Draft version April 29, 2024
Typeset using LaTeX **twocolumn** style in AASTeX631

Flow-Based Generative Emulation of Grids of Stellar Evolutionary Models

Marc Hon [1,2] Yaguang Li [2] and Joel Ong [2]

[1] *Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*

[2] *Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA*

ABSTRACT

We present a flow-based generative approach to emulate grids of stellar evolutionary models. By interpreting the input parameters and output properties of these models as multi-dimensional probability distributions, we train conditional normalizing flows to learn and predict the complex relationships between grid inputs and outputs in the form of conditional joint distributions. Leveraging the expressive power and versatility of these flows, we showcase their ability to emulate a variety of evolutionary tracks and isochrones across a continuous range of input parameters. In addition, we describe a simple Bayesian approach for estimating stellar parameters using these flows and demonstrate its application to asteroseismic datasets of red giants observed by the *Kepler* mission. By applying this approach to red giants in open clusters NGC 6791 and NGC 6819, we illustrate how large age uncertainties can arise when fitting only to global asteroseismic and spectroscopic parameters without prior information on initial helium abundances and mixing length parameter values. We also conduct inference using the flow at a large scale by determining revised estimates of masses and radii for 15,388 *Kepler* field red giants. These new estimates show improved agreement with results from existing grid-based modelling for red giant branch stars and reveal distinct population-level features in the red clump, consistent with predictions from stellar models.

1. INTRODUCTION

Stellar models[1] form a cornerstone of contemporary astrophysics by enabling a direct comparison between observations and theory. A wide variety of variables dictate the evolution of stellar models, including fundamental input parameters (e.g., initial masses and chemical composition), parameters defining the prescription of input physics, the treatment of stellar convection (Kupka & Muthsam 2017), and the presence of stellar rotation (e.g., Ekström et al. 2012). Conversely, stellar models predict a broad range of observables such as stellar luminosities, effective temperatures, chemical abundances, and pulsation periods. Such observables are fundamental to identifying which model best aligns with observed data, enabling properties that can only be inferred from models — such as stellar ages — to be determined (Soderblom 2010).

Corresponding author: Marc Hon
mtyhon@mit.edu

[1] The term 'model' in this work refers only to simulated profiles of stars. This is to disambiguate from its meaning when used in the context of machine learning.

The broad diversity in both input and output parameters for grids of stellar models commonly poses a challenge for using grids to study the evolution and astrophysical properties of stars. A wide range of input parameters results in many free parameters dictating the evolution of models, resulting in the curse of dimensionality that makes the systematic exploration of models across the grid computationally infeasible. The outputs of stellar models, which are often observables, can also be high-dimensional, with complex and non-linear dependencies with input parameters. This complexity in inferring observables from grids of models has naturally led to the adoption of machine learning approaches in forward modelling tasks in grids of stellar models. In machine learning approaches thus far, grid parameters are inferred either in the form of individual numerical values (e.g., Verma et al. 2016; Bai et al. 2019; Bellinger et al. 2020; Mombarg et al. 2021; Panda et al. 2024) or intervals and distributions (e.g., Hon et al. 2020; Garraffo et al. 2021; Lyttle et al. 2021; Scutt et al. 2023). These approaches can be interpreted as a local emulation of the grid, as their predictions focus solely on specific regions of maximum likelihood within the grid. In other words, their task is to find a narrow region where

a grid's output parameters align most closely with the observed data. On the other hand, Li et al. (2022a) present a unique approach to the grid-based modelling task by emulating the full grid using a Gaussian process regressor as a generative surrogate model. Such an approach provided the desirable property of having an emulation that is globally consistent by construction, with a smoothly interpolated input parameter space that can be sampled across the grid to approximate the properties of stellar evolutionary models.

Drawing upon these insights, we present the use of normalizing flows as a highly flexible and expressive global emulator of grids of stellar models. Normalizing flows are a class of probabilistic machine learning techniques that learn invertible transformations between simple base distributions and more complex distributions (Dinh et al. 2015, 2017). Because such transformations are typically learned by training deep neural networks, normalizing flows can conduct the density estimation and sampling of diverse and elaborate high-dimensional distributions. As a result, these approaches have been applied in complex generative tasks and in modern probabilistic inference approaches including high-dimensional statistical modelling (e.g., Ting & Weinberg 2022; Van-Lane et al. 2023) and likelihood-free inference (e.g., Wang et al. 2023). Here, we apply normalizing flows to grids of stellar evolutionary models to emulate the input and output properties of the grids as high-dimensional distributions. By leveraging the expressive power of neural networks to interpolate in high dimensions, we demonstrate the utility of normalizing flows in exploring the complex relations between fundamental properties of the grid with its outputs, and stellar property inference.

2. NORMALIZING FLOWS

The central idea of a normalizing flow is to map a random latent variable $\mathbf{z}$, which is distributed following a simple probability density $p_0(\mathbf{z})$ (typically a multivariate normal), to another variable $\mathbf{y} = f(\mathbf{z})$ that is distributed following a more complex probability density $p_N(\mathbf{y})$. The transformation $f = f_{\theta_N} \circ \dots f_{\theta_2} \circ f_{\theta_1}$ is a series of functions parameterized by $\theta \in \theta_1, \theta_2, \cdots \theta_N$, whose values are learned by deep neural networks. Such a transformation describes how samples $\mathbf{z} \sim p_0(\mathbf{z})$ are mapped to $\mathbf{y} \sim p_N(\mathbf{y})$ following the change of variables formula:

$$p_0(\mathbf{z}) = p_N(\mathbf{y}) \cdot \left| \det \left(\frac{d\mathbf{y}}{d\mathbf{z}} \right) \right| = p_N(f(\mathbf{z})) \cdot \left| \det \left(J_f(\mathbf{z}) \right) \right|, \quad (1)$$

where J_f denotes the Jacobian of the transformation f. Training a normalizing flow to perform density estimation involves the optimization of θ to maximize the likelihood of observed data under the probability density described by the flow. To perform such training, a normalizing flow requires the transformation f to have the following properties:

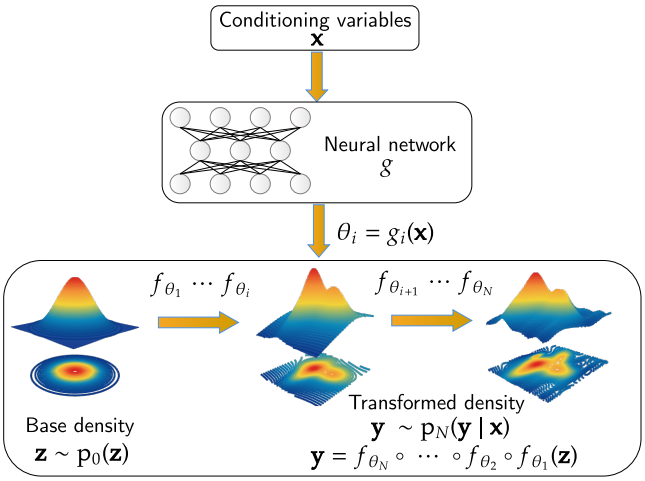

Figure 1. A schematic of conditional normalizing flows. The flow maps random variates $\mathbf{z}$ from a base probability density $\mathbf{p_0}(\mathbf{z})$ (here shown as a 2D normal distribution) to another variable $\mathbf{y}$. The mapping occurs over a series of N invertible transformations $f = f_{\theta_i}, i \in [1, 2, \cdots, N]$, where θ_i is learned by a neural network and is conditioned by contextual inputs $\mathbf{x}$. The probability density of $\mathbf{y}$ is subsequently conditioned by $\mathbf{x}$, such that $y_N \sim p_N(\mathbf{y}|\mathbf{x})$. In this work, the $\mathbf{y}$ corresponds to the output stellar properties of an evolutionary grid of models, while $\mathbf{x}$ corresponds to the input parameters of the grid.

- **Differentiability:** This property enables the factor $|\det(J_f(\mathbf{z}))|$ in Equation 1 to be computed, which ensures that p_N integrates to unity and remains a valid probability density under the transformation. Additionally, the differentiability of the transformation permits efficient gradient-based optimization methods to be used when training the normalizing flow.

- **Invertibility:** This property ensures a one-to-one correspondence between $\mathbf{y}$ and $\mathbf{z}$ under the specified transformations. Consequently, the exact likelihood of samples $\mathbf{y}$ can be computed. The condition of invertibility also necessitates that the dimensionality of both p_0 and p_N are equal.

In literature, there are various classes of transformations that inherit both of these properties. In this work, we use Autoregressive Neural Spline Flows (Durkan et al. 2019), which are Masked Autoregressive Transforms (Papamakarios et al. 2017) combined with Monotonic Rational Quadratic Spline coupling transforms (Durkan

Table 1. Variables from MESA stellar models used for density estimation using the Conditional Normalizing Flow focused on subgiant and dwarf star models, $\mathrm{CNF}_{\mathrm{dwarf}}$.

Conditioning Input Parameters ($\mathbf{x}$)		
Parameter	**Definition**	**Range**
M	Stellar mass	$0.7\,M_\odot \leq M \leq 2.5\,M_\odot$
$\log_{10} Z$	Initial metal fraction	$-4.934 \leq \log_{10} Z \leq -1.291$
Y	Initial helium fraction	$0.23 \leq Y \leq 0.37$
α	Mixing length parameter	$1.0 \leq \alpha \leq 2.7$
$\widehat{f_{\mathrm{ov,env}}}$	Min-max normalized[1] $f_{\mathrm{ov,env}}$, where $f_{\mathrm{ov,env}}$ is the efficiency parameter of convective overshoot in stellar envelope	$2\times10^{-6} \leq f_{\mathrm{ov,env}} \leq 2\times10^{-2}$
$\widehat{f_{\mathrm{ov,core}}}$	Min-max normalized[1] $f_{\mathrm{ov,core}}$, where $f_{\mathrm{ov,core}}$ is the efficiency parameter of convective overshoot in stellar core	$3\times10^{-6} \leq f_{\mathrm{ov,core}} \leq 3\times10^{-2}$
Output Properties ($\mathbf{y}$)		
Property	**Definition**	**Range**
$\widehat{T_{\mathrm{eff}}}$	z-normalized[2] $\log_{10} T_{\mathrm{eff}}$, where T_{eff} is stellar effective temperature	$3344\,\mathrm{K} \leq T_{\mathrm{eff}} \leq 19252\,\mathrm{K}$
$\widehat{\nu_{\mathrm{max}}}$	$\log_{10} \nu_{\mathrm{max}}$, where ν_{max} is the frequency at maximum power estimated by MESA	$300\,\mu\mathrm{Hz} \leq \nu_{\mathrm{max}} \leq 6684\,\mu\mathrm{Hz}$
$\widehat{\Delta\nu}$	$\log_{10} \Delta\nu$, where $\Delta\nu$ is the large frequency separation value computed using a weighted linear fit to $l=0$ mode frequencies around ν_{max}. The weights are determined by fitting a Gaussian envelope[3] centered about ν_{max} to the $l=0$ mode frequencies	$17.94\,\mu\mathrm{Hz} \leq \Delta\nu \leq 232.48\,\mu\mathrm{Hz}$
ϵ	Dimensionless phase offset of the asymptotic relation of p-mode frequencies. Determined as the intercept of the weighted linear fit to $\Delta\nu$.	$0.43 \leq \epsilon \leq 2.41$
$\widehat{\delta\nu_{01}}$	z-normalized[2] $\delta\nu_{01}$, where $\delta\nu_{01}$ is the frequency separation between $l=0$ and $l=1$ frequencies. This quantity is computed as $\nu_0 + 0.5\Delta\nu - \nu_1$, averaged across radial orders using weights determined by fitting a Gaussian envelope[3] centered about ν_{max} to the $l=1$ mode frequencies	$-4.42\,\mu\mathrm{Hz} \leq \delta\nu_{01} \leq 13.42\,\mu\mathrm{Hz}$
$\widehat{\delta\nu_{02}}$	$\log_{10} \delta\nu_{02}$, where $\delta\nu_{02}$ is the small frequency separation value computed as the difference between $l=0$ and $l=2$ frequencies. This quantity is computed as $\nu_0 - \nu_2$, averaged across radial orders using weights determined by fitting a Gaussian envelope[3] centered about ν_{max} to the $l=2$ mode frequencies	$0.31\,\mu\mathrm{Hz} \leq \delta\nu_{02} \leq 20.92\,\mu\mathrm{Hz}$
$\widehat{\delta\nu_{03}}$	$\log_{10} \delta\nu_{03} + 3$, where $\delta\nu_{03}$ is the frequency separation between $l=0$ and $l=3$ frequencies. This quantity is computed as $\nu_0 + 0.5\Delta\nu - \nu_3$, averaged across radial orders using weights determined by fitting a Gaussian envelope[3] centered about ν_{max} to the $l=3$ mode frequencies	$-1.9\,\mu\mathrm{Hz} \leq \delta\nu_{03} \leq 41.0\,\mu\mathrm{Hz}$
$\hat{R}$	$\log_{10} R$, where R is stellar radius in units of $R_\odot$	$0.61\,R_\odot \leq R \leq 5.08\,R_\odot$
$\hat{\tau}$	$\log_{10} \tau$, where τ is stellar age in units of Gyr	$0.0013\,\mathrm{Gyr} \leq \tau \leq 100\,\mathrm{Gyr}$

[1] The quantity is squeezed into an interval $\in [-1, 1]$ by first subtracting its minimum value across the dataset and then dividing by the difference between the maximum and minimum values over the dataset.
[2] The quantity is first subtracted by its mean value over the dataset, then divided by its standard deviation over the dataset.
[3] The width of the envelope is determined as $W = \nu_{\mathrm{max}}^k \cdot e^b$, where $k = 0.9638, b = -1.7145$ based on parametric fits to the widths of *Kepler* asteroseismic targets (Li et al. 2020).

Table 2. Variables from the grid of models accompanying AsfGrid used for training the $\text{CNF}_{\text{asfgrid}}$ Conditional Normalizing Flow.

Conditioning Input Parameters ($\mathbf{x}$)		
Parameter	**Definition**	**Range**
M	Stellar mass	$0.6\,M_\odot \leq M \leq 5.5\,M_\odot$
[Fe/H]	Metallicity, determined from the grid as $\log_{10} Z/Z_\odot$, where $Z_\odot = 0.019$ following Sharma et al. (2016)	$-3\,\text{dex} \leq \text{[Fe/H]} \leq 0.4\,\text{dex}$
E	Evolutionary state, defined as -1 for lower giant branch stars with $\nu_{\max} \geq 300\,\mu$Hz; 0 for hydrogen shell-burning red giant stars with $10\mu\text{Hz} \leq \nu_{\max} < 300\,\mu$Hz; 1 for hydrogen shell-burning giant stars with $\nu_{\max} < 10\,\mu$Hz; 2 for helium-core burning stars	$E \in [-1, 0, 1, 2]$
Output Properties ($\mathbf{y}$)		
Property	**Definition**	**Range**
$\widehat{T_{\text{eff}}}$	$\log_{10} T_{\text{eff}}$, where T_{eff} is stellar effective temperature	$2910\,\text{K} \leq T_{\text{eff}} \leq 22800\,\text{K}$
$\widehat{\Delta\nu}$	$\log_{10} \Delta\nu$, where $\Delta\nu$ is AsfGrid's large frequency separation value computed using radial mode frequencies using GYRE (Townsend & Teitler 2013)	$0.12\,\mu\text{Hz} \leq \Delta\nu \leq 77.42\,\mu\text{Hz}$
$\widehat{\nu_{\max}}$	$\log_{10} \nu_{\max}$, where $\nu_{\max}$ is the frequency at maximum power estimated by MESA as reported by the grid accompanying AsfGrid	$0.05\,\mu\text{Hz} \leq \nu_{\max} \leq 1000\,\mu\text{Hz}$
$\hat{R}$	$\log_{10} R$, where R is stellar radius in units of $R_\odot$	$1.5\,R_\odot \leq R \leq 193\,R_\odot$
$\hat{\tau}$	$\log_{10} \tau$, where τ is stellar age in units of Gyr	$0.05\,\text{Gyr} \leq \tau \leq 53\,\text{Gyr}$

et al. 2019) as implemented by the `zuko`[2] library version 0.3.2 (Rozet et al. 2024). A brief explanation of these transforms and their implementation in this work are provided in Appendix A.

