STELLAR VARIABILITY AND ROTATION IN KEPLER PLANETARY TRANSIT SEARCH DATA

Amy McQuillan
Linacre College

A thesis submitted in candidature for the degree of Doctor of Philosophy
Hilary Term 2013

STELLAR VARIABILITY AND ROTATION IN KEPLER PLANETARY TRANSIT SEARCH DATA

Amy McQuillan
Linacre College

A thesis submitted in candidature for the degree of Doctor of Philosophy
Hilary Term 2013

ABSTRACT

The recent space-based exoplanet transit searches, CoRoT and *Kepler*, have revolutionised the field of stellar variability. In this thesis I exploit the public *Kepler* data to characterise stellar variability, and study rotation periods.

For the study of stellar variability it is a complicated but necessary process to remove instrumental systematics while maintaining intrinsic stellar signal. I was involved in the development of a new correction method for systematics, denoted ARC (Astrophysically Robust Correction). This method relies on the removal of a set of basis functions that are determined to be present in small amounts across many light curves.

Using the first month of *Kepler* data, corrected with the ARC method, I studied the variability properties of main sequence stars as a function of fundamental stellar parameters. I find that the fraction of stars with variability greater than that of the Sun is 60%, and confirm the trend of increasing variability with decreasing effective temperatures. I show tentative evidence that the more active stars have lower proper motions and may be located closer to the galactic plane. I also investigate the frequency content of the variability, showing that there exist significant differences in the nature of variability between spectral types, with a trend towards longer periods at later spectral types.

In order to exploit the full potential of the *Kepler* data for stellar rotation period measurement, I developed a novel method of period detection for use on star spot modulated light curves. Standard approaches to period detection are based on Fourier decomposition or least-squares fitting of sinusoidal models. However, typical stellar light curves are neither sinusoidal nor strictly periodic.

Therefore, I developed an algorithm for period detection based on the autocorrelation function (ACF) of the light curve. Because the ACF measures only the degree of

self-similarity of the light curve at a given time lag, the period remains detectable even when the amplitude and phase of the photometric modulation evolve significantly.

I applied the ACF method for the sample of M-dwarfs observed during the first 10 months of the *Kepler* mission, and detected rotation periods in 1570, ranging from 0.37–69.7 days. The rotation period distribution is clearly bimodal, with peaks at ~ 19 and ~ 33 days, hinting at two distinct waves of star formation. These two peaks form two distinct sequences in period-temperature space, with the period decreasing with increasing temperature.

In a natural continuation to this work I applied measured periods for 1000 stars in each of the F, G and K-dwarf sets observed by *Kepler*, and combined these with the M-dwarf results. The trend of increasing rotation period with increasing mass is clear throughout, as the observations fall along a wide by distinct sequence. Comparison to the rotational isochrones of Barnes (2007) show an overall agreement, although the dataset, which I believe is the largest set of rotation period measurements for main sequence stars, shows addition detail, not captured by the gyrochronology relations. This includes a dip in the rotation period distribution at $\sim 0.6\,M_{\odot}$ and a steep increase in period for the M-dwarfs.

I also applied the ACF method to the *Kepler* exoplanet candidate host stars and used the results to search for evidence of tidal interaction between the star and planet. I show that for the majority of exoplanet host stars, spin-orbit interaction will not have affected the stellar rotation period, permitting the application of gyrochronology for age determination. A comparison of the host stars with a sample of field stars selected to match their temperature and magnitude distribution also indicates no significant difference in the period or amplitude distributions of the two sets. The only notable variation is the lack of planets around the very fast rotators across all spectral types.

DECLARATION

I declare that no part of this thesis has been accepted, or is currently being submitted, for any degree or diploma or certificate or any other qualification in this University or elsewhere. This thesis is the result of my own work unless otherwise stated below.

All material in this thesis was done in collaboration with my supervisor Dr. Suzanne Aigrain. The material in Chapter 2 was also done in collaboration with Prof. Stephen Roberts and Dr. Steven Reece. The techniques were selected and coded by Prof. Stephen Roberts and my role was to identify trends, suggest expected links between trends and their sources and hence the functional form they may take, perform visual verification and provide feedback based on statistical tests. This work has been published in:
Roberts, S., McQuillan, A., Reece, S., Aigrain, S., 2013, *Astrophysically Robust Systematics Removal using Variational Inference: Application to the First Month of* Kepler *Data*, MNRAS, 435, 4, 3639.

The material in Chapter 3 was done in collaboration with Prof. Stephen Roberts, and makes use of the ARC systematics correction method described in Chapter 2. This work has been published in:
McQuillan, A., Aigrain, S., Roberts, S., 2012, *Statistics of Stellar Variability from Kepler - I: Revisiting Quarter 1 with an Astrophysically Robust Systematics Correction*, A&A, 539, A137.
and:
McQuillan, A., Aigrain, S., Roberts, S., 2012, *Statistics of Stellar Variability in* Kepler *Data with ARC Systematics Removal*, New Horizons in Time-Domain Astronomy, Proc. IAU Symposium, 285, 364-365.

The material in Chapters 4 and 5 was done in collaboration with Prof. Tsevi Mazeh. This work has been published in:
McQuillan, A., Aigrain, S., Mazeh, T., 2013, *Measuring the Rotation Period Distribution of Field M-Dwarfs with* Kepler, MNRAS, 432, 2, 1203.
The tables of M-dwarf rotation period data are available online in a machine-readable form at `www.physics.ox.ac.uk/StellarRotation`.

The work in Chapter 7 was done in collaboration with Prof. Tsevi Mazeh, Dr Frédéric Pont and Dr. Caroline Terquem.

ACKNOWLEDGMENTS

During my DPhil, I have received a huge amount of support, encouragement and guidance, without which, this thesis would not have been possible. First and foremost I would like to thank my supervisor Suzanne Aigrain for her continued support of my work since the start of my Mphys project, for offering me this DPhil position and the chance to study at Oxford, for all she has taught me about research, and for her guidance through the many highs and lows of academic life.

I would like to thank Jake, to whom I owe so much for the adventures of the past 9 years; without you I would not have made it this far. Thank you to my family for the emotional and financial support, providing a refuge at all times, and for being so proud of me whatever I do. Thanks to Will for giving a socially awkward 9 year old a microscope and science books, setting me on a path to somewhere I belong.

Thank you to Tsevi Mazeh, Stephen Roberts, Stephen Reece, Frédéric Pont and Caroline Terquem for their contribution to this work, helpful advice and suggestions.

The highlight of my DPhil has been the Oxford Astro family, who've inspired me since day one. Thank you for the pub trips, coffee breaks, parties, games, travelling, music, food and love. Thank you for being there through everything, and getting me back on my feet after I broke one of them ...and then a year later the ankle. Specifically I want to thank Francesco for the 1st year. Sam G for being there through the toughest times, for the puns and songs. Tim for rock nights and curries. Calum for Green Templeton Ball and everything since. Sam D for teaching me so much (skiing, driving, music and more), getting me through the last weeks of thesis and proofreading it all. Adam for cakes and lunches. Taysun for being the perfect, tolerant office mate. Phil for letting me crash the job tour. Garret, Jamie and Boon for all the nights in the Eagle & Child. Ed for the adventure stories and doing my observing over Christmas & New Year! Renée and Sarah M for being caring and inspiring. And to the rest, thank you, it would not have been the same without you!

I would also like to thank Vanessa and Ashling for their help with all my admin needs. I am grateful to the team behind the *Kepler* mission, which provided the data used in this thesis. This DPhil was funded by a Science and Technology Facilities Council (STFC) studentship.

"You said I am as constant as a northern star, and I said constantly in the darkness, where's that at?"

$\sim$ Joni Mitchell, quoting Shakespeare
describing Polaris, which has since been classified as a Cepheid variable.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

In the 4^{th} century B. C., Aristotle believed the stars were eternal and unchanging. This belief was upheld in western culture for the following $\sim$ 1900 years, until 1572, when Tycho Brahe refuted the concept based on his observations of a 'stellae novae', now known as supernovae. During the following centuries, further observations revealed stars with periodic brightness variations, and by the mid 1900s many thousands of variable stars had been discovered, the majority of which were large amplitude pulsators and eclipsing binaries. As observational techniques and technology improved, smaller amplitude variability could be detected, including that of star spot modulation on solar type stars (Radick et al., 1982).

The past decade has seen a rapid increase in the rate of discovery and classification of variable stars, mainly as a result of time-domain photometric surveys whose primary goals were to search for other phenomena, such as microlensing or planetary transits (e.g. the OGLE survey, Udalski et al. 2008, and SuperWASP, Pollacco et al. 2006). The typical precision and time sampling of ground-based surveys confines the associated variability studies to 'classical' variable stars, with amplitudes of a percent or more. In these surveys, the Sun, whose total output never varies by more than around 0.5% peak-to-peak (Fröhlich, 2011), even at the maximum of its activity cycle, would appear as a 'quiet' or 'constant' star. However, space-based transit surveys such as CoRoT (Baglin, 2003) and *Kepler* (Borucki et al., 2010) are sensitive

to 'micro-variability' down to and well below the solar level, on timescales ranging from minutes to months and, over the entire lifetime of *Kepler*, years. These surveys have revolutionised the field of stellar variability, and in this thesis I exploit the *Kepler* data for the study of variability as a function of stellar parameters, and to derive rotation periods for a large sample of main-sequence stars. I place particular focus on new and robust statistical techniques, for both data processing and rotation period measurement.

1.1 STELLAR VARIABILITY

Sources of stellar variability are wide ranging, from the potentially large-amplitude signatures of eclipses, pulsations and star spots, down to the sub-millimagnitude changes induced by granulation. While it is important to understand and recognise these different types of variability, in this thesis I focus predominantly on photometric variations caused by magnetic activity in solar type stars.

1.1.1 Origins of Stellar Activity

All stars with outer convective envelopes display variability as a result of magnetic activity, which results from the dynamo mechanism within the star (see e.g. Ossendrijver, 2003). At the tachocline, where the differentially rotating convective zone meets the uniformly rotating radiative core (Hughes et al., 2007), strong magnetic fields are formed as kinetic energy is converted to magnetic energy. The effect of differential rotation is to stretch and wrap the magnetic field lines around the star, thus converting poloidal flux into toroidal flux (the Ω effect). This toroidal flux is then twisted by the rotation and turbulence, (the α effect) producing a toroidal field. Strong magnetic fields exert pressure and can lead to a decrease in plasma density, causing the region to bubble up to the surface in a phenomena called magnetic buoyancy.

Magnetic fields penetrate the stellar atmosphere causing dark spots, surrounded by bright faculae. The spots appear darker since the magnetic field inhibits convection,

leading to cooler regions on the stellar surface. As these regions evolve and rotate across the stellar surface they cause periodic variations in brightness, which are sometimes of large amplitude. On a smaller amplitude scale, the motion and temperature variations of convective cells leads to granulation, producing a stochastic variation of stellar flux.

Radick et al. (1982) first observed variability in the continuum flux among single solar-type stars, although the precision limit of these early observations was $\sim 0.5\%$, and Dorren & Guinan (1982) showed that photometric variability occurs in anti-phase with variations in H_α and Ca II H & K emission variations. The production of emission lines such as Ca II H & K, is dominated by collisional processes, which reflect the kinetic energy and hence temperature variations on the surface of the star. The S index (Vaughan et al., 1978) is measured as the ratio of the flux in the core of the Ca II H and K lines to the continuum nearby, and is a commonly used stellar activity index.

An alternative index is $R'_{\rm HK}$, defined as the ratio of the chromospheric H and K emission to that of the bolometric emission of the star (Noyes et al., 1984). In the same paper, Noyes et al. (1984) introduced the Rossby number, Ro, and shows that it is empirically related to $R'_{\rm HK}$ by:

$$\log({\rm Ro}) = \log(P_{\rm rot}/\tau_{\rm conv}) = 0.324 - 0.400y - 0.283y^2 - 1.325y^3 \ , \qquad (1.1)$$

where $P_{\rm rot}$ is the rotation period, $\tau_{\rm conv}$ is the convective turnover time, and y is $\log(R'_{\rm HK})$. This has become a common way to parameterise stellar activity levels.

Observed photometric variability is strongly correlated with rotation rate, which drives the stellar dynamo mechanism, creating the magnetic flux from which the star spots originate (Pizzolato et al., 2003). Since angular momentum declines with age (see Section 1.2 for more detail), so does activity level. The timescale for convection is defined by plasma velocity and mixing length, which is mass dependent and can therefore be approximated from $B - V$ or theoretical stellar evolution models.

1.1.2 *Statistics of Variability*

Since the fundamental dependence of activity on rotation period and age were established (Skumanich, 1972), studies have searched for correlations between activity levels and other stellar parameters. Eyer & Grenon (1997) show a temperature dependence, where cooler stars display higher amplitude variability than hotter stars, which translates naturally to a correlation with spectral type.

The recent space-based transit searches CoRoT and *Kepler* have revolutionised the study of stellar variability, with variation precision down to mmag level now obtainable. Hulot et al. (2011) study the low frequency power excess in the CoRoT light curve, which they attribute to stellar activity. After removal of this component, they assign an index to the remaining power and use it to characterise the microvariability and show that it correlates with temperature, displaying the highest values for the coolest stars.

Stellar variability in *Kepler* data has been studied by Basri et al. (2010), who show that somewhat less than half of the dwarf stars surveyed are more variable than the Sun on timescales of up to a month, with the fraction increasing from earlier to later spectral types. They also show a trend of increasing variability with decreasing temperature, and demonstrate clearly a dearth of quiet cool stars. Basri et al. (2011) went on to demonstrate that periodic variable stars have significantly larger amplitudes, as a sample, than aperiodic variables. Ciardi et al. (2011) performed a complementary study considering the likely dwarfs and giants separately. The studies by Basri et al. (2010, 2011) and Ciardi et al. (2011) are described in more detail in Chapter 3. Variability statistics have also been determined using the 10 days of commissioning data (Q0), with the aim of developing methods to characterise and select specific types of variable (Walkowicz & Basri, 2010).

1.2 STELLAR ROTATION

The rotation of the Sun has been studied since 1611, when Galileo observed spots crossing its surface. All stars rotate because material in the clouds of interstellar matter from which they form must spin up as it accretes in order to conserve angular momentum. Of the readily observable properties of stars, the rotation rate is one which evolves significantly on the main sequence: intermediate- and low-mass stars are thought to spin down throughout their lifetimes, losing angular momentum via a magnetised wind (Schatzman, 1962) that is linked to their outer convection zone (Kawaler, 1988; Bouvier et al., 1997), and also leading to mass loss. This magnetised wind can start before the star reaches the main sequence, but while the star is still contracting, the spin-up due to contraction usually dominates over the spin-down due to the wind.

Period measurements for main-sequence stars with well-determined ages remain scarce, particularly at low masses, and gyrochronological ages are thus restricted to a limited range of masses and ages and remain very uncertain (Barnes, 2007).

Over the past decade, open cluster surveys have provided thousands of measurements for low mass stars with ages up to $\sim 650\,\mathrm{Myr}$, an overview of which can be found in Irwin & Bouvier (2009). Notable, more recent additions to the literature on rotation of early main-sequence low-mass stars include studies of M37 (Meibom et al., 2009) and Coma Berenices (Collier Cameron et al., 2009). Together, these data have provided a relatively complete, but complex picture of rotational evolution on the pre-main sequence. In turn, this has led to renewed efforts to develop models which can describe, or even better explain, the observations over the full sub-solar mass range (Barnes & Kim, 2010; Barnes, 2010; Reiners & Mohanty, 2012).

Period measurements for older stars remain very scarce, because they rotate more slowly and are less active than their younger counterparts, making it is very difficult to detect their rotational modulation from the ground. Notable exceptions include 71 main-sequence F, G and K stars observed as part of the Mount Wilson HK project (Barnes, 2003), 1727 mid-F to mid-K stars observed by the CoRoT satellite (Affer et al.,

2012), and 41 low mass (0.1–0.3 $M_\odot$) stars from the *MEarth* survey (Irwin et al., 2011), and $\sim$ 2000 field K and M stars observed the HATNet survey (Hartman et al., 2011). These majority of these stars are assumed to be old since they are not selected to be part of known young clusters, however a range of ages are likely to exist in these samples.

The *Kepler* space mission (Borucki et al., 2010) now offers a unique opportunity to measure rotation periods even for slowly rotating, moderately active stars, thanks to its superior precision and long baseline. A previous study by Harrison et al. (2012) measured rotation periods for 265 stars with $T_{\rm eff} \leq 5200$ K and $\log g \geq 4.0$ dex observed by *Kepler* for 1–2 quarters through the Cycle 1 Guest Observer program.

The few open clusters included in *Kepler*'s field-of-view are particularly important, since their ages can be estimated relatively well, and a small sample of periods has already been published for 71 members of NGC 6811 (Meibom et al., 2011). *Kepler* also observed tens of thousands of field stars, which can yield period measurements. Although they lack individual age estimates, they provide a global picture of stellar spin across our Galactic neighbourhood, and can be used to constrain the period-mass-age relation in a statistical sense. Samples of rotation periods such as these are often very incomplete due to detection limitations, and therefore one must consider the possibility that the observed rotation periods to not represent the bulk of stars at the same age.

1.2.1 *Angular Momentum Evolution*

This section describes current observations and understanding of angular momentum evolution in low mass-stars ($M_\odot$ down to hydrogen burning limit).

1.2.1.1 *Observations*

Before stars arrive on the main sequence (up to $\sim$ 30 Myr for a solar mass star), they display a wide range of rotation periods at a given mass (see e.g. the top left panel of

Figure 1.1). Irwin & Bouvier (2009) provide a review of rotation period data collected from the literature for stars of a wide range of masses in open clusters of increasing age.

For stars close to $1\,M_{\odot}$, the slowest rotating stars ($P \sim 10\,\mathrm{d}$) in the youngest clusters appear to maintain an approximately constant rotation period, spinning up to ($P \sim 8\,\mathrm{d}$) at the zero age main sequence (ZAMS). Over the same duration, the fast rotators ($P \sim 1\,\mathrm{d}$), spin up to ($P \sim 0.2\,\mathrm{d}$), more closely following model predictions of angular momentum conservation during gravitational collapse. By the age of the Hyades (625 Myr), the rotation periods converge onto a well defined sequence.

Following a different behaviour, the lowest mass stars spin up onto the ZAMS (for a $0.4\,M_{\odot}$ star, the ZAMS is reached at $\sim 150\,\mathrm{Myr}$), and start to form an upper envelope in the mass-period plane. A key observation is that all stars spin up, as shown by the fact there are no slowly rotating low-mass stars at 30-100 Myr, and then proceed to spin down at different times, leading to the appearance of two distinct sequences. Eventually, stars converge on the slowly rotating sequence, although even at the age of the Hyades, low-mass stars maintain their angular momentum and display periods $<$ 1 day. More recent rotation period observations of stars in 600 Myr clusters include those in Praesepe (Agüeros et al., 2011; Scholz et al., 2011), which show a significant dispersion in rotation periods still exist at the low-mass end.

1.2.1.2 *Angular Momentum Evolution Models*

No model currently explains all the rotational evolution data, and even the best theories still fail to account for all parameters (e.g. mass loss). The most successful approaches so far are semi-empirical and reasonably simplified formulas.

Stars below $0.5\,M_{\odot}$ have deep convective envelopes which support strong magnetic fields. The stellar wind is magnetically constrained and forced to corotate with the star out to large radii, therefore removing angular momentum from the star. Skumanich (1972) initially showed that the rotation rate of main-sequence stars decreases as $t^{-1/2}$. Since later type stars are observed to spin down more slowly than early types, Stauffer

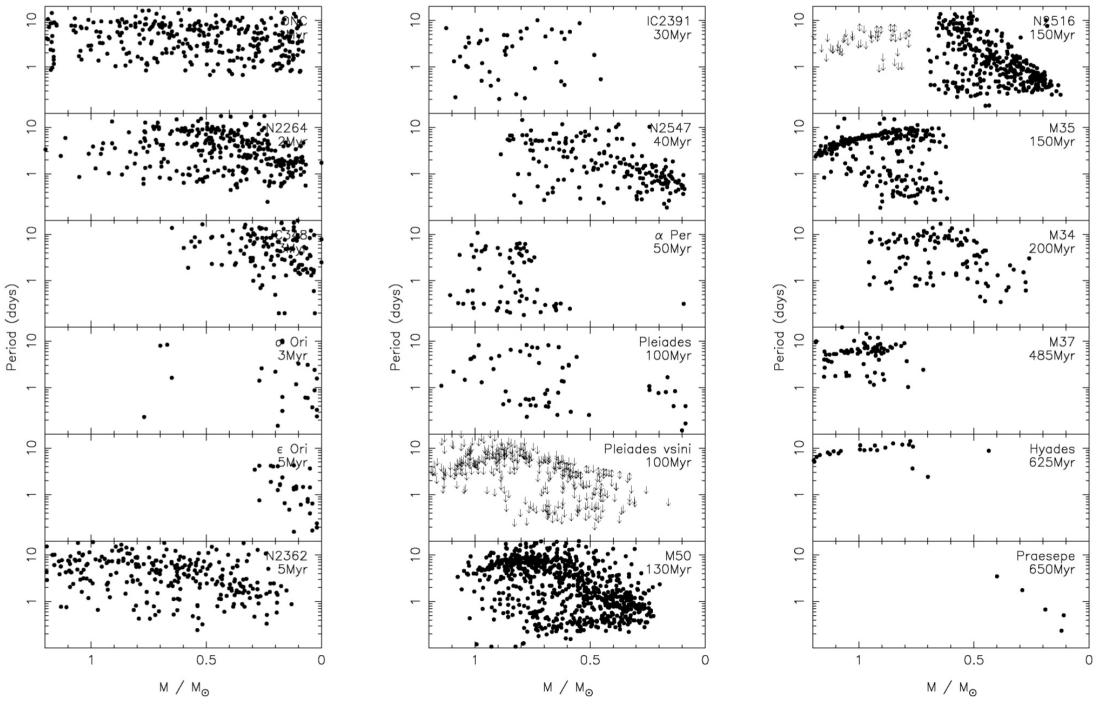

Figure 1.1: Figure from Irwin & Bouvier (2009): Rotation periods for stars with mass $\lesssim 1.2\,M_\odot$ and age $\lesssim 1\,\mathrm{Gyr}$, from open clusters in the literature.

& Hartmann (1986) concluded that only the convective envelope is initially spun down by the wind, and since these contain a larger fraction of the stellar mass and hence moment of inertia for lower mass stars, the spin down in less effective.

Kawaler (1988, 1989) build on the work of Kawaler (1987) and Mestel (1984) to develop a formalism that estimates the rate of angular momentum loss via a magnetised wind. This is based on the assumptions that the strength of the surface B-field scales with rotation rate, the star rotates as a solid body and the angular momentum of the magnetised wind no longer contributes to the star's total angular momentum J, once it has passed the corotation radius:

$$\frac{\mathrm{d}J}{\mathrm{d}t} = -K\left(\frac{R}{R_\odot}\right)^{2-n}\left(\dot{M}\right)^{1-2n/3}\left(\frac{M}{M_\odot}\right)^{-n/3}\times\left(\frac{\Omega}{\Omega_\odot}\right)^{4an/3}\Omega\,, \tag{1.2}$$

where Ω is the star's rotation rate, M is the stellar mass, R is the stellar radius, $\Omega_\odot(\approx 2.6\times10^{-6}\,\mathrm{s})$ is the mean rotation rate of the solar surface, n is an index, which depends on the geometry of the magnetic field, and varies between 2 (for a radial field), and 3/7 (for a dipole field), a is an index related to the scaling of magnetic field with rotation rate, which is usually assumed to be between 1 and 1.5, $\dot{M}$ is the

mass loss rate, and K is a constant which depends on the magnetic flux density at the surface, $B_0/B_\odot$, and is calibrated so that the model predictions match the Sun. If all other quantities are in cgs units, the units of K are [$g^{2n/3}\,cm^2\,s^{-1}$].

Using the simple angular momentum conservation law

$$J = I\Omega = I\frac{v_{\rm eq}}{R}\,, \tag{1.3}$$

where I is the star's moment of inertia and $v_{\rm eq}$ is equator rotational velocity, and assuming I and R do not change significantly over the main sequence, it can be shown that

$$v_{\rm eq} \propto t^{-3/4an}\,. \tag{1.4}$$

Thus for $a = 1$ and $n = 1.5$ this equation yields the Skumanich (1972) law, and conveniently removes the dependance of $\mathrm{d}J/\mathrm{d}t$ on $\dot{M}$. Using these indices, one can also show that mean angular momentum is proportional to M^2, which follows the observed dependence for low mass stars.

Kawaler (1988) shows the dependence of this relation on different values of n and initial J, and demonstrate that for $t \gtrsim 10^8$ yrs, the rotational velocity is independent of initial angular momentum (see Figure 1.2). This model also explains the large spread in rotation velocities of stars in young clusters, if one assumes a correspondingly large spread in initial angular momentum, and perhaps a slight spread in age.

Kawaler (1989) states that by 10^8 yrs, rotation rate is independent of initial angular momentum, and by this time, internal angular momentum redistribution is complete, meaning that low mass stars rotate approximately as solid bodies. These two statements lead to the conclusion that mean rotation period depends only on time, mass and radius, and thus, a calibrated angular momentum evolution law could be used to determine stellar ages where masses and radii are known.

Kawaler (1989) substitutes the relationships between observed colour and stellar properties to derive an equation relating rotation rate, age and colour:

$$\log(v_{\rm rot}) = 0.5\,\log\,(t_9) - 0.390(B-V) + 0.880 \tag{1.5}$$

where $v_{\rm rot}$ is rotational velocity in km s^{-1}, and t_9 is age in Gyrs. More recently, Reiners & Mohanty (2012) revised the Kawaler (1988) formalism arguing there was an error in

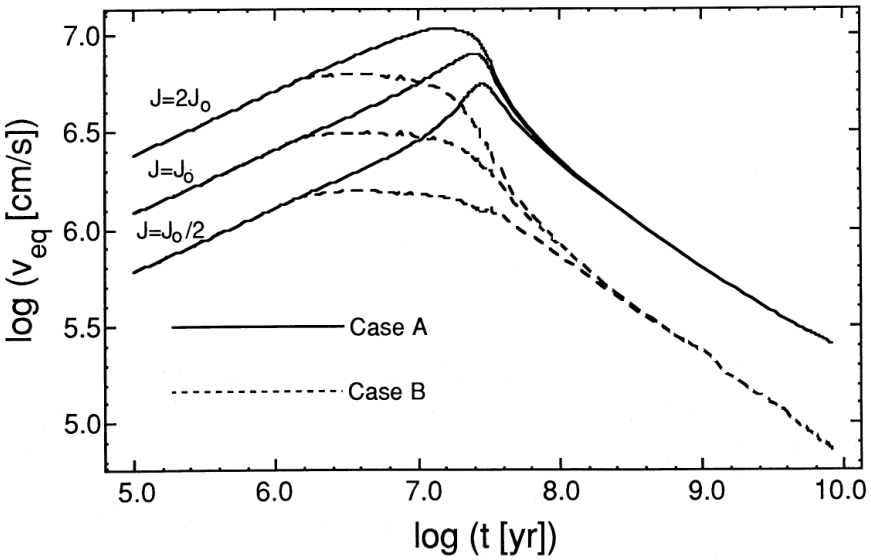

Figure 1.2: Figure from Kawaler (1988): The angular momentum evolution model of Kawaler (1988) for a 1 $M_\odot$ star, for case A: solid body rotation, and case B: local conservation of angular moment in the radiative zones and solid body rotation in the convective envelope. The tracks show several starting values for angular momentum and show that for $t \gtrsim 10^8$ yrs, the rotational velocity is independent of initial angular momentum.

the relation between magnetic field and rotation rate. They suggest that the magnetic flux density, rather than the magnetic flux, should be assumed to scale with rotation rate, changing the dependence of the angular momentum loss rate on stellar radius considerably.

1.2.1.3 Introducing Saturation

The simple models of angular momentum described so far managed to reproduce the bulk of rotational observations available at the time. However, they fail to explain all features seen in Figure 1.1, particularly the very fast rotation of some stars as they arrive on the main sequence. As stars spin up and contract their increased rotational velocity should create strong magnetic fields, which in turn should cause their rotation to slow.

To explain the existence of very rapid rotators in clusters aged $\sim$ 50–100 Myr, Barnes & Sofia (1996) allowed the magnetic field to saturate above a given rotation rate

Ω_{crit}, preventing the increase of angular momentum loss which should limit the rate of spin up. Therefore, stars can spin up as they approach the main sequence, pass the saturation threshold and continue to spin up to very rapid speeds without the increase in angular momentum loss. The same mechanism then causes the rapid rotators to spin down faster than non-saturated stars, to converge onto the main sequence with them.

This value of Ω_{crit} appears to vary for different stellar masses and is therefore tuned to match the observations. This additional consideration also provides a natural explanation for the period dispersion observed in young clusters, whereas the Kawaler (1988, 1989) models converge too quickly.

The saturation requirement also reveals clues to the origin of the magnetic fields. It would appear that fully convective M stars still show saturation in chromospheric emission (Stauffer, 1994), indicating it does not necessarily originate from the radiative-convective boundary as previously believed. This implies that the dependence of Ω_{crit} on mass could be due to the associated variation in convective zone depth.

1.2.1.4 Disk Interactions

An alternative and still debated explanation of the range of periods observed in young clusters is disk interaction (Edwards et al., 1993). The accretion of disk material with high angular momentum onto a pre main sequence (PMS) star is expected to lead to a very large increase in rotation rate. However, this is not supported by the observations, which would suggest a spin down mechanism, potentially through magnetic interaction with the slowly rotating outer disk regions, is acting to counter the increase of rotational velocity caused by accretion.

Edwards et al. (1993) argue that this disk interaction leads to slower rotation rates for stars during the accretion stage. After the disk has dissipated, the star alone must conserve angular momentum and spin up occurs. Stars accreting their disks, or losing them through evaporation at different rates during evolution may therefore end up reaching different maximum rotational velocities before arriving on the main

sequence, leading to the observed period dispersion. Results from a variety of cluster surveys find conflicting evidence for this theory.

Other disk interaction theories include accretion-driven stellar wind (e.g. Matt & Pudritz, 2005) and the simplified disk interaction formulation known as 'disk-locking' (e.g. Koenigl, 1991; Collier Cameron et al., 1995).

1.2.1.5 The I and C Sequences and Gyrochronology

Barnes (2003) combined observations of young open clusters to put forward an empirical formalism to allow calibration of the period-mass-age relation, which he was first to describe as 'gyrochronology'. The progression of cluster observations from young to old (shown in Figure 1.1) display two distinct sequences which Barnes (2003) label as the *I* (interface) and *C* (convective) sequences.

Seen most clearly together in the bottom centre panel of Figure 1.1 (M50), the *I* sequence forms the upper envelope with an overall trend of increasing period with mass. The *C* sequence contains the shorter period objects and has a trend of decreasing period with increasing mass, intersecting the *I* sequence at low masses. The *I* sequence becomes more defined and of steeper gradient with increasing cluster age. The *C* sequence also evolves with time, with initially lengthening periods, followed by a transition of stars over to the *I* sequence. The gap in between the two sequences implies this transition occurs over relatively short timescales.

Stars on the *I* sequence are well described by the Kawaler (1988, 1989) formalism and spin down following the Skumanich (1972) law, whereas stars on the *C* sequence comprise of the fast rotators, for which Barnes & Sofia (1996) proposed that magnetic saturation was required. Barnes (2003) determines that the shape of the *I* sequence depends primarily on the moment of inertia of the entire star, whereas the *C* sequence appears to depend on the moment of inertia of the convective zone only. This leads to the conclusion that the stars begin with their cores and envelopes decoupled and then evolve to become coupled as they progress to the *I* sequence. This process is faster in higher mass stars with thinner convective envelopes. Fully convective stars can only

lie on the *C* sequence. Barnes (2003) promoted this mechanism instead of magnetic saturation, to explain the transition to and from the fast rotation phase.

A further study by Barnes (2010) then negates the proposal that the mass dependence of these rotation periods can be explained simply by the difference between the whole star moment or inertia and that of the convective envelope alone. Barnes (2010) instead revises the formalism to produce an empirical fit to the available data, relating the change in rotation period with time to the convective turnover timescale τ_C:

$$\frac{dP_{rot}}{dt} = \begin{cases} k_C P_{rot}/\tau_C & \text{on the C sequence} \\ \tau_C/k_I P_{rot} & \text{on the I sequence} \end{cases} \tag{1.6}$$

where k_C and k_I are constants calibrated to match observations of young star clusters and the Sun, and τ_C is computed from stellar evolution models (Barnes & Kim, 2010), and encapsulates the mass dependence.

Barnes (2010) then notes that this parameterisation links the period evolution on the two sequences to the Rossby number, with the transition from *C* sequence to *I* sequence occurring at a specific value of $\mathrm{Ro} \equiv \tau_C/P_{rot}$. This links rotational evolution to Ro, which scales with the level of chromospheric activity (Noyes et al., 1984).

1.2.1.6 Rotation Overview

Returning to the discussion of models by Irwin & Bouvier (2009), one sees many of these features brought together to explain the observations available at that time. They assume an initially tight distribution of rotation periods, which are then dispersed by interaction with disks, which hold the rotation at a reduced speed for different disk lifetime durations, before the star begins to spin up. They consider the mechanism for main-sequence angular momentum loss to be a result of a magnetised stellar wind, with saturation.

For the main sequence, Irwin & Bouvier (2009) use the rotational models described in Bouvier et al. (1997); Allain (1998); Irwin et al. (2007) and Irwin (2007), which use a standard non-rotating PMS stellar evolution track to compute the variation of the stellar parameters as a function of time, which can then be used to determine

the angular velocity as a function of time. Their model has four parameters, which can all be a function of mass: the initial angular velocity ω_0 , normalisation of the solar-type angular momentum loss law K, saturation velocity $\omega_{\rm sat}$, and the lifetime of the circumstellar disc $\tau_{\rm disc}$. They find that under the assumption of solid body rotation, their model cannot fit the observations. Addition of the core-envelope decoupling allows the model to fit to the slow rotators more successfully. Their model fails to fit the fully convective regime.

In summary, no single model can fit all the available cluster data with a single prescription. Instead, one must allow for a range of initial rotation rates, disk lifetimes, or core-envelope decoupling, and even with these considerations, the results are not satisfactory. However, the study of rotation in older stars will hopefully prove to be less strongly affected, as the rotation rates evolve to be independent of initial conditions.

1.3 IMPLICATIONS OF VARIABILITY AND ROTATION STUDIES

Using variability and rotation observations to understand the internal structure and evolution of stars is an important objective, however, parameterisation of these phenomena is also valuable to wider astrophysical studies. In this work I focus on 'variability' defined as the amplitude of photometric modulation in the light curve, predominantly caused by star spots.

1.3.1 Consequences for Exoplanet Transit Searches

The impact of solar-like variability on transit searches has been acknowledged for many years, based on observations of the Sun (Borucki et al., 1985). As the photometric precision of exoplanet transit search missions improves, the intrinsic stellar variability becomes the limiting factor. Ground-based photometry is limited to a precision of $\sim$ 100 ppm for 4 m telescopes (Jenkins, 2002, and references therein) due to

atmospheric scintillation, whereas space-based photometry can achieve a precision of order $\sim$ 10 ppm (Jenkins et al., 2010b).

Pont et al. (2006) develop a method to assess the significance of transit detections in correlated or 'red' noise, and various methods have been proposed to filter this noise from transit searches (see e.g. Defaÿ et al., 2001; Jenkins, 2002; Carpano et al., 2003; Aigrain & Irwin, 2004). This is most important on the hours timescales associated with transit signals.

Stellar variability is still not well-understood, and in the pre-*Kepler* era it was even less so. This resulted in predicted transit detection statistics for missions being estimated based on solar data, which was assumed to be similar to 'solar type' stars. To characterise the effect of stellar variability on transit searches with more active stars, Aigrain et al. (2004) developed a model using chromospheric flux measurements as a proxy measure of activity-induced variability. This model takes the solar power spectrum and applies empirically derived scaling laws to produce a range of synthetic light curves into which transit signals could be injected.

Unfortunately, the predicted stellar noise component was based on the solar level as typical, and studies using the *Kepler* data have shown that the Sun is in fact quiet for it's spectral type (Basri et al., 2010, 2011; McQuillan et al., 2012). Gilliland et al. (2011) confirmed that the Combined Differential Photometric Precision, or CDPP (noise metric used by the *Kepler* team) is 30 ppm, or 50% higher than expected due to unforeseen levels of stellar activity. The *Kepler* mission must therefore run for an additional 3 yrs to obtain the same signal to noise confirmation of an Earth analogue.

Even less well-understood is the effect of stellar variability on radial velocity (RV) detection and confirmation of planets. This can induce line shape variations on timescales similar to planetary signals, which are commonly accounted for by adding an estimated 'jitter' term in quadrature to the RV uncertainties before applying planet detection techniques. This is not a robust method to obtain the uncertainties on the final planet parameters derived, and several attempts have been made to estimate the magnitude of RV jitter from observable stellar parameters (Wright, 2005; Isaacson & Fischer, 2010). A recent method to provide a statistical overview of activity effects in RV searches, and potentially reduce contamination in individual cases has been

developed by Aigrain et al. (2012), who propose to use *Kepler* light curves to predict radial velocities.

There is still considerable work to be done on the effects of stellar variability on planet detection and characterisation, which will be be aided by advances in our understanding of stellar activity.

1.3.2 *Know the Star, Know the Planet*

The formation, evolution, environment and habitability of an exoplanet are dependent on the host star, and consequently, any study of the planetary system must start with a thorough understanding of its star. Activity is particularly important since it provides an insight into the rotational and magnetic properties of the star. Gyrochronology has the potential to provide age estimates for single stars and their planetary companions, shedding light on the formation timescales and mechanisms. This will be particularly important for the host stars which are too faint for asteroseismology to be performed.

Asteroseismology is the method of using high cadence photometric observations of pulsation frequencies to determine the origin of the oscillations and hence the stellar density profile. Together with chemical abundance information from spectroscopy, this method can be used to map the internal structure of stars, and infer their ages.

The rotation period can also be used to probe tidal interactions between star and planet (Pont, 2009), and therefore provide constraints on the dynamical history of the system. Measurement of the stellar rotation period combined with model-based stellar radius estimates, can be used to determine the stellar inclination i_*, using spectroscopically measured $v \sin i_*$, where v is rotational velocity. This can then be compared to the orbital inclination of planet, determined by transit fitting.

M-dwarfs are an important target for exoplanet searches, since they are the most numerous stars in our galaxy, and their low mass and small size leads to favourable detection biases in transit and radial velocity surveys. Their low luminosity also allows for shorter period planets, which are more easily detected, to lie in the habitable zone (HZ, defined as orbital radius for water on the planets surface may exist in a liquid

state). The greatest concern for habitability of planets orbiting M-dwarfs is the effect of high level of magnetic activity associated with these stars (Tarter et al., 2007), and potential for tidal locking. This creates flares which could potentially deliver large amount of UV flux to the planet's surface.

1.3.3 Star-Planet Interactions

Spin-orbit interactions between host stars and their planets are ill-understood, due to the lack of host star rotation period observations. By obtaining measurements of the host star rotation period and the planet orbital period, one can search for evidence of tidal interaction between the star and planet. The effect of tidal interaction on the orbits and rotation of binary stars has been widely observed, although the theoretical details are complex and still uncertain (see e.g. Mazeh, 2008, for recent review). The same underlying physics can be applied to planetary systems, although the mass ratio is much greater and the tidal effects on the star are much smaller. Tidal effects lead to the transfer of angular momentum from the planet to the star (provided $P_{\rm orb} < P_{\rm rot}$), which over time can align the rotation axes of the two components, synchronise orbital and rotation periods, and circularise the orbit.

Although spin-orbit interactions have been studied for a number of individual systems, it is much harder to do a statistical study (see Pont, 2009, and references therein) due to the complexity of the tidal and magnetic formulae, poorly constrained parameters and the scarcity of planet host stars with measured rotation periods.

Pont (2009) finds evidence for circularisation of orbits, excess stellar spin due to tidal effects and concludes that tidal spin-up of the host star will produce a strong detection bias against massive, close-in planets, especially in radial velocity surveys. Pont et al. (2011b) extends this work to provide constraints on planet formation models, based on the observations of tidal circularisation of massive, close-in planets, and Husnoo et al. (2012) performs a similar study using new RV data.

1.3.4 *Calibration of Gyrochronology Models*

Over the past decade, open cluster surveys have provided thousands of measurements for low-mass stars with ages up to 650 Myr (Irwin & Bouvier, 2009; Meibom et al., 2011). With good calibration, gyrochronology could be used to measure the absolute age for FGK and early-M single field stars with rotation measurements, to an accuracy of $\sim 10\%$ (Delorme et al., 2011), while producing more self-consistent age measurements than chromospheric activity and isochrone fitting (Barnes, 2007).

The following relation of Barnes (2007) can be used to estimate rotational isochrones from $B - V$:

$$\log t_{\rm gyro} = \frac{1}{n}\left[\log P - \log a - b \log\left(B - V - 0.4\right)\right], \tag{1.7}$$

where $t_{\rm gyro}$ is in Myr, P is rotation period (in days), $n = 0.5189 \pm 0.007$, $a = 0.7725 \pm 0.011$, and $b = 0.601 \pm 0.024$. These parameters were determined by an empirical fit to the solar values, and those of 10 young open clusters, between 30–600 Myr.

More recently Barnes (2010) modified the gyrochronology relation to relate the age, period and convective turnover time, τ:

$$t = \frac{\tau}{k_{\rm C}} \ln\left(\frac{P}{P_0}\right) + \frac{k_{\rm I}}{2\tau}\left(P^2 - P_0^2\right), \tag{1.8}$$

where $k_{\rm C}$ and $k_{\rm I}$ are discussed in Section 1.2.1.5, and P_0 denotes the zero-age main sequence (ZAMS) period of the star. Since this revised relationship depends on many assumptions, I have opted to focus on the Barnes (2007) relationship in this work.

Period measurements for main-sequence stars with well-determined ages remain scarce, and hence gyrochronological ages remain very uncertain. Figure 1.3 illustrates this relation and shows the large areas of unconstrained parameter space.

To calibrate the gyrochronology relation, accurate ages for stars with measured rotation periods are required. The asteroseismic targets will provide ages for a selection of the *Kepler* targets, some of which will have rotation measurements (current PhD project of R. Angus, Oxford). This will be an important progression in the understanding and calibration of gyrochronology models, since it will allow the ages of single stars outside of clusters to be included. Current work on this calibration also

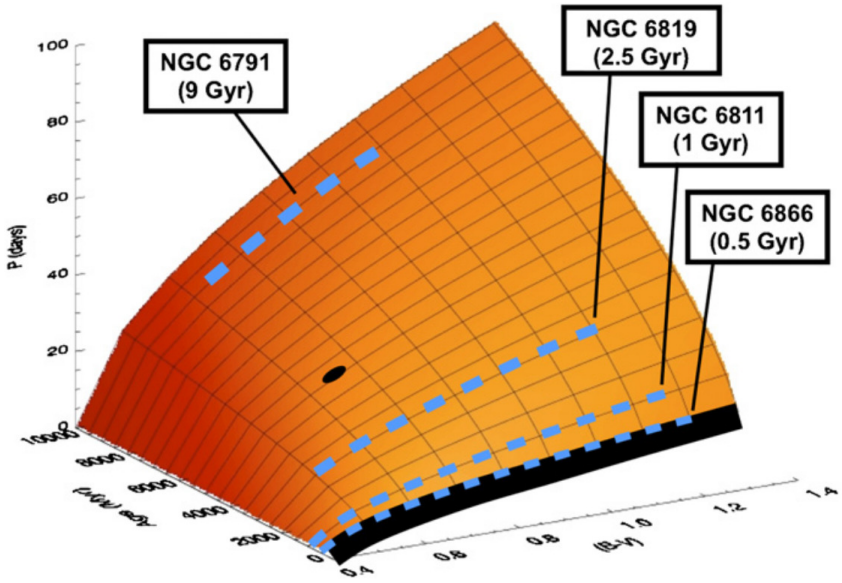

Figure 1.3: Figure from Meibom et al. (2011): Extrapolated colour (mass), age, rotation period surface. The blue lines show the *Kepler* cluster data, the black curve is the area constrained by young cluster data and the black point is the sun.

includes the 4 open clusters (Meibom et al., 2011) within the *Kepler* field of view, for which membership confirmation is being performed.

It is not currently known how well the gyrochronology relationships perform, especially at typical field star ages, and further work is required to establish this. Comparison of ages derived from gyrochronology and stellar evolution models may be able to provide some insight into this.

The work presented in this thesis will contribute to this area of study by providing an unprecedented number of rotation period measurements for field stars. Although independent age measurements for these stars will be scarce (e.g. model comparison from spectroscopic followup), it will be possible to compare gyrochronology models to the general trends seen in the data. This should allow for simple tests to be performed, for example, checking if the slowest rotators suggest unfeasibly old ages using current gyrochronology relations, and seeing the fractions of stars in different age groups seem appropriate for the current understanding of the region observed by *Kepler*.

1.3.5 *Mapping Star Formation Using Variability*

Since variability scales with age, one would expect to see differences in the typical amplitudes and periods of stars in separate waves of star formation. Using *Kepler* data Ciardi et al. (2011) noted tentative evidence that stars closer to the Galactic plane display higher variability, however in McQuillan et al. (2012), described in Section 3, I used a more robust systematics correction and no longer saw evidence of this trend. In Section 5, I discuss a bimodality in rotation periods of the M-dwarf population, which is interpreted as two waves of star formation at separate epochs.

As variability and rotation period work progresses, further investigation of correlations between activity and parameters, such as metallically and proper motion will be performed. This has the potential to open new avenues of study into star formation history and Galactic structure using photometric surveys such as *Kepler* and Gaia.

1.4 THE *kepler* MISSION

The *Kepler* space telescope, launched on 6$^{\text{th}}$ March 2009, is NASA's 10$^{\text{th}}$ Discovery Mission. Its primary aim is to detect extrasolar planets using the transit method, with an emphasis on the discovery of Earth-sized planets in the habitable zone. The results are already shaping our understanding of planetary formation mechanisms, planet frequency, the structure of extra-solar systems, and the characteristics of planet host stars. Among the $\sim$ 150,000 late-type main-sequence stars being observed by *Kepler* it is possible to probe the internal structure of few hundred through asteroseismology.

1.4.1 *Instrument Design*

The *Kepler* telescope has a 0.95 m aperture, and a 1.4 m primary mirror. It stares at a 115 deg^2 field of view (0.28% of the sky), on the edge of the Galactic plane (l = 76.32 deg, b = +13.5 deg), in the constellation of Cygnus. The satellite is in a

heliocentric orbit, and therefore avoids effects associated with geocentric satellites, for example, the high level of cosmic ray events that plagued its predecessor CoRoT. Each month, the satellite stops observations for ~ 1 day to downlink data to Earth, and every 3 months the satellite rotates 90 deg to keep the solar panels pointed at the Sun, introducing a gap in observations of ~ 3 days.

The mission observes $\sim 160,000$ targets, most with a cadence of 29.42 minutes. A subset are observed at an increased cadence of ~ 1 minute for use in asteroseismology studies and planet candidate cases.

The focal plane consists of 42 CCDs with 2200×1024 pixels, which are arranged in 21 modules, with 2 outputs per CCD. The photometer has a soft focus designed for stable photometry. The plate scale is 3.98 arcsec per pixel (van Cleve, 2009) and the astrometric precision for a single 30 minute measure is better than 4 milliarcsec (Monet et al., 2010).

The *Kepler* mission requires photometric precision of 20 ppm on 6.5 hr timescales for a *Kepler* magnitude 12 star, since an Earth-like transit produces a brightness change of 84 ppm and lasts for 13 hours when it crosses the centre of the star, see Jenkins et al. (2010b) for a detailed discussion.

1.4.2 *Instrumental Systematics*

The *Kepler* data contain a variety of instrumental systematics, ranging from high frequency, small amplitude effects of temperature changes, to the large amplitude, long term trend introduced by differential velocity aberration. The majority of instrumental systematics and artefacts in the *Kepler* data arise as a result of the photometric aperture selection. An aperture does not collect all of the flux from its target since inclusion of the target's faint PSF wings degrades the signal to noise. However, this leads to a setup where any effect causing a change of the target focus or position on the CCD results in significant changes in the amount of flux falling within the aperture.

Of the 40 guide stars selected to maintain accurate pointing, a few were intrinsically variable, including an eclipsing binary with a period of 18.25 days and sinusoidal

variable with a period of 2.9 days. The Attitude Determination and Control System (ADCS) consequently moved the spacecraft by small amounts in response to this varying input. The system and guide star selection were updated for Q3 onwards.

Solar radiation torque leads to angular momentum build up in the reaction wheels, which then require thruster firings every 3 days to desaturate them, leading to a 3 day signal in the data. The focus has also been found to vary on timescales of 3 to 6 hours which has been linked to the momentum desaturation activities, since the same signal is detected in the reaction wheel temperature sensors, although the physical mechanism linking the effects is not well understood (Christiansen et al., 2012).

The most prevalent long term trend is due to differential velocity aberration. This occurs because the motion of the satellite leads to an apparent displacement of the target with respect to its real location. Due to *Kepler*'s large FOV, the velocity to line of sight angle varies across the focal plane leading to variations in the relative positions of the targets on the CCD.

