Ensemble asteroseismology of solar-type stars with the NASA *Kepler Mission*

W. J. Chaplin[1], H. Kjeldsen[2], J. Christensen-Dalsgaard[2], S. Basu[3], A. Miglio[4], T. Appourchaux[5], T. R. Bedding[6], Y. Elsworth[1], R. A. García[7], R. L. Gilliland[8], L. Girardi[9], G. Houdek[10], C. Karoff[2], S. D. Kawaler[11], T. S. Metcalfe[12], J. Molenda-Żakowicz[13], M. J. P. F. G. Monteiro[14], M. J. Thompson[12], G. A. Verner[15,1], J. Ballot[16], A. Bonanno[17], I. M. Brandão[14], A.-M. Broomhall[1], H. Bruntt[2,18], T. L. Campante[14,2], E. Corsaro[17], O. L. Creevey[19,20], G. Doğan[2], L. Esch[3], N. Gai[21,3], P. Gaulme[5], S. J. Hale[1], R. Handberg[2], S. Hekker[1], D. Huber[6], A. Jiménez[19,20], S. Mathur[12], A. Mazumdar[22], B. Mosser[23], R. New[24], M. H. Pinsonneault[25], D. Pricopi[26], P.-O. Quirion[27], C. Régulo[19,20], D. Salabert[19,20], A. M. Serenelli[28], V. Silva Aguirre[29], S. G. Sousa[14], D. Stello[6], I. R. Stevens[1], M. D. Suran[26], K. Uytterhoeven[7], T. R. White[6], W. J. Borucki[30], T. M. Brown[31], J. M. Jenkins[32], K. Kinemuchi[33], J. Van Cleve[32], T. C. Klaus[34]

The scientific investigation of solar-type stars has taken a major step forward thanks to the NASA *Kepler Mission* which, in addition to its search for exoplanets, is providing exquisite data on stellar oscillations. We report that the asteroseismic survey undertaken by *Kepler* has met with unprecedented success, yielding clear detections of oscillations in 500 solar-type stars. Distributions of the fundamental properties of these stars show intriguing differences to predictions from models of synthetic stellar populations in the Galaxy. This is important for population studies because the availability of mass estimates for all stars in the ensemble puts us in a position to test theories of stellar evolution to levels not previously possible.

We are entering a golden era for stellar physics, driven by new observations from the ground and space of exceptional quality and scope. An understanding of stars is of central importance to astrophysics. Uncertainties in stellar physics have a direct impact on fixing the ages of the oldest stellar populations (which place tight constraints on cosmologies) as well as on tracing the chemical evolution of galaxies. Stellar astrophysics also plays a crucial role in the current endeavors to detect habitable planets around other stars. In fact, ongoing developments in exoplanet research go hand-in-hand with detailed studies of the host stars (*1,2,3,4*). Accurate data on the host stars are required to determine the sizes of planets discovered by the transit method, to fix the locations of habitable zones around the stars, and to estimate the ages and understand the dynamical histories of these stellar systems. Measurements of the levels of stellar activity, and their variations over time (*5*), provide insights into planetary habitability, the completeness of the survey for extrasolar planets, and on the surface variability shown by our own Sun, which has very recently been in a quiescent state that is unique in the modern satellite era (*6,7*).

New insights are being made possible by asteroseismology, the study of stars by observations of their natural, resonant oscillations (*8,9*). Stellar oscillations are the visible manifestations of standing waves in the stellar interiors. Main-sequence and sub-giant stars whose outer layers are unstable to convection (solar-type stars) display “solar-like” oscillations that are predominantly acoustic in nature, excited by turbulence in the convective envelopes (*10,11*). The dominant oscillation periods are minutes in length and give rise to variations in stellar

brightness at levels of typically just a few parts per million. The frequencies of the oscillations depend on the internal structures of the stars and their rich information content means that the fundamental stellar properties (e.g., mass, radius, and age) can be determined to levels that are difficult to achieve by other means, while the internal structures and dynamics can be investigated in a unique way.

Helioseismology has provided us with an extremely detailed picture of the internal structure and dynamics of the Sun, including tests of basic physics (*12,13,14*). Such investigations are beginning to be possible for other stars. Over the last decade the quality of seismic observations on other solar-type stars has been improving steadily, from ground-based spectroscopy (*15,16,17*) and the French-led *CoRoT* satellite (*18,19*). However, it is the recent launch of the NASA *Kepler Mission* (*1*) that has realized a huge breakthrough in the amount and quality of data for the study of stars based on their oscillations (*2*). *Kepler* is providing ultra-precise observations of variations in stellar brightness (photometry), which are suitable for the study of solar-like oscillations (*20*). During the first seven months of science operations more than 2000 stars were selected for observation for one month each with a cadence rapid enough to perform an "asteroseismic survey" of the solar-type population in the *Kepler* field of view. Here, we present the first results of this survey.

The most important outcome for both the initial and the long-term exploitation of these data is the detection of solar-like oscillations in 500 stars. This represents a dramatic rise, from a pre-*Kepler* total of about 25, in the number of solar-type stars with oscillations detected. The data present a homogeneous seismic ensemble of unprecedented nature that is large enough to allow statistical studies of intrinsic stellar properties and trends for solar-type stars.

The oscillations data are exquisite. Figure 1 shows the frequency spectra of the oscillations exhibited by nine stars from the ensemble. As is evident from the figure, solar-like oscillators present a rich, near-regular pattern of peaks that are the signatures of high-order overtones. The dominant frequency spacing is the so-called large separation, $\Delta\nu$, between consecutive overtones (*21*). The average large separation scales with the square root of the mean density of the star. The observed power in the oscillations is modulated in frequency by an envelope that has an approximately Gaussian shape. The frequency of maximum oscillation power, ν_{max}, scales as $gT_{\rm eff}^{-1/2}$, where $g \propto M / R^2$ is the surface gravity and $T_{\rm eff}$ is the effective temperature of the star (*22,23*).

