Studying Stellar Microvariability using data from the *Kepler Mission*

MSci Project Report

Author:
Ian Woolf

Supervisor:
Dr. Yvonne Unruh

Imperial College London

April 2012

Contents

List of Figures

1 Abstract

This project aimed to use data from the *Kepler* mission to study stellar microvariability, primarily through the detection of intrinsic periods of variation within lightcurves. To achieve this, a study of different mathematical methods was undertaken. It was found that the Lomb-Scargle periodogram and the continuous wavelet transform were appropriate methods to use, but phase dispersion minimisation and information entropy minimisation were rejected due to poor performance. It was demonstrated that fitting Gaussians to be periodogram power spectrum provides a useful estimate of the uncertainty. The methods chosen were proved to work by recovering known values for the Sun, including the eleven year solar cycle, and evidence for the Carrington rotation period. Application of these methods to *Kepler* targets revealed that most of the intrinsic variability has periods of between 1 and 25 days. Interpreting these as rotation periods suggests the target stars are, in general, younger than the Sun and hence rotate faster. Two *Kepler* targets were selected for a more detailed discussion and possible interpretations of their lightcurves were suggested.

2 Introduction

2.1 Motivation

Ever since Galileo used his newly invented telescope to observe the Sun and saw dark spots on its surface around 1611 (Solanki, 2003), it has been known that it is not constant and unchanging. More recently (Willson and Hudson, 1981), scientists have been able to observe many different sources of variation, operating on a wide range of timescales, from minutes to decades.

Any attempt to model the behaviour of stars must include these variations. However, using the Sun as the only test case is not sufficient to draw any general conclusions about stellar behaviour. To test models thoroughly, knowledge of the variations of many more stars is required. The *Kepler* Mission provides a perfect source of data for this analysis.

2.2 Aims

The main aim of this project is to study the microvariability of Sun-like stars using data collected by the *Kepler* Mission. In order to achieve this, various different mathematical methods of finding characteristic periods within the data will be analysed and their reliability and accuracy tested. These methods will then be used on data from the Sun, in an attempt to recover well known results. This will demonstrate the suitability of the techniques for use on *Kepler* data.

These methods will be used to identify periods of interest within the available *Kepler* target stars. Possible interpretations of these periods will be discussed, with reference to the properties of the star in question. The prevalence of different timescales of variability across the data set will be quantified and discussed.

3 Stars

3.1 Introduction

Stars form when clouds of gas (largely molecular hydrogen) undergo a gravitational collapse. As the cloud contracts, the gravitational energy of the particles is converted into kinetic energy (Green et al., 2004, p163), increasing the temperature to such a level that nuclear fusion can occur. Initially, stars burn hydrogen (through either the p-p chain or the CNO bi-cycle - see Prialnik (2000, pp59-63) for more details). The main product of these reactions is helium, but some heavier elements are also produced.

Stars spend most of their lifetimes in this hydrogen-burning phase, so most observed stars are in this period of their lifetime. As a result, these stars produce the most densely populated region of the Hertzsprung-Russell diagram, which is known as the main sequence. This is shown in Figure 1[1]. Each of the dots represents a star. Most of them are in the group stretching from the top left to the bottom right of the diagram; this is the main sequence. Once the stars run out of hydrogen, they begin to fuse heavier elements and have reached the post main sequence phase.

Stars are divided into different spectral types depending on their properties. The major classes, in order of decreasing temperature, are O, B, A, F, G, K and M. More details can be found in Tayler (1994), but Figure 1[1] provides information about the temperatures of the classes. The Sun is a G-type star with a temperature of 5780K Güdel (2007).

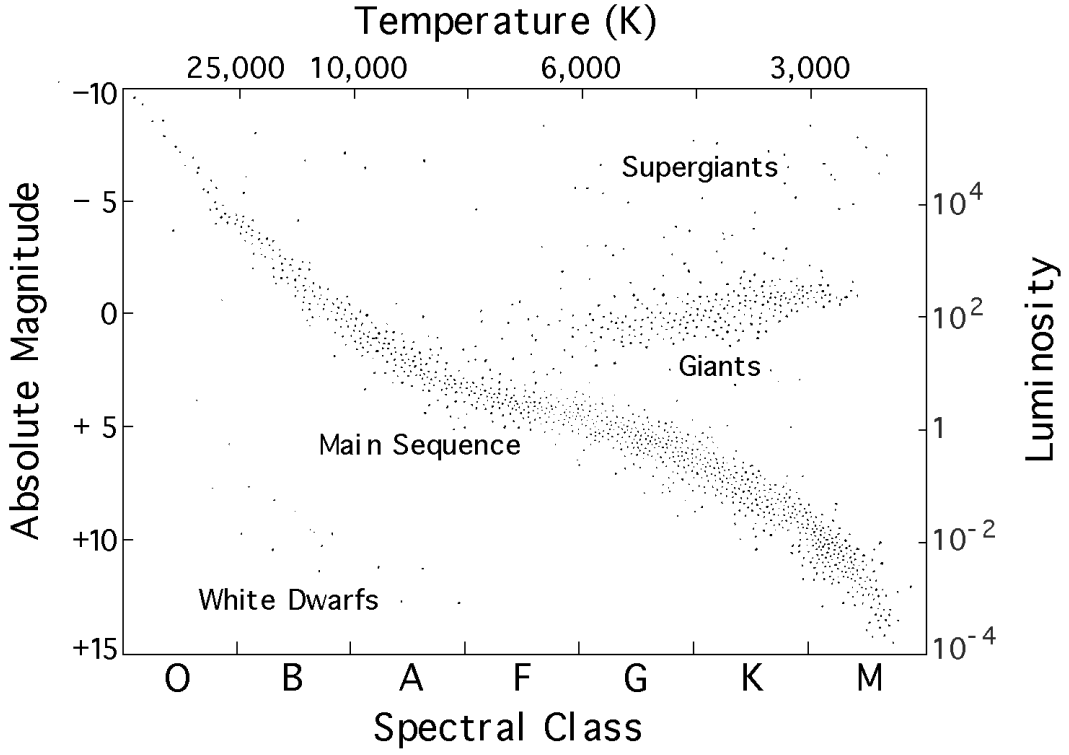


Figure 1: A Hertzsprung-Russell diagram, with the main groups labelled. The luminosity is displayed relative to $L_\odot$. Note that the temperature scale decreases from left to right. The diagram also serves to show the temperatures of different spectral classes of stars.[1]

[1]NASA HEASARC. The Classification of Stars. [online] Available at: http://heasarc.gsfc.nasa.gov/docs/RXTE_Live/class.html [Accessed 09/10/2011].

3.2 Stellar Structure

3.2.1 Internal structure

Stars are made up of several different layers. In the centre of the star is the core, which is the location of heat production from nuclear fusion. Once it has been released, energy from this process travels outwards from the core. There are two methods of energy transfer; radiation and convection.

When radiation is the dominant method of energy transport, photons produced by the fusion reactions random-walk away from the centre by scattering, or absorption and re-emission of photons. The photons produced in the core are very high energy γ-rays, but the interactions redistribute this energy to the surrounding plasma (Green et al., 2004, pp52-3).

As the photons move further out, the temperature decreases until the temperature gradient is sufficient to allow convection (Phillips, 1999, p94). When energy transport by convection dominates, the plasma absorbs the radiation coming from below (either from the core or a radiative layer), heats up and begins to rise. It then radiates the energy again, which causes it to cool, and sink back down again. These convection cells are thought to be the cause of the granulation seen on the solar surface (Green et al., 2004, p54).

The respective sizes and positions of radiative and convective zones within stars depends on their mass. As mass increases, the extent of the outer convection zone decreases. For stars a little more massive than the Sun, they disappear entirely, and an inner convective zone is present instead. (Prialnik, 2000, p139-40).

A diagram showing the distribution of these zones within the Sun is shown in Figure 2. Note the random walk of photons through the radiative layer, and that the convection zone contains more than one set of convection cells. The cells causing granulation do not reach all the way down to the radiative zone.

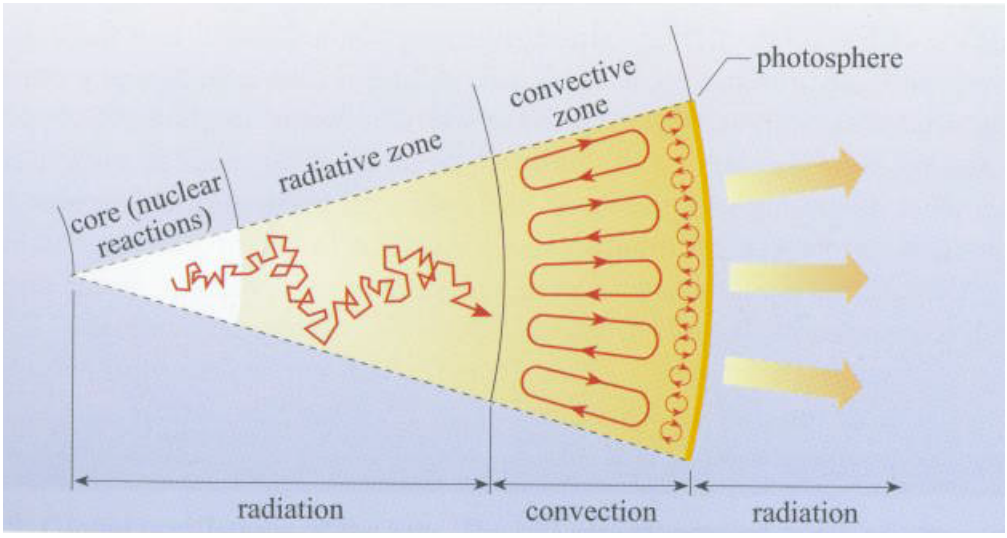


Figure 2: The interior structure of Sun-like stars. From Green et al. (2004, p55)

3.2.2 Atmosphere

When viewing the Sun in the visible spectrum, one is actually looking at the photosphere. This is a layer that radiates similar to a black body with an effective temperature of around 5780K (Güdel, 2007). The Sun's photosphere is around 500km thick (Green et al., 2004, p8), which is a very thin shell. When viewing the Sun from a fixed position, it appears as a circle. In the visible spectrum, areas nearer the circumference are darker than the centre due to an effect called limb darkening. This is mainly a geometrical effect. Lines of sight in the centre of the sun penetrate deeper into the photosphere, as they are closer to the normal to the surface, and therefore more light is visible. Lines of sight towards the limbs reach shallower depths, which are cooler and therefore dimmer (see Green et al., 2004, pp9-10 for a more detailed discussion). This effect varies throughout the spectrum. In some regions, it has no effect, and in others, limb brightening occurs (see e.g. Withbroe (1970)).

Since the *Kepler* mission measures primarily within the visible spectrum (see Section 4), it will detect variation in the photospheres of its target stars. Stellar atmospheres also consist of the chromosphere and the corona but they have little relevence to this analysis.

3.2.3 Differential Rotation

Since they are made of plasma, stars may not have equal angular velocities at all latitudes. In fact, rotation periods around the equator can be significantly shorter than those at the poles. For example, the rotation period of the Sun ranges from around 24 days to around 38 days (Stenflo, 1990). For some stars, including the Sun, differential rotation is a key ingredient of the stellar dynamo. Differential rotation has important implications for lightcurves of stars where variability is present. Two sunspots at different latitudes will have different rotation periods, and so both these periods should be present within the star's lightcurve.

3.2.4 Magnetic Field

Almost all cool stars, including the Sun, have magnetic fields produced by an internal dynamo (Güdel, 2007). Magnetic fields are produced by the motion of charged particles, and in the case of stars there are two types of motion: large scale rotation of the star, and small scale convective motions (Tayler, 1997, p83).

The solar magnetic field is approximately dipolar (Solanki et al., 2006). The solar dynamo excites and de-excites itself on a 22 year cycle, with with poles reversing every 11 years. Figure 3[2] provides a beautiful demonstration of this field reversal.

3.3 Stellar Variability

3.3.1 Starspots & Faculae

Sunspots are a solar feature that have been known about for hundreds of years, because it is possible to observe them with the naked eye. They are regions of the photosphere that appear darker than their surroundings, and they occur because the strong magnetic field

[2]NASA/MSFC/David Hathaway. [online] Available at: http://solarscience.msfc.nasa.gov/dynamo.shtml [Accessed 23/04/2012]

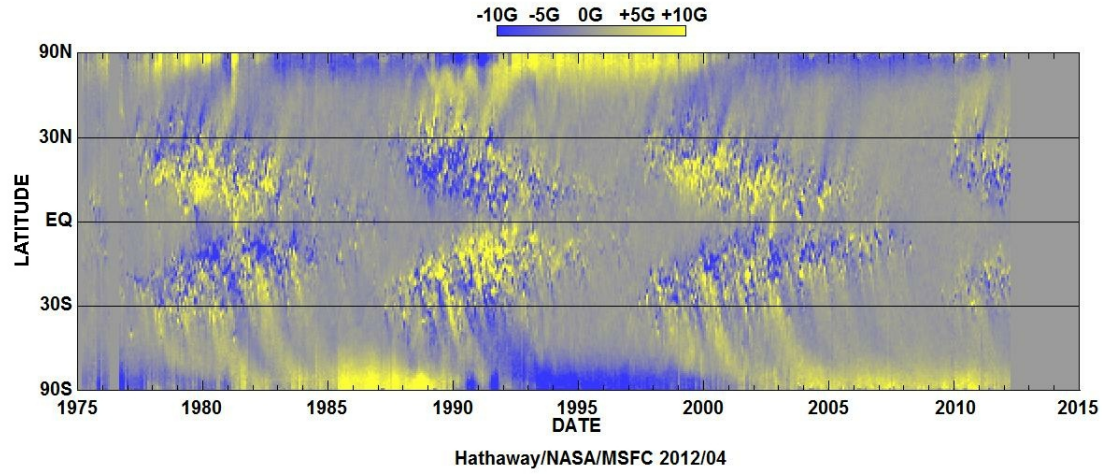


Figure 3: The distribution of the Sun's magnetic field over time. Magnetic field lines directed ito the Sun are blue, those directed out of the sun are yellow. Image credit: NASA/MSFC/David Hathaway[2]

in active regions inhibit the convective transfer of heat from the solar interior (Thomas and Weiss, 2008, p2). The core of a sunspot (called the umbra) ranges from 1000-1900K cooler than the rest of the photosphere. Sunspots generally lie within 40° of the equator (Thomas and Weiss, 2008, p20). According to Strassmeier (2009), starspots can be much larger than sunspots, and appear at higher latitudes. Sunspots have lifetimes ranging from hours to months (Solanki, 2003), but there is evidence to suggest that active regions on other stars can live for many years (Strassmeier, 2009).

Sunspots are accompanied by nearby, bright patches called faculae. Faculae are hotter than the surrounding photosphere (Thomas and Weiss, 2008, p126) and their prevalence varies in phase with the solar cycle, discussed below. The extra luminosity from faculae has a greater effect than the loss caused by spots on older stars, like the Sun, but it appears the reverse is true on younger stars (Radick, 2003).

3.3.2 The Solar Cycle

As mentioned in section 3.2.4, the Sun's magnetic field varies in strength over a twenty-two year cycle. Since sunspots are caused by the magnetic field inhibiting energy transfer through convection, a stronger magnetic field results in more sunspots and faculae. Because of this, total solar irradiance varies in phase with the changes in the magnetic field. Since it is the strength, not direction, of the magnetic field that causes variation, irradiance follows an eleven year cycle rather than twenty-two.

Figure 4[3] contains two plots that demonstrate the solar cycle. The bottom panel shows the area of the solar hemisphere covered by sunspots since 1880. The eleven year cycle is clearly visible. There is also much greater variation during solar maxima, with many sharp spikes. In contrast, solar minima are consistently empty. The top panel shows the so-called butterfly diagram, which shows that sunspot positions vary over the solar cycle,

[3]NASA MSFC, 2011. The Sunspot Cycle. [online] Available at: http://solarscience.msfc.nasa.gov/SunspotCycle.shtml [Accessed 09/10/2011].

starting at high latitudes and moving down towards the equator over the course of the cycle. They fade out at solar minima, before re-appearing again at high latitudes.

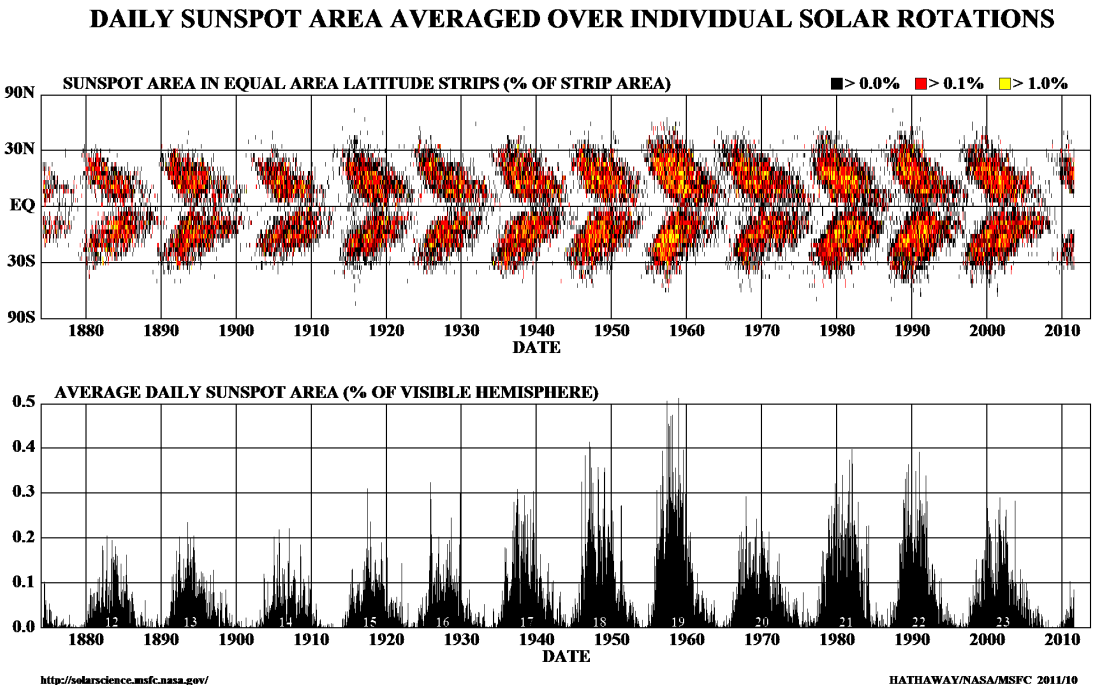


Figure 4: *Upper Panel*: Butterfly diagram showing the position of Sunspots since 1880. *Lower Panel*: Area of the visible solar hemisphere covered by sunspots over the same period.[3]

Many *Kepler* target stars have magnetic fields, so it is likely that they will have similar cycles. The length of the mission should ensure that any given star does not stay at its minimum throughout, and the discovery of the steady increase or decrease in variability would be an interesting result.

