PERIOD ANALYSIS OF STELLAR BRIGHTNESS VARIATIONS USING MULTICHANNEL PHOTOMETRIC OBSERVATIONS

JIANPENG YOU

Division of Astronomy
Department of Physical Sciences
University of Oulu
Finland

Academic dissertation to be presented, with the permission of the Faculty of Science of the University of Oulu, for public discussion in the Auditorium Anttilansali, Linnanmaa, on 14 December, 2007, at 12 o'clock noon.

REPORT SERIES IN PHYSICAL SCIENCES Report No. 45

OULU 2007 ● UNIVERSITY OF OULU

Reviewers
Prof. Werner W. Weiss, University of Vienna, Austria
Prof. Nikolai Piskunov, University of Uppsala, Sweden

Opponent
Doc. Harry J. Lehto, University of Turku, Finland

Custos
Prof. Juri Poutanen, University of Oulu, Finland

ISBN 978-951-42-8642-1
ISBN 978-951-42-8643-8 (PDF)
ISSN 1239-4327

Oulu University Press
Oulu 2007

Abstract

This dissertation presents time series analysis (TSA) on the simultaneous multichannel photometric observations to study stellar differential rotation with a novel three stage, multichannel period analysis (MPA).

The traditional TSA methods are typically applied only to one channel photometric observations. This PhD work devises a new MPA method, based on the traditional TSA technique, to analyze simultaneously all channel data together. The applications of MPA focus on analyzing the photometric observations of brightness variations of stars, studying stellar surface spottedness and differential rotation. Similarly to the solar differential rotation, which can be studied using sunspots, stellar differential rotation can be investigated from photometric variations caused by surface spots, which can be observed instrumentally and analyzed by different TSA techniques.

The PhD work shows that the MPA method has advantages both in recovering physical periodicities and in achieving higher accuracies of the final results over the traditional TSA applied to the single channel data. It also displays that the MPA method is able to detect the stellar differential rotation from the photometric variability and its applications can help understanding the stellar differential rotation, the magnetic field evolution and the physics of dynamos.

Acknowledgments

I would like to express my gratitude to all those who gave me the possibility to complete this PhD thesis.

At first, I want to thank the Department of Physical Sciences, University of Oulu, for giving me permission to commence this thesis and for supporting me continuously to do the necessary research work and to use the facilities during the entire period.

I have furthermore to thank Prof. Juri Poutanen, Prof. Heikki Salo and Dr. Eija Laurikainen for encouraging me to go ahead with my PhD when I was having a difficult time, for helping to arrange financial supports for the final stage of the thesis and, especially Prof. Heikki Salo and Prof. Juri Poutanen looking closely at the final version of the thesis for English style and grammar, correcting both and offering suggestions for improvement.

I am bound to the honorable Dr. Jaan Pelt and to Dr. Lauri Jetsu for their tutorial help in developing the entire package of MPA method. I am deeply indebted to Dr. Gregory W. Henry from the Center of Excellence in Information Systems, Tennessee State University, USA. Without his help, it would not be possible for Paper III to be published and thus to become the essential part in the thesis. I want to thank Dr. Slavek Rucinski from the David Dunlap Observatory, University of Toronto, Canada, for giving me the chance of the four month research experience there. My former colleagues Dr. Andrei Berdyugin and Dr. Rudolf Duemmler helped me in the early time of this thesis work. I want to thank both of them for all their help, support, interest and valuable hints.

I am obliged to so many persons who help me during the process of the thesis work. Please do excuse for not listing them all but only the persons in my memory: Heidi Korhonen, Michael Andersen, Ilya Ilyin, Svetlana Ilyina, Jouko Raitala, Kauko Kauhanen, Terhi Törmänen, Pertti Rautiainen, Raine Karjalainen, Petri Käpylä and, the friends who once share good time with me in Oulu: Chunlei Han and his family, Deqi Huang and his family, Irene and Bryan Dopp, Lijun Chi and Longxiu Pan, Seppo Mäki, Tapani Sillanpää, Wei Xiong and his family, Yangfen Lin and Shaobin Zhang. I would give my special thanks to my wife who has been accompanying me with patient love and bringing up my son, through the entire period of my PhD study and thus enabled me to complete this dissertation.

Finally, I would like to thank Askar Ibragimov, for helping me out with software and thus enabling me to continue this thesis after I returned to Toronto from Oulu. My sincerely thanks also go to Prof. Nikolai Piskunov from Uppsala Observatory, Sweden, and Prof. Werner W. Weiss from Department of Astronomy, University of Vienna, Austria for evaluating the thesis and for a number of useful comments and suggestions. I also give my thanks to Irina Malinovskaya for her help with formatting the thesis and making the cover pages.

List of original papers

Paper I: Jetsu L., Hackman T., Hall D. S., Henry G. W., Kokko M., You J., 2000, Time Series Analysis of V815 Herculis Photometry Between 1984 and 1998, *Astron. Astrophys.*, 362, 223–235

Paper II: You J., Pelt J., Tuominen I., 2000, Period analysis for simultaneous multichannel photometric observations, *Astron. Astrophys. Suppl. Ser.*, 146, 397–406

Paper III: You J., Duemmler R., 2005, Photometric periods of HN Peg from multichannel analysis, *Astron. Astrophys.*, 433, 297–304

Paper IV: You J., 2007, Photometric periods of LQ Hydrae between 1992 and 2000, *Astron. Astrophys.*, 475, 309–315

Paper V: You J., 2007, Complete and componentwise modeling of multiple periods in photometric observations, manuscript, 13 pages

Contents

Chapter 1

Introduction

This PhD thesis presents time series analysis (TSA) of stellar photometric variability with three stage multichannel period analysis (MPA). It focuses on studying stellar differential rotation using modulations in the lightcurves produced by starspots.

Time series analysis (TSA) is used to search periodicities occurring in a process in time. In astronomy, TSA searches for hidden periods and reconstructs corresponding phase process diagrams (PPDs) from observations (typically, photometric observations), to explore the real physical processes. The applications include variable and binary stars, the effect of spottedness on stellar photometry, sunspot variations and magnetic cycles, stellar oscillations and differential rotation modulated by starspots.

Starspots are most commonly produced by stellar magnetic activity. The stellar magnetic fields are generated in the interior and emerge onto the surface causing starspots. These surface spots are tracers of stellar activity and can be relatively long-lived. They are distributed inhomogeneously over the stellar surface and modulate the brightness and spectral line profiles as the star rotates.

The Sun is the only star which surface can be directly observed. Our knowledge of the Sun is the foundation in the study of other stars, which are only point sources in our observations. The stellar surface structure is out of direct observation and can only be reconstructed with the mathematical models (for example, using Doppler imaging techniques). The rotation period of the Sun is about 30 days. At different latitudes it rotates with a different rate (so called differential rotation). The Sun's equator is rotating faster than its poles: the equator rotates approximately four times during the time the poles make three turns. The Sun's rotation can be monitored by sunspots on its surface. Sunspots alter the integrated sunlight by less than 0.1%, so that the effect is too small to be detectable in other main sequence stars. However, due to enhanced magnetic activity in some stars such as BY Dra, RS CVn and T Tau, their starspots are significantly larger. Especially for RS CVn

stars, the spot region may cover as much as 50% of their surfaces. Thus, photometric variability of these star is able to tell us about their surface structure and differential rotation.

In the astronomical observations, stellar brightness variations are closely connected with variations in activity of the stellar magnetic field. The magnetic field governs the structure and the energy balance of the outer layers of the star and varies on a wide range of spatial and temporal scales. The brightness variations in stars can reach an order of 20% over periods of days, or a few weeks. This indicates that the surface of such stars is not uniformly luminous. Pickering (1880) was the first to suggest the starspot idea more than a century ago. He assumed that star rotation carries these variable or 'spotted' regions on and off the visible hemisphere, giving rise to the brightness and color modulations. Nowadays, astronomers find that the starspot model is one hypothesis which can successfully account for the wide diversity of photometric variations in these stars (see Walker et al., 2007; Croll et al., 2006; Jeffers et al., 2007).

Starspot lifetimes may exceed several years. For spotted stars, accurate determination of the rotation periods (usually short, a few days) and their differential rotation rate is important. For example, measurements of the solar differential rotation set an observational boundary condition on the magnitude of the internal differential rotation. In the hydromagnetic dynamo theory, the interaction of internal differential rotation, convection, and turbulence generates and sustains stellar magnetic fields (see Rüdiger and Hollerbach, 2004; Brandenburg and Dobler, 2002; Kitchatinov and Rüdiger, 1999). However, no current dynamo model can fully recreate the spatial and temporal properties of the Solar magnetic field. The study of stellar differential rotation in general may help to solve the problems and improve our understanding of the Sun and the stars. Thus the measurement of stellar differential rotation is of fundamental importance to the magnetic dynamo theory and can be studied with TSA of the photometric observations.

Astronomical photometric observations are usually obtained through different wavelength filters (channels) simultaneously. Traditional TSA includes phase dispersion minimization methods, standard discrete Fourier analysis, string length methods, information entropy minimization method etc. All methods can be classified into the parametric and nonparametric categories. So far, these methods have been typically applied only upon one single channel at a time (observational data from one filter).

It is necessary to devise a TSA method which can use simultaneously all the multichannel data. For example, Jetsu and Pelt (1999) analyzed all channels together after normalizing the data from different channels; however, some information is inevitably lost in this process, in particular if the lightcurves and/or intrinsic errors are very different in the different channels. In this PhD work, we generalize the three stage weighted period analysis

method (Jetsu and Pelt, 1999) to include all channels simultaneously using the original data. Like its predecessor, the new MPA method is based on minimizing the sum of dispersions and the residual sum of squares, now over all channels simultaneously. The three stages are the phase dispersion minimization, the linear modeling and the nonlinear modeling stage. The new method has advantages both in searching for the real periods and in computing the final values with higher accuracies than what can be obtained from the analysis of separate single channels.

In this thesis the new MPA method is applied on the stellar multichannel photometric observation data, focusing on the study of the periodic light variations due to starspots and differential rotation. Chapter 2 reviews the mathematical foundation of TSA and the development of our MPA. Chapter 3 reviews solar and stellar differential rotation, magnetic fields and dynamos with recent development. A summary of the relevant publications is given at the end in both chapters.

Chapter 2

From fundamental time series analysis to multichannel period analysis

Time series analysis (TSA) is commonly needed in astronomy in a wide range of applications. In photometric observations (magnitudes of a variable star), wrapping the data points with a physical period in a so-called phase process diagram (PPD) gives a corresponding light curve. The light curve can show many different shapes corresponding to different types of stellar photometric variability. In many cases we don't know what kind of curve and periodicities to expect. From a wide range of possible periods, we can choose one or more candidate values for the true physical periods. When the approximate values of the period are inspected, our task is either to reject the false possibilities or to derive more accurate values of the true periods, using certain TSA methods. If the data are equally spaced in time, standard discrete Fourier transform (FT) techniques (Press et al., 1994,p. 499ff) can be applied routinely for finding periodicities. However, the astronomical time series are more often unevenly sampled and finite (e.g. due to telescope time availability, weather conditions and moonless night constraints), which makes TSA difficult.

In this PhD thesis, TSA is applied to astronomical photometric observations. For a given observational time series, we need to search for any recurrence in the data series, because any regular or semi-regular periodicity may be associated with physical mechanisms of the underlying object. The photometric observations are usually carried out simultaneously in different wavelength passbands (channels) and the data collection becomes multichannel. We therefore extend the traditional single channel TSA into a three stage multichannel period analysis (MPA), specially designed for multichannel data. We will show that MPA has advantages (Paper II) when analyzing all channels together, if compared with the traditional methods where each

channel is analyzed separately. In particular, this holds for the most complex tasks arising from those photometric data, which have a low accuracy and an insufficient number of points.

Because most astronomical observations are unequally-spaced in time, specialized TSA algorithms have been developed. In particular, standard discrete Fourier analysis becomes complicated because of irregular sampling (Scargle, 1989). This problem becomes even worse when the searched periods exceed one day, due to the typical one day gap between the nightly observations (Scargle, 1982). Lafler and Kinman (1965) and Renson (1978) devised a method, belonging to the general class of 'string length methods'. In their method the period is determined by minimizing the average distance between consecutive observations in the phase curve (observed data as a function of phase, calculated for each trial period). For each trial period, the data have to be reordered according to phase. On the other hand, modified Fourier methods have also been developed to deal with irregular data (see Cuypers, 1987,for a review).

