Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **507,** 1795–1813 (2021)
Advance Access publication 2021 July 20

https://doi.org/10.1093/mnras/stab2030

Determining fundamental parameters of detached double-lined eclipsing binary systems via a statistically robust machine learning method

Bryce A. Remple,[1,2,3]★ George C. Angelou [1] and Achim Weiss[1]

[1]*Max-Planck Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany*
[2]*Ludwig-Maximillians-Universität München, Geschwister-Scholl-Platz 1, D-80539 Munich, Germany*
[3]*Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany*

Accepted 2021 July 12. Received 2021 July 12; in original form 2021 April 21

ABSTRACT

The parameter space for modelling stellar systems is vast and complicated. To find best-fitting models for a star one needs a statistically robust way of exploring this space. We present a new machine-learning approach to predict the modelling parameters for detached double-lined eclipsing binary systems, including the system age, based on observable quantities. Our method allows for the estimation of the importance of several physical effects which are included in a parametrized form in stellar models, such as convective core overshoot or stellar spot coverage. The method yields probability distribution functions for the predicted parameters which take into account the statistical and, to a certain extent, the systematic errors which is very difficult to do using other methods. We employ two different approaches to investigate the two components of the system either independently or in a combined manner. Furthermore, two different grids are used as training data. We apply the method to 26 selected objects and test the predicted best solutions with an on-the-fly optimization routine which generates full hydrostatic models. While we do encounter failures of the predictions, our method can serve as a rapid estimate for stellar ages of detached eclipsing binaries taking full account of the uncertainties in the observables.

Key words: binaries: eclipsing – stars: evolution – stars: fundamental parameters – starspots.

1 INTRODUCTION

Robust stellar modelling of observed stars requires precise knowledge of the fundamental quantities of the objects. Foremost among these quantities are mass and chemical composition, which define the star's evolution and fate. Additionally, radius, surface temperature (or colour), and luminosity (or brightness) relate to the evolutionary stage and the age of the star. Of increasingly higher importance are pulsational properties, as these allow insight into the interior structure via asteroseismic observations and analysis methods. From a theoretical perspective, the models depend on the treatment of physical effects, many of which are addressed with parametrizations designed to approximate the processes at work and which, more often than not, require some form of tuning or calibration. Examples include dynamic convection, diffusive processes, and the influence of magnetic fields. In addition, also unknown details of chemical composition, such as precise helium content, add to the uncertainty of the models. If one does not decide for a fixed set of parameters and physical models, the parameter space for modelling stellar objects becomes vast and complicated, and each new physical process or free parameter considered further increases the dimensionality. However, varying the physically motivated free parameters allows one to find the best-fitting model while simultaneously gaining insight into the physical processes relevant for stars. This is a common method of choice to improve stellar structure theory.

As the relationship between the observations of stars and their internal properties are complicated and difficult to characterize (Bellinger et al. 2016), the task of determining the best parameters for modelling a star is a challenging one. With traditional methods such as iterative optimization or grid-based modelling, one is forced to run more and more models to explore the parameter space sufficiently. This quickly becomes computationally unfeasible. As an example, in Quirion, Christensen-Dalsgaard & Arentoft (2010), they calculate a grid of 7300 stellar tracks totalling more than 5.8 million models while varying only four initial parameters. Alternatively, one could reduce the dimensionality of the problem, by omitting certain physical processes, or by setting a calibrated value for the associated free parameter(s) which may or may not reflect the object being modelled.

In recent years, it has been shown that it is possible to reduce the impact of many of these difficulties by employing machine learning algorithms. These algorithms capture relationships in the data statistically and use them to construct regression models which connect the observational quantities of stars to their chemical, structural, and evolutionary properties (Bellinger et al. 2016; Angelou et al. 2017). Additionally, these regression models can be built from a reasonable number of models and can be used to process entire catalogues of stars at one time (Angelou et al. 2020). Furthermore, machine learning methods can provide an easy way of more robustly considering statistical and systematic uncertainties without computing additional grids (Silva Aguirre et al. 2015; Bellinger et al. 2019).

Previous studies have successfully used machine learning to infer modelling parameters from observational data (Bellinger et al. 2016;

★ E-mail: brempl@mpa-garching.mpg.de

Angelou et al. 2017; Bellinger et al. 2019; Angelou et al. 2020). These studies focused on using fundamental observables along with asteroseismic quantities to analyse single stars. The asteroseismic quantities are necessary to provide additional information on the stars and to break some of the degeneracy inherent in the stellar parameter space. Another way to break some of the degeneracies is by analysing binary systems as we have done in this study. Binaries, specifically detached double-lined eclipsing binaries (DDLEBs), are ideal targets to study because the mass and radius of both components can be measured with a very high accuracy, the components are expected to have very similar initial metallicities, and the components should be co-natal, i.e. the components are the same age. These qualities impose powerful constraints on the possible solutions for the system and restrict the size and complexity of the parameter space that must be considered.

In this paper, we apply machine learning regression in a new way to analyse binary systems and present the results for 26 DDLEB systems. The method utilizes machine learning models to produce robust probability distribution functions (PDF) of the modelling parameters. This is then connected to an on-the-fly optimization search for a deeper analysis of full hydrostatic models. The hydrostatic models are compared to the measured radii of the components as well as the luminosity ratio and effective temperature ratio of the system to determine if the models match the observations. The main purpose of the method is to determine the ages of the systems. Unlike previous machine learning based methods, our algorithm analyses both components of the system at the same time and uses the full observational data of the system to predict the modelling parameters. A machine learning algorithm which focuses on one star at a time, similar to that of Bellinger et al. (2016), is also utilized as a way of comparing our new method to an established method.

Each method is trained on two separate grids of models which consider different physics. The 'canonical grid' includes what could be considered a standard set of physical processes and free parameters for stellar modelling. Specifically, it includes convective core overshoot, diffusion for stars below 1.1 $M_{\odot}$, varying mixing lengths, and, of course, varying metallicity, helium abundance, and mass. The so-called 'extended grid' additionally includes α-enhancement and spot coverage.

The paper is organized as follows: in Section 2 DDLEB studies will be discussed, including traditional methods and previous results from the literature. Following this some background on random forests will be presented. In the subsequent section, Section 3, the details of the methods utilized in this study will be presented. The important aspects of GARSTEC, the stellar evolution code used in this study, will be introduced (Section 3.1), the system selection criteria will be explained (Section 3.2), and the machine learning methods will be discussed in detail (Section 3.3). Next, the results for both methods and both grids will be shown and compared (Section 4). A closer analysis of a chosen few systems will also be done in this section. In Section 5, general conclusions about the new approach regarding effectiveness, reliability, and advantages will be established. Additionally, a special emphasis will be placed on the overshoot parameter values and spot coverage and any trends for these quantities. Finally, we outlay our conclusions in Section 6.

2 BACKGROUND

The use of DDLEB systems as tests of stellar evolution has been common practice for decades. A common use of DDLEB systems is to test the need for or extent of convective core overshoot in stars.[1] One of the classic reference papers is that of Schröder, Pols & Eggleton (1997) who looked at ζ Aurigae systems to establish the necessity or the amount of convective core overshooting.

In Lastennet & Valls-Gabaud (2002) a collection of DDLEB systems were used in a general parameter study wherein the authors attempted to ascertain the age and metallicity of the selected systems. They find that their models can, within errors, fit most of their systems. Crucially, they note a large degeneracy between age and metallicity. According to the authors, the lack of metallicity measurements limits a precise determination of the age regardless of how accurate the other measurements are, and that '...the spectroscopic measure of the metallicity in the components of these systems is essential for this technique to be an efficient test of stellar evolution.'

Torres, Andersen & Giménez (2010) analysed 94 systems and draw the same conclusion as Lastennet & Valls-Gabaud (2002), that metallicity measurements are necessary for precise age determinations of DDLEB systems. In addition, they find that the presence of spots in certain systems is 'confirmed beyond a doubt' and stress that the effect of spots must be addressed for accurate modelling, especially regarding low-mass stars.

Both these issues were considered in Higl & Weiss (2017), where 19 DDLEB systems were studied. All systems were required to have reliable compositional measurements to break the age-metallicity degeneracy. This sets this study apart from most studies of this type where metallicity is varied as a free parameter. For each system a small grid of models was run with the GARSTEC stellar evolution code (Weiss & Schlattl 2008) with three values for the mass, and a selected small number of values for composition, overshoot, spot coverage, and diffusion depending on the applicability of the process to the star in question. This limited number of models reduced the required computational effort, but at the same time did not allow us to identify the truly best-fitting model and the uncertainties in the physical parameters. The models were compared to observations and best models were chosen largely based on age–radius diagrams. They find that the agreement between models and observations for some systems is improved by including one or more of diffusion, overshoot, and spot coverage. We seek to build on this study by more systematically exploring the parameter ranges.

In their investigation, Higl & Weiss find no clear linear relationship between mass and the convective overshoot parameter. The question of whether or not there is a linear relationship between mass and overshoot parameter is unresolved and often debated in the literature. Stancliffe et al. (2015) and Constantino & Baraffe (2018) argue that the uncertainties in the models are too large to claim that a linear relationship exists. On the other hand, in a series of papers by Claret and Torres (Claret & Torres 2016, 2017, 2018, 2019), they argue that they find a linear dependence between mass and overshoot up to around 2 $M_{\odot}$, at which point it flattens out. More recently, in Viani & Basu (2020), evidence for a linear mass-overshoot relationship is also found.

These studies all exploit the constraints provided by DDLEB systems in order to draw conclusions about the stellar models and/or specific modelling parameters. Regardless of the specific aims, these studies all face the same task: to explore the stellar parameter space and to chose the point in that space which produces models that

[1]More generally the extent of a fully or partially mixed zone above the fully convective core is tested. Convective overshoot may be the most likely, but not unique effect to mix material beyond the Schwarzschild boundary

most closely match the observations. One notes from these studies that the dimensionality of the parameter space continues to grow as more physical processes are considered necessary to accurately model stars.

Three traditional ways of exploring the parameter space include grid based search methods (Gai et al. 2011; Chaplin et al. 2014), optimization methods (Metcalfe, Creevey & Christensen-Dalsgaard 2009), and in select cases Markov Chain Monte Carlo Methods (MCMC) (Bazot, Bourguignon & Christensen-Dalsgaard 2012; Jørgensen & Angelou 2019). A grid search requires a grid of tracks to be run which covers the region of the multidimensional parameter space where one believes the solution to be. A key characteristic of this method is the grid density. The resolution of this method is dependent on the density of the grid of parameters used (Bellinger et al. 2016). Thus, to explore a large swath of the parameter space or to explore a many-dimensional parameter space one needs to run many tracks. Additionally, because the method does not scale well one tends to underestimate the systematic uncertainty.

Optimization methods, on the other hand, use an optimization algorithm to search the parameter space for the best model. These methods are robust but are very computationally expensive requiring many tracks to be run in order to explore the parameter space sufficiently. In Metcalfe et al. (2009), a genetic algorithm is used and the authors note that the algorithm is computationally demanding but 'is usually hundreds or even thousands of times more efficient than a complete grid'.

Finally, MCMC allows for a robust exploration of the parameter space and yields reliable probability distributions for target parameters, but incurs a high computational cost to do so (Jørgensen & Angelou 2019). In Bazot et al. (2012) an MCMC method is used to infer the parameters of α Cen A. Their analysis required running 3 Markov chains of length 50 000. As a final note on these traditional methods, a common weakness among optimization and MCMC methods is that the procedure must be repeated in its entirety for each new star that one would like to model.

In recent years computational power has caught up to the algorithmic complexity of machine learning methods leading to machine learning becoming ubiquitous in science (Carleo et al. 2019). Random Forests, the machine learning algorithm utilized in this study, has been a particularly popular choice of algorithm due to its wide applicability and fast training speeds. From classifying the type of compact object in Low-mass X-ray binaries (Pattnaik et al. 2021), to searching for trans-Neptunian objects (Henghes et al. 2021), to predicting modelling parameters of stars (Bellinger et al. 2016; Angelou et al. 2020), random forests have proven to be a useful tool for astrophysicists.

Random forests (RF) is a supervised ensemble learning method. This means that it requires a training data set from which it can learn (supervised) and that it relies on combining the predictions of a population of learning methods, in this case decision trees, in order to make a prediction (ensemble) (Louppe 2014). Each decision tree learns a set of rules for relating the features or input variables to the targets or the variables to predict. Once the forest is trained one can use it to predict the targets of unseen data points. For a theoretical overview of random forests see Mehta et al. (2019) or Hastie, Tibshirani & Friedman (2009).

Much of the groundwork for this study was laid out in Bellinger et al. (2016) where they used random forests to predict the modelling parameters of stars, e.g. age, mixing length, etc., from observational quantities including asteroseismic quantities. They establish the efficacy of using random forests to analyse stellar systems on the main sequence and demonstrate the benefits of this method. They find that random forests are substantially faster than other methods, that the non-linear non-parametric regression which random forests employ allow for a similar level of precision in comparison to other methods with an order-of-magnitude fewer models being required, that the method allows for a wider range of values and a larger combination of stellar parameters to be investigated, and finally that the statistical nature of the method allows for the propagation of statistical errors and a more honest appraisal of the systematic errors. Some of these ideas will be elaborated on more in Section 3.3.

3 METHODS

In this section, we begin by discussing GARSTEC, the stellar evolution code used in this study. A full description of GARSTEC can be found in Weiss & Schlattl (2008), but for the sake of brevity we will focus only on the aspects of the code which are of greatest importance to this study.

3.1 GARSTEC

3.1.1 Convective overshooting

When convective elements, for example plumes or eddies, encounter a stable boundary a number of behaviours are possible (Zahn 1991). One possibility is that the convective material might reach the neutral buoyancy point, otherwise known as the Schwarzschild boundary, with a non-zero velocity, and thus overshoot the convective boundary as determined by mixing length theory (Kippenhahn, Weigert & Weiss 2012). This process was historically referred to as overshooting by the astrophysics community. However, overshoot is applied in stellar models as catch-all for the various physical processes (Angelou et al. 2020) which mix material beyond the stable boundary, thus overshoot may have more than one meaning depending on the context. Because of this there is a push in the literature to refer to the specific overshooting process as penetration, as is done in fluid dynamics, and to refer to the general extraboundary mixing as convective boundary mixing, or CBM (Zahn 1991; Denissenkov et al. 2013; Davis, Jones & Herwig 2019; Angelou et al. 2020).

Regardless of the name used or processes considered, the efficiency of these extra mixing processes is unknown and as such is implemented in stellar evolution codes as a parametrized process. In GARSTEC, convection is handled by the standard mixing-length formulation of Kippenhahn et al. (2012). Convective penetration is implemented via the description by Freytag, Ludwig & Steffen (1996). Their description is based on a diffusive overshoot mixing model which is incorporated into GARSTEC with a diffusion constant, derived from the convective velocity as estimated by mixing-length theory, and an adjustable parameter $f_{\rm ov}$. The standard value of $f_{\rm ov}$ in GARSTEC is 0.02, determined by fitting isochrones to young open clusters.

