On the impact of the structural surface effect on global stellar properties and asteroseismic analyses

Andreas Christ Sølvsten Jørgensen,[1,2]★ Josefina Montalbán,[1] George C. Angelou ,[2] Andrea Miglio ,[1] Achim Weiss ,[2] Richard Scuflaire,[3] Arlette Noels,[3] Jakob Rørsted Mosumgaard [4] and Víctor Silva Aguirre[4]

[1]*School of Physics and Astronomy, University of Birmingham, Edgbaston B15 2TT, UK*
[2]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str 1, D-85748 Garching, Germany*
[3]*Space Sciences, Technologies and Astrophysics Research (STAR) Institute, Université de Liège, 19C Allée du 6 Août, B-4000 Liège, Belgium*
[4]*Stellar Astrophysics Centre (SAC), Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark*

ABSTRACT

In a series of papers, we have recently demonstrated that it is possible to construct stellar structure models that robustly mimic the stratification of multidimensional radiative magnetohydrodynamic simulations at every time-step of the computed evolution. The resulting models offer a more realistic depiction of the near-surface layers of stars with convective envelopes than parametrizations, such as mixing length theory, do. In this paper, we explore how this model improvement impacts on seismic and non-seismic properties of stellar models across the Hertzsprung–Russell diagram. We show that the improved description of the outer boundary layers alters the predicted global stellar properties at different evolutionary stages. In a hare and hound exercise, we show that this plays a key role for asteroseismic analyses, as it, for instance, often shifts the inferred stellar age estimates by more than 10 per cent. Improper boundary conditions may thus introduce systematic errors that exceed the required accuracy of the *PLATO* space mission. Moreover, we discuss different approaches for computing stellar oscillation frequencies. We demonstrate that the so-called gas Γ_1 approximation performs reasonably well for all main-sequence stars. Using a Monte Carlo approach, we show that the model frequencies of our hybrid solar models are consistent with observations within the uncertainties of the global solar parameters when using the so-called reduced Γ_1 approximation.

Key words: asteroseismology – methods: statistical – stars: atmospheres – stars: interiors.

1 INTRODUCTION

In asteroseismic analyses, stellar parameters, as well as the internal physical processes, are determined by comparing observations with theoretical stellar models. To give a holistic depiction of the entire structure and evolution of stars, current stellar models are subject to a set of simplifying assumptions. Stellar models thus assume spherical symmetry, which allows structures to be computed as a function of a single spatial coordinate. They are one-dimensional (1D). Furthermore, to capture the complicated behaviour of multi-dimensional physical processes such as turbulent convection, sim-plified parametrizations are employed. This includes mixing length theory (MLT; Böhm-Vitense 1958) and full-spectrum theory (FST; Canuto & Mazzitelli 1991, 1992; Canuto, Goldman & Mazzitelli 1996). Without these 1D parametrizations, it becomes intractable to compute the details of inherently dynamical processes over the nuclear time-scale. However, the invoked simplifying assumptions do not perfectly capture the behaviours of the relevant hydrodynamic processes. In the case of superadiabatic convection, the resulting inadequate treatment of the surface layers of stars with convective envelopes is known to lead to a systematic offset between observa-tions and the predicted model frequencies. This tension with data, i.e. the aforementioned frequency shift, is the so-called *structural* surface effect.

In addition, model frequencies are computed under the assumption of adiabaticity. The neglect of non-adiabatic energetic and the contributions of turbulent pressure leads to yet another frequency offset known as the *modal* surface effect. The combined structural and modal surface effect has haunted astero- and helioseismology for decades (Brown 1984; Christensen-Dalsgaard, Thompson & Gough 1989; Gough 1990; Aerts 2019).

It is a common practice to deal with the surface effect in the post-processing, using semi-empirical correction relations (e.g. Kjeldsen, Bedding & Christensen-Dalsgaard 2008; Ball & Gizon 2014; Sonoi et al. 2015). However, the versatility and broad applicability of these correction relations throughout the Hertzsprung–Russell (HR) diagram are yet to be fully mapped. Indeed, several studies show that the use of different surface correction relations introduces systematic errors in the inferred stellar parameters from asteroseismic analyses (Nsamba et al. 2018; Jørgensen et al. 2019, 2020). Even if this was not the case, the improper depiction of the boundary layers, from which the surface effect arises, would still introduce systematic offsets in the inferred stellar properties. This is because the surface

★ E-mail: andreas.soelvsten@gmail.com

effect, i.e. the frequency offset, is not the only consequence of an inadequate treatment of superadiabatic convection. Indeed, the improper depiction of the boundary layers has repeatedly been shown to affect the predicted stellar evolution tracks (Salaris & Cassisi 2015; Mosumgaard et al. 2017, 2018; Sonoi et al. 2019).

Multidimensional simulations of radiative magnetohydrodynamics (RHD) (cf. Freytag et al. 2012; Magic et al. 2013; Trampedach et al. 2013) yield a physically more realistic depiction of convection than stellar structure models do. However, such simulations cannot provide the same holistic depiction of stars as stellar models, due to their high computational cost. To overcome this issue, one might combine the advantages of both approaches by implementing the results from the physically more realistic multidimensional simulations into the holistic stellar models from stellar evolution codes. One way to do this is referred to as patching. In this procedure, the outermost layers of a given 1D stellar model are replaced by the average stratification of a multidimensional, often three-dimensional (3D), simulation (Rosenthal et al. 1999; Piau et al. 2014; Sonoi et al. 2015; Ball et al. 2016; Magic & Weiss 2016; Jørgensen et al. 2017; Trampedach et al. 2017; Manchon et al. 2018; Houdek et al. 2019; Jørgensen et al. 2019). Following the terminology introduced by Jørgensen et al. (2018), we will refer to such mean stratifications of the outer superadiabatic layers as ⟨3D⟩ envelopes. We note that the employed ⟨3D⟩ envelopes do by no means cover the entire convective zone. Indeed, they only reach down into the nearly adiabatic region and are therefore often referred to as '3D atmospheres' by other authors.

Due to a high degree of homology between the multidimensional simulations, it is possible to robustly recover the required ⟨3D⟩ envelopes by means of interpolation (Jørgensen et al. 2017; Jørgensen et al. 2019). Patched models can thus be constructed across the HR diagram for any combination of effective temperature (T_{eff}), surface gravity (log g), and metallicity ([Fe/H]).

Patched models do not suffer from the same structural deficiencies as standard stellar models and have repeatedly been shown to overcome the associated contributions to the surface effect (e.g. Rosenthal et al. 1999). The remaining discrepancies between the predicted model frequencies and observations are modal, i.e. the remaining surface effect does not indicate shortcomings of the stellar structure models themselves.

While patching solves some of the structural inadequacies of 1D stellar models, patching only addresses the inadequacies of the model at the last time-step. Throughout the computed stellar evolution, the interior model has thus been subject to incorrect boundary conditions through the simplified assumptions that entered the surface layers. To overcome this issue, Jørgensen et al. (2018) proposed a method for appending ⟨3D⟩ envelopes *at every time-step* and adjusting the interior model accordingly, using the ⟨3D⟩ envelopes as outer boundary conditions. Using the terminology from Jørgensen et al. (2018), we refer to the implementation of the ⟨3D⟩ envelopes into the stellar evolution code as the *coupling* of 1D and 3D models. The resulting hybrid models are thus referred to as coupled models.

In a series of papers, we have explored the properties of coupled models. We have shown that the outermost layers of coupled models perfectly mimic the underlying 3D simulation (Jørgensen et al. 2018, 2019). Furthermore, we have shown that the structures of coupled models are continuous in several physical quantities at the transition between the interior and the appended ⟨3D⟩ envelope (Jørgensen & Angelou 2019). We have demonstrated that coupled models mend the surface effect for the present-day Sun and overcome degeneracies of MLT (cf. Jørgensen & Angelou 2019). Finally, we have shown that the use of coupled models has significant consequences for stellar evolution tracks (cf. Mosumgaard et al. 2020).

In this paper, we continue our exploration of the properties of coupled models, quantifying the implications of the improved boundary conditions across the HR diagram, and demonstrating the general efficacy of our methodology.

The aim of the paper is thus threefold: First, we revisit the case of the present-day Sun (cf. Section 3.2). By employing a Monte Carlo analysis, we quantify the uncertainties that are associated with the model frequencies of coupled models. We hereby aim to contribute to the discussion on whether current hybrid models, including coupled and patched models, perform to the level of precision of the asteroseismic data.

Secondly, most authors, including ourselves, compute the model frequencies of hybrid stellar models, using the so-called gas Γ_1 approximation to avoid the complications that arise from computing model frequencies using adiabatic pulsation codes. However, there is no justification for this approach beyond the fact that it yields reasonable results for the present-day Sun. Whether this approach is generally valid across the HR diagram is hitherto unknown. We will address this issue in Section 4, showing that the so-called gas Γ_1 approximation does, indeed, perform equally well for other low-mass main-sequence stars.

Finally, having discussed the accuracy and proven the versatility of our coupling scheme, we quantify the implications of improving the outer boundary conditions across the HR diagram (cf. Sections 5 and 6). Here, we address both seismic and non-seismic stellar parameters and properties, including the stellar ages.

2 COUPLED STELLAR MODELS

Standard stellar structure models commonly use semi-empirical or theoretical relations between the temperature (T) and the optical depth (τ) to depict the atmospheric stratification above the photosphere. Such $T(\tau)$ relations set the outer boundary conditions for the interior structure (e.g. Weiss & Schlattl 2008; Kippenhahn, Weigert & Weiss 2012). They include Eddington grey atmospheres or the semi-empirical relations by Krishna Swamy (1966) and Vernazza, Avrett & Loeser (1981).

Our coupled stellar models, on the other hand, draw upon ⟨3D⟩ envelopes to set the outer boundary conditions and to depict the outermost layers. We stress that this is the case *at every time-step of the evolution*. The stratification of the ⟨3D⟩ envelopes is determined by interpolation in an existing grid of 3D simulations at every iteration. For this purpose, we use the interpolation scheme by Jørgensen et al. (2017, 2019). This method robustly recovers the accurate mean stratification of the underlying 3D simulations by interpolating in the effective temperature (T_{eff}), surface gravity (g), and metallicity ([Fe/H]). While the low number of available 3D simulations have introduced interpolation errors on the red giant branch (RGB) in previous papers, this issue has now been overcome as demonstrated in Appendix B.

In contrast to $T(\tau)$ relations, the ⟨3D⟩ envelopes stretch into the nearly adiabatic region of the convective zone, placing the outer boundary condition far below the photosphere. Throughout the paper, we set the base of the envelope at a thermal pressure that is 16 times larger than the pressure at the density inflexion at the stellar surface – the same criterion was used in previous papers (Jørgensen et al. 2018, 2019; Jørgensen & Angelou 2019; Mosumgaard et al. 2020). We thus define the point, at which we supply the outer boundary conditions, based on the pressure. This implies that the physical extent of the appended envelope varies from model to model. For the present-day Sun, the outer boundary conditions are placed more than 1000 km below the surface.

By construction, the temperature and thermal pressure stratifications of the resulting coupled models are continuous at the transition between the interior structure and the appended ⟨3D⟩ envelope. All quantities that are derived from the equation of state (EOS) and the opacity tables are, therefore, likewise continuous. Moreover, the implementation ensures that the Stefan–Boltzmann law is fulfilled. Finally, the employed input physic is chosen in such a way as to achieve a high level of consistency between the coupled models and the underlying 3D simulations. For instance, throughout this paper, we use the composition found by Asplund et al. (2009) (AGSS09). We refer to Jørgensen et al. (2018) and Jørgensen & Weiss (2019) for further details on our coupling scheme (cf. the flowchart in fig. 1 of Jørgensen & Weiss 2019).

In this paper, we compute coupled stellar models using the Garching Stellar Evolution Code (GARSTEC; Weiss & Schlattl 2008) and the CLÉS (Code Liégeois d'Évolution Stellaire; Scuflaire et al. 2008) stellar evolution code. We hereby show that the presented results are supported by independent stellar evolution codes. In all cases, we draw upon the Stagger-grid 3D RHD simulations by Magic et al. (2013). Coupled models were computed for the first time by Jørgensen et al. (2018) using GARSTEC. Indeed, results presented on coupled models in previous papers were all computed using GARSTEC, making results from this code an important reference. We have now included the same procedures into the CLÉS stellar evolution code, and we mainly perform computations using CLÉS in this paper.

We compute model frequencies for stellar pulsations, using the Aarhus adiabatic pulsation package, ADIPLS (Christensen-Dalsgaard 2008). Due to the inclusion of turbulent pressure, we compute the stellar oscillation frequencies within the so-called reduced and gas Γ_1 approximations. To highlight the tentative nature of both approaches, we deviate from the common terminology throughout most of this paper by referring to these treatments as the reduced and gas Γ_1 *assumptions*, respectively. For a thorough introduction to both treatments, we refer the reader to Appendix A or to Rosenthal et al. (1999) and Houdek et al. (2017). With the exception of Section 3.2, we deploy the gas Γ_1 assumption throughout this paper. While both the reduced and gas Γ_1 assumptions are the state of the art and widely used (e.g. Sonoi et al. 2015), we note that that the underlying assumptions on how to treat turbulent pressure in adiabatic oscillation codes have only been tested in a limited number of cases (e.g. Houdek et al. 2017). We therefore explore the validity of the gas Γ_1 assumption in Section 4. The use of a fully non-adiabatic time-dependent stellar oscillation code that would overcome the limitations of the reduced and gas Γ_1 assumptions is beyond the scope of this paper.

For all presented models, we draw upon MLT. In standard stellar models, the associated mixing length parameter (α_{MLT}) must bridge the entropy difference between the deep adiabat and the photosphere. When dealing with coupled stellar models, on the other hand, the appended ⟨3D⟩ envelopes cover most of the superadiabatic region, stretching far below the photosphere. However, we still need MLT to bridge the entropy jump between the base of the ⟨3D⟩ envelope and the deep adiabat. In coupled stellar models, MLT is thus used to describe a narrow nearly adiabatic layer. As a result, α_{MLT} plays a different role in coupled stellar models than in standard stellar models, encompassing very different information in the two scenarios. When dealing with coupled stellar models, solar calibrations with different input physics might thus yield significantly different values for α_{MLT}, and these values might by far exceed the values encountered for standard stellar models. For a more detailed discussion on this issue, we refer the reader to Jørgensen & Angelou (2019) and Mosumgaard et al. (2020).

3 THE PRESENT-DAY SUN

While solar calibrations involving coupled models are already to be found in the literature (e.g. Jørgensen & Weiss 2019), the uncertainties on the obtained stellar properties have not yet been quantified, making a direct interpretation less tangible. By performing a Markov chain Monte Carlo (MCMC) analysis, we address this issue by mapping the uncertainties on the derived stellar properties including the individual stellar oscillation frequencies. We do so within both the gas and the reduced Γ_1 assumptions. Uncertainties for standard stellar models have been quantified by, e.g. Bahcall, Serenelli & Basu (2006), Serenelli & Basu (2010), Serenelli, Peña-Garay & Haxton (2013), Vinyoles et al. (2017), and Villante & Serenelli (2020).

3.1 MCMC algorithms

Monte Carlo methods have proven to be exceedingly fruitful techniques for Bayesian inference and are employed within many fields of astrophysics (e.g. Bahcall et al. 2006; Handberg & Campante 2011; Bazot, Bourguignon & Christensen-Dalsgaard 2012; Lund et al. 2017; Vinyoles et al. 2017; Bellinger & Christensen-Dalsgaard 2019; Porqueres et al. 2019a, b). Much can be learned from these studies since they give a thorough mapping of posterior probability distributions rather than solely providing a best-fitting model.

In this paper, we use the algorithm HEPHAESTUS described by Jørgensen & Angelou (2019) to perform the study presented in Section 3.3. HEPHAESTUS is a stellar model optimization and search pipeline that employs an MCMC algorithm based on the MCMC ensemble sampler published by Foreman-Mackey et al. (2013). The underlying procedure for this ensemble sampler was originally designed by Goodman & Weare (2010).

