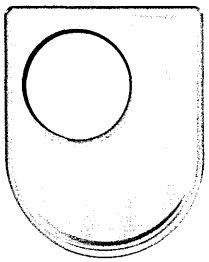
# Populations in the Kepler field

## Thesis

UNRESTRICTED



THE OPEN UNIVERSITY

FACULTY OF SCIENCE

# Populations in the Kepler field

Robert Farmer MPhys

*Submitted for the degree of Doctor of Philosophy*

July 2014

DATE OF SUBMISSION = 31 MARCH 2014

DATE OF AWARD = 27 JUNE 2014

I would first like to thank my supervisors Ulrich Kolb and Andrew Norton for guiding through this process and putting up with my many speelling mistakes. Without their help this thesis would be much shorter and much more incomprehensible.

I'd like to thank my Mum and my Dad, as well as the rest of my family for supporting me through this. Without their support, reassurance, free food and willingness to move furniture I wouldn't be here today.

To the friends I've made here, thank you. Without your support I'm sure I would have gone crazy along time ago. Thanks for all meals out, drinks and the many deep philosophical discussions over lunch. Thanks to the friends back home for the weekends away that broke up the time spent at work. Lets see if we can organise something with less than 6 months notice in the future.

Oh, and just to be clear my name is not Bill, after 3 and a 1/2 years I hope people have realised that.

# Abstract

Using a population synthesis model I have created a synthetic catalogue of stars in the Kepler field of view. This model has then been subjected to the same biases and selection effects inherent in the selection of stars for the Kepler transit survey mission. This produced a synthetic Kepler Input Catalogue (KIC) which was subjected to the Kepler Stellar Classification Program (SCP) method for determining stellar parameters. I achieve a satisfactory match between the synthetic KIC and the real KIC in the $\log g$ vs $\log T_{eff}$ diagram. I find a median difference of $\Delta T_{eff} = +500$ K and $\sim \Delta \log g = -0.2$ dex for main sequence stars, although there is a large variation across parameter space. I find no significant difference between $\Delta T_{eff}$ and $\Delta \log g$ for single stars and the primary star in a binary system. I also re-created the Kepler target selection method and found that the binary fraction is unchanged by the target selection. The fraction of main sequence stars in the sample increases from 75% to 80%, and the giant star fraction decreases from 25% to 20%.

I have then used the synthetic KIC to build a of synthetic sample of eclipsing binaries (EBs) in the Kepler field. Comparing the synthetic catalogue to the Kepler EB catalogue I find that the Kepler EB pipeline introduces significant biases into the derived temperature ratio and fractional radii. I then tested the effect of different initial mass ratio distributions (IMRDs) and initial binary fraction distributions (IBFDs). At this time, all distributions fail to match the data, such that their parameters can not be constrained.

Modelling the population of asteroseismic binaries, where both stars have a detectable asteroseismic signal, have shown a way to constrain the IMRD for equal mass systems. This method is independent of the binary period and orbital orientation. The number of detectable asteroseismic binaries increases from 87 for the IMR parameter $s = -0.5$ to 256 for $s = 1.0$. The number of detectable asteroseismic EBs increases from $34.0 \pm 6.0$ ($s = -0.5$) to $59.0 \pm 6.0$ ($s = 1.0$). This number shows disagreement with the number of actual systems detected (2 for $P_{orb} < 40$ days), which can not be explained by incompleteness alone.

# Contents

**List of Figures** iv

**List of Tables** vii

**1 Introduction** 1

1.1 Population modelling . . . . . 2

1.1.1 Stellar evolution . . . . . 3

1.1.2 Population synthesis . . . . . 8

1.1.3 Observational biases . . . . . 10

1.2 Kepler . . . . . 12

**2 Population synthesis** 16

2.1 Population synthesis model . . . . . 17

2.1.1 Binary evolution . . . . . 17

2.1.2 Galactic model . . . . . 20

2.1.3 Field of view . . . . . 24

2.1.4 Population characteristics . . . . . 26

2.1.5 Magnitude of a system . . . . . 30

2.1.6 Extinction . . . . . 32

2.1.7 Creating a discrete sample . . . . . 33

2.2 Eclipsing Binaries . . . . . 34

2.2.1 Darkening coefficients . . . . . 35

**3 Kepler target selection** 38

3.1 Introduction . . . . . 39
3.2 Kepler target list selection . . . . . 40
3.2.1 Stellar classification . . . . . 41
3.2.2 S/N determination . . . . . 43
3.2.3 Testing the target selection code . . . . . 49
3.3 Synthetic population of Kepler stars . . . . . 51
3.3.1 Sample size . . . . . 52
3.3.2 Distribution in colour-colour diagram . . . . . 52
3.3.3 Stellar parameter distribution . . . . . 55
3.3.4 Post-target selection distributions . . . . . 60
3.3.5 Effect of target selection . . . . . 62
3.3.6 Comparison of SCP and physical parameters . . . . . 65
3.4 Discussion . . . . . 68
3.4.1 Effect of metallicity . . . . . 69
3.4.2 Sensitivity to changes in Galactic parameters . . . . . 71
3.5 Conclusions . . . . . 74

**4 Eclipsing Binaries in the Kepler field** **76**

4.1 Introduction . . . . . 77
4.2 Method . . . . . 79
4.2.1 Updates to the population synthesis method . . . . . 79
4.2.2 DERBs . . . . . 79
4.2.3 Contact Binaries . . . . . 84
4.3 Derivation of an eclipsing binary sample . . . . . 86
4.3.1 Stellar noise . . . . . 86
4.3.2 Deriving an EB sample . . . . . 95
4.3.3 Kepler EB pipeline . . . . . 98
4.3.4 EBAI Training sets . . . . . 99
4.3.5 Comparison of EBAI training sets . . . . . 105
4.4 Results . . . . . 108

4.4.1 Scaling factors . . . . . 108
4.4.2 Number counts . . . . . 108
4.4.3 Period distribution . . . . . 110
4.4.4 EBAI parameters of the synthetic sample . . . . . 114
4.4.5 Variation of the binary fraction . . . . . 116
4.5 Discussion . . . . . 121
4.5.1 Noise model . . . . . 121
4.5.2 Period doubling . . . . . 123
4.5.3 Eclipsing DERBs . . . . . 124
4.5.4 Performance of EBAI . . . . . 124
4.5.5 Effect of eccentricity . . . . . 125
4.5.6 Higher order systems . . . . . 127
4.6 Conclusions . . . . . 128

**5 Asteroseismic binaries** **130**

5.1 Introduction . . . . . 130
5.2 Method . . . . . 134
5.3 Results . . . . . 137
5.3.1 Number of detectable systems . . . . . 137
5.3.2 Asteroseismic binary distributions . . . . . 139
5.3.3 Mass ratio distribution . . . . . 140
5.3.4 Period distribution . . . . . 142
5.3.5 Asteroseismic signal separation . . . . . 143
5.4 Astrophysical contamination . . . . . 144
5.4.1 Resolved binaries . . . . . 145
5.4.2 Mass transferring systems . . . . . 145
5.4.3 Blended systems . . . . . 147
5.5 Asteroseismic eclipsing binaries . . . . . 148
5.6 Discussion . . . . . 149
5.6.1 Dependence on Galactic parameters . . . . . 150

5.6.2 Noise model . . . . . 152
5.7 Conclusions . . . . . 153

**6 Conclusions & Further work** **155**
6.1 Conclusions . . . . . 155
6.2 BiSEPSv3.0 . . . . . 157
6.2.1 Introduction . . . . . 157
6.2.2 Method . . . . . 158
6.2.3 MESA . . . . . 161
6.3 False positives in transit surveys . . . . . 162

**Acronyms** **165**

**Bibliography** **166**

# List of Figures


1.1 Stellar evolution track for a $1M_{\odot}$, $z = 0.02$ star . . . . . 4

1.2 Equipotentials for two equal sized stars in orbit around their centre of mass . . . . . 5

1.3 The light curve of a planetary transit . . . . . 14

2.1 A HRD of non-randomly perturbed systems . . . . . 21

2.2 A HRD of randomly perturbed systems . . . . . 21

2.3 A map of the all-sky extinction ($A_v$), using Hakkila et al. (1997), out to a distance of 1 Kpc. Kepler's field of view is denoted in black . . . . . 25

2.4 A map of the all-sky extinction ($A_v$), using Drimmel et al. (2003), out to a distance of 1 Kpc. Kepler's field of view is denoted in black . . . . . 25

2.5 A comparison of Flower (1996) and Girardi et al. (2002) BCs over a range of $T_{eff}$ . . . . . 32

2.6 The distribution function $\Gamma$ and the sub-sampled $\Gamma$ for a population of 11,000 stars . . . . . 34

3.1 Two examples of Kepler's pixel response function . . . . . 44

3.2 Different stages of adding noise to synthetic Kepler CCD (a) . . . . . 46

3.3 Different stages of adding noise to synthetic Kepler CCD (b) . . . . . 47

3.4 Optimal aperture of a star . . . . . 48

3.5 Normalised magnitude distribution of stars in the target selection . . . 51

3.6 Distribution of objects in a $r - i$ vs $g - r$ colour-colour diagram . . . . . 54

3.7 The distribution of the synthetic KIC sample over the $\log T_{eff}$ -$\log g$ plane . . . . . 58
3.8 The distribution of $\log T_{eff}$ and $\log g$ derived for the synthetic sample using the stellar classification program (SCP) . . . . . 59
3.9 A comparison of the distribution of systems before and after the target selection for the Q2 sample . . . . . 63
3.10 A comparison of the distribution of systems before and after the target selection for the synthetic SCP derived parameters . . . . . 64
3.11 A comparison of the distribution of systems before and after the target selection for the synthetic, real parameters . . . . . 64
3.12 Distribution of the median of the difference between real and SCP - derived effective temperature . . . . . 68
3.13 Distribution of the median of the difference between real and SCP - derived surface gravity . . . . . 69

4.1 Formation probability of DERBs . . . . . 84
4.2 Stellar variability for MS single stars as a function of mass . . . . . 93
4.3 Stellar variability for GB single stars as function of mass . . . . . 93
4.4 Predicted combined noise value for all stars . . . . . 94
4.5 Figure 1 from Jenkins et al. (2010b) . . . . . 94
4.6 Comparison of POLYFIT performance on a case where it falls short . . . 101
4.7 Comparison of POLYFIT performance for a case where it succeeds . . . 102
4.8 Comparison of EBAIs performance on training set 1 . . . . . 103
4.9 Comparison of EBAIs performance on training set 2 . . . . . 104
4.10 Comparison of EBAIs performance on using result from training set 1 on light curves made for training set 2. . . . . 106
4.11 Comparison of EBAIs performance on using result from training set 2 on light curves made for training set 1. . . . . 107
4.12 Period distributions binaries at different points to the pipeline . . . . . 112
4.13 Period distribution of all synthetic eclipsing binaries . . . . . 112

4.14 Period distribution of MS+MS eclipsing binaries . . . . . 113
4.15 Period distribution of giant containing eclipsing binaries systems . . . 113
4.16 $T_2/T_1$ distribution of eclipsing binaries as derived by EBAI . . . . . 115
4.17 The true $T_2/T_1$ distribution of eclipsing binaries . . . . . 116
4.18 $(R_1+R_2)/a$ distribution of eclipsing binaries as derived by EBAI . . . . 117
4.19 The true $(R_1+R_2)/a$ distribution of eclipsing binaries . . . . . 117
4.20 Initial binary fraction distributions . . . . . 118
4.21 Period distribution of all synthetic eclipsing binaries for different binary fractions . . . . . 120
4.22 $T_2/T_1$ distribution of eclipsing binaries as derived by EBAI for different binary fractions . . . . . 120
4.23 $(R_1+R_2)/a$ distribution of eclipsing binaries as derived by EBAI for different binary fractions . . . . . 121

5.1 Schematic of $p$-modes inside star . . . . . 132
5.2 A HRD of detected asteroseismic signals for $s=0.0$ . . . . . 141
5.3 PDMR and IMR of asteroseismic binaries for $s=0.0$ . . . . . 142
5.4 Period distribution of asteroseismic binaries for $s=0.0$ . . . . . 143

# List of Tables


3.1 Extinction coefficients for a 5000 K, $\log g = 4.0, \log(Z/Z_{\odot}) = 0.0$ star, from tables provided in Girardi et al. (2002) . . . . . 43

3.2 Correction terms applied to the calculated KIC magnitudes . . . . . 54

3.3 New correction terms applied to the calculated KIC magnitudes . . . . 55

3.4 The relative distribution of stellar types among the single stars in the synthetic sample, before and after target selection. . . . . 60

3.5 The relative distribution of binary classes in the synthetic sample, before and after target selection . . . . . 61

3.6 The median values of the differences $\Delta T_{eff}$ and $\Delta \log g$ for single stars 66

3.7 The median values of the differences $\Delta T_{eff}$ and $\Delta \log g$ for binaries . . 67

4.1 The number of DERBs and binaries for the synthetic Kepler field of view 83

4.2 Scaling factors for different IMRD and IBFD . . . . . 108

4.3 The number of EBs detected for different values of $s$ . . . . . 109

4.4 The number of each type of EB detected for different values of $s$ . . . . 109

4.5 The number of EBs detected for different values of the initial binary fraction distribution (IBFD) . . . . . 119

4.6 The number of each type of EB detected for different values of the IBFD 119

5.1 Number of target selected asteroseismic detections in the Kepler field . 138

5.2 Number of asteroseismic detections for all systems in the Kepler field . 139

5.3 Number of target selected asteroseismic detections that underwent, or undergoing mass transfer . . . . . 146

5.4 Number of detections for all systems that non-asteroseismic binaries, or undergoing mass transfer . . . . . 146

5.5 Number of predicted blends for different values of $s$ . . . . . 148

5.6 Number of detections for asteroseismic eclipsing binaries . . . . . 150

# Chapter 1

# Introduction

Understanding what we see when we observe the Universe requires a model which can translate what we see into a theory of stellar and Galactic evolution. What kind of star am I looking at? How old is this star? What will happen to this star in the future? Without a valid model we can not place the results into context, in order to answer questions such as: How did the Galaxy form? The way to model the Galaxy is via population synthesis. Here we combine a model of the Galactic distribution of stars with a stellar evolution model providing the stellar parameters as a function of age. We then take the observables measured and calibrate the model to what we can see today. By varying the initial parameter distributions, untill the results of the model match the observations, we can learn how the Galaxy formed from what we see today.

In this work I present my synthetic model of the population of stars in the Kepler field of view. I apply the biases and selection effects inherent to the Kepler mission to derive a synthetic catalogue of stars, which is representative of what the Kepler mission would observe. This allows me to quantify the bias in the Kepler results and determine how this selection process alters the results derived from the Kepler mission. Finally, I model several sub-populations of stars, namely eclipsing binaries and asteroseismic binaries, in the Kepler field. These synthetic populations, when compared to the Kepler data, allow us to constrain the poorly known Galactic parameter of the initial mass ratio distribution (IMRD), which determines how binary

stars form.

With Kepler's primary observing mission finished, the focus has moved to understanding the results Kepler has acquired in the quest for $\eta_{earth}$, the fraction of Sun-like stars with Earth-like planets. Part of this work requires an understanding of the biases and selection effects inherent in the Kepler mission due to the targeted nature of the Kepler input catalogue (KIC). The aim of the target selection was to choose the systems where it was expected to detect an Earth-like planet if it existed in the system. This leads to biasing Kepler to observe what it believes are small cool stars, thus the objects Kepler observes are not fully representative of the Galaxy. This bias would need to be corrected for to translate Kepler's measured $\eta_{earth}$ into a Galactic $\eta_{earth}$. In general, the systems Kepler observes are too faint for spectroscopic follow up, to measure the radial velocity (RV) of the system, which constrains mass of the planet (Mayor & Queloz, 1995). Thus without the RV data, we only have the transit light curve, which on its own can not be used to distinguish planets from binaries. Thus to derive $\eta_{earth}$ we would also need to have an accurate, predictive, model of the binaries in the Kepler field, such that we could *statisticaly* say which are planets and which are binaries.

In section 1.1.1 I review our knowledge of single star and binary star evolution. Section 1.1.2, I look at the Galactic modelling required to produce a population synthesis model. Section 1.1.3 looks at the biases introduced into observations and why these need to be modelled. Finally section 1.2 looks at the Kepler mission and some of its results.

## 1.1 Population modelling

Modelling the structure of the Galaxy successfully requires bringing together several areas of astrophysics. We need to understand stellar evolution such that we understand how a star appears at a given epoch. Complicating matters are binary star systems which can lead to star-star interactions, altering each star's evolution from the single star behaviour. There also needs to be an understanding of the structure

and dynamical evolution of the Galaxy as a function of position and time. Finally, we must also understand the design of the detector, telescope and observing strategy adopted for a particular survey used, such that we can then understand the biases introduced in to the observed objects.

### 1.1.1 Stellar evolution

Stellar evolution itself has been studied for many years, from the earliest ideas on the structure of stars (Schuster, 1903) until the present day 3D hydrodynamical models (Heger et al., 2000; Woosley et al., 2002; Magic et al., 2013). See Lebreton (2000) for a review of stellar evolution.

Briefly a star starts of as a molecular cloud that collapses in on itself to form a pre-main sequence star (Palla & Stahler, 1993). For stars more massive than $\sim 0.07 M_{\odot}$, the star undergoes further collapse until eventually igniting hydrogen fusion in the core, for less massive stars fusion only occurs for a short period of time and is not self-sustaining. At this point it then evolves along its main sequence (MS) track, burning hydrogen (fig. 1.1 red region), until the core contains a significant fraction of helium and transitions to its hertzsprung gap (HG) phase (fig. 1.1 yellow region). The star then evolves onto its giant branch (GB) by undergoing hydrogen burning in a shell around the core (fig. 1.1 green regions) (Iben, 1967). This requires a star to have $M \gtrsim 0.8 M_{\odot}$, otherwise there has been insufficient time since the universe began for the star to have evolved off the main sequence. Eventually the core's temperature and pressure increases untill helium can fuse, forming a core helium burning (CHeB) star (fig. 1.1 cyan region) (Boothroyd & Sackmann, 1988). In time the helium in the core stops fusing and the star forms shells of burning hydrogen and helium surrounding an inert electron-degenerate core, on the asymptotic giant branch (AGB) (fig. 1.1 purple region).

Eventually, if the star is massive enough ($M \gtrsim 8 M_{\odot}$) it can start fusing carbon and the heavier elements in the core. Depending on the initial parameters of the star we can then either end up with a planetary nebula (PN) (fig. 1.1 blue region),

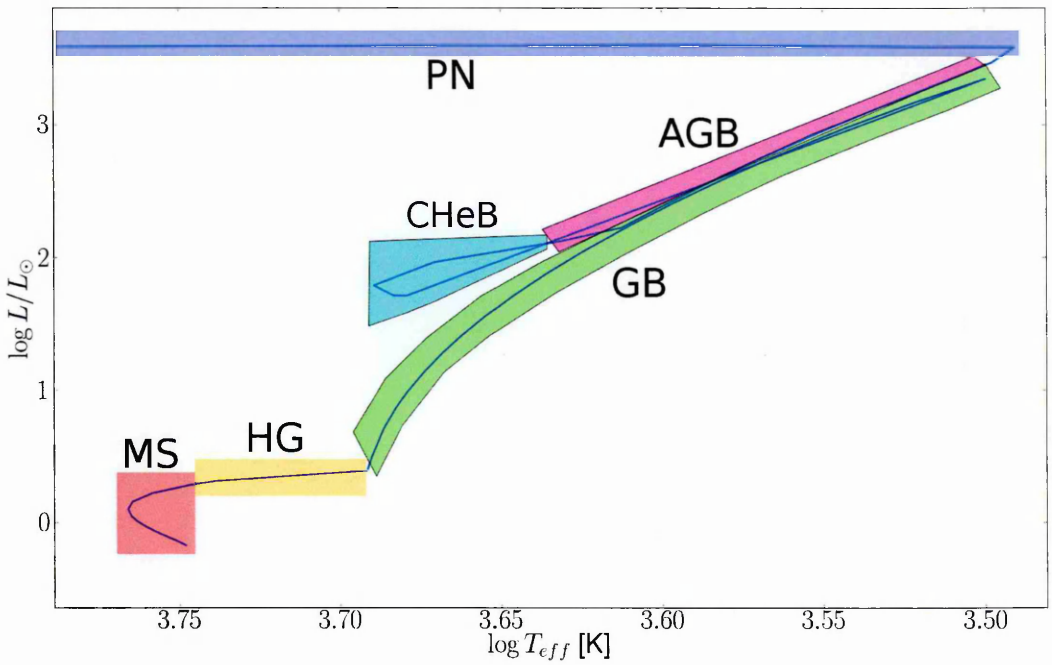



Figure 1.1: Stellar evolution track of a $1M_\odot$, $z = 0.02$ star. Each stage is colour coded and shows their approximate location; red, main sequence; yellow, hertzsprung gap; green, giant branch; cyan, core helium burning; purple, asymptotic giant branch and finally blue, planetary nebula. The final possible outcomes of a star (white dwarf, neutron star and black hole) are not shown.

where the outer layers of the star have been expelled leaving the core of the star behind, which will become a white dwarf (WD) (Kennicutt, 1984) for stars with $M < 8M_\odot$. Or for more massive stars, a supernova leading to a neutron star (NS) ($8 \lesssim M/M_\odot \lesssim 20$) or a black hole (BH) (Lattimer & Prakash, 2001; Shapiro & Teukolsky, 1986). The boundaries between the different final outcomes depends upon the final parameters the star has at the AGB phase (Heger et al., 2003).

Binaries can complicate this picture because of potential interactions between between the components (Iben & Tutukov, 1984). Binary evolution can be loosely split into two pathways: those where both stars have sufficient separation that they do not interact and thus evolve as separate stars (Hurley et al., 2002); and those that do interact (Eggleton, 2006). Those that do interact can transfer mass in two ways either by losing mass due to stellar winds, which can then be partially accreted onto the companion or when one star's radius exceeds its Roche lobe and material flows from one star onto the other. The size of a Roche lobe is defined as the radius of

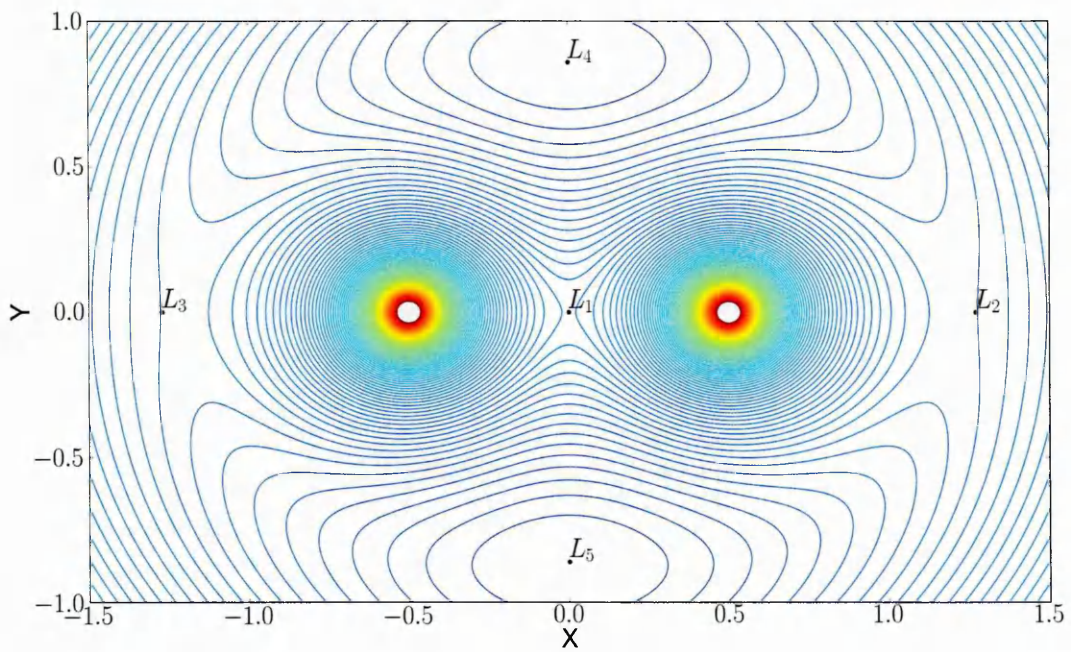



Figure 1.2: Contours showing lines of equipotential for two equal sized components, where each star sits at $(\pm 0.5, 0.0)$. Each of the five Lagrangian points is shown, where the gravitational and centrifugal forces balance between the two masses.

a sphere that has the same volume as the interior of the equipotential that passes through the first Lagrangian (L1) point between the two stars (Eggleton, 1983). There are five Lagrangian points in a two body + negligible mass third body, as seen in fig. 1.2. They represent the locations where the gravitational and centrifugal forces between the two stars balance. A test particle at the L1-3 points exist in a quasi-stable orbit, such that small perturbation to the test particle would move the particle away from the L1-3 point. While at L4 and L5 a test particle would be a in a stable orbit, such that a perturbation would result in the particle moving back to the L4 or L5 point . Considering only the first point, L1, any material that orbits interior to this point is bound to its parent star. However, once it passes the L1 point the gravitational pull of the companion star exceeds that of the parent star and the material is now bound to the companion. Thus if the star expands and its outer envelope orbits beyond the L1 point then the envelope will become bound to the companion, transferring mass between the stars.

During a mass transfer process, the angular momentum of the material may be large enough that it cannot fall directly onto the companion star. Instead the mate-

rial will orbit the star forming an accretion disk (Pringle, 1981). In this disk most material will lose angular momentum via frictional forces and fall onto the star, while some material will gain angular momentum and thus orbit at a larger radius, conserving overall angular momentum (Pringle, 1981). This process will alter the evolution of the two components, rejuvenating the receiving star while slowing the evolution of the donor (Hurley et al., 2002). A stars response to mass transfer depends on the mass ratio of the system and the structure of the stars. Stars with convective envelopes expand when they lose mass, which can lead to dynamical time-scale mass transfer which is unstable (Ivanova et al., 2003). While stars with radiative envelopes contract with mass loss, which can lead to stable mass transfer.

If the mass transfer process is unstable it can lead to a common envelope (CE), where the secondary star ends up orbiting inside the envelope of the donor star. Due to viscous drag, both stars will spiral rapidly inwards heating the CE (Iben & Livio, 1993). There may be two outcomes from this: the second star may undergo enough braking such that it merges with the donor star's core leaving only one star behind; or the envelope may gain enough energy to overcome the binding energy and escape the system. This will leave the core of the donor star behind and the binary system itself in a much shorter orbital period (Schreiber & Gänsicke, 2003). Predominately most binaries do not interact, however Willems & Kolb (2004) suggest that for systems that form WD+MS binaries, that between 10%-25% of systems will transfer mass via Roche lobe overflow or a common envelope phase.

#### 1.1.1.1 Braking forces

The interactions between the two stars can also lead to a series of braking forces which will act to circularise the orbit and to decrease the orbital period.

Using general relativity to solve the motion of a binary system, we find that system will emit gravitational waves (Taylor & Weisberg, 1982). These waves will take energy and angular momentum from the system making the binary stars spiral inwards. The change in the orbital properties is given by (Eggleton, 2006):

$$\frac{\dot{P}}{P} = \frac{3\dot{a}}{2a} = -\frac{3}{t_{GR}} \frac{1 + \frac{73}{24}e^2 + \frac{37}{96}e^4}{\left(1 - e^2\right)^{(7/2)}}$$

$$\frac{\dot{e}}{e} = \frac{-1}{t_{GR}} \frac{\frac{19}{6} + \frac{121}{96}e^2}{\left(1 - e^2\right)^{(5/2)}}$$

$$t_{GR} = 376.8 \frac{(1+q)^2}{q} P^{8/3} M^{-5/3} \qquad \text{(Gyrs)} \tag{1.1}$$

where $P$ is the orbital period, $\dot{P}$ is the change in period, $e$ is orbital eccentricity, $\dot{e}$ is the change in orbital eccentricity, $a$ is the orbital separation and $\dot{a}$ is the change in orbital separation, $t_{GR}$ is the time-scale and $q$ is the mass ratio. It can be seen that gravitational radiation circularises the orbits of the binaries and decreases the orbital separation.

In close binary system tides generated by one star upon the other will cause an exchange between orbital and spin angular momentum. This only occurs if either star has a spin rate not equal to the orbital period, a spin axis not aligned with the orbital plane or if the system is in an eccentric orbit (Hut, 1981). Tides will then produce a torque that acts to circularise the orbit, lock the spin of the stars to the orbital period and align the spin and orbital planes. This is only stable if more than 3/4 of the total angular momentum is in orbital angular momentum. This configuration minimises total energy of the system. If it is not stable, then the stars may spiral inwards at an accelerated rate (Hut, 1980).

Stars are continually emitting charged particles via a stellar wind, these particles become trapped in the magnetic field of the star, forcing the charged particles to rotate at the same rate as the field (Mestel, 1968). As the field is dragging these particles round, the star transfers its angular momentum to the charged particles, extracting angular momentum from the star and slowing its spin down. In a binary system if there is spin-orbit coupling via tides, then this will remove angular momentum from the whole system forcing the binaries to spiral inwards (Parsons et al., 2010).

Thus we can see that the two stars in a close binary can be significantly per-

turbed during their evolution, from where we would expect them to be had they been two separate stars not in a binary. We can also see how the orbital period and eccentricity of the system evolves with time, which needs to be considered in stellar evolution.

### 1.1.2 Population synthesis

Population synthesis models are used to translate the evolution and distribution of stars into observables we can measure, like colours and spectra of stars. There are many models in the literature (Robin et al., 2003; Girardi et al., 2005; Sharma et al., 2011) and a large effort is spent in using population synthesis tools to model galaxies and clusters (Bruzual & Charlot, 2003; Belczynski et al., 2008). However for this work I am interested in the population models that deal with the Milky Way. Here we have much more information available and thus the models can become more complex and more terms considered. The aim of these models is to produce a synthetic catalogue of stars for a given survey, that can then be used to test theories of stellar and Galactic evolution (Boeche et al., 2013; Gao et al., 2013).

Generally, these models are based on a chosen set of stellar isochrones, describing a star's evolution as a function of time. These isochrones are then weighted by a set of initial distribution functions (IDFs) describing the initial parameters of the star and the parameters of the Galaxy (Vanhollebeke et al., 2009). The star is then evolved up to the current epoch and its current magnitude computed. The star can then be tested against a set of visibility constraints, for instance apparent magnitude, spatial location and colours, to determine whether the chosen survey will observe the star.

The Galactic structure is parameterised as a series of discs. Each disc in turn has a series of parameters describing the structure of the disk and the compositions of stars that form in that disc. For instance; the initial mass function (IMF), describes how mass is distributed between stars; the initial binary fraction distribution (IBFD) describes how likely a binary will form; and the star formation rate (SFR) describes

how many stars form each year. These parameters may have temporally and/or spatially varying characteristics; though to simplify matters some parameters may be in common between the discs, such as assuming the IMF is constant. There are many parameters involved, not of all of which are well constrained. The aim therefore is to create a set of self-consistent distributions, that may vary between models but that reproduce the Galaxy as we observe it.

The standard set of parameters used are the Galactic density, star formation rate, IMF and metallicity distributions. Though there may be others, for instance the binary parameters: initial mass ratio distribution (IMRD), initial orbital separation distribution (IOSD) and initial eccentricity distribution (IED). See section 2.1.4 for further details and the preferred choices for this work. The Galactic density is usually taken as a double exponential with a set of scale heights (Bahcall & Soneira, 1980; Dehnen & Binney, 1998). The star formation rate in the Galaxy has some uncertainty to it, versions proposed in the literature vary from an approximately constant rate (Twarog, 1980) to a time varying distribution (Noh & Scalo, 1990). The IMF has a series of possible choices in the literature, log-normal, exponential, and power law (See Chabrier (2003) for a review). Finally the metallicity distribution has been modelled as a Gaussian (Nordström et al., 2004), with different parameters for the different discs (Gilmore & Wyse, 1985). Also included can be an age-metallicity relation (Holmberg et al., 2007), with younger stars being more metal rich.

To model the current population of stars in a field of view we should also consider the stars kinematics and possible radial migration of stars inside the Galaxy (Chiba & Beers, 2000). Here stars which are orbiting inside the Galaxy can migrate from the region they where formed to other regions of the Galaxy. (Haywood, 2008). As these stars migrate into a different region of the Galaxy, they "mix" different populations of stars together. Thus the distribution of objects we detect today, inside a volume element of the Galaxy, may include objects that formed elsewhere from a different set of formation distributions. This process is beyond the scope of this work and not considered further.

Population synthesis models have found many applications. They include constraining stellar evolution quantities such as showing that Cepheids require an enhanced mass loss process (Neilson et al., 2012), determining the IMF at low masses (Schultheis et al., 2006) and a requirement for a faster rate of cooling in low mass WDs than theory predicted (Nelemans et al., 2001). Other applications include predictions of Galactic structure parameters for instance; metallicity gradients in the thick disc (Kordopatis et al., 2013b), formation models of the Galactic bulge (Uttenthaler et al., 2012) and kinematics of the thin disc (Gazzano et al., 2013). This is an area that will be revolutionised by the GAIA mission (Perryman et al., 2001), which will map the location and velocities of $10^9$ stars in this Galaxy.

The treatment of binaries in population models varies. Some authors ignore them (Robin et al., 2003), some treat them in a simplified manner assuming they are simply two single stars (Girardi et al., 2005), while others look at the full evolution (Willems & Kolb, 2002). However, as shown previously, this evolution can be complex and the resulting distribution of binaries is not the same as the distribution of two single stars that are non interacting. Binaries have also been shown in Raghavan et al. (2010) to make up a significant fraction (35% are binaries with a further 15% in higher order systems) of the Galactic population. Thus they can not be ignored.

### 1.1.3 Observational biases

When we observe the Galaxy, we do not observe every star possible, instead we view a biased sample of the Galaxy. Unless we understand how theses biases are introduced, we are unable to translate the statistics of the stars observed into the statistics of the stars of the Galaxy.

There are limits on the flux we can detect from a star: too bright and it will saturate the detector, too faint and we will not resolve it above the noise. Considering the larger of the two issues, namely stars being too faint, the star can either intrinsically faint or it can be distant relative to us, and thus it appears faint. Though of course

both cases are coupled, but we consider them separately for the moment. There is a bias inherent in the fact that we are more likely to observe intrinsically bright stars, as they can be seen over a greater volume of the Galaxy and as a result the sample we observe is not completely representative of the Galaxy as a whole (Butkevich et al., 2005). Instead it favours stars with high luminosities, which will in general imply high mass and/or giant stars. These systems can also be seen to greater distances, thus they probe a different region of the Galaxy compared to the intrinsically fainter objects, which can only be seen closer to the Sun (Gaidos & Mann, 2013).

Observers can also intentionally bias their data, either in the choice of objects observed or in the choice of objects that are then subjected to analysis. For instance, observing in the infra red biases the detection towards cooler, low mass stars (Chabrier et al., 2000), while applying colour cuts introduces biases to certain populations of objects. The cuts are usually intentionally designed to target certain types of stars that are of interest and can provide benefits by reducing the number of "uninteresting" stars, but nevertheless they bias the results.

The choice of field will also bias the observed population. The closer the field is to the plane and the centre of the Galaxy, the more metal rich and younger the stars are (McWilliam, 1997). Observing further from the plane implies that, as a fraction of the total stars observed, the thick disc will increase its contribution (Jurić et al., 2008), even as the total number of stars decreases, thus appearing more metal poor. Depending of the choice of field of view we may be observing certain features in the Galaxy, such as a cluster or a tidal stream (Ivezić et al., 2008), such that the field is not representative of the Galaxy as a whole. Thus the choice of field may bias the observed population's age, metallicity and kinematics (Bond et al., 2010).

To be able to recover the Galactic parameters we must be able to model the biases introduced in the course of an observation. Then we can apply the same set of biases to a chosen population model and determine the goodness of fit to the data. This then allows us to extrapolate the results from a limited survey to the whole Galaxy. For instance, taking the distribution of observed (and modelled) WDs we

can work backwards, with a population model, to determine the star formation rate of the Galaxy (Weidemann, 1990).

## 1.2 Kepler

Kepler was a NASA space-based telescope aimed at detecting transiting Earth-mass planets in the habitable zone of solar like stars, to determine their frequency, $\eta_{earth}$, (Borucki et al., 2010). After 4 years of observing $\approx$ 150,000 stars in the constellation of Cygnus, two reaction wheels failed ending its original mission. During its observational period it has made many discoveries, including 961 confirmed planets (Rowe et al., 2014), ~ 2500 candidate planets (Burke et al., 2014) and several circumbinary planets (Holman et al., 2010; Doyle et al., 2011). Here I present a brief review of the Kepler mission and its results.

Kepler has a 1.4m optical mirror with a field of view of ~ 100 square degrees (Koch et al., 2010). It observed the same of field of view continuously during its 4 year mission, rotating itself every 3 months to align the solar panels with the Sun. These 3 month blocks are defined as a quarter. As of this date 17 quarters worth of data have been released to the public. Kepler has two observing modes, short cadence (SC) with a 60 second exposure and long cadence (LC) which is the sum of 30 SC observations, making a total exposure of 30 minutes (Van Cleve & Caldwell, 2009). It observes in its own filter band, $K_p$, which is an optical band approximately over the range 423 to 897 nm (Koch et al., 2010).

Kepler detects planets by observing the star, continuously, for changes in the flux (Borucki et al., 2011a),. These changes could be due to a transiting exoplanet where a planet passes in front of the star, from our point of view, blocking some of the flux emitted from the star (Charbonneau et al., 2000), as seen in fig. 1.3. Based on the depth and shape of the transit we can then infer the ratio of the radii of the star and planet. The planet's radius can be determined via:

$$\frac{\Delta F}{F} = \left(\frac{R_p}{R_*}\right)^2 \tag{1.2}$$

where $\frac{\Delta F}{F}$ is the maximum change in flux during the transit, $R_p$ is the planet's radius and $R_*$ is the host star's radius which is determined via stellar models (Borucki et al., 2011b). However, detecting a transit alone does not imply that it is a planet (Collier Cameron et al., 2007). Only by follow up work to determine the mass of the object via radial velocities can we unambiguously say that the object is a planet (Léger et al., 2009).

There are several astrophysical phenomenon that cause a change in the light curve. These include: stellar variability, which adds extra noise to the light curve possibly burying a signal; star spots, which act as a dark spot that appears as a transit-like signal; and eclipsing binaries, which can mimic a transit signal (see section 6.3), each of which would could give a false positive signal (Morton & Johnson, 2011; Torres et al., 2011). Thus Kepler has declared its discovery of many *candidate* planets, which have periodic transit-like signals, reserving planet discoveries for those systems that been confirmed with further observations. Due to the faintness of most of Kepler's targets, the majority of the candidate planets are difficult to follow up and confirm their planetary status (Batalha et al., 2010). Kepler also introduced a new term of *validated* planets (Fressin et al., 2011). These planets have not be confirmed as a planet but via modelling the false positive rate derived that is below a threshold such that they are confidant it is a planet. This is done either with BLENDER (Torres et al., 2011) which models all possible eclipsing binary (EB) scenarios, or for systems with multiple transits, showing that the chance alignment of multiple false positives is negligible (Rowe et al., 2014; Lissauer et al., 2014).

Due to bandwidth limitations Kepler did not observe every star in its field of view down to its magnitude limit. At $K_p < 16$ there are $\sim 450,000$ stars in the field, however it only observes 150,000 (Batalha et al., 2010). Thus a target selection process was employed to select those objects for which they had the highest chance of detecting an Earth-sized planet in the habitable zone of the star. The selection process had several aims when determining which stars to observe. Primarily it was aimed at distinguishing between dwarf and giant stars. This is because the transit depth is

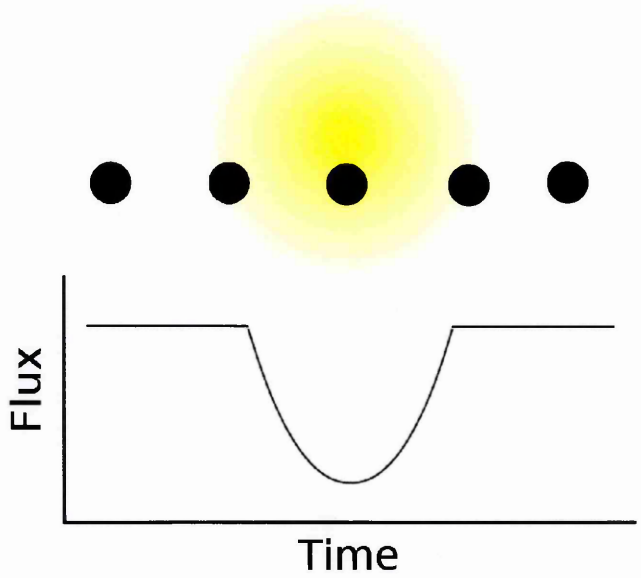



Figure 1.3: Schematic light curve showing the effect of a transiting exoplanet on the total flux output of a star. As the planet passes in front of the star it blocks a fraction ($\Delta F/F$) of the stars flux.

proportional to the radii ratio of the planet and star (equation 1.2), thus it is easier to detect a transiting planet around a small star. There was also a selection of stars which had a lower background contamination, as background binaries can lead to false positives (see section 6.3). This process biases the sample of stars Kepler observes to low mass dwarf stars (Gaidos & Mann, 2013). Thus the Kepler results, like $\eta_{earth}$, must have a correction factor applied for the fact that Kepler's sample of stars is not representative of its field (Howard et al., 2012).

Outside of the primary mission of detecting transiting exoplanets, Kepler has made many contributions to other areas of astrophysics. It has discovered over 2000 EBs (Prša et al., 2011; Slawson et al., 2011), a number of which are triple star systems (Steffen et al., 2011; Rappaport et al., 2013) and even some that are triply eclipsing systems (Carter et al., 2011). Kepler has also discovered a number of binaries with tidally induced pulsations (Welsh et al., 2011; Thompson et al., 2012).

It has also performed asteroseismology on many of its target stars (Gilliland et al., 2010a) with the aim of determining precise stellar parameters (Chaplin et al., 2011c). Here it aims to detect the individual pulsation modes a star can undergo, by detecting the small change in flux due to changes in the surface temperature of the

star due to a pressure wave (Kjeldsen & Bedding, 1995). For main sequence stars, these pulsations modes are of order minutes thus the SC data is needed, which limits its application as there where a limited number of SC slots available each quarter. However, over 500 MS stars have had their pulsation modes mapped and had their stellar parameters derived to < 10% (Chaplin et al., 2014). For evolved stars the pulsation modes are of order hours thus the LC data can be used (Bedding et al., 2011). The difference in pulsation periods is due its dependence on $1/ < \rho >$, where $< \rho >$ is the average density of the star (Chaplin & Miglio, 2013). Kepler has detected solar like oscillations in red giant stars (Bedding et al., 2010), for which over 13,000 now have measured parameters (Stello et al., 2013).

In Chapter 2 I discuses the population model used and the improvements made to it. Chapter 3 discuses the model I implemented to derive the Kepler target list and then I discuss the biases this process introduces into the Kepler results. Chapter 4 details the model of eclipsing binaries I derived and the results of a comparison of this model with the Kepler EB catalogue. In Chapter 5 I derive a model for asteroseismic binaries and make predictions for the number of detectable asteroseismic binaries in the Kepler data set. Finally in Chapter 6 I present my final conclusions and discuss where this work leads too.

# Chapter 2

# Population synthesis

To place the result of a given survey into context we need to compare their measured distributions of stellar properties, which are biased by their observing strategy, with synthetic models. These models can have the same bias applied, then we we can then compare different underlying Galactic distributions and compute a best fit to the biases observed data. These synthetic models therefore need to take into account all relevant physics, from stellar evolution, Galactic evolution and observational constraints to the length of observations, filter band observed and selection process of stars chosen to be observed.

In this chapter I derive the a population of synthetic stars that can then be in future chapters, biased to the Kepler observing strategy. However the work in this chapter is independent of the Kepler field. Its aim is to generate a synthetic catalogue of stars, representative of the Galaxy, for a given observing strategy. This then forms the basis for the work in the other chapters.

To derive a synthetic population of stars, I combine stellar evolution with a Galactic model. I derive a series of evolution tracks for single and binary star systems (section 2.1.1). These tracks are then weighted, at different points in their evolution, by the formation probability of the system and the probability of detection (sections 2.1.2-2.1.6). This forms a continuous distribution of stars inside the Galaxy. This continuous distribution can then be sub-sampled to form a representative catalogue of individual synthetic stars for a given observing strategy (section 2.1.7).

Finally, in section 2.2 I describe the model for describing eclipsing binaries (EBs).

## 2.1 Population synthesis model

To calculate a model for the stellar and binary star population in a field-of-view I added new input physics and functionality to the Binary and Stellar Evolution Population Synthesis (BiSEPS) code. This was originally described in Willems & Kolb (2002, 2004) and later employed by Willems et al. (2006) in a simplified way to study the false positive rate in the exoplanet transit search project SuperWASP (Pollacco et al., 2006) from shallow-eclipsing binaries. BiSEPS in turn is based on the analytical descriptions of stellar and binary evolution by Hurley et al. (2000) and Hurley et al. (2002). This is linked to a comprehensive Galactic model including formation probabilities, Galactic disc structure, interstellar extinction and an assumed star formation rate (SFR). Taken together this derives a representative synthetic catalogue of single and binary stars visible in the Galaxy today.

### 2.1.1 Binary evolution

At the core of the population synthesis scheme is a large library of single star and binary system evolutionary tracks from the ZAMS up to a maximum age of 13 Gyrs. Which provide physical parameters of each system over approximately 100 time steps distributed along each track.

Briefly, the single star evolution (SSE) of Hurley et al. (2000) code uses analytic functions to describe the evolution of the stellar core mass, radius, luminosity and lifetime during each evolutionary phase for a given initial mass and metallicity. Mass loss is included in the form of stellar winds (Hurley et al., 2000). The star is evolved for a time step $\mathrm{d}t$ which is decided by the code to minimise the change in the mass and radius of the star during each time step.

The binary star evolution (BSE) code of Hurley et al. (2002) handles the evolution of binary systems by evolving two single stars with the same $\mathrm{d}t$ and then calcu-

lating the interaction effects between the two stars. As each star may have a different $dt$ as calculated by the SSE code, BSE can roll back each star's evolution and try with a smaller time step, until a point where both stars agree on a $dt$ that minimises the change in their properties. In calculating the interaction between the systems several factors are considered; possible roche-lobe overflow (RLOF), transferring material between each star; magnetic and gravitational braking, acting to decrease the orbital separation; and tidal effects, acting to circularise the system. Interactions are calculated for all binaries, though for most binaries, which are in wide orbits, these effects are negligible. For this work I ignore tidal effects by forcing binaries to be in circular orbits at all times and this will be discussed why, in more detail later. As these effects alter the orbital angular momentum, and are themselves dependant on the orbital angular momentum, limits are placed on its change, however these braking forces and the mass transfer process are only important for systems with short orbital periods which interact.

The evolution code outputs the global properties of each star (mass, radius, temperature, luminosity, evolutionary phase) and of the system (orbital separation, age) at a series of points in time along the evolutionary track. To prevent excessive data output the code was modified to output only a time-weighted averaged value of several time steps. The number of steps averaged together is selected from a uniform random number between an upper and lower bound. For single star systems the limits are [1,25] steps for all evolution phases. For binaries the following limits apply, the lower limit is itself randomly uniformly selected from [1,25] unless the star is either a hertzsprung gap (HG) or has just transitioned into a new evolutionary phase, in which case the limit is 1. The upper limit is dependant on the evolutionary phase of the system, main sequence (MS) 100, HG 25, giant branch (GB) 50 and all others are set to 500 steps. For the case where each star is a in a different evolutionary phase the minimum limit is chosen. These limits are set based on trial and error determinations designed to allow sufficient temporal resolution to model the individual features in a evolution track while also minimising the data output.

The coarser sampling of the binaries is due to the increased initial parameter space required to model the systems. The singles stars and binaries can be described by their initial primary mass and metallicity, however the binaries also need the secondary mass and initial orbital separation. This increases the number of evolution tracks required. Thus the sampling rate is lowered for the binaries to prevent excessive data output. For the binaries we must also consider that the RLOF phase and asymptotic giant branch (AGB)/thermally pulsating asymptotic giant branch (TPAGB) phase (for which both stars may pass through), require high temporal resolution. This is due to the mass loss requiring a smaller $\mathrm{d}t$, which would then require more date to be outputted.

A second set of perturbations was also introduced in the determination of $\mathrm{d}t$ itself. The time step was multiplied by a uniform random number in the range [0.5,1]. This was to prevent aliasing effects when viewing a Hertzsprung-Russell diagram (HRD). This occurs when multiple evolutionary tracks, for different initial masses, form lines roughly perpendicular to the direction of stellar evolution. This is due to each track being sampled at the same relative time in their evolution. Thus this randomisation process aims to perturb where each sample is taken to prevent this aliasing effect.

Fig. 2.1 shows what happens without the randomisation, where we can see different evolutionary tracks being sampled at similar places in their evolution. While fig. 2.2 shows the effect of the randomisation process. We can see the effect of the randomisation is most pronounced during the HG phase, which is a short-lived evolution phase. Thus each $\mathrm{d}t$ step makes up a larger fraction of the evolutionary phase and that there are fewer time steps.

This was not considered in previous versions of BISEPS as those models where only concerned with the distribution of stars over the whole of the parameter space, where the individual placement of each star does not matter. However, for this work where I wish to consider a discrete sub samples of stars (see section 2.1.7). Where it is now important that the distribution of each star is realistically distributed over

the parameter space.

#### 2.1.1.1 Initial parameter space

A newly forming binary system is taken to be fully characterised by the initial masses of its components, the orbital separation, and the star's chemical composition, set here with hydrogen abundance $X = 0.70$ and metallicity ($Z$) (I consider $Z = 0.020$ and $Z = 0.0033$). All systems start with and are forced to have, during their evolution, circular orbits.