3.3.3 Pulsations

Another cause of variation in stars is pulsations, the study of which subject is called asteroseismology. These pulsations are essentially standing sound waves within the star (Ostlie and Carroll, 1996, p549). The simplest modes are purely radial, but more complicated modes exist. The pulsations cause changes in the temperature of the star, which in turn affects its luminosity. Pulsations generally occur on a timescale shorter than those of interest here.

3.3.4 Age-Activity-Rotation Relation

Almost all cool stars, including the Sun, have magnetic fields produced by an internal dynamo (Güdel, 2007). Güdel also presents evidence that the Sun was far more active when it was younger, as more rapid rotation induced a more efficient magnetic dynamo. This resulted in stronger magnetic fields which led to greater variability. Soderblom et al. (1993) discusses the loss of angular momentum over the lifetime of a star, as the star loses angular momentum through its wind. Since rotation affects magnetic activity, there is an overall connection between age, activity and rotation of stars. Young stars tend to

rotate rapidly, resulting in stronger magnetic fields and greater variability. Over time, the rotation slows down, leading to weaker magnetic fields and reduced variability. Barnes (2007) describes a technique called gyrochronology that uses this relationship to determine the age of stars with errors of roughly 15%.

4 The *Kepler* Mission

4.1 Introduction

On 6^{th} March 2009, the *Kepler* Mission was launched. It is part of NASA's Discovery Programme[4], aimed at relatively low cost, science-oriented missions. Its primary purpose, described by Koch et al. (2010), is exoplanet detection. Specifically, it was designed to be able to detect Earth-sized planets within the habitable zones of Sun-like stars (stellar classes F, G and K), and to determine their orbital period and other characteristics. The mission will also generate data of value to other fields, including asteroseismology. In order to achieve this, the photometric output of stars must be continually measured. It measures light primarily in the visible spectrum (although the full range stretches from around $300nm$ to around $1000nm$ - see Koch et al. (2010)).

Figure 5 shows a computer generated image of the *Kepler* Space Telescope[5].

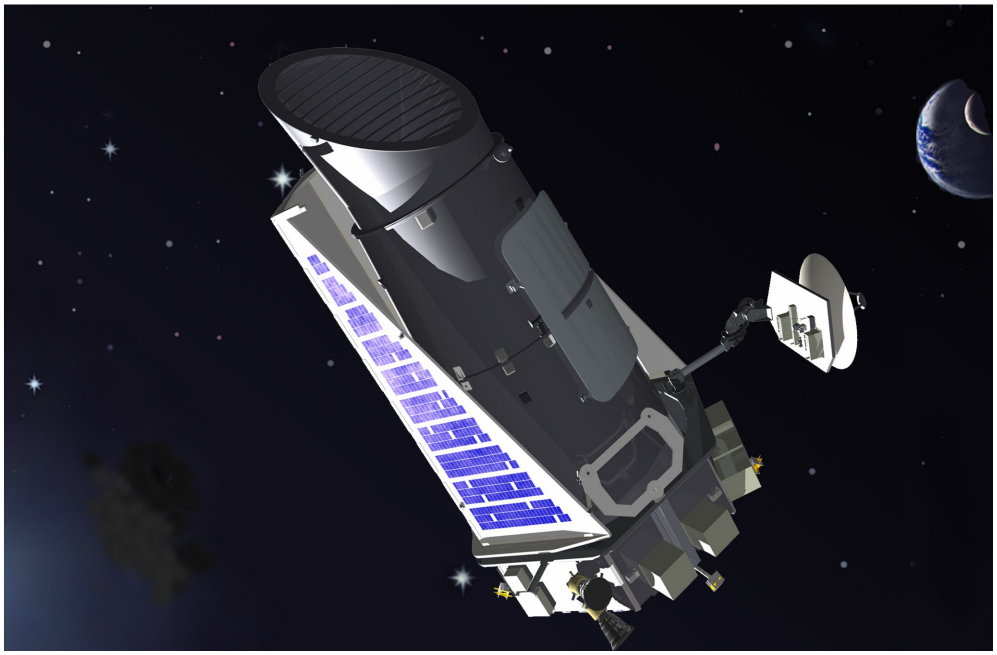


Figure 5: The *Kepler* Space Telescope. Credit: NASA[5]

Figure 6 is a picture of the Milky Way galaxy, showing the position of the Solar System and search space of the *Kepler* mission[6].

[4]http://discovery.nasa.gov/index.cfml

[5]NASA. Artist's rendition of *Kepler* in space. [online] Available at: http://www.redorbit.com/news/space/1650645/after_launch_kepler_prepares_to_carry_out_its_mission/ [Accessed 17/03/2012].

[6]Jon Lomberg. *Kepler* search space. [online] Available at: http://www.jonlomberg.com/profile.html [Accessed 17/03/2012].

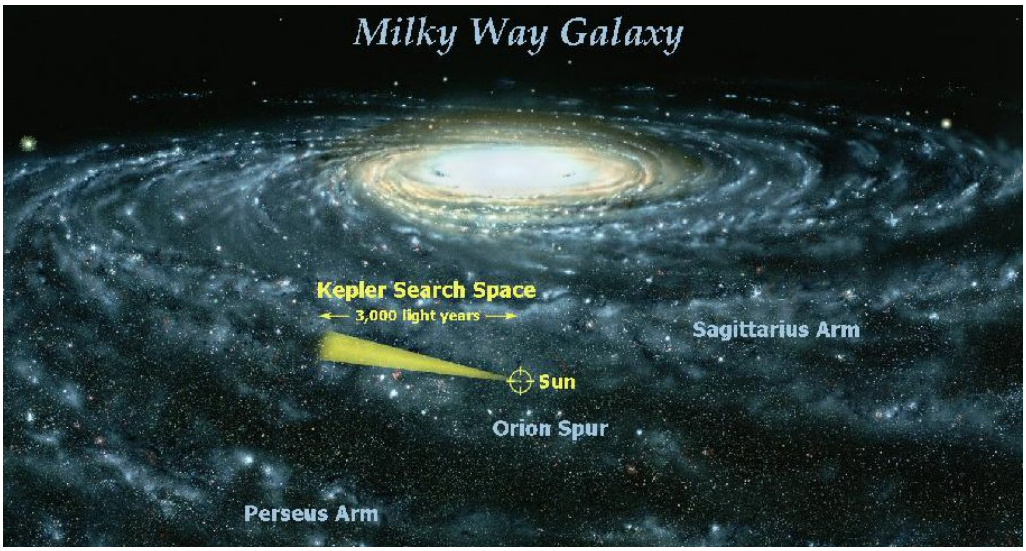


Figure 6: A map of the Milky Way, showing the position of the Solar System and the search space of the *Kepler* Mission. Credit: Jon Lomberg[6]

4.2 Mission Parameters

Koch et al. (2010) describe the scientific goals of the mission, and their influence on mission design. Several requirements were identified to enable the mission to fulful these goals.

The first requirement was a mission length of at least 3 years. This is necessary to ensure measurement of the three planetary transits required for confidence, as an Earth-like planet around a Sun-like star would have an orbital period of about a year. A second requirement was to observe at least 100,000 Sun-like stars. Planetary transits can only be detected photometrically if the planet passes between the star and the Earth. This places severe constraints on the orbital alignment, to the extent that an enormous number of stars must be monitored to find even a few planets. Other requirements concerned high photometric precision and a low level of photometric noise. The mission must be sensitive enough the be able to detect the change in the photometric output of the Sun that would be caused as the Earth passes in front of it.

4.3 Relevance

A mission designed with these goals in mind is clearly also ideal for the study of stellar microvariability. The requirements of these two branches of astronomy are largely identical, because they involve making exactly the same measurements - the changing output of stars - albeit due to different causes. The data set for studying stellar microvariability is also far larger than that for planetary transits, because, for the reasons described above, planetary transits are only detectable on a minority of stars. The inherent variability of stars actually presents a problem for the detection of exoplanets by providing another source of noise. A better understanding of this variability will therefore aid in the detection of exoplanets as it may allow more effective removal of such signals from planetary transit data.

5 Methods for period finding

5.1 Introduction

The main motivation of studying stellar lightcurves is to find intrinsic periodicities. Variation resulting from the rotation of the star, eclipses, or pulsations, should repeat on a given timescale. While these periodicities are sometimes obvious upon a simple visual examination, quantitative mathematical methods are preferred for several reasons. They remove any observer bias, and allow errors to be generated in a repeatable and consistent manner, and allow automation. Several different methods for period finding were used in this project, including periodograms, continuous wavelet transforms, phase dispersion minimisation and information entropy minimisation.

To demonstrate the respective strengths and weaknesses of the various methods, their performance will be tested against several artificial lightcurves. These signals are shown in Figure 7. Lightcurve A is a simple sinusoid with a fixed frequency, Lightcurve B's frequency increases over the course of the signal, Lightcurve C's frequency changes abruptly halfway through the signal, and Lightcurve D is a superposition of a high frequency and a low frequency signal. The frequencies present in each lightcurve, and the corresponding periods, are shown in Table 1.

Signal	Frequencies (1/days)	Periods (days)
A	0.005	200
B	0.0030 to 0.0093	333 to 108
C	0.0064 then 0.0200	157 then 50
D	0.01 and 0.10	100 and 10

Table 1: The frequencies and periods present within the artificial lightcurves shown in Figure 7.

5.2 Periodograms

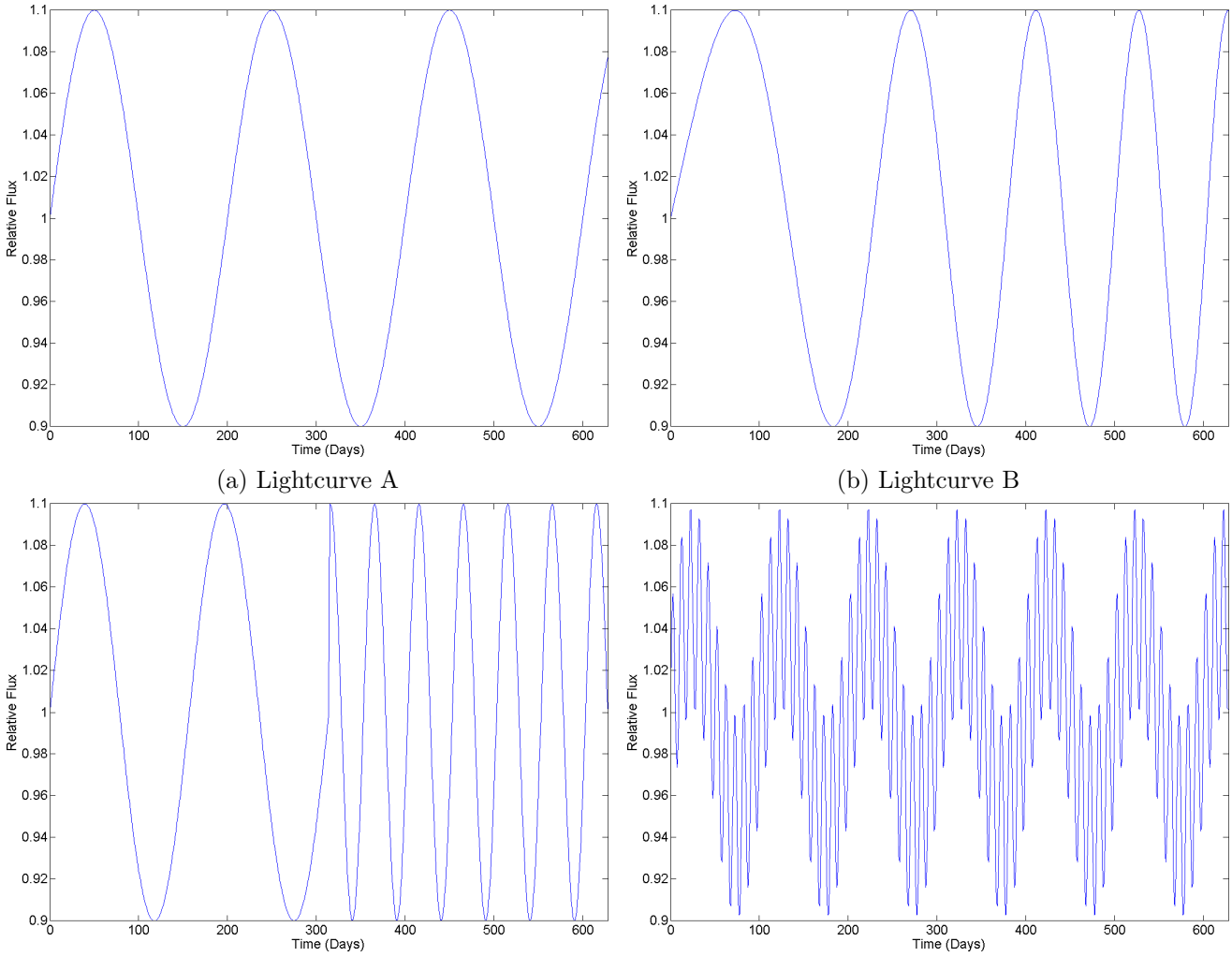
5.2.1 Introduction

Periodograms are a widely used method for finding periodicities within signals. The classical periodogram is simply a normalised discrete Fourier transform. For a data set $\{X(t_i), i = 1, 2...N_0\}$, the classical periodogram (Schuster, 1898) is defined as

$$P_X(\omega) = \frac{1}{N_0}\left|\sum_{j=1}^{N_0} X(t_j)\exp(-i\omega t_j)\right|^2 . \quad (1)$$

This works because the presence of a period within the data means X_t and $\exp(-i\omega t)$ are in phase at that period, and so the sum results in a large value of P_X. Where there is no period, the terms are randomly positive and negative, resulting in a small value of P_X. P_X is often referred to as the power.

Due to statistical difficulties and issues with spectral leakage, Scargle (1982) proposed an alternative definition, namely



(a) Lightcurve A

(b) Lightcurve B

(c) Lightcurve C

(d) Lightcurve D

Figure 7: Four different artificial lightcurves to demonstrate the performance of different methods for period finding.

$$P_X(\omega) = \frac{1}{2}\left\{\frac{\left[\sum_j X_j \cos\omega(t_j-\tau)\right]^2}{\sum_j \cos^2\omega(t_j-\tau)} + \frac{\left[\sum_j X_j \sin\omega(t_j-\tau)\right]^2}{\sum_j \sin^2\omega(t_j-\tau)}\right\}, \tag{2}$$

where

$$\tan(2\omega\tau) = \frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j}. \tag{3}$$

This happens to be exactly equivalent to the reduction of the sum of squares in least-squares fitting of sine waves to the data. It also has simple statistical behaviour and is only slightly more computationally intensive.

Despite this redefinition, improvements can still be made. The Lomb-Scargle Periodogram does not account for measurement errors, and it contains the implicit assumption that the mean of the signal and the mean of the fitted sine wave are the same. To counter both of these issues, Zechmeister and Kürster (2009) introduce a Generalised Lomb-Scargle Periodogram, which solves both these issues.

To provide a measure of the errors of the periodogram, Scargle (1982) describes a False Alarm Probability, which is the probability of a peak of a particular height being generated by a random signal containing no periods. False positives occur with probability p_0 at a power z_0, where

$$z_0 = -\ln\left[1 - (1 - p_0)^{1/N}\right] \tag{4}$$

and N is the number of frequencies searched for the maximum. For a target value of p_0, equation 4 tells you the required power.

To implement the Lomb-Scargle Periodogram, a range of frequencies was calculated to the desired level of sampling. The power at each frequency was calculated using Equation 2. These were then plotted against each other to produce a power spectrum.

