CrossMark

The SPHINX M-dwarf Spectral Grid. I. Benchmarking New Model Atmospheres to Derive Fundamental M-dwarf Properties

Aishwarya R. Iyer[1,6] , Michael R. Line[1] , Philip S. Muirhead[2] , Jonathan J. Fortney[3] , and Ehsan Gharib-Nezhad[4,5]

[1] School of Earth and Space Exploration, Arizona State University, 525 E. University Dr., Tempe, AZ 85281, USA; aiyer13@asu.edu
[2] Department of Astronomy and Institute for Astrophysical Research, Boston University, 725 Commonwealth Ave., Boston, MA 02215, USA
[3] Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High St, Santa Cruz, CA 95064, USA
[4] NASA Ames Research Center, Moffet Field, CA 94035, USA
[5] Bay Area Environmental Research Institute, CA, USA

Abstract

About 70%–80% of stars in our solar and Galactic neighborhood are M dwarfs. They span a range of low masses and temperatures relative to solar-type stars, facilitating molecule formation throughout their atmospheres. Standard stellar atmosphere models primarily designed for FGK stars face challenges when characterizing broadband molecular features in spectra of cool stars. Here, we introduce SPHINX—a new 1D self-consistent radiative–convective thermochemical equilibrium chemistry model grid of atmospheres and spectra for M dwarfs in low resolution ($R \sim 250$). We incorporate the latest precomputed absorption cross sections with pressure broadening for key molecules dominant in late-K, early/main-sequence-M stars. We then validate our grid models by determining fundamental properties ($T_{\rm eff}$, log g, [M/H], radius, and C/O) for 10 benchmark M+G binary stars with known host metallicities and 10 M dwarfs with interferometrically measured angular diameters. Incorporating the Gaussian process inference tool *Starfish*, we account for correlated and systematic noise in low-resolution (spectral stitching of SpeX, SNIFS, and STIS) observations and derive robust estimates of fundamental M-dwarf atmospheric parameters. Additionally, we assess the influence of photospheric heterogeneity on inferred [M/H] and find that it could explain some deviations from observations. We also probe whether the adopted convective mixing length parameter influences inferred radii, effective temperature, and [M/H] and again find that may explain discrepancies between interferometric observations and model-derived parameters for cooler M dwarfs. Mainly, we show the unique strength in leveraging broadband molecular absorption features occurring in low-resolution M dwarf spectra and demonstrate the ability to improve constraints on fundamental properties of exoplanet hosts and brown-dwarf companions.

1. Introduction

M dwarfs are the most ubiquitous stars in our Galaxy, the solar neighborhood, and the universe put together, making up a majority of the combined stellar population by number (Henry et al. 1994). They have incredibly long main-sequence lifetimes —longer than the age of the universe, making them extremely valuable for understanding the dynamical and chemical evolution of the Galaxy (Lépine et al. 2003a, 2003b, 2013; Bochanski et al. 2010, 2011; Lépine & Gaidos 2011; West et al. 2011; Hejazi et al. 2015). Numerous space- and ground-based surveys to find exoplanets have discovered many terrestrial planets orbiting M-dwarf host stars (Nutzman & Charbonneau 2008; Irwin et al. 2009; Gillon et al. 2017). Currently, the Transiting Exoplanet Survey Satellite is continuing to discover thousands of terrestrial exoplanets, most of them orbiting M dwarfs (Muirhead et al. 2018; Ballard 2019) —with an occurrence rate of at least ∼1 rocky planet around such a host (Dressing & Charbonneau 2013, 2015). Because of a high occurrence of Earth-like terrestrial planets in their

habitable zones (Dressing & Charbonneau 2013; Kopparapu et al. 2013; Petigura et al. 2013; Mulders et al. 2015; Gaidos et al. 2016), M dwarfs are valuable targets in the James Webb Space Telescope (JWST) era for understanding the hosting environment, formation, evolution and fundamental atmospheric properties of companion planets.

Fundamental properties of cool stars include effective temperature, gravity, radius, and intrinsic elemental abundances. Processes of stellar formation, dynamics, the formation of molecules, atmospheric opacities, and energy balance depend on accurate knowledge of these fundamental parameters. Previous works have focused on deriving relative elemental abundances by looking at specific lines (e.g., C, O, Fe, etc.) (Fischer & Valenti 2005; Melendez et al. 2009; Chambers 2010; Ramírez et al. 2011, 2014; Teske et al. 2014, 2016; Brewer et al. 2015), which has worked adequately for FGK stars. Moreover, classic methods using synthetic spectral fitting techniques and EW methods have generally focused on narrow spectral regions at higher resolutions (Marfil et al. 2021), which has been particularly problematic as a result of the presence of blended molecular lines when characterizing stars less massive than the Sun. Due to their lower effective temperatures and strategic placement between the main-sequence FGK stars and brown dwarfs, M dwarfs have atmospheres that are plagued with multiple molecules and are known to have significant regions with turbulent convective interiors (Osterbrock 1953). The presence of molecules results in

---

[6] NASA FINESST Fellow.
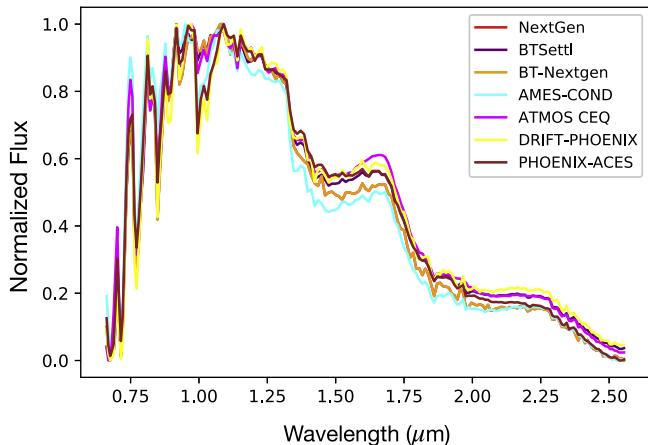
Figure 1. Disagreements between spectra for different stellar model grids available to the community for $T_{eff} = 3000$ K, $\log g = 5.0$, and [Fe/H] = 0.0. Here we see a wide range of differences in spectral shape that may be attributed to model assumptions or differences in the opacity database.

M-dwarf spectra being dominated by broadband absorption features, which influence the overall spectral energy distribution and produce a pseudo-continuum effect that varies with chemical abundances (Veyette et al. 2016). Model grids assuming self-consistent one-dimensional radiative–convective equilibrium conditions available to the community today (Tsuji et al. 1996; Allard et al. 1997, 2000, 2001, 2011; Tsuji 2002; Gustafsson et al. 2008; Helling et al. 2008; Kurucz 2011; Witte et al. 2011; Husser et al. 2013; Phillips et al. 2020) show a wide variation in spectral shape for the same star (see Figure 1 as an example, for a star with $T_{eff} = 3000$ K, $\log g = 5.0$, and solar metallicity), possibly due to differences in atomic and molecular line lists used or the set of model assumptions incorporated. Some of these differences could be further exacerbated by various stellar atmospheric effects that are not yet well understood, such as photospheric heterogeneity in active, fully/partially convective M dwarfs or dust physics (Patience et al. 2012; Apai et al. 2018). Passegger et al. (2022, and references therein) show broad inconsistencies when comparing fundamental parameters of M dwarfs derived from different methods such as pseudo-equivalent width measurements, spectral fitting techniques, empirical calibrations, or machine learning methods relative to observations, despite employing standardization techniques such as consistent synthetic spectral models, wavelength ranges from consistent sets of observations, flux normalizations, dimensionality reduction in spectral fitting by fixing specific parameters, etc. In fact, Passegger et al. (2022) found that different methods employed to characterize M-dwarf abundances show consistency with the overall literature means of benchmark M dwarfs when used without these standardization procedures within their native frameworks.

Naturally, various implications could potentially arise due to these unresolved differences when it comes to modeling M-dwarf atmospheres, such as: (1) improper understanding of the heterogeneous nature of M-dwarf photospheres will continue to induce biases when inferring fundamental atmospheric properties of transiting planets around them despite appropriate corrections, joint star–planet retrievals, and high-fidelity JWST observations (Iyer & Line 2020), or (2) systematic uncertainties in analyzing population trends of effective temperature, metallicity, and alpha enhancement parameters derived from spectral fitting could persist and propagate into studies of Galactic chemodynamical evolution as a result of model deficiencies, especially for early M stars

($T_{eff} > 3500$ K) (Hejazi et al. 2022), or (3) inaccurate estimates of M-dwarf atmospheric C/O ratios—in particular, widely separated primary companions of cooler brown dwarfs (such as Ross 19 AB, Schneider et al. 2021) or exoplanet companions could result in incorrect conclusions about formation mechanisms of these sub-stellar-mass objects (Madhusudhan 2012; Moses et al. 2012; Mordasini et al. 2016). These are only a few consequences that plague our understanding of the complete picture concerning such stellar systems. Improper understanding of M dwarfs would pose a significant setback in answering some of the most fundamental questions about our universe, thereby emphasizing the crucial nature of efforts aimed at characterizing these stars.

In light of these, we attempt to overcome a few challenges by (1) developing a new stellar synthesis model and (2) validating our model-derived stellar fundamental parameters with benchmark M-dwarf targets. In Section 2, we introduce our new stellar atmosphere model grid and describe the low-resolution data used for this analysis. Then, we present our error mitigation approach using *Starfish* (Czekala et al. 2015b, 2018), a Bayesian framework with a Gaussian process kernel accounting for grid-interpolation uncertainties, data systematics, or model uncertainties and our robust estimates of fundamental properties of M dwarfs including T_{eff}, $\log g$, metallicity, C/O, and radius. In Section 3, we validate our metallicity values retrieved from the grid model with empirically measured [M/H] for widely separated M+G binary dwarfs (hereafter, WBS) (Mann et al. 2013) as well as our retrieved radii with those measured with interferometric angular diameters (hereafter, IS) (Boyajian et al. 2012), using the SpeX, SNIFS, and Hubble Space Telescope (HST) STIS observations from Mann et al. (2013, 2015) and references therein, all convolved to a regime of low spectral resolution ($R \sim 120$). In Section 4, we show the influence of including additional parameterizations to retrieve stellar surface photospheric heterogeneity and the influence of a model-assumed mixing length convection parameter on retrieved metallicities, effective temperatures, and radii. Finally, we end with a summary and implications of our work in Section 5.

2. New Model Atmospheres

2.1. 1D Radiative–Convective Equilibrium Model Grid

Our 1D radiative–convective equilibrium model builds upon that from the ScCHIMERA tool primarily developed for extrasolar planet atmospheres described in Arcangeli et al. (2018), Bonnefoy et al. (2018), Kreidberg et al. (2018), Piskorz et al. (2018), Gharib-Nezhad & Line (2019), Gharib-Nezhad et al. (2021a), and Mansfield et al. (2021). The model iteratively solves for vertical profiles of volume mixing ratio, cloud/condensate properties, thermal structure, and the top-of-atmosphere disk-integrated stellar spectrum. We then generate a grid of model spectra and atmospheres,[7] spanning effective temperatures (T_{eff}) from 2000 to 4000 K in steps of 100 K, surface gravity ($\log g$) from 4.0 to 5.25 in steps of 0.25, metallicity ([M/H]) ranging from -1 to 1 in steps of 0.25, and carbon-to-oxygen ratio (C/O) from 0.3 to 0.9 in steps of 0.2, all with uniform spacing. The components of our model pipeline are outlined below.

1. *Radiative Transfer.* We follow the two-stream source function approximation described in Toon et al. (1989)

---

[7] New models are available here: https://doi.org/10.5281/zenodo.7416042.

and Petty ([2006]). We compute radiative transport through the atmosphere constrained by hydrostatic equilibrium in a plane-parallel geometry, solving for the temperature–opacity combination that results in net zero flux divergence across each atmospheric layer (Marley & Robinson [2014]; Hubeny [2017]) using the Newton–Raphson iterative scheme described in McKay et al. ([1989]).

2. *Equation of State.* We use the NASA–CEA package (Gordon & Mcbride [1994]) to compute equilibrium chemical abundances of molecules as well as neutral and singly ionized species. This package can compute equilibrium chemical abundances for thousands of species; however, we only use it to compute abundances for a subset of key molecules and elemental species characteristic of M-dwarf atmospheres (notably metal oxides, hydrides, and neutral and some doubly ionized species).

3. *Opacity Sources.* We take into account opacities relevant to cool stellar atmospheres, including CH_4, CO, CO_2, NH_3, PH_3, HCN, C_2H_2, H_2S, H_2–H_2/He collision-induced absorption (Freedman et al. [2014]; Lupu et al. [2022]), H_2O (Polyansky et al. [2018]), metal hydrides SiH (Yurchenko et al. [2017]), TiH, FeH, CaH, CrH, AlH, MgH, and metal oxides TiO, VO, SiO (*EXOPLINES*, Gharib-Nezhad et al. [2021a], [2021b] and references therein). We also include neutral atomic lines Na, K (Allard et al. [2016], [2019]), Fe I, Mg I, Ca I, C I, Si I, Ti I, O I, ionized atoms Fe II, Mg II, Ti II, C II, Ca II (Kurucz [1993], [2011]; Piskunov et al. [1995]), and bound–bound and bound–free H-continuum (John [1988]). Our molecular opacities are continually updated based on the latest line lists from ExoMol (Tennyson et al. [2016], [2020]). Opacity functions can be dense due to numerous transitions as a function of temperature leading to millions of lines within any spectral window, depending on the resultant resolution. Therefore, to overcome the computational burden caused by integrating over the opacity function in a tedious line-by-line manner, we use the correlated-k resort–rebin approach as described in Amundsen et al. ([2017]). Here, the precomputed cross sections are effectively sampled with line-by-line bins mapped onto a smooth cumulative distribution function of line strengths per bin while keeping the ranking of line strengths consistent through the atmospheric column (Marley et al. [1996]; Burrows et al. [1997]; Fortney et al. [2010]; Grimm & Heng [2015]). The line transmittance integral is computed by performing a Gaussian quadrature sum using the cumulative distribution transformation, choosing tens of quadrature points per spectral bin (we use about 20 points). This opacity transformation allows for faster computation of the transmission function for opacities without significant information loss (Marley et al. [1996]; Burrows et al. [1997]; Showman et al. [2009]; Fortney et al. [2010]; Line et al. [2013], [2014], [2015]). In Figure 2, we show molecular, atomic, and continuum opacities included in our model suite at 3200 K temperature and 1 bar pressure. In fact, the choice of line list used for computing key M-dwarf molecular opacities can show a difference in spectral shape by over a factor of 2 with residual differences in spectral shape of up to 125% and a difference in the thermal profile of up
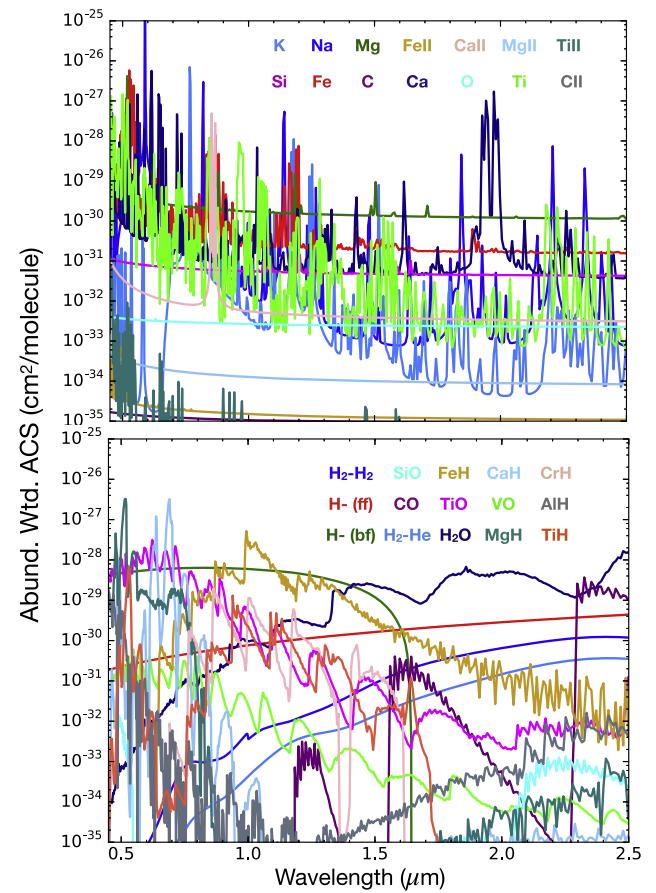

Figure 2. Abundance-weighted absorption cross sections (ACSs) of atomic (top) and molecular (bottom) species included in our model grid from *EXOPLINES* (Gharib-Nezhad et al. [2021a]) at a temperature of 3200 K and a pressure of 1 bar covering a fiducial wavelength grid spanning 0.2–2.5 μm at a resolution $\lambda/\Delta\lambda = 250$.

to 60 K (Gharib-Nezhad et al. [2021a]). Therefore, it is absolutely crucial to use the most up-to-date line lists with appropriate pressure-broadening treatment (see Gharib-Nezhad et al. [2021a]), particularly for key molecules that dominate M-dwarf atmospheres such as metal oxides (TiO, VO, H_2O) and metal hydrides (MgH, CrH, FeH, CaH, SiH, TiH), which is the highlight of our synthesis model.

