The Astrophysical Journal, 988:31 (24pp), 2025 July 20 https://doi.org/10.3847/1538-4357/addd06

Comparing Grid Model Fitting Methodologies for Low-temperature Atmospheres: Markov Chain Monte Carlo versus Random Forest Retrieval

Anna Lueber[1,2] and Adam J. Burgasser[3]
[1] Ludwig Maximilian University, Faculty of Physics, University Observatory, Scheinerstrasse 1, Munich D-81679, Germany; anna.lueber@physik.uni-muenchen.de
[2] University of Bern, Center for Space and Habitability, Gesellschaftsstrasse 6, CH-3012, Bern, Switzerland
[3] Department of Astronomy & Astrophysics, UC San Diego, La Jolla, CA, USA
Received 2024 December 24; revised 2025 May 20; accepted 2025 May 23; published 2025 July 11

Abstract

The atmospheres of low-temperature stars, brown dwarfs, and exoplanets are challenging to model due to strong molecular features and complex gas and condensate chemistry. Self-consistent atmosphere models are commonly used for spectral fitting, but computational limits restrict the production of finely sampled multidimensional parameter grids, necessitating interpolation methods to infer precise parameters and uncertainties. Here, we compare two grid model fitting approaches: a Markov Chain Monte Carlo (MCMC) algorithm interpolating across spectral fluxes, and a random forest retrieval (RFR) algorithm trained on a grid model set. We test these with three low-temperature model grids—Sonora Diamondback, Sonora Elf Owl, and Spectral ANalog of Dwarfs (SAND)—and a sample of 11 L and T dwarf companions to FGKM stars with known distances, compositions, and ages. Diamondback models are optimal for early- and mid-type L dwarfs, Elf Owl for mid and late T dwarfs, and SAND for young L dwarfs and L/T transition objects. The MCMC approach yields higher fit quality and more precise parameters, though best-fit parameters are generally consistent between approaches. RFR analysis is orders of magnitude faster after training. Both approaches yield mixed results when comparing fit parameters to expected values based on the primary (metallicity and surface gravity) or evolutionary models (temperature and radius). We propose modeling low-temperature spectra efficiently by first fitting multiple model sets using RFR, followed by a more accurate MCMC assessment, to accelerate improved grid development.

Unified Astronomy Thesaurus concepts: Brown dwarfs (185); L dwarfs (894); T dwarfs (1679); Stellar atmospheres (1584); Random Forests (1935)

1. Introduction

The atmospheres of the coldest stars, brown dwarfs, and exoplanets ($T_{\rm eff} \lesssim 2500$ K) give rise to complex spectra, shaped by strong molecular bands and continuum features shaped by complex gas and condensate chemistry modulated by a wide range of elemental abundances and nonequilibrium processes. These sources pose unique atmospheric characterization challenges that blend planetary and stellar approaches (F. Allard et al. 1997; A. Burrows et al. 2001; M. S. Marley & T. D. Robinson 2015), and the increased availability of high-fidelity spectra of these sources, particularly of low-temperature brown dwarfs, has driven advances in treatments of condensation and cloud formation (A. S. Ackerman & M. S. Marley 2001; F. Allard et al. 2001; C. Helling et al. 2001, 2008), vertical mixing and nonequilibrium chemistry (B. J. Fegley & R. G. Prinn 1985; C. A. Griffith & R. V. Yelle 1999; D. Saumon et al. 2006), nonsolar abundances (N. Madhusudhan 2012; R. Gerasimov et al. 2024), and modifications to convective heat transport (P. Tremblin et al. 2015, 2016), among others. However, there remain distinct challenges to accurately reproducing the spectra of cool brown dwarfs, including the rapid evolution of clouds at the L dwarf/T dwarf transition (A. Burrows et al. 2006; B. Charnay et al. 2018), molecular abundances that exceed or fall short of model predictions (e.g., CO_2 and PH_3; S. Sorahana & I. Yamamura 2014; S. A. Beiler et al. 2024), the complex spectra of cold Y dwarfs (S. K. Leggett et al. 2021), and the untested chemistry of metal-poor brown dwarfs (A. J. Burgasser et al. 2009; A. C. Schneider et al. 2020; N. Lodieu et al. 2022).

Modeling low-temperature spectra and inferring atmospheric parameters is commonly achieved through comparison to precomputed, self-consistent grids of atmosphere models, either through the identification of a single "best-fit" model (e.g., M. C. Cushing et al. 2008; D. C. Stephens et al. 2009; J. Patience et al. 2012) or through robust inference of parameter distributions by interpolation across model parameters by principal component analysis (Z. Zhang et al. 2021a) or Markov Chain Monte Carlo (MCMC) methods (C. Aganze et al. 2016; A. J. Burgasser et al. 2024). Depending on the source examined and models deployed, these approaches can lead to varying degrees of accuracy in the reproduction of observed spectra, with corresponding systematic uncertainties and biases in inferred parameters (Z. Zhang et al. 2021b; A. J. Burgasser et al. 2025). An alternative approach is retrieval modeling, in which radiative transfer models are computed continuously with varying pressure–temperature profiles and molecular abundances (P. G. J. Irwin et al. 2008; N. Madhusudhan & S. Seager 2009; M. R. Line et al. 2014; B. Burningham et al. 2017; D. Kitzmann et al. 2020; J. W. Xuan et al. 2024). This resource-intensive approach results in accurate fits to observed spectra, but the lack of self-consistent energy balance or chemistry can make it difficult to interpret the atmospheric processes underlying the fit. A third approach that is growing in popularity is the application of various machine learning (ML) models trained on atmosphere grids, which we refer to as ML retrieval (I. P. Waldmann 2016; P. Márquez-Neila et al. 2018; A. D. Cobb et al. 2019; C. Fisher et al. 2020; F. Ardévol Martínez et al. 2022; M. Vasist et al. 2023; T. D. Gebhard et al. 2024). This approach is highly efficient, making it well suited to large and high-resolution spectral surveys that may be prohibitive with direct fit or retrieval methods. However, ML retrievals can also be subject to potential biases inherent in the underlying grid model or the ML algorithm deployed. The challenges associated with deriving inferences from sparsely sampled, high-dimensional parameter spaces have been extensively studied in computer experiment design and analysis. Methodologies like Latin hypercube sampling and emulator-based interpolation, including Gaussian process regression, have been developed to navigate these spaces while minimizing model evaluations (T. J. Santner et al. 2003; K.-T. Fang et al. 2005; J. P. Kleijnen 2015). Applications thereof can be found in, e.g., C. Fisher & K. Heng (2022).

While numerous studies have compared direct fit and retrieval approaches to the fitting of brown dwarf and exoplanet spectra, few comparisons have been made between direct fit and ML retrieval methods with a focus on precision, accuracy, and efficiency in the spectral fits and inferred parameters. In this study, we compare an MCMC-based direct fitting approach to an ML retrieval approach based on the supervised ML method known as random forest (RF; L. Breiman 2001). In Section 2 we describe our empirical spectral sample of 11 benchmark cool stars and brown dwarfs with independently constrained ages and compositions, the three atmospheric model grids used in our fits, and the two model fit approaches explored in this study. In Section 3 we review the results of our fits, evaluating both the quality of reproduction of the spectra and the robustness of the inferred model parameters (uncertainties and physical reliability). We also examine in detail three case studies that span the temperature range of our sample. In Section 4 we discuss our outcomes through the lens of model fitting precision, accuracy, and efficiency. Our main results are summarized in Section 5.

2. Methods

2.1. Empirical Spectral Sample

Our analysis focused on a sample of 11 L and T dwarfs spanning spectral types L0 to T8 that are all widely separated singular companions to main-sequence stars. The sample is summarized in Table 1, which includes prior estimates of effective temperatures ($T_{\rm eff}$), surface gravities ($\log g$), and radii from A. Sanghi et al. (2023) based on bolometric luminosities, system ages, and evolutionary models. We also list the physical properties of the primaries, notably metallicity ([Fe/H]) and age, as inferred by various methods in the cited references. All of the L and T dwarfs in the sample have previously published low-resolution ($\lambda/\Delta\lambda \approx 150$), near-infrared (0.8–2.5 μm) spectra obtained with the SpeX instrument on the 3 m NASA Infrared Telescope Facility (J. T. Rayner et al. 2003), with high values of signal-to-noise ranging within S/N = 40–270 in the 1.25–1.30 μm region, with an average S/N $\approx$ 100. We retrieved these spectra from the SpeX Prism Library Analysis Toolkit (SPLAT; A. J. Burgasser & SPLAT Development Team 2017), and flux-calibrated them using each source's absolute M_J magnitude based on photometry from the Two Micron All Sky Survey (2MASS; M. F. Skrutskie et al. 2006) and the parallax of the primary from Gaia Data Release 3 (Gaia Collaboration et al. 2023). One source, GJ 1048B, is obscured by its primary in the 2MASS J band, so we used its 2MASS K_s-band photometry instead and computed a synthetic J magnitude from the spectrum. Figure 1 displays the sequence of these spectra in normalized F_λ flux densities.

2.2. Spectral Models

We evaluated three state-of-the-art grids of atmosphere models that encompass differing treatments of condensation and cloud formation, nonequilibrium chemistry, and abundance variations, as summarized in Table 2. The Sonora Diamondback model grid (DBack; C. V. Morley et al. 2024) is based on the radiative–convective equilibrium framework described in C. P. McKay et al. (1989) and D. Saumon & M. S. Marley (2008), and assumes radiative–convective and chemical equilibrium throughout the atmosphere. Sonora Diamondback incorporates condensation and cloud formation in both chemistry and opacity using the approach of A. S. Ackerman & M. S. Marley (2001), parameterizing the efficiency of condensation through a sedimentation efficiency parameter ($f_{\rm sed}$). These models include a small range of solar-scaled metallicity variations ($-0.5 \leqslant$ [M/H] $\leqslant +0.5$). The Sonora Elf Owl model grid (EOwl; S. Mukherjee et al. 2024) is based on the PICASO model framework (S. Mukherjee et al. 2023) and is a cloud-free model that includes the effects of nonequilibrium chemistry ("quenching") through vertical mixing as parameterized by the vertical diffusion coefficient (κ_{zz}). This grid also allows for solar-scaled and nonsolar abundances, the latter by varying the C/O ratio relative to an assumed solar $({\rm C/O})_\odot$ = 0.458 (K. Lodders et al. 2009). The Spectral ANalog of Dwarfs (SAND) model grid (R. Gerasimov et al. 2022, 2024; E. Alvarado et al. 2024) is based on the PHOENIX code (P. H. Hauschildt et al. 1997; F. Allard et al. 2011, 2012; R. Gerasimov et al. 2020), and is a chemical equilibrium model that incorporates condensation and cloud formation through a prescribed gravitational settling algorithm ("hybrid"; F. Allard et al. 2012). SAND includes a broad range of metallicities and nonsolar abundances, with the latter quantified by alpha enrichment ([α/Fe]). For ease of computation, we downselected these model grids to streamline our fits. We reduced the number of models by constraining them to ranges 600 K $\leqslant T_{\rm eff} \leqslant$ 2400 K (extended to 3000 K for SAND), 4.5 $\leqslant \log g \leqslant$ 6.0 (units of cm s^{-2}), $-0.5 \leqslant$ [M/H] $\leqslant +0.5$, and 0.5 $\leqslant$ C/O $\leqslant$ 1.5. These constraints are on top of the parameter limits in the original model grids. These constraints result in 864, 5175, and 952 models for Diamondback, Elf Owl, and SAND, respectively. All model spectra were Gaussian-smoothed to the resolution of the SpeX data and mapped onto a common wavelength scale by linear interpolation, and we conducted our fits in surface F_λ flux densities.

2.3. Modeling Approaches

2.3.1. MCMC

Our baseline modeling approach was a direct comparison of data and grid models to identify the best-fitting parameters and a robust estimate of their uncertainties. We deployed a Metropolis–Hastings MCMC algorithm (N. Metropolis et al. 1953; W. K. Hastings 1970), starting from an initial estimate of the best model parameters by first comparing the spectrum to each of the individual models in the grid using a χ^2 fitting statistic,

$$\chi^2 = \sum_{i=1}^{N} \frac{(O[\lambda_i] - \beta M[\lambda_i])^2}{\sigma[\lambda_i]^2}. \quad (1)$$

Table 1
Benchmark Spectral Sample

Secondary							Primary				
Name	SpT	M_J (mag)	M_K (mag)	$T_{\rm eff}$[a] (K)	$\log g$[a] (cm s^{-2})	R[a] ($R_{\rm Jup}$)	Name	SpT	[Fe/H] (dex)	Age (Gyr)	References
LP 465-70B	L0	11.935 ± 0.003	10.771 ± 0.003	2148 ± 11	5.34 ± 0.01	0.99 ± 0.01	LP 465-70	M4	−0.120 ± 0.006	4.5–10	(1–6)
HD 89744B	L0	11.92 ± 0.04	10.65 ± 0.04	2302 ± 19	5.35 ± 0.01	0.98 ± 0.01	HD 89744	F7	−0.15 ± 0.01	1.5–3	(4, 7–9)
GJ 1048B	L1	12.02 ± 0.09[b]	10.53 ± 0.08	2332 ± 24	5.33 ± 0.02	1.00 ± 0.01	GJ 1048	K3.5	0.14 ± 0.01	0.6–2	(4, 10–13)
Gl 618.1B	L2.5	12.84 ± 0.05	11.21 ± 0.04	1907 ± 27	5.34 ± 0.05	0.97 ± 0.02	MCC 760	K7/M0	0.37 ± 0.09	0.5–12	(4, 7, 9, 11, 13)
G 62-33B	L2.5	12.776 ± 0.005	11.232 ± 0.005	1939 ± 16	5.34 ± 0.01	1.08 ± 0.16	HD 116012	K0.5	0.061 ± 0.005	3.3–5.1	(2, 4, 6, 14, 15)
G 196-3B	L3β	13.03 ± 0.05	11.04 ± 0.03	1725 ± 52	4.77 ± 0.15	1.18 ± 0.07	G 196-3	M3	⋯	0.05–0.10	(4, 16–19)
Gliese 584C	L8	14.82 ± 0.02	13.09 ± 0.02	1237 ± 24	5.11 ± 0.09	0.95 ± 0.03	Gliese 584AB	G2+G2	−0.01 ± 0.12	1–2.5	(4, 9, 20–23)
HN Peg B	T2.5	14.57 ± 0.03	13.83 ± 0.03	1056 ± 38	4.53 ± 0.22	1.14 ± 0.08	HN Peg	G0	−0.43 ± 0.01	0.24 ± 0.03	(4, 24–26)
G 204-39B	T6.5	15.193 ± 0.011	15.39 ± 0.03	898 ± 26	4.87 ± 0.16	0.99 ± 0.05	G 204-39	M2.5	−0.04 ± 0.06	0.5–3	(4, 16, 27, 28)
HD 3651B	T7	16.08 ± 0.03	16.64 ± 0.05	785 ± 12	5.22 ± 0.06	0.85 ± 0.02	HD 3651	K0.5	−0.082 ± 0.005	6.9 ± 2.8	(4, 24, 29, 30)
Gliese 570D	T8	16.47 ± 0.05	16.67 ± 0.05	771 ± 18	4.99 ± 0.12	0.93 ± 0.04	Gliese 570A	K4	−0.04 ± 0.03	3.3 ± 1.9	(4, 27, 31–33)

Notes.
[a] Estimated $T_{\rm eff}$, $\log g$, and radii for L and T dwarf secondaries adopted from A. Sanghi et al. (2023) and based on measured bolometric luminosities, system ages, and the evolutionary models of D. Saumon & M. S. Marley (2008) and I. Baraffe et al. (2015).
[b] Estimated from spectrophotometric $J - K_s$ color synthesized from the infrared spectrum.

References. (1) S. L. Hawley et al. (2002); (2) UKIRT Infrared Deep Sky Survey (A. Lawrence et al. 2007); (3) J. I. Gomes et al. (2013); (4) Gaia DR3 (Gaia Collaboration et al. 2023); (5) D. Sprague et al. (2022); (6) J. K. Faherty et al. (2010); (7) J. E. Gizis et al. (2001a); (8) T. J. Dupuy & M. C. Liu (2012); (9) R. Zhang et al. (2024); (10) J. E. Gizis et al. (2001b); (11) J. K. Faherty et al. (2012); (12) R. O. Gray et al. (2006); (13) C. B. Stephenson (1986); (14) D. C. Bardalez Gagliuffi et al. (2014); (15) N. Houk & C. Swift (1999); (16) K. L. Cruz et al. (2009); (17) W. M. J. Best et al. (2021); (18) E. Shkolnik et al. (2009); (19) O. V. Zakhozhay et al. (2023); (20) J. D. Kirkpatrick et al. (2000); (21) A. C. Schneider et al. (2023b); (22) O. Struve & K. L. Franklin (1955); (23) F. van Leeuwen (2007); (24) K. L. Luhman et al. (2007); (25) R. O. Gray et al. (2001); (26) S. A. Barnes (2007); (27) A. J. Burgasser et al. (2006b); (28) G. R. Knapp et al. (2004); (29) E. Marfil et al. (2021); (30) P. C. Keenan & R. C. McNeil (1989); (31) A. Bonfanti et al. (2016); (32) S. K. Leggett et al. (2010); (33) Z. Zhang et al. (2021b).

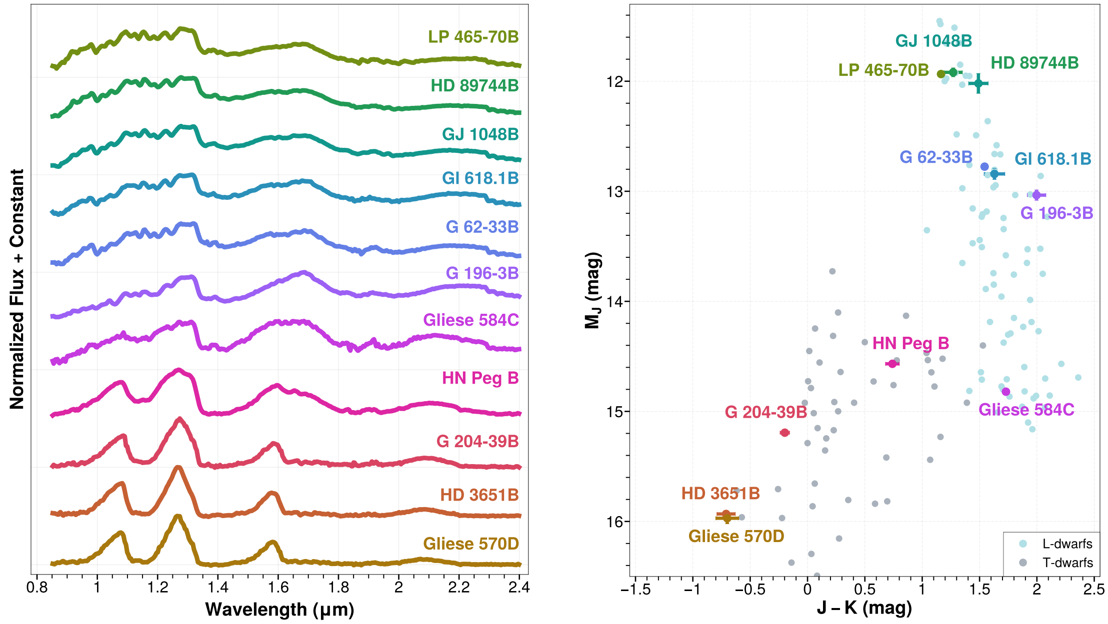

Figure 1. Left: near-infrared spectra of our curated sample of 11 L and T dwarf companions to main-sequence stars (see Table 1). All spectra are normalized at their maximum flux value and offset by a constant. The associated data uncertainties are included in the plot but are too small to be visible. Right: color–magnitude diagram for our curated sample of L and T dwarfs, compared to a broader sample of L dwarfs (light blue circles) and T dwarfs (gray circles) with the measured parallaxes compiled by T. J. Dupuy & M. C. Liu (2012).

