



# Aurora: A Generalized Retrieval Framework for Exoplanetary Transmission Spectra

Luis Welbanks and Nikku Madhusudhan

Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK; [luis.welbanks@ast.cam.ac.uk](mailto:luis.welbanks@ast.cam.ac.uk), [nmadhu@ast.cam.ac.uk](mailto:nmadhu@ast.cam.ac.uk)

Received 2020 August 6; revised 2021 March 5; accepted 2021 March 11; published 2021 June 1

## Abstract

Atmospheric retrievals of exoplanetary transmission spectra provide important constraints on various properties, such as chemical abundances, cloud/haze properties, and characteristic temperatures, at the day–night atmospheric terminator. To date, most spectra have been observed for giant exoplanets due to which retrievals typically assume hydrogen-rich atmospheres. However, recent observations of mini Neptunes/super-Earths, and the promise of upcoming facilities including the James Webb Space Telescope (JWST), call for a new generation of retrievals that can address a wide range of atmospheric compositions and related complexities. Here we report Aurora, a next-generation atmospheric retrieval framework that builds upon state-of-the-art architectures and incorporates the following key advancements: (a) a generalized compositional retrieval allowing for H-rich and H-poor atmospheres, (b) a generalized prescription for inhomogeneous clouds/hazes, (c) multiple Bayesian inference algorithms for high-dimensional retrievals, (d) modular considerations for refraction, forward scattering, and Mie scattering, and (e) noise modeling functionalities. We demonstrate Aurora on current and/or synthetic observations of the hot Jupiter HD 209458 b, mini Neptune K2-18b, and rocky exoplanet TRAPPIST-1 d. Using current HD 209458 b spectra, we demonstrate the robustness of our framework and cloud/haze prescription against assumptions of H-rich/H-poor atmospheres, improving on previous treatments. Using real and synthetic spectra of K2-18b, we demonstrate an agnostic approach to confidently constrain its bulk atmospheric composition and obtain precise abundance estimates. For TRAPPIST-1 d, 10 JWST-NIRSpec transits can enable identification of the main atmospheric component for cloud-free, CO<sub>2</sub>-rich, and N<sub>2</sub>-rich atmospheres and abundance constraints on trace gases, including initial indications of O<sub>3</sub> if present at enhanced levels ( $\sim 10\times$ – $100\times$  Earth levels).

*Unified Astronomy Thesaurus concepts:* Exoplanet atmospheric composition (2021); Exoplanets (498); Hot Jupiters (753); Exoplanet atmospheres (487); Mini Neptunes (1063); Computational methods (1965); Super Earths (1655); Astronomy data modeling (1859); Transmission spectroscopy (2133)

## 1. Introduction

The last decade witnessed a revolution in our understanding of exoplanets and the nature of their atmospheres. Since the detection of the first atmosphere of a transiting exoplanet (Charbonneau et al. 2002), spectroscopic observations of exoplanets have provided wide-ranging insights into the compositions, temperature structures, and physical processes in their atmospheres (see, e.g., Seager & Deming 2010; Madhusudhan 2019; Zhang 2020, for a review). Most of the atmospheric observations have been made using transmission spectroscopy, which is conducted when an exoplanet transits in front of its host star and light from the star passes through the planet’s atmosphere before reaching the observer (Seager & Sasselov 2000). A transmission spectrum, through a wavelength-dependent change in the apparent size of the planet, encodes information about the atmosphere at the day–night terminator of the planet. Particularly, transmission spectroscopy has been key in detecting and quantifying the abundances of multiple molecules and atoms (e.g., Charbonneau et al. 2002; Deming et al. 2013; Kreidberg et al. 2014a; Madhusudhan et al. 2014; Wyttenbach et al. 2015; Sedaghati et al. 2017; Chen et al. 2018; Nikolov et al. 2018; Wakeford et al. 2018), as well as providing important insight into clouds/hazes in exoplanetary atmospheres (e.g., Pont et al. 2008; Sing et al. 2016; Nikolov et al. 2018; Benneke et al. 2019a).

Atmospheric spectra of exoplanets are routinely interpreted using retrieval methods. Introduced in Madhusudhan & Seager (2009), atmospheric retrievals of exoplanets aim to solve the inverse problem—to obtain statistical constraints on the atmospheric properties of an exoplanet from an observed spectrum. A retrieval code is composed of a parametric atmospheric model that computes a synthetic spectrum, coupled with an optimization algorithm that derives the model parameters given the observed spectrum. Although here we focus on retrieval codes for transmission spectroscopy, as discussed below, a plethora of retrieval codes have been developed for other applications (see e.g., Madhusudhan 2018, for a recent review). Retrieval codes have been developed for the analysis of thermal emission spectra of exoplanets (e.g., Madhusudhan & Seager 2009, 2011; Lee et al. 2012; Line et al. 2013, 2014b; Waldmann et al. 2015a; Gandhi & Madhusudhan 2018; Brogi & Line 2019; Gandhi et al. 2019), phase curves (e.g., Changeat & Al-Refaie 2020; Feng et al. 2020; Irwin et al. 2020), spectra of directly imaged exoplanets (e.g., Lee et al. 2013; Barstow et al. 2014; Lupu et al. 2016; Lavie et al. 2017; Nayak et al. 2017; Damiano & Hu 2020), as well as spectra of brown dwarfs (e.g., Line et al. 2014a, 2015; Burningham et al. 2017; Zalesky et al. 2019; Kitzmann et al. 2020; Piette & Madhusudhan 2020) and solar system planets (e.g., Rodgers 2000; Irwin et al. 2001, 2008, 2014; Irwin & Dyudina 2002).

Retrievals of transmission spectra have become ubiquitous in atmospheric characterization studies (see Madhusudhan 2018 for a review). The first retrieval code for exoplanetary atmospheres (Madhusudhan & Seager 2009) performed a



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

grid-based parameter exploration using a large model grid ( $\sim 10^7$  models of 10 parameters each). Subsequent studies adopted more robust statistical optimization algorithms. The next iteration of retrieval codes used Markov Chain Monte Carlo (MCMC) methods (e.g., Tegmark et al. 2004; Foreman-Mackey et al. 2013), providing a better parameter exploration of the parameter space but with limitations in calculating the model evidence for model comparison. Retrieval codes utilizing MCMC methods include Madhusudhan & Seager (2011), Benneke & Seager (2012), Madhusudhan et al. (2014), CHIMERA (e.g., Line et al. 2013; Kreidberg et al. 2014a), MassSpec (de Wit & Seager 2013), ATMO (e.g., Evans et al. 2017; Wakeford et al. 2017), BART (e.g., Blecic 2016; Cubillos 2016), PLATON (Zhang et al. 2019), and METIS (Lacy & Burrows 2020). Concurrently, the retrieval code NEMESIS (Irwin et al. 2008) developed for solar system studies using gradient-descent optimization methods, such as Optimal Estimation (OE), has also seen applications to exoplanetary transmission spectra (e.g., Barstow et al. 2017).

The next generation of retrieval codes came to light with the implementation of the nested sampling algorithm (e.g., Skilling 2006), facilitating more efficient parameter space exploration and calculation of model evidence. Transmission retrieval codes like SCARLET (e.g., Benneke & Seager 2013; Benneke et al. 2019a, 2019b),  $\mathcal{T}$ -REx (e.g., Waldmann et al. 2015b), POSEIDON (e.g., MacDonald & Madhusudhan 2017), AURA (e.g., Pinhas et al. 2018), and petitRADTRANS (Mollière et al. 2019), among others (e.g., Fisher & Heng 2018, 2019; Brogi & Line 2019; Min et al. 2020; Seidel et al. 2020), adopted the MultiNest nested sampling algorithm (Feroz et al. 2009). Although MultiNest has been extensively used, other nested sampling algorithms have been implemented like Nestle (Barbary 2105) in PLATON, Dynesty (Speagle 2020) in PLATON II (Zhang et al. 2020), and PolyChord (Handley et al. 2015a) in  $\mathcal{T}$ -REx III (Al-Refaie et al. 2019).

The extensive availability of computational methods and packages for statistical inference has made it possible for retrieval codes to update their capabilities and include multiple optimization algorithms. For instance, CHIMERA has used OE, Bootstrap Monte Carlo (BMC), MCMC, as well as MultiNest nested sampling (e.g., Line et al. 2013; Colón et al. 2020). NEMESIS has been adapted, beyond OE, to use MultiNest nested sampling (e.g., Krissansen-Totton et al. 2018). Similarly,  $\mathcal{T}$ -REx, through different updates, has used MCMC and diverse nested sampling algorithms (e.g., Waldmann et al. 2015a; Al-Refaie et al. 2019).

Parallel efforts are being made toward exploring the viability of machine-learning algorithms as a replacement for or aid to traditional Bayesian optimization algorithms. Some studies have used the Random Forest algorithm (e.g., Breiman et al. 1984) to train estimators and predict the parameters that better explain an observed spectrum (e.g., Márquez-Neila et al. 2018; Fisher et al. 2020; Guzmán-Mesa et al. 2020; Nixon & Madhusudhan 2020). Other studies have used Deep Belief Neural Networks (albeit in studies of emission spectroscopy, e.g., Waldmann 2016), Generative Adversarial Networks (Zingales & Waldmann 2018), Deep Neural Networks (Soboczenski et al. 2018), and Bayesian Neural Networks (Cobb et al. 2019) in an effort to predict atmospheric properties of exoplanets. A complementary approach has been to use machine learning to help inform the priors in a traditional retrieval (e.g., Hayes et al. 2020). These advancements in retrievals are an active area of research, and future work may

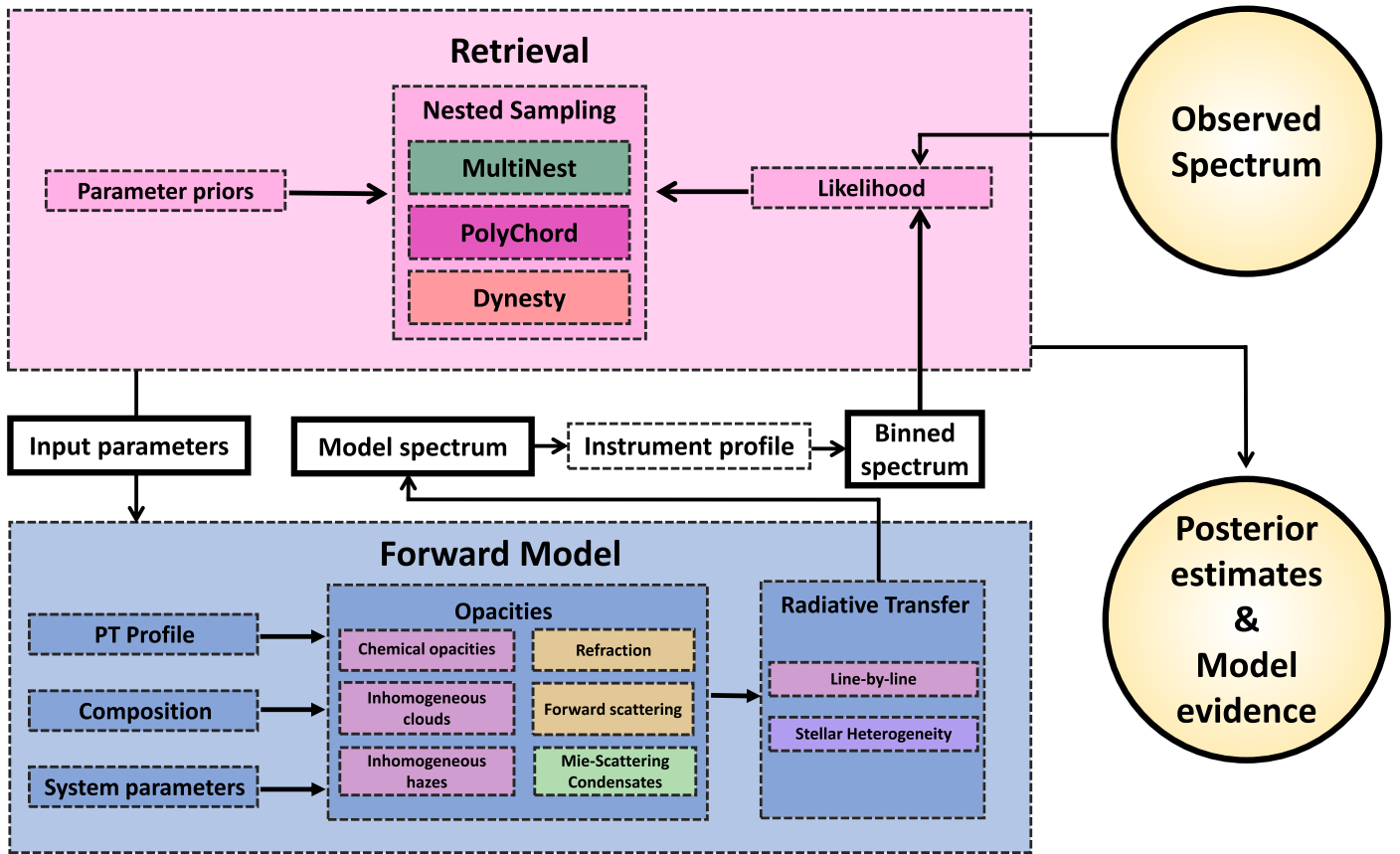
elucidate on the synergies between traditional retrievals and these novel machine-learning techniques.

There have also been developments in model considerations for atmospheric retrievals of transmission spectra. Recent works have investigated the impact of cloud and hazes in atmospheric retrievals (e.g., Line & Parmentier 2016; MacDonald & Madhusudhan 2017; Pinhas et al. 2018; Mai & Line 2019; Barstow 2020). Similarly, studies have investigated the relative importance of various model and data considerations, including temperature structures, clouds, and optical data (e.g., Welbanks & Madhusudhan 2019) over simpler isobaric, isothermal, semianalytic model assumptions. Other works have looked into the effect of uncertainties in the system parameters (e.g., de Wit & Seager 2013; Fisher & Heng 2018; Batalha et al. 2019; Changeat et al. 2020) or temperature and abundance inhomogeneities (Caldas et al. 2019; Changeat et al. 2019; MacDonald et al. 2020) in retrieved properties of the atmosphere. Further efforts have investigated the impact of stellar contamination in transmission spectra (e.g., Pinhas et al. 2018; Iyer & Line 2020; Bruno et al. 2020).

While the studies above have focused primarily on retrievals for giant planets with H-rich atmospheres, some studies have also developed retrieval frameworks for smaller planets where the atmosphere may not be assumed to be H rich a priori. Benneke & Seager (2012) investigated an agnostic retrieval framework for super-Earths, which could have a wide range of atmospheric compositions. They highlight that assuming log-uniform priors for the mixing ratios of the chemical species sampled in a retrieval can lead to a highly asymmetric prior for the last species derived using the unit-sum constraint, which is unfavorable in the absence of a priori knowledge of the dominant species in the atmosphere, e.g., for super-Earths. To overcome this limitation, Benneke & Seager (2012) suggest a reparameterization of the chemical compositions applicable to both H-rich and non-H-rich atmospheres. The parameterization, based on centered-log-ratio transformations (e.g., Aitchison 1986), allows for equal prior probability distributions for all chemical species considered; we discuss this in depth in Section 2.2. In a subsequent work, Benneke & Seager (2013) demonstrate the potential of using Bayesian model comparisons along with high-precision transmission spectra of super-Earths/mini Neptunes to differentiate between cloudy H-rich atmospheres and those of high mean molecular weight, e.g., H<sub>2</sub>O rich.

After this decade of revolutionary work on retrievals, the next generation of retrieval codes is upon us. Such retrievals must incorporate the lessons learned from atmospheric studies of giant exoplanets and also be adaptable to low-mass planets. In preparation for upcoming observations of temperate mini Neptunes and super-Earths, the methods for non-H-dominated atmospheres must be implemented and updated to be compatible with the latest optimization algorithms. Upcoming codes should be able to expand their modeling functionalities motivated by data requirements. Lastly, with the increasing model complexity and data quality, new retrieval codes must be prepared for assessing multidimensional, highly degenerate problems.