4. *Convection.* Our convection model follows mixing length theory (Böhm-Vitense [1958]) to compute heat diffusion through a turbulent medium. We solve for turbulent diffusivity and upward transport velocity using a nonzero convective heat flux in the adiabatic region of the atmosphere. For simplicity, we assume a mixing length parameter value of 1. We use the Schwarzschild criterion to assess whether conditions for convection are met and therefore solve for the convective flux through each stratum and self-consistently update the thermal profile for regions with steep gradients. Mixing length theory also allows for a continuous transition from a fully radiative to a fully convective solution, providing better control over P–T profile convergence and permitting detached convection zones (Hubeny [2017]).

5. *Dust/Clouds.* Dust/cloud particles form when solid grains or liquid droplets condense due to the gas vapor pressure surpassing the local saturation vapor pressure

(Ackerman & Marley 2001; Helling et al. 2008). Contributions from condensate opacity are calculated from the total absorption and scattering efficiencies given the indices of refraction and particle size distributions. Additionally, the particle sizes as a function of altitude are assumed to follow a log-normal distribution (similar to Ackerman & Marley 2001; Sharp & Burrows 2007). For the current grid-model spectra, we assume the M-dwarf atmospheres to be cloud-free; however, this model capability is readily available to expand to late and/or low-surface-gravity M dwarfs, with atmospheres including (but not limited to) Mg_2SiO_4, $MgSiO_3$, $CaTiO_3$, Na_2S, KCl, and Fe.

In Figure 3, we show slices of our model grid versus PHOENIX-ACES[8] (Husser et al. 2013) (smoothed to $R \sim 250$) varying over T_{eff}, $\log g$, and [M/H], which we define as the sum of all elemental abundances relative to hydrogen and helium, with the solar value being 0.0 in log space. Overall, we demonstrate that our models are consistent when compared to the widely used stellar model grids that are available to the community today, with noticeable differences, particularly for the M dwarfs with greater than solar metallicity and early M dwarfs with $T_{eff} > 3500$ K, a likely consequence of using the most recent line lists.

2.2. Low-resolution Data of Benchmark Systems

We analyze M-dwarf data in two groups. The first is widely separated M+G binaries (WBS) with empirically calibrated metallicities ([M/H]) from the host stars (Mann et al. 2013). The spectra obtained were observed with the SpeX spectrograph (Rayner et al. 2003) and SNIFS (Lantz et al. 2004) and cover a combined visible + near-IR (NIR) range spanning 0.32–2.4 μm (Mann et al. 2013). Both sets of observations were carried out by Mann et al. (2013) with the SpeX using the NASA Infrared Telescope Facility (IRTF) on Maunakea with the SXD mode ($0.\!''3 \times 15''$ slit at $R \sim 2000$) and the SNIFS observations with the University of Hawaii 2.2 m telescope (with $R \sim 1300$). Mann et al. (2013) then constructed the combined spectrum we use for our first group of M-dwarf benchmark targets by stitching both sets of data within the overlapping region of 0.81–0.96 μm and performing appropriate corrections for radial velocity (RV) such that the offsets between the two sets are comparable to random errors in their RV measurements (Mann et al. 2013). Our second group focuses on nearby M dwarfs from an interferometric sample (IS) with measured angular diameters from Boyajian et al. (2012). For this group, we obtain data observed by Mann et al. (2015) using SNIFS (Aldering et al. 2002; Lantz et al. 2004) and STIS (Bohlin et al. 2001; Bohlin & Gilliland 2004) on the HST that cover a total optical bandpass of 0.32–0.97 μm with $R \sim 1000$ and NIR Spex (Rayner et al. 2003) covering 0.8–2.4 μm at $R \sim 2000$ (Mann et al. 2015). The SNIFS, SpeX, and STIS data were again stitched, H-band flux-calibrated, and renormalized to minimize the relative offsets between the overlapping regions (see Table 1, and Mann et al. 2015 for further details on the observations and data reduction procedures).

For both groups of M dwarfs, we focus purely on the optical/NIR wavelength range as it is ideal for inferring

molecular abundances and deriving bulk chemical compositions seamlessly (Line et al. 2015). Furthermore, we restrict ourselves only to the low-resolution regime, which provides a way to leverage broadband molecular features over a wide wavelength range without incorporating non-LTE effects or microturbulence—a concern when working with higher-resolution spectra. Additionally, Jofré et al. (2014) showed that varying instrument resolution resulted in metallicity differences of less than 0.05 dex for benchmark FGK stars they analyzed and less than 0.1 dex for stars with T_{eff} below 5000 K. Such a paradigm shift to lower resolutions has also been demonstrated in Galactic abundance synthesis models and brown dwarfs (Line et al. 2015, 2017; Ting et al. 2018; Zalesky et al. 2019, 2022; Zhang et al. 2021a, 2021b). Our model grid spectra are computed at $R \sim 250$ (see Figure 3); however, we smooth both the data and model to $R \sim 120$ for this analysis to match a generic SpeX Prism Library spectrum (Burgasser 2014) as previously demonstrated for benchmark T dwarfs by Zhang et al. (2021a) and mainly to demonstrate the strength in our methodology in the regime of low spectral resolution.

2.3. Grid-model Fitting Routine

We use the Bayesian inference tool *Starfish*[9] (Czekala et al. 2015b, 2018) as our primary grid-model interpolation method, as successfully used for M/G-type stars (Czekala et al. 2015b, 2018; Gully-Santiago et al. 2017), benchmark brown dwarfs (Zhang et al. 2021a, 2021b), and protostars (Greene et al. 2018). We construct a spectral emulator with *Starfish*, specifying the range of our grid of models: 2000 K $\leqslant T_{eff} \leqslant$ 4000 K, $-1 \leqslant$ [M/H] $\leqslant 1$, $0.3 \leqslant$ C/O $\leqslant 0.7$, and $4.0 \leqslant \log g \leqslant 5.5$. Similar to Zhang et al. (2021a), we reduce the spectral resolution of our model spectra for all grid points using a Gaussian kernel matching the resolving power delivered by SpeX in Prism mode with a $0.\!''3$ slit (Rayner et al. 2003) to be $R \sim 120$. The primary function of the emulator is to interpolate by first deconstructing the convolved model grid into eigenspectra using principal component analysis (PCA). In this step, the models are evaluated with a training set and a testing set to generate eigenspectra of PCA components, the mean spectrum over the entire grid, the standard deviation spectrum, as well as associated weights capturing model grid systematics (Czekala et al. 2015b, 2018). Then, for any given set of parameters spanning our grid, a reconstructed (or interpolated) spectrum is computed as a linear combination of these. Emulator reconstruction systematics do not tend to dominate the forward modeling error budget, as demonstrated by Zhang et al. (2021a) in the case of Sonora–Bobcat models (Marley et al. 2021). A Gaussian process kernel is used to solve for an emulator covariance matrix comprising the standardized weights for each PCA eigenspectrum, providing a probability distribution of the weights, a distribution of the interpolated spectra, as well as the propagated interpolation uncertainties (see Czekala et al. 2015b, 2018 for further details).

The key feature of *Starfish* is the construction of a composite covariance matrix that also accounts for the total error budget with both parameter estimation and model evaluation. Components of the full covariance matrix are inferred via retrievals (referred to as hyperparameters) to allow for realistic error analysis driven by the data. In addition to the emulator

---

[8] http://phoenix.astro.physik.uni-goettingen.de/

[9] https://github.com/Starfish-develop/Starfish
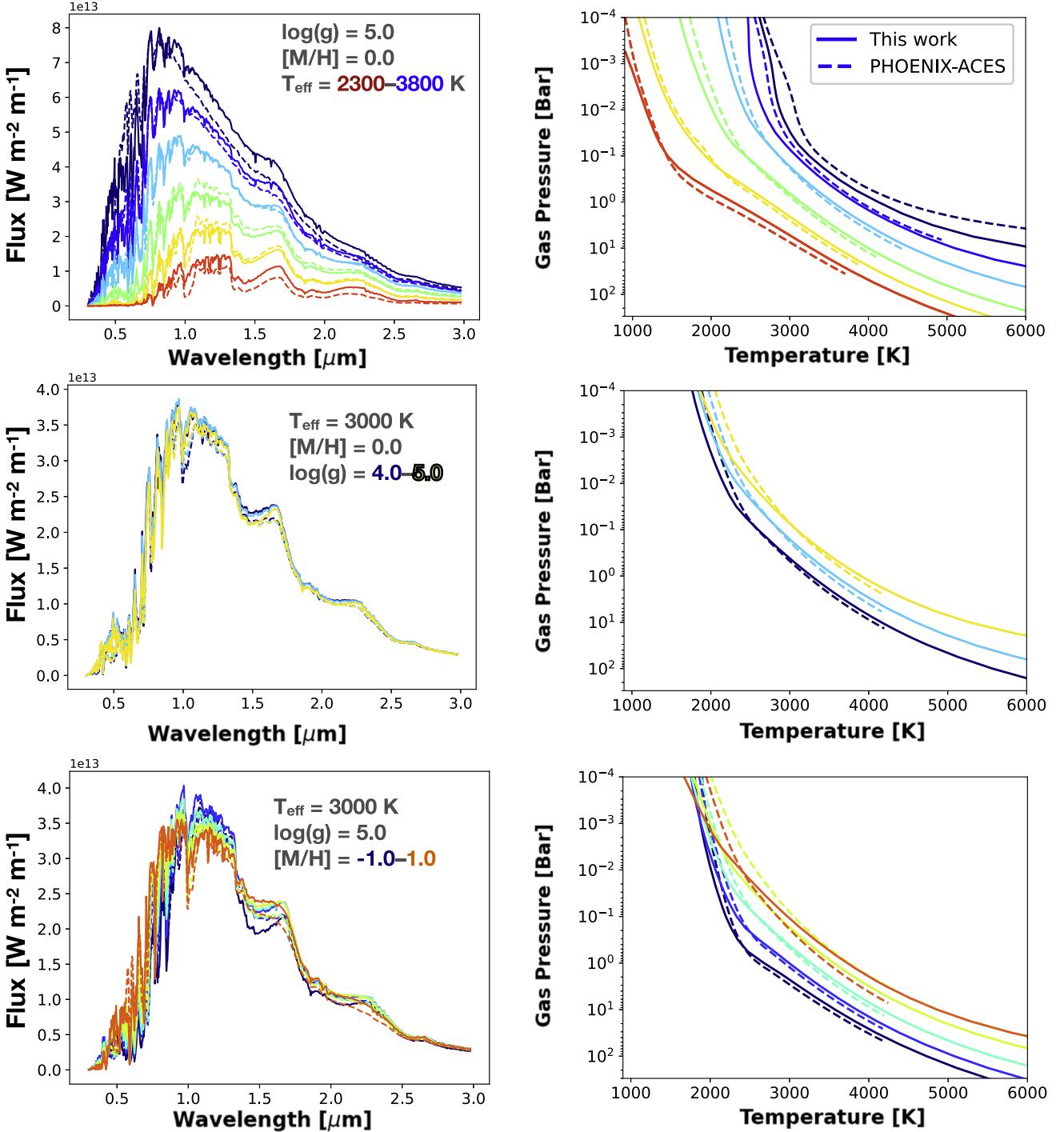
Figure 3. Comparison of a few of our model grid spectra (solid lines) to the PHOENIX-ACES high-resolution model (dashed lines, smoothed to $R \sim 250$) (Husser et al. 2013) in the left panel with (top) varying T_{eff} with fixed $\log g = 5.0$ and [Fe/H], [M/H] = 0.0, (middle) varying $\log g$ with fixed $T_{eff} = 3000$ K, [Fe/H], [M/H] = 0.0, and (bottom) varying [Fe/H], [M/H] with fixed $T_{eff} = 3000$ K and $\log g = 5.0$. Each right panel corresponds to the respective atmospheric structure profile. Here we see that our model grid spectra are in overall agreement with the latest PHOENIX-ACES grid, with most differences arising in varying the stellar metallicity and effective temperature. Our grid also includes a C/O axis (not shown here), ranging from values of 0.3 to 0.9 in steps of 0.2. We include the most up-to-date molecular line-list data (Gharib-Nezhad et al. 2021a) to compute opacities for M-dwarf atmospheres, possibly explaining the main differences.

covariance matrix described above, *Starfish* (Czekala et al. 2015a) also computes a noise covariance matrix, which accounts for differences in the form of data–model uncertainties and observed flux uncertainties within and between wavelength pixels, inferred as parameter a_N. Next, a Gaussian process global covariance matrix is computed, which uses a squared exponential kernel accounting for correlated residuals

caused by oversampling the instrument line-spread function or deficiencies in the atmosphere model as well as systematics such as insufficient opacity data are all inferred as parameters $\log(a_G)$ and $\log(l)$ (Czekala et al. 2015b, 2018; Zhang et al. 2021a). With this method, the traditional evaluation of the likelihood function within an inference problem is modified as a linear combination of these covariance matrices, ultimately

Table 1
H-band Magnitudes, Distances, and Spectral Types of Targets Used for This Analysis: 10 Widely Separated M+G Binaries (WBS) and 10 Targets with Interferometrically Measured Angular Diameters (IS)

Target	*H*-band Magnitude[a,b]	Distance[c] (pc)	Spec. Type[d,e]
Wide Binary Sample (WBS)			
NLTT 11270	10.257 ± 0.021	69.1884 ± 0.1565	M0.4
NLTT 11500	8.749 ± 0.026	36.2962 ± 0.0829	M1.8
NLTT 3725	8.927 ± 0.023	20.4716 ± 0.0479	M3.9
NLTT 738	10.231 ± 0.022	76.7054 ± 0.2795	M2.2
NLTT 8787	6.080 ± 0.038	11.1998 ± 0.0192	M1.6
Gl 100	8.571 ± 0.029	19.6610 ± 0.0210	M2.5
Gl 105	6.793 ± 0.038	7.2221 ± 0.0046	M3.5
Gl 118.2	8.935 ± 0.021	22.5773 ± 0.0325	M3.8
Gl 81.1	7.763 ± 0.021	33.5565 ± 0.0458	M0.1
NLTT 10349	11.043 ± 0.023	77.6536 ± 0.1960	M1.0
Interferometric Targets (IS)			
Gl 436	6.319 ± 0.023	9.7559 ± 0.0089	M2.8
Gl 880	4.800 ± 0.036	6.8676 ± 0.0018	M1.5
Gl 699	4.83 ± 0.03	1.8266 ± 0.0010	M4.2
Gl 526	4.775 ± 0.206	5.4353 ± 0.0015	M1.4
Gl 687	4.77 ± 0.03	4.5500 ± 0.0007	M3.2
Gl 887	3.61 ± 0.23	3.2871 ± 0.0005	M1.1
Gl 411	3.640 ± 0.202	2.548 ± 0.006	M1.9
Gl 581	6.095 ± 0.033	6.2992 ± 0.0020	M3.2
Gl 725A	4.741 ± 0.036	3.5218 ± 0.0008	M3.0
Gl 725B	5.197 ± 0.024	3.5228 ± 0.0013	M3.5

Notes. The WBS spectra were taken from SNIFS (Aldering et al. 2002; Lantz et al. 2004) in the optical and SpeX in the IR (Rayner et al. 2003) by Mann et al. (2013), and the IS spectra were taken from SNIFS (Aldering et al. 2002; Lantz et al. 2004) + STIS (Bohlin et al. 2001; Bohlin & Gilliland 2004) in the optical and SpeX in the IR (Rayner et al. 2003) by Mann et al. (2015).
[a] *H*-band magnitude of the Two Micron All Sky Catalog of Point Sources (Cutri et al. 2003).
[b] GAIA GR2 Collaboration (Gaia Collaboration 2018).
[c] Vizier Online Data Catalog (Ducati 2002).
[d] Mann et al. (2013).
[e] Boyajian et al. (2012).

leaving us to solve for the following eight parameters in our grid-model retrievals: $T_{\rm eff}$, $\log g$, [M/H], C/O, R_*, as physical parameters, and $\log(a_G)$, $\log(l)$, and a_N as hyperparameters (see Figure 4). Note that we essentially fit for the $(R_*/d)^2$ scaling parameter as a proxy for the radius R_* constrained by distances measured from the Gaia Collaboration et al. (2016) and Gaia Collaboration (2018).