Table 2
Parameter Ranges for Spectral Model Grids

Model	$T_{\rm eff}$ (K)	$\log g$ (cm s^{-2})	[M/H] (dex)	(C/O) ($\odot$)[a]	[α/Fe] (dex)	$\log \kappa_{zz}$ (cm^2 s^{-1})	$f_{\rm sed}$	# Models	References
Diamondback	900–2400	4.5–5.5	−0.5 to +0.5	⋯	⋯	⋯	1–8	864	(1)
Elf Owl	600–2400	4.5–5.5	−0.5 to +0.5	0.5–1.5	⋯	2–9	⋯	5175	(2)
SAND	700–3000	4.5–6.0	−0.35 to +0.3	⋯	−0.05 to +0.4	⋯	⋯	952	(3)

Note.
[a] C/O ratio relative to solar abundance, with $(\mathrm{C/O})_\odot = 0.458$ (K. Lodders et al. 2009).
References. (1) C. V. Morley et al. (2024); (2) S. Mukherjee et al. (2024); (3) E. Alvarado et al. (2024).

Here, N is the number of spectral data points, $O[\lambda_i]$ is the observed spectrum scaled to absolute flux density units, $M[\lambda_i]$ is a model spectrum in surface flux density units, $\sigma[\lambda_i]$ is the observed uncertainty spectrum, and β is the optimal scale factor computed as

$$\beta = \frac{\sum_{i=1}^{N} \frac{O[\lambda_i]M[\lambda_i]}{\sigma[\lambda_i]^2}}{\sum_{i=1}^{N} \frac{M[\lambda_i]^2}{\sigma[\lambda_i]^2}} \tag{2}$$

(see M. C. Cushing et al. 2008). Note that the spectral flux density scalings allow us to estimate the source radius from the scale factor as

$$R_\beta = 10\sqrt{\beta} \quad \mathrm{pc} = 2.256 \times 10^{-7}\sqrt{\beta} \quad R_\odot. \tag{3}$$

The parameters of the model with the lowest χ^2 residual were used as the starting point for a single chain of 5000 steps. In each iteration, we drew new parameters using a Gaussian offset with standard deviations of $\sigma_{T_{\rm eff}} = 25$ K, $\sigma_{\log g} = 0.2$ dex, $\sigma_{[\mathrm{M/H}]} = 0.2$ dex, $\sigma_{[\alpha/\mathrm{H}]} = 0.2$ dex, $\sigma_{\mathrm{C/O}} = 0.05$ dex, $\sigma_{\log \kappa_{zz}} = 0.2$ dex, and $\sigma_{f_{\rm sed}} = 0.25$. Models between the grid points were linearly interpolated in logarithmic flux units on a logarithmic parameter grid (i.e., $T_{\rm eff} \Rightarrow \log T_{\rm eff}$). Sequential fits ($i \Rightarrow i + 1$) were compared using the criterion

$$\frac{\chi^2(i+1) - \chi^2(i)}{\mathrm{MIN}[\chi^2]} < \mathcal{U}(0, 0.5) \tag{4}$$

where MIN is the minimum of all χ^2 values in the chain and $\mathcal{U}(0, 0.5)$ is a number drawn from a uniform distribution between 0 and 0.5. If the new fit satisfied this criterion, these parameters were added to the chain; otherwise the previous parameters were added. We also enforced a limit on χ^2 values exceeding $2 \times \mathrm{MIN}[\chi^2]$, at which point the chain reverted to the minimum χ^2 parameter set. We verified that all fits converged using the convergence diagnostic defined by J. Geweke (1992), requiring that the means of the variable parameters in the first 10% and last 50% of each chain differ by less than 3 times their combined variance. We also visually confirmed convergence, and examined each fit to ensure it provided an accurate reproduction of the data.

2.3.2. RF Retrieval

Our second modeling approach uses the supervised ML RF algorithm (T. K. Ho 1998; L. Breiman 2001) integrated in the HELA framework (P. Márquez-Neila et al. 2018), an approach we refer to here as random forest retrieval (RFR). The RF algorithm constructs an ensemble of decision trees, each trained on random subsets of prelabeled training data, in this case a preexisting grid of atmospheric model spectra labeled by the model parameters. This ensemble approach allows for the modeling of complex, nonlinear relationships between the input spectral data and the target atmospheric parameters with improved accuracy and robustness. This methodology represents a type of approximate Bayesian computation (S. A. Sisson et al. 2018). Two of the key advantages of the RFR method are its computational efficiency and its ability to determine feature importance, a metric that ranks the significance of each spectral data point (referred to as "features") in constraining a model parameter. The HELA RFR framework has been applied in multiple retrieval studies, including the analysis of high-resolution spectra of the ultrahot Jupiter KELT-9b (C. Fisher et al. 2020), the analysis of medium-resolution spectra of benchmark brown dwarfs (M. Oreshenko et al. 2020; A. Lueber et al. 2023), the evaluation of information content in James Webb Space Telescope spectra (A. Guzmán-Mesa et al. 2020; A. Lueber et al. 2024a, 2024b), and investigations into model grid sampling strategies (C. Fisher & K. Heng 2022).

The RF algorithm reduces the risk of overfitting by aggregating predictions across multiple decision trees, enhancing the model's generalization capabilities and making it particularly effective for dealing with high-dimensional data sets. To address potential distortions in the predictions caused by sparse sampling, a bias correction was applied (W. G. Cochran 1977; M. Galassi et al. 2011). Within our ML retrieval framework, frequency weights ($w_i \geqslant 0, \forall i$) are calculated as the number of occurrences of training set samples predicted by each decision tree. The weighted mean (μ_w) of a specific parameter is given by

$$\mu_w = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}, \tag{5}$$

where $\sum_{i=1}^{N} w_i = N^*$ represents the effective sample size. The weighted variance across the data set is calculated as the sum of the squared deviations of each data point from the weighted mean, normalized by the effective sample size:

$$\sigma_w^2 = \frac{\sum_{i=1}^{N} w_i (x_i - \mu_w)^2}{\sum_{i=1}^{N} w_i}. \tag{6}$$

This method is particularly suitable for data sets with unequal variances or sampling biases, as it ensures that more reliable data points substantially influence the variability estimate, thereby reducing the impact of less representative samples.

In addition to varying the model parameters, we included the flux scaling factor f, which scales the model surface flux densities scaled at that expected brightness at Earth to the observed flux density of the source:

$$O[\lambda] = f\left(\frac{R}{d}\right)^2 M[\lambda] \tag{7}$$

where d and R represent the distance and radius of the source. As all observed spectra are scaled to their absolute flux densities, we initially assumed $(R/d)^2 = (R_{\rm Jup}/10 \quad {\rm pc})^2$ with a uniform distribution for f ranging from 0.5 to 2.0. This scale factor is reported in terms of the inferred radii in solar units, $R = 0.0103\sqrt{f} \quad R_\odot$. For training, we randomly assigned an individual f value to each spectrum and scaled the fluxes of the corresponding atmospheric model accordingly.

We trained three separate RFR models, one for each of our model grids. To train the RFR model, the atmosphere model grid was partitioned into training and testing sets with an approximate ratio of 80:20 following prior research (P. Márquez-Neila et al. 2018; C. Fisher et al. 2020; A. Lueber et al. 2023). The efficacy of this partitioning can be assessed through real versus predicted (RvP) evaluations, which evaluate the coefficient of determination as a quantitative measure of efficacy,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{S_{\rm res}}{S_{\rm tot}}, \tag{8}$$

with y_i and f_i denoting the real and predicted values for a given parameter, respectively, while $\bar{y}_i$ represents the mean of the given real values. Therefore, $S_{\rm res}$ is the sum of the squares of the residual errors, and $S_{\rm tot}$ is the total sum of the errors. R^2 can vary between negative and positive values, depending on the level of correlation and disparity between inferred and actual values, with perfect alignment yielding $R^2 = 1$ and no correlation yielding $R^2 = 0$. The presence of noise in data usually results in R^2 values below unity, even when the noise-free RvP relationship yields an R^2 of 1. Note that $R^2 < 0$ indicates parameter relationships that are worse than random chance.

The hyperparameters of our RF models were adopted from P. Márquez-Neila et al. (2018). The models were constructed to contain 3000 regression trees with no maximum tree depth constraint. Each tree grows by partitioning the model space until the decrease in variance with further splits is smaller than a fractional value of 0.01. No tree pruning was performed. Additionally, the maximum number of features considered for splitting at each node was limited to the square root of the number of spectral data points used.

3. Results

3.1. MCMC

Figures 2 and 3 display the best-fit models for each of the spectra based on the Diamondback, Elf Owl, and SAND models, while Table 3 summarizes the corresponding posterior parameters. We find that the Diamondback models generally perform best for the early L dwarfs, reproducing the overall spectral energy distributions (SEDs) but not necessarily the detailed molecular features shortward of 1.25 μm. For the mid and late L dwarfs, the SAND models produce equivalent or better fits than Diamondback, better reproducing the redder spectral slopes of these sources, particularly for the young L dwarf G196-3B. The early T dwarf HN Peg B is also best fit by the SAND model despite the poor reproduction of the 1.6 μm CH_4 band in all of the T dwarf fits for this model. The Elf Owl models yield poor fits to L dwarf and early T dwarf spectra due to the absence of condensate opacity, but become the preferred models for the mid and late T dwarfs, particularly as we reach the $T_{\rm eff}$ parameter limits for the Diamondback and SAND models.

The inferred parameters for these fits largely follow expected trends, with declining effective temperatures with later spectral types, and lower surface gravities for lower-mass objects (e.g., the young source G 196-3B). For the

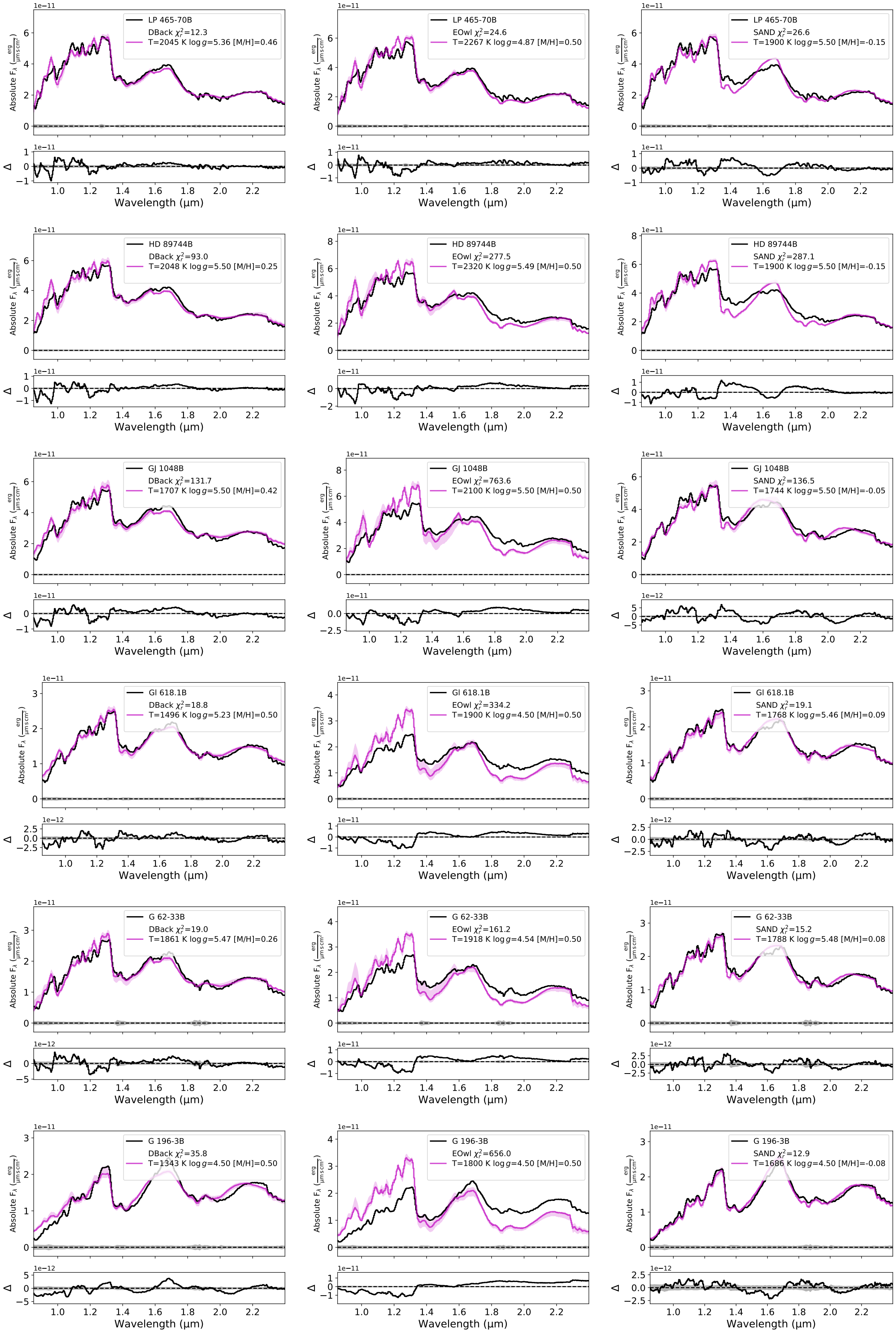

Figure 2. Absolute-flux-calibrated spectra of our benchmark brown dwarfs (black lines) compared to best-fit models (thick magenta lines) and posterior draws (magenta shading) for Diamondback (left), Elf Owl (center), and SAND models (right) based on our MCMC analysis. The bottom panel in each plot compares the difference (Δ = model – data) to the $\pm 1\sigma$ uncertainty spectrum (gray bands). Best-fit parameter values for $T_{\rm eff}$, $\log g$, and [M/H] are indicated in the legend; see full parameter values in Table 3.

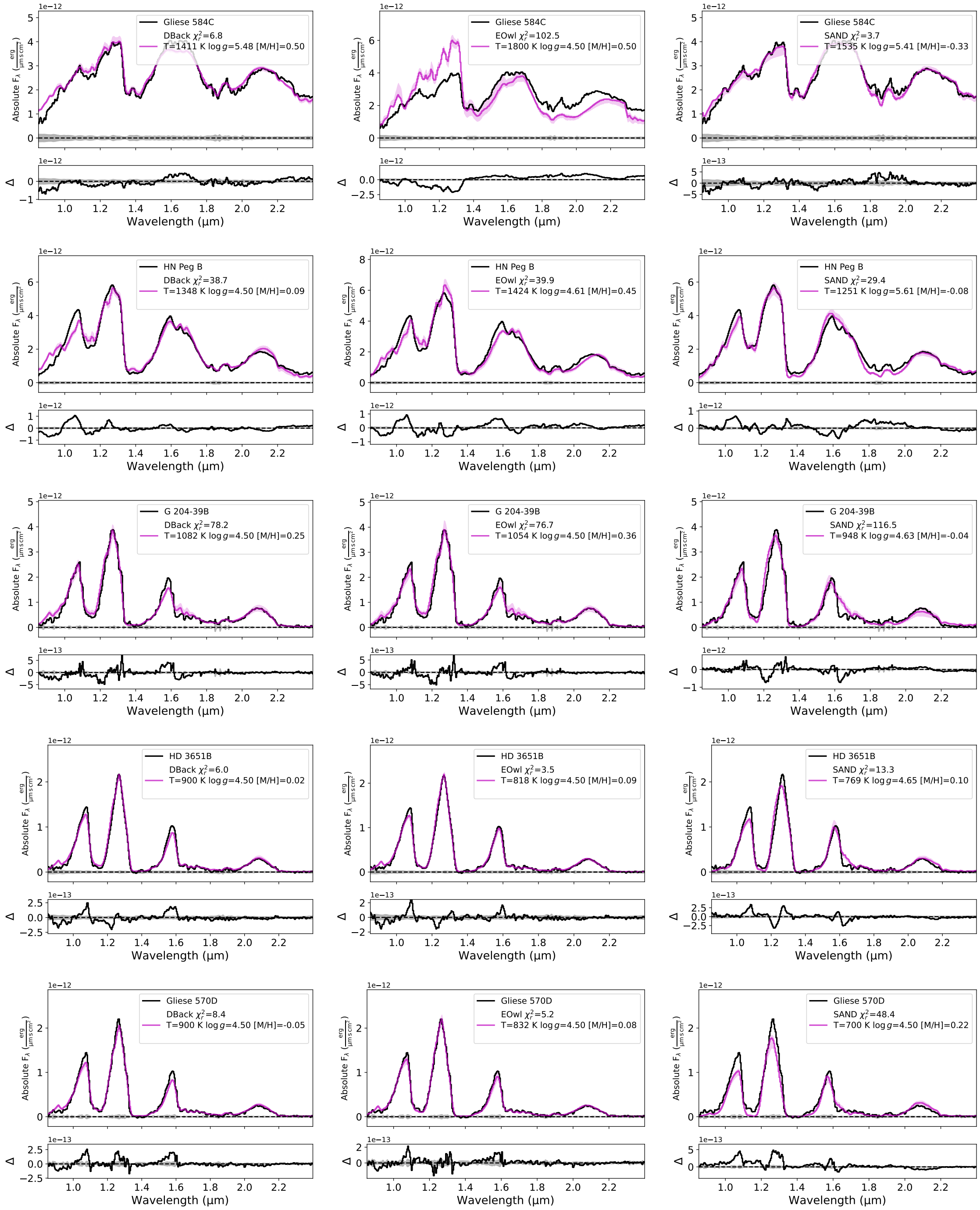

Figure 3. Figure 2 continued.