One complication of this prescription is that problems may occur when running models for stars with very small convective cores. As the pressure scale height, $H_{\rm P}$, grows ever larger towards the centre of the star, the overshooting region would become unrealistically large. This is handled in GARSTEC by reducing $H_{\rm P}$, effectively applying a geometrical cut-off,

$$H_{\rm P}^{*} = H_{\rm P} \min\left\{1, \left(\frac{\Delta R_{\rm CZ}}{2H_{\rm P}}\right)^{2}\right\}, \tag{1}$$

which ensures that the overshooting region is restricted to a fraction of the radial extent of the convective zone, $R_{\rm CZ}$. This is particularly

relevant for models on the lower main-sequence ($M \lesssim 2\,M_\odot$), which have very small convective cores.

The implication of this correction is that the value of the convective overshoot parameter which the user sets, the user-set overshoot (USO) value, is not necessarily the value which the code uses, which we call the effective overshoot (Angelou et al. 2020).

The effective overshoot is defined as the overshoot parameter value which would lead to a convective core overshoot region that is the size of the geometrically cut convective core overshoot region. For more massive models on the main-sequence ($M \gtrsim 2\,M_\odot$), the USO value and the effective overshoot value are the same. However, for the models where it is not, the effective overshoot will better describe the behaviour of overshoot in the code. The relationship between effective overshoot and USO is

$$\text{Effective Overshoot} = \frac{H_P^*}{H_P}\text{USO}. \tag{2}$$

Furthermore, the overshoot routine has been slightly modified in a way that overshoot only occurs in the core convection zone and is turned off for all other convective regions. If there is no convective core present in a model, then the value of both overshoot values is set to zero. Only core overshoot is considered because it is the overshoot region which has the greatest impact on the evolution of the star and we wish to study this effect specifically.

Finally, the quantity which the random forest ultimately predicts is the median effective overshoot (MEO). This is a quantity which is the same for every model in a track. It is calculated by finding the median effective overshoot value of models in a particular track which have a convective core. If there are no models with a convective core, naturally the median effective overshoot is set to 0. The usage of the median effective overshoot is of particular importance for predictions for post-main sequence stars. These stars no longer have a convective core and, thus, no core overshooting is occurring. However, their evolution is still very much influenced by the effects of core overshoot on the main sequence. The median effective overshoot allows for some of this historical context to be available to the random forest. Thus, it can predict an overshoot value which explains the whole evolution and not just what the overshoot would be at that moment in the evolution.

3.1.2 *Diffusion*

Diffusion is a slow process. Under solar conditions the characteristic time it would take for an element to diffuse a solar radius is roughly 6×10^{13} yr (Thoul, Bahcall & Loeb 1994). Since the time-scale for diffusion is large, the effects of diffusion are only important in stars where its effect is not suppressed by a process with a shorter time-scale. Rotational mixing, radiative diffusivity, and stellar wind mass-loss have all been proposed as such a counteracting process (Chaboyer, Demarque & Guenther 1999; Morel & Thévenin 2002). Observationally, for stars with convective envelopes thinner than the Sun, diffusion has been shown to be important. Additionally, it has been shown that including diffusion in models of the sun improves the agreement with observations of, among other things, surface helium abundance (Salaris & Weiss 2001). Therefore, diffusion is included in the training data only for stars whose mass is less than $1.1\,M_\odot$.

As described in Weiss & Schlattl (2008), diffusion is handled using the description of Thoul et al. (1994). This approach assumes all elements are ionized and neglects the effects of radiation pressure but is otherwise quite general. As diffusion gradually reduces the surface abundances of all elements heavier than hydrogen over time, the initial surface abundances are an additional target for the RF. It must predict the appropriate initial abundances such that the composition of the surface matches the present observed metallicity, which may not necessarily be the same for both components at the present age.

3.1.3 *Opacities and equation of state*

All models were run using the OPAL-EoS equation of state (Rogers & Nayfonov 2002). Likewise, OPAL opacities (Iglesias & Rogers 1996), extended by the Alexander and Ferguson molecular opacity tables (Ferguson et al. 2005) for lower temperatures, and by conductive opacities following Cassisi et al. (2007), were used for all models. For α-enhanced mixtures, corresponding opacity tables including this enhancement were used.

3.1.4 *Alpha enhancement*

It has long been known that the chemical composition of stars is a function of place and time of origin. Through a myriad of studies in the 70s and 80s, it became clear that α-element (oxygen, magnesium, silicon, calcium, etc.) abundances are enhanced relative to iron in most, if not all, low metallicity stars (Salaris, Chieffi & Straniero 1993). Additionally, the amount of overabundance is to first order related to the metallicity.

In this way, α-enhancement, like metallicity, is a function of place and time of origin. For this study, we chose to adopt an empirical relation between metallicity and α-enhancement which would allow us to assign α-enhancements based on discrete [Fe/H] bins. The specific relationship used is that found in the Ness et al. (2013) study of stellar populations of the Milky Way. Following the results from panel 3 fig. 19 in that study, we have adopted the following [Fe/H]-α-enhancement relationship:

$$\alpha\text{-enhancement} = \begin{cases} 0.6, & \text{if [Fe/H]} < -1.5 \\ 0.4, & \text{if} -1.5 < \text{[Fe/H]} < -0.5 \\ 0.2, & \text{if} -0.5 < \text{[Fe/H]} < 0.0 \\ 0.0, & \text{else.} \end{cases} \tag{3}$$

In this way the α-enhancement can be assigned to stars based on the metallicity predictions.

3.1.5 *Spot model*

For at least 40 yr, it has been clear that for low-mass stars ($M < 1.1\,M_\odot$), theoretical calculations underestimate the radius measured in observations. Hoxie (1973) noticed a discrepancy of 30 per cent or more. This discrepancy between models and observations is largely attributed to the model's lack of consideration of spot coverage. In this paper, we apply the same spot model as Higl & Weiss (2017) which is laid out in Spruit & Weiss (1986).

Spots inhibit the efficiency of convection near the surface of stars. This causes a drop in the effective temperature of the star. In order to compensate for the reduction in effective temperature, the star is forced to expand to radiate its energy. Following the treatment in Spruit & Weiss (1986), we account for the effects of spots by introducing an effective spot coverage, f_e, defined as,

$$f_e = \left[ba^2 + c\left(1 - a^2\right)\right] f \tag{4}$$

$$f = f_u + f_p \tag{5}$$

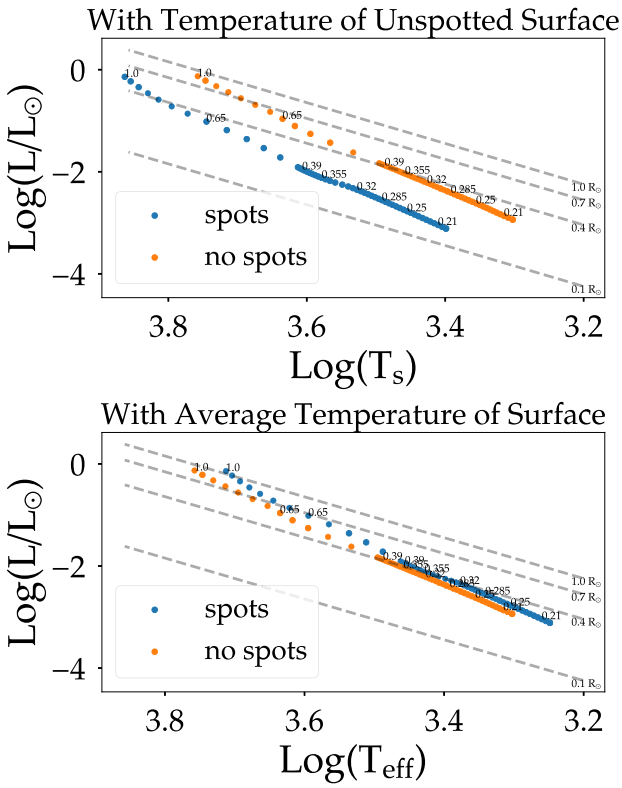


Figure 1. The ZAMS for stars with and without spot coverage are shown. All models with spots have an effective spot coverage of 75 per cent . In the upper panel effective temperature is defined as the temperature of the unspotted surface of the star. In the lower panel effective temperature is the average temperature over the whole surface of the star. Numbers indicate stellar mass in solar units. Lines of constant radius are plotted in dashed grey lines as well.

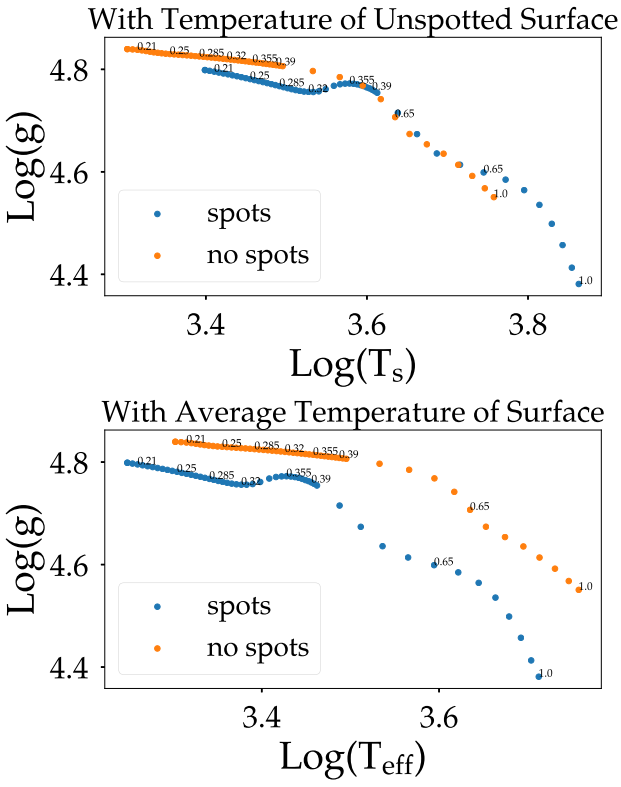


Figure 2. A Kiel diagram of ZAMS with and without spot coverage. See Fig. 1 for an explanation of plot elements.

$$a^2 = \frac{f_u}{f_u + f_p} \tag{6}$$

$$b = 1 - \frac{T_u^4}{T_0^4} \tag{7}$$

$$c = 1 - \frac{T_p^4}{T_0^4}, \tag{8}$$

where f_u, T_u, f_p, and T_p are the fractional areas and temperatures of the umbrae and penumbrae, respectively, and T_0 is the temperature of the unspotted surface.

Fig. 1 shows the effect of a 75 per cent effective spot coverage on the low-mass section of the zero-age main sequence (ZAMS). In the top panel, the stars are plotted with T_s, the temperature of the unspotted surface, and in the bottom panel they are plotted with the average temperature over the whole surface, T_{eff}. Fig. 2 shows the same data but plotted on a Kiel diagram. In this plot, the effect on the radii of the models is even clearer.

Spot coverage may be deduced from observations in a number of ways. Using spectroscopy, one can analyse the TiO lines to derive spot coverage (O'Neal et al. 2004). This technique only works for M stars. Additionally, one can measure light curves for stars and observe variations which are attributed to spots. Synthetic light curves can then be produced with varying spot coverage parameters and matched to the observations as was done in Morales et al. (2009). The authors state, though, that '...such modelling is fraught with difficulties and there is abundant literature discussing problems of indeterminacy and non-uniqueness, particularly when using data of limited quality'. Recently in Pan et al. (2021), Miller et al. (2021), this method was used to deduce the spot coverage of binary systems. In Section 4.2, we look at a system which has a spot coverage estimate from observations and compare it to our results.

3.2 Selecting systems

All systems chosen for this study adhere to five criteria. Criterion 1 is that no component can be outside the mass range $0.25\,M_\odot < M_* < 6.0\,M_\odot$. Criterion number 2 is accurate and precise measurements. The system must have mass and radius measurements which are accurate to within at least 2 per cent. This criterion is in accordance with Andersen (1991) where it is stated that this is the accuracy limit for non-trivial results. The third criterion is that spectroscopic abundances, luminosities, and effective temperatures must be available: a critical test is only possible when all quantities are known (Andersen 1991). Requiring a measured metallicity and not varying it as a free parameter makes this study different from many others (Pols et al. 1997; Schröder et al. 1997; Lastennet & Valls-Gabaud 2002; Meng & Zhang 2014; Claret & Torres 2016), but it has been shown that metallicity is essential for an accurate determination of the age (Lastennet & Valls-Gabaud 2002; Torres et al. 2010). Criterion number 4 is that the system must have a primary and a secondary with different masses: a system with two stars that have the same mass is trivial to find a solution for. All the systems in this study have a mass ratio, $q = M_B/M_A$, that is less than 0.96. The final criterion is that neither component can be on the pre-main sequence

Table 1. Fundamental properties of all systems studied in this work. For each object the first and second line refer to the primary and secondary component, respectively. Data are from the DEBCat catalogue (Southworth 2015). Note, no metallicity uncertainty was reported for RW Lac so an uncertainty of 0.3 was adopted.