Figure 2 shows the stars on a conventional Hertzsprung-Russell diagram, which plots the luminosities of stars – calculated from the estimated radii (see below) and effective temperatures – against $T_{\rm eff}$. We also plot the large separation $\Delta\nu$ against temperature, and, just like the conventional diagram, this asteroseismic version delineates different types of stars and different evolutionary states (the ν_{max} version is similar). Main-sequence stars, burning hydrogen into helium in their cores, lie in a diagonal swathe (from the lower right to top left) on each diagram. Both asteroseismic parameters, $\Delta\nu$ and ν_{max}, decrease along the main sequence toward hotter solar-type stars, where surface gravities and mean densities are lower than in cooler stars (and luminosities are higher). After exhaustion of the core hydrogen, stars eventually follow nearly horizontal paths in the luminosity plot toward lower temperatures as they evolve as sub-giants, before turning sharply upwards to become red giants (*24,25*). The

values of $\Delta\nu$ and ν_{max} decrease comparatively rapidly in size through the sub-giant phase. It is evident from Figure 2 that *Kepler* has detected oscillations in solar-type stars at different evolutionary stages. As we collect more data on the stars, and are able to extract information on the individual oscillation modes, we will be able to test the physics of stars in these different evolutionary stages.

We have detected solar-like oscillations in relatively few stars that have $\Delta\nu$ and ν_{max} larger than the solar values. These stars are intrinsically fainter, and less massive, than the Sun, and we see fewer detections because the intrinsic oscillation amplitudes are lower than in the hotter main-sequence and evolved sub-giant stars. This detection bias means that the most populous cohort in the ensemble is that comprising sub-giants. In time we should get complete frequency spectra for many of these stars. Sub-giants have more complicated oscillation spectra than main-sequence stars. The details of the spectra depend on how, for example, various elements are mixed both within, and between, different layers inside the stars. Seismic analysis of the Sun has already shown that merely reproducing the luminosity and temperature of a star will not guarantee that the internal structure, and hence the underlying physics, is correct. This inspired the inclusion of additional physics, such as the settling over time of chemical elements due to gravity, in stellar models (*12*). The Sun is a relatively simple star compared to some of the solar-type stars observed by *Kepler*.

We have made use of the $\Delta\nu$ and ν_{max} of the ensemble together with photometric estimates of the temperatures to estimate the masses and radii of the stars in a way that is independent of stellar evolutionary models (*26*), and then compared the observed distributions with those predicted from synthetic stellar populations. The synthetic populations were calculated by modelling the formation and evolution of stars in the *Kepler* field of view, which lies in the Cygnus region of the Orion arm of our galaxy, the Milky Way (*26*). This modelling requires descriptions of, for example, the star formation history (including the frequency of occurrence of stars with various masses), the spatial density of stars in the disc of the Milky Way, and the rate at which the galaxy is chemically enriched by stellar evolution (*27*).

Figure 3 shows the observed distributions of masses and radii, together with the synthetic population. Previous population studies have been hampered by not having robust mass estimates on individual stars (*27*). Precise estimates of masses of solar-type stars had been limited principally to stars in eclipsing binaries (*28*). The *Kepler* estimates add significantly to this total, and in numbers that are large enough to do statistical population tests using direct mass estimates, which has not been possible before.

While the distributions of stellar radii in Figure 3 are similar, the same cannot be said for the mass distributions. We have quantified the significance of the differences using statistical tests. Differences in radius were judged to be marginally significant at best. In contrast those in mass were found to be highly significant (>99.99%) (*26*). The observed distribution of masses is wider at its peak than the modeled distribution, and is offset towards slightly lower masses. This result may have implications for both the star formation rate and the initial mass function of stars. Also relevant is mixing or overshooting of material between different layers (including stellar cores), and the choice of the so-called mixing length parameter, which measures the typical lengthscale of the convection and is one of the few free parameters in stellar evolution

theory. It is yet to be tested whether the expected small fraction of unresolved binaries could have contributed to the mass discrepancy.

A coordinated programme of ground-based spectroscopic observations in support of *Kepler* is underway. It is extremely important in that it will provide precise independent constraints on the compositions of these stars (*29*), allowing even tighter constraints to be placed on the estimated masses, and crucially the ages, of the *Kepler* stars. It is now possible to test in great detail the effects of different physics on the evolution of solar-type stars.

References and Notes

1. W. J. Borucki *et al.*, *Science* **327**, 977 (2010); *Science* **325**, 709 (2009).

2. R. L. Gilliland *et al.*, *Pub. Astron. Soc. Pac.* **122**, 131 (2010). The asteroseismology program of *Kepler* is being undertaken by the *Kepler Asteroseismology Science Consortium*.

3. D. Stello, et al., *et al.*, *Astrophys. J.* **700**, 1589 (2009)

4. J. Christensen-Dalsgaard, *et al.*, *Astrophys. J.* **713**, L164 (2010)

5. R. A. García, *et al.*, *Science* **329**, 1032 (2010).

6. D. H. Hathaway, L. Rightmire, *Science* **327**, 1035 (2010).

7. A.-M. Broomhall *et al.*, *Astrophys. J.* **700**, L162 (2009).

8. T. M. Brown, R. L. Gilliland, *Ann. Rev. Astron. & Astrophys.* **32**, 37 (1994).

9. J. Christensen-Dalsgaard, *Sol. Phys.* **220**, 137 (2004).

10. G. Houdek *et al.*, *Astron. & Astrophys.* **351**, 582 (1999)

11. R. Samadi, *et al.*, *Astron. & Astrophys.* **463**, 297 (2007)

12. D. O. Gough, et al., *Science* **272**, 1296 (1996).

13. J. Christensen-Dalsgaard, *Rev. Mod. Phys.* **74**, 1073 (2002).

14. S. V. Vorontsov, *et al., Nature,* **349**, 49 (1991).

15. C. Aerts, J. Christensen-Dalsgaard, M. Cunha, D. W. Kurtz, *Sol. Phys.* **251**, 3 (2008).

16. T. R. Bedding, H. Kjeldsen, in *14th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun* (ASP Conf. Ser., 2008), vol. 384, p. 21.

17. T. R. Bedding et al., *Astrophys. J.* **713**, 935 (2010).

18. T. Appourchaux *et al.*, *Astron. & Astrophys.* **488**, 705 (2008).

19. E. Michel *et al.*, *Science* **322**, 558 (2008).

20. W. J. Chaplin *et al. Astrophys. J.* **713**, L169 (2010)

21. Solar-type stars oscillate in non-radial modes. The radial part is described by the order (i.e., overtone number) *n*, while the surface pattern may be described in terms of a spherical harmonic function of degree, *l* and azimuthal order *m*. The large separation is the separation between consecutive overtones *n* of the same *l*.