1.4.3 The Kepler *Input Catalog*

The stellar properties used for classification when using *Kepler* data in this thesis, namely effective temperature (accurate to 200 K) and surface gravity (accurate to 0.5 dex), come from the *Kepler* Input Catalog (KIC, Brown et al. 2011). These parameters were estimated using Bayesian posterior probability maximisation to match observed colours, estimated from Sloan g, r, i, z filters, 2MASS JHK, and D51 (510nm), to Castelli & Kurucz (2004) stellar atmosphere models. The proper motion values from the KIC are taken from a selection of catalogs[1] where available. Total proper motion is listed on the NASA Star and Exoplanet Database[2] (NStED) as having accuracy of 20 milliarcseconds per year. 39,000 dwarf stars are listed with non-zero total proper motion. The precisions listed in the KIC are the typical value for each parameter, although in reality these may vary by a small amount between stars of different magnitude and spectral type. For a more detailed discussion of the KIC parameters,

[1] *Kepler* Stellar Classification Program, Hipparcos, Tycho-2, UCAC2, 2MASS and USNO-B1.0.
[2] http://nsted.ipac.caltech.edu

see Brown et al. (2011); Batalha et al. (2010); Verner et al. (2011). There is increasing evidence that the KIC values for the radii and $T_{\rm eff}$ are overestimated (e.g. Mann et al., 2012; Verner et al., 2011; Muirhead et al., 2012), and although improved stellar parameter estimates are becoming available for some subsets of stars (e.g. Dressing & Charbonneau, 2013), these were not published in time for use in this thesis.

Data from the *Kepler* mission, once corrected for instrumental systematics, is particularly amenable to statistical variability studies because of the instrument's unprecedented photometric precision, vast field of view and long baseline observations.

1.5 THESIS OUTLINE

As discussed in Section 1.4.2, the raw *Kepler* data contain large amplitude systematic effects. Before the release of the PDC-MAP corrected data, I started work on an alternative correction method, in collaboration with Prof. Stephen Roberts (Oxford Engineering) and Dr S. Aigrain (Oxford Astrophysics). This Astrophysically Robust Correction for systematics (ARC) is described in detail in Chapter 2.

Following the application of the ARC to the Q1 *Kepler* data, I was able to study the variability properties of main-sequence stars as a function of fundamental stellar parameters, using one of the most valuable datasets available at the time. This work is described in Chapter 3, where I examine the variability fraction compared to the solar level, and search for correlations between variability and fundamental stellar parameters.

In order to robustly measure rotation periods in the *Kepler* data, I developed an algorithm for period detection based on the autocorrelation function (ACF) of the light curve. In Chapter 4, I describe the ACF method and show simulated examples with comparison to results derived from the wider used sine-fitting periodogram method.

In Chapter 5, I apply to the ACF method to first 10 months of *Kepler* data for the sample of M-dwarfs observed, and examine the distribution of rotation periods detected. I then proceed to apply the ACF method to a sample of the F, G and K dwarfs observed by *Kepler*, which is described in Chapter 6. Using the results, I am

able to report the periodic fraction and present the mass-period relation for what I believe to be the largest available set of field star rotation period measurements (1000 periodic detections for each spectral type, plus the M-dwarf sample). I then compare this distribution to the gyrochronology models of Barnes (2007).

A final piece of preliminary work on the rotation periods of the *Kepler* exoplanet candidate host stars is detailed in Chapter 7. Using a simple formula for the spin-up timescale, I explore the likelihood of star-planet interactions and search for evidence of tidal effects between the star and its nearest planet. I use these results to comment on the applicability of gyrochronology relations to the host star periods. I also perform a comparison of the rotation and amplitude characteristics of the host and field star populations.

2

REMOVAL OF INSTRUMENTAL SYSTEMATICS IN ENSEMBLES OF TIME SERIES DATA

The material in this chapter was published in Roberts, S., McQuillan, A., Reece, S. & Aigrain, S., 2013, MNRAS, 435, 4, 3639.

In this chapter I present a new method of systematics removal in ensembles of time series data, which we call the Astrophysically Robust Correction method (ARC), developed in collaboration with Prof. Stephen Roberts. In Section 2.1, I examine the types of systematic effects present in time series photometry, and the methods available to remove them. Existing data processing techniques used by the *Kepler* science team are described in Section 2.2 and 2.3. The application of the ARC to *Kepler* Q1 data is presented in Section 2.4.8.

The techniques were selected and coded by Prof. Stephen Roberts and my role was to identify trends, suggest expected links between trends and their sources and hence the functional form they may take, perform visual verification and provide feedback based on statistical tests. Although the majority of the detail contained in this chapter is the work of Prof. Stephen Roberts, a thorough explanation of the ARC is required for this thesis, since it was used to produce the corrected *Kepler* Q1 data, upon which the variability study described in Chapter 3 was performed. Most of the figures in this chapter were produced by Prof. Stephen Roberts for the submitted publication.

2.1 INTRODUCTION

2.1.1 *Systematic Effects*

Systematic effects are those affecting many data points, in the same way, for example by adding an offset or a trend. In this study I define systematic effects to be those affecting many light curves at the same time, although the amplitude of the trend introduced can vary slightly.

All astrophysical data obtained will contain systematic effects, and as the boundaries of instrumental precision are being pushed towards ever increasing sensitivity, the methods for correcting these effects must also advance. Surveys for exoplanet transits monitor many thousands of stars at high precision, with rapid time sampling, and over long durations. These transit data can be used not only for the discovery and characterisation of exoplanets, but also to study a wide range of stellar phenomena, including the variability and rotation on which this thesis focusses.

However, transit survey data are typically affected by a wide range of artefacts, many of which affect all the target stars simultaneously, although to varying degrees. For ground-based surveys such as OGLE (Udalski et al., 2002), HatNET (Bakos et al., 2002), and SuperWASP (Pollacco et al., 2006), the dominant systematics result from atmospheric effects such as differential airmass variations across the field-of-view, as well as seeing and atmospheric transparency variations. These surveys are also affected by instrumental systematics such as pointing jitter combined with inter- and intra-pixel sensitivity variations. In space-based surveys, many of the aforementioned instrumental systematics persist, along with differential velocity aberration due to the motion of the satellite.

When the cause of a systematic effect and its resulting effect on the data are understood well enough, it is possible to formulate explicit models for their removal. For example, if a change in detector temperature introduced a linear trend in detected flux, it would be possible to cotrend against temperature if recorded as part of the engineering telemetry. In many other situations, however, there may be little

information regarding the nature or number of systematic trends in the data.

2.1.2 *Popular Systematics Removal Methods*

The two methods, which are widely used to identify and remove systematics in ground-based transit surveys are known as SysRem (Tamuz et al., 2005) and TFA (Trend Filtering Algorithm, Kovács et al. 2005). SysRem attempts to explain each light curve as a linear superposition of all the other light curves, and adjusts the coefficients of this linear basis model so as to minimise the total squared variance of all the light curves. The resulting algorithm is very similar to Principal Component Analysis (PCA, see e.g. Wall et al. 2003), though it allows for individual weighting of each observation of each star. Because of this, the trends are not strictly orthogonal, and must be identified and removed iteratively. The number of times this is done must then represent a trade-off between removing systematics and preserving real variability. TFA proceeds in a similar fashion, but the trends are identified in a limited subset of the light curves, which are selected by the user, so as to contain a representative sample of the systematics, but minimal intrinsic variability.

The images from which the photometry is extracted, or ancillary sensors, can also be used to measure meteorological and instrumental parameters, which are thought to affect the photometry, for example seeing, airmass, detector temperature. From these, one may construct a linear basis with which to model the systematics, a procedure known as external parameter decorrelation, (EPD, Bakos et al. 2007). EPD is usually applied as a preliminary step prior to running a search for additional trends using (e.g.) SysRem or TFA. EPD does not rely on the availability of large numbers of light curves measured simultaneously with the same instrument, and can therefore be applied in situations where a single star is being monitored.

All three of the methods outlined above make a number of important assumptions:

- that the relationship between the trends and the light curves is linear;
- that the data used to identify the trends (light curves or external parameter data)

is sufficient to describe them completely;

- that the majority of the stars are not significantly variable, so that systematics dominate the global sum of the squared residuals which is used as a figure of merit.

The third assumption becomes problematic in the case of space-based transit surveys such as *Kepler* and CoRoT, which routinely reach sub-millimagnitude precision. As these missions are revealing, many, if not most, stars are intrinsically variable at that level. Intrinsic variability can become an important, if not dominant, contribution to the global figure of merit one is trying to optimise, exacerbating the problem of preserving this variability whilst removing the systematics.

2.2 *kepler*: SYSTEMATICS AND EARLY REMOVAL METHODS

2.2.1 *Origin of Systematics*

Although the focus of this chapter is the design of a systematics removal method which does not rely on prior knowledge of the origin of the systematic effects, it is important to understand the instrumental setup in order to understand the process and verify the results. Details of the *Kepler* satellite and data are described in Section 1.4.

The detector on the *Kepler* satellite is a shutterless photometer using 6 sec integrations and a 0.5 sec readout. For the long cadence data used in this study, 270 integrations are summed for an effective 28.4 min exposure. The time series data is extracted by summing the photoelectron counts within a pre-defined target aperture (Bryson et al., 2010).

The majority of instrumental systematics and artefacts in the *Kepler* data arise as a result of the photometric aperture selection. Using the *Kepler* Input Catalog (KIC) to estimate contamination from background objects, the optimal apertures sizes and shapes were chosen to maximise the signal to noise of each target (Batalha et al., 2010), within a defined spacecraft aperture (see Figure 2.1). This means that their

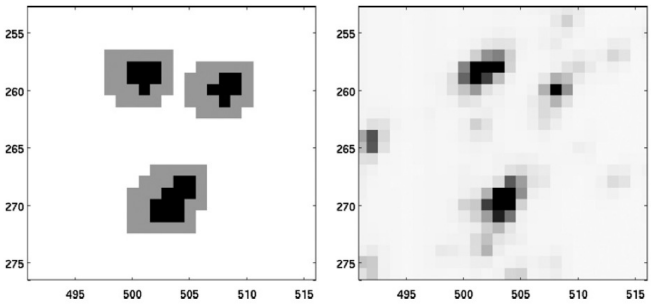

Figure 2.1: Figure from Bryson et al. (2010). Left: optimal apertures for the targets in black and the spacecraft apertures in grey. The additional pixels in the spacecraft aperture make up the 'halo' and 'undershoot column' which are used for calibration and as a margin against uncertainty in the aperture selection. Right: Synthetic image of stars, produced from the KIC, the pixel response function (PRF) and a zodiacal light model.

shapes are specific to each object and are selected to avoid sources of background contamination, while maximising flux from the target. An optimal aperture does not collect all of the flux from its target since inclusion of the target's faint point spread function (PSF) wings degrades the signal to noise. However, this leads to a setup where any effect causing a change of the target focus or position on the CCD leads to significant changes in the amount of flux falling with aperture. Since the aperture shapes are complex and irregular, it is challenging to map the alignment changes onto resulting flux changes.

As discussed in Section 1.4.2, there are several known sources of systematic effects. The differential velocity aberration introduces variations of up to 0.6 pixels over a 93 day period Jenkins et al. (2010a), which can lead to a change in flux of order a few percent. The variable guide stars introduce variations of ~ 1 millipixel, and the focus changes associated with the momentum desaturation activities produce much smaller, high frequency variations.

Figure 2.2 shows 30 random example *Kepler* light curves from Q1, and demonstrates dominant systematic trends which are obvious in all but the most intrinsically variable stars.

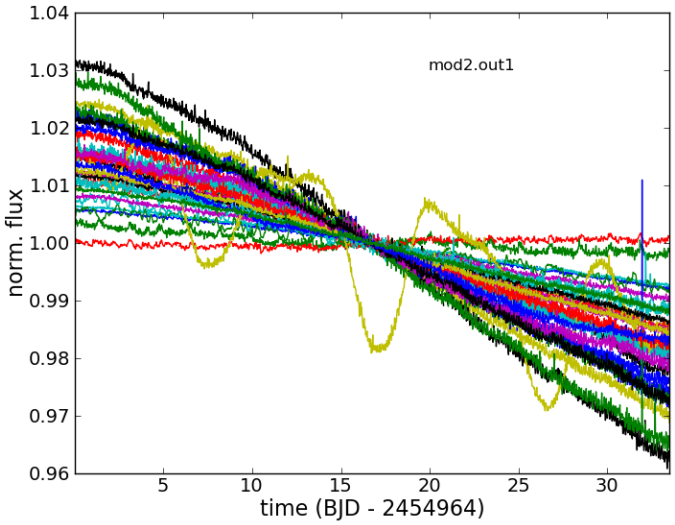

Figure 2.2: A random sample of 30 raw light curves from *Kepler* Q1 data. Each light curve was normalised by dividing it by its median.

2.2.2 *Pre Search Data Conditioning (PDC)*

The *Kepler* Science Processing Pipeline (Jenkins et al., 2010b) consists of several stages, including pixel-level calibration (CAL, Quintana et al. 2010), photometric analysis (PA, Twicken et al. 2010b) and pre-search data conditioning (PDC, Twicken et al. 2010a). From the satellite, raw pixel data is downloaded, and CAL produces calibrated target and background pixels with uncertainties. The PA module fits and removes sky background and cosmic rays, and performs simple aperture photometry, producing a light curve. This output is henceforth referred to as the raw flux, to which systematics removal techniques are applied. The PDC removes instrumental systematics, in a way that is optimised for transit searches. It also corrects discontinuities caused by cosmic rays, removes erroneous outliers and fills gaps in the data.

The PDC cotrends each light curve against ancillary engineering data such as focal-plane temperature and target centroid motion to remove systematics correlated with these measurements. The raw flux and the matrix of ancillary measurements are filtered into low-, mid- and high-pass components. Sections of light curve affected by large amplitude flux changes (e.g. eclipses, flares, microlensing events) are identified, and those not corresponding to known spacecraft anomalies are removed.

A linear least squares fit of the filtered ancillary design matrix columns to the raw flux time series for each target is performed. During this process, a Singular Value Decomposition (SVD) is applied to the matrix of ancillary data to identify strong, independent trends. To mitigate the effects of targets with high levels of intrinsic stellar variability, light curves with a centre-to-peak flux variation above 0.5% of the median target flux are flagged. An attempt is made to fit and remove the intrinsic variability with a superposition of phase shifting harmonics, which are then replaced after cotrending of the residuals. After correction, a decision is made for each target on whether to use the standard cotrending result or the one with harmonics removed and re-added, based on a comparison of the power at short timescales in the raw and corrected flux.

It is this decision point and the attempt to disentangle intrinsic variability and systematic effects, that leads to the largest flaw in the PDC, along with the tendency for the least squares approach to over-fit the light curves. A detailed study of stellar variability in the *Kepler* Q1 data (Chapter 3) shows that this method leads to a dearth of stars at medium levels of intrinsic variability (see Figure 3.1 in Section 3.2), presumably because real variability fell below the threshold and was therefore removed by the correction. Another undesired effect of PDC correction was to insert high frequency noise into the corrected data, because the trends themselves contain noise.

The aim of the PDC was to prepare light curves for transit searches to be performed, and hence in the early days it was not optimised for stellar variability studies.

2.3 *kepler*: PDC-MAP AND ALTERNATIVES

2.3.1 Pre-Search Data Conditioning - Maximum A Posteriori (PDC-MAP)

To address the shortfalls of the PDC, the *Kepler* team began work on a version called PDC-MAP (Section 2.3.1), and independently, the Oxford group began work on an Astrophysically Robust Correction (ARC), described in Section 2.4. Discussion between the groups have since lead to some features from the ARC being incorporated

in the PDC-MAP.

The PDC has recently been updated to the PDC-Maximum A Posteriori (PDC-MAP) (Stumpe et al., 2012; Smith et al., 2012). The two key differences are that cotrending is performed using the ensemble of light curves on the same channel instead of the ancillary engineering data, and that a Bayesian Maximum A Posteriori approach (DeGroot, 1970) is used instead of least squares fitting.

The core assumption is that systematic trends are highly correlated between targets on the same channel. An approximate scale of stellar variability is measured after removal of a 3rd order polynomial, and the most variable stars are excluded from the set used to find the systematics. From the remaining light curves, basis vectors are generated from SVD of the 50% most correlated on each channel, in order to exclude those with strong intrinsic variability. The PDC-MAP also utilises our entropy criterion to further eliminate individual light curves that dominate the basis vector signal (see Section 2.4.3).

Least squares fitting is still used to match basis trends to the target light curves, only this time, prior information is used to constrain the fit and prevent overfitting. This prior information is built on the assumption that nearest targets on the CCD, which are also similar in magnitude, will display the same systematic trends. Therefore, only trends also present in the nearest neighbours, defined in terms of RA, Dec and K_p, will be removed, decreasing the potential for overfitting.

The least squares fit coefficient for each basis vector j to each light curve i is Θ_j^i. The posterior probability density function $p(\Theta|y)$ is maximised for flux values, y, given prior probability density function $p(\Theta)$ generated from fit coefficients of all neighbouring targets, weighted by their distance in parameter space from the target (equation 1 from Stumpe et al. (2012)):

$$p(\Theta|y) = \frac{p(y|\Theta) \cdot p(\Theta)}{p(y)} = \frac{p(y|\Theta) \cdot p(\Theta)}{\int p(y|\Theta)p(\Theta)d\Theta} \tag{2.1}$$

This improvement is robust against intrinsic stellar variability, removing the requirement for harmonic fitting and removal.

An issue that persists with the PDC-MAP is the mixing of high and low frequency trends in the basis functions, whereas the ARC allows them to be separated. Therefore,

a light curve corrected by removal of a long term trend will gain high frequency noise if the accompanying high frequency systematic is not present. However, PDC-MAP does allow for the inclusion of sharply varying trends (like those introduced by the EB guide star, or heater cycling), where the EMD step (described in Section 2.4.5) in the ARC method does not.

2.4 *kepler*: ASTROPHYSICALLY ROBUST CORRECTION (ARC)

2.4.1 *Introduction to ARC*

When the first *Kepler* data was released, I began a study of the stellar variability in Q1 (Chapter 3), from which I concluded that the PDC removed too much real stellar signal to be used in this way. In collaboration with Prof. Stephen Roberts from the Oxford Department of Engineering Science, Dr Suzanne Aigrain and I developed an alternative systematics correction method, the ARC.

Like the methods mentioned in Section 2.1.2, we started with a linear basis model, and identified the trends from the light curves themselves, as is done in SysRem and TFA. However, our approach differs in the following important ways:

- we perform the linear basis regression using the Variational Bayes (VB, described in Appendix Section 2.A.4) method of approximate Bayesian inference to ensure that the procedure is robust to uncertainties in the data, while maintaining computational tractability and scalability to large datasets;
- we apply automatic relevance determination (ARD), sparsity-inducing priors to the coefficients of the basis vectors, which ensures that trends are identified and removed only if there is significant evidence for them in the data;
- we formulate an explicit criterion, based on a measure of the entropy of each trend, to reflect our belief that the trends are systematic i.e. that they are present, at some level, in the majority of light curves, and that the contribution of any single light curve to any given trend should be small;

- we incorporate a de-noising step after the identification of each trend, which uses the empirical mode decomposition (EMD) algorithm, chosen to avoid any intrinsic bias towards trends on particular timescales.

2.4.2 *Method: Candidate Trend Identification*

We start by considering each light curve, $\mathbf{d}_m$ to be composed of a linear combination of the intrinsic stellar signal, $\mathbf{s}_m$, an unknown number, K, of systematic trends, $\mathbf{u}_k$, and a white noise term, $\boldsymbol{\epsilon}_m$:

$$\mathbf{d}_m = \mathbf{s}_m + \sum_{k=1}^{K} a_{mk}\mathbf{u}_k + \boldsymbol{\epsilon}_m \tag{2.2}$$

where the unknown factors a_{mk} represent the contribution of the k^{th} systematic trend to the m^{th} light curve.

We seek to model as much of each light curve, $\mathbf{d}_m$ as possible, using an ensemble of other light curves from the same CCD channel (i.e. those that share the same CCD and readout electronics). We compared results using light curves from across the whole detector to those obtained using individual channels and found the latter to be more effective, since they share the same CCD location and readout mechanism and hence have the most similar systematics. In practice we found that using 200 light curves reduced computational time with a negligible effect on the result. This ensemble is defined as $\mathbf{D}_m = \{\mathbf{d}_{l\neq m}\}$.

We use a Bayesian linear model with inference performed using variational Bayes (see Appendix 2.A) to model each light curve as a sum of candidate trends:

$$\hat{\mathbf{d}}_m = \sum_{l\neq m} w_{ml}\mathbf{d}_l \tag{2.3}$$

where the w_{ml} are a set of weights. This process is repeated for each light curve, resulting in a set of candidate trends.

2.4.3 *Method: Trend Entropy*

One problem encountered in many trend removal algorithms is the tendency for a high amplitude real signal to be removed because of a chance match with another high amplitude real signal in the batch of observations, giving it a very high weighting. To prevent this issue and to enforce the condition that systematic trends are defined to be present in the majority of the light curves, the ARC makes use of an entropy criterion.

If $\hat{\mathbf{d}}_m$ represents a linear combination of 'true' systematic trends, it should be composed of many small contributions from many of the other light curves, rather than a few dominant contributions from a small number of light curves. More formally, the distribution of the weights $\mathbf{w}_m = \{w_{ml}\}$ should have a high Shannon entropy (Shannon, 1951):

$$\mathcal{H}(\mathbf{w}_m) = -\sum_{l \neq m} p_{ml} \log_2 p_{ml}, \tag{2.4}$$

where

$$p_{lk} = \frac{w_{ml}^2}{\sum_{l' \neq m} w_{ml'}^2} \tag{2.5}$$

by analogy with a normalised probability distribution.

Each candidate trend is then ranked according to the entropy $\mathcal{H}(\mathbf{w}_m)$ of the associated set of weights. The $\hat{\mathbf{d}}_m$ with the highest entropy are expected to be mutually similar, since they all represent a *systematic* trend. On the other hand, those with the lowest entropy correspond to cases where a particular light curve was found to be very similar to one or two others, but not to the rest. This is illustrated in the case of the *Kepler* Q1 data by Figure 2.3. A reduced basis set $\mathbf{T}$ is formed by selecting the 10 candidate trends with the highest entropy. The high entropy trends are mutually similar so the exact number selected is arbitrary.

2.4.4 *Method: Principal Component Analysis and Spectral Radius*

The next step is to decompose $\mathbf{T}$ using principal component analysis (PCA). This has two advantages: first, it projects all the members of $\mathbf{T}$ onto a single candidate trend

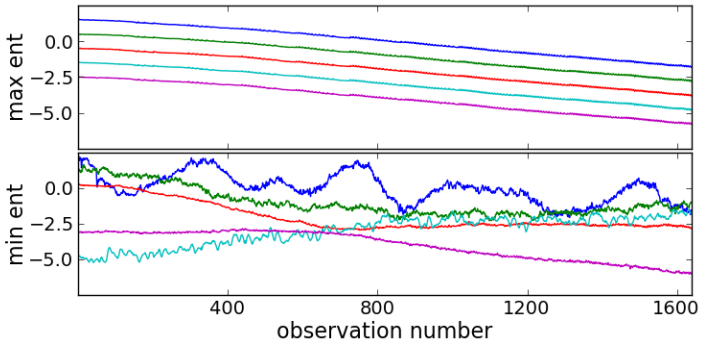

Figure 2.3: 5 Maximum (top) and minimum (bottom) entropy candidate trends for the *Kepler* Q1 data (from all Q1 channels). The trends have been normalised to have zero mean and unit variance. Vertical offsets have been added.

(the first principal component); second, it shows the proportion of the overall variance of $\mathbf{T}$, which this component represents.

$\mathbf{T}$ is decomposed into a pair of orthonormal matrices of singular components, $\mathbf{U}$, and $\mathbf{V}$, and a diagonal matrix $\mathbf{S}$ of singular values, σ_i. The latter are the square roots of λ_i, the eigenvalues of $\mathbf{T}\mathbf{T}^{\mathrm{T}}$.

$$\mathbf{T} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}. \tag{2.6}$$

$\mathbf{U}$ is also the matrix of eigenvectors of $\mathbf{T}\mathbf{T}^{\mathrm{T}}$:

$$\mathbf{T}\mathbf{T}^{\mathrm{T}} = \mathbf{U}\mathbf{S}\mathbf{S}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}. \tag{2.7}$$

The *spectral radius* of each eigenvector,

$$\rho_i = \frac{\lambda_i}{\sum_i \lambda_i}, \tag{2.8}$$

is then a measure of the fraction of the overall variance of $\mathbf{T}$ explained by that eigenvector.

If all the candidate trends in $\mathbf{T}$ are truly self-similar, the spectral radius of the first principle component should be close to unity. The candidate trends will be self-similar if they are true systematics, and therefore a threshold on the spectral radius can be used to determine whether the first principle component of $\mathbf{T}$ should be adopted as one of the systematic trends, $\mathbf{u}_k$. A threshold of $\rho > 0.8$ was found to produce reliable results.

2.4.5 *Method: Empirical Mode Decomposition (EMD)*

Since the discovered systematic trends $\mathbf{u}_k$ are linear combinations of a finite number of noisy light curves, they may still contain significant amounts of noise. Using them directly may therefore introduce noise into the light curves, particularly those of bright stars. A simple smoothing or filtering method to remove this noise would risk removal of real systematic signal, and introduce a frequency selection.

A technique to remove noise without selecting a frequency range, or assuming harmonicity or linearity is Empirical Mode Decomposition (EMD, Huang et al. 1998). This decomposes a time-series into a series of intrinsic modes by detecting successive maxima and minima. These turning points are used to define upper and lower envelopes to the signal by fitting of a spline curve which passes through the turning-point locations. The mid-point of these envelopes is then used as a baseline which is removed from the signal. This procedure is iterated until the baseline function is flat (to within a threshold). The resultant waveform at this point forms an intrinsic mode, which is then removed from the original time series.

The entire process is then repeated, successively extracting intrinsic modes until the residual has no turning points (see Figure 2.4 for example). The intrinsic mode with the largest variance is then adopted as the systematic trend $\mathbf{u}_k$.

2.4.6 *Method: Overview of Steps*

Starting with a batch of light curves, the trend identification and removal process is as follows:

1. identify a set of candidate trends $\{\hat{\mathbf{d}}_m\}$, each associated with entropy $\mathcal{H}(\mathbf{w}_m)$;
2. construct a candidate basis $\mathbf{T}$ from the 10 candidate trends with the highest entropy;
3. apply PCA to $\mathbf{T}$ and measure spectral radius ρ_1 of its first principal component;
4. if $\rho_1 < \rho_{\min}$, stop.

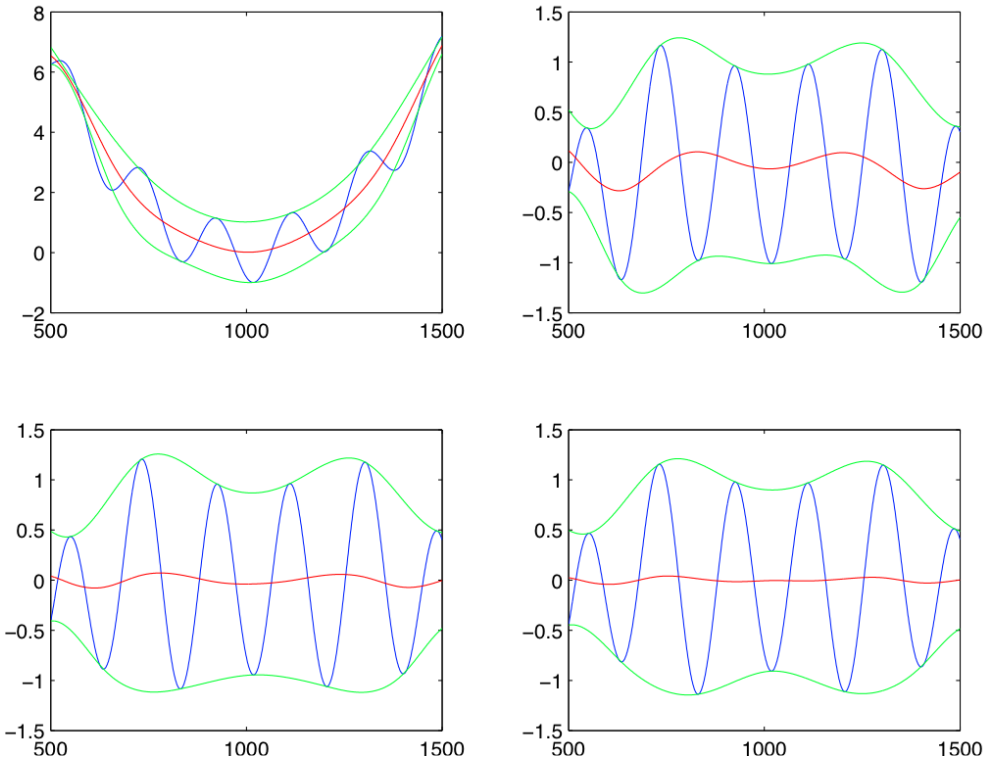

Figure 2.4: The sifting process of EMD shown over four iterations for an example synthetic dataset (top left, top right, bottom left, then bottom right). The original time series (blue, top left) is successively deflated via a baseline (red curve) formed as the mid-point of the envelope functions (green curves).

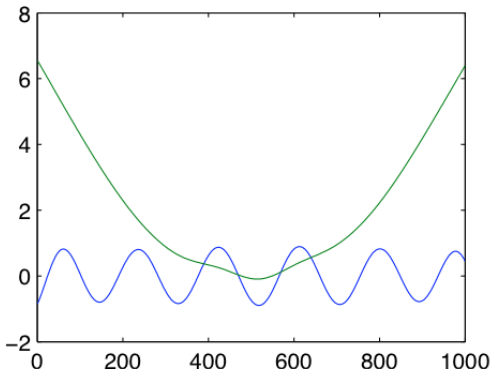

Figure 2.5: The two resultant intrinsic modes for the simple sine and quadratic example shown in Figure 2.4.

5. otherwise, perform EMD on the first principal component, and adopt the intrinsic mode with the largest variance as the next systematic trend u_k;

6. remove all the trends identified so far to produce a set of partially corrected light curves

$$\tilde{\mathbf{d}}_m = \mathbf{d}_m - \sum_{k=1}^{K} a_{mk}\mathbf{u}_k, \tag{2.9}$$

where the a_{mk} are once again found using variational Bayes with ARC priors.

7. return to step (i), with the $\{\tilde{\mathbf{d}}_m\}$ as inputs;

The process continues until $\rho_1 < \rho_{\min}$ at step 4. The choice of stopping threshold $\rho_{\min}$ is a matter of fine-tuning, and a threshold of $\rho_{\min} = 0.8$ was found to produce reliable results. The Bayesian approach used in step 6 ensures that a trend is only removed if there is evidence for it in the data.

2.4.7 *Synthetic example*

In order to demonstrate the ARC, the method was applied to a simple synthetic example dataset. Starting with a set of 200 'light curves' containing only white Gaussian noise, random amounts of two 'systematic trends' were added: an exponential decay with long time constant, and a quadratic term.

The steps of the method are illustrated in Figure 2.6. The trends were identified in a representative subset of 50 'light curves' chosen at random, and the basis $\mathbf{T}$ was constructed from the ten highest-entropy trends at each iteration. Two systematic trends were identified before the stopping criterion was reached using $\rho_{\min} = 0.6$. The stopping criterion is selected for each set through trial and error, and for the simple synthetic case 0.6 was found to be appropriate, whereas a more conservative value of 0.8 was found to be more reliable for the *Kepler* data. The resulting correction is illustrated on an example 'light curve' (not one of the subset in which the trends were identified) in Figure 2.7.

In this synthetic example one may compare the original data, prior to addition of systematic trend artefacts, with the recovered detrended traces, with knowledge of the

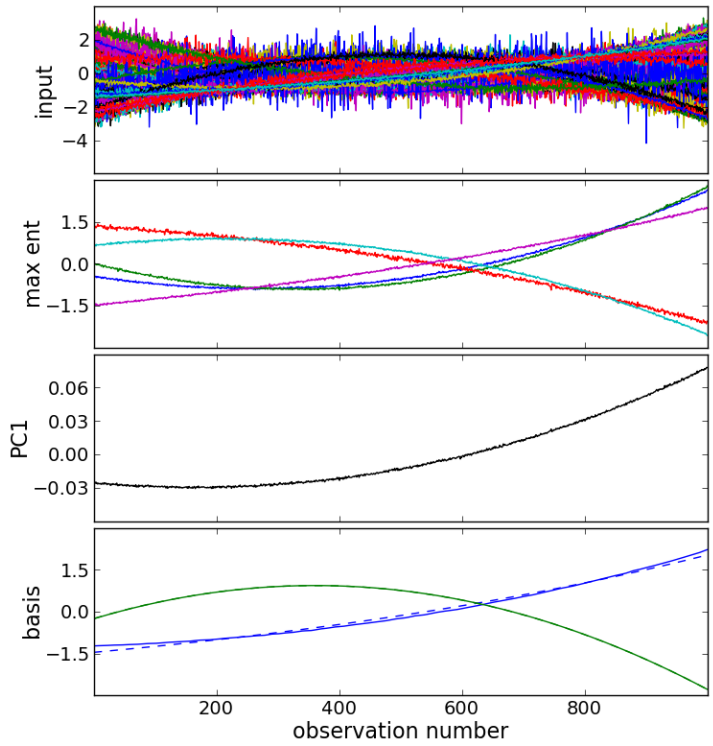

Figure 2.6: Application of the ARC to a synthetic dataset. Top panel: randomly chosen subset of input light curves. Second panel: the five highest entropy candidate trends found during the first iteration. Third panel: the first principal component of the high-entropy trends, which accounts for $> 95\%$ of the variance. Bottom panel: the two basis trends discovered by the algorithm (solid lines) and the corresponding trends injected into the data (dashed lines); note that the green solid and dashed lines are almost undistinguishable.

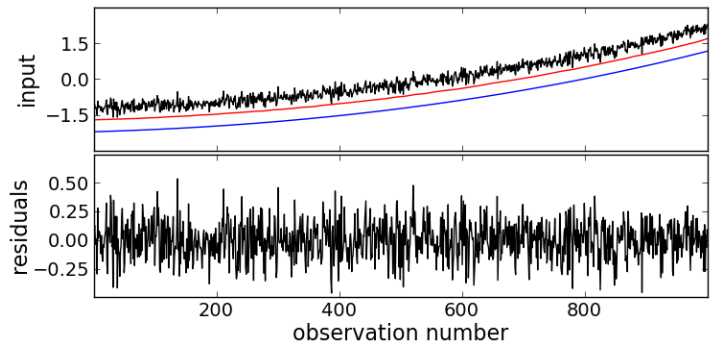

Figure 2.7: Result of the ARC applied to a synthetic dataset. Top panel: example 'light curve', with the systematics model (linear combination of the two identified trends) shown in red, and the actual trends injected in blue (with a vertical offset for clarity). Bottom panel: residuals after trend removal using the identified trends.

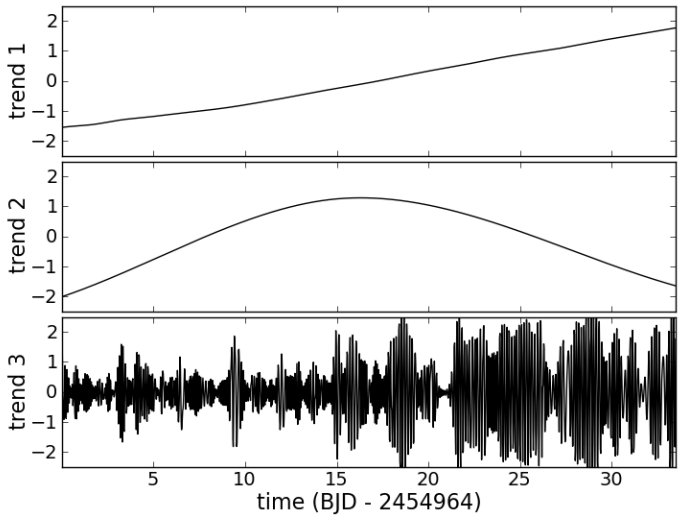

Figure 2.8: Basis trends inferred from the global set of all *Kepler* Q1 light curves.

'ground truth'. The resultant linear correlations between the true 'light curves' and those recovered by the algorithm lie between 0.95 and 0.999, with a median value of 0.994, indicating a very high reconstruction accuracy (the median correlation between the original curves and the trend-corrupted versions was only 0.1).

2.4.8 *Application to* Kepler *Q1 Data*

The method was then applied to the *Kepler* Q1 data, providing a dataset suitable for the study of stellar variability, which is described in detail and compared to the PDC corrected data in Chapter 3. In this section the results are compared to the PDC-MAP corrected data.

A test set of basis trends for all the light curves is identified using a randomly selected subset of 200 light curves from all CCD channels, and using a threshold $\rho_{\min} = 0.8$. The three significant trends are shown in Figure 2.8. The first two have low-complexity, and are roughly linear and quadratic in nature, respectively. The third is a more complex high frequency signal, but is nonetheless present in a very large number of light curves. It likely originates from an artefact due to the satellite's reaction wheel (Jon Jenkins, priv. comm.).

In practice, application of the ARC to individual channels produces better results, since the systematics vary slightly with position on the detector. The full map of basis trends per channel can be seen in Figure 2.9. It is evident that the low frequency trends vary slightly across the detector, presumably due to the differing effect of motion and differential velocity aberration. The high frequency trends are more prevalent in channels near the edges, especially in the top left corner as shown in Figure 2.9, which is mostly likely due to a physical connection between the detector and the source of this artefact in this region.

2.4.9 Method: Implementation

All the code for the ARC was written in Matlab by Prof. Stephen Roberts. The basis discovery phase for all CCD channels takes just under an hour running on a quad-core 2.6 GHz processor under a Unix operating system. The subsequent trend removal over all channels takes approximately five minutes and scales linearly with the number of light curves being detrended. Without loss of generality each light curve is normalised to have zero mean and unit variance prior to analysis, and this step is reversed after processing. The variational Bayes basis model, detailed Appendix 2.A.4, is used in both the trend discovery and trend removal phases of the algorithm, differing only in the details of the structured priors.

In the trend discovery phase of the ARC method, Equation 2.3 is implemented, with successive light curves modelled as linear combinations of one another. In this phase, a single global weight precision hyperparameter, α, is used, over which a Gamma distribution with hyper-hyperparameters is placed (see Appendices 2.A.5 – 2.A.9 for further details and values used).

In the trend removal phase individual precisions associated with the weights of the model are used, as each weight is linked to a putative (pre-discovered) trend component. This allows shrinkage on trends that are not present in a light curve and avoids falsely removing any components for which there is not compelling evidence. Therefore, the procedure for structured priors is implemented.

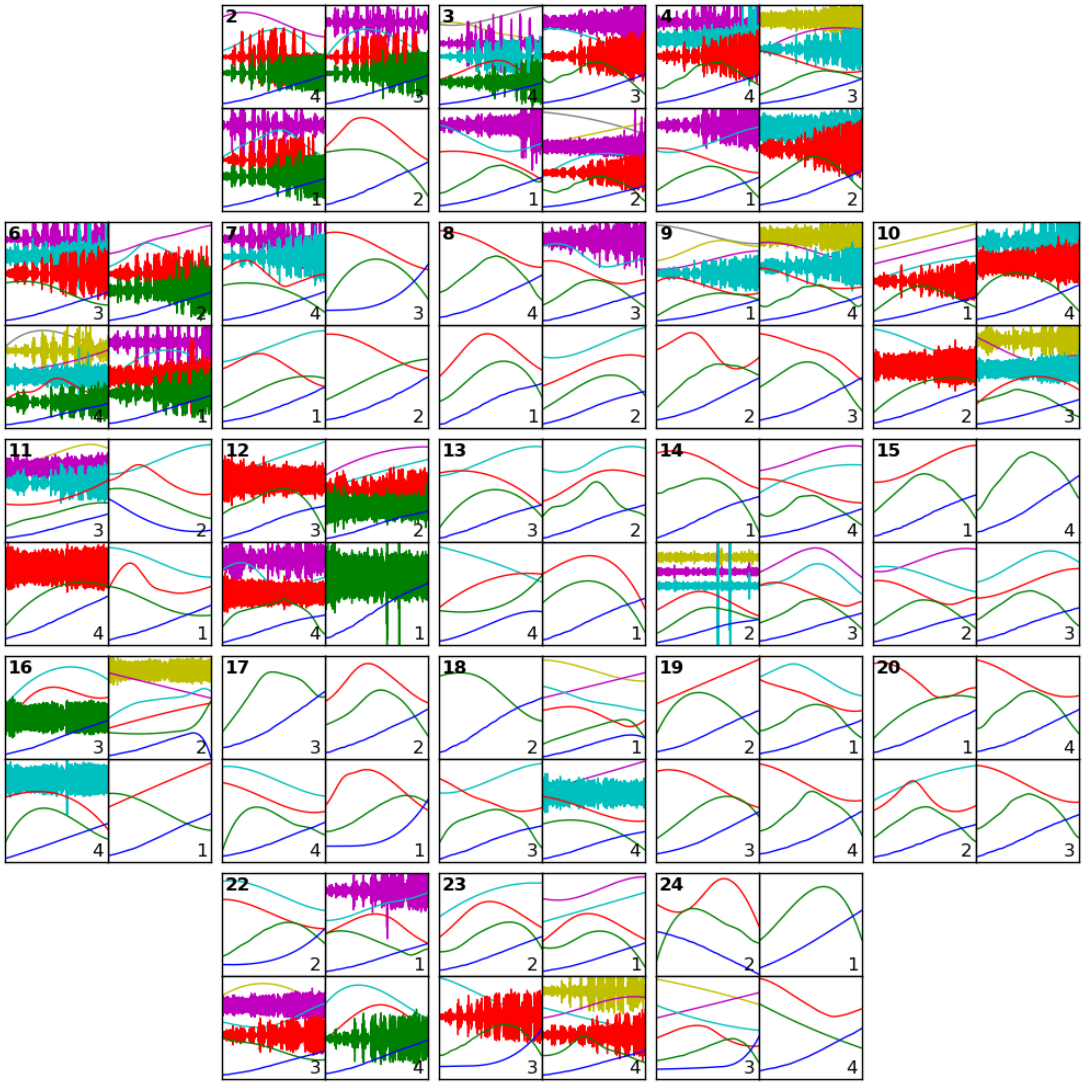

Figure 2.9: Q1 trend basis sets inferred for each channel, labelled by module (labelled in bold at the top left of each group of plots) and output (labelled at the bottom right of each panel). Each module consists of 2 CCDs and 4 outputs channels. The combination of a module and an output is labelled as a 'channel', of which there are 84. The blue, green, red, cyan, magenta, yellow and grey lines show trends 1 to 7 respectively (the number of trends identified varies). A vertical offset has been added between consecutive trends for clarity (the most important trend is always the lowest), and they are arbitrarily scaled in the vertical direction to ensure visibility. The 4 missing modules for the Fine Guidance Sensor.

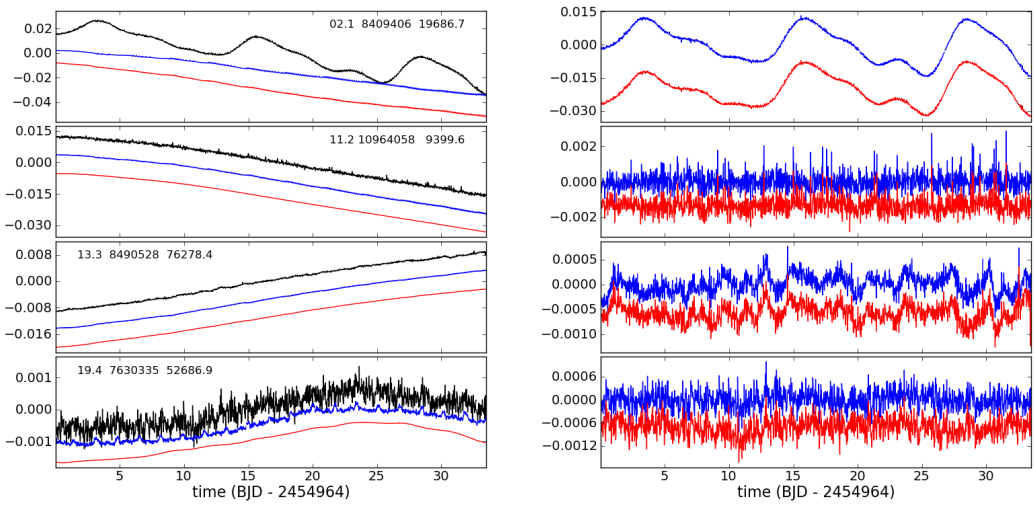

Figure 2.10: Example results from Q1. Each row corresponds to a particular target. In the left column, the raw data and the corrections applied by the PDC-MAP and ARC are shown in black, blue and red, respectively. The module and output channel number, KIC number of the target, and median flux in the raw light curve, are also shown. In the right column, the PDC-MAP and ARC corrected data are shown in blue and red respectively. Each light curve has been normalised by dividing them by their median and subtracting unity. A vertical offset has also been applied between the curves in each panel, to improve clarity.

2.4.10 *Comparison of ARC and PDC-MAP*

A selection of typical light curves from different channels, corrected using the ARC and PDC-MAP are shown in Figure 2.10. To first order they appear very similar, since the updates in the PDC-MAP also introduce a great improvement to the original PDC.

Both PDC-MAP and ARC methods successfully preserve stellar variability in most cases, and remove most of the trends which one would deem to be systematic after visually inspecting many light curves. However, there are some minor differences between the corrections applied.

In the top panel of Figure 2.10, the low-frequency component of the applied corrections differs slightly. It is very difficult to determine, on a dataset of this duration, which correction is more appropriate. In the bottom panel, the PDC-MAP attempts to correct for the effect of periodic pointing anomalies caused by the presence of an eclipsing binary included in the guide stars.

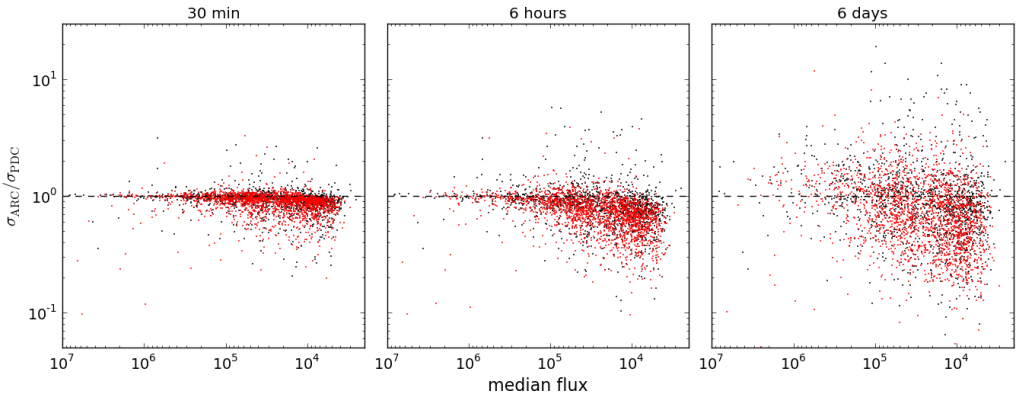

Figure 2.11: Comparison of the light curve scatter on different timescales after correction by PDC-MAP and by the ARC. Each panel shows the ratio of the light curve scatter resulting from the two correction as a function of the median flux in the light curve. The x-axis has been reversed (so brighter stars are to the left) and plotted on a log scale (so it is similar to a magnitude scale). The black points correspond to output channel 1 on module 2, where the high frequency trend featured strongly, and the red points to output channel 1 on module 13, near the centre of the detector, where only long-term trends were significant.

Although the eclipsing binary signature is present among the basis vectors identified by the PDC-MAP pipeline, it remains challenging to correct. It is not captured by the ARC, and this could be because of the EMD step: the sharp edges of the eclipses are not easily modelled by a superposition of intrinsic modes. More importantly, however, in this example the PDC-MAP also introduces significant amounts of high-frequency noise, which was not present in the original light curve.

An investigation of the relative scatter of the corrected light curves on a range of timescales reveals differences between the two methods. The normalised light curves (divided by the raw median flux) were smoothed on the timescales of interest using a median filter of the corresponding width, and the scatter estimated to be σ as 1.48 times the median of the absolute deviations from unity.

Figure 2.11 shows the ratio $\sigma_{\mathrm{ARC}}/\sigma_{\mathrm{PDC}}$ (where the subscript PDC refers to the PDC-MAP) as a function of the median flux, for two example channels, and for three different timescales: 30 minutes (the sampling of *Kepler* long-cadence observations), 6 hours (most relevant for transit detection, and 6 days (relevant, for example, to stellar rotation studies).

Gilliland et al. (2011) investigated the noise properties of the PDC data, and showed that the majority of Sun-like stars observed by *Kepler* appear more variable on 6.5-hour timescales than the Sun. This has a potentially serious impact on *Kepler*'s ability to detect transits of habitable planets. To estimate the variability on transit timescale, Gilliland et al. (2011) used a quantity known as the 6.5-hour combined differential photometric precision (CDPP). The 6-hour σ estimates in this work were obtained in a slightly different fashion to the CDPP, but measure a similar quantity, namely the light curve scatter on typical transit timescales. Although improving the detectability of transits was not the primary motive in developing the ARC, the latter appears to yield slightly lower scatter on transit timescales than the PDC-MAP, and hence the ARC may prove useful for transit searches.

The 'average' power spectra of the bright stars (median raw flux $> 10^5\,\mathrm{e}^-/\mathrm{s}$) before and after correction also reveal differences (Figure 2.12). These are simply averages of the individual power spectra of each light curve in the same two output channels. Again, both systematics correction methods significantly reduce the light curve scatter on most timescales, and their behaviour is very similar at low frequencies (except for the very lowest, but these are poorly constrained due to the limited duration of the dataset). However, the power spectra conclusively demonstrate the fact that the PDC-MAP introduces high-frequency noise into the light curves (see also Murphy, 2012). This effect is more significant in some modules than in others, but the differences with the ARC become noticeable upwards of about 2 cycles per day.

