

Mending the structural surface effect of 1D stellar structure models with non-solar metallicities based on interpolated 3D envelopes

Andreas Christ Sølvesten Jørgensen,^{1★} Achim Weiss,¹ George Angelou¹ and Víctor Silva Aguirre²

¹Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany

²Stellar Astrophysics Centre (SAC), Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark

Accepted 2019 January 30. Received 2019 January 30; in original form 2018 November 8

ABSTRACT

One-dimensional (1D) stellar evolution codes employ rudimentary treatments of turbulent convection. For stars with convective envelopes, this leads to systematic errors in the predicted oscillation frequencies needed for asteroseismology. One way of mending these structural inadequacies is through patching, whereby the outermost layers of 1D models are replaced by the mean stratifications from three-dimensional (3D) simulations. In order to viably implement this approach in asteroseismic analysis, interpolation throughout pre-computed 3D envelopes is required. We present a method that interpolates throughout pre-computed 3D envelopes as a function of effective temperature, surface gravity, and metallicity. We conduct a series of validation tests that demonstrate that the scheme reliably and accurately reproduces the structures of stellar envelopes and apply our method to the Sun as well as two stars observed by *Kepler*. We parametrize the frequency shift that results from patching and show that the functional forms are evolutionary dependent. In addition, we find that neglecting modal effects, such as non-adiabatic energetics, introduces systematic errors in asteroseismically obtained stellar parameters. Both these results suggest that a cautious approach is necessary when utilizing empirical surface corrections in lieu of patching models. Our results have important implications, particularly for characterizing exoplanet systems, where accuracy is of utmost concern.

Key words: asteroseismology – Sun: helioseismology – stars: atmospheres – stars: interiors.

1 INTRODUCTION

Asteroseismology, i.e. the study of stellar oscillations (Leighton 1960; Ulrich 1970), has proven invaluable for the precise determination of stellar properties as well as for exoplanet research (e.g. Christensen-Dalsgaard, Kjeldsen & Brown 2010; Batalha 2014; Lebreton & Goupil 2014). Through the constraints placed on internal structures, both the CoRoT (Baglin et al. 2009) and *Kepler* (Borucki et al. 2010) space missions have profoundly contributed to the field of stellar physics. Comprehensive reviews of helioseismology (Christensen-Dalsgaard 2002) and asteroseismology (Aerts, Christensen-Dalsgaard & Kurtz 2010; Chaplin & Miglio 2013) outline many of the successes. In this work, we focus in particular on solar-like oscillations in the form of p modes. These modes are acoustic oscillations that are stochastically excited by turbulent convection, whereby pressure acts as the restoring force.

The accuracy of seismically inferred stellar properties strongly depends on the underlying stellar structure models and the interpretation of asteroseismic data is hence subject to any model inadequacies. One such deficiency of current one-dimensional (1D) stellar evolution codes is the incorrect depiction of the structure of superadiabatic convective layers, due to a simplistic parametrization of turbulent convection (e.g. Böhm-Vitense 1958). For stars with convective envelopes, such as solar-like stars, this structural discrepancy leads to a deficient representation of the surface layers, altering the predicted stellar p-mode frequencies.

The resulting systematic offset between the model frequencies and asteroseismic observations is referred to as the surface effect (Brown 1984; Christensen-Dalsgaard, Däppen & Lebreton 1988). The different methods for dealing with the surface effect lead to different systematic errors when evaluating stellar parameters (Ball & Gizon 2017; Nsamba et al. 2018), since seismic constraints are used to infer these parameters. Systematic errors in these parameter estimates affect the conclusion drawn in stellar physics

* E-mail: acsj@mpa-garching.mpg.de

and related or dependent fields, such as exoplanet research or galactic archaeology.

In order to mitigate the surface effect, some authors employ a set of frequency ratios suggested by Roxburgh & Vorontsov (2003) that have been shown to be less sensitive to the surface layers (Oti Floranes, Christensen-Dalsgaard & Thompson 2005; Silva Aguirre et al. 2011). Alternatively, many authors rely on empirical or theoretical surface correction relations. A broad variety of these relations can be found in the literature, of which we will refer to the following in the present work: Kjeldsen, Bedding & Christensen-Dalsgaard (2008) suggest a purely empirical power-law correction relation based on analysis of the present Sun. Ball & Gizon (2014) present a theoretically motivated relation, based on an asymptotic analysis by Gough (1990). Sonoi et al. (2015) suggest yet a different functional form based on ten 1D structures, for which the authors have substituted the outermost layers by mean stratifications of three-dimensional (3D) hydrodynamical simulations.

Correcting the structural contribution to the surface effect by substituting the envelope with mean stratifications from hydrodynamic simulations is known as patching and was originally suggested by Schlattl, Weiss & Ludwig (1997) and Rosenthal et al. (1999). This was done based on 1D or at most 2D models. In the last few years, patching has undergone somewhat of a renaissance with several authors adopting this strategy, but using the now available 3D models (Piau et al. 2014; Sonoi et al. 2015; Ball et al. 2016; Magic & Weiss 2016; Jørgensen et al. 2017; Trampedach et al. 2017). However, 3D hydrodynamical simulations are computationally expensive. Rather than computing 3D envelopes that match the restrictions set by a given interior stellar structure, the interior stellar structure was previously chosen in such a way as to match existing 3D simulations. The use of patching has therefore been limited to the analyses of a few special cases, most prominently the present Sun. Thus, until recently, patching has not been a viable alternative to empirical surface correction relations.

Recently, Jørgensen et al. (2017) determined a scheme to interpolate mean stratifications from 3D simulations. This method has proven to accurately mimic the structure of 3D envelopes and has been successfully applied in the analysis of four solar-type dwarfs from the LEGACY and KAGES samples in the *Kepler* field (Silva Aguirre et al. 2015; Davies et al. 2016; Lund et al. 2017; Silva Aguirre et al. 2017). This method allows for the interpolation of the mean 3D structure in the effective temperature (T_{eff}) and the gravitational acceleration (g) but does not include an interpolation in the composition of the substituted envelope, which limits the applicability to stars with near-solar metallicity.

In this paper, we extend the work by Jørgensen et al. (2017, hereafter Paper I) by additionally interpolating in $[\text{Fe}/\text{H}]$. The sparsity of the 3D simulations in this dimension increases the complexity for the original scheme. We show that the method reliably reproduces the mean structure of 3D simulations throughout the parameter space spanned by T_{eff} , $\log g$ as well as $[\text{Fe}/\text{H}]$. In this paper, we employ the 3D hydrodynamic simulations of stellar envelopes by Magic et al. (2013a). Hereafter, we will refer to these simulations as the Stagger grid. More details are supplied in Section 2.

The multidimensional scheme allows us to construct patched stellar models without being restricted to existing 3D simulations. This allows us to map and investigate the structural contribution to the surface effect throughout the parameter space. We can express the dependence of the structural surface effect as a

function of T_{eff} , $\log g$ and $[\text{Fe}/\text{H}]$, and as per Sonoi et al. (2015) parametrize the structural contribution to the surface effect (sans modal effects, see below). With the current method, it is tractable to calculate a vast grid of patched models (PMs) and determine a surface effect formulation from a well-resolved parameter space.

As we neglect non-adiabatic energetics and do not treat the contribution of the turbulent pressure consistently, we do not expect that the obtained parametrization reproduces observations exactly; the current parametrization considers only the structural contribution of the surface effect – ignoring, however, the structural contribution that amounts to an adjustment of the adiabatic index by the turbulent pressure. These neglected contributions are collectively referred to as modal effects (cf. Houdek Trampedach, Aarslev & Christensen-Dalsgaard 2017).

In Sections 2 and 3, we discuss the employed grid of 3D simulations and summarize our patching procedure and interpolation scheme. We test the accuracy of the interpolation scheme and investigate the effect of the obtained interpolation error on the model frequencies based on the solar case. In Section 4, we introduce and compare different parametrizations of the structural surface effect. Using a maximum likelihood estimation (MLE) method, we use one such parametrization to estimate stellar parameters in Section 5. The obtained results are compared to those determined based on surface correction relations from the literature. Section 6 summarizes our conclusions.

2 THE STAGGER GRID

We constructed PMs based on the Stagger grid: a grid of 3D hydrodynamic simulations of stellar envelopes published by Magic et al. (2013a). These 3D simulations all stem from *box-in-a-star* type simulations: that is, these simulations only cover a small representative volume of the outermost convective layers. Since the depth of the simulations only amounts to 0.4–10 per cent of the total stellar radius (Magic et al. 2013a), any sphericity effects as well as changes in $\log g$ are neglected. The envelope models extend from the nearly adiabatic surface layers to the atmosphere, including the photosphere.

The Stagger grid was computed with a successor of the Stein & Nordlund (1998) finite-difference large-eddy simulation code, using the Mihalas–Hummer–Däppen equation of state (MHD-EOS; Hummer & Mihalas 1988), the composition published by Asplund et al. (2009; AGSS09), and the MARCS line opacities listed in Gustavsson et al. (2008).

The Stagger grid contains 206 3D envelopes, of which we excluded 6, since their effective temperature strongly oscillates with time, indicating that the simulations are not yet fully relaxed. In addition, one simulation was excluded due to its subadiabatic stratification: the temperature gradient is lower than the adiabatic gradient throughout the simulation (Collet, private communication).

The grid contains envelopes with T_{eff} between 4000 and 7000 K, $\log g$ between 1.5 and 5.0, and $[\text{Fe}/\text{H}]$ of -4.0 , -3.0 , -2.0 , -1.0 , -0.5 , 0.0 and 0.5 . The spacing in T_{eff} and $\log g$ is 500 K and 0.5 dex, respectively. For $[\text{Fe}/\text{H}] \leq -1$, α -enhancement has been taken into account: $[\alpha/\text{Fe}] = 0.4$ dex (Magic et al. 2013b).

In our substitution of the outermost layers, we employ the spatial average taken over layers of constant geometric depth and time, in order to conserve hydrostatic equilibrium. We will refer to these averaged structures as $\langle 3D \rangle$ envelopes. For a discussion on the influence of the averaging on the evaluated stellar oscillation frequencies, we refer to Magic & Weiss (2016).

3 THE PATCHING AND INTERPOLATION PROCEDURES

Patching refers to the replacement of the outermost layers of 1D stellar models with a (3D) envelope. Using the nomenclature from Paper I, we refer to the lowermost point of the patched envelope as the patching point. At radii that are larger than the distance from the patching point to the centre, all mesh points from the un-patched model (UPM) are discarded and replaced by the mean stratification from 3D simulations. The distance from the patching point to the stellar centre is determined by interpolation from the structure of the UPM, based on the patching quantity.

As argued in Paper I, the temperature (T), density (ρ), total pressure (P), and combinations thereof are all suitable choices for the patching quantity. In theory Γ_1 is also a potential patching quantity. However, in some of the present cases it varies too slowly as function of the distance from the stellar centre at the depth at which the patch is performed. Therefore, in practice, it is not a suitable patching quantity, contrary to the solar case (cf. Paper I).

In this paper, we follow many of the steps outlined in Paper I. We match each UPM with a (3D) envelope of the same effective temperature, gravity, and metallicity. We note that other authors take the opposite approach, choosing the interior model in such a way that it matches the stratification of a given envelope at the patching point (e.g. Ball et al. 2016). This is due to the cost of computing 3D envelopes – without an interpolation algorithm they are limited to the existing 3D simulations. With the interpolation scheme presented in this paper and Paper I, it is tractable to reliably construct (3D) envelopes. This allows us to compute a stellar model with any interior structure that matches observations and then subsequently correct for the shortcomings of the outermost layers, by patching a suitable envelope. Following this approach permits us to compute PMs with any combination of T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$ within a wide region of the parameter space rather than being restricted to those cases, for which 3D simulations happen to exist.

Due to the omission of turbulent pressure and convective back-warming (cf. Trampedach et al. 2013, 2017), the envelopes predicted by stellar evolution codes are less extended than their 3D counterparts, i.e. the radius of the PMs is larger¹ than the radius of the corresponding UPMs. We define the radius of the PMs, as the distance from the stellar centre to the mesh point, at which $T = T_{\text{eff}}$. Matching the UPM with an envelope that has the same stellar parameters, we find the difference between the radius of the PMs and the radius of the UPMs to be of the order of < 0.45 per cent for all 315 models that fulfil our selection criteria in Section 4 – here, we use ρ as the patching quantity. Consequently, the effective temperature of the patched (3D) envelope is consistent with the Stefan–Boltzmann law within a few degrees. An alternative selection of the parameters of the patched envelopes is discussed by Jørgensen et al. (2017).

The gravitational acceleration of each patched envelope is adjusted as a function of depth, in order to correct for the fact that $\log g$ is assumed to be constant throughout the 3D simulations. As mentioned in Section 2, the extent of the envelope is small compared to the stellar radius, and consequently these corrections are small.

The patched (3D) envelopes are constructed by interpolation in the Stagger grid. If models with the same metallicity as the UPM

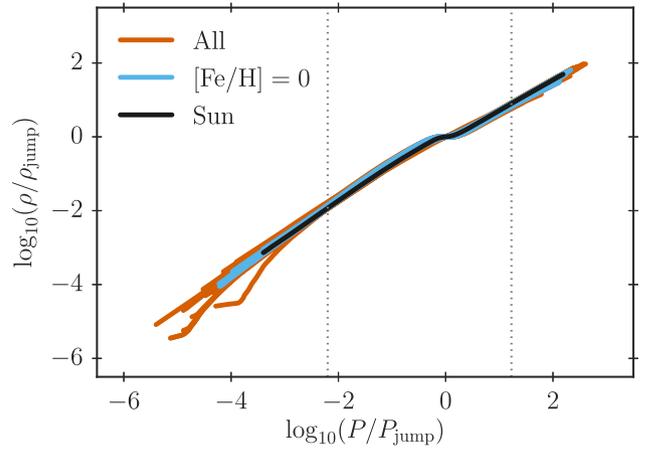


Figure 1. The logarithm of the scaled density as a function of the scaled pressure for 199 Stagger-grid envelope models. The scaling factors (ρ_{jump} and P_{jump}) correspond to the value at the density inversion (jump) near the surface. The vertical dotted lines indicate the employed interpolation range.

exist in the Stagger grid, we interpolate only in the (T_{eff} , $\log g$) plane. This interpolation scheme makes use of the fact that the structure of the employed 3D simulations show a high degree of homology: first, $\log_{10}(\rho)$, $\log_{10}(T)$, Γ_1 , and $\log_{10}(P)$ are scaled by the corresponding value at a local minimum in $\partial \log \rho / \partial P$, i.e. the density inversion near the surface. Secondly, we interpolate the obtained scaled generic stratifications as well as the associated scaling factors. For this purpose, we use the scaled pressure as our coordinate – that is, all quantities are determined for a common set of scaled pressures, requiring interpolation within the mesh of each simulation. Finally, the scaling is inverted. The depth of each mesh point is subsequently established, based on hydrostatic equilibrium.

If, on the other hand, no models with the same metallicity exist, we first construct seven envelopes with the desired T_{eff} and $\log g$ by interpolation in the (T_{eff} , $\log g$) plane. This is performed at each of the seven Stagger-grid metallicities: -4.0 , -3.0 , -2.0 , -1.0 , -0.5 , 0.0 , and 0.5 . For each of these envelopes, we then rescale $\log_{10}(\rho)$, $\log_{10}(T)$, Γ_1 , and $\log_{10}(P)$ by the corresponding value at the density inversion near the surface. Based on these scaled logarithmic stratifications, we evaluate $\log_{10}(\rho)$, $\log_{10}(T)$, and Γ_1 at pre-determined $\log_{10}(P)$ for an envelope with the desired metallicity. The interpolation in $[\text{Fe}/\text{H}]$ is performed using Akima splines (Akima 1970). Based on the scaling factors of the seven models with the same T_{eff} and $\log g$, we determine the scaling factor for the desired parameter values, likewise using Akima spline interpolation. The derivative of the pressure with respect to the density, however, is determined using a third-order polynomial spline. The scaling is then inverted. Fig. 1 shows the scaled density as function of the scaled pressure for all relaxed models in the grid.

When interpolating in $[\text{Fe}/\text{H}]$, the pressure at the deepest point of the interpolated envelopes is a factor of $10^{1.23}$ higher than the pressure at the density jump. This value for the pressure at the deepest point of the interpolated envelope is dictated by the shallowest envelope that enters the interpolation² (cf. Fig. 1). When interpolating in the (T_{eff} , $\log g$) plane only, the pressure at the bottom of the envelope may hence be higher. For each PM, we performed

¹Of course, e.g. large discontinuities in any quantity that enters the determination of the depth from hydrostatic equilibrium may render this statement incorrect.

²We discard the lowermost mesh points of the Stagger-grid envelopes, since we deem these mesh points un-physical: the superadiabatic temperature gradient increases with increasing depth.

Table 1. Interpolation errors in ρ , T , and Γ_1 for each metallicity. The errors listed in the columns denoted by ρ , T , and Γ_1 refer to the highest value within the envelopes: the first value of a given entry is the mean of the highest interpolation error within each envelope, the second is the corresponding median, and the third is the highest value amongst the errors of all envelopes. Consequently, the interpolation errors will be lower than the listed values throughout most of the interpolated structures. The subscript ‘b’ refers to the errors of the corresponding quantity at the base of the envelope.

[Fe/H]	ρ (per cent)	ρ_b (per cent)	T (per cent)	T_b (per cent)	Γ_1 (per cent)	$\Gamma_{1,b}$ (per cent)
−3.0	6.8, 9.3, 37.0	2.3, 3.7, 17.4	7.0, 7.7, 16.7	0.7, 1.1, 4.2	5.8, 6.8, 13.1	0.1, 0.3, 2.2
−2.0	10.0, 11.3, 40.1	6.6, 7.1, 28.3	6.1, 6.3, 12.1	1.0, 1.0, 4.7	3.9, 4.7, 15.4	0.1, 0.2, 0.8
−1.0	11.3, 11.1, 19.7	3.5, 3.3, 6.4	9.7, 9.1, 14.0	1.8, 1.6, 3.0	4.0, 4.0, 6.0	0.2, 0.3, 0.8
−0.5	8.5, 8.5, 12.1	3.6, 3.5, 5.8	7.0, 7.5, 13.0	1.3, 1.4, 2.0	2.7, 3.0, 4.7	0.1, 0.2, 0.5
0.0	4.5, 4.6, 7.6	0.8, 0.9, 2.2	5.7, 5.7, 8.5	0.8, 0.8, 1.4	2.9, 2.7, 3.7	0.1, 0.1, 0.3

the patch at the deepest point of the corresponding interpolated envelopes.

3.1 Testing the interpolation scheme

In order to assess the accuracy of our method, we cross-validate the interpolation scheme at distinct metallicities in the Stagger grid. We reconstruct 95 Stagger-grid envelopes at [Fe/H] = −3.0, −2.0, −1.0, −0.5 and 0.0. We employ a leave many-out strategy for each atmosphere, withholding all models at the target metallicity. As our splines rely on information on each side of the target metallicity, we are precluded from testing the metallicities at the edge of the grid, i.e. [Fe/H] = −4.0 and 0.5. These interpolated models are hence determined based on six interpolated models with the same T_{eff} and $\log g$ but different [Fe/H].

The interpolation error at the base of the envelope is of particular interest, as it affects the selection of the patching point. We determined the median, mean and the highest obtained interpolation error at the base for $\Gamma_1(P)$, $T(P)$, and $\rho(P)$, respectively, from our cross-validation exercise and summarize the results in Table 1. Alongside, we list the corresponding values for the errors obtained at any point within the envelopes, as the accuracy varies throughout the structure. Thus, even small displacements of the density inversion will lead to large residuals in the vicinity thereof, due to the large structural changes that take place near the density inversion. The relative interpolation errors are much smaller throughout most of the structures, including the base, than at the density inversion.

The errors at the base of the interpolated envelopes are illustrated in Fig. 2. In none of the 95 cases, do the interpolation error in $\rho(P)$ at the base exceed 28 per cent. As can be seen from Fig. 2, the interpolation error for the density at the base of the envelope is much lower than the 28 per cent throughout most of the Kiel diagram. Indeed, in Fig. 2, the interpolation scheme only performs so badly in the high- g low- T_{eff} corner of the grid: the mean and median of the errors shown in Fig. 2 are 3.9 per cent and 3.0 per cent, respectively.

Regarding the values stated above, we note that interpolation errors are expected to be smaller than in our cross-validation exercise, i.e. if all models are used. The interpolation errors that enter an analysis of PMs are hence lower than those cited in Table 1.

The highest interpolation errors are obtained for the lowest metallicities. For a fixed [Fe/H], the interpolated values of $T(P)$ and $\rho(P)$ at the base of the envelope are either systematically too high or too low throughout the $(T_{\text{eff}}, \log g)$ plane. Part of this inaccuracy can be attributed to our validation strategy – we withhold all models at a given metallicity. Nevertheless, our method would benefit greatly from an increase in the Stagger-grid resolution.

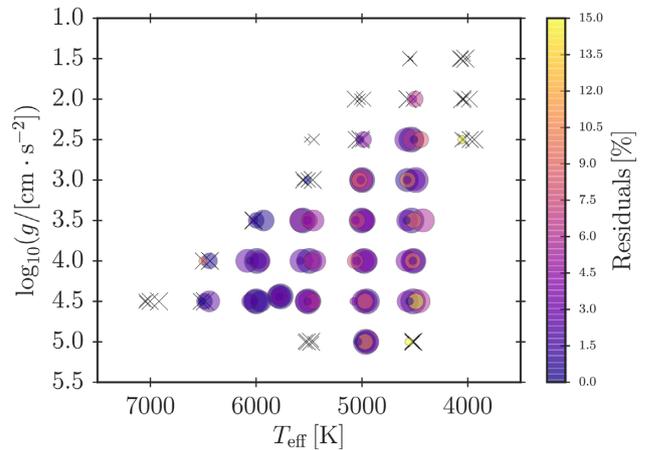


Figure 2. Residual of reconstruction of the density at the bottom of stellar envelopes by interpolation in metallicity. At this mesh point, the pressure is $10^{1.23}$ times higher than at the density inversion near the surface. The larger the marker size is, the larger the metallicity is. The crosses mark the envelopes that could not be reconstructed by interpolation, as they lie on the boundary of the Stagger grid. The colour bar shows the obtained errors in per cent. Errors higher than 15 per cent have all been attributed the same colour.

To test, whether the interpolation errors affect the predicted model frequencies significantly, we constructed patched solar models. The input physics of the underlying un-patched solar calibration model will be specified in Section 4.1. For this test, we computed three different envelopes: the original Stagger-grid envelope and two interpolated envelopes. One of the interpolated envelopes was obtained by interpolation in the $(T_{\text{eff}}, \log g)$ -plane only after excluding the solar envelope model in the grid. The other interpolated envelope has been computed by interpolation in T_{eff} , $\log g$, and [Fe/H] after excluding all models with solar metallicity. We have patched the envelopes to a solar calibration model, using different patching quantities, in order to show the effect of this choice. The resulting model frequency differences, relative to the observed ones, are shown in Fig. 3. Note that these differences reach $14 \mu\text{Hz}$ at a frequency of $4000 \mu\text{Hz}$ for the un-patched solar model. Here, we compare with observations rather than plotting the frequency difference between the UPM and the associated PM to facilitate an easy comparison between Fig. 3 and the results presented in e.g. Paper I, Ball et al. (2016), or Jørgensen et al. (2018).

The structural differences that give rise to the frequency differences in Fig. 3 are shown in Fig. 4. Interpolation after the exclusion of all models with solar metallicity leads to a larger misplacement of the density jump near the surface and lower accuracy in general

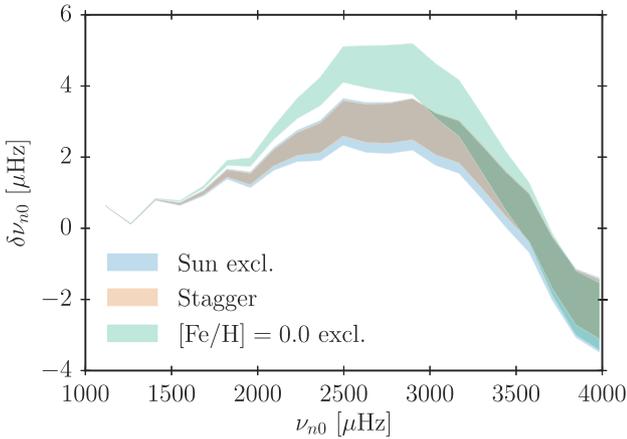


Figure 3. Frequency differences between BiSON observations (Broomhall et al. 2009; Davies et al. 2014) and three PMs. For one model (‘Stagger’), we have patched the Stagger-grid envelope with solar global parameters. One model (‘Sun excl.’) is based on an interpolated envelope that was computed after excluding the solar 3D envelope model from the grid. The last model (‘[Fe/H] = 0 excl.’) has been constructed by patching an interpolated envelope that was computed after excluding all 3D envelopes at solar metallicity from the grid. The width of the shaded area shows the frequency shift that is introduced, when changing the patching quantity: we use P , ρ , or T .

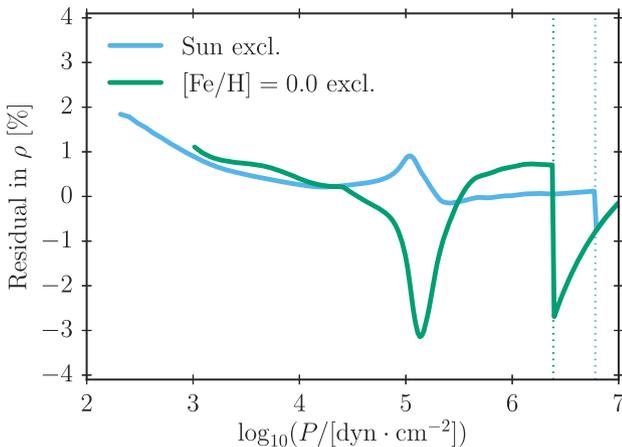


Figure 4. Difference between the density of PMs and the solar envelope in the Stagger grid as a function of the total pressure. The vertical dotted lines indicate the location of the patching point. The colours correspond to Fig. 3. The density has been used as the patching quantity.

than the interpolation in the $(T_{\text{eff}}, \log g)$ plane. Again, this calls for a refinement of the Stagger grid in $[\text{Fe}/\text{H}]$. Furthermore, as can be seen from the figure, the envelope that relies on an interpolation in $[\text{Fe}/\text{H}]$ is shallower (with a depth of 1.1 Mm) than the envelope obtained from interpolation in only T_{eff} and $\log g$ (with a depth of 1.5 Mm). This is because the Stagger-grid envelope with the lowest pressure at its base relative to the pressure at the density inversion dictates the pressure at the base of the interpolated envelopes as described above. Due to this, the envelopes employed in this paper are in general slightly shallower than those presented in Paper I: the interpolated envelopes span roughly four orders of magnitude in pressure (cf. Fig. 1).

4 THE SURFACE CORRECTION

We computed a grid of un-patched stellar structure models, using the Garching Stellar Evolution code (GARSTEC; Weiss & Schlattl 2008). For each UPM, we constructed a PM with the method just described. We then computed stellar oscillation frequencies for all UPMS and PMs in the grid, using the Aarhus adiabatic oscillation package (ADIPLS Christensen-Dalsgaard 2008).

As argued by Sonoi et al. (2015), we can use the frequency difference between the UPMS and the associated PMs to derive relations for the surface correction. We note, however, that the resulting relations cannot be used in the same manner as empirical surface correction relations (e.g. Kjeldsen et al. 2008; Ball & Gizon 2014) as we only parametrize the structural contribution.

As can be seen from the solar case illustrated in Fig. 3, the oscillation frequencies of the PMs are expected to deviate from the true stellar frequencies. The deviation is only of the order of a few microhertz – which is why Sonoi et al. (2015) use the adiabatic oscillation frequencies of PMs to approximate the surface effect. However, we note that this statement only holds true, since we use Γ_1 in the frequency calculations, omitting the corrections introduced to Γ_1 by the turbulent pressure: turbulent pressure alters the linearized expression that relates the relative gas pressure perturbation and the density perturbation (cf. Rosenthal et al. 1999; Houdek et al. 2017). This amounts to reducing Γ_1 by the ratio of the gas pressure to the total pressure, which leads to an additional shift in the predicted model frequencies (cf. Houdek et al. 2017; Jørgensen et al. 2018). Altering Γ_1 accordingly is known as the reduced Γ_1 approximation, whilst the use of the unaltered Γ_1 is known as the gas Γ_1 approximation. Hence, in accordance with e.g. Sonoi et al. (2015), Ball et al. (2016), Magic & Weiss (2016), and Paper I, we employ the gas Γ_1 approximation.

4.1 The stellar model grid

We computed evolutionary tracks of un-patched stellar models with masses between $0.65 M_{\odot}$ and $1.50 M_{\odot}$ in steps of $\Delta M = 0.05 M_{\odot}$ from the pre-main sequence (pre-MS³) to the red giant branch (RGB) up to a $\log g$ of at least 3.0. We use ten different initial compositions: these include the initial composition obtained from the underlying solar calibration⁴ as well as initial compositions that correspond to a metallicity, i.e. $[\text{Fe}/\text{H}]$, of between -0.3 and 0.5 in steps of 0.1. We restrict ourselves to this metallicity range, as it mostly covers the parameter space that is of interest in connection with *Kepler* targets in the LEGACY and KAGES samples.

For the sake of consistency between the UPMS and the patched Stagger-grid envelopes, we use the solar abundance AGSS09. We employ the corresponding OPAL opacities (Ferguson et al. 2005; Iglesias & Rogers 1996) and use the OPAL equation of state (EOS, Rogers & Nayfonov 2002), extending the EOS with the EOS by Hummer & Mihalas (1988) at low temperatures. Furthermore, we

³By which we mean the hydrostatic contraction before the zero age main sequence. Initial pre-MS models are constructed by integration of the stellar structure equations under the assumption of a constant mass–luminosity relation in the stellar interior, in order to obtain the necessary energy generation term. This assumption is consistent with a homologous contraction and is dropped as soon as the evolution proceeds. Pre-MS models typically start at luminosities between $10 L_{\odot}$ and $100 L_{\odot}$ and central temperatures around 10^5 K.

⁴More details on a calibration with similar input physics can be found in the paper by Jørgensen & Weiss (2018).

use the reaction rates suggested by Adelberger (2011) and mixing-length theory for convection (Böhm-Vitense 1958).

The solar calibration yields a mixing length (α_{MLT}) of 1.79. We use this value throughout the grid of stellar models. This is a simplifying assumption: it has been established by independent means that α_{MLT} is not constant with regard to any of the grid parameters (Trampedach et al. 2014; Magic, Weiss & Asplund 2015; Tayar et al. 2017). However, these studies are partly contradictory, i.e. no consensus has been reached on how to correctly vary the mixing length. This being said, it appears that α_{MLT} varies only moderately in the region of the parameter space that is of interest in our analysis (cf. figs 4 and 5 in Trampedach et al. 2014 and fig. 9 in Tayar et al. 2017). Thus, for our purposes, a constant mixing length is a reasonable approximation. Also, although the assumption of a calibrated mixing length is a simplifying approximation, it is commonly used. Adopting this assumption hence allows a point of comparison between modellers. Furthermore, the main objective of this paper is to demonstrate the usefulness and applicability of the interpolation scheme. This is made all that more tractable by reducing the number of dimensions in our modelling.

In connection with our choice of α_{MLT} , it is worth mentioning that the structures of PMs show discontinuities at the patching point. This can be seen from the jumps in the residuals in Fig. 4: the density is used as the patching quantity, so whilst $\rho(r)$ is by construction a smooth function at the patching point, $P(r)$ is not. We also illustrate this in connection with our analysis of *Kepler* stars in Section 5.5. These discontinuities result in a sensitivity of the oscillation frequencies to the patching quantity (cf. Fig. 3) and reflect inconsistencies between 1D stellar models and 3D envelopes – the neglect of turbulent pressure or convective back-warming in 1D models are examples of such inconsistencies. For a detailed discussion of this issue, we refer to Paper I. Since α_{MLT} dictates the adiabat of stellar models, these discontinuities may partly be mended by calibrating the mixing length anew for each PM (cf. Ball et al. 2016). However, ensuring smooth stratifications does not guarantee a physically meaningful structure but may rather mask the mentioned inconsistencies. Also, with our method, we aim to correct the outermost layers of stellar models with the desired interior structure rather than to adjust the interior structure to fit the envelope. Instead of artificially ensuring a smooth transition at the patching point, we therefore introduce a set of selection criteria to discard structure models that we do not deem physically reasonable. These criteria are discussed towards the end of this subsection.

The initial mass fraction of helium ($Y_{\odot,i}$) and heavy elements ($Z_{\odot,i}$) obtained from this calibration is 0.2653 and 0.0149, respectively. The corresponding model of the present Sun was used in Section 3, when testing the effect of interpolation on the frequency computation (cf. Fig. 3).

For bulk compositions different from the Sun, we compute the initial helium mass fraction, using the following linear relation:

$$Y = Y_{\text{BBN}} + \frac{Y_{\odot,i} - Y_{\text{BBN}}}{Z_{\odot,i}} Z, \quad (1)$$

employing the solar calibration values. Y_{BBN} refers to the helium abundance resulting from the big bang nucleosynthesis (BBN; Cyburt et al. 2016). We have set $Y_{\text{BBN}} = 0.245$ (cf. Cassisi, Salaris & Irwin 2003). In contrast, the mass fraction of helium relative to hydrogen is kept fixed in the Stagger-grid simulations. Specifically,

$$A(\text{He}) = \log_{10} \left(\frac{n_{\text{He}}}{n_{\text{H}}} \right) + 12 \quad (2)$$

is kept constant. Here, n_{H} and n_{He} denote the number density of hydrogen and helium, respectively. The relative abundances of the metals correspond to AGSS09 but are scaled to the abundance of Fe (Collet, private communications). As a consequence, the helium abundance in the Stagger-grid envelopes decreases with increasing metallicity, which puts the composition of the 3D simulations somewhat at odds with considerations regarding stellar evolution.

Thus, whilst our initial helium mass fraction is reasonable from an evolutionary perspective, it constitutes an inconsistency between the 1D models and 3D envelopes. As we aim towards modelling *Kepler* stars in Section 5.5, we prefer to include all relevant input physics in order to obtain realistic stellar evolution models rather than artificially mimicking the composition of the 3D simulations. The purpose of the following sections is merely to illustrate the applicability and usefulness of our interpolation scheme. Whilst a different set of choices may likewise constitute an interesting test case, we hence restrict ourselves to cases that are relevant for the analysis of *Kepler* stars, ensuring a clean narrative.

From the evolutionary tracks, we selected 657 un-patched structure models. All GARSTEC models include microscopic diffusion of H, He, Li, C, N, O, Ne, Mg, Si, and Fe. Hence, the surface composition changes during the evolution. We note that the incorporation of the microscopic diffusion of metals may lead to an artificially high depletion of metals, especially for low-metallicity models with masses higher than the mass of the Sun. In order to take this into account, we only selected models, for which $[\text{Fe}/\text{H}] \geq -0.5$, hereby excluding any model, for which the abundance of heavy elements were vastly depleted. This issue could have been avoided by ignoring microscopic diffusion, when constructing the grid. However, as stated above, we prefer to include all relevant input physics.

As mentioned in Section 2, the composition of the Stagger-grid envelopes takes α -enhancement into account for $[\text{Fe}/\text{H}] \leq -1$. We restrict the grid to metallicities, where no α -enhancement is needed, in order for the bulk composition to be trivially related to the metallicity:

$$[\text{Fe}/\text{H}] = \log_{10} \left(\frac{Z_{\text{S}}}{X_{\text{S}}} \right) - \log_{10} \left(\frac{Z_{\odot,\text{S}}}{X_{\odot,\text{S}}} \right). \quad (3)$$

Here, $Z_{\odot,\text{S}}/X_{\odot,\text{S}} = 0.01792$, where Z_{S} and X_{S} denote the mass fraction of metals and hydrogen at the surface, respectively.

For each of the 657 UPMs, we constructed PMs, using the density as the patching quantity. As discussed in Paper I, the frequencies are relatively insensitive to the patching depth, if the patch is performed sufficiently deep within the nearly adiabatic region. In order to ensure that the patched envelope is indeed deep enough, we impose the criterion that the patching point must be placed deeper than the minimum in Γ_1 near the surface. This criterion leads to the omission of 25 models in the high- g , low- T_{eff} corner of the Kiel diagram – that is, main-sequence (MS) stars with $\log g = 4.5$ and $T_{\text{eff}} < 5300$ K.

We excluded 47 additional models, for which the radius of the UPM exceeds the radius of the associated PM: 3D envelopes are more extended than their 1D counterparts. If the radius of the PM is smaller than the radius of the UPM, this must be associated with a large discontinuity in either P or ρ . For all these PMs, $4.0 < \log g < 4.3$.

As shown in Paper I, the radius, at which the patching point is placed, is sensitive to the choice of the patching quantity. Hence, the patching quantity affects the model frequencies as shown in Fig. 3. Since we investigate the structural surface effect, relying therefore on the model frequencies, this ambiguity caused by the choice of the patching quantity is undesirable. To take this into account, we

constructed two additional PMs for each of the 585 UPMs that have so far successfully passed our selection criteria, using P and T as the patching quantity, respectively. We then excluded 270 models, for which the choice between P , T , and ρ as the patching quantity affects the model frequencies by more than $2 \times 10^{-3} \nu_{\max}$, unless the deviation fell below $1.0 \mu\text{Hz}$. For the Sun, $2 \times 10^{-3} \nu_{\max}$ would roughly correspond to twice the highest obtained deviation. The models that are hereby discarded are scattered across the Kiel diagram.

All in all, our selection criteria completely deplete certain regions of the Kiel diagram: all models with $T_{\text{eff}} \gtrsim 6000 \text{ K}$ were discarded as were most cold MS stars with $\log g \approx 4.3$. Meanwhile, no PM with $\log g < 3.15$ was affected by any of our selection criteria.

It is worth to take a closer look at the models that fail to pass the last selection criterion: for some of these models, the patching procedure did not alter the frequencies significantly, which may suggest that the patched envelopes are too shallow in these cases. Our analysis thus calls for deeper 3D simulations – ideally, new simulations should extend into the adiabatic layers. Furthermore, all 47 models with $R_{\text{PM}} < R_{\text{UPM}}$ would have fallen into this category if not previously excluded.

Alternatively, one may require a patching point deeper within the adiabatic region by excluding all models, for which $|\nabla - \nabla_{\text{ad}}|$ exceeds a certain value. Also, one may exclude all models, for which the choice between P , T , and ρ as the patching quantity affects the radius of the patching point by more than a certain fraction of the patching depth. In the case of the Sun, this choice affects the radius of the patching point by 21–27 km, corresponding to roughly 2–3 per cent of the patching depth. However, each additional selection criterion introduces a selection bias: some regions of the Kiel diagram are more harshly depleted than others. As illustrated in Section 4.3, such biases will leave their mark on the parametrization of the structural surface effect. We therefore avoid to introduce further selection criteria. At the end, 315 models passed our selection criteria. Due to microscopic diffusion, the surface metallicity of these models range from -0.49 to 0.45 .

4.2 The surface correction relation

In this paper, we present a parametrization of the structural contribution to the surface effect. We derive surface correction relations based on the frequency difference,

$$\delta\nu = \nu_{\text{PM}} - \nu_{\text{UPM}}, \quad (4)$$

between the model frequencies of the UPM (ν_{UPM}) and the model frequencies of the associated PM (ν_{PM}). As we do not consider modal effects, there is no reason to believe that PMs encapsulate the full surface term. We in fact see this in solar case in Fig. 3, where we do not capture all the features of $\delta\nu$. We investigate how well our parametrization recovers the determination of the stellar parameters in Section 5 in a comparison with Ball & Gizon (2014).

We use the same general functional forms for the surface correction relations as Sonoi et al. (2015). This includes a power-law fit to the frequency correction:

$$\frac{\delta\nu}{\nu_{\max}} = a \left(\frac{\nu_{\text{PM}}}{\nu_{\max}} \right)^b. \quad (5)$$

We alternatively try a Lorentzian fit of the form

$$\frac{\delta\nu}{\nu_{\max}} = \alpha \left(1 - \frac{1}{1 + (\nu_{\text{PM}}/\nu_{\max})^\beta} \right). \quad (6)$$

In the above relations, ν_{\max} denotes the frequency of maximum power. We adopt (Brown 1991)

$$\nu_{\max} = \frac{g}{g_{\odot}} \left(\frac{T_{\text{eff}}}{T_{\text{eff}\odot}} \right)^{-1/2} \nu_{\max\odot}, \quad (7)$$

where g and T_{eff} denote gravitational acceleration and the effective temperature of the UPM, respectively. In this paper, we set $\nu_{\max\odot} = 3090 \mu\text{Hz}$, $T_{\text{eff}\odot} = 5779.57 \text{ K}$, and $g_{\odot} = 4.438$ – that is $R_{\odot} = 6.95508 \times 10^8 \text{ cm}$, $M_{\odot} = 1.9891 \times 10^{33} \text{ g}$. Finally, a , b , α , and β are fitting parameters.

Equation (5) only yields a reasonable description of the frequency difference at frequencies that are lower than or close to ν_{\max} (Sonoi et al. 2015). Whilst all frequencies are included, when determining the fitting parameters in connection with Equation (6), we only include frequencies that are lower than $1.05 \nu_{\max}$, when using equation (5), in accordance with Sonoi et al. (2015). In either case, we perform the fit, using only radial modes ($\ell = 0$), in accordance with Kjeldsen et al. (2008) and Sonoi et al. (2015). We use all radial orders in the relevant frequency range.

Both equations (5) and (6) have two fitting parameters. These are assumed to be functions of the global stellar parameters. For all four fitting parameters, we adopt power-law dependences that include four free parameters, e.g.

$$a = A \left(\frac{T_{\text{eff}}}{T_{\text{eff}\odot}} \right)^B \left(\frac{g}{g_{\odot}} \right)^C \left(\frac{Z_{\text{S}}/X_{\text{S}}}{Z_{\odot,\text{S}}/X_{\odot,\text{S}}} \right)^D. \quad (8)$$

Again, all values are taken from the respective UPM. Equation (8) only deviates from the assumed functional dependence adopted by Sonoi et al. (2015) by the dependence on $Z_{\text{S}}/X_{\text{S}}$, i.e. on $[\text{Fe}/\text{H}]$, since we consider models with non-solar metallicities.

We note that Kjeldsen et al. (2008) have suggested a power-law correction, i.e. a relation on the form of equation (5), using a different definition of the coefficients, a and b . We will state this definition in Section 5, where we include a comparison to Kjeldsen et al. (2008). In this section, however, all coefficients in equations (5) and (6) – that is, a , b , α , and β – are defined equivalently to a in equation (8).

It follows from the above that a fit of $\delta\nu$ to either equation (5) or (6) in a parameter space spanned by T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, and ν/ν_{\max} includes eight free parameters. In this paper, we perform global fits to the entire grid presented in Section 4.1, constraining all eight parameters simultaneously. The advantage of performing such a global fit is that this approach allows for a robust determination of the uncertainties on the predicted frequency corrections based on the variance-covariance matrix (\mathbf{C}) obtained from the fit: for a fixed scaled frequency ν_k/ν_{\max} the associated frequency shift ($\delta\nu_k/\nu_{\max}$) is a function f of eight free parameters (ξ_1, \dots, ξ_8) given by either equation (5) or (6) in combination with equations for a , b , α , and β of the form of equation (8), which provides the Jacobian \mathbf{J} . The variance (σ_k^2) of $\delta\nu_k/\nu_{\max}$ is hence

$$\sigma_k^2 = \mathbf{J}\mathbf{C}\mathbf{J}^T, \quad \mathbf{J} = [\partial f/\partial\xi_1, \dots, \partial f/\partial\xi_8]. \quad (9)$$

The results of our global fit including all 315 PMs and their associated UPMs are summarized in Table 2. How well the Lorentzian fit reproduces the correct frequency shift throughout the Kiel diagram is furthermore illustrated in Fig. 5.

As can be seen from Fig. 5, the fit does not perform equally well throughout the parameter space. Whilst the varying accuracy of an imperfect fit is not surprising, it is worth taking a closer look at this, since it suggests that the surface corrections by Kjeldsen et al. (2008) and Sonoi et al. (2015) are subject to a selection bias: after all, the relation by Kjeldsen et al. (2008) is calibrated based on the

Table 2. Fitting parameters based on a global fit to all 315 PMs that fulfilled our selection criteria. Each fitting parameter (a , b , α , and β) in equations (5) and (6) is itself composed of four fitting parameters (A , B , C , and D) as given in equation (8). The uncertainties denote the square root of the diagonal of the variance–covariance matrix. We have performed the fit, using only radial modes ($\ell = 0$).

	A	B	C	D
a	$(-2.259 \pm 0.050) \times 10^{-3}$	2.808 ± 0.180	$(-7.189 \pm 0.089) \times 10^{-1}$	$(2.513 \pm 0.127) \times 10^{-1}$
b	4.517 ± 0.175	-1.290 ± 0.308	$(2.020 \pm 0.151) \times 10^{-1}$	$(-1.504 \pm 0.214) \times 10^{-1}$
α	$(-3.949 \pm 0.035) \times 10^{-3}$	2.659 ± 0.071	$(-6.888 \pm 0.035) \times 10^{-1}$	$(2.633 \pm 0.052) \times 10^{-1}$
β	5.530 ± 0.153	-1.445 ± 0.217	$(2.271 \pm 0.106) \times 10^{-1}$	$(-1.334 \pm 0.153) \times 10^{-1}$

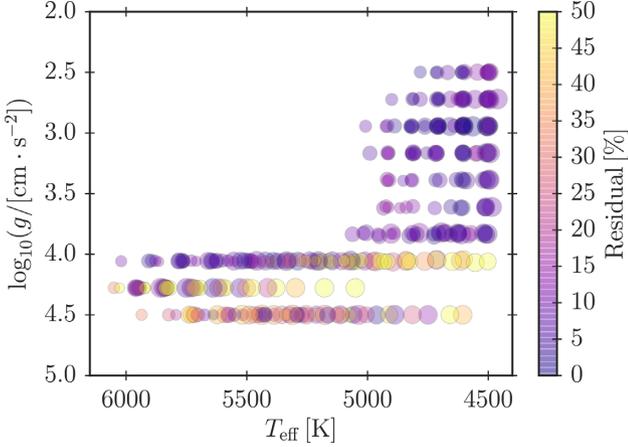


Figure 5. Residual of reconstruction of $\delta\nu$ relative to the actual value of $\delta\nu$ for the Lorentzian fit, i.e. equation (6), summarized in Table 2. The residuals are computed at $1.3\nu_{\max}$ – for the Lorentzian fit, this lies within the covered frequency range. Residuals that exceed 50 per cent are indicated by the same colour: the highest obtained error is 317 per cent. The larger the marker size is, the larger the metallicity is.

Sun only, and the relation by Sonoi et al. (2015) is based on 10 PMs, for which T_{eff} is predominantly larger than 6000 K.

If the fit is performed based on a subsample, excluding further models from the grid, we obtain different fitting parameters. Concretely, if we exclude, say, all RGB stars before evaluating the fitting parameters, we obtain a fit that performs better for MS stars and worse for RGB stars. The ramifications of this sensitivity to the subsample selection are illustrated in Section 4.3, where we reconstruct the present Sun.

4.3 Reconstructing the Sun

As a validation for our global fit, we determine how well the parametrization mimics the frequency difference between the UPM and the associated PM for the present Sun. The solar calibration model of the present Sun, based on which we computed the evolution paths that enter the grid, is itself not in the sample of structure models used to evaluate the fitting parameters. This makes it a suitable test case.

We have used the fit summarized in Table 2 to predict the frequency difference, employing the solar effective temperature, metallicity and gravitational acceleration. A comparison to the actual frequency difference between the UPM and the PM is shown in Fig. 6. Both fits lead to residuals of $\lesssim 2\ \mu\text{Hz}$ within the relevant frequency ranges: the power-law fit is expected to work well up to ν_{\max} , whilst the Lorentzian fit performs well up to the acoustic cut-off frequency. Thus, both fits perform rather well, encapsulating

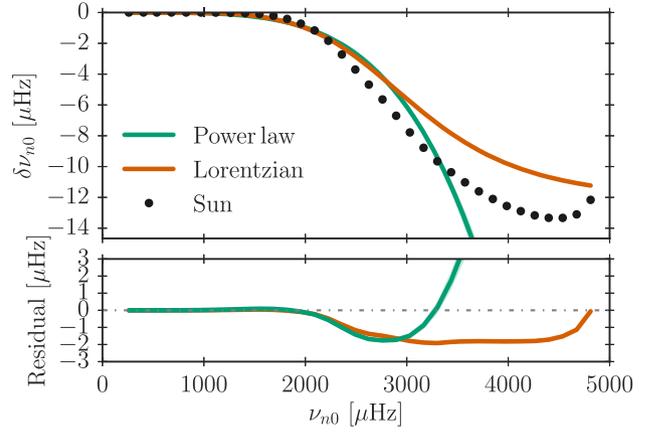


Figure 6. Reproduction of the frequency difference between the solar calibration model and the associated PM, using the power-law and Lorentzian fit, i.e. equations (5) and (6), respectively. The fits have been evaluated based on all successful PMs in the grid (cf. Table 2). The actual frequency difference between the UPM and the PM are labelled ‘Sun’ and marked with circles. The power-law fit is based on low frequencies only.

most of the structural surface effect – however, we note that the residuals are still much larger than the typical uncertainties of the observed frequencies.

As shown by Sonoi et al. (2015), a Lorentzian fit that is derived directly from the solar model yields smaller residuals than the global fit presented in Fig. 6. Based on this notion combined with Fig. 5, one may suspect a fit to a sample of solar-like dwarfs to perform better, when inferring the solar surface effect. Indeed, when recomputing the fit based on different subsamples of the grid and subsequently examining the solar case, a more detailed picture emerges: the performance of the fit depends on the subsample. This is illustrated in Figs 7 and 8.

Besides including all 315 models in the fitting procedure, we have looked at the following subsamples:

- (i) 45 models for which $5500\ \text{K} < T_{\text{eff}} < 5800\ \text{K}$ and $\log g > 4.0$. There are no restrictions on $[\text{Fe}/\text{H}]$. We refer to this sample as *solar-like stars*.
- (ii) 163 models for which $\log g > 4.0$. There are no restrictions T_{eff} on $[\text{Fe}/\text{H}]$. We refer to this sample as *MS stars*.
- (iii) 56 models for which $-0.05 < [\text{Fe}/\text{H}] < 0.05$. There are no restrictions T_{eff} on $\log g$.
- (iv) 152 models for which $\log g < 4.0$. There are no restrictions T_{eff} on $[\text{Fe}/\text{H}]$. We refer to this sample as *RGB stars*.

Figs 7 and 8 include the uncertainties on the predicted frequency difference inferred from the fitting procedure, showing that the different fits are incompatible – in some cases, the line width is broader than the inferred uncertainties. According to Figs 7

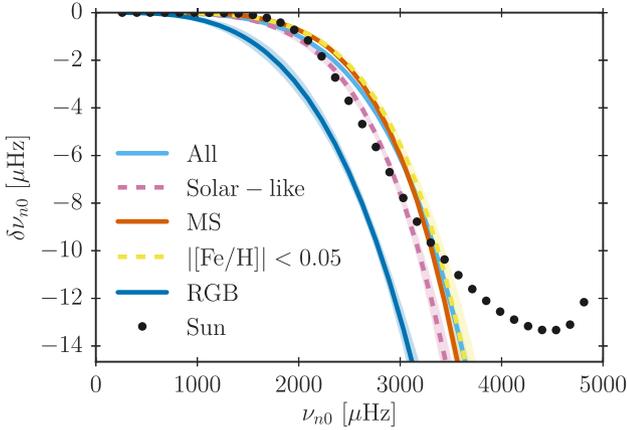


Figure 7. Reproduction of the frequency difference between the solar calibration model and the associated PM, using power-law fits to different subsamples of the grid. Details are specified in the text and in the caption of Fig. 6.

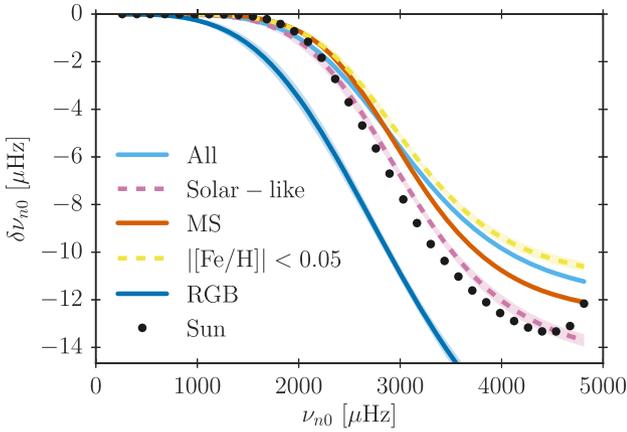


Figure 8. Same as Fig. 7 but for the Lorentzian surface correction relation.

and 8, the best fit is obtained, when the models in the sample close match the target: the residuals become significantly lower, when we restrict ourselves to the 45 ‘solar-like’ stars. The fact that different subsamples lead to different reconstructions of the solar case implies that a single scaling relation that applies to all evolutionary stages is not realizable with the presented approach. Also, it implies that previous studies, such as the inference of a surface correction relation by Sonoi et al. (2015), is subject to a selection bias: in the mentioned case, the authors have selected 10 PMs predominantly with $T_{\text{eff}} > 6000$ K.

For the computation of the fitting coefficients, we have used interpolated (3D) envelopes, since additional PMs add information to the analysis due to the unique interior structure of the associated UPMs. Had we used only the original Stagger-grid simulations, we would have been limited by the small sample size. The lower number of PMs would impose restrictions on the subsample selection and affect the robustness of the inferred fitting coefficients. Furthermore, the distribution of simulations within the Stagger grid would introduce a selection bias. Interpolation gives us control over this selection bias.

5 BEST-FITTING MODELS

Below, we take a closer look at the stellar parameters of two *Kepler* stars, by building on the analysis presented above. As shown in Section 4, the Lorentzian parametrization reliably encodes the structural surface effect, if the fit is based on models, whose global parameters resemble those of the target. Rather than computing further PMs, we hence use this parametrization.

This approach has the advantage that other authors may easily reproduce and implement our results. Furthermore, using this approach, we can investigate the limitations of such parametrizations. For instance, in Section 4, we found that the performance of the Lorentzian fit depends on the subsample, upon which the fit is determined. Using our parametrization allows us to discuss what ramifications this has for the stellar parameter estimates, based on a comparison with Sonoi et al. (2015). This is of interest, since surface correction relations are commonly used.

5.1 Maximum likelihood estimation

Having established a parametrization of the structural surface effect and the associated constraints on the applicability hereof, we can use this parametrization to evaluate stellar parameters based on a grid search, by comparing the corrected model frequencies with observations. We do this by using MLE. For this purpose, we construct a denser grid of un-patched stellar models and attribute a likelihood,

$$\mathcal{L} = \mathcal{L}_{\text{seis}} \mathcal{L}_{\text{spec}}, \quad (10)$$

to each model in the grid. The details of the grid are summarized in Section 5.2. In equation (10), $\mathcal{L}_{\text{seis}}$ denotes the seismic contribution to the likelihood. For simplicity, we assume that the model frequencies are uncorrelated and that the uncertainties are Gaussian. Whilst we note that both assumption are only approximately valid, we also note that at least the latter is commonly used (e.g. Silva Aguirre et al. 2015; Nsamba et al. 2018). The seismic contribution to the likelihood is hence

$$\mathcal{L}_{\text{seis}} = \prod_i^N \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(v_{\text{corr},i} - v_{\text{obs},i})^2}{2\sigma_i^2}\right). \quad (11)$$

Here, $v_{\text{corr},i}$ is the i th corrected model frequency of the UPM, for which we determine the likelihood. In the following, we use equation (6), i.e. our Lorentzian correction, employing fitting parameters that were determined based on an appropriate subregion of the Kiel diagram. The subregion was determined using the grid presented in Section 4 – that is, the models from Section 4 that enter the fit span the same region of the parameter space as the denser grid presented in Section 5.2 does. We elaborated upon this in Section 5.2.

In order to be consistent with the derivation of the frequency correction relation, we only include radial frequencies ($\ell = 0$) in our analysis. For comparison, we additionally repeat the analysis using the surface correction relation by Kjeldsen et al. (2008), Ball & Gizon (2014), and Sonoi et al. (2015).

In equation (11), each of the N observed frequencies ($v_{\text{obs},i}$) is compared to the corresponding corrected model frequency – the number N is determined by the number of observed radial modes and will hence be fixed for a given star. We allow for the radial order of the model frequencies to deviate from observations by up to 2, since the radial orders of the observed frequencies may be incorrectly assigned. We choose the attribution of the radial order that leads to the lowest frequency deviation for the lowest frequency.

Finally, in equation (11), σ_i denotes the uncertainty of the frequency difference. We note that this is not simply the uncertainty of the observed frequency, since the corrected frequencies are themselves a function of the observed frequencies. In order to take this into account, we use equation (9). As we assume the observations to be uncorrelated and the associated uncertainties to be Gaussian, equation (9) reduces to the absolute value of the derivative of $\nu_{\text{obs},i} - \nu_{\text{corr},i}$ with respect to $\nu_{\text{obs},i}$ times the uncertainty on $\nu_{\text{obs},i}$. We consistently ignore the uncertainties on the fitting parameters here, since these are not available for all surface corrections.

When computing the likelihood, we also take spectroscopic constraints on the effective temperature and the metallicity into account. Again, we assume the uncertainties to be Gaussian:

$$\mathcal{L}_{\text{spec}} = \frac{1}{2\pi\sigma_T\sigma_{[\text{Fe}/\text{H}]}} \exp\left(-\frac{\Delta T_{\text{eff}}^2}{2\sigma_T^2} - \frac{\Delta[\text{Fe}/\text{H}]^2}{2\sigma_{[\text{Fe}/\text{H}]}^2}\right). \quad (12)$$

Here, σ_T and $\sigma_{[\text{Fe}/\text{H}]}$ denote the standard deviation of the spectroscopically deduced effective temperature and metallicity, respectively. ΔT_{eff} and $\Delta[\text{Fe}/\text{H}]$ denote the difference between the observationally constrained value and the model prediction for each of the two quantities.

Out of convenience, we work with the natural logarithm of the likelihood, when performing calculations. More specifically, we work with the averaged log-likelihood:

$$\langle \ln \mathcal{L} \rangle = \frac{1}{N+2} \ln \mathcal{L}, \quad (13)$$

for our N frequencies and two spectroscopic observables. The stellar model that maximizes the quantity in equation (13) is considered the best-fitting model in the grid.

In order to ascribe a measure for the goodness of fit, we provide χ^2 of the best-fitting model:

$$\chi^2 = \frac{1}{N+2} \left(\sum_i^N \frac{(\nu_{\text{corr},i} - \nu_{\text{obs},i})^2}{\sigma_i^2} + \frac{\Delta T_{\text{eff}}^2}{\sigma_T^2} + \frac{\Delta[\text{Fe}/\text{H}]^2}{\sigma_{[\text{Fe}/\text{H}]}^2} \right). \quad (14)$$

The parameter values of the best-fitting model do not on their own encode comprehensive information about the likelihood function. We hence likewise provide the weighted mean and the weighted variance of each parameter: having K stellar models in the grid, the weighted mean of a parameter β is

$$\langle \beta \rangle = \frac{\sum_k^K \beta_k \langle \mathcal{L}_k \rangle}{\sum_k^K \langle \mathcal{L}_k \rangle}. \quad (15)$$

Here, $\langle \mathcal{L}_k \rangle$ is the exponential of the quantity in equation (13) for the k th model. Equivalently, the weighted variance is

$$(\sigma_\beta^{\text{mod}})^2 = \frac{\sum_k^K (\beta_k - \langle \beta \rangle)^2 \langle \mathcal{L}_k \rangle}{\sum_k^K \langle \mathcal{L}_k \rangle}. \quad (16)$$

A weighted mean that deviates from the parameter value of the best-fitting model and a large variance imply a likelihood function that deviates from a single narrow peak.

5.2 The stellar model grid

Whilst a few hundred models are sufficient for the purpose of robustly establishing the parametrizations in Section 4, a higher sampling of the parameter space is needed, when evaluating stellar parameters based on a grid search. Using the same input physics as in Section 4.1, we therefore computed a new grid with a higher

sampling rate: we computed 4007 stellar evolution tracks with masses between $0.8 M_\odot$ and $1.3 M_\odot$ in steps of $0.01 M_\odot$. This includes tracks with the same composition as the underlying solar calibration as well as initial compositions on the pre-MS that correspond to a metallicity between $[\text{Fe}/\text{H}] = -0.20$ and 0.55 in steps of 0.01 . Due to microscopic diffusion, the metallicity decrease with age and can hence be lower than the listed initial values. We note that the high-mass stars will be more efficient at draining metals from their photosphere, which implies that the sampling in metallicity depends on the other global parameters: as discussed in Section 4.1, the depletion maybe artificially high at higher masses. For our current analysis, however, this is of no concern, since the Sun and the investigated stars all have masses below $1.2 M_\odot$.

From this set of evolution tracks, we selected a dense grid of stellar structure models. Since we have already dealt with the solar case, we investigate a different region of the Kiel diagram. Applying the same selection criteria as listed in Section 4.1 but demanding at least two model frequencies to be evaluated above $500 \mu\text{Hz}$, we selected 150 719 UPMS with effective temperatures between 5255 and 6120 K, $\log g$ between 3.78 and 4.30 , and metallicities between -0.42 and 0.47 . Due to the constraints set by theoretical evolution tracks, the models are irregularly distributed in the parameter space spanned by T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$. Radial frequencies below $500 \mu\text{Hz}$ are ignored.

We have derived the coefficients for our Lorentzian surface correction relation based on an equivalent region of the Kiel diagram: the surface correction is hence determined using 73 models from the grid presented in Section 4.1, for which T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$ lie in the intervals $[5257, 6050$ K], $[4.05, 4.28]$, and $[-0.49, 0.38]$, respectively. The coefficients are summarized in Table 3.

5.3 Comparisons

We evaluate the best-fitting models in our grid using the MLE procedure based on a Lorentzian fit to PMs in the relevant region of the Kiel diagram as described above. For comparison we repeat the evaluation of the best-fitting stellar parameters, employing surface correction relations by other authors. We briefly elaborate on these below.

Kjeldsen et al. (2008) suggest equation (5), i.e. a fit to a power law, as the functional form of the surface correction, setting $b = 4.90$ and

$$a = \frac{\langle \nu_{\text{obs},i} \rangle - s \langle \nu_i \rangle}{N^{-1} \sum_{i=1}^N [\nu_{\text{obs},i} / \nu_{\text{max}}]}. \quad (17)$$

Here, $s = 1$, $\nu_{\text{obs},i}$ denotes the i th observed frequency and ν_i denotes the corresponding model frequency. In the following, we will refer to the surface correction relation by Kjeldsen et al. (2008) as K08.

Sonoi et al. (2015) suggest to use a surface correction with functional form given by equation (6), i.e. a Lorentzian fit. The parameters α and β are based on 10 PMs and do not include a dependence on metallicity:

$$\log |\alpha| = -7.69 \log T_{\text{eff}} - 0.629 \log g - 28.5, \quad (18)$$

and α is negative, whilst

$$\log \beta = -3.86 \log T_{\text{eff}} + 0.235 \log g + 14.2. \quad (19)$$

We will refer to this surface correction as S15. A comparison between the results obtained, when using our own surface correction relation, and those obtained, when using S15, in principle allows for a discussion on the importance of the inclusion of metallicity. However, different results may mainly reflect the selection bias.

Table 3. Pendant to Table 2, containing fitting parameters based on a global fit to 73 PMs in our target region.

	A	B	C	D
α	$(-1.644 \pm 0.029) \times 10^{-3}$	8.507 ± 0.181	-1.825 ± 0.025	$(8.154 \pm 0.1132) \times 10^{-1}$
β	$(1.839 \pm 0.129)^1$	-7.056 ± 0.665	1.616 ± 0.094	$(-7.966 \pm 0.432) \times 10^{-1}$

Rather than computing coefficients for an assumed functional form using PMs or solar observations, Ball & Gizon (2014) present a surface correction relation that has been deduced from the variational principle by Gough (1990) based on physical considerations:

$$\frac{\delta v_i}{v_{ac}} = \mathcal{I}^{-1} \left(a_{-1} \frac{v_{ac}}{v} + a_3 \frac{v^3}{v_{ac}^3} \right). \quad (20)$$

Here, \mathcal{I} is the mode inertia (cf. Christensen-Dalsgaard 2008), whilst v_{ac} denotes the acoustic cut-off frequency:

$$\frac{v_{ac}}{v_{ac\odot}} = \frac{v_{max}}{v_{max\odot}}, \quad (21)$$

where we set $v_{ac\odot} = 5000 \mu\text{Hz}$ in accordance with Ball & Gizon (2014). We will refer to this surface correction relation as B&G.

The coefficients a_{-1} and a_3 are given by

$$\begin{pmatrix} a_{-1} \\ a_3 \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad (22)$$

where \mathbf{y} denotes the frequency difference between the observed frequencies and the model frequencies in units of the uncertainty of the observed frequencies (σ_i^{obs})

$$y_i = \frac{v_{\text{obs},i} - v_i}{\sigma_i^{\text{obs}}} \quad (23)$$

and the matrix \mathbf{X} is a combination of model frequencies and the associated mode inertias

$$X_{i,1} = \frac{v_{ac}^2 v_i^{-1}}{\mathcal{I}_i \sigma_i^{\text{obs}}}, \quad X_{i,2} = \frac{v_{ac}^{-2} v_i^3}{\mathcal{I}_i \sigma_i^{\text{obs}}}. \quad (24)$$

Our surface correction relation was inferred from PMs without taking modal effects into account. Assuming that the asymptotic analysis that underlies B&G includes all relevant contributions, equation (20) is not subject to this inadequacy. Hence, a comparison between the results obtained when using our surface correction relation, and those obtained when using equation (20), can be used to determine how the neglect of modal effects alters the determination of model parameters. To a certain extent, this also holds true for the comparison with K08.

Since both K08 and B&G rely on the observed frequencies in order to determine the surface correction, these two surface correction relations have limited predictive power in comparison to our parametrization of the structural contribution of the surface effect or to the one by Sonoi et al. (2015). The fact that both Kjeldsen et al. (2008) and Ball & Gizon (2014) adjust the fitting coefficients according to data also implies that a comparison with our surface correction or S15 does not take place on equal footings. In order to investigate the ramifications hereof, we have repeated the evaluation of the best-fitting model based on B&G fixing the coefficients to the values obtained for our solar calibration model: $a_{-1} = 4.27 \times 10^{-12}$ and $a_3 = -6.93 \times 10^{-11}$. In this evaluation of the coefficients, we have used observed BiSON frequencies (Broomhall et al. 2009; Davies et al. 2014). We refer to the use of the solar values for the coefficients as ‘case b’, whilst we refer to the use of equation (22) as ‘case a’.

We note that the incorrect modelling of the surface layers results in wrong boundary conditions for solving the stellar structure equa-

tions throughout the predicted evolution. This leads to systematic errors in the associated stellar parameters, which none of the surface correction relations take into account. We refer to Mosumgaard et al. (2018) for a discussion on how the implementation of information from 3D simulations into 1D models alters the evolution tracks.

Furthermore, we recall that we restrict ourselves to radial modes ($\ell = 0$) in all cases in order to facilitate a meaningful comparison between all surface corrections.

5.4 Hare and hound

In order to evaluate the performance of our Lorentzian surface correction relation across the parameter space, we test how well the relation is able to recover the stellar parameters of PMs. For comparison, we have repeated the same hare and hound exercise, using K08, S15 and B&G, cases a and b.

Since the frequencies and global parameters of the PMs have no intrinsic uncertainties as such, we have attributed artificial uncertainties to mimic observational constraints. Hence, we assume the uncertainty on the effective temperature and metallicity to be 100 K and 0.1, respectively. All frequencies have been attributed an uncertainty of 0.3 μHz . For comparison, the symmetric errors on the radial modes vary between 0.07 and 1.36 μHz , for the two *Kepler* stars presented in Section 5.5. Whilst the chosen value is hence certainly not unreasonable, we note that the use of constant uncertainties is a simplifying approximation: for instance, the uncertainty will tend to reflect the mode amplitude.

We have performed five hare and hound exercises. We have chosen models with T_{eff} between 5600 and 5900 K and mostly with low metallicities. The reason for this selection is the attempt to reliably sample the likelihood for all surface corrections: especially the use of K08 leads to metallicities that are systematically too high and effective temperatures that are much lower than the target value, as discussed below. The global parameters of the five target models for the hare and hound exercises are furthermore selected so that we explore the region of the parameter space that is expected to be relevant for the *Kepler* stars that are to be addressed in Section 5.5. The results obtained for the hare and hound exercises from the MLE are summarized for each surface correction in Tables 4 and 5.

We find our parametrization of the surface effect to recover the stellar parameters rather well, and hence we infer that the Lorentzian fit encodes detailed information about the structural contribution to the surface effect.

S15 has been calibrated on fewer models and with different input physics than our PMs, and a dependence of the coefficients on metallicity was not included. Nevertheless, it likewise performs rather well, when establishing the model parameters in the hare and hound exercise. This may reflect the fact that S15 resembles our surface correction relation in being a parametrization of the structural contribution only. Also, this underlines that a Lorentzian parametrization captures the structural contribution.

In contrast, K08 becomes less reliable the further we move from the solar case. When employing this surface correction, the effective

Table 4. Global parameters from the hare and hound exercise based on PMs. Further parameters are listed in Table 5. The first column contains an identifier. The second through the fourth column contain the global parameters of the PMs. The fifth column lists the used surface correction relation. The sixth through the eighth column summaries, the parameter values for the best-fitting model, the weighted mean, and the square root of the weighted variance. The last column contains the initial metallicity of the best-fitting model. The purpose of including the latter quantity is solely to demonstrate that none of the models are at the edge of our grid of stellar models; hence, we do not provide uncertainties on the initial metallicity. We compare the surface corrections by Ball & Gizon (2014) in combination with equation (22) (B&G, a), the one by Ball & Gizon (2014) with fixed coefficients (B&G, b), the one presented in this paper, the one by Kjeldsen et al. (2008, K08), and the one by Sonoi et al. (2015, S15).

Hare	T_{eff} (K)	$\log g$ (dex)	[Fe/H] (dex)	Surf. Corr.	$T_{\text{eff}}^{\text{mod}}$ (K)	$\log g^{\text{mod}}$ (dex)	[Fe/H] ^{mod} (dex)	[Fe/H] _{init} ^{mod} (dex)
A	5624	4.054	−0.028	K08	5372, 5378 ± 31	4.045, 4.046 ± 0.002	0.081, 0.105 ± 0.046	0.20
				B&G (a)	5596, 5613 ± 183	4.053, 4.053 ± 0.008	0.020, −0.037 ± 0.119	0.13
				B&G (b)	5463, 5487 ± 59	4.045, 4.045 ± 0.005	0.008, −0.024 ± 0.072	0.13
				S15	5596, 5543 ± 101	4.053, 4.049 ± 0.009	0.002, −0.028 ± 0.071	0.13
				This work	5553, 5604 ± 109	4.045, 4.049 ± 0.009	−0.086, −0.076 ± 0.093	0.04
B	5926	4.185	−0.301	K08	5486, 5467 ± 91	4.181, 4.182 ± 0.003	0.096, 0.141 ± 0.068	0.25
				B&G (a)	6077, 6016 ± 90	4.197, 4.196 ± 0.004	−0.334, −0.274 ± 0.092	−0.14
				B&G (b)	5917, 5944 ± 47	4.182, 4.188 ± 0.006	−0.344, −0.285 ± 0.072	−0.16
				S15	5963, 5944 ± 58	4.194, 4.196 ± 0.004	−0.238, −0.183 ± 0.069	−0.06
				This work	5934, 5944 ± 44	4.189, 4.191 ± 0.004	−0.258, −0.249 ± 0.069	−0.08
C	5895	4.053	−0.262	K08	5399, 5426 ± 51	4.035, 4.036 ± 0.004	−0.049, −0.057 ± 0.055	0.07
				B&G (a)	5878, 5792 ± 171	4.056, 4.052 ± 0.008	−0.197, −0.175 ± 0.123	−0.04
				B&G (b)	5508, 5617 ± 121	4.027, 4.035 ± 0.009	−0.266, −0.247 ± 0.054	−0.14
				S15	5953, 5867 ± 79	4.053, 4.053 ± 0.006	−0.349, −0.249 ± 0.099	−0.17
				This work	5819, 5900 ± 74	4.042, 4.052 ± 0.007	−0.343, −0.289 ± 0.059	−0.19
D	5781	4.055	0.06	K08	5412, 5398 ± 53	4.041, 4.037 ± 0.006	0.213, 0.159 ± 0.114	0.33
				B&G (a)	5668, 5681 ± 181	4.049, 4.049 ± 0.008	0.090, 0.060 ± 0.115	0.22
				B&G (b)	5667, 5578 ± 114	4.036, 4.033 ± 0.009	−0.099, −0.072 ± 0.092	0.03
				S15	5710, 5740 ± 63	4.051, 4.051 ± 0.005	0.076, 0.037 ± 0.083	0.21
				This work	5813, 5758 ± 84	4.051, 4.053 ± 0.006	−0.042, 0.042 ± 0.118	0.06
E	5874	4.219	−0.008	K08	5486, 5460 ± 68	4.213, 4.210 ± 0.004	0.322, 0.313 ± 0.046	0.47
				B&G (a)	5995, 5844 ± 183	4.225, 4.222 ± 0.004	−0.075, 0.061 ± 0.166	0.09
				B&G (b)	5874, 5924 ± 64	4.219, 4.217 ± 0.002	−0.008, −0.089 ± 0.083	0.15
				S15	5831, 5834 ± 36	4.224, 4.224 ± 0.001	0.099, 0.100 ± 0.049	0.25
				This work	5886, 5897 ± 23	4.220, 4.219 ± 0.001	−0.019, −0.032 ± 0.025	0.14

temperature of the PM is systematically underestimated by a few hundred degrees. Also, the inferred metallicity is much higher than for the original model.

Using K08 leads to the selection of a model, whose frequencies already closely match the frequencies of the PM without the necessity of a surface correction. This peculiarity also occasionally occurs, when using B&G with fixed coefficient (case b), but contrasts the selection obtained with the other surface correction relations, including B&G in combination with equation (22) (case a). It is exemplified in Fig. 9.

As discussed in Section 5.5, K08 performs better when confronted with *Kepler* data. Therefore, the selection by K08 may partly reflect the neglect of modal effects in our surface corrections relation, i.e. the fact that the frequency difference between the UPM and the associated PM does not fit the expected functional form: as shown in Section 4, the difference between a PM and the associated UPM is not well-described by a power law.

Furthermore, in the case of B&G, the inferred surface correction is lower when the coefficient are fixed to the solar values (case b) than when equation (22) is used to establish the coefficients (case a) for each model anew. It hence stands to reason that the evaluation of the coefficients based on the solar case distorts the model selection in the investigated region of the parameter space. This conclusion seems to agree with the qualitative finding of Section 4.3: no unique set of coefficient values are applicable throughout the Kiel diagram. This agrees with the conclusions drawn by Ball &

Gizon (2017): surface corrections that employ coefficients that are calibrated to the Sun perform comparatively poorly, when fitting *Kepler* observations.

Despite the neglect of the modal effects in the computation of the model frequencies of the PMs, B&G performs rather well, when using equation (22) (case a). However, in case a, B&G also leads to the broadest probability distributions, as a_{-1} and a_3 are treated as free parameters. Thus, the higher accuracy comes at the price of lower precision.

Whilst the different surface corrections recover parameters such as the effective temperature and the metallicity with varying success, the gravitational acceleration is always accurately and precisely reproduced. This stems from the fact that we mainly constrain our models through the oscillation frequencies: the global oscillation properties of stars are determined by both the mass (M) and radius (R) of the star. Thus, both the frequency of maximum amplitude and the large frequency separation are proportional to g and \sqrt{g} , respectively.

5.5 *Kepler* stars

Having established how well our surface corrections reproduce the stellar parameters of PMs in a grid search, we proceed to an analysis of observational data. We search for best-fitting models using our MLE approach for the *Kepler* stars KIC 10516096 (Manon) and

Table 5. Pendant to Table 4 with other stellar parameters. The last column contains the χ^2 of the best-fitting model. The identifiers of the PMs correspond to those in Table 4.

Hare	M (M_{\odot})	R (R_{\odot})	t (Gyr)	Surf. Corr.	M^{mod} (M_{\odot})	R^{mod} (R_{\odot})	t^{mod} (Gyr)	χ^2
A	0.999	1.557	10.8	K08	0.959, 0.97 \pm 0.015	1.541, 1.546 \pm 0.008	13.57, 13.26 \pm 0.48	16.62
				B&G (a)	0.999, 0.992 \pm 0.045	1.557, 1.552 \pm 0.021	10.88, 11.09 \pm 2.14	0.4
				B&G (b)	1.000, 0.954 \pm 0.029	1.54, 1.537 \pm 0.015	12.82, 12.79 \pm 1.01	12.64
				S15	1.000, 0.972 \pm 0.051	1.557, 1.543 \pm 0.026	10.88, 12.06 \pm 2.05	2.49
				This work	0.949, 0.972 \pm 0.052	1.532, 1.543 \pm 0.025	12.35, 11.56 \pm 1.89	1.64
B	0.950	1.305	9.77	K08	0.950, 0.960 \pm 0.019	1.311, 1.315 \pm 0.008	13.64, 13.5 \pm 1.34	11.59
				B&G (a)	1.000, 0.998 \pm 0.018	1.32, 1.321 \pm 0.007	7.50, 7.98 \pm 0.85	6.29
				B&G (b)	1.000, 0.964 \pm 0.033	1.295, 1.310 \pm 0.014	10.35, 9.31 \pm 1.00	13.32
				S15	1.000, 1.005 \pm 0.021	1.319, 1.326 \pm 0.009	8.53, 8.34 \pm 0.65	14.21
				This work	0.970, 0.978 \pm 0.022	1.312, 1.315 \pm 0.01	9.22, 8.95 \pm 0.61	8.53
C	0.999	1.559	8.86	K08	0.919, 0.925 \pm 0.022	1.526, 1.529 \pm 0.012	14.37, 14.0 \pm 0.89	10.84
				B&G (a)	1.019, 1.001 \pm 0.041	1.568, 1.56 \pm 0.019	8.61, 9.61 \pm 1.80	0.36
				B&G (b)	1.000, 0.911 \pm 0.052	1.497, 1.518 \pm 0.026	14.85, 12.93 \pm 2.17	12.05
				S15	1.000, 0.996 \pm 0.036	1.55, 1.556 \pm 0.018	8.67, 9.15 \pm 0.94	6.38
				This work	0.939, 0.993 \pm 0.038	1.530, 1.555 \pm 0.018	10.55, 9.00 \pm 1.19	3.41
D	1.099	1.63	7.78	K08	1.029, 1.004 \pm 0.041	1.604, 1.591 \pm 0.021	11.36, 12.07 \pm 1.17	16.47
				B&G (a)	1.069, 1.064 \pm 0.041	1.618, 1.614 \pm 0.018	8.92, 9.07 \pm 1.72	0.97
				B&G (b)	1.000, 0.973 \pm 0.057	1.581, 1.573 \pm 0.029	10.46, 11.58 \pm 2.04	18.04
				S15	1.000, 1.074 \pm 0.033	1.623, 1.619 \pm 0.016	8.50, 8.46 \pm 0.78	5.19
				This work	1.069, 1.083 \pm 0.039	1.615, 1.622 \pm 0.019	8.19, 8.22 \pm 0.90	5.09
E	1.05	1.319	7.67	K08	1.030, 1.015 \pm 0.02	1.316, 1.311 \pm 0.008	10.69, 11.41 \pm 1.18	6.28
				B&G (a)	1.070, 1.067 \pm 0.018	1.323, 1.324 \pm 0.007	6.51, 7.49 \pm 1.28	2.18
				B&G (b)	1.000, 1.034 \pm 0.016	1.319, 1.312 \pm 0.007	7.67, 7.69 \pm 0.42	19.53
				S15	1.000, 1.081 \pm 0.011	1.330, 1.331 \pm 0.005	7.25, 7.20 \pm 0.23	6.23
				This work	1.050, 1.048 \pm 0.009	1.318, 1.317 \pm 0.003	7.59, 7.57 \pm 0.25	2.76

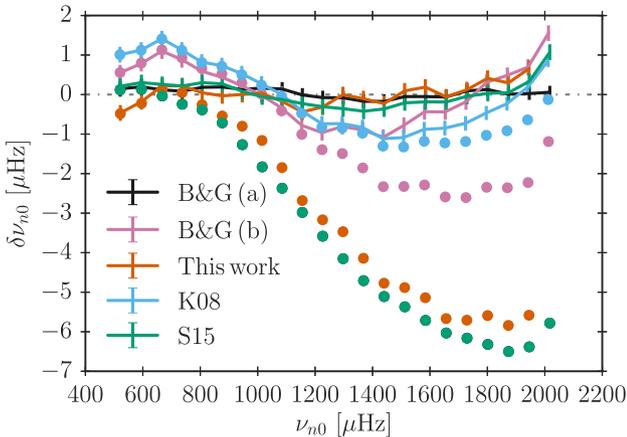


Figure 9. Frequency difference between model frequencies and observations for the PM with the identifier A in Tables 4 and 5, using different surface correction methods. The solid lines show the corrected frequencies. The circular markers in the same colour show the associated un-corrected frequencies of the respective best-fitting models.

KIC 5950854. The observational constraints on the employed radial frequencies as well as on T_{eff} and $[\text{Fe}/\text{H}]$ are taken from Lund et al. (2017). Both stars are in the LEGACY dwarf sample. The obtained stellar parameters from our MLE approach are summarized in Tables 6 and 7, respectively.

An analysis of MS stars with global properties that are rather similar to those of the Sun has already been presented in Paper I. We therefore selected KIC 10516096 and KIC 5950854 in

order to investigate stars at a later evolutionary stage. Furthermore, spectroscopic constraints indicate that the global parameters of these two stars lie well within our computed grid of stellar models, which allows us to properly map the probability distributions.

As can be seen from the tables, the use of B&G in combination with equation (22) (case a) leads to parameter estimates that are in good agreement with the spectroscopic constraints. Lesser agreement is achieved, when using the other surface correction relations and the same relation in case b. Thus, with the exception of B&G case a, the evaluated effective temperature is systematically too low, whilst the metallicity is too high in most cases.

The stellar age estimates vary strongly between the different fits, whilst estimates of $\log g$ and the stellar radius are mutually consistent and precisely determined. The mass estimates also closely match. Again, this seems to reflect the fact that the global oscillation properties encode stringent constraints on g .

We include the projected likelihood for the mass of KIC 10516096 in Fig. 10. As can be seen from the figure, the use of the surface correction relation by B&G, case a, leads to a much broader distribution than the remaining surface correction relations do. This is a general feature that we find for all projections – for both *Kepler* stars as well as in the hare and hound exercise. This can be explained by the fact that equation (22) adjusts the coefficients in order to optimize the fit, i.e. the fit has two additional free parameters, which ascribes artificially high likelihoods to incorrect structures.

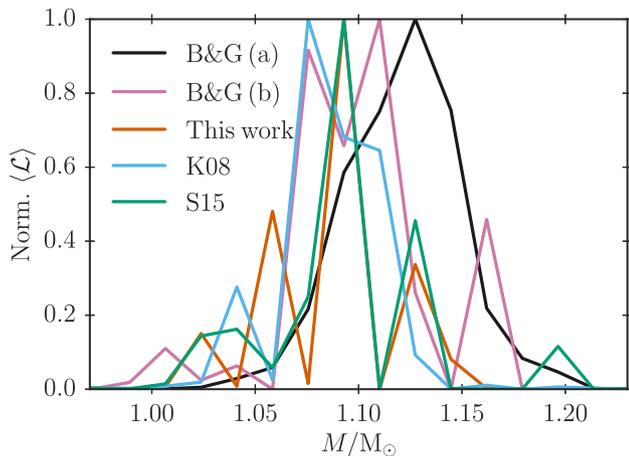
The frequency differences between the obtained best-fitting models and observations are shown in Figs 11 and 12. We note that the surface correction relations that rely on solar coefficients generally predict a smaller frequency correction than the remaining surface correction relations do.

Table 6. Best-fitting parameters for KIC 10516096. According to spectroscopy, $T_{\text{eff}} = 5964 \pm 77$ and $[\text{Fe}/\text{H}] = -0.11 \pm 0.10$, respectively.

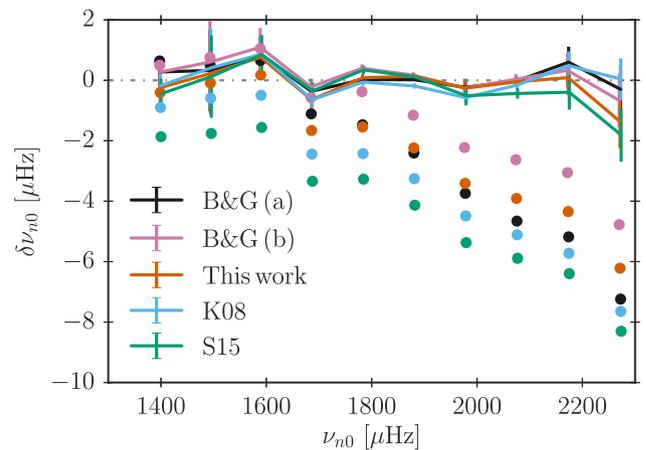
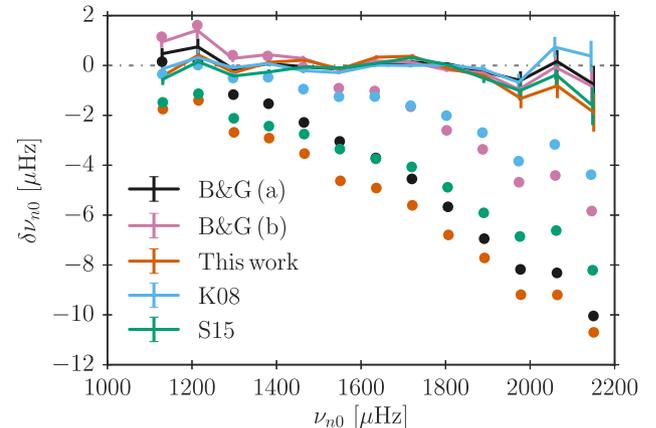
Quantity	Unit	K08	B&G (a)	B&G (b)	S15	This work
χ^2		5.73	2.45	7.06	6.44	7.93
T_{eff}	K	5535, 5592 \pm 63	5887, 5912 \pm 151	5772, 5724 \pm 68	5612, 5657 \pm 81	5760, 5734 \pm 65
$\log g$	dex	4.169, 4.17 \pm 0.004	4.178, 4.179 \pm 0.005	4.175, 4.173 \pm 0.006	4.173, 4.173 \pm 0.006	4.175, 4.173 \pm 0.005
$[\text{Fe}/\text{H}]$	dex	0.376, 0.312 \pm 0.075	0.104, 0.074 \pm 0.143	0.191, 0.211 \pm 0.073	0.336, 0.274 \pm 0.084	0.181, 0.186 \pm 0.099
$[\text{Fe}/\text{H}]_{\text{init}}$	dex	0.52	0.26	0.34	0.48	0.33
M	M_{\odot}	1.080, 1.081 \pm 0.027	1.120, 1.118 \pm 0.026	1.110, 1.101 \pm 0.04	1.100, 1.097 \pm 0.04	1.100, 1.090 \pm 0.029
R	R_{\odot}	1.417, 1.417 \pm 0.011	1.428, 1.425 \pm 0.01	1.427, 1.424 \pm 0.016	1.423, 1.421 \pm 0.015	1.421, 1.417 \pm 0.012
t	Gyr	9.272, 8.979 \pm 0.957	6.592, 6.507 \pm 1.038	7.369, 7.852 \pm 1.248	8.347, 8.244 \pm 1.247	7.641, 7.995 \pm 0.751

Table 7. Best-fitting parameters for KIC 5950854. According to spectroscopy, $T_{\text{eff}} = 5853 \pm 77$ and $[\text{Fe}/\text{H}] = -0.23 \pm 0.10$, respectively.

Quantity	Unit	K08	B&G (a)	B&G (b)	S15	This work
χ^2		2.08	0.6	2.33	3.22	3.4
T_{eff}	K	5702, 5694 \pm 170	5861, 5757 \pm 175	5673, 5647 \pm 148	5650, 5584 \pm 129	5656, 5671 \pm 145
$\log g$	dex	4.23, 4.235 \pm 0.013	4.234, 4.218 \pm 0.034	4.234, 4.234 \pm 0.012	4.235, 4.232 \pm 0.010	4.231, 4.234 \pm 0.011
$[\text{Fe}/\text{H}]$	dex	-0.117, -0.037 \pm 0.161	-0.211, -0.134 \pm 0.158	-0.020, 0.007 \pm 0.159	0.000, 0.019 \pm 0.153	-0.054, -0.033 \pm 0.151
$[\text{Fe}/\text{H}]_{\text{init}}$	dex	0.05	-0.04	0.14	0.16	0.11
M	M_{\odot}	0.930, 0.963 \pm 0.077	0.950, 0.949 \pm 0.063	0.960, 0.963 \pm 0.072	0.960, 0.944 \pm 0.060	0.940, 0.955 \pm 0.065
R	R_{\odot}	1.226, 1.238 \pm 0.031	1.232, 1.256 \pm 0.054	1.24, 1.24 \pm 0.029	1.238, 1.232 \pm 0.024	1.231, 1.236 \pm 0.026
t	Gyr	11.985, 11.356 \pm 3.495	9.971, 11.219 \pm 3.066	11.281, 11.702 \pm 3.337	11.472, 12.79 \pm 2.799	12.078, 11.691 \pm 3.061

**Figure 10.** Projection of the normalized exponential of the averaged log-likelihood for the stellar mass for KIC 10516096. The normalization factor is established in such a way that the likelihood of the maximum likelihood model is 1. We use 30 bins, smoothing the likelihood function.

At this point, we refer to Nsamba et al. (2018) for another recently published determination of the systematics arising from different surface correction methods. Nsamba et al. (2018) determine posterior probability distributions for the stellar mass, radius, and age using different surface correction relations by employing a Markov Chain Monte Carlo algorithm. They then compare the evaluated parameter values with those obtained, when following the same Bayesian inference technique but using frequency ratios suggested by Roxburgh & Vorontsov (2003). Since these frequency ratios have been shown to be rather insensitive to the surface layers, the associated parameter values are a suitable reference value and are taken as a measure of the systematic errors introduced by the use of each surface correction relation. Nsamba et al. (2018) conclude that B&G reproduces the mass, radius, and age that is inferred from frequency ratios. The same conclusion was drawn by Basu &

**Figure 11.** Same as Fig. 9 but for KIC 5950854.**Figure 12.** Same as Fig. 9 but for KIC 10516096.

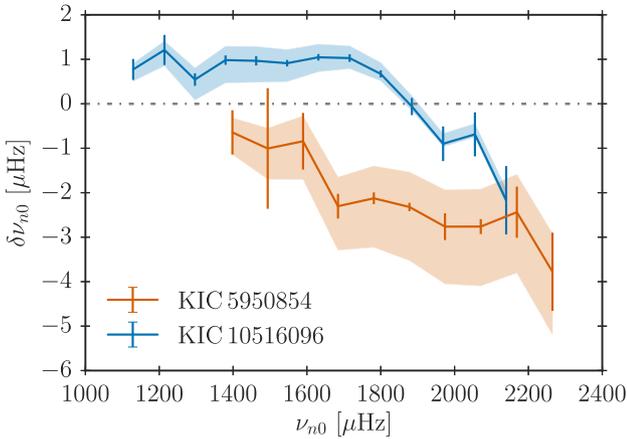


Figure 13. Frequency difference between observations and PM for KIC 5950854 and KIC 10516096. The UPM corresponds to the best-fitting model established using case a of B&G. The solid lines show the frequencies that are obtained, when using P as the patching quantity. The shaded area shows the frequency shift that is introduced, when using ρ or T as the patching quantity. The error bars illustrate the error on the observed frequencies.

Kinnane (2018). This suggests that B&G correctly encodes the structural as well as the modal effects. Thus, we may use parameter values that are based on case a of B&G as a benchmark for the other surface correction relations.

According to Nsamba et al. (2018), K08 leads to larger intrinsic systematics for the mass, radius, and age than B&G does. This is compatible with our results: for both *Kepler* stars K08 leads to lower mass estimates and a higher stellar age. Furthermore, Nsamba et al. (2018) conclude that S15 performs rather well. Again, this agrees with our mass, radius, and age estimates: S15 is rather consistent with case a of B&G. This being said, as regards S15, a direct comparison between our results and those obtained by Nsamba et al. (2018) is misleading, since Nsamba et al. (2018) defines the coefficients differently: β is set to 4.0, whilst α calibrated based on the observed frequencies.

5.6 PMs of *Kepler* stars

Having established that case a of B&G is plausibly the most reliable surface correction relation for establishing stellar parameters, we can use the best-fitting UPMs obtained based on B&G to construct PMs, in order to model the structure of the *Kepler* targets. The associated difference between the adiabatic model frequencies of the PMs and the observations are summarized in Fig. 13.

In order to demonstrate the systematic errors that are introduced by our choice of the patching quantity, Fig. 13 includes three PMs for each *Kepler* star, using P , ρ , or T as the respective patching quantities. As elaborated upon in Section 4, the sensitivity of the oscillation frequencies to the patching quantity stem from discontinuities in the patched structure. This is illustrated in Fig. 14 for KIC 10516096. In this figure, we use the density as the patching quantity, which results in a discontinuity in the temperature and pressure stratifications: whilst $\rho(r)$ is a smooth function by construction, neither $T(r)$ nor $P(r)$ is. Consequently, $T(P)$ in Fig. 14 is not a smooth function. One way to quantify this discontinuity is to evaluate the change in the stellar radius that is introduced by a change in the patching quantity: using the temperature rather than the density as the patching quantity would shift the radius of the

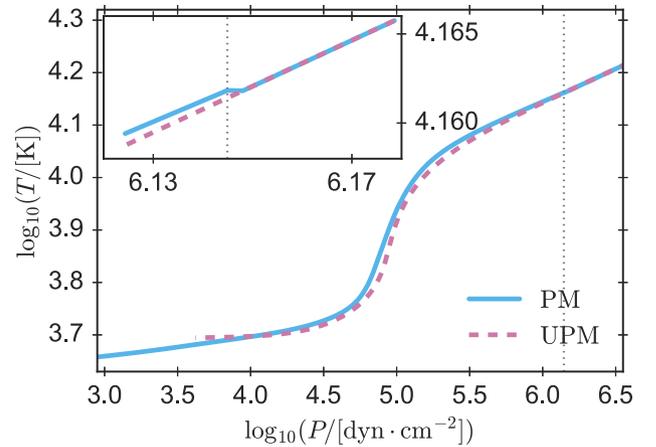


Figure 14. Temperature as a function of pressure for KIC 10516096. The cyan solid line and the dashed magenta line show the stratification of the UPM and PM, respectively. The dotted grey line shows the location of the innermost mesh point of the patched envelope. A zoom-in is included in the upper left corner. For the PM, ρ has been used as the patching quantity. Consequently, there is discontinuity in the temperature stratification at the patching point.

patching point by 18 km. To put this into perspective, we note that the radius of both stars is larger than that of the Sun (cf. Tables 6 and 7) and that patch is performed roughly 2×10^3 km below the surface in either case. In comparison, this shift would be 73 km in the case of KIC 5950854. As a result, the oscillation frequencies of KIC 5950854 are more sensitive to the choice of the patching quantity than those of KIC 10516096, as shown in Fig. 13.

According to Houdek et al. (2017), the modal surface effect counteracts the structural surface effect, increasing the model frequencies – that is at least the case for the Sun. At first glance, Fig. 13 seems to contradict this conclusion in the case of KIC 5950854: the obtained model frequencies of the PM are systematically higher than the observed frequencies. However, in accordance with Ball et al. (2016), Sonoi et al. (2015) and Paper I, we use Γ_1 , when calculating model frequencies, whilst Houdek et al. (2017) modify Γ_1 by the gas pressure as fraction of the total pressure. This is done in order to consistently take the contribution of the turbulent pressure into account and alters the model frequencies (cf. Jørgensen et al. 2018). Thus, the frequency differences between the model frequencies and observations in Fig. 13 do not only include contributions that Houdek et al. (2017) denote as modal.

If we assume that case a of B&G accurately reproduces the stellar structure and frequencies, Fig. 13 rather points towards the conclusion that the structural surface effect alone does not reliably mimic the total surface effect: the deviations are quite significant for KIC 5950854. This is an interesting notion, since it illustrates, why the surface correction in this paper as well as the correction relation by Sonoi et al. (2015) is subject to systematic errors, when estimating stellar parameters.

Whether B&G indeed accurately reproduces the underlying stellar structures is not settled. To answer this question, a detailed analyses of how the structural as well as the modal effect depends on the stellar parameters is needed. This is beyond the scope of this paper. This being said, the comparison by Nsamba et al. (2018) and Basu & Kinnane (2018), in which they infer the systematic error introduced by the B&G performs relatively well.

We note that the model frequencies in Fig. 13 may also reflect inconsistencies in the physical assumptions that enter the UPMs and the patched (3D) envelopes, which would affect the model frequencies. Moreover, due to the restrictions imposed by the interpolation scheme, the patched envelopes may be too shallow, i.e. the patch may not have been performed sufficiently deep within the adiabatic region for the oscillation frequencies to be entirely insensitive to the patching depth (cf. Paper I). In order to exclude this scenario, deeper 3D simulations are required, which again lies beyond the scope of this paper.

Finally, interpolation errors may give rise to further frequency shifts: the structure of the patched envelope may be incorrect. As pointed out in Paper I, the sampling of existing grids of 3D envelopes may hence be too low in some regions of the parameters space. To investigate this possibility, an extension of the Stagger grid is needed.

6 CONCLUSION

We present a method that allows for the interpolation of mean stratifications of 3D simulations of stellar envelopes in metallicity. Hereby, we add an additional parameter to the interpolation scheme published in Paper I, broadening its applicability. The interpolated envelopes can be used to correct the improper structure of current 1D stellar structure models by substituting these layers. We present such PMs for both the Sun and stars in the *Kepler* field.

We show that our interpolation scheme reproduces the correct mean stratification quite accurately throughout the parameter space. Based on the solar case, we note that interpolation errors as well as the limited depth of the interpolated structure may lead to a frequency shift of a few microhertz. In order to improve the performance of the interpolation scheme, a refinement of the Stagger grid of 3D envelopes is needed. Also, our results call for deeper 3D simulations.

Based on our interpolation scheme, we derive a parametrizations of the structural contribution of the surface effect, using the same functional forms as suggested by Sonoi et al. (2015): a power law and a Lorentzian. We find the coefficients of the parametrizations to be sensitive to the selected sample of stellar models. Hence, we conclude that no unique set of coefficients is applicable for all evolutionary stages. Consequently, our results discourages the use of any semi-empirical surface correction that is calibrated to fit the Sun for stars, whose parameters deviate strongly from the solar case. Rather, surface corrections must be calibrated based on stars that closely resemble the target.

Our parametrizations only encode the structural surface effect, neglecting modal contributions (cf. Houdek et al. 2017). Using a simple MLE approach, we derive stellar parameters using both our parametrization and the surface correction by Ball & Gizon (2014) that incorporates structural as well as modal effects. From this comparison, we conclude that the neglect of modal effects alter the parameter estimates. Consequently, without further adjustments, our results do not support the use of our parametrization or the surface correction by Sonoi et al. (2015), when evaluating stellar parameters.

As shown by Jørgensen et al. (2018), it is possible to avoid the necessity of post-evolutionary patching, by including mean 3D envelopes at each time-step, adjusting the boundary conditions of the stellar structure equations accordingly. Having shown that we are able to reliably interpolate in metallicity, we hence plan to implement an interpolation in the stellar composition into our stellar evolution code. In this way, we can take the metallicity

dependence of structure into account on the fly. Including this additional information from 3D simulations may possibly alter the evolution tracks (cf. Mosumgaard et al. 2018).

ACKNOWLEDGEMENTS

We thank J. Christensen-Dalsgaard for his kind assistance with ADIPLS and record our gratitude to Z. Magic and R. Trampedach for supplying the employed mean 3D hydrodynamic simulations as well as for fruitful discussions and helpful insights. Furthermore, we thank R. Collet for sharing his invaluable understanding of the Stagger grid and for his generous assistance. We likewise acknowledge T. Enßlin, C. L. Sahlholdt, and J. R. Mosumgaard for the useful discussions and input. VSA acknowledges support from VILLUM FONDEN (research grant no. 10118) and the Independent Research Fund Denmark (Research grant no. 7027-00096B). Finally, we record our gratitude to our referee, who provided very detailed feedback that showed an exceptional level of care and effort and that improved the paper in many ways.

REFERENCES

- Adelberger E. G. et al., 2011, *RvMP*, 83, 195
Aerts C., Christensen-Dalsgaard J., Kurtz D. W., 2010, *Asteroseismology*. Springer, Dordrecht
Akima H., 1970, *JACM*, 17, 589
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARAA*, 47, 481
Baglin A. et al., 2009, in Pont F., Sasselov D., Holman M.J., eds, *IAU Symp.*, Vol. 253. p. 71
Ball W. H., Gizon L., 2014, *A&A*, 558, A123
Ball W. H., Beeck B., Cameron R. H., Gizon L., 2016, *A&A*, 592, A159
Ball W. H., Gizon L., 2017, *A&A*, 600, A128
Basu S., Kinnane A., 2018, *ApJ*, 869, 11p
Batalha N. M., 2014, *PNAS*, 111, 35
Böhm-Vitense E., 1958, *Zs. Ap.*, 46, 108
Borucki W. J. et al., 2010, *Science*, 327, 977
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., Gilliland R. L., Noyes R. W., Ramsey L. W., 1991, *ApJ*, 368, 599
Brown T. M., 1984, *Science*, 226, 687
Cassisi S., Salaris M., Irwin A. W., 2003, *ApJ*, 588, 862
Christensen-Dalsgaard, J. Däppen W., Lebreton Y., 1988, *Nature*, 336, 634
Christensen-Dalsgaard J., Kjeldsen H., Brown T. M., 2010, *ApJ*, L164, 713
Christensen-Dalsgaard J., 2002, *Rev. Mod. Phys.*, 74, 1073
Christensen-Dalsgaard J., 2008, *Ap&SS*, 316, 113
Cyburt R. H., Fields B. D., Olive K. A., Yeh T.-H., 2016, *Rev. Mod. Phys.*, 88, 1
Davies G. R., Chaplin W. J., Elsworth Y., Hale S. J., 2014, *MNRAS*, 441, 3009
Davies G. R. et al., 2016, *MNRAS*, 456, 2183
Ferguson J. W., Alexander D. R., Allard F., Barman T., Bodnarik J. G., Hauschildt P. H., Heffner-Wong A., Tamanai A., 2005, *ApJ*, 623, 585
Gough D. O., 1990, in Osaki Y., Shibahashi H., eds, *Progress of Seismology of the Sun and Stars*, Lect. Notes Phys., vol. 367. Springer Verlag, Berlin, p. 283
Gustafsson B., Edvardsson B., Eriksson K. et al., 2008, *A&A*, 486, 951
Houdek G., Trampedach R., Aarslev M. J., Christensen-Dalsgaard J., 2017, *MNRAS*, 464, L124
Hummer D. G., Mihalas D., 1988, *ApJ*, 331, 794
Iglesias C. A., Rogers F. J., 1996, *ApJ*, 464, 943
Jørgensen A. C. S., Weiss, 2018, *MNRAS*, 481, 4389
Jørgensen A. C. S., Weiss A., Mosumgaard J. R., Silva Aguirre V., Sahlholdt C., 2017, *MNRAS*, 472, 3264 (Paper I)
Jørgensen A. C. S., Mosumgaard J. R., Weiss A., Silva Aguirre V., Christensen-Dalsgaard J., 2018, *MNRAS*, 481, L35

- Kjeldsen H., Bedding T. R., Christensen-Dalsgaard J., 2008, *ApJ*, 683, L175
- Lebreton, Y. Goupil M. J., 2014, *A&A*, 569A, 21L
- Leighton R. B., 1960, *Proc. IAU symp.*, 12, 321
- Lund M. N. et al., 2017, *ApJ*, 835, 172
- Magic Z., Collet R., Hayek W., Asplund M., 2013, *A&A*, 560, A8
- Magic Z., Weiss A., 2016, *A&A*, 592, A24
- Magic Z., Weiss A., Asplund M., 2015, *A&A*, 573, A89
- Magic Z., Collet R., Asplund M., Trampedach R., Hayek W., Chiavassa A., Stein R. F., Nordlund Å., 2013, *A&A*, 557, A26
- Mosumgaard J. R., Ball W. H., Silva Aguirre V., Weiss A., Christensen-Dalsgaard J., 2018, *MNRAS*, 478, 5663
- 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
- Rogers F. J., Nayfonov A., 2002, *ApJ*, 576, 1064
- 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*, 311, 215
- Schlattl H., Weiss A., Ludwig H. G., 1997, *A&A*, 322, 646
- Silva Aguirre V., Ballot J., Serenelli A. M., Weiss A., 2011, *A&A*, 529, 63
- Silva Aguirre V. et al., 2015, *MNRAS*, 452, 2127
- Silva Aguirre V. et al., 2017, *ApJ*, 835, 173S
- Sonoi T., Samadi R., Belkacem K., Ludwig H.-G., Caffau E., Mosser B., 2015, *A&A*, 583, A112
- Stein R. F., Nordlund Å., 1998, *ApJ*, 499, 914
- Tayar J. et al., 2017, *ApJ*, 840, 12
- Trampedach R., Asplund M., Collet R., Nordlund Å., Stein R. F., 2013, *ApJ*, 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
- Ulrich R. K., 1970, *ApJ*, 162, 993
- Weiss A., Schlattl H., 2008, *Ap&SS*, 316, 99
- Chaplin W. J., Miglio A., 2013, *ARA&A*, 51, 353

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