22. T. M. Brown, R. L. Gilliland, R. W. Noyes, L. W. Ramsey, *Astrophys. J.* **368**, 599 (1991).

23. H. Kjeldsen, T. R. Bedding, *Astron. & Astrophys.* **293**, 87 (1995).

24. J. De Ridder *et al*., *Nature*, **459**, 398 (2009)

25. T. R. Bedding *et al. Astrophys. J.* **713**, L176 (2010)

26. Materials and methods are available as supporting material on *Science* online

27. C. Turon, F. Primas, J. Binney, C. Chiappini, J. Drew, A. Helmi, A. C. Robin, S. G. Ryan, ESA-ESO Working Group on Galactic Populations, Chemistry and Dynamics, *ESA-ESO Working Group Reports* (2008)

28. G. Torres *et al., Astron. Astrophys. Rev*. **18**, 67 (2010)

29. J. Molenda-Zakowicz *et al*., *Astronomische Nachrichten*, **331**, 981 (2010)

30. P. Marigo, *et al.*, *Astron. & Astrophys.* **483**, 883 (2008).

31. L. Girardi, M. A. T. Groenewegen, E. Hatziminaoglou, L. Da Costa, *Astron. & Astrophys.* **436**, 895 (2005).

32. A. Miglio *et al.*, *Astron. & Astrophys.* **523**, L21 (2009).

33. *Kepler* is a *NASA Discovery Class Mission*, which was launched in March 2009, whose funding is provided by NASA's *Science Mission Directorate*. The authors wish to thank the entire *Kepler* team, without whom these results would not be possible.

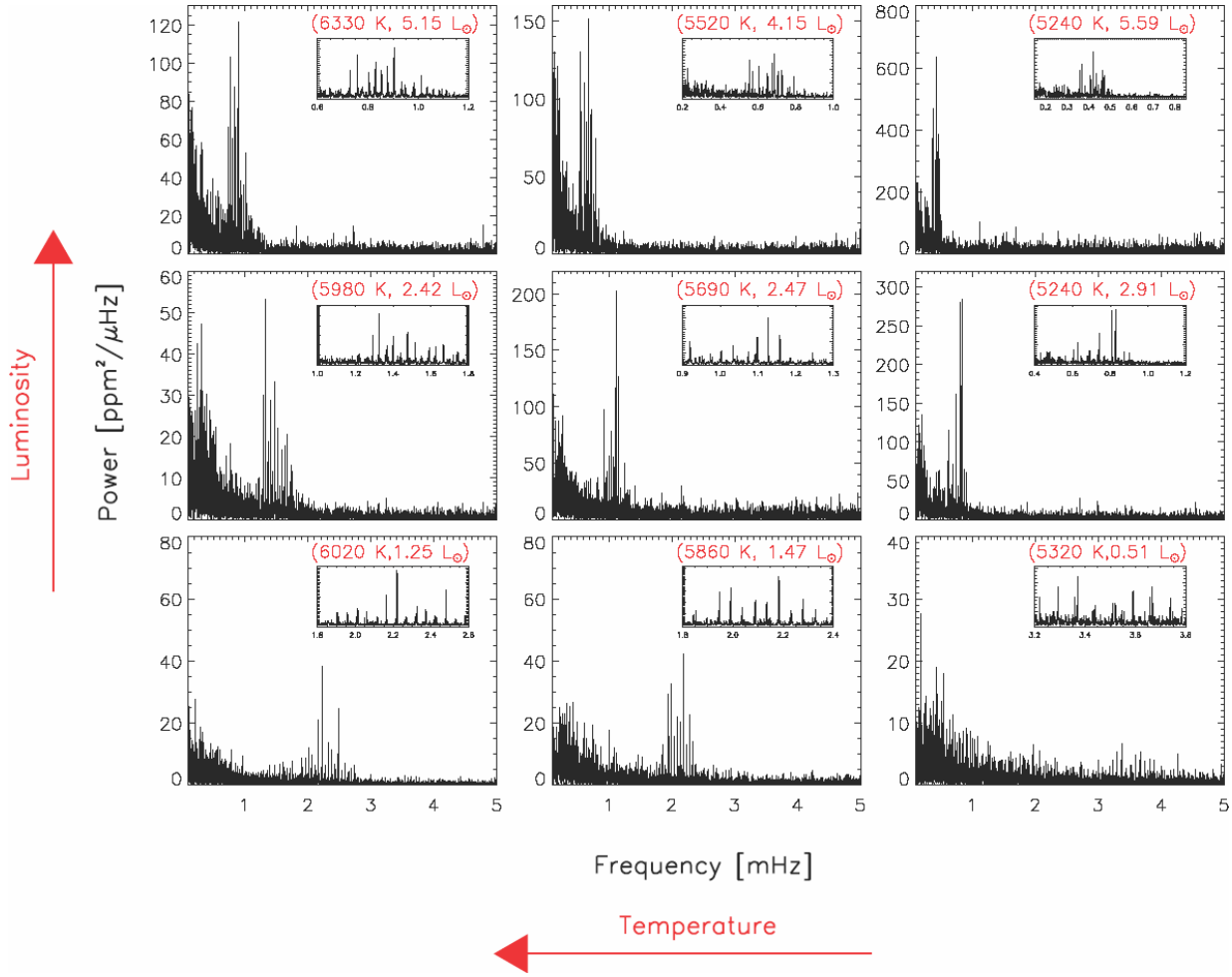

Fig 1.: Frequency spectra of the oscillations exhibited by nine stars from the ensemble. Each spectrum shows a prominent concentration of power due to the oscillations, modulated by a Gaussian-shaped envelope centered on the frequency ν_{max}. Closer inspection reveals that each concentration comprises patterns of near regularly spaced peaks, the most prominent separation $\Delta\nu$ being that between consecutive oscillation overtones (the insets show these patterns more clearly). The rise in power at low frequencies in each panel is due predominantly to stellar noise from surface convection. The stars are arranged by intrinsic brightness (luminosity) and temperature, instrinsically fainter stars clearly showing weaker, less prominent oscillations than their intrinsically brighter cousins.

