

*This article has been accepted for publication in Monthly Notices of the Royal Astronomical Society ©: 2020 The Authors. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.*

# Investigating surface correction relations for RGB stars

Andreas Christ Sølvesten Jørgensen,<sup>1</sup>★ Josefin Montalbán,<sup>1</sup> Andrea Miglio<sup>1</sup>,<sup>1</sup>  
Ben M. Rendle<sup>1</sup>, Guy R. Davies,<sup>1</sup> Gaël Buldgen,<sup>2</sup> Richard Scuflaire,<sup>3</sup> Arlette Noels,<sup>3</sup>  
Patrick Gaulme<sup>4</sup> and Rafael A. García<sup>5,6</sup>

<sup>1</sup>*School of Physics and Astronomy, University of Birmingham, Edgbaston B15 2TT, UK*

<sup>2</sup>*Observatoire de Genève, Université de Genève, 51 Ch. Des Maillettes, CH-1290 Sauverny, Switzerland*

<sup>3</sup>*Space sciences, Technologies and Astrophysics Research (STAR) Institute, Université de Liège, 19C Allée du 6 Août, B-4000 Liège, Belgium*

<sup>4</sup>*Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, D-37077 Göttingen, Germany*

<sup>5</sup>*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*

<sup>6</sup>*AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France*

Accepted 2020 May 22. Received 2020 May 22; in original form 2020 April 24

## ABSTRACT

State-of-the-art stellar structure and evolution codes fail to adequately describe turbulent convection. For stars with convective envelopes such as red giants, this leads to an incomplete depiction of the surface layers. As a result, the predicted stellar oscillation frequencies are haunted by systematic errors, the so-called surface effect. Different empirically and theoretically motivated correction relations have been proposed to deal with this issue. In this paper, we compare the performance of these surface correction relations for red giant branch stars. For this purpose, we apply the different surface correction relations in asteroseismic analyses of eclipsing binaries and open clusters. In accordance with previous studies of main-sequence stars, we find that the use of different surface correction relations biases the derived global stellar properties, including stellar age, mass, and distance estimates. We, furthermore, demonstrate that the different relations lead to the same systematic errors for two different open clusters. Our results overall discourage from the use of surface correction relations that rely on reference stars to calibrate free parameters. Due to the demonstrated systematic biasing of the results, the use of appropriate surface correction relations is imperative to any asteroseismic analysis of red giants. Accurate mass, age, and distance estimates for red giants are fundamental when addressing questions that deal with the chemo-dynamical evolution of the Milky Way galaxy. In this way, our results also have implications for fields such as galactic archaeology that draw on findings from stellar physics.

**Key words:** Asteroseismology – methods: statistical – stars: atmospheres – binaries: eclipsing – open clusters and associations: individual: 6819 and 6791.

## 1 INTRODUCTION

Asteroseismology, i.e. the study of stellar oscillations, has become an invaluable tool for stellar physics, yielding unique insights into stellar structures as well as precise stellar parameters (e.g. Aerts, Christensen-Dalsgaard & Kurtz 2010; Chaplin & Miglio 2013). This makes asteroseismology the backbone of other flourishing disciplines branching from exoplanet research to galactic archaeology (e.g. Miglio et al. 2009; Christensen-Dalsgaard et al. 2010; Batalha et al. 2011; Miglio 2012; Huber et al. 2013; Miglio et al. 2013; Van Eylen et al. 2013; Lundkvist et al. 2018; García & Ballot 2019).

The success of asteroseismology relies on comparisons between state-of-the-art stellar models with observations. However, to give a holistic depiction of stellar structures and their evolution, current stellar models draw on a set of simplifying approximations that result in prominent tensions with data, which greatly complicates such comparisons. One prevailing approximation is the use of mixing length theory (MLT; Böhm-Vitense 1958) or similar parametrizations to model turbulent convection. Another such simplification is the assumption that stars are spherically symmetric. In tandem, these approximations result in an incorrect depiction of the outermost superadiabatic layers of stars with convective envelopes. This model inadequacy affects predicted model frequencies that therefore show a systematic offset relative to observations. The described frequency shift is known as the structural surface effect.

\* E-mail: a.c.s.joergensen@bham.ac.uk

By drawing upon multidimensional, radiative magnetohydrodynamic (MHD) simulations of convection (e.g. Magic et al. 2013; Trampedach et al. 2013), it has been shown that the structural deficiencies that underlie the surface effect can be successfully overcome. This has been demonstrated by different authors and has been accomplished during the post-processing of stellar structure models (Rosenthal et al. 1999; Piau et al. 2014; Sonoi et al. 2015; Ball et al. 2016; Magic & Weiss 2016; Jørgensen et al. 2017; Manchon et al. 2018; Jørgensen et al. 2019). Furthermore, recently, stellar models that use information from multidimensional MHD simulations throughout the stellar evolution and hereby successfully mimic the structure of such simulations have become available (Jørgensen et al. 2018; Jørgensen & Angelou 2019; Jørgensen & Weiss 2019; Mosumgaard et al. 2020).

In addition to the structural surface effect, the frequencies of stellar models suffer from the so-called modal surface effect, denoting yet another systematic frequency offset. The modal surface effect stems from the fact that oscillation frequencies of stars are computed under the assumption of adiabaticity, which does not hold true in the near-surface layers (e.g. Dupret 2004; Christensen-Dalsgaard 2008; Townsend & Teitler 2013; Grosjean et al. 2014). Moreover, the contribution from turbulent pressure, i.e. the pressure stemming from the bulk motion of a convective fluid, is often not correctly accounted for, further contributing to the aforementioned frequency shift (cf. Houdek et al. 2017; Houdek et al. 2019; Jørgensen & Weiss 2019; Schou & Birch 2020, for a detailed discussion of this issue).

To compare model frequencies with observations, many authors rely on empirical surface correction relations that address the combined surface effect. The first such relation was proposed by Kjeldsen, Bedding & Christensen-Dalsgaard (2008). The relation amounts to a power law for which Kjeldsen et al. (2008) calibrated the involved free parameters to mitigate the surface effect of the present-day Sun. However, there is no physical justification for using a power law, and the necessity of calibrating the relation makes it unsuitable for any star whose global parameters deviate strongly from those of the reference star (cf. Ball & Gizon 2017; Jørgensen et al. 2019). To mitigate these drawbacks, one may include a large sample of reference stars to evaluate how the power-law description evolves across the Hertzsprung–Russell (HR) diagram and investigate different functional forms. This was accomplished by Sonoi et al. (2015), who employed three-dimensional MHD simulations to overcome the structural inadequacies of stellar models.

Like the surface correction relation by Kjeldsen et al. (2008), the surface correction relation by Sonoi et al. (2015) is, of course, still subject to a selection bias: both surface correction relations can only be applied to stars whose global parameters are similar to those of the employed reference stars or models (Jørgensen et al. 2019). This being said, Sonoi et al. (2015) cover the global parameters of many main-sequence stars in the *Kepler* field (Borucki et al. 2010). The surface correction relation by Sonoi et al. (2015), meanwhile, suffers from a different drawback: the computation of the reference model frequencies relies on the so-called gas  $\Gamma_1$  approximation (Rosenthal et al. 1999). This means that the contributions of the turbulent pressure and non-adiabatic effects to the reference frequencies are not correctly accounted for. While the gas  $\Gamma_1$  approximation recovers the observed frequencies reasonably well,<sup>1</sup> in the case of

the present-day Sun, it is unclear how this approximation performs for any other star.

In contrast to Kjeldsen et al. (2008) and Sonoi et al. (2015), Ball & Gizon (2014) present a surface correction relation with a physically motivated functional form based on an asymptotic analysis by Gough (1990). While their surface correction relation still includes free parameters, these can be adjusted anew for each target star, circumventing the need for calibrating the parameters based on reference stars.

As regards main-sequence stars, studies by Basu & Kinnane (2018) and Nsamba et al. (2018) show that the use of the surface correction relation by Ball & Gizon (2014) leads to estimates for the stellar mass, radius, and age that are consistent with those obtained from alternative methods for handling the surface effect. One such alternative approach relies on a set of frequency ratios that were originally proposed by Roxburgh & Vorontsov (2003) and that have been shown to be insensitive to the incorrect depiction of the near-surface layers Otí Floranes, Christensen-Dalsgaard & Thompson (2005). Studies by Ball & Gizon (2017) and Nsamba et al. (2018), furthermore, show that the use of the surface correction relations by Kjeldsen et al. (2008) and Sonoi et al. (2015) introduce systematic errors in the stellar parameter estimates when addressing main-sequence stars and subgiants. How well the surface correction relation by Ball & Gizon (2014), or indeed any of the relations and methods mentioned above, perform throughout the HR diagram, including more evolved stages, is yet to be settled (see also Ball, Themeßl & Hekker 2018). In this paper, we address this issue by investigating how the use of different surface correction relations affect stellar parameter estimates for red giants.

For this purpose, we derive stellar parameter estimates adopting different surface correction relations for eight well-constrained eclipsing binaries as well as 19 and 30 red giants in the open clusters NGC 6819 and NGC 6791, respectively (e.g. Grundahl et al. 2008; Basu et al. 2011; Brogaard et al. 2012; Miglio et al. 2012; Jeffries Mark et al. 2013; Sandquist et al. 2013; Brogaard et al. 2015; Bossini et al. 2017; Handberg et al. 2017; Rodrigues et al. 2017; McKeever, Basu & Corsaro 2019).

Accurate parameter estimates for the ages of red giants are essential for establishing the dynamical and chemical evolution of the Milky Way galaxy. The performance of asteroseismology for evolutionary stages beyond the main sequence thus has profound implications for galactic archaeology, making the present study a valuable stepping stone for future analyses. Red giants are, furthermore, of interest to exoplanet research when addressing the dynamics and fate of planetary systems.

Section 2 addresses the underlying Markov Chain Monte Carlo (MCMC) approach as well as the employed stellar models. We introduce the different surface correction relations in Section 3. In Section 4–7, we successively present the analyses of the eclipsing binaries and the two clusters. Our main conclusions are summarized in Section 8.

## 2 AIMS: BAYESIAN INFERENCE OF STELLAR PARAMETERS

To evaluate posterior distributions of stellar parameters, we compare model predictions with observations for each individual star. For this purpose, we employ the open-source code, Asteroseismic Inference on a Massive scale (AIMS; Reese 2016; Lund & Reese 2018; Rendle et al. 2019), which is based on the MCMC ensemble sampler by Goodman & Weare (2010) using the implementation by Foreman-Mackey et al. (2013).

<sup>1</sup> Albeit, even in the case of the Sun, the gas  $\Gamma_1$  approximation only recovers the observed frequencies within a few microhertz.

A large variety of Monte Carlo methods have found their way into modern astrophysics. They have yielded new insights and robust parameter estimates in a variety of astrophysical fields and analyses, ranging from the peak bagging of stellar oscillation frequencies to cosmology (e.g. Gruberbauer et al. 2009; Handberg & Campante 2011; Porqueres et al. 2019a,b). Indeed, the employed oscillation frequencies have been derived from the observed light curves using an open-source MCMC peak-bagging algorithm called PBJAM<sup>2</sup> that likewise builds on the MCMC ensemble sampler by Goodman & Weare (2010). The observed frequencies were subsequently corrected to account for the Doppler shift that arises from the line-of-sight motion of the star relative to the observer (Davies et al. 2014). Three of the eight eclipsing binaries constitute exceptions: the mode identification of KIC 4054 905, KIC 4663 623, and KIC 9540 225 was performed using a maximum a posteriori method based on Gaulme, Appourchaux & Boumier (2009) and Benbakoura et al. (in preparation).

However, due to the high computational cost of successively computing a series of stellar models, Monte Carlo algorithms are not widespread in asteroseismic analyses that seek to derive stellar parameters. Some notable exceptions, meanwhile, exist including analyses of the present-day Sun (Bahcall, Serenelli & Basu 2006; Jørgensen & Christensen-Dalsgaard 2017; Vinyoles et al. 2017) as well as a handful of other main-sequence benchmark stars (e.g. Benomar, Appourchaux & Baudin 2009; Bazot, Bourguignon & Christensen-Dalsgaard 2012; Jørgensen & Angelou 2019). To bypass the high computational cost of MCMC, AIMS computes new samples for the Markov chain, i.e. models with a new combination of global stellar parameters, by interpolation in an existing grid of stellar models (cf. Section 2.1). In this way, AIMS is able to compute a large set of samples with a total computation time of a few hours: for each star, we compute 2000 samples for each of the 800 walkers with 10 different temperatures (see Gregory 2005 for a general introduction to MCMC algorithms). These samples are preceded by a burn-in phase of 4000 samples per walker. Based on thousands or even millions of samples, AIMS is thus able to robustly map the posterior probability distributions for the stellar parameters of evolved stars in the same time as it takes to run a handful of stellar evolution models from the zero-age main sequence (ZAMS) to the red giant branch (RGB).

## 2.1 Stellar model grid

We constructed two grids of stellar models on the RGB, using the CLÉS (Code Liégeois d'Évolution Stellaire; Scuflaire et al. 2008a) stellar evolution code, and computed the associated adiabatic model frequencies, using LOSC (Liège Oscillation Code Scuflaire et al. 2008b). One grid was employed for the modelling of the eclipsing binaries as well as NGC 6819. The second grid was constructed to model the stars in NGC 6791.