Critically, the deep neural networks involved in the estimation of θ can be conditioned on contextual input variables $\mathbf{x}$, such that $\theta \to \theta(\mathbf{x})$. Consequently, the generative model is modified into a Conditional Normalizing Flow (CNF, Winkler et al. 2019) and yields conditional probability densities $p_N(\mathbf{y}|\mathbf{x})$, as illustrated in Figure 1. By incorporating conditioning variables, a CNF leverages the expressiveness of neural networks to estimate conditional and marginal probability distributions across the range of each grid's input parameters. This is performed by optimizing the CNF's neural network to maximize the likelihood of the distribution described by the grid of models under the predicted conditional distribution from the CNF.

3. GRIDS OF EVOLUTIONARY MODELS

3.1. *MESA Grid of Dwarf Star Models*

We train a CNF on a grid of models generated using Modules for Experiments in Stellar Astrophysics (MESA, version r23.01.1; Paxton et al. 2011, 2013, 2015, 2018, 2019). Our prescription for the grid of models closely follows that from Li et al. (2023). More specifically, we adopt the Asplund et al. (2009) solar mixture, whereby $X_\odot = 0.7381, Y_\odot = 0.2485, Z_\odot = 0.0134$ and the metallicity [M/H] is correspondingly defined as $\text{[M/H]} = \log_{10}(Z/X) - \log_{10}(Z_\odot/X_\odot)$. We select opacity tables in MESA following this choice of metal mixture, which are described by a combination of electron conduction opacities (Cassisi et al. 2007), OPAL radiative opacties (Iglesias & Rogers 1993, 1996), low-temperature opacities (Ferguson et al. 2005) and data at the high-temperature Compton-scattering regime (Buchler & Yueh 1976). We use the equation of state provided by MESA, which combines data from OPAL (Rogers & Nayfonov 2002), SCVH (Saumon et al. 1995), PTEH (Pols et al. 1995), HELM (Timmes & Swesty 2000), and PC (Potekhin & Chabrier 2010). We adopt nuclear reaction rates from JINA REACLIB database (Cyburt et al. 2010), using a minimal set of elements specified in MESA's `basic.net`. Both atomic diffusion and gravitational settling are not included in the models. Convection is treated using the Henyey et al. (1965) formalism, whose efficiency is controlled by a mixing length parameter α. We apply convective overshoot in both the envelope and core following the exponential scheme described by Herwig (2000), with the corresponding efficiency parameters $f_{\text{ov,env}}$ and $f_{\text{ov,core}}$, respectively. For the treatment of surface boundary conditions, we adopt the grey model atmosphere in conjunc-

[2] https://zuko.readthedocs.io

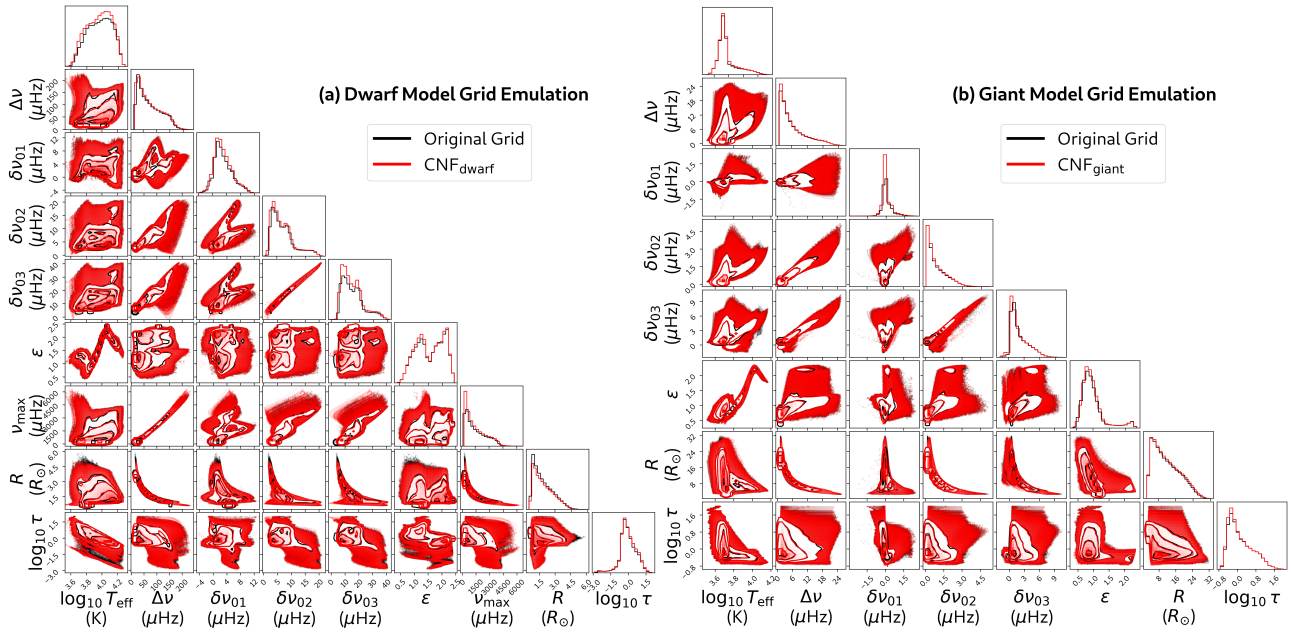

Figure 2. (a) Corner plot showing the original distribution (black) of output parameters $\mathbf{y}$ in the MESA grid of dwarf models and the emulated grid from CNF$_{\rm dwarf}$ in red. **(b)** Same as (a), except the comparison is between the MESA grid of giant models and the emulated grid from CNF$_{\rm giant}$. Descriptions of each parameter are listed in Table 1.

tion with the Eddington $T-\tau$ integration method (Eddington 1926). We use GYRE (version 7.0, Townsend & Teitler 2013) to calculate adiabatic frequencies from the structure profiles computed from MESA. Besides the frequencies of radial modes ($\ell = 0$), we also calculate the frequencies of decoupled (pure p) dipolar ($\ell = 1$) and quadrupolar modes ($\ell = 2$) and octupolar modes ($\ell = 3$) following the Ong & Basu (2020) decoupling approach.

Evolutionary tracks in the grid are varied in $[M,\ \log_{10} Z,\ Y,\ \alpha,\ \widehat{f_{\rm ov,env}},\ \widehat{f_{\rm ov,core}}]$, with the definitions of these input parameters presented in Table 1. These input parameters are varied by Sobol sampling across six dimensions to produce a total of 8,191 unique combinations that constitute the conditioning variable vector $\mathbf{x}$. Each track is evolved from the zero age main sequence up to pre-core helium ignition phase at the tip of the red giant branch. The CNF is trained to estimate an 9D probability density $p(\mathbf{y}|\mathbf{x})$, with $\mathbf{y} = [\widehat{T_{\rm eff}}, \widehat{\Delta\nu}, \widehat{\nu_{\rm max}}, \widehat{\delta\nu_{01}}, \widehat{\delta\nu_{02}}, \widehat{\delta\nu_{03}}, \epsilon, \hat{R}, \hat{\tau}]$ as outputs from the model (see Table 1) at each evolutionary timestamp along the grid. To focus on the density modelling of subgiant and dwarf star models within the grid, we select only models with $\nu_{\rm max} > 300\mu$Hz for this particular CNF, resulting in a total of 443,874 models in this grid. We denote the CNF trained on this grid as CNF$_{\rm dwarf}$.

3.2. *MESA Grid of Giant Star Models*

The same base grid in Section 3.1 is adopted here, with two differences. First, we select only models with $\nu_{\rm max} \leq 300\mu$Hz to focus the density estimation task on models on the red giant branch, which results in a grid of 738,939 models. Next, the CNF uses $\nu_{\rm max}$ as a conditioning variable instead of an output property to simplify selecting emulated models at a specific evolutionary stage, which is useful for the inference task described in Section 5.1. Therefore, we now have $\mathbf{x} = [M,\ \log_{10} Z,\ Y,\ \alpha,\ \widehat{f_{\rm ov,env}},\ \widehat{f_{\rm ov,core}},\ \widehat{\nu_{\rm max}}]$ and $\mathbf{y} = [\widehat{T_{\rm eff}}, \widehat{\Delta\nu}, \widehat{\delta\nu_{01}}, \widehat{\delta\nu_{02}}, \widehat{\delta\nu_{03}}, \epsilon, \hat{R}, \hat{\tau}]$. We denote the CNF trained on this grid as CNF$_{\rm giant}$.

3.3. *AsfGrid Evolutionary Models*

We train a CNF on a grid accompanying the publicly available Python module AsfGrid version 0.06 (Sharma et al. 2016; Stello & Sharma 2022). This is a grid of stellar evolutionary models used for ensemble asteroseismic studies of red giants (e.g., Yu et al. 2018; Zinn et al. 2020, 2022; Li et al. 2023). It adopts a quasi-uniform sampling in M and [Fe/H] across evolutionary tracks, as shown in Figure 1 in Stello & Sharma (2022). To specifically focus on the density modelling of red giant stars within the grid, we only select stellar models with $\nu_{\rm max} \leq 1000\mu$Hz on the red giant branch (`evstate` = 1) and models in the red clump phase (`evstate` = 2) from the grid, resulting in a total of 1,071,210 models. The CNF is trained to estimate the conditional density $p(\mathbf{y}|\mathbf{x})$, where $\mathbf{y} = [\widehat{T_{\rm eff}}, \widehat{\Delta\nu}, \widehat{\nu_{\rm max}}, \hat{R}, \hat{\tau}]$ and $\mathbf{x} = [M, [\mathrm{Fe/H}], E]$, are outputs from the model at each evolutionary timestamp along the grid. The summary of each variable is shown in Table 2. Notably, E is a categorical conditional vari-

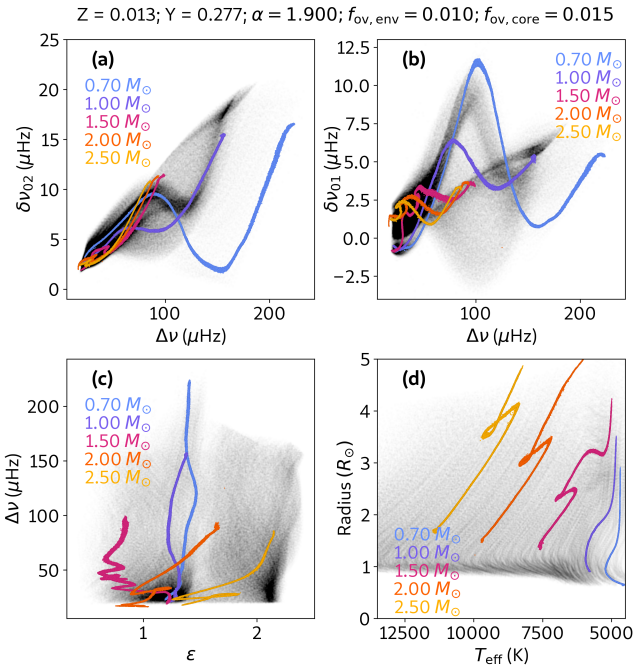

Figure 3. Emulated evolutionary tracks in the grid of MESA dwarf models, predicted as the conditional distribution $p_N(\mathbf{y}|\mathbf{x})$ from $\mathrm{CNF}_{\mathrm{dwarf}}$. Here, $\mathbf{x} = (M, Z, Y, \alpha, f_{\mathrm{ov,env}}, f_{\mathrm{ov,core}})$, where $M \in [0.7, 1.0, 1.5, 2.0, 2.5]\, M_\odot$, with the values of other input parameters listed at the top of the figure. This specific combination of $\mathbf{x}$ is an interpolated value and hence does not exist within the training grid (black). Each panel shows two components from the predicted 9D distribution of output properties in the form of: **(a)** the asteroseismic C-D diagram, **(b)** the $\delta\nu_{01} - \Delta\nu$) diagram, **(c)** the $\Delta\nu - \epsilon$ diagram, **(d)** the 'Hertzsprung-Russell' diagram.

able and is used to focus on sampling models from a particular evolutionary phase. We denote the CNF trained on this grid as $\mathrm{CNF}_{\mathrm{asfgrid}}$.

4. EMULATING GRID OUTPUT PROPERTIES

We first visualize how well the CNF can emulate the global distribution of model outputs from a grid. Using the original, Sobol-sampled distribution of conditioning variables $\mathbf{x}$ from the MESA grid (defined in Table 1), we draw 443,874 samples from $\mathrm{CNF}_{\mathrm{dwarf}}$ and 738,939 samples from $\mathrm{CNF}_{\mathrm{giant}}$. These numbers are chosen to match the number of models in the grid each CNF was trained on. A comparison between the original and emulated distribution of models in the grid is shown in Figure 2. Qualitatively, we find that both CNFs are able to closely emulate the distribution of individual output parameters while also capturing global correlations between the parameters despite the high-dimensionality of the grid.

4.1. *Emulating Evolutionary Tracks*

To examine the emulation ability of the CNF in high dimensions, we use $\mathrm{CNF}_{\mathrm{dwarf}}$ to predict conditional and marginal distributions of evolutionary tracks corresponding in $T_{\mathrm{eff}} - R$ space (a 'Hertzsprung-Russell' diagram) as well as the C-D, $\Delta\nu - \epsilon$, and $\delta\nu_{01} - \Delta\nu$ asteroseismic diagrams. These were deliberately selected to highlight the capability of the CNF in emulating varied and complex evolutionary tracks.

The C-D diagram plots the large frequency separation, $\Delta\nu$, against the small frequency separation, $\delta\nu_{02}$ (Christensen-Dalsgaard 1984). Because $\delta\nu_{02}$ is sensitive to the remaining core hydrogen content of low-mass stars, its value correlates strongly with stellar age when observed as a function of $\Delta\nu$, which approximates mean stellar density. As a result, the position of observed stars within the C-D diagram is a powerful diagnostic for the structure and evolution of low-mass stars along the main sequence (e.g., Roxburgh & Vorontsov 2003; Otí Floranes et al. 2005; White et al. 2011). The $\delta\nu_{01} - \Delta\nu$ diagram in principle has a similar discriminative power in age to the C-D diagram, albeit with a smaller age sensitivity (e.g., Lund et al. 2017). The phase offset ϵ carries information related to the structure of the stellar acoustic mode cavity (Ong & Basu 2019) and has been used for its evolutionary diagnostics (Christensen-Dalsgaard et al. 2014) and mode identification properties (White et al. 2012). Evolutionary tracks in the $\epsilon - \Delta\nu$ diagram, however, are known to vary rapidly and converge for late-stage low-mass main sequence stars, (c.f. White et al. 2011; Ong & Basu 2019), which typically limits the applicability of ϵ to less evolved dwarf stars. Meanwhile, evolutionary tracks of T_{eff} and R for dwarf stars, similar to their trajectories in the Hertzsprung-Russell diagram, vary smoothly with mass and metallicity but can develop sharp morphological features including the Henyey hook and the dip at the base of the red giant branch (Mengel et al. 1979). In Figure 3, we present examples of emulated evolutionary tracks in these four diagrams. For each specific combination of $\mathbf{x}$, we draw a total of 10^6 samples from $\mathrm{CNF}_{\mathrm{dwarf}}$ to yield the conditional distribution $p_N(\mathbf{y}|\mathbf{x})$. In general, we find that $\mathrm{CNF}_{\mathrm{dwarf}}$ is capable of faithfully emulating complex patterns in evolutionary tracks, including intersecting and looping trajectories, sharp transitions, and oscillatory variations.