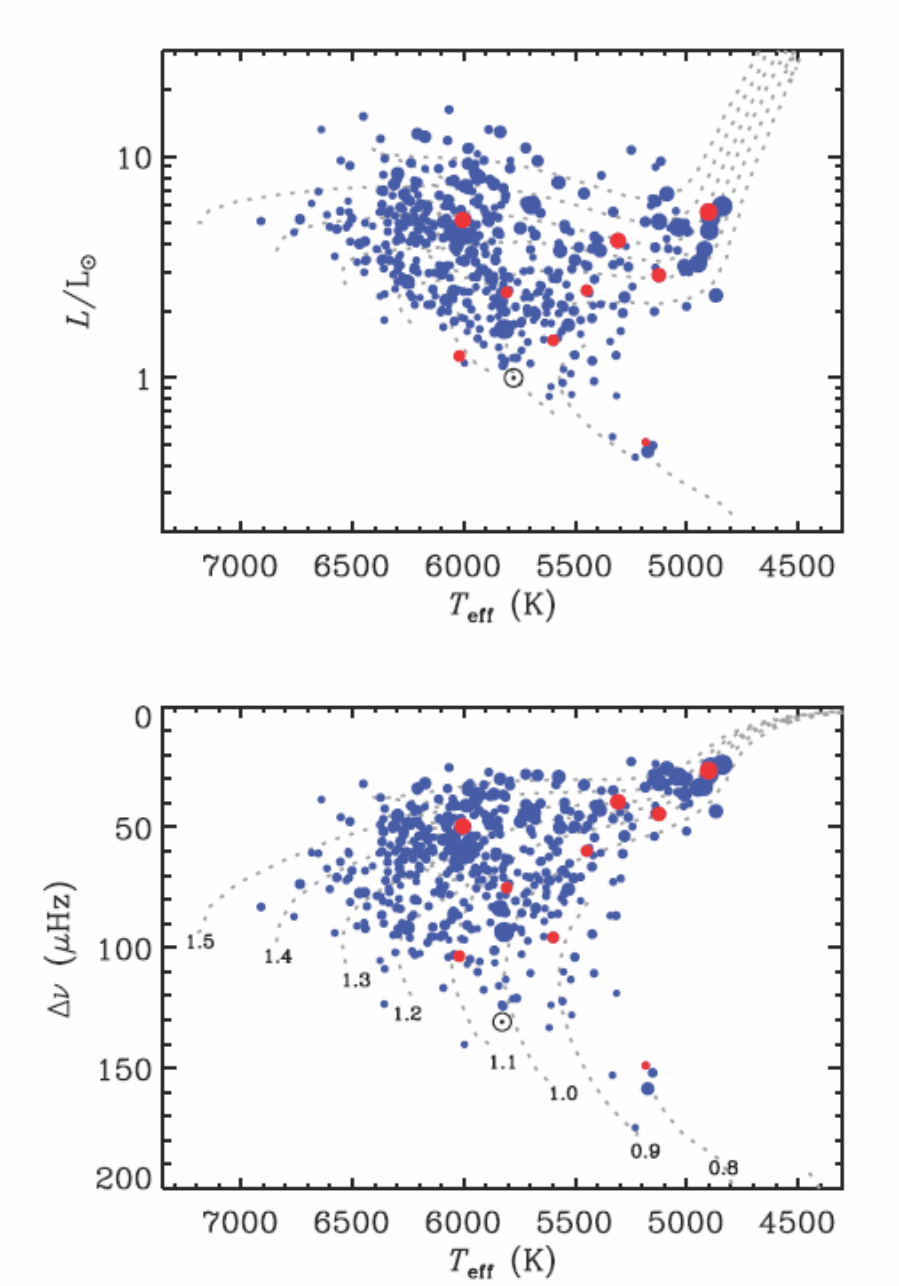

Fig 2.: Top panel: Estimates of the luminosities of the stars (in units of the solar luminosity) of the ensemble of *Kepler* stars showing detected solar-like oscillations, plotted as a function of effective temperature. Stars from Fig. 1 are plotted with red symbols. Bottom panel: average large frequency separations, Δν, against effective temperature. The symbol sizes are directly proportional to the prominence of the detected oscillations (i.e., the signal-to-noise ratios). These ratios depend both on stellar properties (e.g. the photometric amplitudes shown by the oscillations, and the intrinsic stellar backgrounds from convection) and the apparent brightness of the stars. Solar-like oscillations are detected over the range of *Kepler* apparent magnitudes sampled by our survey, i.e., from about 7th down to 12th magnitude. The dotted lines show predicted evolutionary tracks (*30*) for models of different stellar mass (0.8 to 1.5 solar masses, in steps of 0.1). The Sun is marked with its usual symbol [⊙].