System	$M/M_\odot$	$R/R_\odot$	log(g)	log(Teff)	$\log(L/L_\odot)$	[Fe/H]
AD Boo	1.414 ± 0.009	1.612 ± 0.014	4.173 ± 0.008	3.818 ± 0.008	0.64 ± 0.03	0.1 ± 0.15
	1.209 ± 0.006	1.216 ± 0.01	4.351 ± 0.007	3.789 ± 0.008	0.28 ± 0.04	–
ASAS J065134-2211.5	0.956 ± 0.012	0.997 ± 0.004	4.421 ± 0.003	3.74 ± 0.008	−0.09 ± 0.03	0.09 ± 0.13
	0.674 ± 0.005	0.69 ± 0.007	4.589 ± 0.009	3.599 ± 0.012	−0.96 ± 0.05	–
BG Ind	1.428 ± 0.008	2.29 ± 0.017	3.873 ± 0.005	3.803 ± 0.018	0.87 ± 0.07	−0.2 ± 0.1
	1.293 ± 0.008	1.68 ± 0.038	4.099 ± 0.02	3.823 ± 0.015	0.68 ± 0.08	–
BK Peg	1.414 ± 0.007	1.988 ± 0.008	3.992 ± 0.004	3.797 ± 0.006	0.74 ± 0.02	−0.12 ± 0.07
	1.257 ± 0.005	1.474 ± 0.017	4.201 ± 0.01	3.801 ± 0.006	0.49 ± 0.03	–
CoRoT 105906206	2.25 ± 0.04	4.24 ± 0.02	3.53 ± 0.01	3.829 ± 0.01	1.53 ± 0.04	0.0 ± 0.1
	1.29 ± 0.03	1.34 ± 0.01	4.3 ± 0.02	3.789 ± 0.011	0.36 ± 0.05	–
EF Aqr	1.244 ± 0.008	1.338 ± 0.012	4.28 ± 0.007	3.789 ± 0.006	0.36 ± 0.011	0.0 ± 0.1
	0.946 ± 0.006	0.956 ± 0.012	4.453 ± 0.011	3.715 ± 0.009	−0.227 ± 0.02	–
EW Ori	1.173 ± 0.011	1.168 ± 0.005	4.372 ± 0.005	3.783 ± 0.007	0.22 ± 0.03	0.05 ± 0.09
	1.123 ± 0.009	1.097 ± 0.005	4.408 ± 0.005	3.771 ± 0.007	0.12 ± 0.03	–
FL Lyr	1.21 ± 0.008	1.244 ± 0.023	4.331 ± 0.016	3.796 ± 0.008	0.33 ± 0.095	−0.07 ± 0.09
	0.951 ± 0.004	0.9 ± 0.024	4.508 ± 0.023	3.74 ± 0.019	−0.18 ± 0.08	–
HAT-TR-318-007	0.448 ± 0.001	0.455 ± 0.004	4.774 ± 0.0057	3.504 ± 0.015	−1.715 ± 0.06	0.3 ± 0.11
	0.272 ± 0.004	0.291 ± 0.002	4.9442 ± 0.0039	3.491 ± 0.015	−2.151 ± 0.062	–
IM Vir	0.981 ± 0.012	1.061 ± 0.016	4.379 ± 0.014	3.754 ± 0.008	−0.012 ± 0.034	−0.1 ± 0.25
	0.664 ± 0.005	0.681 ± 0.013	4.594 ± 0.017	3.628 ± 0.013	−0.867 ± 0.056	–
KIC 5640750	1.515 ± 0.033	14.06 ± 0.12	2.323 ± 0.007	3.656 ± 0.007	1.871 ± 0.029	−0.29 ± 0.09
	1.292 ± 0.017	1.853 ± 0.023	4.014 ± 0.01	3.767 ± 0.015	0.555 ± 0.06	–
KIC 7821010	1.277 ± 0.017	1.276 ± 0.011	4.332 ± 0.007	3.827 ± 0.01	0.47 ± 0.04	0.1 ± 0.08
	1.221 ± 0.016	1.21 ± 0.014	4.359 ± 0.009	3.817 ± 0.011	0.39 ± 0.05	–
KIC 8410637	1.472 ± 0.017	10.595 ± 0.049	2.556 ± 0.003	3.663 ± 0.008	1.656 ± 0.03	0.02 ± 0.08
	1.309 ± 0.014	1.556 ± 0.01	4.171 ± 0.005	3.783 ± 0.014	0.468 ± 0.058	–
KIC 9540226	1.39 ± 0.031	13.431 ± 0.17	2.326 ± 0.01	3.661 ± 0.007	1.854 ± 0.03	−0.31 ± 0.09
	1.015 ± 0.016	1.034 ± 0.014	4.416 ± 0.01	3.765 ± 0.015	0.042 ± 0.061	–
Kepler-35	0.888 ± 0.005	1.028 ± 0.002	4.3623 ± 0.002	3.749 ± 0.011	−0.027 ± 0.043	−0.34 ± 0.2
	0.809 ± 0.004	0.786 ± 0.002	4.5556 ± 0.0016	3.716 ± 0.008	−0.387 ± 0.02	–
LL Aqr	1.196 ± 0.001	1.321 ± 0.006	4.274 ± 0.004	3.784 ± 0.003	0.332 ± 0.014	0.02 ± 0.04
	1.034 ± 0.001	1.002 ± 0.005	4.451 ± 0.004	3.756 ± 0.004	−0.019 ± 0.016	–
OGLE-LMC-SC9-230659	3.598 ± 0.038	32.832 ± 0.023	1.962 ± 0.005	3.699 ± 0.006	2.783 ± 0.025	−0.24 ± 0.11
	3.429 ± 0.03	23.399 ± 0.032	2.235 ± 0.011	3.702 ± 0.009	2.5 ± 0.036	–
RW Lac	0.928 ± 0.006	1.186 ± 0.004	4.257 ± 0.003	3.76 ± 0.008	0.14 ± 0.03	−0.3
	0.87 ± 0.004	0.964 ± 0.004	4.409 ± 0.004	3.745 ± 0.012	−0.1 ± 0.05	–
SW CMa	2.239 ± 0.014	3.014 ± 0.02	3.8298 ± 0.0065	3.914 ± 0.008	1.566 ± 0.032	0.55 ± 0.15
	2.104 ± 0.018	2.495 ± 0.042	3.967 ± 0.015	3.908 ± 0.008	1.38 ± 0.035	–
TZ For	2.057 ± 0.001	8.341 ± 0.11	2.915 ± 0.023	3.693 ± 0.003	1.57 ± 0.02	0.02 ± 0.05
	1.958 ± 0.001	3.97 ± 0.08	3.539 ± 0.037	3.803 ± 0.005	1.36 ± 0.03	–
V501 Mon	1.646 ± 0.004	1.888 ± 0.029	4.103 ± 0.013	3.876 ± 0.006	1.007 ± 0.027	0.01 ± 0.06
	1.459 ± 0.002	1.592 ± 0.028	4.199 ± 0.016	3.845 ± 0.006	0.743 ± 0.043	–
V530 Ori	1.004 ± 0.007	0.98 ± 0.013	4.457 ± 0.012	3.777 ± 0.007	0.016 ± 0.032	−0.12 ± 0.08
	0.596 ± 0.002	0.587 ± 0.007	4.676 ± 0.01	3.589 ± 0.013	−1.154 ± 0.053	–
V636 Cen	1.052 ± 0.005	1.018 ± 0.004	4.444 ± 0.004	3.771 ± 0.006	0.05 ± 0.03	−0.2 ± 0.08
	0.854 ± 0.003	0.83 ± 0.004	4.532 ± 0.005	3.699 ± 0.009	−0.41 ± 0.04	–
VZ Hya	1.271 ± 0.006	1.314 ± 0.005	4.305 ± 0.005	3.809 ± 0.01	0.48 ± 0.04	−0.2 ± 0.12
	1.146 ± 0.007	1.112 ± 0.007	4.405 ± 0.006	3.799 ± 0.01	0.24 ± 0.04	–
WASP 0639-32	1.155 ± 0.004	1.834 ± 0.023	3.974 ± 0.011	3.801 ± 0.003	0.685 ± 0.018	−0.33 ± 0.1
	0.783 ± 0.003	0.729 ± 0.008	4.607 ± 0.01	3.732 ± 0.006	−0.392 ± 0.028	–
YZ Cas	2.263 ± 0.012	2.525 ± 0.011	3.988 ± 0.004	3.979 ± 0.005	1.672 ± 0.022	0.1 ± 0.06
	1.325 ± 0.007	1.331 ± 0.006	4.311 ± 0.004	3.838 ± 0.015	0.552 ± 0.061	–

or past the helium flash. Consequently, a system satisfying all of these criteria makes for an ideal test of the methods used in this study. All the systems in this study can be found in the DEBCat catalogue (Southworth 2015). The complete list of systems selected can be seen in Table 1.

3.3 Machine learning

The machine learning aspect of this study can be broken up into three steps: generating training data, running that data through one of two RFs, and producing hydrostatic models based on the RF's predictions.

3.3.1 The model grids

Two grids were used as training data for the random forests. They differ from one another both in which physical processes are considered as well as in the ranges that certain parameters are allowed to take. For the 'canonical grid' the following parameter ranges are allowed:

(i) Mass: [0.2–6.0] $M_\odot$
(ii) Initial metallicity: [1e-4–0.03]
(iii) Mixing length parameter: [1.2–2.7]
(iv) Overshoot (USO): [1e-6–0.05]

For the 'extended grid' the parameter ranges are:

(i) Mass: [0.2–6.0] $M_\odot$
(ii) Initial metallicity: [1e-4–0.03]
(iii) Mixing length parameter: [1.566–1.914]
(iv) Overshoot (USO): [1e-6–0.05]
(v) Effective spot coverage: [1e-3–1.0]
(vi) α-enhancement: [0.0, 0.2, 0.4, 0.6] [2]

Note that in both grids diffusion is considered for stars below 1.1 $M_\odot$, but we do not vary the diffusion efficiency. Additionally, the initial helium abundance in both grids is determined from the metallicity via the Y–Z scaling relationship with a slope of 1.4 (Peimbert, Carigi & Peimbert 2001; Balser 2006).

3.3.2 Generating training data

As a first step to any supervised machine learning algorithm, one must acquire the training data from which the algorithm will learn. The training data were produced by generating a grid of 5000 stellar tracks, each with varying initial parameters. The tracks were evolved until they reached the helium flash or an age of 16 Gyr. A separate grid of stellar tracks is generated for both the canonical and extended grids.

The grid of initial conditions was constructed via the Sobol sequence, a quasi-random number generation scheme which picks out points in the hyperrectangle which encloses the parameter space. This method fills the parameter space as quickly and evenly as possible (Bellinger et al. 2016).

With the initial conditions for the 5000 tracks selected and the tracks run, the next step is to combine all the data. One must be careful to sample data points from the tracks intelligently to avoid biasing the random forest towards the phases of evolution which require more models to be computed. To do this, we use the Hungarian algorithm (Kuhn 1955) in the same way as in Bellinger et al. (2016). The algorithm selects evenly sampled points based on central hydrogen content for the main sequence and based on age for the post-main sequence. In the end, one has 64 main sequence data points and 64 post-main sequence data points which are as close to evenly spaced as is possible. If the star never reaches the post-main sequence then 128 main sequence points are selected.

Up to this point both random forest methods involve the same steps, but this is the point where they diverge. As such we will first discuss the single star random forest method. The section following that will be dedicated to the binary random forest method.

[2]The α-enhancement values are assigned based on the metallicity as described by equation (3)

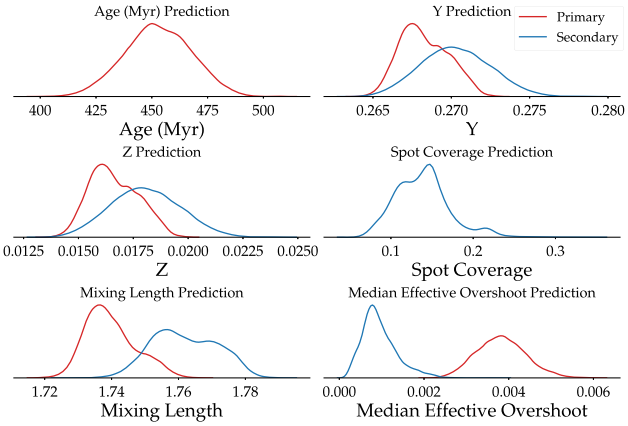


Figure 3. An example of the PDFs generated by the random forest when making a prediction for each of the 10 000 realizations of the observations. Shown here are the PDFs for the YZ Cas system.

3.3.3 Single star random forest

As it is, the data are ready to use in the single star random forest method. Further work must be done to prepare the data for the second method. This will be discussed in the next section (Section 3.3.4).

The single star random forest (SSRF) method is an RF regressor which fits and predicts the targets for the two components individually. Before training the forest, the features and targets for the regressor must be chosen. The selected features are mass, radius, effective temperature, metallicity, luminosity, and surface gravity. The targets are age, initial helium content, initial metallicity, mixing length, median effective overshoot, and effective spot coverage. Additionally, the age of the primary is passed to the secondary as a feature since being co-natal is a known constraint of the problem.

Before the data are passed to the regressor it is first passed through two RF classifiers which label the star based on its evolutionary phase (main sequence, sub-giant branch, or red-giant branch) and whether or not it has a convective core. The training data for the regressor is afterwards limited to models in this evolutionary phase plus the next nearest evolutionary phase and with that type of core. Doing this simplifies the parameter space which the regressor has to regress over.

At this point, the observed quantities of the target system are passed to the random forest so it can make predictions. However, instead of passing just the mean for each measurement of the system and having the RF make one prediction, we assumed a Gaussian error for the measurements, and we used the measurement uncertainties to generate 10 000 realizations of the measurements which are fed to the RF. The RF in turn makes 10 000 predictions for each of the targets for the system. This yields a PDF for each target and in turn an estimation of the error on our predictions. In Fig. 3 an example of such a distribution of predictions is shown. In this way, we can propagate the statistical error through the RF. Because of the systematics we consider one should expect that the uncertainties of this method will be larger than those of other traditional methods.

3.3.4 Binary random forest

The binary random forest (BRF) differs from the SSRF in one key way: the BRF analyses both components of the system at the same time. For this purpose, the training data have to be prepared differently.

For both the SSRF and BRF method the training data initially consist of evenly sampled points from all 5000 tracks. Now this data must be used to generate training data for binary systems. In other words, each data entry in the new training data set must contain information about two stars instead of one. We accomplish this in the following way. One model is picked from the training data at random and is assumed to be the primary. Then, the remaining data are parsed to look for potential secondaries for this star. The criteria used to select the secondaries are

(i) $M_{\rm primary} > M_{\rm secondary}$.
(ii) Both components must have the same α-enhancement.
(iii) The age of the secondary must be within 10 per cent of the age of the primary.
(iv) The initial metallicity of the secondary must be within 10 per cent of the metallicity of the primary.

With the allowed secondaries found, a data entry for each combination of this primary with one of the secondaries is made. This is done for every model of the original training data until a new set of binary training data set is created. It is clear that in this way we force the training data systems to be co-natal and of similar initial chemical composition and, by extension, will force the predictions to be as well. This is due to our underlying assumption about the nature of the binary systems. Since these constraints are imposed, bad predictions will be identified via the validation procedure (Section 3.3.5) instead of by checking for whether the components are co-eval as is usually done in the traditional methods. Having done this, though, the even sampling which has been achieved up to this point is destroyed. Therefore, a new method must be employed to regain the even sampling.

To evenly sample the binary data, the following algorithm was applied. First, an empty histogram for each parameter is prepared with a pre-defined number of bins and a maximum allowed content for each bin. Next, binary data points are selected randomly and the bins which the parameters would fit in are checked. If the inclusion of this data caused any of the parameter bins to overflow, the binary system is thrown away; if not, the binary system is kept. In the end, the histograms will be as flat as possible, and the data will be evenly sampled.

The evenly sampled BRF training data can now be used to train the BRF in the same way as in the SSRF method. The features are mass, radius, effective temperature, metallicity, luminosity, and surface gravity for each component, and the targets are age, initial helium content, initial metallicity, mixing length, median effective overshoot, and effective spot coverage for each component.

As before, we add Gaussian noise to the measurements and feed the 10 000 points to the BRF to get a PDF for each target. For this method no pre-regression classifiers are used as it was determined they did not affect the quality of the predictions.

