaden our understanding of the physics behind the processes governing the active stars. However, the limited number of observational seasons of

the solar-type stars still presently renders the cycle analysis a very complicated, if not infeasible task. On the other hand, challenges in analysing the simulated datasets are related to massive and ever increasing sizes of the datasets. Thus, parallelisation of the existing methods while simultaneously developing new ones benefiting from the aspects of probabilistic machine learning, becomes an increasingly important part of the quickly evolving field of astroinformatics.

References

[1] T.W. Anderson. *The Statistical Analysis of Time Series*. Wiley Series in Probability and Statistics. Wiley, 2011. ISBN 9781118150399. URL https://books.google.fi/books?id=rCOzXIC8ZLkC.

[2] I. L. Andronov. Wavelet analysis of the irregularly spaced time series. *Social Studies of Science*, pages 57–70, February 1999.

[3] I. L. Andronov and L. L. Chinarova. Method of Running Sines: Modeling Variability in Long-Period Variables. *Częstochowski Kalendarz Astronomiczny 2014, ed. Bogdan Wszołek, Vol. X, p. 171-188*, 10:171–188, December 2013.

[4] Prabhu Babu and Petre Stoica. Spectral analysis of nonuniformly sampled data – a review. *Digital Signal Processing*, 20(2):359 – 378, 2010. ISSN 1051-2004. doi: https://doi.org/10.1016/j.dsp.2009.06.019. URL http://www.sciencedirect.com/science/article/pii/S1051200409001298.

[5] T. Bai. Hot Spots for Solar Flares Persisting for Decades: Longitude Distributions of Flares of Cycles 19-23. *The Astrophysical Journal*, 585:1114–1123, March 2003. doi: 10.1086/346152.

[6] S. L. Baliunas, R. A. Donahue, W. H. Soon, J. H. Horne, J. Frazer, L. Woodard-Eklund, M. Bradford, L. M. Rao, O. C. Wilson, Q. Zhang, W. Bennett, J. Briggs, S. M. Carroll, D. K. Duncan, D. Figueroa, H. H. Lanning, T. Misch, J. Mueller, R. W. Noyes, D. Poppe, A. C. Porter, C. R. Robinson, J. Russell, J. C. Shelton, T. Soyumer, A. H. Vaughan, and J. H. Whitney. Chromospheric variations in main-sequence stars. *ApJ*, 438: 269–287, January 1995. doi: 10.1086/175072.

[7] R. V. Baluev. Assessing the statistical significance of periodogram peaks. *MNRAS*, 385:1279–1285, April 2008. doi: 10.1111/j.1365-2966.2008.12689.x.

[8] R. V. Baluev. The impact of red noise in radial velocity planet searches: only three planets orbiting GJ 581? *MNRAS*, 429:2052–2068, March 2013. doi: 10.1093/mnras/sts476.

[9] R. V. Baluev. Detecting multiple periodicities in observational data with the multifrequency periodogram - I. Analytic assessment of the statistical

significance. *MNRAS*, 436:807–818, November 2013. doi: 10.1093/mnras/stt1617.

[10] R. V. Baluev. Detecting multiple periodicities in observational data with the multifrequency periodogram - II. Frequency Decomposer, a parallelized time-series analysis algorithm. *Astronomy and Computing*, 3:50–57, November 2013. doi: 10.1016/j.ascom.2013.11.003.

[11] R. V. Baluev. The Search for Periodic Components in Observational Data. *Astrophysics*, 57:434–455, September 2014. doi: 10.1007/s10511-014-9348-3.

[12] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

[13] Bradley L. Barnhart. *The Hilbert-Huang Transform: theory, applications, development*. PhD thesis, University of Iowa, 2011. URL http://ir.uiowa.edu/etd/2670.

[14] M. S. Bartlett. Smoothing Periodograms from Time-Series with Continuous Spectra. *Nature*, 161:686–687, May 1948. doi: 10.1038/161686a0.

[15] G. Bazilevskaya, A.-M. Broomhall, Y. Elsworth, and V. M. Nakariakov. A combined analysis of the observational aspects of the quasi-biennial oscillation in solar magnetic activity. *Space Science Reviews*, 186(1):359–386, Dec 2014. ISSN 1572-9672. doi: 10.1007/s11214-014-0068-0. URL https://doi.org/10.1007/s11214-014-0068-0.

[16] S. V. Berdyugina and I. Tuominen. Permanent active longitudes and activity cycles on RS CVn stars. *A&A*, 336:L25–L28, August 1998.

[17] S. V. Berdyugina and I. G. Usoskin. Active longitudes in sunspot activity: Century scale persistence. *A&A*, 405:1121–1128, July 2003. doi: 10.1051/0004-6361:20030748.

[18] S. V. Berdyugina, J. Pelt, and I. Tuominen. Magnetic activity in the young solar analog LQ Hydrae. I. Active longitudes and cycles. *A&A*, 394:505–515, November 2002. doi: 10.1051/0004-6361:20021179.