5.2.2 Examples

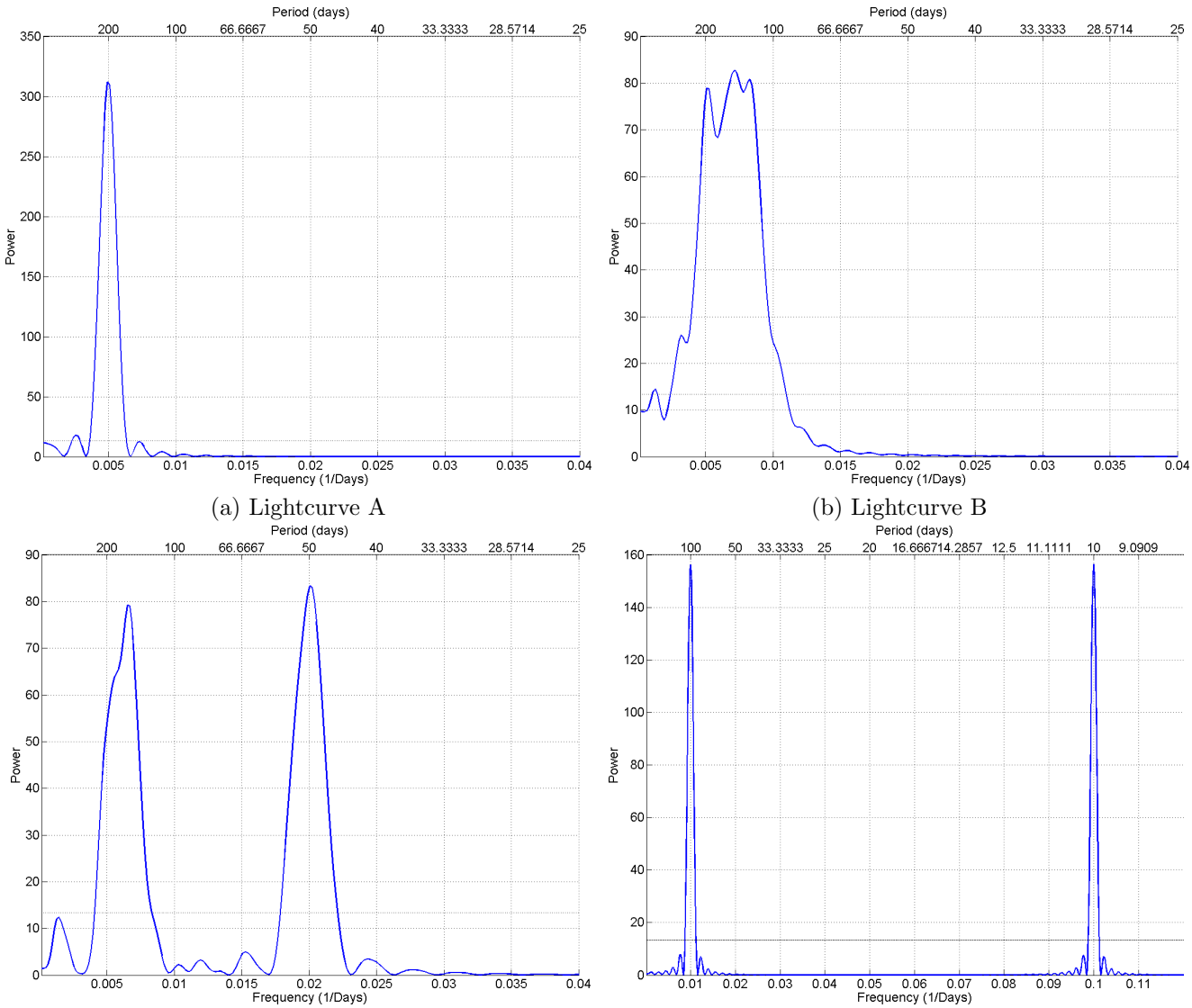
To demonstrate the performance of the periodogram, Figure 8 shows a Lomb-Scargle periodogram for each of the artificial lightcurves from Figure 7. The frequencies of the signals actually present within the lightcurves (and the corresponding periods) are shown in Table 1 for reference. All the results in this section are quoted with uncertainties. These are calculated using the procedure described in Section 5.2.3.

Lightcurve A is a basic sinusoid with a fixed frequency of 5.0×10^{-3} days^{-1}. The periodogram recovers a frequency of $(4.98 \pm 0.68) \times 10^{-3}$days^{-1}, correctly identifying the frequency present. While encouraging, it should be remembered that the signal consists of one signal and no noise, so it is the most trivial of cases.

Much less trivial is lightcurve B. Here, the frequency of the signal changes with time, moving linearly from 3.00×10^{-3} to 9.30×10^{-3} days^{-1} over the course of around 600 days. The resulting periodogram has a very wide, fairly symmetric peak, which returned the result of $(7.17 \pm 2.22) \times 10^{-3}$days^{-1}, meaning the uncertainties quoted for lightcurve B are too small to cover the entire range of frequencies present. The method used to calculate uncertainties provides an underestimate in the case of changing frequencies. This is less important than it might seem initially for two reasons. First, as will be seen in Section 5.3, another method, the Continuous Wavelet Transform, clearly shows if frequencies are changing. The other mitigating factor is the sheer size of the frequency change in lightcurve B: the frequency more than triples across the lightcurve. This is far greater than any frequency change that would be expected from an actual star.

Lightcurve C starts off as a simple sinusoid with a frequency of 6.40×10^{-3} days^{-1}, but halfway through it abruptly switches to a sinusoid with a frequency of 20.00×10^{-3} days^{-1}. The periodogram recovers frequencies of $(6.19 \pm 1.41) \times 10^{-3}$days^{-1} and $(19.99 \pm 1.33) \times 10^{-3}$days^{-1} , so it seems to be performing well. However, the periodogram can provide no indication of when the frequencies are present. A very similar periodogram would be expected from a lightcurve where these two frequencies are superposed throughout.

Lightcurve D contains just such a case, where frequencies of 10.0 and 100.0 days^{-1} are superposed throughout. The periodogram has spikes at $(10.00 \pm 0.68) \times 10^{-3}$days^{-1} and $(100.00 \pm 6.80) \times 10^{-3}$days^{-1} , showing it can recover the correct frequencies, but there is no way of knowing it is not a result of an abrupt change like the one present in lightcurve C.



(a) Lightcurve A

(b) Lightcurve B

(c) Lightcurve C

(d) Lightcurve D

Figure 8: Lomb-Scargle Periodograms of the test signals from Figure 7.

5.2.3 Uncertainties

An advantage of periodograms is that they allow the identification of a specific period within a signal, by locating the peaks in the power spectrum. It is always important to be aware of the limitations of the method and quote results with uncertainties.

The periodograms of the artificial lightcurves (see Figure 8) provide some useful information for this purpose. It is expected that it would be possible to quote the frequency of lightcurve A to much greater accuracy, since lightcurve B does not have a fixed frequency. The peak of lightcurve A, which has a fixed frequency, is far narrower than that of lightcurve B, where the frequency changes throughout the lightcurve. This indicates that the width of the peak is related to the uncertainty. This relationship can be used to estimate the uncertainty in the position of the peak in the power spectrum.

In order to do this, a Gaussian is fitted to the peak (using a Trust-Region-Reflective algorithm, see e.g. Coleman and Li (1993)). The peak of the Gaussian is used as the detected frequency, and the Full Width Half Maximum (FWHM) is used to calculate the

upper and lower bounds of the uncertainty.

Another tactic that was used to estimate the uncertainties in the positions of periodogram peaks was the bootstrap method (e.g. Davison and Hinkley (1997)). This method involves replacing a certain proportion of data points with data points selected randomly from the data. The periodogram is then calculated again and compared to the original. This is repeated a large number of times.

The idea behind this method is to see how robust a result is by changing the data it was calculated from. However, when it was implemented, it was found that the results were incredibly robust, and hardly changed, even when 20% of the data was replaced. The frequencies found by the different repeats were plotted on a histogram, and then a Gaussian fitted to that. Again, the FWHM of the Gaussian was used to estimate the error.

An example will be used to demonstrate this. Using any of the artificial lightcurves in their current form is flawed, because (with the exception of Lightcurve B), they all contain perfect sinusoids in various guises. As a result, it is expected that the periodogram should perform very well, since it is equivalent to a least squares fitting of sinusoids to the data. In contrast, real stars are very different. Many more signals may be present, random noise certainly will be, and the measurements may not be perfectly uniformly distributed in time. In addition, physical signals may not be perfect sinusoids.

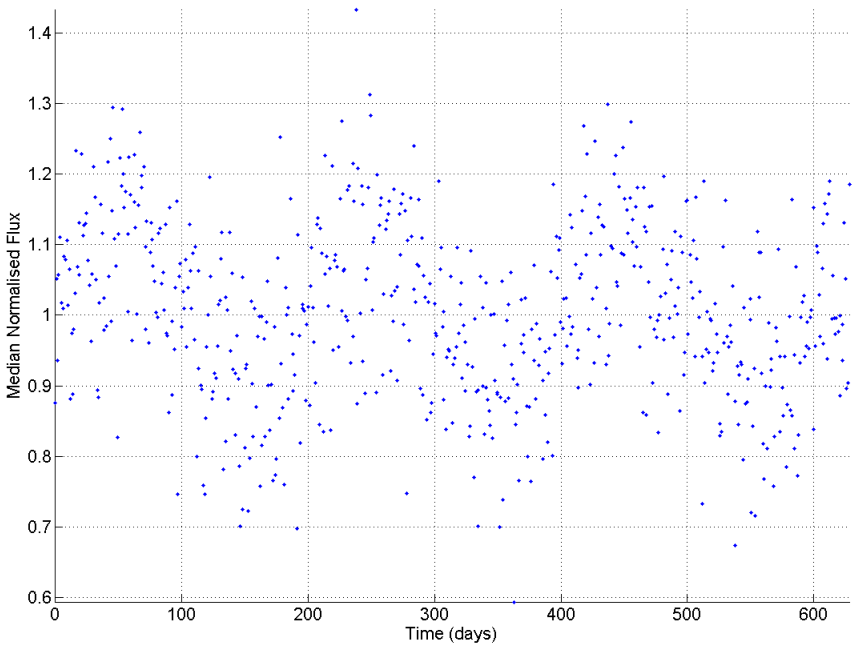


Figure 9: An adapted version of Lightcurve A, used to compare the two possible methods of error analysis.

To reflect these issues, Lightcurve A was adapted to provide a more realistic test subject. First, normally distributed noise (with a mean of 0 and a standard deviation of 0.1) was added to the data. Subsequently, each data point was moved in time by a normally distributed random number (with a mean of 0 and a standard deviation of 0.3 days), 'destroying' the perfect sinusoid within the data. The adapted Lightcurve A is

plotted in Figure 9.

This adapted lightcurve was subjected to Lomb-Scargle periodogram analysis and both methods were used to estimate the uncertainty. First, Gaussians were fitted to each peak of the periodogram and the FWHM calculated to provide the uncertainty. A plot of the periodogram with the fitted Gaussian is shown in Figure 10.

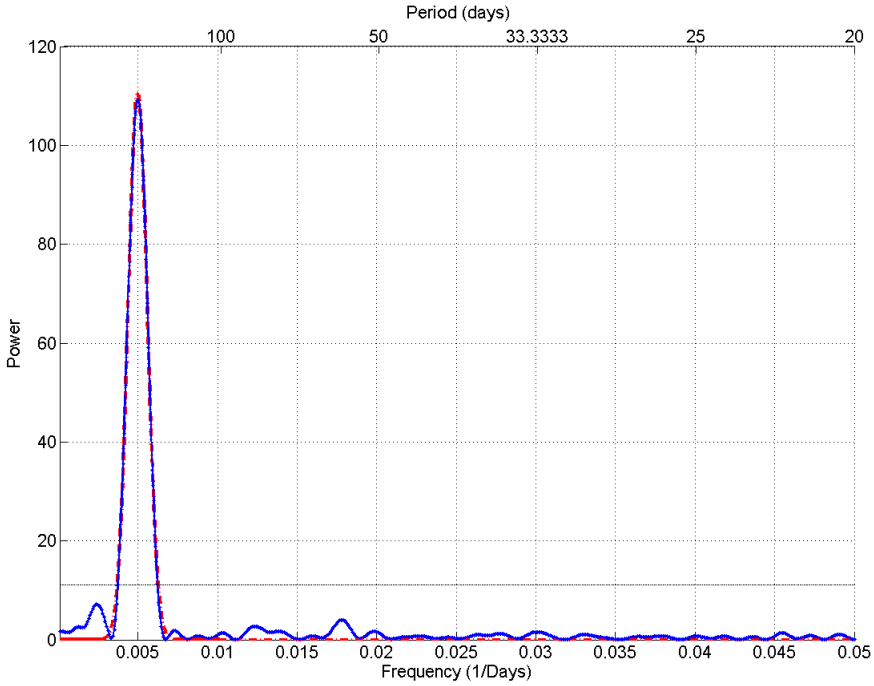


Figure 10: A Lomb-Scargle Periodogram for the adapted Lightcurve A (blue continuous line), with a Gaussian fitted to the peak (red dot-dash line) for error estimation.

The FWHM of the fitted Gaussian gives an value for the frequency as $(4.98 \pm 0.70) \times 10^{-3}\text{days}^{-1}$. Recall that the original lightcurve contained a frequency of $5.00 \times 10^{-3}\text{days}^{-1}$, but since each data point has been shifted by a small random amount, it is now impossible to quote a specific frequency within the signal.

The second method that was tried was the bootstrap method, described above. 20% of points were replaced each time, and 1000 repeats performed. Figure 11 shows a histogram of the distribution of the frequencies of peaks in the Lomb-Scargle periodogram. This has also been fitted with a Gaussian to allow the FWHM to be used to estimate the error.

The fitted Gaussian gives a value of the frequency as $(4.99 \pm 0.06) \times 10^{-3}\text{days}^{-1}$. Given that the frequency of $5.00 \times 10^{-3}\text{days}^{-1}$ no longer exists exactly within the frequency, this is a worryingly small error margin. It is also worth noting that 20% is a very high proportion of data points to replace. 10% is a more common choice, but if used here the periodogram returns the same peak every single time. Admittedly, the possible positions for the peaks are limited by sampling, but given that the perfect signal no longer exists within the lightcurve, the author is very skeptical about the results obtained using the bootstrap method.

As a result of this it was decided to use the first method (fitting Gaussians to periodogram peaks) to estimate the errors in frequencies found by the Lomb-Scargle periodogram. Quite apart from giving suspiciously small errors, it is extremely computation-

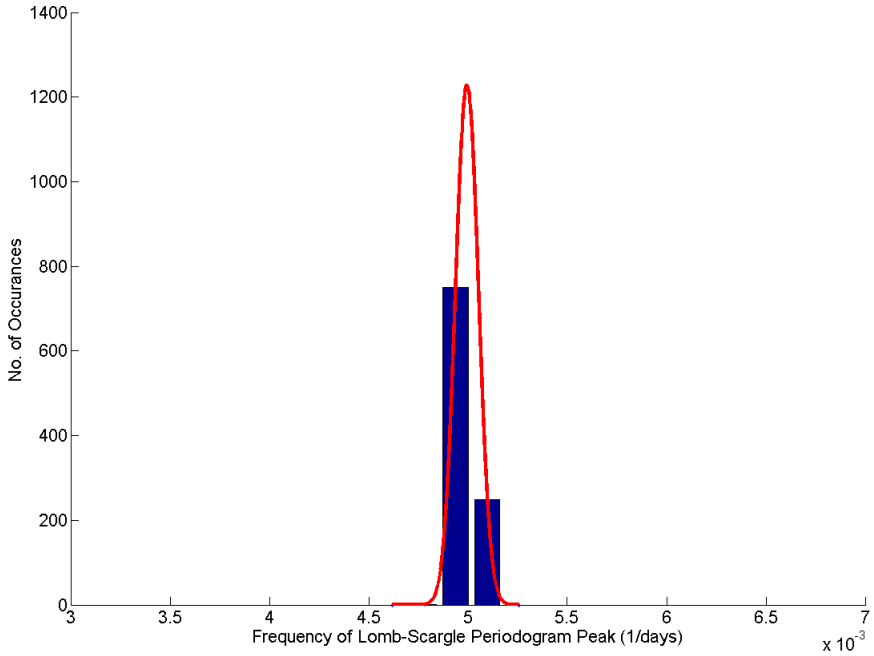


Figure 11: A histogram showing the distribution of peak frequencies from the Lomb-Scargle Periodogram when subjected the the bootstrap method with 1000 repeats and 20% replacement. A Gaussian has been fitted (in red) to allow estimateion of errors.

ally intensive. Another factor to consider is that the levels of accuracy provided by the simpler method are more than sufficient. There is little point trying to measure periods in target stars with supreme accuracy. Obtaining periods with errors of order unity (in days) will be more than enough for the analysis of period distribution throughout the data set, and speculative interpretations of individual targets.

5.3 Continuous Wavelet Transform

5.3.1 Introduction

Another method used to identify periods within a signal is the continuous wavelet transform. This method allows the decomposition of a time series into time-frequency space, allowing simultaneous analysis of the dominant periods that exist within a signal, and how these change over time (Torrence and Compo, 1998).

The Continuous Wavelet Transform of a discrete sequence X_j is defined as the convolution of X_j with a scaled and translated version of the wavelet function, ψ_0:

$$W_j(s) = \sum_{j'=0}^{J-1} x_{j'} \psi^* \left[\frac{(j'-j)\delta t}{s} \right] \tag{5}$$

where s is the wavelet scale, n is the localised time index and δt is the time seperation of data points. The wavelet function must be localised in both time and frequency space and have zero mean (Farge, 1992). A popular choice of wavelet function, and the one used here, is the Mortlet Wavelet. This is a plane wave modulated by a Gaussian and is defined as

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\eta^2/2} \tag{6}$$

where ω_0 is the nondimensional frequency, taken to be 6 to satisfy the conditions described above, and η is the non-dimensional time. Note that Equation 5 contains the term ψ^*, rather than ψ_0 from equation 6, to reflect that it has been normalised and the complex conjugate used.

The continuous wavlet transform was implemented using wavelet software was provided by C. Torrence and G. Compo, which is available at
URL: http://paos.colorado.edu/research/wavelets/.

5.3.2 Examples

A Continuous Wavelet Transform was performed on the artificial lightcurves from Figure 7. The results are shown in Figure 12.