3.3.5 Running validation tracks

At this stage, we have a PDF for the age and the modelling parameters for a system, but we do not know which specific values will lead to the best-fitting hydrostatic model. Furthermore, RFs, like all machine learning algorithms, are fallible, and it is good practice to test the predictions to ensure they are indeed accurate. This process involves running evolutionary tracks, which we call 'validation tracks', based on the RF's predictions.

Two different approaches were used for this purpose. For the SSRF and the BRF canonical-grid results, a straightforward method was utilized. Parameters for the validation tracks were drawn from the distributions which the RFs produced and evolutionary tracks with those parameters were run. The best tracks were chosen based on a reduced χ^2 function. Measured luminosity ratio, effective temperature ratio, and radii of the components are treated as the expected values and compared to the values from appropriate model pairs from the validation tracks. By appropriate model pairs, we mean model pairs which contain one model from the primary and one model from the secondary which are close to each other in age. The ratios of luminosity and effective temperature are used instead of the individual values because the ratios are directly determined from the DDLEB light curves, whereas the individual values require modelling and assumptions to obtain (Higl & Weiss 2017). Hence, the ratios are the more dependable values.

The second approach is more sophisticated and involves an optimization search which is constrained to the region of the parameter space corresponding to the distribution of predicted values from the RF. Since the optimization search is heavily constrained compared to the full parameter space, the computational cost of an optimization search, which was discussed earlier, is not so overbearing. The optimization search looks for the modelling parameters which result in evolutionary tracks which minimize the objective function, which is the reduced χ^2 function already discussed for the previous approach. This approach was used on the results from the BRF trained on the extended grid. The optimizer is part of the HEPHAESTUS code (Jørgensen & Angelou 2019) which was updated for this study.

The optimization algorithm used in the parameter search was a genetic algorithm. Genetic algorithms abstract concepts from Darwin's theory of natural selection to a generalized optimization problems (Charbonneau 1995). By viewing sets of parameters as chromosomes and the individual parameter values as genes, one is able to view the optimization problem through the lens of evolution. The algorithm incorporates three key aspects of evolution: fitness, inheritance, and variation. The fitness of a chromosome is the value of the function to be optimized when it is evaluated with the genes of that chromosome. In our case, the fitness is the value of the reduced χ^2 function of the tracks when run with a particular parameter set. Inheritance is done by finding the most fit chromosomes in a population and selecting, randomly or otherwise, genes from the parents as the genes for the offspring. This step is also referred to as genetic crossover or recombination. Finally, variation is handled by setting a certain mutation probability and mutating/changing a gene's value based on that probability. For more information on genetic algorithms, Goldberg (1989) is a standard textbook on the subject and Metcalfe et al. (2009), Bischoff-Kim & Metcalfe (2011) describe the application of a genetic optimizer in an astrophysical context. A downhill simplex algorithm was also tested, but the genetic optimizer proved to be more suitable for our purposes.

4 RESULTS

As a first step in presenting the results we will discuss the general agreement between the methods and grids. The full tables of predictions and best-fitting validation/optimization models for both methods and both grids can be found in Appendix A. Table A2 shows the results for both methods using the canonical grid and Table A1 shows the results for both methods using the extended grid.

Both methods, regardless of the grid being used as training data, are able to find a solution to a majority of the systems analysed. Based solely on the reduced χ^2 values, the results are as follows: The canonical SSRF fits 17 systems well, the canonical BRF fits 13 systems well, the extended SSRF fits 18 systems well, and the

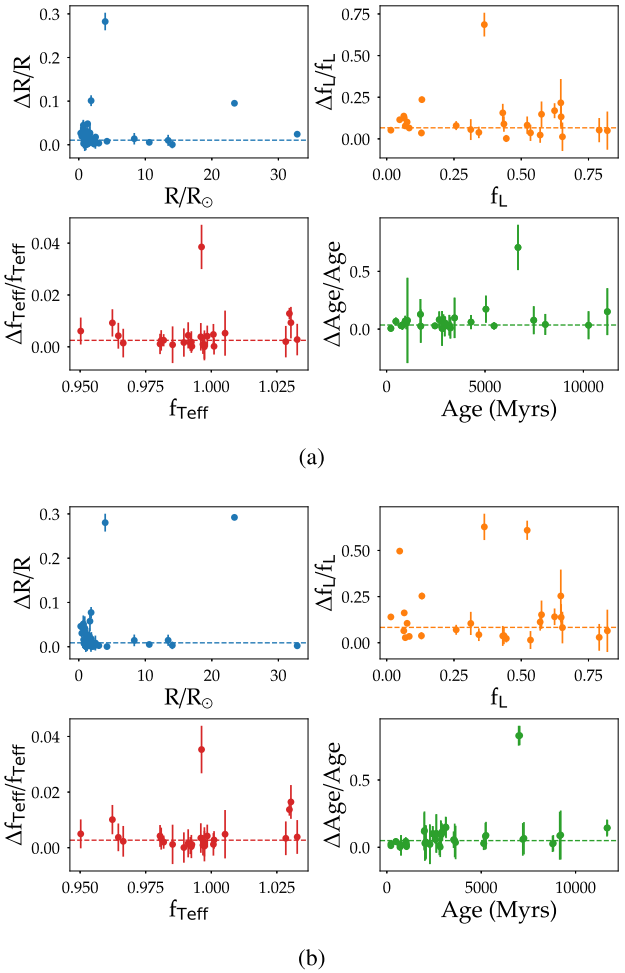


Figure 4. Relative error scatter plots for both the SSRF (a) and BRF (b) methods using the canonical grid. The data points are the relative error of the best-fitting model to the observations. The error bars correspond to the observational error on the quantity. f_L is the luminosity ratio and is defined as $f_L = \log(L_2/L_\odot)/\log(L_1/L_\odot)$. Similarly, the effective temperature ratio is defined as $f_{Teff} = \log(T_{eff2}/T_{Teff1})$. For the age plots, the predicted age from the age distribution and the 1σ deviation of the predicted distribution are treated as the 'observed' value and error, respectively. The dashed line in each plots shows the median absolute deviation of the data.

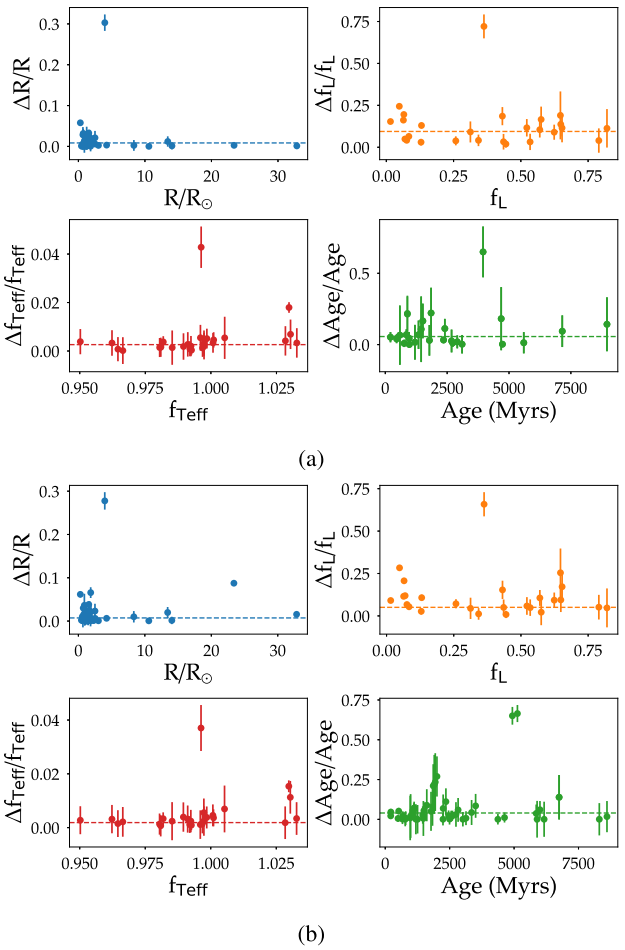


Figure 5. As Fig. 4, but for the extended grid.

extended BRF fits 19 systems well. Whether these solutions are valid is a question we will come to later. In Figs 4 and 5 one can see the relative error of the best-fitting models to the observational/predicted quantities, radius, luminosity ratio (f_L), effective temperature ratio (f_{Teff}), and age, for all methods and grids for all systems. The error bars are the 1σ measurement errors. For the age plots (lower right-hand plot in each panel) in Figs 4 and 5, the age of the best-fitting model is compared to the prediction from the RF to determine an error and the error bars denote the 1σ region of the age prediction. If the method were perfect then all the data points would line on the $y = 0$ line of these plots. The dashed line in the plots shows the median absolute deviation of the data to visualize how much the data deviates from the $y = 0$ line. It is clear that there are a few very strong outliers which pull the line away from $y = 0$. However, most of the data points do in fact lie along $y = 0$, particularly if the error bars are taken into consideration.

Despite the general success of the methods and grids, there are clear differences in the predictions for certain systems. The differences between the grids are largely what one would expect. The canonical grid's validated results are generally older and have a different metallicity as compared to the extended grid. This is not surprising since the extended grid has both spot coverage and α-enhancement included. The spots inflate the radius causing the star to reach the observed radius earlier in its evolution and the α-enhancement leads to different Z values for the same [Fe/H]. Furthermore, the additional parameters in the extended grid allow for solutions to be found for systems which the canonical grid struggles with, e.g. IM Vir, V530 Ori, V636 Cen, and YZ Cas. In most cases, the spot coverage is the key parameter which allows for a solution to be found. For YZ Cas, though, the α-enhancement is the deciding factor (see below).

Comparing the SSRF and BRF methods, two differences in the predictions stand out. First, the BRF generally predicts higher spot coverage, particularly for the secondary. Secondly, the initial metallicity predictions are often different between the methods. The difference is not only in the actual value predicted but also in the difference in metallicity between the components. The SSRF allows for a larger difference in metallicity between the components (see, for example, the predictions for IM Vir with the canonical grid). This is not a coincidence, since no constraints were placed on the metallicity difference between the components for the SSRF method, whereas in the BRF method, a metallicity difference constraint was imposed when combining the single star data into binary system data.

Table 2. The predicted parameters from the random forests as well as those of the best-fitting validation tracks for both methods and both grids for YZ Cas.

Method	Component	Age (Myr)	Y	Z	$\alpha_{\rm MLT}$	MEO	f_e	$M_*/M_\odot$	α-enhancement	Reduced χ^2
Canonical Grid										
SSRF										
Predictions	Primary	449 ± 16	0.267 ± 0.002	0.0156 ± 0.0015	1.954 ± 0.038	0.00423 ± 0.00061	–	–	–	–
	Secondary	449 ± 16	0.272 ± 0.003	0.0193 ± 0.0023	2.126 ± 0.099	0.00067 ± 0.00023	–	–	–	–
Validation	Primary	501	0.267	0.0165	1.925	0.00416	–	2.287	–	44.008
	Secondary	501	0.273	0.0182	2.130	0.00065	–	1.337	–	44.008
BRF										
Predictions	Primary	485 ± 10	0.265 ± 0.001	0.0145 ± 0.0008	1.876 ± 0.039	0.00324 ± 0.00068	–	–	–	–
	Secondary	488 ± 10	0.265 ± 0.001	0.0145 ± 0.0008	1.956 ± 0.024	0.00082 ± 0.00015	–	–	–	–
Validation	Primary	498	0.267	0.0142	1.788	0.00313	–	2.286	–	12.286
	Secondary	498	0.267	0.0141	1.917	0.00077	–	1.331	–	12.286
Extended Grid										
SSRF										
Predictions	Primary	447 ± 14	0.268 ± 0.002	0.0166 ± 0.0011	1.738 ± 0.007	0.00326 ± 0.00051	0.00 ± 0.00	–	–	–
	Secondary	447 ± 14	0.270 ± 0.002	0.0176 ± 0.0015	1.733 ± 0.022	0.00041 ± 0.00032	0.12 ± 0.05	–	–	–
Validation	Primary	478	0.270	0.0149	1.742	0.00319	0.00	2.287	0.2	1.267
	Secondary	475	0.267	0.0156	1.713	0.00035	0.10	1.326	0.2	1.267
BRF										
Predictions	Primary	497 ± 11	0.266 ± 0.001	0.0152 ± 0.0006	1.761 ± 0.010	0.00368 ± 0.00058	0.00 ± 0.00	–	–	–
	Secondary	498 ± 10	0.266 ± 0.001	0.0151 ± 0.0006	1.729 ± 0.007	0.00085 ± 0.00017	0.15 ± 0.03	–	–	–
Validation	Primary	493	0.267	0.0152	1.745	0.00519	0.00	2.279	0.2	1.086
	Secondary	496	0.267	0.0144	1.730	0.00080	0.15	1.311	0.2	1.086

This is the reason that the SSRF occasionally outperforms the BRF. The SSRF has more freedom in a very impactful parameter, and it uses this freedom to find models which the BRF cannot. However, a binary system with two components whose initial metallicity are significantly different is not what one expects theoretically. Thus, if either method predicts that the components should have clearly different initial metallicities, that is a red flag which suggests that the predictions are not an accurate representation of the system in question even if the reduced χ^2 value for the system is very good. If one considers solutions where the difference in the metallicity of the components is $\lesssim 8$ per cent to be appropriate solutions and others to be invalid, then the total number of systems appropriately fit by each method becomes: 10 for the canonical SSRF, 12 for the canonical BRF, 12 for the extended SSRF, and 19 for the extended BRF.

There are a few systems which are of particular interest and warrant a more in-depth discussion. The first of these systems is YZ Cas.

4.1 YZ Cas

YZ Cas consists of a metallic-lined A2 primary ($2.263 \pm 0.012\,M_\odot$) and an F2 dwarf secondary ($1.352 \pm 0.007\,M_\odot$). The large mass difference between the components makes it an important test of our method. Pavlovski et al. (2014) note that stellar evolutionary models are unable to match the properties of the system. Photospheric measurements suggest a roughly solar composition while evolutionary models imply a metallicity of 0.009. They find models with the lower metallicity yield an age estimate of 490–550 Myr. Lastennet & Valls-Gabaud (2002) ascertain that models with $Z = 0.018$, $Y = 0.293$, age = 446, and $Z = 0.02$, $Y = 0.280$, and age = 416, depending on the code used for the isochrones, produce accurate results. Claret & Torres (2018) also find two potential solutions depending on which element mixture they use. In one case they conclude the system has $Z = 0.01$ and an age of 525 and 565 Myr for the primary and secondary, respectively, and with another mixture they determine $Z = 0.012$ and an age of 492 Myr for the primary and 522 Myr for the secondary. Given the number of inconsistent results for this system it is worth looking at how the various methods and grids fair for this system.