The power spectra shown in the bottom panel of Figure 2.12 correspond to the central module of the *Kepler* detector, where the only trends identified by the ARC were long-term trends. There are clearly systematic effects at well-defined frequencies, which are not well captured by either the PDC-MAP or the ARC. The signal at ~ 0.8 cycles/day and harmonics thereof corresponds to the aforementioned eclipsing binary guide star; there are also a number of higher-frequency effects. The fact that neither the PDC-MAP nor the ARC correct these effects well suggests that they may not obey one of the fundamental assumptions underlying both methods. For example, if the contribution of a given trend to a given light curve varies during the quarter, the linear basis model will not be able to capture it.

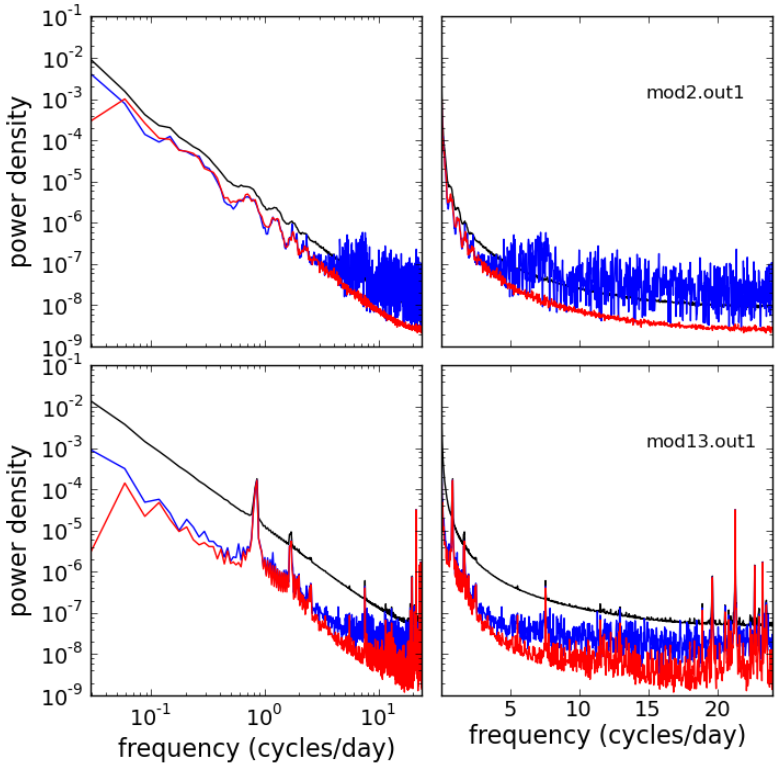

Figure 2.12: Comparison of the 'average' power spectra for bright stars before and after correction with the PDC-MAP and the ARC, in two representative output channels (top and bottom). The black, blue and red curves correspond to the raw, PDC-MAP and ARC data, respectively. The two columns show the same data, but the x-axis scale is logarithmic in one case, and linear in the other.

There are also interesting differences in the average power at a given frequency, and in the relative reduction in this power introduced by the systematics corrections, for the two channels shown. However, it would probably be unwise to read too much into this, as these power spectra are based on relatively small numbers of light curves ($\sim$ 50 in each case). See Chapter 3 for a discussion of the stellar physics represented by the average power spectra of different spectral types.

2.5 DISCUSSION AND FUTURE WORK

2.5.1 Kepler *Q1 Summary*

The ARC performs well, correcting systematic trends in the *Kepler* Q1 data, while preserving 'true' astrophysical variability. It follows established practice in using linear basis models, but incorporates a number of new features, which are intended to ensure its robustness (Bayesian framework, use of shrinkage priors), and its computational efficiency (approximate inference using variational Bayes). A de-noising step is used to prevent the introduction of spurious random noise into the light curves.

The ARC performs significantly better than the original PDC method, as can be seen in Figure 3.1 in Chapter 3. The ARC produces more similar results to the PDC-MAP (see Figure 2.10), but some important differences. The PDC-MAP tends to introduce high-frequency noise into the light curves, whereas the ARC does not. Importantly, this effect remains noticeable on the timescales typical of planetary transits.

There are types of systematics which are not well-corrected by either the PDC-MAP or the ARC, for example the trends introduced by the variable guide stars. The impact of the residual systematic effects should be relatively limited, because they are confined to well-defined frequency ranges. Further work is required on these: one possibility, which we intend to investigate, is to treat the systematics as multiplicative, rather than additive, by working in log flux units. Another possibility is to find an alternative to the EMD step, which cannot capture shapes that are not well-modelled by intrinsic modes, such as the eclipsing binary signature. Alternatively, it may be possible to adapt the EMD step to include more intrinsic modes, although this would need extensive calibration and testing to ensure noise is successfully being removed. The EMD step is important because it also ensures trends are separated into distinct features before being removed. Without it, a long term trend may contain also features of the high frequency systematic, leading to spurious effects if they are removed together from a light curve containing only the long term component.

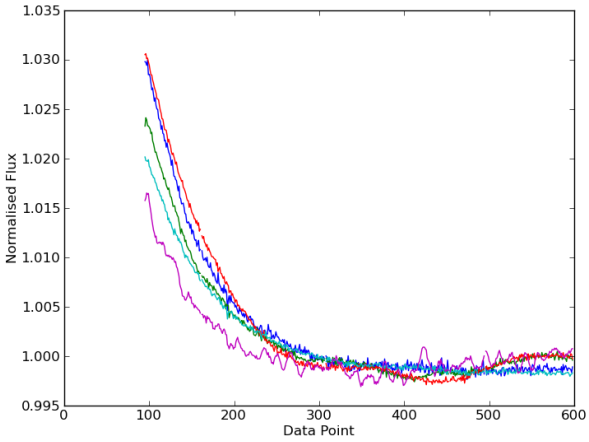

Figure 2.13: Examples of the artefacts seen in the light curves of quarter 2 onwards after data download interruptions. These 5 examples show median normalised data from around the first gap in Q2 data.

2.5.2 *Later Quarters and Jump Correction*

While Q1 is a single month of continuous data, the later quarters consist of three months of data with monthly interruptions for data download. When the satellite resumes observing after each data download, the CCD temperature takes a few days to return to normal, leading to variations in the focus which create sharp features in the light curve (see Figure 2.13). These features look systematic but detailed investigation reveals that are in fact not consistent between light curves, even within the same channel. The ARC, in it's current form, is therefore unable to remove these effects.

These features require an additional jump correction step, which models and removes the exponential decay shape in individual light curves, before the ARC can be implemented. The jump correction algorithm, written by Dr Steven Reece, is still in development. The PDC-MAP does include these jumps in the systematic trends discovered, although they are mixed with trends on other timescales and are a common source of error in the PDC-MAP corrected light curves.

The methods developed for connecting data across the download gaps can be extended to quarter stitching, although additional consideration is required for the potential change in amplitude between quarters. This arises as a result of the different

pixels and aperture associated with the target, following the 90° rotation of the satellite.

Since this stage of our pipeline is still under development, I have chosen to use the ARC corrected Q1 data for the variability work described in Chapter 3, but have opted to use the PDC-MAP corrected data for the rotation studies presented in Chapters 5, 6 and 7. This decision is motivated by the fact that the Q1 variability study is very sensitive to residual systematics, whereas the rotation period determination code if robust against low level systematics in the data, and the PDC-MAP performs much better than the original PDC.

2.5.3 *Alternative Methods*

The ARC and PDC methods are designed to provide a fast, automated and robust treatment of a large ensemble of light curves. For individual light curves, a more accurate correction is often possible by starting with the pixel-level data. Kinemuchi et al. (2012) describes the methods and data products available to the public for manual correction of individual light curves. These include the 16 Cotrending Basis Vectors (CBVs) found for each CCD channel by the PDC-MAP, and instructions of how to fit them to the raw data, using visual inspection to choose which of the CBVs should be applied.

The Appendix of Kinemuchi et al. (2012) also describes the PyKE suite of python software tools developed by the *Kepler* team to analyse *Kepler* light curves. This allows the user to examine the pixel images to identify contamination, and re-extract target pixels to reduce the effects of systematics and artefacts in the light curve.

APPENDIX 2.A VARIATIONAL BAYES

Appendix 2.A.1 Bayesian Linear Basis Model

In order to describe the Variational Bayes method, I will first show how the linear basis model from Equation 2.2 can be re-written and the maximum likelihood solution is obtained. This appendix consists of textbook formulae for explanatory purposes and is not my own derivation of results.

In a general linear basis model, the observations of a dependent variable, d, are modelled as a linear combination of basis functions of an independent variable, t, plus a noise term ϵ:

$$d(t) = \sum_{k=1}^{K} w_k \phi_k(t) + \epsilon, \tag{2.10}$$

where the w's are the factor weights and the ϕ's are the basis functions. The noise term, ϵ, is taken to be drawn from a normal (Gaussian) distribution with zero mean and precision (inverse variance) β.

Defining the column vectors $\mathbf{w} \equiv [w_1, \ldots, w_K]^{\mathrm{T}}$ and $\boldsymbol{\phi}(t) \equiv [\phi_1(t), \ldots, \phi_K(t)]^{\mathrm{T}}$, Equation (2.10) can be written:

$$d(t) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(t) + \epsilon. \tag{2.11}$$

From this, a system of simultaneous linear equations can be used to describe the dataset as a whole:

$$\mathbf{d} = \mathbf{w}^{\mathrm{T}} \boldsymbol{\Phi} + \boldsymbol{\epsilon}, \tag{2.12}$$

where $\boldsymbol{\Phi}$ is an $N \times K$-element matrix with elements $\phi_k(t_n)$ and $\boldsymbol{\epsilon} \equiv [\epsilon_1, \ldots, \epsilon_N]^{\mathrm{T}}$, with N being the number of data points.

The likelihood, i.e. the probability that the model gave rise to the observations, is then simply

$$p(d|\mathbf{w}, \beta) = \mathcal{N}(\mathbf{d}; \mathbf{w}^{\mathrm{T}} \boldsymbol{\Phi}, \beta^{-1} \mathbf{I}) \tag{2.13}$$

where $\mathcal{N}(\mathbf{a}; \mathbf{b}, \mathbf{C})$ is the probability of drawing a vector $\mathbf{a}$ from a multi-variate Gaussian distribution with mean vector $\mathbf{b}$ and covariance matrix $\mathbf{C}$, and $\mathbf{I}$ is the identity matrix.

Appendix 2.A.2 Maximum Likelihood

The maximum likelihood (ML) solution for the weights is given by the standard pseudo-inverse equation (see Chapter 3.1.1 of Bishop 2006 for full derivation):

$$\mathbf{w}_{\mathrm{ML}} = (\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{d}. \tag{2.14}$$

However, the ML approach is prone to over-fitting. Introducing a prior over the weights and obtaining the maximum a posteriori (MAP) solution helps reduce this problem, but this approach still relies on point estimates of the parameters of the models, and consequently fails to take into account some of the intrinsic uncertainties.

A full Bayesian solution is obtained by marginalising over the posterior distributions of the variables, which can be achieved, for example, using sample-based approaches such as Markov-Chain Monte-Carlo (MCMC). These can be computationally intensive however, and they scale poorly to large numbers of parameters, and it is often problematic to establish whether convergence has been reached.

Variational Bayes (VB) is an alternative approach which uses approximate Bayesian inference, and offers computational tractability even on very large data sets. A full tutorial on VB is given in Bishop (2006).

Appendix 2.A.3 The Full Bayesian Model

In a Bayesian framework, the priors over the parameters of the model must be specified. For each of the weights, a zero-mean Gaussian prior with precision α is adopted. The zero mean ensures that the weight associated with a given basis function is only non-zero if the data requires it to be. The Gaussian is the least informative choice of distribution or quantity that can be either positive or negative, and is therefore appropriate when little information is known about the weights a priori. It is also convenient to chose a form for the prior that is conjugate with the likelihood, because it makes the marginalization process analytic. The joint prior over all the weights is

thus:

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{0}, \alpha^{-1}\mathbf{I}). \tag{2.15}$$

The value of α does not enter into the likelihood, but it does control the parameters, and it is therefore known as a hyper-parameter. In a fully Bayesian treatment, rather than fixing α to a specific value, one should marginalise over it, under a so-called hyper-prior. It is convenient to select a hyper-prior for α that is conjugate with the prior over the weights, because then the marginalisation can be done analytically. As the prior is a Gaussian, and α should always be positive (because it represents an inverse variance), a suitable form for the hyper-prior is a Gamma distribution, as the least informative for a positive quantity, with shape parameter a and scale parameter b:

$$p(\alpha) = \mathcal{G}(\alpha; a, b) \equiv \frac{1}{\Gamma(a) b^a}\, \alpha^{a-1}\, e^{-\alpha/b}, \tag{2.16}$$

where $\Gamma(x)$ is the Gamma function.

The noise precision, β, is also a parameter of the model, for which a prior must be adopted. Again, a Gamma distribution is a suitable choice:

$$p(\beta) = \mathcal{G}(\beta; c, d). \tag{2.17}$$

All the parameters and hyperparameters of the model can be represented by a vector $\boldsymbol{\theta} = [\mathbf{w}, \alpha, \beta]^{\mathrm{T}}$. As the weights depend upon the scale α but not the noise precision β, the joint prior distribution over $\boldsymbol{\theta}$ factorises as:

$$p(\boldsymbol{\theta}) = p(\mathbf{w}|\alpha)\, p(\alpha)\, p(\beta). \tag{2.18}$$

The objective in Bayesian regression is to estimate the posterior distribution, which fully describes our knowledge regarding the parameters of the model, $\boldsymbol{\theta}$, given the data $\mathbf{d}$. This is related to the likelihood and prior by Bayes's theorem:

$$p(\boldsymbol{\theta}|\mathbf{d}) = \frac{p(\mathbf{d}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathbf{d})}. \tag{2.19}$$

The denominator is the evidence or marginal likelihood of the data under the model, and is given by

$$p(\mathbf{d}) = \int p(\mathbf{d}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta} \tag{2.20}$$

More than a mere normalisation term, the evidence is a quantitative measure of the extent to which the data supports the model.

Appendix 2.A.4 Variational Bayesian Inference

The functional dependence of the likelihood and posterior distribution on the parameters are generally unknown. An analytical approximation for one of them can be used in order to access the model evidence or posterior distribution over individual parameters. This is know as approximate inference, which is described in detail in Chapter 10 of Bishop (2006).

In variational inference, an analytically tractable distribution $q(\boldsymbol{\theta}|\mathbf{d})$ is introduced, which is used to approximate the posterior distribution $p(\boldsymbol{\theta}|\mathbf{d})$. The log evidence can then be written as the sum of two separate terms by introducing the term $\int q(\boldsymbol{\theta}|\mathbf{d})\mathrm{d}\boldsymbol{\theta}$, which, as the integral of a probability density function, equals 1:

$$\begin{aligned}
\log p(\mathbf{d}) &= \log p(\mathbf{d}) \int q(\boldsymbol{\theta}|\mathbf{d})\,\mathrm{d}\boldsymbol{\theta} && (2.21)\\
&= \int q(\boldsymbol{\theta}|\mathbf{d})\,\log p(\mathbf{d})\,\mathrm{d}\boldsymbol{\theta} && (2.22)\\
&= \int q(\boldsymbol{\theta}|\mathbf{d})\,\log\left[\frac{p(\mathbf{d},\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{d})}\right]\,\mathrm{d}\boldsymbol{\theta} && (2.23)\\
&= \int q(\boldsymbol{\theta}|\mathbf{d})\,\log\left[\frac{p(\mathbf{d},\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{d})}\cdot\frac{q(\boldsymbol{\theta}|\mathbf{d})}{q(\boldsymbol{\theta}|\mathbf{d})}\right]\,\mathrm{d}\boldsymbol{\theta} && (2.24)\\
&= \int q(\boldsymbol{\theta}|\mathbf{d})\log\frac{p(\mathbf{d},\boldsymbol{\theta})}{q(\boldsymbol{\theta}|\mathbf{d})}\,\mathrm{d}\boldsymbol{\theta}\\
&\qquad - \int q(\boldsymbol{\theta}|\mathbf{d})\log\frac{p(\boldsymbol{\theta}|\mathbf{d})}{q(\boldsymbol{\theta}|\mathbf{d})}\,\mathrm{d}\boldsymbol{\theta} && (2.25)\\
&\equiv \mathcal{F}(p,q) + \mathrm{KL}(p,q). && (2.26)
\end{aligned}$$

Equation 2.24 is obtained using the definition of conditional probability, $p(\mathrm{d}|\theta) = p(\mathrm{d}\theta)p(\mathrm{d})$. Equation 2.26 is the fundamental equation of the VB-framework. The first term on the right-hand side, $\mathcal{F}(p,q)$, is known as the (negative) variational free energy. The second, $\mathrm{KL}(p,q)$ is the Kullback-Leibler (KL) divergence between the approximate posterior q and the true posterior p. Importantly, the KL divergence is always positive. $\mathcal{F}(p,q)$ thus provides a strict lower bound on the log evidence. Moreover, because the KL divergence is zero when the two densities are the same, $\mathcal{F}(p,q)$ will become equal to the log evidence when the approximating posterior is equal to the true posterior, i.e. when $q(\boldsymbol{\theta}|\mathbf{d}) = p(\boldsymbol{\theta}|\mathbf{d})$.

The aim of VB learning is therefore to maximise $\mathcal{F}(p,q)$ and so make the approxim-

ate posterior as close as possible to the true posterior. This requires the extremisation of an integral with respect to a functional, which is typically achieved using the calculus of variations.

In order to implement this method, the range of proposal posterior distributions over which the optimisation is performed, are restricted to those to the exponential family. This ensures that extremisation over the function $q(\boldsymbol{\theta})$ can be replaced exactly by extremisation with respect to the parameters $\boldsymbol{\theta}$. From this, one obtains a set of coupled update equations over the parameters which are cycled until a convergence criterion is met.

Furthermore, the posterior distribution can be assumed to be separable, meaning that it can be written as a product of independent functions of different parameters (or subsets of the parameters).

$$q(\boldsymbol{\theta}|\mathbf{d}) = \prod_i q(\theta_i|\mathbf{d}). \tag{2.27}$$

This makes the resultant inference algorithm computationally very rapid, with little loss of information. The aim is to find the distribution $q(\boldsymbol{\theta}|\mathbf{d})$ for which $\mathcal{F}(p,q)$ is largest, which is done by optimising each of the factors in turn. We start by substituting this separable expression into $\mathcal{F}(p,q)$:

$$\mathcal{F}(p,q) = \int \prod_i q(\theta_i|\mathbf{d}) \left[\log p(\mathbf{d},\boldsymbol{\theta}) - \sum_i \log q(\theta_i|\mathbf{d})\right] \mathrm{d}\boldsymbol{\theta}. \tag{2.28}$$

Next, we write down separately the terms which depend on one of the parameters, θ_j:

$$\mathcal{F}(p,q) = \iint q(\theta_j|\mathbf{d}) \prod_{i\neq j} q(\theta_i|\mathbf{d}) \left[\log p(\mathbf{d},\boldsymbol{\theta}) - \log q(\theta_j|\mathbf{d}) - \sum_{i\neq j} \log q(\theta_i|\mathbf{d})\right] \mathrm{d}\theta_j \mathrm{d}\boldsymbol{\theta}_{(i\neq j)}. \tag{2.29}$$

We then rearrange the above equation, so as to isolate the dependency on θ_j:

$$\begin{aligned}
\mathcal{F}(p,q) &= \int q(\theta_j|\mathbf{d}) \left\{ \int \log p(\mathbf{d},\boldsymbol{\theta}) \prod_{i\neq j} q(\theta_i|\mathbf{d}) \, \mathrm{d}\boldsymbol{\theta}_{(i\neq j)} \right\} \mathrm{d}\theta_j \\
&- \int q(\theta_j|\mathbf{d}) \, \log q(\theta_j|\mathbf{d}) \, \mathrm{d}\theta_j \int \prod_{i\neq j} q(\theta_i|\mathbf{d}) \, \mathrm{d}\boldsymbol{\theta}_{(i\neq j)} \\
&- \int \prod_{i\neq j} q(\theta_i|\mathbf{d}) \sum_{i\neq j} \log q(\theta_i|\mathbf{d}) \, \mathrm{d}\boldsymbol{\theta}_{(i\neq j)} \int q(\theta_j|\mathbf{d}) \, \mathrm{d}\theta_j.
\end{aligned} \tag{2.30}$$

The second factor in each of the last two terms in Equation 2.30 is the integral of a probability distribution, and is thus equal to one. Furthermore, the third term is

independent of θ_j, so:

$$\begin{aligned}\mathcal{F}(p,q) &= \int q(\theta_j|\mathbf{d}) \left\{ \int \log p(\mathbf{d},\boldsymbol{\theta}) \prod_{i\neq j} q(\theta_i|\mathbf{d}) \, \mathrm{d}\boldsymbol{\theta}_{(i\neq j)} \right\} \mathrm{d}\theta_j \\ &- \int q(\theta_j|\mathbf{d}) \, \log q(\theta_j|\mathbf{d}) \, \mathrm{d}\theta_j + \text{const}(j) \end{aligned} \tag{2.31}$$

where const(j) represents a term that is constant with respect to θ_j. The quantity in curly braces in the first term of Equation 2.31 is the expectation of $\log p(\mathbf{d},\boldsymbol{\theta})$ under the candidate posterior distribution for all the other parameters, $\prod_{i\neq j} q(\theta_i|\mathbf{d})$. We wish to isolate the part of it, which depends on θ_j. To this end, we define the quantity

$$f(\theta_j) \stackrel{\text{def}}{=} \int \log \left[p(\mathbf{d},\boldsymbol{\theta}) \right]_j \prod_{i\neq j} q(\theta_i|\mathbf{d}) \, \mathrm{d}\boldsymbol{\theta}_{(i\neq j)}, \tag{2.32}$$

where $\left[p(\mathbf{d},\boldsymbol{\theta}) \right]_j$ contains all the terms in $p(\mathbf{d},\boldsymbol{\theta})$, which depend on θ_j. We can then write

$$\int \log p(\mathbf{d},\boldsymbol{\theta}) \prod_{i\neq j} q(\theta_i|\mathbf{d}) \, \mathrm{d}\boldsymbol{\theta}_{(i\neq j)} = f(\theta_j) + \text{const}(j) \tag{2.33}$$

and thus Equation 2.31 becomes:

$$\mathcal{F}(p,q) = \int q(\theta_j|\mathbf{d}) \, \left[f(\theta_j) + \text{const}(j) \right] \, \mathrm{d}\theta_j - \int q(\theta_j|\mathbf{d}) \, \log q(\theta_j|\mathbf{d}) \, \mathrm{d}\theta_j + \text{const}(j). \tag{2.34}$$

The non-constant terms in Equation 2.34 can readily be identified as the negative KL divergence between $q(\theta_j|\mathbf{d})$ and $\exp\left[f(\theta_j) + \text{const}(j) \right]$. This divergence is minimised (and is zero) when the two distributions are the same. Therefore, the free energy $\mathcal{F}(p,q)$ is maximised with respect to θ_j simply by setting

$$\log q^{\text{new}}(\theta_j|\mathbf{d}) = f^{\text{old}}(\theta_j) + \text{ const}(j). \tag{2.35}$$

The additive constant in Equation 2.35 is set by normalising the distribution $q^{\text{new}}(\theta_j|\mathbf{d})$. Thus, if we take the exponential of both sides and normalise, we obtain the updated distribution for parameter θ_j as:

$$q^{\text{new}}(\theta_j|\boldsymbol{d}) = \frac{\exp\left[f^{\text{old}}(\theta_j) \right]}{\int \exp\left[f^{\text{old}}(\theta_j) \right] \mathrm{d}\theta_j}. \tag{2.36}$$

As previously mentioned, we have chosen to use candidate distributions, which belong to the exponential family, so that extremisation over the function $q(\boldsymbol{\theta}|\mathbf{d})$ can be

replaced exactly by extremisation with respect to the parameters $\boldsymbol{\theta}$, and thus we can re-write Equation 2.36

$$q(\theta_j^{\text{new}}|\boldsymbol{d}) = \frac{\exp\left[f(\theta_j^{\text{old}})\right]}{\int \exp\left[f(\theta_j^{\text{old}})\right] \mathrm{d}\theta_j}. \tag{2.37}$$

We note that, as with the EM algorithm, each iteration is *guaranteed* to improve the marginal likelihood of the data under the model.

For the model used in this work,

$$p(\mathbf{d}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}) = p(\mathbf{d}|\mathbf{w}, \beta)\, p(\mathbf{w}|\alpha)\, p(\beta)\, p(\alpha). \tag{2.38}$$

Like in the EM algorithm, each iteration is guaranteed to improve the marginal likelihood of the data under the model.

Appendix 2.A.5 Updating the Weights $\mathbf{w}$

The marginal for the weights is

$$f(\mathbf{w}) = \iint q(\beta|\mathbf{d})\, q(\alpha|\mathbf{d}) \log\left[p(\mathbf{d}|\mathbf{w}, \beta)\, p(\mathbf{w}|\alpha)\right]\, \mathrm{d}\alpha\, \mathrm{d}\beta \tag{2.39}$$

The update equations are hence of the form

$$q(\mathbf{w}|\mathbf{d}) \propto \exp[f(\mathbf{w})] \tag{2.40}$$

Substituting for the terms in Equation 2.39 gives

$$\begin{aligned} f(\mathbf{w}) &= -\int q(\beta|\mathbf{d})\, \frac{\beta}{2}\, (\mathbf{d} - \boldsymbol{\Phi}\mathbf{w})^{\mathrm{T}}(\mathbf{d} - \boldsymbol{\Phi}\mathbf{w})\, \mathrm{d}\beta \\ &\quad -\int q(\alpha|\mathbf{d})\, \frac{\alpha}{2}\, \mathbf{w}^{\mathrm{T}}\mathbf{w}\, \mathrm{d}\alpha \\ &= -\frac{\hat{\beta}}{2}\, (\mathbf{d} - \boldsymbol{\Phi}\mathbf{w})^{\mathrm{T}}(\mathbf{d} - \boldsymbol{\Phi}\mathbf{w}) - \frac{\hat{\alpha}}{2}\, \mathbf{w}^{\mathrm{T}}\mathbf{w} \end{aligned} \tag{2.41}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the expectations of the weight and noise precisions under the proposal distributions over α and β (see next two sections for details of the form of these distributions). The weight posterior is therefore a normal distribution: $q(\mathbf{w}|\mathbf{d}) = \mathcal{N}(\mathbf{w}; \hat{\mathbf{w}}, \hat{\boldsymbol{\Sigma}})$, with

$$\begin{aligned} \hat{\mathbf{w}} &= \hat{\boldsymbol{\Sigma}}\, \hat{\beta}\, \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{d} \\ \hat{\boldsymbol{\Sigma}} &= (\hat{\beta}\, \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \hat{\alpha}\, \mathbf{I})^{-1} \end{aligned} \tag{2.42}$$

Thus, the posterior precision matrix, $\hat{\boldsymbol{\Sigma}}^{-1}$, takes the usual Bayesian form of being the sum of the data precision, and the prior precision, $\hat{\alpha}\mathbf{I}$. If $\hat{\alpha} = 0$, i.e. in the absence of prior on the weights, the ML solution of Equation 2.14 is recovered.

Appendix 2.A.6 Updating the Weight Precision, α

Starting from

$$\begin{aligned} f(\alpha) &= \iint q(\beta|\mathbf{d})\, q(\mathbf{w}|\mathbf{d}) \log\left[p(\mathbf{w}|\alpha)\, p(\alpha)\right] \mathrm{d}\mathbf{w}\, \mathrm{d}\beta \\ &= \int q(\mathbf{w}|\mathbf{d}) \log\left[p(\mathbf{w}|\alpha)\, p(\alpha)\right] \mathrm{d}\mathbf{w} \end{aligned} \tag{2.43}$$

As before, the negative free energy is maximised when

$$q(\alpha|\mathbf{d}) \propto \exp[f(\alpha)] \tag{2.44}$$

By substituting for the terms in Equation 2.43 one finds that the updated weight precision posterior density is a Gamma distribution $q(\alpha|\mathbf{d}) = \mathcal{G}(\alpha; \hat{a}, \hat{b})$ where the updated hyper-hyperparameters, $\hat{a}$ and $\hat{b}$, are given by

$$1/\hat{a} = \hat{\mathbf{w}}^{\mathrm{T}}\hat{\mathbf{w}} + \frac{1}{2}\ \mathrm{Tr}(\hat{\boldsymbol{\Sigma}}) + \frac{1}{a_0} \tag{2.45}$$

$$\hat{b} = \frac{K}{2} + b_0, \tag{2.46}$$

where a_0 and b_0 are the initial estimates for a and b, and K is the number of basis functions. The updated value for α, to be substituted into Equations 2.42, is then simply the mean of this distribution:

$$\hat{\alpha} = \hat{a}\hat{b}. \tag{2.47}$$

Appendix 2.A.7 Updating the Noise Precision, β

Again starting from the marginal

$$\begin{aligned} f(\beta) &= \iint q(\alpha|\mathbf{d})\, q(\mathbf{w}|\mathbf{d}) \log[p(\mathbf{d}|\mathbf{w}, \beta)\, p(\beta)]\, \mathrm{d}\mathbf{w}\, \mathrm{d}\alpha \\ &= \int q(\mathbf{w}|\mathbf{d}) \log[p(\mathbf{d}|\mathbf{w}, \beta)\, p(\beta)]\, \mathrm{d}\mathbf{w}. \end{aligned} \tag{2.48}$$

The negative free energy is then maximised when

$$q(\beta|\mathbf{D}) \propto \exp[f(\beta)] \tag{2.49}$$

By substituting for the terms in Equation 2.48 one finds, as with α, that the posterior distribution over β is of Gamma form, $q(\beta|\mathbf{d}) = \mathcal{G}(\beta; \hat{c}, \hat{d})$. The update equations for $\hat{c}$, $\hat{d}$ and $\hat{\beta}$ are:

$$\begin{aligned} 1/\hat{c} &= \frac{1}{2}(\mathbf{d} - \mathbf{\Phi}\hat{\mathbf{w}})^{\mathrm{T}}(\mathbf{d} - \mathbf{\Phi}\hat{\mathbf{w}}) + \frac{1}{2}\mathrm{Tr}\left(\hat{\mathbf{\Sigma}}\,\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right) + \frac{1}{c_0} \\ \hat{d} &= \frac{N}{2} + d_0 \\ \hat{\beta} &= \hat{c}\,\hat{d} \end{aligned} \tag{2.50}$$

where N is the number of data points.

Appendix 2.A.8 Structured priors

Instead of using the isotropic Gaussian of Equation 2.15, where the distribution over all the weights has a common scale (defined by the single hyperparameter α), the weights can be split into groups, and different groups can be allowed to have different scales in their distributions; each weight w_i can indeed have its own scale hyperparameter. This approach is often referred to as Automatic Relevance Determination (ARD) (Neal, 1998; Bishop, 2006), because by inspecting the inferred scales associated with the weights one can see which (groups of) weights are relevant to the problem at hand. The posteriors for the weights of any basis functions which are not helpful in explaining the data will evolve towards zero-mean distributions with vanishingly small variance. Conversely, the weights of basis functions which are well supported by the data will entertain larger variances in their posterior distributions. This means one may operate with a rich basis set and allow the Bayesian model to select only those basis functions that have explanatory power in the data.

Different weights can be allowed to have different scales, but still take into account domain knowledge which may lead us to believe that certain parameters should have a similar posterior scale, by adopting structured priors (Penny & Roberts, 2002), of

the form

$$p(\mathbf{w}|\{\alpha_g\}) = \prod_{g=1}^{G} \left(\frac{\alpha_g}{2\pi}\right)^{K_g/2} \exp(-\alpha_g E_g(\mathbf{w})) \tag{2.51}$$

where the weights have been split into G groups, with K_g weights in the g^{th} group,

$$E_g(\mathbf{w}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{I}_g\mathbf{w} \tag{2.52}$$

and $\mathbf{I}_g$ is a diagonal matrix with ones in the rows corresponding to the g^{th} group, and zeros elsewhere. Use of structured priors results in VB updates for the posterior weight covariance and weight precision as follows

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}} &= (\hat{\beta}\,\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \sum_{g=1}^{G} \hat{\alpha}_g \mathbf{I}_g)^{-1} \\
1/\hat{a}_g &= E_g(\hat{\mathbf{w}}) + \frac{1}{2}\mathrm{Tr}(\mathbf{I}_g\,\hat{\boldsymbol{\Sigma}}\,\mathbf{I}_j) + \frac{1}{a_0} \\
\hat{b}_g &= \frac{K_g}{2} + b_0 \\
\hat{\alpha}_g &= \hat{a}_g\hat{b}_g.
\end{aligned} \tag{2.53}$$

The other updates are exactly the same as for the global variance scale over the parameters.

Appendix 2.A.9 Hyperparameter Values for the ARC

In trend discovery phase phase of the ARC, a single global weight precision hyper-parameter, α, is used, over which a Gamma distribution with hyper-hyperparameters $a_0 = 10^{-2}$, $b_0 = 10^{-4}$ is placed. The noise precision, β is also Gamma distributed with initial hyper-hyperparameters $c_0 = 10^{-2}$, $d_0 = 10^{-4}$. In the trend removal phase vague priors are set for each weight precision, α_g, and for the noise precision β, choosing $a_0 = 10^{-2}$, $b_0 = 10^{-4}$, $c_0 = 10^{-2}$, $d_0 = 10^{-4}$.

3

STATISTICS OF VARIABILITY

The material in this chapter was published in McQuillan, A., Aigrain, S. & Roberts, S., 2012, A&A, 539, A137.

In this chapter I examine the variability statistics of the *Kepler* Q1 data, after systematics removal using the ARC method described in Chapter 2. In Section 3.1, I detail the motivation behind studies of variability statistics in main sequence stars, and discuss the relevant literature on this topic. I describe the statistics I used to characterise variability in Section 3.2, and examine the solar comparison in Section 3.3. Using the solar level, I divide the main sequence stars into samples of high and low variability and compare their stellar properties and distribution, which is detailed in Section 3.4. A preliminary study of the periodic and stochastic components of variability is presented in Section 3.5.

3.1 INTRODUCTION

3.1.1 Motivation for Stellar Variability Studies

Measuring the basic characteristics of the variability (amplitude, periodicity, etc...) across large samples of stars, and comparing them to stellar parameters such as

age, mass and composition, is a first step towards a better understanding of the underlying phenomena. Many of the latter are ill-understood, because they are related to rotation, convection and magnetism which are challenging to model. By studying the observable properties of activity, one can begin to probe the mechanisms which lead to such effects.

It is also of great importance to understand where the Sun lies in the distribution of stellar variability types and amplitudes. As the most studied star, it provides the foundations for our understanding stellar physics and as a key calibrator for stellar models. However, there is now evidence that the Sun may not be completely 'typical' for it's spectral classification and age, and is in fact the less active on average (see Sections 3.1.2 and 3.3 for further discussion).

Variability statistics also have a crucial impact on exoplanet studies, particularly for radial-velocity searches or radial-velocity confirmation of transiting planet candidates (see e.g. Pont et al., 2011a). Transit searches are also affected by higher levels of stellar variability since it adds correlated noise to the light curves, creating a challenge for transit detection algorithms (Aigrain et al., 2004). One reason cited as a requirement for the extension of the *Kepler* mission was the higher than expected stellar 'noise' (Gilliland et al., 2011), meaning that more transits need to be observed to reach the required signal to noise. The stellar noise term was based on predictions that $\sim 2/3$ of solar-type stars would be as quiet as the Sun (Basri et al., 2005), highlighting again the importance of understanding differences between 'solar-type' stars and the Sun itself.

There are a variety of statistics that can be used to quantify variability, but this chapter focusses primarily on the amplitude of flux variation.

3.1.2 *Previous Work on Variability Statistics*

Since the fundamental dependence of activity on rotation period and age were established (Pizzolato et al., 2003), studies have searched for correlations between activity levels and other stellar parameters. Eyer & Grenon (1997) show a temperature dependence, where cooler stars display higher amplitude variability than hotter stars,

which translates naturally to a correlation with spectral type.

The recent space-based transit searches CoRoT and *Kepler* have revolutionised the study of stellar variability, with variation precision down to mmag level now obtainable. (Hulot et al., 2011) study the low frequency power excess in the CoRoT light curve, which they attribute to stellar activity. After removal of this component, they assign an index to the remaining power and use it to characterise the microvariability and show that it correlates with temperature, displaying the highest values for the coolest stars.

The Quarter 1 (Q1) data, which was made public in June 2010, has already been studied by Basri et al. (2010, hereafter B10), who show that somewhat less than half of the dwarf stars surveyed by *Kepler* are more variable than the Sun on timescales of up to a month, with the fraction increasing from earlier to later spectral types. Basri et al. (2011, hereafter B11) went on to demonstrate that periodic variable stars have significantly larger amplitudes, as a sample, than aperiodic variables. Finally, Ciardi et al. (2011, hereafter C11) performed a complementary study of the same sample using dispersion rather than amplitude as a variability statistic, and studying likely dwarfs and giants separately using the stellar parameters provided in the *Kepler* Input Catalog (KIC, Brown et al. 2011; Batalha et al. 2010). Variability statistics have also been determined using the 10 days of commissioning data (Q0), with the aim of developing methods to characterise and select specific types of variable (Walkowicz & Basri, 2010).

The C11 study use the PDC 'dispersion' which can be downloaded from (NStED) together with other pre-computed statistics, including the light curve median and reduced χ^2. Dispersion is defined as the 1 sigma rms scatter around the median magnitude of the light curve.

Instead of dispersion, B10,11 measured the light curve 'range', $R_{\rm var}$, which is essentially a measure of the peak-to-peak variation. The effect of high-frequency noise was removed either by smoothing the light curve on 10-hour timescales (B10) or by discarding the upper and lower 5 percentiles (B11). The choice of statistic used to study the variability is somewhat arbitrary and does not significantly alter the results.

The biggest variation between these studies is in the choice of systematics cor-

rection. B10, 11 perform their own trend removal and C11 use the *Kepler* team's Pre-Search Data Conditioning (PDC) method (Jenkins et al., 2010b). A detailed description and comparison of systematics treatment in *Kepler* data can be found in Chapter 2. This inspired me to investigate further the apparent bimodality in the variability of dwarf stars observed by *Kepler*, with particular attention to the effect of different systematics correction methods.

The work described in this chapter revisits the studies of C11 and B10,11 following the application of a new astrophysically robust de-trending method (ARC), designed to preserve intrinsic variability signals and remove as fully as possible the systematics (See Chapter 2). We quantify trends, previously identified and new, between variability characteristics and stellar properties, and investigate the nature of the variability in more detail, in order to gain further insight into the underlying mechanisms.

In Section 3.2 we discuss our choice of variability statistics, and in Section 3.3 compare the stars observed by *Kepler* to the Sun. We use this Solar comparison level to divide the stars into low and high variability samples and examine their physical properties in Section 3.4, focussing on their periodic and stochastic nature in Section 3.5.

3.2 AMPLITUDE OF VARIABILITY

Based on the method of B11, I selected $R_{\rm var}$, the range between the 5th and 95th percentile, for the median normalised light curve as the variability statistic. Dispersion and reduced χ^2 provide an appropriate measure of variability that is believed to be primarily stochastic and Gaussian. Pulsations and rotational variability do not meet these criteria and therefore a measurement based on the peak-to-peak variations in the light curve is considered more relevant. Selecting the 5th to 95th percentile range reduces the noise on the peak-to-peak measurement.

$R_{\rm var}$ measurements for the Raw, PDC and ARC data are compared in Figures 3.1 and 3.2. The ARC clearly removes most systematic effects, reducing the lower envelope of points to the photon noise limit, however it does not have the side effect of

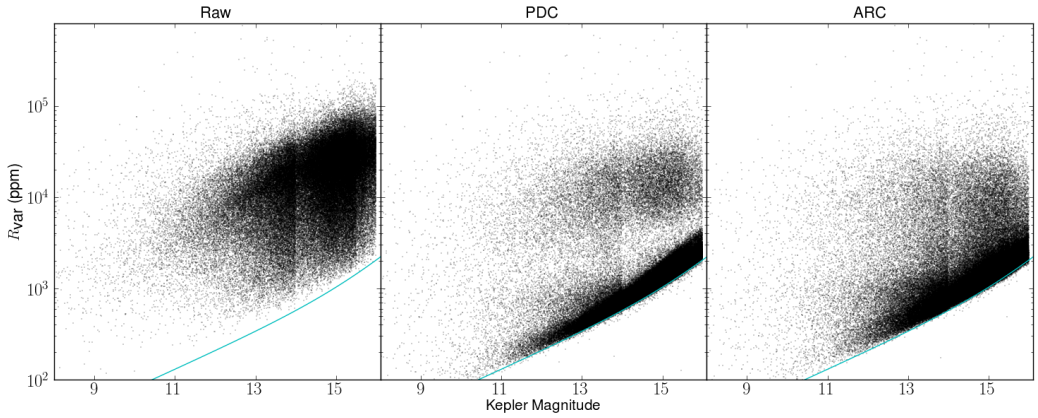

Figure 3.1: $R_{\rm var}$ (described in Section 3.2) against *Kepler* magnitude for the Raw, PDC and ARC *Kepler* Q1 data. The solid line, in the same position on each graph, shows the photometric uncertainty (see Section 3.3). The removal of true stellar variability at medium levels by the PDC can be seen in the dearth of stars around $R_{\rm var} = 10^3 - 10^4$ ppm in the middle panel.

suppressing intermediate amplitude variability (as done by the PDC; this is apparent in the scarcity of points around $R_{\rm var} = 10^3 - 10^4$ ppm in Figure 3.2 and the middle panel of Figure 3.1.) Another unfortunate side effect of the PDC is the introduction of high frequency noise in some light curves, which does not occur with the ARC.

3.3 SOLAR COMPARISON AND VARIABILITY FRACTION

The empirical three-section cut of C11 was used to distinguish between likely dwarfs and giants based on surface gravity $\log g$ and effective temperature $T_{\rm eff}$ in K, where stars defined as dwarfs have:

$$\log g \geq \begin{cases} 3.5 & \text{if } T_{\rm eff} \geq 6000 \\ 4.0 & \text{if } T_{\rm eff} \leq 4250 \\ 5.2 - (2.8 \times 10^{-4} T_{\rm eff}) & \text{if } 4250 < T_{\rm eff} < 6000 \end{cases} \tag{3.1}$$

B10 use a simpler $\log g \geq 4$ cut. It has since been noted that the KIC contains some misidentifications (Koch et al., 2010), but since these are not expected to be numerous enough to affect these results, and for ease of comparison to C11, the KIC values were used without modification. There is no single published value for predicted

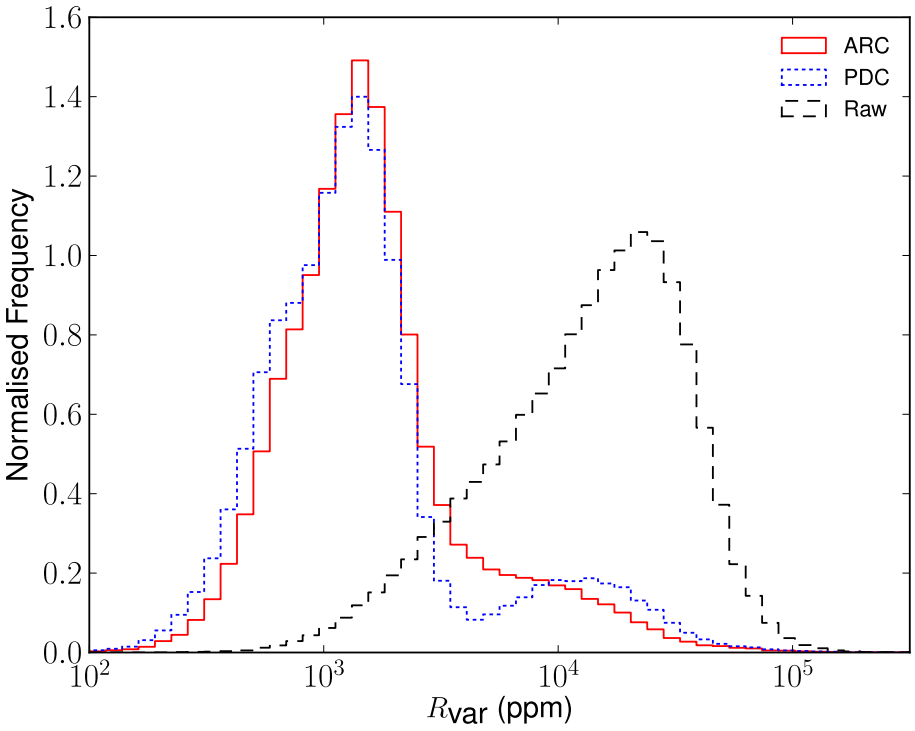

Figure 3.2: Histogram of variability for stars 13 < Kepmag < 14 for the Raw, PDC and ARC *Kepler* Q1 data.

giant contamination of this sample, but C11 suggest giant contamination in the M star sample of 4%.

To compare the properties of the high and low variability stars a cut was made in $R_{\rm var}$ based on a comparison to twice the variability level of the active Sun. This level is somewhat arbitrary but provides an appropriate way to divide stars with approximately solar levels of variability and quieter, from those which are significantly more variable than the active Sun. The solar $R_{\rm var}$ value was calculated from the SOHO/VIRGO summed g+r light curves for the active Sun in the year 2000, because these provide the closest match to the *Kepler* bandpass (B10). The solar $R_{\rm var}$ value was calculated to be 766 ppm from the average obtained using a sliding 33 day section of data over 2 years, centred on the activity maximum. An empirical fit to the median of the photometric uncertainties on the light curves in 0.5 mag bins provides an estimate of noise levels across the range of magnitudes. The equation of the solar equivalent line is composed of the solar $R_{\rm var}$ measurement, photon and background noise terms,

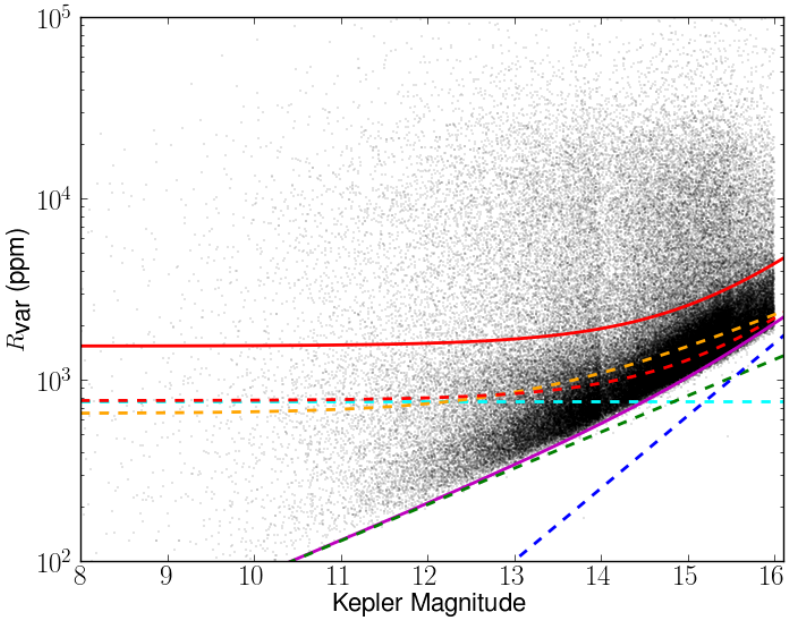

Figure 3.3: Division of high and low variability dwarf stars (red solid line) at twice the solar value (red dashed line). The solar level is calculated from the active Sun level (cyan dashed line) and the noise (magenta solid line), which itself is a combination of background (blue dashed) and photon noise (green dashed). The orange dashed line marks the position of the solar line as determined by B10.

$$y^2 = \text{solar } R_{\rm var}^2 + 10^{0.4(\text{mag}-p_1)^2} + 10^{1.6(\text{mag}-p_2)^2}, \tag{3.2}$$

where best fit values for p_1 and p_2 are 0.4 and 8.0 respectively. See Figure 3.3 for a graphical representation of the division process. The orange dashed line in Figure 3.3 represents the solar line as determined by B10, where extensive visual comparison of solar and *Kepler* light curves is used to estimate a solar level.

The fractions of dwarf stars with variability greater than the solar level are shown in Table 3.1. We find that 60% of dwarf stars are more variable than the active Sun on 33 day timescales. B10 find 46% of dwarf stars to be more variable than the solar level. C11 define stars as significantly variable if the reduced χ^2 is greater than 10, and find the fraction of dwarf stars meeting this criteria to be 18%.

The variability fraction is strongly dependent on the choice and definition of the solar and photometric noise levels. A larger fraction of stars fall into the low variability

Table 3.1: Fractions of stars more variable than once and twice the solar level, and fractions of periodic stars in the high and low variability groups (see Section 3.5.1 for definition). The random uncertainty for each of the measurements is < 0.5%, and is negligible in comparison to difference introduced by the choice of dividing line, demonstrated by the 1 and 2 times the solar value comparisons listed in this table (see Section 3.3 for further discussion).

Sample	$R_{\rm var}$ > Solar	$R_{\rm var}$ > 2×Solar	Periodic Fraction
All Dwarfs	0.60	0.20	0.16
A	0.53	0.42	0.23
F	0.46	0.15	0.17
G	0.56	0.14	0.13
K	0.79	0.33	0.20
M	0.96	0.56	0.25

subset when the division of B10 is used, due to the small difference in the solar value used and more importantly the position of the estimated photometric noise level. My empirical fit to the median of the photometric uncertainties produces a line that more closely follows the shape of the lower envelope in Figure 3.3 than the visual estimate of B10.

The random uncertainty was calculated for each of the variability and periodicity fractions listed in Table 3.1 by performing 10,000 measurements using random samples of 80% of the data, and found these to be < 0.5% and hence negligible in comparison to difference introduced by the choice of dividing line.