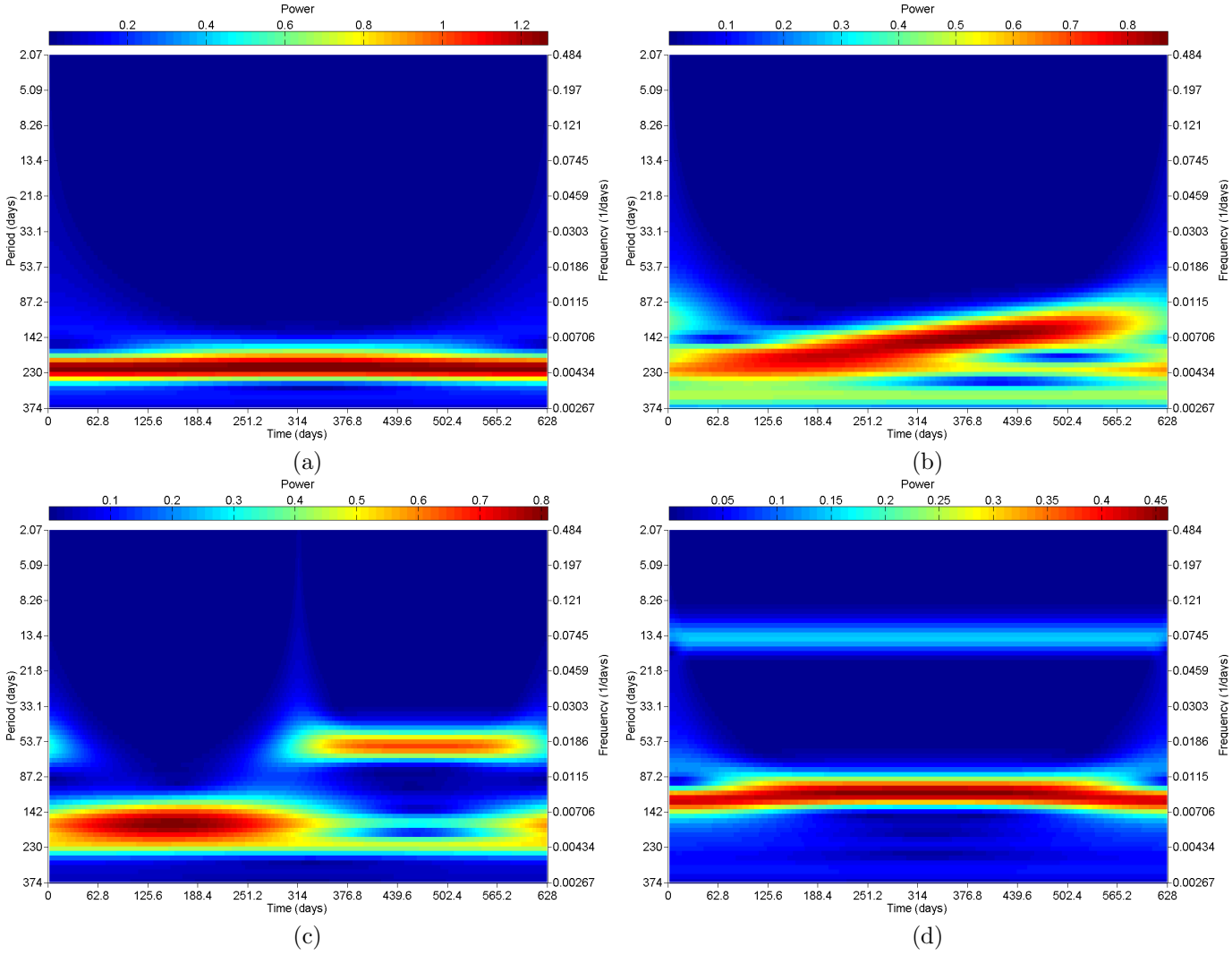


Figure 12: Continuous Wavelet Transforms of the test signals from Figure 7.

The wavelet for lightcurve A is very simple to analyse. A strong band of power is present around the correct frequency of $5.00 \times 10^{-3}\text{days}^{-1}$. The fact that it stretches across the whole time indicates the signal is present throughout, which is correct. The power band is fairly wide, so recovering the exact frequency present is challenging.

Lightcurve B produces a very interesting wavelet. There is a clear upward trend to the band of power, showing the increase in frequency over the course of the signal. It also stars and ends around the correct frequencies, clearly indicating that the wavelet is an appropriate tool for tracking signals that change with time. Again, the wide power band makes the precise detection of signals difficult.

The wavelet's ability to present both frequency and time information simultaneously is also important for lightcurve C. There are areas of high power present in each half of the signal, showing the correct frequency and enabling us to detect the abrupt change.

Lightcurve D shows two separate bands of power at the correct frequencies, across the whole time period. This shows it can cope with superposed signals. The greater power of the band for the lower frequency is to be expected, as the lower frequency signal within the artificial lightcurve is the more powerful of the two. The marked difference between the wavelets for lightcurves C and D, especially considering the similarity of their periodograms, shows the value of the wavelet in interpreting lightcurves containing multiple frequencies.

5.4 Phase Dispersion Minimisation

5.4.1 Introduction

Another common technique for identifying periods within signals is Phase Dispersion Minimisation (PDM) (Stellingwerf, 1978). This involves 'folding' the lightcurve around a trial period. To achieve this, define a phase vector ϕ, given by

$$\phi = t \bmod (\Pi) \tag{7}$$

where Π is the trial period. Plotting the lightcurve against ϕ rather than t will yield a 'folded' lightcurve. The folded lightcurve is then sampled by selecting all the points within a small window in ϕ.

Consider a set of N measurements, where the j^{th} measurement is given by (X_i, t_j). The variance of the whole sample is given by

$$\sigma^2 = \frac{\sum (X_j - \bar{x})^2}{N - 1} \tag{8}$$

where $\bar{x}$ is the mean of the measurements. Each selected window has a variance s_j, and M windows have been selected. Not all points need to be selected, and a point can belong to more than one window. A common choice, also used here, is to fit fixed windows across the whole folded lightcurve. The overall variance for all the samples is given by

$$s^2 = \frac{\sum (n_j - 1) s_j^2}{\sum n_j - M}. \tag{9}$$

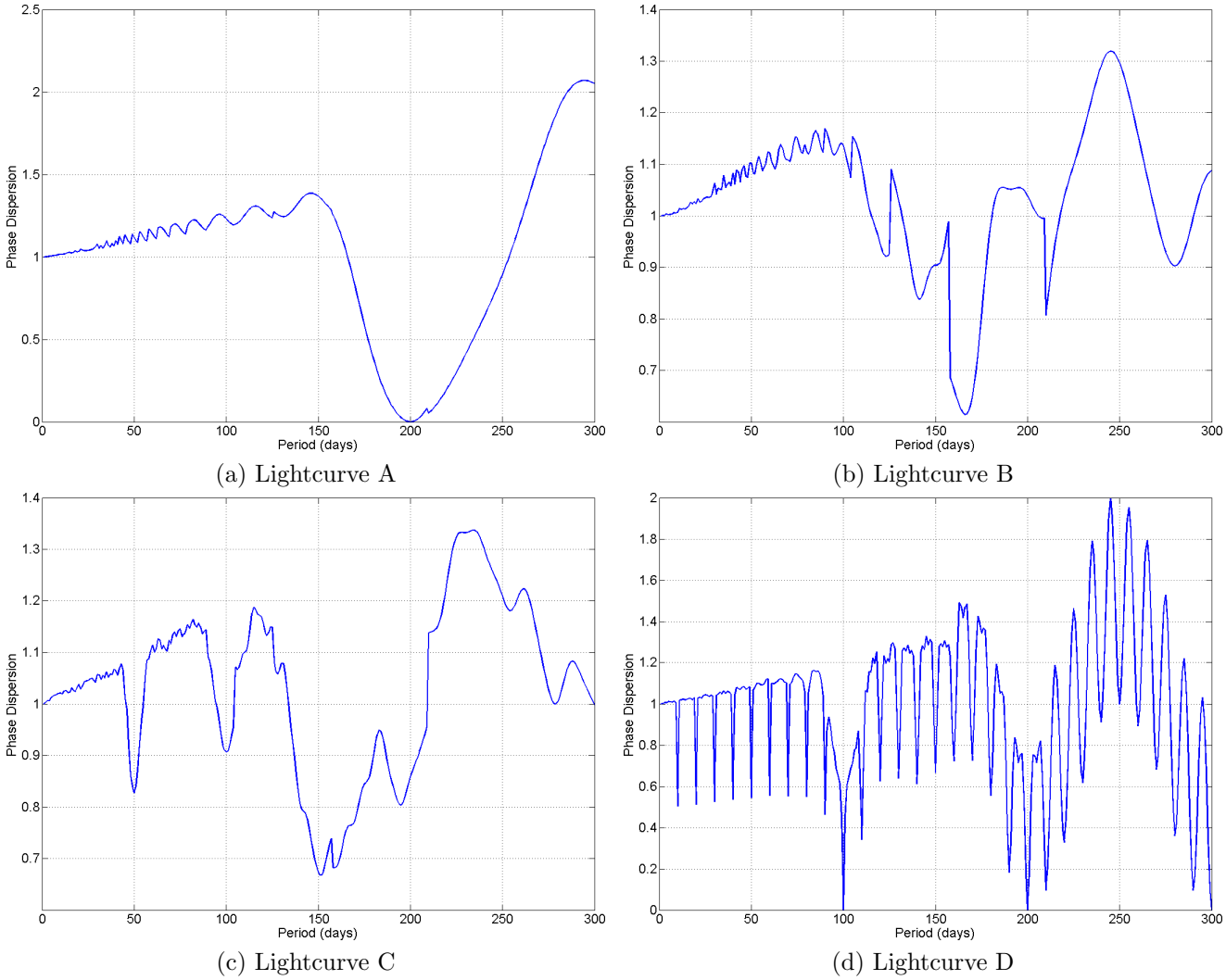
If our chosen value of the trial period Π does not match any periods within the data, we expect $s^2 \approx \sigma^2$, but we have found an existing period, s^2 should be much smaller (Stellingwerf, 1978). We therefore define the quantity θ as

$$\theta = \frac{s^2}{\sigma^2} \tag{10}$$

and look for the minima when θ is plotted against Π to locate periods within the data.

5.4.2 Examples

Figure 13 shows plots of the phase dispersion, θ against the trial period, Π, for the artificial lightcurves shown in Figure 7.



(a) Lightcurve A

(b) Lightcurve B

(c) Lightcurve C

(d) Lightcurve D

Figure 13: Phase Dispersion Minimisation plots of the test signals from Figure 7.

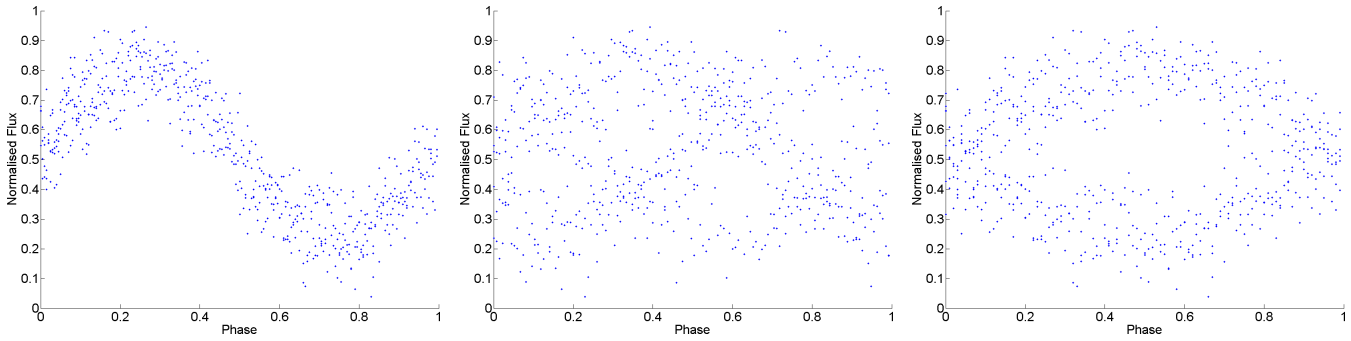
Lightcurve A, the simple sinusoid, performs adequately. The correct period is identified by a minimum at 200 days. However, the trough is extremely wide, considering the lightcurve is a perfect sinusoid with only one frequency and no noise.

The changing frequency of lightcurve B also presents problems for the PDM. There is no clear location of the frequencies within the signal, and a fairly deep and narrow trough is present, which would seem to indicate the presence of a specific period, although none exists within the lightcurve.

Performance is marginally better for lightcurve C. There is a fairly narrow trough at the 50 day period, and a very wide one at the 157 day period. Aliasing is clearly a problem, as the trough at 100 days is every bit as convincing as the correct one at 50. This also manifests in the 'double dip' of the wide peak. One is produced by the genuine period at 157 days, the other at 150 days is another alias. Without the prior knowledge of the signal that benefits us, identifying the correct period would be difficult. As with the

periodograms, the PDM cannot provide information about which signals are significant at different times within the data.

Lightcurve D again demonstrates the problems with aliasing. The period of 10 days is correctly identified, but an alias is produced at every multiple of 10. There is a sharp trough correctly located at 100 days, but another alias at 200 days. The analysis is not too difficult for simple periods used in this example, but extracting useful information is almost impossible if there are several periods whose aliases drift in and out of phase with each other.



(a) 200 Day Period (b) 135 Day Period (c) 100 Day Period

Figure 14: Three examples of folding artificial lightcurve A (with some Gaussian noise added). The left hand plot shows the folding around the correct period of 200 days, the centre plot shows folding around an incorrect period of 135 days, and the right hand plot shows folding around a 100 day alias.

Figure 14 demonstrates why the PDM suffers so badly from aliasing. It is obvious from visual inspection that the left hand plot has found a period within the signal, since each fold fits neatly onto the one before it, leaving large empty areas within the plot. In contrast, when folded around an incorrect period, as in the centre plot, the data points are spread relatively uniformly over the whole plot. In the right hand plot, the lightcurve is folded around a period of 100 days, which is half of the true period. The figure is much more ordered than other incorrect periods, and so the phase dispersion shows another minimum. This also occurs at multiples of the correct period. For instance, folding around 400 days would also show a clearly ordered lightcurve.

5.5 Information Entropy Minimisation

5.5.1 Introduction

Information Entropy Minimisation (IEM) (Cincotta et al., 1995) is similar in many ways to Phase Dispersion Minimisation. Again, the lightcurve is folded by a trial period. Instead of selecting windows in a given range of ϕ, as in PDM, windows are selected to have a given range in ϕ and a given range in u, where the minimum point of the data is mapped to $u = 0$, and the maximum to $u = 1$. Effectively, the folded lightcurve is split into rectangles rather than into columns.

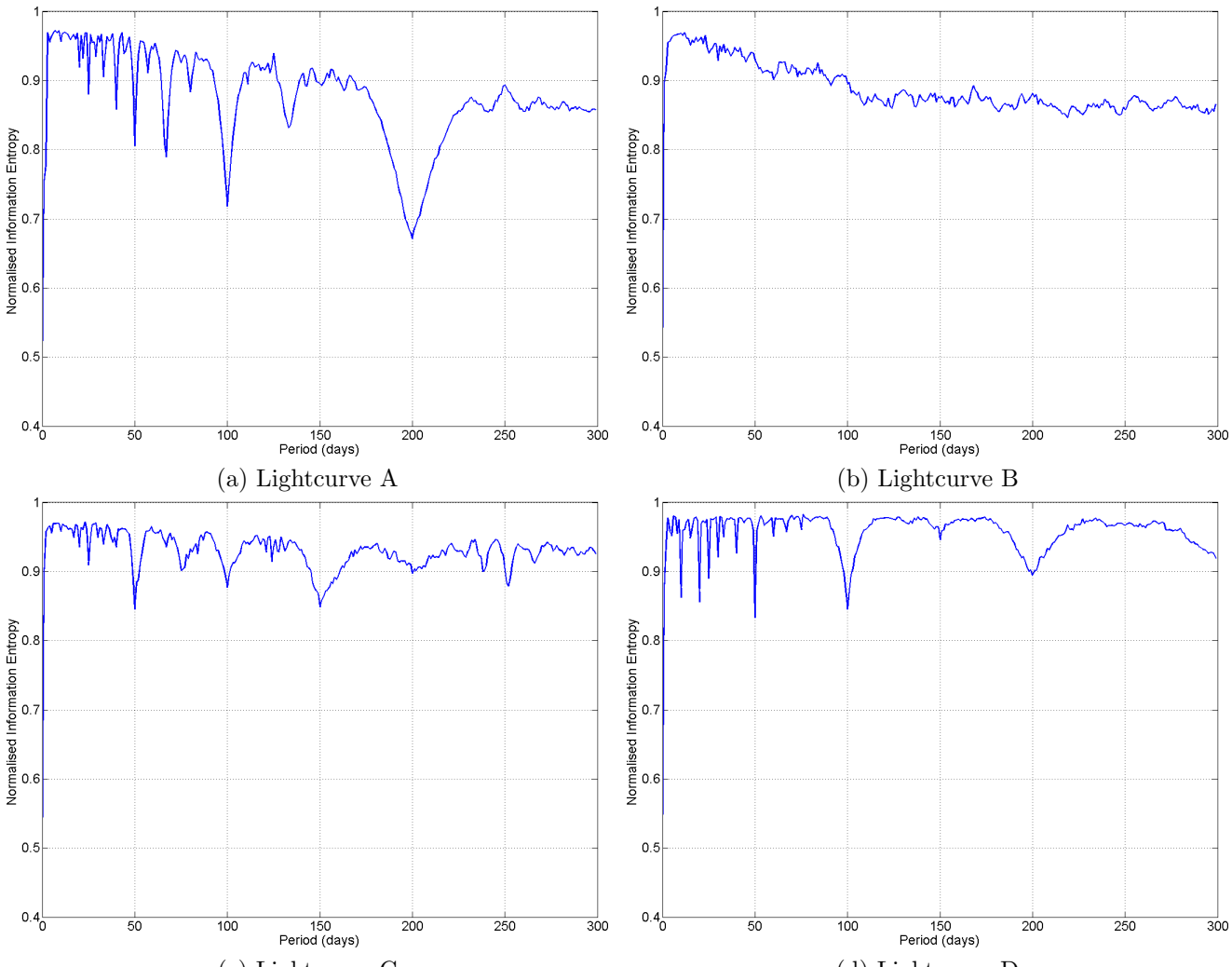
The probability of finding a data point in any given square, μ_i, is simply given by the total number of data points divided by the total number of rectangles. Cincotta et al. (1995) then state that the information entropy is given by

$$S = -\sum_{i=1}^{m} \mu_i \ln(\mu_i) \tag{11}$$

where m is the number of rectangles that the folded lightcurve has been divided into. If the trial period is wrong, the points tend to distribute evenly over the m squares, and so the entropy is high. However, if the trial period is correct, then most of the m rectangles will be empty, while a few will contain many points. This reduces the entropy. In the same way as PDM, IEM identifies correct period by looking for dips in the plot of S against the trial period.