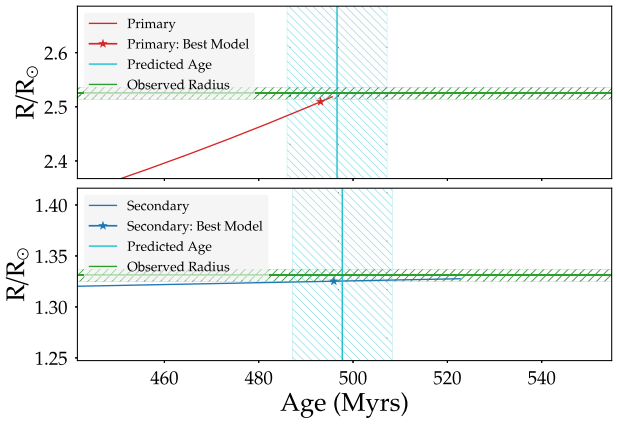
Looking at the results in Table 2, one notices that both methods are able to find a solution when using the extended grid, while neither method finds a good solution using the canonical grid. Since neither component is in a mass range where spot coverage plays an important role, these results imply that α-enhancement is likely important in modelling this system. This could be related to the unusual abundances of the system. Pavlovski et al. (2014) measured the abundances of 20 and 25 elements for the primary and secondary, respectively. They notice an overabundance in elements such as Na, Al, S, Zn, Y, Zr, and Ba for the primary. Additionally, the secondary appears to have an [Fe/H] $\simeq 0.05$ while the primary has an [Fe/H] $\simeq 0.3$. It is possible that increasing the α-enhancement mimics the effect of the overabundance of these mostly non α-elements in the primary. Additionally, in looking at the elemental abundances for the secondary, there does appear to be a slight overabundance of O ([O/H] $= 0.2 \pm 0.14$) and S ([S/H] $= 0.14 \pm 0.05$). This may explain why including α-enhancement improves the modelling of this system despite the fact that the system is likely not α-enhanced.

Focusing on the results from the BRF using the extended grid and the genetic algorithm optimizer, one can see the optimizer found the best solution to be one where the primary's metallicity is slightly higher than the secondary's. This is in agreement with the abundance analysis of Pavlovski et al. (2014).

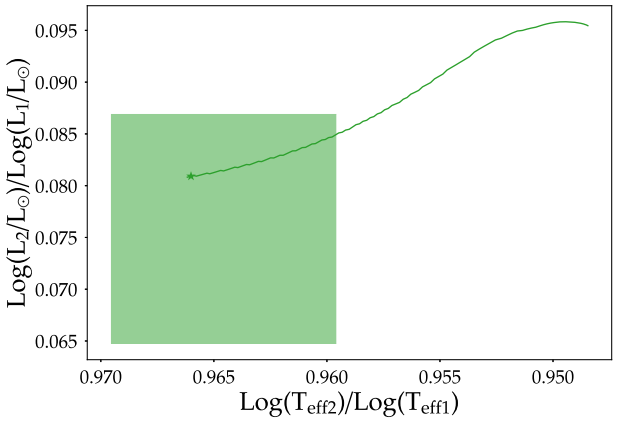
Our result is not fully consistent in both age and metallicity with any of the previous studies' results. The age we determined is in strong agreement with what was stated in Claret & Torres (2018), though they use a lower metallicity. That being said, our result is quite consistent with the observations of the system. The radii of both components are on the lower edge of the 1σ region at the optimized ages (Fig. 6a), which are themselves squarely in the 1σ region of the age prediction from the random forest. Additionally, both the luminosity ratio and effective temperature ratio for the optimized models are in the 1σ region (Fig. 6b).

4.2 IM Vir

IM Vir is a G7 + K7 binary system with a $0.981 \pm 0.012\,M_\odot$ primary and a $0.664 \pm 0.005\,M_\odot$ secondary. Morales et al. (2009) studied the



(a) YZ Cas



(b) YZ Cas

Figure 6. Age–radius diagram (top) and luminosity ratio versus effective temperature ratio plot (bottom) for the BRF optimized result for the YZ Cas system. Hatched or shaded regions represent the 1σ intervals for observed or predicted quantities, and the star marker denotes the best-fitting model.

spectra and light curves of the system and found both components to be larger, by 3.7 per cent and 7.5 per cent, and cooler than models predict making it a good candidate for testing our spot model. They determine, using standard models, that the primary agrees well with a $Z = 0.025$ ([Fe/H] = 0.15) track at an age of about 8 Gyr. Using these parameters, though, the secondary would be implausibly old. The authors then switch to a different set of models which use a correction factor which is applied to the radius (β) and effective temperature ($\beta^{-1/2}$) to account for the chromospheric effects ($\beta_1 = 1.035$ and $\beta_2 = 1.07$). Doing this, Morales et al. achieve a 'near-perfect fit' at an age of 2.4 ± 0.5 Gyr and [Fe/H] ≈ -0.3. They even estimate the spot coverage of the components from their observations and arrive at 5.2 per cent and 9.5 per cent spot coverage for the primary and the secondary, respectively. No errors are reported for these quantities.

In Table 3, one sees that when using the canonical grid, both methods struggle to find a good solution. The SSRF method comes close to a good solution but only when the initial metallicity of the secondary is set 10 per cent lower than the primary's. Using the extended grid, which allows for spot coverage, both methods find very good fits to the observations. The secondary's radius is matched almost perfectly at the optimized age, while the primary's radius is nearing the lower edge of the 1σ region (Fig. 7a). In Fig. 7(b) one can see that the optimized models match both the luminosity and effective temperature ratios from observations.

Nevertheless, the methods differ greatly in their results. The SSRF finds a solution for the system which is 0.3 Gyr younger but with a metallicity 0.06 higher than the BRF solution. However, the ages are actually in agreement well within 1σ and the metallicities are in agreement within 2σ. This is because the uncertainties in many of the key quantities, e.g. age and initial metallicity, are very large for this system. The source of this discrepancy is the very uncertain metallicity measurement for this system ([Fe/H] $= -0.1 \pm 0.25$). In order to better constrain the possible solutions for this system a more accurate metallicity measurement is required.

Together these results corroborate the fact that spot coverage is necessary to model this system. In fact, this system has among the highest spot coverage predictions of any system in this study. Notably, it is much higher than the estimation by Morales et al. (2009). To test what effect the spot coverage might have on the other predicted parameters, another set of predictions was made for this system. In this new set of predictions the spot coverage which the BRF was allowed to predict was restricted; it was not allowed to predict a spot coverage greater than 25 per cent. The BRF predicted the system then to be older and more metal rich with a spot coverage of 8 per cent for both components. This spot coverage is in line with the estimates of Morales et al. (2009). This parameter set does not match the observations as well as the other BRF results. In Fig. 8, one sees the primary's radius is matched well as are the system's luminosity and effective temperature ratios. However, the secondary's radius is roughly 2σ too small. This suggests the spot coverage might be underpredicted. Furthermore, restricting the spot coverage pushes the predictions for the other parameters, most notably the age, even farther from those of Morales et al. (2009).

4.3 KIC 9540226

KIC 9540226 is a binary with a primary ($1.39 \pm 0.031\,M_\odot$) on the red giant branch (RGB) and a main-sequence secondary ($1.015 \pm 0.016\,M_\odot$). Brogaard et al. (2018) compared the measured parameters to calculated theoretical isochrones and determined the age of the system to be 3.1 ± 0.6 Gyr. The difference in evolutionary stage between the primary and the secondary for this system makes it an interesting test of our methods, particularly for overshoot predictions. Furthermore, due to the evolutionary state of the primary it also shows the necessity of considering median effective overshoot.

In Table 4, one can see that both methods using both grids are able to find a solution for this system. Though the SSRF solutions have components whose initial metallicities differ by 10 per cent and 9 per cent for the canonical and extended grid, respectively. As such, these predictions should be met with scepticism.

Between the grids, the predictions from the BRF method differ. The age predictions are consistent within errors, but the metallicity predictions vary. The BRF using the extended grid predicts a higher metallicity as well as including an α-enhancement. Additionally, the mixing lengths for validated tracks are lower for the extended BRF. Furthermore, the extended BRF predicts the secondary to have a fairly high spot coverage. Ultimately, the quality of the fit is the same for both grids using the BRF method. The age result from the optimized models is in complete agreement with that of Brogaard et al. (2018).

Focusing on the BRF extended grid results, the models match the observations well. Specifically, despite the rapid increase of the radius of the primary along the RGB, the age–radius diagram (Fig. 9a) displays a convincing agreement, with the primary's radius only slightly above the 1σ region. Furthermore, the luminosity and effective temperature ratios of the models are in good agreement

Table 3. The predicted parameters from the random forests as well as those of the best-fitting validation tracks for both methods and both grids for IM Vir.

Method	Component	Age (Myr)	Y	Z	$\alpha_{\rm MLT}$	MEO	f_e	$M_*/M_\odot$	α-enhancement	Reduced χ^2
Canonical Grid SSRF										
Predictions	Primary	7414 ± 903	0.273 ± 0.006	0.0180 ± 0.0036	1.851 ± 0.161	0.00000 ± 0.00000	–	–	–	–
	Secondary	7414 ± 903	0.271 ± 0.005	0.0162 ± 0.0031	1.621 ± 0.092	0.00000 ± 0.00000	–	–	–	–
Validation	Primary	8053	0.274	0.0176	1.842	0.00000	–	0.965	–	1.371
	Secondary	8048	0.271	0.0159	1.479	0.00003	–	0.673	–	1.371
BRF										
Predictions	Primary	7783 ± 874	0.273 ± 0.004	0.0179 ± 0.0028	1.879 ± 0.091	0.00000 ± 0.00000	–	–	–	–
	Secondary	7851 ± 873	0.273 ± 0.004	0.0177 ± 0.0028	1.899 ± 0.048	0.00000 ± 0.00000	–	–	–	–
Validation	Primary	6782	0.266	0.0170	1.876	0.00000	–	0.997	–	2.514
	Secondary	6778	0.281	0.0164	1.825	0.00000	–	0.674	–	2.514
Extended Grid SSRF										
Predictions	Primary	5547 ± 1036	0.270 ± 0.004	0.0158 ± 0.0026	1.772 ± 0.018	0.00000 ± 0.00000	0.24 ± 0.08	–	–	–
	Secondary	5547 ± 1036	0.269 ± 0.004	0.0151 ± 0.0027	1.748 ± 0.023	0.00000 ± 0.00000	0.47 ± 0.09	–	–	–
Validation	Primary	5536	0.262	0.0211	1.753	0.00000	0.31	1.005	0.0	1.001
	Secondary	5536	0.277	0.0207	1.768	0.00000	0.38	0.667	0.0	1.001
BRF										
Predictions	Primary	5854 ± 942	0.271 ± 0.004	0.0166 ± 0.0024	1.732 ± 0.014	0.00000 ± 0.00000	0.23 ± 0.06	–	–	–
	Secondary	5817 ± 944	0.271 ± 0.004	0.0164 ± 0.0025	1.776 ± 0.015	0.00000 ± 0.00000	0.52 ± 0.05	–	–	–
Results	Primary	5809	0.262	0.0156	1.722	0.00000	0.26	0.972	0.2	1.000
	Secondary	5812	0.267	0.0156	1.789	0.00000	0.52	0.667	0.2	1.000
BRF (limited Spots)										
Predictions	Primary	7311 ± 722	0.271 ± 0.002	0.0164 ± 0.0016	1.738 ± 0.008	0.00000 ± 0.00001	0.09 ± 0.02	–	–	–
	Secondary	7224 ± 739	0.270 ± 0.003	0.0160 ± 0.0017	1.775 ± 0.020	0.00000 ± 0.00000	0.12 ± 0.01	–	–	–
Validation	Primary	7676	0.272	0.0160	1.743	0.00001	0.08	0.956	0.2	1.827
	Secondary	7679	0.274	0.0156	1.768	0.00000	0.08	0.671	0.2	1.827

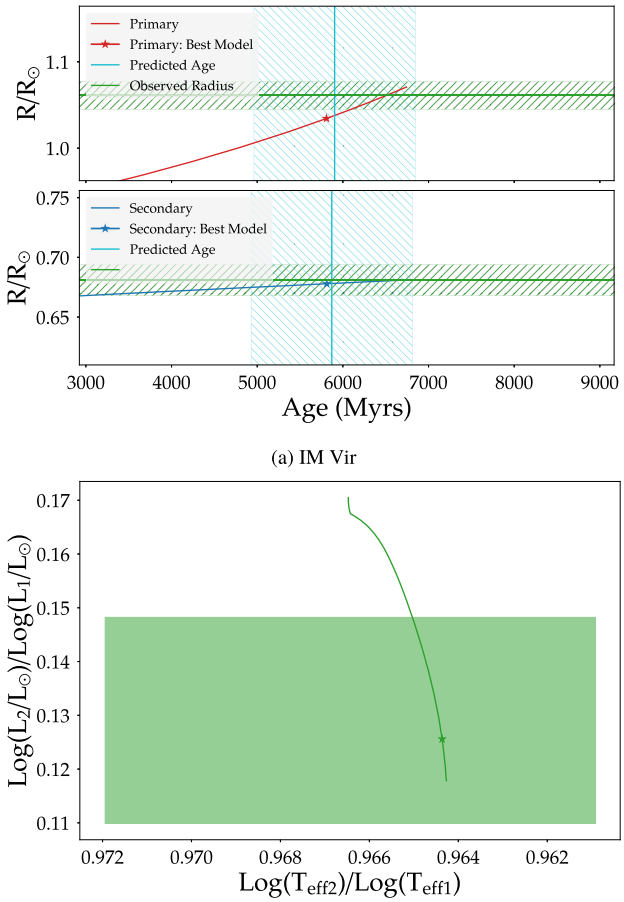
Table 4. The predicted parameters from the random forests as well as those of the best-fitting validation tracks for both methods and both grids for KIC 9540226.