We introduce Aurora, a next-generation retrieval framework for the atmospheric characterization of H-rich and non-H-rich planets. Our code incorporates new key features on the previous retrieval code AURA (Pinhas et al. 2018). First, we



**Figure 1.** Schematic of Aurora’s retrieval framework. The retrieval framework combines an atmospheric forward model with a Bayesian inference and parameter estimation algorithm to derive the model evidence and posterior probability distributions of the model parameters given an observed spectrum. Aurora also has modular capabilities for including additional effects in the atmospheric model if required, e.g., light refraction, forward scattering, and Mie scattering.

reparameterize the volume mixing ratios in the atmosphere expanding the previous scope beyond H-rich atmospheres, adapting methods previously used for super-Earths (e.g., Benneke & Seager 2012) and other areas of compositional data analysis (e.g., Aitchison 1986; Aitchison & Egozcue 2005; Pawlowsky-Glahn & Buccianti 2011). Second, Aurora incorporates the next-generation nested sampling algorithms PolyChord and Dynesty, as well as maintaining compatibility with MultiNest. Third, Aurora includes a new generalized parametric treatment for inhomogeneous clouds and hazes. Compared to previous prescriptions, our new treatment of clouds/hazes is robust against assumptions of whether the atmosphere is H rich or not.

Lastly, Aurora incorporates different modular capabilities that enhance the study of transmission spectra using retrievals and forward models. These include assessing stellar heterogeneity (e.g., Rackham et al. 2017; Pinhas et al. 2018), allowing for underestimated variances in the data (e.g., Foreman-Mackey et al. 2013; Colón et al. 2020), and considering correlated noise using Gaussian processes (e.g., Rasmussen & Williams 2006). Additionally, our forward-modeling capabilities can account for light refraction and forward scattering (Robinson et al. 2017), as well as Mie scattering due to a variety of condensate species (Pinhas & Madhusudhan 2017). Aurora’s modular capabilities can be incorporated in retrievals should observations require it.

In what follows, we present our retrieval framework in Section 2. We benchmark the results of Aurora on current and synthetic observations in Section 3, and present case studies for

characterizing atmospheres of hot Jupiters, mini Neptunes, and rocky exoplanets with the James Webb Space Telescope (JWST). We summarize our conclusions in Section 4 and discuss the implications of our findings and possible avenues for future developments of Aurora.

## 2. Aurora Retrieval Framework

Aurora builds upon the AURA retrieval framework (Pinhas et al. 2018) developed by our group and, among other features, expands the retrieval capabilities to rocky exoplanets without the assumption of H-rich atmospheres. The core retrieval methodology for H-rich atmospheres is explained in Pinhas et al. (2018), with its implementation previously explained in Welbanks & Madhusudhan (2019) and employed in different retrieval studies (e.g., Chen et al. 2018; von Essen et al. 2019; Welbanks et al. 2019; Colón et al. 2020). Here, we reintroduce the basic retrieval methodology of the AURA code and discuss the new enhancements made in Aurora. Figure 1 shows the schematic diagram of the Aurora framework.

Like any contemporary retrieval framework (Madhusudhan 2018), AURA and Aurora comprise a forward model that is interfaced with a Bayesian inference and parameter estimation scheme. The forward model computes a transmission spectrum given a set of parameters for the temperature structure, chemical composition, and presence of clouds/hazes on the planet’s atmosphere. The parameter estimation scheme explores the model’s parameter space in search of regions of high likelihood that can explain a set of observations. The Bayesian inference

scheme estimates the model evidence and posterior probability distributions of the model parameters and is performed using the nested sampling algorithm. In what follows, we describe each of these components for our retrieval framework. We highlight the following key advancements introduced in Aurora:

1. Generalized inhomogeneous cloud/haze parameterization
2. Generalized considerations for H-poor/H-rich compositions
3. Adaptable Bayesian inference algorithms
4. Modular functionalities for considering:
  - (a) Refraction and forward scattering
  - (b) Mie scattering with a library of condensates
  - (c) Error inflation and Gaussian processes to treat correlated noise.

We first discuss the standard features that we retain from the retrieval framework of Pinhas et al. (2018), followed by a description of the new features in the Aurora framework built in this work.

### 2.1. Forward Model

Aurora computes the transmission spectrum of an exoplanet in transit assuming plane-parallel geometry. Our forward model is comprised of a parametric pressure–temperature (P-T) profile; parametric chemical abundances and consideration for multiple sources of opacity including atomic and molecular-line opacity, Rayleigh scattering, and collision-induced absorption (CIA); a treatment for inhomogeneous clouds and hazes; and a line-by-line radiative transfer solver under hydrostatic equilibrium. The forward model can consider light refraction, forward scattering, Mie scattering, and stellar heterogeneity (see Section 2.4).

#### 2.1.1. Pressure–Temperature Profile

The temperature in the atmosphere of an exoplanet as a function of pressure is determined by the P-T profile. We follow the parametric prescription of Madhusudhan & Seager (2009). We choose this profile as it is motivated by the profiles observed in the solar system and has been successfully applied to exoplanet studies (e.g., Madhusudhan & Seager 2009, 2011; Madhusudhan et al. 2014; Blečić et al. 2017). The equations for temperature in this parameterization divide the model atmosphere into three distinct regions:

$$T = T_0 + \left( \frac{\log(P/P_0)}{\alpha_1} \right)^2, \quad P_0 < P < P_1, \quad (1)$$

$$T = T_2 + \left( \frac{\log(P/P_2)}{\alpha_2} \right)^2, \quad P_1 < P < P_3, \quad (2)$$

$$T = T_2 + \left( \frac{\log(P_3/P_2)}{\alpha_2} \right)^2, \quad P > P_3, \quad (3)$$

where we maintain the empirical choice of Madhusudhan & Seager (2009) to set their parameters  $\beta_1 = \beta_2 = 0.5$ . Here,  $T_0$  is the temperature at the top of the model atmosphere  $P_0$  (e.g.,  $10^{-6}$  bar in this work),  $P_1$  is the boundary between the first and second regions,  $P_3$  is the boundary between the second and third regions,  $P_2$  is the pressure in the parameterization that can capture possible thermal inversions if  $P_2 > P_1$ , and  $\alpha_1$  and  $\alpha_2$  are the values that determine the curvature of the profile in the different layers. We restrict our temperature profiles to those

with  $P_2 \leq P_1$  for observations of the day–night terminator where thermal inversions are not expected. Aurora has the option of considering an isothermal profile in which case the free parameter is  $T_0$ . Then, the temperature at all points in the model atmosphere is assumed to be  $T_0$ .

#### 2.1.2. Sources of Opacity

The presence of different chemical species in the atmosphere of an exoplanet is retrieved by considering their contribution to the starlight’s extinction. The extinction coefficient  $\kappa(\lambda, P, T)$  of the atmosphere is a pressure-, temperature-, and wavelength-dependent quantity that contributes to the differential optical depth,

$$d\tau(\lambda, P, T) = \kappa(\lambda, P, T)ds, \quad (4)$$

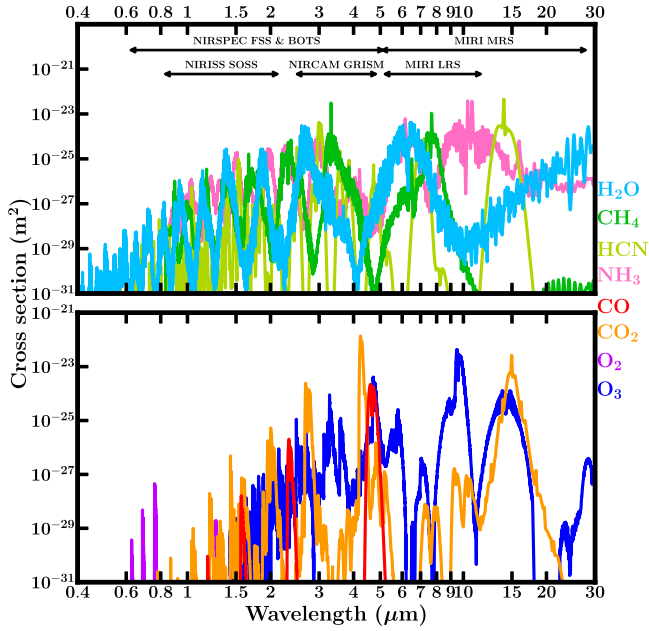
along the line of sight  $s$ . The extinction coefficient for each species is given by  $\kappa_i(\lambda, P, T) = n_i \sigma_i(\lambda, P, T)$ , where  $n_i$  is the number density and  $\sigma_i$  the absorption cross section of the species. The number density,  $n_i$ , is parameterized through the volume mixing ratio,  $X_i = n_i/n_{\text{tot}}$ , where  $n_{\text{tot}}$  is the total number density. The volume mixing ratio of each species is a free parameter and assumed to be uniform in the atmosphere. For H-rich atmospheres, Aurora calculates the volume mixing ratio of H<sub>2</sub> and He by assuming a particular He/H<sub>2</sub> ratio ( $X_{\text{He}}/X_{\text{H}_2}$ ) and the following relations:

$$X_{\text{H}_2} = \frac{1 - \sum_{i,i \neq \text{He}, \text{H}_2} X_i}{1 + \frac{X_{\text{He}}}{X_{\text{H}_2}}}, \quad X_{\text{He}} = X_{\text{H}_2} \frac{X_{\text{He}}}{X_{\text{H}_2}}, \quad (5)$$

where we adopt a solar value of  $X_{\text{He}}/X_{\text{H}_2} = 0.17$  (Asplund et al. 2009) and consider a total of  $n$  species in the model atmosphere. The treatment of the volume mixing ratios when a H-rich atmosphere is not assumed a priori is described in Section 2.2.

Aurora in general considers the opacity sources expected in the atmospheres of hot Jupiters, mini Neptunes, and temperate rocky planets (e.g., Madhusudhan 2012, 2019; Moses et al. 2013). The opacity sources considered in this work are H<sub>2</sub>–H<sub>2</sub> and H<sub>2</sub>–He CIA (Richard et al. 2012) and line opacity due to CH<sub>4</sub> (Yurchenko & Tennyson 2014), CO (Rothman et al. 2010), CO<sub>2</sub> (Rothman et al. 2010), H<sub>2</sub>O (Rothman et al. 2010), HCN (Barber et al. 2014), K (Allard et al. 2016), Na (Allard et al. 2019), N<sub>2</sub> (Rothman et al. 2010), NH<sub>3</sub> (Yurchenko et al. 2011), O<sub>2</sub> (Rothman et al. 2010), and O<sub>3</sub> (Rothman et al. 2010). The opacities for the chemical species are computed following the methods of Gandhi & Madhusudhan (2017), with the updated values of Gandhi & Madhusudhan (2018), and with H<sub>2</sub>-broadened Na and K cross sections as explained in Welbanks et al. (2019).

Aurora also incorporates a continually updated library of cross sections of various other atomic and molecular species (Gandhi et al. 2020). Figure 2 shows the cross section for most of the molecular opacity sources considered in this work for a pressure of 1 bar and a temperature of 300 K, from 0.4 to 30  $\mu\text{m}$ , covering the wavelength range expected to be observable by JWST. The Na and K profiles can be seen in Figure 1 of Welbanks et al. (2019).



**Figure 2.** Top and bottom panels show the absorption cross sections for some of the molecular opacity sources included in this work at a pressure of 1 bar and a temperature of 300 K. The cross sections are shown in the wavelength region where JWST is expected to perform observations. JWST instrument ranges are indicated on the top panel using black lines and arrows.

The resulting extinction coefficient is

$$\kappa(\lambda, P, T) = \sum_i X_i n_{\text{tot}}(P, T) \sigma_i(\lambda, P, T) + X_{\text{H}_2} n_{\text{tot}}^2(P, T) \times [X_{\text{H}_2} \sigma_{\text{H}_2-\text{H}_2}(\lambda, T) + X_{\text{He}} \sigma_{\text{H}_2-\text{He}}(\lambda, T)], \quad (6)$$

where  $\sigma_{\text{H}_2-\text{H}_2}$  and  $\sigma_{\text{H}_2-\text{He}}$  are the  $\text{H}_2$ - $\text{H}_2$  and  $\text{H}_2$ - $\text{He}$  CIA cross sections. The extinction coefficient can be amended to remove  $\text{H}_2$ - $\text{He}$  and  $\text{H}_2$ - $\text{H}_2$  CIA and/or include CIA due to other species. Furthermore, the total extinction coefficient can include  $\text{H}_2$  Rayleigh scattering:

$$\kappa_{\text{H}_2\text{-Rayleigh}}(\lambda, P, T) = X_{\text{H}_2} n_{\text{tot}}(P, T) \sigma_{\text{H}_2\text{ scat}}(\lambda), \quad (7)$$

where the wavelength-dependent cross section in cgs is given analytically by Dalgarno & Williams (1962) as

$$\sigma_{\text{H}_2\text{ scat}}(\lambda) = \frac{8.14 \times 10^{-45}}{\lambda^4} + \frac{1.28 \times 10^{-54}}{\lambda^6} + \frac{1.61 \times 10^{-64}}{\lambda^8} + \mathcal{O}(\lambda^{-10}) \quad (8)$$

and is incorporated up to its third term in Aurora.

Aurora also includes opacity sources relevant for modeling H-poor atmospheres of rocky planets. This library contains CIA cross sections of  $\text{CO}_2$ - $\text{CO}_2$ ,  $\text{N}_2$ - $\text{N}_2$ ,  $\text{O}_2$ - $\text{O}_2$ ,  $\text{O}_2$ - $\text{CO}_2$ ,  $\text{O}_2$ - $\text{N}_2$ , among others obtained from HITRAN (Karman et al. 2019). These additional CIA cross sections are generated within the temperature and wavelength limits available in the HITRAN data. The cross sections for temperatures beyond those limits are set to values at the boundaries. We assume no opacity for wavelengths beyond the database range, as these values are not known. Future efforts, both experimental and theoretical, on extending and revising opacity databases would help obtain cross sections over the full range of wavelengths and temperatures applicable for such planets. Aurora can also

include Rayleigh scattering due to a variety of species including  $\text{O}_2$ ,  $\text{N}_2$ , Ar, Ne,  $\text{CO}_2$ ,  $\text{CH}_4$ ,  $\text{H}_2\text{O}$ , CO, and  $\text{N}_2\text{O}$  (Rao et al. 1977; Snee & Ubachs 2005; Thalman et al. 2014). Rayleigh scattering due to species  $i$  is  $\kappa_{i\text{-Rayleigh}}(\lambda, P, T) = X_i n_{\text{tot}}(P, T) \sigma_{i\text{ scat}}(\lambda)$ . We include CIA due to  $\text{CO}_2$ - $\text{CO}_2$ ,  $\text{N}_2$ - $\text{N}_2$ , as well as Rayleigh scattering due to  $\text{N}_2$ ,  $\text{H}_2\text{O}$ , and  $\text{CO}_2$  in the H-poor models presented in Section 3.4.