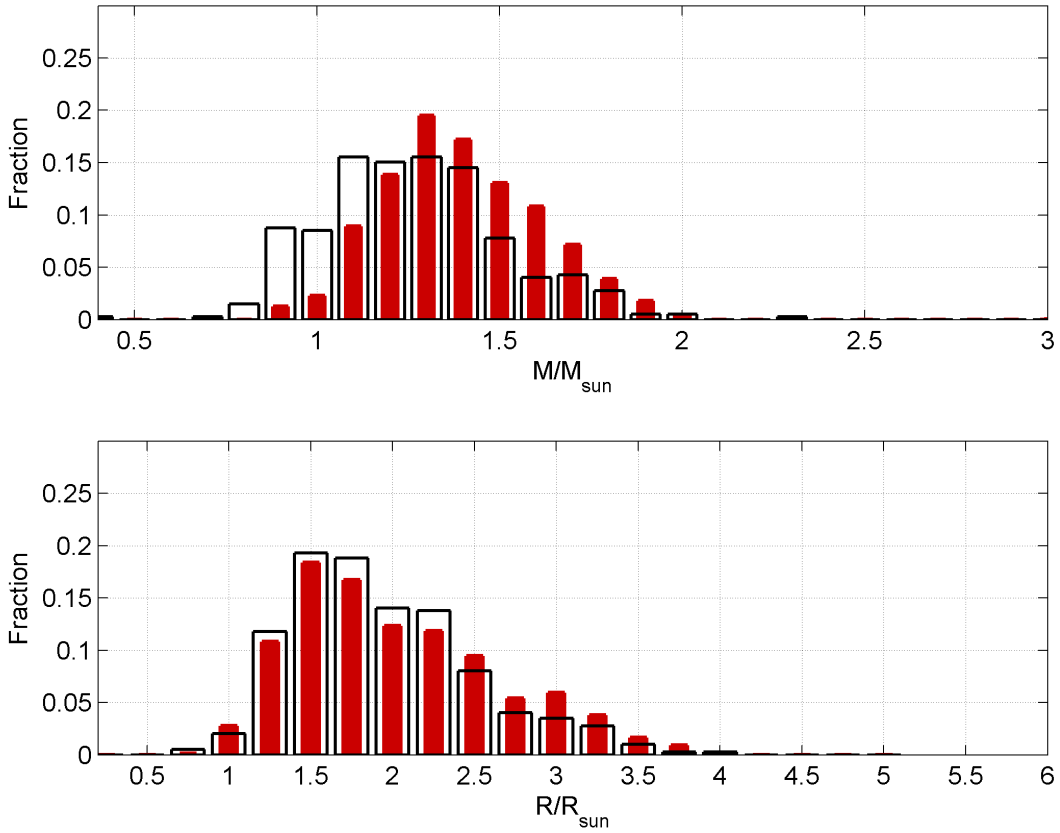

Fig. 3.: Black lines: Histograms of the observed distribution of masses and radii of the *Kepler* ensemble (*26*). In Red: the predicted distributions from population synthesis modelling, after correction for the effects of detection bias (*26*). The population modelling was performed using the TRILEGAL code (*31,32*).

Affiliations

1 School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

2 Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark

3 Department of Astronomy, Yale University, P.O. Box 208101, New Haven, CT 06520-8101, USA

4 Département d'Astrophysique, Géophysique et Océanographie (AGO), Université de Liège, Allée du 6 Août 17 4000 Liège 1, Belgique

5 Institut d'Astrophysique Spatiale, Université Paris XI–CNRS (UMR8617), Batiment 121, 91405 Orsay Cedex, France

6 Sydney Institute for Astronomy (SIfA), School of Physics, University of Sydney, NSW 2006, Australia

7 Laboratoire AIM, CEA/DSM–CNRS–Université Paris Diderot ; IRFU/SAp, Centre de Saclay, 91191 Gif-sur-Yvette Cedex, France

8 Space Telescope Science Institute, Baltimore, MD 21218, USA

9 Osservatorio Astronomico di Padova, INAF, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy

10 Institute of Astronomy, University of Vienna, A-1180 Vienna, Austria

11 Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, USA

12 High Altitude Observatory and, Scientific Computing Division, National Center for Atmospheric Research, Boulder, Colorado 80307, USA

13 Astronomical Institute, University of Wrocław, ul. Kopernika, 11, 51-622 Wrocław, Poland

14 Centro de Astrofísica and Faculdade de Ciências, Universidade do Porto, Rua das Estrelas, 4150-762, Portugal

15 Astronomy Unit, Queen Mary, University of London, Mile End Road, London, E1 4NS, UK

16 Laboratoire d'Astrophysique de Toulouse-Tarbes, Université de Toulouse, CNRS, 14 av E. Belin, 31400 Toulouse, France

17 INAF Osservatorio Astrofisico di Catania, Via S.Sofia 78, 95123, Catania, Italy

18 Observatoire de Paris, 5 place Jules Janssen, 92190 Meudon Principal Cedex, France

19 Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain

20 Instituto de Astrofísica de Canarias, E-38200 La Laguna, Tenerife, Spain

21 Department of Physics, Dezhou University, Dezhou 253023, China

22 Homi Bhabha Centre for Science Education (TIFR), V. N. Purav Marg, Mumbai 400088, India

23 LESIA, CNRS, Université Pierre et Marie Curie, Université, Denis Diderot, Observatoire de Paris, 92195 Meudon cedex, France

24 Materials Engineering Research Institute, Faculty of Arts, Computing, Engineering and Sciences, Sheffield Hallam University, Sheffield, S1 1WB, UK

25 Department of Astronomy, The Ohio State University, 4055 McPherson Laboratory, 140 West 18th Avenue, Columbus, OH 43210, USA

26 Astronomical Institute of the Romanian Academy, Str. Cutitul de Argint, 5, RO 40557, Bucharest, RO

27 Canadian Space Agency, 6767 Boulevard de l'Aéroport, Saint-Hubert, QC, J3Y 8Y9, Canada

28 Instituto de Ciencias del Espacio (CSIC-IEEC), Campus UAB, Facultad de Cìencies, Torre-C5, 08193, Bellaterra, Spain

29 Max Planck Institute for Astrophysics, Karl Schwarzschild Str. 1, Garching, D-85741, Germany

30 NASA Ames Research Center, MS 244-30, Moffett Field, CA 94035, USA

31 Las Cumbres Observatory Global Telescope, Goleta, CA 93117, USA

32 SETI Institute/NASA Ames Research Center, Moffett Field, CA 94035, USA

33 Bay Area Environmental Research Institute/NASA Ames Research Center, Moffett Field, CA 94035, USA

34 Orbital Sciences Corporation/NASA Ames Research Center, Moffett Field, CA 94035, USA

Online Material: Ensemble asteroseismology of solar-type stars with the NASA *Kepler Mission*

1 Observations

The primary objective of the NASA *Kepler Mission* is to detect, by the transit method, Earth-sized planets in the habitable zones of solar-type stars. Photometry of most of the stars is conducted at a long cadence (LC) of 29.4 minutes, but a subset of up to 512 stars can be observed at any one time at a short cadence (SC) of 58.85 s (S1, S2). The cadence of the SC data is rapid enough to allow investigations of solar-like oscillations in main-sequence stars, where dominant periods are of the order of several minutes.