[19] Peter Bloomfield. *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons, Inc., 2 edition, 1 2005. ISBN 9780471889489.

[20] E. Böhm-Vitense. Chromospheric Activity in G and K Main-Sequence Stars, and What It Tells Us about Stellar Dynamos. *ApJ*, 657:486–493, March 2007. doi: 10.1086/510482.

[21] S. Boro Saikia, C. J. Marvin, S. V. Jeffers, A. Reiners, R. Cameron, S. C. Marsden, P. Petit, J. Warnecke, and A. P. Yadav. Chromospheric activity catalogue of 4454 cool stars. Questioning the active branch of stellar activity cycles. *ArXiv e-prints*, March 2018.

[22] R.N. Bracewell. *The Fourier Transform and Its Applications*. Electrical engineering series. McGraw Hill, 2000. ISBN 9780073039381. URL https://books.google.fi/books?id=ZNQQAQAAIAAJ.

[23] A. Brandenburg and M. S. Giampapa. Enhanced Stellar Activity for Slow Antisolar Differential Rotation? *ApJ*, 855:L22, March 2018. doi: 10.3847/2041-8213/aab20a.

[24] A. Brandenburg, S. H. Saar, and C. R. Turpin. Time Evolution of the Magnetic Activity Cycle Period. *ApJ*, 498:L51–L54, May 1998. doi: 10.1086/311297.

[25] A. Brandenburg, S. Mathur, and T. S. Metcalfe. Evolution of Co-existing Long and Short Period Stellar Activity Cycles. *ApJ*, 845:79, August 2017. doi: 10.3847/1538-4357/aa7cfa.

[26] G. L. Bretthorst. *Frequency estimation and generalized Lomb-Scargle periodograms*, pages 309–329. 2003.

[27] Larry G. Bretthorst. *Bayesian Spectrum Analysis and Parameter Estimation*, volume 48. Springer, 1988. ISBN 978-1-4684-9399-3.

[28] B. P. Brown, M. K. Browning, A. S. Brun, M. S. Miesch, and J. Toomre. Global-scale Magnetism (and Cycles) in Dynamo Simulations of Stellar Convection Zones. In C. Johns-Krull, M. K. Browning, and A. A. West, editors, *16th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*, volume 448 of *Astronomical Society of the Pacific Conference Series*, page 277, December 2011.

[29] E. Budding. The interpretation of cyclical photometric variations in certain dwarf ME-type stars. *Ap&SS*, 48:207–223, May 1977. doi: 10.1007/BF00643052.

[30] M. Carbonell, R. Oliver, and J. L. Ballester. Power spectra of gapped time series - A comparison of several methods. *A&A*, 264:350–360, October 1992.

[31] L. Cohen. Time-frequency distributions-a review. *Proceedings of the IEEE*, 77(7):941–981, Jul 1989. ISSN 0018-9219. doi: 10.1109/5.30749.

[32] L. Cohen. *Time-frequency Analysis*. Electrical engineering signal processing. Prentice Hall PTR, 1995. ISBN 9780135945322. URL `http://books.google.fi/books?id=CSKLQgAACAAJ`.

[33] E. Cole, P. J. Käpylä, M. J. Mantere, and A. Brandenburg. An Azimuthal Dynamo Wave in Spherical Shell Convection. *ApJ*, 780:L22, January 2014. doi: 10.1088/2041-8205/780/2/L22.

[34] E. M. Cole, T. Hackman, M. J. Käpylä, I. Ilyin, O. Kochukhov, and N. Piskunov. Doppler imaging of LQ Hydrae for 1998-2002. *A&A*, 581: A69, September 2015. doi: 10.1051/0004-6361/201425440.

[35] Ingrid Daubechies, Jianfeng Lu, and Hau-Tieng Wu. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Applied and Computational Harmonic Analysis*, 30(2):243 – 261, 2011. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2010.08.002. URL `http://www.sciencedirect.com/science/article/pii/S1063520310001016`.

[36] T. J. Deeming. Fourier Analysis with Unequally-Spaced Data. *Ap&SS*, 36: 137–158, August 1975. doi: 10.1007/BF00681947.

[37] E. Distefano, A. C. Lanzafame, A. Lanza, S. Messina, and F. Spada. Activity cycles in members of young loose stellar associations. *ArXiv e-prints*, June 2017.

[38] X. Dumusque, N. C. Santos, S. Udry, C. Lovis, and X. Bonfils. Planetary detection limits taking into account stellar noise. II. Effect of stellar spot groups on radial-velocities. *A&A*, 527:A82, March 2011. doi: 10.1051/0004-6361/201015877.

[39] B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994. ISBN 9780412042317. URL `https://books.google.fi/books?id=gLlpIUxRntoC`.

[40] L. Eyer and P. Bartholdi. Variable stars: Which Nyquist frequency? *A&AS*, 135:1–3, February 1999. doi: 10.1051/aas:1999102.