4.2. *Continuum of Evolutionary Tracks*

Importantly, sampled evolutionary tracks from the CNFs smoothly vary across variations in input grid parameters, in alignment with expectations of stellar evolution. The smoothness of the CNF emulation is at-

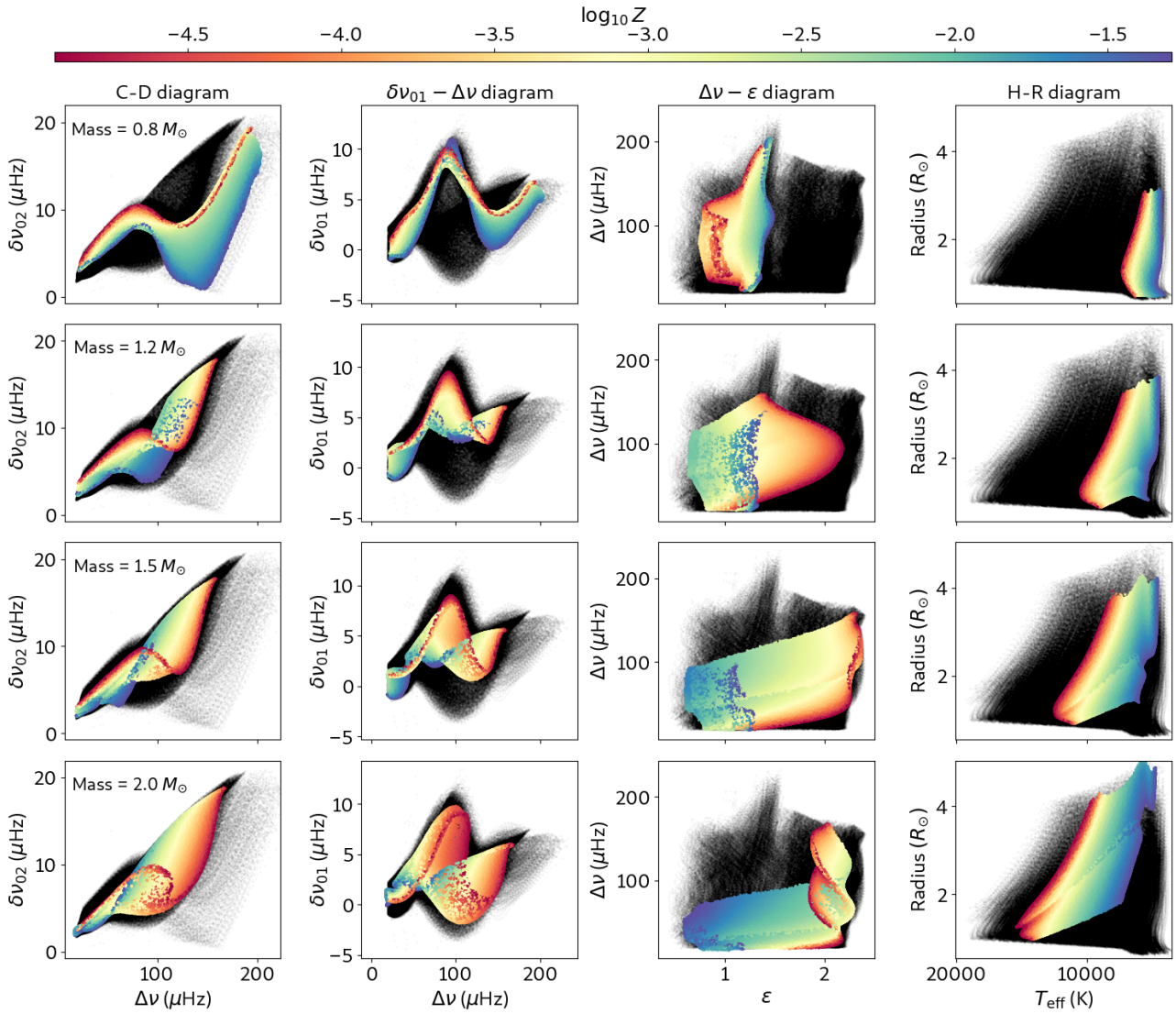

Figure 4. Continua of emulated evolutionary tracks over metallicity by $\mathrm{CNF}_{\mathrm{dwarf}}$, shown as the distribution of $\mathbf{y}$ given $\bar{\mathbf{x}} = (M, Y, \alpha, f_{\mathrm{ov,env}}, f_{\mathrm{ov,core}})$ marginalized over Z, or $p_N(\mathbf{y}|\bar{\mathbf{x}}) \approx \int p(\mathbf{y}|Z, \bar{\mathbf{x}})\, p(Z)\, dZ$. Each row shows $\mathrm{CNF}_{\mathrm{dwarf}}$ conditioned on a specific value of stellar mass, with the adopted values of the other conditional values in $\bar{\mathbf{x}}$ the same as those used in Figure 3. Each column represents a different type of diagram for which evolutionary tracks are commonly visualized.

tributed to the inherent smoothness of the CNF's latent space and the ability of its neural network to interpolate across the space of input parameters. We visualize the smoothness of the emulation by generating a continuum of evolutionary tracks, which is done by marginalizing over the range of specific input parameters.

We present a continuum of evolutionary tracks varied in metallicity in Fig. 4. Metallicity affects radiative opacities, equations of state, and nuclear reaction rates (Mowlavi et al. 1998), resulting a substantial influence on the observable properties of stars as they evolve. As demonstrated in Fig. 4, this influence across a continuous range of observables can be distinctly visualized using the CNF by smoothly interpolating across Z. This marginalization over Z requires minimal modification from the conditional distribution scenario (Fig. 3) by simply sampling with $\log_{10} Z \sim \mathcal{U}(-4.934, -1.291)$, rather than fixing it to a specific value.

To present another example, we next display continuous variations from varying the mixing length parameter, α, in Figure 5. The dimensionless quantity α defines a characteristic distance that a parcel of convective material can travel within the star. Hence, it is particularly important for characterizing the efficiency of energy transport in stars with superadiabatic convective zones, which includes stars with $M \lesssim 1.2 M_{\odot}$ and giant stars (Böhm-Vitense 1958; Henyey et al. 1965). The impact of varying α on stellar observables is of par-

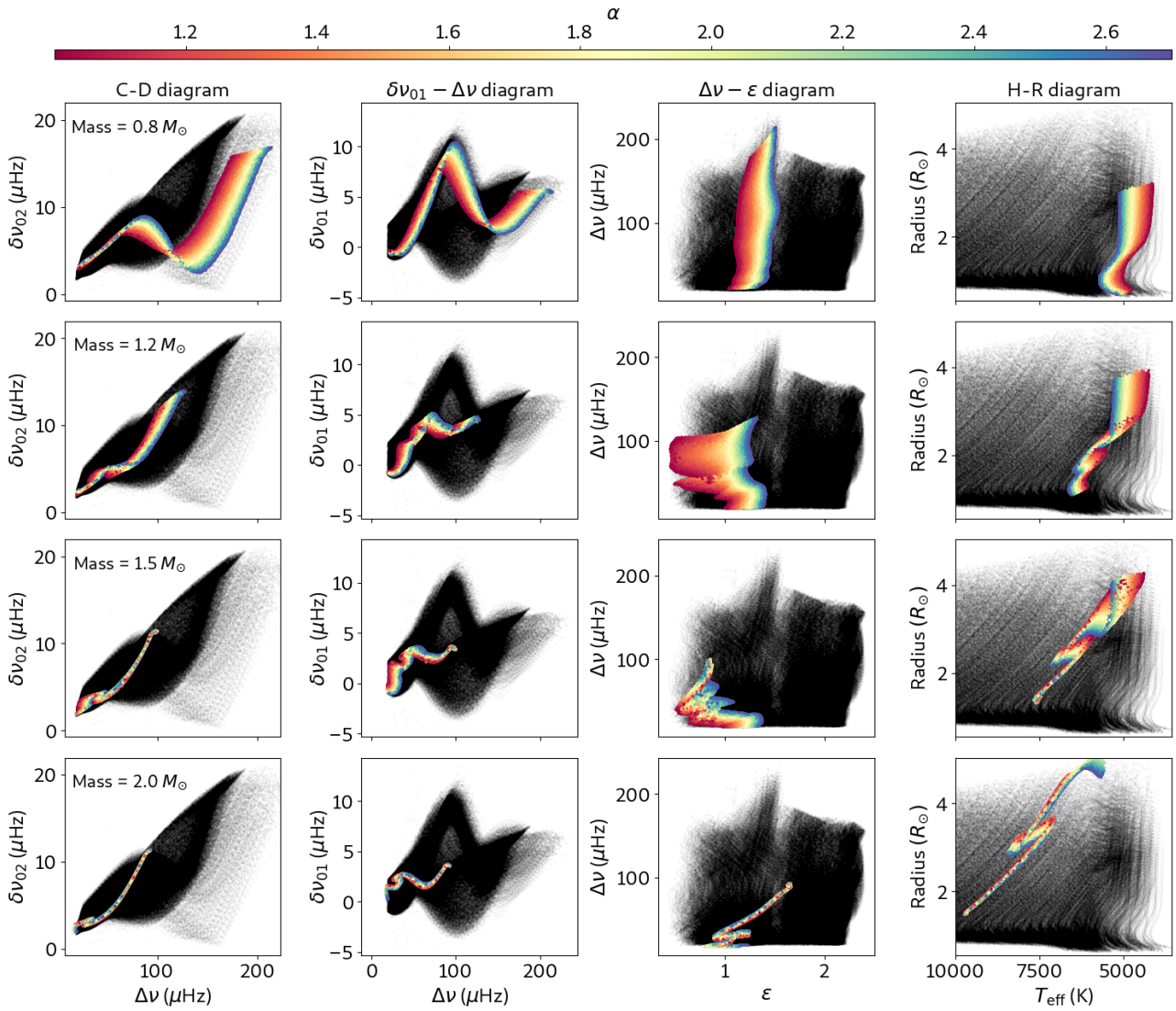

Figure 5. Same as Figure 4, except that the marginalization is now performed over the mixing length parameter α. Therefore, the distributions predicted from $\mathrm{CNF}_{\mathrm{dwarf}}$ here indicate $p_N(\mathbf{y}|\bar{\mathbf{x}}) \approx \int p(\mathbf{y}|\alpha, \bar{\mathbf{x}})\, p(\alpha)\, d\alpha$, with $\bar{\mathbf{x}} = (M, Z, Y, f_{\mathrm{ov,env}}, f_{\mathrm{ov,core}})$ and $Z = 0.013$.

ticular interest, especially given mounting evidence the value of α requires calibration across stars (e.g., Joyce & Chaboyer 2018a,b; Li et al. 2018) and contains dependencies on stellar composition and evolution (e.g., Tanner et al. 2013; Tayar et al. 2017) In Fig. 5, we observe the expected global systematic shift of evolutionary tracks to cooler temperatures in the H-R diagram with lower α due to a lower energy transport efficiency. This effect is markedly less pronounced on the main sequence for stars with $M > 1.2 M_{\odot}$ as a consequence of their interior structure comprising convective cores with radiative envelopes. As described by Joyce & Tayar (2023) and observed in the Figure, the T_{eff} of these stars are nonetheless modified with varying α once they evolve past the main sequence turn-off, corresponding to the development of a convective envelope during the subgiant phase and beyond. Overall, Figs. 4 and 5 demonstrate the ability of CNF-based emulation to distinctly quantify the extent of change of evolutionary tracks with respect to specific input parameters. This ability allows disentangling variations in observables as a result of a change in only one specific parameter, which can be insightful in guiding the optimization of model fitting approaches.

4.3. *Emulating Isochrones*

The CNF can additionally accept constraints on output properties to emulate conditional distributions, which is performed by drawing samples from the CNF and retaining only those that have emulated properties that match user-specified constraints. Since we

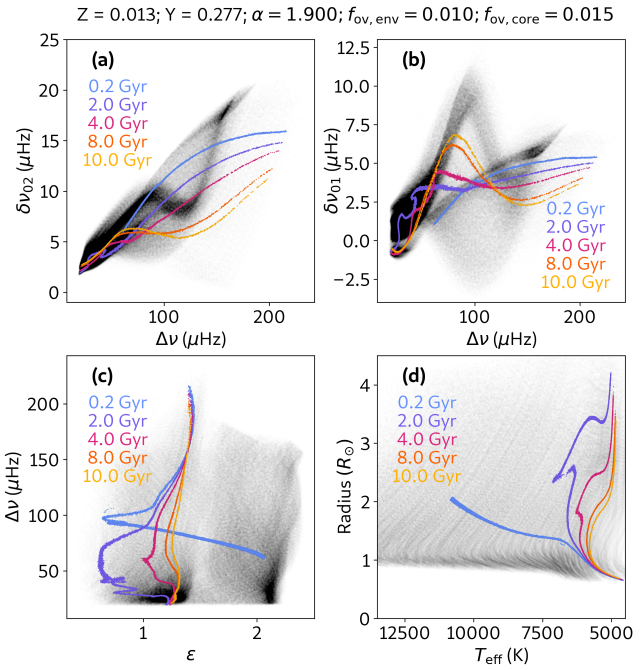

Figure 6. Same as Fig. 3, except here $\mathrm{CNF}_{\mathrm{dwarf}}$ predicts isochrones instead of evolutionary tracks, such that $p_N(\bar{\mathbf{y}}|\tau, \mathbf{x})$. By marginalizing over mass such that $M \sim \mathcal{U}(0.7, 2.5)M_\odot$, we set $\tau \in [0.2, 2, 4, 8, 10]$ Gyr to be a conditional variable, where the vector $\bar{\mathbf{y}}$ is $\mathbf{y}$ excluding τ. Other components of $\mathbf{x}$ are listed above the figure. In practice, $p_N(\bar{\mathbf{y}}|\tau, \mathbf{x})$ is obtained by drawing samples from $p_N(\mathbf{y}|\mathbf{x})$ and accepting samples with an age within a user-defined tolerance (here 0.01 Gyr) of τ.

have stellar age as an output property ($\tau \in \mathbf{y}$) across our trained CNFs, we can emulate isochrones by imposing an age restriction upon generated samples from $p_N(\mathbf{y}|\mathbf{x})$. Fig. 6 demonstrates that emulated isochrones with $\mathrm{CNF}_{\mathrm{dwarf}}$ can indeed replicate true isochrones, evidenced by the observed distributions in the radius-temperature diagram and the stratification of tracks in $\delta\nu_{01}$ and $\delta\nu_{02}$ relative to $\Delta\nu$, which is the defining property of the C-D diagram for age-dating low-mass dwarf stars (e.g., Fig. 7 of White et al. 2011, Fig. 4 of Bellinger et al. 2019).

The ability to emulate the stellar properties of mono-age models is valuable for studying the properties of co-eval systems such as binary stars and stellar clusters, given that fitting the observables of such systems commonly involve varying isochrones across a range of input parameters like mass and metallicity (e.g., An et al. 2007). Similar to our approach with evolutionary tracks, it is straightforward to emulate a continuum of isochrones spanning a range of input parameters by marginalizing over components in $\mathbf{x}$ at specific timestamps. We show such an emulation that marginalizes

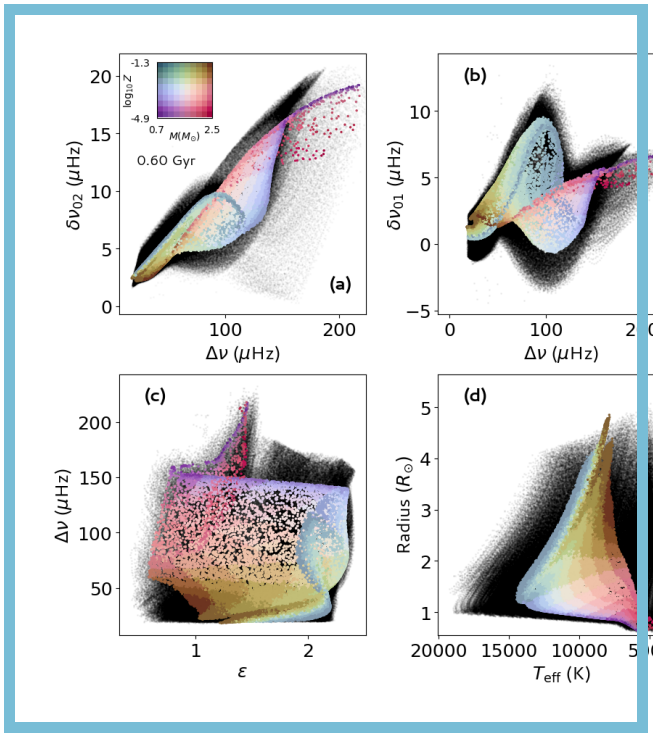

Figure 7. Continua of isochrones in the form of emulated distributions from $\mathrm{CNF}_{\mathrm{dwarf}}$ that are conditioned on age and marginalized over both metallicity (Z) and mass (M). The value of age (τ) is varied first from $0.05 - 1$ Gyr in increments of 0.05 Gyr, then from $1 - 15$ Gyr in increments of 0.15 Gyr. The distribution at each age timestamp represents $p_N(\bar{\mathbf{y}}|\bar{\mathbf{x}}, \tau) \approx \int\int p(\bar{\mathbf{y}}|Z, M, \tau, \bar{\mathbf{x}})\, p(Z)\, p(M|\tau)\, dZ\, dM$, where $\bar{\mathbf{y}}$ is $\mathbf{y}$ excluding τ. Meanwhile, the vector $\bar{\mathbf{x}}$ is $\mathbf{x}$ excluding M and Z, with other input parameter values the same as those in Figure 6. An animated version of this figure is available in the HTML version of the final article.

over both mass and metallicity in Fig. 7 across the age range of $0.1 - 15$ Gyr, which highlights the CNF's ability to capture continuously evolving variations in isochrones over age.