### 2.1.3. A New Cloud and Haze Parameterization

We introduce a new cloud and haze treatment for inhomogeneous cover that considers a total of four distinct spatial areas (sectors) covering the planet. These four areas are (1) a clear, cloud-free and haze-free, area affected only by Rayleigh scattering, (2) an area covered by hazes only, (3) an area covered by a gray cloud deck with Rayleigh scattering above the cloud deck, and (4) an area covered by a gray cloud deck and hazes above it. The total transit depth is a linear superposition of the transit depths of each sector:

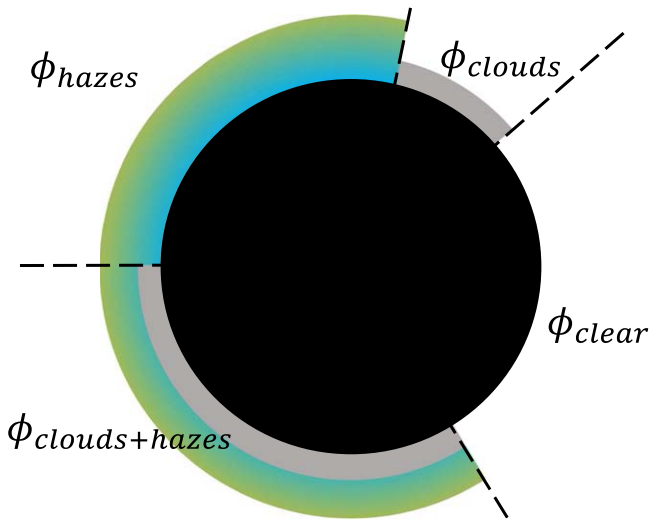
$$\Delta_{\text{planet}}(\lambda) = \phi_{\text{hazes}} \Delta_{\text{hazes}}(\lambda) + \phi_{\text{clouds}} \Delta_{\text{clouds}}(\lambda) + \phi_{\text{clouds+hazes}} \Delta_{\text{clouds+hazes}}(\lambda) + \phi_{\text{clear}} \Delta_{\text{clear}}(\lambda), \quad (9)$$

where the cover fractions are free parameters in the model and  $\phi_{\text{clear}}$  is determined by a unit-sum constraint, i.e.,  $\phi_{\text{clear}} = 1 - \phi_{\text{hazes}} - \phi_{\text{clouds}} - \phi_{\text{clouds+hazes}}$ .

Hazes, e.g., small-size particles resulting from photochemical processes, are implemented into our model atmosphere by parameterizing their effect on the spectrum as deviations from  $\text{H}_2$  Rayleigh scattering (Lecavelier Des Etangs et al. 2008). The parameterization provides a cross section  $\sigma_{\text{hazes}} = a \sigma_0 (\lambda/\lambda_0)^\gamma$ , where  $\gamma$  is the scattering slope,  $a$  is the Rayleigh-enhancement factor, and  $\sigma_0$  is the  $\text{H}_2$  Rayleigh-scattering cross section at the reference wavelength  $\lambda_0$ . We adopt values of  $\sigma_0 = 5.31 \times 10^{-31} \text{ m}^2$  and  $\lambda_0 = 350 \text{ nm}$  for consistency with previous works (e.g., MacDonald & Madhusudhan 2017; Welbanks et al. 2019). Future observations of non-H-rich planets could motivate the use of scattering cross sections for different species. The extinction due to hazes is  $\kappa_{\text{haze}}(\lambda, P, T) = X_{\text{H}_2} n_{\text{tot}}(P, T) \sigma_{\text{hazes}}(\lambda)$ .

The regions of the atmosphere covered by a gray cloud deck are included by adopting a parameter for the cloud top pressure  $P_{\text{cloud}}$ . The optical depth for all pressures higher than  $P_{\text{cloud}}$  is considered to be infinite. The extinction coefficient due to the cloud deck  $\kappa_{\text{clouds}}(P)$  is infinite for  $P > P_{\text{cloud}}$  or zero for  $P < P_{\text{cloud}}$ .

Previous studies have considered the effects of patchy clouds in transmission spectra (e.g., Line & Parmentier 2016; MacDonald & Madhusudhan 2017; Barstow 2020). Our model here generalizes the approach of previous studies while being able to reduce to previous treatments under specific conditions. If the model prefers to consider the presence of clouds and hazes together, the fractions  $\phi_{\text{hazes}}$  and  $\phi_{\text{clouds}}$  approach zero, and we obtain previous treatments for inhomogeneous cover (e.g., MacDonald & Madhusudhan 2017; Welbanks & Madhusudhan 2019). On the other hand, if the combined fraction is zero (e.g.,  $\phi_{\text{clouds+hazes}} = 0$ ), our approach allows us to consider the effect of clouds and hazes separately and distinguish whether the contribution to the spectrum is mostly due to deviations from  $\text{H}_2$  Rayleigh scattering produced by the hazes or muted features due to a gray cloud deck. Lastly, if the combined fraction is zero and so is the haze only fraction (e.g.,  $\phi_{\text{hazes}} = \phi_{\text{clouds+hazes}} = 0$ ), we recover the expression for patchy clouds of Line & Parmentier (2016). By following this



**Figure 3.** Schematic of the four-sector generalized parameterization of inhomogeneous clouds and hazes introduced in this work. The planet is enveloped by its atmosphere, which is divided into four regions. These are (1) a clear, cloud-free and haze-free, sector, (2) a sector with hazes only, (3) a sector with clouds only, and (4) a sector with clouds and hazes.

approach, we obtain a more robust and flexible treatment compared to our previous prescription that combines the effects of clouds and hazes into one sector (e.g., MacDonald & Madhusudhan 2017; Welbanks & Madhusudhan 2019). We present a schematic of our cloud and haze treatment in Figure 3.

We find the generalized treatment of clouds and hazes introduced in this work leads to consistent abundance estimates regardless of whether or not a H-rich atmosphere is assumed. In other words, the existing degeneracies between clouds/hazes and composition are treated equally irrespective of the assumption of the bulk atmospheric composition of the planet. On the other hand, combining clouds and hazes into one individual sector as previously performed (e.g., MacDonald & Madhusudhan 2017; Pinhas et al. 2018; Welbanks & Madhusudhan 2019) can lead to biases and an incomplete exploration of the parameter space that results in distinct solutions when assuming a H-rich atmosphere or not on the same data set. This is mitigated by our new cloud and haze prescription. We discuss these aspects further in the case study of the hot Jupiter HD 209458 b in Section 3.1.

#### 2.1.4. Radiative Transfer

Our model solves line-by-line radiative transfer in transmission geometry in a plane-parallel atmosphere. The model atmosphere is divided into a predetermined number of pressure layers equally spaced logarithmically. The number of layers and their span in pressure space can be arbitrarily established by the user depending on the application. For this work, and based on the empirical results of Welbanks & Madhusudhan (2019), we use 100 pressure layers uniformly spaced in  $\log_{10}(P)$  from  $10^{-6}$  to  $10^2$  bar under hydrostatic equilibrium. Our calculation of hydrostatic equilibrium is performed considering the retrieved composition through the atmospheric mean molecular weight (e.g.,  $\mu = \sum X_i m_i$ , where  $X_i$  and  $m_i$  are the volume mixing ratio and the atomic/molecular mass of species  $i$ , respectively), the retrieved P-T profile, and altitude-dependent gravity.

We solve numerically for the transit depth of the planet:

$$\Delta_\lambda = \frac{R_p^2 + 2 \int_{R_p}^{R_p+H_A} b(1 - e^{-\tau_\lambda(b)}) db - 2 \int_0^{R_p} b e^{-\tau_\lambda(b)} db}{R_{\text{star}}^2}, \quad (10)$$

where  $R_{\text{star}}$  is the stellar radius,  $H_A$  is the maximum height of the planetary atmosphere,  $\tau_\lambda$  is the total slant optical depth and integral of Equation (4),  $b$  is the impact parameter, and  $R_p$  is the radius of the planet. We present Equation (10) as a three-part expression to highlight the fact that the chosen  $R_p$  may not correspond to an optically thick surface. If  $R_p$  corresponds to an optically thick surface, the last integral in Equation (10) evaluates to zero. Otherwise, the integral considers the contribution of the nonoptically thick parts of the atmosphere, below the arbitrarily chosen position in the planet, to the transit depth.

The selected value of  $R_p$ , at a given reference pressure  $P_{\text{ref}}$ , is used to construct a radial distance grid. The distance and pressure grids follow a one-to-one correspondence determined by hydrostatic equilibrium. It is possible in a retrieval to choose a value of  $R_p$  for which the  $P_{\text{ref}}$  parameter will be retrieved, to choose a value of  $P_{\text{ref}}$  for which the associated radius  $R_p$  will be retrieved, or leave both  $R_p$  and  $P_{\text{ref}}$  as free parameters. Welbanks & Madhusudhan (2019) showed that the retrieval results remain mostly unchanged regardless of the choice of free parameter ( $R_p$  and/or  $P_{\text{ref}}$ ). In this work, we choose to keep  $R_p$  as our independent variable for which we retrieve  $P_{\text{ref}}$ .

#### 2.2. Considerations for Non-H-rich Atmospheres

A core assumption present in most atmospheric retrieval codes for hot Jupiters is that the atmosphere is H rich. Such assumption can be appropriate for massive planets that, from a formation perspective, captured a gas mixture of predominantly H and He in cosmic proportions from their protoplanetary nebula (Seager & Deming 2010). However, when characterizing the atmospheres of less massive planets or when pursuing an agnostic approach applicable to atmospheres of general composition, such assumption may need to be relaxed. Instead of assuming a H-rich atmosphere, studies could attempt to retrieve the main gas component of the atmosphere. Such an approach would aim to explore a wider range of atmospheric compositions like  $\text{N}_2$ -rich or  $\text{CO}_2$ -rich atmospheres, and not be constrained to H-rich atmospheres only.

However, when pursuing this agnostic approach, the unit-sum constraint, i.e., the requirement that all the volume mixing ratios in the atmosphere must add up to one, must be incorporated into the statistical modeling appropriately. Incorporating such constraint is nontrivial and has been the subject of study in a subfield of statistical analysis called compositional data analysis (e.g., Pearson 1897; Tanner 1949; Chayes 1960; Aitchison 1986). The tools developed by this subfield of statistics have been implemented in a number of different disciplines like medicine, chemistry, economy, geophysics, among many others (see Aitchison & Egozcue 2005 for a review of the history of compositional data). The concepts of compositional data analysis were introduced to the exoplanet retrieval literature through the work of Benneke & Seager (2012).

Implementing the same methods used for the retrieval of H-rich atmospheres to retrievals in which the main atmospheric

constituent is not known can result in biased results that do not explore all compositions equally. The traditional method would sample the volume mixing ratios of  $n - 1$  species (i.e., minor species) and assign the volume mixing ratio of the  $n$ th species (i.e.,  $H_2$  in the case of a H-rich atmosphere) following the unit-sum constraint. Benneke & Seager (2012) highlight that following this approach will result in a highly asymmetric prior (see Figure 1 of Benneke & Seager 2012) for the  $n$ th species. Under these circumstances, the retrieval is not truly agnostic, and the resulting atmospheric composition will be dependent on which molecule was chosen to be the  $n$ th species.

To circumvent this problem, one must allow for all species to have the same prior probability density in a permutation-invariant prescription. If the prior probability for all species is identical, it is safe for the retrieval to sample over the parameter space of all  $n$  species. The solution is the centered-log-ratio transformation, defined as

$$z_i = \log\left(\frac{X_i}{g(X)}\right), \quad (11)$$

where  $g(X)$  is the geometric mean  $g(X) = (X_1 \dots X_N)^{1/N}$  (Aitchison 1986). The transformed  $z_i$  values, also called compositional parameters, treat all parts of the gas symmetrically.

In Aurora, when not assuming a H-rich atmosphere, we reparameterize the volume mixing ratios ( $X_i$ ) using the centered-log-ratio transformation and obtain the compositional parameters ( $z_i$ ). We assume that the combination of  $H_2$  and He is one single part ( $z_{(H_2+He)}$ ), which we then use to determine the separate  $H_2$  and He volume mixing ratios using a He/ $H_2$  ratio. Then, we sample over the entire transformed space for all  $n$  gas components with the assumption that one of those is a mixture of  $H_2$  and He in solar proportion.

Once sampling is performed in the space of the centered-log-ratio transformation, and to maintain the descriptions above about the treatment of different opacity sources, the inverse transformation (Pawlowsky-Glahn & Buccianti 2011)

$$X_j = \frac{\exp(z_j)}{\sum_{i=1}^N \exp(z_i)} \quad (12)$$

is calculated and the volume mixing ratios  $X_i$ 's are used in our calculations.

It is important to highlight that the compositional parameters ( $z_i$ ) have slightly different properties from their counterparts, the volume mixing ratios ( $X_i$ ). While the typical prior range for  $X_i$  is  $10^{-12} < X_i < 1$ , the limits for  $z_i$  is  $-\infty < z_i < \infty$  where  $-\infty$  is the limit of a species not being present and  $\infty$  means the species is the only one in the atmosphere. While a straightforward expression for the scenario in which all volume mixing ratios are equal is not available, the compositional parameters are present in equal parts when all  $z_i = 0$ . Lastly, the unit-sum constraint for the volume mixing ratios is  $\sum X_i = 1$  and transforms to  $\sum z_i = 0$  for the compositional parameters.

### 2.3. Multialgorithmic Statistical Inferences

The strength in the retrieval approach when assessing the properties of an exoplanet's atmosphere resides in its ability to provide robust statistical estimates of the parameters and models used to explain the observations. As explained in Section 1, many statistical approaches exist in exoplanetary

atmospheric retrievals: grid-based searches (e.g., Madhusudhan & Seager 2009), MCMC (e.g., Madhusudhan & Seager 2011; Benneke & Seager 2012; de Wit & Seager 2013; Line et al. 2013; Madhusudhan et al. 2014; Cubillos 2016; Wakeford et al. 2017; Zhang et al. 2019), nonlinear optimal estimators (e.g., Lee et al. 2013; Barstow et al. 2017), among others (see Madhusudhan 2018, for a review). Of the different approaches available, Bayesian inference tools ease the comparison of models while providing estimates of the posterior distributions of the model parameters. One of these methods, nested sampling (Skilling 2006), has been successfully incorporated into exoplanetary retrieval literature (e.g., Benneke & Seager 2013; Line et al. 2013; Waldmann et al. 2015b; MacDonald & Madhusudhan 2017; Gandhi & Madhusudhan 2018; Krissansen-Totton et al. 2018; Pinhas et al. 2018; Mollière et al. 2019; Zhang et al. 2020) due to its ability to handle high-dimensionality problems, sample the complete parameter space of the model, and use prior information on the model parameters. An overview of the Bayesian approach to inference problems is available in Sivia & Skilling (2006) and Trotta (2008, 2017).

The likelihood of observing the data ( $\mathcal{D}$ ) given a specific set of model parameters ( $\theta_{\mathcal{M}}$ ) for a model ( $\mathcal{M}$ ) is

$$\mathcal{L} = P(\mathcal{D}|\theta_{\mathcal{M}}, \mathcal{M}). \quad (13)$$

Considering the Bayesian approach, where the degree of belief on the model assumptions must be accounted for, one must incorporate the prior distribution ( $\pi$ ) on the model parameters  $\pi = P(\theta_{\mathcal{M}}|\mathcal{M})$ . The marginalized likelihood, also known as evidence, is obtained by integrating the likelihood over the full parameter space:

$$\mathcal{Z} = \int P(\mathcal{D}|\theta_{\mathcal{M}}, \mathcal{M})P(\theta_{\mathcal{M}}|\mathcal{M})d\theta_{\mathcal{M}} = P(\mathcal{D}, \mathcal{M}). \quad (14)$$

The model evidence is the quantity we are interested in evaluating when comparing different models. This is also the quantity different nested sampling algorithms aim to provide. Furthermore, using Bayes' theorem, it is possible to obtain the posterior probability distribution for each parameter given the data

$$P(\theta_{\mathcal{M}}|\mathcal{D}, \mathcal{M}) = \frac{P(\mathcal{D}|\theta_{\mathcal{M}}, \mathcal{M})P(\theta_{\mathcal{M}}|\mathcal{M})}{P(\mathcal{D}, \mathcal{M})}. \quad (15)$$

Aurora uses a likelihood function for data with independently distributed Gaussian errors

$$\mathcal{L}(\mathcal{D}|\theta_{\mathcal{M}}, \mathcal{M}_i) = \prod_k \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{[\mathcal{D}_k - \mathcal{M}_{i,k}]^2}{2\sigma_k^2}\right) \quad (16)$$

for a data set of length  $N$  and computed for each model realization  $\mathcal{M}_i$ . Aurora follows the same binning strategy as AURA (see Section 2.1.6 in Pinhas et al. 2018) where a model spectrum at a much higher resolution than the data is convolved with the point-spread function of the instrument with which the observations were obtained and then binned down to the spectral resolution of the data.