2.4. Priors

Our grid-model fitting pipeline allows for priors on both physical parameters and the hyperparameters. For the effective temperature, surface gravity, metallicity, carbon-to-oxygen ratio, and radius, we use uniform priors spanning the range of the grid (2000 K $\leqslant T_{\rm eff} \leqslant$ 4000 K, $4.0 \leqslant \log g \leqslant 5.0$, $-1 \leqslant$ [M/H] $\leqslant 1$, $0.3 \leqslant$ C/O $\leqslant 0.7$ and $0.01~R_\odot \leqslant R_* \leqslant 0.8~R_\odot$). For efficiency, we truncate the grid ranges for each parameter axis informed by our linear interpolation routine (as done in Zalesky et al. 2019), which precedes the *Starfish* analysis, for each target in our sample. We found that this speeds up the matrix reconstruction by the *Starfish* emulator for

a PCA training set over the grid points by about an order of magnitude. We define physical priors as uniform within the parameter ranges of our subgrid, specific to each target.

For hyperparameters, we follow recommendations from Zhang et al. (2021a) and Czekala et al. (2015b) and use a Gaussian prior on a_N with a mean of 1.0 and width of 0.25 over all positive values to sufficiently describe the "noise covariance matrix" that accounts for covariance in data–model residuals within each pixel. We use uniform priors on both "global covariance matrix" parameters l and $\log(a_G)$ to be between [200, 2000 × 5] km s^{-1} and $(-\infty, \infty)$ respectively for our stitched SNIFS+SpeX (Rayner et al. 2003; Lantz et al. 2004) spectrum (Mann et al. 2013, 2015), which describes the behavior of the covariance values between two wavelengths, the SXD-mode line-spread function, and the amplitude of the Gaussian process kernel to quantify correlated residuals from model systematics. Although we did not rigorously compute the optimum autocorrelation wavelength to infer the appropriate range for the wavelength scale parameter (l) given our choice of data as done in Zhang et al. (2021a), we did assume the average offset in the stitching of the two spectra (SNIFS +SpeX, Rayner et al. 2003; Lantz et al. 2004) to range between 30 and 50 Å as indicated by Mann et al. (2013, 2015) (which corresponds to around 429–829 km s^{-1} using Equation (A3) of Zhang et al. 2021a). We simply choose uniform priors on l to comfortably accommodate this range as well as those prescribed by Zhang et al. (2021a) when working with SpeX 0″.8 and 0″.5 slits. In a future analysis, fine-tuning the inferences on the wavelength scale parameter (l) as well as the amplitude of the Gaussian process kernel $\log(a_G)$ will help in quantifying the impact of other possible sources of systematic noise that could influence the residuals.

2.5. Likelihood Evaluation

With a grid-model fitting process, it is straightforward to consider a likelihood function evaluated between the model spectrum and data assuming the residuals are pixel-independent, meaning that the covariance matrix describing the residuals only has diagonal terms, reflecting the measurement uncertainty in the pixel. The remainder of the uncertainty is captured with an additional free parameter simply inflating the error bars. The true uncertainty budget, however, can be properly characterized by using a nontrivial covariance matrix to capture correlated residuals within a likelihood function to be maximized, thus accounting for pixel-to-pixel variation and quantifying the probability of the data given the model, where large residuals are penalized. In essence, following Equation (7) of Czekala et al. (2015a), the log-likelihood function evaluated within *Starfish* is as follows:

$$\ln p(D|M) = -\frac{1}{2}(R^T C^{-1} R + \ln \det C + N_{\rm pix} \ln 2\pi). \quad (1)$$

Here, R is the residual spectrum, C is the $N_{\rm pix} \times N_{\rm pix}$ covariance matrix, which is a sum of two covariance matrices, and $N_{\rm pix}$ is the noise in each element of the residual spectrum. This first nontrivial covariance matrix comprises the measurement uncertainty in each pixel/element of the residual spectrum and the covariance between each pair of these pixels. In practice, *Starfish* constructs the covariance matrix by parameterizing with a Gaussian process (GP) kernel that describes the individual covariances between pairs of pixels or
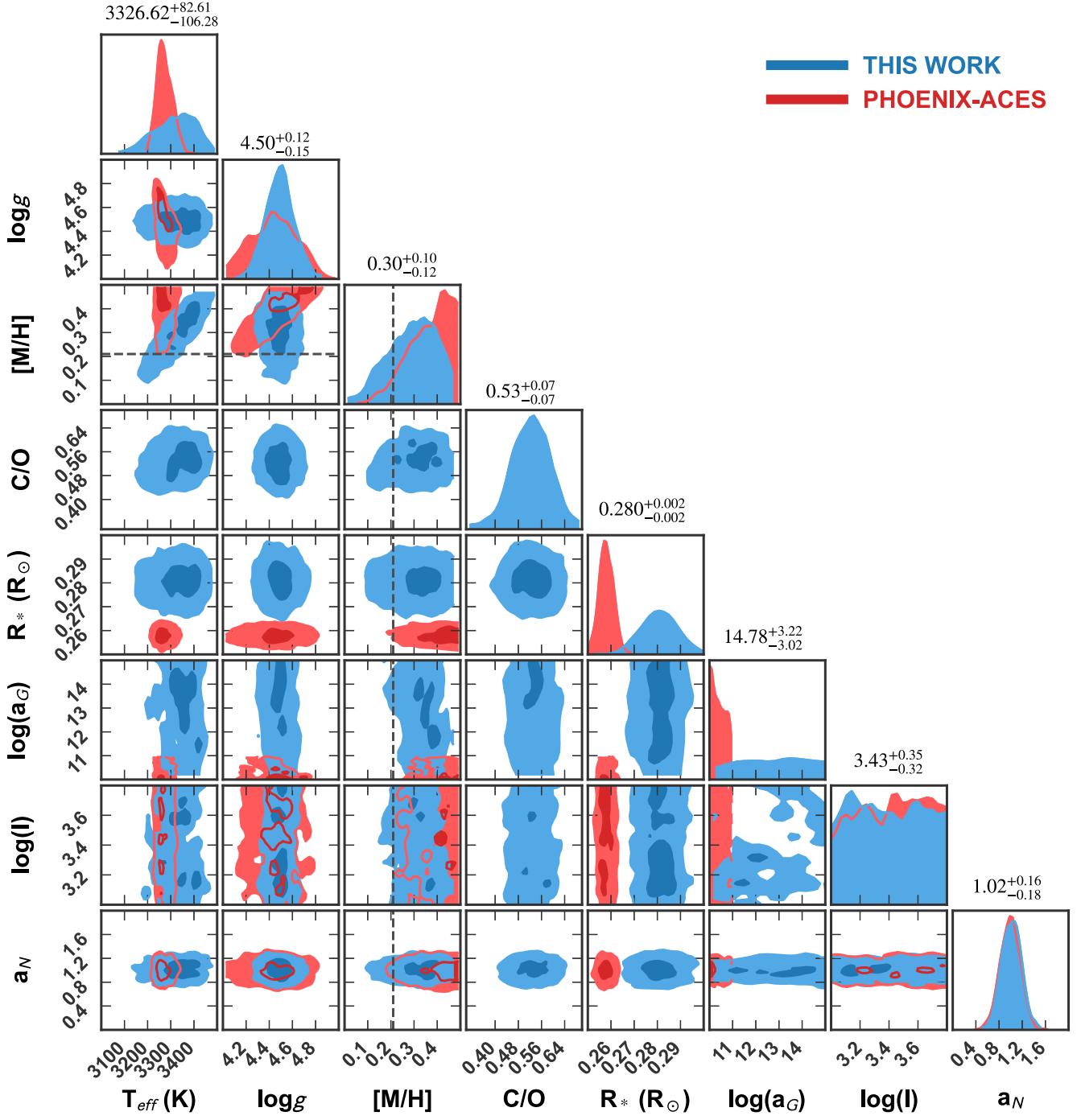
Figure 4. Grid-model fit retrieval results (blue) for NLTT 3725, a widely separated binary G+M3.9, solving for fundamental stellar properties ($T_{\rm eff}$, log g, [M/H], C/O, and R_*) using our model grid compared to PHOENIX-ACES (red) (Husser et al. 2013). Here we find that metallicity values retrieved using our new model grid provide a 1σ improvement in the constraints when compared to empirically derived metallicity ([M/H] = 0.21) (Mann et al. 2013) shown by the dashed vertical line. In addition to physical stellar parameters we also show fits for *Starfish* hyperparameters (log(a_G), log(l), and a_N), which provide additional confidence to our values while incorporating proper sources of data and model uncertainties (see Section 2.3). Our best-fit median and uncertainty values are provided in the inset text (see Table 2).

wavelength bins (via the noise and global covariance matrices with their associated priors described in Section 2.4). Additionally, model deficiencies from assumed physics are taken into account in Equation (1) via mapping differences spanning the model grid (through PCA decomposition) when constructing the emulator object for interpolation (described in Section 2.3). In other words, every model call that is a reconstructed/interpolated spectrum is defined in the following way, as per Equation (23) of Czekala et al. (2015a):

$$f_\lambda(\theta_{*_i}^{\rm grid}) \approx \xi_\mu + \xi_\sigma \sum_{k=1}^{m} w_k(\theta_{*_i}^{\rm grid})\xi_k \qquad (2)$$

where f_λ is an interpolated model, ξ_μ is the mean spectrum of the entire synthetic model grid, ξ_σ is the standard deviation spectrum, and w_k is the weight associated with the kth eigenspectrum. The interpolation/reconstruction uncertainty

Table 2
Fundamental Stellar Parameters Retrieved from Our Grid Models for Targets Used for This Analysis: 10 Widely Separated M+G Binaries and 10 Targets with Interferometrically Measured Angular Diameters

Target	$T_{\rm eff}$ (K)	$\log(g/{\rm cm\ s^{-2}})$	[M/H]	C/O	R_* ($R_\odot$)
Wide Binary Sample (WBS)					
NLTT 11270	$3900.00^{+56.28}_{-58.19}$	$5.32^{+0.15}_{-0.17}$	$-0.24^{+0.14}_{-0.13}$	$0.36^{+0.06}_{-0.06}$	$0.48^{+0.01}_{-0.01}$
NLTT 11500	$3621.12^{+77.27}_{-61.72}$	$4.92^{+0.20}_{-0.21}$	$-0.12^{+0.13}_{-0.13}$	$0.41^{+0.08}_{-0.06}$	$0.61^{+0.10}_{-0.11}$
NLTT 3725	$3326.62^{+82.61}_{-106.28}$	$4.50^{+0.12}_{-0.15}$	$0.31^{+0.10}_{-0.12}$	$0.53^{+0.07}_{-0.07}$	$0.280^{+0.002}_{-0.002}$
NLTT 738	$3630.88^{+117.76}_{-73.35}$	$4.65^{+0.34}_{-0.26}$	$0.05^{+0.13}_{-0.15}$	$0.51^{+0.10}_{-0.10}$	$0.28^{+0.04}_{-0.04}$
NLTT 8787	$3874.58^{+58.35}_{-94.11}$	$5.28^{+0.15}_{-0.24}$	$0.09^{+0.13}_{-0.14}$	$0.39^{+0.07}_{-0.06}$	$0.34^{+0.03}_{-0.03}$
Gl 100	$3570.98^{+77.93}_{-89.81}$	$4.82^{+0.21}_{-0.17}$	$-0.08^{+0.07}_{-0.08}$	$0.34^{+0.08}_{-0.03}$	$0.30^{+0.03}_{-0.03}$
Gl 105	$3466.23^{+56.70}_{-62.52}$	$4.80^{+0.12}_{-0.15}$	$0.18^{+0.06}_{-0.05}$	$0.32^{+0.03}_{-0.01}$	$0.29^{+0.12}_{-0.11}$à
Gl 118.2	$3316.50^{+97.03}_{-106.3}$	$4.77^{+0.19}_{-0.13}$	$0.34^{+0.09}_{-0.13}$	$0.54^{+0.06}_{-0.07}$	$0.10^{+0.05}_{-0.05}$
Gl 81.1	$3996.01^{+2.60}_{-4.45}$	$5.24^{+0.01}_{-0.01}$	$0.00^{+0.03}_{-0.03}$	$0.69^{+0.01}_{-0.01}$	$0.71^{+0.02}_{-0.02}$
NLTT 10349	$3891.32^{+65.43}_{-69.55}$	$5.34^{+0.11}_{-0.17}$	$-0.06^{+0.10}_{-0.11}$	$0.36^{+0.08}_{-0.04}$	$0.42^{+0.02}_{-0.02}$
Interferometric Sample (IS)					
Gl 436	$3633.71^{+41.84}_{-47.42}$	$5.19^{+0.12}_{-0.13}$	$0.18^{+0.05}_{-0.07}$	$0.36^{+0.05}_{-0.04}$	$0.44^{+0.03}_{-0.03}$
Gl 880	$3860.01^{+27.21}_{-44.23}$	$4.77^{+0.13}_{-0.19}$	$-0.03^{+0.13}_{-0.07}$	$0.42^{+0.10}_{-0.07}$	$0.57^{+0.01}_{-0.01}$
Gl 699	$3450.15^{+32.04}_{-62.51}$	$4.97^{+0.13}_{-0.16}$	$0.17^{+0.06}_{-0.09}$	$034^{+0.05}_{-0.03}$	$0.19^{+0.002}_{-0.002}$
Gl 526	$3730.56^{+18.40}_{-20.04}$	$5.33^{+0.06}_{-0.08}$	$-0.33^{+0.04}_{-0.04}$	$0.32^{+0.02}_{-0.01}$	$0.52^{+0.01}_{-0.01}$
Gl 687	$3497.45^{+68.15}_{-71.32}$	$5.0^{+0.20}_{-0.25}$	$0.09^{+0.10}_{-0.12}$	$0.35^{+0.05}_{-0.03}$	$0.41^{+0.01}_{-0.01}$
Gl 887	$3854.64^{+31.69}_{-54.94}$	$5.27^{+0.17}_{-0.23}$	$-0.08^{+0.11}_{-0.10}$	$0.38^{+0.07}_{-0.05}$	$0.49^{+0.001}_{-0.001}$
Gl 411	$3656.81^{+24.31}_{-31.92}$	$5.26^{+0.14}_{-0.15}$	$-0.2^{+0.11}_{-0.06}$	$0.33^{+0.03}_{-0.02}$	$0.38^{+0.001}_{-0.001}$
Gl 581	$3384.15^{+67.20}_{-52.20}$	$5.01^{+0.21}_{-0.13}$	$0.10^{+0.10}_{-0.13}$	$0.36^{+0.05}_{-0.03}$	$0.30^{+0.001}_{-0.001}$
Gl 725A	$3563.91^{+23.70}_{-36.89}$	$5.27^{+0.12}_{-0.20}$	$0.14^{+0.08}_{-0.10}$	$0.33^{+0.03}_{-0.02}$	$0.35^{+0.001}_{-0.001}$
Gl 725B	$3257.01^{+44.80}_{-34.33}$	$5.04^{+0.30}_{-0.34}$	$0.01^{+0.15}_{-0.20}$	$0.37^{+0.07}_{-0.05}$	$0.28^{+0.001}_{-0.001}$

is thus enveloped in a second nontrivial covariance matrix, which is parameterized yet again with GP kernels that describe the probability distribution of weights associated with the PCA eigenspectra. These weights are modeled with joint multivariate Gaussian distribution kernels spanning each dimension of the synthetic grid ($T_{\rm eff}$, $\log g$, [M/H], and C/O in our case), thereby capturing reconstruction/interpolation residuals for each wavelength bin as well as between the bins, in addition to downweighting grid deficiencies. Finally, the likelihood function is evaluated over the defined priors for both the physical parameters and hyperparameters describing the properties of the GP kernel.