Method	Component	Age (Myr)	Y	Z	$\alpha_{\rm MLT}$	MEO	f_e	$M_*/M_\odot$	α-enhancement	Reduced χ^2
Canonical Grid SSRF										
Predictions	Primary	3064 ± 257	0.258 ± 0.002	0.0095 ± 0.0015	1.834 ± 0.080	0.00156 ± 0.00040	–	–	–	–
	Secondary	3064 ± 257	0.265 ± 0.003	0.0125 ± 0.0021	1.689 ± 0.135	0.00000 ± 0.00000	–	–	–	–
Validation	Primary	3029	0.259	0.0103	1.893	0.00153	–	1.429	–	1.002
	Secondary	3024	0.261	0.0114	1.690	0.00000	–	1.015	–	1.002
BRF										
Predictions	Primary	3089 ± 298	0.259 ± 0.002	0.0096 ± 0.0015	1.966 ± 0.102	0.00121 ± 0.00083	–	–	–	–
	Secondary	3097 ± 294	0.260 ± 0.002	0.0096 ± 0.0014	1.969 ± 0.062	0.00003 ± 0.00004	–	–	–	–
Validation	Primary	3136	0.258	0.0096	2.055	0.00100	–	1.404	–	1.006
	Secondary	3136	0.259	0.0092	1.902	0.00000	–	1.019	–	1.006
Extended Grid SSRF										
Predictions	Primary	3030 ± 201	0.259 ± 0.002	0.0097 ± 0.0012	1.768 ± 0.018	0.00348 ± 0.00149	0.00 ± 0.00	–	–	–
	Secondary	3030 ± 201	0.265 ± 0.003	0.0128 ± 0.0020	1.771 ± 0.022	0.00000 ± 0.00000	0.22 ± 0.07	–	–	–
Validation	Primary	3095	0.258	0.0081	1.754	0.00394	0.00	1.378	0.2	1.000
	Secondary	3095	0.262	0.0089	1.835	0.00010	0.20	1.001	0.2	1.000
BRF										
Predictions	Primary	3288 ± 263	0.262 ± 0.002	0.0120 ± 0.0013	1.765 ± 0.021	0.00240 ± 0.00145	0.00 ± 0.00	–	–	–
	Secondary	3261 ± 260	0.263 ± 0.002	0.0119 ± 0.0013	1.724 ± 0.011	0.00003 ± 0.00004	0.37 ± 0.03	–	–	–
Validation	Primary	3210	0.263	0.0119	1.788	0.00058	0.00	1.413	0.2	1.005
	Secondary	3207	0.260	0.0122	1.709	0.00002	0.38	0.990	0.2	1.005

with observations. Even the regular Hertzsprung–Russell Diagram (Fig. 9c) shows a satisfactory fit to the secondary and the primary, which just passed 'the bump' on the RGB.

5 DISCUSSION

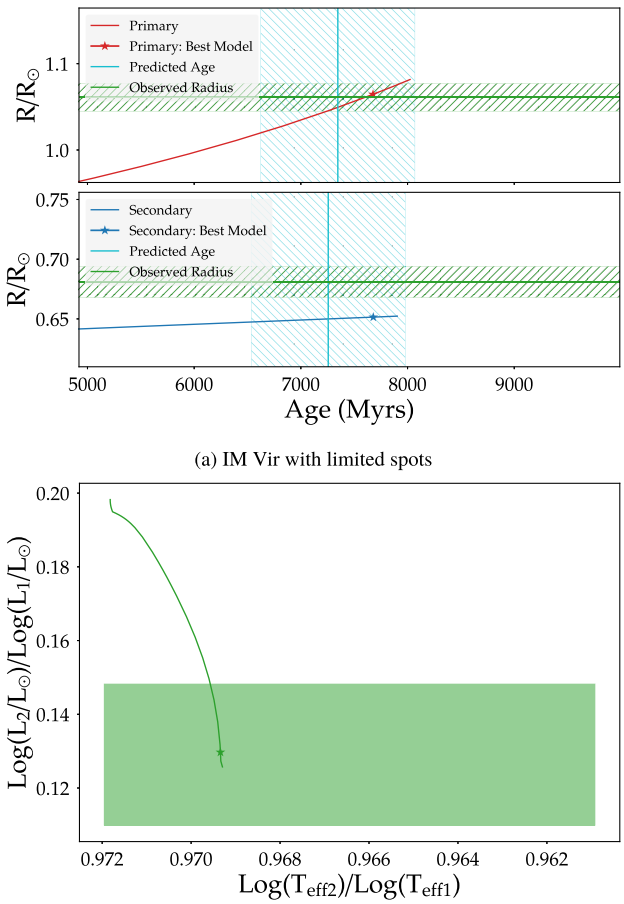
To evaluate the BRF method we have presented, at least three things need to be considered: effectiveness, reliability, and advantages. We will first discuss effectiveness, in other words, does the method work as intended? The primary concern is whether or not the RF is able to learn enough from the training data in order to create accurate regression models. The quality of the regression models determines the quality of the predictions for the validation tracks. As has been shown for a few systems, the RF is able to produce accurate regression models connecting the observational quantities to the targets of interest. It should be made clear that 'accuracy' in this context is judged by whether or not the predictions produce models which match the observations. It is also worth noting that the observed



(a) IM Vir

(b) IM Vir

Figure 7. As Fig. 6 but for the IM Vir system.



(a) IM Vir with limited spots

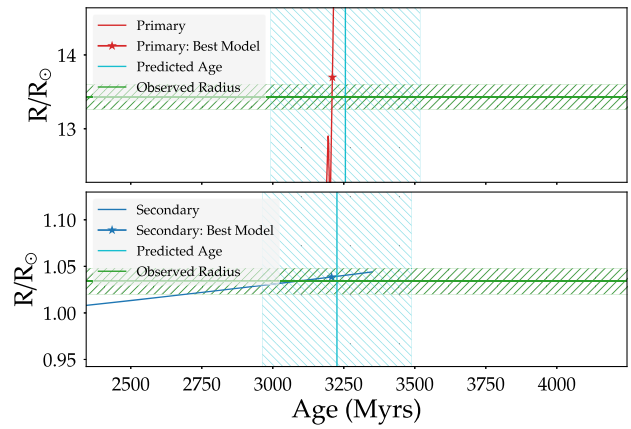
(b) IM Vir with limited spots

Figure 8. As Fig. 6 but for the IM Vir system with limited spot coverage.

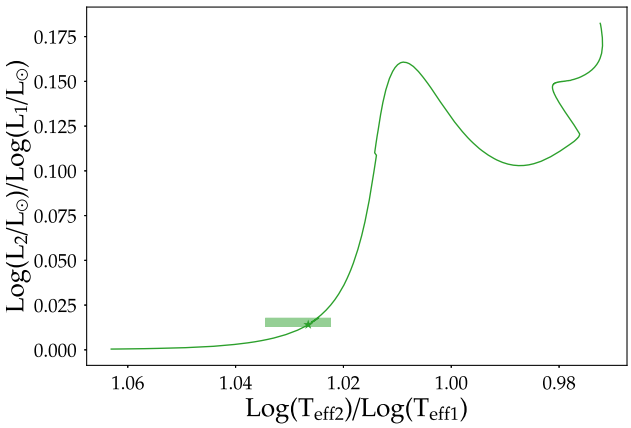
parameters themselves do not have a dependence on our models, however they may be dependent on certain other theoretical models. Thus, one has to either accept the errors as they are reported, attempt to asses the systematic uncertainties oneself, if one believes they are not already included, or ignore these parameters when determining the fit quality. In most cases we have chosen to accept the reported errors. The luminosity and effective temperature of the individual components have already been mentioned as quantities which are model dependent. This is why we have ignored their contribution to the quality of the fit and used the ratios for determining the best-fitting models instead. Finally, there is the possibility that the observed parameters depend on another stellar evolution code. This could be the case, for example, with stellar age determinations. However, we do not believe we use any such quantities in our study. For a more detailed look into how the observations and analysis of these systems is done see Andersen (1991), Torres et al. (2010).

The next aspect is reliability. As one can see by looking at Table A1, the method is quite reliable. For 19 of the 26 systems, the optimization models match the observations very well with the optimized parameters often being within 1 or 2σ of the predicted mean. However, there are some systems for which a good solution is not found. There are a number of reasons why a solution might not be found. One, the system may violate one of our assumptions. A number of assumptions are made about each system including that the components are co-natal and that they are and always have been detached and non-interacting, which may not be the case. Two, it could be that to properly model the star one has to consider the effect of a physical process which was not accounted for in the training data, such as rotation. Three, low-mass ($M_* \lesssim 0.8\,M_\odot$) component predictions are less reliable than higher mass component predictions. The tables in Appendix A show that often when a system has a component with a mass below approximately $0.8\,M_\odot$ all methods and grids are more likely to struggle to find a solution. One example of this is the system HAT-TR-318-007. One can see in Table A2 and Table A1 that none of the methods/grids are able to find a solution to this system. Furthermore, the age of the best-fitting model found does not match the predictions from the random forests. The other two systems which proved to be the most problematic for the extended BRF method were TZ For and OGLE-LMC-SC9-23065, both of which are more evolved systems. TZ For is infamous for being a difficult system to model (Higl et al. 2018). Thus, it is likely that that this system either violates one of our underlying assumptions or that there is a physical process at work which we do not account for. In the Higl et al. (2018) paper they state that using the Y–Z scaling relationship with a slope of 2 prevents the system from being fit and that a reduction of the slope to 1 was found to be necessary in order to fit the system. Thus, the slope of 1.4 used in this study may be too large for this system. In addition, the overshoot value which they find for the best-fitting models is higher than what is predicted by the RFs.

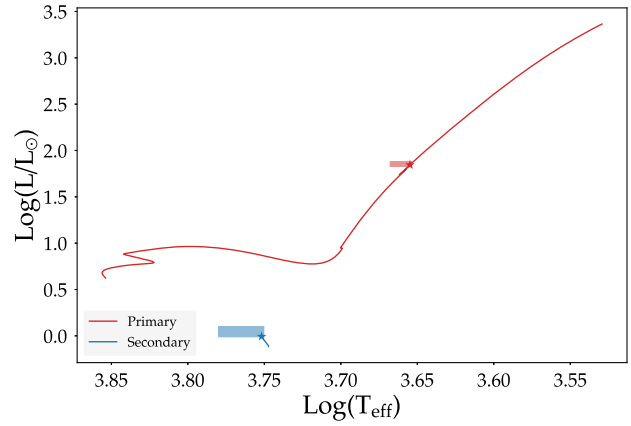
The reason that the BRF and SSRF age predictions are less accurate for stars in this mass range is because of the almost irrelevant evolution over 16 billion years, such that the RF is unable to distinguish between the star when it is young and the star when it is



(a) KIC 9540226



(b) KIC 9540226



(c) KIC 9540226

Figure 9. Top 2 panels are same as in Fig. 6 but for the KIC 9540226 system. Bottom panel is a Hertzsprung–Russell Diagram (HRD) for the BRF optimized result for the KIC 9540226 system. Coloured regions and star markers have the same meaning as in the other panels.

old. As an example, for a 0.37 $M_\odot$ star, the change in radius ($R/R_\odot$), luminosity ($\log L/L_\odot$), and effective temperature ($\log T_{\rm eff}$) over the course of 16 billion years is 0.016, 0.063, and 0.005, respectively. This corresponds to 4.5 per cent of the final radius, 3.1 per cent of the final luminosity, and 0.1 per cent of the final effective temperature. Increasing spot coverage by 3–4 per cent would cause an equal magnitude change to these variables. Given the small changes in the values of the features and the freedom the RF has to explain these small changes by altering other parameters such as metallicity or spot coverage, it is easy to see how the RF could interpret the change of these variables as simply noise. Consequently, the star would look identical to the RF throughout its evolution.

Finally, one must consider the advantages and disadvantages of this method, particularly in comparison to the already established SSRF method. One advantage of the SSRF method is that it is easier to identify failed systems. As explained in the previous section, if the SSRF predicts that the components of a binary system should have significantly different initial metallicities this is an indication that the method has failed to make quality predictions for this system. In the BRF method this is less likely happen due to the metallicity difference constraints that were imposed during the data processing stage. Therefore, one has to run validation tracks in order to determine if the predictions are good.

The BRF, though, has numerous advantages, not the least of which is that is sees the full observations of a system. Because the BRF can analyse the features for both components it should, in theory, be able to make more informed predictions about the targets of the system. Indeed this is the case. The number of systems for which an appropriate solution is found for each method are: 10 for the canonical SSRF, 12 for the canonical BRF, 12 for the extended SSRF, and 19 for the extended BRF. Thus, for both grids the BRF method outperforms the SSRF method. Additionally, the BRF method using the extended grid and the optimizer finds solutions to 50 per cent more systems than any of the other methods.

In comparison to the traditional methods, the main benefit of the BRF method is time. After the one-time cost of generating the training data, the BRF method can be used to predict the targets for a whole catalogue of stars in minutes as supposed to the hours or even days needed by the other methods to predict the parameters for a single star. The speed of this method means it can be leveraged to great effect in analysing some of the vast amounts of data generated from various sky surveys (Abazajian et al. 2009; Gaia Collaboration 2018). The usage of machine learning methods with various applications on the data sets of large surveys has become common practice. Domínguez Sánchez et al. (2018) use machine learning methods to determine the morphology of galaxies using data from the Sloan Digital Sky Survey, Bloom et al. (2012) use machine learning to detect transients and variable stars using the data from a multitude of surveys, and Castro-Ginard et al. (2018) use machine learning to search for open clusters using *Gaia* data to list but a few examples. As long as the data include the necessary parameters for the stars, our method can be used to generate PDFs for the parameters of thousands of DLEB systems quickly. Additionally, it is worth repeating that this method allows for a more honest appraisal of systematic and statistical errors than is possible with most traditional methods.

In addition to comparing methods, this study affords us the opportunity to discuss the prevalence and degree of the physical processes accounted for in the training data. Here we concentrate specifically on the amount of convective core overshoot: Since there is no reason to believe that overshoot is not present at all convective boundaries, one would expect all systems with a convective core to have a non-zero value of the MEO parameter. Indeed, this is reflected in the results for both methods. Moreover, high predicted MEO values were rare. In the training data, the USO parameter was allowed to be as high as 0.05, yet the highest MEO value obtained from the validation tracks is around 0.021.

A few systems in our study are useful for testing the quality of overshoot predictions. BG Ind is one such system where the primary

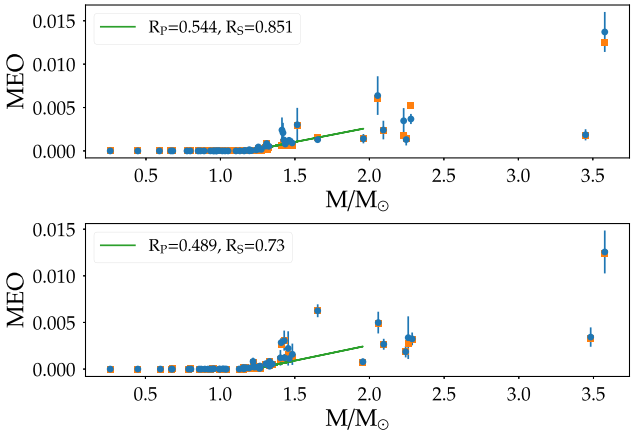


Figure 10. The median effective overshoot (MEO) parameter versus mass plots for the BRF method for both the extended grid (top) and canonical grid (bottom). The blue points and associated error bars are the mean and 1σ deviations from the predictions of the random forest. The orange points are the overshoot values of the corresponding best-fitting validation/optimization models. The green line shows a linear fit to the data in the mass range 1.2 $M_\odot$–2.0 $M_\odot$. The Pearson correlation coefficient, R_P, and Spearman correlation coefficient, R_S for this mass range are also reported.

