Observed Extra Mixing Trends in Red Giants are Reproduced by the Reduced Density Ratio in Thermohaline Zones

ADRIAN E. FRASER,1,2,∗ MERIDITH JOYCE,3,4,2,∗ EVAN H. ANDERS,5,2,∗ JAMIE TAYAR,6,7,8,2,∗ AND MATTEO CANTIELLO9,10,2

1Department of Applied Mathematics, Baskin School of Engineering, University of California, Santa Cruz, CA 95064, USA
2Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106, USA
3Lasker Fellow
4Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
5CIERA, Northwestern University, Evanston IL 60201, USA
6NASA Hubble Fellow
7Department of Astronomy, University of Florida, Bryant Space Science Center, Stadium Road, Gainesville, FL 32611, USA
8Institute for Astronomy, University of Hawai‘i at Mānoa, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
9Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA
10Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA

ABSTRACT

Observations show an almost ubiquitous presence of extra mixing in low-mass upper giant branch stars. The most commonly invoked explanation for this is the thermohaline instability. One dimensional stellar evolution models include prescriptions for thermohaline mixing, but our ability to make direct comparisons between models and observations has thus far been limited. Here, we propose a new framework to facilitate direct comparison: Using carbon to nitrogen measurements from the SDSS-IV APOGEE survey as a probe of mixing and a fluid parameter known as the reduced density ratio from one dimensional stellar evolution programs, we compare the observed amount of extra mixing on the upper giant branch to predicted trends from three-dimensional fluid dynamics simulations. By applying this method, we are able to place empirical constraints on the efficiency of mixing across a range of masses and metallicities. We find that the observed amount of extra mixing is strongly correlated with the reduced density ratio and that trends between reduced density ratio and fundamental stellar parameters are robust across choices for modeling prescription. We show that stars with available mixing data tend to have relatively low density ratios, which should inform the regimes selected for future simulation efforts. Finally, we show that there is increased mixing at low values of the reduced density ratio, which is consistent with current hydrodynamical models of the thermohaline instability. The introduction of this framework sets a new standard for theoretical modeling efforts, as validation for not only the amount of extra mixing, but trends between the degree of extra mixing and fundamental stellar parameters is now possible.

Keywords: stellar evolution, stellar abundances, abundance ratios, stellar interiors, red giant branch, red giant bump, Dredge-up

Corresponding author: Adrian Fraser
adfraser@ucsc.edu
Corresponding author: Meridith Joyce
mjoyce@stsci.edu
Corresponding author: Evan Anders
evanders@northwestern.edu
Corresponding author: Jamie Tayar
jtayar@ufl.edu
∗ These authors contributed equally to this work

1. INTRODUCTION

Observations of globular clusters and low-metallicity field stars show significant changes in the abundance ratios of elements known to be sensitive to mixing, including $^{12}$C/$^{13}$C, lithium, and [C/N], as a star evolves up the red giant branch (Carbon et al. 1982; Pilachowski 1986; Kraft 1994; Shetrone et al. 2019). These changes occur around the red giant branch bump (RGBB) and are largest in the most metal-poor stars (e.g. Gratton...
Large samples of stars that can be used to trace this mixing are now available from a variety of spectroscopic surveys, including GALAH (Buder et al. 2019), APOGEE (Abdurro’uf et al. 2022), and GAIA-ESO (Magrini et al. 2021a). However, observed surface abundance trends are in tension with standard theoretical stellar evolution models, which predict that the surface chemistry should not evolve in this regime.

As low-mass stars ascend the red giant branch, they undergo a series of mixing and homogenizing events as their interior burning and energy transport zones interact. Near the base of the red giant branch, the surface convection zone reaches its deepest level of penetration into the stellar interior, leaving behind a chemical discontinuity from which it recedes in subsequent evolution. This inflection in the convection zone’s movement is known as the “first dredge-up.” The red giant branch bump occurs when the outward-propagating hydrogen burning shell encounters this chemical discontinuity, triggering a structural realignment in which the star’s core contracts and the luminosity drops, causing a disruption to the otherwise monotonic increase in luminosity along the red giant branch (Christensen-Dalsgaard 2015). In one dimensional (1D) stellar models, the sensitivity of the RGBB to physical assumptions makes it a powerful diagnostic of interior mixing processes (e.g. Joyce & Chaboyer 2015; Khan et al. 2018). However, in standard models of red giant stars, there is no mixing between the hydrogen-burning shell and the overlying convective envelope after the first dredge-up, and no change in surface abundances is predicted in this regime. This is in direct conflict with abundance trends found in observations.

The most widely studied candidate mechanism for rectifying this discrepancy is thermohaline mixing, identified in this context by Charbonnel & Zahn (2007) and others. As the hydrogen-burning shell moves into the region chemically homogenized by the first dredge-up, the \(^3\text{He}(^3\text{He}, 2p)^4\text{He}\) reaction creates an inversion of the mean molecular weight \(\mu\). While this \(\mu\) inversion is insufficient to generate a convective region (c.f. Cantiello & Langer 2010), these conditions give rise to the thermohaline instability, a phenomenon perhaps best known in the context of salt water in Earth’s oceans (Stern 1960; Baines & Gill 1969).

Thermohaline mixing is a double-diffusive phenomenon present in fluids that have different diffusivities for heat and chemical composition which in turn make opposing contributions to the vertical density gradient (Turner 1974). Thermohaline mixing occurs in Ledoux-stable regions that have stably stratified temperature gradients but unstable mean molecular weight stratification (see Garaud 2018, for a full review). This process may facilitate the vertical mixing of elements between the hydrogen-burning shell and the stellar convective envelope, thus producing measurable changes in the surface mixing diagnostics after the first dredge-up.

Given that the physical conditions required to trigger the thermohaline instability are in place at around the same time that extra mixing has been observed in red giant stars (e.g. Lagarde et al. 2015), most authors have assumed that all of the observed extra mixing can be attributed to the thermohaline instability (e.g. Kirby et al. 2016; Charbonnel et al. 2020; Magrini et al. 2021b). However, this connection has also been questioned for a number of reasons.

First, reproducing the observed amounts of mixing in this regime with 1D models requires the adoption of much higher efficiency parameters than most fluid simulations would suggest are reasonable or physical (Denissenkov 2010; Denissenkov & Merryfield 2011; Traxler et al. 2011; Brown et al. 2013). Questions have likewise been raised about whether the evolutionary timing of the observed extra mixing is truly consistent with thermohaline models (see e.g. Angelou et al. 2015; Henkel et al. 2017; Tayar & Joyce 2022).

There has also been some debate about how the fluid instability should be parameterized in one dimension, and authors have proposed a variety of different prescriptions informed by numerical simulations (e.g. Traxler et al. 2011; Brown et al. 2013). However, RGB stars have much lower ratios of kinematic viscosity to thermal diffusivity than simulations can reach, of the order \(Pr \sim 10^{-6}\), whereas modern fluid simulations can only probe as low as \(10^{-2} - 10^{-3}\). This has generated skepticism about whether trends from simulations can be accurately extrapolated into stellar regimes. Likewise, models of thermohaline instability that include the presence of a relatively low-amplitude magnetic field can result in much larger diffusivities (Harrington & Garaud 2019), raising the question of whether earlier prescriptions that neglect magnetic fields may be missing key physics.

Given both these observational and theoretical questions, the development of a framework through which we can determine whether signatures from true stellar conditions (observations, \(Pr = 10^{-6}\)) are qualitatively consistent with fluid models is timely and imperative. In this paper, we put forth such a framework: one that allows not only the calibration of individual mixing parameterizations, but also comparison between mixing models. We demonstrate a robust and model-agnostic means of relating the non-dimensional fluid parameters relevant to thermohaline mixing to the observed mix-
ing around the RGB bump and show that this correlation is indeed qualitatively consistent with 1D prescriptions of thermohaline mixing informed by 3D simulations. Further, while previous work (e.g., Charbonnel & Zahn 2007) has used the measurements of the overall amount of extra mixing to tune the overall efficiency of thermohaline mixing prescriptions, our framework allows us to use trends in extra mixing as a function of fundamental stellar parameters to probe trends predicted by various prescriptions.