The prior distributions employed in this study are shown in Table A1 in Appendix A. The priors for the parameters are mostly standard prescriptions adopted from previous studies (e.g., Pinhas et al. 2019; Welbanks et al. 2019). The priors for molecular abundances generally span the complete detectable range unless stated otherwise, with the prior distribution either

log-uniform for the volume mixing ratios for H-rich retrievals or uniform in the corresponding compositional parameters ( $z_i$ ), discussed in Section 2.2, for non-H-rich retrievals. The priors for the parameters associated with other physical properties, e.g., P-T profile and cloud/haze parameters, are also uniform or log-uniform and span the corresponding physically plausible ranges.

### 2.3.1. Next-generation Bayesian Inference Algorithms

The main functionality of a nested sampling algorithm is to obtain the model evidence ( $\mathcal{Z}$ ) while also deriving the posterior probability distributions of the model parameters as a by-product. A full description of the nested sampling algorithm is available in Skilling (2004, 2006), Feroz et al. (2009). In Aurora, we implement three different algorithms, MultiNest (Feroz et al. 2009, 2013) through its implementation PyMultiNest (Buchner et al. 2014), PolyChord (Handley et al. 2015a, 2015b) through its implementation PyPolyChord, and Dynesty (Speagle 2020). Each nested sampling algorithm is different and the in-depth description for each implementation is available in their release papers listed above.

Generally, a nested sampling algorithm generates a number of live points drawn from the prior distribution, which sample the parameter space (Feroz et al. 2009). In each iteration, the point with the lowest likelihood is replaced by a new one which ought to have a larger likelihood. This means that the live points sample the prior volume using continuously shrinking isolikelihood contours, which with every iteration converge to the highest likelihood regions of the parameter space. At each step, every sampled value creates a model realization that results in an evaluation of the likelihood function. The process finishes when a termination condition, like a preset fractional change in the model likelihood, is met. Upon completion, the combination of all sampled points can be used to estimate the model evidence. The procedure to generate new live points can vary between different implementations of the nested sampling algorithm, which are briefly discussed below.

MultiNest has been previously implemented in exoplanet retrievals (e.g., Benneke & Seager 2013; Line et al. 2013; Waldmann et al. 2015b; MacDonald & Madhusudhan 2017; Gandhi & Madhusudhan 2018; Krissansen-Totton et al. 2018; Pinhas et al. 2018; Mollière et al. 2019; Zhang et al. 2020). To draw unbiased samples from the likelihood-constrained prior, MultiNest uses what is called an ellipsoidal rejection sampling scheme. The basis for this scheme is that the replacement point is sought from within the set of ellipsoids described by the full set of live points at any iteration (Feroz et al. 2019). With each iteration, the ellipsoids described by the isolikelihood contours shrink. This procedure is optimal for a small number of parameters but has an exponential scaling with dimensionality.

PolyChord, on the other hand, uses what is called slice-based sampling. In this procedure, the algorithm samples uniformly within the parameter space for which the posterior probability is higher than a given probability level or “slice.” Unlike the exponential scaling problem with MultiNest at higher dimensions, PolyChord’s scaling is  $\sim \mathcal{O}(D^3)$  (Handley et al. 2015b). This makes MultiNest preferred for low-dimensionality problems, while PolyChord is preferred at higher dimensionalities (see Figure 7 in Handley et al. 2015b).

Lastly, Dynesty (Speagle 2020) uses a generalization of nested sampling, in which the number of live points is variable, called dynamic nested sampling (Higson et al. 2019). In

dynamic nested sampling, an initial run with a constant number of live points is used by the algorithm to approximate areas in prior space of the highest likelihood. Then, the algorithm proceeds to iteratively calculate the range of likelihoods where a larger number of live points will have the greatest result in accuracy. In dynamic nested sampling, the number of live points is dynamically allocated to control the resolution at which the prior space is sampled. This would allow for runs that focus on sampling the posterior distribution or better estimate the model evidence. Dynesty allows for both dynamic and static nested sampling. Furthermore, Dynesty has four main approaches to generating samples: uniform sampling (including from ellipsoids like MultiNest), random walks, multivariate slice sampling (similar to PolyChord), and Hamiltonian slice sampling. Each approach has its benefits and impediments and can be better suited for different problem dimensionalities. Speagle (2020) offers an extensive overview of each feature available in Dynesty.

Every algorithm for nested sampling offers different capabilities. While Dynesty is able to handle both static and dynamic sampling, it comes at the cost of multiple tuning parameters that can affect the behavior of a given run. PolyChord is able to handle problems of higher dimensionality more efficiently than MultiNest, but MultiNest still outperforms PolyChord in the number of likelihood evaluations required for problems in low dimensions ( $\lesssim 80$  dimensions, Handley et al. 2015b). Aurora offers the tools to perform retrievals optimizing for evaluation of the model evidence, parameter posterior distributions, or both. The user has the freedom to choose the correct sampling algorithm for their needs depending on the complexity of the problem and its dimensionality.

### 2.3.2. Model Comparison and Detection Significance

The difference in evidence ( $\mathcal{Z}$ ) between models can be used to derive an equivalent detection significance (DS), a figure of merit traditionally used to compare different models. The detection significance is traditionally expressed in units of “sigma” ( $\sigma$ ) and corresponds to the number of standard deviations away from the mean of a normal distribution (Trotta 2008). Expressing a result in “sigmas” does not necessarily mean the detection of new physics or a species in the spectrum of a planet. Instead, it is a useful way to translate the odds in favor of a more complex model into a frequentist metric. The relevance of a model preference can be somewhat arbitrary, and different authors suggest different categories for expressing them. For instance, Trotta (2008) suggests that a difference of  $2.0\sigma$  to  $2.6\sigma$  is weak at best, while Kass & Raftery (1995) suggest that the equivalent of  $\sim 2.1\sigma$  is positive evidence. A way to transform the difference in model evidence to a detection significance was proposed by Benneke & Seager (2013). We perform the comparison of our models by solving Equation (11) in Benneke & Seager (2013) and obtaining a detection significance as

$$DS = \sqrt{2} \operatorname{erfc}^{-1} \left[ \exp \left( \mathcal{W}_{-1} \left( -\frac{1}{B e} \right) \right) \right], \quad (17)$$

where  $\operatorname{erfc}$  is the complementary error function,  $\mathcal{W}_{-1}$  is the Lambert  $\mathcal{W}$  function in its lower branch (i.e.,  $k = -1$  branch),  $e$  is Euler’s number, and  $B$  is the Bayes factor defined as



$B = Z_1/Z_2$ , with the set requirement of  $Z_1 \geq Z_2$  so the Bayes factor is greater than or equal to unity.

#### 2.4. Modular Capabilities

Aurora’s design is modular, ensuring that future capabilities can be easily incorporated into the existing retrieval framework. As part of this modular structure, we include in Aurora preexisting features in AURA (Pinhas et al. 2018), such as the functionality to retrieve stellar properties from a transmission spectrum. Furthermore, we introduce new modular capabilities that aid in the analysis of transmission spectra in the context of retrievals and forward models. These key additions include new considerations for noise modeling and forward models considering light refraction, forward scattering, and Mie scattering.

##### 2.4.1. Stellar Heterogeneity

One of the main features of AURA (Pinhas et al. 2018) was to retrieve stellar properties embedded in the transmission spectrum as well as the planetary properties. Inhomogeneities in the stellar photosphere were modeled by retrieving the areal fraction of the projected stellar disk covered by heterogeneities ( $\delta$ ), hot faculae or cool spots, the heterogeneity temperature ( $T_{\text{het}}$ ), and the photospheric temperature ( $T_{\text{phot}}$ ). Aurora inherits this capability but we do not include it in the present study.

##### 2.4.2. New Noise Modeling Modules

Aurora has the capability to treat noise models that are different from the traditionally assumed white noise. Aurora can consider the possibility of underestimated variance in the data by retrieving an error bar inflation, which is a free parameter (Foreman-Mackey et al. 2013). This approach assumes that the variance is underestimated by a fractional amount  $f$ . The variance of the data is

$$S^2 = \sigma_{\text{obs}}^2 + f^2 \Delta_{\text{mod}}^2, \quad (18)$$

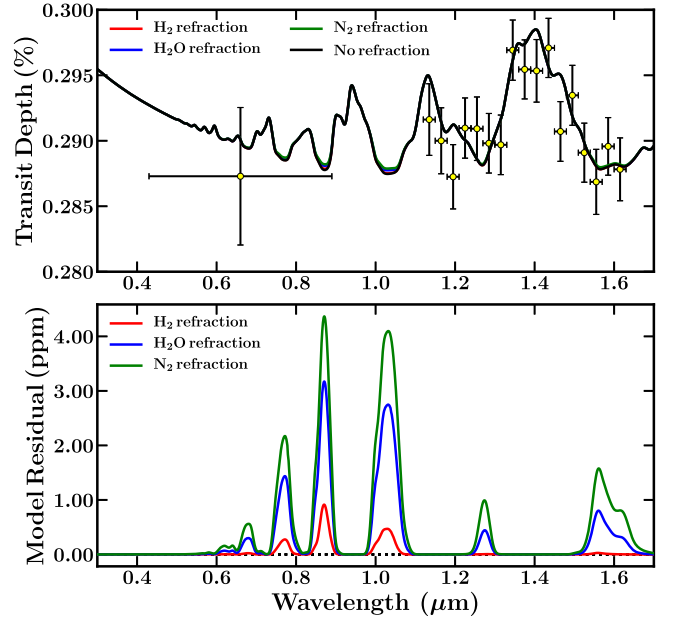
where  $\sigma_{\text{obs}}$  is the error in the observations, and  $\Delta_{\text{mod}}$  is the model’s transit depth. This term replaces the variance term in Equation (16). This functionality has been tested on the recent spectroscopic observations of KELT-11b (Colón et al. 2020).

Aurora also has the capability to consider correlated noise in the data being analyzed. To do so, we have incorporated Celerite (Foreman-Mackey et al. 2017) and George (Ambikasaran et al. 2015) to model the covariance function and compute the likelihood of a Gaussian Process (GP) model (Rasmussen & Williams 2006). The effects of a GP in transmission spectra fall beyond the scope of this paper and we reserve it for a future study.

##### 2.4.3. Refraction and Forward Scattering

In Aurora, we have incorporated the analytic descriptions for forward scattering and refraction of transit spectra proposed by Robinson et al. (2017). These prescriptions have been incorporated in the context of producing forward models and synthetic observations.

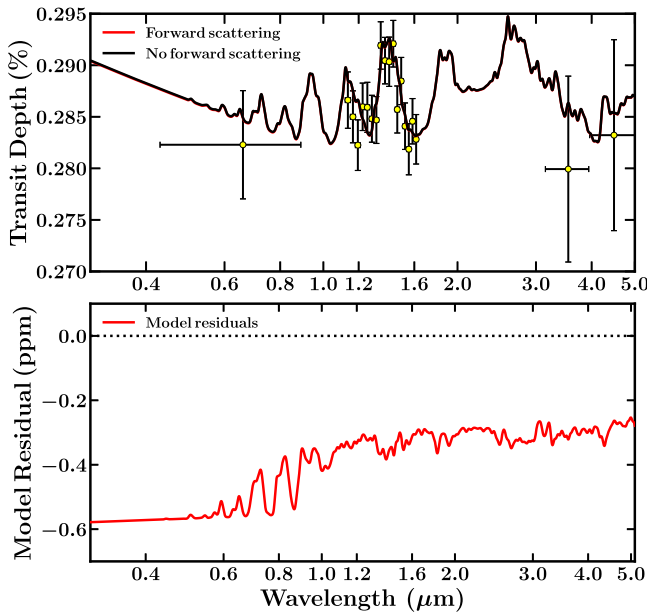
Refraction effects are calculated using the prescription for maximum pressure at which the effect of refraction is large enough to cause a light ray from one side of the planet to come from the far limb (i.e., opposite side) of the host star (Robinson et al. 2017). We incorporate the wavelength-dependent



**Figure 4.** Forward models including the effects of wavelength-dependent refraction for the mini Neptune K2-18b. Top: no-refraction model shown in black while the red, blue, and green lines show the effects of H<sub>2</sub>, H<sub>2</sub>O, and N<sub>2</sub> refraction, respectively. Black error bars with yellow markers show current K2 and HST-WFC3 observations for reference. Bottom: residual of the refraction models relative to the no-refraction model.

refractivity (Robinson et al. 2017) and use it to calculate the maximum pressure probed ( $P_{\text{max}}$ ) at each wavelength following Equation (15) of Robinson et al. (2017). The optical depth for pressures higher than  $P_{\text{max}}$  is set to infinity. Figure 4 shows the effect of considering refraction in forward models of K2-18b. For these forward models, we consider refraction due to H<sub>2</sub>, H<sub>2</sub>O, or N<sub>2</sub>. The forward models are determined by the median retrieved parameters in Section 3.2. Figure 4 shows that the effects of refraction are almost negligible,  $\sim 4$  ppm. Additional models considering the effect of refraction for a rocky exoplanet are shown in Appendix B.

The standard forward model in Aurora combines the absorption and scattering optical depths into the total optical depth as seen in Equation (4). However, it is possible that a portion of the scattered light in the planet’s atmosphere will be directed toward the observer. This portion of light is said to be forward scattered. The additional fraction of light reaching the observer results in attenuation to the transit depth. In Aurora, we can model this by correcting the effective optical depth for the effects of forward scattering. The modified optical depth is  $d\tau_{\text{eff}} = d\tau_{\lambda}(1 - f\tilde{\omega}_o)$ , where  $f$  is the forward-scattering correction factor and  $\tilde{\omega}_o$  is the single scattering albedo (Robinson et al. 2017). We calculate the correction factor  $f$  using the analytic correction expressed in Equation (6) of Robinson et al. (2017) for the Henyey–Greenstein phase function (Henyey & Greenstein 1941). The correction proposed by Robinson et al. (2017) is a function of the stellar radius, the planet–star physical separation, and the asymmetry parameter  $g$ . Figure 5 shows the decrease in transit depth due to considering forward scattering, in the same H<sub>2</sub>-rich forward model for K2-18b described above, assuming  $g = 0.95$  and  $\tilde{\omega}_o = 1$ . The effect of incorporating forward scattering in the model of K2-18b results in a difference of less than 1 ppm. Models considering forward scattering for a rocky exoplanet are shown in Appendix B.



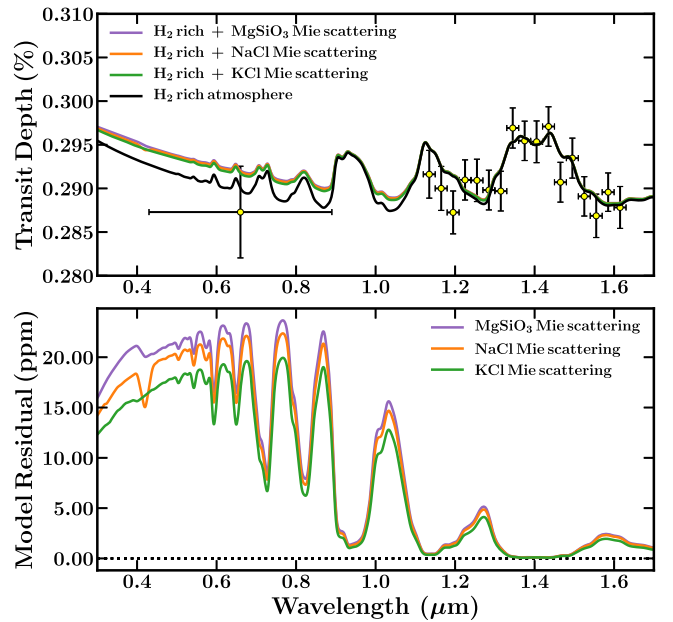
**Figure 5.** Forward-scattering models for the atmosphere of the mini Neptune K2-18b. Top: forward-scattering model shown in red, and in black, the model without forward scattering. Black error bars with yellow markers show current K2, HST-WFC3, and Spitzer observations for reference. Bottom: residual of the forward-scattering mode relative to the non-forward-scattering model.