5.5.2 Examples

Figure 15 shows the results of performing Information Entropy Minimisation on the artificial lightcurves from Figure 7.



(a) Lightcurve A

(b) Lightcurve B

(c) Lightcurve C

(d) Lightcurve D

Figure 15: Information Entropy Minimisation plots of the test signals from Figure 7.

Lightcurve A gives a fairly typical indication of the performance of the IEM. Much like the PDM, it suffers from significant aliasing. A wide trough is present at the correct

period of 200 days, but without prior knowledge it would be difficult to select this one as correct and reject the others as aliases.

The performance of the IEM for Lightcurve B is more encouraging. While there are no obvious troughs in the curve, lightcurve B has a contstantly shifting frequency so it is not unexpected that no specific period is found.

Lightcurves C and D show very similar performance to lightcurve A. There are troughs in the correct positions but they are no deeper or sharper than the many aliases, so it is difficult to correctly identify periods without prior knowledge of the signal. As with all methods discussed here except the Continuous Wavelet Transform, the IEM can provide no information about when periods within the signal are significant.

It should be noted that Cincotta et al. (1995) shows examples of using Information Entropy Minimisation that seem to provide much better performance than that shown above. Periods are located by very narrow dips, and the correct period was much deeper than its aliases. Despite the author's best efforts, these levels of performance proved impossible to reproduce.

5.6 Conclusions

Since finding periods within signals is a common requirement within physics, there are many methods available to use in the analysis of *Kepler* data. Since *Kepler* data is high cadence, there are a significant number of data points available for each target star. As a result, ensuring sensible computation times requires some selectivity in the methods employed, rather than simply applying every method available to the entire data set.

The first conclusion that was reached was that there was no advantage gained by using the Generalised Lomb-Scargle Periodogram instead of the ordinary Lomb-Scargle Periodogram (equation 2).There is almost no difference between the results of the two methods, mainly because the errors in the measurements are all very similar, so weighting points according to their error makes very little difference. In addition, a fast code for performing an ordinary Lomb-Scargle periodogram was available from Press et al. (1992). Computing the Generalised Lomb-Scargle periodogram took significantly longer, and it was decided that the small performance improvement did not justify the extra computation time.

The Continuous Wavelet Transform is also capable of providing useful information about target stars. It is generally harder to recover a precise measurement of a period than using a periodogram, but its performance on the artificial lightcurves (see Figures 7 and 12) demonstrate its superiority for analysing changing periods. This is because, unlike the periodogram, it can provide information in both the frequency and time domains.

The other methods studied, namely Phase Dispersion Minimisation and Information Entropy Minimisation, have many similarities. They are very different to the periodograms and wavelet transform, which are both Fourier-based methods. Unfortunately, their performance was poor and it was impossible to distinguish significant periods from their aliases.

All these factors led to the generation of a fairly standard approach for finding periods. The lightcurves were analysed using a Lomb-Scargle periodogram and a Continuous Wavelet Transform. The former provides specific measurements with errors, while the time-domain information provided by the latter is extremely valuable.

6 Solar Data & Analysis

6.1 Motivation for studying Solar Data

Before starting to work with light curves from the *Kepler* mission, it is important to spend some time considering the the performance of the chosen methods on actual data, rather than the idealised signals used in Section 5. The Sun represents a useful test case. Due to its proximity, its characteristics are well known. By restricting ourselves to basic irradiance data of the same type that will be available for Kepler targets, we can test the accuracy of our analysis methods before subjecting them to our main data.

6.2 Sources of Solar Data

Three sources of data were used for the Total Solar Irradiance (TSI). The first was the PMOD composite[7] (see e.g. Fröhlich (2006)), selected because it provides data over a significant timescale, dating back to 1978. This is beneficial as it allows study over the entire solar cycle. It also enables searches for the well known 11-year solar cycle as a test of the performance of various methods.

The second set of data comes entirely the VIRGO instrument on the SoHO spacecraft[8] (see e.g Fröhlich et al. (1997)). This data is especially useful because it is high cadence, similar to the data from the *Kepler* mission. This makes it relevant to the study of achieving good results without requiring excessive computation time.

The final source of data was from the Total Irradiance Monitor (TIM)[9] (see e.g. Kopp and Lawrence (2005)). This again is daily cadence data with an extremely high signal to noise ratio.

Figure 16 shows the three data sets used in the analysis of the Sun. It very clearly shows the well known eleven year solar cycle over which sunspot numbers (among other things) vary. It is particularly visible in the PMOD data (red) due to the greater timescale available. Both other sets of data are clearly a similar shape to the PMOD lightcurve. Despite being the same shape, the values are clearly displaced vertically from each other. This is a result of being measured by different instruments. For the purposes of this discussion, this difference does not matter, as it is the changes in the value which are of interest.

6.3 The Solar Cycle

A useful first test of our methods is to try and recover the well known 11 year solar cycle. This is clearly visible to the eye in Figure 16 so it should appear very strongly in a periodogram or wavelet of the data.

Figure 17 shows a Lomb-Scargle Periodogram for the full PMOD composite data set. There is a very clear peak at a frequency of $(2.49 \pm 0.38 \times 10^{-4})\text{day}^{-1}$, which corresponds to 4014^{+732}_{-536} days, or $11.00^{+2.01}_{-1.47}$ years. The Lomb-Scargle periodogram has easily recovered the solar cycle, although the uncertainties present are significant. There are mitigating circumstances, however. It is clear from a visual inspection of the PMOD Composite

[7]Data (version d41_62_1111) from PMOD/WRC, Davos, Switzerland. Unpublished data from the VIRGO Experiment on the cooperative ESA/NASA Mission SoHO

[8]Data (version6_002_1102) from VIRGO Team through PMOD/WRC, Davos, Switzerland. Unpublished data from the VIRGO Experiment on the cooperative ESA/NASA Mission SoHO

[9]Data obtained from http://lasp.colorado.edu/sorce/data/data_product_summary.htm

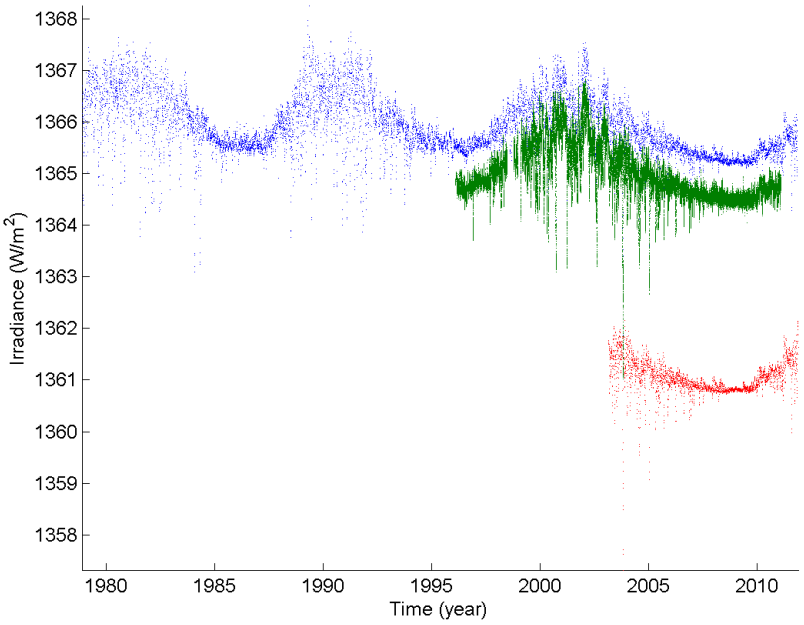


Figure 16: Plot showing the three sets of solar data used to test methods prior to their application to *Kepler* data: the PMOD composite (blue), high cadence data from VIRGO (green), and daily values from the Total Irradiance Monitor (red).

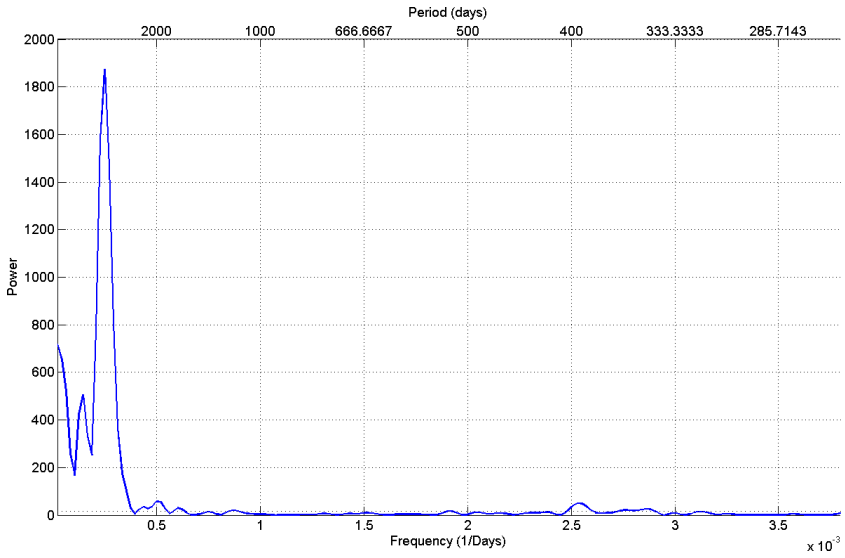


Figure 17: A Lomb-Scargle Periodogram of the full PMOD Composite Solar data set.

data in Figure 16 that the cycle decreases in frequency across the data available here. Indeed, at one point an attempt was made to remove the solar cycle from the data (with the aim of detecting other frequencies more easily), but it proved impossible because of this frequency change. It is unsurprising that a changing frequency results in greater uncertainties in the measurement.

Figure 18 shows the continuous wavelet transform of the full PMOD Composite data set. There are numerous results of interest within the image. Firstly, there is a very

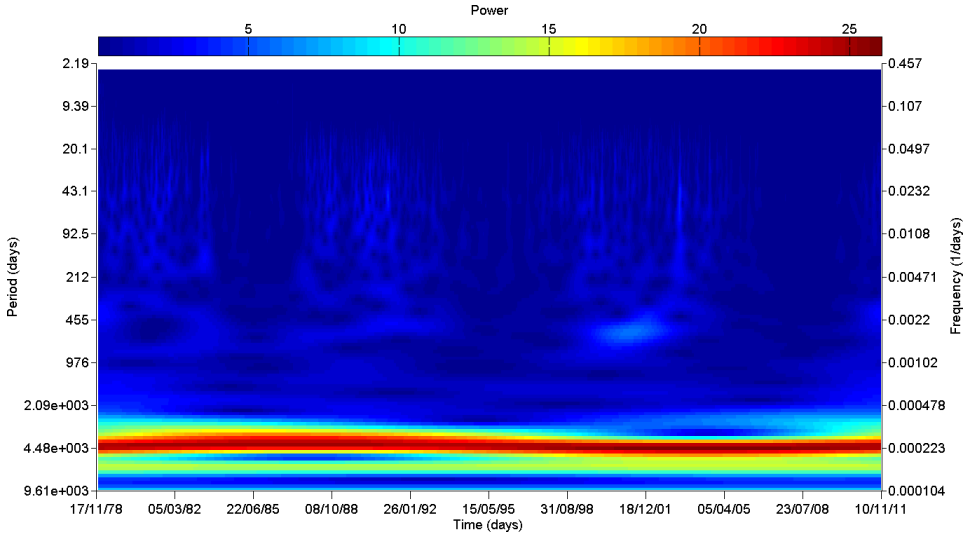


Figure 18: A Continuous Wavelet Transform of the full PMOD Composite Solar data set.

strong period of around 4000 days (about 11 years) for the length of the data set. This slopes downwards slightly, indicating the cycle is getting longer, supporting the earlier statements.. Finally, there is some power between around 10 and around 400 days in three distinct regions throughout the data. Again, comparison to Figure 16 reveals that these correspond to the Solar maxima; the times with the most variability. In contrast, the solar minima show no variability on these timescales.

6.4 Rotation Period

When looking at Figure 18, there is little evidence for a signal of around 27.3 days corresponding to the Carrington rotation period (the rotation period of the Sun when viewed from Earth) (Javaraiah and Gokhale, 2002, p31). There is some power visible during solar maxima, but it occurs over a wide range of periods. Equally, there is no meaningful spike in the periodogram in Figure 17 (although the axis is not long enough to show a spike at this frequency, rest assured there is nothing there). Inability to locate the rotation period on the Sun would be a worrying limitation for the analysis of *Kepler* data.

With some experimentation, it was discovered that the periodogram performs well when the length of the data set is of the order of a few times the desired period. Figure 17 has a data length of the order of 400 times the Carrington rotation period, so failing to find it is not unexpected. In contrast, it is approximately three times the length of the eleven year solar cycle, which is found very strongly.

In order to try and recover the solar rotation period, a different approach was used. The following analysis used the data from the Total Irradiance Monitor. This data set was chosen because it has a timespan that is more than sufficient, and has an extremely high signal no noise ratio. The data set was split up into windows of 80 days each, and a Lomb-Scargle Periodogram performed on each. The significant frequencies in each window were recorded. These were plotted on a histogram, weighted by power (i.e. a frequency in one window with a power of 1000 will appear the same as five different windows containing the same frequency with a power of 200 each time). This is shown in Figure 19. A frequency was classified as significant if it was the frequency with the greatest power in its window, and the power was sufficient that the false alarm probability was less than 0.01.

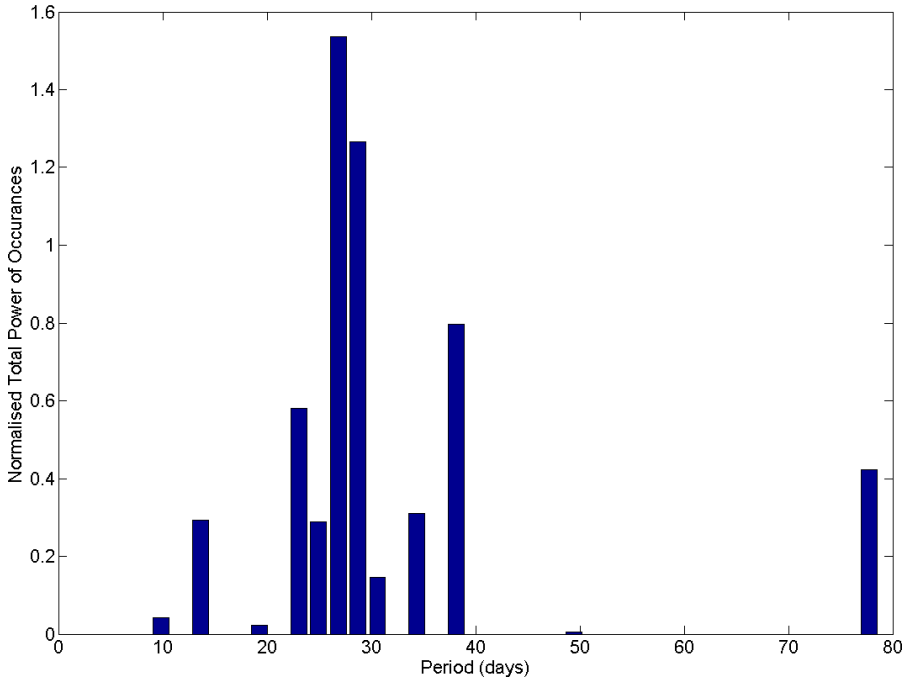


Figure 19: A power-weighted histogram of the most significant frequency in all the 80 days windows in the data from the Total Irradiance Monitor.

It is clear that the most power across all the 80 day windows is around the 27 day mark, so by using 80 day windows is it possible to recover the solar rotation period. However, some windows contained a strongest frequency that was wildly different to this, ranging from 10 to 80 days. Since only a few windows of this length will be available for *Kepler* targets, the histogram approach is unnecessary; periodograms for the different windows can be compared directly. This does demonstrate that it will be difficult to be certain of the rotation period of a target star based on 4 windows, as this would certainly not be enough in the case of the Sun. However, many target stars may show greater irradiance changes due to rotation, which would make the rotation period easier to identify. The presence of large spots would be one feature that could have this effect.

7 Kepler Data Discussion

7.1 Available Data

The data available to use consists of 125 targets that were selected for the *Kepler* Guest Observer Programme[10]. They were selected as they were thought to be part of a cluster, and so would all be of the same age. It was later discovered that the cluster is extremely disperse, to the extent that any results can only be assumed to be true for field stars. The available target stars fall between a right ascension of 300.7° and 301.4°, and a declination of 44° and 44.3°.

Steven Bloemen provided data for *Kepler* quarters 6, 7, 8 and 9, generated from his mask algorithm (see Section 7.2. This data ranges from September 2010 to August 2011.

[10]http://keplergo.arc.nasa.gov/GOprogramScope.shtml

Each quarter is approximately 90 days in length, but quarter 8 is closer to 70. Readings are taken every 30 minutes.

It is important to note the limitations of the *Kepler* data. Some quarters are different lengths, and some have fairly significant gaps between them. For these reasons, it was decided to apply the chosen techniques to each quarter individually, rather than attempting to concatenate them into a single lightcurve. It is also possible that concatenation would reduce the effectiveness of the methods for the same reason that made the solar rotation period difficult to find. When the length of the data set is orders of magnitude larger than a period within the data, it becomes very difficult to identify it, as longer term signals (whether physical or systematics) tend to dominate.

This has some implications for the analysis of *Kepler* lightcurves in Section **??** that should be highlighted. The most important of these is that the figures for different quarters stand alone with regard to normalisation. For instance, the lightcurves are plotted having been median normalised, which allows the reader to easily see proportional changes in the brightness of the target. Since the quarters are treated separately, each quarter is normalised to the median of that quarter, not to the median of the whole lightcurve. A similar limitation exists within the continuous wavelet transforms. The colour represents the power, but only in comparison to the rest of that quarter. It is not possible to track the changing power of a signal across multiple quarters.

