

The BAYesian STellar algorithm (BASTA): a fitting tool for stellar studies, asteroseismology, exoplanets, and Galactic archaeology

V. Aguirre Børsen-Koch,^{1★†} J. L. Rørsted,^{1★‡} A. B. Justesen,¹ A. Stokholm^{1,2},^{1,2} K. Verma,¹ M. L. Winther,^{1★} E. Knudstrup^{1,3},¹ K. B. Nielsen,^{1,3} C. Sahlholdt^{1,4},^{1,4} J. R. Larsen,¹ S. Cassisi,^{5,6} A. M. Serenelli^{1,7,8},^{7,8} L. Casagrande^{1,9},⁹ J. Christensen-Dalsgaard,¹ G. R. Davies,^{1,10}^{1,10} J. W. Ferguson,¹¹ M. N. Lund^{1,11},¹ A. Weiss¹² and T. R. White^{11,13}

¹Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark

²Dipartimento di Fisica e Astronomia, Università degli Studi di Bologna, Via Gobetti 93/2, I-40129 Bologna, Italy

³Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany

⁴Lund Observatory, Department of Astronomy and Theoretical Physics, Box 43, SE-221 00 Lund, Sweden

⁵INAF-Astronomical Observatory of Abruzzo, Via M. Maggini sn, I-64100 Teramo, Italy

⁶INFN - Sezione di Pisa, Largo Pontecorvo 3, I-56127 Pisa, Italy

⁷Instituto de Ciencias del Espacio (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Cerdanyola del Valles, Spain

⁸Institut d'Estudis Espacials de Catalunya (IEEC), Gran Capita 4, E-08034, Barcelona, Spain

⁹Research School of Astronomy and Astrophysics, Mount Stromlo Observatory, The Australian National University, ACT 2611, Australia

¹⁰School of Physics and Astronomy, University of Birmingham, B15 2TT Birmingham, UK

¹¹Physics Department, Wichita State University, KS 67260-0032 Wichita, USA

¹²Max-Planck-Institut für Astrophysics, Karl Schwarzschild Strasse 1, D-85748, Garching, Germany

¹³Sydney Institute for Astronomy (SfA), School of Physics, University of Sydney, NSW 2006, Australia

Accepted 2021 October 1. Received 2021 September 30; in original form 2021 July 19

ABSTRACT

We introduce the public version of the BAYesian STellar Algorithm (BASTA), an open-source code written in PYTHON to determine stellar properties based on a set of astrophysical observables. BASTA has been specifically designed to robustly combine large data sets that include asteroseismology, spectroscopy, photometry, and astrometry. We describe the large number of asteroseismic observations that can be fit by the code and how these can be combined with atmospheric properties (as well as parallaxes and apparent magnitudes), making it the most complete analysis pipeline available for oscillating main-sequence, subgiant, and red giant stars. BASTA relies on a set of pre-built stellar isochrones or a custom-designed library of stellar tracks, which can be further refined using our interpolation method (both along and across stellar tracks or isochrones). We perform recovery tests with simulated data that reveal levels of accuracy at the few percent level for radii, masses, and ages when individual oscillation frequencies are considered, and show that asteroseismic ages with statistical uncertainties below 10 per cent are within reach if our stellar models are reliable representations of stars. BASTA is extensively documented and includes a suite of examples to support easy adoption and further development by new users.

Key words: asteroseismology – methods: numerical – methods: statistical – stars: fundamental parameters.

1 INTRODUCTION

Obtaining reliable properties of field stars is of paramount importance for many fields in astrophysics. An accurate characterization of exoplanets requires precise knowledge of the parent star radius and mass, the ultimate fate and evolutionary remnant of a star can only be determined if we know its initial mass, and the study of the formation and evolution of our Galaxy is incomplete without the distribution of stellar ages across the Milky Way. These are just examples of

how determining fundamental properties of stars has become the cornerstone of research for a variety of subjects.

Empirical methods to determine physical characteristics of stars are restricted to targets where e.g. years of monitoring are feasible (eclipsing binaries), or their bright apparent magnitude allows the measurement of the angular diameter (interferometry). This severely limits the number of stars where these techniques are applicable, and demands the development of methods in which some measured quantities (e.g. stellar effective temperature, surface composition, and luminosity) are compared to model predictions to infer stellar properties (such as age).

The advent of large-scale stellar surveys providing a myriad of data for thousands of stars across the Galaxy has led to the further development of algorithms that can combine different data sets to determine stellar properties. These algorithms vary in the approach

* E-mail: victor@phys.au.dk (VABK); jakob@phys.au.dk (JLR); mark@phys.au.dk (MLW)

† Formerly V. Silva Aguirre

‡ Formerly J. R. Mosumgaard

used to extract the final parameters of stars (e.g. machine-learning, neural networks, Bayesian inference), the method to determine uncertainties (e.g. confidence intervals, Gaussian errors, Monte Carlo sampling), and the set of stellar tracks or isochrones considered in the analysis. Moreover, the possible combinations of input data are different for these codes: while some rely on spectroscopic, photometric, and astrometric data (e.g. *StarHorse* (Queiroz et al. 2018), *MADE* (Das & Sanders 2019), others have the capability of including asteroseismic information [e.g. *PARAM* (Rodrigues et al. 2017), *Isoclassify* (Huber et al. 2017), and *AIMS* (Rendle et al. 2019)]. The latter point is of key importance: due to their dependence on the internal stellar structure, reproducing the observed pulsation properties allows for a determination of the stellar radius, mass, and age of solar-type stars and red giants to a level of precision that cannot be achieved when fitting only atmospheric properties. Asteroseismology has therefore become an invaluable tool for a large variety of studies thanks to the rapidly increasing amount of data available since the launch of the space-missions *CoRoT*, *Kepler*, and *TESS*, which will continue to ramp up as future missions such as *PLATO 2.0* (Rauer et al. 2014) begin to acquire data.

The irruption of asteroseismology in the scene of stellar properties determination poses tremendous challenges to fitting algorithms due to the large variety of oscillation quantities that one can try to reproduce. In stars whose driving mechanism is stochastic excitation from their outer convective envelopes (called solar-like oscillators, see e.g. Chaplin & Miglio 2013, for a review), the so-called global asteroseismic parameters can almost always be determined if oscillations are detected. If the data are of sufficient quality, individual frequencies of oscillation (or combinations of them) can be reproduced in main-sequence stars, as well as modes of mixed character that dominate the information content in subgiants. More evolved red giants present a much richer spectrum of pulsation, where state-of-the-art fitting algorithms reproduce only a subset of the observed frequencies while also considering a probe of the stellar core in the form of a gravity-mode period spacing. It is clear that fully exploiting the richness of asteroseismic data from solar-like oscillators requires algorithms capable of fitting a large variety of oscillation properties, which become relevant at different evolutionary stages and are strongly dependent on the data quality available. Moreover, these data should be supplemented with knowledge of the stellar effective temperature, a measurement of chemical composition, and ideally a determination of luminosity or absolute magnitude from astrometric and photometric data.

With this in mind we have developed the *B*AYesian *S*Tellar Algorithm (*BASTA*), originally introduced in Silva Aguirre et al. (2015). *BASTA* is a fitting tool written in *PYTHON* (Van Rossum & Drake 2009) designed to take advantage of the large variety of data obtained by large-scale ground-based surveys and space missions to precisely characterize stars. It has been built in a flexible way that allows the user to choose any combination of input data to be fit and, to the best of our knowledge, it is the fitting code that includes the largest number of global asteroseismic quantities and individual frequency diagnostics for solar-type stars, subgiants, and red giants. Other codes do not include information from individual oscillation frequencies (e.g. *Isoclassify* and *PARAM*), or glitch properties (e.g. *AIMS*), or simply do not take any asteroseismic input (e.g. *StarHorse* and *MADE*). Moreover, *BASTA* allows to simultaneously reproduce spectroscopic, photometric, astrometric, and asteroseismic data in a self-consistent manner, and is the only code where parallaxes can be fitted directly in addition to e.g. individual oscillation frequencies without the need of transforming

the astrometric information into a luminosity estimate. *BASTA* can run using publicly available compilations of stellar isochrones or tailor-made sets of evolutionary tracks with a wide combination of input physics.

BASTA has been extensively used to determine stellar properties of both asteroseismic and non-seismic exoplanet host stars discovered by *Kepler* (Silva Aguirre et al. 2015; Lundkvist et al. 2016; Bonomo et al. 2019), *K2* (Johnson et al. 2018; Persson et al. 2018; Van Eylen et al. 2018a; Hjorth et al. 2019a; Korth et al. 2019; Lund et al. 2019), *TESS* (Gandolfi et al. 2018; Huber et al. 2019), and *MASCARA* (Talens et al. 2018; Hjorth et al. 2019b). The sample of precise asteroseismic parameters originally derived in Silva Aguirre et al. (2015) has enabled detailed studies of e.g. exoplanet eccentricities (Van Eylen & Albrecht 2015) and the radius gap (Van Eylen et al. 2018b; see also Fulton et al. 2017). Since then, the applications of *BASTA* have been extended to a large variety of studies across fields of astrophysics such as characterization of asteroseismic targets (e.g. Serenelli et al. 2017; Silva Aguirre et al. 2017, 2020b; Stokholm et al. 2019), Galactic archaeology (e.g. Casagrande et al. 2016; Silva Aguirre et al. 2018; Spitoni et al. 2019, 2020; Nissen et al. 2020), open clusters (e.g. Lund et al. 2016; Stello et al. 2016; Arentoft et al. 2019), and the study of physical processes in stars such as rotation, convective overshoot, and magnetic activity (e.g. van Saders et al. 2016; Booth et al. 2017; Hjørringgaard et al. 2017). It has also been shown to be one of the most accurate pipelines available in tests using artificial data of main-sequence stars (Reese et al. 2016).

In this paper we introduce the public version of *BASTA*. We describe the Bayesian approach followed when determining stellar properties, the compilations of stellar track and isochrones available, the main features and capabilities included in the code, and present validation results of the fitting algorithm using artificial data.

2 THE BAYESIAN FRAMEWORK

We use Bayesian statistics for stellar properties inference. In this framework, Bayes' theorem combines our prior knowledge about the model stellar parameters Θ (which includes e.g. mass, radius, and age) with the information given by the data \mathbf{D} (such as measurements of effective temperature, metallicity, and oscillation frequencies), to provide the posterior probability distribution of model stellar parameters,

$$P(\Theta|\mathbf{D}) = \frac{P(\mathbf{D}|\Theta)P(\Theta)}{P(\mathbf{D})}. \quad (1)$$

Here, $P(\mathbf{D}|\Theta)$ or the likelihood is the probability of observing the data given the model parameters, $P(\Theta)$ or the prior is the probability of parameters without seeing the data, and $P(\mathbf{D})$ or the evidence is the total probability of observing the data (which is a normalizing constant).

We define the likelihood assuming Gaussian-distributed uncertainties on all observables except the magnitudes (see Section 4.2.2). *BASTA* is developed with great emphasis on enabling the user to fit a variety of observables including the ones coming from spectroscopy (e.g. effective temperature, T_{eff} , surface metallicity, $[\text{Fe}/\text{H}]$, and logarithm of surface gravity, $\log(g)$), astrometry and photometry (e.g. parallax, ϖ , and apparent magnitudes, m_z), and asteroseismology (e.g. large frequency separation, $\Delta\nu$, frequency of maximum power, ν_{max} , individual oscillation frequencies, ν , and their combinations, r_{01} , r_{10} , r_{02} , r_{012} , and r_{102}). Note that there are correlations among some of these observables which we account for in the fitting process using the corresponding covariance matrix.

To this purpose, we define the full likelihood as the product of likelihoods of groups of observables, D_i ,

$$P(\mathbf{D}|\Theta) = \prod_i P(D_i|\Theta). \quad (2)$$

The group likelihoods are computed using the expression (except for distances, see Section 4.2.2 below),

$$P(D_i|\Theta) = \frac{1}{\sqrt{2\pi|\mathbf{C}_i|}} \exp(-\chi_i^2/2), \quad (3)$$

where $|\mathbf{C}_i|$ is the determinant of the covariance matrix, and

$$\chi_i^2 = \frac{1}{N_i} (\mathbf{O}_{i,\text{obs}} - \mathbf{O}_{i,\text{mod}})^T \mathbf{C}_i^{-1} (\mathbf{O}_{i,\text{obs}} - \mathbf{O}_{i,\text{mod}}). \quad (4)$$

In equation (4), note the division by the number of observables, N_i . Although its inclusion is ad hoc in a statistical sense, it can be useful in artificially reducing the weight of a group of observables (normally the individual oscillation frequencies). In BASTA, the user can choose to turn-off this division.

The user can specify an informative prior on stellar mass as given by the initial mass function (IMF). There are several versions of IMF included in BASTA (Salpeter 1955; Miller & Scalo 1979; Kennicutt, Tamblyn & Congdon 1994; Scalo 1998; Kroupa 2001; Baldry & Glazebrook 2003; Chabrier 2003). To decrease the computation time, the user can pre-select a region of the grid for which the likelihoods are computed. This selection can be made on any available grid properties with user-defined tolerances. Technically, this is equivalent to assuming specific non-informative priors on certain stellar parameters.

We can use the computed posterior probability $P(\Theta|\mathbf{D})$ to derive the marginalized posterior for any model stellar parameter θ using the expression,

$$P(\theta|\mathbf{D}) = \int P(\theta, \Theta'|\mathbf{D}) w_{\Theta'} d\Theta', \quad (5)$$

where Θ' represents all the model parameters except θ . The weight $w_{\Theta'}$ is used to account for the volume of the parameter space occupied by the model characterized by Θ' , i.e. half the distance to its neighbouring points in all dimensions used to generate the grid (see further details in Section 3).

3 GRIDS OF STELLAR MODELS

The functionalities of BASTA rely on the use of collections of stellar evolution tracks or isochrones to extract the properties of stars by means of Bayesian inference. In its current version BASTA runs over publicly available sets of isochrones and tracks as well as custom-computed ones, which we process and store in Hierarchical Data Format version 5 (HDF5) and make available upon request. The functionalities and adopted input physics for each case are described in the following subsections.

3.1 BaSTI isochrones and tracks

Stellar properties can be determined with BASTA making use of the recently updated library ‘a Bag of Stellar Tracks and Isochrones’ (BaSTI¹). This compilation consists of evolutionary tracks and isochrones that are available for four distinct science cases defined by the inclusion of different physical processes as given in Table 1. We give a brief description of its main features in this section, and

¹<http://basti-iac.iaa.es>

Table 1. Science cases of BaSTI tracks and isochrones available in BASTA. Columns show combinations of convective core overshooting (λ_{ov}), microscopic diffusion (D_{diff}), and mass loss (η). See the text for details.

Case	λ_{ov}	D_{diff}	η
1	No	No	No
2	Yes	No	No
3	Yes	No	Yes
4	Yes	Yes	Yes

refer the reader to Hidalgo et al. (2018) and Pietrinferni et al. (2021) for additional details.

The prescription for convective core overshooting consists of instantaneous mixing beyond the region formally defined by the Schwarzschild criterion, keeping the radiative temperature gradient in this region. In the case of main-sequence models with convective core, the overshoot region is defined by the distance $\lambda_{\text{ov}} \times H_p$, where H_p is the local pressure scale height and λ_{ov} is a free parameter. It has been set equal to 0.2, decreasing to zero when the mass decreases below a certain value. The approach used for decreasing λ_{ov} from its maximum value to zero depends on both the chemical composition and stellar mass (see section 2.2 of Hidalgo et al. 2018, for details). During the core-helium burning stage, regardless of the considered science case, core mixing is modelled by accounting for semiconvection and suppression of breathing pulses.

The science case including atomic diffusion follow the prescription of Thoul, Bahcall & Loeb (1994), while mass-loss is taken into account in the formulation of Reimers (1975) with an efficiency of the free parameter η set to 0.3. The temperature stratification in the outer stellar layers is given by the Vernazza, Avrett & Loeser (1981) formulation, but for the case of very-low mass stellar models for which outer boundary conditions based on accurate model atmospheres have been adopted (see Hidalgo et al. 2018; Pietrinferni et al. 2021, for a detailed discussion on this topic). For the adopted physical framework and solar heavy element distribution (see Section 4.2.1 below), the calibration of a Standard Solar Model (SSM) sets the value of the mixing length parameter α_{MLT} , and of the initial solar metallicity and He abundance. At the solar age, the BaSTI SSM² matches the solar luminosity and radius as well as the present $(Z/X)_{\odot}$ abundance ratio with a value for the mixing length parameter $\alpha_{\text{MLT}} = 2.006$ (we refer to Hidalgo et al. 2018, for details about the properties of the BaSTI SSM).