3. Results

We first validate our grid models for both groups of M-dwarf spectra: 10 widely separated M+G binaries with empirically measured [M/H] (WBS) and 10 M dwarfs with interferometrically measured angular diameters (IS). All spectra are *H*-band flux-calibrated and downgraded in resolution to $R \sim 120$ for the entire SpeX SXD bandpass from 0.8 to 2.4 μm to focus solely on the low-resolution regime. For all targets, we perform grid-model fits using the routine mentioned in Section 2 and retrieve fundamental stellar parameters ($T_{\rm eff}$, $\log g$, [M/H], C/O, and R_*, shown in Table 2), while also accounting for the *Starfish* hyperparameters to accurately model errors. We use data from SpeX + SNIFS stitched together for our WBS targets from Mann et al. (2013). We also perform fits using the PHOENIX-ACES (Husser et al. 2013) model grid through our fitting pipeline and find our constraints to be improved overall when compared to the empirically measured values of [M/H] (see Figures 4–6). For six out of 10 targets, our inferred

metallicities are closer to the observational values in units of σ of the retrieved histograms (see Figure 5). For instance, the most significant improvement is for M0.3 target NLTT 112700, where [M/H] derived from our grid model is shifted 0.5σ away (the derived median lies within 1σ of the observations) whereas the value derived by PHOENIX-ACES (Husser et al. 2013) is shifted over 10σ away from the empirical constraint (Mann et al. 2013). Similarly, for M3.9 target Gl 105, we see about a 3σ improvement using our grid models. For Gl 118.2 (M3.8), NLTT 3725 (M3.9), and NLTT 8787 (M1.6), we collectively see about a 1σ improvement. For NLTT 738 (M2.2), we see no significant difference in the retrieved metallicities, where both model values fall under 0.5σ of the observations. However, for Gl 100 (M2.9), Gl 81.1 (M0.1), and NLTT 11500 (M1.8), we see that the PHOENIX-ACES models fare better. For two of the targets in this group (NLTT 10349 and NLTT 11500), we perform additional analyses in the next section to understand the effects of photospheric activity on model-derived [M/H] values.

The overall fits using our model grid and PHOENIX-ACES (Husser et al. 2013) perform well where each target best-fit residual is roughly $\sim30\%$. However, our inferred [M/H] provide for tighter constraints, owing to the inclusion of up-to-date opacities—especially with key molecules that dominate M-dwarf atmospheres such as metal oxides (H_2O, TiO, VO) and metal hydrides (MgH, CaH, CrH, TiH, AlH, and FeH) from the *EXOPLINES* database (Gharib-Nezhad et al. 2021a).

For targets in our second sample group of M-dwarf spectra: 10 IS with measured angular diameters (Boyajian et al. 2012), we focus on validating our model grid by inferring mainly the stellar radius, in addition to other fundamental stellar parameters such as $T_{\rm eff}$, $\log g$, [M/H], and C/O. We perform fits using the same grid-model fitting pipeline with our new
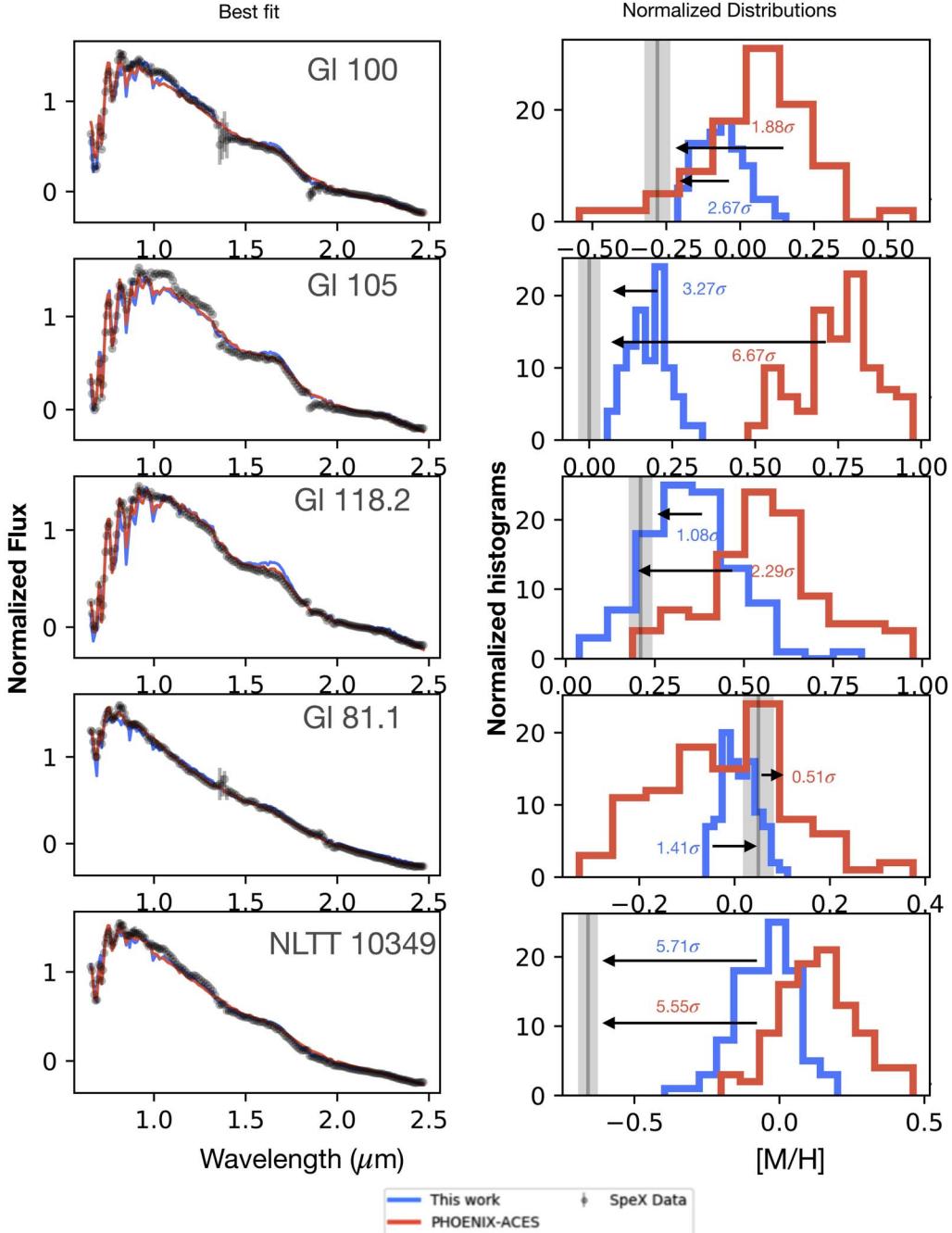
Figure 5. Results for our WBS targets validating with empirically measured metallicity values (Mann et al. 2013). In the left column we show the best-fit spectra for the low-resolution data (Rayner et al. 2003; Lantz et al. 2004; Mann et al. 2013) using SPHINX, our new M-dwarf atmosphere model grid (blue) vs. PHOENIX-ACES (Husser et al. 2013) (red) through the same grid-model fitting pipeline. For each target in the right column, we show posterior probability distributions of [M/H] from our model grid (blue) and from PHOENIX-ACES (Husser et al. 2013) (red). In the inset of these histogram plots, we show in text the deviation (in units of σ width of each histogram) of the retrieved median [M/H] value compared to the empirical values from Mann et al. (2013), and their uncertainty is shown by gray vertical lines. It is worthwhile to note that infrared spectroscopy is subject to significant telluric absorption, especially near the edges of the J, H, and K bands. To calibrate and correct for the absorption, typically a standard star is observed at nearly the same airmass. Nevertheless, the calibration can result in under- or over-correction, which results in systematic errors near the edges of the bands. This may be the reason for the discrepancies seen, for example, in spectra of Gl 100 (top row, left) or NLTT 8787 (bottom row). Accounting for this within the *Starfish* framework (Czekala et al. 2015a), we show improvement in acquired [M/H] constraints using our new model grid over PHOENIX-ACES (Husser et al. 2013) for a majority of targets in this sample.

stellar atmosphere models used for the WBS group (also described in Section 2). Here, we find that our grid-model retrieved radii are very close to interferometrically derived values (Boyajian et al. 2012). In Figure 7, we see that one out of 10 targets—Gl 581 (M3.2)—falls under 1σ, three out of 10 targets—Gl 436 (M2.8), Gl 725A (M3.0), and Gl 687 (M3.2)—

fall within 2σ of the interferometrically retrieved R_*, and an additional two targets—Gl 526 (M1.4) and Gl 880 (M3.7)—fall within 3σ. Note that the observational radii depicted by gray vertical lines in Figure 7 are taken from Boyajian et al. (2012 and references within); however, some targets have been provided with a weighted mean from multiple observations.
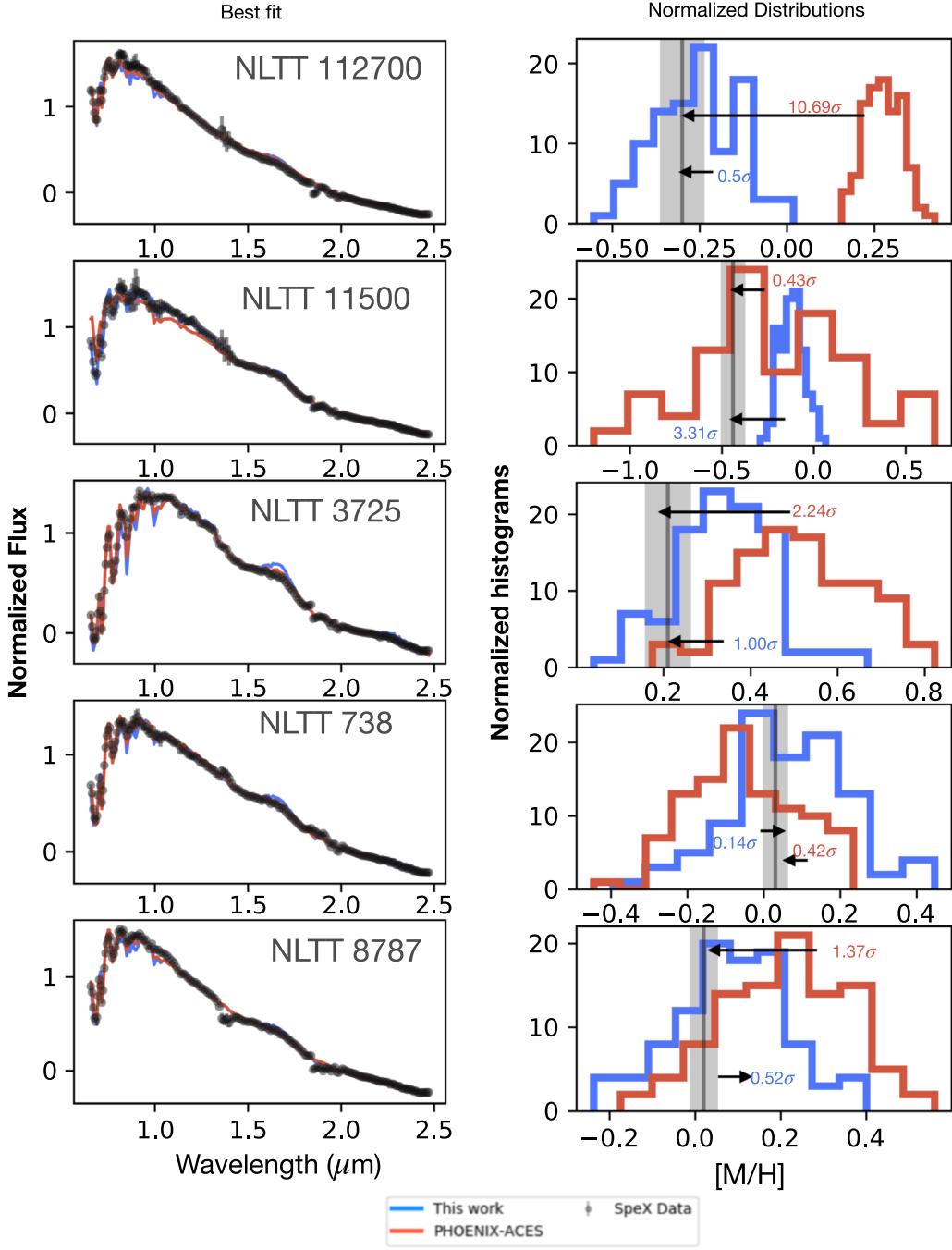
Figure 5. (Continued.)

For Gl 887 (M0.5) specifically, we find that our retrieved radius value is about 9σ off from the mean interferometric radius of 0.4712 ± 0.0086 (Boyajian et al. 2012), but it is closer to 0.4910 ± 0.00140 acquired by Segransan et al. (2003), which is our adopted observational value in Figure 8.

In addition to the retrieved radii, we also compare our grid-model retrieved effective temperatures of these targets to observations (Boyajian et al. 2012 and references within). The overall trend is consistent without any significant correlations with [M/H] (shown in the color bar), at least for our small sample of M dwarfs. We also find that our $T_{\rm eff}$ values are systematically overestimated for a majority of our sample, and R_* deviates slightly from observations only for two targets (particularly Gl

725B or Gl 526, for example). We discuss some possible explanations for these trends in the following section. We also show in Figure 9 that our model retrieved $T_{\rm eff}$ and R_* are overall comparable to values acquired both from Mann et al. (2015) using the BT-Settl version of PHOENIX (Allard 2014) and from interferometric observations done by Boyajian et al. (2012) using the CHARA array (ten Brummelaar et al. 2005) with average scatters of 4% and 5.2%, respectively. Additionally, we also confirm that they are drawn from the same parent distribution with 95% and 99% confidence using a two-sample Kolmogorov–Smirnov (KS) test for cumulative distribution functions of each data set. We also compare literature medians gathered from Passegger et al. (2022) for six targets that are in our IS sample and
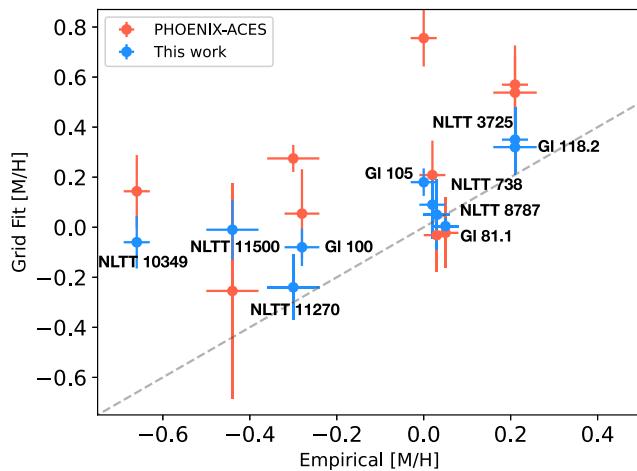
Figure 6. Comparison of our metallicity values (blue) against empirically derived values (Mann et al. 2013). We also contrast against values derived using the PHOENIX-ACES grid models (Husser et al. 2013) (red) within our grid-model fitting routine. Our models fare well overall for all the M+G binary targets in our sample, improving constraints well within 1σ for 60% of targets within our sample. The dashed gray line represents a slope of 1.

five targets from Souto et al. (2022) for T_{eff} and from Souto et al. (2020) for R_* using spectral fitting of APOGEE high-resolution H-band spectra (Majewski et al. 2017) and confirm again with 99% and 95% confidence that these are drawn from the same parent population as our derived radii and effective temperatures.

