universität wien

# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Determining Fundamental Parameters of Red Giants Using Asteroseismic Quantities and Machine Learning"

verfasst von / submitted by

Eva Weiler BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2021 / Vienna, 2021

| | |
|---|---|
| Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet: | A 066 861 |
| Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet: | Master Astronomy |
| Betreut von / Supervisor: | Dr. Thomas Kallinger |

## Acknowledgments

First and foremost I would like to thank my supervisor Dr. Thomas Kallinger, who promoted my scientific autonomy by giving me plenty of scope. Whenever I had questions, he answered them as quickly as possible and took his time to discuss the more challenging ones via video calls. I am also very grateful for the unconditional support from my boyfriend Tobias, who encouraged me throughout the way and who additionally prove-read my thesis which gave it a final touch. Furthermore, I also thank my friends and particularly my roommates for cheering me up whenever I required it the most. My family deserve my heartfelt thanks, too, as the card-playing sessions on the weekends (which were not as frequent as I would have preferred due to the Covid-19 situation) provided the balance I needed to uphold my motivation and delight in working scientifically. I especially thank my parents who support me in every possible way and without whom I would not have been able to pursue my fascination for astronomy.

## Abstract

Radius and mass of solar-like oscillators are frequently derived by means of scaling relations for the large frequency separation, $\Delta\nu$, and the frequency of maximum oscillation power, $\nu_{\max}$. The asteroseismic quantities $\Delta\nu$ and $\nu_{\max}$ are derived from stellar oscillation spectra. They need to be well resolved in order to sufficiently well determine $\Delta\nu$, which becomes quite challenging for short observation periods.

The aim of this thesis is therefore to construct a model that does not depend on $\Delta\nu$ but on other asteroseismic quantities, which are easier to obtain. This way, an alternative method for estimating the radius and mass of solar-like oscillators is provided.

The model is created using a machine learning algorithm, namely a *Random Forest Regressor*, which is fitted to the seismic parameters of more than 6000 red giant stars. These stars were observed with the *Kepler* space telescope and cover different evolutionary stages from the red giant branch to the asymptotic giant branch.

In the end, three models are created: (1) the primary model is based on the seismic parameters $\nu_{\max}$, $\sigma$ and $p_{\text{tot}}$ as well as $T_{\text{eff}}$ and calibrated against the mass and radius that result from the seismic scaling relations, (2) a second model does not rely on temperature measurements as input, and (3) a third model requires also the stellar metallicity in addition to the input of the primary model. The prediction performance of the three models varies with the number of input parameters and give a $RMSE_S$ (i.e. scaled $RMSE$) ranging from 0.64 (model 2) to 0.58 (model 3) for the validation set.The models are tested using evolved solar-like oscillators, for which independent (i.e. non-seismic) mass and radius measurements exist. The first test set includes 24 solar-like oscillators, for which the asteroseismic data has been recently obtained from the ongoing NASA mission *TESS*. 16 stars can be compared with the radii and masses from Aguirre et al. (2020). The masses and radii of the remaining eight stars are estimated in this work for the first time. Furthermore, the models yield good results for six eclipsing binaries and nine metal-poor stars observed by *Kepler*, whereas model 1 systematically overestimates the true radii and masses of the latter. Finally, the average RGB and RC masses of 90 red giants, which belong to the open cluster NGC 6791 and NGC 6819 are predicted. I arrive at an average RGB mass of $(1.051 \pm 0.004)\,M_{\odot}$ and $(1.41 \pm 0.02)\,M_{\odot}$ for the RGB stars in the cluster NGC 6791 and NGC 6810, respectively.

The models work well for intermediate-mass stars, above all for stars with $R_{\text{true}} < 20\,R_{\odot}$. However, they fail in most cases to reproduce the lowest and highest mass stars from the initial data set as well as from the different test sets. Moreover, there seems to be a mass dependency for the radius predictions, as the predicted radii of less massive stars are rather overestimated while the radii of more massive stars are tendentially underestimated by the ML models. Similarly, the masses of low- and high mass stars are over- and underpredicted, respectively.

The models are able to reproduce the overall appearance of the true mass and radius distribution in the analyzed sample, with the different evolutionary stages clearly recognizable.

# Zusammenfassung

Radius und Masse von Sonnen-ähnlichen Pulsatoren werden häufig mithilfe seismischer Skalierungsrelationen der großen Frequenzseparation, $\Delta\nu$, sowie der Frequenz der maximalen Oszillationspower, $\nu_{\rm max}$, ermittelt. Die asteroseismischen Größen, $\Delta\nu$ und $\nu_{\rm max}$, werden über stellare Oszillationsspektren abgeleitet. Diese müssen hoch aufgelöst sein, um $\Delta\nu$ ausreichend gut bestimmen zu können, was mit kürzeren Beobachtungsperioden immer schwieriger wird.
Das Ziel dieser Masterarbeit ist daher, ein Modell zu erstellen, welches nicht von $\Delta\nu$ abhängt, sondern von anderen asteroseismischen Größen, die einfacher abzuleiten sind. Auf diese Weise wird eine alternative Methode geboten, um Radius und Masse von Sonnen-ähnlichen Pulsatoren abzuschätzen.
Das Modell wird mithilfe eines maschinellen Lernenalgorithmus, nämlich eines *Random Forest Regressors*, erstellt, welcher an die seismischen Parameter von mehr als 6000 Roten Riesen angepasst wird. Diese Sterne wurden mit dem *Kepler* Weltraumteleskop beobachtet und decken verschiedene Entwicklungsstadien vom roten bis zum asymptotischen Riesenast ab.
Letztendlich werden drei Modelle erstellt: (1) das erste Modell basiert auf den seismischen Parametern $\nu_{\rm max}$, $\sigma$ und $p_{\rm tot}$ als auch auf $T_{\rm eff}$ und wird mit den Radien und Massen, welche von den seismischen Skalierungsrelationen resultieren, verglichen, (2) ein zweites Modell hängt nicht von Temperaturmessungen ab, und (3) ein drittes Modell verlangt die stellare Metallizität zusätzlich zu den Inputparametern des ersten Modells. Die Vorhersagegenauigkeit der Modelle variiert mit der Anzahl der Inputparameter und erzielt für die Validierungssets $RMSE_S$ (d.h. skalierter $RMSE$) Werte von 0.64 (Modell 2) bis 0.58 (Modell 3). Die Modelle werden an Sonnen-ähnlichen Pulsatoren getestet, für welche unabhängige (d.h. nicht-seismische) Massen- und Radiusmessungen existieren. Der erste Test umfasst 24 Sonnen-ähnlichen Pulsatoren, deren asteroseismischen Daten erst kürzlich von der laufenden NASA Mission *TESS* gemessen worden sind. 16 Sterne können mit den Radius- und Massewerten von Aguirre et al. (2020) verglichen werden. Die Massen und Radien der übrigen acht Sterne werden in dieser Arbeit zum ersten Mal abgeschätzt. Die Modelle erzielen außerdem gute Resultate für sechs Doppelsterne sowie neun metallarme Sterne, welche wiederum mit *Kepler* beobachtet wurden, wobei Modell 1 die Radien und Massen von letzteren systematisch überschätzen. Als letzten Test werden die durchschnittlichen RGB und RC Massen von 90 Roten Riesen abgeschätzt, welche zu den offenen Sternhaufen NGC 6791 und NGC 6819 gehören. Modell 1 sagt eine durchschnittliche RGB Masse von $(1.051 \pm 0.004)\,M_\odot$ und $(1.41 \pm 0.02)\,M_\odot$ für die RGB Sterne in NGC 6791 beziehungsweise NGC 6819 voraus.
Die Modelle funktionieren gut für mittelschwere Sterne, vor allem für Sterne die $R_{\rm true} < 20\,R_\odot$ erfüllen. In den meisten Fällen versagen sie jedoch, die schwersten und leichtesten Sterne des ursprünglichen Datensatzes sowie der verschiedenen Testsets zu reproduzieren. Außerdem scheint es bei der Radiusvorhersage eine Massenabhängigkeit zu geben, da die vorhergesagten Radien von leichten Sternen eher überschätzt, während die Radien von massiveren Sternen tendenziell unterschätzt werden. Ebenso werden die Massen von massearmen sowie massereichen über- beziehungsweise unterschätzt.
Die Modelle können das Aussehen eines berechneten Masse-zu-Radius Plots grundsätzlich gut reproduzieren, vor allem die unterschiedlichen Entwicklungsstadien sind in diesem klar erkennbar.

# Contents

# 1 Introduction

Asteroseismology has proven to be a useful tool as it provides the unique opportunity of looking inside stars. This is achieved by observing the star's brightness, which changes periodically due to oscillations. Stars do not oscillate uniformly but the manner of the pulsations rather depends on the mechanism that drives the oscillations. In this thesis, I only consider stochastically excited oscillations, which are expected in so-called *solar-like oscillators*.
Asteroseismic quantities like the *large frequency separation* $\Delta\nu$ and the *frequency of maximum oscillation power* $\nu_{\mathrm{max}}$ are obtained from stellar oscillation spectra. As

$$\Delta\nu \propto \sqrt{\bar{\rho}} \propto \sqrt{\frac{M}{R^3}} \qquad \text{and} \qquad \nu_{\mathrm{max}} \propto g\, T_{\mathrm{eff}}^{-1/2} \propto \frac{M}{R^2\sqrt{T_{\mathrm{eff}}}}\,,$$

where $\bar{\rho}$ is the mean density and $g$ the surface gravity, one is able to estimate radius and mass of stars using these scaling relations. However, as observation periods shorten it becomes gradually more challenging to derive $\Delta\nu$ as opposed to $\nu_{\mathrm{max}}$ which is obtainable even if a star is observed for a rather short duration.
This master's thesis aims to estimate radius and mass without the explicit need of $\Delta\nu$ measurements. Therefore, a model is created by making use of the *Python* machine learning (ML) library *scikit-learn*. The ML algorithm used to create the model is the so-called *Random Forest Regressor*, which takes an ensemble of decision trees and averages the results of each tree. The model is applied to 6657 stars from the *Kepler Input Cataloge* (KIC) for which the seismic quantities $\nu_{\mathrm{max}}$, $\sigma$ and $p_{\mathrm{tot}}$ as well as the effective temperature $T_{\mathrm{eff}}$ are provided. As the latter is not always at hand, a second model is created that does not need $T_{\mathrm{eff}}$ as input. On the other hand there are sometimes metallicity measurements available, which denotes the fifth input parameter of a third model. Contrary to $\Delta\nu$, these seismic parameters are relatively easy to determine, even for short observations.

In the following sections, I will firstly summarize the basic principles of Asteroseismology. Hereafter, the evolution of Sun-like stars is briefly explained, since this is necessary for a better understanding of the plots that are frequently shown in the course of the analysis. As helioseismology lays the foundation for all achievements of solar-like oscillators, the whole of subsection 4.1 is dedicated to this scientific discipline. Sec. 4.2 deals with the oscillation spectra of solar-like oscillators, where the main emphasis is put on red giant stars. In this section the *nonlinear scaling relations*, which are stated above in simplified terms, are explicitly given.
Different observing strategies are outlined in Sec. 5. As the data set that is used to build the model was obtain by the *Kepler* space telescope, this space mission is described in more detail. The model is also applied to data, that were received with the *TESS* space telescope, which is why this mission is also discussed more thoroughly.
In Sec. 6 the data are finally introduced. More than 6000 KIC stars covering various evolutionary stages are provided. The data are used to fit the model, which aims to predict the radii and masses of previously unparameterized solar-like oscillators. The whole procedure of how this is achieved is outlined in Sec. 7. The results are finally presented in Sec. 8, where I analyze the performance of the three models which require a different number of input parameters and apply the models to data for which, at least in some extent, independent mass and radius measurements exist.

# 2 Asteroseismology

This section gives an overview of Asteroseismology and is a collection of the books 'Asteroseismology' by C. Aerts et al. (2010) and 'Stellar Structure and Evolution' by R. Kippenhahn et al. (2012). Some theoretical principles are also taken from lecture notes provided by Houdek (2005). Linear, adiabatic, and non-radial stellar oscillations are derived in Sec. 2.2. In the course of the derivation some basic principles of asteroseismology, such as the spherical harmonics (Sec. 2.2.1), the different oscillation modes of solar-like pulsators (Sec. 2.2.2) as well as the mechanism that drives theses oscillations (Sec. 2.2.3) are introduced.

## 2.1 Pulsating stars across the HR Diagram

Stars throughout the Hertzsprung-Russell (HR) diagram pulsate either periodically, semi-periodically, or irregularly. When it comes to variability, intrinsic and extrinsic variables must be distinguished. While extrinsic variables denote for example eclipsing binaries or rotating stars, which change their brightness due to obscuration as one star passes the other and solar spots, respectively, the pulsation of intrinsic variables is driven by different physical mechanisms, which categorize the stars in various pulsation classes (Fig. 1).

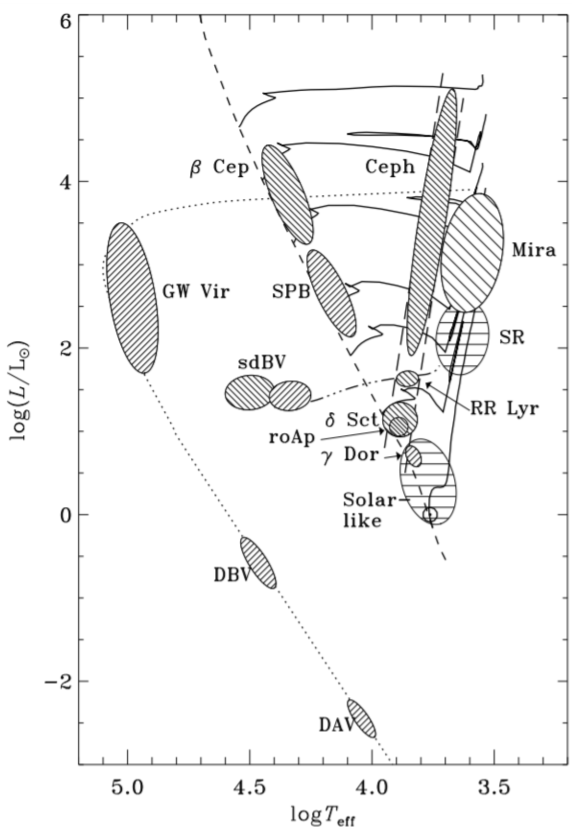



Figure 1: Different pulsation classes in the HR diagram. Figure taken from Cunha et al. (2007).

The brightness variations caused by intrinsic pulsations can be probed at the surface of stars by means of photometry or spectroscopy. Thus the frequencies with which the stars oscillate can be deduced. This has been intensively done for the Sun since the 1960's, when Leighton et al. (1962) first detected the five-minute oscillations on the surface of the Sun, which established a new scientific discipline, namely Helioseismology. Before this term as well as other solar-like oscillators are discussed more thoroughly, some basics will be explained since those are necessary for a better understanding of the subsequent sections.

## 2.2 Basic Principles of Asteroseismology

Since the amplitudes of the oscillations are much smaller than the typical scales of stars, stellar oscillations can be interpreted as small perturbations around a static equilibrium. The qualitative description of stellar oscillations is therefore achieved through perturbation theory.

To begin with, the continuity equations for mass, momentum, and energy are needed. The conservation laws are given in Eulerian coordinates because the physical quantities are described as a function of $\boldsymbol{r} = (r,\, \theta,\, \phi)$, which is the position vector of a given point in space, and time $t$. The Eulerian reference frame therefore correlates with the perspective of a stationary observer. From these conservation laws, which are – up to a simplification concerning the energy equation – stated below, the equilibrium equations of stellar structure can be deduced by neglecting the time derivatives and by assuming spherical symmetry

$$\frac{\mathrm{d}m_0}{\mathrm{d}r} = 4\pi r^2 \rho_0\,, \tag{1a}$$

$$\frac{\mathrm{d}p_0}{\mathrm{d}r} = -\frac{\mathrm{G}m_0\rho_0}{r^2}\,, \tag{1b}$$

$$\frac{\mathrm{d}T_0}{\mathrm{d}r} = -\frac{3\kappa_0\rho_0}{64\pi r^2\sigma T_0^3} L_0\,, \tag{1c}$$

$$\frac{\mathrm{d}L_0}{\mathrm{d}r} = 4\pi r^2 \rho_0 \epsilon_0\,. \tag{1d}$$

Where $r$ is the radius, $m_0$ is the mass, $\rho_0$ is the density, $T_0$ is the temperature, $L_0$ is the luminosity, G is the gravitational constant, $\kappa_0$ is the opacity and $\sigma_0$ is the Stefan-Boltzmann constant. As these are equilibrium quantities the subscript '0' is added.

The mass conservation which is also crucial for the derivation of stellar oscillations is given by

$$\frac{\partial \rho}{\partial t} + \mathrm{div}(\rho \boldsymbol{v}) = 0\,, \tag{2}$$

where bold variables indicate vectors. If we consider pressure and gravity as the only relevant forces acting on a star, the conservation of momentum can be written as

$$\rho\left(\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v}\cdot\boldsymbol{\nabla}\boldsymbol{v}\right) = -\boldsymbol{\nabla}p - \rho\boldsymbol{\nabla}\phi\,, \tag{3}$$

where $\phi$ satisfies the Poisson equation

$$\nabla^2\phi = 4\pi\mathrm{G}\rho\,. \tag{4}$$

The energy conservation, which is simplified by the adiabatic approximation, links $p$ and $\rho$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \Gamma_1 \frac{p}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t}\,, \tag{5}$$

with the 1st adiabatic exponent $\Gamma_1 = (\partial \ln p / \partial \ln \rho)$. This expression is valid because the pulsation period is comparable with the dynamical timescale (at least for p-modes)

$$\tau_{\mathrm{dyn},\odot} = \sqrt{\frac{R_\odot^3}{\mathrm{G}M_\odot}} \sim 20\,\mathrm{min}\,, \tag{6}$$

which in turn is much smaller than the thermal (or Kelvin-Helmholtz) timescale

$$\tau_{\mathrm{KH},\odot} = \frac{\mathrm{G}M_\odot^2}{R_\odot\, L_\odot} \sim 3\cdot 10^7\,\mathrm{yrs}\,, \tag{7}$$

where I put in solar values for clarification. Thus, oscillatory motions are considered adiabatic or, in other words, considered to proceed without energy exchange (Christensen-Dalsgaard, 2002).
If we add a small perturbation to the equilibrium quantity, the pressure, e.g., is written as

$$p(\boldsymbol{r},t) = p_0(\boldsymbol{r}) + p'(\boldsymbol{r},t)\,. \tag{8}$$

While $p'$ denotes the Eulerian perturbation, $p_0$ corresponds to the equilibrium pressure. $\rho(\boldsymbol{r},t)$, $\phi(\boldsymbol{r},t)$ and $\boldsymbol{v}(\boldsymbol{r},t)$ are defined similarly with the only difference that $\boldsymbol{v}_0$ is set to zero, and hence $\boldsymbol{v}(\boldsymbol{r},t) = \boldsymbol{v'}(\boldsymbol{r},t)$.
For the description of the energy equation Lagrangian coordinates are needed. Using the expression

$$\delta p(\boldsymbol{r}) = p(\boldsymbol{r},\boldsymbol{\xi}) - p_0(\boldsymbol{r}) = p'(\boldsymbol{r}) + \boldsymbol{\xi}\cdot\boldsymbol{\nabla}p_0\,, \tag{9}$$

one is able to convert a Lagrangian perturbation $\delta p(\boldsymbol{r})$ into an Eulerian perturbation. $\boldsymbol{\xi}$ quantifies the displacement caused by the perturbation and $\boldsymbol{v} = \partial\boldsymbol{\xi}/\partial t$. In the Lagrangian reference frame the independent variables are $\boldsymbol{r}_0 = (r_0,\, \theta_0,\, \phi_0)$, which is the initial position of a given material point, and time $t$. The spatial component, which labels a certain material point, does not vary in time, hence the Lagrangian reference frame corresponds to the perspective of an observer who follows the motion.
The Eulerian perturbations are inserted in Eq. (2) - (4), while the Lagrangian perturbation is plugged into the energy conservation (5). The perturbed equations are then linearized and the equilibrium equations (1a) - (1d) are subtracted to get a system of partial differential equations for the perturbations. As these equations are linear and the coefficients, which are the unperturbed equilibrium quantities, are time-independent we can use a temporal Fourier-approach to solve them. For example, the pressure perturbation is thus written as

$$p'(\boldsymbol{r},t) = \hat{p}'(\boldsymbol{r})\,h(t)\,, \tag{10}$$

where $\hat{p}'$ is a function of $\boldsymbol{r}$ alone, $h(t) = \exp(\mathrm{i}\omega t)$ and

$$\frac{\partial}{\partial t}\,h(t) = \mathrm{i}\omega\,h(t)\,. \tag{11}$$

$h(t)$ is therefore the eigenfunction of the operator $\frac{\partial}{\partial t}$ and $\omega$ is the corresponding eigenvalue. With this approach the linearized equations become

$$\mathrm{i}\omega\,[\hat{\rho}' + \mathrm{div}(\rho_0\,\hat{\boldsymbol{\xi}})] = 0\,, \tag{12a}$$

$$-\rho_0\,\omega^2\,\hat{\boldsymbol{\xi}} = -\boldsymbol{\nabla}\hat{p}' - \hat{\rho}'\,\boldsymbol{\nabla}\phi_0 - \rho_0\boldsymbol{\nabla}\hat{\phi}'\,, \tag{12b}$$

$$\nabla^2\hat{\phi}' = 4\pi\mathrm{G}\hat{\rho}'\,, \tag{12c}$$

$$\hat{p}' + \hat{\boldsymbol{\xi}}\cdot\boldsymbol{\nabla}p_0 = \Gamma_{1,0}\frac{p_0}{\rho_0}\,(\hat{\rho}' + \hat{\boldsymbol{\xi}}\cdot\boldsymbol{\nabla}\rho_0)\,. \tag{12d}$$

The Lagrangian perturbations in Eq. (12d) are given in Eulerian coordinates using Eq. (9). The term $\boldsymbol{\nabla}\hat{p}'$ in the momentum equation (12b) is the restoring force due to pressure variations and the other two terms, namely $\hat{\rho}'\,\boldsymbol{\nabla}\phi_0$ and $\rho_0\boldsymbol{\nabla}\hat{\phi}$, are the restoring forces due to gravity variations, where the first term is buoyancy and the second term arises from an altered gravitational acceleration. These three restoring forces correspond to three different oscillation modes, which are discussed in Sec. 2.2.2.
Thus a fourth-order system of time-independent differential equations is obtained which now only depends on $\boldsymbol{r}$. As the coefficients in Eq. (12a) - (12d) are the equilibrium quantities, derived under the assumption of spherical symmetry, they depend on $r$ alone. Thus one can separate the radial component from the horizontal $(\theta,\, \phi)$ component (i.e. spatial Fourier analysis)

$$\hat{p}'(r,\,\theta,\,\phi) = \tilde{p}'(r)\,f(\theta,\,\phi)\,, \tag{13}$$

where the function $f(\theta,\,\phi)$ has yet to be determined. In this new form, the fluctuation $\tilde{p}'$ depends only on $r$. The horizontal component of the displacement vector $\hat{\boldsymbol{\xi}} = (\hat{\xi}_r,\,\hat{\xi}_\theta,\,\hat{\xi}_\phi)$ is then $\hat{\boldsymbol{\xi}}_h = \hat{\boldsymbol{\xi}} - \boldsymbol{e}_r\hat{\xi}_r$, where $\boldsymbol{e}_r$ is the unit vector in radial direction. The radial and horizontal components of the gradient in spherical coordinates are defined as

$$\boldsymbol{\nabla} := \left(\frac{\partial}{\partial r},\,\frac{1}{r}\frac{\partial}{\partial \theta},\,\frac{1}{r\sin(\theta)}\frac{\partial}{\partial \phi}\right) := \left(\frac{\partial}{\partial r},\,\boldsymbol{\nabla}_h\right). \tag{14}$$

Taking the horizontal component of the linearized momentum equation (12b) and calculating the horizontal divergence of this equation (which is easy since the equilibrium quantities are independent of $\theta$ and $\phi$), $\nabla\cdot\hat{\boldsymbol{\xi}}_h$ can be eliminated in the continuity equation (12a). Using the adiabatic approximation (12d) $\hat{\rho}'$ can also be eliminated

$$\hat{\rho}' = \rho_0\left(\frac{\hat{p}'}{\Gamma_{1,0}} - A_s\hat{\xi}_r\right), \tag{15}$$

where

$$A_s = \frac{\mathrm{d}\ln\rho_0}{\mathrm{d}r} - \frac{1}{\Gamma_{1,0}}\frac{\mathrm{d}\ln p_0}{\mathrm{d}r} \tag{16}$$

is the *Schwarzschild discriminant* with which one can determine if a star is convective stable ($A_s < 0$) or unstable ($A_s > 0$). Thus the following system of differential equations for linear, adiabatic, non-radial stellar oscillations for the three unknowns, $\hat{p}'$, $\hat{\phi}'$ and $\hat{\xi}_r$, is obtained

$$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\hat{\xi}_r) + \frac{1}{\Gamma_{1,0}}\frac{\mathrm{d}\ln p_0}{\mathrm{d}r}\hat{\xi}_r + \left(\frac{\rho_0}{\Gamma_{1,0}\,p_0} + \frac{1}{\omega^2}\nabla_h^2\right)\frac{\hat{p}'}{\rho_0} + \frac{1}{\omega^2}\nabla_h^2\,\hat{\phi}' = 0\,, \tag{17a}$$

$$\frac{1}{\rho_0}\left(\frac{\partial}{\partial r} + \frac{g_0\rho_0}{\Gamma_{1,0}\,p_0}\right)\hat{p}' - (\omega^2 + g_0A_s)\hat{\xi}_r + \frac{\partial\hat{\phi}'}{\partial r} = 0\,, \tag{17b}$$

$$\left(\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \nabla_h^2\right)\hat{\phi}' - 4\pi\mathrm{G}\rho_0\left(\frac{\hat{p}'}{p_0\,\Gamma_{1,0}} - A_s\hat{\xi}_r\right) = 0\,, \tag{17c}$$

where

$$\nabla_h^2 = \frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2(\theta)}\frac{\partial^2}{\partial\phi^2} \tag{18}$$

is the horizontal component of the Laplace operator. Eq. (17a) corresponds to the linearized continuity equation, Eq. (17b) to the linearized radial equation of motion with $g_0 = -\mathrm{d}\phi_0/\mathrm{d}r$, where the total differential is justified because $\phi_0$ is independent of $\theta$ and $\phi$, and Eq. (17c) represents the linearized Poisson equation in spherical coordinates.
From these equations follows that the derivatives $\partial/\partial\theta$ and $\partial/\partial\phi$ only appear in the combination $\nabla_h^2$. This suggests that $f(\theta,\,\phi)$ (Eq. (13)) is an eigenfunction of $\nabla_h^2$

$$\nabla_h^2\,f(\theta,\,\phi) = \frac{1}{r^2}\Lambda\,f(\theta,\,\phi)\,. \tag{19}$$

### 2.2.1 Spherical Harmonics

As the spherical harmonics are the eigenfunctions of the angular Laplace operator $\nabla_h^2$, one can define

$$f(\theta,\,\phi) \equiv Y_l^m(\theta,\,\phi) = (-1)^m\,c_{lm}\,P_l^m(\cos(\theta))\exp(im\phi)\,. \tag{20}$$

$P_l^m(\cos(\theta))$ is a Legendre Polynomial of degree $l > 0$ and order $m = -l,\,...,\,l$ and $c_{ml}$ is a normalization constant. The eigenvalue $\Lambda$ in Eq. (19) is defined in the following way:

$$\Lambda = l\,(l+1)\,. \tag{21}$$

Finally one can write the pressure perturbation as:

$$p'(r,\theta,\phi,t) = \tilde{p}'(r)\,Y_l^m(\theta,\phi)\,\exp(\mathrm{i}\omega t)\,. \tag{22}$$

From this equation it follows that the horizontal component of the perturbations varies on concentric spheres like spherical harmonics and that the perturbations change temporarily with the frequency $\omega$. We can therefore rewrite Eq. (17a) - (17c) using $\nabla_h^2\,\hat{p}'(r,\theta,\phi) = [l(l+1)]/r^2\,\tilde{p}'(r)$. In the following plot various oscillation modes with degrees up to $l=3$ and the corresponding azimuthal orders $m=-3, ..., 3$ are visualized (Fig. 2).

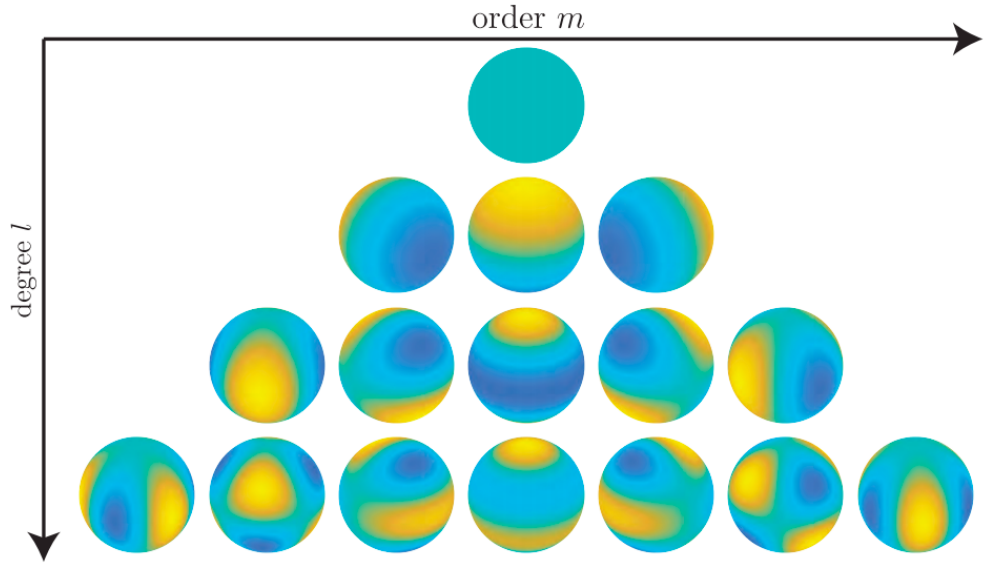



Figure 2: Spherical harmonics for different degrees $l$ and orders $m$. Figure taken from Pfaff et al. (2017).

The first circle represents a radial mode with $l=0$, which defines an oscillatory motion that has no $\boldsymbol{\xi}_h$ component and thus the only displacement happens in radial direction. The equations for adiabatic, radial oscillations can simply be derived by setting $l=0$ in the rewritten Eq. (17a) - (17c). The three spheres in the second row correspond to dipole oscillations with $l=1$. Modes with azimuthal order $m=0$ are in the center of each row. Yellow areas indicate inward motion in radial direction while blue denotes outward motions. It is clearly visible that the number of node lines ($Y=0$) increases with increasing degree while the colored areas become smaller.
The quantum numbers $l$ and $m$ cover two dimensions, namely the variation of the perturbations on concentric spheres. As there are three dimensions, we need a third quantum number $n$ which defines the number of nodes in the radial direction, often referred to as overtone of the mode.