The phase dispersion minimization (PDM) methods (Marraco and Muzzio, 1980) provide an alternative to the modified Fourier techniques and are much faster than the string length methods. The phase of every observation is evaluated for an assumed period, data are binned according to phase, and the sum of the bin variances (the sum of the squared phase dispersions) is computed. The period is then obtained by minimizing this sum. Cincotta et al. (1995) and Cincotta et al. (1999) developed a new method based on the minimization of the information entropy of the phase curve. The idea is that the phase curve, constructed using the correct period, is ordered, while all the other phase curves are not. The advantage is that it can deal with very irregularly sampled data, for example data with concentrations separated by very large gaps.

All the methods of TSA for searching periods can be classified into two basic types: parametric and non-parametric methods. The parametric methods are based on measuring the data dispersion around the phase curve, which is described by a parametric model. The least squares (LS) method with the corresponding LS statistic is the most common of the parametric methods. Non-parametric methods compute the phase dispersion of the observed data without any models. The PDM methods with the corresponding statistics of minimizing the sum of dispersions are usually applied in the non-parametric approach. The main deficiency of parametric modeling is the lack of criteria for choosing the model components. The non-parametric methods are more sensitive to the errors and the noise in the observational data while trying to minimize the sum of dispersions without a specific model. The PDM methods also lack criteria in the selection of optimal binning of the data points when calculating the sum of dispersions.

After detection of a period from an observational series, one needs to

obtain a final value of the period with a reasonable accuracy, using one of the many refinement methods (a different method can be optimal, depending on the properties of the data). The parametric modeling is usually used for the final refinement process. We are seeking the periodicities and are concerned about their accuracy but have less interest in fitting the precise shape of the lightcurve. Fortunately, the lightcurve fitting inaccuracy influences mostly the Residual Sum of Squares (RSS; this is the most common LS statistic) rather than the periodicities we are seeking. Although a slight change in the model can greatly increase or decrease the value of RSS, it does not change much the position of the minimum of RSS.It thus allows the periods and their accuracies to be optimized.

In the following sections, we review the concepts and techniques from the mathematical background of classical TSA to our MPA. We start from the description of the PPDs and the spurious periods which might arise due to characteristics of the data sampling. Both are basic concepts needed to understand the mathematical foundation of TSA. Then we describe the basic idea of PDM methods and its several realizations, the analytics of the wildly used FT and the LS modeling methods for the final refinement of the periods. Finally, we describe the development of MPA and summarize the results of Papers II and V.

2.1 Phase process diagram and spurious periods

Phase Process Diagram

A time series of astronomical observations can be described as $y(t_i), i = 1, 2, 3, ..., N$, where $y(t_i)$ are the observed values at times t_i and N is the number of the observations.

The basic idea for searching an underlying period from the observational data relies on evaluation of the PPD. We assign a trial period P and compute the phase for each data point for this particular trial period

$$\phi_P(t_i) = \mathrm{Frac}\frac{t_i - t_0}{P}, \tag{2.1}$$

where Frac denotes the fractional part of a real number, such as $\mathrm{Frac}(9.7) = 0.7$ and $\mathrm{Frac}(-9.7) = \mathrm{Frac}(10 - 9.7) = 0.3$. The reference time point for computing the phase curve is denoted by t_0. It can be chosen arbitrarily and its influence on the phase curve is only that of shifting the curve in phase; for simplicity, we assign $t_0 = 0$ in the following discussion. For the evaluation of each trial period we plot the implied PPD, which shows all observed values $y(t_i)$, versus the corresponding phase $\phi_P(t_i)$. If the trial

period is close to the true period, all data points lie along the mean curve of the periodic process. Otherwise, the data points in the PPD will be more or less randomly distributed. Figure 2.1 shows the PPDs of the correct period for the multichannel photometric observations of ER Vul (see also You (1999), Chapter 12).

The simplest way to search for the correct period is to build PPDs for a wide range of trial periods $P_l, l = 1, 2, ..., L$, and look for the most promising candidates, e.g. those yielding the smallest summed bin variance.

Considering the number of different PPDs needed for searching the correct periods, we usually compute trials with a fixed step in the frequency domain $f = 1/P$.

$$f_l = (l-1)\Delta f + f_{\min},\ l = 1, 2, ..., L, \qquad f_L = f_{\max}, \tag{2.2}$$

where $[f_{\min}, f_{\max}]$ is the range of frequencies of interest. The step Δf must be chosen so that the maximum shift of a data point in the PPD, when moving from one trial to the next, does not exceed some prescribed value $\Delta\phi$:

$$\Delta f = \frac{\Delta\phi}{t_N - t_1}. \tag{2.3}$$

The maximum of $\Delta\phi$ is 0.5, because a shift by e.g. $\Delta\phi = 0.7$ equals to the shift $\Delta\phi = 0.3$ in the opposite direction. While the number of trials increases with decreasing $\Delta\phi$, a small $\Delta\phi$ may be necessary for not missing some periods, which may not be found in the case of too large step of trial periods. Normally, even for a coarse search, values of $\Delta\phi$ are chosen from 0.05 to 0.1. The number of necessary trials can be found as

$$L = \frac{f_{\max} - f_{\min}}{\Delta f} = \frac{(f_{\max} - f_{\min})(t_N - t_1)}{\Delta\phi}. \tag{2.4}$$

The number L may be rather large, even for the modest data sets. Because checking large number of trials by computing and comparing the corresponding PPDs takes much time, the search range of periods must be narrowed to some degree for computations (see Chapter 2, Section 2.2).

Spurious Periods

A small scatter of the observed data points around the mean curve in the PPD is not a proof for having found the correct period. If P is the true period, then the periods $P_k = kP, k = 2, 3, ...$ also display PPDs with small dispersion. These 'undertones' are easy to identify.

The real *spurious* periods occur, when the time points t_i, $i = 1, 2, ..., N$, happen to be a subset of a regular mesh

$$t_i = k_i\tau + t_1, \quad i = 1, 2, ..., N, \tag{2.5}$$

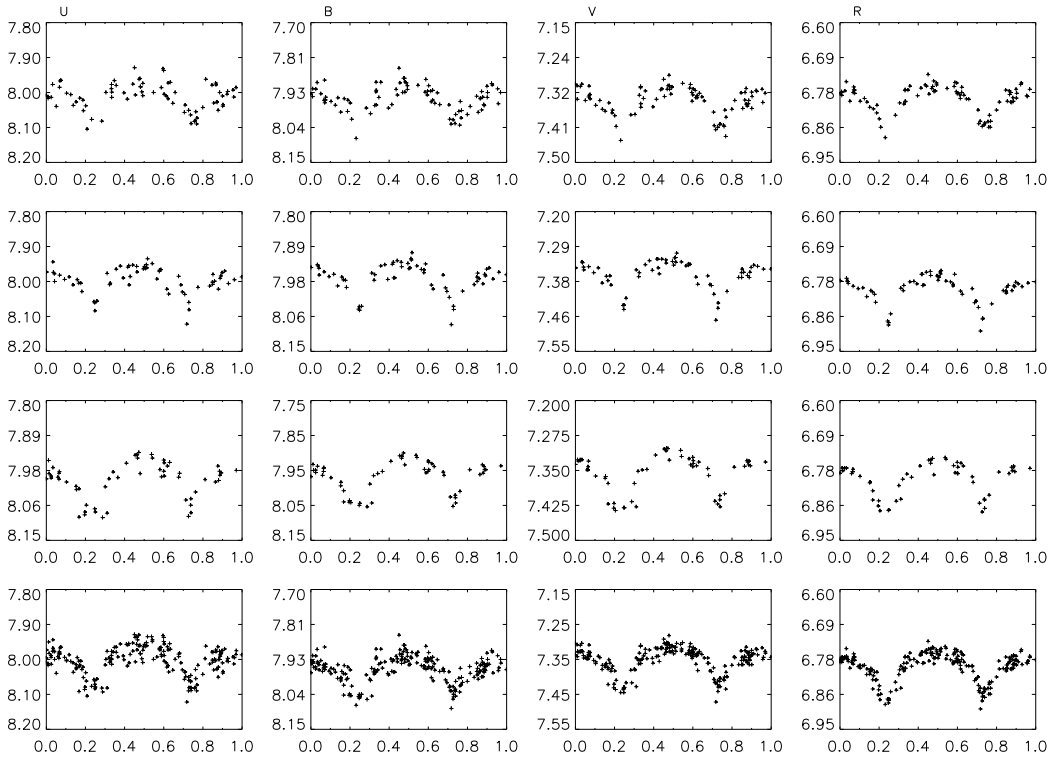

Figure 2.1: PPDs of ER Vul in multiply channels U, B, V and R (from left to right). From top to bottom, the PPDs are in the sequence of three observing seasons and all three seasons combined. The photometric measurements of ER Vul in the UBVR system were taken in Mount Maidanak Observatory. The data span 896 days in the three observing seasons. The corresponding TSA of the data has been performed in the four channels separately (see You, 1999,Chapter 12). The phases were computed as $\phi = \mathrm{Frac}\frac{JD_{\mathrm{hel}}-2449148.4233}{0.69810126}$ for all PPDs. ER Vul is a binary system and its photometric period is the orbital period, which contains two minima in its light curve.

where τ is the mesh step, and the k_i are integers. The time series shows an equal spacing of the data points in time, while τ is the smallest spacing between two consecutive points. t_1 is the time of the first observed data point; for simplicity, we can set $t_1 = 0$. From Eq. (2.1), the computed phases are

$$\phi_P(t_i) = \mathrm{Frac}(\frac{k_i\tau}{P}). \tag{2.6}$$

Also, for another trial period P_s fulfilling

$$\frac{1}{P_s} = \frac{1}{P} + \frac{s}{\tau}, \quad s = 1, 2, ... \tag{2.7}$$

we can compute the phases for the same time series by Eq. (2.1):

$$\begin{aligned}
\phi_{P_s}(t_i) &= \mathrm{Frac}(\frac{k_i\tau}{P_s}) \\
&= (\frac{k_i\tau}{P_s}) - \mathrm{Int}(\frac{k_i\tau}{P_s}) \\
&= k_i\tau(\frac{1}{P} + \frac{s}{\tau}) - \mathrm{Int}(k_i\tau(\frac{1}{P} + \frac{s}{\tau})) \\
&= \frac{k_i\tau}{P} + k_i s - \mathrm{Int}(\frac{k_i\tau}{P}) - \mathrm{Int}(k_i s) \quad + (0 \text{ or } 1)^1 \\
&= \frac{k_i\tau}{P} - \mathrm{Int}(\frac{k_i\tau}{P}) \quad + (0 \text{ or } 1) \\
&= \phi_P(t_i).
\end{aligned} \tag{2.8}$$

In Eq. (2.8), we use the fact that $k_i s$ are integers; Int denotes the integer part of the argument. As a result, the PPD with $\phi_{P_s}(t_i)$ instead of $\phi_P(t_i)$ does not change at all. However in astronomical observations, the time points t_i of the observational data are rarely exactly equally spaced. As a result of this, the dispersion around the real phase curve is less than the dispersion around those with the spurious periods P_s.

The same mechanism also produces another set of spurious periods connected to the true period:

$$\frac{1}{P_s} = \frac{1}{P} + \frac{s}{\tau}, \quad s = -1, -2, ..., \tag{2.9}$$

for which the PPD mimics the real one or its mirror image. The mirror imaging property of the two PPDs can be explained as follows: according to Eq. (2.9), if $P_s < 0$, $|P_s|$ is used because a negative value of the period has no physical meaning. According to Eq. (2.1), the change in sign ($\pm$) of

[1] The difference between 0 and 1 is due to the fact, that the sum of the fractional parts of two numbers can be larger than 1. But this does not affect the final result because the phase is the fractional part of the final value.

the period leads to a change in sign (and thus the sequence) of the phases, and leads to the PPD to be a mirror image. This simply arises from the fact that adding 1 in order to obtain again positive phases in the interval $[0, 1)$ leads to a mirror image of the calculated PPD, i.e. $\phi = -0.1, -0.3, ...$ leads to $\phi = 0.9, 0.7, ...$ According to Eq. (2.8), the phases calculated from P_s are the same as those obtained with the real period P. Thus, the PPDs calculated with P and $|P_s|$ become mirror images.