Diamondback models, the $f_{\rm sed}$ parameter is small for L dwarfs (condensates retained in the photosphere) and large for T dwarfs (condensates depleted in the photosphere), also as expected. We also find the radii of the best-fit models to be generally in the expected range of $0.075\ R_\odot \lesssim R \lesssim 0.105\ R_\odot$ based on evolutionary models (A. Burrows et al. 2001; I. Baraffe et al. 2003). There are, however, notable deviations for individual sources. For example, both Diamondback and SAND fits to the L1 GJ 1048B yield an inflated radius $R = 0.17\ R_\odot$, and a $T_{\rm eff} \approx 1700$ K, which is considerably lower than expected from the source spectral type. This deviation may be due to an interplay between condensate opacity and temperature, where a reduction in the former is compensated by a reduction in the latter. Conversely, the L8 Gliese 584C ($R = 0.064 \pm 0.001\ R_\odot$ from the SAND models) and the T6.5 G 204-39B ($R = 0.065 \pm 0.003\ R_\odot$ and $0.069^{+0.002}_{-0.004}\ R_\odot$ for the Diamondback and Elf Owl models, respectively) have radii somewhat lower than expected for brown dwarfs, suggesting that their inferred $T_{\rm eff}$ may be slightly overestimated. For Gliese 584C, this may reflect a phase of cloud clearing not captured in the SAND models.

3.2. RFR

The performance of the RFR models can be characterized by both the quality of fits and the training metrics of RvP

Table 3
MCMC Fitting Results

Source	SpT	$T_{\rm eff}$ (K)	log g (cm s^{-2})	[M/H]	C/O[a]	[α/Fe]	log κ_{zz} (cm^2 s^{-1})	$f_{\rm sed}$	R ($R_\odot$)	χ^2_r
Diamondback Model Set										
LP 465-70B	L0	2045^{+30}_{-23}	$5.35^{+0.09}_{-0.10}$	$0.41^{+0.07}_{-0.10}$	⋯	⋯	⋯	$1.26^{+0.20}_{-0.21}$	$0.111^{+0.003}_{-0.003}$	12[b]
HD 89744B	L0	2036^{+18}_{-29}	$5.50^{+0.00}_{-0.05}$	$0.22^{+0.06}_{-0.09}$	⋯	⋯	⋯	$1.00^{+0.02}_{-0.00}$	$0.118^{+0.003}_{-0.002}$	93[b]
GJ 1048B	L1	1696^{+11}_{-36}	$5.48^{+0.02}_{-0.10}$	$0.36^{+0.06}_{-0.13}$	⋯	⋯	⋯	$2.04^{+0.31}_{-0.12}$	$0.173^{+0.008}_{-0.002}$	132[b]
Gl 618.1B	L2.5	1499^{+21}_{-15}	$5.13^{+0.09}_{-0.14}$	$0.42^{+0.08}_{-0.12}$	⋯	⋯	⋯	$3.34^{+0.40}_{-0.87}$	$0.157^{+0.004}_{-0.004}$	19[b]
G 62-33B	L2.5	1650^{+211}_{-94}	$5.40^{+0.10}_{-0.26}$	$0.30^{+0.13}_{-0.12}$	⋯	⋯	⋯	$1.39^{+1.80}_{-0.39}$	$0.130^{+0.016}_{-0.024}$	19
G 196-3B	L3β	1343^{+13}_{-13}	$4.50^{+0.05}_{-0.00}$	$0.50^{+0.00}_{-0.07}$	⋯	⋯	⋯	$2.61^{+0.11}_{-0.38}$	$0.206^{+0.003}_{-0.005}$	36
Gliese 584C	L8	1415^{+12}_{-9}	$5.48^{+0.02}_{-0.02}$	$0.50^{+0.00}_{-0.03}$	⋯	⋯	⋯	$8.19^{+0.45}_{-0.72}$	$0.073^{+0.001}_{-0.001}$	7
HN Peg B	T2.5	1348^{+24}_{-23}	$4.59^{+0.12}_{-0.09}$	$0.10^{+0.06}_{-0.06}$	⋯	⋯	⋯	$7.89^{+0.59}_{-0.44}$	$0.068^{+0.002}_{-0.003}$	39
G 204-39B	T6.5	1087^{+25}_{-19}	$4.66^{+0.19}_{-0.15}$	$0.29^{+0.11}_{-0.06}$	⋯	⋯	⋯	$9.57^{+0.32}_{-0.25}$	$0.065^{+0.003}_{-0.003}$	78[b]
HD 3651B	T7.5	900^{+10}_{-0}	$4.58^{+0.10}_{-0.07}$	$0.08^{+0.09}_{-0.07}$	⋯	⋯	⋯	$9.54^{+0.35}_{-0.32}$	$0.067^{+0.000}_{-0.001}$	6
Gliese 570D	T7.5	900^{+6}_{-0}	$4.63^{+0.11}_{-0.09}$	$-0.02^{+0.06}_{-0.04}$	⋯	⋯	⋯	$9.65^{+0.25}_{-0.28}$	$0.065^{+0.000}_{-0.001}$	8
Elf Owl Model Set										
LP 465-70B	L0	2261^{+36}_{-37}	$4.76^{+0.16}_{-0.15}$	$0.42^{+0.08}_{-0.12}$	$1.47^{+0.03}_{-0.06}$	⋯	$2.9^{+0.9}_{-0.5}$	⋯	$0.090^{+0.003}_{-0.003}$	25
HD 89744B	L0	2320^{+5}_{-17}	$5.46^{+0.03}_{-0.18}$	$0.50^{+0.00}_{-0.10}$	$1.50^{+0.00}_{-0.02}$	⋯	$2.2^{+0.4}_{-0.2}$	⋯	$0.089^{+0.001}_{-0.000}$	277
GJ 1048B	L1	2100^{+0}_{-0}	$5.50^{+0.00}_{-0.07}$	$0.50^{+0.00}_{-0.00}$	$1.50^{+0.00}_{-0.00}$	⋯	$4.0^{+0.0}_{-0.0}$	⋯	$0.108^{+0.000}_{-0.000}$	764
Gl 618.1B	L2.5	1900^{+26}_{-0}	$4.50^{+0.17}_{-0.00}$	$0.50^{+0.00}_{-0.02}$	$1.50^{+0.00}_{-0.05}$	⋯	$4.0^{+0.0}_{-0.2}$	⋯	$0.093^{+0.000}_{-0.003}$	334
G 62-33B	L2.5	1918^{+58}_{-18}	$4.60^{+0.22}_{-0.06}$	$0.50^{+0.00}_{-0.06}$	$1.50^{+0.00}_{-0.08}$	⋯	$4.1^{+0.4}_{-0.4}$	⋯	$0.092^{+0.002}_{-0.006}$	161
G 196-3B	L3β	1800^{+41}_{-0}	$4.50^{+0.12}_{-0.00}$	$0.50^{+0.00}_{-0.01}$	$1.50^{+0.00}_{-0.08}$	⋯	$4.0^{+0.0}_{-0.0}$	⋯	$0.100^{+0.000}_{-0.005}$	656
Gliese 584C	L8	1800^{+80}_{-22}	$4.58^{+0.22}_{-0.08}$	$0.50^{+0.00}_{-0.09}$	$1.45^{+0.05}_{-0.13}$	⋯	$4.0^{+0.4}_{-0.5}$	⋯	$0.042^{+0.001}_{-0.004}$	102
HN Peg B	T2.5	1432^{+30}_{-22}	$4.65^{+0.15}_{-0.10}$	$0.45^{+0.05}_{-0.07}$	$1.25^{+0.07}_{-0.06}$	⋯	$2.5^{+0.6}_{-0.4}$	⋯	$0.059^{+0.002}_{-0.003}$	40
G 204-39B	T6.5	1054^{+26}_{-18}	$4.64^{+0.25}_{-0.14}$	$0.36^{+0.04}_{-0.14}$	$1.20^{+0.06}_{-0.11}$	⋯	$2.3^{+0.5}_{-0.3}$	⋯	$0.069^{+0.002}_{-0.004}$	77[b]
HD 3651B	T7.5	808^{+13}_{-14}	$4.50^{+0.08}_{-0.00}$	$0.09^{+0.07}_{-0.06}$	$1.18^{+0.08}_{-0.08}$	⋯	$3.0^{+0.5}_{-0.3}$	⋯	$0.083^{+0.003}_{-0.003}$	4[b]
Gliese 570D	T7.5	812^{+27}_{-21}	$4.57^{+0.09}_{-0.07}$	$0.02^{+0.07}_{-0.06}$	$1.06^{+0.11}_{-0.11}$	⋯	$3.0^{+0.7}_{-0.5}$	⋯	$0.082^{+0.006}_{-0.007}$	5[b]
SAND Model Set										
LP 465-70B	L0	1900^{+2}_{-3}	$5.50^{+0.01}_{-0.01}$	$-0.29^{+0.11}_{-0.06}$	⋯	$0.13^{+0.02}_{-0.01}$	⋯	⋯	$0.132^{+0.000}_{-0.001}$	27
HD 89744B	L0	1900^{+0}_{-0}	$5.50^{+0.00}_{-0.00}$	$-0.15^{+0.00}_{-0.00}$	⋯	$0.15^{+0.00}_{-0.00}$	⋯	⋯	$0.138^{+0.000}_{-0.000}$	287
GJ 1048B	L1	1744^{+0}_{-8}	$5.50^{+0.03}_{-0.00}$	$-0.05^{+0.00}_{-0.00}$	⋯	$0.02^{+0.00}_{-0.00}$	⋯	⋯	$0.171^{+0.002}_{-0.000}$	136
Gl 618.1B	L2.5	1778^{+23}_{-10}	$5.46^{+0.02}_{-0.03}$	$0.14^{+0.07}_{-0.03}$	⋯	$-0.01^{+0.03}_{-0.02}$	⋯	⋯	$0.115^{+0.001}_{-0.002}$	19[b]
G 62-33B	L2.5	1797^{+18}_{-9}	$5.47^{+0.02}_{-0.03}$	$0.13^{+0.05}_{-0.02}$	⋯	$0.00^{+0.03}_{-0.02}$	⋯	⋯	$0.112^{+0.001}_{-0.002}$	15[b]
G 196-3B	L3β	1687^{+32}_{-11}	$4.50^{+0.00}_{-0.00}$	$-0.08^{+0.03}_{-0.06}$	⋯	$0.03^{+0.01}_{-0.01}$	⋯	⋯	$0.129^{+0.002}_{-0.006}$	13[b]
Gliese 584C	L8	1523^{+14}_{-18}	$5.27^{+0.11}_{-0.09}$	$0.05^{+0.08}_{-0.14}$	⋯	$0.03^{+0.14}_{-0.03}$	⋯	⋯	$0.064^{+0.001}_{-0.001}$	4[b]
HN Peg B	T2.5	1250^{+17}_{-8}	$5.59^{+0.10}_{-0.09}$	$-0.08^{+0.11}_{-0.08}$	⋯	$0.13^{+0.03}_{-0.02}$	⋯	⋯	$0.082^{+0.001}_{-0.005}$	29[b]
G 204-39B	T6.5	935^{+22}_{-25}	$4.70^{+0.13}_{-0.05}$	$-0.09^{+0.05}_{-0.04}$	⋯	$0.02^{+0.01}_{-0.01}$	⋯	⋯	$0.099^{+0.006}_{-0.005}$	117
HD 3651B	T7.5	769^{+24}_{-39}	$4.52^{+0.10}_{-0.02}$	$0.10^{+0.01}_{-0.00}$	⋯	$0.02^{+0.01}_{-0.03}$	⋯	⋯	$0.101^{+0.013}_{-0.010}$	13
Gliese 570D	T7.5	700^{+37}_{-0}	$4.50^{+0.04}_{-0.00}$	$0.22^{+0.04}_{-0.12}$	⋯	$-0.02^{+0.04}_{-0.01}$	⋯	⋯	$0.114^{+0.001}_{-0.013}$	48

Notes.
[a] C/O ratio relative to solar abundance, with $(\mathrm{C/O})_\odot = 0.458$ (K. Lodders et al. 2009).
[b] Best-fit model or models (if $\Delta\chi^2_r \leqslant 1$) for this source.

parameters and feature importance. Figure 4 shows the RvP plots for each of the model parameters and for each model. Very high self-consistency is found for effective temperature, with $R^2 > 0.95$ for all of the grids. For the other parameters, R^2 varies widely across the models, generally showing worse performance when fewer parameter values are available. For example, for metallicity R^2 ranges from 0.05 for the SAND models to 0.90 for Elf Owl.

Figures 5–7 display the feature importance diagrams for individual model parameters for the three atmosphere models, compared to the scaled spectrum of HD 89744B to guide assessment of the relevant spectral features. For $T_{\rm eff}$, the Diamondback and SAND models show a consistent structure in feature importance, aligning with the strong molecular absorption bands of H_2O at 0.92, 1.1, and 1.4 μm, and the red optical spectral slope. The Elf Owl models differ slightly, with the strongest feature corresponding to H_2O at 1.4 μm but additional features aligned with the 1.8 μm H_2O band and the 2.3 μm CO band. For surface gravity, all grids consistently indicate that the wavelength range below 1.2 μm is crucial, consistent with the findings of A. Lueber et al. (2023). The lowest wavelengths are particularly sensitive to the presence and strength of the K I line wings centered at 0.77 μm, which are known to be key indicators of an object's surface gravity (N. F. Allard et al. 2005). An additional wavelength range of note in the SAND model grid is the 2.2–2.4 μm region, which likely traces collision-induced H_2 absorption known to be sensitive to atmospheric pressure and thus surface gravity in

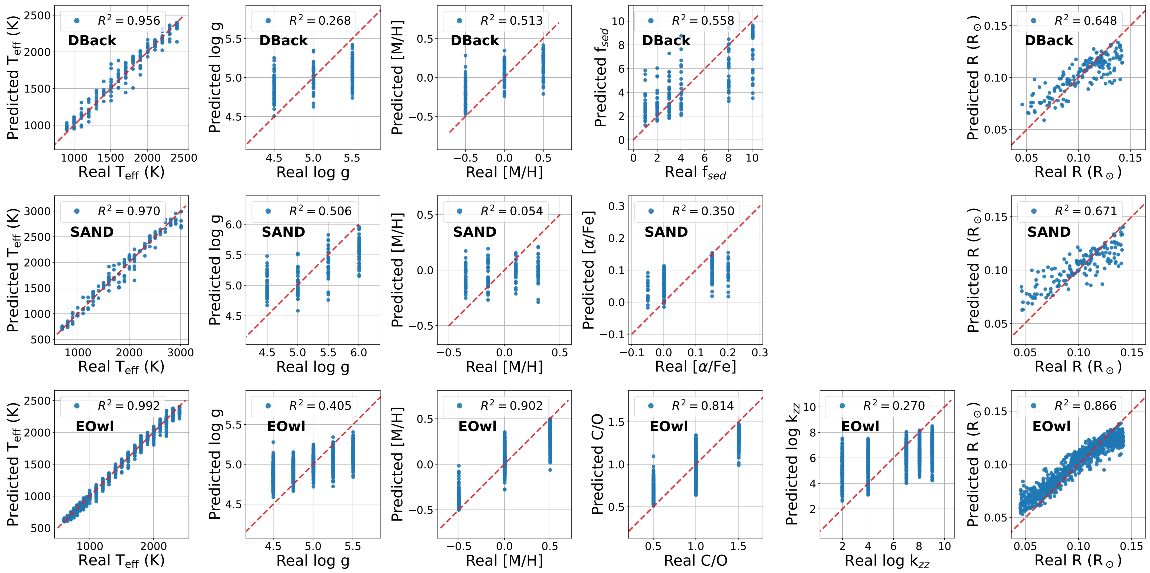

Figure 4. RvP comparison for RF model training and testing on the evaluated model grids. Each row represents one grid with its corresponding model parameters. The red dashed line in each panel indicates perfect agreement ($R^2 = 1$).

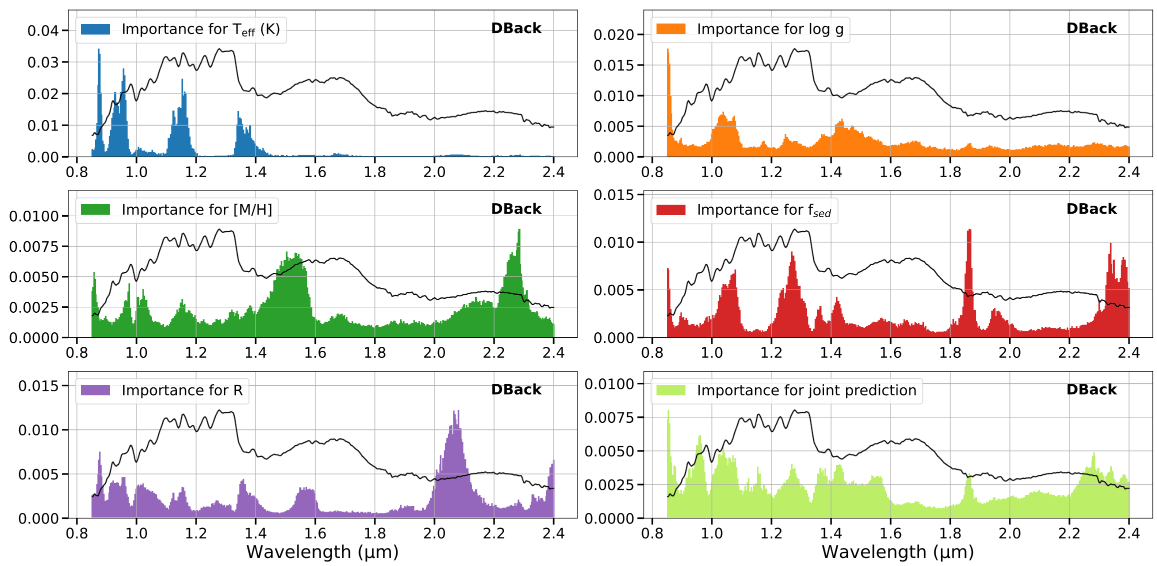

Figure 5. Feature importance plots for the RFR model trained on the Diamondback grid for temperature $T_{\rm eff}$ (blue), surface gravity log g (orange), metallicity [M/H] (green), cloud sedimentation efficiency $f_{\rm sed}$ (red), radius R (calculated from the calibration factor f; purple), and joint retrieval (lime). For comparison, the scaled observed spectrum of HD 89744B (solid, black line) has been added to each plot.