While the fact that *Kepler* obervses its targets from space is a necessity for the requirements of the mission, it does introduce problems not faced by ground-based observatories. There is a chance that the telescope's field of view will drift slightly, and require the mission to make course corrections. These can manifest in the data as jumps or dips as light from a target star is detected by different pixels. Another issue is raised by the need to return data to Earth. Data downlinks take place every 30 days and result in a small gap in the observations. A detailed discussion of instrument noise is given by Gilliland et al. (2011).

7.2 Masks

7.2.1 Lightcurve production

The *Kepler* Telescope detects light from stars using Charged Coupled Detectors, or CCDs. Photons hitting the detector cause electrons to be released, which can then be detected. The number of electrons released is directly proportional to the number of photons that hit the detector (as long as the device isn't saturated), making measurements of flux very straightforward.

CCDs contain arrays of pixels, each detecting the photons that hit them. Since these pixels are small, photons from target stars spread out over many different pixels. Lightcurves are then produced by summing the photon count across many pixels. Selecting which pixels to include, and which to reject, is important for the production of good lightcurves. The pixels selected for any given target are referred to as the mask. A description of the *Kepler* data pipeline is provided by Jenkins et al. (2010).

Figure 20 shows the mask for a target star, kplr008197767 (quarter 6). While it is clear that each pixel has its own trends, the variation around this trend is consistent across all the pixels. Summing the pixels together tends to reduce the trend, and linear trends are not detected by the periodogram or wavelet, so there is no need to remove them.

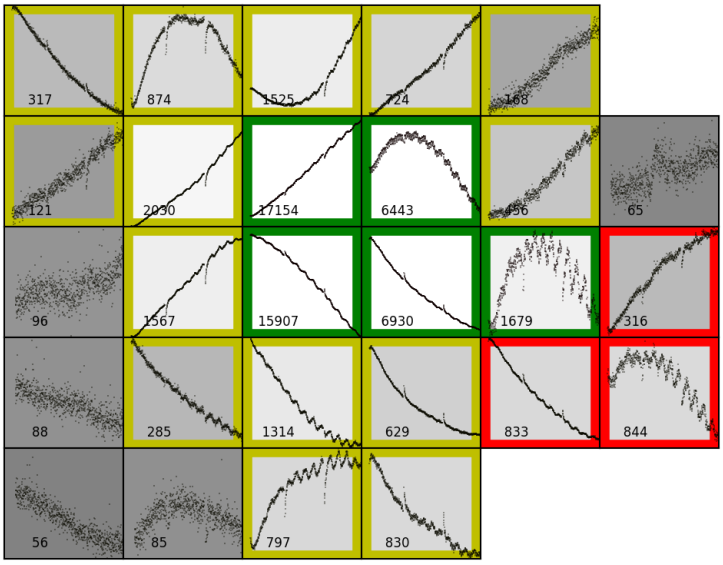


Figure 20: The mask for quarter 6 of kplr00197767. The colour of the pixel boundary indicates when it was used. Green: used by the *Kepler* team and by Steven Bloemen. Red or no border: Rejected by the *Kepler* team and by Steven Bloemen. Blue: used by the *Kepler* team but rejected by Steven Bloemen.

7.2.2 The Importance of Mask Selection

Selecting the right pixels to use in the mask is extremely important. Any variability in a target star should be visible across most (if not all) of the pixels in a mask. If this is not the case, it implies that the variation is not intrinsic to the star, but is being introducted somewhere else. Figure 21 is a good example of this.

Figure 21 shows the lightcurve for quarter 6 of kplr008263752. There are very sharp and regular dips within the lightcurve, which were initially considered as candidates for planetary transits, or a binary star system. However, when the mask was inspected (see Figure 22), there was no evidence for these dips in the centre of the mask, where the flux is greatest. The only pixels where these dips are visible are located on the right hand side of the mask. Any sort of transit, or any variability intrinsic to the star, would be visible in all the pixels. The conclusion is that the dips are some sort of instrumental error.

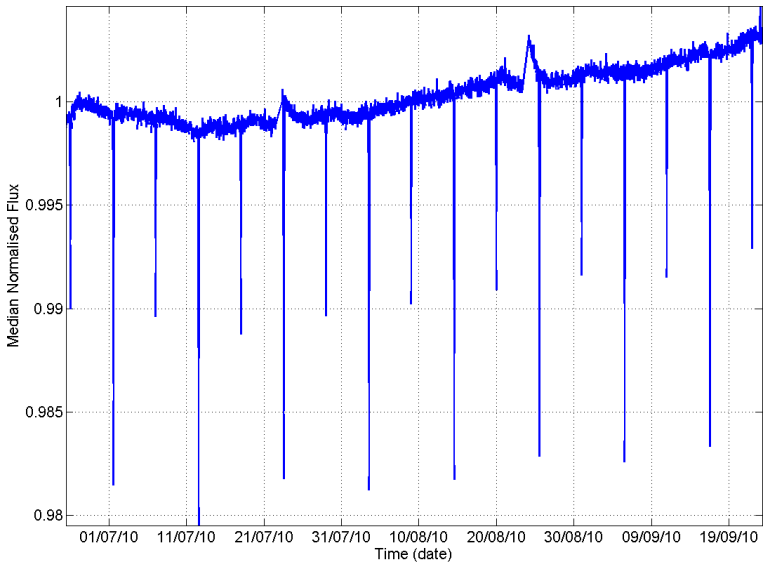


Figure 21: The lightcurve for quarter 6 of target kplr008263752, showing intruiging dips regularly spaced throughout the lightcurve.

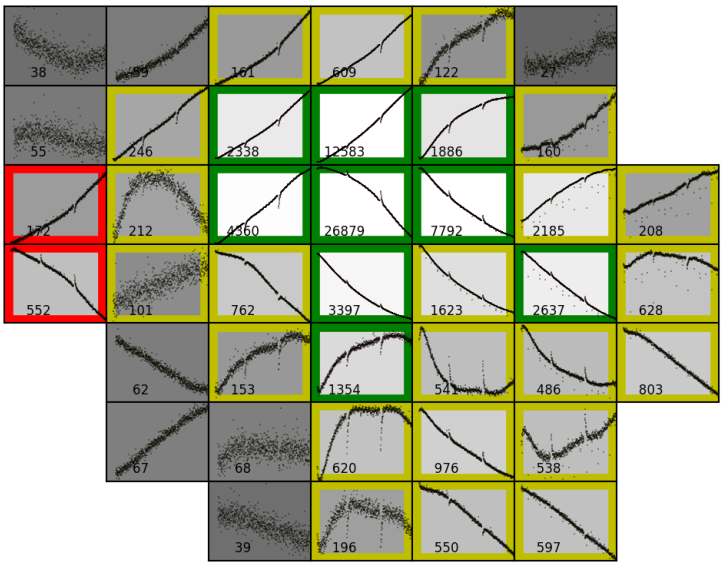


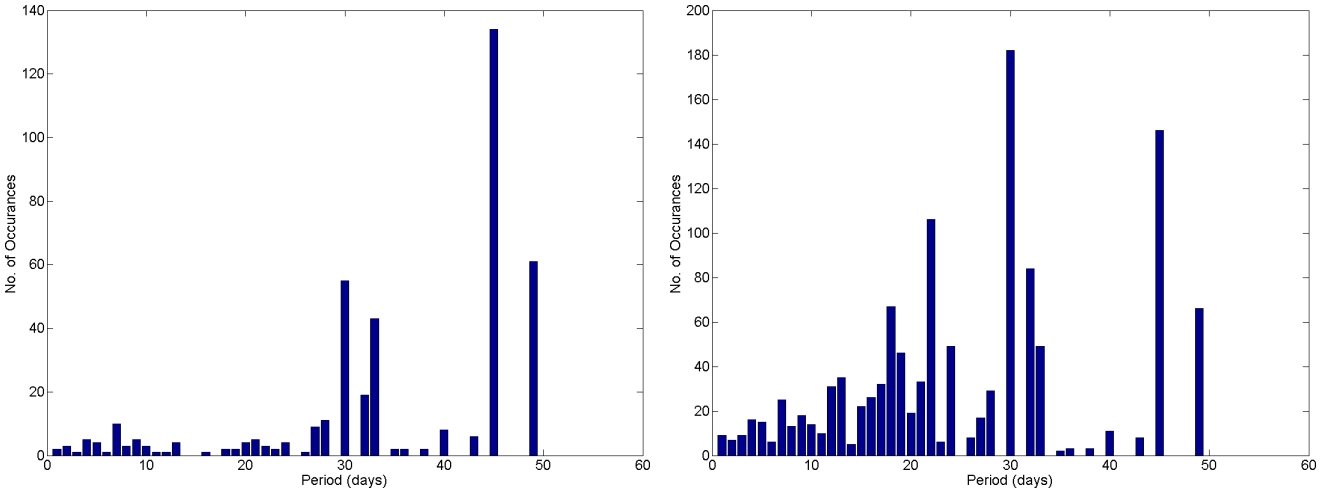
Figure 22: The mask for quarter 6 of target kplr008263752. The dips are clear in pixels to the right hand side of the mask, but there is no evidence of them elsewhere.

8 Kepler Results

8.1 Abundance of different variability types

One of the main aims of this project was to gain an understanding of the distribution of timescales of variability across the data set. To achieve this in an automated fashion, the Lomb-Scargle periodogram was employed, as it allows specific periods to be returned. While this is a fairly simplistic approach (as will be seen in Sections 8.2 and 8.3, the simple number returned from the periodogram often hides important information that can be detected using other methods), it is sufficient for the present purpose. It is far more important is to ensure that the signals detected are physical, not systematic trends.

The basic approach taken involved performing a Lomb-Scargle periodogram of each quarter of *Kepler* data, across all 125 targets available, and record the locations of the peaks with periods of less than 50 days. This was done in an attempt to remove long term systematics which often produce large peaks in the periodogram. Figure 23 shows the distributions of the most powerful period in each quarter of data (left hand plot) and the three most powerful periods in each quarter (right hand plot).



(a) Top one period (b) Top three periods

Figure 23: Distribution of Most Powerful Periods from the Lomb-Scargle Periodogram for each quarter of data for all 125 targets. The left hand plot shows the most powerful period in each quarter, the right hand plot shows the three most powerful periods in each quarter.

Turning first to the single most powerful period in each quarter, and the distribution is very uneven. Long periods (30 days and over) dominate the distribution, but even here the distribution is uneven, with a few periods with many occurances and many periods with none. This would suggest they are the result of systematics, rather than genuine variability within the target stars. Indeed, two of the common frequencies are at 45 and 30 days - one half and one third of the length of each quarter. The 30 day period is likely to correspond to the data downlinks, and the 45 day period could equally derive from a regular *Kepler* activity.

If the search is broadened to the top three periods within each quarter, the distribution looks much more encouraging. Between 1 and 20 days the distribution is reasonably consistent, with no periods that never occur, and no periods that occur very regularly. In

contrast, the distribution above 30 days remains very uneven. This is a good indication that genuine variability is being detected below 20 days, while the majority of periods above 30 days appear to be systematic. This also provides justification for discarding periods greater than 50 days.

This result would suggest that the main source of variation being detected is due to stellar rotation, although it is possible that some of the shortest frequencies are the result of pulsations. This is not a particularly surprising result. Excluding the eleven year solar cycle (which would be difficult to detect in comparable data for the Sun), rotation causes the greatest changes in the brightness of the Sun.

It would also seem to imply that most of the stars in the data set are younger than the Sun. As discussed in Section 3.3.4, younger stars tend to rotate more rapidly. Most of the variability detected here is less than the solar rotation period of around 25 days (Stenflo, 1990). Note that, unlike in Section 4, the Sun's actual rotation period is the relevent comparison, not the Carrington rotation period.

8.2 kplr008329696

Two specific targets have been chosen for a more detailed analysis. The first of these is kplr008329696. According to the *Kepler* Input Catalogue (see Brown et al. (2011)), this target is located at a right ascension of 300.788° and a declination of 44.245°. Its effective temperature is 5632K and its logG is $4.236 cms^{-2}$. Its estimated radius is $1.323\odot$. All things considered, it is very similar to the Sun.

The lightcurves for kplr008329696 for quarters 6, 7, 8 and 9 are shown in Figure 27.

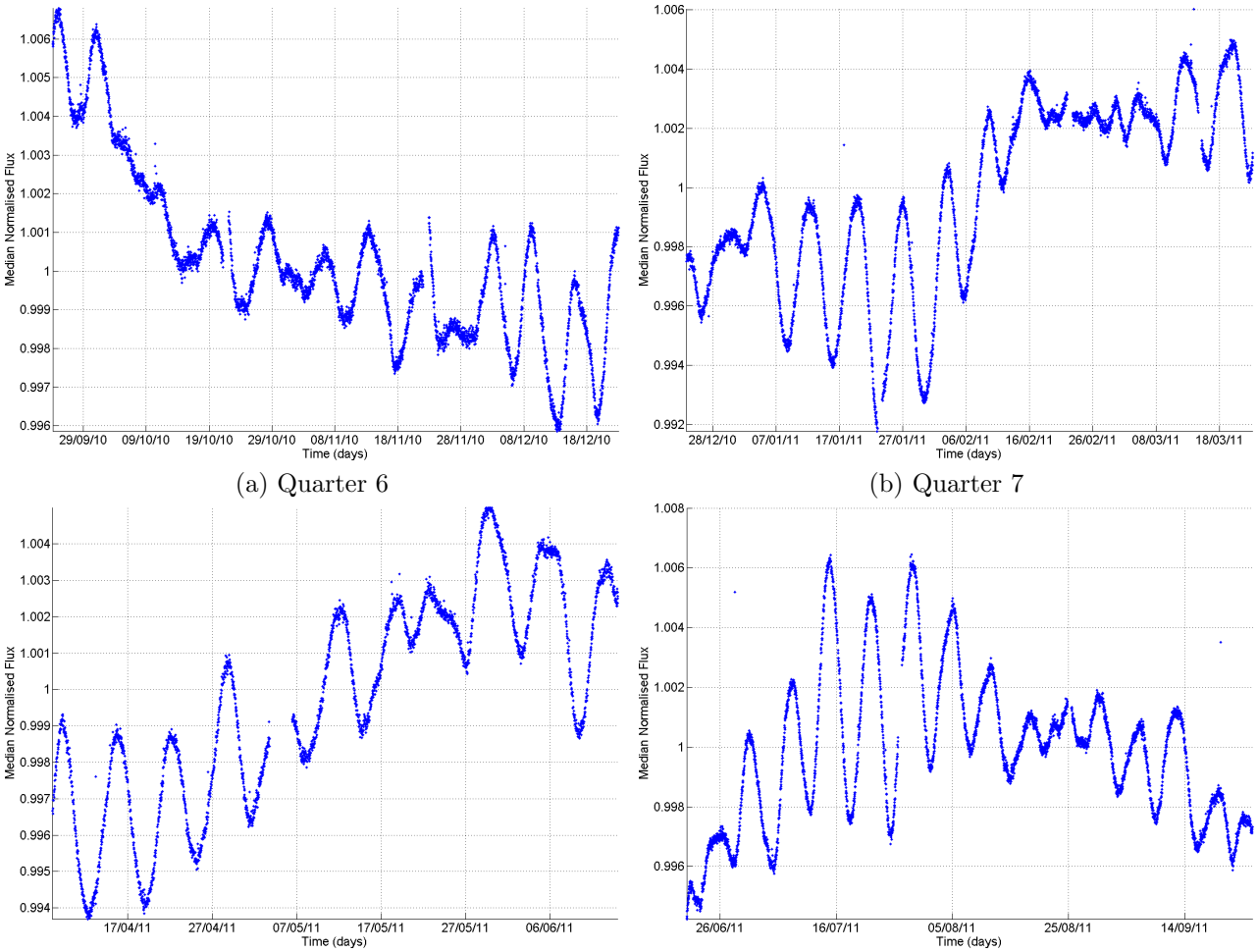
The first thing to notice is the presence of some longer term systematic trends present within the data. No attempt has been made to remove them, as the methods used are able to locate interesting frequencies despite these. The trends themselves can sometimes be detected, but are usually obvious as they have much longer periods .

Figure 25 shows Lomb-Scargle periodograms for each of the available quarters for kplr008329696. Note that although the code used to produce these figures operates in the frequency domain (and hence errors are symmetric here), I will refer to the results found in the time domain, as they are more intuitive. The first thing to notice is that all of the periodograms have a large spike indicating a very long period (ranging from around 50 days to around 130 days). This is very long for a rotation period, and since it is fairly typical of many *Kepler* targets, it is generally rejected as systematic trends. This analysis is supported by the lack of consistency between quarters, while physical signals tend to appear in similar places in several quarters. These systematic trends could stem from a variety of factors, which are discussed in Section 9.

The other area of interest in Figure 25 occurs around the seven day mark. It is especially encouraging that there is activity here in all four quarters of the lightcurve. The actual periods found range from $6.30^{+0.15}_{-0.14}$ days to $8.49^{+0.36}_{-0.33}$ days, but the final three quarters all contain a period of $7.37^{+0.30}_{-0.29}$. The consistency with which this period is found, and the length of the period, make it an excellent candidate for a rotation period.

The periodograms also seem to imply that this period increases in strength throughout the quarters (i.e. the power increases). However, analysis of the Continuous Wavelet Transforms for the same quarters shows this is not really the case, and demonstrates one of the weaknesses of periodogram analysis. These Continuous Wavelet Transforms are shown in Figure 26.