In the two grids, we vary both the initial mass and the metallicity, hereby exploring a 2D parameter space. For the first grid, the initial helium abundance ( $Y_i$ ) is assumed to be related to the initial abundance of heavy elements ( $Z_i$ ) in such a way that an increase in  $Z_i$  is accompanied by an equal increase in  $Y_i$ , i.e.  $\Delta Z_i / \Delta Y_i = 1.0$ . As mentioned above, this grid is applied to model the eclipsing binaries and NGC 6819. For the second grid, we set  $\Delta Z_i / \Delta Y_i$  to 2.0, in order to recover the chemical properties of NGC 6791 (Brogaard et al. 2012). For all models presented in this paper, we do not consider

alpha enrichment, i.e.  $[\alpha/\text{Fe}] = 0.0$  (cf. Section 6 for a discussion hereon).

Both grids cover the evolution from the pre-main sequence to the red giant branch. For the first grid, we have computed stellar models with masses between 0.7 and 2.5  $M_\odot$  in steps of 0.02  $M_\odot$ . The grid includes 23 different values of  $[\text{Fe}/\text{H}]$ , ranging from  $-2.5$  to 0.2 dex. As regards the metallicity, the step-size is not uniform but alters between 0.10 and 0.15 dex. The grid contains stellar models with radii up until 25  $R_\odot$ .

For the second grid, with which we address NGC 6791, we have again computed stellar models with masses between 0.7 and 2.5  $M_\odot$  in steps of 0.02  $M_\odot$ . The grid includes models with metallicities between  $-0.1$  to 0.5 dex in steps of 0.05 dex.

For both grids, the composition of the models is based on the solar mixture evaluated by Asplund et al. (2009). We use the FreeEOS by A. W. Irwin (Cassisi, Salaris & Irwin 2003), the nuclear reaction rates by Adelberger et al. (2011), and the semi-empirical  $T(\tau)$  relation by Vernazza, Avrett & Loeser (1981). We employ OPAL opacities (Iglesias & Rogers 1996) in combination with the low-temperature opacities by Ferguson et al. (2005). We have included both over and undershooting, setting the associated parameters ( $\alpha_{\text{ov}}$  and  $\alpha_{\text{un}}$ ) to 0.1 and 0.2 for over and undershooting, respectively. CLÉS uses instantaneous overshooting. The extent of the over and undershooting regions are  $\alpha_{\text{ov}} H_p$  and  $\alpha_{\text{un}} H_p$ , respectively, where  $H_p$  denotes the pressure scaleheight. In the case of convective core overshooting, we substitute  $H_p$  by the size of the convective core ( $r_{\text{cc}}$ ), if  $H_p > r_{\text{cc}}$ . In the over and undershooting regions, we use the radiative temperature gradient.

## 2.2 Likelihood

When comparing models to data, we include spectroscopic constraints on the effective temperature ( $T_{\text{eff}}$ ) as well as on the metallicity ( $[\text{Fe}/\text{H}]$ ), assuming that these measurements are uncorrelated and that the noise is Gaussian.

In addition, we include the individual radial mode frequencies ( $\ell = 0$ ) as asteroseismic constraints without directly imposing constraints on the radial order of each mode. The individual frequencies are compared to observations after applying a surface correction specified in Section 3. While adding non-radial mode frequencies may help to further constrain the stellar parameters, we limit ourselves to radial modes in this analysis. The reason for this choice is that we perform a differential study in which we seek to compare the different surface correction relations on equal footings. Thus, while the surface correction relation by Ball & Gizon (2014) applies to all modes, the surface correction relations by both Kjeldsen et al. (2008) and Sonoi et al. (2015) are only valid for radial modes.<sup>3</sup> We are, in other words, restricted in our choices regarding the seismic constraints by the prescriptions we seek to compare. As in the case of the spectroscopic constraints, the noise of each radial mode frequency is assumed to be Gaussian and uncorrelated with the remaining observed frequencies. This is a reasonable approximation for the considered modes.

Regarding priors, we only include the constraints on the parameters that enter the surface correction relations specified in Section 1. In other words, we do not include any further prior restrictions

<sup>2</sup>Cf. <https://github.com/grd349/PBJam>.

<sup>3</sup>This being said, as regards the power-law correction relation by Kjeldsen et al. (2008), some authors have included non-radial modes by addressing the effect of the mode inertia (e.g. Huber et al. 2013).

on the global stellar properties other than those that are indirectly imposed through the limited extent of our grids.

Following the described approach, however, we often find the posterior distributions of different stellar parameters to be multimodal. Furthermore, for one of the explored surface correction relations (see Section 4), the associated échelle diagrams show that the inferred surface effect often exceeds the observed large frequency separation ( $\Delta\nu$ ) and is a significant fraction of the frequency of maximum power ( $\nu_{\max}$ ). In other words, for these cases, the free parameters that enter the surface correction relation are chosen by AIMS, in such a way as to shift the model frequencies substantially. We deem such solutions to be un-physical. We will discuss these findings further in Sections 4 and 5.

One viable approach to address both issues mentioned above would be to make the priors on the global stellar parameters more informative. Alternatively, one may make the likelihood more informative. We settled for the latter approach, including  $\nu_{\max}$  into the likelihood, using the scaling relation by Brown et al. (1991) and Kjeldsen & Bedding (1995):

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

Here,  $M$  and  $R$  denote the stellar mass and photospheric radius, respectively.

The individual stellar acoustic mode frequencies tightly constrain the mean density. By including  $\nu_{\max}$  into the likelihood, we introduce constraints on  $\log g$  and thus lift degeneracies between the inferred stellar mass and radius, which explains the improved inference. Studies by Handberg et al. (2017) and Zinn et al. (2019) show that equation (1) is both accurate and precise for stars on the RGB with the considered metallicities (see also Viani et al. 2017) – Zinn et al. (2019) show that this statement holds true to a 2 per cent based on Gaia parallaxes (Gaia Collaboration 2016). Below, we discuss and compare the results obtained with and without constraints on  $\nu_{\max}$  to determine the implications of adding this quantity into the likelihood.

The fact that we obtain more accurate results by including  $\nu_{\max}$  goes to show the importance of additional constraints, including non-seismic measurements, which complement the individual observed frequencies. We note that we could have introduced other (non-seismic) constraints into the likelihood – of course, each of these come with their own caveats. One option would be to use the stellar luminosity. However, adding such constraints or exploring the ramifications of doing so would not contribute to answering the scientific question that we address in this paper: how does the use of specific surface correction relations affect the inferred stellar properties. Indeed, the use of additional or alternative non-seismic constraints might obscure the influence of the surface correction relations by dominating the likelihood. In other words, while good non-seismic constraints are often invaluable for asteroseismic analyses, they do not help us to discriminate between the different surface correction relations. As stated above, we therefore restrict ourselves to including weak spectroscopic constraints on  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$  in addition to the constraints on  $\nu_{\max}$ .

### 2.3 Goodness of fit

Throughout this paper, we discuss the goodness of fit for the maximum a posteriori models, i.e. the best-fitting models within each run. As a measure for the goodness of fit, we refer to the

reduced  $\chi^2$ -value of the best-fitting models:

$$\chi_{\text{red}}^2 = \frac{1}{N} \sum_{i=1}^N \frac{(x_{\text{mod},i} - x_{\text{obs},i})^2}{\sigma_i^2}. \quad (2)$$

Here, the sum runs over all  $N$  observational constraints, and the subscripts ‘mod’ and ‘obs’ refer to the model and observational values, respectively. The models, for  $\chi_{\text{red}}^2 \approx 1$ , reliably recover the observational constraints. While models that yield  $\chi_{\text{red}}^2 \gg 1$  simply constitute a poor fit, values of  $\chi_{\text{red}}^2 \ll 1$  imply that the data is overfitted.

As an alternative measure for the goodness of fit, we directly refer to the evaluated likelihood function from which we compute the so-called Bayesian information criterion (BIC):

$$\text{BIC} = \ln(N_s)k - 2 \ln(\hat{\mathcal{L}}), \quad (3)$$

where  $N_s$  denotes the combined number of samples from the Markov chains,  $k$  is the number of free parameters, and  $\hat{\mathcal{L}}$  denotes the maximized value of the likelihood function. Better models lead to lower values of the BIC: the second term in equation (3) rewards approaches, i.e. combinations of stellar models and surface correction relations that yield high values for the maximized likelihood. The first term in equation (3), meanwhile, penalizes models that reach this goal due to a high degree of complexity.

## 3 SURFACE CORRECTION RELATIONS

The surface effect describes a systematic offset ( $\delta\nu$ ) between the uncorrected adiabatic model frequencies ( $\nu_{\text{mod}}$ ) and observations ( $\nu_{\text{obs}}$ ):

$$\delta\nu = \nu_{\text{obs}} - \nu_{\text{mod}}. \quad (4)$$

As discussed in the introduction, this frequency offset comes about as a result of the structural inadequacies of state-of-the-art stellar models as well as simplifying approximations that enter the frequency computation.

When determining the likelihood of each model in the Markov Chain by comparing model frequencies to observations, we correct the model frequencies, taking the surface effect into account. We do so using different surface correction relations. We discuss each of these surface correction relations below.

To mitigate the surface effect, Kjeldsen et al. (2008) propose to fit the surface effect by a power law, calibrating the exponent ( $b$ ) based on the present-day Sun:

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

Throughout the rest of this paper, we refer to the surface correction relation in equation (5) as K08. We note that AIMS uses  $\nu_{\text{mod}}$  rather than  $\nu_{\text{obs}}$  when adopting equation (5) under the assumption that  $\delta\nu \ll \nu_{\text{obs}}$ .

When following the approach suggested by Kjeldsen et al. (2008),  $b$  is commonly set to the solar calibrated value of 4.9 (e.g. Nsamba et al. 2018), while  $a$  is adjusted based on the observed frequencies. Drawing upon the analysis of 3D simulations by Sonoi et al. (2015), one may alternatively vary both  $a$  and  $b$  as a function of the global stellar parameters:

$$\log |a| = 8.13 \log T_{\text{eff}} - 0.670 \log g - 30.2, \quad (6)$$

and  $a$  is negative, while

$$\log b = -3.16 \log T_{\text{eff}} + 0.184 \log g + 11.7. \quad (7)$$

Sonoi et al. (2015) argue in favour of substituting the power-law fit by a Lorentzian surface correction relation as this yields a better fit to the frequency shift derived from 3D MHD simulations within the gas  $\Gamma_1$  approximation mentioned above:

$$\frac{\delta\nu}{\nu_{\max}} = \alpha \left( 1 - \frac{1}{1 - (\nu_{\text{obs}}/\nu_{\max})^\beta} \right). \quad (8)$$

In the following, we refer to the surface correction relation in equation (8) as S15. According to the analyses by Sonoi et al. (2015),

$$\log |\alpha| = -7.69 \log T_{\text{eff}} - 0.629 \log g - 28.5, \quad (9)$$

and  $\alpha$  is negative, while

$$\log \beta = -3.86 \log T_{\text{eff}} + 0.235 \log g + 14.2. \quad (10)$$

As in the case of equation (5), AIMS draws on  $\nu_{\text{mod}}$  rather than  $\nu_{\text{obs}}$  when correcting model frequencies.

Finally, Ball & Gizon (2014) presents a surface correction relation based on an asymptotic analysis by Gough (1990):

$$\frac{\delta\nu_i}{\nu_{\text{ac}}} = \mathcal{I}^{-1} \left( a_{-1} \frac{\nu_{\text{ac}}}{\nu_{\text{mod}}} + a_3 \frac{\nu_{\text{mod}}^3}{\nu_{\text{ac}}^3} \right). \quad (11)$$

Here,  $a_{-1}$  and  $a_3$  are free parameters,  $\nu_{\text{ac}}$  denotes the acoustic cut-off frequency that scales linearly with  $\nu_{\text{max}}$ , and  $\mathcal{I}$  is the mode inertia. We note that we substitute  $\nu_{\text{ac}}$  by a reference frequency that is related to the dynamical time-scale in AIMS. In the following, we refer to the surface correction relation in equation (11) as BG14.

When faced with free parameters in any of the surface correction relations mentioned above, AIMS selects that combination of parameters that minimizes the discrepancy between model frequencies and observations. This is done by following the procedure suggested by Ball & Gizon (2014). In other words, when these parameters ( $a$ ,  $b$ ,  $\alpha$ ,  $\beta$ ,  $a_{-1}$ , or  $a_3$ ) are kept free, they are optimized for every single stellar model rather than being randomly sampled by the MCMC. In this way, the free parameters of the surface correction relations adjust based on the pseudo-random walk through the parameters space spanned by the stellar mass and composition.

We note that the free parameters that enter the surface correction relations depend on both  $\nu_{\text{mod}}$  and  $\nu_{\text{obs}}$ , which implies that the corrected model frequencies are correlated both internally and with  $\nu_{\text{obs}}$ . Following common practise, we do not take these correlations into account.

The following section gives an overview of which parameters are kept fixed and which parameters are varied in different approaches.

