

Development of a Stellar Model-Fitting Pipeline for Asteroseismic Data from the Kepler Mission

Summary: In the past two decades, helioseismology has revolutionized our understanding of the interior structure of the Sun. Asteroseismology will place this knowledge into context, by providing structural information for hundreds of solar-type stars. Solar-like oscillations have already been detected from the ground in several stars, and NASA's Kepler mission will soon unleash a flood of stellar pulsation data. Deriving reliable asteroseismic constraints from these observations will require a significant improvement in our analysis methods. We propose to develop an objective model-fitting pipeline for asteroseismic data from the Kepler mission. The cornerstone of our automated approach is an optimization method using a parallel genetic algorithm (GA). This method was developed for white dwarf asteroseismology, and led to a dramatic improvement in our ability to match the observed pulsation frequencies with theoretical models. We now aim to extend this success to models of main-sequence stars, and to explore the information content of the observables by combining the GA with a local analysis using singular value decomposition (SVD). We have adapted the Aarhus stellar evolution and pulsation codes to interface with the GA. These models were developed for helioseismology and are the source of 'Model S', which has been used extensively as a reference model for solar inversions. We have applied a prototype of our pipeline to Sun-as-a-star data, yielding a standard solar model within reasonable tolerances. We now propose to validate the method on stars with various masses and at different evolutionary stages using archival ground-based data. We will then exploit the better resolution and coverage of Kepler data to include rotation velocity as an additional fitting parameter, which will be essential for faster rotators to avoid possible ambiguity in the mode identification. Finally, we will provide our pipeline as a web-based service to the broader community through a TeraGrid Science Gateway.

1 Significance & Objectives

Most of what we can learn about stars comes from observations of their outermost surface layers. We are left to infer the properties of the interior based on our best current understanding of the constitutive physics. The exception to this general rule arises from observations of pulsating stars, where seismic waves probe deep through the interior and bring information to the surface in the form of light and radial velocity variations. The most dramatic example is the Sun, where such observations have led to the identification of millions of unique pulsation modes, each sampling the solar interior in a slightly different and complementary way. The inverted radial profile of the sound speed from these data led to such precise constraints on the standard solar model that, before the recent controversy over heavy element abundances (Asplund et al., 2004), the observations and theory agreed to better than a few parts per thousand over 90 percent of the solar radius (Christensen-Dalsgaard, 2002). Although such detailed inversions are not currently possible for other stars, pulsation data does allow us to determine the global properties and to probe the gross internal composition and structure, providing valuable independent tests of stellar evolution theory. The reason for this qualitative difference is simple: if we could move the Sun to the distance of the nearest star, many of the pulsation modes would no longer be detectable. We would lose much of our spatial resolution, so only those modes with the lowest spherical degree ($\ell \lesssim 3$) would produce significant variations in the total integrated light or the spectral line profiles. These are also the modes that probe deepest into the stellar interior, collectively sampling the physical conditions from the core to the photosphere.

Recent improvements in our ability to make high-precision radial velocity measurements from the ground have been driven largely by efforts to detect extra-solar planets. These advances in technology have also led to the first unambiguous detections of solar-like oscillations in other stars. Scintillation in the Earth's atmosphere severely limits our ability to detect the parts-per-million light variations due to these pulsations. However, space-based photometric programs designed to detect extra-solar planet transits also have the sensitivity to document the stellar pulsation signals. Solar-like oscillations have now been detected in more than a dozen main-sequence and subgiant stars (including α Cen A & B and β Hyi), and in numerous giants (for a recent review see Bedding & Kjeldsen, 2007). The oscillation amplitudes and the frequency of maximum power in these stars agree reasonably well with our theoretical expectations—but in some cases the mode lifetimes are significantly shorter than expected (Stello et al., 2004), suggesting that our knowledge of the convective driving and damping physics is incomplete. The field is progressing very rapidly and many new observations will become available in the next few years, particularly after the launch of NASA's Kepler mission in 2009. With this in mind, it will be a distinct advantage to have in place the computational methods that will allow us to maximize the science return of these data. This will lead us quickly to a deeper understanding of the solar oscillations in the context of similar pulsations in other stars, and will provide new insights into the formation and evolution of stellar and planetary systems.

The likely excitation mechanism for solar-like oscillations is turbulent convection near the surface, creating a broad envelope of power with a peak that scales with the acoustic cutoff frequency (Brown et al., 1991). Within this envelope a large fraction of the predicted low-degree pulsation modes are excited to detectable amplitudes, leading to readily identifiable patterns (see Figure 1). Without any detailed modeling, these overall patterns (the so-called large and small separations, $\Delta\nu_0$ and δ_{02}) immediately lead to an estimate of the mean density of the star and can indicate the presence of interior chemical gradients that reflect the stellar age. But a full analysis must include a detailed comparison of the individual frequencies with theoretical models. One complication with such a comparison is the existence of so-called *surface effects*, which appear as systematic

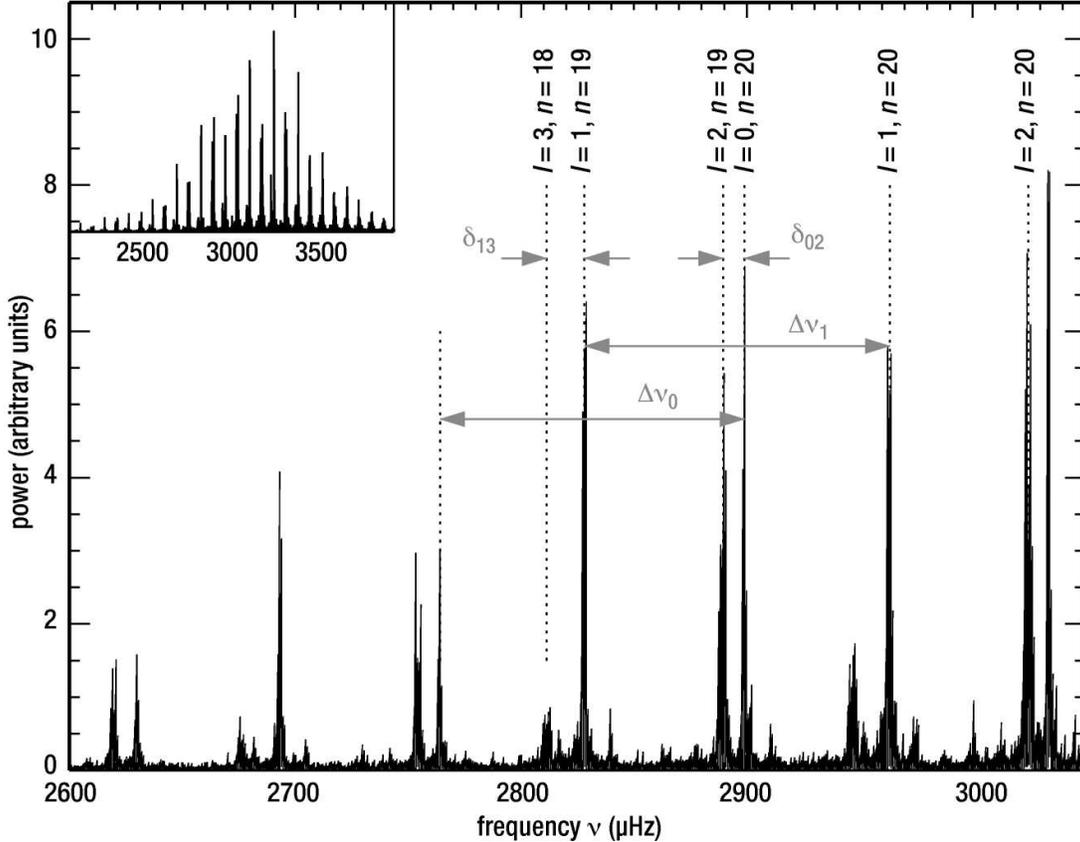


Figure 1: A portion of the solar power spectrum as observed by BiSON, showing the characteristic large and small separations $\Delta\nu$ and $\delta\nu$. The complete power spectrum from these Sun-as-a-star observations is shown in the inset (adapted from Elsworth & Thompson, 2004).

differences between the observed and calculated oscillation frequencies that grow larger towards the acoustic cutoff frequency. Surface effects arise primarily due to incomplete modeling of the near-surface layers of the star where convection plays a major role (Christensen-Dalsgaard & Thompson, 1997), and they are evident even in the best standard solar models like ‘Model S’. Addressing this inherent deficiency in our 1D models would require (among other things) that we substitute the results of extensive 3D calculations for the parameterized mixing-length treatment of convection that is currently used in nearly all stellar evolution codes. Alternately, we can make an empirical correction to the calculated frequencies following Kjeldsen et al. (2008), who recently devised a method for calibrating surface effects using solar data, and then scaling by the mean stellar density for other models.