This paper is organized as follows: we begin by summarizing the formalism and stellar structure quantities relevant to thermohaline mixing (Sec. 2). This is followed by a description of various 1D mixing prescriptions commonly adopted in stellar evolution calculations (Sec. 3). We then introduce a suite of 1D MESA simulations and calculate the relevant fluid parameters in the thermohaline region for a range of mass and metallicity assumptions (Secs. 4 and 5). Finally, we compare an observational proxy of extra mixing, the decrease in [C/N], by various prescriptions.

The instability driving thermohaline mixing requires a Ledoux-stable inversion of the mean molecular weight $\mu$ stratification in the presence of a stable temperature gradient. The stability of the temperature gradient is given by the Schwarzschild criterion:

$$\nabla_{\text{rad}} - \nabla_{\text{ad}} < 0,$$

where the temperature gradient $\nabla \equiv d\ln P/d\ln T$ (pressure $P$ and temperature $T$) has an adiabatic value $\nabla = \nabla_{\text{ad}}$ and saturates to $\nabla = \nabla_{\text{rad}}$ in hydrostatically stable regions where the flux is carried radiatively. The Ledoux criterion for convective stability is (Ledoux 1947)

$$\nabla_{\text{rad}} - \nabla_{\text{ad}} - \frac{\phi}{\delta} \nabla_{\mu} < 0,$$

which must be satisfied despite the inversion of the mean molecular weight. The Ledoux criterion accounts for the composition gradient $\nabla_{\mu} = d\ln \mu/d\ln P$, where $\delta = -(\partial \ln \rho/\partial \ln T)_{P,\mu}$ and $\phi = (\partial \ln \rho/\partial \ln \mu)_{P,T}$ (where $\rho$ is density).

The stabilizing influence of the temperature gradient relative to the destabilizing influence of the $\mu$ gradient is measured by the density ratio,

$$R_0 \equiv \nabla - \nabla_{\text{ad}} \frac{\phi}{\delta} \nabla_{\mu},$$

where $R_0 < 1$ implies the $\mu$ gradient is sufficiently unstable to drive convection, and $R_0 > 1$ implies the fluid is stably-stratified (i.e., no convection). As reviewed by Garaud (2018), fluids with $R_0 > 1$ can be prone to double-diffusive instabilities whenever the thermal diffusivity, $\kappa_T$, is greater than the compositional diffusivity, $\kappa_\mu$. Specifically, the instability driving thermohaline mixing acts whenever

$$1 < R_0 < 1/\tau,$$

(Baines & Gill 1969) where

$$\tau \equiv \kappa_\mu / \kappa_T.$$ 

Note that in stellar radiation zones, typically $\tau \lesssim 10^{-6}$. This means that very slight inversions of $\mu$ (large $R_0$) can drive thermohaline mixing, even when the temperature gradient is strongly stable according to the Schwarzschild criterion.

Throughout this paper, we express the density ratio $R_0$ in terms of the reduced density ratio

$$r \equiv \frac{R_0 - 1}{\tau - 1},$$

per, e.g., Traxler et al. (2011); Brown et al. (2013). In terms of $r$, the condition for thermohaline instability, Eq. (4), is

$$0 < r < 1,$$

where $r \leq 0$ is the threshold for convection and $r \geq 1$ corresponds to scenarios where the $\mu$ inversion is too weak to drive the thermohaline instability.

3. PARAMETERIZED THERMOHALINE MODELS

Equation (7) can be readily evaluated at any radial location in a model star generated with a 1D stellar

---

1 Note that $R_0 > 1$ is equivalent to the Ledoux criterion Eq. (2) only if $\nabla = \nabla_{\text{rad}}$. Thermohaline mixing primarily mixes chemicals, but does produce some minimal thermal mixing (see, e.g., Fig. 4 of Brown et al. 2013); thus, $\nabla \neq \nabla_{\text{rad}}$. This thermal mixing is often ignored in mixing prescriptions in 1D stellar evolution programs, however.
structure and evolution program. However, predicting the efficiency of thermohaline mixing is much more challenging. A diffusive approximation is commonly taken for the turbulent mixing of chemicals such that the total mixing of chemical species is given by the sum of the molecular diffusivity and a turbulent mixing coefficient, $D_{th}$. This quantity can be converted to the compositional Nusselt number discussed in the fluid dynamics literature, $\text{Nu}_\mu$, using the formula

$$D_{th} = (\text{Nu}_\mu - 1) \kappa_\mu. \quad (8)$$

We call any model that predicts $D_{th}$ as a function of stellar structure variables (e.g. gradients and molecular diffusivities of chemicals and heat) a parameterized mixing model or mixing prescription. Efforts to develop thermohaline mixing prescriptions for use in models of stellar interiors date back many decades, see Garaud (2018) for a full review. Such mixing prescriptions have been implemented in a variety of 1D stellar evolution programs (see Lattanzio et al. 2015, and references therein), enabling studies of the effects of thermohaline mixing in stars across the Hertzsprung-Russell diagram.

Here, we briefly summarize the most commonly used mixing models or mixing prescriptions. Efforts to develop thermohaline mixing theories have been made to refine these mixing prescriptions by performing numerical experiments with multidimensional simulations to more accurately parameterize mixing efficiency (Denissenkov 2010; Traxler et al. 2011). Traxler et al. (2011) and Brown et al. (2013) performed 3D hydrodynamic simulations across a broad range of parameters. Not only did they find orders of magnitude less mixing than what is predicted by the Kippenhahn model with the efficiency parameter required in Charbonnel & Zahn (2007) to find agreement with observations ($C_t = 1000$), they also developed new mixing prescriptions that fit their simulations much more closely. In the case of Traxler et al. (2011), the authors derived a mixing law by fitting an analytic function of the form

$$D_{th} = \kappa_\mu \left( \frac{Pr}{\tau} \right)^{1/3} \left( ae^{-br} [1 - r]^c \right), \quad (11)$$

to their simulation results, where

$$Pr = \frac{\nu}{\kappa_T} \quad (12)$$

is the Prandtl number, with $\nu$ the kinematic viscosity, and $a$, $b$, and $c$ are constants which they fit to data.

While Traxler et al. (2011) clearly showed their simulations are inconsistent with the mixing efficiency implied by the Kippenhahn model with $\alpha_{th}, C_t \sim 10^2 - 10^3$, it is important to note that their simulations generally explored $Pr, \tau \sim 10^{-1}$, whereas these fluid parameters are generally of the order $10^{-6}$ in the radiative interiors of RGB stars. Thus, a fair question is whether mixing efficiency might increase to these larger values as $Pr$ and $\tau$ approach $10^{-6}$. However, Traxler et al. (2011) varied these parameters by an order of magnitude in their simulations, and investigated trends in $D_{th}$. They found that mixing should not increase in this fashion, as indicated by the dependence of $D_{th}$ on $Pr$ and $\tau$ in Eq. (11), which makes an argument that these models can be made to fit the observational data difficult to justify.

Brown et al. (2013) note that the model in Eqn. (11) performs well at high $R_0$, but underestimates mixing at $r = 10^{-6}$, which is a representative value for the thermohaline-unstable region of RGB stars, and the same $\alpha_{th} = 2$ assumed in Cantiello & Langer (2010).