The detectability of these secondary effects remains to be confirmed. Current observations using Hubble Space Telescope (HST) and ground-based observatories do not possess the precision necessary to identify them. In the meantime, Aurora possesses the capabilities to model these effects in transmission spectra of exoplanets in the context of forward models. The implementation of these models in the context of retrievals remains a possibility for future studies should the data require so.

#### 2.4.4. Mie-scattering Forward Models

Aurora includes Mie scattering in the forward models due to condensates with different particle sizes and compositions adopted from Pinhas & Madhusudhan (2017). The effective cross section for these species is calculated using their extinction and scattering coefficients, along with the corresponding asymmetry parameter  $g$  following Equation (11) of Pinhas & Madhusudhan (2017), obtained using Mie theory.

Figure 6 shows the spectroscopic features of Mie scattering for different compositions in the  $H_2$ -rich atmosphere of K2-18b. The models assume the retrieved chemical abundances from the results in Section 3.2. The models shown include  $H_2$  Rayleigh scattering and  $H_2$ - $H_2$  and  $H_2$ -He CIA. In black, we show the  $H_2$ -rich atmosphere only, while in purple, orange, and green the effects of Mie scattering due to  $MgSiO_3$ , NaCl, and KCl are shown, respectively. The assumed abundances for the condensate species is  $10^{-16}$ , similar to expectations for NaCl and KCl from equilibrium chemistry calculations (e.g., Woitke et al. 2018) for the equilibrium temperature of the planet of  $T_{eq} \sim 290$  K (e.g., Welbanks et al. 2019), with a particle size of  $4.89 \times 10^{-2} \mu m$  ( $< 0.1 \mu m$ , e.g., Adams et al. 2019; Lavvas et al. 2019). As shown in the bottom panel of Figure 6, the maximum difference between the clear  $H_2$ -rich model and the models considering Mie scattering is  $\sim 25$  ppm, within the



**Figure 6.** Top: forward models including the effects of Mie scattering for the mini Neptune K2-18b. The condensates shown are  $MgSiO_3$  (purple), NaCl (orange), and KCl (green). In black, the  $H_2$ -rich-only model is shown. Black error bars with yellow markers show current K2 and HST-WFC3 observations for reference. Bottom: model residuals.

precision limits of current observations. Future observatories with high-precision measurements in the optical wavelengths may be able to distinguish the effects of these condensate species in the atmospheres of exoplanets.

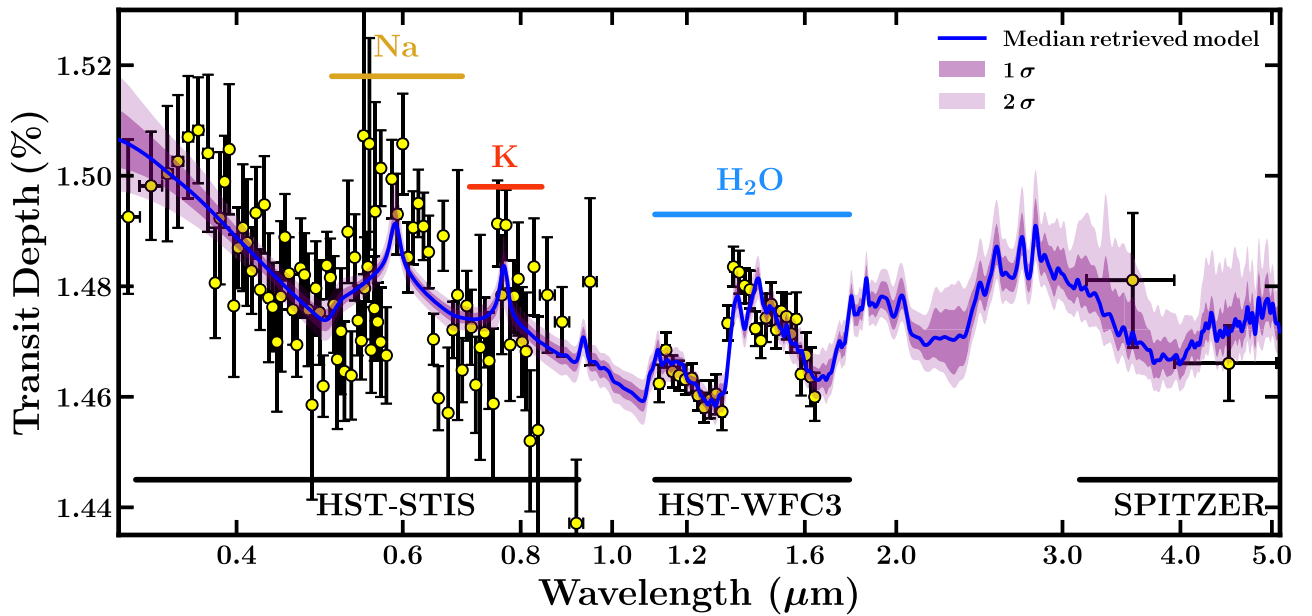
### 3. Results

We validate Aurora’s new retrieval features on real and synthetic spectrophotometric observations. First, we validate our H-rich and non-H-rich approaches as well as the new prescription for inhomogeneous cloud and haze cover on the prototypical hot Jupiter HD 209458 b (Charbonneau et al. 2000; Henry et al. 2000) using observations from Sing et al. (2016). Next, we test the different nested sampling algorithms included in Aurora using the most recent observations of K2-18b (Foreman-Mackey et al. 2015) from Benneke et al. (2019b) and investigate the robustness of the retrieved abundance estimates comparing them to previous works (e.g., Benneke et al. 2019b; Welbanks et al. 2019; Madhusudhan et al. 2020). Lastly, we investigate future atmospheric constraints of mini Neptunes and rocky exoplanets using synthetic observations.

#### 3.1. Validation of Aurora on Hot Jupiter HD 209458 b

We perform a series of retrievals on the transmission spectrum of HD 209458 b from Sing et al. (2016), composed of spectrophotometric observations with HST-STIS, HST-WFC3, and Spitzer. We use the standard model setup described in Welbanks & Madhusudhan (2019), Pinhas et al. (2019), and Welbanks et al. (2019). Our sources of opacity include  $H_2$ - $H_2$  and  $H_2$ -He CIA,  $H_2$  Rayleigh scattering, and line opacity due to  $H_2O$ , Na, K,  $CH_4$ ,  $NH_3$ , HCN, CO, and  $CO_2$ . We conduct a range of retrievals with different cloud and haze prescriptions, and assumptions of whether or not the atmosphere is H rich.

We perform retrievals using four models with different considerations for clouds and hazes allowed by our generalized



**Figure 7.** Validation of Aurora’s retrieval framework on the spectrum of HD 209458 b. Retrieved median model (blue line),  $1\sigma$  and  $2\sigma$  confidence intervals (purple shaded regions), and spectroscopic observations (black error bars with yellow markers) for the highest evidence model (model 3; see Table C1). Horizontal lines at the bottom of the figure show the wavelength coverage of HST-STIS, HST-WFC3, and Spitzer. Yellow, orange, and blue horizontal lines show the approximate wavelength regions where Na, K, and  $\text{H}_2\text{O}$  spectral features are expected, respectively.

prescription explained in Section 2.1.3. Model 0 considers a clear atmosphere (i.e.,  $\phi_{\text{clouds}} = \phi_{\text{hazes}} = \phi_{\text{clouds+hazes}} = 0$ ). Model 1 considers one sector for a clear atmosphere and one sector for the combined effects of clouds and hazes (i.e.,  $\phi_{\text{clouds}} = \phi_{\text{hazes}} = 0$ ). Model 2 considers one sector for a clear atmosphere, one sector for the presence of clouds only, and one sector for the presence of hazes only (i.e.,  $\phi_{\text{clouds+hazes}} = 0$ ). Model 3 considers one sector for a clear atmosphere, one sector for clouds only, one sector for hazes only, and one sector for the combined presence of clouds and hazes (i.e., the full inhomogeneous cloud and haze prescription introduced in this work). For each cloud and haze model above, we perform a retrieval assuming a H-rich atmosphere and a retrieval relaxing such assumption. In summary, we perform eight retrievals in this section with the models above, four assuming a H-rich atmosphere and four not assuming a H-rich atmosphere.

### 3.1.1. A Generalized Cloud and Haze Prescription

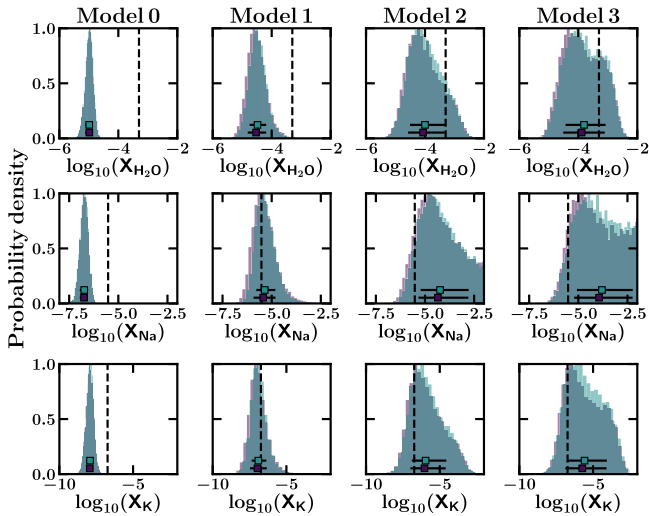
We consider a generalized cloud and haze prescription in order to explore a larger parameter space than available when restricting the presence of clouds and hazes to one sector only (i.e., model 1). We find that assuming a H-rich atmosphere or not can result in different solutions when restricting the clouds and hazes to the same region, as in model 1 (e.g., MacDonald & Madhusudhan 2017). This is not the case for any of the other models in this section (i.e., model 0, model 2, or model 3). When assuming a H-rich atmosphere, we find that, using any of the models for inhomogeneous cloud/haze cover, the spectrum of HD 209458 b can be explained by two possible scenarios. The first is the known solution with median values of subsolar<sup>1</sup>  $\text{H}_2\text{O}$  of  $\log_{10}(X_{\text{H}_2\text{O}}) \sim -4.5$ ,

$\log_{10}(X_{\text{Na}}) \sim -5.2$ ,  $\log_{10}(X_{\text{K}}) \sim -7.0$ , and a cloud and haze cover of roughly 50% (e.g., MacDonald & Madhusudhan 2017; Pinhas et al. 2019; Welbanks & Madhusudhan 2019; Welbanks et al. 2019); the second is a physically implausible solution with high Na abundances that make up for  $\sim 20\%$  of the atmosphere’s composition and an atmosphere fully covered by clouds and hazes (e.g.,  $\log_{10}(X_{\text{H}_2\text{O}}) \sim -2.5$ ,  $\log_{10}(X_{\text{Na}}) \sim -0.7$ ,  $\log_{10}(X_{\text{K}}) \sim -2.4$ ). Both modes are simultaneously retrieved by model 2 and model 3, regardless of whether or not a H-rich composition is assumed. However, when treating clouds and hazes together (i.e., model 1), while assuming a H-rich atmosphere results in the two modes as discussed above, relaxing the H-rich assumption results in only the high Na abundance solution. Therefore, model 1 may be susceptible to potential biases in retrieved solutions when the dominant atmospheric composition may not be assumed to be H rich a priori. On the other hand, models 2 and 3 provide more generalized parameterizations that do not depend strongly on the H-rich assumption. In what follows, we restrict the prior space of the  $\log_{10}$  abundances of Na, K and CO to an upper limit of  $-1.5$ , consistent with assumptions in previous studies (e.g., MacDonald & Madhusudhan 2017; Pinhas et al. 2019; Welbanks & Madhusudhan 2019; Welbanks et al. 2019). We implement this upper limit by rejecting the unphysical solutions.

### 3.1.2. Effect of Cloud and Haze Prescriptions

We rerun all eight cases with the new constraints on the abundances of Na, K and CO. We present the complete set of retrieved parameters for the four cloud and haze models assuming a H-rich atmosphere and not assuming a H-rich atmosphere in Table C1 included in Appendix C. Figure 7 shows the median retrieved spectrum for model 3, which results in the highest model evidence, while Figure 8 shows the  $\text{H}_2\text{O}$ , Na, and K posterior distributions for all four cloud and haze models.

<sup>1</sup> We clarify that in this context, we refer to abundances of  $\text{H}_2\text{O}$  as subsolar by assessing them relative to expectations from thermochemical equilibrium for solar elemental abundances (Asplund et al. 2009). For a solar composition, the expectation is a  $\text{H}_2\text{O}$  abundance of  $\log_{10}(X_{\text{H}_2\text{O}}) \sim -3.3$  for a planet with the equilibrium temperature of HD 209458 b (Madhusudhan 2012).



**Figure 8.** Posterior distributions for  $\text{H}_2\text{O}$ , Na and K abundances under the four different cloud and haze models. The purple (blue) distributions correspond to retrievals (not) assuming a H-rich atmosphere. Solar abundance expectations are shown using vertical black dashed lines. The markers in the posterior distributions show the median value (square marker) and the  $1\sigma$  (error bars) range covering  $\sim 68.27\%$  of the samples around the median value.

Considering a cloud-free atmosphere (model 0) results in tight  $\text{H}_2\text{O}$  abundance constraints with precisions smaller than 0.5 dex. Regardless of the treatment for the main gas constituent in the atmosphere, both cloud-free retrievals result in subsolar  $\text{H}_2\text{O}$  abundances with abundance estimates smaller than the models considering clouds and hazes. These low abundances are the consequence of having a larger observable atmosphere (i.e., larger atmospheric column), unocculted by clouds, in which a small abundance of  $\text{H}_2\text{O}$  can contribute enough to explain the observations (see e.g., Welbanks & Madhusudhan 2019).

Contrary to the cloud-free solutions, the cloudy and hazy scenarios (i.e., models 1, 2, and 3) result in higher  $\text{H}_2\text{O}$  abundances, although with still generally subsolar values. Models 1, 2 and 3 are consistent in their retrieved parameters when assuming a H-rich atmosphere and when relaxing this assumption. The retrieved  $\text{H}_2\text{O}$  abundances are consistent with each other and within  $1\sigma$  between all three cloud and haze models. The same is true for the Na and K abundances. Model 1, consisting of one fraction combining clouds and hazes as in MacDonald & Madhusudhan (2017), results in tighter constraints relative to models 2 and 3. These tighter constraints indicate that part of the parameter space explored by the other two prescriptions was not considered in model 1. The increase in model evidence for model 2 and model 3 relative to model 1 indicates that the increased parameter space contains previously unsampled regions of high likelihood. The retrieved P-T profiles are consistent between models 1, 2, and 3, with a retrieved temperature near the photosphere for model 3 assuming a H-rich atmosphere of  $T_{100\text{ mbar}} = 1308^{+345}_{-278}$  K, consistent with previous studies (e.g., Welbanks & Madhusudhan 2019; Welbanks et al. 2019). On the other hand, the retrieved P-T profile for model 0 is tightly constrained at colder temperatures (e.g.,  $T_{100\text{ mbar}} = 852^{+20}_{-12}$  K for the retrieval assuming a H-rich atmosphere) and inconsistent with the planet’s equilibrium temperature ( $T_{\text{eq}} \sim 1450$  K, e.g., Welbanks et al. 2019).