The overall goal of this proposal is the development of an objective and automated method of fitting stellar models to the asteroseismic data soon expected to emerge from NASA’s Kepler mission. This will lead to reliable determinations of stellar radii to help characterize the extra-solar planetary systems discovered by the mission, and stellar ages to reveal how such systems evolve over time. For the asteroseismic targets that do not contain planetary companions it will allow a uniform determination of fundamental physical properties for hundreds of solar-type stars, providing a new window into stellar structure and evolution. To realize these broad goals, we are proposing to implement some powerful optimization and analysis methods (**section 2**) to facilitate a massive

exploration of asteroseismic models for solar-type stars in an effort to unleash the full potential of the observations. We have already developed a prototype of this model-fitting pipeline, and we have successfully used it to derive a standard solar model from Sun-as-a-star observations. We now need to ensure that our empirical treatment of surface effects is sufficient for stars with various masses and at different evolutionary stages (**section 3.1**). The longer time baseline and continuous coverage of space-based data will then allow us to include rotation velocity as an additional fitting parameter, which will be essential for the faster rotators to avoid possible ambiguity in the mode identification (**section 3.2**). This plan for the development of a model-fitting pipeline will clearly facilitate the interpretation of data from NASA’s Kepler mission while contributing to NASA’s strategic goals (**section 4**) and it is a natural extension of our previous work in this area (**section 5**). Finally, the unprecedented quantity of asteroseismic data that will emerge from the Kepler mission motivates us not only to release the source code of the pipeline, but also to develop a web-based interface tied to TeraGrid computing resources through a Science Gateway, to encourage a uniform analysis of all data without requiring a detailed operational knowledge of the code. This work is part of our data sharing plan (**section 5.3**).

2 Technical Approach

The Kepler mission will soon yield precise high cadence time-series photometry of hundreds of pulsating stars every few months for at least 3.5 years. We will then face the challenge of determining the fundamental properties of these stars from the data, by attempting to match them with the output of computer models. The traditional approach to this task is to make informed guesses for each of the model parameters, and then to adjust them iteratively until an adequate match is found. The volume of asteroseismic data that will emerge from the Kepler mission calls for a more automated approach to modeling that initially explores a broad range of model parameters in an objective manner. The cornerstone of our model-fitting approach is a global optimization method using a parallel genetic algorithm (**section 2.1**). The result of the global search provides the starting point for a local analysis using singular value decomposition (**section 2.2**), which will also allow us to explore the information content of the observables and the impact of including other observational constraints.

2.1 Global Search: Parallel Genetic Algorithm

An optimization scheme based on a genetic algorithm (GA; Charbonneau, 1995; Metcalfe & Charbonneau, 2003) can avoid the problems inherent in many traditional approaches. Using only observations and the constitutive physics of the model to restrict the range of possible values for each parameter, genetic algorithms provide a relatively efficient means of searching globally for the optimal model. Although it is more difficult for GAs to find *precise* values for the optimal set of parameters efficiently, they are well suited to search for the *region* of parameter space that contains the global minimum. In this sense, the GA is an objective means of obtaining a good first guess for a more traditional local hill-climbing method, which can narrow in on the precise values and uncertainties of the optimal parameters.

Genetic algorithms (Goldberg, 1989; Davis, 1991; Holland, 1992; Mitchell, 1996), arguably still the most popular class of evolutionary algorithms (Michalewicz, 1992; Bäck, 1996), were inspired by Charles Darwin’s notion of biological evolution through natural selection (Darwin, 1859). The basic idea is to solve an optimization problem by *evolving* the global solution, starting with an initial set of purely random guesses. The evolution takes place within the framework of the model, with the individual parameters serving as the genetic building blocks. Selection pressure is imposed by some

goodness-of-fit measure between the model and observations. Several books have been written to describe how these ideas can be applied in a computational setting (Goldberg, 1989; Davis, 1991), but we provide a basic overview below.

2.1.1 Computational Details. To begin, the GA samples the parameter space at a fixed number of points defined by a uniform selection of randomly chosen values for each parameter. The GA evaluates the model for each set of parameters, and the predictions are compared to observations. Each point in the “population” of trials is subsequently assigned a *fitness* based on the relative quality of the match. A new “generation” of sample points is then created by selecting from the current population of points according to their computed fitness, and then modifying their defining parameter values with two genetic operators in order to explore new regions of parameter space.

Rather than modifying the parameter values directly, the genetic operators are applied to encoded representations of the parameter sets. The simplest way to encode them is to convert the numerical values of the parameters into a string of digits. The string is analogous to a chromosome, and each digit is like a gene. For example, a point defined by two parameters with numerical values $a_1 = 0.123$ and $b_1 = 0.456$ could be encoded into the string 123456.

The two basic genetic operators are *crossover*, which emulates sexual reproduction, and *mutation* which emulates somatic defects. The crossover procedure pairs up the strings, chooses a random position for each pair, and swaps the two strings from that position to the end. For example, suppose that the encoded string above is paired with another point having $a_2 = 0.567$ and $b_2 = 0.890$, which encodes to the string 567890. If the second position between numbers on the string is chosen, the strings become:

$$\begin{array}{l} 12\boxed{3456} \rightarrow 12\boxed{7890} \\ 56\boxed{7890} \rightarrow 56\boxed{3456} \end{array}$$

To help keep favorable combinations of parameters from being eliminated or corrupted too hastily, this operation is not applied to all of the pairs. Instead, it is assigned a fixed occurrence probability (p_c) per selected pair.

The mutation operator spontaneously replaces a digit in the string with a new randomly chosen value. In our example above, if the mutation operator is applied to the fourth digit of the second string, the result might be:

$$563\boxed{4}56 \rightarrow 563\boxed{2}56$$

Such digit replacement occurs with a small probability (p_m), often dubbed the *mutation rate*.

After both operators have been applied, the strings are decoded back into sets of numerical values for the parameters. In this example, the new first string 127890 becomes $a'_1 = 0.127$ and $b'_1 = 0.890$, and the new second string 563256 becomes $a'_2 = 0.563$ and $b'_2 = 0.256$. Note that mutation in this case has caused a significant “jump” in parameter space, from the value $b'_2 = 0.456$ that would have been generated by the crossover operation only. The new genetically-shuffled set of points replaces the original set, and the process is repeated until a termination criterion is met.

2.1.2 Stochastic & Parallel. It should be clear already from this brief introductory discussion that the operation of a GA involves a number of random processes, so that the resulting search algorithm is stochastic in nature. Consequently, there is always a finite probability that the GA will not find the globally optimal solution in a given run. This probability decreases gradually, of course, as the evolution is pushed through more and more generations. Alternately, one can run the GA for fewer generations, but do so several times with different random initialization. This form of higher-level Monte Carlo simulations makes it possible to establish the validity of the optimal set of model parameters with an acceptable degree of confidence.

Metcalfe & Charbonneau (2003) developed a fully parallel and distributed hardware/software implementation of the popular PIKAIA genetic algorithm that was originally written by Charbonneau (1995). Metcalfe et al. (2000) used this modeling tool in the context of white dwarf asteroseismology, which ultimately led to a number of interesting physical results, including: [1] a precise estimate of the astrophysically important ($^{12}\text{C} + ^4\text{He} \rightarrow ^{16}\text{O}$) nuclear reaction rate (Metcalfe, 2003), [2] the first unambiguous detection of a crystallized core in a massive pulsating white dwarf (Metcalfe et al., 2004), and [3] asteroseismic confirmation of a key prediction of diffusion theory in white dwarf envelopes (Metcalfe et al., 2005). The impact of this method on the analysis of pulsating white dwarfs suggests that seismological modeling of other types of stars could also benefit from this approach.