### 2.2.2 Oscillation Modes

In order to get a qualitative understanding of the oscillatory behavior, some approximations are introduced which reduce the fourth-order system of differential equations to a second-order system.
Firstly, the equations Eq. (17a) - (17c) are rewritten, after replacing $\nabla_h^2\,\hat{p}'(r,\theta,\phi) = [l(l+1)]/r^2\,\tilde{p}'(r)$, using two characteristic frequencies, namely the *Lamb frequency*

$$S_l^2 = \frac{l\,(l+1)\,c_s^2}{r^2} \simeq k_h^2 c_s^2\,, \tag{23}$$

where $c_s$ denotes the sound speed and $k_h^2$ is the horizontal component of the wave vector

$|\boldsymbol{k}| = k_r + \boldsymbol{k}_h$. The second expression results from the local approximation of a plane wave $\exp(\mathrm{i}\boldsymbol{k}\boldsymbol{r})$. As well as the *Brunt-Väisälä frequency*

$$N^2 = -g\,A_s = g\left(\frac{\mathrm{d}\ln\rho}{\mathrm{d}r} - \frac{1}{\Gamma_1}\frac{\mathrm{d}\ln p}{\mathrm{d}r}\right). \tag{24}$$

From now on the tilde for fluctuating quantities as well as the subscript '0' for equilibrium quantities are dropped. The equations above are dispersion relations, which connect the frequency with the phase velocity. The phase velocity of the *Lamb frequency* is therefore the sound speed while the phase velocity of the *Brunt-Väisälä frequency* is proportional to the gravitational acceleration.
Secondly, the so-called *Cowling approximation* (Cowling, 1941), which neglects the perturbation of the gravitational potential, i.e. $\phi' = 0$, is applied. Thus all terms multiplied with $\phi'$ vanish. The system of differential equations further simplifies if one introduces the pressure scale height

$$H_p^{-1} = -\frac{\mathrm{d}\ln p}{\mathrm{d}r} \tag{25}$$

and assumes that for oscillations with high radial orders ($n \gg$, i.e. many nodes in radial direction) the quantity $\left|H_p^{-1}\,\xi_r\right| \ll |\mathrm{d}\xi_r/\mathrm{d}r|$, which becomes obvious considering the sketch below

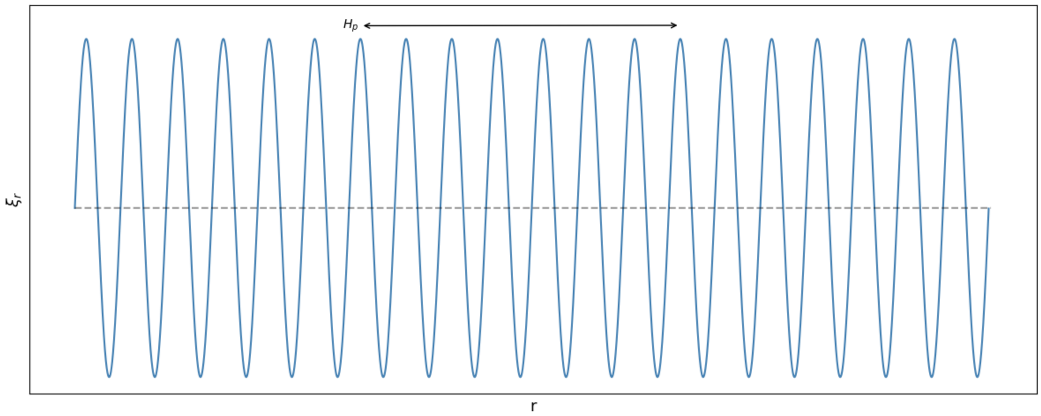



Figure 3: Scale height in comparison with the variation of the radial displacement for a mode with a high radial order. The units of $\xi_r$ and $r$ are arbitrary.

Regarding all of this, the oscillation equations (17a) - (17c) become

$$\frac{\mathrm{d}\xi_r}{\mathrm{d}r} = \frac{1}{\rho c_s^2}\left(\frac{S_l^2}{\omega^2} - 1\right)p', \tag{26a}$$

$$\frac{\mathrm{d}p'}{\mathrm{d}r} = \rho\,(\omega^2 - N^2)\,\xi_r\,. \tag{26b}$$

By taking the derivative of Eq. (26a) one can combine the two equations

$$\frac{\mathrm{d}^2\xi_r}{\mathrm{d}r^2} = -\frac{\omega^2}{c_s^2}\left(1 - \frac{N^2}{\omega^2}\right)\left(\frac{S_l^2}{\omega^2} - 1\right)\xi_r = -K_s(r)\,\xi_r\,. \tag{27}$$

If $\xi_r$ oscillates as a function of $r$ hence depends on the sign of $K_s(r)$: A positive sign of $K_s(r)$ results in an oscillating character of the mode while a negative sign produces an exponentially decreasing solution.

There are two possibilities for the former case to happen:

1. $|\omega| > |N|$ and $|\omega| > |S_l|$
2. $|\omega| < |N|$ and $|\omega| < |S_l|$

and two possibilities that $K_s(r)$ is negative and that the solution decays exponentially:

1. $|N| < |\omega| < |S_l|$
2. $|S_l| < |\omega| < |N|$

The condition where $K_s(r) = 0$ is called *turning point*.

#### 2.2.2.1 p-Modes

For high-frequency modes, which satisfy the first condition for an oscillation solution $|\omega| > |N|, |S_l|$, the restoring force is dominated by pressure (the first term on the right hand side of Eq. (12b)), which is why these modes are called pressure- or *p-modes*.
Usually, the observed p-modes have a frequency that is much higher than the *Brunt-Väisälä* frequency, $|\omega| \gg |N|$. In this case $K_s(r)$ becomes

$$K_s(r) \sim \frac{1}{c_s^2}(\omega^2 - S_l^2)\,. \tag{28}$$

If we compare this with the dispersion relation of a plane sound wave

$$\omega^2 = c_s^2|\boldsymbol{k}|^2 = c_s^2\left[k_r^2 + \frac{l(l+1)}{r^2}\right] = c_s^2\,k_r^2 + S_l^2 \tag{29}$$

we get

$$k_r^2 = \frac{1}{c_s^2}(\omega^2 - S_l^2)\,, \tag{30}$$

and it follows that $K_s(r) = k_r$ (Christensen-Dalsgaard, 2002). Therefore, the dynamic behavior of p-modes is solely determined by the sound speed which is why these modes correspond to acoustic waves. Considering once more Eq. (28) it can be deduced that the modes 'turn' ($k_r = 0$, i.e. the wave vector has only horizontal components) when $\omega = S_l(r_t)$, which is equivalent to

$$\frac{c_s^2(r_t)}{r_t^2} = \frac{\omega^2}{l\,(l+1)}\,. \tag{31}$$

The *Lamb frequency* therefore defines the frequency (to a given degree $l$) at which the p-mode is refracted back to the surface.
Physically, the refraction happens due to the increase in sound speed

$$c_s = \sqrt{\frac{\gamma\,k_\mathrm{B}\,T}{\mu m_u}} \tag{32}$$

with decreasing radius and therefore increasing temperature, where $k_\mathrm{B}$ is the Stefan-Boltzmann constant, $\mu$ is the mean molecular weight, $m_u$ is the atomic mass unit and $\gamma$ is the adiabatic index. For this definition of $c_s$ an ideal gas is assumed. The p-modes are therefore trapped inside an acoustic cavity, where the extent of this cavity depends on the angular degree $l$: While modes with high angular degrees are refracted close to the photosphere, modes with low angular degrees penetrate the star much deeper, as is also shown in Fig. 4.

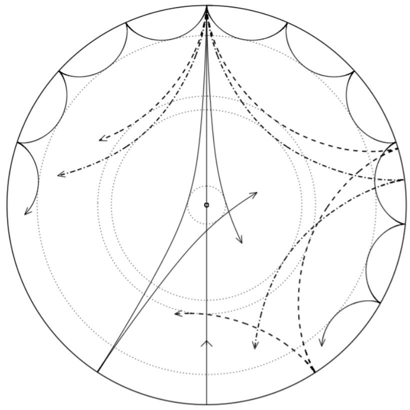


Figure 4: Propagation diagram of p-modes within a solar-like star. Figure taken from Cunha et al. (2007).

The figure above represents the propagation of rays within a solar-like star. The straight line through the center of the star marks the $l = 0$ or radial mode. The dotted circles define the turning points $r_t$ at which the modes are refracted, where the largest circle corresponds to a p-mode with an angular degree of 75 while the smallest circle determines the turning point for modes with $l = 2$.
The reflection near the surface – which cannot be directly derived from the equations above but should not be left unmentioned – happens because of the steep decrease in density. More specifically, the reflection occurs when $\omega = \omega_c$, where $\omega_c$ is the acoustic cut-off frequency

$$\omega_c = \frac{c_s}{2H}\,, \tag{33}$$

assuming an isothermal atmosphere. This means that waves are reflected when the frequency of the mode becomes comparable with the order of the density scale height $H$ which decreases with increasing radius.
There are two more things to say about p-modes: Firstly, p-modes with the same $\omega/[l\,(l+1)]$ ratio are refracted at the same radius $r_t$. Secondly, p-modes with higher radial orders have larger frequencies. Modes with the same angular degree $l$ can therefore have different frequencies depending on the radial order of the mode.

#### 2.2.2.2 g-Modes

The modes which satisfy the second condition for an oscillating solution, $|\omega| < |N|$ and $|\omega| < |S_l|$, are called gravity modes or *g-modes*. This term already suggests that the main restoring force is gravity, or more specifically buoyancy (the second term on the right hand side of Eq. (12b)). Naturally, these modes are gravity waves.
For high-order g-modes we can assume that $\omega^2 \ll S_l^2$ and thus Eq. (27) becomes

$$K_s(r) = \frac{1}{\omega^2}(N^2 - \omega^2)\frac{l\,(l+1)}{r^2}\,. \tag{34}$$

$K_s(r)$ is zero if $N = \omega$, which defines the turning point of g-modes, and the modes decay exponentially if $\omega > N$.
The extend of the resonant cavity does not depend on the angular degree, it rather depends on the inner border of the convection zone. As can be deduced from Eq. (24) the *Brunt-Väisälä frequency* is only positive if the *Schwarzschild discriminant* (Eq. (16)) is negative. Hence g-modes only arise in convectively stable regions.

The propagation of a g-mode within a Sun-like star is shown in Fig. 5.

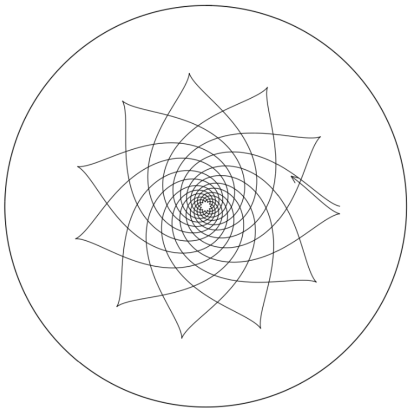


Figure 5: Propagation diagram of a g-mode within a Sun-like star. Figure taken from Cunha et al. (2007).

Since solar-like oscillators have a convective envelope where $N^2 < 0$, g-modes are trapped in the interior of the star.
Assuming a fully ionized gas, one can also rewrite the *Brunt-Väisälä frequency* as

$$N^2 = \frac{g^2\rho}{p}(\nabla_{\rm ad} - \nabla + \nabla_\mu)\,, \tag{35}$$

where

$$\nabla_{\rm ad} = \left(\frac{\partial \ln T}{\partial \ln p}\right)_{\rm ad}, \quad \nabla = \frac{{\rm d}\ln T}{{\rm d}\ln p} \quad \text{and} \quad \nabla_\mu = \frac{{\rm d}\ln \mu}{{\rm d}\ln p}\,.$$

According to that, $N^2$ also varies with the chemical composition, more specifically $N^2$ increases as the mean molecular weight increases. This means that the cavity for g-modes expands with age, and thus so-called *mixed modes*, which not only show characteristics of a p-mode but also those of a g-mode, are observed in red giants. In the Sun, however, only p-modes are clearly detected so far. If g-modes exist on the solar surface is still a matter of debate.
Fig. 6 summarizes the discussion of the two oscillation modes observed in stars.

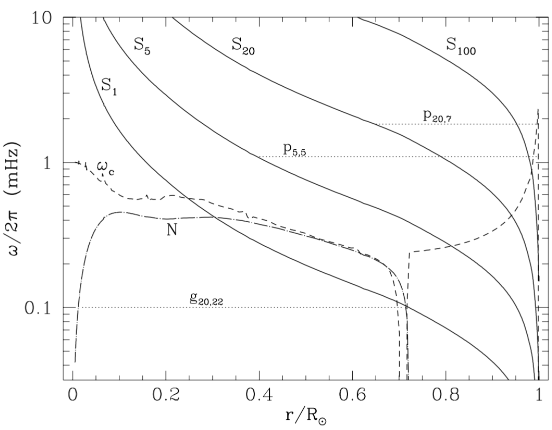



Figure 6: Propagation diagram of different oscillation modes for a solar model. The horizontal lines indicate trapped regions for two p-modes with a radial order of $n = 5$ and $n = 7$, respectively, and an angular degree of $l = 5$ and $l = 10$ as well as one g-mode with $n = 20$ and $l = 22$. Figure taken from Di Mauro (2012).

Modes with different angular degrees, which are indicated as a subscript of the *Lamb frequency* $S_l$, are plotted for a solar model. $S_l$ decreases with increasing $r$ and decreasing sound speed. Close to the center of stars the *Lamb frequency* becomes very large and tends to infinity. The peak of the *Brunt-Väisälä frequency*, visualized as dashed line, results from the increase in helium abundance near the stellar center, where nuclear fusion happens. $\omega_c$ again denotes the acoustic cut-off frequency.

#### 2.2.2.3 f-Modes

For reasons of completeness I briefly introduce another oscillation mode, namely the f-mode, for which the restoring force (the third term on the right hand side of Eq. (12b)) is again gravity.
The f-mode can be considered as surface gravity mode or highest possible g-mode. This mode only depends on the surface gravity and not on $\rho$ (as g-modes do). Hence the dispersion relation can be written as

$$\omega^2 = g\,k_h = g\frac{l\,(l+1)}{r^2}\,. \tag{36}$$

For very large angular degrees ($l \gtrsim 20$) one can rewrite this expression

$$\omega^2 = g\frac{l}{R} = \frac{\mathrm{G}M}{R^3}l \tag{37}$$

where $R$ is the surface radius of the star.

### 2.2.3 Driving Mechanisms

Last but not least we need to discuss the mechanism that drives solar-like oscillations to complete the basics of Asteroseismology.
A driving mechanism is a mechanism that puts energy into the oscillations. Such a mechanism is necessary considering the fact that stellar oscillations are not truly adiabatic but, as shown by Balmforth (1992), intrinsically damped. The oscillations of solar-like oscillators are stochastically excited by turbulent near-surface convective motions (Houdek et al., 1999) and are therefore only observable in stars which are cool enough to harbour an outer convective envelope. Considering Fig. 1 this is the case for stars on the right hand side of the classical instability strip, which is marked by two parallel dashed lines. The driving mechanism for solar-like oscillators is quite different from the $\kappa$ mechanism, which excites oscillations coherently in stars located on the left hand side of the instability strip.

# 3 Stellar Evolution of Evolved Solar-like Stars

For a better understanding of solar-like oscillation spectra, the evolution of solar-type stars needs to be discussed beforehand. As theoretical input I mainly used the lecture notes by O. Pols (2011) and again the book by Kippenhahn et al. (2012).
In this master's thesis only low and intermediate-mass stars up to about $3.5\,M_\odot$ are considered, and therefore the evolution of high-mass stars will not be discussed in this section. Furthermore, since the data set only comprises evolved stars, the description of stellar evolution begins with the exhaustion of hydrogen in the core, hence after the main sequence phase.
Fig. 7 shows a seismic HR diagram for the stars in the data set, where $\nu_{\max}$ is indirectly proportional to the luminosity (Kallinger et al., 2018)

$$\nu_{\max} \propto M\,L^{-1}\,T_{\rm eff}^{7/2}\,. \tag{38}$$

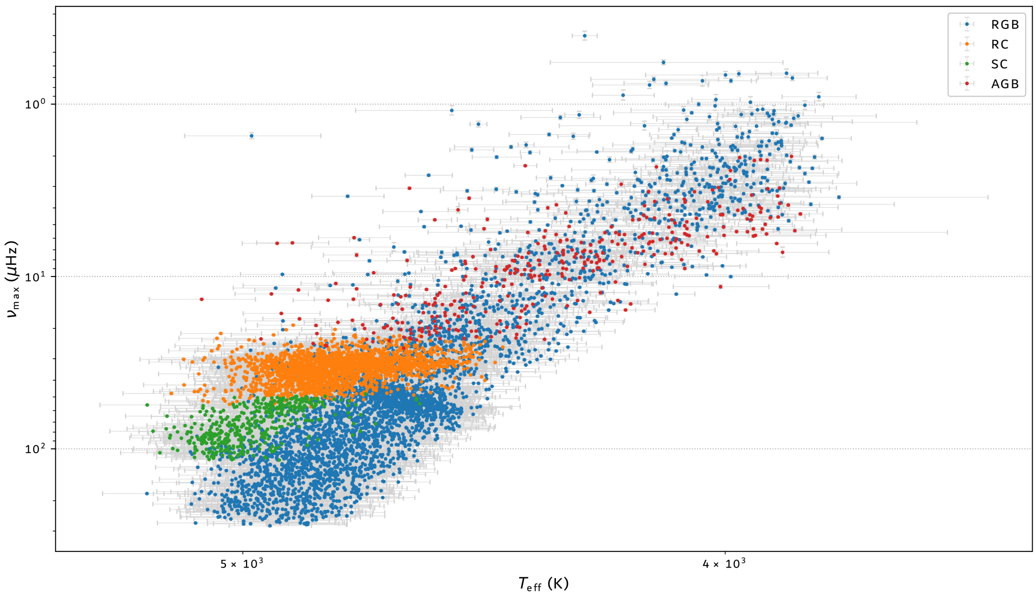



Figure 7: Seismic HR diagram of 6657 evolved *Kepler Input Catalog* (KIC) stars. The evolutionary stages are indicated by different colors. As stars ascend the *Red Giant Branch* (RGB) their surface temperature only slightly cools down while the radius and hence luminosity strongly increases. After He-ignition at the *tip of the RGB* lower- and intermediate-mass stars settle in the *red clump* (RC) and *secondary clump* (SC), respectively. As the helium supply is exhausted the star evolves further on the *Asymptotic Giant Branch* (AGB).

The colors correspond to stars of different evolutionary stages which are described in the following. For further information regarding the classification of the evolutionary stages of these stars, see Kallinger et al. (2012).
Initially, the cores of intermediate-mass stars ($M \gtrsim 2\,M_\odot$) are in thermal and hydrostatic equilibrium as H-burning in the shell begins, which surrounds the isothermal core. As mass is gradually added to the core, the *Schönberg-Chandrasekhar limit* (Schönberg and Chandrasekhar, 1942)

$$\frac{M_c}{M} < 0.10 \tag{39}$$

is exceeded at some point. This limit defines an upper limit of the core mass $M_c$ for which thermal equilibrium can be maintained. The pressure within the core can then no longer withstand the mass of the envelope and the core begins to contract rapidly also establishing

a temperature gradient. To prevent the burning shell from becoming thermally unstable, the envelope above the H-burning shell expands in response to the contracting core. The expanding envelope cools, leading to an inwards increasing convection zone and to an increase in luminosity. At this evolutionary stage the star is said to ascend the *Red Giant Branch* (RGB).
Helium is ignited as the core reaches a temperature of about $10^8$ K, which prevents it from further contracting and consequently stops the envelope expansion. The stars then leave the RGB and temporarily end up in the *secondary clump* (SC).
In contrast to intermediate-mass stars the cores of low-mass stars ($M \lesssim 2M_\odot$) become degenerate at the base of the RGB before the *Schönberg-Chandrasekhar limit* is reached, which makes this limit redundant. As a result, low-mass stars stay in hydrostatic and thermal equilibrium throughout the H-shell burning, and hence throughout the RGB phase. While the core slowly contracts on the RGB the envelope expands, cools and becomes more and more convective. As the core is already degenerate before the temperature for helium ignition is reached, the ignition happens under unstable conditions. This results in a thermonuclear runaway and is referred to as *Helium flash*. At this point the core stops contracting and starts expanding until helium can be burnt in a thermally stable manner, lifting the degeneracy. Simultaneously, due to the burning H-shell, the radius of the envelope and the luminosity decreases. After the *Helium flash*, which defines the tip of the RGB, stars with approximately the same metallicity as the Sun settle in the so-called *red clump* (RC) while stars with lower metallicities settle in the *horizontal branch* (HB). RC and HB define compact regions in the HR diagram, which result from the fact that the core mass of all low-mass stars is roughly the same when helium burning starts. This can also be seen in Fig. 7 as well as in Fig. 8.
As the stellar evolution of low- as well as intermediat-mass stars after He-exhaustion is qualitatively similar, the two cases are now discussed mutually. The He-burning in the core is followed by He-shell burning. Thus two actively burning shells now surround the CO core as the star ascends the so-called *Asymptotic Giant Branch* (AGB). As the core contracts the layer between the two burning shells expands while the radius of the outer envelope, above the H-burning shell, decreases. The expanding layer forces the H-burning layer to burn at a very low level. The luminosity early in the AGB evolution is therefore determined by the He-burning shell. In a later stage of the AGB-phase, He-burning in the He-shell becomes unstable which causes a *helium shell flash*, also called periodic *thermal pulses* (TP). The duration of this *TP-AGB* phase is on one hand limited by the decreasing envelope mass due to the two burning shells and dust-driven winds, and on the other hand by the increase in core mass. Only the most massive stars can ignite carbon in the center, which is not the case for the stars in the KIC data set. Once the dust-driven winds end, the star further evolves on the *post-Asymptotic Giant Branch*.
As all stars from the main-sequence to high up the AGB have an outer convective envelope, the oscillations of these stars are stochastically excited and therefore solar-like.
The discovery that the properties of the pulsations in semiregular variables (SR, see Fig. 1) and in solar-like oscillators are pretty similar has led to the assumption that the oscillations of these stars are also driven by convective motions near the surface (Yu et al., 2020). In fact, even the pulsations of Mira stars, which are basically coherently driven, are influenced by convection (Cunha et al., 2020). This is why all stars considered in this master's thesis are assumed to be solar-like oscillators.

Fig. 8 shows a mass-to-radius (M-R) plot for the 6657 *Kepler Input Catalog* (KIC) stars used in this master's thesis. The plot descriptively summarizes the evolutionary stages defined above.

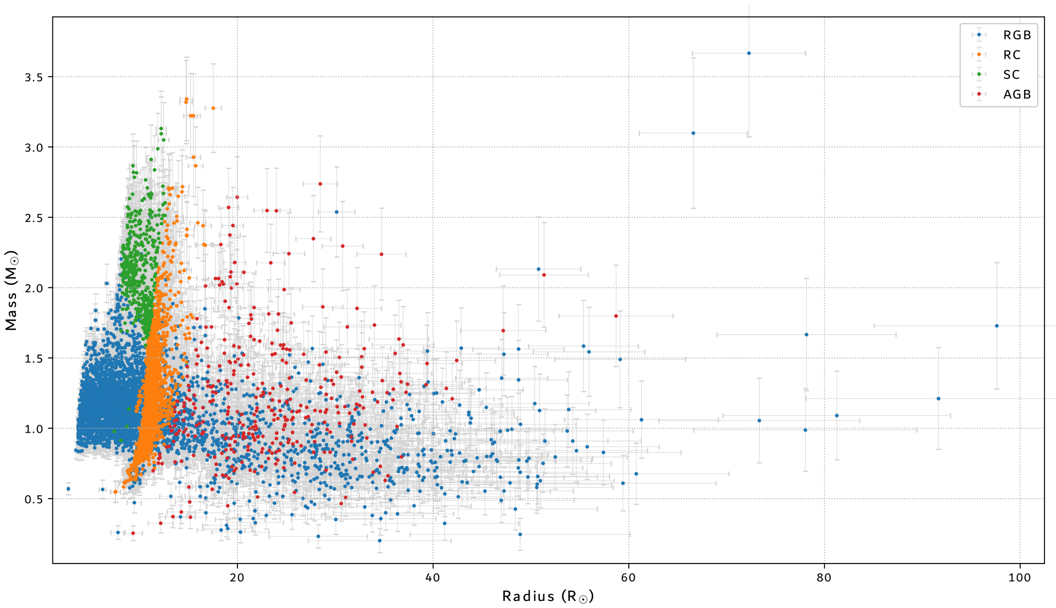



Figure 8: M-R plot for the stars in the KIC data set. Each color corresponds again to an evolutionary stage.

As can be seen, nearly all RGB stars have masses between $0.9 < M_\odot < 2$ which can be ascribed to observational reasons: Stars with masses $M < 0.9\,M_\odot$ are rarely observed because their main sequence evolution lasts very long while stars with masses $\mathrm{M} > 2\,\mathrm{M}_\odot$ evolve very quickly on the RGB, which also reduces the number of detection (Mosser, 2011). Red clump stars are clearly in contrast to the other evolutionary stages: Stars with different masses have approximately the same radius. The RC is also pretty striking in Fig. 7, where stars covering a large temperature range have roughly the same luminosity. This can be attributed to the fact that the core mass is pretty similar for all stars when the *helium flash* takes place. Contrary to this, SC stars are more massive and helium is not ignited under degenerate conditions, which results in different core masses at He-ignition and hence in a spread within the M-R plot. AGB stars lose mass due to dust-driven winds which is why they occupy the intermediate-mass region in the plot.

# 4 Solar-like Oscillation Spectra

With all this information gathered we can finally discuss solar-like oscillation spectra. The most prominent solar-like oscillator is the Sun, for obvious reasons. As already mentioned in Sec. 2.2.3, however, all stars with an outer convective envelope, that are stars with spectral classes from F to M, or, in other words, stars on the cooler side of the instability strip in Fig. 1, show solar-like oscillations. Important quantities such as the large and small frequency separation can be deduced from the oscillation spectra of these stars.
Firstly I will give a short historical overview of how the scientific discipline Helioseismology was established. The insights gained by studying the Sun for decades are fundamental for the understanding of other solar-like oscillators. In Sec. 4.2 the oscillation spectra of solar-like oscillators are discussed, mainly focusing on red giants.

## 4.1 Helioseismology

When Leighton et al. (1962) discovered the previously mentioned vertical velocities in the Ca $\lambda$ 6103 line, or five-minute oscillations, with a 'spectroheliograph' various attempts have been made to describe this oscillatory motions.
Frazier (1968) performed spectroscopic observations of the center of the solar disk, which he analyzed by means of a two-dimensional Fourier technique. The advantage of this technique was that one not only obtains spatial properties of the oscillations, but also temporal ones. Thus Frazier could display his results in a so-called diagnostic diagram or $\mathrm{k_h}-\omega$ plane, where $k_\mathrm{h}$ denotes the horizontal wave number and $\omega$ the frequency. This approach revealed that the oscillatory power is separated from the convection power within these diagrams, even at lower altitudes. Up to this discovery, the five-minute oscillations were considered a local phenomena in the solar atmosphere but the observation of Frazier led to the assumption that oscillations arise from subphotospheric layers and are therefore uncorrelated with the motion of convection cells by the time they reach the surface.
By analyzing the dispersion relation of a plane sound wave

$$\omega^2 = c_s^2\,|\boldsymbol{k}|$$

Ulrich (1970) showed that the five-minute oscillations result from trapped standing acoustic waves. He demonstrated that there are two boundaries for which $k_r = 0$ is satisfied, meaning that waves which propagate in radial direction are refracted on the one hand near the photosphere and on the other hand well below the photosphere due to the increase in sound speed. The derivation for trapped sound waves is also sketched in Sec. 2.2.2.1 where Fig. 4 shows the propagation diagram of p-modes.
While Frazier and others such as Deubner (1975) and Rhodes et al. (1977) mainly observed oscillations with low $n$, high $l$ value modes, Claverie et al. (1979), who performed observations with an integral light spectrometer which cancels out high degree modes, studied low $l$ value oscillations and firstly proved the existence of global modes of oscillations.

Nowadays, thanks to ground-based missions such as the Global Oscillation Network Group (GONG)[1] and the Birmingham Solar Oscillation Network (BiSON)[2] as well as space missions like the Solar and Heliospheric Observatory (SOHO)[3], which offer a continuous observation of the Sun, many more modes were identified, as is also shown in the solar power spectrum below (Fig. 9).

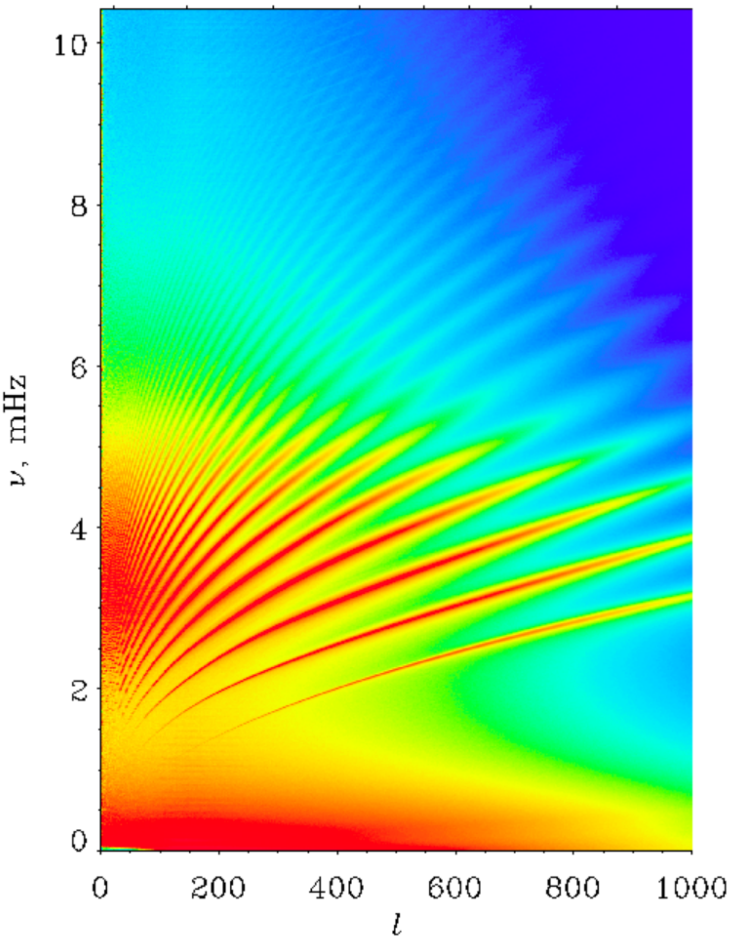



Figure 9: Power spectrum of the Sun with angular degree values up to 1000. The colors correspond to the oscillation power, where red is highest. The figure is produced from MDI observations on SOHO. Figure taken from MDI-SOHO.[4]

Modes with an angular degree up to $l = 1000$ are identified for the Sun. The ridges show different modes with the same radial order and the color indicates the oscillation power, where most of the power is concentrated around 3 mHz, corresponding to the five-minute oscillations.