In addition to tension regarding the choice of overall mixing efficiency (see e.g. Ulrich 1972; Kippenhahn et al. 1980; Charbonnel & Zahn 2007; Cantiello & Langer 2010; Traxler et al. 2011, for further discussion), there have also been questions about the appropriate trends in the amount of mixing as a function of fluid parameters (e.g. $R_0$ and $\tau$) and therefore the stellar structure variables on which they depend (Garaud 2018). Thus, recent work has sought to refine these mixing prescriptions by performing numerical experiments with multidimensional simulations to more accurately parameterize mixing efficiency (Denissenkov 2010; Traxler et al. 2011).
low $R_0$, particularly in the stellar regime of low $Pr$ and $\tau$. They develop a semi-analytical model for thermohaline mixing,

$$D_{th} = \kappa_\mu C^2 \frac{\lambda^2}{\tau l^2 (\lambda + \tau l^2)}, \quad (13)$$

where $\lambda$ is the growth rate of the fastest-growing linearly unstable mode, $l$ is its horizontal wavenumber, and $C \approx 7$ was fit to data from 3D hydrodynamic simulations. Both $\lambda$ and $l$ are functions of $Pr$, $\tau$, and $R_0$, and are obtained by finding the roots of a cubic and quadratic polynomial (their Eqs. 19 and 20). The orange curve in Fig. 1 shows $D_{th}/\kappa_\mu$ vs. $r$ calculated according to Eq. (13) for $Pr = \tau = 10^{-6}$, representative values for the thermohaline-unstable regions of RGB stars. Note that $D_{th}/\kappa_\mu \rightarrow 0$ as $r \rightarrow 1$ as expected, since the thermohaline instability becomes stable for $r \geq 1$. We see that Eq. (10) with $\alpha_{th} = 2$ agrees with this prescription for some values of $r$, suggesting that significantly larger values of $\alpha_{th}$ are not consistent with 3D hydrodynamic simulations. While the general dependence of $D_{th}/\kappa_\mu$ on $r$ is significantly different between these two models, they do both feature monotonically decreasing values of $D_{th}/\kappa_\mu$ versus $r$. This prescription is implemented in MESA and has since been used in Bauer & Bildsten (2019) and other works.

Harrington & Garaud (2019) extended the work of Brown et al. (2013) by performing 3D magnetohydrodynamic (MHD) simulations of thermohaline mixing with initially uniform, vertical magnetic fields of various strengths to approximate the effects of magnetic fields from external processes including, for instance, a global dipole field or a large-scale magnetic field left behind by a dynamo acting in the receding convective envelope. They found that magnetism strictly increases mixing efficiency, sometimes dramatically. They developed a mixing prescription that accounts for this effect by building on the model of Brown et al. (2013). The strength of the magnetic field is introduced into their model through their parameter $H_B$, which is proportional to the square of the magnetic field strength and depends on other stellar structure variables (see Eq. 19 of Harrington & Garaud 2019). Their mixing prescription is of the form

$$D_{th} = \kappa_\mu K_B \frac{w_f^2}{\tau (\lambda + \tau l^2)}, \quad (14)$$

where $w_f$ is obtained by solving a quartic polynomial that includes the magnetic field strength through $H_B$, and $K_B \approx 1.24$ is directly related to the constant $C$ in Eq. (13).

This mixing prescription agreed remarkably well with their 3D simulations, which were limited to $r = 0.05$ but ranged in magnetic field strength over several orders of magnitude. The prescription, which has not yet been implemented in MESA at the time of this writing, has two asymptotic limits, one where $w_f^2 \propto B_0^2$ when the magnetic field strength $B_0$ is large, and one which reduces to the model of Brown et al. (2013) when $B_0$ is small. We note that the threshold $B_0$ separating these limits depends on the same stellar structure variables that enter into $r$; thus, a given magnetic field strength might simultaneously be considered “large” in one region of a star and “small” in another, implying that simplifications of this model may not be simultaneously appropriate for all regions within a single star.

The purple curve in Fig. 1 shows $D_{th}/\kappa_\mu$ vs. $r$ calculated according to Eq. (14) for the same parameter choices as the orange curve, and with $H_B = 10^{-6}$, appropriate for the thermohaline zone of a 1.1 $M_\odot$ star at $[\text{Fe/H}] = -0.2$ and a magnetic field whose strength is $\mathcal{O}(100 \text{ G})$. Note that this magnetic field strength dramatically increases mixing efficiency relative to the hydrodynamic values, particularly at larger values of $r$, whereas the model predicts the same mixing as the Brown model for $r \lesssim 10^{-5}$. For larger values of $r$, the dependence of $D_{th}/\kappa_\mu$ on $r$ is profoundly different than either of the hydrodynamic models, with $D_{th}/\kappa_\mu$ increasing with $r$, even as the thermohaline instability approaches marginal stability as $r \rightarrow 1$.

Given the variance of the predictions of these models, we focus in this paper on showing how observations can be used to suggest the trends in mixing that models should hope to explain rather than on trying to calibrate the overall mixing efficiency for a particular model, as has been done before, in order to provide a framework in which we can distinguish between mixing prescriptions.

4. STELLAR EVOLUTION MODELS

We use MESA stable release version 21.12.21 to conduct 1D numerical simulations of stars incorporating the effects of thermohaline mixing for metallicities ranging from $[\text{Fe/H}] = -1.4$ to 0.4 ($Z = 0.00068$ to 0.038) and masses from 0.9 to 1.7 $M_\odot$ at resolutions of 0.2 dex and 0.2 $M_\odot$, respectively. We adopt the solar abundance scale of Grevesse & Sauval (1998) and the corresponding opacities of Iglesias & Rogers (1996). We use an Eddington T-σ relation for the atmospheric surface boundary conditions. We adopt the mixing length theory (MLT) prescription of Cox (1980) with a fixed value of $\alpha_{\text{MLT}} = 1.6$ times the pressure scale height ($H_p$). We use the Ledoux criterion for convective stability and neglect the effects of convective overshoot (Ledoux 1947).
Figure 1. Prescriptions of the compositional diffusivity due to thermohaline mixing $D_{th}$ normalized by the molecular diffusivity $\kappa_{\mu}$, are plotted against the reduced density ratio $r$. For each prescription, we use $Pr = \tau = 10^{-6}$, consistent with the conditions in these regions of RGB stars. We plot two hydrodynamic models, the Brown et al. (2013) model (orange) and the Kippenhahn et al. (1980) model with $\alpha_{th} = 2$ (green). In both cases, the mixing efficiency decreases with $r$. The Harrington & Garaud (2019) model (HG19) is also shown; it includes magnetic fields, which cause mixing efficiency to increase with $r$ for these parameters. The plotted curve for the HG19 model depends on $H_B$, which depends on the stellar structure and magnetic field strength; the plotted value is characteristic of the structure in the thermohaline zone of a 1.1 $M_\odot$ star at $[\text{Fe/H}] = -0.2$ with a magnetic field whose strength is $O(100\, \text{G})$. The purple-to-yellow color gradient plotted in the background denotes the range of $r$ values that we measure in our grid of 1D stellar evolution models, which are displayed in Fig. 2.

We use the pp_extras.net nuclear reaction network, which contains 12 isotopes. More details are available in Appendix A, and the exact configuration of our physical and numerical parameter choices is available on Zenodo\(^2\).

Simulations are evolved at $1.25\times$ the default mesh (structural) resolution and $2\times$ the default time resolution on the pre-main sequence and main sequence. Once the models ascend the red giant branch and reach a surface gravity $\log g \leq 3$, resolutions are increased to $2\times$ the default spatial resolution and $10\times$ the default temporal resolution, respectively. Optimal resolution values were determined according to the convergence tests detailed in Appendix B.