This kind of spurious periods in astronomical observational data is most commonly connected with one sidereal day (0.99717 solar days), because usually the star is consistently observed at the same hour angle(s) every night.

$$\frac{1}{P_s} = \frac{1}{P} + s \cdot 1.00284, \quad s = ..., -2, -1, 1, 2, ... \tag{2.10}$$

and also sometimes with annual (due to visibility of the object) and lunar month cycles (due to the time allocation policy at the observatories). Some other restrictions (such as mechanics for some instruments, especially in space experiments) may cause observations to be made at certain periodic moments or intervals. These time gaps will produce the corresponding series of spurious periods.

The existence of this kind of spurious period series in discrete FT spectra is called aliasing. Fig. 2.2 shows the spurious periods which occur in the power spectrum of discrete FT of the photometric observations in channel V of V815 Her (see You, 1999,Chapter 4). The identification of these spurious periods can be achieved by analyzing the interplay of the actual frequencies present in the data and those in the so-called transformed data window (see Chapter 2, Section 2.3).

2.2 Methods of phase dispersion minimization

In general, when performing TSA on photometric observation data, the information on the shape of the light curve and the period are both unknown. Thus, an ideal method for the search for periodicity should be independent on any assumption about the shape of the PPD, i.e. one should use a non-parametric approach. The phase dispersion minimization (PDM) methods are the most commonly used non-parametric methods nowadays. The idea behind the PDM methods is simple: for the true period a correlation exists between the proximity of the phases of two observations and the proximity of the observed process values. For an arbitrarily chosen period this correlation is much less probable.

In most cases, the observational data set contains information about the

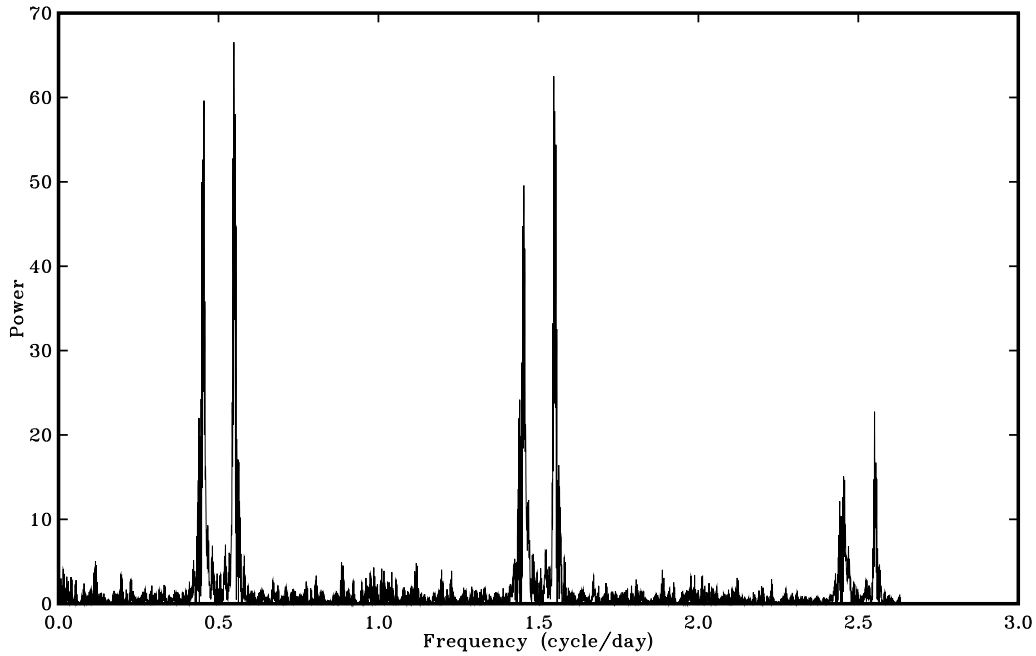

Figure 2.2: Discrete FT power spectrum of photometric observations in channel V of V815 Her. The highest peak in the spectrum is at the frequency of 0.5481 cycles/day, which corresponds to the real period of 1.82454 day. Here we can see that a whole series of spurious peaks is produced due to the daily observational gap of $\tau = 0.99717$ days. All the peaks in the spectrum arise from the convolution of the true periodicity of 1.82454 days with the transformed data window peaking at 1.00284 cycles/day (see Eq. 2.34 in Sect. 2.3). The first peak at a frequency of 0.4547 cycles/day is the mirror image of the peak of the real frequency with $s = -1$ and $\tau = 0.99717$ days in Eq. (2.9). The double peaks around 1.5 cycles/day and around 2.5 cycles/day are spurious: the right peaks are explained by Eq. (2.38) and the left peaks by Eq. (2.39), respectively.

accuracy of each data point, i.e. the observed time series is

$$y(t_i), \quad \sigma(t_i), \quad i = 1, 2, 3, ..., N. \tag{2.11}$$

The $y(t_i)$ are the observed values and the $\sigma(t_i)$ are the errors of the corresponding observations $y(t_i)$. N is the number of the observations. Formally, the following statistic can be used to discriminate between the cases of correct and arbitrarily chosen periods:

$$\begin{aligned} D(P) &= \frac{\sum_{i=1}^{N}\sum_{j=i+1}^{N} g(t_i, t_j, P) w_{i,j} [y(t_i) - y(t_j)]^2}{\sum_{i=1}^{N}\sum_{j=i+1}^{N} g(t_i, t_j, P) w_{i,j}}, \\ w_{i,j} &= \frac{w_i w_j}{w_i + w_j}. \end{aligned} \tag{2.12}$$

The weights w_i are usually calculated as

$$w_i = \frac{1}{\sigma(t_i)^2}. \tag{2.13}$$

$D(P)$ is the weighted sum of the variances[2] of pairs of observations in the PPD corresponding to the period P. The function $g(t_i, t_j, P)$ defines the selection law, which determines whether to take the pair of t_i and t_j into account for computing $D(P)$ or not, based on the trial period P. There have been many different realizations or approximations of $g(t_i, t_j, P)$, which result in different PDM methods (Lafler and Kinman, 1965; Burke et al., 1970; Penfold, 1975; Renson, 1978; Stellingwerf, 1978; Dworetsky, 1983; Cuypers, 1986; Schwarzenberg-Czerny, 1989, 1991, 1997; Davies, 1990).

For the true period, a correlation exists between the proximity of the phases of two observations and the proximity of the observed process values. The selection law $g(t_i, t_j, P)$ is defined to pair two observations for computing $D(P)$ only when

$$t_i - t_j \approx kP, \quad k = 0, \pm 1, \pm 2, ... \tag{2.14}$$

or in terms of the phases

$$\phi_P(t_i) \approx \phi_P(t_j). \tag{2.15}$$

The selection law is thus commonly defined as

$$g(t_i, t_j, P) = \begin{cases} 1, & |\phi_P(t_i) - \phi_P(t_j)| \leq \varepsilon \ \text{ or } \ |\phi_P(t_i) - \phi_P(t_j)| > 1 - \varepsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{2.16}$$

for some small fixed value $\varepsilon > 0$. ε determines the maximum distance in phase by which t_i and t_j can be binned for computing $D(P)$, e.g. $\varepsilon = 0.1$.

[2]Nevertheless, the method is called phase *dispersion* minimization, because a minimum of the variance is also a minimum of the dispersion.

The smaller value for this parameter, the less pairs of (t_i, t_j) are taken into account.

For obtaining a simple statistic for computation, various different techniques (see below) are used to choose $g(t_i, t_j, P)$. These particular realizations or approximations to the general PDM statistic $D(P)$ differ by the arrangement of the computations and by the approximations used, which result in different PDM methods. In most cases, these methods give slightly different results in the search for periods.

For the computation of the general dispersion $D(P)$, a selection window $L(t_i, t_j)$ is generally introduced and $D(P)$ is set to:

$$D(P) = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} g(t_i, t_j, P) L(t_i, t_j) w_{i,j} [y(t_i) - y(t_j)]^2}{\sum_{i=1}^{N} \sum_{j=i+1}^{N} g(t_i, t_j, P) L(t_i, t_j) w_{i,j}} \ , \tag{2.17}$$

where the selection window $L(t_i, t_j)$ is determined for binning the pairs of t_i and t_j according to their distance in time. $L(t_i, t_j)$ is generally computed as:

$$L(t_i, t_j) = \begin{cases} 1, & T_{\min} \leq |t_i - t_j| \leq T_{\max} \ , \\ 0, & \text{other.} \end{cases} \tag{2.18}$$

The degree of smoothing of the spectrum of $D(P)$ obtained by the use of the selection window $L(t_i, t_j)$ depends on $T_{\max}$. When $T_{\max} = t_N - t_1$, then all the observation pairs are taken into account when computing $D(P)$, and maximal resolution is attained. For a smaller value of $T_{\max}$, only a subset of pairs is considered and the computed function of $D(P)$ is smoother. This value of the upper limit $T_{\max}$ must be defined for the computation of $D(P)$. Too large a $T_{\max}$ makes two far separated points t_i and t_j binned into a pair, which could be un-correlated, if e.g. the amplitude of the variation changes over a long time scale or long-term trends are present in the data. Grouping these far separated points not only consumes more computation time, but also may lead to wrong periods, even if the periods are stable. A typical choice of $T_{\max}$ depends on the maximum trial period $P_{\max}$, e.g. $T_{\max} = 10 P_{\max}$.

If the time points t_i and t_j are too close in time, i.e. $|t_i - t_j|$ is less than the minimum trial period $P_{\min}$, the corresponding terms in $D(P)$ do not change significantly for all trial periods so that it does not help to discriminate between them. To avoid such unnecessary computations, the lower limit $T_{\min}$ makes such pairs drop out of the general sum of $D(P)$. A typical choice of $T_{\min}$ depends on the minimum trial period $P_{\min}$, e.g. $T_{\min} = 0.9 P_{\min}$.

Additionally to the minimization of $D(P)$ as given in Eq. (2.17) by using Eqs. (2.16) and (2.18) directly, the following PDM methods are also used:

Stellingwerf method

The Stellingwerf method (Stellingwerf, 1978) uses a particular binning method in phase for two observations t_i and t_j to compute the phase dispersion $D(P)$. The full range [0, 1) of phases is divided into K equal subranges

(bins) I_k

$$I_k = [\frac{k-1}{K}, \frac{k}{K}), \quad k = 1, 2, ..., K. \tag{2.19}$$

The approximation to the normal selection law $g(t_i, t_j, P)$ is to consider the phases to be adjacent when the distance between them does not exceed $\Delta_{\rm bin}+$ 1 times the bin length $\frac{1}{K}$. The bins may overlap and periodic boundary conditions are used at $\phi_P = 1$. When $\Delta_{\rm bin} = 0$, it becomes the well-known statistic proposed by Jurkevich (1971). The role of $\Delta_{\rm bin}$ is more or less the same as that of ε in Eq. (2.16) and $\frac{\Delta_{\rm bin}+1}{2K}$ is comparable to ε. The Stellingwerf method is well documented in the literature. This method is widely used in the astronomical community.

Lafler-Kinman method

The Lafler-Kinman method is one of the oldest and most popular PDM methods in the astronomical community. The original paper by Lafler and Kinman (1965) was one of the first to introduce PDM methods into the practice of astronomical research. The Lafler-Kinman criterion chooses $g(t_i, t_j, P)$ such that $g(t_i, t_j) = 1$ only when $\phi_p(t_i)$ and $\phi_p(t_j)$ are immediate neighbors in the sequence

$$\phi_P(t_{i1}) \leq \phi_P(t_{i2}) \leq ... \leq \phi_P(t_{iN}) \tag{2.20}$$

and otherwise $g(t_i, t_j, P) = 0$. To determine Eq. (2.20), the Lafler-Kinman method formally regards the criterion from Eq. (2.16) as $\varepsilon = \frac{0.5}{N}$, where N is the number of data points in the observations. This small effective value for ε causes the Lafler-Kinman spectrum of $D(P)$ to be usually erratic.