For our combined sample of 20 M dwarfs, we also investigate spectral fitting residuals from *Starfish* (Czekala et al. 2015b, 2018) similar to Zhang et al. (2021b). This allows us to explore model deficiencies or problematic observational contributions and how they affect spectral residuals as a function of wavelength. In Figure 10, we first stack all the data–model residuals (top panel, in black) from fits of each target normalized by their respective observed peak I-band flux. Overplotted in the same panel are stacked 1σ and 2σ dispersions over 5000 random draws from the *Starfish* (Czekala et al. 2015b, 2018) composite covariance matrix, also normalized by the I-band peak flux of each target. In the bottom panel of the same figure, we show one spectrum from each group of observations (WBS and IS) as representative of that group for reference. The residuals from all fits do not exceed 35%; however, they show a significant wavelength-dependent trend with prominent features pointing to data stitching and telluric contamination at specific bands. We also see that the largest deviations appear at shorter wavelengths ($<1\,\mu$m), a region filled with bulky molecular opacity features (see Figure 2), which shows possible sensitivity to model assumptions as well —such as considering a uniform non-heterogeneous photosphere or picking a solar-calibrated mixing length value for convection for low-mass stars, or the lack of pressure-broadening data on key M-dwarf molecules and their influence on spectra (as discussed in Gharib-Nezhad et al. 2021a). While it appears challenging to decouple and quantify the individual contribution of model or data systematics, we can establish confidence in our retrieved fundamental stellar parameter values as we validate our models with benchmark M dwarfs while accounting for this combination of error sources in a robust manner.

4. Discussion

In this work, we introduce a 1D self-consistent radiative–convective thermochemical equilibrium model atmosphere grid to characterize M dwarfs, focusing purely on low resolution. We set out with the goal to infer fundamental stellar properties for these exoplanet-hosting M dwarfs, and we show that we are able not only to acquire reasonable estimates for the stellar effective temperature, surface gravity, metallicity, and radius but also to extend to estimating their carbon-to-oxygen ratio with improved constraints. Knowing the C/O ratio of an exoplanet-hosting star with increasing accuracy has vast implications in refining planet formation theories and estimating the C/O ratios of orbiting planets themselves, as they are necessary initial conditions for disk chemistry models (Fortney 2012).

Low-mass stars, such as M dwarfs in particular, fall within the transition region from radiative to fully convective interiors around 0.28–0.33 $M_\odot$ (Chabrier & Baraffe 1997). Around this region of the main sequence, numerous challenges lie for both observational and theoretical efforts. We find for our WBS sample of 10 widely separated binary companion M dwarfs that our retrieved bulk [M/H] values are consistent with empirically calibrated values (for 60% of our sample); however, there are a few deviations that leave room for further exploration. We also find trends consistent with the findings of Mann et al. (2015), Boyajian et al. (2012), and Dieterich et al. (2021) for our IS targets with interferometric measurements, where T_{eff} is overestimated for a majority in our sample. We perform a few tests to investigate these trends further.

4.1. Influence of Photospheric Heterogeneity

First, we assess the influence of photospheric heterogeneity in the retrieved abundances. To answer this question, we include a parameterization (Rackham et al. 2017, 2018; Apai et al. 2018; Pinhas et al. 2018; Zhang et al. 2018; Iyer & Line 2020) for the stellar spectrum to be a linear combination of two spectra—one representative of the calm photosphere and one representative of a cooler spot region on the photosphere, which are both weighted by the fractional coverage in area of the spots:

$$F_* = F_{phot}(1 - f_{spot}) + F_{spot}f_{spot} \qquad (3)$$

where F_* is the final stellar model spectrum, F_{phot} is the spectrum of the stellar photosphere region alone, F_{spot} is the spectrum of the stellar spot region alone, and f_{spot} is the fractional coverage in area of the spots. Here, both F_{phot} and F_{spot} are model spectra as functions of their respective T_{eff}, $\log g$, [M/H], C/O, and scaled $(R_*/d)^2$ values evaluated (generated with the *Starfish* interpolation emulator, see Section 2.3) at each instantiation within our retrieval. All parameters are sampled as a single instance from their respective posteriors for each likelihood evaluation within our routine, except for the effective temperature T_{eff} alone, which is drawn twice: as T_{phot} for the temperature of the photosphere region and as T_{spot} for the temperature of the spot region, to ultimately produce the final model spectrum F_*. Following this, we end up with three additional parameters for our grid-model retrievals, namely, T_{phot}, T_{spot}, and f_{spot}. Both the temperature values are bound by uniform priors spanning the range of our model grid (but with the condition that T_{spot} is
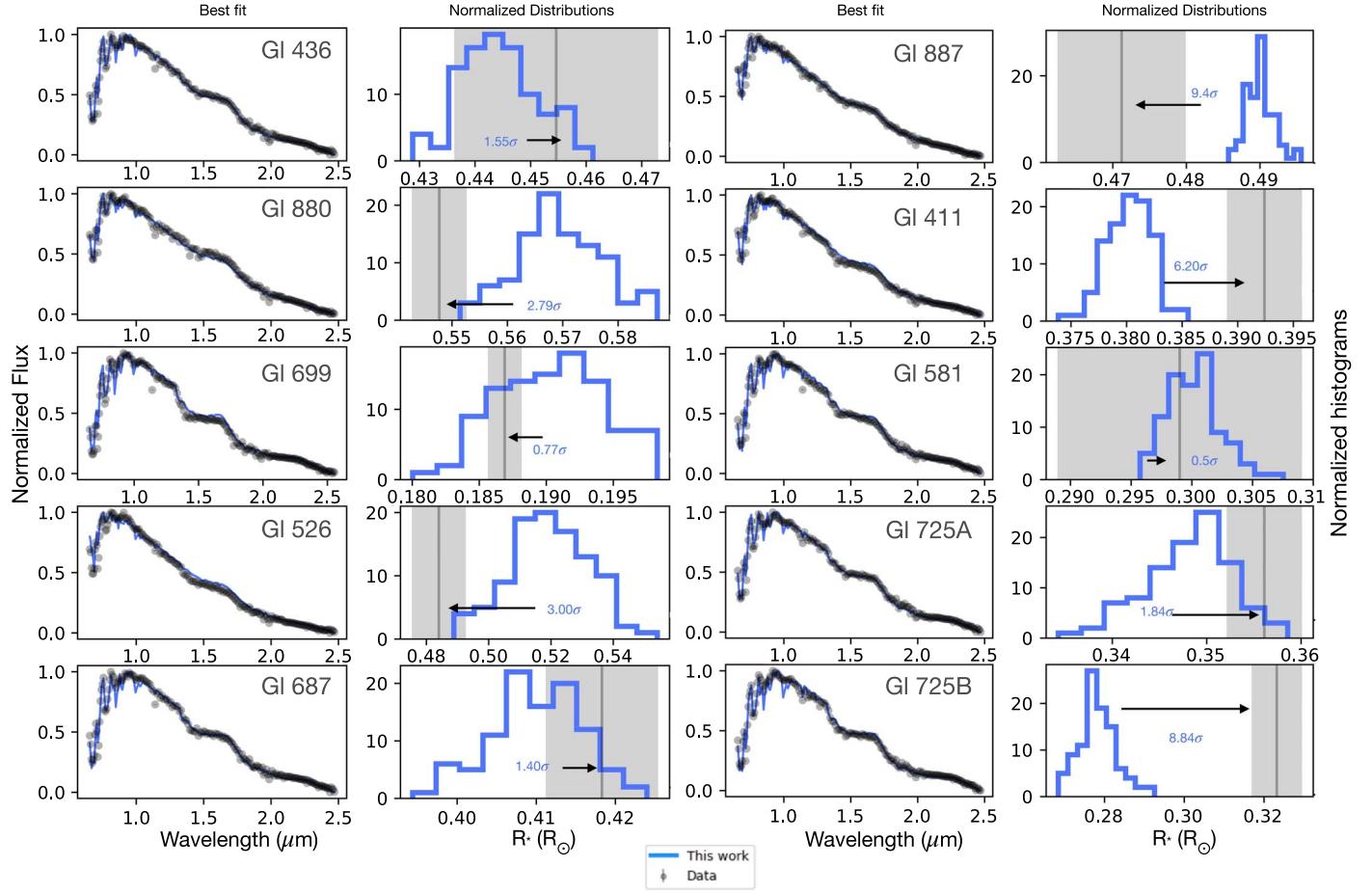
Figure 7. Results for our IS targets with interferometrically measured angular diameters (Boyajian et al. 2012). In the first and third columns we show best-fit spectra for the low-resolution data (Rayner et al. 2003; Lantz et al. 2004; Mann et al. 2015) using our new M-dwarf atmosphere model grid (blue) through our grid-model fitting pipeline. In the second and fourth columns for each adjacent target we show posterior probability distributions of R_* from both fits, corresponding to the same color. In the inset of these histogram plots, we show in text the deviation (in units of σ width of each histogram) of retrieved median R_* value compared to the observed value and uncertainty shown by the gray vertical line (Boyajian et al. 2012).

always lower than T_{phot}), and f_{spot} is taken as a uniform prior between 0 and 1 because we do not have abundant a priori knowledge of physically reasonable spot covering fractions on M-dwarf photospheres (Iyer & Line 2020).

We perform retrievals including this "mixed model" parameterization for two targets with the largest σ deviation from observations in Figure 5, NLTT 10349 (M1.0) and NLTT 11500 (M1.8). In Figure 11, NLTT 10349 shows an improvement from 3σ to 2σ in our ability to constrain [M/H] relative to the empirically derived value. At the same time, NLTT 11500 does not show any further improvements in [M/H] upon the inclusion of stellar spots. However, the retrieval does appear to be artificially pushing the spot covering fraction to be significantly high. We also see that the log g posterior is driven up to higher values, necessitating further investigations while potentially exploiting the mass/metallicity correlation for a mid-M dwarf (see Figure 11).

We also test the influence of including stellar spot parameterization on the retrieved radii. For the case of Gl 725B (M3.5)—a target in our sample with the most significant deviation from the interferometric measurements (IS)—we find that the inclusion of stellar heterogeneity ever so slightly improves accuracy in the retrieved radius. The retrieved effective temperature, however, is significantly more

overestimated than before, with a significant deviation from interferometry. This could be due to a simplified stellar spot parameterization causing degeneracies with other fundamental stellar properties, which do not always prove to be accurate for active M dwarfs (Iyer & Line 2020) (extra figures provided in our zenodo link). It is also well established by now that there is very little knowledge of physically reasonable coverage fractions of stellar photospheric heterogeneities or of their distribution throughout the photosphere of an M dwarf, which adds to numerous problems that exist in characterizing the active environment of such stars (Rackham et al. 2022). Spots occur in stars primarily due to pockets of magnetic flux tubules stretching outwards beyond the photosphere, suppressing the flux in that region as a result of convection processes that produce an overall cooler effect in this region. Therefore, efforts to model this more accurately, including complementary efforts from the observational side, could provide a great starting point for fine-tuning priors on such correlated parameters and improving parameterizations within a retrieval framework.

4.2. Influence of the Convective Mixing Length Parameter

Here we explore the influence of the convective mixing length parameter (Böhm-Vitense 1958) on values retrieved
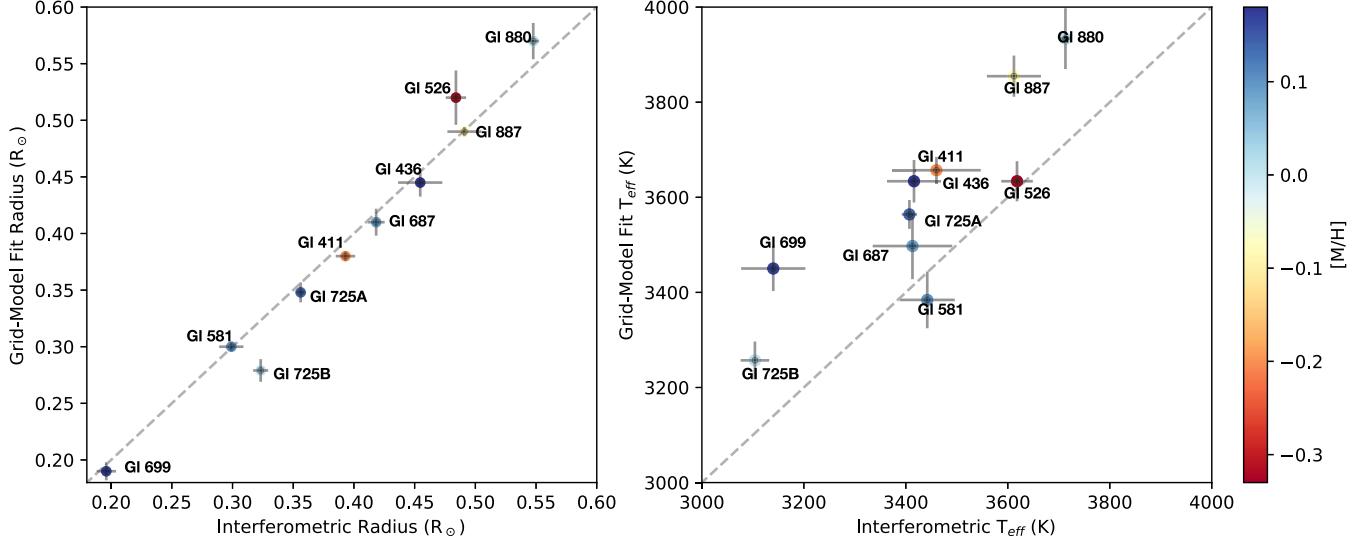
Figure 8. Comparison of our retrieved radius (left) and effective temperatures (right) for the IS targets to those measured and derived interferometrically using the CHARA array (Boyajian et al. 2012). Here, we show that our radii values are quite consistent with the observations, but our grid-model-derived effective temperatures are overestimated. We also see here that there is no correlation with the derived metallicities of these stars (shown in the color bar) to explain any of the deviations from observed values. We explore the effects of convective mixing length and photospheric heterogeneity to possibly explain some of these trends (see Section 4).
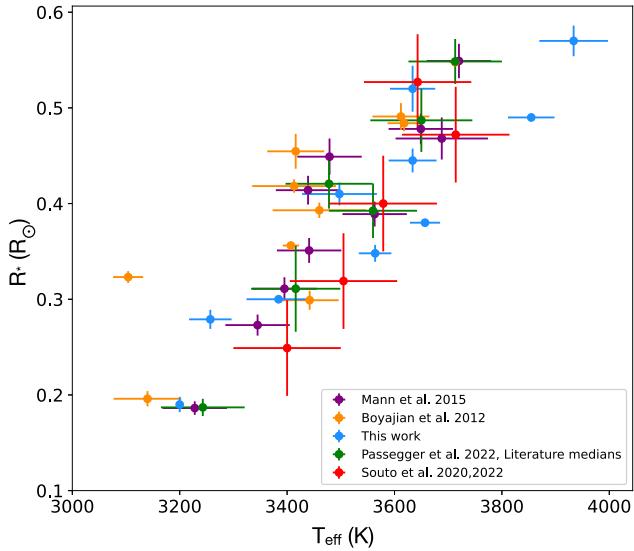
Figure 9. Comparison of our model-derived T_{eff} and R_* for the IS group targets. We overplot values derived from other works (Boyajian et al. 2012; Mann et al. 2015; Souto et al. 2020, 2022; Passegger et al. 2022) and show that our grid-model retrieved values are consistent within the scatter, and a KS-2 sample test confirms they are drawn from the same parent distribution.

from our stellar atmosphere model. We define the mixing length parameter as

$$\alpha = \frac{l_{mix}}{H_p} \qquad (4)$$

where l_{mix} is the convective mixing length and H_p is the pressure scale height; the ratio α is a free parameter that is often based on solar-calibrated values (Chabrier & Baraffe 1997). All our models in the grid in Section 2 are calculated assuming $\alpha = 1$ for simplicity. In Figure 8, we find that most of our retrieved radii are consistent with the interferometry values, except that for Gl 725B (M3.5) is underestimated. In the same figure (right panel), we also find that their effective

temperatures are overestimated relative to the interferometry values. To address this, we perform a test by changing the mixing length value to $\alpha = 0.5$ and generate a mini-grid of models to assess whether it could improve our predictions and explain the systematic deviations seen in both stellar radius and effective temperature for this target. We choose this value as recommended by Mann et al. (2015), who found that reducing the mixing length parameter yielded a reduction in their T_{eff} values of about 2% and an increase in R_* of 2.5% for low-mass stars $< 0.3\,M_\odot$. This is also consistent with findings of Barklem et al. (2002) using the MARCS model (Gustafsson et al. 2008), where simultaneous fits to Hα and Hβ in spectra of metal-poor dwarf stars were sensitive to the mixing length value assumed in the model ($\alpha = 0.5$ instead of 1.5), which varied the effective temperature by up to 40 K.