T dwarfs (J. L. Linsky 1969; A. J. Burgasser et al. 2002a). Feature importance patterns for [M/H] across all model grids, as well as for C/O in the Elf Owl models and alpha enhancement [α/Fe] in the SAND models, show similar structures aligned with the strong molecular absorption bands of H_2O, CO, and CH_4. Peaks are also found at wavelengths below 1.2 μm, again likely due to the influence of alkali line features. Similarly, the parameter log κ_{zz} in the Elf Owl models has the greatest feature importance shortward of 0.9 μm and in the CO band, indicating comparable abundance effects from vertical mixing. The parameter $f_{\rm sed}$ in the Diamondback models shows several discrete peaks that overlap with the 1.05 and 1.25 μm flux peaks, where condensate grain scattering is most prominent, and the 1.8 μm H_2O and 2.3 μm CO bands, which are relatively enhanced in the dustiest L dwarfs (A. C. Schneider et al. 2023a). Finally, for the radius scaling parameter f, all model grids behave differently. The most significant feature in the Diamondback model grid is centered at 2.1 μm, a pseudocontinuum region of the L dwarfs and a flux peak for the T dwarfs. The Elf Owl model grid shows numerous discrete features associated with flux peaks at 1.05, 1.25, and 1.58 μm, and absorption features from H_2O at 1.4 μm and CO at 2.3 μm. The SAND model grid shows different features associated with the $<$0.9 μm red optical slope, the 2.1 μm flux peak, and the 1.6 μm CH_4 absorption band. These distinctions in feature importance indicate that a direct comparison between different model grids should be treated with caution.

Figures 8 and 9 display the best-fit models for each of the spectra based on the Diamondback, Elf Owl, and SAND models, while Table 4 summarizes the corresponding posterior parameters. As observed in prior work on RFRs (A. Lueber et al. 2023), the

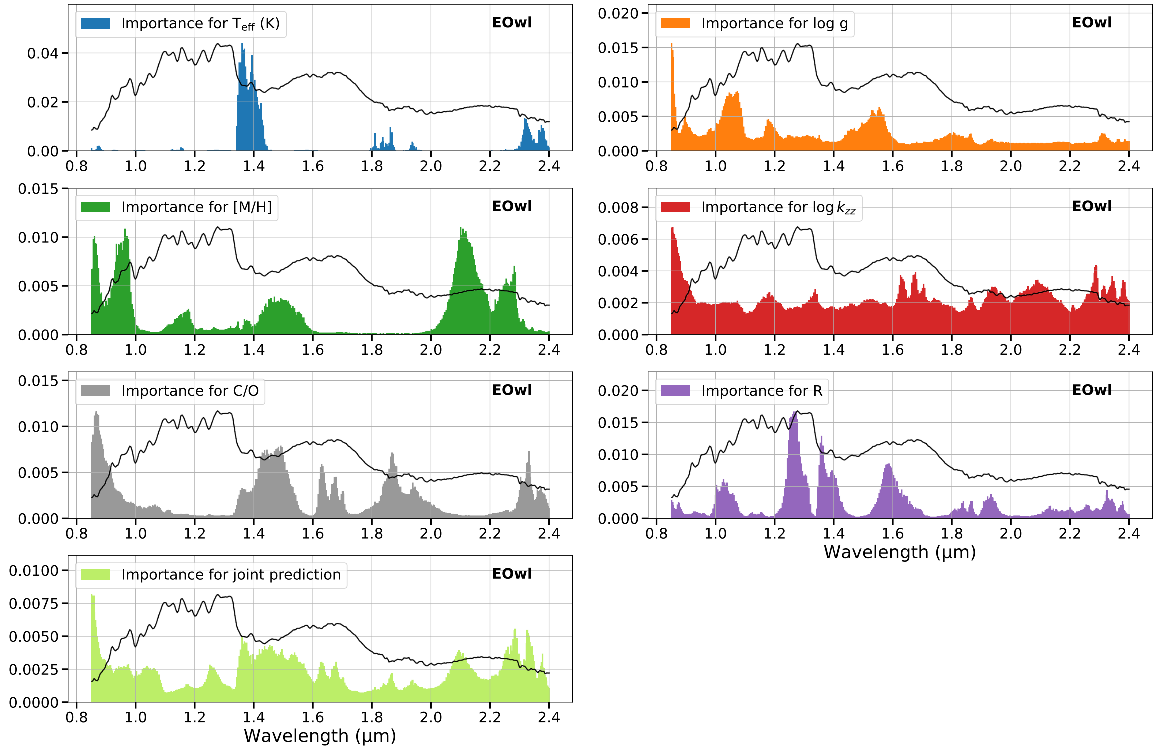

Figure 6. Same as Figure 5 but for the Elf Owl model grid—for temperature $T_{\rm eff}$ (blue), surface gravity $\log g$ (orange), metallicity [M/H] (green), vertical eddy diffusion coefficient $\log \kappa_{zz}$ (red), carbon-to-oxygen ratio C/O (gray), radius R (purple), and joint retrieval (lime).

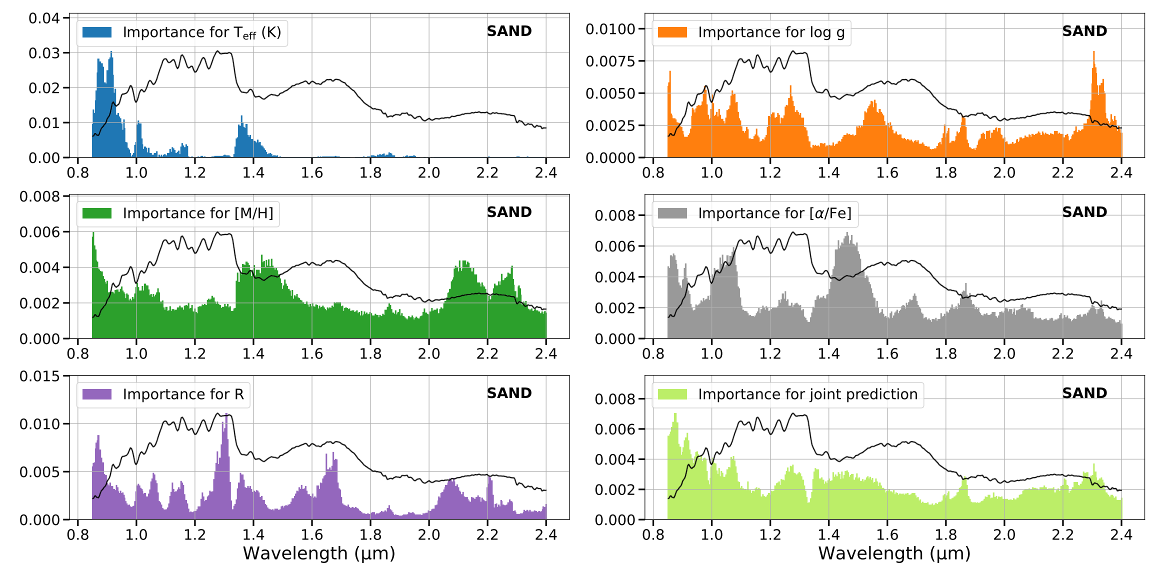

Figure 7. Same as Figure 5 but for the SAND model grid—for temperature $T_{\rm eff}$ (blue), surface gravity $\log g$ (orange), metallicity [M/H] (green), alpha element enrichment $[\alpha/\mathrm{Fe}]$ (gray), radius R (purple), and joint retrieval (lime).

median retrieved spectra show significant differences between the three grids, with both over- and underestimates of spectral fluxes as compared to the measured spectra of up to an order of magnitude. Nevertheless, each source can be reasonably well fit by one of the model grids, following the same pattern as the MCMC fits: L dwarfs and early T dwarfs are best fit by the Diamondback models, mid- and late-type L dwarfs (most notably G 196-3B) by the SAND models, and mid- and late-type T dwarfs by the Elf Owl models. Detailed fits of molecular features shortward of 1.3 μm in the early L dwarfs are again problematic for all three models, with the L1 GJ 1048B showing the worst fits ($\chi_r^2 > 100$), while the SAND models are unable to properly reproduce the 1.6 μm CH_4 band. We also see for these fits that the Diamondback and Elf Owl models diverge for the latest T dwarfs, with the latter providing clearly superior fits. As in the MCMC fits, we see the anticipated trends among best-fit parameters from the RFR modeling, with effective temperatures decreasing toward later spectral types, lower surface gravities for the young brown dwarfs G 196-3B and HN Peg B, and $f_{\rm sed}$ increasing across the L/T transition for the Diamondback models. Furthermore, while

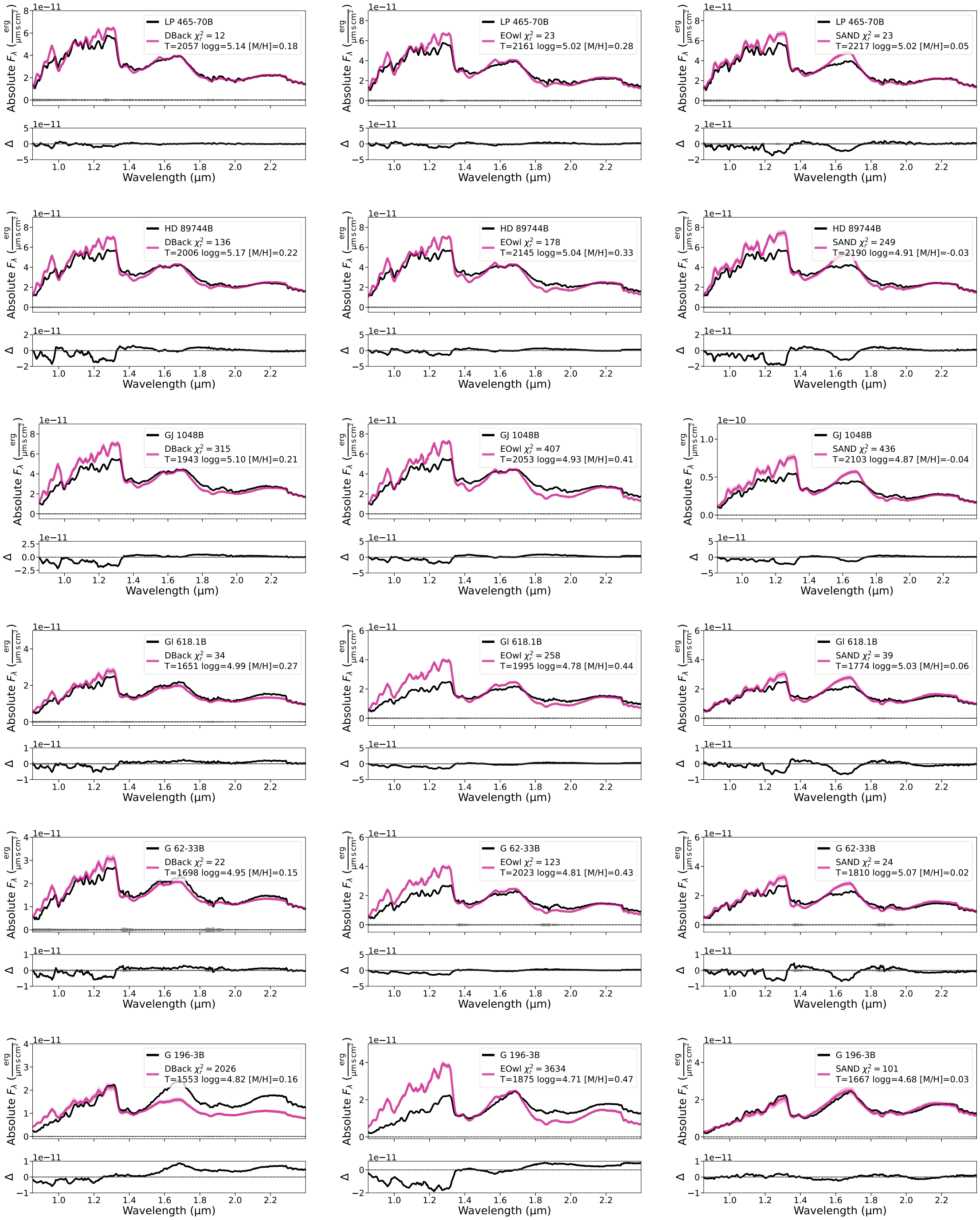

Figure 8. Same as Figure 2 but for RFR spectral fits for the Diamondback (DBack, left), Elf Owl (EOwl, middle), and SAND (right) model grids. The bottom panel in each plot compares the difference (Δ = model – data) to the $\pm 1\sigma$ uncertainty spectrum (gray bands). Magenta curves display the median fits while shaded areas display the 1σ uncertainties based on posterior draws.

the best-fit radius for GJ 1048B from the Diamondback fits, $R = 0.130^{+0.008}_{-0.006}\ R_\odot$, is higher than expected from evolutionary models, it is not as extreme as that in the MCMC fit.

3.3. Comparison of MCMC and RFR Approaches

3.3.1. Comparison of Fit Quality

The MCMC and RFR approaches generally yield comparable results in terms of preferred model grid and inferred parameters, albeit with some informative differences. First, the quality of the best fits as quantified by χ_r^2 is consistently better for the MCMC approach. In several cases the difference is substantial; for example, SAND fits to the L3β G 196-3B yield $\chi_r^2 = 101$ for RFR versus $\chi_r^2 = 13$ for MCMC, the latter being an obviously better fit (see Figures 2 and 8). On the other hand, Elf Owl fits for T6.5 G 204-39B are significantly better for the RFR approach ($\chi_r^2 = 38$) than for the MCMC approach ($\chi_r^2 = 77$). There are also cases where the best-fit model set differs between approaches; for example, the best MCMC fit for HN Peg B comes from the SAND models, while the best RFR fit comes from the Diamondback model, both having identical $\chi_r^2 = 29$. As such, while both approaches yield

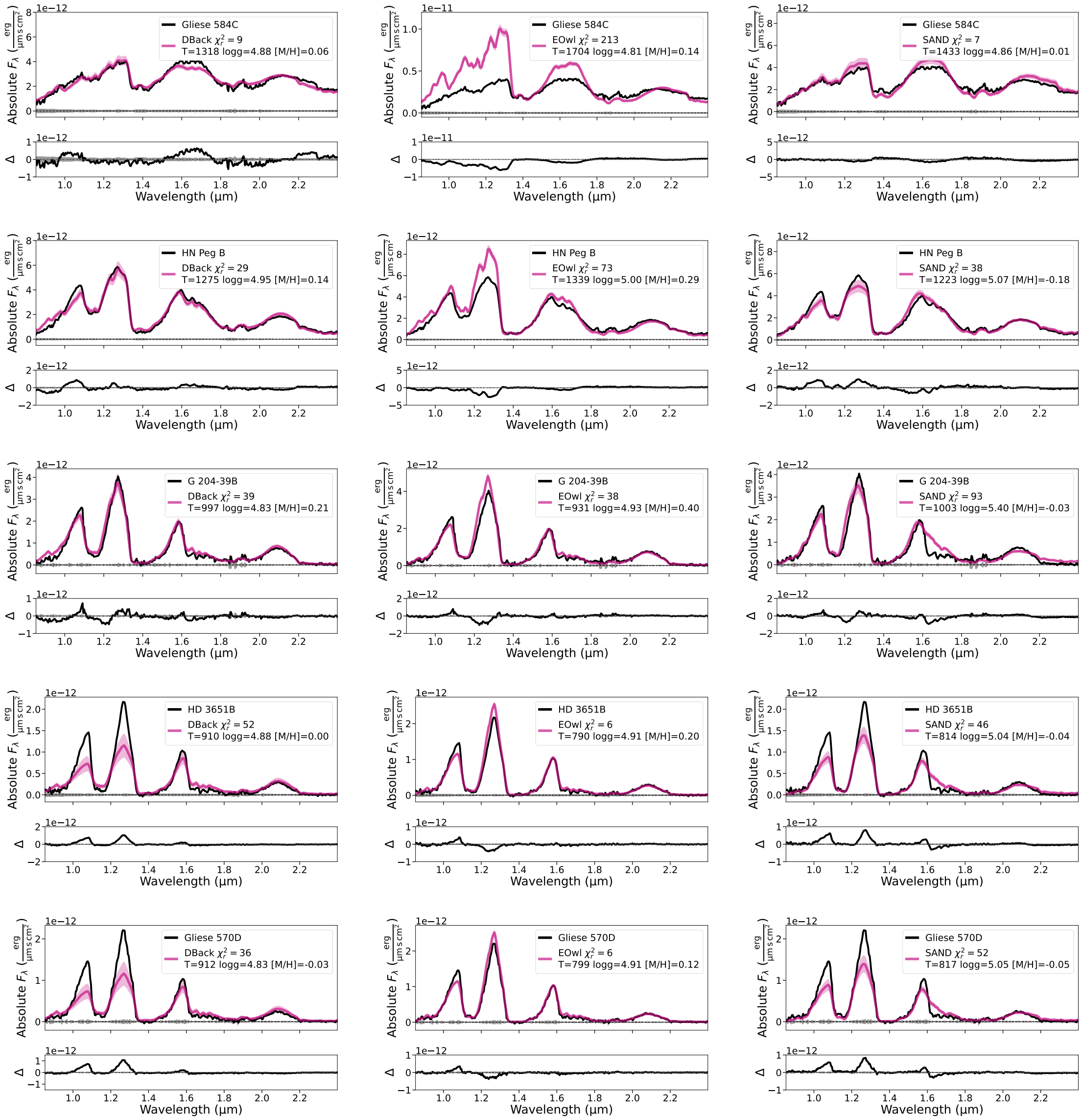

Figure 9. Figure 8 continued.

broadly comparable fits, the MCMC approach provides on average a better quality of fit.

3.3.2. Comparison of Fit Parameters

Figure 10 compares the fit parameters and 1σ uncertainties for the MCMC and RFR approaches for each of our sources, with the best-fit model for each approach highlighted. There is a clear difference in the inferred parameter uncertainties, with the RFR uncertainties being significantly larger. For example, typical uncertainties in $T_{\rm eff}$ are $\sim$5%–10% for RFR and $\sim$1% for MCMC, while for $\log g$ uncertainties are $\sim$0.5 dex for RFR and $\sim$0.1 dex for MCMC. The lower uncertainties for the MCMC parameters may partly reflect the better fidelity in reproducing the observed spectra, but may also indicate underestimation of parameter uncertainties in the approach applied here. Within their mutual uncertainties, the parameter values are largely consistent for the best-fit models, with a few notable exceptions. The aforementioned low $T_{\rm eff}$ and inflated radius from MCMC Diamondback model fits to the L1 GJ 1048B are significantly distinct from the parameters from the RFR SAND model fits for this source, which have a more reasonable (but more uncertain) $T_{\rm eff} = 2100 \pm 200$ K and radius $R = 0.116^{+0.019}_{-0.027}\,R_\odot$. In contrast, RFR Elf Owl model fits yield a marginally larger and more realistic radius ($R = 0.093^{+0.024}_{-0.025}\,R_\odot$) for the T6.5 G 204-39B as compared to the MCMC fits. We also find that the surface gravities from the RFR Elf Owl model fits to the late T dwarfs HD 3651B and Gliese 570D are higher than the MCMC values for these sources and more in line with those expected for older brown dwarfs, while the surface gravity for the best-fit RFR Diamondback model fits for HN Peg B is lower than that for the MCMC SAND model fits for this source and more in line with this system's young age. Even among these variances, the best-fit MCMC and RFR model parameters are generally consistent with each other, albeit within the relatively large uncertainties of the RFR values.