Overall, the CNF emulator shows great potential as an instructive tool for visualizing how input parameters and output properties within a grid of stellar models interact with one another. To highlight this aspect, we present a publicly accessible interactive demonstration of $\mathrm{CNF}_{\mathrm{dwarf}}$ that allows users to explore the conditional distributions in the evolutionary diagrams presented in this work, with details in Appendix B.

5. STELLAR PARAMETER INFERENCE

Because the CNF interprets the distribution of models within a grid as a density, it can act as a generative prior in Bayesian inference tasks. Given samples $\mathbf{x}$ drawn from a prior distribution $p(\mathbf{x})$, the role of the CNF is to directly map the samples into $\mathbf{y}$, which exist in the

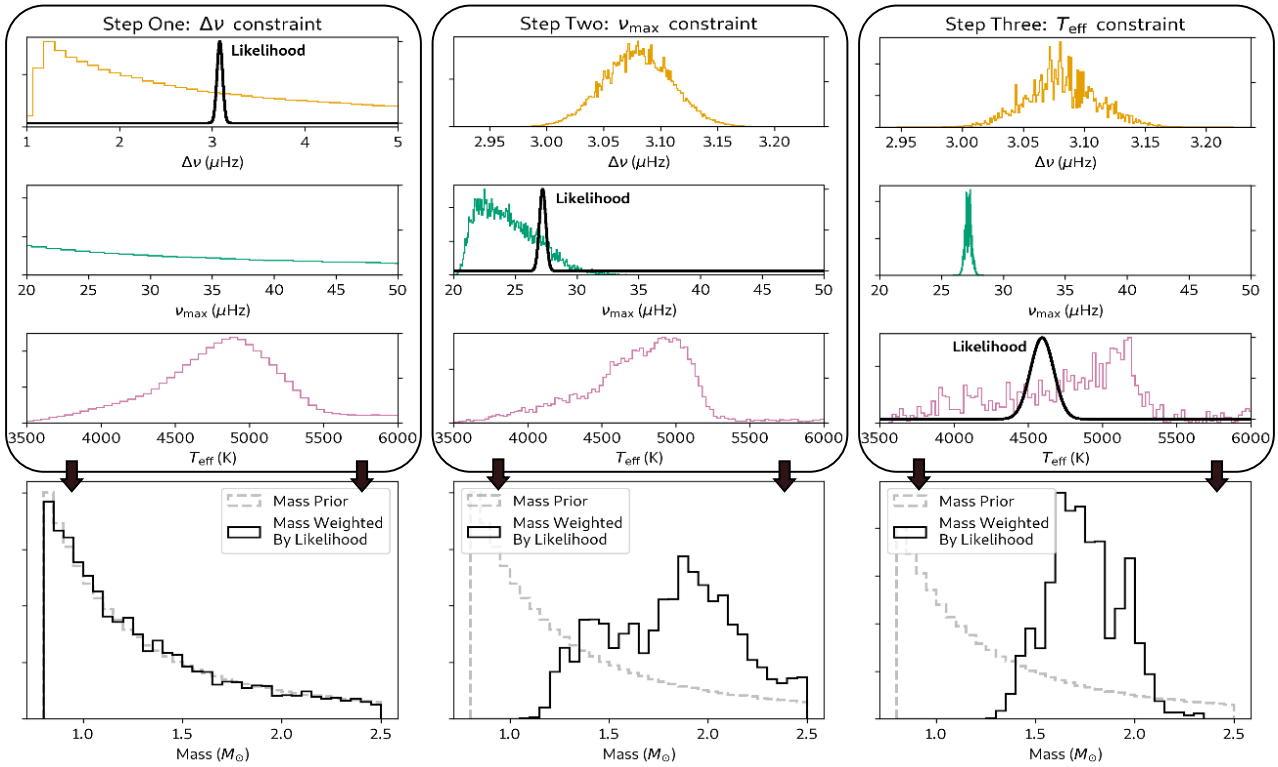

Figure 8. Sampling a mass posterior distribution from $\text{CNF}_{\text{asfgrid}}$ using constraints from observables $\Delta\nu$, $\nu_{\max}$, and T_{eff}. The upper three panels in each column (steps one to three) contain histograms of emulated observables from the CNF using a mass prior that is shown in the bottom panels. In step one, a constraint of $\Delta\nu = 3.08 \pm 0.03\,\mu$Hz forms the applied likelihood. In step two, a constraint of $27.2 \pm 0.3\,\mu$Hz is applied. In step three, a constraint of 4593 ± 78 K is applied. In the bottom panel of each step, samples from the prior mass distribution are directly weighted by the applied likelihood and are resampled using SIR to approximate a posterior distribution. Note that from step two onwards, the distributions of emulated observables also change due the weighting of a likelihood from the previous step.

space of output properties (or observables). These samples can subsequently be weighted using some likelihood of observed data, $p(\mathbf{y}|\mathbf{x})$. Because the CNF provides a one-to-one mapping between samples $\mathbf{x}$ and samples $\mathbf{y}$, the weights from the likelihood can be directly applied to samples drawn from the prior distribution. Following Bayes theorem, we would expect that weighting the prior with the data likelihood following this approach would thus provide an approximation to a posterior distribution.

Formally, this approach is known as Sampling/Importance Resampling (SIR, Rubin 1988). Following the description by Smith & Gelfand (1992), SIR seeks to estimate the distribution of a target density $h(\theta) = k(\theta)/\int k(\theta)\,d\theta$ using the following:

1. Draw samples θ_i, with $i = 1, 2, \cdots, n$ from a proposal density $g(\theta)$.

2. Calculate sample importance weights $q_i = \omega_i / \sum_{j=1}^{n} \omega_j$, where $\omega_i = k(\theta_i)/g(\theta_i)$.

3. Draw samples θ^* from the discrete distribution over $\{\theta_1, \theta_2, \cdots, \theta_n\}$ with probability mass q_i on θ_i.

The drawn samples $\theta^* = \{\theta_1^*, \theta_2^*, \cdots, \theta_N^*\}$ will consequently be approximately distributed following the target density $h(\theta)$. The selection of the proposal density $g(\theta)$ follows from standard importance sampling, in that we typically seek a $g(\theta)$ that covers the support of $h(\theta)$.

Importantly, Bayes' theorem can be used to express the unnormalized density $k(\theta)$ as $k(\theta) = \mathcal{L}(\theta)\,p(\theta)$, where $p(\theta)$ is a prior density and $\mathcal{L}(\theta)$ is the data likelihood. Because $k(\theta_i) = \omega_i\,g(\theta_i)$ in step 2 of SIR, we may equate the prior density $p(\theta)$ with the proposal density $g(\theta)$ to obtain a special case in which the importance weights q_i are the data likelihood, such that $q_i = \mathcal{L}(\theta_i)/\sum_{j=1}^{n} \mathcal{L}(\theta_j)$. As a result, $h(\theta)$ approximates the posterior density following the SIR approach.

In summary, using SIR, we can approximate Bayesian posterior densities of input parameters $\mathbf{x}$ drawn from a prior density. The advantages of using the CNF for Bayesian inference are the following:

- **The resolution of the posterior distribution is not limited by the inherent resolution of the original grid**. Techniques for computing posterior distributions from a grid of models commonly involve sampling or integrating over a volume of the input parameter space discretized by the grid itself, which can heavily depend on the concentration of models within that volume. In contrast, Sections 4.1 and 4.2, demonstrate that the CNF can smoothly interpolate across $\mathbf{x}$, such that a prior distribution $p(\mathbf{x})$ passed through the CNF yields a continuous space of emulated observables that is not limited by the original grid's discretization.

- **There exists a one-to-one mapping between samples from the prior distribution, samples in the observable space, and samples from the posterior distribution**. This provides a direct approach for examining how specific observed data points, including outliers or extreme values, shape the resulting posterior distribution. An example of this mapping is shown in Fig. 8, in which weighting from a data likelihood is applied sequentually to samples drawn from a prior distribution.

In this study, we perform stellar parameter inference using CNF-emulated grids on two tasks. The first uses $\text{CNF}_{\text{giant}}$ to estimate the ages and input parameters of *Kepler* red giants in the open clusters NGC 6791 and 6891. The second task uses $\text{CNF}_{\text{asfgrid}}$ to re-estimate the masses and radii of *Kepler* red giants analyzed by Yu et al. (2018). In each task, we assume that the observables forming the vector $\bar{\mathbf{y}}$ are mutually independent and Gaussian distributed, such that the likelihood $\mathcal{L}$ can be expressed by

$$\mathcal{L} = \prod_{y \in \bar{\mathbf{y}}} \frac{1}{\sigma_{y_{\text{obs}}}\sqrt{2\pi}} \exp \frac{(y_{\text{obs}} - y_{\text{CNF}})^2}{2\sigma^2_{y_{\text{obs}}}}. \quad (2)$$

Alongside the trained CNFs presented in this study, we present the `modelflows`[3] Python package that implements SIR for inferring stellar parameters. Details on its usage can be found at in https://github.com/mtyhon/modelflows.

5.1. *Kepler Open Clusters*

Here, we aim to demonstrate the uncertainties encountered in calculating stellar ages as a consequence of the competing influences across the grid's input parameters on the ages of models. We perform this demonstration by applying $\text{CNF}_{\text{giant}}$ on red giant branch stars within the open clusters NGC 6791 and NGC 6819 observed by the *Kepler* mission. We use broad, uniformly sampled priors for Z, Y, α, $f_{\text{ov,env}}$, and $f_{\text{ov,core}}$, while the distribution for M was drawn from a standard Salpeter initial mass function independent of age and metallicity. We obtain posterior distributions for each star by sampling from the prior distributions, emulating the grid properties using $\text{CNF}_{\text{giant}}$, and applying SIR. Using this approach, samples are drawn until the discrete distribution θ^* over $\{\theta_1, \theta_2, \cdots, \theta_n\}$ contains a minimum of 50,000 non-zero importance weights. Because the CNF maintains a one-to-one mapping between input parameters and output observables (as described in Section 5), we obtain posterior distributions of τ simply by examining the ages of samples forming the posterior distributions for $\bar{\mathbf{x}}$.

For NGC 6791, we adopt observables $\Delta\nu$, ν_{max}, and T_{eff} of the 27 red giant branch stars from the analysis by McKeever et al. (2019) for our likelihood function $\mathcal{L}$. We additionally impose a metallicity constraint for the cluster of $[\text{Fe/H}] = +0.35 \pm 0.04$ dex based on the Open Cluster Chemical Abundances and Mapping Survey (OCCAM, Donor et al. 2020), which presented abundance analyses from APOGEE Data Release 16. For NGC 6819, we use measurements of $\Delta\nu$, ν_{max}, $\delta\nu_{02}$ and T_{eff} for 20 red giant branch stars from the study by Handberg et al. (2017), removing any reported overmassive stars or non-cluster members in the process. We adopt a bulk metallicity constraint of $[\text{Fe/H}] = +0.05 \pm 0.03$ dex from OCCAM.

By combining samples from the posterior distributions across stars from each open cluster, we present the heatmaps describing correlations between Z, Y, α, and age in Fig. 9. We observe the following:

- The strong influence of initial helium abundances Y on stellar ages τ. An increase in Y results in an increase to a star's mean molecular weight, which increases its luminosity and decreases its lifetime along the main sequence. Panels (a1) and (a2) in Fig. 9 show that in the absence of a strong prior in Y, there exists models with a diverse range of initial helium abundances whose output properties may provide reasonable fits to the seismic and spectroscopic parameters of the observed stars. The combined spread in ages for these models spans a range of about 10 Gyr for NGC 6791 and about 2.5 Gyr for NGC 6819 within our grid of models.

[3] https://pypi.org/project/modelflows/

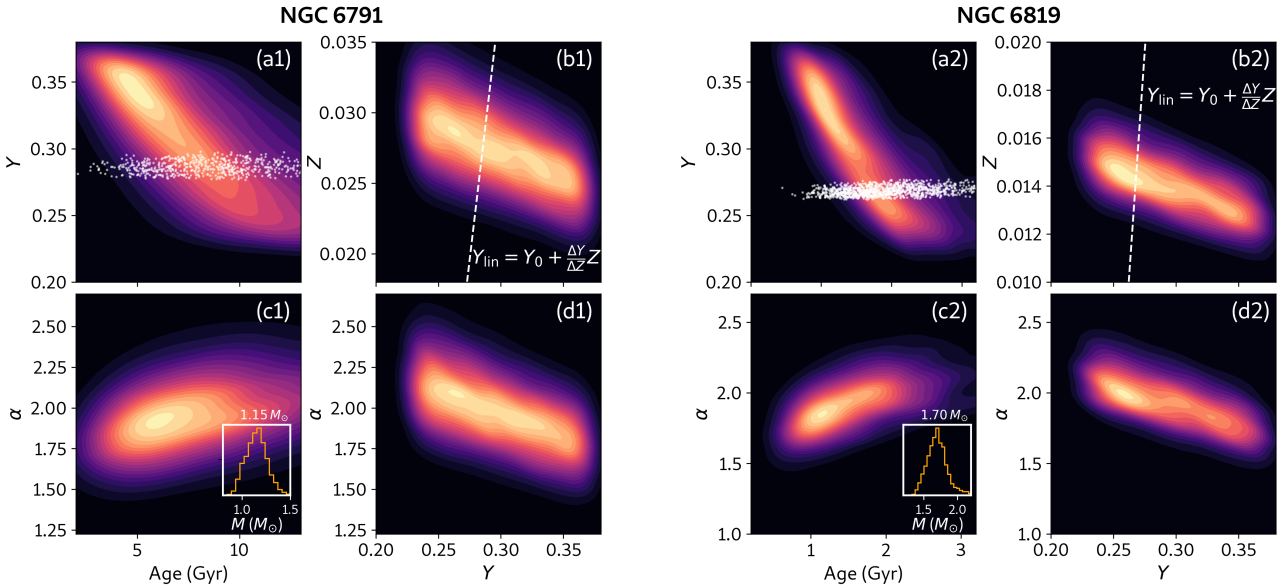

Figure 9. 2D kernel density estimates of stacked posterior distributions of input parameters for red giant branch stars in open clusters NGC 6791 and NGC 6819, as inferred by $\mathrm{CNF}_{\rm giant}$. Insets in panels (c1) and (c2) correspond to the posterior distribution of stellar mass for NGC 6791 and NGC 6819, respectively. The dashed lines in panels (b1) and (b2) shows the region in the $Z - Y$ parameter space explored using a linear helium enrichment relation $Y_{\rm lin}$, here shown with $Y_0 = 0.249$ and $\Delta Y/\Delta Z = 1.33$. White points in (a1) are samples from the joint posterior distribution for NGC 6791 that have $|Y - Y_{\rm lin} < 0.001|$, which demonstrates the age boundaries implicitly assumed when adopting $Y_{\rm lin}$. White samples in (b2) are similar, but for NGC 6819.

- Despite such large variations in age, the combined mass posterior distributions are reasonably well-constrained, with a spread of about $0.12\,M_\odot$ for both clusters. The median mass of $M = 1.15\,M_\odot$ for NGC 6791 here is in good agreement with the reported $M = 1.15 \pm 0.01\,M_\odot$ from McKeever et al. (2019). Meanwhile, the $M = 1.70\,M_\odot$ median mass reported for NGC 6819 here is skewed to a higher mass (and thus younger age) relative to its value of $M = 1.61 \pm 0.02\,M_\odot$ reported from literature (Miglio et al. 2012; Handberg et al. 2017). This difference can be attributed to the additional degrees of freedom in our grid for examining models at the extreme ends of Y and α.

- An expected anti-correlation between Z and Y in panels (b1) and (b2). Given that the ratio Z/X is constrained by the observed bulk metallicity of each cluster, an increase in Y must lead to diminishing Z, with its spread determined by the uncertainty in [Fe/H]. In these panels, we also visualize the linear helium enrichment relation, a commonly used prior for initial helium abundances, and how such a prior reduces the parameter space for stellar ages in panels (a1) and (a2).