We also repeat this exercise with a value of $\alpha = 2$ following Baraffe et al. (2015), who primarily favor mixing length values above 1.6 for cooler and denser stars with larger convective envelopes. [M/H] could potentially be sensitive to the choice of mixing length based on relevant model calibrations; however, Song et al. (2020) find that the typical scatter in [Fe/H] values due to α is well within the uncertainty levels of metallicity itself.

We also pick Gl 699 for this exercise (Barnard's star, which is known to have subsolar metallicity, [Fe/H] = -0.36, Rojas-Ayala et al. 2012), where our model appears to derive a supersolar [M/H] (0.17) and overestimated T_{eff} (3450 K, see Figure 8). In Figure 12, the model grid assuming $\alpha = 0.5$ fares better than the $\alpha = 2$ models in terms of the retrieved effective temperature, getting it closer to the interferometry value within a shift of 1σ and 2σ from the median respectively for Gl 725B and Gl 699. We do not see any changes in the retrieved radii for either case; however, the retrieved [M/H] value changes to better match the empirically derived value from Mann et al. (2015) under 1σ for Gl 725B and 2σ for Gl 699 compared to observations. In particular, the median retrieved [M/H] with the $\alpha = 0.5$ model results in a subsolar metallicity for Barnard's star as expected. We also do not see a significant difference in
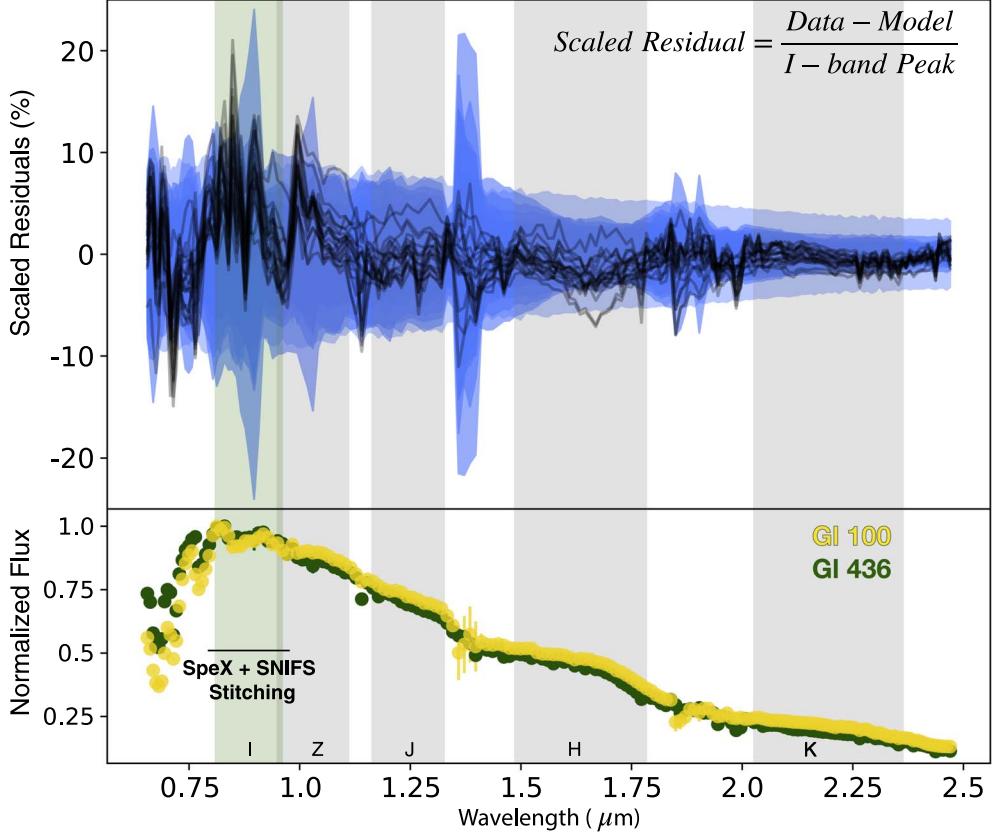
Figure 10. Top: stacked residuals from our spectral grid-model fits for all targets in our samples (black), including the widely separated binary sample (WBS) and interferometry sample (IS) with a total of 20 M dwarfs. The blue shaded regions are stacked 1σ and 2σ regions of 5000 random draws from the *Starfish* (Czekala et al. 2015b, 2018) covariance matrix for each target. Bottom: spectra of Gl 436 (dark green) and Gl 100 (yellow) for reference as representative of both groups within our sample, convolved to $R \sim 120$. Both spectra and residuals are normalized by the observed peak *I*-band flux of each target. In the gray vertical shaded regions we show the filters of *ZJHK* bands defined by the SpeX Instrument (Rayner et al. 2003). The green shaded region shows the SpeX+SNIFS overlap. We see that the largest deviations from our fits occur primarily at shorter wavelengths of $<1\,\mu$m, which could be due to a combination of data discrepancies and model deficiencies.

the retrieved surface gravity parameter, log g, for either case, consistent with findings from Mann et al. (2015) where the mixing length parameter did not appear to perturb derived masses significantly. Hence, for very low-mass targets (<0.3 $M_\odot$) such as Gl 725B and Gl 699, we find that effective temperature and metallicity are sensitive to the mixing length parameter assumed. Reducing α from the solar-calibrated value ($\alpha = 1.88$, Mann et al. 2015) to a lower value ($\alpha = 0.5$) is in essence equivalent to reducing the convective envelope and increasing the radiative interior, at the same time lowering the surface temperature and influencing the molecular opacities in accordance. Consistently, we see a slight reduction in model-derived effective temperatures and metallicities for both Gl 725B and Gl 699 that fall within the fully convective regime ($<0.3\,M_\odot$, Figure 12). For the $>0.3\,M_\odot$ target Gl 436, reducing the mixing length parameter showed a dramatic reduction in effective temperature and metallicity, with no noteworthy change in the radius (see supplementary figures). In line with findings of Mann et al. (2015, see their Table 4), we also find that the relative difference in retrieved effective temperature due to reducing the convective parameter is likely larger for $>0.3\,M_\odot$ M dwarfs. However, analyzing a larger sample and the resulting population-level trends will allow for a more thorough assessment (forthcoming in the second paper of this series).

We also investigated the effects of stellar spots for the case of Gl 725B alone while keeping the mixing length parameter

constant at 1. We found that the model-derived radius is slightly increased by an additional 1σ, thereby shifting the histogram closer to the observational value (supplementary figures). No significant changes occurred in the other derived parameters. Therefore, we recommend that including mixed models that consider the effects of both convection and stellar photospheric heterogeneity could be one way of understanding the nuances that influence retrieved fundamental properties of M dwarfs, particularly when moving toward later spectral types.

4.3. Carbon-to-oxygen Ratio of M Dwarfs

The carbon-to-oxygen ratio is an extremely valuable composition that can inform about the balance of carbon and oxygen species chemistry or refractory elements in atmospheres of cool stars and their possible formation scenarios such as fragmentation (Offner et al. 2010) or core collapse (Gagné et al. 2018). C/O ratios also shed light on the density, formation, and migration history of rocky planets or planetary-mass companions harbored by M dwarfs (Gaidos 2000). Several retrieval techniques using low-resolution NIR spectra have acquired robust C/O constraints for cold brown dwarfs (Line et al. 2015, 2017; Zalesky et al. 2019, 2022; Gonzales et al. 2020); in addition, pEW (Nakajima & Sorahana 2016) and synthetic spectral model techniques (Gizis et al. 2016; Veyette et al. 2016) have been employed for M dwarfs. This
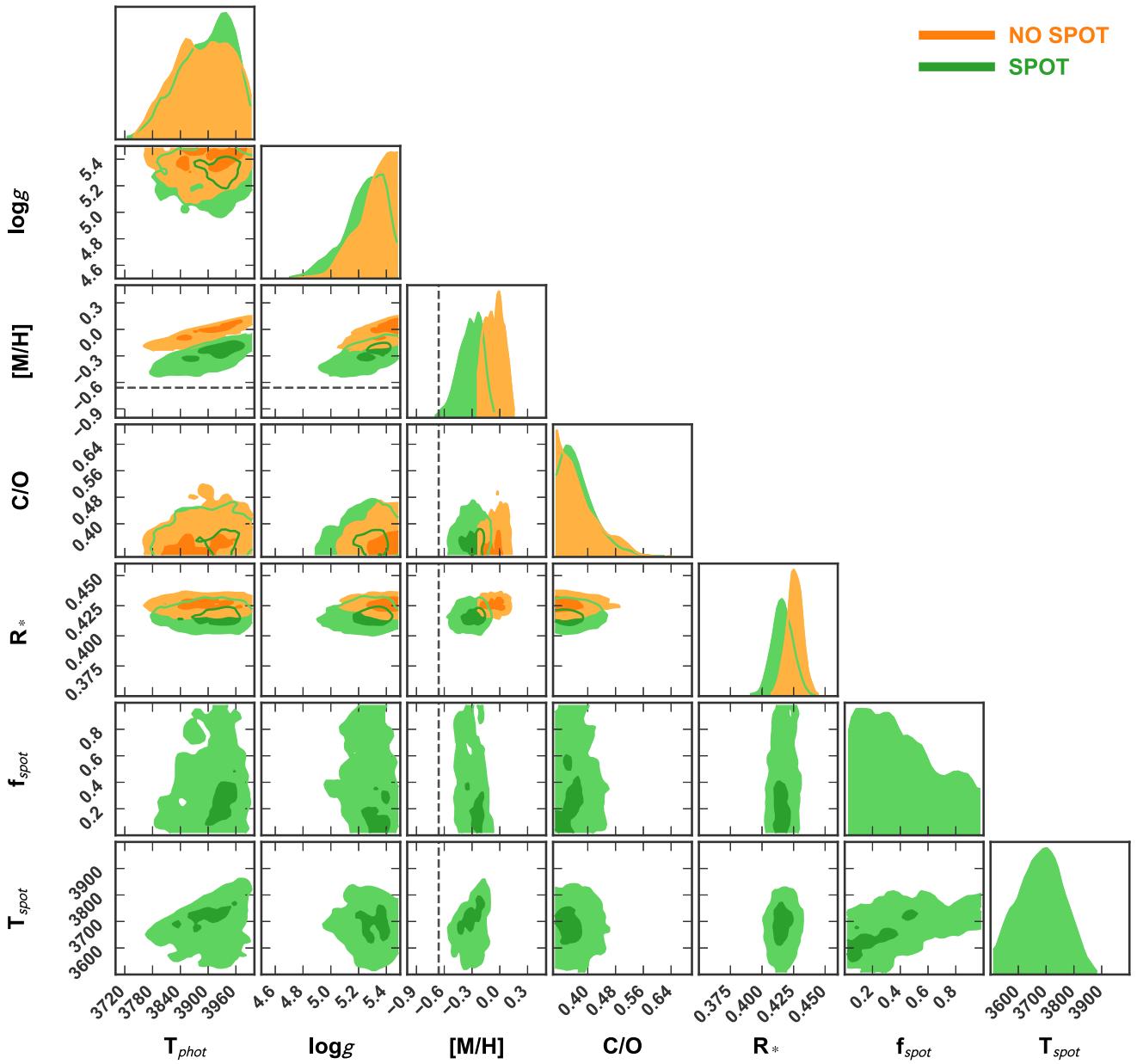
Figure 11. Posterior probability distributions for the two WBS (metallicity calibration) targets with the largest deviations from empirical measurements (see Figure 5) —NLTT 10349 (top) and NLTT 11500 (bottom). Here we compare the results of retrieved stellar properties with (green) and without (orange) including parameterization of stellar surface heterogeneity. We find for NLTT 10349 that we are able to further improve the [M/H] from 3σ–2σ closer to the empirically derived [M/H] shown in inset as the black vertical dotted line. For NLTT 11500, however, there is no such change, except for the retrieved $\log g$. The empirical [M/H] for this target is -0.44 (Mann et al. 2013), which is beyond the range of the plot and therefore not shown.

work, however, is a first in the simultaneous and systematic determination of [M/H] and C/O constraints of benchmark M dwarfs from low-resolution spectra. In Figure 13, we overlay our grid-model-derived [M/H] and C/O values with FGK stars in the local population (Hinkel et al. 2014) (gray), showing overall consistency. Most of our low-mass star sample falls below the solar C/O value of 0.55. We also find that our average C/O constraints of $\sim 10\%$ have not been previously achieved even for main-sequence G stars, specifically in low-resolution ($R \sim 120$) spectra.

Deriving the C/O ratios of FGK stars has been challenging yet feasible due to limited blended lines in the spectra and not accounting for non-LTE corrections (Fortney 2012). Delgado Mena et al. (2021) have computed C/O values for over 1000

FGK stars and found that C/O ratios derived from high-resolution ($R > 100{,}000$) spectra are extremely sensitive to the choice of oxygen line or stellar atmosphere models used to compute them, leading to systematic shifts that need to be scaled with a solar reference value, particularly for cool stars (K type). In even cooler M dwarfs, the bulk of C goes into the CO opacity and the O into metal oxides (TiO/VO) with dominant absorption features. Given that this falls typically in the optical/NIR region of the spectra that is more susceptible to model assumptions as seen from the above analysis, we plan in the future to perform forward model Bayesian retrievals (Gonzales et al. 2020; Zalesky et al. 2022) to directly solve for the radiative–convective atmospheres of M dwarfs without having to rely on self-consistent grid-model assumptions.
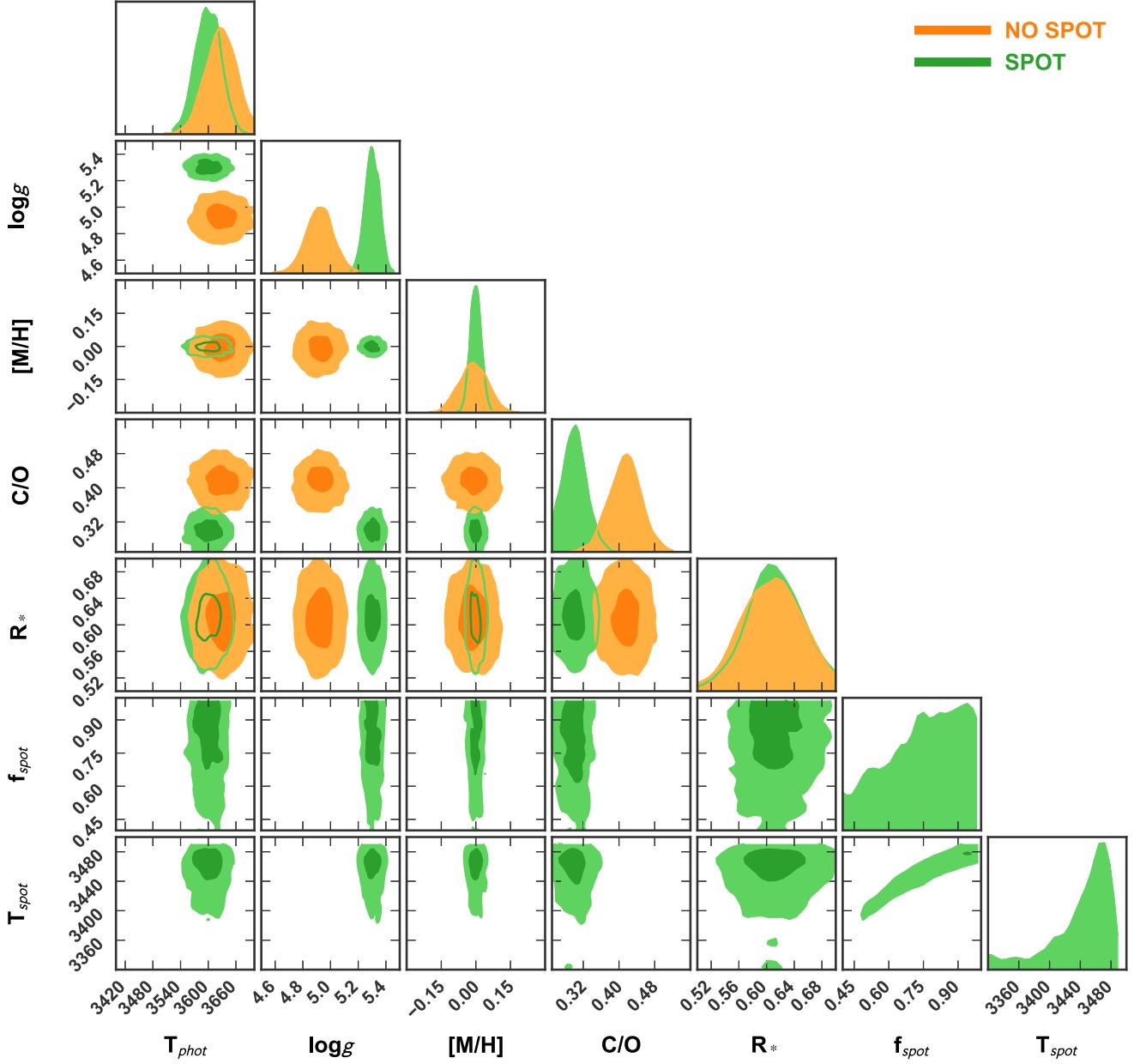
Figure 11. (Continued.)