2.2 Local Analysis: Singular Value Decomposition

Once the GA brings us close enough to the optimal solution, we can treat the model locally as if it were linear. In general, we would like to find the parameters \mathbf{P} that yield predictions that are as close as possible to the actual measurements \mathbf{O} . This is a typical inverse problem whose solution \mathbf{P} (in the pure linear case) falls neatly out from:

$$\mathbf{P} = \mathbf{P}_0 + \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\delta\mathbf{B}. \quad (1)$$

Here $\delta\mathbf{B}$ is simply the difference between what we actually observe \mathbf{O} and the model observables \mathbf{B} which result from our first guess \mathbf{P}_0 , divided by the measurement errors ϵ . $\mathbf{U}\mathbf{W}\mathbf{V}^T$ is the *Singular Value Decomposition* (SVD) of the derivative matrix $\mathbf{D} = \frac{\partial\mathbf{B}}{\partial\mathbf{P}}\epsilon^{-1}$.

We will use SVD in our local analysis because it describes the relationship between the observables and the parameters as a set of linear transformation vectors with an importance value \mathbf{W} assigned to each of them. The matrices \mathbf{U} and \mathbf{V} are orthonormal vectors that span the observable and parameter spaces respectively. Thus, SVD can be used to understand how the observables with their associated errors constrain each of the parameters. In addition, it provides an effective inversion technique by allowing us to filter the least important information from the observables (some of which may be dominated by noise).

Table 1: Eclipsing binary parameter sets

Set	M_A	M_B	R_A	R_B	T_{BA}	i	a
P_R	0.90	0.15	0.80	0.20	0.40	89.5	0.02
P_0	0.92	0.16	0.83	0.19	0.38	89.6	0.02
P_{F1}	0.88	0.14	0.69	0.17	0.39	99.5	0.02
P_{F2}	0.89	0.15	0.76	0.19	0.39	89.6	0.02

To demonstrate the capabilities of SVD, we apply the technique to a detached eclipsing binary system. The parameters \mathbf{P} of the system are two masses (M_A, M_B), two radii (R_A, R_B), an effective temperature ratio (T_{BA}), the inclination of the system (i) and the separation of the components (a). The observables are semi-amplitudes of the radial velocity curves (K_A, K_B), the depth of the primary eclipse (D_P) and the ratio of the depths of both eclipses (D_{SP}), the duration of the *total* eclipse (L_T) and of the *flat* part of the eclipse (L_F), and finally the period of the system (p). The simple model that relates these parameters to the observables is a set of analytical equations. Suppose that our real model parameters (\mathbf{P}_R) are those given in the first row of Table 1, and our initial guesses of the parameters (\mathbf{P}_0) are given in the second row. We calculate \mathbf{D} and decompose it into solution and observable vectors using SVD.

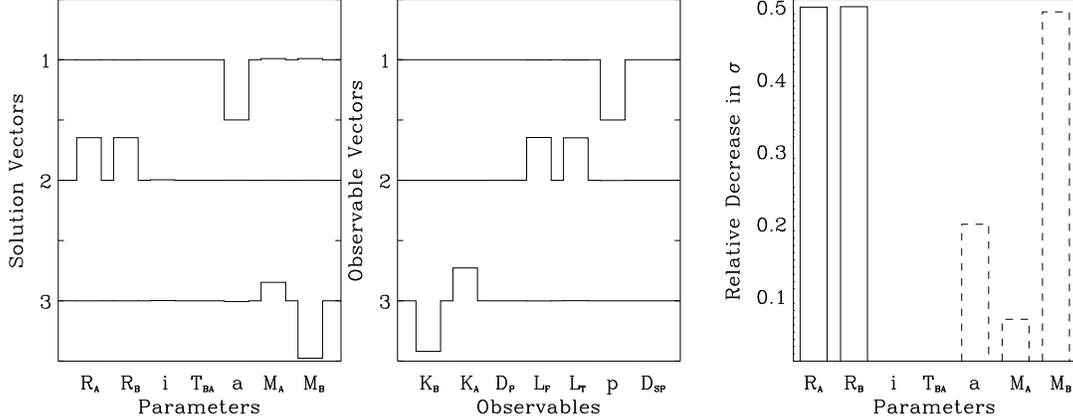


Figure 2: Information from the decomposition matrices of the derivative matrix \mathbf{D} . *Left*: Three solution vectors \mathbf{V} showing how much each parameter is constrained. *Center*: Three observable vectors \mathbf{U} quantifying the amount of information contributed by each observable. *Right*: The impact of reducing ϵ_{D_P} (solid) and ϵ_{K_A} (dashed) by a factor of two on the parameter uncertainties.

2.2.1 Relationship Vectors. Figure 2 shows the first three solution vectors (\mathbf{V} , left panel) and observable vectors (\mathbf{U} , center panel) of \mathbf{D} . The size of the singular values (\mathbf{W}) indicate the relative importance of the corresponding vectors. So, the largest singular value has the smallest associated error. The vector corresponding to the largest singular value (W_1) is vector 1 in each panel. Figure 2 shows that the best-constrained parameter combination consists primarily of a , and to a lesser extent the two masses (left panel, vector 1). By inspecting the observable vector (center panel, vector 1), it is clear that p is the only component of the most reliable observable combination. We know that p is directly related to a and $M_A + M_B$ through Newton’s form of Kepler’s Third Law. Given that p has the smallest associated error, it should naturally give a well-determined value of a . M_A and M_B are mainly constrained by the radial velocity measurements (left and center panels, vector 3).

The second observable vector (center panel) shows that L_F and L_T are also important for constraining the solution. The second solution vector (left panel) shows that R_A and R_B are the parameters which are primarily determined by L_F and L_T . Note that the sixth vector (not shown) indicates that L_F and L_T also play a role in determining i , while the fourth vector (not shown) reveals that D_P is also important for determining R_A and R_B .

2.2.2 Uncertainties. Suppose that the errors in our measurements ϵ_i were reduced by a factor of two. By decreasing the ϵ_i individually, we can investigate how each of the parameter uncertainties σ_j consequently improves (this allows us to examine the relative impact of obtaining more precise measurements for each observable). Note that simultaneously reducing all of the ϵ_i by a factor of two will decrease all of the σ_j by a factor of two. Figure 2 (right panel) summarizes the relative decrease in each σ_j due to the reduction of ϵ_{D_P} (solid), and ϵ_{K_A} (dashed) by a factor of two.

The decrease in ϵ_{D_P} is entirely responsible for the reduction of both σ_{R_A} and σ_{R_B} . As noted above, the fourth vectors of \mathbf{U} and \mathbf{V} showed this relationship. Similarly, ϵ_{K_A} is responsible for all of the reduction of σ_{M_B} (see vector 3 in the left and center panels). However, ϵ_{K_A} is also *partially* responsible for reducing both σ_{M_A} and σ_a by 0.1 and 0.2 respectively. If we were interested in improving the precision of M_A further, we would need a combined decrease in the errors (not shown)—in this case, a reduction of ϵ_{K_B} , ϵ_{L_F} and ϵ_{L_T} together.

2.2.3 Solution. Returning to the model-fitting problem, given a set of observations \mathbf{O} we would like to find the best set of parameters \mathbf{P} that reproduce \mathbf{O} . Suppose that \mathbf{O} have no measurement errors and the model is physically complete. Using Eq.(1) with the \mathbf{P}_0 given in Table 1, we successfully calculate our fit parameters to be identical to \mathbf{P}_R . Now suppose that \mathbf{O} have measurement errors, i.e. there does not exist a set \mathbf{B} that will match the measurements exactly. If we use Eq.(1) as is, we will miscalculate the fit parameters to be \mathbf{P}_{F1} . However, if we discard the least important vector (setting the largest W^{-1} to 0) and then use Eq.(1), we will calculate the fit parameters \mathbf{P}_{F2} , which are better estimates than either \mathbf{P}_{F1} or \mathbf{P}_0 .