The initial parameter space for binaries is divided into 50 logarithmically spaced equidistant bins of initial masses $M_1$ and $M_2$ between 0.1 and $20M_\odot$ and into 250 logarithmically spaced equidistant bins of initial semi-major axes $a$ between $3R_\odot$ and $10^6 R_\odot$. By symmetry, only objects where $M_1 \geq M_2$ are evolved. I do not consider the initial eccentricity distribution (IED) of binaries at the moment due to computational reasons. The inclusion of an extra initial distribution increases the runtime and storage requirements exponentially.

Single star tracks are obtained from the primary star tracks in very wide, non-interacting binaries (with initial $a = 10^7 R_\odot$ and $M_2 = 0.1M_\odot$). The single star parameter space is divided into 10000 logarithmically spaced equidistant bins of initial mass between 0.1 and $20M_\odot$. Where the lower limit is set by the SSEs lower mass limit and the upper limit is set by the assumption that there are very few very massive stars.

### 2.1.2 Galactic model

Underpinning the spatial distribution of the synthetic stars calculated is a simple kinematic model of the Galaxy, described in detail in Willems et al. (2006). The Galaxy is assumed to comprise a young thin disc and an older thick disc. Each discs number density is modelled as a double exponential of the form:

$$\Omega(R,z) = n_o \exp\left(\frac{-R}{h_R}\right) \exp\left(\frac{-|z|}{h_z}\right) \tag{2.1}$$

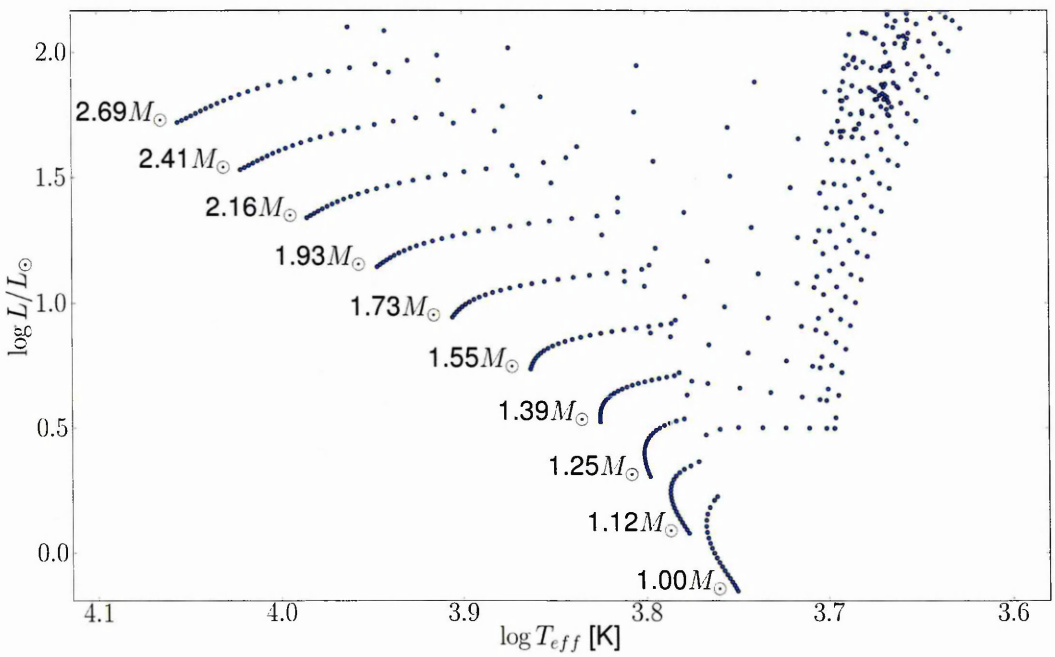



Figure 2.1: A HRD showing a series of evolutionary tracks for solar metallicity objects when there are no perturbations in the size of the time steps (d$t$) or number of bins averaged together

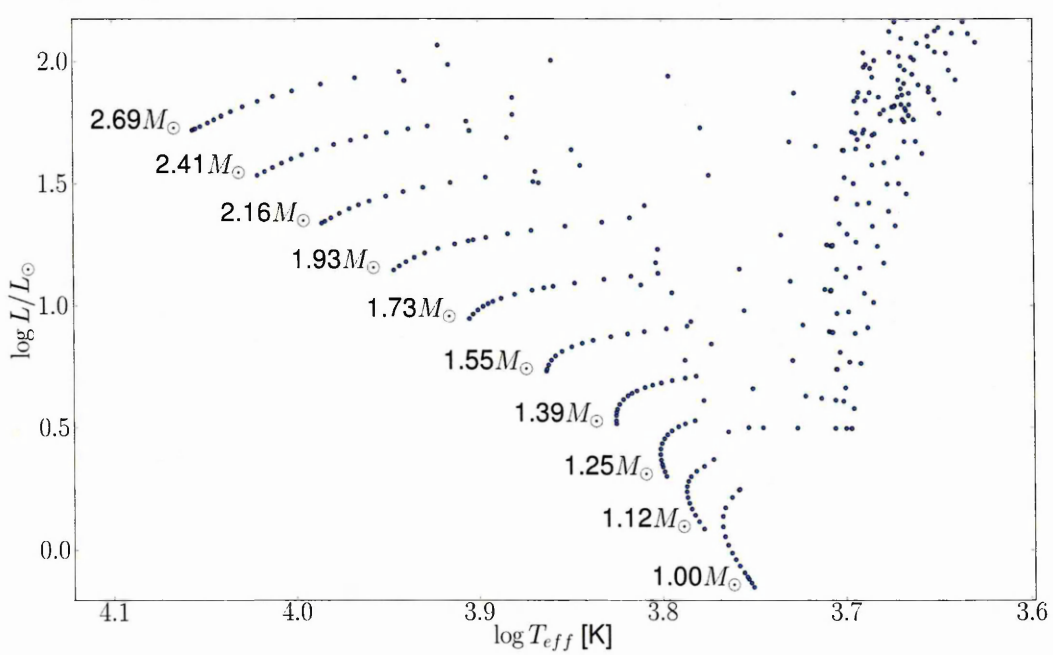



Figure 2.2: Same as fig. 2.1 except now d$t$ and the number of points averaged together is randomized.

with $h_R = 2.5$ kpc and $h_z = 300$ pc for the thin disc and $h_R = 3.8$ kpc, $h_z = 1$ kpc for the thick disc (Gilmore & Reid, 1983; Du et al., 2003). The integral is normalised to unity, thus $n_o = 1/4\pi h_R^2 h_z$.

Star formation is assumed to proceeded for the first 3 Gyrs after the formation of the Galaxy in the thick disc at a constant rate. The age and star formation rate of the thick disc is ill-determined with assumptions of a star burst 8 Gyr ago (Ibukiyama & Arimoto, 2002) or a variable rate over the first several Gyrs (Chiappini et al., 1997). Due to this uncertainty, I choose a flat SFR for simplicity (Twarog, 1980). The thin disk is then assumed to form stars for the next 10 Gyrs after the thick disc has stopped forming stars (Wood & Oswalt, 1998).

During the respective star forming periods the SFR is taken to be constant, such that one star or binary with component mass $M > 0.8M_\odot$ is produced per year (Weidemann, 1990). Combining the SFR with the choice of IMF (see section 2.1.4) leads to a formation rate of 7.6 stars $\mathrm{yr}^{-1}$.

Each system, at each point of its evolution, is weighted by the integral of the SFR between the minimum and maximum age of the system such that it is alive in the current epoch. For instance if at a point in the stars evolution takes 2 Gyrs to evolve to its current state and then stays in this state for a further 1 Gyr, then we integrate the SFR between 2 and 3 Gyrs ago. As anything that forms more than 3 Gyrs ago will no longer be in this phase and anything that forms less than 2 Gyrs ago will not have had sufficient time to evolve in this state.

To capture the essence of the metallicity evolution with Galactic age I go beyond Willems et al. (2006) and assume that thick disc stars have a metallicity $Z = 0.0033$ (Gilmore et al., 1995) while stars forming in the thin disc have a solar metallicity value of $Z = 0.020$ (Haywood, 2001). There is evidence for age-metallicity relation and Galactic radius-metallicity relation in the discs (Nordström et al., 2004). However as I am computationally limited to a few metallicities I decided to stay with a constant metallicity in each disc. Evolving these stars with a lower metallicity, creates a population of stars that both age quicker and are bluer compared to the solar-

like population.

To obtain the total number of systems in a given survey field the stellar density, equation 2.1 is numerically integrated over Galactic longitude, latitude and distance $(l, b, d)$ by translating it from galactocentric $(R, z)$ to heliocentric coordinates $(l, b, d)$ via:

$$R = \left(d^2 \cos^2 b - 2dR_\odot \cos b \cos l + R_\odot^2\right)^{\frac{1}{2}}$$

$$z = d \sin b + z_\odot, \tag{2.2}$$

where $R_\odot$ = 8.5kpc is the radial distance of the Sun from the Galactic centre (Reid, 1993) and $z_\odot$ = 30pc is the height of the Sun above the Galactic plane (Chen et al., 2001).

For each system the integral over distance is carried out between the minimum and maximum distance, $d_{min}$ and $d_{max}$, this system can be detected to by the survey I wish to model. If the survey is magnitude limited these are determined by:

$$d = 10^{(m-M+5-A_\lambda)/5} \tag{2.3}$$

Where $m$ is the lower or upper magnitude limit of the survey, $M$ the absolute magnitude of the system and $A_\lambda$ is the extinction along a line of sight at $(l, b)$ integrated between $[0, d]$ in the filter band of the survey.

As the extinction itself depends on the distance, I calculate the distance limits for the integral iteratively. Rearranging equation 2.3 in terms of the apparent magnitude I can then perform a bisection search over the distance, recomputing the extinction as needed, to find where the apparent magnitude is within 0.1% of the desired value.

I upgraded the BISEPS extinction routine described in Willems et al. (2006) which is based on the model proposed by Hakkila et al. (1997) to that of Drimmel et al. (2003). This calculates the Galactic extinction from a 3D dust model of the Galaxy that has been scaled using data from the COBE/DIRBE NIR instrument to provide extinction values along lines of sight in the $V$-band (see also section 2.1.5

below).

Fig. 2.3 shows the extinction map at 1Kpc for Hakkila et al. (1997), while fig. 2.4 shows it for Drimmel et al. (2003). The most striking difference is the improvement in spatial resolution of fig. 2.4 compared to fig. 2.3. As Hakkila et al. (1997) map was made of a combination of other authors maps the resolution varies over the whole sky, leading to the blockiness seen in fig. 2.3. I also found that along several lines of sight the extinction would decrease, likely a bug where different maps had not been combined properly. Drimmel et al. (2003) allows the use of a much higher resolution, based on the size of the COBE pixels, of $0.35^o \times 0.35^o$ over the whole sky, which reduces the blockiness. However Drimmel et al. (2003) has limitations in that the model has limited applicability along the Galactic plane and in the Galactic centre. This is due to the dust model not including the central bulge as well as uncertainties in how far the spiral arms extend into the centre (Drimmel et al., 2003). This however is not an issue for the moment as Kepler does not observe the plane or the centre of the Galaxy.

Kepler's field of view, for the primary mission, is at $l = 75° \pm 8°$ and $b = 13° \pm 8°$, covering ~ 100 square degrees. This field can be seen on highlighted in figs. 2.3 & 2.4. The choice of field was driven by the need to maximise the number of stars observed while at the same time minimising the number of bright giants present in the Galactic plane from being observed (Van Cleve & Caldwell, 2009). It can also be seen in figs. 2.3 & 2.4 that the choice of field for Kepler places it in a region of low extinction.

### 2.1.3 Field of view

The integration boundaries for Galactic longitude and latitude in equation 2.1 are determined by the location of the field of view of the detector system. For practical reasons I define integration regions bounded by lines of constant $l$ and $b$ that enclose the region.

The extinction in equation 2.3 is then computed over several regions, typically

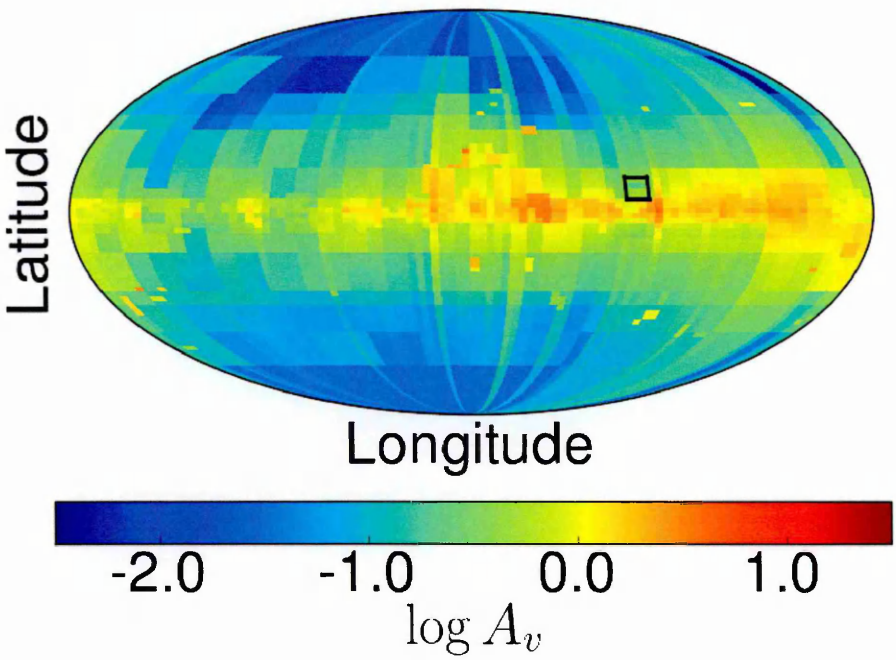



Figure 2.3: A map of the all-sky extinction ($A_v$), using Hakkila et al. (1997), out to a distance of 1 Kpc. Kepler's field of view is denoted in black

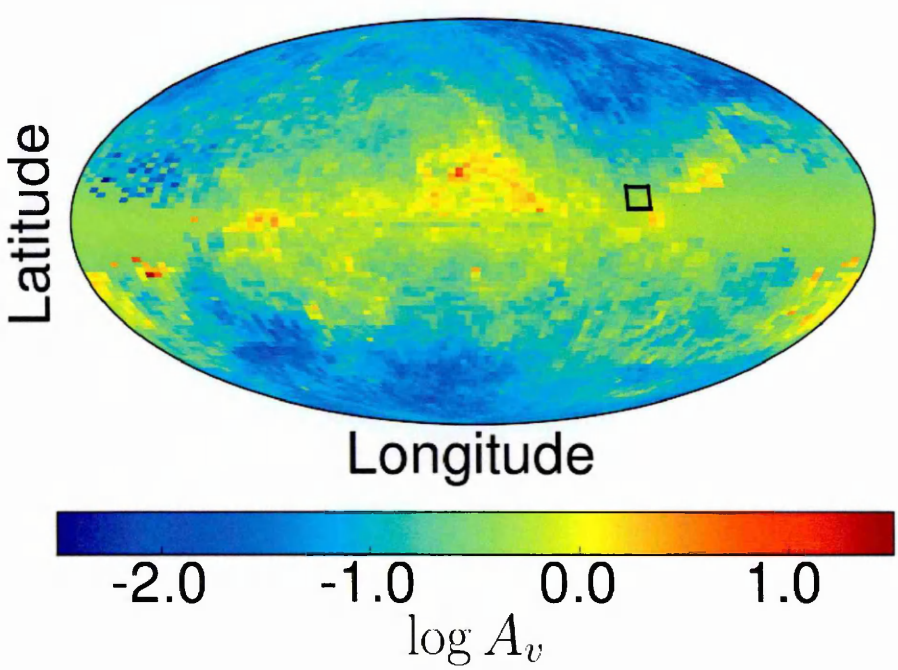



Figure 2.4: A map of the all-sky extinction ($A_v$), using Drimmel et al. (2003), out to a distance of 1 Kpc. Kepler's field of view is denoted in black

a few square degrees each, inside the integration bounds and averaged together. At this point we are only concerned with the bulk parameters of the region and not the fine structure of the field. These regions are designed such that the extinction can be treated as a constant over this region.

With the average extinction in the field of view defined I then compute the distance limits to be used in equation 2.1. With the $l$ and $b$ defined by the detector and $d$ now defined via the magnitude and extinction I then perform a numerical integration over equation 2.1. This computes the total stellar density for a system of absolute magnitude $M$, such that the system is be within the detection limits of the survey.

The numerical volume integration for each region makes use of a Romberg integral following Press et al. (1992). This divides the volume up into at least $2^5$ sub-intervals in each of the directions $l, b$ and $d$, and iteratively increases the number of sub-intervals by factors of 2 up to a maximum of $2^{10}$ intervals. This is done until the change in the integral, when increasing the number of intervals, changes by less than 0.1%. If this condition is not met once $2^{10}$ sub-intervals are reached the integral obtained for $2^{10}$ sub-intervals is used. I found that decreasing the cut-off to below 0.1% did not significantly alter the results, while it markedly increased the computational runtime. Very few integration regions need more than $2^5$ intervals.

As part of the integration routine I can model a region that is not defined as parallel to $l$ and $b$ axis, such as the Kepler detector which has its CCDs orientated as boxes rotated $\sim 45^o$ to the $l$ and $b$ axes. For this I weight each of the sub-intervals by the proportion that the sub-interval overlaps with the CCD, using the Convex-Intersect routine from O'Rourke (1995). This assumes that the Galactic density is a constant over the sub-interval, which with a minimum resolution of $2^5$ is met.

### 2.1.4 Population characteristics

The system-specific observable volumes are then multiplied by weighting factors determined from the distribution functions of newly-formed stars and binaries.

This then calculates the total number of each type of binary and single star that are visible in a field of view in the current epoch. For single stars I consider only the initial mass function (IMF), while for binaries I consider the IMF, initial mass ratio distribution (IMRD) and initial orbital separation distribution (IOSD). There is then a final distribution the initial binary fraction distribution (IBFD) which applies to both singles and binaries.

The IMF describes the distribution of the formation mass for stars. Its functional form relates to how proto-stellar gas clouds fragment into multiple components which will then go on to form stars (Padoan & Nordlund, 2002). I use a IMF modelled as a 3-component power law (Kroupa, 2001):

$$\begin{aligned} IMF &= 0 && M_1 < 0.1 \\ IMF &\propto M^{-1.3} && 0.1 \leq M_1 < 0.5 \\ IMF &\propto M^{-2.2} && 0.5 \leq M_1 < 1.0 \\ IMF &\propto M^{-2.7} && 1.0 < M_1 \end{aligned} \tag{2.4}$$

where $M_1$ is the mass of the initially more massive component of the system (recall single stars are modelled as binaries with a very low mass companion). Their are uncertainties in the choice of the IMF, for instance some propose a log-normal distribution (Miller & Scalo, 1979) or a power-law distribution (Salpeter, 1955). In general the different IMFs agree at the higher mass end, $M > 1M_\odot$, while disagreeing at the lower mass range where measurements are less certain.

To derive the number of systems that form each year, I assume that 1 star with $M > 0.8M_\odot$ forms each year. This is based on the formation rate of white dwarfs (WDs) from Weidemann (1990) and by assuming that the SFR is a constant then this implies the birth rate of WDs is the same as the birth rate of the progenitors of the WDs. Then by integrating the IMF between $0.8M_\odot$ and $8M_\odot$, I derive the number of systems that form (7.6 stars $yr^{-1}$). The upper mass limit of the integral is set by the largest star that can form a WD, and that there is a stellar evolution track for, in the library. However increasing the upper mass limit changes the result very little due to

the IMF being negligible at high masses. By choosing the Kroupa (2001) law, and in combination with the other input parameters, I derive a realistic number of systems in a field of view.

The IMRD describes the distribution of mass between the two stars in a binary system, when it forms. The IMRD can be expressed in the form:

$$IMR(q) \propto q^s$$

where q is the mass ratio $M_2/M_1$, $M_2$ is the mass of the initially least massive component, and $s$ is a free parameter. I consider $s = 0$ (a flat distribution) as the canonical value for this work. There is little agreement in the literature as to the precise form of the IMRD, though it appears to vary as a function of the primary mass. Janson et al. (2012) finds for M-dwarfs a best fit is a flat distribution, Duquennoy & Mayor (1991) favours a Gaussian distribution that peaks at $q = 0.25$ for G dwarfs while (Raghavan et al., 2010) suggest a distribution that favours equal sized objects ($s \rightarrow 1$). Finally, Goodwin (2013) argues that this can all be explained if instead of selecting stars via their primary mass in the IMF, instead the system mass is used. Combined with a flat mass ratio distribution this can then give rise to an apparent mass dependant IMRD.

The IOSD is modelled as log flat in orbital separation distribution (Kroupa & Petr-Gotzens, 2011):

$$\begin{aligned} IOSD &= 0 && a/R_\odot < 3 \\ IOSD &\propto a^{-1} && 3 \leq a/R_\odot < 10^6 \end{aligned} \tag{2.5}$$

where $a$ is the initial orbital separation. The lower limit is a simplistic cut (Hurley et al., 2002), while binaries beyond the upper limit are likely to be disrupted by passing intergalactic stars (Heggie, 1975).

Raghavan et al. (2010) found that ~ 50% of systems were single star systems, with binaries making ~ 30% and the rest being higher order systems. This therefore gives us a single star fraction of 50% and as BISEPS can't model higher order multiple systems this implies the binary fraction is 50%. There is evidence that the binary

volume:

$$P_s = IMF(M_1)M_1 \ln 10 \Delta M_1 \tag{2.12}$$

Combining equations 2.8, 2.9 and 2.11 leads to:

$$P_b = 3\ln 10 \times \left(M_1 q a\right) \times \left(IMF(M_1)IMR(q)IOSD(a)\right) \times \left(\Delta \log M_1 \Delta \log M_2 \Delta \log a\right) \tag{2.13}$$

This value $P$ can then be combined with the value of the time integral and Galactic volume integral to form the total number density of a stars evolutionary track at each point in its lifetime.

### 2.1.5 Magnitude of a system

To derive the apparent magnitude of a system we need to compute the absolute magnitude, which is based on the stars intrinsic luminosity. I first compute the relevant quantities in terms of the *bolometric* values. These are based on the total flux of the star integrated over all wavelengths. These bolometric quantities can then be computed for an individual filter band via bolometric corrections (BCs).

The systems bolometric luminosity, $L$, is given as an output of the SSE code. This is then translated into an absolute bolometric magnitude $M_{bol}$ via:

$$M_{bol} - M_{bol,\odot} = -2.5 \log (L/L_\odot) \tag{2.14}$$

where $M_{bol,\odot}$ and $L_\odot$ is the absolute bolometric magnitude and bolometric luminosity of the Sun, respectively. With the absolute magnitude I then translate this into an bolometric apparent magnitude and then into the apparent magnitude in the filter band the observations are made in. This requires the use of bolometric corrections, which translate the flux measured in a filter with the expected flux from in another filter. These values are highly dependant on the stellar parameters, temperature, surface gravity and metallicity, due to the BC wavelength dependence.

BISEPS as of Willems et al. (2006) only had V band BCs provided by a polyno-

mial fit to $T_{eff}$ from Flower (1996). I use the tables provided by Girardi et al. (2002) to expand the possible filter sets BISEPS can deal with. This expands the possible filter sets to include the Johnson-Cousins-Glass *UBVRIJHK*, Strömgren *uvbyβ*, Sloan *ugriz* and Kepler $K_p$ bands. The BCs are provided as a function of $T_{eff}$ and $\log g$ in the form of tables for different metallicities. They are based on the synthetic ATLAS9 spectra for stars between 3900K < $T_{eff}$ < 50000K and 0 < $\log g$ [dex] < 5 (Castelli et al., 1997), and the BDdusty1999 atmosphere models (Allard et al., 2001) for stars with 700K < $T_{eff}$ < 3900K. Stars hotter than $T_{eff}$ > 50000K are treated as black-bodies. M giants are treated separately by using the empirical spectra of Fluks et al. (1994). These stellar spectra are integrated over the filter response curve to derive the bolometric corrections for a star in any filter system (Girardi et al., 2002).

I perform a bi-linear interpolation over $T_{eff}$ and $\log g$ for tabulated metallicities either side of a target metallicity and then a linear interpolate between the two metallicities. If the metallicity lies exactly on the boundary of the tables (for instance for Solar metallicity) then I use only that table. If the parameters of a star place it outside of the range provided by the tables in Girardi et al. (2002) then the closest point inside the tables is used rather than risk extrapolating the data.

With the BC defined for a specific $T_{eff}$, $\log g$ and metallicity I then calculate the absolute magnitude $M_x$ of a star for a specific filter $x$ as

$$M_x = -2.5 \log\left(\frac{L}{L_\odot}\right) - M_{Bol,\odot} - BC_x \tag{2.15}$$

where $L$ is the star's bolometric luminosity, as delivered by the evolutionary model, $M_{Bol,\odot}$ is the Sun's bolometric magnitude and $BC_x$ is the BC for a star in filter band $x$. I calculate $M_{Bol,\odot}$ in a self-consistent way from:

$$M_{Bol,\odot} = M_{V,\odot} + BC_{V,\odot} \tag{2.16}$$

The Sun is taken to have $T_{eff}$ = 5777K and $\log g$ = 4.44 [dex], giving $BC_{V,\odot} = -0.06$ (Girardi et al., 2002). Defining the visual apparent magnitude of the Sun to be

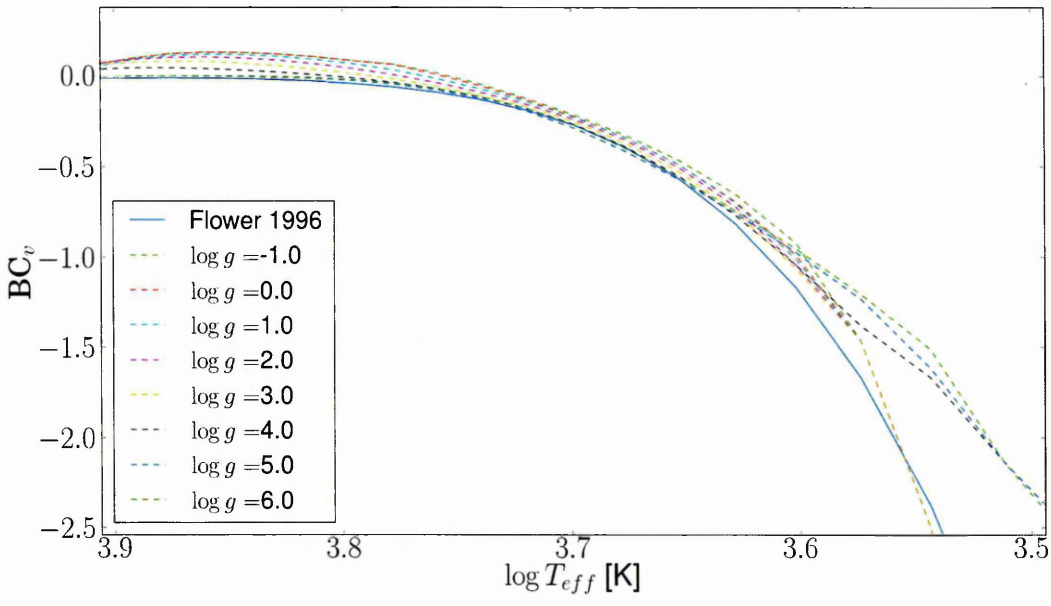



Figure 2.5: A comparison of Flower (1996) and Girardi et al. (2002) BCs over a range of $T_{eff}$ in the Johnson $V$ band. Note Flower (1996) has no log $g$ or metallicity dependence.

$V_\odot = -26.76$ implies $M_{V,\odot} = 4.81$, and hence $M_{Bol,\odot} = 4.75$ (Torres, 2010). For binary systems I calculate each set of BCs separately, compute the filter dependant luminosity of each star, sum the two luminosities together and use this luminosity to compute the absolute magnitude of the system in the required filter band.

Comparing the results by Flower (1996) to Girardi et al. (2002), in fig. 2.5 we can see that that there is a reasonable agreement between the two sets of corrections for $\log T_{eff} > 3.6$K, though the differences increase with increasing radius of the star. Below $\log T_{eff} < 3.6$K the differences increase to at least an order of magnitude difference. This is not unexpected as Flower (1996) notes there are uncertainties in the $T_{eff}$ measurements for cool objects which is also then combined with the limited sample size used, increasing the overall uncertainty.

### 2.1.6 Extinction

The extinction $A_\lambda$ in a given filter band is computed from the extinction $A_V$ in the visual band, given by Drimmel et al. (2003), via the relation $A_\lambda / A_V = \Lambda$, where $\Lambda$ is a filter dependant coefficient (Girardi et al., 2008). $\Lambda$ is calculated in a similar fashion

to the BCs by taking a bilinear interpolation of $T_{eff}$ and $\log g$ and then interpolating over $A_v$ (Girardi et al., 2008). I interpolate over $A_v$ due to $\Lambda$'s slight dependence on it via the "Forbes effect", where extinction is stronger the higher the flux, while the metallicity dependence is minimal and thus ignored (Girardi et al., 2008).

### 2.1.7 Creating a discrete sample

The result of the above volume integration and weighting with initial distributions is a multi-dimensional, continuous (albeit binned) distribution function $\Gamma$ that characterizes the content of the field-of-view at the current epoch. The total number ($N$) of stars and binary systems in the field is given by the sum over $\Gamma$. To obtain a representative synthetic sample of systems which can then be subjected to further selection processes I create a discrete synthetic sample of $N$ stars from this continuous distribution.

To this end I draw a random sample with replacement, of $N$ objects from the distribution function $\Gamma$. Each object in the sample is placed randomly at a location ($l$, $b$, $d$) inside the field of view, based on the Galactic density, extinction and absolute magnitude of the system. Comparing the continuous distribution $\Gamma$ to the sub sampled distribution shows no discernible difference as seen in fig. 2.6.

This sub sample then represents a magnitude limited complete synthetic catalogue of stars in a field of view. Where we have the stellar parameters (mass, radius etc), the system properties for binaries (orbital separation) and the Galactic parameters (4D spacial and temporal location). This catalogue then forms the basis for further work, such that we can apply measurement biases to better model that of a measured catalogue of stars. This can be done in multiple ways for instance; including noise terms in the magnitude; applying selection effects, like a colour cut; selecting systems that would be predicted to show variability, for instance EBs or asteroseismic signals.

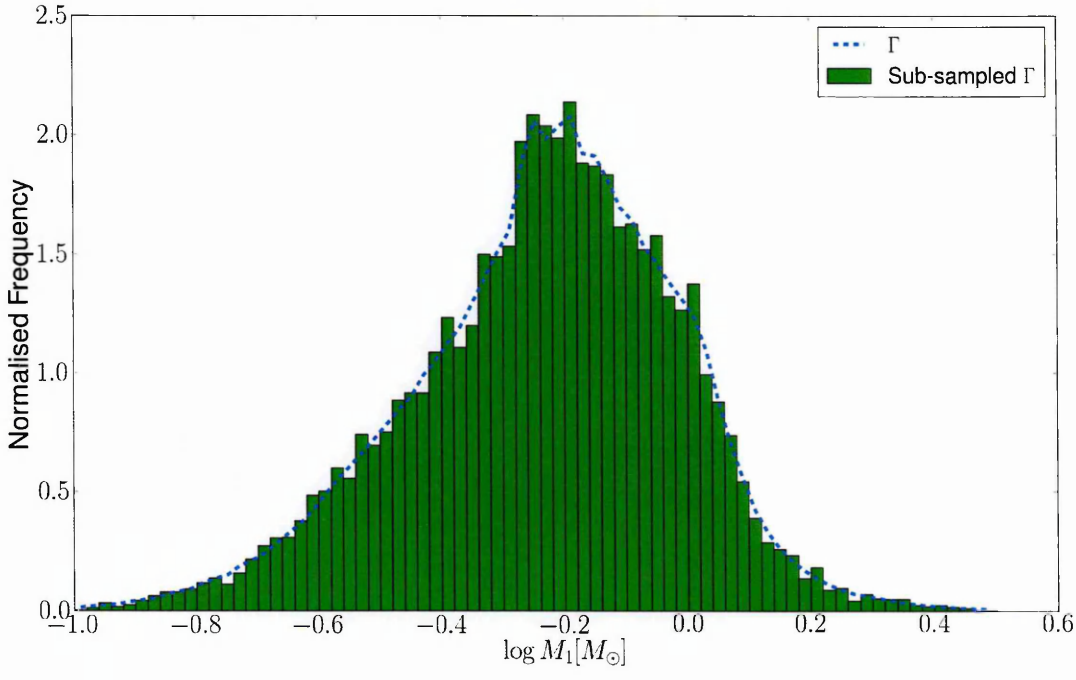



Figure 2.6: The distribution function Γ, solid bars, and the sub-sampled Γ, dotted line, for a population of ~ 11,000 stars, showing no discernible difference between the two methods.

## 2.2 Eclipsing Binaries

Eclipsing binaries (EB) occur when one star in a binary system passes in front of its companion star along our line of sight. As the star passes in front there is a gradual decrease in flux as the star orbits the companion, reaching a maximum before decreasing again until the star no longer blocks any of the flux from its companion. Eclipses come in two types, grazing eclipses and total eclipses. In a grazing eclipse, the eclipsing star passes only partially over its companion's disk from our point of view. While in a total eclipse it passes entirely (though it may not cover the companion completely) over the disk.

Modelling the light curve of the system during the eclipse, is a complex process (Southworth et al., 2004). The basic model of this assumes that there are two circles each of radius $R_1$ and $R_2$ and both have uniform flux over their surface which allows us to express the change in flux $\Delta F/F$ purely geometricly (Mandel & Agol, 2002). However there are complications to this picture; a star does not present a uniform flux to an observer over its disk, thus the change in flux is not equal to

the area covered during the eclipse (Wilson & Devinney, 1971); the stars projection along our line of sight may not be a circle of a constant radius, if for instance the star is distorted due to its companion (Morris, 1985). There are also many other factors that may alter the light curve; for instance one star may be reflecting the flux from its companion (Wilson, 1990); star spots and stellar variability will change the flux emitted by the star during the eclipse and in eccentric binaries (not limited to eclipsing systems) the companion can induce pulsations in its companion (Thompson et al., 2012).

Therefore I use the JKTEBOP, as described in Southworth et al. (2004) and based on the EBOP model by Popper & Etzel (1981). This computes a light curve for a given binary system by treating each star as a bi-axial spheroid and projecting this shape onto each star. This projection can then be integrated over concentric circles for each star, taking into account limb darkening, gravity darkening, reflection effects and ellipsoidal modulation to derive a light curve for the system. This light curve can then be subjected to a noise term and/or background flux which dampens the eclipse signal.

### 2.2.1 Darkening coefficients

To derive the light curve I take into account two effects that alter the stellar luminosity, as seen by an observer. Firstly limb darkening this is where the flux at the edge of the star appears less than the flux at the centre. Gravity darkening is also considered this is where the star under the influence of centrifugal force, due to its rotation, bulges at its equator. This alters the gravitational potential differentially between the poles and the equator changing the local temperature and hence luminosity.

The surface brightness of the disc of a star appears to get fainter the further from the centre we look. The light we see from the star, at a certain wavelength, will have been emitted at a optical depth $\tau$ below the photosphere. For light emitted at the centre of the disc, it only has to travel in the radial direction a distance of $\tau$ within the stellar atmosphere. But any light emitted at the edge of the stellar disc, that we see,

must have travelled at some angle $\lambda$ relative to the radial vector. Thus the photons emitted at the edge must have travelled a longer distance of $s = \tau / \cos\lambda = \tau / \mu$ where $\mu = \cos\lambda$. Therefore we can describe the change in intensity as $\frac{I(\mu)}{I(1)}$ where $I(\mu)$ is the intensity at a point on the surface of the apparent stellar disc and $I(1)$ is the intensity at the centre of the disc.

Various relationships have been proposed for $\frac{I(\mu)}{I(1)}$:

Linear (Al-Naimiy, 1978)

$$\frac{I(\mu)}{I(1)} = 1 - c_1(1-\mu) \tag{2.17}$$

Quadratic (Wade & Rucinski, 1985)

$$\frac{I(\mu)}{I(1)} = 1 - c_1(1-\mu) - c_2(1-\mu)^2 \tag{2.18}$$

Logarithmic (Klinglesmith & Sobieski, 1970)

$$\frac{I(\mu)}{I(1)} = 1 - c_1(1-\mu) - c_2\mu\log\mu \tag{2.19}$$

Power law (Claret, 2004)

$$\frac{I(\mu)}{I(1)} = 1 - c_1(1-\mu^{1/2}) - c_2(1-\mu) - c_3(1-\mu^{3/2}) - c_4(1-\mu^2) \tag{2.20}$$

where $c_i$ are the limb darkening coefficients (LDCs). The more complicated formulae fit the data better, at the expense of introducing more free parameters in the form of the LDCs. These coefficients are calculated from stellar atmosphere models, which means they have a dependence on $T_{eff}$, $\log g$, metalicity and also on the wavelength being observed at. As modelling the entire stellar atmosphere is very complex and time consuming, many authors (van Hamme, 1993; Claret, 2004; Sing, 2010) have produced tables of coefficients, for various limb darkening laws and wavelengths, that can be interpolated over to cover a range of $T_{eff}$, $\log g$ and metallicity

I used a quadratic limb darkening law of Wade & Rucinski (1985) The LDCs were

taken from the tables provided in Claret & Bloemen (2011) who provided them as a function of $T_{eff}$, $\log g$, metallicity, microturbelent velocity, and wavelength band. Setting microturbelent velocity to 2km $s^{-1}$ I perform a bi-linear interpolation over $T_{eff}$ and $\log g$ at a metallicity value either side of the targeted metallicity. Then I apply a linear interpolation between metallicity values, to derive the LDCs in the appropriate band pass. Claret & Bloemen (2011) derived the LDCs using both the ATLAS9 and the PHOENIX model atmospheres, I choose to use the ATLAS9 model over its whole range, 3500K < $T_{eff}$ < 50000K, and the PHOENIX model for a sub-range, 2000K < $T_{eff}$ < 3500K. Note the BC tables go to $T_{eff}$ = 700K while the limit for the LDC is $T_{eff}$ = 2000K, this is not an issue as the BISEPS code does not evolve objects with $M < 0.1 M_{\odot}$ and thus the minimum $T_{eff}$ is approximately 2500K.

As a star rotates it may become non-spherical as the centrifugal force forces material at the equator away from the rotation axis. At the equator the ratio of these forces with respect to gravity is greatest thus the radius increases at the equator. As the radius has increased, the surface gravity at the equator decreases and this lowers the emitted flux of the material at the equator making the equator appear dimmer. This can be expressed as a relationship between the flux and the surface gravity as $F \propto g^{\beta}$ where the parameter is usually assumed to be $\beta = 1.0$ (von Zeipel, 1924) for radiative atmospheres and $\beta = 0.3$ (Lucy, 1967) for convective atmospheres.

I however use the work of Claret & Bloemen (2011) which derives $\beta$ as a function of $T_{eff}$, $\log g$, metallicity, microturbelent velocity and wavelength band. Under the same conditions and assumptions made for the LDCs, I bilinearly interpolate over $T_{eff}$ and $\log g$ and then linearly interpolate over metallicity to derive the gravity darkening coefficient (GDC).

# Chapter 3

# Kepler target selection

In this Chapter I create a synthetic Kepler input catalogue (KIC) and subject it to the Kepler Stellar Classification Program (SCP) method for determining stellar parameters such as the effective temperature $T_{eff}$ and surface gravity $g$. I then compare this to my synthetic population model to determine the level of biases and selection effects inherent in the Kepler results.

I find a satisfactory match between the synthetic KIC and the real KIC in the $\log g$ vs $\log T_{eff}$ diagram, while there is a significant difference between the actual physical stellar parameters and those derived by the SCP of the stars in the synthetic sample. I find a median difference $\Delta T_{eff} = +500$ K and $\sim \Delta \log g = -0.2$ dex for main-sequence stars, and $\sim \Delta T_{eff} = +50$ K and $\Delta \log g = -0.5$ dex for giants, although there is a large variation across parameter space. The error in the temperature means most stars are hotter than expected thus their habitable zones will be further out than predicted. For a MS star the median difference in $g$ would equate to a $\sim 3\%$ increase in stellar radius and a consequent $\sim 3\%$ overestimate of the radius for any transiting exoplanet, which is comparable to the median uncertainty in planetary radius for Kepler. I find no significant difference between $\Delta T_{eff}$ and $\Delta \log g$ for single stars and the primary star in a binary system. I also re-created the Kepler target selection method and found that the binary fraction is unchanged by the target selection. Binaries are selected in similar proportions to single star systems; the fraction of MS dwarfs in the sample increases from about 75% to 80%,

and the giant star fraction decreases from 25% to 20%. A number of giant stars were misclassified as dwarfs. This implies that a bias-corrected $\eta_{earth}$ for Kepler will be underestimated due to an increase in the number of null detections, as a planet is harder to detect around a giant star.

## 3.1 Introduction

The work in this chapter is based on results presented in Farmer et al. (2013).

The NASA Kepler mission (Borucki et al., 2010) is designed to detect transiting exo-earths in habitable zones around solar-like stars. To achieve this goal Kepler monitored around 150,000 stars for 4 years. The target stars were selected from a larger list, the KIC, according to a set of criteria that rank stars in order of the likelihood to display detectable transits of exo-earths in the habitable zone (Batalha et al., 2010). The KIC covers the 116 square degrees of the Kepler field (Koch et al., 2010) and contains about 450,000 stars with magnitude brighter than $K_p = 16$ (where $K_p$ is the magnitude in the Kepler band). This catalogue was established to derive physical parameters for objects in Kepler's field of view and to allow the selection of a set of optimal targets that would maximise Kepler's chance of detecting an Earth-sized transit around a Sun-like star (Brown et al., 2011). The KIC itself was compiled from a ground-based survey using broad-band Sloan Digital Sky Survey (SDSS) filters with a flux precision of 2%.

Kepler's Stellar Classification Program (SCP) (Brown et al., 2011) derived basic physical parameters of all KIC stars, chiefly the effective temperature $T_{eff}$, surface gravity $g$, and metallicity $\log Z/Z_{\odot}$, and, by comparison with suitable stellar models, the stellar mass, radius and age, using only the observed broad-band magnitudes and colours of these stars as an input. The target selection in turn is based on these stellar classification program (SCP)-derived stellar parameters.

These SCP -derived parameters may suffer from random and systematic uncertainties introduced because the measured magnitudes of a star may differ from its true, intrinsic magnitudes, and because colours alone will not always unambigu-

ously deliver appropriate estimates of the physical parameters. This will in turn translate into a bias of the statistical properties of samples drawn from Kepler data, including the exoplanet candidate sample itself, or the sample of binary stars with Kepler light curves.

It is therefore important to critically examine the performance of the SCP approach, and the consequences of any inherent systematic bias for the actual Kepler target list, and for subsamples created from Kepler data. To this end I create a synthetic version of the KIC, using the BISEPS code which is then validated against the actual KIC in colour-magnitude space. I then employ the SCP technique to derive "apparent" stellar parameters for all stars in the synthetic sample, i.e. exclusively from their magnitudes in different colour bands. Then I investigate the difference between the actual, physical parameters of the synthetic stars, and their SCP -derived parameters.

Due to bandwidth limitations Kepler does not observe every object in the KIC, instead a target list is drawn up that aims to maximise the science return on the targets observed. This list is determined on the basis of the SCP -derived parameters and the expected flux levels, aiming to increase the fraction of Sun-like stars and decrease the fraction of giants in the sample. I aim to reproduce the target selection procedure and apply it to the synthetic sample of the KIC, to quantify the resulting bias against giants on the basis of the actual, physical parameters of the synthetic KIC stars.

In section 3.2 I model the processes involved in deriving the Kepler target list. Sections 3.3.1-3.3.3 I evaluate the effectiveness of my target selection model. In sections 3.3.4-3.3.6 I evaluate the effectiveness of the SCP and the target selection process. Finally in sections 3.4 and 3.5 I discuss my results and present my conclusions.

## 3.2 Kepler target list selection

Out of the possible 450,000 stars in the Kepler field, only ~ 150,000 can be observed at any one time due to bandwidth limitations. Therefore Kepler uses a tailored target

list selected according to a number of criteria designed to maximise the likelihood for the detection of Earth-like transits in the star's habitable zone (Batalha et al., 2010). To be able to generate a synthetic target list from the synthetic sample that would reproduce the actual Kepler target list, I created my own model of the Kepler detector system and target selection method. Following the procedures as set out in Brown et al. (2011), Bryson et al. (2010b) and Batalha et al. (2010) this entailed the following principal steps: (a) derive estimates of the system parameters from broad-band colours using the SCP routine, (b) construct a model of the expected S/N measured by each pixel, and then combine (a) and (b) to calculate the likelihood of detecting Earth-like transits in the star's habitable zone.

In essence, to compile the target list the stars are ranked in terms of the minimum radius $R_{p,min}$ of a planet that can still be detected securely in the absence of intrinsic stellar noise within the nominal 3.5 yr mission. The radius $R_{p,min}$ is obtained by requiring that the relative transit depth in flux $F$, $\Delta F/F = (R_p/R_*)^2$, where $R_*$ is the stellar radius, exceeds a suitable multiple of the light curve noise $\sigma_{tot}$. This becomes

$$R_{p,min} = R_* \sqrt{\frac{7.1\sigma_{tot}}{r}} \tag{3.1}$$

(Equation 7 of Batalha et al., 2010) where $r$ is a crowding metric and discussed below. Choosing to set the noise level to $7.1\sigma$ also implies that there would only be one statistical false positive signal due to random fluctuations in the data set (Batalha et al., 2010).

I now discuss the different factors in equation 3.1.

### 3.2.1 Stellar classification

The determination of physical parameters of all KIC stars, including the stellar radius $R_*$, is the remit of the SCP . This uses a Bayesian posterior probability estimation method to derive a star's $T_{eff}$, $\log g$, $\log Z/Z_\odot$, luminosity, mass and radius from its observed colours (Brown et al., 2011).

The two-step procedure is based on two sets of input models. Stellar atmo-

sphere models of Castelli & Kurucz (2004) were combined with filter response functions to determine the expected colours. This was defined for objects between 3500K $< T_{eff} <$ 50000K, $0 < \log g$ cm s$^{-2} < 5.5$ and $-3.5 < \log(Z/Z_{\odot}) < 0.5$ (although not every gravity is available at every temperature), while stellar evolution tracks of Girardi et al. (2000), assuming a constant star formation rate and solar metallicity, link these with the stellar mass and radius.

Bayesian priors based on the $T_{eff}$, $\log g$ distributions of stars observed by the *Hipparcos* satellite (Perryman et al., 1997), the $\log Z$ distribution from Nordström et al. (2004) and a Galactic distribution model from Cox & Pilachowski (2000) are employed to focus the search in parameter space. The claimed advantages of a Bayesian approach is that the chosen priors, a set of known distributions of the parameters, rule out implausible systems which a standard $\chi^2$ minimisation technique might obtain. However shortcomings were noted in Brown et al. (2011); the metallicity distribution is only valid statistically and not on an individual star basis, $T_{eff}$ is unreliable for the hottest and coolest objects and there are systematic errors in $\log g$ for objects with $g - r > 0.65$.

For each object in the BISEPS synthetic sample I supply the calculated *g,r,i,z* and *D51* magnitudes as an input for the SCP code, to estimate the object's physical parameters in the same way as the SCP did for the stars in the real KIC[1] (Brown et al., 2011). The SCP code takes into account magnitude uncertainties, and for simplicity I assume a value of 0.02 mags in each band for all stars, which is the quoted photometric precision for objects with $K_p < 15$, as measured by the SCP (Brown et al., 2011). As the KIC required excessive exposure times in the $u$ band it is excluded from the fitting process by selecting a large photometric uncertainty for it. I also found that the J, H and K magnitudes had little effect on the results, and thus excluded these bands as well, to reduce the number of unnecessary fit parameters and save CPU time (see section:3.3.2).

For simplicity I followed the SCP approach and adopt a single value of $\Lambda$ for all

[1]http://www.cfa.harvard.edu/kepler/kic/kic/index.html

| Filter | Extinction coefficient(Λ) |
|---|---|
| g | 1.193 |
| r | 0.868 |
| i | 0.681 |
| z | 0.490 |
| D51 | 0.999 |

Table 3.1: Extinction coefficients for a 5000 K, $\log g = 4.0, \log(Z/Z_\odot) = 0.0$ star, from tables provided in Girardi et al. (2002)

stars in each filter band, neglecting the real dependence on $T_{eff}, \log g, Z$ (Girardi et al., 2002). I chose the coefficients of a 5000 K, $\log g = 4.0, \log(Z/Z_\odot) = 0.0$ star (Girardi et al., 2002), which are given in Table 3.1. These coefficients are for the same star as used in the generation of the KIC (Brown et al., 2011)

### 3.2.2 S/N determination

Determining the expected S/N for an observation requires knowledge of Kepler's noise characteristics, a model for which exists in Bryson et al. (2010a), however the tools required are not publicly available and therefore I re-derive them here.

To calculate the S/N expected for each synthetic system from its $K_p$ magnitude and the system's RA and dec we require a model of Kepler's focal plane geometry (FPG) which is derived here.