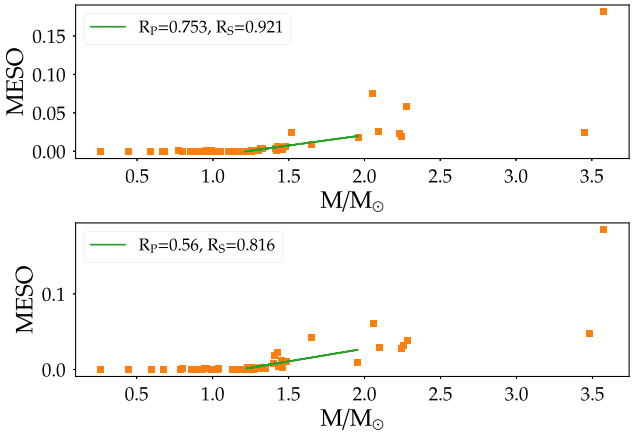


Figure 11. Same as Fig. 10 but for the median effective step overshoot (MESO) parameter. MESO was not predicted by the random forests so no error bars are plotted.

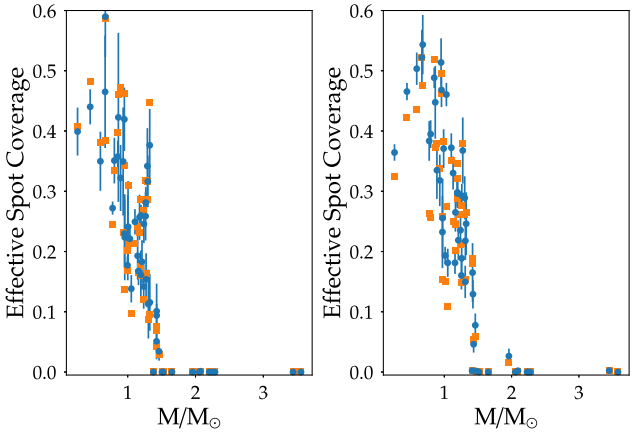


Figure 12. Effective spot coverage versus mass plots for both methods for the extended grid. The blue points and associated error bars are the mean and 1σ deviations from the predictions of the random forest. The orange points are the overshoot values of the corresponding best-fitting validation/optimization models. The left-hand panel is the SSRF and the right-hand panel is the BRF.

is at the end of the main sequence where overshoot should have a large effect on its evolution. BG Ind is matched well by both methods and grids with moderate amounts of overshooting in both components. For the BRF using the canonical grid the MEO values are 0.003 and 0.0004 for the primary and secondary, respectively, while using the extended grid the MEO values are 0.0006 and 0.0002. This trend of the primary having a higher MEO value than the secondary is present in most systems. The only exception is SW CMa where the secondary's MEO value is larger than the primary's. Furthermore, the greater the mass difference between the components, the greater the difference in MEO values. Other good test case systems, such as KIC 5640750, were not well matched by either method.

Another point which can be addressed by the overshoot data is the question whether a linear relationship between mass and overshoot is found (see discussion in Section 3.1.1). Fig. 10 shows plots of the MEO parameter value versus mass for the best-fitting tracks for each component for both methods. The plots also show a linear fit to the data in the mass range of 1.2 $M_\odot$ − 2.0 $M_\odot$. The Pearson correlation coefficient, R_P, and Spearman correlation coefficient, R_S, are also reported for this mass range. Based on the Spearman coefficients, there is indeed a positive correlation between mass and overshoot. However, given the Pearson correlation coefficient values as well as the uncertainties in the method, a linear relationship can be neither confirmed nor ruled out. Fig. 11 is the same plot but for the median effective step overshoot (MESO). This is calculated in the same way as was the median effective overshoot but the exponential overshoot parameter used in GARSTEC has been converted to a overshoot parameter for a step overshoot formalism. This is done to match the quantity which is plotted in other studies. One can see that the plots are quite similar, but with the step overshoot values being roughly 10 times higher. Importantly, the Pearson and Spearman coefficients are both higher in the case of Fig. 11. However, the Pearson correlation coefficient is still not large enough to confirm a linear relationship.

In addition to overshoot, spot coverage is also of particular interest in this study. Both methods predict some amount of spot coverage for all stars where spot coverage is possible. As one would expect, the low-mass stars are predicted to have a higher spot coverage, and after a certain mass the stars are predicted to have no spot coverage. These trends are plainly visible in Fig. 12. Furthermore, one can see that the spot coverage is predicted to be quite high for some systems (as high as 57 per cent). It remains to be seen how realistic these high spot coverage predictions are and how ubiquitous, or not, spots are in low mass systems. O'Neal et al. (2004) observed spot coverages of up to 42 per cent, which is lower than the highest spot coverage predicted in this study, but places the predictions within the realm of possibility. Given the simplicity of the spot model and the non-varied parameters included in equation (4), a rough agreement with observations and realistic values is all one can expect.

There are a few key systems in this study which are promising candidates for needing spot coverage, including FL Lyr. Lastennet & Valls-Gabaud (2002) were unable to find a satisfactory solution for this system which they suspected was due to a lack of a spot model in their study. Both methods used in our study were able to find solutions to this system by including spot coverage. IM Vir is also a notable example because of the study by Morales et al. which analysed this system with a spot coverage model. This allows us to compare our results with another result that also accounts for spots. Our methods obtain an age that is roughly twice as high as Morales et al. (2009) and

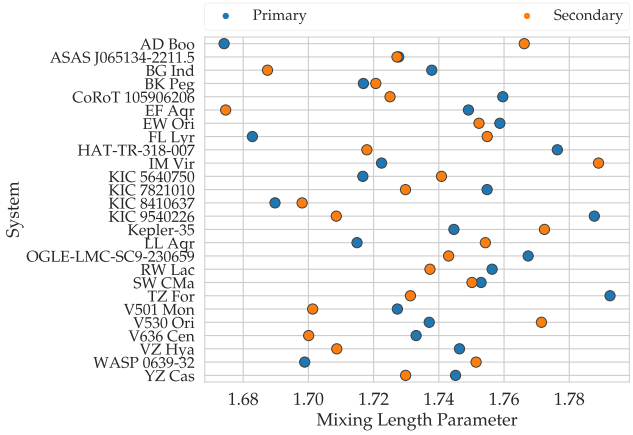


Figure 13. A plot of the mixing length parameters of the optimized models for the BRF method for every system. The points are colour-coded to denote which value is the primary and which is the secondary.

has a much higher spot coverage. Unfortunately, the large uncertainty in the metallicity measurements prevents constraining the solution further. Overall, the spot model allowed lower mass systems to be fit well, however, more work needs to be done in this area to better understand which systems have spot coverage and how much spot coverage is reasonable.

Spot coverage is a physically motivated parameter for lower mass systems so the fact that including it in our models improves the solution for certain systems is expected. However, for a number of systems in this study the validated tracks have a very high spot coverage. Extreme values for the predicted parameters is a possible indicator that something is amiss with the modelling. Spots are just one manifestation of plasma/magnetic physics observable at the surface. The fact that the spot coverage is very high for some systems suggests that other physical processes or parametrizations are necessary. For instance, it is possible that there are effects on the EOS that we are not considering or that the spot model we have implemented is too rudimentary.

As a final note on the parameters of the validated models of this study, the mixing length parameter will be briefly discussed. There are a number of studies (Trampedach et al. 2014; Magic, Weiss & Asplund 2015) which have calibrated the 1D mixing length parameter from 3D simulations. They find that the mixing length parameter varies with the star's parameters including effective temperature, surface gravity, and metallicity. Many studies (see Eggenberger et al. 2004; Joyce & Chaboyer 2018; Jørgensen & Angelou 2019, and others) have attempted to model α Cen and have found that the mixing lengths must differ between the components. Specifically, the secondary must have a higher mixing length than the primary. This motivated us to explore the mixing length values of the optimized models of the BRF method for our systems. In Fig. 13, one can see the mixing length parameter values for the primary and secondary of each system for the optimized validation tracks of the BRF method. The components of AD Boo have the highest difference between their mixing length parameters at roughly 0.1. Additionally, for 17 of 26 systems the mixing length parameter is smaller for the secondary than for the primary.

6 CONCLUSIONS

In this study, we have introduced a new machine learning approach for predicting the modelling parameters of DDLEB systems. The method uses a random forest to build regression models which relate the observable quantities of a system to the parameters needed as input for evolutionary tracks. The random forest is connected to an on-the-fly genetic optimizer to produce hydrostatic models based on the predictions. The method is unique because the random forest analyses both components of a system at the same time and thus has the full set of observables for the system from which to inform its predictions, hence why it is referred to as the binary random forest (BRF) method. A random forest method where only one component is analysed at a time, the SSRF method, is also used for comparison.

Two grids were used as training data: a canonical grid where mass, metallicity, core overshooting, and mixing length were varied and an extended grid where also α-enhancement and spot coverage were varied. Once trained, each RF was used to analyse 26 chosen detached double-lined eclipsing binary (DDLEB) system. The DDLEB systems were selected based on five criteria: masses of both components must be between 0.25 $M_\odot$ and 6 $M_\odot$, mass and radius measurements accurate to within 2 per cent, measured luminosities, effective temperatures, and metallicities, $q = M_2/M_1 < 0.96$, and no component can be on the pre-main sequence or past the helium flash. The random forests use the measured masses, radii, luminosities, effective temperature, surface gravity, and metallicities as inputs, or features, and predicts the initial helium content, initial metallicity, spot coverage, median effective overshoot, mixing length, and age of each component.

These machine learning methods have a number of advantages over existing approaches. First, RFs can be trained in seconds and are therefore much less computationally expensive than other methods. Secondly, these methods allow for an investigation of wider range and combination of parameters to be explored. Finally, by passing 10 000 Gaussian perturbed realizations of the measurements of the system to the random forest and having the random forest make a prediction for each, the statistical error is propagated through the method and an estimation of the statistical error on the predictions is possible.

We find that the BRF is unable to reliably predict parameters for stars under $\approx 0.8\,M_\odot$, because of their insignificant evolution with time, but for stars more massive than this the BRF method using the extended grid works a majority of the time. In total, the method finds a solution for 19 of 26 systems. For the stars which are above this mass and which are not fit well, it is possible that other physical processes, such as rotation, are needed to model them properly. It is also possible that these systems violate one of our assumptions (co-eval and no interaction).

One target of the RF was median effective overshoot (MEO). The BRF predicted some amount of overshoot for all stars which have a convective core. Moreover, particularly high MEO values were rarely predicted and rarely were needed in the best-fitting models. Furthermore, based on the MEO values from the optimized tracks for each system, we find a positive correlation between mass and core overshoot, but whether that relationship is linear or not remains unanswered.

Another target of particular interest was spot coverage. The inclusion of spot coverage allowed us to fit systems which previous studies had failed to fit, for example, FL Lyr. In the case of IM Vir, we also obtained very different results compared to a previous study where spot coverage was also taken into account. Even limiting spot coverage to more align with what was used in the previous study does not reconcile the differences in predictions.

Ultimately, the new approach is successful but there is room for improvement. Specifically, improving the results for low-mass stars would be important.

ACKNOWLEDGEMENTS

This reasearch was supported by the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094 - 390783311.

DATA AVAILABILITY

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

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Andersen J., 1991, A&AR, 3, 91
Angelou G. C., Bellinger E. P., Hekker S., Basu S., 2017, ApJ, 839, 116
Angelou G. C., Bellinger E. P., Hekker S., Mints A., Elsworth Y., Basu S., Weiss A., 2020, MNRAS, 493, 4987
Balser D. S., 2006, AJ, 132, 2326
Bazot M., Bourguignon S., Christensen-Dalsgaard J., 2012, MNRAS, 427, 1847
Bellinger E. P., Angelou G. C., Hekker S., Basu S., Ball W. H., Guggenberger E., 2016, ApJ, 830, 31
Bellinger E. P., Hekker S., Angelou G. C., Stokholm A., Basu S., 2019, A&A, 622, A130
Bischoff-Kim A., Metcalfe T. S., 2011, MNRAS, 414, 404
Bloom J. S. et al., 2012, PASP, 124, 1175
Brogaard K. et al., 2018, MNRAS, 476, 3729
Carleo G., Cirac I., Cranmer K., Daudet L., Schuld M., Tishby N., Vogt-Maranto L., Zdeborová L., 2019, Rev. Mod. Phys., 91, 045002
Cassisi S., Potekhin A. Y., Pietrinferni A., Catelan M., Salaris M., 2007, ApJ, 661, 1094
Castro-Ginard A., Jordi C., Luri X., Julbe F., Morvan M., Balaguer-Núñez L., Cantat-Gaudin T., 2018, A&A, 618, A59
Chaboyer B., Demarque P., Guenther D. B., 1999, ApJ, 525, L41
Chaplin W. J. et al., 2014, ApJS, 210, 1
Charbonneau P., 1995, ApJS, 101, 309
Claret A., Torres G., 2016, A&A, 592, A15
Claret A., Torres G., 2017, ApJ, 849, 18
Claret A., Torres G., 2018, ApJ, 859, 100
Claret A., Torres G., 2019, ApJ, 876, 134
Constantino T., Baraffe I., 2018, A&A, 618, A177
Davis A., Jones S., Herwig F., 2019, MNRAS, 484, 3921
Denissenkov P. A., Herwig F., Bildsten L., Paxton B., 2013, ApJ, 762, 8
Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, MNRAS, 476, 3661
Eggenberger P., Charbonnel C., Talon S., Meynet G., Maeder A., Carrier F., Bourban G., 2004, A&A, 417, 235
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
Freytag B., Ludwig H. G., Steffen M., 1996, A&A, 313, 497
Gai N., Basu S., Chaplin W. J., Elsworth Y., 2011, ApJ, 730, 63
Gaia Collaboration 2018, A&A, 616, A1
Goldberg D. E., 1989, Genetic algorithms in search, optimization and machine learning, Addison-Wesley, USA
Hastie T., Tibshirani R., Friedman J., 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media, Berlin
Henghes B. et al., 2021, PASP, 133, 014501
Higl J., Weiss A., 2017, A&A, 608, A62
Higl J., Siess L., Weiss A., Ritter H., 2018, A&A, 617, A36
Hoxie D. T., 1973, A&A, 26, 437
Iglesias C. A., Rogers F. J., 1996, ApJ, 464, 943
Joyce M., Chaboyer B., 2018, ApJ, 864, 99
Jørgensen A. C. S., Angelou G. C., 2019, MNRAS, 490, 2890
Kippenhahn R., Weigert A., Weiss A., 2012, Stellar Structure and Evolution (second edition). Springer, Heidelberg
Kuhn H. W., 1955, Naval Res. Logist. Quart., 2, 83
Lastennet E., Valls-Gabaud D., 2002, A&A, 396, 551
Louppe G., 2014, PhD thesis, Université de Liège, Liège, Belgium
Magic Z., Weiss A., Asplund M., 2015, A&A, 573, A89
Mehta P., Bukov M., Wang C.-H., Day A. r. G. R., Richardson C., Fisher C. K., Schwab D. J., 2019, Phys. Rep., 810, 1
Meng Y., Zhang Q. S., 2014, ApJ, 787, 127
Metcalfe T. S., Creevey O. L., Christensen-Dalsgaard J., 2009, ApJ, 699, 373
Miller A., Kounkel M., Boggio C., Covey K., Price-Whelan A. M., 2021, PASP, 133, 044201
Morales J. C., Torres G., Marschall L. A., Brehm W., 2009, ApJ, 707, 671
Morel P., Thévenin F., 2002, A&A, 390, 611
Ness M. et al., 2013, MNRAS, 430, 836
O'Neal D., Neff J. E., Saar S. H., Cuntz M., 2004, AJ, 128, 1802
Pan Y., Fu J.-N., Zhang X., Wang J., Zong W., Li C., Zhang B., 2021, PASP, 133, 044202
Pattnaik R., Sharma K., Alabarta K., Altamirano D., Chakraborty M., Kembhavi A., Méndez M., Orwat-Kapola J. K., 2021, MNRAS, 501, 3457
Pavlovski K., Southworth J., Kolbas V., Smalley B., 2014, MNRAS, 438, 590
Peimbert M., Carigi L., Peimbert A., 2001, Astrophys. Space Sci. Suppl., 277, 147
Pols O. R., Tout C. A., Schröder K.-P., Eggleton P. P., Manners J., 1997, MNRAS, 289, 869
Quirion P.-O., Christensen-Dalsgaard J., Arentoft T., 2010, ApJ, 725, 2176
Rogers F. J., Nayfonov A., 2002, ApJ, 576, 1064
Salaris M., Weiss A., 2001, A&A, 376, 955
Salaris M., Chieffi A., Straniero O., 1993, ApJ, 414, 580
Schröder K.-P., Pols O. R., Eggleton P. P., 1997, MNRAS, 285, 696
Silva Aguirre V. et al., 2015, MNRAS, 452, 2127
Southworth J., 2015, in Rucinski S. M., Torres G., Zejda M., eds, ASP Conf. Ser. Vol. 496, DEBCat: A Catalog of Detached Eclipsing Binary Stars. Astron. Soc. Pac., San Francisco, p. 164
Spruit H. C., Weiss A., 1986, A&A, 166, 167
Stancliffe R. J., Fossati L., Passy J. C., Schneider F. R. N., 2015, A&A, 575, A117
Thoul A. A., Bahcall J. N., Loeb A., 1994, ApJ, 421, 828
Torres G., Andersen J., Giménez A., 2010, A&AR, 18, 67
Trampedach R., Stein R. F., Christensen-Dalsgaard J., Nordlund, Å., Asplund M., 2014, MNRAS, 445, 4366
Viani L. S., Basu S., 2020, ApJ, 904, 22
Weiss A., Schlattl H., 2008, Ap&SS, 316, 99
Zahn J. P., 1991, A&A, 252, 179