- A positive correlation of mixing length α with stellar age in panels (c1) and (c2). This is a trend also observed by other model-fitting studies (e.g., Joyce & Chaboyer 2018b, Li et al. 2024 in review), which emerges as a result of the competing influence between Y and α in the fitting of stellar observables. For instance, the increase in $T_{\rm eff}$ due to an increase in α in a model (c.f. Fig. 5) needs to be balanced with the decrease in $T_{\rm eff}$ from a higher Y, among other observables. This anti-correlation between Y and α — evident in panels (d1) and (d2) — indicates that assuming a constant value of α indirectly sets a prior on Y, and this indirectly influences the age range of models that may be explored from the grid.

To date, there exists a broad range of ages that have been reported for NGC 6791 ($6 - 13$ Gyr) and for NGC 6819 ($1.5 - 3$ Gyr), which partially stems from the wide variety of input parameters priors for grid-based modelling (see e.g., discussions and comparisons by Basu et al. 2011; McKeever et al. 2019; Li et al. 2023). The demonstration in Fig. 9 shows that without enforcing a strong prior across input parameters, the age of a co-eval population determined across its individual members remains highly ambiguous. Rather, a more judicious approach may be to explicitly include a goodness-of-fit metric that penalizes strong differences in input parameters across co-eval members, such as those implemented by Joyce & Chaboyer (2018b).

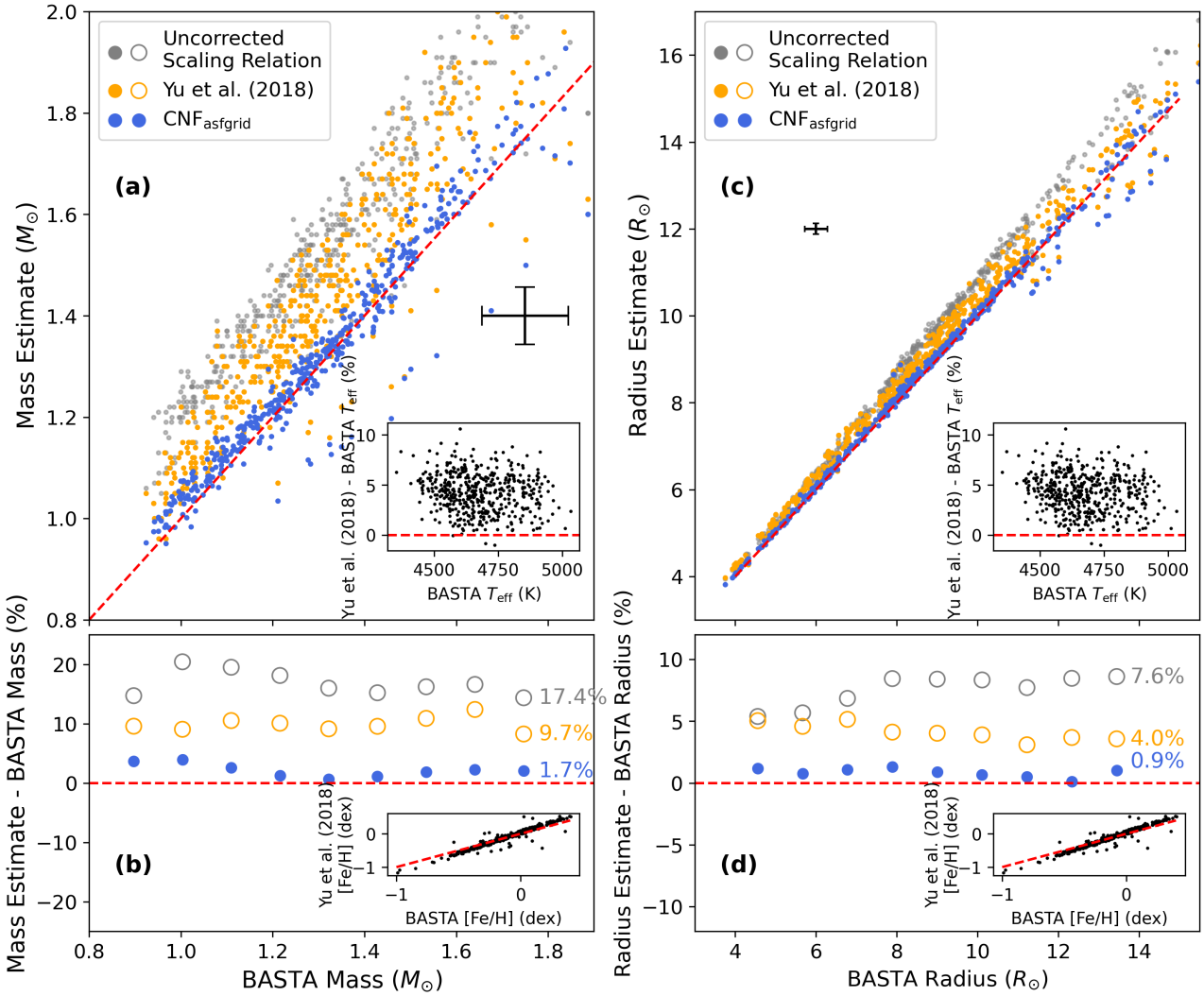

Figure 10. Mass (a-b) and radius (c-d) estimates from CNF$_{asfgrid}$ for 1,199 *Kepler* field red giant branch stars. The one-to-one relation (red dotted line) is shown with respect to results from the BAyesian STellar Algorithm (BASTA) reported by Silva Aguirre et al. (2018). Also shown are mass and radius estimates using the uncorrected scaling relations in gray (Eqn. 3) as well as the values reported from the Yu et al. (2018) catalog. Fractional residual plots are shown in panels (b) and (d), with the median fractional deviation across each estimate source labelled accordingly. Representative errorbars are shown in panels (a) and (c), with inset panels in (a-d) showing differences in spectroscopic parameters $T_{\rm eff}$ and [Fe/H] adopted by the Silva Aguirre et al. (2018) and Yu et al. (2018) studies.

These results are also indicative of the insufficiency of spectroscopic and global asteroseismic observables alone in placing strong constraints on Y, α, and stellar ages. Rather, further independent constraints are needed from cluster-bound eclipsing binaries (Brogaard et al. 2012; McKeever et al. 2019; Brogaard et al. 2021), and also from the measurement of helium abundances using detailed asteroseismic measurements of interior stellar structure (Verma et al. 2019; Nsamba et al. 2021; Verma et al. 2022).

5.2. *Kepler Field Red Giants*

We use CNF$_{asfgrid}$ to estimate the conditional distribution $p(\mathbf{y}|\mathbf{x})$, where $\mathbf{x} = [T_{\rm eff}, \Delta\nu, \nu_{\rm max}, R, \tau]$ and $\mathbf{y} = [M, [{\rm Fe/H}], E]$. Other than M,R, and τ, all values in these vectors adopt measurements reported by Yu et al. (2018) for 15,388 *Kepler* red giants. We obtain posterior distributions for M and R of each star by sampling from CNF$_{asfgrid}$ using a standard Salpeter initial mass function independent of age and metallicity. Similar to Section 5.1, samples are drawn until the discrete distribution θ^* over $\{\theta_1, \theta_2, \cdots, \theta_n\}$ contains a minimum of 50,000 non-zero importance weights. The reported point

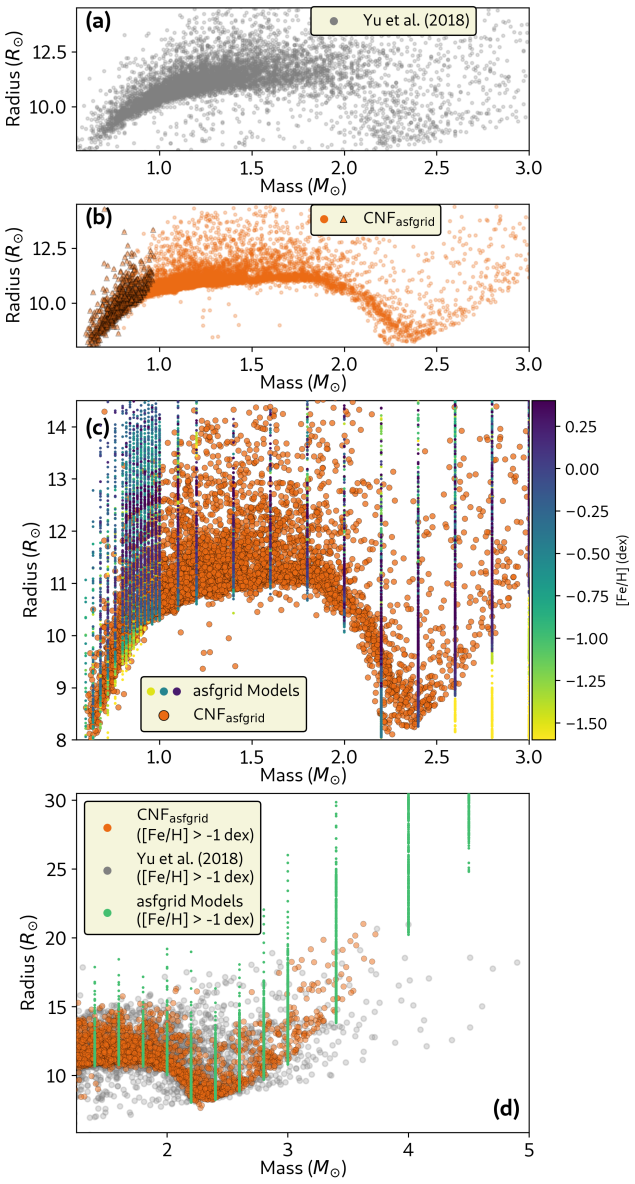

Figure 11. Estimated masses and radii of 7,703 helium core-burning (CHeB) stars from the Yu et al. (2018) dataset of *Kepler* field red giant stars. **(a)** Estimates reported by Yu et al. (2018), which applied AsfGrid's native interpolation method. **(b)** Estimates using the median of the posterior distribution of each star from CNF$_{asfgrid}$. The triangles indicate stars whose model ages are greater than 13.8 Gyr, a heuristic used by Li et al. (2022b) to identify very low-mass red clump stars. **(c)** The CNF$_{asfgrid}$ estimates from panel (b) plotted alongside the CHeB models from the AsfGrid's grid, which are colored by their metallicity [Fe/H]. **(d)** A comparison of estimates between those from CNF$_{asfgrid}$ and from AsfGrid's native interpolation at the high-mass end for stars with [Fe/H]> -1 dex, which comprises 99% of the CHeB stars from the Yu et al. (2018) sample. Also shown are AsfGrid models with the same metallicity range.

Table 3. Estimates of 15,388 *Kepler* red giants from the Yu et al. (2018) catalog based on the emulation of the grid accompanying AsfGrid. Uncertainties are reported as the deviation of the 16th and 84th percentile values from the median. The column 'Ev' refers to the adopted evolutionary stage of the star, where 1 is for hydrogen shell-burning (RGB) red giants, whereas 2 is for helium core-burning (HeB) stars. The full version of this table is available in a machine-readable format in the online journal, with a portion shown here for guidance regarding its form and content.

KIC	Mass ($M_\odot$)	Radius ($R_\odot$)	Ev
757137	$1.417^{+0.123}_{-0.089}$	$12.799^{+0.486}_{-0.247}$	1
892760	$0.959^{+0.066}_{-0.067}$	$10.459^{+0.338}_{-0.308}$	2
893214	$1.500^{+0.075}_{-0.070}$	$11.185^{+0.221}_{-0.216}$	1
1026084	$1.533^{+0.167}_{-0.105}$	$11.204^{+0.481}_{-0.263}$	2
...	...	...	...
12885373	$1.264^{+0.044}_{-0.084}$	$9.116^{+0.163}_{-0.176}$	1
12934574	$1.020^{+0.063}_{-0.065}$	$10.743^{+0.264}_{-0.261}$	2

estimates for M and R are determined from the median of the posterior distribution based on 50,000 draws from θ^*, with uncertainty intervals defined as the interquartile range between the 25th and 75th percentiles. We present our new estimates in Table 3 and highlight the differences of these estimates with previously reported values in the following.

5.2.1. *Red Giant Branch Stars*

We seek to demonstrate that inferring stellar parameters using our flow-based emulation approaches are consistent with grid-based modelling approaches from literature. We compare our mass and radii estimates with those reported by Silva Aguirre et al. (2018) on 1,199 red giant branch stars using the BAyesian STellar Algorithm (BASTA, Aguirre Børsen-Koch et al. 2022), which determines Bayesian posterior distributions using grid-based sampling based on Markov Chain Monte Carlo approaches. The Silva Aguirre et al. (2018) study uses asteroseismic measurements from Yu et al. (2018), but adopts spectroscopic measurements $T_{\rm eff}$ and [Fe/H] from the Data Release 13 of the Sloan Digital Sky Survey (Albareti et al. 2017). We adopt the Yu et al. (2018) reported $T_{\rm eff}$ and [Fe/H] values because we predict M and R over all stars in the Yu et al. (2018) catalogue instead of only a subset that overlaps with the Silva Aguirre et al. (2018) study.

Fig. 10 shows a comparison of mass and radii estimates, which shows that the $\mathrm{CNF}_{\mathrm{asfgrid}}$ results closely match the BASTA values up to a 1-2% offset in mass and a 1% offset in radius. These small discrepancies are expected, given differences in the input spectroscopic values used (Fig. 10 inset). Furthermore, both results are based on different evolutionary codes (BASTA by default uses BaSTI isochrones (Pietrinferni et al. 2004), while AsfGrid uses MESA models), which can lead to slight variations in the output properties predicted across grids (e.g., Silva Aguirre et al. 2020). Besides a comparison with BASTA, we also include a comparison with the values reported directly from the Yu et al. (2018) that are based on the asteroseismic scaling relations:

$$M = \left(\frac{\nu_{\max}}{\nu_{\max,\odot}}\right)^{3}\left(\frac{\Delta\nu}{f_{\Delta\nu}\Delta\nu_{\odot}}\right)^{-4}\left(\frac{T_{\mathrm{eff}}}{T_{\mathrm{eff},\odot}}\right)^{3/2}$$
$$R = \left(\frac{\nu_{\max}}{\nu_{\max,\odot}}\right)\left(\frac{\Delta\nu}{f_{\Delta\nu}\Delta\nu_{\odot}}\right)^{-2}\left(\frac{T_{\mathrm{eff}}}{T_{\mathrm{eff},\odot}}\right)^{1/2}, \quad (3)$$

where $\nu_{\max,\odot} = 3090\mu$Hz, $\Delta\nu_{\odot} = 135.1\mu$Hz, and $T_{\mathrm{eff},\odot} = 5777$ K. By definition, the 'uncorrected' scaling relations have $f_{\Delta\nu} = 1$. Meanwhile, the scaling relations used in the Yu et al. (2018) work applies $f_{\Delta\nu}$ computed by the AsfGrid code, which are determined by an interpolation over ($\Delta\nu$, $\nu_{\max}$, [Fe/H], T_{eff}) within its accompanying grid (Sharma et al. 2016; Stello & Sharma 2022).

Compared to AsfGrid's native interpolated solution, $\mathrm{CNF}_{\mathrm{asfgrid}}$'s result in Fig. 10 shows greater consistency in mass and radii with BASTA (which uses grid-based sampling), as evidenced by its overall smaller fractional deviation and dispersion relative to the BASTA values. We credit this improvement to the direct sampling of mass and radii values from the emulated grid, which circumvents the need to utilize scaling relations (Eqn. 3) and their associated correction factors.