In short, HEPHAESTUS engages several walkers that map the space spanned by the selected global parameters of stellar models. In this process, each walker constructs a Markov chain. For each entry in a Markov chain, the associated walker computes the evolution of a star up until a certain age using GARSTEC. The global parameters of each of the models, including the stellar age, are randomly drawn from proposal distributions around the parameters of the previous samples in the Markov chains of a subset of the other walkers. By comparing seismic and non-seismic properties of the final structure model from the computed evolution track to observations, HEPHAESTUS evaluates the posterior probability of the constructed model – we specify the likelihood in Section 3.2. Based on this comparison, HEPHAESTUS either rejects or accepts the investigated models as an entry in the Markov chain. Following this procedure, the density of the accumulated samples across the parameter space converges towards the posterior probability distribution of the stellar parameters of the target star – that is, after an appropriate burn-in phase. In Section 6, we perform a hare and hound exercise based on another MCMC based code called Asteroseismic Inference on a Massive scale (AIMS; Reese 2016; Lund & Reese 2018; Rendle et al. 2019). AIMS bypasses the high computational cost of MCMC by computing new samples by interpolation in an already existing grid of stellar models. Within a few hours, AIMS is thus able to investigate millions of a new combination of global stellar parameters and compare the stellar properties with observational constraints, mapping the posterior probability distribution. Like HEPHAESTUS, AIMS is based on the MCMC ensemble sampler by Goodman & Weare (2010) using the implementation by Foreman-Mackey et al. (2013).

3.2 Method: solar calibrations, likelihood, and priors

To produce a solar calibration model, GARSTEC uses a Newton solver to optimize for the structure model to fit observational constraints on the present-day Sun (Weiss & Schlattl 2008). This is an iterative procedure: GARSTEC computes several stellar evolution tracks of $1.0\,M_\odot$ stars, adjusting the mixing length parameter (α_{MLT}) and initial composition on the pre-main sequence, until the code recovers the solar luminosity ($L_\odot$), the solar radius ($R_\odot$), and the surface composition of the Sun at the present solar age. The result of this iterative calibration is a single structure model that recovers the required properties within a specified accuracy. While we thus arrive at a model of the present-day Sun, the Newton solver approach does not map the uncertainties of the global solar parameters into uncertainties on the properties of the final structure model. To do so, we perform an MCMC analysis based on the same criteria as used in standard solar calibrations. In our analysis, we thus explore a 3D parameter space, spanned by α_{MLT} as well as the initial hydrogen (X_i), and heavy metal (Z_i) abundances.

Like in a normal solar calibration, we keep the mass and the stellar age fixed to $1.0\,M_\odot$ and 4.57 Gyr, respectively. Furthermore, we evaluate our model based on $L_\odot$, $R_\odot$, and the surface composition, i.e. $Z_{S\odot}/X_{S\odot}$. By only including these three observational constraints in our likelihood, we reliably map the uncertainties that are introduced when performing a standard solar calibration. We thus vary three parameters (X_i, Z_i, and α_{MLT}) to fit three observables ($L_\odot$, $R_\odot$, and $Z_{S\odot}/X_{S\odot}$).

To facilitate an easy comparison with the literature, we use the same constraints on $L_\odot$ as Bahcall et al. (2006). As regards the solar radius, we draw upon Brown & Christensen-Dalsgaard (1998). We use AGSS09 and set the uncertainty on [Fe/H] to be 0.05 dex. This is equivalent to the uncertainties of the most abundant metals as well as on iron.

We employ broad uniform priors for all three parameters. Since a solar calibration based on coupled stellar models from GARSTEC yields a mixing length parameter of 4.9 (Jørgensen & Weiss 2019), we restrict ourselves to map the parameters space for α_{mlt} between 4.0 and 8.0. As regards the initial chemical composition, we require that the initial helium content is larger or equal to the primordial value from big bang nucleosynthesis (i.e. $Y_i \geq Y_{BBN} = 0.245$; Planck Collaboration XIII 2016). The discussed observational constraints are listed in the upper panel of Table 1.

3.3 Results and discussion

We have performed an analysis with 32 walkers, accumulating 7488 samples, after discarding a burn-in phase. The obtained posterior probability distributions on α_{MLT}, X_i, and Z_i are summarized alongside the observational constraints in Table 1.

We note that our analysis yields a broad posterior probability distribution for α_{MLT}. This is consistent with an analysis of Alpha Centauri A and B, for which Jørgensen & Angelou (2019) found that the structure and evolution of our coupled models are rather insensitive to the value taken by α_{MLT}. This is because the mixing length parameter only dictates the structure of a narrow nearly adiabatic layer, as discussed in Section 2.

We computed stellar oscillations for all 7488 realizations of the present-day Sun in our MCMC analysis. This allowed us to construct the posterior probabilities of the model frequencies. Fig. 1 shows a comparison between the resulting posterior distributions and observations from the Birmingham Solar Oscillation Network (BiSON: Broomhall et al. 2009; Davies et al. 2014) within the

Table 1. Summary of our solar MCMC analysis. The uppermost three rows contain the employed constraints. The lowermost rows contain a summary of the posterior probability distributions of the obtained stellar parameters, including the median and the percentiles of the 68 per cent credibility interval. We set $Z_{S\odot}/X_{S\odot} = 0.0180$.

R	$695,508 \pm 26^a$ km
[Fe/H]	0.00 ± 0.05^b
L	$(3.842 \pm 0.0154^c) \times 10^{33}$ erg s^{-1}
α_{MLT}	$5.77^{+0.99}_{-0.71}$
X_i	$0.7203^{+0.0076}_{-0.0069}$
Z_i	$0.014\,97^{+0.000\,96}_{-0.001\,04}$

Notes. ^{a}Seismic constraint by Brown & Christensen-Dalsgaard (1998). ^{b}Based on the composition by Asplund et al. (2009). ^{c}Constraint from Bahcall et al. (2006) and references therein.
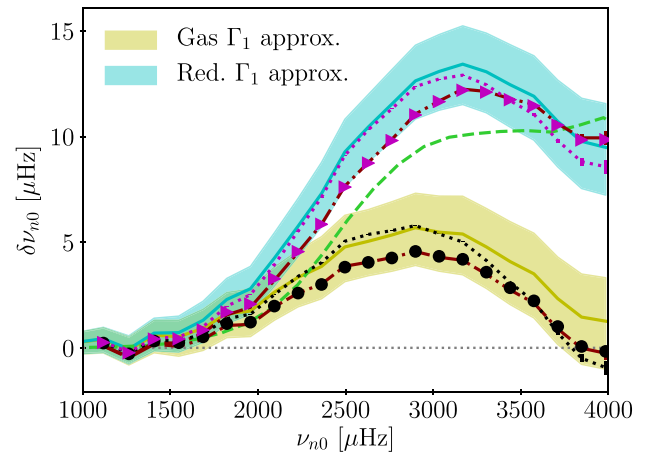

Figure 1. 68 per cent credibility intervals on the residuals $\delta\nu_{n\ell}$) between the solar model frequencies and BiSON observations within the so-called gas Γ_1 (yellow) and reduced Γ_1 (cyan) approximations. The corresponding solid line indicates the median. The plot is based on the 7488 samples from our MCMC analysis. The dashed green line shows the modal effect computed by Houdek et al. (2017) for the present-day Sun (courtesy of G. Houdek). The dots and triangles show the results obtained from a solar calibration using the CLÉS stellar evolution code (case a in Table 2). The dash–dotted dark red lines indicate the results that were obtained from the GARSTEC solar calibration model presented by Jørgensen & Weiss (2019). Finally, the dotted black and purple lines indicate the frequencies obtained when increasing the solar radius for the CLÉS model by 6 km (case b in Table 2).

gas and reduced Γ_1 assumptions. Fig. 1 also includes the modal effect determined by Houdek et al. (2017). To include non-adiabatic effects in the comparison between the adiabatic model frequencies and observations, one simply has to subtract these modal effects from the model frequencies within the reduced Γ_1 assumption (cf. Houdek et al. 2017). This is illustrated in Fig. 2. We note that we do not include any uncertainties on the modal surface effect. This is because such error bars are currently not available and because the computation of such uncertainties lies beyond the scope of this paper.

In addition to the results of the MCMC analysis, Figs 1 and 2 include the results from the solar calibration by Jørgensen & Weiss (2019) as well as two solar calibration models that have been computed using the CLÉS stellar evolution code. We summarize key numbers for these solar calibrations in Table 2.

The GARSTEC solar calibration model by Jørgensen & Weiss (2019) recovers observations within $2\,\mu$Hz at all frequencies. One of the
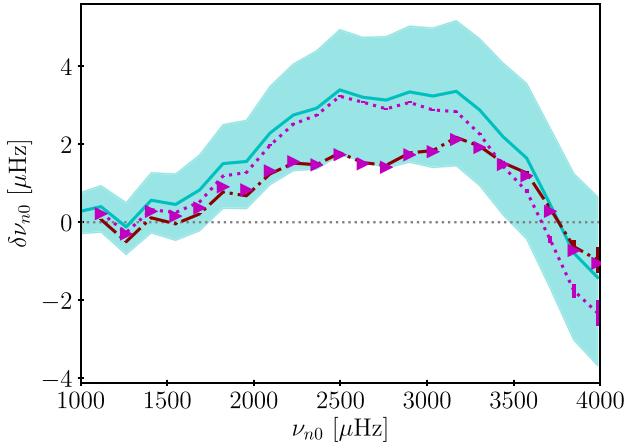
Figure 2. Pendant to Fig. 1 including only the frequencies computed within the so-called reduced Γ_1 approximation after subtracting the modal contribution to the surface effect.

Table 2. Summary of solar calibrations. All models are calibrated to recover the solar radius by Brown & Christensen-Dalsgaard (1998) (695 508 km). For the GARSTEC model, $1\,M_\odot = 1.9891 \times 10^{33}$ g, $G = 6.6738 \times 10^{-8}$ cm^3 g^{-1} s^{-2}, and $Z_{S\odot}/X_{S\odot} = 0.0179$. For the CLÉS models, $1\,M_\odot = 1.9884 \times 10^{33}$ g, $G = 6.6743 \times 10^{-8}$ cm^3 g^{-1} s^{-1}, and $Z_{S\odot}/X_{S\odot} = 0.0181$. While GARSTEC aims to fit a luminosity of 3.816×10^{33} erg s^{-1}, CLÉS aims for a luminosity of 3.828×10^{33} erg s^{-1}.

Model	R (km)	L (erg s^{-1})	α_{mlt}	Z_i	X_i
GARSTEC	695 494	3.816×10^{33}	4.876	0.0149	0.7215
CLÉS (a)	695 565	3.830×10^{33}	3.935	0.0151	0.7186
CLÉS (b)	695 571	3.830×10^{33}	3.935	0.0151	0.7186

CLÉS models (case a in Table 2) performs equally well, while the median GARSTEC model from the MCMC analysis and the other solar CLÉS model (case b in Table 2) yield slightly larger residuals. For comparison, the residuals of standard stellar models exceed the residuals shown in Fig. 2 by one order of magnitude (e.g. Model S; Christensen-Dalsgaard et al. 1996).

The remaining residuals of our coupled models are still orders of magnitude larger than the measurement uncertainties. They may hence point towards missing input physics. For instance, as discussed by Magic & Weiss (2016), the neglect of magnetic fields in the 3D simulations plays a role for the seismic properties of patched models. The same holds true for the employed solar composition, the EOS, the opacity tables, and the boundary condition for the p modes in the pulsation code. However, considering the inferred uncertainties on the model frequencies, we might at least partly account for the remaining residuals based on the uncertainties on the solar global parameters (L, R, and Z_S/X_S) alone. We also note that this finding brings the model frequencies of various patched models in the literature into line: While frequencies of published solar patched models differ by a few microhertz, this might at least partly reflect differences in the global stellar parameters. For further discussions on this topic, we refer the reader to Jørgensen et al. (2017, 2019) and Schou & Birch (2020).

Furthermore, based on the same notion, we can explain the discrepancies between the different models in Figs 1 and 2. For instance, the difference between the median of the MCMC run and the solar model presented by Jørgensen & Weiss (2019) can be explained in terms of the difference in the adopted luminosity. The solar calibration
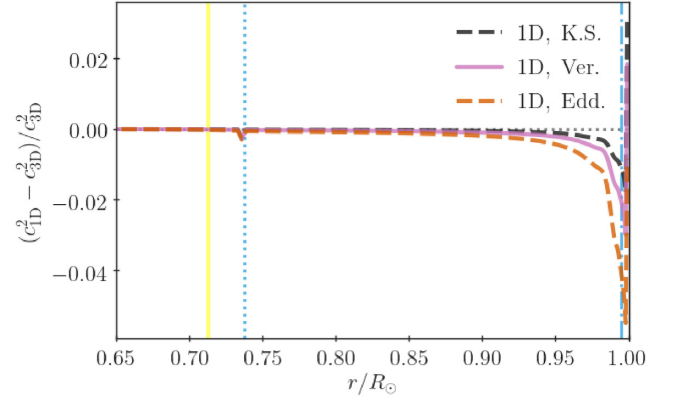
Figure 3. Relative difference in the squared sound speed between a CLÉS solar model and three standard solar models. The standard models employ an Eddington grey atmosphere (1D, Edd.) or the semi-empirical relations by Krishna Swamy (1966) (1D, K.S.) and Vernazza et al. (1981) (1D, Ver.). The deployed solar calibrations underlie the analysis in Section 5. In contrast to the solar models presented in Figs 1, 2, and 4, they do not include atomic diffusion. The dash–dotted cyan line at the right edge of the panel indicates the position of the lowermost meshpoint in the appended ⟨3D⟩ envelope. The dotted cyan line shows the position of the lower convective boundary in the coupled model. The nearby small peak in the sound speed difference indicates a rather insignificant shift in the lower convective boundary that arises from the use of simple model atmospheres. The shaded yellow area indicates observational constraints on the lower convective boundary by Basu & Antia (1997). The discrepancy between model predictions and observations for the location of the lower convective boundary is a well-known tension that arises for AGSS09. For further comparisons between the structures of coupled and standard (solar) models, we refer the reader to figs 2–4 in Jørgensen et al. (2018), figs 2, 3, and 5 in Jørgensen & Weiss (2019), figs 3 and 7 in Jørgensen & Angelou (2019), and fig. 5 in Mosumgaard et al. (2020). For a detailed depiction of the interior structure of coupled models, we refer the reader to figs 1 and 10 in Jørgensen & Angelou (2019).

model presented by Jørgensen & Weiss (2019) is thus constructed assuming the solar luminosity to be 3.816×10^{33} erg s^{-1}, in order to recover the effective temperature of the solar envelope simulation in the Stagger grid (5768.5 K), while the solar luminosity used in the MCMC analysis is 3.842×10^{33} erg s^{-1}. Similarly, the differences between the frequencies of the solar calibration model presented by Jørgensen & Weiss (2019) and the CLÉS solar calibration models can be explained in terms of the differences in the adopted luminosity and photospheric radius (cf. Table 2). The differences between the median MCMC model and the discussed solar calibrations are thus all well within the error bars that were determined by the MCMC analysis.

Finally, we turn to a discussion on the interior solar structure. The deep adiabat of the Sun, i.e. the entropy in solar adiabatic convective zone, is determined by the global solar parameters. It is therefore almost fully independent of whether we append a ⟨3D⟩ envelope or use a standard 1D atmosphere to set the outer boundary conditions (cf. Fig. 3). This is not to say that the improved boundary conditions do not affect the structure below the appended ⟨3D⟩ envelope. Indeed, as discussed by Jørgensen & Weiss (2019), the use of coupled models improves the overall sound speed profile in the upper convective layers (cf. Fig. 3). Meanwhile, the use of ⟨3D⟩ envelopes as the upper boundary conditions does, for instance, not affect the location of the base of the convective envelope significantly. Indeed, the depth of the convection zone relative to the solar radius is rather insensitive to the adiabat for a fixed EOS and fixed opacity tables, as discussed by Christensen-Dalsgaard (1997b).
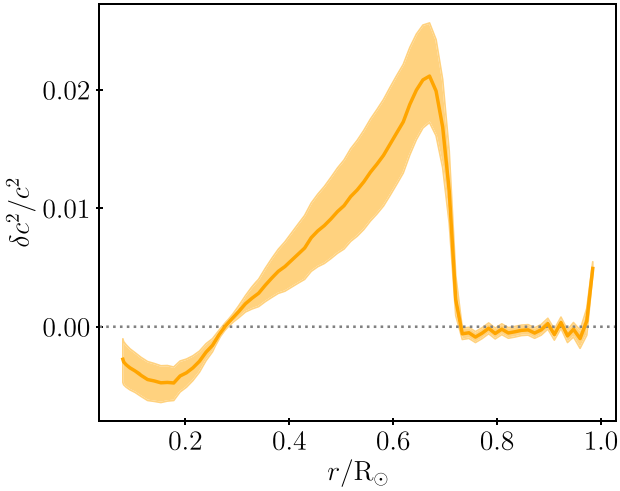
Figure 4. Squared difference between the sound speed profile of the solar model and the inferred sound speed by Basu & Antia (2008). The shaded area shows the 68 per cent credibility intervals that we infer from our MCMC analysis. The solid line indicates the median of the sound speed profile distribution. The uncertainties in the plot solely refer to the uncertainties introduced by the global stellar parameters. For a discussion on the errors that stem from the inversion method, we refer to Degl'Innoccenti et al. (1997) and Vinyoles et al. (2017).

In this paper, we employ AGSS09. As shown by Serenelli et al. (2009), this composition leads to a particularly strong disagreement with observations near the base of the convective envelope: The sound speed profiles of the stellar models are incompatible with the sound speed profile inferred from helioseismic constraints. The use of ⟨3D⟩ envelopes does not solve this shortcoming. Indeed, while the use of ⟨3D⟩ envelopes makes patched models and our coupled models superior to standard stellar models, the improved outer boundary conditions do not solve all tensions with seismic measurements. We illustrate this for the sound speed profile in Fig. 4. The tension at the lower boundary of the convection zone may, however, be addressed by including overshooting (e.g. Schlattl & Weiss 1999; Baraffe et al. 2017; Jørgensen & Weiss 2018) or altering the opacities (e.g. Christensen-Dalsgaard 1997a, b; Montalbán et al. 2004; Montalban et al. 2006; Christensen-Dalsgaard et al. 2009; Christensen-Dalsgaard, Gough & Knudstrup 2018, and references therein). Finally, earlier measurements of the solar mixture (Grevesse & Noels 1993) lead to better agreement with helioseismology. For a recent discussion of this pending issue, we refer the reader to Buldgen, Salmon & Noels (2019).

4 THE GAS Γ_1 ASSUMPTION

As can be seen from the solar models presented above, the gas Γ_1 assumption recovers observations within a few microhertz in the case of the present-day Sun. While this still corresponds to several standard deviations of the observed frequencies, it is a sizeable improvement over the uncorrected frequencies of standard stellar models. Many authors have therefore assumed that the gas Γ_1 assumption performs reasonably well across the HR diagram. However, there is no justification for this approach beyond the fact that it yields reasonable results for the present-day Sun.

As shown by Houdek et al. (2017), the reduced Γ_1 assumption appropriately accounts for the *adiabatic* contribution of the turbulent pressure to the eigenfrequencies in the case of the present-day

Sun. We can thus recover the observed frequencies by computing the adiabatic frequencies within the *reduced* Γ_1 assumption and subsequently adding the modal effect (cf. Fig. 2). It follows that the difference between the reduced and gas Γ_1 assumptions should correspond to the modal effect across the HR diagram, if the gas Γ_1 assumption indeed recovers the observed frequencies for stars other than the Sun, and if the assumptions that underlie the reduced Γ_1 assumption hold true for these stars. To establish the validity of the gas Γ_1 assumption, we therefore computed the frequency difference between the gas and reduced Γ_1 assumptions at the frequency of maximum power (ν_{max}) for different stellar parameters. We then compared this with the modal effect at ν_{max} presented in fig. 5 of Houdek et al. (2019). Note that the modal effect presented by Houdek et al. (2019) has been computed from fully non-adiabatic calculations by subtracting adiabatic frequencies that were computed within the reduced Γ_1 assumption. By comparing the difference between the reduced and gas Γ_1 assumptions to the results in Houdek et al. (2019), we are thus directly comparing the gas Γ_1 assumption to the outcome of a fully non-adiabatic time-dependent treatment. The inferred *absolute* errors of the gas Γ_1 assumption do hence not depend on the validity of the assumptions that underlie the reduced Γ_1 assumption.

For the computation of ν_{max}, we adopt (Brown et al. 1991)

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

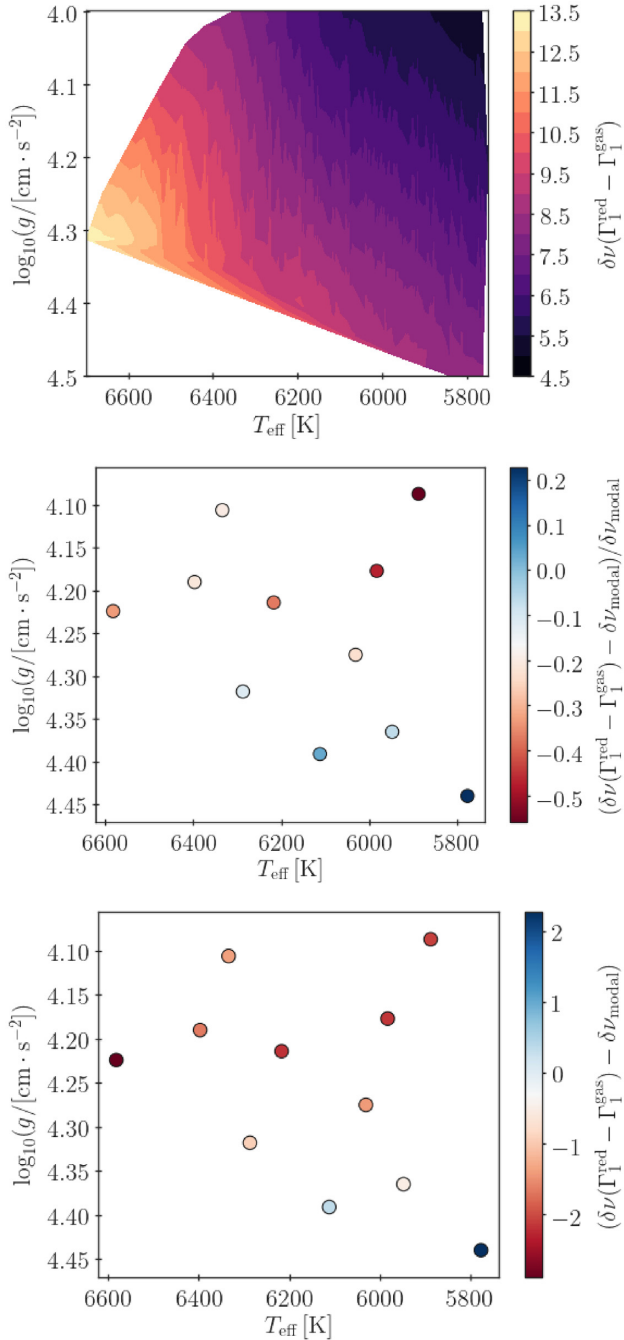where $\nu_{max\odot} = 3090\,\mu\text{Hz}$ (Huber et al. 2011; Hekker 2020) and $T_{eff\odot} = 5777$ K.

We have computed the model frequencies within the gas and reduced Γ_1 assumption for a grid of coupled models of main-sequence stars with effective temperatures between 5750 and 6700 K and with log g between 4.0 and 4.5 dex. We hereby cover the same region of the Kiel diagram as explored by Houdek et al. (2019). All models in the grid are computed without diffusion so that models that enter the analysis have solar metallicity ([Fe/H] = 0.0). The composition is based on the solar abundances evaluated by Asplund et al. (2009) and a solar calibration that was likewise performed without including diffusion ($X_i = 0.7301$, $Z_i = 0.013214$, and $\alpha_{MLT} = 1.82$).

The frequency difference between the reduced and gas Γ_1 assumptions at ν_{max} is shown in the upper panel of Fig. 5. From a qualitative comparison with the results in the paper by Houdek et al. (2019), one can see that the difference between the reduced and gas Γ_1 assumptions shows the same overall trends across the Kiel diagram as the modal effect does. Combining our results with those listed in Tables 1 and 2 in the paper by Houdek et al. (2019), we find that the difference between the reduced and gas Γ_1 assumptions recovers the modal surface effect within 50 per cent of the modal effect across the sampled region of the parameter space. The corresponding absolute error that results from the use of the gas Γ_1 assumption is thus at most 2.9 μHz across the explored region of the parameter space. These findings are illustrated in the two lower panels of Fig. 5.

While a discrepancy of up to 2.9 μHz (or 50 per cent) is sub-stantial, we note that the gas Γ_1 assumption recovers the solar observations with a similar accuracy (cf. Fig. 1). Indeed, for a large fraction of the sampled parameter space, the gas Γ_1 assumption even performs better at ν_{max} than in the case of the Sun. We thus conclude that the gas Γ_1 assumption performs as well for other low-mass main-sequence stars as it does for the Sun. While we are thus able to demonstrate the fitness of the gas Γ_1 assumption beyond the Sun, we do not directly provide a physical justification for the underlying assumptions. Rather, our results indirectly gain a physical
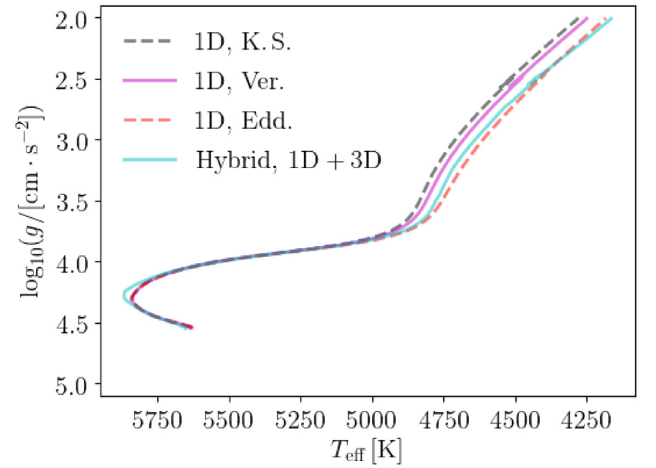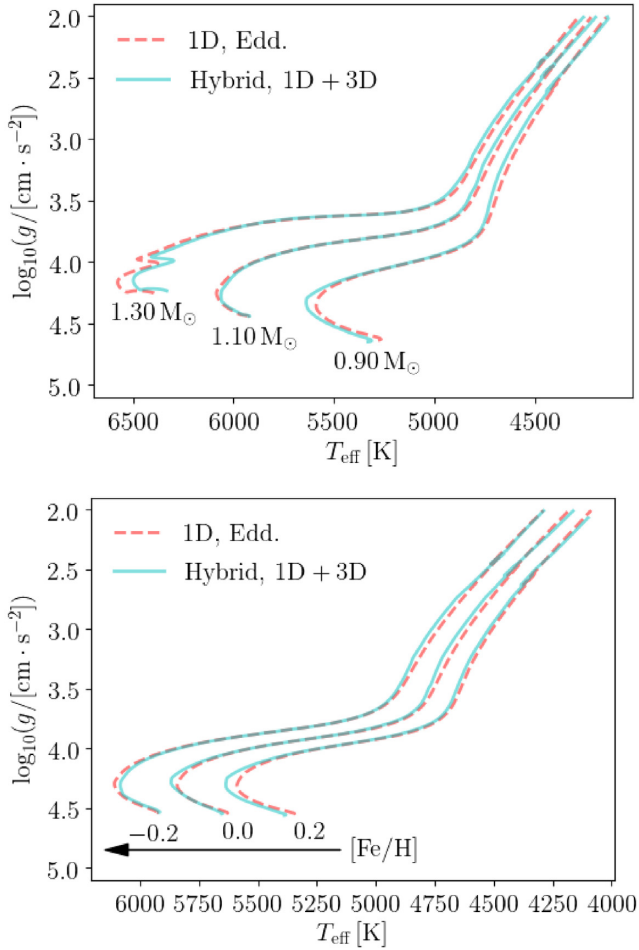
Figure 6. Kiel diagram, showing the theoretical stellar evolution track of a $1\,M_\odot$ star using different outer boundary conditions. Here, we include our coupled stellar models (Hybrid, 1D + 3D) as well as standard stellar models that employ Eddington grey atmospheres (1D, Edd.) or the semi-empirical relations by Krishna Swamy (1966) (1D, K.S.) and Vernazza et al. (1981) (1D, Ver.).

tracks. This is illustrated in Fig. 6 for a $1\,M_\odot$ star. The figure includes the evolution of a coupled stellar model as well as of three standard stellar models. The standard models are based on different $T(\tau)$ relations that are commonly found in the literature: Eddington grey atmospheres and the semi-empirical relations by Krishna Swamy (1966) and Vernazza et al. (1981). Each of the stellar evolution tracks in Fig. 6 passes through the present-day Sun by default. Each track is thus based on a distinct solar calibration, for which we employ the same outer boundary conditions. As can be seen from the figure, the use of $\langle$3D$\rangle$ envelopes affects both the predicted turn-off point (TO) and the evolution on the RGB.

In this section, we further quantify the impact of $\langle$3D$\rangle$ envelopes on the inferred global stellar properties by comparing our coupled models with standard stellar models at different masses, ages, and metallicities. Based on Fig. 6, we note that the simple Eddington grey atmosphere does a better job than its semi-empirical counterparts at recovering the evolution of the coupled models. We therefore perform the majority of the following comparisons, using standard stellar models that employ Eddington grey atmospheres. A selection of the resulting evolution tracks are shown in Fig. 7.

Other authors have included information from 3D simulations into stellar evolution codes by varying $\alpha_{\rm mlt}$ across the Kiel diagram (Trampedach et al. 2014; Magic, Weiss & Asplund 2015; see also Appendix C). It is worth noting that the resulting changes in the stellar evolution tracks are qualitatively consistent with the results presented in Fig. 6 (see Mosumgaard et al. 2020, for a more detailed discussion). Both Mosumgaard et al. (2018) and Sonoi et al. (2019) thus find that the predicted variation in $\alpha_{\rm mlt}$ leads to higher effective temperatures on the RGB than standard stellar models with constant $\alpha_{\rm mlt}$ and Eddington grey atmospheres [cf. figs 3 and 4 in Mosumgaard et al. (2018) and fig. 15 in Sonoi et al. (2019)].

Meanwhile, we note that the use of coupled models leads to a shift in the TO that is not observed when using a variable mixing length parameter (e.g. Mosumgaard et al. 2018; Sonoi et al. 2019). This might imply that the resolution of the Stagger grid is too low in the corresponding region of the HR diagram for our interpolation scheme to perform well (cf. Appendix B). If so, the position of the TO for our coupled models might be subject to interpolation

Figure 5. Upper panel: Kiel diagram, showing the frequency difference between the so-called reduced and gas Γ_1 approximations at $\nu_{\rm max}$. The plot is based on 418 stellar structure models. Middle panel: Here, we subtracted the modal effect determined by Houdek et al. (2019) from the results shown in the upper panel and plotted the relative differences. Lower panel: Pendant to middle panel but showing absolute rather than relative errors.

justification through the fully non-adiabatic calculations by Houdek et al. (2019), to which we compare.

5 GLOBAL STELLAR PROPERTIES

As discussed by Jørgensen & Weiss (2019), Jørgensen & Angelou (2019), and Mosumgaard et al. (2020), the use of $\langle$3D$\rangle$ envelopes as the outer boundary layers affects the predicted stellar evolution

Figure 7. Upper panel: Kiel diagram, showing the theoretical stellar evolution track of 0.90, 1.10, and 1.30 M$_\odot$ stars using different outer boundary conditions. Here, we include our coupled stellar models (Hybrid, 1D + 3D) as well as standard stellar models that employ Eddington grey atmospheres (1D, Edd.). Lower panel: Evolution tracks of 1.00 M$_\odot$ standard and coupled models for different metallicities.

errors. On the other hand, we note that the use of a varying mixing length parameter comes with its own caveats. First, the varying mixing length parameter is calibrated based on the existing 3D RHD simulations and is then varied across the HR diagram by interpolation in these calibrated values. The varying mixing length parameter approach is itself thus subject to the assumptions that enter through the chosen interpolation algorithm and the low resolution of the underlying grids. Secondly, it has been shown by, e.g. Trampedach & Stein (2011) that the mixing length parameter not only varies as a function of the global stellar parameters but also as a function of depth. The use of a constant mixing length parameter throughout the interior structure is thus a simplifying assumption. The procedures by Mosumgaard et al. (2018) and Sonoi et al. (2019) do not account for this and do hence not recover the stratification of the underlying 3D simulations (cf. Jørgensen et al. 2017; Mosumgaard et al. 2018). Meanwhile, as shown by Sonoi et al. (2019), a shift near the TO similar to that in Fig. 6 appears between tracks computed using MLT and FST (see also Mazzitelli, D'Antona & Caloi 1995; D'Antona et al. 2002). Since the stratifications predicted by MLT and FST are somewhat different, the finding by Sonoi et al. (2019) tells us that a shift in the TO may arise, if the variation of the mixing length

parameter with depth changes throughout the HR diagram – in this picture, from a more MLT-like to a more FST-like behaviour. In this scenario, the shift in the TO that arises from the use of coupled models (cf. Fig. 6) might be a physical feature rather than stemming from an interpolation error. To shed light on this issue, however, further 3D simulations are needed. This is beyond the scope of this paper. We thus restrict ourselves to raise caution regarding the behaviour of our coupled models near the TO. However, we also note that any interpolation errors that might occur at the TO neither affect the previous nor the subsequent evolution of the stellar models.

All models in this section were computed using the CLÉS stellar evolution code. They have all been computed without atomic diffusion, in order to ensure a constant metallicity along the stellar evolution tracks. Our coupled models include turbulent pressure in the appended ⟨3D⟩ envelope, while we ignore the contribution of the turbulent pressure to the hydrostatic equilibrium in the deep interior. In contrast, the GARSTEC models in Section 3.2 include turbulent pressure throughout the stellar structure calibrated based on the appended ⟨3D⟩ envelopes (Jørgensen & Weiss 2019). This being said, the contribution of the turbulent pressure to the total pressure is small below the ⟨3D⟩ envelope compared to its contribution within the envelope. Furthermore, as shown by Jørgensen & Weiss (2019), the stellar evolution track would be left unaffected, even if the turbulent pressure were to be ignored altogether (cf. fig. 7 in Jørgensen & Weiss 2019).

### 5.1 Comparing models at solar metallicity

In this section, we investigate stars at solar metallicity. For this purpose, we constructed a grid of coupled models with masses between 0.88 and 1.32 M$_\odot$ with a step-size of 0.01 M$_\odot$. For all models, [Fe/H] = 0. The resulting stellar evolution tracks are illustrated in Fig. B1 in Appendix B. A subsample of structure models in this grid is used in the analysis presented in Section 4. For comparison, we have constructed a grid of standard stellar models with masses between 0.80 and 1.50 M$_\odot$ with a step-size of 0.01 M$_\odot$. Again, we only include models for which [Fe/H] = 0.

For main-sequence stars, we find that the predicted ages are strongly affected by the outer boundary conditions when considering fixed masses and radii. Across the main sequence, the age differences lie close to or even exceed the 10 per cent accuracy. If one were to infer the stellar age based on tight constraints on the stellar mass and radius, coupled stellar models would hence lead to different age estimates than their standard stellar counterparts. This finding is illustrated in Fig. 8.

If we instead compare coupled and standard models with the same mass and luminosity, we again find that the largest discrepancies in age occur on the main sequence and near the TO. This is illustrated in Fig. 9. However, in this comparison, the age discrepancy is rather low for a large fraction of main-sequence stars. In accordance with this, Jørgensen & Angelou (2019) and Mosumgaard et al. (2020) both find that asteroseismic analyses based on both coupled and standard stellar models indeed yield mutually consistent age estimates for target stars on the main sequence when choosing a suitable likelihood. We thus note that the established age difference arises from changes in the properties, based on which the age is pinned down. The outer boundary conditions do not fundamentally change the evolutionary time-scales. For a star with a given mass and chemical composition, the age is hence largely independent of the boundary conditions.

Independently of the parameters that enter our comparison, the same age is obtained for standard and coupled *solar* models since both grids are based on solar calibrations. The two underlying
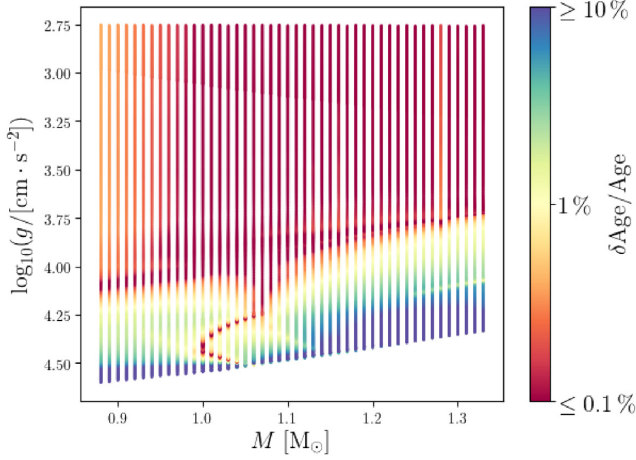
Figure 8. Relative difference in age between coupled and standard stellar models with the same mass and radius. All models were computed at solar metallicity. For clarity, we have limited the colour scale to distinguish models, for which the age difference lies at or above 10 per cent and at or below 0.1 per cent.
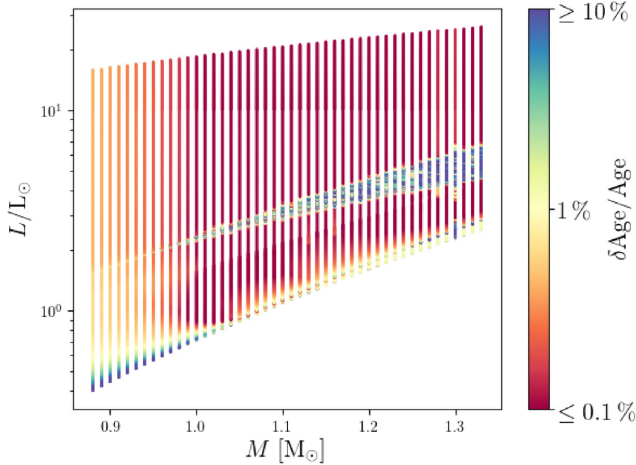

Figure 9. Pendant to Fig. 8: Relative difference in age between coupled and standard stellar models with the same mass and luminosity.

solar calibrations yield the same initial hydrogen and heavy metal abundance within $\times 10^{-4}$ and $\times 10^{-6}$, respectively. The calibrated mixing length parameter is 1.67 and 1.82 for the standard and coupled models, respectively. We use the values from the solar calibrations throughout the respective grids but note that this is a simplifying approximation (cf. Appendix C). Nevertheless, this assumption is commonly used, and adopting it thus allows for a point of comparison with the literature.

On the RGB, we find that the absolute and relative discrepancy in age is much smaller than that on the main sequence when comparing coupled and standard stellar with the same mass and radius (or luminosity). However, as can be seen from Figs 6 and 7, the effective temperature on the RGB as a function of the surface gravity is significantly altered by the use of coupled stellar models. We thus find the age estimates of coupled and standard stellar models of RGB stars to differ substantially when comparing fixed positions in the Kiel diagram.

In Fig. 10, we compare standard and coupled models with the same mass and age to quantify the resulting difference in the effective
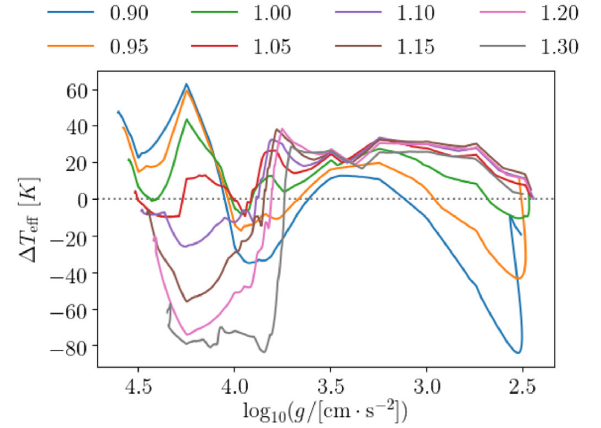

Figure 10. Difference in effective temperature between standard and coupled stellar models (see also Fig. 7). Positive residuals imply that the effective temperatures of the coupled models are higher than those of their standard model counterpart. We compare models with the same masses and ages. The mass is specified in the legend in units of the solar mass. On the abscissa, we specify the surface gravity of the coupled models, although we compare models at *the same age*. The surface gravity of the corresponding standard model at the same age thus deviates from this value. We do so in order to indicate the evolutionary phase of the star. For each star, we have included the evolution up until $\log g = 2.5$.
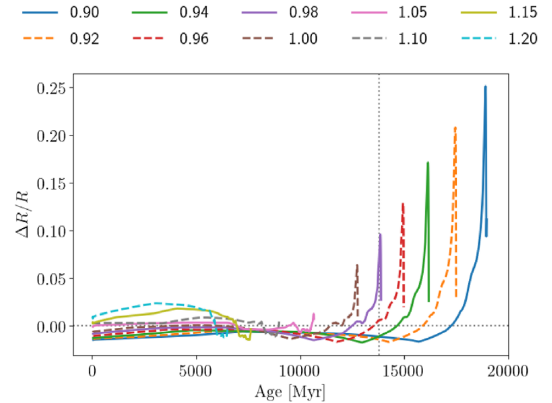
Figure 11. Difference between the radius of standard stellar models and our coupled models as a function of age relative to the radius of the coupled model. Positive residuals imply that the coupled models have larger radii than their standard stellar counterparts. We compare models with the same masses and ages. The mass is specified in the legend in units of the solar mass. For each star, we have included the evolution up until $\log g = 2.5$. The vertical dotted grey line indicates the present age of the Universe. The peak deviation is achieved close to $\log g = 2.5$, corresponding to the red-giant luminosity bump (cf. Christensen-Dalsgaard 2015). At the same $\log g$, a decrease in $\Delta T_{\rm eff}$ occurs in Fig. 10.

temperature. While coupled models of RGB stars with low masses are found to be colder than their standard stellar counterparts, coupled models of RGB stars with masses above $1\,{\rm M_\odot}$ are warmer than the standard stellar models. The opposite is true on the main sequence.

In Fig. 11, we likewise compare standard and coupled models with the same mass and age. Here, we investigate how the outer boundary conditions affect the predicted stellar radii. For models with masses below roughly $1\,{\rm M_\odot}$, we find large differences in the predicted stellar radii on the RGB. For instance, the discrepancy in the predicted radius reaches 6 per cent for a $1.0\,{\rm M_\odot}$ star (at $\log g \approx 2.5$). This finding implies that standard stellar models of red giants attribute different
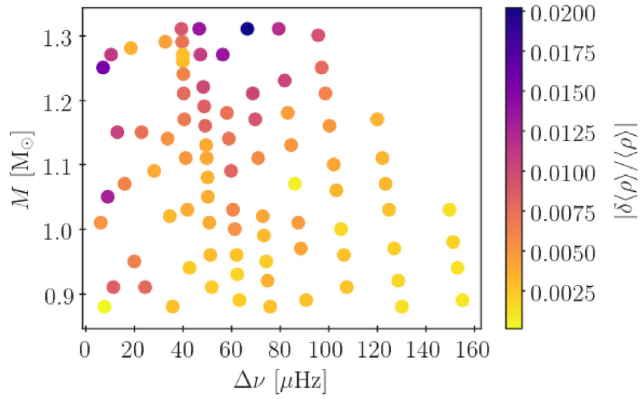
Figure 12. Relative difference in the mean density between standard and coupled stellar models as a function of the large frequency separation and the mass for a selection of models at solar metallicity. The employed models have been chosen such that they cover the Kiel diagram in a regular pattern, i.e. such that different masses and all evolutionary stages are represented. We include coupled models for which $\log g \geq 2.75$. To ensure that we are comparing models with the exact same masses and large frequency separations, we interpolate in the corresponding global stellar parameters of the computed standard stellar models.

seismic properties to the star than the corresponding coupled model with same age and mass would – especially, for low masses. We discuss this further below.

Based on Fig. 11, we note that the deviations in the stellar radius between coupled and standard stellar models are more complex than what one might anticipate based on patched stellar models. As discussed in the introduction, patched models are standard stellar models, for which the outermost layers are substituted by averaged RHD simulations *after* computing the stellar evolution (e.g. Rosenthal et al. 1999). Due to turbulent pressure and convective back warming (Trampedach et al. 2013, 2017), 3D simulations of convective envelopes are more extended than their 1D counterparts. The radius of patched models thus always exceeds that of the underlying standard stellar model. However, the improved boundary conditions do not leave the interior unaffected and alter the stellar evolution tracks. This is how a coupled stellar model can end up being smaller than a standard stellar model with the same mass and age.

To illustrate how the use of coupled models affects the global seismic properties, we computed the mean large separation ($\Delta\nu$) from the individual frequencies for a subset of coupled and standard stellar models:

$$\Delta\nu = \langle \nu_{n,\ell=0} - \nu_{n-1,\ell=0} \rangle. \tag{2}$$

We took the average over all frequencies between half and three halves of the frequency of maximum power (ν_{max}). In Fig. 12, we compare standard and coupled stellar models with the same mass and $\Delta\nu$. Because the evolution of coupled and standard stellar models differs, it stands to reason that standard and coupled stellar models with the same mass and $\Delta\nu$ will differ in some other global properties. However, even without any impact of the outer boundary conditions on the predicted evolution tracks, the models would necessarily differ in other global properties. This is because the use of $\langle 3D \rangle$ envelopes partly mends the surface effect, which shifts the individual model frequencies and thus $\Delta\nu$. The model frequencies of the standard stellar models, on the other hand, have not been corrected to take the surface effect into account. Indeed, we find the coupled models to have higher mean densities, as shown in Fig. 12.
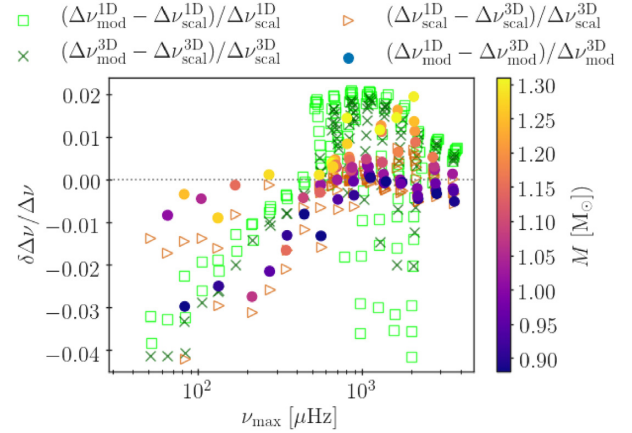
Figure 13. Relative difference in the large frequency separation as a function of the frequency of maximum power. We include large frequency separations that are computed using the scaling relation given in equation (3) as well as from the individual frequencies of the models. These are denoted by the subscripts 'scal' and 'mod', respectively. At low (high) ν_{max}, the highest residuals stem from the models with lowest (highest) masses, i.e. those models for which the highest residuals in the stellar radius are obtained (cf. Fig. 11).

Moreover, we find that the use coupled models leads to a higher value of $\Delta\nu$ when comparing coupled and standard stellar models with the same value of ν_{max}. Again, the explanation for this finding is twofold. First, the use of $\langle 3D \rangle$ envelopes partly mends the surface effect, shifting $\Delta\nu$. Secondly, while the coupled and standard stellar models that enter the comparison share the same ν_{max}, they do not share many other global properties. After all, ν_{max} is computed based on equation (1) and is thus sensitive to any shifts in mass, radius, and effective temperature between coupled and standard stellar models. If it is indeed the case that we are comparing models with different masses and radii, it follows that the standard and coupled stellar models in our comparison would also not lead to the same $\Delta\nu$ if we were to compute $\Delta\nu$ from a simple scaling relation (e.g. Handberg et al. 2017; Rodrigues et al. 2017; Sahlholdt & Silva Aguirre 2018, for a discussion on scaling relations):

$$\Delta\nu_{scal} = \Delta\nu_{\odot} \left(\frac{M}{M_{\odot}} \right)^{1/2} \left(\frac{R}{R_{\odot}} \right)^{3/2}, \tag{3}$$

where we set $\Delta\nu_{\odot} = 135.1\,\mu\text{Hz}$ (Huber et al. 2011). Indeed, when computing $\Delta\nu$ based on equation (3), we arrive at discrepancies in $\Delta\nu$ between the standard and coupled stellar models that are as large as the deviations in $\Delta\nu$ obtained from the individual frequencies. In both cases, the deviations between the standard and coupled models are thus of the order of 10^{-3}–10^{-2} times $\Delta\nu$. We show this in Fig. 13.

5.2 Comparing models across metallicities

To evaluate the impact of metallicity on the conclusions drawn above, we have computed a set of coupled and standard stellar models with [Fe/H] between -0.5 and 0.4 in steps of 0.1. In all cases, we have fixed the stellar mass to $1.0\,M_{\odot}$ and do not include diffusion. In Fig. 14, we compare the effective temperature of standard and coupled models at different evolutionary stages. For this purpose, we compare structure models with the same surface gravity. At all metallicities, we find that our coupled stellar models yield higher effective temperatures on the RGB than the standard stellar models do ($\log g \leq 3.75$). The same conclusion is drawn for all masses at solar metallicity from Fig. 8 in Section 5.1.
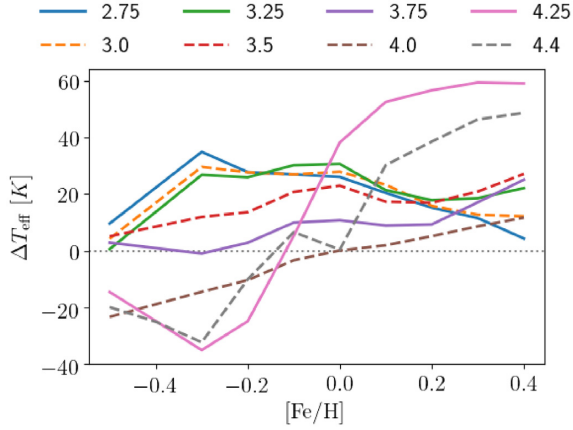
Figure 14. Difference in effective temperature at fixed metallicities and surface gravities for stars with $1\,M_\odot$. Positive residuals imply that the effective temperature of the coupled model is higher than that of its standard model counterpart. The surface gravity ($\log g$) is specified in the legend. To avoid complications arising from the RGB bump, we only include models for which $\log g >= 2.75$.
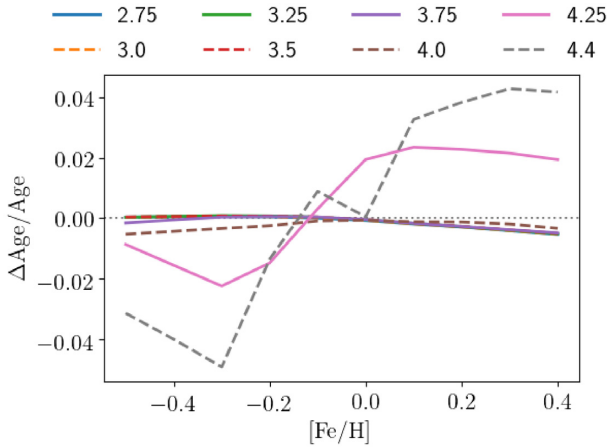
Figure 15. Relative difference in age at fixed metallicities and surface gravities for stars with $1\,M_\odot$. Positive residuals imply that our coupled model is older than its standard model counterpart. The surface gravity ($\log g$) is specified in the legend. To avoid complications arising from the RGB bump, we only include models for which $\log g >= 2.75$.

For the main sequence, a more nuanced picture emerges. At supersolar metallicities, the coupled stellar models yield higher effective temperatures on the main sequence and close to the TO than the standard stellar models do. For subsolar metallicities, we find the opposite behaviour.

Fig. 15 shows the difference in age between standard and coupled stellar models of $1\,M_\odot$ stars with different surface gravities as a function of metallicity. To ensure that we are comparing models with the exact same surface gravities, we interpolate in the global stellar parameters of the computed standard stellar models. We find that the largest absolute and relative age differences are obtained on the main sequence and close to the TO, which implies that the use of standard stellar models affects isochrones and thus age estimates for clusters. The largest error on the main sequence is thus of the order of 4 per cent.

As regards Figs 14 and 15, we note that there is no difference in effective temperature or age on the main sequence ($\log g = 4.4$) at

solar metallicity by construction. At this point of its evolution, the corresponding $1\,M_\odot$ star lies close to the present-day Sun, based on which the initial conditions of both grids were determined.

6 HARE AND HOUND EXERCISE

In this section, we perform an artificial asteroseismic analysis, in which we examine how well we can infer the global stellar properties of coupled models based on a grid of standard stellar models. For this purpose, we employ AIMS (cf. Section 3.1). The aim of our hare and hound exercise is to evaluate the magnitude of the systematic biases that are introduced on the inferred parameters when using standard rather than coupled stellar models. While coupled models give a more physically realistic depiction of stars, it is yet to be demonstrated that coupled models also yield more accurate parameter estimates. We do not aim to settle this issue here. However, under the assumption that the properties of coupled models more closely represent those of real stars, our analysis can give us an idea of how well standard stellar models perform in actual asteroseismic analyses.

As we have repeatedly addressed asteroseismic analyses of main-sequence stars in previous papers (Jørgensen et al. 2019; Mosum-gaard et al. 2020), we turn our attention to TO and RGB stars. This choice is important, since it affects how to construct a grid of stellar models for AIMS to interpolate in (cf. Rendle et al. 2019, for a detailed discussion of this issue).

We use models with the same input physics as used in Sections 4 and 5. As regards the coupled stellar models, we consider a subsample from the grid, consisting of 68 models at solar metallicity with $\log g < 4.2$. As regards the grid of standard stellar models, we include models with a metallicity between -0.5 and 0.5 dex in steps of 0.1 dex. The mass of the standard models ranges between 0.7 and $2.5\,M_\odot$ in steps of $0.02\,M_\odot$ – this is thus a different grid than the one used throughout Section 5.

6.1 Likelihood

We strive to recover a set of non-seismic properties (L, [Fe/H], and $T_{\rm eff}$) as well as the individual model frequencies of coupled stellar models using standard stellar models. For each of the considered coupled stellar models, we computed adiabatic oscillation frequencies within the gas Γ_1 assumption using ADIPLS. To account for the surface effect when using standard stellar models, we included the surface correction relation by Sonoi et al. (2015):

$$\frac{\delta\nu}{\nu_{\rm max}} = \alpha\left(1 - \frac{1}{1-(\nu_{\rm obs}/\nu_{\rm max})^\beta}\right).\tag{4}$$

Here, we let α be a free parameter but require it to be negative. We do so in order to recover the notion that the adiabatic frequencies are assumed to overestimate the observed frequencies across the HR diagram analogously to the case of the present-day Sun (Houdek et al. 2017; Houdek et al. 2019).

In equation (4), we adopt

$$\log\beta = -3.86\log T_{\rm eff} + 0.235\log g + 14.2\tag{5}$$

from Sonoi et al. (2015). We note that equations (4) and (5) have been calibrated based on the *radial* modes ($\ell = 0$) of patched models. We thus limit ourselves to only include radial modes in the likelihood – the results discussed in Section 6.1.1 constitute an exception to allow for a more direct comparison with Rendle et al. (2019).

The reason for choosing the surface correction relation by Sonoi et al. (2015) is that it has been derived within the gas Γ_1 assumption based on patched models rather than based on observations. For the

present-day Sun, this surface correction relation therefore recovers frequencies that closely resemble those obtained within the gas Γ_1 assumption in Section 3.3 (cf. Fig. 1). So far as that we believe that the surface correction relation by Sonoi et al. (2015) *generally* yields a good parametrization of the model frequencies of patched models, our coupled and standard stellar models therefore treat the surface effect consistently. Based on an analysis of hundreds of patched models, for which the frequencies have been computed within the gas Γ_1 assumption, Jørgensen et al. (2019) indeed demonstrate that a Lorentzian parametrization of the associated structural surface effect also performs well for giants and subgiants (see also Manchon et al. 2018).

This being said, the surface correction relation by Sonoi et al. (2015) is based on only 10 patched models. Moreover, these 10 models predominantly correspond to main-sequence stars and subgiants with $T_{\rm eff} \geq 6000$; log g is only lower than 3.5 in 2 out of the 10 samples. As a result, this surface correction relation is subject to a selection bias, which might affect the inferred surface effect (Jørgensen et al. 2019, 2020). However, the use of any other surface correction relation than that by Sonoi et al. (2015) would be problematic, since they do not recover the systematic frequency offset that haunts the gas Γ_1 assumption.

To include the theoretical frequencies and the remaining artificial observables into the likelihood, we have ascribed artificial statistical errors to the properties of the coupled models. For all model properties, we assume the noise to be Gaussian. This assumption is commonly used in the literature (e.g. Silva Aguirre et al. 2015; Nsamba et al. 2018). The likelihood ($\mathcal{L}$) thus takes the form

$$\mathcal{L} = (2\pi)^{-N/2} |\mathbf{C}|^{-N/2} \times$$
$$\exp\left(-\frac{1}{2} \sum_{i=1}^{N} \left[x_{1{\rm D},i} - x_{3{\rm D},i}\right]^T \mathbf{C}^{-1} \left[x_{1{\rm D},i} - x_{3{\rm D},i}\right]\right). \quad (6)$$

Here, the sum runs over all N properties ($x_{3{\rm D},i}$) that we aim to recover from the coupled models, $x_{1{\rm D},i}$ denotes the corresponding model predictions from the standard stellar models, and $\mathbf{C}$ denotes the covariances. In this paper, we assume that the observed quantities, including the non-seismic constraints as well as the individual frequencies, are uncorrelated. Consequently, the expression inside the exponential in equation (6) reduces to the expression for $-\chi^2/2$.

Based on Lund et al. (2017), we assume that a standard deviation of 0.1 μHz can be achieved at $\nu_{\rm max}$ for $\nu_{\rm max} = 2500$ μHz and that the same relative uncertainty can be achieved at $\nu_{\rm max}$ in general. For the remaining frequencies, we assume that the error increases quadratically with the frequency difference to $\nu_{\rm max}$, in order to mimic the decreasing amplitude of the modes:

$$\sigma(\nu_{n,\ell=0}) = \frac{\nu_{\rm max}}{2500} \left(0.1 + 3.6 \frac{(\nu_{n,\ell=0} - \nu_{\rm max})^2}{\nu_{\rm max}^2}\right). \quad (7)$$

Based on the assumption of a Gaussian envelope for the frequency amplitudes (e.g. Mosser et al. 2012; Rodrigues et al. 2017), we only consider frequencies that deviate less than thrice the standard deviation ($\sigma_{\rm env}$) of the Gaussian distribution from $\nu_{\rm max}$ (Mosser et al. 2012):

$$\sigma_{\rm env} = \frac{0.66 \nu_{\rm max}^{0.88}}{2\sqrt{2 \ln 2}}. \quad (8)$$

As regards the remaining constraints, we set the uncertainty on $T_{\rm eff}$ and [Fe/H] to be 100 K and 0.1 dex, respectively. The uncertainty on the luminosity is set to be 3 per cent.

6.1.1 The Sun

The chosen likelihood closely matches that used in analyses of real targets (e.g. Jørgensen et al. 2020). However, to further validate that the chosen likelihood leads to meaningful results, we ran the hare and hound exercise for the present-day Sun. For this purpose, we used the model frequencies of one of the coupled solar CLÉS model in Section 3.2 (case a in Table 2). However, since the Sun is not in our grid, we used the grid by Rendle et al. (2019). We note that this grid is based on the solar mixture found by Grevesse & Sauval (1998) and that it is thereby not fully consistent with the assumptions that enter our solar calibration model.

Rendle et al. (2019) show that they are able to recover the global properties of the Sun based on observed solar frequencies. This was accomplished using radial and non-radial modes ($\ell = 0$, 1, and 2) in combination with non-seismic constraints. When following this approach, we find that we obtain the global solar parameters with equivalent accuracy based on the model frequencies of coupled stellar models. $M = (1.001 \pm 0.003)\,{\rm M}_\odot$ and age $= (4534 \pm 91)$ Myr. Meanwhile, a lower accuracy is achieved when treating the Sun as a star on the lines described in Section 6.1 using only radial modes ($\ell = 0$). Here, we find that $M = (0.971 \pm 0.003)\,{\rm M}_\odot$ and age $= (4808 \pm 128)$ Myr. The discrepancies in mass and age thus correspond to 3 and 5 per cent, respectively. The impaired accuracy of the fit is a natural consequence of the fact that we include less informative constraints in the likelihood.

We note that we can use the grid by Rendle et al. (2019) to fit the special case of the Sun because the grid is based on a solar calibration and because the coupled solar models in Section 6.1 demonstrably recover the true solar structure with high accuracy. On the other hand, the input physics that underlies our grid of coupled stellar models differs significantly from the assumptions that enter the grid by Rendle et al. (2019). This is especially true for the composition profiles – that is, whether or not, say, atomic diffusion is included. Performing an hare and hound exercise based on other coupled models of main-sequence stars would thus lead to results that would be very hard to interpret. The analysis of the RGB stars presented below, on the other hand, does not suffer from this obstacle, since we use a grid of 1D standard stellar models that is fully consistent with the coupled stellar models, whose properties we seek to recover. The only difference lies in the treatment of the superadiabatic surface layers.

6.2 Goodness of fit

In the following, we quote the model parameters of the maximum posteriori models, to which we refer as the best-fitting models. We furthermore quote the 1σ uncertainties based on the associated 68 per cent credibility intervals derived from the posterior probability distributions. By doing so, we assume that the posterior probability distribution is well approximated by a Gaussian.

To access the goodness of fit for the best-fitting standard model, we quote the reduced χ^2-value:

$$\chi^2_{\rm red} = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_{1{\rm D},i} - x_{3{\rm D},i})^2}{\sigma^2_{3{\rm D},i}}. \quad (9)$$

Best-fitting standard models, for which $\chi^2_{\rm red} \approx 1$, reliably recover the required properties of the coupled stellar models. While values of $\chi^2_{\rm red} \ll 1$ point towards overfitting, values of $\chi^2_{\rm red} \gg 1$ reveal a poor fit. We thus discard all models, for which $\chi^2_{\rm red} > 4$.
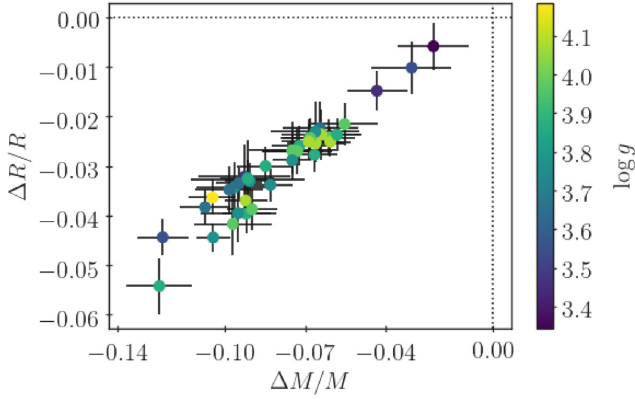
Figure 16. Relative difference in mass and radius between the inferred values and the underlying parameters of the coupled models. Positive values imply that the mass and radius have been overestimated.

Note that we neither use the reduced χ^2 to establish the best-fitting models nor to determine uncertainties. For this purpose, we use the mapped posterior probability density. Instead, we merely use χ^2 to discard targets from the analysis.

6.3 Results

Based on the reduced χ^2 values of the best-fitting standard stellar models, we discard 28 coupled models. This leaves us with a sample of 40 models. Since we raise caution about the inferred properties of coupled models near the TO in Section 5, it is worth mentioning that the majority of the 40 coupled models in our sample are giants and subgiants that lie far from the TO. Moreover, we note that the few coupled models that lie close to TO do not skew the sample or bias the qualitative and quantitative conclusions that are drawn in this section.

We find that the structural changes that are introduced by the improved boundary conditions are so large that we cannot accurately infer the global stellar properties of the underlying coupled stellar models from our grid of standard stellar models. We thus find that the mass and radius are consistently underestimated. We summarize these findings in Fig. 16.

As can be seen from Fig. 16, there is a clear correlation between the discrepancies in the inferred masses and radii. This reflects the fact that the best-fitting standard models approximately recover the mean densities of the underlying coupled models, since they are required to recover the individual mode frequencies.

In all 40 cases, the surface correction relation by Sonoi et al. (2015) lowers the model frequencies. Moreover, the relative change in $\Delta\nu$ as a result of the surface correction is roughly constant as a function of ν_{max} and is of the order of 10^{-2} for all models. The obtained behaviour of the inferred surface effect is thus similar to that obtained from asteroseismic analyses based on actual observations (e.g. Rodrigues et al. 2017, who use standard stellar models and scaling relations, and Jørgensen et al. 2020, who use standard stellar models).

We find that the ages of the coupled model are systematically overestimated and this often by more than 10 per cent. On average, the deviation in age is 8.8 ± 1.8 per cent. Furthermore, we find that the metallicity is systematically underestimated – on average by (-0.23 ± 0.04) dex. We illustrate both of these findings in Fig. 17. The effective temperature is systematically overestimated – on
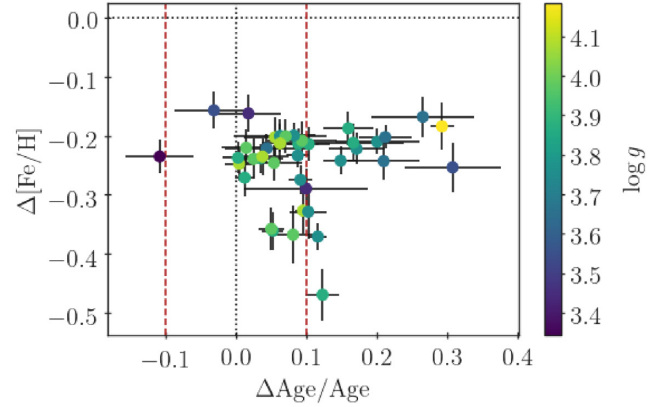

Figure 17. Absolute difference in metallicity and fractional difference in age between the inferred values and the underlying parameters of the coupled models. Positive values imply that the metallicity and age have been overestimated. The dashed red line divides the sample into those models that recover the correct age within 10 per cent and those that do not. The plot contains 40 stars that passed our selection criteria.
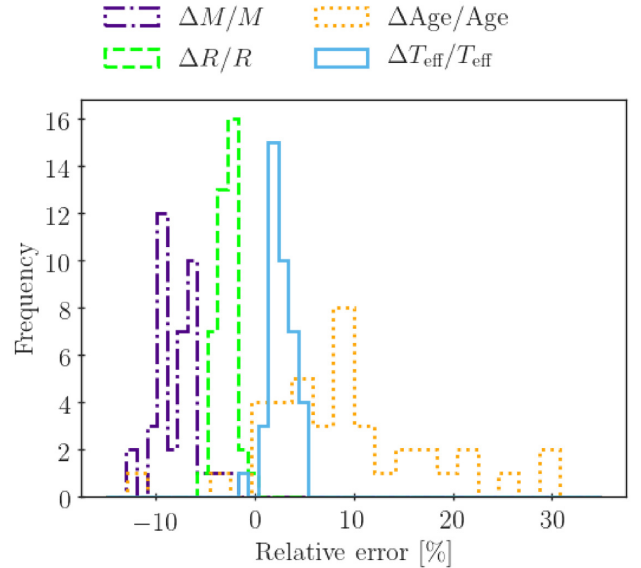
Figure 18. Relative error in the inferred stellar parameters for the stellar mass, radius, age, and effective temperature based on the samples presented in Figs 16 and 17.

average by (150 ± 28) K, i.e. 1.5σ, where σ denotes the attributed observational error. We illustrate this latter statement in Fig. 18.

Tayar et al. (2017) have evaluated the shift in the mixing length parameter that is necessary to recover observation constraints on over 3000 red giants based on standard stellar models (see also Appendix C). Based on their analysis, Tayar et al. (2017) conclude that the omission of this correction can affect isochrone ages by as much as a factor of 2, even when considering target stars with near-solar metallicity. Whether or not the asserted variation in α_{mlt} is indeed physical (Salaris et al. 2018), the results by Tayar et al. (2017) demonstrate the huge impact of the chosen input physics of stellar models on the derived parameter estimates. While the obtained errors in the stellar age in Fig. 17 are high, they do thus not seem implausible. Stellar ages obtained from asteroseismic analyses based on standard stellar models may thus suffer from significant systematic errors. This is also reflected in the age uncertainties determined from
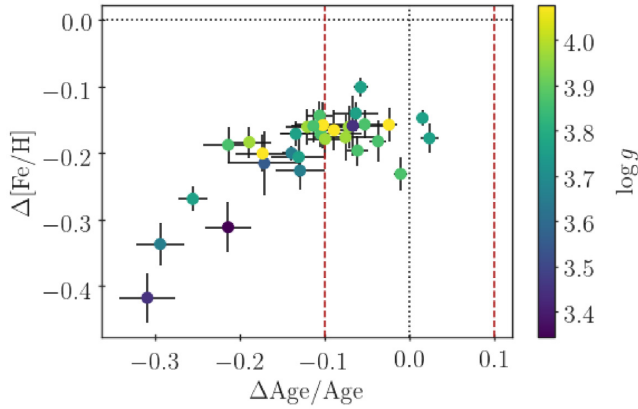
Figure 19. As Fig. 17 for a different likelihood: Here, we have substituted the constraint on the stellar luminosity with constraints on both the stellar mass and radius. The plot contains 31 stars that passed our selection criteria. These are all also in the subsample in Figs 16 and 17. Regarding the four outliers at low metallicity, we note that these are all early RGB stars with $T_{\mathrm{eff}} \lesssim 4800\,\mathrm{K}$ with $\log g \lesssim 3.7$ dex, for which the radius and mass estimates closely match the values of the underlying coupled models. The outliers are discussed further in the bulk text.

methods such as those used by Bellinger et al. (2016) and Angelou et al. (2017) where the input physics is varied widely. Even when studying the present-day Sun as a star by attempting to recover the solar properties based on observations, changes in the input physics of the models can play a significant role as shown by, e.g. Rendle et al. (2019).

As demonstrated in Fig. 8, the best-fitting standard stellar models of RGB stars should recover the correct age, if the models were to accurately predict the stellar masses, radii, and metallicities. It follows that the systematic errors in the inferred ages mirror the incorrectly deduced masses and radii, which in turn reflect the chosen constraints. We thus repeated the analysis, substituting the constraint on the luminosity with a constraint on the stellar radius. In practice, such constraints are available from interferometric measurements or dynamical studies of binaries. Based on studies of eclipsing binaries by White et al. (2013) and Gaulme et al. (2016), we assumed that an error of 2 per cent is feasible. Doing so, however, we arrive at similar qualitative mismatches. This is due to the fact that the accuracy, with which the radius is recovered, is already of the order of 2 per cent without including the radius in the likelihood.

For some stars, dynamical studies provide robust observational constraints on the stellar mass. In the best-case scenario, the statistical errors are of the order of 1 per cent or lower (Gaulme et al. 2016; Pourbaix & Boffin 2016). Adopting these optimistic uncertainties on the mass, we repeated the analysis. This time the errors in the obtained masses and radii are recovered within 6 per cent and 3 per cent, respectively, for all considered coupled models. Meanwhile, a large fraction of the RGB stars do still not recover the correct age within 10 per cent, as shown in Fig. 19.

Comparing Figs 17 and 19, we note that the inferred stellar ages go from being too high in Fig. 17 to being too low in Fig. 19. The inferred ages are hence sensitive to the change in the likelihood function. This implies that the individual frequencies do not dominate the likelihood. The observed age deviation can therefore not be explained by our treatment of the surface effect alone. The age deviations demonstrably reflect the changes in the global stellar parameters that arise from the improved boundary conditions (cf. Section 5). This is
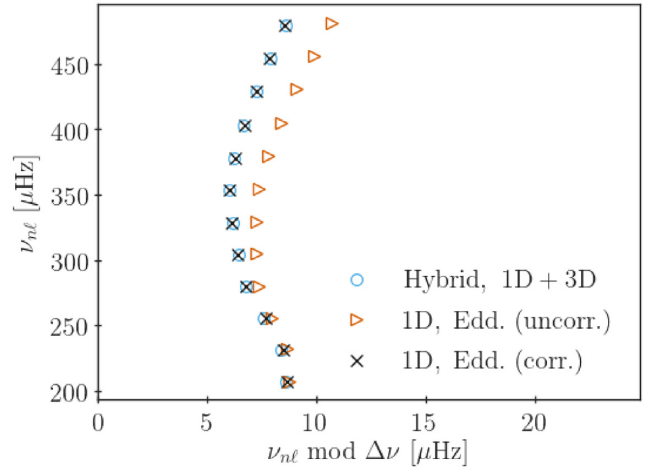
Figure 20. Echelle diagram showing the adiabatic radial mode frequencies of a coupled model within the so-called gas Γ_1 approximation and the uncorrected and corrected adiabatic frequencies of the associated best-fitting standard model found by AIMS using the surface correction relation by Sonoi et al. (2015). For the presented coupled model, $M = 0.91\,\mathrm{M_\odot}$, $T_{\mathrm{eff}} = 4724\,\mathrm{K}$, and $\log g = 3.45$ dex. The model corresponds to the outlier, for which $\Delta\mathrm{[Fe/H]} = -0.41 \pm 0.04$ dex in Fig. 19. As can be seen from the figure, the inferred surface effect behaves as expected. The same holds true for the other cases.

not to say that the frequencies are not well fitted. On the contrary, the Echelle diagrams look as expected. We show this in Fig. 20.

The samples, for which we infer the largest age discrepancies in Fig. 19, also lead to the largest discrepancies in metallicity. Fig. 19 thus contains 4 outliers, for which the deviations in metallicity lie between -0.27 and -0.41 dex, and for which the deviations in ages exceed that of the remaining 27 samples. There are, meanwhile, several (four) other models with similar $\log g$ (< 3.7 dex) yielding better age and metallicity estimates than the outliers. However, the higher accuracy in age and metallicity comes at the cost of lower accuracy in both the mass and radius. To fit the required stellar properties for the outliers, AIMS thus compensated for the impact of the different boundary conditions of standard and coupled stellar models on the global stellar properties by adjusting the stellar metallicity. Such offsets in metallicity are well known to affect stellar age estimates and can therefore explain the associated large deviations in age (e.g. Worthey 1994, 1999). However, even without taking these outliers into account, the mean deviation in age is still 5.8 ± 1.5 per cent. Moreover, ignoring the outliers, the offset in metallicity is roughly constant – this holds true for both Figs 17 and 19. The bias in age does not hence generally scale with the bias in metallicity.

We also repeated the analysis, including constraints on T_{eff}, [Fe/H], and ν_{max}. Here, we set the error on ν_{max} to 1 per cent. Once again, we reach the same qualitative conclusions regarding the offsets in mass, radius, metallicity, and age.

Independently of the constraints that enter our likelihood, we thus always end up with the same qualitative conclusion: The outer boundary conditions have an impact on the outcome of asteroseismic analyses through the resulting change in the global stellar parameters. However, the ability of standard stellar models to recover given properties of coupled stellar models is, of course, sensitive to these constraints. Indeed, it is well known that the stellar parameters of asteroseismic analyses reflect the chosen likelihood (e.g. Silva Aguirre et al. 2013; Basu & Kinnane 2018; Nsamba

et al. 2018). To give the non-seismic constraints higher impact, one might shift to using global seismic constraints, such as the large frequency separation, rather than individual frequencies. Using this approach, Jørgensen & Angelou (2019) are able to achieve mutually consistent parameter estimates for main-sequence stars based on coupled and standard stellar models – the individual frequencies are meanwhile not perfectly recovered. Alternatively, one might introduce more information from seismic constraints by drawing upon higher degree modes. This has been shown to be a successful strategy by Mosumgaard et al. (2020), who recover very similar global properties for *Kepler* stars using both coupled and standard stellar models. However, we also note that the modifications that are required for AIMS to correctly handle higher degree modes for more evolved stars lie beyond the scope of this paper.

Note that the findings by Jørgensen & Angelou (2019) and Mosumgaard et al. (2020) are consistent with our hare and hound exercise for the present-day Sun in Section 6.1.1. Moreover, this statement does not contradict the conclusions drawn in this section, as we do not address main-sequence stars here.

6.4 Discussion

A direct comparison between actual observations and the individual model frequencies of coupled models is hampered by inaccurate nature of the gas Γ_1 assumption (cf. Figs 1 and 5). The precision of the observed frequencies thus, by far, surpasses the accuracy of the model frequencies. For main-sequence stars, this issue can be avoided by circumventing the surface effect altogether. Rather than comparing observations to individual model frequencies, one might employ the frequency ratios proposed by Roxburgh & Vorontsov (2003). These ratios are insensitive to the surface layers, as shown by Otí Floranes, Christensen-Dalsgaard & Thompson (2005). With this in mind, coupled stellar models can successfully be applied in analysis of real stars as shown by Jørgensen & Angelou (2019) and Mosumgaard et al. (2020). However, it is yet not settled, whether the use of frequency ratios is a safe and viable strategy beyond the main sequence, due to the occurrence of mixed modes.

7 CONCLUSION

In this paper, we discuss coupled stellar models that combine state-of-the-art 1D standard stellar models with 3D simulations of the outermost layers of convective envelopes. Our work is a continuation of a series of papers, in which we have established the robustness and versatility of our method (Jørgensen et al. 2017, 2018, 2019; Jørgensen & Angelou 2019; Jørgensen & Weiss 2019; Mosumgaard et al. 2020). Our results can be summarized as follows:

(i) **We show that the uncertainties on the global solar parameters that enter solar calibrations allow for shifts in the individual model frequencies of the order of 1–2 μHz** (cf. Section 3.2). With this finding, we are able to explain the differences between the model frequencies obtained from different coupled and patched solar models that are presented in the literature (cf. Schou & Birch 2020). Moreover, we note that the remaining residuals between our coupled solar calibration models and observations lie below 2 μHz at all frequencies. Coupled stellar models have thus reduced the surface effect to become comparable to the established error bars.

(ii) **We demonstrate that the so-called gas Γ_1 approximation generally performs well for low-mass main-sequence stars (cf. Section 4).** For all stellar models within the explored region of the parameter space, the errors that are introduced by the so-called gas

Γ_1 approximation lie within 2.9 μHz at ν_{max}. The applicability of this approach beyond the case of the present-day Sun has previously not been validated.

(iii) **We find that the improved outer boundary layers of coupled models impact the predicted stellar properties across the HR diagram (cf. Section 5).** At fixed mass, age, and metallicity, the deviation between the effective temperatures of coupled stellar models and their standard stellar counterparts thus exceeds 80 K in some cases. The discrepancy in the stellar radius meanwhile ranges from a few per cent to 25 per cent. Discrepancies in the mean density and large frequency separation reach 2 and 3 per cent, respectively.

(iv) **In a hare and hound exercise, we demonstrate that the dissonance between standard and coupled stellar models affects the outcome of asteroseismic analyses (cf. Section 6).** In this exercise, we attempt to recover the global stellar parameters of the coupled stellar models ($3.4 \lesssim \log g \lesssim 4.1$) by drawing upon standard stellar models. We show that the inferred stellar properties deviate significantly from the ground truth. The deviation in the inferred stellar age thus often exceeds 10 per cent, which corresponds to the desired accuracy of the *PLATO* space mission – both for the core objectives of the mission and for the purpose of galactic archaeology (Rauer 2013; Miglio et al. 2017).

Although coupled stellar models give a more realistic depiction of the stellar surface layers than standard stellar models do, it is not settled whether coupled models also yield more accurate parameter estimates. However, our results demonstrate that the treatment of superadiabatic convection not only affects the model frequencies, but also alters the predicted global stellar parameters. In the light of the high-quality asteroseismic data from current and up-coming Earth-bound surveys and space missions, it is hence *not* enough to address the surface effect when attempting to deal with the shortcomings of standard stellar models. One must also consider the impact of a simplified depiction of superadiabatic convection on stellar evolution. Our results thus strongly advocate a synergy of state-of-the-art stellar evolution codes and multidimensional simulations of magnetohydrodynamics. Our coupled stellar models show a possible way towards achieving this synergy and thereby provide essential improvements towards the next generation of stellar models. As the next step in our exploration of coupled models, we will produce grids of coupled stellar models to be used in asteroseismic analyses of *Kepler*, *TESS*, and *PLATO* target stars.

DATA AVAILABILITY

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

REFERENCES

Aerts C., 2019, preprint (arXiv:1912.12300)
Angelou G. C., Bellinger E. P., Hekker S., Basu S., 2017, ApJ, 839, 116
Angelou G. C., Bellinger E. P., Hekker S., Mints A., Elsworth Y., Basu S., Weiss A., 2020, MNRAS, 493, 4987
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, Annu. Rev. Astron. Astrophys., 47, 481
Bahcall J. N., Serenelli A. M., Basu S., 2006, ApJS, 165, 400
Ball W. H., Gizon L., 2014, A&A, 568, A123
Ball W. H., Beeck B., Cameron R. H., Gizon L., 2016, A&A, 592, A159
Baraffe I., Pratt J., Goffrey T., Constantino T., Folini D., Popov M. V., Walder R., Viallet M., 2017, ApJ, 845, L6
Basu S., Antia H. M., 1997, MNRAS, 287, 189
Basu S., Antia H. M., 2008, Phys. Rep., 457, 217
Basu S., Kinnane A., 2018, ApJ, 869, 8
Bazot M., Bourguignon S., Christensen-Dalsgaard J., 2012, MNRAS, 427, 1847
Bellinger E. P., Christensen-Dalsgaard J., 2019, ApJ, 887, L1
Bellinger E. P., Angelou G. C., Hekker S., Basu S., Ball W. H., Guggenberger E., 2016, ApJ, 830, 31
Böhm-Vitense E., 1958, Z. Astrophys., 46, 108
Broomhall A.-M., Chaplin W. J., Davies G. R., Elsworth Y., Fletcher S. T., Hale S. J., Miller B., New R., 2009, MNRAS, 396, L100
Brown T. M., 1984, Science, 226, 687
Brown T. M., Christensen-Dalsgaard J., 1998, ApJ, 500, L195
Brown T. M., Gilliland R. L., Noyes R. W., Ramsey L. W., 1991, ApJ, 368, 599
Buldgen G., Salmon S., Noels A., 2019, Frontiers Astron. Space Sci., 6, 42
Canuto V. M., Mazzitelli I., 1991, AJ, 370, 295
Canuto V. M., Mazzitelli I., 1992, AJ, 389, 724
Canuto V. M., Goldman I., Mazzitelli I., 1996, ApJ, 473, 550
Christensen-Dalsgaard J., 1997a, in Bedding T. R., Booth A. J., Davis J., eds, Proc. IAU Symp. 189, Constraints on Stellar Interior Physics from Helioseismology. Kluwer, Dordrecht, p. 285
Christensen-Dalsgaard J., 1997b, in Pijpers F. P., Christensen-Dalsgaard J., Rosenthal C. S., eds, Effects of Convection on the Mean Solar Structure, SCORe'96: Solar Convection and Oscillations and their Relationship. Kluwer Acad. Publishers, Dordrecht, p. 3
Christensen-Dalsgaard J., 2008, Ap&SS, 316, 113
Christensen-Dalsgaard J., 2015, MNRAS, 453, 666
Christensen-Dalsgaard J., Thompson M. J., Gough D. O., 1989, MNRAS, 238, 481
Christensen-Dalsgaard J. et al., 1996, Science, 272, 1286
Christensen-Dalsgaard J., di Mauro M. P., Houdek G., Pijpers F., 2009, A&A, 494, 205
Christensen-Dalsgaard J., Gough D. O., Knudstrup E., 2018, MNRAS, 477, 3845
D'Antona F., Montalbán J., Kupka F., Heiter U., 2002, ApJ, 564, L93
Davies G. R., Broomhall A.-M., Chaplin W. J., Elsworth Y., Hale S. J., 2014, MNRAS, 439, 2025
Degl'Innocenti S., Dziembowski W. A., Fiorentini G., Ricci B., 1997, Astropart. Phys., 7, 77
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Freytag B., Steffen M., Ludwig H. G., Wedemeyer-Böhm S., Schaffenberger W., Steiner O., 2012, J. Comput. Phys., 231, 919
Gaulme P. et al., 2016, ApJ, 832, 121
Goodman J., Weare J., 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Gough D. O., 1990, in Osaki Y., Shibahashi H., eds, Comments on Helioseismic Inference, in Lecture Notes in Physics, Progress of Seismology of the Sun and Stars, Vol. 367. Springer-Verlag, Berlin, p. 283
Grevesse N., Noels A., 1993, in Hauck B., Paltani S., Raboud D., eds, Perfectionnement de l'Association Vaudoise des Chercheurs en Physique. p. 205
Grevesse N., Sauval A. J., 1998, Space Sci. Rev., 85, 161
Handberg R., Campante T. L., 2011, A&A, 527, A56

Handberg R., Brogaard K., Miglio A., Bossini D., Elsworth Y., Slumstrup D., Davies G. R., Chaplin W. J., 2017, MNRAS, 472, 979
Hekker S., 2020, Frontiers Astron. Space Sci., 7, 3
Houdek G., Trampedach R., Aarslev M. J., Christensen-Dalsgaard J., 2017, MNRAS, 464, L124
Houdek G., Lund M. N., Trampedach R., Christensen-Dalsgaard J., Handberg R., Appourchaux T., 2019, MNRAS, 487, 595
Huber D. et al., 2011, ApJ, 743, 143
Jørgensen A. C. S., 2019, Combining Stellar Structure and Evolution Models with Multi-Dimensional Hydrodynamic Simulations (Dissertation, LMU München: Faculty of Physics). available at: http://nbn-resolving.de/urn:nbn:de:bvb:19-250031
Jørgensen A. C. S., Angelou G. C., 2019, MNRAS, 490, 2890
Jørgensen A. C. S., Weiss A., 2018, MNRAS, 481, 4389
Jørgensen A. C. S., Weiss A., 2019, MNRAS, 488, 3463
Jørgensen A. C. S., Weiss A., Mosumgaard J. R., Silva Aguirre V., Sahlholdt C. L., 2017, MNRAS, 472, 3264
Jørgensen A. C. S., Mosumgaard J. R., Weiss A., Silva Aguirre V., Christensen-Dalsgaard J., 2018, MNRAS, 481, L35
Jørgensen A. C. S., Weiss A., Angelou G., Silva Aguirre V., 2019, MNRAS, 484, 5551
Jørgensen A. C. S. et al., 2020, MNRAS, 495, 4965
Kippenhahn R., Weigert A., Weiss A., 2012, Stellar Structure and Evolution, 2nd edn., Springer, Berlin
Kjeldsen H., Bedding T. R., Christensen-Dalsgaard J., 2008, AJ, 683, L175
Krishna Swamy K. S., 1966, AJ, 145, 174
Ludwig H. G., Freytag B., Steffen M., 1997, in Pijpers F. P., ChristensenDalsgaard J., Rosenthal C. S., eds, A Calibration of Mixing Length Theory Based on RHD Simulations of Solar-Type Convection, SCORe'96: Solar Convection and Oscillations and their Relationship. Kluwer Acad. Publishers, Dordrecht, p. 59
Ludwig H.-G., Freytag B., Steffen M., 1999, A&A, 346, 111
Lund M. N., Reese D. R., 2018, in Campante T. L., Santos N. C., Monteiro M. J. P. F. G., eds, Asteroseismology and Exoplanets: Listening to the Stars and Searching for New Worlds, Vol. 49. Springer Int. Publ. AG, p. 149
Lund M. N. et al., 2017, AJ, 835, 172
Magic Z., Weiss A., 2016, A&A, 592, A24
Magic Z., Collet R., Asplund M., Trampedach R., Hayek W., Chiavassa A., Stein R. F., Nordlund Å., 2013, A&A, 557, A26
Magic Z., Weiss A., Asplund M., 2015, A&A, 573, A89
Manchon L., Belkacem K., Samadi R., Sonoi T., Marques J. P. C., Ludwig H. G., Caffau E., 2018, A&A, 620, A107
Mazzitelli I., D'Antona F., Caloi V., 1995, A&A, 302, 382
Miglio A. et al., 2017, Astron. Nachr., 338, 644
Montalbán J., Miglio A., Noels A., Grevesse N., di Mauro M. P., 2004, in Danesy D., ed., ESA SP-559: SOHO 14 Helio- and Asteroseismology: Towards a Golden Future. ESA, Noordwijk, p. 574
Montalban J., Miglio A., Theado S., Noels A., Grevesse N., 2006, Commun. Asteroseismol., 147, 80
Mosser B. et al., 2012, A&A, 537, A30
Mosumgaard J. R., Silva Aguirre V., Weiss A., Christensen-Dalsgaard J., Trampedach R., 2017, in Monteiro M., Cunha M., Ferreira J., eds, Seismology of the Sun and the Distant Stars II, Vol. 160. EPJ Web Conf., Les Ulis Cedex A, France, p. 03009
Mosumgaard J. R., Ball W. H., Silva Aguirre V., Weiss A., Christensen-Dalsgaard J., 2018, MNRAS, 478, 5650
Mosumgaard J. R., Jørgensen A. C. S., Weiss A., Silva Aguirre V., Christensen-Dalsgaard J., 2020, MNRAS, 491, 1160
Nsamba B., Campante T. L., Monteiro M. J. P. F. G., Cunha M. S., Rendle B. M., Reese D. R., Verma K., 2018, MNRAS, 477, 5052
Otí Floranes H., Christensen-Dalsgaard J., Thompson M. J., 2005, MNRAS, 356, 671
Piau L., Collet R., Stein R. F., Trampedach R., Morel P., Turck-Chièze S., 2014, MNRAS, 437, 164
Planck Collaboration XIII, 2016, A&A, 594, A13
Porqueres N., Kodi Ramanah D., Jasche J., Lavaux G., 2019a, A&A, 624, A115

Porqueres N., Jasche J., Lavaux G., Enßlin T., 2019b, A&A, 630, A151

Pourbaix D., Boffin H. M. J., 2016, A&A, 586, A90

Rauer H., 2013, in European Planetary Science Congress. p. EPSC2013

Reese D. R., 2016, AIMS: Asteroseismic Inference on a Massive Scale, Astrophysics Source Code Library, record ascl:1611.014

Rendle B. M. et al., 2019, MNRAS, 484, 771

Rodrigues T. S. et al., 2017, MNRAS, 467, 1433

Rosenthal C. S., Christensen-Dalsgaard J., Nordlund Å., Stein R. F., Trampedach R., 1999, A&A, 351, 689

Roxburgh I. W., Vorontsov S. V., 2003, A&A, 411, 215

Sahlholdt C. L., Silva Aguirre V., 2018, MNRAS, 481, L125

Salaris M., Cassisi S., 2015, A&A, 577, A60

Salaris M., Cassisi S., Schiavon R. P., Pietrinferni A., 2018, A&A, 612, A68

Schlattl H., Weiss A., 1999, A&A, 347, 272

Schlattl H., Weiss A., Ludwig H.-G., 1997, A&A, 322, 646

Schou J., Birch A. C., 2020, preprint (arXiv:2004.13548)

Scuflaire R., Théado S., Montalbán J., Miglio A., Bourge P. O., Godart M., Thoul A., Noels A., 2008, Ap&SS, 316, 83

Serenelli A. M., Basu S., 2010, ApJ, 719, 865

Serenelli A. M., Basu S., Ferguson J. W., Asplund M., 2009, ApJ, 705, L123

Serenelli A., Peña-Garay C., Haxton W. C., 2013, Phys. Rev. D, 87, 043001

Silva Aguirre V. et al., 2013, AJ, 769, 141

Silva Aguirre V. et al., 2015, MNRAS, 452, 2127

Sonoi T., Samadi R., Belkacem K., Ludwig H.-G., Caffau E., Mosser B., 2015, A&A, 583, A112

Sonoi T., Ludwig H. G., Dupret M. A., Montalbán J., Samadi R., Belkacem K., Caffau E., Goupil M. J., 2019, A&A, 621, A84

Tanner J. D., Basu S., Demarque P., 2013, ApJ, 778, 117

Tayar J. et al., 2017, AJ, 840, 17

Trampedach R., Stein R. F., 2011, ApJ, 731, 78

Trampedach R., Asplund M., Collet R., Nordlund Å., Stein R. F., 2013, AJ, 769, 18

Trampedach R., Stein R. F., Christensen-Dalsgaard J., Nordlund Å., Asplund M., 2014, MNRAS, 445, 4366

Trampedach R., Aarslev M. J., Houdek G., Collet R., Christensen-Dalsgaard J., Stein R. F., Asplund M., 2017, MNRAS, 466, L43

Vernazza J. E., Avrett E. H., Loeser R., 1981, ApJS, 45, 635

Villante F. L., Serenelli A., 2020, preprint (arXiv:2004.06365)

Vinyoles N. et al., 2017, ApJ, 835, 202

Weiss A., Schlattl H., 2008, Ap&SS, 316, 99

White T. R. et al., 2013, MNRAS, 433, 1262

Worthey G., 1994, ApJS, 95, 107

Worthey G., 1999, in Hubeny I., Heap S., Cornett R., eds, ASP Conf. Ser. Vol. 192, Spectrophotometric Dating of Stars and Galaxies. Astron. Soc. Pac., San Francisco, p. 283

APPENDIX A: FREQUENCY COMPUTATION

In this paper, we compute stellar model frequencies using the adiabatic oscillation code ADIPLS. Such codes compute stellar oscillations by considering perturbations around the equilibrium state. Within the assumption of adiabaticity, the relative Lagrangian perturbation in the density ($\delta\rho/\rho$) and the relative Lagrangian perturbation in the thermal pressure ($\delta P_{th}/P_{th}$) are closely related:

$$\frac{\delta\rho}{\rho} = \frac{1}{\Gamma_1}\frac{\delta P_{th}}{P_{th}}. \tag{A1}$$

Here, Γ_1 refers to the first adiabatic index. Note that equation (A1) does not include any reference to the turbulent pressure (P_{turb}). In contrast, the equilibrium structure fulfils hydrostatic equilibrium and is hence dictated by the total pressure $P = P_{th} + P_{turb}$. Both the thermal and total pressures hence enter the frequency calculations. However, adiabatic oscillation codes do not distinguish between different forms of pressure. To circumvent this issue, we need to introduce assumptions that allow us to describe the oscillations solely in terms of relative Lagrangian perturbations in the total pressure.

Rewriting equation (A1), we find that (cf. Houdek et al. 2017)

$$\frac{\delta\rho}{\rho} = \frac{1}{\Gamma_1}\frac{P}{P_{th}}\left(\frac{\delta P}{P} - \frac{\delta P_{turb}}{P}\right). \tag{A2}$$

As argued by Houdek et al. (2017) based on non-adiabatic pulsation calculations, perturbations in the turbulent pressure mainly contribute to the non-adiabatic effects and hence to the modal effects – that is, at least, in the case of the present-day Sun. It follows that the Lagrangian perturbation of the turbulent pressure can be neglected when computing the *structural* contribution to the surface effect. By setting δP_{turb} to zero in equation (A2), we find that

$$\frac{\delta\rho}{\rho} \approx \frac{1}{\Gamma_1^{red}}\frac{\delta P}{P}, \tag{A3}$$

where $\Gamma_1^{red} = \Gamma_1 P_{th}/P$. Neglecting the Lagrangian perturbation of the turbulent pressure thus amounts to altering the first adiabatic index by the ratio between the thermal and total pressures. In the literature, this approach is therefore referred to as the reduced Γ_1 approximation.

Alternatively, we may assume that the relative Lagrangian perturbation of the thermal pressure equals the relative Lagrangian perturbation of turbulent pressure. The relative Lagrangian perturbation of the thermal pressure therefore amounts to the relative Lagrangian perturbation of the total pressure, i.e. $\delta P_{th}/P_{th} = \delta P/P$. Following this assumption, equation (A1) takes the form

$$\frac{\delta\rho}{\rho} \approx \frac{1}{\Gamma_1}\frac{\delta P}{P}. \tag{A4}$$

This approach is known as the gas Γ_1 approximation in the literature.

APPENDIX B: OVERCOMING INTERPOLATION ERRORS

As discussed by Mosumgaard et al. (2020), coupled stellar models are subject to interpolation errors, due to the insufficient resolution of the underlying grids of 3D RHD simulations. In addition, we note that the stellar evolution codes rely on linear interpolation algorithms when computing the temperature and turbulent pressure profiles of coupled stellar models (cf. flowchart in fig. 1 in Jørgensen & Weiss 2019). These algorithms recover the structure of the ⟨3D⟩ envelopes with lower accuracy than the algorithms presented by Jørgensen et al. (2017), since Jørgensen et al. (2017) interpolate in the (T_{eff}, $\log g$) plane by constructing piecewise cubic, continuously differentiable surfaces (see also chapter 5 in Jørgensen 2019). Both the insufficiently low resolution of the grid of 3D simulations and the linear interpolation ultimately lead to kinks in the evaluated stellar evolution tracks.

To overcome the errors that result from the linear interpolation, we have used a piecewise cubic interpolation to produce a denser artificial grid of ⟨3D⟩ envelopes. This denser grid is now employed by our stellar structure and evolution code to construct stellar models using linear interpolation in this paper.

Our artificial grid of ⟨3D⟩ envelopes is shown in Fig. B1 together with stellar evolution tracks. We find that the use of the artificial denser grid greatly improves the stellar evolution tracks, especially on the RGB. The unphysical kinks on the RGB have been reduced to such an extent that we deem an analysis of red giants with our coupled models both meaningful and feasible. Although the unphysical kinks have partly disappeared, the tracks still show small irregularities. The issue of the kinks has thus not been entirely solved by the use of the artificial grid, and an extension of the Stagger grid is still much desired to further improve the interpolation (see also Jørgensen et al.
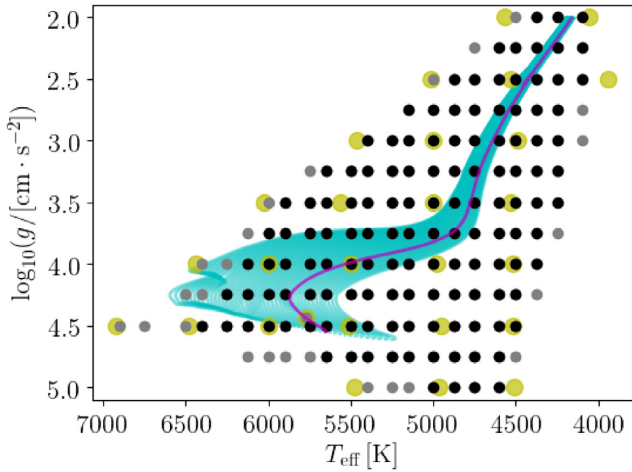
Figure B1. Artificial grid of 3D RHD simulations. The black dots denote the simulations that are available at all metallicities. The grey dots show the extent of the grid in the (T_{eff}, $\log g$) plane at solar metallicity. The purple track shows the evolution of a $1\,M_\odot$ star at solar metallicity. The cyan lines show the evolution tracks of all models that enter the analysis in Section 5 at solar metallicity. These tracks span masses between 0.88 and $1.32\,M_\odot$. The locations of the original 3D simulations in the Stagger grid are marked with yellow circles.

2017, 2019). The Stagger grid has a spacing in $\log g$ and T_{eff} of 0.5 dex and 500 K, respectively. On the RGB especially, additional 3D simulations at intermediate values of $\log g$ and T_{eff} are required.

The use of such artificial grids also allows for a straightforward rudimentary interpolation in metallicity at every time-step of the evolution and at a low computational cost. One can compute artificial grids with any metallicity, including intermediate metallicities that do not exist in the underlying grid of 3D simulations, by employing the interpolation scheme presented by Jørgensen et al. (2019). At every time-step, the stellar evolution code can then select that artificial grid of models that matches the current metallicity of the convective envelope, achieving a discrete resolution in [Fe/H]. We have introduced this scheme in both GARSTEC and the CLÉS stellar evolution code.

Since the stellar evolution code chooses that artificial grid, whose metallicity most closely matches the composition of the convective envelope at every time-step, the code can to some extent adapt to composition changes that arise from atomic diffusion. In this manner, we address changes in metallicity from one time-step to the next. In this paper, we have computed artificial grids with a resolution in [Fe/H] of 0.1 dex, but with the exception of Sections 3.2 we do not include atomic diffusion.

In connection with the chemical composition of the models, we note that the Stagger grid does not contain models with different helium contents at fixed metallicity. While one could partly account for this by introducing an offset between the surface gravity of the appended ⟨3D⟩ envelope and the interior (Tanner, Basu & Demarque 2013), this issue limits the degree, to which coupled models can in practice account for variations in the chemical profile. We also note that the helium abundance in the 3D simulations decreases with increasing metallicity. This means that the composition of the ⟨3D⟩ envelopes is slightly at odds with the chemical evolution of the galaxy when addressing stars with a composition that deviates from that of the present-day Sun. These considerations call for further extensions of the Stagger grid to increase its versatility in connection with stellar evolution calculations.

Like for the interpolation in the (T_{eff}, $\log g$) plane, additional 3D simulations at intermediate values are still needed to improve the interpolation in metallicity. For a discussion on how well the interpolation in metallicity performs across the Stagger grid, we refer the reader to Jørgensen et al. (2019).

Rather than introducing an artificial grid of 3D simulations, one might as well include higher order interpolation schemes directly into the stellar evolution code. Although this might come at a higher computational cost, it is doubtlessly a viable solution. A thorough exploration of different higher order interpolation techniques is, however, beyond the scope of this paper. Here, we have settled for the mentioned approach based on artificial grids. From this exercise, we can already conclude that the coupling scheme by Jørgensen et al. (2018) and Jørgensen & Weiss (2019) in tandem with the interpolation scheme by Jørgensen et al. (2017) performs very well across the entire parameter space. The accuracy of the interpolation scheme by Jørgensen et al. (2017) is, meanwhile, affected by the order of the interpolation that underlies its implementation as well as the employed triangulation (cf. Mosumgaard et al. 2020).

APPENDIX C: VARYING THE MIXING LENGTH PARAMETER

In this paper, we assume that the mixing length parameter is constant throughout the computed stellar evolution. However, different studies demonstrate the need to vary α_{MLT} across the HR diagram when attempting to encapsulate the correct properties of stellar structures and their evolution using standard stellar models. Indeed, by varying α_{MLT}, one might partly counteract the impact of the improper outer boundary layers of standard stellar models. This being said, we note that it is still not settled how to vary the mixing length parameter across the HR diagram and throughout the stellar interior (e.g. Ludwig, Freytag & Steffen 1997, 1999; Schlattl, Weiss & Ludwig 1997; Trampedach et al. 2014; Magic et al. 2015; Tayar et al. 2017; Sonoi et al. 2019; Angelou et al. 2020).

Tayar et al. (2017) investigate over 3000 red giants with $\log g$ between approximately 1.5 and 3.5, in order to evaluate what changes are needed in α_{MLT} for standard stellar models to recover the observational constraints from both APOGEE and *Kepler*. They conclude that α_{MLT} is sensitive to [Fe/H]. Concretely, they state that a change in the metallicity of 1.0 dex requires a change in α_{MLT} of 0.2 on the RGB.
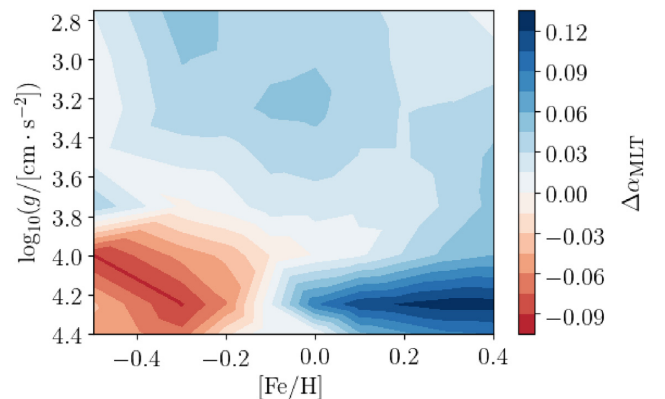
Figure C1. Variation in the mixing length parameter that is required for the standard stellar models to recover the effective temperature at different $\log g$ for a $1.0\,M_\odot$ coupled model. We only include models for which $\log g \geq 2.75$ to avoid ambiguities that arise from the RGB bump.
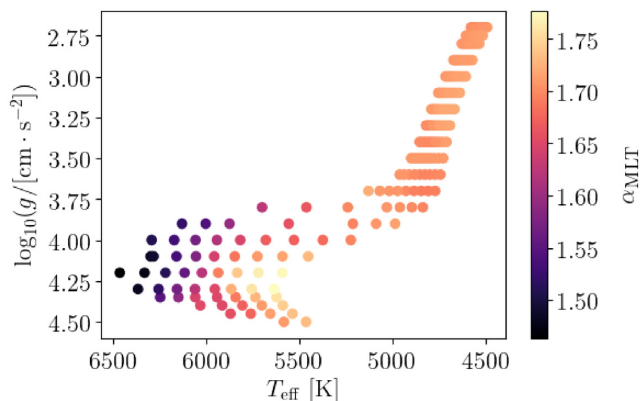
Figure C2. Kiel diagram, showing the variation in α_{MLT} that is needed for the standard stellar models to recover the same position in the Kiel diagram as their coupled stellar model counterparts. For all models, [Fe/H] = 0.0. We compare standard and coupled stellar models with the same mass. Two outliers on the subgiant branch with log g = 3.9 as well as one outlier at log g = 3.8 were excluded from the plot. The mixing length parameter obtained for standard stellar model from a solar calibration with the same input physics is 1.67.

To address the statement by Tayar et al. (2017) and to contribute to the discussion regarding the metallicity dependence of the mixing length parameter, we have computed a set of standard stellar models, varying the α_{MLT} around the solar calibrated value. Mirroring the approach by Tayar et al. (2017), we have then selected that value of α_{MLT} that recovers the value T_{eff} of coupled stellar models for different values of log g. In all cases, we have compared coupled and standard stellar models with the same mass. For simplicity, we have fixed the mass to 1.0 $M_{\odot}$. We varied $\Delta\alpha_{MLT}$ between −0.18 and +0.18 in steps of 0.02. The results are summarized in Fig. C1.

As can be seen from the figure, the variation of α_{MLT} on the main sequence is consistent with the variation found by Tayar et al. (2017). However, for red giants, we do not find the same behaviour. Indeed, on the RGB, we only observe a limited variation in $\Delta\alpha_{MLT}$ with metallicity. This makes sense in the light of the results presented in Section 5.2, where we find that the discrepancy in T_{eff} varies less on the RGB than on the main sequence as a function of metallicity. In Section 5.2, we, furthermore, find that the discrepancy in T_{eff} does not change sign on the RGB, i.e. the standard stellar models are consistently (≤40 K) too cold.

Our results are consistent with those of Salaris et al. (2018), who re-analysed the stars addressed by Tayar et al. (2017) and found

that Tayar et al. (2017) had not accounted for alpha enhancement. When only considering stars with scaled solar metal mixture, i.e. low α-enhancement, Salaris et al. (2018) do not recover the metal dependence of the calibrated mixing length parameter. Meanwhile, Salaris et al. (2018) note that the metal dependence of α_{MLT} found by Tayar et al. (2017) is reintroduced when α-enhanced stars are included.

Trampedach et al. (2014), Magic et al. (2015), and Sonoi et al. (2019) have investigated the variation of α_{MLT} across the Kiel diagram based on 3D RHD simulations. To investigate the variation of α_{MLT} across the Kiel diagram and to compare to Trampedach et al. (2014), Magic et al. (2015), and Sonoi et al. (2019), we follow a similar approach to that used for the analysis in connection with Fig. C1. We thus compute a set of standard stellar models, for which we vary α_{MLT} around the solar calibrated value. This time, we vary the stellar mass but keep the metallicity fixed to the solar value. For the standard stellar models, we thus vary $\Delta\alpha_{MLT}$ between −0.30 and +0.30 in steps of 0.02 and vary the mass from 0.90 to 1.3 $M_{\odot}$ in steps of 0.05. Again, we compare standard and coupled stellar models with the same mass. Following this procedure, our analysis of the variation in the (T_{eff}, log g) plane matches the approaches that underlie the investigation of the metallicity dependence presented in Fig. C1. The results are summarized in Fig. C2.

We find the same overall qualitative trends as Trampedach et al. (2014), Magic et al. (2015), and Sonoi et al. (2019) (cf. figs 4 and 2 and table A.1 in Trampedach et al. 2014; Magic et al. 2015; Sonoi et al. 2019, respectively). Thus, we find that lower values of α_{MLT} are generally needed at higher effective temperatures when considering stars at the same log g. To quantify this statement, we compared our results for the variation of α_{MLT} with the corresponding results obtained by Magic et al. (2015) and Sonoi et al. (2019). To facilitate a meaningful comparison, we limited ourselves to those cases, for which the underlying 3D simulations used by Magic et al. (2015) and Sonoi et al. (2019) lie within the region covered in Fig. C2. There are 10 such cases, excluding the present-day Sun. In all 10 cases, we find that we infer the same values for α_{MLT} as Magic et al. (2015) and Sonoi et al. (2019) do within 0.04. Furthermore, in all 10 cases, we find the inferred shift in α_{MLT} to have the same sign as the corresponding shifts found by Magic et al. (2015) and Sonoi et al. (2019). The fact that our results mirror those found by other authors that use different methods further underlines the validity and flexibility of our coupling scheme.

This paper has been typeset from a T$_{\rm E}$X/L$^{\rm A}$T$_{\rm E}$X file prepared by the author.