We will apply this method to observations of solar-type stars, with the \mathbf{P}_0 specified by the result of the genetic algorithm search. We will also repeat the local analysis using various combinations of the fundamental physical ingredients for the equations of state, opacities, and different prescriptions for mixing to quantify their contributions to the final parameter uncertainties. We are currently working to optimize the prototype local analysis by studying the transformation matrices between the observable and solution spaces over the full range of parameters, and we still need to automate this procedure and combine it with a local hill-climbing method for the pipeline.

3 Methodology

We have recently adapted the Aarhus stellar evolution code (ASTEC; Christensen-Dalsgaard, 2007) and adiabatic pulsation code (ADIPLS; Christensen-Dalsgaard, 2008) to interface with the parallel genetic algorithm. These are essentially the same models that were developed for the analysis of helioseismic data, and are the source of ‘Model S’ in Christensen-Dalsgaard et al. (1996), which has been used extensively as a reference model for solar inversions. Using these models for the analysis of pulsations in solar-type stars will provide some internal consistency in our understanding of solar-like oscillations. Each model evaluation involves the computation of a stellar evolution track from the zero age main-sequence through a mass-dependent number of internal time steps, terminating just prior to the beginning of the red giant stage. Rather than calculate the pulsation frequencies of each model along the track, we exploit the fact that the average frequency spacing of consecutive radial orders (the large separation $\Delta\nu_0$ in Figure 1) is a monotonically decreasing function of age (Christensen-Dalsgaard, 1993). Once the evolution track is complete, we start with a pulsation analysis of the model at the middle time step and then use a binary decision tree—comparing the observed and calculated values of $\langle\Delta\nu_0\rangle$ —to select older or younger models along the track. In practice, this recipe allows us to interpolate the age between the nearest two time steps by running the pulsation code on just 8 models from each track.

We have already developed a prototype model-fitting pipeline based on the codes mentioned above, and we have optimized the efficiency of the GA search by passing synthetic data through the fitting procedure (this phase of the project was funded by an NSF Fellowship to Metcalfe, prior to being hired by NCAR). The GA optimizes four adjustable parameters including the stellar mass (M_*) from 0.75 to 1.75 M_\odot , an initial metallicity (Z) from 0.002 to 0.05 (equally spaced in $\log Z$), an initial helium mass fraction (Y) from 0.22 to 0.32, and a mixing-length parameter (α) from 1 to 3. The stellar age (τ) is optimized internally during each model evaluation. The GA uses two-digit decimal encoding, so there are 100 possible values for each parameter within the ranges specified above. Each run of the GA evolves a population of 128 models through 200 generations to find the optimal set of parameters, and we execute 4 independent runs with different random initialization to ensure that the best model is truly the global solution (tests with synthetic data indicate that 2-3 out of the 4 runs typically find the global solution within 200 generations). This method requires about 10^6 model evaluations, compared to 10^8 models for a complete grid at the

same sampling density, making the GA 100 times more efficient than a complete grid (currently 1 week of computing time, compared to 2 years for a grid). The GA approach also gives us the flexibility to improve the physical ingredients in the future, while the physics of a grid are fixed.

We have validated the GA method using Sun-as-a-star data from the BiSON experiment (Chaplin et al., 1999) to verify that the model parameters emerging from the prototype resemble a standard solar model within reasonable tolerances (see additional details below). We now propose to refine this prototype into a general-purpose tool for the analysis of asteroseismic data from the Kepler mission, which will involve:

- **Phase 1:** Using archival ground-based data on other solar-type stars to validate the prototype pipeline for various stellar masses (e.g. α Cen A & B, at ~ 1.1 and $\sim 0.9 M_{\odot}$ respectively) and for different evolutionary stages (e.g. the “future Sun” β Hyi, at ~ 7 Gyr).
- **Phase 2:** Exploiting the better frequency resolution and complete coverage of Kepler data to include rotation velocity as an additional fitting parameter, which will be essential for the faster rotators where the splitting can lead to ambiguity in the mode identification.

Our experience with the development of the prototype suggests that we will be able to validate the pipeline with other solar-type stars and generalize it for space-based data during the proposed timeline (see section 4). We provide additional details for each of these phases of the project in the subsections below.

3.1 Validating the Treatment of Surface Effects

The biggest challenge to comparing the oscillation frequencies from theoretical models with those observed in solar-type stars are the systematic errors due to *surface effects*. The mixing-length parameterization of convection that is used in most stellar models is insufficient to describe the near-surface layers, and this leads to a systematic difference of several μHz (about 0.1% for a solar model) between the observed and calculated frequencies. The offset is independent of the spherical degree (ℓ) of the mode and grows larger towards the acoustic cutoff frequency. The 3D simulations of convection that might in principle reduce this discrepancy for individual stars are far too computationally expensive for the model-fitting approach that we are proposing. Instead, we adopt the method for empirical correction of surface effects described by Kjeldsen et al. (2008), which used the discrepancies between ‘Model S’ and GOLF data for the Sun (Lazrek et al., 1997) to calibrate the model. Although they demonstrated the method by applying it to models of several other stars (including α Cen A & B and β Hyi), our experience using it with solar data suggests that a *global* exploration of the models will present additional challenges. In this section we describe the successful adaptation of this method for our slightly different solar models and data, and we outline the steps that will be necessary to validate the method for other stars.

3.1.1 Calibration of the empirical model. The models that we have adopted for the pipeline include a slightly different set of physical ingredients than what was used to produce ‘Model S’ from Christensen-Dalsgaard et al. (1996). To characterize the surface effects, we need a model that matches the ‘Model S’ frequencies as closely as possible. Since finding such a model involves the comparison of two sets of model frequencies—both of which include a mixing-length parameterization of convection—there are no surface effects to consider. We assume that typical asteroseismic data from the Kepler mission will include twelve frequencies for each of the radial ($\ell = 0$), dipole ($\ell = 1$), and quadrupole ($\ell = 2$) modes, with consecutive radial orders in the range $n = 14$ -25. We further assume that we will have some constraints on the effective temperature and

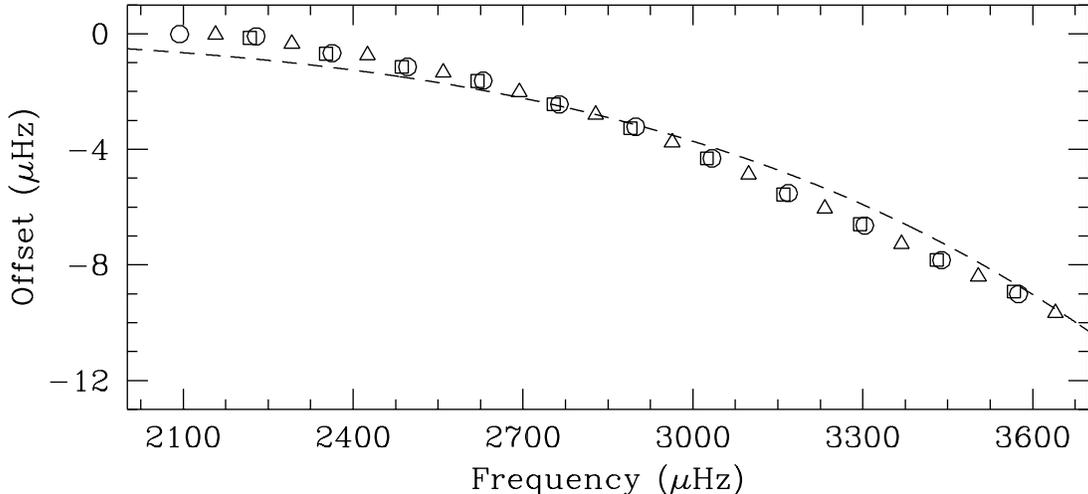


Figure 3: The offset due to surface effects between our fit to ‘Model S’ frequencies and the BiSON data for modes with $\ell = 0$ (circles) $\ell = 1$ (triangles) and $\ell = 2$ (squares). Also shown is the power law fit following Kjeldsen et al. (2008) to the radial modes (dashed line), which is used for all modes. This empirical surface correction can be applied to other stars with appropriate scaling.