5.2.2. *Red Clump Stars*

We estimate masses and radii for 7,703 red clump stars from the Yu et al. (2018) dataset using $\mathrm{CNF}_{\mathrm{asfgrid}}$, which we present in Fig. 11. In theory, low-mass red clump stars ignite core helium at similar core masses, which manifest as a sharp lower edge in the M-R plane corresponding to the zero-age core helium burning stage (Li et al. 2021). Figs. 11b-c show that the estimates from $\mathrm{CNF}_{\mathrm{asfgrid}}$ are strongly consistent with model predictions of the sharp edge, with smoothly interpolated estimates across regions in the grid for which mass is sparsely sampled ($M > 1.2M_{\odot}$). Beyond the sharp edge, the $\mathrm{CNF}_{\mathrm{asfgrid}}$ estimates for $M \gtrsim 1.8M_{\odot}$ reveal well-defined structures not easily observed in in the Yu

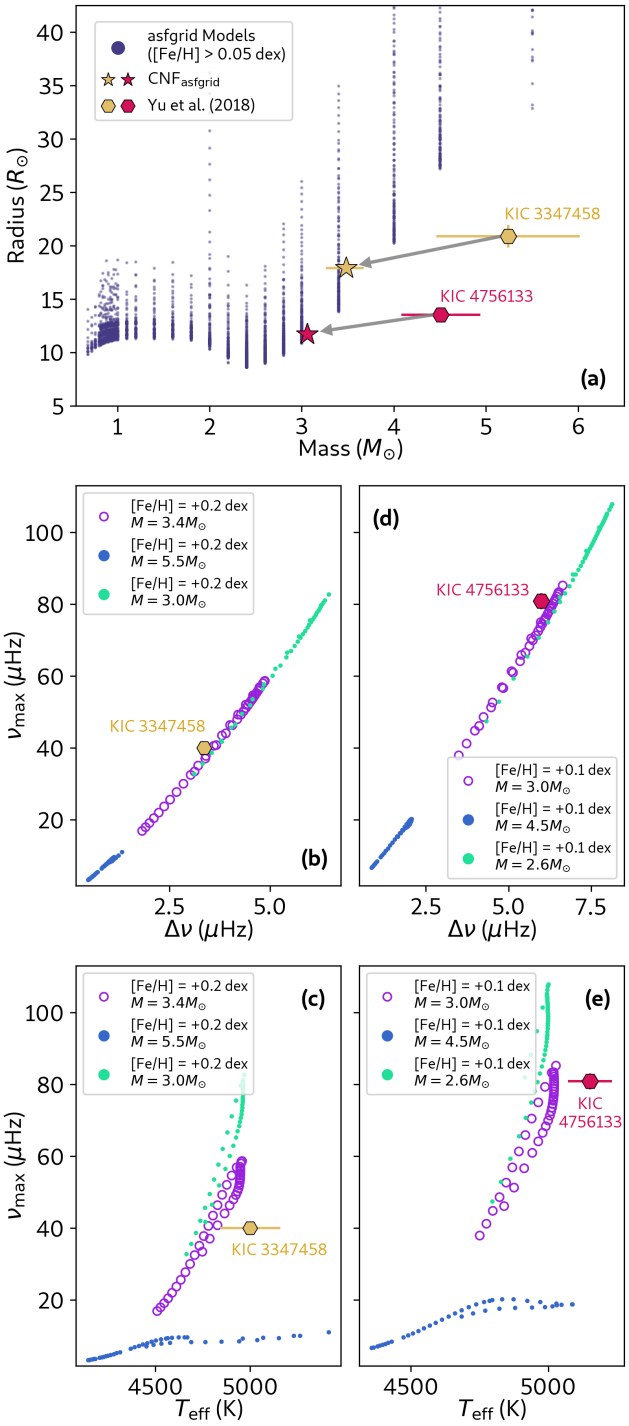

Figure 12. The revision of two high-mass ($M \geq 4.5\,M_{\odot}$) helium core-burning stars to lower masses using $\mathrm{CNF}_{\mathrm{asfgrid}}$. **(a)** Shown are new mass and radius estimates of KIC 3347458 ([Fe/H] $= +0.18 \pm 0.3$ dex) and KIC 4756133 ([Fe/H] $= +0.11 \pm 0.15$ dex) from $\mathrm{CNF}_{\mathrm{asfgrid}}$, which are now within boundaries of the original grid. The grid does not inherently contain any models with $M \geq 4.5\,M_{\odot}$ having $R < 25\,R_{\odot}$. **(b-c)** An examination of models from the grid that have the closest observables ($\nu_{\max}$, $\Delta\nu$, T_{eff}) to those from KIC 3347458. Models from the $M = 3.4\,M_{\odot}$ track form the closest match to the observed star. **(d-e)** The same as panels (b-c), but for KIC 4756133, where models from the $M = 3.0\,M_{\odot}$ track form the closest match across all observables.

et al. (2018) results (Fig. 11a). In particular, red clump stars are predicted to transition to lower luminosities and radii as a consequence of diminishing core masses for stars with total masses at and slightly above the helium flash limit at $M \sim 1.7 - 2.0 M_\odot$ (Girardi 1999). This transition with mass is expected to be gradual yet distinct. Beyond the helium flash limit, helium core-burning stars are predicted to become larger and more luminous from an increase in helium core mass, leading to the presence of the Vertical Substructure (VS) feature observed in Milky Way red clump populations (Girardi 2016). The estimates from $\mathrm{CNF}_{\mathrm{asfgrid}}$ distinctly show these population-level features as demonstrated in Figs. 11c-d, consistent with predictions from models.

These new results suggest revisions to earlier estimates of the masses and radii of *Kepler* helium core-burning stars, especially for those at higher masses. Fig. 11d shows that a large number of stars with a previously estimated $M > 3 M_\odot$ have radii too small relative to the lower boundary of AsfGrid's grid of models in $M - R$ space. According to the grid, such stars are not as massive as previously reported, instead belonging to the VS population representing the transition of helium core-burning stars to higher radii and luminosities. Fig. 12 demonstrates this revision for two *Kepler* red giants, whose previously reported masses of $M > 4.5\,M_\odot$ are now estimated to be within $3.0{-}3.5\,M_\odot$ from $\mathrm{CNF}_{\mathrm{asfgrid}}$. To confirm that these new estimates are indeed consistent with the grid, we present in Figs. 12b-e the observables ($\Delta\nu$, ν_{max}, T_{eff}) of several evolutionary tracks from the grid that are approximately matched in metallicity to the two *Kepler* red giants but are varied in mass. In both examples, we find that tracks having masses similar to the $\mathrm{CNF}_{\mathrm{asfgrid}}$ estimate (lower mass) contain the model forming the closest match of their output properties to the observables of the two red giants in consideration. Meanwhile, tracks having masses similar to the previous estimate (higher mass) do not have any output properties within close proximity to the observables of the two red giants. This result verifies the revised lower-mass estimates for these stars from AsfGrid's grid of models, and at the same time indicates that these stars have predicted mean densities that deviate significantly from those expected by the scaling relations. Note that these results only preclude the existence of actual oscillating helium core-burning stars with $M > 4\,M_\odot$ in the context of AsfGrid's grid of models. Only by further consideration of various input physics from other grids of stellar models alongside more detailed asteroseismic measurements of these stars (e.g., Crawford et al. 2024) can we completely rule out the possibility of very high-mass stars existing in the *Kepler* field.

Table 4. Comparison of the interpolation accuracy of output observables from $\mathrm{CNF}_{\mathrm{giant}}$ versus a simple nearest neighbour (NN) approach. Errors are measured as fractional mean absolute residuals.

Observable	$\mathrm{CNF}_{\mathrm{giant}}$ (%)	NN (%)
T_{eff}	0.009	0.231
$\Delta\nu$	0.106	1.559
$\delta\nu_{01}$	3.092	14.277
$\delta\nu_{02}$	0.339	3.387
$\delta\nu_{03}$	0.262	3.402
ϵ	0.388	6.278
R	0.068	2.367
τ	0.371	20.92

6. INTERPOLATION ACCURACY

Because interpolation is a key functionality of the CNF, we investigate the interpolation accuracy of the CNFs applied in Section 5. Note that the following results are specific to the CNFs and grids published in this work only, and will thus vary when other grids of models and CNF architectures are used.

6.1. *MESA Grid of Giant Star Models*

To validate the accuracy of $\mathrm{CNF}_{\mathrm{giant}}$, we generate a new grid of models with the same input physics and range of input parameters but using a different random number seed to generate its Sobol sequence. This validation grid also has 8,191 unique combinations of input parameters $\mathbf{x} = [M,\ \log_{10} Z,\ Y,\ \alpha,\ \widehat{f_{\mathrm{ov,env}}},\ \widehat{f_{\mathrm{ov,core}}},\ \widehat{\nu_{\mathrm{max}}}]$, but none are identical to the combinations within the grid used to train $\mathrm{CNF}_{\mathrm{giant}}$. For each combination of input parameters from the validation grid, we draw 10,000 samples of output observables from $\mathrm{CNF}_{\mathrm{giant}}$ and compute each observable's median to obtain a point estimate. We quantify the interpolation accuracy based on how much such point estimates deviate from their corresponding ground truth values from the validation grid. The results of this exercise is shown in Fig. 13, which demonstrates that the interpolation error across most of the grid spans only a fraction of a percent. The interpolations, however, are consistently more erroneous at $M \gtrsim 2.0\,M_\odot$ and at $\log_{10} \nu_{\mathrm{max}} \lesssim 0.5$, which corresponds to models at which their speed of evolution is typically more rapid than the rest of the models in the grid. Therefore, the output observables of these rapidly-

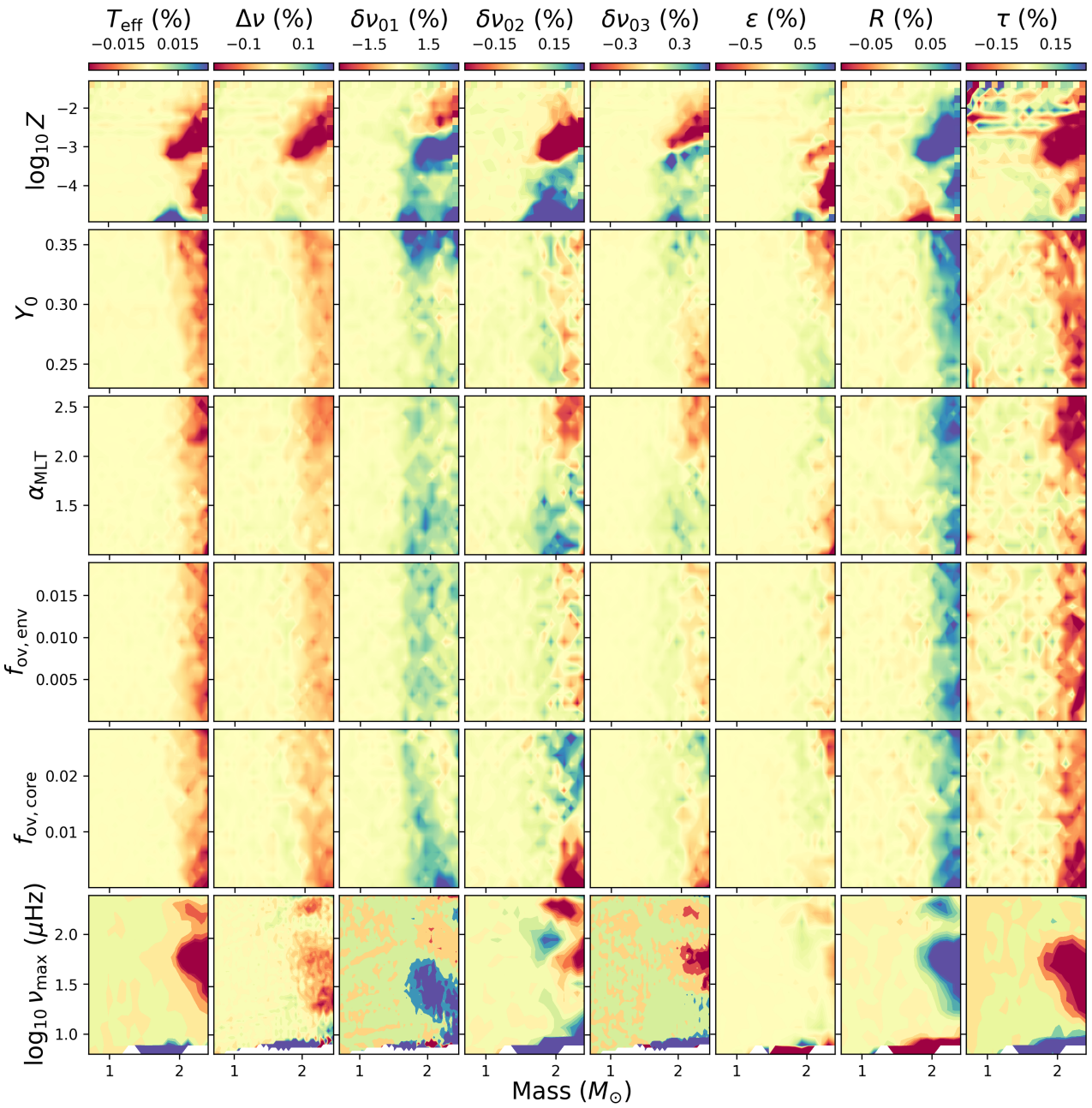

Figure 13. Interpolation errors of output observables from $\text{CNF}_{\text{giant}}$ as a function of input parameters. Each panel corresponds to a 20×20 grid varied in mass and a second input parameter, which varies from row to row. The color within each panel corresponds to the median fractional deviation (true minus predicted) of a specific output observable at that combination of input parameters, with the specific observable varied across columns.

evolving models yield a more 'diffuse' density that contribute less to the likelihood optimization and are thus given lower emphasis during training.

A summary of the interpolation accuracy across the whole validation grid for $\text{CNF}_{\text{giant}}$ is tabulated in Table 4. A notable outlier in performance is $\delta\nu_{01}$, which we attribute to the strong degeneracy in such an observable for giant models in across the grid (c.f. Fig. 2). To benchmark our results against another method that can perform a multi-dimensional interpolation using multi-dimensional input parameters, we also compare the validation accuracy when using a nearest neighbour interpolation using `interpolate.griddata` as implemented in `scipy` (Virtanen et al. 2020). As demonstrated in Table 4, the errors from CNF are about an order of magnitude

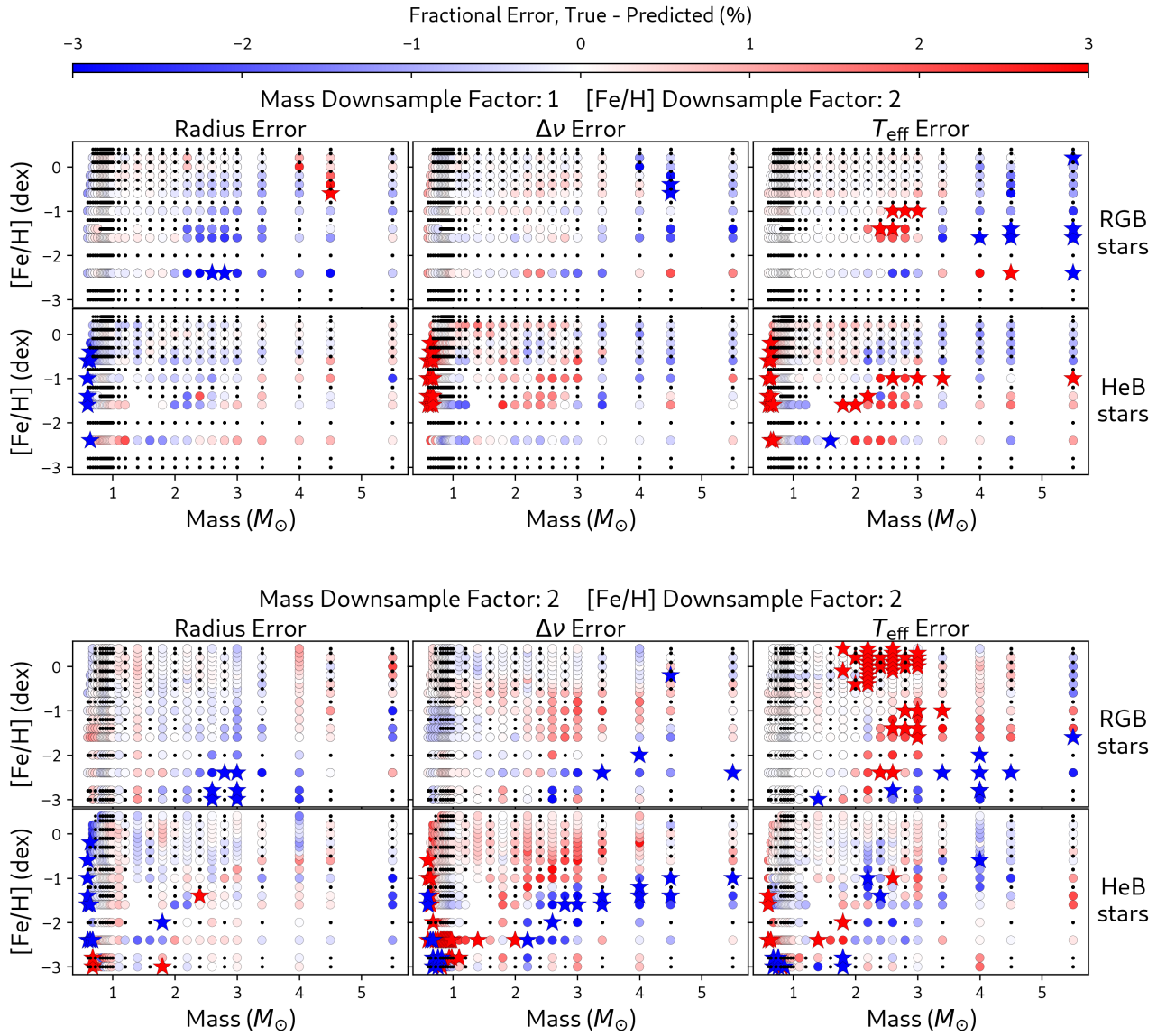

Figure 14. Interpolation errors of output properties [R, $\Delta\nu$, $T_{\rm eff}$] from $\mathrm{CNF}_{\rm asfgrid}$ as a function of input parameters M, [Fe/H], and evolutionary state. The upper series of panels corresponds to the scenario when $\mathrm{CNF}_{\rm asfgrid}$ is trained on a grid downsampled in [Fe/H] by a factor of two, while the lower series of panels is for a training grid downsampled in both mass and [Fe/H] by a factor of two. For each series, models in the hydrogen shell-burning phase (RGB) and in the helium-core burning phase (HeB) of evolution are considered separately. Black points in each panel correspond to combinations of input parameters used for training the CNF in each scenario, while colored circles are combinations that are held out in training and used for validating the CNF's performance. Points highlighted with a star are combinations of input parameters that have a fractional error exceeding 3%.

smaller across all observables compared to the nearest neighbour method.