3.3.3. Comparison to Evolutionary Model Predictions

Figure 10 compares the fit parameters to the expected evolutionary-model-based estimates of $T_{\rm eff}$, $\log g$, and radius from A. Sanghi et al. (2023; see Table 1). While there is broad

Table 4
RFR Results

Source	SpT	$T_{\rm eff}$ (K)	log g (cm s^{-2})	[M/H]	C/O[a]	[α/Fe]	log κ_{zz} (cm^2 s^{-1})	$f_{\rm sed}$	R ($R_\odot$)	χ_r^2
Diamondback Model Set										
LP 465-70B	L0	2057^{+243}_{-157}	$5.14^{+0.36}_{-0.64}$	$0.18^{+0.33}_{-0.18}$	…	…	…	$2.9^{+1.1}_{-1.9}$	$0.112^{+0.020}_{-0.023}$	12[b]
HD 89744B	L0	2006^{+294}_{-106}	$5.17^{+0.33}_{-0.17}$	$0.22^{+0.28}_{-0.22}$	…	…	…	$2.7^{+1.3}_{-1.7}$	$0.122^{+0.015}_{-0.027}$	136[b]
GJ 1048B	L1	1943^{+57}_{-143}	$5.10^{+0.40}_{-0.35}$	$0.21^{+0.29}_{-0.21}$	…	…	…	$2.4^{+0.6}_{-1.4}$	$0.130^{+0.008}_{-0.006}$	315[b]
Gl 618.1B	L2.5	1651^{+49}_{-109}	$4.99^{+0.51}_{-0.49}$	$0.27^{+0.23}_{-0.27}$	…	…	…	$2.2^{+0.8}_{-1.2}$	$0.124^{+0.015}_{-0.010}$	34[b]
G 62-33B	L2.5	1698^{+102}_{-98}	$4.95^{+0.56}_{-0.45}$	$0.15^{+0.35}_{-0.65}$	…	…	…	$2.6^{+0.4}_{-1.6}$	$0.120^{+0.020}_{-0.014}$	22[b]
G 196-3B	L3β	1553^{+47}_{-53}	$4.82^{+0.68}_{-0.32}$	$0.16^{+0.34}_{-0.46}$	…	…	…	$2.4^{+0.6}_{-1.4}$	$0.125^{+0.012}_{-0.010}$	2026
Gliese 584C	L8	1318^{+130}_{-143}	$4.88^{+0.12}_{-0.38}$	$0.06^{+0.44}_{-0.56}$	…	…	…	$2.9^{+1.1}_{-1.9}$	$0.090^{+0.028}_{-0.024}$	9
HN Peg B	T2.5	1275^{+225}_{-175}	$4.95^{+0.55}_{-0.45}$	$0.14^{+0.36}_{-0.64}$	…	…	…	$7.0^{+3.0}_{-3.0}$	$0.080^{+0.028}_{-0.025}$	29[b]
G 204-39B	T6.5	997^{+103}_{-97}	$4.83^{+0.17}_{-0.33}$	$0.21^{+0.29}_{-0.21}$	…	…	…	$8.6^{+1.4}_{-0.6}$	$0.082^{+0.024}_{-0.016}$	39[b]
HD 3651B	T7	910^{+10}_{-10}	$4.88^{+0.48}_{-0.38}$	$0.00^{+0.50}_{-0.50}$	…	…	…	$5.9^{+2.1}_{-2.9}$	$0.060^{+0.010}_{-0.012}$	52
Gliese 570D	T8	912^{+12}_{-12}	$4.83^{+0.67}_{-0.33}$	$-0.03^{+0.19}_{-0.47}$	…	…	…	$6.0^{+2.0}_{-3.0}$	$0.059^{+0.011}_{-0.012}$	36
Elf Owl Model Set										
LP 465-70B	L0	2161^{+239}_{-161}	$5.02^{+0.48}_{-0.27}$	$0.28^{+0.22}_{-0.28}$	$1.30^{+0.20}_{-0.30}$	…	$5.8^{+3.1}_{-3.8}$	…	$0.100^{+0.015}_{-0.018}$	23
HD 89744B	L0	2145^{+155}_{-145}	$5.04^{+0.46}_{-0.29}$	$0.33^{+0.17}_{-0.33}$	$1.45^{+0.05}_{-0.05}$	…	$5.4^{+2.6}_{-3.4}$	…	$0.105^{+0.021}_{-0.019}$	178
GJ 1048B	L1	2053^{+147}_{-153}	$4.93^{+0.32}_{-0.43}$	$0.41^{+0.09}_{-0.41}$	$1.47^{+0.03}_{-0.03}$	…	$4.3^{+3.7}_{-2.3}$	…	$0.114^{+0.015}_{-0.014}$	407
Gl 618.1B	L2.5	1995^{+205}_{-195}	$4.78^{+0.22}_{-0.28}$	$0.44^{+0.06}_{-0.06}$	$1.41^{+0.09}_{-0.41}$	…	$4.7^{+3.3}_{-2.7}$	…	$0.090^{+0.020}_{-0.020}$	258
G 62-33B	L2.5	2023^{+277}_{-223}	$4.81^{+0.20}_{-0.31}$	$0.43^{+0.07}_{-0.07}$	$1.44^{+0.06}_{-0.06}$	…	$4.7^{+3.3}_{-2.7}$	…	$0.088^{+0.017}_{-0.020}$	123
G 196-3B	L3β	1875^{+125}_{-175}	$4.71^{+0.29}_{-0.21}$	$0.47^{+0.03}_{-0.03}$	$1.36^{+0.14}_{-0.36}$	…	$4.9^{+3.1}_{-2.9}$	…	$0.098^{+0.017}_{-0.019}$	3634
Gliese 584C	L8	1704^{+96}_{-104}	$4.81^{+0.19}_{-0.31}$	$0.14^{+0.36}_{-0.14}$	$1.15^{+0.35}_{-0.65}$	…	$6.0^{+3.0}_{-4.0}$	…	$0.057^{+0.007}_{-0.008}$	213
HN Peg B	T2.5	1339^{+161}_{-139}	$5.00^{+0.50}_{-0.50}$	$0.29^{+0.21}_{-0.29}$	$1.27^{+0.23}_{-0.27}$	…	$5.8^{+2.2}_{-3.8}$	…	$0.073^{+0.015}_{-0.015}$	73
G 204-39B	T6.5	931^{+69}_{-81}	$4.93^{+0.43}_{-0.43}$	$0.40^{+0.10}_{-0.40}$	$1.36^{+0.14}_{-0.36}$	…	$5.2^{+2.8}_{-3.2}$	…	$0.093^{+0.024}_{-0.025}$	38[b]
HD 3651B	T7	790^{+110}_{-90}	$4.91^{+0.34}_{-0.41}$	$0.20^{+0.30}_{-0.20}$	$1.21^{+0.29}_{-0.21}$	…	$4.5^{+3.5}_{-2.5}$	…	$0.092^{+0.032}_{-0.026}$	6[b]
Gliese 570D	T8	799^{+101}_{-99}	$4.91^{+0.34}_{-0.41}$	$0.12^{+0.39}_{-0.12}$	$1.16^{+0.34}_{-0.66}$	…	$5.2^{+2.8}_{-3.2}$	…	$0.090^{+0.028}_{-0.025}$	6[b]
SAND Model Set										
LP 465-70B	L0	2217^{+157}_{-317}	$5.02^{+0.48}_{-0.52}$	$0.05^{+0.25}_{-0.20}$	…	$0.03^{+0.12}_{-0.03}$	…	…	$0.099^{+0.029}_{-0.018}$	23
HD 89744B	L0	2190^{+210}_{-289}	$4.91^{+0.59}_{-0.41}$	$-0.03^{+0.33}_{-0.32}$	…	$0.02^{+0.13}_{-0.07}$	…	…	$0.106^{+0.027}_{-0.019}$	249
GJ 1048B	L1	2103^{+197}_{-203}	$4.87^{+0.63}_{-0.37}$	$-0.04^{+0.34}_{-0.32}$	…	$0.01^{+0.01}_{-0.06}$	…	…	$0.116^{+0.019}_{-0.027}$	436
Gl 618.1B	L2.5	1774^{+126}_{-174}	$5.03^{+0.47}_{-0.53}$	$0.06^{+0.24}_{-0.21}$	…	$0.06^{+0.09}_{-0.11}$	…	…	$0.115^{+0.023}_{-0.019}$	39
G 62-33B	L2.5	1810^{+190}_{-110}	$5.07^{+0.43}_{-0.57}$	$0.02^{+0.28}_{-0.37}$	…	$0.06^{+0.09}_{-0.11}$	…	…	$0.112^{+0.026}_{-0.022}$	24
G 196-3B	L3β	1667^{+33}_{-67}	$4.68^{+0.32}_{-0.18}$	$0.03^{+0.09}_{-0.18}$	…	$0.00^{+0.00}_{-0.05}$	…	…	$0.129^{+0.012}_{-0.013}$	101[b]
Gliese 584C	L8	1433^{+167}_{-233}	$4.86^{+0.65}_{-0.36}$	$0.01^{+0.29}_{-0.36}$	…	$0.12^{+0.08}_{-0.12}$	…	…	$0.076^{+0.019}_{-0.017}$	7[b]
HN Peg B	T2.5	1223^{+177}_{-123}	$5.07^{+0.43}_{-0.57}$	$-0.18^{+0.05}_{-0.17}$	…	$0.04^{+0.11}_{-0.09}$	…	…	$0.089^{+0.030}_{-0.024}$	38
G 204-39B	T6.5	1003^{+197}_{-103}	$5.40^{+0.60}_{-0.40}$	$-0.03^{+0.33}_{-0.32}$	…	$0.06^{+0.09}_{-0.11}$	…	…	$0.088^{+0.026}_{-0.035}$	93
HD 3651B	T7	814^{+86}_{-114}	$5.04^{+0.47}_{-0.54}$	$-0.04^{+0.14}_{-0.32}$	…	$0.06^{+0.14}_{-0.07}$	…	…	$0.086^{+0.028}_{-0.028}$	46
Gliese 570D	T8	817^{+83}_{-117}	$5.05^{+0.45}_{-0.55}$	$-0.05^{+0.15}_{-0.30}$	…	$0.06^{+0.14}_{-0.06}$	…	…	$0.085^{+0.026}_{-0.027}$	52

Notes.
[a] C/O ratio relative to solar abundance, with $({\rm C/O})_\odot = 0.458$ (K. Lodders et al. 2009).
[b] Best-fit model or models (if $\Delta\chi_r^2 \leqslant 1$) for this source.

agreement in the spectral type trends among these values, there are several deviations that merit scrutiny. Across the early L dwarf sequence, the best-fit $T_{\rm eff}$ and radii from the Diamondback models are consistently lower and larger, respectively, compared to the expected values—more so for the MCMC fits. This trend reverses for the T dwarfs, although the RFR Elf Owl model fits for the mid and late T dwarfs are consistent with expected temperatures and radii. The SAND models show particularly good agreement with expected $T_{\rm eff}$ and radii across the full sample, even when providing the lowest-quality fit to the spectral data, with a notable match to the inflated radius expected for the young L3β G196-3B. With regard to surface gravities, the Diamondback and Elf Owl fits are consistently lower than expectations for the L dwarfs, in line with inflated radii, and are essentially constant for the T dwarfs. The SAND models show differing behaviors between the MCMC and RFR approaches across the sample, the latter generally underpredicting $\log g$ for L dwarfs and overpredicting $\log g$ for T dwarfs. Overall, the early L dwarfs show the largest deviations between fit parameters and evolutionary-model-based estimates, whereas parameters for the mid and late L dwarfs and T dwarfs are generally consistent with these estimates.

3.3.4. Comparison to System Parameters

As benchmark companions, our sample has independent constraints on the atmosphere parameters of metallicity and surface gravity based on the elemental composition and ages

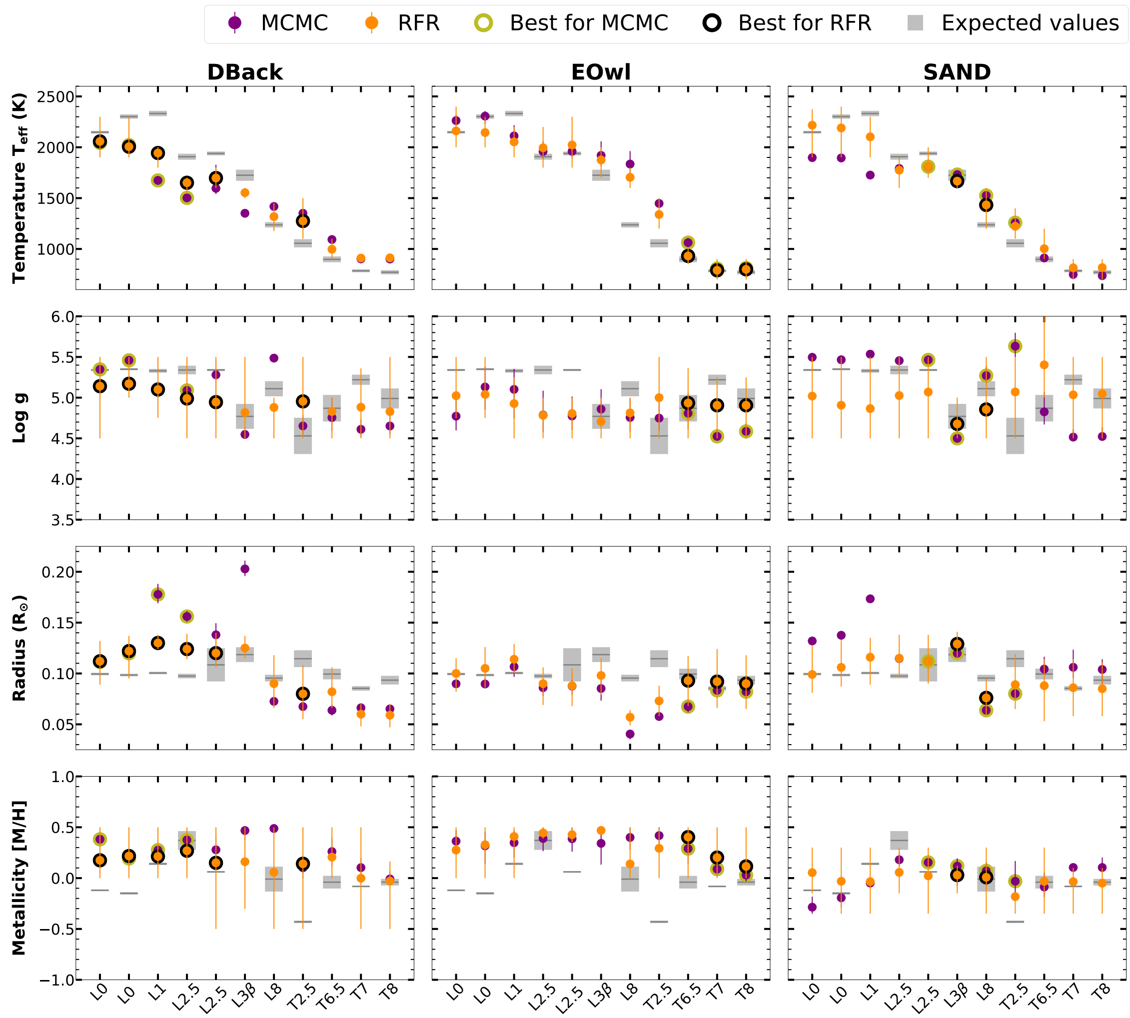

Figure 10. Model fit parameters and 1σ uncertainties for our full L and T dwarf sequence, comparing the MCMC (purple points) and RFR (orange points) approaches, and the Diamondback (DBack, left), Elf Owl (EOwl, middle), and SAND (right) model grids. The parameters shown from top to bottom are the effective temperature ($T_{\rm eff}$ in units of K), surface gravity ($\log g$ in log units of cm s^{-2}), radius (R in units of $R_{\odot}$), and metallicity ([M/H] in log units relative to solar). For each method, the models with the best χ^2 values for each object are highlighted with circles (green for MCMC, black for RFR). Gray bars indicate the expected values from evolutionary models ($T_{\rm eff}$, $\log g$, and radius) or primaries (metallicity) as listed in Table 1.

of the host primaries. For metallicity, Figure 10 shows that there are mixed outcomes when comparing between fit metallicities and primary iron abundances, with about half of the fits yielding metallicities that are too high. The Elf Owl models consistently yield supersolar metallicities for the L dwarfs and early/mid-type T dwarfs for both approaches, often reaching the limits of the parameter grid. This positive metallicity skew and similarly inflated C/O ratios likely compensate for missing condensate opacity in the Elf Owl model grid, as the two late T dwarfs HD 3651B and Gliese 570D show no such skew. In contrast, the SAND models yield metallicities that are largely consistent with the iron abundances of their primaries, even when the model does not provide the best fit overall. Metallicities from the Diamondback models are too uncertain for us to make a reliable comparison. The better performance of the SAND models for metallicity inference is somewhat surprising given their low RvP value for this parameter (Figure 4).

For surface gravities, while these are skewed low for the L dwarfs they are generally in line with the ages of the host systems across our sample. High surface gravities for old systems such as LP 465-70B (Diamondback $\log g = 5.35^{+0.09}_{-0.10}$ for MCMC, $\log g = 5.14^{+0.36}_{-0.64}$ for RFR) are consistent with ages in the $\tau \gtrsim 2$ Gyr range, while low surface gravities for young systems such as G 196-3B (SAND $\log g \leqslant 4.5$ for MCMC, $\log g = 4.68^{+0.32}_{-0.18}$ for RFR) are consistent with ages in the $\tau \lesssim 1$ Gyr range, based on evolutionary models for the effective temperatures of these sources (A. Burrows et al. 2001; I. Baraffe et al. 2003). There are again exceptions. The T2.5 dwarf HN Peg B, part of the 240 ± 30 Myr HN Peg system (S. A. Barnes 2007), has a best-fit MCMC SAND model with $\log g = 5.59^{+0.10}_{-0.09}$, consistent with an age $\gtrsim$ 5 Gyr.[4] Similarly, the best-fit MCMC Elf Owl models for the late T dwarfs HD 3651B and Gliese 570D yield $\log g = 4.5$–4.6 implying ages $\lesssim$ 500 Myr, considerably younger than their primaries.

3.4. Case Studies

Here we examine in detail fits to three representative objects that encompass our sample: the early L dwarf LP 465-70B, the L/T transition object Gliese 584C, and the late T dwarf Gliese

[4] The RFR Diamondback models yield $\log g$ values that are too uncertain for a robust age test.