[1] `https://gong.nso.edu/`
[2] `http://bison.ph.bham.ac.uk/`
[3] `https://sohowww.nascom.nasa.gov/`
[4] `http://soi.stanford.edu/` Accessed: 2021-01-12

The solar power spectrum illustrated in Fig. 10 resolves the section of highest oscillation power a bit better.

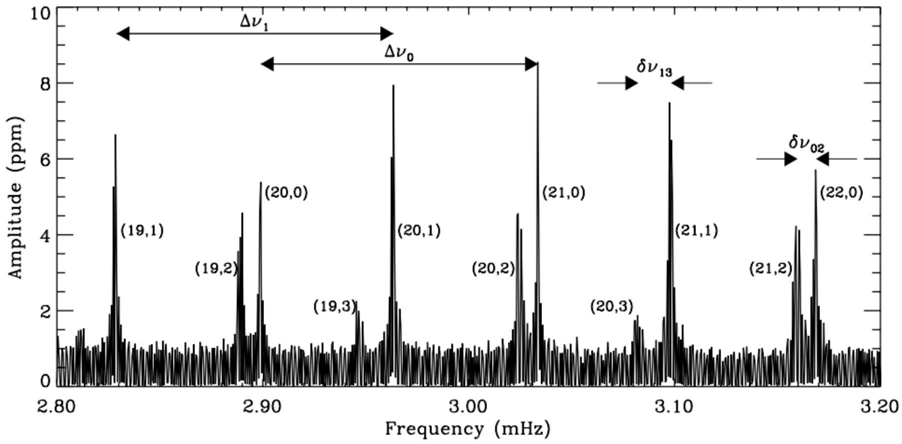



Figure 10: Sequence of the solar power spectrum, where the brackets indicate the radial order $n$ and angular degree $l$ of the mode. The data were obtained from the VIRGO instrument on SOHO. The figure is taken from Bedding and Kjeldsen (2003).

The modes with the highest amplitude, identifiable as peaks, are acoustic waves with a high-radial order but low-angular degree. The peaks appear to be almost equally spaced. Furthermore, the *large* and *small frequency separation* which are defined in the next section, are also indicated in the figure.

## 4.2 Asteroseismology of Solar-like Oscillators

In the early 1990's Brown et al. (1991) discovered solar-like oscillations on another solar-type star, namely Procyon. Eleven years later Frandsen et al. (2002) firstly detected solar-like oscillations on a red giant star, $\xi$ Hya. Since then the number of solar-like oscillators has increased vastly thanks to various observing campaigns, which are discussed in Sec. 5.
Main sequence stars with approximately the same properties as the Sun, hereafter called solar-type stars, primarily show evenly spaced peaks in their frequency spectra. These peaks are pure p-modes of high-radial order and low-angular degree (Chaplin and Miglio, 2013). Tassoul (1980) showed that these high overtone p-modes follow the asymptotic relation

$$\nu_{nl} \sim \Delta\nu \left( n + \frac{l}{2} + \alpha \right) , \tag{40}$$

where $\alpha$ is a constant of order unity and

$$\Delta\nu = \nu_{nl} - \nu_{n-1\,l} \tag{41}$$

is the *large frequency separation*, which can also be written as

$$\Delta\nu = \left( 2 \int_0^R \frac{\mathrm{d}r}{c_s(r)} \right)^{-1} . \tag{42}$$

The *large frequency separation* therefore denotes the inverse of the sound travel time of a p-mode across the stellar diameter (Aerts et al., 2010).
Following Eq. (40), modes of $(n,\, l)$ and $(n{-}1,\, l{+}2)$ have the same frequency. This degeneracy is lifted by the *small frequency separation* (Christensen-Dalsgaard, 2004)

$$\delta\nu_{n,\,l} = \nu_{n,\,l} - \nu_{n-1,\,l+2} \,. \tag{43}$$

Considering once more Fig. 10, this regular, repeated pattern indicated by the asymptotic relation is indeed visible. Modes of the same angular degree and consecutive radial order are evenly spaced according to the *large frequency separation*, while modes of $(n, l)$ and $(n-1, l+2)$ have truly different frequencies following Eq. (43). As the frequency spacing behaves like the pulsation period (which in turn is comparable with the dynamical timescale ((6))) $\Delta\nu$ is a proxy for the mean stellar density (Ulrich, 1986)

$$\Delta\nu \propto \sqrt{\bar{\rho}}\,. \tag{44}$$

The small frequency separation, in contrast, arises from the fact, that low-degree modes reach the deep interior of the star where they respond to the sound speed gradient in the core (Christensen-Dalsgaard and Thompson, 2011). The small frequency separation is therefore most sensitive to the conditions within the core and gives hints for the evolutionary state of a star.

The asymptotic relation (Eq. (40)), however, becomes gradually less accurate for more evolved stars. This can be ascribed to the increase in central condensation, and hence gravitational acceleration with stellar age, which leads to a higher *Brunt-Väisälä frequency* (Eq. (24)) (Christensen-Dalsgaard and Thompson, 2011). Consequently, coupling between a p-mode and a g-mode of the same degree can occur, provided that the region between the two cavities in which the modes propagate, is not too large. This gives rise to *mixed modes*, which were already discussed in Sec. 2.2.2.2.

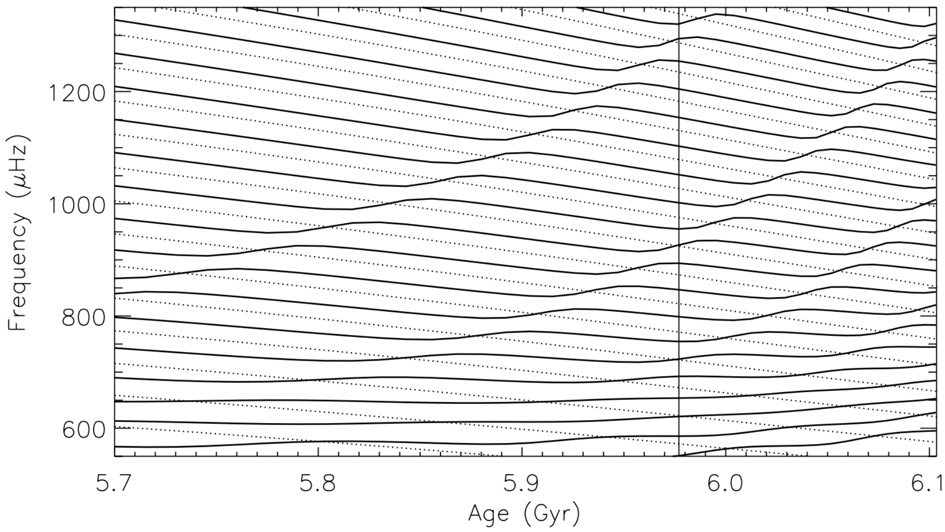



Figure 11: p-modes of angular degree $l = 0$ (dotted lines) decrease monotonically with age while $l = 1$ modes (black lines) couple with non-radial g-modes of the same degree. The age of the solar-type star Gemma is indicated as a vertical line. Figure taken from Metcalfe et al. (2010).

In Fig. 11 the theoretical frequencies of $l = 0$ (dotted line) and $l = 1$ (black lines) modes of the star KIC 11026764 (a.k.a. Gemma) are plotted against age (Metcalfe et al., 2010). Gemma is a solar-type star aged 5.77 Gyr, which is indicated by a vertical line. The $l = 0$ p-mode, which does not interact with g-modes as these are non-radial modes (Christensen-Dalsgaard and Thompson, 2011), decreases linearly with stellar age due to the increasing radius. Thus, the asymptotic relation (Eq. (40)) still applies for radial p-modes. The $l = 1$ p-mode, however, can couple with the $l = 1$ g-mode, leading to an increase in frequency as the coupling happens.

The existence of these mixed modes leads to rather complex oscillation spectra of red giants (see Fig. 12).

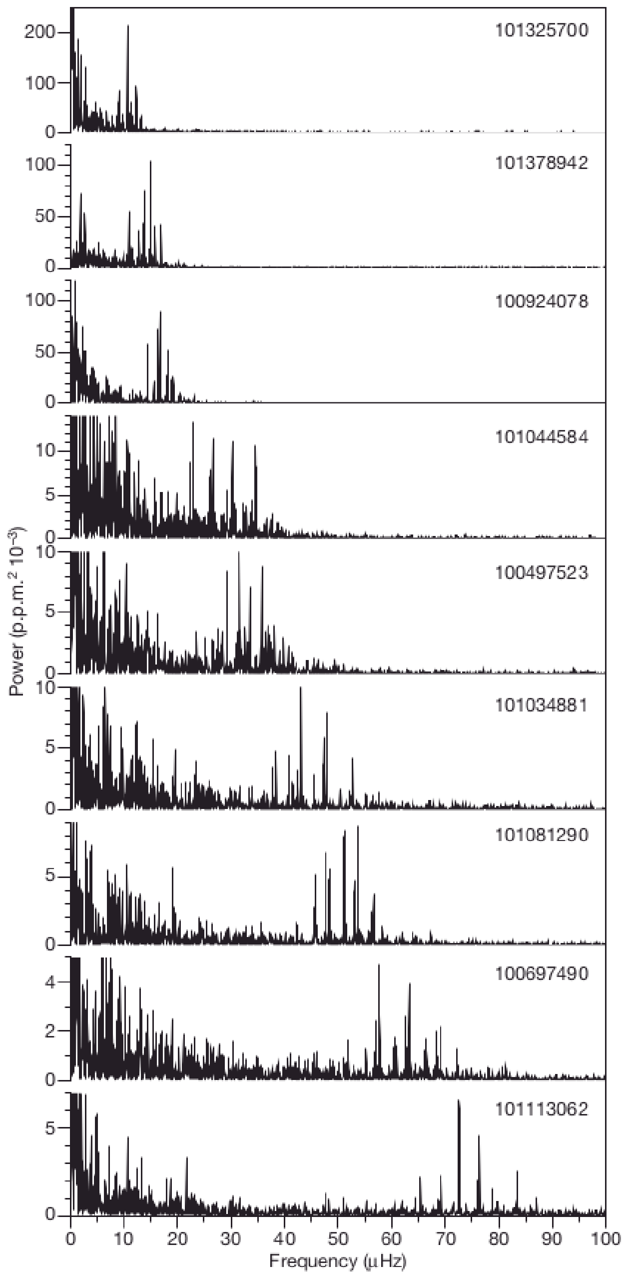



Figure 12: Oscillation spectra of nine red giant stars observed with the space telescope CoRoT. The radius of the stars decreases from top to bottom. Figure taken from De Ridder et al. (2009).

As one can see, the frequency of maximum oscillation power $\nu_{\max}$ is shifted to higher frequencies from top to bottom. This happens because $\nu_{\max}$ scales as the acoustic cut-off frequency $\nu_c = \omega_c/2\pi$ (Eq. (33)) (Belkacem et al., 2011)

$$\nu_{\max} \propto \nu_c \propto M\, R^{-2}\, T_{\text{eff}}^{-1/2} \tag{45}$$

and because the radius of the red giant stars decreases from top to bottom.

### 4.2.1 Nonlinear Seismic Scaling Relations

Using the two scaling relations for $\Delta\nu$ and $\nu_{\rm max}$ one is able to estimate radius and mass of solar-type stars. If scaled with solar values, Eq. (44) and Eq. (45) become

$$\frac{\Delta\nu}{\Delta\nu_\odot} = \left(\frac{M}{M_\odot}\right)^{1/2} \left(\frac{R}{R\odot}\right)^{-3/2} , \tag{46}$$

$$\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}} = \frac{g}{g_\odot} \left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right)^{-1/2} , \tag{47}$$

where $g$ and $g_\odot$ is the surface gravity for a star and the Sun, respectively. These scaling relations work quite well for main sequence stars, but overestimate the radii and masses of more evolved stars. Kallinger et al. (2018) therefore adjusted the scaling relations using seismic data of six eclipsing binaries with a red giant component, for which independent radius and mass measurements exist, as well as 60 RGB and RC stars in the open clusters NGC 6791 and NGC 6819. By doing so they obtain *nonlinear seismic scaling relations*:

$$\frac{g}{\sqrt{T_{\rm eff}}} = \left(\frac{\nu_{\rm max}}{\nu_{\rm max,\odot}}\right)^{1.0075\pm0.0021} , \tag{48}$$

$$\sqrt{\bar\rho} = \frac{\Delta\nu}{\Delta\nu_\odot} \left[\eta - (0.0085 \pm 0.0025) \cdot \log^2(\Delta\nu/\Delta\nu_\odot)\right]^{-1} , \tag{49}$$

which can be used to estimate the masses of RGB, RC, SC and AGB stars. These are the evolutionary stages considered in this thesis. In the above equations, the surface gravity $g$, $T_{\rm eff}$ and $\bar\rho$ are given in solar units, $\nu_{\rm max} = (3140 \pm 5)\,\mu$Hz, $\Delta\nu_\odot = (135.08 \pm 0.02)\,\mu$Hz, $\eta = 1$ for RGB stars and $\eta = 1.04 \pm 0.01$ for RC, SC, and AGB stars.

# 5 Obtaining Asteroseismic Data

In this section I will give a rough overview of observational techniques. The NASA missions *Kepler* and *TESS* are discussed in more detail as they are important for this master's thesis. Stellar oscillations generally cause fluctuations in physical parameters, which become noticeable in form of:

1. intensity variations
2. radial velocity variations.

Brightness variations can be detected by means of photometric time series, while radial velocity variations are measured via spectroscopic time series.
Since solar-like oscillators show comparatively small amplitudes (a few ppm in the Sun to some hundred ppm in red giants) they are difficult to detect by photometric ground-based observations. Instead, solar-like oscillators are observed spectroscopically by means of spectrographs like HARPS[5], UVES[6], CORALIE[7], SARG[8] and UCLES[9], which are at different observing sites, namely Chile, Spain and Australia. These spectrographs, among some others, achieve the necessary precision of about $5\,\mathrm{m\,s^{-1}}$ for the detection of solar-like oscillations (Christensen-Dalsgaard, 2002).
When it comes to photometry, the precision required for solar-like oscillations ($\sim$ few $\mu$mag) is only achieved from space (Cunha et al., 2007). The micro-satellite MOST (Walker et al., 2003), operated by the Canadian Space Agency and launched in 2003, continuously observed bright stars up to 60 days, which enabled the detection of solar-like oscillations.
Asteroseismology improved a lot due to the french-led mission CoRoT (Baglin, 2003) as well as NASA's *Kepler* mission (Borucki et al., 2010). The ongoing NASA mission *TESS* (Ricker et al., 2015) as well as the future ESA mission PLATO (Rauer et al., 2016) are also expected to observe a large number of solar-like oscillators.

## 5.1 *Kepler*

The main objective of the *Kepler* mission was to detect exoplanets by means of the so-called transit method which photometrically measures the brightness variation caused by an exoplanet while passing in front of its host star. Since oscillations are also detectable as changes in the star's intensity, both exoplanet search as well as Asteroseismology can be conducted simultaneously. The latter was carried out by the Kepler Asteroseismic Science Consortium (KASC).
*Kepler* was launched in March 2009 and operated until May 2013 when two of its reaction wheels stopped working. It was a Schmidt telescope with an aperture of $\sim$ 1 m and it had a field of view that covered $\sim$ $115\,\mathrm{deg}^2$ (Koch et al., 2004). The viewing field pointed towards a region between the constellations Cygnus and Lyrae, which enabled continuous observations of about 150,000 stars (Borucki et al., 2010). The data, obtained by 42 CCDs mounted in the focal plane of the telescope, was read out every 6.05 seconds and summed up on board for 30 minutes (long cadence) (Koch et al., 2004). This sampling rate allows us to estimate the *Nyquist frequency*

$$\nu_{\mathrm{Nyq}} = \frac{1}{2\Delta_t}\,, \tag{50}$$

which is the highest usable frequency in the oscillation spectra. Thus, the frequency of a signal has to be sampled twice to be considered as intrinsic frequency.

[5] `https://www.eso.org/sci/facilities/lasilla/instruments/harps.html`
[6] `https://www.eso.org/sci/facilities/paranal/instruments/uves.html`
[7] `https://plone.unige.ch/EULER/technic/documentation/missions/coralie-007`
[8] `http://www.tng.iac.es/instruments/sarg/`
[9] `https://aat.anu.edu.au/science/instruments/current/ucles/overview`

For a sampling rate of 30 min (long cadence) we get $\nu_{\rm Nyq} \approx 280\,\mu$Hz. As main sequence stars and most subgiants oscillate above this frequency (Huber and Zwintz, 2020) *Kepler* mainly observed stars that are ascending the RGB (Fig. 13).

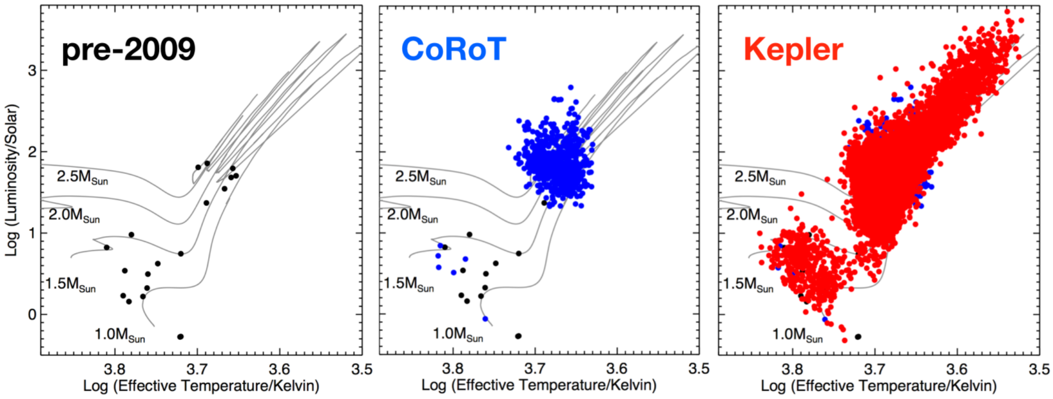



Figure 13: Increase in detections of solar-like oscillators since the space missions CoRoT and *Kepler*. Long cadence observations by *Kepler* enabled the detection of solar-like oscillations in about 20,000 RGB stars while short cadence observations yielded a few hundred solar-like oscillation detections in subgiant stars. Figure taken from Huber and Zwintz (2020).

With its photometric precision of 20 ppm for an exposure time of 6.5 hrs, *Kepler* was able to detect solar-like oscillations in more than 20,000 low- and intermediate-mass RGB stars (Borucki et al., 2010). A limited set of stars was also summed up every minute (short cadence), which enabled the detection of solar-like oscillations in about 500 main-sequence and subgiant stars (Huber and Zwintz, 2020). Fig. 13 also shows the increase in the number of detected solar-like oscillators thanks to CoRoT and *Kepler*.
The *Kepler* mission was continued as K2 mission (Howell et al., 2014) after the loss of the reaction wheels. During the K2 mission more than 20 fields along the ecliptic were observed for 80 days each. The K2 mission ended in 2018 as the fuel for the spacecraft was exhausted.

## 5.2 *TESS*

As *Kepler*, *TESS* (Transiting Exoplanet Survey Satellite) was also primarily constructed to detect expolanets. In contrast to *Kepler*, however, *TESS* observes bright, nearby stars, which is achieved through smaller aperture sizes of the four different cameras mounted on the spacecraft. These cameras cover a $24^\circ \times 94^\circ$ field of view which are observed for at least 27 days, whereas the ecliptic poles are viewed continuously (Ricker et al., 2015).
*TESS* launched in April 2018 and operates since then in a 2:1 lunar resonance orbit. Every 30 minutes a full-frame image (FFI) is obtained. The sampling rate is therefore the same as for *Kepler* and many evolved solar-like oscillations ascending the RGB will be observed. Additional 200,000 stars will be sampled every two minutes (short cadence) which shifts the *Nyquist frequency* (Eq. (50)) to 4.2 mHz, enabling the detection of solar-like oscillations in subgiants, which fills the gap of detected solar-like oscillators in the low-mass HR diagram (see also Fig. 13). These stars are listed in the Asteroseismic Target List (ATL) (Schofield et al., 2019).

Fig. 14 shows the expected distribution and number of targets in the ATL in comparison to already detected, less evolved solar-like oscillators by *Kepler*.

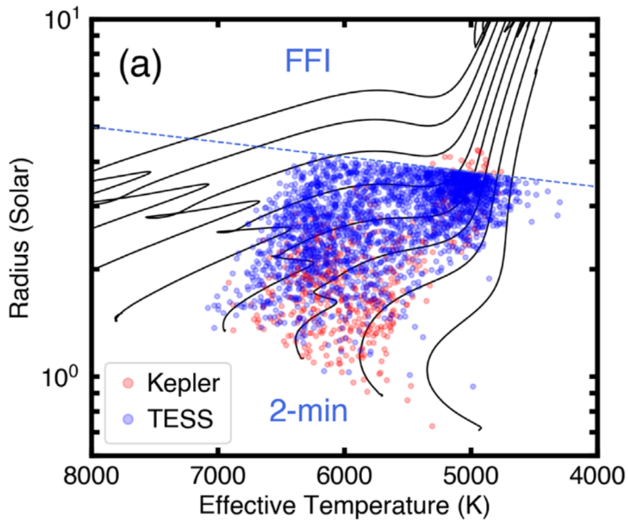



Figure 14: Expected detections of solar-like oscillators by *TESS*. Figure taken from Schofield et al. (2019).

It is worth mentioning that the comparatively short observation periods of *TESS* makes it gradually more difficult to obtain the *large frequency separation* as the frequency resolution, $\Delta_\nu$, is indirectly proportional to the length of the time series:

$$\Delta_\nu \propto \frac{1}{2\pi T} \,. \tag{51}$$

Since the *large frequency separation* is in general pretty challenging to obtain, an independent method that does not need $\Delta\nu$ is desirable. If no *large frequency separation* measurements are available, the seismic scaling relations cannot be applied which makes the estimation of radius and mass challenging. This is exactly why a machine learning (ML) algorithm, that does not need $\Delta\nu$ as input parameter but other seismic parameters introduced in the following section, is created.

# 6 Data

The data set used to create a model by means of ML consists of more than 6000 evolved stars. These stars are taken from the APOKASC catalog (Pinsonneault et al., 2019), which not only contains asteroseismic data from the *Kepler* mission (Borucki et al., 2010) but also information determined from high resolution near-infrared spectra from the APOGEE project (Majewski et al., 2017). The APOKASC catalog includes 6676 targets altogether. Proceeding from these, Kallinger (2019) extracted oscillation mode parameters for 6179 stars. As the effective temperatures are not available for 15 stars, I end up with 6164 stars from the APOKASC catalog. The seismic parameters of additional 493 evolved red giants, again observed with the *Kepler* telescope, are provided by my supervisor. The effective temperatures of these stars are taken from the LAMOST catalog, which is a spectroscopic survey of the galactic anticentre (Xiang et al., 2017). The total data sample therefore comprises 6657 red giants, which are plotted in Fig. 7. For all 6657 stars in the sample the seismic analysis pipeline of Kallinger et al. (2012) delivers the following global seismic parameters (except $T_{\text{eff}}$, which is taken from the APOKASC and LAMOST catalog) as well as their observational errors:

| | |
|---|---|
| $\nu_{\text{max}}$ | frequency of maximum oscillation power |
| $\sigma$ | width of Gaussian fit to the oscillation power excess |
| $p_{\text{tot}}$ | root of the integral below the Gaussian fit |
| $T_{\text{eff}}$ | effective temperature |
| $\Delta\nu$ | large frequency separation |

Table 1: Starting parameters for each star.

Below is the power density spectrum of KIC 4448777, which is not part of the data set, but which is adduced as an example how different parameters listed above are obtained.

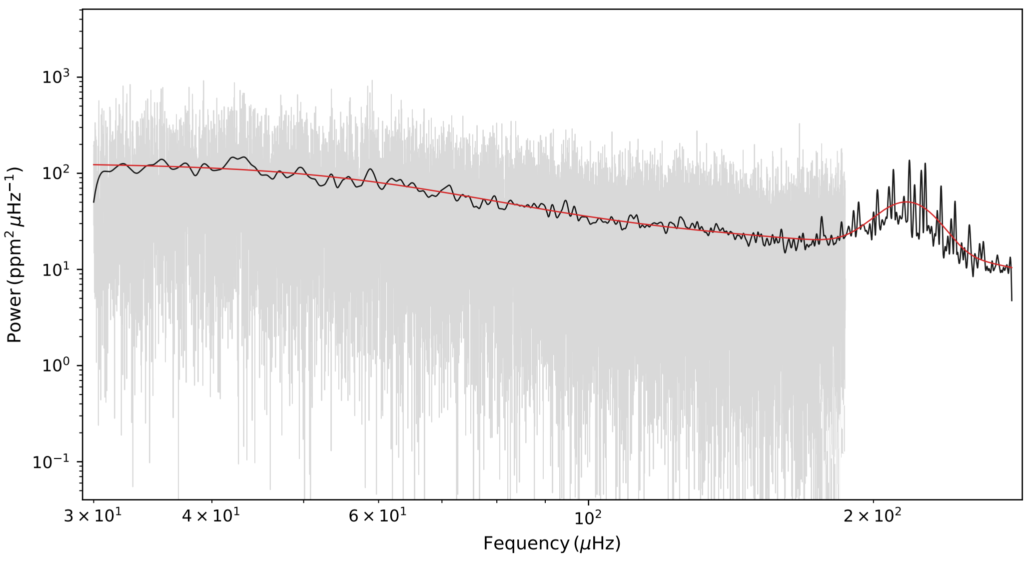



Figure 15: Power Density Spectrum of KIC 4448777.

The gray lines represent the raw data while the black indicates the smoothed data to which the function (red)

$$P(\nu) = P_n + P_g \exp\left(\frac{-(\nu - \nu_{\text{max}})^2}{2\sigma^2}\right) + \sum_i \frac{\xi a_i^2/b_i}{1 + (\nu/b_i)^4} \tag{52}$$

is fitted. The first term corresponds to the instrumental noise, the second term is a proxy for the power excess of the stellar pulsation and the last term indicates the stellar granulation

signal with $\xi = 2\sqrt{2}/\pi$. For further information regarding the modeling of power density spectra, see Kallinger et al. (2014). For an understanding of the parameters that are used in this thesis, it is sufficient to only consider the power excess, which is described by a Gaussian function: $\nu_{\rm max}$ is therefore the position of the peak, $\sigma$ is the width and $p_{\rm tot}$ is the root of the integral of the Gaussian function.
As was already pointed out in Sec. 3, all stars are assumed to be solar-like oscillators, which is why we can use the *nonlinear scaling relations* (see Sec. 4.2.1) to estimate radius and mass. The resulting masses and radii are plotted in Fig. 8 in Sec 3. With the values for radius and mass the parameter set for the ML algorithm is complete. The following table shows exemplary the first five entries of this data set

| | Source | | | | Target | |
|---|---|---|---|---|---|---|
| KIC | $\nu_{\rm max}$ ($\mu$Hz) | $\sigma$ ($\mu$Hz) | $p_{\rm tot}$ (ppm) | $T_{\rm eff}$ (K) | Radius ($R_\odot$) | Mass ($M_\odot$) |
| 1160789 | $25.32 \pm 0.19$ | $4.66 \pm 0.15$ | $366 \pm 8$ | $4730 \pm 72$ | $10.82 \pm 0.71$ | $0.82 \pm 0.11$ |
| 1161447 | $36.82 \pm 0.22$ | $5.76 \pm 0.18$ | $248 \pm 5$ | $4776 \pm 86$ | $10.92 \pm 0.61$ | $1.23 \pm 0.14$ |
| 1161618 | $33.91 \pm 0.22$ | $5.68 \pm 0.16$ | $295 \pm 5$ | $4742 \pm 72$ | $11.00 \pm 0.61$ | $1.14 \pm 0.13$ |
| 1162220 | $11.04 \pm 0.99$ | $1.95 \pm 0.79$ | $741 \pm 2$ | $4190 \pm 52$ | $17.85 \pm 1.34$ | $0.92 \pm 0.14$ |
| 1162746 | $27.85 \pm 0.21$ | $4.53 \pm 0.15$ | $312 \pm 7$ | $4798 \pm 76$ | $10.29 \pm 0.67$ | $0.83 \pm 0.11$ |

Table 2: Initial values of five exemplary stars for ML. The previously known parameters are referred to as ‘Source’ while the term ‘Target’ denotes parameters which should be predicted by the ML algorithm and hence corresponds to radius and mass.

Fig. 16 shows the whole parameter set where one quantity is plotted against the other with the mass color-coded.