We use asteroseismic results on solar-type stars that were observed by *Kepler* during the first seven months of science operations. About 2000 stars, down to *Kepler* apparent magnitude $Kp \simeq 12$, were selected as potential solar-type targets based upon parameters in the *Kepler* Input Catalog (S3, S4). Each star was observed for one month at a time in SC mode. Time series were prepared for asteroseismic analysis in the manner described in S5, using procedures that work on the raw lightcurves which were developed for application to GOLF/SoHO data (S6). Additive corrections were applied to correct thermal drifts; any sudden jumps or discontinuities were removed; while outliers were removed by clipping at the 5σ level.

2 Estimation of average seismic parameters

The frequency power spectra of "solar-like" oscillations in solar-type stars present a pattern of peaks with near regular frequency separations (see Figs. 1 and 2). The mode powers are modulated by an envelope that has a bell-shaped (i.e., Gaussian) appearance in many stars for which solar-like oscillations have been observed (including the Sun). Different techniques have been devised and applied to the *Kepler* SC data to detect signatures of solar-like oscillations (e.g., S7, S8, S9, S10, S11, S12, S13, S14). Some techniques rely on extracting signatures of the near-regular frequency separations of the oscillations, while others search for signatures of the Gaussian-like power excess due to the oscillations. These methods have been tested intensively on artificial datasets (e.g., as part of the asteroFLAG hare-and-hounds exercises; see S15, S16). The primary data products from these automated searches are estimates of the average large frequency separation, $\Delta\nu$, and the frequency of maximum oscillations power, $\nu_{\rm max}$.

The large frequency separation, $\Delta\nu$, is the spacing between consecutive overtones of the same spherical angular degree, l. When the signal-to-noise ratios in the seismic data are insufficient to allow robust extraction of individual oscillation frequencies, it is still possible to extract estimates of the average large frequency separation for use as the seismic input data. Indeed, this is the case for many *Kepler* stars. The average large separation scales to good approximation with the square root of the mean density of a star (S17). This gives the following scaling relation

$$\left(\frac{\Delta\nu}{\Delta\nu_\odot}\right) = \left(\frac{M}{\rm M_\odot}\right)^{0.5} \left(\frac{R}{\rm R_\odot}\right)^{-1.5}, \qquad (1)$$

with M and R the stellar mass and radius, and $\Delta\nu_\odot$ the large separation of the Sun.

The frequency of maximum power in the oscillations power spectrum, $\nu_{\rm max}$, is related to the acoustic cut-off frequency of a star (S18, S19, S20, S21), which in turn scales to very good approximation as $M\,R^{-2}\,T_{\rm eff}^{-1/2}$, $T_{\rm eff}$ the effective temperature of the star. This gives the following scaling

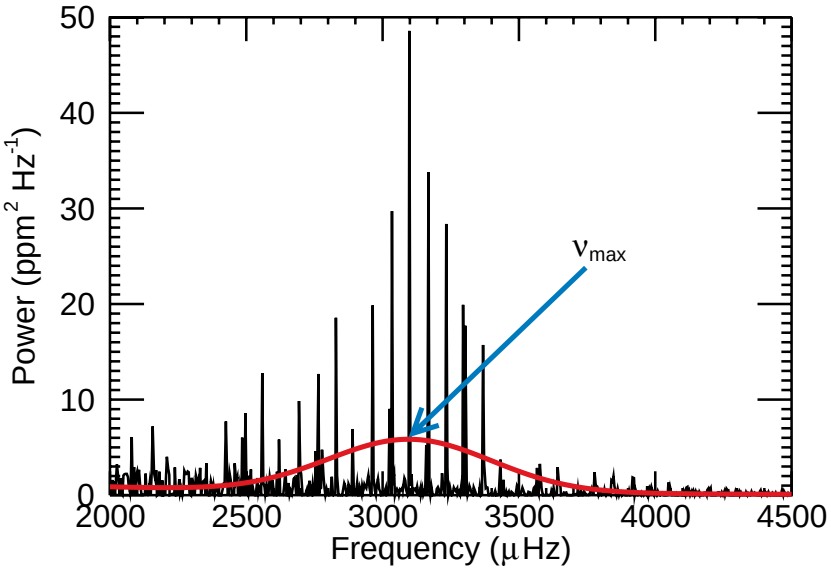

Figure 1: Frequency spectrum of low-degree oscillations shown by the Sun (in Sun-as-a-star photometry data from the VIRGO/SPM instrument on board the ESA/NASA SoHo spacecraft). The red line follows the Gaussian-like power envelope of the observed oscillations, with the frequency of maximum power marked by $\nu_{\rm max}$.

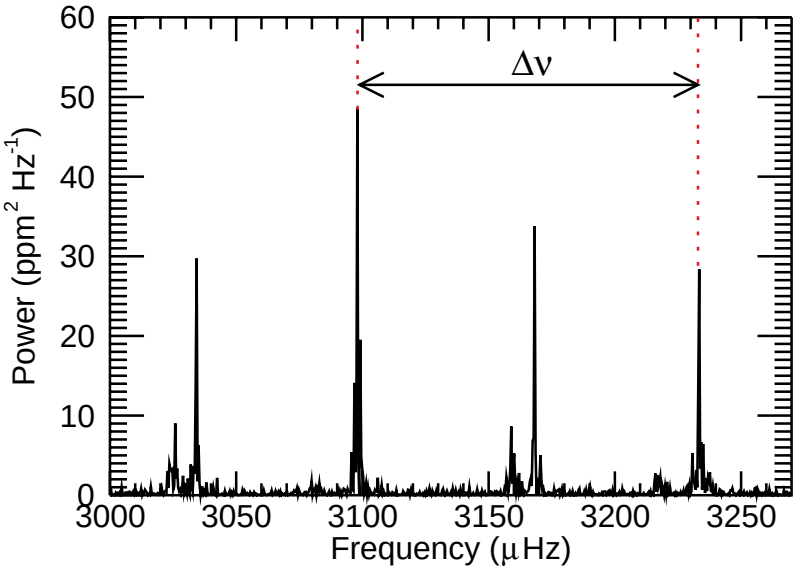

Figure 2: Zoom on the central frequency region of the oscillations spectrum plotted in Fig. 1, to show the large frequency separation, $\Delta\nu$.

relation:

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

3 Estimation of M and R

3.1 Direct method of estimation

If $\Delta\nu$, $\nu_{\rm max}$ and $T_{\rm eff}$ are known, Equations (1) and (2) represent two equations in two unknowns, and we may therefore re-arrange and solve for M and R in what we call the "direct" method of estimation of the stellar properties. This gives (S22):

$$\left(\frac{R}{\rm R_\odot}\right) = \left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}}\right)\left(\frac{\Delta\nu}{\Delta\nu_\odot}\right)^{-2}\left(\frac{T_{\rm eff}}{\rm T_{eff,\odot}}\right)^{0.5}, \quad (3)$$

and

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

The results presented in the paper were estimated by the direct method. We used $\nu_{\rm max,\odot} = 3150\,\mu$Hz and $\Delta\nu_\odot = 134.9\,\mu$Hz. Estimates of $T_{\rm eff}$ were derived by An et al., (in preparation) from the multicolor photometry available in the *Kepler* Input Catalog. The median fractional precision in the temperatures is about 1 %.