### 3.1 Choosing surface correction relation parameters

In this paper, we deal with nine distinct ways of addressing the surface effect. We hence both employ different surface correction relations and vary the prescriptions for the different parameters involved. A summary can be found in Table 1.

When using K08, we either fix  $b$  to the solar calibrated value of 4.9 as found by Kjeldsen et al. (2008), establish  $b$  using equation (7), or let  $b$  adjust freely. In the latter case, we require that  $b \geq 0$ , since the surface effect increases with increasing frequency. This is due to the frequency dependence of the upper turning point of the oscillation. High-frequency modes thus probe shallower near-surface layers than low-frequency modes do, i.e. the eigenfunctions of low-frequency modes are evanescent in the near-surface layers (cf. Christensen-Dalsgaard & Thompson 1997, for a detailed discussion). In all three cases, we let  $a$  vary freely, requiring that  $a \leq 0$ . This requirement builds on the assumption

**Table 1.** Summary of the different surface correction relations that are investigated in this paper.

Surf. corr.	Eq.	Parameters
K08 (a)	(5)	$a \leq 0, b = 4.9$
K08 (b)	(5)	$a \leq 0, b$ from equation ((7))
K08 (c)	(5)	$a \leq 0, b \geq 0$
S15 (a)	(8)	$\alpha \leq 0, \beta = 4.0$
S15 (b)	(8)	$\alpha \leq 0, \beta$ from equation ((10))
S15 (c)	(8)	$\alpha \leq 0, \beta \geq 0$
BG14 (a)	(11)	$a_{-1} = 0, a_3$ is free
BG14 (b)	(11)	$a_{-1}$ and $a_3$ are free
NoSC (None)	–	–

that the combined surface effect is negative as in the case of the present-day Sun and other main-sequence stars (e.g. Brown 1984; Christensen-Dalsgaard, Däppen & Lebreton 1988; Houdek et al. 2017; Houdek et al. 2019).

When dealing with S15, we similarly distinguish between three different approaches: in the first approach, we fix  $\beta$  to 4.0 as done by Nsamba et al. (2018). In the second approach,  $\beta$  is determined using equation (10). Finally, we let  $\beta$  vary, solely requiring that  $\beta \geq 0$ . In all cases, we let  $\alpha$  adjust freely, requiring that  $\alpha \leq 0$ , following the arguments given above.

As regards the cases of K08 and S15 for which we let both parameters adjust freely, we stress that this path is not commonly taken in literature. This is because the sole physical justification for K08 and S15 lies in the calibrated parameters. Indeed, as shown below, K08 and S15 do not perform well, when no such calibration has taken place. By allowing both parameters to adjust freely, we thus are able to highlight the limitations of K08 and S15.

We, furthermore, investigate two different cases based on BG14: in one case, we only include the cubic term, while we allow both coefficients ( $a_{-1}$  and  $a_3$ ) to vary freely in the second approach. In other words, we include both a one- and a two-term version of BG14.

For comparison, we include the case where no surface correction is taken into account. In the following, we refer to this as NoSC.

## 4 ECLIPSING BINARIES

Using AIMS, we have determined the stellar parameters of eight red giant branch stars in the *Kepler* field: KIC 4054 905, KIC 4663 623, KIC 5786 154, KIC 7037 405, KIC 8410 637, KIC 8430 105, KIC 9540 226, and KIC 9970 396. All eight stars are members of eclipsing and spectroscopic binaries, which allows for accurate dynamical measurements of their masses and radii. This property makes the investigated eight giants suitable benchmark stars for asteroseismic analyses (e.g. Gaulme et al. 2016; Li et al. 2018). We thus compare our results with the conclusions from dynamical studies by Gaulme et al. (2016), Brogaard et al. (2018), and Benbakoura et al. (in preparation). The observational constraints are summarized in Table 2.

For all investigated nine treatments of the surface effect, the best-fitting model, i.e. the model with highest posterior probability as well as the median of the obtained posterior probability distributions recover the correct masses and radii of most or all stars within  $3\sigma$ . This includes NoSC, i.e. the case where no surface correction relation has been implemented. KIC 4054 905 constitutes a prominent exception: the attempt to model this star without any surface correction relation fails to recover the correct mass and radius within  $10\sigma$ , while all surface correction relations perform

**Table 2.** Summary of the observational constraints that were employed to model the eclipsing binaries investigated in this paper. This includes the number of radial modes to which we refer as  $\Delta n$ . The observed frequencies were deduced using PBJAM or using the method presented by Gaulme et al. (2009). The spectroscopic constraints on  $T_{\text{eff}}$ ,  $[\text{Fe}/\text{H}]$  and  $\nu_{\text{max}}$  stem from Gaulme et al. (2016), Brogaard et al. (2018), Li et al. (2018), and Benbakoura et al. (in preparation).

	$\Delta n$	$T_{\text{eff}}$ [K]	$[\text{Fe}/\text{H}]$	$\nu_{\text{max}}$
KIC 4054 905	6	$4790 \pm 190$	$-0.72 \pm 0.31$	$48.15 \pm 0.21$
KIC 4663 623	7	$4803 \pm 91$	$0.16 \pm 0.04$	$46.51 \pm 2.34$
KIC 5786 154	5	$4747 \pm 100$	$-0.06 \pm 0.06$	$29.75 \pm 0.16$
KIC 7037 405	5	$4500 \pm 80$	$-0.27 \pm 0.05$	$21.75 \pm 0.14$
KIC 8410 637	7	$4699 \pm 91$	$0.16 \pm 0.05$	$46.00 \pm 0.19$
KIC 8430 105	8	$5042 \pm 68$	$-0.49 \pm 0.04$	$76.70 \pm 0.57$
KIC 9540 226	6	$4662 \pm 91$	$-0.16 \pm 0.08$	$27.88 \pm 0.17$
KIC 9970 396	6	$4860 \pm 80$	$-0.35 \pm 0.1$	$63.70 \pm 0.16$

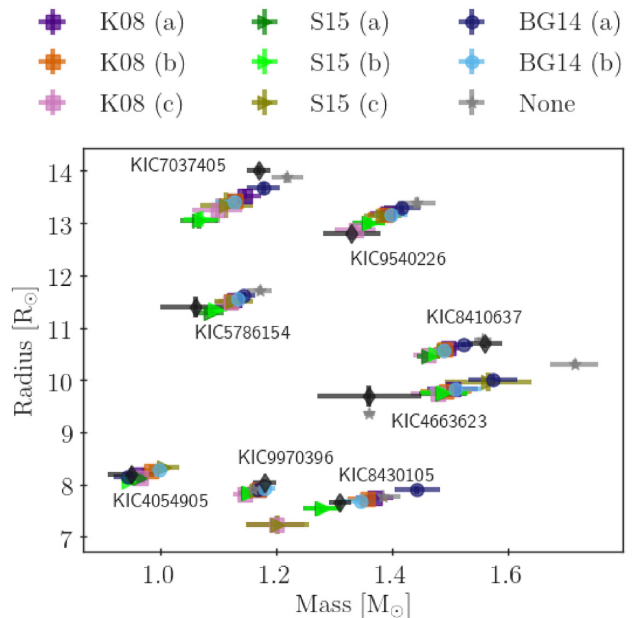


**Figure 1.** Absolute differences between dynamically inferred stellar parameters and the results obtained from stellar models when employing different surface correction relations. The differences are given in units of the standard deviation of the differences, i.e.  $\sigma_M$  and  $\sigma_R$  include both the observational errors and the uncertainty of the inferred model properties. The outlier, for which deviations in both mass and radius exceed  $10\sigma$ , is associated with KIC 4054 905.

well. These results are illustrated in Figs 1 and 2. Here,  $\sigma$  includes the errors on the model parameters that are inferred from the MCMC as well as the errors on the dynamically inferred parameters. In other words,  $\sigma$  denotes the combination of errors that can be derived from the law of propagation of errors.

Furthermore, we find that S15 systematically yields lower mass and radius estimates than both K08 and BG14 do. NoSC and the one-term correction by BG14 (case a) systematically yield higher mass and radius estimates than the other approaches. These systematic trends are also found for the clusters and thus reappear in Sections 5 and 6. Indeed, the analyses of the two clusters show the same relative performance of the different surface correction relations.

There is thus a scatter in the mass and radius estimates that are obtained based on the different surface correction relations. The lowest scatter is found for KIC 9970 396. On first glance,



**Figure 2.** Comparison between dynamically inferred stellar parameters (black markers) and the results obtained from stellar models when employing different surface correction relations. The markers show the location of the best-fitting model, while the error bars correspond to 68 per cent credibility intervals. The likelihood includes  $\nu_{\text{max}}$ . When repeating the analysis using the medians rather than the best-fitting values, we reach the same quantitative and qualitative conclusions. Note that NoSC dramatically overestimates the mass and radius of KIC 4054 905, yielding values that lie closer to KIC 4663 623. This is the outlier in Fig. 1.

Fig. 2 might suggest that a similarly low scatter is obtained for KIC 4054 905. However, in reality, NoSC overestimates the mass by more than  $10\sigma$ , as mentioned above, yielding mass and radius estimates that more closely matches the dynamical constraints on KIC 4663 623 than those on KIC 4054 905.

We partly ascribe the fact that all nine treatments of the surface effect perform similarly well to the included spectroscopic constraints and the constraints on  $\nu_{\text{max}}$ . Thus, when no constraints on  $\nu_{\text{max}}$  are taken into account, a slightly different picture emerges: when equation (1) is not imposed, NoSC performs poorly, overestimating the radius obtained by Gaulme et al. (2016) and Brogaard et al. (2018) by up to a factor of two. The resulting errors in the inferred masses or radii exceed  $5\sigma$  in six out of eight cases – reaching  $60\sigma$  for the radius of KIC 9970 396. Moreover, only the two-term correction by BG14 (case b) does not have one or more outliers, for which the error in either the mass or radius exceeds  $3\sigma$ . Indeed, the two-term correction by BG14 (case b) recover the dynamical constraints on all binaries within  $2.4\sigma$ . For most of the remaining surface correction relations,<sup>4</sup> KIC 8430 105 is one of the outliers. However, we note that the power spectrum for KIC 8430 105 is noisy and that the mode identification might be skewed due to the influence of binarity and magnetic activity (cf. Magic & Weiss 2016; Pérez Hernández et al. 2019, for further discussions on the influence of magnetic activity on stellar oscillations).

In addition, when excluding  $\nu_{\text{max}}$  from the constraints and freely adjusting both parameters related to K08 (case c), the inferred surface effect is larger than the observed large frequency separation

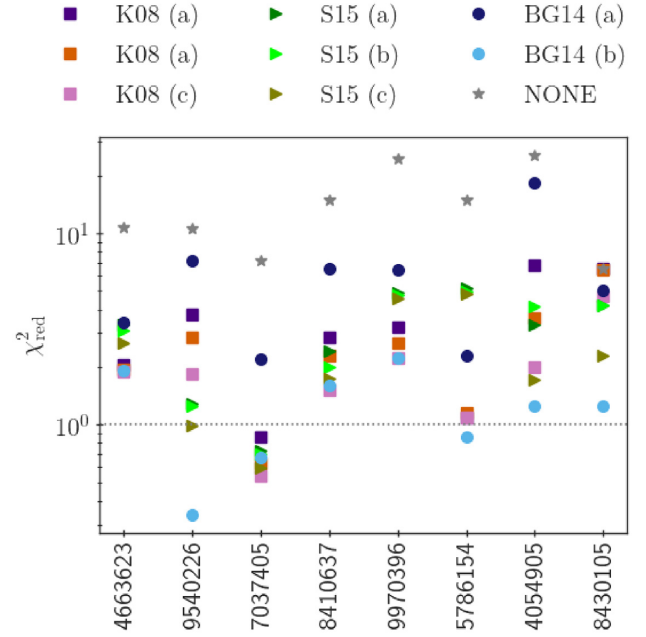
<sup>4</sup>In this particular case, NoSC does very well, recovering the dynamical constraints within  $1\sigma$ .

for KIC 8410 637, KIC 8430 105, and KIC 9540 226. The surface effect is even a substantial fraction of the observed  $\nu_{\max}$ . Both the uncorrected theoretical values for  $\Delta\nu$  and  $\nu_{\max}$  are substantially larger than the observed values. This highly un-physical behaviour of K08 (case c) is likewise obtained for seven out of the nineteen stars that enter the analyses of NGC 6819 in Section 5. We attribute this behaviour to the fact that the physical justification for K08 solely lies in the calibration of the involved parameters. Without sufficient constraints on  $a$  and  $b$  in equation (5), a power-law description of the surface effect becomes unreliable.