[41] Y. Fan and F. Fang. A Simulation of Convective Dynamo in the Solar Convective Envelope: Maintenance of the Solar-like Differential Rotation and Emerging Flux. *ApJ*, 789:35, July 2014. doi: 10.1088/0004-637X/789/1/35.

[42] F. Feng, M. Tuomi, and H. R. A. Jones. Agatha: disentangling periodic signals from correlated noise in a periodogram framework. *MNRAS*, 470: 4794–4814, October 2017. doi: 10.1093/mnras/stx1126.

[43] S. Ferraz-Mello. Estimation of Periods from Unequally Spaced Observations. *AJ*, 86:619, April 1981. doi: 10.1086/112924.

[44] P. Flandrin, G. Rilling, and P. Goncalves. Empirical mode decomposition as a filter bank. *Signal Processing Letters, IEEE*, 11(2):112–114, Feb 2004. ISSN 1070-9908. doi: 10.1109/LSP.2003.821662.

[45] G. Foster. The cleanest Fourier spectrum. *AJ*, 109:1889–1902, April 1995. doi: 10.1086/117416.

[46] G. Foster. Time Series Analysis by Projection. I. Statistical Properties of Fourier Analysis. *AJ*, 111:541, January 1996. doi: 10.1086/117805.

[47] G. Foster. Time Series Analysis by Projection. II. Tensor Methods for Time Series Analysis. *AJ*, 111:555, January 1996. doi: 10.1086/117806.

[48] G. Foster. Wavelets for period analysis of unevenly sampled time series. *AJ*, 112:1709, October 1996. doi: 10.1086/118137.

[49] Emily Fox and David B. Dunson. Multiresolution gaussian processes. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 737–745. Curran Associates, Inc., 2012. URL `http://papers.nips.cc/paper/4682-multiresolution-gaussian-processes.pdf`.

[50] F. A. M. Frescura, C. A. Engelbrecht, and B. S. Frank. Significance of periodogram peaks and a pulsation mode analysis of the Beta Cephei star V403Car. *MNRAS*, 388:1693–1707, August 2008. doi: 10.1111/j.1365-2966.2008.13499.x.

[51] Ebrahim Ghaderpour and Spiros D. Pagiatakis. Least-squares wavelet analysis of unequally spaced and non-stationary time series and its applications. *Mathematical Geosciences*, 49(7):819–844, Oct 2017. ISSN 1874-8953. doi: 10.1007/s11004-017-9691-0. URL https://doi.org/10.1007/s11004-017-9691-0.

[52] M. Ghizaru, P. Charbonneau, and P. K. Smolarkiewicz. Magnetic Cycles in Global Large-eddy Simulations of Solar Convection. *ApJ*, 715:L133–L137, June 2010. doi: 10.1088/2041-8205/715/2/L133.

[53] D. L. Gilman, F. J. Fuglister, and J. M. Mitchell Jr. On the power spectrum of "red noise". *Journal of the Atmospheric Sciences*, 20(2):182–184, 1963. doi: 10.1175/1520-0469(1963)020<0182:OTPSON>2.0.CO;2. URL https://doi.org/10.1175/1520-0469(1963)020<0182:OTPSON>2.0.CO;2.

[54] T. Hackman, M. J. Mantere, M. Lindborg, I. Ilyin, O. Kochukhov, N. Piskunov, and I. Tuominen. Doppler images of II Pegasi for 2004-2010. *A&A*, 538:A126, February 2012. doi: 10.1051/0004-6361/201117603.

[55] T. Hackman, J. Pelt, M. J. Mantere, L. Jetsu, H. Korhonen, T. Granzer, P. Kajatkari, J. Lehtinen, and K. G. Strassmeier. Flip-flops of <ASTROBJ>FK Comae Berenices</ASTROBJ>. *A&A*, 553:A40, May 2013. doi: 10.1051/0004-6361/201220690.

[56] T. Hackman, J. Lehtinen, L. Rosén, O. Kochukhov, and M. J. Käpylä. Zeeman-Doppler imaging of active young solar-type stars. *A&A*, 587:A28, March 2016. doi: 10.1051/0004-6361/201527320.

[57] D. S. Hall. Learning about stellar dynamos from long-term photometry of starspots. In I. Tuominen, D. Moss, and G. Rüdiger, editors, *IAU Colloq. 130: The Sun and Cool Stars. Activity, Magnetism, Dynamos*, volume 380 of *Lecture Notes in Physics, Berlin Springer Verlag*, page 353, 1991. doi: 10.1007/3-540-53955-7_156.

[58] J. Hartikainen and S. Särkkä. Kalman filtering and smoothing solutions to temporal gaussian process regression models. In *2010 IEEE International Workshop on Machine Learning for Signal Processing*, pages 379–384, Aug 2010. doi: 10.1109/MLSP.2010.5589113.