5. Summary

Numerous challenges plague the path of characterizing M-dwarf atmospheres. In this work, we attempt to address on a fundamental level the need for improved stellar atmosphere models and the implications of working with outdated line lists, which leads to problems that percolate into exoplanet characterization, Galactic and chemical evolution, and ultimately our understanding of the universe. With such far-reaching applications of concern, here are our key conclusions.

1. We present SPHINX: a new 1D radiative–convective thermochemical equilibrium atmosphere model grid to characterize M-dwarf atmospheres, permitting physical plausibility to solve for fundamental stellar parameters—T_{eff}, log g, [M/H], and R_* using a Bayesian inference covariance matrix-error analysis method invoking Gaussian processes (*Starfish*, Czekala et al. 2015b, 2018). We also

include the most up-to-date compilation of molecular absorption cross-section data from *EXOPLINES* (Gharib-Nezhad et al. 2021a).

2. We have benchmarked our stellar model grid with a grid-fitting routine described in Section 2.3 on low-resolution SpeX+SNIFS spectra of 10 widely separated G+M binary targets and evaluated against empirically measured [M/H] (Mann et al. 2013). Mainly, we compare how our model grid fares against PHOENIX-ACES (Husser et al. 2013) and find that we are not only consistent with the widely used model grid but also able to improve the [M/H] constraints relative to observations for 60% of our sample targets.

3. We also benchmark our stellar atmosphere model grid by retrieving R_* and T_{eff} for low-resolution SpeX+SNIFS +STIS spectra of 10 targets with interferometrically measured angular diameters using the CHARA Array

Iyer et al.
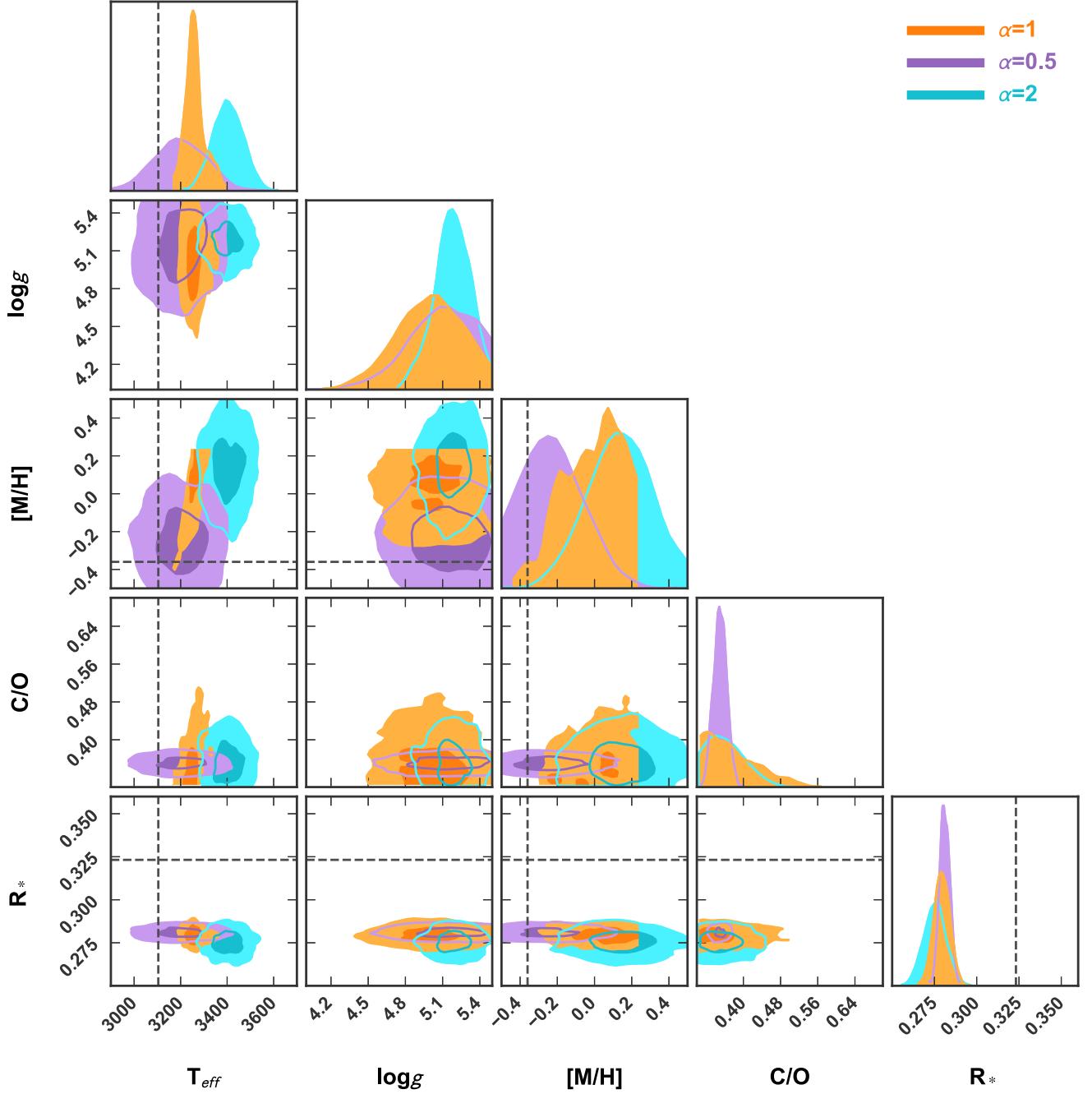
Figure 12. Posterior probability distributions comparing fits for Gl 725B (top) and Gl 699 (bottom) using a mini-atmosphere grid generated for $\alpha = 0.5$ (purple), $\alpha = 1$ (orange), and $\alpha = 2$ (cyan), both targets being from our IS sample (radius calibration). Here we see that the retrieved radii are no different; however, the retrieved $T_{\rm eff}$ and [M/H] derived using the model with lower mixing length parameter ($\alpha = 0.5$) match within 1σ and 2σ for Gl 725B and Gl 699, respectively, compared to observations (Boyajian et al. 2012). At least for these two cases, it appears that a smaller mixing length parameter α seems to be appropriate.

(Boyajian et al. 2012). Here we find that our retrieved parameters are consistent (with 95% and 99% confidence using a two-sample KS test, see Section 3) compared to observations as well as to values derived using BT-Settl models (Allard et al. 2000; Mann et al. 2015) with average scatters of 4% and 5.2% respectively for $T_{\rm eff}$ and R_*. We also find that our derived values are consistent within reasonable scatter with high-resolution spectral fitting as well as with machine learning methods as surveyed and compiled in the form of literature median values by Passegger et al. (2022) (see Figure 9).

4. For all 20 M dwarfs in this study, we extend our grid to provide the stellar C/O ratio, which is extremely valuable for assessing the behavior of the carbon chemistry within the M dwarf as well as serving as a valuable input parameter for models of planetary formation as well as disk evolution (Fortney 2012). We find that simultaneously constrained C/O values using our methodology for M dwarfs are overall consistent with those for neighboring FGK main-sequence stars (see Figure 13). In a future study, we plan to expand our grid to provide N/C and α/M ratios for M dwarfs.
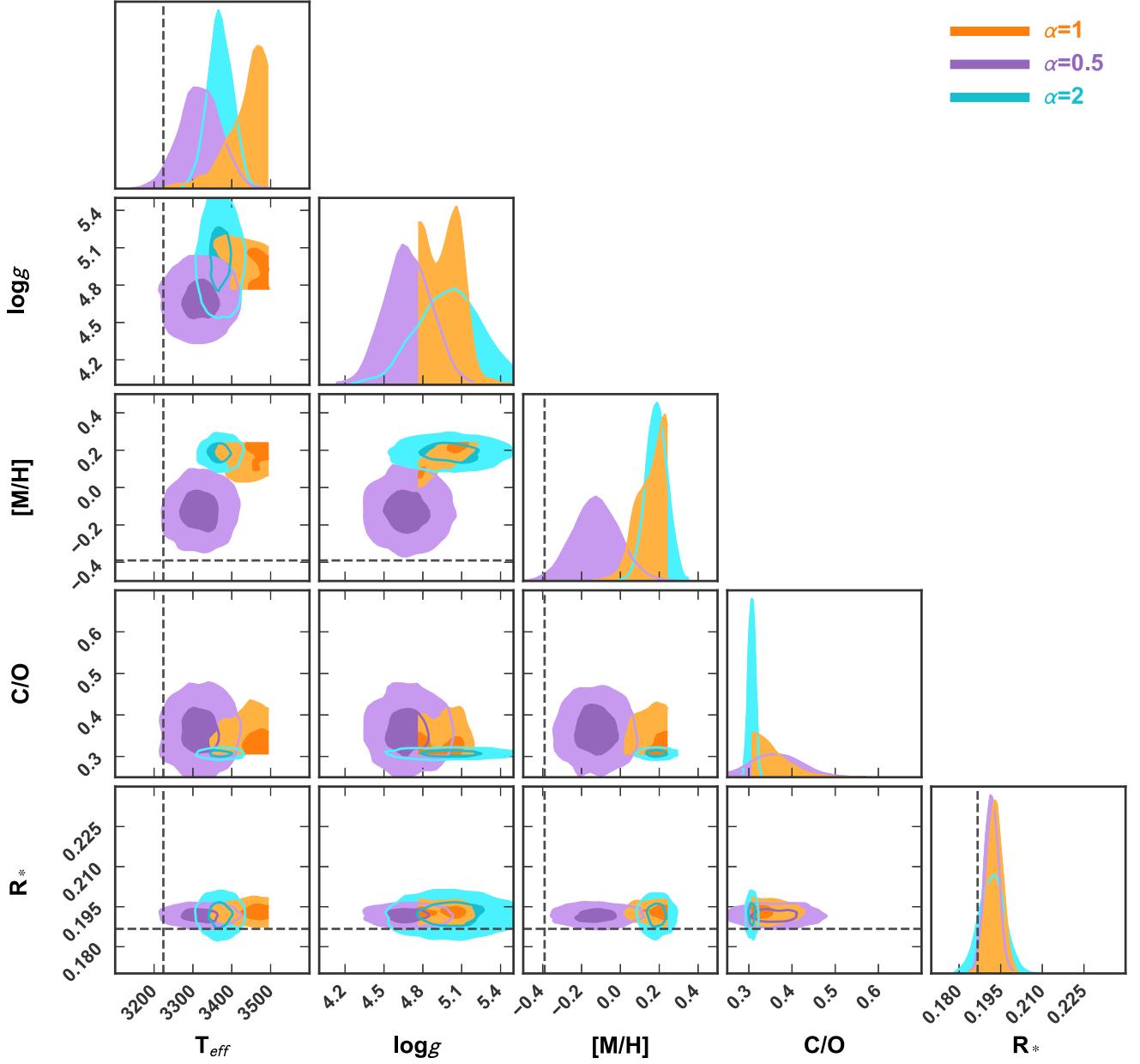
Figure 12. (Continued.)

5. We also explore the possibility of stellar photospheric heterogeneity and the inclusion of a parameterization of a simple linear combination for retrieving specifically the stellar spot temperature T_{spot} and the fractional coverage in area of the spots distributed on the photosphere f_{spot}, in addition to the stellar effective temperature T_{eff}. Assuming stellar photospheric heterogeneity as a hypothesis works well for certain targets and does not affect others. Therefore, a consensus must be reached to understand the degree to which stellar heterogeneity would affect M-dwarf abundances when focusing on the low-resolution ($R \sim 120$) regime. We stick purely to low resolution because we do not make assumptions for non-LTE within our model grid and therefore present an elegant method of leveraging low-resolution spectra and acquiring constraints on bulk chemical properties such as C/O and [M/H].

6. Lastly, we address the effect of general model assumptions that may also influence fundamental stellar properties that are retrieved, such as the mixing length parameter α. Here, we find that low-mass M dwarfs, particularly those that fall within the fully convective regime, are likely to be more sensitive to the assumed mixing length parameter than slightly hotter targets. However, assumed mixing length (or in other words, convection) has a potential for degeneracies manifesting from changing internal chemistry, ultimately affecting the bulk metallicity and effective temperature. Therefore, care must be taken especially when modeling low-mass stars that are around 0.3 $M_{\odot}$ or lower. In a future study (paper III of this series), we plan to perform free Bayesian retrievals (without self-consistent grid models and accompanying assumptions) on deriving appropriate atmospheric structures driven by low-resolution M-dwarf
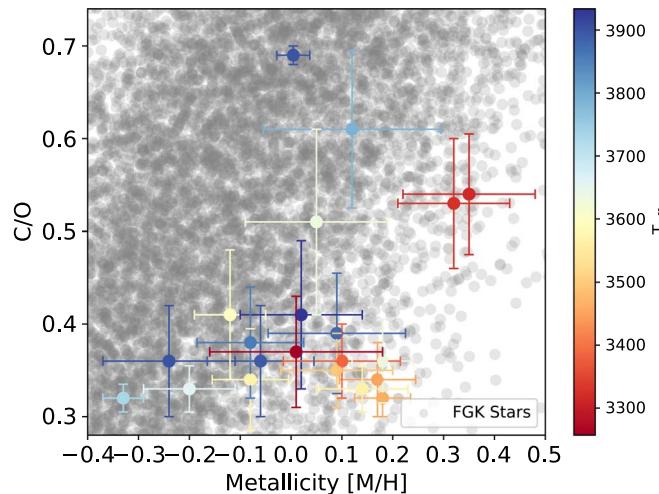
Figure 13. Comparison of our grid-model-retrieved metallicity values with C/O for both sample groups (WBS + IS). We also show the derived effective temperature for each target in the color bar. In gray, we show estimates of metallicity and C/O for the local FGK population from Hinkel et al. (2014). Our derived values are consistent with those for the local population.

data, and allow for arbitrary determinations of species abundances, which would truly stress-test model assumptions such as these.