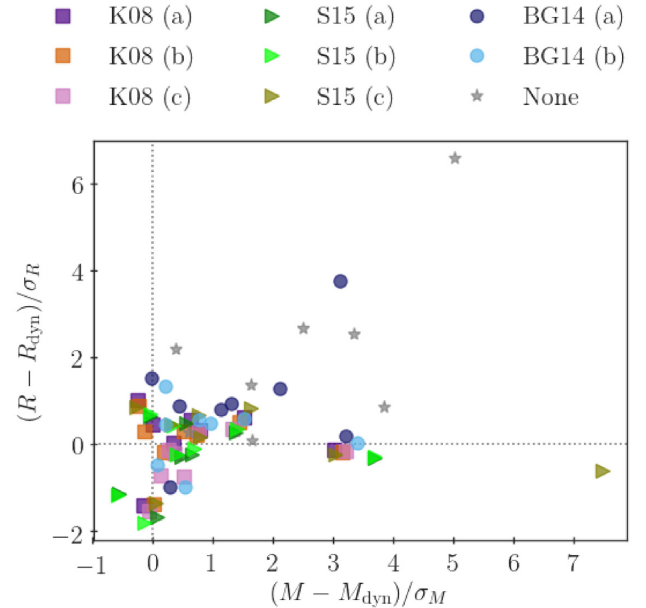
As shown by Ball & Gizon (2017) and Nsamba et al. (2018), the use of solar calibrated values in K08 systematically shifts the stellar parameter estimates when addressing other main-sequence stars. Moreover, calibrating the involved parameters based on 3D MHD simulations as suggested by Sonoi et al. (2015; equation 7) has several caveats: first, current calibrations suffer from the use of the gas  $\Gamma_1$  approximation, as discussed in the introduction. Secondly, the calibration by Sonoi et al. (2015) does not take deviations from solar metallicity into account (cf. Manchon et al. 2018; Jørgensen et al. 2019, who address this issue using 3D MHD simulations). These points generally discourage from the use of K08 and also apply to S15. Nevertheless, disregarding the case of KIC 8430 105, all surface correction relations explored in this paper perform equally well in recovering the global stellar mass and radius of the considered eclipsing binaries. Neglecting the surface effect altogether by not including a surface correction relation, however, does not yield the correct stellar parameters. Indeed, we note that NoSC generally tends to overestimate the stellar mass and radius of the eclipsing binaries, while we have not spotted similarly clear trends for the remaining treatments of the surface effect. To shed more light on this issue, additional data points, i.e. more eclipsing binaries, are needed.

Regarding the goodness of fit, we find several consistent trends. The best-fitting models that were selected using the one-term correction by BG14 (case a) and NoSC lead to higher reduced  $\chi^2$ -values and BICs than any of the other treatments of the surface effect do. For all eight stars, the best-fitting model with the lowest reduced  $\chi^2$ - and BIC-value is found by either using the two-term correction by BG14 (case b) or the free fit based on K08 (case c); the associated scatter is higher for case c of K08. For all investigated binaries, the two-term correction by BG14 (case b) thus results in good fits to data and often yields better fits than the remaining surface correction relations do in terms of the reduced  $\chi^2$ - and BIC-value. We illustrate this in Fig. 3.

Since we are dealing with eclipsing binaries, accurate and precise alternative observational constraints are available. Among other parameters, the stellar radii have been well determined. We have thus repeated the analysis, replacing the constraint on  $\nu_{\max}$  with a constraint on the stellar radius. We find that this greatly reduces the scatter between the mass estimates obtained from the use of different surface correction relations. By including the radius into the observation constraints, we also reduce the mean absolute error in the inferred properties for all treatments of the surface effect. We illustrate this in Figs 4 and 5. Once again, this goes to show the importance of additional (non-seismic) constraints in asteroseismic analyses. The demonstrated improvement that arises from including the radius into the likelihood is, furthermore, of particular interest, since the upcoming third data release (DR3) by *Gaia* (Gaia Collaboration 2016) will provide the community with robust measurements of stellar radii. We note, however, that NoSC still does not perform as well as any of the surface correction relations do, even if we include the radius among the constraints.



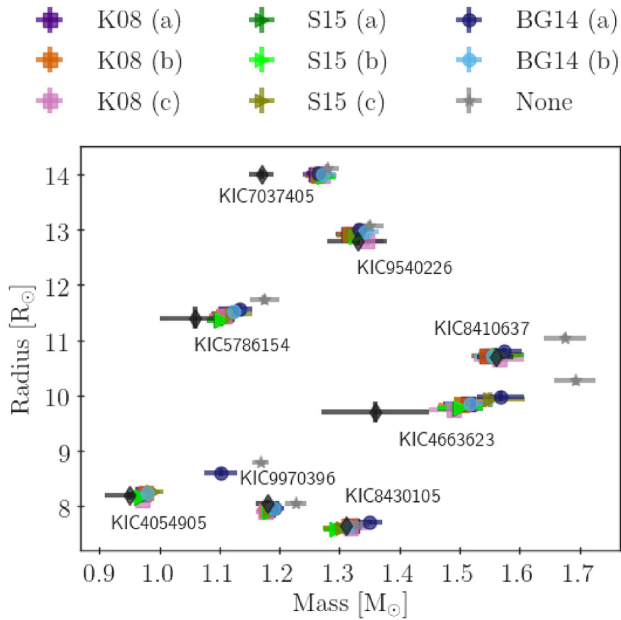
**Figure 3.** Reduced  $\chi^2$ -values for all eight investigated eclipsing binaries (cf. Equation 2). The labels on the abscissa indicate respective the KIC-numbers. The likelihood includes  $\nu_{\max}$ .



**Figure 4.** As Fig. 1, but employing the stellar radius rather than  $\nu_{\max}$  in the likelihood. The accumulation of models, for which  $(M - M_{\text{dyn}})/\sigma_M \approx 3$  and  $(R - R_{\text{dyn}})/\sigma_R \approx 0$ , is associated with KIC 7037 405, for which we use the dynamical constraints by Brogaard et al. (2018). For these models, much lower residuals are obtained when using the dynamical constraints by Gaulme et al. (2016).

This underlines the vital role played by the treatment of the surface effect.

As mentioned above, the employed constraints on the stellar radius reduce the scatter between the inferred masses. To this end, the constraints on the stellar radius do not help us much to further discriminate between the different surface correction relations.



**Figure 5.** As Fig. 2, but employing the stellar radius rather than  $\nu_{\max}$  in the likelihood. Both the one-term correction of BG14 (case a) and NoSC dramatically overestimates the mass KIC 4054 905, yielding a value that lies closer to KIC 9970 396.

On the other hand, this result goes to show that we can easily construct a likelihood that is not dominated by the bias introduced by the employed surface correction relation. In this connection, it is worth highlighting the consequences for the inferred properties of KIC 7037 405. With the likelihood that includes the stellar radius, all surface correction relations unanimously point towards a mass of KIC 7037 405, which is in better agreement with the dynamical measurements by Gaulme et al. (2016;  $1.25 \pm 0.03 M_{\odot}$ ) than with the constraints by Brogaard et al. (2018;  $1.17 \pm 0.02 M_{\odot}$ ), despite the fact that we use the constraints on the radius from Brogaard et al. (2018). We note that this finding is, furthermore, in agreement with other asteroseismic studies (cf. fig. 14 in Buldgen et al. 2019). However, we also note that the discrepancy between our models and the dynamical measurements by Brogaard et al. (2018) might partly come down to the chosen input physics. The models, for instance, follow a fixed relationship between the metal abundance and the helium content (cf. Section 2.1), which might introduce a bias. However, we are partly able to delve into this issue since KIC 7037 405 also lies within the grid that was computed for the analysis of NGC 6791 in Section 6. As mentioned above, this grid employs a different value for  $\Delta Z_i / \Delta Y_i$ . We have thus repeated the analysis using another relationship between  $Z_i$  and  $Y_i$ . We find that we arrive at the same conclusion: our results prefer the dynamical constraints by Gaulme et al. (2016) over those of Brogaard et al. (2018). Moreover, we have repeated the analysis, altering the constraints on  $[\text{Fe}/\text{H}]$ . This likewise leaves the conclusion unaffected.

In so far as we believe that our models recover the correct stellar structures, our analysis of KIC 7037 405 exemplifies how an asteroseismic analysis is able to complement the classical methods irrespective of the employed surface correction relation. On the other hand, we note that surface correction relations do not account for all physical inadequacies of stellar models but only deal with the surface effect and that asteroseismic analyses may naturally fail to deliver accurate stellar parameter estimates if the underlying

stellar models are incorrect (see e.g. Jørgensen et al. 2019; Rendle et al. 2019, for a comparison of models that are based on different physical assumptions).

It is worth mentioning that the best-fitting models, as well as the mean of the associated posterior distributions, tend to systematically underestimate the measured effective temperatures for all binaries. Thus, the mean absolute deviation between measurements and model predictions exceeds 100 K in all cases. This holds true whether or not we introduce  $\nu_{\max}$  or the radius as an additional observational constraint.

As regards the goodness of fit, we obtain slightly higher values for both the reduced  $\chi^2$  and the BIC when including the radius in the likelihood than shown in Fig. 3. While an independent constraint on the stellar radius thus improves the agreement of the predicted stellar mass with dynamical measurements, it leads to a slightly worse recovery of the individual frequencies than when the individual frequencies dominate the likelihood.

To further discriminate between the different treatments of the surface effect, more samples are needed. We therefore turn to an analysis of the open clusters NGC 6819 and NGC 6791 below.

## 5 NGC 6819

Using AIMS and employing all nine treatments of the surface effect listed in Table 1, we have derived stellar parameters for 19 RGB stars in the open cluster NGC 6819.<sup>5</sup> We have adopted non-seismic constraints from Handberg et al. (2017) and set  $[\text{Fe}/\text{H}] = 0.02 \pm 0.1$  dex for all cluster members.

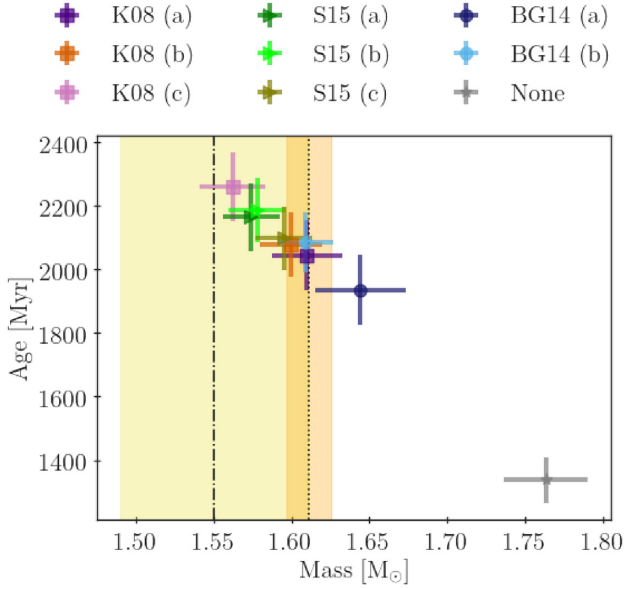
We computed the mean mass of the 19 RGB stars based on each of the nine treatments of the surface effect. The results are summarized in Fig. 6 and in Table 3. When repeating the analysis without including  $\nu_{\max}$  into the likelihood, all treatments lead to similar mean masses and mass scatter to those obtained with  $\nu_{\max}$ . We note that the presented averages are taken over the best-fitting values and weighted by the respective variance of the associated posterior probability distribution. We also note that we reach the same quantitative and qualitative conclusions if we repeat the analysis using the medians rather than the best-fitting values. The applicability of this procedure rests on several assumptions, including the notion that the posterior probability distribution of the stellar mass is well-described by a Gaussian. While this is a good approximation for the vast majority of the stars presented in this paper, we note that cases exist for which the posterior mass distribution is less symmetric or even multimodal.

Thus, the mean mass of the cluster is estimated to be

$$\bar{M} = \frac{\sum_i^N M_i / \sigma_i^2}{\sum_i^N 1 / \sigma_i^2}, \quad (12)$$

where  $M_i$  and  $\sigma_i$  denote the best-fitting value of the mass and the corresponding estimate of the standard deviation for each of the  $N$  samples. Due to the low number of samples, the uncertainty on the mean has been computed by following the procedure outlined by Chaplin et al. (1998; see also Miglio et al. 2012). When dealing with small sample sizes, the estimate of the standard deviation that follows from the sample might be biased and often underestimates the standard deviation of the underlying normally distributed

<sup>5</sup>KIC 4937 576, KIC 5023 732, KIC 5023 931, KIC 5024 240, KIC 5024 297, KIC 5024 312, KIC 5024 405, KIC 5024 512, KIC 5024 583, KIC 5111 718, KIC 5111 940, KIC 5112 072, KIC 5112 361, KIC 5112 734, KIC 5112 744, KIC 5112 880, KIC 5112 948, KIC 5113 041, and KIC 5113 441.



**Figure 6.** The mean mass and age obtained from nine different treatments of the surface effect using equations (12) and (13) for NGC 6819. The orange shaded area indicates to the 68 per cent credibility interval of the asteroseismic constraints by Handberg et al. (2017). The corresponding mean is marked by the dotted black line. We note that the constraints obtained by Handberg et al. (2017) are likewise obtained from an asteroseismic study. Classical measurements by Sandquist et al. (2013) and Brogaard et al. (2015) lead to a broader confidence interval that agrees well with the results obtained based on all surface correction relations ( $1.55 \pm 0.06 M_{\odot}$ ). These constraints are shown with the shaded yellow area. The corresponding mean is indicated with the dash-dotted line. The likelihood includes  $\nu_{\max}$ .

**Table 3.** Summary of the inferred mean mass and age ( $\tau$ ) of the RGB stars in NGC 6819 based on the different surface correction relations that are investigated in this paper. The likelihood includes  $\nu_{\max}$ .