When comparing the retrievals assuming H-rich atmospheres, model 3 has the highest model evidence with a value of  $\log(\mathcal{Z}) = 958.40$ . Using model 3 as our reference, model 0 is strongly disfavored at  $4.6\sigma$ ; model 1 is disfavored at  $1.8\sigma$ ; and model 2 is weakly disfavored at  $1.4\sigma$ . A similar interpretation is available when comparing the non-H-rich retrievals among themselves.

### 3.1.3. H-rich versus Non-H-rich Assumptions

We also compare retrievals assuming a H-rich atmosphere against retrievals not assuming a H-rich atmosphere. Relaxing the assumption of a H-rich atmosphere requires an additional parameter to retrieve the volume mixing ratio of a mixture of  $\text{H}_2$  and He in solar proportions. This additional parameter results in a decrease in model evidence relative to retrievals assuming a H-rich atmosphere. Retrievals using model 3 favor assuming a H-rich atmosphere at  $2.82\sigma$  over not assuming a H-rich atmosphere. Despite this decrease in evidence, retrievals not assuming a H-rich atmosphere find that 99.9% of the atmosphere is made up of  $\text{H}_2$  and He. By not assuming a H-rich atmosphere a priori, our models are able to robustly confirm that the data corresponds to the atmosphere of a H-rich planet. Our results indicate that assuming a H-rich atmosphere is appropriate for the spectrum of HD 209485 b as expected. These results demonstrate for gas giants that both approaches, whether or not assuming a H-rich atmosphere, are consistent and that the retrieved parameter estimates are robust against either methodology.

### 3.1.4. Assessing the Highest Evidence Model

The highest evidence model (i.e., model 3, H-rich assumption) results in retrieved abundance estimates for  $\text{H}_2\text{O}$ , Na, and K that are consistent with previous results (e.g., Barstow et al. 2017; MacDonald & Madhusudhan 2017; Pinhas et al. 2019; Welbanks & Madhusudhan 2019; Welbanks et al. 2019). However, their precisions are wider with abundances of  $\log_{10}(X_{\text{H}_2\text{O}}) = -3.89^{+0.78}_{-0.62}$ ,  $\log_{10}(X_{\text{Na}}) = -3.95^{+1.71}_{-1.27}$ , and  $\log_{10}(X_{\text{K}}) = -5.73^{+1.65}_{-1.15}$ . The wider estimates result in a median  $\text{H}_2\text{O}$  abundance that is still subsolar based on expectations of thermochemical equilibrium, but consistent with a solar value to within  $1\sigma$ . Importantly, while the  $\text{H}_2\text{O}$  abundance is largely subsolar, both Na and K abundances are significantly super-solar, implying a relative depletion in  $\text{H}_2\text{O}$  compared to Na and K as found in Welbanks et al. (2019). The retrieved cloud and haze parameters indicate a nonclear atmosphere covered by clouds and hazes with a cloud deck located above the expected photosphere. The retrieved fractions are  $\phi_{\text{clouds}} = 0.34^{+0.18}_{-0.20}$ ,  $\phi_{\text{hazes}} = 0.27^{+0.09}_{-0.10}$ , and  $\phi_{\text{clouds+hazes}} = 0.24^{+0.19}_{-0.16}$ .

Noteworthy, too, are the retrieved values for the Rayleigh-enhancement factor. Model 3 (H-rich) retrieves a Rayleigh-enhancement of  $\log_{10}(a) = 3.28^{+1.01}_{-1.13}$ , while model 2 (H-rich) retrieves  $\log_{10}(a) = 2.88^{+0.91}_{-0.85}$ . Both retrieved Rayleigh-enhancement factors have median values and upper limits smaller than the retrieved median value for model 1 (e.g.,  $\log_{10}(a) = 4.35^{+0.71}_{-1.01}$  for the H-rich case). This may indicate a tendency to over estimate the Rayleigh-enhancement factor in the hazes when using model 1 (e.g., MacDonald & Madhusudhan 2017). If true, this possibility must be accounted for when studying the nature of super-Rayleigh slopes as performed in recent studies (e.g., Ohno & Kawashima 2020). Similarly, although consistent with each other, the retrieved

median value for the scattering slope  $\gamma$  is higher for model 1 than for models 2 and 3. The constraints from the H-rich retrievals are  $\gamma = -14.04^{+4.53}_{-3.94}$  for model 1,  $\gamma = -16.57^{+3.15}_{-2.37}$  for model 2, and  $\gamma = -16.15^{+3.36}_{-2.60}$  for model 3. We note that the interpretation of the Rayleigh-enhancement factor ( $a$ ) should be done in conjunction with the value for the scattering slope ( $\gamma$ ) as these parameters are correlated. Lastly, the retrieved cloud top pressure for the gray cloud deck is consistent within  $1\sigma$  between all approaches with a retrieved value of  $\log_{10}(P_{\text{cloud}}) = -4.72^{+0.99}_{-0.84}$  for the model with the highest evidence.

Finally, we compare model 2 to model 3. The retrieved parameters are consistent between the two approaches and have similar precisions. Due to their similar performance and relatively small difference in model evidence, we consider both approaches interchangeable for the effects of this work. In what follows we consider model 2 (i.e., one sector for a clear atmosphere, one sector for clouds only and one sector for hazes only) as our preferred model due to its smaller number of parameters and similar performance to model 3 (i.e., full inhomogeneous cloud and haze prescription). We utilize model 2 as our approach for inhomogeneous cloud and hazes in the remaining of the results section unless otherwise stated.

### 3.2. Testing Multiple Nested Sampling Algorithms

We validate the different nested sampling algorithms in Aurora by performing retrievals using the same model and the same data. We discuss three nested sampling algorithms in Section 2.3.1. Four retrievals are performed, one using MultiNest, one using PolyChord, one using Dynesty in its static nested sampling mode, and one using Dynesty in its dynamic nested sampling mode. We use the observed transmission spectrum of K2-18b from Benneke et al. (2019b) including K2-band photometry, HST-WFC3 G141 grism spectra, and Spitzer IRAC photometric observations. The model considers an isothermal and clear atmosphere. We assume a H-rich atmosphere and consider the absorption due to H<sub>2</sub>-H<sub>2</sub> and H<sub>2</sub>-He CIA, H<sub>2</sub>O, CH<sub>4</sub>, NH<sub>3</sub>, CO, and CO<sub>2</sub>. In total, the model has seven free parameters: five molecular species, one parameter for the temperature of the isotherm, and one parameter for the reference pressure. The retrieved parameters are used to produce the forward models in Sections 2.4.3 and 2.4.4.

When initializing the nested sampling algorithms, different parameters responsible for the algorithm's settings can be modified. Examples of such parameters are the maximum number of iterations in the sampling algorithm (PyMultiNest), parameters to increase the number of posterior samples produced (PyPolyChord), or the maximum number of likelihood evaluations before terminating (Dynesty). We keep most settings for the nested sampling algorithms to their default values. We only modify parameters needed for a direct comparison, e.g., the number of live points used to sample the prior distributions.

MultiNest was set up with 2000 live points. PolyChord was also set up with 2000 live points and 7 repeats. The number of repeats is specific to PolyChord's settings, and it corresponds to the length of the sampling chain used to generate a new live point. The longer the chain, the less correlated the live points and the more reliable the evidence inference is; however, the run takes longer to complete. The default value for the number of repeats used by PolyChord is  $5\times$  the number of dimensions

in the problem, that is,  $5\times$  the number of model parameters. Because we do not need an estimate for the model evidence in this exercise, as we are not comparing the model evidence between samplers, we do not need to choose a significantly larger number of repeats. We find that for our atmospheric model with seven free parameters (i.e., seven dimensions), our choice of seven repeats (i.e.,  $1\times$  the number of dimensions) is sufficient.

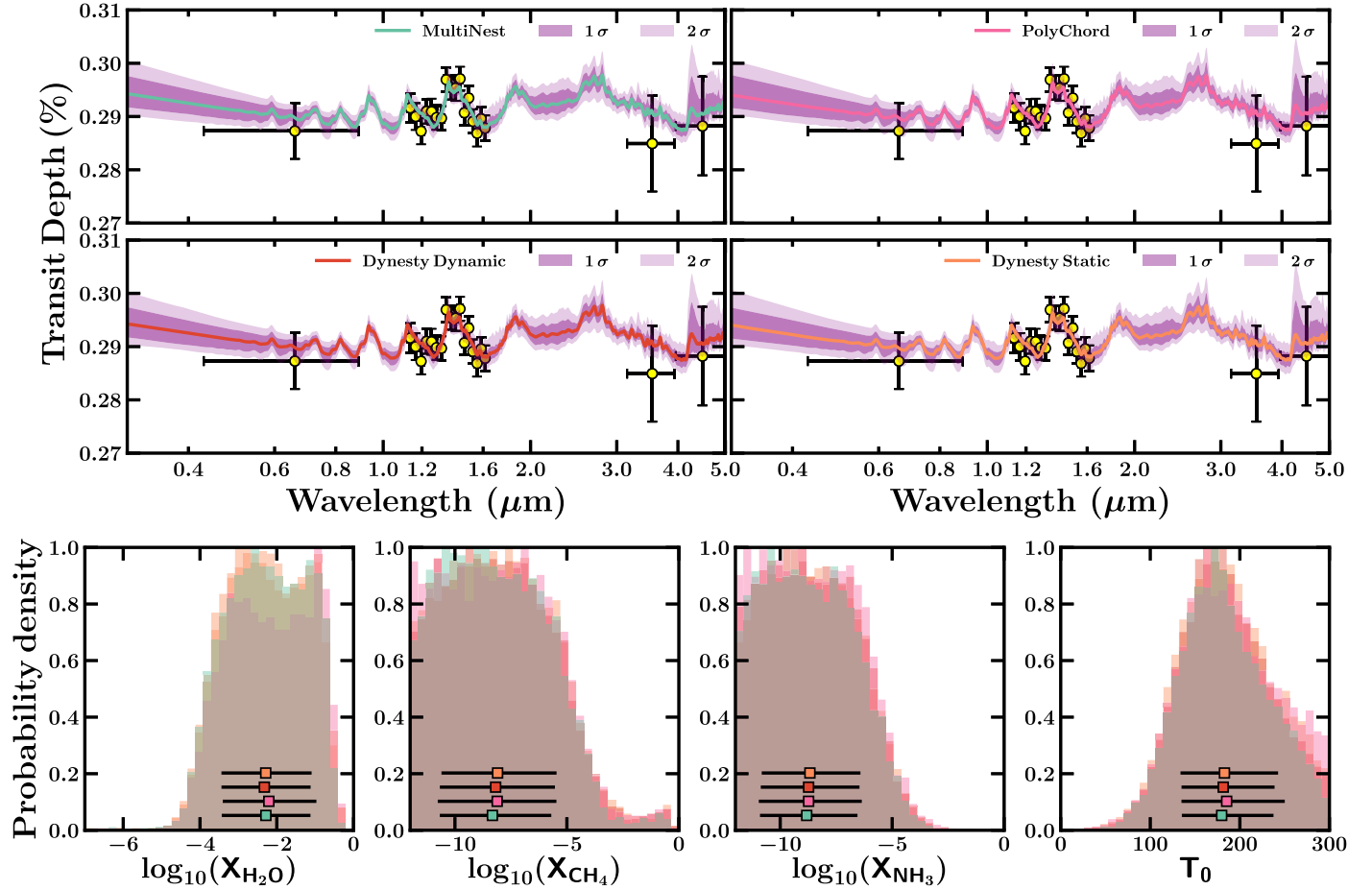
For Dynesty, two separate runs were performed: one using static nested sampling and the other using dynamic nested sampling. For the static Dynesty run, we used 2000 live points. Similarly, the dynamic Dynesty run had an initial number of 2000 live points. The Dynesty runs were set up to generate the new live points using multiellipsoidal decomposition with uniform sampling; this is so that their sampling methods were similar to MultiNest.

Figure 9 presents the retrieved spectra for the data of K2-18b when using each of the different nested sampling algorithms in Aurora, along with the posterior distributions for the parameters of interest when comparing to previous works (e.g., Benneke et al. 2019b; Welbanks et al. 2019; Madhusudhan et al. 2020). The complete list of retrieved parameters is shown in Table 1.

All four nested sampling algorithms included in Aurora retrieve values consistent with each other and with previous works. The retrieved H<sub>2</sub>O abundance for MultiNest is  $\log_{10}(X_{\text{H}_2\text{O}}) = -2.28^{+1.16}_{-1.15}$ , for PolyChord it is  $\log_{10}(X_{\text{H}_2\text{O}}) = -2.21^{+1.24}_{-1.20}$ , for Dynesty in its static run is  $\log_{10}(X_{\text{H}_2\text{O}}) = -2.29^{+1.20}_{-1.15}$ , and for Dynesty in its dynamic run it is  $\log_{10}(X_{\text{H}_2\text{O}}) = -2.32^{+1.20}_{-1.12}$ . These H<sub>2</sub>O abundances are consistent within  $1\sigma$  with the works of Benneke et al. (2019b), Welbanks et al. (2019), and Madhusudhan et al. (2020). In agreement with previous works, the retrieved CH<sub>4</sub> and NH<sub>3</sub> median abundances are lower than expectations from chemical equilibrium. Comparing between the four sampler setups, the posterior distributions obtained for the parameters studied are largely consistent with each other. The retrieved precision on the model parameters is consistent between samplers, with a precision on the H<sub>2</sub>O abundance of  $\sim 1.2$  dex.

Our comparison shows that despite the differences in sampling algorithms, the parameter posterior distributions are mostly independent of the method employed for a problem of this dimensionality (i.e., seven model parameters) and with current data quality. We assess the different run times for each of the nested sampling algorithms by performing these retrievals on one thread of a four-core Intel Core i7-4700MQ CPU at 2.40 GHz. The fastest Nested algorithm under these conditions was MultiNest, followed by the static run of Dynesty ( $\sim 3\times$  longer), the dynamic run of Dynesty ( $\sim 5\times$  longer), and PolyChord ( $\sim 6\times$  longer). These run times are not necessarily representative of the full potential of each sampler. Different parameters in the setup of each algorithm can result in different run times.

In general, we have shown that Aurora includes the capabilities to retrieve the atmospheric properties of an exoplanet spectrum using different nested sampling algorithms. For current data and models, MultiNest is an optimal tool that retrieves the posterior distribution of our parameters efficiently. As data increases and possible degeneracies between the parameters in our models are exacerbated, or as the dimensionality of our problems and models increases, PolyChord and Dynesty are tools that offer alternative ways to



**Figure 9.** Comparison between retrievals with different nested sampling algorithms. Top four panels show the retrieved median model (colored line),  $1\sigma$  and  $2\sigma$  confidence intervals (shaded purple regions), and K2, HST-WFC3, and Spitzer observations (black error bars with yellow markers). Bottom row shows the posterior distributions for  $\text{H}_2\text{O}$ ,  $\text{CH}_4$ ,  $\text{NH}_3$ , and  $T_0$ , the isothermal temperature. The posterior distributions and colored lines are shown in green for MultiNest, pink for PolyChord, red for Dynesty in its dynamic mode, and coral for Dynesty in its static mode. All retrievals are consistent with each other.