In conclusion, we present our new stellar atmosphere model grid and emphasize that low-resolution spectra show immense potential in acquiring robust constraints on fundamental properties of vital exoplanet-hosting M dwarfs. Without invoking complicated non-LTE models or parameterizations of turbulent convection from 3D to 1D, we derive M-dwarf effective temperatures ($T_{\rm eff}$), surface gravity ($\log g$), bulk chemical properties such as metallicity ([M/H]), carbon-to-oxygen ratio (C/O), and the stellar radius (R_*), all within a reasonable scatter of values acquired from previous works (model-derived, empirical/observational, and machine learning methods). We also gain insight into model-sensitive parameters that extend our caution into understanding and modeling cooler M dwarfs that fall within the substellar boundary and implications for understanding brown dwarfs (late M and L) as we enter the JWST era. In future works, we will extend this analysis to a larger homogeneous sample of M-dwarf spectra to obtain broad population-level trends in composition and fundamental parameters.

Supplementary figures are available at: https://doi.org/10.5281/zenodo.7416042.

ORCID iDs

Aishwarya R. Iyer ● https://orcid.org/0000-0003-0971-1709
Michael R. Line ● https://orcid.org/0000-0001-6247-8323
Philip S. Muirhead ● https://orcid.org/0000-0002-0638-8822
Jonathan J. Fortney ● https://orcid.org/0000-0002-9843-4354
Ehsan Gharib-Nezhad ● https://orcid.org/0000-0002-4088-7262

References

Ackerman, A. S., & Marley, M. S. 2001, ApJ, 556, 872
Aldering, G., Adam, G., Antilogus, P., et al. 2002, Proc. SPIE, 4836, 61
Allard, F. 2014, in Proc. of the IAU 8, The BT-Settl Model Atmospheres for Stars, Brown Dwarfs and Planets (Cambridge: Cambridge Univ. Press), 271
Allard, F., Hauschildt, P. H., Alexander, D. R., & Starrfield, S. 1997, ARA&A, 35, 137
Allard, F., Hauschildt, P. H., Alexander, D. R., Tamanai, A., & Schweitzer, A. 2001, ApJ, 556, 357
Allard, F., Hauschildt, P. H., & Schweitzer, A. 2000, ApJ, 539, 366
Allard, F., Homeier, D., & Freytag, B. 2011, in ASP Conf. Ser. 448, 16th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. C. Johns-Krull, M. K. Browning, & A. A. West (San Francisco, CA: ASP), 91
Allard, N., Spiegelman, F., & Kielkopf, J. 2016, A&A, 589, A21
Allard, N., Spiegelman, F., Leininger, T., & Molliere, P. 2019, A&A, 628, A120
Amundsen, D. S., Tremblin, P., Manners, J., Baraffe, I., & Mayne, N. J. 2017, A&A, 598, A97
Apai, D., Rackham, B. V., Giampapa, M. S., et al. 2018, arXiv:1803.08708
Arcangeli, J., Désert, J.-M., Line, M. R., et al. 2018, ApJL, 855, L30
Ballard, S. 2019, AJ, 157, 113
Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, A&A, 577, A42
Barklem, P., Stempels, H., Prieto, C. A., et al. 2002, A&A, 385, 951
Bochanski, J. J., Hawley, S. L., Covey, K. R., et al. 2010, AJ, 139, 2679
Bochanski, J. J., Hawley, S. L., & West, A. A. 2011, AJ, 141, 98
Bocquet, S., & Carter, F. W. 2016, JOSS, 1, 6
Bohlin, R. C., Dickinson, M. E., & Calzetti, D. 2001, AJ, 122, 2118
Bohlin, R. C., & Gilliland, R. L. 2004, AJ, 128, 3053
Böhm-Vitense, E. 1958, ZA, 46, 108
Bonnefoy, M., Perraut, K., Lagrange, A. M., et al. 2018, A&A, 618, A63
Boyajian, T. S., von Braun, K., van Belle, G., et al. 2012, ApJ, 757, 112
Brewer, J. M., Fischer, D. A., Basu, S., Valenti, J. A., & Piskunov, N. 2015, ApJ, 805, 126
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Burgasser, A. J. 2014, arXiv:1406.4887
Burrows, A., Marley, M., Hubbard, W. B., et al. 1997, ApJ, 491, 856
Chabrier, G., & Baraffe, I. 1997, arXiv:astro-ph/9704118
Chambers, J. E. 2010, ApJ, 724, 92
Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, VizieR On-line Data Catalog: II/246

Czekala, I., Andrews, S. M., Mandel, K. S., Hogg, D. W., & Green, G. M. 2015a, Starfish: Robust spectroscopic inference tools, Astrophysics Source Code Library, ascl:1505.007

Czekala, I., Andrews, S. M., Mandel, K. S., Hogg, D. W., & Green, G. M. 2015b, ApJ, 812, 128

Czekala, I., Gully-Santiago, M., Gullikson, K., et al. 2018, iancze/Starfish: ca. release w/ Czekala et al. 2015. Version v0.3.0, doi:10.5281/zenodo.4288316

Delgado Mena, E., Adibekyan, V., Santos, N. C., et al. 2021, A&A, 655, A99

Dieterich, S. B., Simler, A., Henry, T. J., & Jao, W.-C. 2021, AJ, 161, 172

Dressing, C. D., & Charbonneau, D. 2013, ApJ, 767, 95

Dressing, C. D., & Charbonneau, D. 2015, ApJ, 807, 45

Ducati, J. R. 2002, CDS/ADC Collection of Electronic Catalogues, 2237, 0

Fischer, D. A., & Valenti, J. 2005, ApJ, 622, 1102

Fortney, J. J. 2012, ApJL, 747, L27

Fortney, J. J., Shabram, M., Showman, A. P., et al. 2010, ApJ, 709, 1396

Freedman, R. S., Lustig-Yaeger, J., Fortney, J. J., et al. 2014, ApJS, 214, 25

Gagné, J., Allers, K. N., Theissen, C. A., et al. 2018, ApJL, 854, L27

Gaia Collaboration 2018, VizieR On-line Data Catalog: I/345

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1

Gaidos, E. 2000, Icar, 145, 637

Gaidos, E., Mann, A. W., Kraus, A. L., & Ireland, M. 2016, MNRAS, 457, 2877

Gharib-Nezhad, E., Iyer, A. R., Line, M. R., et al. 2021a, ApJS, 254, 34

Gharib-Nezhad, E., Iyer, A. R., Line, M. R., et al. 2021b, EXOPLINES: Molecular Absorption Cross-Section Database for Brown Dwarf and Giant Exoplanet Atmospheres, v0.0, Zenodo, doi:10.5281/zenodo.4458189

Gharib-Nezhad, E., & Line, M. R. 2019, ApJ, 872, 27

Gillon, M., Triaud, A. H., Demory, B.-O., et al. 2017, Natur, 542, 456

Gizis, J. E., Marks, Z., & Hauschildt, P. H. 2016, MNRAS, 455, 3824

Gonzales, E. C., Burningham, B., Faherty, J. K., et al. 2020, ApJ, 905, 46

Gordon, S., & Mcbride, B. J. 1994, Computer program for calculation of complex chemical equilibrium compositions and applications. Part 1: Analysis, 19950013764, NASA, https://ntrs.nasa.gov/search.jsp?R=19950013764

Greene, T. P., Gully-Santiago, M. A., & Barsony, M. 2018, ApJ, 862, 85

Grimm, S. L., & Heng, K. 2015, ApJ, 808, 182

Gully-Santiago, M. A., Herczeg, G. J., Czekala, I., et al. 2017, ApJ, 836, 200

Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357

Hejazi, N., De Robertis, M. M., & Dawson, P. C. 2015, AJ, 149, 140

Hejazi, N., Lépine, S., & Nordlander, T. 2022, ApJ, 927, 122

Helling, C., Ackerman, A., Allard, F., et al. 2008, MNRAS, 391, 1854

Helling, C., Dehn, M., Woitke, P., & Hauschildt, P. H. 2008, ApJL, 675, L105

Henry, T. J., Kirkpatrick, J. D., & Simons, D. A. 1994, AJ, 108, 1437

Hinkel, N. R., Timmes, F. X., Young, P. A., Pagano, M. D., & Turnbull, M. C. 2014, AJ, 148, 54

Hubeny, I. 2017, MNRAS, 469, 841

Husser, T. O., Wende-von Berg, S., Dreizler, S., et al. 2013, A&A, 553, A6

Irwin, J., Charbonneau, D., Berta, Z. K., et al. 2009, ApJ, 701, 1436

Iyer, A. R., & Line, M. R. 2020, ApJ, 889, 78

Jofré, P., Heiter, U., Soubiran, C., et al. 2014, A&A, 564, A133

John, T. 1988, A&A, 193, 189

Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, ApJ, 765, 131

Kreidberg, L., Line, M. R., Parmentier, V., et al. 2018, AJ, 156, 17

Kurucz, R. L. 1993, PhyST, 1993, 110

Kurucz, R. L. 2011, CaJPh, 89, 417

Lantz, B., Aldering, G., Antilogus, P., et al. 2004, Proc. SPIE, 5249, 146

Lépine, S., & Gaidos, E. 2011, AJ, 142, 138

Lépine, S., Hilton, E. J., Mann, A. W., et al. 2013, AJ, 145, 102

Lépine, S., Rich, R. M., & Shara, M. M. 2003a, AJ, 125, 1598

Lépine, S., Shara, M. M., & Rich, R. M. 2003b, AJ, 126, 921

Line, M. R., Knutson, H., Wolf, A. S., & Yung, Y. L. 2014, ApJ, 783, 70

Line, M. R., Marley, M. S., Liu, M. C., et al. 2017, ApJ, 848, 83

Line, M. R., Teske, J., Burningham, B., Fortney, J. J., & Marley, M. S. 2015, ApJ, 807, 183

Line, M. R., Wolf, A. S., Zhang, X., et al. 2013, ApJ, 775, 137

Lupu, R., Freedman, R., Gharib-Nezhad, E., & Molliere, P. 2022, High Resolution Opacities for H2/He Atmospheres, v1, Zenodo, doi:10.5281/zenodo.6600976

Madhusudhan, N. 2012, ApJ, 758, 36

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94

Mann, A. W., Brewer, J. M., Gaidos, E., Lépine, S., & Hilton, E. J. 2013, AJ, 145, 52

Mann, A. W., Feiden, G. A., Gaidos, E., Boyajian, T., & Von Braun, K. 2015, ApJ, 804, 64

Mansfield, M., Line, M. R., Bean, J. L., et al. 2021, NatAs, 5, 1224

Marfil, E., Tabernero, H., Montes, D., et al. 2021, A&A, 656, A162

Marley, M. S., & Robinson, T. D. 2015, ARA&A, 53, 279

Marley, M. S., Saumon, D., Guillot, T., et al. 1996, Sci, 272, 1919

Marley, M. S., Saumon, D., Visscher, C., et al. 2021, ApJ, 920, 85

McKay, C. P., Pollack, J. B., & Courtin, R. 1989, Icar, 80, 23

Melendez, J., Asplund, M., Gustafsson, B., & Yong, D. 2009, ApJL, 704, L66

Mordasini, C., van Boekel, R., Mollière, P., Henning, T., & Benneke, B. 2016, ApJ, 832, 41

Moses, J., Madhusudhan, N., Visscher, C., & Freedman, R. 2012, ApJ, 763, 25

Muirhead, P. S., Dressing, C. D., Mann, A. W., et al. 2018, AJ, 155, 180

Mulders, G. D., Pascucci, I., & Apai, D. 2015, ApJ, 814, 130

Nakajima, T., & Sorahana, S. 2016, ApJ, 830, 159

Nutzman, P., & Charbonneau, D. 2008, PASP, 120, 317

Offner, S. S. R., Kratter, K. M., Matzner, C. D., Krumholz, M. R., & Klein, R. I. 2010, ApJ, 725, 1485

Osterbrock, D. E. 1953, ApJ, 118, 529

Passegger, V., Bello-García, A., Ordieres-Meré, J., et al. 2022, A&A, 658, A194

Patience, J., King, R., De Rosa, R., et al. 2012, A&A, 540, A85

Petigura, E. A., Marcy, G. W., & Howard, A. W. 2013, ApJ, 770, 69

Petty, G. W. 2006, A First Course in Atmospheric Radiation (Madison, WI: Sundog Publishing)

Phillips, M. W., Tremblin, P., Baraffe, I., et al. 2020, A&A, 637, A38

Pinhas, A., Rackham, B. V., Madhusudhan, N., & Apai, D. 2018, MNRAS, 480, 5314

Piskorz, D., Buzard, C., Line, M. R., et al. 2018, AJ, 156, 133

Piskunov, N., Kupka, F., Ryabchikova, T., Weiss, W., & Jeffery, C. 1995, A&AS, 112, 525

Polyansky, O. L., Kyuberis, A. A., Zobov, N. F., et al. 2018, MNRAS, 480, 2597

Rackham, B., Espinoza, N., Apai, D., et al. 2017, ApJ, 834, 151

Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, ApJ, 853, 122

Rackham, B. V., Espinoza, N., Berdyugina, S. V., et al. 2022, arXiv:2201.09905

Ramírez, I., Melendez, J., & Asplund, M. 2014, A&A, 561, A7

Ramírez, I., Meléndez, J., Cornejo, D., Roederer, I. U., & Fish, J. R. 2011, ApJ, 740, 76

Rayner, J. T., Toomey, D. W., Onaka, P. M., et al. 2003, PASP, 115, 362

Rojas-Ayala, B., Covey, K. R., Muirhead, P. S., & Lloyd, J. P. 2012, ApJ, 748, 93

Schneider, A. C., Meisner, A. M., Gagné, J., et al. 2021, ApJ, 921, 140

Segransan, D., Kervella, P., Forveille, T., & Queloz, D. 2003, A&A, 397, L5

Sharp, C. M., & Burrows, A. 2007, ApJS, 168, 140

Showman, A. P., Fortney, J. J., Lian, Y., et al. 2009, ApJ, 699, 564

Song, N., Alexeeva, S., Sitnova, T., et al. 2020, A&A, 635, A176

Souto, D., Cunha, K., Smith, V. V., et al. 2020, ApJ, 890, 133

Souto, D., Cunha, K., Smith, V. V., et al. 2022, ApJ, 927, 123

ten Brummelaar, T. A., McAlister, H. A., Ridgway, S. T., et al. 2005, ApJ, 628, 453

Tennyson, J., Yurchenko, S. N., Al-Refaie, A. F., et al. 2016, JMoSp, 327, 73

Tennyson, J., Yurchenko, S. N., Al-Refaie, A. F., et al. 2020, JQSRT, 255, 107228

Teske, J. K., Cunha, K., Smith, V. V., Schuler, S. C., & Griffith, C. A. 2014, ApJ, 788, 39

Teske, J. K., Khanal, S., & Ramírez, I. 2016, ApJ, 819, 19

Ting, Y.-S., Conroy, C., Rix, H.-W., & Asplund, M. 2018, ApJ, 860, 159

Toon, O. B., McKay, C. P., Ackerman, T. P., & Santhanam, K. 1989, JGRD, 94, 16287

Tsuji, T. 2002, arXiv:astro-ph/0209495

Tsuji, T., Ohnaka, K., & Aoki, W. 1996, A&A, 305, L1

Veyette, M. J., Muirhead, P. S., Mann, A. W., et al. 2016, ApJ, 828, 95

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261

Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9

West, A. A., Morgan, D. P., Bochanski, J. J., et al. 2011, AJ, 141, 97

Witte, S., Helling, C., Barman, T., Heidrich, N., & Hauschildt, P. H. 2011, A&A, 529, A44

Yurchenko, S. N., Sinden, F., Lodi, L., et al. 2017, MNRAS, 473, 5324

Zalesky, J. A., Line, M. R., Schneider, A. C., & Patience, J. 2019, ApJ, 877, 24

Zalesky, J. A., Saboi, K., Line, M. R., et al. 2022, ApJ, 936, 44

Zhang, Z., Liu, M. C., Marley, M. S., Line, M. R., & Best, W. M. J. 2021a, ApJ, 916, 53

Zhang, Z., Liu, M. C., Marley, M. S., Line, M. R., & Best, W. M. J. 2021b, ApJ, 921, 95

Zhang, Z., Zhou, Y., Rackham, B. V., & Apai, D. 2018, AJ, 156, 178