Surf. corr.	$\langle M \rangle \pm \sigma_{\bar{M}}$	$\langle \tau \rangle \pm \sigma_{\bar{\tau}}$ [Myr]
K08 (a)	$1.610 \pm 0.023$	$2045 \pm 110$
K08 (b)	$1.600 \pm 0.020$	$2077 \pm 102$
K08 (c)	$1.562 \pm 0.021$	$2260 \pm 108$
S15 (a)	$1.574 \pm 0.018$	$2164 \pm 106$
S15 (b)	$1.578 \pm 0.0182$	$2187 \pm 103$
S15 (c)	$1.595 \pm 0.018$	$2098 \pm 98$
BG14 (a)	$1.644 \pm 0.029$	$1935 \pm 110$
BG14 (b)	$1.609 \pm 0.018$	$2084 \pm 94$
NoSC (None)	$1.763 \pm 0.027$	$1337 \pm 70$

population. To account for this, we compute the uncertainty from the unbiased estimate of the variance from the sample and apply a correction factor  $t[N-1]$  drawn from the Student's  $t$  distribution with  $N-1$  degrees of freedom:

$$\sigma_{\bar{M}} = t[N-1] \sqrt{\frac{\sum_i^N (M_i - \bar{M})^2 / \sigma_i^2}{(N-1) \sum_i^N 1/\sigma_i^2}}. \quad (13)$$

In the present case, the correction factor, i.e.  $t[N-1]$ , is of the order of unity, largely leaving the uncertainties unaltered. We note that equation (13) yields larger uncertainties than what would be obtained from the law of propagation of error based on the uncertainties on the samples alone. This implies that the scatter of the samples are not dominated by random errors, which might

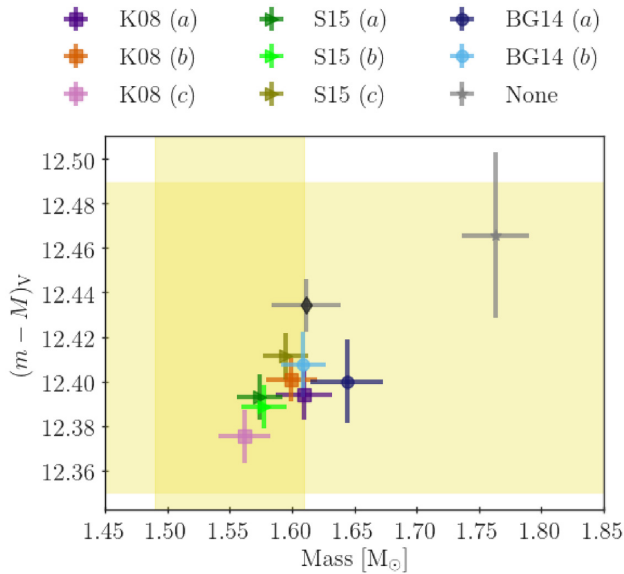
imply that AIMS underestimates the true errors on the sampled stellar parameters – at least, within the Gaussian approximation. This circumstance might reflect the fact that the grid is 2D, as additional free parameters might lead to broader posterior distributions of the sampled properties. As specified in Section 2.1, we thus use a fixed mixing length parameter, fixed parameters associated with over and undershooting, and a fixed relation between the initial helium abundance and the initial abundance of heavy elements.

Based on dynamical measurements of eclipsing binaries by Sandquist et al. (2013), Brogaard et al. (2015) determine the mean mass of the RGB stars in the NGC 6819 to be  $1.55 \pm 0.06 M_{\odot}$ . Using asteroseismic scaling relations and empirical corrections for  $\Delta\nu$  and  $\nu_{\max}$ , Handberg et al. (2017) finds the mean mass of RGB stars in the cluster to be  $1.61 \pm 0.02 M_{\odot}$  (see also e.g. Miglio et al. 2012). We note that only the models for which no surface correction relation is taken into account fail to fall within one standard deviation of the mean evaluated by either Brogaard et al. (2015) or Handberg et al. (2017). This holds true with and without including  $\nu_{\max}$  into the likelihood.

This being said, such a direct comparison with the mean cluster masses of RGB stars from Brogaard et al. (2015) and Handberg et al. (2017) might be slightly skewed since our analysis only includes 19 stars. Our results may thus be subject to a selection bias. Instead, we note that Handberg et al. (2017) also supply mass estimates for the individual members of the cluster, yielding a weighted mean mass of  $1.61 \pm 0.01 M_{\odot}$  for the considered 19 giants. With the exception of NoSC, we find that all treatments of the surface effect yield parameter estimates that agree with this result within  $2\sigma$  whether or not we include  $\nu_{\max}$  in the likelihood. Moreover, on a star by star basis, all models recover the observational constraints on the effective temperature within 100 K irrespective of the employed surface correction relation. Again, this statement does not apply to NoSC but does remain valid whether or not we include  $\nu_{\max}$  in the likelihood. We, furthermore, note that we find no obvious trends between the inferred mass of the individual cluster members and the associated value of  $\nu_{\max}$  when including the latter in the likelihood. This holds true for all nine treatments of the surface effect, and it also holds true for the stars in NGC 6791. Moreover, we find the mass scatter to be seemingly uniform as a function of  $\nu_{\max}$ , i.e. along the RGB.

As can be seen from Fig. 6, the different treatments of the surface effect that employ S15 tend to lead to lower average masses than the two-term surface correction relation by BG14 does. This is consistent with our analysis of the eclipsing binaries, where S15 is likewise found to underestimate the classical constraints on the stellar masses. According to a differential study by Nsamba et al. (2018), this behaviour is not found for main-sequence stars, where S15 and BG14 on average yield equally robust mass and age estimates. The fact that the cases a and b of S15 perform better for main-sequence stars than for more evolved stars presumably reflects the fact that the involved parameters are calibrated based on a study that primarily includes patched models of main-sequence stars. Thus, Sonoi et al. (2015) solely include one red giant branch star in their calibration sample.

As regards S15 we, furthermore, note that we have excluded KIC 5112 880 from the sample when treating  $\alpha$  and  $\beta$  as free parameters (case c). This is because case c of S15 yields an age estimate for this star that exceeded the age of the Universe by a factor of two. Once again, this behaviour might reflect the fact that there is no physical justification for S15 beyond the calibrated surface correction relation parameters.



**Figure 7.** The mean mass and distance modulus obtained from nine different treatments of the surface effect for NGC 6819. The error bars on the distance modulus correspond to the sample standard deviation and thus reflect the scatter of the individual distance moduli. The black marker summarizes the results in Handberg et al. (2017) based on the same 19 stars. The yellow shaded area corresponds to the 68 per cent credibility interval for the classical constraints on the mass and distance modulus presented by Sandquist et al. (2013) and Brogaard et al. (2015). The likelihood includes  $\nu_{\max}$ .

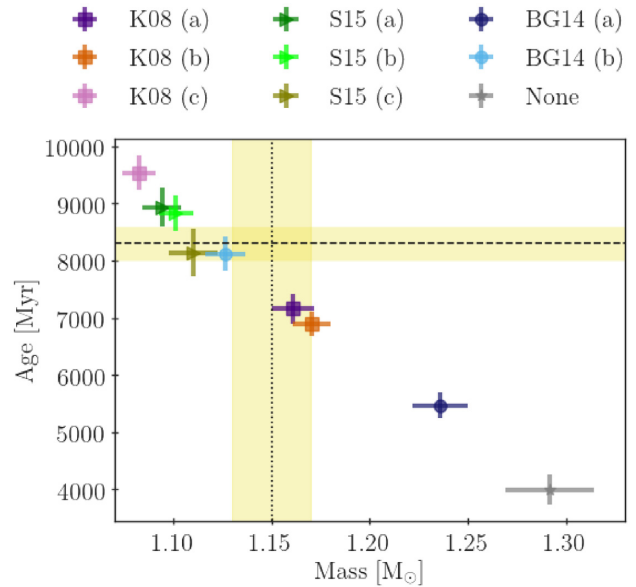
From Fig. 6, it is also apparent that the discrepancies in the estimated mean stellar mass translate into substantial systematic offsets in the estimated mean age of the stars: excluding the case where the surface effect is left unaccounted for the choice for the treatment of the surface effect alone affects the estimated age of the cluster by up to 17 per cent. As discussed in Section 2.2, tighter non-seismic constraints might lift degeneracies and hereby partly mend but not eliminate this issue. The obtained systematic offsets thus illustrate the vital role that a proper understanding of the surface effect, i.e. superadiabatic convection plays for galactic archaeology.

### 5.1 Distance modulus

To investigate how the treatment of the surface effect alters other inferred physical properties of the cluster, we have computed the apparent distance modulus,  $(m - M)_V$ , of NGC 6819 based on the prescription by Torres (2010). For this purpose, we use the  $V$ -band magnitudes by Milliman et al. (2014) in combination with the bolometric corrections by Casagrande & VandenBerg (2014).

In Fig. 7, we compare the obtained results with classical measurements by Sandquist et al. (2013) and Brogaard et al. (2015), according to whom the apparent distance modulus and mean mass of the RGB stars in the cluster are  $12.42 \pm 0.07$  and  $1.55 \pm 0.06 M_{\odot}$ , respectively. As can be seen from the figure, the influence of the treatment of the surface effect on the distance modulus is as pronounced as its influence on the age and mass estimates. Indeed, the spread of obtained mean apparent distance moduli are similar to the internal scatter.

We find that all surface correction relations recover the mean distance modulus of the cluster within  $1 \sigma$  whether or not we include  $\nu_{\max}$  in the likelihood.



**Figure 8.** As Fig. 6, but for NGC 6791. Here, we compare with the results listed in Brogaard et al. (2012). The horizontal dashed line and the yellow shaded horizontal area indicate an age estimate based on isochrones.

Based on the same 19 giants, Handberg et al. (2017) finds the apparent distance modulus of NGC 6819 to be  $12.43 \pm 0.01$ , which likewise agrees with the results presented by Sandquist et al. (2013) and Brogaard et al. (2015) within  $1 \sigma$ . Notably, all the investigated treatments of the surface effect, furthermore, recover the quoted distance modulus by Handberg et al. (2017) within  $2 \sigma$  – again this holds true with and without constraints on  $\nu_{\max}$ .

On a star by star basis, we find the mass to be correlated with the attributed distance modulus. We thus find that the stars that have been assigned a lower mass are also assigned a lower distance modulus. We ascribe this behaviour to the correlation between seismically inferred masses and radii. The described correlation between the mass and distance modulus is observed with and without constraints on  $\nu_{\max}$ . We address this issue further in Section 6, where we deal with NGC 6791.

## 6 NGC 6791

In this section, we present an analysis of 30 red giants<sup>6</sup> in the open cluster NGC 6791 to further validate and extend upon the conclusions drawn in Section 5. Based on Brogaard et al. (2012), we set  $[\text{Fe}/\text{H}] = 0.29 \pm 0.1$  dex. Non-seismic constraints were adopted from Basu et al. (2011).

As for NGC 6819, we have computed the mean mass for the cluster based on the output from AIMS for all 30 considered members. We have repeated this for all nine surface correction relations in Table 1. The results are summarized in Fig. 8 as well as in Table 4. The figure includes an estimate for the absolute mass of stars on the lower red giant branch by Brogaard et al. (2012);

<sup>6</sup>KIC 2435 987, KIC 2436 688, KIC 2437 270, KIC 2437 402, KIC 2437 972, KIC 2438 140, KIC 2569 618, KIC 2570 244, KIC 2436 097, KIC 2437 653, KIC 2437 933, KIC 2437 976, KIC 2438 333, KIC 2570 094, KIC 2436 540, KIC 2436 818, KIC 2437 240, KIC 2437 488, KIC 2437 957, KIC 2438 038, KIC 2570 172, KIC 2569 360, KIC 2437 816, KIC 2436 824, KIC 2436 814, KIC 2437 444, KIC 2437 507, KIC 2436 900, KIC 2436 209, and KIC 2436 332.

**Table 4.** Summary of the inferred mean mass and age ( $\tau$ ) of the RGB stars in NGC 6791 based on the different surface correction relations that are investigated in this paper. The likelihood includes  $\nu_{\max}$ .