We study four grids of stellar evolution simulations with different thermohaline mixing prescriptions. One grid employs the Brown et al. (2013) prescription\(^3\), while the other three employ the Kippenhahn et al. (1980) prescription with coefficients $\alpha_{th} \in [0.1, 2, 700]$. The Kippenhahn $\alpha_{th} = 0.1$ model has inefficient mixing and represents a regime where thermohaline mixing is present but weak. We study $\alpha_{th} = 2$ because (1) it is consistent with the default implementation in MESA and discussion in the instrument paper (Paxton et al. 2013); (2) it is used in previous work (Cantiello & Langer 2010; Tayar & Joyce 2022); and (3) it is consistent with findings from 2D and 3D hydrodynamical simulations in stellar regimes (Denissenkov et al. 2010; Traxler et al. 2011; Brown et al. 2013). We also study a more traditional $\alpha_{th} = 700$, which is of the order of literature values calibrated using stellar observations (Lattanzio et al. 2015; Charbonnel & Zahn 2007).

These three choices for $\alpha_{th}$ in the Kippenhahn prescription also correspond to three different relationships between the timescale over which thermohaline mixing acts, $t_{th}$, and the evolutionary timescale of the star, $t_{evol}$. We take the thermohaline timescale to be the diffusive timescale associated with thermohaline mixing, $t_{th} = d^2/D_{th}$, where $d$ is the radial depth of the thermohaline zone. In the case where $\alpha_{th} = 0.1$, $t_{evol} \ll t_{th}$, meaning the timescale for homogenization of the mixing zone is large. In the case where $\alpha_{th} = 700$, $t_{evol} \gg t_{th}$ and the homogenization timescale is short. In the intermediate case ($\alpha_{th} = 2$), the timescales are comparable; this is likewise true for the assumptions made in the Brown model, though their prescription does not involve $\alpha_{th}$. Given the range of mixing timescales probed, these models should confer some insight as to how (or whether) the inferred fluid parameters, including the reduced density ratio $r$, depend on both the input mixing timescale and the stellar parameters.

4.1. Method for Extracting $r$

We wish to extract a parameter that characterizes the instability of the mean molecular weight gradient to mixing in the thermohaline region above the burning shell. This can be measured by the reduced density ratio $r$. We define a selection criterion that averages over many mass shells and many evolutionary time steps to ensure that our measured values of $r$ are representative of thermohaline mixing during the relevant evolutionary regime.

To measure $r$ in our MESA simulations, we first restrict to the appropriate evolutionary phase. We ex-

\(^2\) MESA inlists will be made available upon publication.

\(^3\) Although the Brown prescription contains no free parameters in their original conception, a multiplicative factor on $D_{thorn}$ has been introduced in the MESA implementation. Brown’s model is reproduced by assigning the quantity $\text{thermohaline\_coeff} = 1$, as was done here.
clude all models for which MESA does not detect thermohaline mixing within \( m_{\text{max}} \leq m_i < 1.1 m_{\text{max}} \), where \( m_i \) is the mass coordinate of the \( i \)th mass shell and \( m_{\text{max}} \) is the mass coordinate coinciding with the instantaneous peak of the nuclear energy generation. The thermohaline zone extends from a maximum mass coordinate \( m_{\text{heavy}} \) to a minimum mass coordinate \( m_{\text{light}} \) with stratification \( \Delta m = m_{\text{heavy}} - m_{\text{light}} \). We exclude the first 21 models in which the thermohaline zone spans at least 10 mass shells. We then compute the evolution of \( \Delta m \) of the \( j \)th model, \( \delta m^j = \Delta m^j - \Delta m^{j-1} \) for model \( j \) and the previous 20 models and compute \( \langle \delta m \rangle = (1/20) \sum_{j=20}^{0} \delta m^j \) to determine whether the mass of the thermohaline region has evolved appreciably over the past several timesteps. We expect \( \langle \delta m \rangle \) to be relatively large when the thermohaline zone is developing and small when it is in a relatively steady state. We then measure \( \epsilon = |\langle \delta m \rangle / \max(\Delta m^j)| \). If \( \epsilon < 5 \times 10^{-3} \), we consider the model to have reached a steady state of thermohaline mixing (classified as “good” or “stable”), at which point we compute \( r \).

To compute the reduced density ratio \( r = (R_0 - 1)/(\tau^{-1} - 1) \), we take the volume average \( \bar{r} = \sum r_i dV_i / \sum dV_i \) over a subset of mass bins \( i \) of the thermohaline zone. We volume-average the reduced density ratio \( r \) over the mass range bounded by \( m_{\text{heavy}} + 0.1 \Delta m < m_i \leq m_{\text{heavy}} + (0.1 + 1/3) \Delta m \). In the volume average, we set the volume element \( dV_i = 4\pi r_i^2 \Delta r_i \) and perform integration using the composite trapezoidal rule as implemented in NumPy (vander Walt et al. 2011). We stop extracting \( r \) after we have collected measurements over 1000 models, which captures the behavior of the saturated thermohaline zone and its eventual merging with the convective envelope. For each stellar evolution simulation, we report the median of the volume-averaged \( r \) over all of the stable models in which measurements were taken. Results are discussed in terms of the logarithm of this quantity, \( \log_{10} r \).

A movie demonstrating the evolution of a thermohaline front and the reduced density ratio selection algorithm is available in Appendix C.

5. RESULTS FROM NUMERICAL EXPERIMENTS

Figure 2 compares results from four physical configurations describing thermohaline mixing in MESA: the upper left panel shows results from the Brown model; the remaining three show results from the Kippenhahn prescription with \( \alpha_{\text{th}} \) varying as indicated. The reduced density ratio \( \log_{10} r \) is shown as a function of mass and metallicity and indicated on the color bar and grid labels.

In all cases, the most notable trend is that \( \log_{10} r \) decreases along the diagonal from high masses and metallicities (upper left) to low masses and metallicities (lower right). There is particularly high qualitative similarity between the Brown model and Kippenhahn model with \( \alpha_{\text{th}} = 2 \), which correspond to similar thermohaline mixing timescales. The case with the lowest mixing parameterization is the Kippenhahn \( \alpha_{\text{th}} = 0.1 \) case, and there the span of \( \log_{10} r \) values is smallest. We also note that, unlike in the other three cases, \( \log_{10} r \) does not scale precisely monotonically with either mass or \( [\text{Fe}/\text{H}] \) in the Kippenhahn \( \alpha_{\text{th}} = 700 \) case. While there is no clear relationship between the spread of \( \log_{10} r \) values observed when using the Kippenhahn prescriptions and the values of \( \alpha_{\text{th}} \) adopted in each, there is a clear relationship between the median values of \( \log_{10} r \) and \( \alpha_{\text{th}} \): the reduced density ratios are larger when mixing is highly efficient (i.e. \( t_{\text{th}} \ll t_{\text{vol}} \)). Most importantly, the overall behavior of \( \log_{10} r \) as a function of mass and \( [\text{Fe}/\text{H}] \) is consistent regardless of the theoretical assumption adopted. This robustness across 1D thermohaline mixing model assumptions suggests that \( r \) may be useful as a mixing diagnostic in physical data sets. We explore its application to observations subsequently.