It is important to note that for many stars in the low variability sample, the lower limit of $R_{\rm var}$ is due to the photometric precision of the *Kepler* satellite, shown as the magenta line in Figure 3.3. However, due to the position of the dividing line (solid red line in Figure 3.3) and its dependence on the photometric precision, there are still a large number of stars that are significantly more variable than the photometric noise limit for their magnitude. There are ∼1,500 stars in the low variability sample that are above twice the photometric noise level. Using only stars with *Kepler* magnitude < 14 the fraction of dwarf stars with greater than solar level of variability and twice solar level are 0.41 and 0.22 respectively. This shows a decrease in the fraction greater than solar level, however, the M dwarfs have the highest levels of variability, and since these are faintest, few are included in the bright sample.

3.4 STELLAR PROPERTIES OF THE LOW AND HIGH VARIABILITY SAMPLES

B10,11 and C11 highlighted various relationships between variability and stellar properties. These are revisited here using the ARC data. As C11 noted, there is a decrease in variability with temperature which cannot be entirely explained by the increased noise levels in the fainter cool stars (see Figure 3.4). Pearson's correlation coefficient of effective temperature and $\log(R_{\rm var})$ for the whole dwarf sample is -0.31, which is weak but non-negligible.

To examine whether the high and low variability samples belong to different stellar populations, I investigated the possible differences in their spatial distribution and kinematics. Due to dynamical heating of the galactic disk over long timescales, older populations of stars tend to have larger galactic scale heights and larger velocity dispersions than younger populations (see e.g. Freeman & Bland-Hawthorn, 2002, and references therein). This can be tested by examining the distribution of the two samples in galactic coordinates, shown in Figure 3.5.

The division in $R_{\rm var}$ introduces a magnitude bias between the high and low samples, with a greater proportion of the low variability objects at higher magnitudes. To reduce any possible effect of this bias on our comparison tests, I selected high and low variability subsets with approximately equal magnitude distributions. This was done by dividing the stars into 1 mag bins (with the exception of stars brighter than 8th and fainter than 16th in *Kepler* magnitude, which were treated separately), and choosing a random set from the more populated bin to match the number in the smaller one. This process, used for the spatial distribution and proper motion comparison tests, also serves to select an equal number of stars in the high and low variability sets.

When using the PDC data, a difference in spatial distribution of the two subsets is evident, as noted by C11. The low variability stars appear approximately uniformly distributed, whereas the high variability stars are concentrated towards low galactic latitudes. Using the PDC data alone, one could hypothesise that the variable sample may correspond to a younger, thin disk population, and the quiet sample to an older

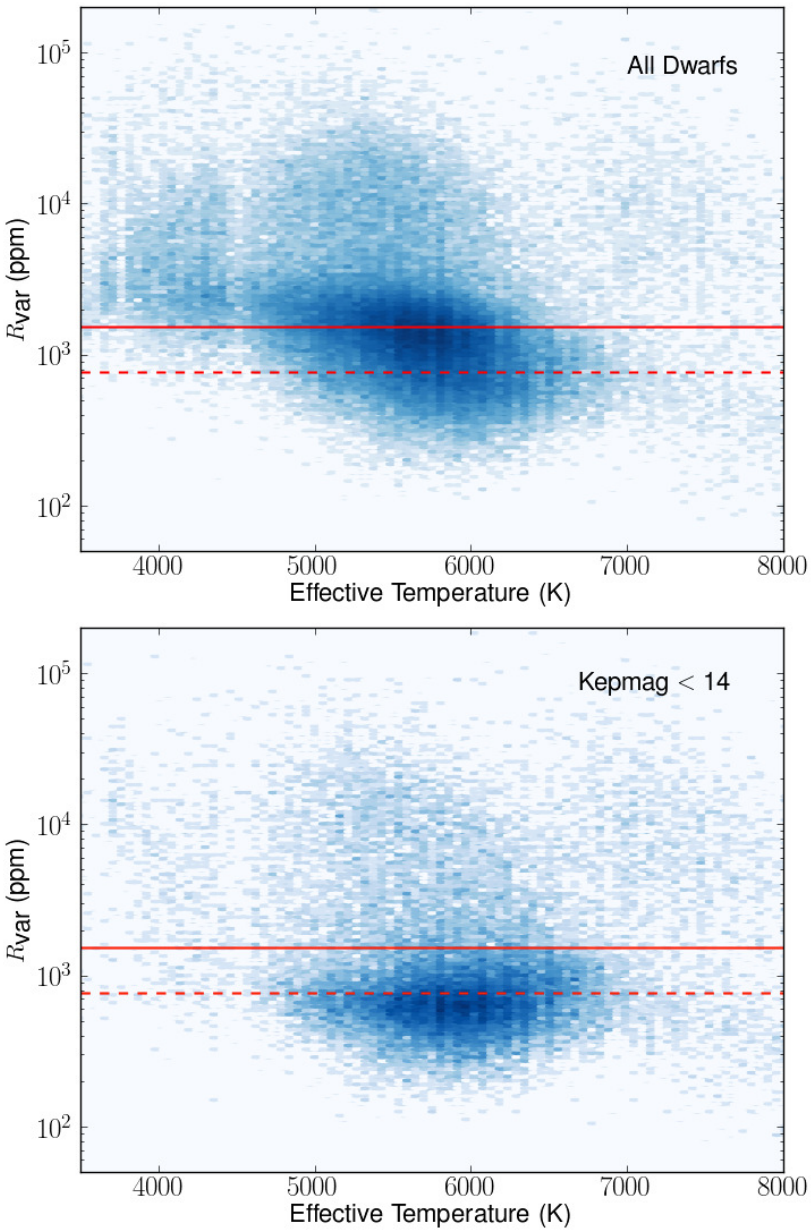

Figure 3.4: Density plot of effective temperature against $R_{\rm var}$, showing the decrease in variability with increasing temperature for all the selected dwarf stars (*top*) and for those with *Kepler* magnitude < 14 (*bottom*). This illustrates that the dearth of low variability stars at low temperatures is not a result of the increased noise floor of the cool, faint stars. The dashed line shows the solar variability level and the solid line is twice the solar level.

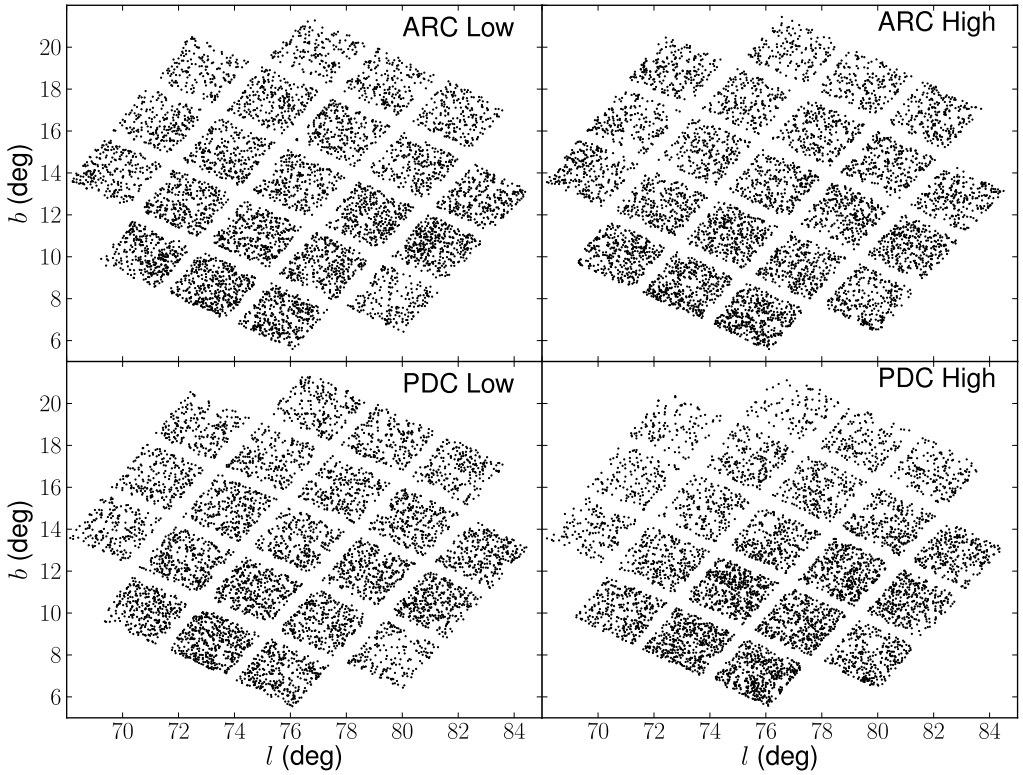

Figure 3.5: The spatial distribution of low (*left*) and high (*right*) variability stars for ARC (*top*) and PDC (*bottom*). Each has been subject to a random selection of equal numbers of stars per magnitude bin, before a random selection of 6000 for each panel was selected. The PDC panels show a slight tendency for the high variability sample to be more concentrated towards the galactic plane, whereas this effect is not significant in the ARC, suggesting it is more effective at removing contamination related systematics.

population with larger scale height. C11 hinted at this conclusion but argued that this could also be an artefact of increased crowding at low galactic latitudes, causing higher levels of photometric dispersion.

By visually comparing the spatial distribution of the high and low variability samples from the PDC and ARC, it becomes evident that the effect is dependent on the choice of reduction method (Figure 3.5). A two-sample Kolmogorov-Smirnov (KS) test shows a p-value of 1×10^{-15} for the null hypothesis that the high variability galactic latitudes for the ARC and PDC corrections come from the same distributions. This suggests that contamination effects were the dominant cause of the apparent higher variability close to the galactic plane, and that the ARC is better at removing the associated systematics. There is still some sign that active F stars are more

Table 3.2: Statistics of galactic latitude (b) distributions, including two-sample Kolmogorov-Smirnov tests, median, median absolute deviation (MAD) and skew of the low (L) and high (H) variability samples.

Sample	b KS	Med Lb	Med Hb	MAD Lb	MAD Hb	Skew Lb	Skew Hb
A	0.016	11.68	12.00	1.76	1.57	0.70	0.54
F	1.4×10^{-9}	12.52	11.79	1.92	1.70	0.35	0.59
G	0.002	12.88	12.65	2.08	1.95	0.21	0.28
K	0.1	13.29	13.32	2.21	2.08	0.07	0.04
M	1.0	13.65	13.58	2.21	2.08	-0.004	-0.016

concentrated at lower latitudes (Figure 3.6), but it is too subtle to draw any strong conclusions.

The giant contamination in the F stars is higher than other spectral classes because they span the temperature range where the giant branch intersects the main sequence on the HR diagram. Giant stars are typically observed to be more variable than dwarf stars (Gilliland 2008, B10, C11), which may lead to the apparent concentration of variable F stars at low galactic latitudes, if the variable and luminous giants are visible out to greater distances in the galactic plane, and hence appear with higher density in these regions. Spectroscopic followup would be required to determine the exact level of giant contamination, but it is believed to be a few percent at most (Koch et al. 2010, B10).

As a more quantitative measure, two-sample KS tests were performed. The results are shown in Table 3.2, and confirm that the galactic latitudes of the low and high variability F stars are very unlikely to be drawn from the same distribution, compared to a high probability for the M stars. The distribution shape parameters, also shown in Table 3.2, display the trends of increasing median galactic latitude with later spectral type and the subtle differences between high and low variability subsets, although these are not sufficient to prove the stars belong to different populations. The spatial distribution plots also serve to show that the correction applied using ARC affects all areas of the CCD equally, despite being implemented on each CCD channel separately.

To examine the effect of crowding, I used the contamination fraction available from the *Kepler* mission archive, which provides an approximate measure of the fraction of

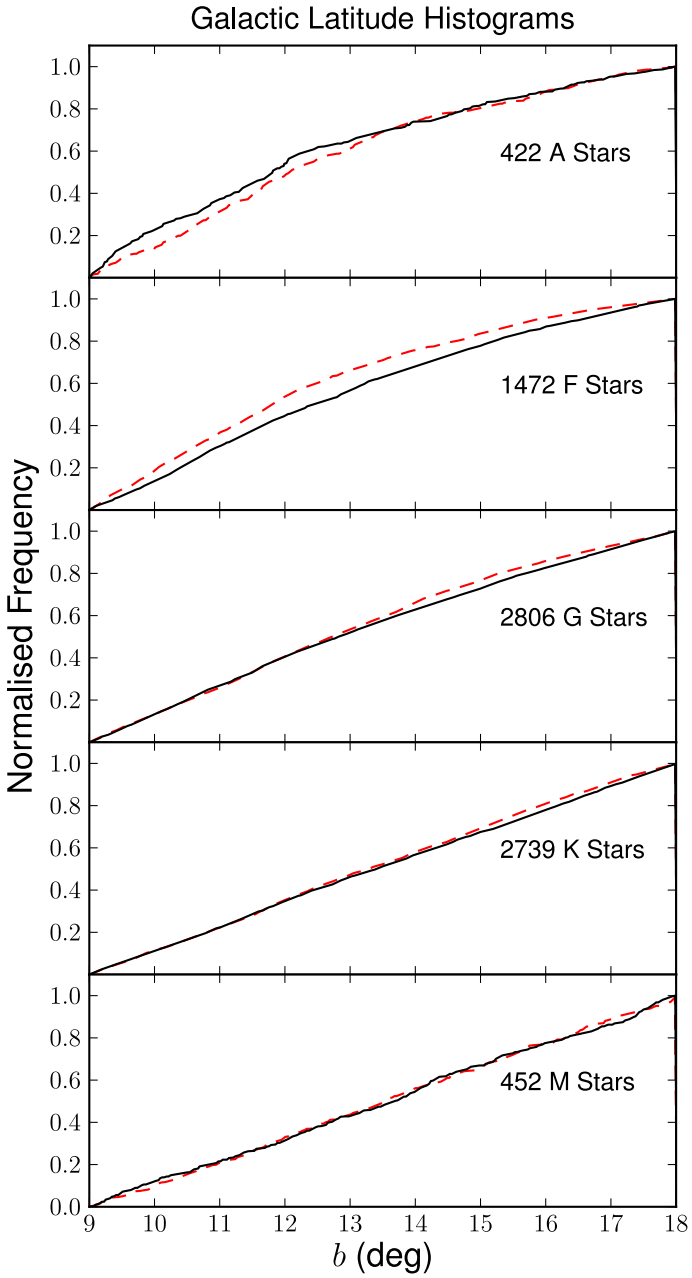

Figure 3.6: Histogram showing the galactic latitude distribution of low (solid line) and high (dashed line) variability stars for each spectral type. Sample selection ensured the orientation of the *Kepler* field on the plane did not introduce biases. An equal sample of high and low variability stars from each magnitude bin was also selected. The increased number of high variability F stars at low galactic latitudes may arise from giant contamination of the sample (see Section 3.4).

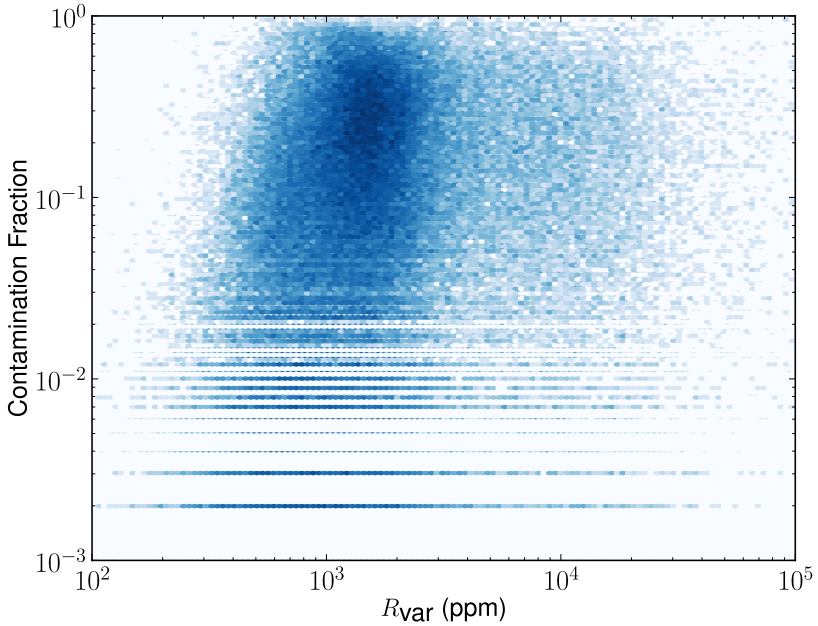

Figure 3.7: A weak correlation can be seen between contamination fraction and R_{var}, for non-zero contamination values.

flux attributed to the target in a 21 $\times$ 21 pixel aperture. Values range from 0, implying no contamination, to 1, which indicates the flux is essentially all background. There is a weak correlation between contamination fraction and R_{var}, shown in Figure 3.7. We tested whether placing a constraint on the contamination fraction, using only targets in the range 0.1 to 0.2, altered the appearance of Figure 3.5, but it did not.

A potential indication that the low and high variability subsets belong to different stellar populations can be seen in the proper motion distributions (Figure 3.8), for stars with total proper motion greater than zero. It shows a weak but noteworthy trend for the high variability group to have lower proper motion than the low variability group. I also tested the significance of these differences using two-sample KS tests, and parameterised the shape of the distributions, the results of which can be seen in Table 3.3. The KS test results show that with the exception of A stars, and to some extent F, the proper motion values for the high and low variability groups are very unlikely to be drawn from the same distribution, while the shape parameters highlight more subtle differences in shape and trends between spectral types. This is consistent with the view that higher variability stars are younger (West et al., 2008), and therefore

Table 3.3: Statistics of total proper motion distributions, including two-sample Kolmogorov-Smirnov tests, median, median absolute deviation (MAD) and skew of the low (L) and high (H) variability samples.

Sample	*pm* KS	Med *pm* L / H	MAD *pm* L / H	Skew *pm* L /H
A	0.61	0.0057 / 0.0063	0.0032 / 0.0030	4.61 / 1.57
F	0.09	0.0077 / 0.0076	0.0038 / 0.0033	9.57 / 12.66
G	0.0009	0.010 / 0.0091	0.0045 / 0.0037	12.06 / 4.45
K	0.002	0.013 / 0.011	0.0061 / 0.0045	8.75 / 9.41
M	1.5×10^{-6}	0.033 / 0.020	0.013 / 0.0091	2.41 / 4.48

have lower proper motions (see e.g. Freeman & Bland-Hawthorn, 2002, and references therein). This conclusion is dependent on the stars being at the same distance and location on the sky. In these samples, the magnitude and spatial distributions are approximately equal for the high and low variability subsets, which should satisfy this condition.

One caveat applicable to the proper motion distributions is that the giant contamination in the M dwarfs (described by C10) could increase the apparent number of high variability stars with low proper motion. The potential effects of Malmquist bias were also considered in these results, but by selecting equal magnitude distributions and comparing each spectral type separately, any effect introduced should be negligible. Given the large uncertainties on the KIC proper motions, we are not able to draw strong conclusions as to whether the proper motion difference indicates an age difference in the samples.

3.5 PERIOD AND STOCHASTIC VARIABILITY IN Q1

I investigated the nature of the variability, comparing the low and high variability samples, using a combination of visual examination, the statistics described in Section 3.2, and periodogram analysis. The typical characteristics of the light curves also vary between spectral types as shown in the examples in Figure 3.9. Visual examination of the light curves shows that A and F stars contain many pulsators with a large range of amplitudes, while later types show rotational modulation and more stochastic variability. This trend is also apparent in the periodicity and stochasticity

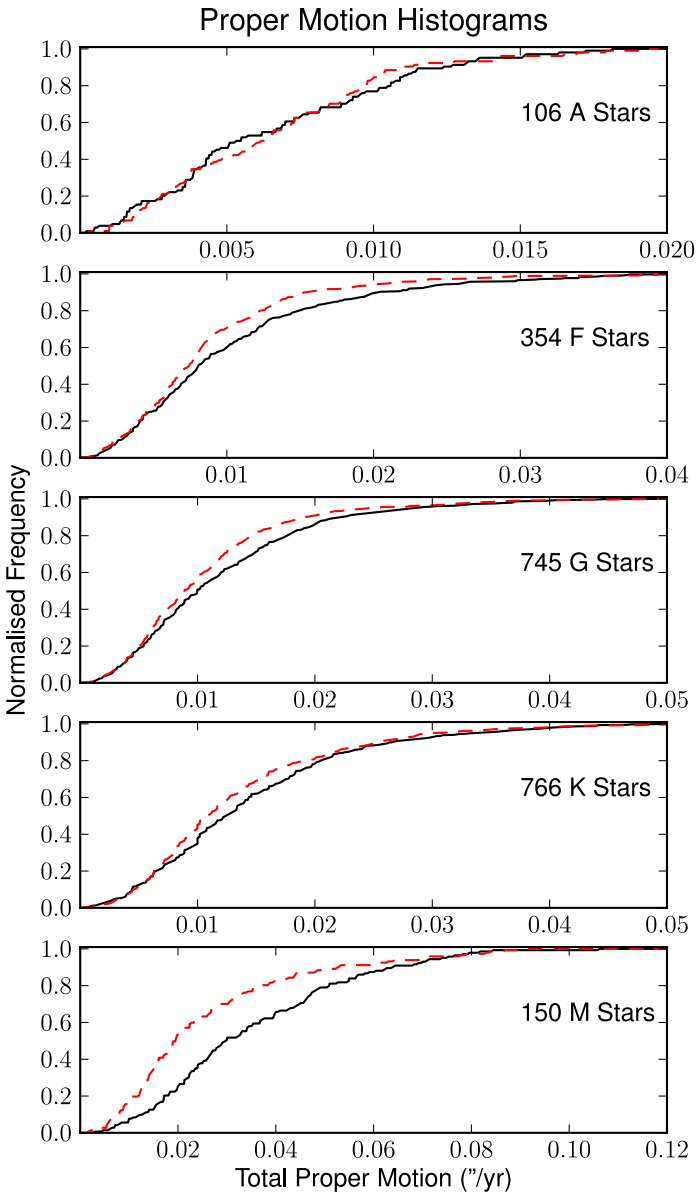

Figure 3.8: Distribution of proper motion values for the low (solid line) and high (dashed line) variability samples, where proper motion > 0, in a box of even galactic latitude and longitude distribution, with equal numbers of high and low variability stars selected in each magnitude bin. The uncertainty on the proper motion values is 0.02"/year but given the high numbers in each sample these results are still significant.

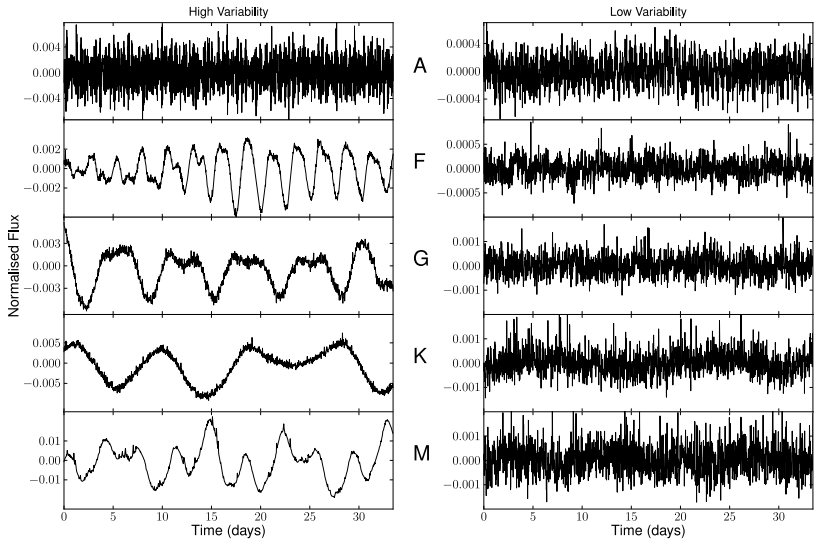

Figure 3.9: Typical light curves from the high and low variability groups (*left and right respectively*) for each spectral type.

tests described in this section.

3.5.1 *Periodicity*

For each light curve, I computed a periodogram by least-squares fitting of a sinusoid plus a constant at each trial period. The periodogram is expressed in terms of the statistic

$$S = \left(\chi_0^2 - \chi^2\right) / \chi_0^2, \tag{3.3}$$

where χ_0^2 is the reduced chi-squared of the light curve with respect to a constant value, and χ^2 is the reduced chi-squared with respect to the best-fit sinusoid. I used 500 logarithmically-spaced periods between 0.01 and 100 days. However, only periods between 1 hour (twice the sampling rate) and 16 days (half the time span of the data) were considered valid. This approach was chosen over limiting the initial search space to this range, because that would have led to many spurious detections close to the upper limit. Note that some objects with true periods above 16 days may be detected at harmonics of their true period.

The dearth of stars with periods detected close to the 16 day limit is a bias introduced by the stringent period identification method. Many objects that would

occupy this period range have broad periodogram peaks at around the 16 day limit and the selection method has been designed to only accept periodogram peaks that drop to the median level before reaching the edges of the allowed range. This prevents false detections of periods at the length of the permitted dataset, and misidentification of the highest periodogram continuum point as a genuine peak.

Bretthorst (1998) describes the Bayesian theory of modelling data using a sinusoid plus white noise. In this formalism, my metric can be shown to be

$$S = 2C\left(\omega\right) / \left(N\sigma_{\mathrm{rms}}^2\right), \tag{3.4}$$

where $C\left(\omega\right) = \left(R(\omega)^2 + I(\omega)^2\right)/N$, is the squared magnitude of the discrete Fourier transform of the data, N is the number of data points in the light curve, $R(\omega) = \sum_{i=1}^{N} \cos(\omega t_i)$, $I(\omega) = \sum_{i=1}^{N} \sin(\omega t_i)$ and σ_{rms} is the rms scatter of the data. The data is assumed to be mean-subtracted.

As discussed in B11, the choice of a threshold for periodicity detection is best made empirically. I selected $S = 0.3$ as an appropriate threshold between clear-cut periodic variability and non-periodic or ambiguous cases. This threshold is slightly lower than the value of $S = 0.4$ used in the Monitor project (see e.g. Irwin et al., 2006), which is reasonable given the vastly superior time-sampling of the *Kepler* data.

This threshold can be compared to a power spectrum cut, since the power spectrum is given by $\mathrm{PS}(\omega) = 4C(\omega)/N$ where, following Kjeldsen & Bedding (1995), I normalised the power spectra such that a sinusoidal oscillation of amplitude A gives rise to a peak of height A^2 in the power spectrum. The threshold selected is then equivalent to a power spectrum threshold of $S_{\mathrm{PS}} = 0.6\sigma_{\mathrm{rms}}^2$. By comparison Kjeldsen & Bedding (1995) give an expression for the noise level in the power spectrum of $\sigma_{\mathrm{PS}} = 4\sigma_{\mathrm{rms}}/N$.

Using the Q1 data, $N \gg 1$, so my threshold is conservative when comparing to white noise. However, white noise is not the dominant factor defining the ability to detect periodicities. To test the appropriateness of my periodicity threshold on data with realistic noise properties, I ran a set of simulations where we injected periodic signals into actual *Kepler* light curves. 1000 Q1 light curves were then randomly selected, with low variability ($R_{\mathrm{var}} < R_{\mathrm{var,Sun}}$) and no significant period

($S < 0.25$), into which sinusoidal signals were injected, with random periods uniformly distributed between 2 and 16 days, and random amplitudes ranging from 0.1 to 10 times the high-frequency noise level (measured as the scatter in the difference between consecutive flux measurements), with a distribution of amplitudes relative to the noise level that was uniform in log. I then stored the best-fit value of S and the corresponding period, and found that S was always < 0.3 when the best-fit period differed by $> 10\%$ from the injected value, suggesting a false detection rate $< 10^{-3}$. As one would expect, the missed detection rate is period- and amplitude-dependent: there were only 5 cases in these simulations where the recovered period was within 10% of the injected one but S was < 0.3, and these were all for $P_{\rm injected} > 12$ days and amplitudes significantly smaller than the noise. However, these simulations did not include a large enough number of low-amplitude, difficult to detect signals to make a more detailed estimate of the completeness. Since the period search in Q1 data was intended as a preliminary exercise, and the results will become much more reliable when additional data is included, no further simulations were performed at this stage.

Using the criteria outlined here, 16% of the dwarf stars are determined to be periodic, and the fraction for each spectral type is given in Table 3.1. These numbers vary slightly from those quoted in B11 and C11. This indicates that my threshold for periodicity is more stringent, designed to detect only genuine periodicities (or strong harmonics) within the limits of the frequency resolution. The periodic fraction will undoubtably increase with longer datasets and more efficient quasi-periodic signal detection methods.

Histograms of the detected periods for each spectral class are shown Figure 3.10. When interpreting these, one should bear in mind that our period sensitivity is non-uniform, and that the largest peak in the periodogram is not necessarily at the true period but can be at one of its harmonics. The distribution is also truncated close to the maximum period due to the definition of 'peak' required for selection. These period distributions should thus be taken as indicative only (the period sensitivity will improve vastly with longer time coverage). Nonetheless, differences between the histograms are obvious, with the typical period clearly increasing towards later

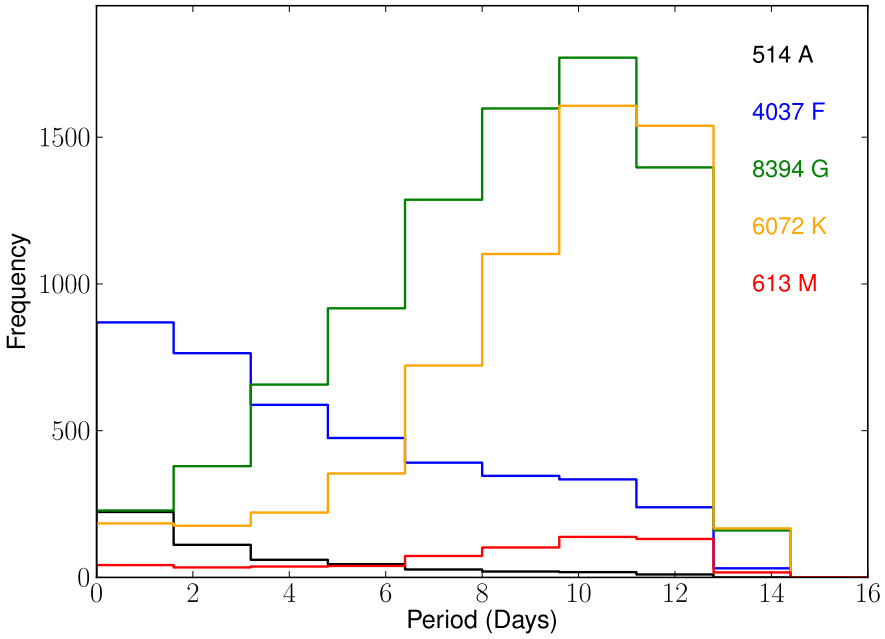

Figure 3.10: Period distribution for each spectral type, for stars where a significant period between 1hr and 16d has been detected (see Section 3.5.1 for method and caveats), showing an increase in period towards later spectral types.

spectral types.

The majority of A stars, and about half of F stars, have very short periods (< 2 days). These are likely to be pulsators (many of the A and F stars are located in the instability strip), although they could also be unpublished close binaries (where ellipsoidal variations and mutual heating induce sinusoidal variations) or active stars with very short rotation periods. On the other hand, G, K and M stars, along with the rest of the F stars, have significantly longer periods (≥ 5 days), as one might expect from rotational modulation of active regions. This distinction is confirmed by the appearance of the light curves (Figure 3.9). The trend of longer periods towards later spectral types is also visible amongst the late type objects.

3.5.2 *Degree of Stochasticity*

A significant difference in the nature of variability across different spectral types is shown by the degree of stochasticity. I performed a simple measure of this based on the number of peaks, $N_{\rm pk}$, in the sine-fitting periodogram which have power greater than 10% the maximum power. Only stars above the periodicity threshold described in Section 3.5.1 were selected. An example $N_{\rm pk}$ determination is shown in Figure 3.11.

The histograms showing the $N_{\rm pk}$ distribution for each spectral type are displayed in Figure 3.12. The light curves of pulsating stars are dominated by near-sinusoidal variability at one, or a few, clear dominant periodicities, corresponding to low values of $N_{\rm pk}$. Rotational variables are less well modelled by sine-fitting. They typically have a more complex periodogram, with significant power at multiple harmonics of the dominant period (see e.g. Boisse et al., 2009) and some stochastic variability (with power at all frequencies, as seen in the case of the Sun, see e.g. Aigrain et al. 2004). This results in somewhat larger values of $N_{\rm pk}$. The high variability group show a tendency towards slightly lower levels of stochasticity than the low variability group, which may arise from the strongly periodic pulsating stars that typically have high amplitude variations. It is important to note that the $N_{\rm pk}$ statistic is not a quantitive measure of stochasticity, but intended for use as a comparison between the low and high dispersion samples.

Alternative metrics for measuring timescale, periodicity and stochasticity were demonstrated by Walkowicz & Basri (2010), who use time separation between points where the differential light curve crosses zero, and the time separation between changes in the sign of the slope in the light curve. This method, without smoothing of the light curve, is slightly more sensitive to noise than the $N_{\rm pk}$ approach.

3.5.3 *Harvey Model Fitting*

I also studied the typical stochastic properties of each spectral class by fitting a model to the average power spectra of each. The Fast Fourier Transform (FFT) of each light

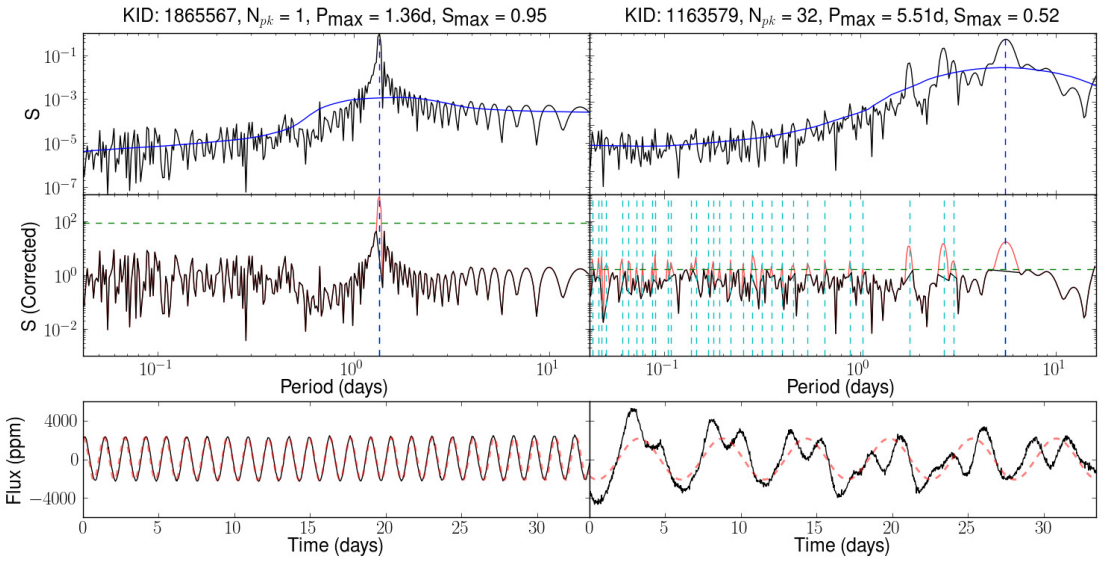

Figure 3.11: Graphical representation of the calculation of $N_{\rm pk}$ for stable-periodic (*left*) stars and those with more stochastic periodicity (*right*). The *top panels* show the original power spectrum, smoothed spectrum, and the maximum peak (dashed line). The original power spectrum is divided by the smoothed one to get the corrected spectrum in the *centre panels*. Peaks above 10% of the maximum value (horizontal dashed line) are then counted. The bottom panels shows the light curve (solid line) with best fit sinusoid (dashed).

curve was computed and the median, 10th and 90th percentile at each frequency were used to create the plots in Figure 3.13. These spectra were smoothed using a nonlinear median boxcar filter to remove the effects introduced by the variable pointing stars (evidence of which can be seen in the peaks of the cyan line in Figure 3.13).

Autoregressive (AR) models are commonly used to describe stochastic processes and take the form,

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-1} + \epsilon_t, \tag{3.5}$$

where x_t and x_{t-1} are data values at time index t and t-1 respectively, c is a constant, p is the order of the model, ϕ are the parameters and ϵ is white noise. Such a process has a spectral density

$$P(\nu) = \sum_{i=1}^{N} P_i = \sum_{i=1}^{N} \frac{A_i}{1 + (B_i \nu)^{C_i}}, \tag{3.6}$$

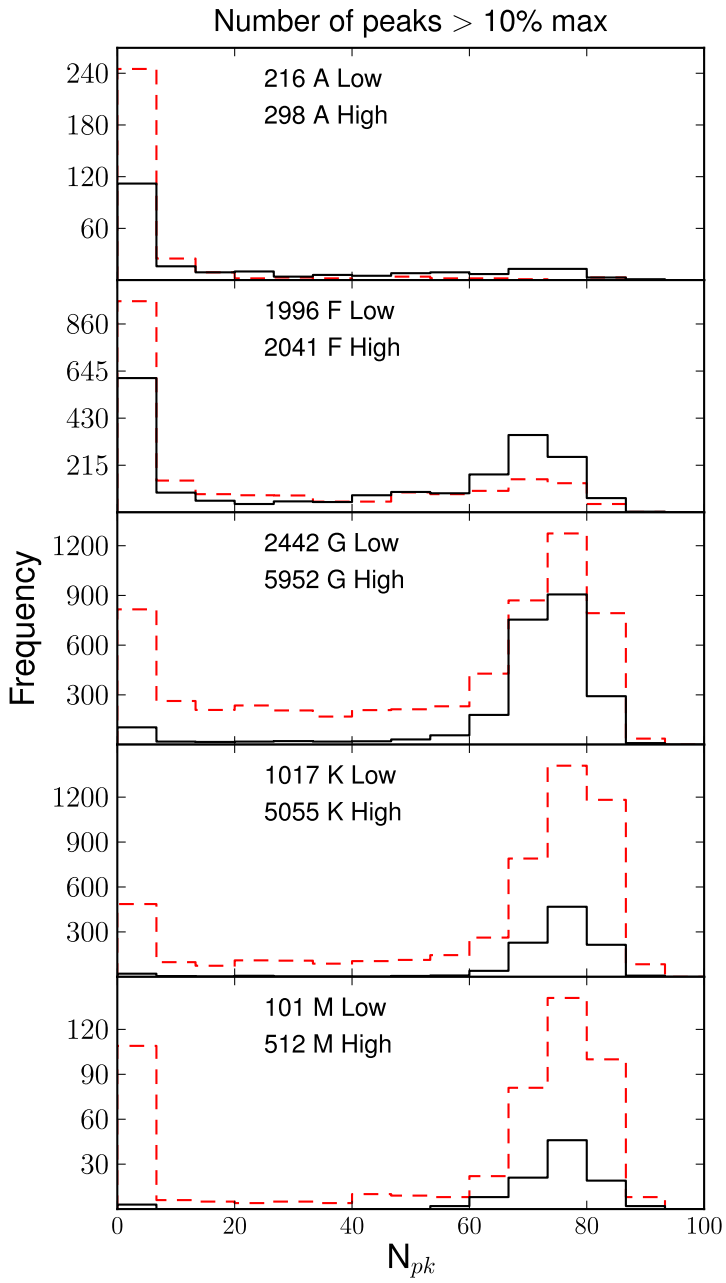

Figure 3.12: Distribution of the number of peaks with power greater than 10% of the maximum peak for low (solid line) and high (dashed line) variability samples, where $S \geq 0.3$.

Table 3.4: Best fit parameters for the Harvey models shown in Figure 3.13. SpT denotes the spectral type of the star, Background is the fixed constant included in the model, and the following columns show the values A, B and C described in Equation 3.6, for the 1^{st} and 2^{nd} model components.

SpT	Background	A_1	B_1	C_1	A_2	B_2	C_2
A	1.9e7	4.1e9	1.0e6	1.0	2.9e6	1.6e3	4.8
F	4.3e7	4.3e9	1.2e5	1.7	9.1e6	1.6e3	4.8
G	1.2e8	1.1e10	1.6e5	2.1	1.8e7	1.6e3	4.8
K	2.0e8	1.2e11	2.8e5	2.3	4.2e7	1.6e3	4.8
M	3.2e8	1.9e12	3.6e5	2.8	1.2e8	1.6e3	4.8

where ν is frequency, A_i is the amplitude of the ith component, B_i is its characteristic timescale, and C_i is the slope of the power law. By considering the power spectra as a sum of N different AR(1) spectra, one can fit a model of this form, first introduced by Harvey (1985) and previously used to model stellar spectra (see e.g. Aigrain et al., 2004; Michel et al., 2009).

The Harvey model consists of a constant background level, a component to describe the intrinsic power at each frequency, and a second component to describe the extra power at $\sim 8 \times 10^{-5}$ Hz introduced by the onboard motor vibrations. The timescale and slope of the motor component have been set to the mean value found when the parameters are allowed to vary, because there is no reason they should be different between spectral types. The amplitude for each is allowed to vary because it will have a greater effect for the fainter stars.

The median power spectrum and fitted powerlaw models describe the typical stochastic background variability of the stars. Clear trends through the spectral classes can be seen in the fit parameters, listed in Table 3.4. As expected, the background constant increases steadily towards the fainter classes. The amplitude, timescale and slope of the powerlaw fit also increases towards later spectral types. This may arise from the slower typical rotation periods associated with later type stars, creating large, slowly evolving active regions. The timescale of the A stars does not fit this trend, which is most likely due to the large number of pulsating stars shifting the median power towards slightly longer timescales, and the fact that the frequency resolution is not sufficient to resolve the powerlaw turnoff.

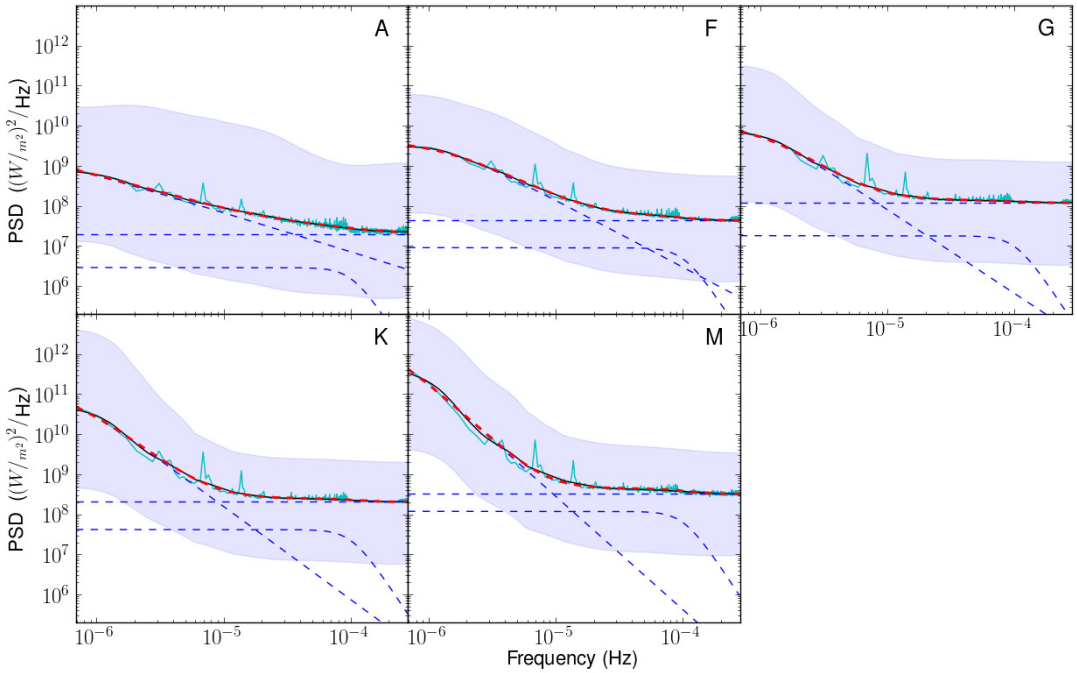

Figure 3.13: Harvey model fits to the median periodogram for each spectral type. The blue shaded area represents the region between 10-90th percentiles of variability. The cyan line shows the unsmoothed median spectrum, and the fit (red dashed line) is made to the smoothed spectra (black line). The individual components of the fit are shown as blue dashed lines (see text for full explanation).

3.6 DISCUSSION

Following B11, I quantified the variability level of each star as the interval between the 5^{th} and 95^{th} percentile of its light curve, and focused my study on the main-sequence stars, separating them from giants using the T_{eff}-dependent $\log g$ cut defined by C10. A division between low and high variability was made using a magnitude dependent cut in R_{var} at twice the equivalent variability measurement for the active Sun, resulting in a split of 80% low and 20% high. I examined the relationship between the nature of the variability and physical properties of the stars, and compared results to that of C11, who use the PDC data and B10,11, who use a polynomial fitting correction method. As suggested by B10 and C11, there is clear evidence for a decrease in variability with increasing temperature (Figure 3.4), which cannot be entirely explained by selection effects or the noise floor at fainter magnitudes. This is equivalent to an increase of

variability with increasing $B - V$, as noted by Isaacson & Fischer (2010), who use $S_{\rm HK}$ as an indicator of chromospheric activity.

The spatial distribution and proper motion of the stars were used to look for an indication that the low and high variability samples come from different stellar populations. C11 suggest that the high variability sample belongs to a younger population, existing closer to the galactic plane than the older low variability sample. I note a weak correlation between contamination fraction and variability, which could potentially increase the apparent variability in the more contaminated regions close to the galactic plane. The proper motion distributions displayed in Figure 3.8, indicate that, for the M stars at least, more active stars tend to have slightly lower proper motions. This may suggest they are younger, although distance estimates and radial velocity measurements would be needed to interpret the proper motion distributions robustly, and it should be noted that giant contamination could affect this result.

There may exist interesting further correlations between the variability level and other parameters listed in the KIC. For example, I compared the surface gravity and metallicity at different variability levels, and did find apparently statistical significant differences (in K-S test terms), however, these parameters have larger uncertainties than $T_{\rm eff}$, and are effected by non-negligible biases as a function of Galactic position (because of crowding and extinction, see e.g. Verner et al., 2011, and references therein), which cannot easily be disentangled from actual differences between populations. Any biases affecting the $\log g$ information in the KIC may also have an indirect effect on the present study because they could affect the separation between dwarfs and giants. Verner et al. (2011); Brown et al. (2011) investigated the reliability of the KIC parameters in detail, and concluded that the $T_{\rm eff}$ and $\log g$ estimates are reliable within the stated uncertainties. This implies that any biases should have only a very minor effect on the C11 dwarf/giant cut used.

One of the most interesting results in this work is the period distribution for each spectral type, displayed in Figure 3.10. Although this should be considered a preliminary test due to the short timespan of the dataset, there is clear evidence that, for light curves with a clear periodic component, the period increases towards later spectral types. The fraction of periodic stars also varies with spectral class, increasing

towards later types to a maximum in the K stars, before reducing again in M stars. The relatively high number of stars in which we were able to identify a periodic or quasi periodic modulation in this dataset alone suggests that, when further quarters of *Kepler* data are added, it should be possible to measure periods for a significant fraction of all *Kepler* stars – and thus calibrate the evolution of angular momentum for intermediate and low-mass stars on the main sequence to an unprecedented level. This direction of study is explored in more detail in Chapters 5 and 6.

The stochastic component of the variability was investigated by the following two methods. First, I measured the number of periodogram peaks with power greater than 10% of the maximum for the stars which pass the periodicity selection threshold. For all spectral type, the distribution of this statistic, $N_{\rm pk}$, has a peak at low values, corresponding to clearly periodic light curves dominated by a single frequency, or a small number of frequencies, as would be expected for pulsating stars. However, there is also another peak at high $N_{\rm pk}$, whose amplitude increases towards later spectral types, reaching a maximum in the K stars and decreasing slightly in the M stars (as did the periodicity fraction). This peak corresponds to quasi-periodic light curves, as expected for rotational variables with evolving active regions.

I also parameterised the stochastic component of variability by fitting auto-regressive models (also known as Harvey models) to the median power spectrum for each spectral type. Because of the relatively limited frequency resolution (caused by the short duration of the dataset used in this study), the fitted model parameters cannot be considered definitive, but they enable a preliminary inter-comparison. I find that the typical amplitude, timescale and power-law index all increase towards later types.

This is consistent with other tests I performed, and supports a broadly coherent picture of main-sequence variability. The hotter, earlier spectral classes show a lower variability level on the whole, and the variables tend to show clearly periodic behaviour on short time-scales, as expected from pulsations. The shape of their power spectra suggest that these stars possess smaller active regions that evolve more quickly. By contrast, the cooler, later type stars show larger amplitude variability on longer time-scales, with quasi-periodic rather than periodic behaviour, and appear to possess

more slowly evolving, larger active regions. The K stars have the highest fraction of light curves with a significant periodicity, but these are also the most complex, with the largest number of significant frequencies per object.

As more *Kepler* quarters become available and systematics treatment is improved, similar investigations could provide vastly improved constraints on the periodicity measurements, and reveal changes in the variability though time (i.e. pulsation and activity cycles).

The question of variability measurements using other metrics and timescales, for example the CDPP on transit timescales calculated by Gilliland et al. (2011), is also a possible extension to this work.

Finally, the results presented in this chapter also have implications for planet detection and followup, in both photometric and radial velocity programs (Aigrain et al., 2012), which I will attempt to quantify in future work.

4

MEASURING STELLAR ROTATION: THE ACF METHOD

The material in this chapter has been published, together with the material in Chapter 5, in McQuillan, A., Aigrain, S. & Mazeh, T., 2013, MNRAS, 432, 2, 1203.

In this chapter I describe a new method developed to measure rotation periods from star spot modulation in light curves, using the autocorrelation function (ACF). In Section 4.1 I briefly describe the motivation behind stellar rotation observations. I then discuss existing methods of rotation period measurement in Section 4.1.2, and then present the ACF method in Section 4.2.