570D. All three objects show robust fits to one of the model sets with the parameters listed in Tables 3 and 4. In addition to the spectral fits shown above, we display joint posterior parameter distributions for both MCMC and RFR approaches for these three sources in Figures 11 through 16 of the Appendix.

3.4.1. LP 465-70B and Diamondback Models

The spectrum of the L0 LP 465-70B is best fit by the Diamondback models, reflecting the important role of condensates in shaping the infrared spectrum of this source. Both MCMC and RFR fits yield the same $\chi_r^2 = 12$, albeit with different deviations from the data. The MCMC fit accurately reproduces the SED but fails to match detailed absorption features shortward of 1.2 μm, particularly around the 0.99 μm FeH band; the RFR fit overpredicts the 1.1–1.3 μm continuum and shows similar feature divergences shortward of 1.0 μm. All of the fit parameters are formally consistent between the models, with only a 12 K difference in $T_{\rm eff}$ and nearly identical radii. The condensation parameter $f_{\rm sed}$ is inferred to be low for both fits ($1.26^{+0.20}_{-0.21}$ for MCMC, $2.9^{+1.1}_{-1.9}$ for RFR), indicating significant condensate opacity in the atmosphere of this L dwarf. Figures 11 and 12 demonstrate that these parameters are well constrained, with single peaks in the marginalized distributions that are set in part by the parameter limits for $\log g$, [M/H], and $f_{\rm sed}$. We find only one (expected) negative correlation between $T_{\rm eff}$ and radius. The inferred temperatures are $\sim$100 K cooler and the inferred radii 13% larger than the evolutionary model estimates of A. Sanghi et al. (2023), while the MCMC surface gravity is consistent with the older age of this system ($\tau \geqslant 4.5$ Gyr). Both fits also yield a high metallicity ([M/H] $= +0.41^{+0.07}_{-0.10}$ for MCMC), in contradiction to the subsolar iron abundance inferred for the primary ([Fe/H] $= -0.120 \pm 0.020$; D. Sprague et al. 2022). This offset reflects the positive metallicity bias seen among the Diamondback and Elf Owl model fits of all of the L dwarfs in our sample.

3.4.2. Gliese 584C and SAND Models

The spectrum of the L8 Gliese 584C is well reproduced by the SAND models, which provide an intermediary between the condensate-infused Diamondback models (a close second-best fit with a large value of $f_{\rm sed}$) and the condensate-free Elf Owl models (which poorly fit). Both MCMC and RFR analyses produce best fits with similar χ_r^2 values. The MCMC fit is slightly superior ($\chi_r^2 = 4$ for MCMC versus $\chi_r^2 = 7$ for RFR) with only a modestly reduced 1.25 μm flux peak, while the RFR fit shows considerable variance around the three main flux peaks at 1.25, 1.6, and 2.1 μm, which inflates the final parameter uncertainties. Both analyses nevertheless yield equivalent $T_{\rm eff}$ and effectively solar metallicities, the latter consistent with iron abundance measurements of the G dwarf primaries ([Fe/H] $= -0.01 \pm 0.12$; R. Zhang et al. 2024). Both [M/H] and [α/H] are poorly constrained in these fits, and we find a strong negative correlation between these parameters, which may inflate their uncertainties. We also see a complex correlation between $\log g$ and [M/H] in the RFR fits that drives up the uncertainties for these parameters. There is a significant difference in the inferred surface gravities between the analyses, being higher and more consistent with the age of the system for the MCMC fits. In contrast, the inferred MCMC radius of $R = 0.064 \pm 0.001\ R_\odot$ is too small to be physical, while the slightly larger RFR radius of $R = 0.076^{+0.019}_{-0.017}\ R_\odot$ is also low but formally consistent with the expected model-dependent radius. The small radii inferred from these fits reflect the $\sim$200 K higher effective temperature inferred in this analysis as compared to the $T_{\rm eff} = 1237 \pm 24$ K model estimate from A. Sanghi et al. (2023), although prior spectral model fits (T. Nakajima et al. 2004; Z. Zhang et al. 2020; A. Sanghi et al. 2023) have found $T_{\rm eff}$ in the 1300–1500 K range. Our ability to accurately reproduce the spectrum with Gliese 584C with the SAND models indicates progress in modeling the complex L/T transition (A. J. Burgasser et al. 2002b; C. Helling 2005; P. Tremblin et al. 2016), but the large variances and uncertainties in temperature, surface gravity, and metallicity suggest that additional spectral data are needed to tightly constrain the physical properties.

3.4.3. Gliese 570D and Elf Owl Models

The spectrum of the T8 dwarf Gliese 570D is best fit by the Elf Owl models, with the Diamondback models being a close second and the SAND models well behind due to poor reproduction of the 1.6 μm CH_4 band. Again, both MCMC and RFR analyses yield comparable best fits ($\chi_r^2 = 5$ and 6, respectively), with the former showing $\sim$10% deviations at the 1.05 and 1.55 μm flux peaks while the latter shows slightly larger deviations at 1.05 and 1.25 μm. Temperatures and metallicities are well constrained and consistent between the MCMC and RFR analyses, and are also consistent with previously inferred values for the brown dwarf and the K4 dwarf primary (T. R. Geballe et al. 2001; A. J. Burgasser et al. 2006a; M. R. Line et al. 2015; Z. Zhang et al. 2021b; A. Lueber et al. 2023). The radii are also equal to within 10% and match model-based expectations. While all of the fit parameters are consistent within the RFR uncertainties, we do find a notable offset in surface gravity between the analysis methods, with the lower MCMC $\log g = 4.57^{+0.09}_{-0.07}$ significantly below the model-based estimate $\log g = 4.99 \pm 0.12$. This low surface gravity implies an evolutionary model age of $\tau = 450$–650 Myr (A. Burrows et al. 2001; I. Baraffe et al. 2003), which is significantly younger than the inferred age of the primary ($\tau = 3.3 \pm 1.9$ Gyr; Z. Zhang et al. 2021b), and reflects the general underestimation of surface gravities seen in Elf Owl model fits of T dwarfs (S. A. Beiler et al. 2024). This offset may reflect modeling errors in pressure-dependent chemistry or the strength of collision-induced H_2 absorption (J. L. Linsky 1969). The Elf Owl models do account for nonequilibrium chemistry through vertical mixing, which is quantified by the $\log \kappa_{zz}$ diffusion coefficient and found to be relatively weak in the MCMC fits ($\log \kappa_{zz} = 3.0^{+0.7}_{-0.5}$). We do note a slight positive correlation between $\log \kappa_{zz}$ and the C/O ratio, the latter being consistent with solar in both analyses. This trend is consistent with increased CO and decreased CH_4 abundances in a vertically mixed or oxygen-enriched atmosphere (D. Saumon et al. 2006; S. K. Leggett et al. 2007).

4. Discussion

The central focus of this study is a critical analysis of two different approaches to modeling low-temperature spectra using precomputed forward atmosphere models. Here we evaluate our outcomes through the lenses of accuracy, precision, and efficiency.

4.1. Accuracy: Quality of Spectral Fits

One of the key outcomes of this study is that the choice of model set is more important in reproducing an observed spectrum than the model fitting approach. For our L and T dwarf sample, both MCMC and RFR models were able to converge to a best-fit model that largely reproduced the overall SED and detailed features, with on average a lower value of χ^2 for the MCMC fits. Even cases with persistently high χ^2 values (e.g., HD 89744B, GJ 1048B, and G 204-39B) show a reasonable visual agreement, suggesting that these values are driven by the high S/N of the spectra or uncorrected calibration issues. Much greater variance in fit quality is seen between the model sets themselves, with typically one or two of the Diamondback, Elf Owl, and SAND models providing a clearly superior fit depending on the spectral type of the source, in some cases by a factor of 30 difference in χ_r^2 (e.g., G 196-3B). These differences reflect choices in the underlying gas chemistry (including nonequilibrium effects), treatments of condensation, line broadening profile (relevant for the 0.77 μm K I doublet), and adopted molecular opacities, as well as potentially unmodeled physical processes. As such, the methodology of fitting appears to be less critical for ensuring an accurate fit than evaluating a broad selection of models.

4.2. Accuracy: Parameter Values

Another measure of accuracy considered in this study is the agreement between the model fit parameters and the properties inferred from the primaries to our benchmark companions. Here, our results are mixed and likely driven by shortcomings in the models rather than the model fitting process. Our best MCMC and RFR fits based on the Diamondback and Elf Owl models consistently yield significantly supersolar metallicities, even for systems with solar or subsolar iron abundance primaries. This metallicity bias is absent for the late T dwarfs, and for G 196-3B and Gliese 584C, which are best fit by the SAND models. Our analysis suggests that metallicity may not be robustly measurable with these sets of models and/or with low-resolution near-infrared spectra, although the range of metallicities sampled (primary iron abundances $-0.43 \leqslant$ [Fe/H] $\leqslant +0.37$) is arguably narrow. The other benchmark metric, age or surface gravity, shows better agreement across the sample albeit with larger variance between the MCMC and RFR approaches. For the early L dwarfs and late T dwarfs, the MCMC method yields significantly higher and lower surface gravities, respectively, than the RFR method, the latter diverging from expected values. However, these differences may be driven in part by the relatively large uncertainties in the RFR $\log g$ values (0.3–0.4 dex, about 20% of the parameter range), while the disagreement between the fit and expected $\log g$ for the late T dwarfs can be ascribed to a known gravity bias in the Elf Owl models.

A separate accuracy assessment comes through comparison of fit parameter to evolutionary model expectations for $T_{\rm eff}$ and radius. Here we see significant departures for the early L dwarfs, with low temperatures and inflated radii based on the best-fit Diamondback models. The much better reproduction of $T_{\rm eff}$ and radius for the T dwarfs suggests this to be a condensation effect, although we cannot rule out biases in the evolutionary models caused by condensation as well. Overall, we find mixed outcomes in reproducing independent physical parameters, which indicate continued issues in the spectral models, particularly for the L dwarfs. The large variance between MCMC and RFR parameters does not indicate that one method is preferred over the other.

4.3. Precision: Parameter Uncertainties

The largest difference found between the MCMC and RFR approaches was the precision of the inferred parameters, considerably larger for the latter approach. This difference can be ascribed to the use of coarse-grid sampling in this study. The MCMC approach, based on linear interpolation, can in principle reach any parameter value assuming the interpolation is valid. The RFR approach, based on decision trees and ensemble learning, constructs predictive models from the provided model grid, which may lead to less accurate parameter estimates when the grid is not sampled at a sufficiently fine scale. A key measure of this is the RvP statistic R (Figure 4), which tracks predictive accuracy. While high self-consistency was found for $T_{\rm eff}$ ($R^2 > 0.95$ for all grids), predictability was very low for other parameters that were more coarsely sampled. Recent advancements, such as those outlined in A. Lueber et al. (2023), suggest that incorporating interpolation between grid points for a more refined parameter mapping could significantly improve the precision of RFR and reduce parameter uncertainties. In addition, while model grids are usually constructed through uniform sampling of the parameters (either linearly or log-linearly), C. Fisher & K. Heng (2022) showed that this method is suboptimal for parameter predictions, and that random or Latin hypercube sampling performs better for predicting parameters in higher-dimensional models. This study also emphasized that traditional retrievals with interpolation on a linear grid can produce biased posterior distributions, especially for parameters with nonlinear effects on the spectrum such as the C/O ratio, cloud properties, and metallicity. As such, the small parameter uncertainties from the MCMC approach may mask systematic errors and biases in the grid sampling or interpolation, while the large parameter uncertainties for the RFR approach represent a fundamental limit for coarse model grids. Furthermore, while the present study is based on high-S/N spectra, it is important to recognize that the effectiveness of ML methodologies like RFR can be adversely affected by noise, particularly in instances of low-S/N observations. The work conducted by I. Reis et al. (2019) offers a comprehensive analysis regarding the impact of various levels and types of injected noise in an RFR training set and the precision of the resultant RFR outcomes.

4.4. Efficiency: Computational Effort

The computational efficiency of a fitting method is an important factor for practical applications, particularly when dealing with large spectral data sets such as those provided by the Sloan Digital Sky Survey (D. G. York et al. 2000) or expected from Euclid (R. J. Laureijs et al. 2010) and SPHEREx (B. P. Crill et al. 2020). MCMC is a relatively inefficient method of sampling a posterior parameter space, and therefore requires significant computational resources. The major time sink for RFR, on the other hand, is in the training stage; the subsequent prediction-based model determines parameters relatively quickly. Our study found that MCMC fits take, on average, 30–50 times more computational time per source than RFR fitting (i.e., minutes versus seconds; Table 5). Therefore,

Table 5
Average Computational Time of MCMC Fitting and RFRs

	DBack (mm:ss)	EOwl (mm:ss)	SAND (mm:ss)
MCMC Fitting			
Total	04:31	22:31	02:11
RFRs			
Training	00:05	00:30	00:04
Testing	<00:01	00:02	<00:01
Total	~00:05	00:32	~00:04

RFR computations were performed on a MacBook Pro (2019) with a 2.4 GHz eight-core Intel Core i9 processor and 32 GB of random-access memory (RAM), while MCMC calculations were done on a MacBook Pro (2021) with a 3.2 GHz eight-core M1 Max processor with 64 GB of RAM. For large-scale or time-sensitive studies, the efficiency of RFR makes it a more practical approach, given the offsets in quality of fit and parameter precision. On the other hand, the better performance of MCMC makes it valuable for single fits requiring higher accuracy despite its higher computational cost.

4.5. Combining MCMC and RFR Approaches

Based on the comparative strengths and weaknesses of these two approaches, and the existing variance in the quality of fits among present-day atmosphere models for low-temperature spectra, we recommend a three-phase hybrid approach for fitting precomputed models to source spectra: first, a preparation phase in which data and/or models are interpolated onto a common wavelength scale and a set of RFR models are generated with each trained on a different atmosphere model set; second, an RFR fitting phase, with spectra fit to each of the model sets to identify the optimal set and initial parameters for each spectrum; third, an MCMC fitting phase in which the parameters are fine-tuned and precise estimates of uncertainty are inferred. This general framework allows for efficient evaluation of the optimal atmosphere model for a given source, or multiple models depending on error tolerance. It also has numerous avenues for optimization—for example, using subgrids of models tuned to the expected parameters of a sample, applying alternative supervised learning models or Bayesian fitting approaches, incorporating additional model calculations to improve parameter sensitivity or interpolation error, and iterative inclusion of additional fitting parameters such as element abundance variance (e.g., M. J. Veyette et al. 2016; S. A. Beiler et al. 2024) once an initial optimal parameter set is identified, among others. More importantly, this approach is agnostic to the source type or resolution of the data, making it applicable to exoplanet spectra (see C. Fisher & K. Heng 2018; A. Lueber et al. 2024b) and high-resolution data (see C. Fisher et al. 2020; N. P. Gibson et al. 2022) with appropriate modifications.

5. Summary

This study has assessed two approaches to fitting atmospheric models to low-temperature L and T dwarf spectra: MCMC and RFR. These approaches were applied using three sets of atmosphere models (Sonora Diamondback, Sonora Elf Owl, and SAND) and a sample of near-infrared spectra for 11 benchmark companions for which independent assessments of age and metallicity were available from their primaries and estimates of temperature and radius were available from evolutionary models. The goal of this study was to determine the relative trade-offs in these approaches in terms of accuracy, precision, and efficiency, and their future application to large spectral samples from current or future surveys. Our primary findings are as follows:

1. All 11 sources had a least one model set that provided a robust fit to near-infrared spectroscopy. In general, Diamondback provided the best fits for early L dwarfs, Elf Owl for mid and late T dwarfs, and SAND for late L and early T dwarfs and the young L dwarf G 196-3B. Absence of condensate opacity and errors in specific molecular gas opacities (e.g., CH_4 for SAND) were the general causes of poor fits.
2. The MCMC fits generally matched the observed spectra better than the RFR fits, and both fitting methods generally converged to the same best model set, with some individual exceptions (e.g., G 196-3B, HN Peg B).
3. MCMC also provided more accurate parameter estimates (smaller uncertainties) than RFR, although the actual parameters were generally consistent between the methods. In both cases, there was mixed agreement between the fit parameters and expected values based on the primaries (metallicity, surface gravity) or evolutionary models (radius, effective temperature).
4. The RFR approach is a far more efficient algorithm, with fits taking seconds as compared to the minutes of the MCMC approach. The RFR approach also identifies features of importance that align with temperature, surface gravity, and metallicity sensitive features, potentially useful for more directed parameter fitting schemes.
5. Based on this analysis, we propose a method for combining the strengths of these approaches to address the variances inherent in existing precomputed models. Specifically, we advocate a first-pass RFR fit to identify the optimal model set and initial parameters, followed by a more precise evaluation using MCMC fits when needed. This hybrid approach is flexible and can be adapted to different ML and Bayesian analysis approaches, as well as different data sets and source types.

Our combined approach joins a growing set of ML-assisted strategies that aim to significantly increase the efficiency of fitting precomputed models to stellar, brown dwarf, and exoplanet spectra. Continued advancements in these methodologies, coupled with improved interpolation techniques and more finely computed model grids, are needed to enhance our ability to accurately and precisely constrain the multiple physical parameters and processes that shape the spectra of low-temperature atmospheres.

Acknowledgments

The authors recognize and acknowledge the significant cultural role and reverence that the summit of Maunakea has with the indigenous Hawaiian community, and that the NASA Infrared Telescope Facility stands on Crown and Government Lands that the State of Hawai'i is obligated to protect and preserve for future generations of indigenous Hawaiians. Portions of this work were conducted at the University of

California, San Diego, which was built on the unceded territory of the Kumeyaay Nation, whose people continue to maintain their political sovereignty and cultural traditions as vital members of the San Diego community. A.L. acknowledges partial financial support from the Swiss National Science Foundation and the European Research Council (via a Consolidator Grant to Kevin Heng; grant No. 771620), as well as administrative support from the Center for Space and Habitability. This study would not have been possible without the open-source code HELA (P. Márquez-Neila et al. 2018), which can be found on the Exoclimes Simulation Platform: https://github.com/exoclime.

Facility: IRTF (SpeX).

Software: astropy (Astropy Collaboration et al. 2013, 2018), Matplotlib (J. D. Hunter 2007), NumPy (S. van der Walt et al. 2011), pandas (W. McKinney 2010), SciPy (P. Virtanen et al. 2020), SPLAT (A. J. Burgasser & SPLAT Development Team 2017), scikit.learn (F. Pedregosa et al. 2011).

Appendix

Here, we provide in Figures 11 through 16 the joint posterior parameter distributions for our best-fit MCMC and RFR analyses of three exemplary brown dwarfs in our sample: the early-type L dwarf LP 465-70B, the L/T transition object Gliese 584C, and the late T dwarf Gliese 570D (Section 3.4). Each figure shows the marginalized individual parameter distributions as histograms, and the joint distributions as 2D contour plots, with common colors assigned to each parameter. The optimal parameter values and estimated uncertainties are provided in Table 3 for MCMC and Table 4 for RFR.