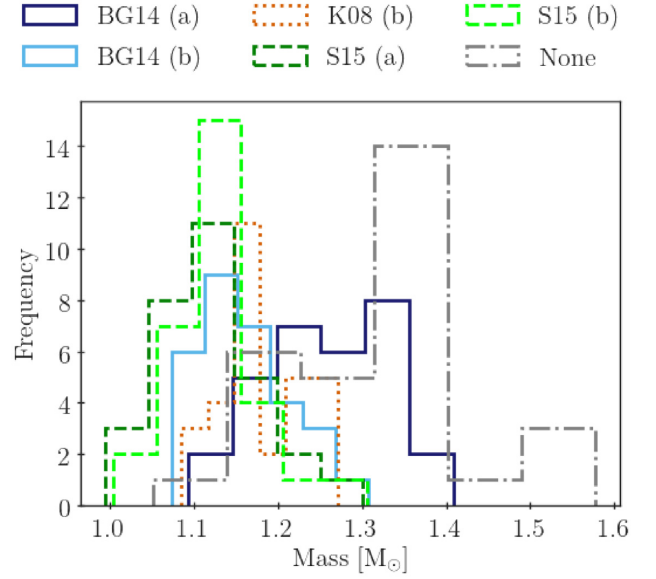
Surf. corr.	$\langle M \rangle \pm \sigma_M$	$\langle \tau \rangle \pm \sigma_\tau$ [Myr]
K08 (a)	$1.161 \pm 0.011$	$7157 \pm 254$
K08 (b)	$1.170 \pm 0.010$	$6893 \pm 219$
K08 (c)	$1.082 \pm 0.008$	$9530 \pm 300$
S15 (a)	$1.094 \pm 0.010$	$8930 \pm 340$
S15 (b)	$1.101 \pm 0.009$	$8834 \pm 305$
S15 (c)	$1.110 \pm 0.012$	$8141 \pm 411$
BG14 (a)	$1.236 \pm 0.014$	$5466 \pm 227$
BG14 (b)	$1.126 \pm 0.010$	$8122 \pm 300$
NoSC (None)	$1.292 \pm 0.023$	$4000 \pm 252$

$1.15 \pm 0.02 M_\odot$ ) as well as an age estimate ( $8.3 \pm 0.3$  Gyr) based on isochrones from the same paper. While the results by Brogaard et al. (2012) are based on an analysis of eclipsing binaries, we note that the cited age estimate is model dependent. As discussed in Section 3.2 of the paper by Brogaard et al. (2012), the assumed abundances of different elements and the treatment of heavy element diffusion lead to additional systematic and statistical errors on the age estimate. Analogous to the case of NGC 6819 we, furthermore, note that the mass estimate from our analysis might be subject to a selection bias. After all, we only consider 30 stars. We also note that analyses by other authors yield slightly different mass estimates for the red giant branch stars in the cluster. This includes asteroseismic analyses by Basu et al. (2011) and Miglio et al. (2012) – that is,  $1.20 \pm 0.01 M_\odot$  and  $1.23 \pm 0.02 M_\odot$ , respectively. This being said, both Basu et al. (2011) and Miglio et al. (2012) have computed their mass estimates directly from the scaling relations. Going beyond such scaling relations by using stellar models would shift the mass estimates and potentially improve the agreement with the values obtained from eclipsing binaries by Brogaard et al. (2012). Thus, the asteroseismic analysis by McKeever et al. (2019) based on stellar models yields a mean mass estimate for the red giant branch stars in the cluster of  $1.15 \pm 0.01 M_\odot$ , which closely matches that by Brogaard et al. (2012). This latter analysis notably also employs the two-term correction by BG14 (case b) but includes higher degree modes ( $\ell = 2$ ).

In short, the statistical errors of the reference age in Fig. 8 might be underestimated, while the reference mass might be subject to systematic errors. Taking these circumstances into account, we note that the majority of the investigated surface correction relations lead to mass and age estimates that agree reasonably well with the literature. The case without any surface correction is clearly an example of an exception to this statement.

By comparing Fig. 8 to Fig. 6 in Section 5, we note that the relative performance of the different surface correction relations is the same for both clusters. In other words, the ordering from the lowest to the highest mean mass or age is the same for both clusters. For instance, NoSC yields much higher masses and much lower ages and the remaining eight treatments of the surface effect do. The same conclusion was drawn from the analysis of the eclipsing binaries in Section 4. This underlines that the use of improper surface correction relations does, indeed, lead to systematic errors in the obtained global parameters.

As can, furthermore, be seen from the error bars in Fig. 8, the one-term correction by BG14 (case a) and the case without any surface correction relation yield broader mass distributions than the remaining seven treatments of the surface effect do – the same is



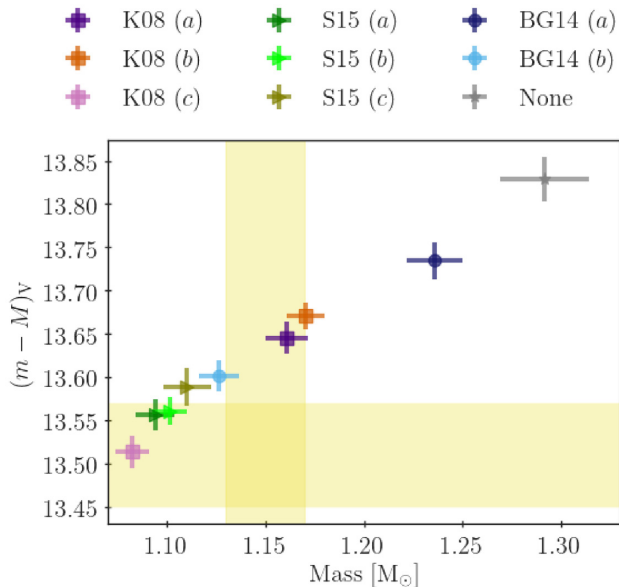
**Figure 9.** Mass distributions for six of the nine treatments of the surface effect discussed in this paper for NGC 6791. The remaining three approaches were excluded to avoid that the plot became too crowded. For each histogram, we employ six bins. These bins are therefore broader in mass for the cases with larger mass scatter. The likelihood includes  $\nu_{\max}$ .

found for NGC 6819, although to a lesser extent. Since we are dealing with a cluster, this larger mass scatter indicates that the one-term correction by BG14 (case a) and the case without any surface correction relation are less reliable than the other approaches. We further illustrate this in Fig. 9.

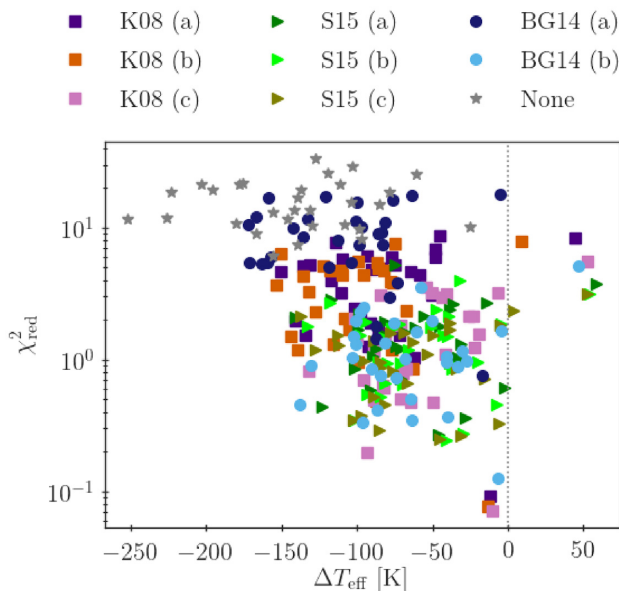
As in the case of NGC 6819, we have computed the distance modulus of the cluster. For this purpose, we used V-band magnitudes by Basu et al. (2011) in combination with the bolometric corrections by Casagrande & VandenBerg (2014). With the exceptions of the one-term correction by BG14 (case a), cases a and b of K08, and NoSC, all treatments of the surface effect agree within  $2\sigma$  with the value obtained from the study of eclipsing binaries by Brogaard et al. (2012) ( $13.51 \pm 0.06$ ). We quantify this result in Fig. 10. Once again, we note that the one-term correction by BG14 (case a) and NoSC perform worse than the remaining seven procedures. This notion is consistent with the results obtained in both Sections 4 and 5. Only the two-term correction by BG14 recovers the stellar mass within  $1\sigma$  in addition to yielding a mean distance modulus that is consistent with Brogaard et al. (2012) within  $2\sigma$ .

While the inferred effective temperatures from AIMS closely match spectroscopic constraints in the case of NGC 6819, the inferred effective temperatures are too high for most cluster members in NGC 6791 irrespective of how we treat the surface effect. Again, we find that the one-term correction by BG14 and NoSC lead to the worst goodness-of-fit scores, indicating that these approaches do not recover observations well. Meanwhile, in several cases, the other surface corrections lead to slight overfitting. We illustrate and quantify the discussed features in Fig. 11.

On a star by star basis, we find the mass estimates to be correlated with the attributed distance modulus, as mentioned in Section 5. We, furthermore, find the mass to be correlated with the offset between the effective temperature of the model and the spectroscopic constraint on  $T_{\text{eff}}$ . We thus find stars that are assigned a lower mass to yield a higher offset in  $T_{\text{eff}}$ . There is, meanwhile, no clear correlation between the offset in  $T_{\text{eff}}$  and the goodness of



**Figure 10.** As Fig. 7, but for NGC 6791. Here, we compare with the mass interval and the constraints on the distance modulus provided by Brogaard et al. (2012).



**Figure 11.** The reduced  $\chi^2$  for all stars in NGC 6791 as a function of the deviation between the measured and predicted effective temperature. Positive values for the temperature difference indicate that the spectroscopic constraint on the effective temperature is higher than the inferred value. The likelihood includes  $\nu_{\max}$ .

fit, if we solely consider models that share the same treatment of the surface correction relation. Overall, however, such a correlation exists, since the surface correction relations that yield the highest masses also lead to higher  $\chi^2$ - and BIC-values – that is, case a of BG14 and NoSC.

In this connection, we note that the evaluated  $T_{\text{eff}}$  is sensitive to the constraint on  $[\text{Fe}/\text{H}]$ . The inferred values of  $T_{\text{eff}}$  and thereby of the distance moduli can thus be improved by changing the

imposed metallicity of the cluster: changing the constraint on  $[\text{Fe}/\text{H}]$  by 0.1 dex increases the inferred temperature by roughly 50 K. However, we note that the mass and age estimates are likewise sensitive to this observational constraint. We can thus increase the mean mass and decrease the mean age estimates by assuming a systematically lower metallicity for the cluster members. An increase in the imposed metallicity by 0.1 dex thus decreases the mean mass estimate by roughly  $0.02 M_{\odot}$  for all surface correction relations. In other words, in the presented scenario, the improvement of the distance modulus that is achieved by increasing the metallicity comes at the cost of lowering the mean mass, which for most surface correction relations increases the tension with the dynamical constraints by Brogaard et al. (2012).

In the same manner, as regards the distance modulus, we can improve the agreement with Brogaard et al. (2012) by including alpha enrichment for all models presented in this paper,  $[\alpha/\text{Fe}] = 0.0$  (cf. Section 2.1). We thus repeated the analysis adopting  $[\alpha/\text{Fe}] = 0.1$ . In this case, we find that all treatments of the surface effect but the one-term correction by BG14 (case a) and NoSC recover the mean distance modulus by Brogaard et al. (2012) within  $2\sigma$ . However, once again, this comes at the cost of lower mean masses and higher mean ages. The mean mass thus decreases by roughly  $0.01$ – $0.03 M_{\odot}$ .

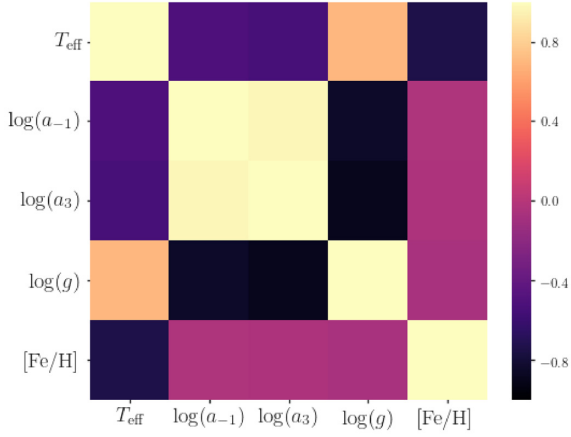
The effective temperature can also be shifted by choosing a different model atmosphere. To investigate this, we have repeated the analysis, substituting the semi-empirical model atmospheres by Vernazza, Avrett & Loeser (1981) with those by Krishna Swamy (1966). Doing so, we find that the offset in  $T_{\text{eff}}$  increases by roughly 50 K, leading to even higher model temperatures. As a result, none of the investigated treatments of the surface effect yields mean distance moduli that lie within  $2\sigma$  of the constraints by Brogaard et al. (2012) when employing the model atmospheres by Krishna Swamy (1966). Moreover, when using the model atmospheres by Krishna Swamy (1966), only the cases a and b of S15 and case c of K08 lie within  $1\sigma$  of the mass estimate by Brogaard et al. (2012).

## 7 DISCUSSION

Throughout the presented analysis of the eclipsing binaries, NGC 6819 and NGC 6791, the two-term correction by BG14 performs well. As argued by Ball & Gizon (2014) based on Gough (1990), the underlying functional form is physically motivated, which may explain the success of this approach. Encouragingly, we furthermore find that the best-fitting values of both parameters involved in BG14 ( $a_1$  and  $a_3$ ) are strongly correlated with the global stellar parameters, predominantly  $\log g$ , of the best-fitting models. Thus, the inferred surface effect evolves in a predictable manner through the HR diagram. This is illustrated in Fig. 12.

As regards K08 and S15, we likewise find the involved parameters ( $a$ ,  $b$ ,  $\alpha$ , and  $\beta$ ) to be correlated with the global stellar parameters, when  $a$ ,  $b$ ,  $\alpha$ , and  $\beta$  are kept free.