luminosity from the Kepler Input Catalog (KIC) combined with a parallax from Kepler astrometry. We applied the GA method described above to a set of 36 oscillation frequencies from ‘Model S’, errors from BiSON data (scaled up by a factor of 10), and the solar effective temperature and luminosity with reasonable errors. The resulting optimal model from the GA had: $M_* = 0.99 M_\odot$, $Z = 0.019$, $Y = 0.277$, $\alpha = 2.04$, and $\tau = 4.56$ Gyr. Following Kjeldsen et al. (2008), we fit a power law to the differences between the radial modes of this model and the corresponding frequencies from BiSON data to characterize the surface effects (see Figure 3). We then applied the GA method to the BiSON data, using this empirical surface correction for all modes and scaling it for each model, leading to the optimal parameters: $M_* = 1.01 M_\odot$, $Z = 0.022$, $Y = 0.279$, $\alpha = 2.12$, and $\tau = 4.56$ Gyr. Note that we have not yet optimized this GA result with a local analysis, since the prototype local pipeline is still under development. With the successful validation of our prototype GA pipeline using solar data, we now need to ensure that our adopted treatment of surface effects yields reasonable optimal models when applied to stellar data.

3.1.2 Validation for different stellar masses. α Cen is the nearest stellar system to the Sun, containing primary and secondary components with masses near 1.1 and $0.9 M_\odot$ respectively. Since our stellar models and the empirical correction for surface effects have both been calibrated using a main-sequence star at $1.0 M_\odot$, the α Cen system will help validate the models with interior physical conditions that differ slightly from those of the Sun. Its proximity and multiple nature make it an excellent second test of our pipeline, because it has very well determined properties including stellar radii from interferometry (Kervella et al., 2003). There are also strong constraints on the component metallicities and effective temperatures (Porto de Mello et al., 2008), while the initial composition and age of the two stars are presumably identical. We will use the published oscillation frequencies of α Cen A (Bouchy & Carrier, 2002; Bedding et al., 2004) and α Cen B (Carrier & Bourban, 2003; Kjeldsen et al., 2005) with the additional constraints from interferometry, spectroscopy, and the binary nature of the system to validate our prototype pipeline and the empirical correction for surface effects. We will use SVD to investigate the role of the additional constraints in determining

the optimal model parameters, and to decide the relative weight that should be assigned to each observable. Kjeldsen et al. (2008) successfully applied their empirical correction for surface effects to a set of stellar models that broadly resemble the components of the α Cen system, so we have good reason to believe that our implementation will also succeed. Our key objective will be to minimize the systematic errors on the optimal model parameters, just as we did for solar data. To achieve this goal, the empirical correction must not only work *well* for models in a certain region of the search space—it must work *best* for the model that simultaneously matches all of the independent observational constraints within their uncertainties.

3.1.3 Validation at different evolutionary stages. The G2 subgiant β Hyi has long been studied as a “future Sun”, with an age near 7 Gyr. It has been characterized almost as extensively as the α Cen system, including recent interferometric measurements of its diameter (Thévenin et al., 2005; North et al., 2007) and dual-site asteroseismic observations that determined its mean density with an accuracy of 0.6% (Bedding et al., 2007). These data included the detection of several $\ell = 1$ modes that deviate from the expected frequency spacing, suggesting that they are “mixed modes” behaving like g -modes in the core and p -modes in the envelope. This is expected for evolved stars like β Hyi because as they expand and cool the p -mode frequencies decrease, while the g -mode frequencies increase as the star becomes more centrally condensed. This leads to a range of frequencies where these modes can overlap and exchange their character, manifested as so-called *avoided crossings* and mixed modes. This behavior changes very quickly with stellar age, and propagates from one mode to the next as a star continues to evolve. Consequently, the particular mode affected yields a very strong constraint on the age of the star (see Christensen-Dalsgaard, 2004). We will use the published oscillation data for β Hyi (Bedding et al., 2007) along with the constraints from interferometry and spectroscopy to validate the empirical treatment of surface effects for stars that are significantly more evolved than the Sun. We will devise an automated method to recognize mixed modes in the data set and to incorporate them into the optimization of stellar age along each track. Again we will use SVD to assign weights to the different observations, and our objective will be to minimize the systematic errors on the optimal model parameters. By the end of the first phase of the project our pipeline will be validated across a significant fraction of the H-R diagram where solar-like oscillations are excited.

3.2 Fitting for the Rotation Velocity

In the absence of rotation, the frequency of a stellar oscillation mode is determined by the radial order (n) and the spherical degree (ℓ). For non-radial modes ($\ell > 0$), rotation breaks the spherical symmetry and the $2\ell+1$ modes with different azimuthal order (m) are separated into multiplets with a frequency splitting that depends on the rotation rate. The detectability of the individual components depends on the inclination of the rotation axis relative to our line of sight, and on the size of the frequency splitting ($\sim 0.4 \mu\text{Hz}$ for the Sun) relative to the intrinsic width of the Lorentzian profile ($\sim 1 \mu\text{Hz}$ for the Sun) arising from the stochastic excitation of the oscillations (see Gizon & Solanki, 2003; Ballot et al., 2006, 2008). For solar-type stars that host transiting planets, the inclination of the orbital plane must be $\sim 90^\circ$ and we can assume that the rotation axis is perpendicular to our line of sight—making the characterization of multiplet structure less ambiguous. For other stars, the ratio of amplitudes within the multiplet helps constrain the inclination—which we can also use with planet-hosting stars to determine whether the rotation axis is actually perpendicular to the orbital plane, providing an interesting test of planetary formation models. To resolve the rotational splitting, a data set must have a time baseline at least as long as the rotation period of the star, so ground-based observations spanning a week or two can usually only detect statisti-

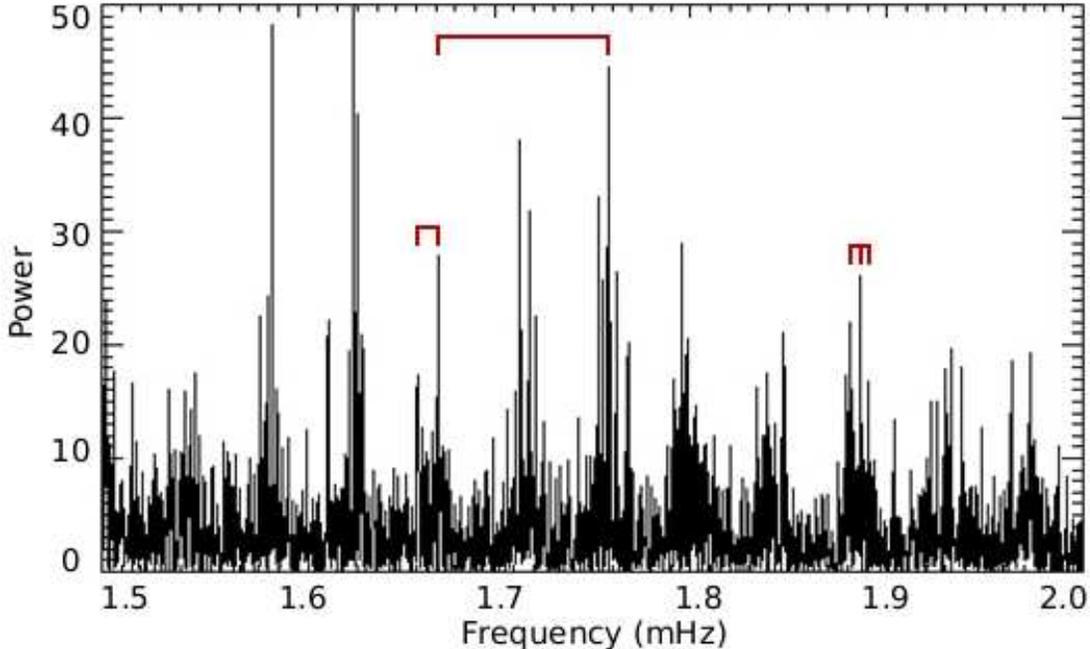


Figure 4: A power spectrum of the CoRoT observations of HD 49933, showing the characteristic large separation ($\Delta\nu_0$, top), small separation (δ_{02} , left) and rotational splitting (right). Adapted from an ESA press release in May 2007 (<http://www2.cnrs.fr/en/904.htm>).

cal signatures, as Bazot et al. (2007) recently showed for α Cen A. The longer time baseline and continuous monitoring that is possible with space-based observations will make rotational splitting a new observable. This potential was first demonstrated by Fletcher et al. (2006), who detected rotational splitting in α Cen A using a 50 day time series from the WIRE satellite. More recently, the CoRoT satellite monitored HD 49933 for 60 days and Appourchaux et al. (2008) simultaneously constrained the rotational splitting to be in the range 2.7-3.4 μHz and the inclination to be between 50-62° from these data (see Figure 4). In this section we describe our plans to extend the fitting method to include an additional parameter for the rotation velocity of the star, and to validate it using both synthetic data and observations from Kepler.