We obtained from the direct method a median fractional uncertainty of just over 10 % in M and about 5.5 % in R. That the fractional uncertainties on M are larger than those on R is apparent from the propagation of the uncertainties on the observables, i.e.,

$$\left(\frac{\delta R}{R}\right)^2 = \left(\frac{\delta\nu_{\rm max}}{\nu_{\rm max}}\right)^2 + \left(2\frac{\delta\Delta\nu}{\Delta\nu}\right)^2 + \left(0.5\frac{\delta T_{\rm eff}}{T_{\rm eff}}\right)^2. \quad (5)$$

and

$$\left(\frac{\delta M}{M}\right)^2 = \left(3\frac{\delta\nu_{\rm max}}{\nu_{\rm max}}\right)^2 + \left(4\frac{\delta\Delta\nu}{\Delta\nu}\right)^2 + \left(1.5\frac{\delta T_{\rm eff}}{T_{\rm eff}}\right)^2, \quad (6)$$

The direct method is very attractive because it provides mass and radius estimates that are independent of stellar evolutionary models, and this is clearly of great benefit for instructive comparisons with the population synthesis models. However, the direct method does give larger uncertainties on M and on R than would be obtained from a "grid-based" method of estimation (here about twice as large; see Section 3.2 below). Although the uncertainties are larger than for the grid-based method, this does mean that they are expected to largely capture any uncertainties in Equations (1) and (2) due to, for example, metallicity effects.

For the present, the lack of precise independent constraints on the metallicities (e.g., on [Fe/H]) means that the grid method is vulnerable to systematic bias in the estimates of M (although not R). Once we have those tight independent constraints on all the stars – from complementary ground-based observations being made in support of *Kepler* – we will be able to take full advantage of the grid-based method and the significantly better precision it offers, as we now go on to discuss.

3.2 Grid-based method of estimation

We also applied the so-called grid-based method to estimation of the M and R of the *Kepler* stars. This is essentially the well-used approach in stellar astronomy of matching the observations to stellar evolutionary tracks, but with the powerful diagnostic information contained in the seismic $\Delta\nu$ and

$\nu_{\rm max}$ also brought to bear (e.g., as per Equations 1 and 2). Properties of stars are determined by searching among a grid of stellar evolutionary models to get a "best fit" for a given observed set of input parameters, typically $\{\Delta\nu, \nu_{\rm max}, T_{\rm eff}, [{\rm Fe/H}]\}$. While the direct method assumes that all values of temperature are possible for a star of a given mass and radius, we know from stellar evolution theory that only a narrow range of $T_{\rm eff}$ is allowed for a given M and R. This prior information is implicit in the grid-based approach, and means that estimated uncertainties on M and R are lower than for the direct method because a narrower range of outcomes is permitted.

Descriptions of the various grid-based pipelines used in the analysis of *Kepler* data may be found in, for example, S16, S23, S24, and S25. In addition to $\Delta\nu$, $\nu_{\rm max}$ and $T_{\rm eff}$, the grid-based methods also used as input the [Fe/H] provided in the *Kepler* Input Catalog. The KIC [Fe/H] are derived from the *Kepler* photometry, but since the available wavelength bands are not optimally sensitive to metallicity the uncertainties are very large (taken to be 0.5 dex).

The grid-based estimates of M are sensitive to choices made in construction of the grid of stellar models, i.e., they are to some extent model dependent (see S25 for an in-depth discussion). The biggest sensitivity is to the metallicity. One might have thought that the large uncertainties assumed on the KIC [Fe/H] would have largely captured the range of possibilities, albeit at the cost of reducing the precision in the estimated M (even then, we obtained a typical fractional median error on M of about 6 %). However, without tight prior constraints on metallicity, the estimated masses are still affected by the helium abundance, Y, of models in the grid (often treated in different ways). This is why it is extremely important to have precise independent constraints on [Fe/H] for all the stars. Inconsistencies resulting from the poorly known metallicity and the treatment of Y are then removed, and it will be possible to take full advantage of the superior precision in M offered by the grid-based approach.

As shown by the extensive tests of the method (S16, S25) the grid-based estimates of R are, in contrast, largely insensitive to the choice of grid, i.e., they may be regarded as being essentially model independent, at the typical level of precision given by the one-month *Kepler* survey data. We found that estimates of R for the *Kepler* stars returned by the direct and grid-based methods were indeed consistent.

4 Population synthesis modelling

We estimated the properties of the stellar population observed by *Kepler* using the code TRILEGAL (S26), designed to simulate photometric surveys in the Galaxy. In TRILEGAL, several model parameters (such as the star-formation history and the morphology of different galactic components) were calibrated to fit Hipparcos data for the immediate solar neighbourhood (S27), as well as star counts from a wide area (with 2MASS; S28), and a few deep photometric surveys, i.e., CDFS (S29), and DMS (S30). We adopted the standard parameters describing the components of the Galaxy and simulated the stellar population in the sky area observed in each of the 21 five-square-degree *Kepler* sub-fields of view, considering for each of them an average interstellar extinction at infinity (S31). The extinction is assumed to be caused by an exponential dust layer with a scale height, above and below the galactic plane, equal to 110 parsec. The photometry in TRILEGAL was simulated with the known wavelength response function of *Kepler*, and the synthetic population was magnitude-selected, using the same range as the observed sample (see Section 1 above).