**Table 1**  
Retrieved Parameters for the Spectrum of K2-18b Using Different Nested Sampling Algorithms as Explained in Section 3.2

Parameter	MultiNest	PolyChord	Dynesty Static	Dynesty Dynamic
Chemical Species				
$\log_{10}(X_{\text{H}_2\text{O}})$	$-2.28^{+1.16}_{-1.15}$	$-2.21^{+1.24}_{-1.20}$	$-2.29^{+1.20}_{-1.15}$	$-2.32^{+1.20}_{-1.12}$
$\log_{10}(X_{\text{CH}_4})$	$-8.33^{+2.63}_{-2.35}$	$-8.11^{+2.64}_{-2.65}$	$-8.09^{+2.63}_{-2.51}$	$-8.18^{+2.65}_{-2.49}$
$\log_{10}(X_{\text{NH}_3})$	$-8.82^{+2.27}_{-2.09}$	$-8.73^{+2.37}_{-2.22}$	$-8.67^{+2.26}_{-2.15}$	$-8.73^{+2.27}_{-2.14}$
$\log_{10}(X_{\text{CO}})$	$-6.89^{+3.53}_{-3.33}$	$-6.63^{+3.58}_{-3.71}$	$-6.70^{+3.54}_{-3.42}$	$-6.71^{+3.42}_{-3.36}$
$\log_{10}(X_{\text{CO}_2})$	$-7.49^{+3.42}_{-2.97}$	$-7.31^{+3.28}_{-3.15}$	$-7.25^{+3.23}_{-3.12}$	$-7.21^{+3.17}_{-3.13}$
$T_0$ (K)	$179.70^{+57.86}_{-44.47}$	$185.27^{+64.87}_{-50.45}$	$182.81^{+59.85}_{-49.08}$	$181.45^{+57.28}_{-46.58}$
$\log_{10}(P_{\text{ref}})$ (bar)	$-0.86^{+0.37}_{-0.44}$	$-0.89^{+0.39}_{-0.44}$	$-0.86^{+0.37}_{-0.44}$	$-0.86^{+0.37}_{-0.43}$

perform retrievals. The user in Aurora has the freedom to choose the optimal tool for the problem at hand.

### 3.3. Application to Mini Neptune K2-18b

We validate Aurora’s retrieval framework on current observations of K2-18b (Benneke et al. 2019b), including K2, HST-WFC3, and Spitzer, spectrophotometric data. Unlike the previous section, we perform retrievals not assuming a H-rich atmosphere. Using a full Bayesian approach, we test the validity of previous assumptions of a H-rich atmosphere when analyzing the most recent transmission spectrum of this mini

Neptune. Then, we perform retrievals on HST-STIS and JWST-NIRSpec synthetic observations of the same planet and assess the constraints on the chemical abundances and cloud/haze properties.

#### 3.3.1. Case Study: Current Observations of K2-18b

While previous studies have assumed that K2-18b has a H-rich atmosphere (e.g., Benneke et al. 2019b; Tsiaras et al. 2019; Welbanks et al. 2019; Madhusudhan et al. 2020), the robustness of this assumption has remained untested. Here, we apply the new functionality of Aurora, to retrieve atmospheric

**Table 2**  
Retrieved Parameters for Current Observations of K2-18b, Assuming and Not Assuming a H-rich Atmosphere

	Parameter	H-rich Retrieval w/o O <sub>2</sub> , O <sub>3</sub>	H-rich Retrieval w O <sub>2</sub> , O <sub>3</sub>	No H-rich Retrieval w/o O <sub>2</sub> , O <sub>3</sub>	No H-rich Retrieval w O <sub>2</sub> , O <sub>3</sub>
Chemical Species	$\log_{10}(X_{\text{H}_2+\text{He}})$	N/A	N/A	$-0.09^{+0.09}_{-5.71}$	$-0.14^{+0.13}_{-6.64}$
	$\log_{10}(X_{\text{H}_2\text{O}})$	$-2.10^{+1.07}_{-1.20}$	$-2.12^{+1.09}_{-1.24}$	$-1.20^{+1.15}_{-1.81}$	$-1.22^{+1.18}_{-2.03}$
	$\log_{10}(X_{\text{N}_2})$	$-6.29^{+3.38}_{-3.45}$	$-6.59^{+3.49}_{-3.40}$	$-5.53^{+3.24}_{-3.23}$	$-5.66^{+3.28}_{-3.25}$
	$\log_{10}(X_{\text{CH}_4})$	$-8.16^{+2.54}_{-2.41}$	$-8.20^{+2.60}_{-2.31}$	$-6.71^{+3.84}_{-2.57}$	$-6.59^{+3.28}_{-2.79}$
	$\log_{10}(X_{\text{HCN}})$	$-8.07^{+2.53}_{-2.46}$	$-8.03^{+2.51}_{-2.45}$	$-6.78^{+2.89}_{-2.48}$	$-6.81^{+3.05}_{-2.61}$
	$\log_{10}(X_{\text{NH}_3})$	$-8.73^{+2.16}_{-2.03}$	$-8.64^{+2.11}_{-2.08}$	$-7.58^{+2.30}_{-2.02}$	$-7.61^{+2.39}_{-2.10}$
	$\log_{10}(X_{\text{CO}})$	$-6.64^{+3.23}_{-3.34}$	$-6.64^{+3.18}_{-3.24}$	$-5.76^{+3.11}_{-3.10}$	$-5.78^{+3.14}_{-3.13}$
	$\log_{10}(X_{\text{CO}_2})$	$-7.16^{+3.00}_{-2.99}$	$-7.25^{+2.97}_{-2.99}$	$-6.16^{+2.87}_{-2.91}$	$-6.27^{+2.95}_{-2.88}$
	$\log_{10}(X_{\text{O}_2})$	N/A	$-6.33^{+3.40}_{-3.39}$	N/A	$-5.72^{+3.28}_{-3.21}$
$\log_{10}(X_{\text{O}_3})$	N/A	$-7.44^{+2.96}_{-2.84}$	N/A	$-6.26^{+3.06}_{-2.92}$	
P-T Parameters	$T_0$ (K)	$162.12^{+73.73}_{-67.26}$	$166.50^{+72.47}_{-67.41}$	$192.49^{+68.43}_{-77.90}$	$194.86^{+67.88}_{-79.54}$
	$\alpha_1$	$1.32^{+0.43}_{-0.48}$	$1.32^{+0.43}_{-0.48}$	$1.20^{+0.49}_{-0.52}$	$1.17^{+0.50}_{-0.51}$
	$\alpha_2$	$1.20^{+0.50}_{-0.54}$	$1.19^{+0.51}_{-0.54}$	$1.17^{+0.51}_{-0.55}$	$1.16^{+0.51}_{-0.55}$
	$\log_{10}(P_1)$ (bar)	$-1.64^{+1.61}_{-1.65}$	$-1.74^{+1.62}_{-1.63}$	$-1.77^{+1.58}_{-1.61}$	$-1.69^{+1.56}_{-1.64}$
	$\log_{10}(P_2)$ (bar)	$-4.00^{+1.74}_{-1.32}$	$-4.03^{+1.70}_{-1.27}$	$-4.11^{+1.65}_{-1.22}$	$-4.09^{+1.65}_{-1.24}$
	$\log_{10}(P_3)$ (bar)	$0.56^{+0.95}_{-1.33}$	$0.52^{+0.97}_{-1.29}$	$0.50^{+0.97}_{-1.29}$	$0.58^{+0.93}_{-1.28}$
	$\log_{10}(P_{\text{ref}})$ (bar)	$-1.27^{+0.54}_{-0.62}$	$-1.23^{+0.54}_{-0.61}$	$-1.27^{+0.72}_{-0.58}$	$-1.19^{+0.76}_{-0.62}$
	Cloud-Haze Parameters	$\log_{10}(a)$	$0.76^{+4.28}_{-2.98}$	$1.03^{+4.40}_{-3.19}$	$1.55^{+4.54}_{-3.46}$
$\gamma$		$-10.17^{+6.91}_{-6.39}$	$-9.97^{+6.78}_{-6.31}$	$-10.06^{+6.98}_{-6.16}$	$-9.88^{+6.71}_{-6.18}$
$\phi_{\text{hazes}}$		$0.27^{+0.25}_{-0.17}$	$0.27^{+0.24}_{-0.18}$	$0.28^{+0.25}_{-0.18}$	$0.29^{+0.26}_{-0.19}$
$\log_{10}(P_{\text{cloud}})$ (bar)		$-1.57^{+1.99}_{-2.46}$	$-1.38^{+1.89}_{-2.21}$	$-1.23^{+1.88}_{-2.16}$	$-1.19^{+1.88}_{-2.19}$
$\phi_{\text{clouds}}$		$0.30^{+0.24}_{-0.19}$	$0.29^{+0.25}_{-0.18}$	$0.30^{+0.26}_{-0.19}$	$0.29^{+0.25}_{-0.19}$
$\log(\mathcal{Z})$		179.15	179.08	176.84	176.49

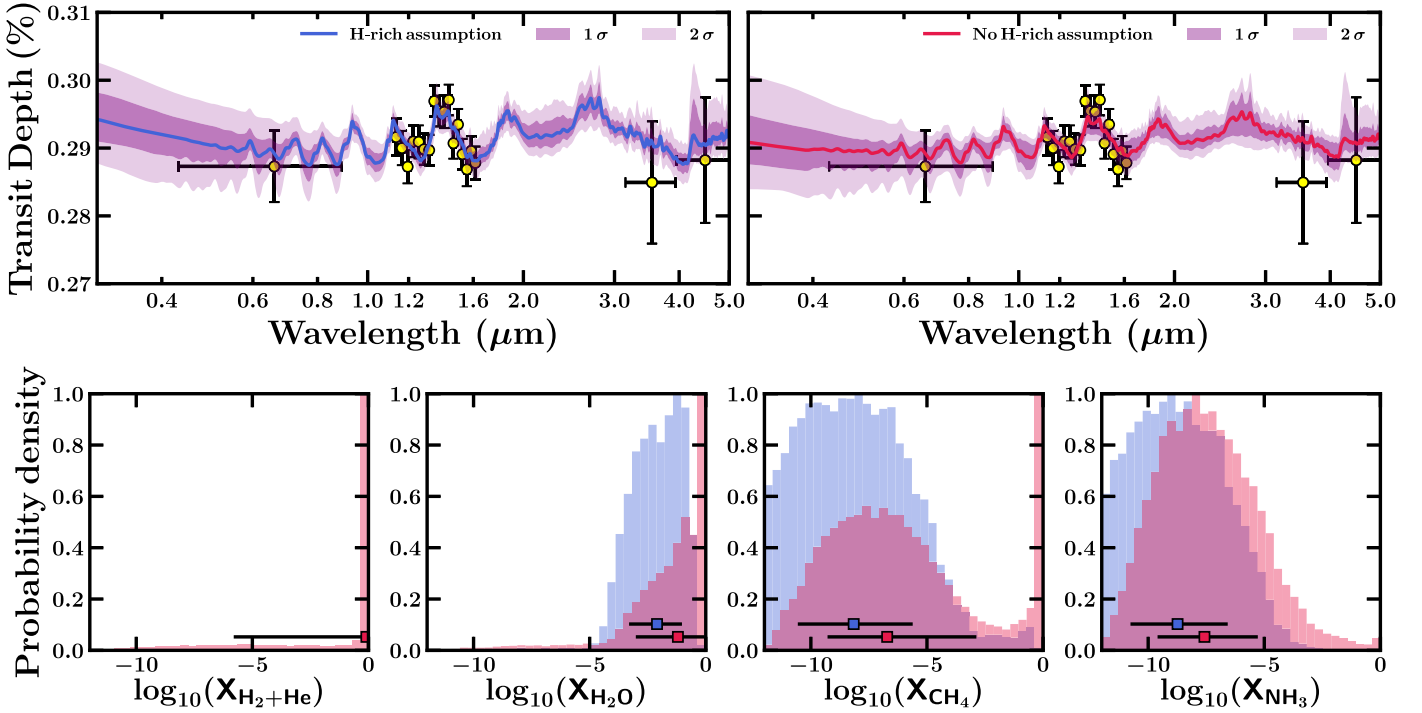
**Note.** All retrievals were computed using MultiNest. N/A means that the parameter in question was not considered in the model by construction.

properties of an exoplanet without assuming a H-rich atmosphere to the broadband transmission spectrum of K2-18b. As discussed in Sections 3.2 and 3.3, we include spectroscopic observations from the HST-WFC3 G141 grism (1.05–1.7  $\mu\text{m}$ ), as well as photometric data in the Spitzer IRAC 3.6 and 4.5  $\mu\text{m}$  bands, and K2 optical bands (0.43–0.89  $\mu\text{m}$ ) from Benneke et al. (2019b). We also redo an analysis assuming the planet has a H-rich atmosphere and with more model considerations than the results in Section 3.2. By performing both sets of retrievals, we can compare their model evidences and assess if the assumption of a H-rich atmosphere is preferred by our retrievals. Furthermore, we expand on previous studies and consider the possibility of O<sub>2</sub> and O<sub>3</sub> absorption for illustration.

The retrievals on the full broadband spectrum of K2-18b consider absorption due to H<sub>2</sub>O, N<sub>2</sub>, CH<sub>4</sub>, HCN, NH<sub>3</sub>, CO, and CO<sub>2</sub>, and H<sub>2</sub>–H<sub>2</sub> and H<sub>2</sub>–He CIA. A second set of retrievals expands the number of absorbers included by considering O<sub>2</sub> and O<sub>3</sub> absorption. Our models employ a full parametric P-T profile, include the presence of H<sub>2</sub> Rayleigh scattering, and follow our new inhomogeneous cloud and haze treatment using two distinct cloud/haze sectors (i.e., model 2 in Section 3.1). We perform retrievals by assuming the atmosphere is H rich as well as relaxing this assumption. The retrieved parameters are shown in Table 2. Figure 10 shows the retrieved spectra and a subset of the retrieved posterior distributions for the highest evidence models assuming a H-rich atmosphere and not assuming a H-rich atmosphere.

We first assess the retrievals when assuming a H-rich atmosphere. The retrieved H<sub>2</sub>O abundances are consistent when considering the possibility of O<sub>2</sub> and O<sub>3</sub> absorption and when not. With retrieved H<sub>2</sub>O abundances of  $\log_{10}(X_{\text{H}_2\text{O}}) = -2.10^{+1.07}_{-1.20}$  (not considering O<sub>2</sub> or O<sub>3</sub>) and  $\log_{10}(X_{\text{H}_2\text{O}}) = -2.12^{+1.09}_{-1.24}$  (considering O<sub>2</sub> and O<sub>3</sub>), the results are in agreement with the estimates of Welbanks et al. (2019), Benneke et al. (2019b), and Madhusudhan et al. (2020). Likewise, both retrievals find a depletion of CH<sub>4</sub> and NH<sub>3</sub> despite the strong absorption of these species in the HST-WFC3 and Spitzer bands, in agreement with retrievals in previous studies. The limited spectral information in the optical wavelengths results in weak constraints on the cloud and haze parameters. The use of a more complex cloud and haze parameterization (i.e., more parameters) relative to previous studies (e.g., Welbanks et al. 2019; Madhusudhan et al. 2020) does not result in better constraints on the presence of clouds and hazes. The derived parameters are largely consistent with a clear atmosphere, i.e., small retrieved cloud cover fractions and haze cover fractions, and cloud-deck top pressures mostly near or below the photosphere.

The retrieved parameters when not assuming a H-rich atmosphere are consistent, within 1 $\sigma$ , with the retrieved parameters when assuming a H-rich atmosphere discussed above. Although consistent, this second approach results in wider and higher abundance estimates for all the chemical species considered. The retrieved H<sub>2</sub>O abundances have median values almost 1 dex higher than those obtained when



**Figure 10.** Retrieval of current K2-18b observations with and without the assumption of a H-rich atmosphere. Top: retrieved model spectra and observations for the two cases. Each panel shows the retrieved median model (blue for H rich and red for non H rich),  $1\sigma$  and  $2\sigma$  confidence intervals (shaded purple regions), and the K2, HST-WFC3, and Spitzer observations (black error bars with yellow markers). These cases do not consider  $O_2/O_3$  as these molecules were not preferred by the data (see Table 2). Bottom row shows the posterior distributions for  $H_2+He$ ,  $H_2O$ ,  $CH_4$ , and  $NH_3$  for the two retrievals.

assuming a H-rich atmosphere. These retrieved abundances are  $\log_{10}(X_{H_2O}) = -1.20^{+1.15}_{-1.81}$  (not considering  $O_2$  or  $O_3$ ) and  $\log_{10}(X_{H_2O}) = -1.22^{+1.18}_{-2.03}$  (considering  $O_2$  and  $O_3$ ).