The BaSTI stellar models are available both for the Caffau et al. (2011) solar heavy element mixture, and for an α -element enhanced mixture³ ($[\alpha/\text{Fe}] = +0.4$). For each selected metallicity, the corresponding initial helium abundance has been obtained by adopting an He-enrichment ratio equal to $\Delta Y/\Delta Z = 1.31$, and a primordial He abundance equal to $Y_p = 0.247$. From the calibration of an SSM the resulting initial solar abundances are $Z_{\text{ini}} = 0.01721$ and $Y_{\text{ini}} = 0.2695$. From the compilation of 21 metallicities initially available with the release of the BaSTI models, we have increased the resolution in chemical composition and age by interpolating metallicity points for all science cases using the available online routine provided by the BaSTI team.

The current release of the BaSTI library also contains evolutionary tracks spanning masses from $0.1 M_{\odot}$ to $15 M_{\odot}$. For a subset of

²We note that the BaSTI SSM has been calculated starting from the pre-main sequence and including diffusion of He and heavy elements.

³The computation of a stellar model grid for an α -depleted mixture is in progress.

these (ranging from $\sim 0.7 M_{\odot}$ to $4.0 M_{\odot}$ with the lower limit being metallicity dependent) the full interior structure is also provided throughout the evolution. The number of structures stored⁴ at each mass varies with metallicity, ranging from ~ 300 individual models at the high-mass end to more than 9000 structures at the low-mass end. These are stored in the standardized `fgong` format described in the website⁵ of the Aarhus Red Giants Challenge (Silva Aguirre et al. 2020a). The availability of interior structures allows us to compute oscillation frequencies and determine asteroseismic observables as described in Section 4.1.

Our grids of BaSTI isochrones and tracks reach ages up to 16 Gyr, which allow us to properly construct the posterior distributions of old stars without risking the appearance of an edge effect (see e.g. Valle et al. 2014, 2015).

3.2 Custom-computed evolutionary tracks

While the use of publicly available compilations of tracks and isochrones (such as BaSTI) makes it easy to compare the BASTA results with those from other fitting codes, it poses a limit on our flexibility to explore different combinations of input physics and numerical implementations. For this reason, we developed algorithms that calculate and process user-defined grids of evolutionary tracks with any combination of input physics and make them readily usable with BASTA. The limitations on the input physics are given by the features available in an evolutionary code, and currently we support the usage of GARSTEC (Weiss & Schlattl 2008) and the latest publicly available version of MESA (Paxton et al. 2011, 2013, 2015).

In the case of GARSTEC, our running version of the code has experienced a number of developments since the published version of Weiss & Schlattl (2008). Updated nuclear reactions from Solar Fusion II are available (Adelberger et al. 2011), and electron screening of nuclear reactions now also covers the intermediate regime using the prescriptions of Dewitt, Graboske & Cooper (1973) and Graboske et al. (1973). The definition of convective boundaries follows the recipe of Gabriel et al. (2014), and includes a treatment of semiconvective mixing as described in Silva Aguirre et al. (2011). GARSTEC can couple on-the-fly distilled information from the *Stagger* grid of 3D-hydrodynamical simulations of stellar atmospheres during the evolution (Jørgensen et al. 2018; Mosumgaard et al. 2018, 2019; Jørgensen & Weiss 2019). The prescription in GARSTEC for overshoot consists of a diffusive process with a diffusion constant given by

$$D(z) = D_0 \exp\left(\frac{-2z}{f H_p}\right), \quad (6)$$

where the constant D_0 is derived from the mixing length theory convective velocities, z is the radial distance from the edge of the convective zone, f is a free efficiency parameter, and H_p is the pressure scale height. For small convective cores the overshooting efficiency is limited using a geometrical cut-off factor g_{cut} that scales the local pressure scale height as follows:

$$H_p = \min\left\{H_p, H_p \left(\frac{R_{cz} f}{g_{\text{cut}} H_p}\right)^2\right\}. \quad (7)$$

Here R_{cz} is the radial thickness of the convective zone. The default value of the free parameter is $g_{\text{cut}} = 2$, and it can be modified to

allow finer control over the size of small convective cores, a desired feature in e.g. studies constraining the extent of convective cores using asteroseismic data (Silva Aguirre et al. 2013; Deheuvels et al. 2016).

We have two different approaches for grid sampling. In the first, which is the conventional approach, we compute tracks on a predefined mesh of stellar parameters. The mesh points are typically equally spaced along each parameter. The stellar model grids calculated in this manner are known as Cartesian grids. In the second approach, we sample the parameter space uniformly using a quasi-random number generator described in Sobol (1967). Note that quasi-random number generators perform better than pseudo-random number generators, specifically in high-dimensional spaces, as the latter provides more clumpy distributions. We refer to grids computed in this way as Sobol grids.

We can generate Cartesian and Sobol grids over a number of stellar parameters including mass (M_{ini}), initial helium abundance (Y_{ini}), initial metallicity ($[\text{Fe}/\text{H}]_{\text{ini}}$), alpha enhancement ((α/Fe)), mixing-length (α_{MLT}), and overshoot and mass-loss (η). The parameters used to generate the grid define the dimension of the weight w_{Θ} included in our marginalized posterior distribution (see equation 5). Note that we construct grids over $[\text{Fe}/\text{H}]_{\text{ini}}$ (instead of initial metal mass fraction), because this quantity is well constrained by the observed metallicity, and hence allows convenient choice of the parameter space over which one needs to calculate a grid to model the observed star. We can either treat Y_{ini} as a free parameter similar to other stellar parameters, or determine it from the $[\text{Fe}/\text{H}]_{\text{ini}}$ assuming values for the primordial helium abundance (default $Y_p = 0.248$; Fields et al. 2020), and the helium-to-metal enrichment ratio (default $\Delta Y/\Delta Z = 1.4$). The default values can be changed by the user. In the case of GARSTEC, we have implemented alpha enhanced and depleted mixtures of the Asplund et al. (2009) solar abundances, ranging from $[\alpha/\text{Fe}] = -0.2$ to $[\alpha/\text{Fe}] = 0.6$ in steps of 0.1. We adopt consistent opacity tables from OPAL for high-temperatures (Iglesias & Rogers 1996) and the Ferguson et al. (2005) opacities in the low-temperature regime.

4 AVAILABLE FITTING PARAMETERS

One of the core features of BASTA is its ability to handle inhomogeneities in the available data as long as they are contained in the grid model parameters. BASTA can easily deal with heterogeneous input and provide a robust set of stellar properties based on the likelihood of the models in its grids. In the following sections, we describe the quantities available for fitting asteroseismic, photometric, spectroscopic, and astrometric data. A complete and up-to-date list of code parameters can be found in the code documentation (available on GitHub).

4.1 Asteroseismology

The wealth of data from main-sequence and red giant stars provided by asteroseismic space missions has driven the development of BASTA towards the study of solar-like oscillators. These are stars whose oscillations are excited by the same mechanism as in the Sun, and comprise the vast majority of targets where asteroseismic quantities are available. For solar-like oscillators, there are several observables that the user can select to be reproduced by the models and fit with BASTA. In all grids of stellar models currently supported (see Section 3), theoretical oscillation frequencies have been computed using the Aarhus adiabatic oscillation package (ADIPLS; Christensen-Dalsgaard 2008).

⁴The complete data base of interior structures is publicly available at the BaSTI URL repository.

⁵https://github.com/vaguirrebkoeh/aarhus_RG_challenge

4.1.1 Global asteroseismic quantities

The number of asteroseismic properties that can be extracted from the power spectrum of a given star depends on the length of the observations, the target's apparent magnitude, and its evolutionary state (since the time-scale for oscillations scales with the intrinsic luminosity, see e.g. Kjeldsen & Bedding 1995, 2011). If the data reveal the signal of the oscillations, two basic seismic observables that can be readily extracted are the average large frequency separation $\langle \Delta\nu \rangle$ and the frequency of maximum power ν_{\max} . These quantities, also known as the global asteroseismic parameters, scale approximately with stellar properties as follows:

$$\left(\frac{\langle \Delta\nu \rangle}{\langle \Delta\nu_{\odot} \rangle} \right)^2 \simeq \frac{\bar{\rho}}{\bar{\rho}_{\odot}}, \quad (8)$$

$$\frac{\nu_{\max}}{\nu_{\max, \odot}} \simeq \frac{M}{M_{\odot}} \left(\frac{R}{R_{\odot}} \right)^{-2} \left(\frac{T_{\text{eff}}}{T_{\text{eff}, \odot}} \right)^{-1/2}, \quad (9)$$

where $\langle \Delta\nu_{\odot} \rangle$, $\nu_{\max, \odot}$, and $T_{\text{eff}, \odot}$ are the values measured in the Sun. As $\langle \Delta\nu \rangle$ and ν_{\max} are normalized to these reference solar values, they can be specified by the user. By default, BASTA adopts $\langle \Delta\nu_{\odot} \rangle = 135.1 \mu\text{Hz}$ and $\nu_{\max, \odot} = 3090 \mu\text{Hz}$ from Huber et al. (2011).

There is extensive literature devoted to the testing and validating the scaling relations by using independent constraints in stellar masses and radii from e.g. binary stars, clusters, interferometry, and parallaxes (see e.g. Huber et al. 2012, 2017; Miglio et al. 2012; Silva Aguirre et al. 2012; White et al. 2013; Gaulme et al. 2016; Brogaard et al. 2018; Sahlholdt & Silva Aguirre 2018; Sahlholdt et al. 2018, to name a few). There is general agreement that the scaling relations are accurate to within a few percent, but as it is clear from equations (8) and (9) these extrapolations from the solar values do not take into account variations with chemical composition nor the evolutionary stage of the star.

To address some of these issues, several prescriptions have been proposed to correct the $\langle \Delta\nu \rangle$ scaling relation, while the equation for ν_{\max} is purely empirical and no corrections across the range of interest in T_{eff} and $[\text{Fe}/\text{H}]$ have been derived yet (but see Belkacem et al. 2011, for an initial explanation on the theory behind equation 9). We have implemented in BASTA four additional determinations of the average large frequency separation that aim at further decreasing the level of systematic deviation in the scaling relation equation (8). A detailed comparison between the performances of each prescription can be found in Viani et al. (2019).

The first two determinations consist of a linear fit as a function of radial order to the model individual frequencies of $\ell = 0$ weighted by a Gaussian centred at ν_{\max} , the only difference being the adopted Full Width at Half Maximum of $0.25\nu_{\max}$ (White et al. 2011) or $0.66\nu_{\max}^{0.88}$ (Mosser et al. 2012a). The average large frequency separation $\langle \Delta\nu \rangle$ is determined from the slope of this fit and is meant to mimic as close as possible the manner in which this quantity is derived from the observations. This average large frequency separation is available for the grids of stellar models where we compute the individual oscillation frequencies (see Sections 3.1 and 3.2). As a by-product of this calculation we obtain the dimensionless offset ϵ that can be used to correct for differences between the observed and model radial order (see Section 4.1.2 below).

The other two determinations are those of Sharma et al. (2016) and Serenelli et al. (2017) who, based on the same principle of the White et al. (2011) approach, have computed a correction factor across the Hertzsprung–Russell Diagram for the value of $\langle \Delta\nu \rangle$ obtained from equation (8) that depends on the mass, metallicity,

effective temperature, and evolutionary state (hydrogen-core or shell burning or core-helium burning) of the star. These corrections can be computed for any grid of stellar tracks or isochrones, as they are independent of the availability of theoretical oscillation frequencies.

As a final remark, we note that when defining the solar reference values $\langle \Delta\nu_{\odot} \rangle$ and $\nu_{\max, \odot}$ in equations (8) and (9) there is an implicit assumption that fitting a target with those values of $\langle \Delta\nu \rangle$ and ν_{\max} (and solar temperature and metallicity) will result in a star of $1 M_{\odot}$ and $1 R_{\odot}$ (but not necessarily solar age, as this depends on the input physics used to construct the models). To ensure this level of consistency, all theoretical values of $\langle \Delta\nu \rangle$ in our grids of models are re-scaled by a factor given by the fraction $\langle \Delta\nu_{\odot} \rangle / \langle \Delta\nu_{\odot} \rangle_{\text{grid}}$, where $\langle \Delta\nu_{\odot} \rangle_{\text{grid}}$ is the average large frequency separation computed from the individual $\ell = 0$ modes of a solar model computed with the same input physics of the corresponding grid using the full width at half maximum (FWHM) of White et al. (2011).

4.1.2 Individual oscillation frequencies

If the time-series of observations is of sufficient signal-to-noise ratio and resolution, it is possible to extract individual oscillation frequencies characterized by radial order n and angular degree ℓ (see e.g. Davies et al. 2016; Lund et al. 2017). For stellar disc-integrated observations, as it is the case of the space missions, geometrical cancellation suppresses the signal from all modes except those with low degree $\ell \leq 3$. If the asymptotic theory can be applied to describe the oscillation (e.g. Tassoul 1980), modes of odd and even degree are separated by $\sim \langle \Delta\nu \rangle / 2$. As a result, the observed oscillation spectrum contains one mode of each degree $\ell = 1, 2, 3$ between two consecutive $\ell = 0$ modes. Departures from this asymptotic description occur e.g. in the presence of mixed modes (see Section 4.1.5).

Most current stellar evolutionary models use a rudimentary description of the outer-most layers of stars with convective envelopes (e.g. the mixing-length theory of Böhm-Vitense 1958), leading to systematic frequency shifts in the oscillation modes when compared to observations (Christensen-Dalsgaard et al. 1996). In order to correct for this effect, a so-called surface correction needs to be applied to the model frequencies. There are three prescriptions for the surface corrections implemented in BASTA and all include power-law dependence on frequency. They aim at obtaining the corrected model frequencies $\nu_{n, \ell}^{\text{cor}}$ from the original model frequencies $\nu_{n, \ell}^{\text{model}}$ by determining the corresponding fitting coefficients of the power-law correction to make them as close as possible to the observed frequencies $\nu_{n, \ell}^{\text{obs}}$.

The first prescription is the empirical power-law correction from Kjeldsen, Bedding & Christensen-Dalsgaard (2008):

$$r \nu_{n, \ell}^{\text{cor}} - \nu_{n, \ell}^{\text{model}} = \frac{a}{Q} \left(\frac{\nu_{n, \ell}^{\text{model}}}{\nu_0} \right)^b, \quad (10)$$

in which a and r are the fitting coefficients, b is a fixed exponent, and ν_0 is a reference frequency, typically chosen to be the frequency of maximum power ν_{\max} . Q is the ratio between the inertia of the mode and the theoretical inertia that a radial ($\ell = 0$) mode would have at that frequency, determined by linear interpolation.

The other implemented corrections are the two physically motivated surface corrections from Ball & Gizon (2014), first their *cubic* correction:

$$\nu_{n, \ell}^{\text{cor}} - \nu_{n, \ell}^{\text{model}} = \frac{a_3 (\nu_{n, \ell}^{\text{model}} / \nu_0)^3}{I}, \quad (11)$$

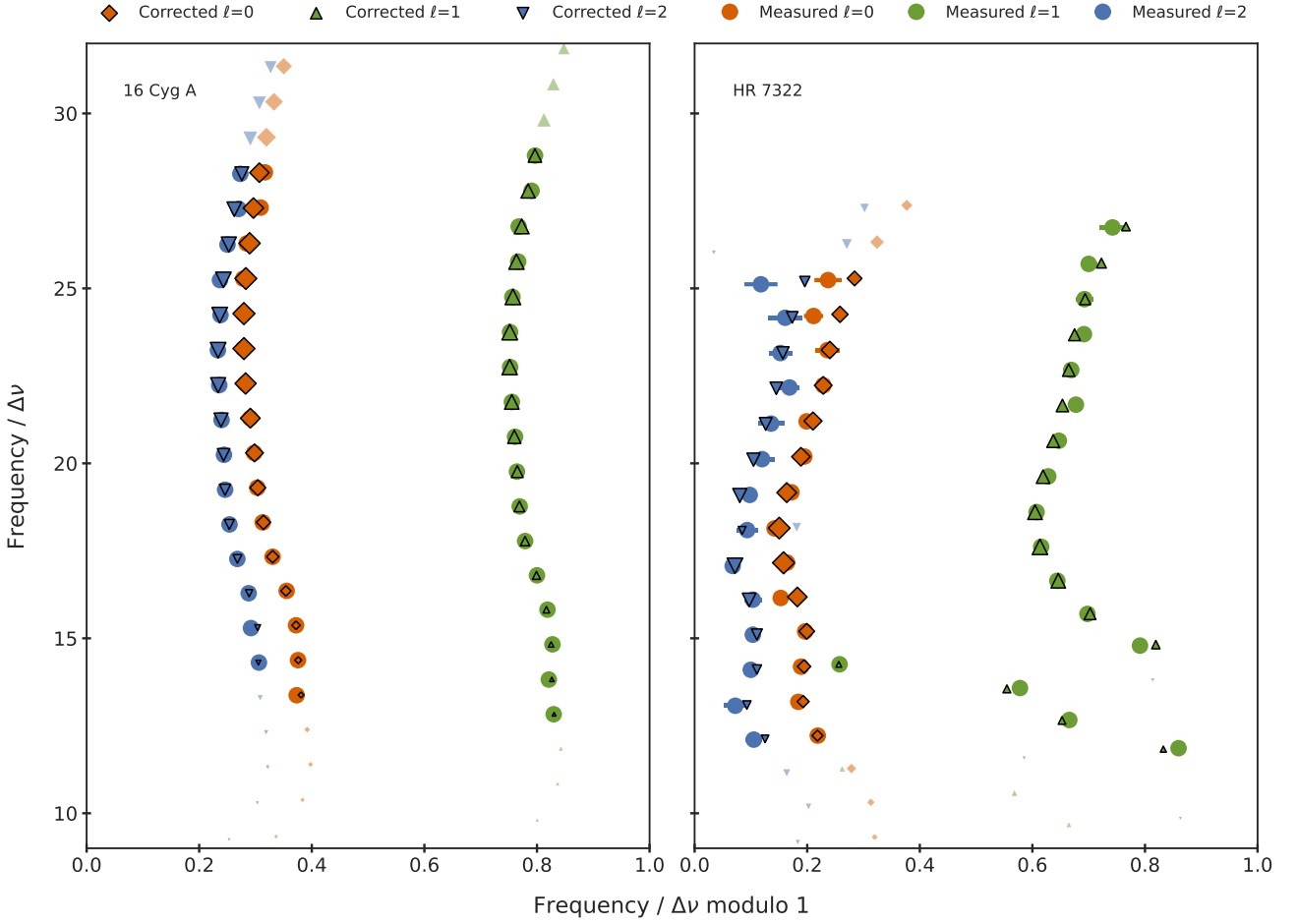


Figure 1. Échelle diagram of the main-sequence star 16 Cyg A and the subgiant star HR 7322. The coloured circles show the observed oscillation modes (red: $\ell = 0$, green $\ell = 1$, blue: $\ell = 2$) from Lund et al. (2017) and Stokholm et al. (2019), respectively, while the coloured symbols with a black outline show the modes predicted from model with the highest likelihood (same colour coding), corrected using the combined surface correction from Ball & Gizon (2014) and scaled using the observed $\Delta\nu$ of the star. The size of the symbols from the model is scaled inversely with their normalized mode inertias: the larger the symbol, the greater the probability of the mode being observed. The lighter coloured symbols with no outline are not matched to any observation, but still predicted by the model.

and secondly their *combined* correction, adding the cubic correction from above to an inverse term:

$$v_{n,\ell}^{\text{cor}} - v_{n,\ell}^{\text{model}} = \frac{a_{-1}(v_{n,\ell}^{\text{model}}/v_0)^{-1} + a_3(v_{n,\ell}^{\text{model}}/v_0)^3}{I}. \quad (12)$$

In these corrections, a_{-1} and a_3 denotes the fitting coefficients, and I is the scaled mode inertia, typically normalized at the stellar surface as (Aerts, Christensen-Dalsgaard & Kurtz 2010),

$$I = \frac{4\pi \int_0^R [|\tilde{\xi}_r|^2 + \ell(\ell+1)|\tilde{\xi}_h|^2] \rho_0 r^2 dr}{M [|\tilde{\xi}_r(\tilde{R})|^2 + \ell(\ell+1)|\tilde{\xi}_h(\tilde{R})|^2]}, \quad (13)$$

where ξ_r and ξ_h are the radial and horizontal components of the displacement, ρ_0 is the unperturbed stellar density, M is the total stellar mass, and R is the photospheric radius.

When fitting individual oscillation frequencies, it is necessary to correctly match each observed mode with its corresponding theoretical counterpart of identical radial order n as the surface correction and thus the computation of the likelihood of the model depend on the difference in frequency between the observed and modelled frequency of the same radial order and angular degree. Identifying the radial order in the observations of main-sequence

stars is relatively straight forward using the dimensionless offset ϵ (White et al. 2012), and BASTA has an option for the user to apply a suitable correction to the radial order of the observed frequencies, if desired. An example of the best-fitting model found by BASTA when fitting the individual oscillation frequencies of the *Kepler* target 16 Cyg A is shown in Fig. 1.

4.1.3 Frequency combinations

Combinations of frequencies have been long used in asteroseismic analysis to isolate the signature of a given stellar region and extract detailed information about the structure of a star (e.g. Roxburgh & Vorontsov 2003; Oti Floranes, Christensen-Dalsgaard & Thompson 2005; Cunha & Metcalfe 2007; Silva Aguirre et al. 2011). A simple example of this is the large separation between modes of same angular degree and consecutive overtone, $\Delta\nu_\ell(n) = \nu_{n,\ell} - \nu_{n-1,\ell}$, which is related to the mean stellar density (cf. equation 8). In BASTA we have included a number of combinations as fitting parameters such as the small-frequency separation:

$$d_{02}(n) = \nu_{n,0} - \nu_{n-1,2}, \quad (14)$$

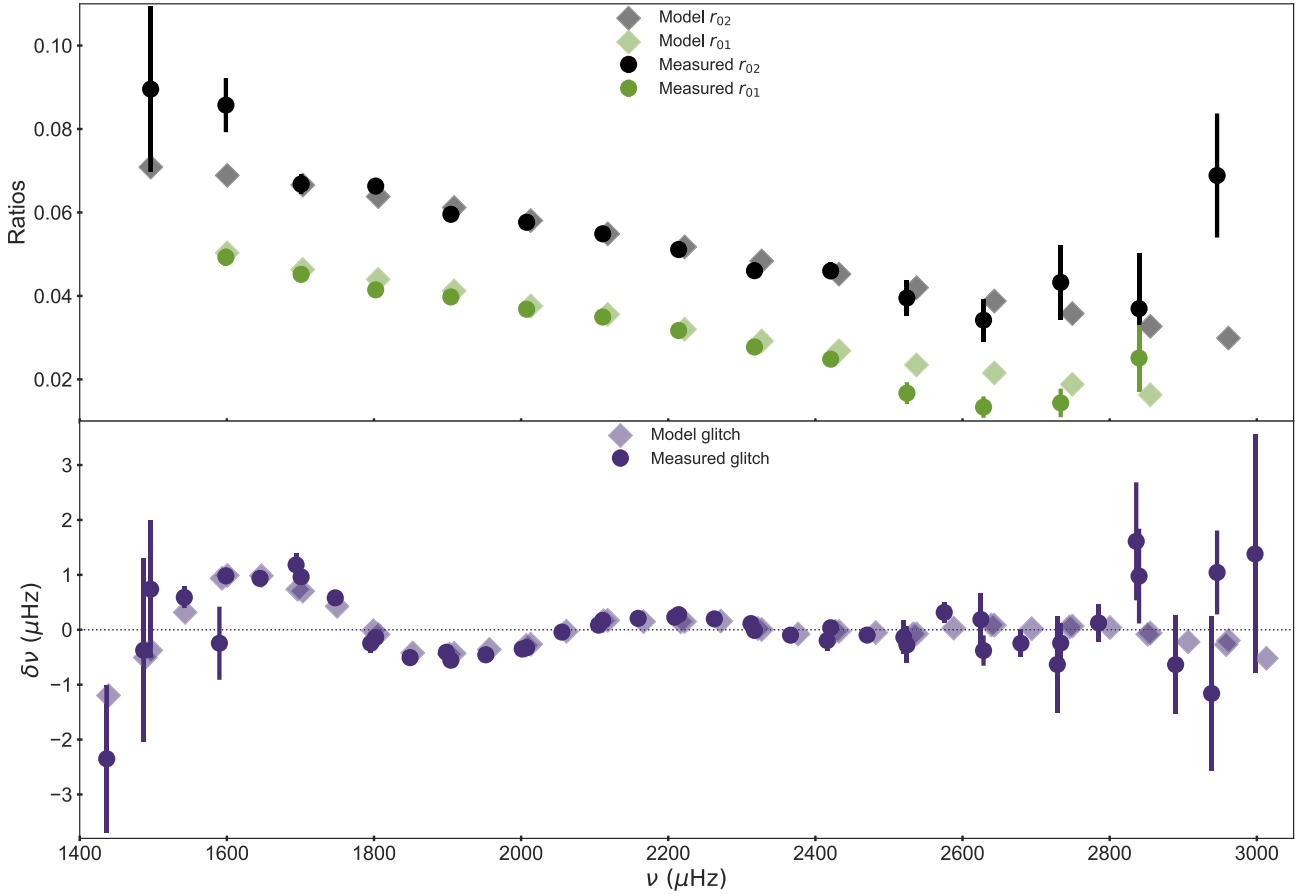


Figure 2. Ratios and glitch signatures for 16 Cyg A as a function of oscillation frequencies. In the top panel, the black and green circles represent the observed ratios r_{02} and r_{01} , respectively, while the black and green diamonds show the corresponding quantities for the best-fitting model (see the legend). In the bottom panel, the circles represent the sum of the observed glitch signatures from the helium ionization zone and the base of convection zone, while the diamonds show the same for the best-fitting model.

the 5-point small-frequency separations:

$$d_{01}(n) = \frac{1}{8}(v_{n-1,0} - 4v_{n-1,1} + 6v_{n,0} - 4v_{n,1} + v_{n+1,0}) \quad (15)$$

$$d_{10}(n) = -\frac{1}{8}(v_{n-1,1} - 4v_{n,0} + 6v_{n,1} - 4v_{n+1,0} + v_{n+1,1}), \quad (16)$$

the frequency separation ratios:

$$r_{02}(n) = \frac{d_{02}(n)}{\Delta v_1(n)} \quad (17)$$

$$r_{01}(n) = \frac{d_{01}(n)}{\Delta v_1(n)}, \quad (18)$$

$$r_{10}(n) = \frac{d_{10}(n)}{\Delta v_0(n+1)}, \quad (19)$$

and the set of ratios r_{010} , r_{012} , and r_{102} :

$$r_{010} = \{r_{01}(n), r_{10}(n), r_{01}(n+1), r_{10}(n+1), \dots\}, \quad (20)$$

$$r_{012} = \{r_{01}(n), r_{02}(n), r_{01}(n+1), r_{02}(n+1), \dots\}, \quad (21)$$

$$r_{102} = \{r_{02}(n), r_{10}(n), r_{02}(n+1), r_{10}(n+1), \dots\}. \quad (22)$$

There are strong correlations between combinations including five individual frequencies, and Deheuvels et al. (2016) showed that the set r_{010} results in almost singular covariance matrices with large condition numbers that can lead to overfitting the data as recently suggested by Roxburgh (2018). Instead, the latter study recommends

the usage of r_{01} or r_{10} in combination with r_{02} to form the series r_{012} or r_{102} , respectively. Here, the underlying problem is in the use of conventional numerical methods to estimate the inverse of covariance matrices, which turns out to be highly inaccurate if these are ill-conditioned. We overcome this issue by using the Moore–Penrose pseudo-inverse (see e.g. Strang 2006) of the covariance matrices in the likelihood. The pseudo-inverse is calculated using singular value decomposition and sets singular values to zero when these are below a threshold that is defined relative to the largest singular value.

The approach devised in BASTA to handle frequency combinations attempts to give the user as much flexibility as possible while keeping the statistical approach robust, and allows fitting any of these quantities (d_{01} , d_{10} , d_{02} , r_{01} , r_{10} , r_{02} , r_{010} , r_{012} , r_{102}) as desired. In all cases the user only needs to provide the individual oscillation frequencies and BASTA will calculate the needed combinations. It is possible to supply the code with the necessary correlations across terms, and if these are not given then BASTA calculates them by doing 10 000 Monte Carlo realizations drawn from random Gaussian distributions of the individual frequencies. An example of the fit to the frequency ratios obtained for 16 Cyg A is shown in the top panel of Fig. 2.

4.1.4 Acoustic glitches

There are regions inside solar-like oscillating stars where the sound speed changes at length scales substantially shorter than the local

wavelengths of the acoustic waves. These regions are known as acoustic glitches in the stellar structure, and the two most prominent are the helium ionization zone and the base of the envelope convection zone. Glitches in the acoustic structure of stars leave tiny signatures in the observed oscillation frequencies currently detectable from space borne missions (see e.g. Miglio et al. 2010; Mazumdar et al. 2012) as well as from ground-based facilities (see e.g. Bedding et al. 2010; Grundahl et al. 2017). The bottom panel of Fig. 2 shows an example of glitch signatures as seen in the oscillation frequencies of 16 Cyg A observed by the *Kepler* satellite.

The detection of the glitch signatures can provide useful information about stellar interior, e.g. they can be used to measure the location of the base of envelope convection zone as well as to infer the surface helium abundance (see e.g. Mazumdar et al. 2014; Verma et al. 2014, 2017, 2019; Verma & Silva Aguirre 2019). The perturbation to the oscillation frequencies due to acoustic glitches can be derived using the asymptotic theory of stellar oscillations (see e.g. Houdek & Gough 2007),

$$\delta\nu = A_{\text{He}}\nu e^{-8\pi^2\Delta_{\text{He}}^2\nu^2} \sin(4\pi\tau_{\text{He}}\nu + \psi_{\text{He}}) + \frac{A_{\text{CZ}}}{\nu^2} \sin(4\pi\tau_{\text{CZ}}\nu + \psi_{\text{CZ}}), \quad (23)$$

where the two terms on the right-hand side represent contributions from the helium and base of the convection zone glitches, respectively. The parameters A_{He} , Δ_{He} , τ_{He} , and ψ_{He} depend on the properties of the helium ionizing layers, whereas A_{CZ} , τ_{CZ} , and ψ_{CZ} depend on the properties of the base of the convection zone. The parameter τ_{CZ} is of particular importance as it provides an estimate of the acoustic depth of the base of the convective envelope. Another quantity of interest is the average amplitude of helium glitch signature,

$$\langle A_\nu \rangle = \frac{\int_{\nu_1}^{\nu_2} A_{\text{He}}\nu e^{-8\pi^2\Delta_{\text{He}}^2\nu^2} d\nu}{\int_{\nu_1}^{\nu_2} d\nu} = \frac{A_{\text{He}}[e^{-8\pi^2\Delta_{\text{He}}^2\nu_1^2} - e^{-8\pi^2\Delta_{\text{He}}^2\nu_2^2}]}{16\pi^2\Delta_{\text{He}}^2[\nu_2 - \nu_1]}, \quad (24)$$

which has been used in the past to measure the envelope helium abundance of solar-type stars.

We have implemented in *BASTA* the capability to fit the observed glitch parameters. The helium and convection zone glitch parameters are extracted from oscillation frequencies using the Method A of Verma et al. (2017) and Verma et al. (2019). Briefly, in Method A, we modelled the total oscillation frequency,

$$f(n, \ell) = \sum_{k=0}^4 b_k(\ell)n^k + \delta\nu, \quad (25)$$

where the first term represents the smooth component of oscillation frequencies while the second term arises from glitches (see equation 23). The polynomial coefficients $b_k(\ell)$ are determined together with the glitch parameters by fitting equation (25) to the oscillation frequencies using non-linear optimization method based on Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (see e.g. Fletcher 1987). To estimate the uncertainties in the glitch parameters, we repeat the fitting process using 1000 realizations of the individual oscillation frequencies and estimate the full covariance matrix. The uncertainties in the glitch parameters correspond to the square root of the diagonal terms in the matrix, which are consistent with the error bars obtained from the Hessian matrix.

In the current implementation, we do not use the parameters associated with the base of convection zone glitch in the stellar properties determination as it is typically difficult to reliably extract

them from the contemporary frequency precision for two reasons: (1) the small amplitude of the base of convection zone signature, and (2) the issue caused by aliasing (Mazumdar & Antia 2001). Having said that, it is straightforward to modify the current implementation to use the glitch parameters related to the base of convection zone.

4.1.5 Mixed modes and the frequency matching routine

As stars evolve beyond the core-hydrogen burning phase their oscillation pattern develop irregularities: when the helium core contracts as a product of stellar evolution, the frequency of the g modes increases and interactions between the p-mode behaviour near the surface and the g-mode behaviour near the core take place. Modes can experience mixed properties and exchange nature, causing these so-called mixed modes to deviate from the regular oscillation pattern of the p modes and thus to be visible in an échelle diagram as avoided crossings.

Mixed modes have a substantial diagnostic potential as they are sensitive to properties of the stellar core. However, mixed modes complicate the analysis of individual frequencies as this departure from simple asymptotic theory results in the presence of more than one non-radial mode of a given angular degree between two consecutive $\ell = 0$ modes, and some of these theoretically predicted mixed modes do not reach observable amplitudes. This complicates the matching of model modes to observed modes, which can lead to incorrect computations of the surface correction and of the likelihood evaluation of the given model.

To address this issue we note that the amplitude of a mode can be roughly estimated from the mode inertia (see equation 13). The frequency fitting procedure in *BASTA* uses a mode matching algorithm, where the observed modes are matched to their most-likely counterpart in the model using their frequencies as well as their inertias as a proxy of the likelihood of observability. We describe the matching procedure and assumptions in the following paragraphs.

The frequency matching routine is based on mode counting. Even though the extracted frequencies can be uncertain at times as they depend on the power spectrum background and systematic effects in the pipeline, these effects do not change the relative ordering of the modes. An $\ell = 1$ mode and an $\ell = 0$ mode will not exchange order due to effects such as the surface effect. Due to the physical nature of the radial modes, avoided crossings do not occur in the pattern of the $\ell = 0$ modes. We therefore use the observed radial modes to define a number of frequency intervals. If higher angular degree modes are observed outside the frequency range encompassed by the radial modes, the frequency binning is extended to lower and higher frequencies in steps of $\Delta\nu$.

BASTA counts how many modes of a given angular degree $\ell \neq 0$ are present in each frequency bin between two consecutive radial modes. When there is an equal number of model and observed modes, the modes are matched based only on frequency. We note in passing that this is the typical case for main-sequence stars that do not present avoided crossings. If there are more observed modes than modelled modes in a given range, the model is rejected as it fails to accurately describe the observed pattern.

If, between two $\ell = 0$ modes, j modelled modes exist with inertias $I_1 < \dots < I_j$ and k observed modes of same angular degree with $j > k$, we need to select k modelled modes to match one-to-one to the observed modes. This is usually the case when avoided crossing takes place in the model, but not all mixed modes have high enough amplitudes to be observed. Intuitively, one might think to just pick the k modelled modes with the lowest inertia as they should have the

highest likelihood of being observed. However, small differences in inertia might cause this to result in a miss-matching. Instead, BASTA selects two inertia thresholds a and b (with $a < b$) and subdivides the modelled modes into a set A with inertias less than a , a set B with inertia between a and b , and a set C for inertias greater than b . The set A thus contains modes that are likely to be detected, while set C contains modes that are unlikely to be observed, and set B contains modes that are somewhere in between.

By picking all modes in A and a subset of B such that k modes are selected in total, the modes can be matched one-to-one in frequency to the observed modes. Specifically, the thresholds a and b are chosen based on the k 'th smallest modelled inertia I_k : a is $I_k/10$ and b is $10I_k$. This ensures that $|A| \leq k \leq |A| + |B|$, where $|A|$ is the number of modes in the set. Each possible match is evaluated based on the total L_1 distance in frequency space and the subset of B that minimizes this metric is chosen. This match between the observed and modelled modes will then be used in the following surface effect and model likelihood computation.

Fig. 1 shows the échelle diagrams for two examples of asteroseismic fitting to individual frequencies using this matching algorithm. The code can nicely reproduce the observed oscillation pattern of the main-sequence star 16 Cyg A, and also follow the rapid evolution of a distinct dipole mixed mode in the bright F6 subgiant star HR 7322. Further examples of matching the mixed-mode pattern in subgiant stars are given in Section 6 and Appendix B below.

4.1.6 Period spacing

The observed power spectra of red-giant stars exhibit a complex pattern due to the presence of mixed modes. As described above, mixed modes behave as gravity modes in the inner regions of the star and as acoustic modes in the outer layers, and hence their observations provide a unique opportunity to probe the conditions deep in the stellar core (see e.g. Beck et al. 2011; Bedding et al. 2011; Mosser et al. 2012b). The detection of the mixed modes makes it possible to measure the gravity mode period spacing, and currently measurements of the asymptotic period spacing for the dipole modes are available for several thousands of *Kepler* red-giant stars (see e.g. Stello et al. 2013; Mosser et al. 2014; Datta et al. 2015; Vrad, Mosser & Samadi 2016).

We can use the observed dipole mode asymptotic period spacing in BASTA as a quantity to be fitted. The corresponding asymptotic period spacing for the models is computed according to the formula,

$$\Delta P_1 = \sqrt{2}\pi^2 \left(\int \frac{N}{r} dr \right)^{-1}, \quad (26)$$

where N is the Brunt-Väisälä frequency, r the radial coordinate, and the integration is performed over the radiative interior. A few necessary considerations regarding the integral in equation (26): (1) the integrand has a numerical singularity, and (2) tabulated values of r have variable step size. These make rectangle and trapezoidal rules for the integration inaccurate (particularly if the step size is not very small), and the Newton-Cotes formulas with higher order accuracy inapplicable. For this reason, we use an adaptive Gauss-Kronrod quadrature method (7-points Gauss rule combined with 15-points Kronrod rule) to compute the integral with high accuracy (Kronrod 1965). This requires the values of the integrand at intermediate r (other than the tabulated values), which is obtained using the basis spline interpolation.

4.2 Atmospheric properties

4.2.1 Surface chemical composition

Fitting the observed surface chemical composition requires certain assumptions about the relation between measured number density ratios of a given element and the metal mass fraction used to construct stellar models. For a given grid of evolutionary tracks or isochrones characterized by hydrogen, helium, and metal mass fractions ($X + Y + Z = 1$), BASTA assumes the following mapping:

$$[M/H] = \log_{10} \left(\frac{(Z/X)_{\text{model}}}{(Z/X)_{\odot}} \right) \quad (27)$$

$$[\text{Fe}/H] = [M/H] - \text{corr}([\alpha/\text{Fe}]). \quad (28)$$

The equations above depend on the solar heavy element distribution adopted to construct the grid, and a correction factor $\text{corr}([\alpha/\text{Fe}])$ which is determined by comparing stellar tracks with and without the inclusion of alpha-elements enhancement (see Salaris, Chieffi & Straniero 1993). These correction factors, valid when all alpha-elements are equally enhanced, change according to the considered solar mixture. We have determined them for the fixed alpha enhancements of $[\alpha/\text{Fe}] = +0.4$ in the Grevesse & Noels (1993) and Grevesse & Sauval (1998) solar mixtures, as well as all the available values of $[\alpha/\text{Fe}]$ for the Asplund et al. (2009) solar composition (see Section 3.2). We adopt $\text{corr}([\alpha/\text{Fe}]) = 0.3016$ as given by Pietrinferni et al. (2021) when dealing with the Caffau et al. (2011) compilation, which is appropriate for an alpha-enhancement value of $[\alpha/\text{Fe}] = +0.4$.

Under these assumptions, all custom-constructed grids used in BASTA are mapped from the input $[\text{Fe}/H]_{\text{ini}}$ and $[\alpha/\text{Fe}]$ into $[M/H]$ using equation (28), and then the ratio $(Z/X)_{\text{model}}$ is determined using equation (27). The BaSTI models on the other hand follow the inverse procedure where we use their initial mass fractions of X and Z as given in their data base to determine $[M/H]$ following equation (27), and $[\text{Fe}/H]$ is then calculated for the corresponding case of $[\alpha/\text{Fe}]$ using equation (28) and the solar mixture of Caffau et al. (2011). This procedure ensures that all surface abundance ratios are computed in a consistent manner, but it is up to the user to ensure that the combination of input values of $[\text{Fe}/H]_{\text{ini}}$ and $[\alpha/\text{Fe}]$ are reasonable for observed stars. We note in passing that the surface abundance by mass of elements such as ^3He , ^{12}C , ^{13}C , ^{14}N , and ^{16}O are also available in our custom-computed grids.

4.2.2 Synthetic photometry, parallaxes, and distances

All grids of stellar models available in BASTA (cf. Section 3) can be mapped from the theoretical H-R diagram to various photometric systems using bolometric corrections (BCs) tables provided by Hidalgo et al. (2018). This allows us to determine synthetic magnitudes in more than 15 photometric systems including those that are of relevance for asteroseismic and exoplanet studies (e.g. *Kepler* and *TESS* passbands), compilations of bright stars (Tycho-2 and *Hipparcos*), large photometric surveys (e.g. 2MASS, Skymapper, sloan, and VISTA), and naturally all Gaia data releases. A list of the photometric systems currently available in BASTA is provided in Table A1.

BASTA supports the inclusion of parallaxes as a fitting parameter together with at least one apparent magnitude m_{ζ} . The evaluation is then done by comparing the grid-model absolute magnitude M_{ζ} , computed from the bolometric luminosity and effective temperature using bolometric corrections, to the measured absolute magnitude computed from the apparent magnitude m_{ζ} , an estimate

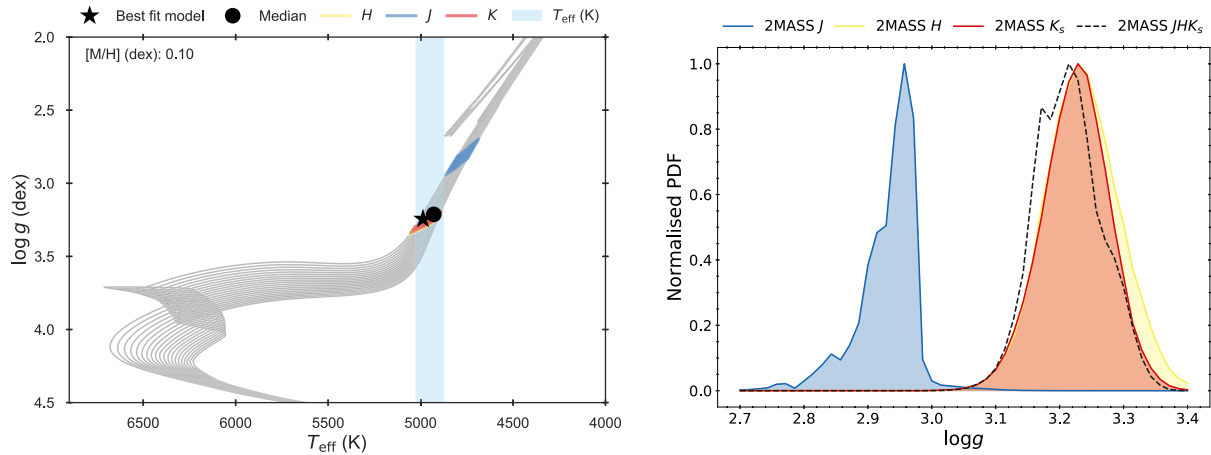


Figure 3. Left: Kiel diagram depicting BaSTI isochrones colour-coded according to their correspondence to each fitted parameter. Right: Resulting $\log(g)$ probability density function as predicted in each filter, as well as the joint distribution obtained from equation (2); dashed line). See the text for details.

of extinction, and the distance modulus from the observed parallax $\mu = 5 \log(d) - 5$. The computation of the observed absolute magnitude undergoes multiple transformations and thus the assumption of this value being normally distributed like the other fitting parameters is weak. Instead this parameter is evaluated by constructing the likelihood distribution of M_ζ and including it in the calculation of the posterior using equation (2).

The procedure when parallax is included as a fitting parameter is as follows. In the first step, BASTA constructs prior distributions in distance and apparent magnitude using their observed values. The distance distribution is analytically calculated from the measured parallax and its associated uncertainty using the exponentially decreasing space density prior with a scale length of 1.35 kpc described in Bailer-Jones (2015) and Astraatmadja & Bailer-Jones (2016). For the apparent magnitude m_ζ , we assume a normal distribution with the observed values of m_ζ and its uncertainty as the mean and standard deviation, respectively.

In the second step, BASTA samples over the distance and apparent magnitude distributions to calculate the reddening. When multiple distributions are considered it is difficult to properly sample the tails of all distributions if one just draws samples following the distributions and let the number density of the samples determine the probability density function. If we draw samples from two normal distributions, then the odds of drawing values at e.g. 3 standard deviations away from the mean in both distributions is less than 1 in 100 000. This means that the tails of the resulting distribution would be artificially steep. To overcome this, BASTA draws the samples systematically over a large range of values for each given parameter.

For a number N of samples in distance, BASTA draws half of them linearly across a range of $d \in [10^{(-0.4)}, 10^{(4.4)}]$ pc, and the other half as quantiles of the normal distribution around the mode of the distance distribution derived from the observed parallax as described above. Similarly, we produce K apparent magnitude samples distributing half of them linearly across the limiting magnitudes published by each relevant survey (or else assume $m_\zeta \in [-10, 25]$), and the rest from a normal distribution around the observed apparent magnitude.

For each of the $N \times K$ pairs of distance and apparent magnitude we determine the reddening $E(B - V)$ using the latest version of the Bayestar dust map (currently Bayestar19, see Green et al. 2019), and if the target falls outside of its coverage we apply the simpler map of Schlegel, Finkbeiner & Davis (1998). Since Bayestar provides multiple estimates of the colour excess at each distance, we determine the individual reddening values using the

median and standard deviation of those samples. To transform the reddening values into absorption A_ζ estimates at a given filter ζ , we use extinction coefficients from table 6 in Schlafly & Finkbeiner (2011) assuming the Cardelli, Clayton & Mathis (1989) relation $A_\zeta = 3.1 \times E(B - V)$. For filters not contained in that compilation, we default to the temperature and metallicity dependent extinction coefficients from Casagrande & Vandenberg (2014).

We compute the absolute magnitudes M_ζ from our $N \times K$ groups of distance, apparent magnitude, and absorption using the distance modulus. This sample is converted into a probability distribution function by weighting each obtained absolute magnitude by the underlying observed distance and apparent magnitude probability distributions. This distribution is then included in the computation of the posterior (see equation 2). We note that in cases where the parallax uncertainty is smaller than 5 per cent and the absolute magnitude prior is symmetric (i.e. if the distance between the median and each quantile is within a predefined threshold), we fit a Gaussian function to the probability distribution of absolute magnitudes and use the analytical expression in the computation of the likelihood.

Fig. 3 shows an example of the distance sampling procedure. The fitting includes effective temperature, metallicity, and parallax using the 2MASS filters. For the purpose of this example we have modified the apparent J -magnitude to be in disagreement with the other two magnitudes, as it can be seen in the Kiel diagram. The right-hand panel of Fig. 3 shows the probability density function of $\log(g)$ predicted by each of the filters. If BASTA did not sample the tails of the magnitude distributions far from their median values and standard deviations, the full likelihood as defined in equation (2) will be zero. The designed sampling scheme avoids this singularity and provides a robust statistical solution.

In addition to fitting parallax directly, distances can be determined in BASTA independent of any parallax input as long as one photometric apparent magnitude is provided to the code. In this case, we solve for the distance and absorption iteratively using the magnitudes and the dust map until convergence is reached (normally within three iterations). If multiple apparent magnitudes are given as input we derive a distance in each passband and determine a joint distance by multiplying the individual probability distribution functions.

5 INTERPOLATION

If the resolution of the grid used with BASTA is lower than what is desired, the grid can be interpolated to match a user-defined

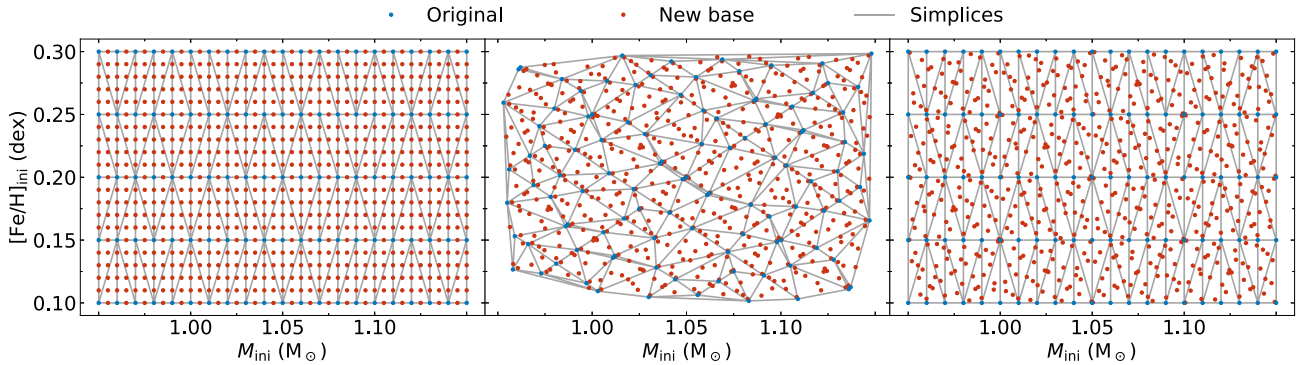


Figure 4. Example of how new tracks are assigned for interpolation across tracks for the case of a base consisting of M_{ini} and $[\text{Fe}/\text{H}]_{\text{ini}}$. The simplices show the tessellation of the original grid, where each new track is interpolated to by using the tracks forming the simplex it is contained within. The left-hand panel shows the assignment of tracks for a Cartesian input grid with a starting resolution of 0.05 dex in initial metallicity and $0.01 M_{\odot}$ in initial mass, and a desired interpolated resolution of 0.01 dex and $0.005 M_{\odot}$, respectively. The middle panel shows the assignment of tracks for a Sobol input grid with an increase in the number of tracks by a factor of 2. The right-hand panel shows the assignment of tracks for a Cartesian input grid but with Sobol assignment of tracks with an increase in resolution by a factor of 2. See the text for details.

resolution. BASTA includes this option and can interpolate along or across stellar evolution tracks or isochrones (or both along and across, which we refer to as the combined option), and every parameter of the grid can be included in this interpolation. The individual procedures for each type of interpolation are described in the following sections.

To minimize computation time, the interpolation of tracks or isochrones can be limited to only a section of the original grid given by limits in any of the grid parameters. When fitting multiple stars simultaneously, these limits can be applied on a star-by-star basis (and thus producing one interpolated subgrid per star), or as a single set of limitations for the full collection of stars (producing only one new subgrid of tracks or isochrones).

5.1 Interpolation along tracks or isochrones

For interpolation along the tracks or isochrones the user must define two relevant quantities. The first is the desired resolution in a stellar property between consecutive points in the track, normally an observed quantity that will be fitted (e.g. large frequency separation, or individual oscillation frequencies). The second relevant quantity is a (smoothly varying) grid parameter to be used as the independent variable in the interpolation. We refer to this parameter as the ‘base parameter’, where typical examples are age, central hydrogen content, or central density for stellar evolution tracks, or the initial mass for the case of isochrones. The user can choose which base parameter will be used, or BASTA will consider age or initial mass as default for stellar tracks or isochrones.

Once these two quantities are defined, the number of points along each interpolated track or isochrone is estimated as the number of points needed to satisfy the desired resolution assuming an equal spacing in the base parameter. The interpolation is performed as a 1-dimensional function using either a linear or cubic method via `scipy.interpolate.interpld` (from the `SCIPY` package; Virtanen et al. 2020) on a track-by-track basis. The new base along with the interpolated parameters are stored and replace the original track. Therefore, one should consider these as completely new evolutionary tracks, instead of simple refinements of the original.

5.2 Interpolation across tracks or isochrones

Interpolation across tracks or isochrones can be applied to grids of identical microphysics. The base for this interpolation is formed by

the parameters used to construct the original grid (see Section 3), and their spacing gives the original grid resolution. For a Cartesian interpolated grid a desired resolution can be set for each of these parameters, and the code determines the minimal amount of tracks needed to satisfy this resolution one parameter at the time. For a Sobol-sampled interpolated grid the user defines a multiplicative increase in the number of tracks of the selected section of the grid. BASTA then automatically determines a homogeneous distribution of new tracks in the base parameters that meets the required increase in the number of tracks. To retain this homogeneity, the resulting grid consists solely of interpolated tracks and the original ones are excluded. An example of these tracks assignments can be seen in Fig. 4. We note that a Sobol-sampled interpolated grid can be constructed starting from a Cartesian grid, as shown in the right-hand panel of the figure.

The enveloping tracks used to interpolate each new track are determined from a tessellation of the base parameters using `scipy.spatial.Delaunay` (based on the `QHULL` package; Barber, Dobkin & Huhdanpaa 1996). The number of tracks considered to envelop a new track corresponds to the number of parameters in the base plus one. In the example shown in Fig. 4 this requires three enveloping tracks for each new track (there are two parameters in the base, M_{ini} and $[\text{Fe}/\text{H}]_{\text{ini}}$). The tessellation divides the original grid into triangles, as it can be seen in the example.

In addition to the enveloping tracks, the user must define a single additional smoothly varying quantity that runs along the evolution of the track to perform the interpolation (e.g. age, central hydrogen content, central density). We note that, when applying the combined interpolation method, this quantity does not need to be the same as the one selected for interpolation along the track. For example, the user can choose central hydrogen content when interpolating across tracks, and the subsequent interpolation along tracks can be performed with age as the independent variable.

To avoid extrapolations, the range of the smoothly varying parameter in the new track is limited to an interval that is contained by all enveloping tracks, and its spacing is determined as the mean of the spacing in the enveloping tracks. Using this basis, each parameter in the new track is interpolated separately using `scipy.interpolate.LinearNDInterpolator` (that relies on a new tessellation of the points along each enveloping track). Individual oscillation frequencies are treated as independent parameters, and are therefore interpolated individually.

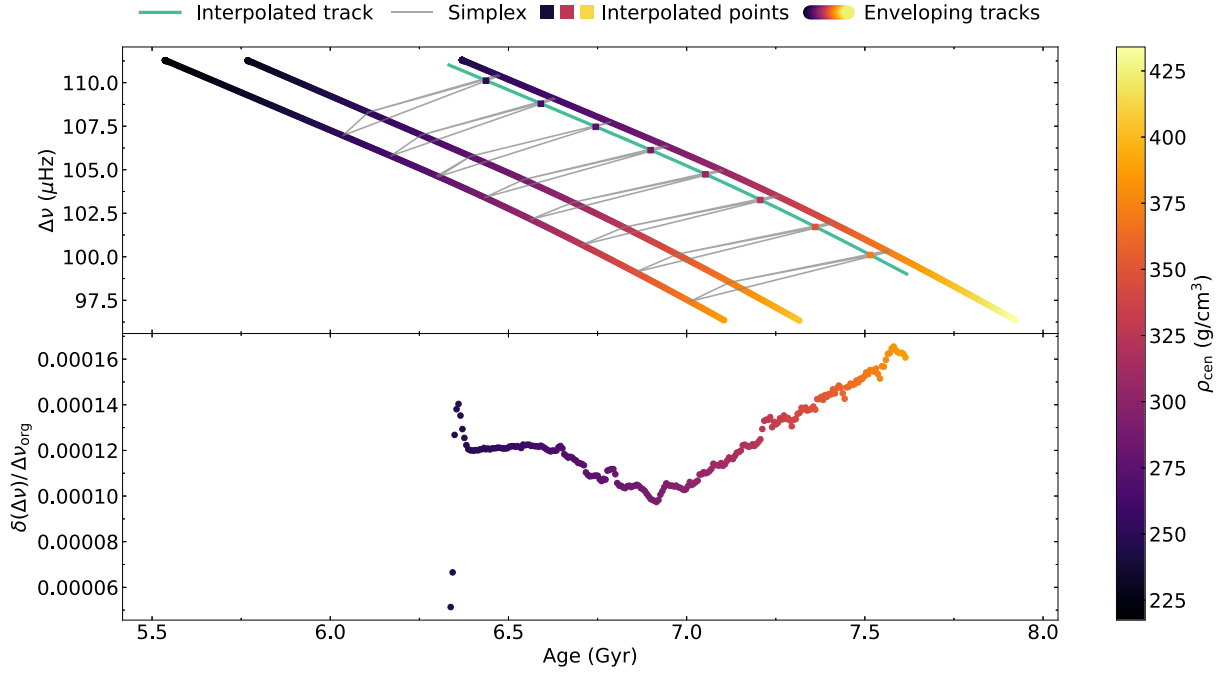


Figure 5. Example of the performance of our interpolation scheme. Top: large frequency separation as a function of age for the interpolated track (mint green) and the enveloping tracks used for the interpolation. The latter have been colour-coded according to the base quantity used along the track (the central density ρ_{cen}). Bottom: fractional difference in $\Delta\nu$ between the original and interpolated track. See the text for details.

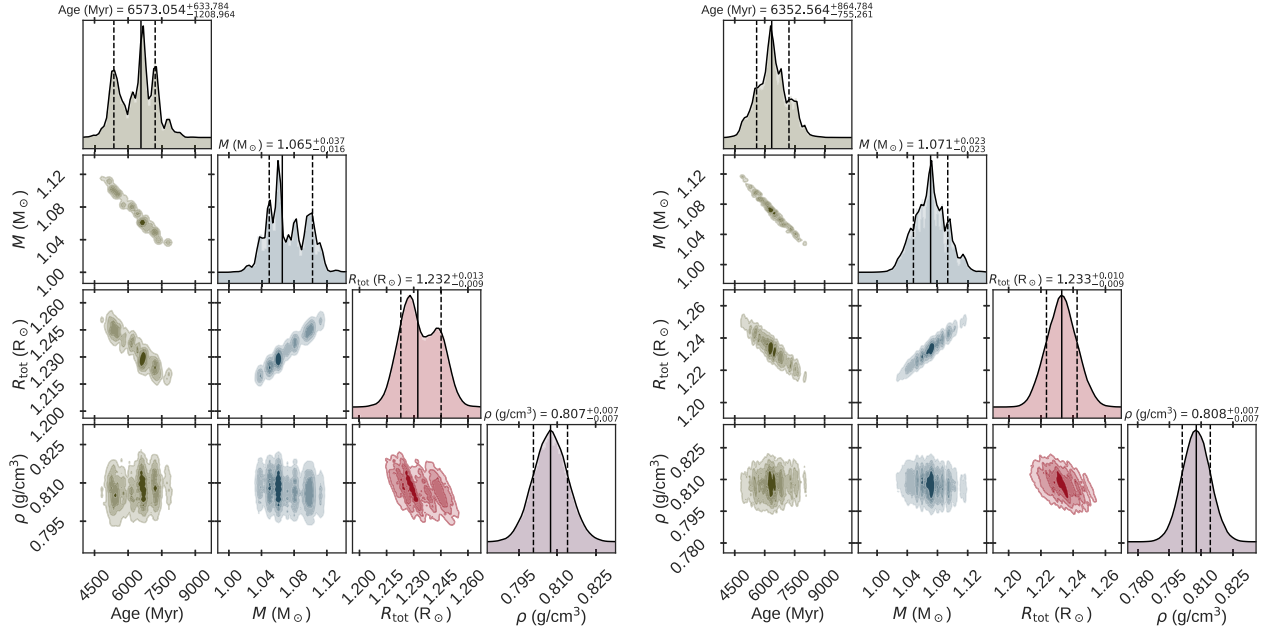


Figure 6. Posterior distributions of stellar parameters for the *Kepler* target 16 Cyg A obtained fitting the set $(T_{\text{eff}}, [\text{Fe}/\text{H}], \Delta\nu, \nu_{\text{max}})$ with a Sobol grid of different resolutions. Left: original grid. Right: interpolated grid. See the text for details.

As a test of our interpolation procedure, Fig. 5 shows the results of reconstructing an evolutionary track extracted from the Sobol grid presented in the middle panel of Fig. 4. The top panel depicts the evolution of the large frequency separation $\Delta\nu$ as a function of age for the three enveloping tracks determined from the tessellation of the base parameters, as well as the track reconstructed with our combined interpolation approach. The bottom panel presents the outcome of the combined interpolation procedure as the fractional difference in $\Delta\nu$

between the interpolated track and the original track. Deviations are at the 10^{-4} level, an order of magnitude smaller than the uncertainty in $\Delta\nu$ measured for the best *Kepler* targets (see fig. 6 in Yu et al. 2018).

The impact of our interpolation procedure can be seen in Fig. 6, where we compare the obtained posterior PDFs for mass, radius, age, and density when fitting data of the *Kepler* target 16 Cyg A. For this example we considered as input parameter the effective

temperature and metallicity from Ramírez, Meléndez & Asplund (2009) and the large frequency separation and frequency of maximum power derived by Lund et al. (2017). These data were then fitted to a Sobol grid in its original resolution, and an interpolated grid constructed with the combined method and an increase in resolution by a factor of 5 in the number of tracks (across tracks) and spacing of $0.1 \mu\text{Hz}$ between the lowest observed $\ell = 0$ frequency (along tracks). The derived quantities are in very good agreement (and certainly within their respective uncertainties), but the resulting distributions are significantly smoother after the interpolation procedure.

Before closing this section we note that when the combined method is chosen, the interpolation across tracks or isochrones is performed before interpolation along the tracks or isochrones to increase computation efficiency. We have tested the inverse case (along before across), and confirmed the differences in the recovery procedure presented in Fig. 5 are of the same magnitude regardless of the order of the interpolation. This inverse case (along before across) is still available for usage in BASTA at a much larger computational cost.

6 VALIDATION WITH ARTIFICIAL DATA

We performed a thorough end-to-end validation of BASTA to quantify the robustness of our pipeline in retrieving stellar properties. For this purpose we produced artificial data from models extracted from a grid and used the same grid to fit the observables, which is equivalent to assuming that the underlying stellar models are true representations of the observations. By quantifying the deviations in our derived parameters from the true values we can estimate the level of accuracy of our pipeline, as these will depend exclusively in the reliability of BASTA. In addition, our obtained uncertainties in stellar properties provide an estimate of the typical statistical precision, which in turn depends on the assumed observational errors and the combination of input quantities fitted. We emphasize that this exercise allows us to test the accuracy and statistical precision obtained for a given set of observables and their measured uncertainty, but we cannot account for deviations between the physical and observed properties of a real star and those predicted by a grid of stellar models. This additional systematic uncertainty undoubtedly exists and remains to be quantified, but it depends on our many shortcomings in theory of stellar evolution and goes well beyond the scope of this paper.

For this particular exercise we constructed a Sobol grid of stellar models using GARSTEC comprising 3000 evolutionary tracks, a mass range between 0.8 and $1.5 M_{\odot}$, initial metallicity $-0.5 < [\text{Fe}/\text{H}]_{\text{ini}} < +0.5$, the Asplund et al. (2009) solar mixture, and an enrichment law $\Delta Y/\Delta Z = 1.4$. The tracks were evolved from the pre-main sequence to a value of the large frequency separation $\Delta\nu = 10 \mu\text{Hz}$, roughly corresponding to the lower RGB. We determined individual oscillation frequencies of angular degree $\ell = 0, 1, 2, 3$ for stars down to a value of $\Delta\nu = 30 \mu\text{Hz}$, and only radial modes for more evolved stars. The coverage of the grid was selected to ensure that we can test our fitting procedures in various types of stars (main sequence, subgiants, and red giants) which have different observed quantities to be fitted (individual frequencies, mixed-modes, period spacing).

We assume the following observational uncertainties (taken from Serenelli et al. 2017 and Lund et al. 2017 for main-sequence and subgiant targets, and Yu et al. 2018 for red giant stars): 70 K in T_{eff} , 0.1 dex in $[\text{Fe}/\text{H}]$, 0.1 dex in $\log(g)$, 0.5 per cent of the observed value in $\Delta\nu$, and 2 per cent in ν_{max} . These are typical uncertainties in the global asteroseismic properties for *Kepler* stars observed for more than 50 d (see Serenelli et al. 2017). The asymptotic period spacing uncertainty is assumed to be 1.5 per cent, which corresponds to the

average uncertainty measured by Vrad et al. (2016) for RGB stars in the range $10 \mu\text{Hz} \leq \Delta\nu \leq 30 \mu\text{Hz}$. For the individual frequencies of oscillation we adopt a two-step process where we first identify which modes would be detected given an assumed observation length, and then determine their uncertainties following a recipe derived from *Kepler* targets.

For a given theoretical mode calculation we estimate which modes would likely be measurable based on results from the main-sequence stars included in the LEGACY study (Lund et al. 2017). From the LEGACY data we first estimate the typical minimum and maximum frequencies of measured $\ell = 0$ modes in units of ν_{max} , and derive simple linear relations of these frequencies as a function of ν_{max} . The typical frequency intervals of measurable $\ell = 0$ modes are found to range from $\pm 0.4 \nu_{\text{max}}$ at $\nu_{\text{max}} = 1000 \mu\text{Hz}$ to approximately $\pm 0.2 \nu_{\text{max}}$ at $\nu_{\text{max}} = 4000 \mu\text{Hz}$. We then estimate the expected relative amplitude of $\ell = 0$ modes at these frequency limits, using a relation for the envelope width of the assumed Gaussian modulation of mode amplitudes around ν_{max} . This relation is determined from fits to LEGACY data, and includes both a dependence on ν_{max} and T_{eff} (Lund et al., in preparation). By including mode visibilities from Lund et al. (2017), we can then estimate the corresponding relative amplitudes of non-radial modes and assess which of these exceeds the limit for detectability set by the $\ell = 0$ modes. Following Ball et al. (2018), we include information on the mode inertia for non-radial modes, where we divide amplitudes by the square-root of the Q -factor (ratio of the mode inertia relative to the $\ell = 0$ inertia at the corresponding frequency, see also equation 10).

Concerning frequency uncertainties for the modes deemed measurable from the above procedure we use a polynomial relation between the frequency uncertainties in units of ν_{max} of $\ell = 0$ modes from the LEGACY data and the corresponding relative frequency away from ν_{max} . At ν_{max} the typical minimal uncertainty is found to be of the order $\sim 7.6 \times 10^{-5} \nu_{\text{max}}$, which for a $\nu_{\text{max}} = 2000 \mu\text{Hz}$ star corresponds to $\sigma_{\nu} \sim 0.15 \mu\text{Hz}$. Away from ν_{max} the typical uncertainties increase by factors of 5 ($\nu \sim -0.4 \nu_{\text{max}}$) to 10 ($\nu \sim +0.4 \nu_{\text{max}}$).

With the observational uncertainties in all relevant observational quantities defined, the validation procedure was designed as follows. We selected 443 models from the grid to be used as artificial targets mimicking the observed distribution in large frequency separation of the *Kepler* main-sequence and subgiant sample of Lund et al. (2017; see Fig. 7). This sample reaches values of $\Delta\nu \simeq 20 \mu\text{Hz}$, which we extended in same proportion as the last bin to $\Delta\nu \simeq 10 \mu\text{Hz}$ to encompass the base of the RGB. We considered various sets of input quantities (see Table 2) and generated the synthetic data by drawing a sample from a normal Gaussian distribution with the model value of the relevant parameter as the mean and a standard deviation given by the observational uncertainty described above. In that manner, each quantity in the input set does not exactly correspond to the model value but it has been perturbed by typical observational uncertainties. This is the closest we can simulate observations of a real star, under the assumption that the underlying grid of models is a true representation of the physics at play in stellar evolution.

The distribution of the recovered fractional values in density, radius, mass, and age for the main-sequence artificial targets are shown in Fig. 8 (equivalent figures for the subgiant and RGB targets are given in Appendix B). As we can see, the inferred properties are generally in good agreement with the corresponding underlying input parameters. The inclusion of asteroseismic data dramatically improves the quality of the recovered properties compared to the cases where only atmospheric parameters are used.

As is also visible in Fig. 8, the inferred properties do not match exactly with the underlying model values. To ensure that this is a

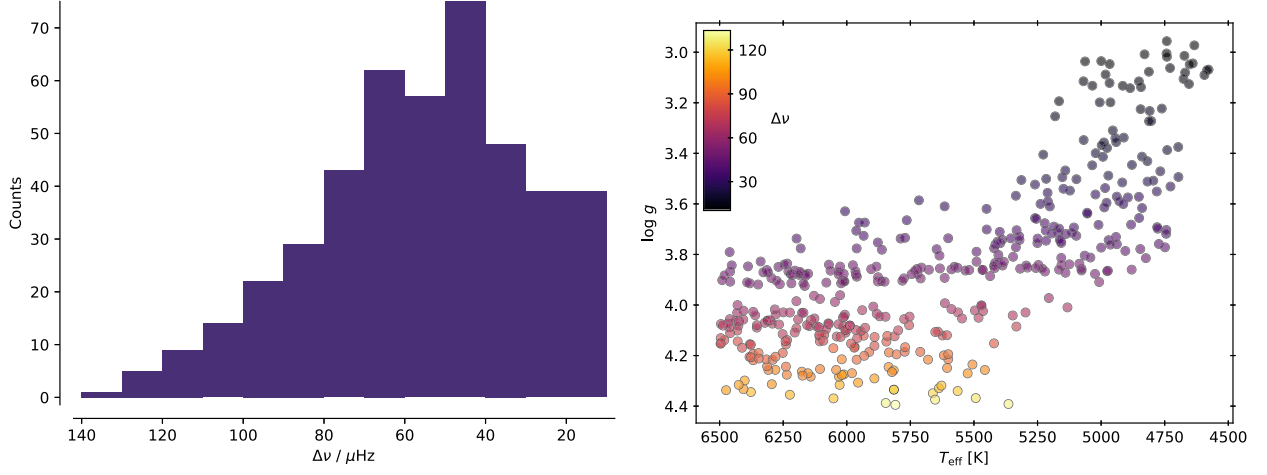


Figure 7. Left: distribution in large frequency separation of the synthetic data used for the validation procedure. Right: Kiel diagram depicting the distribution of our synthetic targets across the grid, colour-coded by the value of the large frequency separation.

Table 2. Fractional uncertainties in stellar properties across evolutionary phases determined for different sets of input. The values are calculated as the average of the 16th and 84th percentiles normalized by the derived median. See the text for details.

Phase	Input set	$\delta\rho/\rho$	$\delta R/R$	$\delta M/M$	$\delta \text{Age}/\text{Age}$
MS	T_{eff} , [Fe/H], $\log(g)$	0.3223	0.1179	0.0558	0.2880
$\Delta\nu \geq 60$ (185 stars)	T_{eff} , [Fe/H], $\Delta\nu$, ν_{max}	0.0101	0.0124	0.0364	0.1620
	T_{eff} , [Fe/H], ν_i	0.0005	0.0010	0.0029	0.0177
SG	T_{eff} , [Fe/H], $\log(g)$	0.3290	0.1356	0.0947	0.3262
$60 < \Delta\nu < 30$ (178 stars)	T_{eff} , [Fe/H], $\Delta\nu$, ν_{max}	0.0098	0.0142	0.0393	0.1175
	T_{eff} , [Fe/H], ν_i	0.0001	0.0002	0.0005	0.0009
RGB	T_{eff} , [Fe/H], $\log(g)$	0.3240	0.1283	0.1291	0.4635
$\Delta\nu \leq 30$ (80 stars)	T_{eff} , [Fe/H], $\Delta\nu$, ν_{max}	0.0101	0.0213	0.0592	0.2227
	T_{eff} , [Fe/H], ν_i	0.0009	0.0075	0.0217	0.0937
	T_{eff} , [Fe/H], $\Delta\nu$, ΔP_1	0.0097	0.0313	0.0915	0.3447
Full Sample	T_{eff} , [Fe/H], $\log(g)$	0.3253	0.1269	0.0847	0.3350
	T_{eff} , [Fe/H], $\Delta\nu$, ν_{max}	0.0099	0.0147	0.0417	0.1551
	T_{eff} , [Fe/H], ν_i	0.0004	0.0018	0.0053	0.0247

result of finite precision of the data, we reduced the assumed standard uncertainties by a factor of 10^{-4} and repeated the above validation. BASTA recovered the original model in 440 out of 443 cases, and in 3 cases we pick the model just before or after which is expected due to the perturbation to the model parameters that defines the input observables. These results confirm that the differences are due to finite data precision and are not an artefact of combined effects of priors and weights.

Since we have determined stellar properties assuming typical observed uncertainties in various sets of input, we can use the BASTA results to quantify the statistical uncertainty obtained across evolutionary phases for various sets of input. Table 2 summarizes the results, where we have listed for each stellar property the average of the 16th and 84th percentiles normalized by the median derived by BASTA. It is clear from the table how the accuracy increases as asteroseismic quantities are included in comparison to just the spectroscopic input, as well as the additional gain from including individual frequencies. This exercise also shows that, if our stellar models are a good representation of the observed stars, statistical uncertainties below the 10 per cent level in age are within reach thanks to the advent of asteroseismology.

Another interesting point arising from the validation is the extremely high precision in the recovered stellar properties for subgiant

stars when individual oscillation frequencies are fitted. As mixed-modes evolve rapidly and are very sensitive to the conditions in the stellar core, such a small statistical uncertainty is not surprising as long as the code is capable of identifying the correct underlying model. We remind the reader that in subgiant stars the number of model non-radial oscillation frequencies between two radial modes exceeds the number of observed frequencies as their visibility depends in the inertia. The procedure we described in Section 4.1.5 can successfully identify which mixed-modes should be visible and thus avoids an artificial increase in the statistical uncertainties due to missidentified modes. We include a few representative Échelle diagrams from the validation procedure in Appendix B, where the accuracy of the algorithm is demonstrated.

7 NOTES ABOUT CODE DEVELOPMENT AND CONTRIBUTING

The core developers of BASTA have customized the code following the needs of our own research fields. As the code begins to be used by scientists around the world, we expect that the inclusion of additional features will become highly desirable. Naturally the amount of proposed improvements to the code will increase proportionally to the number of users, and it will go beyond the available time of the

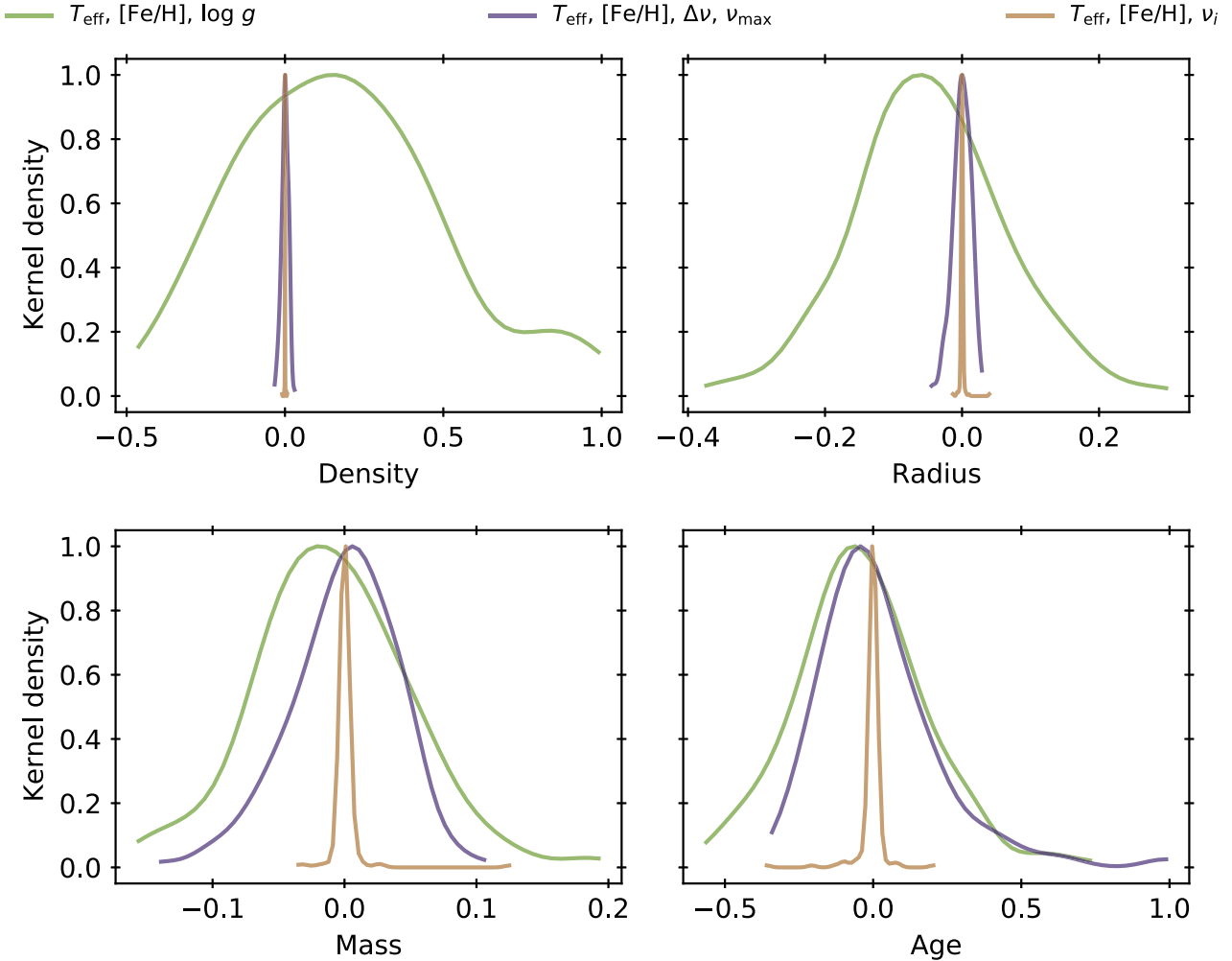


Figure 8. Distribution of normalized differences between derived stellar properties with BASTA and the underlying stellar model for the main-sequence targets. Values are computed as $(\text{Solution}-\text{Model}) / \text{Model}$. See the text for details.

core developers to implement them and maintain a stable version of BASTA.

For these reasons we rely on the `git` version-control system and the `GitHub` repository⁶ to encourage users to develop and share their contributions with the rest of the users. New features, additions, or extensions should be in separate branches or forks so that they do not need the direct involvement of the core developing team. Once the new additions are completed, the code developers will be happy to handle the pull request and make it part of the stable version of BASTA after proper testing has been completed. A more detailed contribution guide can be found on the `GitHub` repository along with the installation instructions and a tracker for issues/bugs/suggestions. We encourage the users to follow the project on `GitHub` (please ‘watch’ and select ‘custom or releases’ as a minimum) to get notifications and updates on new developments. We welcome any participation including pull requests.

8 CONCLUDING REMARKS

We have introduced the public version of the BASTA, an open-source code that determines stellar properties using a set of observables

and a grid of stellar models or isochrones. It is flexible in its input and can combine a large number of spectroscopic, photometric, astrometric, and asteroseismic input to extract properties of stars under a statistically robust Bayesian scheme. The large number of functionalities included in BASTA, combined with the various sets of publicly available or custom-computed grids of stellar models or isochrones, make it the most versatile pipeline for stellar analysis currently available.

We have thoroughly described the type of input that can be given to retrieve stellar properties, and discussed the assumptions made when predicting these quantities from the underlying grids of evolutionary models. We have performed an extensive validation test that confirms the reliability of BASTA in determining accurate stellar properties, and use these results to quantify the typical statistical uncertainties obtained for various combinations of fitting parameters. Our results show that asteroseismic ages with statistical uncertainties below the 10 per cent level are achievable for data sets of the quality obtained by the *Kepler* satellite, and the only limiting factor is the reliability of our stellar evolution and pulsation models.

We hope to have provided the community with a useful analysis tool for stellar properties, which is specifically designed to meet the challenges of large inhomogeneous data sets available in the era of large-scale stellar surveys. Its capability of combining a wide range of input data makes it an invaluable tool to fully exploit

⁶<https://github.com/BASTAcode/BASTA>

the potential of current stellar catalogues (by simultaneously fitting e.g. TESS, APOGEE, 2MASS, and Gaia eDR3 data), and its future developments will make it ready for the challenge of the next data deluge from surveys such as PLATO 2.0 (Rauer et al. 2014), WEAVE (Dalton et al. 2014), 4MOST (de Jong et al. 2019), and the Legacy Survey of Space and Time at the Vera Rubin Observatory (LSST; Ivezić et al. 2019).

ACKNOWLEDGEMENTS

Funding for the Stellar Astrophysics Centre is provided by The Danish National Research Foundation (Grant agreement No. DNRF106). VAB-K acknowledges support from VILLUM FONDEN (research grant 10118), the Independent Research Fund Denmark (Research grant 7027-00096B), and the Carlsberg foundation (grant agreement CF19-0649). SC acknowledges support from Premiale INAF MITiC (PI: B. Garilli), from Istituto Nazionale di Fisica Nucleare (INFN) (Iniziativa specifica TAsP), from Progetto Mainstream INAF (PI: S. Cassisi) and grant AYA2013-42781P from the Ministry of Economy and Competitiveness of Spain. AMS is partially supported by grant PID2019-108709GB-I00 of the MICINN of Spain. AS acknowledge support from the European Research Council Consolidator Grant funding scheme (project ASTEROCHRONOMETRY, G.A. n. 772293, <http://www.asterochronometry.eu>).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

The code is available via <https://github.com/BASTAcode/BASTA>.

The following softwares were used during this research: ASTROPY (Astropy Collaboration 2013), CORNER (Foreman-Mackey 2016), DUSTMAPS (Green 2018), EMCEE (Foreman-Mackey et al. 2013), H5PY (Collette 2013), HEALPY (Górski et al. 2005), MATPLOTLIB (Hunter 2007), NUMPY (Harris et al. 2020), SCIKIT-LEARN (Pedregosa et al. 2011), SCIPY (Virtanen et al. 2020), and TQDM (da Costa-Luis et al. 2021).

REFERENCES

Adelberger E. G. et al., 2011, *Rev. Mod. Phys.*, 83, 195
 Aerts C., Christensen-Dalsgaard J., Kurtz D. W., eds, 2010, *Asteroseismology*, Springer, Berlin
 Arentoft T. et al., 2019, *A&A*, 622, A190
 Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, 47, 481
 Astraatmadja T. L., Bailer-Jones C. A. L., 2016, *ApJ*, 832, 137
 Astropy Collaboration, 2013, *A&A*, 558, A33
 Bailer-Jones C. A. L., 2015, *PASP*, 127, 994
 Baldry I. K., Glazebrook K., 2003, *ApJ*, 593, 258
 Ball W. H., Gizon L., 2014, *A&A*, 568, A123
 Ball W. H., Chaplin W. J., Schofield M., Miglio A., Bossini D., Davies G. R., Girardi L., 2018, *ApJS*, 239, 34
 Barber C. B., Dobkin D. P., Huhdanpaa H., 1996, *ACM Trans. Math. Softw.*, 22, 469
 Beck P. G. et al., 2011, *Science*, 332, 205
 Bedding T. R. et al., 2010, *ApJ*, 713, 935
 Bedding T. R. et al., 2011, *Nature*, 471, 608
 Belkacem K., Goupil M. J., Dupret M. A., Samadi R., Baudin F., Noels A., Mosser B., 2011, *A&A*, 530, 142
 Bessell M. S., 1990, *PASP*, 102, 1181
 Bessell M. S., Brett J. M., 1988, *PASP*, 100, 1134
 Bessell M., Murphy S., 2012, *PASP*, 124, 140
 Bessell M. S., Castelli F., Plez B., 1998, *A&A*, 333, 231

Bessell M., Bloxham G., Schmidt B., Keller S., Tisserand P., Francis P., 2011, *PASP*, 123, 789
 Böhm-Vitense E., 1958, *Z. Astrophys.*, 46, 108
 Bonomo A. S. et al., 2019, *Nature Astron.*, 3, 416
 Booth R. S., Poppenhaeger K., Watson C. A., Silva Aguirre V., Wolk S. J., 2017, *MNRAS*, 471, 1012
 Brogaard K. et al., 2018, *MNRAS*, 476, 3729
 Caffau E., Ludwig H. G., Steffen M., Freytag B., Bonifacio P., 2011, *Sol. Phys.*, 268, 255
 Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245
 Casagrande L., Vandenberg D. A., 2014, *MNRAS*, 444, 392
 Casagrande L. et al., 2016, *MNRAS*, 455, 987
 Chabrier G., 2003, *PASP*, 115, 763
 Chaplin W. J., Miglio A., 2013, *ARA&A*, 51, 353
 Christensen-Dalsgaard J., 2008, *Ap&SS*, 316, 113
 Christensen-Dalsgaard J. et al., 1996, *Science*, 272, 1286
 Cohen M., Wheaton W. A., Megeath S. T., 2003, *AJ*, 126, 1090
 Collette A., 2013, Python and HDF5., O'Reilly Media, Inc., Sebastopol, CA
 Cunha M. S., Metcalfe T. S., 2007, *ApJ*, 666, 413
 da Costa-Luis C. et al., 2021, tqdm: A fast, Extensible Progress Bar for Python and CLI. Zenodo. Available at: <https://doi.org/10.5281/zenodo.5517697>
 Dalton G. et al., 2014, in Ramsay S. K., McLean I. S., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V. SPIE, Bellingham, p. 91470L
 Das P., Sanders J. L., 2019, *MNRAS*, 484, 294
 Datta A., Mazumdar A., Gupta U., Hekker S., 2015, *MNRAS*, 447, 1935
 Davies G. R. et al., 2016, *MNRAS*, 456, 2183
 de Jong R. S. et al., 2019, *The Messenger*, 175, 3
 Deheuvels S., Brandão I., Silva Aguirre V., Ballot J., Michel E., Cunha M. S., Lebreton Y., Appourchaux T., 2016, *A&A*, 589, A93
 Dewitt H. E., Graboske H. C., Cooper M. S., 1973, *ApJ*, 181, 439
 Doi M. et al., 2010, *AJ*, 139, 1628
 Dotter A., Chaboyer B., Jevremović D., Kostov V., Baron E., Ferguson J. W., 2008, *ApJS*, 178, 89
 Ferguson J. W., Alexander D. R., Allard F., Barman T., Bodnarik J. G., Hauschildt P. H., Heffner-Wong A., Tamanai A., 2005, *ApJ*, 623, 585
 Fields B. D., Olive K. A., Yeh T.-H., Young C., 2020, *J. Cosmology Astropart. Phys.*, 2020, 010
 Fletcher R., 1987, *Practical Methods of Optimization*, 2nd edn. Wiley, New York
 Foreman-Mackey D., 2016, *J. Open Source Softw.*, 1, 24
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
 Fulton B. J. et al., 2017, *AJ*, 154, 109
 Gabriel M., Noels A., Montalbán J., Miglio A., 2014, *A&A*, 569, A63
 Gaia Collaboration, 2021, *A&A*, 650, C3
 Gandolfi D. et al., 2018, *A&A*, 619, L10
 Gaulme P. et al., 2016, *ApJ*, 832, 121
 Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
 Graboske H. C., Dewitt H. E., Grossman A. S., Cooper M. S., 1973, *ApJ*, 181, 457
 Green G., 2018, *J. Open Source Softw.*, 3, 695
 Green G. M., Schlafly E., Zucker C., Speagle J. S., Finkbeiner D., 2019, *ApJ*, 887, 93
 Grevesse N., Noels A., 1993, in Prantzos N., Vangioni-Flam E., Casse M., eds, *Origin and Evolution of the Elements*. p. 15
 Grevesse N., Sauval A. J., 1998, *Space Sci. Rev.*, 85, 161
 Grundahl F. et al., 2017, *ApJ*, 836, 142
 Harris C. R. et al., 2020, *Nature*, 585, 357
 Hidalgo S. L. et al., 2018, *ApJ*, 856, 125
 Hjorth M. et al., 2019a, *MNRAS*, 484, 3522
 Hjorth M. et al., 2019b, *A&A*, 631, A76
 Hjørringgaard J. G., Silva Aguirre V., White T. R., Huber D., Pope B. J. S., Casagrande L., Justesen A. B., Christensen-Dalsgaard J., 2017, *MNRAS*, 464, 3713
 Houdek G., Gough D. O., 2007, *MNRAS*, 375, 861
 Huber D. et al., 2011, *ApJ*, 743, 143

- Huber D. et al., 2012, *ApJ*, 760, 32
 Huber D. et al., 2017, *ApJ*, 844, 102
 Huber D. et al., 2019, *AJ*, 157, 245
 Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
 Iglesias C. A., Rogers F. J., 1996, *ApJ*, 464, 943
 Ivezić Ž. et al., 2019, *ApJ*, 873, 111
 Johnson M. C. et al., 2018, *MNRAS*, 481, 596
 Jordi C. et al., 2010, *A&A*, 523, A48
 Jørgensen A. C. S., Weiss A., 2019, *MNRAS*, 488, 3463
 Jørgensen A. C. S., Mosumgaard J. R., Weiss A., Silva Aguirre V., Christensen-Dalsgaard J., 2018, *MNRAS*, 481, L35
 Kennicutt Robert C. J., Tamblyn P., Congdon C. E., 1994, *ApJ*, 435, 22
 Kjeldsen H., Bedding T. R., 1995, *A&A*, 293, 87
 Kjeldsen H., Bedding T. R., 2011, *A&A*, 529, L8
 Kjeldsen H., Bedding T. R., Christensen-Dalsgaard J., 2008, *ApJ*, 683, L175
 Korh J. et al., 2019, *MNRAS*, 482, 1807
 Kronrod A., 1965, *Nodes and Weights of Quadrature Formulas: Sixteen-place Tables*. Consultants Bureau, New York. Available at: <https://books.google.dk/books?id=CP9BAAAAIAAJ>
 Kroupa P., 2001, *MNRAS*, 322, 231
 Lund M. N. et al., 2016, *MNRAS*, 463, 2600
 Lund M. N. et al., 2017, *ApJ*, 835, 172
 Lund M. N. et al., 2019, *AJ*, 158, 248
 Lundkvist M. S. et al., 2016, *Nature Commun.*, 7, 11201
 Maíz Apellániz J., 2006, *AJ*, 131, 1184
 Maíz Apellániz J., Weiler M., 2018, *A&A*, 619, A180
 Mazumdar A., Antia H. M., 2001, *A&A*, 377, 192
 Mazumdar A., Michel E., Antia H. M., Deheuvels S., 2012, *A&A*, 540, A31
 Mazumdar A. et al., 2014, *ApJ*, 782, 18
 Miglio A. et al., 2010, *A&A*, 520, L6
 Miglio A. et al., 2012, *MNRAS*, 419, 2077
 Miller G. E., Scalo J. M., 1979, *ApJS*, 41, 513
 Mosser B. et al., 2012a, *A&A*, 537, A30
 Mosser B. et al., 2012b, *A&A*, 540, A143
 Mosser B. et al., 2014, *A&A*, 572, L5
 Mosumgaard J. R., Ball W. H., Silva Aguirre V., Weiss A., Christensen-Dalsgaard J., 2018, *MNRAS*, 478, 5650
 Mosumgaard J. R., Sølvsten Jørgensen A. C., Weiss A., Silva Aguirre V., Christensen-Dalsgaard J., 2020, *MNRAS*, 491, 1160
 Nissen P. E., Christensen-Dalsgaard J., Mosumgaard J. R., Silva Aguirre V., Spitoni E., Verma K., 2020, *A&A*, 640, A81
 Oti Floranes H., Christensen-Dalsgaard J., Thompson M. J., 2005, *MNRAS*, 356, 671
 Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, *ApJS*, 192, 3
 Paxton B. et al., 2013, *ApJS*, 208, 4
 Paxton B. et al., 2015, *ApJS*, 220, 15
 Pedregosa F. et al., 2011, *J. Mach. Learn. Res.*, 12, 2825
 Persson C. M. et al., 2018, *A&A*, 618, A33
 Pietrinferni A. et al., 2021, *ApJ*, 908, 102
 Queiroz A. B. A. et al., 2018, *MNRAS*, 476, 2556
 Ramírez I., Meléndez J., Asplund M., 2009, *A&A*, 508, L17
 Rauer H. et al., 2014, *Exp. Astron.*, 38, 249
 Reese D. R. et al., 2016, *A&A*, 592, A14
 Reimers D., 1975, *Mém. Soc. R. Sci. Liège*, 8, 369
 Rendle B. M. et al., 2019, *MNRAS*, 484, 771
 Rodrigues T. S. et al., 2017, *MNRAS*, 467, 1433
 Roxburgh I. W., 2018, preprint ([arXiv:1808.07556](https://arxiv.org/abs/1808.07556))
 Roxburgh I. W., Vorontsov S. V., 2003, *A&A*, 411, 215
 Rubele S. et al., 2012, *A&A*, 537, A106
 Sahlholdt C. L., Silva Aguirre V., 2018, *MNRAS*, 481, L125
 Sahlholdt C. L., Silva Aguirre V., Casagrande L., Mosumgaard J. R., Bojsen-Hansen M., 2018, *MNRAS*, 476, 1931
 Salaris M., Chieffi A., Straniero O., 1993, *ApJ*, 414, 580
 Salpeter E. E., 1955, *ApJ*, 121, 161
 Scalo J., 1998, in Gilmore G., Howell D., eds, *ASP Conf. Ser. Vol. 142, The Stellar Initial Mass Function (38th Herstonceux Conference)*. Astron. Soc. Pac., San Francisco, p. 201
 Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, 737, 103
 Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
 Serenelli A. et al., 2017, *ApJS*, 233, 23
 Sharma S., Stello D., Bland-Hawthorn J., Huber D., Bedding T. R., 2016, *ApJ*, 822, 15
 Silva Aguirre V., Ballot J., Serenelli A. M., Weiss A., 2011, *A&A*, 529, 63
 Silva Aguirre V. et al., 2012, *ApJ*, 757, 99
 Silva Aguirre V. et al., 2013, *ApJ*, 769, 141
 Silva Aguirre V. et al., 2015, *MNRAS*, 452, 2127
 Silva Aguirre V. et al., 2017, *ApJ*, 835, 1
 Silva Aguirre V. et al., 2018, *MNRAS*, 475, 5487
 Silva Aguirre V. et al., 2020a, *A&A*, 635, A164
 Silva Aguirre V. et al., 2020b, *ApJ*, 889, L34
 Sobol I. M., 1967, *USSR Comp. Math. Math. Phys.*, 7, 86
 Spitoni E., Silva Aguirre V., Matteucci F., Calura F., Grisoni V., 2019, *A&A*, 623, A60
 Spitoni E., Verma K., Silva Aguirre V., Calura F., 2020, *A&A*, 635, A58
 Stello D. et al., 2013, *ApJ*, 765, L41
 Stello D. et al., 2016, *ApJ*, 832, 133
 Stockholm A., Nissen P. E., Silva Aguirre V., White T. R., Lund M. N., Mosumgaard J. R., Huber D., Jessen-Hansen J., 2019, *MNRAS*, 489, 928
 Strang G., 2006, *Linear Algebra and its Applications*. Thomson, Brooks/Cole, Belmont, CA
 Talens G. J. J. et al., 2018, *A&A*, 612, A57
 Tassoul M., 1980, *ApJS*, 43, 469
 Thoul A. A., Bahcall J. N., Loeb A., 1994, *ApJ*, 421, 828
 Valle G., Dell'Omodarme M., Prada Moroni P. G., Degl'Innocenti S., 2014, *A&A*, 561, A125
 Valle G., Dell'Omodarme M., Prada Moroni P. G., Degl'Innocenti S., 2015, *A&A*, 575, A12
 Van Eylen V., Albrecht S., 2015, *ApJ*, 808, 126
 Van Eylen V. et al., 2018a, *MNRAS*, 478, 4866
 Van Eylen V., Agentoft C., Lundkvist M. S., Kjeldsen H., Owen J. E., Fulton B. J., Petigura E., Snellen I., 2018b, *MNRAS*, 479, 4786
 van Saders J. L., Ceillier T., Metcalfe T. S., Silva Aguirre V., Pinsonneault M. H., García R. A., Mathur S., Davies G. R., 2016, *Nature*, 529, 181
 Van Rossum G., Drake F. L., 2009, *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA
 Verma K., Silva Aguirre V., 2019, *MNRAS*, 489, 1850
 Verma K. et al., 2014, *ApJ*, 790, 138
 Verma K., Raodeo K., Antia H. M., Mazumdar A., Basu S., Lund M. N., Silva Aguirre V., 2017, *ApJ*, 837, 47
 Verma K., Raodeo K., Basu S., Silva Aguirre V., Mazumdar A., Mosumgaard J. R., Lund M. N., Ranadive P., 2019, *MNRAS*, 483, 4678
 Vernazza J. E., Avrett E. H., Loeser R., 1981, *ApJS*, 45, 635
 Viani L. S., Basu S., Corsaro E., Ball W. H., Chaplin W. J., 2019, *ApJ*, 879, 33
 Virtanen P. et al., 2020, *Nature Methods*, 17, 261
 Vrandar M., Mosser B., Samadi R., 2016, *A&A*, 588, A87
 Weiss A., Schlattl H., 2008, *Ap&SS*, 316, 99
 White T. R., Bedding T. R., Stello D., Christensen-Dalsgaard J., Huber D., Kjeldsen H., 2011, *ApJ*, 743, 161
 White T. R. et al., 2012, *ApJ*, 751, L36
 White T. R. et al., 2013, *MNRAS*, 433, 1262
 Wright E. L. et al., 2010, *AJ*, 140, 1868
 Yu J., Huber D., Bedding T. R., Stello D., Hon M., Murphy S. J., Khanna S., 2018, *ApJS*, 236, 42

APPENDIX A: PHOTOMETRIC SYSTEMS

Table A1 gives a list of the currently available synthetic colours for all tracks and isochrones used with BASTA. Additional sets are added as new passbands become available.

Table A1. Available photometric systems.

Photometric system	Calibration	Passbands	Zero-points
<i>UBVRIJHKLM</i>	Vegamag	Bessell & Brett (1988); Bessell (1990)	Bessell, Castelli & Plez (1998)
<i>HST-WFPC2</i>	Vegamag	SYNPHOT	SYNPHOT
<i>HST-WFC3</i>	Vegamag	SYNPHOT	SYNPHOT
<i>HST-ACS</i>	Vegamag	SYNPHOT	SYNPHOT
2MASS	Vegamag	Cohen, Wheaton & Megeath (2003)	Cohen et al. (2003)
DECam	ABmag	DES collaboration	0
<i>Gaia DR1</i>	Vegamag	Jordi et al. (2010)	Jordi et al. (2010)
<i>Gaia DR2</i>	Vegamag	Maíz Apellániz & Weiler (2018)	Maíz Apellániz & Weiler (2018)
<i>Gaia eDR3</i>	Vegamag	Gaia Collaboration (2021)	Gaia Collaboration (2021)
<i>JWST-NIRCam</i>	Vegamag	<i>JWST</i> User Documentation	SYNPHOT
SAGE	ABmag	SAGE collaboration	0
Skymapper	ABmag	Bessell et al. (2011)	0
Sloan	ABmag	Doi et al. (2010)	Dotter et al. (2008)
Strömgren	Vegamag	Maíz Apellániz (2006)	Maíz Apellániz (2006)
VISTA	Vegamag	ESO	Rubele et al. (2012)
Tycho+ <i>Hipparcos</i>	ABmag	Bessell & Murphy (2012)	Bessell & Murphy (2012)
<i>Kepler</i>	ABmag	Kepler collaboration	0
<i>TESS</i>	ABmag	TESS collaboration	0
<i>WISE</i> W1 & W2	Vegamag	WISE Collaboration	Wright et al. (2010)

APPENDIX B: VALIDATION FIGURES FOR SUBGIANT AND RGB TARGETS

We include in this section the validation figures for the subgiant (Fig. B1) and RGB targets (Fig. B2), equivalent to Fig. 8. We also show a few representative Échelle diagrams in Figs B3–B6 obtained in

the validation procedure showing the performance of the frequency matching algorithm described in Section 4.1.5 in the presence of mixed-modes.

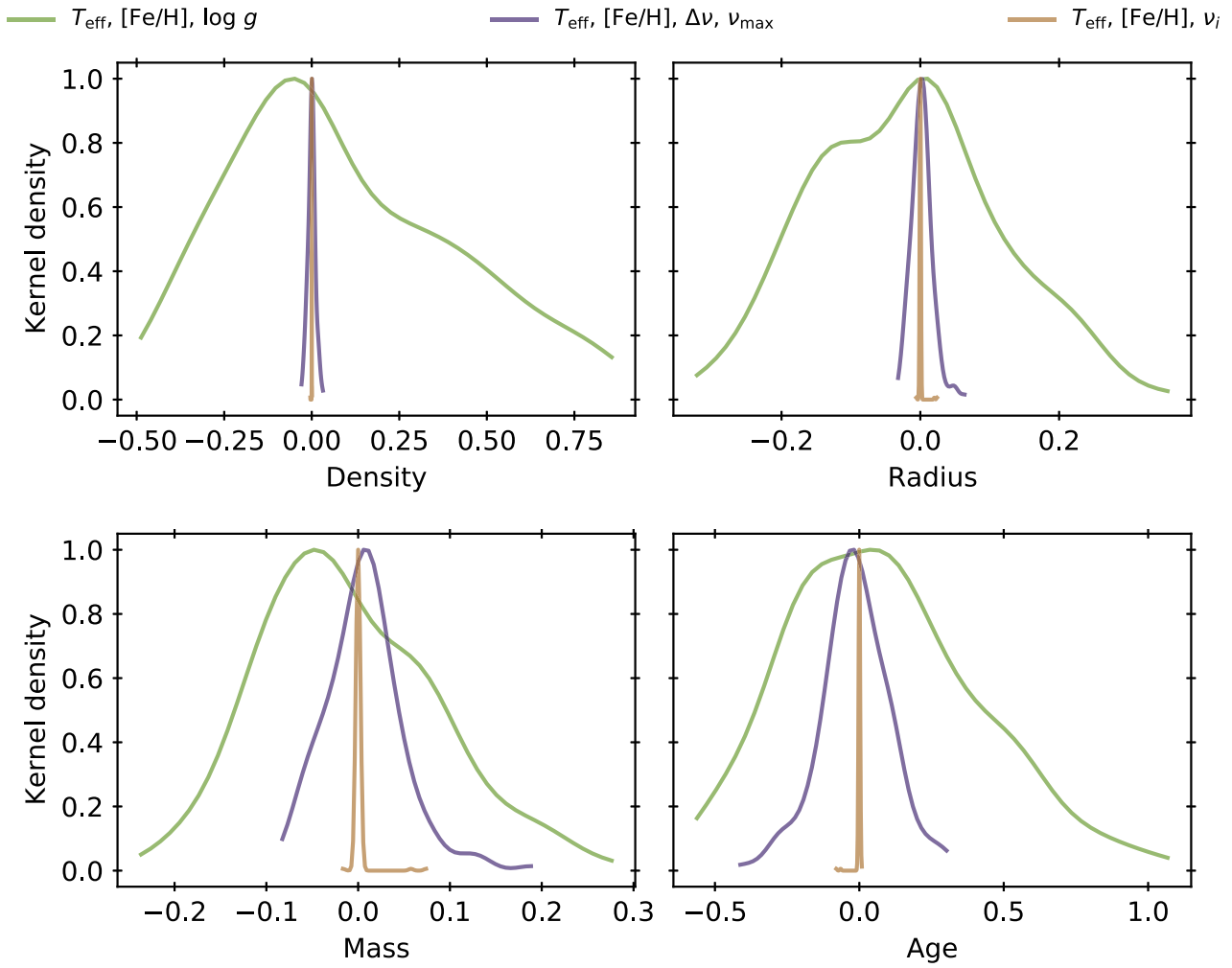


Figure B1. Same as Fig. 8 for the targets in the subgiant phase.

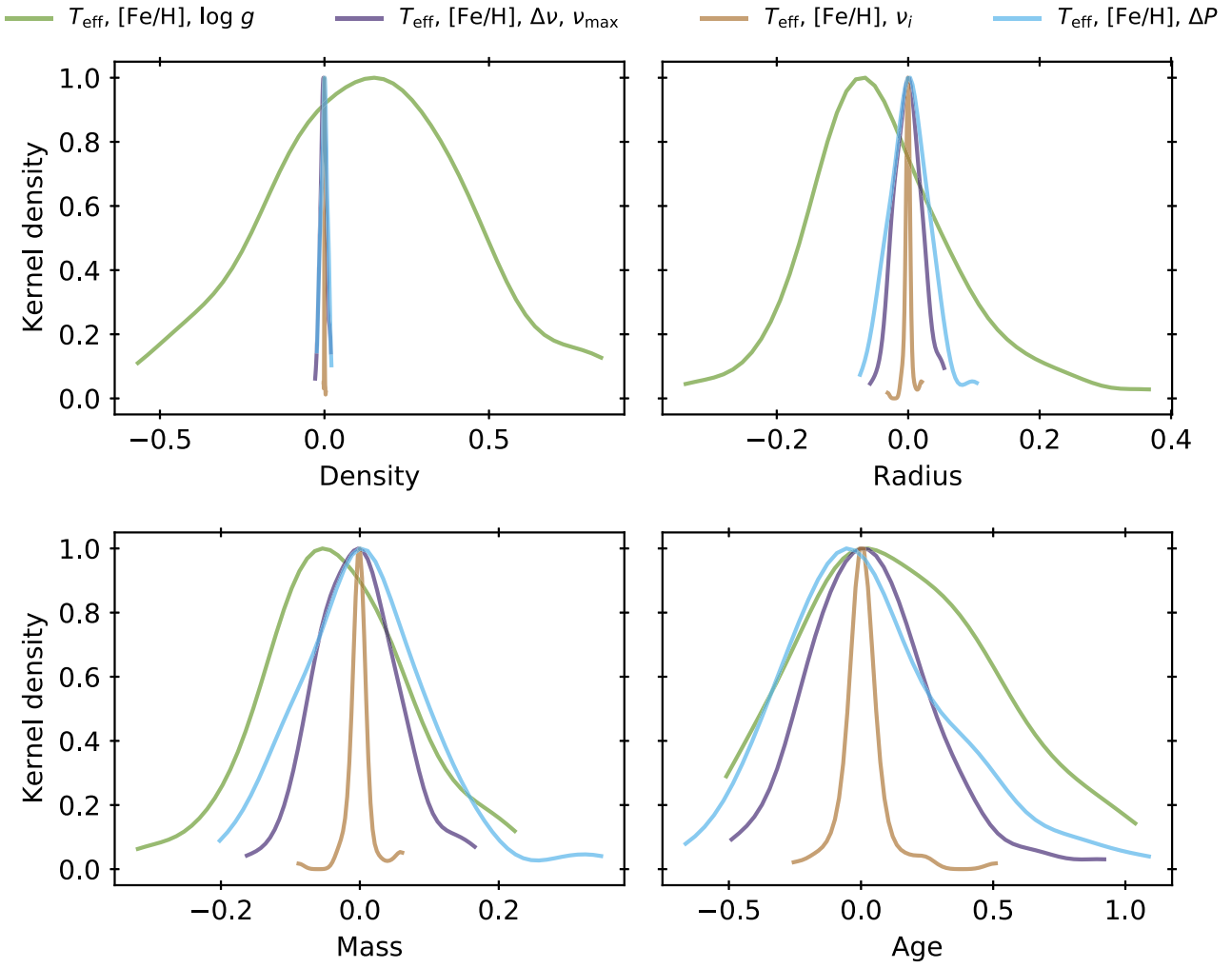


Figure B2. Same as Fig. 8 for the targets in the RGB phase.

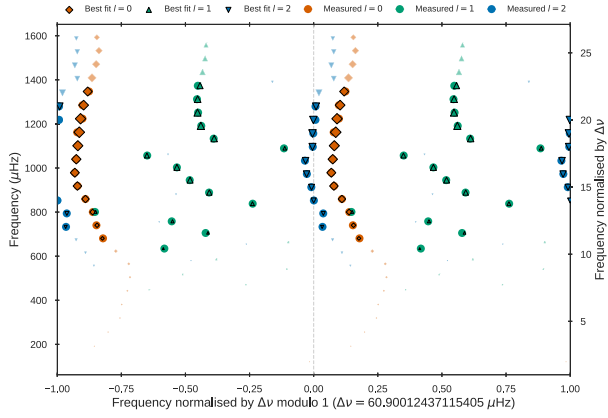


Figure B3. Échelle diagram of a subgiant validation star of $\Delta\nu \simeq 60\mu\text{Hz}$.

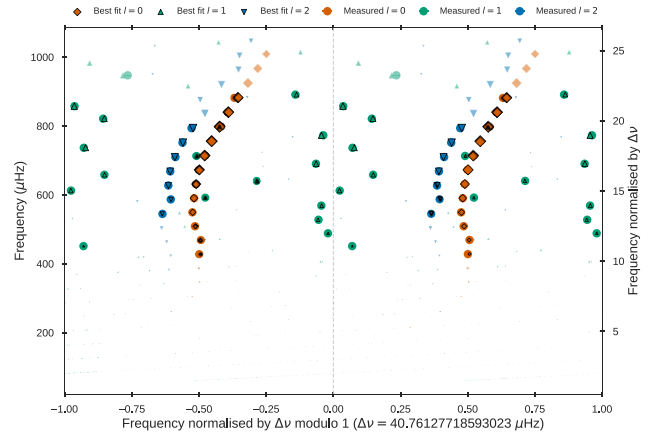


Figure B5. Échelle diagram of a subgiant validation star of $\Delta\nu \simeq 40\mu\text{Hz}$.

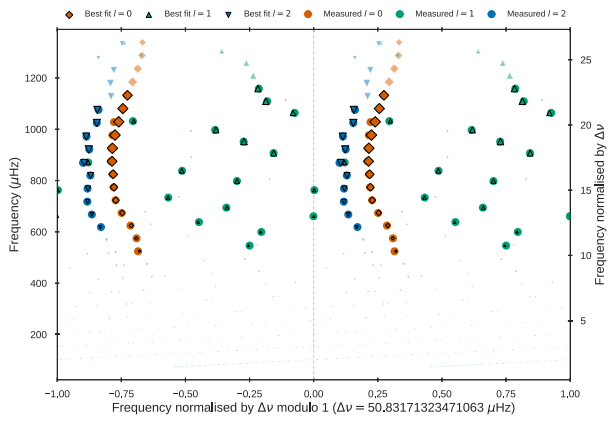


Figure B4. Échelle diagram of a subgiant validation star of $\Delta\nu \simeq 50\mu\text{Hz}$.

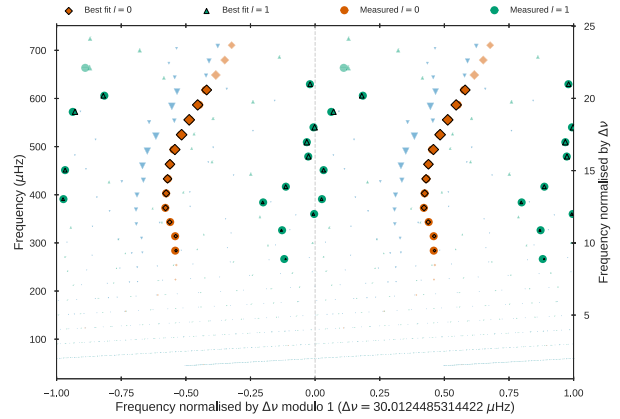


Figure B6. Échelle diagram of a subgiant validation star of $\Delta\nu \simeq 30\mu\text{Hz}$.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.