[59] David H. Hathaway. The solar cycle. *Living Reviews in Solar Physics*, 7(1):1, Mar 2010. ISSN 1614-4961. doi: 10.12942/lrsp-2010-1. URL https://doi.org/10.12942/lrsp-2010-1.

[60] Monson H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc., 1 edition, 4 1996. ISBN 0471594318.

[61] F. R. Holzinger and M. Benedikt. Online instantaneous frequency estimation utilizing empirical mode decomposition and hermite splines. In *2014*

22nd European Signal Processing Conference (EUSIPCO), pages 446–450, Sept 2014.

[62] J. H. Horne and S. L. Baliunas. A prescription for period analysis of unevenly sampled time series. *ApJ*, 302:757–763, March 1986. doi: 10.1086/164037.

[63] B. Huang and A. Kunoth. A different view on the vector-valued empirical mode decomposition (VEMD). *ArXiv e-prints*, February 2015.

[64] Norden E. Huang and Zhaohua Wu. A review on hilbert-huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46(2):n/a–n/a, 2008. ISSN 1944-9208. doi: 10.1029/2007RG000228. URL http://dx.doi.org/10.1029/2007RG000228.

[65] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 454(1971): 903–995, 1998. ISSN 1364-5021. doi: 10.1098/rspa.1998.0193.

[66] P. Huijse, P. A. Estévez, F. Förster, S. F. Daniel, A. J. Connolly, P. Protopapas, R. Carrasco, and J. C. Príncipe. Robust Period Estimation Using Mutual Information for Multiband Light Curves in the Synoptic Survey Era. *ApJS*, 236:12, May 2018. doi: 10.3847/1538-4365/aab77c.

[67] Daubechies Ingrid. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909–996, 1988. doi: 10.1002/cpa.3160410705. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160410705.

[68] R. L. Jennings. Symmetry breaking in a nonlinear -dynamo. *Geophysical and Astrophysical Fluid Dynamics*, 57:147–189, 1991. doi: 10.1080/03091929108225233.

[69] L. Jetsu and J. Pelt. Three stage period analysis and complementary methods. *A&AS*, 139:629–643, November 1999. doi: 10.1051/aas:1999411.

[70] M. J. Käpylä, P. J. Käpylä, N. Olspert, A. Brandenburg, J. Warnecke, B. B. Karak, and J. Pelt. Multiple dynamo modes as a mechanism for long-term solar activity variations. *A&A*, 589:A56, May 2016. doi: 10.1051/0004-6361/201527002.

[71] P. J. Käpylä, M. J. Mantere, and A. Brandenburg. Cyclic Magnetic Activity due to Turbulent Convection in Spherical Wedge Geometry. *ApJ*, 755:L22, August 2012. doi: 10.1088/2041-8205/755/1/L22.

[72] P. J. Käpylä, A. Brandenburg, N. Kleeorin, M. J. Käpylä, and I. Rogachevskii. Magnetic flux concentrations from turbulent stratified convection. *A&A*, 588:A150, April 2016. doi: 10.1051/0004-6361/201527731.

[73] Z. Kővári, A. Künstler, K. G. Strassmeier, T. A. Carroll, M. Weber, L. Kriskovics, K. Oláh, K. Vida, and T. Granzer. Time-series Doppler images and surface differential rotation of the effectively single, rapidly rotating K-giant KU Pegasi. *A&A*, 596:A53, November 2016. doi: 10.1051/0004-6361/201628425.

[74] Z. Kővári, K. Oláh, L. Kriskovics, K. Vida, E. Forgács-Dajka, and K. G. Strassmeier. Rotation-differential rotation relationships for late-type single and binary stars from Doppler imaging. *Astronomische Nachrichten*, 338:903–909, October 2017. doi: 10.1002/asna.201713400.

[75] D. M. Kipping. An analytic model for rotational modulations in the photometry of spotted stars. *MNRAS*, 427:2487–2511, December 2012. doi: 10.1111/j.1365-2966.2012.22124.x.

[76] O. Kochukhov, M. J. Mantere, T. Hackman, and I. Ilyin. Magnetic field topology of the RS CVn star II Pegasi. *A&A*, 550:A84, February 2013. doi: 10.1051/0004-6361/201220432.

[77] C. Koen. Significance testing of periodogram ordinates. *ApJ*, 348:700–702, January 1990. doi: 10.1086/168277.

[78] C. Koen. The Nyquist frequency for irregularly spaced time-series: a calculation formula. *MNRAS*, 371:1390–1394, September 2006. doi: 10.1111/j.1365-2966.2006.10762.x.

[79] D. Y. Kolotkov, S. A. Anfinogentov, and V. M. Nakariakov. Empirical mode decomposition analysis of random processes in the solar atmosphere. *A&A*, 592:A153, August 2016. doi: 10.1051/0004-6361/201628306.

[80] R. W. Komm, F. Hill, and R. Howe. Empirical Mode Decomposition and Hilbert Analysis Applied to Rotation Residuals of the Solar Convection Zone. *ApJ*, 558:428–441, September 2001. doi: 10.1086/322464.