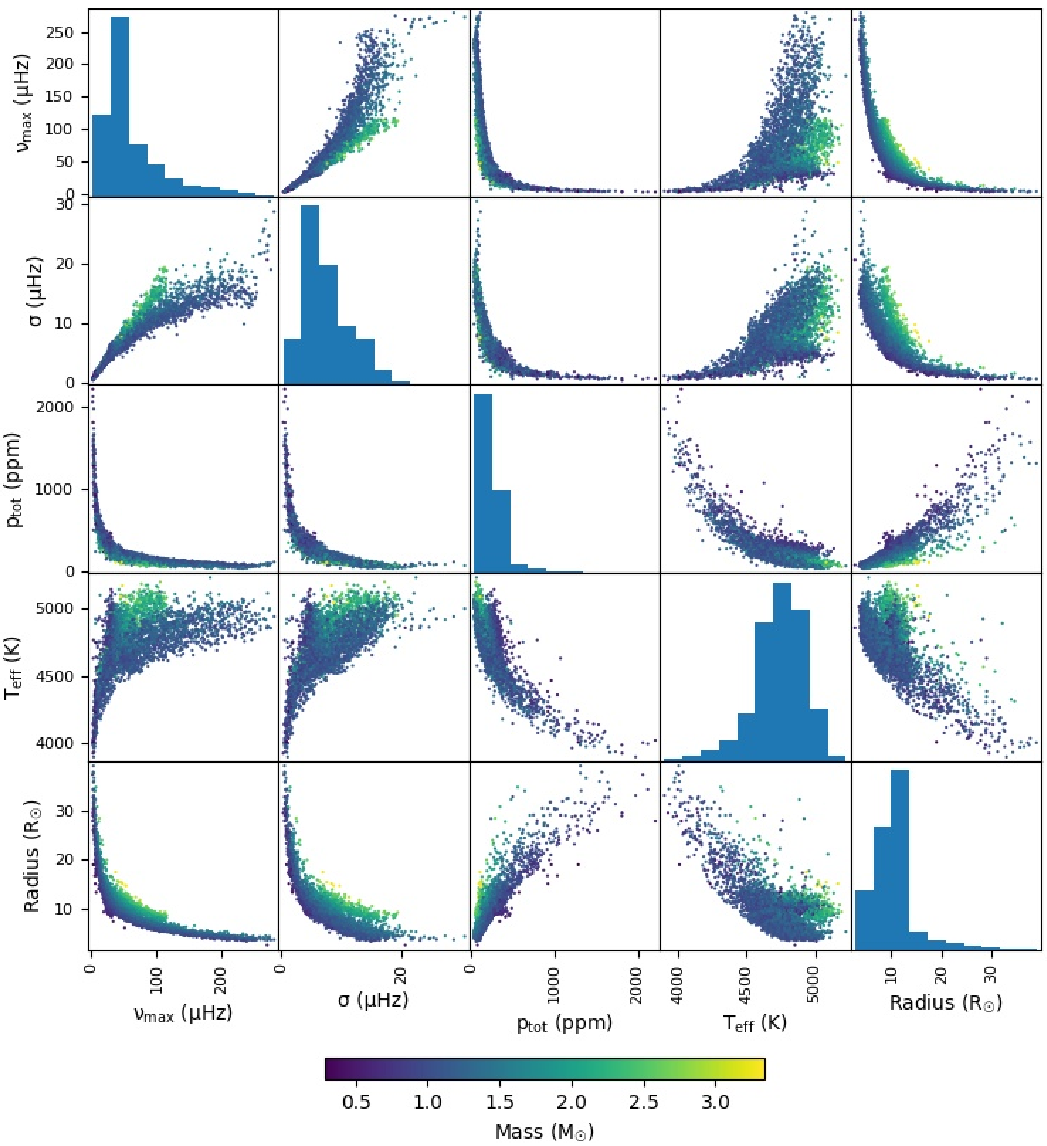



Figure 16: Initial values plotted against each other with the mass color-coded.

This plot also demonstrates why we cannot simply derive an analytic expression to get values for radius and mass, provided that no $\Delta\nu$ measurements are available. The parameter space is quite large and moreover the parameters depend on each other. Thus, finding an analytic expression is challenging. Machine learning, in contrast, works very well for such tasks.

In Fig. 16 one can see that there is a tight relation between $\nu_{\max}$ and radius, which is comprehensible given the fact that oscillations need less time to travel through a small star than through a big star. From the last column we can infer that all the other parameters also depend strongly on the radius.

The subplots $\nu_{\max}$ vs. radius, $\sigma$ vs. $\nu_{\max}$ and $\sigma$ vs. radius show that higher mass stars follow a different relation than lower mass stars. This happens because higher mass stars are more luminous than lower mass stars and they have a certain minimum luminosity. Thus, according to Eq. (38) there also exists an upper limit for $\nu_{\max}$ for these stars.

The diagonal in Fig. 16, where one variable is plotted against itself, displays the distribution of this quantity. From the distribution of $\nu_{\mathrm{max}}$ we can deduce that most stars have a frequency of maximum oscillation power less than $100\,\mu$Hz. Furthermore the data set includes mostly stars with a radius less than $20\,\mathrm{R}_{\odot}$ and with effective temperatures around $4500 - 5000\,$K.

# 7 Machine Learning

Whether in medicine, economy, engineering or natural sciences: Machine learning (ML) has proven to be an important tool in various scientific disciplines.
The basic idea of ML is that a computer can learn by itself through experience without any User interaction, where this 'experience' normally concerns a data set (Garreta and Moncecchi, 2013; Müller, 2017): The learning is achieved by algorithms, so-called *estimators* which adapt progressively to a data set and thereby improve their fitting performance (Alpaydin, 2004). This way, the User obtains a model, which is able to predict the outcome of the data, to which it was previously fitted. In order to implement ML, the *Python* library *scikit-learn* is used, that includes many different estimators as well as other ML tools (Pedregosa et al., 2011).
In this section I will introduce some theoretical terminologies concerning ML, such as supervised and unsupervised learning. Furthermore, I will explain in Sec. 7.2 how *regression problems*, corresponding to problems where the values to predict are continuous, are solved. Therefore, the outcome of four different learning algorithms is discussed, using the KIC data set which was introduced in the previous section. In doing so, I will show in a transparent way, why the *random forest* algorithm is selected for predicting the radii and masses of solar-like oscillators. The random forest is an *ensemble* algorithm, meaning that it is composed of numerous *decision tree* algorithms. As the latter is more intuitive than the random forest algorithm, its functionality is explained in Sec. 7.2.1. Based on this, Sec. 7.2.2 finally discusses the random forest algorithm, and it is shown how the best random forest model for the KIC data set is derived.

## 7.1 Supervised and Unsupervised Learning

There are two ways to perform ML, namely either supervised or unsupervised. Supervised learning means that the algorithm learns from pairs of input (*source*) and known output (*target*) values, $([x_1, \ldots, x_n]^T, y_n)$ (Buitinck et al., 2013). Emanating from the source data $\boldsymbol{x}$, the learning algorithm establishes general rules which lead to the corresponding target values $\boldsymbol{y}$ (Garreta and Moncecchi, 2013). A model is thus created which can be applied to previously unknown data, where only the source but not the target values are known. Unsupervised learning algorithms on the other hand are only learning from source values. This learning procedure aims to recognize specific patterns in the data. As this is not done in this work, it will not be discussed in more detail.
The source data can be a composition of various quantities, called *features* that describe the properties of the data points, called *samples*. As the samples in this work correspond to stars, the features, namely $\nu_{\max}$, $\sigma$, $p_{\text{tot}}$ and $T_{\text{eff}}$ describe their physical properties. The target variables can also have more than one dimension, in which case the model predicts more than one quantity – this is referred to as *multioutput*. As this master's thesis aims to predict the radii and masses of solar-like oscillators in a supervised way (which is why radius and masses are calculated beforehand using the nonlinear scaling relations), we are faced with such a multioutput problem. Fur such issues a simultaneous prediction of the target values can be advantageous as the model can identify and use the correlations between the target values itself (Borchani et al., 2015).
A common problem that often arises when supervised methods are concerned is *overfitting*. Overfitting occurs when the resulting model, previously fitted to some data, becomes too complex so that it predicts this data perfectly but fails to predict the outcome of unknown data (Müller, 2017). To control overfitting, the initial data set is split into a *training* and a *test set*. The model is built up only by the training set and as soon as the target variables are represented in a good way, it is applied to the test set, aiming to predict its target values using the rules that were previously learned. By checking the prediction quality of both the training and the test set it can be ensured that overfitting does not occur. Usually the prediction ability of the test sample is a bit worse compared to that of the training set. Another similar but contrary problem is *underfitting*. It arises when the model is too simple so that it neither predicts the target values in the training set nor in the test set accurately.

However, this can again be avoided by controlling the performance of both sets.
One can explicitly express the prediction quality using the *root mean squared error*

$$RMSE\,(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\,, \tag{53}$$

which is a loss function frequently applied to regression problems (Pedregosa et al., 2011). Here, $n$ is the number of samples, while $y$ and $\hat{y}$ are the true and predicted values, respectively. As this metric is highly dependent on the data set which it is applied to, I additionally compute the *mean absolute percentage error*

$$MAPE\,(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{54}$$

for the test sets. The closer the $MAPE$ and $RMSE$ values are to zero the better is the prediction. The $MAPE$ returns a relative error which is why it is dimensionless, contrary to the $RMSE$ that has the same unit as the predictions. The advantage of a relative error (for this master's thesis) is that it enables an easier comparison between the achieved radius and mass prediction performances. However, as it also has drawbacks such as a preference towards underestimated predictions, $RMSE$ is used over $MAPE$ in the model-selection process as the former is indifferent to the direction of errors and hence unbiased.

## 7.2 Learning Algorithms

There are many different supervised learning algorithms and it is not trivial to choose one among them. On the one hand the estimator itself should not be too complex in order to prevent overfitting, on the other hand it should not be too simple either so that underfitting is avoided. Thus, the choice of the estimator depends on the data set it is applied to. For a rather simple data set a less complex estimator, corresponding to less complex models, can be used. One could also compare the balance between over- and underfitting with trading off predictability against interpretability, as it is often not necessary, and recommendable, to select the least interpretable model with the (theoretical) highest performance (Fig. 17).

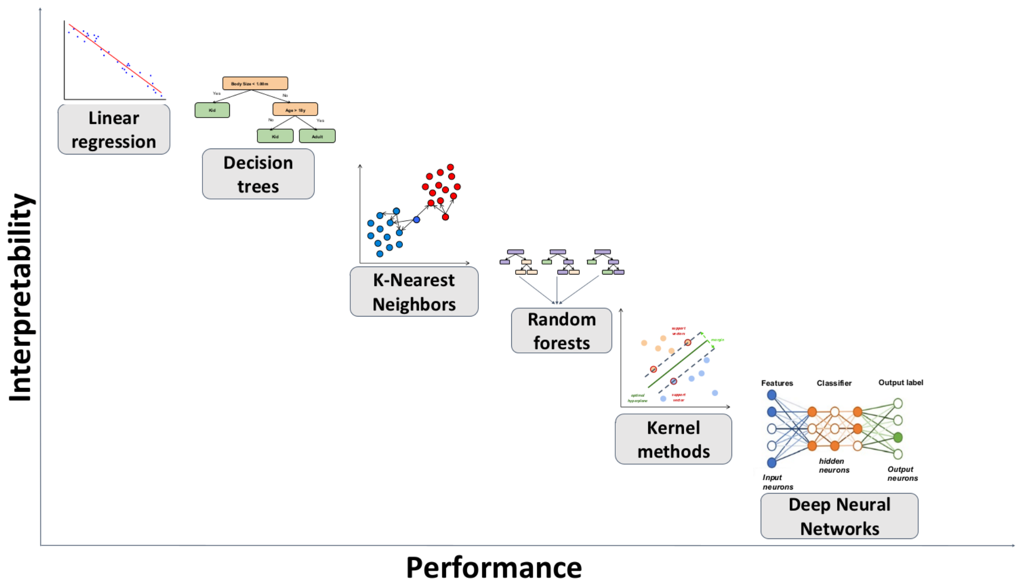



Figure 17: Various ML algorithms are plotted in an interpretability vs predictability plot. Figure taken from Badillo et al. (2020).

As can be seen, one of the most interpretable learning algorithms are linear regression models. In contrast to more complex models, however, they have a rather low prediction performance. While a decision tree model is also comparatively easy to interpret, it is often not able to predict the target values as accurately as, for example, random forest models, which use an ensemble of decision trees for the prediction.
In the following I will present four plots (Fig. 19, Fig 20) which show the performance of different estimators, namely linear regression, decision tree, random forest and neural network. All of these estimators are supervised learning algorithms. However, before the performances of the models can be checked, the hyper-parameters, which are given as arguments to the estimator rather than learned by it, have to be adjusted accordingly. Hyper-parameters are different for each estimator and the number of those usually increases with the complexity of the estimator: While there are only three hyper-parameters to adjust for the *LinearRegression* estimator there are 23 hyper-parameters for the neural network. The hyper-parameters of an estimator can be optimized using the *scikit-learn* function *GridSearchCV* or *RandomizedSearchCV*. The difference between those *meta-estimators* is that while *GridSearchCV* defines the best hyper-parameters by exhaustively searching all possible combinations within a given parameter grid, *RandomizedSearchCV* draws $n$ random combinations from given hyper-parameter distributions, where $n$ is a User-specified number of iterations (Buitinck et al., 2013). Hyper-parameters influence the estimators complexity and the optimization of the these could thus lead to overfitting. To avoid this, an additional *validation set* (often abbreviated as 'val.'), which is a percentage of the training set, is defined. As the reduction of samples within the training set might lead to a decrease in model performance, we take advantage of a scheme referred to as *cross-validation* Pedregosa et al. (2011). Fig. 18 shows exemplary a five-fold cross validation.

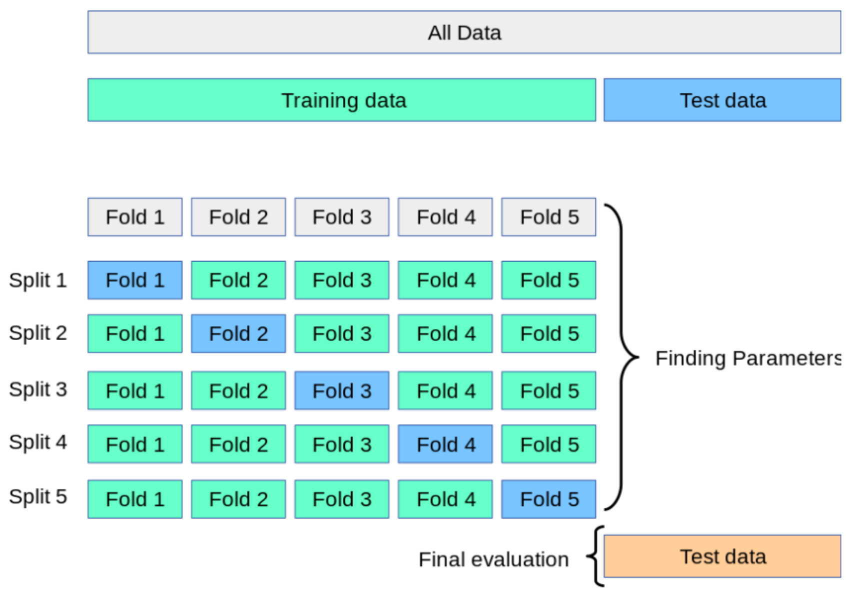



Figure 18: Five-fold cross-validation procedure for hyper-parameter optimization. Figure taken from Pedregosa et al. (2011).

Following the procedure of Fig. 18, the training set is split into $k$ folds (i.e. $k$ subgroups of the data set), where $k-1$ folds are used to train the model while the remaining fold is used for validation. The training and validation set are split and re-shuffled $k$ (e.g. five) times for one hyper-parameter combination, where the $RMSE$ of the training and validation set is stored for each iteration. As radius and mass are predicted simultaneously (i.e. *multioutput*), the $RMSE$ returns the averaged $RMSE$ for the radius and mass predictions and has therefore to be scaled by the mean true target value.

A snippet of the *Python* code, demonstrating how this is implemented, is shown below.

```
from sklearn.metrics import mean_squared_error

def RMSE_S(y_true, y_pred):
    return mean_squared_error(y_true, y_pred, squared=False, multioutput=[1./
                                              np.mean(y_true.radius), 1./np.mean(
                                              y_true.mass)])
```

Here the keyword 'squared' specifies if the *mean squared error* ($MSE$) or the $RMSE$ is calculated. Via the keyword 'multioutput', one is able to determine, how the different scores for the multioutput should be weighted. An index $S$ is appended, whenever this scaled $RMSE_S$ is applied. This way it can be ensured that the resulting $RMSE_S$ is dimensionless and that both the performance measures for radius and for the mass predictions are treated equally. Thus, the procedure just described results in k different $RMSE_S$ values which are averaged and from which the error is estimated using the standard deviation. The lowest $RMSE_S$ defines the best hyper-parameter combination for a given estimator, which is then finally applied to the test data.
The hyper-parameters for each of the four estimators are optimized in this way. The *GridSearchCV* is used for the linear regression and the decision tree, as their hyper-parameter space is comparatively small and the computation time can manage to go through all given hyper-parameter combinations. For the remaining two estimators, the meta-estimator *RandomizedSearchCV* is used. In the plots the $RMSE_S$ of the validation sets as well as the $MAPE$ and $RMSE$ of the test set are given in each legend. The $MAPE$ and $RMSE$ of the test set are calculated separately for radius and mass, which is why I do not have to scale the $RMSE$. As the test set is only evaluated once, I do not give an error estimation.

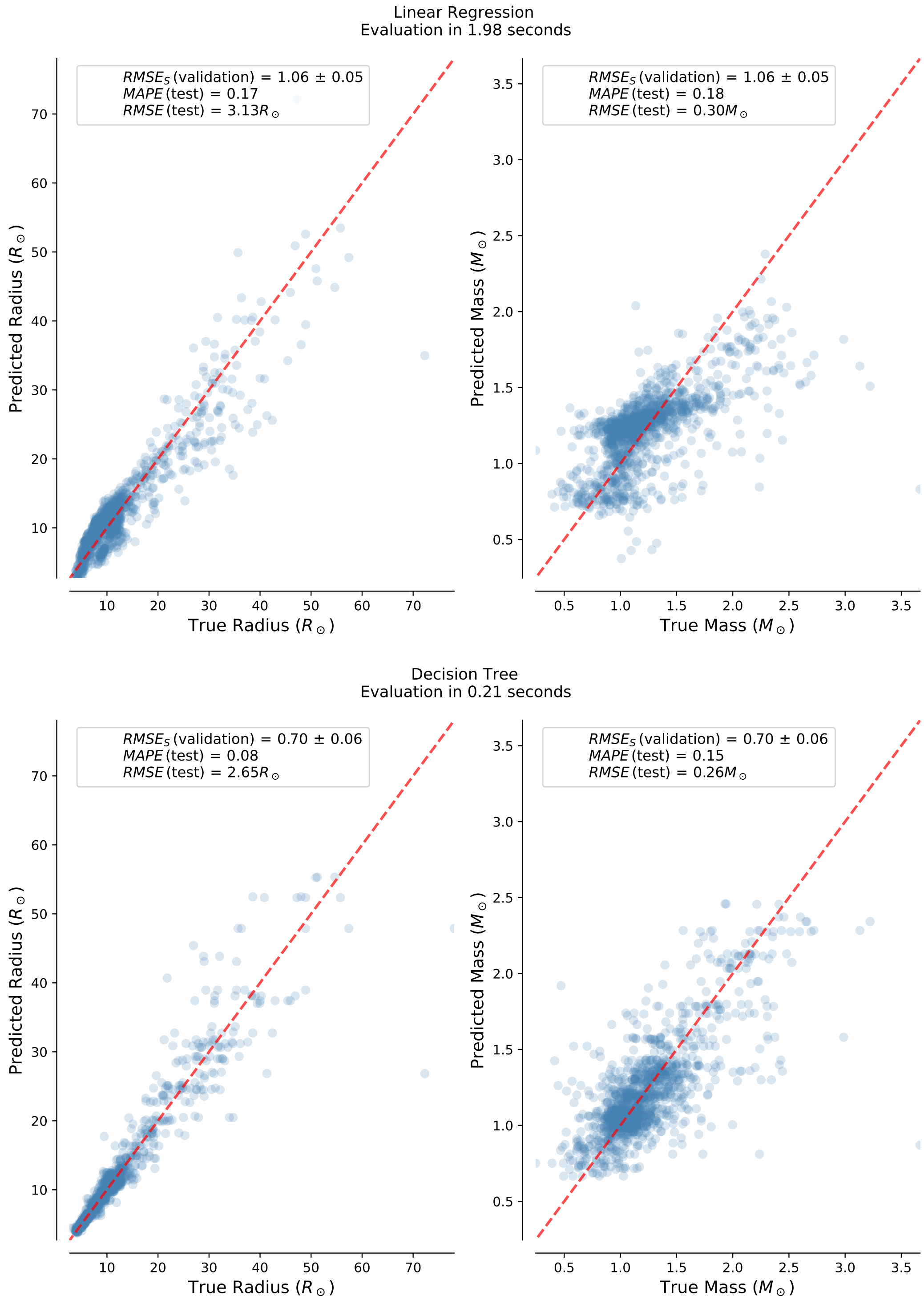



Figure 19: Results for supervised learning methods using the best hyper-parameter combination for each estimator. The $RMSE_S$ of the validation as well as the $MAPE$ and $RMSE$ of test set are given in the legend. On the x-axis are the true values plotted, on the y-axis the predicted ones. Top: Prediction performance of a linear regression model. Compared to the $MAPE$ and $RMSE$ of the other models, neither the prediction of the radius ($MAPE = 0.17$, $RMSE = 3.13\,R_{\odot}$) nor the prediction of the mass ($MAPE = 0.18$, $RMSE = 0.30\,M_{\odot}$) is very good. Bottom: The target values are predicted using a decision tree model. The performance of the validation set ($RMSE_S = 0.70 \pm 0.06$) is considerably better compared to the linear regression model. The $MAPE$ and $RMSE$ of the test set improve too.

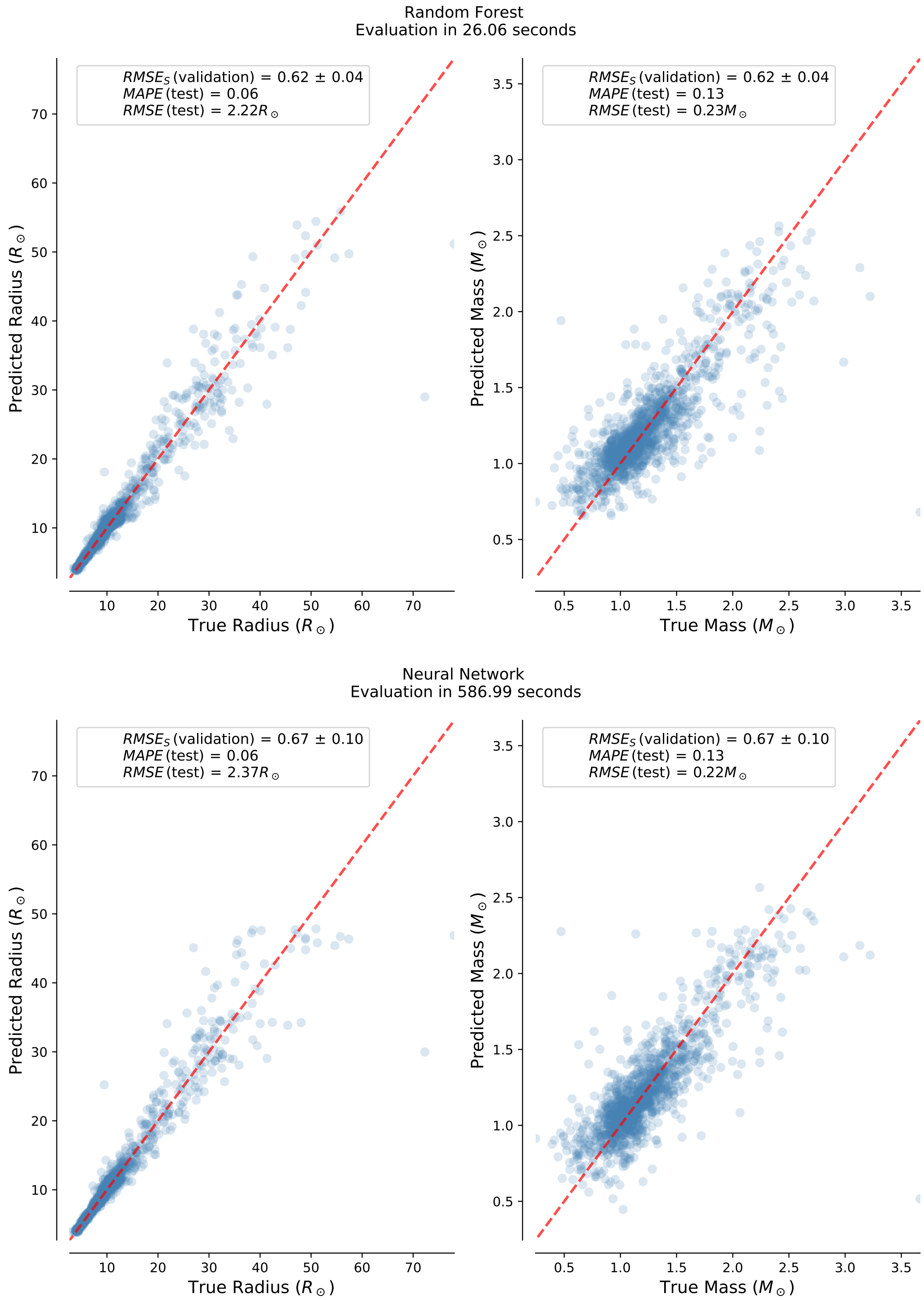



Figure 20: Results for supervised learning methods using the best hyper-parameter combination for each estimator. Top: Predicted values are plotted against true values using the random forest model. Bottom: A neural network method is used to predict radii and masses. The $RMSE_S$ of the validation as well as the $MAPE$ and $RMSE$ of test set are again given in the legends.

The relatively poor $RMSE_S$ (val.) $= 1.06 \pm 0.05$ of the validation set of the linear regression estimator indicates, that the resulting model is too simple for our data set and therefore causes underfitting. The performance of the validation set increases ($RSME_S$ (val.) $= 0.70 \pm 0.06$) when using a decision tree estimator. It becomes even better ($RSME_S$ (val.) $= 0.62 \pm 0.04$) as a random forest model is used. However, the prediction quality of the validation set does not improve but actually declines ($RSME_S$ (val.) $= 0.67 \pm 0.10$) for the neural network estimator, contrary to the proposition in Fig. 17. This leads to the conclusion, that the random forest model describes the data sufficiently well, and that there is not much to gain by using a more complex model. This is why the random forest regressor is chosen as the final estimator for the radius and mass prediction of solar-like oscillators. Furthermore, as can be gathered from the title of each figure, this estimator is approximately 23 times faster than the neural network.
From the figures above it can be concluded, that the $MAPE$ and $RMSE$ of the test set vary accordingly with $RMSE_S$ of the validation set. We note that while the $MAPE$ is always better for the radius prediction, the $RMSE$ is better for the mass prediction. This is because the $RMSE$ is highly dependent on the data set: The stellar masses cover a smaller value range than the radii, which therefore leads to a smaller $RMSE$ value. Hence, the $RMSE$ of the target values can only be compared to the $RMSE$ of the corresponding target value for a different model, whereas the $MAPE$ is used to compare the radius prediction performance with the mass prediction performance of the same model. That means, that all four models indeed predict the radii better than the masses.

### 7.2.1 Decision Tree

As the random forest model is relatively challenging to interpret, this section demonstrates how one decision tree in general works, using the *scikit-learn* command *DecisionTreeRegressor* and the KIC data set (Sec. 6).
As already pointed out in Sec. 7.1, a model is created by learning general rules from source and target variables which are in the present case simple decision rules. Fig. 21 shows a very simple decision tree for a multioutput problem (i.e. the tree aims to predict radius as well as mass) with depth two. The first rectangle is the *child node*, the two rectangles in the second row define so-called *decision nodes* at which the data are further split using one feature as splitting criterion, and the rectangles in the third row correspond to *leaf nodes*. The model partitions a child node or decision nodes by minimizing the *mean squared error* ($mse$) of the predictions within these nodes. A decision tree is able to predict the target values for unknown data by following the path from the child node to the leaf nodes. The leaf nodes therefore contain the results, which are the averaged target values of all training samples within that node. The first and second entry in the square brackets correspond to the averaged radius and mass, respectively. The parameter X[0] in this figure is equivalent to the feature $\nu_{\max}$, while X[1] indicates the feature $\sigma$.
The training set comprises 80% of the original 6657 KIC stars (hence 5325 stars) which are used to build the model while the remaining 1332 stars belong to the test set. In the child node the training set is split for the first time by requiring that $\nu_{\max} \leqslant 10.461$. Stars for which this condition is true end up in the left decision node, while stars which do not fulfill this condition end up in the right one.

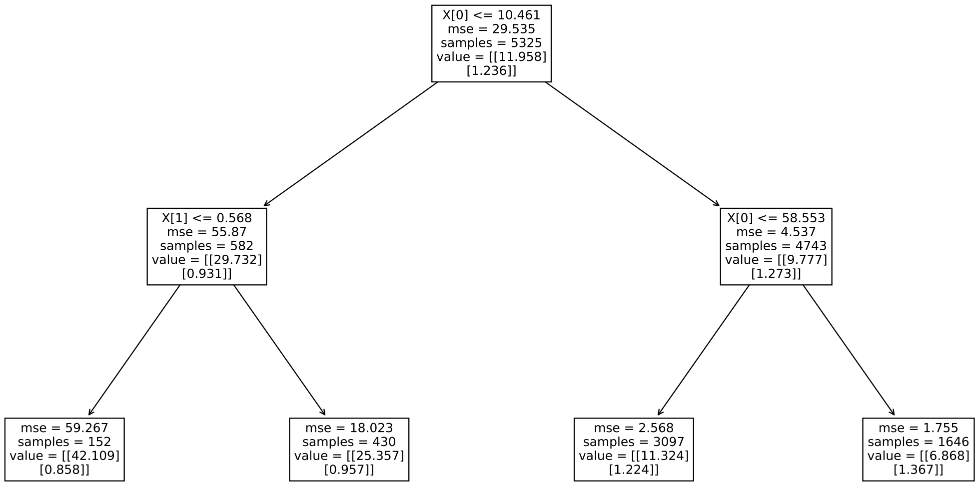



Figure 21: Decision tree with depth two evaluated for the KIC data set.

Let us now study the leftmost leaf node in more detail. According to the figure, there are 152 stars with $\nu_{\max} \leqslant 10.461\,\mu$Hz and $\sigma \leqslant 0.568\,\mu$Hz. The average radius of these stars is $42.109\,R_{\odot}$ while the average mass is $0.858\,M_{\odot}$. In this example only $\nu_{\max}$ and $\sigma$ are used to split the nodes. However, other features can be used for splitting the nodes as well. One can explicitly quantify the importance of a feature for building the tree by calling an adequately named *scikit-learn* function. This function specifies how often one feature is used for partitioning the nodes relative to the other features.
This decision tree model, with a maximum depth of two, is not able to predict the radii and masses of the stars accurately. For example, the predicted radius and mass for any star which satisfies $\nu_{\max} \leqslant 10.461\,\mu$Hz and $\sigma \leqslant 0.568\,\mu$Hz would be $42.109\,R_{\odot}$ and $0.858\,M_{\odot}$, respectively. Thus the model obviously underfits the data. In order to improve the prediction quality one needs to create a deeper tree. The depth of a decision tree cannot be arbitrarily deep, though, since this would lead to overfitting (Fig. 22).