We also simulated observing each of the stars in the synthetic population for one month at a time with *Kepler* (S32), in order to predict whether the stars would show detectable solar-like oscillations. Those synthetic stars judged to have detectable oscillations were added to a final list for direct comparison with the observations. In brief, we used the known M, R and $T_{\rm eff}$ of each synthetic star to predict the total mean power we would expect to observe in its solar-like oscillations, using well-established scaling relations that describe the seismic parameters in terms of the stellar properties

(e.g., S18, S20). From the distance of each synthetic star, and the known luminosity, we could calculate an apparent magnitude and, from that, the noise that would be expected in observations (using the description in S1). We then applied statistical tests to the resulting S/N estimate to estimate the probability of detecting solar-like oscillations, assuming observations lasting one month. The tests were based on well-established approaches used in helioseismology and asteroseismology (S33, S34). Synthetic stars were included in the final synthetic distribution if there was judged to be a better than 9 in 10 chance of making a detection.

5 Comparison of observed and synthetic distributions

We applied the Kolmogorov-Smirnov (K-S) test in order to quantify differences between the observed and synthetic distributions. The K-S test returns an estimate of the probability that the observed and synthetic distributions come from the same, parent population. We applied the test in a way that took into account the statistical uncertainties on the observed masses and radii. Take, for example, the test as applied to the masses. To each observed mass we added a random offset drawn from a Gaussian distribution having a standard deviation equal to the estimated uncertainty on the mass. The K-S test was then applied to the resulting, perturbed set of observed masses. The test was repeated 5000 times, each time on a fresh set of perturbed masses (made with a new set of random numbers). The distribution of results from the 5000 K-S tests then allowed us to judge the full range of probabilities commensurate with the observational uncertainties. (We used results from artificial data to calibrate the impact of uncertainties of different fractional sizes on the returned K-S probabilities, in cases where the artificial "observations" were known to come from the same population as the artificial "synthetic" data.)

The K-S probabilities for mass were in every case lower than 10^{-5}, indicating that differences between the observed and synthetic distributions were highly significant. K-S probabilities for radius gave values that in contrast went up to approximately 20 %, with a typical value between 5 and 10 %. These levels of probability (in what is a null hypothesis test) are only marginally significant at best.

References

S1. Gilliland, R., Jenkins, J. M., Borucki, W. J., et al., 2010, ApJ, 713, 160L
S2. Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al., 2010, ApJ, 713, L120
S3. Batalha, N. M., Rowe, J. F., Gilliland R. L., et al., 2010, ApJ, 713, L109
S4. Koch, D. G., Borucki, W. J., Basri, G., 2010, ApJ, 713, L79
S5. García, R. A., Hekker, S., Stello, D., et al., 2010, MNRAS, submitted
S6. García, R. A., Turck-Chièze, S., Boumier, P., et al., 2005, A&A, 442, 385
S7. Campante, T. L., Karoff, C., Chaplin, W. J., Elsworth, Y., Handberg, R., Hekker, S., 2010, MNRAS, 408, 542
S8. Christensen-Dalsgaard, J., Arentoft, T., Brown, T. M., Gilliland, R. L., Kjeldsen, H., Borucki, W. J., Koch, D., 2008, J. Phys. Conf. Ser., 118, 012039
S9. Hekker, S., Broomhall, A.-M., Chaplin, W. J., et al., 2010, MNRAS, 402, 2049
S10. Huber, D., Stello, D., Bedding, T. R., Chaplin, W. J., Arentoft, T., Quirion, P.-O., Kjeldsen, H., 2009, CoAst, 160, 74
S11. Karoff, C., Campante, T. L., Chaplin, W. J., 2010, AN, in the press (arXiv:1003.4167)
S12. Mosser, B., Appourchaux, T., 2009, A&A, 508, 877
S13. Mathur, S., García, R. A., Régulo C., et al., 2010, A&A, 511, 46
S14. Roxburgh, I. W., 2009, A&A, 506, 435
S15. Chaplin, W. J., Appourchaux, T., Arentoft, T., et al., 2008, AN, 329, 549
S16. Stello, D., Chaplin, W. J., Bruntt, H., 2009, ApJ, 700, 1589

S17. Christensen-Dalsgaard, J. 1993, in ASP. Conf. Ser. 42, Proc. GONG 1992, Seismic Investigation of the Sun and Stars, ed. T. M. Brown (San Francisco, CA: ASP), 347
S18. Kjeldsen, H., Bedding, T. R., 1995, A&A, 293, 87
S19. Bedding, T. R. & Kjeldsen, H. 2003, PASA, 20, 203
S20. Chaplin, W. J., Houdek, G., Appourchaux, T., Elsworth, Y., New, R. & Toutain, T. 2008, A&A, 485, 813
S21. Mosser, B., Belkacem, K., Goupil, M.-J., et al., 2010, A&A, 517, 22
S22. Kallinger, T., Weiss, W. W., Barban, C., et al., 2010, A&A, 509, 77
S23. Basu, S., Chaplin, W. J., Elsworth, Y., 2010, ApJ, 710, 1596
S24. Quirion, P.-O., Christensen-Dalasgaard, J., Arentoft, T., 2010, ApJ, in the press (arXiv:1009.5131)
S25. Gai, N., Basu, S., Chaplin, W. J., Elsworth, Y., 2010, ApJ, submitted (arXiv:1010.0834)
S26. Girardi, L., Bressan, A., Bertelli, G & Chiosi C., 2000, A&ASS, 141, 371
S27. Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49
S28. Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, 2MASS All Sky Catalog of point sources
S29. Arnouts, S., Vandame, B., Benoist, C., et al. 2001, A&A, 379, 740
S30. Osmer, P. S., Kennefick, J. D., Hall, P. B., Green, R. F. 1998, ApJS, 119, 189
S31. Schlegel D. J., Finkbeiner, D. P., Davis, M., 1998, ApJ, 500, 525
S32. Chaplin, W. J., Kjeldsen, H., Bedding, T. R., et al., ApJ, submitted
S33. Appourchaux, T., 2004, A&A, 428, 1039
S34. Broomhall, A.-M., Chaplin, W. J., Elsworth, Y., Appourchaux, T., New, R., 2010, MNRAS, 406, 767