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Table S1 The predicted parameters from the random forests as well as those of the best-fitting validation tracks for both methods using the extended grid.

Table S2 The predicted parameters from the random forests as well as those of the best-fitting validation tracks for both methods using the canonical grid.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: TABULATED RESULTS

Here are the collated results for both methods and both grids for every system. Table A2 show the results for both methods using the canonical grid and Table A1 shows the results for both methods using the extended grid.

Table A1. The predicted parameters from the random forests as well as those of the best fitting validation tracks for both methods using the extended grid. The full table is available online.

System	Method	Age (Myr)	Y	Z	α_{MLT}	MEO	f_e	$M_*/M_\odot$	α-enhancement	Reduced χ^2
AD Boo	SSRF Predictions	1436 ± 134	0.272 ± 0.003	0.0196 ± 0.0019	1.725 ± 0.036	0.00036 ± 0.00016	0.10 ± 0.05	–	–	–
		1436 ± 134	0.272 ± 0.003	0.0191 ± 0.0020	1.747 ± 0.015	0.00001 ± 0.00000	0.18 ± 0.04	–	–	–
	SSRF Validation	1436	0.273	0.0191	1.669	0.00028	0.07	1.426	0.0	1.006
		1437	0.273	0.0194	1.736	0.00000	0.16	1.211	0.0	1.006
	BRF Predictions	1422 ± 80	0.274 ± 0.002	0.0206 ± 0.0014	1.713 ± 0.013	0.00209 ± 0.00115	0.16 ± 0.05	–	–	–
		1414 ± 77	0.275 ± 0.002	0.0210 ± 0.0015	1.753 ± 0.014	0.00014 ± 0.00009	0.30 ± 0.02	–	–	–
	BRF Validation	1418	0.272	0.0199	1.674	0.00067	0.19	1.420	0.0	1.000
		1417	0.274	0.0196	1.766	0.00007	0.29	1.194	0.0	1.000
ASAS J065134-2211.5	SSRF Predictions	5627 ± 416	0.275 ± 0.002	0.0187 ± 0.0016	1.749 ± 0.015	0.00000 ± 0.00000	0.23 ± 0.06	–	–	–
		5627 ± 416	0.275 ± 0.002	0.0193 ± 0.0014	1.749 ± 0.029	0.00000 ± 0.00000	0.59 ± 0.07	–	–	–
	SSRF Validation	5668	0.277	0.0211	1.733	0.00000	0.34	0.953	0.0	1.002
		5673	0.274	0.0204	1.765	0.00000	0.59	0.670	0.0	1.002
	BRF Predictions	5547 ± 328	0.275 ± 0.002	0.0190 ± 0.0014	1.725 ± 0.020	0.00000 ± 0.00000	0.26 ± 0.07	–	–	–
		5532 ± 323	0.275 ± 0.002	0.0189 ± 0.0014	1.751 ± 0.018	0.00000 ± 0.00000	0.54 ± 0.05	–	–	–
	BRF Validation	5641	0.271	0.0210	1.728	0.00000	0.15	0.974	0.0	1.006
		5641	0.279	0.0217	1.727	0.00000	0.48	0.680	0.0	1.006
BG Ind	SSRF Predictions	2745 ± 214	0.266 ± 0.002	0.0153 ± 0.0016	1.751 ± 0.036	0.00036 ± 0.00021	0.05 ± 0.03	–	–	–
		2745 ± 214	0.263 ± 0.003	0.0126 ± 0.0018	1.700 ± 0.024	0.00047 ± 0.00047	0.11 ± 0.06	–	–	–
	SSRF Validation	2746	0.266	0.0152	1.701	0.00033	0.04	1.425	0.2	1.016
		2747	0.267	0.0137	1.714	0.00034	0.09	1.305	0.2	1.016
	BRF Predictions	2511 ± 86	0.264 ± 0.001	0.0137 ± 0.0008	1.778 ± 0.021	0.00079 ± 0.00021	0.05 ± 0.01	–	–	–
		2488 ± 87	0.264 ± 0.001	0.0137 ± 0.0007	1.706 ± 0.017	0.00061 ± 0.00020	0.22 ± 0.03	–	–	–
	BRF Validation	2506	0.266	0.0132	1.738	0.00066	0.05	1.436	0.2	1.007
		2501	0.264	0.0127	1.687	0.00020	0.15	1.313	0.2	1.007
BK Peg	SSRF Predictions	2344 ± 52	0.269 ± 0.001	0.0172 ± 0.0008	1.719 ± 0.027	0.00085 ± 0.00027	0.09 ± 0.02	–	–	–
		2344 ± 52	0.269 ± 0.002	0.0170 ± 0.0012	1.738 ± 0.009	0.00005 ± 0.00002	0.28 ± 0.03	–	–	–
	SSRF Validation	2434	0.269	0.0157	1.667	0.00089	0.07	1.426	0.2	4.355
		2429	0.270	0.0163	1.712	0.00008	0.28	1.267	0.2	4.355
	BRF Predictions	2563 ± 78	0.269 ± 0.001	0.0175 ± 0.0007	1.718 ± 0.016	0.00127 ± 0.00029	0.13 ± 0.02	–	–	–
		2549 ± 82	0.270 ± 0.001	0.0176 ± 0.0007	1.725 ± 0.006	0.00024 ± 0.00004	0.29 ± 0.02	–	–	–
	BRF Validation	2542	0.270	0.0178	1.717	0.00093	0.18	1.426	0.2	1.776
		2544	0.271	0.0178	1.721	0.00017	0.28	1.260	0.2	1.776
CoRoT 105906206	SSRF Predictions	806 ± 32	0.275 ± 0.002	0.0216 ± 0.0017	1.722 ± 0.018	0.00080 ± 0.00036	0.00 ± 0.00	–	–	–
		806 ± 32	0.274 ± 0.003	0.0209 ± 0.0021	1.761 ± 0.021	0.00010 ± 0.00006	0.32 ± 0.04	–	–	–
	SSRF Validation	794	0.277	0.0203	1.701	0.00088	0.00	2.202	0.0	1.010
		793	0.277	0.0193	1.771	0.00004	0.32	1.297	0.0	1.010
	BRF Predictions	781 ± 34	0.274 ± 0.002	0.0210 ± 0.0016	1.739 ± 0.028	0.00347 ± 0.00147	0.00 ± 0.00	–	–	–
		778 ± 34	0.274 ± 0.002	0.0206 ± 0.0016	1.722 ± 0.014	0.00055 ± 0.00024	0.29 ± 0.04	–	–	–
	BRF Validation	786	0.274	0.0206	1.760	0.00176	0.00	2.230	0.0	1.026
		775	0.271	0.0199	1.725	0.00035	0.26	1.303	0.0	1.026

Table A2. The predicted parameters from the random forests as well as those of the best-fitting validation tracks for both methods using the canonical grid. The full table is available online.

System	Method	Age (Myr)	Y	Z	α_{MLT}	MEO	$M_*/M_\odot$	Reduced χ^2
AD Boo	SSRF Predictions	1928 ± 228	0.275 ± 0.004	0.0214 ± 0.0031	1.913 ± 0.145	0.00081 ± 0.00037	–	–
		1928 ± 228	0.274 ± 0.003	0.0205 ± 0.0023	1.956 ± 0.097	0.00068 ± 0.00055	–	–
	SSRF Validation	1939	0.283	0.0225	1.981	0.00076	1.413	1.004
		1938	0.280	0.0216	1.904	0.00056	1.221	1.004
	BRF Predictions	2029 ± 260	0.275 ± 0.003	0.0218 ± 0.0023	1.992 ± 0.159	0.00122 ± 0.00041	–	–
		2028 ± 255	0.276 ± 0.003	0.0214 ± 0.0023	1.914 ± 0.061	0.00082 ± 0.00040	–	–
	BRF Validation	1993	0.270	0.0214	2.116	0.00119	1.431	1.010
		1991	0.273	0.0209	1.867	0.00075	1.221	1.010
ASAS J065134-2211.5	SSRF Predictions	7629 ± 732	0.278 ± 0.002	0.0209 ± 0.0015	1.831 ± 0.100	0.00000 ± 0.00000	–	–
		7629 ± 732	0.278 ± 0.002	0.0209 ± 0.0015	1.640 ± 0.048	0.00000 ± 0.00000	–	–
	SSRF Validation	8392	0.274	0.0213	1.717	0.00000	0.944	8.520
		8388	0.283	0.0230	1.581	0.00007	0.684	8.520
	BRF Predictions	7790 ± 537	0.278 ± 0.002	0.0209 ± 0.0012	1.847 ± 0.091	0.00000 ± 0.00000	–	–
		7767 ± 532	0.278 ± 0.002	0.0210 ± 0.0015	1.928 ± 0.039	0.00000 ± 0.00000	–	–
	BRF Validation	8556	0.280	0.0186	1.886	0.00000	0.933	10.500
		8557	0.283	0.0183	1.872	0.00006	0.680	10.500
BG Ind	SSRF Predictions	2828 ± 215	0.265 ± 0.003	0.0145 ± 0.0022	1.887 ± 0.087	0.00109 ± 0.00048	–	–
		2828 ± 215	0.261 ± 0.002	0.0117 ± 0.0017	1.816 ± 0.186	0.00048 ± 0.00030	–	–
	SSRF Validation	2733	0.264	0.0133	1.869	0.00104	1.425	1.060
		2736	0.266	0.0120	1.700	0.00032	1.300	1.060
	BRF Predictions	2614 ± 137	0.263 ± 0.002	0.0131 ± 0.0015	1.840 ± 0.095	0.00305 ± 0.00108	–	–
		2602 ± 139	0.263 ± 0.002	0.0129 ± 0.0014	1.902 ± 0.079	0.00057 ± 0.00032	–	–
	BRF Validation	2676	0.263	0.0129	1.927	0.00304	1.429	2.723
		2676	0.266	0.0127	1.901	0.00048	1.307	2.723
BK Peg	SSRF Predictions	2526 ± 52	0.268 ± 0.002	0.0166 ± 0.0012	1.551 ± 0.058	0.00269 ± 0.00068	–	–
		2526 ± 52	0.265 ± 0.002	0.0146 ± 0.0015	1.557 ± 0.117	0.00017 ± 0.00008	–	–
	SSRF Validation	2527	0.266	0.0151	1.607	0.00262	1.428	1.007
		2522	0.269	0.0136	1.546	0.00013	1.252	1.007
	BRF Predictions	2578 ± 80	0.267 ± 0.002	0.0160 ± 0.0012	1.680 ± 0.065	0.00284 ± 0.00049	–	–
		2599 ± 80	0.267 ± 0.002	0.0157 ± 0.0012	1.901 ± 0.053	0.00033 ± 0.00014	–	–
	BRF Validation	2722	0.268	0.0158	1.712	0.00265	1.409	1.194
		2726	0.270	0.0143	1.834	0.00027	1.266	1.194
CoRoT 105906206	SSRF Predictions	793 ± 30	0.275 ± 0.003	0.0215 ± 0.0024	2.043 ± 0.089	0.00194 ± 0.00075	–	–
		793 ± 30	0.272 ± 0.004	0.0190 ± 0.0030	1.690 ± 0.154	0.00031 ± 0.00018	–	–
	SSRF Validation	777	0.279	0.0205	1.961	0.00191	2.220	1.110
		774	0.279	0.0196	1.805	0.00022	1.343	1.110
	BRF Predictions	755 ± 57	0.273 ± 0.003	0.0198 ± 0.0024	1.960 ± 0.130	0.00337 ± 0.00229	–	–
		753 ± 57	0.273 ± 0.003	0.0196 ± 0.0023	1.874 ± 0.048	0.00055 ± 0.00019	–	–
	BRF Validation	740	0.279	0.0206	2.050	0.00270	2.260	1.055
		740	0.275	0.0202	1.918	0.00053	1.349	1.055

This paper has been typeset from a TEX/LATEX file prepared by the author.