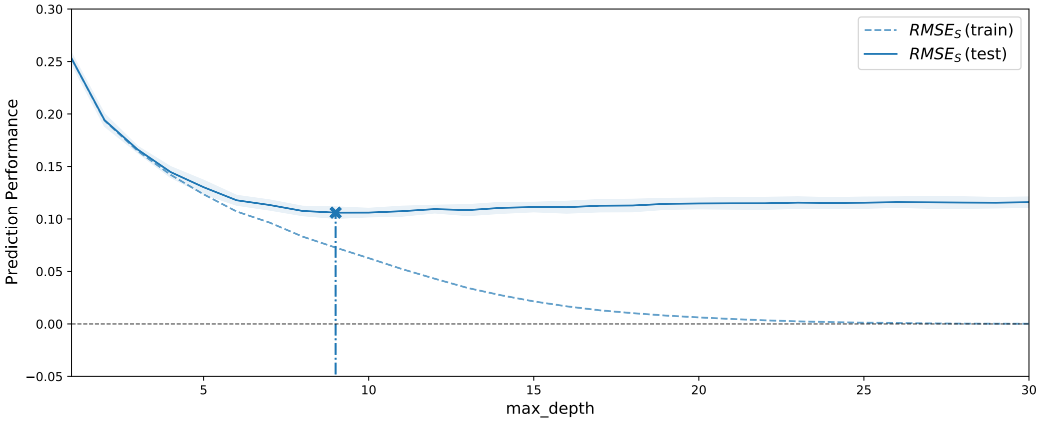



Figure 22: The prediction performance for decision trees with depth 1 to 30. While a decision tree with depth $\lesssim 5$ underfits the data, a decision tree with depth $\gtrsim 20$ causes overfitting. An appropriate choice would be a maximum depth of nine.

Up to a depth of 5 the model underfits while a model with maximum depth $> 20$ overfits. Overfitting occurs when the training set is predicted perfectly ($RMSE_S$=0) while the prediction ability of the test set decreases. This is why it is important to always check the training and test score, as was already pointed out in Sec. 7.1. The best choice, corresponding to the lowest $RMSE_S$ value, for the maximum depth of this decision tree is nine.

### 7.2.2 Random Forest

In order to increase the maximum depth but avoid overfitting at the same time, we can use an ensemble of decision trees. This is enabled by the so-called random forest (RF) estimator, where the randomness is introduced via *bootstrapping*: From $n$ data points $n$ elements are one after the other randomly pulled and replaced. This results in a new data set in which some of the original data points appear more than once while others are left out. For each of these slightly different data sets a decision tree is constructed. Overfitting can be reduced by averaging the results of the individual trees (Müller, 2017). Another possibility to make this estimator even more random is to randomly choose a subset from all features so that each node of a tree has a different choice of features to split the data. Although this makes the different trees quite diverse, which again reduces overfitting and could lead to a better prediction performance, it is recommendable to consider all features at each split for regression problems (Müller, 2017).
As already mentioned, the hyper-parameters for the RF estimator are optimized using the *RandomizedSearchCV*. For the RF estimator there are 18 different hyper-parameters but only a few of those have to be adjusted to receive better results, while other values can be left at the default values from *Python* (Müller, 2017). The most important hyper-parameters are given in Tab. 3.

| Hyper-parameter | Definition | Range |
|---|---|---|
| $n\_estimators$ | number of decision trees used | $100 - 700$ |
| $max\_depth$ | maximum depth of the tree | $5 - 200$ |
| $min\_samples\_split$ | minimum number of samples required to split a node | $2 - 20$ |
| $min\_samples\_leaf$ | minimum number of samples within a leaf node | $1 - 20$ |

Table 3: Important hyper-parameters to adjust for the RF estimator. The ranges in the last column define the interval from which the hyper-parameters for the *RandomForestRegressor* are randomly picked.

In the last column I specified the range from which a random integer is picked for each iteration of the *RandomizedSearchCV*. As the *RandomizedSearchCV* is iterated 100 times, we get 100 different random hyper-parameter combinations for which the $RMSE_S$ of the five validation sets (i.e. five folds) is calculated (see Fig. 18), which are averaged in each run.

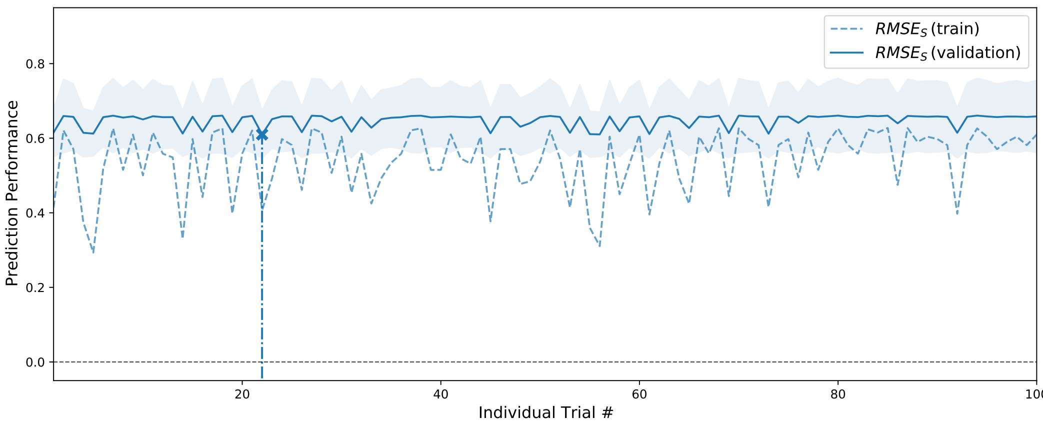



Figure 23: $RMSE_S$ for 100 different hyper-parameter combinations. The best hyper-parameter combination from the *RandomizedSearchCV* achieves $RMSE_S = 0.61 \pm 0.06$.

It can be seen, that the prediction quality of the training set is indeed better than the one of the validation set, as was proposed in Sec. 7.1. The best $RMSE_S$ (val.) $= 0.61 \pm 0.06$ of the validation set is marked and corresponds to the following hyper-parameter combination:

| | |
|---|---|
| $n\_estimators$ | 502 |
| $max\_depth$ | 90 |
| $min\_samples\_split$ | 3 |
| $min\_samples\_leaf$ | 2 |

Fig. 23 highlights another advantage of the RF model, namely the robustness of the model: Even though the hyper-parameters are selected randomly 100 times, the performance of the model does not vary much but levels off at around 0.65. The neural network, in contrast, reacts strongly to changing hyper-parameters (see Fig. 24).

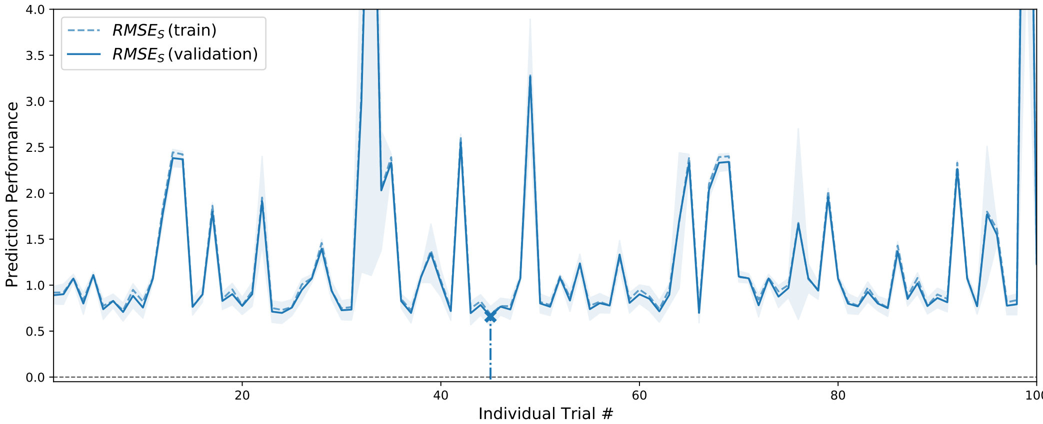



Figure 24: $RMSE_S$ for 100 different hyper-parameter combinations of the neural network estimator. Contrary to the RF estimator the $RMSE_S$ varies a lot due to changing combinations of hyper-parameters.

Having set the hyper-parameters, the following procedure has to be carried out to get the desired RF model. The results of this *Python* script and specifically of the RF model are discussed in the next section.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
from joblib import dump

# Define MAPE and RMSE
def mape(y_true, y_pred):
    return np.mean(np.abs((y_true - y_pred) / y_true))

def rmse(y_true, y_pred):
    return np.sqrt(mean_squared_error(y_true, y_pred))

# Import the KIC data set
data = pd.read_csv('KIC_data_set.csv', header=0)

# Exclude radius and mass from features
# The features are: fmax, sigma, ptot, Teff
X = features.drop('radius', axis=1)
X = X.drop('mass', axis = 1)

# Define radius and mass as target variables --> multioutput
y = pd.concat((features.radius, features.mass), axis=1)


# Split the data into training (0.80) and test (0.20) sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                       random_state=42)

# Initializing the RandomForestRegressor using the best hyper-parameter
                                       combination
rf = RandomForestRegressor(bootstrap=True, n_estimators=502, max_depth=90,
                                       min_samples_split=3, min_samples_leaf=2
                                       )

# Create a model using the training set
rf.fit(X_train, y_train)

# Predict the target values of the test set using the learned model
y_pred = rf.predict(X_test)

# Print MAPE and RMSE
print(mape(y_test, y_pred), rmse(y_test, y_pred))

# Print feature importanced
print(rf.feature_importances_)

# Save the RF model
dump(rf, 'Radius_Mass_Prediction.joblib')
```

# 8 Creating the Models Using the KIC Data Set

In the end not one but three RF models for different initial conditions are created: The first model (model 1) takes, as already mentioned, four features, namely $\nu_{\text{max}}$, $\sigma$, $p_{\text{tot}}$ and $T_{\text{eff}}$. One model (model 2) does not need the temperature $T_{\text{eff}}$ as input, since the effective temperatures can sometimes not be derived. On the other hand, there are occasionally metallicity measurements available which is why a third model (model 3) takes the [Fe/H] ratio as an additional feature. All three models are soon to be published on *Github*[10].
In the following the results of the three RF models are presented. All models are constructed based on the same KIC training set, only varying in the number of features.

## 8.1 Model 1

As the best hyper-parameters were already deduced in Sec. 7.2.2, I can immediately present the feature importances of the model, which quantify how important the different features for splitting the nodes and therefore building the trees are:

| | |
|---|---|
| $\nu_{\text{max}}$ | 0.873 |
| $\sigma$ | 0.096 |
| $p_{\text{tot}}$ | 0.015 |
| $T_{\text{eff}}$ | 0.016 |

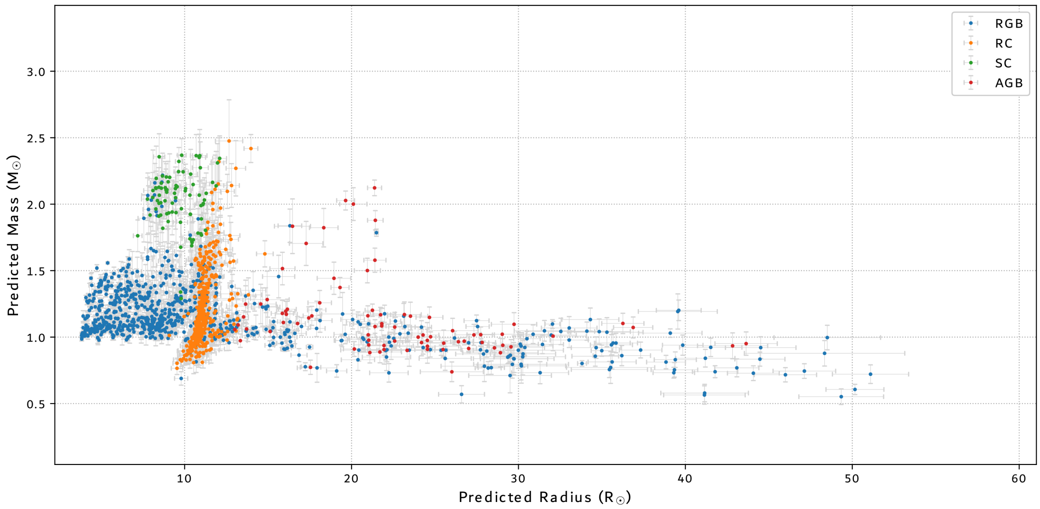
Apparently $\nu_{\text{max}}$ is by far the most important parameter with a relative importance of 87.3%. This is not very surprising when considering Fig. 16, which shows a very tight relation between $\nu_{\text{max}}$ and the radius. The model is saved using the *dump* command from the *Python* library *joblib*[11]. It is then imported into a file called 'Radius_Mass_Prediction.py', where it defines a method of the class RadiusMassPrediction(). Thus, one is able to predict radius and mass of previously unknown stars by simply executing:

```
from Radius_Mass_Prediction import *
radius , mass = RadiusMassPrediction(pd.DataFrame(features)).Prediction()
```

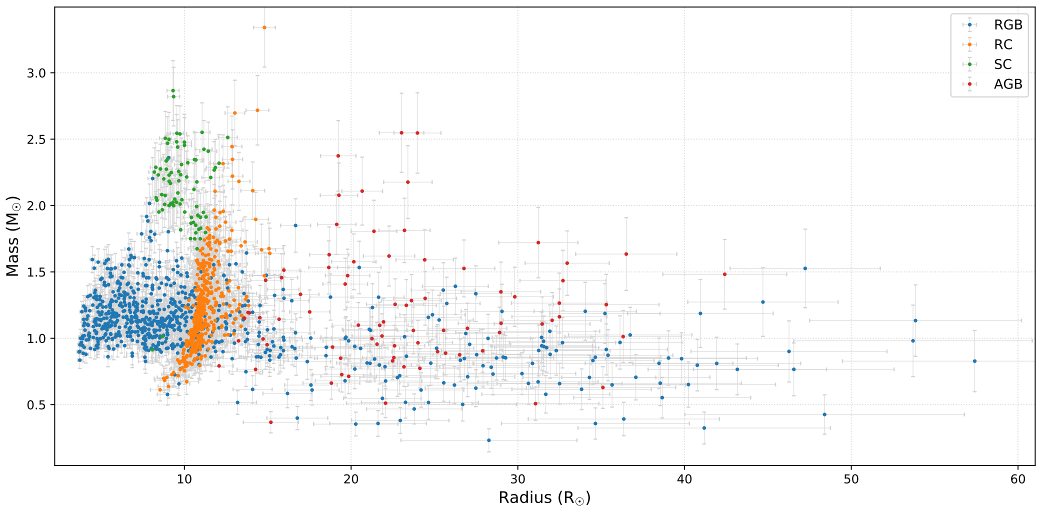
This code is run 100 times, where the features are randomly varied according to a normal distribution with the mean and the standard deviation corresponding to the observed value and its uncertainty, respectively. The individual predictions build up a distribution for the radius and mass from which average values and their uncertainties (i.e. the standard deviation of the distribution) are determined.

[10] https://github.com/
[11] https://joblib.readthedocs.io

These predictions for the stars in the test set are visualized in a M-R plot in Fig. 25a.




(a)




(b)

Figure 25: (a) Predicted M-R plot of the test set. (b) True M-R plot of the test set.

As a reminder, Fig. 25a shows the averaged predicted target values of the test set, which comprises only 20% of the whole data set. For comparison, a similar plot as Fig. 8 is displayed but here only with the 1332 stars contained in the test set (Fig. 25b). Considering both figures we note that all stars with radii $> 60\,R_{\odot}$ are missing in the plots, meaning that these stars are part of the training set. This is fortunate, since stars with $R > 60\,R_{\odot}$ would never be predicted if the model was not trained on them. At first sight one notices, that the scatter in mass for stars with radii $> 20\,R_{\odot}$ is much smaller in the predicted M-R plot than in the true one. The radius range becomes also a bit smaller in Fig. 25a, though apparently less severely compared to the mass range. Emphasizing a positive effect, on the other hand, the different evolutionary stages, such as the RC stars which suffer from mass loss after the RGB-tip, as well as the more massive SC stars, are clearly visible and the overall appearance of the true M-R plot seems to be reproduced adequately. To quantify this, the performance parameters $MAPE$ and $RMSE$, given for radius and mass separately, are given below.

| | $MAPE$ (test) | $RMSE$ (test) |
|---|---|---|
| Radius | $0.051 \pm 0.001$ | $(1.56 \pm 0.05)\,R_\odot$ |
| Mass | $0.100 \pm 0.001$ | $(0.173 \pm 0.004)\,M_\odot$ |

According to the achieved $MAPE$, the performance of the radius predictions is better than the performance of the mass predictions. To be able to comprehend this in a more descriptive way, the ratio of the predicted to the true values plotted against the true target values is shown in Fig. 26.

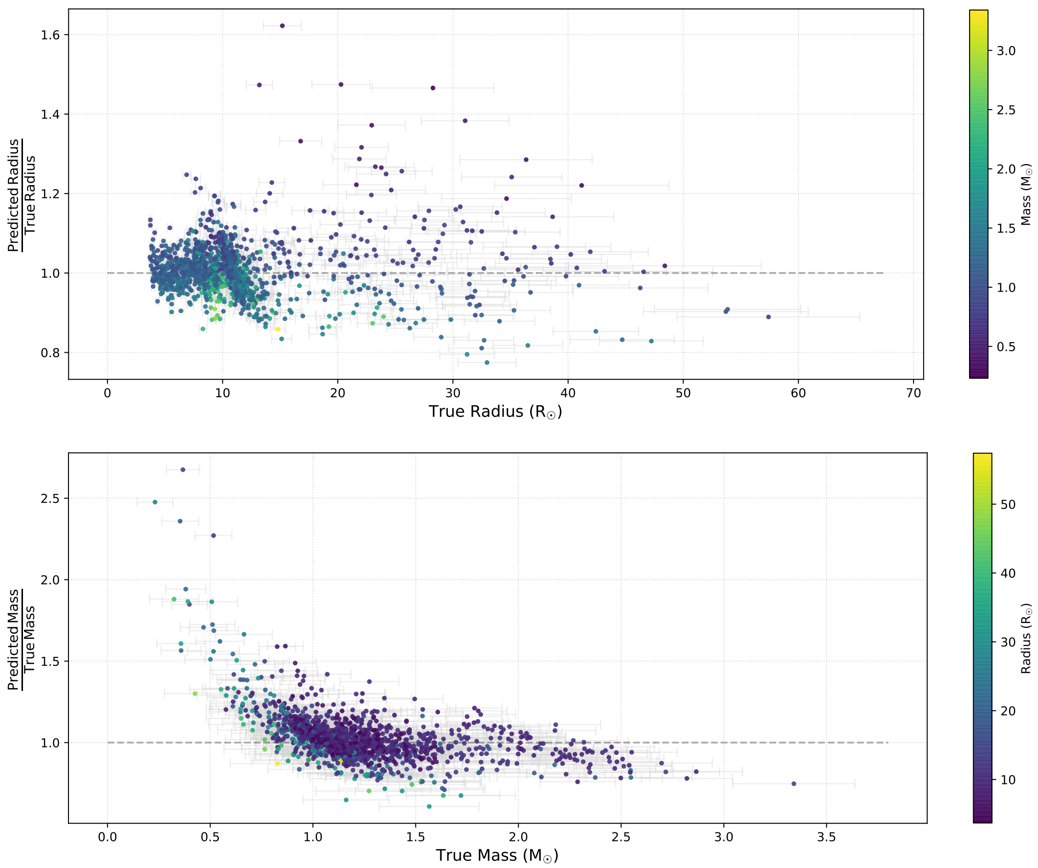



Figure 26: The ratio of predicted-to-true target values vs. the true values using the input parameters $\nu_{\max}$, $\sigma$, $p_{\rm tot}$, $T_{\rm eff}$. Top: The result for predicted radii with the calculated mass color-coded. Bottom: The result for predicted masses with the true radius color-coded.

From the first plot we can deduce that the prediction works quite well for stars with $R_{\rm true} < 15\,R_\odot$. This seems evident, as this is the case for 90% of all stars in the sample: The model fits obviously best where the majority of the data lies. It is quite striking that the predicted radii of higher-mass stars (green and yellow) tend to be underestimated while lower-mass stars (blue and violet) are more likely to be overestimated.
The second plot shows that the model works best for stars with $\sim 1\,M_\odot < M_{\rm true} < 2\,M_\odot$, where only the masses of some larger stars ($R_{\rm true} > 20$) are noticeably underestimated. The predicted masses for most of the stars with masses $< 1.0\,M_\odot$ are overestimated for stars with both small and large radii. However, the predicted masses of small stars ($R_{\rm true} \sim 10\,R_\odot$, violet) are less drastically overestimated than the predicted masses of larger stars ($R_{\rm true} > 10\,R_\odot$). This can also be seen when comparing Fig. 25a to Fig. 25b: While there are many stars with $M_{\rm true} < 0.5\,M_\odot$ in the true M-R plot, there are no predicted masses within the same mass range in the predicted M-R plot which explains this severe mass overestimation.

Less massive RGB and RC stars with $R_{\rm true} < 20 R_\odot$ are also overestimated by the model, though in a less extreme extent. The masses of stars with $M_{\rm true} > 2\,M_\odot$, on the other hand, are underestimated, especially for stars with $R_{\rm true} < 20 R_\odot$, which makes sense as there are no such massive stars in the higher radius range ($R_{\rm true} > 20 R_\odot$).
To summarize, the model overestimates the radii as well as the masses of the least massive stars and underestimates the target values of the highest mass stars.

## 8.2 Model 2

In this section the prediction ability of the model with only a reduced set of features, namely $\nu_{\rm max}$, $\sigma$ and $p_{\rm tot}$, will be discussed.
As a whole new model is created, the best parameters have again to be determined following the procedure outlined in Sec. 7.2. Fig. 27 shows the outcome of the *RandomizedSearchCV*.

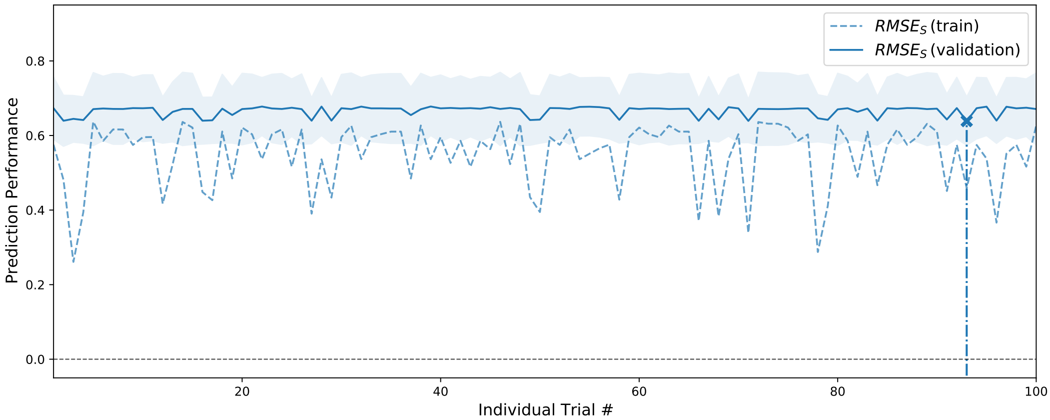



Figure 27: $RMSE_S$ of the validation set after 100 iterations of the *RandomizedSearchCV*. The best prediction performance $RMSE_S = 0.64 \pm 0.07$ is highlighted. It corresponds to the best hyper-parameter combination for model 2 which *RandomizedSearchCV* found.

The best $RMSE_S$ (val.) $= 0.64 \pm 0.07$ of the validation set corresponds to the following hyper-parameter combination:

| | |
|---|---|
| $n\_estimators$ | 354 |
| $max\_depth$ | 116 |
| $min\_samples\_split$ | 10 |
| $min\_samples\_leaf$ | 1 |

The RF model with this set of hyper-parameters is saved, and the corresponding feature importances are:

| | |
|---|---|
| $\nu_{\rm max}$ | 0.886 |
| $\sigma$ | 0.096 |
| $p_{\rm tot}$ | 0.018 |

Two of the three parameters gain importance, which seems natural as there is one less parameter in total.

The *Python*-script:

```
from Radius_Mass_Prediction import *
radius , mass = RadiusMassPrediction(pd.DataFrame(features)).
                                        Prediction_without_Teff()
```

is run 100 times while the features of the test set are randomly varied along a normal distribution. Here the method 'Prediction_without_Teff()' loads the previously stored model. These 100 predicted radius and mass values of the test set are averaged and plotted in Fig. 28.

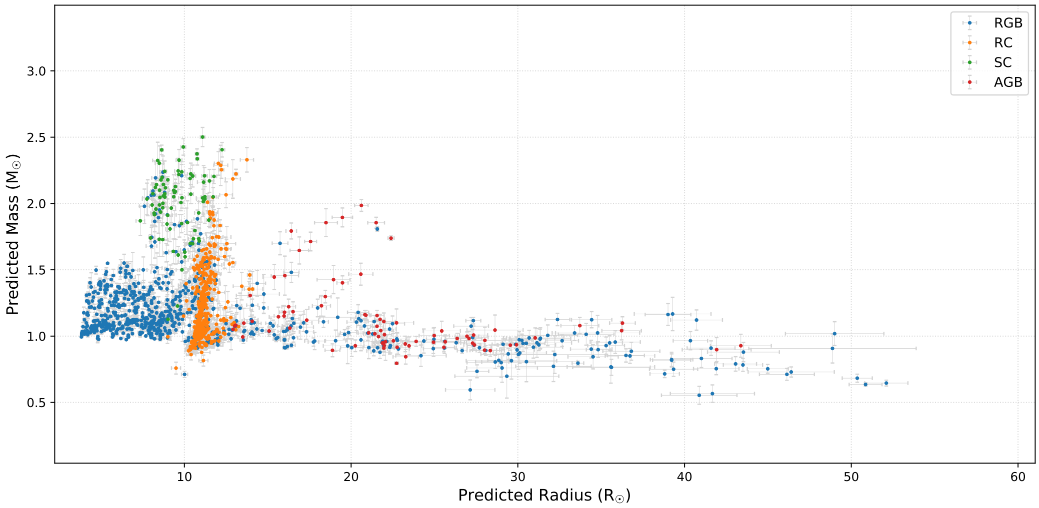



Figure 28: Predicted M-R plot of the test set for model 2.

The resulting $MAPE$ and $RMSE$ of the averaged target values are:

| | $MAPE$(test) | $RMSE$(test) |
|---|---|---|
| Radius | $0.053 \pm 0.001$ | $(1.630 \pm 0.058)\, R_\odot$ |
| Mass | $0.105 \pm 0.001$ | $(0.178 \pm 0.002)\, M_\odot$ |

Even though a different choice of hyper-parameters is used, and the quality of the prediction declines a bit for the validation ($\Delta RMSE_S\,(\text{val.}) = +0.03$) as well as the test set ($\Delta MAPE\,(\text{test}_R) = +0.002$, $\Delta MAPE\,(\text{test}_M) = +0.005$, $\Delta RMSE\,(\text{test}_R) = +0.07\, R_\odot$, $\Delta RMSE\,(\text{test}_M) = +0.05\, M_\odot$), Fig. 28 closely resembles the predicted M-R plot from the previous section (Fig. 25a): The mass range is stronger quenched than the radius range when compared with Fig. 25b. Low-mass RC stars are seemingly worse predicted than in Fig. 25a. The different evolutionary stages are again clearly recognizable and, what is more, intermediate-mass stars seem to be reproduced well for stars with $R_{\text{true}} < 20\, R_\odot$.

Fig. 29 shows again the radio of the predicted to the true value plotted against the true value.

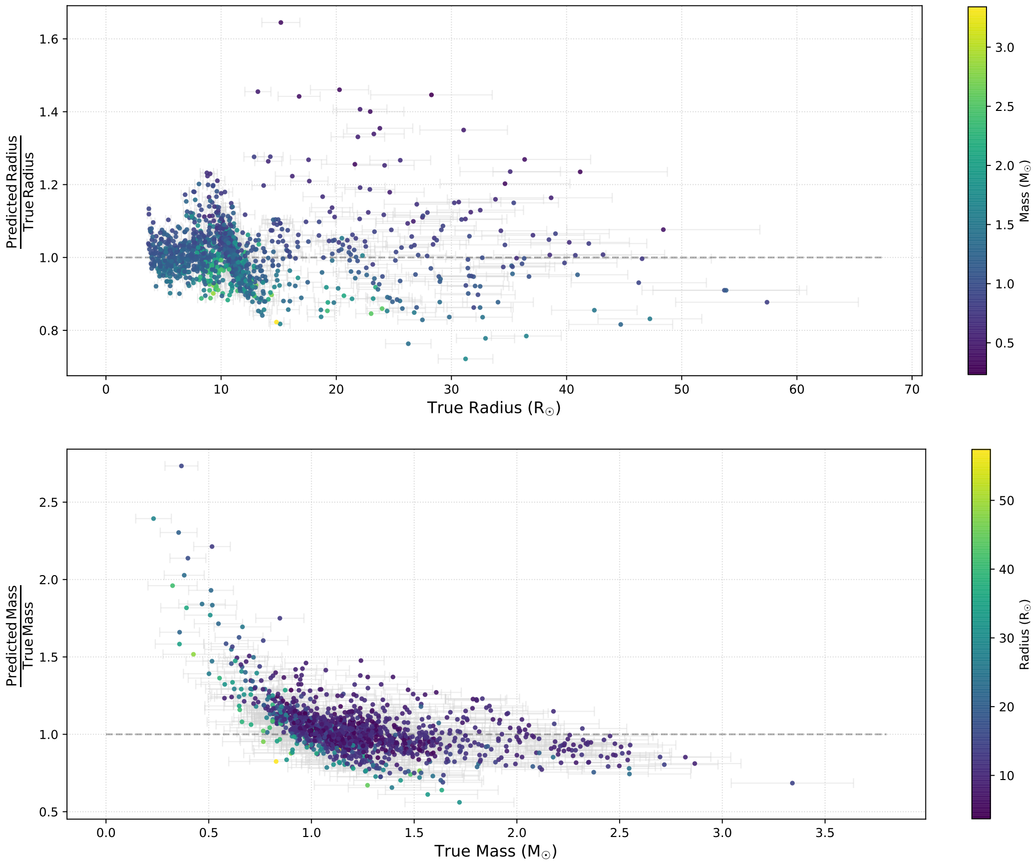



Figure 29: The predicted values obtained by model 2 are compared with the true values. Up: The ratio predicted-to-true radius is plotted against the predicted radius with the true mass color-coded. Down: The ratio predicted-to-true mass is plotted against the predicted mass with the true radius color-coded.

Here we can observe the same trends as in the previous section: Although the model works best for stars with $R_{\text{true}} < 20\,R_\odot$, there seems to be a mass dependency for the predicted radii in this range. This mass dependency can be identified by the radius underprediction of higher-mass stars and the radius overprediction of lower-mass stars.
As before, the predicted masses of stars with $M_{\text{true}} < 1\,M_\odot$ are overestimated whereas the masses of stars with $M_{\text{true}} > 2\,M_\odot$ are underestimated.
Finally it can be said, that even though the effective temperature is not considered as a feature for this model, the results do not change a lot compared to the predicted values in Sec. 8.1. This can be explained by the fact that the construction of the model depends much more on $\nu_{\text{max}}$ than on $T_{\text{eff}}$ (see table at the beginning of Sec. 8.1). Thus, since both the first and the second model depend most on this important feature, they do not differ by much.

## 8.3 Model 3

In this section the results for a model which requires a fifth feature, namely the metallicity, are presented. The [Fe/H] measurements for each star are partly taken from the APOKASC (Pinsonneault et al., 2018) and partly from the LAMOST catalogue (Xiang et al., 2017). For reasons of completeness, Fig. 30 shows the metallicity plotted against all other features, where the first plot displays, as in Fig. 16, the frequency distribution of the metallicity.

According to this distribution, most of the stars have approximately the same [Fe/H] ratio as the Sun (Fig. 30).

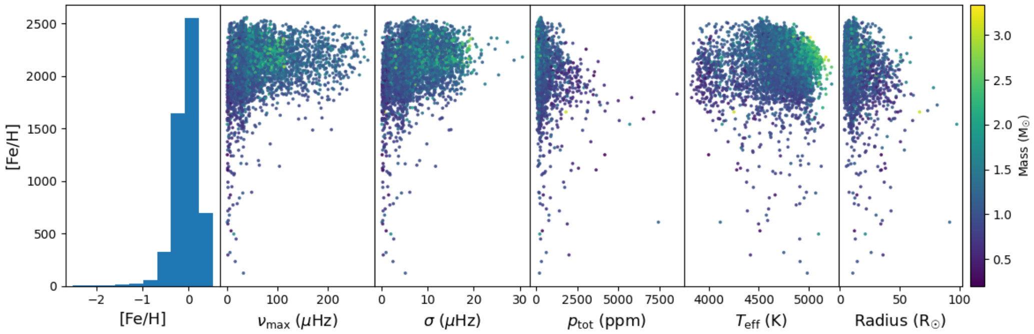



Figure 30: Metallicity plotted against the other four features. The plot on the far left shows the frequency distribution of [Fe/H].

Apparently the metallicity does not correlate as strongly with the radius as $\nu_{\max}$. This leads to the assumption that the feature will play a less important role when it comes to splitting the nodes of the decision trees. However, in order to obtain the feature importances, the model has to be build in the first place and therefore the best hyper-parameter combination must again be derived. This is once more achieved by randomly selecting 100 hyper-parameter combinations using the *RandomizedSearchCV*.

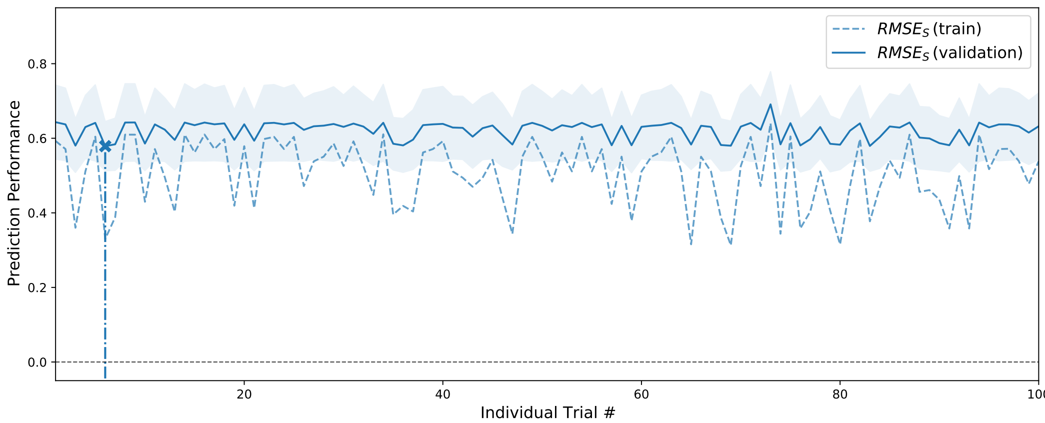



Figure 31: $RMSE_S$ of the validation set after 100 iterations of the *RandomizedSearchCV*. Again the best model with $RMSE_S = 0.58 \pm 0.07$ is highlighted.

The best $RMSE_S$ (val.) $= 0.58 \pm 0.07$ corresponds to the following hyper-parameter combination:

| | |
|---|---|
| $n\_estimators$ | 283 |
| $max\_depth$ | 177 |
| $min\_samples\_split$ | 12 |
| $min\_samples\_leaf$ | 2 |

with the following feature importances:

| | |
|---|---|
| $\nu_{\max}$ | 0.872 |
| $\sigma$ | 0.09 |
| $p_{\rm tot}$ | 0.013 |
| $T_{\rm eff}$ | 0.012 |
| [Fe/H] | 0.013 |

The importance of $\nu_{\text{max}}$ stays nearly the same compared to the importance of 0.873 in Sec. 8.1. The additional feature [Fe/H] is about as important for building the decision tree as $p_{\text{tot}}$ and $T_{\text{eff}}$. This model is again saved and it defines a third and final method in the class *RadiusMassPrediction()*.
After determining the best hyper-parameter combination, another set of mass and radius estimates is created using the code below and the results are shown in Fig. 32.