[81] H. Korhonen, S. V. Berdyugina, I. V. Ilyin, K. G. Strassmeier, and T. Hackman. Spot evolution and active longitudes on FK Com: more than a decade of detailed surface mapping. In *Revista Mexicana de Astronomia y Astrofisica Conference Series*, volume 36 of *Revista Mexicana de Astronomia y Astrofisica, vol. 27*, pages CD323–CD327, August 2009.

[82] F. Krause and K. H. Raedler. *Mean-field magnetohydrodynamics and dynamo theory*. 1980.

[83] G. E. Kron. The Probable Detecting of Surface Spots on AR Lacertae B. *PASP*, 59:261, October 1947. doi: 10.1086/125964.

[84] J. R. Kuhn. Recovering spectral information from unevenly sampled data - Two machine-efficient solutions. *AJ*, 87:196–202, January 1982. doi: 10.1086/113096.

[85] J. Lafler and T. D. Kinman. An RR Lyrae Star Survey with Ihe Lick 20-INCH Astrograph II. The Calculation of RR Lyrae Periods by Electronic Computer. *ApJS*, 11:216, June 1965. doi: 10.1086/190116.

[86] J. Lehtinen, L. Jetsu, T. Hackman, P. Kajatkari, and G. W. Henry. The continuous period search method and its application to the young solar analogue HD 116956. *A&A*, 527:A136, March 2011. doi: 10.1051/0004-6361/201015454.

[87] J. Lehtinen, L. Jetsu, T. Hackman, P. Kajatkari, and G. W. Henry. Spot activity of <ASTROBJ>LQ Hydra</ASTROBJ> from photometry between 1988 and 2011. *A&A*, 542:A38, June 2012. doi: 10.1051/0004-6361/201219185.

[88] J. Lehtinen, L. Jetsu, T. Hackman, P. Kajatkari, and G. W. Henry. Activity trends in young solar-type stars. *A&A*, 588:A38, April 2016. doi: 10.1051/0004-6361/201527420.

[89] M. Lindborg, M. J. Mantere, N. Olspert, J. Pelt, T. Hackman, G. W. Henry, L. Jetsu, and K. G. Strassmeier. Multiperiodicity, modulations and flip-flops in variable star light curves. II. Analysis of II Pegasus photometry during 1979-2010. *A&A*, 559:A97, November 2013. doi: 10.1051/0004-6361/201321695.

[90] A. F. Linnell Nemec and J. M. Nemec. A test of significance for periods derived using phase-dispersion-minimization techniques. *AJ*, 90:2317–2320, November 1985. doi: 10.1086/113936.

[91] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Ap&SS*, 39:447–462, February 1976. doi: 10.1007/BF00648343.

[92] X. Luo, K. Wang, X. Zhang, L. Deng, Y. Luo, and C. Luo. Rapid Light-curve Changes and Probable Flip-flop Activity of the W UMa-type Binary V410 Aur. *AJ*, 154:99, September 2017. doi: 10.3847/1538-3881/aa8325.

[93] M. J. Mantere, P. J. Käpylä, and J. Pelt. Role of longitudinal activity complexes for solar and stellar dynamos. In A. G. Kosovichev, E. de Gouveia Dal Pino, and Y. Yan, editors, *Solar and Astrophysical Dynamos and Magnetic Activity*, volume 294 of *IAU Symposium*, pages 175–186, July 2013. doi: 10.1017/S1743921313002457.

[94] T. S. Metcalfe and J. van Saders. Magnetic Evolution and the Disappearance of Sun-Like Activity Cycles. *Sol. Phys.*, 292:126, September 2017. doi: 10.1007/s11207-017-1157-5.

[95] A. Mortier and A. Collier Cameron. Stacked Bayesian general Lomb-Scargle periodogram: Identifying stellar activity signals. *A&A*, 601:A110, May 2017. doi: 10.1051/0004-6361/201630201.

[96] A. Mortier, J. P. Faria, C. M. Correia, A. Santerne, and N. C. Santos. BGLS: A Bayesian formalism for the generalised Lomb-Scargle periodogram. *A&A*, 573:A101, January 2015. doi: 10.1051/0004-6361/201424908.

[97] Kevin P Murphy. *Machine learning: a probabilistic perspective*. Cambridge, MA, 2012.

[98] R. W. Noyes, N. O. Weiss, and A. H. Vaughan. The relation between stellar rotation rate and activity cycle periods. *ApJ*, 287:769–773, December 1984. doi: 10.1086/162735.

[99] K. Oláh, Z. Kővári, K. Petrovay, W. Soon, S. Baliunas, Z. Kolláth, and K. Vida. Magnetic cycles at different ages of stars. *A&A*, 590:A133, June 2016. doi: 10.1051/0004-6361/201628479.