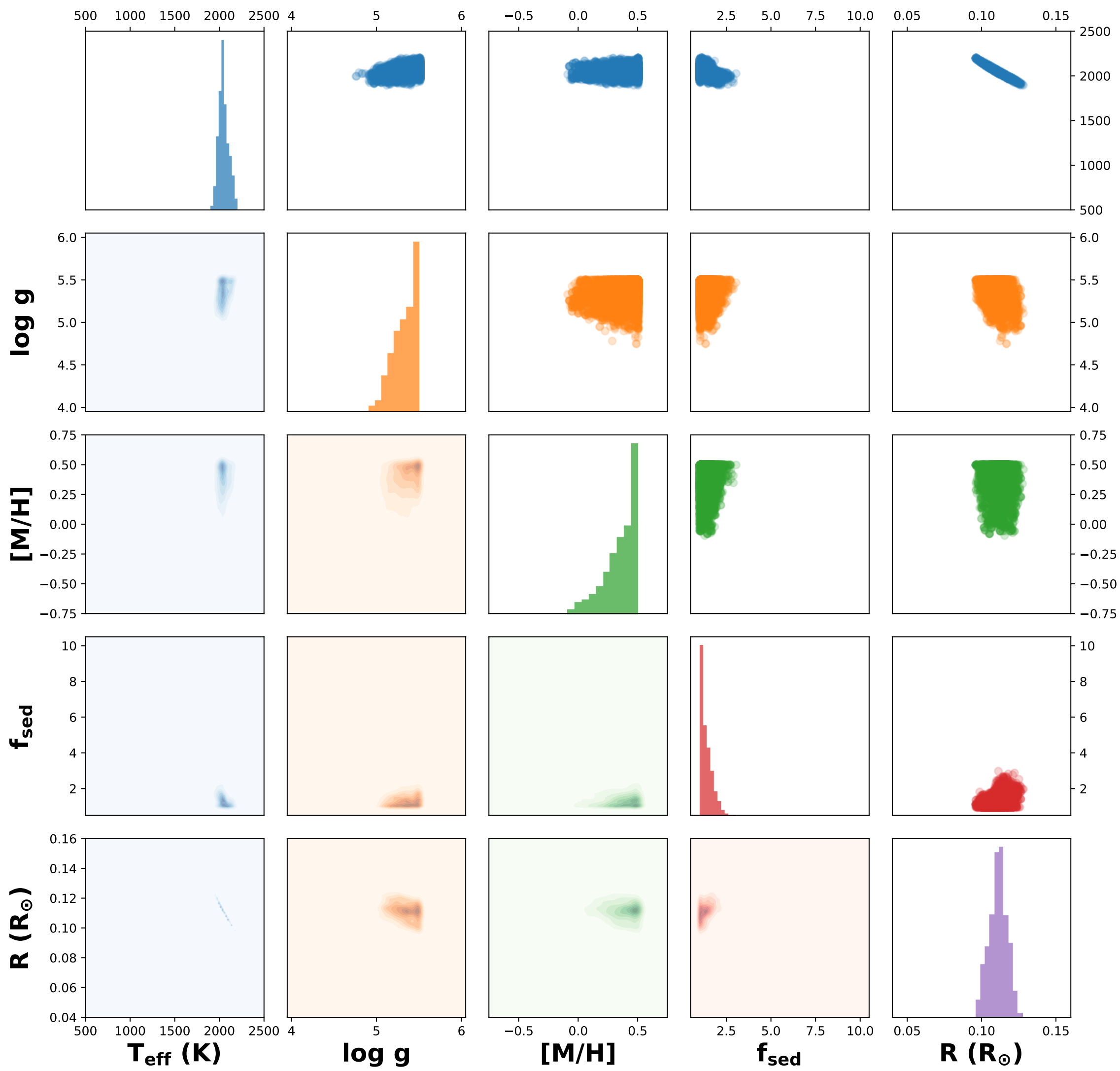

Figure 11. Posterior parameter distributions from our MCMC Diamondback fits of the near-infrared spectrum of the L0 dwarf LP 465-70B. The diagonal plots display the marginalized 1D parameter distributions for temperature $T_{\rm eff}$ (blue), surface gravity $\log g$ (orange), metallicity [M/H] (green), cloud sedimentation efficiency $f_{\rm sed}$ (red), and radius R (purple). The other panels display the joint parameter distributions: the lower left panels display contour plots of the probability, and the upper right panels display the individual parameters in the MCMC chain.

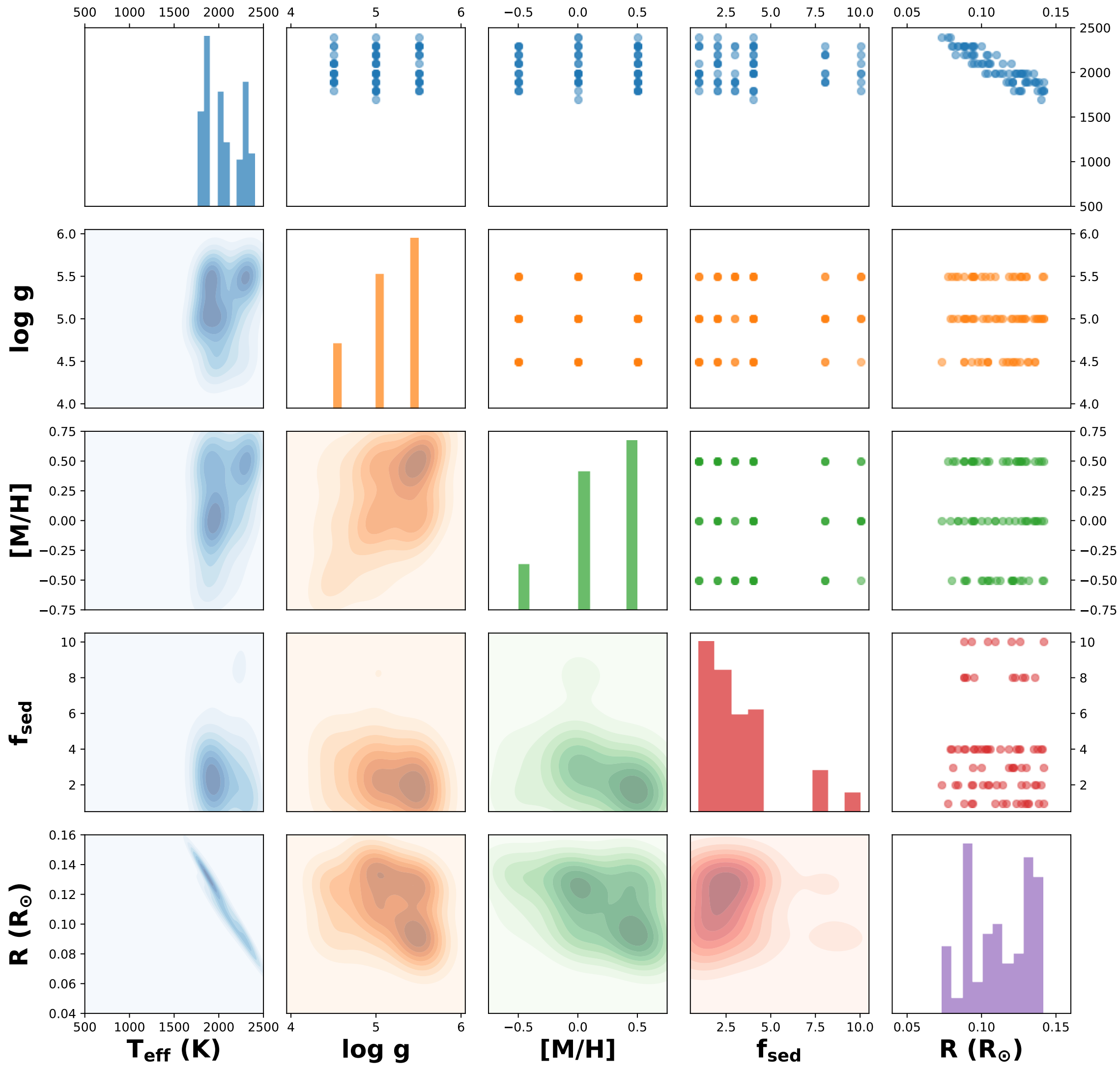

Figure 12. Posterior parameter distributions for our Diamondback RFR of the L0 dwarf LP 465-70B. Parameters and panels are as defined in Figure 11, with the upper right panels showing the predictions of individual trees within the RF ensemble.

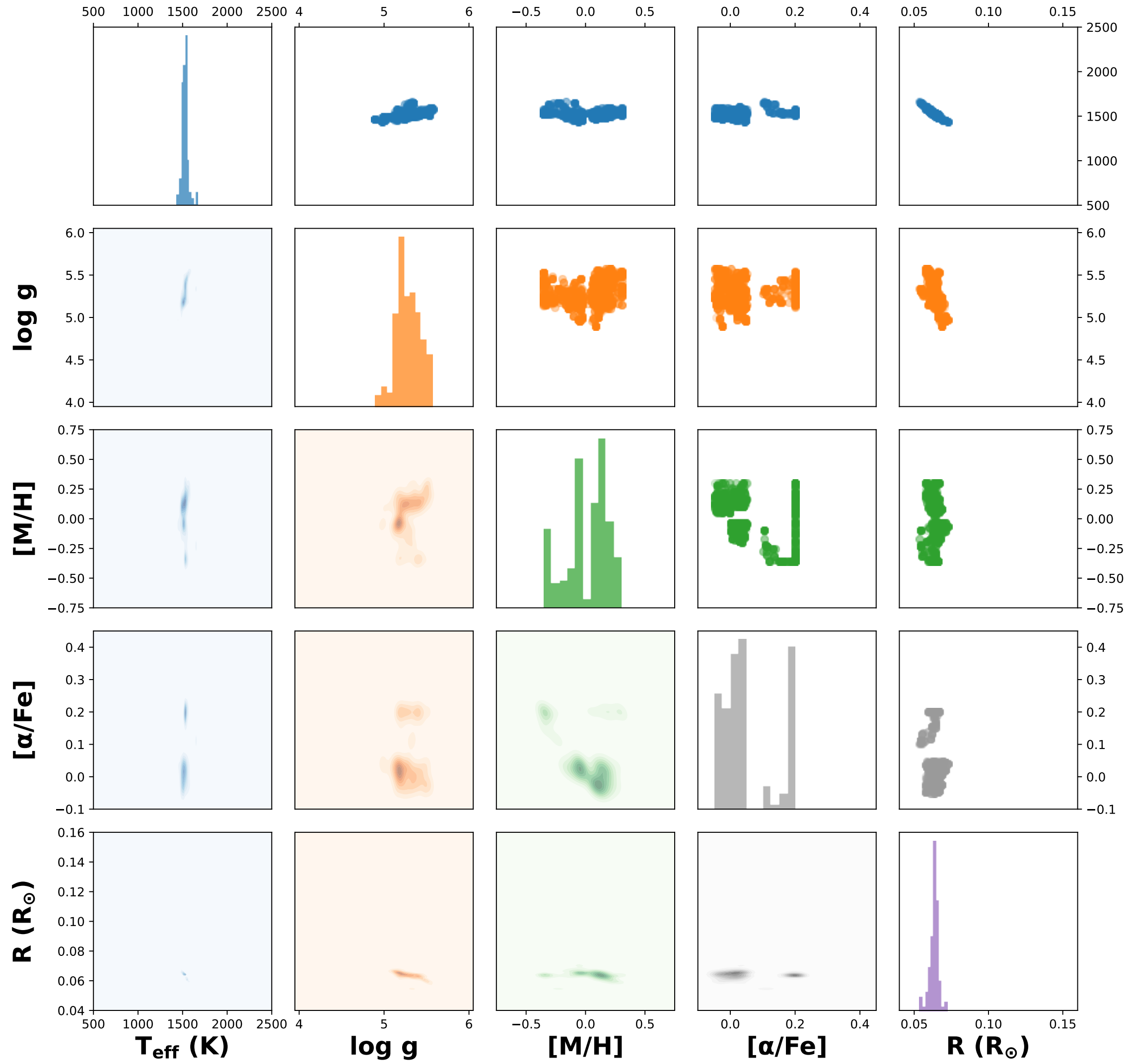

Figure 13. Same as Figure 11 but for MCMC SAND fits of the near-infrared spectrum of the L8 dwarf Gliese 584C. The parameters shown are the temperature $T_{\rm eff}$ (blue), surface gravity $\log g$ (orange), metallicity [M/H] (green), alpha element enrichment [α/Fe] (gray), and R (purple).

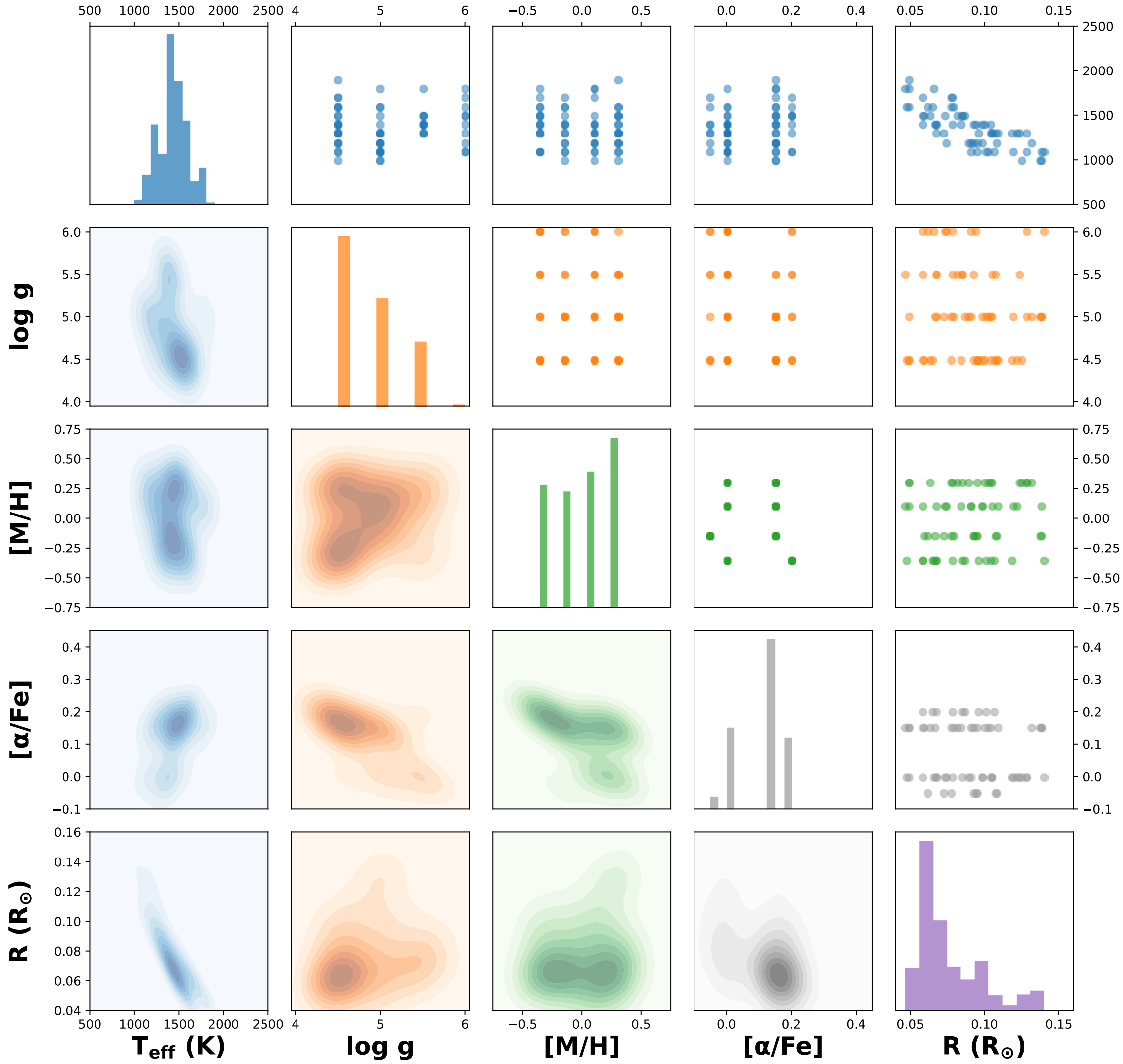

Figure 14. Same as Figure 13 but for the SAND RFR of the near-infrared spectrum of the L8 dwarf Gliese 584C.

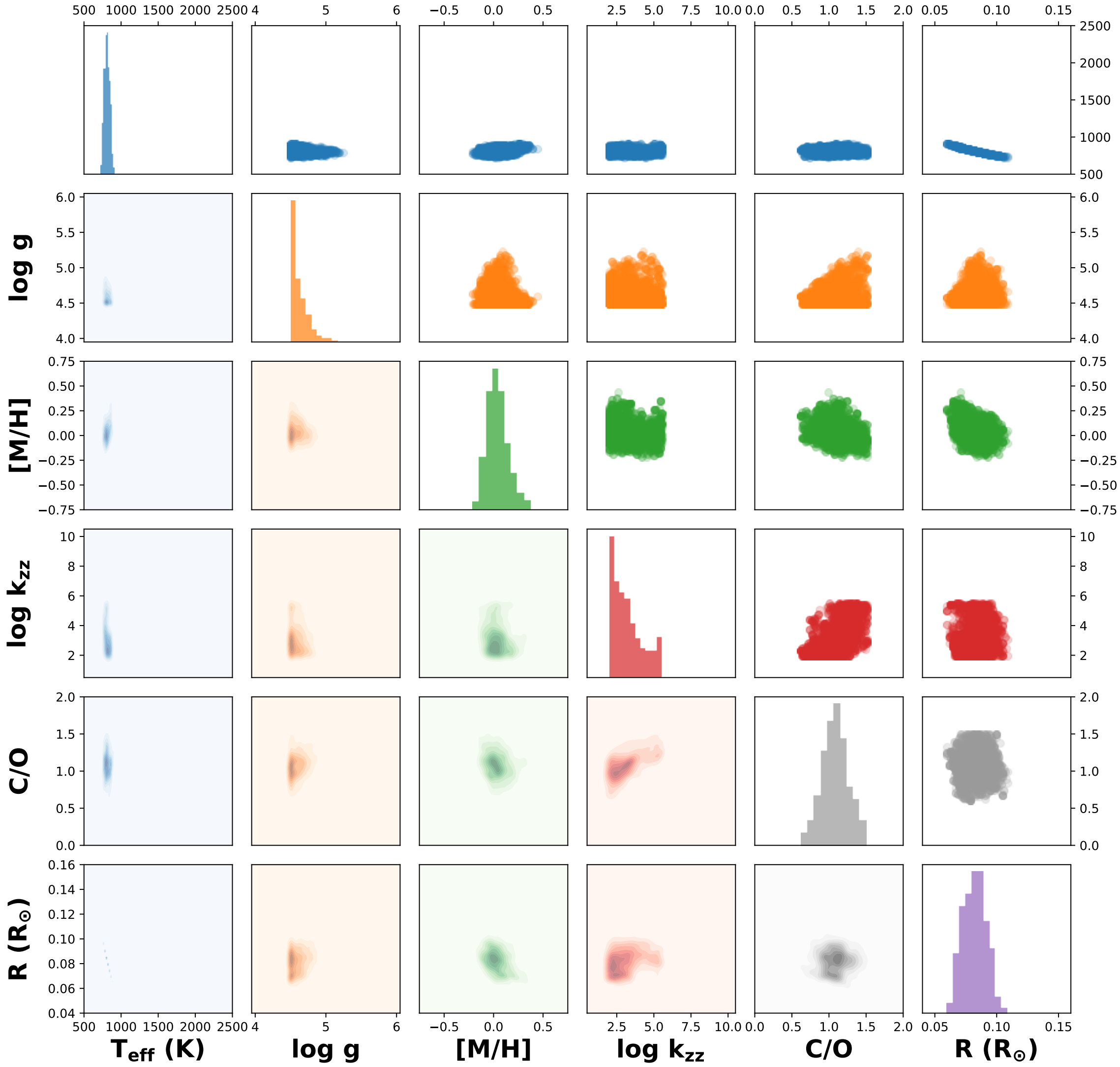

Figure 15. Same as Figure 11 but for MCMC Elf Owl fits of the near-infrared spectrum of the T8 dwarf Gliese 570D. The parameters shown are the temperature $T_{\rm eff}$ (blue), surface gravity $\log g$ (orange), metallicity [M/H] (green), vertical eddy diffusion coefficient $\log \kappa_{zz}$ (red), carbon-to-oxygen ratio C/O (gray), and radius R (purple).