The Lafler-Kinman statistic for data normalized to the overall variance becomes:

$$D(P) = \frac{1}{N-1} \sum_{i=1}^{N-1} [y(t_i^{'}) - y(t_{i+1}^{'})]^2 \ , \tag{2.21}$$

where the observations $y(t_i^{'})$ are sorted in phases

$$\phi_P(t_{i1}^{'}) \leq \phi_P(t_{i2}^{'}) \leq ... \leq \phi_P(t_{iN}^{'}) \ . \tag{2.22}$$

Here, the times $t_i^{'}$ are in the phase sorted sequence.

String-length method

The sum of dispersions $D(P)$ to be minimized is simply the sum of lengths of line segments joining successive points in a PPD:

$$\begin{aligned} D(P) = \sum_{i=1}^{N-1} ([y(t_i^{'}) - y(t_{i+1}^{'})]^2 + [\phi_P(t_i^{'}) - \phi_P(t_{i+1}^{'})]^2)^{1/2} \\ +([y(t_1^{'}) - y(t_N^{'})]^2 + [\phi_P(t_1^{'}) - \phi_P(t_N^{'}) + 1]^2)^{1/2} \end{aligned} \tag{2.23}$$

where the observations are sorted in phases. The string-length method was introduced by Burke et al. (1970) and evaluated by Dworetsky (1983). It is especially suitable for data sets with small number of observations.

2.3 Fourier transform

The discrete Fourier transform (DFT) is a common parametric method used for a fast search of periodicity, even if it has its shortcomings in the case of irregularly sampled data. In DFT, the correlations between the data points are measured based on a series of given basic functions, usually sine and cosine functions. DFT is classically presented by Deeming (1975) and later by Lomb (1976) and Scargle (1982). A recent widely used software package, called Period04, designed for the multi-period analysis of pulsating stars by Lenz and Breger (2005) uses a combination of DFT and Marquardt-modelling of sinusoids. A most recent improved statistical treatment to determine the significance of a peak in a DFT, called SigSpec, is presented by Reegen (2007). His concept of spectral significance can be understood as an improvement of the Lomb-Scargle periodogram.

The full FT of a function $y(t)$ is defined as the complex expression

$$\mathcal{Y}(f) = \int_{-\infty}^{+\infty} y(t) e^{i2\pi ft} dt \ . \tag{2.24}$$

With this definition, the inverse transform[3] is

$$y(t) = \int_{-\infty}^{+\infty} \mathcal{Y}(f) e^{-i2\pi ft} df \ . \tag{2.25}$$

If the function $y(t)$ is a finite data segment, some important conclusions can be deduced like:

1. FT of a finite data segment is different from zero nearly everywhere.
2. A shorter data segment leads to a smoother spectrum $\mathcal{Y}(f)$.
3. A longer data segment can reveal more details in the spectrum $\mathcal{Y}(f)$.

The spectral resolution is bounded by the width of the sinc smoothing kernel, which is defined as

$$\mathrm{sinc}(\pi t) = \frac{\sin(\pi t)}{\pi t} \tag{2.26}$$

and thus the spectral resolution is limited by the length of the data segment.

[3]Here we follow the notation by Deeming; readers may find $e^{-i2\pi ft}$ instead of $e^{i2\pi ft}$ in the integral part of Eq. (2.24) from other literature. In this case the inverse transform will use $e^{i2\pi ft}$ instead of $e^{-i2\pi ft}$ in Eq. (2.25).

In astronomy, an observed time series is not only finite but also discrete, which can be defined as $y(t_i), i = 1, 2, 3, ..., N$. We can specify an arbitrary trial period $P = f^{-1}$, and compute correlations with cosine,

$$C(f) = \sum_{j=1}^{N} y(t_i) \cos(2\pi f t_j) \ , \tag{2.27}$$

and sine,

$$S(f) = \sum_{j=1}^{N} y(t_i) \sin(2\pi f t_j) \ , \tag{2.28}$$

functions. The power periodogram can be obtained by combining these two correlations into one statistic, expressed generally as

$$\Theta(f_l) = \frac{A(f_l)^2 C(f_l)^2 + B(f_l)^2 S(f_l)^2}{D}, \quad l = 1, 2, ..., L, \tag{2.29}$$

where the weighting functions $A(f)$, $B(f)$ and normalization coefficient D are to be specified.

The power $\Theta(f)$ can be computed as the Deeming periodogram (Deeming, 1975) or as the Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982), which is also widely used. The traditional Deeming periodogram is computed with $A = 1$, $B = 1$, and $D = N/2$:

$$\Theta_{\rm Deeming}(f) = \frac{2[C(f)^2 + S(f)^2]}{N} \ . \tag{2.30}$$

The Lomb-Scargle periodogram is computed by setting

$$\begin{aligned} A(f) &= \frac{1}{\{\sum_{j=1}^{N} \cos^2 2\pi f(t_j - \tau(f))\}^2} \ , \\ B(f) &= \frac{1}{\{\sum_{j=1}^{N} \sin^2 2\pi f(t_j - \tau(f))\}^2} \ , \\ D &= 2\sigma^2 \ , \end{aligned} \tag{2.31}$$

where σ is the standard deviation of the original data around the mean. The value of τ is computed from

$$\tan(4\pi f\tau) = \frac{\sum_{j=1}^{N} \sin(4\pi f t_j)}{\sum_{j=1}^{N} \cos(4\pi f t_j)}. \tag{2.32}$$

The main statistical properties of $\Theta(f)$ can be obtained under the assumption, that the observed data points are statistically independent, randomly distributed with equal dispersion and zero mean. From a knowledge of the

distribution of the $\Theta(f)$ values for purely random data, a false alarm probability can be deduced and consequently the significance of the peaks in the $\Theta(f)$ spectrum can be estimated (Scargle, 1982).

The Deeming periodogram is a sampling method, which assumes that the properties of the periodogram do not change much when unequally spaced data are used instead of equally spaced data. The Lomb-Scargle periodogram can be derived from the standard least-squares principle (Lomb, 1976). The actual difference of the two periodograms is rather small. Even for small data sets, the corresponding spectra are rather similar.

The convolution theorem from standard Fourier theory states that, if

$$z(t) = x(t) \cdot y(t) \tag{2.33}$$

then

$$\mathcal{Z}(f) = \mathcal{X}(f) * \mathcal{Y}(f) = \int_{-\infty}^{+\infty} \mathcal{X}(f - f^{'})\mathcal{Y}(f^{'})df^{'} \ , \tag{2.34}$$

where $\mathcal{Z}(f)$, $\mathcal{X}(f)$ and $\mathcal{Y}(f)$ are the corresponding FTs of $z(t)$, $x(t)$ and $y(t)$, and $*$ means convolution as described by the rightmost part of Eq. (2.34). This is true in particular, when $y(t)$ is the continuous process to be observed, and $w(t)$ is the data window:

$$w(t) = \sum_{j=1}^{N} \delta(t - t_j) \quad j = 1, 2, ..., N, \tag{2.35}$$

where $\delta(t - t_j)$ is the Dirac function. The FT data window $W(f)$ is called the transformed data window or sometimes the spectral window. From Eq. (2.34) the final FT is

$$\mathcal{Y}(f) = Y(f) * W(f) \ , \tag{2.36}$$

where

$$W(f) = \sum_{j=1}^{N} e^{i2\pi f t_j} \ . \tag{2.37}$$

In astronomy, the FT is performed on finite and discrete observed data points, which are usually unequally distributed in time. Both the data length and the data spacing have important effects on the accuracy. If the final FT turns out to be concentrated in certain regions of frequency, it suggests certain characteristic time scales present in the data. The power of the final periodogram $\Theta(f)$ is actually given by the convolution of the true physical process with the spectral window (transformed data window). The pathology of the data distribution is all contained in the spectral window, which can be calculated from the data spacing alone, and does not depend directly on the data themselves (see Eq. 2.37 and Deeming 1975). The interference of the spectral window can be described as one of two types: due to the finite

length of the data and due to the data spacing, which is called aliasing. For continuously recorded data, aliasing does not exit, while for equally spaced data, it exists in its most extreme form.

It is usually thought that the largest recoverable frequency (or the smallest recoverable period) in the data depends on the minimum distance between consecutive observations in time. For unequally spaced data, the transformed data window depends on the overall structure of the time spacing, and does not require time points to be arbitrarily close. The possibility of finding a period smaller than the data gaps still exists.

The spectral window is, therefore all-important. In any practical application for astronomical data, the particular data spacing involved and the particular spectral window it gives rise to are known. Typically, the periodogram $\Theta(f)$ shows a series of well defined peaks, which have a central peak at frequency f_0 and other subsidiary peaks at

$$f_k = f_0 \pm k \cdot \Delta f, \quad k = 1, 2, 3, ...\{f_0 \pm k \cdot \Delta f > 0\} \tag{2.38}$$

and mirror imaging peaks at

$$f_k = \mid f_0 - k \cdot \Delta f \mid, \quad k = 1, 2, 3, ...\{f_0 - k \cdot \Delta f < 0\} \ , \tag{2.39}$$

where Δf arises from the peculiarities in the data spacing. As already mentioned, in astronomical observations, such peculiarities often arise due to a one day, one synodic month, one calendar month, or one year gaps in the data. All these facts require us to compare the peaks in the periodogram $\Theta(f)$ with the shape of the spectral window for identification of true periodicities in the observed data.

2.4 Modeling and period refinement

Given an observational time series, for a specific trial period we can compute the parameters for some parametric model and characterize the distribution of the corresponding residuals by some statistic $\Theta(P)$. By computing $\Theta(P)$ for a grid of trials $P_l, l = 1, 2, ..., L$, we obtain a periodogram or spectrum whose minima or maxima (depending on the nature of the statistic $\Theta(P)$) indicate probable values for the real periods. For instance, the modeling spectrum

$$\Theta(P) = \sum_{i=1}^{N} [y(t_i) - \sum_{r=1}^{R} \hat{C}_r(P)\psi_r(ft_i)]^2 \tag{2.40}$$

is a conceptual statistic for periodicity analysis, which can be derived from the maximum likelihood principle. Here, the parametric model $\hat{C}_r(P)\psi_r(ft_i)$ consists of the function $\psi_r(ft_i)$ and the amplitudes $\hat{C}_r(P)$, in which $P = 1/f$ is the trial period.

When the modeling uses trigonometric functions (Vaníček, 1969; Lomb, 1976), the general form of this statistics is

$$\Theta(P) = \sum_{i=1}^{N} \{y(t_i) - a_0 - \sum_{r=1}^{R} [a_r \cos(2\pi r f t) + b_r \sin(2\pi r f t)]\}^2 \ . \qquad (2.41)$$

The particular case of $R = 1$ or 2 is the most popular.

Given an observed time series, the FT and the PDM methods can normally give the real time scales of the periodicity with the help of the corresponding data window, but the precision of the periods obtained from these methods is limited by the frequency step used, and thus, the final optimized values and error estimates for the periods cannot be obtained. After the identification of the periods from the grid searches in equal frequency steps, the final value of the periods is normally obtained from a refinement with modeling of a parametric model.

When $M(t, c_1, c_2, ..., c_k)$ denotes a general model for the data $y(t_i)$ with K parameters to be estimated, the estimates for the model parameters $\hat{c}_1, \hat{c}_2, ..., \hat{c}_k$ are computed as those values for which the weighted residual sum of squares

$$WRSS = \sum_{i=1}^{N} w(t_i)[y(t_i) - M(t_i, c_1, c_2, ..., c_k)]^2 \qquad (2.42)$$

attains a minimum. After the model parameters are estimated, we can compute the residuals of the fit

$$e(t_i) = y(t_i) - M(t_i, \hat{c}_1, \hat{c}_2, ..., \hat{c}_k), i = 1, 2, ..., N \qquad (2.43)$$

and predict process values for an arbitrary time point t:

$$\hat{y}(t) = M(t, \hat{c}_1, \hat{c}_2, ..., \hat{c}_k). \qquad (2.44)$$

According to the standard statistical theory, the estimates and the values of predicted process have certain statistical properties - mean, standard deviation (variance), skewness and kurtosis (see Press et al., 1994,p. 610). We use only the standard deviation for the estimated parameters and derive confidence limits within the normal theory for the predicted process (see Press et al., 1994,p. 689). In practice, very rarely more sophisticated characteristics such as skewness and kurtosis are used.