6. OBSERVED MIXING SIGNATURES

As discussed in Section 3, we are trying to distinguish between different models of thermohaline mixing based not on the amount of mixing they predict, but on their trends as a function of the reduced density ratio \( r \). As discussed in Section 5, this requires stars of a wide range of masses and metallicities. It is therefore quite convenient that modern spectroscopic surveys have recently begun collecting measurements of mixing diagnostics for large samples of stars whose masses are also well constrained. We choose for this work to use the carbon-to-nitrogen ratios, \([\text{C}/\text{N}]\), measured from the Apache Point Galactic Evolution Experiment (APOGEE, Majewski et al. 2017). APOGEE is a Sloan Digital Sky Survey III and IV (Blanton et al. 2017) project using the 2.5 meter Sloan Telescope (Gunn et al. 2006) and the APOGEE spectrograph (Wilson et al. 2019) to obtain medium resolution (\( R \sim 22,500 \)) spectra of large numbers of stars across the galaxy (Zasowski et al. 2017; Beaton et al. 2021; Santana et al. 2021). These spectra are homogeneously reduced and analyzed using the ASPCAP pipeline (Nidever et al. 2015; Zamora et al. 2015; García Pérez et al. 2016) and the resulting stellar parameters are then calibrated using asteroseismic, cluster, and field data (Holtzman et al. 2015, 2018; Jönsson et al. 2020). We choose to use the APOGEE data because this calibration work has already been done and
Figure 2. The reduced density ratio \( \log_{10} r \) is extracted as discussed in Section 4 for four grids of stellar models with differing prescriptions for thermohaline mixing. Results for \( \log_{10} r \) are shown as a function of stellar mass and metallicity \([\text{Fe/H}]\), with high values of \( \log_{10} r \) in brighter colors (yellow) and low values of \( \log_{10} r \) in darker colors (purple). The model name and mixing efficiency, \( \alpha_{\text{th}} \) (where applicable) constitute the physical configuration and are indicated in the panel labels.
The difference in average $[\text{C}/\text{N}]$ (indicated by box color, with negative values in orange and positive values in purple) between stars significantly below the RGB bump and those significantly above the bump is shown as a function of stellar mass and metallicity. The gradient is consistent with previous work, with lower mass, lower metallicity stars having more extra mixing (purple), but there is clearly an unphysical ‘unmixing’ trend (orange) that needs to be removed (see text). We highlight only bins with a sufficient number of stars both below and above the bump.

Figure 3. The difference in average $[\text{C}/\text{N}]$ (indicated by box color, with negative values in orange and positive values in purple) between stars significantly below the RGB bump and those significantly above the bump is shown as a function of stellar mass and metallicity. The gradient is consistent with previous work, with lower mass, lower metallicity stars having more extra mixing (purple), but there is clearly an unphysical ‘unmixing’ trend (orange) that needs to be removed (see text). We highlight only bins with a sufficient number of stars both below and above the bump.

an asteroseismic overlap sample is already available to provide stars with precise and accurate masses, though similar work could likely be done with, for example, the lithium abundances measured by the GALAH survey (Buder et al. 2019) or the $^{12}\text{C}/^{13}\text{C}$ data estimated from the APOGEE data using the Brussels Automatic Code for Characterizing High acCuracy Spectra (BACCHUS, Masseron et al. 2016) pipeline (C. Hayes, submitted).

We also note that the evolution of $[\text{C}/\text{N}]$ is in some ways simpler for these low-mass stars than other mixing diagnostics. Unlike lithium, its abundance at the surface does not change significantly during the main sequence due to the effects of rotational and other mixing processes (Iben 1967). Its initial ratio seems to be somewhat metallicity dependent (Shetrone et al. 2019), with higher values at lower metallicity. As stars reach the first dredge up, there is a strong, rapid, mass–dependent change in the surface $[\text{C}/\text{N}]$ ratio (Masseron & Gilmore 2015; Martig et al. 2016; Ness et al. 2016; Spoo et al. 2022). The $[\text{C}/\text{N}]$ ratio at the surface then remains constant until stars reach the red giant branch bump, after which there seems to be another rapid drop in the $[\text{C}/\text{N}]$ ratio, particularly in stars of low metallicity (e.g. Gratton et al. 2000; Shetrone et al. 2019); it is this drop that has been associated with thermohaline mixing. For stars of particularly low metallicities, there are some suggestions of Upper RGB extra mixing (Shetrone et al. 2019), but this is not well motivated theoretically and is distinct from the processes we are discussing here.

To estimate the amount of extra mixing in these stars near the bump—which thermohaline models suggest should correlate with the mixing coefficient $D_{\text{th}}$ described above—Shetrone et al. (2019) estimated the drop in $[\text{C}/\text{N}]$ just above the red giant branch bump. Their work used $\alpha$-element enhanced, and therefore old and low-mass ($\sim 0.9 M_\odot$), first ascent red giant branch stars and binned them in bins of 0.2 dex in metallicity. The location of the red giant branch bump was identified empirically as an overdensity of stars at a particular surface gravity in each bin. They then identified the log $g$ regime around the red giant branch bump and fit a hyperbolic tangent function to measure the location and size of the drops in the $[\text{C}/\text{N}]$ ratio. For simplicity, we have reproduced their results in Table 1.

We add to their analysis a sample of higher metallicity stars with asteroseismic masses from the APOGEE–Kepler overlap sample (APOKASC, Pinsonneault et al. 2014, 2018). We do this because, according to our analysis in Section 5, higher mass, higher metallicity stars probe larger values of the reduced density ratio, $r$. We first bin the stars in mass (0.2 $M_\odot$) and metallicity (0.2 dex). For consistency with Pinsonneault et al. (2018) and Shetrone et al. (2019), we use the Data Release 14 Abolfathi et al. (2018) carbon and nitrogen abundances. We note however that while the abundance scale seems to shift between releases, the rank ordering does not change very much (Spoo et al. 2022), which means that the conclusions of this analysis are not strongly affected by the choice of Data Release or seismic parameters.

Unlike in the Shetrone et al. (2019) analysis (e.g. their Figure 2), there is not a sufficient number of stars near the bump in each bin to detect and measure the extra mixing directly in the asteroseismic sample. Instead, we define a ‘pre-mixing’ bin of stars between log $g$ of 3.4 and 2.8 dex whose oscillations have identified them as first ascent red giants (Elsworth et al. 2019), as well as a ‘post-mixing’ bin of RGB stars with surface gravities between 2.3 and 1.0 dex. We then compute the average $[\text{C}/\text{N}]$ of stars in each of the pre-mixing and post-mixing bins. If both bins had at least three stars, then the difference between the pre-mixing and post-mixing average
Table 1. Observed extra mixing drops in bins of mass and metallicity, corrected for the 0.1456 dex of unmixing observed that we assume is due to systematic errors. We also include the reduced density ratios calculated for each of these bins using the variety of models discussed in Section 5.

| M  | [Fe/H] | Δ[C/N]_{APK,cor} | Δ[C/N]_{Shet,cor} | r_{Brown,1} | r_{Kip,0.1} | r_{Kip,2} | r_{Kip,700} |
|----|--------|-----------------|-----------------|--------------|-------------|-----------|-------------|
| 0.9 | -1.4   | ...             | 0.73            | 0.00013      | 0.00012     | 0.00015   | 0.00066     |
| 0.9 | -1.2   | ...             | 0.67            | 0.00016      | 0.00013     | 0.00017   | 0.00073     |
| 0.9 | -1.0   | ...             | 0.48            | 0.00018      | 0.00014     | 0.00017   | 0.00078     |
| 0.9 | -0.8   | 0.52            | 0.36            | 0.00020      | 0.00016     | 0.00020   | 0.00099     |
| 0.9 | -0.6   | 0.29            | 0.27            | 0.00026      | 0.00018     | 0.00026   | 0.00140     |
| 0.9 | -0.4   | 0.16            | 0.21            | 0.00029      | 0.00019     | 0.00031   | 0.00143     |
| 1.1 | -0.4   | 0.13            | ...             | 0.00037      | 0.00024     | 0.00036   | 0.00147     |
| 1.3 | -0.4   | 0.07            | ...             | 0.00045      | 0.00027     | 0.00045   | 0.00126     |
| 1.5 | -0.4   | 0.10            | ...             | 0.00054      | 0.00029     | 0.00053   | 0.00160     |
| 0.9 | -0.2   | 0.20            | ...             | 0.00037      | 0.00023     | 0.00036   | 0.00181     |
| 1.1 | -0.2   | 0.11            | ...             | 0.00048      | 0.00028     | 0.00045   | 0.00156     |
| 1.3 | -0.2   | 0.09            | ...             | 0.00059      | 0.00032     | 0.00055   | 0.00161     |
| 1.5 | -0.2   | 0.10            | ...             | 0.00069      | 0.00034     | 0.00065   | 0.00173     |
| 1.1 | 0.0    | 0.14            | ...             | 0.00059      | 0.00032     | 0.00054   | 0.00224     |
| 1.3 | 0.0    | 0.05            | ...             | 0.00072      | 0.00037     | 0.00064   | 0.00210     |
| 1.5 | 0.0    | 0.09            | ...             | 0.00085      | 0.00039     | 0.00078   | 0.00225     |
| 1.7 | 0.0    | 0.02            | ...             | 0.00107      | 0.00042     | 0.00093   | 0.00275     |
| 1.1 | 0.2    | 0.12            | ...             | 0.00072      | 0.00037     | 0.00068   | 0.00226     |
| 1.3 | 0.2    | 0.00            | ...             | 0.00094      | 0.00043     | 0.00083   | 0.00317     |
| 1.5 | 0.2    | 0.01            | ...             | 0.00112      | 0.00047     | 0.00098   | 0.00267     |
| 1.7 | 0.2    | 0.00            | ...             | 0.00142      | 0.00051     | 0.00119   | 0.00318     |
| 1.1 | 0.4    | 0.04            | ...             | 0.00100      | 0.00045     | 0.00092   | 0.00289     |
| 1.3 | 0.4    | 0.01            | ...             | 0.00124      | 0.00052     | 0.00104   | 0.00319     |