We note that the discussed correlations partly reflect the intrinsic correlations between the global stellar parameters, say, between  $T_{\text{eff}}$  and  $\log g$ . These intrinsic correlations come about as red giant branch stars in any given cluster will share similar properties – indeed, this notion underlies the analysis presented in the previous sections. In other words, the intrinsic correlations reflect a selection bias. Notably, repeating the analysis solely based on a single cluster, the results occasionally point towards correlations between the free surface correction relation parameters and  $[\text{Fe}/\text{H}]$ . If this result were to be genuine, it would be quite intriguing, since such correlations

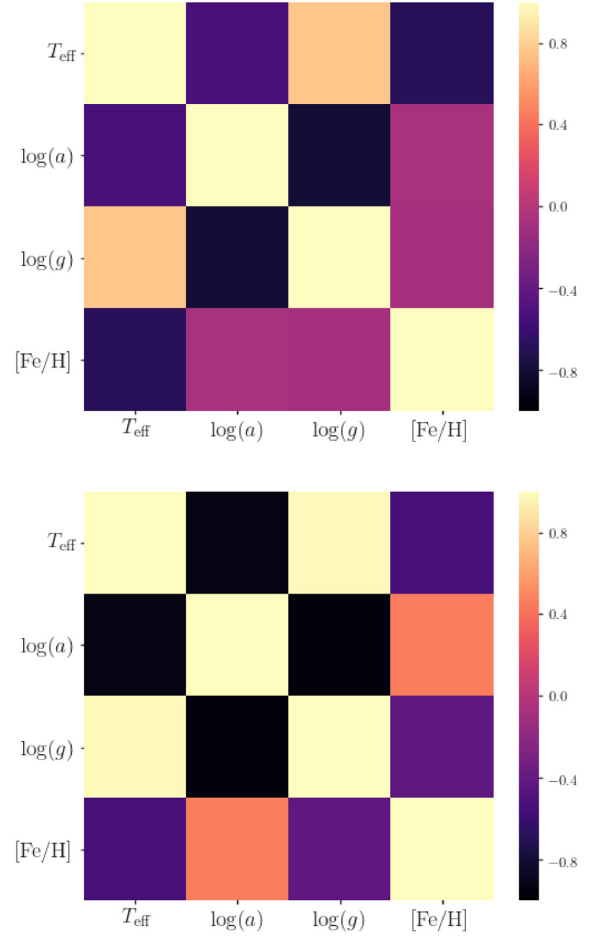


**Figure 12.** Correlations between the surface correction relation parameters of BG14 (case b) and the global stellar parameters of the stars in NGC 6819 and NGC 6791 in combination with the eight eclipsing binaries in Section 4. The correlation matrix has been computed based on the best-fitting models of each star. The analysis thus includes 57 red giant branch stars. The likelihood includes  $\nu_{\max}$ .

are not taken into account by equations (6), (7), (9), and (10). However, these correlations are washed away when including more data. This is exemplified in Fig. 13: based on NGC 6819 alone,  $a$  seems to be strongly correlated with  $\log g$ ,  $T_{\text{eff}}$ , and  $[\text{Fe}/\text{H}]$  when considering case b of K08 (lower panel). The picture changes when including an additional cluster and the eclipsing binaries (upper panel). Weaker correlations with  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$  are obtained. This is not to say that the metallicity might not play a role for the parameters involved in the surface correlation relations. However, while correlations between the surface correction relation parameters and the global stellar parameters might express dependencies on these global parameters, these correlations also indirectly express a selection bias.

Moreover, our results suggest that equations (6), (9), and (10) systematically underestimate the best-fitting values of  $a$ ,  $\alpha$ , and  $\beta$ , while equation (7) overestimates  $b$ . This was to be anticipated as Sonoi et al. (2015) have calibrated these relations mainly based on main-sequence stars for which  $T_{\text{eff}}$  lies above 6000 K. The population that is studied in the present paper thus strongly differs from that based on which Sonoi et al. (2015) derived their expressions for  $a$ ,  $b$ ,  $\alpha$ , and  $\beta$ .

In order to further quantify these statements, one may attempt to calibrate the parameters in K08 and S15 based on a large sample of well-constrained stars, such as eclipsing binaries. The use of such benchmark targets allows adjusting the parameters based on well-established non-seismic constraints (cf. Fig. 5). However, since the obtained values may, to a large extent, reflect other input physics, such as the chosen  $T(\tau)$  relation, and since there is no physical justification for the functional form of K08 and S15, a calibration of these surface correction relations is not guaranteed to be widely applicable. Indeed, as discussed by Jørgensen et al. (2019), any attempt to establish a global calibration of K08 and S15 may be subject to a selection bias. In other words, the parameters of surface correction relations encode physical inadequacies of the stellar models, and they thus reflect the associated input physics as well as the global stellar parameters, since the mentioned inadequacies are sensitive to these properties. It thus stands to reason that a proper calibration of K08 and S15 must be performed based on a suitable set of



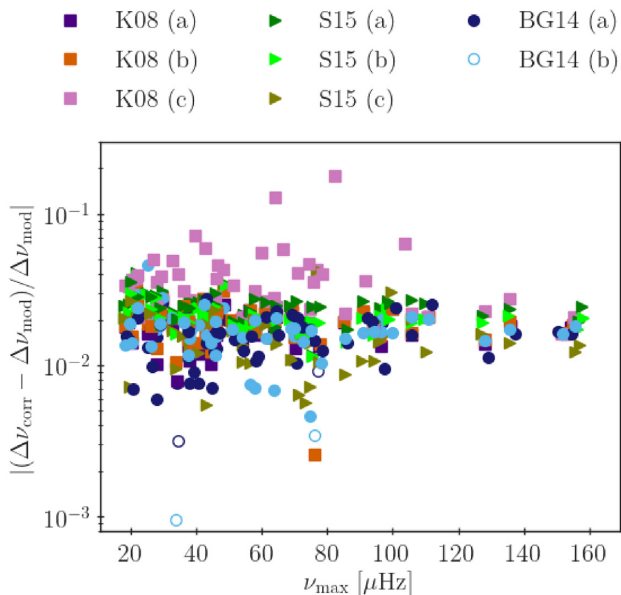
**Figure 13.** As Fig. 12, but for K08 (b). The upper panel is based on the global stellar parameters of the stars in NGC 6819 and NGC 6791 in combination with the eight eclipsing binaries in Section 4. For the lower panel, we have only included the stars from NGC 6819.

benchmark stars every time the input physics is significantly altered, and every time a yet uncovered region of the HR diagram is explored.

While we do not dive further into the coefficients and exponents that enter the surface correction relations, it is worth taking a closer look at the consequences of using the surface correction relations for the predicted physical stellar properties. We thus note that the relative shift in the large frequency separation is constant across all values of  $\nu_{\max}$ . This is illustrated in Fig. 14 (see also Rodrigues et al. 2017). For nearly all cases, the large separation is lower when including a correction of the surface effect.<sup>7</sup>

Case c of K08 constitutes a notable exception to this rule and deviates by one order of magnitude in the size of the relative shift in  $\Delta\nu$  from the other surface correction relations. The fact that this correction significantly differs in the predicted surface effect is consistent with the notion that case c of K08 leads to lower mass estimates than the remaining surface correction relations do, including the cases a and b of K08, as shown in Figs 6 and 8.

<sup>7</sup>The very few exceptions all employ the one- or two-term correction by BG14. Note that we indirectly enforced this behaviour for K08 and S15 by imposing priors on the involved parameters (cf. Table 1).



**Figure 14.** Relative difference in the large frequency separation between that obtained from corrected frequencies ( $\Delta\nu_{\text{corr}}$ ) and that obtained from uncorrected frequencies ( $\Delta\nu_{\text{mod}}$ ). The plot is based on all 57 red giant branch stars presented in this paper. One outlier with a relative difference below  $10^{-3}$  were removed from the plot for clarity – it used case a of K08. We note that  $\Delta\nu_{\text{corr}} < \Delta\nu_{\text{mod}}$  in nearly all cases. Exceptions only occur for BG14 and have been indicated with open symbols. The averages constitutes unweighted means taken over the model frequencies that correspond to the observed modes.

## 8 CONCLUSION

In this paper, we perform asteroseismic analyses of several red giant branch stars imposing different surface correction relations (cf. Kjeldsen et al. 2008; Ball & Gizon 2014; Sonoi et al. 2015). We demonstrate that the use of each of these surface correction relations biases the inferred stellar properties. In accordance with the studies of main-sequence stars by Ball & Gizon (2017) and Nsamba et al. (2018), our results thus show that the use of different surface correction relations leads to systematic offsets in the derived global stellar properties, including stellar ages, masses, and distances.

In Section 4, we present an asteroseismic analysis of eight giants that are situated in eclipsing binaries. From comparisons with the results of dynamical studies, we conclude that we are able to recover the global stellar parameters equally well independent of the employed surface correction relation, if we impose sufficiently informative non-seismic constraints. Even with such non-seismic constraints, however, there is a scatter between the inferred stellar properties that are favoured by the different surface correction relations. Throughout most of this paper, we ensure that the constraints on the individual frequencies dominate the likelihood in order to discriminate between the different surface correction relations. Meanwhile, we also demonstrate how the use of different additional constraints that complement the individual frequencies reduces the impact of the chosen surface correction relation. When including dynamical constraints on the stellar radius or constraints on  $\nu_{\text{max}}$ , most treatments of the surface effect match the dynamical constraints on the masses of the binary members within  $2\sigma$  and  $3\sigma$ , respectively. Without any of these constraints, however, only the two-term correction by Ball & Gizon (2014) is able to recover the dynamical measurements of the stellar masses

and radii for all stars in the sample within  $3\sigma$ . Especially, the approach of ignoring the surface effect altogether yields poor results. Moreover, in several cases, the power-law description by Kjeldsen et al. (2008) leads to un-physically large frequency offsets in the un-corrected frequencies. We attribute this behaviour to the fact that there is no physical justification for the power-law correction beyond the calibration of the involved coefficient and exponent.

Based on our analysis of eclipsing binaries, we thus generally recommend including available non-seismic constraints in asteroseismic analyses to lower any systematic errors that might be introduced by the chosen treatment of the surface effect. Surveys and missions, such as Gaia, provide such constraints.

In Sections 5 and 6, we address a large sample of red giant branch stars from two open clusters and investigate the attributed mean properties of the populations. Through this analysis, we illustrate that the use of different surface correction relations biases mass, age, and distance estimates for the cluster members. In other words, improper treatments of the surface effect lead to systematic errors in the obtained global parameters. In this connection, we note that the different surface correction relations perform in the same manner relative to each other for both clusters. For instance, for both clusters, the one-term correction relation by Ball & Gizon (2014) yields higher mean mass and lower mean age estimates than the remaining surface correction relations do. Furthermore, when using the surface correction relation from Sonoi et al. (2015), one on average obtains lower mass estimates than when employing the two-term correction by Ball & Gizon (2014). We also note that the obtained results are inconsistent with existing constraints on the mass and age of the clusters when no surface correction is imposed. These conclusions are consistent with our analysis of the eclipsing binaries in Section 4.

Finally, we illustrate that the attempt to calibrate a surface correction relation based on a sample of target stars introduces correlations in the derived coefficients and exponents, reflecting the underlying sample. We note that this agrees with the results obtained by Jørgensen et al. (2019), who show that the underlying sample leads to a selection bias. In combination with the results presented by Jørgensen et al. (2019), our results thus discourage from the use of calibrated surface correction relations in connection with targets that do not closely resemble the stars, based on which the relations were calibrated. Overall, our results are thus not in favour of using the surface correction relations by Kjeldsen et al. (2008) and Sonoi et al. (2015) without a proper calibration of the parameters involved. Such a calibration must be based on the input physics of the employed models as well as the global parameters of the target star.

The surface correction relation by Ball & Gizon (2014), on the other hand, does not rely on any such calibration. Nevertheless, our results discourage from the use of the one-term correction by Ball & Gizon (2014), since the use of this surface correction relation does not recover the correct stellar properties and leads to models with a poor goodness-of-fit score. Moreover, in the presented analysis of the two clusters, the one-term correction by Ball & Gizon (2014) leads to the largest mass scatter among the investigated surface correction relations. This finding indicates that the one-term correction by Ball & Gizon (2014) is less reliable than the other surface correction relations. On the other hand, the two-term correction by Ball & Gizon (2014) performs very well throughout the analysis presented in this paper. It recovers the classical constraints on the binaries and both clusters, and it leads to models with an excellent goodness-of-fit score.

Neglecting the surface effect altogether is demonstrably not a viable strategy, as it significantly skews the obtained estimates for the global stellar parameters. When addressing the binary members in Section 4, the approach of neglecting the surface effect altogether thus yields parameter estimates that strongly deviate from the dynamical constraints whether or not we include additional (non-seismic) constraint to complement the individual frequencies. This approach also leads to larger mass scatter on a star by star basis when addressing the two clusters.