[100] S. J. O'Toole, C. G. Tinney, and H. R. A. Jones. The impact of stellar oscillations on doppler velocity planet searches. *MNRAS*, 386:516–520, May 2008. doi: 10.1111/j.1365-2966.2008.13061.x.

[101] O. Özdarcan, T. A. Carroll, A. Künstler, K. G. Strassmeier, S. Evren, M. Weber, and T. Granzer. Time-series Doppler imaging of the red giant HD 208472. Active longitudes and differential rotation. *A&A*, 593:A123, October 2016. doi: 10.1051/0004-6361/201628545.

[102] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Series in Electrical Engineering. McGraw-Hill, 1984. ISBN 9780070484689. URL `https://books.google.fi/books?id=3NRQAAAAMAAJ`.

[103] J. R. Parks, R. J. White, F. Baron, J. D. Monnier, B. Kloppenborg, G. Henry, G. Scheafer, X. Che, E. Pedretti, N. Thureau, M. Zhao, T. ten Brummelaar, H. McAlister, S. T. Ridgway, N. Turner, J. Sturmann, and L. Sturmann. First Images of Cool Starspots on a Star Other than the Sun: Interferometric Imaging of λ Andromedae. *ArXiv e-prints*, August 2015.

[104] J. Pelt. Phase dispersion minimization methods for estimation of periods from unequally spaced sequences of data. In E. J. Rolfe, editor, *Statistical Methods in Astronomy*, volume 201 of *ESA Special Publication*, November 1983.

[105] J. Pelt. *Non-parametric methods for shift and periodicity delection in irregularly measured data*, page 249. 1997.

[106] J. Pelt, I. Tuominen, and J. Brooke. Century-scale persistence in longitude distribution in the Sun and in silico. *A&A*, 429:1093–1096, January 2005. doi: 10.1051/0004-6361:20041357.

[107] J. Pelt, N. Olspert, M. J. Mantere, and I. Tuominen. Multiperiodicity, modulations and flip-flops in variable star light curves. I. Carrier fit method. *A&A*, 535:A23, November 2011. doi: 10.1051/0004-6361/201116882.

[108] F. Pont, S. Zucker, and D. Queloz. The effect of red noise on planetary transit detection. *MNRAS*, 373:231–242, November 2006. doi: 10.1111/j.1365-2966.2006.11012.x.

[109] B. Porat. *Digital Processing of Random Signals: Theory and Methods*. Dover Books on Electrical Engineering. Dover Publications, 2008. ISBN 9780486462981. URL `https://books.google.fi/books?id=1vAuDQAAQBAJ`.

[110] W. H. Press and G. B. Rybicki. Fast algorithm for spectral analysis of unevenly sampled data. *ApJ*, 338:277–280, March 1989. doi: 10.1086/167197.

[111] Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006. ISBN 0-262-18253-X.

[112] R.T. Rato, M.D. Ortigueira, and A.G. Batista. On the hht, its problems, and some solutions. *Mechanical Systems and Signal Processing*, 22(6):1374 – 1394, 2008. ISSN 0888-3270. doi: https://doi.org/10.1016/j.ymssp.2007.11.028. URL http://www.sciencedirect.com/science/article/pii/S0888327007002701. Special Issue: Mechatronics.

[113] N. Rehman and D. P. Mandic. Multivariate empirical mode decomposition. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 466(2117):1291–1302, 2010. ISSN 1364-5021. doi: 10.1098/rspa.2009.0502. URL http://rspa.royalsocietypublishing.org/content/466/2117/1291.

[114] T. Reinhold and L. Gizon. Rotation, differential rotation, and gyrochronology of active Kepler stars. *A&A*, 583:A65, November 2015. doi: 10.1051/0004-6361/201526216.

[115] T. Reinhold, A. Reiners, and G. Basri. Rotation and differential rotation of active Kepler stars. *A&A*, 560:A4, December 2013. doi: 10.1051/0004-6361/201321970.

[116] T. Reinhold, R. H. Cameron, and L. Gizon. Evidence for photometric activity cycles in 3203 Kepler stars. *ArXiv e-prints*, May 2017.

[117] Sami Remes, Markus Heinonen, and Samuel Kaski. Nonstationary spectral kernels. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4642–4651. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/7050-non-stationary-spectral-kernels.pdf.

[118] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Processing Magazine*, 8(4):14–38, Oct 1991. ISSN 1053-5888. doi: 10.1109/79.91217.

[119] D. H. Roberts, J. Lehar, and J. W. Dreher. Time Series Analysis with Clean - Part One - Derivation of a Spectrum. *AJ*, 93:968, April 1987. doi: 10.1086/114383.

[120] L. Rosén, O. Kochukhov, and G. A. Wade. First Zeeman Doppler Imaging of a Cool Star Using all Four Stokes Parameters. *ApJ*, 805:169, June 2015. doi: 10.1088/0004-637X/805/2/169.

[121] S. H. Saar and A. Brandenburg. Time Evolution of the Magnetic Activity Cycle Period. II. Results for an Expanded Stellar Sample. *ApJ*, 524:295–310, October 1999. doi: 10.1086/307794.