[C/N] is plotted in Figure 3. Because of the calibration and choices in the analysis pipeline (see e.g. Holtzman et al. 2018; Jönsson et al. 2020; Smith et al. 2021), the scale of the abundances, particularly for carbon and nitrogen, is somewhat uncertain. There sometimes exist small trends with surface gravity and temperature that are not fully removed in the calibration process. This is notable in our measurement results here; in the highest mass, highest metallicity bins, we formally measure ‘unmixing’ near the red giant branch bump, i.e. an increase rather than a decrease in the average [C/N] near the red giant branch bump, which is inconsistent both with theoretical expectations and with measurements from other sources. Following discussions with the APOGEE team (C. Hayes, private communication), we have decided to correct for these effects by correcting the bin with the most ‘unmixing’ to have 0 mixing, and subtracting that change from all of the other bins under the assumption that the systematic measurement errors are consistent for stars of similar temperatures and gravities. Because we are most interested here in the trend in mixing amounts as a function of the stellar parameters, we do not expect this shift to alter the results of this analysis, but we emphasize that care should be taken by future users of this data.

7. RESULTS

Given the measured amounts of mixing described in Section 6 and the reduced density ratios $r$ computed for stars of various masses and metallicities described in Section 5, it is now possible to compare the observations to the predictions of various thermohaline models from Section 2 and to assess whether the observed mixing is qualitatively consistent with any such theoretical prescription. We show in Figure 4 the corrected changes in [C/N] compared to the inferred reduced density ratios on axes analogous to those of Figure 1, where the y axis represented the rate of mixing. The four panels correspond to the four modeling configurations described in Section 4.

We first note that the observed trends are not strongly sensitive to the assumed 1D mixing model: mixing decreases with increasing reduced density ratio regardless of parameterization. Besides this, there are similarities and differences between the data and the theoretical pre-
8. CONCLUSIONS
Thermohaline mixing has long been considered the most likely candidate for explaining the evolution of the surface chemistry of low–mass upper red giant branch stars. In this analysis, we have shown that:

1. prescriptions informed by three-dimensional simulations of the thermohaline instability suggest that mixing rates should strongly depend on the reduced density ratio, $r$, but the shape of the correlation varies from model to model;

2. one-dimensional stellar evolution models suggest that the average reduced density ratio, $r$, should vary as a function of mass and metallicity, with stars of lower masses and lower metallicities having smaller $r$;

dictions, which are reproduced in Figure 5 for comparison. A key finding is that the observed mixing is strongly correlated with the fluid parameters as predicted; this is true for stars with different masses and metallicities but similar reduced density ratios. We observe a decrease in the amount of mixing as the density ratio increases, which is consistent with standard 1D prescriptions of thermohaline mixing but inconsistent with the prescription from Harrington & Garaud (2019), which was informed by magnetohydrodynamic simulations. We also find that the range of average reduced density ratios probed by the observational data we have available here is much smaller than the range of density ratios simulated by and studied within the theoretical 3D fluid dynamics community (e.g. Brown et al. 2013), although we note that the full range of ratios do appear in each individual simulation (See Appendix C).
3. one-dimensional stellar evolution models suggest that the average reduced density ratio, $r$, is not strongly dependent on the assumed parameterized thermohaline mixing model (e.g. Brown vs Kippenhahn), and regardless of the choice of $\alpha_{th}$ in the case of Kippenhahn;

4. one-dimensional stellar evolution models suggest that the average reduced density ratio, $r$, occupies only a small range ($10^{-4} < r < 10^{-3}$) of the parameter space formally unstable to thermohaline mixing, $0 < r < 1$;

5. observations suggest that the amount of mixing is strongly correlated with the reduced density ratio $r$ in a way that qualitatively agrees with predictions from three dimensional simulations;

6. observations indicate that the mixing rate and reduced density ratio $r$ are inversely correlated, a finding that is consistent with currently available 1D prescriptions informed by 3D simulations, but not consistent with the prescription put forth by Harrington & Garaud (2019).

Most importantly, we find that our proposed framework for connecting observations of extra mixing in red giants to the fluid simulations of the thermohaline instability through the medium of one dimensional stellar evolution models is feasible and robust. It motivates more rigorous exploration into whether red giant branch extra mixing should be associated with the thermohaline instability, and it will facilitate the translation of observational information into theoretical simulations. As observational data sets improve, abundance measurements will only become more capable of constraining stellar interior mixing, the relevant fluid parameters, and potentially even the magnetic fields in these regions.

However, the present study is largely a proof of concept; there is room for significant development in all aspects of this method. From an observational perspective, we have so far only considered the change in $\text{[C/N]}$ as an observational mixing diagnostic, and only in a fairly restricted set of stars. Looking at higher or lower masses, using a variety of mixing diagnostics, or probing stars including globular clusters, binary stars, and dwarf galaxies—where the base composition and mixing history may be different—could all be illuminating expansions of this work. There is also the potential to look at the timing of the extra mixing, as well as mixing rate as a function of time. Further, it may be possible to investigate whether mixing can be connected to the sorts of asteroseismic diagnostics that probe the interior structure of a star, including its density profile (Kjeldsen & Bedding 1995), chemical discontinuities (Verma et al. 2017), internal rotation (Gehan et al. 2018), and magnetic fields (Bugnet et al. 2021).

On the modeling side, we have explored a coarse but reasonable range of possible thermohaline conventions that could have impacted our conclusions, but this pa-
rameter exploration was certainly not exhaustive. It is worthwhile and necessary to test whether different 1D modeling assumptions impact the direction of this trend, particularly in the case of other mixing-related physical assumptions. Key variations in this regard include the choice of prescription for the Mixing Length Theory (MLT) formalism and value for the mixing length parameter, $\alpha_{\text{MLT}}$, the treatment of convective boundary mixing and convective overshoot, and the choice of atmospheric surface boundary conditions—all of which are well known to affect thermodynamic quantities in the regime we study here (Tayar et al. 2017; Joyce & Chaboyer 2018a,b; Viani et al. 2018). For similar thermodynamic reasons, it is also important to explore more extreme metallicity regimes, as the global metal content dictates the behavior of $\nabla \mu$. Further, the need to introduce rotational mixing alongside thermohaline mixing in 1D stellar models to achieve the desired observational reproductions is likewise well established in the literature (c.f. Charbonnel & Lagarde 2010), making the consideration of rotational effects an obvious candidate for future theoretical work.