To place the synthetic star on Kepler's focal plane I obtained its pixel coordinates by extrapolating those of the closest match in the actual Kepler data set, based purely on the star's RA and dec[2]. Stars near the centre of the field have an almost circular pixel response function (PRF), while near the edge the PRF is elongated towards the centre of the field (Bryson et al., 2010b). Thus the PRF s are both a function of CCD and pixel location of the system. Bryson et al. (2010b) defines a set of 5 PRF s for each CCD, four in the corners and one in the centre. Each of these PRF s gives the flux distribution over a $n \times n$ pixel array, with usually $n = 11$ but occasionally $n = 15$, for a star centred in the middle of the grid. Examples are shown in fig. 3.1, highlighting the differences between the PRF depending on location in the

[2]http://keplergo.arc.nasa.gov/ContributedSoftwarePyKEP.shtml

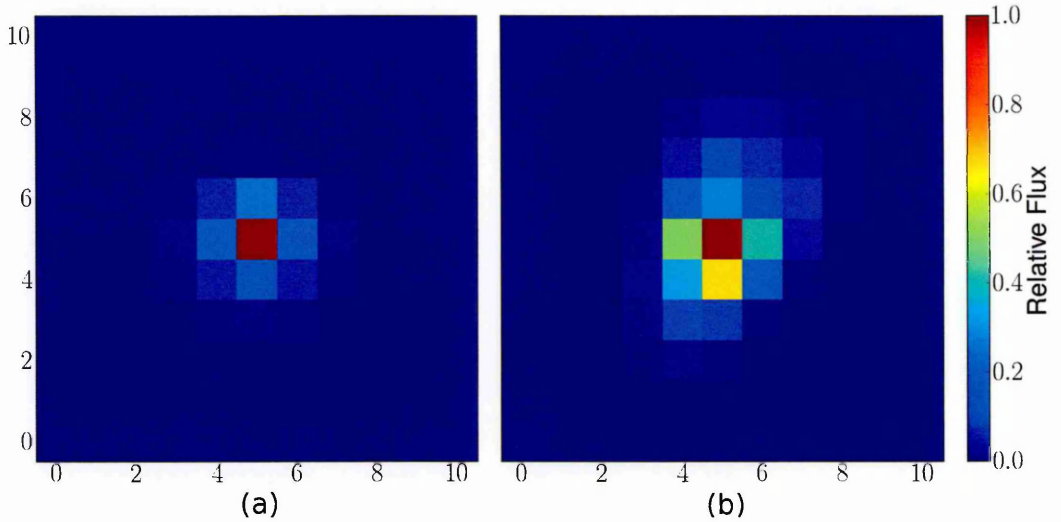



Figure 3.1: The flux distribution for two of Kepler's CCDs, highlighting the differences in the PRF depending on their location in Kepler's focal plane. Figure (a) is near the centre of the focal plane and thus the flux is concentrated in the centre, while (b) is near the edge of the focal plane and its PRF is elongated along the axis of deflection of the Kepler CCDs. The axis in both cases is a nominal pixel location

CCD array. To derive the PRF of a synthetic system I linearly interpolate between the 2 nearest corner PRFs and the central PRF .

In this way I build up a full frame image (FFI) of all synthetic stars in the Kepler field down to a limiting magnitude of $K_p = 19$. This was chosen as the assumed zodiacal light emission equates to a $19^{\text{th}}$ magnitude star on each pixel (Jenkins et al., 2004).

With the FFI in place I can determine the noise per pixel, as described in Caldwell et al. (2010) and summarised here. For each synthetic star I calculate the PRF and subtract this from the FFI to obtain an image of the system on its own as well as of the background around the system, including the zodiacal light.

I convert the flux to electrons via

$$f_{kep} = 10^{-0.4(K_p-12)} \times f_{12} \tag{3.2}$$

where $f_{12} = 1.74 \times 10^5 e^- s^{-1}$ is the photoelectric signal for a G2 V star with $K_p = 12$ (Jenkins et al., 2010b). The resulting flux distribution across part of one of Kepler's CCD, for the synthetic sample, can be seen in fig. 3.2a I then apply smearing to each

image, by summing the flux of each pixel in each column, multiplying by the read time of 0.52s, dividing by the number of rows and adding this to each pixel, the effect of which can be seen in fig. 3.2b.

At this point I apply a saturation model by 'rolling over' the electrons which are above the well depth (Caldwell et al., 2010), as seen in fig. 3.3a. This is done by performing a 50/50 split of the overflowing electrons, moving half of them up the pixel column and half down the pixel column, with each subsequent overflow moving electrons in the same direction; until such a point that the number of electrons per pixel is at most the well depth (Van Cleve & Caldwell, 2009). A charge transfer efficiency (CTE) model is then applied with a value of 0.99993 for the parallel reads and 0.99995 for the serial reads (Van Cleve & Caldwell, 2009) as seen in fig. 3.3b. The effect of the CTE in fig. 3.3b can bee seen as a slight increase in the number of electrons to the right and up from the peak saturated pixels seen in fig. 3.3a. For simplicity I assume the the change due to the CTE is small and does not require us to recompute which pixels are saturated.

With both images now expressed in electrons and the various systematics applied I calculate the S/N ratio for each pixel using

$$S/N = \frac{S}{\sigma_{ccd}} \tag{3.3}$$

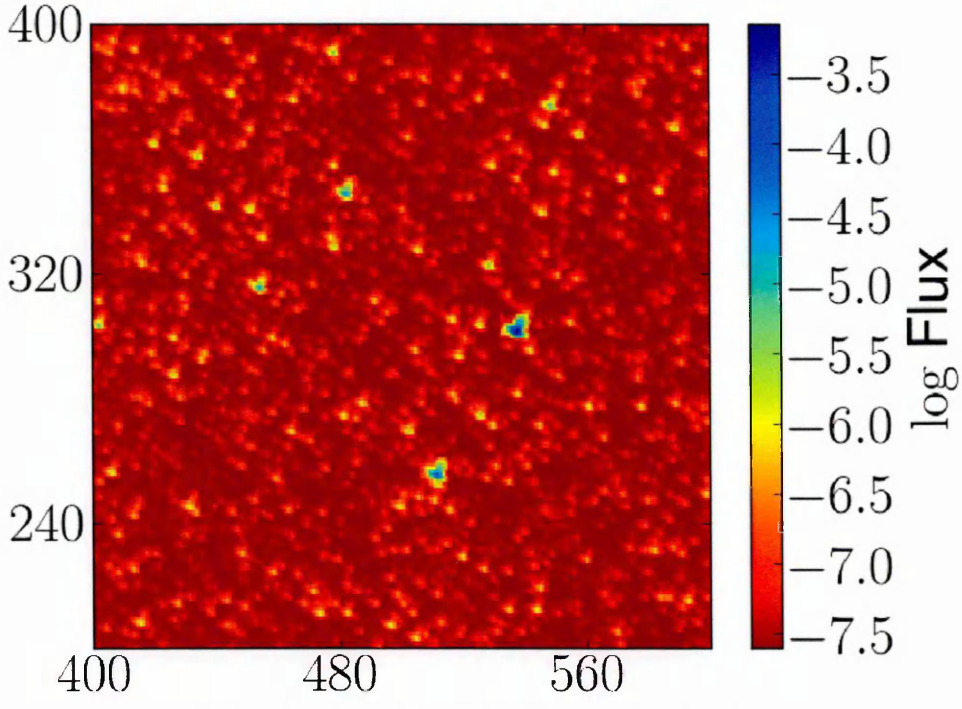
and

$$\sigma_{ccd} = \sqrt{S + Bg + \sigma^2_{read} + \sigma^2_{quant}} \tag{3.4}$$

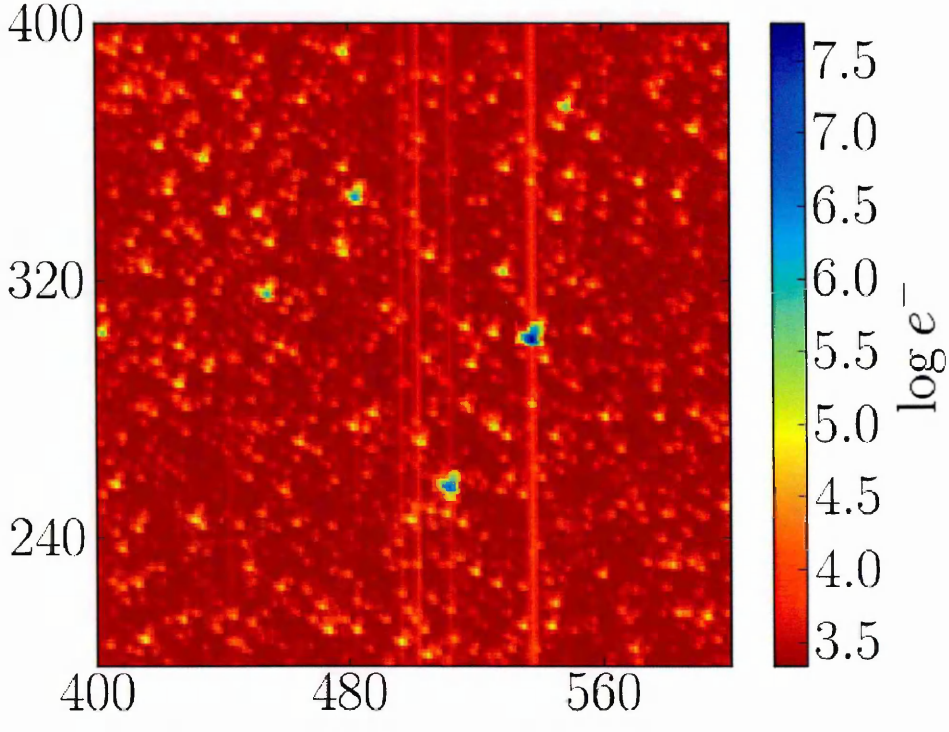
In this version of the CCD equation the signal $S$ and background $Bg$ are given in electrons, while the read noise $\sigma_{read} \sim 100e^-$ per read is CCD dependant (Van Cleve & Caldwell, 2009) and the quantisation error $\sigma_{quant}$ is given by (Bryson et al., 2010a):

$$\sigma_{quant} = \sqrt{\frac{1}{12}}\left(\frac{\mathrm{W}}{2^{N_{bits}-1}}\right). \tag{3.5}$$

Here the CCD well depth W is of order $\sim 10^6 e^-$ per pixel (though it is also CCD dependant, see Van Cleve & Caldwell (2009)), and $N_{bits} = 14$ denotes the number of

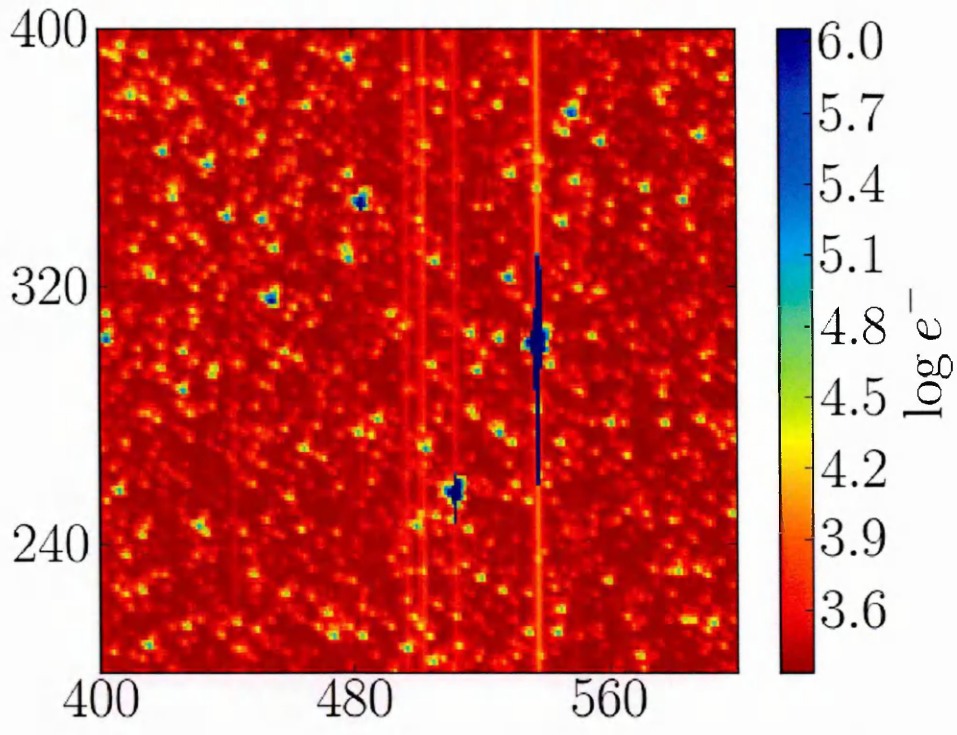



(a) Distribution of the flux for a set of synthetic stars.

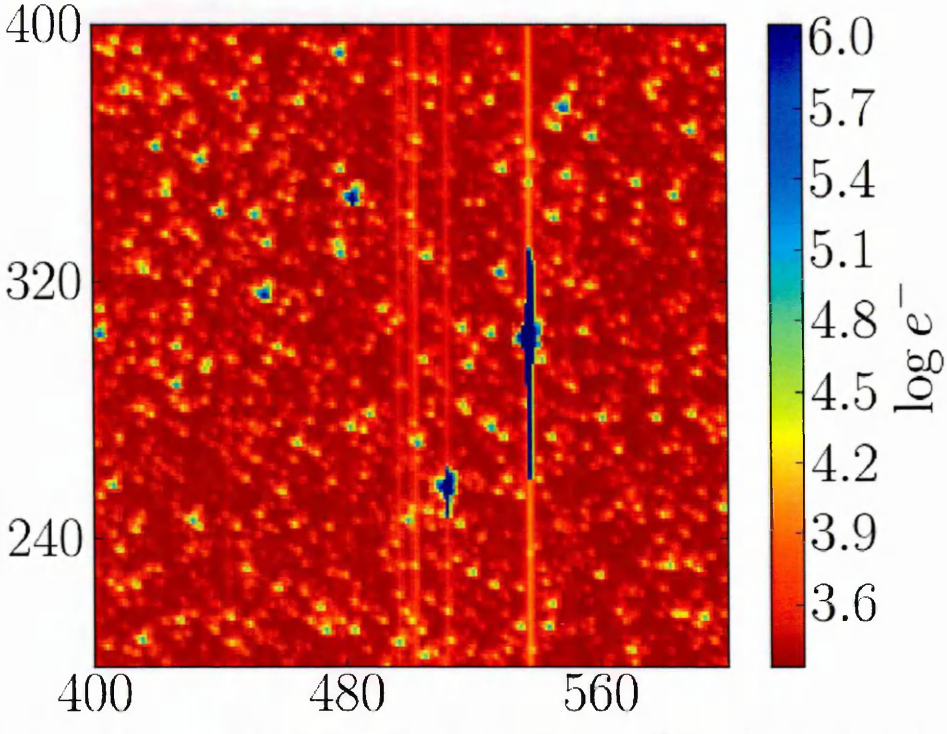



(b) The distribution of charge after I have applied smearing, which acts to distribute charge along a column.

Figure 3.2: Each plot represents a section of a Kepler CCD as I generate a synthetic image. (a) The raw flux image and (b) image after application of smearing.




(a) The distribution of charge after I have taken into account saturation effects, which provides a maximum amount of charge that can be held by any pixel; any extra charge is moved to the next pixel in the column.




(b) The distribution of charge with CTE taken into account. This models the charge that gets left behind during readout process and increases the amount of charge for pixels above and to the right of a pixel.

Figure 3.3: (a) with the saturation model applied, (b) Final image after application of a charge transfer efficiency (CTE) model. In all cases the axes indicate the pixel locations.

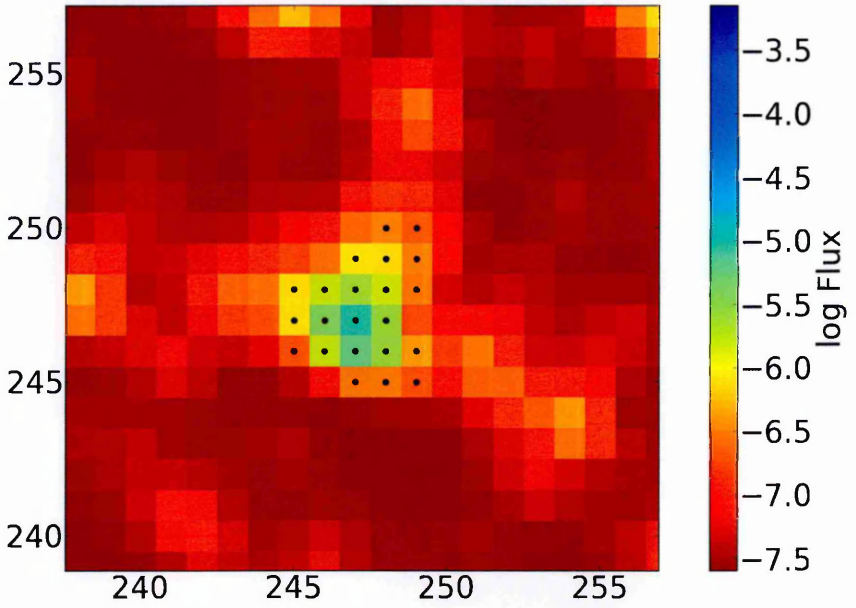



Figure 3.4: Optimal aperture (black dots) of a star. Note how the aperture is non uniform, unlike the PRFs in fig. 3.1. This is due to the background stars reducing the S/N of the target star

bits the data is quantised to. This gives $\sigma_{quant} \sim 30e^{-}$ per pixel.

The pixels are ranked in order of decreasing S/N and summed in quadrature, until the sum of the S/N is maximised, thus defining the optimum aperture for the star. This is repeated for each star with $K_p \leq 16$. A representative optimal aperture is shown in fig. 3.4.

The total photometric error $\sigma_{tot}$ is obtained from the S/N value, scaled by the total number of individual integrations while the system was in transit over the envisaged 3.5 years of the mission. This number is the product of the 270 integrations co-added together in one long-cadence (30 min) observation, the number $N_{sample}$ of long cadence observations that fit in a single transit, and the number $N_{tr}$ of transits in 3.5 years. The total photometric error is thus:

$$\sigma_{tot} = \frac{N}{S \times \sqrt{270 N_{sample} N_{tr}}}. \tag{3.6}$$

For randomly distributed inclinations of circular orbits the average transit duration is $t_0\pi/4$, where $t_0$ is the duration of a transit that is central across the star. $N_{sample}$ is

then $t_0(\pi/4)/30$ min. The central transit duration is calculated as

$$t_0 = 2R_* \sqrt{\frac{a}{GM_*}}, \tag{3.7}$$

with the stellar mass $M_*$ derived from the SCP , and with the semi-major axis $a$ taken at three different locations, $5R_*$, $0.5H_*$ and $H_*$. The quantity $H_*$ is the characteristic distance of the habitable zone (HZ) for the star in consideration and is given by $0.95\sqrt{L_*/L_\odot}$(Batalha et al., 2010).

The final term required for evaluating equation 3.1 is the crowding metric, $r$, which is given by Batalha et al. (2010)

$$r = \frac{F_*}{F_* + F_{bg}} \tag{3.8}$$

where $F_*$ is the flux from the star in the optimal aperture before addition of the systematics, and $F_{bg}$ is the flux from the background in the optimal aperture before addition of the systematics but after the zodiacal light has been added.

### 3.2.3 Testing the target selection code

With $R_{p,min}$ calculated for each synthetic star in the field-of-view I can draw up a ranked list of stars in order of increasing $R_{p,min}$. The subset of systems with a detectable terrestrial sized transit in the habitable zone, i.e. $R_{p,min} \leq 2R_E$ (where $R_E$ denotes the radius of the Earth) includes a large number of objects, ~ 60%, that are too faint ( $K_p > 15.0$) for radial velocity follow up. Thus an additional prioritisation scheme is employed, the details of which are given in table 1 of Batalha et al. (2010). In essence, the highest priority stars are those with $R_{p,min} \leq 2R_E$ in the HZ, with a magnitude bright enough to perform high precision radial velocity on ( $K_p \leq 14$), followed by those with $14 \leq K_p \leq 16$. Then there are those with detectable Earth-sized planets at $a = 0.5H_*$ or $a = 5R_*$ (these deliver a larger number of transits over the lifetime of the mission), and finally those with $R_{p,min} < 2R_e$ in the HZ around the faintest stars. Batalha et al. (2010) divides the sample into 13 classification groups,

with the 11 highest priority groups making up the target list.

To test the target selection code I apply it to the actual KIC and compare the target list we obtain with the Kepler Quarter 2 (Q2) data set which I use as a proxy for the actual Kepler target list. We chose Q2 as the catalogue of objects defined in Batalha et al. (2010) is not publicly available, and because both Quarter 0 and Quarter 1 were affected by commissioning of the Kepler instrument. In Q0 only $\sim 50,000$ stars were observed (Borucki et al., 2011a), while Q1 had over-sized apertures (Borucki et al., 2011b), thus a reduction in the number of faint, $K_p = 15-16$, stars. I chose not to use later quarters either because after each quarter some targets are removed due to follow-up work, or added due to the guest observation program. Tenenbaum et al. (2013) shows that in the first 12 quarters, 60% of objects were observed for all 12 quarters. A further 15% were observed for 10 quarters; these predominately are systems falling on the CCD module that failed during quarter 4, and thus were only observable for 75% of the time.

I show the magnitude distributions of the synthetic target list and that of the actual target list in fig. 3.5a. The discrepancy seen is primarily due to giant stars, here defined as stars with $\log g < 3.5$, highlighted in fig. 3.5b. I could not attribute these differences to inadequacies in my implementation of the target selection and SCP code and rather suspect that at least some differences exist because the actual Q2 list will have had some objects added or removed from the original list of objects as defined in Batalha et al. (2010).

To achieve a better agreement I applied a series of ad-hoc corrections to my target selection criteria:

1. For faint objects ($14 < K_p < 16$), if $R_{p,min} \leq 2.0 R_E$ for $a = H_*$, I redefine the selection criterion to $R_{p,min} \leq 2.4 R_E$. This increases the number of faint dwarfs.

2. All objects that saturate at least one pixel are included, if they have not already been placed into one of the groups in Batalha et al. (2010). This predominantly increases the number of bright giants.

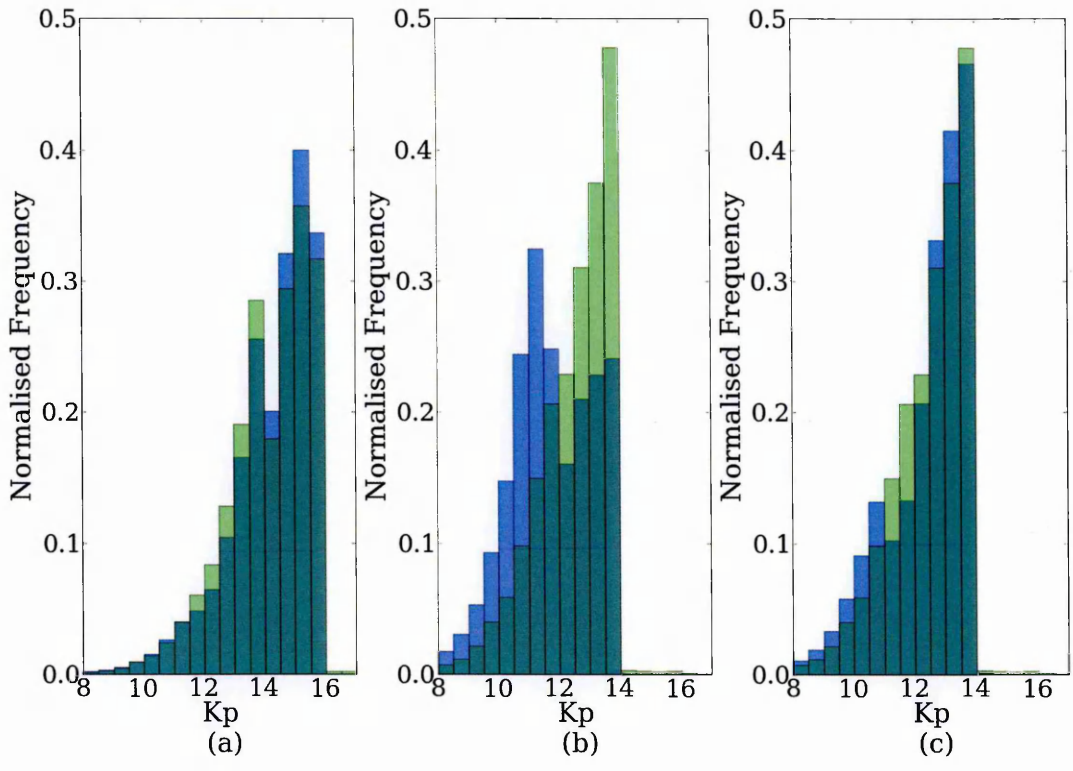



Figure 3.5: Normalised magnitude distribution of stars selected by my procedure (blue). Left: all systems, no ad-hoc corrections. Middle: only giants, defined as objects with KIClog $g < 3.5$, no ad-hoc corrections. Right: only giants, but with ad-hoc corrections. Shown in green in panel (a) is the Q2 target catalogue, while in panels (b) & (c) it is the giants from the Q2 catalogue.

3. All objects with $3 < R/R_{\odot} < 10$ and magnitude $K_p < 14$ are included, if they have not already been placed into one of the groups in Batalha et al. (2010). This is purely ad-hoc and is designed to increase the number of bright giants.

With these corrections in place I consider the match between the reproduced and actual target list satisfactory (see fig. 3.5c) and sufficient for the study of system properties presented in the following sections.

## 3.3 Synthetic population of Kepler stars

I now present the synthetic Kepler field population, covering both the synthetic KIC and the synthetic target catalogue which emerges from it. I compare the actual physical properties of the synthetic stars with the properties these stars appear to have when analysed with the SCP method.

### 3.3.1 Sample size

Using the population synthesis parameters described in Chapter:2, I obtain a total of ~ 353,000 objects in the synthetic KIC, compared to the ~ 416,000 objects in the real KIC. Increasing the Galaxy-wide SFR from the default value of 1 star $yr^{-1}$ with $M > 0.8M_{\odot}$ to 1.2 stars $yr^{-1}$ increases the number of systems in the synthetic sample to ~ 425,000. Changing the global scale factor in this way to achieve a better match with the observed KIC does not affect the relative distribution of the stars in the synthetic sample, but it can play a role in the target selection due to its effect on the background flux. For the following work I use the increased value of the Galaxy-wide star formation rate (SFR).

### 3.3.2 Distribution in colour-colour diagram

The distribution of KIC objects with $K_p \leq 16.0$ in the $r-i$ vs $g-r$ colour-colour diagram is shown in fig. 3.6. The left panel shows the synthetic KIC (fig. 3.6a) while the right panel displays the real KIC(fig. 3.6c) In $(g-r)$-$(r-i)$ colour space, effective temperature decreases from left to right and metallicity acts essentially perpendicular to the main band of systems, with higher metallicities having lower $r-i$. The fork at $g-r \sim 1.5$ is where the dwarfs (top branch) split from the giants (lower branch), and is located at $T_{eff} \sim 3500$ K. The distributions of the synthetic and real KIC display a reasonable agreement in the overall shape, however I found that when I applied the SCP code to the synthetic sample, the resulting derived physical parameters were very sensitive to the precise location of the stars in the colour-colour diagram.

I therefore implemented a set of corrections to force a yet better agreement between the colour-colour distributions, the result of which can be seen in the middle panel of fig. 3.6. I applied a set of three correction terms: a linear offset in each filter band, a colour-dependant term, and a Gaussian perturbation in each filter band. The rationale for this approach is provided by Pinsonneault et al. (2012) who found a linear offset and a colour dependant difference term when comparing the magnitudes measured by the KICand by the SDSS. The Gaussian perturbation applied to

all magnitudes on the other hand acts to widen the main band in the colour-colour diagram, mimicking a more realistic, continuous metallicity distribution (rather than a bimodal one) and the effect of photometric uncertainties.

I tested the corrections presented in Pinsonneault et al. (2012), which translate the KIC based magnitudes into SDSS based magnitudes, however these lead to unsatisfactory fits in the resulting SCP derived parameters. This suggests that the effects of a systematic shift between SDSS and KIC magnitudes, the simplified metallicity distribution, and superimposed photometric uncertainties can not be separated into three independent corrections that would stand on their own.

To derive the corrections I applied a least squares minimisation procedure, fitting the linear offset and colour terms simultaneously, using the distributions of the synthetic and real KIC in the following colour-magnitude diagrams: $g$ vs $(g-r)$, $r$ vs $(r-i)$, $i$ vs $(r-i)$, $z$ vs $(i-z)$, and $D51$ vs $(r-D51)$. For the Gaussian terms I also used a least squares minimisation procedure to find its width for each filter band, fitting in colour-colour space. I draw a random number from a standard normal distribution, using the same random number for each filter, scale it by the estimated width of a Gaussian centred on the magnitude derived for the object in question, and repeated this for each system. This was performed for the colour-colour distributions in $(g-r)$ vs $(r-i)$ and $(z-r)$ vs $(r-D51)$, while not allowing $r$ to vary, to derive Gaussian width coefficients for $g, i, z$ and $D51$. The procedure leads to the coefficients quoted in Table 3.2. The corrections are derived and applied to the model magnitudes after the application of the extinction model (section 2.1.2) and reddening (Table 3.1). The magnitudes for stars from the KIC, being observational derived, already have extinction and reddening applied.

Comparing the three colour-colour diagrams in fig. 3.6 it can be see that the corrections have had the desired effect. The agreement between the corrected synthetic sample (middle panel) and the reference sample (right panel) has improved in two important aspects: there is a better match of the location of the peak density, and the width of the main band has also increased. Whilst there are still some areas of

| Filter | Linear offset (l) | Colour term (c) | $\sigma^2$ |
|---|---|---|---|
| g | -0.01819 | $0.02535(g-r)$ | 0.01921 |
| r | -0.01192 | $0.05728(r-i)$ | 0.0 |
| i | -0.02209 | $0.09656(r-i)$ | 0.00995 |
| z | -0.01313 | $0.08599(i-z)$ | 0.02611 |
| D51 | -0.0222 | $-0.0571(r-D51)$ | 0.00001 |

Table 3.2: Correction terms applied to the calculated KIC magnitudes according to $X'_j = X_j + l + c + \sigma^2\phi$, where $X_j$ is the magnitude in filter bands $j = g, r, i, z, D51$ and $\phi$ is a random number drawn from a standard normal distribution.

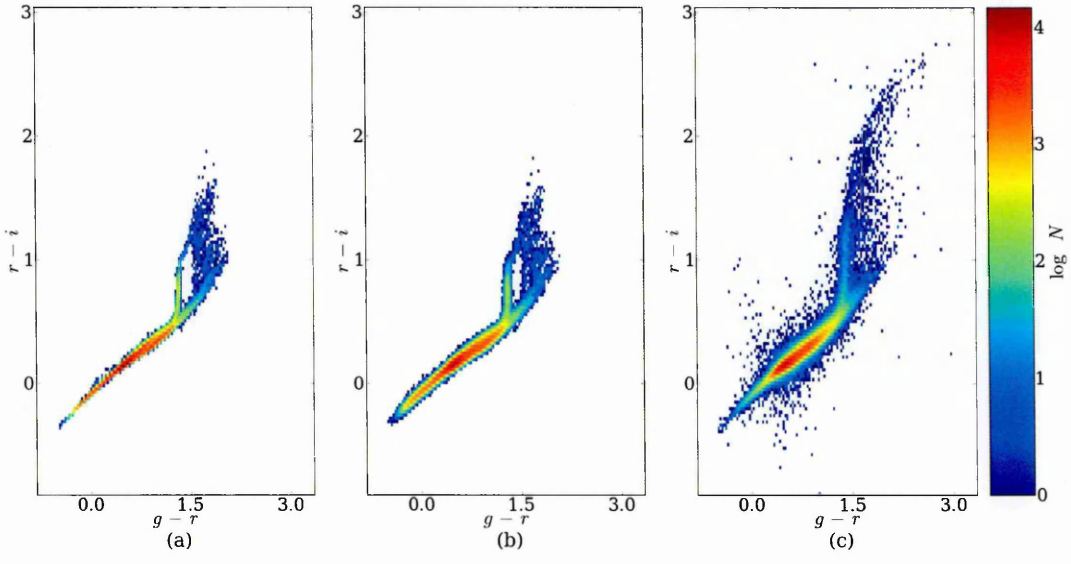



Figure 3.6: Distribution of objects in a $r-i$ vs $g-r$ colour-colour diagram. (a) The synthetic KIC sample obtained by population models (b) the synthetic KIC sample after the application of magnitude corrections terms (c) the real KIC Q2 data set.

improvement, for example there appear to be too many objects with $g-r \leq 0.6$ in the synthetic sample, which would translate into too many 'hot' dwarfs after target selection, and there is a lack of the reddest dwarfs, with $r-i > 1.5$, the bulk features of the synthetic sample are in satisfactory quantitative agreement with the reference sample for the purpose of the analysis presented below.

After publication of Farmer et al. (2013) a issue was found in the derivation of the bolometric correction (BC), specifically the metallicity used would always be the most metal poor for which I had data, which was $[M/H] = -2.5$. Thus the colours of the stars would be not be correct. However, because I had applied the colour corrections in table 3.2 which acts to smooth out differences between the synthetic sample and the KIC in colour-colour space, this issue is not as severe as it appears.

| Filter | Linear offset (l) | Colour term (c) | $\sigma^2$ |
|---|---|---|---|
| g | 0.00687 | -0.00675$(g-r)$ | 0.00703 |
| r | -0.0179 | 0.08653$(r-i)$ | 0.0 |
| i | -0.0202 | 0.09685$(r-i)$ | 0.01758 |
| z | -0.0195 | 0.09975$(i-z)$ | 0.00074 |
| D51 | -0.03211 | -0.06871$(r-D51)$ | 0.00055 |

Table 3.3: New correction terms applied to the calculated KIC magnitudes according to $X'_j = X_j + l + c + \sigma^2\phi$, where $X_j$ is the magnitude in filter bands $j = g, r, i, z, D51$ and $\phi$ is a random number drawn from a standard normal distribution.

Comparisons of the synthetic SCP parameters with the old and new BCs, after each set of colour corrections have been applied, show no appreciable difference. The corrected set of colour corrections are in table 3.3:

For the rest of this chapter I use the first set of colour corrections and the improperly derived BCs, while subsequent chapters use the new correction terms and the proper BCs.

### 3.3.3 Stellar parameter distribution

Based on the corrected magnitudes we subjected all objects in the synthetic sample to analysis with the SCP code, and thus determined their 'apparent' physical parameters, as obtained by the SCP . I can compare the actual physical properties (as determined by the population model) and the SCP -derived properties of synthetic KIC stars, and check if there are significant differences between the two. By implication, I expect that any such differences would also be present in the real KIC. For this comparison I focus on the distribution of synthetic KIC stars in the $\log T_{eff}$ - $\log g$ diagram, as these are the most reliable parameters derived from the SCP .

I first present the distribution of the actual parameters of the synthetic sample (fig. 3.7), broken up by evolutionary type. For the binaries in the sample I show the location of the primary star (except in panel e, see below). The systems occupy a region with a bird-like shape with two prominent 'wings' and a long 'neck' towards large $g$ and small $T_{eff}$. The location of this region is outlined in black in panels a-e of the figure. The 'neck' in fact consists of two narrow, essentially parallel branches

which result from the bimodal metallicity distribution in my population model. The lower branch is occupied by the lower metallicity, $Z = 0.0033$, main sequence (MS) stars, while the solar metallicity MS stars are in the upper branch. The high $T_{eff}$ 'wing' is comprised of higher-mass MS stars while the other, lower $T_{eff}$ 'wing' is comprised solely of evolved stars.

Panel a of fig. 3.7 shows the distribution of MS stars ('dwarfs'), while panel b shows Hertzsprung gap and giant branch (GB) systems. In panel c I display core helium burning (CHe) systems, and panel d shows asymptotic and thermally pulsing giant branch (AGB) systems. In fig. 3.7e I show the distribution of the secondary components in binary systems; comparing with fig. 3.7a, and in particular with the black outline, it can be seen that in general the secondaries are more clustered at the low $T_{eff}$, high $g$ end of the diagram. This implies that they in general have a lower mass or are less evolved than their primary companions, reflecting the fact that they were the lower mass component at birth of the binary.

Figure 3.7f shows the distribution of white dwarfs (WD) that are in a binary system. The synthetic sample contains no single WDs, but there are a very small number of binaries with a neutron star component (249 for the adopted input parameter set). I do not investigate the distribution of these NS systems further as the model currently treats them in a simplistic way.

I now turn to the corresponding distribution of the synthetic sample over the 'apparent', SCP -determined values for $\log T_{eff}$ and $\log g$, shown in fig. 3.8. To aid the comparison with the previous figure a grey-shaded area indicates the region the synthetic sample occupies in fig. 3.7.

Panels a-d in fig. 3.8 display the same stellar subtypes as panels a-d of fig. 3.7. The 'neck', made up of low-mass MS stars, is wider in fig. 3.8a than in fig. 3.7a, and obviously is not bimodal. The 'neck' is also at roughly constant $g$, while in the actual parameter space (fig. 3.7a) $g$ increases with decreasing $T_{eff}$. The 'wing' of higher-mass, more evolved MS systems (towards large $T_{eff}$) in fig. 3.8a is shorter than its analogue in fig. 3.7a. Comparing panel b in Figs. 3.8b and fig. 3.7b reveals that giant

branch stars extend over a similar range in $T_{eff}$ and $g$, however in fig. 3.8b some of the giants appear at low $T_{eff}$ along the 'neck', with a small gap between the bulk of the GB stars and these outliers. The CHe systems in fig. 3.8c are less constrained in $T_{eff}$-$g$ space than in fig. 3.7c while also having a small population in the 'neck'. Finally, the AGB systems in fig. 3.8d appear mostly in the 'neck' rather than the expected low $g$ 'wing' as seen in fig. 3.7d. The MS stars, or 'dwarfs' (fig. 3.7a & 3.8a) are well constrained by the requirement $\log g > 3.5$. However the more evolved objects (fig. 3.7b-d & 3.8b-d) are not constrained by $\log g$ alone. So a selection based purely by $\log g$ will be able to include or exclude dwarfs, but not giants. This has ramifications for the bulk characteristics of the exoplanet candidate systems (Gaidos & Mann, 2013). If the classification is purely based on $\log g$, then the giants that are misclassified as dwarfs will increase the number of systems with a null detection, as their radius is larger. This is because the size of the transit is proportional to the ratio of the radii, thus a larger star implies a smaller transit, thus a planet is harder to detect. Thus the bias corrected planet occurrence rate will be lower than it truly is.

There is no analogous version for panel e of fig. 3.7 as the SCP treats all objects as single stars. Instead fig. 3.8e shows how systems with a WD component would appear after the SCP analysis. I find that the resulting distribution is not significantly different from systems without a WD, confirming that there is no systematic way to identify WD systems from KIC parameters alone. This lack of difference is due to the fact that the WD's luminosity is at least a factor of 100 less than its companion's luminosity, thus its flux is negligible for the colour bands that determine the solution in $T_{eff}$ and $g$.

Finally fig. 3.8f shows the real KICstars (for Q2), with the black contour outlining the distribution of the SCP -processed synthetic KIC, demonstrating satisfactory agreement in terms of overall shape and distribution. The only significant difference remaining is the lack of a continuous giant branch track towards the lowest $g$ values.

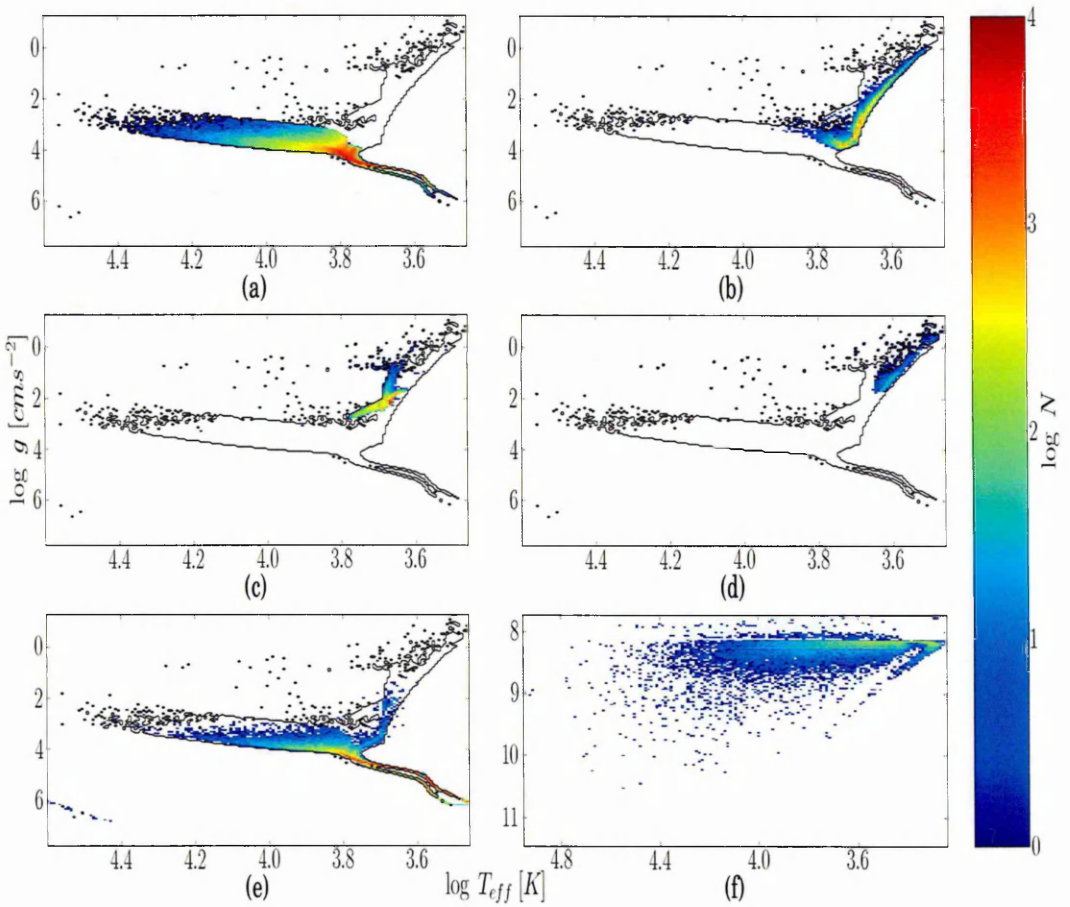



Figure 3.7: The distribution of the synthetic KIC sample over the $\log T_{eff}$ - $\log g$ plane, for different system types. $T_{eff}$ and $g$ are the actual physical parameters of the population model stars. In case of binaries the location of the primary is shown in panels a-d. The black contour outlines the region occupied by the combined synthetic sample. (a) Main-sequence (MS) stars; 'dwarfs'; (b) Hertzsprung gap and giant branch (hereafter GB) stars; (c) core helium burning (CHe) stars, (d) asymptotic and thermally pulsing giant branch (AGB) stars. (e) secondary components of a binary system, excluding systems containing a white dwarf (WD) or neutron star (NS); (f) WDs (these are all in binaries; there are no single WDs in the synthetic sample).

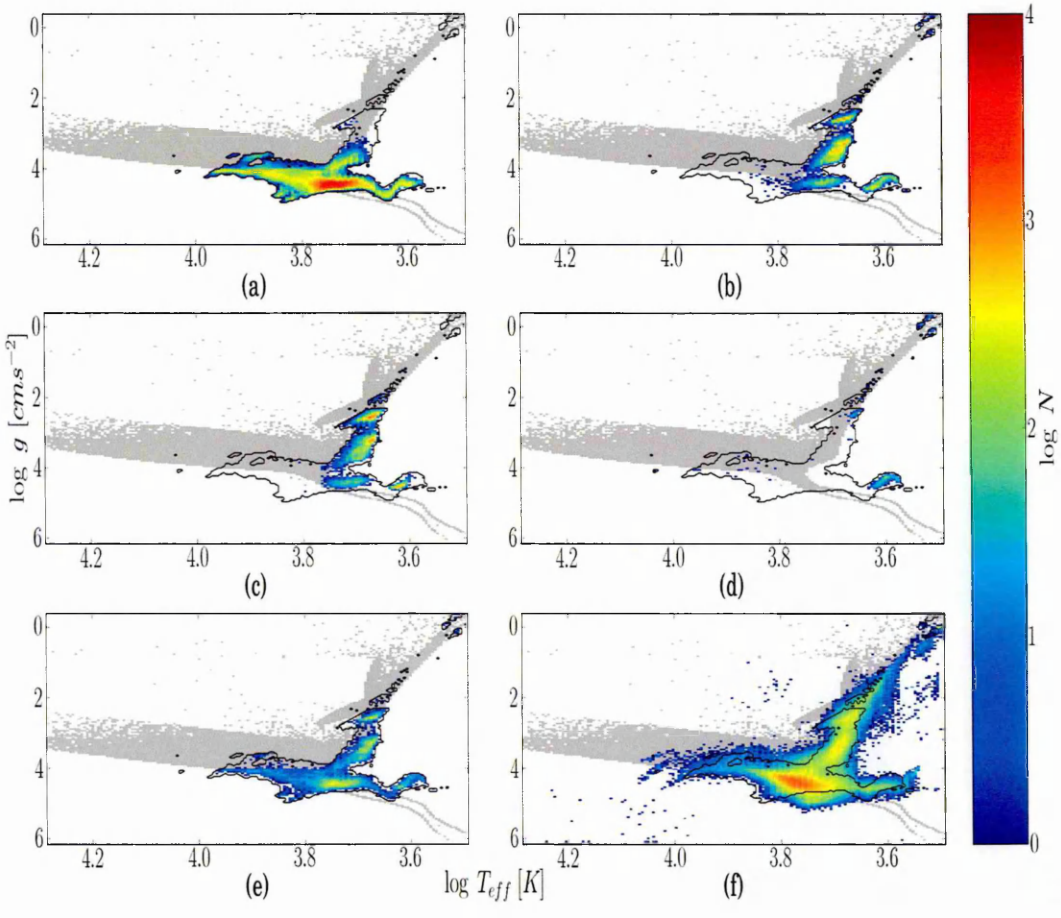



Figure 3.8: The distribution of $\log T_{eff}$ and $\log g$ derived for the synthetic sample using the procedure in the SCP , for different stellar types : (a) MS stars, (b) GB stars, (c) CHe stars, (d) AGB stars and (e) systems containing a WD. Panel (f) shows the distribution of the Q2 data set from the SCP . In all panels the contour indicates the region covered by the combined synthetic distribution, while the grey-shaded area indicates the region covered by the distributions seen in fig. 3.7

| Type | Singles | | |
|---|---|---|---|
| | Pre | Post | Relative difference |
| MS | 73.7% | 79.6% | +8.0% |
| GB | 15.7% | 10.2% | -35% |
| CHe | 10.1% | 9.4% | -6.9% |
| AGB | 0.4% | 0.6% | +50% |
| Total number | 208,697 | 104,664 | -50% |

Table 3.4: The relative distribution of stellar types among the single stars in the synthetic sample, before and after target selection.

### 3.3.4 Post-target selection distributions

After applying the target selection code described in Section 3.2.1 to the synthetic population of stars we can investigate how the target selection criteria affect the different evolutionary types of systems compared to their intrinsic distribution.

In terms of total number of objects, the synthetic KICsample was made up of 424,511 objects (208,697 single stars and 215,814 binary systems). This is reduced to 214,747 objects (104,663 single and 110,084 binaries) after target selection. The real KIC data set contains 405,789 stars while the Q2 catalogue contains 165,434 objects. Thus the specific synthetic sample I chose to work with has 5% more objects than the KICto begin with, and 20% more objects after target selection compared to the Q2 dataset. The pre-target selection number of objects could be matched perfectly by fine-tuning the underlying global Galactic SFR, but this would not affect the fraction of stars being selected as a target - ~ 50% for the synthetic vs ~ 40% for the Q2 stars.

I find that the binary fraction of the sample remains largely unaltered near the 50% level after the application of the target selection, thus I conclude that the target selection procedure does not select binaries differently than it does single stars. The synthetic sample contains slightly more binaries than single stars, due to binaries being inherently more luminous and thus a magnitude-limited sample will probe a larger volume of the Galaxy; however this difference is negligible.

Tables 3.4 and 3.5 show how the relative contribution from the different stellar and binary types change after the target selection. The relative fraction of MS and

| Type | Binaries | | |
|---|---|---|---|
| | Pre | Post | Relative difference |
| MS & MS | 68.1% | 74.8% | +9.8% |
| GB & MS | 11.7% | 6.3% | -46% |
| WD & MS | 8.0% | 8.5% | +6.3% |
| CHe & MS | 5.7% | 5.5% | -3.5% |
| WD & GB | 3.3% | 1.7% | -49% |
| WD & CHe | 2.0% | 1.85% | -7.5% |
| GB & GB | 0.28% | 0.17% | -39% |
| Total number | 215,814 | 110,084 | -49% |

Table 3.5: The relative distribution of binary classes in the synthetic sample, before and after target selection. Note this list has been truncated, the remaining types make up < 0.2% individually and 1% combined.

MS+MS objects increases by ~ 10%, while the fraction of systems containing a giant decreases by ~ 40%. The original aim of the Kepler target selection was to prioritize Sun-like stars (Batalha et al., 2010), while also removing giant stars where Earth-sized transits are harder to detect (Borucki et al., 2011a). My analysis shows that the target selection largely succeeded in this goal, and our simulations allow one to quantify the bias this procedure introduces to the stellar sample.

The fraction of single CHe stars is almost unchanged after the target selection, most likely due to the fact that most of them are misclassified into the dwarf region of $\log T_{eff}$ - $\log g$ space. The fraction of single AGB stars increases by 50% but is overall very small. The CHe+MS binary systems are also unaffected by the target selection, while binaries containing an AGB star, or CHe+GB systems, are too rare to draw conclusions from.

The change in the contribution of binaries containing a WD depends on the nature of the WD's companion. WD+MS systems are essentially unaffected. The fraction of WD+GB systems on the other hand is almost halved, which is again a consequence of the fact that the giant dominates the combined flux in the *griz* magnitudes, and thus the SCP parameter estimation is not significantly altered by the presence of the WD.

In figs. 3.8b-d there are systems that contain a giant star that sit in the 'neck', that have been misclassified as dwarfs. I define these misclassified systems as sys-

tems which are GB or AGB, or GB/AGB + GB/AGB systems as having SCP derived parameters $T_{eff,scp} < 5000$ K and $\log g_{scp} > 4.0$. In this region there are 32,929 single stars misclassified and 25,228 binary misclassified systems in the non target selected sample. In the target selected sample this is reduced to 8,622 singles and 6,494 binaries that are misclassified. This equates to 26% of the assumed MS dwarfs, in the range $T_{eff,scp} < 5000$ K and $\log g_{scp} > 4.0$, are in fact misclassified giant stars in the non target selected sample. In the target selected sample this rate increases to 35%. However, these misclassified systems make up only ~ 6% of the overall target list. The number of these misclassified systems can be estimated from asteroseismology, as seen in Huber et al. (2013). Here ~ 90% of the giants identified in a sample of planet candidate stars from Mann et al. (2012) could be identified via their asteroseismology signal as being giants, when compared to the luminosity classification in Mann et al. (2012).

### 3.3.5 Effect of target selection

The impact of the target selection can be visualised in the $\log T_{eff}$ - $\log g$ diagram by showing the ratio of the number of systems per ($\log T_{eff}$, $\log g$) bin post- to pre-target selection, for three different samples. Figure 3.9 compares the actual Q2 target list with the real KIC, fig. 3.10 considers the synthetic sample in SCP parameters, and fig. 3.11 in real parameters.

As can be seen from fig. 3.9 the target selection increased the fraction of cool dwarfs and decreased the fraction of the hotter dwarfs and of giants. The change in the density between the 'neck' and the 'wings' is due to objects in the 'neck' having $N_{tr} > 3$ for objects in their HZ.

The synthetic sample in SCP -derived parameters (fig. 3.10) has a population of dwarfs in the 'neck' which is comparable to those in fig. 3.9. The population of target-selected objects in the high $T_{eff}$ 'wing' partially matches those found in fig. 3.9, though I have many more objects there. They have SCP mass ~ $1-2M_{\odot}$ and radii ~ $1.5-4R_{\odot}$, allowing the detection of a planet at $5R_{\odot}$ that would transit 3

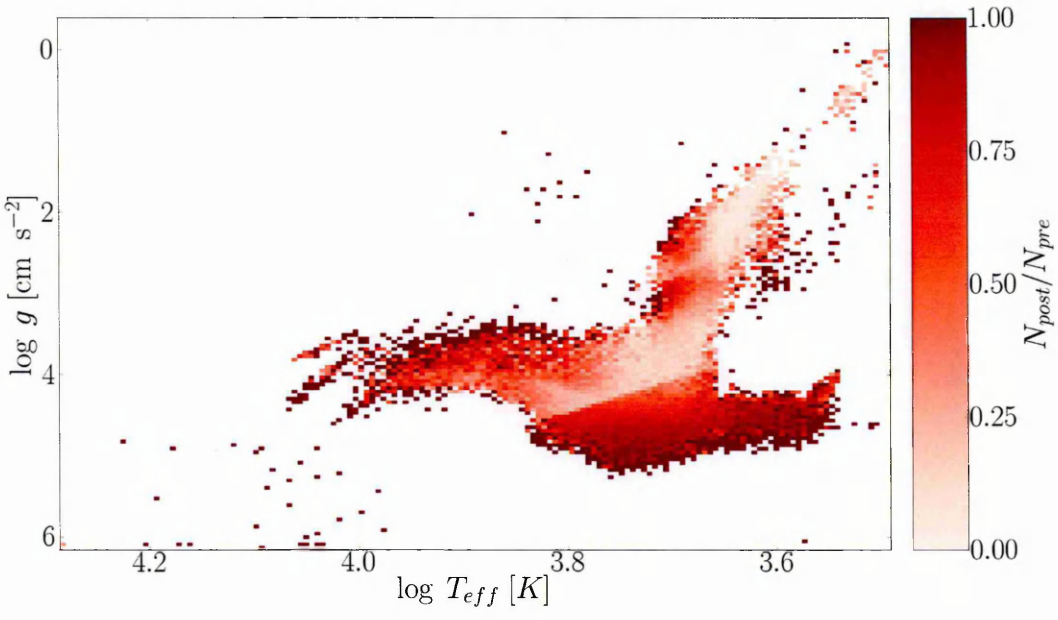



Figure 3.9: A comparison of the distribution of systems before and after the target selection, in $\log T_{eff}$ - $\log g$ space, for the actual Q2 star sample.

times in 3.5 yrs. The giants in the low $g$ 'wing' are again more marked relative to the real KIC. These are partly made up of giants that have survived the target selection criteria of Batalha et al. (2010) and partly due to the ad-hoc correction I applied to increase the number of objects with real radii $3 < R_{\odot} < 10$ (these are predominantly CHe stars). The population of giants at the lowest $g$ values is due to the ad-hoc correction that adds objects that saturate at least one pixel.

Figure 3.11 finally reveals how the synthetic sample is target-selected as a function of actual, physical parameters. The population of dwarfs in the 'neck' of fig. 3.11 matches well with the population in the 'neck' of fig. 3.9. The 'hot' dwarfs are still present in fig. 3.11. Note that the large number of target-selected objects in the GB and AGB 'wing' are due to their misclassification by the SCP (objects seen in fig. 3.8b-d in the 'neck'). They have SCP -derived $\log g$ values of 4.2-4.6 which implies an SCP -derived mass $M = 0.5 - 0.8 M_{\odot}$; hence these objects were in fact classified into my highest priority target group. The overpopulation of giants noted in fig. 3.10 is less pronounced in fig. 3.11, but here they reside in the CHe region (see fig. 3.7c) and the extreme end of the AGB region (see fig. 3.7d).

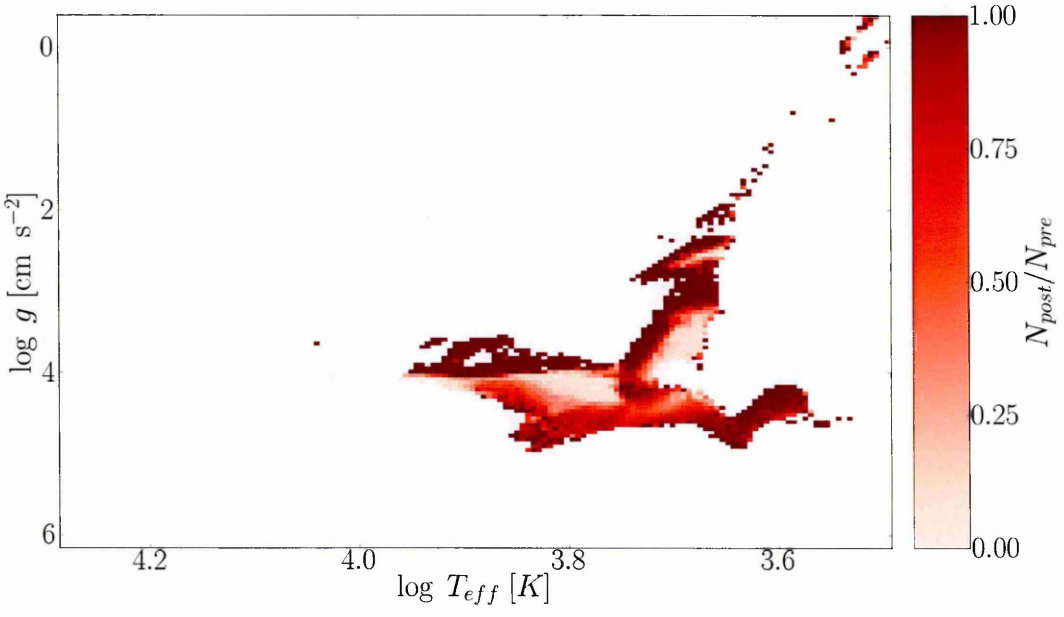



Figure 3.10: Same as fig. 3.9, but for the synthetic sample, using SCP -derived parameters.

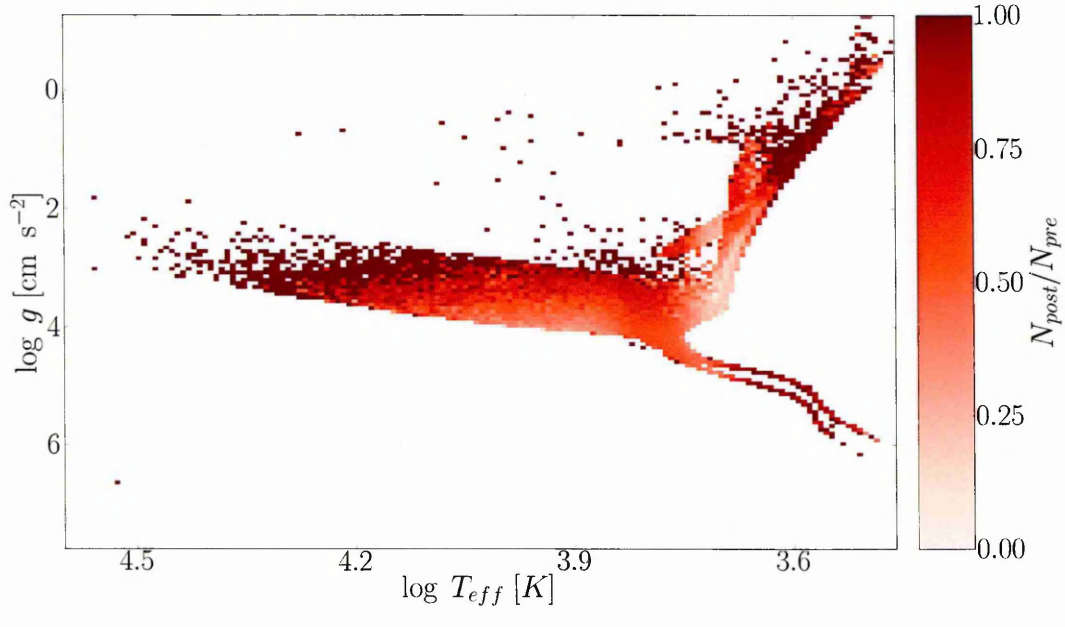



Figure 3.11: Same as fig. 3.9, but for the synthetic sample, using their correct, physical parameters.

### 3.3.6 Comparison of SCP and physical parameters

For a closer inspection of the differences between the real, physical parameters and the SCP -derived parameters I compare the synthetic sample and the KIC before target selection, as this maximises the number of objects to derive results from. For each object in the sample I determine the difference between the real and SCP -derived effective temperature, $\Delta T_{eff} = T_{eff,real} - T_{eff,SCP}$, and surface gravity, $\Delta \log g = \log g_{real} - \log g_{SCP}$. In case of a binary system only the primary star is considered. I then adopt a suitable binning of the $\log T_{eff}$ -$\log g$ plane and determine, for each bin, the median values of $\Delta T_{eff}$ and $\Delta \log g$ for all objects that fall into a given bin.

Figure 3.12 shows $\Delta T_{eff}$ as a function of $\log T_{eff}$ and $\log g$. The largest differences are seen in the hottest dwarfs. This is not unexpected as the SCP had a $T_{eff}$ limit of 50,000 K (Brown et al., 2011). Figure 3.13 displays the distribution of $\Delta \log g$ over the $\log T_{eff}$ - $\log g$ plane. The population of giants that are in the 'neck' and misclassified as dwarfs are clearly visible, having the largest $\Delta \log g$.

Tables 3.6 and 3.7 show the median values of $\Delta T_{eff}$ and $\Delta \log g$ across the whole parameter space, and the corresponding standard deviation, $\sigma$, binned on evolutionary type. MS systems (MS, MS+MS & WD+MS) have the largest values of $\Delta T_{eff} \sim 500$ K as well as the largest standard deviations, which is caused by the hot dwarfs. The evolved systems (GB, CHe & AGB-containing systems) all have relatively small values, $\Delta T_{eff} < 100$ K. Although the standard deviations are still much larger than the median values, they are $< 1/2$ those of the dwarfs.

The dwarfs have the better estimates for $\log g$, with values around $\Delta \log g \sim -0.25$, while evolved systems have $\Delta \log g$ values spread from $-0.5$ to $-3.10$ with standard deviations approximately twice that of the dwarf systems.

Comparing the differences between the single star systems and binaries of a similar evolutionary state, there are small differences in the median shift, however these are too small to be statistically significant. Therefore I conclude that the SCP has treated the binaries in a similar fashion to the single stars.

| Type | Singles | | | |
|---|---|---|---|---|
| | $\Delta T_{eff}$ [K] | | $\Delta \log g$ [dex] | |
| | Median | $\sigma$ | Median | $\sigma$ |
| MS | 492 | 918 | -0.23 | 0.39 |
| GB | 61 | 197 | -0.42 | 0.96 |
| CHe | 74 | 214 | -1.01 | 0.67 |
| AGB | -23 | 3760 | -3.01 | 1.14 |

Table 3.6: The median values of the differences $\Delta T_{eff}$ and $\Delta \log g$ (with the corresponding standard deviation $\sigma$), between the real, physical parameters and the SCP -derived parameters for the synthetic single stars.

As on average the SCP -derived $g$ is larger than the real $g$, the SCP will therefore also return a radius that is smaller than the real radius, and consequently any derived planet radius will be smaller than it really is. If $\log g$ for a MS star is overestimated by the average value of 0.23 dex and assuming all else being equal then the implied stellar radius is too small by ~ 3% and hence for a measured transit depth $\Delta F/F = (R_p/R_*)^2$ the planet radius$R_p$ is underestimated by ~ 3%. While confirmed Kepler planets will have stellar radii determined by other means, usually by spectroscopy (Batalha et al., 2011), most systems are too faint, and they are too numerous, for affordable, individual follow up (Batalha et al., 2010), thus their radii will be uncorrected in the first instance and any derived planetary distributions skewed.

This is a comparable to the quoted estimated uncertainties for the planet candidates in the Q1-Q8[3] sample, which has a median value of ~ 5% (Burke et al., 2014), though the distribution of uncertainties is skewed to those with large uncertainties such that the mean value is ~ 10%. The uncertainties are also believed to be underestimates of the true uncertainties due to correlations between fitting parameters which where not taken into account in the uncertainty calculations. This bias determined here is also comparably to the quoted uncertainty in systems which have had further follow up work and modelling (Cochran et al., 2011; Borucki et al., 2012) and those with asteroseismic measurements of the host stars (Huber et al., 2013), which provide more robust values for the stellar parameters.

Other authors find similar results. In the SCP paper, Brown et al. (2011) com-

[3]http://archive.stsci.edu/kepler/confirmed_planets/search.php

| Type | Binaries | | | |
|---|---|---|---|---|
| | $\Delta T_{eff}$ [K] | | $\Delta \log g$ [dex] | |
| | Median | $\sigma$ | Median | $\sigma$ |
| MS & MS | 558 | 931 | -0.24 | 0.41 |
| GB & MS | 56 | 268 | -0.48 | 0.72 |
| WD & MS | 471 | 615 | -0.29 | 0.38 |
| CHe & MS | 58 | 229 | -1.06 | 0.65 |
| WD & GB | 53 | 175 | -0.47 | 0.78 |
| WD & CHe | -6 | 708 | -3.10 | 0.98 |
| GB & GB | -8 | 258 | -0.5 | 0.82 |

Table 3.7: Same as Table 3.6, but for binaries and only considering the primary star.

pared the KIC estimates for some 35 stars with spectroscopic measurements, and noted for dwarfs with $T_{eff} = 4500-6500$ K a temperature difference $\Delta T_{eff} = \pm 200$ K and surface gravity difference $\Delta \log g = -0.4$ dex. Sampling the synthetic dwarfs over this $T_{eff}$ range I find a median value of $\Delta T_{eff} = +423$ K with $\sigma = 231$ K and $\Delta \log g = -0.14$ dex, $\sigma = 0.33$ dex.

Pinsonneault et al. (2012) modelled SDSS stars in the Kepler field with the infra red flux method (IRFM) and derived an average deviation of $\Delta T_{eff} = +215 \pm 100$ K for dwarf stars between 4000 K and 6500 K. Taking the synthetic systems over a similar temperature range and only considering dwarfs I find a median $\Delta T_{eff} = +413$ K, $\sigma = 240$ K, consistent with Pinsonneault et al. (2012).

Mann et al. (2012) found from medium-resolution spectra of 382 stars, $\Delta T_{eff} = -110^{+15}_{-35}$ K for dwarfs and $\Delta T_{eff} = -150^{+10}_{-35}$ K for giants. Following their selection of objects with $K_p - J > 2.0$ my synthetic sample gives $\Delta T_{eff} = -140$ K, $\sigma = 116$ K for dwarfs and $\Delta T_{eff} = +44$ K, $\sigma = 213$ K for giants, still consistent with Mann et al. (2012).

Dressing & Charbonneau (2013) used a set of Dartmouth stellar evolution and atmosphere models, with $M < 1 M_{\odot}$ and $T_{eff} < 7000$ K, to model 3897 KIC objects with $T_{eff,KIC} < 4000$ K to derive from the KIC photometry improved stellar parameters. They found for a typical cool star in their sample that $\Delta T_{eff} = -130$ K and that the radius is 69% of the KIC radius. Following their selection criteria I find that $\Delta T_{eff} = -180$ K,$\sigma = 200$ K and that the radii are 62%,$\sigma = 27\%$ of the KIC radius.

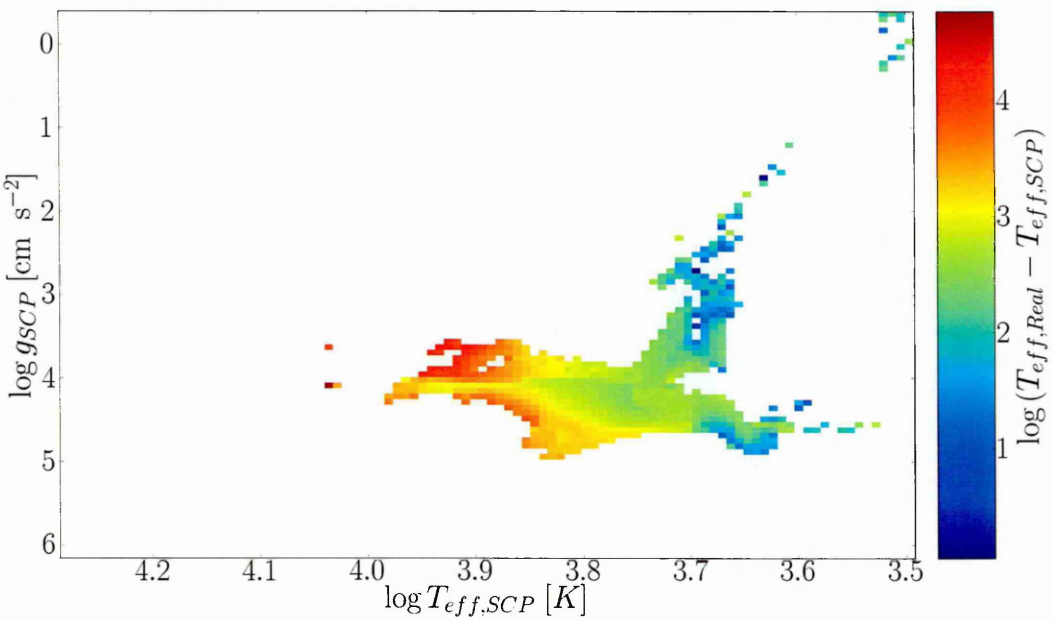



Figure 3.12: Distribution of the median of the difference between real and SCP - derived effective temperature, $\Delta T_{eff} = T_{eff,real} - T_{eff,SCP}$, per bin of $\log T_{eff}$ - $\log g$.

The apparent differences between the various authors can be explained as each author focuses on systems with different temperature and different selection criteria, such that they sample different regions of the parameter space. As an example the apparent differences between Mann et al. (2012) and Pinsonneault et al. (2012) can be explained as Mann et al. (2012) focuses on systems with $T_{eff} < 4000$ K while Pinsonneault et al. (2012) considers systems with $T_{eff} > 4000$ K.

## 3.4 Discussion

I have created a synthetic model of the KIC and of the corresponding target-selected subsample by adapting a full stellar and binary star population synthesis model to the specific circumstances of the Kepler field and the Kepler detector. In order to do so I necessarily had to adopt a number of simplifications and ad-hoc assumptions. Here I discuss the potential impact these may have on my results, and what improvements further work should consider.

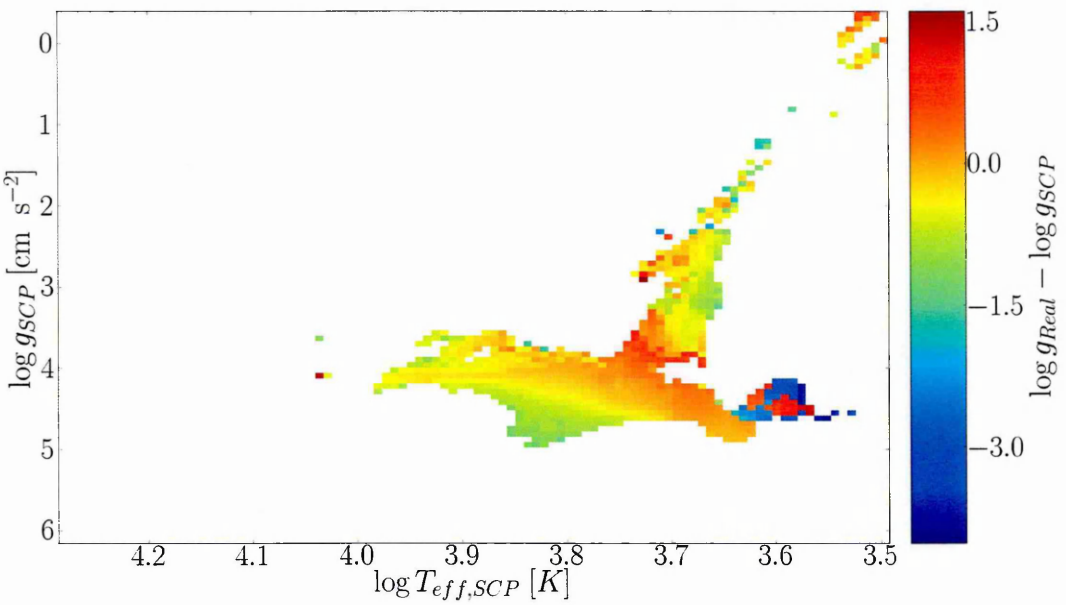



Figure 3.13: Distribution of the median of the difference between real and SCP - derived surface gravity, $\Delta \log g = \log g_{real} - \log g_{SCP}$, per bin of $\log T_{eff}$-$\log g$.

### 3.4.1 Effect of metallicity

The lack of a realistic metallicity distribution, I believe, is the most limiting simplification of the model. The current design of the population synthesis procedure makes the inclusion of an initial metallicity that continuously varies with Galactic epoch computationally too expensive (however see section 6.2 for a possible solution). The adopted bimodal model highlights the variation with $Z$, but does neither span the full range of metallicities implied by SCP fits, nor cover the bracketed range in a continuous fashion. To mimic the effect of a continuous metallicity distribution I had to introduce small, random perturbations of the calculated stellar magnitudes. This approach however cannot fully capture the effect of metallicity on evolutionary time-scales and system appearance — metal-poor stars have a shorter MS life and are less luminous than stars with solar metallicity, ultimately resulting in differences in their respective distributions in the colour-colour diagrams I set out to match.

To test the significance of the single metallicity value I used for the thin disc, $Z = 0.02$, I modelled a small FOV corresponding to one Kepler CCD channel (roughly 6500 systems) with a thin disc metallicity of $Z = 0.014$, (Asplund et al., 2009; Nordström et al., 2004). I found no statistically significant differences in the stellar pa-

rameter distribution, i.e. the differences were smaller than those seen by random sampling with 6500 systems of the underlying probability distribution function $\Gamma$.

Changing the metallicity alters the colour distribution in Fig.3.6. Lowering the metallicity leads in general to lower effective temperatures, thus shifts the stellar flux to longer wavelengths. In Fig.3.6a, lowering the metallicity to $Z = 0.014$ shifts objects with $g - r < 1.5$ upwards by 0.02, towards larger $r - i$, and systems with $g - r > 1.5$ are shifted rightwards, to larger $g - r$ by 0.2.

However, as discussed above, to force a better agreement between the synthetic sample and the real KIC I apply a series of colour-correction terms to the synthetic stars. The colour distribution for the lower thin disc metallicity will therefore simply result in a slightly different set of corrections to achieve the best fit to the actual KIC distribution seen in Fig.3.6c, thus effectively eliminating the underlying differences. I conclude that the results are not sensitive to the choice of single-value metallicity in the thin disc.

In this context I note that the SCP itself is inconsistent in its use of metallicity. In assigning a metallicity to a given object the SCP disregards the metallicity from the stellar input models (Castelli & Kurucz (2004) and Girardi et al. (2000)) and exclusively relies on solar metallicity ($Z = 0.02$) models for the BC (Brown et al., 2011).

I note that the resulting SCP-derived parameters of the synthetic stars are sensitive to the colour corrections, so great care has to be taken not to introduce spurious features into the synthetic distributions. I expect that the introduction of a realistic metallicity distribution, whilst keeping a Gaussian perturbation approach to model photometric uncertainties, would reduce these corrections to a term dependant on the difference between the SDSS magnitudes and the KIC filter system. Such a term could then be independently constrained by the findings of Pinsonneault et al. (2012).

### 3.4.2 Sensitivity to changes in Galactic parameters

For the purpose of the current study I chose to keep the population parameters fixed. There is considerable uncertainty in some of them, and I will present my work on constraining the initial mass ratio distribution (IMRD) in Chapters 5 & 4. I will however briefly discuss some of the choices and impact of the other parameters.

I tested the sensitivity of my model to the assumed Galactic distribution of stars by decreasing all scale heights and radial scale lengths by 25%, and recalculating the synthetic sample for one CCD channel. This has the effect that for a given line of sight the sample of stars are 'further' away (in units of scale lengths) from the Galactic centre, so the proportion of old disc stars, which are predominately low mass MS and WD systems, increases from 14% to 20%. After re-normalising the stellar density, to remain consistent with the KIC number count, I however find this effect to be negligible for the Kepler field, with only small differences, of order the random sampling noise, from the results for the original scales lengths. The length scale reduction considered in this test is large compared to the range proposed in the recent literature. In particular, Jurić et al. (2008) found scale heights that are consistent with my adopted values, except their thick disc scale height is 10% smaller. In fact, using their values, I also find no significant differences to the results presented here.

The total binary fraction is, somewhat arbitrarily, set at 50%. This allows me to study the differential effect of the SCP and the target selection on the binary content in general. In reality the binary fraction is likely to be a function of stellar mass, being ~ 35% for M dwarfs (Reid & Gizis, 1997) increasing to ~ 75% for G dwarfs (Duquennoy & Mayor, 1991). The choice of binary fraction becomes a more important concern when considering Kepler's sample of eclipsing binaries or the false positive transit signal and will be discussed further in Chapter 4.

The high-mass end of the IMF is well constrained, and indeed the overall normalisation of the SFR is based on this. However, the shape of the IMF below ~ 0.5 $M_{\odot}$ is more uncertain, and I have indeed tested if this offered a way to boost the number of faint objects in the target-selected synthetic sample. I found that varying the

low-mass IMF within physically reasonable limits does not significantly change the magnitude distribution of the synthetic sample.

The adopted flat IMRD is preferred by many population synthesis studies, including those by Girardi et al. (2005) (but note that these authors add binaries to the population of single stars in an ad-hoc way, while my model treats evolving binaries self-consistently). I assessed the effect of the IMRD by re-calculating a Kepler subfield population with an IMRD that favours equal mass companions ($n(q) \propto q$) and one that favours unequal mass ratios ($n(q) \propto q^{-1}$); $q$ is the mass ratio.

In the former case the population of systems with near-identical components increases. The SCP will assign the correct stellar parameters of one component to the combined object. The net effect is that a magnitude-limited sample such as the KIC will include relatively more of these objects as they are intrinsically brighter and hence can be seen out to larger distances. The apparent binary fraction does indeed increase to ~ 55% (it was ~ 50% in my standard model with a flat IMRD) in the KIC sample, and remains unchanged after target selection.

In the case of favouring unequal mass ratios on the other hand the SCP will pick out the correct parameters for the primary, but the apparent binary fraction is only ~ 3%. This is because favouring unequal component masses requires high-mass stars which are rare overall.

In both cases the apparent binary fraction remains unchanged after target selection. After correcting for the change in the total number of systems, to obtain the appropriate background flux, I find that after target selection the two different IMRDs increase (decrease) the fraction of dwarfs (giants), by ~ 5pp, in the same way as for the flat IMRD (see Tables 3.4 and 3.5). I also found that the median $\Delta T_{eff}$ and $\Delta \log g$ are the same as for the flat IMRD, within the quoted uncertainties.

In the population model I have ignored the fact that binaries form with eccentric orbits and circularise on a finite, system-dependent time-scale. Instead I kept binary orbits circular at all times. This seems justified as Hurley et al. (2002) showed that the circularisation time-scale for interacting binaries is short enough to not al-

ter bulk properties of the binary population from one where the eccentricity is kept at 0. There is also no suggestion that the eccentricity of a binary orbit would have any effect on the system's detectability in the Kepler field.

The adopted Galactic absorption model obviously affects the make-up of the synthetic KIC, but with the smallest scale modelled by Drimmel et al. (2003) being $0.35^o \times 0.35^o$ I deem this well suited to resolve the statistics of the larger Kepler field. I note that the Kepler team assumes a smooth, exponentially decaying absorbing disk (Brown et al., 2011) which on average returns a larger extinction for a given distance than Drimmel et al. (2003). The Kepler team quote that most of the target-selected stars are within 1 kpc from the Sun, with ~ 50% of objects suffering a V band extinction $A_V < 0.4$. In contrast, in my model only 30% of the target-selected stars are at < 1 kpc, while 70% are at < 2 kpc, which also corresponds to $A_V < 0.4$. Using the Kepler extinction model reduces the number of objects seen with $Kp < 16$ by ~ 5%, while leaving the underlying distributions the same, within the limits of random sampling.

In this model I have assumed that the SFR is a constant, thus differing SFRs can lead to to different results. An increased SFR in the past would lead to a larger fraction of older stars being present in the sample. This would lead to a relative increase in the number of evolved systems and those evolved systems would have lower masses. This would then alter the absolute numbers of each type of system presented in the pre/post target selection process. However, the relative change is based on the efficiency of the target selection process. This is dependant on the the local density of stars on the CCD and on the SCP derived parameters which depend on the objects colour. Thus the first term is independent of star formation rate, as I scale the total number of stars to normalise the total counts to Kepler. The second term is altered by the colour corrections. These act to smooth out any changes in the colour-colour distributions, thus minimising the change in the colour and hence the relative change in objects pre/post target selection.

## 3.5 Conclusions

I have created a comprehensive population synthesis model of the Kepler field, taking into account single and binary star evolution. I have also modelled the selection effects inherent in the Kepler objects of interest, the SCP parameter estimation, Kepler's instrumental noise and the targeted selection of systems with the highest chance of detecting an Earth-like planet round a Sun-like star in the HZ. The main output of this procedure is a synthetic catalogue of systems in the Kepler field. This catalogue was the basis for a comparison between the real physical parameters of the catalogue stars, as indicated by the population model, and the corresponding SCP-derived parameters. Such a comparison over the bulk of the Kepler field is only possible with a full theoretical population model; purely observational tests of the SCP performance will always be limited to a small sample of stars on the basis of bespoke spectral fitting. Using the synthetic sample I also investigated the effect of the target selection method on the underlying distributions in both SCP and real parameter space.

I found satisfactory agreement between the synthetic KIC and the real KIC in colour-colour space, and between my target selection method and the Q2 target selection. My simulations highlight a difference between the physical parameters of the stars in the synthetic sample and those derived by the SCP for the synthetic sample. I conclude that this systematic difference does also exist for the SCP-derived parameters of the objects in the real KIC. Specifically, for systems containing a MS star, the SCP-derived parameters deviate on average by $\sim \Delta T_{eff} = 500$ K and $\sim \Delta \log g = -0.2$ dex from the real physical parameters. In case of GB stars the deviation is $\sim \Delta T_{eff} = 50$ K and $\sim \Delta \log g = -0.5$ dex. This has the remarkable consequence that the SCP-derived stellar radii of MS stars are on average too small by $\sim 3\%$. If these radii are used to estimate the radius of any planet observed to be transiting then the planet radius will be $\sim 3\%$ too small. This bias is of the same order as the median quoted uncertainties in the Kepler data.

After correcting for selection effects I find that these results are consistent with

differences highlighted by other authors, on the basis of observational consideration of subsamples. The average deviation for a given stellar type is observed regardless of if the star is single or in a binary.

The models confirm that the Kepler target selection procedure increases the fraction of main-sequence stars, from about 75% to 80%, and decreases the fraction of giants, from 25% to 20%, relative to the KIC. I have also found a small proportion of the giants are misclassified as dwarfs, which implies that the Kepler derived $\eta_{earth}$ will be larger than predicted, due to the increase in the number of systems that may have a null detection. In fact, my population synthesis approach is the only way to quantify this bias; the figures demonstrate that the change is only moderate. It has also been shown that a number of the giants in the original Kepler target list were inserted in an ad-hoc manner as opposed to being based on their transit detectability (Pinsonneault priv.comm), which is similar to the ad-hoc corrections presented here.

The bias introduced into the target-selected sample is roughly the same for single stars and binary systems. I also found that the target selection has a negligible effect on the binary fraction, and that it does not alter the relative fractions of systems with different stellar evolution types, when compared to the single star population.

# Chapter 4

# Eclipsing Binaries in the Kepler field

In this Chapter I create a model which generates realistic light curves for eclipsing binaries (EBs) in the Kepler field. I then apply the Kepler target selection and Kepler's EB detection criteria to derive the population of EBs predicted to be detected by Kepler. I then investigate the dependence of these results to different initial mass ratio distributions (IMRDs) and initial binary fraction distributions (IBFDs).

I find that the Kepler EB parameter determination of the temperature ratio and fraction radii, has a large bias in it, when comparing the true parameters to the predicted parameters for synthetic light curves. This has implications for predicted rate of false positives assumed by Kepler, which relies on the Kepler EB catalogue as an input. This bias I believe is due to the POLYFIT code which re-samples the light curve into equidistant phase points. This is due to POLYFIT fitting the eclipse depths poorly which then results in a poor fit to the temperature ratio. Comparing the synthetic EB population to the Kepler EB catalogue I find a approximate match in terms of the number of objects found, however my models have too many short period systems. At this time I can not determine which IMRD or IBFD model fits the data best, due to the overall poor fit to the data. However, I do show that there is sufficient difference between the different IMRD, but not IBFD, models that with further modeling it should be possible to determine the IMRD from the Kepler data.

## 4.1 Introduction

During the first 2 quarters of Kepler's observation it has discovered over 2000 EBs, where one star passes in front of the other from our line of sight. Increasing the number of known EBs has a range of benefits for astrophysics; from understanding stellar evolution (Torres & Ribas, 2002), through precise determination of stellar parameters like the mass and radii (Hilditch et al., 2005); understanding systems that can mimic planetary transits (Fressin et al., 2013), required to complete Kepler's primary mission (Borucki et al., 2011b) and understanding poorly known binary evolution parameters (Chabrier et al., 2007), for instance tidal interactions (Claret & Cunha, 1997).

Kepler has detected many interesting types of EB such as post common envelope binaries (Almenara et al., 2012), eccentric binaries with tidally induced pulsations the so called "Heartbeat" stars (Thompson et al., 2012), systems with $\delta$-Scuti-like oscillations (Lehmann et al., 2013) and triple star systems (Steffen et al., 2011). Systems with pulsations are especially useful as they allow the derivation of independent constraints on the system parameters. Thus the EBs with pulsations can have their parameters determined via the pulsations and via their eclipses. This allows us to bridge our knowledge of stellar parameters from the two different fields (Charpinet et al., 2008).

The process of compiling the Kepler EB catalogue has multiple steps. First, the light curves are subjected to Kepler's transit planet search (TPS) pipeline (Jenkins et al., 2010a) which searches the light curves for single events that look like transits. These single events are then phase folded at different orbital periods to determine whether they pass a maximum multiple event statistic, which determines whether multiple single events can be considered as a periodic event. Those that pass the test are then classified as threshold crossing events (TCE s) and contain both EBs and planetary transits. The TCE s are then filtered for systems that are planets, based on the size and duration of transits as well modelling the flux centroids to determine changes in the centre-of-light distribution (Batalha et al., 2011) to exclude systems

that are clearly not planets. The filtered systems are then subject to further filtering and modelling to determine whether they are an EB or a pulsating system (Prša et al., 2011). Those that are EBs are then fed into the Kepler EB catalogue. EBs can also be added to the catalogue if they are already known as an EB or if after further observations a candidate planetary systems turns out to an EB (Prša et al., 2011).

To determine the parameters of the EBs, the light curves for each system where passed through the Eclipsing Binaries via Artificial Intelligence (EBAI) code (Prša et al., 2008). EBAI is a neural network based algorithm which after being trained on a set of light curves with known binary & stellar parameters is designed to determine the parameters from light curves with unknown parameters. This process introduces a bias into the determined distribution. Prša et al. (2011) claim, based on a a set of synthetic light curves that where used to train EBAI, that ~ 90% of their systems will have errors < 10%. However, the parameter distributions as determined by EBAI in Prša et al. (2011) do not resemble those in Prša et al. (2008), which are based on OGLE data.

Thus this chapter is concerned with building a synthetic population of EBs, applying a model of the detection process and using the EBAI code to determine the Kepler EB parameters of the synthetic EBs. The aim of this work is to understand the size of the bias introduced and possible selection effects inherent in the Kepler EB pipeline. By applying the EB model and the Kepler selection effects to the synthetic stars, I can perform a statistical comparison between different underlying populations of *non*-eclipsing binaries with the Kepler EB catalogue to constrain initial Galactic distribution functions.

In this chapter I update my population synthesis method, to handle a new classification of binaries not previously considered (section 4.2.1). I then derive a synthetic EB sample by synthesising eclipse light curves for the EBs, which are then combined with a model for the noise in the light curve (section 4.3.1). The noisy light curves are then passed though a set of constraints aimed to determine which would be "detected" and then classified as EBs (section 4.3.3). Finally, the "detected"

light curves are passed to EBAI to compute their Kepler derived stellar parameters which are then compared to the true parameters (section 4.4.4).

## 4.2 Method

To create the synthetic EB catalogue I have updated the model from that used in section 3.2. The extinction model has been improved and there is now a treatment for binaries with very low mass companions. I have also attempted to include contact binary evolution into the binary star evolution (BSE).

### 4.2.1 Updates to the population synthesis method

The extinction coefficient $A_\lambda$ that translate the $A_v$ extinction, as computed in Drimmel et al. (2003), from the $V$-band extinction to the Kepler filter band is now computed via interpolation of tables given in terms of $T_{eff}$, $\log g$ and $A_v$ (Girardi et al., 2008). This is instead of the single value for all stars as used in section 3.2.1. This change is due to improvements in the code that now allow $A_\lambda$ to be derived for individual stars. Due to changes discussed later in section 4.2.2 the star formation rate (SFR) is now normalised to produce 8.3 stars $yr^{-1}$, to maintain consistency with the Kepler star counts.

### 4.2.2 DERBs

One of the assumptions made when using the initial formation distributions (section 2.1.4) is that the stellar evolution track library, derived in section 2.1.1, contains all possible initial systems that the distribution functions are normalised over. This however is not the case for the IMRD. The IMRD in section 2.1.4 is normalised between 0 and 1 as:

$$IMR(q) = (s+1)\, q^s \tag{4.1}$$

where $q = M_2/M_1$. For $q \to 0$ this requires $M_2 \to 0$ for a given $M_1$, but the single star evolution (SSE) code has a lower mass limit of $0.1 M_\odot$. This means that for a given $M_1$

there exists only stellar evolution tracks for systems with $q$ between $0.1/M_1$ and 1. Any system with $q < 0.1/M_1$ is counted in the total number of systems formed, but is not included in the outputted $\Gamma$ distribution as there is no corresponding track. The reason these "missing" systems become important now, is that varying the IMRD to a negative value ($s \sim -1$) implies most binaries form with an extreme mass ratios for which I do not have stellar evolution tracks for and thus the predicted number of systems drop.

The way around this problem is to extend the stellar evolution modelling to systems below the lower mass limit of $0.1M_\odot$. This however would be a considerable piece of work in itself, thus a simpler approach was taken. Stars with masses below $0.1M_\odot$ undergo little stellar evolution on Galactic time scales, predominately undergoing cooling (Chabrier et al., 2000). They would also in general be much fainter than their companions, especially in the optical Kepler band pass, as they emit predominantly at infra-red wavelengths (Chabrier et al., 2000). So to first order I assume that the binaries with a low mass companion undergo no binary induced evolution (for example mass transfer) or interactions (for example tides) and that the luminosity of the system can be taken as that of the primary alone. This allows the use of the single star evolution tracks as templates for these systems, but with the integrals over the initial distribution functions taken for a binary (see equation 4.4).

These systems I classify as DERBs, Diffuse Extreme Ratio Binaries, where the companion has a mass between $0.07 < M_2/M_\odot < 0.1$, where the limits will be discussed later. These are "diffuse" as they represent systems where the secondary has a unknown but bounded mass. As these are modelled solely on the single star tracks, the period evolution can not be tracked nor can interactions between the stars be derived, thus the integral over the initial orbital separation distribution (IOSD) is taken as 1. Therefore a DERB represents systems at all initial orbital separations, for a given initial $M_1$ and $M_2$. Note we must consider all possible primary stars, which have a DERB companion. This is due to the requirement that the integrals are normalised and that we have stellar evolution tracks covering the the entirely of the

distribution functions validity, even if the formation or detection of the system is improbable.

I take each single star track, from section 2.1.1, and weight them by the initial mass function (IMF) (equation 2.4) and then weight them by a modified version of equation 4.1.The modified IMRD is expressed as:

$$IMR(q) = \frac{s+1}{1-q_{min}^{s+1}} q^s \tag{4.2}$$

This is equation 4.1 but now normalised between $q_{min}$ and 1.0, where $q_{min} = \frac{0.07}{M_1}$. This now assumes the minimum mass of the secondary, for which the IMRD is valid, is $0.07M_\odot$, rather than $0.0M_\odot$. This limit was chosen as it is the limit for hydrogen burning and commonly taken as the lower mass limit of a star. The DERBs are thus systems with $0.07/M_1 < q < 0.1/M_1$ and binaries are defined over the range $0.1/M_1 \leq q \leq 1.0$.

Stars in this range of masses have considerable uncertainty regarding their formation probability. Their is evidence for the IMF turning over at low masses (Scholz et al., 2013), though I make the simplifying assumption that the distributions chosen for the higher mass stars are still valid in this region. There is evidence that the Brown dwarfs (BD), which have $M < 0.07M_\odot$ are described by a different set of initial distribution functions, believed to be due to the difference in formation mechanism (Scholz et al., 2013). Thus by assuming that stars with $M > 0.07M_\odot$ follow the same distributions as those with $M > 0.1M_\odot$ I am implicitly assuming that these low mass stars have formed in the same manner as the higher mass stars.

This parametrisation requires a change in the integration methods used in section 2.1.1. Previously, each evolution track corresponded to a point in initial parameter space, that is then used to represent all systems in a bin containing the initial configuration of this track. The integral over the multi dimensional initial parameter space was trivial as each distribution was independent from each other. Now however the IMRD is dependant on both $M_1$ and $M_2$ thus the integration method described in section 2.1.4 is no longer valid. This used a numerical trapezium rule

but now I solve the multi dimensional integral with a combination of analytical and numerical methods. The IMF for single stars and the IOSD for binaries are solved analytically. For the binaries and the DERBs the integral is transformed from one that is integrated over $M_1$ and $q$ in equation 2.13 into one that is integrated over $M_1$ and $M_2$.

The integral over the IMF and IMRD for the binaries and the DERBs is expressed as:

$$D = \int\int IMF(m_1) \times IMR(q) \mathrm{d}q \mathrm{d}M_1 \tag{4.3}$$

Subsisting equation 4.2 into 4.3 and expanding the $q$ terms:

$$D = \int\int IMF(m_1) \times \frac{s+1}{1-\left(\frac{0.07}{M_1}\right)^{s+1}} \left(\frac{M_2}{M_1}\right)^s \mathrm{d}q \mathrm{d}M_1 \tag{4.4}$$

Changing the volume element of $\mathrm{d}q\mathrm{d}M_1$ into $\mathrm{d}M_1\mathrm{d}M_2$, as the evolution tracks are defined for bins in $M_1$ and $M_2$, requires a co-ordinate transformation of $M_1 \rightarrow M_1$ and $q \rightarrow M_2/M_1$. Thus the Jacobian is:

$$\begin{vmatrix} \frac{\mathrm{d}M_1}{\mathrm{d}M_1} & \frac{\mathrm{d}M_1}{\mathrm{d}M_2} \\ \frac{\mathrm{d}q}{\mathrm{d}M_1} & \frac{\mathrm{d}q}{\mathrm{d}M_2} \end{vmatrix} = \left|\frac{1}{M_1}\right| \mathrm{d}M_2 \mathrm{d}M_1$$

Therefore overall I have:

$$D = \int\int IMF(m_1) \times \frac{s+1}{1-\left(\frac{0.07}{M_1}\right)^{s+1}} \left(\frac{M_2}{M_1}\right)^s \left|\frac{1}{M_1}\right| \mathrm{d}M_2 \mathrm{d}M_1 \tag{4.5}$$

I solve equation 4.5 numerically using a two-dimensional Romberg integral (Press et al., 1992). The limits are handled in a similar fashion to the section 2.1.3. During the integration routine I determine the value of $D$ (equation 4.5) over a small region of $M_1, M_2$ space. This region is then weighted by a function that determines the fraction of that region that falls within the integration limits. The limits are thus that the minimum and maximum mass of the primary is $0.1 < M_1/M_\odot < 20$ and that the secondary is less massive than the primary, $M_2 \leq M_1$. For the binaries there is an

| s | Number | | | |
|---|---|---|---|---|
| | Singles | Binaries | DERBs | Total |
| -0.5 | 186,687 | 198,174 | 11,767 | 396,628 |
| 0.0 | 186,731 | 217,244 | 5,202 | 409,177 |
| 0.5 | 186,798 | 231,146 | 2,220 | 420,164 |
| 1.0 | 187,129 | 237,299 | 819 | 425,247 |

Table 4.1: The number of DERBS and binaries for different values of $s$ for the synthetic Kepler field of view. The final column shows the sum of the two totals, showing how the total number of systems remains constant with different values of $s$

extra region, the minimum and maximum mass of the secondary is $0.1 < M_2/M_\odot < 20$. While for the DERBs the secondary mass limit is $0.07 < M_2/M_\odot < 0.1$.

For the moment I have only considered binaries with the secondary star having $M < 0.1M_\odot$, their are of course single stars and primary stars in binaries in the same mass range. I ignore the low mass single stars as a single star with this mass would be almost undetectable at optical wavelengths. Thus the low mass stars would only alter the synthetic population via the normalisation of the IMF (equation 2.4), but this is a constant for all values of $s$. Therefore I can simply alter the SFR to compensate for the reduction in the number of stars. Binaries where the primary has $M < 0.1M_\odot$ must have a secondary with $M < 0.1M_\odot$. These would also be undetectable in the optical. However the normalisation would now be dependant on $s$ and as such should not be ignored. For the moment I ignore this affect as I due not have an evolutionary model for these low mass systems. This is an area for further work.

Fig. 4.1 shows the distribution of formation probabilities for different values of the IMRD parameter $s$. It can be seen that the DERBs, which have previously been ignored, contribute a non-negligible component to newly forming systems. Table 4.1 shows that the new normalisation scheme keeps the number counts between the different IMRDs relativity constant, unlike in section 3.4.2. The resulting difference in the number of binaries is due to the fact that equal mass binaries will be more luminous than non equal mass systems, thus we see a greater fraction of their Galactic volume.

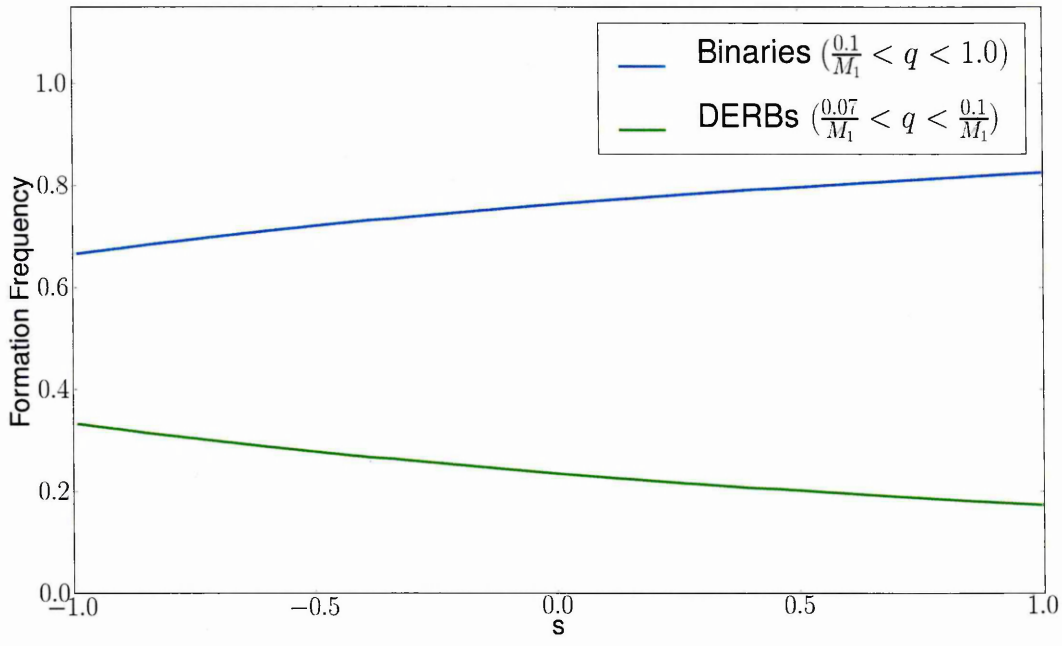



Figure 4.1: The relative formation probabilities for binaries (excluding DERBS) and DERBs as a function of the IMRD parameter $s$ for an arbitrary field. The distributions include a Kroupa based IMF (equation 2.4).

### 4.2.3 Contact Binaries

In the Kepler EB catalogue there are 469 (~ 20%) contact binaries out of 2165 detected EBs (Slawson et al., 2011). These systems therefore represent a significant population of objects in the Kepler EB catalogue, which currently can not be included in the synthetic EB catalogue. I thus attempted to introduce a simplified model of them into the model, however this was unsuccessful and not pursued further.

Following the work in Yakut & Eggleton (2005), I forced the system to undergo a thermal-timescale roche-lobe overflow (RLOF) from the secondary star, when the system is in contact. The system was allowed to continue evolving as it would under normal RLOF conditions. This is a simplification of the actual process undergoing in contact binaries; Lucy (1968) suggests that there should be energy transport from the primary to the secondary such that secondary appears brighter than it would if not in contact. I ignore this detail, as the main quantities relevant for population synthesis work is the total luminosity of the system and the lifetime of the system. I assume the total luminosity would be unchanged and the lifetime determined by

the thermal time-scale. This simplified approach however fails to produce a usable stellar evolution track. The time step d$t$ which varies as function of the change in the mass, radius and angular momentum, is forced towards a hard-coded minimum (100 years) which is then too small to perform stellar evolution over the giga-years necessary for contact systems (Stępień & Gazeas, 2012).

To solve this I would need a full stellar evolution code where the mass and energy distribution in the outer envelopes can be modelled. This would then allow a more complex treatment of the mass and energy transfer at the point of contact, as well as the response of each star to the changes in mass and energy. Surprisingly, a few systems in my simplified model undergo what appeared to be thermal relaxation oscillations (TRO), where heat is transported from the more massive component to the least massive. This leads the least massive star expanding and reversing the mass transfer back to the more massive component; the systems orbit then expands and contact is broken (Yakut & Eggleton, 2005). The more massive component then expands as it is no longer transferring heat to its companion and the cycle continues. This shows that it may be possible with further work to model contact binaries using BISEPS with a more robust treatment of mass and energy transfer.

Without a model for the contact binaries the synthetic EB catalogue will have $\sim$ 20% fewer systems than the the true Kepler EB catalogue. However this $\sim$ 20% fewer EBs does not equate to $\sim$ 20% fewer binaries overall. This is because the contact binaries have a large range of possible detectable eclipse inclinations, thus they will be preferentially detected. Rucinski (1998) suggest that there are $\sim$ 1% the number of contact binaries compared to MS stars. Extrapolated to the number of systems in the synthetic Kepler input catalogue (KIC), this implies there are $\sim$ 3000 contact binaries in the field, which is small enough to ignore for the purposes of the normalisation. Extrapolating from the Kepler result of 496 contact binaries and given that contact binaries have a minimum inclination of $\sim 35°$ for a grazing eclipse (Morris, 1999), implies there should be of order $\sim$ 1000 contact binaries. Considering this assumes we can detect all eclipses, thus underestimates the minimum inclination

and therefore underestimates the total we see, this would appear consistent with the number of systems found in Rucinski (1998).

## 4.3 Derivation of an eclipsing binary sample

Starting from the population method presented in Chapter 2 and the Kepler target selection in section 3.2 I now present the method for deriving a synthetic EB sample which is detectable by Kepler. This process is divided into several steps: derive a sample of EBs, generate a synthetic light curve with a representative level of noise, compute the detectability of the eclipses, determine whether the detected eclipses would be flagged as EB and finally compute the Kepler derived parameters of the EB.

### 4.3.1 Stellar noise

Almost all stars are variable. They may be undergoing large scale variations in flux output (Miras (Feast et al., 1989), Cepheids (Freedman et al., 1994)) on time scales of days and years and small scale variations (granulation (Michel et al., 2008), solar like asteroseismic variations (Chaplin et al., 2011c)) on time scales of minutes and hours. Thus these variations in the flux act as a noise source reducing the strength of the eclipse signal and in low signal cases may entirely mask the eclipse. I have built a model of the expected variability so that I can compare the signal to the noise for the eclipse and determine the detectability of the eclipse. This model is partially based on the derivation of the combined differential photometric precision (CDPP), the estimate of all noise sources in Kepler (Gilliland et al., 2011). This work does not claim to employ a definitive model for stellar variability. Rather it should be considered a order of magnitude estimate, suitable for estimating the average S/N of a light curve.

I apply two types of noise to the synthetic light curves, the intrinsic variability of the stars and the measurement uncertainties. The noise due to the measurement is

taken from section 3.2.2. While the noise due to the stellar variability is derived now. For this work I concentrate only on the small scale variations and assume that the larger scale variations have been removed from the data. For main sequence (MS) stars I consider three variability modes; granulation, solar like oscillations (asteroseismic signals) and chromospheric activity. For giant branch (GB) like stars (which I define as stars in any phase from the hertzsprung gap (HG) to the thermally pulsating asymptotic giant branch (TPAGB)) I consider only the granulation and asteroseismic variability, as evidence for magnetic activity is limited in evolved stars (Konstantinova-Antova et al., 2010). For WD and NS I assume the variability is 0, due to their low luminosity compared to a MS or GB companion. For simplicity I ignore for the moment any binary induced variability. Star spots are also ignored for this work to simplify the calculations, but is an area of further work. These are regions on a stars surface of high magnetic field, which inhibits convection and then lowers the local temperature thus appearing dark. These can then appear as transit-like features in a light curve of a star.

To compute the "noise" in a light curve due to variability I wish to derive the RMS ($\sigma_{var}$) of the variance, $V$, of the variability mode in the *time* domain. However, the individual modes (granulation, activity, asteroseismic oscillations) are defined in terms of a characteristic time-scales, $\tau$, and RMS amplitudes in the *frequency* domain. To translate between the two domains we can use Parseval's theorem (Press et al., 1992) which shows that the power in the frequency domain is the same as in the time domain. Thus to compute the total power of the oscillations in then time domain we simply need to compute the power $P$ in frequency domain. To do this we integrate the power spectral density $P(\nu)$ over a set of measurement frequencies $\nu$. For Kepler I set the range of measurements frequencies to be between 1 LC observation (30 minutes or $555.0\mu\text{Hz}^{-1}$) to 1 quarter (3 months or $0.13\mu\text{Hz}^{-1}$)

Thus we have that:

$$P = \eta^2 \int_{0.13\mu\text{Hz}}^{555.0\mu\text{Hz}} \frac{2\sigma^2\tau}{1+(2\pi\nu\tau)^2} d\nu \tag{4.6}$$

where $\eta$ is a function designed to suppress the measured power the closer the oscillation frequency ($\nu' = 1/\tau$) is to the Nyquist frequency ($555.0\mu\text{Hz}^{-1}$).

$$\eta = \text{sinc}\left(\pi/2\left(\frac{\nu'}{555\mu\text{Hz}}\right)\right) \tag{4.7}$$

The total "noise" term is thus the quadrature sum of each individual terms:

$$\sigma_{var} = \sqrt{(P_{gran} + P_{ast} + P_{chrom})} \quad \text{ppm} \tag{4.8}$$

for a MS star, where $\sigma_{var}$ is the RMS of the variance $V$. The $P_{gran}$ is the power due to the granulation, $P_{ast}$ is the power due to the solar like asteroseismic signals and $P_{chrom}$ is the power due to the chromospheric activity. The power terms are already expressed in terms of square quantities.

We can however make several simplifications to equation 4.6 for the granulation and asteroseismic signals.

The asteroseismic oscillations are comprised of $p$ mode oscillations, where the pressure acts to restore the star after a perturbation due to sound waves (Bahcall et al., 2001). The star then oscillates in either the radial (l), radial (n) and azimuthal (m) modes or as a combination of the three. The total power of the oscillations is then due to the contribution from each of these modes. We assume that the power can be modelled as a Gaussian centred on the frequency at which the observed modes have their largest amplitudes $\nu_{max}$. We then only consider the oscillations around this frequency, as opposed to all frequencies as in equation 4.6. $\nu_{max}$ is defined as:

$$\nu_{max} = \nu_{max,\odot}\left(\frac{R}{R_\odot}\right)^{-2}\left(\frac{T_{eff}}{T_{eff,\odot}}\right) \tag{4.9}$$

where $\nu_{max,\odot}$ is the solar value, $\nu_{max,\odot} = 3150\mu\text{Hz}^{-1}$.

The asteroseismic signal is then assumed to be contained within the envelope of the Gaussian centred at $\nu_{max}$ and with a FWHM of $\nu_{max}/2$ (Kjeldsen et al., 2008). Assuming that most of the power then is within the FWHM, we can add the amplitudes from the $l = 0, 1, 2, 3$ modes, each of which scale as a function of $A^2_{max}$ (Chaplin

et al., 2011c). Then we add in the contributions from the $n$ modes (The m mode contribution is assumed to be averaged over in the l and n contributions). Where there are $\nu_{max}/\Delta\nu$ total segments contributing to the average power. Therefore overall the total observed oscillation power is (Chaplin et al., 2011b):

$$P_{ast} = 1.55 A_{max}^2 \eta^2 \frac{\nu_{max}}{\Delta\nu} \tag{4.10}$$

with $A_{max}$ the maximum amplitude of the asteroseismic signal and $\Delta\nu$ is the frequency difference between radial, $n$, overtones (Chaplin et al., 2011c):

$$\Delta\nu = \Delta\nu_{\odot} \left(\frac{M}{M_{\odot}}\right)^{0.5} \left(\frac{R}{R_{\odot}}\right)^{-1.5} \tag{4.11}$$

with $\Delta\nu_{\odot}$ being the solar value, 134.9$\mu$Hz.

The maximum amplitude of the asteroseismic signal can then be computed as:

$$A_{max} = 2.5\beta \left(\frac{R}{R_{\odot}}\right)^2 \left(\frac{T_{eff}}{T_{eff,\odot}}\right)^{0.5} \tag{4.12}$$

Where $\beta = 1 - e^{(-(T_{red}-T_{eff})/\Delta T)}$, $T_{red} = T_{red,\odot}(L/L_{\odot})^{-0.093}$, $T_{red,\odot} = 8907$K and $\Delta T = 1550$K. $\beta$ is a fitted correction factor used because the predictions of the amplitude get progressively worse as a star nears the $\delta$-Scuti instability strip.

Combining all the terms leads to (Chaplin et al., 2011b):

$$P_{ast} \approx 225 \text{ppm}^2 \beta^2 \eta^2 \left(\frac{R}{R_{\odot}}\right)^{3.5} \left(\frac{T_{eff}}{T_{eff,\odot}}\right)^{1.25} \tag{4.13}$$

For the granulation term, it can be shown that $\tau_{gran}/\tau_{gran,\odot} = \nu_{max,\odot}/\nu_{max}$ (Huber et al., 2009). By considering the granulation at $\nu = \nu_{max}$ the denominator in equation 4.6 is constant for all stars. Substituting in the solar value of $\tau_{gran,\odot} = 210s$ and $\nu_{max,\odot} = 3150\mu\text{Hz}^{-1}$ (Chaplin et al., 2011b):

$$P_{gran} \approx 0.1 \sigma_{gran} \tau_{gran} \text{ppm}^2 \mu\text{Hz}^{-1} \tag{4.14}$$

where the $\sigma_{gran} \times \tau_{gran}$ term can be expressed in terms of the mass, radius and luminosity (Kjeldsen & Bedding, 2011):

$$\sigma_{gran} \times \tau_{gran} \propto \left(\frac{L}{L_\odot}\right)^2 \left(\frac{M}{M_\odot}\right)^{-3} \left(\frac{T_{eff}}{T_{eff,\odot}}\right)^{-5.5} \tag{4.15}$$

which can be re-expressed in terms of $\nu_{max}$ and substituting into equation 4.9:

$$P_{gran} \approx 0.1 \left(\frac{\nu_{max}}{\nu_{max,\odot}}\right)^{-2} \text{ppm}^2 \mu\text{Hz}^{-1} \tag{4.16}$$

Thus for both $P_{ast}$ and $P_{gan}$ we can approximate their power over the whole frequency range, ignoring the frequency dependency for simplicity, as $P_{ast|gran} \times \nu_{max}$.

The power in the chromospheric activity can not be simplified so we must use the full expression in equation 4.6. $\sigma_{chrom}$ is RMS amplitude of the activity and is estimated from the the Ca H and K emission index $R'_{HK}$ (Gilliland et al., 2011). $\tau_{chrom}$ is the time scale of the activity, which for simplicity is assumed as the solar value of $\tau_{chrom,\odot} = 8$ days (Gilliland et al., 2011). Thus $\sigma_{chrom}$ can be expressed as, based on an empirical fit to the *SOHO* data from the Sun (Gilliland et al., 2011):

$$\sigma_{chrom} = 6 \times 10^{(10.5+1.75 \log_{10} R'_{HK})} \tag{4.17}$$

$R'_{HK}$ is defined in Noyes et al. (1984), based on the measurement of 41 main sequence stars, as:

$$R'_{HK} = 6 * 10^{-2} e^{-0.9 p_{rot}/\tau_c} \tag{4.18}$$

with $p_{rot}$ is the rotation period, in days and $\tau_c$ is the convective overturn time in days (Noyes et al., 1984):

$$\log \tau_c = \begin{cases} 1.362 - 0.166x + 0.025x^2 - 5.323x^3 & x > 0 \\ 1.362 - 0.14x & x < 0 \end{cases}$$

With $x = 1 - (B - V)$, and $(B - V)$ is the colour of the star. For consistency with Noyes et al. (1984) I set $(B - V) = \left(\log_{10}(M/M_{\odot}) - 0.28\right) / -0.42$.

$p_{rot}$ is based on the measured spin down rate of stars in the Hyades cluster (Aigrain et al., 2004):

$$\log p_{rot} - 0.5 \log\left(\frac{t}{625\text{Myr}}\right) = \begin{cases} -0.669 + 2.58(B-V) & 0.45 \leq B - V < 0.62 \\ 0.725 + 0.326(B-V) & 0.62 \leq B - V < 1.30 \end{cases}$$

With $t$ being the age of the star and 625Myr is the age of the Hyades.

For stars outside the valid range of $(B - V)$ I set $(B - V)$ to the nearest valid value. This means it is only defined for stars with masses between $\sim 0.5 M_{\odot}$ and $\sim 1.23 M_{\odot}$. The age of each of the stars in the synthetic sample occupies a bin in time, therefore I set the age to a random uniformly selected value inside the time bin. I ignore the factor of $\eta$ for the $P_{chrom}$, as the time-scale $\tau_{chrom} >>$ Nyquist frequency, thus $\eta \sim 1.0$.

From fig. 4.2 it can be seen the predominant noise source for the low and high mass main sequence stars is from the granulation. While stars with intermediate masses $M \sim 1.0 M_{\odot}$ the asteroseismic and the granulation terms and contribute equally. The variability due to the chromospheric activity is negligible for all systems due to the long timescale (8 days) of the activity compared Kepler's LC sampling rate (30 mins). The cut at high masses for the chromospheric activity is due to the edge where $(p_{rot})$ is defined, this is due to the rotation rates saturating (Aigrain et al., 2004). While the maximum mass for the asteroseismic signal is due to stars on the ZAMS being inside the $\delta$-Scuti instability strip.

McQuillan et al. (2014) compiled a catalogue of 34,000 MS objects with detectable rotation periods in Kepler, based on an automated analysis. From these objects they found that the amplitude of variability was between 950 ppm and 22,700 ppm, which is larger than that seen in fig. 4.2. Part of this difference may be due to the longer time scale of observations, 3 years, made by McQuillan et al. (2014), compared to the assumed 3 months used here, allowing longer period modulations

to be visible. They also show that the amplitude increases with decreasing temperature and mass, which is not seen in fig. 4.2. Further work is needed to integrate these results into the model.

For giant stars I ignore equation 4.6 and use the measured values from Hekker et al. (2012):

$$V = \left(3.5\nu_{max}^{-1.28} + 1.4\nu_{max}^{-1.5}\right) \times 10^7 \text{ppm}^2 \tag{4.19}$$

Where V is the variance ($\sigma_{var}^2$) in ppm$^2$ and $\nu_{max}$ is taken from the scaling relation shown in equation 4.9. The first term is due to granulation while the second is due to the asteroseismic signal.

In fig. 4.3 we can see the variability in evolved stars as a function of mass. The variability generally increases with stellar evolution (and hence age), due to $\nu_{max}$ decreasing as the radius increases.

The gap between the the thin and thick disc components is due to the star formation history. In principle there can be any evolved star with any initial mass in the thin disc visible today as it is still forming stars. However in the thick disc there exists both a minimum and maximum mass of a system that can still be in each evolutionary phase as it is no longer producing any new stars.

For binary systems I compute the total noise as the luminosity fraction weighted sum of the two components. Fig. 4.4 shows the combination of the stellar variability and the CCD noise as compared to the results presented in Jenkins et al. (2010b) (fig. 4.5) for the Kepler measured noise. We can see that the MS dwarfs are limited by instrumental noise, due to its dependence on magnitude, as well as following the upper and lower envelopes. This is because the the intrinsic variability is less than the instrumental noise. For the giant stars the noise is independent of the magnitude, as they have a much higher intrinsic variability, thus they are limited by their intrinsic variability.

Comparing with fig. 4.5 from Jenkins et al. (2010b) we can see that the giants match well, with a noise independent of the magnitude, with most objects scatted around the $10^3 - 10^4$ range. However the MS objects fare less well. A significant num-

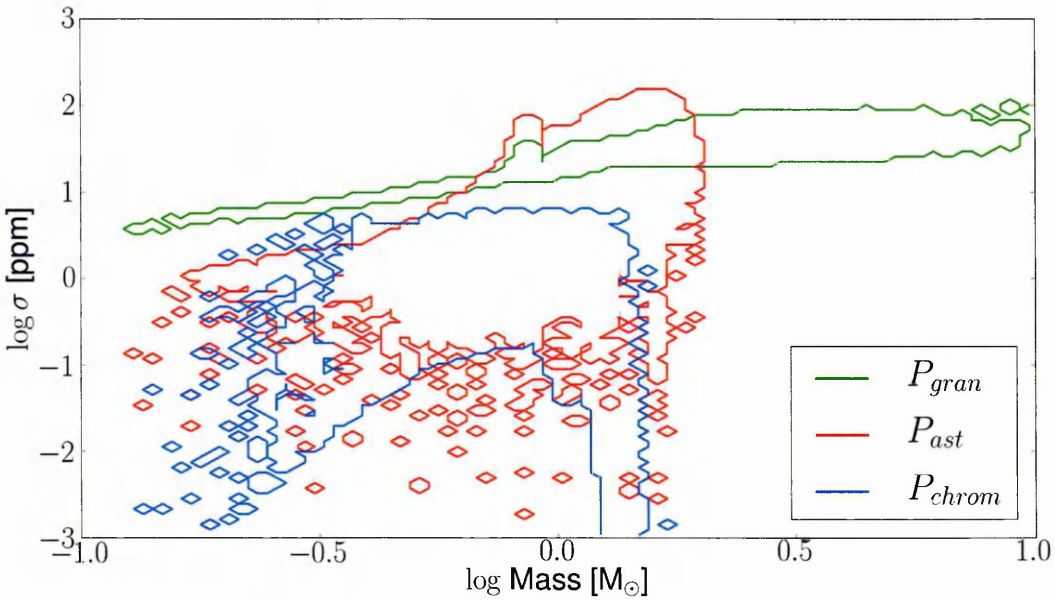



Figure 4.2: Expected noise power in Kepler light curves, as a function of mass for MS single stars. Green is the power due to granulation (equation 4.16) , red is the power due to solar like oscillations (equation 4.13) and blue is the power due to chromospheric activity (equation 4.6).

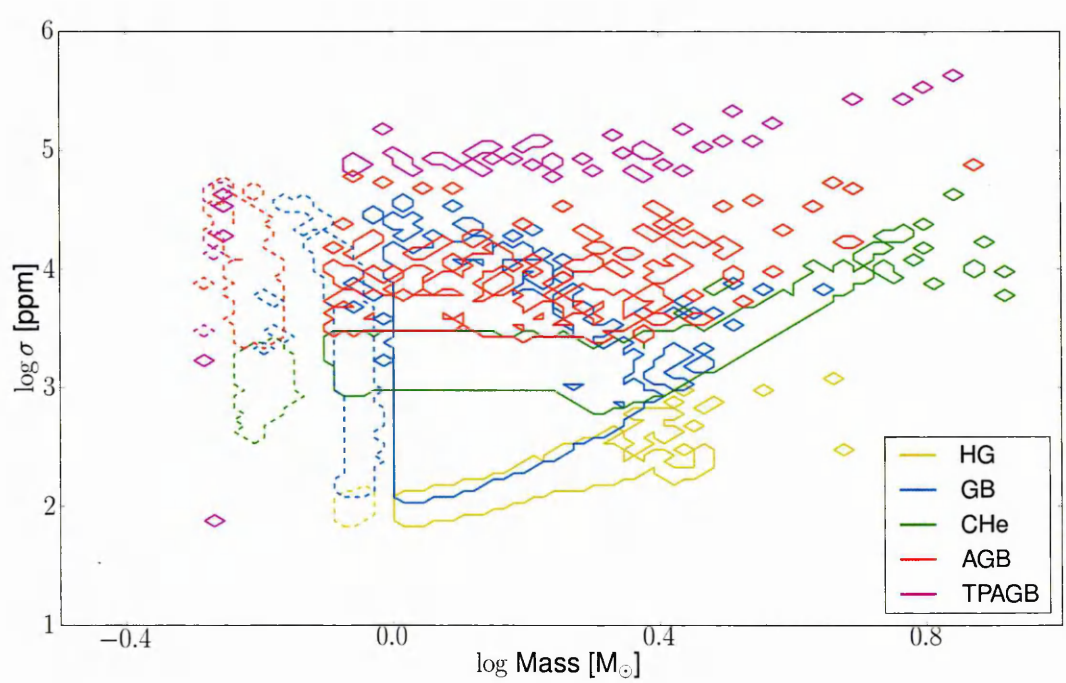



Figure 4.3: Power of the solar like oscillations and granulation for stars between the HG and TPAGB phases. Each colour represents a different evolutionary state, while the dotted lines represent thick disc stars and the solid lines represent the thin disc stars.

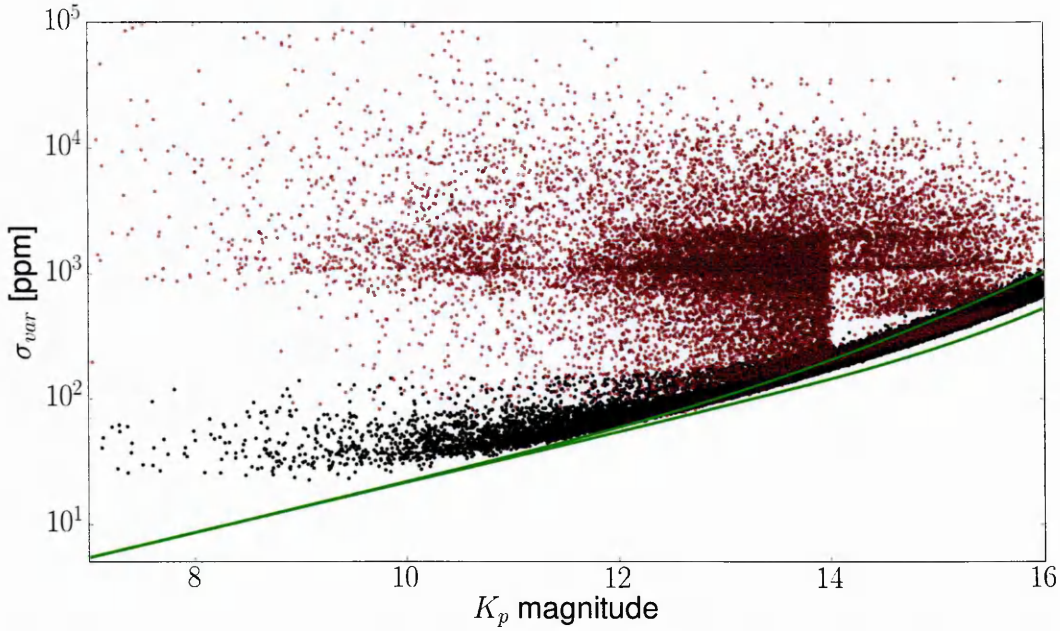



Figure 4.4: The total predicted noise for target selected single stars. The noise is the quadrature sum of the CCD noise (section 3.2.2) and the stellar variability noise (equations 4.8 & 4.19). Black points are MS stars while red are GB stars. Green lines are the upper and lower envelopes for the measured noise from Kepler (Jenkins et al., 2010b) .

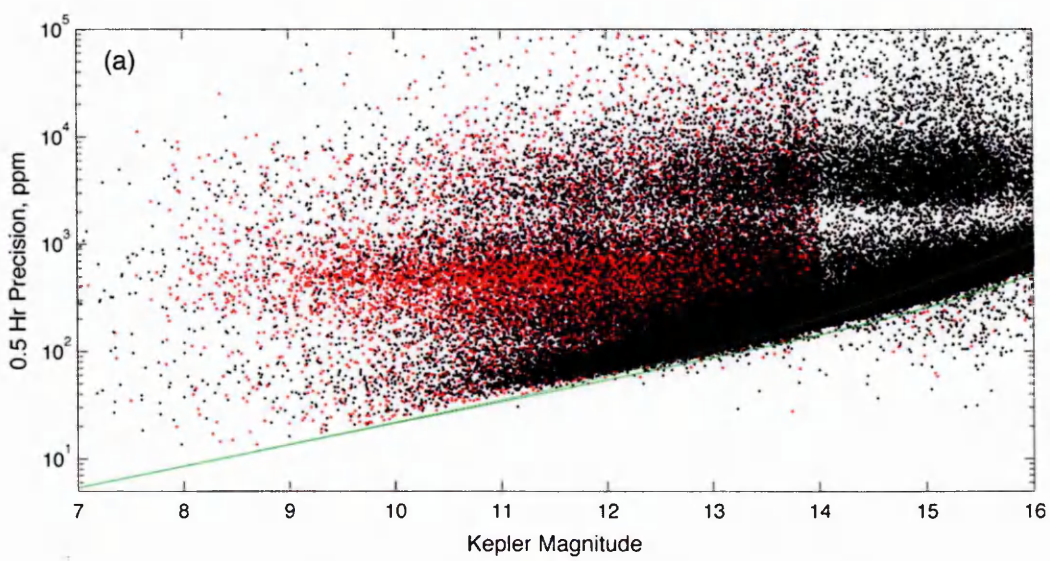



Figure 4.5: Figure 1 from Jenkins et al. (2010b). Original caption: Scatter plot of the 0.5 hr-to-0.5 hr precision. The dashed and solid green curves are the upper and lower envelopes of the measurement uncertainties, respectively, propagated through the data processing steps used to construct the flux time series. Dwarf stars ($\log g > 4$) appear as black points, while giants ($\log g < 4$) appear as red points. There is a strong separation between the dwarfs and the giants in their photometric behaviour at both timescales. Kepler is delivering near-intrinsic measurement-limited noise performance across the full dynamic range of its target stars.

ber of the dwarfs in Jenkins et al. (2010b) are limited by the instrumental noise and exist within the upper and lower envelopes. However, there are also populations of dwarfs with much larger noise, $\sim 10^3 - 10^4$ppm, that appear magnitude independent that are missing in fig. 4.4.

Possible reasons for this discrepancy are now discussed. The dwarfs in fig. 4.5 are based on the KIC $\log g$ values, thus some will be misclassified giants. However there are not enough of these systems to explain all of the discrepancy. Binary systems will have a noise based on the sum of the two noise values and a GB+MS system will be dominated by the noise from the GB star. However these represent at most $\sim$ 15% of the overall population (Table 3.5), so they can not completely explain the discrepancy either. The most probable source is a missing intrinsic variability, as the missing noise would need to be magnitude independent, while at the same time not affecting every MS object, as Jenkins et al. (2010b) shows most dwarfs lie inside the upper and lower noise envelopes.

Possible contenders would be the assumption that I can ignore the large scale variability for pulsating objects, where the variability is no longer a background contribution but comprises a significant portion of the light curve signal. There is also possible binary induced variability from accretion steams, accretion discs and tidal distortions. The other possibility is noise from other nearby stars that contaminate the optimal aperture of the star. Debosscher et al. (2011) found from a search of Q1 Kepler data that for $\sim$ 50% of their variable objects they detected the variability could be ascribed to contamination from other stars. Finally the variability due to the rotating systems detected in McQuillan et al. (2014) would provide a significant number of systems with noise in the $10^3 - 10^4$ ppm range.

### 4.3.2 Deriving an EB sample

To derive the synthetic EB catalogue I take each target selected binary from the sub-sampled $\Gamma$ distribution of all stars, and give the system a random inclination between 0 and $\pi/2$. Using JKTEBOP (section 2.2) I generate a synthetic light curve of

the system using a quadratic limb-darkening law, gravity darkening and the noise estimate from section 4.3.1 and compute the eclipse depths, $\Delta F/F$.

For this work I consider only detached and semi-detached EBs, ignoring ellipsoidal and contact EBs. I define semi-detached EBs as those undergoing RLOF with a detectable eclipse, while ellipsoidal system those with a detectable ellipsoidal modulation but no detectable eclipses. Contact EBs are those with $R_1 + R_2 = a$, but there are none in the synthetic sample, all other systems as classified as detached.

I ignore ellipsoidal systems as they share similar light curves to contact systems, mainly the large changes in flux at 1/4 and 3/4 phases due to tidal distortions. Thus the classification between ellipsoidal and contact, for which I have no evolutionary model for, is less certain. The same can be said for the distinction between detached and semi detached systems, especially with regards to the boundary between the classifications. However as BISEPS has an evolutionary model for both systems I can compare the synthetic EBs to the Kepler EBs in a combined detached+semi-detached group and ignore the classification.

To determine the "detectability", $D$, of an eclipse I assume the detection process is a Gaussian-like process. I perform a rejection sampling, using the S/N of the eclipse, against a Gaussian distribution (Fressin et al., 2013). I compute the probability of detection for a given S/N using equation 4.20, this is then compared to a random uniformly selected number in the range [0,1]. If $D(S/N)$ is greater than the randomly drawn number then I select the eclipse as being "detected". The cumulative distribution function of a Gaussian can be expressed as (Fressin et al., 2013):

$$D(S/N) = 0.5 + 0.5\,\mathrm{erf}\left(\frac{(S/N - 7.1)}{\sqrt{2}}\right) \tag{4.20}$$

where 7.1 is the favoured Kepler S/N, erf is the error function and I derive the S/N later. With equation 4.20, an eclipse with a $S/N = 7.1$ would have a 50% chance of being detected. I perform the sampling process equal to the number of times that the eclipse is predicted to be seen during Kepler's Q1 & Q2 period, which is 120 days (Tenenbaum et al., 2013).

I then apply a set of criteria to determine whether Kepler would determine the system to be an EB, based purely on the synthetic light curve. Firstly, I select systems with at least 3 primary and secondary eclipses detected. Then I select those that have 3 eclipses (primary or secondary) where the eclipse depth is larger than $2R_{jupiter}$ (Slawson et al., 2011) using:

$$\Delta F/F = \left(R_p/R_*\right)^2 \tag{4.21}$$

Where $R_p = 2R_{jupiter}$ and $R_*$ is the radius of the of the synthetic star (assuming it was a single star not a binary star) using the stellar classification program (SCP) method in section 3.2.1. The requirement for 3 detections implies a maximum period for the synthetic EBs of 40 days, I also apply a minimum cut of 1 day to prevent issues with possible ellipsoidal/contact systems.

I derive the apparent $(S/N)$ of an eclipse from star 1 eclipsing star 2 using the following:

$$S/N = \frac{(\Delta F/F)}{\sqrt{\sigma_{ccd}^2 + \sigma_{var}^2}} \tag{4.22}$$

where $\Delta F/F$ is the eclipse depth from JKTEBOP, computed as the difference in the flux between 1$^{st}$ contact and mid eclipse and in the absence of noise. The eclipse depth is modified to include flux from background sources which acts to make $\Delta F/F$ smaller. $\sigma_{ccd}$ is the combined noise term from the CCD ( see equation 3.4), which takes into account the noise from the signal, background and instrumental effects. I assume for simplicity that all terms are constant during the eclipse. Finally, $\sigma_{var}$ is the noise during the eclipse from the stellar variability. The S/N is then scaled by the number of 30 minute samples (assuming LC observations) Kepler would acquire during the eclipse. The duration of an eclipse in seconds can be expressed as:

$$T_D = \frac{P}{\pi}\sin^{-1}\left(\frac{\sqrt{(R_1+R_2)^2 - (a\cos i)^2}}{a}\right) \tag{4.23}$$

where $P$ is the orbital period in seconds, $a$ is the orbital separation in meters and

$R_{1,2}$ are the radii of the stars in meters. Kepler therefore makes $T_D$/ (30mins) samples during the eclipse, thus the overall S/N of an eclipse is:

$$(S/N)_{total} = (S/N) \times \sqrt{T_D/\,(30\text{mins})} \tag{4.24}$$

The $(S/N)$ is computed for both the primary and secondary eclipses and then used in equation 4.20 to derive the detectability of each for each eclipse.

### 4.3.3 Kepler EB pipeline

With a synthetic sample of detected EBs in hand the next step is to pass these light curves through the Kepler pipeline to determine their Kepler derived parameters. The Kepler EB pipeline is made up of two components. First, the light curve is passed through POLYFIT (Prša et al., 2008) which acts to re-sample the light curve into a uniformly distributed set of phase points. This is done by modeling the light curve as a set of polynomials with a series of "knots" where the polynomials meet. This re-sampled light curve is then passed though EBAI (Prša et al., 2008), which derives a set of binary parameters based on knowledge it gains from training itself on light curves with known parameters. Fortunately, both sets of codes are publicly available[1].

There has been a third code recently introduced (Matijevič et al., 2012) which derives from the shape of the light curve a parametrised value, $c$, between [0,1]. This parametrised value corresponds to the systems morphology (detached [0,0.5], semi [0.5,0.7], contact [0.7,0.8] or ellipsoidal/uncertain [0.8,1.0]). The mapping between $c$ and the classification was estimated by comparing the $c$ values with a manual classification of the light curves. Thus these classifications are only an approximation, providing a single reproducible way of classifying systems, which is broadly similar to the morphology. Integrating this code in to my pipeline is left for future work.

EBAI (Prša et al., 2008) is an artificial neural network code designed to learn the shape of a light curve and from that derive the parameters of the system. The code

[1] http://phoebe-project.org/

is first trained on a set of representative light curves, where the parameters of the system are known. On each iteration the code attempts to map the inputs (flux at each phase point) via a set of "hidden" parameters to a set of outputs (Temperature ratios, mass ratios etc) and on each iteration the estimation is refined. Once the code has been trained on known systems, it can then be run on unknown data, processing hundreds of thousands of light curves per second (Prša et al., 2008).

Based on the information available from a light curve different parameters are derived for different systems.For detached and semi detached systems EBAI derives the temperature ratio $T_2/T_1$, fractional radii $(R_1+R_2)/a$, radial and tangential eccentricity $e\sin\omega, e\cos\omega$ and the sine of the inclination $\sin i$. For contact an ellipsoidal and contact systems it derives $T_2/T_1$, photometric mass ratio $q_{ph}$, $\sin i$ and the fill out factor $F_f$. The difference parameters that are determined are due to differences in the information content of the light curves. For contact and ellipsoidal systems the mass ratio, $q_{ph}$, can be determined from the ellipsoidal modulations. They are also all circularized due to their small separations, thus the eccentricity terms are assumed zero. The temperature ratio for all systems is based on the eclipse depth ratio which is a function of the surface brightness ratio. For the detached and semi detached systems the fill out factor is replaced by the fractional radii, due to the wider separations between the bodies. The detached systems may not be circularized, hence the eccentricity contributions.

### 4.3.4 EBAI Training sets

To use EBAI I first generated a training set of light curves following the procedures in Prša et al. (2008, 2011). I created two training sets, each of which has a different set of distributions. Firstly, I randomly sampled the binary output I generated in section 2.1 to select a set of *possible* binary systems and secondly I sampled from a flat IMRD to give a set of *probable* systems. For each sample I select 35,000 systems and constrain the periods to be within 1 day and 1 year.

For each system BISEPS supplies the the $T_2/T_1$ and $(R_1+R_2)/a$. I select a incli-

nation from a random uniform distribution between $[\cos^{-1}(R_1 + R_2)/a, \pi/2]$, which guarantees the system to eclipse. I then give each system a uniform randomly sampled argument of periastron between $[0, 2\pi]$ and an eccentricity that is selected from a Gaussian with $\mu = 0$ and $\sigma = 0.05$. For training set 1 there is no constraint on the eccentricity, while training set 2, which more closely mimics a realistic distribution, there is a constraint that the eccentricity must be 0 for $P_{orb} < 10$ days (Duquennoy & Mayor, 1991). Note these eccentricities are not from any binary evolution, they are only inserted into the calculation of the light curves and are there to maintain consistency with Prša et al. (2008) use of 5 parameters for EBAI to derive. I then apply the following cuts, based on table 1 from Prša et al. (2008):

- $0.1 < T_2/T_1 < 1.0$
- $0.01 < (R_1 + R_2)/a < 0.5$
- $-0.3 < e\sin\omega \;\&\; e\cos\omega < 0.3$
- $0.85 < \sin i < 1.0$

For each system I generate a light curve using JKTEBOP, with quadratic limb darkening, gravity darkening and reflection effects applied. I generate 201 phase points and at each point I apply a noise term. This is sampled from a Gaussian with $\mu = 0$ and a uniformly randomly selected $\sigma$ between 0.0 and 5.0%. This is only a training set of data so the noise levels only have to be approximately what would be expected from the Kepler data set. For systems where the average noise is greater than $7.1 \times (\Delta F/F)$ I ignore the system and select another system as the eclipse is deemed undetectable.

Each light curve, after the addition of the noise, is passed through POLYFIT which re-samples the light curve and smooths out the noise. This is modelled using a set of 4 separate quadratic polynomials, allowing POLYFIT to determine the "knot" points between the quadratic polynomials automatically (Prša et al., 2008). The choice of 4 polynomials was chosen by following the procedure in Prša et al. (2011), which selected this value as a compromise between the runtime and the precision of the

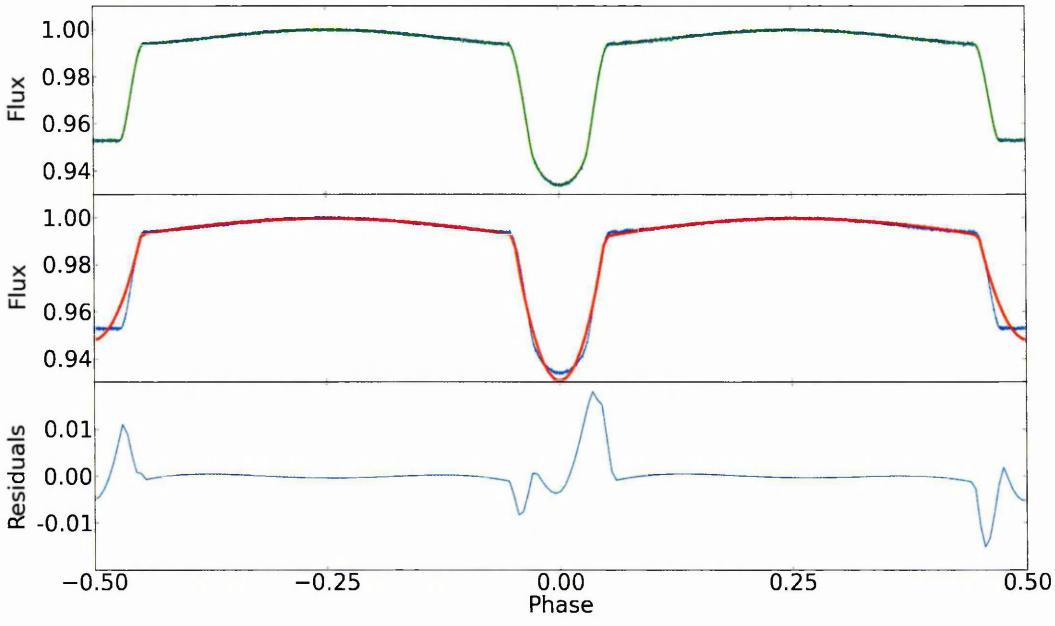



Figure 4.6: Comparison of POLYFIT fits to a input light curve, where POLYFIT performs poorly. Top panel: In blue the noisy light curve for a synthetic star, Green the non-noisy light curve. Middle panel: Blue same as top panel, Red POLYFIT fit to the light curve. Bottom panel: The residuals between the POLYFIT fit and the non-noisy fit from the top panel.

fit. POLYFIT was then set to output the light curves with 201 equidistant points in phase space.

Figs. 4.6 & 4.7 show cases where POLYFIT has succeeded in fitting the light curve and where it has been less successful. In Fig. 4.6 we can see how POLYFIT misses the fact that the secondary eclipse is flat and how it fits the depth of both the primary and secondary eclipses poorly compared with the out of transit fit. This is significant as EBAI uses the relative depths of the eclipses to determine the $T_2/T_1$ parameter. Though fig. 4.7 shows a better fit by POLYFIT to the data, which implies that the bias inserted by POLYFIT depends on the type of system being considered. The most significant difference between the two fits is that fig. 4.6 the secondary is totally eclipsed by the primary. Thus this would suggest this effect is largest for unequal sized binaries.

This processed light curve is then passed to EBAI along with the system parameters and set to training mode. I run 35,000 light curves, through 10,000,000 iterations of the training mode, across 24 nodes using MPI, with 40 hidden parameters. This

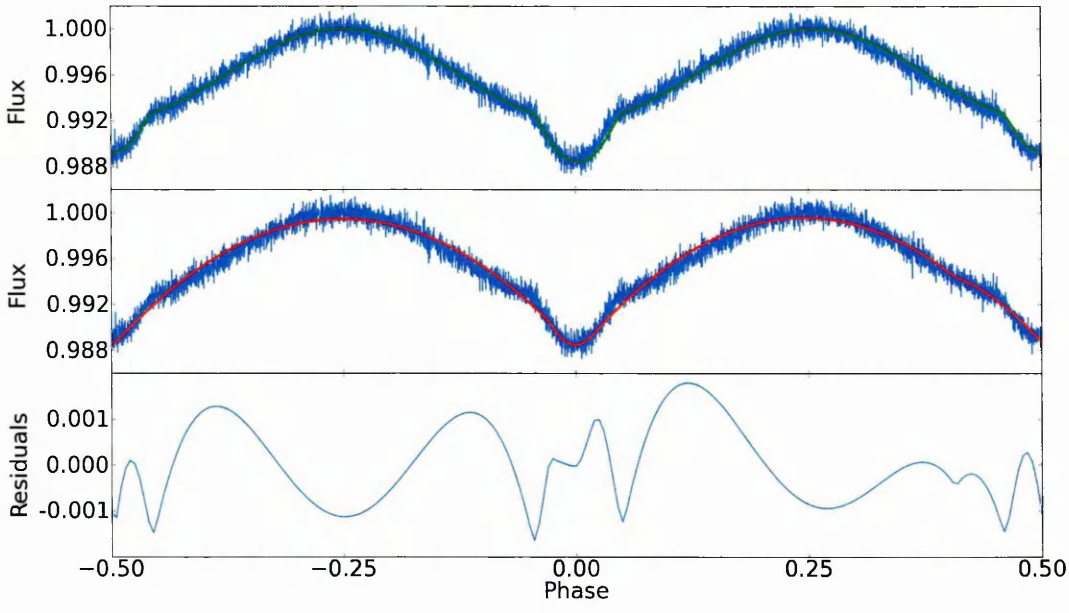



Figure 4.7: Similar to Fig. 4.6, except where POLYFIT has derived a more robust fit.

is similar to the setup that the original Kepler training set used and therefore should introduce a similar level of bias into my data. As noted in Prša et al. (2008) splitting up the problem over MPI speeds up the computation significantly, however combining results from multiple individual processors (each processor gets a chunk of the light curves), leads to slightly different results than if computed on a single processor due to the recombination of the results from the different nodes. The output of the training run is then fed back in to EBAI in recognition mode to compare its predicted system parameters to the true parameters.

The results of training set 1 & 2 can be seen in Fig. 4.8 & 4.9. Fig. 4.8 shows EBAI performing well on the fits for all but the temperature ratio $T_2/T_1$. The $T_2/T_1$ and $r_1+r_2/a$ plots show how the selection of *possible* systems has weighted the distribution towards systems with a large amount of temporal resolution, for instance during a RLOF, which generate many output points in the data set. The selection process then favours these semi detached systems.

Fig. 4.9 shows a slightly poorer fit to the eccentricity based distributions, primarily as less systems exhibit effects due to eccentricity as I set e=0 for $P_{orb} < 10$ days. The $r_1 + r_2/a$ fit is good, while the inclination fit is reasonable, however the $T_2/T_1$ fit

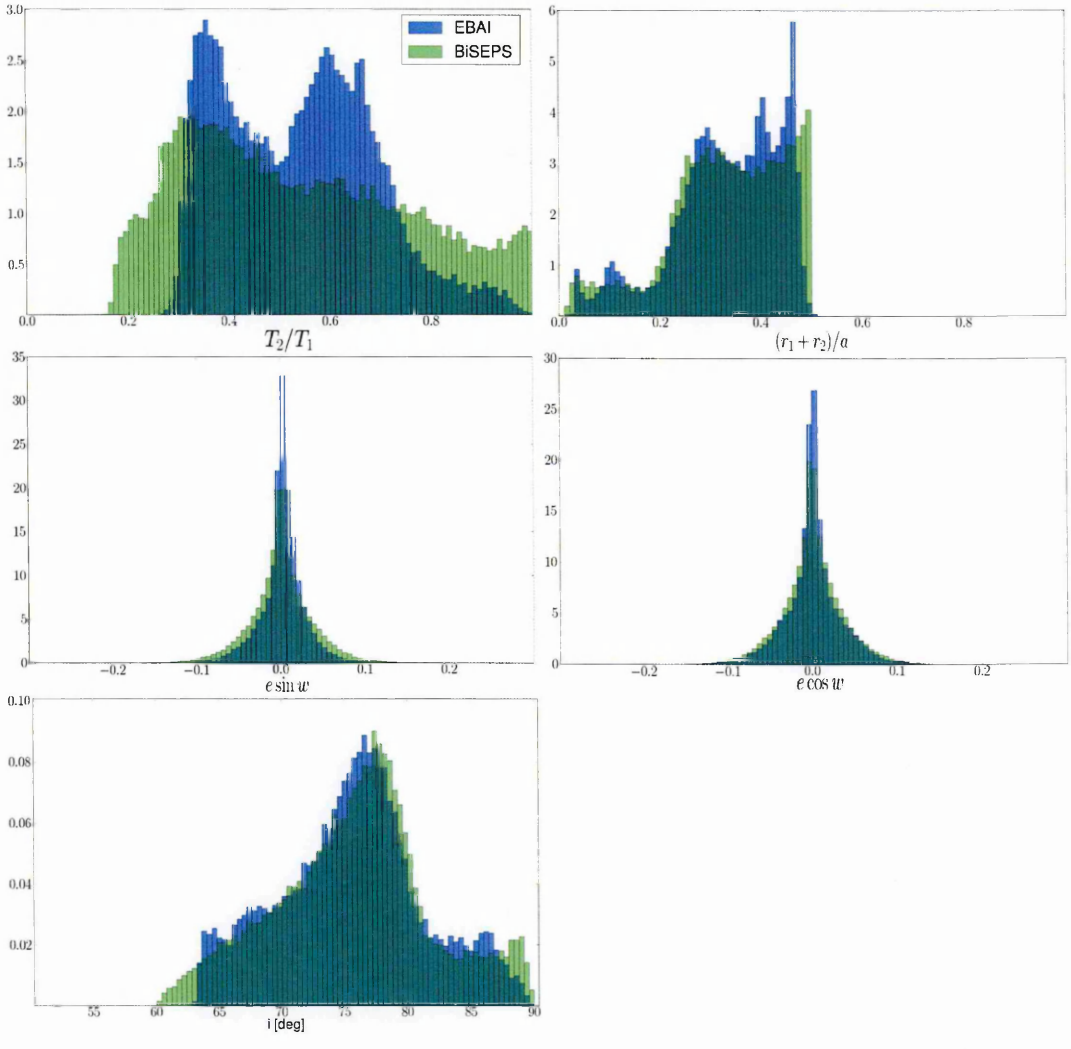



Figure 4.8: Comparison between EBAI derived parameters and the true parameters for EBs in training set 1. The EBs were selected from a set of *possible* binary systems. There are 35,000 light curves in each sample. In blue is the EBAI result, green is the true parameters.

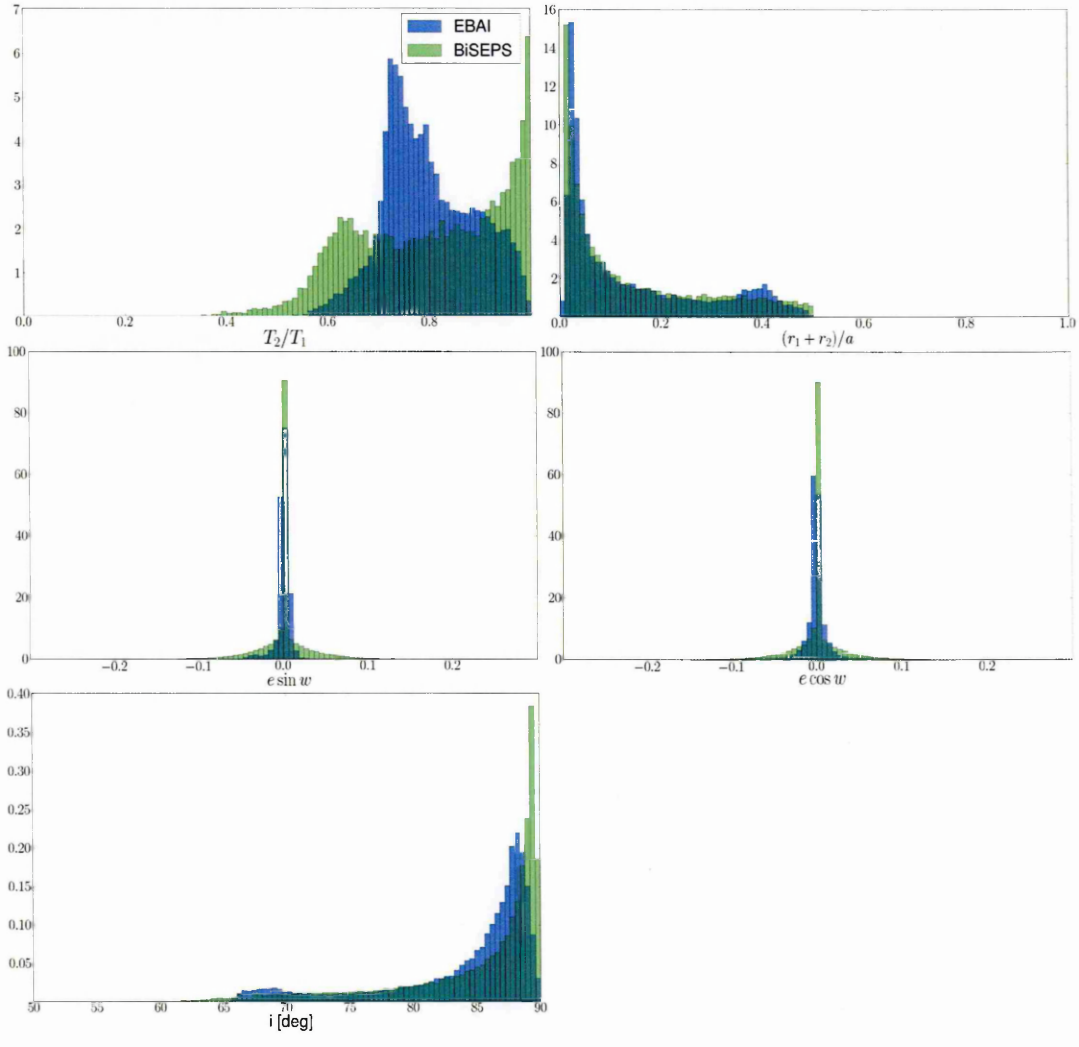



Figure 4.9: Comparison between EBAI derived parameters and the true parameters for EBs in training set 2. The EBs were selected from a set of *probable* binary systems (based on a flat IMRD distribution). There are 35,000 light curves in each sample. In blue is the EBAI result, green is the true parameters.

is again poor and exhibits the completely wrong trend with regards to the number counts as a function of $T_2/T_1$. It also shows that the selection of *probable* systems selects those with long lifetimes (wide non interacting systems), thus favouring detached systems.

The poorness of the temperature ratio fit I believe is due the POLYFIT fit to the light curve. POLYFIT fits the eclipse minima poorly in fig. 4.6, which is where the information on the temperature ratio is contained via the eclipse depth ratio.

### 4.3.5 Comparison of EBAI training sets

To determine which is the best training set to use on the actual sample of synthetic light curves I tested each training set with the light curves from the other set. This way I test how effective EBAI is in recognising parameter distributions given light curves that do not follow the same distribution as it was trained on.

Fig. 4.10 shows the results of the training set 1 recognising the light curves from training set 2. The temperature ratio and fractional radii distributions show broad agreement, though both introduce artefacts. The temperature ratio distribution gains a dual peaked distribution not present in the real distribution and the fractional radii distribution shows a peak at $\sim 0.1$, both of which are present in the Kepler EB catalogue (see Fig. 4.18).

Fig. 4.11 shows the results of the training set 2 recognising the light curves from training set 1. Again a spurious double peaked temperature ratio distribution appears in the EBAI results, while the fractional radii fits the input distribution below 0.2, while above 0.2 EBAI derives a narrower and more peaked distribution.

On the basis that both training sets have issues recognising completely different distributions to those they where trained on, I decided to use training set 2 to analyse my synthetic EB catalogue. This is on the basis that it does a reasonable job at recognising the light curves it was trained on and I expect the synthetic EB catalogue to resemble those more than the light curves from training set 1.

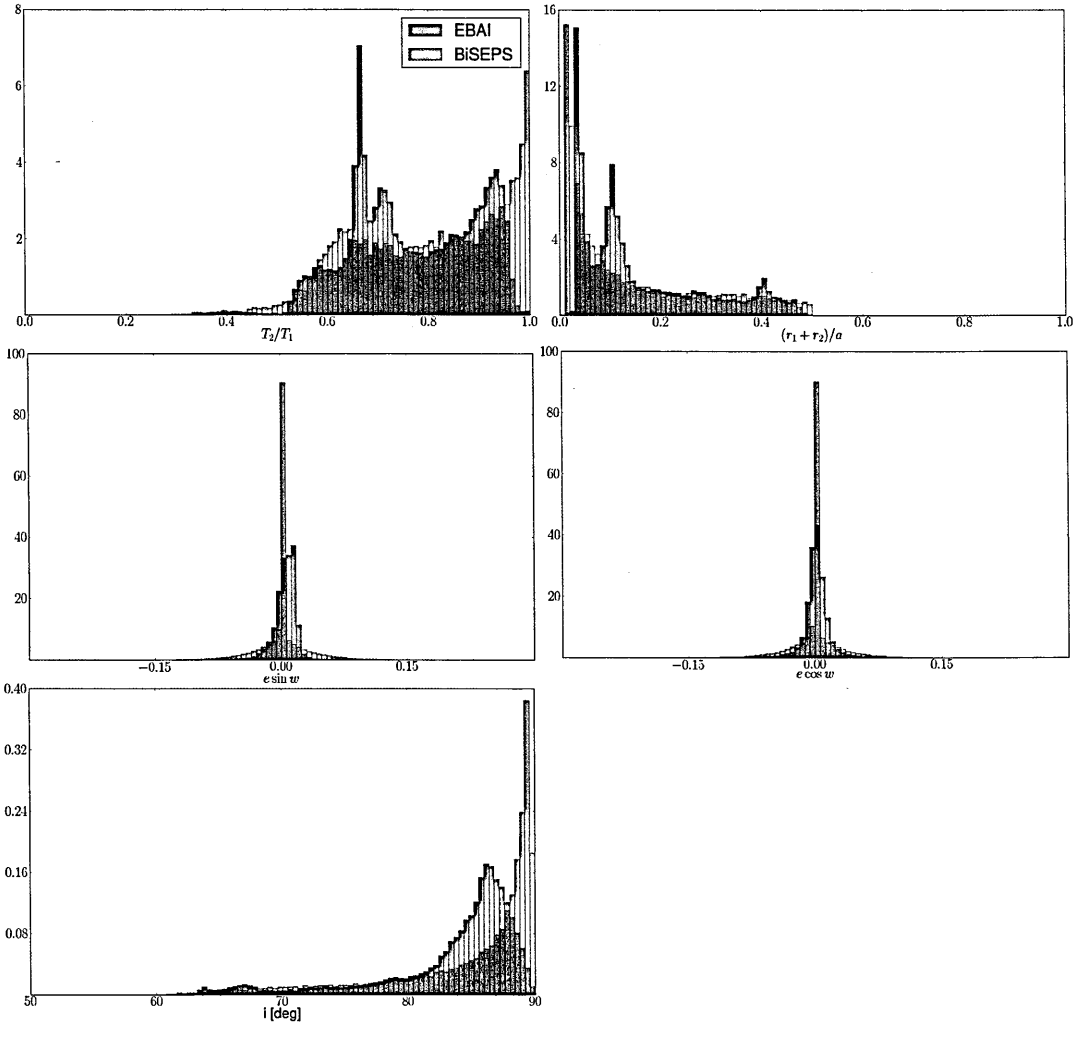



Figure 4.10: Comparison of EBAIs performance on using results from training set 1 on light curves made for training set 2. In blue is the EBAI result, green are the true parameter distributions.

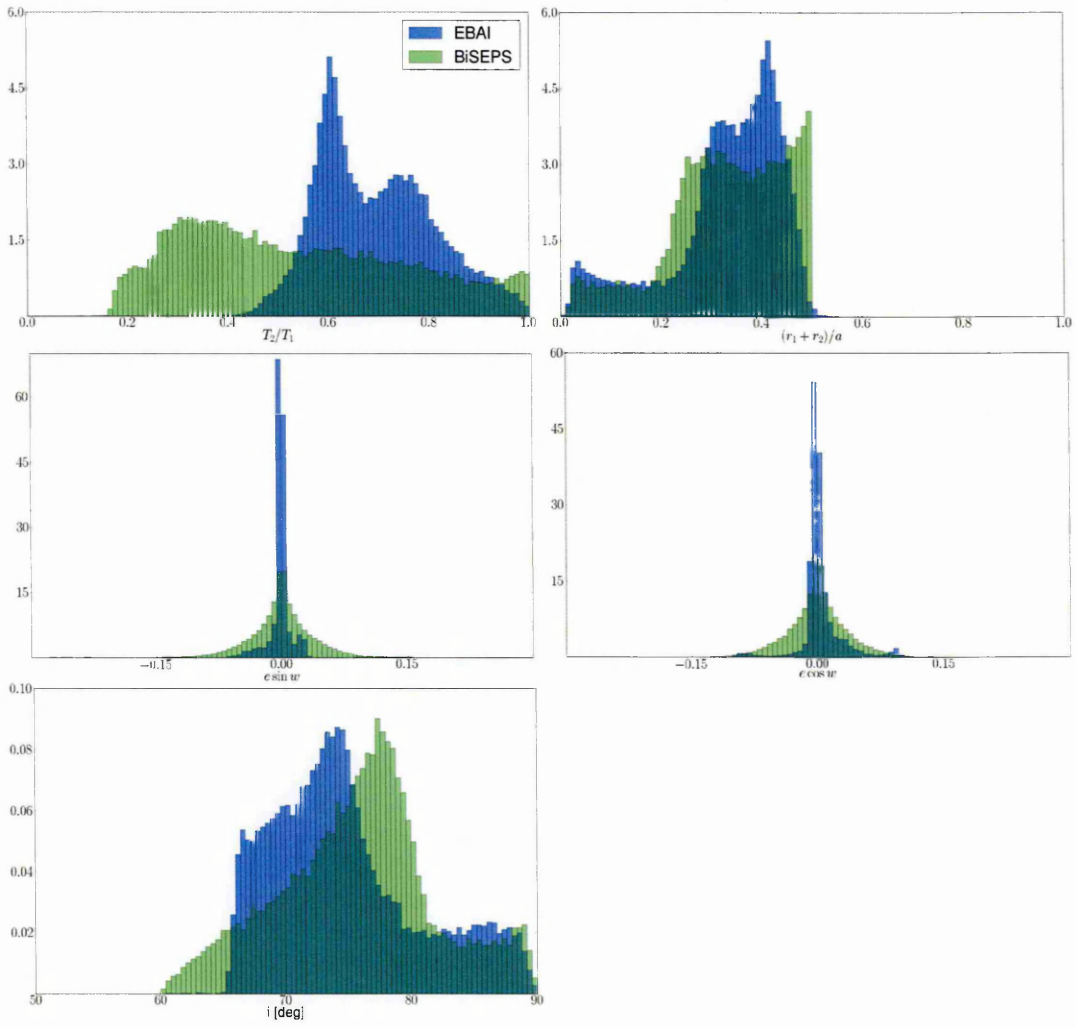



Figure 4.11: Comparison of EBAIs performance on using the results from training set 2 on light curves made for training set 1.In blue is the EBAI result, green are the true parameter distributions.

| s | Scaling factors | |
|---|---|---|
| | Target | Non-target |
| -0.50 | 0.63 | 0.80 |
| 0.00 | 0.77 | 1.00 |
| $\log M_1$ | 0.74 | 0.95 |
| 0.50 | 0.83 | 1.07 |
| 1.00 | 0.82 | 1.06 |

Table 4.2: Scaling factors required to normalise the different IMRD and IBFD distributions. See section 4.4.5 for the IBFD. The factors are defined as $N_{Kepler}/N_{biseps}$, where $N_{Kepler}$ is the number of Kepler systems (target or non-target) and $N_{biseps}$ is the number of BISEPS systems (target or non-target)

## 4.4 Results

I now investigate the effects of the choice of IMRD parameter on the synthetic EB sample to see whether it would be possible to determine the IMRD from the Kepler EB data. For this work I take 4 values of the IMR $s$ parameter; -0.5,0.0,0.5,1.0. I also consider the case (for $s = 0$) of both a uniform binary fraction of 50% and one where the binary fraction is a function of the primary mass.

### 4.4.1 Scaling factors

Due to each model generating a different number of total stars and a different number of target selected stars, I introduced a set of scaling factors to normalise the results. I normalise the total counts against the number of objects in the Kepler Q2 catalogue. For non-target selected systems there are 415,943 stars and for target selected systems there are 165,434 stars in the Q2 catalogue. The factors can be seen in table 4.2. The counts for each distribution are multiplied by these values to derive the Kepler equivalent counts.

### 4.4.2 Number counts

For each background population model I generated 10 synthetic EB catalogues as described in section 4.3, by resampling from the inclination distribution, to determine average value for the distributions. I chose to repeat the model 10 times as a

| s | Number of detectable EBs | | | |
|---|---|---|---|---|
| | Total | MS+MS | GB+MS | GB+GB |
| -0.50 | $835 \pm 27$ | $740 \pm 24$ | $93 \pm 7$ | $2 \pm 1$ |
| 0.00 | $1035 \pm 15$ | $913 \pm 16$ | $118 \pm 6$ | $4 \pm 1$ |
| 0.50 | $1174 \pm 24$ | $1029 \pm 20$ | $141 \pm 9$ | $4 \pm 2$ |
| 1.00 | $1280 \pm 35$ | $1136 \pm 36$ | $137 \pm 10$ | $8 \pm 34$ |
| Kepler | 1244 | | | |

Table 4.3: The number of EBs detected for different evolutionary types as a function of the IMRD parameter $s$. Values are quoted as Kepler equivalent, taking into account the difference in size between the true and synthetic target lists. The numbers quoted are of detached and semi detached systems with periods between 1-40 days. GB contains both HG and GB stars. Uncertainties are 1 standard deviation.

| s | Percentage of EBs (%) | |
|---|---|---|
| | Detached | Semi detached |
| -0.50 | $90.51 \pm 3.04$ | $9.49 \pm 0.68$ |
| 0.00 | $89.62 \pm 1.46$ | $10.38 \pm 0.74$ |
| 0.50 | $88.79 \pm 2.02$ | $11.21 \pm 0.58$ |
| 1.00 | $89.50 \pm 2.89$ | $10.50 \pm 0.55$ |
| Kepler | 94.35 | 5.64 |

Table 4.4: The number of EBs detected as detached or semi detached systems. Values are quoted as percentages of the total number of EB in the synthetic sample. Periods are constrained to be within 1-40 days. The Kepler results have been rescaled to include only the detached and semi detached systems. Uncertainties are 1 standard deviation.

compromise between runtime and precision of the final results, given that for the moment I am concerned with determining the validity of the method, as opposed to the true EB distribution functions. Uncertainties quoted are 1 standard deviations, with 1 degree of freedom, on the number of objects in each bin in parameter space.

From table 4.3 we can see the differences in the number of detached and semi detached EBs in the synthetic sample. There is a trend for an increasing number of EBs as $s$ increases. This is due to the fact that for a given $M_1$ an equal mass component, which will typically be the same size, will be both more luminous and more likely to eclipse than an unequal mass component. On average the fraction of MS+MS among EBs is $\sim 88\%$, which is constant over the different $s$ values.

Table 4.4 shows the split between detached and semi detached systems. The number of detached systems slightly decreases with increasing $s$. We can see that

none of the models reproduce the Kepler results. though this may be due to a number of reasons either the uncertainty in Kepler's classification or the period distribution of the synthetic EBs, generating too many short period systems (see section 4.4.3)

### 4.4.3 Period distribution

Fig. 4.12 shows the period distribution for different stages in the pipeline, for the $s = 0$ case. We can see that the initial period distribution (IPD) and present day period distribution (PDPD) is log flat for $P_{orb} > 1$ day. While for $P_{orb} < 1$ day, the distribution drops to 0 as the period decreases. The reason this distribution is not log flat over all periods, as implied by the log flat IOSD, is that these are systems visible today. Thus only the systems that have not merged are in the synthetic samples. We can see that the target selection has no effect on the distributions, which is as expected as the target selection has no period dependence or even the concept of a binary system. The EB period distribution is shown for comparison. We can see that the EB selection effect has increased the fraction of short period compared to longer period systems. This is unsurprising as shorter period systems will have more eclipses for which to sample against equation 4.20, thus low S/N systems will be more detectable at short periods.

Fig. 4.13 shows the period distribution of the synthetic EB catalogues in comparison to the actual Kepler EB catalogue. Note the period is not derived from EBAI, instead it is from BISEPS and is thus the true orbital period of the system. This means the period distribution is testing only the model for detection of EBs, based on equation 4.20. We can see the effect of the binary evolution, and hence age of the system, on the short period drop in systems. This is due to the braking forces (section 1.1.1.1) and stellar evolution into a giant star forcing short period systems to merge and thus no longer be visible in the sample.

From fig 4.13 it can be seen that the none of the models fit the Kepler data. We can see that the different IMR laws all show the same period distribution, within un-

certainties. It therefore appears unlikely that even with a "fixed" period distribution that there would be sufficient difference between the different IMRDs to distinguish between them. From the model the period distribution is $\sim P_{orb}^{-1}$, while the Kepler data suggests an approximately log flat distribution.

Splitting the period distribution up into MS+MS binaries (fig. 4.14) and GB containing systems (fig. 4.15) we can see where the differences between the model and data are. Fig. 4.14 follows fig. 4.13 closely which is not surprising as MS+MS systems make up ~ 90% of all the EBs. The GB containing systems period distribution has a larger deviation per bin, compared with the MS+MS systems, due to there being less systems. As $s$ increases the period distribution at $\log P < 0.4$ flattens out, as it becomes more difficult to have two equal sized GBs in a stable short period orbit.

A possible solution to this would be to increase the noise for short period systems. With a lower S/N then less systems will be detected. This may be related to the increase in semi-detached systems seen in table 4.4 compared to Kepler. If I included a binary noise term (accretion, tidal distortion etc.) this would be strongly dependant on the orbital period. However, if I increase the noise then the total number of EBs will decrease, which will cause a poorer fit with the total EBs in table 4.3. Thus I may also need to find a way to decrease the noise for longer period systems to compensate. Gaulme et al. (2014) suggests that the asteroseismic signal in GB systems with $P_{orb} < 40$ days is suppressed due to induced magnetic activity in the GB star.

Age may also be a factor in this. If the stars where older, via a higher SFR in the past, then the braking forces applied to short period systems will have longer to act. This will drive systems to shorter orbital periods, as well as increase the rate of mergers and common envelope systems. This may be able to drive down the short period end, by moving systems to below the $P > 1$ day cut, though we still need to maintain consistency with the total number of EBs through some other mechanism.

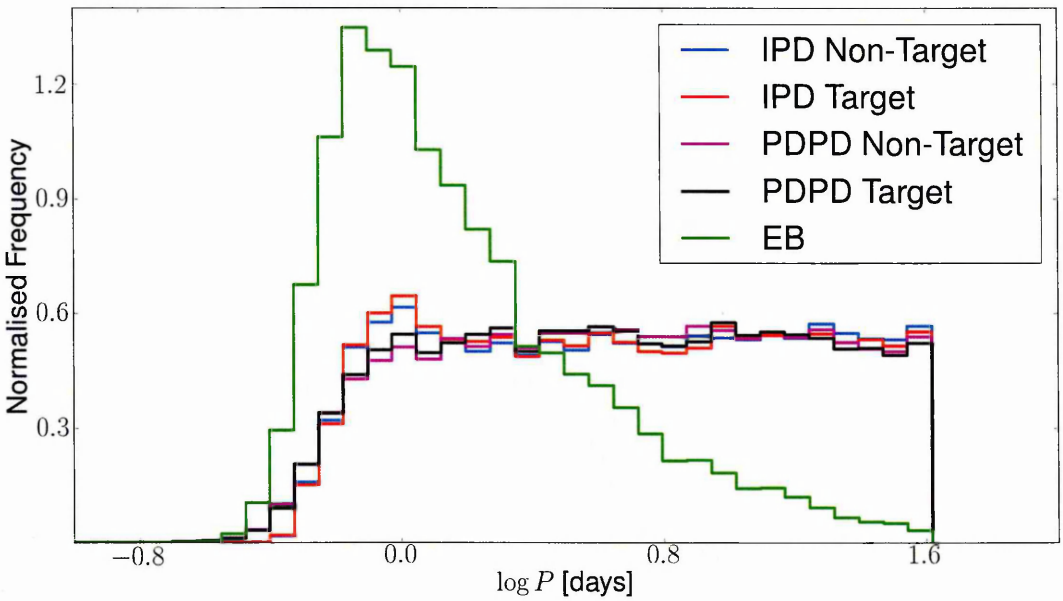



Figure 4.12: The period distribution of the synthetic binaries at different points in the pipeline, for the $s = 0$ model. The IPD is the initial period distribution for all binaries in the synthetic input catalogue, while PDPD is the present day period distribution of all binaries in the synthetic catalogue. The EB period distribution is also shown for the detached and semi detached systems.

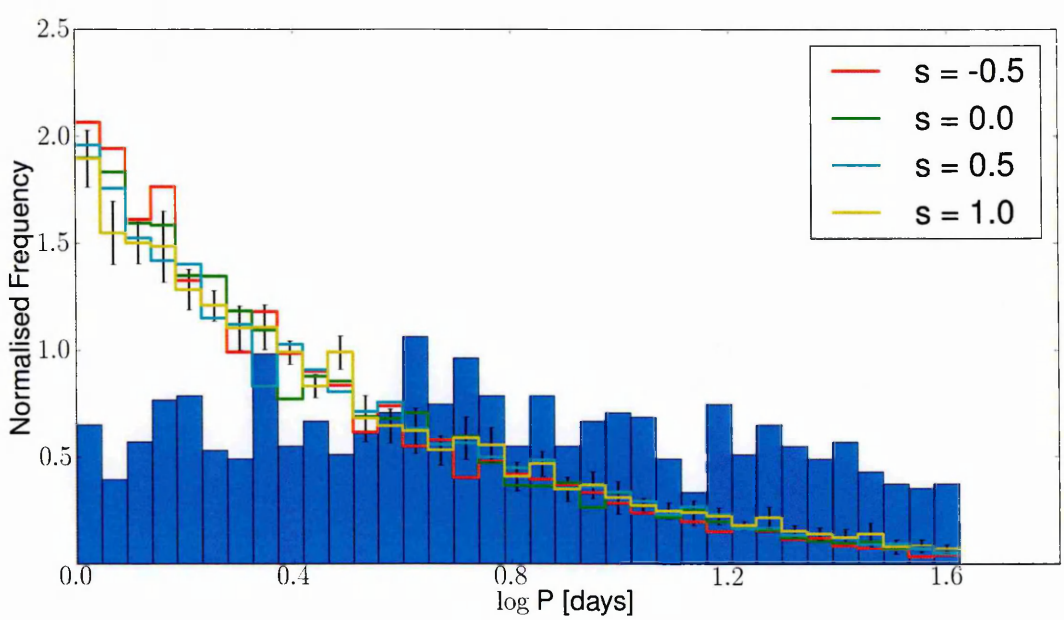



Figure 4.13: Filled histogram: the period distribution of the Kepler EB catalogue, for detached and semi detached systems with period between 1-40 days. Coloured lines: the period distribution for different IMRD values of $s$, for detached and semi detached systems with periods between 1-40 days, averaged over 10 runs. Representative 1 standard deviation uncertainty bars are shown.

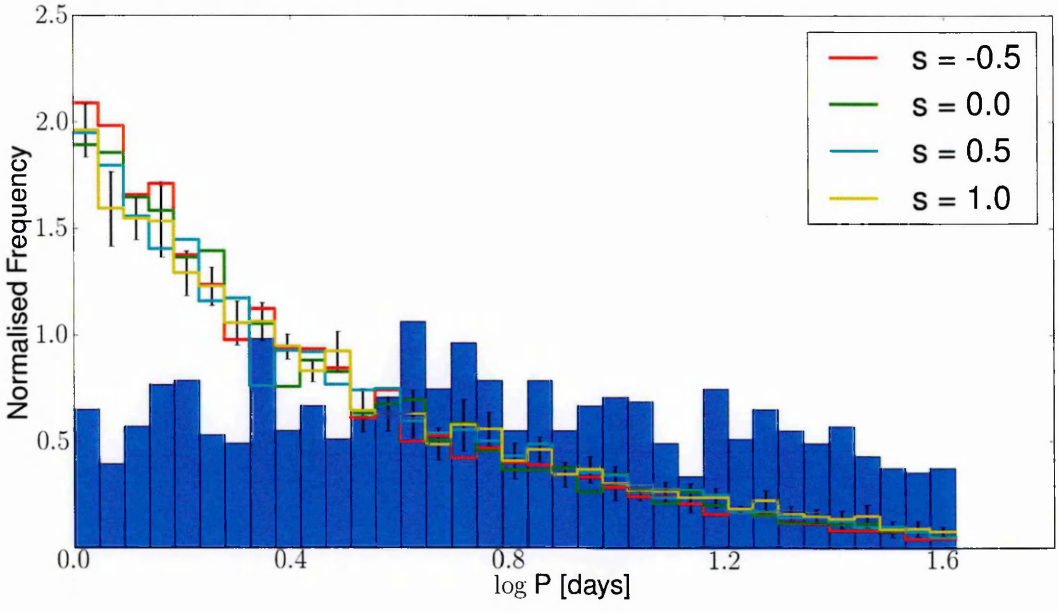



Figure 4.14: Similar to Fig.:4.13, except the synthetic stars are filtered to be MS+MS systems. Note there is no information in the Kepler EB catalogue about stellar classification thus there are no extra filters applied.

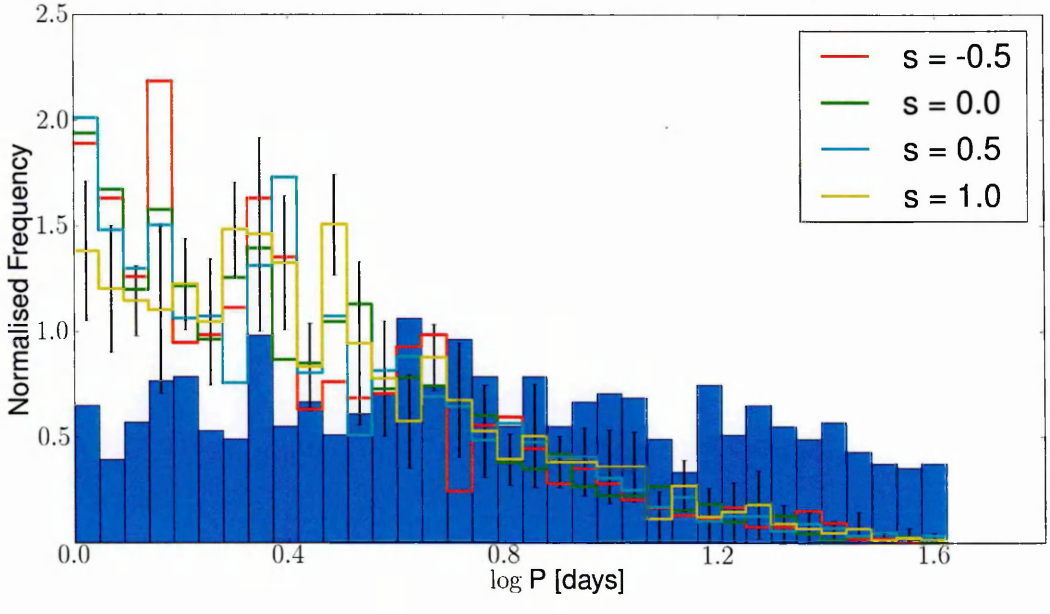



Figure 4.15: Similar to Fig.:4.13, except the synthetic stars are filtered to be only systems in which at least 1 star has evolved off of the MS and is a GB. Note there is no information in the Kepler EB catalogue about stellar classification thus there are no extra filters applied.

### 4.4.4 EBAI parameters of the synthetic sample

With the sample of synthetic EBs I investigate how EBAI affects the results and whether it can be used to determine the IMRD. Fig. 4.16 shows the temperature ratio distribution for the choice of IMR values and the temperature ratio distribution of the actual Kepler EB catalogue. There is a peak, present in all the IMRs, at $T_2/T_1 \sim 0.75$ and $T_2/T_1 \sim 0.9$ with the distribution being U-shaped in between. All the distributions peak at the same point, though with different heights, suggesting that there is something in the underlying EBAI training set that is causing this. These spurious peaks can be seen in both training sets, even when running the training set through the light curves of the other training set (figs. 4.8-4.11)

Clearly, none of these are a viable fit to the data, however there is a difference between the different IMR laws which suggests that it is possible to determine the IMR despite the bias introduced by EBAI. To be able to determine the true IMRD I would need to match the training set closer to that used in the actual Kepler EB catalogue, or run the actual Kepler EB catalogue through my training set.

Fig. 4.17 shows the true distribution of the temperature ratio distribution for the synthetic EBs compared with the Kepler catalogue. There is a peak at $T_2/T_1 = 1.0$ as these systems are the most luminous, we see a greater fraction of the Galactic volume they occupy. These systems also have a higher transit probabilities: assuming both components are in the same evolutionary state, we then have $T_2/T_1 \sim M_2/M_1 \sim R_2/R_1$, and equal radii maximise the range of inclinations of that can eclipse. This holds as long as both stars have the same exponent in the mass-temperature relation $M \propto T^{\alpha}$.

There is a slight peak at $T_2/T_1 \sim 0.65$ which increases with decreasing $s$. As most of the primaries are approximately solar mass, then this peak corresponds to a secondary of $M_2 \sim 0.7$. However, at this point ($M_2 \lesssim 0.7 M_{\odot}$) the efficiency of hydrogen burning decreases, thus both stars no longer have the same mass-temperature relation (Baraffe et al., 1998). As the mass decreases, $\Delta T_{eff}/\Delta M$ decreases as well. Thus at $T_2/T_1 \lesssim 0.7$ there are a wider range of possible mass ratios that can correspond

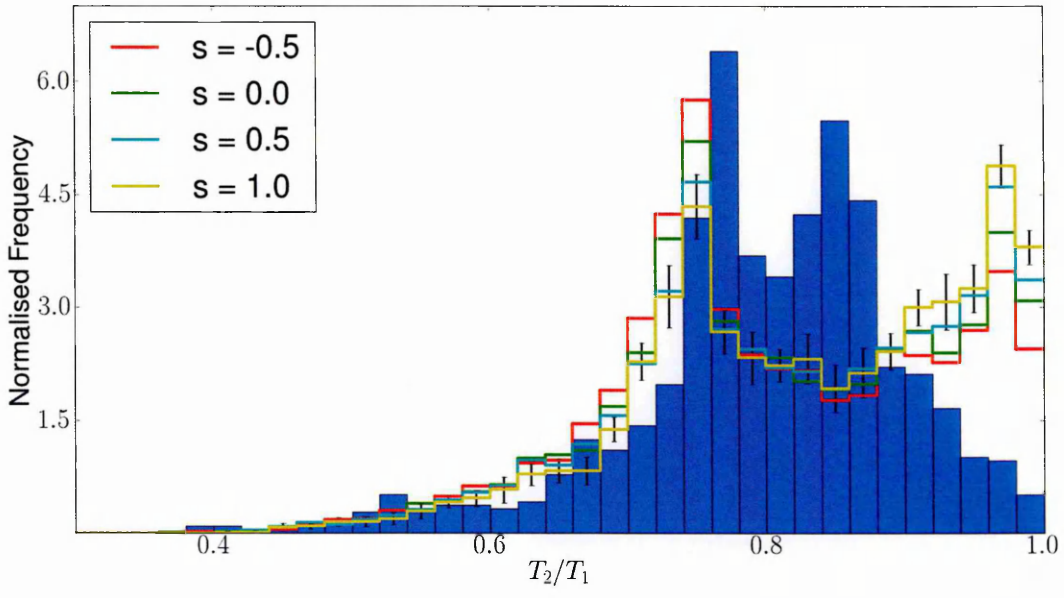



Figure 4.16: Filled histogram: the $T_2/T_1$ distribution of the Kepler EB catalogue, for detached and semi detached systems with period between 1-40 days. Coloured lines: the $T_2/T_1$ distribution of the synthetic catalogue as derived by EBAI, for different IMR values of $s$. Systems are filtered to include only detached and semi detached systems with periods between 1-40 days, with the results being averaged over 10 runs. Representative 1 standard deviation error bars are shown.

to a single temperature ratio, thus they appear to "pile" up at the extreme temperature ratios. As $s$ decreases we get more extreme mass ratio systems and this effect becomes larger. The spread in the size and location of the peak is due to the different IMRDs and how strongly they each weight these extreme systems. In this region the secondaries are small compared to the primary, so the different transit probabilities and luminosities contribute less to the shape and spread of the peak than at $T_2/T_1 = 1$.

Interestingly it appears that the bias EBAI introduces, may in fact help with determining the IMRD. EBAI condenses the EBs over a smaller range of $T_2/T_1$ values, than they truly have. Thus the peaks become more pronounced and as each bin has more values it may be possible to drive the uncertainties down.

Figs. 4.18 & 4.19, show the fractional radii distribution as derived by EBAI and the true distribution. Again, we can see that the synthetic EBs do not fit the Kepler data. It also appears that we can not distinguish between the different IMRDs as they are all within uncertainties of each other. Fig. 4.19 shows the true distribution of

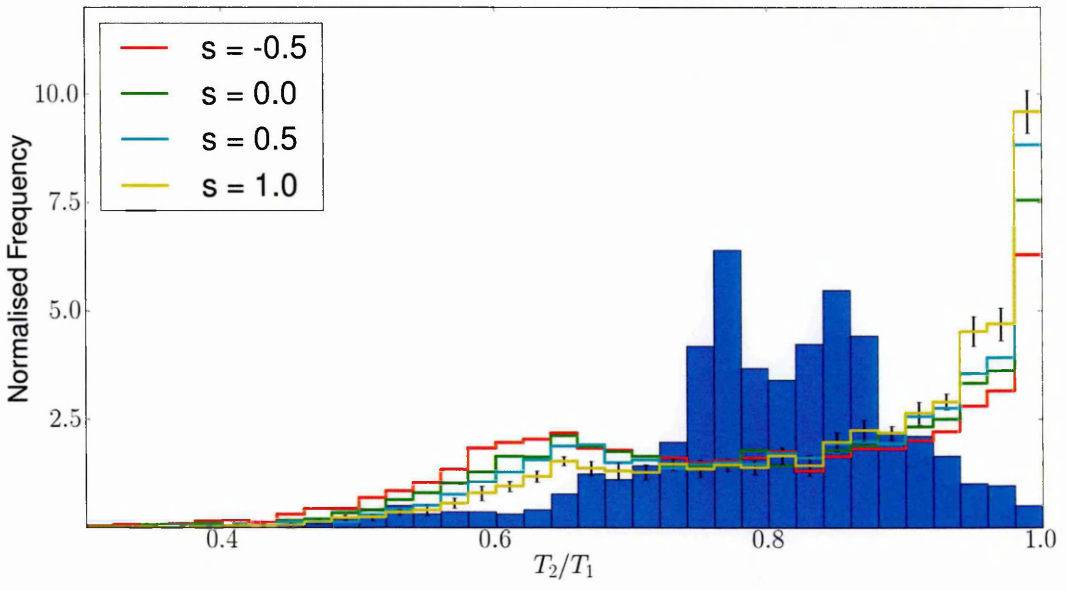



Figure 4.17: Filled histogram: the $T_2/T_1$ distribution of the Kepler EB catalogue, for detached and semi detached systems with period between 1-40 days. Coloured lines: the true $T_2/T_1$ distribution of the synthetic catalogue as derived by BISEPS, for different IMR values of $s$. Systems are filtered to include only detached and semi detached systems with periods between 1-40 days, with the results being averaged over 10 runs. Representative 1 standard deviation error bars are shown.

$(R_1+R_2)/a$ as compared to the Kepler values. The EBAI fractional radii distribution shows a strong similarity with the true fractional radii distribution, thus it appears that EBAI has successfully fitted the distribution. Which is what we expect from the training sets (figs. 4.8 & 4.9). However we have been unsuccessful in reproducing Kepler's peak at ~ 0.1, which was seen in figs. 4.8 & 4.10. This suggest that I need to compare the Kepler EBs to the synthetic EBs, with the same EBAI training set.

I do not show the $e\sin\omega, e\cos\omega$ or inclination plots based on the fact that the parameters are all chosen randomly and thus do not convey any information about the underlying distributions.

### 4.4.5 Variation of the binary fraction

I now investigate whether it is possible to constrain the IBFD from the EB data. There is evidence from Raghavan et al. (2010) that the binary fraction increases with spectral type, with the most massive stars having the highest binary fractions. However translating the results of Raghavan et al. (2010) which describe the current bi-

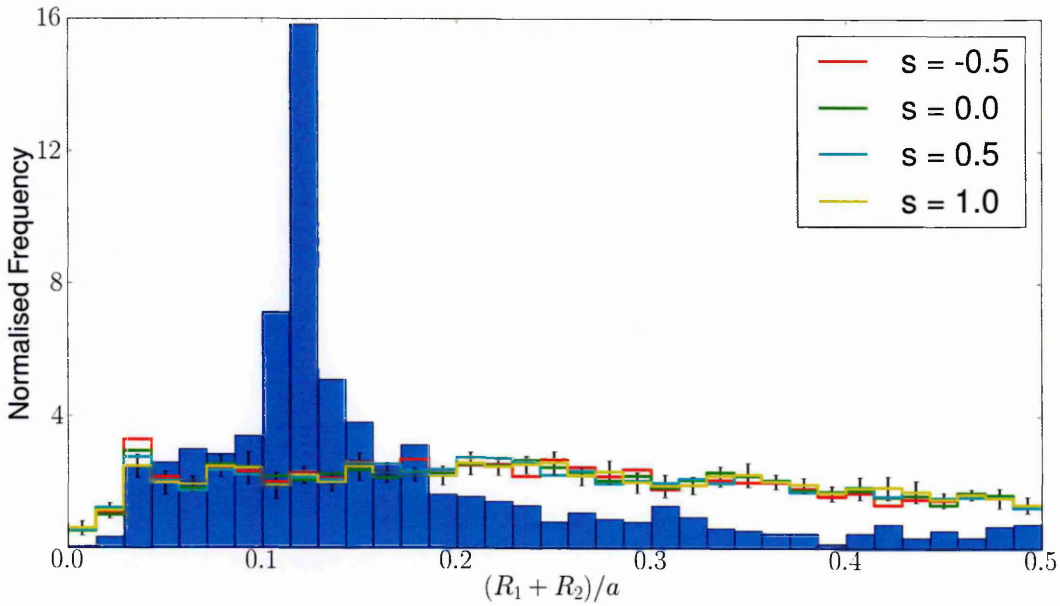



Figure 4.18: Similar to Fig. 4.16, except for the $(R_1 + R_2)/a$ distribution, as derived by EBAI.

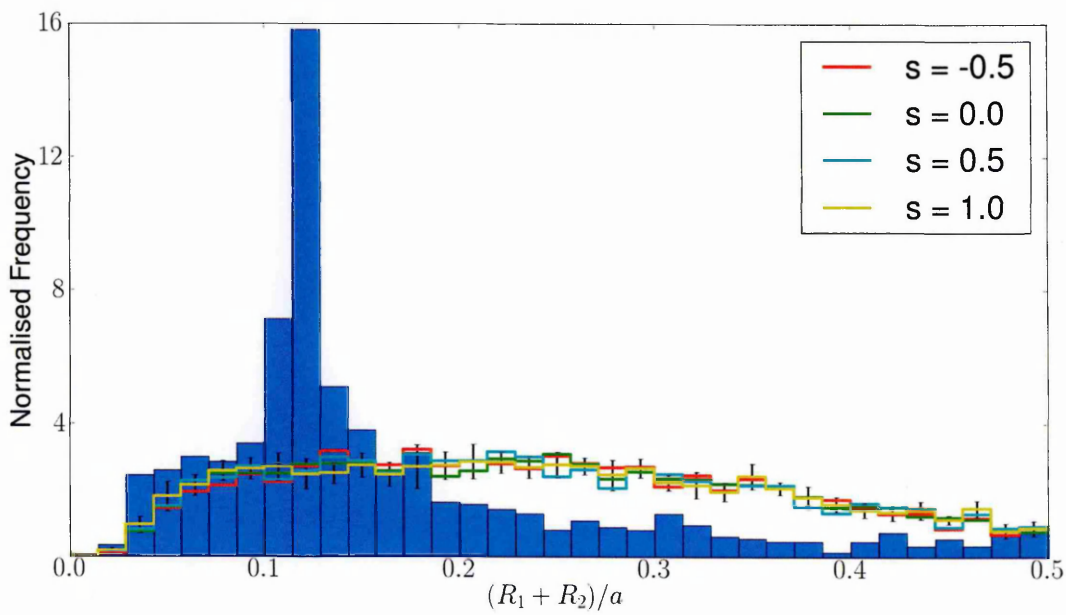



Figure 4.19: Similar to Fig. 4.17, except for the $(R_1 + R_2)/a$ distribution, as derived from BISEPS.

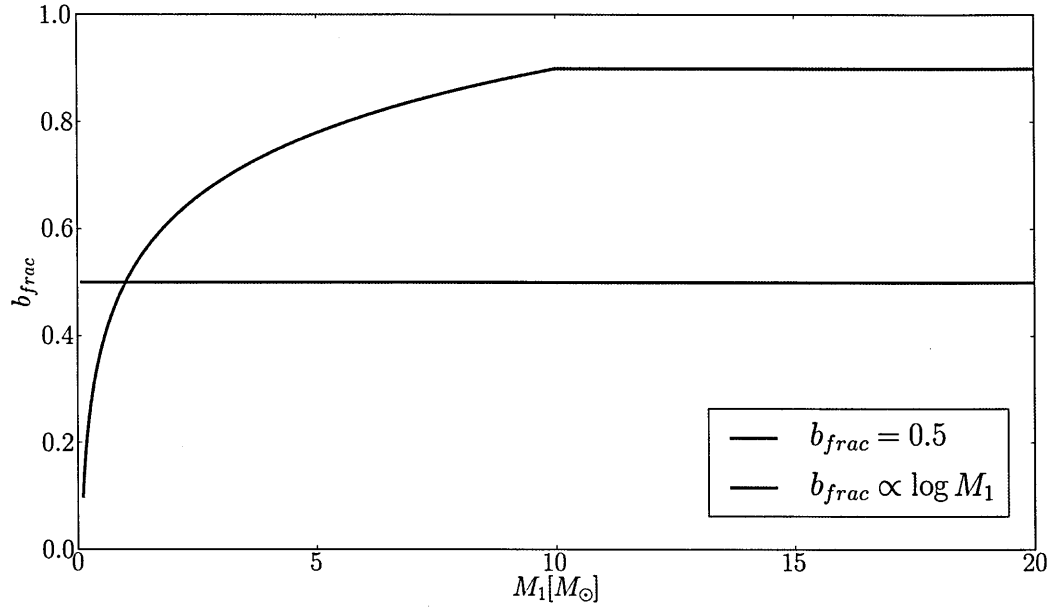



Figure 4.20: The two different IBFDs considered here.

nary fraction as a function of spectral type into the initial binary fraction as a function of mass, is non trivial. Therefore I guess a form of the IBFD that should give similar results to that of Raghavan et al. (2010). For the exercise here the choice of functional form is less about finding a realistic IBFD and more about determining whether we can see the difference between the different cases.

I test the following distributions; $b_{frac} = 0.5$ constant for all masses and a flat $\log M_1$ distribution, which can be seen in fig. 4.20:

$$b_{frac} = min(0.9, 0.4 \times \log M_1 + 0.5) \tag{4.25}$$

For both cases I assume a $s = 0$ IMR law and all other distributions are the same as previously stated.

From table 4.5 we can see the number counts for different functions of the IBFD. The model labelled $b_{frac} = 0.5$ is the same case as the model labelled $s = 0$ from table 4.3. For the $\log M_1$ distribution we can see that there is a increase in the number of EBs, due to the increase in high mass systems which are more luminous. The closest match in number counts to the Kepler data is the $\log M_1$ distribution.

In table 4.6 we can see the break down between detached and semi detached

| $b_{frac}$ | Number targeted systems | | | |
|---|---|---|---|---|
| | Total | MS+MS | GB+MS | GB+GB |
| 0.5 | $1035 \pm 15$ | $913 \pm 16$ | $118 \pm 6$ | $4 \pm 1$ |
| $\log M_1$ | $1297 \pm 27$ | $1144 \pm 27$ | $148 \pm 9$ | $6 \pm 1$ |
| Kepler | 1244 | | | |

Table 4.5: Similar to table 4.3, the number of EBs detected for different evolutionary types as a function of the adopted IBFD. Values are quoted as the total number of objects in the target selected sample, in Kepler equivalent counts. The numbers quoted are for detached and semi detached systems with periods between 1-40 days. GB contains both HG and GB stars. Uncertainties are 1 standard deviations.

| s | Percentage of EBs (%) | |
|---|---|---|
| | Detached | Semi detached |
| 0.5 | $89.65 \pm 1.46$ | $10.38 \pm 0.74$ |
| $\log M_1$ | $87.91 \pm 2.42$ | $12.09 \pm 0.80$ |
| Kepler | 94.35 | 5.64 |

Table 4.6: The fraction of EBs detected as detached or semi detached systems for different values of the IBFD. Values are quoted as percentages of the total number of EB in the synthetic sample. Periods are constrained to be within 1-40 days. The Kepler EB catalogue sample has been rescaled to include only the detached and semi detached systems. Uncertainties are 1 standard deviations.

systems. It can bee seen that the $\log M_1$ has a higher proportion of semi detached systems. This is because the $\log M_1$ distribution favours higher mass systems. These higher mass EBs must orbit at a shorter orbital separation, to stay within the period bin. Thus these systems are more likely to fill their Roche lobes and undergo mass transfer. Neither model fits the Kepler result, possible to due to noise issues discussed previously, though Prša et al. (2011) cautions against relying too heavily on the classification of systems due to degeneracies in the light curves. If we considered the classification parameter $c$ (section 4.3.3), then we could be more certain that we are comparing the same types of systems.

We can see in fig. 4.21, that similar to fig. 4.13, that we can not distinguish between the different IBFD distributions based on the period alone.

Fig. 4.22 shows the temperature ratio distribution, while fig. 4.23 shows the fractional radii distribution, both of which show that the IBFD is not resolvable in the data, and neither of them fits the Kepler EB catalogue data.

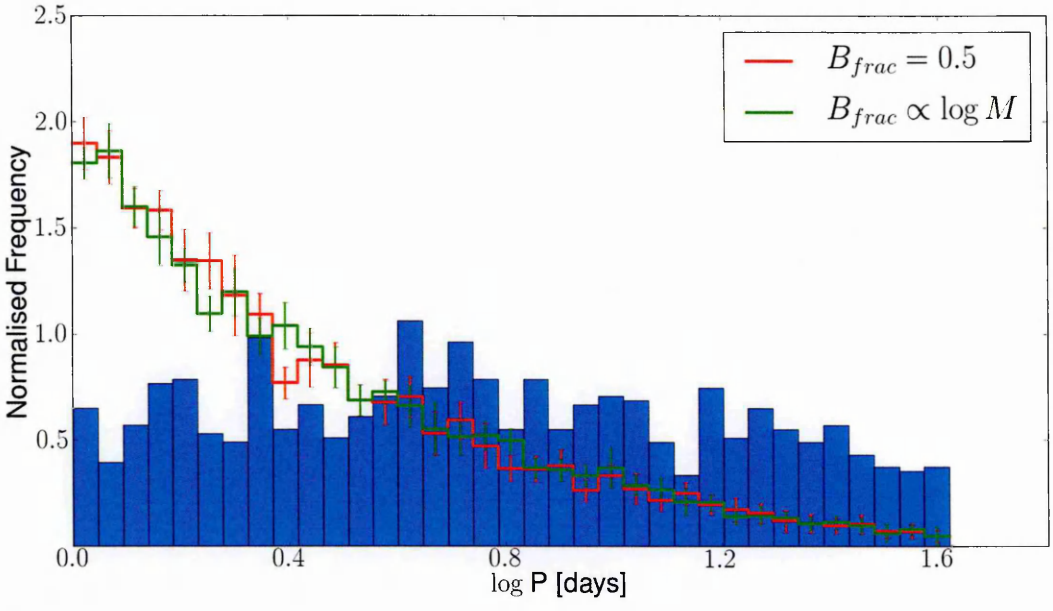



Figure 4.21: Filled histogram: the period distribution of the Kepler EB catalogue, for detached and semi detached systems with period between 1-40 days. Coloured lines: the period distribution for different binary fractions. Systems are filtered to include only detached and semi detached systems with period between 1-40 days, with the results being averaged over 10 runs. One standard deviation error bars are shown.

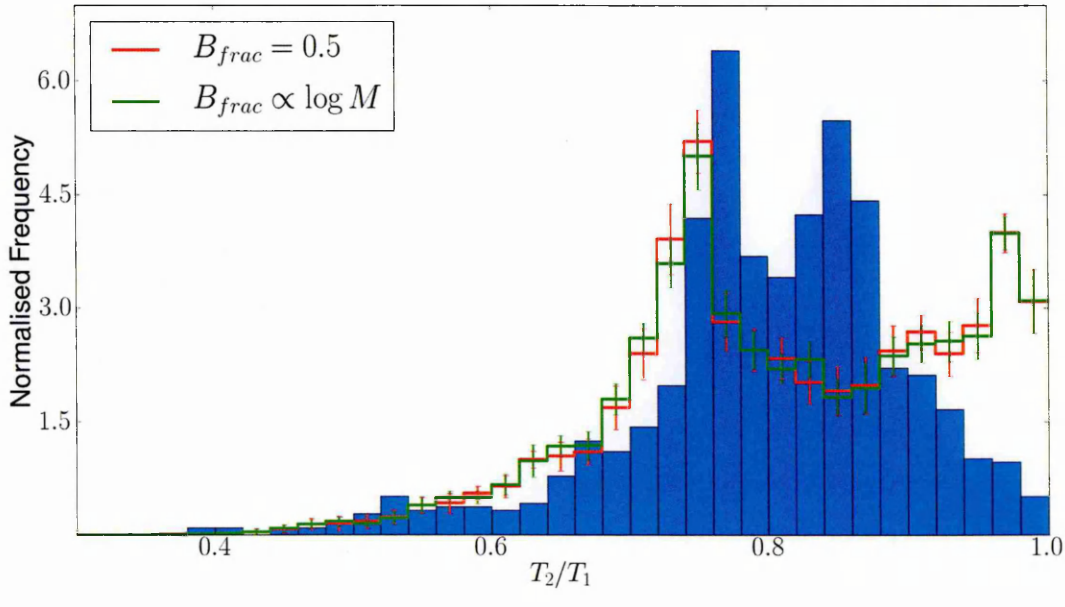



Figure 4.22: Filled histogram, the $T_2/T_1$ distribution of the Kepler EB catalogue, for detached and semi detached systems with period between 1-40 days. Coloured lines, the $T_2/T_1$, as derived from EBAI, distribution for different binary fractions. Systems are filtered to include only detached and semi detached systems with period between 1-40 days, with the results being averaged over 10 runs. One standard deviation error bars are shown.

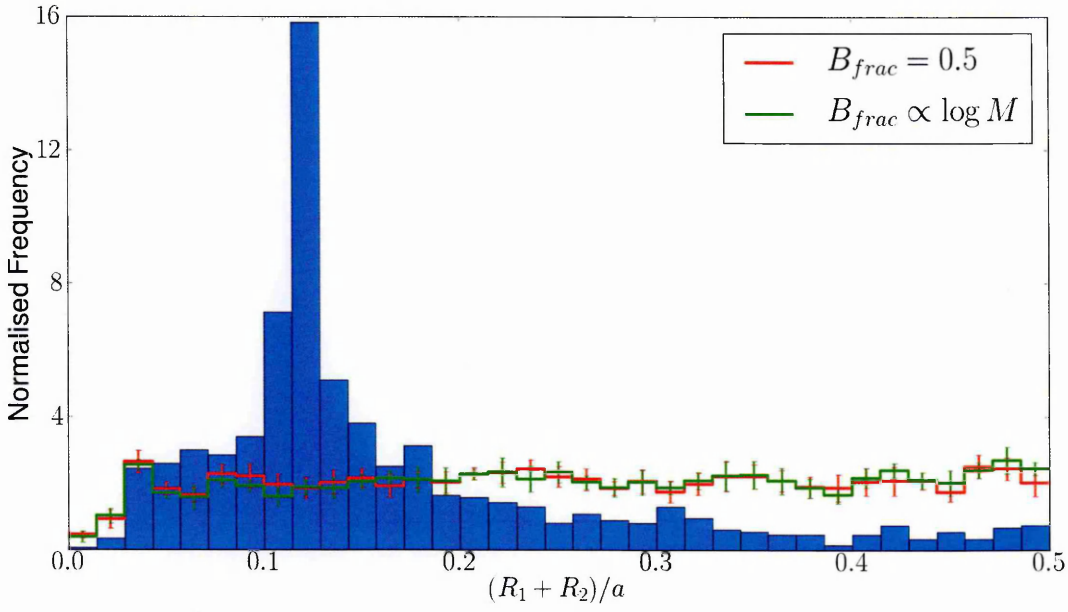



Figure 4.23: Similar to Fig. 4.22, except for the $(R_1 + R_2)/a$ distribution, as derived by EBAI, for different binary fractions.

## 4.5 Discussion

I have created a model for the detection of eclipsing binaries in the Kepler field and passed the resulting synthetic light curves through a recreated Kepler EB pipeline. I have then varied the Galactic distribution functions that underpin the population models to determine whether it is possible to distinguish between different input parameters from a comparison with the Kepler EB catalogue. I found that my models currently do not reproduce the Kepler results, but they do show as a proof of concept that it is possible to distinguish different initial distribution functions. I now discuss the different models I used and how they affect the results.

### 4.5.1 Noise model

The resulting number of synthetic EB systems that are detected depend crucially on the amount of noise inserted into the light curve. I tested the sensitivity of the results by introducing a random multiplier into the intrinsic variability in an attempt to more closely mimic the results of Jenkins et al. (2010b) for the actual Kepler data. For the $s = 0$ case I scaled the noise from the intrinsic variability by a random number

chosen from a log-flat distribution between 1.0 and 10.0. This has the effect that most dwarfs are no longer limited by the instrumental noise but by their intrinsic variability.

I found that by increasing the noise in this way the number of EBs decreased, which would result in a poorer fit of the number of EBs in the Kepler catalogue. The period distribution, with the increased noise, becomes even more weighted towards short period systems. Therefore simply increasing the noise for all systems will not work as we end up with too many short period systems. This suggests that to improve the overall agreement the noise specifically for short period systems needs to be increased (or at least systems that would preferentially be in short periods). This way I can decrease the number of short period systems, while preserving the longer period systems. One way of looking at this is to compare the results for the $\log M_1$ IBFD with the flat IBFD, where we see that the fraction of semi detached systems increases. This occurs as these systems are preferentially more massive than in the flat IBFD case. As these are more massive systems then on average they must have smaller orbital separations, such that the period cut of $P_{orb} < 40$ days is preserved. With smaller orbital separations these systems would be more likely to be in the semi detached state.

Therefore further investigation is needed to determine whether the noise should be increased for either short period systems in general or heavier systems which have preferentially small separations. Accretion might be responsible for the increased noise in semi detached systems (Patterson, 1981). Alternatively the noise could be increased for high mass systems. For systems with $M > 2M_{\odot}$ equation 4.8 is dependent only on the granulation term, as the stars are now in the $\delta$-Scuti instability strip.

Another consideration is the case of blended noise sources. This is where the flux from a background star is measured inside the optimal aperture of the target star. If the background star is a variable object, then its variability will be superimposed on the target star. Its amplitude will be suppressed by a factor based on the

relative flux contribution in the optimal aperture. Debosscher et al. (2011) found ~ 4000 variable objects, based on an automated analysis of the Kepler light curves. In this sample 50% of the variable objects were due to contamination from background sources. This sample only includes those systems where they could detect and classify the variability mode. Thus the number of systems which are contaminated, but the variability is small enough to only be considered "noise", would be expected to be higher.

During an eclipse a star blocks out an amount of $\Delta F/F$ of the flux from its companion. Therefore it should also block out $\Delta F/F$ of the noise, assuming the intrinsic variability is evenly distributed on the star's surface. However currently the noise from the companion is assumed to be constant. The noise during the eclipse thus should be rescaled down, based on the $\Delta F/F$ that is blocked. This will increase the number of systems where one component is significantly more variable than the second component, for instance GB+MS systems. As we will now be more likely to detect the eclipse from the less variable star, this will increase the number of synthetic EBs, independently of the orbital period, which will allow the synthetic catalogue to maintain consistency with the Kepler EB number counts.

### 4.5.2 Period doubling

A factor that has be ignored in this work so far is period doubling. To determine the period of an EB the primary and secondary eclipses must be determined in the light curve, either by an automated code or by eye (Prša et al., 2011). The light curve is then phase folded on a trial period, based on the separation of the eclipses, and a goodness of fit statistic computed. However if the secondary eclipse is undetectable, either by being smaller than the noise or missing due to interruptions in the data, then a primary eclipse may be mistakenly selected as the secondary. This would then lead to the EB having a derived orbital period twice that of the true orbital period. As the two eclipses would also be the same size therefore the apparent number of equal sized systems would increase.

This could be added to the model by setting synthetic systems with undetectable secondaries, as having twice their true orbital period. We would also need to compute a new synthetic phase folded light curve with two primary eclipses visible to pass to POLYFIT and EBAI. Modelling this, would shift the synthetic period distribution to longer periods and shift the $T_2/T_1$ distribution towards $T_2/T_1 = 1$. While the period shift would improve the overall fit, the improvement due to the shift in the temperature ratio is less certain. This is an area for further work.

Currently the model predicts less than 10 EBs would be expected to have a detected primary and undetectable secondary eclipse. This implies that period doubling is not a significant effect. However it may become more significant if the noise is increased, as discussed in section 4.5.1, such that eclipses become less detectable.

### 4.5.3 Eclipsing DERBs

To maintain the correct number of systems when the IMRD is varied, a new normalisation scheme was used. As part of this a new classification of binary was devised to handle limitations in the stellar evolution code. These systems with a very low mass companion are merely a mathematical tool designed to maintain the normalisation of the IMRD. We would expect some of these systems to eclipse and that the eclipses would be detectable by Kepler. However as BISEPS has no evolutionary model for them we cannot determine their masses or orbital separation. Therefore they can not be added to the synthetic EB catalogue. These systems could explain part of the suppression in the number of EBs seen in the low $s$ distributions. The solution to this is to replace the evolution code with one that can handle low mass systems. This would also allow the inclusion of the low mass ($M < 0.1M_{\odot}$) singles and primaries in binaries,which would benefit the normalisation of the SFR.

### 4.5.4 Performance of EBAI

The POLYFIT code was designed to re-sample the existing light curves and produce a light curve that has points equidistant in phase space. However I showed that using

the parameter choices of Prša et al. (2008, 2011) it can misidentify the depth of the secondary and sometimes the primary eclipse, by up to 20% of the eclipse depth. This is significant as the ratio of the eclipse depths is what drives EBAI to derive the temperature ratio which is the worst performing of the EBAI fits. However, as I am predominately concerned with comparing my results with the Kepler EB catalogue results (and not the true parameters) in principle this is not as serious an issue, as both sets will be affected similarly.

From section 4.4.4 we can see that the choice of EBAI training set can introduce biases into the resulting distributions. I chose for simplicity to use a training set that was based on a binary distribution similar to that in the synthetic EB catalogues. This was based on the assumption that it should perform better on systems it has already seen. However, I have found that none of the synthetic catalogues fit the Kepler EB catalogue satisfactorily. Testing the catalogues with the other training set shows an even poorer fit to the Kepler data. Thus, to be able to compare to the Kepler EB catalogue I need a training set that more closely resembles that used in Prša et al. (2011). This may relate to the input distributions of systems used in the training sets or how the light curves are generated (noise, limb darkening, third light etc). It is also possible that there are unspecified changes in the method between Prša et al. (2008) and Prša et al. (2011) that need to be taken into account in order to reproduce the training set. A solution to this problem would be to use the same training set on both the synthetic and Kepler EB catalogues. This could be achieved either by using Prša et al. (2011) training set or by downloading the phase folded light curves for the Kepler EB catalogue and use the training sets derived here. This way it would not matter if there are differences in the method or parameters used to generate the training set.

### 4.5.5 Effect of eccentricity

I have ignored for the moment the eccentricity distribution of binaries due to the computational costs involved in modelling an extra distribution function. However,

for a study of EBs the eccentricity does play an important role. In the Kepler EB catalogue, 60% of the detached and semi detached systems with $1 < P_{orb}/\text{days} < 40$ have $e > 0.1$, thus they make up a significant population of short period EBs.

From the evolution point of view short period binaries have eccentric orbits only for a short time compared to their lifetime, due to tidal circularization (Hurley et al., 2002). However when considering the EB sample, eccentric systems are more probable due to increased transit probabilities. Depending on the precise alignment of the orbit the duration of an eclipse may be shorter or longer which will affect the probability as well (Burke, 2008). Thus eccentric binaries will have a different selection function to non-eccentric systems. To compute this selection function I would need to consider their eccentricity induced binary evolution, as well as the evolution of the eccentricity and period, via tidal forces (section 1.1.1.1). This can only be achieved by including an additional distribution function in the phase space covered by the stellar evolution tracks.

The "noise" for eccentric systems will also be different from that of circular systems. Roche lobe overflow mass transfer can occur in eccentric binaries, under certain conditions, even though in principle the stars should have circularized before Roche lobe overflow (Davis et al., 2013). This mass transfer rate is variable depending on the orbital phase of the system, introducing another source of noise in short period systems, that will be time varying over the orbital cycle. Tidal distortions caused by an eccentric orbit will induce tides on the star (section 1.1.1.1), introducing an extra source of noise as the stars become deformed and then re-equilibrate. These terms may be counter balanced by the fact that as the star is no longer spherically symmetric the asteroseismic signal will vary (Springer & Shaviv, 2013). Thus there are several changes needed in the noise model for close, eccentric binaries. Introducing a fully comprehensive EB model of eccentric binaries is left for future work.

### 4.5.6 Higher order systems

The seminal study of Raghavan et al. (2010) found ~ 9% of all systems contain three stars and a further ~ 3% have more than three. Thus higher order systems make up a non-negligible fraction of the stellar population of the Galaxy, a contribution currently not considered in the model. The contribution of a third light source in these systems will decrease the measured eclipse depth, as well as introduce another source of stellar noise. Higher order systems also pose an issue with the normalisation of the model. The IBFD used here assumes that 50% of systems are binaries. But based on the Raghavan et al. (2010) results, what should really be assumed is that 50% of systems are single stars, the remaining being binaries and higher order systems. Thus the number of binaries will decrease.

The full solution to this would require introducing these systems as a new population in the model. However each additional star would require a new set of initial distributions (mass, orbital separation and eccentricity) that would increase the runtime. In principle I could assume that all higher order systems are hierarchical, for instance a close binary with a distant third body. This way would assume that the third body plays no evolutionary role, third bodies can of course alter the eccentricity and inclination via the Kozai mechanism (Kozai, 1962). I would then select a binary from the synthetic catalogue and select a single star from the stellar track library, where the selected single star is based on a set of third body initial distribution functions (IDFs) and constrained to be the same age as the binary. However, this approach would not work as the probability of selecting a system in the synthetic catalogue depends on the total luminosity. This alters the fraction of the Galactic volume the system is visible in. A third body would thus allow the system to be seen to a greater distance and therefore the probability of the binary in the first place changes.

Therefore attempting to model a triple star system without performing the full integral over all the initial distributions would not be possible. To get round this would require the new model as described in chapter 6.2. Here each system is cho-

sen based on its IDFs, evolved and then its detectability is determined. The runtime of this method is then independent of the number of initial distribution functions. Triple stars would then be selected based on a distribution function. then only the stars selected by the IDF are evolved. This way only the necessary triple star systems are evolved rather than all possible triple stars, as would be required with the current model.

This approach would still require the assumption of the third body being at a large separation from the binary such that it plays no role in the evolution. This is because the BSE code does not handle three stars simultaneously. This is a reasonable assumption as systems where the third body does interact with the binary, Kozai-Lidov (KL) oscillations can be induced. These drive rapid changes in the eccentricity which can force the binary to merge on short time scales (Antognini et al., 2014).

## 4.6 Conclusions

I have created a model for generating synthetic eclipsing binaries catalogues with different initial distribution functions. I have modelled the noise predicted during an eclipse and used this to determine whether an eclipsing binary would be detected by Kepler. Those systems that were detected were then passed though the same procedures as the Kepler EB catalogue to derive Kepler based binary parameters.

I have shown that, at this time, my model is incomplete and thus unable to accurately reproduce the Kepler results. Thus I can not constrain any of the initial distribution functions from my population models. However I have shown as a proof of concept that it is possible to distinguish between different IMRDs but not different IBFDs. To distinguish between the different IMRDs requires the use of the EBAI code, which I have shown has issues in reproducing the correct parameters.

I have shown that the EBAI derived parameters are biased and not representative of the true distributions. Hence the only way to be able to interpret the Kepler EB

parameters is to compare the distributions against a theoretical population model. Additional I have shown that the model and the Kepler data need to be compared using the same EBAI training set, such that the same bias is introduced into both results. This bias in the Kepler derived parameters has implications for the rates of false positives, which are based on the sample of Kepler EBs.

# Chapter 5

# Asteroseismic binaries

In this Chapter I derive a model for the distribution and detectability of asteroseismic binaries. These are systems where we can detect pulsations in both stars in a binary and in general implies both objects are giants and have similar luminosities. This enables us to calibrate the asteroseismology derived parameters against those derived from eclipse measurements or radial velocities.

I predict that there should be between 87 ($s = -0.5$) and 256 ($s = 1.0$) asteroseismic detectable binaries in the Kepler field, though the fraction that may be recovered by analysis of the light curves may be considerable less. I also determine that the rate of false positives for these asteroseismic binaries is $\lesssim 20\%$. I find that there is a no period dependence for the systems detected, except for a short period cut of $\sim 30$ days. Assuming that the asteroseismic signal is not suppressed in the presence of eclipses, I find that there should be between $34.0 \pm 6.0$ ($s = -0.5$) to $59.0 \pm 6.0$ ($s = 1.0$) asteroseismic eclipsing binaries (EBs) which disagrees strongly with the number of systems detected, 2 for $P_{orb} < 40$ days.

## 5.1 Introduction

The numbers and distribution of asteroseismic binaries was presented in Miglio et al. (2014). The numbers of astrophysical contaminants and the asteroseismic EB are additions to the work in Miglio et al. (2014).

Asteroseismology, the study of the pulsation modes inside stars, has been revolutionised by Kepler. By combining the high precision with the almost uninterrupted temporal coverage of the Kepler field it has been possible to detect the pulsation modes in 13,000 red giants (Stello et al., 2013) and measure the stellar properties; like mass, radius and temperature to < 10% (Huber et al., 2013; Chaplin et al., 2014).

The pulsations themselves can be split into two types. There are solar like pulsations ($p$ modes) caused by pressure gradients which act to restore the star after a perturbation caused by sound waves (Bahcall et al., 2001). Then there are $g$ modes where gravity is the predominant restoring force, where the pulsations are due to gas packets rising and falling due to buoyancy (Bedding et al., 2011). For both types of oscillations, what we detect is a series of different pulsation frequencies based on which frequency can form a standing wave inside the star and then also the harmonics of this frequency. The oscillation peaks for $p$ modes are spaced equally in frequency, while for $g$ modes they are approximately equally spaced in period (Bedding et al., 2011).

The $p$ modes form due to turbulence inside convective regions inside a star. This turbulence causes pressure fluctuations to form which drive sound waves, which can then set up a standing wave pattern which we see as a $p$ mode oscillation, as shown in fig. 5.1. These oscillations change the density of the material at the nodes of the standing wave, which we then see as a change in the luminosity due to the change in the temperature of the material. This is a stochastic process, the oscillations are due to many randomly perturbed regions which we assume act independently of each other. Therefore the change in luminosity we see is an average over many turbulent regions, with each region oscillating at a different phase (Kjeldsen & Bedding, 2011).

The $g$ modes are ignored for this work, as they are not relevant for the types of objects I wish to model. The pulsations arise due to gas bubbles raising and falling due to buoyancy. They are predominately generated in the core of the stars and

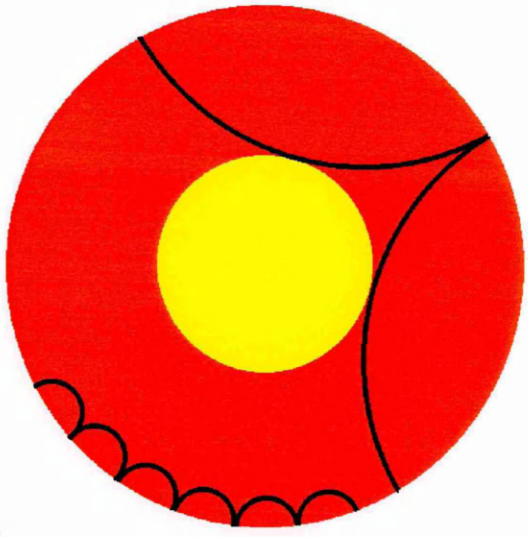


Figure 5.1: A schematic showing two sets of $p$-modes, denoted by black lines, travelling through the convective envelope of a star, which is denoted in red. The core is denoted in yellow. Only those $p$ modes that can form a standing wave inside the envelope are detectable due to the change in pressure at the surface nodes. Note figure is not to scale

trapped due to the convective envelopes of stars and thus difficult to detect on the surface (Bedding et al., 2011).

However, they can be detected in some objects in a *mixed* mode configuration. In systems where the $g$ modes in the core have frequencies similar to those of the $p$ modes, the two sets of oscillations can couple allowing the detection of the $g$ modes on the surface of the star (Handler et al., 2009; Bedding et al., 2011). Detecting these modes enables the distinction between shell hydrogen burning stars and core helium burning (CHeB) stars, due to the dependency the $g$ mode frequencies have on the size of the core (Bedding et al., 2011).

By measuring the individual pulsation frequencies detected from a star, in the same way earthquakes on Earth provide information on the Earth's interior, the interior structure of the star can be determined. This is because the pulsations ($p$ modes) are related to the local sound speed and thus track changes in density in the stellar interior (Brown & Gilliland, 1994). This was originally applied to the Sun, which is then called helioseismology, and was used to provide information on the interior of the Sun. Examples of this include the size of convective envelope (Brown

& Gilliland, 1994), the luminosity evolution of the Sun (Bahcall et al., 2001) and differential rotation in the Solar envelope (Schou et al., 1998). The pulsations manifest themselves as small, local, changes in the temperature and radius of the star, though only the temperature change is significant (Kjeldsen & Bedding, 1995). The combination of many small changes over the whole disc of a star then show as changes in the total luminosity of the star.

For this work I consider the case where both stars in a binary system are pulsating with $p$ modes, such that there are two *detectable* asteroseismic signals in the light curve. For this to happen however requires that both stars have similar luminosities, otherwise the pulsation signal from the fainter star is washed out. Primarily the pulsations will be detected in giant stars. This is because the pulsation signals are easier to detect in giants due to the frequencies being lower than in main sequence (MS) stars. This then allows the long cadence (LC) Kepler data to be used, which is available for all the objects Kepler observes as opposed to the short cadence (SC) data which is limited in quantity and temporal coverage.

Detecting these pulsating binary systems allows us to compare the parameters derived from asteroseismology with those derived from other binary methods, e.g. radial velocities or eclipses, thus allowing us to calibrate the asteroseismology models, for instance against the the 34 red giant eclipsing binaries found in the OGLE data set (Nataf et al., 2012). These systems can also be detected over the whole period range binaries can exist in, except for a short period cut where we are incapable of having two giant stars in a stable orbit. This is something not seen in any other binary detection method. As both stars must be giants and also have similar luminosities then they must also be of a similar mass and age. Thus these objects gives us a handle on the distribution of equal sized systems providing a constraint on the initial mass ratio distribution (IMRD). Discovering and modelling these systems is beyond the scope of this work.

In section 5.2 I discuss the model used to determine the detection criteria for the asteroseismic signals. Section 5.3.1 looks at the predicted number of systems

detected via this method, while sections 5.3.2-5.3.4 look at the distributions of the detected systems. Section 5.4 looks at various false positives and contaminants in the predicted sample. In section 5.5 the systems common to both the asteroseismic and EB catalogues is derived. While in sections 5.6 & 5.7 I discuss the results and present my conclusions.

## 5.2 Method

Using the method presented in Chaplin et al. (2011b) and used in Miglio et al. (2014), the detectability of a star's asteroseismic signal is derived. This value is computed from the scaling relations, which assume we can scale the results derived from the Sun to other stars (Kjeldsen & Bedding, 1995). These relations use the mass, radius and temperature of the star to derive the strength of the asteroseismic signal. This is then combined with a prediction of the noise for the star to derive the S/N of the signal. With the S/N for the signal a false alarm probability can be computed for the detection of the asteroseismic signal. Based on a cut, designed to minimise false positives, those systems with a low false alarm probability are selected to form a synthetic asteroseismic catalogue.

This work relies on the computation of the asteroseismic signal and granulation, which acts a a noise term for this work, as presented in section 4.3.1. The two principle components measured from a asteroseismic signal is the frequency of the peak amplitude signal, $\nu_{max}$ which scales as the acoustic frequency of the star, and the large frequency separation, $\Delta\nu$ which scales as the mean density of the star. Thus our aim in this work is to compute these quantities and use them to derive the strength of the overall asteroseismic signal. $\nu_{max}$ is the frequency at which the observed modes have their largest amplitudes (Chaplin et al., 2011c). This is assumed to be scaled from the solar value:

$$\nu_{max} = \nu_{max,\odot}\left(\frac{M}{M_\odot}\right)\left(\frac{R}{R_\odot}\right)^{-2}\left(\frac{T_{eff}}{T_{eff,\odot}}\right)^{-0.5} \tag{5.1}$$

where $\nu_{max,\odot}$ is the solar value of 3150 $\mu$Hz.

The second quantity $\Delta\nu$ is the frequency difference between radial overtones (Chaplin et al., 2011c):

$$\Delta\nu = \Delta\nu_{\odot}\left(\frac{M}{M_{\odot}}\right)^{0.5}\left(\frac{R}{R_{\odot}}\right)^{-1.5} \tag{5.2}$$

where $\Delta\nu_{\odot}$ is the solar value of 134.9$\mu$Hz.

The maximum amplitude of the asteroseismic signal can be computed as:

$$A_{max} = 2.5\beta\left(\frac{R}{R_{\odot}}\right)^{2}\left(\frac{T_{eff}}{T_{eff,\odot}}\right)^{0.5} \text{ppm} \tag{5.3}$$

where $\beta = 1 - exp(-(T_{red} - T_{eff})/\Delta T)$, $T_{red} = T_{red,\odot}(L/L_{\odot})^{-0.093}$, $T_{red,\odot} = 8907$K and $\Delta T = 1550$K. The parameter $\beta$ is a fitted correction factor used because the predictions of the amplitude get progressively worse as a star nears the $\delta$-Scuti instability strip. For any star inside the instability strip I set $\beta = 0.0$, thus they have no asteroseismic signal.

We then assume that the power due to the asteroseismic signal is entirely contained within a Gaussian envelope, with a mean of $\nu_{max}$ and FWHM of $\nu_{max}/2$. The contributions from the different $l\&n$ modes can be combined, where each scales as a function of the maximum amplitude, $A^2_{max}$, (Chaplin et al., 2011c). Therefore the total observed oscillation power and hence the signal, is:

$$P_{tot} = 1.55A^2_{max}\eta^2\frac{\nu_{max}}{\Delta\nu}B^2_{k,i} \quad \text{ppm}^2\mu\text{Hz}^{-1} \tag{5.4}$$

Here $\eta = sinc\left(\pi/2\left(\frac{\nu_{max}}{\text{Nyquist}}\right)\right)$ is a correction factor designed to reduce the strength of the signal the closer the pulsation frequency $\nu_{max}$ is to the Nyquist frequency of the observations. The quantity $B_{k,i} = L_i/(L_1 + L_2)$, is the luminosity fraction derived separately for each star, in single stars this is set as 1. This factor compensates for extra background flux in the measured signal due to the companion star. As $B_k$ enters into equation 5.4 as a squared quantity, only in systems where the luminosity of both components is similar can the asteroseismic signal be detected. We can see that even for two stars with the same luminosity the asteroseismic signal is 1/4 the

strength it would have if it were a single star.

Finally, the noise is calculated as

$$B_{tot} = \left(b_{instr} + P_{gran}\right)\nu_{max} \tag{5.5}$$

where $b_{instr}$ is the instrumental noise and $P_{gran}$ is the power due to the granulation variability:

$$P_{gran} = 0.1\left(\frac{\nu_{max}}{\nu_{max,\odot}}\right)^{-2} B_k^2 \quad \text{ppm}^2\mu\text{Hz}^{-1} \tag{5.6}$$

The instrumental noise, $b_{instr}$, is expressed as

$$b_{instr} = 2 \times 10^{-6}\sigma^2\Delta t \quad \text{ppm}^2\mu\text{Hz}^{-1} \tag{5.7}$$

with $\Delta t$ = 58.85s and $\sigma$, a semi-empirical RMS noise measured from Kepler data (Gilliland et al., 2010b):

$$\sigma = \frac{10^6}{c}\left(c + 9.5 \times 10^5 \left(14/K_p\right)^5\right)^{1/2} \tag{5.8}$$

and $c = 1.28 \times 10^{0.4(12-K_p)+7}$, which is the expected number of detections per cadence. Stellar activity is also important to consider in these systems, party for the increased noise (section 4.3.1) and also as it has been shown to to suppress the asteroseismic signal (Chaplin et al., 2011a). For the moment however, this is ignored.

The S/N ratio is thus:

$$S/N = P_{tot}/B_{tot} \tag{5.9}$$

This value can then be compared to a threshold $(S/N)_{thres}$ which is set such that there is a 1% chance of a false positive, based on a $\chi^2$ distribution with $2N_d$ degrees of freedom. Here $N_d$ is the number of frequency bins measured in the observational period. Only systems where the calculated probability that $S/N$ is greater than $(S/N)_{thres}$, is greater than 90%, do we assume that there is a detection. For binary systems I calculate this quantity for both stars and then require that both stars

have a probability greater than 90%.

As the frequency of the asteroseismic signal in a MS star is of order minutes, SC (cadence 1 minute) data is required to sample the signal. Thus for systems containing a MS star the amount of time available to observe the system is 30 days, the time approximately allocated for objects in SC mode (Gilliland et al., 2010b). For giants, the frequency is lower, thus the asteroseismic signal can be detected in the LC data (cadence 30 minutes). Thus the observational period for these is set to full Kepler observing period of 4 years (Miglio et al., 2014). For binaries where there are both types of stars in the same system the length of observations is set to the shortest period, 30 days. There is a final constraint applied in that the calculated $\nu_{max}$ for both stars must be less than the Nyquist frequency for the observing period selected.

## 5.3 Results

The aim of this work is to predict whether it is possible to distinguish different Galactic distribution functions based on the number of asteroseismic binaries Kepler could detect. None of these objects have been published yet, but there have been some detections (Chaplin priv.comm).

For this work I consider 4 sets of the IMRD $s$ parameter, ($s = -0.5, 0.0, 0.5, 1.0$). I also consider two initial binary fraction distributions (IBFDs) set as either a constant value (0.5) or log flat ($\log M_1$), see equation 4.25 and Chapter 4 for choice of parameters. All other Galactic parameters are as described in Chapter 2. I also include DERBs in these models; these act as single stars for the purpose of this work as the secondary will always be too faint to be detectable.

### 5.3.1 Number of detectable systems

Firstly, I consider the number of detections of asteroseismic binaries. The number of which can be seen in table 5.1 for both LC and SC detections, including cases where only one star is detectable. The numbers quoted are Kepler equivalent numbers,

| | | Number of detections | | | |
|---|---|---|---|---|---|
| $s$ | $b_{frac}$ | Both (LC) | Either (LC) | Both (SC) | Either (SC) |
| -0.5 | 0.5 | 87 | 10797 | 3 | 713 |
| 0.0 | 0.5 | 114 | 11637 | 6 | 740 |
| 0.0 | $\log M_1$ | 182 | 14914 | 5 | 902 |
| 0.5 | 0.5 | 181 | 11877 | 4 | 700 |
| 1.0 | 0.5 | 256 | 12081 | 5 | 714 |

Table 5.1: Number of predicted detections where we see the asteroseismic signal in both stars or in only one star, considering both the LC (favouring giant stars) and the SC (MS stars) data for the target selected sample, as a function of the IMRD and IBFD. These numbers have been scaled to show the Kepler equivalent counts, compensating for the variations in number of systems between BISEPS and Kepler (see table 4.4.1).

compensating for the fact that BISEPS generates different numbers of total objects and different numbers of target selected systems compared to Kepler (section 4.4.1). Thus these are the "true" counts as expected for Kepler, assuming a 100% recovery rate.

We can see in the table that the prediction is of order one hundred LC asteroseismic binaries and that there is a clear increasing trend in the number of detections as a function of the IMR $s$ parameter. This is due to the increase in similar mass systems, where both stars also have similar luminosities. The SC asteroseismic binaries detections show no overall trend however there are only a few detections, thus it is difficult to draw a conclusion from this. The low numbers are due to the reduced observing time for these systems in SC mode.

The number of LC systems with 1 detection is of order the total number of systems ($\sim 13{,}000$) with detected asteroseismic signals in the Kepler field (Stello et al., 2013), while the SC numbers are of the order of the number with well determined parameters ($\sim 500$) (Chaplin et al., 2014). This suggests that the analysis of the presumed single stars may need to take into account possibly undetected companions which may alter the derived parameters for the star (Springer & Shaviv, 2013). On the other hand the models may over predict the detectability, as this sample does not include the single stars, yet we predict more detections than have currently been made.

| | | Number of detections | | | |
|---|---|---|---|---|---|
| $s$ | $b_{frac}$ | Both (LC) | Either (LC) | Both (SC) | Either (SC) |
| -0.5 | 0.5 | 214 | 34895 | 4 | 1063 |
| 0.0 | 0.5 | 265 | 37569 | 9 | 1101 |
| 0.0 | $\log M_1$ | 418 | 46860 | 6 | 1360 |
| 0.5 | 0.5 | 417 | 38760 | 7 | 1060 |
| 1.0 | 0.5 | 561 | 39721 | 8 | 1060 |

Table 5.2: Similar to table 5.1 except now for all systems in the Kepler field down to $K_p \leq 16$

The difference in numbers between the flat and $\log M_1$ binary fractions is due to the $\log M_1$ preference for high mass stars. These stars are more luminous than lower mass stars so there is an increased probability in seeing them. They also have a lower $\nu_{max}$ (as they are larger) thus being easier to detect as an asteroseismic binary via equation 5.4.

In table 5.2 we can see the number of detections for the complete sample of stars in the Kepler field down to $K_p \leq 16$. The idea behind this is that Kepler observed additional stars not in the original target selection list due to the guest observation program. Thus table 5.2 provides an upper estimate on the total number of possible detections, having assumed that Kepler observed all stars in its field of view.

We can see that the the number of SC detections, for both single and double detections, increases by a third. The LC detections in both stars approximately double, while the LC detections in a single star are 3× as large. This is due to the change in the number of objects considered (target list is ~ 50% of all the Kepler objects), combined with the bias to observe MS systems in the target selected sample.

### 5.3.2 Asteroseismic binary distributions

I now discuss the distribution of the asteroseismic binaries, for $s = 0.0$, in $\log T_{eff}$ - $\log L$ space in fig. 5.2. Note the number of systems plotted is different to that of table 5.1 as the table shows the Kepler equivalent numbers. Firstly, looking at the grey dots, which shows the distribution of asteroseismic single stars and asteroseismic DERBs, we can see the imprint of the thin/thick disc. The thick line running through

the giant branch and the large bulge at $\log L \sim 2$, $\log T_{eff} = 3.7 - 3.8$, is due to thick disc objects. These are constrained in the $\log T_{eff}$ - $\log L$ space as there is a limited range of masses they can have while still being a giant during the current epoch. This is due to the lack of star formation in the thick disc after 3 Gyrs. The bulk of the asteroseismic systems have $\log L \sim 2$, $\log T_{eff} = 3.7$, which is the thin disc CHeB branch. Predominately the asteroseismic binaries have at least one component in the CHeB phase, ~ 80% of the total compared to only 15% in the synthetic binary sample. This relative increase is due to increased detectability due to the increased $P_{tot}$, as CHeB stars have lower values of $\nu_{max}$ and $\Delta\nu$.

There are a few detections in the MS region and HG region. The number of MS objects is limited due to the shorter period of observations. The number of HG stars is limited as the S/N is small. From equation 5.4 there is an overall $R^{3.5}$ dependence, thus the total signal, $P_{tot}$, will be smaller for HG stars compared to GB/CHeB stars. We also can see that HG stars will have a higher $\nu_{max}$, due to its $R^{-1.5}$ dependence. This implies that the $\eta$ term in equation 5.4 will be smaller (than in GB/CHeB stars), as $\nu_{max}$ will be closer to the Nyquist frequency.

### 5.3.3 Mass ratio distribution

I now consider the mass ratio distribution of the detected asteroseismic binaries, as seen in fig. 5.3. The mass ratio distribution is very strongly peaked at 1, which is unsurprising as we can only detect systems with luminosity ratios close to 1 which would generally imply equal masses. The systems where the mass ratio is less than ~ 0.9 are either systems that have undergone mass transfer or where the primary is descending along the giant branch or ascending the asymptotic giant branch (AGB), while its companion is still ascending or descending the giant branch, as seen for some objects in fig. 5.2. Thus it has a similar luminosity as its companion, but is at a later stage of evolution and therefore must have a higher initial mass to have evolved quicker.

We can also see in fig. 5.3 that the present day mass ratio distribution (PDMRD)

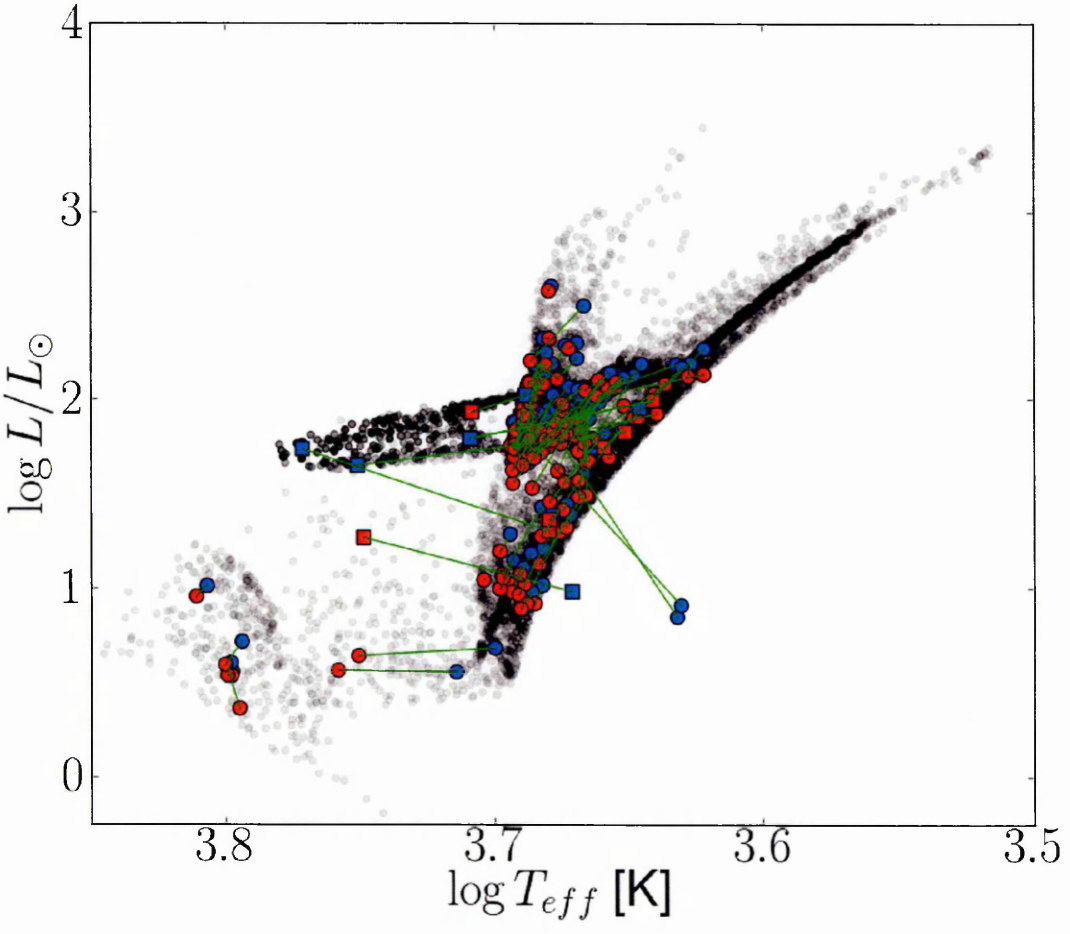



Figure 5.2: A Hertzsprung-Russell Diagram of the detected asteroseismic signals. In grey are the detections from the single stars and the DERBs (which in this instance act as single stars). Coloured dots represent where we see detections in both stars (for LC and SC modes). Green lines connect the individual components of the binary. Blue dots are the initially more massive component while red dots are the initial secondary component. Circles are thin disc systems while squares are thick disc. This is plotted for $s = 0.0$, $b_{frac} = 0.5$ target selected systems.

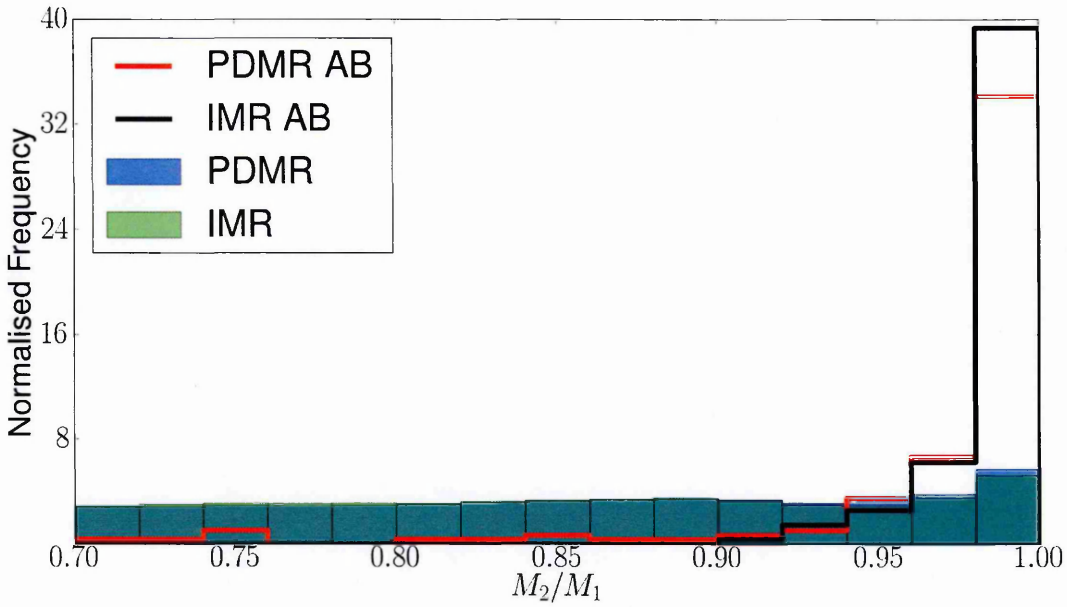



Figure 5.3: The present day mass ratio (PMR) and initial mass ratio distribution (IMR) of target selected asteroseismic binaries (AB) and all target selected binaries for the $s = 0.0$ distribution. On this scale the PDMR and IMR for all binaries are identical. The other IMRDs and IBFDs show similar distributions.

is similar to the IMRD for the asteroseismic binaries. This is also unsurprising as most stars in general do not undergo mass transfer, thus mass loss is limited to winds. As both stars are of similar masses then the winds will be of equal strength thus they will lose mass at similar rates, preserving the mass ratio. The asteroseismic binary PDMR therefore provides a proxy for the IMR of the asteroseismic binaries. The IMR and PDMR distribution of all binaries are flat for the modelled $s = 0$. The distributions are similar as the majority ($\sim 98\%$) of systems are wide detached systems that do not undergone mass transfer.

### 5.3.4 Period distribution

Fig. 5.4 shows the period distribution of the asteroseismic and non-asteroseismic binaries for $s = 0$. The non-asteroseismic binaries have a log flat period distribution with the drop at long periods due to the maximum initial orbital separation of $10^6 R_\odot$. Anything further separated has expanded its orbit due to mass loss during its evolution. The distribution of the asteroseismic binaries is generally similar to the non-asteroseismic binaries, though they differ at short orbital periods. The de-

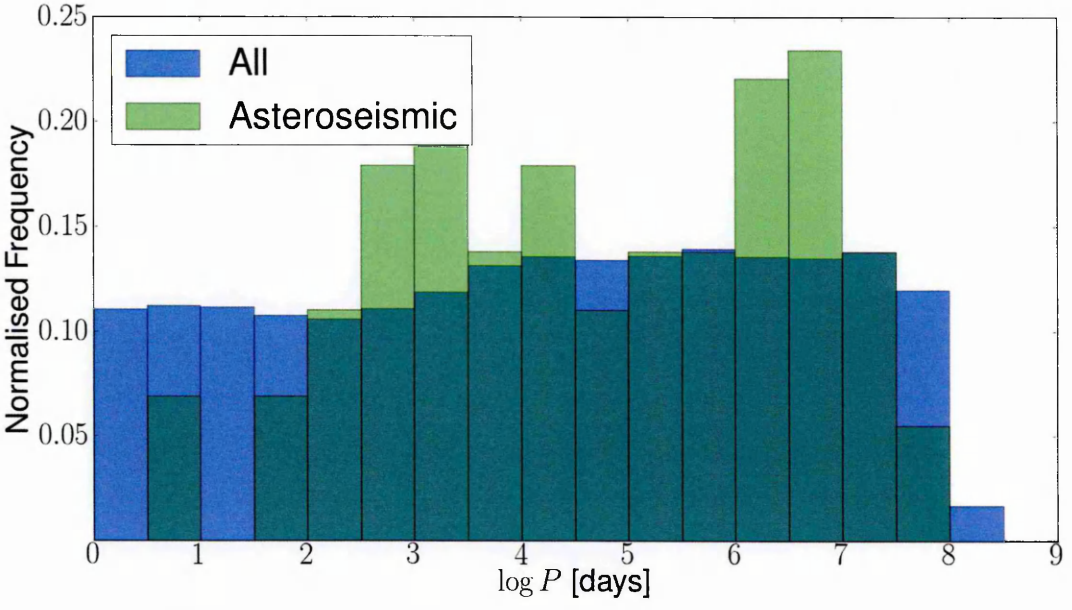



Figure 5.4: The period distribution of asteroseismic binaries (green) and all binaries (blue) for the target selected sample with $s = 0.0$. Other values for the IMRD and the IBFD show similar distributions.

viations are due to the low number of systems per bin. There are two populations of binaries, those with $P_{orb} \gtrsim 30$ days and those with $P_{orb} \lesssim 30$ days. Those with longer periods are constrained by the requirement to have two giants in the same system while not undergoing mass transfer during its evolution. Those with shorter periods have undergone mass transfer, during its evolution from a MS to a HG star, see section 5.4.2. The other IMRD and IBFD distributions show similar results.

The two distributions being the same is expected as there is no explicit dependence on the period in equation 5.9. The only period-dependant constraint is whether a system is wide enough to support two giants in a stable configuration. Thus we can see that asteroseismic binaries provides information on binaries independent of their orbital separation over the whole range of periods, something not seen in other binary detection methods.

### 5.3.5 Asteroseismic signal separation

To detect a asteroseismic binary we must be able to distinguish between the two sets of pulsation frequencies. For example, if we had two exactly identical stars then

$\nu_{max}$ and $\Delta\nu$ would be equal and the peaks in frequency space would be aligned. We would then only be able to see one set asteroseismic signals. Thus in system where $\nu_{max}$ is similar for both stars the extraction of the signal will be more difficult due to the overlapping signals.

Let us assume that the asteroseismic signal has a Gaussian shaped envelope (section 5.2), with a mean of $\nu_{max}$ and a FWHM of $\nu_{max}/2$. Then we can define those binaries where the separation between each $\nu_{max}$ is less than the sum of the two FWHMs to have overlapping asteroseismic signals:

$$|\nu_{max,1} - \nu_{max,2}| < \frac{\nu_{max,1} + \nu_{max,2}}{2} \tag{5.10}$$

where the subscripts denote each star. This defines the minimum separation for the signals to not overlap. For the target selected $s = 0.0$ case all the asteroseismic binaries, for both SC and LC, have overlapping asteroseismic signals.

We can also use a more stringent constraint such that:

$$|\nu_{max,1} - \nu_{max,2}| < \frac{\min(\nu_{max,1}, \nu_{max,2})}{2} \tag{5.11}$$

So that one of the star's asteroseismic peak will be within the FWHM of the Gaussian envelope of the second star. The number of target selected, $s = 0.0$, systems with the more stringent constraint is 57% for the LC detections and 85% for the SC detections. Thus the number of *recoverable* asteroseismic binaries may be much lower than predicted in table 5.1.

## 5.4 Astrophysical contamination

I now consider various types of systems that complicate the analysis of the asteroseismic binaries. These consist of binaries where we can resolve the individual stars, systems which have or are undergoing mass transfer, and finally blended single stars with detectable asteroseismic signals.

### 5.4.1 Resolved binaries

As the asteroseismic binaries can be at any period, some may have wide enough orbital separations that they are resolvable as individual stars. This would complicate the analysis as any loss of flux (for instance part of the star's PSF moving out of the optimal aperture via jitter) would disproportionately affect one star over the other. They would also complicate the analysis of the optimal aperture and the target list as the binaries can no longer be treated as point sources, something that was not considered in section 3.2.1.

From the period distribution and the distance distribution of the synthetic asteroseismic catalogue, we can compute the projected sky separation of the binaries. We find that ~ 2.5% of binaries, have projected sky separations wider than $4''$, the size of one Kepler pixel. This is the same for both asteroseismic and non-asteroseismic binaries. Therefore the numbers of resolvable asteroseismic binaries is small, of order 1-2 systems, and can therefore be ignored.

### 5.4.2 Mass transferring systems

One of the advantages of detecting binaries via their asteroseismic signal is that both objects must be similar luminosities, which for almost all systems implies similar masses, thus we can derive the IMRD from them. However systems that have or are undergoing mass transfer complicate this. Mass being transferred can introduce extra noise, not considered in equation 5.9. The mass transfer process can also drive systems into unexpected regions of the HRD where we may not expect that the relations used in equation 5.9 would still be valid. Also a system that has undergone or is undergoing mass transfer implies that it has a comparatively short orbital period. This means we need to consider extra noise sources, like tidally induced pulsations (Thompson et al., 2012) and stars no longer being spherical (Springer & Shaviv, 2013).

Tables 5.3 & 5.4 show the numbers of predicted asteroseismic binaries that underwent or are undergoing mass transfer (assuming that equation 5.9 holds). We can

| | | Number of detections | | | |
|---|---|---|---|---|---|
| $s$ | $b_{frac}$ | Both (LC) | Either (LC) | Both (SC) | Either (SC) |
| -0.5 | 0.5 | 2 | 163 | 0 | 25 |
| 0.0 | 0.5 | 2 | 233 | 0 | 24 |
| 0.0 | $\log M_1$ | 2 | 312 | 0 | 33 |
| 0.5 | 0.5 | 0 | 301 | 0 | 21 |
| 1.0 | 0.5 | 3 | 316 | 0 | 37 |

Table 5.3: Similar to table 5.1 except now for systems that non-asteroseismic binaries, or are undergoing mass transfer, and are in the target list

| | | Number of detections | | | |
|---|---|---|---|---|---|
| $s$ | $b_{frac}$ | Both (LC) | Either (LC) | Both (SC) | Either (SC) |
| -0.5 | 0.5 | 3 | 236 | 0 | 292 |
| 0.0 | 0.5 | 6 | 315 | 0 | 420 |
| 0.0 | $\log M_1$ | 8 | 403 | 0 | 552 |
| 0.5 | 0.5 | 3 | 411 | 0 | 447 |
| 1.0 | 0.5 | 8 | 403 | 1 | 562 |

Table 5.4: Similar to table 5.1 except now for all systems in the Kepler field with $K_p \leq 16$ that non-asteroseismic binaries, or are undergoing mass transfer.

see that the numbers are small compared to tables 5.1 & 5.2 for systems where we detect both asteroseismic signals. For the cases of single detections the LC results are also small compared to tables 5.1 & 5.2. For the non target selected SC systems the number of mass transfer systems is comparable to the total number detected. This may explain why table 5.2 has more detections than Chaplin et al. (2014); if the mass transfer process significantly suppresses the asteroseismic signal or otherwise increases the noise such that a signal is undetectable. Then the total number of synthetic detections decreases to values less than the number of actual detections in Chaplin et al. (2014).

All the LC asteroseismic binaries that have had mass transfer through case A and then case B processes. The primary star, during its main sequence lifetime, transfers mass on a nuclear time scale, and continues transferring mass during its evolution into a giant on the same time scale (Case A). As the star expands due to its evolution into a giant, it would normally form a common envelope with its companion. However because of the mass transfer (Case B) the orbit period increases, such that the system avoids forming a common envelope.

### 5.4.3 Blended systems

The final set of contaminants considered here are blended systems. These are two singles stars (or DERBs) that are within 1 Kepler pixel of each other, with a detectable asteroseismic signal from both stars predicted. I search the synthetic catalogue for all singles and DERBs that are within $4''$ of each other, then apply the asteroseismic detection code, to determine their number. I alter the $B_k$ factor in equations 5.4 & 5.6, to be equal to the apparent luminosity fraction of each star, as opposed to the intrinsic luminosity fraction.

Table 5.5 shows the predicted number of blends, in Kepler equivalent counts, for the different IMRDs and IBFDs considered here. There is a trend for increasing number of blends as $s$ increases, though $s = -0.5$ is an outlier. The $s = -0.5$ case has a large number of blends due to the larger fraction of DERBs (table 4.1).

As $s$ increases the population of DERBs shifts towards the primary having a lower mass, such that $q \rightarrow 1$. The lower mass implies these systems live longer thus we see an increase in the fraction of GB and CHeB primaries in DERBs as $s \rightarrow 1$, which are the preferential systems to see an asteroseismic signal in.

Also significant is the number of blends containing a thick disc star. Overall the thick disc makes up $\sim 10\%$ of systems, however in the case of blended systems thick disc stars contributes $\sim 50\%$ of the blends. This is likely due to the large CHeB component in BISEPS, as seen in fig 5.2. This in turn is due to the single metallicity used for the thick disc. Decreasing the metallicity moves the bulge towards lower temperatures and decreases the total number of expected CHeB stars.

Comparison with table 5.1 shows that the number of blends with a target selected system is small for all values of $s$. The non target selected systems also show a similar trend. Note this work includes DERBs and the new normalisations scheme (equation 4.5). These where not present in the study of Miglio et al. (2014), which is why the results preferred here are slightly larger.

| | | Number of detections | | | |
|---|---|---|---|---|---|
| $s$ | $b_{frac}$ | Target All | Target Thick | Non-target All | Non-target Thick |
| -0.5 | 0.5 | 20 | 13 | 45 | 31 |
| 0.0 | 0.5 | 4 | 2 | 10 | 5 |
| 0.0 | $\log M_1$ | 4 | 2 | 10 | 5 |
| 0.5 | 0.5 | 8 | 4 | 20 | 10 |
| 1.0 | 0.5 | 11 | 8 | 26 | 19 |

Table 5.5: The Kepler equivalent number of blends predicted for different IMRD and the IBFD. Blends are where two singles stars (or DERBs) are within 1 Kepler pixel (4″) and the combined asteroseismic signal is detected using equation 5.9. Target blends are where at least one star is in the target list, while non target contains all stars with $K_p \leq 16$. The contribution due to one of the stars in the blend being a thick disc object is also shown. These are all LC detections.

## 5.5 Asteroseismic eclipsing binaries

I now cross reference the synthetic systems in the asteroseismic and EB catalogues. These are systems that have both detectable asteroseismic signals and at least 3 detectable eclipses (see section 4.3). It is assumed that the presence of the eclipses does not alter the detectability of the asteroseismic signal.

In table 5.6 we can see the number of asteroseismic EBs as a function of the IMRD and IBFD. There are no predicted systems with detectable asteroseismic signal in both stars and eclipses for SC systems due to the low number of asteroseismic binaries detected in SC mode. Only a few LC asteroseismic binaries have detectable eclipses, primarily due to the requirement that the orbital period must be less than 40 days to be a detected EB (section 4.3). For systems with only one detectable asteroseismic signal there should be of order 50 systems, with a slight dependence on $s$.

Gaulme et al. (2013) found 70 systems that exist in both of Kepler's RG and EB catalogues, with 47 of these objects showing detectable oscillations in a single star. However, only 13 could be confirmed to be an EB with a detectable asteroseismic signal. With 12 of them being GB+MS systems and one being a GB+GB system, where there is a single detected asteroseismic signal. The remaining 34 are systems where a giant is blended with the EB, either as a background object or a third body in a triple system. The Gaulme et al. (2013) sample should be compared with the LC

category with a single detectable asteroseismic signal, which contain systems with a detectable asteroseismic GB plus any companion and detectable eclipses. Thus the model predicts there should be between 35 and 60 asteroseismic EBs, with $P_{orb} < 40$ days. However in the Gaulme et al. (2013) only two asteroseismic EBs have $P_{orb} < 40$ days. Thus it appears as the model is over estimating the number of asteroseismic EBs. As well, the period distribution of EBs in section 4.4.3 places too high a weight on short period systems ($P_{orb} < 10$ days), where it is more difficult to have a giant containing system. Thus we would expect the estimates in table 5.6 to be an underestimation of the total number of systems, if we increased the number of longer period systems.

This suggests that either Gaulme et al. (2013) has missed a large number of short period systems, possibly if the system was not classified as a giant based on the stellar classification program (SCP) parameters. Or that we need a way to dampen the asteroseismic signal in the synthetic EBs such that we reduce the number of detectable asteroseismic EBs. Gaulme et al. (2014) proposes just such a mechanism, for systems with $0.16 < (R_1 + R_2)/a < 0.24$, where this equates to $P_{orb} < 41$ days. For 9 of the systems in the Gaulme et al. (2013) sample, which all had $P_{orb} < 110$ days, quasi-periodic modulations were detected. These were attributed to stellar activity on the giant star, as the giant contributes the bulk of the luminosity in the system. The proposed mechanism is that these systems that have had their spin-orbit synchronised, which requires the giant to be spun up. This then creates a dynamo mechanism which generates a magnetic field. This field could dampen the $p$ mode oscillations, as seen in magnetically active MS objects (Chaplin et al., 2011a).

## 5.6 Discussion

I have shown using a model for the detectability of the asteroseismic signal generated by two stars in a binary, that it is possible to detect a large sample of asteroseismic binaries in the Kepler field. I have also shown there is a clear dependence on the IMRD and a weaker dependence on the IBFD, such that the number of detected

| | | Number of detections | | | |
|---|---|---|---|---|---|
| $s$ | $b_{frac}$ | Both (LC) | Both (SC) | Either (LC) | Either (SC) |
| -0.5 | 0.5 | $1.0 \pm 1.0$ | $0.0 \pm 0.0$ | $34.0 \pm 6.0$ | $10.0 \pm 3.0$ |
| 0.0 | 0.5 | $2.0 \pm 1.0$ | $0.0 \pm 0.0$ | $44.0 \pm 8.0$ | $11.0 \pm 3.0$ |
| 0.0 | $\log M_1$ | $2.0 \pm 1.0$ | $0.0 \pm 0.0$ | $57.0 \pm 5.0$ | $10.0 \pm 2.0$ |
| 0.5 | 0.5 | $1.0 \pm 1.0$ | $0.0 \pm 0.0$ | $60.0 \pm 8.0$ | $13.0 \pm 4.0$ |
| 1.0 | 0.5 | $3.0 \pm 2.0$ | $0.0 \pm 0.0$ | $59.0 \pm 6.0$ | $15.0 \pm 4.0$ |

Table 5.6: The number of systems that are in both the synthetic asteroseismic and synthetic EB catalogues. Results are the average of 10 runs of the EB pipeline, with uncertainties quoted as 1 standard deviation. Number counts are split between systems with both stars having detectable asteroseismic signal or either (but not both) having a detectable asteroseismic signal.

objects can provide constraints on both quantities. I have also shown the size of possible contaminants in the asteroseismic binary sample, and that of most of the other asteroseismic populations, is a small quantity. The only exception is due to MS+MS asteroseismic binaries that have/or are undergoing mass transfer. Finally I considered the number of detectable asteroseismic EBs and showed how this number is inconsistent with the measured number of asteroseismic EBs. I now consider the various assumptions made in this work and how they can be improved in future work.

### 5.6.1 Dependence on Galactic parameters

As seen in table 5.5 the thick disc contributes significantly more blended systems than would be expected based on its total contribution to the number of stars. The thick disc only makes up 10% of the total population of synthetic stars, yet accounts for 50% of the blends. In terms of detectable asteroseismic binaries the thick disc makes up 20% of the population.

Setting the thick disc arbitrarily to $Z = 0.008$, the midpoint in log space between the thin and thick disc metallicity values originally adopted, increases the number of asteroseismic LC detections by 25%, but decreases the total number of thick disc objects by 10%. With a higher $Z$ value for the thick disc, the CHeB branch, seen as

the bulge in fig. 5.2, shifts towards cooler temperatures and also decreases its spread in temperature. The minimum mass of a primary star in a binary to be in the CHeB phase for the higher $Z$ value is $0.92M_{\odot}$, while for the lower $Z$ value it is $0.87M_{\odot}$. This is because the higher the $Z$ value the slower the evolution of the star. Thus the very lowest mass systems in the higher $Z$ disc have had insufficient time to evolve onto the CHeB phase, which reduces the total number of systems. Due to the slower evolution the systems that are CHeB will be in that phase for longer, increasing the chances of detecting the system and thus subsequently detecting the asteroseismic signal. The masses of the CHeB are also larger which increases the $P_{tot}$ (equation 5.4) via its overall $M^{0.5}$ dependence.

Clearly selecting only a single value for the metallicity for the thick disc (though this also applies to the thin disc as well) alters the predicted number counts significantly. Kordopatis et al. (2013a) found, based on ~ 132,000 stars from the *RAVE* survey, a thick disc average value of [M/H]=-0.45, which for a solar $\alpha$-abundance equates to a $Z = 0.007$. Thus the true thick disc number counts should be somewhere in between the $Z = 0.0033$ (our default value) and $Z = 0.008$ (higher $Z$ value tested here) counts. Kordopatis et al. (2013a) found an average [M/H]=-0.09 for thin disc systems, which for a solar $\alpha$-abundance equates to a $Z = 0.016$. Thus we would expect the number of thin disc detections to decrease, as the average $Z$ value is less than that considered here. Therefore the results predicted here are dependant on the metallicity value in the discs and a full metallicity distribution is needed to be able to refine the estimates of the total number counts. However, I would expect the refined estimate to be of the same order of magnitude of that presented here.

Setting the thick disc scale height to be 10% smaller, as preferred in Jurić et al. (2008) increases the number of target selected thick disc objects by 5%, which is of order the random sampling noise. The number of detectable asteroseismic binaries decrease by ~ 2%, which is smaller than the random sampling noise. This suggests that the results presented here are insensitive to the small changes in the disc structure as proposed in the literature.

There is also the possibility that the initial distribution function (IDF), which I have assumed are the same for all binaries, may be different for wide binaries. As the asteroseismic binaries are insensitive to the period, then wide binaries make up the predominant fraction of systems. Models of star formation have shown that close in binaries form from interactions with other stars during their formation (Bate et al., 2003) and not from fragmentation as assumed for wider binaries. Observational, as well, White & Ghez (2001) has shown that in young binary star systems the mass ratio distribution is higher for short period binaries, which they attribute to formation processes. Thus if the mass ratio is lower for wide binaries then the luminosity ratio will be more extreme and there will be less asteroseismic binaries overall. This may explain some of the discrepancy between the model counts and the number of detected systems. Once a sufficient number of these systems are discovered and the period of the system determined in some manner, then we will able to test the formation theories for binaries over range of periods.

### 5.6.2 Noise model

The instrumental noise expressed in equation 5.8 and the RMS value of $\sigma$ are both derived from an empirical fit in Gilliland et al. (2010b). This fit was applied to the lower envelope of the noise measured for 512 SC stars measured during Q1 observations. Thus this noise unlikely to be representative of the expected noise for all stars (see section 4.3.1). For giants the instrumental noise, given by the lower envelope in fig. 4.4, is much smaller than the granulation and asteroseismic variability terms, thus this term can be ignored. However in MS stars where the overall noise budget is smaller the instrumental noise becomes the dominate noise term. Thus for these systems it would be better to use a full noise model, as used in section 3.2.2, to derive the instrumental noise.

Another concern is magnetic activity, which has been shown to suppress the asteroseismic signal in MS stars (Chaplin et al., 2011a). The magnetic fields act as a restoring force, reducing the strength of the oscillations. This suppression is largest

for the MS objects, thus reducing the number of SC detections. It has also been postulated to suppress the asteroseismic signal in close binaries with a tidally locked asteroseismic GB star (Gaulme et al., 2014). Overall we would therefore expect the results presented here to be an upper limit on the number of detectable asteroseismic binaries and asteroseismic EBs.

## 5.7 Conclusions

I have created a model for the detection of the asteroseismic signal in single and binary stars. For the binary stars I consider how the flux from the companion can weaken the detected asteroseismic signal and how the signals can overlap making detection harder. However I have shown that it is still possible to detect both asteroseismic signals in a binary, with a large sample. These objects will predominately have near-equal masses, and I have shown how varying the IMRD and IBFD alters the number of these systems that are detected.

Comparing the results for detection of an asteroseismic signal in one of the two component stars to the Kepler asteroseismic catalogue. The number of predicted LC and SC detections is of order that of the number of actual stars with detectable asteroseismic signals. Considering that the synthetic systems only contain binaries and not single stars, it would appear that the synthetic sample has over estimated the total number counts. This may be due to unconsidered noise terms in short period binaries or the effect of mass transfer reducing the detectability of an asteroseismic signal.

The number of asteroseismic EBs is also inconsistent with that of the published sample of asteroseismic EBs. This may also be due to unconsidered noise terms in short period binaries or the suppression of the asteroseismic signal in short period systems.

These asteroseismic binaries and EB allows us to calibrate the stellar parameters derived from the signal with those derived from radial velocities and eclipses. These systems also allow us to detect and probe the binary parameters over the whole of

the binary period range, something not seen in other detection methods.

# Chapter 6

# Conclusions & Further work

In this chapter I present a summary of the work I have done in this thesis. I then present the main results and conclusions of this work. Finally, I briefly describe two areas of further work that flow naturally from the research presented in this thesis.

## 6.1 Conclusions

In this work I have successfully built a synthetic population of single and binary stars and applied the Kepler specific selection biases to this population. This has allowed me to quantify the bias introduced by the Kepler target selection process and the offsets in the parameter estimation of the the stellar classification program (SCP). Using this synthetic population I have generated a model for the sample of eclipsing binaries in the Kepler field. I have shown that it is possible to distinguish between different initial mass ratio distributions (IMRDs), but not initial binary fraction distributions (IBFDs). At this time I can not say which is the preferred choice due to systematic differences between the model and the data. I have then applied a model for the detectability of a new class of binaries, asteroseismic binaries, and showed how the number of detections can be used to distinguish between different IMRDs.

The SCP was shown to cause a systematic offset in the estimated parameters of the stars in the Kepler input catalogue (KIC). The median values for dwarfs is $\sim \Delta T_{eff} = 500$ K and $\sim \Delta \log g = -0.2$ dex and for giants this is $\sim \Delta T_{eff} = 50$ K and

$\sim \Delta \log g = -0.5$ dex. However this offset varies across the parameter space. I also found that this effect was similar between both primary stars in binaries and single stars. I predict that there is also a population of giants that have been misclassified as dwarfs. The size of this offset was also found to be consistent with other authors, after correcting for their selection effects. Kepler's target selection process was shown to increase the fraction of dwarfs from 75% to 80% and to decrease the fraction of giants from 25% to 20%. The selection effects and SCP offsets between single stars and binaries show no significant differences.

Comparison of my synthetic sample of eclipsing binaries model with the the Kepler eclipsing binary catalogue shows that further work is needed to understand the disagreements. The period distributions can not be reconciled at this time, which I believe is due to the limited noise model used. Further work is needed to introduce new noise terms, such as accretion, and to modify the existing terms, such as the asteroseismic in short period giant systems. Comparison of the temperature ratio and fractional radii distribution show that the Kepler eclipsing binary (EB) pipeline introduces systematics into the measured distributions. Comparing my model with the Kepler eclipsing binary catalogue shows disagreements that I attribute to differences in the training set used. However, different IMRD distributions show a detectable difference in the EBAI-derived distributions. This therefore presents a way to constrain the IMRD in future work. I could not show a significant difference between different IBFDs from the EBs alone, though the number of distributions used was limited.

Predictions for the number of detectable asteroseismic binaries show that there should be of order 100 systems in the Kepler field, though this number is dependant on the IMRD and IBFD. These systems are predominately detected in the long cadence (LC) data, with a core helium burning (CHeB) star as a component. Possible sources of contamination have been considered and I have shown that their numbers are negligible. I also combined the synthetic eclipsing binaries with the synthetic asteroseismic binaries and find a factor of 25 more than have been detected,

which may hint at new physics in short period giant-containing binaries.

The primary limitation of my model currently is the lack of a realistic metallicity distribution. This showed itself in Chapter 3 by forcing the introduction of colour corrections to reproduce the results of the SCP . The use of a single metallicity explains the large extent of the old disk CHeB clump in Chapter 5, not seen in other population models (Miglio et al., 2014). This is due to the metallicity determining the lifetime of the CHeB stars as well as their location and extent in the Hertzsprung-Russell diagram (HRD). I showed that it does have a statistically significant effect on the numbers of asteroseismic binaries.

## 6.2 BiSEPSv3.0

### 6.2.1 Introduction

As noted in section 3.4 one of the main shortcomings in my model is the lack of a realistic metallicity distribution. We also saw in Chapter 4 that the model is also lacking an eccentricity distribution. Both sets of distributions could be introduced into the model in its current form by suitable binning in initial parameter space and then weighting the result by a probability distribution function. However this would be cost prohibitive with current computing power. Currently using 50 bins in the primary and secondary mass, 300 in orbital separation and ~ 100 bins in time generates over 10 million binary configurations for a single metallicity. Adding only 10 bins in metallicity and 10 in eccentricity would bring this value in excess of 1 billion. This is more data than it is feasible to process, it is also exceedingly wasteful. Only a small sub-sample of the most probable systems from the data set are actually used, many systems are either too short lived, too faint or too improbable to form to ever be sub-sampled into the output distribution. Thus a new method is needed to generate only the systems needed for the sub-sampled $\Gamma$ distribution.

The idea is this: instead of evolving all possible systems and then computing which are most likely, we generate the most likely systems to form then evolve only

those systems. This has several benefits. First, the old method's runtime and storage requirements grows linearly with the number of initial distribution functions as well as the number of distinct disc components in the Galaxy. The new method would only scale as a function of the number of systems regardless of the complexity of the underlying Galactic distribution functions. Secondly, by also reducing the number of systems that need to be evolved with the stellar evolution codes, it would also become possible to replace the evolution code with a newer more advanced (possibly slower) code, for instance MESA (Paxton et al., 2011).

### 6.2.2 Method

I now describe the computational method in broad steps. I start with defining a disc in the Galaxy; this is not just a density profile, but represents a complete set of probability distribution functions that represent star formation in a region. First I compute the total number of stars formed in the field of view. I integrate the Galactic density profile over the field of view, possibly with CCD sub structure (see section 2.1.3), between 0 and 50 kpc. This acts to encompass the whole Galaxy along a line of sight. I then integrate the time distribution function, which is the relative local star formation rate in the disc. This describes the relative contribution this disk makes to the star formation rate, thus it would be possible to have multiple disks forming stars at the same time in the same region. This is then scaled by the actual star formation rate to derive the total number of stars that have ever formed in the field of view.

I place stars randomly via rejection sampling according to the distribution function. The Galactic spatial and star formation distributions are then sampled to place each star in a 4D space-time location. Then the initial mass function (IMF) is sampled to derive the primary mass and then using the IBFD the systems status as a binary or a single star is determined. If the system is a binary then I sample from the IMRD, initial orbital separation distribution (IOSD) and initial eccentricity distribution (IED). At this point the metallicity and extinction value, $A_v$ is computed

for both single stars and binaries.

Each system now has a set of initial parameters that describe everything about it, from what properties it had when it formed and when and where it formed. This can then be fed into the single star evolution (SSE) and binary star evolution (BSE) codes to derive the system parameters at the current epoch. The current epoch parameters are then used to give the absolute and apparent magnitude of the system. The extinction corrected apparent magnitude determines whether the system is observable with the chosen observing strategy.

The disadvantage of this method is requirement to evolve every star that has ever formed, even if the system dies before the current epoch or is too faint to be visible. To reduce the number of systems evolved several strategies can be used. Firstly I consider binning the initial parameter space, then I consider using a low resolution population model to remove systems which would never be visible.

The first thing to realise is that it is not necessary to evolve every star separately. Although improbable that any two stars will have the same initial parameters, two stars with similar initial parameters will have similar evolutionary sequences. Therefore I bin the output from the selection of the stellar parameters (mass, metallicity etc) and evolve only one representative system, where the parameters are randomly selected inside the bins. For each star in a bin, information is saved only at the point in the evolution corresponding to its age. Testing on a single Kepler CCD field there were $\sim 65,000$ single stars in the mass bin $M = 0.10 - 0.11 M_{\odot}$ which compresses to just one star that is then evolved. This works well for the single stars with only the mass and metallicity to consider, but is less successful for the binaries that have multiple distributions (primary and secondary mass, metallicity, eccentricity and orbital separation).

Secondly I can filter systems based on whether their luminosity would be sufficient to place the system inside the detection limits. Using BISEPS I generate a coarse grid of evolution tracks distributed over the input parameter space. I then fit a set of polynomials to the luminosity and age (as not all systems will survive to

the present epoch) as a function of the input parameters. The polynomials provide a fast and reliable estimate on whether a given system would be visible today given its initial parameters and its age. If the system would not be visible or would not survive till the present epoch then it can be discarded before performing the full stellar evolution. This removes many of the faint distant objects, as well as high mass old stars, which would not survive to the present epoch, from the population model.

The binaries can not be modelled in exactly the same way due to mass transferring systems. As a post mass transfer binary depends strongly on the input parameters, we can not use a coarse grid approach. Thus only long period systems which we can guarantee will not interact can be approximated using this approach. Short period binaries may need to simply all be modelled and filtered afterwards, lest we discard systems that become visible during or after mass transfer.

Higher order systems (triples, quadruples) can also be taken into account with this new scheme. So long as triple star systems can be described as a binary and a distant non-interacting third body. We can include the third star as an extra set of distribution functions. First we formulate a formation probability of a triple star system; Raghavan et al. (2010) found ~ 10% of systems are triples. Then we derive the binary parameters as previously mentioned and then derive those of the third body, treating it as a single star. It may have the same IMF as single stars or we may need to use an alternate. We can then evolve the binary and single star separately. Once both stars have been evolved to the target age we can compute the system's total luminosity and determine if its detected. This is not a comprehensive model of triple stars, as we are ignoring interacting ones, but it would provide a handle on EBs with a third body diluting their eclipses depths as well as the amount of extra noise in a light curve due to the extra star.

Removing the necessity to integrate the $\Gamma$ distribution (see section 2.1.2) over each disk, allows the inclusion of a more complex disk structure, such as the central bulge and the halo, or sub-populations of the thin disc with different characteristics. Parameters like the IMF, star formation rate (SFR) and metallicity could now

be modelled as a function of location in the Galaxy as the recent literature suggests (Robin et al., 2003; Girardi et al., 2005; Sharma et al., 2011). This would then allows the model to include a realistic metallicity distribution and allow the model to adhere closer to the Galaxy.

### 6.2.3 MESA

By limiting the number of systems evolved I could also replace the SSE and BSE codes with a detailed cutting edge stellar structure and evolution code. MESA is an open source 1D stellar evolution code[1] with an active user community. It solves the full coupled stellar structure equations, taking into account the nuclear reaction rates, EOS opacities, convection and diffusion. Using the MESA code in the population synthesis work would allow the integration of the newest models of the physics of stars, for instance the treatment of convection (Paxton et al., 2013). It would also allow the study of new observables not possible with SSE, such as the elemental abundances, and also allow generating full asteroseismic signals (from the packaged ADIPLS and GYRE codes).

There is some work however still to be done to MESA itself before it would be suitable for inclusion. Binaries have only recently been introduced into the code and would require further testing. Performance wise, MESA is much slower than SSE, thus several strategies are needed to use MESA in population synthesis work. The MESA code itself could be tuned for performance. The input parameters chosen for the stellar models can then also be tuned for different types of stars. As an example stars with $M < 0.8M_{\odot}$ undergo little fusion so the fusion reactions for high $Z$ elements can be ignored. The exact boundaries would need to be tested for where the limit of each approximation can be made. Finally we can consider reducing the initial parameter space further, by increasing the initial parameter bin size such that even fewer systems have to be evolved.

[1] `http://mesa.sourceforge.net`

## 6.3 False positives in transit surveys

I now describe an extension to the model of the eclipsing binaries presented in this work. The main source of uncertainty in planet finding transit surveys, like the Kepler mission, is determining whether the transit signal is from a planet or from an eclipsing binary with a shallow eclipse depth (Morton & Johnson, 2011). Generally we can assume that anything with a transit depth $(\Delta F/F) \gtrsim 1\%$ is not a planet and must be due to a star. But this does not imply that transiting systems with $\Delta F/F \lesssim 1\%$ are due to a planet. Thus we can classify binary systems as false positives if they have $\Delta F/F \lesssim 1\%$, that absent other information like timing analysis or masses from radial velocities would be considered a planet. There are two ways for an EB to have shallow eclipse; either it is intrinsically shallow or the eclipse is blended with another source e.g., Colón et al. (2012).

As most of Keplers' targets are too faint for follow up work (Batalha et al., 2010), the only way to derive the statistics of planets is to take into account the expected rate of false positives. Then we can determine the bias these systems introduce into the results. This bias may not be a constant, but may have a period dependency and/or a dependency on the measured eclipse depth as we probe different populations of false positive systems.

Claimed rates of false positives seen by Kepler have ranged from < 10% (Morton & Johnson, 2011), 10-20% (Fressin et al., 2013) to 35-40% (Santerne et al., 2012). The apparent discrepancy is due to differences in the modelling used, where each author has looked at different types of false positives or in different period/eclipse depth ranges. Thus each sample probes a different set of conditions. One way to resolve this is to have a model that can take into account all types of false positives and give the period and eclipse depth distributions as well.

There are many different configurations of false positives to consider, ranging from systems with intrinsically small eclipses to eclipses that appear small due to a background blended star. Systems where the eclipse itself is small include:

- Grazing eclipses, where the star eclipses only a small region on the limb of its

companion (Morton & Johnson, 2011)

- Extreme radii ratio binaries, for instance MS+WD or MS+GB (Farmer & Agol, 2003)
- Eccentric binaries orientated such that only the secondary eclipse is visible (Santerne et al., 2013)

In cases where the EB is blended with a background source the eclipse itself may be deep enough to classify as stellar in origin, but because of the contribution of flux from the third star the eclipse will appear shallower. There are several possible configurations of this:

- EB blended with non-bound background source (Bryson et al., 2013)
- EB blended with a bound companion

As Kepler is aiming to measure the rate of exo-Earths we can also consider false positives where a large transiting planet appears as an Earth sized transit (Fressin et al., 2013):

- Transiting planet blended with a non-bound background star
- Transiting planet blended with the flux from a bound companion

Kepler has also detected a significant number of exoplanets in multi-planet systems. Thus we should also consider the likelihood of having either multiple false positives, such that there are no planets or that there is only one planet but one or more false positives (Lissauer et al., 2014; Rowe et al., 2014).

Currently I could compute the rates for the grazing eclipses, extreme ratio eclipses and blends with a non-bound companion, subject to the refinements in the EB model discussed in Chapter 4. This is because these are simply extensions to the existing EB model and require no new physics to be implemented.

The eccentric binaries would require a full model of the eccentricity distribution and evolution, while the blends with a bound companion would require the triple

star model presented in section 6.2. For the false positives which include a planet we can not directly compute these rates as we have no population model for planets. However, we could combine the measured planet distributions with the binary model presented here to place bounds on these rates (Fressin et al., 2013).

# Acronyms

**AGB** asymptotic giant branch.

**BC** bolometric correction.

**BH** black hole.

**BSE** binary star evolution.

**BiSEPS** Binary and Stellar Evolution Population Synthesis.

**CDPP** combined differential photometric precision.

**CE** common envelope.

**CHeB** core helium burning.

**CTE** charge transfer efficiency.

**EB** eclipsing binary.

**EBAI** Eclipsing Binaries via Artificial Intelligence.

**FFI** full frame image.

**FPG** focal plane geometry.

**GB** giant branch.

**GDC** gravity darkening coefficient.

**HG** hertzsprung gap.

**HRD** Hertzsprung-Russell diagram.

**IBFD** initial binary fraction distribution.

**IDF** initial distribution function.

**IED** initial eccentricity distribution.

**IMF** initial mass function.

**IMRD** initial mass ratio distribution.

**IOSD** initial orbital separation distribution.

**IPD** initial period distribution.

**KIC** Kepler input catalogue.

**LC** long cadence.

**LDC** limb darkening coefficient.

**MS** main sequence.

**NS** neutron star.

**PDMRD** present day mass ratio distribution.

**PDPD** present day period distribution.

**PN** planetary nebula.

**PRF** pixel response function.

**RLOF** roche-lobe overflow.

**RV** radial velocity.

**SC** short cadence.

**SCP** stellar classification program.

**SDSS** Sloan Digital Sky Survey.

**SFR** star formation rate.

**SSE** single star evolution.

**TCE** threshold crossing events.

**TPAGB** thermally pulsating asymptotic giant branch.

**TPS** transit planet search.

**TRO** thermal relaxation oscillations.

**WD** white dwarf.

# Bibliography


Aigrain S., Favata F., Gilmore G., 2004, A&A, 414, 1139

Al-Naimiy H. M., 1978, Ap&SS, 53, 181

Allard F., Hauschildt P. H., Alexander D. R., Tamanai A., Schweitzer A., 2001, ApJ, 556, 357

Almenara J. M. et al., 2012, MNRAS, 420, 3017

Antognini J. M., Shappee B. J., Thompson T. A., Amaro-Seoane P., 2014, MNRAS

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Bahcall J. N., Pinsonneault M. H., Basu S., 2001, ApJ, 555, 990

Bahcall J. N., Soneira R. M., 1980, ApJS, 44, 73

Baraffe I., Chabrier G., Allard F., Hauschildt P. H., 1998, A&A, 337, 403

Batalha N. M. et al., 2011, ApJ, 729, 27

Batalha N. M. et al., 2010, ApJ, 713, L109

Bate M. R., Bonnell I. A., Bromm V., 2003, MNRAS, 339, 577

Bedding T. R. et al., 2010, ApJ, 713, L176

Bedding T. R. et al., 2011, Nature, 471, 608

Belczynski K., Kalogera V., Rasio F. A., Taam R. E., Zezas A., Bulik T., Maccarone T. J., Ivanova N., 2008, ApJS, 174, 223

Boeche C. et al., 2013, A&A, 559, A59

Bond N. A. et al., 2010, ApJ, 716, 1

Boothroyd A. I., Sackmann I.-J., 1988, ApJ, 328, 653

Borucki W. J. et al., 2010, Science, 327, 977

Borucki W. J. et al., 2011a, ApJ, 728, 117

Borucki W. J. et al., 2011b, ApJ, 736, 19

Borucki W. J. et al., 2012, ApJ, 745, 120

Brown T. M., Gilliland R. L., 1994, ARA&A, 32, 37

Brown T. M., Latham D. W., Everett M. E., Esquerdo G. A., 2011, AJ, 142, 112

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Bryson S. T. et al., 2013, PASP, 125, 889

Bryson S. T. et al., 2010a, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7740, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Bryson S. T. et al., 2010b, ApJ, 713, L97

Burke C. J., 2008, ApJ, 679, 1566

Burke C. J. et al., 2014, ApJS, 210, 19

Butkevich A. G., Berdyugin A. V., Teerikorpi P., 2005, MNRAS, 362, 321

Caldwell D. A. et al., 2010, ApJ, 713, L92

Carter J. A. et al., 2011, Science, 331, 562

Castelli F., Gratton R. G., Kurucz R. L., 1997, A&A, 318, 841

Castelli F., Kurucz R. L., 2004, ArXiv Astrophysics e-prints

Chabrier G., 2003, PASP, 115, 763

Chabrier G., Baraffe I., Allard F., Hauschildt P., 2000, ApJ, 542, 464

Chabrier G., Gallardo J., Baraffe I., 2007, A&A, 472, L17

Chaplin W. J. et al., 2014, ApJS, 210, 1

Chaplin W. J. et al., 2011a, ApJ, 732, L5

Chaplin W. J. et al., 2011b, ApJ, 732, 54

Chaplin W. J. et al., 2011c, Science, 332, 213

Chaplin W. J., Miglio A., 2013, ARA&A, 51, 353

Charbonneau D., Brown T. M., Latham D. W., Mayor M., 2000, ApJ, 529, L45

Charpinet S., Van Grootel V., Reese D., Fontaine G., Green E. M., Brassard P., Chayer P., 2008, A&A, 489, 377

Chen B. et al., 2001, ApJ, 553, 184

Chiappini C., Matteucci F., Gratton R., 1997, ApJ, 477, 765

Chiba M., Beers T. C., 2000, AJ, 119, 2843

Claret A., 2004, A&A, 428, 1001

Claret A., Bloemen S., 2011, A&A, 529, A75+

Claret A., Cunha N. C. S., 1997, A&A, 318, 187

Clark B. M., Blake C. H., Knapp G. R., 2012, ApJ, 744, 119

Cochran W. D. et al., 2011, ApJS, 197, 7

Collier Cameron A. et al., 2007, MNRAS, 380, 1230

Colón K. D., Ford E. B., Morehead R. C., 2012, MNRAS, 426, 342

Cox A. N., Pilachowski C. A., 2000, Physics Today, 53

Davis P. J., Siess L., Deschamps R., 2013, A&A, 556, A4

Debosscher J., Blomme J., Aerts C., De Ridder J., 2011, A&A, 529, A89

Dehnen W., Binney J., 1998, MNRAS, 294, 429

Doyle L. R. et al., 2011, Science, 333, 1602

Dressing C. D., Charbonneau D., 2013, ApJ, 767, 95

Drimmel R., Cabrera-Lavers A., López-Corredoira M., 2003, A&A, 409, 205

Du C. et al., 2003, A&A, 407, 541

Duquennoy A., Mayor M., 1991, A&A, 248, 485

Eggleton P., 2006, Evolutionary Processes in Binary and Multiple Stars. Cambridge University Press

Eggleton P. P., 1983, ApJ, 268, 368

Farmer A. J., Agol E., 2003, ApJ, 592, 1151

Farmer R., Kolb U., Norton A. J., 2013, MNRAS, 433, 1133

Feast M. W., Glass I. S., Whitelock P. A., Catchpole R. M., 1989, MNRAS, 241, 375

Flower P. J., 1996, ApJ, 469, 355

Fluks M. A., Plez B., The P. S., de Winter D., Westerlund B. E., Steenman H. C., 1994, A&AS, 105, 311

Freedman W. L. et al., 1994, ApJ, 427, 628

Fressin F. et al., 2013, ApJ, 766, 81

Fressin F. et al., 2011, ApJS, 197, 5

Gaidos E., Mann A. W., 2013, ApJ, 762, 41

Gao S., Just A., Grebel E. K., 2013, A&A, 549, A20

Gaulme P., Jackiewicz J., Appourchaux T., Mosser B., 2014, ApJ, 785, 5

Gaulme P., McKeever J., Rawls M. L., Jackiewicz J., Mosser B., Guzik J. A., 2013, ApJ, 767, 82

Gazzano J.-C., Kordopatis G., Deleuil M., de Laverny P., Recio-Blanco A., Hill V., 2013, A&A, 550, A125

Gilliland R. L. et al., 2010a, PASP, 122, 131

Gilliland R. L. et al., 2011, ApJS, 197, 6

Gilliland R. L. et al., 2010b, ApJ, 713, L160

Gilmore G., Reid N., 1983, MNRAS, 202, 1025

Gilmore G., Wyse R. F. G., 1985, AJ, 90, 2015

Gilmore G., Wyse R. F. G., Jones J. B., 1995, AJ, 109, 1095

Girardi L., Bertelli G., Bressan A., Chiosi C., Groenewegen M. A. T., Marigo P., Salasnich B., Weiss A., 2002, A&A, 391, 195

Girardi L., Bressan A., Bertelli G., Chiosi C., 2000, A&AS, 141, 371

Girardi L. et al., 2008, PASP, 120, 583

Girardi L., Groenewegen M. A. T., Hatziminaoglou E., da Costa L., 2005, A&A, 436, 895

Goodwin S. P., 2013, MNRAS, 430, L6

Hakkila J., Myers J. M., Stidham B. J., Hartmann D. H., 1997, AJ, 114, 2043

Handler G. et al., 2009, ApJ, 698, L56

Haywood M., 2001, MNRAS, 325, 1365

Haywood M., 2008, MNRAS, 388, 1175

Heger A., Fryer C. L., Woosley S. E., Langer N., Hartmann D. H., 2003, ApJ, 591, 288

Heger A., Langer N., Woosley S. E., 2000, ApJ, 528, 368

Heggie D. C., 1975, MNRAS, 173, 729

Hekker S. et al., 2012, A&A, 544, A90

Hilditch R. W., Howarth I. D., Harries T. J., 2005, MNRAS, 357, 304

Holman M. J. et al., 2010, Science, 330, 51

Holmberg J., Nordström B., Andersen J., 2007, A&A, 475, 519

Howard A. W. et al., 2012, ApJS, 201, 15

Huber D. et al., 2013, ApJ, 767, 127

Huber D., Stello D., Bedding T. R., Chaplin W. J., Arentoft T., Quirion P.-O., Kjeldsen H., 2009, Communications in Asteroseismology, 160, 74

Hurley J. R., Pols O. R., Tout C. A., 2000, MNRAS, 315, 543

Hurley J. R., Tout C. A., Pols O. R., 2002, MNRAS, 329, 897

Hut P., 1980, A&A, 92, 167

Hut P., 1981, A&A, 99, 126

Iben, Jr. I., 1967, ARA&A, 5, 571

Iben, Jr. I., Livio M., 1993, PASP, 105, 1373

Iben, Jr. I., Tutukov A. V., 1984, ApJS, 54, 335

Ibukiyama A., Arimoto N., 2002, A&A, 394, 927

Ivanova N., Belczynski K., Kalogera V., Rasio F. A., Taam R. E., 2003, ApJ, 592, 475

Ivezić Ž. et al., 2008, ApJ, 684, 287

Janson M. et al., 2012, ApJ, 754, 44

Jenkins J. M. et al., 2010a, ApJ, 713, L87

Jenkins J. M. et al., 2010b, ApJ, 713, L120

Jenkins J. M., Peters D. J., Murphy D. W., 2004, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 5497, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, S. C. Craig and M. J. Cullum, ed., pp. 202–212

Jurić M. et al., 2008, ApJ, 673, 864

Kennicutt, Jr. R. C., 1984, ApJ, 277, 361

Kjeldsen H., Bedding T. R., 1995, A&A, 293, 87

Kjeldsen H., Bedding T. R., 2011, A&A, 529, L8

Kjeldsen H. et al., 2008, ApJ, 682, 1370

Klinglesmith D. A., Sobieski S., 1970, AJ, 75, 175

Koch D. G. et al., 2010, ApJ, 713, L79

Konstantinova-Antova R. et al., 2010, A&A, 524, A57

Kordopatis G. et al., 2013a, MNRAS, 436, 3231

Kordopatis G. et al., 2013b, A&A, 555, A12

Kozai Y., 1962, AJ, 67, 591

Kroupa P., 2001, MNRAS, 322, 231

Kroupa P., Petr-Gotzens M. G., 2011, A&A, 529, A92

Lattimer J. M., Prakash M., 2001, ApJ, 550, 426

Lebreton Y., 2000, ARA&A, 38, 35

Léger A. et al., 2009, A&A, 506, 287

Lehmann H., Southworth J., Tkachenko A., Pavlovski K., 2013, A&A, 557, A79

Lissauer J. J. et al., 2014, ApJ, 784, 44

Lucy L. B., 1967, ZAp, 65, 89

Lucy L. B., 1968, ApJ, 153, 877

Magic Z., Collet R., Asplund M., Trampedach R., Hayek W., Chiavassa A., Stein R. F., Nordlund Å., 2013, A&A, 557, A26

Mandel K., Agol E., 2002, ApJ, 580, L171

Mann A. W., Gaidos E., Lépine S., Hilton E. J., 2012, ApJ, 753, 90

Matijevič G., Prša A., Orosz J. A., Welsh W. F., Bloemen S., Barclay T., 2012, AJ, 143, 123

Mayor M., Queloz D., 1995, Nature, 378, 355

McQuillan A., Mazeh T., Aigrain S., 2014, ArXiv e-prints

McWilliam A., 1997, ARA&A, 35, 503

Mestel L., 1968, MNRAS, 138, 359

Michel E. et al., 2008, Science, 322, 558

Miglio A., Chaplin W. J., Farmer R., Kolb U., Girardi L., Elsworth Y., Appourchaux T., Handberg R., 2014, ApJ, 784, L3

Miller G. E., Scalo J. M., 1979, ApJS, 41, 513

Morris S. L., 1985, ApJ, 295, 143

Morris S. L., 1999, ApJ, 520, 797

Morton T. D., Johnson J. A., 2011, ApJ, 738, 170

Nataf D. M., Gould A., Pinsonneault M. H., 2012, Acta Astron., 62, 33

Neilson H. R., Langer N., Engle S. G., Guinan E., Izzard R., 2012, ApJ, 760, L18

Nelemans G., Yungelson L. R., Portegies Zwart S. F., Verbunt F., 2001, A&A, 365, 491

Noh H.-R., Scalo J., 1990, ApJ, 352, 605

Nordström B. et al., 2004, A&A, 418, 989

Noyes R. W., Hartmann L. W., Baliunas S. L., Duncan D. K., Vaughan A. H., 1984, ApJ, 279, 763

O'Rourke J., 1995, Computational Geometry in C. Cambridge University Press

Padoan P., Nordlund Å., 2002, ApJ, 576, 870

Palla F., Stahler S. W., 1993, ApJ, 418, 414

Parsons S. G. et al., 2010, MNRAS, 407, 2362

Patterson J., 1981, ApJS, 45, 517

Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, ApJS, 192, 3

Paxton B. et al., 2013, ApJS, 208, 4

Perryman M. A. C. et al., 2001, A&A, 369, 339

Perryman M. A. C. et al., 1997, A&A, 323, L49

Pinsonneault M. H., An D., Molenda-Żakowicz J., Chaplin W. J., Metcalfe T. S., Bruntt H., 2012, ApJS, 199, 30

Pollacco D. L. et al., 2006, PASP, 118, 1407

Popper D. M., Etzel P. B., 1981, AJ, 86, 102

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in FORTRAN: The Art of Scientific Computing, Vol. 2. Cambridge University Press

Pringle J. E., 1981, ARA&A, 19, 137

Prša A. et al., 2011, AJ, 141, 83

Prša A., Guinan E. F., Devinney E. J., DeGeorge M., Bradstreet D. H., Giammarco J. M., Alcock C. R., Engle S. G., 2008, ApJ, 687, 542

Raghavan D. et al., 2010, ApJS, 190, 1

Rappaport S., Deck K., Levine A., Borkovits T., Carter J., El Mellah I., Sanchis-Ojeda R., Kalomeni B., 2013, ApJ, 768, 33

Reid I. N., Gizis J. E., 1997, AJ, 113, 2246

Reid M. J., 1993, ARA&A, 31, 345

Robin A. C., Reylé C., Derrière S., Picaud S., 2003, A&A, 409, 523

Rowe J. F. et al., 2014, ApJ, 784, 45

Rucinski S. M., 1998, AJ, 116, 2998

Salpeter E. E., 1955, ApJ, 121, 161

Santerne A. et al., 2012, A&A, 545, A76

Santerne A., Fressin F., Díaz R. F., Figueira P., Almenara J.-M., Santos N. C., 2013, A&A, 557, A139

Scholz A., Geers V., Clark P., Jayawardhana R., Muzic K., 2013, ApJ, 775, 138

Schou J. et al., 1998, ApJ, 505, 390

Schreiber M. R., Gänsicke B. T., 2003, A&A, 406, 305

Schultheis M., Robin A. C., Reylé C., McCracken H. J., Bertin E., Mellier Y., Le Fèvre O., 2006, A&A, 447, 185

Schuster A., 1903, ApJ, 17, 165

Shapiro S. L., Teukolsky S. A., 1986, Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects. Wiley-Interscience

Sharma S., Bland-Hawthorn J., Johnston K. V., Binney J., 2011, ApJ, 730, 3

Sing D. K., 2010, A&A, 510, A21+

Slawson R. W. et al., 2011, AJ, 142, 160

Southworth J., Maxted P. F. L., Smalley B., 2004, MNRAS, 351, 1277

Springer O. M., Shaviv N. J., 2013, MNRAS, 434, 1869

Stępień K., Gazeas K., 2012, Acta Astron., 62, 153

Steffen J. H. et al., 2011, MNRAS, 417, L31

Stello D. et al., 2013, ApJ, 765, L41

Taylor J. H., Weisberg J. M., 1982, ApJ, 253, 908

Tenenbaum P. et al., 2013, ApJS, 206, 5

Thompson S. E. et al., 2012, ApJ, 753, 86

Torres G., 2010, AJ, 140, 1158

Torres G. et al., 2011, ApJ, 727, 24

Torres G., Ribas I., 2002, ApJ, 567, 1140

Twarog B. A., 1980, ApJ, 242, 242

Uttenthaler S., Schultheis M., Nataf D. M., Robin A. C., Lebzelter T., Chen B., 2012, A&A, 546, A57

Van Cleve J. E., Caldwell D. A., 2009, Kepler Instrument Handbook (KSCI-19033)

van Hamme W., 1993, AJ, 106, 2096

Vanhollebeke E., Groenewegen M. A. T., Girardi L., 2009, A&A, 498, 95

von Zeipel H., 1924, MNRAS, 84, 684

Wade R. A., Rucinski S. M., 1985, A&AS, 60, 471

Weidemann V., 1990, ARA&A, 28, 103

Welsh W. F. et al., 2011, ApJS, 197, 4

White R. J., Ghez A. M., 2001, ApJ, 556, 265

Willems B., Kolb U., 2002, MNRAS, 337, 1004

Willems B., Kolb U., 2004, A&A, 419, 1057

Willems B., Kolb U., Justham S., 2006, MNRAS, 367, 1103

Wilson R. E., 1990, ApJ, 356, 613

Wilson R. E., Devinney E. J., 1971, ApJ, 166, 605

Wood M. A., Oswalt T. D., 1998, ApJ, 497, 870

Woosley S. E., Heger A., Weaver T. A., 2002, Reviews of Modern Physics, 74, 1015

Yakut K., Eggleton P. P., 2005, ApJ, 629, 1055