[122] A. Saha and A. K. Vivas. A Hybrid Algorithm for Period Analysis from Multiband Data with Sparse and Irregular Sampling for Arbitrary Light-curve Shapes. *AJ*, 154:231, December 2017. doi: 10.3847/1538-3881/aa8fd3.

[123] M. Sandsten. *Time-Frequency Analysis of Non-Stationary Processes. An Introduction.* Lund University, Centre for Mathematical Sciences, 2013. URL http://www.maths.lth.se/matstat/kurser/masm26/2013/TIMEFREQkompendie.pdf.

[124] Simo Särkkä. *Bayesian Filtering and Smoothing*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2013. doi: 10.1017/CBO9781139344203.

[125] J. D. Scargle. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *ApJ*, 263:835–853, December 1982. doi: 10.1086/160554.

[126] C. J. Schrijver, J. Cote, C. Zwaan, and S. H. Saar. Relations between the photospheric magnetic field and the emission from the outer atmospheres of cool stars. I - The solar CA II K line core emission. *ApJ*, 337:964–976, February 1989. doi: 10.1086/167168.

[127] M. Schrinner, L. Petitdemange, and E. Dormy. Oscillatory dynamos and their induction mechanisms. *A&A*, 530:A140, June 2011. doi: 10.1051/0004-6361/201016372.

[128] Arthur 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. ISSN 0272-7528. doi: 10.1029/TM003i001p00013. URL http://dx.doi.org/10.1029/TM003i001p00013.

[129] A. Schwarzenberg-Czerny. Accuracy of period determination. *MNRAS*, 253:198–206, November 1991. doi: 10.1093/mnras/253.2.198.

[130] A. Schwarzenberg-Czerny. Fast and Statistically Optimal Period Search in Uneven Sampled Observations. *ApJ*, 460:L107, April 1996. doi: 10.1086/309985.

[131] A. Schwarzenberg-Czerny. The distribution of empirical periodograms: Lomb-Scargle and PDM spectra. *MNRAS*, 301:831–840, December 1998. doi: 10.1046/j.1365-8711.1998.02086.x.

[132] M. J. Shensa. The discrete wavelet transform: wedding the a trous and mallat algorithms. *IEEE Transactions on Signal Processing*, 40(10):2464–2482, Oct 1992. ISSN 1053-587X. doi: 10.1109/78.157290.

[133] S. K. Solanki, B. Inhester, and M. Schüssler. The solar magnetic field. *Reports on Progress in Physics*, 69:563–668, March 2006. doi: 10.1088/0034-4885/69/3/R02.

[134] Arno Solin and Simo Särkkä. Explicit link between periodic covariance functions and state space models. In *Artificial Intelligence and Statistics*, pages 904–912, 2014.

[135] R. F. Stellingwerf. Period determination using phase dispersion minimization. *ApJ*, 224:953–960, September 1978. doi: 10.1086/156444.

[136] M. Stix. *The Sun: An Introduction*. Astronomy and Astrophysics Library. Springer Berlin Heidelberg, 2004. ISBN 9783540207412. URL https://books.google.fi/books?id=wxHN9jP-mNMC.

[137] R. W. Tanner. Spurious Periods in Spectroscopic Binaries. *JRASC*, 42:177, August 1948.

[138] J. Taylor and S. Hamilton. Some tests of the Vaníček Method of spectral analysis. *Ap&SS*, 17:357–367, August 1972. doi: 10.1007/BF00642907.

[139] D. J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, Sept 1982. ISSN 0018-9219. doi: 10.1109/PROC.1982.12433.

[140] Michalis K Titsias. Variational learning of inducing variables in sparse gaussian processes. In *International Conference on Artificial Intelligence and Statistics*, pages 567–574, 2009.

[141] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin. A complete ensemble empirical mode decomposition with adaptive noise. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4144–4147, May 2011. doi: 10.1109/ICASSP.2011.5947265.

[142] Richard Turner and Maneesh Sahani. Probabilistic amplitude and frequency demodulation. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 981–989. Curran Associates, Inc., 2011. URL http://papers.nips.cc/paper/4284-probabilistic-amplitude-and-frequency-demodulation.pdf.

[143] I. G. Usoskin, R. Arlt, E. Asvestari, E. Hawkins, M. Käpylä, G. A. Kovaltsov, N. Krivova, M. Lockwood, K. Mursula, J. O'Reilly, M. Owens, C. J. Scott, D. D. Sokoloff, S. K. Solanki, W. Soon, and J. M. Vaquero. The Maunder minimum (1645-1715) was indeed a grand minimum: A reassessment of multiple datasets. *A&A*, 581:A95, September 2015. doi: 10.1051/0004-6361/201526652.

[144] J. T. VanderPlas. Understanding the Lomb-Scargle Periodogram. *ArXiv e-prints*, March 2017.