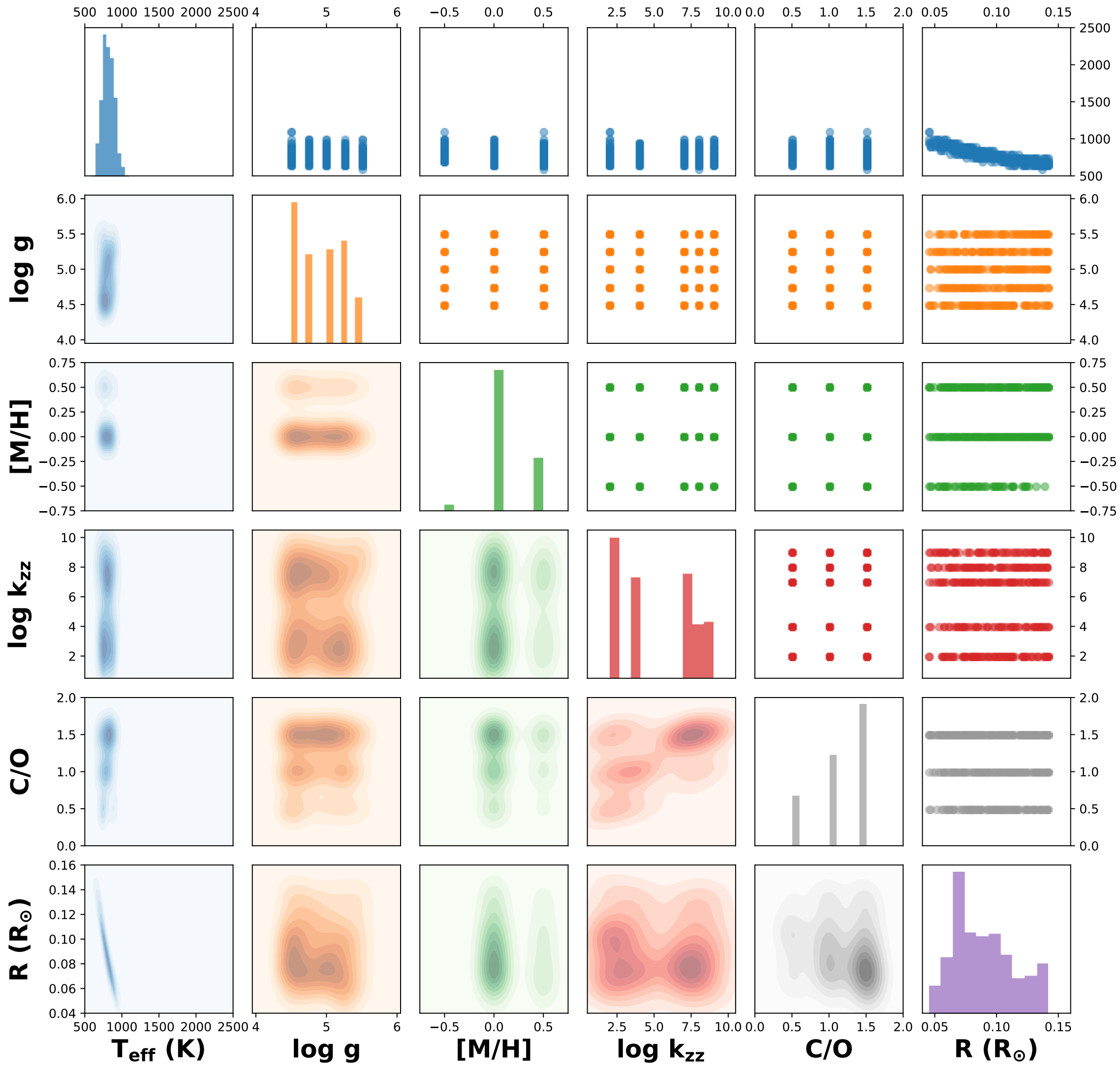

Figure 16. Same as Figure 15 but for the Elf Owl RFR of the near-infrared spectrum of the T8 dwarf Gliese 570D.

ORCID iDs

Anna Lueber https://orcid.org/0000-0001-6960-0256
Adam J. Burgasser https://orcid.org/0000-0002-6523-9536

References

Ackerman, A. S., & Marley, M. S. 2001, ApJ, 556, 872
Aganze, C., Burgasser, A. J., Faherty, J. K., et al. 2016, AJ, 151, 46
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., 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, F., Homeier, D., & Freytag, B. 2012, RSPTA, 370, 2765
Allard, N. F., Allard, F., & Kielkopf, J. F. 2005, A&A, 440, 1195
Alvarado, E., Gerasimov, R., Burgasser, A. J., et al. 2024, RNAAS, 8, 134
Ardévol Martínez, F., Min, M., Kamp, I., & Palmer, P. I. 2022, A&A, 662, A108
Astropy Collaboration, Price-Whelan, A. M., Sipöcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Baraffe, I., Chabrier, G., Barman, T. S., Allard, F., & Hauschildt, P. H. 2003, A&A, 402, 701
Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, A&A, 577, A42
Bardalez Gagliuffi, D. C., Burgasser, A. J., Gelino, C. R., et al. 2014, ApJ, 794, 143
Barnes, S. A. 2007, ApJ, 669, 1167
Beiler, S. A., Mukherjee, S., Cushing, M. C., et al. 2024, ApJ, 973, 60
Best, W. M. J., Liu, M. C., Magnier, E. A., & Dupuy, T. J. 2021, AJ, 161, 42
Bonfanti, A., Ortolani, S., & Nascimbeni, V. 2016, A&A, 585, A5
Breiman, L. 2001, MachL, 45, 5
Burgasser, A. J., Burrows, A., & Kirkpatrick, J. D. 2006a, ApJ, 639, 1095
Burgasser, A. J., Geballe, T. R., Leggett, S. K., Kirkpatrick, J. D., & Golimowski, D. A. 2006b, ApJ, 637, 1067
Burgasser, A. J., Gerasimov, R., Kremer, K., et al. 2024, ApJL, 971, L25
Burgasser, A. J., Kirkpatrick, J. D., Brown, M. E., et al. 2002a, ApJ, 564, 421
Burgasser, A. J., Marley, M. S., Ackerman, A. S., et al. 2002b, ApJL, 571, L151
Burgasser, A. J., Schneider, A. C., Meisner, A. M., et al. 2025, ApJ, 982, 79
Burgasser, A. J. & SPLAT Development Team 2017, in ASI Conf. Ser. 14, 3rd International Workshop on Spectral Stellar Libraries, ed. P. Coelho, L. Martins, & E. Griffin, 7
Burgasser, A. J., Witte, S., Helling, C., et al. 2009, ApJ, 697, 148
Burningham, B., Marley, M. S., Line, M. R., et al. 2017, MNRAS, 470, 1177
Burrows, A., Hubbard, W. B., Lunine, J. I., & Liebert, J. 2001, RvMP, 73, 719
Burrows, A., Sudarsky, D., & Hubeny, I. 2006, ApJ, 640, 1063
Charnay, B., Bézard, B., Baudino, J. L., et al. 2018, ApJ, 854, 172
Cobb, A. D., Himes, M. D., Soboczenski, F., et al. 2019, AJ, 158, 33

Cochran, W. G. 1977, Sampling Techniques (3rd ed.; New York: Wiley)
Crill, B. P., Werner, M., Akeson, R., et al. 2020, Proc. SPIE, 11443, 114430I
Cruz, K. L., Kirkpatrick, J. D., & Burgasser, A. J. 2009, AJ, 137, 3345
Cushing, M. C., Marley, M. S., Saumon, D., et al. 2008, ApJ, 678, 1372
Dupuy, T. J., & Liu, M. C. 2012, ApJS, 201, 19
Faherty, J. K., Burgasser, A. J., Walter, F. M., et al. 2012, ApJ, 752, 56
Faherty, J. K., Burgasser, A. J., West, A. A., et al. 2010, AJ, 139, 176
Fang, K.-T., Li, R., & Sudjianto, A. 2005, Design and Modeling for Computer Experiments (New York: Chapman and Hall/CRC)
Fegley, B. J., & Prinn, R. G. 1985, ApJ, 299, 1067
Fisher, C., & Heng, K. 2018, MNRAS, 481, 4698
Fisher, C., & Heng, K. 2022, ApJ, 934, 31
Fisher, C., Hoeijmakers, H. J., Kitzmann, D., et al. 2020, AJ, 159, 192
Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1
Galassi, M., Davies, J., Theiler, J., et al., 2011 GNU Scientific Library—Reference Manual, Version 1.15, https://www.gnu.org/software/gsl/doc/html/intro.html
Geballe, T. R., Saumon, D., Leggett, S. K., et al. 2001, ApJ, 556, 373
Gebhard, T. D., Angerhausen, D., Konrad, B. S., et al. 2024, A&A, 681, A3
Gerasimov, R., Burgasser, A. J., Caiazzo, I., et al. 2024, ApJ, 961, 139
Gerasimov, R., Burgasser, A. J., Homeier, D., et al. 2022, ApJ, 930, 24
Gerasimov, R., Homeier, D., Burgasser, A., & Bedin, L. R. 2020, RNAAS, 4, 214
Geweke, J. 1992, in Bayesian Statistics 4, ed. A. D. J. M. Bernardo, J. O. Berger, & A. Smith (Oxford: Oxford Univ. Press), 169
Gibson, N. P., Nugroho, S. K., Lothringer, J., Maguire, C., & Sing, D. K. 2022, MNRAS, 512, 4618
Gizis, J. E., Kirkpatrick, J. D., Burgasser, A., et al. 2001a, ApJL, 551, L163
Gizis, J. E., Kirkpatrick, J. D., & Wilson, J. C. 2001b, AJ, 121, 2185
Gomes, J. I., Pinfield, D. J., Marocco, F., et al. 2013, MNRAS, 431, 2745
Gray, R. O., Corbally, C. J., Garrison, R. F., et al. 2006, AJ, 132, 161
Gray, R. O., Napier, M. G., & Winkler, L. I. 2001, AJ, 121, 2148
Griffith, C. A., & Yelle, R. V. 1999, ApJL, 519, L85
Guzmán-Mesa, A., Kitzmann, D., Fisher, C., et al. 2020, AJ, 160, 15
Hastings, W. K. 1970, Bimka, 57, 97
Hauschildt, P. H., Baron, E., & Allard, F. 1997, ApJ, 483, 390
Hawley, S. L., Covey, K. R., Knapp, G. R., et al. 2002, AJ, 123, 3409
Helling, C. 2005, AN, 326, 627
Helling, C., Dehn, M., Woitke, P., & Hauschildt, P. H. 2008, ApJL, 675, L105
Helling, C., Oevermann, M., Lüttke, M. J. H., Klein, R., & Sedlmayr, E. 2001, A&A, 376, 194
Ho, T. K. 1998, ITPAM, 20, 832
Houk, N., & Swift, C. 1999, Michigan Spectral Survey, 5 (Ann Arbor, MI: Univ. Michigan), 0
Hunter, J. D. 2007, CSE, 9, 90
Irwin, P. G. J., Teanby, N. A., de Kok, R., et al. 2008, JQSRT, 109, 1136
Keenan, P. C., & McNeil, R. C. 1989, ApJS, 71, 245
Kirkpatrick, J. D., Reid, I. N., Liebert, J., et al. 2000, AJ, 120, 447
Kitzmann, D., Heng, K., Oreshenko, M., et al. 2020, ApJ, 890, 174
Kleijnen, J. P. 2015, in Int. Workshop on Simulation, ed. J. Pilz (Cham: Springer), 3
Knapp, G. R., Leggett, S. K., Fan, X., et al. 2004, AJ, 127, 3553
Laureijs, R. J., Duvet, L., Escudero Sanz, I., et al. 2010, Proc. SPIE, 7731, 77311H
Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599
Leggett, S. K., Burningham, B., Saumon, D., et al. 2010, ApJ, 710, 1627
Leggett, S. K., Saumon, D., Marley, M. S., et al. 2007, ApJ, 655, 1079
Leggett, S. K., Tremblin, P., Phillips, M. W., et al. 2021, ApJ, 918, 11
Line, M. R., Fortney, J. J., Marley, M. S., & Sorahana, S. 2014, ApJ, 793, 33
Line, M. R., Teske, J., Burningham, B., Fortney, J. J., & Marley, M. S. 2015, ApJ, 807, 183
Linsky, J. L. 1969, ApJ, 156, 989
Lodders, K., Palme, H., & Gail, H. P. 2009, LanB, 4B, 712
Lodieu, N., Zapatero Osorio, M. R., Martín, E. L., Rebolo López, R., & Gauza, B. 2022, A&A, 663, A84
Lueber, A., Heng, K., Bowler, B. P., et al. 2024a, A&A, 690, A357
Lueber, A., Kitzmann, D., Fisher, C. E., et al. 2023, ApJ, 954, 22
Lueber, A., Novais, A., Fisher, C., & Heng, K. 2024b, A&A, 687, A110
Luhman, K. L., Patten, B. M., Marengo, M., et al. 2007, ApJ, 654, 570
Madhusudhan, N. 2012, ApJ, 758, 36
Madhusudhan, N., & Seager, S. 2009, ApJ, 707, 24
Marfil, E., Tabernero, H. M., Montes, D., et al. 2021, A&A, 656, A162
Marley, M. S., & Robinson, T. D. 2015, ARA&A, 53, 279
Márquez-Neila, P., Fisher, C., Sznitman, R., & Heng, K. 2018, NatAs, 2, 719
McKay, C. P., Pollack, J. B., & Courtin, R. 1989, Icar, 80, 23
McKinney, W. 2010, in Proc. 9th Python in Science Conf., ed. S. van der Walt & J. Millman (Austin, TX: SciPy), 56
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, JChPh, 21, 1087
Morley, C. V., Mukherjee, S., Marley, M. S., et al. 2024, ApJ, 975, 59
Mukherjee, S., Batalha, N. E., Fortney, J. J., & Marley, M. S. 2023, ApJ, 942, 71
Mukherjee, S., Fortney, J. J., Morley, C. V., et al. 2024, ApJ, 963, 73
Nakajima, T., Tsuji, T., & Yanagisawa, K. 2004, ApJ, 607, 499
Oreshenko, M., Kitzmann, D., Márquez-Neila, P., et al. 2020, AJ, 159, 6
Patience, J., King, R. R., De Rosa, R. J., et al. 2012, A&A, 540, A85
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Rayner, J. T., Toomey, D. W., Onaka, P. M., et al. 2003, PASP, 115, 362
Reis, I., Baron, D., & Shahaf, S. 2019, AJ, 157, 16
Sanghi, A., Liu, M. C., Best, W. M. J., et al. 2023, ApJ, 959, 63
Santner, T. J., Williams, B. J., & Notz, W. I. 2003, The Design and Analysis of Computer Experiments, 1 (Berlin: Springer)
Saumon, D., & Marley, M. S. 2008, ApJ, 689, 1327
Saumon, D., Marley, M. S., Cushing, M. C., et al. 2006, ApJ, 647, 552
Schneider, A. C., Burgasser, A. J., Bruursema, J., et al. 2023a, ApJL, 943, L16
Schneider, A. C., Burgasser, A. J., Gerasimov, R., et al. 2020, ApJ, 898, 77
Schneider, A. C., Munn, J. A., Vrba, F. J., et al. 2023b, AJ, 166, 103
Shkolnik, E., Liu, M. C., & Reid, I. N. 2009, ApJ, 699, 649
Sisson, S. A., Fan, Y., & Beaumont, M. 2018, Handbook of Approximate Bayesian Computation (Boca Raton, FL: CRC Press)
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Sorahana, S., & Yamamura, I. 2014, ApJ, 793, 47
Sprague, D., Culhane, C., Kounkel, M., et al. 2022, AJ, 163, 152
Stephens, D. C., Leggett, S. K., Cushing, M. C., et al. 2009, ApJ, 702, 154
Stephenson, C. B. 1986, AJ, 92, 139
Struve, O., & Franklin, K. L. 1955, ApJ, 121, 337
Tremblin, P., Amundsen, D. S., Chabrier, G., et al. 2016, ApJL, 817, L19
Tremblin, P., Amundsen, D. S., Mourier, P., et al. 2015, ApJL, 804, L17
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
van Leeuwen, F. 2007, A&A, 474, 653
Vasist, M., Rozet, F., Absil, O., et al. 2023, A&A, 672, A147
Veyette, M. J., Muirhead, P. S., Mann, A. W., & Allard, F. 2016, ApJ, 828, 95
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Waldmann, I. P. 2016, ApJ, 820, 107
Xuan, J. W., Hsu, C.-C., Finnerty, L., et al. 2024, ApJ, 970, 71
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
Zakhozhay, O. V., Osorio, M. R. Z., Béjar, V. J. S., et al. 2023, A&A, 674, A66
Zhang, R., Liu, M. C., & Zhang, Z. 2024, ApJ, 960, 105
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., Liu, M. C., Hermes, J. J., et al. 2020, ApJ, 891, 171