Finally, while previous and ongoing work has endeavored to find a theoretical justification for why thermohaline mixing should be so efficient that it accounts for the entire amount of extra mixing observed after the RGB bump (by adding e.g. magnetic fields or rotation), this work sets a new target for theoretical efforts: in order for thermohaline mixing to be the primary source of extra mixing, not only should there be a mixing prescription that agrees with 3D simulations and also reproduces the amount of extra mixing, but such a prescription should also reproduce trends in this extra mixing as a function of fundamental stellar parameters like mass and metallicity. For instance, while Harrington & Garaud (2019) demonstrate that magnetic fields of moderate strength can dramatically enhance the efficiency of thermohaline mixing—plausibly to levels that explain the full amount of extra mixing observed after the RGB bump—this work shows that their mixing prescription does not predict trends in mixing rate as a function of the density ratio that are consistent with observations.

In short, this paper has demonstrated the viability of comparing observed signatures of extra mixing on the red giant branch to the predictions of models informed directly by fluid simulations, and it has done so in the framework of parameters used in those simulations. We therefore anticipate and welcome future explorations that will produce better understanding of whether extra mixing should indeed be associated with the thermohaline instability and how observations of real stars—which probe a fluid regime well outside the regime we are currently able simulate—can provide constraints on the physics of that instability.

The first four authors of this manuscript contributed equally to this work.

A. Fraser contributed the majority of text and was responsible for Secs. 2 and 3.

M. Joyce wrote and tested the MESA modeling templates, including inlists, evolutionary and structural output, and run_star_extras functionality, wrote the data analysis and parameter extraction scripts in collaboration with E.H. Anders (Python), and contributed text.

E.H. Anders computed and analyzed stellar structure models, created Figs. 2 and 4-7, and contributed text.

J. Tayar was responsible for all analysis related to observations, constructed the initial manuscript template, and contributed text.

Author order within this set was decided according to height in heels, descending.

The authors thank C. Hayes, A. Jermyn, and M. Pinsonneault for helpful discussions that contributed to this work and T. White for providing the script that translates between [Fe/H] and Z. The authors acknowledge the thermohaline working group at the KITP program “Probes of transport in stars” for fruitful discussion that led to the conception of this idea. The authors likewise thank the Space Telescope Science Institute for providing accommodation and infrastructure support during the week-long meeting during which the majority of this paper was written. AF thanks Pascale Garaud for many discussions of thermohaline mixing which helped improve his understanding of the theory of that field. MJ thanks John Bourke for proofreading and helpful discussion of numerics. EHA thanks MJ for her mentorship in teaching him how to run and interact with MESA models. Support for this work was provided by NASA through the NASA Hubble Fellowship grant No.51424 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. JT also acknowledges support from 80NSSC20K0056. AF acknowledges support from NSF Grant No. AST-1908388. MJ acknowledges the Space Telescope Science Institute’s Lasker Data Science Fellowship, the Kavli Institute for Theoretical Physics at UC Santa Barbara, and the MESA developers team.
EHA acknowledges the support of a CIERA Postdoctoral fellowship. The Center for Computational Astrophysics at the Flatiron Institute is supported by the Simons Foundation. Computations were conducted with support from the NASA High End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center on Pleiades with allocation GID s2276 which is provided through NASA HTMS grant 80NSSC20K1280. This research was supported in part by the National Science Foundation under Grant No. NSF PHY-1748958.

**Software:** Astropy (Astropy Collaboration et al. 2013, 2018), Matplotlib (Hunter 2007), NumPy (van der Walt et al. 2011), SciPy (Virtanen et al. 2020)

**Facilities:** Du Pont (APOGEE), Sloan (APOGEE)

### REFERENCES

Abdur’uf, Accetta, K., Aerts, C., et al. 2022, ApJS, 259, 35, doi: 10.3847/1538-4365/ac4414

Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, ApJS, 235, 42, doi: 10.3847/1538-4365/aa9e8a

Angelou, G. C., D’Orazi, V., Constantino, T. N., et al. 2015, MNRAS, 450, 2423, doi: 10.1093/mnras/stv770

Angulo, C., Arnould, M., Rayet, M., et al. 1999, Nuclear Physics A, 656, 3, doi: 10.1016/S0375-9474(99)00030-5

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Baines, P. G., & Gill, A. E. 1969, Journal of Fluid Mechanics, 37, 289, doi: 10.1017/S0022112069000553

Bauer, E. B., & Bildsten, L. 2019, ApJ, 872, 96, doi: 10.3847/1538-4357/ab0028

Beaton, R. L., Oelkers, R. J., Hayes, C. R., et al. 2021, AJ, 162, 302, doi: 10.3847/1538-4357/ac260c

Blench, P., Mers, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28, doi: 10.3847/1538-4388/aa7567

Brown, J. M., Garaud, P., & Stellmach, S. 2013, ApJ, 768, 34, doi: 10.1088/0004-637X/768/1/34

Buder, S., Lind, K., Ness, M. K., et al. 2019, A&A, 624, A19, doi: 10.1051/0004-6361/201833218

Bugnet, L., Prat, V., Mathis, S., et al. 2021, A&A, 650, A53, doi: 10.1051/0004-6361/202039159

Cantiello, M., & Langer, N. 2010, A&A, 521, A9, doi: 10.1051/0004-6361/201014305

Carbon, D. F., Romanishin, W., Langer, G. E., et al. 1982, ApJS, 49, 207, doi: 10.1086/190796

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

Charbonnel, C., & Lagarde, N. 2010, A&A, 522, A10, doi: 10.1051/0004-6361/201014432

Charbonnel, C., & Zahn, J.-P. 2007, Astronomy & Astrophysics, 467, L15, doi: 10.1051/0004-6361:20077274

Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102, doi: 10.3847/0004-637X/823/2/102

Christensen-Dalsgaard, J. 2015, MNRAS, 453, 666, doi: 10.1093/mnras/stv1656

Chugunov, A. I., Dewitt, H. E., & Yakovlev, D. G. 2007, PhysRevD, 76, 025028, doi: 10.1103/PhysRevD.76.025028

Cox, J. P. 1980, Theory of stellar pulsation

Cyburt, R. H., Anthor, A. M., Ferguson, R., et al. 2010, ApJS, 189, 240, doi: 10.1086/0067-0049/189/1/240

Denissenkov, P. A. 2010, ApJ, 723, 563, doi: 10.1088/0004-637X/723/1/563

Denissenkov, P. A., & Merryfield, W. J. 2011, ApJL, 727, L8, doi: 10.1088/2041-8205/727/1/L8

Denissenkov, P. A., Pinsoneau, M., Terndrup, D. M., & Newsham, G. 2010, ApJ, 716, 1269, doi: 10.1088/0004-637X/716/2/1269

Elsworth, Y., Hekker, S., Johnson, J. A., et al. 2019, MNRAS, 489, 4641, doi: 10.1093/mnras/stz2356

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

Fuller, G. M., Fowler, W. A., & Newman, M. J. 1985, ApJ, 293, 1, doi: 10.1086/163208

Garaud, P. 2018, Annual Review of Fluid Mechanics, 50, 275, doi: 10.1146/annurev-fluid-122316-045234

García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, AJ, 151, 144, doi: 10.3847/0004-6256/151/6/144

Gehan, C., Mosser, B., Michel, E., Samadi, R., & Kallinger, T. 2018, A&A, 616, A24, doi: 10.1051/0004-6361/201832822