4.1 INTRODUCTION

Measuring rotation rates for large numbers of stars over a wide range of masses and ages is a long-standing goal in stellar astronomy, not only to understand the physical mechanisms, but also to calibrate the relationship between period $P_{\rm rot}$, age t and stellar mass M, enabling age estimates to be made for individual stars (Kawaler, 1989; Barnes, 2003).

4.1.1 *Previous Work*

Until the 1990s, stellar rotation rates were mainly measured from spectroscopy, via the rotational broadening of absorption lines. These rotational velocity measurements provided key insights, particularly the well-known spin-down law for Sun-like stars, $P_{\rm rot} \propto \sqrt{t}$ (Skumanich, 1972). However, these measurements yielded only model-dependent constraints on the rotation rate, and were limited to relatively fast rotators. Using modern wide-field detectors, it is possible to measure $P_{\rm rot}$ directly, by monitoring the brightness of large numbers of stars simultaneously, and detecting quasi-periodic brightness variations, which arise as magnetically active regions on the star's surface rotate in and out of view.

The development of the photometric approach to rotation period determination has led to a vast increase in the number of rotational studies and period determinations in the last decade. These studies are too numerous to list here, for a representative example see the Monitor project (see Irwin & Bouvier, 2009, and references therein). Most of them focus on open clusters, because stars belonging to a given open cluster can be assumed to have a common, and relatively well known age (e.g. from isochrone fitting). Some rotation measurements have been made for older field stars, for example Baliunas et al. (1996); Kiraga & Stepien (2007); Hartman et al. (2011) and Goulding et al. (2012). However, ground-based observations are limited in precision by the Earth's atmosphere and in time-sampling by day-night variations, weather and telescope scheduling constraints. This means that photometric rotation periods can be measured from the ground only for very active, young stars. Thus rotation period measurements have remained very scarce for stars older than the Hyades ($\sim 650\,$Myr). Space-based transit surveys such as CoRoT and *Kepler* now provide an ideal dataset to search for rotation periods in older stars with lower activity levels and longer periods.

4.1.2 *Common Methods of Period Measurement*

Standard approaches to period detection in light curves are based on Fourier decomposition or, for irregularly sampled data, least-squares fitting of sinusoidal models and variants thereof (Scargle, 1982; Zechmeister & Kürster, 2009). However, typical stellar light curves are neither sinusoidal nor strictly periodic, probably because of the clumpy and time-evolving nature of the underlying active region distribution. Residual instrumental systematics are often present as well.

These effects can all lead to a complex periodogram structure, with spurious peaks from jumps and long term systematics, and multiple or split peaks from spot evolution or differential rotation. It is therefore challenging to determine which peak corresponds to the rotation period, without a priori knowledge of the range of rotation periods expected. Consequently, Fourier-domain methods are not always the best suited to make the most of *Kepler*'s many thousands of spot modulated light curves, which display a wide range of rotation periods. I present an alternative approach based on the autocorrelation function (ACF) of the light curves. To my knowledge, this is the first time that the ACF is used as the primary tool to detect stellar rotation periods, although Affer et al. (2012) used it as a secondary verification tool.

Because the ACF measures only the degree of self-similarity of the light curve at a given time lag, it is expected to be more robust to active region evolution than the periodogram, which implicitly assumes a stable, sinusoidal signal. he ACF method should also be capable of producing more robust results in cases with residual instrumental systematics, because correlated noise, long-term trends and discontinuities give rise to monotonic trends in the ACF.

4.2 THE AUTOCORRELATION FUNCTION - METHOD

4.2.1 *The Autocorrelation Function*

In signal processing, the ACF takes the standard form

$$r_k = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}, \tag{4.1}$$

(see e.g. Shumway & Stoffer, 2010) where r_k is the autocorrelation coefficient at lag k, for time series x_i ($i = 1, \ldots, N$). Each lag k corresponds to $\tau_k = k\Delta t$, where Δt is the cadence. In my implementation, the light curves are median normalised before the ACF is computed, and I only search for periods less than half the length of the dataset, i.e. $k < N/2$.

The period measurement involves three steps: identifying peaks in the ACF, selecting the peak associated with the mean rotation period, if any, and evaluating the uncertainty on the period.

The presence of high-frequency noise in the light curves leads to numerous local extrema in the ACF. To address this problem, the ACF needs to be smoothed before the peaks of interest can be identified (see Figure 4.1). I experimented with a number of smoothing approaches and eventually settled on convolution with a Gaussian kernel as being the one which gave the most satisfactory results when applied to both simulated examples and real *Kepler* data. The kernel width needs to be adjusted so minimise the ACF noise without smoothing out features associated with the rotational variability. In the case of the *Kepler* data, a window size of 56 lags (corresponding to 1.2 days) and a full-width at half-maximum of 26 lags ($\sim$ 0.5 days) were found to be suitable. This does not prevent detection of short period objects, which are instead limited by the cadence of the *Kepler* data.

If the light curve contains a clear rotational modulation signal, this process yields a series of clear, regularly spaced peaks of gradually decreasing height, as seen in Figure 4.2. The first peak corresponds to the interval between patterns in the light curve, which evolve gradually, but are clearly repeated, and is thus identified as the rotation period. Some of the light curves contain long term trends and discontinuities,

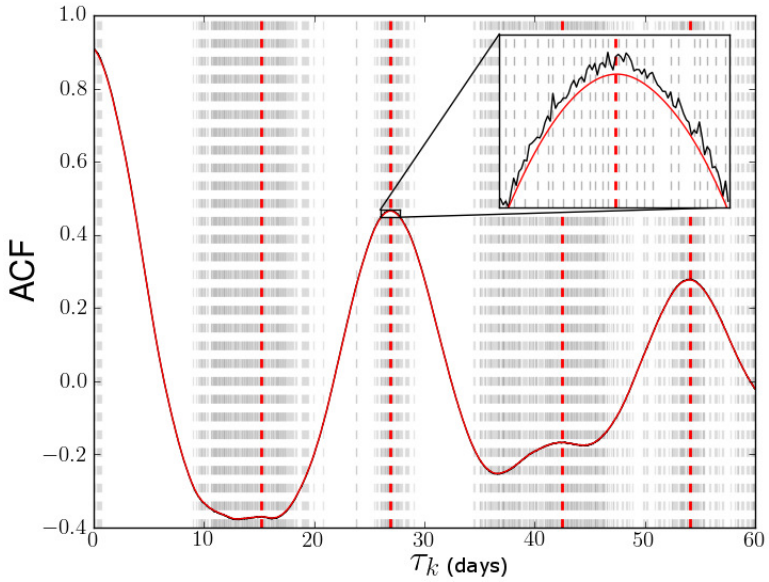

Figure 4.1: Example of an unsmoothed ACF (black curve) and corresponding smoothed curve (red), which is indistinguishable except in the detailed section shown in the inset. The many small peaks detected in the unsmoothed curve are shown as grey vertical dashed lines and the peaks detected from the smoothed version are shown as red vertical dashed lines.

as a result of imperfections in the systematics correction. These introduce power at low frequencies, and thus affect the behaviour of the ACF for large lags, but the first ACF peak still corresponds to the correct period (as identified by visual examination of the light curve).

In the remaining steps, I make use of the height of the ACF peaks. However, correlated noise and residual systematics can introduce underlying long term trends, which mean the absolute peak height is no longer a good diagnostic. To mitigate this effect, the height of each peak is measured relative to the two adjacent minima, and the mean of the two measurements is adopted as the 'local height' of the peak, denoted by $h_{\rm P}$. Since the ACF values range from -1 to 1, $h_{\rm P}$ has a positive value, of maximum 2.

If a star has two dominant active regions located on opposite hemispheres, each causes a series of dips in the light curve, approximately in anti-phase with each other. This gives rise to a partial correlation at half the mean rotation period, leading to a peak in the ACF at $\tau_k = P/2$. However, the peak at $P/2$ is typically smaller than that at P, because the two active regions do not generally give rise to identical dips. I

therefore adopt the following condition: if h_P of the second peak is greater than that of the first, the second peak is selected instead. The right panel of Figure 4.4 shows a synthetic example of this effect, where this method has detected the correct peak.

In some cases, correlated noise and residual systematics produce an underlying slope at small τ_k, causing a shift the position of the first peak associated with the rotation period. This occurs because a peak on a slope will have its maxima shifted in the direction of increasing gradient, as seen in the left panel of Figure 4.4. To avoid this bias, a more robust period measurement is obtained using the median of the intervals $\Delta\tau_k$ between consecutive ACF peaks associated with the rotation period. To avoid selection of erroneous peaks, only those located at or close to (within 20%) of integer multiples of τ_k of the selected peak are used. A further step removes erroneous multiple detections around a single ACF peak, ensuring only 1 data point per peak is used in the period measurement. Since the accuracy of the peak positions can decrease for very large τ_k, a maximum of 10 peaks was selected for measurement of the median period and uncertainty.

The period uncertainty is defined as the scatter of the $\Delta\tau_k$. Specifically,

$$\sigma_P = \frac{1.483 \times \mathrm{MAD}}{\sqrt{N-1}} \tag{4.2}$$

where N is the number of peaks, and MAD is the median of the absolute deviations from the median $\Delta\tau_k$. This 'MAD-estimated scatter' is equivalent to the standard deviation for a Gaussian distribution, but is more robust to outliers. In cases where I identified only one peak matching the selection criteria, I adopt the peak position as the period and the half width at half local peak height as the period uncertainty.

4.2.2 *Comparison of the ACF and Periodogram*

I compare the ACF method to the method most commonly used to search for rotation periods in stellar light curves, namely least squares fitting of sinusoids over a grid of trial periods (Irwin et al., 2006; Zechmeister & Kürster, 2009). The amplitude, phase and zero-point of the sinusoid are free to vary. The sine-fitting periodogram is

expressed in terms of the statistic

$$S = \left(\chi_0^2 - \chi^2\right) / \chi_0^2, \tag{4.3}$$

where χ_0^2 is the reduced chi-squared of the light curve with respect to a constant value, and χ^2 is the reduced chi-squared with respect to the best-fit sinusoid. This method is described in more detail in Section 3.5.1.

Figure 4.2 shows two synthetic time-series curves, together with their ACFs and least-squares sine curve fitting periodograms (Zechmeister & Kürster, 2009). The left column shows a strictly periodic signal, for which the periodogram displays a clear pronounced peak. The ACF displays an oscillatory behaviour, with regularly spaced peaks located at multiples of the period. The amplitude of these peaks decays gradually because of the definite duration of the time-series. The right column shows the effect of introducing correlated noise to the signal in the left hand column.

Figure 4.3 demonstrates how the ACF and the periodogram are affected by varying signal phase and amplitude, systematics effects and noise. In cases where the phase and amplitude of the signal vary with time, the correct period is detected but the ACF peak amplitude varies within an envelope, corresponding to the amplitude variation of the signal. The peak width and level of symmetry can also vary. In this case, the periodogram produces two peaks on both sides of the correct period.

If the signal contains multiple minima and maxima per period, as can occur for spotted stars with more than one dominant active region, the ACF often shows alternating low and high ACF peaks, due to a partial correlation between the sets of maxima or minima. The ACF algorithm is built to identify these cases (see Section 4.2.1), and in this case selected the right period. On the other hand, the periodogram picked half the right period.

A jump or long term trend in the signal introduces a long term trend in the ACF, and since this can take many shapes, it is important to look at the local variations in peak height when performing diagnostics on the ACF. These long term trends introduce a long-period peak in the periodogram, which can lead to a wrong identification of long periodicity. It is therefore expected that the ACF method will be more reliable in these cases.

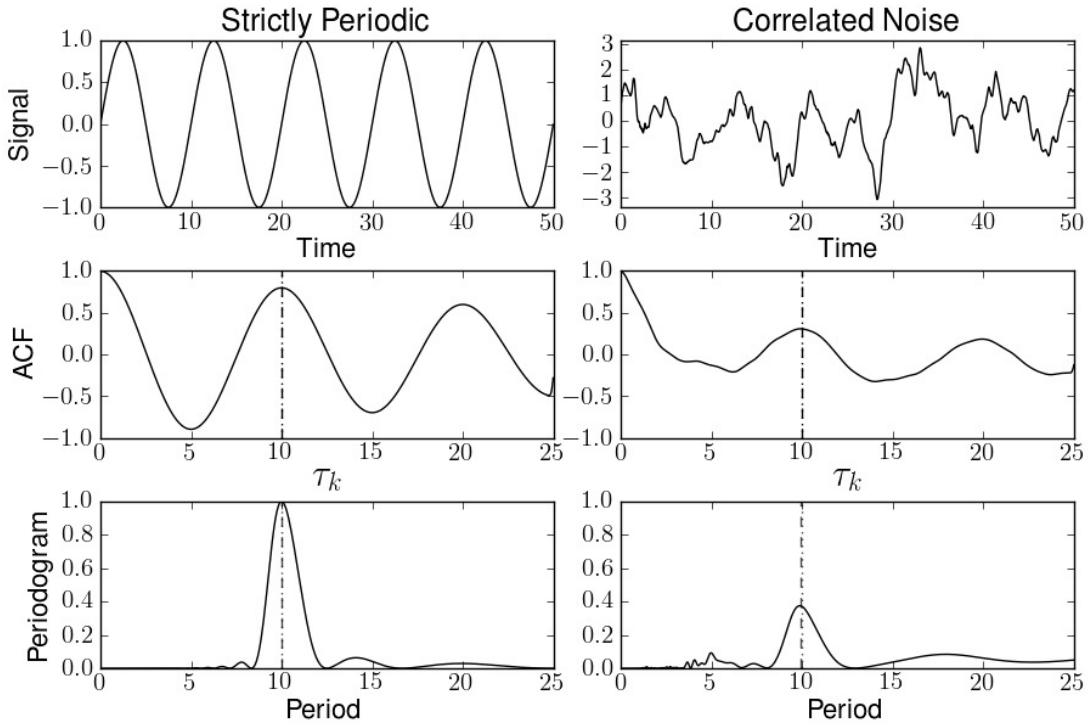

Figure 4.2: Simple synthetic signal of amplitude 1 (top row), and corresponding ACF (centre row) and periodogram (bottom row). The right hand column shows the effect of introducing noise of amplitude 0.9, correlated on a 2 time unit timescale. On the ACF and the periodogram panels, the input period used to generate the signal is shown as a vertical dotted line, detected period from each corresponding method is marked as the over-plotted dashed line.

Noise and systematics are present in many stellar light curves displaying rotational modulation, and must be accounted for when attempting to determine rotation periods using the ACF. The effect of combining these factors is shown in the right column of Figure 4.3, which has phase and amplitude modulation, white and correlated noise, a linear trend and a jump.

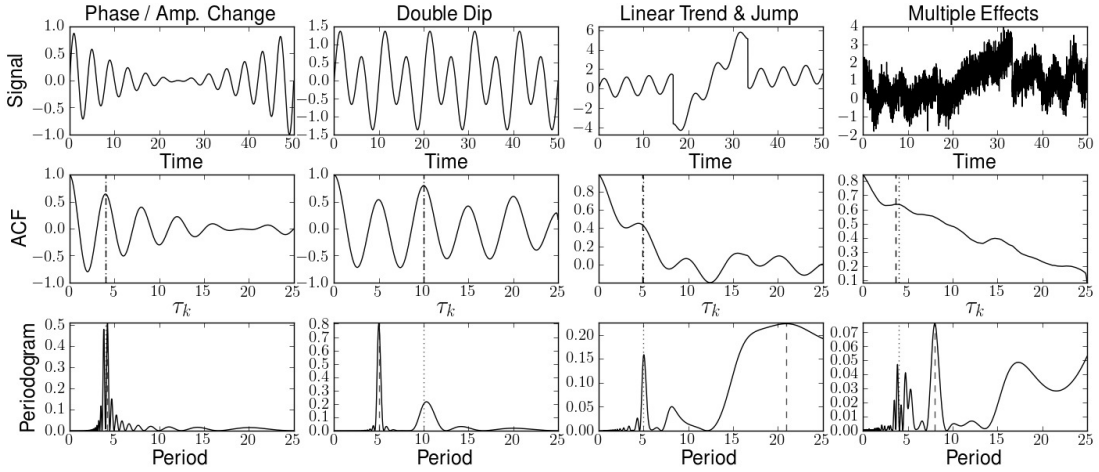

Figure 4.3: Synthetic examples showing the effects of varying signal phase and amplitude and injecting noise and systematics into the sine curves (top row), on the ACFs (middle row) and periodograms (bottom row). The 'Multiple Effects' signal comprises of the 'Phase/Amp. Change' signal from the first column, with injected white noise, correlated noise, a linear trend and a jump. On the ACF and the periodogram panels, the input period used to generate the signal is shown as a vertical dotted line, detected period from each corresponding method is marked as the dashed line, overlaying the dotted line in most cases.

4.3 THE AUTOCORRELATION FUNCTION - TESTS

4.3.1 *Tests with Realistic Light Curves with Injected Simulated Stellar Modulation*

In this section I show the results of a small number of representative, realistic light curves, containing realistic spot-induced variability, and noise properties, but where the ground truth is known. This is not intended as a thorough study of the ACF over the full parameters space, since this would be too time consuming at this stage, given that the parameter space of spot and noise models is vast. This is something I plan to evaluate more fully in future work.

To generate synthetic light curves I used a simple spot model code, described in Aigrain et al. (2012), which takes the following input parameters: number of spots, light curve amplitude, characteristic spot half-life (λ) in rotation periods, light curve duration, fractional differential rotation, inclination and time sampling.

This model assumes the spots are small enough to ignore projection effects, spot

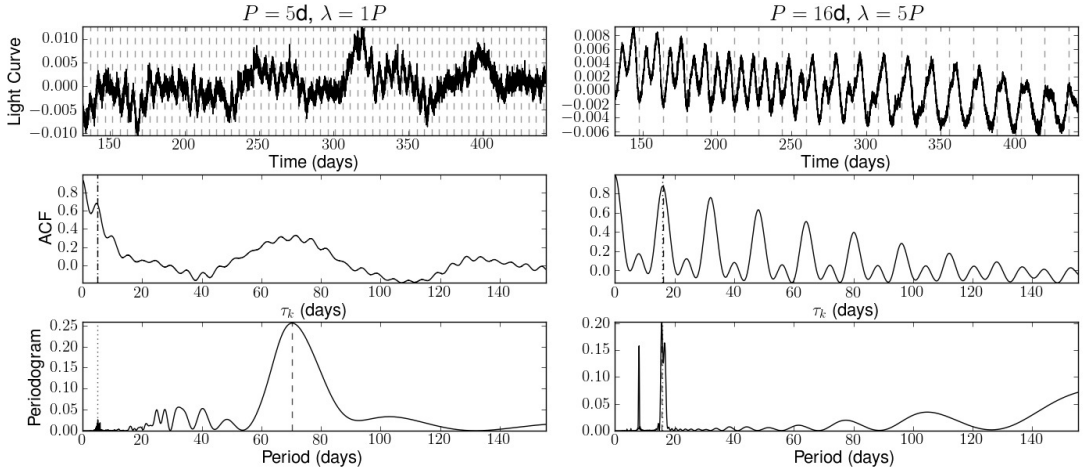

Figure 4.4: Examples of synthetic light curves and corresponding ACFs and periodograms for two different combinations of period, P, and spot half-life, λ, indicated in the plot titles. The right panel shows a short rotation period with fast evolution and in the left panel, the spot distribution was selected to give a double-dip light curve. The dashed lines on the light curve plots indicate intervals at the period detected by the ACF. On the ACF and periodogram plots, the input period is marked by a dotted line and the detected period is marked by a dashed line.

shapes and overlapping. Limb darkening is also ignored. The relative drop in flux due to a single point-like, dark spot rotating on the stellar surface is

$$F(t) = f\text{MAX}\{\cos\beta(t); 0\}\,, \tag{4.4}$$

where $f = 2(1-c)(1-\cos\alpha)$ gives the relative flux drop for a spot at the disc centre, c is the contrast ratio between the spot and the unspotted photosphere, α is the angular radius on the surface of the star, and $\beta(t)$ describes the angle between the spot normal and the line of sight. The observed stellar flux is given by

$$\Psi(t) = \Psi_0[1 - F(t)]\,, \tag{4.5}$$

where Ψ_0 is the flux in the absence of spots.

The observations of Jackson & Jeffries (2012) show that the photometric variability caused by rotational modulation of active region is better explained by a large number of small spots, than a small number of large spots. They reach this conclusion after finding no difference in the stellar properties or chromospheric activity levels of two large groups of low mass stars in the young cluster NGC 2516, which display very

different levels of photometric spot modulation. Therefore, an axisymmetric covering of small spots is a logical explanation.

They suggest a spot number of 2500-5000, however, since they observe younger stars I opt for a smaller spot number of 200. This also increases the speed of synthetic light curve generation and the main difference in output is the zero level, which is not used in this study. I selected an amplitude of 1% and an inclination angle of 75°. The time sampling and duration match that of the *Kepler* Q1-4 data.

Since the effects of the noise and residual systematics of the *Kepler* data are an important factor when analysing the ACF, these are introduced to the synthetic light curves. To generate correlated noise in the light curves, white Gaussian noise was smoothed using a boxcar filter.

A random selection of the *Kepler* M-dwarf light curves were visually examined and 10 were selected which appeared to contain only noise and systematic effects, with negligible intrinsic stellar signal. For each synthetic output, I created 10 light curves with realistic noise and systematics by adding each of the selected quiet *Kepler* light curves.

The two synthetic examples in Figure 4.4 demonstrate the ability of the ACF to recover the input period from quite difficult cases, where the periodogram failed to identify the right period.

The many synthetic cases tried, led us to believe that the ACF method and the periodogram are less reliable for light curves with very fast spot evolution, short periods ($<$ 7 days) and strong systematics. This is caused by the steep initial slope in the ACF, combined with the smoothing algorithm, which masks the peaks. Similarly for fast evolution, strong systematics and long periods ($>$ 35 days) the ACF peaks are too strongly affected by underlying trends. The periodogram suffers mainly from spurious peaks at long periods.

As the spot half-life is increased to 1 period, the periodic signal strength increases, and by λ = 5 periods, the majority of periods are correctly detected, across the range of input values. It should be noted that these synthetic examples were designed to test the limits of the ACF and therefore all have a low signal to noise. A recent study of a sample of Sun-like stars observed by CoRoT (Mosser et al., 2009) found that spot

lifetime of the order of the rotation period are not atypical for Sun-like stars. On the other hand, the spot lifetime is also known to increase with increased activity level (Hall & Henry, 1994).

I tested the effect on the ACF and periodogram detections when the light curve is scrambled. When the quarter orders are changed the correct period is recovered by the ACF providing the period is significantly less than the 33 day shortest section length. For periods close to the shortest section length both the ACF and periodogram are shifted slightly from the input period. For periods longer than the shortest section length the period is not recovered by either the ACF or periodogram.

4.4 DISCUSSION

In this chapter, I have presented the ACF method for detection of rotation periods in star spot modulated time series photometry. Because the ACF measures only the degree of self-similarity of the light curve at a given time lag, the period remains detectable even when the amplitude and phase of the photometric modulation evolve significantly during the time-span of the observations. The ACF method is also capable of producing robust results in cases with residual instrumental systematics, because correlated noise, long-term trends and discontinuities give rise to monotonic trends in the ACF.

I conclude that the periodogram remains a valuable technique for period detection, however the clarity and robustness of the ACF method makes it my tool of choice for the measurement of stellar rotation in light curves. The ACF method works independently of periodogram based methods and is ideal for determining rotation period statistics for long datasets.

In Chapter 5 I describe the application of the ACF method to the M-dwarf sample observed by *Kepler*, which demonstrates the ability of this new approach.

The ACF method I have developed is a valuable addition to the existing rotation period detection methods, because of its ability to cope with non-harmonic, evolving signals and with residual instrumental effects. If it can be automated further, its

application to *Kepler* data would enable the first truly systematic survey of rotation rates across a very large sample of stars spanning a wide range of ages and spectral types. As it stands, it requires a visual check for every object, but this is fairly streamlined, and the method can already be applied to many thousands of targets. This is what I have done in the remainder of the thesis, producing what I believe are by far the largest samples of rotation periods for field M-dwarfs (Chapter 5), main sequence Sun-like stars (Chapter 6), and planet host stars (Chapter 7) available to date.

Much of the future development of the ACF method will rely on extensive simulations, which could not be carried out during the time frame of this DPhil. Instead I opted to prioritise the application of the method to the *Kepler* data, relying on visual verification of results, since rotation in the *Kepler* data is currently a very active topic.

As well as the mean rotation period determination, the ACF may provide a new approach to the measurement of active region evolution, by calibrating the decay of the ACF local peak heights. Another application may be to measure differential rotation using a metric based on the widths of the ACF peaks, although this will need to first be disentangled from the effects of evolution. These studies would require further tests on synthetic light curves.

An important step to enable the application of the ACF method to the entire dataset, is the automation of the period detection verification stage. There are clear indications that ACF statistics such as local peak height, spacing and decay timescale relate to the strength of the period detection. However, determining a threshold using these parameters will again require extensive simulations and evaluation of detection efficiencies.

5

ROTATION OF THE *kepler* M-DWARFS

The material in this chapter has been published, together with the material in Chapter 4, in McQuillan, A., Aigrain, S. & Mazeh, T., 2013, MNRAS, 432, 2, 1203.

In this chapter I present the results obtained by applying the autocorrelation function (ACF) method of rotation period determination to the *Kepler* M-dwarf sample. I describe the target selection and light curve preprocessing in Section 5.2. The period measurement is detailed in Section 5.3, with a comparison to results obtained using a periodogram approach included in Section 5.4. An investigation of the dependency of period on fundamental stellar parameters is described in Section 5.5, and a detailed examination of the relationship between period and stellar mass is presented in Section 5.6.

5.1 INTRODUCTION

In this chapter I focus specifically on the rotation of M-dwarfs, which are of particular interest amongst main sequence stars due to their internal structure. M-dwarfs are largely or fully convective, and hence it's not clear that the processes which generate the large scale magnetic field, and hence drive the angular momentum evolution, in the solar type stars, can occur in M dwarfs. By comparing the rotational properties

and photometric variability levels of M-dwarfs with those of solar type stars, this can be seen. It is also valuable to compare rotational properties of stars either side of the convective boundary, which lies in the M-dwarf regime at $\sim 0.35\,M_{\odot}$. Although the *Kepler* sample (according to the KIC values) does not extend to masses this low, the results can be compared to those in the literature from lower mass samples. The few examples of old (field) M-dwarf rotation periods in the literature (Irwin et al., 2011; Kiraga & Stepien, 2007) indicate trends in the period-mass plane that are quite distinct from the behaviour of earlier type stars, involving a steep increase towards longer rotation periods.

M-dwarfs spin down very slowly and even at the age of the Hyades some are still rotating very rapidly. This provides an opportunity to study a wider range of evolutionary states than for earlier spectral types, which converge more quickly onto a single sequence, losing information about their initial states.

M-dwarfs are an important target for exoplanet searches, since they are the most numerous stars in our galaxy, and their low mass and small size leads to favourable detection biases in transit and radial velocity surveys. Their low luminosity also allows for shorter period planets, which are more easily detected, to lie in the habitable zone. The greatest concern for habitability of planets orbiting M-dwarfs is the effect of high level of magnetic activity associated with these stars (Tarter et al., 2007), and potential for tidal locking.

This high level of magnetic activity leads to large amplitude light curve modulation due to star spots, which benefits rotation period detection. Since the M-dwarf sample observed by *Kepler* is a manageable size for visual confirmation of detections, it presents the ideal dataset with which to begin the application of the ACF method.

5.2 TARGET SELECTION AND PREPROCESSING

Following Ciardi et al. (2011), M-dwarfs were selected among the *Kepler* targets based on the effective temperature $T_{\rm eff}$ and surface gravity $\log g$ reported in the *Kepler* Input Catalog, or KIC (Brown et al., 2011). These parameters were estimated using Bayesian

Table 5.1: Number of objects included at each stage of the study. The column labelled 'All dM' includes both M-dwarfs selected from the KIC parameters and the previously known M-dwarfs without KIC parameters. The totals in the second column include the known EBs and planet candidate host stars, which are also listed separately in the last two columns. The possible giants and double-period stars were removed from the final results reported in Tables 5.2 & 5.5. Note: removed 4 contact EBs from Q1-4 EB selection.

Stage	All dM	EBs	Pl. host
KIC selection	2937	15	57
Q1-4 light curves	2483	9	51
Period detected	1730	7	42
Possible giants	121	0	0
Double-period	39	0	0
Binaries or pulsators	112	-	0
Final periodic total	1570	7	42

posterior probability maximisation (Brown et al., 2011), matching observed colours, estimated from Sloan g, r, i, z filters, 2MASS JHK, and D51 (510 nm), to the stellar atmosphere models of Castelli & Kurucz (2004), resulting in typical uncertainties of 200 K for $T_{\rm eff}$ and 0.5 dex for $\log g$, respectively. These are the typical errors for each parameter, but the actual errors may vary by a small amount between stars of different magnitude and spectral type. For a more detailed discussion of the KIC parameters, see Brown et al. (2011), Batalha et al. (2010) and Verner et al. (2011). There is increasing evidence that the KIC values for the radii and $T_{\rm eff}$ are overestimated (e.g. Mann et al., 2012; Verner et al., 2011; Muirhead et al., 2012), which means the values of $T_{\rm eff}$ should be interpreted with care. A revised set of $T_{\rm eff}$ estimates were recently published by Dressing & Charbonneau (2013), but were too late to be incorporated in this work, so I will revisit this work in the future using the revised parameters. A preliminary test confirmed they do not alter the conclusions of this chapter.

This chapter focusses on M-dwarfs stars, which were selected following Ciardi et al. (2011) as having $T_{\rm eff} \leq 4000$ K and $\log g \geq 4.0$ dex, manually adding 25 known M-dwarfs (Ciardi et al., 2011) with missing KIC parameters. Table 5.1 summarises the number of stars considered at each stage of the study.

5.2.1 *Obtaining and pre-processing the light curves*

This study is based on public release 14, which was available when the present analysis was performed. The release included data from Quarter 1 (Q1) to Quarter 4 (Q4) of *Kepler* observations, which took place over $\sim$ 310 days between May 13$^{\text{th}}$ 2009 and March 19$^{\text{th}}$ 2010. Approximately 156 000 targets were observed during this time, most with a cadence of 29.42 minutes. These data are publicly available and were downloaded from the *Kepler* mission archive (http://archive.stsci.edu/kepler) at the Space Telescope Science Institute (STScI).

Public release 14 included a re-processing of Q1–Q4 with a new version of the presearch data conditioning (PDC) pipeline known as PDC-MAP. The purpose of the PDC is to remove the majority of instrumental glitches and systematic trends. In contrast with earlier versions, the new PDC-MAP uses a Bayesian approach to do so while retaining most real (astrophysical) variability (Smith et al., 2012; Stumpe et al., 2012). It works very well for Q3 and Q4, but some artefacts remain in some light curves in Q1 and Q2. Improving the correction of instrumental effects further remains an important goal to enable the full exploitation of *Kepler*'s potential in terms of stellar astrophysics. However, in the vast majority of the cases I have examined, the residual artefacts in the PDC-MAP data do not prevent the detection of rotational modulation, so it is considered suitable for the present work. See Chapter 2 for a detailed discussion of the correction of systematic effects in the *Kepler* data.

Additional quarters of PDC-MAP data have been made publicly available since this work. I am planning to analyse these in future work, in the hope that it will improve the sensitivity to long periods, but I do not expect it to affect the main conclusions of this chapter.

The ACF calculation requires the light curves to be regularly sampled and normalised to zero. I divided the flux in each quarter by its median and subtracted unity. Gaps in the light curve longer than the *Kepler* long cadence were filled using linear interpolation with added white Gaussian noise. This noise level was estimated using the variance of the residuals following subtraction of a smoothed version of the

flux. This method of gap filling and quarter stitching was only found to introduce spurious jumps in light curves for which the systematics correction had failed, and were therefore already problematic. In the vast majority of cases, the amplitude of M-dwarf activity mean that any quarter stitching effects have a negligible impact on the period detection.

To smooth the flux I applied an iterative nonlinear filter which consists of a median filter followed by a boxcar filter, both with 11-point windows, with iterative 3-sigma clipping of outliers. The sample also includes a number of previously identified eclipsing binaries and planetary transit candidates (Prša et al., 2011; Batalha et al., 2012). In those cases I cut out and interpolated over the eclipses/transits before performing the rotation period search.

5.3 PERIOD MEASUREMENT

5.3.1 The ACF

The ACF method is described in Chapter 4. Given the scientific importance and manageable size of the sample under study here, I opted to perform a visual examination of all the light curves in order to verify the period detected by the ACF. The light curves were compared to dashed vertical lines at intervals of the detected period (see e.g. Figure 5.1). In order to classify the detection as valid, features in the light curve must be present at intervals matching the dashed lines, across several periods, preferably in more than 1 quarter. In cases where the ACF period detection clearly matches the visually detected period in the light curve, I denote these period detections to be 'correct', although it is acknowledged this test of correctness differs from that of simulations where the true period is known with certainty.

Particular attention was paid to the question of whether the detected period was clearly the rotation period, or could be $P/2$, as described in Section 4.2.1. In 73% of periodic light curves, features were identified which could be 'tracked' visually, i.e. a particular shape of spot crossing that repeats throughout the light curve. When there

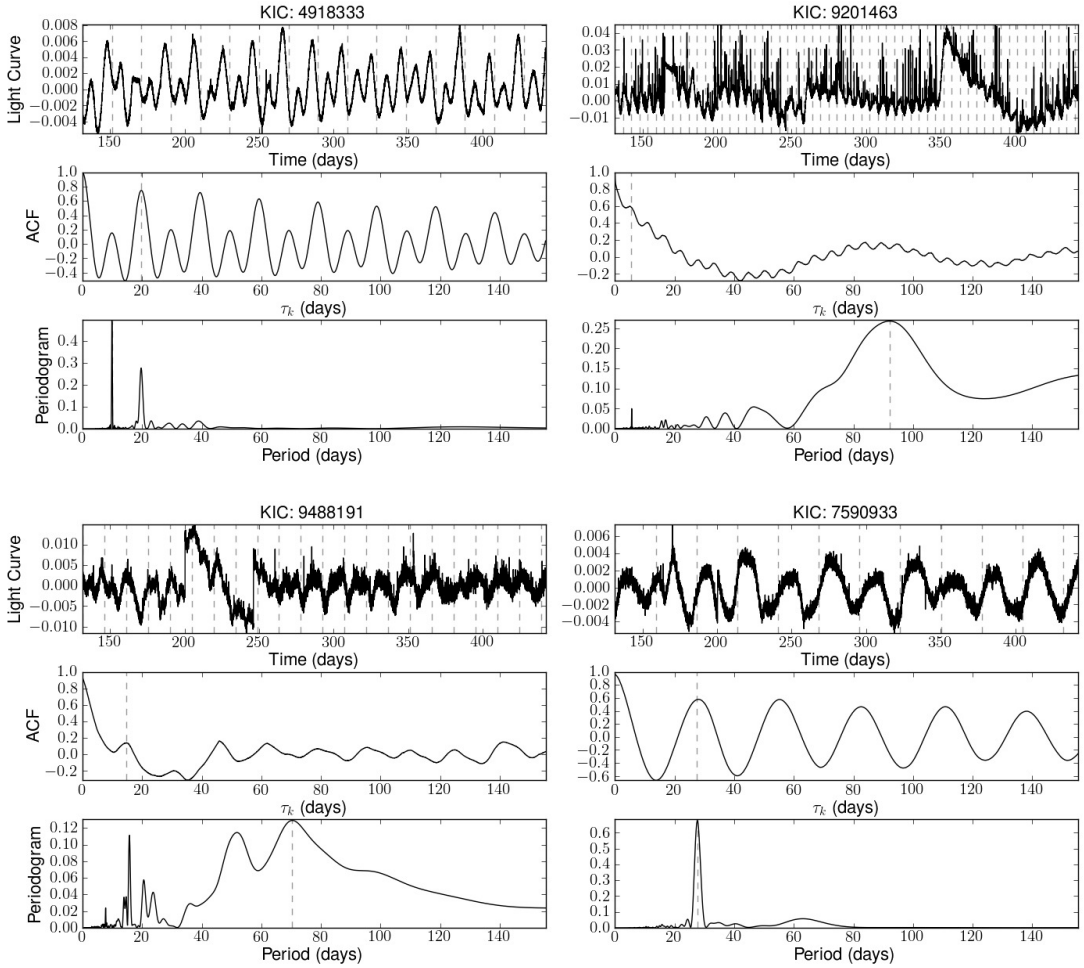

Figure 5.1: Examples of *Kepler* light curves and corresponding ACFs. The dashed lines on the light curve plots indicate intervals at the period detected by the ACF (dashed line on ACF plot). KIC 4918333 shows two active regions, creating a double dip effect. The correct period was automatically detected by selecting the second (higher) ACF peak. KIC 9201463 shows that the ACF is robust against flares and significant systematics. KIC 9488191 contains significant residual systematics and the effect on the ACF is apparent, however, the correct period is still detected. KIC 7590933 shows an example which has been marked as 'possible harmonic' or 'HM' in Table 5.2, since it is not clear whether the rotation period corresponds to the first or second ACF peak, even though the ACF peak height indicates it should be the first (see text for further discussion).

are more than one such sets of features, and they evolve gradually relative to each other, as shown for example in the top left panel of Figure 5.1, the rotation period is very clear. When only one such set is visible and the first ACF peak is not higher than the second (e.g. the bottom right panel of Figure 5.1), we cannot be so certain, but the simplest explanation is nonetheless that the detected period is correct. In around 1% of the cases where period is reported, there is very tentative evidence in the light curve that the detected period is a harmonic (i.e. $P/2$), but this cannot be certain, so I simply flagged the corresponding objects as 'possible harmonics'. Such cases may be resolved once additional quarters of *Kepler* data are incorporated into the analysis. It should be noted that these cases are not numerous enough to explain the two sequences seen in the period-temperature diagram (Figure 5.6), and are spread across the entire period range.

In 3.6% (57) of the periodic cases an incorrect peak was identified. This was most frequently a result of noise introducing extra peaks, or very large residual systematics changing the relative peak heights. These cases were manually corrected to select peaks corresponding to the period identified by eye.

In the future, I plan to apply the same analysis to a much larger sample of *Kepler* targets (including F, G and K dwarfs), for which an automated detection method must be developed. The present, visually inspected subset will then prove valuable, to validate any threshold to be applied automatically to a larger sample. Here I merely note that, in the present study using 10 months of data, it is possible to recover 91% of the detections, at the cost of a false alarm rate of 10%, by selecting objects with

$$h_{\rm P} > {\rm MAX}\left(0.15, \frac{\sigma_{\rm P}}{51\ {\rm days}}\right) , \tag{5.1}$$

where $h_{\rm P}$ is local ACF peak height, defined in Section 4.2.1.

5.3.2 *Periodic Sample*

The ACF method was applied and visual inspection steps performed as described in the previous section to the *Kepler* light curves of the objects selected as likely M-dwarfs.

Table 5.2: M-dwarfs with detected rotation periods. This table is available in its entirety, in a machine-readable form at www.physics.ox.ac.uk/StellarRotation or on the data disc provided to the examiners. A portion is shown here for guidance regarding its form and content. $T_{\rm eff}$ and $\log g$ are from the KIC and M was derived from $T_{\rm eff}$ using the 600 Myr isochrone of Baraffe et al. (1998). The average amplitude of variability per period bin of the light curve, $R_{\rm per}$, is included. The meaning of the flags are: 'EB': known eclipsing binary (Prša et al., 2011); 'PL': planet host candidate (Batalha et al., 2012); 'PB': ultra-stable periodic behaviour, indicating possible binary, pulsator or young object; 'HM': the reported period may be a harmonic of the true period; 'NF': no flag.

KIC	$T_{\rm eff}$ (K)	$\log g$ (g/cm^3)	M ($M_\odot$)	$P_{\rm rot}$ (days)	$\sigma_{\rm P}$ (days)	$R_{\rm per}$ (mmag)	Flag
1162635	3899	4.62	0.5037	15.509	0.064	10.7	NF
1430893	3956	4.41	0.5260	17.144	0.046	10.4	NF
1572802	3990	4.48	0.5394	0.368	0.000	74.8	PB
1721911	3833	4.58	0.4781	28.403	0.394	3.9	NF
1866535	3878	4.50	0.4955	25.052	0.136	4.0	NF

The results are reported in Tables 5.2 to 5.5. Full versions of these tables are available in HTML and machine-readable form at the URL www.physics.ox.ac.uk/StellarRotation, together with plots of every light curve, its ACF, and sine-fitting periodogram. Table 5.2 reports all the period measurements, except for two groups of objects, listed separately in Tables 5.3 and 5.4, which were excluded from the sample for reasons detailed below. Table 5.5 lists all the objects which passed the target selection criteria, but for which no period was detected. The final number of likely M-dwarfs with detected rotation periods is 1570 (out of the 2483 light curves analysed, or 2362 objects later found to be giants are discounted).

5.3.3 *Excluding Non-Rotators*

Rotational modulation of star spots is not the only cause of periodic, or quasi-periodic, variability in stellar light curves. The visual examination stage was therefore important in identifying groups of stars whose variability did not appear to be caused by rotation, or by rotation alone.

The *Kepler* light curves have already been searched extensively for planetary

Table 5.3: Objects identified as likely giants from their light curves. Errors in the KIC can lead to false $\log g$ or $T_{\rm eff}$ values, such that these stars were not originally classified as giants. This table is available in its entirety, in a machine-readable form at www.physics.ox.ac.uk/StellarRotation or on the data disc provided to the examiners. A portion is shown here for guidance regarding its form and content.

KIC	M ($M_\odot$)	$\log g$ (g/cm^3)	$T_{\rm eff}$ (K)	$R_{\rm var}$ (mmag)
1026895	0.5343	4.53	3977	8.99
1160867	0.4474	4.57	3753	8.29
1431599	0.5068	4.5	3907	12.68
1576043	0.4889	4.52	3861	7.76
2010137	0.4881	4.48	3859	7.01

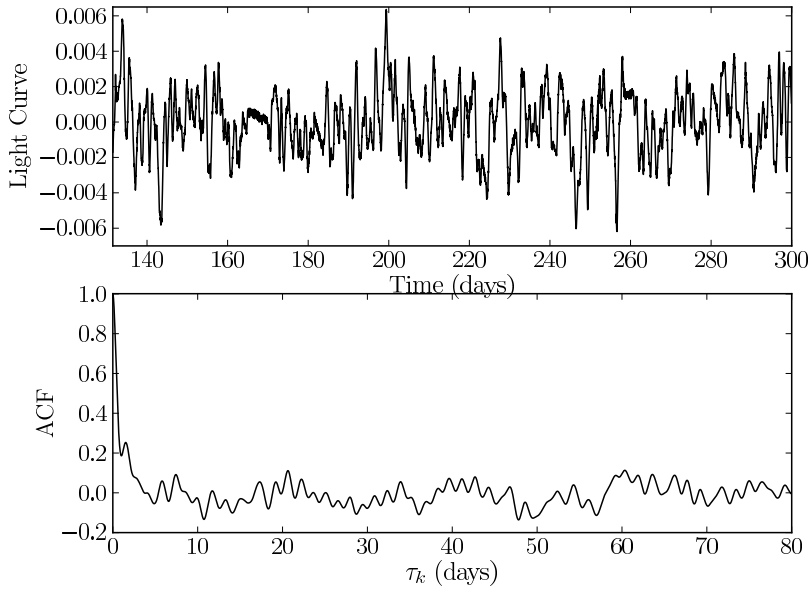

Figure 5.2: KIC 2010137 is an example of a star identified as a likely giant based on its light curve, which shows stochastic variability with a clear, dominant time-scale, but no repeatability, and does not resemble the signal expected from rotation.

transits (Borucki et al., 2011; Batalha et al., 2012) and stellar eclipses (Prša et al., 2011). These objects were included in this sample, after removing the transits and eclipses, as described earlier. Measuring the photometric rotation periods of known binaries is useful to check for differences between the rotational properties of close binaries and of apparently single stars. The rotation of planet-host stars is an important topic in itself, which is examined in more detail in Chapter 7.

Table 5.4: Objects with two distinct periods in their light curves. This table is available in its entirety, in a machine-readable form at www.physics.ox.ac.uk/StellarRotation or on the data disc provided to the examiners. A portion is shown here for guidance regarding its form and content.

KIC	M ($M_\odot$)	$\log g$ (g/cm^3)	$T_{\rm eff}$ (K)	$R_{\rm var}$ (mmag)
892376	0.4704	4.47	3813	13.84
1569863	0.4688	4.45	3809	33.94
2557669	0.4585	4.36	3782	37.7
3646734	0.4372	4.47	3726	25.89
3735772	0.4414	4.48	3737	85.71

One distinct group of 121 stars show a clear ACF peak at relatively short τ_k ($\sim$ 1–12 days), but there are few additional peaks at integer multiples of the first, and the light curve appears stochastic rather than truly periodic. Such an example is shown in Figure 5.2. I initially thought that these objects may be rapid rotators with very rapidly evolving active regions, or a hitherto unidentified type of pulsating star. I examined their KIC parameters, and noted that they appear redder, and have lower proper motions, than the rest of the sample, suggesting they might be giants. Indeed, if the $J - H < 0.75$ cut advocated by Ciardi et al. (2011) is applied, it removes 117 stars from the M-dwarf sample, of which 103 belong to this group of objects displaying stochastic behaviour. I therefore concluded that these objects are likely giants, removed them from the M-dwarf sample and list them separately in Table 5.3.

A further 39 objects display evidence for two, very distinct periods in their light curves, with one period several times longer than the other. This could potentially result from the light of two different variable stars being included in a single photometric aperture, which may be revealed by examination of the pixel level data. Since the nature of the periodicity is unknown, and the determination of the KIC parameters used to select these objects as M-dwarfs could be affected by the presence of a close companion, I excluded these objects from the sample, and listed them separately in Table 5.4. I have not included periods for them, because the presence of two distinct signals in their light curves makes the identification of either period more challenging.

Finally, 109 stars show unusually stable periodic behaviour compared to the bulk of the period stars observed, with periods typically < 7 days, and very little or no

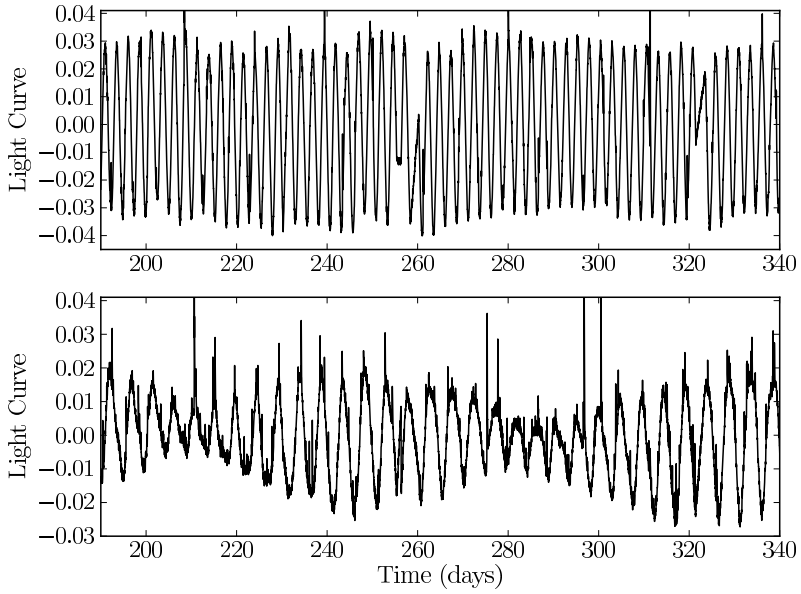

Figure 5.3: Examples of the objects classified as 'PB'. KIC 5516671 (top) shows a very stable light curve and KIC 3103752 (bottom) shows the 'beat pattern' discussed in the text.

evidence of any evolution over the full Q1–Q4 duration. Some of these also show a beat pattern characteristic of the light curves containing two, mutually similar periodicities. See Figure 5.3 for examples. These objects could be members of close binary systems, where the active region pattern is stabilised over long timescales because of the presence of a companion. Their rotation periods and amplitudes are certainly similar to those of the known eclipsing binaries in the sample. If they are binaries, the beat patterns may indicate differential rotation on one of the stars, or two slightly different rotation periods for the two components of the binary. I am not aware of any type of main-sequence M-dwarf that would be expected to pulsate in this period range, but cannot rule out binarity or pulsation without spectroscopy. I therefore kept these objects in Table 5.2, but flagged them as 'binaries or pulsators', indicating their period may result from a phenomenon other than rotation.

Table 5.5: Objects with no rotation period detection. This table is available in its entirety, in a machine-readable form at www.physics.ox.ac.uk/StellarRotation or on the data disc provided to the examiners. A portion is shown here for guidance regarding its form and content.

KIC	M ($M_\odot$)	$\log g$ (g/cm^3)	$T_{\rm eff}$ (K)	$R_{\rm var}$ (mmag)
1160684	0.5244	4.48	3952	3.35
1292688	0.5146	4.86	3927	7.48
1569682	0.5154	4.69	3929	4.99
1718059	0.526	4.49	3956	3.13
1718071	0.5225	4.54	3947	5.77

5.4 COMPARISON OF THE AFC AND SINE-FITTING PERIODOGRAM

I computed sine-fitting periodograms for the sample of *Kepler* M-dwarfs using 1000 logarithmically-spaced periods between 0.1 and 155 days.

In general, similar results are obtained from the ACF or periodogram methods, however, I consider the ACF clearer and more reliable. A useful feature of the ACF method is that it enables automatic identification of cases where the detected period is a half the mean rotation period. An example of this is shown for KIC 4918333 in Figure 5.1, which clearly shows that the mean rotation period corresponds to the second ACF peak, whereas the first periodogram peak is the highest.

Similarly, for KIC 9488191 in Figure 5.1, one can determine easily from the ACF that the rotation period corresponds to the first peak at $\sim$ 15 days and not the highest peak in the periodogram at $\sim$ 70 days. Although long period peaks could be excluded from the periodogram to allow the correct peak to be selected, this would require prior knowledge of the possible range of rotation periods to avoid removal of genuine long period signatures.

A more systematic comparison between the ACF and periodogram results is shown in Figure 5.4. Here a one-to-one comparison is made of the periods detected using the ACF and the sine-fitting periodogram, for all the cases where the light curves were determined to show periodic variations from visual examination.

In 222 (14%) of the periodic cases, the periodogram does not detect the correct period (to within 10%). The most common discrepancy arises from cases where

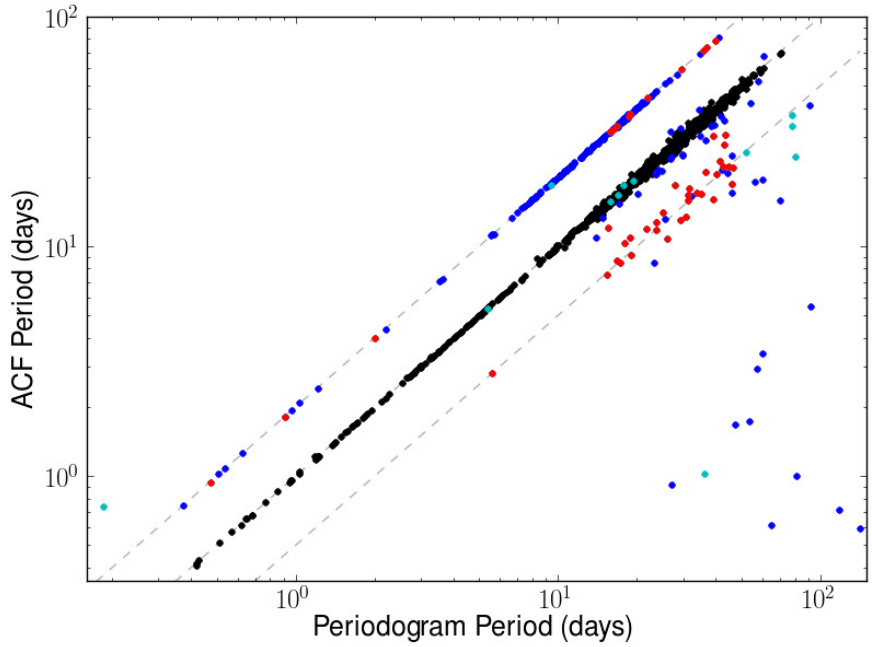

Figure 5.4: Comparison of the periods detected using the ACF and the periodogram, for all stars with a visually confirmed periodicity and ACF detection. The correct ACF detections where the periodogram detections are within 10% of the ACF period are shown in black (1298), the cases where the periodogram period is not within 10% of the ACF period are shown in dark blue (215). Cases where the ACF initially selected the wrong period but the periodogram selected the correct one are marked by red points (45). Cases where both the ACF and periodogram detected the wrong period are marked as cyan points (12). The grey dashed lines show where $P_{\mathrm{ACF}} = P_{\mathrm{LS}}$, $P_{\mathrm{ACF}} = 2 \times P_{\mathrm{LS}}$ and $P_{\mathrm{ACF}} = 0.5 \times P_{\mathrm{LS}}$.

the periodogram has detected half the period of the ACF (170 cases). There is no obvious way of identifying these cases automatically using the periodogram alone (for example, I tried to use the relative heights of the periodogram peaks, without success). In 11 cases the periodogram selects erroneous long period peaks and the ACF selects the correct shorter period peak.

On the other hand, the ACF yielded 57 incorrect ACF period detections (3.6%), 45 cases occur where the ACF method selects the wrong period, and the periodogram selects the correct period. These cases are marked as red points in the Figure 5.4. For the remaining 12, both the ACF and periodogram detect the wrong period (cyan points).

In one case (KIC 10553513, not included in the periodic sample), the large residual

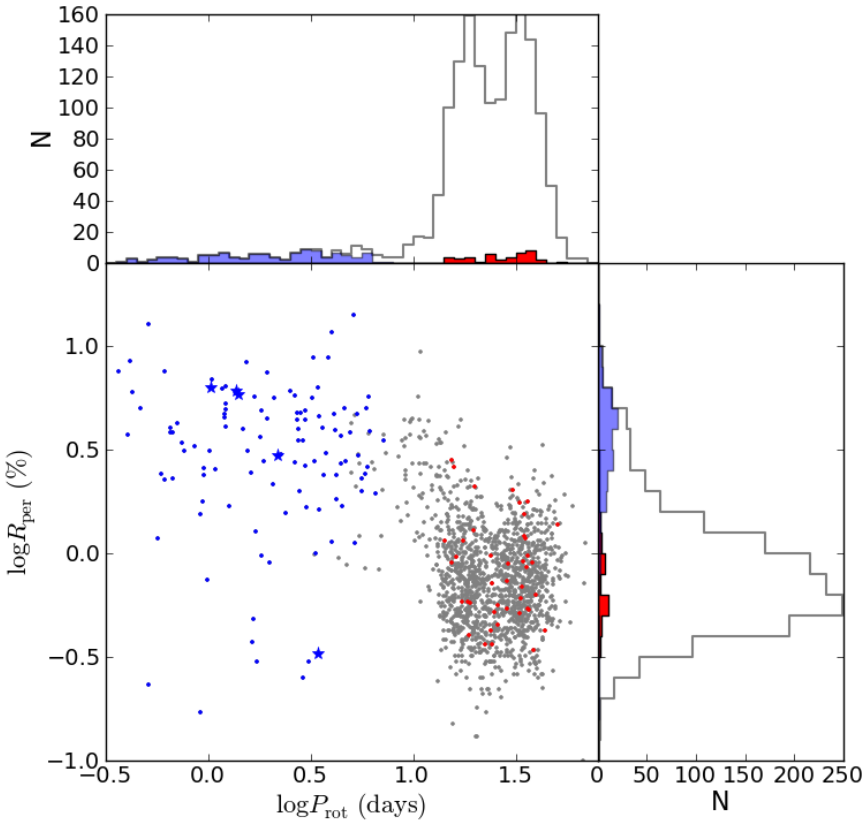

Figure 5.5: Period versus amplitude for the rotating *Kepler* field M-dwarfs. Blue dots represent objects with $P_{\rm rot} < 10$ days, which also display unusually stable modulation patterns in their light curves, and blue stars known, short-period eclipsing binaries (Prša et al., 2011). Red dots represent the host stars of candidate transiting planets (Batalha et al., 2012). All the other M-dwarfs with detected rotation periods are shown as grey dots. The histograms of each parameter are shown along the corresponding axis, with matching colours. Two long-period binaries are not indicated in the plot.

systematics prevent either the ACF or the periodogram from detecting the rotation period which is visible in the light curve. The only case for which the periodogram is able to detect the period and the ACF is not (even with manual correction) is KIC 5480340 (not included in the periodic sample). This is due to the extremely short period (0.25 days). The steep gradient in this region of the ACF prevents peak detection, which I plan to resolve in a future version of the algorithm.

I conclude that the periodogram remains a valuable technique for period detection, however the clarity and robustness of the ACF method makes it my tool of choice for the measurement of stellar rotation in light curves. The ACF method works

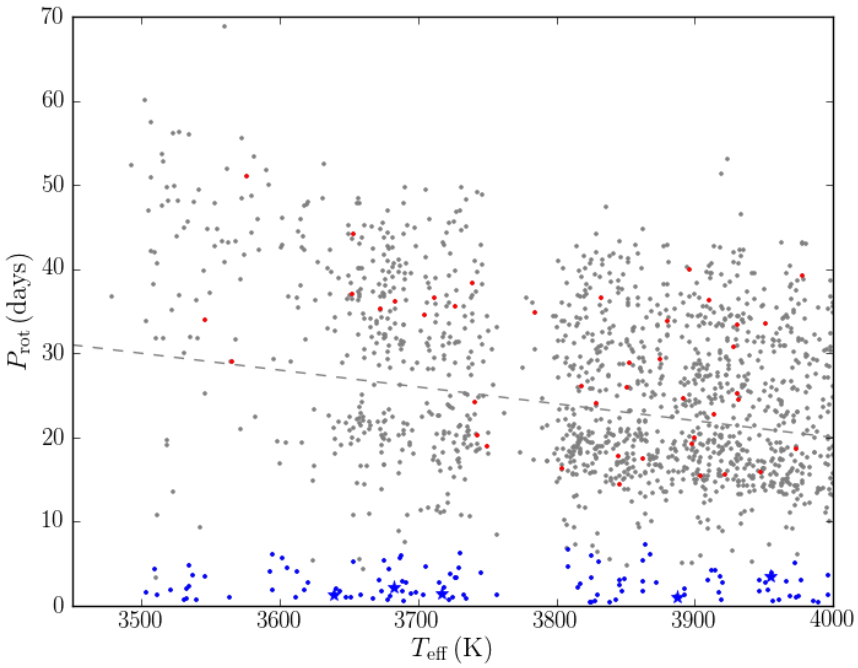

Figure 5.6: Period versus effective temperature for the rotating *Kepler* field M-dwarfs. The symbols and colours as the same as in Figure 5.5. The apparent dearth of objects around $T_{\rm eff} = 3775\,{\rm K}$ is not apparent in any of the KIC colours, and is therefore thought to be a result of the KIC stellar parameter estimation procedure, rather than a real effect. The dashed line marks the location of the cut made by eye between the fast and slow rotators.

independently of periodogram based methods and is ideal for determining rotation period statistics for large datasets.

5.5 PERIOD, AMPLITUDE AND TEMPERATURE

I now examine the period distribution of the sample as a function of amplitude (Figure 5.5) and effective temperature (Figure 5.6). A number of interesting features are immediately apparent.

5.5.1 Bimodality

First, the period distribution is clearly bimodal for $P_{\rm rot} > 10\,{\rm days}$, with peaks of approximately equal height at ~ 19 and $\sim 33\,{\rm days}$. This bimodality appears stat-

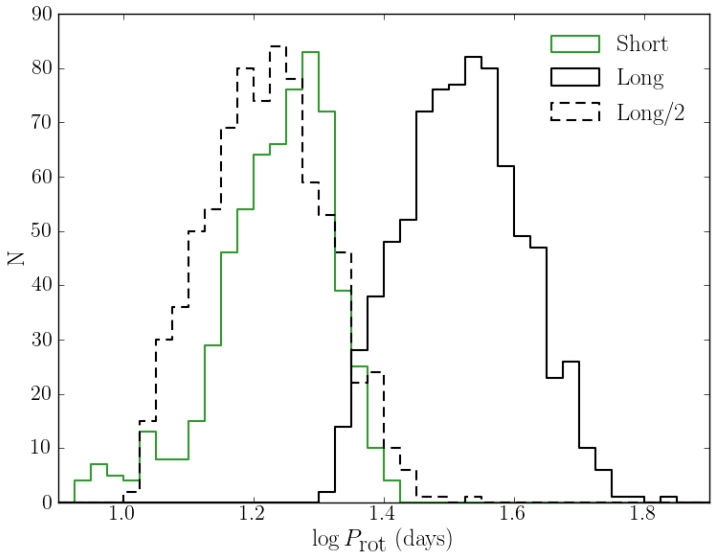

Figure 5.7: Histogram of the short and long period M-dwarfs, as defined by the cut marked with the dashed line in Figure 5.6. The dashed histogram marks the long periods divided by 2, and provides further evidence that the short period set are unlikely to be erroneous half-period measurements of the long period population.

istically significant in log period space, with a Hartigan's dip test p-value of 0.0003 (Hartigan & Hartigan, 1985). The two peaks of the rotation period distribution form two distinct sequences in period-temperature space (Figure 5.6). The period decreases with increasing temperature, in much the same way, for both sequences. This variation as a function of temperature implies the gap between the sequences is not a result of systematics in the *Kepler* light curves leading to missed detections in a particular period range. I define the short and long period samples based on the line, plotted somewhat arbitrarily, shown in Figure 5.6.

Before interpreting this result, I address the possibility that the bimodal distribution is spurious: an error of a factor of two in the periods of about half the objects could give rise to the observed distribution. Such errors are not uncommon in rotational studies based on ground-based data (see e.g. Collier Cameron et al., 2009). However, they are less likely to be prevalent in the present study, for the following reasons. One can exclude a scenario where twice the true period is measured for the objects belonging to the longer period peak, because of the continuous sampling of this data. The alternative scenario is that half the true period is measured for the

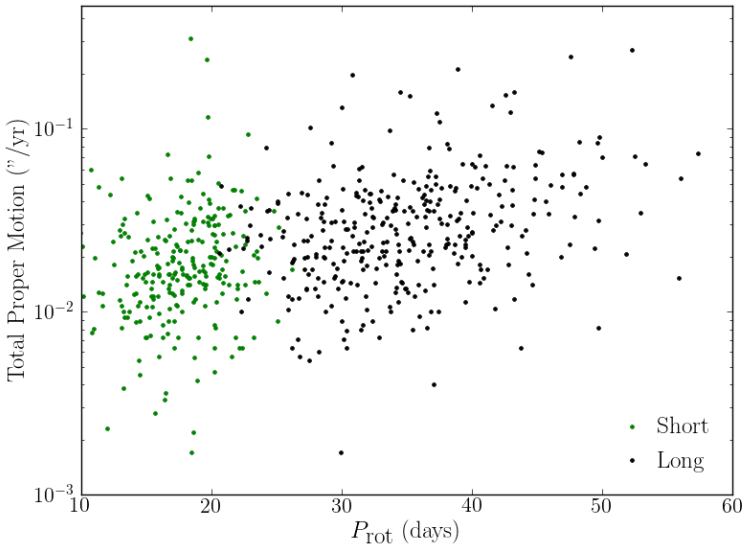

Figure 5.8: Total proper motion against Period for the short period (green) and long period (black) M-dwarfs, as defined by the cut marked with the dashed line in Figure 5.6.

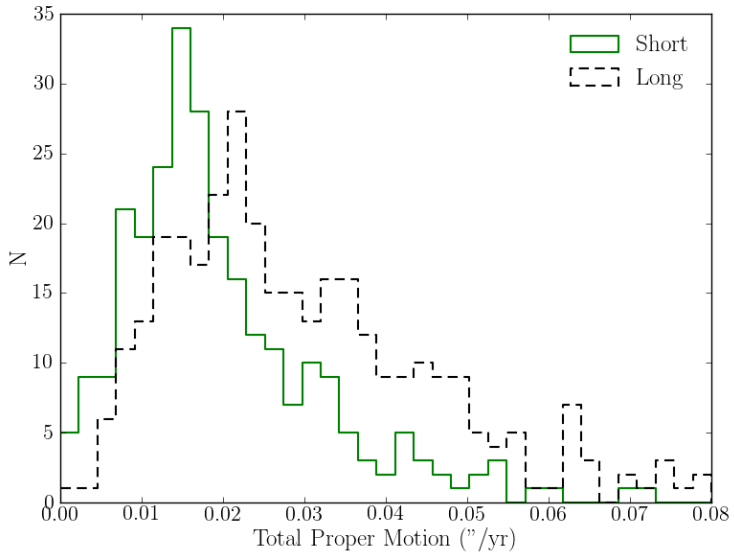

Figure 5.9: Histogram of the total proper motion for the short period and long period M-dwarfs, as defined by the cut marked with the dashed line in Figure 5.6. The uncertainty on the proper motion values is 0.02"/year but given the high numbers in each sample these results are still significant. The known and 'possible' binaries have been removed from this sample (blue points and stars in Figure 5.6).

objects belonging to the shorter period peak. This kind of problem arises when the brightness distribution of the stellar surface is bimodal, due to concentrations of active regions on opposite hemispheres, resulting in 'double-dip' light curves. However, the ACF peak selection routine was specifically designed to address this problem – this is one of the strengths of the method. In fact, for 210 of the objects belonging to the shorter period peak, a 'double-dip' light curve lead to the second ACF peak being automatically selected as the rotation period. Unless these stars have quadrupolar surface brightnesses, it is very unlikely that their periods are underestimated by a factor of two.

5.5.2 *Comparison of Sample Properties*

The case for genuine bimodality is further strengthened by comparing the detected distribution to that which would arise if either set were wrong by a factor of 2. Figure 5.7 shows that division of the long period set by 2 does not reproduce the short period set (and equivalently neither does multiplication of the short period group). A quantitative two-sample Kolmogorov-Smirnov (KS) test of the short period sample and the long period sample divided by 2 gives a p-value 9×10^{-14}, confirming they are very unlikely to be drawn from the same distribution.

Figure 5.8 shows a weak correlation between period and total proper motion for the two samples. There is also tentative evidence of a difference in the proper motion distributions of the short and long period samples. Figure 5.9 shows the histograms for stars in the short and long period samples with non-zero proper motion KIC measurements (266 short, 345 long). The proper motion values from the KIC are taken from a selection of catalogs[1] where available. Total proper motion is listed on NStED as having accuracy of 20 milliarcseconds per year, but the large number of objects in each sample support the difference between them. We do not draw strong conclusions from this test, since only a small number of the targets have proper motion measurements and the potential biases are not well understood. A more accurate test

[1] *Kepler* Stellar Classification Program, Hipparcos, Tycho-2, UCAC2, 2MASS and USNO-B1.0.

would be to compare tangential velocity, taking into account the distance to the target, but distances to the *Kepler* targets are hard to derive with the current data. Since we are considering only M-dwarfs within a narrow range of magnitudes, we consider them the distances to be similar enough to perform a basic test using proper motion, which should at least reveal a clear difference if, for example, one sample consisted predominantly of misclassified giants.

I also checked for differences in the Galactic latitude and *Kepler* magnitude distributions of the two samples but did not see any clear differences. By using the galactic latitude measurements and an approximate distance estimate from average apparent and observed M-dwarf magnitudes, I conclude that the observed population lies within a small fraction of the scale height of the thin disk, and therefore a variation of properties with Galactic latitude is not expected.

For the remainder of this study, I therefore assume the sample contains two distinct stellar populations. The fact that these have different median periods *and* different proper motion characteristics suggests that they may result from two distinct waves of star formation: as already discussed, low-mass main-sequence stars spin down as they age, and older populations tend to have larger velocity dispersions than younger populations, due to dynamical heating of the Galactic disk over long timescales (see Freeman & Bland-Hawthorn, 2002, and references therein). The ratio of the two peaks in the period distribution corresponds to an age ratio of ~ 3, if one assumes that M-dwarfs spin down as $t^{0.5}$, as observed for Sun-like stars (Skumanich, 1972). Recent rotational studies of low-mass stars in open-clusters suggest that the index of the spin-down low might be closer to 0.6 (see e.g. Meibom et al., 2009), which would imply an age ratio of ~ 2.5. I note that an interpretation in terms of thin and thick disk populations is unlikely because there should be many fewer thick disk than thick objects in the *Kepler* sample, which is essentially magnitude-limited.

Further kinematic and distance information is required to draw stronger conclusions about the nature of the two populations. The *Kepler* mission itself may, in the medium term, provide this information: the pixel-level data can be used to monitor the centroid of each star over the full lifetime of the mission, which should enable the measurement of proper motions and parallaxes for some targets. Monet et al. (2010)

shows that astrometric precision for a single 30 minute measure is $<$ 4 milliarcseconds. In the longer term, the GAIA mission (de Bruijne, 2012) will provide this information with greater precision and for more objects.

While the bulk of the objects (89%) have periods in the range 10 – 50 days, a small fraction (8%) of the periods are $<$ 7 days. Almost all these rapid rotators display unusually stable light curves over the full 10 month dataset (blue points in Figs. 5.5 and 5.6, label 'PB' in Table 5.2). Most of them also have relatively large amplitudes ($>$ 2%, compared to 0.2–2% for the bulk of the sample). Two possible explanations spring to mind for these rapid rotators: they may be significantly younger than the rest of the sample (for example members of the young disk population, see Section 5.6), or they may be close binary systems, which have become fast rotators due to spin-orbit interactions. The 5 known short-period eclipsing binaries in the sample (shown as stars in Figure 5.5) all display synchronised rotation signals. However, I cannot conclusively distinguish between the two possibilities without spectroscopy. Another possible route of further investigation would be to compare the properties of the fast rotators in this sample to those of the fast rotators in young clusters.

Finally, all the planet-host candidates (Batalha et al. 2012, red points in Figs. 5.5 and 5.6) have $P_{\rm rot} > 10$ days and amplitude $<$ 2%. If the period and amplitude distributions of the planet-host candidates were the same as those of the rest of the sample, one would have expected $\sim$ 3 of them to have $P_{\rm rot} < 10$ days and/or amplitude $>$ 2% (of all the M-dwarfs with rotation periods, 10% fall outside these limits, and there are 42 candidate planet host stars in the sample). The number of objects concerned is too small to draw firm conclusions, but it does suggest that either the search for transits in *Kepler* light curves may be less complete around active, rapidly rotating stars, or that there are fewer transiting planets around these stars (which could be the case of the rapid rotators are close binaries). Either way, this should be taken into account when inferring the incidence of planets around different types of stars. This is studied in further detail in Chapter 7.

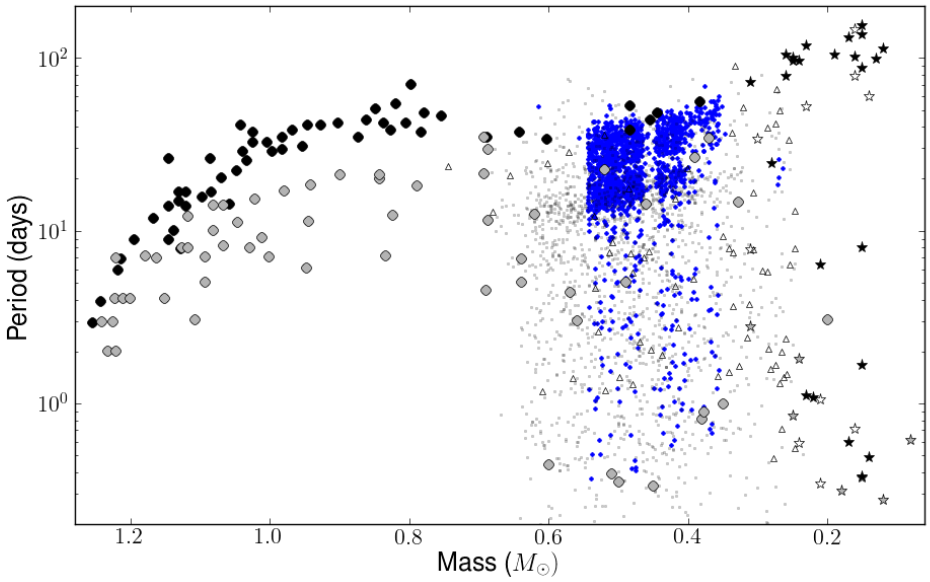

Figure 5.10: Period versus mass for low-mass field stars. Based on data from Baliunas et al. (1996) and Kiraga & Stepien (2007) (circles) and *MEarth* data from Irwin et al. (2011) (stars), with grey and black symbols representing objects with young and old disk kinematics, respectively. Additional M-dwarf periods from the WFCAM Transit Survey (Goulding et al., 2012), for which no kinematic classification is available, are shown as white triangles, and the sample of Hartman et al. (2011) are shown as small grey dots. The new results from the present study are shown as blue points.

5.6 PERIOD-MASS RELATION

To compare the *Kepler* M-dwarf results with other rotational studies of low-mass stars, I estimated their masses from the KIC effective temperatures. Much of the previous work on rotational evolution (Kawaler, 1989; Barnes, 2003, 2007; Meibom et al., 2009; Collier Cameron et al., 2009) uses a directly observed colour index, such as $B-V$ or $J-K$, instead of model-dependent mass estimates, to confront models with observations. I opted for masses for three reasons. First, angular momentum evolution models generally depend on fundamental parameters rather than colours. Second, neither of the colour indices most frequently used, $B-V$ and $J-K$, is well matched to these targets, which are very faint in B, and have almost constant $J-K$ over the mass range of interest (in contrast to G and K-dwarfs, where $J-K$ is a steeper function of mass). Third, stellar parameters based on multi-colour photometry should

be less sensitive to reddening than any single colour index.

The masses used in this study were obtained by interpolating the 630 Myr isochrone of Baraffe et al. (1998). I checked that the results are essentially unchanged if the age is increased by a factor of up to 10. While useful for a general comparison, it should be stressed that these masses are very uncertain and should not be used for individual objects. The 200 K formal uncertainties on $T_{\rm eff}$ alone translate into mass uncertainties ranging from 0.15 $M_{\odot}$ (for $M \sim 0.3\,M_{\odot}$) to 0.1 $M_{\odot}$ (for $M \sim 0.6\,M_{\odot}$). Additionally, any systematic errors in $T_{\rm eff}$, discussed in Section 5.2 are automatically propagated through to the masses.

Figure 5.10 shows the period-mass relation for field stars with masses below 1.3 $M_{\odot}$, based on a compilation of literature sources (Baliunas et al., 1996; Kiraga & Stepien, 2007; Hartman et al., 2011; Irwin et al., 2011; Goulding et al., 2012). The *Kepler* M-dwarf sample (small blue dots on Figure 5.10) ties in smoothly with the existing data at both higher and lower masses (circles, stars and triangles). The shorter period sequence is also in good agreement with the data from Hartman et al. (2011), which may have shown the bimodality if the survey had allowed for longer period measurements. The tight sequence followed by old F, G and K stars is clearly seen on Figure 5.10, as are the younger stars gradually evolving towards that sequence. This evolution is relatively well reproduced by simple models of angular momentum loss via a magnetised wind (Kawaler, 1988, 1989), and forms the basis for rotational age-dating, or gyrochronology (Barnes, 2003, 2007).

The picture is more complex for M-dwarfs, which evolve more slowly, and hence have not converged onto a common sequence, even after several Gyr. However, the new, much larger sample of early M-dwarfs strengthens two interesting features, which were only hinted at by previously published results.

First, one might have expected to see some kind of transition around 0.35 $M_{\odot}$ (Chabrier & Baraffe, 1997; Scholz et al., 2011), which corresponds to the transition between fully convective stars and stars with a radiative core, because the so-called interface dynamo, which powers the magnetic field of Sun-like stars, cannot operate in the absence of a radiative core. There is no evidence for a sudden change in the period-mass relation at this point, although the periods of the upper envelope of the

distribution do increase considerably below $0.35\,M_{\odot}$. Other studies have shown that below this threshold the rapid rotators persist even among kinematically old stars, as the *MEarth* sample of Irwin et al. (2011) illustrates, but above it, all the fast rotators are kinematically young, as Kiraga & Stepien (2007) pointed out. The *Kepler* sample as it stands cannot shed much additional light on this point because it currently lacks kinematic and distance information. Compared to similar studies mentioned in this chapter, the *Kepler* sample shows a lack of fast rotators.

Second, a transition that was not expected, but is clearly seen, occurs in the upper envelope of the period-mass relation around 0.6–$0.55\,M_{\odot}$, where the slope of the relation suddenly changes sign. To my knowledge, this intriguing feature has not yet been discussed in the literature, and it will need to be accounted for in future modelling work. In particular, this transition will need to be incorporated into empirical gyrochronological relations, if the latter are to be applied to M-dwarfs.

The moderately slow rotation of field M-dwarfs has been problematic for some time. The latest generation of theoretical models of angular momentum loss via a magnetised wind (Reiners & Mohanty, 2012) naturally explains the wide range of rotation rates observed for M-dwarfs in early main-sequence clusters. However, the behaviour of field M-dwarfs can only be reproduced by assuming that the critical rotation rate (above which the magnetic field saturates) is not universal, but depends on mass and perhaps even on age. With the new results, upper-envelope of the period-mass relation is now much better defined, which should prove valuable in testing and calibrating future refinements of the models.

5.7 DISCUSSION

Rotation periods were measured for 1570 main-sequence M-dwarf stars with masses between 0.3 and $0.55\,M_{\odot}$, measured using a new method based on the autocorrelation function, applied to the first 10 months of data from *Kepler*. The fraction of objects in which periods are detected, 63.2%, is remarkably high. For comparison, Irwin et al. (2011) detected periods in 15% of their 273-object sample. The contrast can be

explained by the unprecedented precision and continuous sampling of the *Kepler* data, and by the performance of the ACF method.

The bulk of these period detections fall into two distinct groups, with periods in the range 10–25 and 25–80 days with peaks at $\sim$ 19 and $\sim$ 33 days respectively. I suggest that these correspond to two stellar populations with different median ages. The comparison of available non-zero proper motions for the two main samples further supports the hypothesis that they have different ages. The more slowly rotating group has typically higher proper motion values, as would be expected for an older population. Within each group there is also a weak correlation between period and proper motion. This study shows the potential use of variability statistics as a probe of star formation history, although further data on distance and kinematics are required to draw sound conclusions about the nature of these regions.

In each of the two groups of stars, which form the two peaks of the period distribution, the rotation period tends to increase with decreasing mass. By combining these results and existing rotation data from the literature over the mass range 0.1–1.3 $M_{\odot}$, I have shown that this relation extends over the whole M-dwarf regime (0.1–0.55 $M_{\odot}$) but is in stark contrast to the behaviour of K-stars, whose period decreases with decreasing mass. To my knowledge, this dichotomy between K and M-dwarfs has not been noticed before, and is not predicted by current models.

A small fraction ($\sim$ 8%) of the M-dwarfs display short ($<$ 7 days), stable periods, and marginally enhanced variability compared to the rest of the sample. The most likely explanations are that these are short-period binaries or very young stars.

When combining the *Kepler* sample with data from the literature, there appears to be no evidence for a discontinuity in the upper envelope of the period-mass distribution around 0.35 $M_{\odot}$, even though stars below this mass are expected to remain fully convective. Such a break might have been expected if the development of a radiative core played a key role in driving a large-scale magnetic field, as proposed for example by Barnes (2003). I note that the *Kepler* and *MEarth* field M-dwarfs lie in the 'rotational gap' defined by Barnes (2010). This suggests that the Rossby number, which plays a key role in controlling the rotational evolution of G and K stars, may be less important for low mass stars.

There are however still changes at the fully convective boundary, with a steep increase in the periods of the slowest rotators below 0.35 $M_{\odot}$, and the persistence of fast rotators even for a kinematically old sample. The *Kepler* sample also displays a relative lack of fast rotators compared to other studies.

Analysis of additional quarters of *Kepler* data may reveal long periods for some of the objects without detections in the present sample. I am also involved in work to improve the systematics correction with an alternative to the PDC-MAP (see Chapter 2), and to enhance the performance of the ACF method. Residual systematics and quarter joins are the limiting factors in the current analysis because they introduce steep variations in the ACF. By optimising the ACF smoothing parameters after initial period detection, and working on methods to separate the periodic signal from the long term ACF trends, I hope reach a higher level of precision and clarity.

This study is the first large-scale investigation of rotation in field M-dwarfs, and provides the first useful constraints on the period-mass relation for these objects after they have settled onto a common rotational sequence. This work also demonstrates the power of *Kepler* data and of the ACF method for rotation studies, and paves the way for a truly systematic survey of rotation rates on the main sequence, from mid-F to mid-M spectral types, which I examine in the next chapter.

6

ROTATION PERIODS OF FIELD DWARFS

In this chapter I present a preliminary study of the rotational properties of the F, G and K dwarf stars observed by *Kepler*. In Section 6.2 I describe the data selection process and period determination, followed by a discussion of the amplitude-period distribution in Section 6.3. Finally, in Section 6.4 I present the mass-period plot, and compare to the rotational isochrones of Barnes (2007).

6.1 INTRODUCTION

The natural progression of this study, following the detailed examination of the M-dwarf sample, is to apply the ACF method to the F, G and K dwarf stars observed by *Kepler*. However, without automation of the ACF method, it is unfeasible to visually confirm the results for $\sim 130,000$ stars.

In this short Chapter, I have therefore opted to run the ACF code on a large subset of each spectral type, and perform the visual verification stage until a sample of 1000 periodic objects is obtained in each set. This allows the continuation of trends observed in the M-dwarf sample to be studied, along with a more detailed comparison to previous rotation period observations. This work is also provides the foundation for the comparison of planet host stars to single stars (Chapter 7).

Table 6.1: Number of objects included at each stage of the study, from initial selection. Visual examination was performed on the ACF output in a random order until 1000 periodic stars were found. The periodic fraction is calculated from the visual examination total and the number of periodic stars. The final column shows the periodic fraction for stars brighter than *Kepler* magnitude14.

SpT	Initial	Visual	Periodic	Periodic %	Periodic % kepmag < 14
M	2423	2423	1521	62.7	81.7
K	6000	2553	1000	39.7	45.2
G	7000	5731	1000	17.5	26.0
F	8000	3975	1000	25.3	31.2

6.2 DATA SELECTION AND PERIOD DETERMINATION

From the KIC have selected a random sample of F, G and K dwarfs which are assumed to be single stars without planets (i.e. they are not listed as planet candidates or known eclipsing binaries). The number of stars selected of each spectral type, listed in Table 6.1, was chosen based on roughly estimated periodic fraction from visual inspection of a small subset, to ensure that 1000 period detections could be made from the sample. The dwarf/giant cut and spectral classes were defined using the method of Ciardi et al. (2011), as described in Chapter 3.

In each case Q1–4 data were obtained and prepared following the method described in Chapter 5. The ACF method (Chapter 4) was run to determine periods, which were verified visually, in a random order, until a set of 1000 periods of each spectral type were found.

The periodic fractions are shown in Table 6.1. The M-dwarf results are included, but without the planet host stars and known binaries. From this point on, a random selection of 1000 periodic M-dwarfs were used in this study. The periodic fraction are discussed in Section 6.5. A sample portion of the period detection and non detection results are shown in Tables 6.3 to 6.8 in the appendix of this chapter. These tables are available in their entirety on the data disc provided to the examiners.

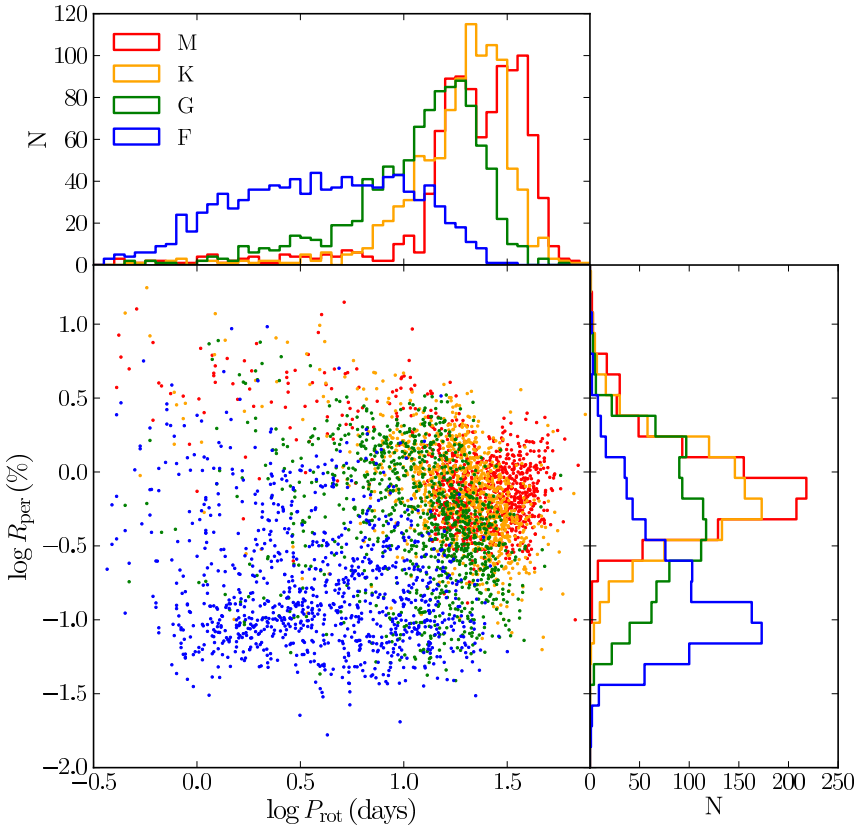

Figure 6.1: Period versus amplitude for the *Kepler* field M, K, G and F dwarfs. The histograms of each parameter are shown along the corresponding axis, with matching colours.

6.3 PERIOD AND AMPLITUDE

I first examined the period distribution of the sample as a function of amplitude, as shown in Figure 6.1. The bimodality, which is clearly present in the M-dwarf sample is not obviously a feature of the other spectral classes. This may simply be due to the mass dependance of the periods, masking the effect in the wide mass range of the K-dwarfs, and the steep gradient of the trend in the G and F dwarfs. There are subtle signs on the bimodality for the K-dwarfs in mass-period distribution shown in Figure 6.2, but a closer examination is required to confirm this.

If the bimodality does indeed only occur in the late type stars, the question of its origin should be posed. It could potentially be an observational bias, since the earlier

spectral types are brighter and hence seen at larger distances. If the bimodality results from two different epochs of star formation, then the signature could be blurred when observing over a wider range of distances. The *Kepler* field lies just above the Galactic plane so an increase in distance corresponds to an increase in Galactic scale height and hence probes a typically older population. The functional form of this increase in age with Galactic scale height is not well known, and considerable further work would be required to access if the angular momentum evolution of low mass stars could produce such a bimodal distribution over the observed Galactic age distribution. In this work I simply suggest this as a possible explanation requiring further analysis. Another possibility is that the bimodality is specific to late type stars, although the physical origin of this is not known.

The trend of increasing period towards later spectral types, accompanied by increased amplitude of variability has long been known, and is shown clearly in Figure 6.1. In each spectral class there exist objects with short periods and high amplitudes. As discussed in Chapter 5, a likely explanation of this phenomena is that these objects are either young stars or non-eclipsing binaries, when spin-orbit interactions have artificially increased rotation periods.

The amplitudes for the periodic fraction of F-dwarfs are considerably lower than for other spectral types. It is possible that the brighter stars show intrinsic variability to lower amplitudes, since it is not lost in the photon noise. Alternatively, this variability may result from a phenomenon other than rotation, such as pulsation (see Section 6.4.4).

6.4 GYROCHRONOLOGY AND THE PERIOD-MASS RELATION

6.4.1 Mass Estimation

Mass, M, is calculated from the KIC T_{eff} using the stellar evolution models of Baraffe et al. (1998), using isochrone no. 1 for $M < 0.7 M_{\odot}$ and isochrone no. 3 for higher masses, and assuming an age of 1 Gyr. I checked that the change in results is negligible

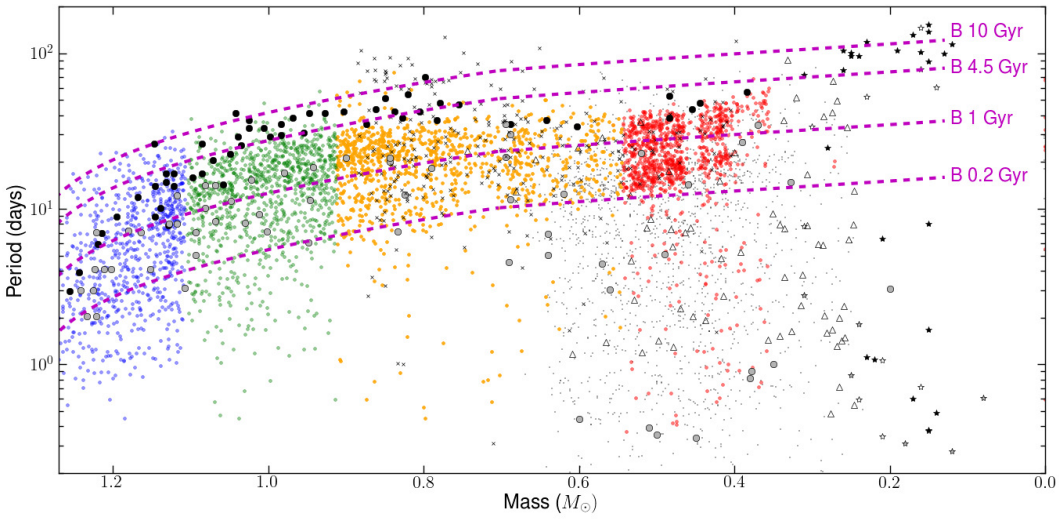

Figure 6.2: Period versus mass for the F, G, K & M field stars (blue, green yellow and red respectively). With data from Baliunas et al. (1996) and Kiraga & Stepien (2007) (circles) and *MEarth* data from Irwin et al. (2011) (stars), with grey and black symbols representing objects with young and old disk kinematics, respectively. Additional M-dwarf periods from the WFCAM Transit Survey (Goulding et al., 2012), for which no kinematic classification is available, are shown as white triangles. Periods from Hartman et al. (2011) are shown as small grey dots, and periods from Harrison et al. (2012) are shown as crosses. The magenta lines show the 0.2, 1, 4.5 and 10 Gyr rotational isochrones Barnes (2007),

if the age is varied by a factor of up to 10. The systematic errors in $T_{\rm eff}$, discussed in Section 5.2, are automatically propagated through to the masses.

6.4.2 *The Mass-Period Relation*

The mass-period plot for the F–M stars is show in Figure 6.2, along with data from the literature, as described in the caption. I have also plotted the 0.2, 1, 4.5 and 10 Gyr rotational isochrones, calculated using Equation 1.7.

As mentioned in Section 6.3, there is tentative evidence that the bimodality exists beyond the M-dwarfs, although interestingly, the upper sequence appears more populated for the K and G dwarfs. Figure 6.3 shows some evidence in the K-dwarfs of a secondary peak to the distribution around 12 days, which is the projected location of the lower sequence of the M-dwarf set. This second peak is no longer visible for

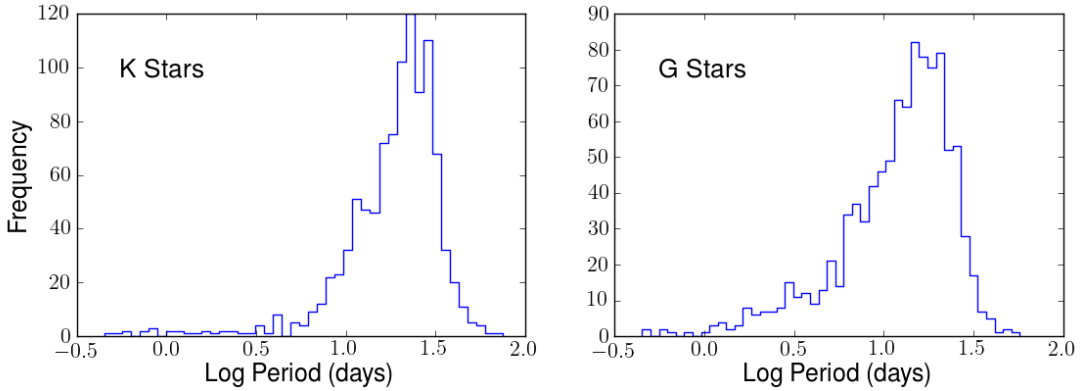

Figure 6.3: Histograms of log period for the G and K stars, showing that the bimodality does not persist to the G stars, but the K stars show some evidence, around 12 days, of a secondary peak to the distribution.

the G-dwarfs.

This potentially favours the suggestion that these brighter stars are seen to greater distances and hence the contribution of a younger sample would decrease against an older background population. This is of course highly speculative at this stage, and I plan to follow up these ideas in future work.

6.4.3 *Gyrochronological Age Approximations*

As discussed in Section 1.3.4, gyrochronology relations are currently not well calibrated. It is however interesting to test where the isochrones of Barnes (2007) lie on the period-mass plot for the M to F spectral types.

The gyrochronology relation of Barnes (2007) (Equation 1.7) uses rotation period measurements and $B-V$ to determine ages. Since the KIC does not include $B-V$, I have used the $T_{\rm eff}$ conversion of Sekiguchi & Fukugita (2000):

$$\begin{aligned}(B-V)_0 = {} & t_0 + t_1 \log\, T_{\rm eff} + t_2(\log\, T_{\rm eff})^2 + t_3(\log\, T_{\rm eff})^3 \\ & + f_1[{\rm Fe/H}] + f_2[{\rm Fe/H}]^2 + d_1[{\rm Fe/H}] \log\, T_{\rm eff} \\ & + g_1 \log\, g + e_1 \log\, g \log\, T_{\rm eff}\,, \end{aligned} \tag{6.1}$$

where the fit parameters are shown in Table 6.2.

Table 6.2: Fit parameters for Equation 6.1

Parameter	Value
t_0	-813.3175 ± 42.5
t_1	684.4585 ± 34.3
t_2	-189.923 ± 9.23
t_3	17.40875 ± 0.827
f_1	1.2136 ± 0.038
f_2	0.0209 ± 0.0006
d_1	-0.294 ± 0.010
g_1	-1.166 ± 0.028
e_1	0.3125 ± 0.0076

The rotation isochrones closely follow the *Kepler* star period distribution, although they fail to match the decrease in rotation periods around $0.6M_{\odot}$ (described in Section 5.6), and the sharp increase in rotation periods below $0.4M_{\odot}$. The $0.6M_{\odot}$ dip is predominantly formed by a tail of slow-rotating G and K stars, which do not appear to have any unusual properties indicating a different rotational behaviour should be expected.

6.4.4 Caveats

It should be noted that the period determination stage for the F, G and K stars was not performed to the same level of precision as for the M-dwarf sample, due to the preliminary nature of this section. Therefore, the steps performed to remove potential giant contamination, re-run visual examination for verification, and include a detailed breakdown of excluded objects, such as those with multiple periods, were omitted.

Another potential issue is that of contamination by pulsating stars. For the M-dwarf sample this was not a concern since no pulsations are expected in the period or amplitude range of interest. However, stars towards the earlier spectral classes lie increasingly close to the intersection of the instability strip with the main sequence in the Hertzsprung-Russell (HR) diagram. This is the region in the HR diagram which is populated by pulsating variables. Therefore, contamination by pulsating stars cannot be ruled out and no effective method to distinguish between pulsation and fast, stable star spot modulation has been developed.

6.5 DISCUSSION

This chapter presents a first look at the distribution of rotation periods for a large number of field stars observed by *Kepler*. I have examined a subset of the *Kepler* F, G and K dwarf light curves and determined periods using the ACF method for 1000 of each spectral type, along with corresponding periodic fractions.

Following a simplified version of the approach used for the M-dwarf sample (Chapter 5), I explored the relationship between period and amplitude. This study builds on the work described in Chapter 3, where I characterised variability as a function of stellar parameters. The trends concluded from that work are displayed clearly in the period-amplitude plot in Figure 6.1; later spectral types have longer rotation periods and higher amplitude variability. The periodic fraction of objects also decreases towards earlier spectral types, although it appears to increase in the F-dwarfs. The periodic fractions for a magnitude limited subset (*Kepler* magnitude < 14) show a slightly higher number of periodic stars, but maintain this trend. This could potentially arise from contamination by pulsating stars on the instability strip, as discussed in Section 6.4.4, or be due to the different internal structure of the F-dwarfs. An alternative explanation may be that the brighter stars show intrinsic variability to lower amplitudes, since it is not lost in the photon noise.

The mass-period distribution (Figure 6.2) displays what is to my knowledge, the largest sample of rotation period measurements for field stars. These results show a densely populated upper band, with a less populated lower band below, split approximately where the dearth of M-dwarf periods is located. The bimodality is not as clear in the K–F stars, although there are tentative evidence, shown in Figure 6.3 that it extends slightly into the K stars. There also appears to be a distinct lower envelope to the rotation periods of the late-F and early-G stars, which decreases with increasing mass, although more data points are required to confirm this. For all spectral types there exist anomalously short period rotators, which are likely to be very young stars, or binary systems where spin-orbit interactions have increased or prevented evolution of rotation periods.

The period measurements of Hartman et al. (2011) are more sparse and lacking the bulk of the long rotation periods seen in the ACF results. This is likely a result of ground-based measurement, which makes longer periods harder to detect. I chose not to include the results of Affer et al. (2012) in Figure 6.2, since these contain a large number of long period (> 60 day) detections, which I consider unlikely to be real given the short duration and high noise levels of the CoRoT data. The Harrison et al. (2012) results are derived from a year of *Kepler* data, and show a higher proportion of longer periods than the ACF results. Although this dataset has a shorter baseline than that used for the ACF study, they examined a smaller number of targets and performed their own data reduction, which is likely to perform better than the PDC-MAP. For this reason I am inclined to believe the longer periods and hope that with further improvement of the *Kepler* systematics removal and quarter joining, the ACF will be able to probe these regions successfully.

Rotational isochrones calculated using the formalism of Barnes (2007) fit the overall trend of the data well, following the approximate slope of the sequence, at least above $0.4 M_{\odot}$. However, they fail to account for the decrease in the upper envelope of rotation periods around $0.6 M_{\odot}$ and the sudden increase in periods in the M-dwarfs. Considering the age estimates presented by the isochrones, it appears that the bulk of stars are in the oldest regions (> 1 Gyr), which would be expected for the sample of stars observed by *Kepler*. The increased fraction of older stars towards earlier spectral types may result from these brighter objects being observed to greater distances, further out of the Galactic plane. In this case however, it is more likely that the rotational isochrones are incorrect, due to poor calibration for older main-sequence stars. High mass stars should not be seen above the 10 Gyr isochrone, as the measurements of Harrison et al. (2012) are. One would also not expect to see such a large number of very young stars (< 2 Myr) in the *Kepler* field, and therefore many these fast rotators are likely to require an alternative explanation than youth for their rapid rotation.

The spread of F-dwarfs periods, in comparison to the tighter sequence observed for the later spectral types, is not well modelled by the rotational isochrones. This may indicate that contamination from variability sources other than rotational modulation

(e.g. pulsations) is a problem in this sample, or that the physical mechanisms inside these stars deviate more considerably from the predicted relations.

One may not expect to see so many young stars ($<$ 1 Gyr), especially in the M-dwarf sample, which calls into question the calibration of the gyrochronology relation. Alternatively, the source of this anomaly may be the temperature to colour conversion stage. When additional age and rotation information becomes available for the asteroseismic targets and *Kepler* clusters, I hope to re-examine this issue.

With the automation of the ACF method, the full sample of dwarf stars observed by *Kepler* can be studied in this way. I will also seek to address the pulsation contamination based on the stability of the light curve modulation over many *Kepler* quarters.

APPENDIX 6.A PERIOD DETECTION AND NON-DETECTION TABLES

Table 6.3: F-dwarfs with detected rotation periods. This table is available in its entirety on the data disc provided to the examiners. A portion is shown here for guidance regarding its form and content. $T_{\rm eff}$ and $\log g$ are from the KIC and M was derived from $T_{\rm eff}$ using the 600 Myr isochrone of Siess et al. (2000). The average amplitude of variability per period bin of the light curve, $R_{\rm per}$, is included.

KID	$T_{\rm eff}$ (K)	$\log g$ (dex)	M ($M_\odot$)	$P_{\rm rot}$ (days)	$\Delta P_{\rm rot}$ (days)	Amp. (mmag)
1028120	6613	4.32	1.4460	5.946	0.076	1.26
1162715	6167	3.94	1.2659	6.559	0.010	12.37
1292666	6284	3.81	1.3125	18.584	0.269	7.05
1430741	6947	3.97	1.5857	1.124	0.010	2.24
1434286	6267	4.04	1.3057	11.719	0.175	5.48

Table 6.4: G-dwarfs with detected rotation periods. This table is available in its entirety on the data disc provided to the examiners. See Table 6.3 caption for descriptions of column headings.

KID	$T_{\rm eff}$ (K)	$\log g$ (dex)	M ($M_\odot$)	$P_{\rm rot}$ (days)	$\Delta P_{\rm rot}$ (days)	Amp. (mmag)
1725133	5768	4.49	1.1113	1.798	0.000	2.38
1868918	5954	4.20	1.1826	12.986	1.216	0.42
1869933	5303	4.44	0.9392	15.530	0.011	21.69
1873987	5731	4.34	1.0973	14.477	0.025	4.17
2016452	5389	4.21	0.9704	24.112	1.145	6.07

Table 6.5: K-dwarfs with detected rotation periods. This table is available in its entirety on the data disc provided to the examiners. See Table 6.3 caption for descriptions of column headings.

KID	$T_{\rm eff}$ (K)	$\log g$ (dex)	M ($M_\odot$)	$P_{\rm rot}$ (days)	$\Delta P_{\rm rot}$ (days)	Amp. (mmag)
1297058	5135	4.52	0.8792	27.381	0.712	10.98
1571364	4404	4.60	0.6322	20.587	0.210	5.70
1722506	4270	4.43	0.5895	10.758	0.010	30.48
1866182	4711	4.07	0.7331	10.625	0.167	14.35
1867210	4918	4.53	0.8035	16.705	0.143	4.96

Table 6.6: F-dwarfs with no rotation period detection. This table is available in its entirety on the data disc provided to the examiners. A portion is shown here for guidance regarding its form and content.

KID	$T_{\rm eff}$ (K)	$\log g$ (dex)	M ($M_\odot$)	Amp. (mmag)
891901	6051	4.41	1.2204	2.96
893234	6666	4.18	1.4679	1.80
1026475	6682	4.17	1.4745	0.59
1026647	6518	4.17	1.4070	1.37
1026669	6244	4.21	1.2965	1.07

Table 6.7: G-dwarfs with no rotation period detection. This table is available in its entirety on the data disc provided to the examiners. A portion is shown here for guidance regarding its form and content.

KID	$T_{\rm eff}$ (K)	$\log g$ (dex)	M ($M_\odot$)	Amp. (mmag)
892977	5861	4.33	1.1468	2.10
893004	5921	4.50	1.1699	1.85
1026321	5580	4.69	1.0407	1.91
1027030	6000	4.01	1.2005	0.72
1430333	5732	4.05	1.0977	1.31

Table 6.8: K-dwarfs with no rotation period detection. This table is available in its entirety on the data disc provided to the examiners. A portion is shown here for guidance regarding its form and content.

KID	$T_{\rm eff}$ (K)	$\log g$ (dex)	M ($M_\odot$)	Amp. (mmag)
893033	4707	4.64	0.7318	10.55
1025578	4928	4.53	0.8069	2.36
1026287	4911	4.55	0.8011	3.27
1162467	5224	4.52	0.9108	9.82
1162725	5217	4.48	0.9083	2.95

7

ROTATION PERIODS OF PLANET CANDIDATE HOST STARS

In this chapter I examine the rotational properties of the *Kepler* exoplanet host stars. In Section 7.2 I describe the data selection and period determination process for the host stars and a comparison sample of F–M field dwarfs. This comparison is shown in Section 7.3. To test whether any of these planets are likely to show signs of tidal interaction, I first estimate the masses of the exoplanet candidates (Section 7.4), which together with stellar mass and orbital period, provide an estimates of the spin-up timescale for the system (Section 7.5). By comparing the predicted spin-up timescale to the level of spin-obit synchronisation, I selected 16 stars which are likely to show evidence for tidal interaction, and examine the in further detail in Section 7.6. Finally, I use the gyrochronology to estimate the relative ages of the exoplanet candidate systems.

7.1 INTRODUCTION

The formation, environment and habitability of an exoplanet are dependent on the properties of the host star. Since the primary aim of the *Kepler* mission is to detect and characterise exoplanet systems, information about the activity and rotation of the host stars is extremely valuable.

Spin-orbit interactions between host stars and their planets are ill-understood, due to the lack of host star rotation periods observations. Using the ACF method described in Chapter 4, I have measured rotation periods for 548 (33%) of the 1646 *Kepler* exoplanet candidate host stars, which are classified as dwarfs using the cut of Ciardi et al. (2011).

By obtaining measurements of the host star rotation period and the planet orbital period, I am able to search for evidence of tidal interaction between the star and planet. The effect of tidal interaction on the orbits and rotation of binary stars has been widely observed, although the theoretical details are complex and still uncertain (see e.g. Mazeh, 2008, for recent review). The same underlying physics can be applied to planetary systems, although the mass ratio is much greater and the tidal effects on the star are much smaller. Tidal effects lead to the transfer of angular momentum from the planet to the star (provided $P_{\rm orb} < P_{\rm rot}$), which over time can align the rotation axes of the two components, synchronise orbital and rotation periods, and circularise the orbit.

Although spin-orbit interactions have been studied for a number of individual systems, it is much harder to do a statistical study (see Pont, 2009, and references therein) due to the complexity of the tidal and magnetic formulae, poorly constrained parameters and the scarcity of planet host stars with measured rotation periods. To determine whether a system is likely to display signs of tidal interaction, I estimate the spin-up timescale and use this to select the systems for which this value is anomalously short, i.e. falling well below the bulk of estimations, and examine their light curves and rotation periods for signs of synchronisation.

For the study of tidal interaction timescales, I chose to adopt the methodology of Terquem et al. (1998), which is similar to Pont (2009), using a simple scaling relation between the strength of tidal effects, the mass ratio and the orbital separation to estimate the timescale for spin-up:

$$t^{*}_{\rm sp}({\rm Gyr}) = 1.725 \times 10^{-6} \left(\frac{M_{\rm p} + M_{*}}{M_{\rm p}} \right)^{2} \left(\frac{P_{\rm orb}}{1 \text{ day}} \right)^{3}, \tag{7.1}$$

where $M_{\rm p}$ and M_{*} are the planet and stellar masses respectively, and $P_{\rm orb}$ is the orbital period of the planet. This relation is calibrated for a solar type star, and therefore the

timescale will be shorter for stars with a proportionally deeper convective envelope since tidal interactions would have more effect. Since the dependence scales strongly with planet mass and orbital period, this relationship provides a good estimate of whether tidal interaction is likely to have led to any observable synchronisation during the lifetime of the star. For this study, it is also assumed that the innermost planet observed is actually the closest planet existing in the system.

Pont (2009) finds evidence for circularisation of orbits, excess stellar spin due to tidal effects and concludes that tidal spin-up of the host star will produce a strong detection bias against massive, close-in planets, in radial velocity surveys, since fast rotation broadened spectral lines. Pont et al. (2011b) extends this work to provide constraints on planet formation models, based on the observations of tidal circularisation of massive, close-in planets, and Husnoo et al. (2012) performs a similar study using new RV data.

Activity levels also affect transit detection, as discussed in Section 1.3.1. The contribution of activity and rotation to detection statistics has not to my knowledge been fully explored, and I will therefore present a preliminary comparison of the host star rotation and activity levels to that of the non-host stars (Section 7.3).

7.2 DATA SELECTION AND PERIOD DETERMINATION

This section of work requires both the light curves of the *Kepler* planet candidates, and a comparison sample of dwarf stars of spectral type F – M, which are assumed to be single stars without planets. The latter are described in Chapter 6 and were selected to provide a sample with approximately the same magnitude and temperature distribution and the host star sample.

This selection was performed by comparing the distribution of field stars in magnitude-temperature space, to that of the host stars. As demonstrated in Figure 7.1, I selected a random sample of field stars from each grid cell, to match the number present in the host star distribution (with available $T_{\rm eff}$ and *Kepler* magnitude values). Using this method, I retained a sample of 1631 field stars for comparison.

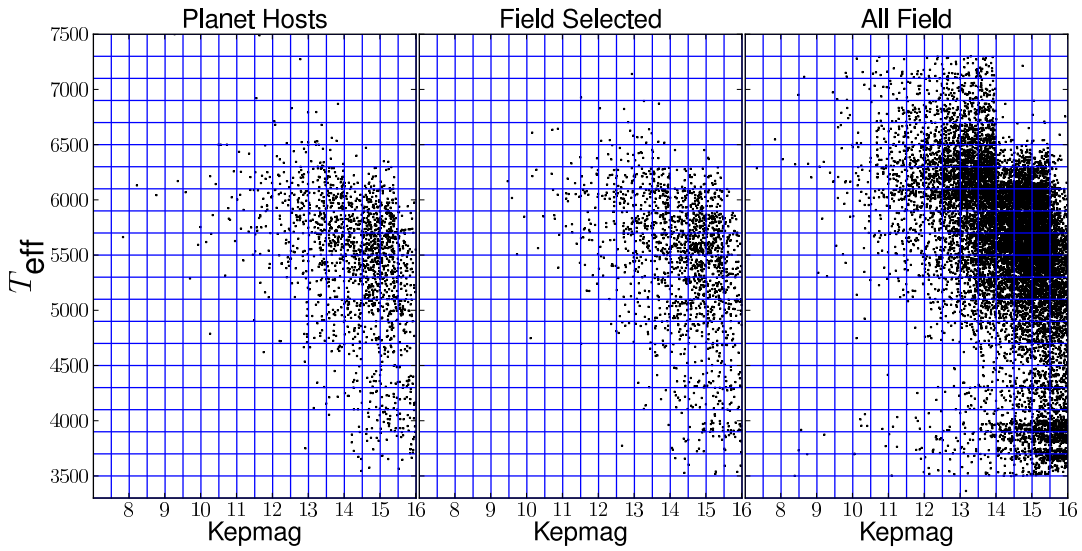

Figure 7.1: The comparison set of field stars were selected by assigning a grid (blue lines) in magnitude-temperature space, and selecting a random sample of field stars to match the number of host stars in each cell, where available.

A list of the *Kepler* objects of interest (KOIs) was obtained from the *Kepler* mission archive[1]. On Feb 13$^{\rm th}$ 2013, this list contained 2740 planets or planet candidates, orbiting 2036 stars. Of these, 1716 had Q1–4 long cadence light curves available. A further 22 stars were removed from the sample due to binarity. Since we are only considering main sequence stars in this study, a further 48 stars were removed because they did not meet the criteria for dwarf stars, using the cut of Ciardi et al. (2011).

The ACF method was applied to the Q1–4 light curves and the results verified through visual inspection, as described for the *Kepler* M-dwarf sample in Chapter 5. The results for the periodic host star sample are shown in Table 7.4 and the non-periodic stars in Table 7.5 in Appendix 7.B.

7.3 COMPARISON OF HOST AND NON-HOST ROTATION

Rotation periods were detected for 548 (33%) of the exoplanet host stars (dwarf stars only), which is similar to the average fraction found for the large sample of field stars (30.7%), described in Chapter 6. See Table 6.1 for breakdown of periodic fractions.

[1]http://archive.stsci.edu/kepler/koi

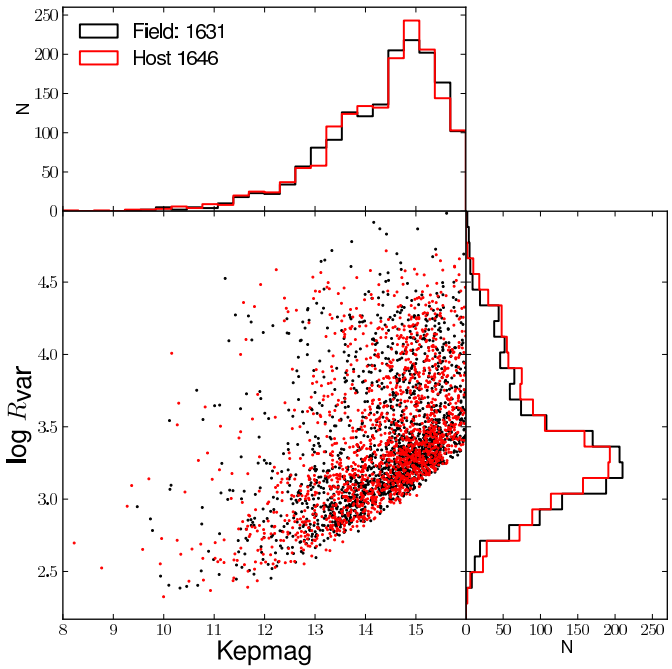

Figure 7.2: Amplitude of variability, R_{var} vs. *Kepler* magnitude for the field stars (black) and the planet hosts (red).

To compare the amplitude of variability of the planet hosts to the single stars, I used an amplitude statistic R_{var}. This is the range between the 5^{th}–95^{th} percentile of differential flux values in ppm (see Section 3.2 further details). Figure 7.2 shows R_{var} as a function of *Kepler* magnitude for the single field stars (black) and the planet hosts (red). It shows no evidence that the host stars are less active than the non hosts, if anything it appears to be the opposite. The distribution of magnitudes of the comparison sample was selected to match that of the host stars, so any resulting bias can be ruled out.

To compare the planet host stars to the single stars more accurately, I selected a subset of the field stars is to match the magnitude and temperature distribution of the host stars. Figure 7.3 shows distribution of planet host that is on the whole, very similar to that of the single stars. One minor difference is that it appears uncommon for the planet hosts to have very short periods, below the lower envelope of the bulk of single stars. This is quantified in the difference in the short period region between the log period distributions shown in Figure 7.3, which show an Anderson-Darling

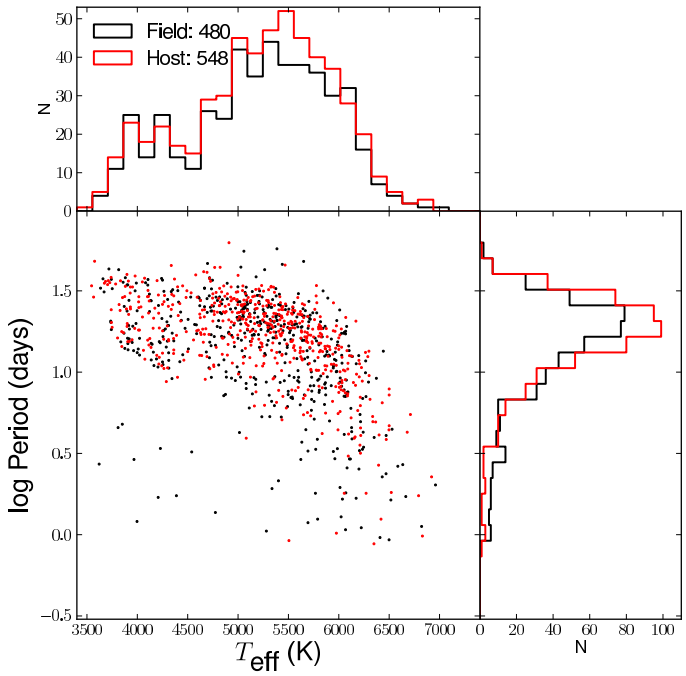

Figure 7.3: Period vs. effective temperature for the single M, K, G and F dwarfs (black) and planet host stars (red).

(Scholz & Stephens, 1987) test result of having a 2% chance of being drawn from the same distribution.

In the short period regime, stars are potentially close non-transiting binaries where the interaction between stars has left them rotating faster. They may also be young stars which have not spun down. This group tends to have high amplitude variability which could potentially lead to fewer planet transit detections, although this does not appear to be the case based on Figure 7.2. Alternatively it could show that planets are less common around this type of star or system.

A more thorough study focussing on individual spectral types is required to determine the differences, if present, between the host planet stars and the field stars.

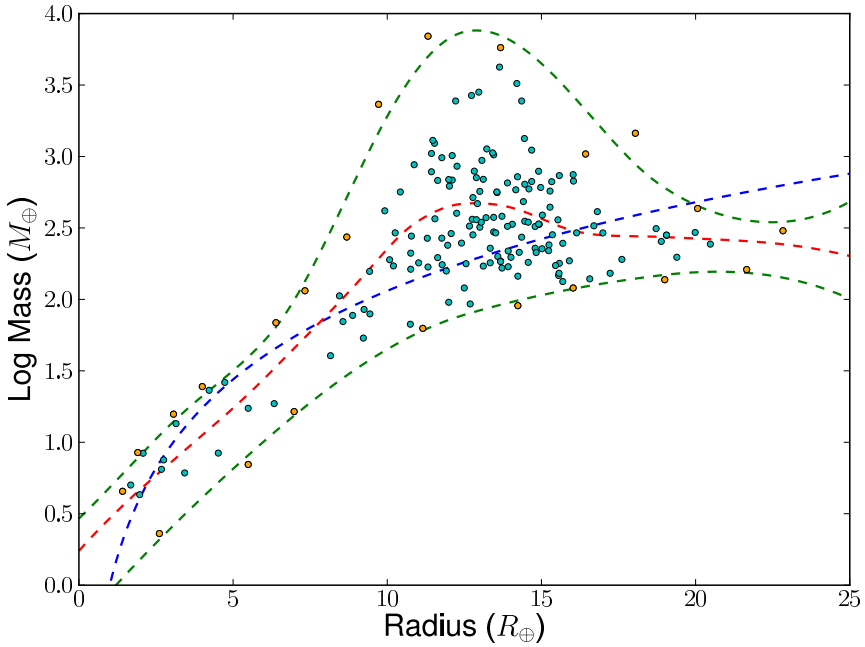

Figure 7.4: Regression fits to the exoplanet data (cyan/orange points), including the best fit (red line), fits (green lines) to the envelope edges (orange points), and the power-law fit of Wolfgang & Laughlin (2011) (blue line).

7.4 PLANET CANDIDATE MASS ESTIMATION

Since the tidal interaction timescale is a function of the mass ratio, both the mass of the planet and the star had to be estimated. The stellar mass is estimated using the models of Siess et al. (2000) (see Section 6.4.1 for details).

Estimating the planetary mass is more challenging because no mass-radius relationship exists that is consistent from rocky planets to gas giants. Mordasini et al. (2012) describe theoretical formation mechanisms for exoplanets and present an accompanying mass-radius relation, however, this relation involves many assumptions about the composition of the planet.

Empirical mass-radius relations have been derived for subsets of the planet candidates, such as that presented by Lissauer et al. (2011) for Earth and Saturn:

$$\frac{M}{M_\oplus} = \left(\frac{R}{R_\oplus}\right)^{2.06}, \tag{7.2}$$

which was tested and applied by Wolfgang & Laughlin (2011). However, this is only valid for relatively small radius and high density planets and does not characterise

Table 7.1: Values of the empirical planet mass-radius relation fit described in Chapter 5.

Radius $R_\oplus$	log Mass (best) $M_\oplus$	log Mass (min) $M_\oplus$	log Mass (max) $M_\oplus$
1.000	0.468	-0.047	0.684
2.000	0.672	0.184	0.912
3.000	0.863	0.404	1.131
4.000	1.048	0.614	1.323
5.000	1.239	0.814	1.501
6.000	1.443	1.006	1.708
7.000	1.662	1.190	2.002
8.000	1.881	1.360	2.406
9.000	2.110	1.515	2.856
10.000	2.354	1.651	3.281
11.000	2.548	1.766	3.617
12.000	2.651	1.857	3.826
13.000	2.672	1.923	3.881
14.000	2.644	1.980	3.812
15.000	2.564	2.030	3.651
16.000	2.482	2.073	3.430
17.000	2.449	2.111	3.185
18.000	2.443	2.144	2.955
19.000	2.436	2.172	2.780
20.000	2.426	2.190	2.661
21.000	2.416	2.193	2.583
22.000	2.403	2.179	2.544
23.000	2.382	2.145	2.545
24.000	2.349	2.088	2.590

the uncertainties in the mass calculation.

To obtain mass estimates and errors for the *Kepler* exoplanet candidates, I made a simple empirical fit to the data from 175 exoplanets with reliable mass and radius measurements[2]. These data are listed in Table 7.3 and shown as points in Figure 7.4. The orange points indicate the planets that were manually selected to to fit a representative envelope for the distribution.

The best fit lines, shown in Figure 7.4, are obtained using the 'locfit' package in R (Loader, 1999). This performs local regression using 3^{rd} order polynomials for the best fit and lower bound, and 4^{th} for the upper bound. This fit allows for an acceptable preliminary estimate of the masses with uncertainties. The best fit values and uncertainties are listed in Table 7.1. The distributions of planet mass and radius as

[2]Downloaded from `http://exoplanets.org` on 18^{th} Feb 2013

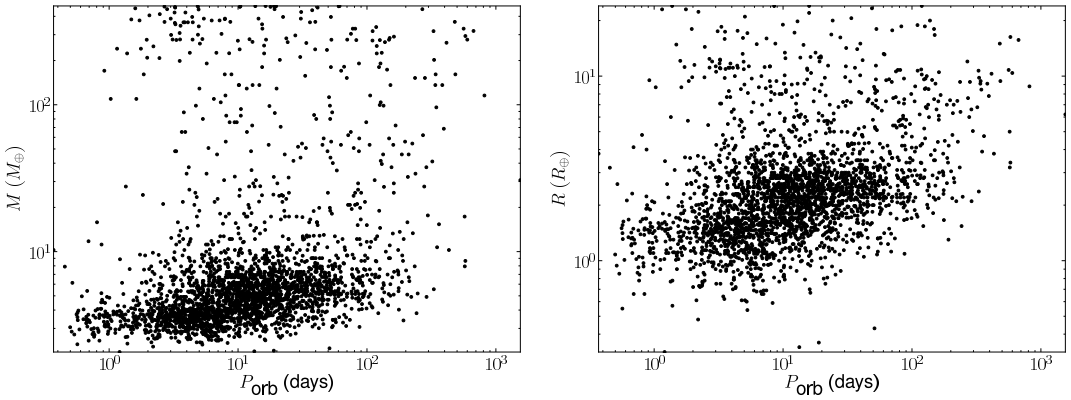

Figure 7.5: Exoplanet candidate mass and radius as a function orbital period, using mass estimates from the fit shown in Figure 7.4.

a function of orbital period are shown in Figure 7.5, and demonstrate that the sample mainly comprises of super-Earths, with a smaller subset of giant planets.

7.5 SYNCHRONISATION TESTS

Using Equation 7.1 (Terquem et al., 1998), and the empirical planetary mass-radius relation derived in Section 7.4, estimates for the spin-up timescale were determined. These results are shown in Figure 7.6, with the accompanying error bars associated with uncertainty on the planet mass.

In this case I have used only the innermost planet, since the $t^*_{\rm sp}$ scales so strongly with decreasing orbital separation. It should be noted that if the second planet is much more massive, then it may also affect the star, but for simplification in this case I will examine only the effect of the inner-most planet.

The spin-up relationship is only applicable for systems where the orbital period of the planet is shorter than the rotation period of the star, since angular momentum transfer always acts to slow down the faster body and speed up the slower one. Therefore only points on or to the right of the 1:1 ratio line are considered candidates for host star spin-up. There is a dearth of points to the left of the 1:1 line at short $t^*_{\rm sp}$ ($< 10^3$ Gyrs).

The overall trend of decreasing spin-up time with increasing $P_{\rm rot}/P_{\rm orb}$ is a result

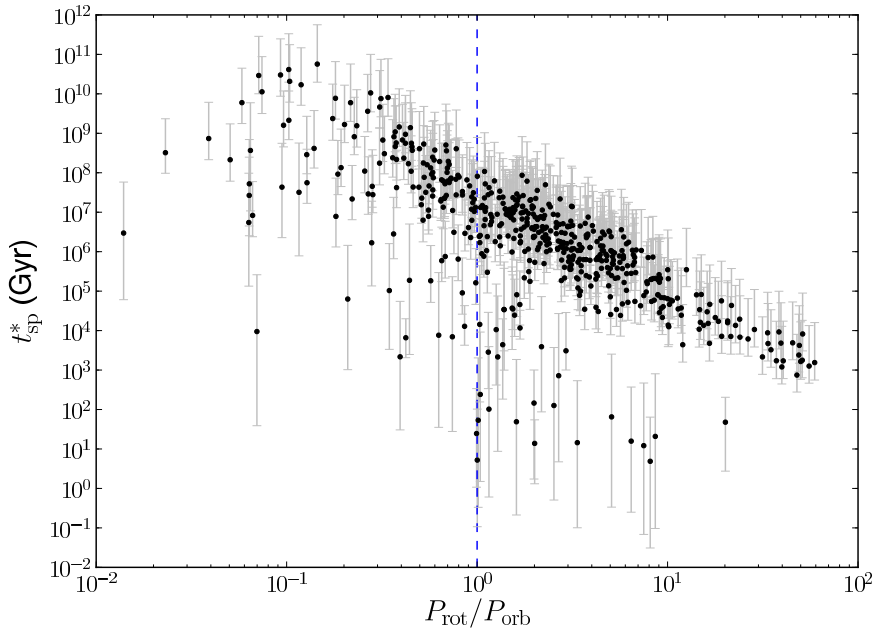

Figure 7.6: Estimated timescale for spin-up of star by planet, compared to the ratio of stellar rotation period and planet orbital period. The blue dashed line marks a 1:1 ratio. The error bars are associated only with uncertainty in planet mass, defined by the empirical fits to the envelope on the radius-mass plot in Figure 7.4.

of the strong dependence of $t^*_{\rm sp}$ on the orbital period, which is demonstrated in bottom left panel of Figure 7.7. This also shows that the stars with short spin-up timescales (bottom centre and bottom right of panel) have the shortest orbital periods, as expected.

The distribution of stellar rotation periods is shown in the top right panel of Figure 7.7, and demonstrates that $P_{\rm rot}$ has little effect on $t^*_{\rm sp}$, although the anomalously short spin-up timescales are associated with the fastest rotators.

The $T_{\rm eff}$ distribution (top left panel of Figure 7.7) follow a similar trend to the rotation periods, as would be expected, showing that the hottest stars tend to be the fastest rotators. The planetary mass distribution (bottom right panel of Figure 7.7) shows that the majority of systems lying below the bulk of points, have considerably higher mass.

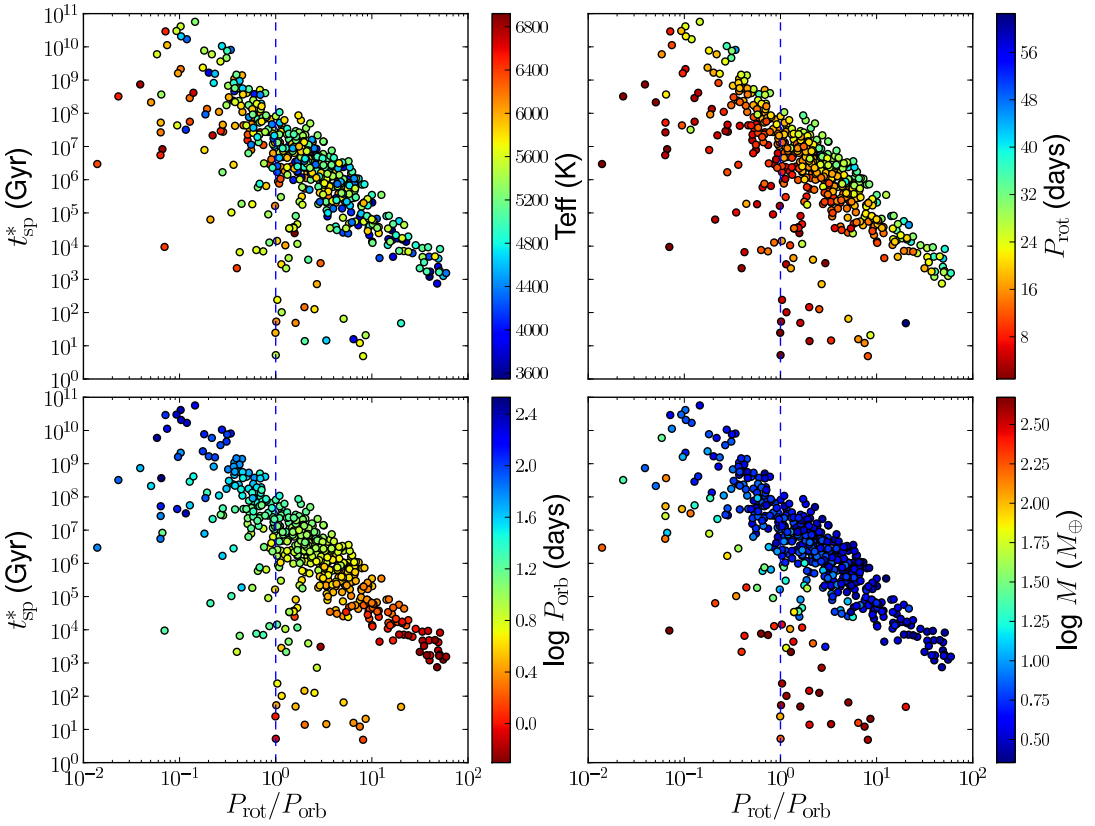

Figure 7.7: Same data points as Figure 7.6, but with a colour gradient showing the distribution of T_{eff} (top left), P_{rot} (top right), P_{orb} (bottom left), and M_{pl} (bottom right).

7.6 EVIDENCE OF TIDAL INTERACTION

The *Kepler* exoplanet candidate systems which are most likely to show evidence of tidal interaction are located in the lower-right section of Figure 7.6, where $P_{rot} < P_{orb}$. The 16 systems matching this criteria are listed in Table 7.2 and labelled in Figure 7.9. In the bottom panel of Figure 7.9 the positions of these systems are marked on the period - temperature plot, showing that these stars are often found to rotate faster than the bulk of single stars. A direct comparison of rotation periods shows that 12 of the 16 systems lie below the median period for field stars within 100 K of the planet host star T_{eff}. The two systems (KIC 5115978 and KIC 5475431) with the shortest rotation periods are also the systems with the closest spin-orbit synchronisation.

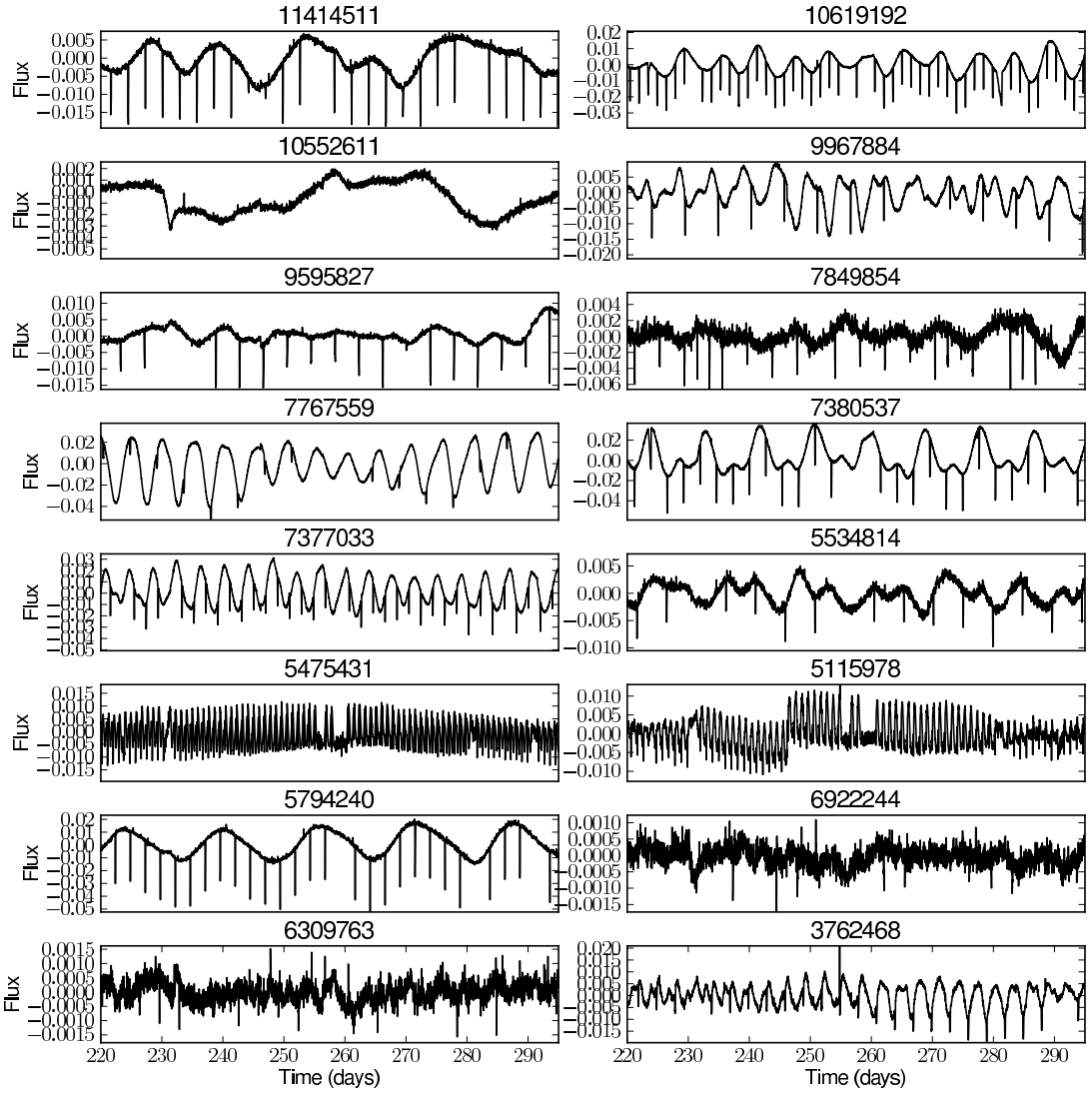

Figure 7.8: Light curves of the 16 stars selected as likely spin-up candidates. These plots show only a segment of the light curve, in order to make the fast rotators visible, however this magnification results in several of the panels (10552611, 6309763 & 6922244) looking non-periodic. The full light curve displays periodicity at a borderline detection level.

Table 7.2: Details of the 16 stars selected as likely candidates for tidal interaction.

KID	$T_{\rm eff}$	$\log g$	M	$P_{\rm rot}$	$\Delta P_{\rm rot}$	Amp.	$t^*_{\rm sp}$	$M_{\rm pl}$	$P_{\rm orb}$
	(K)	(dex)	($M_\odot$)	(days)	(days)	(mmag)	(Gyr)	($M_\oplus$)	(days)
3762468	6094	4.58	1.2372	3.045	0.005	12.14	53.449	2.586	3.004
5115978	5976	4.43	1.1912	1.022	0.000	8.83	24.607	2.040	1.028
5475431	5505	4.97	1.0129	0.920	0.000	13.24	5.225	2.232	0.918
5534814	5794	4.47	1.1212	12.301	0.061	5.43	126.211	2.670	4.859
5794240	3948	4.54	0.4903	15.816	0.021	25.99	15.795	2.318	2.455
6309763	6122	4.55	1.2482	5.231	1.280	0.95	48.870	2.661	3.252
6922244	6213	4.17	1.2842	7.009	2.456	0.79	145.642	2.488	3.522
7377033	5081	4.57	0.8602	3.923	0.005	33.26	13.840	2.443	1.957
7380537	4674	4.82	0.7207	9.032	0.005	38.14	14.446	2.564	2.689
7767559	5436	4.37	0.9875	5.088	0.010	28.71	102.204	2.598	4.409
7849854	5734	4.46	1.0985	15.387	0.214	3.25	12.131	2.609	2.052
9595827	5504	4.72	1.0125	19.841	0.030	9.45	64.994	2.628	3.905
9967884	5689	4.54	1.0815	5.640	0.051	11.30	241.084	2.586	5.428
10552611	4910	4.18	0.8007	62.527	14.501	4.26	47.749	2.444	3.108
10619192	5634	4.49	1.0608	12.046	0.025	22.10	4.874	2.581	1.486
11414511	5431	4.44	0.9857	24.280	0.159	9.88	20.817	2.651	2.817

7.7 GYROCHRONOLOGY

The ages of exoplanetary systems are of great interest to the astrophysical community, since they provide constraints on planet formation and evolution mechanisms, and shed light on the question of habitability. Asteroseismology can provide ages for some exoplanet host stars where pulsations are detected, although they require very high precision observations, detailed individual modelling. Gyrochronology, when reliably calibrated, has the potential to allow age determination for a large number of stars, including those with exoplanets, provided the rotational evolution does not differ from stars without planets.

From the distribution of $t^*_{\rm sp}$ shown in Figure 7.6, one may conclude that in the majority of cases, the spin-up effect on the star is negligible, implying gyrochronology relations will hold for these stars.

Using the gyrochronology relation of Barnes (2007), (Equation 1.7 discussed in Section 1.3.4), and $B-V$ estimates from $T_{\rm eff}$ (described in Section 6.4.3), the approximate ages were calculated for 539 of the *Kepler* exoplanet candidate stars with measured rotation periods. Since the field star results presented in Chapter 6 indicate

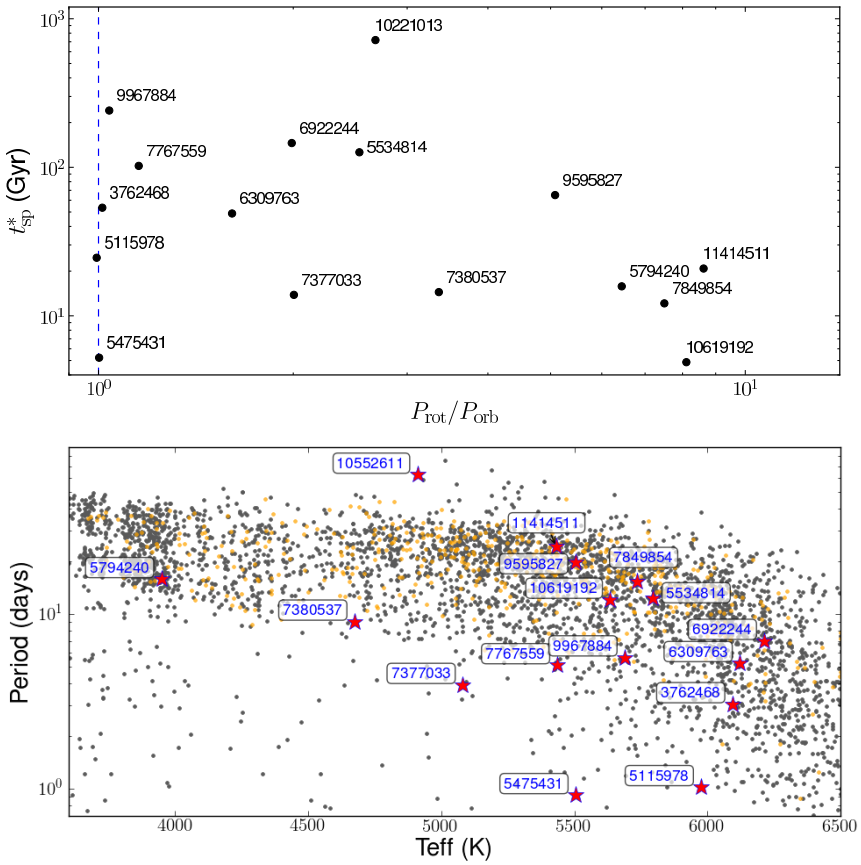

Figure 7.9: The top panel shows the bottom right corner of Figure 7.6, with the 16 selected objects labelled. The bottom panel shows the location of these stars on the temperature-period plot, where the F–M dwarfs with measured periods (described in Chapter 6) are plotted in grey, with the planet host stars plotted in yellow.

that the gyrochronology relation may not be adequately calibrated for this work, I have chosen not to list individual host star ages. Figure 7.10 shows the period-radius distribution for with a colour scale for age. These ages should only be used as indicative relative to the other host stars and not taken at their quoted value.

The host stars with $T_{\rm eff} < 5700$K, and $P_{\rm rot} < 6$ days are excluded because the relationship is not a good match to the data in this region (based on Figure 6.2 from Chapter 6). Figure 7.10 shows no indication that period-radius distribution of exoplanet candidates is dependent on age.

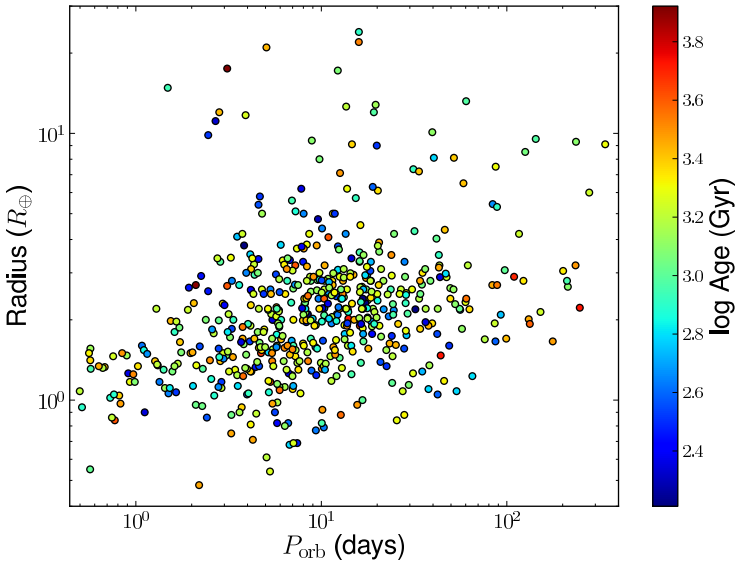

Figure 7.10: Orbital period versus planet radius, coloured by age estimates from the gyrochronology relation of Barnes (2007). Note: known calibration issues (see Section 6.5) mean that some slow rotators appear older than 100 Gyr which is obviously incorrect. This plot does however provide comparative ages for the stars.

7.8 DISCUSSION

7.8.1 Comparison of Host and Field Stars

I have applied the ACF method to the *Kepler* exoplanet candidate host stars and detected periods for 548 (33%) of them. This is consistent with the periodic fraction for the F-M dwarf stars, which I determine to 29% for a sample with matching temperature and magnitude distribution. One may hypothesise that rotation period detection may be fractionally higher for stars with transiting planets, since the tendency for spin-orbital inclination alignment would lead to a favourable observation bias towards spot-crossing events as opposed to pole-on stellar views.

The amplitude of photometric variability does not appear to differ between the host stars and the single field star sample. It was initially suspected that the planet hosts would display lower levels of variability due to the detection bias that transits are harder to detect in more variable light curves, but this does not appear to be the case

when the selected comparison sample has a matching temperature and magnitude distribution.

However, this work does not conclusively prove there is no difference between the exoplanet candidate host stars and the field stars, since it requires the assumption that the field stars do not have exoplanets. While this cannot be proved using existing data, the assumption that the field stars do not have large, close-in undetected planets is reasonable, since these are likely to have been detected by transits, RV or astrometry in the majority or orbital inclinations. These are the only planets likely to alter the rotation of their host star, due to their size and proximity.

The distribution of periodicities detected in the host and comparison samples are also very similar, with the exception of the short period region (< 6 days), where very few planets are detected. In Chapter 5, I suggest these fast rotators are likely binaries, with increased rotation periods from to spin-orbit interactions, or potentially they are very young stars. Although circumbinary planets have been detected in the *Kepler* sample (e.g. Doyle et al., 2011; Welsh et al., 2012), they are uncommon and since the potential binaries in this sample are non-transiting, there is little chance of observing a transit in the system, assuming all bodies are coplanar. If these objects are very young single stars, then variability on these timescales and amplitudes could hinder transit detection. It is therefore not possible to conclude from this data whether the dearth of host stars at with very short periods is a genuine result or an observational bias.

Overall this comparison demonstrates that in the vast majority of cases the presence of a planet does not act to alter the period or level of photometric variability of the host star. This is an encouraging result because it allows the application of standard stellar physics and models to the host stars.

7.8.2 Tidal Interactions

Motivated by the work of Pont (2009) and Pont et al. (2011b), I performed some preliminary calculations to estimate the spin-up timescale, $t^*_{\rm sp}$, of the star as a result of angular momentum transfer from the close-in planets. These results depend on

approximate planetary mass estimates from an empirical fit to the mass-radius data of confirmed planets, and are derived using a formulation for t^*_{sp} that is calibrated for solar type stars. These caveats should be borne in mind when interpreting the results, but it provides an order of magnitude approximation for the scale of the tidal effects.

Stars with short t^*_{sp} are found to have rotation-orbit period ratios of 1 or greater. The spin-up relationship is only applicable for systems where the orbital period of the planet is shorter than the rotation period of the star, since angular momentum transfer always acts to slow down the faster body and speed up the slower one. The dearth of stars at short t^*_{sp} and period ratios less than 1 is therefore expected.

The 16 stars with $t^*_{sp} < 600$ Gyrs were selected as likely candidates for tidal interaction, since they fall considerably below the bulk of points. They cover a range of T_{eff} but typically have a short period giant planet, as would be expected for interaction to occur.

The bottom panel of Figure 7.9 shows the location of these stars on the T_{eff}-period distribution, and demonstrates that they are typically below or in the lower portion of the main periodic sequence, indicating that they are rotating more rapidly than would be expected for their spectral type. This is consistent with the interpretation that these are likely to have been spun up by their planets.

7.8.3 Gyrochronology

The bulk of host stars in the bottom panel of Figure 7.9 show the same T_{eff} - period distribution as the F–M field stars without planets. This implies that in the vast majority of cases gyrochronological relations will apply to host stars without alteration, with the exception of the $\sim 2\%$ which appear to have been spun up by their planet.

As a preliminary exercise, I plotted the distribution of planet radius and orbital period as a function of age (Figure 7.10) for the planets with measured rotation periods. Although no relationship with age is apparent, it demonstrates the type of tests which can easily be performed using gyrochronologically determined ages for large number of exoplanets.

7.8.4 *Future Work*

This chapter demonstrates the potential for spin-orbit interaction studies of exoplanets. The limiting factor is currently the level of uncertainty on the system parameters, which will improve in time as more followup observations are performed. Despite these uncertainties, a more thorough study of this nature could be performed using the currently available data, although the time constraints did not permit its inclusion in this thesis. The simple $t^*_{\rm sp}$ of Terquem et al. (1998) may not be well suited for application to all spectral types and planet masses, however the more complex formulations involve many assumptions about ill-understood parameters. For example, a key parameter in tidal effect equations is the tidal factor Q (Goldreich & Soter, 1966), which describes the effectiveness of tidally induced energy in the interior of a body. This is poorly constrained and expected to vary by orders of magnitude for different planet compositions.

Using more precise data on exoplanet systems, it would be interesting to populate a plot similar to that of Figure 3 in (Pont et al., 2011b), which displays $M_{\rm pl}/M_*$ as a function of $a/R_{\rm pl}$, where a is the orbital separation. By dividing the data sample into stars which show potential spin-up effects and those which don't, one could shed light on the boundaries of tidal interaction and planet destruction, as the planet orbital radius reduces to conserve angular momentum.

APPENDIX 7.A CONFIRMED PLANETS

Table 7.3: The confirmed planets with accurately measured masses and radii, from `http://exoplanets.org`, that were used to obtain the empirical mass-radius relation in Figure 7.4.

Name	Period (days)	Radius $R_\oplus$	Mass $M_\oplus$
WASP-44 b	2.4238039	1.140	0.890
WASP-32 b	2.718659	1.180	3.553
WASP-26 b	2.7566	1.320	1.017
WASP-1 b	2.5199449	1.516	0.918
WASP-45 b	3.1260876	1.160	1.007
HAT-P-19 b	4.008778	1.132	0.292
HAT-P-16 b	2.77596	1.289	4.202
HAT-P-28 b	3.257215	1.212	0.628
WASP-18 b	0.9414529	1.267	10.201
HAT-P-32 b	2.150009	2.037	0.951
HAT-P-29 b	5.723186	1.107	0.779
HAT-P-38 b	4.640382	0.825	0.268
WASP-77 A b	1.3600309	1.210	1.760
HD 17156 b	21.21663	1.020	3.303
WASP-50 b	1.9550959	1.153	1.472
WASP-11 b	3.722465	0.910	0.540
HAT-P-25 b	3.652836	1.190	0.567
WASP-22 b	3.53269	1.120	0.559
WASP-78 b	2.17517656	1.700	0.884
XO-3 b	3.1915426	1.217	13.285
HAT-P-15 b	10.863502	1.072	1.952
WASP-79 b	3.6623817	1.700	0.888
WASP-61 b	3.8559	1.240	2.055
WASP-35 b	3.161575	1.320	0.717
WASP-62 b	4.411953	1.390	0.562
WASP-49 b	2.7817387	1.115	0.378
KELT-2 A b	4.1137912	1.286	1.521
WASP-63 b	4.37808	1.430	0.378
CoRoT-19 b	3.89713	1.290	1.108
WASP-12 b	1.09142245	1.790	1.361
CoRoT-18 b	1.9000693	1.310	3.487

Continued on next page

Table 7.3 – continued from previous page

Name	Period	Radius	Mass
CoRoT-12 b	2.828042	1.440	0.919
CoRoT-7 b	0.853585	0.150	0.016
WASP-23 b	2.9444256	0.962	0.872
CoRoT-5 b	4.0378962	1.388	0.463
CoRoT-1 b	1.5089557	1.490	1.030
CoRoT-4 b	9.20205	1.190	0.717
CoRoT-13 b	4.03519	0.885	1.311
CoRoT-14 b	1.51214	1.090	7.695
HAT-P-24 b	3.35524	1.242	0.686
HAT-P-9 b	3.92289	1.400	0.777
HAT-P-20 b	2.875317	0.867	7.290
HAT-P-33 b	3.474474	1.827	0.766
HAT-P-39 b	3.54387	1.571	0.599
XO-5 b	4.1877537	1.030	1.153
XO-2 b	2.615838	0.983	0.567
HAT-P-35 b	3.646706	1.332	1.054
HAT-P-30 b	2.810595	1.340	0.711
HAT-P-13 b	2.91625	1.281	0.857
WASP-36 b	1.5373653	1.269	2.269
55 Cnc e	0.736546	0.186	0.026
WASP-13 b	4.353011	1.389	0.479
HD 80606 b	111.4367	1.029	3.891
WASP-19 b	0.78884	1.386	1.133
WASP-43 b	0.813475	0.930	1.776
HAT-P-22 b	3.21222	1.080	2.151
WASP-66 b	4.086052	1.390	2.313
OGLE-TR-211 b	3.67724	1.360	0.757
OGLE-TR-132 b	1.689868	1.200	1.180
OGLE-TR-113 b	1.4324752	1.093	1.260
OGLE-TR-111 b	4.0161	1.051	0.550
WASP-34 b	4.3176782	1.220	0.583
OGLE2-TR-L9 b	2.4855335	1.610	4.574
OGLE-TR-182 b	3.9791	1.130	1.025
WASP-31 b	3.405909	1.537	0.479
HAT-P-21 b	4.124481	1.024	4.078
HATS-1 b	3.446459	1.302	1.865
GJ 436 b	2.64385	0.377	0.073
HAT-P-36 b	1.327347	1.264	1.839
WASP-41 b	3.052401	1.200	0.932
WASP-42 b	4.9816877	1.063	0.497

Continued on next page

Table 7.3 – continued from previous page

Name	Period	Radius	Mass
WASP-25 b	3.764825	1.220	0.578
WASP-55 b	4.465633	1.300	0.571
HAT-P-3 b	2.899703	0.899	0.596
Qatar-2 b	1.3371182	1.144	2.484
WASP-15 b	3.7521	1.379	0.543
HAT-P-12 b	3.2130598	0.959	0.211
HAT-P-26 b	4.234516	0.565	0.059
WASP-16 b	3.1186009	1.008	0.842
WASP-39 b	4.055259	1.270	0.284
WASP-14 b	2.243752	1.281	7.692
WASP-37 b	3.577469	1.160	1.794
HAT-P-27 b	3.0395589	1.038	0.617
WASP-24 b	2.3412127	1.300	1.091
HAT-P-4 b	3.056536	1.274	0.672
WASP-17 b	3.735433	1.932	0.509
XO-1 b	3.941534	1.206	0.918
WASP-38 b	6.871814	1.094	2.690
HAT-P-2 b	5.6334729	1.157	8.871
HD 149026 b	2.8758911	0.654	0.361
HAT-P-18 b	5.508023	0.995	0.197
GJ 1214 b	1.58040482	0.239	0.020
HAT-P-14 b	4.627669	1.150	2.236
OGLE-TR-10 b	3.101278	1.250	0.620
TrES-3 b	1.30618581	1.336	1.910
TrES-4 b	3.553945	1.783	0.925
OGLE-TR-56 b	1.2119189	1.363	1.390
HAT-P-31 b	5.005425	1.070	2.169
HAT-P-5 b	2.788491	1.254	1.056
WASP-58 b	5.01718	1.370	0.891
CoRoT-16 b	5.35227	1.170	0.538
WASP-3 b	1.846834	1.290	2.012
CoRoT-17 b	3.7681	1.020	2.460
CoRoT-23 b	3.6313	1.050	3.085
CoRoT-11 b	2.99433	1.430	2.348
CoRoT-9 b	95.2738	1.050	0.844
CoRoT-6 b	8.886593	1.166	2.954
Kepler-8 b	3.52254	1.419	0.586
HAT-P-37 b	2.79736	1.178	1.174
Kepler-43 b	3.024095	1.200	3.231
Kepler-9 b	19.243158	0.842	0.249

Continued on next page

Table 7.3 – continued from previous page

Name	Period	Radius	Mass
Kepler-9 c	38.90861	0.823	0.169
Kepler-4 b	3.21346	0.357	0.077
Kepler-10 b	0.837495	0.126	0.014
TrES-1 b	3.030065	1.067	0.752
Kepler-12 b	4.4379637	1.695	0.432
TrES-2 b	2.47063	1.224	1.201
Kepler-20 b	3.6961219	0.170	0.027
Kepler-20 c	10.854092	0.274	0.050
Kepler-20 d	77.61184	0.245	0.024
Kepler-14 b	6.790123	1.136	8.406
Kepler-7 b	4.885525	1.478	0.438
Kepler-16 b	228.776	0.754	0.333
CoRoT-10 b	13.2406	0.970	2.755
WASP-48 b	2.143634	1.670	0.984
CoRoT-8 b	6.21229	0.570	0.216
CoRoT-2 b	1.7429935	1.466	3.275
CoRoT-3 b	4.2568	1.010	21.855
HAT-P-7 b	2.2047304	1.363	1.802
Kepler-45 b	2.455239	0.960	0.511
Kepler-35 b	131.458	0.728	0.127
Kepler-41 b	1.855558	0.841	0.494
WASP-67 b	4.61442	1.400	0.419
Kepler-15 b	4.942782	0.960	0.662
Kepler-34 b	288.822	0.764	0.220
Kepler-40 b	6.87349	1.170	2.177
Kepler-6 b	3.234723	1.323	0.669
Kepler-39 b	21.0874	1.220	18.181
Kepler-11 b	10.30375	0.176	0.014
Kepler-11 c	13.02502	0.281	0.042
Kepler-11 d	22.68719	0.306	0.019
Kepler-11 e	31.9959	0.403	0.026
Kepler-11 f	46.68876	0.233	0.007
HAT-P-41 b	2.694047	1.685	0.800
HAT-P-11 b	4.8878162	0.422	0.083
Kepler-18 c	7.64159	0.490	0.054
Kepler-18 b	3.504725	0.490	0.022
Kepler-18 d	14.85888	0.623	0.052
Kepler-17 b	1.4857108	1.330	2.479
Kepler-5 b	3.54846	1.431	2.117
Kepler-44 b	3.24674	1.240	1.021

Continued on next page

Table 7.3 – continued from previous page

Name	Period	Radius	Mass
HD 189733 b	2.21857567	1.138	1.144
HAT-P-34 b	5.452654	1.197	3.334
Qatar-1 b	1.420033	1.164	1.090
TrES-5 b	1.4822446	1.209	1.778
HAT-P-23 b	1.212884	1.368	2.096
WASP-2 b	2.1522254	1.077	0.909
WASP-7 b	4.954658	0.915	0.919
WASP-46 b	1.43037	1.310	2.102
HAT-P-17 b	10.338523	1.010	0.530
HD 209458 b	3.52474859	1.359	0.690
WASP-47 b	4.1591399	1.150	1.136
HAT-P-40 b	4.457243	1.730	0.620
HAT-P-8 b	3.076337	1.500	1.293
HAT-P-1 b	4.4652934	1.242	0.532
WASP-21 b	4.322482	1.070	0.300
WASP-6 b	3.361006	1.224	0.521
WASP-52 b	1.7497798	1.270	0.457
WASP-10 b	3.0927616	1.080	3.191
WASP-59 b	7.919585	0.775	0.859
WASP-4 b	1.3382299	1.341	1.223
HAT-P-6 b	3.852985	1.330	1.060
WASP-29 b	3.922727	0.792	0.243
WASP-5 b	1.6284229	1.140	1.624
WASP-8 b	8.158715	1.038	2.138

APPENDIX 7.B PERIOD DETECTION AND NON-DETECTION TABLES

Table 7.4: Host stars with detected rotation periods. This table is available in its entirety on the data disc provided to the examiners. A portion is shown here for guidance regarding its form and content. $T_{\rm eff}$ and $\log g$ are from the KIC and M was derived from $T_{\rm eff}$ using the 600 Myr isochrone of Siess et al. (2000). $t^{*}_{\rm sp}$ is calculated using the relation of Terquem et al. (1998), $M_{\rm pl}$ is estimated using the mass-period relationship described in Section 7.4 and $P_{\rm pl}$ is taken from `http://archive.stsci.edu/kepler/koi` The average amplitude of variability per period bin of the light curve, $R_{\rm per}$, is included.

KID	$T_{\rm eff}$ (K)	$\log g$ (dex)	M ($M_\odot$)	$P_{\rm rot}$ (days)	$\Delta P_{\rm rot}$ (days)	Amp. (mmag)	$t^{*}_{\rm sp}$ (Gyr)	$M_{\rm pl}$ ($M_\oplus$)	$P_{\rm pl}$ (days)
757450	5101	4.48	0.8672	19.351	0.049	18.79	3897.084	2.207	8.885
1161345	5836	4.14	1.1372	8.010	0.010	11.31	312148.950	0.898	4.287
2142522	6046	4.41	1.2184	10.299	0.586	1.12	68536729.210	0.496	13.324
2161536	4013	4.48	0.5099	27.054	0.212	9.74	14700173.525	0.605	16.855
2165002	5118	4.57	0.8732	22.681	0.217	8.60	16757959.647	0.799	16.568

Table 7.5: Host star with no rotation period detection. This table is available in its entirety on the data disc provided to the examiners. A portion is shown here for guidance regarding its form and content.

KID	$T_{\rm eff}$ (K)	$\log g$ (dex)	M ($M_\odot$)	Amp. (mmag)
893033	4707	4.64	0.7318	10.55
1025578	4928	4.53	0.8069	2.36
1026287	4911	4.55	0.8011	3.27
1162467	5224	4.52	0.9108	9.82
1162725	5217	4.48	0.9083	2.95

8

CONCLUSIONS AND FUTURE WORK

8.1 ASTROPHYSICALLY ROBUST CORRECTION FOR SYSTEMATICS

The raw *Kepler* data contain large amplitude systematic effects and the *Kepler* science team's PDC correction is not well suited to variability studies. Alternative attempts to clean the data (Basri et al., 2010, 2011) also suffer significant problems. Before the release of the PDC-MAP corrected data, I started work on an alternative correction method, in collaboration with Prof. Stephen Roberts and Dr Suzanne Aigrain. This Astrophysically Robust Correction for systematics (ARC), which is presented in Chapter 2, is based on the removal of a set of basis functions that are determined to be present in small amounts across many light curves, therefore effectively removing systematics while leaving the true variability signal unchanged. It is uses a Bayesian linear basis model and includes shrinkage priors for robustness, variational inference for speed, and a de-noising step based on empirical mode decomposition to prevent the introduction of spurious noise into the corrected light curves.

This new method was applied to a synthetic dataset to demonstrate it's performance, before being applied to the *Kepler* Q1 data. A detailed comparison of the ARC and the PDC correction methods was also performed. The ARC is considerably better than the PDC, and has some advantages over the PDC-MAP, for example, the ARC has significantly lower RMS, over a range of timescales, than the PDC-MAP. This

systematics correction developed for the *Kepler* light curves has potential to be used on a wide range of time series data containing systematic trends.

8.2 VARIABILITY IN THE *kepler* Q1 DATA

Following the application of the ARC to the *Kepler* Q1 data, I was able to study the variability properties of main sequence stars as a function of fundamental stellar parameters, using one of the most valuable datasets available at the time. This work is described in Chapter 3. I showed that the fraction of stars with variability greater than that of the Sun is 60%, which is marginally consistent with previous studies, and confirmed the trend of increasing variability with decreasing effective temperatures.

Low and high variability samples were defined by a cut corresponding to twice the variability level of the active Sun, and the properties of the stars belonging to each sample were compared. I showed tentative evidence that the more active stars have lower proper motions and may be located closer to the galactic plane.

I also investigated the frequency content of the variability, and found clear evidence for periodic or quasi-periodic behaviour in 16% of stars, and showed that there exist significant differences in the nature of variability between spectral types. Of the periodic objects, most A and F stars have short periods ($<$ 2 days) and highly sinusoidal variability, suggestive of pulsations, whilst G, K and M stars tend to have longer periods ($>$ 5 days, with a trend towards longer periods at later spectral types) and show a mixture of periodic and stochastic variability, indicative of activity.

Finally, I used auto-regressive models to characterise the stochastic component of the variability, and show that its typical amplitude and time-scale both increase towards later spectral types, which was interpreted as a corresponding increase in the characteristic size and life-time of active regions.

8.3 PERIOD DETECTION USING THE AUTOCORRELATION FUNCTION

In order to exploit full potential of the *Kepler* data for stellar rotation period measurement, it was important to identify the most appropriate period detection algorithm. Standard approaches to period detection in light curves are based on Fourier decomposition or, for irregularly sampled data, least-squares fitting of sinusoidal models. However, typical stellar light curves are neither sinusoidal nor strictly periodic, probably because of the clumpy and time-evolving nature of the underlying active region distribution.

I therefore developed an algorithm for period detection based on the autocorrelation function (ACF) of the light curve. Because the ACF measures only the degree of self-similarity of the light curve at a given time lag, the period remains detectable even when the amplitude and phase of the photometric modulation evolve significantly during the time-span of the observations.

In Chapter 4, I describe the ACF method and how it can be used to robustly measure rotation periods in stellar light curves. I showed simulated examples and compared the results to those derived from the wider used sine-fitting periodogram method.

8.4 STELLAR ROTATION IN THE *kepler* M-DWARFS

In Chapter 5, I present the results of the ACF method on the first 10 months of *Kepler* data (Q1–4), for the sample of M-dwarfs observed. Of the 2483 stars examined, I detected rotation periods in 1570 (63.2%), which considerably increased the sample of known field M-dwarf periods.

The periods range from 0.37–69.7 days, with amplitudes ranging from 1.0–140.8 mmags. The rotation period distribution is clearly bimodal, with peaks at $\sim$ 19 and $\sim$ 33 days, hinting at two distinct waves of star formation, a hypothesis that is supported by the fact that slower rotators tend to have larger proper motions. The two peaks of the rotation period distribution form two distinct sequences in

period-temperature space, with the period decreasing with increasing temperature.

The period-mass distribution of our sample shows no evidence of a transition at the fully convective boundary. On the other hand, the slope of the upper envelope of the period-mass relation changes sign around $0.55\,M_{\odot}$, below which period rises with decreasing mass.

8.5 ROTATION PERIODS OF THE F,G & K FIELD STARS

In Chapter 6 I proceeded by applying the ACF method to a sample of the F, G and K stars observed by *Kepler*. Visual verification of the detected periods was performed in a random order until a sample of 1000 rotation periods for each spectral type were obtained. The periodic fraction decreases from the M stars (63%) to the K (40%) and G (18%) dwarfs, before increasing again for the F-dwarfs (25%). This could potentially arise from contamination by pulsating stars on the instability strip, or be due to the different internal structure of the F-dwarfs.

This F–M dwarf sample is, to my knowledge, the largest sample of rotation period measurements for field stars. The sample displays a clear mass-period relation, and once systematic errors in the colour to mass conversion are allowed for, matches the sequence of period measurements from the literature. The bimodality is not clearly seen in the F–K stars. For all spectral types there exist anomalously short period rotators, which are likely to be very young stars, or binary systems where spin-orbit interactions have increased rotation periods.

Rotational isochrones calculated using the formalism of Barnes (2007) fit the overall trend of the data well, although they fail to account for the decrease in rotation periods around $0.6\,M_{\odot}$ and the sudden increase in periods in the M-dwarfs. Considering the age estimates presented by the isochrones, it appears that the bulk of stars are in the oldest regions ($> 1\,$Gyr), which would be expected for the sample of stars observed by *Kepler*. However, the fraction of young stars ($< 1\,$Gyr), especially in the M-dwarf sample, seems unreasonably high, which may reflect poor calibration of the gyrochronology relation or an issue with the conversion between $B - V$ colour and

effective temperature which is required to apply the Barnes (2007) relations to our sample.

8.6 ROTATION OF THE *kepler* PLANET CANDIDATE HOST STARS

As described in Chapter 7, I applied the ACF method to the *Kepler* exoplanet candidate host stars and detected periods for 559 (33%) of them. This is consistent with the periodic fraction for the F-M dwarf stars, which I determined to 29% for a sample with matching temperature and magnitude distribution. The amplitude of photometric variability does not appear to differ between the host stars and the single field star sample.

The distribution of periodicities detected in the host and comparison samples are also very similar, with the exception of the short period region ($<$ 6 days), where very few planets are detected. This would be expected if these are, as suggested in Chapter 5, non-eclipsing binaries or very young stars. Overall this comparison demonstrates that in the vast majority of cases the presence of a planet does not act to alter the period or level of photometric variability of the host star. This is an encouraging result because it allows the application of standard stellar physics and models to the host stars.

I performed some preliminary calculations to estimate the spin-up timescale, $t^*_{\rm sp}$, of the star as a result of angular momentum transfer from the close-in planets. These results depend on approximate planetary mass estimates from an empirical fit to the mass-radius data of confirmed planets, and are derived using a formulation for $t^*_{\rm sp}$ that is calibrated for solar type stars. These caveats should be borne in mind when interpreting the results, but it provides an order of magnitude approximation for the scale of the tidal effects.

There are 16 stars with $t^*_{\rm sp} < 600\,$Gyrs, which I selected as likely candidates for tidal interaction, since they fall considerably below the bulk of points. They cover a range of $T_{\rm eff}$ but typically have a short period giant planet, as would be expected for interaction to occur. They are typically below or in the lower portion of the main

periodic sequence on the mass-period diagram, indicating that they are rotating more rapidly than would be expected for their spectral type. This is consistent with the interpretation that these are likely to have been spun up by their planets.

The vast majority of host stars show the same $T_{\rm eff}$ - period distribution as the F–M field stars without planets. This implies that in nearly all cases, gyrochronological relations will therefore apply to host stars without alteration, with the exception of the $\sim 2\%$ of stars which appear to have been spun up by their planet.

As a preliminary exercise, I plotted the distribution of planet radius and orbital period as a function of age for the planets with measured rotation periods. Although no relationship with age is apparent, it demonstrates the type of tests which can easily be performed using gyrochronologically determined ages for large number of exoplanets.

8.7 FUTURE WORK

In the near future, I will perform a more thorough study of the rotational properties of the *Kepler* exoplanet host stars, building on the work in Chapter 6 and 7. I intend to continue work on the ACF method and run extensive simulations in order to quantify detection efficiency. I will seek to automate the ACF period verification stages for use on large datasets, and proceed by applying this technique to the whole *Kepler* sample. It will be important at this stage to formulate a method for distinguishing between rotation and pulsation effects, which may be possible either using statistics of the ACF or by examination of the phase-folder light curve (see e.g. Affer et al., 2012).

With good calibration, gyrochronology could be used to measure the absolute age for F,G, K and early-M single field stars with rotation measurements, to an accuracy of $\sim 10\%$ (Delorme et al., 2011), while producing more self-consistent age measurements than chromospheric activity and isochrone fitting (Barnes, 2007). To calibrate the gyrochronology relation, accurate ages for stars with measured rotation periods are required. The asteroseismic targets and the 4 open clusters within the *Kepler* field of view will provide age measurements, and by compiling rotation and age

measurements from all previous gyrochronology work, I will be able to place these studies in the context of field star rotation. Comparison to ages derived from stellar evolution models may be able to provide further constraints on the gyrochronology relations.

Chapter 7 demonstrates the potential for spin-orbit interaction studies of exoplanets. The limiting factor is currently the level of uncertainty on the system parameters, which will improve in time as more followup observations are performed. The simple spin-up timescale formulation of Terquem et al. (1998) may not be well suited for application to all spectral types and planet masses, however the more complex versions involve many assumptions about ill-understood parameters. Using more precise data on exoplanet systems, it would be interesting to plot $M_{\rm pl}/M_*$ as a function of $a/R_{\rm pl}$, where a is the orbital separation. By dividing the data sample into stars which show potential spin-up effects and those which don't, one could shed light on the boundaries of tidal interaction and planet destruction, as the planet orbital radius reduces to conserve angular momentum.

REFERENCES

Affer, L., Micela, G., Favata, F. & Flaccomio, E., 2012. The rotation of field stars from CoRoT data. *MNRAS*, **4**24(1), 11–22.

Agüeros, M. A., Covey, K. R., Lemonias, J. J., Law, N. M., Kraus, A., Batalha, N., Bloom, J. S., Cenko, S. B., Kasliwal, M. M., Kulkarni, S. R., Nugent, P. E., Ofek, E. O., Poznanski, D. & Quimby, R. M., 2011. The Factory and the Beehive. I. Rotation Periods for Low-mass Stars in Praesepe. *ApJ*, **7**40, 110.

Aigrain, S., Favata, F. & Gilmore, G., 2004. Characterising stellar micro-variability for planetary transit searches. *A&A*, **4**14, 1139–1152.

Aigrain, S. & Irwin, M., 2004. Practical planet prospecting. *MNRAS*, **3**50, 331–345.

Aigrain, S., Pont, F. & Zucker, S., 2012. A simple method to estimate radial velocity variations due to stellar activity using photometry. *MNRAS*, **4**19, 3147–3158.

Allain, S., 1998. Modelling the angular momentum evolution of low-mass stars with core-envelope decoupling. *A&A*, **3**33, 629–643.

Baglin, A., 2003. COROT: A minisat for pionnier science, asteroseismology and planets finding. *Advances in Space Research*, **3**1, 345–349.

Bakos, G. Á., Kovács, G., Torres, G., Fischer, D. A., Latham, D. W. et al., 2007. HD 147506b: A Supermassive Planet in an Eccentric Orbit Transiting a Bright Star. *ApJ*, **6**70, 826–832.

Bakos, G. Á., Lázár, J., Papp, I., Sári, P. & Green, E. M., 2002. System Description and First Light Curves of the Hungarian Automated Telescope, an Autonomous Observatory for Variability Search. *PASP*, **1**14, 974–987.

Baliunas, S., Sokoloff, D. & Soon, W., 1996. Magnetic Field and Rotation in Lower Main-Sequence Stars: an Empirical Time-dependent Magnetic Bode's Relation? *ApJL*, **4**57, L99.

Baraffe, I., Chabrier, G., Allard, F. & Hauschildt, P. H., 1998. Evolutionary models for solar metallicity low-mass stars: mass-magnitude relationships and color-magnitude diagrams. *A&A*, **3**37, 403–412.

Barnes, S., 2003. On the Rotational Evolution of Solar- and Late-Type Stars, Its Magnetic Origins, and the Possibility of Stellar Gyrochronology. *ApJ*, **5**86(1), 464–479.

Barnes, S., 2007. Ages for Illustrative Field Stars Using Gyrochronology: Viability, Limitations, and Errors. *ApJ*, **6**69(2), 1167–1189.

Barnes, S., 2010. A Simple Nonlinear Model for the Rotation of Main-sequence Cool Stars. I. Introduction, Implications for Gyrochronology, and Color-Period Diagrams. *ApJ*, **7**22(1), 222–234.

Barnes, S. & Kim, Y. C., 2010. Angular Momentum Loss from Cool Stars: An Empirical Expression and Connection to Stellar Activity. *ApJ*, **7**21(1), 675–685.

Barnes, S. & Sofia, S., 1996. On the Origin of the Ultrafast Rotators in Young Star Clusters. *ApJ*, **4**62, 746.

Basri, G., Borucki, W. J. & Koch, D., 2005. The Kepler Mission: A wide-field transit search for terrestrial planets [review article]. *New Astronomy Reviews*, **4**9, 478–485.

Basri, G., Walkowicz, L. M., Batalha, N., Gilliland, R. L., Jenkins, J., Borucki, W. J., Koch, D., Caldwell, D., Dupree, A. K., Latham, D. W., Marcy, G. W., Meibom, S. & Brown, T., 2011. Photometric Variability in Kepler Target Stars. II. An Overview of Amplitude, Periodicity, and Rotation in First Quarter Data. *AJ*, **1**41, 20.

Basri, G., Walkowicz, L. M., Batalha, N., Gilliland, R. L., Jenkins, J., Borucki, W. J., Koch, D., Caldwell, D., Dupree, A. K., Latham, D. W., Meibom, S., Howell, S. & Brown, T., 2010. Photometric Variability in Kepler Target Stars: The Sun Among Stars- -a First Look. *ApJL*, **7**13, L155–L159.

Batalha, N. M., Borucki, W. J., Koch, D. G., Bryson, S. T., Haas, M. R., Brown, T. M., Caldwell, D. A., Hall, J. R., Gilliland, R. L., Latham, D. W., Meibom, S. & Monet, D. G., 2010. Selection, Prioritization, and Characteristics of Kepler Target Stars. *ApJL*, **7**13, L109–L114.

Batalha, N. M., Rowe, J. F., Bryson, S. T., Barclay, T., Burke, C. J. et al., 2012. Planetary Candidates Observed by Kepler, III: Analysis of the First 16 Months of Data. *ApJ*, **s**ubmitted. Available at arXiv:1202.5852.

Bishop, C. M., 2006. *Pattern Recognition and Machine Learning*. Springer.

Boisse, I., Moutou, C., Vidal-Madjar, A., Bouchy, F., Pont, F. et al., 2009. Stellar activity of planetary host star HD 189 733. *A&A*, **4**95, 959–966.

Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T. et al., 2010. Kepler Planet-Detection Mission: Introduction and First Results. *Science*, **3**27(5968), 977–980.

Borucki, W. J., Koch, D. G., Basri, G., Batalha, N., Brown, T. M. et al., 2011. Characteristics of Planetary Candidates Observed by Kepler. II. Analysis of the First Four Months of Data. *ApJ*, **7**36, 19.

Borucki, W. J., Scargle, J. D. & Hudson, H. S., 1985. Detectability of extrasolar planetary transits. *ApJ*, **2**91, 852–854.

Bouvier, J., Forestini, M. & Allain, S., 1997. The angular momentum evolution of low-mass stars. *A&A*, **3**26, 1023–1043.

Bretthorst, G. L., 1998. Bayesian spectrum analysis and parameter estimation, Lecture Notes in Statistics. Tech. rep., Springer-Verlag, Available at http://bayes.wustl.edu/.

Brown, T. M., Latham, D. W., Everett, M. E. & Esquerdo, G. A., 2011. Kepler Input Catalog: Photometric Calibration and Stellar Classification. *AJ*, **1**42(4), 112.

Bryson, S. T., Tenenbaum, P., Jenkins, J. M., Chandrasekaran, H., Klaus, T., Caldwell, D. A., Gilliland, R. L., Haas, M. R., Dotson, J. L., Koch, D. G. & Borucki, W. J., 2010. The Kepler Pixel Response Function. *ApJ*, **7**13, L97–L102.

Carpano, S., Aigrain, S. & Favata, F., 2003. Detecting planetary transits in the presence of stellar variability. Optimal filtering and the use of colour information. *A&A*, **4**01, 743–753.

Castelli, F. & Kurucz, R. L., 2004. New Grids of ATLAS9 Model Atmospheres. In *Modelling of Stellar Atmospheres*, vol. 210 of *IAU Symp.*. Available at arXiv:astro-ph/0405087.

Chabrier, G. & Baraffe, I., 1997. Structure and evolution of low-mass stars. *A&A*, **3**27, 1039–1053.

Christiansen, J. L., Van Cleve, J. E., Jenkins, J. M., Caldwell, D. A., Barclay, T. et al., 2012. Kepler Data Characteristics Handbook. Tech. rep., KSCI-19040-003.

Ciardi, D. R., von Braun, K., Bryden, G., van Eyken, J., Howell, S. B., Kane, S. R., Plavchan, P., Ramírez, S. V. & Stauffer, J. R., 2011. Characterizing the Variability of Stars with Early-release Kepler Data. *AJ*, **1**41(4), 108.

Collier Cameron, A., Campbell, C. G. & Quaintrell, H., 1995. Rotational evolution of magnetic T Tauri stars with accretion discs. II. Approach to the main sequence. *A&A*, **2**98, 133.

Collier Cameron, A., Davidson, V. A., Hebb, L., Skinner, G., Anderson, D. R. et al., 2009. The main-sequence rotation-colour relation in the Coma Berenices open cluster. *MNRAS*, **4**00(1), 451–462.

de Bruijne, J. H. J., 2012. Science performance of Gaia, ESA's space-astrometry mission. *Ap&SS*, **3**41, 31–41.

Defaÿ, C., Deleuil, M. & Barge, P., 2001. A Bayesian method for the detection of planetary transits. *A&A*, **3**65, 330–340.

DeGroot, M., 1970. *Optimal Statistical Decisions*. Wiley-Interscience.

Delorme, P., Collier Cameron, A., Hebb, L., Rostron, J., Lister, T. A., Norton, A. J., Pollacco, D. & West, R. G., 2011. Stellar rotation in the Hyades and Praesepe: gyrochronology and braking time-scale. *MNRAS*, **4**13, 2218–2234.

Dorren, J. D. & Guinan, E. F., 1982. Evidence for starspots on single solar-like stars. *AJ*, **8**7, 1546–1557.

Doyle, L. R., Carter, J. A., Fabrycky, D. C., Slawson, R. W., Howell, S. B. et al., 2011. Kepler-16: A Transiting Circumbinary Planet. *Science*, **3**33, 1602–.

Dressing, C. D. & Charbonneau, D., 2013. The Occurrence Rate of Small Planets around Small Stars. *ArXiv e-prints*.

Edwards, S., Strom, S. E., Hartigan, P., Strom, K. M., Hillenbrand, L. A., Herbst, W., Attridge, J., Merrill, K. M., Probst, R. & Gatley, I., 1993. Angular momentum regulation in low-mass young stars surrounded by accretion disks. *AJ*, **1**06, 372–382.

Eyer, L. & Grenon, M., 1997. Photometric Variability in the HR Diagram. In R. M. Bonnet, E. Høg, P. L. Bernacca, L. Emiliani, A. Blaauw, C. Turon, J. Kovalevsky, L. Lindegren, H. Hassan, M. Bouffard, B. Strim, D. Heger, M. A. C. Perryman & L. Woltjer, eds., *Hipparcos - Venice '97*, vol. 402 of *ESA Special Publication*, 467–472.

Freeman, K. & Bland-Hawthorn, J., 2002. The New Galaxy: Signatures of Its Formation. *ARA&A*, **4**0, 487–537.

Fröhlich, C., 2011. Total Solar Irradiance: What Have We Learned from the Last Three Cycles and the Recent Minimum? *Space Sci. Rev.*, 133.

Gilliland, R. L., 2008. Photometric Oscillations of Low-Luminosity Red Giant Stars. *AJ*, **1**36, 566–579.

Gilliland, R. L., Chaplin, W. J., Dunham, E. W., Argabright, V. S., Borucki, W. J., Basri, G., Bryson, S. T., Buzasi, D. L., Caldwell, D. A., Elsworth, Y. P., Jenkins, J. M., Koch, D. G., Kolodziejczak, J., Miglio, A., van Cleve, J., Walkowicz, L. M. & Welsh, W. F., 2011. Kepler Mission Stellar and Instrument Noise Properties. *ApJS*, **1**97, 6.

Goldreich, P. & Soter, S., 1966. Q in the Solar System. *Icarus*, **5**, 375–389.

Goulding, N. T., Barnes, J. R., Pinfield, D. J., Kovács, G., Birkby, J., Hodgkin, S., Catalán, S., Sipőcz, B., Jones, H. R. A., Del Burgo, C., Jeffers, S. V., Nefs, S., Gálvez-Ortiz, M. C. & Martin, E. L., 2012. J-band variability of M dwarfs in the WFCAM Transit Survey. *MNRAS*, **4**27, 3358–3373.

Hall, D. S. & Henry, G. W., 1994. The Law of Starspot Lifetimes. *International Amateur-Professional Photoelectric Photometry Communications*, **5**5, 51.

Harrison, T. E., Coughlin, J. L., Ule, N. M. & López-Morales, M., 2012. Kepler Cycle 1 Observations of Low-mass Stars: New Eclipsing Binaries, Single Star Rotation Rates, and the Nature and Frequency of Starspots. *AJ*, **1**43, 4.

Hartigan, J. A. & Hartigan, P. M., 1985. The dip test of unimodality. *The Annals of Statistics*, **1**3, 70–84.

Hartman, J. D., Bakos, G. Á., Noyes, R. W., Sipőcz, B., Kovács, G., Mazeh, T., Shporer, A. & Pál, A., 2011. A Photometric Variability Survey of Field K and M Dwarf Stars with HATNet. *AJ*, **1**41, 166.

Harvey, J. W., 1985. Future missions in Solar, heliospheric and space plasma physics. *ESA SP-235*.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C. & Liu, H. H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *RSPSA*, **4**54, 903–995.

Hughes, D. W., Rosner, R. & Weiss, N. O., 2007. *The Solar Tachocline*. Cambridge University Press.

Hulot, J. C., Baudin, F., Samadi, R. & Goupil, M. J., 2011. A quantitative analysis of stellar activity based on CoRoT photometric data. *ArXiv e-prints*.

Husnoo, N., Pont, F., Mazeh, T., Fabrycky, D., Hébrard, G., Bouchy, F. & Shporer, A., 2012. Observational constraints on tidal effects using orbital eccentricities. *MNRAS*, **4**22, 3151–3177.

Irwin, J., 2007. *Ph.D. thesis*. Ph.D. thesis, University of Cambridge.

Irwin, J., Aigrain, S., Hodgkin, S., Irwin, M., Bouvier, J., Clarke, C., Hebb, L. & Moraux, E., 2006. The Monitor project: rotation of low-mass stars in the open cluster M34. *MNRAS*, **3**70, 954–974.

Irwin, J., Berta, Z. K., Burke, C. J., Charbonneau, D., Nutzman, P., West, A. A. & Falco, E. E., 2011. On the Angular Momentum Evolution of Fully Convective Stars: Rotation Periods for Field M-dwarfs from the MEarth Transit Survey. *ApJ*, **7**27(1), 56.

Irwin, J. & Bouvier, J., 2009. The rotational evolution of low-mass stars. In *The Ages of Stars*, vol. 258 of *IAU Symp.*, 363–374.

Irwin, J., Hodgkin, S., Aigrain, S., Hebb, L., Bouvier, J., Clarke, C., Moraux, E. & Bramich, D. M., 2007. The Monitor project: rotation of low-mass stars in the open cluster NGC2516. *MNRAS*, **3**77, 741–758.

Isaacson, H. & Fischer, D., 2010. Chromospheric Activity and Jitter Measurements for 2630 Stars on the California Planet Search. *ApJ*, **7**25, 875–885.

Jackson, R. J. & Jeffries, R. D., 2012. Why do some young cool stars show spot modulation while others do not? *MNRAS*, **4**23, 2966–2976.

Jenkins, J. M., 2002. The Impact of Solar-like Variability on the Detectability of Transiting Terrestrial Planets. *ApJ*, **5**75, 493–505.

Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., Twicken, J. D., Bryson, S. T. et al., 2010a. Initial Characteristics of Kepler Long Cadence Data for Detecting Transiting Planets. *ApJ*, **7**13, L120–L125.

Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., Twicken, J. D., Bryson, S. T. et al., 2010b. Overview of the Kepler Science Processing Pipeline. *ApJ*, **7**13, L87–L91.

Kawaler, S. D., 1987. Angular momentum in stars - The Kraft curve revisited. *PASP*, **9**9, 1322–1228.

Kawaler, S. D., 1988. Angular momentum loss in low-mass stars. *ApJ*, **3**33, 236–247.

Kawaler, S. D., 1989. Rotational dating of middle-aged stars. *ApJ*, **3**43, L65–L68.

Kinemuchi, K., Barclay, T., Fanelli, M., Pepper, J., Still, M. & Howell, S. B., 2012. Demystifying Kepler Data: A Primer for Systematic Artifact Mitigation. *PASP*, **1**24, 963–984.

Kiraga, M. & Stepien, K., 2007. Age-Rotation-Activity Relations for M Dwarf Stars. *ACTAA*, **5**7, 149–172.

Kjeldsen, H. & Bedding, T. R., 1995. Amplitudes of stellar oscillations: the implications for asteroseismology. *A&A*, **2**93, 87–106.

Koch, D. G., Borucki, W. J., Basri, G., Batalha, N. M., Brown, T. M. et al., 2010. Kepler Mission Design, Realized Photometric Performance, and Early Science. *ApJ*, **7**13, L79–L86.

Koenigl, A., 1991. Disk accretion onto magnetic T Tauri stars. *ApJ*, **3**70, L39–L43.

Kovács, G., Bakos, G. & Noyes, R. W., 2005. A trend filtering algorithm for wide-field variability surveys. *MNRAS*, **3**56, 557–567.

Lissauer, J. J., Ragozzine, D., Fabrycky, D. C., Steffen, J. H., Ford, E. B. et al., 2011. Architecture and Dynamics of Kepler's Candidate Multiple Transiting Planet Systems. *ApJS*, **1**97, 8.

Loader, C., 1999. *Local Regression and Likelihood*. Springer, New York.

Mann, A. W., Gaidos, E., Lépine, S. & Hilton, E. J., 2012. They Might be Giants: Luminosity Class, Planet Occurrence, and Planet-Metallicity Relation of the Coolest Kepler Target Stars. *ApJ*, **7**53, 90.

Matt, S. & Pudritz, R. E., 2005. Accretion-powered Stellar Winds as a Solution to the Stellar Angular Momentum Problem. *ApJ*, **6**32, L135–L138.

Mazeh, T., 2008. Observational Evidence for Tidal Interaction in Close Binary Systems. In M.-J. Goupil & J.-P. Zahn, eds., *EAS Publications Series*, vol. 29 of *EAS Publications Series*, 1–65.

McQuillan, A., Aigrain, S. & Roberts, S., 2012. Statistics of stellar variability from Kepler. I. Revisiting Quarter 1 with an astrophysically robust systematics correction. *A&A*, **5**39, A137.

Meibom, S., Barnes, S. A., Latham, D. W., Batalha, N., Borucki, W. J. et al., 2011. The Kepler Cluster Study: Stellar Rotation in NGC 6811. *ApJL*, **7**33(1), L9.

Meibom, S., Mathieu, R. D. & Stassun, K. G., 2009. Stellar Rotation in M35: Mass-Period Relations, Spin-Down Rates, and Gyrochronology. *ApJ*, **6**95(1), 679–694.

Mestel, L., 1984. Angular Momentum Loss During Pre-Main Sequence Contraction. In S. L. Baliunas & L. Hartmann, eds., *Cool Stars, Stellar Systems, and the Sun*, vol. 193 of *Lecture Notes in Physics, Berlin Springer Verlag*, 49.

Michel, E., Samadi, R., Baudin, F., Barban, C., Appourchaux, T. & Auvergne, M., 2009. Intrinsic photometric characterisation of stellar oscillations and granulation. Solar reference values and CoRoT response functions. *A&A*, **4**95, 979–987.

Monet, D. G., Jenkins, J. M., Dunham, E. W., Bryson, S. T., Gilliland, R. L., Latham, D. W., Borucki, W. J. & Koch, D. G., 2010. Preliminary Astrometric Results from Kepler. *ApJ*, **s**ubmitted, available at arXiv:1203.1383.

Mordasini, C., Alibert, Y., Georgy, C., Dittkrist, K.-M., Klahr, H. & Henning, T., 2012. Characterization of exoplanets from their formation. II. The planetary mass-radius relationship. *A&A*, **5**47, A112.

Mosser, B., Baudin, F., Lanza, A. F., Hulot, J. C., Catala, C., Baglin, A. & Auvergne, M., 2009. Short-lived spots in solar-like stars as observed by CoRoT. *A&A*, **5**06, 245–254.

Muirhead, P. S., Hamren, K., Schlawin, E., Rojas-Ayala, B., Covey, K. R. & Lloyd, J. P., 2012. Characterizing the Cool Kepler Objects of Interests. New Effective Temperatures, Metallicities, Masses, and Radii of Low-mass Kepler Planet-candidate Host Stars. *ApJL*, **7**50, L37.

Murphy, S. J., 2012. An examination of some characteristics of Kepler short- and long-cadence data. *MNRAS*, **4**22, 665–671.

Neal, R. M., 1998. *Assessing relevance determination methods using DELVE*. Neural Networks and Machine Learning. Springer-Verlag.

Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K. & Vaughan, A. H., 1984. Rotation, convection, and magnetic activity in lower main-sequence stars. *ApJ*, **2**79, 763–777.

Ossendrijver, M., 2003. The solar dynamo. *A&A Rev.*, **1**1, 287–367.

Penny, W. & Roberts, S., 2002. Bayesian Multivariate Autoregressive Models with Structured Priors. *IEE Proceedings on Vision, Signal & Image Processing*, **1**49(1), 33–41.

Pizzolato, N., Maggio, A., Micela, G., Sciortino, S. & Ventura, P., 2003. The stellar activity-rotation relationship revisited: Dependence of saturated and non-saturated X-ray emission regimes on stellar mass for late-type dwarfs. *A&A*, **3**97, 147–157.

Pollacco, D. L., Skillen, I., Collier Cameron, A., Christian, D. J., Hellier, C. et al., 2006. The WASP Project and the SuperWASP Cameras. *PASP*, **1**18, 1407–1418.

Pont, F., 2009. Empirical evidence for tidal evolution in transiting planetary systems. *MNRAS*, **3**96, 1789–1796.

Pont, F., Aigrain, S. & Zucker, S., 2011a. Reassessing the radial-velocity evidence for planets around CoRoT-7. *MNRAS*, **4**11, 1953–1962.

Pont, F., Husnoo, N., Mazeh, T. & Fabrycky, D., 2011b. Determining eccentricities of transiting planets: a divide in the mass-period plane. *MNRAS*, **4**14, 1278–1284.

Pont, F., Zucker, S. & Queloz, D., 2006. The effect of red noise on planetary transit detection. *MNRAS*, **3**73, 231–242.

Prša, A., Batalha, N., Slawson, R. W., Doyle, L. R., Welsh, W. F., Orosz, J. A., Seager, S., Rucker, M., Mjaseth, K., Engle, S. G., Conroy, K., Jenkins, J., Caldwell, D., Koch, D. & Borucki, W., 2011. Kepler Eclipsing Binary Stars. I. Catalog and Principal Characterization of 1879 Eclipsing Binaries in the First Data Release. *AJ*, **1**41, 83.

Quintana, E. V., Jenkins, J. M., Clarke, B. D., Chandrasekaran, H., Twicken, J. D., McCauliff, S. D., Cote, M. T., Klaus, T. C., Allen, C., Caldwell, D. A. & Bryson, S. T., 2010. Pixel-level calibration in the Kepler Science Operations Center pipeline. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 7740 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*.

Radick, R. R., Mihalas, D., Hartmann, L., Worden, S. P., Africano, J. L., Klimke, A. & Tyson, E. T., 1982. The photometric variability of solar-type stars. I - Preliminary results for the Pleiades, Hyades, and the Malmquist Field. *PASP*, **9**4, 934–944.

Reiners, A. & Mohanty, S., 2012. Radius-dependent Angular Momentum Evolution in Low-mass Stars. I. *ApJ*, **7**46, 43.

Scargle, J. D., 1982. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *ApJ*, **2**63, 835–853.

Schatzman, E., 1962. A theory of the role of magnetic activity during star formation. *Annales d'Astrophysique*, **2**5, 18.

Scholz, A., Irwin, J., Bouvier, J., Sipőcz, B. M., Hodgkin, S. & Eislöffel, J., 2011. Rotation periods for very low mass stars in Praesepe. *MNRAS*, **4**13, 2595–2605.

Scholz, F. W. & Stephens, M. A., 1987. K-sample Anderson-Darling Tests. *Journal of the American Statistical Association*, **8**2(399), 918–924.

Sekiguchi, M. & Fukugita, M., 2000. A Study of the B-V Color-Temperature Relation. *AJ*, **1**20, 1072–1084.

Shannon, C. E., 1951. Prediction and Entropy of Printed English. *The Bell System Technical Journal*, **3**0, 50–64.

Shumway, R. H. & Stoffer, D. S., 2010. *Time Series Analysis and Its Applications: With R Examples*. Springer.

Siess, L., Dufour, E. & Forestini, M., 2000. An internet server for pre-main sequence tracks of low- and intermediate-mass stars. *A&A*, **3**58, 593–599.

Skumanich, A., 1972. Time Scales for CA II Emission Decay, Rotational Braking, and Lithium Depletion. *ApJ*, **1**71, 565.

Smith, J. C., Stumpe, M. C., van Cleve, J. E., Jenkins, J. M., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Kolodziejczak, J. J., McCauliff, S. D., Morris, R. L. & Twicken, J. D., 2012. Kepler Presearch Data Conditioning II - A Bayesian Approach to Systematic Error Correction. *PASP*, **s**ubmitted. Available at arXiv:1203.1383.

Stauffer, J., 1994. Angular Momentum Evolution of Young Main Sequence Stars. In J.-P. Caillault, ed., *Cool Stars, Stellar Systems, and the Sun*, vol. 64 of *Astronomical Society of the Pacific Conference Series*, 163.

Stauffer, J. B. & Hartmann, L. W., 1986. The rotational velocities of low-mass stars. *PASP*, **9**8, 1233–1251.

Stumpe, M. C., Smith, J. C., Van Cleve, J. E., Twicken, J. D., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Jenkins, J. M., Kolodziejczak, J. J., McCauliff, S. D. & Morris, R. L., 2012. Kepler Presearch Data Conditioning I - Architecture and Algorithms for Error Correction in Kepler Light Curves. *PASP*, **1**24, 985–999.

Tamuz, O., Mazeh, T. & Zucker, S., 2005. Correcting systematic effects in a large set of photometric light curves. *MNRAS*, **3**56, 1466–1470.

Tarter, J. C., Backus, P. R., Mancinelli, R. L., Aurnou, J. M., Backman, D. E. et al., 2007. A Reappraisal of The Habitability of Planets around M Dwarf Stars. *Astrobiology*, **7**, 30–65.

Terquem, C., Papaloizou, J. C. B., Nelson, R. P. & Lin, D. N. C., 1998. On the Tidal Interaction of a Solar-Type Star with an Orbiting Companion: Excitation of g-Mode Oscillation and Orbital Evolution. *ApJ*, **5**02, 788.

Twicken, J. D., Chandrasekaran, H., Jenkins, J. M., Gunter, J. P., Girouard, F. & Klaus, T. C., 2010a. Presearch data conditioning in the Kepler Science Operations Center pipeline. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 7740 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*.

Twicken, J. D., Clarke, B. D., Bryson, S. T., Tenenbaum, P., Wu, H., Jenkins, J. M., Girouard, F. & Klaus, T. C., 2010b. Photometric analysis in the Kepler Science Operations Center pipeline. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 7740 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*.

Udalski, A., Paczynski, B., Zebrun, K., Szymanski, M., Kubiak, M., Soszynski, I., Szewczyk, O., Wyrzykowski, L. & Pietrzynski, G., 2002. The Optical Gravitational Lensing Experiment. Search for Planetary and Low-Luminosity Object Transits in the Galactic Disk. Results of 2001 Campaign. *Acta Astronomica*, **5**2, 1–37.

Udalski, A., Szymanski, M. K., Soszynski, I. & Poleski, R., 2008. The Optical Gravitational Lensing Experiment. Final Reductions of the OGLE-III Data. *Acta Astronomica*, **5**8, 69–87.

van Cleve, J. & Caldwell, D. A., 2009. *Kepler* Instrument Handbook, KSCI 19033-001. Tech. rep., NASA Ames Research Center.

Vaughan, A. H., Preston, G. W. & Wilson, O. C., 1978. Flux measurements of CA II H and K emission. *PASP*, **9**0, 267–274.

Verner, G. A., Chaplin, W. J., Basu, S., Brown, T. M., Hekker, S. et al., 2011. Verification of the Kepler Input Catalog from Asteroseismology of Solar-type Stars. *ApJ*, **7**38, L28.

Walkowicz, L. M. & Basri, G., 2010. First thoughts on stellar variability from Kepler commissioning data. In A. G. Kosovichev, A. H. Andrei & J.-P. Roelot, eds., *IAU Symposium*, vol. 264 of *IAU Symposium*, 469–474.

Wall, J. V., Jenkins, C. R., Ellis, R., Huchra, J., Kahn, S., Rieke, G. & Stetson, P. B., 2003. *Practical Statistics for Astronomers*. Cambridge University Press.

Welsh, W. F., Orosz, J. A., Carter, J. A., Fabrycky, D. C., Ford, E. B. et al., 2012. Transiting circumbinary planets Kepler-34 b and Kepler-35 b. *Nature*, **4**81, 475–479.

West, A. A., Hawley, S. L., Bochanski, J. J., Covey, K. R., Reid, I. N., Dhital, S., Hilton, E. J. & Masuda, M., 2008. Constraining the Age-Activity Relation for Cool Stars: The Sloan Digital Sky Survey Data Release 5 Low-Mass Star Spectroscopic Sample. *AJ*, **1**35, 785–795.

Wolfgang, A. & Laughlin, G., 2011. Combining Kepler and HARPS Occurrence Rates to Infer the Period-Mass-Radius Distribution of Super-Earths/Sub-Neptunes. *ArXiv e-prints*.

Wright, J. T., 2005. Radial Velocity Jitter in Stars from the California and Carnegie Planet Search at Keck Observatory. *PASP*, **1**17, 657–664.

Zechmeister, M. & Kürster, M., 2009. The generalised Lomb-Scargle periodogram. *A&A*, **4**96(2), 577–584.