3.2.1 Extending the fitting method. From a practical standpoint, it will be straightforward to include the rotation velocity as an additional fitting parameter. The ASTEC code already has the ability to specify a rotation velocity, and the ADIPLS code is already configured to calculate oscillation frequencies using the second-order formalism of Gough & Thompson (1990), which may be necessary for the faster rotators. The efficiency of the GA search does not depend strongly on the number of free parameters, so extending the method from four (M_* , Z , Y , α) to five (the previous four plus v_{rot}) will only marginally increase the computational demands of the fitting method. Our experience making such a transition for models of white dwarf stars suggests that we will simply need to reduce the maximum mutation rate and allow each run of the GA to continue for a larger number of generations. For example, the optimal efficiency for the 3 and 4 parameter fits of Metcalfe et al. (2000, 2003) was achieved in 200-250 generations, while it was 300-400 generations for the 5 and 6 parameter fits of Metcalfe et al. (2001, 2005). Thus, the faster computer hardware that will be available through the TeraGrid by the second phase of the project will compensate for the expected increase in the optimal number of generations. The simulations of Ballot et al. (2008)

suggest that it will be straightforward to fit the frequencies of solar-type stars with rotation rates between 2-10 times the solar rate. For slower rotation it will be difficult to resolve the splitting within the intrinsic line widths, while faster rates will lead to ambiguity between rotational splitting and the small separation (δ_{02} in Figure 1). In the models that are used to fit peaks in the oscillation spectrum prior to our analysis, the rotation velocity is weakly correlated with the inclination. Any independent constraints on the system inclination will be applied by others during this phase of the data analysis, prior to our application of the stellar model-fitting pipeline. The inclination will primarily affect which components of the multiplet are detectable, leading to a possible ambiguity in the azimuthal order (m) assigned to each observed frequency. We will modify our code to accept an additional fitting parameter for the rotation velocity, and we will configure it to search a broad range of rotation rates. We will calculate the second-order corrected frequencies for each component of each multiplet and select the best match for each observed frequency. This approach will be essential for the faster rotators ($v_{\text{rot}} > 10 v_{\odot}$), where the $\ell = 2$ splitting can be large enough to overlap with the radial modes.

3.2.2 Testing with synthetic data. Beyond the mechanics of modifying our codes to include an additional fitting parameter for rotation, we will again need to pass synthetic data through the fitting procedure to optimize the efficiency of the method. This will involve one team member calculating the theoretical oscillation frequencies for a specific set of model parameters, and then giving a subset typical of what will be available from actual observations to another team member who does not know the source parameters. The frequencies are passed through the complete optimization method in an attempt to recover the source parameters without any additional information. The results of such *hare & hound* (H&H) exercises are used to quantify the success rate of the optimization method (the fraction of independent runs that lead to the known source parameters), and to improve its efficiency (minimize the number of model evaluations) if possible. We will conduct a series of H&H exercises spanning a representative range of model parameters. We will begin by using the exact frequencies from the model, and then add various levels of random Gaussian noise from the $\sim 0.1 \mu\text{Hz}$ expected from the best space-based observations up to the $\sim 1 \mu\text{Hz}$ that is more typical of current ground-based campaigns. We will also include some classical observables such as effective temperature and luminosity with their associated uncertainties, using SVD to help us assign weights. This will allow us to determine an optimal balance between the classical and asteroseismic observables at various signal-to-noise levels before we deal with the additional complications inherent in the real observations.

3.2.3 Validation with space-based data. NASA’s Kepler satellite is currently scheduled for launch in February 2009, and it will operate for a baseline mission of 3.5 years with a possible 2 year extension. Kepler will monitor 100,000 solar-type stars with a cadence of 30 minutes, and a revolving selection of 512 stars at 1 minute cadence for asteroseismology. New target lists can be uploaded every 30 days, and light curves resulting from the on-board photometric pipeline will be downloaded every 90 days. Within 90 days of each download, we will have access to the data through our involvement with the Kepler Asteroseismic Science Consortium (KASC), which is authorized by NASA and organized through the University of Aarhus in Denmark. Thus, we expect to have data for several hundred solar-type stars prior to the end of the first phase of this project. Each of the asteroseismic targets will continue to be monitored for the duration of the mission at 30 minute cadence to provide astrometry for a parallax, which is needed to derive an absolute luminosity. These data will also yield precise photometry that can be used for detailed spot modeling and the characterization of latitudinal differential rotation—a technique that was recently demonstrated with data from the MOST satellite (Walker et al., 2007). We will select several stars

from the initial Kepler data release showing various degrees of rotational splitting to validate our extended pipeline. We will focus on those stars that have an independent determination of the rotation period from detailed spot modeling, which will provide a check of the pipeline results. As a contingency plan, in case the Kepler mission does not follow the expected timeline, we will use the data on HD 49933 from the CoRoT satellite, which will be publicly available before the end of 2008. By the end of the second phase of the project we will have validated our pipeline to document asteroseismic signatures of stellar rotation from the high precision data that will emerge from Kepler and other space-based instruments.

4 Impact & Relevance

By the end of this project, we will have developed and validated an optimal, efficient, robust and unbiased model-fitting pipeline to target the vast quantities of asteroseismic data that will emerge from the Kepler mission. We are proposing this development at a time when it is just feasible using supercomputers, but computational power is expected to roughly quadruple during the course of our work. As a consequence, when we release our model-fitting pipeline for general use, we expect that it will be usefully deployed by individual investigators on small distributed computer clusters or high-end desktop workstations. Our timeline is targeted to maximize the science return from NASA's Kepler mission, though our pipeline should also be useful for analyses of existing ground-based data and other satellite data streams. It is primarily this large quantity of data that necessitates an automated approach to model-fitting, especially if we want to avoid the limitations (and computational inefficiency) of simply publishing a big grid of models.

The pipeline that we are proposing addresses several of NASA's strategic sub-goals. These specific sub-goals and science outcomes are:

- **Sub-goal 3B:** "Understand the Sun and its effects on the Earth and the solar system." Science outcome 3B.1: "progress in understanding the fundamental physical processes of the space environment from the Sun to Earth, to other planets, and beyond to the interstellar medium."
- **Sub-goal 3C:** "Advance scientific knowledge of the origin and history of the solar system, the potential for life elsewhere, and the hazards and resources present as humans explore space." Science outcomes 3C.1: "progress in learning how the Sun's family of planets and minor bodies originated and evolved," and 3C.2: "progress in understanding the processes that determine the history and future of habitability in the solar system, including the origin and evolution of Earth's biosphere and the character and extent of prebiotic chemistry on Mars and other worlds."
- **Sub-goal 3D:** "Discover the origin, structure, evolution, and destiny of the universe, and search for Earth-like planets." Science outcomes 3D.3: "progress in understanding how individual stars form and how those processes ultimately affect the formation of planetary systems," and 3D.4: "progress in creating a census of extra-solar planets and measuring their properties."

NASA's Kepler mission is designed to discover Earth-sized habitable planets. Using the techniques outlined in this proposal, our model-fitting pipeline will be able to characterize the planet-hosting stars with asteroseismology. This is essential to convert precise transit photometry into an absolute radius for the planetary body. In addition, accurate rotation rates and ages will provide clues

about the formation and evolution of the planet-hosting systems. The determination of accurate stellar parameters for a broad array of solar-type stars will provide us with new insights about stellar structure and evolution, including convection zone dynamics and long-term magnetic activity cycles, and will provide a broader context for our understanding of the Sun and our own solar system.

5 Plan of Work

A prototype of the GA portion of the pipeline is fully operational on the BlueGene/L system at NCAR, and has been validated with Sun-as-a-star observations. The local portion of the pipeline is currently in the final stages of development on a desktop platform. A postdoc (to be hired) will port it to the BlueGene/L system and automate the local analysis procedure for the pipeline before gradually taking a more central role in the validation efforts.

5.1 Key Milestones

Each of the two major phases of this project will require 18 months to complete, including the publication of results in a refereed journal. We will also publicize the availability of our pipeline at international meetings. The key milestones during each year will be:

- **Year 1:** Finish porting the local portion of the pipeline to the BlueGene/L architecture and integrate it with the GA pipeline for complete automation of the procedure.
- **Year 2:** Complete the validation of our treatment of surface effects through successful application of the prototype pipeline to ground-based data on α Cen A & B and β Hyi.
- **Year 3:** Successfully extend the fitting method to include rotation velocity as an additional parameter and validate with early data from the Kepler mission (contingency plan: use public data from the CoRoT satellite).

Our experience with the development of the prototype suggests that this is a reasonable timeline. These milestones rely on existing data for validation that are either already public (α Cen A & B and β Hyi) or will be available before they are needed (Kepler data released to the KASC, or CoRoT observations of HD 49933).

5.2 Management Structure

Much of the development work involving the Co-I in Denmark has already been completed during 18 months of funding through an NSF Astronomy & Astrophysics Postdoctoral Fellowship (to Metcalfe, prior to being hired by NCAR). Continued collaboration will be done primarily by email, but we also anticipate that Jørgen will continue his regular extended visits to Boulder each summer, funded through the HAO Visiting Scientist program.

Travis Metcalfe (PI): developed the parallel version of the PIKAIA genetic algorithm (GA), and ported it to the BlueGene/L architecture at NCAR which is now on the TeraGrid. His experience using the GA for asteroseismology started with his Ph.D. thesis on white dwarfs at the University of Texas-Austin. He is also the primary “experimental user” of the BlueGene/L system, and knows the technical aspects of parallel computing that will ensure the highest possible throughput for our pipeline. He will be responsible for further development of the GA, and for working with a postdoc (to be hired) on the local portion of the pipeline development.

Jørgen Christensen-Dalsgaard (Co-I): lead developer of the Aarhus stellar evolution code (ASTECC) and the adiabatic pulsation code (ADIPLS) that are being used for this project. He will be responsible for making any necessary updates to the interface and constitutive physics of these codes, and for coordinating the hare & hound exercises during the second phase of development. He leads the Kepler Asteroseismic Science Consortium (KASC), which is hosted at the University of Aarhus, and he has organized a group of collaborators (including Metcalfe) to help analyze the forthcoming data.

Postdoc with expertise in SVD (to be hired): will port the local portions of the pipeline to the BlueGene/L system, and will automate the local analysis using the GA result as a starting point. Will begin to coordinate the global analyses during the first phase of development, and will lead the validation efforts during the second phase. The postdoc will also play a prominent role in helping to write the journal papers and presenting the results at conferences.

5.3 Data Sharing Plan

Hundreds of scientists from around the globe will be involved in the analysis of asteroseismic data from the Kepler mission, but few will have the technical expertise or the computational resources to run our stellar modeling pipeline. We will release the source code for our pipeline during the first phase of the project, in time for the initial data release from Kepler. This will allow scientists with parallel computing access to apply our pipeline to their targets of choice. However, we are also currently developing a web-based user interface and workflow management software to fully automate the pipeline and link it to TeraGrid computing resources through a Science Gateway project. This development is supported at NCAR by the TeraGrid consortium, so we do not request any funds or computing resources from NASA for this work. We mention it here because it is a central component of our plan to make the proposed model-fitting pipeline generally available and broadly utilized. We will apply for a medium resource allocation (MRAC) from the TeraGrid to support the operation of the Science Gateway once it has completed development. The source code for some components of the proposed pipeline have already been released, e.g. the parallel genetic algorithm is available at <http://www.cisl.ucar.edu/css/staff/travis/mpikaia/> while the stellar pulsation package has for a long time been available at <http://www.phys.au.dk/~jcd/adipack.n/>. We have always made our software available to other investigators, so we anticipate no problems continuing this policy for all software and documentation developed as a result of this proposal.

References

- Appourchaux, T., Michel, E., Auvergne, M., Baglin, A., Toutain, T., Benomar, O. & Chaplin, W. (2008) *CoRoT sounds the stars: p-mode parameters of Sun-like oscillations on HD49933*. Astronomy & Astrophysics, in preparation.
- Asplund, M., Grevesse, N., Sauval, A. J., Allende Prieto, C. & Kiselman, D. (2004) *Line formation in solar granulation. IV. [O I], O I and OH lines and the photospheric O abundance*. Astronomy & Astrophysics, 417:751–768.
- Bäck, T. (1996) *Evolutionary Algorithms in theory and practice*. (Oxford: University Press).
- Ballot, J., García, R. A. & Lambert, P. (2006) *Rotation speed and stellar axis inclination from p modes: how CoRoT would see other suns*. Monthly Notices of the Royal Astronomical Society, 369:1281–1286.
- Ballot, J., Appourchaux, T., Toutain, T. & Guittet, M. (2008) *On deriving p-mode parameters for inclined solar-like stars*. Astronomy & Astrophysics, [arXiv:0803.0885].
- Bazot, M., Bouchy, F., Kjeldsen, H., Charpinet, S., Laymand, M. & Vauclair, S. (2007) *Asteroseismology of α Centauri A. Evidence of rotational splitting*. Astronomy & Astrophysics, 470:295–302.
- Bedding, T. R. & Kjeldsen, H. (2007) *Observing Solar-like Oscillations*. In: American Institute of Physics Conference Series. vol. 948, pp. 117–124.
- Bedding, T. R., Kjeldsen, H., Butler, R. P., McCarthy, C., Marcy, G. W., O’Toole, S. J., Tinney, C. G. & Wright, J. T. (2004) *Oscillation Frequencies and Mode Lifetimes in α Centauri A*. Astrophysical Journal, 614:380–385.
- Bedding, T. R., Kjeldsen, H., Arentoft, T., Bouchy, F., Brandbyge, J., Brewer, B. J., Butler, R. P., Christensen-Dalsgaard, J., Dall, T., Frandsen, S., Karoff, C., Kiss, L. L., Monteiro, M. J. P. F. G., Pijpers, F. P., Teixeira, T. C., Tinney, C. G., Baldry, I. K., Carrier, F. & O’Toole, S. J. (2007) *Solar-like Oscillations in the G2 Subgiant β Hydri from Dual-Site Observations*. Astrophysical Journal, 663:1315–1324.
- Bouchy, F. & Carrier, F. (2002) *The acoustic spectrum of alpha Cen A*. Astronomy & Astrophysics, 390:205–212.
- Brown, T. M., Gilliland, R. L., Noyes, R. W. & Ramsey, L. W. (1991) *Detection of possible p-mode oscillations on Procyon*. Astrophysical Journal, 368:599–609.
- Carrier, F. & Bourban, G. (2003) *Solar-like oscillations in the K1 dwarf star alpha Cen B*. Astronomy & Astrophysics, 406:L23–L26.
- Chaplin, W. J., Elsworth, Y., Isaak, G. R., Miller, B. A. & New, R. (1999) *Skew-symmetric solar P modes in low-l BiSON data*. Monthly Notices of the Royal Astronomical Society, 308:424–430.
- Charbonneau, P. (1995) *Genetic Algorithms in Astronomy and Astrophysics*. Astrophysical Journal Supplement, 101:309–344.
- Christensen-Dalsgaard, J. (1993) *On the Asteroseismic HR Diagram*. In: Brown, T. M. (ed.), ASP Conf. Ser. 42: GONG 1992. Seismic Investigation of the Sun and Stars. pp. 347–350.

- Christensen-Dalsgaard, J. (2002) *Helioseismology*. *Reviews of Modern Physics*, 74:1073–1129.
- Christensen-Dalsgaard, J. (2004) *Physics of solar-like oscillations*. *Solar Physics*, 220:137–168.
- Christensen-Dalsgaard, J. (2007) *ASTEC—the Aarhus STellar Evolution Code*. *Astrophysics & Space Science*, [arXiv:0710.3114].
- Christensen-Dalsgaard, J. (2008) *ADIPLS—the Aarhus adiabatic oscillation package*. *Astrophysics & Space Science*, [arXiv:0710.3106].
- Christensen-Dalsgaard, J. & Thompson, M. J. (1997) *On solar p-mode frequency shifts caused by near-surface model changes*. *Monthly Notices of the Royal Astronomical Society*, 284:527–540.
- Christensen-Dalsgaard, J., Dappen, W., Ajukov, S. V., Anderson, E. R., Antia, H. M., Basu, S., Baturin, V. A., Berthomieu, G., Chaboyer, B., Chitre, S. M., Cox, A. N., Demarque, P., Donatowicz, J., Dziembowski, W. A., Gabriel, M., Gough, D. O., Guenther, D. B., Guzik, J. A., Harvey, J. W., Hill, F., Houdek, G., Iglesias, C. A., Kosovichev, A. G., Leibacher, J. W., Morel, P., Proffitt, C. R., Provost, J., Reiter, J., Rhodes, Jr., E. J., Rogers, F. J., Roxburgh, I. W., Thompson, M. J. & Ulrich, R. K. (1996) *The Current State of Solar Modeling*. *Science*, 272:1286–1292.
- Darwin, C. (1859) *On the origin of species by means of natural selection*. (London: John Murray).
- Davis, L. (1991) *Handbook of genetic algorithms*. (New York: Van Nostrand Reinhold).
- Elsworth, Y. P. & Thompson, M. J. (2004) *Asteroseismology: Asteroseismology of Sun-like stars*. *Astronomy and Geophysics*, 45:14–19.
- Fletcher, S. T., Chaplin, W. J., Elsworth, Y., Schou, J. & Buzasi, D. (2006) *Frequency, splitting, linewidth and amplitude estimates of low- l p modes of α Cen A: analysis of Wide-Field Infrared Explorer photometry*. *Monthly Notices of the Royal Astronomical Society*, 371:935–944.
- Gizon, L. & Solanki, S. K. (2003) *Determining the Inclination of the Rotation Axis of a Sun-like Star*. *Astrophysical Journal*, 589:1009–1019.
- Goldberg, D. E. (1989) *Genetic algorithms in search, optimization and machine learning*. (Reading: Addison-Wesley).
- Gough, D. O. & Thompson, M. J. (1990) *The effect of rotation and a buried magnetic field on stellar oscillations*. *Monthly Notices of the Royal Astronomical Society*, 242:25–55.
- Holland, J. H. (1992) *Adaptation in natural and artificial systems*. (Cambridge, MA: MIT Press), 2nd ed.
- Kervella, P., Thévenin, F., Ségransan, D., Berthomieu, G., Lopez, B., Morel, P. & Provost, J. (2003) *The diameters of alpha Centauri A and B. A comparison of the asteroseismic and VINCI/VLTI views*. *Astronomy & Astrophysics*, 404:1087–1097.
- Kjeldsen, H., Bedding, T. R., Butler, R. P., Christensen-Dalsgaard, J., Kiss, L. L., McCarthy, C., Marcy, G. W., Tinney, C. G. & Wright, J. T. (2005) *Solar-like Oscillations in α Centauri B*. *Astrophysical Journal*, 635:1281–1290.
- Kjeldsen, H., Bedding, T. R. & Christensen-Dalsgaard, J. (2008) *Correcting Stellar Oscillation Frequencies for Near-Surface Effects*. *Astrophysical Journal Letters*, submitted.

- Lazrek, M., Baudin, F., Bertello, L., Boumier, P., Charra, J., Fierry-Fraillon, D., Fossat, E., Gabriel, A. H., García, R. A., Gelly, B., Gouiffes, C., Grec, G., Pallé, P. L., Pérez Hernández, F., Régulo, C., Renaud, C., Robillot, J.-M., Roca Cortés, T., Turck-Chièze, S. & Ulrich, R. K. (1997) *First Results on its P Modes from GOLF Experiment*. *Solar Physics*, 175:227–246.
- Metcalfe, T. S. (2003) *White Dwarf Asteroseismology and the $^{12}\text{C}(\alpha,\gamma)^{16}\text{O}$ Rate*. *Astrophysical Journal Letters*, 587:L43–L46.
- Metcalfe, T. S. & Charbonneau, P. (2003) *Stellar structure modeling using a parallel genetic algorithm for objective global optimization*. *Journal of Computational Physics*, 185:176–193.
- Metcalfe, T. S., Nather, R. E. & Winget, D. E. (2000) *Genetic-Algorithm-based Asteroseismological Analysis of the DBV White Dwarf GD 358*. *Astrophysical Journal*, 545:974–981.
- Metcalfe, T. S., Winget, D. E. & Charbonneau, P. (2001) *Preliminary Constraints on $^{12}\text{C}(\alpha,\gamma)^{16}\text{O}$ from White Dwarf Seismology*. *Astrophysical Journal*, 557:1021–1027.
- Metcalfe, T. S., Montgomery, M. H. & Kawaler, S. D. (2003) *Probing the core and envelope structure of DBV white dwarfs*. *Monthly Notices of the Royal Astronomical Society*, 344:L88–L92.
- Metcalfe, T. S., Montgomery, M. H. & Kanaan, A. (2004) *Testing White Dwarf Crystallization Theory with Asteroseismology of the Massive Pulsating DA Star BPM 37093*. *Astrophysical Journal Letters*, 605:L133–L136.
- Metcalfe, T. S., Nather, R. E., Watson, T. K., Kim, S.-L., Park, B.-G. & Handler, G. (2005) *An asteroseismic test of diffusion theory in white dwarfs*. *Astronomy & Astrophysics*, 435:649–655.
- Michalewicz, Z. (1992) *Genetic algorithms + Data structures = Evolution program*. (Berlin: Springer).
- Mitchell, M. (1996) *Introduction to genetic algorithms*. (Cambridge, MA: MIT Press).
- North, J. R., Davis, J., Bedding, T. R., Ireland, M. J., Jacob, A. P., O’Byrne, J., Owens, S. M., Robertson, J. G., Tango, W. J. & Tuthill, P. G. (2007) *The radius and mass of the subgiant star β Hyi from interferometry and asteroseismology*. *Monthly Notices of the Royal Astronomical Society*, 380:L80–L83.
- Porto de Mello, G. F., Lyra, W. & Keller, G. R. (2008) *The Alpha Centauri Binary System: Atmospheric Parameters and Element Abundances*. *Astronomy & Astrophysics*, [arXiv:0804.3712].
- Stello, D., Kjeldsen, H., Bedding, T. R., de Ridder, J., Aerts, C., Carrier, F. & Frandsen, S. (2004) *Simulating stochastically excited oscillations The mode lifetime of ξ Hya*. *Solar Physics*, 220:207–228.
- Thévenin, F., Kervella, P., Pichon, B., Morel, P., di Folco, E. & Lebreton, Y. (2005) *VLT/VINCI diameter constraints on the evolutionary status of δ Eri, ξ Hya, η Boo*. *Astronomy & Astrophysics*, 436:253–262.
- Walker, G. A. H., Croll, B., Kuschnig, R., Walker, A., Rucinski, S. M., Matthews, J. M., Guenther, D. B., Moffat, A. F. J., Sasselov, D. & Weiss, W. W. (2007) *The Differential Rotation of κ^1 Ceti as Observed by MOST*. *Astrophysical Journal*, 659:1611–1622.