In general, the model $M(t, c_1, c_2, ..., c_k)$ with parameters to be estimated contains a polynomial function, a trigonometric function and a transient function. In real applications to photometric observation data, the trigonometric functions are most important for iteratively refining the periods (frequencies) of different components.

The polynomial part of a time series can be described by a polynomial model

$$M_{\rm polynomial}(t) = c_m t^{m-1} + c_{m-1} t^{m-2} + ... + c_2 t + c_1. \qquad (2.45)$$

The number of the polynomial parameters is m, and the degree of the corresponding polynomial is $m-1$. When $m = 1$, only the mean C_1 of the original data is described, and for $m = 2$, a linear trend $c_2 t + c_1$ is considered. In practice $m = 1$ is the most common case, while $m \geq 2$ is rarely applied. Some stellar light variations contain a steady long-term increasing or decreasing trend. For the accuracy in seeking short-term periodicities in the data, the long-term trend is usually removed by subtracting the mean in each divided data segment before the application of the modeling. Because we are seeking periodicities hidden in the observations, any higher degree of the polynomial may affect the periods sought with the trigonometric part. If the data length is not sufficient for the time scale of periodicities under seeking, to distinguish the long-term polynomial trends from long-term periodicities has to be resolved based upon the verification of the applied mathematical model with stellar physics.

Trigonometric functions are entered into the model as

$$\mathcal{T}(t, f, R) = \sum_{r=1}^{R} [a_r \cos(2\pi r f t) + b_r \sin(2\pi r f t)] \ . \qquad (2.46)$$

The results give the estimates of the parameters (a_r, b_r) with standard deviations $(\delta a_r, \delta b_r)$. Phases and amplitudes are also computed from a_r and b_r. If the subterm is denoted by

$$\mathcal{T}_r = a_r \cos(2\pi r f t) + b_r \sin(2\pi r f t) \ , \qquad (2.47)$$

then the amplitude is computed as

$$A_r = \sqrt{a_r^2 + b_r^2} \qquad (2.48)$$

and the phase as

$$\phi_r = \arctan(-b_r/a_r) \ , \qquad (2.49)$$

adjusted by $\pm\pi$, if necessary, due to the ambiguity of the arctangent function. The phase ϕ_r is called cos-phase, because it allows a representation in the form

$$\mathcal{T}_r = A_r \cos(2\pi r f t + \phi_r) \ . \qquad (2.50)$$

Introducing the transient function in the modeling is quite rare in the search for periods in astronomical data. But functions can be added to the model to specify the transient part, if required in some special cases, and the corresponding parameters and their standard deviations can also be estimated.

In summary, for modeling the observed data, a time series model can contain all three parts as

$$M(t) = M_{\text{polynomial}}(t) + M_{\text{trigonometric}}(t) + M_{\text{transient}}(t), \tag{2.51}$$

where the most important part, from which the periodicities are sought, is

$$M_{\text{trigonometric}}(t) = \sum_{j=1}^{n} \mathcal{T}(t, f_j, R_j). \tag{2.52}$$

The total number of parameters L in the complete model $M(t)$ is

$$L = m + n + \sum_{j=1}^{n}(2R_j) + E, \quad j = 1, 2, ..., n, \tag{2.53}$$

where $m - 1$ is the degree of the polynomial part, n is the number of model periods included into the trigonometric part, R_j is the number of harmonics in the trigonometric component of the jth period and E is the number of parameters in the function used for the transient part in the model.

The total number of parameters of the model L must be smaller than the number of data points. However, modeling precision depends not only on the data number but also, greatly on the accuracy of the observations: the higher the accuracy, the fewer is the number of the observations required for the same successful modeling. It can occur that too few observations per parameter produces an overfit, manifesting as spurious results.

The parameters in the modeling process are classified as linear and nonlinear. Linear parameters are simply multipliers attached to the expressions, such as c_1, c_2, ..., c_N, ..., in

$$c_1 f_1(t) + c_2 f_2(t) + ... + c_N f_N(t). \tag{2.54}$$

The nonlinear parameters occur such as P in the expressions like

$$y(t, P) = \cos(2\pi t/P) \ . \tag{2.55}$$

Linear modeling uses fixed trial frequencies in equal steps and thus optimizes only the linear parameters. The solution of the linear modeling has been the standard problem in solving the normal equations (see Press et al., 1994,p. 665). But due to the fixed trial frequencies, the final value of the period cannot be optimized and its error estimate cannot be reached. Nonlinear modeling avoids such a shortcoming by varying the period as a nonlinear parameter in a certain range working together with linear modeling for the harmonic amplitudes of the corresponding period.

For multidimensional minimum seeking algorithms, such as the Downhill Simplex Method (see Press et al., 1994,p. 402) and the gradient-search

by Marquardt (see Bevington, 1969,p. 235), the final results often depend greatly on the initial value of the period. On the other hand, the standard one-dimensional minimum seeking algorithm – Brent method (see Press et al., 1994,p. 395) gives a stable solution once the correct bracketing for the minimum is given. The Brent algorithm reaches the convergence to the minimum of a one-dimensional curve by inverse parabolic interpolation.

For most cases the linear modeling can be applied for getting an initial value for the subsequent non-linear refinement. We can nearly always be sure that a local minimum found by the linear modeling is also the global one, which is reliable within its accuracy. But if this is not the case, the final value of the periods may be so sensitive to changes of the initial values that the final value becomes unreliable. At this point, the best choice for optimizing the periods is to compare the initial estimates from various methods (see Sections 2.2 and 2.3).

The main deficiency of the data modeling is the lack of criteria for choosing the model components. Thus, all preliminary information about the physical process must be used in building of the model. The observed data inevitably contain noise. The interplay of different cyclic processes described in the model with a combination of several harmonic components, the data noise, and the data irregular sampling, all distort the modeling procedure. Additionally, any model can only approximate the real process. As a consequence, the parametric approach includes the following unavoidable limitations:

1. Some information is inevitably lost when noise is added to the process values: a precise computation does not always guarantee the recovery of precise physical periods.

2. The model components cannot always be recovered with sufficient precision in a one-by-one manner, because of the interplay between the various components.

3. Every prediction is correct only as much as the underlying assumptions about the model are correct.

4. To be reliable, the periods that we are seeking from the data should be many times shorter than the characteristic length of changes in the polynomial part of the model.

Fortunately, all the above limitations mostly influence the attainable minimum value of $WRSS$ rather than the periods themselves, which are our main interest. A slight change in the model can greatly increase or decrease the value of $WRSS$, but usually does not change much the position of the minimum of $WRSS$, especially in the case of only one period. This is the reason for us to rely on modeling to get the final value of the periods with their error estimates. Also, this is the reason which allows us to seek the periods one by one from a data set containing multiple periods, using subtraction procedures.

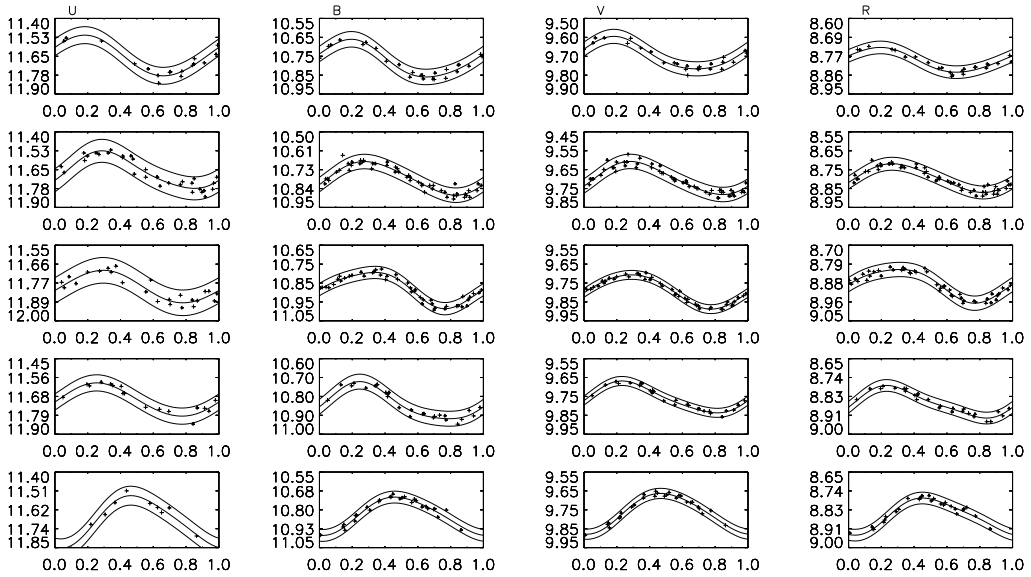

Figure 2.3: Modeling light curves of ET Dra in five seasons in multichannel UBVR photometric observations. From left to right, the modeling curves are in the sequence of U, B, V and R. From top to bottom, they are in the sequence of the seasons 1, 2, 3, 4 and 5. All the curves show well defined sinusoid shapes with clear minima and maxima. The lines represent the LS model for the modeling curves and their 3σ limits. The shapes of the light curves in the same season appear identical in the different channels. But there is a clear phase shift of the light curves in time. The photometric measurements of ET Dra in the UBVR system were carried out at Mount Maidanak Observatory. The data span five seasons and covers a time interval of 1417 days. The TSA application for the data and the derivation of the period ($P = 14.072 \pm 0.016$ days) of ET Dra can be found in (see You, 1999,Chapter 11). The counting of cycles starts at $\mathrm{HJD}_0 = 2448178.1329$, which is the date of the first observation in the data.

In the routine of the refinement of periods we can input multiple periods simultaneously and may take into account transient phenomena for the process. But in many applications, we compute best modeling for only one period with higher reliability. In the modeling procedure, we can always build the model with one single period to describe the initial data. The parameters of the model can be estimated, and the modeling residuals can be computed. Then, the residuals can be considered as an initial data set to repeat the previous step, and so on. For most cases, the models used for refining periods contain only one period with two harmonics $R = 2$ in the trigonometric part, and a $m = 1$ polynomial component as usual. In the period search range, for each trial period all linear parameters are estimated and the corresponding $WRSS$ is computed; then the $WRSS$ is minimized by varying the period. Figure 2.3 shows modeling light curves of multichannel UBVR photometric observations of ET Dra.

It has been argued that if a data set contains two periods too near to each other, the modeling of only one period will greatly affect the detection of the second one due to the incompleteness of the model. A simulation was done to demonstrate the validity of TSA in such kind of situation (see You, 1999,Chapter 6), using artificially generated data. Conclusions can be drawn as following. The periodicities hidden in a process keep their identity so strongly that distorting the process by subtracting other periodicities usually does not affect much their neighbors, even if they are very close to each other. One criterion of how close the two periods can be for the subtraction procedure to stay valid is that the two neighboring periods need to be resolved as two peaks in the power spectrum (the FT spectrum in the case studied in You, 1999). Then, the modeling and the subtracting by the modeling procedure can be performed. Subtracting one period with an incomplete model will inevitably add extra noise to the residuals, and this noise will magnify as the procedure continues for possible further periods. Thus, the one-by-one process does not always recover the multiple periods with sufficient precision, especially if the amplitudes are small. This problem becomes more and more evident as larger noise and more periodicities are added into the real process.

2.5 Multichannel period analysis and review of Papers II and V

Astronomical photometric observations are usually obtained through different wavelength passbands (channels) simultaneously. Traditional time series analysis typically uses only one single channel. For a maximum use of information, one can analyze separate channels and then use a weighted mean of the obtained individual periods as the final period (You, 1999). This ap-

proach may introduce a large error in the final average due to the errors in separate channels. For analyzing all channel data combined together, Jetsu and Pelt (1999) and Lyytinen et al. (2002) first normalized the data of separate channels and applied the TSA afterward. However, such a normalization approach can cause discrepancies which arise from the distortion and the phase lag of light curves.

To keep the shapes of light curves and the phase lags in different channels untouched, a new MPA method was devised in this PhD work, extending the method of three stage analysis (Jetsu and Pelt, 1999) for the simultaneous multi-channel observations. The three stages are the phase dispersion minimization (PDM), the linear modeling (LM) and the nonlinear modeling (NLM) stage. MPA is based on minimizing the weighted sum of dispersions in the first stage of PDM search and the weighted residual sum of squares (WRSS) in the second and the third stages of LN and NLM modeling, over all channels simultaneously. Both the full analytical details and the computational testing results have been presented in Papers II and V. Here, we give a brief review of the articles.