6.2. *AsfGrid's Evolutionary Models*

We adopt a hold-out strategy to validate $\mathrm{CNF}_{\rm asfgrid}$ by training a new CNF on a downsampled version of the original grid and performing the interpolation on the tracks that were held out. We consider first a scenario in which the metallicity of the original grid is downsampled by a factor of two, then another in which both mass and metallicity are downsampled by a factor of two. By following the same procedures as described in Section 6.1, we visualize in Fig. 14 the interpolation error in $\Delta\nu$, R, and $T_{\rm eff}$ as a function of grid input parameteters, namely evolutionary state, mass, and metallicity. Similar to Section 6.1, the largest interpolation errors often occur for sparse regions within the grid, which belongs to models with high masses ($M > 2.5\,M_\odot$) or those with low metallicity ([Fe/H] < -2.0 dex). The consistently large errors for helium core-burning stars at the

extreme low-mass end ($M = 0.6\,M_\odot$) of the grid is also a consequence of sparsity in the grid, with such evolutionary tracks having as few as 10 models. Meanwhile, the systematically large $T_{\rm eff}$ underestimation for RGB stars with $2.0\,M_\odot < M < 3.0\,M_\odot$ and [Fe/H] > 0.5 dex, following the downsampling of both mass and [Fe/H], is likely the effect of the [Fe/H] downsampling as it disappears when [Fe/H] is not subject to downsampling.

Overall, we find that interpolation errors of R, $\Delta\nu$, and $T_{\rm eff}$ are bounded to within 1 percent within the regions of interest for most of the stars in asteroseismic surveys of the Milky Way, which is $0.7\,M_\odot \lesssim M \lesssim 3.0\,M_\odot$ and [Fe/H] > -2.5 dex (e.g., Pinsonneault et al. 2018; Hon et al. 2021; Theodoridis & Tayar 2023). We remind the reader that these results are strictly a lower bound to the true performance of $\rm CNF_{asfgrid}$, given that the CNFs in this exercise are trained on a downsampled grid that has only between 25-50% of models from the original grid accompanying AsfGrid.

7. CONCLUSION

In this study, we have presented conditional normalizing flows (CNFs) as emulators for grids of stellar evolutionary models. This generative approach is capable of capturing the complex relations between the input parameters $\mathbf{x}$ of a grid with its output properties $\mathbf{y}$. We have trained three CNFs, namely $\rm CNF_{dwarf}$, $\rm CNF_{giant}$, and $\rm CNF_{asfgrid}$, to emulate three grids of models that have multi-dimensional input parameters and predicted output observables. Our main results in this work are the following:

- The CNF can efficiently interpolate within the input parameter space, yielding highly flexible conditional distributions $p_N(\mathbf{y}|\mathbf{x})$ that take the form of evolutionary tracks in the emulated grid, as demonstrated with $\rm CNF_{dwarf}$. Marginalizing across a range of input parameters yields a continuum of evolutionary tracks, which captures the smooth variation of output observables with respect to input parameters that is expected from stellar evolutionary models.

- Being able to accept constraints on both $\mathbf{x}$ and $\mathbf{y}$, the CNF is capable of emulating evolutionary trajectories of models having a fixed output parameter, such as isochrones. The CNF presents an instructive tool for visualizing the relations between input and output parameters within the grid. This concept is illustrated through an interactive demonstration of $\rm CNF_{dwarf}$ that is hosted on an online repository.

- The CNF can be used as a generative prior for Bayesian inference on a grid of models using Sampling/Importance Resampling. The posterior distribution of grid parameters obtained using this approach is not resolution-limited by the original grid, and contains samples having a one-to-one mapping with samples from the prior distribution and with samples in the output parameter space. We presented the `modelflows` package to enable stellar parameter inference through this approach using $\rm CNF_{giant}$ and $\rm CNF_{asfgrid}$.

- Using $\rm CNF_{giant}$, we inferred stellar parameters for red giants in open clusters NGC 6791 and NGC 6819 using measurements from global asteroseismic parameters and spectroscopy. In both cases, large uncertainties in the age of the cluster were demonstrated as a result of unconstrained priors in initial helium abundances and the value of the mixing length parameter.

- We applied $\rm CNF_{asfgrid}$ as an emulator to the grid accompanying the AsfGrid code and inferred revised masses and radii for 15,388 *Kepler* red giant stars from the Yu et al. (2018) sample. Our estimated masses and radii of a subset of red giant branch stars show a robust consistency with BASTA grid-based modelling, with only offsets of $\sim 2\%$ in mass and $\sim 1\%$ in radius. Meanwhile, our estimates on helium core-burning stars recovered distinct characteristics of secondary clump and vertical substructure populations in mass-radius space, consistent with predictions from stellar models. These new estimates indicate an overestimation of previously reported masses above $3.5\,M_\odot$ for helium core-burning stars in the *Kepler* field.

- We benchmarked the CNFs used for stellar parameter inference by comparing the accuracy of their output observables to those from a held-out validation grid. In most regions within the grid, the emulation error incurred is below a fraction of a percent, with exceptions only for extreme cases where models within the original grid are highly sparse or substantially overlapping.

- The trained CNFs, experiments, and interactive visuals in this work are publicly available at https://github.com/mtyhon/modelflows.

The use of generative approaches for statistical inference using simulations is an emerging domain in machine learning. Normalizing flows, especially, have seen

use in simulation-based inference tasks (Tejero-Cantero et al. 2020) and even integration in modern sampling techniques such as Markov Chain Monte Carlo methods (Hoffman et al. 2019). While this study has focused on the use of normalizing flows on grids of evolutionary stellar models, the ability of normalizing flows to interpolate in high dimensions and generative highly expressive conditional distributions is generally valuable across other areas in astronomy in which data can be structured in the form of a high-dimensional grid.

The groundwork for this study was performed at the 'Stellar Astrophysics in the Era of Gaia, Spectroscopic, and Asteroseismic Surveys' workshop at the Munich Institute for Astro-, Particle and BioPhysics (MIAPbP) center held during August 2023.

APPENDIX

A. MASKED AUTOREGRESSIVE FLOW AND NEURAL SPLINE FLOWS

The masked autoregressive flow (Papamakarios et al. 2017) learns joint, multi-dimensional densities by considering them as a product of one-dimensional densities, such that the density of the i-th variable is conditioned on the $i-1$ variables preceding it. These one-dimensional densities are parameterized as single Gaussians:

$$p(x_i|\mathbf{x}_{1:i-1}) = \mathcal{N}(x_i|\mu_i, (\exp\alpha_i)^2), \tag{A1}$$

where $\mu_i = f_\mu(\mathbf{x}_{1:i-1})$ and $\alpha_i = f_\alpha(\mathbf{x}_{1:i-1})$. Here, f_μ and f_α are functions that determine the mean and variance of the i-th conditional density given the $i-1$ variables. Such functions are implemented with neural networks. Sampling from the autoregressive flow proceeds through recursive sampling across i:

$$x_i = z_i \exp\alpha_i + \mu_i, \tag{A2}$$

where z_i is the base distribution, defined as $z_i \sim \mathcal{N}(0,1)$ for one dimensions or a multivariate normal for higher dimensions. By construction, this transformation is easily inverted, such that $z_i = (x_i - \mu_i)\exp(-\alpha_i)$. In addition, the autoregressive structure of this flow yields a triangular Jacobian for the transformation, such that the logarithm of its absolute determinant is easily computed as $\Sigma_i f_{\alpha_i}(\mathbf{x}_{1:i-1})$. In practice, the autoregressive property of one-dimensional densities is gained through the use of binary masks within the computational layers of the neural network. Following the construction of the Masked Autoencoder Distribution Estimator (MADE, Germain et al. 2015), such masks are constructed with a specific ordering to enable individual output nodes of the neural network to predict $p(x_1), p(x_2|x_1), \cdots, p(x_i|\mathbf{x}_{1:i-1})$ within a single pass of data. Therefore, μ_i and α_i across all i can be output simultaneously from a single network, while maintaining the autoregressive property of the flow.

A neural spline flow applies the concept of a coupling transform (Dinh et al. 2017), namely a transformation that maps an input $\mathbf{z}$ to $\mathbf{x}$ by performing the following operations:

1. Split $\mathbf{z}$ into $[\mathbf{z}_{1:i-1}, \mathbf{z}_{i:D}]$, where D is the dimensionality of $\mathbf{z}$.
2. Compute $x_j = g_{\theta_j}(z_j)$ for $j = i, \cdots, D$ in parallel, where $g_{\theta_\mathbf{j}}$ is a function parameterized by $\theta_\mathbf{i}$ that is learned by a neural network.
3. Set $\mathbf{z}_{1:i-1} = \mathbf{x}_{1:i-1}$ and return $\mathbf{x} = [\mathbf{x}_{1:i-1}, \mathbf{x}_{i:D}]$.

Note that following the above approach, vector components $\mathbf{x}_{1:i-1}$ are not directly transformed. A coupling transform satisfies the conditions for a normalizing flow, in that it has a triangular Jacobian matrix, with the determinant as the product of its diagonal elements. Constructing a neural spline flow proceeds by specifying $g_{\theta_\mathbf{i}}$ as a monotonic rational quadratic spline function. Briefly, the function splits the input domain into K bins over which the spline is defined. Here, $\theta_\mathbf{i}$ comprises $= [\theta_\mathbf{i}{}^w, \theta_\mathbf{i}{}^h, \theta_\mathbf{i}{}^d]$, which parameterize the splines by the following:

- $\theta_\mathbf{i}{}^w$ is of length K and determines the width of the bins.
- $\theta_\mathbf{i}{}^h$ is of length K and determines the height of the bins.
- $\theta_\mathbf{i}{}^d$ is of length $K-1$ and determines the derivatives at the internal knots of the spline.

The coupling is combined with an autoregressive transform by defining splines with parameters $\theta_{1:i-1}$ that act element-wise on $\mathbf{x}_{1:i-1}$ (such that this part is no longer not transformed), which are learnable by direct optimization using stochastic gradient descent. Meanwhile, $\theta_{i:D} = \mathrm{NN}(\mathbf{z}_{1:i-1})$, where NN is an autoregressive neural network. Together, these form an Autoregressive Neural Spline Flow (AR-NSF) transform. For further details on the splines, we refer the reader to Durkan et al. (2019).

For $\mathrm{CNF}_{\mathrm{dwarf}}$ and $\mathrm{CNF}_{\mathrm{giant}}$, we use ten AR-NSF transforms with a ten layer multi-layer perceptron with a width of 256 neurons for their neural networks. For $\mathrm{CNF}_{\mathrm{asfgrid}}$, we use eight AR-NSF transforms with a eight layer multi-layer perceptron with a width of 512 neurons for the neural network. We use the default number of spline bins ($K = 8$) implemented by `Zuko` for each transform. The normalizing flows are trained using the Adam optimizer (Kingma & Ba 2015) with learning rate annealing on plateau, reaching convergence in approximately 1,500-2,000 iterations.

B. INTERACTIVE VISUALIZATION

For demonstrative purposes, we develop an interactive plotting widget using Plotly v2.27 (Plotly Technologies Inc. 2015) that showcases the ability of $\mathrm{CNF}_{\mathrm{dwarf}}$ to emulate dwarf and subgiant stellar evolutionary tracks. This source code to run this widget locally is available at https://github.com/mtyhon/modelflows, with details on an online-only version on the same page. A screenshot of the widget is shown in Figure 15, demonstrating samples drawn from $\mathrm{CNF}_{\mathrm{dwarf}}$ emulating models across two evolutionary diagrams simultaneously. The widget has the following functionalities:

- Interpolation across input parameters $\mathbf{x} = [M,\, Z,\, Y,\, \alpha]$ controlled by the user with sliders to provide on-the-fly emulation of models across a range of $\mathbf{x}$.
- Visualization of either the C-D diagram, the H-R diagram, or the $\Delta\nu - \epsilon$ diagram. These three diagrams are currently supported at time of release, with more in future releases. Drawn samples from $\mathrm{CNF}_{\mathrm{dwarf}}$ are persistent across a change of diagram, which is performed using the drop-down menu below the sliders above each panel.
- Highlighting of targets across diagrams by lasso or box select. Samples are colored orange if highlighted in the left panel, while they are colored purple of highlighted in the right panel. The selection of samples in one panel will automatically highlight the same set of samples in the other diagram. This consistency persists even with a change of diagram. The text box at the bottom of each panel displays the age range of each highlighted selection.