```
from Radius_Mass_Prediction import *
radius , mass = RadiusMassPrediction(pd.DataFrame(input_parameters)).
                                              Prediction_with_FeH()
```

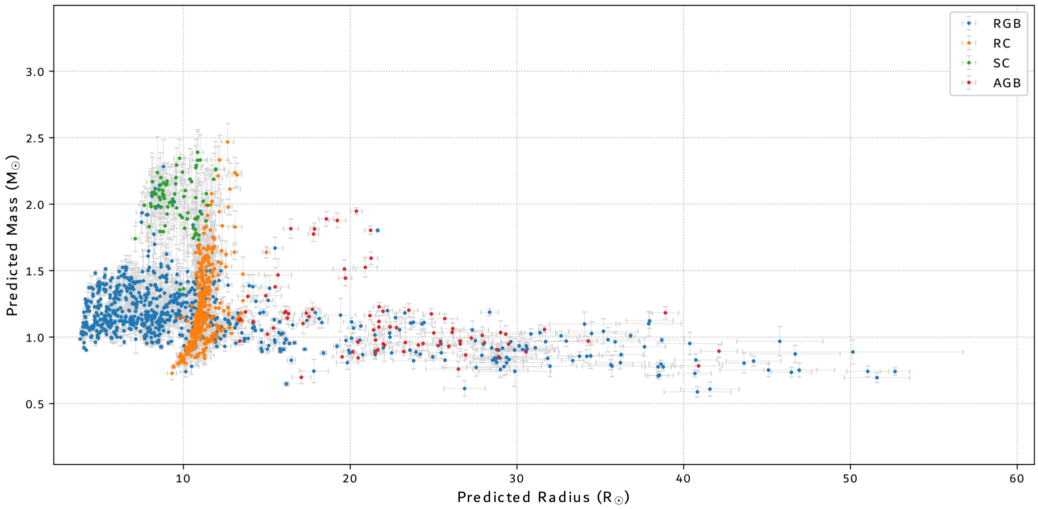



Figure 32: M-R plot for the test sample using model 3.

At first sight, the overall mass distribution is seemingly better reproduced than when using the other models. This assumption is also encouraged by the $MAPE$ and $RMSE$ of the averaged target values:

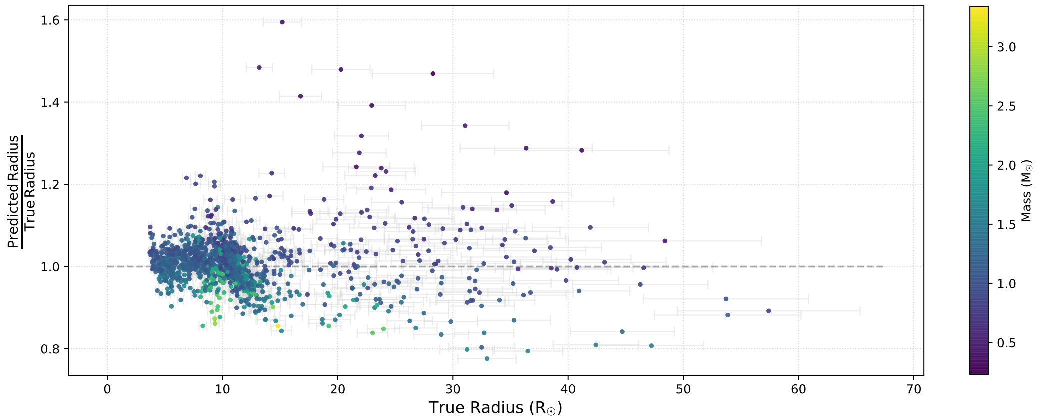
| | $MAPE$ (test) | $RMSE$ (test) |
|---|---|---|
| Radius | $0.0460 \pm 0.0005$ | $(1.503 \pm 0.047)\, R_{\odot}$ |
| Mass | $0.093 \pm 0.001$ | $(0.168 \pm 0.004)\, M_{\odot}$ |

These are the best $MAPE$ and $RMSE$ achieved by any of the three models. It is noteworthy that the $MAPE$ of the radius predictions is better than the $MAPE$ of the mass predictions. As this is also the case for the other two models I can conclude that the radii are always better predicted than the masses by the three models.
The following table shows the differences in the $MAPE$ (test) and $RMSE$ (test) (given for radius and mass separately) between the third and the first model (index $3, 1$) as well as between the third and the second model (index $3, 2$).

| | $\Delta MAPE_{3,1}$ | $\Delta RMSE_{3,1}$ | $\Delta MAPE_{3,2}$ | $\Delta RMSE_{3,2}$ |
|---|---|---|---|---|
| Radius | -0.05 | $-0.057\, R_{\odot}$ | -0.052 | $-0.127\, R_{\odot}$ |
| Mass | -0.007 | $-0.005\, R_{\odot}$ | -0.012 | $-0.055\, R_{\odot}$ |

The difference in prediction performance between the third and the second model is always bigger than between the third and the first, leading to the deduction that the prediction quality varies with the number of features: The best performance is achieved with the highest number of features while the least number of features leads to the worst performance.

The following plots show again the ratio of predicted-to-true against the true value.



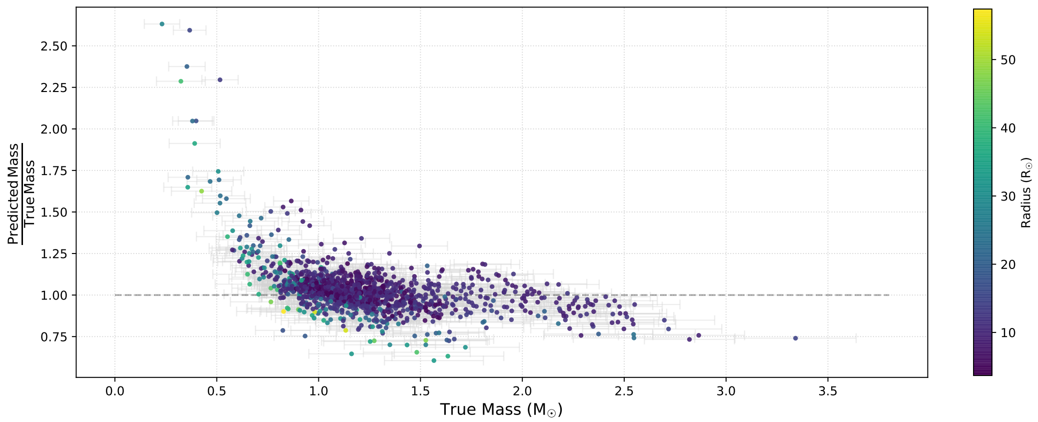




Figure 33: Top: The ratio predicted radius to true radius is plotted against the predicted radius with the true mass color-coded. Bottom: The ratio predicted mass to true mass is plotted against the predicted mass with the true radius color-coded.

The values of the predicted radii do not alter a lot compared to Fig. 26 and Fig. 28: The radii of high-mass stars are still underestimated while the radii of low-mass stars are overestimated. Stars with radii up to $R_{\text{true}} = 20\,R_{\odot}$ are represented best, whereas the overall scatter of the predicted-to-true ratio diminishes in this radius range compared to the previous results.
In the second plot of Fig. 33 it can be seen, that the masses of stars with $R_{\text{true}} \sim 10\,R_{\odot}$ and $M_{\text{true}} < 1\,M_{\odot}$ are slightly better predicted and hence less overpredicted. However, the masses of stars with bigger radii within the same true mass regime are still strongly overestimated.
To summarize, all three models are not very accurate in reproducing the target values of the most as well as the least massive stars in the data set. Therefore, the mass range is always a bit quenched in the predicted M-R plot compared to the true one. Quantitatively speaking, the predicted mean mass of all stars which satisfy $R_{\text{pred}} > 20\,R_{\odot}$ is $\langle M\rangle_{\text{pred}} = 0.96\,M_{\odot}$, where the scatter around that mean is $\pm 0.28\,M_{\odot}$. This value is a bit underestimated compared with the true mean mass of $\langle M\rangle_{\text{true}} = 0.98\,\text{M}_{\odot}$ within the same true radius range. It is noteworthy that the corresponding standard deviation of true mean mass $\pm 0.38\,M_{\odot}$ is significantly larger than $\pm 0.28\,M_{\odot}$. Considering the second and weakest model this scatter even reduces by a factor of 2: $(0.96 \pm 0.19)\,M_{\odot}$.

Some model dependent systematic is also observable in the evolutionary stages for stars with $R < 20\,R_\odot$:

| evolutionary stage | $\langle M\rangle_{\rm true}$ $(M_\odot)$ | $\langle M\rangle_{\rm pred}$ $(M_\odot)$ | | |
|---|---|---|---|---|
| | | model 1 | model 2 | model 3 |
| RGB | $1.16 \pm 0.22$ | $1.20 \pm 0.20$ | $1.20 \pm 0.22$ | $1.20 \pm 0.21$ |
| RC | $1.24 \pm 0.33$ | $1.22 \pm 0.29$ | $1.23 \pm 0.27$ | $1.23 \pm 0.29$ |
| SC | $2.13 \pm 0.31$ | $2.02 \pm 0.21$ | $2.00 \pm 0.25$ | $2.03 \pm 0.19$ |
| AGB | $1.23 \pm 0.44$ | $1.14 \pm 0.29$ | $1.13 \pm 0.27$ | $1.15 \pm 0.32$ |

This table summarizes the predicted averaged masses and the corresponding standard deviations for stars with $R < 20R_\odot$. While all models tend to overestimate the mean mass of RGB stars they underestimate the mean mass of SC stars. The predicted masses of the RC stars are overestimated for low-mass and underestimated for high-mass stars. Last but not least, the models also rather underestimate the masses of AGB stars. Keeping all of this in mind, the predicted mass range is considerably smaller than the true mass range and the most as well as the least massive stars are not as accurately reproduced as the intermediate-mass stars which are indeed predicted well considering Fig. 26, Fig. 29 and Fig. 33.

# 9 Testing the Models

In this section the previously constructed models will be tested using data which are not part of the original KIC data set. All stars within these independent test sets are solar-like oscillators. The choice of model used for predicting the target values depends on the measurements that are at hand for each test set. Firstly, all three models are applied to 16 TIC stars, observed with the *TESS* telescope. Secondly, model 1 predicts the target values of six eclipsing red giant stars, which are compared with dynamically inferred radii and masses. Furthermore, the masses and radii of nine metal-poor stars will be estimated by model 1 and model 3 in Sec. 9.3. As these stars belong to the KIC data set, the predicted target values can be compared with the calculated ones. Finally, in Sec. 9.4 the mean RGB and RC mass of the stars within the open clusters NGC 6791 and NGC 6819 are derived using model 1.

## 9.1 TESS

In the following I will use all models to predict the radii and masses of solar-like oscillators which were observed with the *TESS* space telescope. There are 16 stars from the *TESS Input Cataloge* (TIC) for which I have the measurements of $\nu_{\rm max}$, $\sigma$, $p_{\rm tot}$, $T_{\rm eff}$ and [Fe/H] at disposal. The data – except for $p_{\rm tot}$ measurements – are taken from Aguirre et al. (2020). For additional eight TIC stars are only the values for $\nu_{\rm max}$, $\sigma$ and $p_{\rm tot}$ at hand, enabling the prediction of the target values for altogether 24 TIC stars with model 2. The corresponding results for model 2 are given in Sec. 9.1.1. In Sec. 9.1.2 the performance of model 1 when applied to the TIC data set will be analyzed while Sec. 9.1.3 discusses the predictions of the target values obtained from model 3. This way the prediction performance should increase from Sec. 9.1.1 to Sec. 9.1.3.

### 9.1.1 Model 2

First of all, I display the predicted target values of 24 TIC stars in Fig. 34, where I use model 2 for the predictions.

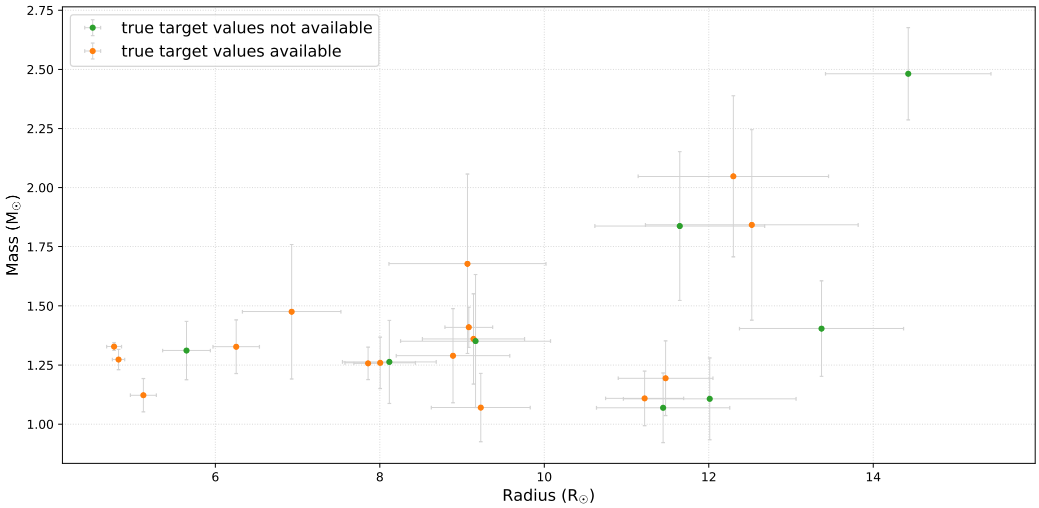



Figure 34: M-R plot for 24 solar-like oscillators, which were observed with the *TESS* telescope. Orange data points indicate that for these stars the true reference target values are also available.

Eight of these stars (marked in green) cannot be compared with any reference target values and are estimated in this work for the first time. In Aguirre et al. (2020) are true reference target values given for the remaining 16 stars (marked in orange). As in the previous sections, this enables a comparison between true and predicted target values (Fig. 35, 35). A description of the results is given in the caption of each figure.

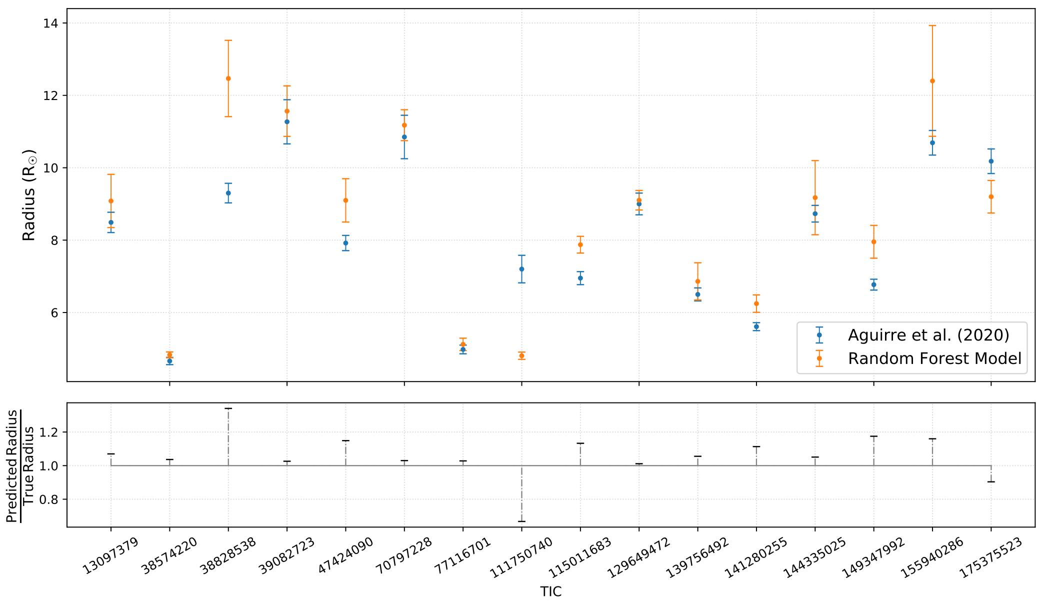



Figure 35: The predicted radii of 16 TIC stars (orange) from model 2 are compared with the true reference values obtained from Aguirre et al. (2020).

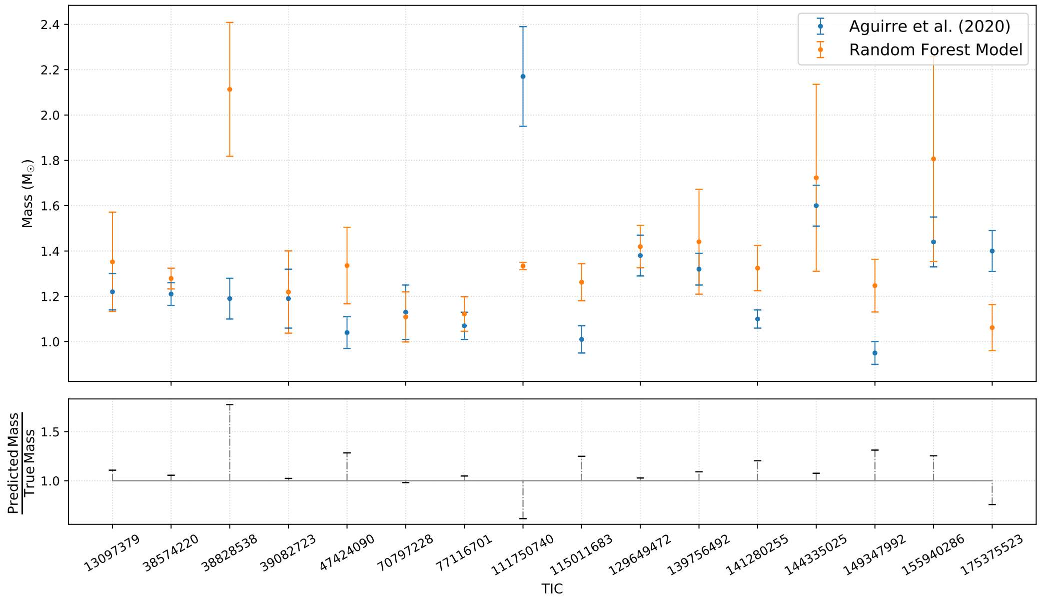



Figure 36: Predicted masses (orange) from model 2 are compared with the true reference values (blue) obtained from Aguirre et al. (2020).

The following table summarizes the achieved $MAPE$ and $RMSE$ of these 16 TIC stars.

| | $MAPE$ | $RMSE$ |
|---|---|---|
| Radius | $0.117 \pm 0.011$ | $(1.416 \pm 0.211)\, R_\odot$ |
| Mass | $0.195 \pm 0.020$ | $(0.426 \pm 0.061)\, M_\odot$ |

Obviously, we cannot compare these results with the $MAPE$ and $RMSE$ values of the KIC data set because the number of samples within the two data sets strongly varies. However, a comparison of the $MAPE$ and $RMSE$ within the same data set is always possible.

### 9.1.2 Model 1

As already mentioned in the introduction of this section, there are $T_{\text{eff}}$ measurements for the 16 TIC star available (Aguirre et al., 2020), which is why model 1 can be used for predicting the target values. In Fig. 37 and Fig. 38 I compare these predicted values with the target values stated in Aguirre et al. (2020). At the end of this subsection is again a table given that summarizes the achieved performance measures of these stars.

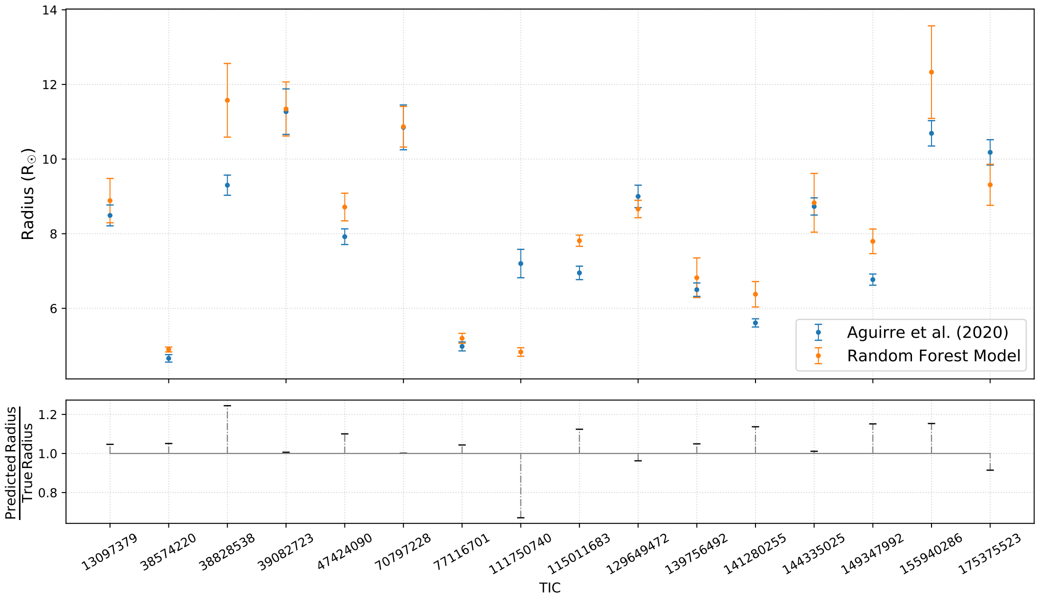



Figure 37: Predicted radii of 16 TIC stars (orange) from model 1 are compared with the radii from Aguirre et al. (2020) (blue). The reproduction of the true target values is slightly better than in Fig. 35, where the absolute average improvement of the predicted-to-true ratios is -0.017. It is noteworthy, that the predictions of TIC 39082723, TIC 707907228 and TIC 144335025 with predicted-to-true ratios of 1.006, 1.002 and 1.011, respectively, agree strongly with the true radii from Aguirre et al. (2020). However, the predicted radius of TIC 38828538 is still overestimated by a factor of 1.25 while the predicted radius of TIC 1175074 is heavily underestimated.

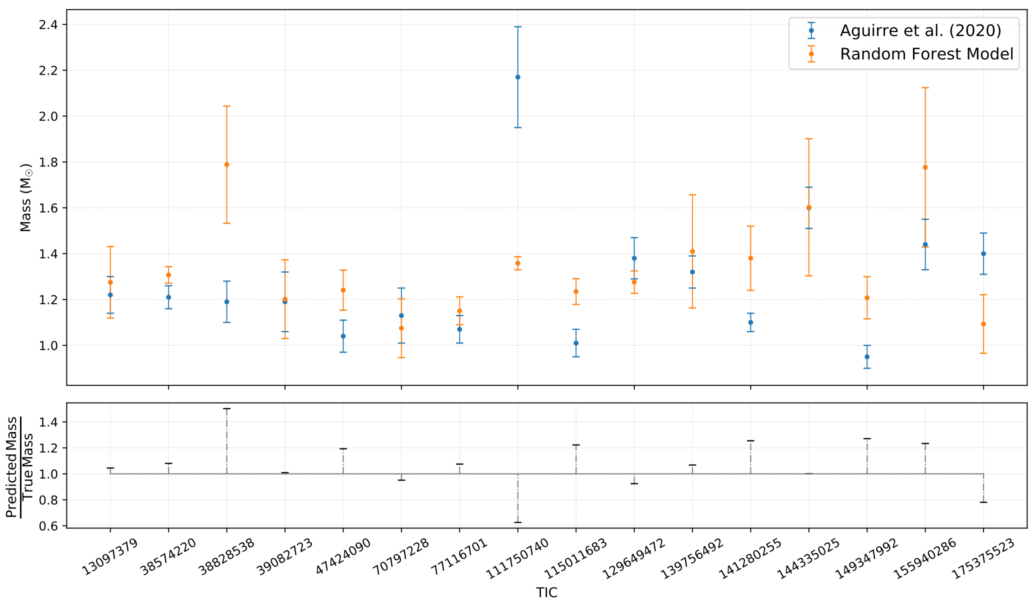



Figure 38: Predicted masses of 16 TIC stars (orange) from model 1 are compared with the masses from Aguirre et al. (2020) (blue). The same trends as in Fig. 37 can be seen: The three stars TIC 39082723, TIC 707907228 and TIC 144335025 reproduce the masses from Aguirre et al. (2020) best, where the mass of the latter agrees with a predicted-to-true ratio of 1.001 strongest with the corresponding true mass.

The $MAPE$ and $RMSE$ of the predicted target values when compared with the true ones, from the paper of Aguirre et al. (2020), are listed in the following table.

| | $MAPE$ | $RMSE$ |
|---|---|---|
| Radius | $0.110 \pm 0.011$ | $(1.198 \pm 0.191)\, R_\odot$ |
| Mass | $0.183 \pm 0.020$ | $(0.351 \pm 0.049)\, M_\odot$ |

The $MAPE$ and $RMSE$ values improve compared with the previous results in Sec. 9.1.1, where the absolute improvement of the radius and mass $MAPE$ is -0.007 and -0.012, respectively, while the absolute improvement of the two distinct $RMSE$ values for radius and mass are $RMSE_{\mathrm{R}} = -0.218\, R_\odot$ and $RMSE_{\mathrm{M}} = -0.075\, M_\odot$, respectively.

### 9.1.3 Model 3

As the radii and masses for most of the stars are overestimated and since the third model performs best in this low-mass regime, it is expected that the results further improve when taking the metallicity into account. The predicted target values are again compared with the the radii and masses stated in Aguirre et al. (2020) (Fig. 39, Fig. 40).

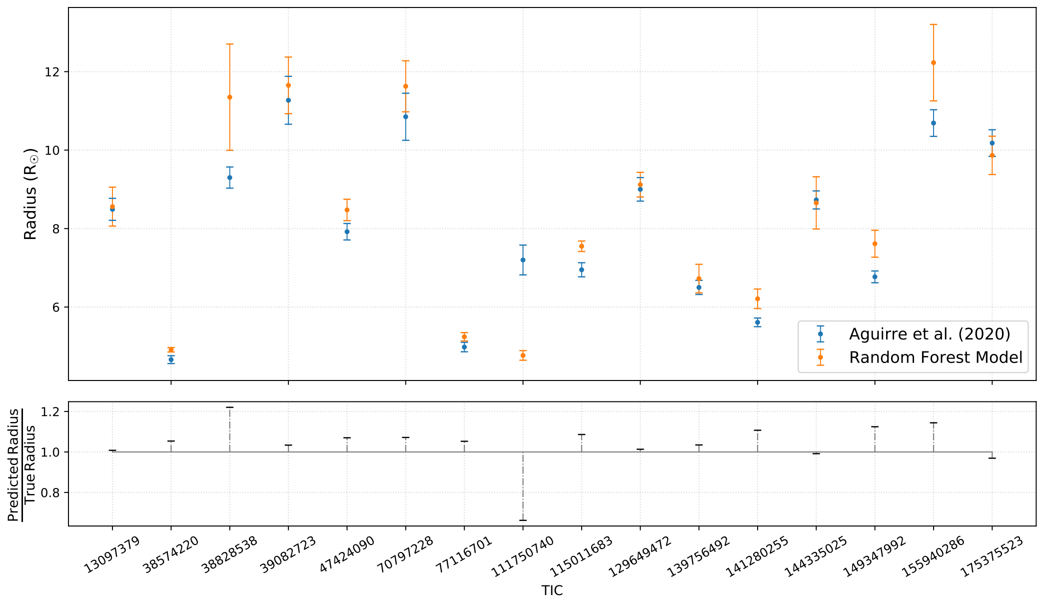



Figure 39: Predicted radii of 16 TIC stars (orange) from model 3 are compared with the radii from Aguirre et al. (2020) (blue). As the response of the radius predictions to the additional feature (i.e. the metallicity) is not that crucial, it is not very surprising that the only noticeable alteration in comparison with Fig. 37 is the diminished predicted-to-true ratio of TIC 70797228 by an absolute factor of 0.07.

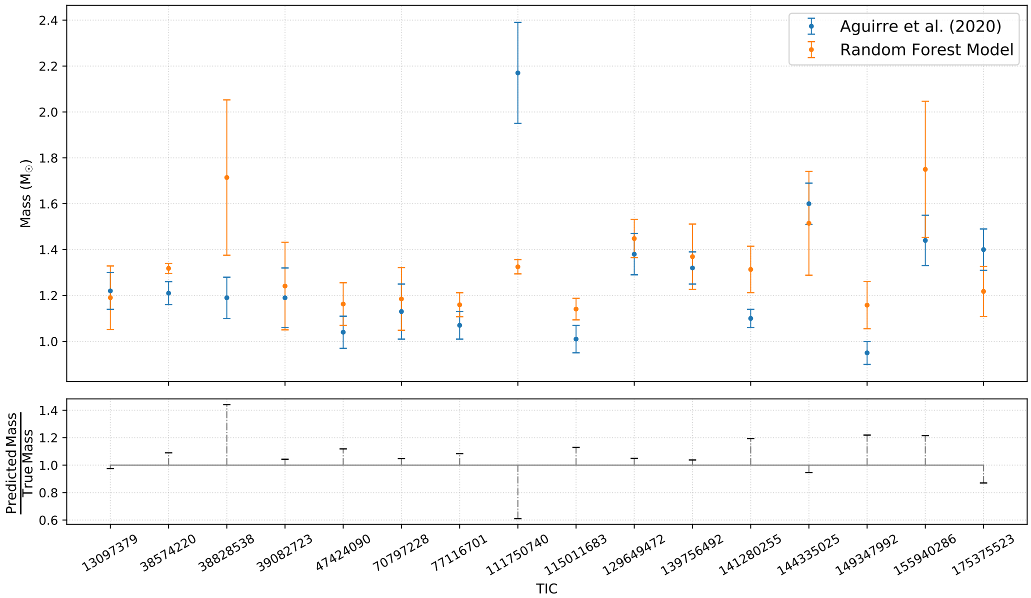



Figure 40: Predicted masses of 16 TIC stars (orange) from model 3 are compared with the masses from Aguirre et al. (2020) (blue). The variations of the mass predictions are also rather small compared to Fig. 38. Nonetheless, some things change: The mass of the star TIC 13097379 is now under- while the masses of TIC 7079228 and TIC 29649472 are overpredicted.

The $MAPE$ and $RMSE$ of the predicted target values when compared with the reference ones, obtained from the paper of Aguirre et al. (2020), are listed in the table below.

| | $MAPE$ | $RMSE$ |
|---|---|---|
| Radius | $0.097 \pm 0.011$ | $(1.097 \pm 0.212)\, R_\odot$ |
| Mass | $0.153 \pm 0.018$ | $(0.316 \pm 0.047)\, M_\odot$ |

As expected, the performance measures further improve when using model 3 instead of model 1 (Sec. 9.1.2), where the $MAPE$ of the radius and mass predictions improve by an absolute factor of -0.013 and -0.03, respectively. The average improvement for the $RMSE$ of the radius predictions is -0.101 $R_\odot$ while it is -0.035 $M_\odot$ for the $RMSE$ of the mass predictions.

## 9.2 Binaries

In this section model 1 is applied to six eclipsing red giant stars, which show solar-like oscillations. As these stars are double-lined spectroscopic binaries that also eclipse, the dynamical masses and radii can be determined, which was done by Gaulme et al. (2016), Themeßl et al. (2018) and Brogaard et al. (2018). Their results are summarized by Kallinger et al. (2018), who used these stars to derive the nonlinear scaling relations (Sec. 4.2.1).

Thus, I have from seismology independently derived target values at disposal, which are compared to the predicted radii and masses from model 1 in Fig. 41.

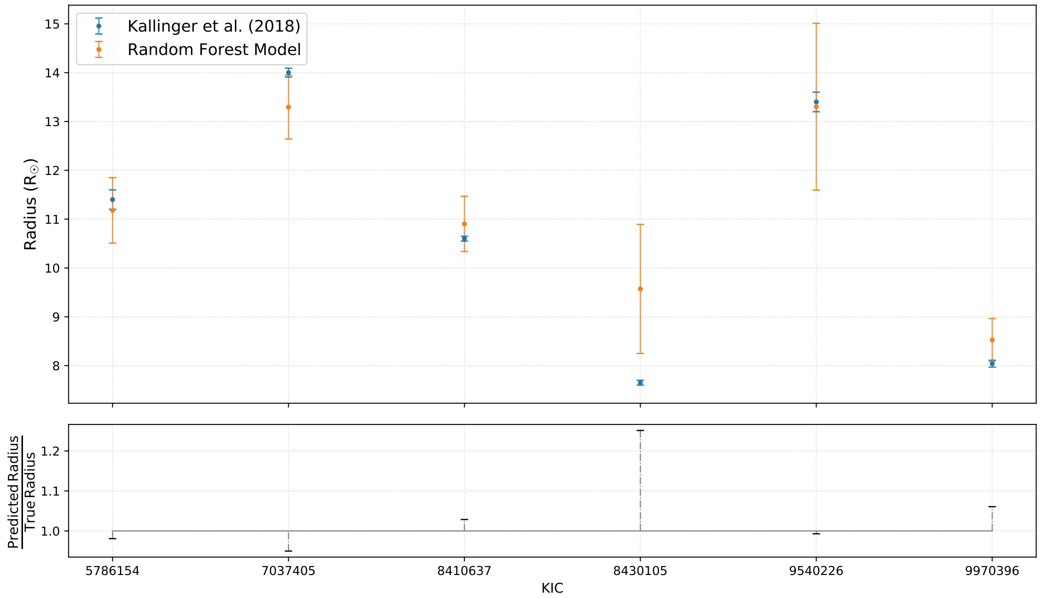



Figure 41: Predicted radii of six eclipsing binaries compared with dynamically inferred radii listed in Kallinger et al. (2018).

The overall trend of the true radii is with a mean relative error of 7% adequately reproduced, only the radius of the star KIC 8430105 is not predicted very accurately. However, the model delivers within $1\,\sigma = \pm 0.89\,R_\odot$ acceptable results for the other stars. Fig. 42 shows the predicted masses and the corresponding true dynamically inferred masses.

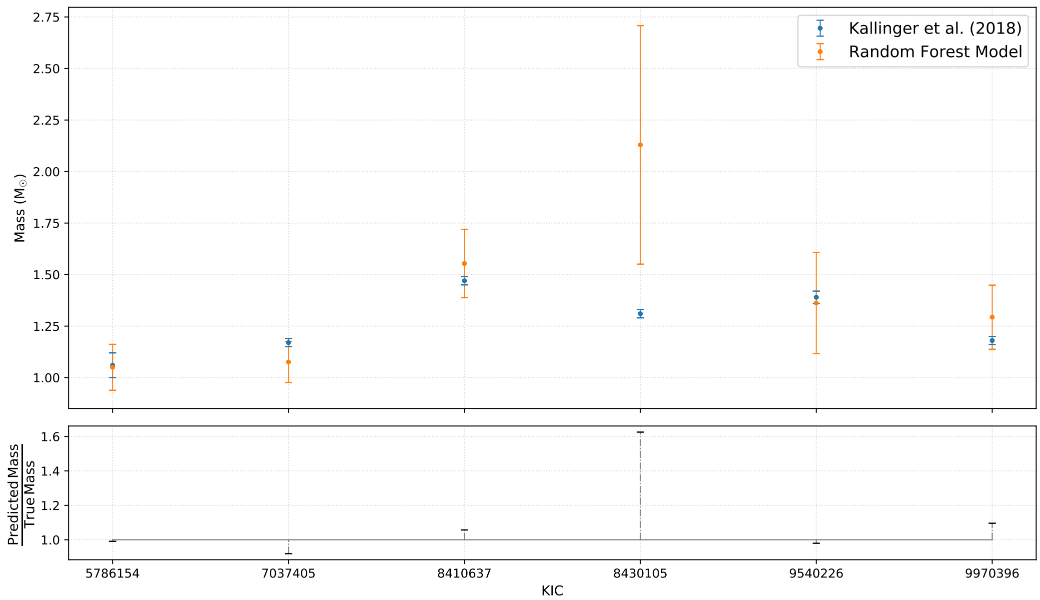



Figure 42: Predicted masses of six eclipsing binaries compared with dynamically inferred masses which are taken from Kallinger et al. (2018).

The masses are also predicted well except for the star KIC 8430105, whose mass is overestimated by a factor of 1.6. The true masses of all the other stars, except KIC 9970396, agree within the error bars of the predicted masses, the mean relative error being 15%.