In Paper II, the test of MPA was done by applying the method to two groups of simulated data. Each group contains twelve simulated data sets, which represent typical multi-channel photometric observations. The testing results have been compared with the analyses of the single channels. The comparisons show that MPA has advantages both in searching the real periods and in computing the final values with higher accuracies upon those obtained by analyzing only the separate single channels.

1) The first stage of PDM search is able to detect existing periods in a wide range of time scale for a given data set. In this rough search stage, the identification of spurious periods is done with the data window of the spacing of observation time points. A more general mathematical expression is also derived for the identification of spurious periods originating from the interplay of a correct period with periodicities in the observing times. The PDM can identify and provide initial estimates for all real periods.

2) The second stage of LM is an intermediate stage, which refines the periods from PDM with much smaller equal step size increments of trial frequencies. The modeling in LM stage is linear because the periods are fixed and only the amplitudes of the harmonics are modeled. It is solved by the conventional linear least square (LS) solution. The purpose of the LM is to provide best initial values and correct bracketing (the search range for periods) for the final optimization in the NLM stage.

3) The third NLM stage does not use multidimensional nonlinear minimization directly, because of the above discussed pitfalls of such an approach. Instead, each minimization step is splitted into a linear multidimensional step (for model amplitudes) and a nonlinear one-dimensional step (for the period). In practice, the final period is found by a one-dimensional Brent minimiza-

tion, performed using a subroutine that for each fixed period returns the best fitting harmonic model amplitudes obtained by LS solution. The procedure continues until the period yielding the minimum of WRSS is found. Brent's algorithm also gives the error estimate for the period computed from the WRSS hypersurface.

Paper II shows that the three stage MPA performs successfully for simultaneous multi-channel photometric observations. The advantage of MPA over traditional TSA can be summarized as following:

1. Because the multichannel analysis method only makes summing of $D(P)$ of Eq. (2.12) or $WRSS$ of Eq. (2.42) over individual channels, this allows the physical light curves in different channels to vary, as long as the periods are identical in all channels. Thus MPA has a wider field of application than the normalization approach which makes the implicit assumption that the light curves are identical in different channels except for their amplitude scales. Also, there is no need for separate processing of raw data before the application of MPA.

2. The application of MPA on the simulated data demonstrates that it gives more accurate results than the analysis of the single channel data, even when compared to that individual channel which has the best signal to noise ratio. For most cases, MPA is efficient and accurate in detecting periods in the multichannel data, especially for noisy data, for which the traditional approach of analyzing separate channels may lead to erroneous results.

   As the noise level increases in the data, the simulated period may not be recovered from the analyses of channels separately. On the other hand, combining all channels into our multichannel PDM search, the real period can be recovered as the deepest minimum in the $D(P)$ periodogram. The reason here is the fact that $D(P)$ is a simple summing of $D^{\mathrm{c}}(P)$ of all channels. For the large noise cases in $D^{\mathrm{c}}(P)$ periodogram, the large noise combined with the interference between the two selection criteria $L(t_{ij})$, $g(t_{ij}, P)$ (see Paper II for explanations) and the data gaps in time, can cause deep spurious minima. Because these spurious minima usually occur at different trial frequencies in different channels, they are suppressed when summing into $D(P)$, while the minima of the real periods accumulate into an even deeper minimum.

3. If some channels are totally corrupted and contain only noise, the application of MPA should omit these channels. Thus for such kind of data, it is still necessary to analyze all channels separately in order to make an initial identification of the real periods. The subsequent application of MPA then gives the final optimized value of the periods with reasonable error estimations.

4. The recovered period values and their error estimates rely on two aspects of the original observational data: the observation accuracies and the number of observations. In particular, increased number of data points in MPA means usually improved accuracy. Nevertheless, the former aspect needs also attention: correct error estimation depends crucially on knowing the true errors of the original data points.

5. The basic structure of the three stage method (PDM, LM and NLM) is very logical. For a given data set, we may not know where the true periods are located. Thus, the role of PDM is to detect existing periods in a wide range of time scales with the help of the data window to identify the spurious periods. LM refines these periods with much denser trial frequencies and provides best initial values and correct bracketing for NLM. The final optimized value of the periods and all other modeling parameters with their error estimations are given by NLM. Note that for some special cases not all three stages are needed: the LM and/or NLM stages can be used to refine the period and compute a model also in a case where the data have been analyzed earlier (for example this has been done in paper IV).

Paper V tackles another problem in the modeling procedure when multiple periodicities are encountered in time series analysis for the photometric observations. For example, multiple periodicities are encountered in both Papers III and IV and the final optimization of the periods then depends on the modeling the lightcurves. The paper V tries to compare the accuracy between complete modeling and componentwise modeling for the multiple periods. The complete modeling is performed by building a full model including all detected periods and varying them simultaneously in a multi-dimensional period space. The componentwise modeling fits first a model containing only one detected period and interprets the effects from all other periods as random noise. Both mathematical formalism and statistical simulations for the two modeling types are presented in the paper. The work in Paper V supports the subtracting and optimizing technique of recovering multiple periods, as is done in the MPA of the photometric observations in Papers III and IV.

Chapter 3

Applications of multichannel period analysis to photometric observations of brightness variations of stars

3.1 Solar and stellar differential rotation

Differential rotation means that different parts of a rotating object move with different angular velocities (rates of rotation). The phenomenon usually occurs when the rotating object is not solid. Most astrophysical objects are not solid (such as stars, accretion disks and galaxies etc.) and thus usually show differential rotation. The examples in our solar system include the Sun, Jupiter and Saturn. The Sun and stars are giant balls of gaseous plasma and their different parts spin at different rates.

Mathematically, solar and stellar rotation can be described with the differential rotation rate. In the case of the Sun, the rate of rotation is fastest at the equator (latitude $\theta = 0°$) and decreases as latitude increases. The rotation rate is described by the equation:

$$\Omega(\theta) = \Omega_{eq} - \Delta\Omega \sin^2 \theta \tag{3.1}$$

where $\Omega(\theta)$ is the angular velocity in degrees per day at latitude θ, Ω_{eq} is the equatorial angular velocity and $\Delta\Omega = \Omega_{eq} - \Omega_{pole}$ is the difference between equator and pole. Both Ω_{eq} and $\Delta\Omega$ are constants such as for the Sun:

$$\Omega_{eq} \approx 14.18° \text{ /day} , \quad \Delta\Omega \approx 3.15° \text{ /day} .$$

At the equator the solar rotation period is 25.38 days, which is called the sidereal rotation period.

For the convenience to describe the degree of star's differential rotation, we define a differential rotation parameter α as

$$\alpha = \Delta\Omega/\Omega_{eq}. \tag{3.2}$$

For the Sun $\alpha \approx 0.2$ and the lap time is 115 days, which is the time the equator needs to overtake the poles by one full rotation. The above equations have been commonly used for characterizing the observations of stellar differential rotation (Donati and Collier Cameron, 1997; Kovári et al., 2004).

For the Sun, the rotation constants have been measured by following the motion of various features (tracers) on the solar surface (St. John, 1918). The first and most widely used tracer are sunspots. At the start of a new sunspots cycle, the sunspot groups form at middle latitudes. As the sunspot activity rises and then declines, the sunspot groups subsequently form at latitudes closer to the equator. Because the equator is rotating faster than the poles, the mean rotation period marked by sunspots progresses from longer to shorter periods. When a large sunspot group crosses the visible hemisphere of the Sun, the observed variation of the solar integrated flux is less than 1%. Comparing with up to 30% variations for the RS CVn and BY Dra stars, the starspots may be very different from sunspots. The sensitivity of the ground-based stellar photometry used is not sufficient to detect solar-type spots on stars. Thus, the properties of the starspots detected by photometric methods differ from those of sunspots. Only modern space telescopes can observe solar and stellar low-amplitude light variations: it has only recently become possible to track the solar differential rotation by the modulation in the flux caused by sunspots (Hempelmann, 2003).

3.2 Stellar magnetic fields and dynamos

Magnetism pervades the Universe. Planets, stars and entire galaxies all have associated magnetic fields. Magnetic fields are generated by the motion of electrically conducting fluids, which is the so-called dynamo effect. Stellar magnetic fields are thought to be generated by magnetohydrodynamic dynamos. The precise details of what drives the motion and what the fluid consists of, have not been fully understood. Magnetic fields are also crucial in the formation of stars.

The basic theory of stellar dynamo magnetic field generation was initially developed by Parker (1955) and further elaborated by Steenbeck et al. (1966) and Parker (1971). Recent progress in the theory of solar and stellar dynamos has been reviewed by Brandenburg and Dobler (2002) and Rüdiger and Hollerbach (2004). Particular emphasis is placed on the mean-field theory which tries to describe the collective behavior of the magnetic field. The dynamo theory can describe the global geometry and behavior of solar and stellar magnetic fields by the stability of dynamo modes with different

symmetries (Brandenburg et al., 1989). For instance, the 11-year sun spot cycle agrees with an axisymmetric mean-field dynamo dipole-like mode. Similarly, starspot cycles may be associate with an axisymmetric quadrupole-like mode, in which case the symmetry is with respect to the equator.

Solar and starspots are closely associated with their dynamo magnetic fields. High-latitude spots on very active stars can be explained by non-linear models for flux-tube instability (Schuessler et al., 1996). In a rapid rotator with a dynamo operating near the base of the convection zone, the effect of the Coriolis force makes flux tubes to emerge nearly parallel to the rotation axis. This produces high-latitude activity even if the dynamo amplifies the field only at low latitudes. Such a model is capable to reproduce the sunspot behavior as well (Caligari et al., 1995). In addition to rapid rotation, the depth of the convection zone, stratification and magnetic field strength play an equally important role in the poleward deflection of rising flux-tubes. In particular, the size of the stellar core affects the magnetic curvature force and the rise time, while gravity determines the strength of the buoyancy force. Cool stars with external convection envelopes exhibit various magnetic phenomena similar to those observed on the Sun.

Starspots originate from stellar dynamo magnetic fields and thus provide the best window from which we can see through stellar magnetism and understand the stellar dynamos. On the other hand, the Sun is a slow rotator with low activity. Studying stars rotating as slow as the Sun help us to know what to expect from the Sun in the future. A butterfly diagram is a combination of two effects: stellar surface differential rotation and systematic latitudinal shift of star spot regions over an activity cycle. Monitoring over the period of at least one activity cycle can yield a stellar butterfly diagram. The groundbased automatic photometric telescopes (Henry, 1999) are precise enough for observations of the butterfly diagrams (Strassmeier et al., 1997). A remarkable progress in observational facilities and numerical techniques for studying starspots has been achieved in the last two decades. High precision stellar photometry gives us a deep insight into the nature of starspots and underlying magnetic fields (Walker et al., 2007; Croll et al., 2006). Differential rotation is a key parameter for stellar dynamos. A theoretical study of its dependence on the rotation rate and spectral type is important for interpreting observations. A noticeable progress in such modeling was achieved by Kitchatinov and Rüdiger (1999).

3.3 Differential rotation and multichannel period analysis

Differential rotation is a fundamental ingredient of hydromagnetic dynamo theory. Measurements of the solar differential rotation set an observable

boundary condition to the amount of the internal differential rotation. In the hydromagnetic dynamo theory, the interaction of internal differential rotation, convection, and turbulence generates and sustains the magnetic fields. However, no current dynamo model can fully recreate the spatial and temporal properties of the Solar magnetic field. The study of stellar differential rotation may help to solve the problems and improve our understanding of the Sun and the stars. Precise measurements from the observations of stellar differential surface rotation are important for understanding the connections between stellar dynamo theory, differential rotation, and stellar surface spots. The MPA may provide the answers using observations of the stellar photometric variability.

It is not yet possible to directly observe stellar surface differential rotation from the migration of starspots. The study of the stellar surface differential rotation has so far been based on two basic approaches: the changes in Doppler-imaging maps (Donati and Collier Cameron, 1997; Korhonen et al., 2000) and the changes in the rotation period as modulated by starspots in the lightcurves (Reglero et al., 1986). For stars with significant spot coverage, the presence of differential rotation can be well deduced with the second approach. This approach has been based on broadband photometric measurements (Hall, 1991) and on time variations in Ca II H&K emission-line fluxes (Vaughan et al., 1981; Baliunas et al., 1983, 1985; Donahue et al., 1996).