Gratton, R. G., Sneden, C., Carretta, E., & Bragaglia, A. 2000, A&A, 354, 169

Grevesse, N., & Sauval, A. J. 1998, SSRv, 85, 161, doi: 10.1023/A:1005161325181

Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332, doi: 10.1086/500975

Harrington, P. Z., & Garaud, P. 2019, The Astrophysical Journal, 870, L5, doi: 10.3847/2041-8213/aaf812
Ulrich, R. K. 1972, ApJ, 172, 165, doi: 10.1086/151336
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science and Engineering, 13, 22, doi: 10.1109/MCSE.2011.37
Verma, K., Raodeo, K., Antia, H. M., et al. 2017, ApJ, 837, 47, doi: 10.3847/1538-4357/aa5da7
Viani, L. S., Basu, S., Joel Ong J. M., Bonaca, A., & Chaplin, W. J. 2018, ApJ, 858, 28, doi: 10.3847/1538-4357/aab7eb
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: https://doi.org/10.1038/s41592-019-0686-2
Wilson, J. C., Hearty, F. R., Skrutskie, M. F., et al. 2019, arXiv e-prints. https://arxiv.org/abs/1902.00928
Zamora, O., García-Hernández, D. A., Allende Prieto, C., et al. 2015, AJ, 149, 181, doi: 10.1088/0004-6256/149/6/181
Zasowski, G., Cohen, R. E., Chojnowski, S. D., et al. 2017, AJ, 154, 198, doi: 10.3847/1538-3881/aa8df9
Table 2. Mappings between [Fe/H] values and MESA input values of (X, Y, Z).

| [Fe/H] | X      | Y      | Z       |
|--------|--------|--------|---------|
| 0.400  | 0.66214302 | 0.29971262 | 0.03814436 |
| 0.200  | 0.69253197  | 0.28229599  | 0.02517204  |
| 0.000  | 0.71318414  | 0.27045974  | 0.01635613  |
| -0.200 | 0.72686070  | 0.26262137  | 0.01051793  |
| -0.400 | 0.73576323  | 0.25751912  | 0.00671765  |
| -0.600 | 0.74149343  | 0.25213501  | 0.00427157  |
| -0.800 | 0.74515509  | 0.25081612  | 0.00270849  |
| -1.000 | 0.74748410  | 0.25008616  | 0.00171429  |
| -1.200 | 0.74896112  | 0.24995509  | 0.00108379  |
| -1.400 | 0.74989606  | 0.24941926  | 0.00068468  |

APPENDIX

A. MESA

The MESA EOS is a blend of the OPAL (Rogers & Nayfonov 2002), SCVH (Saumon et al. 1995), FreeEOS (Irwin 2004), HELM (Timmes & Swesty 2000), PC (Potekhin & Chabrier 2010), and Skye (Jermyn et al. 2021) EOSes.

Radiative opacities are primarily from OPAL (Iglesias & Rogers 1993, 1996), with low-temperature data from Ferguson et al. (2005) and the high-temperature, Compton-scattering dominated regime by Poutanen (2017). Electron conduction opacities are from Cassisi et al. (2007).

Nuclear reaction rates are from JINA REACLIB (Cyburt et al. 2010), NACRE (Angulo et al. 1999) and additional tabulated weak reaction rates Fuller et al. (1985); Oda et al. (1994); Langanke & Martínez-Pinedo (2000). Screening is included via the prescription of Chugunov et al. (2007). Thermal neutrino loss rates are from Itoh et al. (1996).

We create 1D stellar models and evolve them from the pre-main sequence until roughly the end of hydrogen shell burning. We study stellar masses between 0.9 and 1.7 $M_{\odot}$ in steps of 0.2 $M_{\odot}$. We study metallicities [Fe/H] ranging from -1.4 to 0.4 in steps of 0.2 dex. To convert from metallicity units to MESA input $Y$ and $Z$ units, we assume a linear helium enrichment law (per e.g., Choi et al. 2016, sec 3.1) in which we adopt a Big-Bang $Y_p = 0.2485$ and $\Delta Y/\Delta Z = 1.3426$ according to Table 1 of Tayar et al. (2022). The algorithm we use to calculate $X$, $Y$, and $Z$ from these values is identical to the one used in https://github.com/aarondotter/initial_xa_calculator; we use the opacity tables of Grevesse & Sauval (1998) and [$\alpha$/Fe]. The specific [Fe/H] to (X, Y, Z) conversions used here are shown in Table 2.

B. RESOLUTION TESTING

We performed resolution tests for models with [Fe/H] $\in \{-1.2, -0.4, 0.4\}$ and $M \in \{0.9, 1.3, 1.7\}$ using the Brown thermohaline mixing prescription. We studied a grid of mesh_delta_coeff and time_delta_coeff values which span from 0.1 to 1.0 over five log-space steps. This means that we evolved simulations with both spatial and temporal resolutions ranging from $1 \times$ to $10 \times$ the default resolutions. For each simulation, we evolve the 1D stellar model as described in Sec. 4, then decrease (or increase) the spatial and temporal resolution on the red giant branch once log $g \leq 3$. We measure the inverse density ratio $r$ in each of these models, and in Fig. 6 we plot the absolute value of the percentage difference between that $r$ value and the reference $r_{ref}$ value reported for that case in Fig. 2. We calculate the percentage difference to be $100(1 - r/r_{ref})$.

We find that small values of the mesh coefficient (high spatial resolution) combined with large values of the time coefficient (large timesteps) result in large errors. This occurs because the front of the thermohaline zone, and sometimes the full thermohaline zone, becomes numerically unstable, and large oscillations in $R_0$ lead to large errors in the $r$ calculation. Furthermore, we find that when the thermohaline front is not properly numerically resolved, it does not propagate upwards in mass coordinate and so does not connect with the convective shell. This likely has important implications for the evolution (or lack thereof) of surface abundances in these models.
Figure 6. We plot a 3x3 grid of colormaps corresponding to a grid of mass $M \in [0.9, 1.3, 1.7]$ and $[\text{Fe/H}] \in [-1.2, -0.4, 0.4]$. At each mass and $[\text{Fe/H}]$, we simulate a 5x5 grid of MESA models with varying spatial and temporal resolution. We plot in color the percent difference between the measured value of the reduced density ratio $r$ and its reference value reported in Fig. 2. The resolution of the grids of simulations presented in Fig. 2 are marked by black stars. Cases with $r$ measurements within 5\% of the reference values are colored in green, while points with larger differences are colored pink. Grey pixels are simulations for which either there was a supercomputer error or the algorithm failed to identify a thermohaline zone.

C. MOVIE OF THERMOHALINE FRONT EVOLUTION

In Fig. 7, we plot the stellar structure vs. mass coordinate in the simulation which employs the Brown model and has $M = 1.1M_\odot$ and $[\text{Fe/H}] = -0.2$. We limit the x-limits of the plot to the mass coordinate energy generation peak
Figure 7. We plot various stellar structure quantities (indicated in the table) vs. mass coordinate \((M/M_\ast)\) for a \(M_\ast = 1.1M_\odot\) star with \([\text{Fe/H}] = -0.2\) which employs the Brown et al. (2013) mixing prescription. We zoom in near the hydrogen burning shell, as indicated by the pink \(\text{eps}_\text{nuc}\) line which shows the nuclear energy generation rate. The thermohaline region can be identified as the place where there is a large amount of mixing (the light green \(\log_\text{D}_\text{mix}\) is large), and we shade the bulk of this region in grey. As described in section 4, we only measure \(r\) over 1/3 of this region by mass, and the region over which we take this measurement is shaded in a darker grey. Within this region, we plot \(r\) in dark green. A zoom-in on \(\log_{10}r\) within the dark grey region is shown in the inset, and the measured value of \(r\) identified by the algorithm is plotted as a dark green horizontal line. Additional plotted lines include \(R_0\) and \(1/\tau\) as described in Sec. 3. Various quantities are quoted in text at the top of the image, including the star’s age in years, the thermohaline mixing timescale in seconds, the radial extent of the thermohaline zone in terms of both length (\(dr\)) and mass coordinate (\(dm\)), and a flag indicating that this model is identified as “good” by our algorithm. An animated version of this figure is available online in the published HTML version of this article and on YouTube at https://youtu.be/XLU8aS2q5-o.

of the hydrogen burning shell on the left \(m_{\text{max}}\), and to \(1.1m_{\text{max}}\) on the right. An animated version of this figure is available online in the published HTML version of this article and on YouTube at https://youtu.be/XLU8aS2q5-o.