## 9.3 Metal-poor Stars

Another independent test for our models concerns nine stars that were observed with the *Kepler* telescope. These stars are part of the original KIC test set, however, they have a very low [Fe/H] ratio which suggests a more thorough investigation. Since some authors, such as Epstein et al. (2014), assume that the seismic scaling relations (Eq. (46) and Eq. (47)) also depend on the stellar metallicity, a comparison of the predicted target values using the first and third model seems reasonable. More specifically, Epstein et al. (2014) state that the seismic masses of nine thick-disk and halo stars, for which the masses are thanks to astrophysical priors known to range from about 0.8 to 0.9 $M_{\odot}$, are systematically overestimated by $\langle \Delta M \rangle = 0.17 \pm 0.05\, M_{\odot}$. If the masses of these stars are calculated with the nonlinear scaling relations (derived by Kallinger et al. (2018)) the results fit better into the mass range that is expected for halo and thick-disk stars. Thus, the systematic overestimation could either result from a metallicity effect or from inaccurate linear, solar-calibrated seismic relations. The radius predictions from model 1 are plotted in Fig. 43.

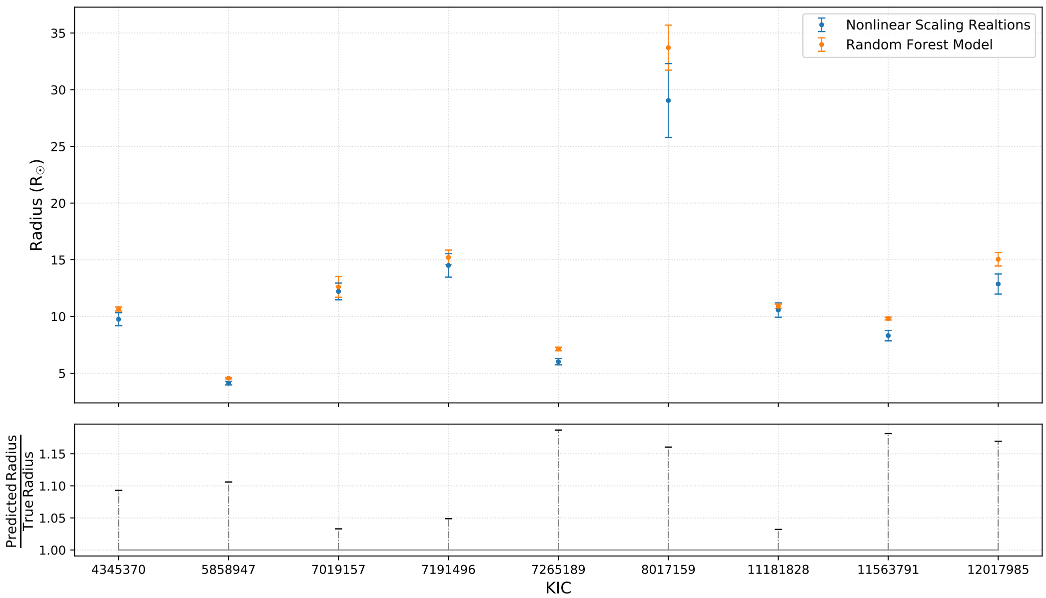



Figure 43: Predicted radii of nine metal-poor stars using model 1.

As one can see, the radii are with a mean relative error of 11% predicted quite well. However, it is striking, that the predicted radii of all stars are overestimated. This can be explained considering the results in Sec. 8, in which I state that the radii of low-mass stars are tendentially overestimated.

Considering Fig. 44, theses metal-poor stars can indeed be considered as low-mass stars ($M < 1\,M_\odot$).

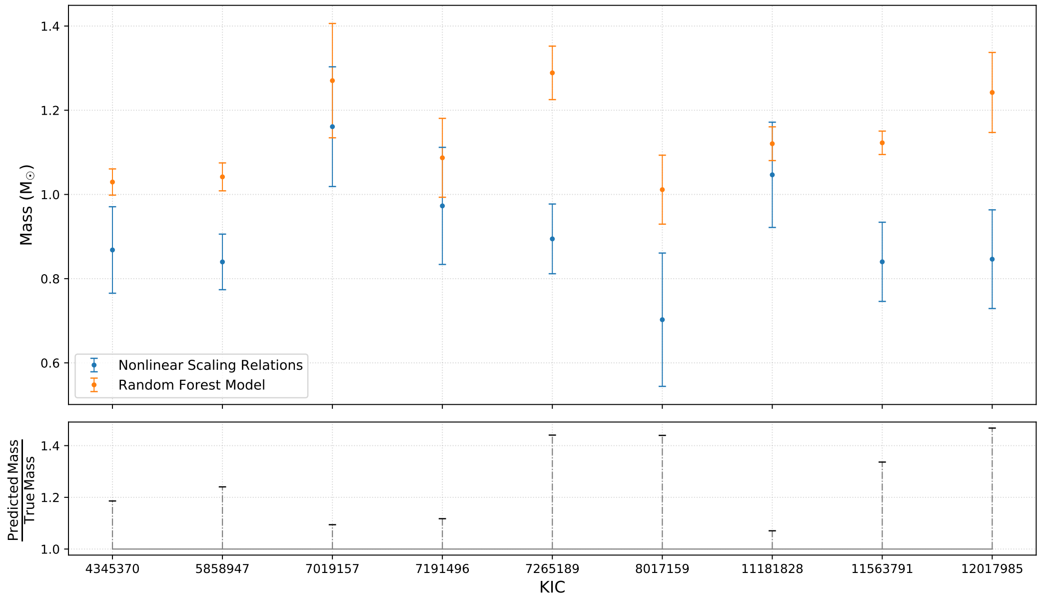



Figure 44: Predicted masses of nine metal-poor stars using model 1.

Again, a systematic overestimation of the mass is evident. It is unclear whether the average discrepancy between the predicted and true mass of $\langle \Delta M \rangle = 0.23 \pm 0.09\,M_\odot$ is a systematic effect, which can be attributed to the metallicity or if this is a model-dependent effect or both: As can be deduced from the analysis of the three models, the least massive stars are overestimated by the model. Only model 3 is able to better reproduce the stars in the lowest mass regime. As there are metallicity measurements available, the target values are also predicted using model 3 (Fig. 45).

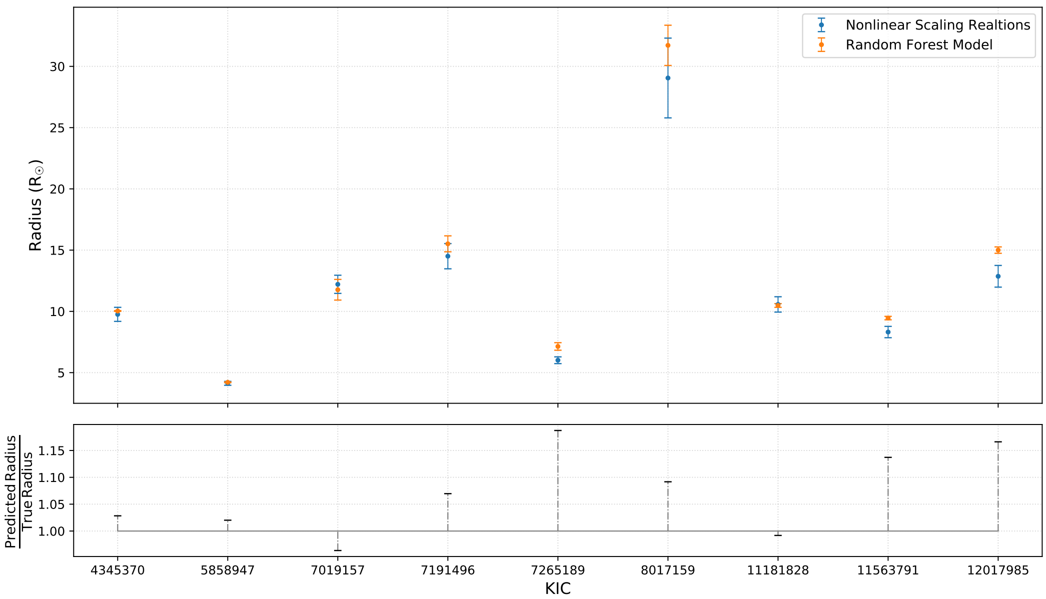



Figure 45: Predicted radii of nine metal-poor stars using the model 3.

The prediction of the radius is not really affected by the additional input parameter. The mass prediction, however, improves for most stars as can be seen from Fig. 46.

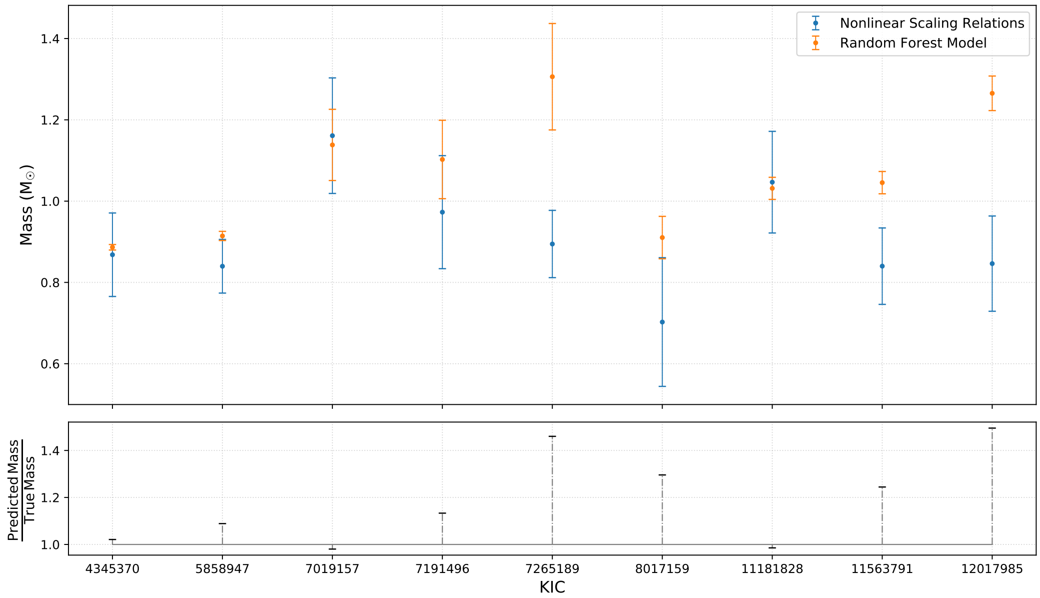



Figure 46: Predicted masses of nine metal-poor stars using model 3.

Four of the nine stars, which are also predicted quite badly using the model 1, are still overestimated by more than 10%. The overall prediction improves though: While the $RMSE$ of the first model is $RMSE_{\mathrm{M}} = (0.260 \pm 0.027)\, M_{\odot}$, the third model achieves $RMSE_{\mathrm{M}} = (0.226 \pm 0.026)\, M_{\odot}$. Model 3 apparently uses the correlation between the mass and the metallicity to enhance its prediction performance ($\Delta RMSE = -0.034\, M_{\odot}$) and thus diminishes the discrepancy between the predicted and true masses. However, as an incorrectness concerning the linear seismic scaling relations cannot definitely be ruled out, the answer to the question if the systematic overestimation in the work of Epstein et al. (2014) really can be ascribed to a metallicity effect is still open for future work.

## 9.4 NGC 6791 and NGC 6819

In this section I test the ML models on stars within the clusters NGC 6791 and NGC 6819. The former is a very old and metal-rich open cluster, whereas the latter is a rather young open cluster with [Fe/H] $\sim 1$ (Kallinger et al., 2018).
There are 51 evolved solar-like oscillators belonging to the cluster NGC 6791 and 39 evolved solar-like oscillators belonging to the cluster NGC 6819 at disposal, shown in Fig. 47.

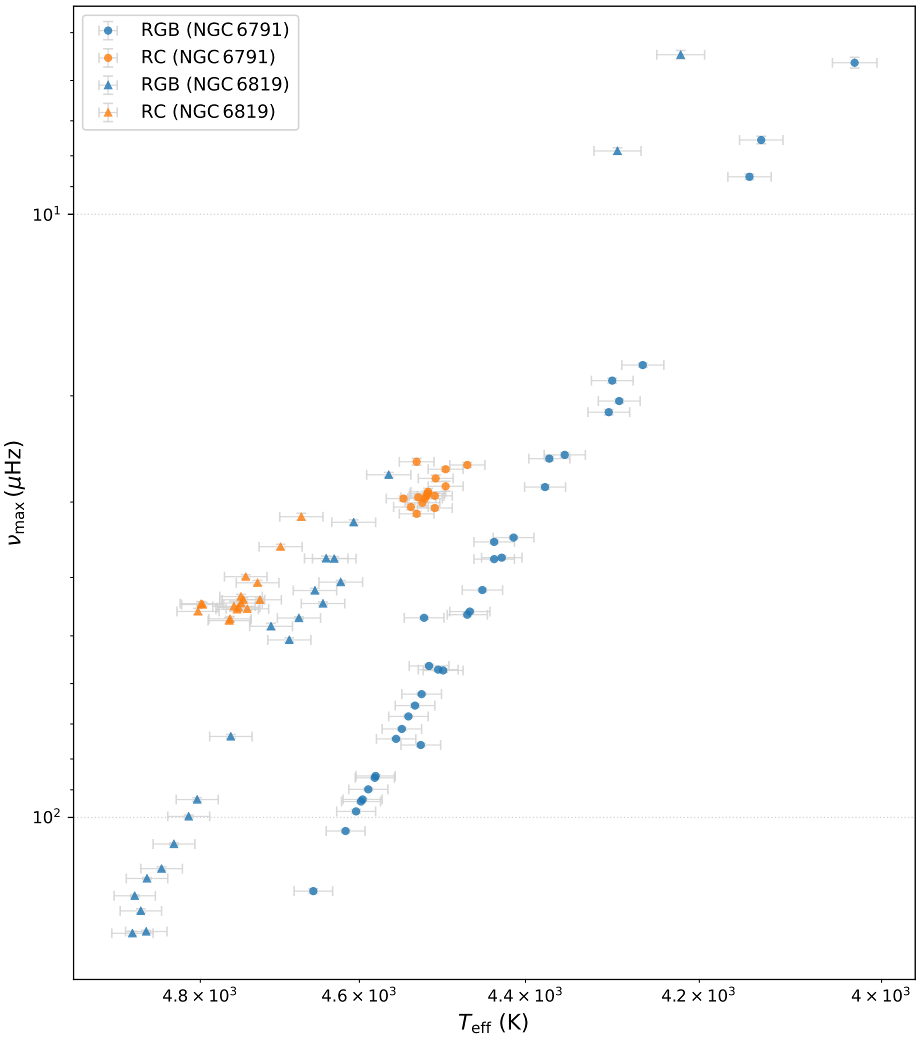



Figure 47: Cluster members of NGC6791 (circles) and NGC 6819 (triangles) plotted in a seismic HR diagram. Orange colors correspond to RC stars, blue ones to RGB stars.

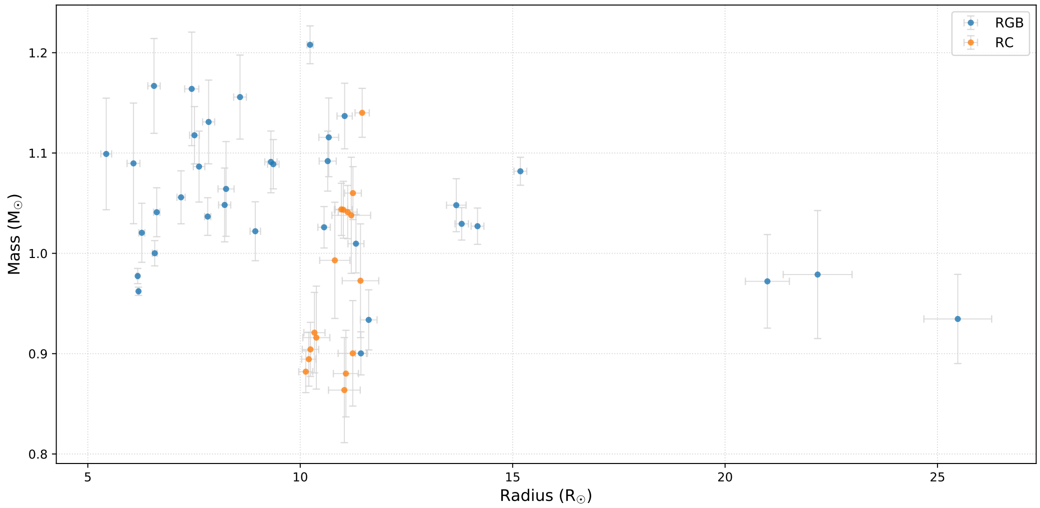
The RC stars are clearly separated from the less evolved RGB stars. The displacement of the stars belonging to different clusters within the diagram presumably corresponds to different RGB masses, confirmed by Tab. 4. As RGB stars evolve quite fast, one can assume that the masses of all RGB stars within the clusters are more or less the same. Kallinger et al. (2018) derived the averaged mass for both the RGB as well as RC stars using the nonlinear scaling relations, which are compared to the predicted averaged masses of our model. Since there are no [Fe/H] measurements available, I use again the model 1 for the prediction.
Tab. 4 shows that the average mass of the RGB and RC stars in NGC 6791 are noticeably underestimated by an absolute value of 0.048 $M_{\odot}$ and 0.08 $M_{\odot}$, respectively. The predicted

averaged RGB mass for the second cluster, NGC 6819, is underestimated by 0.04 $M_\odot$, while the predicted RC mass is overestimated by 0.01 $M_\odot$.

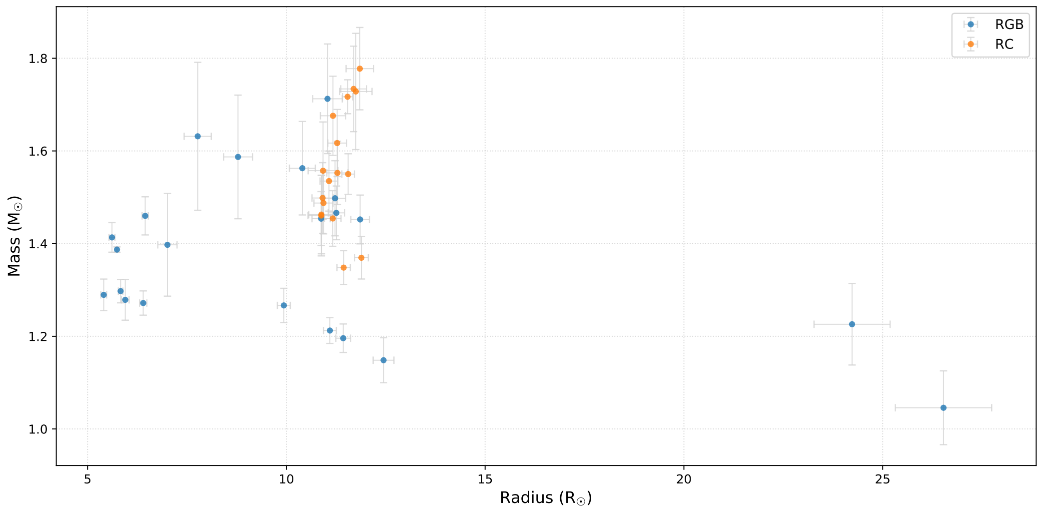
| | NGC 6791 | | | NGC 6819 | |
|---|---|---|---|---|---|
| | $M_{\rm RGB}\,(M_\odot)$ | $M_{\rm RC}\,(M_\odot)$ | $\Delta M\,(M_\odot)$ | $M_{\rm RGB}\,(M_\odot)$ | $M_{\rm RC}\,(M_\odot)$ |
| Kallinger et al. (2018) | $1.10 \pm 0.03$ | $1.02 \pm 0.05$ | $0.08 \pm 0.03$ | $1.45 \pm 0.06$ | $1.54 \pm 0.07$ |
| this work | $1.052 \pm 0.004$ | $0.94 \pm 0.01$ | $0.11 \pm 0.01$ | $1.41 \pm 0.02$ | $1.55 \pm 0.02$ |

Table 4: Predicted averaged $M_{\rm RGB}$ and $M_{\rm RC}$ for stars in the open clusters NGC 6791 and NGC 6819 compared with the mean masses obtained by Kallinger et al. (2018).

For reasons of completeness, the resulting M-R plots for both clusters are shown in Fig. 48a, where the RC stars (orange) located at $\sim$ 11 $R_\odot$ stand clearly out, and Fig. 48b.




(a)




(b)

Figure 48: (a) M-R plot for 51 stars in NGC 6791. (b) M-R plot for 39 stars in NGC 6819.

The RC stars are again confined to a small space within the M-R plot. There seems to be no dramatic mass loss for these stars though as it is the case for the RC stars in NGC 6791. This is also indicated in Tab. 4 where the average mass of the RC stars increases compared with the averaged mass of the RGB stars.

# 10 Discussion

In this final section I want to get to the bottom of the question, why the radii and masses of high-mass stars are rather overestimated while the target values of low-mass stars tend to be underestimated by all three models?

Firstly, I want to review a statement in Borchani et al. (2015) (see Sec. 7.1) which says that a RF model is able to identify the correlations between the target values in the case of multioutput problems. Considering the tests in Sec. 9.2, Sec. 9.3 and Sec. 9.1 for the three models, this statement is rather encouraged than dismissed: When the radius of a star is over- or underestimated, the corresponding mass is also either over- or underestimated and vice versa which could explain the presumable mass dependency for the radius predictions.

Secondly, in order to understand to prediction-process more thoroughly, Fig. 21 shall be once more considered (Fig. 49).

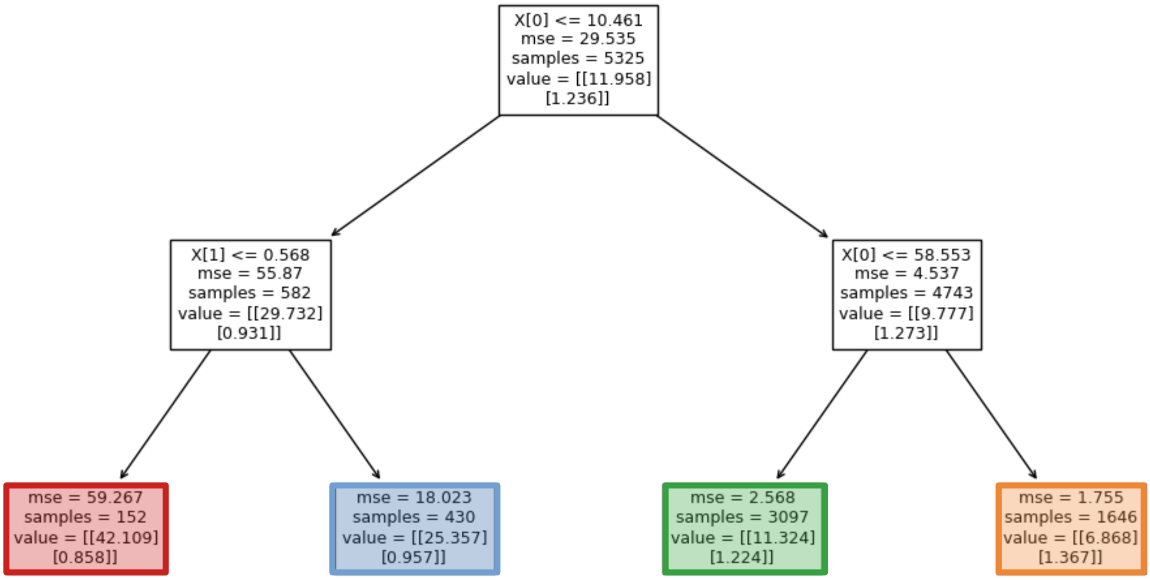



Figure 49: Decision tree with depth two. The colors indicate which stars in Fig. 50 belong to which leaf node.

If the stars in the orange leaf node are considered, which fulfill $\nu_{\mathrm{max}} > 58.533$, it can be seen in Fig. 50 that the predicted averaged mass $M_{\mathrm{pred}} = 1.367\, M_\odot$ is strongly overestimated for low-mass stars while this decision tree model underestimates the true masses of high-mass stars. On the other hand, the predicted radius $R_{\mathrm{pred}} = 16.868\, R_\odot$ is for most low-mass stars below the true radius value while the true radii of higher mass stars are underestimated. The same effect can be seen in all other groups (green, blue and red). Hence this leads to an over- and underestimation of the target values for low-mass and high-mass stars, respectively. Of course, this is an extreme example and the decision tree, which is the underlying estimator for the random forest algorithm, obviously underfits the data, however, a deeper decision tree still makes the same error but in a much less severe extent. Furthermore, the reason that the mass prediction is always worse than the radius prediction is that the features have a much higher dependency on the radius than on the mass and thus the model splits the data rather vertically (i.e. into smaller radius intervals) than horizontally within a M-R plot (Fig. 50).

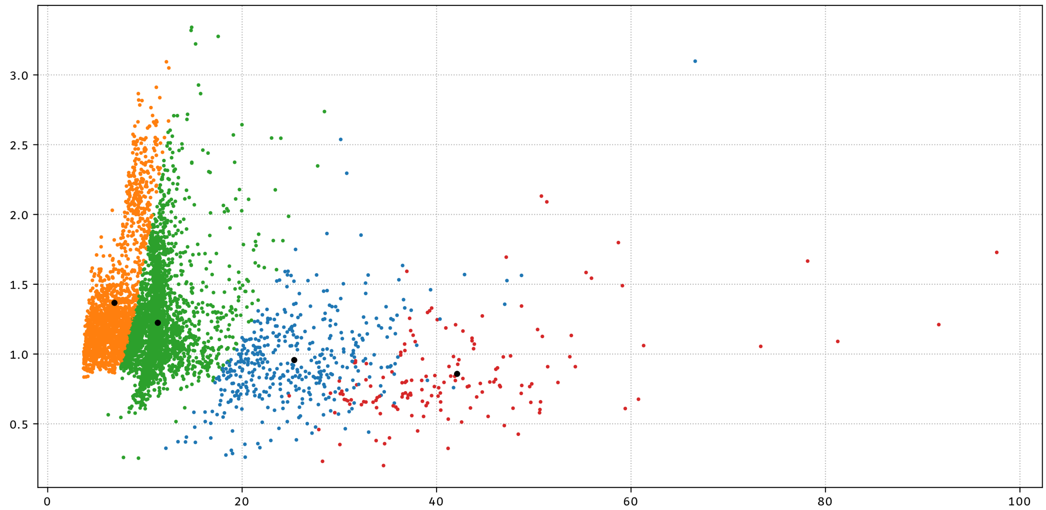



Figure 50: M-R plot of the training set, where the colors correspond to the data samples within the leaf nodes in Fig. 49. The four black dots in the figure indicate the predicted masses and radii for each leaf node (given as 'value' in square brackets in each leaf node in in Fig. 49).

In conclusion, the three models work best for intermediate-mass stars where the third model with $RMSE_\mathrm{R} = (1.503 \pm 0.047)\,R_\odot$ and $RMSE_\mathrm{M} = (0.168 \pm 0.004)\,M_\odot$ achieves the best results. All three models fail to reproduce the least as well as most massive stars. Future missions as well as the ongoing space mission *TESS* will likely lead to a higher availability of low- and high-mass solar-like oscillators. By adding these stars to the training set, the RF models can continually be improved. The results of the three models show, that the radius predictions are always better than the mass predictions. Furthermore there is a mass dependency for the radius predictions observable, such that the radii of low-mass stars are overestimated while the radii of high-mass stars are underestimated. The models perform well on the test data sets, including binaries as well as metal-poor and open cluster stars. As the metal-poor stars are low-mass stars, the mass predictions considerably improve ($\Delta RMSE_\mathrm{M} = -0.034\,M_\odot$) when using the third model, which also takes the metallicity as input. The RGB masses $M_\mathrm{RGB} = (1.051 \pm 0.004)\,M_\odot$ and $M_\mathrm{RGB} = (1.41 \pm 0.02)\,R_\odot$ of the open clusters NGC 6791 and NGC 6819, respectively, underestimate the RGB masses derived by Kallinger et al. (2018) absolutely by $\sim 0.04\,M_\odot$. As the TIC data set is most independent from the training set, it is reassuring that the radius and mass predictions achieve with $RMSE_\mathrm{R} = (1.097 \pm 0.212)\,R_\odot$ and $RMSE_\mathrm{M} = (0.316 \pm 0.047)\,M_\odot$ (when compared with the target values from Aguirre et al. (2020) and when using model 3) acceptable results.

With this work I hope to contribute to a facilitated estimation of fundamental parameters of red giant stars, even though there are not many measurements at hand. This often coincides with short observation periods such as the 27-day observations from *TESS*.

# List of Figures

# References


Aerts, C., J. Christensen-Dalsgaard, and D. W. Kurtz (2010). *Asteroseismology.*

Aguirre, V. S., D. Stello, A. Stokholm, J. R. Mosumgaard, W. H. Ball, S. Basu, D. Bossini, L. Bugnet, D. Buzasi, T. L. Campante, L. Carboneau, W. J. Chaplin, E. Corsaro, G. R. Davies, Y. Elsworth, R. A. García, P. Gaulme, O. J. Hall, R. Handberg, M. Hon, T. Kallinger, L. Kang, M. N. Lund, S. Mathur, A. Mints, B. Mosser, Z. Ç. Orhan, T. S. Rodrigues, M. Vrard, M. Yıldız, J. C. Zinn, S. Örtel, P. G. Beck, K. J. Bell, Z. Guo, C. Jiang, J. S. Kuszlewicz, C. A. Kuehn, T. Li, M. S. Lundkvist, M. Pinsonneault, J. Tayar, M. S. Cunha, S. Hekker, D. Huber, A. Miglio, M. J. P. F. G. Monteiro, D. Slumstrup, M. L. Winther, G. Angelou, O. Benomar, A. Bódi, B. L. D. Moura, S. Deheuvels, A. Derekas, M. P. D. Mauro, M.-A. Dupret, A. Jiménez, Y. Lebreton, J. Matthews, N. Nardetto, J. D. do Nascimento, F. Pereira, L. F. R. Díaz, A. M. Serenelli, E. Spitoni, E. Stonkutė, J. C. Suárez, R. Szabó, V. V. Eylen, R. Ventura, K. Verma, A. Weiss, T. Wu, T. Barclay, J. Christensen-Dalsgaard, J. M. Jenkins, H. Kjeldsen, G. R. Ricker, S. Seager, and R. Vanderspek (2020, jan). Detection and characterization of oscillating red giants: First results from the TESS satellite. *The Astrophysical Journal 889*(2), L34.

Alpaydin, E. (2004). *Introduction to machine learning.* Adaptive computation and machine learning series. Cambridge, Mass. [u.a.]: MIT-Press.

Badillo, S., B. Banfai, F. Birzele, I. I. Davydov, L. Hutchinson, T. Kam-Thong, J. Siebourg-Polster, B. Steiert, and J. D. Zhang (2020). An introduction to machine learning. *Clinical Pharmacology & Therapeutics 107*(4), 871–885.

Baglin, A. (2003, January). COROT: A minisat for pionnier science, asteroseismology and planets finding. *Advances in Space Research 31*(2), 345–349.

Balmforth, N. J. (1992, April). Solar pulsational stability - I. Pulsation-mode thermodynamics. MNRAS *255*, 603–649.

Bedding, T. R. and H. Kjeldsen (2003, January). Solar-like Oscillations. PASA *20*(2), 203–212.

Belkacem, K., M. J. Goupil, M. A. Dupret, R. Samadi, F. Baudin, A. Noels, and B. Mosser (2011, June). The underlying physical meaning of the $\nu_{max}$ - $\nu_c$ relation. A&A *530*, A142.

Borchani, H., G. Varando, C. Bielza, and P. Larrañaga (2015). A survey on multi-output regression. *Wiley interdisciplinary reviews. Data mining and knowledge discovery 5*(5), 216–233.

Borucki, W. J., D. Koch, G. Basri, N. Batalha, T. Brown, D. Caldwell, J. Caldwell, J. Christensen-Dalsgaard, W. D. Cochran, E. DeVore, E. W. Dunham, A. K. Dupree, T. N. Gautier, J. C. Geary, R. Gilliland, A. Gould, S. B. Howell, J. M. Jenkins, Y. Kondo, D. W. Latham, G. W. Marcy, S. Meibom, H. Kjeldsen, J. J. Lissauer, D. G. Monet, D. Morrison, D. Sasselov, J. Tarter, A. Boss, D. Brownlee, T. Owen, D. Buzasi, D. Charbonneau, L. Doyle, J. Fortney, E. B. Ford, M. J. Holman, S. Seager, J. H. Steffen, W. F. Welsh, J. Rowe, H. Anderson, L. Buchhave, D. Ciardi, L. Walkowicz, W. Sherry, E. Horch, H. Isaacson, M. E. Everett, D. Fischer, G. Torres, J. A. Johnson, M. Endl, P. MacQueen, S. T. Bryson, J. Dotson, M. Haas, J. Kolodziejczak, J. Van Cleve, H. Chandrasekaran, J. D. Twicken, E. V. Quintana, B. D. Clarke, C. Allen, J. Li, H. Wu, P. Tenenbaum, E. Verner, F. Bruhweiler, J. Barnes, and A. Prsa (2010, February). Kepler Planet-Detection Mission: Introduction and First Results. *Science 327*(5968), 977.