Figure 26 in part supports the results found by the Lomb-Scargle periodograms shown



(a) Quarter 6

(b) Quarter 7

(c) Quarter 8

(d) Quarter 9

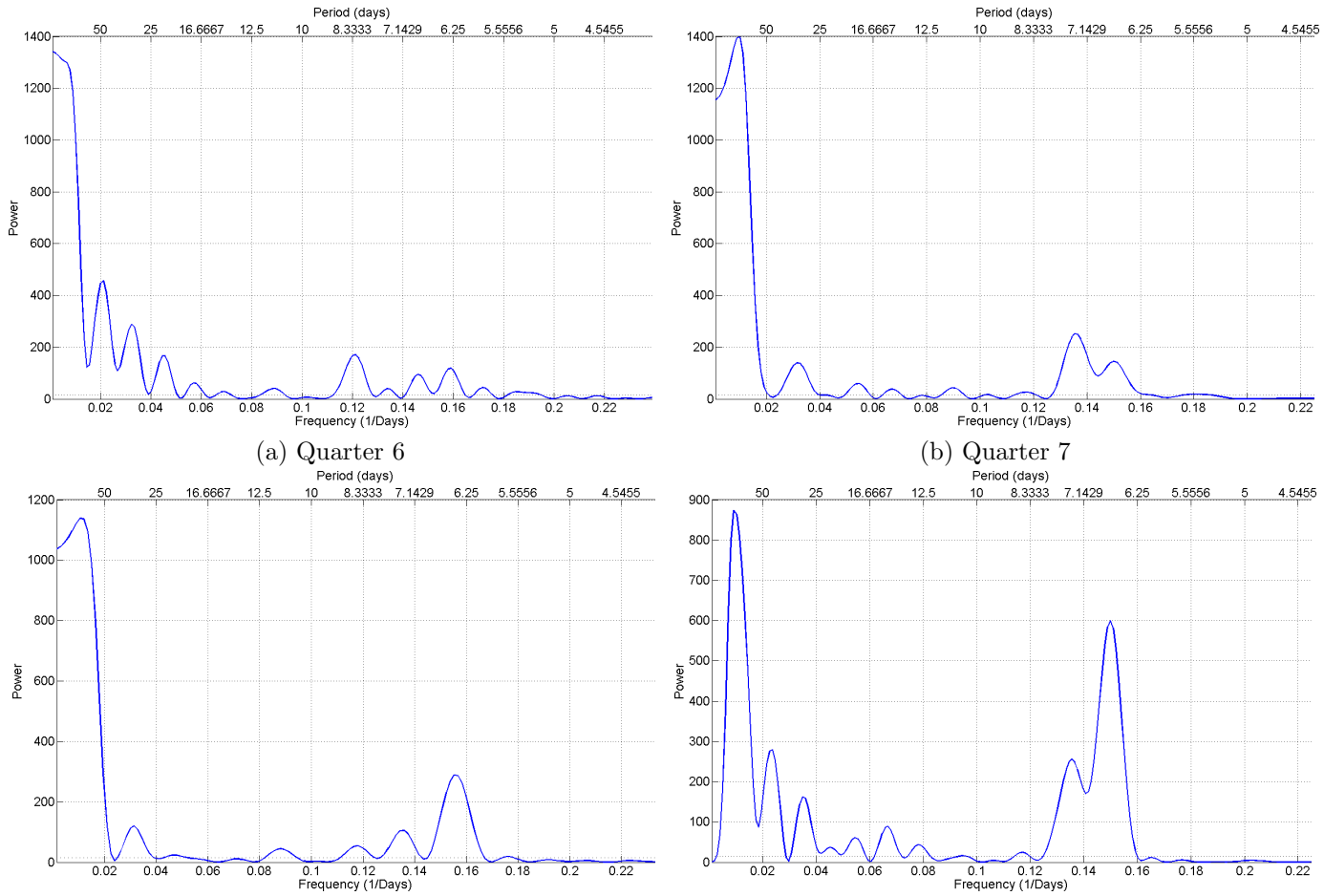
Figure 24: Lomb-Scargle Periodograms for quarters 6, 7, 8 and 9 of target kplr008329696.

in Figure 25. There is certainly power present around the seven day mark in all four quarters. However, the time information also provided by the Continuous Wavelet Transform shows that this period is not constant and unchanging throughout, but instead varies significantly. The regions with greatest power are near the start of quarters nine and eleven.

This result has important consequences for the interpretation of kplr008329696. From periodogram analysis alone, a possible conclusion would be that the star contained a single long-lived active region whose spot coverage increases over the course of the data available here. With the extra time information provided by Wavelet analysis, it is more likely that active regions with lifetimes of a few rotations are present. This presence of many different periods around the seven day mark could be explained by differential rotation within the star, and active regions appearing at slightly different latitudes.

8.3 kplr008264134

Another target selected for further analysis is kplr008264134. This star is again similar to the Sun, with a radius of 1.61$\odot$ and an effective temperature of 4970K. It is located at a right ascension of 300.882° and a declination of 44.133°. Again, these values are taken



(a) Quarter 6

(b) Quarter 7

(c) Quarter 8

(d) Quarter 9

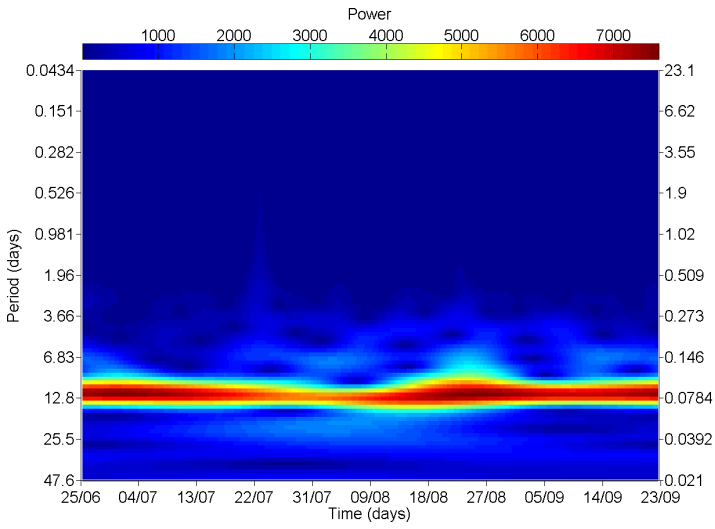
Figure 25: Lomb-Scargle Periodograms for quarters 6, 7, 8 and 9 of target kplr00826413.

from the *Kepler* Input Catalogue (see Brown et al. (2011)). The lightcurves for quarters 7, 8, 9 and 10 of kplr008264134 are shown in Figure 27.

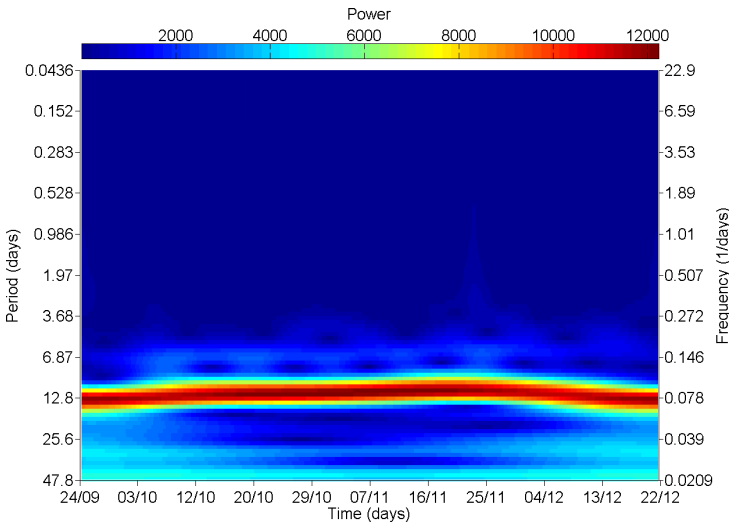
It is clear that the lightcurves for kplr008264134 (Figure 27) are generally less effected by the longer term systematics than those for kplr008329696 (Figure 24). As discussed in Section 7.2, these are a result of trends in the individual CCD pixels. These tend to cancel out to some extent when they are combined into a lightcurve, but some lightcurves cancel them out better than others. The result of this is that, while the most powerful frequency found for kplr008329696 was generally a systematic, this might not be the case for kplr008264134.

Lomb-Scargle periodograms for kplr008264134 are shown in Figure 28. Previous speculation that systematics should be less powerful than physical signals appear to be justified. In quarter 9 especially, the lack of systematic trends produces an exceptionally clear periodogram with very little power anywhere above around 20 days.

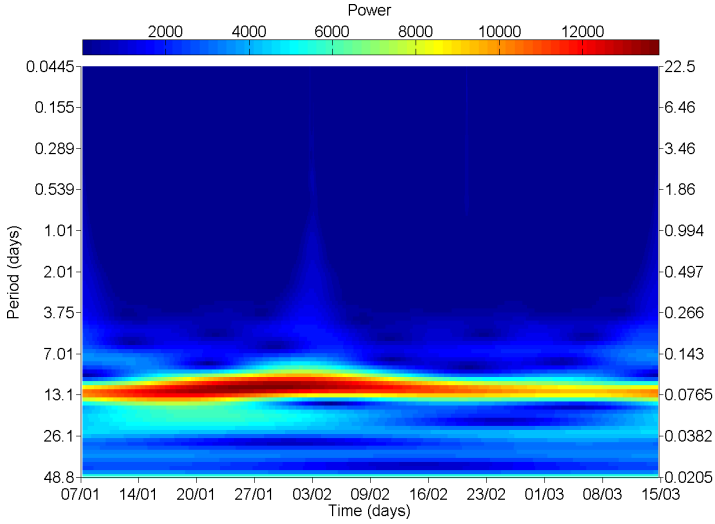
Throughout all four quarters, the periodograms show a very strong period of $11.14^{+0.99}_{-0.85}$ days. The final three quarters also show a much weaker period of $5.58^{+0.37}_{-0.30}$ days. This is supported by a visual inspection of the lightcurve, which suggests there is another period present. This is easiest to see in the unusually shaped peaks and troughs within the more obvious signal. For example, the first three peaks of quarter 9 are all different shapes, which implies the presence of a second signal moving out of phase with the main one.



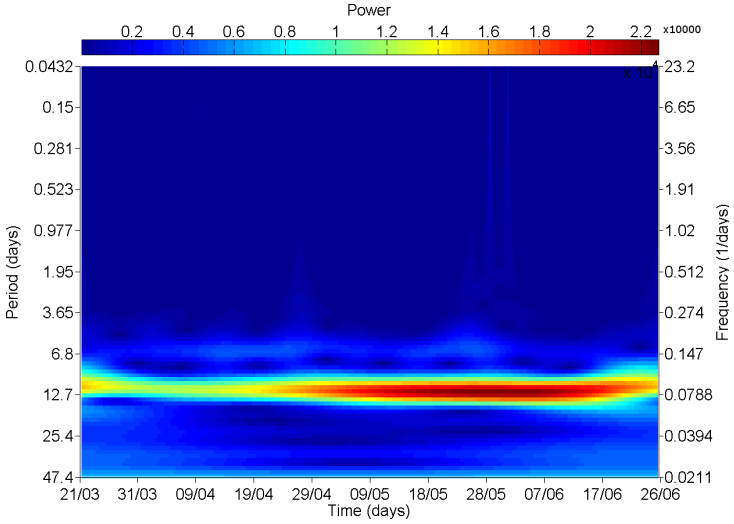
(a) Quarter 6



(b) Quarter 7



(c) Quarter 8



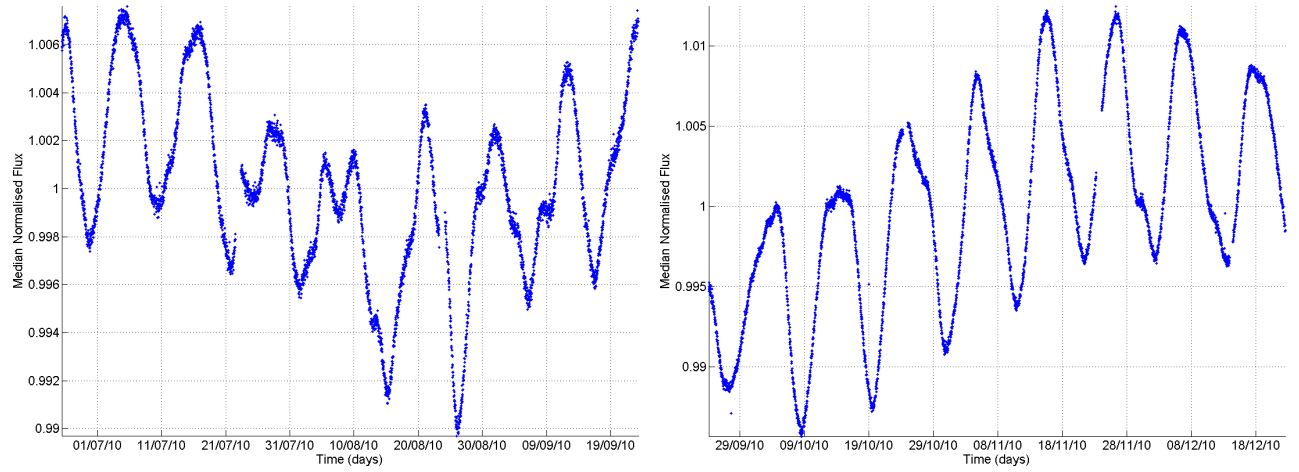
(d) Quarter 9

Figure 26: Continuous Wavelet Transforms for quarters 6, 7, 8 and 9 of target kplr00826413.

The location of this signal highlights the advantage of using mathematical techniques such as the Lomb-Scargle periodogram for this analysis. The strongest frequencies in both kplr008329696 and kplr008264134 are immediately obvious upon visual inspection, and a fairly accurate period could be determined by eye. However, while an observer may identify the presence of a second period within kplr008264134, reliably finding its timescale would be impossible.

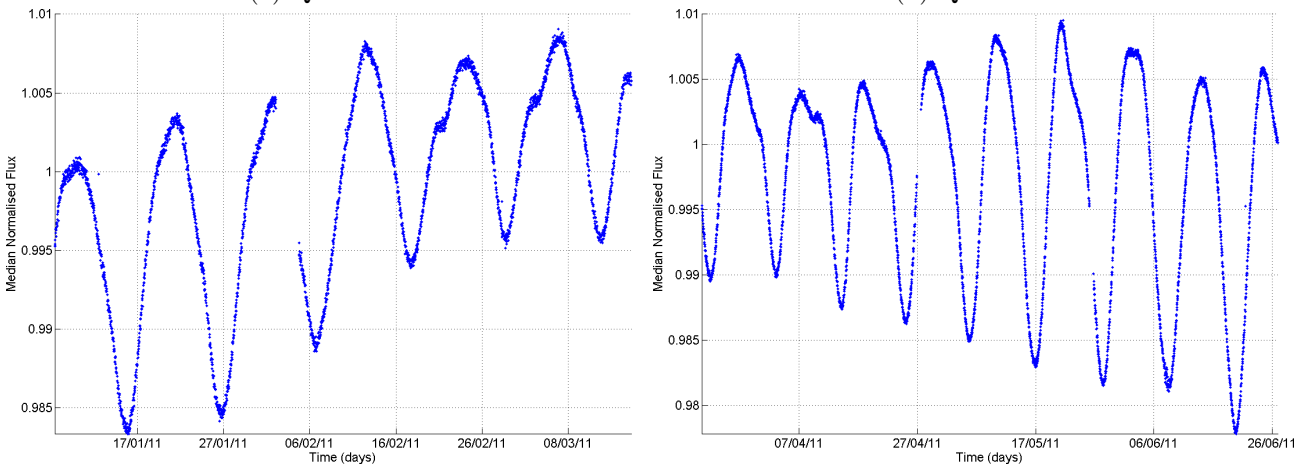
Figure 29 shows the Continuous Wavelet Transforms for the four quarters of kplr008264134. There is a very clear band of power corresponding to the 11.14 day period located by the periodogram. As discussed in Section 7, it is not possible to track the strength of the signal from one quarter to another, but it clearly decreases in strength across quarter 8 and increases again over quarter 9, which is confirmed by a visual inspection.

There is limited evidence to support the 5.58 day period found by the periodogram. There is some power in the correct region of most of the quarters, but it is nothing like as clear as the 11.14 day period. Given the respective strengths of the spikes in the



(a) Quarter 6

(b) Quarter 7



(c) Quarter 8

(d) Quarter 9

Figure 27: Lomb-Scargle Periodograms for quarters 6, 7, 8 and 9 of target kplr008264134.

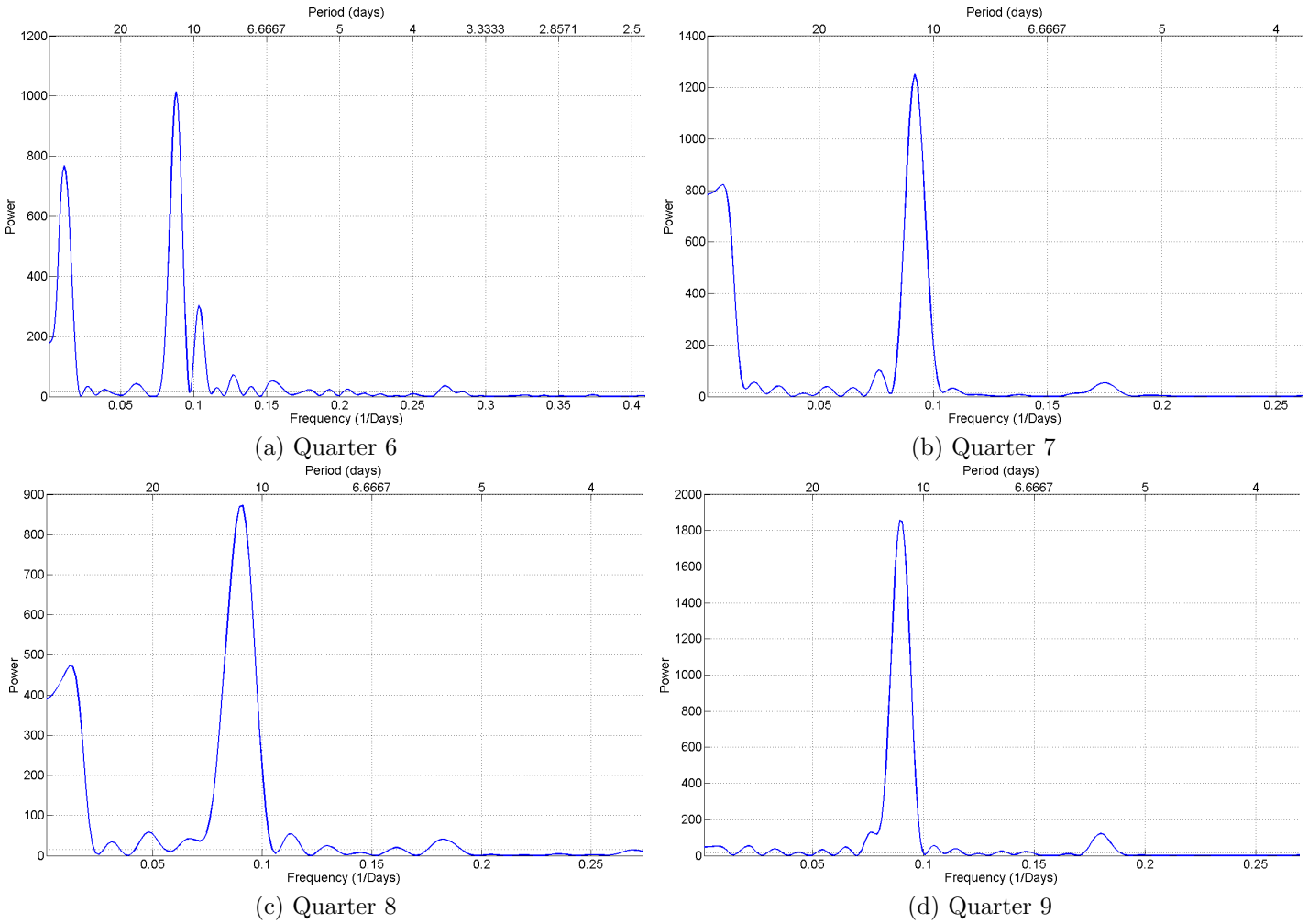
periodograms, this is not a particularly surprising result.

To a level of accuracy far greater than the uncertainties in the periodogram, the shorter period is half the length of the longer one. A possible interpretation of these results is that kplr008264134 is a star with an 11.14 day rotation period and a large active region. If there is a smaller active region on the opposite side of the star, then this could introduce variation on the timescale of half the rotation period, as seems to be the case.

9 Conclusions & Further Work

The main aim of this project was to study stellar microvariability in Sun-like stars using data from the *Kepler* Mission. It was hoped that an understanding could be gained of the timescales of variability and how they are distributed throughout the available target stars.

Initially, various mathematical methods of finding regular periods within data were studied. These included periodogram analysis, Continuous Wavelet Transfroms, Phase Dispersion Minimisation (PDM) and Information Entropy Minimisation (IEM). To test the effectiveness and reliability of these methods, and to highlight any weaknesses, they were tested on a series of artificial lightcurves.



(a) Quarter 6

(b) Quarter 7

(c) Quarter 8

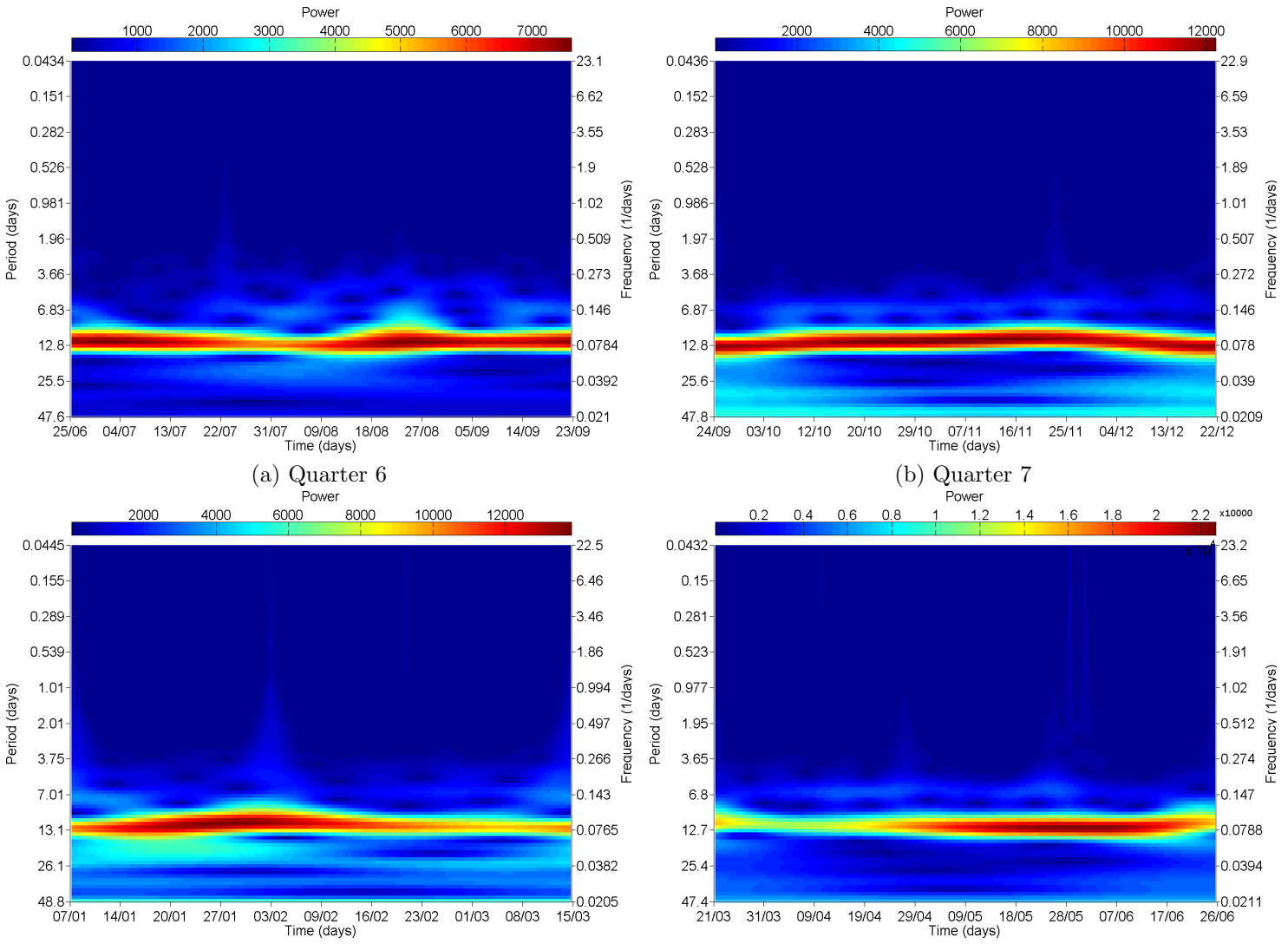
(d) Quarter 9

Figure 28: Lomb-Scargle Periodograms for quarters 6, 7 , 8 and 9 of target kplr00826413.

It was discovered that periodogram analysis was a useful technique. The Lomb-Scargle periodogram (Scargle, 1982) was selected over the more basic classical periodogram because of its improved performance. A further modification, the Generalised Lomb-Scargle periodogram (Zechmeister and Kürster, 2009), was rejected because its marginal performance improvements did not justify the extra computation time. It was also found that fitting Gaussians to the peaks of the periodogram was a suitable method of estimating the uncertainty in located frequencies, while the bootstrap method seemed to drastically underestimate the uncertainties. It was also vastly more compuationally expensive to perform.

The major weakness of the Lomb-Scargle periodogram was the lack of time-domain information provided. This was mitigated by the positive performance of the Continuous Wavelet Transform (Torrence and Compo, 1998), which was able to provide time-domain and frequency-domain information simultaneously, allowing frequencies that change throughout the data to be tracked.

The performances of the PDM and IEM methods were less encouraging. While it was possible to identify some frequencies, it was almost impossible to distinguish between the true frequency and its aliases. However, the folded lightcurves used in this method proved to be a useful visual aid to confirm frequencies found using the periodogram or



(a) Quarter 6

(b) Quarter 7

(c) Quarter 8

(d) Quarter 9

Figure 29: Continuous Wavelet Transforms for quarters 6, 7, 8 and 9 of target kplr00826413.

the wavelet.

The Lomb-Scargle Periodogram and the Continuous Wavelet Transform were applied to data from the Sun to demonstrate their ability to recover known results. The eleven year solar cycle was recovered from the PMOD Composite data, and evidence for the Carrington rotation period was found within data from the Total Irradiance Monitor.

Before applying the Lomb-Scargle Periodogram and the Continuous Wavelet Transform to the *Kepler* data, it was decided to leave the data in its 90 day quarters rather than attempting to concatonate it into a single lightcurve. This was partly due to the prescence of gaps between some of the quarters, but also because it becomes harder to detect frequencies when the data set is orders of magnitude longer than the relevent period, as demonstrated by the inability to find the solar rotation period in the PMOD composite.

When plotting distributions of the most powerful and the three most powerful periods within each quarter of data, it appeared that most periods over 30 days were due to

systematics. There was a fairly even spread of periods between one and twenty days. This suggests that much of the variability detected was due to stellar rotation, although some of the shorter periods may be from pulsations.

Two targets were selected for a more detailed analysis. kplr008329696 is a star slightly larger and hotter than the sun, and it was found that a period of $7.37^{+0.30}_{-0.29}$ was consistently found in three of the four quarters, making it an excellent candidate for the rotation period. Time-domain information from the Continuous Wavelet Transform seemed to suggest that the variation could be caused by large active regions with lifetimes of a few roations.

kplr008264134 is slightly larger and slightly cooler than kplr008329696. It was found to have a strong period of $11.14^{+0.99}_{-0.85}$ and a much weaker one of $5.58^{+0.37}_{-0.30}$, which is half of the strong period. A possible interpretation of this would be one large and one small active region, on opposite sides of the star.

This project could be extended in a number of ways. Firstly, access to a much greater proportion of *Kepler*'s data set would allow the conclusions drawn about the distribution of variability timescales to be tested much more thoroughly. It would also be beneficial to find a reliable way of removing systematics from the data, to allow intrinsic variability to dominate the figures.

Another possible continuation of the project would be to focus more on the interpretation of specific targets. Performing Lightcurve Inversion (e.g. citetharmon2000) on *Kepler* lightcurves would allow surface maps of the star to be generated to test the possible interpretations suggested here. Lightcurve Inversion does not produce unique results, so they must be biased towards more physically likely ones. An ambitious extension would be to try and design this bias to be target specific, based on the knowlege we have about the properties of the star.

Acknowledgements

I would like to thank my supervisor during this project, Dr. Yvonne Unruh. Her insight and enthusiasm have proved invaluable.

I am also deeply endebted to my project partner, Alex Chaushev. He is never lacking in ideas, and without his knowledge of programming the project would have ground to a halt on several occasions. We have shared many late nights coding together, and I have enjoyed most of them!

I am very grateful to Dr. Gaitee Hussain, for her enthusiastic support throughout.

My thanks must also go to Dr. Steven Bloemen, for the use of his data, and C. Torrence and G. Compo for their Matlab wavelet code.

Finally, I must thank my parents, Anne and Dave, and my sister Jane, without whose support I would never have never been in a position to undertake this project. I must also thank my wonderful girlfriend Jenny, for her support and proofreading, but mostly for her patience when I spout jargon at her in an attempt to sound intelligent!

References

S. A. Barnes. Ages for Illustrative Field Stars Using Gyrochronology: Viability, Limitations, and Errors. *Astrophysics Journal*, 669:1167–1189, November 2007. doi: 10.1086/519295.

T. M. Brown, D. W. Latham, M. E. Everett, and G. A. Esquerdo. Kepler Input Catalog: Photometric Calibration and Stellar Classification. *The Astronomical Journal*, 142:112, October 2011. doi: 10.1088/0004-6256/142/4/112.

P. M. Cincotta, M. Mendez, and J. A. Nunez. Astronomical Time Series Analysis. I. A Search for Periodicity Using Information Entropy. *Astrophysics Journal*, 449:231, August 1995. doi: 10.1086/176050.

Thomas F. Coleman and Yuying Li. An interior trust region approach for nonlinear minimization subject to bounds. Technical report, Ithaca, NY, USA, 1993.

A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge, 1997. URL `http://statwww.epfl.ch/davison/BMA/`. ISBN 0-521-57391-2.

M. Farge. Wavelet Transforms and their Applications to Turbulence. *Annual Reviews in Fluid Mechanics*, 24(1):395–458, 1992.

C. Fröhlich. Solar irradiance variability since 1978. *Space Science Reviews*, 125:53–65, 2006. ISSN 0038-6308. URL `http://dx.doi.org/10.1007/s11214-006-9046-5`. 10.1007/s11214-006-9046-5.

C. Fröhlich, D. Crommelynck, C. Wehrli, M. Anklin, S. Dewitte, A. Fichot, W. Finsterle, A. Jimnez, A. Chevalier, and H. Roth. In-flight performance of the virgo solar irradiance instruments on soho. *Solar Physics*, 175:267–286, 1997. ISSN 0038-0938. URL `http://dx.doi.org/10.1023/A:1004929108864`. 10.1023/A:1004929108864.

R. L. Gilliland, W. J. Chaplin, E. W. Dunham, V. S. Argabright, W. J. Borucki, G. Basri, et al. Kepler Mission Stellar and Instrument Noise Properties. *Astrophysical Journal Supplement Series*, 197:6, November 2011. doi: 10.1088/0067-0049/197/1/6.

S.F. Green, M.H. Jones, S.J. Burnell, et al. *An Introduction to the Sun and Stars*. Open University, 2004. ISBN 9780521546222.

M. Güdel. The sun in time: Activity and environment. *Living Reviews in Solar Physics*, 4(3), 2007. URL `http://www.livingreviews.org/lrsp-2007-3`.

J. Javaraiah and K. Gokhale. *The Sun's Rotation*. Number v. 1. Nova Science, 2002. ISBN 9781590332504.

J. M. Jenkins, D. A. Caldwell, H. Chandrasekaran, J. D. Twicken, S. T. Bryson, Quintana, et al. Overview of the Kepler Science Processing Pipeline. *Astrophysics Journal Letters*, 713:L87–L91, April 2010. doi: 10.1088/2041-8205/713/2/L87.

D. G. Koch, W. J. Borucki, G. Basri, N. M. Batalha, T. M. Brown, D. Caldwell, J. Christensen-Dalsgaard, W. D. Cochran, et al. Kepler mission design, realized photometric performance, and early science. *The Astrophysical Journal Letters*, 713(2):L79, 2010. URL `http://stacks.iop.org/2041-8205/713/i=2/a=L79`.

G. Kopp and G. Lawrence. The total irradiance monitor (tim): Instrument design. *Solar Physics*, 230:91–109, 2005. ISSN 0038-0938. URL http://dx.doi.org/10.1007/s11207-005-7446-4. 10.1007/s11207-005-7446-4.

D. A. Ostlie and W. C. Carroll. *Modern Stellar Astrophysics*. Addison-Wesley Publishing Company, 1996. ISBN 9780201598809.

A.C. Phillips. *The Physics of Stars*. Manchester Physics Series. John Wiley, 1999. ISBN 9780471987987.

W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical recipes in C (2nd ed.): the art of scientific computing*. Cambridge University Press, New York, NY, USA, 1992. ISBN 0-521-43108-5.

D. Prialnik. *An Introduction to the Theory of Stellar Structure and Evolution*. Cambridge University Press, 2000. ISBN 9780521650658.

R. R. Radick. Variability of sunlike stars. *Journal of Atmospheric and Solar-Terrestrial Physics*, 65:105–112, January 2003. doi: 10.1016/S1364-6826(02)00284-5.

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

A. Schuster. On the Investigation of Hidden Periodicities with Application to a Supposed 26 Day Period of Meteorological Phenomena. *Terrestrial Magnetism (Journal of Geophysical Research)*, 3:13–41, 1898. doi: 10.1029/TM003i001p00013.

D. R. Soderblom, J. R. Stauffer, K. B. MacGregor, and B. F. Jones. The evolution of angular momentum among zero-age main-sequence solar-type stars. *Astrophysics Journal*, 409:624–634, June 1993. doi: 10.1086/172694.

S. K. Solanki. Sunspots: An overview. *Astronomy and Astrophysics Review*, 11:153–286, 2003. ISSN 0935-4956. URL http://dx.doi.org/10.1007/s00159-003-0018-4. 10.1007/s00159-003-0018-4.

S. K. Solanki, B. Inhester, and M. Schüssler. The solar magnetic field. *Reports on Progress in Physics*, 69:563–668, March 2006. doi: 10.1088/0034-4885/69/3/R02.

R. F. Stellingwerf. Period determination using phase dispersion minimization. *Astrophysics Journal*, 224:953–960, September 1978. doi: 10.1086/156444.

J. O. Stenflo. Time invariance of the sun's rotation rate. *Astronomy & Astrophysics*, 233: 220–228, July 1990.

K. G. Strassmeier. Starspots. *Astronomy and Astrophysics Review*, 17:251–308, 2009. ISSN 0935-4956. URL http://dx.doi.org/10.1007/s00159-009-0020-6. 10.1007/s00159-009-0020-6.

R.J. Tayler. *The Stars: Their Structure and Evolution*. Cambridge University Press, 1994. ISBN 9780521458856.

R.J. Tayler. *The Sun As a Star*. New York, NY, USA, 1997. ISBN 9780521468374.

J.H. Thomas and N.O. Weiss. *Sunspots and starspots*. Cambridge astrophysics series. Cambridge University, 2008. ISBN 9780521860031.

C. Torrence and G. P. Compo. A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, 79:61–78, January 1998. doi: 10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2.

R. C. Willson and H. S. Hudson. Variations of solar irradiance. *Astrophysics Journal Letters*, 244:L185–L189, March 1981. doi: 10.1086/183508.

George L. Withbroe. Solar xuv limb brightening observations. *Solar Physics*, 11: 42–58, 1970. ISSN 0038-0938. URL `http://dx.doi.org/10.1007/BF00156549`. 10.1007/BF00156549.

M. Zechmeister and M. Kürster. The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. *Astronomy & Astrophysics*, 496:577–584, March 2009. doi: 10.1051/0004-6361:200811296.