REFERENCES

Aguirre Børsen-Koch, V., Rørsted, J. L., Justesen, A. B., et al. 2022, MNRAS, 509, 4344, doi: 10.1093/mnras/stab2911

Albareti, F. D., Allende Prieto, C., Almeida, A., et al. 2017, ApJS, 233, 25, doi: 10.3847/1538-4365/aa8992

An, D., Terndrup, D. M., Pinsonneault, M. H., et al. 2007, ApJ, 655, 233, doi: 10.1086/509653

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481, doi: 10.1146/annurev.astro.46.060407.145222

Bai, Y., Liu, J., Bai, Z., Wang, S., & Fan, D. 2019, AJ, 158, 93, doi: 10.3847/1538-3881/ab3048

Basu, S., Grundahl, F., Stello, D., et al. 2011, ApJL, 729, L10, doi: 10.1088/2041-8205/729/1/L10

Bellinger, E. P., Hekker, S., Angelou, G. C., Stokholm, A., & Basu, S. 2019, A&A, 622, A130, doi: 10.1051/0004-6361/201834461

Bellinger, E. P., Kanbur, S. M., Bhardwaj, A., & Marconi, M. 2020, MNRAS, 491, 4752, doi: 10.1093/mnras/stz3292

Böhm-Vitense, E. 1958, ZA, 46, 108

Brogaard, K., VandenBerg, D. A., Bruntt, H., et al. 2012, A&A, 543, A106, doi: 10.1051/0004-6361/201219196

Brogaard, K., Grundahl, F., Sandquist, E. L., et al. 2021, A&A, 649, A178, doi: 10.1051/0004-6361/202140911

Buchler, J. R., & Yueh, W. R. 1976, ApJ, 210, 440, doi: 10.1086/154847

Cassisi, S., Potekhin, A. Y., Pietrinferni, A., Catelan, M., & Salaris, M. 2007, ApJ, 661, 1094, doi: 10.1086/516819

Christensen-Dalsgaard, J. 1984, in Space Research in Stellar Activity and Variability, ed. A. Mangeney & F. Praderie, 11

Christensen-Dalsgaard, J., Silva Aguirre, V., Elsworth, Y., & Hekker, S. 2014, MNRAS, 445, 3685, doi: 10.1093/mnras/stu2007

Crawford, C. L., Bedding, T. R., Li, Y., et al. 2024, MNRAS, 528, 7397, doi: 10.1093/mnras/stae473

Cyburt, R. H., Amthor, A. M., Ferguson, R., et al. 2010, ApJS, 189, 240, doi: 10.1088/0067-0049/189/1/240

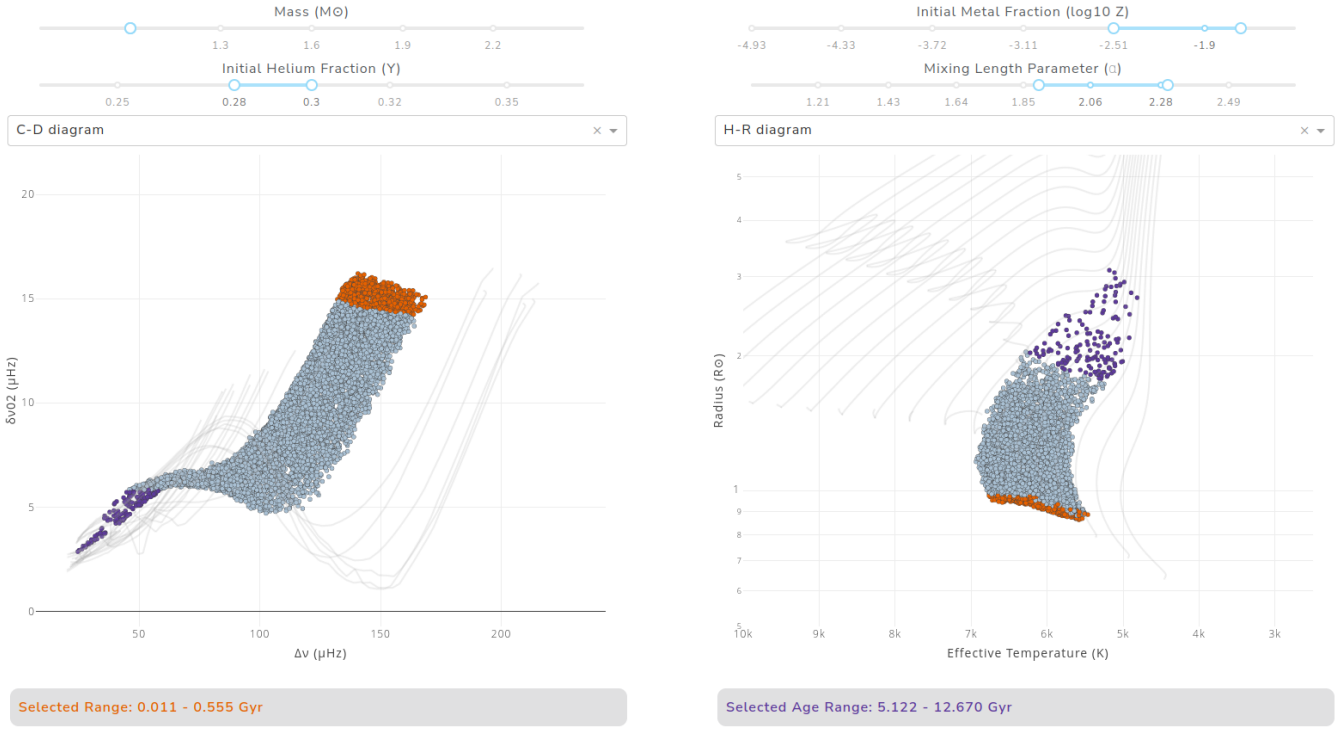

Figure 15. A screenshot of the interactive display tool to visualize evolutionary tracks using $\mathrm{CNF}_{\mathrm{dwarf}}$. The points correspond to on-the-fly interpolated samples drawn from $\mathrm{CNF}_{\mathrm{dwarf}}$, with sliders at the top to vary input parameters. Shown here are emulated models in the C-D diagram and the H-R diagram for $M = 1\,M_{\odot}$, $Y = 0.28 - 0.30$, $\log_{10} Z$=-2.509 − -1.609, $\alpha = 1.9 - 2.3$. The purple and orange points are user-highlighted samples with a one-to-one correspondence between the two panels. The age range of the two groups of highlighted samples are shown at the bottom. Background tracks are from solar metallicity MIST tracks for the H-R diagram, and near-solar metallicity tracks from the MESA grid in this work for asteroseismic diagrams.

Dinh, L., Krueger, D., & Bengio, Y. 2015, in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings, ed. Y. Bengio & Y. LeCun

Dinh, L., Sohl-Dickstein, J., & Bengio, S. 2017, in 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings (OpenReview.net)

Donor, J., Frinchaboy, P. M., Cunha, K., et al. 2020, AJ, 159, 199, doi: 10.3847/1538-3881/ab77bc

Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. 2019, arXiv e-prints, arXiv:1906.04032, doi: 10.48550/arXiv.1906.04032

Eddington, A. S. 1926, The Internal Constitution of the Stars

Ekström, S., Georgy, C., Eggenberger, P., et al. 2012, A&A, 537, A146, doi: 10.1051/0004-6361/201117751

Ferguson, J. W., Alexander, D. R., Allard, F., et al. 2005, ApJ, 623, 585, doi: 10.1086/428642

Garraffo, C., Protopapas, P., Drake, J. J., Becker, I., & Cargile, P. 2021, AJ, 162, 157, doi: 10.3847/1538-3881/ac0ef0

Germain, M., Gregor, K., Murray, I., & Larochelle, H. 2015, in Proceedings of Machine Learning Research, Vol. 37, Proceedings of the 32nd International Conference on Machine Learning, ed. F. Bach & D. Blei (Lille, France: PMLR), 881–889

Girardi, L. 1999, MNRAS, 308, 818, doi: 10.1046/j.1365-8711.1999.02746.x

—. 2016, ARA&A, 54, 95, doi: 10.1146/annurev-astro-081915-023354

Handberg, R., Brogaard, K., Miglio, A., et al. 2017, MNRAS, 472, 979, doi: 10.1093/mnras/stx1929

Henyey, L., Vardya, M. S., & Bodenheimer, P. 1965, ApJ, 142, 841, doi: 10.1086/148357

Herwig, F. 2000, A&A, 360, 952, doi: 10.48550/arXiv.astro-ph/0007139

Hoffman, M., Sountsov, P., Dillon, J. V., et al. 2019, arXiv e-prints, arXiv:1903.03704, doi: 10.48550/arXiv.1903.03704

Hon, M., Bellinger, E. P., Hekker, S., Stello, D., & Kuszlewicz, J. S. 2020, MNRAS, 499, 2445, doi: 10.1093/mnras/staa2853

Hon, M., Huber, D., Kuszlewicz, J. S., et al. 2021, ApJ, 919, 131, doi: 10.3847/1538-4357/ac14b1

Iglesias, C. A., & Rogers, F. J. 1993, ApJ, 412, 752, doi: 10.1086/172958
—. 1996, ApJ, 464, 943, doi: 10.1086/177381
Joyce, M., & Chaboyer, B. 2018a, ApJ, 856, 10, doi: 10.3847/1538-4357/aab200
—. 2018b, ApJ, 864, 99, doi: 10.3847/1538-4357/aad464
Joyce, M., & Tayar, J. 2023, Galaxies, 11, 75, doi: 10.3390/galaxies11030075
Kingma, D. P., & Ba, J. 2015, in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, ed. Y. Bengio & Y. LeCun
Kupka, F., & Muthsam, H. J. 2017, Living Reviews in Computational Astrophysics, 3, 1, doi: 10.1007/s41115-017-0001-9
Li, T., Bedding, T. R., Huber, D., et al. 2018, MNRAS, 475, 981, doi: 10.1093/mnras/stx3079
Li, T., Davies, G. R., Lyttle, A. J., et al. 2022a, MNRAS, 511, 5597, doi: 10.1093/mnras/stac467
Li, Y., Bedding, T. R., Li, T., et al. 2020, MNRAS, 495, 2363, doi: 10.1093/mnras/staa1335
Li, Y., Bedding, T. R., Stello, D., et al. 2021, MNRAS, 501, 3162, doi: 10.1093/mnras/staa3932
Li, Y., Bedding, T. R., Murphy, S. J., et al. 2022b, Nature Astronomy, 6, 673, doi: 10.1038/s41550-022-01648-5
Li, Y., Bedding, T. R., Stello, D., et al. 2023, MNRAS, 523, 916, doi: 10.1093/mnras/stad1445
Lund, M. N., Silva Aguirre, V., Davies, G. R., et al. 2017, ApJ, 835, 172, doi: 10.3847/1538-4357/835/2/172
Lyttle, A. J., Davies, G. R., Li, T., et al. 2021, MNRAS, 505, 2427, doi: 10.1093/mnras/stab1368
McKeever, J. M., Basu, S., & Corsaro, E. 2019, ApJ, 874, 180, doi: 10.3847/1538-4357/ab0c04
Mengel, J. G., Sweigart, A. V., Demarque, P., & Gross, P. G. 1979, ApJS, 40, 733, doi: 10.1086/190603
Miglio, A., Brogaard, K., Stello, D., et al. 2012, MNRAS, 419, 2077, doi: 10.1111/j.1365-2966.2011.19859.x
Mombarg, J. S. G., Van Reeth, T., & Aerts, C. 2021, A&A, 650, A58, doi: 10.1051/0004-6361/202039543
Mowlavi, N., Meynet, G., Maeder, A., Schaerer, D., & Charbonnel, C. 1998, A&A, 335, 573, doi: 10.48550/arXiv.astro-ph/9804155
Nsamba, B., Moedas, N., Campante, T. L., et al. 2021, MNRAS, 500, 54, doi: 10.1093/mnras/staa3228
Ong, J. M. J., & Basu, S. 2019, ApJ, 885, 26, doi: 10.3847/1538-4357/ab425f
—. 2020, ApJ, 898, 127, doi: 10.3847/1538-4357/ab9ffb
Otí Floranes, H., Christensen-Dalsgaard, J., & Thompson, M. J. 2005, MNRAS, 356, 671, doi: 10.1111/j.1365-2966.2004.08487.x
Panda, S. K., Dhanpal, S., Murphy, S. J., Hanasoge, S., & Bedding, T. R. 2024, ApJ, 960, 94, doi: 10.3847/1538-4357/ad0a97
Papamakarios, G., Pavlakou, T., & Murray, I. 2017, arXiv e-prints, arXiv:1705.07057, doi: 10.48550/arXiv.1705.07057
Paxton, B., Bildsten, L., Dotter, A., et al. 2011, ApJS, 192, 3, doi: 10.1088/0067-0049/192/1/3
Paxton, B., Cantiello, M., Arras, P., et al. 2013, ApJS, 208, 4, doi: 10.1088/0067-0049/208/1/4
Paxton, B., Marchant, P., Schwab, J., et al. 2015, ApJS, 220, 15, doi: 10.1088/0067-0049/220/1/15
Paxton, B., Schwab, J., Bauer, E. B., et al. 2018, ApJS, 234, 34, doi: 10.3847/1538-4365/aaa5a8
Paxton, B., Smolec, R., Schwab, J., et al. 2019, ApJS, 243, 10, doi: 10.3847/1538-4365/ab2241
Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, ApJ, 612, 168, doi: 10.1086/422498
Pinsonneault, M. H., Elsworth, Y. P., Tayar, J., et al. 2018, ApJS, 239, 32, doi: 10.3847/1538-4365/aaebfd
Plotly Technologies Inc. 2015, Collaborative data science, Montreal, QC: Plotly Technologies Inc. https://plot.ly
Pols, O. R., Tout, C. A., Eggleton, P. P., & Han, Z. 1995, MNRAS, 274, 964, doi: 10.1093/mnras/274.3.964
Potekhin, A. Y., & Chabrier, G. 2010, Contributions to Plasma Physics, 50, 82, doi: 10.1002/ctpp.201010017
Rogers, F. J., & Nayfonov, A. 2002, ApJ, 576, 1064, doi: 10.1086/341894
Roxburgh, I. W., & Vorontsov, S. V. 2003, A&A, 411, 215, doi: 10.1051/0004-6361:20031318
Rozet, F., Divo, F., & Schnake, S. 2024, probabilists/zuko: Zuko 1.1.0, Zenodo, doi: 10.5281/ZENODO.7625672
Rubin, D. 1988, in Bayesian statistics 3. Proceedings of the third Valencia international meeting, 1-5 June 1987, Clarendon Press, 395–402
Saumon, D., Chabrier, G., & van Horn, H. M. 1995, ApJS, 99, 713, doi: 10.1086/192204
Scutt, O. J., Murphy, S. J., Nielsen, M. B., et al. 2023, MNRAS, 525, 5235, doi: 10.1093/mnras/stad2621
Sharma, S., Stello, D., Bland-Hawthorn, J., Huber, D., & Bedding, T. R. 2016, ApJ, 822, 15, doi: 10.3847/0004-637X/822/1/15
Silva Aguirre, V., Bojsen-Hansen, M., Slumstrup, D., et al. 2018, MNRAS, 475, 5487, doi: 10.1093/mnras/sty150
Silva Aguirre, V., Christensen-Dalsgaard, J., Cassisi, S., et al. 2020, A&A, 635, A164, doi: 10.1051/0004-6361/201935843
Smith, A. F. M., & Gelfand, A. E. 1992, The American Statistician, 46, 84

Soderblom, D. R. 2010, ARA&A, 48, 581, doi: 10.1146/annurev-astro-081309-130806

Stello, D., & Sharma, S. 2022, Research Notes of the American Astronomical Society, 6, 168, doi: 10.3847/2515-5172/ac8b12

Tanner, J. D., Basu, S., & Demarque, P. 2013, ApJ, 767, 78, doi: 10.1088/0004-637X/767/1/78

Tayar, J., Somers, G., Pinsonneault, M. H., et al. 2017, ApJ, 840, 17, doi: 10.3847/1538-4357/aa6a1e

Tejero-Cantero, A., Boelts, J., Deistler, M., et al. 2020, The Journal of Open Source Software, 5, 2505, doi: 10.21105/joss.02505

Theodoridis, A. T., & Tayar, J. 2023, Research Notes of the American Astronomical Society, 7, 148, doi: 10.3847/2515-5172/ace7af

Timmes, F. X., & Swesty, F. D. 2000, ApJS, 126, 501, doi: 10.1086/313304

Ting, Y.-S., & Weinberg, D. H. 2022, ApJ, 927, 209, doi: 10.3847/1538-4357/ac5023

Townsend, R. H. D., & Teitler, S. A. 2013, MNRAS, 435, 3406, doi: 10.1093/mnras/stt1533

Van-Lane, P., Speagle, J. S., & Douglas, S. 2023, arXiv e-prints, arXiv:2307.08753, doi: 10.48550/arXiv.2307.08753

Verma, K., Hanasoge, S., Bhattacharya, J., Antia, H. M., & Krishnamurthi, G. 2016, MNRAS, 461, 4206, doi: 10.1093/mnras/stw1621

Verma, K., Raodeo, K., Basu, S., et al. 2019, MNRAS, 483, 4678, doi: 10.1093/mnras/sty3374

Verma, K., Rørsted, J. L., Serenelli, A. M., et al. 2022, MNRAS, 515, 1492, doi: 10.1093/mnras/stac1860

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2

Wang, B., Leja, J., Villar, V. A., & Speagle, J. S. 2023, ApJL, 952, L10, doi: 10.3847/2041-8213/ace361

White, T. R., Bedding, T. R., Stello, D., et al. 2011, ApJ, 743, 161, doi: 10.1088/0004-637X/743/2/161

White, T. R., Bedding, T. R., Gruberbauer, M., et al. 2012, ApJL, 751, L36, doi: 10.1088/2041-8205/751/2/L36

Winkler, C., Worrall, D., Hoogeboom, E., & Welling, M. 2019, arXiv e-prints, arXiv:1912.00042, doi: 10.48550/arXiv.1912.00042

Yu, J., Huber, D., Bedding, T. R., et al. 2018, ApJS, 236, 42, doi: 10.3847/1538-4365/aaaf74

Zinn, J. C., Stello, D., Elsworth, Y., et al. 2020, ApJS, 251, 23, doi: 10.3847/1538-4365/abbee3

—. 2022, ApJ, 926, 191, doi: 10.3847/1538-4357/ac2c83