Various numerical techniques have been applied to study the properties of starspots. These methods include light-curve modeling and inversions (surface imaging), Doppler and Zeeman-Doppler imaging, molecular line diagnostics, asteroseismology, etc. Detailed reconstructions of starspot sizes, shapes, and positions have been attempted by the 'surface imaging' technique. It shows that the shape and amplitude of both the brightness and color variations yield typical 'effective' spot temperatures near 3400 ± 200K, generally 1000 - 1200K cooler than the unspotted regions. This indicates that the spotted areas are huge compared to those of sunspots. Starspots reconstructed so far tend to show an extensive coverage (up to 40% of the stellar disk), and the presence of polar spots is frequently inferred. Such large cool spots when passing over stellar disks cause large variations in brightness (up to 0.6 mag) and strong line profile distortions in the spectra of rapidly rotating stars. Analysis of time series of spectral line profile variations using Doppler imaging techniques provides the spot distribution over the visible stellar surface. Doppler images reveal that spots on cool rapidly rotating stars are preferably formed at higher latitudes, from 30° up to the visible pole, in contrast to sunspots which on average appear at latitudes below 30°. Magnetic field measurements suggest that the large starspots represent active regions consisting of smaller, mixed polarity spots.

In RS CVn and BY Dra variables, differential surface rotation is a consis-

tent explanation of 'wave migration' or changing periods observed in the light curves (Vogt, 1981; Dorren and Guinan, 1984). Surface differential rotation has been attributed as being the cause of changing or dual periodicities in the nearby young solar analogue HN Peg in Paper III and also in chromospheric records of some sample of stars in Wilson (1978). These stars have been monitored on a nightly basis since 1980 (Baliunas et al., 1985). The rotation of the Sun differs by 3% within the restricted latitude zones of sunspots (Labonte, 1984). However, because the latitudes of the star rotation tracers are undetermined from the chromospheric records alone (Gilliland, 1984; Baliunas et al., 1985), it is difficult to compare the solar and stellar results. For spotted stars, accurate determinations of the differential rotation periods (usually short, a few days) are possible. Early estimates of differential rotation where obtained by Vogt (1983).

Two types of behavior in the time series of chromospheric records imply surface differential rotation. Firstly, the periods significantly vary from one season to the next. Active areas presumably form at a different latitudes and hence show a different period, when one season is compared with another. Secondly, two or more periods can be present simultaneously in a given season, which suggests two or more active areas rotating differentially with respect to each other, probably because they formed at different latitudes within the season. The growth and decay of active areas forming at different longitudes can mimic the signal of the multiple periods in the seasonal time series (see Paper III for HN Peg and Paper IV for LQ Hydrae).

The stellar photometry is commonly done by collecting stellar photons through different filters (passbands) at the same time. For this kind of multichannel data, traditional TSA methods analyze first the separate bands and then combine these analyses to get a final result, but this introduces arbitrary errors. In another approach all passband data are normalized to allow a simultaneous analysis, but it can cause discrepancies due to distortion of the light curves. In order to maximally use the information and to reach higher precision, we devised in this PhD work the multichannel period analysis (MPA) package for the simultaneous analysis of multichannel data (see section 2.5 and Paper II). The aim of MPA is to reach higher accuracy, not only for the primary period (modulation of the primary starspot region), but also for the additional secondary periods (modulation of the secondary spotted regions). The more precise measurements of stellar rotation periods give more precise estimates of differential rotation rates and thus give us more solid foundation for interpreting stellar magnetic fields and hydromagnetic dynamo theory.

3.4 Summary of Papers I, III and IV

The MPA application to real photometric observations in this PhD work was conducted in Paper III for HN Peg and in Paper IV for LQ Hydrae, while TSA for V815 Herculis in Paper I still uses the earlier normalization method by Jetsu and Pelt (1999). Here, we briefly review the work reported in Papers III and IV and make conclusions about the usefulness of MPA.

HN Peg is a nearby young solar analogue exhibiting variations in the Ca II H&K emission flux. We chose it as a useful candidate for testing the MPA method, because of the ambiguous period variations reported by previous authors, likely to indicate the presence of multiple periods (see the introduction part of Paper III).

Paper III is the first application of MPA to real observations. All the three stages (PDM, LM and NLM) have been fully performed for HN Peg. The first stage of PDM search shows its ability to identify multiple periods by analyzing the residuals after subtracting the previously known periods. The following stages of LM and NLM can constrain the periods to a satisfactory degree through a one-dimensional refinement. The final period refinement optimizes only one period in the original data and interprets the effects from other periods as random noise.

Paper IV for LQ Hydrae is the second application of MPA to real observations. Because of the closeness of multiple periods of LQ Hydrae, the original data cannot be always used for optimizing the periods, except for the first significant period which corresponds to the largest amplitude. This makes the main difference from the case of HN Peg. For the case of LQ Hydrae, the direct use of LM spectra simplifies the computational analysis and shows the method's powerful ability to identify the multiple periods by subtracting the most significant minimum in the multichannel spectrum. Also, the NLM constrains the periods with higher accuracy than what could be obtained from a single channel.

In Paper III, the MPA detects two photometric periods of HN Peg, with the longer one almost constant near 5.1 days and the shorter one fluctuating within 4.4–4.8 days. The spot modulation by the differential rotation of HN Peg may consist of two distinct latitude zones of spot activity corresponding to the two periods. The spots corresponding to 5.1 days would be confined to within a very narrow range of latitudes, and the other spot latitude range corresponding to 4.4–4.8 days would be broader. The suggested existence of two latitude zones of activity in this young solar analogue matches closely what is seen in the Sun.

In Paper IV on LQ Hydrae, the MPA detects a primary period of 1.60–1.61 days, existing in all observing seasons, while additional secondary periods during different seasons expand the entire period range to be three times wider (1.585–1.625 days). This may tell us that the primary long-lived

spots on LQ Hydrae are constrained between two latitudes corresponding to two different rotation periods of 1.60 and 1.61 days, while some new surface structures temporary exist at different latitudes corresponding to the larger period range. It is important that these additional spot structures are also detectable with our MPA measurements.

In summary, the applications of the MPA to the two stars demonstrate that the method allows us to reach high precision in the analysis of multichannel photometric observations. Papers III and IV are fundamental for further application of MPA to other stellar photometric time series. It is very interesting and important to see how the measured periods change and further investigate their cyclicity. But for further investigation of the starspot cyclicities, longer datasets are required (usually from a few years to several decades). Because the photometric data lengths used in the two papers are not yet long enough to allow us to make firm conclusions, we haven't touched these long-term cyclicities.

3.5 Discussion and future prospects

A key mystery in stellar magnetohydrodynamics is the cause of stellar differential rotation. The question whether TSA for stellar photometry is an appropriate tool to detect stellar differential rotation must be clarified for both the future stellar photometry and the TSA techniques. In this section, we attempt to give a brief answer to that, not only for rapidly rotating stars with high activity levels, but also for slowly rotating stars with low activity.

Rapidly rotating stars are typically covered by spots, often located near the rotation poles (Strassmeier and Linsky, 1996). These spots modulate the stellar broad-band photometry. The modulations are easily detected in the photometric variations. This allows our TSA analysis to measure the photometric period and the differential rotation. The photometric TSA methods are mostly applied to rapidly rotating active stars such as in the case of HN Peg and LQ Hydrae. When extended cool spots dominate the light curve and cause observable amplitudes to be stable over many stellar rotations, photometric TSA is a powerful tool to analyze differential rotation of the active star.

The amplitudes of photometric variability of low-activity stars like the Sun cannot be detected by groundbased telescopes with typical detection limit of one millimag. The photometric TSA becomes difficult to apply to these slow rotators concerning activity and differential rotation. Knowledge of the activity and differential rotation of these stars has depended on CaII H+K observations (Baliunas et al., 1985, 1995; Donahue et al., 1996). The question is still open whether the future photometric precision allows TSA to be successfully applied to the stars with solar-like activity levels as what has

been done for HN Peg and LQ Hydrae in this PhD work. At present, the only practical method for detecting the presence of surface differential rotation for slowly-rotating stars relies on the indirect measurement of the stellar butterfly diagram such as Doppler imaging (Jeffers et al., 2007; Marsden et al., 2005).

The photometric period analysis strongly depends on the data length applied. The change in mean latitude of the star surface inhomogeneities modulates its light variations and allows us to apply MPA to determine the rotation period change correlated with the stellar long-term activity cycle. However, searching for the correct period is hampered by the evolution of star's active regions which introduces variability in the amplitude and phase of the modulation. Detecting this evolution becomes the most difficult and unavoidable task in the photometric MPA. The most important part of the photometric MPA concerns the division of the entire data into reasonable segments before its application. Too small sub-divisions give rise to an insufficient number of data points for MPA. It occurs usually in this kind of situation that MPA recovers only the primary period (or not any periods at all) from the sub-divisions, while their error estimates are so large that they appear to agree with the mean value within their errors. The information obtainable from the original data is thus not fully recovered. On the other hand, accumulating more data in a longer time span may also prevent meaningful periods from being derived. This is because in the presence of multiple periods, the variations of the periods, the phase shifts, and the long-term variation of the mean brightness of the star, all tangle together intrinsically. For a maximum use of information and the correct MPA application, we have to know how to balance all these factors: we need to consider the stability of star's cool spots and bright faculae, as well as the rapid evolution of spots on time-scales comparable to the star's rotation period.

Mathematically, MPA uses the subtraction procedure to pick up one-by-one the multiple periods (for HN Peg and LQ Hydrae in this PhD work). Questions arise if the first significant period is precise enough, and if not, how accurate is the second one, the third one, etc. It is clear that subtracting one period with an incomplete model will inevitably add extra noise to the residuals, and the noise will increase as the procedure continues. Thus, the periods may not be recovered with sufficient precision in the one-by-one manner, especially those with small amplitudes. This will become more evident as larger noise and more periodicities are added into the real process. It is obvious that not all periods can be recovered due to the insufficient data sampling and the observational errors. The best we can do is to be cautious when performing the subtraction procedure during MPA, and to use all existing knowledge about the star's physics. The PPDs provide the best check for a correct period. When folding the data points with the period found, distinguishing the pattern of data point distribution in the light curve, whether

randomness or noisiness, gives us the degree of confidence of the period. The photometric MPA of HN Peg and LQ Hydrae show its ability in analyzing the patterns of the stellar surface differential rotation when viewed through modulated star spots. The MPA technique demonstrates itself as a suitable method for studying stellar differential rotation and monitoring the stellar butterfly diagram with satisfactory accuracy.

Starspot region lifetimes may exceed several years. The degree of spottedness varies with time, which shows up as slow changes in the mean luminosity on time scales of years. It has also been found that some of these changes are cyclic, which indicate the presence of cycles in both the spot formation rate and the stellar magnetic activities, which can be predicted with the dynamo modes. Concerning this long-term spot activity cycle (such as 11-year of the sunspots), the data length in respect to the the cycle length must be considered. For homogeneously sampled data with daily gaps and seasonal gaps typical to astronomical photometry, it is commonly required that the data length is at least ten-times than the period to be detected, to ensure a correct TSA performance. Also, a sufficient amount of data should be provided to obtain a clear PPD plot. In the case of LQ Hydrae in Paper IV, the longest data segment for MPA has a time span of 237 days but contains only 181 observations in V channel. This is due to the daily observational gap, weather and instrumental problems. This time span is about 148 rotations of LQ Hydrae with the period of 1.6 days. Combining all seasonal observations (for about half a year) of LQ Hydrae for MPA makes the assumption that the modulation of the differential rotation is not yet too much affected by the evolution of star's active region, such as causing variability in the mean magnitude, amplitude and phase.

Very high precision data may give us a clear period even when the data length is only two or three times the period. But for a star with long-term activity cycles this may not be enough. Even if the photometric data have good precision and sampling coverage in time, at least three cycle lengths should be covered. This is the reason why our MPA technique has not been applied to study the stellar long-term activity cycles. Clearly, this will become one of the future tasks of our photometric MPA. Space automatic photometric telescopes will accumulate stellar photometric data with unprecedented high precision. The MPA method in principle can be applied to the data obtained by the space projects such as MOST[1] (Microvariablity and Oscillations of Stars, see Walker et al. 2003) and COROT[2] (COnvection, ROtation and planetary Transits, see Weiss et al. 2004).