Brogaard, K., C. J. Hansen, A. Miglio, D. Slumstrup, S. Frandsen, J. Jessen-Hansen, M. N. Lund, D. Bossini, A. Thygesen, G. R. Davies, W. J. Chaplin, T. Arentoft, H. Bruntt, F. Grundahl, and R. Handberg (2018, May). Establishing the accuracy of asteroseismic mass and radius estimates of giant stars - I. Three eclipsing systems at [Fe/H] $\sim$ -0.3 and the need for a large high-precision sample. MNRAS *476*(3), 3729–3743.

Brown, T. M., R. L. Gilliland, R. W. Noyes, and L. W. Ramsey (1991, February). Detection of Possible p-Mode Oscillations on Procyon. ApJ *368*, 599.

Buitinck, L., G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux (2013). API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122.

Chaplin, W. and A. Miglio (2013, 03). Asteroseismology of solar-type and red-giant stars. *Annual Review of Astronomy and Astrophysics 51.*

Christensen-Dalsgaard, J. (2002, November). Helioseismology. *Reviews of Modern Physics 74*(4), 1073–1129.

Christensen-Dalsgaard, J. (2004, April). Physics of solar-like oscillations. Sol. Phys. *220*(2), 137–168.

Christensen-Dalsgaard, J. and M. J. Thompson (2011, August). Stellar hydrodynamics caught in the act: Asteroseismology with CoRoT and Kepler. In N. H. Brummell, A. S. Brun, M. S. Miesch, and Y. Ponty (Eds.), *Astrophysical Dynamics: From Stars to Galaxies*, Volume 271 of *IAU Symposium*, pp. 32–61.

Claverie, A., G. R. Isaak, C. P. McLeod, H. B. van der Raay, and T. R. Cortes (1979, December). Solar structure from global studies of the 5-minute oscillation. Nature *282*, 591–594.

Cowling, T. G. (1941, January). The non-radial oscillations of polytropic stars. MNRAS *101*, 367.

Cunha, M. S., C. Aerts, J. Christensen-Dalsgaard, A. Baglin, L. Bigot, T. M. Brown, C. Catala, O. L. Creevey, A. Domiciano de Souza, P. Eggenberger, P. J. V. Garcia, F. Grundahl, P. Kervella, D. W. Kurtz, P. Mathias, A. Miglio, M. J. P. F. G. Monteiro, G. Perrin, F. P. Pijpers, D. Pourbaix, A. Quirrenbach, K. Rousselet-Perraut, T. C. Teixeira, F. Thévenin, and M. J. Thompson (2007, November). Asteroseismology and interferometry. A&A Rev. *14*(3-4), 217–360.

Cunha, M. S., P. P. Avelino, and W. J. Chaplin (2020, September). From Solar-like to Mira stars: a unifying description of stellar pulsators in the presence of stochastic noise. MNRAS.

De Ridder, J., C. Barban, F. Baudin, F. Carrier, A. P. Hatzes, S. Hekker, T. Kallinger, W. W. Weiss, A. Baglin, M. Auvergne, R. Samadi, P. Barge, and M. Deleuil (2009, May). Non-radial oscillation modes with long lifetimes in giant stars. Nature *459*(7245), 398–400.

Deubner, F. L. (1975, November). Observations of low wavenumber nonradial eigenmodes of the sun. A&A *44*(2), 371–375.

Di Mauro, M. P. (2012, 12). Helioseismology: A fantastic tool to probe the interior of the sun. *Lecture Notes in Physics*.

Epstein, C. R., Y. P. Elsworth, J. A. Johnson, M. Shetrone, B. Mosser, S. Hekker, J. Tayar, P. Harding, M. Pinsonneault, V. Silva Aguirre, S. Basu, T. C. Beers, D. Bizyaev, T. R. Bedding, W. J. Chaplin, P. M. Frinchaboy, R. A. García, A. E. García Pérez, F. R. Hearty, D. Huber, I. I. Ivans, S. R. Majewski, S. Mathur, D. Nidever, A. Serenelli, R. P. Schiavon, D. P. Schneider, R. Schönrich, J. S. Sobeck, K. G. Stassun, D. Stello, and G. Zasowski (2014, April). Testing the Asteroseismic Mass Scale Using Metal-poor Stars Characterized with APOGEE and Kepler. ApJ *785*(2), L28.

Frandsen, S., F. Carrier, C. Aerts, D. Stello, T. Maas, M. Burnet, H. Bruntt, T. C. Teixeira, J. R. de Medeiros, F. Bouchy, H. Kjeldsen, F. Pijpers, and J. Christensen-Dalsgaard (2002, October). Detection of Solar-like oscillations in the G7 giant star $\xi$ Hya. A&A *394*, L5–L8.

Frazier, E. N. (1968, January). A Spatio-Temporal Analysis of Velocity Fields in the Solar Photosphere. ZAp *68*, 345.

Garreta, R. and G. Moncecchi (2013). *Learning Scikit-learn : Machine Learning in Python: Experience the Benefits of Machine Learning Techniques by Applying Them to Real-world Problems Using Python and the Open Source Scikit-learn Library.* Community Experience Distilled. Packt Publishing.

Gaulme, P., J. McKeever, J. Jackiewicz, M. L. Rawls, E. Corsaro, B. Mosser, J. Southworth, S. Mahadevan, C. Bender, and R. Deshpande (2016, December). Testing the Asteroseismic Scaling Relations for Red Giants with Eclipsing Binaries Observed by Kepler. ApJ *832*(2), 121.

Houdek, G. (2005). Lecture course in Theory of Pulsating Stars.

Houdek, G., N. J. Balmforth, J. Christensen-Dalsgaard, and D. O. Gough (1999, November). Amplitudes of stochastically excited oscillations in main-sequence stars. A&A *351*, 582–596.

Howell, S. B., C. Sobeck, M. Haas, M. Still, T. Barclay, F. Mullally, J. Troeltzsch, S. Aigrain, S. T. Bryson, D. Caldwell, W. J. Chaplin, W. D. Cochran, D. Huber, G. W. Marcy, A. Miglio, J. R. Najita, M. Smith, J. D. Twicken, and J. J. Fortney (2014, April). The K2 Mission: Characterization and Early Results. PASP *126*(938), 398.

Huber, D. and K. Zwintz (2020, January). Solar-Like Oscillations: Lessons Learned & First Results from TESS. In C. Neiner, W. W. Weiss, D. Baade, R. E. Griffin, C. C. Lovekin, and A. F. J. Moffat (Eds.), *Proceedings of the conference Stars and their Variability Observed from Space*, pp. 457–463.

Kallinger, T. (2019, June). Release note: Massive peak bagging of red giants in the Kepler field. *arXiv e-prints*, arXiv:1906.09428.

Kallinger, T., P. G. Beck, D. Stello, and R. A. Garcia (2018, August). Non-linear seismic scaling relations. A&A *616*, A104.

Kallinger, T., J. De Ridder, S. Hekker, S. Mathur, B. Mosser, M. Gruberbauer, R. A. García, C. Karoff, and J. Ballot (2014, October). The connection between stellar granulation and oscillation as seen by the Kepler mission. A&A *570*, A41.

Kallinger, T., S. Hekker, B. Mosser, J. De Ridder, T. R. Bedding, Y. P. Elsworth, M. Gruberbauer, D. B. Guenther, D. Stello, S. Basu, R. A. García, W. J. Chaplin, F. Mullally, M. Still, and S. E. Thompson (2012, May). Evolutionary influences on the structure of red-giant acoustic oscillation spectra from 600d of Kepler observations. A&A *541*, A51.

Kippenhahn, R., A. Weigert, and A. Weiss (2012). *Adiabatic Spherical Pulsations*, pp. 519–528. Berlin, Heidelberg: Springer Berlin Heidelberg.

Koch, D. G., W. Borucki, E. Dunham, J. Geary, R. Gilliland, J. Jenkins, D. Latham, E. Bachtell, D. Berry, W. Deininger, R. Duren, T. N. Gautier, L. Gillis, D. Mayer, C. D. Miller, D. Shafer, C. K. Sobeck, C. Stewart, and M. Weiss (2004, October). Overview and status of the Kepler Mission. In J. C. Mather (Ed.), *Optical, Infrared, and Millimeter Space Telescopes*, Volume 5487 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp. 1491–1500.

Leighton, R. B., R. W. Noyes, and G. W. Simon (1962, March). Velocity Fields in the Solar Atmosphere. I. Preliminary Report. ApJ *135*, 474.

Majewski, S. R., R. P. Schiavon, P. M. Frinchaboy, C. Allende Prieto, R. Barkhouser, D. Bizyaev, B. Blank, S. Brunner, A. Burton, R. Carrera, S. D. Chojnowski, K. Cunha, C. Epstein, G. Fitzgerald, A. E. García Pérez, F. R. Hearty, C. Henderson, J. A. Holtzman, J. A. Johnson, C. R. Lam, J. E. Lawler, P. Maseman, S. Mészáros, M. Nelson, D. C. Nguyen, D. L. Nidever, M. Pinsonneault, M. Shetrone, S. Smee, V. V. Smith, T. Stolberg, M. F. Skrutskie, E. Walker, J. C. Wilson, G. Zasowski, F. Anders, S. Basu, S. Beland, M. R. Blanton, J. Bovy, J. R. Brownstein, J. Carlberg, W. Chaplin, C. Chiappini, D. J. Eisenstein, Y. Elsworth, D. Feuillet, S. W. Fleming, J. Galbraith-Frew, R. A. García, D. A. García-Hernández, B. A. Gillespie, L. Girardi, J. E. Gunn, S. Hasselquist, M. R. Hayden, S. Hekker, I. Ivans, K. Kinemuchi, M. Klaene, S. Mahadevan, S. Mathur, B. Mosser, D. Muna, J. A. Munn, R. C. Nichol, R. W. O'Connell, J. K. Parejko, A. C. Robin, H. Rocha-Pinto, M. Schultheis, A. M. Serenelli, N. Shane, V. Silva Aguirre, J. S. Sobeck, B. Thompson, N. W. Troup, D. H. Weinberg, and O. Zamora (2017, September). The Apache Point Observatory Galactic Evolution Experiment (APOGEE). AJ *154*(3), 94.

Metcalfe, T. S., M. J. P. F. G. Monteiro, M. J. Thompson, J. Molenda-Żakowicz, T. Appourchaux, W. J. Chaplin, G. Doğan, P. Eggenberger, T. R. Bedding, H. Bruntt, O. L. Creevey, P. O. Quirion, D. Stello, A. Bonanno, V. Silva Aguirre, S. Basu, L. Esch, N. Gai, M. P. Di Mauro, A. G. Kosovichev, I. N. Kitiashvili, J. C. Suárez, A. Moya, L. Piau, R. A. García, J. P. Marques, A. Frasca, K. Biazzo, S. G. Sousa, S. Dreizler, M. Bazot, C. Karoff, S. Frandsen, P. A. Wilson, T. M. Brown, J. Christensen-Dalsgaard, R. L. Gilliland, H. Kjeldsen, T. L. Campante, S. T. Fletcher, R. Hand berg, C. Régulo, D. Salabert, J. Schou, G. A. Verner, J. Ballot, A. M. Broomhall, Y. Elsworth, S. Hekker, D. Huber, S. Mathur, R. New, I. W. Roxburgh, K. H. Sato, T. R. White, W. J. Borucki, D. G. Koch, and J. M. Jenkins (2010, November). A Precise Asteroseismic Age and Radius for the Evolved Sun-like Star KIC 11026764. ApJ *723*(2), 1583–1598.

Mosser, B. (2011, December). Red giants unveiled. In G. Alecian, K. Belkacem, R. Samadi, and D. Valls-Gabaud (Eds.), *SF2A-2011: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, pp. 225–229.

Müller, A. C. (2017). *Einführung in Machine Learning mit Python : Praxiswissen Data Science* (1. Auflage. ed.). [Köln] Heidelberg: O'Reilly dpunkt.verlag GmbH.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12*, 2825–2830.

Pfaff, F., G. Kurz, and U. Hanebeck (2017, 11). Filtering on the unit sphere using spherical harmonics.

Pinsonneault, M. H., Y. P. Elsworth, J. Tayar, A. Serenelli, D. Stello, J. Zinn, S. Mathur, R. A. Garcia, J. A. Johnson, S. Hekker, D. Huber, T. Kallinger, S. Meszaros, B. Mosser, K. Stassun, L. Girardi, T. S. Rodrigues, V. S. Aguirre, D. An, S. Basu, W. J. Chaplin, E. Corsaro, K. Cunha, D. A. Garcia-Hernandez, J. Holtzman, H. Jonsson, M. Shetrone, V. V. Smith, J. S. Sobeck, G. S. Stringfellow, O. Zamora, T. C. Beers, J. G. Fernand ez-Trincado, P. M. Frinchaboy, F. R. Hearty, and C. Nitschelm (2019, February). VizieR Online Data Catalog: APOKASC-2 catalog of Kepler evolved stars (Pinsonneault+, 2018). *VizieR Online Data Catalog*, J/ApJS/239/32.

Pinsonneault, M. H., Y. P. Elsworth, J. Tayar, A. Serenelli, D. Stello, J. Zinn, S. Mathur, R. A. García, J. A. Johnson, S. Hekker, D. Huber, T. Kallinger, S. Mészáros, B. Mosser, K. Stassun, L. Girardi, T. S. Rodrigues, V. Silva Aguirre, D. An, S. Basu, W. J. Chaplin, E. Corsaro, K. Cunha, D. A. García-Hernández, J. Holtzman, H. Jönsson, M. Shetrone, V. V. Smith, J. S. Sobeck, G. S. Stringfellow, O. Zamora, T. C. Beers, J. G. Fernández-Trincado, P. M. Frinchaboy, F. R. Hearty, and C. Nitschelm (2018, December). The Second APOKASC Catalog: The Empirical Approach. ApJS *239*(2), 32.

Pols, O. (2011, September). Lecture course in Stellar Structure and Evolution.

Rauer, H., C. Aerts, J. Cabrera, and PLATO Team (2016, September). The PLATO Mission. *Astronomische Nachrichten 337*(8-9), 961.

Rhodes, E. J., J., R. K. Ulrich, and G. W. Simon (1977, December). Observations of nonradial p-mode oscillations on the sun. ApJ *218*, 901–919.

Ricker, G. R., J. N. Winn, R. Vanderspek, D. W. Latham, G. Á. Bakos, J. L. Bean, Z. K. Berta-Thompson, T. M. Brown, L. Buchhave, N. R. Butler, R. P. Butler, W. J. Chaplin, D. Charbonneau, J. Christensen-Dalsgaard, M. Clampin, D. Deming, J. Doty, N. De Lee, C. Dressing, E. W. Dunham, M. Endl, F. Fressin, J. Ge, T. Henning, M. J. Holman, A. W. Howard, S. Ida, J. M. Jenkins, G. Jernigan, J. A. Johnson, L. Kaltenegger, N. Kawai, H. Kjeldsen, G. Laughlin, A. M. Levine, D. Lin, J. J. Lissauer, P. MacQueen, G. Marcy, P. R. McCullough, T. D. Morton, N. Narita, M. Paegert, E. Palle, F. Pepe, J. Pepper, A. Quirrenbach, S. A. Rinehart, D. Sasselov, B. Sato, S. Seager, A. Sozzetti, K. G. Stassun, P. Sullivan, A. Szentgyorgyi, G. Torres, S. Udry, and J. Villasenor (2015, January). Transiting Exoplanet Survey Satellite (TESS). *Journal of Astronomical Telescopes, Instruments, and Systems 1*, 014003.

Schofield, M., W. J. Chaplin, D. Huber, T. L. Campante, G. R. Davies, A. Miglio, W. H. Ball, T. Appourchaux, S. Basu, T. R. Bedding, J. Christensen-Dalsgaard, O. Creevey, R. A. García, R. Handberg, S. D. Kawaler, H. Kjeldsen, D. W. Latham, M. N. Lund, T. S. Metcalfe, G. R. Ricker, A. Serenelli, V. Silva Aguirre, D. Stello, and R. Vanderspek (2019, March). The Asteroseismic Target List for Solar-like Oscillators Observed in 2 minute Cadence with the Transiting Exoplanet Survey Satellite. ApJS *241*(1), 12.

Schönberg, M. and S. Chandrasekhar (1942, September). On the Evolution of the Main-Sequence Stars. ApJ *96*, 161.

Tassoul, M. (1980, August). Asymptotic approximations for stellar nonradial pulsations. ApJS *43*, 469–490.

Themeßl, N., S. Hekker, J. Southworth, P. G. Beck, K. Pavlovski, A. Tkachenko, G. C. Angelou, W. H. Ball, C. Barban, E. Corsaro, Y. Elsworth, R. Handberg, and T. Kallinger (2018, August). Oscillating red giants in eclipsing binary systems: empirical reference value for asteroseismic scaling relation. MNRAS *478*(4), 4669–4696.

Ulrich, R. K. (1970, December). The Five-Minute Oscillations on the Solar Surface. ApJ *162*, 993.

Ulrich, R. K. (1986, July). Determination of Stellar Ages from Asteroseismology. ApJ *306*, L37.

Walker, G., J. Matthews, R. Kuschnig, R. Johnson, S. Rucinski, J. Pazder, G. Burley, A. Walker, K. Skaret, R. Zee, S. Grocott, K. Carroll, P. Sinclair, D. Sturgeon, and J. Harron (2003, September). The MOST Asteroseismology Mission: Ultraprecise Photometry from Space. PASP *115*(811), 1023–1035.

Xiang, M. S., X. W. Liu, H. B. Yuan, Z. Y. Huo, Y. Huang, C. Wang, B. Q. Chen, J. J. Ren, H. W. Zhang, Z. J. Tian, Y. Yang, J. R. Shi, J. K. Zhao, J. Li, Y. H. Zhao, X. Q. Cui, G. P. Li, Y. H. Hou, Y. Zhang, W. Zhang, J. L. Wang, Y. Z. Wu, Z. H. Cao, H. L. Yan, T. S. Yan, A. L. Luo, H. T. Zhang, Z. R. Bai, H. L. Yuan, Y. Q. Dong, Y. J. Lei, and G. W. Li (2017, May). LAMOST Spectroscopic Survey of the Galactic Anticentre (LSS-GAC): the second release of value-added catalogues. MNRAS *467*(2), 1890–1914.

Yu, J., T. R. Bedding, D. Stello, D. Huber, D. L. Compton, L. Gizon, and S. Hekker (2020, March). Asteroseismology of luminous red giants with Kepler I: long-period variables with radial and non-radial modes. MNRAS *493*(1), 1388–1403.