The surface effect refers to a frequency offset that partly arises from an incomplete description of the outermost superadiabatic layers of stars with convective envelopes. When addressing the surface effect, it is therefore worth noting that the underlying structural shortcomings affect the outer boundary conditions for the evaluated interior equilibrium structure (e.g. Kippenhahn, Weigert & Weiss 2012). As a result, the surface effect is a symptom of a model inadequacy that also affects the predicted stellar evolution tracks. This is demonstrated by, e.g. Mosumgaard et al. (2020) based on 3D magnetohydrodynamic simulations by Magic et al. (2013) and based on the method by Jørgensen et al. (2018). Surface correction relations do not account for such changes in the global stellar parameters. In other words, surface correction relations do not deal with all aspects that are associated with the structural shortcomings that give rise to the surface effect. However, to address this issue, one needs to go beyond some of the assumptions that enter state-of-the-art stellar models as shown by, e.g. Jørgensen et al. (2018) and (Mosumgaard et al. 2018). This is beyond the scope of the present paper. In this paper, we rather demonstrate that the correct treatment of the surface effect is crucial for the outcome of asteroseismic analyses when employing state-of-the-art stellar evolution codes.

## ACKNOWLEDGEMENTS

We thank Karsten Brogaard, J. Ted Mackereth, and Martin B. Nielsen for their input and useful discussions. We, furthermore, acknowledge the diligent and helpful feedback of our anonymous referee. The research leading to this paper has received funding from the European Research Council (ERC grant agreement No. 772293 for the project ASTEROCHRONOMETRY). GB acknowledges funding from the SNF AMBIZIONE grant No. 185805 (Seismic inversions and modelling of transport processes in stars) and funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 833925, project STAREX). RAG acknowledges the support received from the PLATO CNES grant.

## NOTE ADDED IN PROOF

In connection with the discussion of Fig. 10, we computed the mean distance moduli for all nine treatments of the surface effect based on the spectroscopic values for the effective temperature in combination with the stellar radii inferred by modelling. More specifically, we used the stellar radii from the models that enter Fig. 10. We did so to further substantiate that the offset in the distance modulus is related to the mismatch in the effective temperature. Indeed, in this scenario, all nine treatments, except for case c of K08 and NoSC, lie within 1 standard deviation of the constraints by Brogaard et al. (2012).

## REFERENCES

Adelberger E. G. et al., 2011, *Rev. Mod. Phys.*, 83, 195  
Aerts C., Christensen-Dalsgaard J., Kurtz D. W., 2010, *Asteroseismology*, Springer Verlag

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, 47, 481  
Bahcall J. N., Serenelli A. M., Basu S., 2006, *ApJS*, 165, 400  
Ball W. H., Gizon L., 2014, *A&A*, 568, A123  
Ball W. H., Gizon L., 2017, *A&A*, 600, A128  
Ball W. H., Beeck B., Cameron R. H., Gizon L., 2016, *A&A*, 592, A159  
Ball W. H., Themeßl N., Hekker S., 2018, *MNRAS*, 478, 4697  
Basu S., Kinnane A., 2018, *ApJ*, 869, 8  
Basu S. et al., 2011, *ApJ*, 729, L10  
Batalha N. M. et al., 2011, *ApJ*, 729, 27  
Bazot M., Bourguignon S., Christensen-Dalsgaard J., 2012, *MNRAS*, 427, 1847  
Benomar O., Appourchaux T., Baudin F., 2009, *A&A*, 506, 15  
Böhm-Vitense E., 1958, *Z. A.*, 46, 108  
Borucki W. J. et al., 2010, *Science*, 327, 977  
Bossini D. et al., 2017, *MNRAS*, 469, 4718  
Brogaard K. et al., 2012, *A&A*, 543, A106  
Brogaard K., Sandquist E., Jessen-Hansen J., Grundahl F., Frandsen S., 2015, *Asteroseismology of Stellar Populations in the Milky Way*, Astrophysics and Space Science Proceedings, vol. 39. Springer International Publishing, Switzerland, p. 51  
Brogaard K. et al., 2018, *MNRAS*, 476, 3729  
Brown T. M., 1984, *Science*, 226, 687  
Brown T. M., Gilliland R. L., Noyes R. W., Ramsey L. W., 1991, *ApJ*, 368, 599  
Buldgen G. et al., 2019, *MNRAS*, 482, 2305  
Casagrande L., VandenBerg D. A., 2014, *MNRAS*, 444, 392  
Cassisi S., Salaris M., Irwin A. W., 2003, *ApJ*, 588, 862  
Chaplin W. J., Miglio A., 2013, *ARA&A*, 51, 353  
Chaplin W. J., Elsworth Y., Isaak G. R., Lines R., McLeod C. P., Miller B. A., New R., 1998, *MNRAS*, 300, 1077  
Christensen-Dalsgaard J., 2008, *Ap&SS*, 316, 113  
Christensen-Dalsgaard J., Thompson M. J., 1997, *MNRAS*, 284, 527  
Christensen-Dalsgaard J., Däppen W., Lebreton Y., 1988, *Nature*, 336, 634  
Christensen-Dalsgaard J. et al., 2010, *ApJ*, 713, L164  
Davies G. R., Handberg R., Miglio A., Campante T. L., Chaplin W. J., Elsworth Y., 2014, *MNRAS*, 445, L94  
Dupret M. A., 2004, in Kurtz D. W., Pollard K. R., eds, *ASP Conf. Ser. Vol. 310, Non-adiabatic asteroseismology of near-main sequence variable stars*. Astron. Soc. Pac., San Francisco, p. 458  
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  
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306  
Gaia Collaboration, 2016, *A&A*, 595, A1  
García R. A., Ballot J., 2019, *Liv. Rev. Sol. Phys.*, 16, 4  
Gaulme P., Appourchaux T., Boumier P., 2009, *A&A*, 506, 7  
Gaulme P. et al., 2016, *ApJ*, 832, 121  
Goodman J., Weare J., 2010, *Commun. Appl. Math. Comput. Sci.*, 5, 65  
Gough D. O., 1990, in Osaki Y., Shibahashi H., eds, *Lecture Notes in Physics*, Vol. 367, *Comments on Helioseismic Inference*. Springer, Berlin, p. 283  
Gregory P. C., 2005, in Gregory P. C., ed., *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica'*. Support. Cambridge Univ. Press, Cambridge  
Grosjean M., Dupret M. A., Belkacem K., Montalbán J., Samadi R., Mosser B., 2014, *A&A*, 572, A11  
Gruberbauer M., Kallinger T., Weiss W. W., Guenther D. B., 2009, *A&A*, 506, 1043  
Grundahl F., Clausen J. V., Hardis S., Frandsen S., 2008, *A&A*, 492, 171  
Handberg R., Campante T. L., 2011, *A&A*, 527, A56  
Handberg R., Brogaard K., Miglio A., Bossini D., Elsworth Y., Slumstrup D., Davies G. R., Chaplin W. J., 2017, *MNRAS*, 472, 979  
Houdek G., Trampedach R., Aarslev M. J., Christensen-Dalsgaard J., 2017, *MNRAS*, 464, L124  
Houdek G., Lund M. N., Trampedach R., Christensen-Dalsgaard J., Handberg R., Appourchaux T., 2019, *MNRAS*, 487, 595  
Huber D. et al., 2013, *Science*, 342, 331  
Iglesias C. A., Rogers F. J., 1996, *ApJ*, 464, 943  
Jeffries Mark W. J. et al., 2013, *AJ*, 146, 58  
Jørgensen A. C. S., Angelou G. C., 2019, *MNRAS*, 490, 2890

- Jørgensen A. C. S., Christensen-Dalsgaard J., 2017, *MNRAS*, 471, 4802
- Jørgensen A. C. S., Weiss A., 2019, *MNRAS*, 488, 3463
- Jørgensen A. C. S., Weiss A., Mosumgaard J. R., Silva Aguirre V., Sahlholdt C. L., 2017, *MNRAS*, 472, 3264
- Jørgensen A. C. S., Mosumgaard J. R., Weiss A., Silva Aguirre V., Christensen-Dalsgaard J., 2018, *MNRAS*, 481, L35
- Jørgensen A. C. S., Weiss A., Angelou G., Silva Aguirre V., 2019, *MNRAS*, 484, 5551
- Kippenhahn R., Weigert A., Weiss A., 2012, *Stellar structure and evolution*. Springer-Verlag, Berlin
- Kjeldsen H., Bedding T. R., 1995, *A&A*, 293, 87
- Kjeldsen H., Bedding T. R., Christensen-Dalsgaard J., 2008, *ApJ*, 683, L175
- Krishna Swamy K. S., 1966, *ApJ*, 145, 174
- Li T., Bedding T. R., Huber D., Ball W. H., Stello D., Murphy S. J., Bland-Hawthorn J., 2018, *MNRAS*, 475, 981
- Lundkvist M. S., Huber D., Silva Aguirre V., Chaplin W. J., 2018, preprint ([arXiv:1804.02214](https://arxiv.org/abs/1804.02214))
- Lund M. N., Reese D. R., 2018, in Campante T. L., Santos N. C., Monteiro M. J. P. F. G., eds, *Astrophysics and Space Science Proceedings, Vol. 49, Asteroseismology and Exoplanets: Listening to the Stars and Searching for New Worlds*. Springer, Berlin, p. 149
- Magic Z., Weiss A., 2016, *A&A*, 592, A24
- Magic Z., Collet R., Asplund M., Trampedach R., Hayek W., Chiavassa A., Stein R. F., Nordlund Å., 2013, *A&A*, 557, A26
- Manchon L., Belkacem K., Samadi R., Sonoi T., Marques J. P. C., Ludwig H. G., Caffau E., 2018, *A&A*, 620, A107
- McKeever J. M., Basu S., Corsaro E., 2019, *ApJ*, 874, 180
- Miglio A. et al., 2009, *A&A*, 503, L21
- Miglio A. et al., 2012, *MNRAS*, 419, 2077
- Miglio A. et al., 2013, *MNRAS*, 429, 423
- Miglio A., 2012, *Astrophysics and Space Science Proceedings, Vol. 29, Red Giants as Probes of the Structure and Evolution of the Milky Way*. Springer-Verlag, Berlin, p. 11
- Milliman K. E., Mathieu R. D., Geller A. M., Gosnell N. M., Meibom S., Platais I., 2014, *AJ*, 148, 38
- Mosumgaard J. R., Ball W. H., Silva Aguirre V., Weiss A., Christensen-Dalsgaard J., 2018, *MNRAS*, 478, 5650
- Mosumgaard J. R., Jørgensen A. C. S., Weiss A., Silva Aguirre V., Christensen-Dalsgaard J., 2020, *MNRAS*, 491, 1160
- Nsamba B., Campante T. L., Monteiro M. J. P. F. G., Cunha M. S., Rendle B. M., Reese D. R., Verma K., 2018, *MNRAS*, 477, 5052
- Oti Floranes H., Christensen-Dalsgaard J., Thompson M. J., 2005, *MNRAS*, 356, 671
- Pérez Hernández F., García R. A., Mathur S., Santos A. R. G., Régulo C., 2019, *Frontiers Astron. Space Sci.*, 6, 41
- Piau L., Collet R., Stein R. F., Trampedach R., Morel P., Turck-Chièze S., 2014, *MNRAS*, 437, 164
- Porqueres N., Kodi Ramanah D., Jasche J., Lavaux G., 2019a, *A&A*, 624, A115
- Porqueres N., Jasche J., Lavaux G., Enßlin T., 2019b, *A&A*, 630, A151
- Reese D. R., 2016, *Astrophysics Source Code Library*, record ascl:1611.014
- Rendle B. M. et al., 2019, *MNRAS*, 484, 771
- Rodrigues T. S. et al., 2017, *MNRAS*, 467, 1433
- Rosenthal C. S., Christensen-Dalsgaard J., Nordlund Å., Stein R. F., Trampedach R., 1999, *A&A*, 351, 689
- Roxburgh I. W., Vorontsov S. V., 2003, *A&A*, 411, 215
- Sandquist E. L. et al., 2013, *ApJ*, 762, 58
- Schou J., Birch A. C., 2020, preprint ([arXiv:2004.13548](https://arxiv.org/abs/2004.13548))
- Scuflaire R., Théado S., Montalbán J., Miglio A., Bourge P. O., Godart M., Thoul A., Noels A., 2008a, *Ap&SS*, 316, 83
- Scuflaire R., Montalbán J., Théado S., Bourge P. O., Miglio A., Godart M., Thoul A., Noels A., 2008b, *Ap&SS*, 316, 149
- Sonoi T., Samadi R., Belkacem K., Ludwig H.-G., Caffau E., Mosser B., 2015, *A&A*, 583, A112
- Torres G., 2010, *AJ*, 140, 1158
- Townsend R. H. D., Teitler S. A., 2013, *MNRAS*, 435, 3406
- Trampedach R., Asplund M., Collet R., Nordlund Å., Stein R. F., 2013, *ApJ*, 769, 18
- Van Eylen V., Lindholm Nielsen M., Hinrup B., Tingley B., Kjeldsen H., 2013, *ApJ*, 774, L19
- Vernazza J. E., Avrett E. H., Loeser R., 1981, *ApJS*, 45, 635
- Viani L. S., Basu S., Chaplin W. J., Davies G. R., Elsworth Y., 2017, *ApJ*, 843, 11
- Vinyoles N. et al., 2017, *ApJ*, 835, 202
- Zinn J. C., Pinsonneault M. H., Huber D., Stello D., Stassun K., Serenelli A., 2019, *ApJ*, 885, 166

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