[1]MOST is a scientific satellite of the Canadian Space Agency and designed to perform high-precision stellar photometry from space.

[2]COROT is a space mission led by the French Space Agency with the collaboration of many European countries and Brazil. Its main scientific goals are the study of stellar variations and the detection of extrasolar planets.

Bibliography

Baliunas, S. L., Donahue, R. A., Soon, W. H., et al. (1995). Chromospheric variations in main-sequence stars. *Astrophysical Journal*, 438:269–287.

Baliunas, S. L., Hartmann, L., Noyes, R. W., et al. (1983). Stellar rotation in lower main-sequence stars measured from time variations in H and K emission-line fluxes. II - Detailed analysis of the 1980 observing season data. *Astrophysical Journal*, 275:752–772.

Baliunas, S. L., Horne, J. H., Porter, A., et al. (1985). Time-series measurements of chromospheric CA II H and K emission in cool stars and the search for differential rotation. *Astrophysical Journal*, 294:310–325.

Bevington, P. R. (1969). *Data reduction and error analysis for the physical sciences*. New York: McGraw-Hill.

Brandenburg, A. and Dobler, W. (2002). Solar and stellar dynamos - latest developments. *Astronomische Nachrichten*, 323:411–416.

Brandenburg, A., Krause, F., Meinel, R., Moss, D., and Tuominen, I. (1989). The stability of nonlinear dynamos and the limited role of kinematic growth rates. *Astronomy and Astrophysics*, 213:411–422.

Burke, Jr., E. W., Rolland, W. W., and Boy, W. R. (1970). A Photoelectric Study of Magnetic Variable Stars. *Journal of the Royal Astronomical Society of Canada*, 64:353–369.

Caligari, P., Moreno-Insertis, F., and Schussler, M. (1995). Emerging flux tubes in the solar convection zone. 1: Asymmetry, tilt, and emergence latitude. *Astrophysical Journal*, 441:886–902.

Cincotta, P. M., Helmi, A., Mendez, M., Nunez, J. A., and Vucetich, H. (1999). Astronomical time-series analysis – II. A search for periodicity using the Shannon entropy. *MNRAS*, 302:582–586.

Cincotta, P. M., Mendez, M., and Nunez, J. A. (1995). Astronomical Time Series Analysis. I. A Search for Periodicity Using Information Entropy. *Astrophysical Journal*, 449:231–235.

Croll, B., Walker, G. A. H., Kuschnig, R., et al. (2006). Differential Rotation of ϵ Eridani Detected by MOST. *The Astrophysical Journal*, 648:607–613.

Cuypers, J. (1986). A new method of detecting frequency variations in variable stars applied to the Beta Cephei star Delta Ceti. *Astronomy and Astrophysics*, 167:282–286.

Cuypers, J. (1987). New observations and frequency analysis of the Beta Cephei star Tau1 LUPI. *Astronomy and Astrophysics Supplement Series*, 69:445–449.

Davies, S. R. (1990). An improved test for periodicity. *MNRAS*, 244:93–95.

Deeming, T. J. (1975). Fourier Analysis with Unequally-Spaced Data. *Astrophysics and Space Science*, 36:137–158.

Donahue, R. A., Saar, S. H., and Baliunas, S. L. (1996). A Relationship between Mean Rotation Period in Lower Main-Sequence Stars and Its Observed Range. *Astrophysical Journal*, 466:384–391.

Donati, J.-F. and Collier Cameron, A. (1997). Differential rotation and magnetic polarity patterns on AB Doradus. *MNRAS*, 291:1–19.

Dorren, J. D. and Guinan, E. F. (1984). Starspots, Differential Rotation and a Possible Six-Year Spot Cycle on Lambda Andromedae. In Baliunas, S. L. and Hartmann, L., editors, *Cool Stars, Stellar Systems, and the Sun*, volume 193 of *Lecture Notes in Physics, Berlin Springer Verlag*, pp. 259–263.

Dworetsky, M. M. (1983). A period-finding method for sparse randomly spaced observations of 'How long is a piece of string?'. *MNRAS*, 203:917–924.

Gilliland, R. L. (1984). Deterministic Time Series Modelling of Ca II H and K Stellar Index Data. In Baliunas, S. L. and Hartmann, L., editors, *Cool Stars, Stellar Systems, and the Sun*, vol. 193, *Lecture Notes in Physics*, Berlin, Springer Verlag, pp. 146–149.

Hall, D. S. (1991). Learning about stellar dynamos from long-term photometry of starspots. In Tuominen I., Moss D., Rudiger G., editors, IAU Colloq. 130, *The Sun and Cool Stars: activity, magnetism, dynamos*, pp. 353–369.

Hempelmann, A. (2003). Wavelet analysis of stellar differential rotation. III. The Sun in white light. *Astronomy and Astrophysics*, 399:717–721.

Henry, G. W. (1999). Techniques for Automated High-Precision Photometry of Sun-like Stars. *PASP*, 111:845–860.

Jeffers, S. V., Donati, J.-F., and Collier Cameron, A. (2007). Magnetic activity on AB Doradus: temporal evolution of star-spots and differential rotation from 1988 to 1994. *MNRAS*, 375:567–583.

Jetsu, L. and Pelt, J. (1999). Three stage period analysis and complementary methods. *Astronomy and Astrophysics Supplement Series*, 139:629–643.

Jurkevich, I. (1971). A Method of Computing Periods of Cyclic Phenomena. *Astrophysics and Space Science*, 13:154–167.

Kitchatinov, L. L. and Rüdiger, G. (1999). Differential rotation models for late-type dwarfs and giants. *Astronomy and Astrophysics*, 344:911–917.

Korhonen, H., Berdyugina, S. V., Hackman, T., Strassmeier, K. G., and Tuominen, I. (2000). Study of FK Comae Berenices — II. Spot evolution from 1994 to 1997. *Astronomy and Astrophysics*, 360:1067–1076.

Kovári, Z., Strassmeier, K. G., Granzer, T., Weber, M., Oláh, K., and Rice, J. B. (2004). Doppler imaging of stellar surface structure. XXII. Time-series mapping of the young rapid rotator LQ Hydrae. *Astronomy and Astrophysics*, 417:1047–1054.

Labonte, B. J. (1984). Is stellar differential rotation observable? *Astrophysical Journal*, 276:335–340.

Lafler, J. and Kinman, T. D. (1965). An RR Lyrae Star Survey with Ihe Lick 20-INCH Astrograph II. The Calculation of RR Lyrae Periods by Electronic Computer. *Astrophysical Journal Supplement*, 11:216–222.

Lenz, P. and Breger, M. (2005). Period04 User Guide. *Communications in Asteroseismology*, 146:53–136.

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:447–462.

Lyytinen, J., Johansson, P., Jetsu, L., et al. (2002). Time series analysis of V511 Lyrae photometry. *Astronomy and Astrophysics*, 383:197–201.

Marraco, H. G. and Muzzio, J. C. (1980). An improved method to derive periods of cyclic phenomena. *PASP*, 92:700–701.

Marsden, S. C., Waite, I. A., Carter, B. D., and Donati, J.-F. (2005). Doppler imaging and surface differential rotation of young open cluster stars - I. HD 307938 (R58) in IC 2602. *MNRAS*, 359:711–724.

Parker, E. N. (1955). Hydromagnetic Dynamo Models. *Astrophysical Journal*, 122:293–314.

Parker, E. N. (1971). The Generation of Magnetic Fields in Astrophysical Bodies. VII. Dynamical Considerations. *Astrophysical Journal*, 168:239–249.

Penfold, J. E. (1975). Two-colour observations of RR Lyrae variables. *The Observatory*, 95:44–50.

Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1994). *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge University Press, 2nd ed.

Reegen, P. (2007). SigSpec. I. Frequency- and phase-resolved significance in Fourier space. *Astronomy and Astrophysics*, 467:1353–1371.

Reglero, V., Fabregat, J., and de Castro, A. (1986). Stromgren uvby-Hbeta Photometry of the Solar Type Star HD 206860. *Informational Bulletin on Variable Stars*, 2904:1–4.

Renson, P. (1978). A method for finding the periods of variable stars. *Astronomy and Astrophysics*, 63:125–129.

Rüdiger, G. and Hollerbach, R. (2004). *The magnetic universe: geophysical and astrophysical dynamo theory*. Wiley-VCH.

Scargle, J. D. (1982). Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *Astrophysical Journal*, 263:835–853.

Scargle, J. D. (1989). Studies in astronomical time series analysis. III - Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data. *Astrophysical Journal*, 343:874–887.

Schuessler, M., Caligari, P., Ferriz-Mas, A., Solanki, S. K., and Stix, M. (1996). Distribution of starspots on cool stars. I. Young and main sequence stars of one solar mass. *Astronomy and Astrophysics*, 314:503–512.

Schwarzenberg-Czerny, A. (1989). On the advantage of using analysis of variance for period search. *MNRAS*, 241:153–165.

Schwarzenberg-Czerny, A. (1991). Accuracy of period determination. *MNRAS*, 253:198–206.

Schwarzenberg-Czerny, A. (1997). The Correct Probability Distribution for the Phase Dispersion Minimization Periodogram. *Astrophysical Journal*, 489:941–945.

St. John, C. E. (1918). The Present Condition of the Problem of Solar Rotation. *PASP*, 30:319–325.

Steenbeck, M., Krause, F., and Rädler, K. H. (1966). A Calculation of the Mean Electromotive Force in an Electrically Conducting Fluid in Turbulent Motion Under the Influence of Coriolis Forces. *Zeitschrift Naturforschung Teil A*, 21:369–376.

Stellingwerf, R. F. (1978). Period determination using phase dispersion minimization. *Astrophysical Journal*, 224:953–960.

Strassmeier, K. G., Bartus, J., Cutispoto, G., and Rodono, M. (1997). Starspot photometry with robotic telescopes: Continuous UBV and V(RI)_C photometry of 23 stars in 1991-1996. *Astronomy and Astrophysics Supplement series*, 125:11–63.

Strassmeier, K. G. and Linsky, J. L., editors (1996). *Stellar surface structure: proceedings of IAU Symposium 176.*

Vaníček, P. (1969). Approximate Spectral Analysis by Least-Squares Fit. Successive Spectral Analysis. *Astrophysics and Space Science*, 4:387–391.

Vaughan, A. H., Preston, G. W., Baliunas, S. L., Hartmann, L. W., Noyes, R. W., Middelkoop, F., and Mihalas, D. (1981). Stellar rotation in lower main-sequence stars measured from time variations in H and K emission-line fluxes. I - Initial results. *Astrophysical Journal*, 250:276–283.

Vogt, S. S. (1981). A method for unambiguous determination of starspot temperatures and areas - Application to II Pegasi, BY Draconis, and HD 209813. *Astrophysical Journal*, 250:327–340.

Vogt, S. S. (1983). Spots, spot-cycles, and magnetic fields of late-type dwarfs. In Byrne, P. B. and Rodono, M., editors, *ASSL Vol. 102: IAU Colloq. 71: Activity in Red-Dwarf Stars*, pp. 137–155.

Walker, G., Matthews, J., Kuschnig, R., et al. (2003). The MOST Asteroseismology Mission: Ultraprecise Photometry from Space. *PASP*, 115:1023–1035.

Walker, G. A. H., Croll, B., Kuschnig, R., et al. (2007). The Differential Rotation of κ^1 Ceti as Observed by MOST. *The Astrophysical Journal*, 659:1611–1622.

Weiss, W. W., Aerts, C., Aigrain, S., et al. (2004). Additional science potential for COROT. In Favata, F., Aigrain, S., and Wilson, A., editors, *Stellar Structure and Habitable Planet Finding*, vol. 538 *ESA Spec. Publ.*, pp. 435–444.

Wilson, O. C. (1978). Chromospheric variations in main-sequence stars. *Astrophysical Journal*, 226:379–396.

You, J. (1999). *Time Series Analysis of Brightness Variations of Stars*. Licentiate thesis, Astronomy Division, University of Oulu.