[145] J. T. VanderPlas and Ž. Ivezić. Periodograms for Multiband Astronomical Time Series. *ApJ*, 812:18, October 2015. doi: 10.1088/0004-637X/812/1/18.

[146] P. Vaníček. Further Development and Properties of the Spectral Analysis by Least-Squares. *Ap&SS*, 12:10–33, July 1971. doi: 10.1007/BF00656134.

[147] A. H. Vaughan and G. W. Preston. A survey of chromospheric CA II H and K emission in field stars of the solar neighborhood. *PASP*, 92:385–391, August 1980. doi: 10.1086/130683.

[148] S. Vaughan. A simple test for periodic signals in red noise. *A&A*, 431: 391–403, February 2005. doi: 10.1051/0004-6361:20041453.

[149] S. Vaughan. A Bayesian test for periodic signals in red noise. *MNRAS*, 402:307–320, February 2010. doi: 10.1111/j.1365-2966.2009.15868.x.

[150] Aki Vehtari and Janne Ojanen. A survey of bayesian predictive methods for model assessment, selection and comparison. *Statist. Surv.*, 6:142–228, 2012. doi: 10.1214/12-SS102. URL https://doi.org/10.1214/12-SS102.

[151] M. Vetterli and C. Herley. Wavelets and filter banks: theory and design. *IEEE Transactions on Signal Processing*, 40(9):2207–2232, Sep 1992. ISSN 1053-587X. doi: 10.1109/78.157221.

[152] R. Vio, P. Andreani, and A. Biggs. Unevenly-sampled signals: a general formalism for the Lomb-Scargle periodogram. *A&A*, 519:A85, September 2010. doi: 10.1051/0004-6361/201014079.

[153] R. Vio, M. Diaz-Trigo, and P. Andreani. Irregular time series in astronomy and the use of the Lomb-Scargle periodogram. *Astronomy and Computing*, 1:5–16, February 2013. doi: 10.1016/j.ascom.2012.12.001.

[154] M. Viviani, J. Warnecke, M. J. Käpylä, P. J. Käpylä, N. Olspert, E. M. Cole-Kodikara, J. J. Lehtinen, and A. Brandenburg. Transition from axi- to nonaxisymmetric dynamo modes in spherical convection models of solar-like stars. *ArXiv e-prints*, October 2017.

[155] J. Warnecke, M. Rheinhardt, S. Tuomisto, P. J. Käpylä, M. J. Käpylä, and A. Brandenburg. Turbulent transport coefficients in spherical wedge dynamo simulations of solar-like stars. *A&A*, 609:A51, January 2018. doi: 10.1051/0004-6361/201628136.

[156] O. C. Wilson. Chromospheric variations in main-sequence stars. *ApJ*, 226: 379–396, December 1978. doi: 10.1086/156618.

[157] N. J. Wright and J. J. Drake. Solar-type dynamo behaviour in fully convective stars without a tachocline. *Nature*, 535:526–528, July 2016. doi: 10.1038/nature18638.

[158] N. J. Wright, J. J. Drake, E. E. Mamajek, and G. W. Henry. The Stellar-activity-Rotation Relationship and the Evolution of Stellar Dynamos. *ApJ*, 743:48, December 2011. doi: 10.1088/0004-637X/743/1/48.

[159] Zhaohua Wu and Norden E. Huang. A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 460(2046):1597–1611, 2004. ISSN 1364-5021. doi: 10.1098/rspa.2003.1221. URL http://rspa.royalsocietypublishing.org/content/460/2046/1597.

[160] Zhaohua Wu and Norden E. Huang. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 01(01):1–41, 2009. doi: 10.1142/S1793536909000047. URL http://www.worldscientific.com/doi/abs/10.1142/S1793536909000047.

[161] N. B. Xiang and Z. N. Qu. Ensemble Empirical Mode Decomposition of the Magnetic Field of the Sun as a Star. *AJ*, 151:76, March 2016. doi: 10.3847/0004-6256/151/3/76.

[162] H. Yoshimura. Nonlinear astrophysical dynamos - The solar cycle as a nonlinear oscillation of the general magnetic field driven by the nonlinear dynamo and the associated modulation of the differential-rotation-global-convection system. *ApJ*, 220:692–711, March 1978. doi: 10.1086/155953.

[163] M. Zechmeister and M. Kürster. The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. *A&A*, 496:577–584, March 2009. doi: 10.1051/0004-6361:200811296.

Erratum

Publication IV

Eq. (4) should read $g = \mathbb{I}(|t_i - t_j| < t_{\mathrm{coh}})\mathbb{I}\left(\min\left(\mathrm{frac}(\nu|\mathrm{t_i} - \mathrm{t_j}|), 1 - \mathrm{frac}(\nu|\mathrm{t_i} - \mathrm{t_j}|)\right) < \epsilon\right)$

Aalto-DD 224/2018

ISBN 978-952-60-8292-9 (printed)
ISBN 978-952-60-8293-6 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Computer Science
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS