Buoyancy-driven oscillations in helio- and asteroseismology

by

James Stevenson Kuszlewicz

A Thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

College of Engineering and Physical Sciences

School of Physics and Astronomy

University of Birmingham

April, 2017


PER
ARDUA
AD
ALTA

Abstract

This thesis focuses on the application of asteroseismology to red giants observed with *Kepler* alongside searching for solar g-modes using the Birmingham Solar Oscillations Network (BiSON). In the case of the Sun, solar gravity modes are highly sought after because they can shed light on the inner rotation profile of the Sun. This thesis contains work showing how the low frequency regime of BiSON data has been cleaned enabling the search to be made in BiSON data without instrumental artefacts. Moving onwards along the stars evolution, thanks to space mission such as *Kepler* and CoRoT tens of thousands of red giant stars have been observed allowing huge ensemble investigations. The ability to use high-quality, long datasets as constraints to shorter and noiser datasets has been investigated through fitting the background power of 6000 *Kepler* red giants. Red giants also offer the opportunity to study the inclination angle distribution of stars to confirm that the distribution conforms to the expected isotropy used in many simulations. This can be extended to inferring the obliquity through asteroseismology, as applied to a red-giant, M-dwarf eclipsing binary. This offering a means to probe obliquity distributions in in a different regime to that using traditional spectroscopic techniques.

To my parents and sister.

Acknowledgments

First and foremost, I would like to thank my supervisor Bill Chaplin, for his invaluable advice and enthusiasm for the subject that has helped so much over the past three and a half years. I would also like to thanks Guy Davies for having to put up with all of the interesting debates about work we would have over the course of my PhD.

I would also like to thank Caitlin and Tom especially for having to put up with more than their fair share of moaning but for also making my time during my PhD much more enjoyable. I am also very grateful to Ben, Mat, Ollie, Diego, Hugo, Andrea, Yvonne, Mikkel, Tiago, Nadege, Daniel, Rasmus, Steve and Warrick for all of the helpful discussions and good times during my PhD. Lastly, but by no means least, a special thanks you to my Mum, Dad and Sister, along with the rest of my family, for the unwavering support shown over my time at school and university.

Contents

Abstract **iii**

Dedication **iv**

Acknowledgments **v**

List of figures **xxx**

List of tables **xxxii**

List of Publications **xxxii**

Statement of Collaboration **xxxiii**

1 An Introduction to Helio- and Asteroseismology **1**

1.1 Brief Overview of Stellar Evolution 3

1.2 Solar-like Oscillations 5

1.3 Modes of mixed character 13

1.4 Structure of the Thesis 14

2 BiSON and the Quest for Solar g-modes **16**

2.1 BiSON Data 17

2.2 Window Function 24

2.3 Modelling the BiSON Background Limit Spectrum 29

2.4 Understanding the Cause of Diurnal Peaks 36

2.5 Improved Removal of Diurnal Peaks 42
2.6 Conclusion 58

3 Inference and characterisation of Red Giants from their background power **60**

3.1 Data sets 62
3.2 Data Preparation 63
3.3 Background Model 63
3.4 Fitting Methodology 67
3.5 Fitting Process 69
3.6 Cleaning the data 79
3.7 Non-parametric priors 83
3.8 Exploring the results from the *Kepler* fit 85
3.9 Kepler-as-K2 105
3.10 Can we predict evolutionary state from background parameters? 110
3.11 Conclusions 122

4 Generation of Artificial *Kepler* Red Giant Spectra **124**

4.1 Background Spectrum 126
4.2 Mode Frequencies 129
4.3 Mode Amplitudes 135
4.4 Mode Linewidths 142
4.5 Rotational Splitting 147
4.6 Example Spectra and Comparisons to Real Data 149
4.7 Extension to Doppler Velocity 150

5 Hierarchical Inference in the era of *Kepler* **159**

5.1 Inferring the obliquity distribution with *Kepler* KOIs 161

5.2 How to determine the angle of inclination using asteroseismology 164
5.3 Applications to *Kepler* 166
5.4 Conclusion 180

6 Inferring the Distribution of Stellar Inclination Angles **182**
6.1 Overview of methodology 184
6.2 Mixed Mode Identification 184
6.3 Mode fitting 186
6.4 Hierarchical modelling 189
6.5 Issues with the data 197
6.6 Observed excess at 90° - investigating potential biases 201
6.7 Using artificial data 211
6.8 Conclusion 216

7 KOI-3890 **218**
7.1 Introduction 218
7.2 Observations 220
7.3 Close companions 221
7.4 Simultaneous radial velocity and transit fitting 224
7.5 Global stellar parameters 227
7.6 Determination of obliquity 231
7.7 Conclusions 235

8 Conclusion **237**

Appendix A Appendix **240**
A.1 Data of Finite Extent 240
A.2 Derivation of Single Station Window Function 246

A.3 Simplification of Equation A.16 247
A.4 Sunrise-sunset Equation 248
A.5 Adding "Realism" 250
A.6 Peak detection - the case of single spikes 253
A.7 Derivation of the likelihood function for rebinned data 255
A.8 Gelman-Rubin convergence criterion 255
A.9 Derivation of Mode Visibilities 258
A.10The Hierarchical Process 260
A.11Hierarchical process 260
A.12Mode Detection 268
A.13What does lack of secondary eclipse tell us? 270

References **286**

Listing of figures

1.1 A Hertzsprung-Russell diagram showing a 1 $M_\odot$ and 3 $M_\odot$ evolutionary track compute using MESA (Paxton et al., 2011). The letters on the 1 $M_\odot$ are referred to in the text when needed, the inset plot on the right hand side shows the position of the RGB bump in the 1 $M_\odot$ track. Figure from Hekker & Christensen-Dalsgaard (2016). 6

1.2 An example background-subtracted power spectrum of the red giant branch star KIC 8564976. The frequency of maximum power is shown by the dotted blue line and the arrows show the approximate positions of the radial modes and the large frequency separation $\Delta\nu$. The heavily smoothed power spectrum is shown in red (amplified by a few orders of magnitude to make it visible) in order to show the power excess caused by the oscillations themselves. 9

2.1 Power spectrum of 21 years of BiSON data where the strong diurnal peaks can be seen at low frequency as well as the Gaussian-like envelope of the p-mode region at high frequency. 18

2.2 Power spectrum of a sine wave with frequency ν_s = 5mHz with the BiSON window imposed in black. In red is the power spectrum of same sine wave with the BiSON window imposed but with mean subtracted from the data. The continuum arises from short timescale gaps in the window function which in turn contributes to broadband noise in the power spectrum. 21

2.3 Power spectrum of the BiSON dataset (smoothed with a boxcar of width 1 μHz) having not subtracted the mean (black) and subtracted the mean (red). 22

2.4 The resultant plot following the convolution of the sine wave with added mean and the BiSON window function. The main peak corresponding to the sine wave frequency at 5mHz can be clearly seen. The inset box in the top right hand corner shows the very low frequency structure caused by the convolution of the zeroth bin of the data with the window function. 23

2.5 The difference between subtracting the mean when there is and is not as signal present in the data with a period longer than the length of the time series. The case without the very low frequency signal is shown in black and the case with the very low frequency signal in red. 25

2.6 The right hand planel shows the power spectrum of the window function for 23 years of BiSON data over the range of 1 to 5000μHz. Present in the spectrum are a set of large periodic peaks characteristic of diurnal signatures and their overtones, due to gaps in the time series. The left hand panel shows a zoomed in view of the peaks present at 57.87 and 81.02μHz. 26

2.7 Artificial window function power spectrum simulated including the addition of changing length of day, interruptions and multiple stations. The addition of noise is the most striking influence the interruptions possesses. The left hand plot shows the entire power spectrum of the artificial window function and the left hand plot shows the peaks at 81 μHz and 57.8 μHz. 29

2.8 BiSON limit spectrum constructed using the 23 year BiSON data. The full model is shown in the solid red line, the granulation components are shown in the blue, green and red dotted lines and the sum of the granulation components multiplied by the frequency response (shown in pink) is shown in the second blue dotted line. The white noise and pink noise are the yellow and turquoise dotted lines respectively. 35

2.9 Power spectrum of BiSON data smoothed with a 28 point moving mean filter shown in grey and the limit spectrum convolved with the window function power spectrum is shown in red. 37

2.10 An example of contemporaneous data taken from two different BiSON stations. The title shows the time period over which the data was taken and each coloured trace denotes a different day smoothed with a box car of width 20 bins. The top plot is data taken from the station in Carnarvon and the bottom plot is data taken from the station at Las Campanas. 38

2.11 Examples of the footprint applied to 1 day of SolarFLAG data, each with a different long-term granulation signal added where the underlying trend (caused by the footprint) is shown in red. The top left plot has no long-term signal added, whereas the top right plot contains a granulation signal with a characteristic timescale of 1 day and the bottom plot contains granulation with a characteristic timescale of 30 days. 40

2.12 Examples of the footprint applied to 1 day of SolarFLAG data, each with a different long-term granulation signal added where the underlying trend (caused by the footprint) is shown in red. The top left plot has no long-term signal added, whereas the top right plot contains a granulation signal with a characteristic timescale of 1 day and the bottom plot contains granulation with a characteristic timescale of 30 days. 42

2.13 Examples of 7 days of data taken from the BiSON station at Sutherland before (left hand plot) and after (right hand plot) the PCA has been applied. 46
2.14 Two plots showing the difference in the velocity residuals as a function of time for one day of data taken at Las Campanas caused by masking out different regions of data. The top plot is created using the standard regions where bad data is masked whereas the bottom plot has an additional region at the end of the day masked out where the differential extinction fit was not as good. (Figures courtesy of Steve Hale). 47
2.15 Two examples of power spectra (smoothed with a boxcar of width 1 μHz) constructed from data before (black) and after (red) the PCA has been applied. The right hand plot is for the BiSON station at Mount Wilson, and the left hand plot is for Izaña. 47
2.16 The fraction of days PCA is applied to each BiSON station is shown in the left hand plot, any zero points are a result of data not being taken with that station at that particular time. On the right hand plot is the explained variance ratio for each station. Any time spent at 1 shows that only 1 day of data was present for the 7 days analysed. 48
2.17 A comparison between the various datasets used for solar analyses. The black shows the BiSON data before the PCA is applied, the red shows it afterwards and the green is approximately 15 years of GOLF data. All power spectra have been smoothed by a boxcar of width 1μHz. 50
2.18 Plots comparing the median noise levels (in amplitude) of the individual BiSON stations (before the PCA was applied) and GOLF over a time period of 1992 to 2013. 51

2.19 Plots comparing the median noise levels (in amplitude) of the individual BiSON stations (after the PCA was applied) and GOLF over a time period of 1992 to 2013. 53

2.20 Plots comparing the median noise levels (in amplitude) of the combined BiSON data (before and after the PCA was applied) and GOLF over a time period of 1992 to 2013. The left hand panel shows the 100-200 μHz region and the right hand panel shows the 200-800 μHz region. 53

2.21 PSPS for full BiSON datasets (black) and subset (red). The PSD has been normalised by the highest value of each power spectrum for aesthetic purposes. 55

2.22 PSPS for full BiSON dataset (top left) and subset (top right) as well as the last 16 years of GOLF data. The significance levels according to 90%, 95% and 99%, given by the green, blue and red dashed lines respectively according to $N = 20$ bootstrapped datasets. 56

2.23 Upper limits on g-mode amplitudes as a function of frequency for BiSON (black) and GOLF (red). 58

3.1 Power spectrum of KIC 3533464 shown in black and in the red is the same spectrum smoothed with a boxcar over a region of 1μHz. 63

3.2 Two more examples of red giant power spectrum observed by *Kepler*. The left-hand plot is of a high $\nu_{\rm max}$ star that shows the slight rising component at low frequency commonly associated with "activity", whereas the right-hand plot shows a star with a much lower $\nu_{\rm max}$ that does not show any evidence of "activity". Note the change in scale of the y-axis due to the large difference in $\nu_{\rm max}$ of the two stars. 66

3.3 The initial model created during the initial guess (KIC 10351196) on the left hand side shows the rebinned power spectrum (over 1μHz) in black and the approximate model from the iterative algorithm in red. On the right-hand side is the final fit to the same star with the power spectrum in black, rebinned power spectrum in blue and models constructed from random samplings of the posterior PDFs in green. The $\nu_{\rm max}$ posterior PDF is also inset to show the precision of the fitting process. Also note the difference in scales between the left and right panels, this is purely for demonstrative purposes. 75

3.4 A comparison between using different summary statistics with regards to two different distributions, where the coloured line represents the value of the different quantities used. The red shows the median, the green shows the mean and the blue shows the mode. 80

3.5 A comparison between using different summary statistics with regards to two different distributions, where the coloured point represents the median of the distribution and the coloured line represents the region encompassed within the calculated uncertainties. The red shows the combination of the median and highest posterior density (HDP) region, the blue shows the median and the interquartile range and the green shows the median and the standard deviation. 80

3.6 A poor fit to the power spectrum of KIC 1720241. The histogram of $\nu_{\rm max}$ shows that it has not converged well to a solution and the many spikes in the data can be seen. 82

3.7 The left-hand plot shows amplitude of the first granulation component as a function of $\nu_{\rm max}$ shown before and after ordering the timescales such that the activity component is less than the first component which is less than the second component. The right-hand plot is the same as that on the left but added in red is the first component amplitude having applied both the timescale and amplitude ordering. 83
3.8 Plots of the background fit parameters as a function of $\nu_{\rm max}$ coloured by the natural logarithm of their white noise value. The orange contours show the mixture of Gaussians fitted to the data whilst the background contours show the uninformative priors applied to the data in the background fitting. 86
3.9 Same as Fig 3.8. 87
3.10 The ratio of the second granulation frequency plotted as a function of $\nu_{\rm max}$ is shown on the left panel coloured by evolutionary state (blue is RGB, green is RC and red is SC). The right-hand side plot is the height of the oscillation envelope plotted as a function of $\nu_{\rm max}$ coloured by evolutionary state as well. . 88
3.11 Comparison plots of the background models for those fitted in this work (in red) and those constructed from the scaling relations (green) for three red clump stars KIC 8489832, KIC 8611114 and KIC 10683647. The left column shows the background models and the right column shows the signal-to-noise spectrum. 90
3.12 Same as 3.11 but for 3 RGB stars with similar $\nu_{\rm max}$ values. 91
3.13 Comparison histograms of the two samples of clump stars made at a cut-off of $b_2/\nu_{\rm max} > 1.2$ to be considered part of the "weird" sample. 92
3.14 The width of the oscillation envelope as a function of $\nu_{\rm max}$ for the stars fitted in this chapter. The red line is the scaling relation given in 3.11 which is valid for stars with $\nu_{\rm max} < 100$ μHz. 94

3.15 The fit to the full sample where the red line shows the power law relation with the 1σ calculated from the residuals shown by the red dotted lines. The blue shaded regions show the 1σ and 2σ bounds including the additional modelled variance term. 95
3.16 The same as Fig 3.15 where the left hand panel shows the RGB sample and the right hand panel shows the RC and SC sample. 96
3.17 The residuals of the power law for $\delta\nu_{\rm env}$ (in log-space) as a function of $\nu_{\rm max}$ for the RGB sample coloured by $T_{\rm eff}$ (left-hand plot) and [Fe/H] (right-hand plot). 97
3.18 The residuals of the power law for $\delta\nu_{\rm env}$ (in log-space) as a function of $\nu_{\rm max}$ for the RC and SC sample coloured by $T_{\rm eff}$ (left-hand plot) and [Fe/H] (right-hand plot). 98
3.19 The fit to the full sample where the red line shows the power law relation with the 1σ calculated from the residuals shown by the red dotted lines. The blue shaded regions show the 1σ and 2σ bounds including the additional modelled variance term. 99
3.20 The same as Fig 3.19 where the left hand panel shows the RGB sample and the right hand panel shows the RC and SC sample. 100
3.21 The residuals of the power law for $H_{\rm env}$ (in log-space) as a function of $\nu_{\rm max}$ for the RGB sample coloured by $T_{\rm eff}$ (left-hand plot) and [Fe/H] (right-hand plot). 100
3.22 The residuals of the power law for $H_{\rm env}$ (in log-space) as a function of $\nu_{\rm max}$ for the RC and SC sample coloured by $T_{\rm eff}$ (left-hand plot) and [Fe/H] (right-hand plot). 100
3.23 The predicted $A_{\rm max}$ are shown in blue plotted as a function of $\nu_{\rm max}$ where the original data is shown in black. The RGB sample is on the left-hand side whereas the clump sample is shown on the right-hand plot. 104

3.24 The residuals of the fit to the $A_{\rm max}$ data shown for the RGB sample as a function of $T_{\rm eff}$ on the left-hand plot and [Fe/H] on the right-hand plot. 104
3.25 Same as Fig 3.24 but for the clump sample. 104
3.26 The white noise extracted from the background fitting plotted as a function of *Kepler* magnitude. The red dotted line shows the upper limit derived in Jenkins et al. (2010b) and the red solid line shows the lower limit on the precision also derived in Jenkins et al. (2010b). 106
3.27 A comparison of the uncertainties in $\nu_{\rm max}$ from the *Kepler*data in blue and the *Kepler*-as-K2 data in red. 107
3.28 The fractional uncertainties on $\nu_{\rm max}$ shown as a function of $\nu_{\rm max}$ for the *Kepler* data (left hand panel) and the Kepler-as-K2 sample (right-hand panel). 108
3.29 A comparison of the fractional uncertainties on $\nu_{\rm max}$ shown as a function of $\nu_{\rm max}$. The *Kepler* data is shown in blue, and the Kepler-as-K2 sample is shown in red rescaled by $\sqrt{N}$ to make it comparable with the original *Kepler* data. . . . 109
3.30 The difference between the determined $\nu_{\rm max}$ from the *Kepler*data and *Kepler*-as-K2 data divided by their error (added in quadrature). 109
3.31 A histogram showing the distribution of $\nu_{\rm max}$ values that have been drawn from the modelled prior distributions. 111
3.32 Estimate background profiles for KIC 8564976 on the left and EPIC 201127270 on the right. The data is shown in black with the spectrum rebinned over 1 μHz shown in blue and the estimated model in red. 111
3.33 A comparison of the labels used in the classification task before and after consensus labelling was applied, in both plots blue points are RGB, green are RC and red are SC. The left-hand panel shows the labels from Elsworth et al. (2016), whereas the right-hand panel shows the results of consensus labelling and the final labels used. 114

3.34 The logarithmic loss function plotted as a function of epoch for the training of the classifier. The points indicate the mean of the log-loss at that epoch and the errorbars show the standard deviation. In blue is the training set and the test set is given in red. The loss function flattening out is a good sign that the data is not being overfitted and that the model has been well trained. 116
3.35 Confusion matrices for the classifier when applied to the test data. The left-hand panel shows the unnormalised confusion matrix where the values displayed and the colour bar refer to the number of stars in each category. The right-hand panels shows the number of stars in each category normalised by the number of stars in each class. 118
3.36 Width of the oscillation envelope plotted as function of $\nu_{\rm max}$, coloured by evolutionary state in the same way as previous plots, for the entire sample of stars. The yellow crosses denote stars that have been misclassified in the test set (either as false positives or false negatives). 118
3.37 Feature importance for the features used in the classification task. The F-score is defined as the number of times a feature is used to split the data across all trees, giving a measure of how important each feature is to the fitting process. 119
3.38 Probability distributions corresponding to the probability that the star belongs to a certain class. The distribution for RGB is in blue, RC in red and SC in green. The SC distribution cannot be seen due to the fact that it is so heavily concentrated at zero. 121
3.39 Confusion matrices for the classifier when applied to the test data. The left-hand panel shows the unnormalised confusion matrix where the values displayed and the colour bar refer to the number of stars in each category. The right-hand panels shows the number of stars in each category normalised by the number of stars in each class. 121

3.40 Width of the oscillation envelope plotted as function of $\nu_{\rm max}$ for the *Kepler*-as-K2 sample, coloured by evolutionary state in the same way as previous plots, for the entire sample of stars. The yellow crosses denote stars that have been misclassified in the test set (either as false positives or false negatives). 122

4.1 The background granulation model shown in black and decomposed into its constituent parts normalised to unity to show the effect of sampling. Shown in red is the η^2 function showing the effect of sampling, in blue is the first granulation component, the second component is shown in green and finally the shot noise is shown in yellow. 129

4.2 The left-hand panel shows an echelle plot consisting of the radial mode frequencies generated from the asymptotic expression given in equation 4.13 (green diamonds) and the extracted frequencies from the power spectrum of Kepler-56 (red diamonds). The black dotted line indicates the position of $\nu_{\rm max}$ and the blue dotted line shows the frequencies predicted by the asymptotic expression neglecting curvature. The right-hand panel shows the difference in frequency between successive radial modes using the same colours as the left-hand plot. The gradient seen in some of the frequencies is a clear indication of the presence of curvature. 132

4.3 A linear fit (shown in red) made to the data analysed in Vrard et al. (2016) identified as red giants. 135

4.4 A comparison of the bandpasses for the *Kepler* mission in black and the TESS mission in red as a function of wavelength. Both transmission curves have been normalised such that their maximum value is unity. 136

4.5 Radial mode amplitudes calculated according to equation 4.26 as a function of frequency for a Kepler-56 like star. The black dashed line shows the interpolation function used whilst the red diamonds mark the location of the radial mode frequencies. The blue dashed vertical line shows the location of the *Kepler* long-cadence Nyquist frequency beyond which we do not calculate the oscillation properties. 140

4.6 Mode heights normalised by their relative visibility evaluated at each predicted mixed mode frequency, plotted as a function of frequency for a Kepler-56 like star. The red and blue diamonds show the heights of the radial modes and nominal p-modes respectively and the solid lines show the heights of the $\ell = 1$ mixed modes calculated using equation 4.29 with a linewidth of 0.3 muHz and an observing length of 4 years. 141

4.7 Mode linewidths for the radial modes fitted in Chapter 6 as a function of reduced frequency. The red solid line shows the fit of equation 4.32 to the data and the grey lines show random draws from the posterior distributions of the parameters. 144

4.8 Mode linewidths of the radial, nominal p-mode and mixed modes as a function of frequency. The dashed blue line shows the fitted model for the widths whereas the red and blue diamonds show the linewidths of the radial modes and nominal p-modes respectively and the solid lines show the linewidths of the $\ell = 1$ mixed modes calculated using equation 4.36. 146

4.9 A comparison of the mixing function ζ using the formulation derived in Goupil et al. (2013) (red) and the updated version given by Deheuvels et al. (2015) (black). 148

4.10 Two artificial power spectra generated simulating a Kepler-56-like star for 1 year of observations with the spectrum smoothed with a boxcar filter of width 1μHz shown in red. The left hand plot would be as observed by *Kepler* and the right by TESS. The granulation power at $\nu_{\rm max}$ and shot noise are shown in addition to give an idea of the different properties of the two space missions. . 151

4.11 Two artificial power spectra generated simulating a Kepler-56-like star for 30 days of observations with the spectrum smoothed with a boxcar filter of width 10μHz shown in red. The left hand plot would be as observed by *Kepler* and the right by TESS. The granulation power at $\nu_{\rm max}$ and shot noise are shown in addition to give an idea of the different properties of the two space missions. . 151

4.12 Example time series and power spectrum of 7CMa generated using the values given in the text and Table 4.3. 154

4.13 Example time series and power spectrum of 7CMa generated using the values given in the text with a window function applied corresponding to 8 hours of observations a night. Note the change in scale of the power spectrum compared to Fig 4.12. 156

4.14 Example power spectra simulating KELT-11. The top panel shows the power spectrum with no window function applied for 14 days of observations. The bottom panel shows the power spectrum for 14 days of data with the window function applied (on the left) and for 7 days of data (on the right). Note the change in scale between the plots and the corruption caused by the window function. 158

5.1 A demonstration of the geometry involved in determining the obliquity. The left-hand plot shows the observer-oriented system where the z-axis is along the observer's line-of-sight. The right-hand plot shows the orbit-oriented coordinate system, this is a rotation of $\pi/2 - i_{\rm o}$ anti-clockwise in the x-axis from the observer-oriented coordinates. The angles are given as follows: ψ is the obliquity of the system, ϕ is the azimuthal angle, $i_{\rm o}$ is the orbital inclination angle, $i_{\rm s}$ is the stellar inclination angle and λ is the projected obliquity. The unit vectors $\mathbf{n}_{\rm o}$ and $\mathbf{n}_{\rm s}$ denote the directions of the stellar rotation and orbital angular momentum axes respectively. (Figure reproduced from Campante et al. 2016a) 164

5.2 A contour plot showing the relative mode power as a function of inclination angle for an $\ell = 1$ mode. The amplitude was taken to be unity, the linewidth was 0.2 μHz and the rotational splitting was 0.4 μHz. 167

5.3 Limit spectra for an $\ell = 1$ mode observed at different inclination angles (shown in black), the outer components are show in red, the central components are given in green. The angle of inclination used is labelled along with using a mode linewidth of 0.2 μHz and a rotational splitting of 0.4 μHz for an observing length of 4 years. 167

5.4 The same limit spectra given in Fig 5.3 but this time with χ^2_2 noise applied. . 168

5.5 Probability density functions describing the Fisher distribution given by Eqn 5.8 for various values of κ. 170

5.6 An example posterior PDF for KIC 3544595 where the histogram of 80,000 samples is shown in black. The red line shows the kernel density estimate using 1,000 randomly chosen samples (without replacement). 173

5.7 Kernel density estimates of the combined posterior distributions of $\cos i_s$ coloured according to the properties of the systems. Those with only one observed transit planet are shown in blue, those with multiple observed transiting planets are shown in red and the combined distribution for all the systems is given by the black solid line. 173

5.8 Kernel density estimates of the concatenated posterior distributions of $\cos i_s$ as extracted from the KOI in the asteroseismic sample. The different types of system are colour-coded and the prior used in the fitting process is given by the dashed black line. 174

5.9 Posterior probability distributions of κ (the concentration parameter) for the combined sample under the Fisher distribution model. The three different sub-samples are coloured as before and the posterior distribution for the asteroseismic sample is also shown as the dashed line for reference. 180

5.10 Posterior probability distributions of f and κ for the mixture model. The different sub-samples are coloured as before the the asteroseismic sample result is shown again for reference. 181

6.1 The top panel shows a log-log plot of the power spectrum for KIC 5553307 in black and a subsequent smoothed spectrum (1 μHz boxcar) in red. The bottom panel shows a single radial order of the above spectrum with key features overplotted. The red dotted lines show the position of the radial modes identified using the red giant universal oscillation pattern (Mosser et al., 2011). The green dotted lines show the position of the quadrupole ($\ell = 2$) modes which were tuned by eye using the small frequency separation d_{02} (which is the separation in frequency between a radial mode of order n and a quadrupole mode of order $n - 1$). The blue diamonds give the predicted positions of the $\ell = 1$ mixed modes using equation 6.3 for a period spacing $\Delta\Pi_1 = 87.6$ s and coupling $q = 0.2$. The cyan and magenta triangles denote the predicted positions of the rotationally split components ($m = -1$ and $m = +1$ respectively) using equation 6.4 with $\nu_{s,\mathrm{max}} = 0.3$ μHz. 185

6.2 The left hand panel shows a fit to an mixed mode in the intermediate angle star KIC 3973247, the data is shown in black and the best fit in red. The right hand panel shows the same but this time for a mixed mode in the low angle star KIC 3645589. 189

6.3 The left hand panel shows the individual posterior PDFs in angle for KIC 3531478 (in black) and the concatenated posterior for the star (in red). The right hand panel shows the same thing except this time for the star KIC 8645227. 190

6.4 The final concatenated distribution in angle for all the stars in the sample, the isotropic distribution is shown as the red dotted line. The shaded orange region is the 1-σ bounds bootstrapped estimate. 190

6.5 Posterior PDF for the concentration parameter κ used in the first incarnation of the model. 192

6.6 Example distributions of the Fisher distribution including negative κ values in $i_{\rm s}$. The solid lines show the analytical expression whilst the histograms shows the result of Monte Carlo simulations. 193
6.7 Example distributions of the updated model for different values of κ in both $i_{\rm s}$ and $\cos i_{\rm s}$. The solid lines show the analytical expression whilst the histograms shows the result of Monte Carlo simulations. 194
6.8 A triangle plot from the hierarchical analysis using the updated model. 196
6.9 Declination plotted against right ascension for the stars in the sample, where the points are coloured according to their distance (calculated using asteroseismic scaling relations) in the left-hand plot and by *Kepler* magnitude in the right-hand plot. 197
6.10 Declination plotted against right ascension for the stars in the sample, where the points are coloured according to their average inclination angle as fitted in this work. 198
6.11 Concatenated posterior distributions for the real stars with the sample split according to the *Kepler* magnitude (on the left-hand side) and reduced-splitting (right-hand side). The red shows the distribution for stars above the respective median value and the blue for those below. The isotropic distribution is overplotted in black. 200
6.12 The minimum and maximum inclination angle achievable for a given signal to noise. The red can be interpreted as the maximum angle and the black as the minimum angle achievable. 202
6.13 Height over background plotted as a function of the reduced splitting for both the simulated data (coloured x markers) and the real data (shown in black). The colours of the simulated data denotes the signal-to-noise ratio, ranging from 1.2 (blue) to 194 (in yellow). 204

6.14 Fractional difference of fitted amplitude plotted as a function of H/B coloured by the signal-to-noise input for the simulations. 205
6.15 The fractional difference between the fitted amplitude and true amplitude for the simulations, where in each figure in the tile shows a different linewidth value used. The blue line denotes the mean value of the fractional difference averaged over the realisations and the red envelope shows the standard deviation of the distribution of fractional differences for each H/B value. 206
6.16 Fractional difference of fitted linewidth plotted as a function of H/B coloured by the linewidth input for the simulations. 208
6.17 The fractional difference between the fitted linewidth and true linewidth for the simulations, where in each figure in the tile shows a different linewidth value used. The blue line denotes the mean value of the fractional difference averaged over the realisations and the red envelope shows the standard deviation of the distribution of fractional differences for each H/B value. 209
6.18 Fractional difference of fitted inclination angle plotted as a function of H/B coloured by the lienwidth input for the simulations. 210
6.19 The fractional difference between the fitted inclination angle and true angle for the simulations, where in each figure in the tile shows a different linewidth value used. The blue line denotes the mean value of the fractional difference averaged over the realisations and the red envelope shows the standard deviation of the distribution of fractional differences for each H/B value. 210
6.20 Reduced splitting as a function of the height over background ratio for the real data (in red) and artificial data (in blue). 212

6.21 The left hand panel shows the concatenated posteriors for the real stars in black and the posteriors with the proposed cuts in H/B and Γ made in blue. The right hand panel shows the same thing but for the artificial stars. The isotropic distribution is overplotted in red in both cases. 214

6.22 A comparison between the uncertainties extracted from the posteriors of the individual modes (in red) and the concatenated posteriors for each star (in blue). The left-hand panel shows this for the real stars and the right-hand panel shows this for the artificial sample. 215

7.1 An image of the field surrounding KOI-3890 (identified by the red cross) by the Aladin Sky Atlas (`http://aladin.unistra.fr`). The field of view of image is 59.21" provided and the star in close proximity is another red giant KIC 8564966. 222

7.2 Target pixel files for KIC 8564966 (left-hand panel) and KOI-3890 (right-hand panel). The mask used to extract the photometry is shown in the blue highlighted region. 222

7.3 A comparison between the power spectra of KOI-3890 shown in black and KIC 8564966 shown in red zooming in on the region around the oscillations. 223

7.4 Signal residue outputs from the BLS algorithm for KOI-3890 (on the left) and KIC 8564966 (on the right). The peak around 150 days in the left hand plot shows the presence of the eclipses with that period, whereas for KIC 8564966 there is no such peak, again indicating a lack of contamination. 224

7.5 The best fitting simultaneous transit and radial velocity observations. The upper panel shows the detrended lightcurve, folded on the period from the fitting procedure, with the out of transit lightcurve normalised to 1. The middle panel shows the radial velocity signal, observed by TRES as the black points, and the best fitting model in red, with errobars inflated by factor 20 observe them. Lower panel shows the residuals of radial velocity fit (with non-inflated errorbars). 226

7.6 The background fit to the power spectrum of KOI-3890. The power spectrum is shown in black and in blue is the spectrum rebinned over 1 μHz. The green lines show random draws from the posterior distributions of the fitted parameters. Inset is the posterior probability density function for $\nu_{\rm max}$. 228

7.7 The power spectrum of KOI-3890 zoomed in to show the three radial orders used in the determination of the inclination angle. For each $\ell = 1$ mixed mode the blue triangles show the position of the $m = 0$ component, the cyan marker shows the position of the $m = -1$ and the magenta marker shows the position of the $m = +1$. Note the large change in scale as a function radial order. . . . 233

7.8 Kernel density estimate (KDE) of the angle of inclination posterior probability distribution. The median value is shown by the solid red line whilst the 68.3% highest posterior density (HPD) credible interval is encompassed by the dashed red lines. 234

7.9 Kernel density estimate (KDE) of the Monte-Carlo estimate of the obliquity distribution for KOI-3890. The plot is annotated in the same manner as Fig 7.8. 235

A.1 The left hand panel shows a zoomed in view of the diurnal peaks at 81.02 and 57.87 μHz in the BiSON window power spectrum. The right hand panel shows the equivalent peaks in the artificial window function power spectrum showing how the effect of adding in the change in length of day has affected the shape and structure of the peaks. 251

A.2 Gap distribution of 23 years of BiSON data with $\Gamma = 1$ distribution overplotted in red with a mean of 42 minutes. 253

Listing of tables

2.1 Longitude and latitude values of BiSON stations 29

2.2 Parameter values for the limit spectrum constructed in figure 2.8. The labels corresponding to the various granulation components are to be used as a guide for the approximate timescales to which they correspond. 34

2.3 Parameters for the construction of the footprint. 41

3.1 Parameters for the Gaussian priors. 70

3.2 Parameters for the uniform priors used in background fit. In the case of using short-cadence data instead of long-cadence the parameters marked with an asterisk would need to be increased to ensure the optimal solution is found. . 70

3.3 Results of the fitting procedure for $\delta\nu_{\rm env}$. 97

3.4 Results of the fitting procedure for $H_{\rm env}$. 101

3.5 Results of the fitting procedure for $A_{\rm max}$. 102

4.1 A comparison of the relative visibilities for *Kepler*, TESS and SONG. 136

4.2 Derived parameters from the fit to the radial mode linewidths. 145

4.3 Literature values of $\log g$ and $T_{\rm eff}$ for 7CMa along with the $\nu_{\rm max}$ value calculated from equation 4.40. The respective sources are given beneath the table. 153

5.1 Results from the full Bayesian hierarchical analysis on each sample (reproduced from Campante et al. (2016a)). 177

6.1 Input parameters into time domain simulations, those with no upper bound denote fixed values over all simulations. ($\nu_{\rm bw}$ is the frequency bin width) . . . 203

7.1 Spectroscopic parameters derived from the TRES observations for KOI-3890. . 221

7.2 Final values from simultaneous transit and radial velocity fit, along with the priors applied in the fit. Uniform priors are in range [a, b] in the final column, while normal priors [μ,σ] values are shown. 227

7.3 Stellar properties derived from grid-modelling procedure. 230

List of Publications

Campante T. L., et al., 2016a, ApJ, 819, 85

Campante T. L., et al., 2016b, ApJ, 830, 138

Campante T. L., et al., 2017, preprint, (arXiv:1704.01794)

Davies G. R., et al., 2017, A&A, 598, L4

Kuszlewicz J., Davies G. R., Chaplin W. J., 2015, in European Physical Journal Web of Conferences. p. 06041, doi:10.1051/epjconf/201510106041

Miglio A., et al., 2016, MNRAS, 461, 760

North T. S. H., et al., 2017, MNRAS, 465, 1308

Quinn S. N., et al., 2015, ApJ, 803, 49

Statement of Collaboration

All the work presented in this thesis is my own, some of the work presented was performed with the help of collaborators.

- Chapter 5: the work presented in the chapter forms part of the paper Campante et al. (2016a). The work that is covered in Chapter 5 appears in Section 4 of Campante et al. (2016a).
- Chapter 7: the radial velocity data for KOI-3890 was collected and analysed by Dave Latham and Allyson Bieryla (both Harvard). The transiting fitting in Section 7.4 was performed by Thomas North.

1

An Introduction to Helio- and Asteroseismology

Over the past decade, there has been a space-based revolution courtesy of the *Kepler* and CoRoT missions propelling asteroseismology to the forefront of astrophysics. The study of solar-like oscillations, those excited and damped by near-surface turbulent convection had long been confined to the Sun save a few rare cases [e.g Arcturus (Smith et al., 1987), Procyon (Brown et al., 1991; Arentoft et al., 2008; Bedding et al., 2010) and αCenB (Kjeldsen et al., 2005)]. The long-period, high quality datasets available courtesy of the space-based missions have resulted in a golden-age for the field with oscillations detected in hundreds of solar-type stars and, the big surprise of the missions, thousands of red giants. In this work, the focus will be on stars showing solar-like oscillations as opposed to the classical pulsators, e.g. RR Lyrae, Cepheids (Leavitt & Pickering, 1912) or δ Scutis (see Aerts et al. 2010 and references

therein).

Staying close to home the Sun, our nearest star, provides a great opportunity to study such an object in a way that is still not possible for other stars. Helioseismology provides a means to uncover information about the Sun using its intrinsic oscillations excited by near surface turbulent convection. Oscillations were first found in the Sun in the 1960s (Leighton et al., 1962) which showed periods of above five minutes. These were later found to be acoustic pressure (p) modes. Since that first initial discovery a lot of further observational work was carried out (Deubner, 1975; Claverie et al., 1979) which coupled with theoretical work showed that these oscillations were global modes of oscillations. It should also be noted that the discovery that the Sun supported whole-Sun low-degree modes was made by the group in Birmingham from the work by Claverie et al. (1979).

The detailed study of the properties of these global modes provides information that can subsequently be used to update solar (Christensen-Dalsgaard et al., 1996; Christensen-Dalsgaard, 2002), and therefore stellar models. The importance of such measurements is exemplified by the role of helioseismology in the solar neutrino problem (Bahcall, 1972), whereby the neutrino fluxes observed from the Sun were too low compared to the predicted fluxes from models. It was thought that the issue lay in the solar models, however this was inconsistent with results from helioseismology and hence confirmed the issue lay within particle physics as opposed to the solar models (Elsworth et al., 1990).

The use of ground-based networks to study the Sun has been highly successful in the detection and characterisation of solar p-modes in the global effort to learn more about the internal structure of the Sun (e.g. Broomhall et al. 2012; Davies et al. 2014a). The Birmingham Solar Oscillations Network (BiSON) is one such network (Chaplin et al., 1996a). It is a network of six stations involved in Sun-as-a-star observations, in other words the investigation of low-degree modes (disk integrated data). The stations are placed all around the world with the aim of trying to provide continuous observations of the Sun such that

the data has as high a duty cycle (or fill) as possible. This is not always possible due to mechanical faults or atmospheric interruptions (such as clouds), but even so the fill is in the region of 80% for the last 20 years.

Before the era of *Kepler* (Borucki et al., 2010) and CoRoT (Baglin et al., 2006) there were a few detections of solar-like oscillations, the majority of which were from ground-based observations (Brown et al., 1991; Kjeldsen et al., 2005; Arentoft et al., 2008) in addition to a small number from the first cohort of space-based missions (WIRE, MOST, SMEI; see Bedding 2014 for more information). It was not until the introduction of *Kepler* and CoRoT that the number of stars with detected oscillations greatly increased, resulting in new applications of asteroseismology that were not previously possible. Since then, asteroseismology has become a valuable tool when applied to exoplanet systems due to its ability to better constrain the star's fundamental parameters (e.g. Silva Aguirre et al. 2015), along with the ability to make inferences on galactic structure through galactic archaeology (e.g. Miglio et al. 2013). Asteroseismology offers the ability to extract information about the evolutionary state of the star (when in the red giant phase of its evolution) through the oscillations themselves (more on this topic will be discussed later) (Stello et al., 2013; Elsworth et al., 2016), which is very difficult to attain otherwise.

1.1 Brief Overview of Stellar Evolution

The main body of this work is occupied with solar-like oscillations in red giants and so it is therefore worth taking the time to briefly explain how a star evolves from the main-sequence onto the red giant branch and how this impacts asteroseismology.

The majority of red giants observed by *Kepler* have masses lower than $2M_{\odot}$ and so we will only concern ourselves with the evolution of low mass stars, generally defined as those stars that will ignite helium under degenerate conditions[1].

[1]The secondary clump stars observed by *Kepler* will have masses $M > 2M_{\odot}$ and so their evolution will be briefly mentioned.

Let us start with a 1 $M_\odot$ star at zero-age main-sequence (point A on Fig 1.1) and follow its evolution onto the red giant branch. For stars with radiative cores ($M \leq 1.1 M_\odot$) the dominant mechanism of energy production is of course the fusing of Hydrogen to ^{4}He via the proton-proton (pp) chain[2]. The resulting reactions cause an increase in the mean molecular weight (μ) which affects the hydrostatic equilibrium of the star by decreasing the gas pressure. In order to counter the increasing μ and decreasing pressure, the core contracts increasing the temperature which leads to an increase in the luminosity of the star over its time on the main-sequence, this is observed as part A to B in Fig 1.1.

As the star evolves off the main sequence it moves onto the subgiant branch where the core contraction continues and the envelope starts to expand. This results in the almost constant luminosity due to the cooling of the envelope. The next part of the star's evolution is most relevant to the rest of this work. After the subgiant branch the star develops a large convective envelope as a result of the increased opacity due to the lower temperatures and now sits on the red-giant branch (point D in Fig 1.1).

On the red-giant branch the star is burning hydrogen in a thin shell above a degenerate He core. The material produced by the shell burning falls onto the degenerate core and the shell moves outwards in radius. As the mass of the degenerate core increases the core contracts which results in an increase in core temperature[3]. This increase in temperature also affects the shell and so it contracts in response to the temperature increase, thereby increasing the rate of energy production. Soon the density gradient between the core and envelope is so large that they are basically decoupled from one another. This is important since it means that the stellar luminosity is completely determined by the core properties and not of the envelope (or the total mass of the star). The star will continue to evolve through the RGB bump (Riello et al., 2003; Christensen-Dalsgaard, 2015) and proceed through the Helium

[2]This is also true for stars with small convective cores, the contribution from the CNO cycle becomes progressively more important as the mass increases.

[3]In principle contraction does not induce heating in degenerate conditions. The contracting core releases energy lost during contraction, part of which heats the H-shell which in turn heats the core.

Flash (see Hekker & Christensen-Dalsgaard 2016 for more details). After this it reaches the next point of discussion, the red clump.

Once the star begins helium core burning (as a result of the flash), it continues to also burn hydrogen in a shell around the no longer inert helium core. After the helium flash the density of the shell burning region decreases (along with its temperature) which results in a decrease in the star's luminosity. In addition, the core continues to expand as a result of the deposition of ashes from the hydrogen shell-burning and so the envelope contracts. This phase of the star's evolution is the red clump, since all stars will have very similar luminosities and core masses. It is also relatively long lived and so one expects to see a lot of stars there.

The final phase of evolution that is relevant in this work is the secondary clump. This is populated by higher mass stars ($M \geq 2M_{\odot}$) which can ignite helium-core-burning under non-degenerate conditions. The core masses of stars in the secondary clump can greatly differ and so whilst they form a clump this is not quite at the same position as the red clump since they tend to have slightly higher effective temperatures.

1.2 Solar-like Oscillations

Solar-like oscillations are both driven and damped by near-surface turbulent convection and so this shows a necessary condition for a star to show solar-like oscillations. The oscillations come about due to two distinct restoring forces, those where the pressure gradient is the restoring force (i.e. of an acoustic nature) which are pressure (p) modes; and those where buoyancy is the restoring force which are gravity (g) modes. These two different types of mode are also sensitive to different regions of the star; the p-modes propagate throughout the majority of the star[4], whereas the g-modes propagate in the radiative regions of the star. In the case of the Sun and solar-type stars although p-modes probe the entirety of the star their sensitivity to the core regions is lower due to the acoustic waves spending much less

[4]In the case of radial modes the waves propagate throughout the entirety of the star.

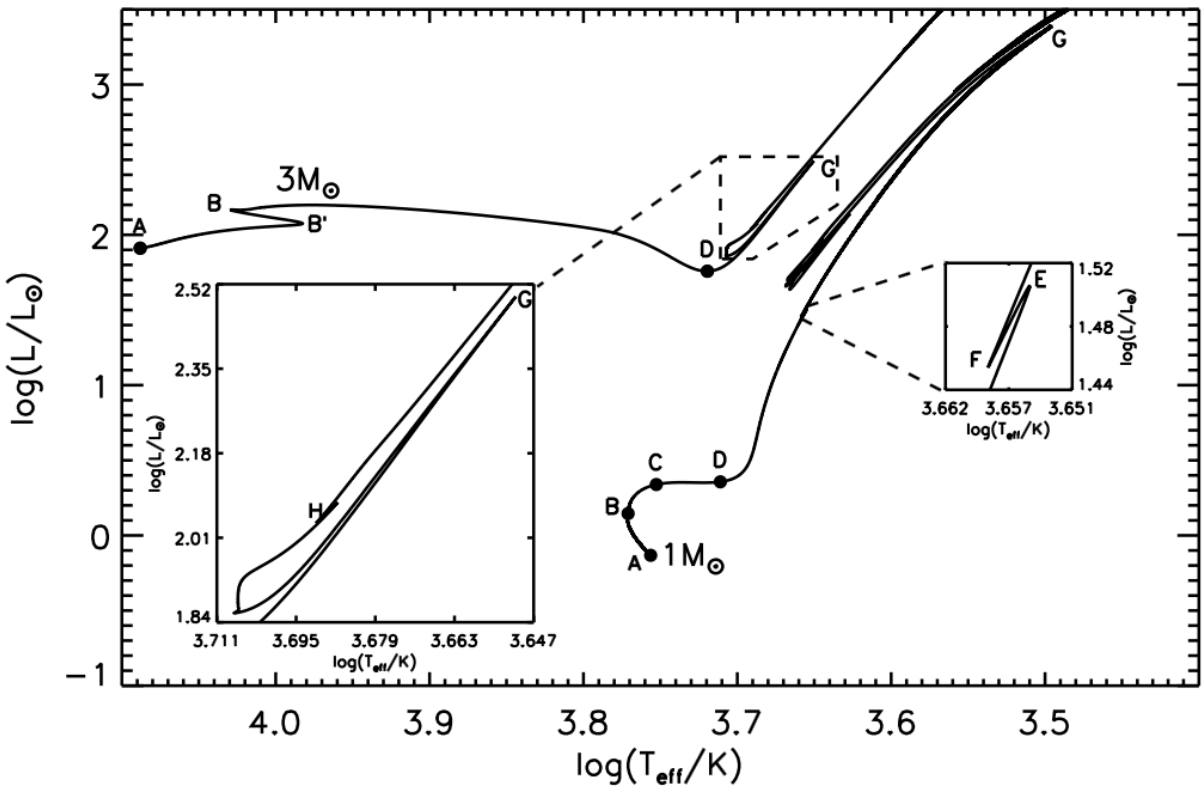


Figure 1.1: A Hertzsprung-Russell diagram showing a 1 $M_\odot$ and 3 $M_\odot$ evolutionary track compute using MESA (Paxton et al., 2011). The letters on the 1 $M_\odot$ are referred to in the text when needed, the inset plot on the right hand side shows the position of the RGB bump in the 1 $M_\odot$ track. Figure from Hekker & Christensen-Dalsgaard (2016).

time in the hotter, deeper regions compared to the cooler envelope whereby radial modes will probe deeper and the lower turning point increases in radius as a function of mode degree. Therefore their diagnostic potential for the core is limited. G-modes, however, probe these inner radiative regions and the core therefore providing diagnostic potential in those regions[5].

Stars can be very simply thought of as balls of gas under which the force of gravity is balanced by radiation pressure thereby maintaining the system in hydrostatic equilibrium. The oscillations themselves can be considered as perturbations about the equilibrium state, and can be described by spherical harmonics, $Y_\ell^m(\theta, \phi)$, of degree ℓ and order m where (θ, ϕ) are spherical polar coordinates θ is colatitude and ϕ is longitude.

Each mode of oscillation can be characterised by three wavenumbers: the radial order, n, can be equated to the number of nodes in the radial direction (from the centre of the star to the surface), the degree, ℓ, which gives the total number of node lines and an azimuthal order, m, which gives the orientation (i.e. the number of lines crossing the equator).

Due to the fact that both the instruments we consider, the spectrometers of BiSON, and the CCDs of *Kepler*, gather disc-integrated data, geometric effects mean that low-degree modes, $\ell \leq 3$, dominate. BiSON can also detect $\ell = 4$ in long datasets, however they are very weak (Chaplin et al., 1996b).

There are many properties of the oscillations that are of interest, such as the frequencies, amplitudes, linewidths, all of which help shape the data that is processed and analysed. The dominant global asteroseismic parameters of interest are the frequency of maximum power, $\nu_{\rm max}$, and large frequency separation $\Delta\nu$, which are important diagnostics in terms of mass, radius and evolutionary state.

[5]In the case of higher mass stars g-modes are much more visible due to the shrinking convective envelope and larger radiative interiors.

1.2.1 $\nu_{\rm max}$ & $\Delta\nu$

The power excess caused by the oscillations is an important diagnostic that is commonly extracted in almost every asteroseismic analysis. The frequency at which this power excess peaks is referred to as the frequency of maximum power, $\nu_{\rm max}$, which has been shown to scale very well the with acoustic cut-off frequency (under the approximation of an isothermal atmosphere)[6]

$$\nu_{\rm ac} = \frac{c}{4\pi H_p}, \tag{1.1}$$

where c is the sound speed and H_p is the pressure scale height. The surface gravity and effective temperature of stars decrease as they evolve which causes both $\nu_{\rm max}$ and the acoustic cut-off frequency to move towards lower frequencies. As a result there is an assumed relation between $\nu_{\rm max}$, the effective temperature of the star and the surface gravity (e.g. Brown & Gilliland 1994)

$$\nu_{\rm max} \propto g T_{\rm eff}^{-1/2}, \tag{1.2}$$

where g is the surface gravity of the star and $T_{\rm eff}$ is its effective temperature.

There are more than a few methods for determining $\nu_{\rm max}$ which either involve a full background fit to the power spectrum of the data where $\nu_{\rm max}$ is defined as the centre of a Gaussian envelope (e.g. Kallinger et al. 2014, which will be explained in Chapter3), or from the smoothed power spectrum (e.g. Huber et al. 2009). An example is shown in Fig 1.2 for a red giant star illustrating the position of the global asteroseismic parameters $\nu_{\rm max}$ and $\Delta\nu$.

In addition to $\nu_{\rm max}$, the other important asteroseismic index is the large frequency separation $\Delta\nu$. This is defined as the difference between modes of the same ℓ but of subsequent

[6]It is important to note that the scaling is essentially empirical and is not yet theoretically firmly understood.

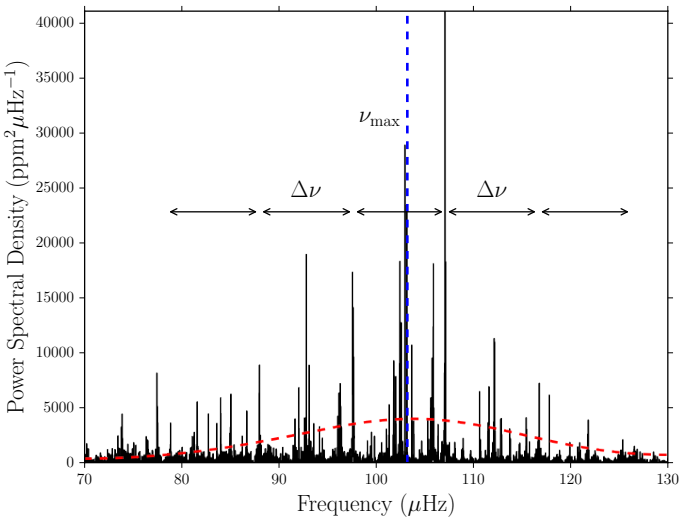


Figure 1.2: An example background-subtracted power spectrum of the red giant branch star KIC 8564976. The frequency of maximum power is shown by the dotted blue line and the arrows show the approximate positions of the radial modes and the large frequency separation $\Delta\nu$. The heavily smoothed power spectrum is shown in red (amplified by a few orders of magnitude to make it visible) in order to show the power excess caused by the oscillations themselves.

radial order, i.e. $\Delta\nu \approx \nu_{n+1,\ell} - \nu_{n,\ell}$ (for large n) which is in turn linked to the acoustic radius of the star through

$$\Delta\nu = \left(2\int_0^R \frac{\mathrm{d}r}{c}\right)^{-1}. \tag{1.3}$$

where the sound speed of the medium is given by c, R is the stellar radius. It can also be shown that the large frequency separation scales, to a good approximation, with the mean density of the star (Ulrich, 1986). This follows from the idea that the fundamental period of a pulsating star (such as the Cepheids) should give a measure of the acoustic travel time which is in turn dependent upon the mean density of the star. So it can be expected that $\Delta\nu$ will behave like the fundamental pulsation period, therefore it should scale with the mean density of the star.

The measurement of $\Delta\nu$ can be perturbed by variations of the local sound speed due to rapid density changes in the interior of the star, such as at the base of the convective zone

(e.g. Verma et al. 2017) or the presence of the helium second-ionisation zone (e.g Miglio et al. 2010; Broomhall et al. 2014; Vrard et al. 2015). But these perturbations give greater insight into the stellar structure, helping add extra constraints to stellar models.

Since both $\nu_{\rm max}$ and $\Delta\nu$ are dependent on the mass and radius of the star, these relationships can be inverted to provide estimates of the mass and radius of the star in the form of scaling relations

$$\frac{R}{R_\odot} = \left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}}\right)\left(\frac{\Delta\nu}{\Delta\nu_\odot}\right)^{-2}\left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^{1/2}, \tag{1.4}$$

$$\frac{M}{M_\odot} = \left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}}\right)^{3}\left(\frac{\Delta\nu}{\Delta\nu_\odot}\right)^{-4}\left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^{3/2}, \tag{1.5}$$

where the solar reference values are given by $\Delta\nu_\odot = 134.9$ μHz and $\nu_{\rm max,\odot} = 3090$ μHz. These provide approximate masses and radii for red giants but departures from the scaling relations are seen (see e.g. Frandsen et al. 2013 or Gaulme et al. 2013 for more detail). This is mainly due to the assumption of homology in the construction of the scaling relations, that all stars are scaled versions of the Sun. Certainly in the case of red giants where the internal structure is not simply a scaled version of the Sun deviations of up to 10% can be seen in mass (e.g. Belkacem et al. 2013).

1.2.2 Asymptotic frequency spacings

The properties of the oscillations can be extracted from expressions derived from an asymptotic formulation, which, whilst demonstrating a good agreement with observations for main-sequence stars, shows significant deviations when the stars have evolved off the main-sequence (due to the presence of modes having a mixed p- and g-mode properties). Following on from the previous section, the asymptotic formalism is applicable in the regime where $n/\ell \rightarrow \infty$and so in the case of the Sun, for example, this formalism is acceptable as the order of the modes

around the frequency of maximum oscillation power ($\nu_{\max}$) is much greater than one[7].

This asymptotic formulation is given as follows

$$\nu_{n,\ell} = \left(n + \frac{\ell}{2} + \varepsilon\right)\Delta\nu - \delta\nu_{0,\ell}. \tag{1.6}$$

where n and ℓ are the radial order and degree respectively, ε is a phase term. This equation can be interpreted simply and shows that the radial ($\ell = 0$) modes are evenly-spaced in frequency according to the large frequency separation, whereas $\ell = 1$ modes can be seen to have an approximate position of $\nu_{n,0} + \Delta\nu/2$, and for $\ell = 2$ modes this should be $\nu_{n,0} + \Delta\nu$.

In the case of this thesis, the focus will be mainly on red giants (especially with regards to the acoustic p-modes) and so the asymptotic expression has been updated for red giants by Mosser et al. (2011) under the universal red-giant oscillation pattern. This has the following form

$$\nu_{n,\ell} = \left[n + \frac{\ell}{2} + \varepsilon(\Delta\nu) + \frac{\alpha_\ell}{2}\left(n - \frac{\nu_{\max}}{\Delta\nu}\right)^2\right]\Delta\nu - \delta\nu_{0,\ell} \tag{1.7}$$

where α describes the curvature found in the frequencies of the radial modes (this is explained. This is equivalent to a linearly increasing $\Delta\nu$ as demonstrated later in Chapter 4[8].

There are however deviations from this degeneracy due to the sound-speed gradient in the core regions and so this is incorporated into the small frequency-separation $\delta\nu_{\ell,\ell+2}$, where

$$\delta\nu_{\ell,\ell+2} = \nu_{n,\ell} - \nu_{n-1,\ell+2} \approx -(4\ell + 6)\frac{\Delta\nu}{4\pi^2\nu_{n,\ell}}\int_0^R \frac{\mathrm{d}c}{\mathrm{d}r}\frac{\mathrm{d}r}{r} \tag{1.8}$$

G-modes, on the other hand, are a fundamentally different type of mode whose restoring force is buoyancy as opposed to being of an acoustic nature like p-modes. Due to the difference

[7]The JWKB approximation (Jeffreys, 1925; Kramers, 1926; Wentzel, 1926) can be used to derive eigenvalue conditions for high order non-radial oscillations giving rise to the asymptotic regime mentioned above.

[8]Some terms are given as a function of $\Delta\nu$ in order to reflect their scaling of the form $A + \log_{10}\Delta\nu$ obtained by Mosser et al. (2011)

in character, g-modes possess different propagation cavities to p-modes in addition to being equally spaced in period rather than frequency. The asymptotic formalism is also slightly different and can be used under very different circumstances and so in the case of solar g-modes the asymptotic approximation can be used for low-degree modes with a frequency $\nu_{nl} \leq 200\mu$Hz (see Tassoul (1980) for a full derivation of the asymptotic expression)[9]. The resulting expression (to second-order) for the period, denoted by $\Pi_{n\ell} = 1/\nu_{n\ell}$, is given by

$$\Pi_{n\ell} \approx \langle \Pi_{n\ell} \rangle = \frac{\Pi_0}{\sqrt{\ell(\ell+1)}} \left(n + \frac{\ell}{2} - \frac{1}{4} + \theta \right) + \frac{P_0^2}{\langle \Pi_{n\ell} \rangle} \frac{\ell(\ell+1)V_1 + V_2}{\ell(\ell+1)} \ , \tag{1.9}$$

where, following Appourchaux et al. (2010),

$$\Pi_0 = \frac{2\pi^2}{\int_0^{r_{cz}} (N/r) \mathrm{d}r} \ . \tag{1.10}$$

where r_{cz} is the radius of the convective zone, N is the Brunt-Väisälä frequency and is defined as follows:

$$N^2 = g \left[\frac{1}{\gamma_1 p} \frac{dp}{dr} - \frac{1}{\rho} \frac{d\rho}{dr} \right] \ , \tag{1.11}$$

where γ_1 is the first adiabatic exponent. The Brunt-Väisälä frequency is the buoyancy frequency and so corresponds to the the frequency at which a parcel of gas will oscillate if displaced from its equilibrium position.

The similarities between the two different types of mode should be apparent in the sense that both have a comb-like structure (one in frequency, the other in period). The g-mode period (equation 1.9) is closely linked to the Brunt-Väisälä frequency, N, as a result of the term Π_0. More importantly, from an observational point of view, modes of the same ℓ will be spaced approximately evenly in period according to $\Pi_0/\sqrt{(\ell(\ell+1))}$ and this means that the comb-like structure in period is a good observational signature of g-modes.

[9]For g-modes this corresponds to the regime of high radial order.

1.3 Modes of mixed character

In the case of more evolved stars, the oscillation spectra get far more complicated due to the introduction of coupling between p and g-modes. As the star proceeds to evolve, a very large density gradient is created between the core and the surrounding region which in turn results in a large increase in the buoyancy frequency, N. The frequencies of the p- and g-modes are no longer clearly separated because as N increases into the p-mode regime. The two cavities of propagation for p-modes and g-modes effectively couple together, leading to mode bumping (see Unno et al. 1989 for a more detailed discussion) whereby the frequencies of the coupled modes are shifted from their expected regular spacing. The g-mode will couple to a p-mode with a similar frequency and so the mode will have a mixed character and propagate as a gravity wave in the radiative regions and as an acoustic wave in the convective regions. This gives a unique opportunity to extract information about both the core and envelope from the properties of the modes.

An asymptotic framework for modes of a mixed character was provided by Shibahashi (1979) and later by Unno et al. (1989). For the sake of brevity the expression will not be derived and following Mosser et al. (2012b) the frequency for a dipole (i.e. $\ell = 1$) mixed mode that is coupled to a p-mode is given by

$$\nu = \nu_{n_p,\ell=1} + \frac{\Delta\nu}{\pi} \arctan\left[q \tan\pi\left(\frac{1}{\Delta\Pi_1\nu} - \varepsilon_g\right)\right] \tag{1.12}$$

where $\nu_{n_p,\ell=1}$ is the frequency of a pure dipole p-mode, $\Delta\Pi_1 = \Delta\Pi_0/\sqrt{2}$ is the $\ell = 1$ period spacing, q is the coupling between the p- and g-mode and ε_g is another phase term (although is commonly assumed to be zero). The more complicated nature of the above equation helps show the difficulty that is present when identifying mixed modes in red giant stars. This provides a very simple overview of mixed modes, for a more detailed description see Chapter 4.

1.4 Structure of the Thesis

This work in this thesis is organised as follows:

Chapter 2 details the progress made searching for solar g-modes using the Birmingham Solar Oscillations Network (BiSON) that is to be presented in Kuszlewicz et al. (in prep). I cover the initial problems with searching the low-frequency regime for g-modes such as the diurnal peaks along with how to remove them. In addition I also provide upper limits on the g-mode amplitudes.

Chapter 3 describes the ensemble characterisation of a set of red giants stars observed with *Kepler* through their background power in the power spectrum. This chapter provides updated scaling relations in fundamental global oscillation parameters with 4 years of *Kepler* data as well as a possible new way in which to classify the evolutionary state of red giant stars.

Chapter 4 describes the creation of artificial red giant power spectra that were used in Chapter 6. The process and approximations used to generate the spectra are also given.

Chapter 5 details the statistical analysis presented in Campante et al. (2016a). This revolves around inferring the obliquity distribution of a set of asteroseismic exoplanet hosts through the use of hierarchical Bayesian inference.

Chapter 6 describes work to be presented in Kuszlewicz et al. (in prep) which involves exploring the stellar inclination angle distribution for a sample of *Kepler* red giants. Through the use of hierarchical Bayesian modelling the distribution is modelled to check for consistency with the expected isotropic distribution. Additional biases are also explored to explain the observed bias seen at 90°.

The final chapter, Chapter 7, details work to be presented in Kuszlewicz et al. (in prep) explaining the determination of the obliquity of a red giant, M-dwarf eclipsing binary using asteroseismology.

This thesis will focus on the use of probabilistic techniques and their application to asteroseismic datasets. All of the work can be linked together through the data analysis techniques and (in most cases) the use of buoyancy-driven oscillations to derive the asteroseismic quantities being analysed.

2

BiSON and the Quest for Solar g-modes

The quest for solar g-modes has been the motivation behind many missions since they would provide a wealth of information about the solar core, something that we know surprisingly little about. As of yet there have been no unambiguous detections of solar g-modes, although García et al. (2007) has reported detections using the Global Oscillations at Low Frequency (GOLF) instrument (Gabriel et al., 1995) aboard SOHO which have yet to be reproduced.

One of the main reasons for solar g-modes being so hard to detect is their very small amplitudes, this is due to their evanescent nature in the convective envelope giving rise to small amplitudes at the surface. Appourchaux (2003b) established theoretical upper limits on g-mode amplitudes in the Sun of around 3 mms^{-1} at 200 μHz, this is in stark contrast to the amplitudes of the p-modes which are in the tens of cms^{-1} range. This upper limit was set using 9 years of BiSON data from Chaplin et al. (2002) assuming a quadruplet structure of the modes. More recent work was performed by Belkacem et al. (2009) who attained

g-modes amplitudes of ~ 3 mms^{-1} for an $\ell = 2$ mode at 100μHz. This is very close to the GOLF detection limit and should be in the range of BiSON as well (although as we shall soon see, it is not quite that simple). However the prediction of g-mode amplitudes are very difficult given the limited knowledge about the g-mode properties themselves. Therefore it can be seen that the detection of g-modes or, at the very least, observational upper limits would help constrain the g-mode properties. There has also been extensive theoretical work looking at predicting g-mode frequencies (e.g. Provost et al. 2000; Mathur et al. 2007) and amplitudes (e.g.Belkacem et al. 2009).

2.1 BiSON Data

The BiSON stations that are used to collect the data analysed in this work are situated at Mount Wilson (USA), Las Campanas (Chile), Izaña (Tenerife), Sutherland (South Africa), Carnarvon and Narrabri (Australia). The corresponding station positions in terms of latitude and longitude are given in Table 2.1. For more information on the instruments used (the resonant-scattering spectrometer) and the way in which the data are calibrated see Brookes et al. (1978) and Elsworth et al. (1995).

The dataset analysed in this work was collected over a period starting from 16th April 1992 until 22nd July 2013 resulting in just over 20 years of data in total. Due to the presence of gaps in the data there is not full coverage and so the fill (or duty cycle), defined as the percentage of the timeseries where data is taken is $\sim$75%. The resulting power spectrum smoothed with a boxcar of width 1 μHz is shown in Fig 2.1.

The region around 3000 μHz contains the 5-minute oscillations of the solar p-modes. Due to the nature of the instrument (disk-integrated) all observed p-modes will correspond to p-modes with order $\ell \leq 3$, with the possibility of some very weak $\ell = 4, 5$ modes also being present (Chaplin et al., 1996b). However, this is not the region of interest when searching for g-modes, instead the low-frequency regime is more important.

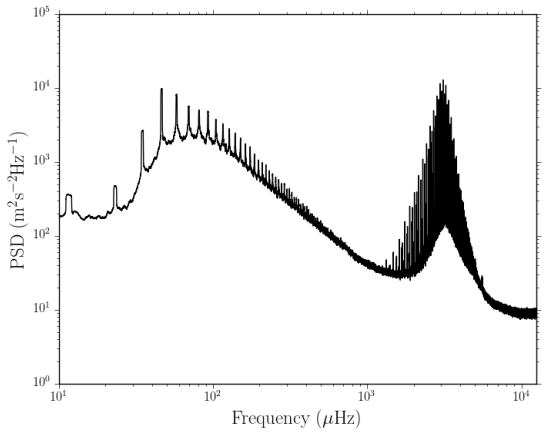


Figure 2.1: Power spectrum of 21 years of BiSON data where the strong diurnal peaks can be seen at low frequency as well as the Gaussian-like envelope of the p-mode region at high frequency.

2.1.1 Low-frequency Regime

In the search for g-modes, the region to be analysed lies in the frequency range $50 - 200$ μHz and so if we look back to Fig 2.1 some clear structure is already present. At this point it is important to reiterate the fact that the fill of the data is $\sim$75% which means there are a considerable number of gaps present in the data. This will contribute towards contamination seen in the power spectrum which will be the main focus of the next few sections. Understanding the nature of this contamination is important as it can help identify and clear out the cause. An explanation of the way in which the power spectrum is computed and calibrated is given in Appendix A.1.

It is necessary to consider the idea of a "window function" which tells us where there is and is not data, by either being 1 in the former and 0 in the latter case. The final dataset can be thought of as the initial uninterrupted data (that would ideally want to be received) multiplied by the window function (which introduces gaps due to interruptions) in the time domain. In the frequency domain this multiplication can be seen as a convolution

$$\mathcal{F}[x(t) \times w(t)] = X(\nu) * W(\nu), \tag{2.1}$$

where $x(t)$ is the data, $w(t)$ is the window function and the capital letters denote their respective Fourier transforms.

This results in there being some contamination present in the power spectrum as a result of the convolution of the data with the window function, this is commonly seen as sidebands in ground-based data (e.g. Arentoft et al. 2008). At higher frequencies, where the amplitudes of the signal are larger these sidebands are easy to account for, however in the low-frequency regime their cause is far from trivial. For low amplitude modes very close to the noise level these sidebands can also cause problems for detection and assessing upper limits on amplitudes.

2.1.2 Preparation of the Power Spectrum

In order to start searching for g-modes and to ensure that our data is as well prepared as possible it is necessary to perform a few more actions in addition to the calibration of the power spectrum.

Removal of DC Component

If there is a non-zero mean present within the dataset then this will manifest itself in the power spectrum as power at zero frequency (i.e. in the zeroth bin). Whilst this does not seem to be an issue, in the case of data with gaps the DC component can cause unwanted artefacts and should be removed to stop this from happening. The calculation of the mean for a gapped time series involves only points with data, and the mean is removed from only those points, so that the gaps remain as zeros.

The removal of the mean may seem like a trivial point to bring up, but it does have consequences when analysing the power spectrum. The window function appears in the power spectrum as a result of the multiplication in the time domain or convolution in the frequency domain. The positions of the peaks in the window function power spectrum and

in the power spectrum of the data can be compared in order to try and determine how the window function "bleeds through" into the power spectrum. The diurnal peaks appear at the same frequencies in the window function power spectrum and the power spectrum of the data. Due to the convolution operation in the frequency domain the DC component is most likely responsible for the appearance of the diurnal peaks. The reasoning behind this is that when a function is convolved with a unit impulse at zero frequency then the original function is returned. Due to the fact that the DC component dominates the low frequency part of the spectrum (in terms of power) the same idea can be applied and so the diurnal peaks appear as the result of the convolution of the window function power spectrum with the DC component.

The implication is that by removing the mean from the time series the manifestation of the window function in the spectrum should disappear as the DC component should then be zero. This idea will be tested for a few different scenarios in order to consolidate the rationale behind the removal of the mean from the time series.

Subtracting the Mean

A simple example would be that of a sine wave with a frequency, ν_s, of 5mHz which is very different with respect to the diurnal peak frequency (i.e. $\nu_s \gg 11.57\mu$Hz) with an added constant offset with the BiSON window function imposed upon it with length of observations of a year. The power spectrum is then created in the two cases when the mean is and is not subtracted from the dataset.

The result of subtracting the mean (as shown in figure 2.2) is very striking. The contamination in the window function at low frequency has been completely nullified by the subtraction of the mean (before the power spectrum was constructed). This is rather remarkable, although it simply shows that what was expected to happen really does. But this is only for a simple example and so should be tested on the BiSON data in order to see how

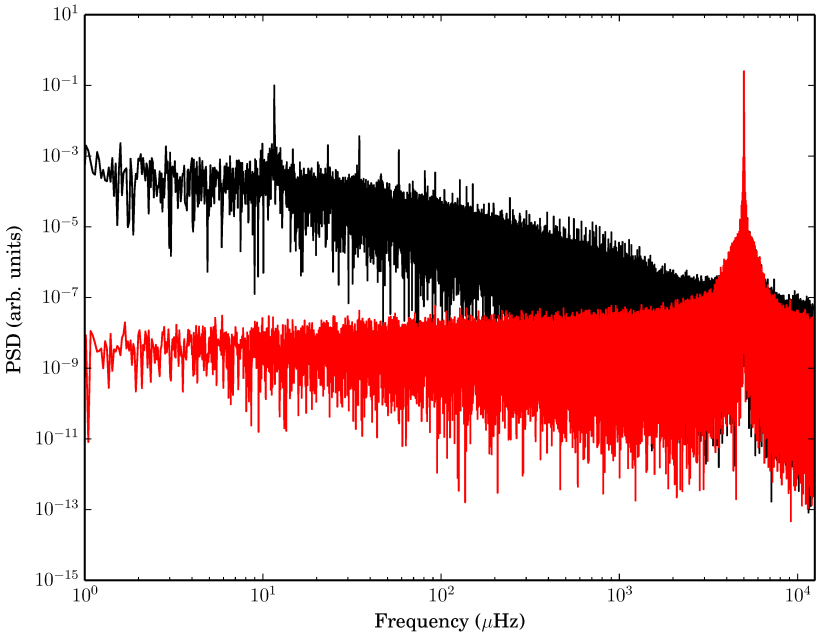


Figure 2.2: Power spectrum of a sine wave with frequency $\nu_s = 5$mHz with the BiSON window imposed in black. In red is the power spectrum of same sine wave with the BiSON window imposed but with mean subtracted from the data. The continuum arises from short timescale gaps in the window function which in turn contributes to broadband noise in the power spectrum.

it is affected.

A common technique applied to power spectral analysis is the use of tapers to improve the estimation of the power spectral density. An example is Welch's method (Welch, 1967) whereby the periodogram of subsections of the data are taken and averaged together to reduce the variance in the frequency domain. Unfortunately this would not work in our case since it reduces the frequency resolution by a factor corresponding to the number of spectra averaged over. Another option would be to window each chunk of contiguous data, ensuring that the data smoothly tends to zero where there are gaps. This will attempt to reduce the variance in the power spectrum, however in the case of many short chunks of data and small timescale gaps (such as those present in the BiSON datasets) the applications of windows offers no improvement to the data quality and so will not be used further in this work.

The same procedure detailed above was followed for the BiSON data and the mean was

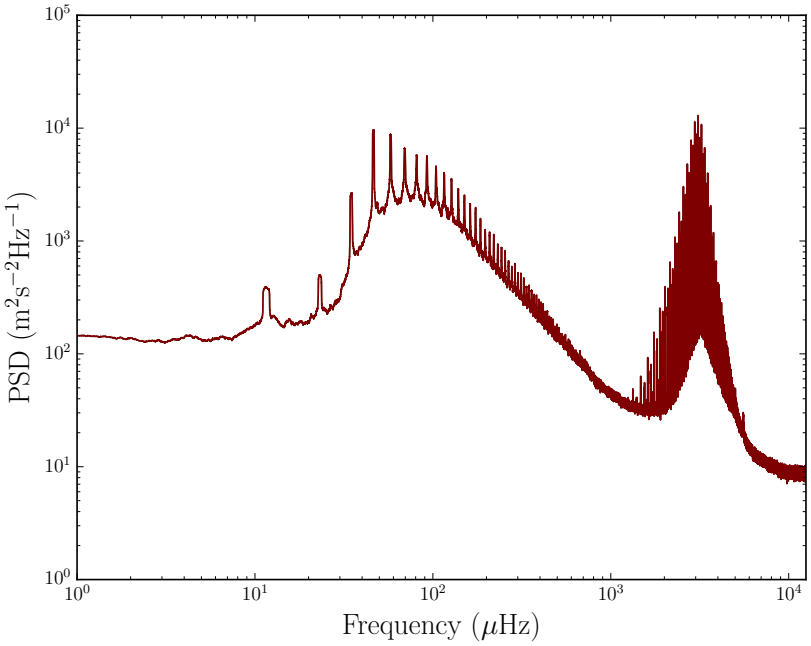


Figure 2.3: Power spectrum of the BiSON dataset (smoothed with a boxcar of width 1 μHz) having not subtracted the mean (black) and subtracted the mean (red).

subtracted and the results are shown in Fig 2.3. In this case it can be seen that the mean subtraction makes no difference to the power spectrum and most importantly the diurnal peaks are still present in the spectrum. The DC component of the power spectrum is zero and so this means that there must be something else that has a similar contribution and allows the window function to appear in the power spectrum. In order to identify where this might be, let us go back to basics and work through what each component of the power spectrum does, the details of which are given in Appendix A.1.

Equation A.14 again shows what has been previously discussed and that the second term on the right hand side contributes to the overall power spectrum because of the DC component. If the value of M was zero, as would be the case if the mean was subtracted from the time series, then this term would not be seen in the power spectrum. Therefore in order to remove the presence of the window function the DC component must be removed. A demonstration of equation A.14 can be seen in figure 2.4 for the case of the simple sine wave.

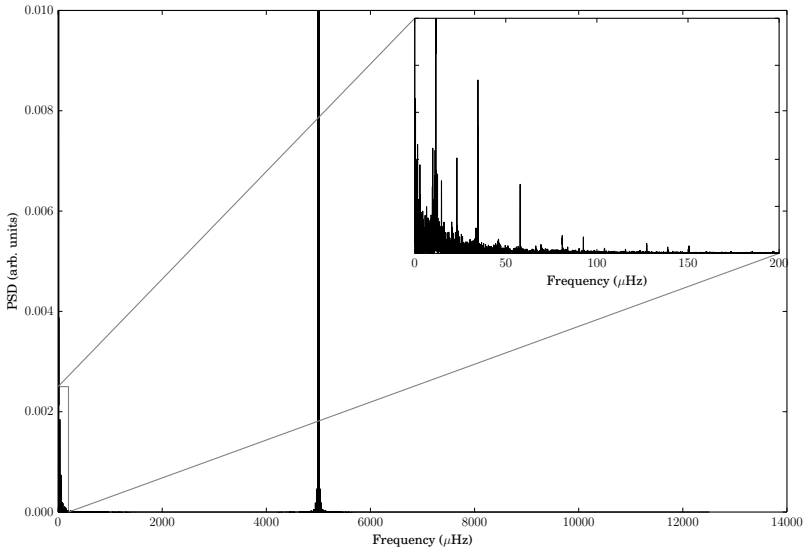


Figure 2.4: The resultant plot following the convolution of the sine wave with added mean and the BiSON window function. The main peak corresponding to the sine wave frequency at 5mHz can be clearly seen. The inset box in the top right hand corner shows the very low frequency structure caused by the convolution of the zeroth bin of the data with the window function.

This presents an interesting question as to how the peaks continue to remain in the BiSON power spectrum having subtracted the mean from the time series. For this it is necessary to go back to the sine wave example and consider what happens if there is another sine wave present in the time series with a frequency that would place it in the zeroth bin also but is not commensurate with the bin width (i.e. $\nu/\nu_{\rm bw} < 1$). A sine wave with a frequency of 0.02μHz (corresponding to a period of ≈ 1.58 years) was added to the original sine wave and so this is not commensurate with the bin width, $\nu_{\rm bw} = 0.03171\mu$Hz.

The resultant graph is shown in figure 2.5 and so this tells us that if there is any sort of signal present in the dataset with a period that is longer than the data set the diurnal structure will remain in the power spectrum even if the DC component is removed. This could be due to instrumental drifts or another process that up until this time had not previously been considered. As BiSON has been used predominantly for looking at solar p-modes, the very low-frequency variations of the instrument have not been a major concern. However at

this point it seems that even the very smallest of variations can contribute to the presence of the window function in the power spectrum.

It is common to remove long-term trends in data by subtracting a smoothed version of the data or a polynomial fit (e.g. Basri et al. 2010, 2011). This could work in our case but this would ignore any other properties of the trends themselves. If they were for example, instrumental, then they could be coherent over certain timescales or be similar day-to-day. This is additional information that will help distinguish between instrumental variation and long-term solar variation and will be covered later on in this chapter[1].

It may be possible through sine-wave fitting to extract the signals with periods greater than the length of observation however this remains to be seen and has not been done as of the time of writing. Although there can be contributions from many such signals and the ambiguity present if the period of the signals are much greater than the length of the time series is very great. A different strategy must be used to counter the presence of these signals and these strategies are discussed in section 6 and 7.

2.2 Window Function

A discussion of the basics of the window function and how it causes contamination within the data has already been made, but it is also necessary to understand the structure present in the window function and where it comes from. It is with this knowledge that a solution regarding its removal can be formulated.

At low frequency the window function is a major contributor to the overall shape and structure, as well as noise properties of the power spectrum. This is not limited to low frequency either and even at high frequencies around the p-mode oscillations the effects of the window function are seen in terms of sidebands off the modes themselves. However, for the purposes of g-mode searches the focus is on the low frequency where the effects are much

[1]There have been a number of instrument changes that have resulted in decreasing noise levels. Certainly this should be investigated with regards to the low frequency background signal.

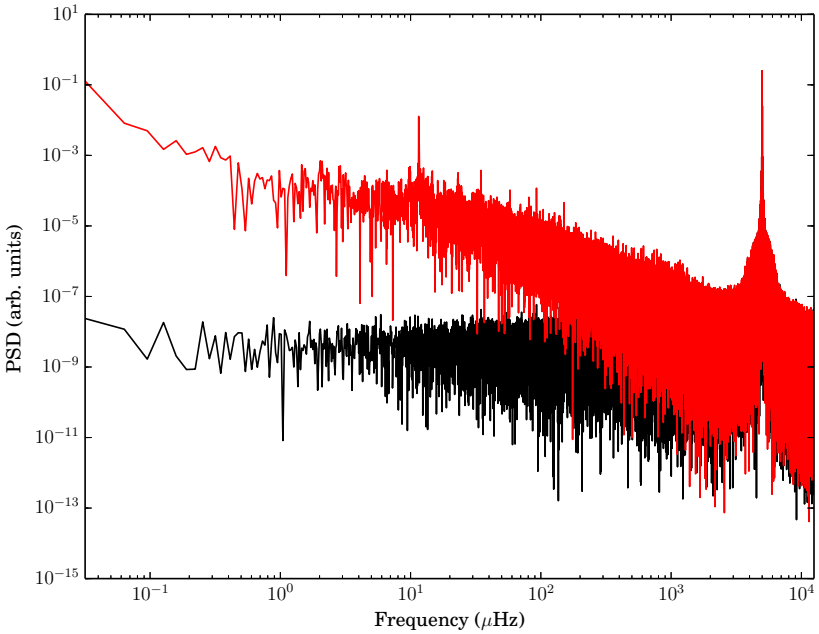


Figure 2.5: The difference between subtracting the mean when there is and is not as signal present in the data with a period longer than the length of the time series. The case without the very low frequency signal is shown in black and the case with the very low frequency signal in red.

harder to disentangle from the signal.

2.2.1 Simple Simulations

Let us start by going back to basics and looking at the very simplest ideas regarding the window function and build up towards the six station scenario that is present in BiSON. Two simple cases will be considered: the first is a year of continuous observations (i.e. no gaps in the data) and the second revolves around simulating a single station.

The window function is a function used to describe where there are and are not data, so it is 1 when there are data and 0 otherwise. The observed, gapped data can therefore be thought of as the underlying signal multiplied by this window function, this idea will be very useful later on in this chapter. Therefore, in the first case mentioned above the window function would be all 1s, whereas for the second case for a single station the window function

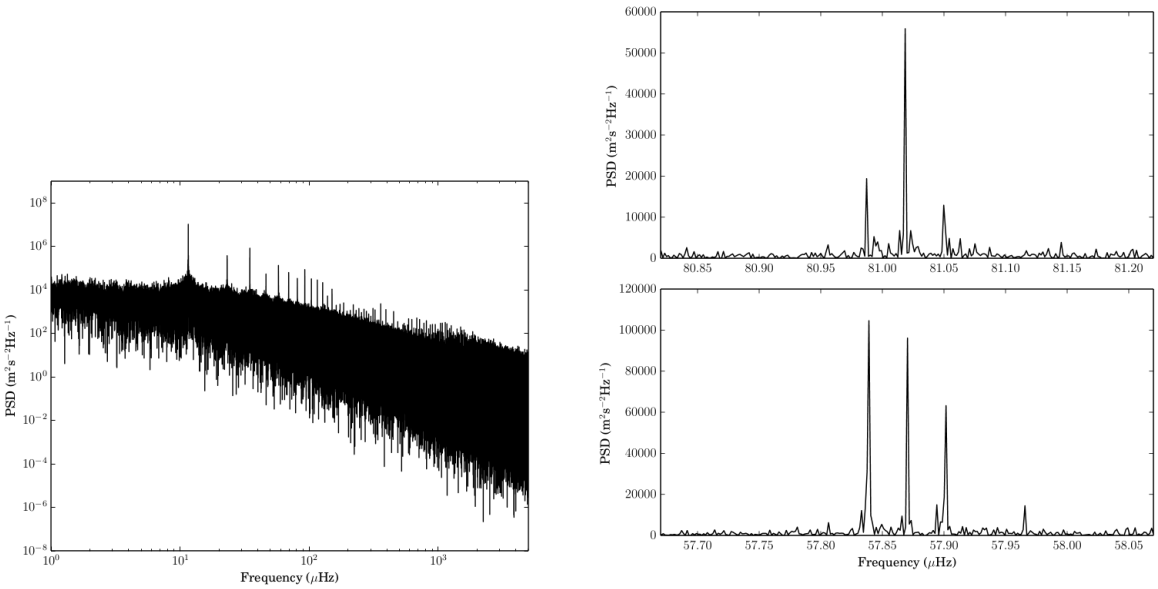


Figure 2.6: The right hand planel shows the power spectrum of the window function for 23 years of BiSON data over the range of 1 to 5000μHz. Present in the spectrum are a set of large periodic peaks characteristic of diurnal signatures and their overtones, due to gaps in the time series. The left hand panel shows a zoomed in view of the peaks present at 57.87 and 81.02μHz.

would be 1 for approximately 10 hours and 0 for approximately 14 hours (simulating the effects of day and night). The way in which the multiple different components of the window function interact with the data will contribute to a greater understanding of the BiSON data at low frequency.

In the case of the window function for continuous observations, this is simply a rectangular function with a width equal to the length of the observation.

$$\mathrm{rect}(t) = \begin{cases} 0 & \text{if } |t| > \frac{T}{2} \\ 1 & \text{if } |t| \leq \frac{T}{2} \end{cases} \tag{2.2}$$

The Fourier transform of the above function is given by

$$\mathcal{F}[\mathrm{rect}(t)] = \int_{-T/2}^{T/2} \mathrm{d}t \; \mathrm{rect}(t) \exp(-2\pi i \nu t) = \frac{T \sin(\pi \nu T)}{\pi \nu T} = \frac{\mathrm{sinc}(\pi \nu / \nu_{\mathrm{bw}})}{\nu_{\mathrm{bw}}}, \tag{2.3}$$

where the length of observation T is equal to the reciprocal of the bin-width $\nu_{\rm bw}$, i.e. the frequency resolution in the power spectrum. The limits of integration have been shifted from $0 \to T$ to $-T/2 \to T/2$ for ease of calculation, but otherwise the results are the same.

Hence the power spectrum of this simple window function is given by

$$S(\nu) = |\mathcal{F}[\mathrm{rect}(t)]^2 = \frac{\mathrm{sinc}^2(\pi\nu/\nu_{\rm bw})}{\nu_{\rm bw}^2}. \tag{2.4}$$

An important feature of Fourier transforms is the "uncertainty principle" that is essentially associated with, whereby a function is "narrow" in time then it will be "wide" in frequency space (and vice versa). This can lead us to deduce the following useful property; the length of the observation will impose a restriction on the minimum frequency resolution possible in the power spectrum (as demonstrated above). Crucially this idea also has an effect on the modes of oscillation (the details of which will be left to later chapters) where if the lifetime of the mode is greater than the length of the observing run then it will appear in the power spectrum as a sinc^2 function and be unresolved. The difference between unresolved and resolved modes is a little more subtle than given here but the important idea to note is that g-modes are theorised to have very long lifetimes, far longer than any dataset available. So if all the power is contained within a single frequency bin, the signal is much more susceptible to noise than if the power was spread over multiple bins.

In the case of a single station window function, the power spectrum is not as simple as the first case. This is due to the fact that the window function is now composed of two distinct components: the first being a rectangular function contributing to the daily component of the window (i.e. the 1s and 0s), and the second is a finite series of delta functions which give the rectangular functions the appropriate position in time. This finite series of delta functions has a spacing T of 24 hours and each delta function is situated at half way through the observing day. The result is a set of delta functions in the power spectrum (these are actually sinc^2 functions due to the finite frequency resolution) evenly-spaced by $1/T$, these

are at the positions of the so-called diurnal peaks seen in the BiSON power spectrum. The rectangular functions will contribute a slightly different type of structure in the form of a sinc^2-like envelope, modulating the amplitudes of the peaks in the power spectrum. The derivation of this function is given in Appendix A.2.

2.2.2 Multiple Stations

Added realism can be included in the form of account for the changing length of day and interruptions, this work is included in Appendix A.5. The final addition to the simulations is to account for the fact that BiSON is a network of six stations and a window function must be simulated accounting for six stations rather than just one. This ensures that the simulations are as "realistic" as possible, not including instrumental effects that are currently not well parametrised. The longitudes and latitudes of each station can also be used to ensure maximal realism, which are given in Table 2.1.

Rather than simulating the combined window function at once, it makes much more sense to simulate each individual station and then combine them all at the end. This is a simple procedure and involves setting the window function to 1 anywhere it is not zero. The combined window function simulated for the six stations can be seen in Fig 2.7. The broad structure of the BiSON window function can be reproduced in terms of the diurnal peaks and the general power law decrease in the noise as a function of frequency. However, in the artificial simulation there is a "knee" observed (at ~ 200 μHz) before which the power spectrum is approximately flat[2], a feature not seen in the real case. Also the base of the diurnal peaks are considerably wider than the simulated case indicating that there is more power in the peaks than can be reproduced simply from our method.

These differences are to be expected, the gap distributions should ideally be tailored to each individual station rather than using the same distribution for each. But even such

[2]The reason for this is not currently known.

Station	Latitude (°)	Longitude (°)
Mount Wilson	34.224156N	118.05851W
Las Campanas	29.021595S	70.689876W
Izana	28.300158N	16.512065W
Sutherland	32.38038S	20.810525E
Carnarvon	24.86723S	113.70375E
Narrabri	30.309838S	149.565296E

Table 2.1: Longitude and latitude values of BiSON stations

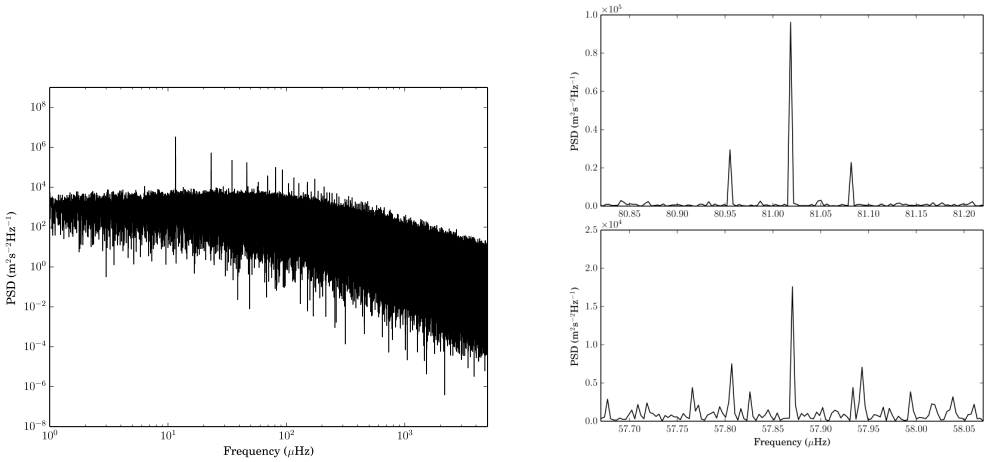


Figure 2.7: Artificial window function power spectrum simulated including the addition of changing length of day, interruptions and multiple stations. The addition of noise is the most striking influence the interruptions possesses. The left hand plot shows the entire power spectrum of the artificial window function and the left hand plot shows the peaks at 81 μHz and 57.8 μHz.

simple added complexity gives some insight as to what these peaks are due to.

2.3 Modelling the BiSON Background Limit Spectrum

It has been shown in 2.1.2 that a non-negligible DC component can contribute to structure in the power spectrum, whether this is from a non-zero mean or a period longer than the length of the time series. The mean of the data is always subtracted before the power spectrum is constructed so the former is not applicable, however a very long period signal could be present. This could be due to instrumental effects or any other such signal that could be present in the data over such a long timescale. Therefore the idea would be to model this effect in the power spectrum by convolving the window function with a model of the BiSON

background along with a flexible zeroth bin (the limit spectrum) which would account for this very long time-scale signal. The convolution of the zeroth bin with the diurnal peaks will resolve the diurnal peaks in the correct place and hopefully with a similar set of amplitudes. The purpose of the limit spectrum would be to encapsulate all of the available knowledge about the noise structure present within the data. Such noise would include granulation, solar activity and importantly at low frequency, the effect of the calibration on the data.

There had also been testing using high-pass filtering to remove the long-term trend, however this removed far too much of the granulation signal. In addition, find a filter that has a sharp enough cut-off is very difficult and it did not improve our data as we would have hoped. Therefore we have instead resorted to modelling the background.

2.3.1 Description of Model

The model itself consists of a few different components, the first that will be explained is granulation and "activity". The effects of active regions, sunspots and any process relating to long-period variability in the Sun is gathered under the catch all term of "activity", whether this is best or not it provides a reasonable approximation to the underlying process. The implicit assumption here is that these features can be modelled together using a Harvey-profile (Harvey, 1985). It is assumed that the contributions from granulation and activity can be described by random fluctuations in some quantity $x(t)$ which have some variance $\langle \delta x(t)^2 \rangle = \sigma^2$. The auto-covariance can then be described in terms of a single-sided exponentially decaying function with a characteristic timescale τ given by $x(t) = \exp(-t/|\tau|)$, where τ corresponds to the relaxation time of the granulation components. The power spectrum of such a signal is given by

$$P(\nu) = \frac{4\sigma^2\tau}{1 + (2\pi\nu\tau)^2}, \tag{2.5}$$

which is simply a zero-centred Lorentzian. This model however does not give the required

flexibility needed to fit the BiSON data, partly because it is unknown whether the low-frequency region is dominated by granulation of instrumental effects. Instead the following model was adopted

$$P(\nu) = \sum_{i=0}^{2} \frac{4\sigma_i^2 \tau_i}{1 + (2\pi\tau_i\nu)^{c_i}}, \tag{2.6}$$

where the summation is made over 3 granulation components, σ is the rms velocity defined previously, τ is the characteristic timescale of each component, and c is the exponent.

Due to the fact that the exponent value is difficult to determine in many solar-like stars it is normally treated as a free parameter (e.g. Handberg & Campante 2011). Therefore the value of c was allowed to vary in order to provide the best fit to the data. This does bring into question the process behind the granulation signal if the exponent is not 2 or 4 (the processes behind those examples are discussed later), and so it is used more to provide a better find as opposed to investigating the underlying processes.

By using the variant of the Harvey-profile shown in equation 2.6 and a variable exponent implicit assumptions are being made about the nature of the granulation signal. An exponent of 2 (as shown in equation 2.5) describes a signal with an autocovariance that is exponentially decaying (as explained earlier). Values of the exponent that differ from 2 start to become more difficult to explain in terms of physical processes, but there are some exponents that can be explained (such as 4) (e.g. Hekker et al. 2011 or Kallinger et al. 2014).

As has already been mentioned previously, there is a large degeneracy in the granulation model at very low frequency ($\nu \leq 80\mu$Hz) due to the effect of the calibration. There is, therefore, some degree of uncertainty in terms of visuals in the parameters of the granulation components at very low frequency. In essence it becomes harder to determine whether the fit is a good one or not due to there being no uncorrupted data present below 80μHz and so the granulation components fitted can only be used to improve the shape of the limit as opposed to drawing meaningful inferences about granulation from the parameters.

In addition to the granulation components there is a very clear turn-over at very low-frequency due to the calibration of the data[3]. As a result this needs to be included in the model and so two filters were used in order to try and incorporate this feature. The first is a standard high-pass filter

$$r_{\rm s}(\nu) = 1 - \mathrm{sinc}^2 \nu_{\rm s}, \tag{2.7}$$

where $\nu_{\rm s}$ is the characteristic frequency of the filter. Due to the nature of the filter response there is extensive ringing in the passband around unity which is undesirable due to the want to further minimise any artefacts put into the data. Also, the high-pass filter is not steep enough to give the required turnover seen in the spectrum and so another filter was added to accomplish this. The second filter took the form of a Butterworth filter (Butterworth, 1930)

$$r_{\rm b}(\nu) = \frac{G_0^2}{1 + (\nu/\nu_b)^{2n}}, \tag{2.8}$$

where G_0 is the gain at zero frequency, n is the order of the filter and $\nu_{\rm b}$ is the cutoff frequency.

The two response functions are added together and the combined response is set equal to one wherever it is greater than or equal to unity. The response is then smoothed due to the fact that there is a kink present where the two responses are combined, and finally the region that is spoiled by the smoothing (at very low frequency) is replaced with the response before the smoothing (this does not extend as far as the region where the kink is). The final frequency response function will be denoted by $r(\nu)$. This method of combining the filter responses is cumbersome but performs adequately. A better method might involve using a tanh function for example to smoothly combine the signals.

Finally there are two extra components to add in, the white noise level and a pink noise

[3]The use of the third-order polynomial in the calibration and subsequent division by the sensitivity causes this feature, see Elsworth et al. (1995) for more details.

component. This second component is empirical in nature as it provides a good fit to both high and low frequency. The physical reasoning is linked to the combination of the individual station data which will lead to discontinuities being put into the data, giving rise to a pink noise-like component. This is modelled as follows

$$n_{\rm p}(\nu) = \frac{a}{\nu^{\alpha}}, \tag{2.9}$$

where a is a constant and α is the exponent. The reason for there being pink noise present is thought to be as a result of the concatenation process used to combine multi-site data sets (see Davies et al. (2014b) for more details on the calibration and concatenation process). It is thought that during the concatenation process drifts present in the individual station time series are combined and so there will be slight discontinuities present in these long-period drifts. As a result this will lead to noise in the power spectrum whose gradient should be larger the longer the time series, this idea will be returned to later in the chapter.

2.3.2 Final Model and Inclusion of the Window Function

Having discussed the individual component it is necessary to combine all of them to create the final model of the limit spectrum. This model is defined as such

$$L(\nu) = r(\nu)\left[\sum_{i=0}^{2}\frac{4\sigma_i^2\tau_i}{1+(2\pi\tau_i\nu)^{c_i}}\right] + n_{\rm p}(\nu) + n_{\rm w}(\nu). \tag{2.10}$$

Two of the granulation components will have timescales below the turnover seen in the spectrum. They are still needed since there is still power present before the calibration is applied so in order to best summarise the effect of the calibration a realistic model is needed.

The model of the limit spectrum can be seen to be a good fit of the data (using maximum likelihood estimation, MLE) in the low-frequency regime as shown in Fig 2.8. Although it is clear that the presence of the diurnal peaks does alter the shape of the background and so

Limit Components	τ (s)	σ (ms^{-1})	c
"Granulation"	934.85	0.449	2.3
"Meso/Supergranulation"	2.25×10^4	2.5	2.3
"Active Regions"	1.0×10^6	0.274	1.626
Noise Components	a	α	-
Pink Noise	2.5	0.3	-
Frequency Response	G_0	ν_c (μHz)	n
Butterworth	1.0	6.5×10^1	3.0
Frequency Response	G_0	ν_h (μHz)	n
High-pass	-	5.0×10^2	-

Table 2.2: Parameter values for the limit spectrum constructed in figure 2.8. The labels corresponding to the various granulation components are to be used as a guide for the approximate timescales to which they correspond.

they should be included in any subsequent modelling (the parameters of the fit are shown in Table 2.2).

The next course of action is to include the window function in the limit spectrum, such that the power spectrum can be divided through by this new limit (including the window function) and the result will, hopefully, be white noise plus any signal left over (i.e. g-modes).

The question is, what operation allows us to include the window function in the computed limit spectrum? The answer to which is the convolution operation. If the limit spectrum is convolved with the window function then we would hope to see the peaks in the window function appear in the correct place in the convolved product, however it is not quite this simple.

The diurnal peaks in the power spectrum appear in exactly the same position in frequency as the diurnal peaks in the window function, this suggests that the dominant factor in the convolution between the data and the window is the bin at zero frequency as has already been demonstrated earlier. Therefore the only addition to the limit spectrum model is that the value of the zeroth bin must be set to such a value that when it is convolved with the window function, the peaks appear at approximately the correct amplitude. Obviously there

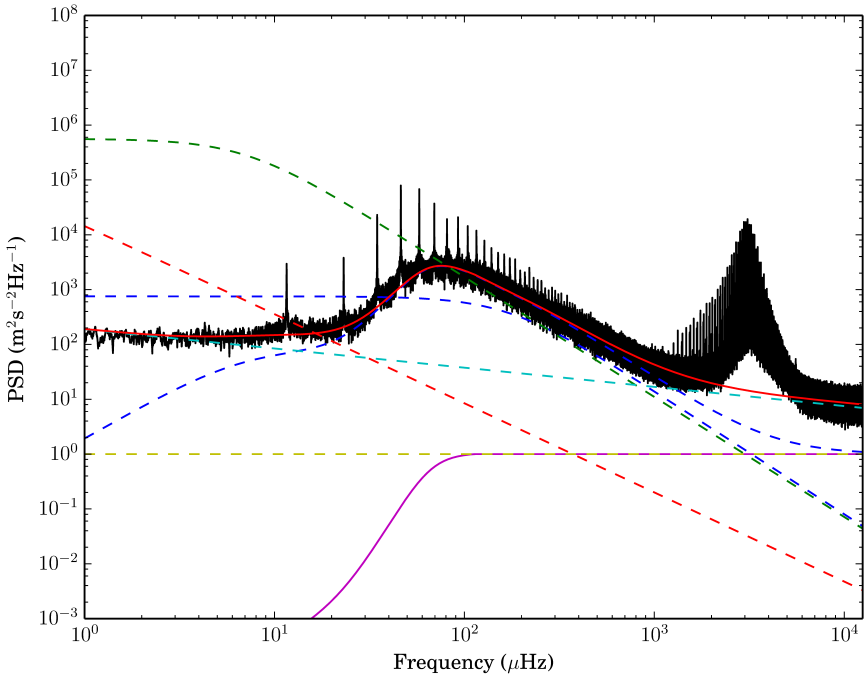


Figure 2.8: BiSON limit spectrum constructed using the 23 year BiSON data. The full model is shown in the solid red line, the granulation components are shown in the blue, green and red dotted lines and the sum of the granulation components multiplied by the frequency response (shown in pink) is shown in the second blue dotted line. The white noise and pink noise are the yellow and turquoise dotted lines respectively.

is not going to be an exact correspondence because there is no noise in the limit spectrum whereas there would be the usual χ^2_2 noise present in the data.

Before the convolution is performed it remains to be seen how the model is affected by the convolution, in other words, what components are involved in the convolution. The granulation noise must be included in the convolution as it is part of the signal that is observed and so the window function will be imposed upon it, however in the case of the frequency response and the pink noise they both should not be included in the convolution process. The reasoning behind this is that these processes happen after the data has already had the window imposed upon it and therefore should not be included in the convolution. Also in the case of the pink noise, it is a result of the combination of the data from individual stations and so the data has already been taken before the pink noise inadvertently ends up

in the data. The frequency response instead "shapes" the spectrum after the convolution takes place.

Therefore our final model including the window function is

$$L_{\text{final}}(\nu) = r(\nu)\left[\sum_{i=0}^{2}\frac{4\sigma_i^2\tau_i}{1+(2\pi\tau_i\nu)^{c_i}} + n_{\text{w}}(\nu)\right] \otimes |W(\nu)|^2 + n_{\text{p}}(\nu). \tag{2.11}$$

The resultant fit is shown in Fig 2.9 and it should be clear that it is not too bad in terms of getting the broad stucture and the peaks in the correct position. Beyond that the fit does not do a very good job since the peak amplitudes are very different and the hump caused by the calibration is difficult to produce. The parameters of the fit are not included here since this is simply a demonstration of how this method cannot work to remove the diurnal peaks. In addition the value of the zeroth bin would have to be so large in order to produce noticeable peaks in the power spectrum that a very obvious shift would be present in the time domain.

As a result it is clear that our understanding of what is causing the peaks is lacking and that a different path should be taken to try and remove these peaks, possibly in a non-parametric way that enables the data to inform us rather than making too many assumptions.

2.4 Understanding the Cause of Diurnal Peaks

It is clear from the previous sections that the underlying cause of the diurnal peaks is not well understood. As a result of this it is advisable to test ideas using simulated data to see if through these the diurnal peaks can be made to appear, for example through inserting long period signals.

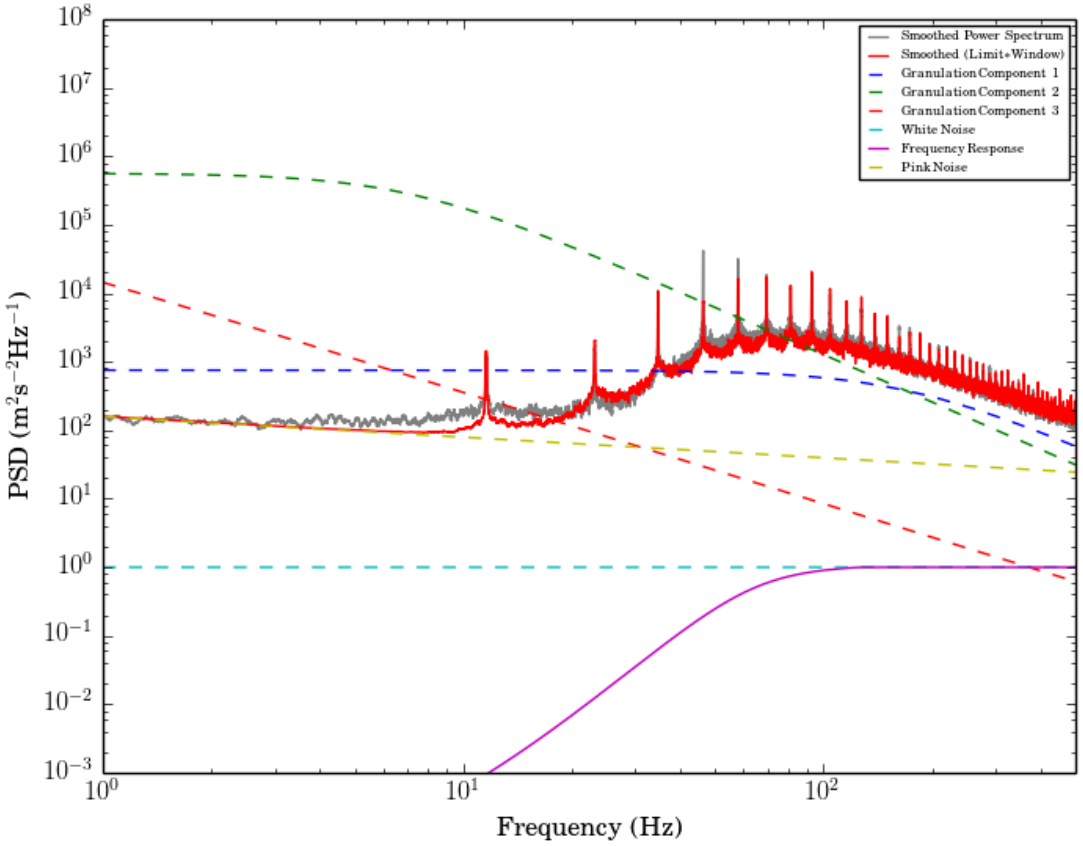


Figure 2.9: Power spectrum of BiSON data smoothed with a 28 point moving mean filter shown in grey and the limit spectrum convolved with the window function power spectrum is shown in red.

2.4.1 SolarFLAG Simulations

The data used in this section came from SolarFLAG (Chaplin et al., 2006)[4], from which we have an artificial solar dataset to use. Using realistic artificial data provides the best opportunity to try and create the diurnal peaks in the power spectrum and understand their cause.

An important component of determining the cause of the diurnal peaks is to inspect the time series from different stations over similar time periods. In other words the coherency of

[4]The data is produced in the time domain making use of theoretical and observed parameters to make the artificial data as realistic as possible.

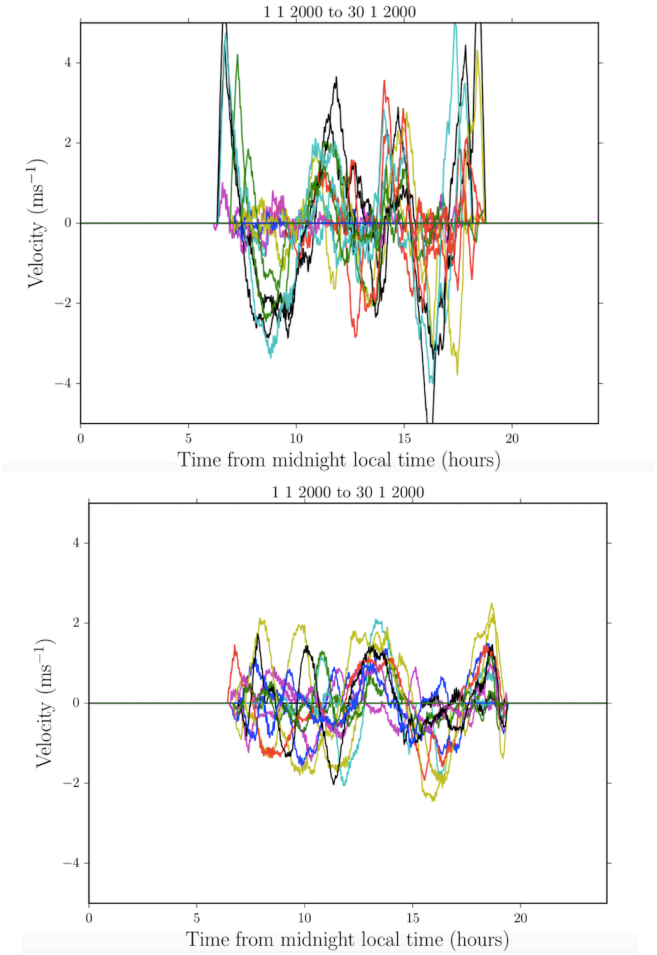


Figure 2.10: An example of contemporaneous data taken from two different BiSON stations. The title shows the time period over which the data was taken and each coloured trace denotes a different day smoothed with a box car of width 20 bins. The top plot is data taken from the station in Carnarvon and the bottom plot is data taken from the station at Las Campanas.

any long term trends needs to be investigated. If they were coherent over long time-scales in multiple stations then one may be tempted to associate them with a solar origin. But if they are seen in one station but not another which is observing at the same time then any large trends would be instrumental in origin.

2.4.2 Footprint

The basic idea that can be gathered from the previous sections is that whatever source is contributing to the presence of the diurnal peaks, it must be some sort of low-frequency quasi-

periodic signal. It is known that some of the BiSON stations that use equatorial mounts for the instruments exhibit low frequency oscillations in the time series (known as a footprint) which Chaplin et al. (2002) attempted to model as follows

$$s(t) = \begin{cases} s_{\rm am}(t) & \text{for } t \leq \text{ local noon,} \\ s_{\rm pm}(t) & \text{for } t \geq \text{ local noon,} \end{cases} \tag{2.12}$$

where

$$\begin{aligned} s_{\rm am}(t) = s_1 \cos\left\{ 2\pi\nu_1 \exp\left[-\frac{(t - t_{\rm start})}{\tau_{\rm am}} \right] + \phi_1 \right\} \\ + s_2 \cos\left(2\pi\nu_2 \sin\omega_{\rm HA} + \phi_2 \right), \end{aligned} \tag{2.13}$$

and

$$\begin{aligned} s_{\rm pm}(t) = s_3 \cos\left\{ 2\pi\nu_3 \exp\left[-\frac{(t_{\rm end} - t)}{\tau_{\rm pm}} \right] + \phi_3 \right\} \\ + s_2 \cos\left(2\pi\nu_2 \sin\omega_{\rm HA} + \phi_2 \right). \end{aligned} \tag{2.14}$$

The components are characterised by amplitudes s_1, s_2, s_3, frequencies ν_1, ν_2, nu_3 and phases ϕ_1, ϕ_2, ϕ_3. In addition the exponentially decaying functions are described by time constants $\tau_{\rm am}$ and $\tau_{\rm pm}$ along with two parameters that signal the beginning and end of the day, $t_{\rm start}$ and $t_{\rm end}$.

A few examples of the footprint are shown in Fig 2.11, the almost "chirp" like signal of increasing frequency is a result of the cos(exp) term in each component. Certainly this is a trend that would be difficult to fit out for a number of reasons, mainly due to the nature of granulation which is also shown in Fig 2.11. When there is no granulation signal present, the data that is observed can be reproduced by fitting the footprint model to the data. Unfortunately this is not going to be the case since granulation will be present on timescales similar to that of the trend itself and so will cause deviations from the expected model.

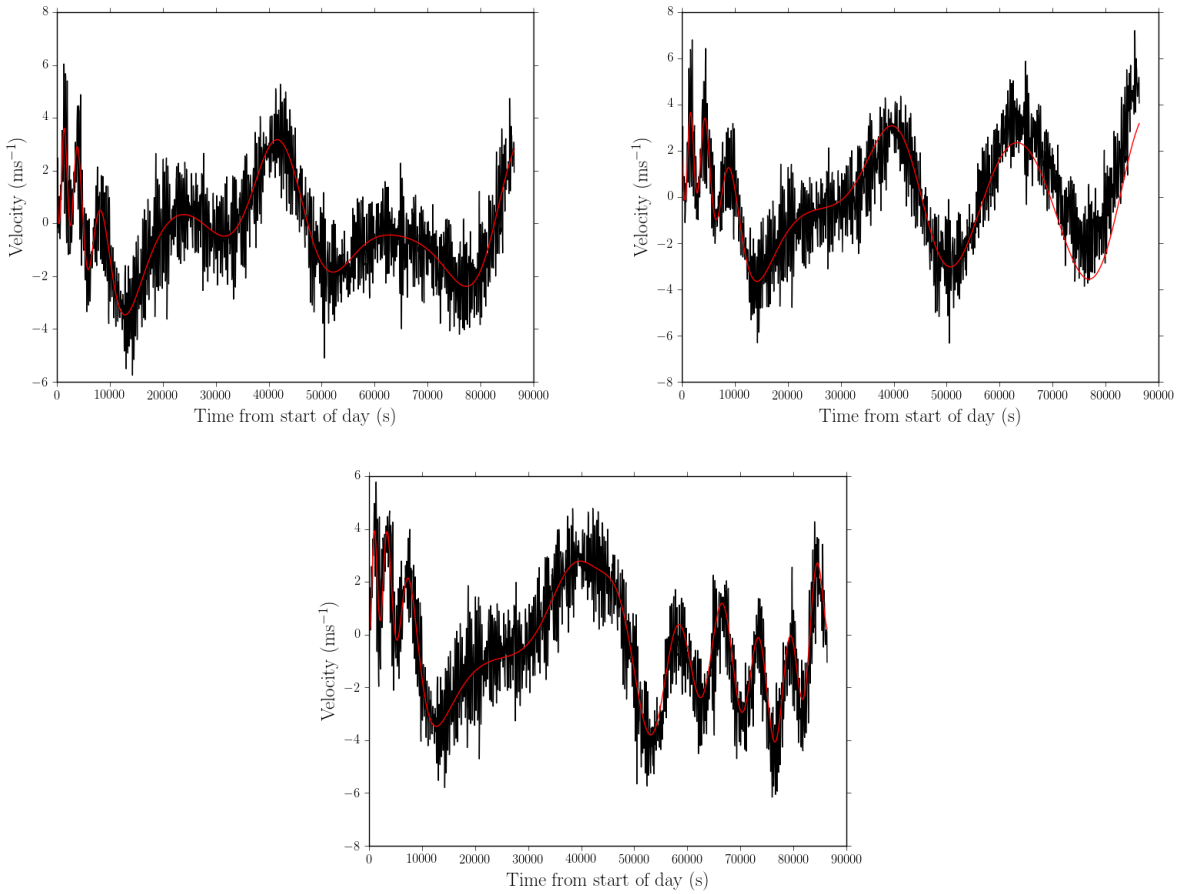


Figure 2.11: Examples of the footprint applied to 1 day of SolarFLAG data, each with a different long-term granulation signal added where the underlying trend (caused by the footprint) is shown in red. The top left plot has no long-term signal added, whereas the top right plot contains a granulation signal with a characteristic timescale of 1 day and the bottom plot contains granulation with a characteristic timescale of 30 days.

This can be clearly seen in Fig 2.11 when the granulation timescale is approximately 1 day, roughly in accordance with a supergranulation-like signal. Therefore it may not be wise to fit the data on a daily basis since the effect of granulation will be large, it would be better to fit an average over a length of time that should average out the granulation signal.

The complicated nature of the footprint will make a simple least-squares fit to the data very difficult and prone to error and so instead a non-parametric approach will be pursued using principal component analysis (PCA).

Table 2.3: Parameters for the construction of the footprint.

Parameter	Value
s_1 (ms^{-1})	1.76
s_2 (ms^{-1})	1.93
s_3 (ms^{-1})	1.73
ϕ_1	0.4
ϕ_2	0.4
ϕ_3	0.4
ν_1 (hours^{-1})	3.51
ν_2 (hours^{-1})	0.72
ν_3 (hours^{-1})	4.26
τ_{am} (hours)	1.85
τ_{pm} (hours)	20.87

2.4.3 Effect on the power spectrum

Before the footprint is removed it is important to understand its manifestation in the power spectrum to see if it is the feature that is causing the diurnal peaks to appear in the power spectrum. Having ruled out a signal with a period longer than the length of the timeseries, due to the fact that it would be visible in the timeseries due to an unphysically large amplitude, it seems like a good idea to test the footprint as a possible cause.

In order to simulate this the solarFLAG data had a supergranulation component added (with a characteristic timescale of 1 day) which is not included in the original data and then the footprint was applied to every day of data, shown in Fig 2.12. On top of this the window function for 10 years of data taken from Izaña was applied in order to make the data more realistic. The footprint inserted into the data can be seen in the left hand panel of Fig 2.12. Obviously applying the footprint with the same parameters every day is not realistic. However it can be seen in the power spectrum created from this data that the diurnal peaks occur, although this is greatly exaggerated compared to the real data. This does hammer home the point that the footprint is causing the diurnal peaks to appear in the power spectrum.

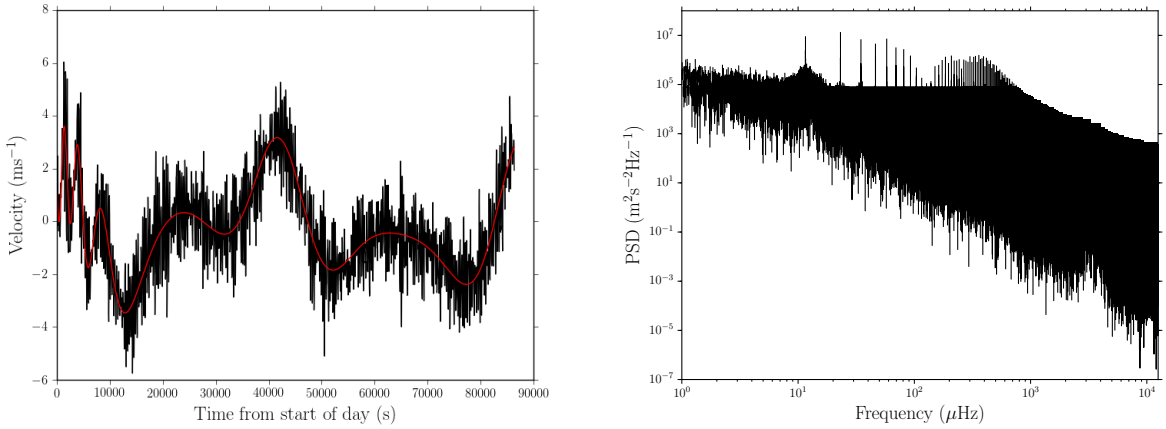


Figure 2.12: Examples of the footprint applied to 1 day of SolarFLAG data, each with a different long-term granulation signal added where the underlying trend (caused by the footprint) is shown in red. The top left plot has no long-term signal added, whereas the top right plot contains a granulation signal with a characteristic timescale of 1 day and the bottom plot contains granulation with a characteristic timescale of 30 days.

Looking back at the BiSON power spectrum clearly the footprint is not present everyday or in every station and so this would lead to a reduced effect which would certainly lead to the diurnal peaks having a decreased amplitude at higher frequency so that they do not swamp the oscillations. As a result a method is needed to remove the footprints to see if this removes the diurnal peaks from the power spectrum of the real data.

2.5 Improved Removal of Diurnal Peaks

2.5.1 Principal Component Analysis (PCA)

Due to the nature of the data taken with BiSON there will be larger trends in the data that are not necessarily solar in origin. As shown in the previous section these are coherent over timescales of a week and could be explained by the footprint issue as a result of the equatorial mount used in some stations (Chaplin et al., 2002). A method is therefore needed that can remove these coherent trends and do so in a robust and fast way.

As the title of this subsection suggests principal component analysis (for a good overview see Jolliffe 1986) has been chosen to perform this job. The reason for this is that the purpose

of PCA is to identify patterns in data and is used as a form of dimensionality reduction in larger datasets whilst minimising the loss of information. In other words, given a dataset, the job of PCA is to find a projection of the data such that the variance is maximising along the new axes.

This is ideal for our purpose whereby 7 days[5] of data will be considered at a time from a single station. This is provided in the form of a $N \times K$ matrix, where N is the number of data points in one day of data and K is the number of days in the dataset (i.e. 7 in this case). Therefore a search can be made for the trend in the data that contributes most to the variance of the given dataset and remove it, thereby removing additional instrumental effects and ideally lowering our low frequency noise level. It is also common to normalise data when performing PCA such that it has zero mean and unit variance, however in this case that would be far from ideal. When trying to find and characterise trends that are not guaranteed to be present in all days this normalisation would completely ruin the data, as data without the trends would be brought to the same amplitude as data with the trend present.

The method that was adopted will be applied to 7 day chunks of data. The reasoning is that over this range no coherent solar signal should be seen and so any trends seen over multiple days should be instrumental in origin[6]. Also, 7 days is enough to identify and model the trend without it being smoothed out by taking too many days (if the trend is not coherent over such long timescales). The data is also filtered (with a boxcar of width 800s) before the PCA is applied in order to suppress the high frequency variations and ensure that only the low-frequency trends are modelled. Once PCA is applied the primary component is taken to be the trend since it will contain the most variance. This trend component is then subtracted

[5]Since this idea is a proof of concept, the 7 day period itself is rather arbitrary and chosen as it appears to be a reasonable amount of time over which the instrumental trends should be coherent. Certainly this is something that could be tweaked and optimised in the future.

[6]This could affect the g-modes since they have periods on this timescale, however the assumption is that the g-modes are of low enough amplitude that they should be unaffected. This is an idea that needs to be looked into in the future.

from each day of data separately and a check is made to ensure that the variance is reduced in each day of data in the subset. If the variance is reduced then clearly the trend has been reduced and so this is ideal, however if the variance is not reduced then the assumption is made that the footprint is not present in that day of data and so it is left alone.

The length of data making up each subset could be different for each station and change as a function of time, however this would require much more investigation and so to prove the concept the 7 days chunks were kept.

An example of the PCA being applied to the data is given in Fig 2.13 for the BiSON station Sutherland. The rough trends seen in the data and their almost coherent nature are quite obvious and thanks to applying the PCA the majority of the trends have been removed. This is by no means perfect and additional modifications could be made to improve the method, but as a proof of concept this is a step in the right direction.

Differential extinction (the effect of atmospheric refraction across the solar disc, which is largest at the beginning and end of the day) can have an effect on the PCA either directly from the fit in the calibration process being caught out by other features or by it not being taken out completely. This would commonly be observed as a large increase or decrease in velocity at the beginning and end of the day. The amount of data used in the fitting process can also cause artefacts in the residuals which the PCA is applied to. "Flipping" of the residuals can occur due to the instability of the fit to the differential extinction and this will cause the PCA some issues. If, for example the data contains 6 days of data which show the same coherent trend and then 1 day shows the opposite because of "flipping", the PCA might approximate the coherent trend fairly well but the trend will not be removed from the opposite case since the variance will not be reduced by doing so. As a result a trend caused by the footprint can still be left in the data therefore contributing to the presence of the diurnal peaks. This is not too much of an issue because the majority of the peaks have been greatly reduced in amplitude. Certainly there is some evidence to suggest that a small

amount of power is being input into the very low-frequency part of the power spectrum, but this will not effect the area of interest. An example of this "flipping" is shown in Fig 2.14, however this is not as large an effect as observed in some of the data and so it might not simply be a calibration issue but something for fundamental.

A small amount of power will also be removed from the low frequency region of the power spectrum as a result of the granulation signal. This is caused by the granulation signal causing slight deviations from the footprint-like signal, as a result some granulation signal will also be removed when PCA is applied. This could be mitigated by running the PCA with larger subsets of the data, thereby smoothing out the effect of granulation however this has yet to be tested.

There are a few useful metrics that can be used to assess the performance of the PCA: the explained variance ratio gives information about how much of the variance is contained within the trend from the PCA and the fraction of days that PCA is applied to the data.

The explained variance ratio gives an idea as to how much variance the coherent trends contribute to the data. This can be used as a diagnostic for the individual stations themselves since it gives information as to how influential the trends can be in the data. The only stations thought to be affected by the footprint were Carnarvon (A& B), Sutherland and Narrabri, however it is clear that this is not the case. Certainly Izaña, which is the most stable (and oldest) station requires the least input and has the lowest contribution from trends as expected. But all the other stations require intervention from the PCA and around 45% of the variance is removed as a result of coherent trends. This is surprising as it shows that there are other coherent trends in the data that are not a result of the mount used. It also shows the benefits of using a non-parametric approach over fitting out the footprint,

An alternative to using the PCA to remove the trends could involve the use of Gaussian Processes (GPs) (Rasmussen & Williams, 2006). This would enable the oscillations to be modelled as a distinct noise process alongside accounting for the trends to be modelled using

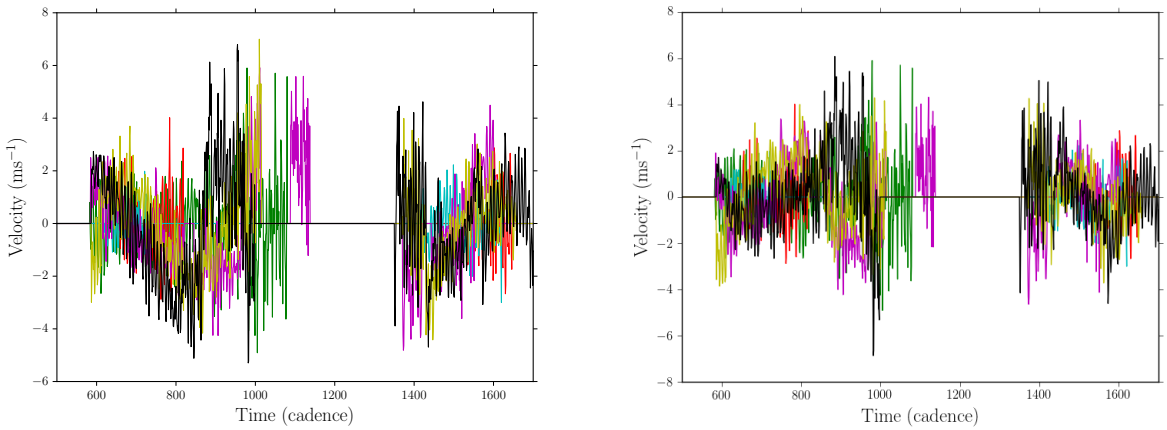


Figure 2.13: Examples of 7 days of data taken from the BiSON station at Sutherland before (left hand plot) and after (right hand plot) the PCA has been applied.

a different kernel, for example a polynomial. In addition the g-modes could also be included by fitting for a given set of sinusoids such that the properties of the g-modes would remain as unaffected from the analysis as possible. This would be similar to ideas already used for detrending *Kepler* and K2 data (Luger et al., 2016; Aigrain et al., 2016).

Once the PCA has been applied to the data the hope is that all (or at the very least most) of the diurnal peaks will have been removed and examples of the data before and after are shown in Fig 2.15 for Mount Wilson and Izaña. The diurnal peaks have been very much suppressed beyond a frequency of $\sim$70 μHz which is a great improvement. Some additional very low frequency noise is added into the data for Mount Wilson, but this is less of a worry since the method clearly does the intended job. Given that the method works on individual stations it would be best to combine the data and compare the full datasets before and after the PCA is applied.

2.5.2 Combining Individual Station Data

Simply performing the principal component analysis on a single station timeseries is useful but ideally the combined multi-station dataset is wanted. Normally the station data is combined on a daily basis, making use of precise timings of each station to know when stations are

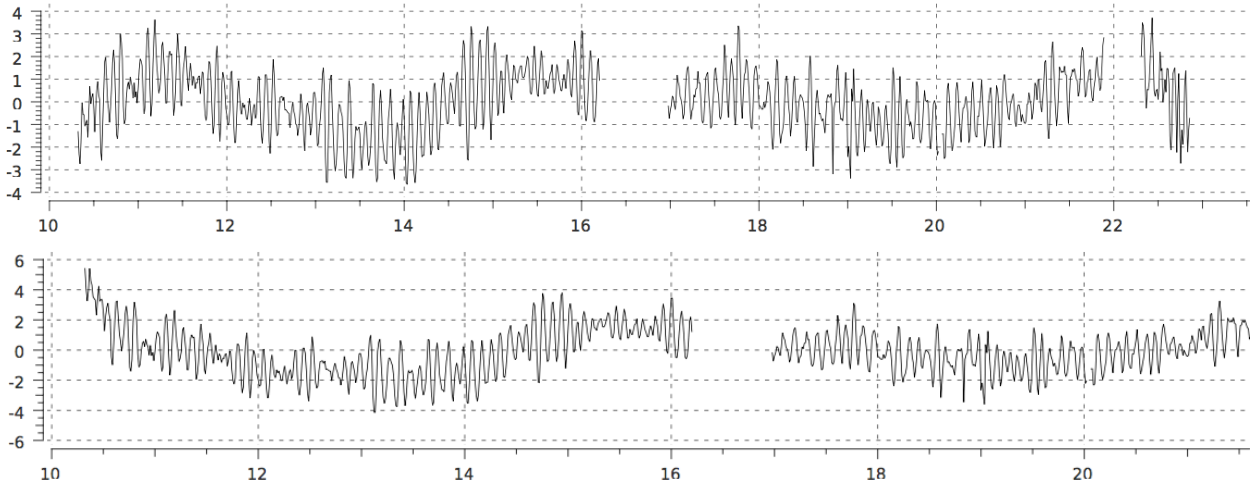


Figure 2.14: Two plots showing the difference in the velocity residuals as a function of time for one day of data taken at Las Campanas caused by masking out different regions of data. The top plot is created using the standard regions where bad data is masked whereas the bottom plot has an additional region at the end of the day masked out where the differential extinction fit was not as good. (Figures courtesy of Steve Hale).

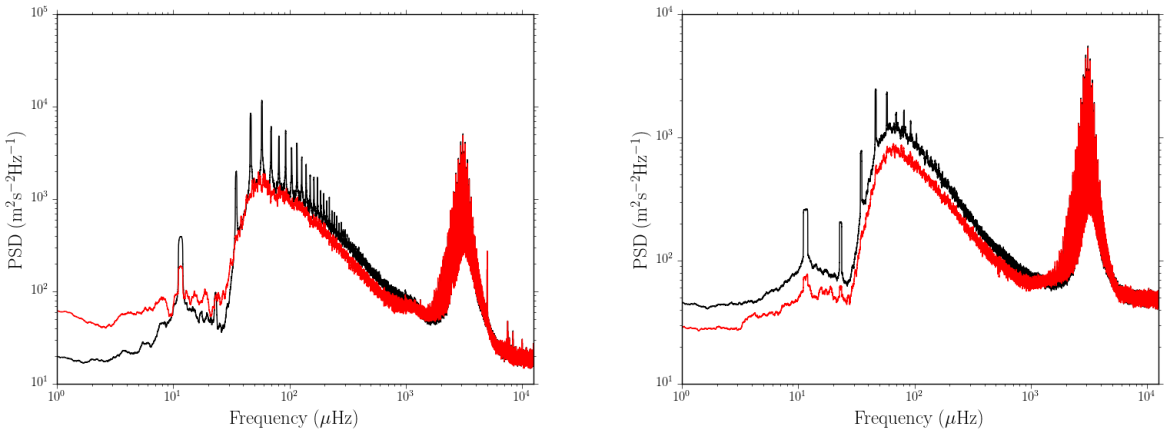


Figure 2.15: Two examples of power spectra (smoothed with a boxcar of width 1 μHz) constructed from data before (black) and after (red) the PCA has been applied. The right hand plot is for the BiSON station at Mount Wilson, and the left hand plot is for Izaña.

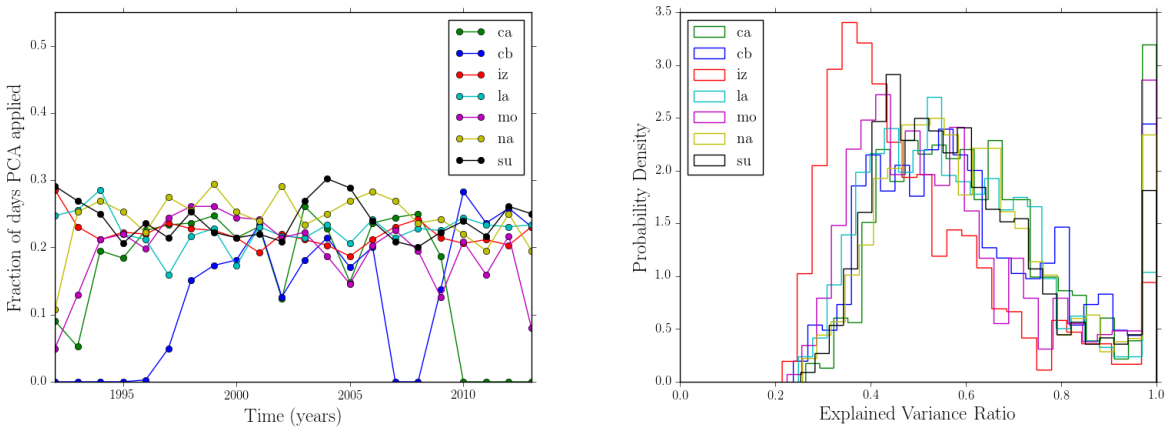


Figure 2.16: The fraction of days PCA is applied to each BiSON station is shown in the left hand plot, any zero points are a result of data not being taken with that station at that particular time. On the right hand plot is the explained variance ratio for each station. Any time spent at 1 shows that only 1 day of data was present for the 7 days analysed.

overlapping and how to combine them properly.

The procedure here is a bit simpler and can avoid the need to combine stations a day at a time. The PCA is applied to a year of data for each station and then the data for each year is combined separately (each combined year can then be stitched together later).

To begin with the overlaps between each station are found and then the overlaps are enumerated over recording the duration of the overlap along with the start and end times. The overlapping segments are combined according to a Figure-of-Merit (FoM) defined as the inverse of the integrated power contained at low frequency

$$w = 1/\int_{\nu_1}^{\nu_2} P \mathrm{d}\nu, \tag{2.15}$$

where the integral is taken from 0 to 500 μHz. This results in a weighting that is smaller if the low-frequency noise is lower and vice versa in an attempt to reduce the low frequency noise (this is also another reason why there is a difference between the real data at low-frequency). The FoM can only be calculated for overlapping segments longer than 125 points (due to the low frequency resolution of power spectra with such few points) such that there

are a few frequency bins to integrate over to derive the FoM. If the overlapping segment is less than 125 data points the inverse of the variance in the time domain is taken as a proxy.

This method has been seen to insert more discontinuities into the data which can be seen by the increased higher-frequency noise due to a larger pink noise amplitude. Certainly if the data was to be prepared using the original pipeline then the signal to noise at higher-frequency (in the regime of low-frequency p-modes) could be improved and possibly surpass current values. In order to stitch the stations together the year long datasets are simply concatenated together as a final step.

A comparison between the old and new datasets are shown in Fig 2.17, the data are smoothed to give a better idea of the differences between the two methods. Clearly the diurnal peaks above 70 μHz have been removed which is a great start to improving the low-frequency noise, although as a result the noise levels have increased slightly at high frequency along with some granulation power being removed. Removing the diurnal peaks shows that the coherent trends (whatever they are caused by, most likely instrumental noise) are the problem in the data and that the PCA is an efficient method of removal. This opens up the low-frequency regime to full analysis to look for g-modes and in the future, low-frequency p-modes.

There is also the issue of the difference in the granulation background between BiSON and GOLF which can be clearly seen in Fig 2.17. There is an assumption to be made that the background granulation signal should not differ much between the two different instruments, and certainly not to such a large degree. Since GOLF observes at a different height in the solar atmosphere it may be expected that the amplitudes could differ, but certainly not the granulation timescales. Therefore there may be the possibility that the background seen in BiSON is instrumentation dominated rather than solar dominated, however that is beyond the scope of this work. In addition the PCA could also distort any potential g-mode signal present and so the effect of the additional analysis would have to be tested on artificial data

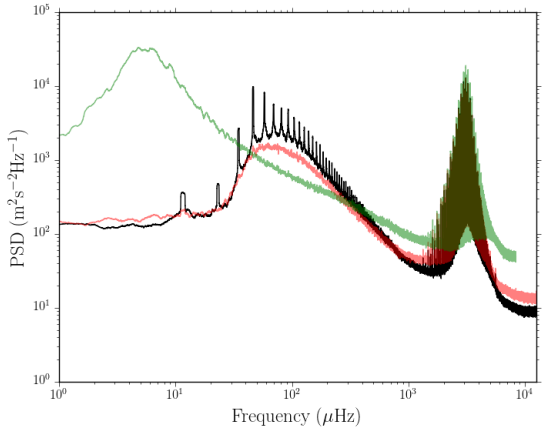


Figure 2.17: A comparison between the various datasets used for solar analyses. The black shows the BiSON data before the PCA is applied, the red shows it afterwards and the green is approximately 15 years of GOLF data. All power spectra have been smoothed by a boxcar of width 1μHz.

to assess any damage it might do.

2.5.3 Improvement of Low-frequency Noise Levels

Before the search for g-modes can begin, it is important to look at how the noise levels in given frequency bands have been improved as a result of the PCA. To do this the data was split into 50 day windows during which the median noise level was computed. The frequency bands used were $100-200$ μHz which is part of the g-mode region of interest, $200-800$ μHz which is important for low-frequency p-modes, $800-1300$ μHz which covers the frequency region where the lowest-frequency p-modes so far have been found (e.g. Davies et al. 2014b). This procedure can also be carried out for GOLF data and compared directly to the BiSON noise levels.

Analysis of this kind is important, not only for assessing the improvement in noise levels, but also for seeing how the noise levels change as a function of time. This can help diagnose problems with datasets but also if the current length of dataset is optimum for its intended purpose (as will be seen later).

It is assumed throughout this work that the long-term trends are due to instrumental and

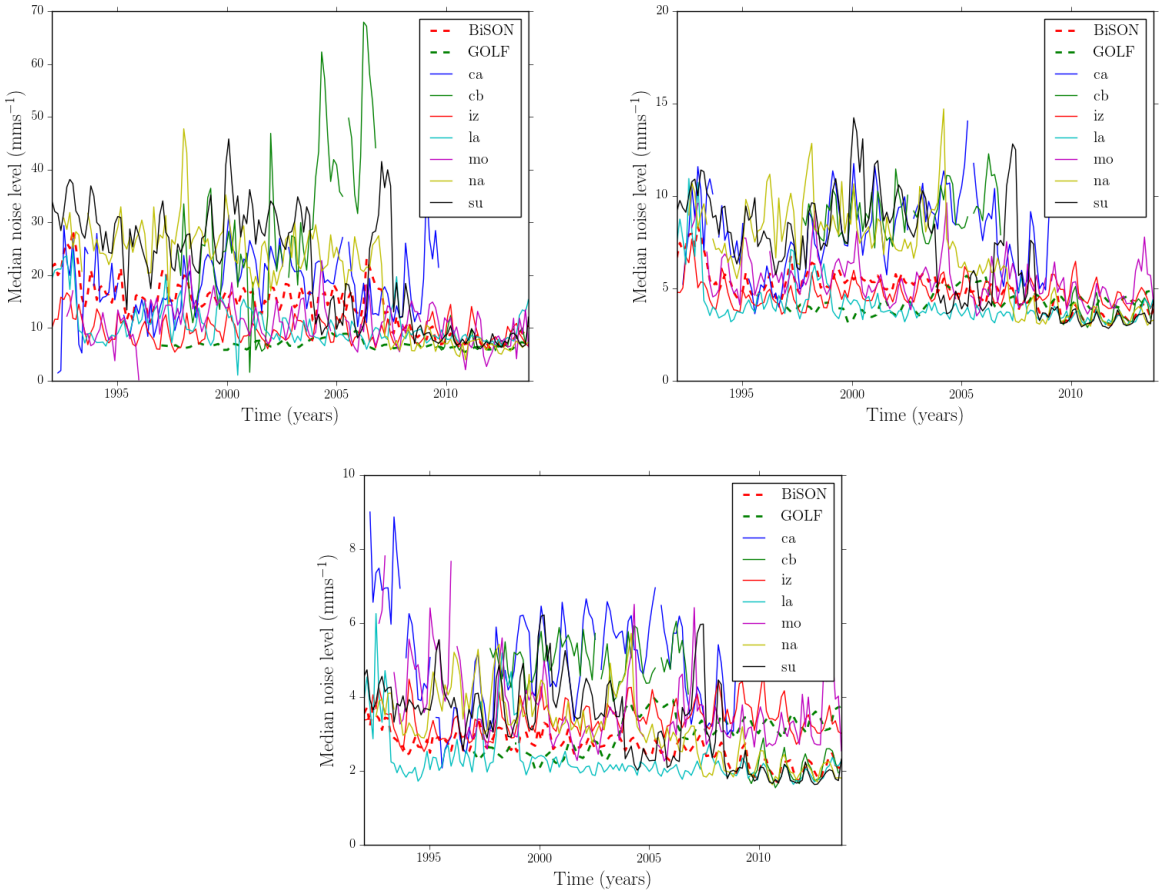


Figure 2.18: Plots comparing the median noise levels (in amplitude) of the individual BiSON stations (before the PCA was applied) and GOLF over a time period of 1992 to 2013.

not due to solar activity. Correlating data gained from the instruments and comparing these against other observational traces of granulation activity could help shed some additional light on the nature of the trends. This should be checked in the future as it will help give an idea as to the magnitude of the trends and whether the PCA is taking out too much or not enough noise.

To start, the analysis applied to the BiSON data before the PCA was applied is shown in Fig 2.18. It can clearly be seen that there is a sudden reduction in noise after approximately 2007 partly due to improved instrumentation. Otherwise it is clear that before about 2007 the stations Sutherland, Narrabri and Carnarvon all have much higher noise levels that the other stations. It is thought that due to the mounts used in these stations this contributes to a higher noise level which can be seen here. Certainly the noise levels are higher than GOLF and if the aim is to detection g-modes with theoretical upper limits of $\sim$3 mms^{-1} then an improvement in the noise levels needs to be made.

This analysis can be performed again, but this time for the data once the PCA has been applied as shown in Fig 2.19. Whilst the reduction in noise level is not as obvious on an individual basis, apart from Sutherland and Narrabri, the combined data shows a definite improvement at low frequency. This is ideal and will contribute towards BiSON pushing to the detection limit of g-modes in the future.

In order to make the improvements a little more obvious Fig 2.20 shows the same plots as above (for the first two frequency regions) but this time only showing the combined datasets. The biggest improvements in the very-low frequency regions have been made in the older datasets before the new instrumentation was installed. Any improvement in the noise levels is a step in the right direction and this reduction by almost a factor of 2 (in power) in parts is very valuable. This provides a means of showing that the PCA, along with removing the diurnal peaks, has helped improved the BISON low-frequency noise levels and beyond that also shows that with additional testing and tweaks they could be improved even more.

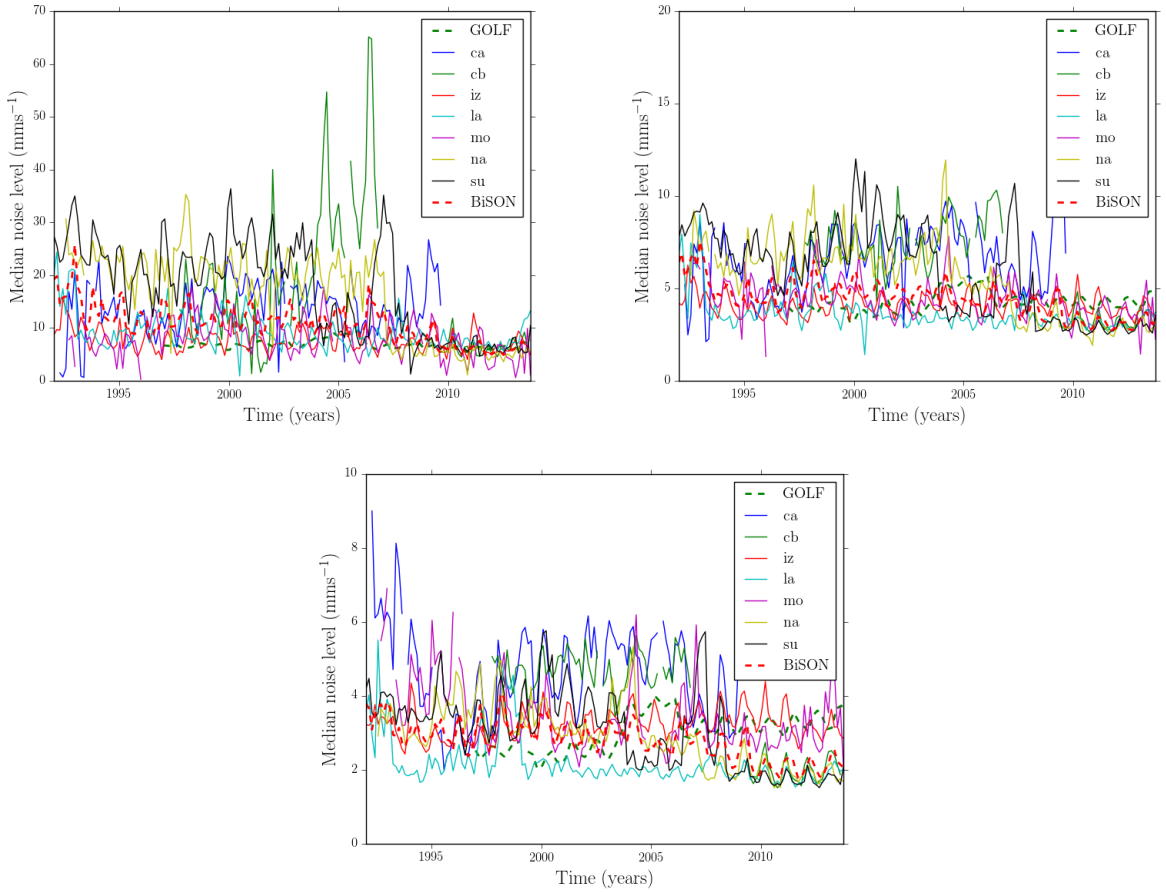


Figure 2.19: Plots comparing the median noise levels (in amplitude) of the individual BiSON stations (after the PCA was applied) and GOLF over a time period of 1992 to 2013.

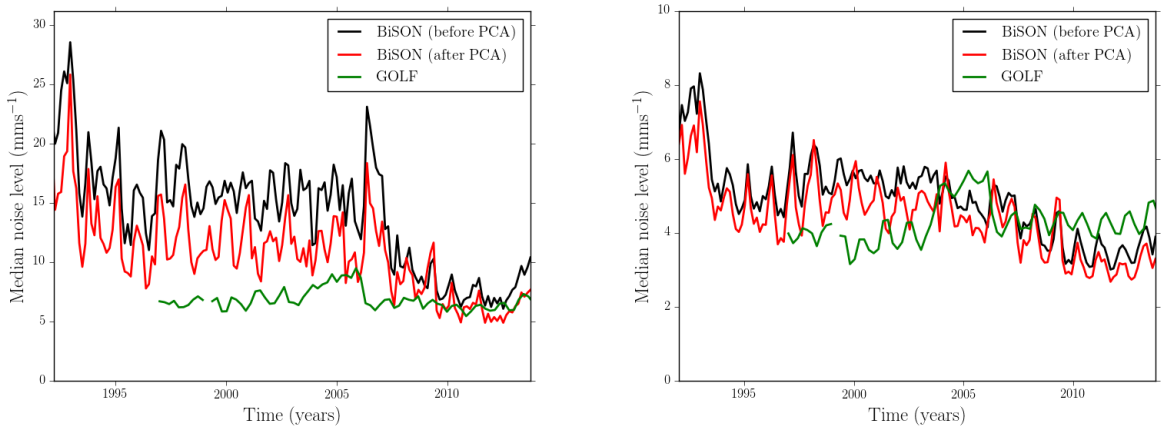


Figure 2.20: Plots comparing the median noise levels (in amplitude) of the combined BiSON data (before and after the PCA was applied) and GOLF over a time period of 1992 to 2013. The left hand panel shows the 100-200 μHz region and the right hand panel shows the 200-800 μHz region.

2.5.4 Search for g-modes

Now that the dataset has been cleaned and the diurnal peaks removed the search for g-modes can finally begin. This will be performed in two different steps, the first will be making use of the periodic nature of the g-modes (in period) by computing the power spectrum of the power spectrum (in period) and searching for a peak at the period spacing, the second will be searching for individual peaks in the power spectrum and assessing statistical significance.

The use of the power spectrum of the power spectrum in period (PSPS) applied to the GOLF data by García et al. (2007) who subsequently found a set of statistically significant peaks at a period of ∼23.5 mins. Making the assumption that the modes contributing to these peaks are $\ell = 1$ then the period spacing of the g-modes would be approximately 24 mins which is in-line with current solar models (e.g. Provost et al. 2000 or Mathur et al. 2007).

Performing this type of analysis on the BiSON data before the PCA was applied would not be helpful since the diurnal peaks would dominate the PSPS and possible g-mode signal would be swamped. Therefore the PSPS can be applied to the newly cleaned dataset for both the entire length (21 years) and a shorter subset (∼5 years) which corresponds to data taken after the sudden decrease in low-frequency noise levels.

In order to test the significance of any peaks a method similar to that proposed by Baluev (2008) whereby the significance levels are computed using bootstrapping. The data is permuted N times and the PSPS computed, the maximum peak is recorded for each permutation and the significance levels computed from the percentiles of those maximum peaks. This gives an approximate method of computing peak significance and will be applied to both GOLF and BiSON data.

The data needs some preparation before the PSPS can be applied in order to try and reduce extra noise sources being introduced (due to granulation for example). Therefore the signal-to-noise (or background-divided) spectrum is used and created by dividing the

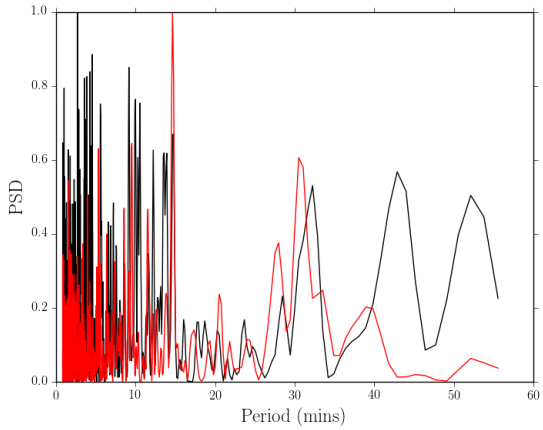


Figure 2.21: PSPS for full BiSON datasets (black) and subset (red). The PSD has been normalised by the highest value of each power spectrum for aesthetic purposes.

power spectrum through by a smoothed version of the power spectrum (approximating the background) using a boxcar of width 20 μHz. Consistency is maintained with previous works (García et al., 2007) by searching over the same range of frequencies, 25-140 μHz, which lies within the asymptotic region for high-order g-modes. The PSPS for both the longer and shorter subset of data is shown in Fig 2.21, where a clear peak at 30 minutes is seen rather than at the expected period of 24 minutes, the reason for this peak is unclear.

The bootstrap confidence intervals were computed using the `astroML` package[7]. This is performed by permuting the data N times and each time calculating the PSPS and recording the maximum value. After doing this for the N permutations the false alarm probability is estimated by computing the 10%, 5% siginificance levels of the distribution of the maximum values. In order to do this properly the value of N needs to be reasonably large, it is possible in this case that it is not large enough to give a reliable estimate of the significance levels. This could also have a knock on effect when looking at the results from GOLF, since the significance levels were computing in a different manner to that given here, possibly leading to the observed discrepancy.

The PSPS for the long and short subsets and the last 16 years of GOLF data are shown in

[7]https://pypi.python.org/pypi/astroML/

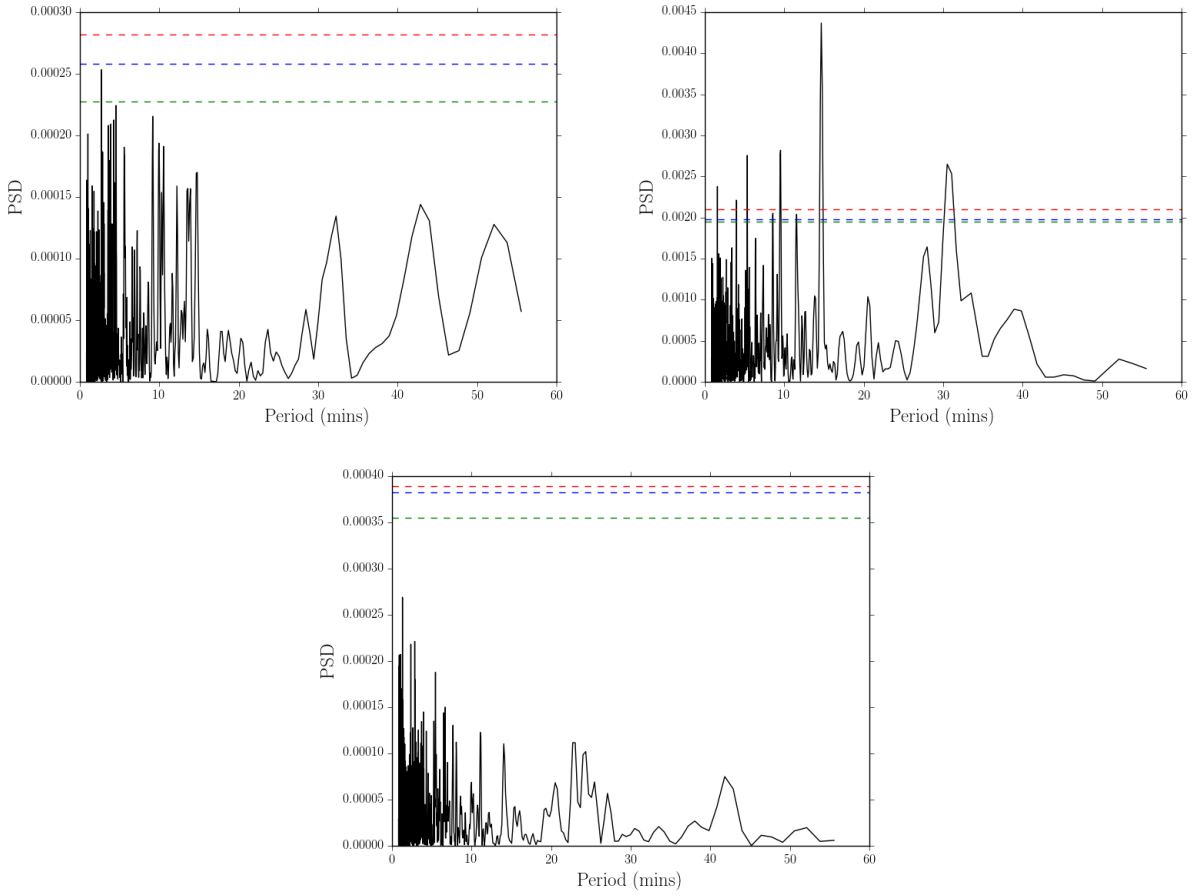


Figure 2.22: PSPS for full BiSON dataset (top left) and subset (top right) as well as the last 16 years of GOLF data. The significance levels according to 90%, 95% and 99%, given by the green, blue and red dashed lines respectively according to $N = 20$ bootstrapped datasets.

Fig 2.22 along with the bootstrapped significance levels. It is clear that in the shorter, less noisy subset of the BiSON data the peak at 30 minutes is significant, however its position does not correspond to any known or theorised peak. In addition the peak observed at ∼24 minutes is shown to not be significant when processed under the same conditions as the BiSON data, showing that perhaps there has still not been an ensemble detection of solar g-modes.

Due to the fact that no ensemble detection has been made in the BiSON data the next step is to place an observational upper limit on the g-mode amplitudes, this is done using statistical detection tests (see Appendix A.6 for more details). The limit is first defined in the background-divided spectrum as the power value such that the probability of observing a noise spike above that value is less than 10% (i.e. a 10% false alarm probability). This value can then be multiplied back up by the estimated background to give the threshold in the power spectrum, from this the estimated background is then subtracted off given that the modes will be sat upon the background. Since g-modes are thought to be unresolved in the power spectrum (due to their large lifetimes and narrow linewidths) the amplitude can be simply calculated through $P = A^2/2$ where P is the mode power and A is the amplitude. Not forgetting that the inferred amplitudes must be fill-corrected, and so the amplitude must be divided by the square root of the fill to account for this (the fills are 0.8 and 0.96 for BiSON and GOLF respectively).

The observational thresholds are shown in Fig 2.23 for both BiSON (in black) and GOLF (in red). The differences in the background profiles of the data become apparent in these thresholds, the turnover as a result of the calibration can be seen in the BiSON thresholds. In terms of upper limits, whilst GOLF is lower than BiSON for the majority of the g-mode region this disparity has been reduced and improved by the removal of the diurnal peaks. There is a point at $\sim$ 290 μHz where the limits cross over and BiSON has a lower upper limit than GOLF. This is important when considering the detection of possible low frequency

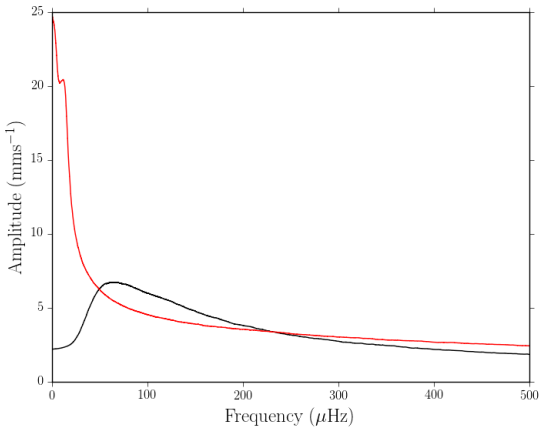


Figure 2.23: Upper limits on g-mode amplitudes as a function of frequency for BiSON (black) and GOLF (red).

p-modes and solar mixed modes.

The established theoretical upper limit on g-mode amplitudes set by Appourchaux (2003b) was around 3 mms^{-1} at 200 μHz assuming a quadruplet structure, whereas Belkacem et al. (2009) attained g-modes amplitudes of ~ 3 mms^{-1} for an $\ell = 2$ mode at 100μHz. Our upper limits are established to be 6.7 mms^{-1} at 100 μHz and 4.3 mms^{-1} at 200 μHz assuming a single spike. The upper limits extracted from GOLF data using the same method are 4.7 mms^{-1} at 100 μHz and 3.7 mms^{-1} at 200 μHz, again assuming a single spike. These upper limits will be reduced when considering both the doublet structure of the $\ell = 1$ and quadruplet structure of the $\ell = 2$ modes.

2.6 Conclusion

In conclusion, the source of the diurnal peaks found in the very-low frequency region of the BiSON power spectrum have been attributed to long-term coherent instrumental trends. Through principal components analysis (PCA) these peaks have been removed from the data, greatly improving the quality of the data at very-low frequency.

The analysis performed by (García et al., 2007) was repeated for both the GOLF and BiSON data and no significant peak at the expected period was found in either, indicating

no ensemble detection of g-modes. In the BiSON data a significant peak was detected at a period of ~ 31 minutes, but this does not correspond to any known physical signal. The underlying cause of this peak has yet to be investigated.

Upper limits have been placed on g-mode amplitudes using BiSON and GOLF data. For the BiSON data this is 6.7 mms^{-1} at 100 μHz and 4.3 mms^{-1} at 200 μHz assuming a single spike. Whilst for GOLF the limits are 4.7 mms^{-1} at 100 μHz and 3.7 mms^{-1} at 200 μHz again assuming a single spike. These upper limits still lie above limits established by previous work and limits from theory. However, with appropriate tuning of the PCA and the method behind combining the data this threshold could be lowered to bring the BiSON noise levels into the regime of g-mode amplitudes. It is also necessary to test the effect of the PCA on the amplitude of g-modes using artificial data to ensure that our limits are reliable.

As a result of applying the PCA the low frequency regime of the BiSON data has finally been opened up to be analysed for g-modes and low-frequency p-modes.

3

Inference and characterisation of Red Giants from their background power

The idea of obtaining global properties of an underlying distribution is at the heart of the majority of analyses, and asteroseismology is no different. The wealth of red giants at our disposal means there is the potential to explore the ensemble properties of stars in much more detail. Relating to red giants, this has been done in the context of background granulation parameters (Mathur et al., 2011; Mosser et al., 2012a; Kallinger et al., 2014) and global oscillation properties (Huber et al., 2010; Mosser et al., 2012b; Vrard et al., 2016) but there is still much more that can be done with these data. It is common when possessing such a large amount of information to try to infer scaling laws from these background and oscillation parameters (e.g. Mosser et al. 2012a or Kallinger et al. 2014), generally with $\nu_{\rm max}$, due to the relative ease of its determination. This is performed in a deterministic sense, i.e. we fit a

power law to the data and use that power law as the predictor for new data or as first guesses in subsequent analyses. Analysing the datasets in this way throws away a tremendous amount of diagnostic potential contained within these data and the aim of the work presented in this chapter is to use this power to not only improve predictions but to also update these scaling laws by appealing to their stochastic nature and including the scatter by using hierarchical Bayesian modelling e.g. Angus & Kipping (2016).

An important idea is, can knowledge of high-quality datasets be used to inform fits to low-quality or shorter datasets? By fitting the background power spectra of a set of high-quality template stars, the extracted data can be used as prior distributions on fits to lower-quality datasets, or at the very least provide information that can improve initial guesses for the fitting process or a more physically motivated way in which to initialise walkers in the MCMC part of the fits. This is important due to the generally lower signal-to-noise and worse frequency resolution found in shorter datasets. This idea will be explored throughout this chapter and is relevant for using data from *Kepler* and applying it to data from K2.

In addition to updating basic scaling laws, the data can also be applied to other problems within asteroseismology. One of the main problems within stellar astrophysics is determining the evolutionary state of red giant stars, whether they are red giant branch (RGB), red clump (RC) or secondary clump (SC) stars. Knowledge of the internal processes and core mechanisms would reveal the evolutionary state very easily[1]. Discrimination using characteristics that can be measured with traditional astronomical data, such as luminosity, effective temperature and surface gravity can be difficult due to the fact that RGB and RC can have similar observable characteristics whilst possessing vastly different internal conditions.

With the aid of asteroseismology it is possible to classify the evolutionary state of a star by making use of its oscillations, either through the period spacing $\Delta\Pi_1$ (Mosser et al., 2014)

[1]RGB stars are burning hydrogen (H), RC core-helium burning in an electron-degenerate core and SC are core-helium burning in a non-degererate core (at least when core Helium burning begins, but becomes degenerate if the mass is not too large).

or through the empirical properties of the oscillations themselves (Elsworth et al., 2016). This works well for long, high-quality datasets where the oscillations are resolved, but for the missions that have shorter amounts of data available (e.g. K2 which has a baseline of 80 days of observations per campaign, or TESS which has a maximum observing length of 1 year and a minimum of 27 days) the oscillations will be much harder, if not impossible, to resolve. The ability therefore to deduce the evolutionary state of a star from its background parameters could prove valuable and is an idea that will be addressed later in this chapter.

3.1 Data sets

In the process of performing the background fits the set of template stars must be chosen carefully such that they provide as much information as possible for future applications. In addition they must also be long enough to be cut down to simulate shorter datsets.

To achieve this goal a subset of red giants observed by *Kepler* (Stello et al., 2013) from the APOKASC sample (Pinsonneault et al., 2014), specifically the Data Release 1 datasets, were used. All data were collected in long-cadence (LC) at a cadence, Δt, of 29.4 min for a total of 4 years and were reduced using the method described in Handberg & Lund (2014). The reason for picking the APOKASC sample is due to the overlap with some of the *Kepler* targets with the spectroscopic campaign APOGEE (Majewski et al., 2010; Hayden et al., 2014). This enables the spectroscopic parameters of all the stars fitted in this work to be used in the subsequent exploration of the data and allows any trends in, for example, metallicity [Fe/H] or effective temperature $T_{\rm eff}$ to be explored.

In order to assess the performance, the methods we will develop will be tested on "*Kepler* as K2", i.e. running on a reduced *Kepler* data set (down to 80 days from 4 years). This gives the opportunity to test for any biases present in any fitted parameters by comparing the 80 day parameters to those determined from the full 4-years.

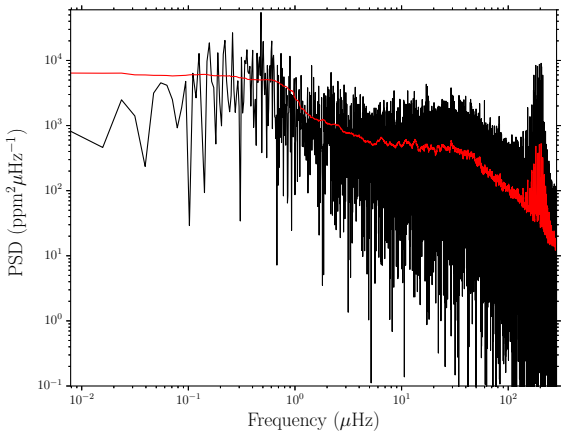


Figure 3.1: Power spectrum of KIC 3533464 shown in black and in the red is the same spectrum smoothed with a boxcar over a region of 1μHz.

3.2 Data Preparation

An example of the red giant power spectra being fitted in this work is shown in Fig 3.1. The oscillations can clearly be seen at a frequency of $\sim 195\mu$Hz and the rising granulation background can be seen extending to low frequency. The important idea to note is that in the preparation, the data are effectively high-pass filtered and so part of the very-low frequency region of the spectrum will be contaminated by this filtering. This can be seen below $\sim 1\mu$Hz in Fig 3.1 where the granulation power appears to fall off. In addition any long-term instrumental effects will also manifest themselves in the first few bins of the power spectrum. This behaviour is far from ideal and in an attempt to negate any of the above effects the power spectra are cut off at 1μHz. The applied cut in frequency will also limit the fitting to stars with $\nu_{\rm max} > 5\mu$Hz, but stars with $\nu_{\rm max} < 5\mu$Hzthese are fairly rare and so it is not too much of an issue for our purposes.

3.3 Background Model

Arguably the most important part of this work is formulating a good model of the granulation background and oscillations such that the priors extracted from the data are meaningful. The

model used consists of three Harvey-like profiles (Harvey, 1985) described by zero-frequency centred Lorentzians with a given amplitude and characteristic frequency, a Gaussian envelope to describe the power contained within the oscillations and a flat background representing the white noise. The reasoning behind each component of the model will be explained shortly, but first let us state the full model

$$\mathcal{B}(\nu) = \eta^2(\nu) \left(\sum_{i=0}^{2} P_{\mathrm{gran},i} + P_{\mathrm{osc}} \right) + W \ , \tag{3.1}$$

where $P_{\mathrm{gran},i}$ is the ith granulation component, P_{osc} describes the contribution from the oscillations, W is the white noise and $\eta(\nu)$ accounts for the amplitude modulation due to the sampling interval. The $\eta(\nu)$ is an important quantity (see Campante 2012 for more details) to include as it accounts for the fact that there is a finite sampling of the data (which in turn gives rise to the Nyquist frequency, ν_{nyq}) and manifests itself as a sinc^2 function in the frequency domain with zeros corresponding to the sampling frequency[2] as given by (e.g. Chaplin et al. 2014b)

$$\eta = \mathrm{sinc} \left(\frac{\pi}{2} \frac{\nu}{\nu_{\mathrm{nyq}}} \right) . \tag{3.2}$$

Each granulation component is described by the following form

$$P_{\mathrm{gran},i} = \frac{\sigma_i}{1 + (\nu / b_i)^{c_i}}, \tag{3.3}$$

where σ_i denotes the height of the ith component, b is the characteristic frequency and c is the exponent which controls the steepness of the drop-off of the Lorentzian-like profile. The model given in equation 3.1 is a variant of model H described by Kallinger et al. (2014). In the original work performed by Kallinger et al. (2014) Bayesian model comparison was performed to select the most appropriate model for background fitting. The decision was

[2]For *Kepler* LC observations this is $\sim 568\mu$Hz

mainly between two models, one with an exponent set to 4 (roughly corresponding to a process that has a rising and decaying correlation) or one with the exponent set as a free parameter. Although the model comparison suggested, marginally, that the model with the free exponent (model H) was more suitable than the model set to 4 (model F), model F was adopted instead. This was mainly due to ease of use and practicality, however it is apparent that an exponent of 4 is not always suitable. As a result the decision in this work was to let the exponent vary as a free parameter. Fixing the exponent will also have a knock on effect on the amplitude and timescale of the granulation components and so any scaling relations created will depend on the model used. It also provides an opportunity to explore the exponent as a function of stellar parameters to deduce where any dependencies may lie.

Another slight difference between the model presented in equation 3.3 and that of Kallinger et al. (2014) is the use of the height instead of the more commonly used amplitude of the granulation component. For an exponent of 4, σ is related to the amplitude a via $2\sqrt{2}a^2/(\pi b)$, where the factor $\xi = 2\sqrt{2}/\pi$ ensures the granulation component is correctly normalised. However this is not valid when c is not an integer (see Kallinger et al. 2014) and so to get around this the height of the granulation component σ is fitted instead which automatically incorporates the normalisation factor. In the case of c being a non-integer a can be derived by rearranging the formulation for the normalisation constant ξ

$$\frac{1}{\xi} = \frac{1}{b}\int_0^\infty \frac{1}{1+\left(\frac{\nu}{b}\right)^c}\mathrm{d}\nu, \tag{3.4}$$

where

$$\sigma = \frac{\xi a^2}{b}. \tag{3.5}$$

This normalisation constant is designed to ensure that the square of the amplitude is equal to the variance in the time-series.

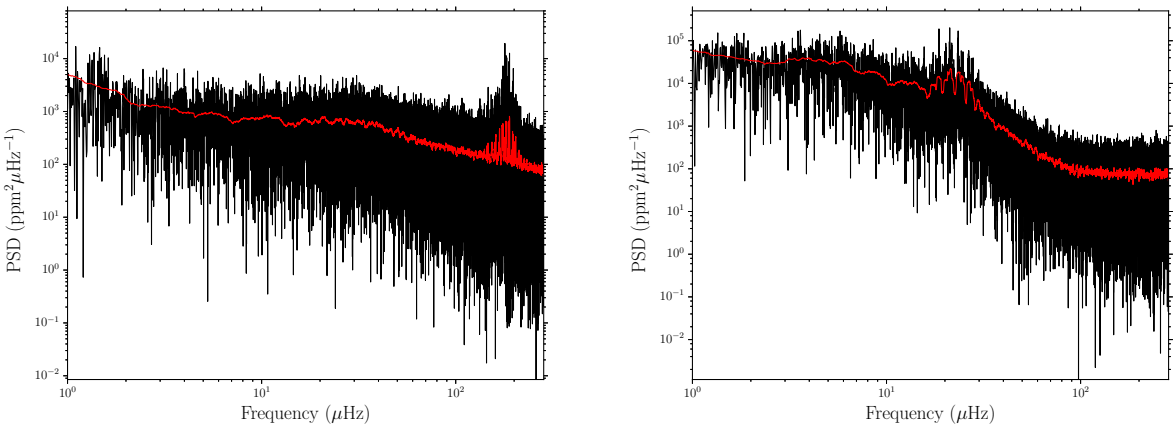


Figure 3.2: Two more examples of red giant power spectrum observed by *Kepler*. The left-hand plot is of a high $\nu_{\rm max}$ star that shows the slight rising component at low frequency commonly associated with "activity", whereas the right-hand plot shows a star with a much lower $\nu_{\rm max}$ that does not show any evidence of "activity". Note the change in scale of the y-axis due to the large difference in $\nu_{\rm max}$ of the two stars.

The third component is treated almost like a "nuisance component" (to draw from the Bayesian language) whereby it is incorporated into the model to mop up any low frequency signal but the values taken by the component is not important[3]. As a result the characteristic frequency is limited to no greater than $\sim 50\mu$Hz. An example is shown in Fig 3.2 for two stars with very different $\nu_{\rm max}$ values (183μHz and 23μHz respectively), where for the low $\nu_{\rm max}$ star the granulation background appears to level off at low frequency but starts to rise slightly for the higher $\nu_{\rm max}$ star.

The other major constituent part of the model is to describe the oscillations, done so using a Gaussian envelope with height $H_{\rm env}$, full-width at half-maximum (FWHM) $\delta\nu_{\rm env}$ and central frequency $\nu_{\rm max}$ as given by

$$P_{\rm osc} = H_{\rm env} \exp\left(-\frac{(\nu - \nu_{\rm max})^2}{2\sigma^2}\right), \tag{3.6}$$

where the width of the Gaussian σ is related to the FWHM $\delta\nu_{\rm env}$ by $\delta\nu_{\rm env} = 2\sqrt{2\ln 2}\sigma$.

[3]This is not strictly true, but this will be explained in more detail later in the chapter.

3.4 Fitting Methodology

Having covered the model that is going to be used in the fitting process let us now cover the procedure that is going to be adopted in this work. The use of Bayesian statistics has quickly risen to prominence within the field and numerous works have explained this in detail (see Appourchaux 2008; Appourchaux et al. 2009; Kallinger et al. 2014 and Davies & Miglio 2016 for example). As a result these approaches will be followed and MCMC will be used for parameter estimation (by sampling the joint probability distributions), using `emcee`[4] (Foreman-Mackey et al., 2013). The advantage of `emcee` over other MCMC implementations comes in the form of the stretch move which gives the algorithm its affine-invariant property. In addition it explores the parameter space using an ensemble of walkers, thereby enabling faster traversal, where the proposal distribution for each walker is based on the positions of every other walker.

Given this Bayesian approach both prior distributions and a likelihood function need to be defined, this will be covered in the next section.

Before jumping straight into the methodology, let us first give a brief overview of Bayes's theorem. The objective of this analysis is to make inferences about some parameters θ (which will be part of a chosen model) given some observed data y. This can be done according to Bayes' rule

$$p(\theta|y) \propto p(\theta)p(y|\theta). \tag{3.7}$$

The first term on the right-hand side of the above equation, $p(\theta)$, is the prior probability which encompasses our prior knowledge about the parameters of interest. This is subsequently modified by the likelihood function, $p(y|\theta)$, which gives us information on the probability of the data given the input set of parameters. The most important term is that on

[4] `emcee` is pure-Python implementation of the affine invariant MCMC ensemble sampler proposed by Goodman & Weare (2010).

the left-hand side, $p(\theta|y)$, which is the posterior probability which represents our knowledge of the parameters given the data. In reality the proportionality should be replaced with an equality with the marginalised likelihood as the denominator, which represents the integral of numerator over all of parameter space. The marginalised likelihood is important when considering model selection, but when using MCMC it is not needed and the unnormalised quantities are used instead.

Likelihood function

The properties of the noise in the power spectrum follows a χ^2_2 (χ^2 with 2 degrees of freedom) distribution and so the likelihood function needs to be chosen accordingly. The appropriate function (e.g. see Duvall & Harvey 1986; Anderson et al. 1990) is as follows

$$\ln(L) = -\sum_i \left[\ln M_i(\theta) + \frac{P_i}{M_i(\theta)} \right], \tag{3.8}$$

where the summation is made over each element in the frequency array, $M_i(\theta)$ is the model evaluated at a given frequency for a set of parameters θ and P is the power value at a given frequency[5].

As explained in the previous section it is not simply the power that is dealt with, but the rebinned power spectrum. The process of rebinning means that the data are averaged over s bins, which in the case of this work is taken to be equivalent to 1μHz. This value was judged to provide a good trade-off between effective frequency resolution and computational speed. As a result of fitting to the rebinned power spectrum the statistics of the data have changed and this needs to be incorporated into the likelihood function defined in equation 3.8. Therefore the likelihood function becomes (Appourchaux, 2003a)

[5]The log-likelihood ($\ln L$) is commonly used due to numerical stability.

$$-\ln \mathcal{L}(P,\theta) \approx s \sum_i \left[\ln M_i(\theta) + \frac{P_i}{M_i(\theta)} \right], \tag{3.9}$$

where s is the number of bins rebinned over. This is a very simple modification to the original likelihood function (the derivation of which is given in Appendix A.7) and results in a large speed up in computation time due to using a fraction $1/s$ of the original dataset.

Priors

The second important set of quantities are the prior distributions. These distributions will describe (as the name suggests) any prior knowledge about the data before the fitting is applied. Generally these come in the form of uniform priors that stop the sampling algorithm from straying into unphysical regimes, however a few additional priors were also added (that are weakly informative and in the form of Gaussian priors) to help with this. The idea of these priors was to stop the fit from pursuing any non-physical solutions whilst remaining uninformative to the rest of the data, which is important if new inferences wish to be drawn from the results.

The prior distributions for each parameter are shown in tables 3.1 and 3.2. It is common in astrophysics for parameters to show power law dependencies and asteroseismology is no different. Generally the amplitude and frequency parameters show a power law dependence with $\nu_{\rm max}$ and so these are used as the means in the Gaussian priors. The priors used were established over several runs of older versions of code and built up through experience. They were also chosen deliberately such that they have no influence on the fitting procedure, only that they stop the walkers from exploring unphysical regions of parameter space.

3.5 Fitting Process

Having covered the setup of the statistics behind the fitting process let us move on to the procedures used behind the scenes. This involves how to initialise the ensemble of walkers

Table 3.1: Parameters for the Gaussian priors.

Parameter	μ	σ
$\ln \sigma_1$ (dex)	$\ln 2 \times 10^7 - 2 \ln \nu_{\rm max}$	$\ln 8 \times 10^6$
$\ln \sigma_2$ (dex)	$\ln 1.5 \times 10^7 - 2.4 \ln \nu_{\rm max}$	$\ln 3 \times 10^6$
b_1 (μHz)	${\nu_{\rm max}}^{0.95}/2.75$	1×10^3
b_2 (μHz)	$\nu_{\rm max}$ /0.95	1.5×10^3
$\ln H_{\rm env}$ (dex)	$\ln 4 \times 10^7 - 2.6 \ln \nu_{\rm max}$	$\ln 5 \times 10^6$
$\delta\nu_{\rm env}$ (μHz)	0.1	150

Table 3.2: Parameters for the uniform priors used in background fit. In the case of using short-cadence data instead of long-cadence the parameters marked with an asterisk would need to be increased to ensure the optimal solution is found.

Parameter	Lower bound	Upper Bound
$\ln \sigma_1$ (dex)	-3	20
$\ln \sigma_2$ (dex)	-3	20
$\ln \alpha$ (dex)	-3	20
b_1 (μHz)	0.1	283.0*
b_2 (μHz)	0.1	283.0*
β (μHz)	0	50
c_1	2.0	6.0
c_2	2.0	6.0
c_3	0.0	6.0
$\nu_{\rm max}$ (μHz)	1	350*
$\ln H_{\rm env}$ (dex)	-3	20
$\delta\nu_{\rm env}$ (μHz)	0.1	150*
W (ppm$^2\mu$Hz^{-1})	$W_{\rm est}/2$	$4W_{\rm est}$

used in `emcee`, assessing convergence and getting a good estimate of $\nu_{\rm max}$ from which to create the prior distributions.

3.5.1 Initialising the fits

The sampling algorithm `emcee` provides an ensemble of walkers to explore the parameter space and so each of these walkers needs to be initialised at a point in parameter space. It is commonly suggested that this is done in a small Gaussian ball about the first guesses of the parameters. This idea would work well for an individual star but when performing ensemble analyses it is important for the walkers to retain no memory of their initial positions. This should never happen provided the walkers have achieved convergence, but initialising all the walkers in one place can also be a hindrance if the initial guesses are not good enough.

The purpose of performing ensemble analyses is to make inferences about the underlying population distribution or the observed distribution, and so a requirement should be that no solution is favoured over another. In other words the results of previous works should be incorporated into the prior distributions but should not bias the way in which the fits are performed. Therefore the decision was made to "over-disperse" the walkers according to a Normal distribution with a given mean and standard deviation. This may enable solutions to be found that were not possible in other analyses due to idiosyncrasies in their methodologies[6], but also ensures that the fit can still converge to an appropriate solution even if the initial guesses are very bad.

The walkers were initialised with mean values of the priors given in Table 3.1 but with standard deviations a factor of 5 lower than those in the table, this results in faster convergence whilst maintaining an over-dispersed set of walkers compared to the case. For the exponents the walkers are initialised at a value of 4 corresponding to the values used in Kallinger et al. (2014) and a standard deviation of 0.1. This gives the walkers the oppor-

[6]This could be due to overly-constrained priors, or getting stuck in local minima if gradient-descent type algorithms are used.

tunity to explore a parameter space that is not arbitrarily constrained, since the value of the exponent is unknown although it is assumed to lie close to 4. It was decided to use 400 walkers and run 400 iterations with 6 temperatures, these values were chosen through trial and error and gave a good trade-off between computation time and efficiency[7]. The large number of temperatures gives the walkers the opportunity to explore a large region of the parameter space. However it is noted that sometimes when including parallel tempering it can struggle when the signal-to-noise is very high, and so a better defined temperature ladder would be needed (Vousden et al., 2016).

The walkers were not completely over-dispersed (according to the priors given Table 3.1) and scattered with smaller standard deviations for a couple of reasons. Completely over-dispersing the walkers will result in loss of memory of their initial positions can also result in highly unphysical solutions being obtained. Whilst it is a good idea to over-disperse the walkers to enable the optimal solution to be obtained, unphysical solutions should be avoided at all costs (as the name suggests). Given the large range of which the walkers would be dispersed and the potentially highly multi-modal structure of the likelihood it could be possible for the walkers to get stuck in minima far away from the desired solution with no way of transitioning back to a more likely minimum. As a result in this work the walkers are over-dispersed but not by such a large amount that the fit will take an unreasonable amount of time to achieve convergence[8].

Initial Estimate of $\nu_{\max}$

All of the initialisation procedures above rely on a good determination of $\nu_{\max}$ in order to attain good (and fast) convergence. To do this, a quick and easy model can be created from

[7]Parallel tempering is designed to improve the sampling of multimodal distributions by raising the likelihood to the inverse of a set of temperatures (defined by a set ladder) which results in "smoothing" out the likelihood the higher the temperature. See Neal (1996) for more detail.

[8]The fit is judged to have taken too long to converge if it takes over 20 iterations of the fitting procedure (as explained later in the chapter). The assumption is made that if the fit has not converged in this time it is unlikely to if the fit is run for any longer.

the scaling relations for a given $\nu_{\rm max}$ and then fitted to the data. Depending on the $\nu_{\rm max}$ values used this can yield a reasonable estimate, as will now be shown based on an approach by Davies (2015a), whilst an improved method will be shown towards the end of the chapter.

Rather than using the established scaling relations it was decided to estimate the mean of the parameter distributions from old runs of the code (in the same way the prior distributions were created). In order to estimate the mean of the prior distributions power laws were fitted to each parameter as a function of $\nu_{\rm max}$ using the random sample consensus (RANSAC) algorithm (Fischler & Bolles, 1981) which provides a way to perform robust linear regression on the data whilst reducing the impact of outliers. This meant that for a given $\nu_{\rm max}$ value the background parameters could be estimated, apart from white noise which was estimated by taking the mean of the last 50 bins in the power spectrum. To enhance the computational efficiency the spectrum was also rebinned over 1 μHz and the accompanying factor in the likelihood function was also accounted for as shown above. The choice of 1 μHz was again made as a compromise between computational efficiency and frequency resolution, which gave good results for low $\nu_{\rm max}$ stars and high $\nu_{\rm max}$ stars alike.

Having prepared the data an iterative algorithm was applied to calculate the best $\nu_{\rm max}$ value. The model was created for a given value of $\nu_{\rm max}$ and the mean parameter values extracted from the power law fits to the prior distributions. The log-likelihood for that model was then calculated. Steps of 2 μHz were taken in frequency to provide a rough estimate of the $\nu_{\rm max}$ value of the star. As discussed above, the third granulation component is not always needed and was not included in this iterative procedure, therefore any low-frequency structure can provide added difficulty to this simple algorithm. To account for this a tweak to the original procedure was added where the power spectrum was cut off at 0μHz, 0.1μHz, 0.5μHz, 1μHz and 5μHz, then the $\nu_{\rm max}$ that gave the best log-likelihood value from all of those runs was taken as the estimate of $\nu_{\rm max}$. An overview of the algorithm is given in Algorithm 1.

Algorithm 1 Estimating $\nu_{\rm max}$

if $W < 1 \times 10^4$ **then**
  $L_{\rm max} = 1 \times 10^{12}$
  $\nu_{\rm max,est} = 0$
  for $i = [0.0, 0.1, 0.5, 1, 5]$ **do**
    Process data such that frequency and power only taken above i μHz
    Rebin processed power spectrum over 1 μHz
    for $j = 0$ **to** 500 **in steps of** 2 **do**
      Evaluate log-likelihood ($L_{\rm eval}$) for model constructed at $\nu_{\rm max} = (j + 1)$ μHz
      if $L_{\rm eval} < L_{\rm max}$ **then**
        $\nu_{\rm max,est} = (j + 1)\mu$Hz
        $L_{\rm max} = L_{\rm eval}$
      end if
    end for
  end for
end if

An example of the algorithm applied to the star KIC 10351196 is shown in Fig 3.3, a high signal-to-noise star that is representative of our sample. The algorithm does a good job of picking out a $\nu_{\rm max}$ close to the actual value . In the case of binary stars with close $\nu_{\rm max}$ values the companion can contribute towards "washing out" the observed flux from the primary star which in turn lowers the granulation and oscillations amplitudes compared to what would be expected. In such a case the procedure may not do a good job and estimate a higher $\nu_{\rm max}$ than there should be. However, these stars would not want to be included if they were obvious binaries and the parameters derived would be a combination of the two stars and therefore add biases into the method.

3.5.2 Checking convergence

A hugely important part of any method involving MCMC is to assess convergence, in other words to make sure that the walkers have converged to an equilibrium distribution which is the posterior distribution from which we want to sample. There are a variety of ways to check for convergence, such as Geweke's z-score (Geweke, 1992), the Raferty-Lewis diagnostic

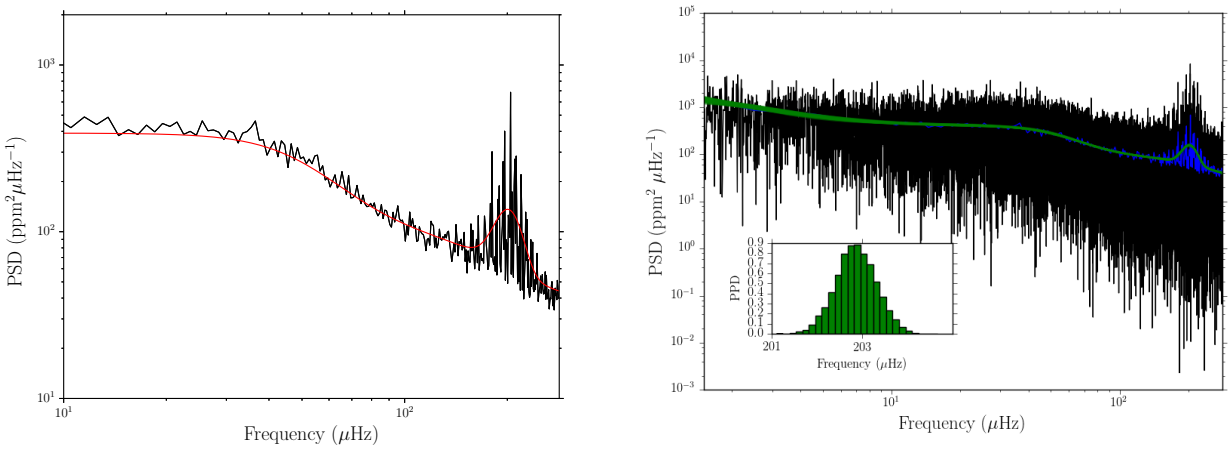


Figure 3.3: The initial model created during the initial guess (KIC 10351196) on the left hand side shows the rebinned power spectrum (over 1μHz) in black and the approximate model from the iterative algorithm in red. On the right-hand side is the final fit to the same star with the power spectrum in black, rebinned power spectrum in blue and models constructed from random samplings of the posterior PDFs in green. The $\nu_{\rm max}$ posterior PDF is also inset to show the precision of the fitting process. Also note the difference in scales between the left and right panels, this is purely for demonstrative purposes.

(Raftery & Lewis, 1995; Gamerman & Lopes, 2006) and the Gelman-Rubin convergence criterion (Brooks & Gelman, 1998; Gelman et al., 2013). For this work the Gelman-Rubin convergence criterion was adapted due to its ease of computation from our chains and well defined behaviour when run in tests over sub-samples of the data.

The Gelman-Rubin convergence criterion assesses the mixing and stationarity using variances computed between and within the chains used in the sampling (a derivation of the statistic is given in Appendix A.8). The basic idea is that the statistic looks at the difference between the inter- and intra-chain variances and checks that they are both similar in size, this is summarised by the potential scale reduction factor $\hat{R}$ (PSRF)

$$\hat{R} = \sqrt{\frac{1}{W}\left(\frac{n}{n-1}W + \frac{1}{n}B\right)}. \tag{3.10}$$

where W and B are the within and between-sequence variances respectively.

The PSRF helps show whether the fit needs to be run for longer in order to achieve convergence. It is expected that the PSRF will tend to 1 in the limit that the amount of data

tends to infinity and so if the PSRF is large then there would be evidence to suggest that the parameter estimation could be improved by running the fit for longer as convergence has not yet been achieved.

An acceptable value for the PSRF at convergence is a tricky topic and can vary substantially depending on the type of fit you are performing and the subtleties of this process which can directly affect the time taken to converge. Angus & Kipping (2016) used a value of $\hat{R} = 1.002$ and an autocorrelation time[9] of greater than 35 for convergence in their analysis of stellar flicker. However in our case due to the fact that our samplings are deliberately from wide priors and the walkers are over-dispersed, the time for convergence will increase and the chances of us achieving such a small value of $\hat{R}$ are very small. Therefore, a target value of $\hat{R} = 1.01$ was taken after our first production run. If this was not achieved then another run of the same number of iterations was performed and the criteria of convergence adjusted such that $\hat{R} = 1.01 + 0.005i$, where i is the number of additional iterations. This meant that the criterion was relaxed if more iterations were needed, this would have been due to either a poor fit (due to additional structure in the power spectrum), or due to the presence of a binary (or simply just the washout due to a binary companion). This cannot continue indefinitely in these cases and so if this occurs then one final iteration is made with double the number of iterations in order to make one final push for convergence to an optimal solution. There are cases where the algorithm fails and the fits are all inspected by eye to deduce whether the fit has been a success.

An additional step that is made after each iteration is to "reassign" walkers in poor locations, which is equivalent to pruning according to a walker's log-probability. The bottom half of the walkers, calculated according to their log-probability, are taken to be "bad" and are then scattered according to the median of the good walker locations with a standard deviation calculated from the MAD of the good walkers. The idea of this step is to mitigate

[9]Samples taken in steps of an autocorrelation time are thought to no longer be correlated.

some effects of over-dispersing the walkers. It was mentioned before that the walkers are deliberately not completely over-dispersed such that local-maxima and unphysical solutions are avoided. This cannot, however, be guaranteed in all cases and sometimes a few stray walkers will find local minima that are very difficult to escape from. In such conditions it is advantageous to prune the walkers and relocated them to the minima occupied by the majority of the walkers. This helps improve convergence and stops stray walkers from finding unphysical solutions. In addition, this will not have a very big effect on any good walkers caught up in the relocating process, since they will be relocated close to their original position anyway. This could be linked back to parallel tempering, which when the signal-to-noise is very good, can struggle in this manner.

Algorithm 2 Production run background fits

Take a subsample of the APOKASC red giants with good signal to noise covering wide range in $\nu_{\max}$
for Each star in sample i **do**
    Initialise walkers according to priors given in Table 3.1
    Run MCMC procedure on given star (burn-in then first production run)
    Compute $\hat{R}$ for production run
    if $\hat{R} < 1.01$ **then**
        Run final iteration and compute summary statistics
    else if $\hat{R} > 1.01$ **then**
        for $j \leftarrow 0, 19$ **do**
            Run additional production run with n iterations
            if $\hat{R} > 1.01 + 0.005j$ **then**
                Continue loop
            else
                Run final iteration and compute summary statistics
            end if
        end for
    end if
end for

3.5.3 Extraction of summary statistics

The subject of extracting summary statistics from posterior probability distributions is an important topic that warrants a little bit of discussion. On the one hand, a true Bayesian would say that the summary statistic is the posterior itself, that it contains all the relevant information about the parameter in question and no summary statistic can appropriately summarise it properly. However most people are accustomed to seeing some sort of summary value and uncertainty that appropriately summarises the quantities of interest. Therefore it is worth spending a little time to explain which quantities were adopted and the justification.

The first quantity to decide upon is the value of the summary statistic, i.e. the quantity to present, for which there are a few possibilities, namely the mean, median or mode of the posterior distribution. These different quantities are plotted for two distributions as an illustration of the advantages and pitfalls of using each one, in Fig. 3.4. The two distributions chosen were a Gaussian and an exponential distribution. The scale and location parameters do not matter, hence the lack of labels on the plot (as this is just to aid the explanation). As expected for a Gaussian distribution, the mean, median and mode all lie in the same place and take the same value, however for the exponential distribution this is not the case where the mode lies at zero and the mean and median lie further out in the tail.

An important feature of a summary statistic is that one wants it to be invariant under any transformation of the data, in other words if the log of the above distributions were to be taken then the value taken as the summary statistic should be in the same place. The only summary statistic that has this character under any transformation is the median due to its definition of being the middle of the distribution, whereas the mean and mode are influenced by the transformation. Therefore it was deemed best to adopt the median as the quantity of choice for the first part of the summary statistic.

Now that the "point estimate" has been decided upon, the representation for the uncertainties must be chosen. Equivalently there are also three possible estimates that we shall

consider: the standard deviation of the posterior, the interquartile range and the highest posterior density (HPD) 68.3% credible region. The standard deviation and interquartile range are both fairly self explanatory, for the interquartile range the 15th and 84th percentile are taken to represent the 68.3% according to 1 σ for a Gaussian distribution. The HPD is taken as the smallest region of the probability distribution that contains 68.3% of the probability density.

There is quite a difference between all the above measures; let us see how they compare given two different distributions: a Gaussian and a negative exponential. The results of choosing each of the different methods in conjunction with the median value are shown in Fig 3.5. For a Gaussian each method produces the same uncertainties, but for a negative exponential they are much more different. The point being made is that under the assumption the posterior distribution is Gaussian each method is identical and perfectly admissible, however in the event of a distinctly non-Gaussian or multi-modal distribution both the standard deviation and interquartile range are skewed by the non-Gaussian nature of the posterior. In the end the method needs to be able to cope with such distributions and so the HPD was chosen due to this property. Given the assessment of convergence applied to these data the posteriors would be expected to be Gaussian-like (if convergence is achieved), although there is no guarantee.

3.6 Cleaning the data

In order to extract the best priors possible from the data they need to be cleaned first, and this involves weeding out the poor fits. Another issue that needs to be tackled is label switching between the granulation components, because they are not required to be ordered during the fitting process and as a result a little more post-processing may be needed.

The cleaning started broadly with a cut made in both $\nu_{\rm max}$ and its uncertainty, of $\nu_{\rm max} > 5\mu$Hz and $\sigma_{\nu_{\rm max}} < 1$ μHz. This ensures that all of the fits are of a good enough quality

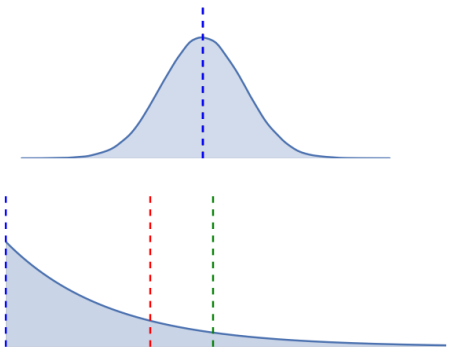


Figure 3.4: A comparison between using different summary statistics with regards to two different distributions, where the coloured line represents the value of the different quantities used. The red shows the median, the green shows the mean and the blue shows the mode.

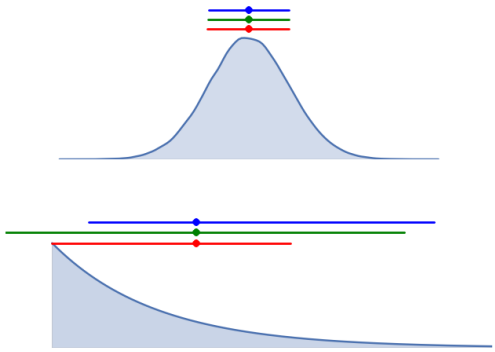


Figure 3.5: A comparison between using different summary statistics with regards to two different distributions, where the coloured point represents the median of the distribution and the coloured line represents the region encompassed within the calculated uncertainties. The red shows the combination of the median and highest posterior density (HDP) region, the blue shows the median and the interquartile range and the green shows the median and the standard deviation.

to be used in the analysis. Due to the decreasing effective frequency resolution $\Delta\nu/\nu_{\rm bw}$, low $\nu_{\rm max}$ stars will be more difficult to fit especially considering the additional rebinning of the data and so the cut was made at $\nu_{\rm max} > 5\mu$Hz. With regards to the uncertainties it would be expected for the high-quality nature of the data that the uncertainty of $\nu_{\rm max}$ should be very low and so $\sigma_{\nu_{\rm max}} = 1$ μHz is taken as the upper limit[10]. However this does not catch all bad fits as there could be some that have converged to a solution with a very low uncertainty but are completely unphysical due to artefacts in the power spectrum. This idea is covered next.

Unfortunately not all of the fits to the data can end in success, whether that is because of poor initialisation or unexpected features in the power spectrum, such as spikes or signatures of rotation that have not been properly removed from the time series. These can be very difficult to remove in a clean and quick way using cuts in parameters or their errors, so instead every fit was checked by eye to ensure no poor fits slipped through. An example of such a poor fit is shown in Fig 3.6 where there are multiple additional spikes present that will throw off the fit, these are most likely due to background classical pulsators and other sources of contamination[11]. In addition, there seem to be multiple contaminating stars and this is a problem in some red giant data. Ideally when $\nu_{\rm max}$ is close to the Nyquist frequency the model should be adjusted to cope with the power that is reflected. Due to the small fraction of stars with such a high $\nu_{\rm max}$ this feature was not included, but in the majority of cases the $\nu_{\rm max}$ can be extracted easily. It would be obvious later on in the analysis if stars are in fact super-Nyquist as their parameters would not lie on the expected relation.

Label Switching

Due to the fact that the main two granulation components have identical priors, it is possible for the two components to switch places and the second component take the place of

[10]This is also chosen due to some trial and error as to when the fits have failed and when they have succeeded. In almost all cases where the fit has succeeded the uncertainties lie well below 1 μHz.

[11]Approximately 3% of fits were subject to such contamination.

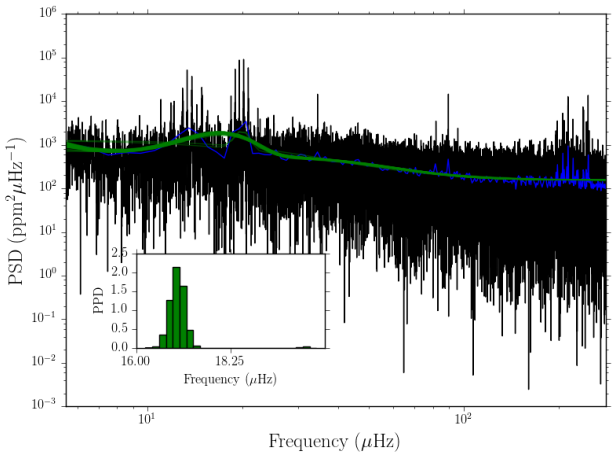


Figure 3.6: A poor fit to the power spectrum of KIC 1720241. The histogram of $\nu_{\rm max}$ shows that it has not converged well to a solution and the many spikes in the data can be seen.

the first component at low frequency and vice versa. In addition the third component for "activity" can also take the place of the first granulation component. Therefore this needs to be addressed as otherwise any inference drawn from the population distributions would not accurately represent the data.

One way of dealing with the issue of label switching is by reordering the characteristic frequency such that $\beta < b_1 < b_2$ and then the amplitudes and exponents accordingly. Figure 3.7 shows the data for the first granulation component amplitude as a function of $\nu_{\rm max}$ before and after applying this reordering in characteristic frequency. It becomes apparent where the first granulation component is not needed in the fit due to its very low amplitude (for a $\nu_{\rm max} < 100$ μHz). The reason that σ_1 is used instead of σ_2 is due to a strange feature in some of the stars (namely red clump stars), whereby the second granulation frequency appears to take on a value closer to $2\times\nu_{\rm max}$ rather than the expected value of $\sim\nu_{\rm max}$ and so this makes the plot much more messy and so σ_1 is used as a proxy instead.[12]

This method of dealing with the label switching does not fully resolve the problem because the situation now occurs where the "activity" component takes the place of the first granulation component when it should not. A proposed idea is to look at the relative size

[12]The stars showing this strange feature will be discussed in more detail later in this section.

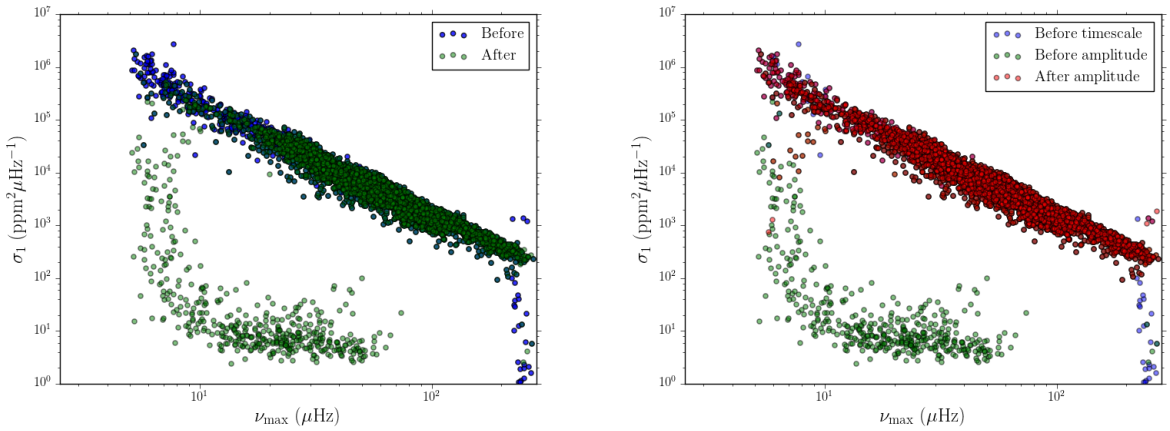


Figure 3.7: The left-hand plot shows amplitude of the first granulation component as a function of $\nu_{\max}$ shown before and after ordering the timescales such that the activity component is less than the first component which is less than the second component. The right-hand plot is the same as that on the left but added in red is the first component amplitude having applied both the timescale and amplitude ordering.

of the amplitudes of each granulation component and if $\sigma_1 < \sigma_2$ the activity component is taking the place of the first component and so the two amplitude values can be swapped as a correction. This is not the only situation in which this could happen, the other situation would be if σ_2 was pushed very low and the other two components essentially "moved" up a label to account for this. However this cannot happen due to a prior place on the maximum frequency of the activity component, $\max(\beta) \leq 50\mu$Hz, the main source of label switching is between the activity and first components.

The culmination of this preprocessing is demonstrated in Fig 3.7 whereby the effect of exchanging the characteristic frequencies and then the amplitudes can be seen and the extent to which this cleans up the data.

3.7 Non-parametric priors

Having fitted all of the stars in our sample[13], run the preprocessing steps and removed any poor fits[14], the next step was to extract the relevant information from the data so that

[13]The total size of the sample was 6661 stars.
[14]300 stars were removed from the sample due to poor fits.

they can be implemented as prior distributions in subsequent fits. Rather than creating 1-dimensional prior distributions, e.g. $p(\nu_{\rm max})$ or $p(H_{\rm env})$, N-dimensional priors were instead created over all the parameters in the model to fully capture the diverse behaviour of each parameter.

The prior probability distribution over all parameters can be written as the joint distribution $p(\theta) = p(\theta_n, \theta_{n-1}, ..., \theta_1)$. Independence between each parameter is normally invoked which results in the simplified form $p(\theta) = p(\theta_n)p(\theta_{n-1})...p(\theta_1)$, however in this case the correlations and dependencies on the parameters are wanted as they can provide additional prior knowledge in the fitting process.

In order to produce these priors a supervised machine learning scheme known as Gaussian Mixture Models (hereafter GMMs) was used, which will be described in more detail in the next section (see e.g. Bishop 2006 for a more comprehensive overview). An assumption has been made when producing the plots shown in Figs 3.8 and 3.9 that modelling the data using 2D Gaussians is equivalent to looking at the 2D projection of the N-dimensional Gaussians fitted to the entirety of the data; this is for simplicity and ease of visualisation. Therefore a given number of (in our case 2D) Gaussians are fitted to the data using the expectation-maximisation (EM) algorithm. A brief description of the mixture model will be given in Appendix A.8.1[15].

3.7.1 Application to *Kepler* data

The data extracted from the fits to the *Kepler* data were preprocessed as detailed above and the poor fits extracted by eye. Having done this, each parameter can be plotted as a function of $\nu_{\rm max}$ and modelled according to the best-fitting mixture of 2D Gaussians (as demonstrated in Figs 3.8 and 3.9). Performing this sort of analysis enables the large-scale structure of the parameter space to be investigated and additional trends identified. For

[15]The python package `scikit-learn` was used to perform this inference.

example, all parameters apart from the exponents of the granulation components show a power law dependence akin to those found in previous works. In addition, extra structure can be seen that was not previously identified. An example is in the plots showing b_2, where a second strand of data can be seen that is offset from the main trend in the data. This is certainly very interesting and further investigation will be left to later in this chapter.

The model fitted using the GMMs can be used in a variety of ways and enables checks to be made when new stars are fitted to identify whether they are in the expected part of parameter space. The parameter space could be sampled from and this used as a more robust means of estimating $\nu_{\rm max}$ (as will be shown later). Another potential application is in initialising the walkers in a more physically motivated manner, which comes about because once a $\nu_{\rm max}$ value has estimated the conditional probability of some parameter given the estimate $\nu_{\rm max}$ value, $p(\theta|\nu_{\rm max})$, can easily be calculated for the fitted mixture of Gaussians.

3.8 Exploring the results from the *Kepler* fit

Here we look in detail at several parameters of interest and explore the relationship between these parameters and other known stellar parameters, such as metallicity, i.e. [Fe/H], effective temperature and evolutionary state.

A good place to start is the height of the oscillation envelope. This is due to the obvious difference in the values when coloured by evolutionary state (e.g. Elsworth et al. 2016), as shown in Fig 3.10. It immediately becomes apparent that the second strand in $H_{\rm env}$ is caused by the red clump and secondary clumps stars. Coincidentally this is also the case for the second granulation component amplitude and width of the oscillation envelope. There is also an interesting feature present in the second granulation frequency. There is an assumed scaling between b_2 and $\nu_{\rm max}$ which is approximately linear (from Kallinger et al. 2014) which accounts for the stars at $b_2/\nu_{\rm max} \approx 1$. However the accumulation of stars (the majority of which are red clump) at values of $b_2/\nu_{\rm max} > 1$ is an interesting feature that warrants more

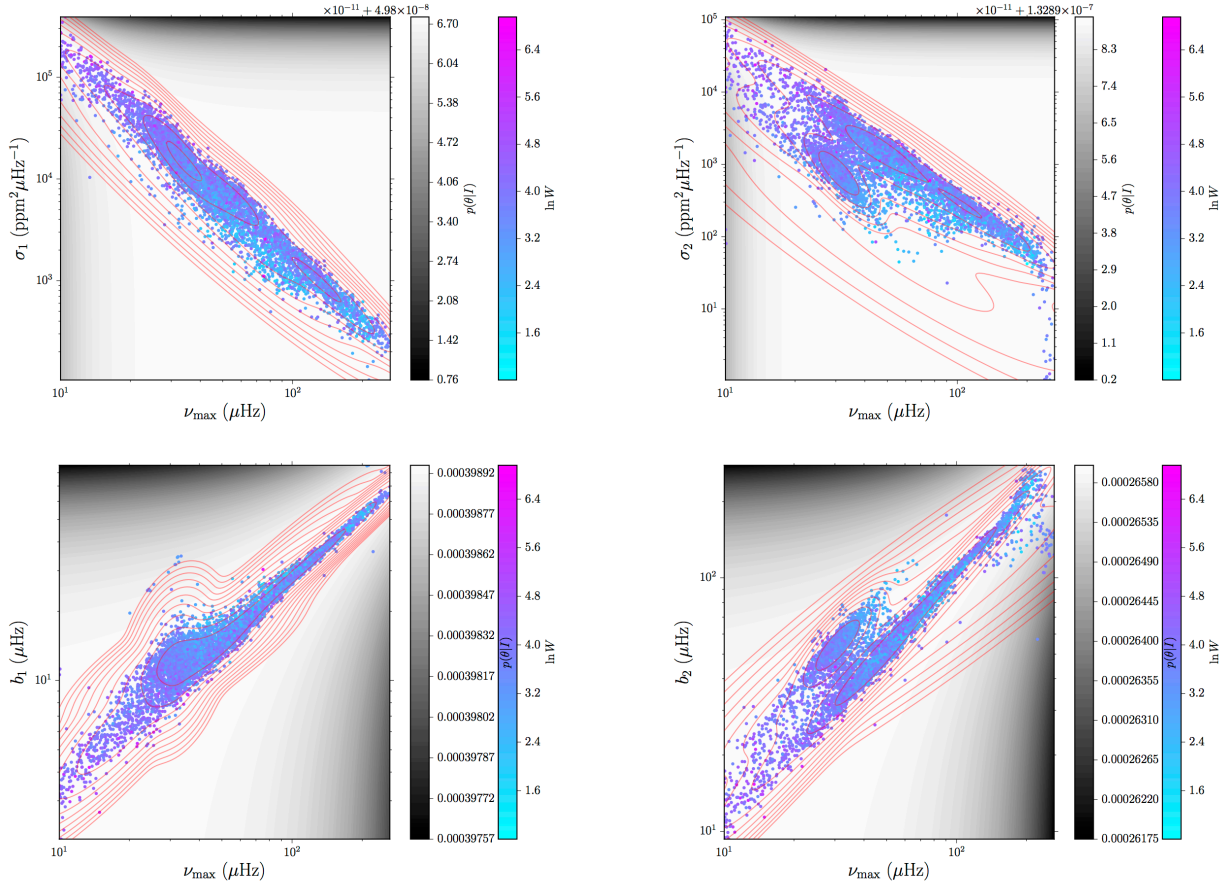


Figure 3.8: Plots of the background fit parameters as a function of $\nu_{\max}$ coloured by the natural logarithm of their white noise value. The orange contours show the mixture of Gaussians fitted to the data whilst the background contours show the uninformative priors applied to the data in the background fitting.

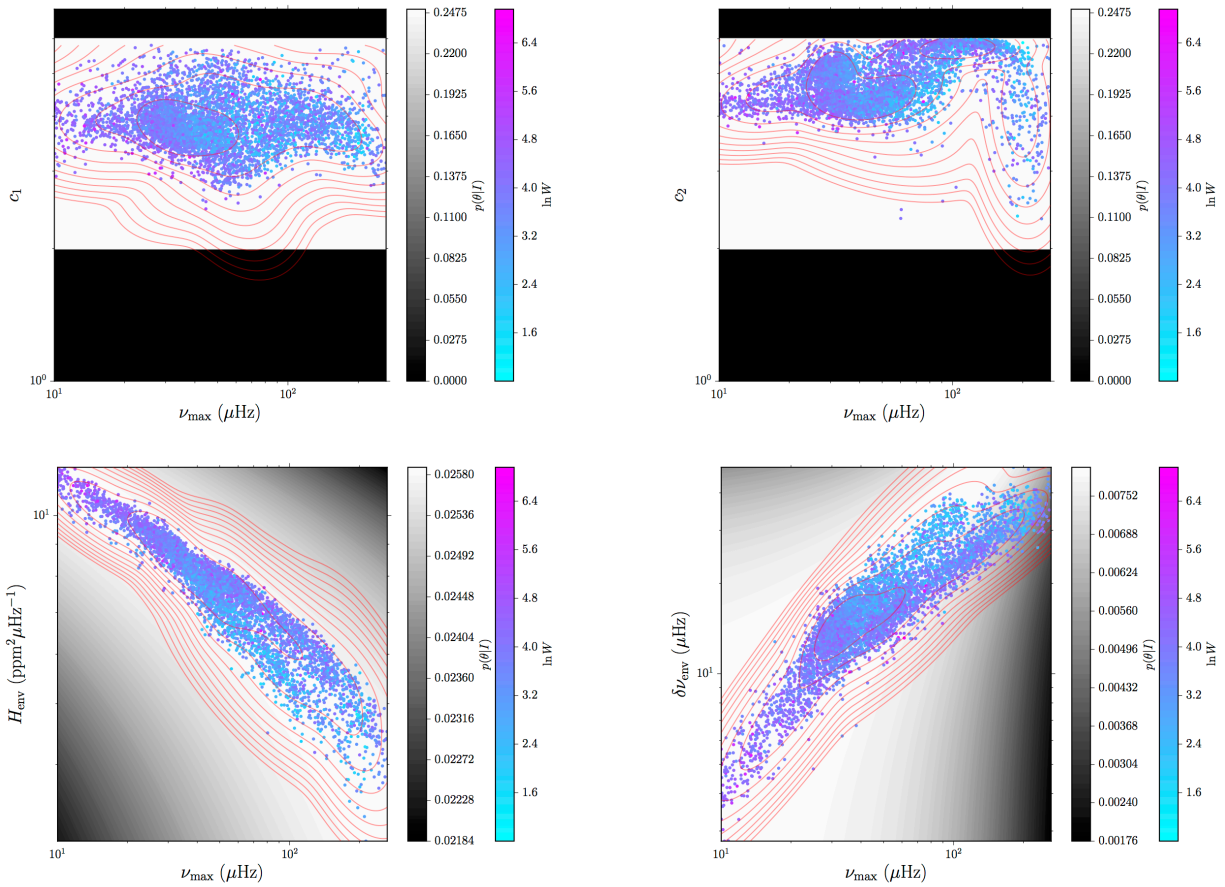


Figure 3.9: Same as Fig 3.8.

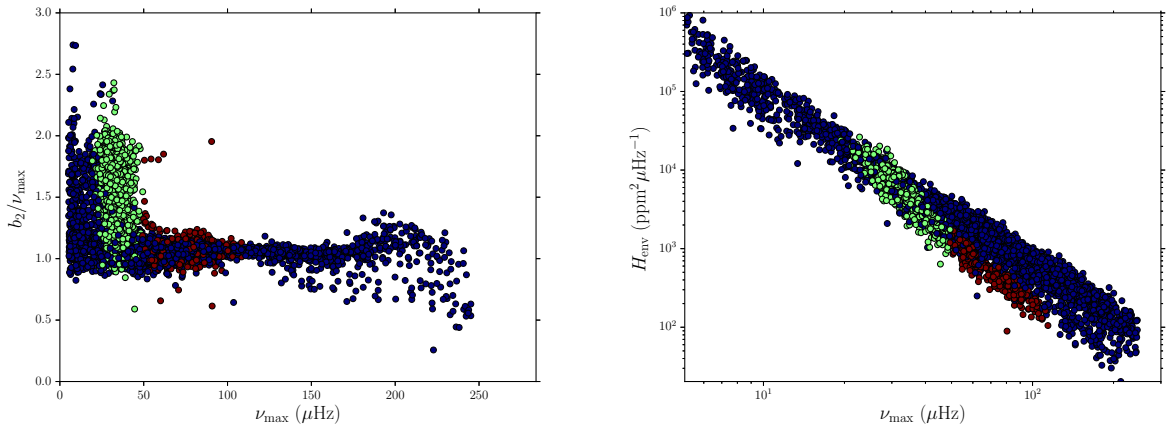


Figure 3.10: The ratio of the second granulation frequency plotted as a function of $\nu_{\rm max}$ is shown on the left panel coloured by evolutionary state (blue is RGB, green is RC and red is SC). The right-hand side plot is the height of the oscillation envelope plotted as a function of $\nu_{\rm max}$ coloured by evolutionary state as well.

exploration. There are also some RGB stars that also show this feature and so this may be evolutionary state dependent, but again will require more investigation.

The easiest way in which to observe the odd granulation profile of some RC stars is to compare them to those where no oddity is expected, i.e. RGB stars with a similar $\nu_{\rm max}$. Therefore the fits were scrutinised for three RC stars and three RGB stars (with similar $\nu_{\rm max}$ values), whose data are shown below in Figs 3.11 and 3.12. These data were compared as follows: firstly the background was reconstructed from our background fits and another set constructed using the scaling relations given in Kallinger et al. (2014), then a rebinned version of the power spectrum was divided through by each model and the resulting signal-to-noise spectrum inspected.

Good agreement can be seen between the two different models in the case of red giant branch stars (Fig 3.12) which is reassuring. However, when applied to some red clump stars the differences between the models become apparent and clearly the scaling relations do not do a very good job. The key feature in Fig 3.11 is the slight "knee" seen just beyond the oscillations that could be indications of another granulation component. Whether this is a separate granulation component to the two already included or simply an increased

characteristic frequency is not clear and would require further research that will be performed in the future.

An explanation for this additional "knee" could therefore be sought in terms of differences in stellar properties such as effective temperature or metallicity. However for all intents and purposes the two populations (those showing normal granulation properties and those showing a high b_2) are very similar. Fig 3.13 shows the distributions of $\nu_{\max}$, $\Delta\nu$, [Fe/H] and $T_{\rm eff}$[16] for the two different sets of stars. The cut-off for a star being classed as "weird" was made at $b_2/\nu_{\max} > 1.2$ since it is expected that b_2 should lie close to $\nu_{\max}$. It can be seen that there is very little difference in $T_{\rm eff}$ between the two samples and the metallicity distributions are fairly similar. The main difference comes in slightly different $\Delta\nu$ and $\nu_{\max}$ distributions which could imply different mass and radius distributions (both of which would require detailed modelling to provide). The difference between the two distributions is slight and may hint at discrepancies since there may be slight excesses at low masses for the "weird" stars and at high masses for normal stars. However this is not enough to adequately separate the distributions, therefore more work will need to be done in order to obtain the cause of such properties[17].

Although any solid conclusions cannot be drawn from these initial observations this is certainly an area that needs investigating (in the future, but will not be done here), firstly through model comparison to see if this is a genuine third component or whether the second component characteristic frequency is changing (which makes less sense) and then to understand the physical mechanisms behind what is happening.

[16] $\Delta\nu$ was taken as the `COR_DELTA_NU` column from the APOKASC catalogue, [Fe/H] was taken as the `DR12_FE_H` column and $T_{\rm eff}$ was taken as `DR12_TEFF_COR`.

[17] The masses given in the catalogue were not used as there were not enough to adequately characterise the sample, only 1686 of the 5957 stars in the sample had masses.

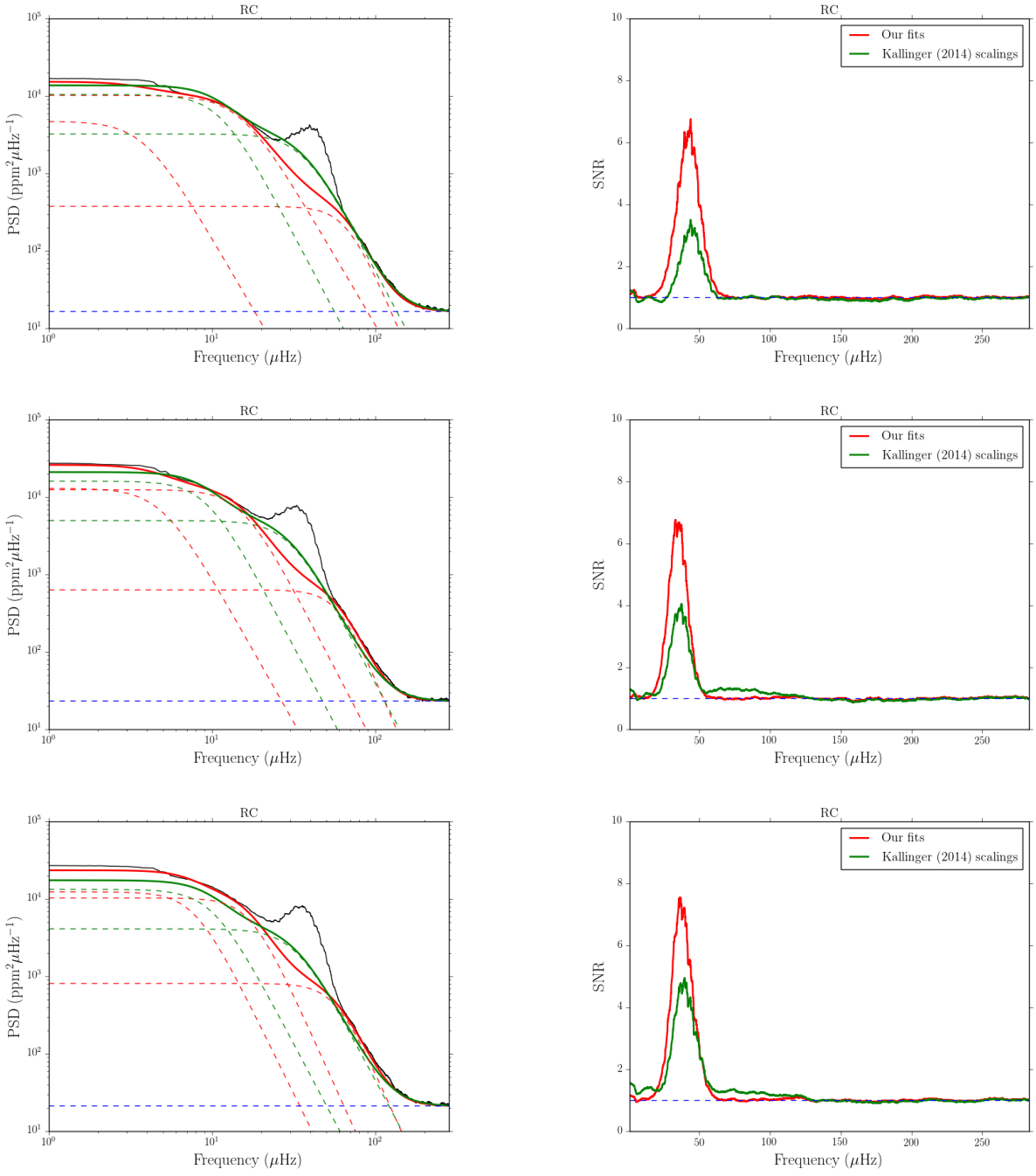


Figure 3.11: Comparison plots of the background models for those fitted in this work (in red) and those constructed from the scaling relations (green) for three red clump stars KIC 8489832, KIC 8611114 and KIC 10683647. The left column shows the background models and the right column shows the signal-to-noise spectrum.

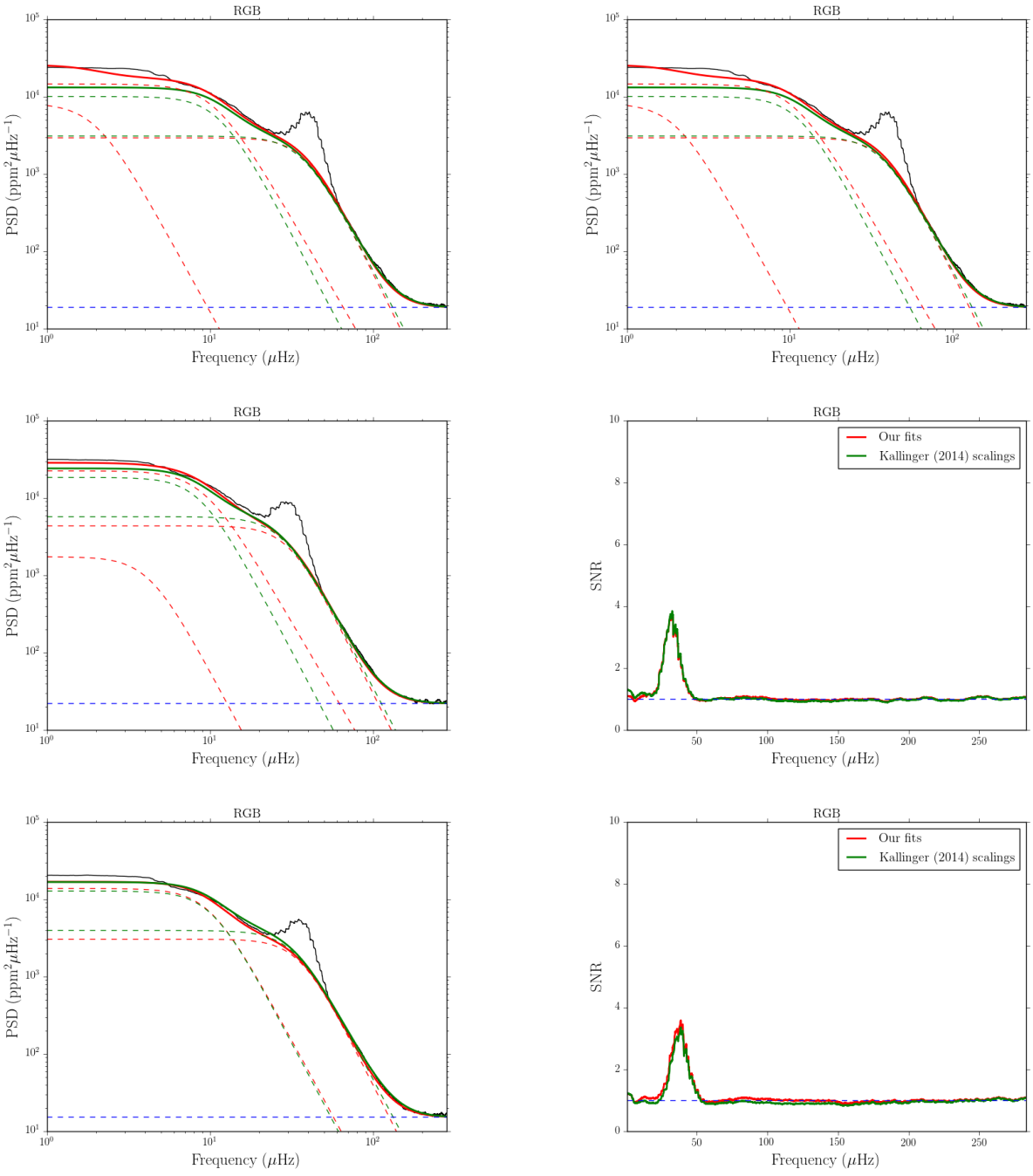


Figure 3.12: Same as 3.11 but for 3 RGB stars with similar $\nu_{\rm max}$ values.

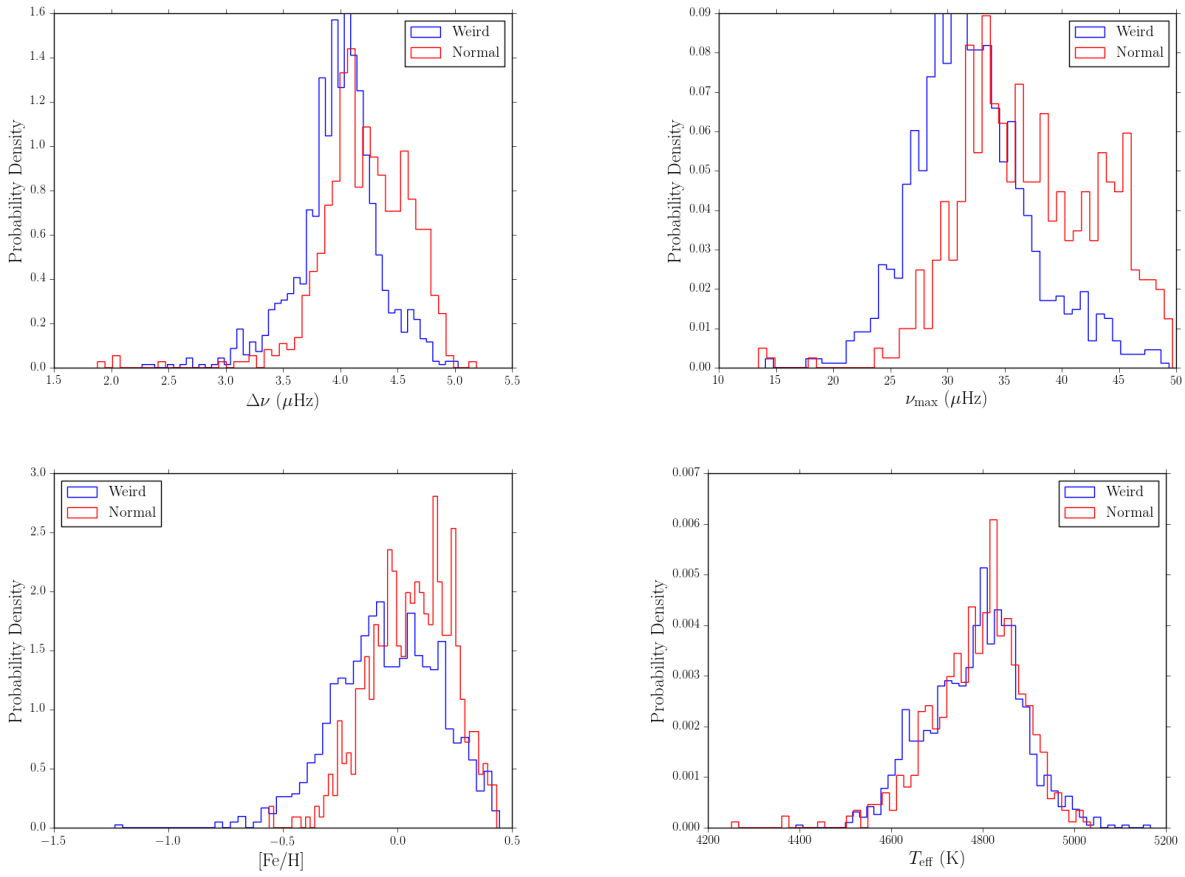


Figure 3.13: Comparison histograms of the two samples of clump stars made at a cut-off of $b_2/\nu_{\rm max} > 1.2$ to be considered part of the "weird" sample.

3.8.1 $\delta\nu_{\rm env}$ AND $H_{\rm env}$

There are established scaling relations for many asteroseismic quantities, such as the width of the oscillation envelope and its height, however these were extracted using just under 2 years of *Kepler* data rather than the full 4 years. The parameters extracted from the data presented here provides an opportunity to test and update these scaling relations, including evolutionary state information. The first scaling relation to test is that for the width of the oscillation envelope $\delta\nu_{\rm env}$, which is currently given by (Mosser et al., 2012a)

$$\delta\nu_{\rm env} = 0.66\nu_{\rm max}^{0.88}. \tag{3.11}$$

This relation is plotted over our data in Fig 3.14. It is immediately apparent that it is inconsistent with our data due to the presence of two populations in the data. This is because the scaling relation is only valid for stars with $\nu_{\rm max} < 100$ μHz, however even in this case it is clear that the secondary clump stars are the main cause for this apparent bias towards larger envelope widths. Now that there are many more targets and 4 years worth of data it is prudent to derive our own relations given the data in order to provide better consistency with the data now available. This is done in a Bayesian manner and taking advantage of the power law scalings observed in the data.

The model that will be used is as follows

$$\log_{10}(\delta\nu_{\rm env}) = \alpha + \beta\log_{10}(\nu_{\rm max}), \tag{3.12}$$

where α and β are the parameters of the power law. Assuming Gaussian uncertainties the overall model can be given by

$$\log_{10}(\delta\nu_{\rm env}) \sim \mathcal{N}(\mu = \alpha + \beta\log_{10}(\nu_{\rm max}), \sigma^2). \tag{3.13}$$

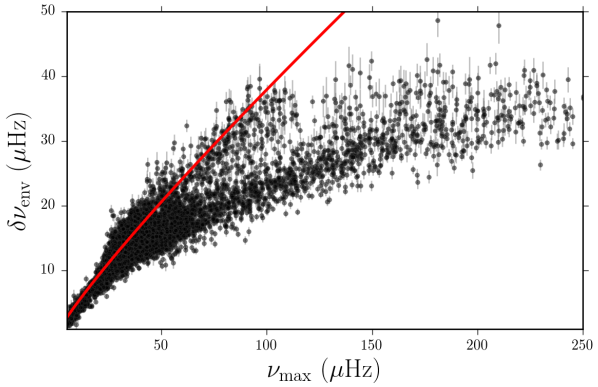


Figure 3.14: The width of the oscillation envelope as a function of $\nu_{\rm max}$ for the stars fitted in this chapter. The red line is the scaling relation given in 3.11 which is valid for stars with $\nu_{\rm max} < 100$ μHz.

The variance σ^2 is not quite as simple as the uncertainties on $\delta\nu_{\rm env}$, but must also include uncertainties on $\nu_{\rm max}$, therefore

$$\sigma_i^2 = \sigma_{y,i}^2 + \beta^2\sigma_{x,i}^2 + S^2, \tag{3.14}$$

where S^2 is a free parameter in the model and accounts for the intrinsic scatter present within the data. The idea here is to give some idea as to the underlying scatter in the data in addition to the measured uncertainties. In reality, S is a nuisance parameter to mop up the excess scatter to enable a better fit, however the value could be important from a physical perspective by giving a magnitude to the extra scatter which could come from physical sources (such as effective temperature or metallicity).

This can be applied to the data in a few different ways, firstly to the entire sample in order to try and reproduce the work of Mosser et al. (2012a) and secondly just using the RGB sample (using evolutionary state classifications from Elsworth et al. 2016) and thirdly to the RC and SC sample. If there are any differences between the samples in terms of a fitted relation then it may make more sense to have a relation for each evolutionary state rather than an incorrect relation for all stars.

A summary of the fits is shown in Figs 3.15 and 3.16 and the parameters fitted are given

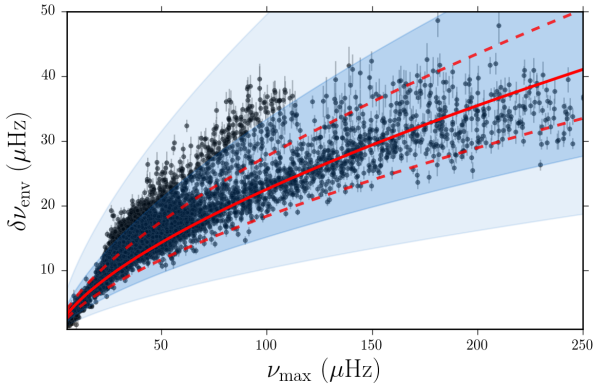


Figure 3.15: The fit to the full sample where the red line shows the power law relation with the 1σ calculated from the residuals shown by the red dotted lines. The blue shaded regions show the 1σ and 2σ bounds including the additional modelled variance term.

in Table 3.4. When fitting the combined sample the fitted relation is very different giving

$$\delta\nu_{\rm env,total} = (1.113 \pm 0.007)\nu_{\rm max}^{(0.654\pm0.004)}, \tag{3.15}$$

It is clear from Fig 3.15 that the higher $\nu_{\rm max}$ stars (RGB) are contributing the most to keeping the exponent low. In addition, it can be seen that the secondary clump stars are mainly responsible for the necessary increase in variance, indicative of the fact that the RGB and clump stars should be treated as two separate populations. This justifies the modelling of the RGB and clump populations separately, it can also be seen as the drop in variance when the two populations are modelled separately.It should also be noted that it appears as though for the same $\nu_{\rm max}$ stars that are helium core burning seem to have a larger $\delta\nu_{\rm env}$ which would result in more detected orders. The reasons for this are not clear and warrant further investigation in future work.

Clearly from both the above fits and the plots presented, the relationship between $\nu_{\rm max}$ and $\delta\nu_{\rm env}$ is very different for RGB and clump stars. It should be noted that in the case of the RGB sample there are a few stars with a $\nu_{\rm max}$ just below 100μHz that seem to have a larger envelope than expected that might therefore be misclassified secondary clump stars.

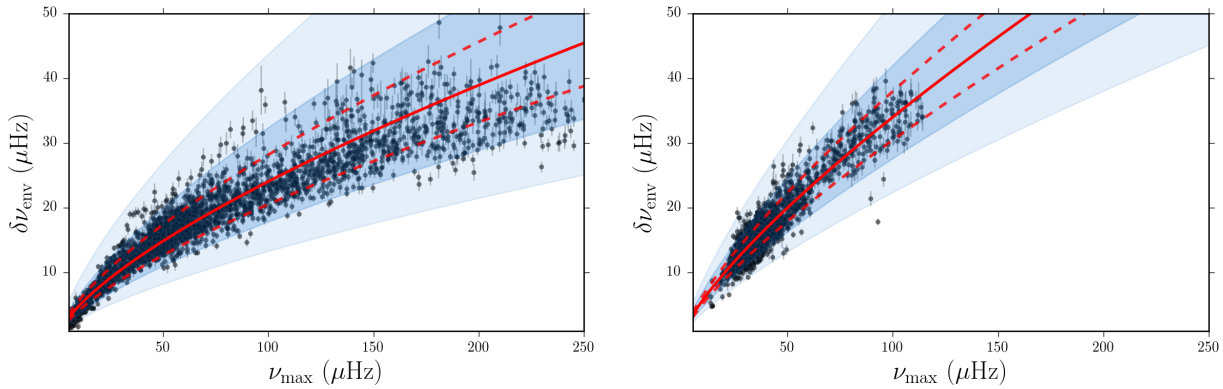


Figure 3.16: The same as Fig 3.15 where the left hand panel shows the RGB sample and the right hand panel shows the RC and SC sample.

The purpose of adding in the additional variance term is to help capture artefacts such as those so that the fit is unaffected. Interestingly, in all cases the additional variance term provides a non-negligible contribution to the fit. It appears that there is scatter in the relations themselves that cannot simply be described in terms of the uncertainties on each data point. This makes sense, since one would expect scatter in the relations due to the underlying physics. In order to account for this (in a basic manner) dependencies on physical parameters, such as [Fe/H] or $T_{\rm eff}$ can be explored. Studying the residuals of the fit (as shown in Fig 3.17), a slightly non-trivial dependency of $\delta\nu_{\rm env}$ on $T_{\rm eff}$ can be seen. There is already a known temperature relation with $\nu_{\rm max}$ [18] (since it is proportional to the ratio of the sound speed and the pressure scale height) and so this will be the dominant trend seen in the $T_{\rm eff}$ residuals. This trend could easily be factored into the variance calculation above in order to reveal any other observed smaller trend, however that is beyond the scope of this work.

The residuals are not flat for the RGB sample and there is some structure present, especially at both low and high $\nu_{\rm max}$. The deviations at low $\nu_{\rm max}$ will be a result of the decreasing frequency resolution and the fit struggling at lower frequencies. At high frequency this is more likely due to the non-uniform distribution of stars with respect to $\nu_{\rm max}$, which instead

[18] $\nu_{\rm max} \propto g/\sqrt{T_{\rm eff}}$

Table 3.3: Results of the fitting procedure for $\delta\nu_{\rm env}$.

Parameters	Total	RGB	Clump (RC and SC)
α (μHz)	1.112 ± 0.007	0.986 ± 0.007	1.00 ± 0.01
β	0.654 ± 0.004	0.694 ± 0.004	0.766 ± 0.007
$\log_{10}(S)$	0.082 ± 0.001	0.061 ± 0.001	0.043 ± 0.01

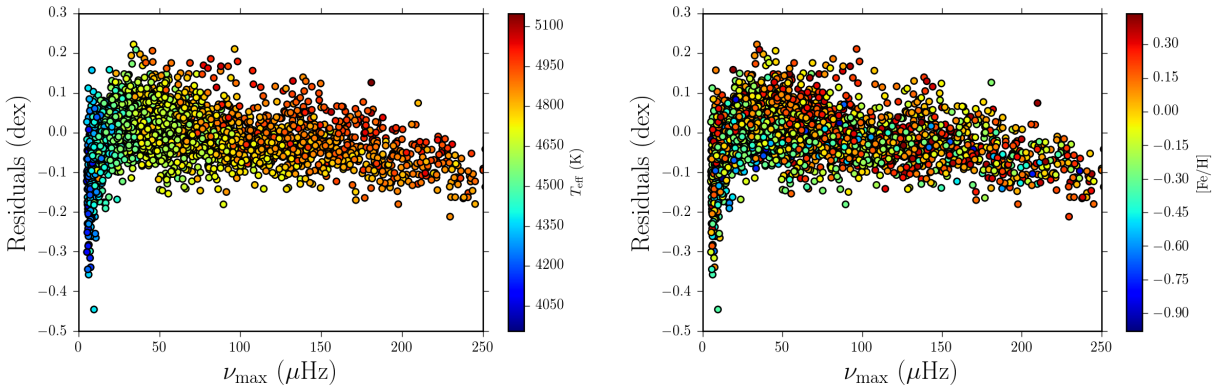


Figure 3.17: The residuals of the power law for $\delta\nu_{\rm env}$ (in log-space) as a function of $\nu_{\rm max}$ for the RGB sample coloured by $T_{\rm eff}$ (left-hand plot) and [Fe/H] (right-hand plot).

peaks close to 50μHz (the RGB bump). The apparent "saturation" in $\delta\nu_{\rm env}$ was noted by Mosser et al. (2012a) and the deviation from their scaling relation was explained possibly through a change in the relation between $\nu_{\rm max}$ and the acoustic cut-off frequency. However in our case with more data this feature is less apparent and is more likely due to having fewer stars in that part of the parameter space (due to the relative speed of that particular phase of the star's evolution).

When looking at the residuals of the RC and SC sample, Fig 3.18 they show no obvious structure indicating a much better fit as the sample conforms to the expected power law relation. There may be a small apparent trend in $T_{\rm eff}$ which is anti-correlated with the residuals, however this could appear more prominent due to the presence of the (much hotter) secondary clump at higher $\nu_{\rm max}$. Also in [Fe/H] there do not seem to be any apparent trends in the residuals which shows that the power law relation with $\nu_{\rm max}$ seems to be adequate for the majority of analyses.

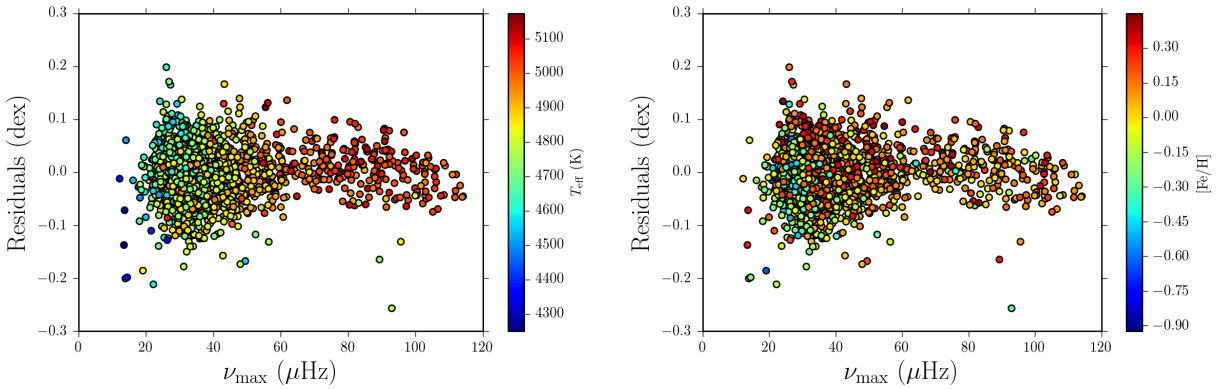


Figure 3.18: The residuals of the power law for $\delta\nu_{\rm env}$ (in log-space) as a function of $\nu_{\rm max}$ for the RC and SC sample coloured by $T_{\rm eff}$ (left-hand plot) and [Fe/H] (right-hand plot).

The other parameter that can be investigated is the height of the oscillation envelope, $H_{\rm env}$. The archival scaling relation is given by (Mosser et al., 2012a)

$$H_{\rm env} = 2.03 \times 10^{7} \nu_{\rm max}^{-2.38}. \tag{3.16}$$

The same method is adopted as for the envelope width including the extra variance term to account for any intrinsic scatter in the data. This resulting scaling relations are given in Figs 3.19 and 3.20 whilst the parameter values are displayed in Table 3.4.

Again the parameters fitted to the entire sample are inconsistent with the previously derived scaling relations with a shallower power law and lower multiplicative factor. The intrinsic scatter of the whole sample is dominated by the RGB as opposed to $\delta\nu_{\rm env}$ whereby the added variance was caused by the clump following a very different relation. One of the reasons for this large scatter is caused by the observed increase in scatter towards higher $\nu_{\rm max}$ values. It is known that the effects of binarity for mass fractions close to 1 can cause the observed oscillation power to sit lower than expected as a result of the washout caused by the companion. As a result the increased scatter could be a result of this as well as modes showing some form of suppression (e.g. Fuller et al. 2015; Stello et al. 2016; Cantiello et al. 2016), although this would not be symmetric since it would reduce the height of the envelope.

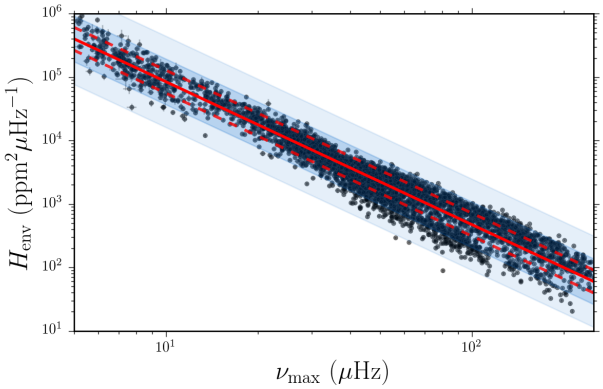


Figure 3.19: The fit to the full sample where the red line shows the power law relation with the 1σ calculated from the residuals shown by the red dotted lines. The blue shaded regions show the 1σ and 2σ bounds including the additional modelled variance term.

In contrast to $\delta\nu_{\rm env}$, the residuals for the RGB sample show much better agreement with the power law form used to fit the data and there is much less structure present. There are some additional trends present in $T_{\rm eff}$ for the RGB sample but not in [Fe/H]. The trends seen in temperature are expected and those stars with larger $T_{\rm eff}$ are expected to show lower oscillation amplitudes (as shown later), however the trend seems to be a composite of the expected trend with $\nu_{\rm max}$ and this additional factor. Therefore an additional temperature term may be needed to properly quantify the temperature gradient observed in the data, that is accounted for in the amplitudes of the oscillations $a_{\rm max}$.

The clump sample shows much more structure in the residuals suggesting that the RC and SC obey different relations with regards to the height of the envelope. There also appears to be a trend in both $T_{\rm eff}$ and [Fe/H] which seem to be anti-correlated (the link between the amplitude of oscillations and metallicity has been seen before, e.g. Samadi et al. 2010 on CoRoT targets). This also suggests the possibility that there are additional factors needed to properly model the height of the envelope along with $\nu_{\rm max}$. The difference in properties between the RC and SC also become more apparent and in the future should be kept separate when considering the amplitude of the oscillations.

An important parameter linked to $H_{\rm env}$ is the maximum amplitude of oscillation $a_{\rm max}$

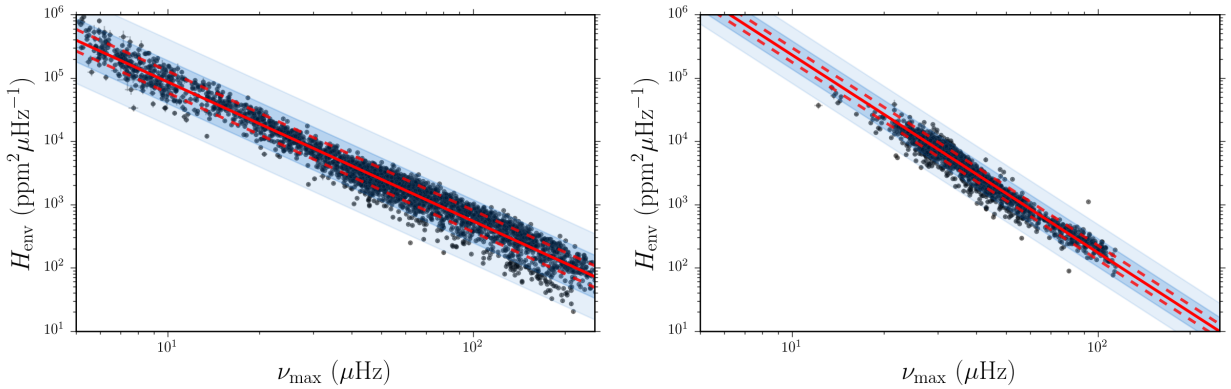


Figure 3.20: The same as Fig 3.19 where the left hand panel shows the RGB sample and the right hand panel shows the RC and SC sample.

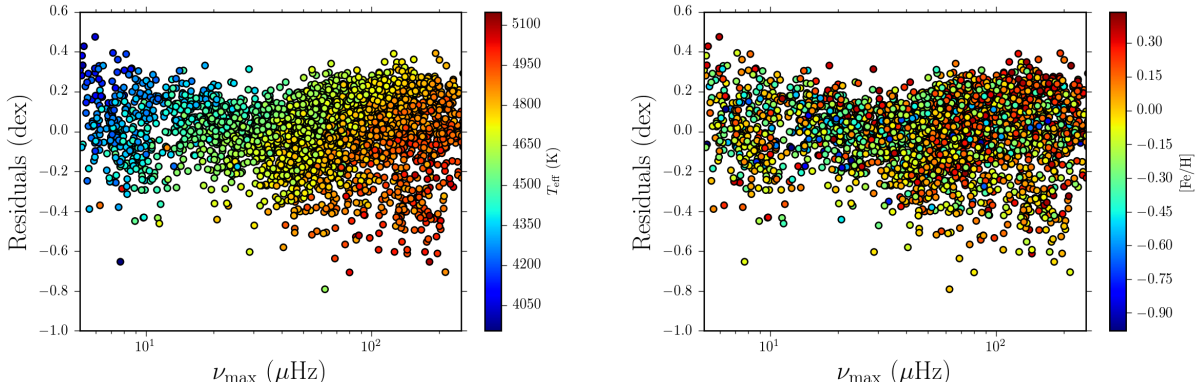


Figure 3.21: The residuals of the power law for $H_{\rm env}$ (in log-space) as a function of $\nu_{\rm max}$ for the RGB sample coloured by $T_{\rm eff}$ (left-hand plot) and [Fe/H] (right-hand plot).

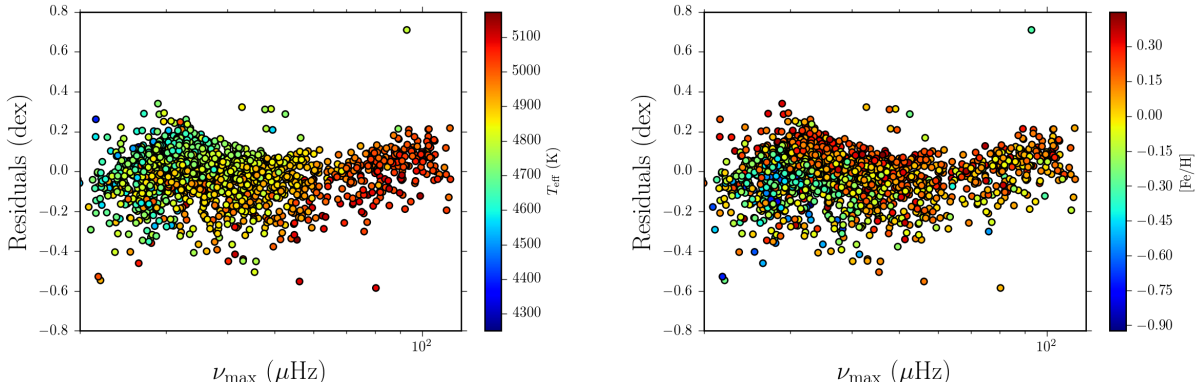


Figure 3.22: The residuals of the power law for $H_{\rm env}$ (in log-space) as a function of $\nu_{\rm max}$ for the RC and SC sample coloured by $T_{\rm eff}$ (left-hand plot) and [Fe/H] (right-hand plot).

Table 3.4: Results of the fitting procedure for $H_{\rm env}$.

Parameters	Total	RGB	Clump (RC and SC)
α ($\times 10^7 {\rm ppm}^2 \mu{\rm Hz}^{-1}$)	1.53 ± 0.05	1.37 ± 0.04	31.0 ± 2.0
β	-2.255 ± 0.008	-2.201 ± 0.008	-3.12 ± 0.02
$\log_{10}(S)$	0.180 ± 0.002	0.170 ± 0.002	0.110 ± 0.002

which is defined as follows

$$A_{\rm max} = \sqrt{\frac{H_{\rm env} \Delta\nu}{\xi}}, \tag{3.17}$$

where ξ is the total visibility of the modes ($\xi = 3.14$) which is explained in more detail in chapter 4. The reason for also looking at $a_{\rm max}$ is that it is more closely linked to the stellar parameters themselves rather than being an inferred, convenient term in background fitting.

The relation used is that in Campante et al. (2014) and designed to show the differences between the populations and place some constraints on the red giant stars that are consistent with previous works (e.g. Corsaro et al. 2013). This is as follows

$$\left(\frac{A_{\rm max}}{A_{\rm max,\odot}}\right) = \beta \left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}}\right)^{-s} \left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^{3.5s-r}, \tag{3.18}$$

where β is a factor that means the amplitudes do not need to pass through the solar value, $A_{\rm max,\odot} = 2.5$ppm is the solar maximum amplitude and $\nu_{\rm max,\odot} = 3090 \mu$Hz is the solar $\nu_{\rm max}$ value. This formulation is advantageous because rather than involving quantities such as the luminosity and mass, only observable quantities are needed which reduces the uncertainties in the derived amplitudes. This can in turn be linked to the theoretical work (e.g. Belkacem et al. 2011a) which makes use of the relation

$$A \propto \left(\frac{L}{M}\right)^{s} \propto \nu_{\rm max}^{-s} T_{\rm eff}^{3.5s}. \tag{3.19}$$

The form of the likelihood function will be similar to that used above but will require a

Table 3.5: Results of the fitting procedure for $A_{\rm max}$.

Parameters	RGB	Clump (RC and SC)
β (ppm)	0.932 ± 0.02	0.107 ± 0.004
s	0.654 ± 0.004	1.17 ± 0.01
r	8.20 ± 0.15	6.48 ± 0.10
$\log_{10}(S)$	0.041 ± 0.002	0.048 ± 0.001

few additional terms due to the addition of the $T_{\rm eff}$ in the relation. First of all, let us rewrite the model such that it is linear in log-space:

$$\log_{10}(A_{\rm max}) = \log_{10}(\beta) - s\log_{10}\left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}}\right) + (3.5s - r)\log_{10}\left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right) + \log_{10}\left(A_{\rm max,\odot}\right). \quad (3.20)$$

Due to the $T_{\rm eff}$ term, the total uncertainties need to be incorporated into the likelihood function following

$$\sigma^2_{tot} = \sigma^2_{\log_{10}(A_{\rm max})} + [(\log_{10}\beta)^2 + s^2]\sigma^2_{\log_{10}(\nu_{\rm max})} + (3.5s - r)\sigma^2_{\log_{10}(T_{\rm eff})} + \sigma^2, \quad (3.21)$$

where the errors in the logarithm of each quantity are needed, and σ^2 denotes the term describing the intrinsic scatter in the relation. The data can then be fitted using the same method as above (MCMC) and using the different subsets to gauge any differences between the populations.

The results for the fits to the two samples are shown in Table 3.5. The fitted parameters can be converted into the quantities used in the expression for $A_{\rm max}$, s and $3.5s - r$ in order to look at the $\nu_{\rm max}$ and temperature dependencies. For the RGB sample $s \approx 0.654$ which is slightly lower than the same value found for main-sequence stars in Campante et al. (2014), but this still indicates that the $\nu_{\rm max}$ dependence on the amplitude is similar although not

consistent (within uncertainties). On the contrary $3.5s-r \approx -0.94$ in Campante et al. (2014) whereas in our RGB sample $3.5s - r \approx -7.2$ which is a much larger inverse dependence with temperature. The same is the case for the clump sample where the temperature dependence has an exponent of ~ -2.4, which is much less steep than the RGB. However the reason for this can be seen in Fig 3.23 where the RC and SC clearly show two different relations and by fitting the two together this will result in a shallower slope in $\nu_{\rm max}$ and therefore affect the temperature dependence through s.

For the RGB sample the residuals shown in Fig 3.24 show very little obvious structure other than a few points at high $\nu_{\rm max}$ (which could be binaries or have supressed modes of oscillation). The temperature gradient appears to decrease with $\nu_{\rm max}$ as expected and any other temperature dependence has been effectively removed (unlike in the case of $H_{\rm env}$ where no temperature term was used). This also leads to an interesting gradient seen in the residuals in [Fe/H] whereby the gradient is really quite obvious and strong. It is possible to account for a gradient such as this by adding in a term to the variance that is dependent upon the metallicity value, however this is work to be performed in the future.

In stark contrast, the residuals for the RC and SC stars (Fig 3.25) show marked trends, especially in the SC which does not appear to fit the relation well. There is also structure at low $\nu_{\rm max}$ where the amplitude seems to have a lower gradient than expected, leading to a small knee at ~ 30 μHz. The temperature gradient still appears to be present but this is likely due to its poor removal during the fitting process, and there are no trends observed in [Fe/H]. In the future it would be wise to consider the RC and SC clump as separate populations when looking at amplitudes. This is due to the nature of the SC and the fact that $A_{\rm max} \propto L/M$, where L is the luminosity of the star and M is the mass. The SC lies at a minimum in the luminosity of the clump due to it also lying at a minimum in the mass of the He core. This coupled with the increased mass results in a lower amplitude than for the rest of the clump, and so will contribute to the SC possessing a different relation.

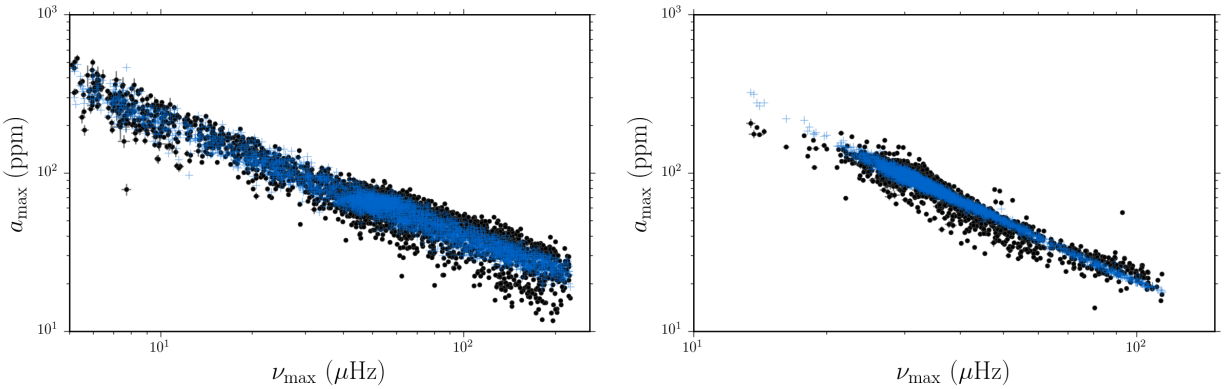


Figure 3.23: The predicted $A_{\rm max}$ are shown in blue plotted as a function of $\nu_{\rm max}$ where the original data is shown in black. The RGB sample is on the left-hand side whereas the clump sample is shown on the right-hand plot.

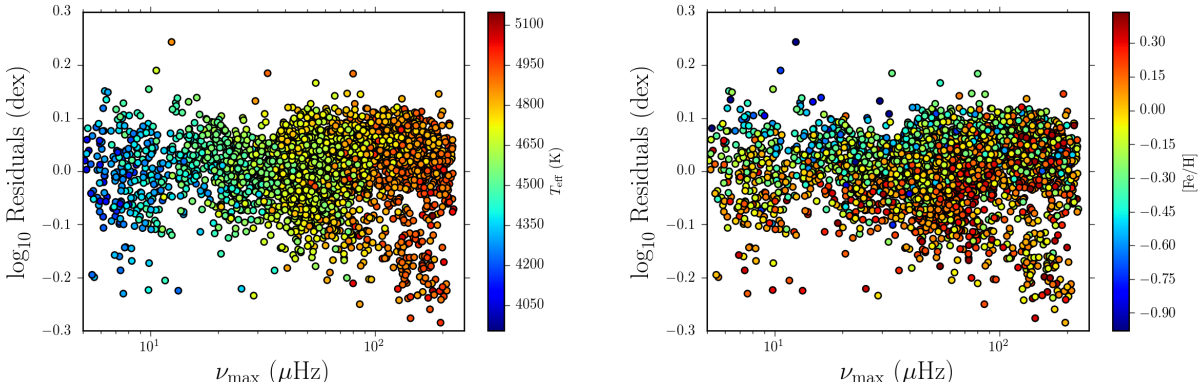


Figure 3.24: The residuals of the fit to the $A_{\rm max}$ data shown for the RGB sample as a function of $T_{\rm eff}$ on the left-hand plot and [Fe/H] on the right-hand plot.

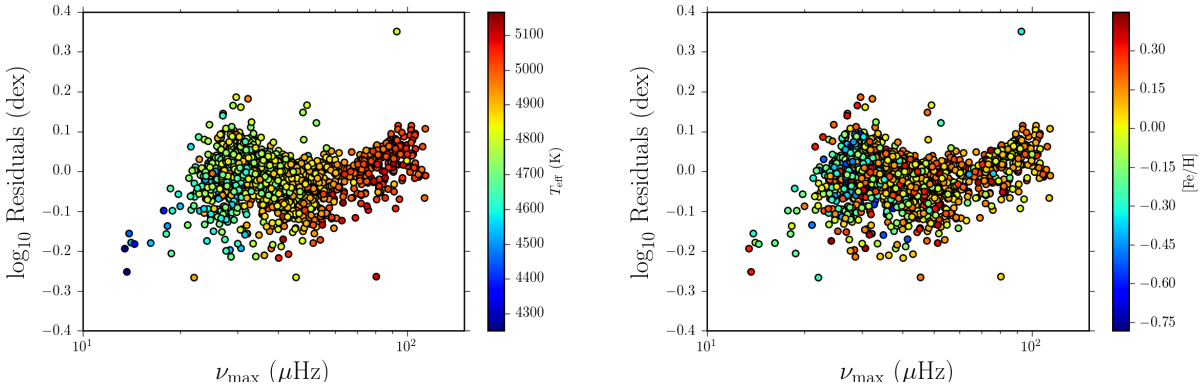


Figure 3.25: Same as Fig 3.24 but for the clump sample.

3.8.2 White noise vs. *Kepler* magnitude

A very useful sanity check when fitting such a large number of stars is to compare the results to those of previous works to check for consistency between the different methodologies. One such check involves looking at the white noise level as a function of the *Kepler* magnitude (K_p) and the relationship should be similar to Fig 1. in Jenkins et al. (2010b). The white noise extracted from the background fitting was converted back into the rms scatter in the timeseries by inverting equation 18 in Chaplin et al. (2011b). It is known that the precision over a 6.5 hour timescale taken from the timeseries for the red giants would be dominated by granulation and oscillations as opposed to white noise (North et al., 2017). As a result it would not be surprising if our values lay lower overall than those extracted from Jenkins et al. (2010b).

The plot produced from the background fitting, shown in Fig 3.26, reproduces the observed relationship and this shows again that the fitting process is performing as expected. There are a few stars that lie below the lines, however this could be due to incorrect *Kepler* magnitudes or a result of the relation being empirical and fitted by eye (or the reasons stated in the last paragraph). This reproduction of the white noise level provides some merit to the method and shows that the outputted parameters can be thought of as valid. However for stars with a high $\nu_{\rm max}$ (close to the Nyquist) it becomes progressively harder to estimate the white noise from the power spectrum due to the timescales of the granulation increases and the oscillations increasing in frequency.

3.9 Kepler-as-K2

Since the failure of two reaction wheels aboard *Kepler* and the subsequent re-purposing to become the K2 mission (Howell et al., 2014) the observational strategy shifted greatly. During the nominal *Kepler* mission observing runs were made in quarters (always observing the same patch of sky), however during K2 the observing has been split up into multiple 80

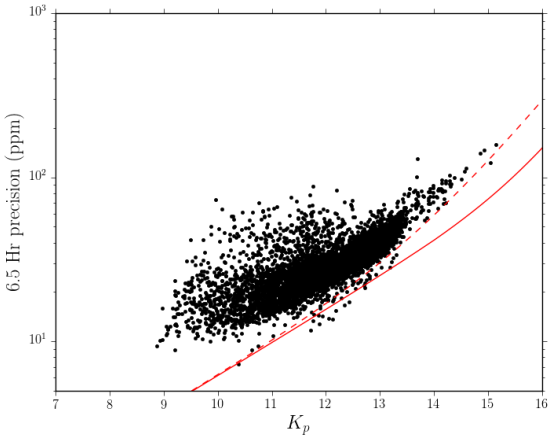


Figure 3.26: The white noise extracted from the background fitting plotted as a function of *Kepler* magnitude. The red dotted line shows the upper limit derived in Jenkins et al. (2010b) and the red solid line shows the lower limit on the precision also derived in Jenkins et al. (2010b).

day campaigns which observe changing fields in the ecliptic plane. As a result the datasets from K2 will only be 80 days as opposed to the 4 years from the nominal mission. It is therefore also important to assess the quality of the methodology on K2-like datasets, this can help assess biases and give an idea as to the quality of the fitting process.

The original *Kepler* datasets were cut down to simulate K2 data, i.e. into 80 day chunks (the typical length of K2 data). Only the first 80 days of the *Kepler* mission were taken and the fitting procedure applied to those data. The hope is that all the values are consistent with one another and that the only difference being an increase in the uncertainties as a result of the shorter length of dataset. The fitting procedure was identical to that of the original *Kepler* data with no additional priors introduced.

The main parameter that will be looked at in this section is $\nu_{\rm max}$ and how its determination is affected by the degraded data quality. Since all other parameters have been seen to scale as a function of $\nu_{\rm max}$ this is the most important parameter to check the consistency of.

The best place to start is looking at the uncertainties on $\nu_{\rm max}$ between the two different datasets. It would be expected that the uncertainty should decrease with the length of the dataset according to $\sqrt{N}$ where N is the number of points in the dataset. Fig 3.27 shows

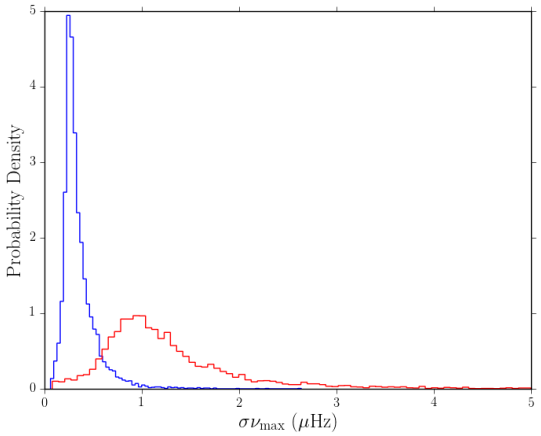


Figure 3.27: A comparison of the uncertainties in $\nu_{\rm max}$ from the *Kepler*data in blue and the *Kepler*-as-K2 data in red.

the histograms of the uncertainties from each set of fits. It should be apparent that the *Kepler* fits have much lower uncertainties than the *Kepler*-as-K2 and this is to be expected. Simply knowing the uncertainties helps, but it is important to make sure that they scale in accordance with the expected length of timeseries. The best way to view this is in terms of fractional uncertainty as a function of $\nu_{\rm max}$. This is shown in Figs 3.28 and 3.29 where it can be seen that the uncertainties agree well in that the $\sqrt{N}$ dependence is obeyed. The *Kepler*-as-K2 sample uncertainties are slightly underestimated given the length of the timeseries, however this will be due to additional noise sources in K2 data that is difficult to account for when modelling it using *Kepler* data.

After considering the uncertainties, the next step is to look at the differences between the $\nu_{\rm max}$ values determined in each dataset. Rather than simply looking at the differences, it would benefit us to look at the differences with respect to the measured uncertainties as this will give an idea as to whether there are any biases present or uncertainties underestimated. The relative differences are shown in Fig 3.30 and the median of the distribution appears to be very close to zero which is a good sign[19]. Due to the slightly asymmetric nature of the

[19]There is of course a caveat present in this analysis: this is only strictly applicable if the two datasets are independent of one another, but the *Kepler*-as-K2 data is part of the original dataset. Given that the

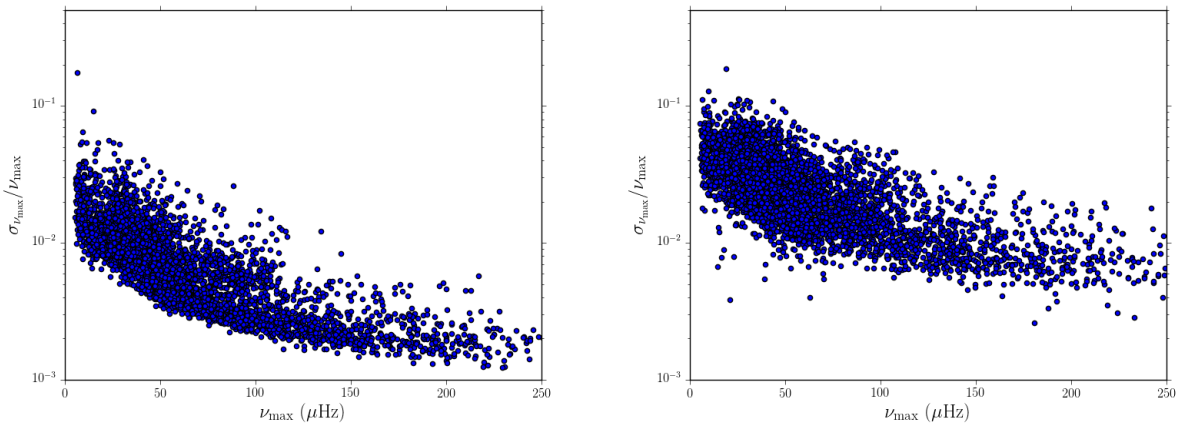


Figure 3.28: The fractional uncertainties on $\nu_{\rm max}$ shown as a function of $\nu_{\rm max}$ for the *Kepler* data (left hand panel) and the Kepler-as-K2 sample (right-hand panel).

distribution (which implies the *Kepler*-as-K2 fits underestimate $\nu_{\rm max}$ slightly) the standard deviation was dropped in favour of the 68.3% credible interval calculated as the highest posterior density (HPD), which described the distribution as $0.06^{+1.14}_{-1.18}$. A value of 0 ± 1 would imply that there are no biases present and that the difference in $\nu_{\rm max}$ values between the two datasets can be completely associated with the increased uncertainties. The fact that the real values are very close to 1 shows that the differences in $\nu_{\rm max}$ are almost completely within the respective uncertainties, but this could not always be the case and there are additional uncertainties not taken into account. This is the same reason that the *Kepler*-as-K2 $\nu_{\rm max}$ uncertainties do not quite scale correctly.

3.9.1 Updating the $\nu_{\rm max}$ determination

The $\nu_{\rm max}$ determination can be updated such that rather than simply using the mean of the prior distributions a slightly more involved method can be used. Since the N-dimensional Gaussian has been used to model the observed *Kepler* data, it can also be used to recreate the observed distributions which can subsequently be sampled from . This can enable a

Kepler-as-K2 is only 80 days out of the ∼1460 days of the *Kepler* data, the assumption that the smaller chunks are independent is adequate.

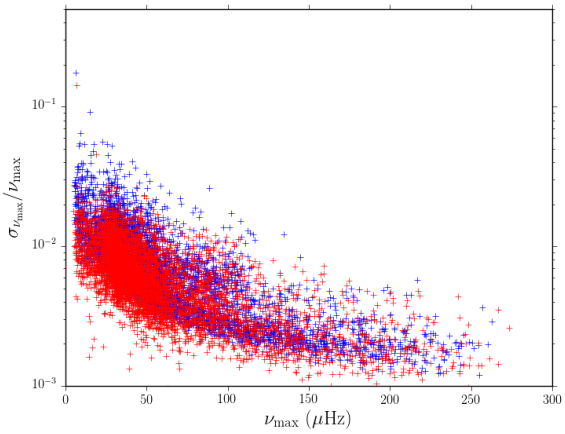


Figure 3.29: A comparison of the fractional uncertainties on $\nu_{\rm max}$ shown as a function of $\nu_{\rm max}$. The *Kepler* data is shown in blue, and the Kepler-as-K2 sample is shown in red rescaled by $\sqrt{N}$ to make it comparable with the original *Kepler* data.

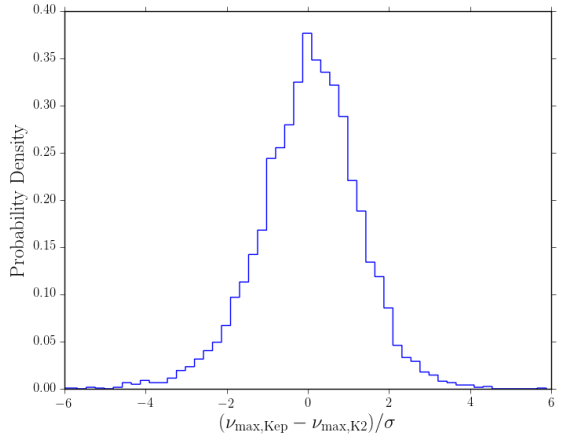


Figure 3.30: The difference between the determined $\nu_{\rm max}$ from the *Kepler*data and *Kepler*-as-K2 data divided by their error (added in quadrature).

fast determination of $\nu_{\rm max}$ that is based upon fits to real data (provided the stars show the expected behaviour), which should be more robust than the other method proposed due to its ability to cope with power spectra that deviate from what is expected.

To start with, 50,000 samples are drawn from the N-dimensional Gaussian, all of which are checked to make sure that the $\nu_{\rm max}$ values sampled lie within the range expected for long-cadence observations. The log-likelihood is then calculated for each parameter sample, making use of the data rebinned over 1μHz. Due to the distribution of $\nu_{\rm max}$ values in the data, this method is much more sensitive to $\nu_{\rm max}$ values close to the peak seen in Fig 3.31 at $\nu_{\rm max} \sim 30$ μHz. In order to increase the robustness at higher $\nu_{\rm max}$ the samples were initially increased from 10,000 to the current level of 50,000. In order to choose the best model, the set of parameters with the maximum likelihood value were taken. Two examples are shown in Fig 3.32 where this method is applied to a *Kepler* target observed for 4 years and a K2 target observed for 80 days. In both cases the method does a good job of estimating $\nu_{\rm max}$ which can then be fed into the fitting procedure.

In the future it could be possible to extend this method to oscillation detection whereby models are constructed with and without oscillations in addition to a flat white noise model. Since the log-likelihood is calculated it would be possible to construct the BIC for the maximum likelihood parameters of each model set and then construct the Bayes factor. From this, model comparison could be used to deduce whether oscillations are detected.

3.10 Can we predict evolutionary state from background parameters?

Given the large amount of data extracted from the *Kepler* data we can also start to address some of the more pressing questions before future space missions start, one such question is regarding classifying the evolutionary state of a star. In previous works this has relied upon the presence and analysis of the oscillations themselves (Stello et al., 2013; Vrard et al., 2016; Elsworth et al., 2016) which provides some metric akin to the period spacing from which

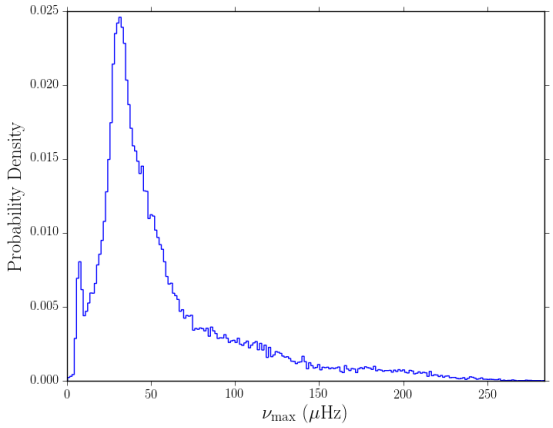


Figure 3.31: A histogram showing the distribution of $\nu_{\rm max}$ values that have been drawn from the modelled prior distributions.

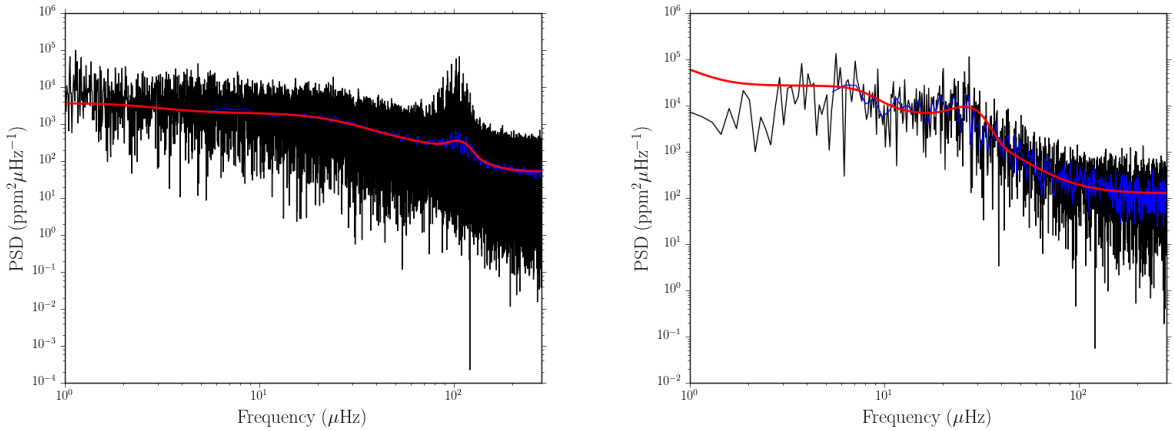


Figure 3.32: Estimate background profiles for KIC 8564976 on the left and EPIC 201127270 on the right. The data is shown in black with the spectrum rebinned over 1 μHz shown in blue and the estimated model in red.

the stars can be classified. This is extremely valuable as no other technique can distinguish between RGB and RC to the extent that asteroseismology enables.

It would be ideal to continue this classification to future missions such as TESS where we would have many more stars, but shorter time series that would almost certainly curtail the majority of current methods. As a result, the development of a technique that could classify the evolutionary state of a star based on more general features would prove beneficial. An example would be using the parameters extracted from the background fitting procedure which only rely on the oscillations being present and utilise the hump of power rather than the individual modes. This could then provide an easy means to provide coarse characterisation of a large number of stars without the need to analyse the modes themselves.

The application of machine learning techniques to such classification tasks is well documented and we shall proceed using supervised machine learning. By supervised we mean that the goal of the algorithm is to correctly predict the label of some known data (i.e. the evolutionary state) given a set of input features (e.g. the background parameters).

3.10.1 Target Labels

Apart from the input features, the other main component of any classification (or for that matter, any supervised machine learning task) is the target labels that the chosen algorithm is trying to predict. The labels in this classifier will, as the title of the section suggests, be the evolutionary state of the star. These are initial given as nominal variables with "RGB" denoted a red giant branch star, "RC" denoting red clump and "SC" denoting secondary clump stars. For use with the classification algorithm these were remapped onto integers from 0 to 2 (this mapping is required of some algorithms). The labels were not one-hot encoded[20] as the algorithm being used (`xgboost`) does not require one-hot encoder labels for multi-class classification problems.

[20]One-hot encoding would creating a column for each evolutionary state which contained a binary variable, 1 if the star has that evolutionary state and zero otherwise.

Whilst a key piece of information that asteroseismology can extract to derive the evolutionary state is the period spacing, this requires good signal-to-noise in the data and the presence of the mixed-modes in order for the analysis to work. For example, in the case of "clean" dipole red giants the period spacing cannot be determined due to it being low enough (when combined with a low $\Delta\nu$) that all the $\ell = 1$ mixed modes are contained within a wide central hump of power and are indistinguishable from each other, this is a result of radiative damping (Grosjean et al., 2014). This is also a problem for low $\nu_{\rm max}$ stars where $\Delta\nu$ is only a few multiples of the frequency resolution and so the period spacing cannot currently be accurately determined. As a result, the use of the period spacing in such an analysis is not viable on a large enough scale in order to check the evolutionary states used as our labels. However in the case of very low $\nu_{\rm max}$ stars they are either on the RGB or AGB (asymptotic giant branch) and are definitely not RC or RC.

As opposed to using the period spacing, a slightly different direction must be taken. In the APOKASC catalogue (Pinsonneault et al., 2014) there are 2 evolutionary state classifications that provide values for the majority of the stars in the sample which we will combine with the results from Elsworth et al. (2016), `CONS_EVSTATES` and `KALLINGER_EVSTATES`. A variant of a majority-rule type algorithm could be used whereby if 2 out of the 3 labels agree then that label is taken as being the "correct" value. However this could still introduce noisy labels into the training set and this will affect the accuracy of any algorithm. Instead a safer alternative is only selecting stars that have the same evolutionary state from all the classifications. This way the effect of noisy labels can be mitigated, and this should help remove some of the more obvious outliers without having to cherry-pick the data (which would only exacerbate any previous selection effects).

The result of this consensus labelling is shown in Fig 3.33 which shows a few differences between the two methods. The first is that there are a few secondary clump (SC) stars in the left-hand panel of the figure that clearly look out of place amongst the RGB, whereas if

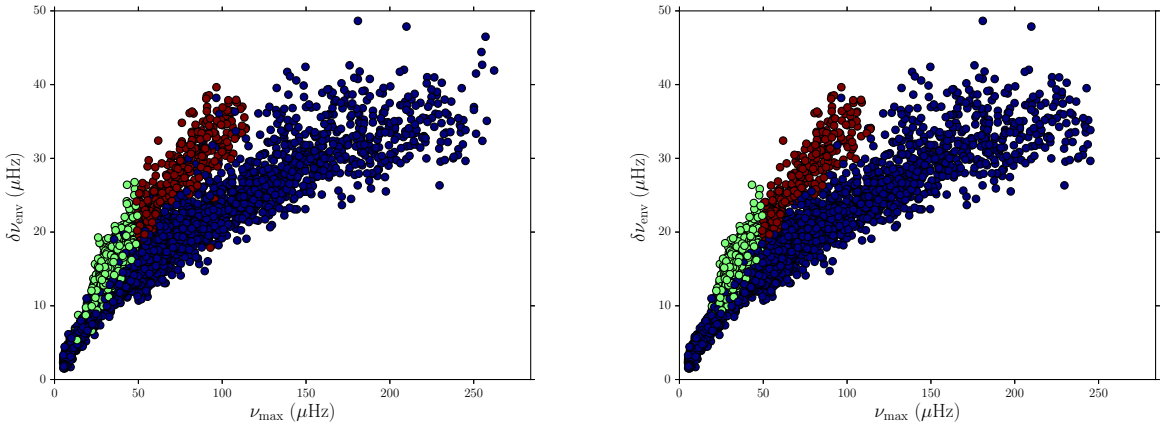


Figure 3.33: A comparison of the labels used in the classification task before and after consensus labelling was applied, in both plots blue points are RGB, green are RC and red are SC. The left-hand panel shows the labels from Elsworth et al. (2016), whereas the right-hand panel shows the results of consensus labelling and the final labels used.

all of the information available is used these stars no longer appear there and are not in our sample. This is a slight downside as it means the sample is reduced from $\sim$ 6000 stars to $\sim$ 5000 stars.

3.10.2 Using a single classifier

To begin, a single classifier algorithm will be used to assess the viability of the method and assess the importance of the features to be used. The input features used will consist of all the parameters extracted in the background fitting procedure described above, apart from the white noise level. The reason for leaving out the white noise is that is not intrinsic to the star, therefore if this method is to be generalised to different missions then any mission-specific parameters should be dropped.

A very popular and easy-to-use algorithm used in classification problems is `xgboost` (extreme gradient boosting[21]) developed by Chen & Guestrin (2016), which is a variant of the popular gradient boosting algorithms (which is in itself a modification of the very popular

[21] `xgboost` is a fast, flexible gradient boosting library that builds upon the pre-existing notion of decision trees

decision tree algorithms). The basic idea behind the algorithm is to combine an ensemble of weak learners into one strong learner that is then used to make predictions about the data (typically the weak learners are decision trees). For a more in-depth description of how this class of classifier works see, e.g. Hastie et al. (2001).

The classifier was trained making use of a 3:1 train-test split[22] initially in order to provide a test set to check for over-fitting and assess the performance of the model. In addition, 5-fold stratified[23] cross-validation was performed in order to determine the optimum number of iterations for the classifier, otherwise the default values of `xgboost` were used (this could be updated if the classification accuracy needs to be improved significantly). Stratified K-fold cross-validation was applied instead of typical K-fold cross-validation to ensure the ratio of classes in each validation set was the same. This is typical in multi-class classification problems, especially if the classes are not present in equal numbers.

In order to train the classifier the objective function is `multi:softprob`, which amounts to using the softmax objective function for multiclass classification. A slight difference is that rather than returning the class with the maximum probability (which is typically returned when using the softmax objective), the class probabilities are returned for each predicted label. Consider a N-dimensional vector $\sigma(\mathbf{z})$, the softmax function is defined as

$$\sigma(\mathbf{z})_n = \frac{e^{z_n}}{\sum_{n=1}^{N} e^{z_n}} \text{ for } j = 1, ..., N, \tag{3.22}$$

where the purpose of the function is to map all real values onto the range of (0, 1), which is helpful since they can then be interpreted as probabilities. This is useful as it gives some insight as to the quality of the prediction and where the approximate decision boundaries are relative to the expected positions. See Reverdy & Leonard (2015) and references therein

[22]By this we mean that 75% of the data was used to train the classifier and the other 25% was used to test the performance of the model.

[23]The data was divided into 5 chunks for hyperparameter tuning which were made sure to contain equal proportions of each class (the stratification).

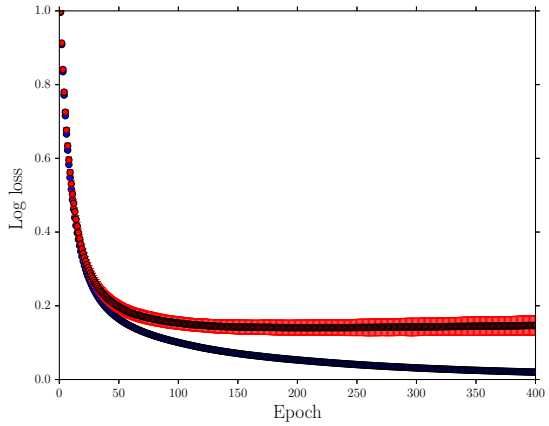


Figure 3.34: The logarithmic loss function plotted as a function of epoch for the training of the classifier. The points indicate the mean of the log-loss at that epoch and the errorbars show the standard deviation. In blue is the training set and the test set is given in red. The loss function flattening out is a good sign that the data is not being overfitted and that the model has been well trained.

for more information of softmax objective functions, for example.

The evaluation metric used during the cross-validation was taken to be the multiclass logarithmic loss, which is closely linked to the negative log-likelihood of the multinomial distribution given by

$$-\ln \mathcal{L} = -\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} \ln p_{ij}. \tag{3.23}$$

p_{ij} is defined as the probability that the ith data point belongs to class j and

$$y_{ij} = \begin{cases} 1 & \text{if } y_i = \text{class } j. \\ 0 & \text{if } y_i \neq \text{class } j. \end{cases} \tag{3.24}$$

The reason for using this evaluation metric is due to its ability to heavily penalise errors in the predictions through the use of the log predicted probability given above. This is important considering the nature of the task.

The training of the model is summarised in Fig 3.34 where the loss function is plotted as a function of epoch (or iteration number). The blue points show the training set and it can

be seen that as the number of iterations increase the log-loss tends to zero, whilst the test set (shown in red) bottoms out just below 0.2. This is important as it shows that the model is not overfitting the data on the test set and so the model will perform well on unseen data.

A useful diagnostic is the confusion matrix, which is generally used to assess the performance of an algorithm by displaying the true positive rates along with the false positive and false negative rates. In our case it helps reaffirm a lot of common knowledge which shows the classifier is performing well. It can be seen from Fig 3.35 that no RC stars are misclassified as SC or vice versa which is promising since they are fairly well separated in parameter space[24]. The other boundaries however are not quite as clear, since there is a small amount of confusion between SC and RGB which occurs due to a boundary shared between the stars where there is a small amount of overlap, but the largest confusion is between the RGB and RC classes. This is to be expected since they overlap in the majority of parameter spaces. The width of the envelope ($\delta\nu_{\rm env}$) is one of the few where there is more separation. This is shown in Fig 3.36 where the majority of stars that have been misclassified as either RGB or RC lie along the boundary between the two, which again is to be expected since there is no feature that shows a clear delineation between the two classes. The number of false positives and negatives is quite high at around 5%, but this is not too bad considering the accuracy of the classifier is $\sim 94.6\%$. The fuzzy boundary between the RGB and RC is accentuated as a result and since only global parameters are supposed to be used as input features this is a boundary that could only be improved with more data or better initial labels.

Additionally in a situation where the classes are not in equal measure the classification accuracy could be somewhat biased, in other words the quality of the classifier could be overestimated if there is a class imbalance.

In the construction of this classifier all the parameters (except the white noise) of the background fits were used. Whether this is necessary or not is another question. It could be

[24]The difference between RC and SC in the initial labels is made with a $\nu_{\rm max}$ cut, in reality it would be best to only use two labels and combine RC and SC into a helium core burning class.

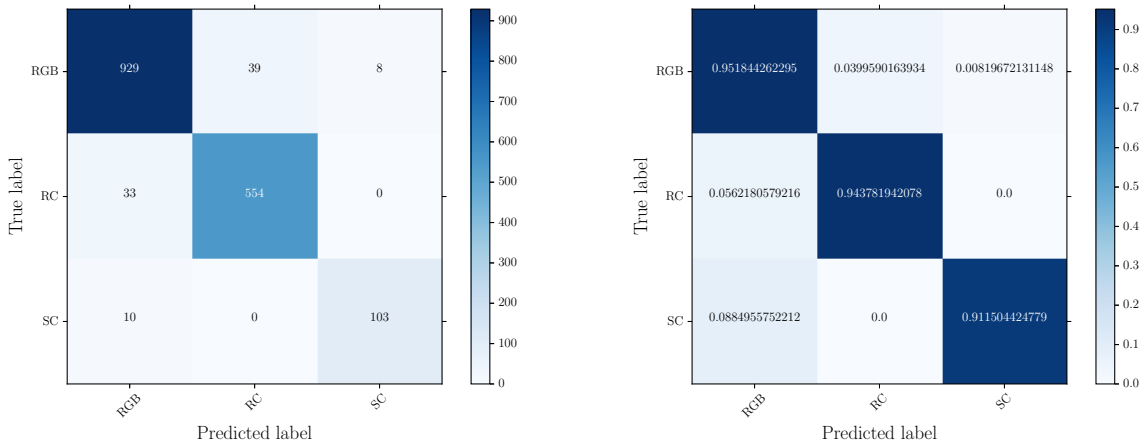


Figure 3.35: Confusion matrices for the classifier when applied to the test data. The left-hand panel shows the unnormalised confusion matrix where the values displayed and the colour bar refer to the number of stars in each category. The right-hand panels shows the number of stars in each category normalised by the number of stars in each class.

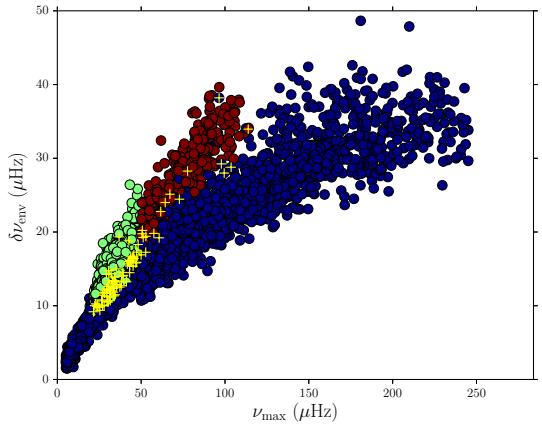


Figure 3.36: Width of the oscillation envelope plotted as function of $\nu_{\rm max}$, coloured by evolutionary state in the same way as previous plots, for the entire sample of stars. The yellow crosses denote stars that have been misclassified in the test set (either as false positives or false negatives).

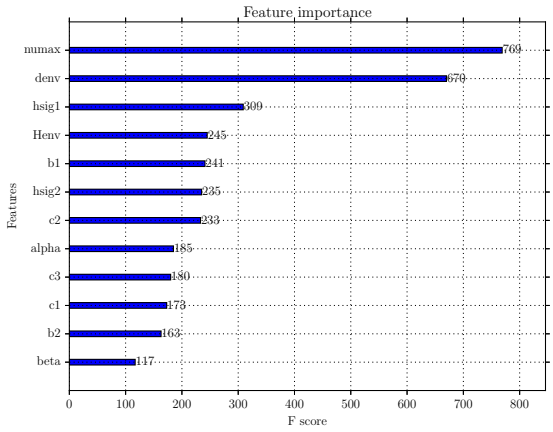


Figure 3.37: Feature importance for the features used in the classification task. The F-score is defined as the number of times a feature is used to split the data across all trees, giving a measure of how important each feature is to the fitting process.

possible that there are only a few features that provide information to the classifier whilst the rest are not very useful. To keep track of this the feature importance[25] can be used, as seen in Fig 3.37. As expected the most important features are $\nu_{\rm max}$, $\delta\nu_{\rm env}$, $H_{\rm env}$ and the first granulation component height σ_1 because they are feature spaces where the classes are most easily separated (see Fig 3.36). At the other end of the importance scale it is reassuring to see that the so-called nuisance "activity" component appears low down since it should not have a direct impact on the evolutionary state of the star. In addition, all of the exponents are in the bottom half of the feature importances which is useful given the contentiousness of their apparent values. Using feature importance gives the opportunity to cut down the feature set whilst maintaining the high accuracy of the classifier and is certainly an avenue that could be pursued in the future.

Given that the data of the stars in this sample have been cut down and fitted to simulate data from K2, the same classifier can be used to predict the evolutionary state of the stars given these noisier parameters. This is the final step of the current section, however this is when the uncertainties on the parameters become important and need to be accounted for.

[25]This is calculated by performing the fitting process many times and holding out a selected feature and assessing the performance of the classifier.

In most machine learning tasks known uncertainties on the inputs are not normally available to use within the algorithm istlef (see Bi & Zhang 2004 for application to support vector machines) and so this must be accounted for in a different manner. Due to the very low fractional uncertainties on the parameters extracted from the *Kepler* data ($\sim$ 1%), it was assumed during the training phase that there were no uncertainties on the input features. However when fitting to *Kepler*-as-K2 data, the uncertainties are considerably larger and must be taken into account when making predictions. The easiest way to do this is by bootstrapping, whereby N (which was taken to be 1000) rounds of predictions are made on the input features which are themselves perturbed from their median value according to their uncertainties. Each of the N rounds contributes to building up a distribution of the probability that a star belongs to a specific class, from which the median value can be taken as the summary statistic (for reasons given earlier) and the class label derived accordingly. For the stars that are a considerable distance from a decision boundary this process will make little difference to the predicted class labels, however those near decision boundaries the bootstrapping will result in a smoothed interpretation allowing for the fact that the uncertainties can allow a star to straddle two different class labels.

An example of the distribution resulting from the bootstrapping is shown in Fig 3.38 where it is clear that this particular star is RGB. This method of incorporating the uncertainties boosts the classification accuracy for the *Kepler*-as-K2 data from $\sim$76% to 81% which is a big increase in accuracy terms. However it is clear when looking at the misclassified stars in Fig 3.40 that the quality of the fits were much lower than for the original data, so certainly additional data cleaning would be needed to weed out additional poor fits, not just those that were poor from the original data. There is also a much larger degree of scatter which hinders the classification, however still achieving 81% accuracy with a dataset that is over 17 times shorter is very helpful for the future missions to show that this sort of classification can be performed.

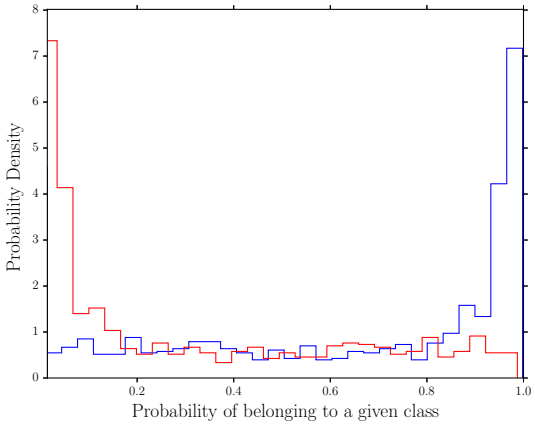


Figure 3.38: Probability distributions corresponding to the probability that the star belongs to a certain class. The distribution for RGB is in blue, RC in red and SC in green. The SC distribution cannot be seen due to the fact that it is so heavily concentrated at zero.

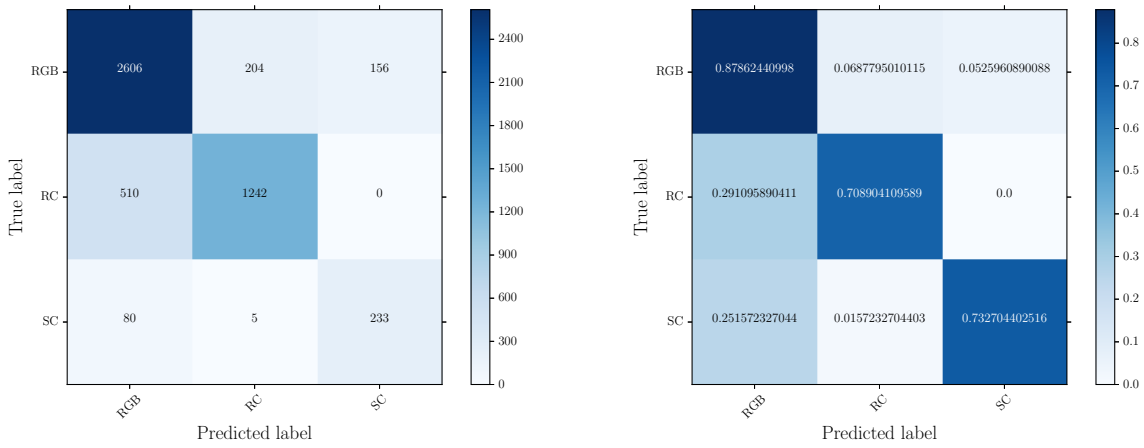


Figure 3.39: Confusion matrices for the classifier when applied to the test data. The left-hand panel shows the unnormalised confusion matrix where the values displayed and the colour bar refer to the number of stars in each category. The right-hand panels shows the number of stars in each category normalised by the number of stars in each class.

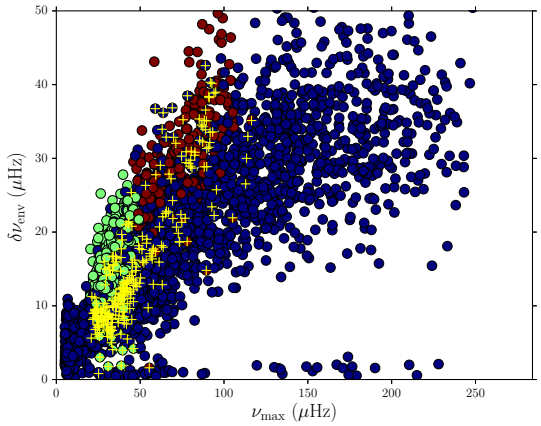


Figure 3.40: Width of the oscillation envelope plotted as function of $\nu_{\rm max}$ for the *Kepler*-as-K2 sample, coloured by evolutionary state in the same way as previous plots, for the entire sample of stars. The yellow crosses denote stars that have been misclassified in the test set (either as false positives or false negatives).

3.11 Conclusions

In conclusion, a robust method has been provided to enable extra value to be extracted from large-scale ensemble analyses through the use of MCMC and non-parametric priors. These methods help provide estimates of background and oscillation parameters for a large number of stars such that further analysis can be made. Inference has then been made on these data including searching for physical effects on the scatter of oscillation parameters as well as those belonging to granulation, such as $T_{\rm eff}$ and [Fe/H], as well as looking for the effect of evolutionary state on well known relations.

The extra robustness added through careful initialisation of the walkers in the MCMC fitting helps achieve the optimal solution and those that might not previously have been accessed. The prime example being the red clump stars that show unusual granulation timescales, which under normal circumstances would not be found. These stars show no obvious difference in stellar parameters from those stars that do not show such unusual features leading to the possibility that all clump stars may show this feature but it is only visible in certain cases.

The addition of new scaling relations for $\delta\nu_{\rm env}$, $H_{\rm env}$ and $A_{\rm max}$ constructed from the full *Kepler* datasets help remove any artefacts in previous analyses seen through the use of shorter datasets. Relations have been provided as a function of evolutionary state, which can in themselves help break degeneracies caused by evolutionary state. This also provides the opportunity to highlight trends and gradients present within the data that have not been previously studied leading to better predictions of these quantities in the future.

Finally it has been shown that background power in the power spectrum can be used as features to classify the evolutionary state of stars using random forests. For *Kepler* data this can achieve an accuracy of 95%, whilst on shortened *Kepler* datasets (simulating K2) the an accuracy of $\sim$ 80% can be reached. This could help provide classifications in datasets where the oscillations themselves cannot be resolved well enough for other methods to be applicable.

4

Generation of Artificial *Kepler* Red Giant Spectra

An important aspect of any discipline is being able to generate artificial data, whether this is to help test theory or to aid the building and fitting of models. In our case the concern is creating artificial power spectra of red giants. The underlying motivation is that we want to predict what red giant power spectra should look like under specific conditions, such as a given inclination, or as observed with future space missions such as TESS (Ricker et al., 2015) or PLATO (Rauer et al., 2014).

A number of works make use of generating artificial power spectra of artificial targets, such as SolarFLAG (Chaplin et al., 2006) and AsteroFLAG (Chaplin et al., 2008). Both of these produce helio- and asteroseismic time series respectively rather than power spectra as will be done here. In addition, the FLAG methods have a large degree of computation involved

with the construction of each dataset due to the way in which, for example the granulation components are calculated (due to the computations being made in the time-domain). The reason behind this was a hare-and-hounds exercise as preparation for the launch of *Kepler* and the focus was primarily on main-sequence targets. Only a few asteroseismic targets had been observed at this time, for example, Procyon (Arentoft et al., 2008; Mosser et al., 2008; Bedding et al., 2010), α Centauri (Thévenin et al., 2002; Bedding et al., 2003; Bazot et al., 2007) and Arcturus (Merline, 1995; Retter et al., 2003; Tarrant et al., 2007). This meant that in terms of the theoretical grounding these data were the closest approximation to what would have been expected.

Using the same procedure of simulation in the time domain for red giants would be costly in terms of computing time and necessary computing power due to the increased number of modes observed (through $\ell = 1$ mixed modes) and complex coupling giving rise to the mixed modes. There has been previous work (Grosjean et al., 2014) whereby power spectra have been generated from the underlying physics and this can give a great deal of information when compared to the observed data. However, the reasons for creating the artificial data are not for testing the underlying physics (such as mode damping rates) but more so for testing analysis codes, in line with the aims from the FLAG projects and so the power spectra need to be produced quickly, accurately and in large number.

The purpose of the artificial power spectra generated in this chapter is to be analysed as part of the artificial sample used in Chapter 6. As a result the artificial simulations need to be able to reproduce high $\nu_{\rm max}$ red giants to fairly good precision, as a result the creation process will not be as rigorous as it perhaps would be if this work was to be applied to all stars. Therefore the level of precision is lower in some cases, for example in the creation of the $\ell = 2$ modes, due to the fact that they are not analysed in the later work.

Since the launch of the space missions CoRoT and *Kepler* extensive ensemble analyses have been carried out on red giants (Huber et al., 2010; Hekker et al., 2011; Mathur et al., 2011;

Mosser et al., 2012a; Kallinger et al., 2014). These analyses enable the creation of scaling relations for given parameters either as a function of $\nu_{\rm max}$ of $\Delta\nu$. This vastly simplifies the process of generating asteroseismic data and means that from the basic quantities $\Delta\nu$ and $\nu_{\rm max}$ most of the power spectrum can be generated. The use of such empirically derived relations to produce artificial power spectra will be described throughout this chapter along with applications to different instruments. All that is needed for the simulations is an estimate of $\nu_{\rm max}$, all other parameters can be approximately calculated from this initial estimate. But in the case that parameters, such as $\Delta\Pi_1$, are known these can also be used as inputs. These simulations are restricted to red giant branch stars and do not extend to red clump stars, therefore the fact that for the same $\nu_{\rm max}$ $\Delta\Pi_1$ can be degenerate between the red giant branch and red clump is not a problem here.

4.1 Background Spectrum

There are two main components of the power spectrum: the background due to granulation and shot noise, and the oscillation modes. Firstly the focus will be on the granulation background and how this is simulated in the power spectrum. An explanation of the reasoning behind the specific form of the background model is given in Chapter 3.

The background profile is created following model F of Kallinger et al. (2014) which consists of two Harvey-like profiles (Harvey, 1985) describing the granulation and a flat background due to white noise. There is also evidence for a very low frequency term needed when fitting *Kepler* data which results in a frequency dependent noise term added into the background model. A good example is shown by the final model used in Kallinger et al. (2014). There have been studies into the low-frequency noise properties of *Kepler* (Gilliland et al., 2015) and through updated methods the low-frequency signal is contaminated less by the data processing (Stumpe et al., 2014b). Due to the lack of an easy way to use scaling relations, however, the decision was to leave out this component for the time being. Consequently,

disentangling any instrumental noise with very-low frequency stellar signal (such as activity from starspots or faculae) would prove to be very difficult and generally the frequency-dependent term is designed to just "mop up" this unknown signal[1]. In addition, the frequency signal should only impact stars with a $\nu_{\rm max}$ close to, or above, the *Kepler* long-cadence Nyquist frequency if the relationships in the scaling relations are to be carried forward. Any "activity" component should only occupy a small, very low frequency region of the power spectrum therefore it will be ignored. Therefore, the background model is as follows

$$\mathcal{B}(\nu) = \eta^2 \left(\sum_{i=0}^{1} \frac{\xi a_i^2 / b_i}{1 + (\nu / b_i)^4} \right) + W, \tag{4.1}$$

where η^2 denotes the sinc-squared due to the sampling of the data, a_i and b_i are the amplitudes and characteristic frequencies of the ith granulation component and $\xi = 2\sqrt{2}/\pi$ is a normalisation constant for the Harvey-like profiles with an exponent of 4. Finally, the white noise level is given by W which is unaffected by the sinc-squared modulation.

In order to calculate the values of the granulation amplitudes and characteristic frequencies, the scaling relations given by Kallinger et al. (2014) are used. These are defined for a given $\nu_{\rm max}$ and are as follows

$$a_1 = a_2 = 3382 \nu_{\rm max}^{-0.609}, \tag{4.2}$$

$$b_1 = 0.317 \nu_{\rm max}^{0.970}, \tag{4.3}$$

$$b_2 = 0.948 \nu_{\rm max}^{0.992}. \tag{4.4}$$

As a result it would be expected that the characteristic frequency of granulation increases with increasing $\nu_{\rm max}$ and the corresponding amplitudes decrease[2].

[1]Since the granulation signal is lower for stars with a high $\nu_{\rm max}$ this low-frequency signal is much more apparent than for the lower $\nu_{\rm max}$ cases where the granulation is larger than the possible instrumental signal

[2]The scaling relations derived in the previous chapter were not adopted in this part of the thesis since they were derived after this work had been completed. There is however, no reason why they cannot be

The white noise component, W, was modelled according to the shot noise given by Jenkins et al. (2010b) using the following formula (Chaplin et al., 2011b)

$$W = 2 \times 10^{-6} \sigma^2 \Delta t, \tag{4.5}$$

where Δt is the cadence of the instrument (29.4 minutes for long-cadence *Kepler* observations) and the rms value of the noise, σ, is defined as follows

$$\sigma = 1 \times 10^6 \sqrt{c + 7 \times 10^7}/c, \tag{4.6}$$

where c is given by

$$c = 3.46 \times 10^{0.4(12-K_p)+8}, \tag{4.7}$$

and K_p is the magnitude of the star in the *Kepler* passband (sourced from the *Kepler* Input Catalogue or KIC, Huber et al. 2014). This only places a lower limit on the shot noise due to it being an empirical relation, rather than giving an accurate prediction for a given magnitude value and so a way to improve the simulations would be to use extracted values from the data instead, however that is beyond the scope of this more simplified model[3].

An example of the background model for a Kepler-56-like star is shown in Fig 4.1. There is also a contribution from above the Nyquist frequency in the form of aliases which are reflected about the Nyquist frequency back into the region of interest (Murphy et al., 2013; Chaplin et al., 2014b). This is, however, not accounted for in order to simplify the computation of mode frequencies but this could easily be added in as an extension to the code. Only for stars with a $\nu_{\rm max}$ very close to $\nu_{\rm nyq}$ do the aliases have non-negligible amplitudes, otherwise they can be ignored.

adopted into the code at a later date.

[3]The predictions from the previous chapter could also be used to improve the predictions of the white noise, but again this work was performed before the work of the previous chapter was completed.

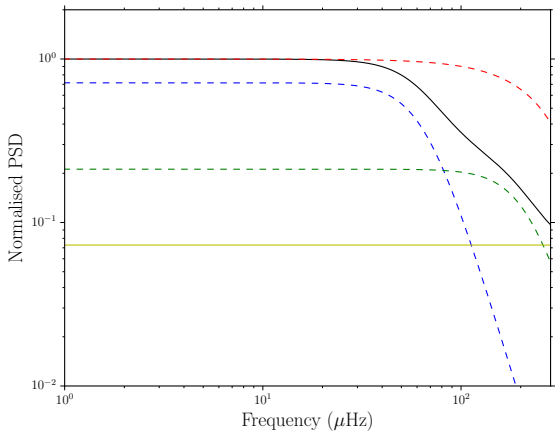


Figure 4.1: The background granulation model shown in black and decomposed into its constituent parts normalised to unity to show the effect of sampling. Shown in red is the η^2 function showing the effect of sampling, in blue is the first granulation component, the second component is shown in green and finally the shot noise is shown in yellow.

4.2 Mode Frequencies

Having covered the background we can move on to the oscillations. It is important to note that unlike the model in Chapter 3, the oscillations are not explicitly modelled as a Gaussian envelope, instead they are modelled as individual modes. The properties of the oscillations have been split into three sections with each covering a different aspect. In this section the mode frequencies and how they are calculated for a given set of input data will be discussed. This will build on information already given in Chapter 1, but the important points will be re-emphasised.

However first it is necessary to give an overview of the profile of the oscillation modes. Due to their stochastically excited and damped nature they are commonly modelled using a Lorentzian profile (see for example, Handberg & Campante 2011; Campante et al. 2016a). This results in a generalised model given as follows

$$\mathcal{M}(\nu) = \sum_{m=-\ell}^{\ell} \frac{\mathcal{E}_{\ell,m}(i) H_\ell}{1 + \left(\frac{2}{\Gamma}\right)^2 (\nu - \nu_0 - m\nu_s)^2}, \tag{4.8}$$

where $\mathcal{E}$ is the mode visibility, H is the mode height, Γ is the mode linewidth, ν_0 is the central frequency of the mode and ν_s is the rotational splitting. The summation of the azimuthal order m will dictate the form of the mode profile given the degree, so a radial mode will contain only one central component, whereas a dipole mode will be a triplet and so on. The mode height can also be refined in terms of quantities that are easier (for our purposes) to estimate

$$H = \frac{2A^2}{\pi\Gamma}, \tag{4.9}$$

where A is the mode amplitude (the way A, and other important quantities, are estimated will be explained later in this section). This is the basic mode profile that will be used throughout this chapter to generate the artificial power spectrum. Having covered the mode model, let us move onto describing the mode frequencies.

4.2.1 $\ell = 0$ & $\ell = 2$

To first order, the asymptotic relation for mode frequencies is given by

$$\nu_{n,\ell} = \left(n + \frac{\ell}{2} + \varepsilon\right)\Delta\nu - \delta\nu_{n,\ell}, \tag{4.10}$$

where n is the radial order, ℓ is the mode degree, ε is a phase term, $\Delta\nu$ is the large frequency separation and $\delta\nu_{n,\ell}$ is the small frequency separation. In the case of radial modes ($\ell = 0$) the asymptotic expression above reduces to

$$\nu_{n,\ell=0} = (n + \varepsilon)\,\Delta\nu, \tag{4.11}$$

which translates to frequencies that are evenly spaced according to $\Delta\nu$ with some shift given by $\varepsilon\Delta\nu$. This equation makes the assumption that both $\Delta\nu$ and ε do not change with radial order, i.e. that they are constant throughout the power spectrum. This is a reasonable

first order approximation, however in practice curvature is visible in the echelle diagram of both solar-type and red-giant stars and so this will be taken into account as well.

Before curvature can be taken into account the value of ε also needs to be predicted from a given $\Delta\nu$. This is done using the following equation taken from Mosser et al. (2011) who derived ε for a large number of CoRoT red giants and created a scaling relation as a function of $\Delta\nu$ in the form a power-law

$$\varepsilon = 0.634 + 0.546 \log_{10} \Delta\nu. \tag{4.12}$$

An extensive investigation into the asymptotic large separations of red giant stars observed by both CoRoT and *Kepler* was conducted by Mosser et al. (2013), who found and characterised curvature seen in the power spectrum. The curvature in the radial mode frequencies was observed to always have the same sign, i.e. positive curvature, which represents a positive gradient in $\Delta\nu$ with increasing frequency. An empirical addition to equation 4.11 that mimics the second-order correction to the asymptotic expression given by Tassoul (1980) was provided and equation 4.11 can be rewritten as

$$\nu_{n,\ell=0} = \left(n + \varepsilon + \frac{\alpha}{2}\left[n - n_{\max}\right]^2\right)\Delta\nu, \tag{4.13}$$

where the $n_{\max} = \nu_{\max}/\Delta\nu$ and the curvature term α is given according to

$$\alpha = 0.015\Delta\nu^{-0.32}, \tag{4.14}$$

and $n_{\max}$ is the radial order defined at $\nu_{\max}$ [4].

The effect of curvature on the frequencies and how it manifests itself in the frequency differences is shown in Fig 4.2. It can be seen how positive curvature results in a positive gradient in $\Delta\nu$, but also how small in size the curvature generally is which makes it difficult

[4]Unlike the radial order n, $n_{\max}$ does not have to be an integer (Mosser et al., 2011; Vrard et al., 2015).

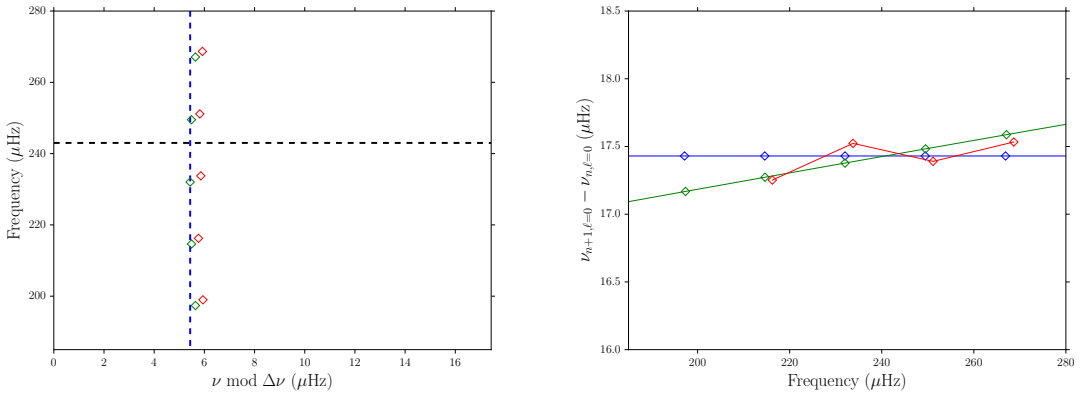


Figure 4.2: The left-hand panel shows an echelle plot consisting of the radial mode frequencies generated from the asymptotic expression given in equation 4.13 (green diamonds) and the extracted frequencies from the power spectrum of Kepler-56 (red diamonds). The black dotted line indicates the position of $\nu_{\rm max}$ and the blue dotted line shows the frequencies predicted by the asymptotic expression neglecting curvature. The right-hand panel shows the difference in frequency between successive radial modes using the same colours as the left-hand plot. The gradient seen in some of the frequencies is a clear indication of the presence of curvature.

to reliably extract with the small number of radial orders on display in red giants[5].

Having calculated the radial mode frequencies the $\ell = 2$ frequencies can also be calculated in a simple manner as follows,

$$\nu_{n-1,\ell=2} = \nu_{n,\ell=0} - \delta\nu_{n,\ell=2}. \tag{4.15}$$

where $\delta\nu_{n,\ell=2}$ is calculated using the scaling relation given in (Corsaro et al., 2012)

$$\delta\nu_{n,\ell=2} = 0.121\Delta\nu + 0.035. \tag{4.16}$$

The $\ell = 2$ mode of order n appears at almost $\Delta\nu$ higher in frequency than the radial mode of the same order. What has been done here is to simply compute the $\ell = 2$ frequency of the previous order using a well defined scaling relation (for the sake of simplicity). The $\ell = 2$ frequencies can now be calculated just from knowledge of the $\ell = 0$ frequencies and

[5]The error bars on the frequencies in Fig 4.2 would be too small to seen on the plot and so were not shown.

the small-frequency separation. Higher degree modes have been observed, up to $\ell = 3$, (e.g. Mosser et al. 2012a) but due to the relatively small amount of power contributing to the power spectrum and the lack of scaling relations for their mode amplitudes and linewidths they have been ignored in these simulations[6]

It is worth saying that the scaling relations used in this chapter are taken to be deterministic due to the way in which they were constructed. The intrinsic scatter seen in the relations are not taken into account, such as that due to effective temperature or metallicity. If these were added in the future it should be possible to reproduce the intrinsic scatter in the population when creating the artificial power spectra. It can be seen in Chapter 3 that both these quantities can affect the scaling relations in non-trivial ways (e.g. White et al. 2011).

4.2.2 $\ell = 1$

Due to the mixed nature of the $\ell = 1$ modes their frequencies are a little more difficult to determine and cannot be done so using equation 4.10[7]. Instead the following equation must be used which combines knowledge of g- and p-modes to create an asymptotic expression for mixed modes (Mosser et al., 2012b)

$$\nu = \nu_{n_p,\ell=1} + \frac{\Delta\nu}{\pi} \arctan\left[q \tan\pi\left(\frac{1}{\Delta\Pi_1 \nu} - \varepsilon_g\right)\right], \tag{4.17}$$

where $\nu_{n_p,\ell=1}$ is the nominal p-mode frequency, q is the coupling factor, $\Delta\Pi_1$ is the $\ell = 1$ period spacing and ε_g is a phase term. It can be seen that the frequency of the mixed modes ν occurs on both sides of equation 4.17 and so in order to solve it efficiently the equation can be rearranged as follows

[6]The purpose of these artificial spectra is for extracting the angle of inclination from the $\ell = 1$ mixed modes and so only approximate relations are needed to the high degree modes. In the case of $\ell = 3$ these have been ignored since they contribute so little power and would not affect this process.

[7]In fact all non-radial modes will exhibit a mixed behaviour, mixed $\ell = 2$ modes have been observed in *Kepler* data (Deheuvels et al., 2017). For simplicity and due to a lack of expressions describing the mixed higher degree modes they were assumed to be p-like

$$\tan\left[\frac{\pi\left(\nu-\nu_{n_p,\ell=1}\right)}{\Delta\nu}\right]=q\tan\left[\pi\left(\frac{1}{\Delta\Pi_1\nu}-\varepsilon_g\right)\right], \tag{4.18}$$

where the roots of the equation for a given set of mixed mode parameters give the mixed mode frequencies.

The nominal p-mode frequency is the frequency the p-mode would take if it was not mixed, which can be approximated as the separation $\delta\nu_{01}$ defined as the separation of the $\ell=0$ and $\ell=1$. It is assumed that this quantity does not change as a function of frequency and is set to $\Delta\nu/2$ (assuming the value from the asymptotic expression), this is not actually the case but again due to a lack of relations leading to an easy prediction it was assumed to be fixed. The same goes for the coupling factor, q, which was fixed to $q=0.2$ as deemed appropriate for red giant stars. This was shown in recent work by Mosser et al. (2016) who obtained $q\approx 0.17$ for red giant stars. Additionally the phase term $\varepsilon_{\rm g}$ was set to zero, due to the fact that this simply provides only a small shift in the observed frequencies and there (as of yet) exists no scaling relation for the quantity.

The final property of the mixed modes that needs to be addressed is the ($\ell=1$) period spacing, $\Delta\Pi_1$. To provide this estimate a linear fit was made to the data given in Vrard et al. (2016) so that a value can be estimated for a given $\Delta\nu$. This is a simple way of providing an estimate, a better way would be to model the intrinsic scatter in the data so that the relation is not deterministic but this is beyond our scope for the time being. Unfortunately due to the effects of radiative damping it is not possible to measure $\Delta\Pi_1$ for red giants high on the RGB (Grosjean et al., 2014). The fit to the data is shown in Figure 4.3 and the resultant "scaling-relation" is given by

$$\Delta\Pi_1=1.95\Delta\nu+56.2, \tag{4.19}$$

The solving of equation 4.18 proceeds in an iterative manner whereby the points above

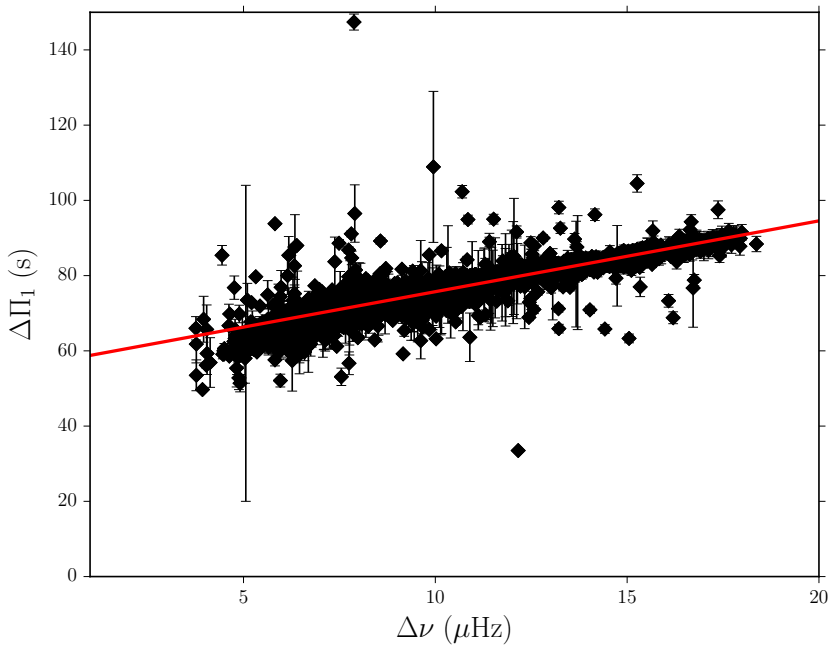


Figure 4.3: A linear fit (shown in red) made to the data analysed in Vrard et al. (2016) identified as red giants.

and below the roots (in frequency) of the equation are found. At this point the function is interpolated between those two frequencies to help improve the resolution of the determination. This results in a very quick and efficient method for obtaining the frequencies to a high precision.

4.3 Mode Amplitudes

Now that the mode frequencies can be calculated, the next parameter to consider is the mode amplitude. Once the radial mode amplitude is known the amplitudes of both the $\ell = 1$ and $\ell = 2$ modes can be easily calculated and this follows from estimates of relative mode visibilities which will be briefly explained in Appendix A.9 (e.g. Handberg & Campante 2011; Lund et al. 2016).

The calculation of the visibilities is the first step towards deriving the $\ell = 0$ amplitudes, the next is through the quantity, $A_{\rm max}$. This is the amplitude of the radial mode at $\nu_{\rm max}$ and

Table 4.1: A comparison of the relative visibilities for *Kepler*, TESS and SONG.

Visibilities	*Kepler*	TESS	SONG
$\tilde{V}^2_{\rm tot}$	3.16	2.94	4.09
$\tilde{V}^2_{\ell=0}$	1.0	1.0	1.0
$\tilde{V}^2_{\ell=1}$	1.54	1.46	1.82
$\tilde{V}^2_{\ell=2}$	0.58	0.46	1.04

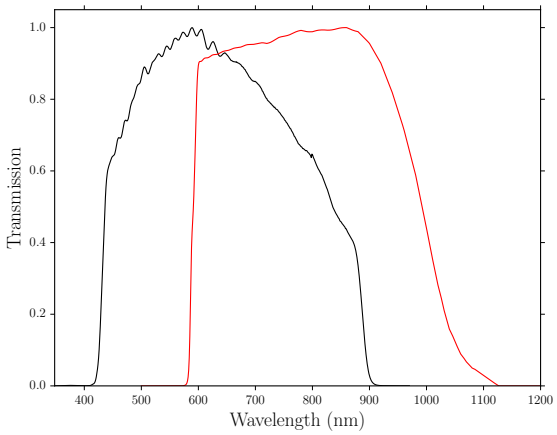


Figure 4.4: A comparison of the bandpasses for the *Kepler* mission in black and the TESS mission in red as a function of wavelength. Both transmission curves have been normalised such that their maximum value is unity.

can be defined as follows (cf. Eqn 3.2),

$$A_{\rm max} = \sqrt{\frac{H_{\rm env}\Delta\nu}{\tilde{V}^2_{\rm tot}}}, \tag{4.20}$$

where $H_{\rm env}$ is the height of the Gaussian envelope describing the oscillations. Equation 4.20 can be derived by considering $H_{\rm env}$ as being the average amount of power contained in the oscillations at $\nu_{\rm max}$ and so the total amount of power contained in a $\Delta\nu$-wide region around $\nu_{\rm max}$ is given by $H_{\rm env}\Delta\nu$. This tells us the power contained in all the modes in that region, i.e. in $\ell = 0, 1, 2$, but we only want the power contained in the radial modes. To convert this into the power contained in the radial modes the power must be divided through by the total relative visibility (this is because the relative visibility for radial modes is equal to unity) which leads to the quantity inside the square root in equation 4.20. The final step is to realise that the power is equal to the square of the amplitude and so square rooting the power gives us the amplitude of a radial mode at $\nu_{\rm max}$.

The quantity $H_{\rm env}$ is estimated using scaling-relations and calculated according to Mosser et al. (2012a)

$$H_{\rm env} = 2.03 \times 10^7 \nu_{\rm max}^{-2.38}. \tag{4.21}$$

This works well in practice for *Kepler* data since $H_{\rm env}$ has been obtained for many stars (e.g. Mosser et al. 2012a).

In addition to deriving $A_{\rm max}$ through quantities extracted from modelling the power spectrum it can also be derived from the fundamental stellar parameters (Huber et al., 2011)

$$A_{\rm max} = \frac{A_{\rm max,\odot}}{c_k}\frac{L^{0.838}}{M^{1.32}T}, \tag{4.22}$$

where L and M are the solar-scaled luminosity and mass respectively. The quantity $A_{\rm max,\odot} = 2.53$ ppm is the corresponding solar value, $T = T_{\rm eff}/T_{\rm eff,\odot}$ where $T_{\rm eff}$ is the effective

temperature of the star, $T_{\rm eff,\odot} = 5777$K is the solar effective temperature and the bolometric correction c_k as defined by Ballot et al. (2011)

$$c_k = \left(\frac{T_{\rm eff}}{T_0}\right)^{\alpha}, \tag{4.23}$$

where $\alpha = 0.8$ and $T_0 = 5934$K. In order to calculate the amplitude of the oscillations the same method as Campante et al. (2016b) was adopted whereby the *Kepler* bolometric correction was used and the 0.85 factor in the amplitudes accounts for the difference in the *Kepler* and TESS bandpasses. This is a useful method in the event that the stellar parameters are known to a good degree as opposed to relying simply in an estimate of $\nu_{\rm max}$.

In addition, if fundamental parameters of the stars are known the asteroseismic scaling relations (Chaplin et al., 2011a), given in equations 4.24 and 4.25, can be rearranged to derive $\Delta\nu$ and $\nu_{\rm max}$ from these fundamental properties.

$$\left(\frac{R}{R_\odot}\right) \approx \left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}}\right)\left(\frac{\Delta\nu}{\Delta\nu_\odot}\right)^{-2}\left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^{0.5}, \tag{4.24}$$

$$\left(\frac{M}{M_\odot}\right) \approx \left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}}\right)^{3}\left(\frac{\Delta\nu}{\Delta\nu_\odot}\right)^{-4}\left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^{1.5}. \tag{4.25}$$

When generating the artificial spectra only asteroseismic parameters are used as input and not fundamental stellar parameters, therefore when using equation 4.22 the mass and luminosity needs to be calculated given the chosen asteroseismic parameters. The relation $L \propto R^2 T_{\rm eff}^4$ can then be used in conjunction with equation 4.24 to compute the luminosity and equation 4.25 to compute the mass. It is however known that the above relation can overestimate the mass by around $10-15\%$ (Chaplin et al., 2011a) and so this would result in a slight underestimate of $A_{\rm max}$. This is due to the assumption made in the scaling relations that stars are homologous, that the stars scale in exactly the same way, which is not the case. A simple example can be demonstrated in terms of the radial order at $\nu_{\rm max}$, $n_{\rm max} = \nu_{\rm max}/\Delta\nu$.

It is known that $\nu_{\rm max} \propto \Delta\nu^{0.75}$ which results in $n_{\rm max} \propto \nu_{\rm max}^{0.25}$ and the scaling should be very weak. Scaling by the solar values and plugging in the values for the star KIC 8565976 with $\nu_{\rm max} \approx 104$ μHz, the estimated $n_{\rm max,est} \approx 10$ compared to the value for the actual star $n_{\rm max,actual} \approx 11$. Although this is only a small change it certainly shows that the scaling from main-sequence stars to red giants is not perfect and it is this imperfection in the scaling relations that causes deviations in estimated masses and radii.

Normally the oscillation envelope is described as Gaussian-like in power, but notice that the amplitude has been calculated rather than the power. This is done for ease of derivation, since the scaling relations are given in terms of mode amplitude as opposed to mode power and so the envelope can be calculated as the square-root of a Gaussian-like profile in amplitude and then this can be fed into equation 4.8 to compute the mode profile. To obtain the radial mode amplitude at the correct frequency the above square root profile must first be constructed as a function of frequency

$$a_{\ell=0}(\nu) = a_{\rm max}\left[\exp\left(-\frac{(\nu-\nu_{\rm max})^2}{2\sigma^2}\right)\right]^{0.5}. \tag{4.26}$$

Evaluating this function at the frequency of the radial modes leads to the appropriate amplitude. The resultant profile is shown in Fig 4.5. The width of the Gaussian σ is linked to the full-width at half maximum denoted by $\delta\nu_{\rm env}$ through $\delta\nu_{\rm env} = 2\sqrt{\ln 2}\sigma$. $\delta\nu_{\rm env}$ is an asteroseismic quantity that has been studied in large ensemble analyses and so there exists a scaling relation, again from Mosser et al. (2012a)

$$\delta\nu_{\rm env} = 0.66\nu_{\rm max}^{0.88}. \tag{4.27}$$

The amplitudes of the $\ell = 1$ and $\ell = 2$ modes follow-on very simply and can be calculated by evaluating equation 4.26 at the respective frequency and multiplying by the corresponding relative visibility, i.e.

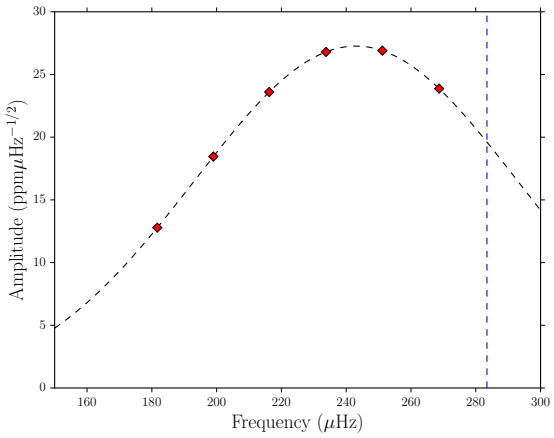


Figure 4.5: Radial mode amplitudes calculated according to equation 4.26 as a function of frequency for a Kepler-56 like star. The black dashed line shows the interpolation function used whilst the red diamonds mark the location of the radial mode frequencies. The blue dashed vertical line shows the location of the *Kepler* long-cadence Nyquist frequency beyond which we do not calculate the oscillation properties.

$$a_{\ell=1,2} = \tilde{V}_{\ell=1,2} a_{\ell=0}(\nu_{\ell=1,2}). \tag{4.28}$$

When creating the $\ell = 1$ mixed modes the height was used to calculate the mode profiles. The equation used is a variant of that used in (Fletcher et al., 2006) and can be seen in Basu & Chaplin (2016)

$$H = \frac{2\tilde{V}^2_{\ell=1} a_{\ell=0}(\nu_{\ell=1})^2}{\pi T \Gamma_{\ell=0}(\nu_{\ell=1}) + 2Q}, \tag{4.29}$$

where $\Gamma_{\ell=0}(\nu_{\ell=1})$ is the radial mode linewidth evaluated at the nominal p-mode frequency and $a_{\ell=0}(\nu_{\ell=1})$ is the amplitude of the radial mode evaluated at the nominal p-mode frequency[8]. The parameter Q is defined in Appendix A.9.1 as the ratio of the inertia of the non-radial mode (in our case the $\ell = 1$, I_1) relative to the radial mode (I_0) at the same frequency (ν_1).

In addition, the amplitude $a_{\ell=0}(\nu_{\ell=1})$ is the amplitude of the radial mode evaluated at the

[8]The units of equation 4.29 are currently ppm^2 so to convert them into the correct units of ppm$^2\mu$Hz^{-1} equation 4.29 must be divided by the bin-width.

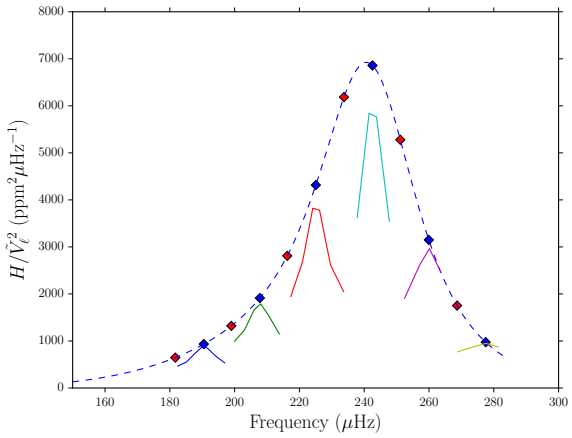


Figure 4.6: Mode heights normalised by their relative visibility evaluated at each predicted mixed mode frequency, plotted as a function of frequency for a Kepler-56 like star. The red and blue diamonds show the heights of the radial modes and nominal p-modes respectively and the solid lines show the heights of the $\ell = 1$ mixed modes calculated using equation 4.29 with a linewidth of 0.3 muHz and an observing length of 4 years.

mixed mode frequency and the linewidth is obtained in the same manner, i.e. $\Gamma_{\ell=0}(\nu_{\ell=1})$ is the radial mode linewidth evaluated at the mixed mode frequency.

The key features of this formulation are the limiting cases (see Fletcher et al. (2006) for more details)

$$H(\nu) = \begin{cases} \frac{2\tilde{V}^2_{\ell=1} a_{\ell=0}(\nu_{\ell=1})^2}{\pi T \Gamma_{\ell=0}(\nu_{\ell=1})} \text{ for } T >> 2\tau_{\ell=0}(\nu_{\ell=1}), \\ \frac{\tilde{V}^2_{\ell=1} a_{\ell=0}(\nu_{\ell=1})^2}{Q} \text{ for } T << 2\tau_{\ell=0}(\nu_{\ell=1}), \end{cases} \tag{4.30}$$

whereby the mode is unresolved if $T << 2\tau_{\ell=0}(\nu_{\ell=1})$ and resolved in the other case.

Therefore for a p-dominated mode as observed by *Kepler* it will, most likely, be resolved and so the first case in equation 4.30 is valid. This nicely reduces down to the common expression used for mode profiles. Whereas in the case of a g-dominated mode it is most likely unresolved and so its linewidth is most likely less than a single bin. As a result the peak height is suppressed by a factor that is inversely proportional to the factor Q.

The heights normalised by their respective relative visibilities as a function of frequency are plotted in Fig 4.6 for a Kepler-56 like star providing a comparison between the radial mode

heights, nominal p-mode heights and mixed mode heights. It should be the case that all of the mixed mode heights (when normalised as explained above) should lie below the dotted blue line showing the interpolation function for the radial mode heights, however this is not always true. Any deviations from that are most likely to be caused by the steep increase in linewidth beyond $\nu_{\rm max}$ which we will explain in the following section as we move on to cover the calculation of the mode linewidths.

4.4 Mode Linewidths

After considering the amplitudes, the linewidths of the modes are the next parameter to calculate. It is known that the mode linewidth is dependent upon the effective temperature of the star, Chaplin et al. (2009) proposed $\Gamma \propto T_{\rm eff}^4$ for radial modes based on predictions from pulsation computations, whilst Baudin et al. (2011) showed using CoRoT targets that the exponent was closer to 16 for main-sequence stars and was almost flat for red giants. Considerable theoretical work has been performed (Houdek et al., 1999, 2001; Samadi, 2011; Belkacem et al., 2012) investigating both the excitation mechanisms and the damping rates of solar-like oscillations as the two are inexplicably linked.

As a result of this aforementioned complex dependence there is no simple scaling relation of mode linewidth with $\nu_{\rm max}$. In order to infer some sort of scaling the best place to start is the relation derived by Corsaro et al. (2012), where the mean radial mode linewidth was estimated from the collapsed echelle diagram by fitting a Lorentzian profile as given below

$$\Gamma = \Gamma_0 \exp\left[\frac{(T_{\rm eff} - T_{\rm eff,\odot})}{T_0}\right], \tag{4.31}$$

where Γ is the mean radial mode linewidth, Γ_0=1.39 μHz and T_0 = 601 K.

This gives a scaling as a function of effective temperature but neglects a lot of the other features of the mode linewidth that is known to occur in stars. The mode linewidth has been observed to change as a function of frequency (e.g. Chaplin et al. 1997; Corsaro et al.

2012; Appourchaux et al. 2014), due to the fact that the modes are slightly more sensitive to the superficial layers of the star as a function of their frequency. In other words as the frequency of the mode increases it is slightly more sensitive to the near-surface layers, leading to a shorter lifetime and therefore enhanced linewidth due to the increased turbulence as a results of convection.

The above model can be updated according to the data used in Chapter 6 to try and account for the more complex features observed and so the linewidths extracted from the radial modes were fitted according to the following model (Appourchaux et al., 2014)

$$\ln(\Gamma) = [\alpha \ln(\nu/\nu_{\rm max}) + \ln \Gamma_\alpha] - \left[\frac{\ln \Delta\Gamma_{\rm dip}}{1 + \left(\frac{2\ln(\nu/\nu_{\rm max})}{\ln(W_{\rm dip}/\nu_{\rm max})} \right)} \right], \tag{4.32}$$

where ν is the frequency of the mode, α is the exponent of the power law, Γ_α is the multiplicative factor in the power law, $\Delta\Gamma_{\rm dip}$ is the height (or depth) of the Lorentzian profile and $W_{\rm dip}$ is the width of the Lorentzian. It is also important to note that sign from the original expression has been changed such that the last term in the brackets is subtracted rather than added due to the stability of the fitting procedure (Lund, 2016). In addition, the argument in the denominator of equation 4.32 has been changed to $\nu/\nu_{\rm max}$ from $\nu/\nu_{\rm dip}$ where $\nu_{\rm dip}$ is the frequency of the dip in the mode linewidths. This has been done because rather than assuming that there is a $\nu_{\rm dip}$ for each star it has been redefined in terms of a normalised version $\nu_{\rm max}$. This is reasonable as it is assumed that the plateau in damping rates should occur at approximately $\nu_{\rm max}$, the idea being that since all the stars in the sample used in the fitting have a similar $\nu_{\rm max}$ their properties should be similar (such as $\Delta\Pi_1$) and so $\nu/\nu_{\rm max}$ results in all the fitted frequencies being folded onto the same scale. It therefore seems wise to recast the expression in terms of parameters that are useful for an ensemble of stars rather than an individual case.

Equation 4.32 may seem like a rather odd expression but it provides a great deal of flexi-

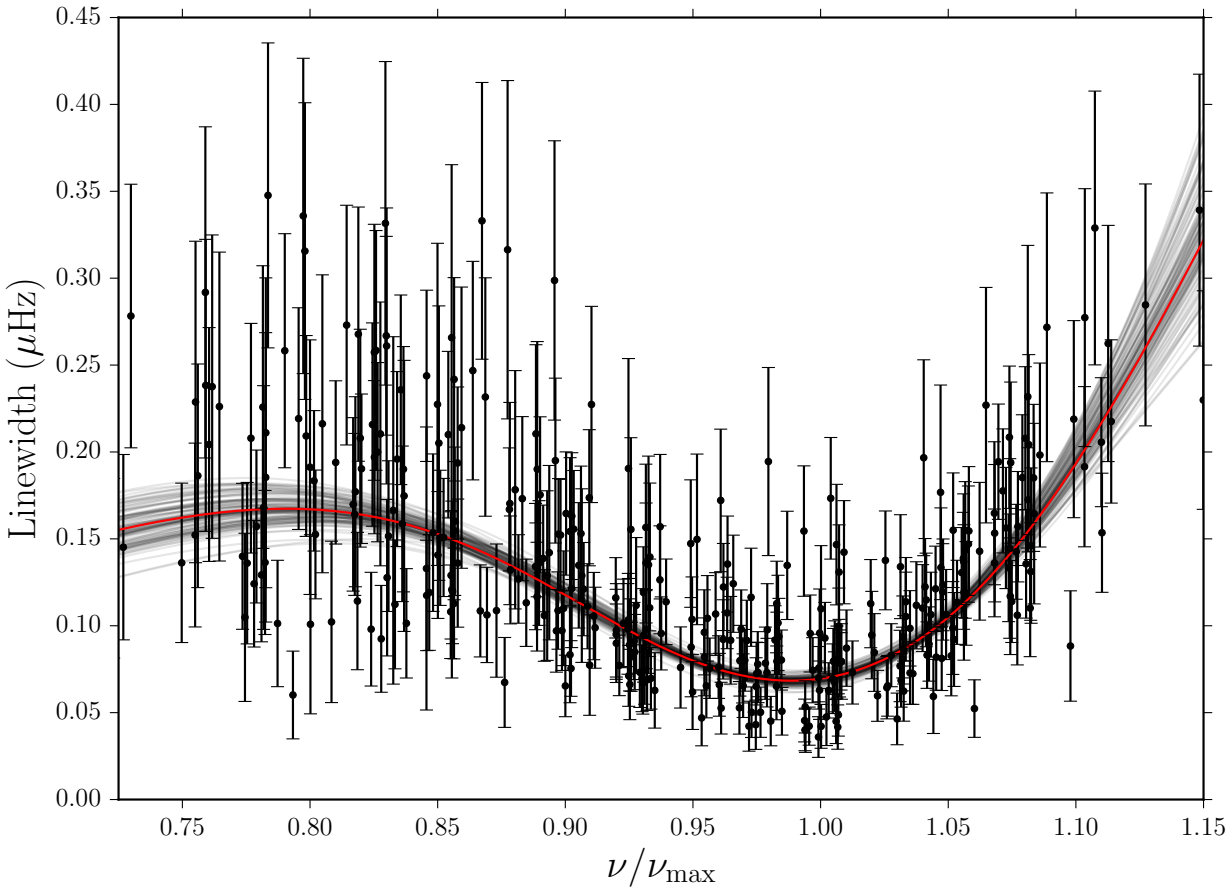


Figure 4.7: Mode linewidths for the radial modes fitted in Chapter 6 as a function of reduced frequency. The red solid line shows the fit of equation 4.32 to the data and the grey lines show random draws from the posterior distributions of the parameters.

bility. The basis of the expression is a power law for the mode linewidth as a function of the approximate radial order $\nu/\nu_{\rm max}$ with the addition of a depression modelled by a Lorentzian profile. The reasoning behind this dip is covered extensively by Belkacem et al. (2011b) and is related to a plateau seen in the damping rates.

The mode linewidths were fitted using MCMC assuming Gaussian uncertainties (i.e. a Gaussian likelihood function) and the result of the fit can be seen in Fig 4.7 as well as the parameters fitted to the data which are shown in Table 4.2. The value of $\Delta\Gamma_{\rm dip}$ may seem to be very large compared to the scale of the data, but if the relationship between this and $T_{\rm eff}$ given in Fig 6 of Appourchaux et al. (2014) was extrapolated to the approximate temperature

Table 4.2: Derived parameters from the fit to the radial mode linewidths.

Parameter	Value
α	$2.95^{+0.31}_{-0.32}$
Γ_α (μHz)	$0.52^{+0.11}_{-0.08}$
$\Delta\Gamma_{\rm dip}$ (μHz)	$7.52^{+1.45}_{-1.16}$
$W_{\rm dip}$	$0.79^{+0.02}_{-0.02}$

of our sample ($\sim$ 5000K) then a value of 7.5μHz is very reasonable. It was also noted by Appourchaux et al. (2014) that the location of the dip in linewidth and location of maximum mode height coincide, but not at $\nu_{\rm max}$ and in fact occur about half a radial order higher. This would affect the properties of the fit parameters slightly, but given the approximate nature and the desire to simply extract some sort of scaling relation this observation was neglected. The fitting procedure was not performed in log-space and the model was raised to an exponential to maintain the Gaussian uncertainties in linear space.

Before we define the mixed mode linewidth let us introduce the quantity ζ, which is defined as the ratio of the inertia in the g-mode cavity to the inertia in the whole of the cavity

$$\zeta = \frac{(I_{\rm g})_1}{I_1} \approx \frac{(I_{\rm g})_1}{(I_{\rm g})_1 + (I_{\rm p})_1}. \tag{4.33}$$

Following Deheuvels et al. (2015) this can be defined as

$$\zeta \approx \left[1 + \frac{1}{q} \frac{\cos^2\left[\pi\left(\frac{1}{\nu\Delta\Pi_1} - \varepsilon_g\right)\right]}{\cos^2\left[\pi\left(\frac{\nu - \nu_{n_p,\ell=1}}{\Delta\nu}\right)\right]} \frac{\nu^2\Delta\Pi_1}{\Delta\nu} \right]^{-1}, \tag{4.34}$$

where $\zeta \approx (Q-1)/Q$.

The mixed $\ell = 1$ mode linewidths are assumed to be modulated by a term proportional to the inertia (mixing function) in a manner that is proportional to the amount of coupling between the g- and p-modes. This comes about by assuming that the linewidth decreases as the mode becomes more g-dominated, therefore the linewidth is inversely proportional to the inertia. If we assume

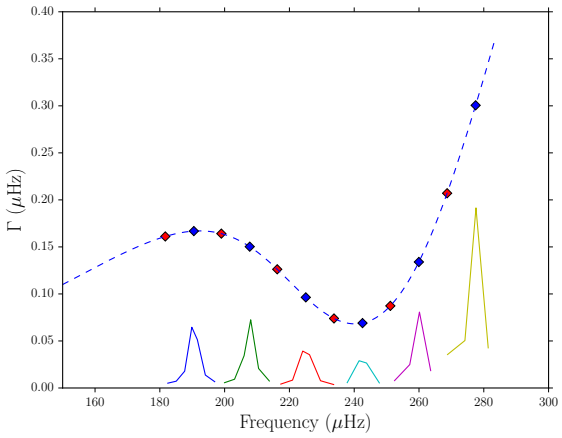


Figure 4.8: Mode linewidths of the radial, nominal p-mode and mixed modes as a function of frequency. The dashed blue line shows the fitted model for the widths whereas the red and blue diamonds show the linewidths of the radial modes and nominal p-modes respectively and the solid lines show the linewidths of the $\ell = 1$ mixed modes calculated using equation 4.36.

$$\frac{\Gamma_1}{\Gamma_0} \approx \frac{I_0}{I_1} \approx \frac{I_p}{I_p + I_g} = 1 - \zeta, \tag{4.35}$$

then the mode linewidths can be approximated through

$$\Gamma_{\ell=1} = \Gamma_{\ell=0}(\nu_{n_p,\ell=1})(1 - \zeta), \tag{4.36}$$

where $\Gamma_{\ell=0}(\nu_{n_p,\ell=1})$ is the width of the radial mode evaluated at the nominal p-mode frequency and ζ is the mixing function. This means that the most p-dominated mode has the largest linewidth (as expected) because its inertia is lowest and so the factor $(1 - \zeta)$ is closest to unity and consequently the most g-dominated modes have the smallest linewidth corresponding to the inertia tending to unity and $\zeta \to 0$.

The properties of the $\ell = 2$ modes are assumed to be the same as the radial modes with regards to the widths, but evaluated at the frequency of the $\ell = 2$ modes. Since there is little information available concerning the feature of $\ell = 2$ modes in red giants this assumption is adequate. It is also worth noting that the $\ell = 2$ modes are assumed to be pure p-modes and

not mixed in any manner. This is not strictly true since mixing in $\ell = 2$ modes has been observed (e.g. Deheuvels et al. 2012), however for the purposes of these artificial spectra we are not concerned with the structure of the $\ell = 2$ modes and so stick to the simpler representation.

4.5 Rotational Splitting

The final contribution to the modes to calculate is the effect of rotation, in other words the rotational splitting of the modes. The effect of rotation is to lift the degeneracy in frequency between modes of the same degree ℓ and different azimuthal order m. Therefore radial modes are unaffected since they only have a single m component ($m = 0$). The rotational splitting of the mixed $\ell = 1$ modes will be described first following the work of Goupil et al. (2013)

$$\nu_s = [\zeta(1 - 2\mathcal{R}) + 2\mathcal{R}]\, \nu_{s,\max}, \tag{4.37}$$

where $\mathcal{R}$ is the ratio of the average rotation rates of the envelope to the core, $\nu_{s,\max}$ is the maximum rotational splitting and ζ is the mixing function. It is important to note that the ζ used here is that defined in Deheuvels et al. (2015) and not Goupil et al. (2013). The above expression for the rotational splittings comes about by decomposing the rotational splittings into two contributions, one from the core and one from the envelope. The integrated kernels of each contribution then define the weighting of the core and envelope to the observed rotational splittings.

The reason we chose to use this updated version is due to some assumptions made by Goupil et al. (2013) in their original version (see Appendix B of Deheuvels et al. 2015 for a detailed explanation of the modifications made). These amount to modifying simplifications made to the p- and g-mode phases. As can be seen in Fig 4.9 by removing the simplifications the shape of the mixing function is changed slightly and in the updated expression the depth of ζ is increased which amounts to the mode closest to the nominal p-mode having slightly

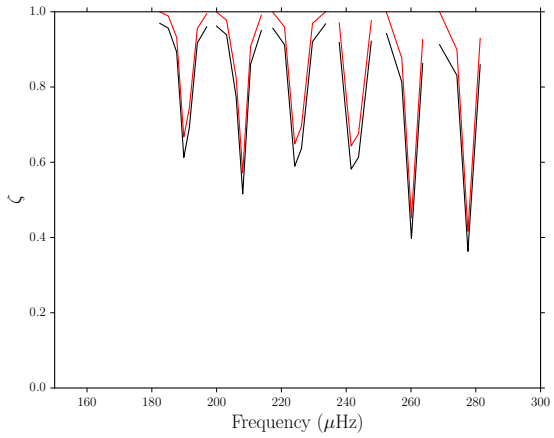


Figure 4.9: A comparison of the mixing function ζ using the formulation derived in Goupil et al. (2013) (red) and the updated version given by Deheuvels et al. (2015) (black).

more p-mode-like behaviour. The updated expression also brings the calculated expression closer to those inferred from observation and leads to improvements when fitting the mixed mode frequencies (e.g. Davies & Miglio 2016). The rotational splittings will take a form similar to that of the inertia derived in equation 4.34, where the p-dominated modes will have the lowest splitting value and the g-dominated modes will have the largest (under the assumption that the core is rotating faster than the envelope).

Due to there being no scaling relations for the envelope or core rotation rates (which go into the variable $\mathcal{R}$), it was decided to set $\mathcal{R} \sim 0.01$ since spectra computed with this value appear close to the observed power spectra. The maximum splitting defaults to 0.5 μHz if no value is given. This gives a certain degree of flexibility and can be tailored to individual cases if other parameters of the star are known, such as $v \sin i$ and radius which could be used to derive the rotation period and therefore approximate the envelope rotation rate.

In the case of the rotational splittings for $\ell = 2$ modes the splitting for the nominal p-mode were set to be the same as the $\ell = 1$ modes (assuming pure p-like modes). Given the fact that there have been no studies into the rotational splittings of $\ell = 2$ modes or whether they do have a mixed nature the assumption made about the splittings is reasonable.

The final quantity which is linked to the rotational splittings is the angle of inclination, the way in which this affects the mode properties will be discussed in Chapter 6. The value can either be set to a chosen value or drawn from an isotropic distribution.

4.6 Example Spectra and Comparisons to Real Data

4.6.1 *Kepler*, K2 and TESS

In the previous sections the process of creating the artificial spectra has been described and now it is time to show some applications of such data. This is very useful when predicting properties of the power spectrum for future missions as well as comparing to real data.

The starting point will be to consider a high $\nu_{\rm max}$ RGB star with properties very close to Kepler-56 (Huber et al., 2013a) over two different lengths of observations: the first will be for 1 year and the second will be 30 days. The reason for picking these values is that they correspond to both the minimum and maximum observing time of a star during the TESS mission[9]. Comparing the artificial spectra generated according to the TESS specifications with those from *Kepler* gives us a good idea of how much information can be extracted given different observing lengths for red giants observed with TESS.

The shot noise values for *Kepler* are calculated according to equation 4.5 whilst for TESS they are calculated according to the noise model described in Sullivan et al. (2015) assuming no systematic noise and that Kepler-56 has an I_c-band magnitude of 9.

A comparison between a Kepler-56-like star as observed by *Kepler* and TESS for a period of a year (the longest continuous observing time for TESS) is shown in Fig 4.10. The differences between the two are quite striking and are a consequence of the higher shot noise levels for TESS which, combined with the lower amplitude of the oscillations, result in a much lower signal-to-noise level. This is only a problem for the higher $\nu_{\rm max}$ stars, whereas for the lower

[9]Whilst the minimum observing time is technically 27 days, 30 days gives a good enough example of what the data should look like.

$\nu_{\rm max}$ stars where the amplitude of the oscillations are much larger, the high shot noise levels are less of an issue. This will however, have consequences regarding the ability to extract the finer asteroseismic details (so to speak), such as angle of inclination or period spacing due to the lower probable signal-to-noise of the data and loss of frequency resolution (especially in 30 days).

The high shot noise level is only a problem for the fainter stars, and so represents the more difficult case. Therefore these plots should be thought of almost as a worst-case scenario for red giants, whereby the majority of the data should be better and much easier to perform asteroseismic analysis upon.

In addition to the year-long observations, the same power spectra were generated for a 30 day observing period which corresponds to approximately the shortest observing time for TESS and they can be seen in Fig 4.11. In both cases the oscillations are hard to observe and so this would present a significant challenge for the extraction of global asteroseismic parameters either through fitting the background to the power spectrum or through other detection pipelines. Certainly for lower $\nu_{\rm max}$ stars this would be less of a problem since the amplitude of the oscillations would be larger, however the high shot noise could still prove difficult for the fitting procedures.

4.7 Extension to Doppler Velocity

Stellar oscillations can not only be observed in intensity with missions such as *Kepler* and *K2*, but also in Doppler velocity using high-resolution spectrographs such as SONG and HARPS. This is analogous to the Doppler velocity measurements of the Sun taken with BiSON (Hale et al., 2016). It is therefore appealing to simulate power spectra in velocity as well as intensity so that an approximation of the observed power spectrum can be created, to give an idea of what the power spectrum should look like before the observations are made (given the highly competitive nature of proposals).

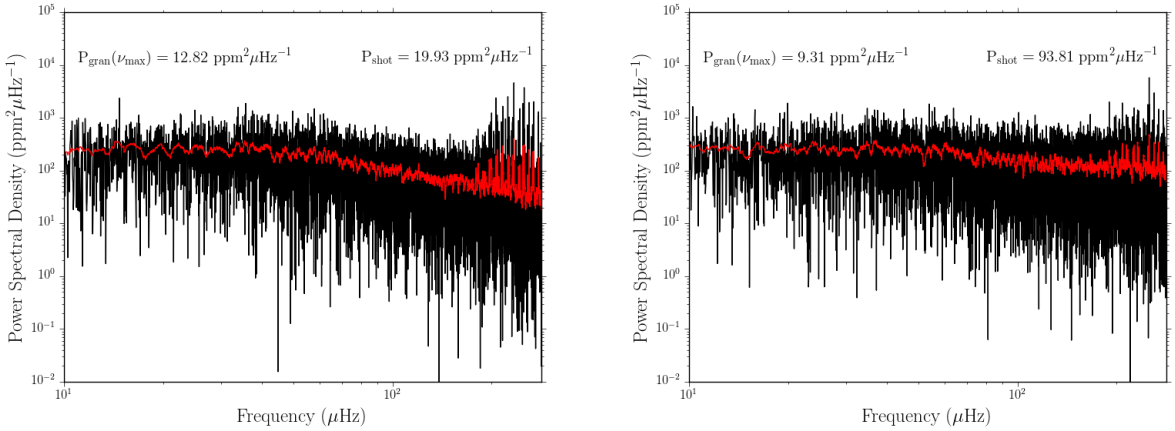


Figure 4.10: Two artificial power spectra generated simulating a Kepler-56-like star for 1 year of observations with the spectrum smoothed with a boxcar filter of width 1μHz shown in red. The left hand plot would be as observed by *Kepler* and the right by TESS. The granulation power at $\nu_{\rm max}$ and shot noise are shown in addition to give an idea of the different properties of the two space missions.

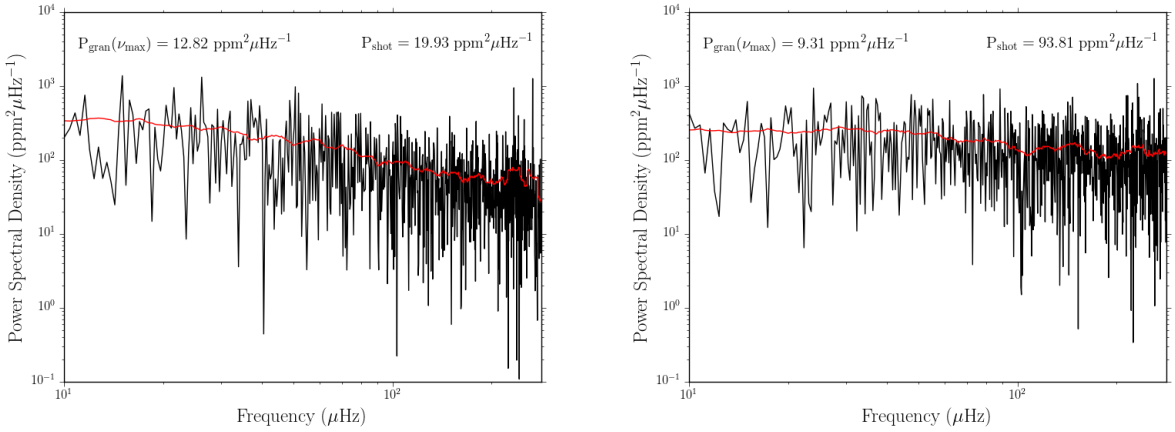


Figure 4.11: Two artificial power spectra generated simulating a Kepler-56-like star for 30 days of observations with the spectrum smoothed with a boxcar filter of width 10μHz shown in red. The left hand plot would be as observed by *Kepler* and the right by TESS. The granulation power at $\nu_{\rm max}$ and shot noise are shown in addition to give an idea of the different properties of the two space missions.

As has been shown in the previous sections, there are a large number of scaling relations that have been derived from intensity data. Therefore there is a need to convert these relations from intensity to velocity for use with radial velocity observations. The frequencies and timescales of the modes and granulation are assumed to stay the same, which is a suitable assumption considering there is nothing to the contrary observed in the Sun through intensity or velocity. Therefore the only parameters that need to be suitably scaled are the height of the Gaussian envelope describing the oscillations, $H_{\rm env}$ and the granulation amplitudes a_i. The scalings are as follows (see Basu & Chaplin 2016 for detailed explanation)

$$H_{\rm env,v} = \frac{H_{\rm env,I}}{400} \left(\frac{T_{\rm eff}}{5777 \text{ K}} \right)^{0.5}, \tag{4.38}$$

$$a_v = 0.13 \text{ ms}^{-1} \frac{a_I}{40} \left(\frac{T_{\rm eff}}{5777 \text{ K}} \right)^{0.5}, \tag{4.39}$$

where subscript v is the quantity observed in velocity (in units of ms^{-1}) and subscript I is observed in intensity.

In addition to scaling certain stellar oscillation and granulation parameters it is also necessary to update the spatial response function which will differ from the values used in intensity, the updated values are given in Table 4.1 following Kjeldsen et al. (2008).

4.7.1 SONG

The Stellar Observations Network Group (SONG) is a Danish-led project aimed at developing an inexpensive yet scientifically unique network of robotic telescopes around the world, with the task of studying the internal structure and evolution of stars (Grundahl et al., 2008). There is currently one node operational at the Observatorio del Teide on Tenerife and another being built in China. The part of the project that interests us in this context is the ability to perform time-series asteroseismology using an Iodine cell which is analogous to the way in

Table 4.3: Literature values of $\log g$ and $T_{\rm eff}$ for 7CMa along with the $\nu_{\rm max}$ value calculated from equation 4.40. The respective sources are given beneath the table.

$\log g$	$T_{\rm eff}$ (K)	$\nu_{\rm max}$ (μHz)
3.25[a]	4792	224
3.11[b]	4744	163
3.4[c]	4830	315.8
3.11[d]	4761	163
3.14[e]	4825	173
2.96[e]	4625	117
3.15[e]	4804	178
3.14[e]	4825	173

[a] (Wittenmyer et al., 2011)
[b] (da Silva et al., 2006)
[c] (Hekker & Meléndez, 2007)
[d] (Mortier et al., 2013)
[e] (Wittenmyer et al., 2016)

which we observe the Sun using BiSON (but with a Potassium cell instead).

Application to 7CMa

This subsection describes the work performed simulating artificial red giant power spectra in velocity space rather than intensity and so there are a few differences that need to be accounted for. The aim is to simulate red giant power spectra for a given star that is to be observed with SONG, which in our case is the radial-velocity planet-host 7CMa (Wittenmyer et al., 2011). The ground-based nature of the observations requires the application of a realistic window function such that tests can be performed to see to what accuracy the global asteroseismic parameters can be retrieved.

In order to calculate the global asteroseismic parameters of the star, the literature stellar parameters were used in combination with a rearranged form of the asteroseismic scaling relations given below.

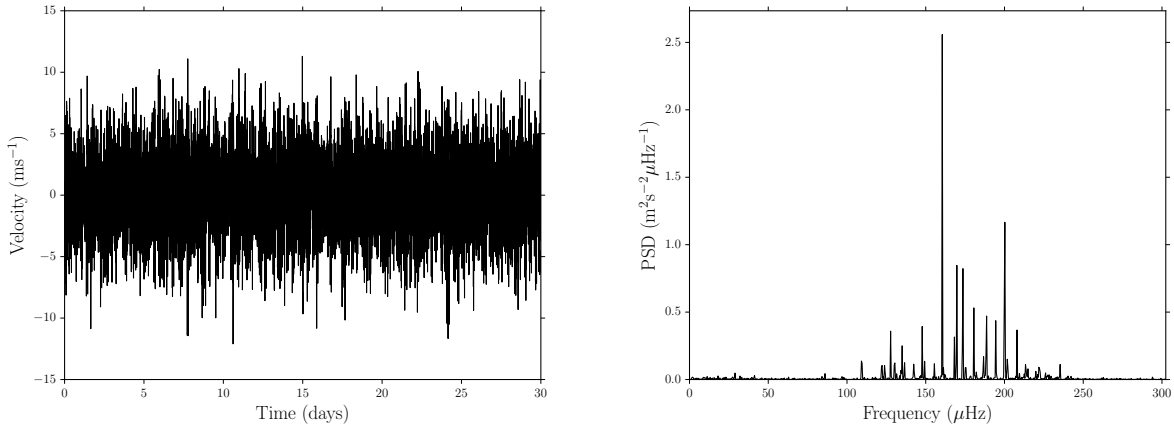


Figure 4.12: Example time series and power spectrum of 7CMa generated using the values given in the text and Table 4.3.

$$\left(\frac{\nu_{\max}}{\nu_{\max,\odot}}\right) = \left(\frac{g}{g_{\odot}}\right)\left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^{-1/2}. \tag{4.40}$$

The $\nu_{\max}$ values calculated from the literature values of $\log g$ and $T_{\rm eff}$ are given in Table 4.3. It can be seen that there is a large degree of scatter in the calculated $\nu_{\max}$ values but some seem to cluster around the 170 μHz mark, as a result the value of 173 μHz (corresponding to the first value from Wittenmyer et al. 2016 in the table) was chosen as the $\nu_{\max}$ value for the construction of the artificial power spectra (using the method described above), where $\Delta\nu$ was derived from the following scaling relation (Mosser et al., 2012a)

$$\Delta\nu = 0.276\nu_{\max}^{0.751}, \tag{4.41}$$

and the other parameters were derived as shown above. An example time series and power spectrum is shown in Fig 4.12.

Due to the fact that the data is generated as a power spectrum as opposed to a time series from which the power spectrum is then produced[10], generating the time series is not quite

[10]Generating a time series makes it easy to apply window functions and simulate gaps as opposed to performing a convolution in the frequency domain.

as simple as taking the square root of the power spectrum and taking the inverse Fourier transform. The reason for this is that any phase information in the construction of the power spectrum has been ignored if the inverse Fourier transform is taken then all the signal in the data will have the same phase (i.e. zero) and so the time series will not have the desired properties of the stochastic oscillations that we are trying to simulate. Therefore in order to generate the time series properly the method outlined by Fierry Fraillon et al. (1998) is followed which, given a power spectrum $P(\nu)$ can be converted into the following quantity, i.e. the complex spectrum

$$h(\nu) = g(\nu)\sqrt{\frac{P(\nu)}{2}} + ig^{'}(\nu)\sqrt{\frac{P(\nu)}{2}}. \tag{4.42}$$

The $g(\nu)$ and $g^{'}(\nu)$ denote two independent Normally distributed random variables (with zero mean and unit variance) and $h(\nu)$ is equivalent to the Fourier transform of our desired time series. This helps alleviated the issue above whereby all the data has the same phase and so helps give a more realistic time series. To then derive the time series the inverse Fourier transform of the quantity $h(\nu)$ in equation 4.42 can be taken. It is very important to remember that the time series will not be properly normalised and so the Parseval's theroem can be applied in reverse and noting that the square root of the ratio is needed as it is being applied to the time series and not the power spectrum.

Bearing in mind the fact that SONG only has one operational node at this time, using a time series with 100 percent coverage is not wise. Therefore the time series is multiplied by a window function corresponding to 8 hours of observation each night for 30 days. The corresponding time series and power spectrum are shown in Fig 4.13 in which the corruption caused by the window function can be seen. For more detail regarding the effect of the window function on the data see Chapter 2.

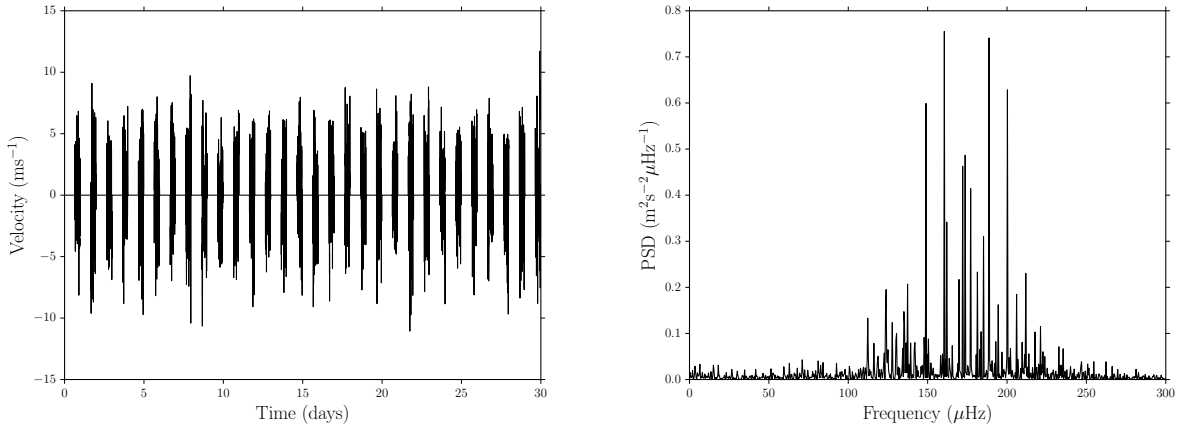


Figure 4.13: Example time series and power spectrum of 7CMa generated using the values given in the text with a window function applied corresponding to 8 hours of observations a night. Note the change in scale of the power spectrum compared to Fig 4.12.

4.7.2 HARPS

Not only can we use instruments like SONG to observe stars in Doppler velocity, but we can also use other instruments such as the High Accuracy Radial velocity Planet Searcher (HARPS). HARPS is a high resolution spectrograph designed for the explicit task of discovering exoplanets via radial velocity measurements at the ESO La Silla 3.6m telescope.

Application to KELT-11

HARPS has previously been used for time-series asteroseismology of the Sun-as-a-star observations (Kjeldsen et al., 2008; Dumusque et al., 2015) and the proposal here was to use HARPS to observe the subgiant system KELT-11 (Pepper et al., 2016) with hopes of deriving a better estimate of the stellar radius using asteroseismology.

In order to estimate $\nu_{\rm max}$ for the star we used a variant of the asteroseismic scaling relations that are independent of the stellar mass, as given by equation (20) in (Campante et al., 2016b)

$$\nu_{\rm max} = \nu_{\rm max,\odot} \left(\frac{R}{R_\odot}\right)^{-1.85} \left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^{0.92} . \tag{4.43}$$

In the above equation we use the following solar reference values $\nu_{\max,\odot} = 3090$ μHz, $T_{\text{eff},\odot} = 5777$ K and the stellar parameters are calculated using the scaling relations given in (Campante et al., 2016b) and the $\log g$ and T_{eff} from (Pepper et al., 2016). This gave $R = 2.72$ $R_\odot$ and $T_{\text{eff}} = 5370$ K. Working through the above equation with the given values the value of $\nu_{\max}$ is determined to be approximately 453 μHz.

The time series and power spectra were created in the same manner as they were for 7CMa, however this time the observing time and length of observations were changed slightly. Due to the highly competitive nature of HARPS proposals the length of the observing run was shortened from 30 days to either 7 or 14 days and the observing time reduced to 6.5 hours per night with a cadence of 180 seconds.

The simulated power spectra are shown in Fig 4.14 whereby the artefacts introduced into the spectra by the window function can be clearly seen. With such limited dataset lengths the frequency resolution greatly suffers and with such low fill (around 27%) the quality of the data is far from optimal. However with the application of proper techniques and fitting procedures it should still be possible to leverage some information even from the worst cases, but this is beyond the scope of this section and the thesis as a whole.

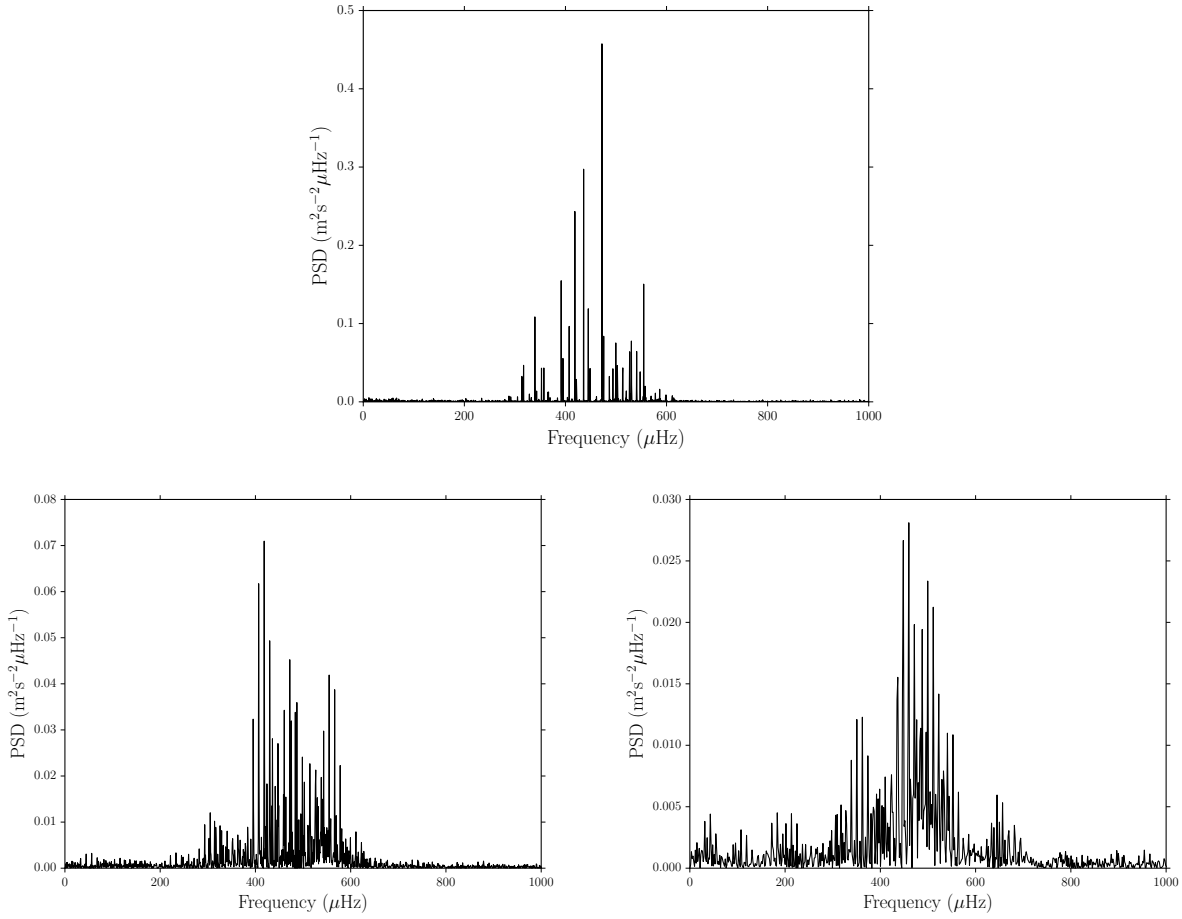


Figure 4.14: Example power spectra simulating KELT-11. The top panel shows the power spectrum with no window function applied for 14 days of observations. The bottom panel shows the power spectrum for 14 days of data with the window function applied (on the left) and for 7 days of data (on the right). Note the change in scale between the plots and the corruption caused by the window function.

5

Hierarchical Inference in the era of *Kepler*

The advent of space-based missions such as *Kepler* and CoRoT has enabled the analyses of many hundreds of stellar and exoplanetary systems. Knowledge of the properties of individual systems is very useful, but learning about the overall distribution of those properties in the context of a wider population can also be very powerful. With such large amounts of data available there is now the ability to look for the underlying population distributions from these "noisy" estimates. This idea has been put forward by Hogg et al. (2010) in the context of inferring the true eccentricity distribution of exoplanet systems given noisy estimates from radial-velocity data, but has also been used by Blanton et al. (2003) to estimate the galaxy luminosity function and Bovy et al. (2009) to model the velocity distribution in the Galactic disk (all of which applied to non-*Kepler* data). There have also been examples of hierarchical analysis applied to data from *Kepler*, for example Foreman-Mackey et al. (2014) who made

inferences about exoplanet populations given noisy, censored[1] data, and Angus & Kipping (2016) who constrained the intrinsic scatter in the data relating "stellar flicker" to both surface gravity and stellar density.

A standard question to start with could be, for example, "Given the distribution of stellar inclination angles for host stars in systems with transiting planets, what can be inferred about the obliquity distribution?". This type of question is suggestive of Bayesian hierarchical modelling, whereby we view each star having an inclination angle j and obliquity θ_j from which θ_j could be drawn from some common population distribution[2]. The important and extremely useful feature of this type of analysis is that the observed data (in this case the stellar angle of inclination) can be used to infer properties of the population distribution (i.e. the obliquity distribution) even though the obliquities of the system are not themselves observed. This is very powerful and gives us the ability to probe the population distributions for different sorts of planetary systems, for example, the difference in obliquity distributions between single and multiple planet systems.

The work presented here is inspired by the approach used in Hogg et al. (2010) for the importance sampling representation of the hierarchical modelling, and Morton & Winn (2014) for the representation of the obliquity given knowledge of the angle of inclination of the star. Let us firstly go through the model being applied to the data, the hierarchical method and then consider the data themselves. This same methodology will also be applied later in this thesis in the context of inferring the population inclination angle distribution for low luminosity red giant stars observed with *Kepler*. All details of the hierarchical process and its derivation are given in Appendix A.10.

[1]Data where the measurement is only partially known.

[2]The notation used above (j and θ_j) is simply for illustrative purposes and those quantities will not be denoted as such from now on.

5.1 Inferring the obliquity distribution with *Kepler* KOIs

5.1.1 Overview - what is the obliquity?

The obliquity, ψ, is defined as the angle between the rotation axis of the star and the orbital axis of the planet. It is an important property of any planetary (or stellar in the case of binaries) system due to its ability to provide important diagnostic constrains with regards to planet formation theory along with planetary migration and the tidal evolution of systems (Schlaufman, 2010; Winn et al., 2010). This is due to its dependence on factors such as the alignment of the primordial disk and dynamical interactions either within the system or from external bodies. For an excellent review of this topic see Winn & Fabrycky (2015) and references therein.

Although ψ is an important quantity, it is difficult to measure in practice due to its reliance on two measured angles. The first angle is the sky-projected obliquity which is generally referred to as λ. This is important because it gives information about how close the stellar rotation axis and orbital momentum vector are in the plane of the sky. Information on this projection is hard to come by since the other quantities explained here give us no access to it. For transiting exoplanetary and binary systems the Rossiter-McLaughlin (RM) effect can be used to derive λ through the modulation of stellar absorption lines as a result of the transiting body (e.g. Winn 2007). There are also other methods that can be used to determine λ such as Doppler tomography (Collier Cameron et al., 2010a,b; Gandolfi et al., 2012; Bourrier et al., 2015; Zhou et al., 2016), gravity-darkening (Barnes, 2009; Barnes et al., 2011, 2013; Ahlers et al., 2014; Masuda, 2015; Ahlers et al., 2015) and from the interaction of transits and star spots e.g. (Désert et al., 2011).

Two other angles are needed before the obliquity can be calculated, the angle inclination of the star i_s and, in the case of a transiting system, the planetary inclination angle i_o. The latter can be determined through modelling of the transits observed in the light curve as seen

in Chapter 7, whilst the former is a little more tricky. The angle of inclination of the star is defined as the angle between the star's rotation axis and our line of sight. It is possible to estimate it using photometry via spot-modelling, however this relies on the presence of star spots modulating the photometric signal from the star (Dumusque, 2014). Another possible method involves combining the above photometry of star spots and spectroscopy through the use of the rotation period of the star $P_{\rm rot}$, the radius of the star $R_\star$ and the rotational velocity of the star with respect to our line of sight $v \sin i_{\rm s}$. Work of this kind has been carried out by Abt (2001) and has been updated recently by Morton & Winn (2014) making it more robust statistically, especially in the unphysical regime where $\sin i_{\rm s} > 1$. It is possible for $\sin i_{\rm s} > 1$ if $v \sin i_{\rm s}$ is greater than the estimated v from the radius and rotation period.

5.1.2 Geometry of the Obliquity

Simply explaining the obliquity in terms of angles without a proper description of the geometry of the system is unhelpful and will only lead to confusion, therefore this section will go over the geometry behind the obliquity and the other angles involved.

There are two coordinate systems that are useful when explaining the geometry behind the obliquity, "observer-oriented" and "orbit-oriented". Each system gives a different insight into the angles that are being inferred as a result of a given measurement and Fig 5.1 demonstrates each orientation. In the case of the observer-oriented coordinate system the orbital angular momentum unit vector $\mathbf{n}_0$ can be described completely by the angle between the planet's orbital axis and line of sight, i_o, and can be extracted for the majority of transiting systems. The determination of the vector of stellar rotation $\mathbf{n}_s$ is a little more complicated and cannot simply be determined from the angle between the stellar rotation axis and line of sight, i_s. Since it carries no information about the projected component of $\mathbf{n}_s$ another angle is needed in the form of λ to fully define $\mathbf{n}_s$.

It is common practice to specify the intervals over which these angles are defined, partly in

an attempt to avoid degeneracies, but also to aid interpretation. Both i_o and i_s are defined between 0 and $\pi/2$, this is because unless there is information on the projected angle given by λ in the case of, for example, i_s the angle θ will look identical to the angle $\pi - \theta$. Additionally, λ is defined between $-\pi$ and π, which results in ψ being defined between 0 and π, where values between 0 and $\pi/2$ denote prograde orbits and those above $\pi/2$ denote retrograde orbits. Stars with low obliquities, $|\psi| \approx 0$, are consistent with spin orbit alignment whilst those with high obliquities are not, e.g. XO-3 (Fabrycky & Winn, 2009). The final angle, the azimuthal component ϕ varies between $-\pi$ and π where π is defined as being along the line of sight.

The determination of ψ from i_o, i_s and λ is given by Fabrycky & Winn (2009)

$$\sin i_s \sin \lambda = \sin \psi \sin \phi, \tag{5.1}$$

$$\cos \psi = \sin i_s \cos \lambda \sin i_o + \cos i_s \cos i_o, \tag{5.2}$$

$$\sin \psi \cos \phi = \sin i_s \cos \lambda \cos i_o - \cos i_s \sin i_o. \tag{5.3}$$

Since the aim is to estimate ψ, equation 5.2 gives the opportunity to do so using measurements of i_o, λ and i_s without any knowledge of ϕ, however it is actually equation 5.3 that is of more use. For a transiting system it is known that $\sin i_o \approx 1$ and so equation 5.3 reduces to

$$\sin \psi \cos \phi = \cos i_s. \tag{5.4}$$

.

The reason for the sign change in the above equation is due to the fact that when deducing the angle of inclination (using asteroseismology) there is no means to distinguish between i_s and $180 - i_\mathrm{s}$. This means that $\cos i_s = -\cos i_s$ as a result and so the sign change is a result

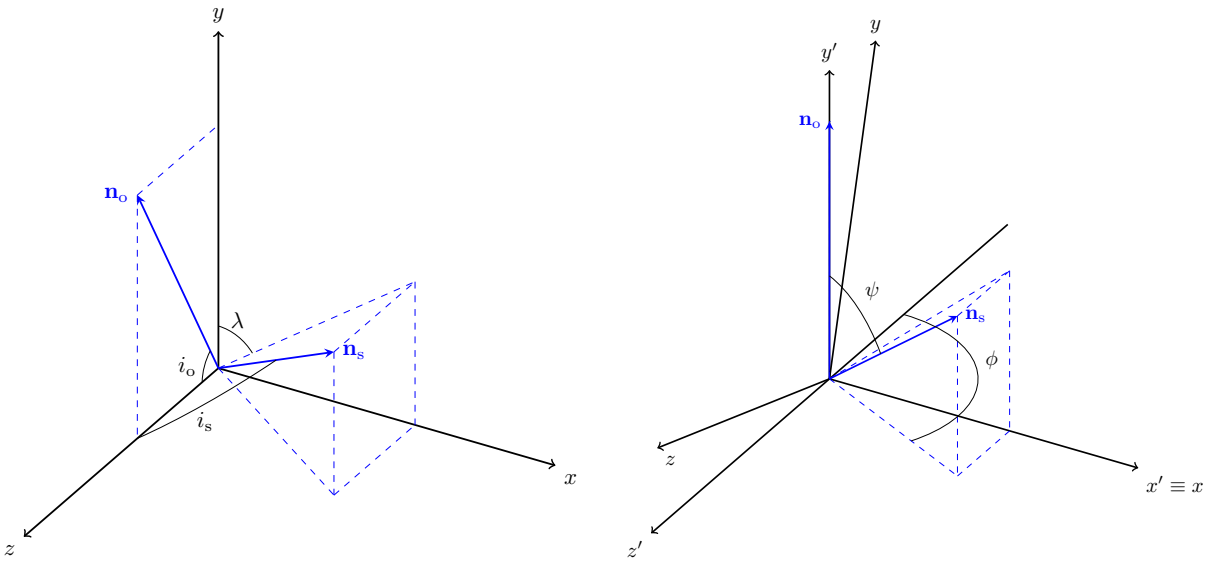


Figure 5.1: A demonstration of the geometry involved in determining the obliquity. The left-hand plot shows the observer-oriented system where the z-axis is along the observer's line-of-sight. The right-hand plot shows the orbit-oriented coordinate system, this is a rotation of $\pi/2 - i_o$ anti-clockwise in the x-axis from the observer-oriented coordinates. The angles are given as follows: ψ is the obliquity of the system, ϕ is the azimuthal angle, i_o is the orbital inclination angle, i_s is the stellar inclination angle and λ is the projected obliquity. The unit vectors $\mathbf{n}_o$ and $\mathbf{n}_s$ denote the directions of the stellar rotation and orbital angular momentum axes respectively. (Figure reproduced from Campante et al. 2016a)

of that. This is extremely useful and enables the obliquity to be inferred simply given the inclination angle of the star and some assumptions made about the transiting body. The limited-scope for obtaining measurements of the RM effect, due to its scaling with $(R_p/R_s)^2$ leading to a decidedly small anomalous Doppler shift, can result in λ being very difficult (not to mention time-consuming) to extract for a large sample of stars. Therefore, having a method that does not require a measurement of λ is much simpler, although it does not provide sensitivity to prograde or retrograde orbits (which is not a problem when trying to infer the population distribution).

5.2 How to determine the angle of inclination using asteroseismology

Asteroseismology gives us a great opportunity to measure the inclination angle of a star that is independent of other techniques (such as through the rotation period, radius and $v \sin i$).

The main caveat is that in order to deduce the inclination angle non-radial modes need to be seen that are split by the presence of rotation (Pesnell, 1985; Gizon & Solanki, 2003), and if the effects of rotation are not observed then the inclination angle cannot be extracted.

As explained in Chapter 1 the modes seen in solar-like oscillations can be characterised according to their radial order n, degree ℓ and azimuthal order m. The degree of the oscillation mode that is important for determining the angle of inclination is in our case $\ell = 1$. This is partially due to the fact that modes with degree ℓ are split into $2\ell + 1$ azimuthal components under the influence of rotation. Radial modes ($\ell = 0$) are unsuitable due to being unaffected by rotation and $\ell = 2$ are much more difficult to extraction rotational splittings from due to the dependence of the rotational splitting on degree and their increased mode linewidth in addition to the lower signal-to-noise due to lower mode visibility (for main-sequence stars).

To first order, the rotational splitting of a mode with radial order n, degree ℓ and azimuthal order m can be defined by (Ledoux, 1951)

$$\nu_{n\ell m} \approx \nu_{n\ell 0} + m\nu_{\rm s}, \tag{5.5}$$

where $\nu_{n\ell 0}$ is the frequency of the central component and $\nu_{\rm s}$ is the rotational splitting. An approximation is made in the above equation whereby the rotation splitting can also be given by (assuming a constant velocity throughout the star)

$$\nu_{\rm s} \approx m\frac{\Omega}{2\pi}(1 - C_{n\ell}), \tag{5.6}$$

where Ω is the angular velocity of the star and $C_{n\ell}$ is the Ledoux constant. The nature of the mode will affect the value of the Ledoux constant, for p-modes $C_{n\ell} << 1$ and for g-modes $C \approx 0.5$ which will come into play more when considering rotational splittings in red giants (as shown in Chapter 4).

An important assumption in the derivation of the angle of inclination is the equipartition

of energy between the mode components, i.e. between those of same $|m|$, which in turn simplifies the expression used for the modes. This equipartition of energy manifests itself in the mode visibility, $\mathcal{E}_{\ell m}(i_{\rm s})$, which is dependent on the degree, azimuthal order and angle of inclination of the mode (the derivation of which is given in Appendix A.11.2).

Let us look at a few examples to confirm the picture of how angle of inclination affects the mode profile. The consequences of these equations mean that if a star was oriented at approximately 90° to our line of sight then only the outer components of the $\ell = 1$ modes would be seen; alternatively if the star was oriented at approximately 0° to our line of sight (i.e. pole on) then only the central $m = 0$ component would be seen. Therefore the observed profile of the non-radial modes can shed light on the angle of inclination of the star before the fitting process is even performed.

An example showing how the amplitude of the individual mode components change as a function of angle is shown in Fig 5.2. However, in Figs 5.3 and 5.4 both the limit spectrum and limit spectrum with noise added are shown. It is important to see that for low angles (from $\sim 0 - 30°$) the modes look very similar and differentiating between those angles is difficult, especially at lower signal to noise. The same is apparent for intermediate angles where, although when looking at the limit spectrum the differences are obvious, the differences in the power spectrum with noise applied are much less so[3].

5.3 Applications to *Kepler*

5.3.1 A statistical approach to constraining spin-orbit alignment

There have a been a number of projects investigating the spin-orbit alignment of both exo-planet (e.g. Chaplin et al. 2013) and binary systems (e.g. Albrecht et al. 2007). The stellar inclination angles for many *Kepler* candidates were derived by Hirano et al. (2012) who were

[3]Although this does depend on the difference in magnitude between the rotational splitting and mode linewidth.

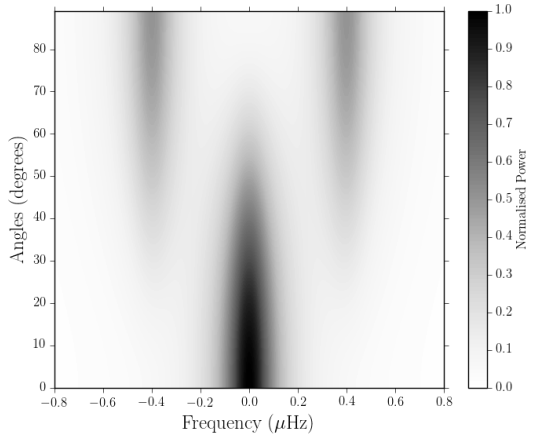


Figure 5.2: A contour plot showing the relative mode power as a function of inclination angle for an $\ell = 1$ mode. The amplitude was taken to be unity, the linewidth was 0.2 μHz and the rotational splitting was 0.4 μHz.

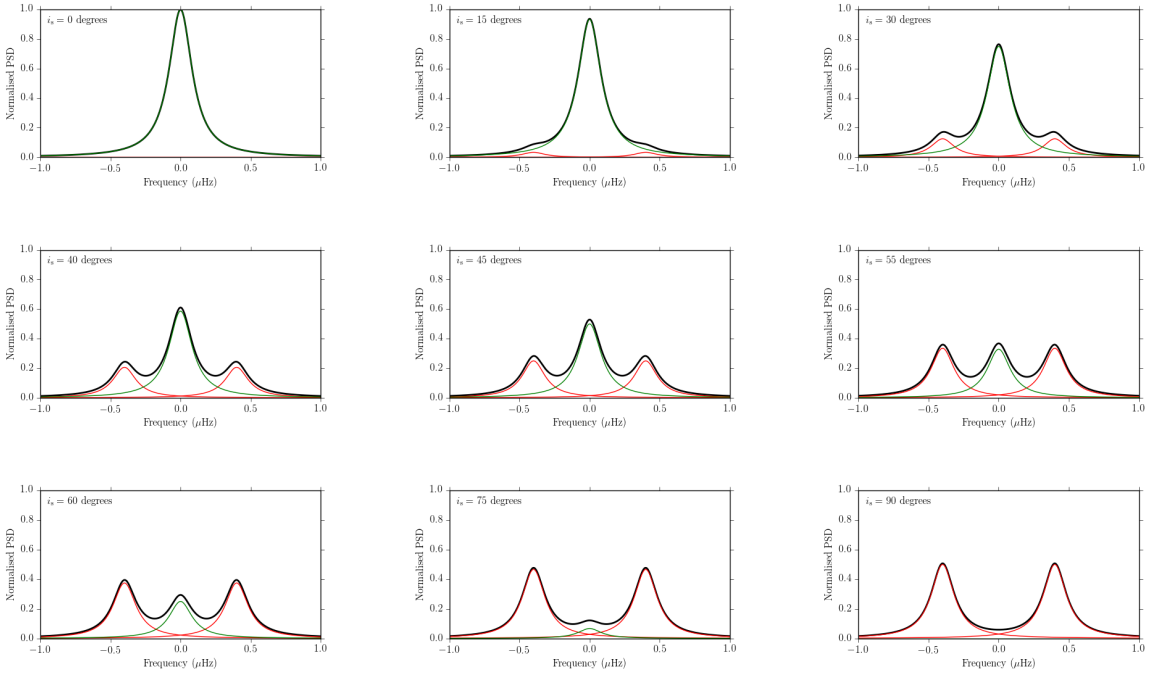


Figure 5.3: Limit spectra for an $\ell = 1$ mode observed at different inclination angles (shown in black), the outer components are show in red, the central components are given in green. The angle of inclination used is labelled along with using a mode linewidth of 0.2 μHz and a rotational splitting of 0.4 μHz for an observing length of 4 years.

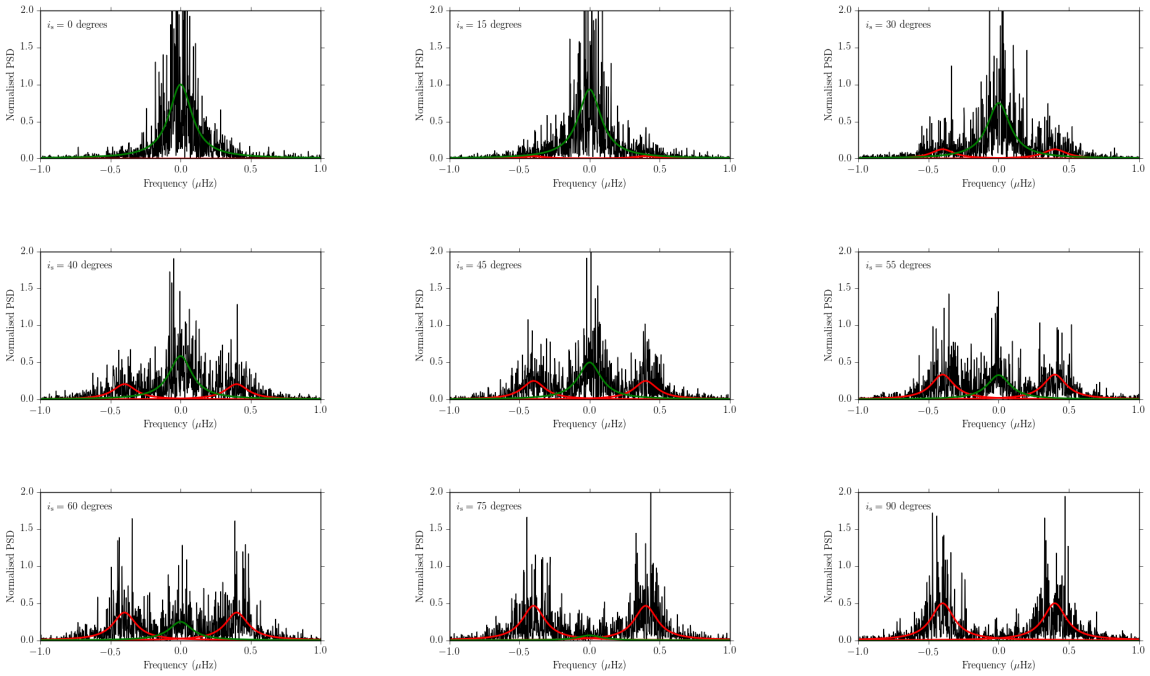


Figure 5.4: The same limit spectra given in Fig 5.3 but this time with χ^2_2 noise applied.

looking for spin-orbit misalignment in individual systems. They found at least one system (KOI-261) which exhibited possible spin-orbit misalignment and in their follow-up work (Hirano et al., 2014) they found three multiple-planet systems that showed possible spin-orbit misalignment.

Looking at the properties of individual systems is extremely useful, but (as noted previously) in the era of large photometric surveys such as *Kepler* and CoRoT there is the opportunity to look at the sample as a whole and look at ensemble properties. The first in-depth look in a statistical sense was performed by Hirano et al. (2014) who, in addition to deriving inclination angles for *Kepler* candidates, also showed that since the stellar inclination angles distribution of the host stars in their sample deviated from isotropy the obliquity distribution also could not be isotropic. This comes about because if the inclination angle distribution is not isotropic and all the systems are transiting, then the stellar rotation axis and planetary orbital axis are correlated. If this is the case then some stellar inclination

angles are more likely to be seen than others (such as those close to 90°), therefore showing the obliquity distribution cannot be isotropic.

This statistical result was followed up by Morton & Winn 2014 who expanded the initial sample of 25 KOIs from Hirano et al. (2014), to 70 KOIs and performed a similar analysis. Their conclusions align with those from Hirano et al. (2014): multi-transiting systems observed with *Kepler* were observed to be better aligned than those that are single-transiting systems.

The hope with this chapter is to show how asteroseismology can enter the fold and help contribute to this area of research, to complement other techniques and add another independent method for deriving stellar inclination angles, adding to work already performed by e.g. Benomar et al. (2014).

5.3.2 Modelling the obliquity distribution

In section 5.1.2 an equation was derived that enabled the calculation of the obliquity given the angle of inclination of the star and the azimuthal angle. This is the first step towards modelling the distribution of ψ for a given inclination angle. As a refresher, here is the equation again

$$\cos i_{\rm s} = \sin\psi \cos\phi, \tag{5.7}$$

where it is assumed that the inclination angle of the planet orbit[4] is close to 90° and ϕ is the azimuthal angle of the stellar rotation axis. Any small deviation from $i_o = 90°$ will not have a large impact on equation 5.7 so long as $\sin i_o \approx 1$.

The first distribution that would seem appropriate for ψ is an isotropic distribution, where $p(\psi) \propto \sin\psi$. This was shown to be inadequate by Winn et al. (2006) who analysed the sky-projected angle λ for 3 systems and showed that even with so few points they were

[4]Defined as the angle between the plane of the orbit and the stellar rotation axis.

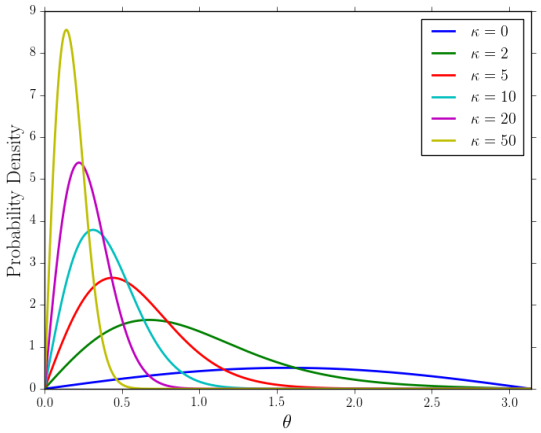


Figure 5.5: Probability density functions describing the Fisher distribution given by Eqn 5.8 for various values of κ.

incompatible with an isotropic distribution. This can, as alluded to earlier, be interpreted as a correlation between $\mathbf{n}_s$ and $\mathbf{n}_o$, and so this might be expected to be due to the mechanisms behind planetary formation and migration. Instead of modelling ψ as an isotropic distribution we will follow the work of (Fabrycky & Winn, 2009) and use a Fisher distribution (adopting the same notation as Morton & Winn 2014)

$$f_\psi(\psi|\kappa) = \frac{\kappa}{2\sinh\kappa}\exp(\kappa\cos\psi)\sin\psi, \tag{5.8}$$

where κ is a concentration parameter. The best way to visualise this distribution is analogous to a normal distribution on a sphere with zero mean and as such κ describes the width and how peaked the distribution becomes. The important properties of this distribution lie in its limits and so when $\kappa \to 0$ the distribution tends towards isotropy, i.e. $f_\theta \propto \sin\psi$, whereas for large κ it tends to a Rayleigh distribution with width $\sigma = \kappa^{-1/2}$ as seen in Fig 5.5. In addition ϕ is assumed to be uniformly distributed between 0 and 2π. This is uninformative since there is no means to infer this angle from the data and there also should not be any preferential direction either.

The above distributions certainly help our aim to derive a model for $\cos i_\mathrm{s}$, but their

final distributions are not what is needed. The derivation of the final model is given in Appendix A.11.3 and results in the model given below

$$f_{\cos i_s}(z|\kappa) = \frac{2\kappa}{\pi \sinh \kappa} \int_z^1 \frac{\cosh\left(\kappa\sqrt{1-y^2}\right)}{\sqrt{1-y^2}\sqrt{1-(z/y)^2}} \mathrm{d}y. \tag{5.9}$$

A choice of prior on the concentration parameter κ is also needed to finalise the inference, for which we had a choice of a uniform prior or the prior given in Fabrycky & Winn (2009) whereby $p(\kappa) \propto (1+\kappa^2)^{-3/4}$. The latter prior has the properties that $p(\kappa) \to$ constant. Having compared the two choices of priors it was noted that it made very little difference to the posterior distributions and so a uniform prior was taken on κ.

This hierarchical inference was carried out in a grid-based manner in which a grid of κ values, 0 to 200, (and later for the mixture models f varying from 0 to 1, which will be covered later) were searched over and the posterior probability calculated at each point. This was done for simplicity and ease of computation, however due to the more inefficient exploration of the parameter space the uncertainties will be larger than those if it were to be explored using MCMC. Although this is dependent on whether the grid spacing is much less than the width of the posterior distribution.

5.3.3 Asteroseismic sample

Asteroseismology gives a great opportunity to measure the angle of inclination of stars therefore enabling the inference of the obliquity for transiting systems. Therefore to start with an asteroseismic sample will be used, consisting of 25 solar-type KOIs (see Table 1 of Campante et al. (2016a) for the full data) for which the angle of inclination could be extracted[5]. Huber et al. (2013b) provided a full characterisation of 66 planet-candidate hosts observed with *Kepler* using asteroseismology. Whilst there are a considerable number of *Kepler* exoplanet

[5]None of the asteroseismic analysis in the paper Campante et al. (2016a) was carried out in this work, only the statistical analysis.

host stars that show solar-like oscillations, in order to determine the angle of inclination there must be some signatures of rotation present in the power spectrum (as explained in section 5.2). The smaller sample size introduced as a result is not a problem since every system will help provide constraints on the obliquity distribution in some sense, whether that is ruling out isotropy or showing some more distinct alignment.

Within the sample of 25 stars, 16 were multiple-planet systems. It is also worth noting that of all the multiple-planet systems only Kepler-93 (Ballard et al., 2014) and Kepler-410A (Van Eylen et al., 2014) are not multi-transiting. The other stars in the asteroseismic are assumed to be single-planet systems. The sample was also limited to stars with $T_{\rm eff} < 6200$ K. This is due to the fact that the mode linewidth increases with effective temperature and so leads to the rotationally split components being blended together resulting in an inability to derive the rotational splitting and consequently angle of inclination. In addition it is interesting to note that there is only one hot-Jupiter (HAT-P-7, Pál et al. 2008) system, whereas the rest of the sample appears to be much smaller in planetary radius.

To perform the hierarchical inference, the posterior PDFs of the inclination angle for each star are needed as an input. The asteroseismic analysis provided 80,000 samples (from the MCMC fitting) from each star but this would make the computation prohibitively slow and so it was decided to randomly draw 1000 samples from each star as a compromise. An example of this is shown in Fig 5.6 for KIC 3544595 and the good agreement between the reduced sampling and the original distribution[6]; this was also the approach adopted in Morton & Winn (2014).

In order to produce a combined distribution for all the stars in our sample, the posterior distributions were concatenated together and this resultant distribution is shown in Fig 5.7. The isotropic distribution is given as the black dotted line (flat in $\cos i_s$) and the fact that

[6]It is possible for some additional structure to be added into the posteriors when only using 1000 samples, however the large uncertainties inherent to the distributions themselves act to cancel out any effect due to this.

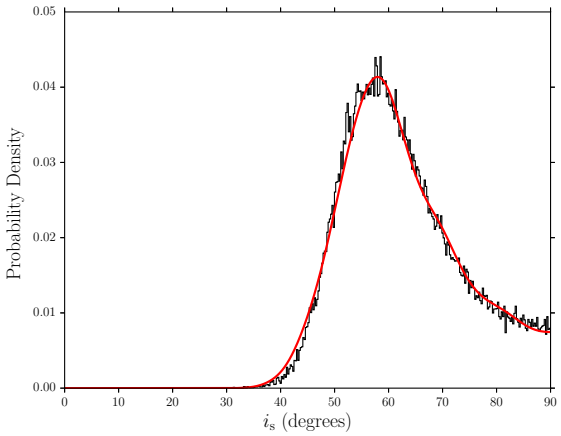


Figure 5.6: An example posterior PDF for KIC 3544595 where the histogram of 80,000 samples is shown in black. The red line shows the kernel density estimate using 1,000 randomly chosen samples (without replacement).

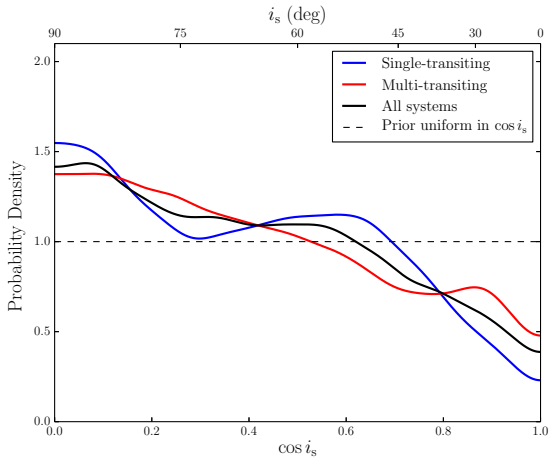


Figure 5.7: Kernel density estimates of the combined posterior distributions of $\cos i_s$ coloured according to the properties of the systems. Those with only one observed transit planet are shown in blue, those with multiple observed transiting planets are shown in red and the combined distribution for all the systems is given by the black solid line.

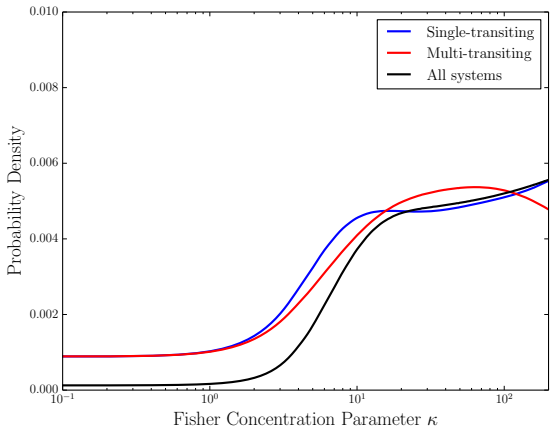


Figure 5.8: Kernel density estimates of the concatenated posterior distributions of $\cos i_s$ as extracted from the KOI in the asteroseismic sample. The different types of system are colour-coded and the prior used in the fitting process is given by the dashed black line.

the distributions are not isotropic would suggest that the angle of inclination of the stars in our sample have a greater tendency to be aligned at 90° with respect to our line of sight. Combined with the transiting nature of these systems this would imply that the two quantities ($\mathbf{n_s}$ and $\mathbf{n_o}$) are correlated, as suggested by previous analyses.

The posterior distribution for the concentration parameter κ is shown in Figure 5.8 for the entire sample and also for systems with only a single transiting planet[7] or multiple transiting planets. The logarithmic scale helps show the fact that although the asteroseismic sample cannot constrain the values of κ very well (as seen in Table 5.1) it can certainly show that the sample is inconsistent with isotropy. The prior can be seen to be dominating at large κ where the asteroseismic sample does not offer much constraint to the concentration parameter, which comes as a result of the small sample size and large uncertainties on the inclination angle. Simply being able to rule out isotropy is a valuable result and helps confirm findings from other works.

The results presented here however, do not agree with all works as they cannot support the

[7]The wording here is important as a single transiting planet may have more planets that are not necessarily orbiting in the same plane, however it is assumed that it is unlikely for a system with many transiting planets to be misaligned.

conclusions of Morton & Winn (2014), i.e. that planetary systems with a single transiting planet were found to have larger obliquities than those with multiple transiting planets. It is worth noting however that the sample used was over three times larger and only partially overlapped with our sample, therefore in order to try and put tighter constraints on the obliquity distribution, the sample size must be increased and to do this we include the sample used from Morton & Winn (2014) with a few changes that will be described below.

5.3.4 Morton & Winn sample

The sample of stars originally analysed in Morton & Winn (2014) can also be included in the analysis to create the largest sample of stars for which this type of analysis has been performed. The posteriors will be amalgamated into a combined sample for further analysis. The sample includes the majority of the stars analysed in Hirano et al. (2014) and an additional 41 KOIs (Kepler Objects of Interest) which have both rotation periods and spectroscopic parameters.

In order to prepare the data from Morton & Winn (2014), the values of R, $P_{\rm rot}$, and $v \sin i$ were taken from those given in Table 1 of the same paper. Then the same methodology was adopted whereby v was estimated from R and $P_{\rm rot}$. The posterior distribution of $\cos i_{\rm s}$ given the distributions of v ($p_{\rm V}$) and $v \sin i_{\rm s}$ ($p_{{\rm V}_s}$) follows

$$p(\cos i_{\rm s}|D) \propto \int_0^\infty p_{{\rm V}_s}(v) p_{\rm V}\left(\frac{v}{\sqrt{1-\cos^2 i_{\rm s}}}\right) {\rm d}v. \tag{5.10}$$

This provides a statistical framework for determining the distribution of $\cos i_{\rm s}$ from v and $v \sin i$ as opposed to simply dividing $v \sin i$ by v and taking the inverse of the result. For consistency with the asteroseismic sample, 80,000 samples were generated in total for each star and from that 1,000 samples were randomly sampled for use in the inference.

5.3.5 Combining different datasets

The beauty of hierarchical inference is that it automatically copes with heteroscedastic data (i.e. where the variance is not the same from point to point) and so additional datasets can be incorporated into the analysis to see if they provide additional constraints. To do this the 62 additional systems from Morton & Winn (2014) that did not overlap with our sample were added (where 39 were single-transiting and 23 were multiple planet systems). In what follows, the asteroseismic sample will be denoted by a subscript a and the Morton & Winn (2014) sample using a subscript m.

In order to take full advantage of the two datasets a likelihood function can be constructed that accounts for them, i.e.

$$\ln \mathcal{L} = \sum_i \ln \mathcal{L}_{a,i} + \sum_j \ln \mathcal{L}_{m,j}, \tag{5.11}$$

where i denotes the number of stars in the asteroseismic sample and j the number in the Morton & Winn (2014) sample and the individual likelihood functions are those given by equation A.69. This combination of the likelihoods assumes no covariance between the methods, i.e. that they are independent. The assumption should be upheld as the data is generated in two different ways using different input data, and the measurement uncertainty is accounted for in the posterior distributions.

The posterior distribution for the concentration parameter when applied to the combined sample is shown in Fig 5.9. The shape of the distributions obtained for the single-transiting and all the systems together look remarkably similar to those in Morton & Winn (2014). This suggests that the asteroseismic sample does not additionally constrain the model which will in part be due to the low number of stars and larger uncertainties in the posterior inputs. The dashed black line in Fig 5.9 accentuates this, where the difference in constraints offered by the two different samples (at least for the single systems) is clear. The interpretation for

Table 5.1. Results from the full Bayesian hierarchical analysis on each sample (reproduced from Campante et al. (2016a)).

Model	Single-transiting			Multi-transiting			All		
	κ	f	E	κ	f	E	κ	f	E
Asteroseismic Sample									
Single-Fisher	106^{+92}_{-94}	···	···	101^{+96}_{-88}	···	···	107^{+93}_{-89}	···	···
Combined Sample									
Single-Fisher	$8.7^{+6.7}_{-5.4}$	···	1.10×10^5	118^{+82}_{-83}	···	5.11×10^7	$11.5^{+7.5}_{-5.7}$	···	1.61×10^9
Mixture	126^{+74}_{-86}	$0.21^{+0.24}_{-0.18}$	1.24×10^7	116^{+84}_{-82}	$0.08^{+0.24}_{-0.08}$	6.16×10^6	129^{+71}_{-80}	$0.13^{+0.16}_{-0.12}$	7.97×10^{11}

Note. — We quote the median and 95.4 % HPD credible region for both κ and f. The Bayesian model evidence, E, is reported for each cohort of systems in the combined sample.

all systems has to be carefully made due to the fact that there is not an equal number of single and multi-transiting systems and the sample is clearly dominated by the single systems. Therefore the posteriors for all systems are shown for completeness rather than for specific inference.

This does however mean that to some degree the asteroseismic data has a precision that is too low to help constrain the distribution. The sample size also has a slight effect given that the asteroseismic sample is a third of the Morton & Winn sample.

The use of a different prior also highlights some thoughts on the methodology that need to be expressed. The use of an uninformative prior (which in our case is really weakly informative, see Gelman et al. 2009) is desirable because the idea is to stop any unwanted or unphysical solutions whilst not inserting any biases into the procedure. The multi-transiting systems are of interest in this case since in Morton & Winn (2014) (Fig 6) the posterior probability distribution for κ can be seen to peak at $\kappa \approx 19.1$ and proceed to tail off towards large κ. However when a uniform prior on κ is applied in our work, the posterior distribution for the multi-transiting systems (in both samples) does not peak but flattens off over large κ values. This would be indicative of a lack of constraint offered by the data, because in both cases the prior must be doing the work and influencing the posterior. The data is therefore not informative enough to offer any firm conclusions, other than the tendency for alignment.

In addition to a single Fisher distribution as the model for the obliquity distribution, a mixture model can also be used. This consists of the sum of an isotropic distribution and the aforementioned Fisher distribution. The physical motivation behind this model is that the Fisher distribution describes one migration mechanism, for example disk migration that is expected to be common for systems with multiple transiting planets, and an additional mechanism is accounted for by the isotropic distribution which allows much higher obliquities, for example scattering which may be more common in systems with hot Jupiters.

The mixture model is defined as follows

$$f_\alpha(\cos i_\mathrm{s}) = f + (1-f)p_{\cos i_\mathrm{s}}(\cos i_\mathrm{s}|\kappa), \tag{5.12}$$

where the first component describes the isotropic distribution (which is flat in $\cos i_\mathrm{s}$) and the second component describes the previously derived model for the Fisher distribution.

The posterior distributions for the parameters f and κ are shown in Fig 5.10 and results given in Table 5.1. The most apparent result, and a reassuring one as a sanity check, is that there is no suggestion of isotropy from the multi-transiting systems. This also confirms the results from the single Fisher model and the currently accepted mechanisms for planet migration in multiple planet systems. In contrast the single-transiting systems do suggest some degree of isotropy $f \approx 0.21$ but then the concentration parameter for the Fisher distribution cannot be well-constrained.

Whilst these results may hint at there being two different migration mechanisms in single-transiting systems and only one for multi-transiting, the two models used here need to be compared in a statistically consistent manner. To do so the Bayesian evidence was computed for each model (which is performed by integrating the likelihood over all parameter space) and the models compared by making use of the Bayes factor, the ratio of the two evidences for each model, F_B.

In order to draw conclusions from the use of the Bayes factor the scale introduced by Kass & Raftery (1995) is used, whereby the logarithm of the Bayes factor is analysed. For the single-transiting sample the logarithm (base-10) of the Bayes factor is 2.05 in favour of the mixture model, and so according to Kass & Raftery (1995) this is decisive evidence (in favour of the mixture model). This is in contrast to the multi-transiting sample where the Bayes factor is 0.92 in favour of the single Fisher model, which is defined as substantial evidence. These results from the model comparison also adhere to what has been seen in the posterior distributions, for example the fact that the multi-transiting sample always produced a well-aligned result no matter the model.

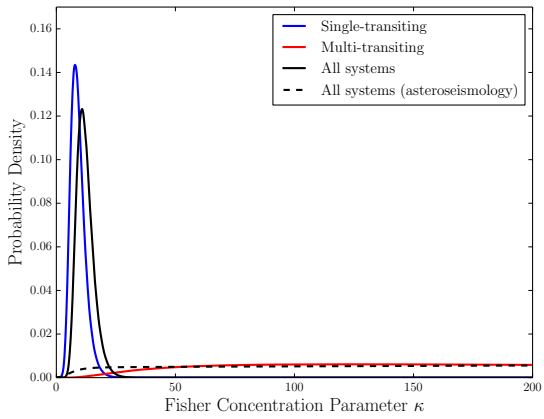


Figure 5.9: Posterior probability distributions of κ (the concentration parameter) for the combined sample under the Fisher distribution model. The three different sub-samples are coloured as before and the posterior distribution for the asteroseismic sample is also shown as the dashed line for reference.

In the case of the single-transiting systems the lack of constraint on κ whilst the relatively good constraint on f would suggest that the isotropic does a reasonably good job of describing part of the data, but perhaps the other part is not well described with the Fisher distribution. Before anything more can be said on this subject this same analysis would need to be carried out with a large sample size, since the small sample and large uncertainties will hinder any attempt to derive tight constraints.

5.4 Conclusion

In conclusion, Bayesian hierarchical inference has been applied to the problem of inferring the obliquity distribution of a large sample of stars. There is agreement with previous works that the stellar rotation and orbital axes are possibly correlated, indicated by the deviation from isotropy of the stellar inclination angle distribution. But also some disagreement with Morton & Winn (2014) when using only the asteroseismic data, likely due to the differences in the small samples.

The obliquity distributions of both single- and multi-transiting systems have been derived where for single-transiting systems a mixture model is favoured over a single Fisher distri-

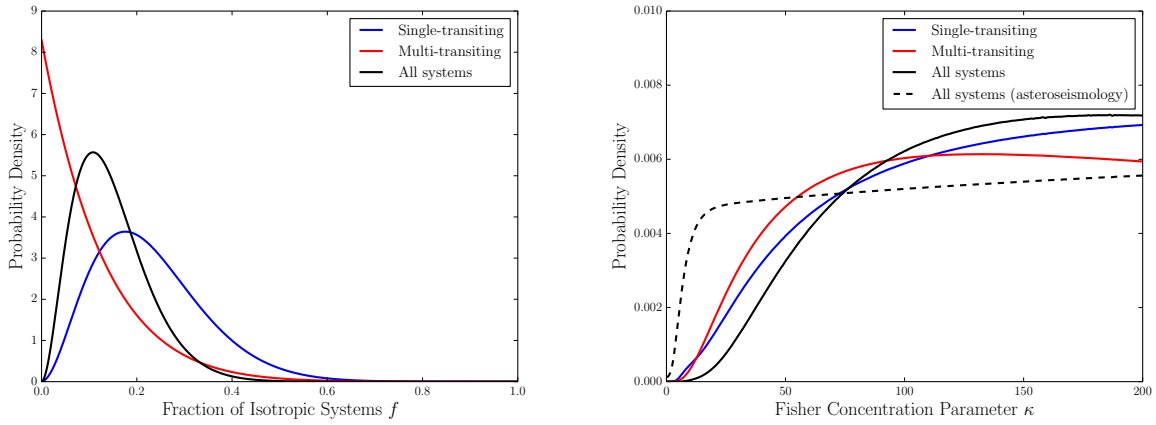


Figure 5.10: Posterior probability distributions of f and κ for the mixture model. The different sub-samples are coloured as before the the asteroseismic sample result is shown again for reference.

bution, whereas the opposite is the case of multi-transiting systems. This may suggest that different migration mechanisms are at play for the different system types. The high obliquity systems would need a migration mechanism that provides a way of exciting the orbits to large obliquities, perhaps in the form of planet-planet scattering (Rasio & Ford, 1996; Nagasawa et al., 2008). The multi-transiting systems agree with the observations made by Albrecht et al. (2013b) and show that the obliquities are low, as shown by the large concentration value κ inferred in the hierarchical analysis.

The application of this updated method to the upcoming space missions such as TESS and PLATO will offer many more stars to add to this sample. The fact that the stars, especially for TESS, will be much brighter will greatly improve the ability to perform spectroscopic follow-up and will result in many stars (hopefully) for which $v \sin i_{\rm s}$ is available. Coupled with the asteroseismic targets, this opens up the possibility of greatly increasing the sample size and putting much tighter constraints on the population obliquity distribution and shedding more light on planet formation mechanisms.

6

Inferring the Distribution of Stellar Inclination Angles

Stellar inclination angles, that is the angle between our line-of-sight and the stellar rotation axis, have long been a highly sought after quantity in many fields, such as in the study of exoplanet systems (e.g. Chaplin et al. 2013; Hirano et al. 2014; Campante et al. 2016a), eclipsing binaries (e.g. Albrecht et al. 2013a, 2014) or clusters (e.g. Jackson & Jeffries 2010; Corsaro et al. 2017). Previous work has been made to try and infer the distribution of inclination angles predominantly using spectroscopic quantities combined with stellar parameters. For example the most common way is to use knowledge of the radius of the star $R_\star$ the star's rotation period $P_{\rm rot}$ and projected rotational velocity $v \sin i$ (e.g. Abt 2001). This is done by assuming that the star's rotational velocity is well approximated according to

$$v = \frac{2\pi R_\star}{P_{\rm rot}}, \tag{6.1}$$

from which the angle of inclination can be calculated

$$i = \arcsin\left(\frac{v \sin i}{v}\right). \tag{6.2}$$

However rather than calculating the angle itself using this method, the sine of the angle is normally calculated instead which can lead to values of $\sin i$ greater than 1.

Morton & Winn (2014) also presented some work on how to simulate the distribution of $\cos i$ given the same quantities making use of a Bayesian framework and calculating the likelihood of $\cos i$ given a certain set of relevant inputs (which has been explained in Chapter 5).

Both of the above methods however are indirect and provide reasonable estimates when there are no alternatives. Asteroseismology gives an opportunity to extract the angle of inclination from the non-radial modes of oscillation in the presence of visible rotational splitting. This is especially useful for the faint stars observed by *Kepler* for which no reliable $v \sin i$ or $P_{\rm rot}$ could be extracted. The way in which we provide the estimate of the angle of inclination was described in Chapter 5, and so here we shall proceed to describe how we put all the data together and performed the relevant analyses.

As part of their analysis of the clusters NGC 6819 and NGC 6791 Corsaro et al. (2017) derived the inclination angles of a small sample of field stars (30 in total) as a control sample to check that an isotropic distribution was obtained for a set of random stars. Although they said this was achieved we will highlight in this chapter some of the pitfalls of deriving the angle of inclination through asteroseismology and also the need to interpret the results with care.

6.1 Overview of methodology

The initial sample consisted of 121 red giant stars observed by *Kepler* (taken from the sample of 13,000 red giants in Stello et al. 2013) that possessed a $\nu_{\rm max}$ in the range of $231 - 269$ μHz and had 4 years of observations in long-cadence ($\Delta t = 40$s). No stars that had a $\nu_{\rm max}$ value above (or close to) the long cadence Nyquist frequency, given by 283.4 μHz, were used in this analysis due to the added complications of Nyquist aliases (e.g. Murphy et al. 2013; Chaplin et al. 2014b). All of the photometric timeseries were reduced using the pipeline developed by Jenkins et al. (2010b) and power spectra were obtained using the Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982). The sample size was reduced to 90 due to a few either showing suppressed $\ell = 1$ mixed modes (Mosser et al., 2012a; García et al., 2014; Fuller et al., 2015; Mosser et al., 2017) or highly complex spectra from which the relevant modes could not be successfully disentangled, possibly due to the effects of rotation. The effect of the reduced sample size is also discussed later by applying the same method to artificial power spectra with a known inclination angle distribution.

The extraction of the inclination angle has been covered already in this work and uses the same method discussed in Chapters 7 and 5 and so will not be discussed further here.

6.2 Mixed Mode Identification

The initial identification of the radial modes was performed using the universal pattern described by Mosser et al. (2011) which is dependent only on the large frequency separation, $\Delta\nu$ (the separation in frequency between modes of the same degree and subsequent radial order). This was necessary to make sure that the initial placement of our window (in frequency) in the power spectrum was correct and aid the mode identification in lower signal-to-noise stars. In this work we are interested in fitting individual $\ell = 1$ mixed modes rather than performing a global fit, the reason being that there are many mixed modes per order and with the

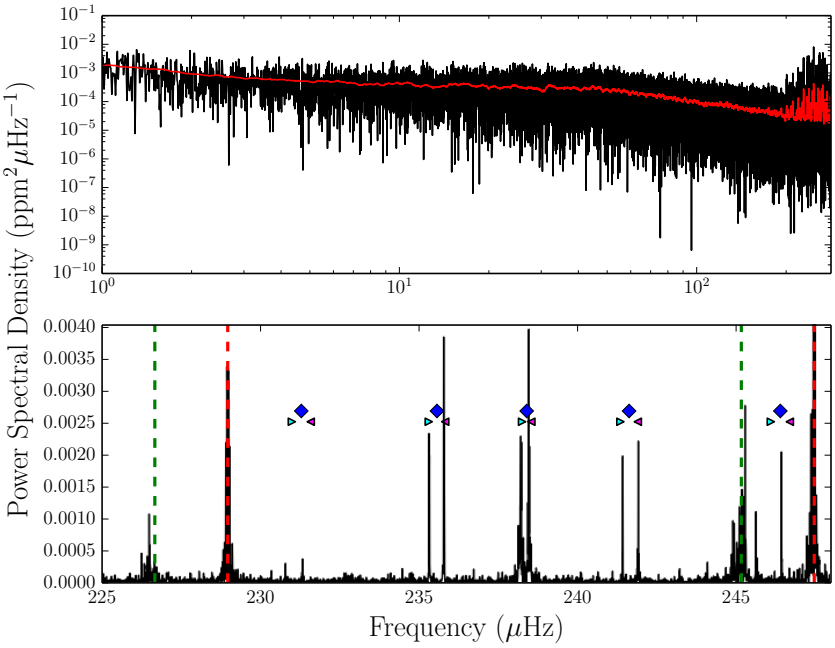


Figure 6.1: The top panel shows a log-log plot of the power spectrum for KIC 5553307 in black and a subsequent smoothed spectrum (1 μHz boxcar) in red. The bottom panel shows a single radial order of the above spectrum with key features overplotted. The red dotted lines show the position of the radial modes identified using the red giant universal oscillation pattern (Mosser et al., 2011). The green dotted lines show the position of the quadrupole ($\ell = 2$) modes which were tuned by eye using the small frequency separation d_{02} (which is the separation in frequency between a radial mode of order n and a quadrupole mode of order $n - 1$). The blue diamonds give the predicted positions of the $\ell = 1$ mixed modes using equation 6.3 for a period spacing $\Delta\Pi_1 = 87.6$ s and coupling $q = 0.2$. The cyan and magenta triangles denote the predicted positions of the rotationally split components ($m = -1$ and $m = +1$ respectively) using equation 6.4 with $\nu_{s,\max} = 0.3$ μHz.

high quality *Kepler* data this enables us to use many measures of the angle of inclination as opposed to fitting one overall angle per order (e.g. Davies & Miglio 2016). An example spectrum is shown in figure 6.1 including details of how the methodology works, along with examples of rotational splittings.

When considering these $\ell = 1$ mixed modes we follow Mosser et al. (2012b) and so the mixed mode frequencies can be described according to

$$\nu = \nu_{n_p,\ell=1} + \frac{\Delta\nu}{\pi} \arctan\left[q \tan\pi\left(\frac{1}{\Delta\Pi_1\nu} - \varepsilon_g\right)\right], \tag{6.3}$$

where $\nu_{n_p,\ell=1}$ is the nominal p-mode frequency, $\Delta\nu$ is the large frequency separation, q is the coupling between the p- and g-modes, $\Delta\Pi_1$ is the $\ell = 1$ period spacing and ε_g is a phase term. For simplicity we assume that ε_g is 0, in accordance with and Mosser et al. (2012b). This was combined with the formulation for the rotational splittings of $\ell = 1$ mixed modes as described in Goupil et al. (2013)

$$\frac{\nu_{\rm s}}{\nu_{\rm s,max}} = \zeta \left(1 - 2\frac{\langle\Omega\rangle_{\rm env}}{\langle\Omega\rangle_{\rm core}}\right) + 2\frac{\langle\Omega\rangle_{\rm env}}{\langle\Omega\rangle_{\rm core}}, \tag{6.4}$$

where $\nu_{\rm s}$ is the rotational splittings, $\nu_{\rm s,max}$ is the maximum splitting, ζ is the mode inertia, $\langle\Omega\rangle_{\rm core}$ is the angular rotational velocity averaged over the core regions and $\langle\Omega\rangle_{\rm env}$ is the angular rotational velocity averaged over the envelope.

The above expressions were combined into the form of an interactive environment (GUI) (Bossini, 2015) whereby each of the parameters above could be adjusted and the predicted frequencies updated in real time. This initial work was further adapted by myself to work out any bugs present but also extend the code slightly to give a means of detecting and outputting the requisite mixed modes.

In the event that $\nu_s \geq 1/\Delta\Pi_1$, the components of different multiplets can overlap creating a complex spectrum. Stars for which this is the case have most likely not been included since identifying modes with consistent parameters would be very difficult.

6.3 Mode fitting

The method used to identify the presence of the mixed modes is given in Appendix A.12. Having identified the presence of a $\ell = 1$ mixed mode we have the opportunity to extract its underlying parameters and, like the majority of the analysis performed in this thesis, this was done in a Bayesian manner (using MCMC to sample the posterior distributions of the parameters).

The framework used is identical to that given in Chapter 3 in terms of the likelihood

function used, the mixed modes are fitted individually with fitting regions selected by the user. The model fitted for each mode is a triple Lorentzian incorporating the angle of inclination (e.g. Handberg & Campante 2011)

$$\mathcal{M}(\nu; A, \nu_0, \Gamma, \nu_{\rm s}, i) = \sum_{m=-\ell}^{\ell} \frac{2\mathcal{E}_{1m}(i)A^2}{\pi\Gamma\left[1 + \frac{4}{\Gamma^2}(\nu - \nu_0 - m\nu_s)^2\right]} + B, \tag{6.5}$$

where A is the mode amplitude, Γ is the mode linewidth, $\mathcal{E}$ is the relative mode visibility, nu_0 is the central frequency of the mode, ν_s is the rotational splitting and B is the background. The background is assumed to be constant over the small range in which the mode is fitted in order to simplify the model which is a reasonable assumption to make. It has been noted that asymmetry in the rotational splittings have been observed in red giants (e.g. Di Mauro et al. 2016), however this has not been included in the model. Asymmetry will be most likely to affect stars of intermediate angle rather than those close to 90 degrees (due to no central component being visible). The effects of asymmetry, if it is a considerable issue, will be made obvious in the final distribution since the it should deviate from the expected values, but not at 90 degrees.

The varying linewidths of the mixed modes due to their varying inertia can cause problems for any fitting procedure and so the initial guesses used to position the walkers in the MCMC needed to be fairly accurate. The reason for this is that the posterior PDF for the frequency of the mode can be multi-modal, containing two or three peaks for any mode with an angle of inclination above $\sim$20 degrees. As a result if the initial guess for the central frequency is closer to the outer components than the central one, the fitting procedure can mistakenly identify the central frequency as that of the $m = \pm 1$ component instead of the $m = 0$. This problem is exacerbated by the diminishing linewidth of the g-dominated modes where the comparative distance in parameter space between the frequency of the mode components is much larger (as a result of the smaller linewidths) and therefore makes the peaks in the central frequency posterior PDF much more marked.

The use of MCMC gives the posterior distribution for each parameter. This is particularly useful in the context of the inclination angle. Since the modes are fitted individually there will be inclination angle PDFs for each mode of the star and this needs to be combined in order to produce a posterior PDF for the entire star. The simplest method was used for this work whereby the PDFs for the modes were concatenated to give the posterior PDF for the entire star. This will usually result in a larger uncertainty in the final angle than if all the modes were fitted at the same time. When all the modes are fitted together, the angle of inclination inferred is the angle that is most consistent with all modes, which will reduce region of parameter space explored. Whereas if all modes are fitted individually there is no knowledge of the other modes , but it gives the opportunity to explore any biases that may be present in more detail. A few example of the fits made to the mixed modes are shown in Fig 6.2 for an intermediate and low angle of inclination star.

An example of the posterior PDFs for KIC 3531478 are shown in Fig 6.3 where the difference between the posterior distributions for the individual modes and the concatenated distribution can be seen. In the event of the majority of the modes being consistent with a small range of angles the concatenated PDF should possess the mean value of the distributions and reduced uncertainties, as shown in Fig 6.3. This works well for cases at intermediate and high angles of inclination, but when the inclination angle is much lower problems can occur as a result of the prior distribution. This is shown in the right hand panel of Fig 6.3 for KIC 8645227 whereby the angle of inclination is clearly low but the influence of the prior is apparent due to the non-negligible mass of probability at higher angles. This is caused by low or poorly defined splittings, which make it difficult to distinguish between a mode with low splitting from one with no splitting and a low angle. As a result, the fitting procedure will always result in larger errors for lower angles than higher angles (where the prior works in its favour) and the non-negligible component of the PDF at higher angles for stars determined to be low angles will also bump up the PDF at 90° in the final distribution.

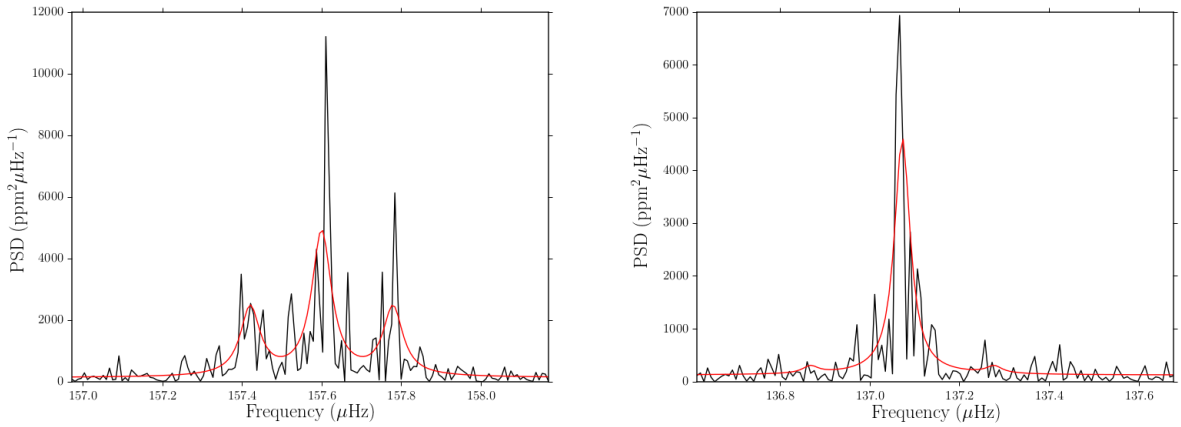


Figure 6.2: The left hand panel shows a fit to an mixed mode in the intermediate angle star KIC 3973247, the data is shown in black and the best fit in red. The right hand panel shows the same but this time for a mixed mode in the low angle star KIC 3645589.

The final concatenated distribution for the entire sample of stars is shown in Fig 6.4 and it can clearly be seen that there is an excess at 90° that should not be expected. This is puzzling, however previous discussions regarding the priors need to be accounted for and this can be done within a Bayesian Hierarchical framework very similar to that described in Chapter 5 in order to deconvolve the effects of the prior.

There could be an issue of misidentifying the modes an in fact confusing a high angle with a small splitting and a very low angle star. Each fitted mode was identified to check for consistency throughout the star, in addition the number of stars seen with a very low angle is close to the expected number for our sample size. Therefore it is unlikely for this to be the case.

6.4 Hierarchical modelling

In order to infer the angle of inclination distribution for the population of stars analysed we must once again turn to Bayesian Hierarchical modelling. The method used closely resembles that used in Chapter 5 and is based upon that described in Hogg et al. (2010), however this time a slightly different model will be used in order to infer whether or not the distribution

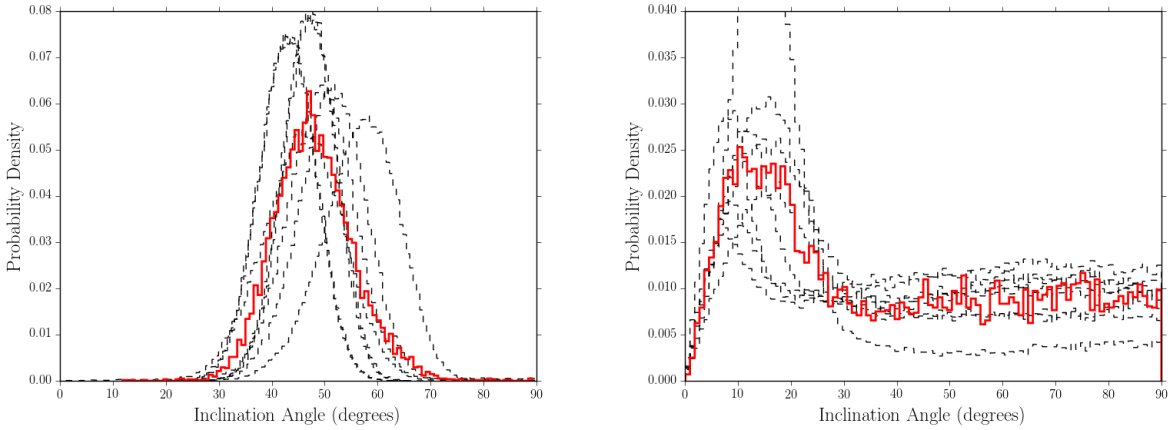


Figure 6.3: The left hand panel shows the individual posterior PDFs in angle for KIC 3531478 (in black) and the concatenated posterior for the star (in red). The right hand panel shows the same thing except this time for the star KIC 8645227.

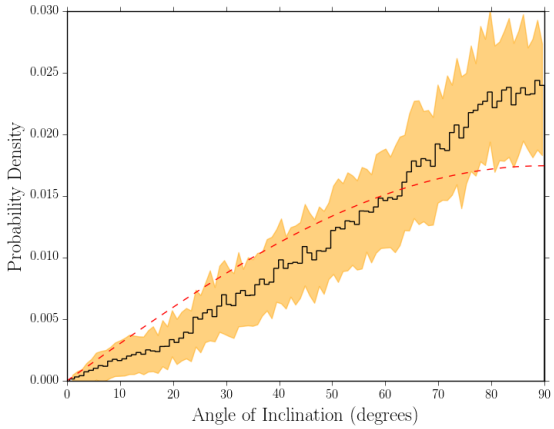


Figure 6.4: The final concatenated distribution in angle for all the stars in the sample, the isotropic distribution is shown as the red dotted line. The shaded orange region is the 1-σ bounds bootstrapped estimate.

is isotropic.

6.4.1 Original model

Whilst the objective of Chapter 5 was to infer the obliquity distribution given a distribution of inclination angles we instead desire the distribution of inclination models. The original model used to describe this distribution was the Fisher distribution as given in Morton & Winn (2014), however this does not give a great deal of flexibility. Instead the model used here is a variant of the Fisher distribution which has a location parameter, this accounts for the distribution being peaked at an angle that is not 90 degrees. In other words rather than just stipulating as to whether the observed distribution is or is not consistent with isotropy we would also want to be able to quantify by how much.

The Fisher distribution with no location parameters is defined as follows

$$f_\theta(\theta|\kappa) = \frac{\kappa}{2\sinh\kappa}\exp\left(\kappa\cos\theta\right)\sin\theta\ , \tag{6.6}$$

where κ is the concentration parameter and θ is the angle of inclination. It is important to note that Morton & Winn (2014) did not consider the regime $\kappa < 0$ because it was not needed for their analysis. Whilst $\kappa \to \infty$ results in an excess at zero degrees, $\kappa \to -\infty$ results in an excess at 180 degrees which gives a bit more flexibility to the model.

The form of the prior used can influence the parameter space in which we perform the inference. For example, the isotropic distribution, which constitutes the uninformative prior for the angle of inclination, which is flat in $\cos i_\mathrm{s}$ is a much more sensible parameterisation in which to perform the inference than i_s where the prior is of the form $p(i_\mathrm{s}) \propto \sin i_\mathrm{s}$. A more practical reason for performing the inference in $\cos i_\mathrm{s}$ is that the isotropic prior in i_s tends to zero as $i_\mathrm{s} \to 0$ and so as this quantity is in the denominator of the likelihood function it would introduce numerical instability unnecessarily into the method.

To transform equation 6.6 we use the following equation (as defined in Chapter 5)

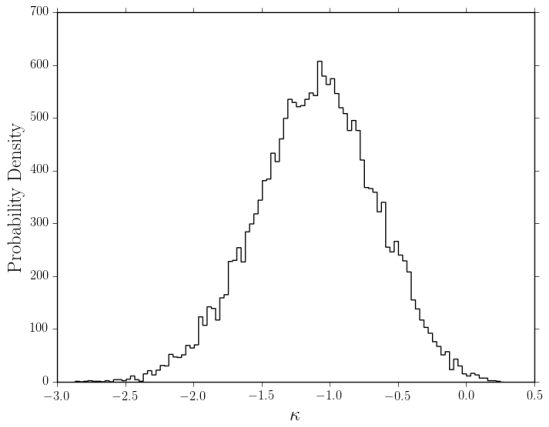


Figure 6.5: Posterior PDF for the concentration parameter κ used in the first incarnation of the model.

$$f_Y(y) = \sum_{k=1}^{n(y)} \left| \frac{\mathrm{d}}{\mathrm{d}y} g_k^{-1}(y) \right| f_X \left[g_k^{-1}(y) \right] \ , \tag{6.7}$$

where $y = g(x)$ and the summation is made over the number of solutions to $g_k^{-1}(y)$. Inserting $g(x) = \cos x$ into the above equation gives

$$f_{\cos i_\mathrm{s}}(\cos i_\mathrm{s}|\kappa) = \frac{\kappa}{\exp \kappa - 1} \exp \left(\kappa \cos i_\mathrm{s} \right) . \tag{6.8}$$

As a result the corresponding likelihood function (from Chapter 5) can be reduced to

$$\mathcal{L}_\kappa \approx \prod_{n=1}^{N} \frac{1}{K} \sum_{k=1}^{K} f_{\cos i_\mathrm{s}}(c_{nk}) \ , \tag{6.9}$$

since the isotropic prior is flat in $\cos i_\mathrm{s}$.

The choice of prior on κ was taken to be the same as that given in Fabrycky & Winn (2009) where $p(\kappa) \propto (1 + \kappa^2)^{-3/4}$. In order to sample the parameter space MCMC was used once again and supplied the posterior PDFs from which the rest of the inference could be made.

The concentration parameter inferred from this simple model is given by $\kappa = -1.09^{+0.45}_{-0.44}$ with the posterior PDF shown in Fig 6.5. This shows that the data is bunched towards 90°.

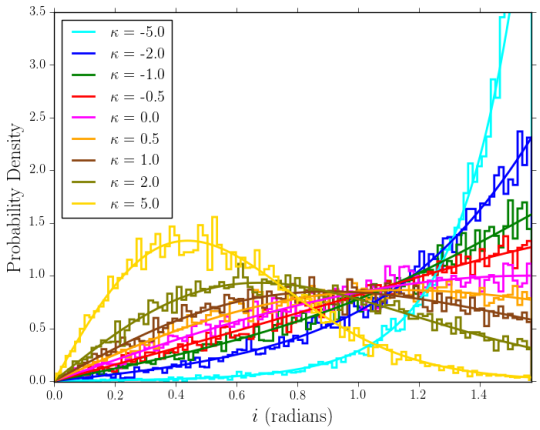


Figure 6.6: Example distributions of the Fisher distribution including negative κ values in i_s. The solid lines show the analytical expression whilst the histograms shows the result of Monte Carlo simulations.

The result is consistent with isotropy only at the 2.5σ level which is certainly a bit worrying, however this is due to the form of the model when κ is negative. This is illustrated in Fig 6.6 and shows that the Fisher distribution with negative κ does not accurately reproduce the shape of the data at high angles and so a better model is needed that can better reproduce this excess at 90°. The behaviour of the Fisher distribution for negative κ is due to the truncation of model at 90 degrees, since should normally be defined up to 180 degrees. As a result we would prefer a model that is not affected by truncation and also well reproduces the observed data.

6.4.2 Updating the Fisher model

The original formulation of the Fisher distribution is a valid model if we assume our angle of inclination distribution is, or very close to, isotropic. In order to add more flexibility into the model we add a location parameter that enables the peak of the distribution to be shifted in the region $\mu \in [0, \pi/2]$. This means that κ no longer has to go negative into a regime where, although it better fits the data, it does not properly represent the data in terms of shape. As a result the form of the model is updated to give

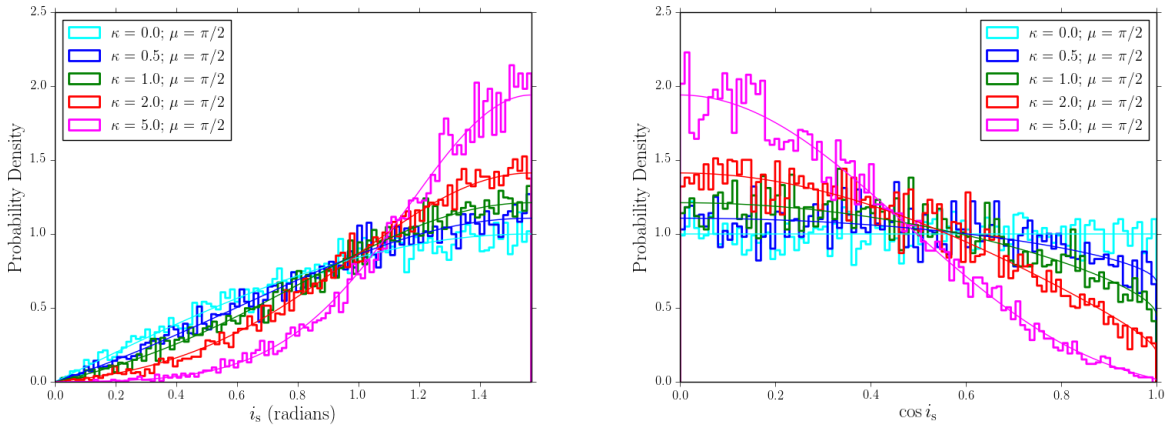


Figure 6.7: Example distributions of the updated model for different values of κ in both i_s and $\cos i_s$. The solid lines show the analytical expression whilst the histograms shows the result of Monte Carlo simulations.

$$f_{i_s}(i_s|\mu,\kappa) = \frac{\kappa}{2\sinh\kappa}\exp\left(\kappa\cos\left(i_s-\mu\right)\right)\sin i_s \ , \tag{6.10}$$

where the additional parameter μ is the aforementioned location parameter. Examples of the distribution are shown in Fig 6.7.

Following the method given above for transforming the original Fisher Distribution we can transform this updated model using equation 6.7 and the fact again that $g(x) = \cos x$ to derive the following unnormalised probability distribution

$$f_Y(y|\mu,\kappa) = \frac{\kappa}{2\sinh\kappa}\exp\left(\kappa y\cos\mu + \kappa\sqrt{1-y^2}\sin\mu\right) \tag{6.11}$$

where $y = \cos i_s$. It is important to again stress that this distribution is unnormalised and so numerical integration is needed to ensure proper normalisation.

This can again be plugged into the hierarchical method to see how it describes the data. For an isotropic distribution the location parameter would be expected to lie at $\sim 57.2°$ and so this gives an expected value to make inferences from.

The results from the hierarchical model can be seen in Fig 6.8 and the final parameters were found to be $\kappa = 2.09^{+1.45}_{-1.32}$ and $\mu = 1.37^{+0.20}_{-0.95}$. This time the model is consistent with

isotropy within 1.6σ which is a significant improvement over the simpler model, however this is due to the increased parameter uncertainties.

Clearly it would be a good idea to compare the models to see which best represents the data being analysed and help provide some context for the results. This is done using the model evidence, performed by integrating the likelihood over all of parameter space, which in this case was performed using brute-force integration due to there only being 1 and 2 dimensional problems. As a result the odds ratio between the two models can be computed

$$O_{21} = \frac{p(D|M_2)}{p(D|M_1)} \frac{p(M_2)}{p(M_1)}, \tag{6.12}$$

where the second half of the fraction on the right hand side is the prior odds ratio, in other words the prior knowledge that one model is better than the other. It is common place to set this equal to 1 if both models are equally likely and since we do not know which should be favoured this was also set to 1. Therefore only the Bayes factor is left which was 26.25 for the simple Fisher distribution (M_1) and 15.29 for the model with the location parameter (M_2). The Bayes factor is given by the ratio of these quantities and is ~ 0.58 in favour of M_2 and so is 1.7 in favour of M_1. This means that the simple model is favoured but the value of the Bayes factor is not worth more than a mention according to the summary given in Kass & Raftery (1995). Since the use of negative κ is not particularly useful in the sense that the distribution does not maintain its shape, it simply bumps up against 90°, it would be worth replacing the evidence of M_1 with the evidence calculated for only positive κ values, therefore making a fairer comparison. The Bayes factor in this case is 109 which is strong evidence in favour of the model with the location parameter. This due to the inflexible nature of the Fisher distribution for distributions that peak towards 90°.

Due to the favouring of the more complex model this will be adopted in the event of any more applications of this analysis being needed.

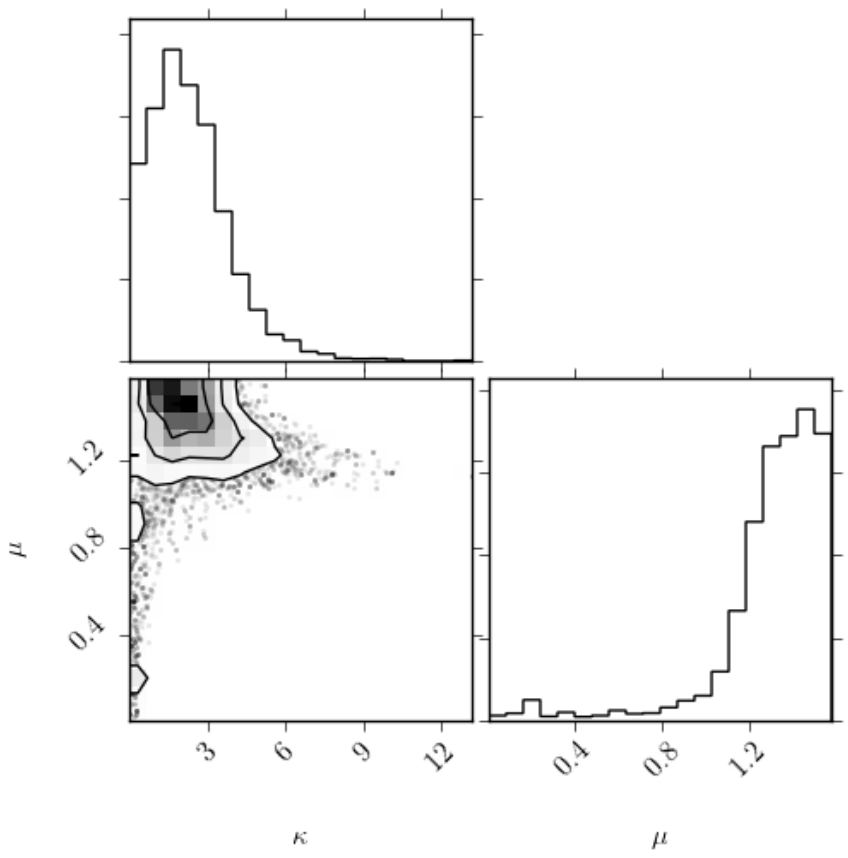


Figure 6.8: A triangle plot from the hierarchical analysis using the updated model.

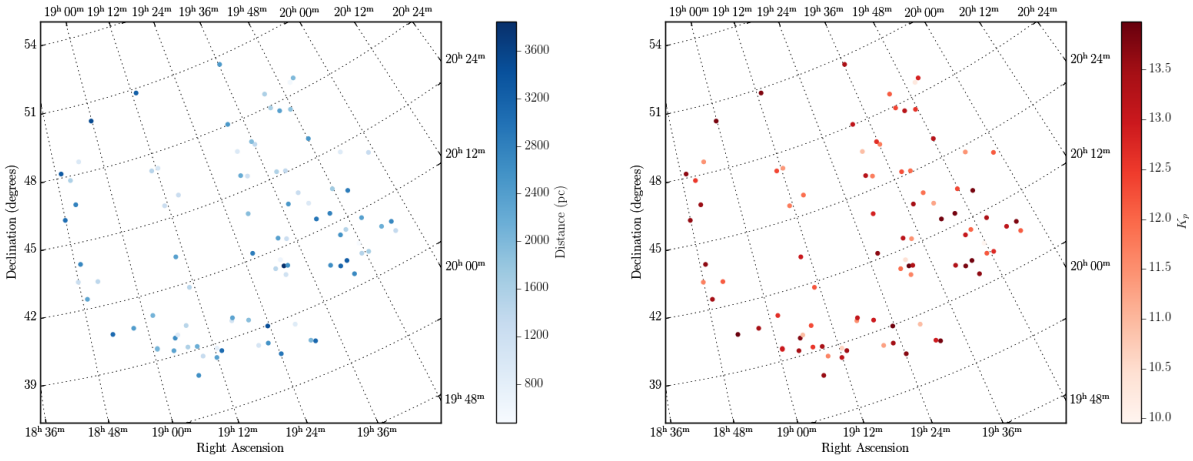


Figure 6.9: Declination plotted against right ascension for the stars in the sample, where the points are coloured according to their distance (calculated using asteroseismic scaling relations) in the left-hand plot and by *Kepler* magnitude in the right-hand plot.

6.5 Issues with the data

The hierarchical method has shown that the distribution is consistent with isotropy at the 1.6-σ level for the Fisher distribution with added location parameter through the application of model comparison. However there is still an observed excess at 90° that is causing the concentration parameter to have probability mass above zero. It is therefore necessary to check if there are any biases present in our sample, this could be as a result of, say, the position of the stars in the sky, their *Kepler* magnitude, or stellar/oscillation mode properties.

6.5.1 Location of sample

Before anything is done to the original datasets the positions of the stars in the *Kepler* field must be checked along with their distances for any "clumping". By this we mean to check whether there are stars in similar positions or with similar distances that may possess similar inclination angles, as this would clearly push our inclination angle distribution away from an isotropic distribution. The position on the sky was taken from the revised KIC (Huber et al., 2014) along with the star's K_p.

By looking at the distribution of stellar positions given in Figure 6.9 it can be seen that

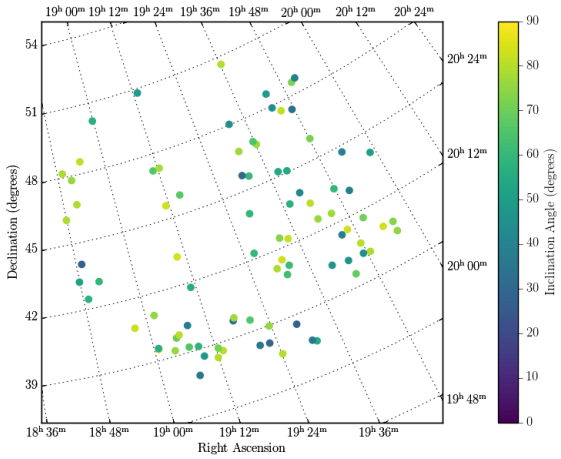


Figure 6.10: Declination plotted against right ascension for the stars in the sample, where the points are coloured according to their average inclination angle as fitted in this work.

there is no obvious clumping of stars and so it is unlikely that there is a clump of similarly aligned stars in the sample being analysed. This same plot can be created but instead colouring by the extracted inclination angle, as given in Fig 6.10. The findings are still the same, there is no obvious clumping and so it can be assumed that the positions of the stars in the *Kepler* field are not causing any deviations from isotropy.

6.5.2 Magnitude-limited sample

Having seen that the initial sample gives a slight excess of stars at 90° the reasons behind such deviations could lie in the properties of the stars involved. First of all we shall split the sample up according to their *Kepler* magnitude (K_p) to see if the white noise level (which is correlated with the apparent magnitude) affects the determination of the inclination angle. The sample was therefore split up into a bright and faint dataset about the median K_p value of 12.9. The resultant datasets contained 45 stars each and the hierarchical inference was then performed on each set and the combined sample to check for any inconsistencies. The two datasets should then equate to better (lower magnitude) and worse (higher magnitude) signal-to-noise sets.

The best way in which to compare these two samples is to simply plot the concatenated

posteriors for each, as shown in Fig 6.11, because this gives a more direct way in which to compare them. The impact of signal-to-noise on the sample is interesting since there are small changes in the structure as a function of angle. Certainly there are differences at low angles where for the brighter stars lower angles can be achieved. However for the fainter stars there is a peak at $\sim 30^\circ$ which demonstrates the fact that the minimum angle is limited by the signal-to-noise ratio. This idea is explained in more detail in section 6.6.1

In addition there is also a slight difference at very high angles as well where the fainter stars peak at around 80° degrees as opposed to 90° for the brighter sample. This can also be thought of as the signal-to-noise limit the highest angle that can be achieved as well, since the higher the background noise the higher the amplitude of the inner component of the mode could be thereby pushing the inclination angle lower than it maybe should be. This idea is covered later in a bit more detail showing limiting inclination angles for given signal-to-noise.

It is also possible that this new structure is a result of the uncertainties on the posteriors and due to the smaller sample size as a result of splitting the initial data. The differences, whilst being present, are also not significant enough to be convincing as a cause for the observed excesses in the concatenated inclination angle distribution.

6.5.3 Reduced-splitting sample

In addition to investigating the effect of signal-to-noise, the nature of the mode and its relationship to the angle of inclination of the star, can also be investigated. The data was split according to what we have called the "reduced splitting" by which we mean the ratio of the mode linewidth to the rotational splittings. This will separate the data out into the more p-dominated mixed modes which will possess a larger ratio and the g-dominated mixed modes which will have a smaller ratio. This is an important parameter in the determination of the angle of inclination because the rotational splitting is key in its determination. Therefore if a mode as a large reduced splitting the mode linewidth is much larger than the rotational

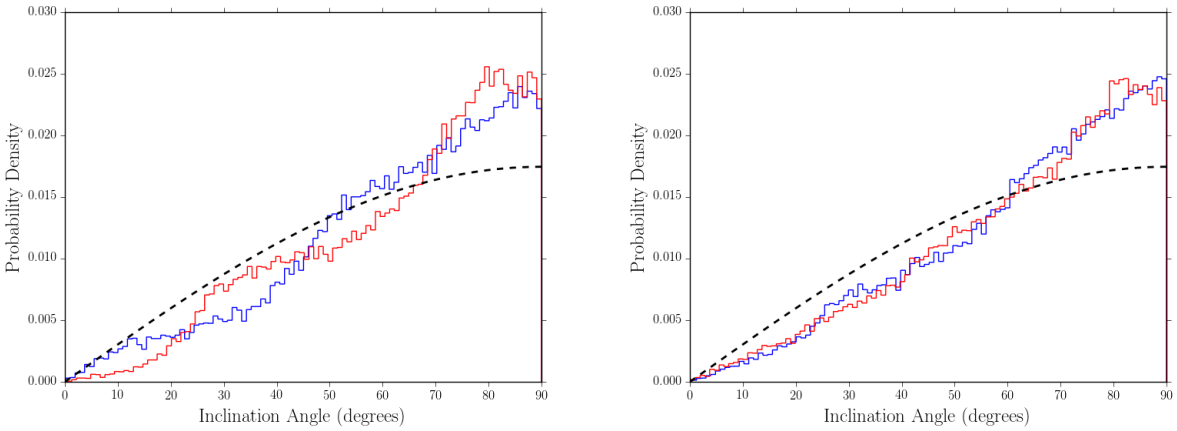


Figure 6.11: Concatenated posterior distributions for the real stars with the sample split according to the *Kepler* magnitude (on the left-hand side) and reduced-splitting (right-hand side). The red shows the distribution for stars above the respective median value and the blue for those below. The isotropic distribution is overplotted in black.

splitting and so the angle of inclination would be harder to extracted since the signature of rotation could be absorbed by the enhanced mode linewidth. This is contrary to the case of a small reduced splitting where the angle of inclination determination would be much easier as the mode components would be well separated.

In the same way the magnitude-limited samples were split, these new datasets were created by splitting the data according to the median reduced-splitting of the entire sample (thereby ensuring the sample sizes are the same). This now means that each mode is being treated individually rather than as part of a star, therefore the interpretation of the result has to be carefully considered due to the nature of the hierarchical model. But this can be fixed by creating posteriors for each star that contain only the p- or g-dominated mixed modes and that is what has been done here. The division in reduced splitting was made at 0.195 according the median as previously described.

The two posteriors can be seen in the right-hand panel of Fig 6.11 with the more p-dominated modes being shown in red and the more g-dominated being shown in blue. In general here seems to be very little difference between the two samples indicating that the unresolved nature of the modes is not causing any biases in the context of the distribution

as a whole, but this could be down to larger uncertainties masking this issue. As a result the inclusion of unresolved modes cannot be discounted as a possible bias in the data.

6.6 Observed excess at 90° - investigating potential biases

Although the posteriors extracted from the sample of stars is consistent with isotropy, there is still the problem of why there is an excess close to $\sim 90°$. This needs to be investigated as there could be potential biases present in the data that get smeared out due to the large uncertainties present.

6.6.1 Limiting cases for angle of inclination

Gizon & Solanki (2003) stated that the angle of inclination could not be reliably determined below $i < 30°$ using their methodology. However in the era of *Kepler* this should be able to be pushed lower due to the high quality of the data. As an added measure it is important to work out the limiting cases of the angle of inclination determination since it affects both the low and high angle stars. The derivation of this is given in Appendix A.11.4.

The limiting cases are shown in Fig 6.12 where the upper limit is shown in red and the lower limit is shown in black. Clearly this conforms with the common sense view, that as the signal-to-noise increases the extremes of the distribution can be better retrieved. This idea can also be used to check the quality of the fits performed by making sure that, for the given H/B ratio the inclination angle does not go above or below the bounds.

6.6.2 Data set & method

In order to explore potential biases, modes of oscillation were simulated in the time domain following De Ridder et al. (2006) with the range of parameters given in Table 6.1. The aim of this was to see how both the signal-to-noise and change in mode linewidth affects the determination of the angle of inclination.

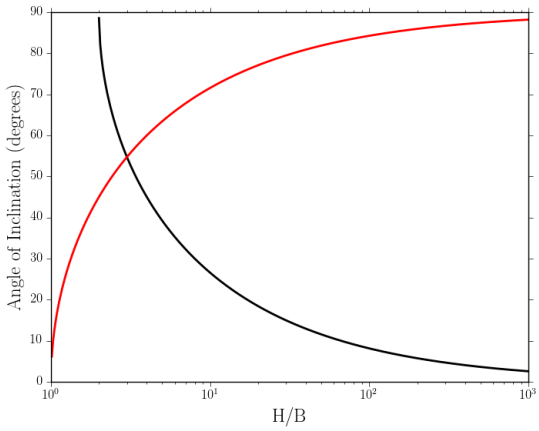


Figure 6.12: The minimum and maximum inclination angle achievable for a given signal to noise. The red can be interpreted as the maximum angle and the black as the minimum angle achievable.

Only the amplitude and linewidth were allowed to vary in order to improve the computational efficiency of the simulations, but also due to the fact that it is thought that these will have the largest effect on our extracted angle. There would be some argument to vary the rotational splitting as well due to the fact that it is through the rotationally split modes that we determine the angle, however we already know that rotational splittings are observed in all cases and so this is not a priority. Varying the mode linewidth will automatically take care of this, since this will provide a range of reduced splitting values in the dataset. The inclination angle was also fixed at 45° to represent a "worst case scenario" in terms of signal-to-noise due to the presence of the triplet. The background value was computed for $\sigma = 10$ppm and using the following equation

$$B = 2\sigma^2 \Delta t / 1 \times 10^6, \tag{6.13}$$

where B is the background in units of $\mathrm{ppm}^2 \mu\mathrm{Hz}^{-1}$ and Δt is the cadence of the observations (taken to be 29.4 min according to *Kepler* long-cadence observations).

In the case of both the amplitude and linewidth 6 values were chosen evenly-spaced between the lower and upper bounds. These provide a compromise in terms of time taken to run the

Table 6.1: Input parameters into time domain simulations, those with no upper bound denote fixed values over all simulations. ($\nu_{\rm bw}$ is the frequency bin width)

Parameter	Lower Bound	Upper Bound
A (ppm)	0.5	15
Γ (μHz)	$0.1\nu_{\rm bw}$	$20\nu_{\rm bw}$
Inclination Angle (degrees)	45.0	-
Rotational Splitting (μHz)	0.25	-
Background (ppm$^2\mu$Hz^{-1})	0.352	-

simulations and coverage of parameters space. The best way to display the coverage is in terms of the reduced-splitting (given in this work by Γ/ν_s, the ratio of the mode linewidth to its rotational splitting) and the height over background ratio as shown in figure 6.13. To also account for the unresolved nature of some of the modes the heights were calculated following Fletcher et al. (2006)

$$H = \frac{2A^2}{\pi\Gamma T + 2}, \tag{6.14}$$

where H is the mode height[1] and T is the length of the observation (which in this case is 4 years).

Following the creation of the data the method was very simple and consisted of fitting the mode (using MCMC) with our standard model (e.g. Handberg & Campante 2011).

$$\mathcal{M}(\nu) = \sum_{m=-1}^{m=+1} \frac{H\mathcal{E}_{\ell=1,m}(i)}{1 + \frac{4}{\Gamma^2}(\nu - m\nu_s)^2}, \tag{6.15}$$

where H is calculated using equation 6.14 and $\mathcal{E}$ is the mode visibility which is in turn dependent upon the inclination angle i.

[1]This is currently given in ppm^2, to convert to the correct units of ppm$^2\mu$Hz the height needs to be divided by the bin width.

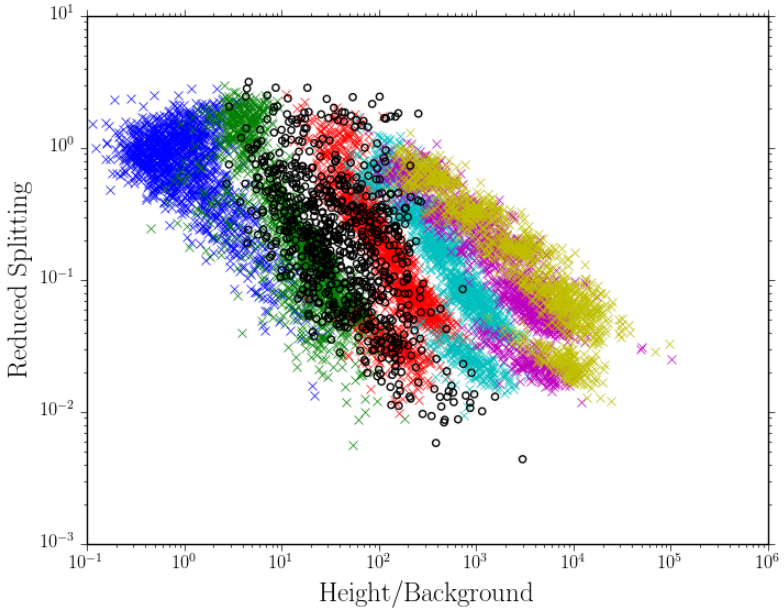


Figure 6.13: Height over background plotted as a function of the reduced splitting for both the simulated data (coloured x markers) and the real data (shown in black). The colours of the simulated data denotes the signal-to-noise ratio, ranging from 1.2 (blue) to 194 (in yellow).

6.6.3 Exploring the biases

By looking at figure 6.13 it can be seen that in general the two signal-to-noise ratios that best reproduce the real *Kepler* data are those corresponding to the amplitudes of 1.0 ppm and 2.5 ppm, which is a height over background (H/B) calculated from the data of approximately 10-1000. So we shall therefore restrict ourselves to looking at those relevant subsets of the parameter space.

Amplitude

The first parameter we shall consider is the amplitude of the mode and so from initial thinking it would seem reasonable that the more p-dominated modes will have better fitted amplitudes and the unresolved modes will have poorly determined amplitudes.

Figure 6.14 shows the fractional difference in amplitude as a function of H/B (coloured by input signal-to-noise) and it can be seen that in general the fractional difference is close

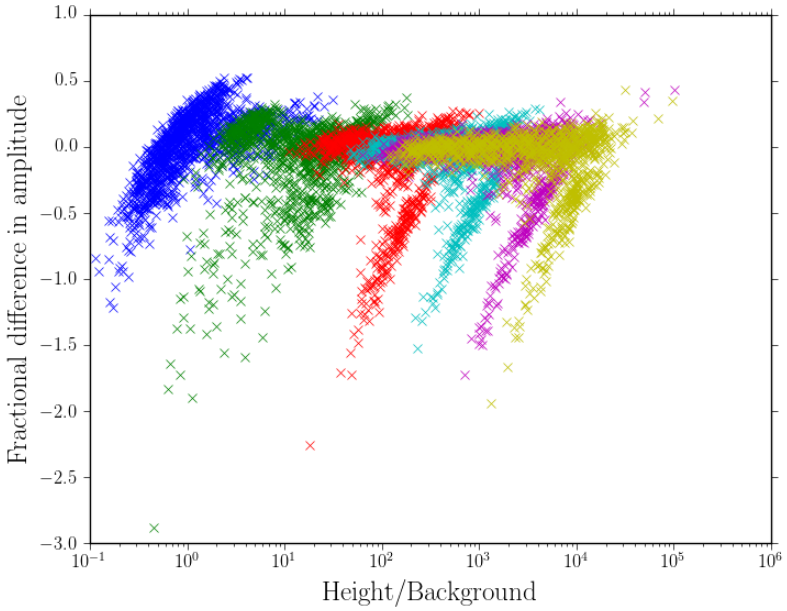


Figure 6.14: Fractional difference of fitted amplitude plotted as a function of H/B coloured by the signal-to-noise input for the simulations.

to zero with some scatter. However there are extended tails where the fitted amplitude had been greatly underestimated, and this corresponds to the unresolved modes ($\Gamma \lesssim 2\nu_{\rm bw}$). This comes about due to the fact that as the mode linewidth becomes smaller and tends towards the unresolved regime the standard Lorentzian profile is no longer a good fit to the data. Technically speaking the correct model (in all cases) is the Lorentzian profile convolved with the sinc-squared corresponding to the length of the timeseries and this would take into account the unresolved nature (i.e. if all the power in the Lorentzian is contained in on or two bins). Therefore it is no real surprise that the fits perform badly for unresolved modes.

Linewidth

Having seen the reasonable performance with regards to amplitudes for resolved modes, let us now have a look at the linewidth and how it is affected. Figure 6.16 gives a good insight as to why the amplitudes were underestimated for many of the modes. Instead of colouring by H/B the points have been coloured by linewidth to give an idea as to how each set performs.

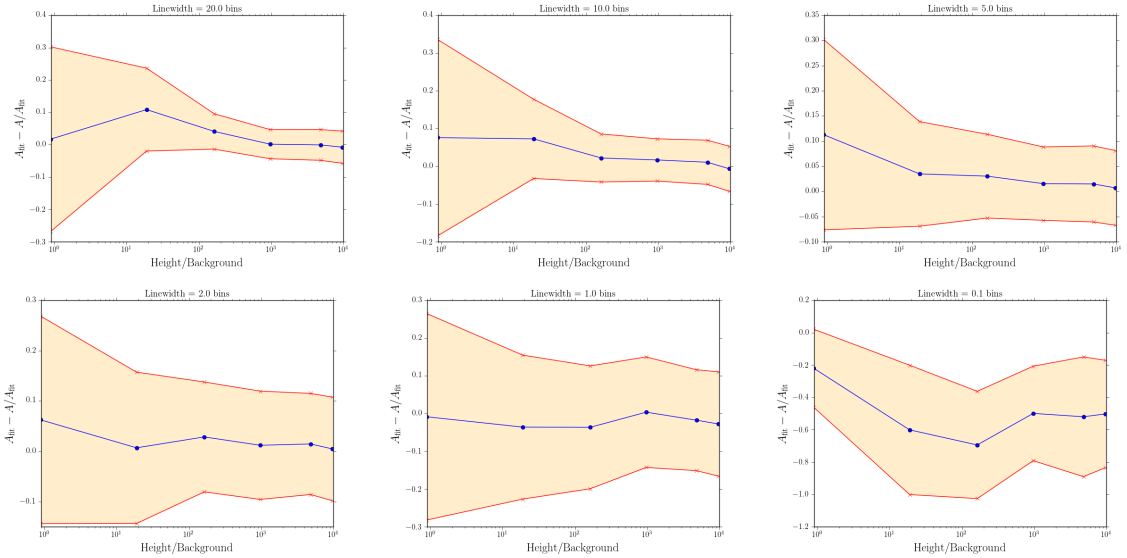


Figure 6.15: The fractional difference between the fitted amplitude and true amplitude for the simulations, where in each figure in the tile shows a different linewidth value used. The blue line denotes the mean value of the fractional difference averaged over the realisations and the red envelope shows the standard deviation of the distribution of fractional differences for each H/B value.

It can be clearly seen that for the lowest SNR set of simulations that the correct linewidth was never recovered, however as our data does not lie in this region it is of little concern. However we can see that for the highly unresolved modes the input linewidth is greatly overestimated and so this would correspond to an underestimation in the amplitudes. We can see that fitting to most unresolved modes is a very bad idea. It also appears as though the modes with $\Gamma = 1.0\ \nu_{\rm bw}$ tend to have slightly overestimated linewidths and so this would also make some degree of sense as we are fitting an unresolved mode with a profile that is only applicable to resolved modes.

The idea of underestimating the linewidth leads to an overestimation in the amplitude can be shown quite simply by firstly considering the amount of power contained in a Lorentzian profile

$$\mathcal{P}(\nu) = \int_{-\infty}^{\infty} \frac{2A^2}{\pi\Gamma\left[1 + \left(\frac{4}{\Gamma^2}\right)(\nu - \nu_0)^2\right]} \mathrm{d}\nu. \tag{6.16}$$

However we only utilise the positive frequency part of the power spectrum and so this can be rewritten

$$\mathcal{P}(\nu) = 2\int_0^\infty \frac{2A^2}{\pi\Gamma\left[1+\left(\frac{4}{\Gamma^2}\right)(\nu-\nu_0)^2\right]}\mathrm{d}\nu, \tag{6.17}$$

and this can be evaluated to give

$$\mathcal{P}(\nu) = \frac{2A^2}{\pi}\left[\arctan(\infty) - \arctan(0)\right] = A^2. \tag{6.18}$$

We have also used the parameterisation of the height, H, that can be derived by considering that the profile has a maximum when $\nu = \nu_0$ and so evaluating the Lorentzian at ν_0 gives a value for the height

$$H = \frac{2A^2}{\pi\Gamma}, \tag{6.19}$$

which follows on from our parameterisation, but it handy to show nonetheless. If we rearrange the above equation in terms of the total power in the profile we can see that the amplitude and linewidth are anti-correlated through the term involved H

$$A^2 = \frac{\pi H\Gamma}{2}, \tag{6.20}$$

in the sense that in order to conserve the total power in the profile if the amplitude is underestimated, then the linewidth must increase in order to conserve power. The difference between the input linewidths and extracted linewidths can be seen in Fig 6.16 and Fig 6.17. It is certainly clear that for both low signal-to-noise modes and modes that are unresolved it is not possible to problem extract the mode linewidth using our current model. This is important (as mentioned above) when extracting the angle of inclination and so there should be an imposed cut in signal-to-noise to ensure the fit to the mode is good.

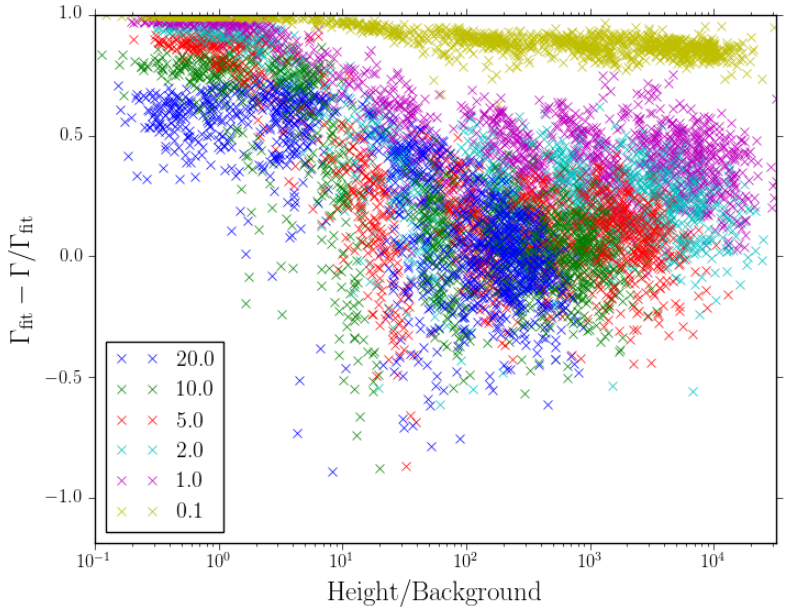


Figure 6.16: Fractional difference of fitted linewidth plotted as a function of H/B coloured by the linewidth input for the simulations.

An important argument when considering inclination angles is the idea of equipartition of energy, whereby we have observed the mode over enough lifetimes to see its average properties. For unresolved modes this clearly is not the case and so should not be considered for this work[2] This can be seen through the poor determination of both the amplitude and linewidth. However it would still be wise to see how the angle of inclination varies as a function of linewidth.

Inclination Angle

The quantity of interest from these simulations is the inclination angle and how it is affected by the change in linewidth and signal-to-noise. It can be seen in figure 6.18 that overall the angle of inclination can be well determined apart from the lower signal-to-noise cases. Interestingly enough, even though the expression for equipartition is not valid for the unresolved

[2]Less than 2% of modes in the sample of stars were unresolved, so even though unresolved modes must be discarded this does not limit our sample.

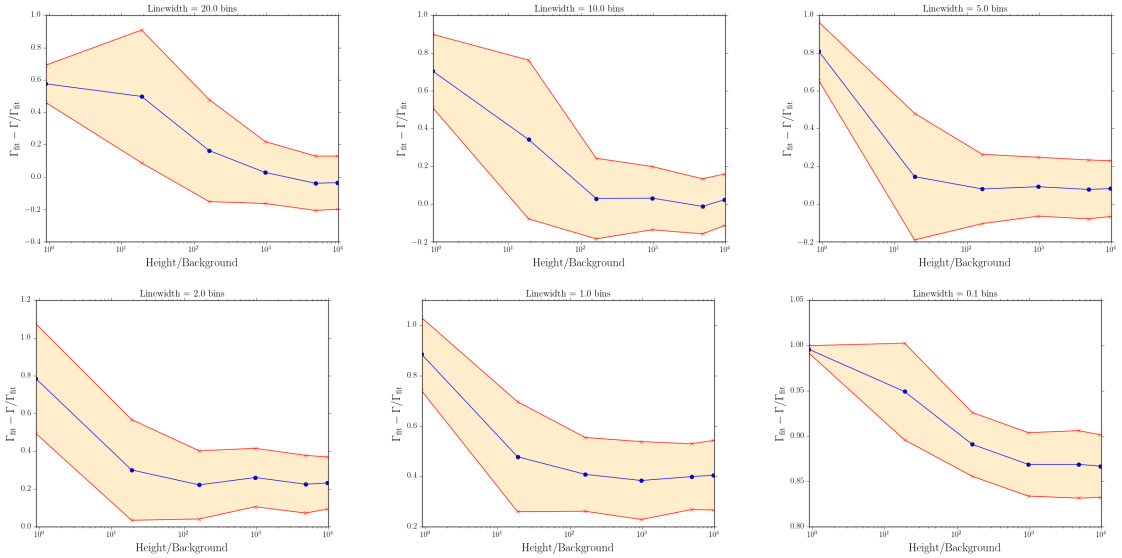


Figure 6.17: The fractional difference between the fitted linewidth and true linewidth for the simulations, where in each figure in the tile shows a different linewidth value used. The blue line denotes the mean value of the fractional difference averaged over the realisations and the red envelope shows the standard deviation of the distribution of fractional differences for each H/B value.

modes, it does not do a bad job of obtaining the correct angle.

When looking at figure 6.18, the uncertainties are of the order 20% and so hide any possible bias present in the data. As would be expected, when the H/B increases the determination of the angle also improves. We can therefore draw from this that the inclination angle can be determined well over the region of parameter space we operate in.

However, this is where we have to be very careful about how to interpret these results. Although the inclination angle can be well determined, even for the most unresolved modes, the mode linewidth clearly cannot be retrieved for the unresolved cases. This should be a red flag and indicates that the model being fitted to the g-dominated modes is not suitable and so the extracted angle of inclination should not be trusted. The model fitted to unresolved modes is a Lorentzian, but it should also be convolved with the sinc^2 corresponding to the frequency resolution, this should enable the correct extraction of the angle but was not carried out in this work.

It is also possible to make some approximate cuts in the parameter space of H/B and

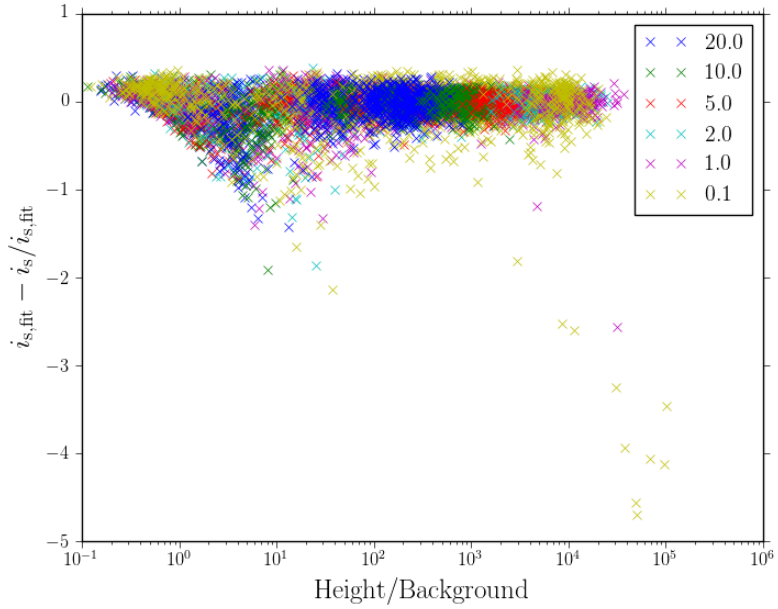


Figure 6.18: Fractional difference of fitted inclination angle plotted as a function of H/B coloured by the lienwidth input for the simulations.

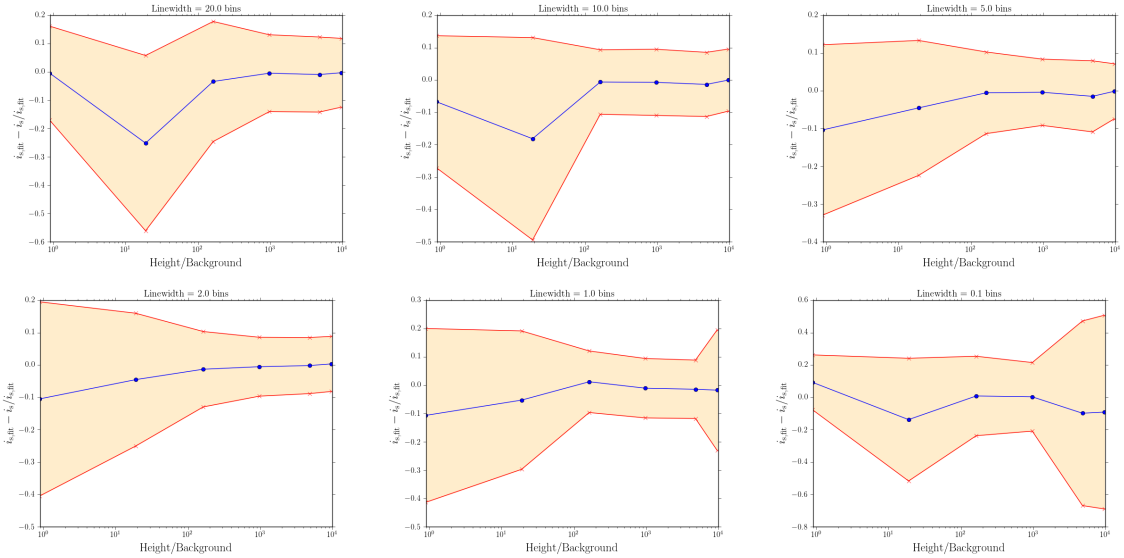


Figure 6.19: The fractional difference between the fitted inclination angle and true angle for the simulations, where in each figure in the tile shows a different linewidth value used. The blue line denotes the mean value of the fractional difference averaged over the realisations and the red envelope shows the standard deviation of the distribution of fractional differences for each H/B value.

linewidth in order to ensure that the properties of the modes are correctly extracted and therefore that the angle of inclination determined has as few biases as possible. From Fig 6.17 the linewidth cannot be extracted properly for the sample with linewidths below 5 bins (< 0.04 μHz) and so this can be used as a conservative lower limit in that regard. In addition, there is also a cut in H/B required since the linewidths cannot be extracted well for the lowest value chose, but it can be mostly for the second highest case. Therefore another cut can be made according to the mean H/B value of the modes created from the second lowest amplitude value which gives H/B$>$ 18.9. It is also worth mentioning that the reason the linewidth is harder to extract correctly for the largest two cases is likely due to the fitting window becoming too small and influencing the quality of the fit. However since the largest linewidth is not representative of our real data this is not too much of an issue.

6.7 Using artificial data

Although some cuts can be applied to the data from the informative nature of the individual fits it would be best to see if an overall distribution can be recovered. In order to do this, for each star in the sample, an artificial spectrum was generated (according to the method given in Chapter 4 with the same properties however the angle of inclination was instead drawn from an isotropic distribution. This will enable us to see if there is anything present in the real data that is causing any potential biases or in the method itself.

A comparison between the artificial dataset and the real data is shown in Fig 6.20. It is clear that the artificial data covers the correct region of parameter space that overlaps adequately with the real data. Therefore the conclusions drawn about the method from the artificial data can be applied to the real data reliably.

The artificial stars were analysed and peak-bagged in exactly the same way as the real set and the concatenated posterior for the sample is shown in Fig 6.21. The final posterior for the artificial data conforms very well with the isotropic distribution which then highlights

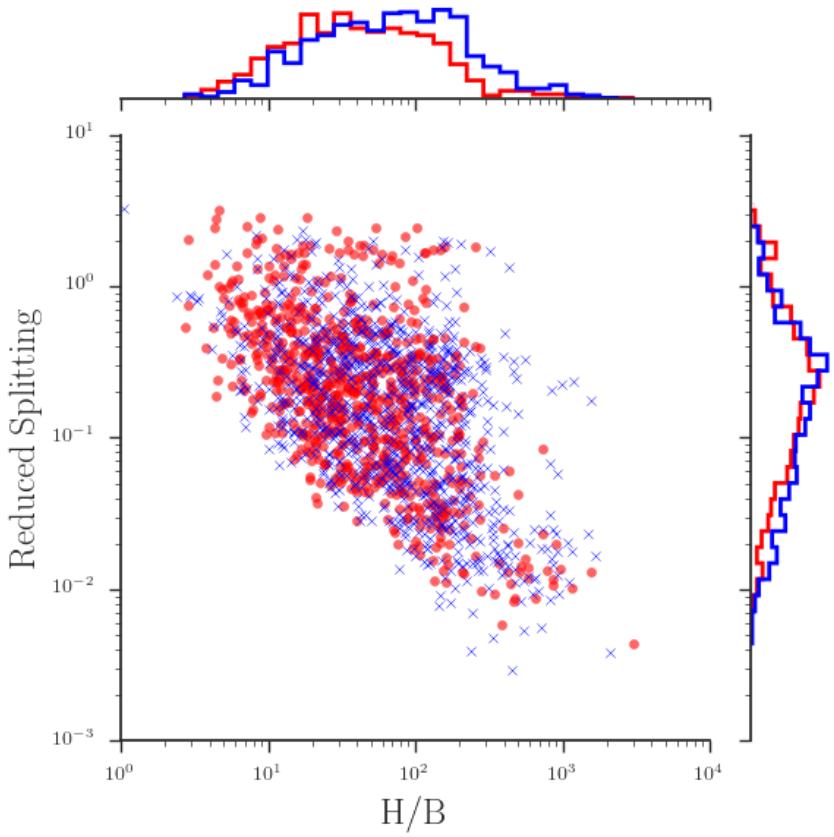


Figure 6.20: Reduced splitting as a function of the height over background ratio for the real data (in red) and artificial data (in blue).

where the observed issues might lie. There can be no issue with the method of extracting the angle of inclination itself, since if a bias was introduced through the method then this would cause deviations from the isotropic distribution put into the data originally. The reasons for this difference will be explained in the next section.

At the very least this exercise shows that the method itself works and that the issue may lie elsewhere, therefore let us move on to discuss the possible causes of the observed discrepancy.

6.7.1 Observational bias towards 90 degrees

Therefore the issue must either arise from features present in the data that have not been accounted for in the artificial data or down to selection effects. The first point is unlikely since if this was caused by, for example by asymmetric splitting, the effect would have to be large enough that it would either be observable in the spectrum or that the fit (when inspected) would obviously not provide a good fit to the data. Each mode fitted in the real data has been inspected by eye to ensure that this is not a problem.

Firstly the cuts mentioned in the previous section can be applied to the data to see if it changes the structure of the final posteriors, if there is some change then this gives an idea of the properties of the modes being fitted. As can be seen in 6.21 these cuts make very little difference to the overall structure of either sample posteriors, therefore showing that very few of the modes in the sample were affected by the cut.

In addition we can check the number of stars expected to be in a given inclination angle range under the assumption of an isotropic distribution. This is done simply by integrating $p(i_{\rm s}) \propto \sin i_{\rm s}$ over the desired range. In this case we would like to check the expected number of stars with inclination angles greater than 80°, $\int_{80^\circ}^{90^\circ} \sin i_{\rm s} {\rm d}i_{\rm s} = 0.1736$. This gives a total of 15 stars for a sample size of 90 stars. Comparing this against the number of stars within the real data is a bit tricky. If we take median inclination angles above 80° in the real data then this is 17 stars. But if instead we take those whose mode of their inclination angle

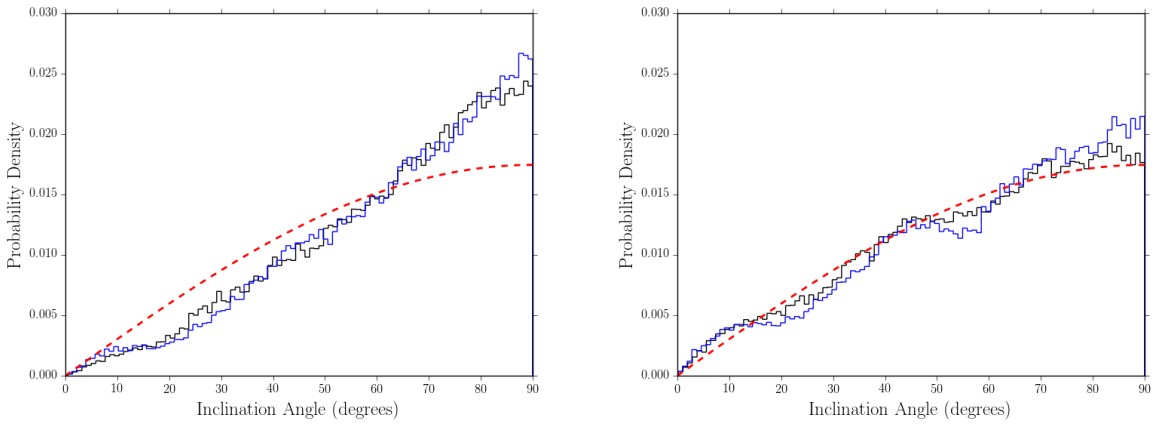


Figure 6.21: The left hand panel shows the concatenated posteriors for the real stars in black and the posteriors with the proposed cuts in H/B and Γ made in blue. The right hand panel shows the same thing but for the artificial stars. The isotropic distribution is overplotted in red in both cases.

distribution lies above 80° then there are 27 stars. Therefore getting a handle on exactly how big the excess is can prove to be difficult.

6.7.2 Fractional uncertainties on individual modes and stars

A key diagnostic to check the method is definitely introducing no biases is to look at the uncertainties on the individual modes compared to those on the concatenated posteriors for each stars. If there were large differences in the uncertainties between the modes themselves and the concatenated posteriors created from the modes this would be indicative of the individual modes having slightly different fitted inclination angles.

The uncertainties on the inclination angle were taken once the cuts of $\Gamma > 5\nu_{\rm bw}$ and $H/B > 18.9$ were made in order to ensure this was applied to the most up-to-date distribution. Fig 6.22 shows the resultant distributions for both the real and artificial data. Both show the same shape of distributions whereby the median uncertainty on the individual modes is slightly smaller than the uncertainty on the concatenated stellar inclination. This would indicate a small amount of scatter present in the median inclination angles extracted for each mode, which when all the posteriors are combined results in a slightly wider final posterior.

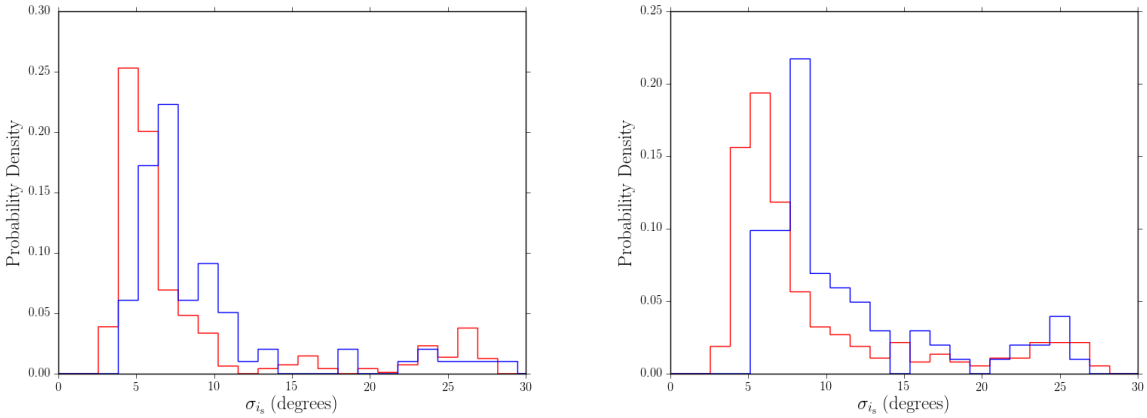


Figure 6.22: A comparison between the uncertainties extracted from the posteriors of the individual modes (in red) and the concatenated posteriors for each star (in blue). The left-hand panel shows this for the real stars and the right-hand panel shows this for the artificial sample.

The indications of this are that the method by which the individual mode posteriors are combined is sub-optimal and could be improved, possibly by applying Bayesian hierachical analysis with a suitable model. The difference in the medians of the two distributions in both cases is no more than 2° which indicates that this difference in the individual modes cannot be particularly large. As a result the method still holds up and is not responsible for the deviations from isotropy observed in the real data.

6.7.3 Selection effects

An important point to remember is that the initial sample contained 120 stars, of which 90 had modes that could be easily identified and fitted. Therefore the effective sample size has been reduced and this could cause deviations from the expected distribution, especially if the removed stars tended to have low angles of inclination and could not be identified due to lower signal to noise (for example). The amount of structure expected is large in the observed distribution given our small sample size, however this reduction in the sample size (because it is not random) could be responsible for an observed bias. It also helps that 90° is the easiest angle to observed, coupled with the fact that it is the most common. As a result

it can be inferred that the stars not included could be more likely to possess lower angles, thus biasing the result. However this cannot be proven since there is no way to extract the angle from those stars under the current method. The other reason for not being able to identify the modes is that the star shows suppressed dipole modes (e.g. García et al. 2014; Mosser et al. 2017) and so as the mixed modes have much lower visibilities than expected in this case they generally cannot be seen above the noise to be identified.

There is another effect which is, arguably more important and has been alluded to throughout this chapter, the idea of a 90° attractor. This comes about because angles of 0 and 90° are the easiest to see visually, since the former is a single peak and the latter is a doublet, both of which will have larger amplitudes than the triplet observed at intermediate angles. However, due to the assumption of isotropy observing an angle close to 0° has a very low probability which therefore creates a bias towards 90°. This bias is observational and caused by the fact that, for a given signal-to-noise, stars at 90° are easier to observe. This would help explain why the excess at 90° is seen in the real data but not the artificial data. Therefore this would likely be a selection effect that is caused in the mode identification stage rather than the fitting stage.

6.8 Conclusion

In conclusion, the angles of inclination were extracted for a sample of 90 *Kepler* red giants, using individual fits to $\ell = 1$ mixed modes. The posterior PDFs for each mode were concatenated to create a PDF for each star and then subsequently concatenated to produce the overall inclination angle distribution for the sample. Through the use of hierarchical Bayesian analysis the data was found to be consistent with an isotropic distribution at the 1.6-σ level although an excess of probability mass was observed at 90°.

The observed excess at 90° was investigated firstly through the observed spatial distribution of the *Kepler* magnitudes, distances and inclination angles of the stars in the sample. No

clumping of stars were found and there was no indication that the spatial distribution of the stars was causing the excess. Potential biases in the method were investigated through the creation of artificial modes covering a range of signal-to-noise and reduced splitting space. Whilst the angle of inclination could be retrieved for modes with $\Gamma > 5\nu_{\rm bw}$ and $H/B > 18.9$, for unresolved modes the angle of inclination cannot be trusted since the model being fitted to the data is not adequate (due to the poor retrieval of the mode linewidth in such cases). As a result the angle of inclination should be interpreted with care for modes with linewidths < 0.04 μHz.

As a further step, the biases were investigated by generating an artificial set of stars with identical properties to the real stars (including imposing an isotropic angle of inclination distribution). These artificial stars were fitted using the same procedure as the real data and the input isotropic distribution was extracted, showing not biases in the methodology.

The excess of stars at 90° is likely due to the idea of a 90° attractor. There is an observational bias towards 90° attractor due to the doublet the mode profiles take at that angle. Given the signal-to-noise ratio in the data, the most likely angle to observe is 90° assuming an isotropic distribution. As a result this is the most likely cause of excess seen in the real data.

7

KOI-3890

7.1 Introduction

Determinations of the true obliquity ψ (the angle between the binary orbital axis and the stellar rotation axis) in binary systems have been long sought after, as they can shed light on formation mechanisms (e.g. Bonnell et al. 1992; Bate et al. 2010) and the dynamical evolution of the system (Mazeh & Shaham, 1979).

The measurement of ψ remains a difficult problem due to the need to derive three angles from the system, all of which must be done using different methods. Namely the angle of inclination of the star $i_{\rm s}$ (that is the angle between the rotation axis of a star and our line-of-sight), the orbital inclination angle $i_{\rm o}$ (the angle normal to the place of the orbit) and the sky-projected obliquity λ (defined as the angle between the projections of the orbital and rotation axes on the sky). The angle $i_{\rm o}$ is determined from photometry by modelling the

light curve, λ is commonly found using the Rossiter-McLaughlin (RM) effect (McLaughlin, 1924; Rossiter, 1924) (see e.g.Winn 2007 or Gaudi & Winn 2007 for and overview of the effect), which can be especially difficult for eclipsing binaries (Albrecht et al., 2007). But other methods do exist such as the photometric RM effect (Shporer et al., 2011, 2012; Groot, 2012).

There are a number of methods designed to measure the inclination angle of the star $i_{\rm s}$. For example, from knowledge of the rotational velocity of the star $v \sin i_{\rm s}$, the rotation period $P_{\rm rot}$ and stellar radius R, e.g. Abt (2001), or from star spot modelling, e.g. Dumusque (2014). More recently Morton & Winn (2014) have provided a statistical framework for deriving the distribution of $\cos i_{\rm s}$ given the inputs $v \sin i_{\rm s}$, $P_{\rm rot}$, and R which was used in the process of deriving the obliquity distribution of 70 *Kepler* Objects of Interest (KOIs).

Asteroseismology is coming to the forefront as a reliable, independent method for determining stellar inclination angles in stars showing solar-like oscillations (as highlighted in Chapter 6). Pesnell (1985) initially theorised how it was possible to extract the angle of inclination in stars showing slow rotation which was later expanded on by Gizon & Solanki (2003); Ballot et al. (2006, 2008). This highlights an important requirement for the extraction of the angle of inclination, the star must be rotating fast enough to see the effects of rotation in the power spectrum. In the special case of a transiting planet, the combination of $i_{\rm o}$ and $i_{\rm s}$ gives access to the obliquity of the system without the need for λ (as will be seen later). Chaplin et al. (2013) presented the first use of asteroseismology in deriving the obliquity for the exoplanet hosts Kepler-50 and Kepler-65.

The most extensive investigation into the obliquities of eclipsing binaries was performed by the BANANA project (Albrecht et al., 2007, 2009, 2011, 2013a, 2014). During these investigations 5 systems were studied making use of the RM effect to derive the sky-projected obliquities of each star in the binary. All the binaries analysed were close-in with periods in the range $\sim 6 - 16$ days, of which 2 were misaligned (CV Vetorum and DI Herculis) and the

rest were consistent with alignment.

Meanwhile the large number of solar-type stars targeted by the space missions such as *Kepler* (Borucki et al., 2010) or CoRoT (Baglin et al., 2006) enables us to move away from more bespoke estimates towards ensemble cases. Campante et al. (2016a) showed how asteroseismology can contribute in an ensemble sense to inferring the obliquity distribution of different populations of main-sequence exoplanet host stars. In this work we show how asteroseismology can be used to place constraints on the obliquity of the red-giant eclipsing binary KOI 3890 through the determination of the stellar inclination angle.

The system being analysed in this chapter, KOI-3890 (KIC 8564976), has been studied in many previous works (Lillo-Box et al., 2014, 2015; Rowe et al., 2015). Lillo-Box et al. (2015) provided characterisation of 13 KOIs through the use of radial velocities of which KOI-3890 was included, however due to the long period of the orbit there was insufficient phase coverage to properly solve for the orbit. Additional radial velocity follow-up has been gathered here to improve upon this and provide characterisation of the host star as well as estimating the obliquity of the system.

7.2 Observations

This work made use of long-cadence (LC) observations from the *Kepler* space telescope and consisted of data taken from Q1-Q17, corresponding to just under 4 years of data (Jenkins et al., 2010a,b). The time-series used in the eclipse fitting consisted of the PDC data, the preparation of which is discussed in the next section. The power spectra used for the asteroseismic analysis was made from PDC data (Smith et al., 2012; Stumpe et al., 2012, 2014a) detrended with a moving-median filter of length 20 days. This removed any long term variations in the photometric time-series; the transit duration of the eclipsing binary was short enough that it was not affected and the period (of 152.8 days) was long enough such that transits do not have an impact on the asteroseismic analysis if not removed.

Table 7.1: Spectroscopic parameters derived from the TRES observations for KOI-3890.

Parameter	Result
$T_{\rm eff}$ (K)	4726±52
$\log g$	2.92±0.10
[Fe/H]	-0.13±0.08
$v \sin i$ (kms^{-1})	4.1±0.5

In addition to the space-based *Kepler* observations spectroscopic data were acquired using the TRES optical echelle spectrograph on the 1.5-metre Tillinghast telescope at the Fred L. Whipple Observatory to measure radial velocities for the system. The Stellar Parameter Classification pipeline was used to derive the atmospheric parameters (Buchhave et al., 2012) and a subsequent additional derivation was performed using the asteroseismic $\log g$ as a prior on the spectroscopic value. It was these latter values that were adopted due to the slight disagreement between the inferred $\log g$ values from the original analysis. Estimates of $T_{\rm eff}$, $\log g$, [m/H] and $v \sin i$ were obtained, and [M/H] (the relative metal abundance) was assumed to be is equivalent to the metallicity [Fe/H]. In total 9 radial velocity measurements were taken over 192 days, derived from multi-order fitting to spectral templates.

7.3 Close companions

KOI-3890 is currently flagged as a possible eclipsing binary (Coughlin et al., 2014) and both Coughlin et al. (2014) and Lillo-Box et al. (2014) established that it is an isolated KOI. The follow-up by Lillo-Box et al. (2015) was unable to characterise the system fully due to poor phase coverage of the radial-velocity follow-up. However Lillo-Box et al. (2015) did calculate the blended source confidence (the probability that the system was isolated, i.e. the probability that it is not a blend) for the system which was 0.983, therefore we can use this as an additional reason to help validate its status as an eclipsing binary and not a blend. Fig 7.1 shows why the blended source confidence is an important quantity for us to have access to, due to there being another star close to KOI-3890 in the field, KIC 8564966. In

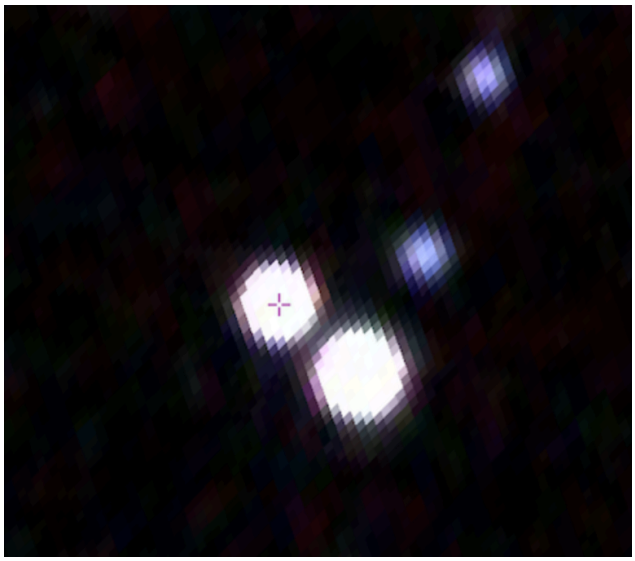


Figure 7.1: An image of the field surrounding KOI-3890 (identified by the red cross) by the Aladin Sky Atlas (`http://aladin.unistra.fr`). The field of view of image is 59.21" provided and the star in close proximity is another red giant KIC 8564966.

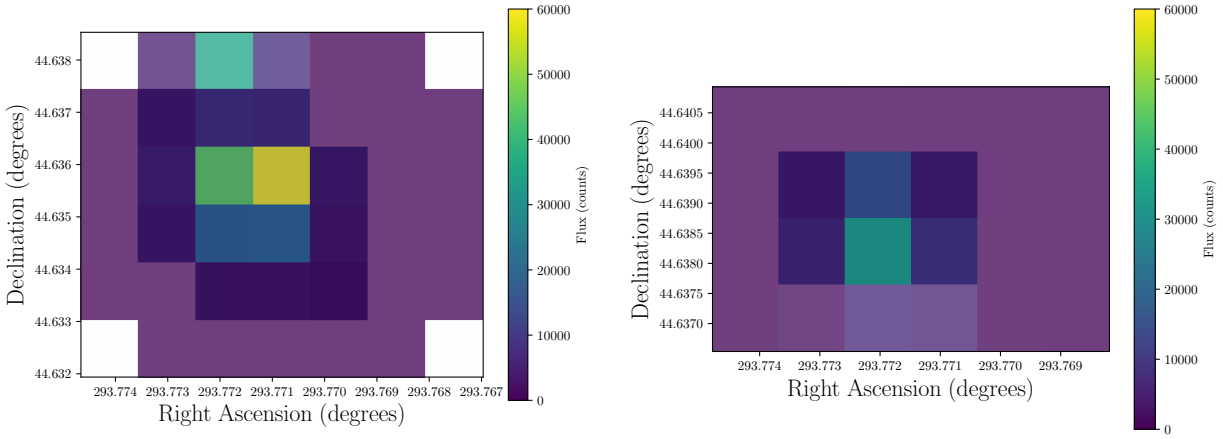


Figure 7.2: Target pixel files for KIC 8564966 (left-hand panel) and KOI-3890 (right-hand panel). The mask used to extract the photometry is shown in the blue highlighted region.

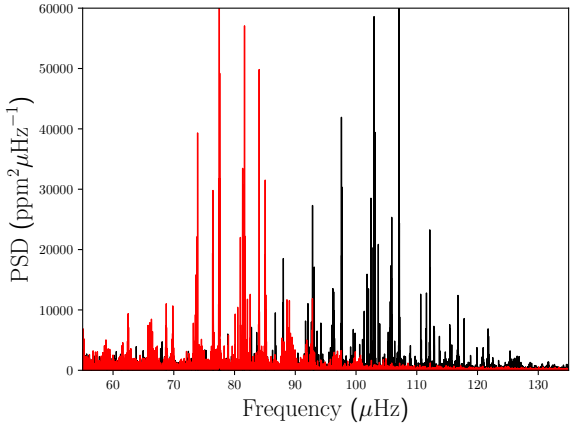


Figure 7.3: A comparison between the power spectra of KOI-3890 shown in black and KIC 8564966 shown in red zooming in on the region around the oscillations.

addition to providing the blended source confidence, Lillo-Box et al. (2014) also showed that there was not a companion with 6 arc-seconds of the star and that KOI-3890 is not a blend. This can also be confirmed by looking at the pixel masks used to extract the photometric data of each star, shown in Fig 7.2. The pixel masks show no overlap between the two stars and as a result it is reasonable to assume very little contamination from the neighbouring star.

Each star was also checked to see which the eclipse signatures come from. This was done by applying the box-least squares (BLS) algorithm (Kovács et al., 2002) to the time-series of each star. The BLS algorithm consists of sliding a box-shaped transit with varying width and depth across a phase-folded version of the time-series and deducing which combination of parameters fits the data best by computing a proxy of the signal to noise. The results are shown in Fig 7.4 where the clear excess at the correct period can be seen around KIC 85764976 and not around the nearby star.

Fortunately, the nearby star is also a red giant showing solar-like oscillations which enables us to better characterise the effect (if any) of contamination of the power spectrum due to the presence of this nearby star, the power spectra of the two stars are shown in Fig 7.3.

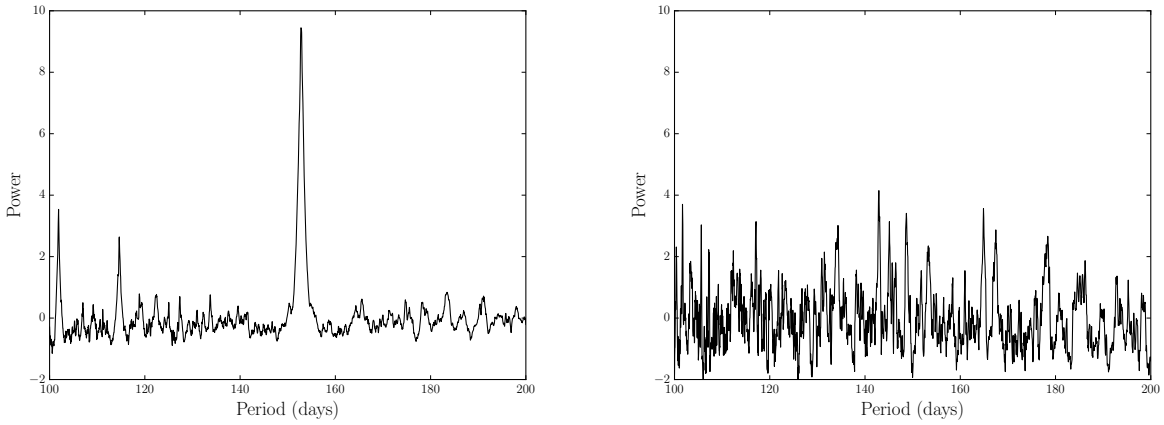


Figure 7.4: Signal residue outputs from the BLS algorithm for KOI-3890 (on the left) and KIC 8564966 (on the right). The peak around 150 days in the left hand plot shows the presence of the eclipses with that period, whereas for KIC 8564966 there is no such peak, again indicating a lack of contamination.

Through the observation of tens of thousands of red giants with CoRoT (Baglin et al., 2006) and *Kepler* (Borucki et al., 2010) we have gained the ability to identify and predict the oscillation modes in red giants with very good precision (Mosser et al., 2011). The modes of oscillation are clearly visible for each star (as seen in Fig 7.3) and are suitably well separated that any mode not belonging to the star in question would be readily apparent. In other words, since there are no modes appearing at unexpected frequencies and so there does not appear to be any contamination between the two stars.

7.4 Simultaneous radial velocity and transit fitting

The data used in the transit fitting was taken from Q2-Q17 *Kepler* PDC-MAP data (Smith et al., 2012; Stumpe et al., 2012, 2014a). The detrending was performed on a quarter by quarter basis, using a 20-day moving median filter, alongside an iterative 4-σ clipping to remove outliers[1]. Due to being only 30 days in total Quarter 1 was ignored, and no eclipses occurred during this time. Another eclipse was only partially observed (at T[BJD]$\sim$ 55641) simply because it fell in a gap between Quarters 8 and 9, as a result it was removed before

[1]The 4-σ clipping was chosen deliberately such that it did not affect the transits, due to the relatively small eclipse depths.

the fitting procedure.

The radial velocity observations[2] (middle panel of Figure 7.5) indicate a rather eccentric orbiting companion, however the values listed in the NASA Exoplanet Archive (Akeson et al., 2013) for KOI-3890 are for a circular orbit. This is due to the lack of secondary transits observable and in turn distorts the other parameters for the system, as listed in the Exoplanet Archive.

The radial velocity and transit signals were fitted simultaneously using the Markov chain Monte Carlo implementation `emcee` (Foreman-Mackey et al., 2013), where the radial velocity signal was the constraint on the eccentricity of the system (since no secondary eclipses were observed). Due to the fact that no secondary eclipses were observed it was deemed appropriate to treat the system in the same way one would a transiting planetary system. The transit signal was modelled using `batman` (Kreidberg, 2015), using quadratic limb darkening parameters taken from Sing (2010) for a star of $\log g = 3, T_{\rm eff} = 4750$K,[Fe/H]=-0.1, (very similar to spectroscopic parameters in Table 7.1), in the *Kepler* bandpass. The radial velocity signature was modelled using the Python/C codes `ajplanet`[3].

Table 7.2 shows the median values for each parameter in the fit, along with errorbars associated with the 68% credible intervals. Due to the eccentric nature of the orbit, the time of periastron T_0 is a separate parameter from the time of central transit T_c.

The results given in Table 7.2 provide us with some degree of insight into the properties of the binary system. Certainly due to the low value of $R/R_\star$ the companion is considerably smaller ($\sim$ 20 times) than the red giant primary and we shall proceed to derive the global stellar parameters in section 7.5.

[2]The radial velocity and transit fitting procedure detailed here was performed by Thomas North as part of the work to be published in Kuszlewicz et al. (2017)

[3]https://github.com/andres-jordan/ajplanet

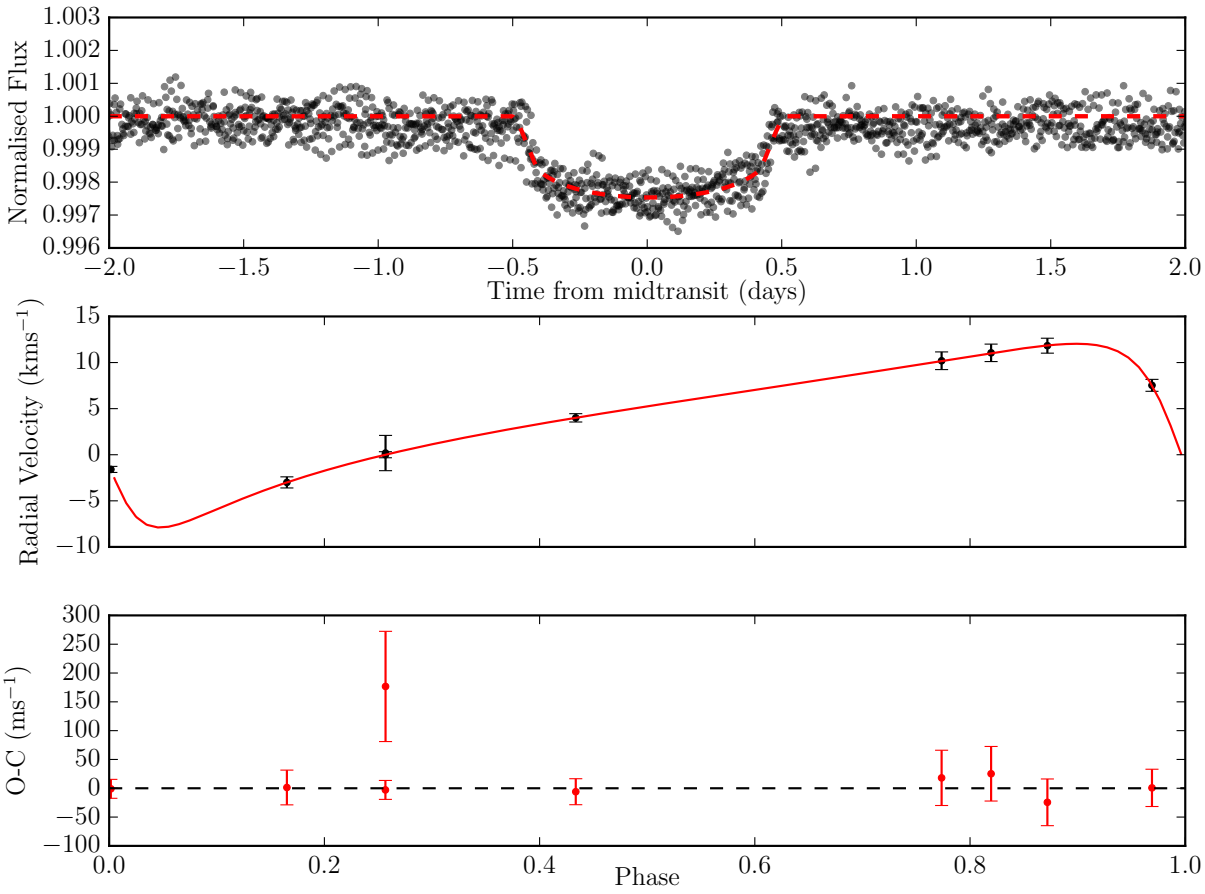


Figure 7.5: The best fitting simultaneous transit and radial velocity observations. The upper panel shows the detrended lightcurve, folded on the period from the fitting procedure, with the out of transit lightcurve normalised to 1. The middle panel shows the radial velocity signal, observed by TRES as the black points, and the best fitting model in red, with errobars inflated by factor 20 observe them. Lower panel shows the residuals of radial velocity fit (with non-inflated errorbars).

Table 7.2: Final values from simultaneous transit and radial velocity fit, along with the priors applied in the fit. Uniform priors are in range [a, b] in the final column, while normal priors [μ,σ] values are shown.

Parameter	Result	Prior Type	Range
Period(days)	$152.82605^{0.00025}_{0.00026}$	Uniform	[152,153]
T_c[BJD-2400000]	$55030.4085^{0.0015}_{0.0015}$	Norm	[55030.18,1]
$a/R_\star$	$18.00^{0.39}_{0.39}$	Norm	[18,5]
$R/R_\star$	$0.0492^{0.0003}_{0.0003}$	Norm	[0.05,0.01]
i_o[deg]	$84.11^{0.23}_{0.24}$	Uniform	[80,90]
V_{sys}[kms^{-1}]	$4.005^{0.013}_{0.013}$	Uniform	[0,10]
K[kms^{-1}]	$9.965^{0.037}_{0.037}$	Uniform	[0,15]
T_0[BJD-2400000]	$57324.36^{0.07}_{0.07}$	Uniform	[57320,57330]
e	$0.609^{0.001}_{0.001}$	Uniform	[0,1]
w[deg]	$108.6^{0.4}_{0.4}$	Uniform	[0,180]

7.5 Global stellar parameters

In order to better characterise the system the next step is to determine the global stellar parameters. To do this we employ the methods of asteroseismology, which involves the extraction of the parameters $\nu_{\rm max}$ (the frequency of maximum power) and the large-frequency separation $\Delta\nu$ (see Chaplin & Miglio (2013) and references therein for an extensive overview of red giant and solar-type asteroseismology).

The extraction of $\nu_{\rm max}$ was performed using the method described by Davies & Miglio (2016) which involved fitting the background of the solar signal in the power spectrum using Model H from Kallinger et al. (2014). This consists of two Harvey-like profiles to model the granulation background (based on that given in Harvey 1985), a Gaussian envelope for the oscillations and an instrumental background. The central frequency of the Gaussian profile was taken to be $\nu_{\rm max}$ and this fit is shown in Fig 7.6. This fitting process proceeded under a Bayesian framework also making use of the affine-invariant ensemble sampler `emcee` (Foreman-Mackey et al., 2013) and the likelihood function given in Anderson et al. (1990) using the method shown in Chapter 3. All amplitude parameters (including the height of the Gaussian envelope) were fitted in log-space whereas the rest were fitted in linear-space, all

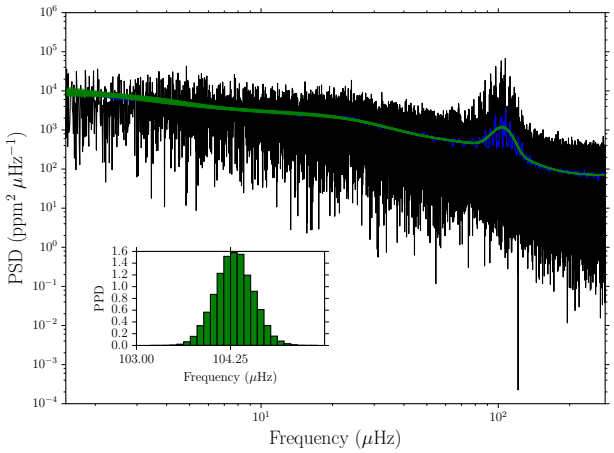


Figure 7.6: The background fit to the power spectrum of KOI-3890. The power spectrum is shown in black and in blue is the spectrum rebinned over 1 μHz. The green lines show random draws from the posterior distributions of the fitted parameters. Inset is the posterior probability density function for $\nu_{\rm max}$.

with uniform priors. The values quoted are given as the median of the posterior distribution and the 68.3% credible interval, determined by the highest posterior density (HPD), giving a fitted value of $\nu_{\rm max} = 104.3 \pm 0.3$ μHz.

The average large frequency separation $\langle \Delta\nu \rangle$ was determined using the autocorrelation (hereafter ACF) of the power spectrum (Campante et al., 2010; Karoff et al., 2010; Huber et al., 2009). The regular spacing of the modes manifest themselves as a set of equally spaced peaks in the autocorrelation where the highest peak is a general indication of $\langle \Delta\nu \rangle$.

To derive $\langle \Delta\nu \rangle$ from the ACF the work of Kiefer et al. (2015) was followed by fitting a triple-Lorentzian model to the data in a region around the highest peak using a non-linear least-squares algorithm. This gave $\langle \Delta\nu \rangle = 9.57 \pm 0.21$ μHz, where the central frequency of the triple Lorentzian was taken to be $\langle \Delta\nu \rangle$ with a conservative estimate of the error being given by the half width at half maximum of the central peak.

The global stellar parameters of the red giant primary can be obtained through grid-based modelling (e.g. Stello et al. 2009; Basu et al. 2010; Gai et al. 2011; Chaplin et al. 2014a) using the Bayesian grid-modelling code `PARAM` (Rodrigues et al., 2014, 2017), which finds the best fit stellar evolutionary model (from a chosen grid) given a set of input parameters . In

this work PARAM was run using models generated using MESA[4] (for a detailed discussion of the input physics see Rodrigues et al. 2017).

The set of inputs to the grid-modelling consisted of $\{[\mathrm{Fe/H}], T_{\mathrm{eff}}, \Delta\nu, \nu_{\mathrm{max}}, \Delta\Pi_1\}$, where the period spacing $\Delta\Pi_1 = 75.9 \pm 0.6$ s was taken from the analysis by Vrard et al. (2016). The inclusion of $\Delta\Pi_1$ with a small fractional uncertainty will result in a tighter constraint on the mass of the star due to the information that it provides about the size of the stellar core. This dependence has been observed both theoretically and observationally (Stello et al., 2013; Montalbán et al., 2013; Mosser et al., 2014; Vrard et al., 2016; Lagarde et al., 2016). The outputs from the grid modelling are shown in Table 7.3 including the credible intervals taken as the 68.3% highest posterior density. The errors retrieved on mass are extremely small as a result of the inclusion of the the period spacing since this offers such tight constraints on the core mass (even with fairly large uncertainties), but are nonetheless in line with uncertainties derived from other works (e.g. Pérez Hernández, F. et al. 2016 who found average uncertainties on mass of 2%).

Given that we have derived a mass and radius of the red giant using asteroseismology it would be worth briefly discussing how it relates to other determinations such as the revised KIC (Huber et al., 2014), in which the mass and radius are given as $M_{\mathrm{KIC}} = 1.761^{+0.335}_{-0.728}\ M_{\odot}$ and $R_{\mathrm{KIC}} = 7.506^{+0.793}_{-1.748}\ R_{\odot}$. The masses in the KIC were derived using average relations between literature values of $\mathrm{T_{eff}}$, $\log g$ and [Fe/H] and masses from Padova isochrones (Girardi et al., 2000), whereas radii were derived from luminosities and $\mathrm{T_{eff}}$. It is however known that for red giants the masses and radii are biased (systematically) as the isochrone grids do not include Helium core burning models, resulting in low mass stars (such as KOI-3890) being matched to higher mass models (which subsequently include helium core burning). The results obtained in this work are consistent within the errors derived from the KIC but the much tighter constraints offered by asteroseismology imply a much less massive star.

[4]Modules for Experiments in Stellar Astrophysics (Paxton et al., 2011, 2013, 2015).

Table 7.3: Stellar properties derived from grid-modelling procedure.

	Primary	Secondary
Radius ($R_\odot$)	$6.12^{+0.09}_{-0.08}$	0.301 ± 0.005
Mass ($M_\odot$)	1.14 ± 0.03	0.248 ± 0.004
Age (Gyr)	6.4 ± 0.6	-

The mass and radius of the secondary in the binary system can also be inferred given the known mass and radius of the primary star. It is unfortunate that the system is not a double-lined spectroscopic binary so there is not an independent measure of the masses and radii of the two stars (Frandsen et al., 2013; Brogaard et al., 2016; Gaulme et al., 2016) to compare. But the results from asteroseismology can be used to gain an insight into the properties of the secondary.

Lillo-Box et al. (2015) previously obtained radial velocity data from this system however given the limited phase coverage could do no more than put a lower limit on the eccentricity $e > 0.33$ and mass of the companion $M_c > 0.097 \pm 0.014 M_\odot$. Using our data which has considerably larger phase coverage there is agreement that the system is in fact highly eccentric. To infer the mass of the secondary companion knowledge of the radial velocity semi-amplitude, K, orbital period, P, eccentricity e and the mass of the primary $M_\star$ to find M_c can be applied through e.g. (Cumming et al., 1999)

$$K = \frac{2\pi G}{P}^{1/3} \frac{M_c \sin i_c}{(M_\star + M_c)^{2/3}} \frac{1}{(1-e^2)^{1/2}}. \tag{7.1}$$

From the above equation and the values given in Table 7.2 a mass of $M_c = 0.248 \pm 0.004$ $M_\odot$ is obtained (which is considerably larger than the lower limit set by Lillo-Box et al. 2015), where the uncertainty has been calculated from the standard deviation of Monte-Carlo simulations drawn from the distributions found on the other parameters.

The radius of the companion is much simpler and can be found directly from $R/R_\star$ (given in Table 7.2) and our derived stellar radius from the grid-modelling. Using these values

$R = 0.301 \pm 0.005\ R_\odot$ and so along with the derived mass for the secondary this places it firmly in the M-dwarf regime. As a slight aside, due to $a/R_\star \sim 0.5$ AU we would expect the companion will eventually be engulfed by the red giant primary and so perhaps this could be a progenitor system for some sdB M-dwarf binary systems (e.g Østensen et al. 2010).

The lack of a secondary eclipse is also important and could give us some extra information about the system, the derivation behind this is given in Appendix A.13.

7.6 Determination of obliquity

The angle of inclination of the binary orbit, i_o, is only one piece of the puzzle towards determining ψ. The next is i_s, which is described in chapter 5.

To extract the angle of inclination the methods described in Appourchaux et al. (2012), Chaplin et al. (2013), Benomar et al. (2014), Campante et al. (2016a) are followed (but instead applied to red giants rather than main-sequence stars) where Lorentzian profiles (including rotational splitting) are fitted to the individual mixed modes. This was performed using the same framework as the background fitting (described in Chapter 3) making use of MCMC sampling to explore the parameter space. Uniform priors were applied to all parameters except the angle of inclination for which an isotropic prior ($p(i_\mathrm{s}) \propto \sin i_\mathrm{s}$) was applied under the common assumption that stars are isotropically distributed in angle.

The modes fitted are shown in Fig 7.7, and were identified aligning the asymptotic expression for mixed modes (Mosser et al., 2015) by eye to each radial order. Only those modes consistent with the resultant expression were fitted. In addition those modes with posterior distributions that were too noisy or resulted in poor fits were also disregarded which explains one or two gaps in Fig 7.7[5].

The use of MCMC to fit the individual mixed modes enabled us to extract the posterior

[5]The mode identification here is currently disputed since the rotational splittings are not consistent with the mode identification made and there is the possibility that the modes are overlapping each other. This observation was made after the main body of this work had been performed and so will need to be properly checked in the future to ensure the situation can be resolved.

probability density function (PDF) of the angle of inclination for each fitted mode. In order to turn this into a posterior PDF for the entire star all of the posteriors for the individual modes were concatenated into one combined posterior and this distribution is given in Fig 7.8. This is made under the assumption that all of the modes in the star possess the same angle which, assuming the core and envelope are aligned, is valid. The angle of inclination of the primary star was therefore found to be $i_{\rm s} = 80.6^{+9.4}_{-6.9}$ degrees. Combined with the value of $i_{\rm p}$ this results in the system being consistent will alignment, however this does not have to be the case.

The stellar obliquity, ψ is defined as the angle between the stellar rotation axis and the planetary orbital axis and given by (following Fabrycky & Winn 2009)

$$\sin\psi\cos\phi = \sin i_{\rm s}\cos\lambda\cos i_{\rm p} - \cos i_{\rm s}\sin i_{\rm p}, \tag{7.2}$$

where ϕ is the azimuthal angle. λ could not be determined from our radial velocity data due to the size of the effect (~ 10 ms^{-1}) being close to the uncertainty on each measurement taken. Considering that $\sin i_{\rm p} \sim 1$ the above equation can be greatly simplified. But we can also make use of another simplification, the degeneracy present when measuring $i_{\rm s}$ using asteroseismology. There is no means for us to distinguish between $i_{\rm s}$ and $180 - i_{\rm s}$, which means that for our case $\cos i_{\rm s} = -\cos(i_{\rm s})$. As a result equation 7.2 can be reduced to

$$\cos i_{\rm s} = \sin\psi\cos\phi. \tag{7.3}$$

Equation 7.3 is much simpler, but the ability to distinguish prograde and retrograde orbits is lost as well as the extra sensitivity to the spin-orbit alignment (or misalignment) that λ offers. However some semblance as to the value of the obliquity can still be gained through the use of Monte-Carlo simulations of the above equation rearranged such that ψ is the subject. During the simulations it was assumed that ϕ is uniformly distributed between $-\pi$

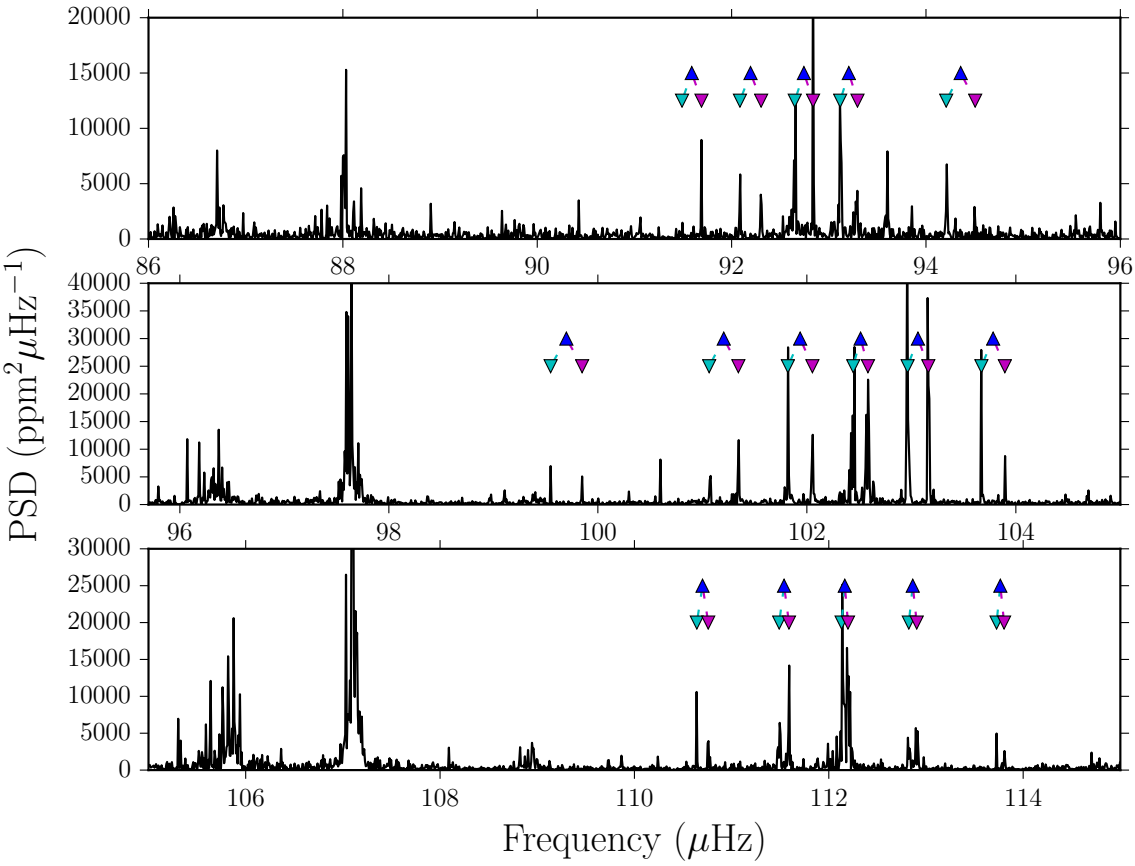


Figure 7.7: The power spectrum of KOI-3890 zoomed in to show the three radial orders used in the determination of the inclination angle. For each $\ell = 1$ mixed mode the blue triangles show the position of the $m = 0$ component, the cyan marker shows the position of the $m = -1$ and the magenta marker shows the position of the $m = +1$. Note the large change in scale as a function radial order.

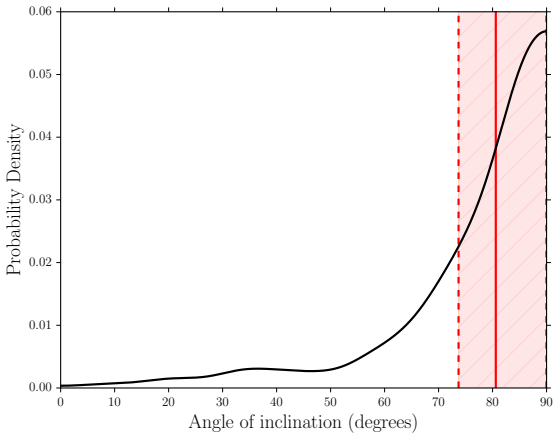


Figure 7.8: Kernel density estimate (KDE) of the angle of inclination posterior probability distribution. The median value is shown by the solid red line whilst the 68.3% highest posterior density (HPD) credible interval is encompassed by the dashed red lines.

and π since no access to the angle is possible.

Using the above method the obliquity of the system is determined to be $\psi = 12.8^{+8.0}_{-12.8}$ degrees as shown in Fig 7.9. Normally a long tail in the PDF would be expected showing the possibility that even though the difference between the two spin-orbit angle is small the system could be misaligned. But due to the isotropic prior imposed on the angle of inclination this ensures that the prior probability goes to zero as the angle tends to zero, therefore resulting in the PDF in Fig 7.9 tending to zero as the obliquity tends to 90 degrees.

Since the angle of inclination of the star is close to 90 degrees the effect of the projected obliquity λ is diminished. As a result the estimate of the obliquity provides a reasonable constraint. A limitation of this method however is that we cannot uncover the full 3D geometry of the system, such as whether the orbit of the M-dwarf is retrograde of prograde. Knowledge of λ from the RM effect would be required in that case. Therefore we can say that it is highly probable that the system is well-aligned.

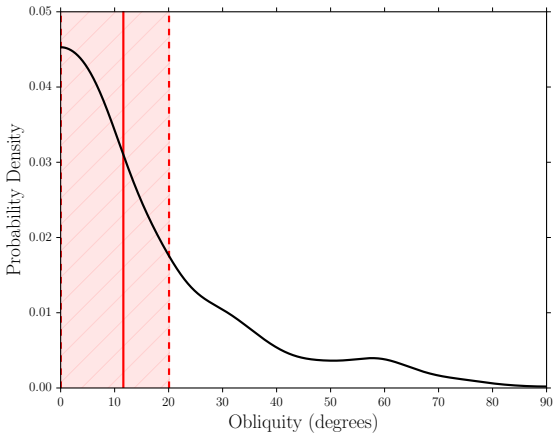


Figure 7.9: Kernel density estimate (KDE) of the Monte-Carlo estimate of the obliquity distribution for KOI-3890. The plot is annotated in the same manner as Fig 7.8.

7.7 Conclusions

An estimate of the obliquity for the red-giant eclipsing binary system KOI-3890 (KIC 8564976) has been provided through the use of eclipse fitting and asteroseismology. KOI-3890 was found to be consistent with spin-orbit alignment with $\psi = 12.8^{+8.0}_{-12.8}$ degrees. This is the first application using asteroseismology to red giants and eclipsing binaries, currently giving the best constraint on the obliquity of a red giant binary system (certainly using asteroseismology). In addition, this gives the ability to derive obliquities of systems with very long periods where radial velocity observations would be time-consuming.

Given the large number of eclipsing binaries, especially involving red giants (e.g. Gaulme et al. 2013), found with *Kepler* and ensemble analysis would be perfectly suited. This would nullify the issue of not being able to determine λ and enable a study of the overall distribution of obliquities (e.g. Campante et al. 2016a). For a transiting system this can be done using only the angle of inclination of the star (Morton & Winn, 2014) and so this shows how the obliquity distribution of eclipsing binaries could be investigated in a similar way. This can pave the way for ensemble obliquity studies of eclipsing binary systems in *Kepler* data as, through the use of asteroseismology, we provide an independent measure of the stellar

inclination angle which other methods cannot provide. Recently some of the issues present with the determination of the obliquity using the RM effect have been highlighted by Bourrier et al. (2016). The advantage of proceeding in an ensemble manner is the ability to disregard λ and use the reduced obliquity value that we can determine here in the same manner as Campante et al. (2016a).

In addition characterisation of the eclipsing binary system has also been provided with the help of stellar grid-modelling on top of the eclipse fitting, radial velocity measurements and asteroseismic analysis. Having been thought of as an eclipsing binary system this has since been confirmed and we have shown that the system is a red giant primary with a M-dwarf companion with respective masses and radii of $M_\star = 6.12^{+0.09}_{-0.08} M_\odot$, $R_\star = 1.14 \pm 0.03 R_\odot$, $M_c = 0.248 \pm 0.004 M_\odot$ and $R_c = 0.301 \pm 0.005 R_\odot$.

8

Conclusion

In conclusion the main aim of this thesis was to use buoyancy-driven oscillations to further our knowledge about both the Sun and red giant stars through the use of probabilistic techniques.

The first step was looking at the Sun using data from BiSON to continue the search for g-modes. Although the search did not result in any detections a large amount of knowledge was gained about the low-frequency regime of BiSON along with the probable causes for a lot of the excess noise. The source of the noise was found to most likely be instrumental giving us the chance to remove this and drastically improve the low-frequency noise levels. Through the use of Principal Component Analysis (PCA) the coherent instrumental trends were modelled over a 7 day period for each station and subsequently removed. The recombination of the data resulted in the noise levels being reduced and, more importantly, the diurnal peaks originally thought to be caused by gaps in our data, were almost completely removed in the power spectrum of each station dataset. As a result the BiSON data has been cleaned and

opened up to give the ability to search for g-modes and attempt to reproduce the results already seen from the GOLF mission.

Moving onwards from helioseismology to asteroseismology, using high quality data to inform inference on low signal-to-noise data is going to become increasing important with the 80 day campaigns of K2 and the 30 day minimum observing periods of TESS. Many red giants have been observed by *Kepler* and will be observed in future missions due to their large amplitude oscillations and so they are the perfect targets for asteroseismic missions as you can probe much further distances. In this chapter we fitted the background profile of the red giants in the APOKASC sample and proceeded to update the asteroseismic scaling relations showing that there are some differences as a result of using the full 4 years of data. A by-product of these results was observing the differences in the oscillation envelope width and envelope height as a function of evolutionary state. This in turn led to the use of the parameters extracted from the background of the power spectrum to classify the evolutionary state of the star with very good success on the *Kepler data* and good success on *Kepler* data degraded to simulate data from K2.

Asteroseismology also provides a means to determine the angle of inclination of the star. This is useful for the exoplanet community since in the case of transiting planets it enables the obliquity of the system to be determined. The distribution of which provides information regarding possible formation mechanisms for systems with only a single transiting planet and those with multiple. In addition the angle of inclination enables us to use asteroseismology to test well held assumptions about the galaxy. One such assumption is that stars are isotropically distributed in angle with respect to an observer. We made use of red giants observed with *Kepler* and Bayesian hierarchical inference to show that the angle of inclination distribution of our sample was consistent with isotropy, but that there were additional biases present most likely caused by the selection of our sample.

The obliquity is also useful in the study of binary systems giving the ability to determine

(in the case of an eclipsing system) whether the system is well-aligned. This was explored in the case of the red-giant M-dwarf eclipsing binary (KOI-3890). Through asteroseismology we were able to determine the mass and radius of the primary red giant star and then provide estimates of the secondary properties. The angle of inclination was also determined showing that the system was well-aligned, however such a system may be interesting as a potential progenitor to binary systems with an sdB star.

A

Appendix

A.1 Data of Finite Extent

In order to fully appreciate the nuances and problems faced when analysing gapped time-series it is worth giving a quick recap of how the data from BiSON (or from *Kepler*) is analysed and some of the theory behind it. The vast majority of work is dedicated to the estimation and detection of astrophysical signals in the presence of noise. This could be the situation where the signal is buried deep within the noise (such as solar g-modes), or in the case of eclipsing binaries where in general the signal is very obvious and the noise does not provide too much of a challenge to the analysis.

In order to properly analyse the signal it would be wise to be able to decompose it using a set of basis vectors that spans the entirety of the signal space. An extremely useful property of the signals that are observed is that they are, in general, periodic and so the basis functions

that can be used in the decomposition are sine and cosine functions. This gives rise to the well known Fourier transform, which can be written as follows (e.g. Appourchaux 2014)

$$X(\nu) = \int_{-\infty}^{+\infty} \exp(-2\pi i \nu t) x(t) \mathrm{dt}, \tag{A.1}$$

where $x(t)$ is the signal in the time domain and ν is frequency. The transform itself can be thought of as a mapping from time t to frequency ν. For continuous signals the transform is defined as an integral from $-\infty$ to ∞ however for observed signals this is not necessary. This is a result of the observations being discrete not continuous (due to there being sampling involved). For example in the case of *Kepler* where a sequence of images are taken at a regular interval and then stacked. This sampling results in the creation of discrete data from a continuous underlying signal and this helps greatly simplify the computation of the Fourier Transform.

In the case of a discrete signal, the signal can be thought of as containing N samples evenly spaced in time (assuming regular sampling), with a sampling time equal to the cadence Δt. This results in equation A.1 becoming (e.g. Appourchaux 2014)

$$X_{DFT}(\nu) = \Delta t \sum_{n=1}^{N} x(t_n) \exp(-2\pi i \nu t_n). \tag{A.2}$$

It is important to remember that the Fourier Transform outputs a complex variable that has a magnitude and a phase. Due to the stochastic nature of solar-like oscillations there is no need to hold onto the phase information of the signal. Therefore, instead of computing the standard Fourier transform of the data (as would be done for coherent signals), the power spectrum of the data (which essentially throws away the phase information) is computed. The power spectrum gives the power contained at a given frequency as opposed to the amplitude and phase and is defined as follows

$$S(\nu) = |X_{DFT}(\nu)|^2 . \tag{A.3}$$

In addition to these concepts, there are a few fundamental parameters that need to be defined before we delve into the world of signal processing. These are the sampling frequency, defined as the inverse of the cadence, $\nu_{\rm samp} = 1/\Delta t$. The second is the Nyquist frequency, $\nu_{\rm nyq} = \nu_{\rm samp}/2$, which is defined as the maximum frequency for which a signal can be completely determined for a given sampling frequency $\nu_{\rm samp}$ (Nyquist, 1928; Shannon, 1949). In the majority of the work carried out we shall not concern ourselves with what happens when a signal is undersampled, i.e. $\nu > \nu_{\rm nyq}$; for more information about the super-Nyquist regime see Murphy et al. (2013); Chaplin et al. (2014b); Murphy (2015).

The process of sampling a continuous signal and decomposing it using a discrete basis set can lead to some unfortunate issues that must be taken into account. If, and only if, the frequency of the signal coincides exactly with a frequency of the basis set will it project onto a single basis vector. Otherwise all other frequencies will result in a non-zero projection onto all basis vectors over the entire set. This is the phenomenon known as spectral leakage and is highly undesirable, however it is simply a consequence of trying to interpret a continuous signal as a discrete one. It is important to realise that if a signal should be contained with a single bin and it has a frequency that is not commensurate with the basis vectors then the power will be spread over multiple bins and the power in the expected bin decreased as a result. This is vital to understand the signals created by g-modes which are expect to have long lifetimes and therefore small linewidths.

A.1.1 Calibration of the Power Spectrum

The calibration of the power spectrum is an important topic and can have a knock on effect to any related amplitude or height measurements extracted from the data. Therefore the means by which the power spectrum is calibrated will be briefly explained.

The starting point is Parseval's theorem, whilst the importance of this to Fourier analysis warrants a section devoted to itself we shall settle for a short explanation to emphasise why it is important. Parseval's theorem provides a relation to the squared, integrated values of a function to its Fourier transform

$$\int_{-\infty}^{+\infty} \mathrm{d}t \, |f(t)|^2 = \int_{-\infty}^{+\infty} \mathrm{d}\nu \left| \tilde{f}(\nu) \right|^2 \tag{A.4}$$

where ν denotes frequency and t is time. Another way of thinking about this is to think in terms of a system of waves, whereby $|f(t)|^2$ is a quantity proportional to the energy density at a time t and $\left|\tilde{f}(\nu)\right|^2$ is a quantity proportional to the energy density at a given frequency ν. Parseval's theorem dictates that the total energy is the same whether the integral is expressed in time or frequency. This is an expression of conservation of energy so that for any physical system, energy is conserved under the transform to and from its corresponding Fourier pairs (i.e. from $t \rightarrow \nu$).

Equation A.4 is valid for continuous signals (hence the integral), but for discrete signals the sampling must be accounted for. As a result, Parseval's theorem when applied to the discrete Fourier transform (DFT) gives

$$\sum_{n=0}^{N-1} |f[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} ||F[k]|^2 \,, \tag{A.5}$$

where F is the DFT of f (which is of length N). This can also be defined in terms of the power spectrum $P(\nu)$ defined $|A|^2 = AA^*$, where A is the DFT and A^* is the complex conjugate of A (not forgetting that the Fourier transform is a complex quantity. Therefore in terms of the DFT[1], the power spectrum is defined

$$P(\nu) = \frac{|\mathrm{DFT}[f(t)]|^2}{N^2}, \tag{A.6}$$

[1]The denominator in the equation below depends on the software being used. For example, in the case of `IDL` the factor N^2 is not need, but in the case of `python` it is needed.

where N is the number of points in the input signal. This process will give us both positive and negative frequencies and considering the data being analysed is purely real we are only interested in those frequencies that are positive, therefore the summation will be only to $N/2$ rather than N as before giving a one-sided power spectrum.

Now equation A.5 can be reformulated in terms of the power spectrum according to

$$\frac{1}{N}\sum_{n=0}^{N-1}|f[n]|^2 = \sum_{k=0}^{N-1}P(k), \tag{A.7}$$

which results in the sum over the power in the power spectrum being equal to the mean squared power in the time domain.

Parseval's theorem provides a handy sanity check when using Fourier transforms. Many different software packages and coding languages use different conventions for normalising the Fourier transforms and so the above equation can provide an easy way to check the data are calibrated properly. It can also be used to directly calibrate the power spectrum, but it is a good idea to understand what process is going on in the background so that the data can also be calibrated manually if needed. As discussed before, the power spectrum of a time series gives an estimate of the power contained at a given frequency for a certain input signal.

The main purpose of the calibration is to obtain a power spectrum in units of power spectral density, using the BiSON data as an example where the units in the time domain are (ms^{-1}), the units of power spectral density would be ($\mathrm{m}^2\mathrm{s}^{-2}\mathrm{Hz}^{-1}$). The current calibration assumes that there are no gaps present in the data, i.e. that the duty cycle is 100%. Unfortunately this is unlikely to be the case, especially in the case of a ground-based network where weather or instrumental effects may cause gaps in the data. The correction made to the calibration is very simple and results in dividing the power spectrum by D where D is the duty cycle

Finally one last correction is made to put the power spectrum into units of $\mathrm{m}^2\mathrm{s}^{-2}\mathrm{Hz}^{-1}$ rather than the $\mathrm{m}^2\mathrm{s}^{-2}\mathrm{bin}^{-1}$ calculated thus far. The correction involves dividing through by the bin-width of the power spectrum (in Hz) and so the final calibration is given by

$$\mathrm{PSD}[\mathrm{m^2 s^2 Hz^{-1}}] = \begin{cases} \frac{|A_0|^2}{DN^2\nu_{\mathrm{bw}}} & \text{for } n = 0 \\ 2\frac{|A_n|^2}{DN^2\nu_{\mathrm{bw}}} & \text{for } n = 1, ..., \frac{N}{2} - 1 \end{cases} . \tag{A.8}$$

A.1.2 The Basics of the DC Component

The data, $y(t)$ can be thought of as some signal, $x(t)$, plus a constant value (i.e. the mean), m, and so it can be written

$$y(t) = x(t) + m \, . \tag{A.9}$$

The Fourier transform is in fact a linear transformation and so the addition of the components in the time domain leads to the addition of their Fourier transforms in the frequency domain. This means that the Fourier transform of the data, $Y(\nu)$, can be written as

$$Y(\nu) = X(\nu) + M(\nu) \, , \tag{A.10}$$

where $M(\nu)$ is the Fourier transform of the constant, which is given by

$$M(\nu) = M \int_{-\infty}^{\infty} \exp(-2\pi i \nu t) \mathrm{d}t = \delta(\nu) \, , \tag{A.11}$$

where M is a constant and $\delta(\nu)$ is a unit impulse defined by

$$\delta(x) = \begin{cases} 1 & x = 0 \\ 0 & x \neq 0 \end{cases} . \tag{A.12}$$

The Fourier transform of the mean value results in a spike with height M at zero frequency and a width corresponding to the frequency resolution. If we look to the same formulation for the power spectrum

$$|Y(\nu)|^2 = |X(\nu)|^2 + |M\delta(\nu)|^2 + 2|X(\nu)M\delta(\nu)| \ . \tag{A.13}$$

The mixed terms in the above equation will be ignored as they have no major effect on the shape of the power spectrum for our example. However one major component has of course been forgotten and that is the convolution with the window function and so the final form of the equation is as follows

$$|Y(\nu)|^2 = |X(\nu)|^2 \otimes |W(\nu)|^2 + |M\delta(\nu)|^2 \otimes |W(\nu)|^2 \ , \tag{A.14}$$

where $|W(\nu)|^2$ is the power spectrum of the window function.

A.2 Derivation of Single Station Window Function

Since the power spectrum of the continuous case has been shown above, the derivation of the single stations case follows as such. Let us consider a finite series of delta functions separated in time by a value T (where T is now 24 hours as opposed to the length of the timeseries)

$$f(t) = \sum_{m=-N}^{m=N} \delta(t - mT), \tag{A.15}$$

where m is an integer. The Fourier transform of the above expression is given by

$$\mathcal{F}[f(t)] = \sum_{m=-N}^{N} \exp(-2\pi i\nu mT). \tag{A.16}$$

This can be simplified to give a more useful quantity (see Appendix A.3 for the derivation)

$$\mathcal{F}[f(t)] = \sum_{m=0}^{2N} \exp(-2\pi i\nu mT) = \frac{\sin[(2N+1)\pi\nu T]}{\sin(\pi\nu T)}. \tag{A.17}$$

where $2N+1$ is the number of days in the datasets. In order to gain the complete expression for the single station window function the above equation can be combined with that derived

in equation 2.4 to obtain

$$|W(\nu)|^2 = \left\{ \frac{\sin(\pi\nu/\nu_{\rm d})}{\nu_{\rm d}} \right\}^2 \left\{ \frac{\sin[(2N+1)\pi\nu T]}{\sin(\pi\nu T)} \right\}^2, \tag{A.18}$$

where $\nu_{\rm d} = 1/T_{\rm d}$ where $T_{\rm d}$ is the amount of time the window function is at 1 each day.

This simple model can be interpreted as having peaks with a central frequency of (m/T)Hz in addition to a sinc^2 modulation in power, giving us the basis from which the more complicated models can be built.

A.3 Simplification of Equation A.16

Equation A.16 is not a particularly useful quantity and it is not immediately apparent as to what this would mean for the power spectrum. Therefore, let us go about simplifying the equation into a more useful for, starting with an identity for summing over a finite geometric series (that does not start at zero)

$$\sum_{n=m}^{N} r^n = \frac{\left(r^m - r^{N+1}\right)}{1-r}. \tag{A.19}$$

It should be clear that r^n from equation A.19 is $\exp(-2\pi i\nu nT)$ from equation A.16 and so by inserting this into equation A.19 (with a change of limits of $m = 0$ to $m = 2N$)

$$\sum_{m=0}^{2N} \exp(-2\pi i\nu mT) = \frac{1 - \exp\left[-2\pi i\nu(2N+1)T\right]}{1 - \exp(-2\pi i\nu T)}. \tag{A.20}$$

A small amount of algebraic manipulation leads to the finally simplified version

$$\mathcal{F}[f(t)] = \sum_{m=0}^{2N} \exp(-2\pi i\nu mT) = \frac{\sin[(2N+1)\pi\nu T]}{\sin(\pi\nu T)}. \tag{A.21}$$

A.4 Sunrise-sunset Equation

Typically the best way to obtain the times at which sunrise and sunset occur through calculating the hour angle

$$\cos\omega_0 = -\tan\phi \times \tan\delta, \tag{A.22}$$

where ω_0 is the hour angle at sunrise or sunset, ϕ is the latitude of the observer and δ is the solar declination. The hour angle is defined to be zero at solar noon, positive in the morning and negative in the afternoon.

Whilst equation A.22 approximates the hour angle, for use with data given by astronomical almanacs a few changes must be made. This is partly because the above equation neglects the size of the solar disc on the sky and the effect of refraction when the solar disc is close to the horizon. In addition the time of sunrise and sunset need to be defined and in our case we define it as the rising and setting of the solar limb (upper). This leads to a more general equation (Cooper, 1969)

$$\cos\omega_0 = \frac{\sin a - \sin\phi \times \sin\delta}{\cos\phi \times \cos\delta} \tag{A.23}$$

where a is the altitude of the centre of the solar disk and set to a value of $a = 0.83°$ (Cooper, 1969). The solar declination δ is defined according to the equation below

$$\delta = -23.44° \cos\left[\frac{360°}{365}(n+10)\right]. \tag{A.24}$$

These equations simply give the hour angle and are not enough to deduce the sunrise and sunset times. For that the following equations are needed, firstly with the equation for solar noon (in Julian days)

$$J_{\text{noon}} = 2451545.0009 + n - \frac{l}{360^\circ}, \tag{A.25}$$

where J_{noon} is the mean solar time (in Julian days) calculated at a given longitude l with n being the number of days since 1st January 2000 at 12:00. This reaffirms common sense and shows that those to the West experience noon later (relative to UTC) than those to the East.

Now that both local solar noon and the hour angle can be calculated, the sunrise and sunset times can be calculated given the following formulae

$$J_{\text{rise}} = J_{\text{noon}} - \frac{\omega_0}{360^\circ}, \tag{A.26}$$

and

$$J_{\text{set}} = J_{\text{noon}} + \frac{\omega_0}{360^\circ}. \tag{A.27}$$

The addition of this modification to the window function means that it becomes possible to simulate the changing length of the day and add an extra stage of "realism" into the creation of the artificial window functions. There is however one component that has not been account for, which comes in the form of the equation of time. In other words accounting for the fact the solar noon changes over the year, commonly known as the analemma. To add this effect in equation A.25 can be modified slightly to give

$$J_{\text{noon}} = 2451545.0009 + n - \frac{l}{360^\circ} + 0.0053 \sin M - 0.0069 \sin\left(2\lambda\right), \tag{A.28}$$

where M is the solar mean anomaly and λ is the ecliptic longitude[2].

[2]The python package `PyEphem` was used for all the computations of sunrise and sunset times

A.5 Adding "Realism"

Whilst the construction of a simple window function provides an easy means to try and understand how the major components interact and manifest themselves in the power spectrum, this is not adequate for simulating the BiSON window function. Apart from the multiple-station element of the network there are also additional components that need to be added in to understand the extra structure they may add into the power spectrum. The hope is that, through the modelling of the BiSON window the features that it introduces can be understood and through that an easier method to remove them from the data.

The first such components come in the form of varying the length of the observing day, since the length of the day is not constant throughout the year, and the variation in solar noon (i.e. the analemma). Therefore these must be calculated in order to correct the window function (a full derivation and demonstration of the calculation is given in Appendix A.4). For the purposes of these simulations the start time was chosen to be midnight on 1st January 2000 for a period of 10 years. The effect of changing the length of the observing day is to modulate the amplitude of the diurnal peaks in a non-trivial manner. Since the effect is modelled as a boxcar with varying width this will be a sinc-squared function in the power spectrum, thereby causing the modulation of the amplitudes.

The result of adding in these terms to the window function gives us peaks of the form seen in the right hand panel of Fig A.1. The observed structure seen in the real data can be reproduced in a broad sense, certainly the lack of broadband noise will affect the presentation of the peaks in the artificial case. This can almost be thought of as the limit spectrum for the window function, before any broadband noise has been introduced through interruptions and gaps. This is a step in the right direction and when coupled with interruptions and combined to simulate multiple stations the amount of realism should stack up to give a good model.

The additional sidebands of the diurnal peaks are seen as a result of this change in length

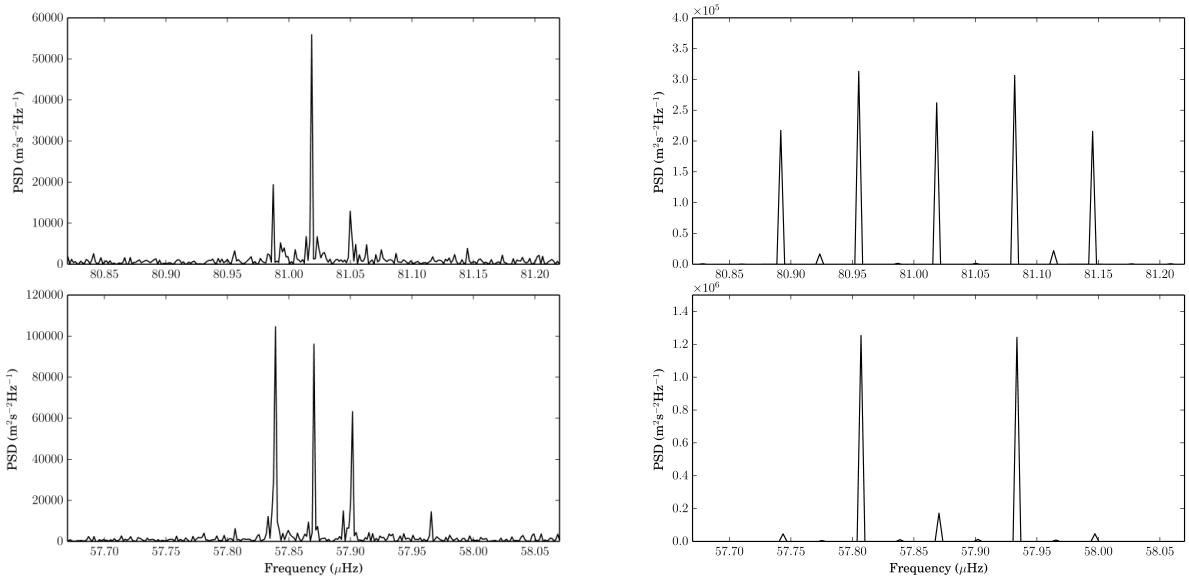


Figure A.1: The left hand panel shows a zoomed in view of the diurnal peaks at 81.02 and 57.87 μHz in the BiSON window power spectrum. The right hand panel shows the equivalent peaks in the artificial window function power spectrum showing how the effect of adding in the change in length of day has affected the shape and structure of the peaks.

of the observing day and change in solar noon. These peaks occur at frequencies of 1/year, 2/year and so on either side of the main diurnal peaks. Therefore any seasonal variation will induce sidebands with respect to the diurnal peaks, however the analytic structure for these peaks has not been derived due to its increased complexity. The idea is remarkably close to work performed by Murphy et al. (2013) whereby the (very slightly) unevenly sampled data taken by *Kepler* can be decomposed into a modulation of the sampling on top of an even sampling rate. This would provide a useful model if it were not for our additional component of change in solar noon which would severely complicate the model.

A.5.1 Interruptions

A slightly different component of the observed window function that is missing from the simulations is the inclusion of interruptions that are not associated with the typical day-night cycle. These come in the form of mechanical faults with the instrument or bad data

taken due to the presence of clouds or other such weather effects. These quasi-random gaps will contribute to the smaller-scale noise in the time-series and so will result in enhanced levels of broadband noise in the power spectrum.

In practice, this amounts to setting the amplitude of the window function equal to zero for a chosen mean duration at random positions throughout the time series. These quantities are describing using probability distributions, due to the (mostly) random nature of the gaps in the data. A point is made here that in the real data the gaps are quasi-random due to periods of time where the station is taken off-line, whether this be for upgrades or such-like and so this results in a greater number of gaps around 24 hours or subsequent multiples. Therefore only the smaller gaps were modelled in these simulations.

In order to model the centre of the gaps (in time), a uniform distribution was used over the length of the time-series. This does have the unfortunate effect of gaps appearing when the station is not observing and so the number of interruptions put into the artificial data would be expected to be larger than the number really observed. The length of the gaps were taken from 10 years of BiSON data over the same period as the simulations (i.e 10 years from 1st January 2000 to 2010). The duration was modelled as a gamma distribution

$$f(\nu; k, \theta) = \frac{\nu^{k-1} \exp\left(-\nu/\theta\right)}{\theta^k \Gamma(k)} \text{ for } \nu > 0 \text{ and } k, \theta > 0, \tag{A.29}$$

where $k > 0$ is the shape parameter and θ is the scale parameter. The distribution of gaps is given by a $\Gamma = 1$ distribution which simplifies the above equation to

$$f(\nu; k, \theta) = \frac{\exp(-\nu/\theta)}{\theta}, \tag{A.30}$$

where the mean value θ is 42 minutes. As can be seen in Fig A.2, the fit is by no means perfect due to the existence of slightly longer period gaps. These simulations are supposed to give us an idea as to how the various features present in the BiSON window function interact

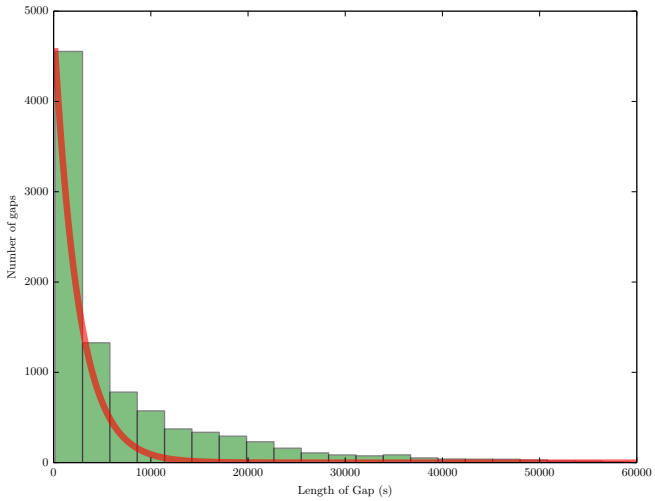


Figure A.2: Gap distribution of 23 years of BiSON data with $\Gamma = 1$ distribution overplotted in red with a mean of 42 minutes.

with the data. In the case of interruptions the above fit does not do a particularly good job at describing what is seen, however for the purposes of the simulations it does an adequate job.

A.6 Peak detection - the case of single spikes

In order to set upper limits on solar g-mode amplitudes it is necessary to work out the threshold in power that constitutes a detection. This will be done using the false-alarm, or H0 approach.

The noise statistics of the power spectrum (as stated earlier) correspond to χ^2_2 and so the probability that the observed power in a particular bin (in frequency) will be greater than or larger to a value $P(\nu)$ is

$$p(P(\nu)) = \frac{1}{\langle P(\nu) \rangle} \exp\left(-\frac{P(\nu)}{\langle P(\nu) \rangle}\right) . \tag{A.31}$$

An important assumption made is that our estimate of the power contained within the

mode $\langle P(\nu)\rangle$ (i.e. the model) is a good estimate of the underlying limit spectrum. This enables us to change the quantity we use from the mode power to the relative mode power, and this simplifies the above equation

$$p(s_\nu) = \exp\left(-s_\nu\right) \ , \tag{A.32}$$

where s_ν is the relative power $P(\nu)/\langle P(\nu)\rangle$. This equation gives us the probability of observing a mode in a single bin, but if we want to search over N bins then it becomes a little more involved. Firstly it is best to work in terms of failure to observe the spike which from above becomes $1 - p(s_\nu)$ and so the failure to observe a spike over N bins becomes $\left[1 - p\left(s_\nu\right)\right]^N$, leaving the probability observing at least one spike in N bins as

$$p_N = 1 - \left[1 - p\left(s_\nu\right)\right]^N \ . \tag{A.33}$$

However this only gives us the probability, whereas in our current situation we want to work out the threshold in relative power, s_ν, given some arbitrary threshold probability, $p_{\rm thresh}$, i.e. $p(s'_\nu \leq s_\nu) = p_{\rm thresh}$. This is done with the help of the following equation (Appourchaux, 2004)

$$p(s'_\nu \geq s_\nu) = \int_{s_\nu}^{\infty} \exp(-s'_\nu) \mathrm{d}s'_\nu \ , \tag{A.34}$$

where we set $p_{\rm thresh} = p(s'_\nu \geq s_\nu)$ and invert the above equation to solve for s_ν which is done by recognising that equation A.34 is the upper incomplete gamma function.

This test is a good starting point, however ideally we would extend the test to account for the doublet of quadruplet structure of the $\ell = 1$ and $\ell = 2$ modes respectively. This would result in a lower threshold for detection, and will be pursued in the future.

A.7 Derivation of the likelihood function for rebinned data

The likelihood function for data averaged over s bins is given by (Appourchaux, 2003a)

$$\mathcal{L}(P_i, \theta) = \frac{s^{s-1}}{(s-1)!} \frac{P_i^{s-1}}{M_i(\theta)^s} \exp\left[-\frac{sP_i}{M_i(\theta)}\right]. \tag{A.35}$$

Taking the natural logarithm gives

$$\ln \mathcal{L}(P_i, \theta) = \ln\left(\frac{s^{s-1}}{(s-1)!}\right) + (s-1)\ln P_i - s\left(\ln M_i(\theta) + \frac{P_i}{M_i(\theta)}\right). \tag{A.36}$$

The first two terms on the left-hand side are constants with respect to each set of data and therefore the negative log-likelihood can be approximated as

$$-\ln \mathcal{L}(P, \theta) \approx s \sum_i \left[\ln M_i(\theta) + \frac{P_i}{M_i(\theta)}\right], \tag{A.37}$$

where s is the number of bins rebinned over.

A.8 Gelman-Rubin convergence criterion

The Gelman-Rubin convergence criterion assesses the mixing and stationarity using variances computed between and within the chains used in the sampling. This is performed on the chains themselves ψ_{ij} where $i = 1, ..., n$ denotes its position in the length of the chain and $j = 1, ..., m$ denotes the chain. The within-sequence and between-sequence variances, W and B respectively can be computed according to

$$W = \frac{1}{m} \sum_{j=1}^{m} s_j^2, \tag{A.38}$$

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\bar{\psi}_{.j} - \bar{\psi}_{..})^2, \tag{A.39}$$

where

$$\bar{\psi}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\psi_{ij}, \tag{A.40}$$

$$\bar{\psi}_{..} = \frac{1}{m}\sum_{j=1}^{m}\bar{\psi}_{.j}, \tag{A.41}$$

and

$$s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\psi_{ij} - \bar{\psi}_{.j})^2. \tag{A.42}$$

The basic idea is that the statistic looks at the difference between the inter- and intra-chain variances and checks that they are both similar in size, this is summarised by the potential scale reduction factor $\hat{R}$ (PSRF)

$$\hat{R} = \sqrt{\frac{1}{W}\left(\frac{n}{n-1}W + \frac{1}{n}B\right)}. \tag{A.43}$$

A.8.1 Overview of Gaussian Mixture Models

Considering the data we have from the fits it can be assumed that our data should be well described by a set of 2D Gaussians and so we shall use this as our underlying assumption[3]. The most important quantity to consider is the likelihood function and so the likelihood of a chosen data point x_i for a Gaussian mixture model is

$$p(x_i|\theta) = \sum_{j=1}^{M} c_j \mathcal{N}(\mu_j, \sigma_j), \tag{A.44}$$

where we have used $\mathcal{N}(\mu_j, \sigma_j)$ to denote a Normal distribution with mean μ_j and variance σ_j^2, c_j is a normalisation constant for the distributions subject to the constraint

[3]Although it should be noted that the more data we have, the more structure there is likely to be. However in our case this assumption is upheld to a good degree.

$$\sum_{j=1}^{M} c_j = 1. \tag{A.45}$$

In this formulation it is assumed that the data have uncertainties that are negligible with respect to the smallest standard deviation of the Gaussian mixture components, which given the small uncertainties present is an adequate assumption. The likelihood function for the entire dataset is given by

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{N} \left[\sum_{j=1}^{M} c_j \mathcal{N}(\mu_j, \sigma_j) \right], \tag{A.46}$$

and so the log-likelihood is given by

$$\ln \mathcal{L} = \sum_{i=1}^{N} \ln \left[\sum_{j=1}^{M} c_j \mathcal{N}(\mu_j, \sigma_j) \right]. \tag{A.47}$$

It is this equation that can be used with the EM algorithm to determine the best-fit parameters θ. This can easily be extended to n dimensions by considering μ as an n-dimensional vector and σ becomes an $n \times n$ covariance matrix Σ.

There is also the question of how many Gaussians do we need in the Mixture, this is decided *a priori* due to the nature of the algorithm and so in order to choose the best model the Bayesian Information Criterion (BIC) is used (since this is the method used in the `scikit-learn` implementation) which is simply defined as follows

$$\mathrm{BIC} \equiv -2 \ln \mathcal{L}^0 + k \ln M, \tag{A.48}$$

where $\ln \mathcal{L}^0$ is the maximum of the log-likelihood, k is the number of model parameters and the number of data points is given by M. Note how BIC naturally incorporates "Occam's Razor" by penalising models with more parameters, and so the model with the smallest BIC is taken to be the best choice and that is how we choose the number of Gaussians in the

mixture model up to a maximum value of 20.

A.9 Derivation of Mode Visibilities

The visibility of a mode of a given degree, ℓ, can be given by (Dziembowski, 1977; Gizon & Solanki, 2003; Ballot et al., 2011)

$$V_\ell = \sqrt{(2\ell+1)\pi}\int_0^1 P_\ell(\mu)W(\mu)\mu \mathrm{d}\mu, \tag{A.49}$$

where $W(\mu)$ is a weighting function which is dependent upon the centre-to-limb distance of the star, μ, as well as the instrument and P_ℓ is the Legendre Polynomial of degree ℓ. As opposed to working with the visibility it is easier to instead compute the relative visibility, defined as the ratio of the visibility of a mode of degree ℓ to the radial mode visibility

$$\tilde{V}^2_{\mathrm{tot}} = \sum_{\ell=0}^{\infty}\frac{V_\ell^2}{V_{\ell=0}^2}. \tag{A.50}$$

Although the above sum is taken to ∞ for the purposes of these simulations we are only interested in $\ell = 0, 1, 2$ and so the contribution from modes with degree greater than 2 are assumed to be negligible with respect to those below, in other words

$$\tilde{V}^2_{\mathrm{tot}} \approx \sum_{\ell=0}^{2}\frac{V_\ell^2}{V_{\ell=0}^2}. \tag{A.51}$$

For *Kepler* $\tilde{V}^2_{\mathrm{tot}}$ can be approximated to 3.16 (Ballot et al., 2011), whereas for TESS due to the fact that it has a different bandpass (as shown in Fig 4.4) $\tilde{V}^2_{\mathrm{tot}} \approx 2.94$ (see Campante et al. 2016b for more details). This is because of the impact of the different bandpasses on the weighting function, $W(\mu)$, which in turn also depends on properties such as limb-darkening that depend upon the bandpass in which the star is observed. For example, the effect of limb-darkening is greatly reduced when the bandpass is redder due to the inverse wavelength dependence of the temperature gradient of the spectral radiance. The differences between

the *Kepler* and TESS bandpasses can be seen in Fig 4.4 and the extent to which the TESS bandpass extends into the red can be clearly seen. The reason for the shift towards the red for TESS is due to its primary purpose of finding planets around stars smaller than the Sun, i.e. M-dwarfs.

In photometry the oscillations are observed as variations in flux, and therefore temperature. Assuming that observations are made in a narrow wavelength band and the oscillations are a linear perturbation the relative flux variation can be shown to have the following proportionality

$$\frac{\delta F_\lambda}{F_\lambda} \propto \frac{1}{\lambda}. \tag{A.52}$$

This follows from the approximation of the star as a blackbody radiator and use of the Planck function (see Basu & Chaplin 2016 or Kjeldsen & Bedding 1995 for a more comprehensive overview). As a result it can be seen that over a narrow-wavelength range, the smaller the wavelength the larger the amplitude of the oscillations. For TESS this is very important since the bandpass extends further into the red than *Kepler* and so the oscillations will have a reduced amplitude compared to *Kepler* by a factor of ~ 0.85 (Campante et al., 2016b). Table 4.1 shows the mode visibility for *Kepler*, TESS and the spectroscopic telescope SONG (Tenerife node).

A.9.1 Derivation of Q

The parameter Q is defined as the ratio of the inertia of the non-radial mode (in our case the $\ell = 1$, I_1) relative to the radial mode (I_0) at the same frequency (ν_1).

$$Q = \frac{I_1}{I_0(\nu_1)}. \tag{A.53}$$

Since I_1 is the inertia of the mixed dipole mode it seems reasonable to assumed that it is

the combination of the mode inertia in the p- and g-mode cavity, such that

$$I_1 \approx (I_1)_{\rm g} + (I_1)_{\rm p}, \tag{A.54}$$

and if we also assume that the radial mode inertia is equivalent to the inertia in the p-mode cavity then equation A.53 can be rewritten as

$$Q \approx \frac{(I_1)_{\rm g} + (I_1)_{\rm p}}{(I_1)_{\rm p}}. \tag{A.55}$$

A.10 The Hierarchical Process

A.11 Hierarchical process

The place to start with most Bayesian methods is the likelihood function and prior distributions that enable the evaluation of the target posterior distribution. The likelihood function provided by Hogg et al. (2010) is certainly an ideal place to start and so let us continue by deriving this and explaining the inference it allows us to make.

Consider the situation where for each star, labelled with subscript n, there are K-samples from a posterior probability density function (PDF) under the assumption of a given (uninformative) prior $p_0(\theta_n)$, where θ are the parameters of the model fitted to the data[4] (in our case this would take the form of inclination angle PDFs for n stars). This will take the form of

$$p(\theta_n|\mathbf{y}_n) \propto p_0(\theta_n)p(\mathbf{y}_n|\theta_n), \tag{A.56}$$

where $\mathbf{y}_n$ denotes the data for each star n and the same is applied to the parameters θ.

[4]It is important for the prior to be uninformative, or at the very least weakly informative, over the parameter space being explored, otherwise there is a risk that the posterior being inferred extends beyond the range of the prior.

This provides a framework for evaluating the likelihood of individual stars and not the sample as a whole. Under the assumption that there are no covariances between the parameters, the total likelihood for all the n stars can be written as

$$\mathcal{L} \equiv p(\mathbf{y}_n|\theta_n) = \prod_{n=1}^{N} \mathcal{L}_n. \tag{A.57}$$

Making the implicit assumption of independence between the stars in the sample assumes that any underlying instrumental effects (that could cause some degree of correlation between all the observations) are taken to be negligible. This assumption is generally upheld in the context of the stellar parameters themselves, it is only through instrumental noise that the observations could become correlated in some manner (such as intra-pixel variations on the CCD).

Possessing the total likelihood would be helpful if all the stars were fitted at the same time, through which the prior on the angle of inclination, i_s, could be modelled using some parametrised model on $\cos i_\mathrm{s}$, $f_\alpha(\cos i_\mathrm{s})$[5]. This is not currently computationally tractable and so the case where there are K posterior samples of i_s from each of the n stars in the sample will be considered along with the likelihood function under those conditions.

In this newly described case, the stellar parameters are considered nuisance parameters[6] and will be marginalised out, leaving behind only the parameters relating to the parametrised model for $\cos i_s$. The desired probability distribution, $p(\mathbf{y}_n|\alpha)$ (which in our case would be the probability of the measured inclination angle given the parameters describing the population distribution) is the likelihood of the data given the parameters of the model on $\cos i_\mathrm{s}$, and therefore a new "marginalised" likelihood can be formulated using the following equation, c.f. Eqn 7 of Hogg et al. (2010),

[5]The reason for fitting in $\cos i_s$ will become apparent later once its relationship to ψ has been introduced
[6]These are parameters in the model that have to be fitted but we do not care about their value.

$$p(\mathbf{y}_n|\alpha) = \prod_{n=1}^{N} \int_{\theta} p(\mathbf{y}_n|\theta_n) p(\theta_n|\alpha) \mathrm{d}\theta_n, \tag{A.58}$$

This marginalised likelihood is not ideal since it is computationally difficult N-dimensional integral, therefore we follow Hogg et al. (2010) and use the following likelihood function

$$p(y|\alpha) = \prod_{n=1}^{N} \frac{1}{K} \sum_{k=1}^{K} \frac{f_{\alpha}(i_{\mathrm{s},nk})}{p_0(i_{\mathrm{s},nk})}. \tag{A.59}$$

where $f_{\alpha}(i_{\mathrm{s},nk})$ is the model for the population distribution, N is the number of stars and K is the number of posterior samples in the initial data. The derivation of equation A.59 from equation A.58 is given in Appendix A.11.1.

A.11.1 Creation of the hierarchical model

The marginalised likelihood given in equation A.58 contains some features that makes its computation rather difficult. This mainly comes in the form of the integral which will be over the entire parameter space, so for the large parameter space involved in these fitting procedures a brute-force computation is prohibitively slow. Therefore an alternative method must be sought. This section describes the creation of the hierarchical model and a likelihood function that is much easier to compute.

The target posterior distribution for this inference task is

$$p(\alpha, \theta|y) \propto p(\alpha, \theta) p(y|\alpha, \theta), \tag{A.60}$$

where α denotes the hyperparameters describing the form of the prior distribution and θ are the parameters of the model to be fitted to the data. The above posterior is formed from the joint prior distribution $p(\alpha, \theta)$ and the likelihood function $p(y|\alpha, \theta)$. Equation A.60 can be simplified by expanding out the joint prior distribution to give

$$p(\alpha, \theta) = p(\alpha)p(\theta|\alpha), \tag{A.61}$$

which follows from the basic tenets of conditional probability and results in the following new form of equation A.60

$$p(\alpha, \theta|y) = p(\alpha)p(\theta|\alpha)p(y|\theta). \tag{A.62}$$

This should now start to look familiar to the marginal likelihood given in Equation A.58. It has been assumed that the likelihood function, $p(y|\alpha, \theta)$, only depends upon the original set of parameters θ. In other words, since α only affects y through θ there is no dependence of the likelihood function on α. Our target posterior for the purposes of this inference is $p(\alpha|y)$, the probability of the hyperparameters given the data, which is the posterior distribution of the hyperparameters of our parametrised model for the $\cos i_{\rm s}$ distribution. Now, the nuisance parameters need to be marginalised out (the parameters of the original fits to the data)

$$p(\alpha|y) = \int p(\theta, \alpha|y)\mathrm{d}\theta, \tag{A.63}$$

which, thanks to our previous simplifications, becomes

$$p(\alpha|y) = p(\alpha) \int p(\theta|\alpha)p(y|\theta)\mathrm{d}\theta. \tag{A.64}$$

This integral will be N-dimensional, where N corresponds to the number of parameters used in the original fitting. Evaluating this in a brute force manner would be far too computationally expensive and so taking advantage of the following trick (e.g. MacKay 1999) provides an importance resampled estimate

$$\int f(x)p(x)\mathrm{d}x = \int \frac{f(x)p(x)}{q(x)}q(x)\mathrm{d}x \approx \frac{1}{K}\sum_k \frac{f(x)^{(k)}p(x)^{(k)}}{q(x)^{(k)}}. \tag{A.65}$$

There are objections to the use of this approximation as it is effectively a Monte-Carlo approximation of the integral, which is a purely frequentist technique. Moreover this is due to the arbitrary choice of the importance sampling distribution $q(x)$ (O'Hagan, 1987), however $q(x) \propto p(\theta)p(y|\theta)$ can be chosen as our sampling distribution which is not arbitrary since it has already sampled in the original fitting (as a result there is no need to reproduce a full Bayesian Monte Carlo sampling e.g. Ghahramani & Rasmussen 2002). Therefore the distributions can be plugged into equation A.65 to give (using subscript 0 to denote quantities from the original fits)

$$p(\alpha|y) \propto p(\alpha) \int \frac{p(\theta|\alpha)p_0(y|\theta)}{p_0(y|\theta)p_0(\theta)} \mathrm{d}\theta, \tag{A.66}$$

and now we can make use of the summation approximation to give

$$p(\alpha|y) \propto p(\alpha) \prod_{n=1}^{N} \frac{1}{K} \sum_{k=1}^{K} \frac{p(\theta|\alpha)^{(k)}}{p_0(\theta)^{(k)}}. \tag{A.67}$$

It should become apparent that the hyperparameters of the model, α, depend solely upon the angle of inclination (assuming no covariances between parameters) and so $p_0(\theta)$ can be simplified to $p_0(i_\mathrm{s})$. Although not strictly true for solar-type stars[7] (in red giants this is much less of a problem), as suggested by Hogg et al. (2010) a multivariate model could instead be used for $f_\alpha(\cos i)$. This same idea can also be applied to $p(\theta|\alpha)$ whereby it only depends on the inclination angle and so becomes our model for the inclination angle distribution

$$p(\theta|\alpha) = f_\alpha(i_\mathrm{s}), \tag{A.68}$$

where $f_\alpha(i_\mathrm{s})$ is our model for the population distribution. Plugging the above into equation A.67 a final form for the marginal likelihood is obtained

[7]In which there is an observed covariance with rotational splitting e.g. Lund et al. (2014); Campante et al. (2016a)

$$p(y|\alpha) = \prod_{n=1}^{N} \frac{1}{K} \sum_{k=1}^{K} \frac{f_\alpha(i_{\mathrm{s},nk})}{p_0(i_{\mathrm{s},nk})}. \tag{A.69}$$

In order to estimate the posterior distribution of the hyperparameters given the data, the likelihood must be multiplied by a prior distribution for the hyperparameters and then this provides an approximation of the posterior distribution that can be sampled from

$$p(\alpha|y) \propto p(\alpha) \prod_{n=1}^{N} \frac{1}{K} \sum_{k=1}^{K} \frac{f_\alpha(i_{\mathrm{s},nk})}{p_0(i_{\mathrm{s},nk})}. \tag{A.70}$$

In this case a prior for the hyperparameters $p(\alpha)$ must be chosen, but given that the model for population distribution has yet to be defined this will be left till later in the chapter.

A.11.2 Deriving the Angle of Inclination using Asteroseismology

The mode visibility modulates the height of the mode according to ℓ, m and i_s, and so it is through the fitting of the mode amplitudes of the different components that the angle of inclination can be determined[8]. The visibility itself is given by (Gizon & Solanki, 2003)

$$\mathcal{E}_{\ell m}(i_\mathrm{s}) = \frac{(\ell - |m|)!}{(\ell + |m|)!} P_\ell^{|m|}(\cos i_\mathrm{s})^2, \tag{A.71}$$

subject to the condition

$$\sum_m \mathcal{E}_{\ell m}(i_\mathrm{s}) = 1, \tag{A.72}$$

where $P_\ell^{|m|}(\cos i_\mathrm{s})$ is a Legendre polynomial. Since the interest is in $\ell = 1$ modes the above expression can be simplified to

$$\mathcal{E}_{1,0}(i_\mathrm{s}) = \cos^2(i_\mathrm{s}), \tag{A.73}$$

[8]This comes about due to the fact that our observations are averaged on the disk.

and

$$\mathcal{E}_{1,\pm1}(i_{\rm s}) = \frac{1}{2}\sin^2(i)_{\rm s}. \tag{A.74}$$

A.11.3 Transforming the Model into $\cos i_{\rm s}$

The starting point is recognising that for an equation such as $Z = XY$, where $X = \sin\theta$ and $Y = \cos\phi$, the distribution of Z is given by (Rohatgi, 1976)

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z/x) \frac{1}{|x|} {\rm d}x. \tag{A.75}$$

However, both X and Y in the above equation are not defined in terms of probability distributions that are currently known. The distributions in terms of $\sin\psi$ and $\cos\phi$ are needed instead of those in ψ and ϕ. The question that needs to be asked is; given a probability distribution $p(x)$, where x is some random variable, what is the distribution of $p(y)$, where $y = f(x)$? For this, the following identity is needed (e.g. Ivezić et al. 2014)

$$p(y) = \sum_{k=1}^{n(y)} p\left[f_k^{-1}(y)\right] \left| \frac{{\rm d}f_k^{-1}(y)}{{\rm d}y} \right|, \tag{A.76}$$

where the subscript k denotes different solutions to the equation $y = f(x)$. In our case $y = \sin x$ or $y = \cos x$, therefore let us first start by deriving the PDF for the sine of a Fisher-distributed angle, where $y = \sin x$ and x is distributed according to equation 5.8. There are two solutions in the range 0 to π for $y = \sin x$ corresponding to $x = \arcsin y$ and $x = \pi - \arcsin y$ and so plugging these into equation A.76

$$f_{\sin\psi}(y|\kappa) = \frac{\kappa}{\sinh\kappa} \frac{y}{\sqrt{1-y^2}} \cosh(\kappa\sqrt{1-y^2}). \tag{A.77}$$

Following the same procedure for $\cos\phi$, where $f_\phi = 1/2\pi$,

$$f_{\cos\phi}(x) = \frac{2}{\pi\sqrt{1-x^2}}. \tag{A.78}$$

Both of the above distributions have been normalised over the range $(0, 1)$ rather than $(-1, 1)$ for the sake of simplicity. Plugging these identities into equation A.76 the ideal model for the hierarchical analysis is obtained

$$f_{\cos i_s}(z|\kappa) = \frac{2\kappa}{\pi \sinh\kappa} \int_z^1 \frac{\cosh\left(\kappa\sqrt{1-y^2}\right)}{\sqrt{1-y^2}\sqrt{1-(z/y)^2}} \mathrm{d}y. \tag{A.79}$$

As detailed in Morton & Winn (2014) the integral is defined from z to 1 as this is how it is bounded due to the ranges of both X and Y.

A.11.4 Limiting Cases of Angle of Inclination

Let us begin with the equation for the mode heights, given simply as

$$H_{\ell,m} = \mathcal{E}_{\ell,m}(i)H, \tag{A.80}$$

where $\mathcal{E}_{\ell,m}(i)$ is the mode visibility, H is the mode height and $H_{\ell,m}$ is the height of each component. In the conservative sense a mode component can be considered no longer detectable if its height is lower than the background. In reality this limit would be determined by deriving the mode detection probability for a given star. This is not taken into account and only a rough estimate is used to show how the signal-to-noise can affect the angle of inclination retrieved. Therefore the conservative equality is given by

$$\mathcal{E}_{\ell,m}(i)H \geq B, \tag{A.81}$$

for a mode component to be detectable. This does not take into account any detection methods and is used as a rough estimate to show how the signal-to-noise can affect the angle

of inclination retrieved. Therefore the limiting angle, $i_{\rm limit}$, is given by

$$\mathcal{E}_{\ell,m}(i_{\rm limit})H = B. \tag{A.82}$$

This can be split up into two separate cases, firstly applied to the outer mode components which defines a lower limit on the angle of inclination and secondly applied to the central component which defines an upper limit. These two cases give

$$B = \begin{cases} \frac{1}{2}\sin^2(i_{\rm limit})H & \text{for } m = \pm 1, \\ \cos^2(i_{\rm limit})H & \text{for } m = 0, \end{cases} \tag{A.83}$$

where the case of $m = \pm 1$ described the limiting case of the outer components tending to the background in height and $m = 0$ is for the central component.

A.12 Mode Detection

Simply identifying the mixed modes by eye is not robust enough to decide whether or not that feature seen in the power spectrum is actually a mixed mode given the human brain's extraordinary ability to see patterns where there are none. In order to provide some degree of mode detection we followed the work of Appourchaux et al. (2012).

Firstly we try and use the above method of mode identification to provide the best fit of the mixed mode asymptotic expression by eye to the data. The power spectrum is then smoothed over n bins where n can be decided by the user but for our purposes is set to a default value of 12 bins (for use with *Kepler* data, corresponding to ~ 0.1 μHz for 4 years of data). The reason behind performing the smoothing is to try and maximise the power contained within modes spread over many bins. This is valid in the context of p-dominated modes and some more g-dominated modes that have linewidths larger than a bin width, however in the context of the unresolved modes this will smooth out the power contained

in the single bin and make them more difficult to detect. But given the susceptibility of unresolved modes to being affected by noise means that there is not guarantee they could be detected even if present[9].

Rather than simply picking one single value to smooth over and use for the detection test we choose to smooth over a range of values. This is to try to alleviate the issue of the presence of modes with power in only a few bins. As a result the number of smoothed realisations is given as an input and the number of bins smoothed over for each realisation is taken from 1 to n in steps of 2.

Once the smoothed realisations have been made the detection test can be applied to the data and consists of a Frequentist approach whereby we choose a false alarm probability of 0.1 over a given window of $\Delta\nu/2$. This results in there being $\Delta\nu/(2n\nu_{\rm bw})$ independent bins, where n is the number of bins smoothed over for each realisation. Our detection probability in a single bin is therefore given by

$$p_{\rm fap} = \frac{p_{\rm fap}^{\rm win}}{N_{\rm ind}}, \tag{A.84}$$

where $N_{\rm ind}$ is the number of independent bins defined above and $p_{\rm fap}^{\rm win}$ is the false alarm probability in the given window, for which we have set a value of 0.1.

Computing the false alarm probability is not the final step in this procedure as we really want the power level that corresponds to our false alarm probability as this is the observable that we can tie directly back to the power spectrum. This is done using the following equation

$$p_{\rm fap} = \frac{1}{\Gamma(n)} \int_{x_{\rm fap}}^{+\infty} u^{n-1} e^{-u} {\rm d}u, \tag{A.85}$$

where u is a dummy symbol and Γ is the gamma function. But rather than calculating $p_{\rm fap}$ from $x_{\rm fap}$ we want to invert the equation and calculate $x_{\rm fap}$ for a given $p_{\rm fap}$ value. From

[9]It will be explained later as to why this does not matter for unresolved modes due to the properties required of the mode in order to reliably extract their angle of inclination.

which we then select the bins in the given window that lie above $x_{\rm fap}$ for each realisation due to smoothing. This helps build up a set of points where power has been detected above the desired threshold.

In order to identify the individual mode frequencies we make use of the clustering algorithm DBSCAN (Ester et al., 1996) (Density-based spatial clustering of applications with noise) which is a density-based clustering algorithm that also enables us to identify points that could be noise (Davies, 2015b). However just giving the algorithm a 1-D set of data is not enough and in order to transform the data into 2 dimensions we gave each frequency a small y-value drawn from a Normal distribution with mean zero and standard deviation of 0.01. The application of DBSCAN also requires the setting of two parameters, the minimum number of points to form a cluster and the maximum distance from a point within the cluster. These parameters were set such that the method worked well for the majority of stars but was also tweaked to account for differing stellar parameters, such as a lower $\Delta\nu$ value which would result in the frequencies being closer together and so not separated into the appropriate clusters but instead grouped together. Viewing the results of the algorithm by eye enables the suitability of the chosen parameters to be assessed.

There are some alternative methods that could be adopted such as binning all of the frequency values of the detected peaks into a histogram and fitting them with a Gaussian mixture model. However in this case a large number of smoothed realisations would be needed to ensure that the peaks seen in the histogram would tend towards being Normally-distributed.

A.13 What does lack of secondary eclipse tell us?

In the lightcurve of KOI-3890 no secondary eclipse is seen and so it is necessary to assess whether this is due to the architecture of the system or due to the size of the eclipse. Firstly we shall look at the architecture of the system by computing the impact parameter of the

secondary eclipse, following Winn (2010)

$$b_{\rm sec} = \frac{a\cos i_{\rm s}}{R_\star}\left(\frac{1-e^2}{1-e\sin\omega}\right). \tag{A.86}$$

If we plug in the parameters obtained in the transit fitting process then we get $b_{\rm sec} = 2.75$. The eclipse would be expected to be observed if $b \lessapprox 1$, but given the large value of $b_{\rm sec}$ we would not expect to observe the secondary eclipse.

As a sanity check we can compute the expected secondary eclipse depth to ensure that it should be detectable if the impact parameter was within the correct range.

Let us first start by defining the flux from the system both in and out of (secondary) eclipse

$$F_{\rm out} \propto R_1^2 T_1^4 + R_2^2 T_2^4, \tag{A.87}$$

$$F_{\rm in} \propto R_1^2 T_1^4, \tag{A.88}$$

where subscripts 1 and 2 denote the primary and secondary respectively. The fractional depth can then be calculated as

$$\delta_{\rm s} = \frac{F_{\rm out} - F_{\rm in}}{F_{\rm out}}, \tag{A.89}$$

which using the quantities in equations A.87 and A.88 gives

$$\delta_{\rm s} \approx \frac{R_1 T_1^4 + R_2^2 T_2^4 - R_1 T_1^4}{R_1^2 T_1^4 + R_2^2 T_2^4}. \tag{A.90}$$

Plugging in the derived quantities for the primary and an assumed effective temperature of $3000K$ and radius of 0.3 $R_\odot$ for the secondary, the predicted depth of the secondary eclipse is $\delta_{\rm s} \approx 390$ppm.

An upper limit on the secondary eclipse can be derived from the timeseries in the following manner: firstly the root-mean-square (rms) scatter, measured to be 388 ppm in the real data,

of the out-of-eclipse portion of the timeseries can be extracted to approximate the noise level in the data. Then to retrieve the noise level over the length-scale of the transit ($\sim$ 1 day) the value is reduced by a factor of $\sqrt{N}$, which for *Kepler* $N = 48$. In order to constitute a detection it is common to use 10-σ above the noise (North et al. 2017 and references therein) which gives a minimum threshold of 560 ppm. We must also not forget that multiple eclipses have been observed which will again (assuming the noise is white) reduce the noise level by a factor $\sqrt{N_{\text{eclipses}}}$. In this case 8 eclipses were observed out of a total 10 due to the presence of gaps where the eclipses would have been. If we assume that the full 10 secondary eclipses would be present in the data then the final minimum threshold becomes 177 ppm, which is well below the predicted secondary eclipse depth. In reality the non-white nature of the noise will affect this prediction and result in a threshold that lies above the predicted value, however this does provide a minimum threshold for the secondary eclipse.

References

Abt H. A., 2001, AJ, 122, 2008

Aerts C., Christensen-Dalsgaard J., Kurtz D. W., 2010, Asteroseismology

Ahlers J. P., Seubert S. A., Barnes J. W., 2014, ApJ, 786, 131

Ahlers J. P., Barnes J. W., Barnes R., 2015, ApJ, 814, 67

Aigrain S., Parviainen H., Pope B. J. S., 2016, MNRAS, 459, 2408

Akeson R. L., et al., 2013, PASP, 125, 989

Albrecht S., Reffert S., Snellen I., Quirrenbach A., Mitchell D. S., 2007, A&A, 474, 565

Albrecht S., Reffert S., Snellen I. A. G., Winn J. N., 2009, Nature, 461, 373

Albrecht S., Winn J. N., Carter J. A., Snellen I. A. G., de Mooij E. J. W., 2011, ApJ, 726, 68

Albrecht S., Setiawan J., Torres G., Fabrycky D. C., Winn J. N., 2013a, ApJ, 767, 32

Albrecht S., Winn J. N., Marcy G. W., Howard A. W., Isaacson H., Johnson J. A., 2013b, ApJ, 771, 11

Albrecht S., et al., 2014, ApJ, 785, 83

Anderson E. R., Duvall Jr. T. L., Jefferies S. M., 1990, ApJ, 364, 699

Angus R., Kipping D. M., 2016, ApJ, 823, L9

Appourchaux T., 2003a, A&A, 412, 903

Appourchaux T., 2003b, in Sawaya-Lacoste H., ed., ESA Special Publication Vol. 517, GONG+ 2002. Local and Global Helioseismology: the Present and Future. pp 131–138

Appourchaux T., 2004, A&A, 428, 1039

Appourchaux T., 2008, Astronomische Nachrichten, 329, 485

Appourchaux T., 2014, A crash course on data analysis in asteroseismology. p. 123

Appourchaux T., Samadi R., Dupret M.-A., 2009, A&A, 506, 1

Appourchaux T., et al., 2010, A&A, 18, 197

Appourchaux T., et al., 2012, A&A, 543, A54

Appourchaux T., et al., 2014, A&A, 566, A20

Arentoft T., et al., 2008, ApJ, 687, 1180

Baglin A., et al., 2006, in 36th COSPAR Scientific Assembly.

Bahcall J. N., 1972, Comments Nucl. Part. Phys., Vol. 5, p. 59 - 64, 5, 59

Ballard S., et al., 2014, ApJ, 790, 12

Ballot J., García R. A., Lambert P., 2006, MNRAS, 369, 1281

Ballot J., Appourchaux T., Toutain T., Guittet M., 2008, A&A, 486, 867

Ballot J., Barban C., van't Veer-Menneret C., 2011, A&A, 531, A124

Baluev R. V., 2008, MNRAS, 385, 1279

Barnes J. W., 2009, ApJ, 705, 683

Barnes J. W., Linscott E., Shporer A., 2011, ApJS, 197, 10

Barnes J. W., van Eyken J. C., Jackson B. K., Ciardi D. R., Fortney J. J., 2013, ApJ, 774, 53

Basri G., et al., 2010, ApJ, 713, L155

Basri G., et al., 2011, AJ, 141, 20

Basu S., Chaplin W. J., 2016, Asteroseismic Data Analysis: Foundations and Techniques. Princeton University Press,

Basu S., Chaplin W. J., Elsworth Y., 2010, ApJ, 710, 1596

Bate M. R., Lodato G., Pringle J. E., 2010, MNRAS, 401, 1505

Baudin F., et al., 2011, A&A, 529, A84

Bazot M., Bouchy F., Kjeldsen H., Charpinet S., Laymand M., Vauclair S., 2007, A&A, 470, 295

Bedding T. R., 2014, Solar-like oscillations: An observational perspective. p. 60

Bedding T. R., Butler R. P., McCarthy C., Kjeldsen H., Marcy G. W., O'Toole S. J., Tinney C. G., Wright J., 2003, in Thompson M. J., Cunha M. S., Monteiro M. J. P. F. G., eds, Vol. 1, Asteroseismology Across the HR Diagram. pp 303–306 (arXiv:astro-ph/0209359)

Bedding T. R., et al., 2010, ApJ, 713, 935

Belkacem K., Samadi R., Goupil M. J., Dupret M. A., Brun A. S., Baudin F., 2009, A&A, 494, 191

Belkacem K., Samadi R., Goupil M. J., 2011a, Journal of Physics: Conference Series, 271, 012047

Belkacem K., Goupil M. J., Dupret M. A., Samadi R., Baudin F., Noels A., Mosser B., 2011b, A&A, 530, A142

Belkacem K., Dupret M. A., Baudin F., Appourchaux T., Marques J. P., Samadi R., 2012, A&A, 540, L7

Belkacem K., Samadi R., Mosser B., Goupil M.-J., Ludwig H.-G., 2013, in Shibahashi H., Lynas-Gray A. E., eds, Astronomical Society of the Pacific Conference Series Vol. 479, Progress in Physics of the Sun and Stars: A New Era in Helio- and Asteroseismology. p. 61 (arXiv:1307.3132)

Benomar O., Masuda K., Shibahashi H., Suto Y., 2014, PASJ, 66, 94

Bi J., Zhang T., 2004, in NIPS.

Bishop C. M., 2006, Pattern recognition and machine learning

Blanton M. R., et al., 2003, ApJ, 592, 819

Bonnell I., Arcoragi J.-P., Martel H., Bastien P., 1992, ApJ, 400, 579

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

Bossini D., 2015, personal communication

Bourrier V., et al., 2015, A&A, 579, A55

Bourrier V., Cegla H. M., Lovis C., Wyttenbach A., 2016, preprint, (arXiv:1611.07985)

Bovy J., Hogg D. W., Roweis S. T., 2009, ApJ, 700, 1794

Brogaard K., et al., 2016, Astronomische Nachrichten, 337, 793

Brookes J. R., Isaak G. R., van der Raay H. B., 1978, MNRAS, 185, 1

Brooks S. P., Gelman A., 1998, Journal of Computational and Graphical Statistics, 7, 434

Broomhall A.-M., Chaplin W. J., Elsworth Y., Simoniello R., 2012, MNRAS, 420, 1405

Broomhall A.-M., et al., 2014, MNRAS, 440, 1828

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

Brown T. M., Gilliland R. L., Noyes R. W., Ramsey L. W., 1991, ApJ, 368, 599

Buchhave L. A., et al., 2012, Nature, 486, 375

Butterworth S., 1930, Wireless Engineer, 7

Campante T. L., 2012, PhD thesis, PhDT Thesis, Universidade do Porto

Campante T. L., Karoff C., Chaplin W. J., Elsworth Y. P., Handberg R., Hekker S., 2010, MNRAS, 408, 542

Campante T. L., et al., 2014, ApJ, 783, 123

Campante T. L., et al., 2016a, ApJ, 819, 85

Campante T. L., et al., 2016b, ApJ, 830, 138

Cantiello M., Fuller J., Bildsten L., 2016, ApJ, 824, 14

Chaplin W. J., Miglio A., 2013, ARA&A, 51, 353

Chaplin W. J., et al., 1996a, Sol. Phys., 168, 1

Chaplin W. J., Elsworth Y., Howe R., Isaak G. R., McLeod C. P., Miller B. A., New R., 1996b, MNRAS, 280, 1162

Chaplin W. J., Elsworth Y., Isaak G. R., McLeod C. P., Miller B. A., New R., 1997, MNRAS, 288, 623

Chaplin W. J., Elsworth Y., Isaak G. R., Marchenkov K. I., Miller B. A., New R., Pinter B., Appourchaux T., 2002, MNRAS, 336, 979

Chaplin W. J., et al., 2006, MNRAS, 369, 985

Chaplin W. J., et al., 2008, Astronomische Nachrichten, 329, 549

Chaplin W. J., Houdek G., Karoff C., Elsworth Y., New R., 2009, A&A, 500, L21

Chaplin W. J., et al., 2011a, Science, 332, 213

Chaplin W. J., et al., 2011b, ApJ, 732, 54

Chaplin W. J., et al., 2013, ApJ, 766, 101

Chaplin W. J., et al., 2014a, ApJS, 210, 1

Chaplin W. J., Elsworth Y., Davies G. R., Campante T. L., Handberg R., Miglio A., Basu S., 2014b, MNRAS, 445, 946

Chen T., Guestrin C., 2016, preprint, (arXiv:1603.02754)

Christensen-Dalsgaard J., 2002, Reviews of Modern Physics, 74, 1073

Christensen-Dalsgaard J., 2015, MNRAS, 453, 666

Christensen-Dalsgaard J., et al., 1996, Science, 272, 1286

Claverie A., Isaak G. R., McLeod C. P., van der Raay H. B., Cortes T. R., 1979, Nature, 282, 591

Collier Cameron A., Bruce V. A., Miller G. R. M., Triaud A. H. M. J., Queloz D., 2010a, MNRAS, 403, 151

Collier Cameron A., et al., 2010b, MNRAS, 407, 507

Cooper P., 1969, Solar Energy, 12, 333

Corsaro E., et al., 2012, ApJ, 757, 190

Corsaro E., Fröhlich H.-E., Bonanno A., Huber D., Bedding T. R., Benomar O., De Ridder J., Stello D., 2013, MNRAS, 430, 2313

Corsaro E., et al., 2017, Nature Astronomy, 1, 0064

Coughlin J. L., et al., 2014, AJ, 147, 119

Cumming A., Marcy G. W., Butler R. P., 1999, ApJ, 526, 890

Davies G. R., 2015a, personal communication

Davies G. R., 2015b, personal communication

Davies G. R., Miglio A., 2016, preprint, (arXiv:1601.02802)

Davies G. R., Broomhall A. M., Chaplin W. J., Elsworth Y., Hale S. J., 2014a, MNRAS, 439, 2025

Davies G. R., Chaplin W. J., Elsworth Y., Hale S. J., 2014b, MNRAS, 441, 3009

De Ridder J., Arentoft T., Kjeldsen H., 2006, MNRAS, 365, 595

Deheuvels S., et al., 2012, ApJ, 756, 19

Deheuvels S., Ballot J., Beck P. G., Mosser B., Østensen R., García R. A., Goupil M. J., 2015, A&A, 580, A96

Deheuvels S., Ouazzani R.-M., Basu S., 2017, preprint, (arXiv:1705.10326)

Désert J.-M., et al., 2011, ApJS, 197, 14

Deubner F.-L., 1975, A&A, 44, 371

Di Mauro M. P., et al., 2016, ApJ, 817, 65

Dumusque X., 2014, ApJ, 796, 133

Dumusque X., et al., 2015, ApJ, 814, L21

Duvall Jr. T. L., Harvey J. W., 1986, in Gough D. O., ed., NATO Advanced Science Institutes (ASI) Series C Vol. 169, NATO Advanced Science Institutes (ASI) Series C. pp 105–116

Dziembowski W., 1977, Acta Astron., 27, 203

Elsworth Y., Howe R., Isaak G. R., McLeod C. P., New R., 1990, Nature, 347, 536

Elsworth Y., Howe R., Isaak G. R., McLeod C. P., Miller B. A., New R., Wheeler S. J., 1995, A&AS, 113, 379

Elsworth Y., Hekker S., Basu S., Davies G., 2016, preprint, (arXiv:1612.04751)

Ester M., Kriegel H.-P., Sander J., Xu X., 1996. AAAI Press, pp 226–231

Fabrycky D. C., Winn J. N., 2009, ApJ, 696, 1230

Fierry Fraillon D., Gelly B., Schmider F. X., Hill F., Fossat E., Pantel A., 1998, A&A, 333, 362

Fischler M. A., Bolles R. C., 1981, Commun. ACM, 24, 381

Fletcher S. T., Chaplin W. J., Elsworth Y., Schou J., Buzasi D., 2006, MNRAS, 371, 935

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Foreman-Mackey D., Hogg D. W., Morton T. D., 2014, ApJ, 795, 64

Frandsen S., et al., 2013, A&A, 556, A138

Fuller J., Cantiello M., Stello D., Garcia R. A., Bildsten L., 2015, Science, 350, 423

Gabriel A. H., et al., 1995, Sol. Phys., 162, 61

Gai N., Basu S., Chaplin W. J., Elsworth Y., 2011, ApJ, 730, 63

Gamerman D., Lopes H. F., 2006, Markov chain Monte Carlo: stochastic simulation for Bayesian inference. CRC Press

Gandolfi D., et al., 2012, A&A, 543, L5

García R. A., Turck-Chièze S., Jiménez-Reyes S. J., Ballot J., Pallé P. L., Eff-Darwich A., Mathur S., Provost J., 2007, Science, 316, 1591

García R. A., et al., 2014, A&A, 563, A84

Gaudi B. S., Winn J. N., 2007, ApJ, 655, 550

Gaulme P., McKeever J., Rawls M. L., Jackiewicz J., Mosser B., Guzik J. A., 2013, ApJ, 767, 82

Gaulme P., et al., 2016, preprint, (arXiv:1609.06645)

Gelman A., Jakulin A., Grazia Pittau M., Su Y.-S., 2009, preprint, (arXiv:0901.4011)

Gelman A., Carlin J., Stern H., Dunson D., Vehtari A., Rubin D., 2013, Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis, https://books.google.co.uk/books?id=ZXL6AQAAQBAJ

Geweke J., 1992, in IN BAYESIAN STATISTICS. University Press, pp 169–193

Ghahramani Z., Rasmussen C. E., 2002, in Advances in neural information processing systems. pp 489–496

Gilliland R. L., Chaplin W. J., Jenkins J. M., Ramsey L. W., Smith J. C., 2015, AJ, 150, 133

Girardi L., Bressan A., Bertelli G., Chiosi C., 2000, A&AS, 141, 371

Gizon L., Solanki S. K., 2003, ApJ, 589, 1009

Goodman J., Weare J., 2010, Comm. App. Math. and Comp. Sci.

Goupil M. J., Mosser B., Marques J. P., Ouazzani R. M., Belkacem K., Lebreton Y., Samadi R., 2013, A&A, 549, A75

Groot P. J., 2012, ApJ, 745, 55

Grosjean M., Dupret M.-A., Belkacem K., Montalban J., Samadi R., Mosser B., 2014, A&A, 572, A11

Grundahl F., Arentoft T., Christensen-Dalsgaard J., Frandsen S., Kjeldsen H., Rasmussen P. K., 2008, in Journal of Physics Conference Series. p. 012041 (arXiv:0705.0851), doi:10.1088/1742-6596/118/1/012041

Hale S. J., Howe R., Chaplin W. J., Davies G. R., Elsworth Y. P., 2016, Sol. Phys., 291, 1

Handberg R., Campante T. L., 2011, A&A, 527, A56

Handberg R., Lund M. N., 2014, MNRAS, 445, 2698

Harvey J., 1985, in Rolfe E., Battrick B., eds, ESA Special Publication Vol. 235, Future Missions in Solar, Heliospheric & Space Plasma Physics.

Hastie T., Tibshirani R., Friedman J., 2001, The Elements of Statistical Learning. Springer Series in Statistics, Springer New York Inc., New York, NY, USA

Hayden M. R., et al., 2014, AJ, 147, 116

Hekker S., Christensen-Dalsgaard J., 2016, preprint, (arXiv:1609.07487)

Hekker S., Meléndez J., 2007, A&A, 475, 1003

Hekker S., et al., 2011, MNRAS, 414, 2594

Hirano T., Sanchis-Ojeda R., Takeda Y., Narita N., Winn J. N., Taruya A., Suto Y., 2012, ApJ, 756, 66

Hirano T., Sanchis-Ojeda R., Takeda Y., Winn J. N., Narita N., Takahashi Y. H., 2014, ApJ, 783, 9

Hogg D. W., Myers A. D., Bovy J., 2010, ApJ, 725, 2166

Houdek G., Balmforth N. J., Christensen-Dalsgaard J., Gough D. O., 1999, A&A, 351, 582

Houdek G., et al., 2001, MNRAS, 327, 483

Howell S. B., et al., 2014, PASP, 126, 398

Huber D., Stello D., Bedding T. R., Chaplin W. J., Arentoft T., Quirion P.-O., Kjeldsen H., 2009, Communications in Asteroseismology, 160, 74

Huber D., et al., 2010, ApJ, 723, 1607

Huber D., et al., 2011, ApJ, 743, 143

Huber D., et al., 2013a, Science, 342, 331

Huber D., et al., 2013b, ApJ, 767, 127

Huber D., et al., 2014, ApJS, 211, 2

Ivezić Z., Connolly A. J., VanderPlas J. T., Gray A., 2014, Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data, stu - student edition edn. Princeton University Press, http://www.jstor.org/stable/j.ctt4cgbdj

Jackson R. J., Jeffries R. D., 2010, MNRAS, 402, 1380

Jeffreys H., 1925, Proceedings of the London Mathematical Society, s2-23, 428

Jenkins J. M., et al., 2010a, ApJ, 713, L87

Jenkins J. M., et al., 2010b, ApJ, 713, L120

Jolliffe I., 1986. Springer Verlag

Kallinger T., et al., 2014, A&A, 570, A41

Karoff C., Campante T. L., Chaplin W. J., 2010, preprint, (arXiv:1003.4167)

Kass R. E., Raftery A. E., 1995, Journal of the American Statistical Association, 90, 773

Kiefer R., Schad A., Herzberg W., Roth M., 2015, A&A, 578, A56

Kjeldsen H., Bedding T. R., 1995, A&A, 293, 87

Kjeldsen H., et al., 2005, ApJ, 635, 1281

Kjeldsen H., et al., 2008, ApJ, 682, 1370

Kovács G., Zucker S., Mazeh T., 2002, A&A, 391, 369

Kramers H. A., 1926, Zeitschrift für Physik, 39, 828

Kreidberg L., 2015, PASP, 127, 1161

Lagarde N., Bossini D., Miglio A., Vrard M., Mosser B., 2016, MNRAS, 457, L59

Leavitt H. S., Pickering E. C., 1912, Harvard College Observatory Circular, 173, 1

Ledoux P., 1951, ApJ, 114, 373

Leighton R. B., Noyes R. W., Simon G. W., 1962, ApJ, 135, 474

Lillo-Box J., Barrado D., Bouy H., 2014, A&A, 566, A103

Lillo-Box J., Barrado D., Mancini L., Henning T., Figueira P., Ciceri S., Santos N., 2015, A&A, 576, A88

Lomb N. R., 1976, Ap&SS, 39, 447

Luger R., Agol E., Kruse E., Barnes R., Becker A., Foreman-Mackey D., Deming D., 2016, AJ, 152, 100

Lund M. N., 2016, private communication

Lund M. N., et al., 2014, A&A, 570, A54

Lund M. N., et al., 2016, preprint, (arXiv:1612.00436)

MacKay D. J. C., 1999, MIT Press, Cambridge, MA, USA, Chapt. Introduction to Monte Carlo Methods, pp 175–204, http://dl.acm.org/citation.cfm?id=308574.308666

Majewski S. R., Wilson J. C., Hearty F., Schiavon R. R., Skrutskie M. F., 2010, in Cunha K., Spite M., Barbuy B., eds, IAU Symposium Vol. 265, Chemical Abundances in the Universe: Connecting First Stars to Planets. pp 480–481, doi:10.1017/S1743921310001298

Masuda K., 2015, ApJ, 805, 28

Mathur S., Turck-Chièze S., Couvidat S., García R. A., 2007, ApJ, 668, 594

Mathur S., et al., 2011, ApJ, 741, 119

Mazeh T., Shaham J., 1979, A&A, 77, 145

McLaughlin D. B., 1924, ApJ, 60

Merline W. J., 1995, PhD thesis, THE UNIVERSITY OF ARIZONA.

Miglio A., et al., 2010, A&A, 520, L6

Miglio A., et al., 2013, MNRAS, 429, 423

Montalbán J., Miglio A., Noels A., Dupret M.-A., Scuflaire R., Ventura P., 2013, ApJ, 766, 118

Mortier A., Santos N. C., Sousa S. G., Adibekyan V. Z., Delgado Mena E., Tsantaki M., Israelian G., Mayor M., 2013, A&A, 557, A70

Morton T. D., Winn J. N., 2014, ApJ, 796, 47

Mosser B., et al., 2008, A&A, 478, 197

Mosser B., et al., 2011, A&A, 525, L9

Mosser B., et al., 2012a, A&A, 537, A30

Mosser B., et al., 2012b, A&A, 540, A143

Mosser B., et al., 2013, A&A, 550, A126

Mosser B., et al., 2014, A&A, 572, L5

Mosser B., Vrard M., Belkacem K., Deheuvels S., Goupil M. J., 2015, A&A, 584, A50

Mosser B., Pinçon C., Belkacem K., Takata M., Vrard M., 2016, preprint, (arXiv:1612.08453)

Mosser B., et al., 2017, A&A, 598, A62

Murphy S. J., 2015, MNRAS, 453, 2569

Murphy S. J., Shibahashi H., Kurtz D. W., 2013, MNRAS, 430, 2986

Nagasawa M., Ida S., Bessho T., 2008, ApJ, 678, 498

Neal R. M., 1996, Statistics and Computing, 6, 353

North T. S. H., et al., 2017, MNRAS, 465, 1308

Nyquist H., 1928, Transactions of the American Institute of Electrical Engineers, Volume 47, Issue 2, pp.~617-624, 47, 617

O'Hagan A., 1987, Journal of the Royal Statistical Society. Series D (The Statistician), 36, 247

Østensen R. H., et al., 2010, MNRAS, 408, L51

Pál A., et al., 2008, ApJ, 680, 1450

Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, ApJS, 192, 3

Paxton B., et al., 2013, ApJS, 208, 4

Paxton B., et al., 2015, ApJS, 220, 15

Pepper J., et al., 2016, preprint, (arXiv:1607.01755)

Pesnell W. D., 1985, ApJ, 292, 238

Pinsonneault M. H., et al., 2014, ApJS, 215, 19

Provost J., Berthomieu G., Morel P., 2000, A&A, 353, 775

Pérez Hernández, F. García, R. A. Corsaro, E. Triana, S. A. De Ridder, J. 2016, A&A, 591, A99

Raftery A. E., Lewis S. M., 1995, in In Practical Markov Chain Monte Carlo (W.R. Gilks, D.J. Spiegelhalter and. Chapman and Hall, pp 115–130

Rasio F. A., Ford E. B., 1996, Science, 274, 954

Rasmussen C. E., Williams C. K. I., 2006, Gaussian Processes for Machine Learning. MIT Press

Rauer H., et al., 2014, Experimental Astronomy, 38, 249

Retter A., Bedding T. R., Buzasi D., Kjeldsen H., 2003, in Thompson M. J., Cunha M. S., Monteiro M. J. P. F. G., eds, Vol. 1, Asteroseismology Across the HR Diagram. pp 383–386 (arXiv:astro-ph/0208518)

Reverdy P., Leonard N. E., 2015, preprint, (arXiv:1502.04635)

Ricker G. R., et al., 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Riello M., Piotto G., Recio-Blanco A., Cassisi S., Salaris M., 2003, in Piotto G., Meylan G., Djorgovski S. G., Riello M., eds, Astronomical Society of the Pacific Conference Series Vol. 296, New Horizons in Globular Cluster Astronomy. p. 232 (arXiv:astro-ph/0303295)

Rodrigues T. S., et al., 2014, MNRAS, 445, 2758

Rodrigues T. S., et al., 2017, MNRAS,

Rohatgi V. K., 1976, An introduction to probability theory and mathematical statistics. Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York [etc.]

Rossiter R. A., 1924, ApJ, 60

Rowe J. F., et al., 2015, ApJS, 217, 16

Samadi R., 2011, in Rozelot J.-P., Neiner C., eds, Lecture Notes in Physics, Berlin Springer Verlag Vol. 832, Lecture Notes in Physics, Berlin Springer Verlag. p. 305 (arXiv:0912.0817), doi:10.1007/978-3-642-19928-8_11

Samadi R., et al., 2010, A&A, 509, A16

Scargle J. D., 1982, ApJ, 263, 835

Schlaufman K. C., 2010, ApJ, 719, 602

Shannon C. E., 1949, IEEE Proceedings, 37, 10

Shibahashi H., 1979, PASJ, 31, 87

Shporer A., et al., 2011, AJ, 142, 195

Shporer A., Brown T., Mazeh T., Zucker S., 2012, New A, 17, 309

Silva Aguirre V., et al., 2015, MNRAS, 452, 2127

Sing D. K., 2010, A&A, 510, A21

Smith P. H., McMillan R. S., Merline W. J., 1987, ApJ, 317, L79

Smith J. C., et al., 2012, PASP, 124, 1000

Stello D., et al., 2009, ApJ, 700, 1589

Stello D., et al., 2013, ApJ, 765, L41

Stello D., Cantiello M., Fuller J., Garcia R. A., Huber D., 2016, PASA, 33, e011

Stumpe M. C., et al., 2012, PASP, 124, 985

Stumpe M. C., Smith J. C., Catanzarite J. H., Van Cleve J. E., Jenkins J. M., Twicken J. D., Girouard F. R., 2014a, PASP, 126, 100

Stumpe M. C., Smith J. C., Catanzarite J. H., Cleve J. E. V., Jenkins J. M., Twicken J. D., Girouard F. R., 2014b, Publications of the Astronomical Society of the Pacific, 126, 100

Sullivan P. W., et al., 2015, ApJ, 809, 77

Tarrant N. J., Chaplin W. J., Elsworth Y., Spreckley S. A., Stevens I. R., 2007, MNRAS, 382, L48

Tassoul M., 1980, ApJS, 43, 469

Thévenin F., Provost J., Morel P., Berthomieu G., Bouchy F., Carrier F., 2002, A&A, 392, L9

Ulrich R. K., 1986, ApJ, 306, L37

Unno W., Osaki Y., Ando H., Saio H., Shibahashi H., 1989, Nonradial oscillations of stars

Van Eylen V., et al., 2014, ApJ, 782, 14

Verma K., Raodeo K., Antia H. M., Mazumdar A., Basu S., Lund M. N., Silva Aguirre V., 2017, ApJ, 837, 47

Vousden W. D., Farr W. M., Mandel I., 2016, MNRAS, 455, 1919

Vrard M., et al., 2015, A&A, 579, A84

Vrard M., Mosser B., Samadi R., 2016, A&A, 588, A87

Welch P., 1967, IEEE Transactions on Audio and Electroacoustics, 15, 70

Wentzel G., 1926, Zeitschrift für Physik, 38, 518

White T. R., Bedding T. R., Stello D., Christensen-Dalsgaard J., Huber D., Kjeldsen H., 2011, ApJ, 743, 161

Winn J. N., 2007, in Afonso C., Weldrake D., Henning T., eds, Astronomical Society of the Pacific Conference Series Vol. 366, Transiting Extrapolar Planets Workshop. p. 170 (arXiv:astro-ph/0612744)

Winn J. N., 2010, preprint, (arXiv:1001.2010)

Winn J. N., Fabrycky D. C., 2015, ARA&A, 53, 409

Winn J. N., et al., 2006, ApJ, 653, L69

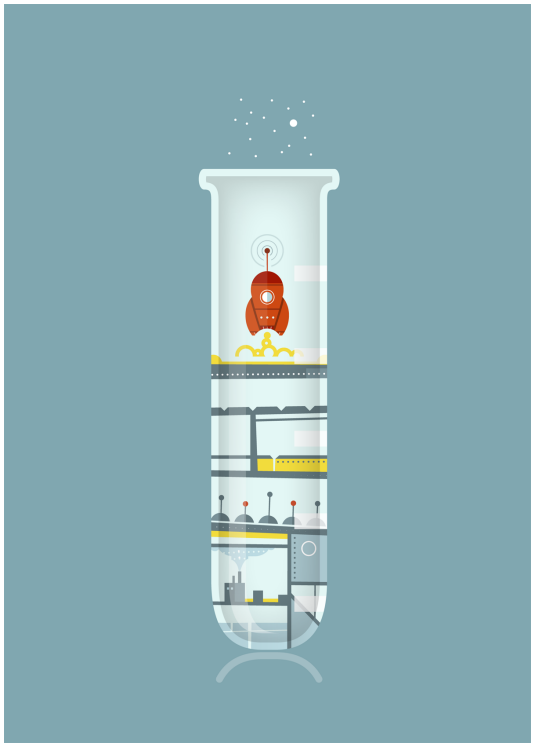
Winn J. N., Fabrycky D., Albrecht S., Johnson J. A., 2010, ApJ, 718, L145

Wittenmyer R. A., Endl M., Wang L., Johnson J. A., Tinney C. G., O'Toole S. J., 2011, ApJ, 743, 184

Wittenmyer R. A., Liu F., Wang L., Casagrande L., Johnson J. A., Tinney C. G., 2016, AJ, 152, 19

Zhou G., Latham D. W., Bieryla A., Beatty T. G., Buchhave L. A., Esquerdo G. A., Berlind P., Calkins M. L., 2016, MNRAS, 460, 3376

da Silva L., et al., 2006, A&A, 458, 609



This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, *Science Experiment 02*, was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD dissertation with this look *&* feel has been released under the permissive AGPL license, and can be found online at github.com/suchow/Dissertate or from its lead author, Jordan Suchow, at suchow@post.harvard.edu.