Despite the higher  $H_2O$  abundance estimates, the retrievals indicate that the main component of the atmosphere is  $H_2$  and He with retrieved abundances of  $\log_{10}(X_{H_2+He}) = -0.09^{+0.09}_{-5.71}$  (not considering  $O_2$  or  $O_3$ ) and  $\log_{10}(X_{H_2+He}) = -0.14^{+0.13}_{-6.64}$  (considering  $O_2$  and  $O_3$ ). The retrieved  $H_2+He$  abundance estimates correspond to a median of  $\sim 72\%$ – $81\%$  and allow for  $H_2+He$  abundances of less than 1% within  $1\sigma$  as shown in Figure 10. Assuming a solar He/ $H_2$  ratio of 0.17 (Asplund et al. 2009), the retrieved median  $H_2+He$  abundance estimate of  $\log_{10}(X_{H_2+He}) = -0.09$  ( $H_2$  and He volume mixing ratio of  $\sim 81\%$ ) indicates a  $\log_{10}(X_{H_2}) = -0.16$  ( $H_2$  volume mixing ratio of  $\sim 69\%$ ) and  $\log_{10}(X_{He}) = -0.93$  (He volume mixing ratio of  $\sim 12\%$ ).

All other chemical abundances are poorly constrained, with most uncertainties greater than 3 dex. Similarly, the cloud and haze parameters remain unconstrained. Overall, retrieving the main gas constituent in the atmosphere of K2-18b using current observations results in a  $H_2$ - and He-rich atmosphere ( $\sim 72\%$ – $81\%$  median volume mixing ratio) with strong  $H_2O$  absorption ( $\sim 6\%$  median volume mixing ratio), consistent with previous retrieval studies (e.g., Benneke et al. 2019b; Welbanks et al. 2019; Madhusudhan et al. 2020).

The highest model evidence corresponds to the retrieval assuming a H-rich atmosphere and not considering absorption due to  $O_2$  or  $O_3$ . Neither approach, assuming a H-rich atmosphere or not, favors the presence of  $O_2$  and  $O_3$  absorption in the atmosphere of K2-18b. In the H-rich approach, the additional parameter space due to considering the presence of these extra two absorbers dilutes the model evidence to a  $1.17\sigma$  equivalent level. Likewise, the non-H-rich approach results in a

decrease in model evidence equivalent to a  $1.54\sigma$  level when considering absorption due to  $O_2$  or  $O_3$ .

Increasing the parameter space to retrieve the abundance of  $H_2$  and He results in a decrease in model evidence. When not considering  $O_2$  and  $O_3$  absorption, the model evidence for the H-rich assumed retrieval is higher at a  $2.68\sigma$  level compared to the retrieval without a priori assumptions on the bulk composition of the atmosphere. Similarly, the H-rich assumption is preferred at a  $2.79\sigma$  level over the non-H-rich assumed model when considering absorption due to  $O_2$  and  $O_3$ . This preference of almost  $3\sigma$  for the H-rich model should not be interpreted as a detection of a H-rich atmosphere on K2-18b but instead must be understood as an indication that the additional parameter space explored by the non-H-rich approach is of lower likelihood. On the other hand, the fact that both retrieval approaches infer a H-rich atmosphere can be interpreted as a demonstration that current data favor a H-rich atmosphere on K2-18b.

### 3.3.2. Future Spectroscopic Observations: K2-18b

In order to investigate the abundance constraints that may be possible with future observations, we generate synthetic HST-STIS and JWST-NIRSpec observations of K2-18b based on the retrieved median  $H_2O$  abundance for our highest evidence model in Section 3.3.1. We choose abundances for  $CH_4$  and  $NH_3$  that are  $\sim 1\times$  solar ( $\log_{10}(X_{CH_4}) = -3.3$ ,  $\log_{10}(X_{NH_3}) = -3.9$ ; e.g., Woitke et al. 2018; Madhusudhan et al. 2020), consistent with their apparent depletion relative to the retrieved  $\sim 10\times$  solar  $H_2O$  abundance (see Section 3.3.1, e.g., Madhusudhan et al. 2020). The input model also includes absorption due to HCN, CO, and  $CO_2$ , with an input nominal abundance of 1 ppm. We generate a model spectrum at a constant spectral resolution of  $R = 5000$  between 0.3 and



5.5  $\mu\text{m}$ . Given that current observations of K2-18b do not place strong constraints on the presence of clouds and hazes, we use input values for the cloud and haze prescription that fall within  $1\sigma$  of the retrieved parameters in Section 3.3.1. These input parameters are a Rayleigh-enhancement factor  $a = 10$ , a slope  $\gamma = -10$ , a gray cloud deck with a top pressure in bar of  $\log_{10}(P_{\text{cloud}}) = -1.6$ , and a 25% cover due to the hazes and 30% cover due to clouds. The input P-T profile is set by the retrieved parameters for the highest evidence model in Section 3.3.1.

The synthetic JWST observations are generated using PANDEXO (Batalha et al. 2017). We generate observations for a transmission spectrum of K2-18b observed with JWST-NIRSpec using its three high-resolution gratings (G140H/F100LP, G235H/F170LP, and G395H/F290LP) in the sub-array SUB2048 mode, i.e., a total of three transits. Further details on the model inputs to PANDEXO are described in Appendix D.1. We also model synthetic HST-STIS observations covering the optical wavelengths from  $\sim 0.3$  to  $1.0 \mu\text{m}$ . Comparing an observed HST-WFC3 transmission spectrum of K2-18b (Benneke et al. 2019b) with that of HD 209458 b (Deming et al. 2013), it is seen that nine transits of K2-18b provide data of comparable quality, in terms of precision per spectral bin, to one transit of HD 209458 b. Because there are no HST-STIS observations of K2-18b available, we derive a synthetic HST-STIS spectrum of K2-18b by scaling the uncertainties and resolution from an observed HST-STIS spectrum of HD 209458 b (Sing et al. 2016) in the same proportion as that of the HST-WFC3 spectra between the two planets. We note that the resulting synthetic observations of K2-18b would require a significantly larger number of transits with HST-STIS than the nine observed with HST-WFC3. Nevertheless, we consider this optimistic scenario as a test case to demonstrate our retrievals. Our synthetic observations have Gaussian-distributed uncertainties with a mean precision of  $\sim 72$  ppm in the STIS G430L band and  $\sim 71$  ppm in the STIS G750L band.

The resulting synthetic observations are shown in Figure 11. The synthetic HST-STIS and JWST-NIRSpec observations provide a spectral range of  $\sim 0.3$ – $5.0 \mu\text{m}$ , encoding information about the presence of clouds, hazes, and absorption due to different species like  $\text{H}_2\text{O}$ ,  $\text{CH}_4$ , and  $\text{NH}_3$ . We perform a retrieval on these observations considering absorption due to  $\text{H}_2\text{O}$ ,  $\text{CH}_4$ ,  $\text{HCN}$ ,  $\text{NH}_3$ ,  $\text{CO}$ ,  $\text{CO}_2$ ,  $\text{N}_2$ ,  $\text{O}_2$ , and  $\text{O}_3$ , and  $\text{H}_2$ – $\text{H}_2$  and  $\text{H}_2$ – $\text{He}$  CIA. We employ the same cloud and haze prescription employed in Section 3.3.1. We do not assume the bulk composition of the atmosphere and retrieve it instead.

Figure 11 shows the retrieved posterior probability distributions for  $\text{H}_2+\text{He}$ , the detected species  $\text{H}_2\text{O}$ ,  $\text{CH}_4$ , and  $\text{NH}_3$ , and some relevant cloud/haze parameters. The full posterior distribution is shown in Appendix E, Figure E1. The bottom half in Figure 11 also shows (in gray) the probability distributions obtained with current K2, HST-WFC3, and Spitzer spectrophotometric observations (first column in Table 2). Comparing both gray and green probability distributions, it is possible to appreciate that abundance estimates will be largely improved using JWST observations. For the assumed synthetic model and data considerations, we retrieve the abundances to be consistent with the input values at  $\log_{10}(X_{\text{H}_2\text{O}}) = -1.66^{+0.50}_{-0.55}$ ,  $\log_{10}(X_{\text{CH}_4}) = -2.94^{+0.35}_{-0.37}$  and  $\log_{10}(X_{\text{NH}_3}) = -3.79^{+0.36}_{-0.40}$ . The corresponding detection significances of the molecules are  $\sim 5\sigma$ ,  $\sim 7\sigma$ , and  $\sim 3\sigma$  for  $\text{H}_2\text{O}$ ,  $\text{CH}_4$ , and  $\text{NH}_3$  respectively. In principle, even better abundance

precisions and detection significances may be attained by combining with other observations (e.g., JWST-MIRI, JWST-NIRISS, HST-WFC3) or considering data of higher resolution. We also note that these precisions and detection significances are dependent on the input model assumptions:  $\sim 10\times$  solar  $\text{H}_2\text{O}$  and  $\sim 1\times$  solar  $\text{CH}_4$  and  $\text{NH}_3$ . Nevertheless, these results demonstrate the capability of Aurora to precisely retrieve the true input values of a mini Neptune like K2-18b.

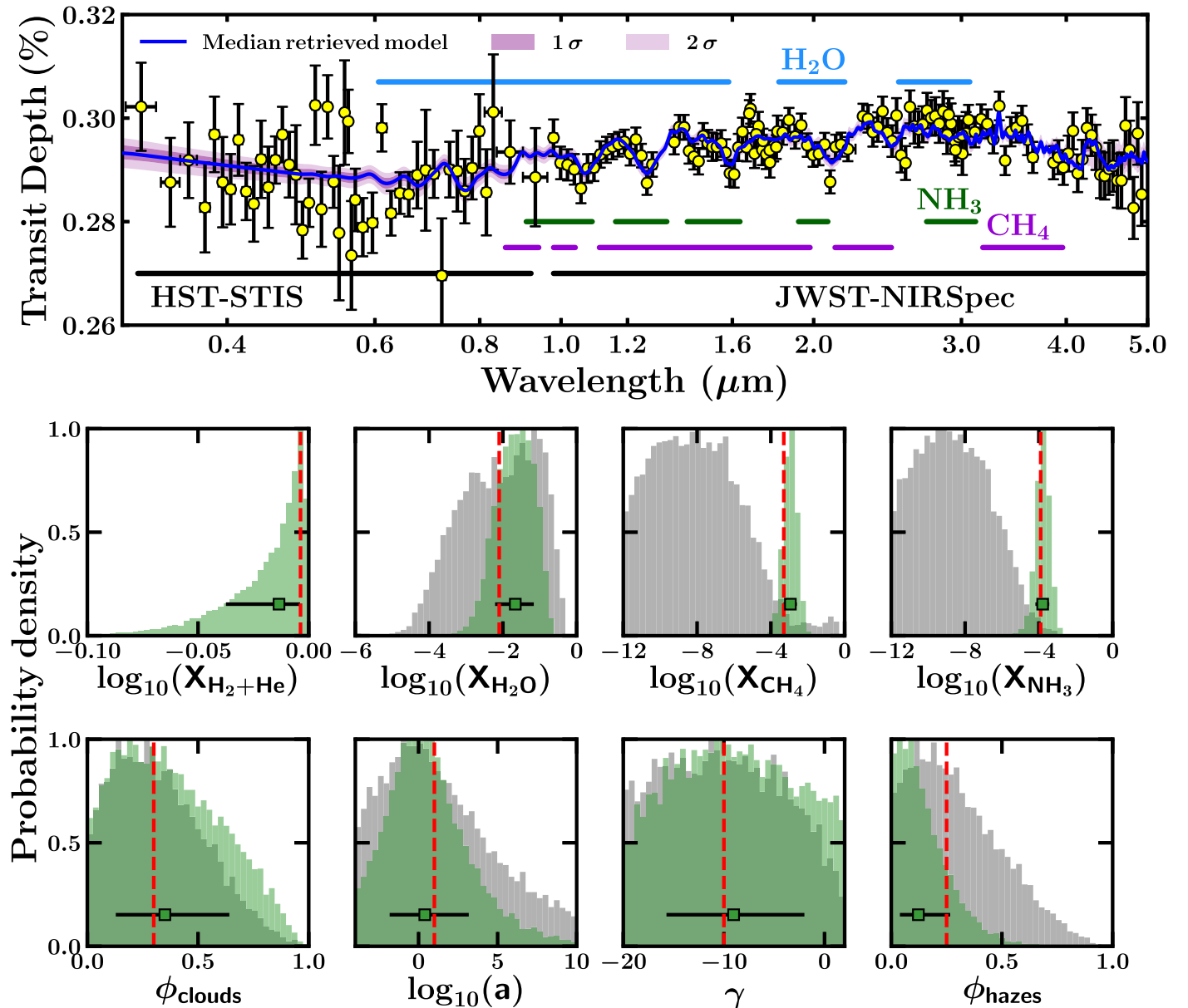
Furthermore, the retrieval on synthetic data demonstrates Aurora’s ability to retrieve the bulk atmospheric composition of a mini Neptune like K2-18b. With a retrieved abundance of  $\log_{10}(X_{\text{H}_2+\text{He}}) = -0.013^{+0.010}_{-0.024}$ , Aurora correctly identifies  $\text{H}_2$  and  $\text{He}$  as the bulk composition of the atmosphere as shown in Figure 11. The retrieved median abundance indicates that  $\text{H}_2$  and  $\text{He}$  account for more than 97% of the atmosphere’s composition, consistent with the input model. Future observations with JWST will make it possible to unequivocally retrieve the bulk gas composition in the atmosphere of K2-18b, improving on present-day estimates derived using current K2, HST-WFC3, and Spitzer observations.

The cloud and haze parameters in the input model are motivated by current constraints on K2-18b using existing data, as discussed above, which indicate a relatively cloud-/haze-free atmosphere. Under these cloud/haze assumptions, the retrieved abundance estimates and their detection significances are not strongly affected when only JWST-NIRSpec observations are considered in our retrievals. The retrieved cloud and haze parameters are mostly unconstrained, consistent with the cloud-/haze-free input model, and similar to constraints obtained with current data (e.g., gray posterior distributions in Figure 11). Even when both HST-STIS and JWST-NIRSpec observations are considered, the cloud/haze constraints are only marginally improved, as shown in Figure 11 and Appendix E, as expected considering the low cloud/haze cover in the input model. Regardless of the cloud/haze constraints, however, the chemical abundances are still derived precisely as discussed above.

In principle, further spectroscopic observations, including those with other JWST instruments like NIRISS and MIRI, could help obtain better constraints than those reported here. At the same time, it could also be important to revisit the model assumptions in present retrievals (e.g., by considering higher-dimensional models, temporal variability, etc.) when confronted with observations of higher quality (e.g., higher-resolution, better signal-to-noise ratio, broader wavelength coverage) expected in the near future. We discuss these implications and the prospect for future works in Section 4.

### 3.4. Application to Rocky Exoplanets

We demonstrate Aurora’s capabilities to identify the composition of atmospheres that are not H rich. We use synthetic JWST-NIRSpec observations of the rocky exoplanet (i.e., terrestrial-size exoplanet) TRAPPIST-1 d ( $R_p = 0.788 R_{\oplus}$ ,  $M_p = 0.388 M_{\oplus}$ ; Gillon et al. 2017; Agol et al. 2021) to validate Aurora’s retrieval capability for H-poor atmospheres. Of all seven TRAPPIST-1 planets, TRAPPIST-1d is the closest to Earth in terms of insolation (see Figure 12). This makes TRAPPIST-1d an appealing candidate for characterization with JWST, especially in the context of planets residing in their habitable zone. This opportunity has been recognized by the JWST Guaranteed Time Observations (GTO) program by



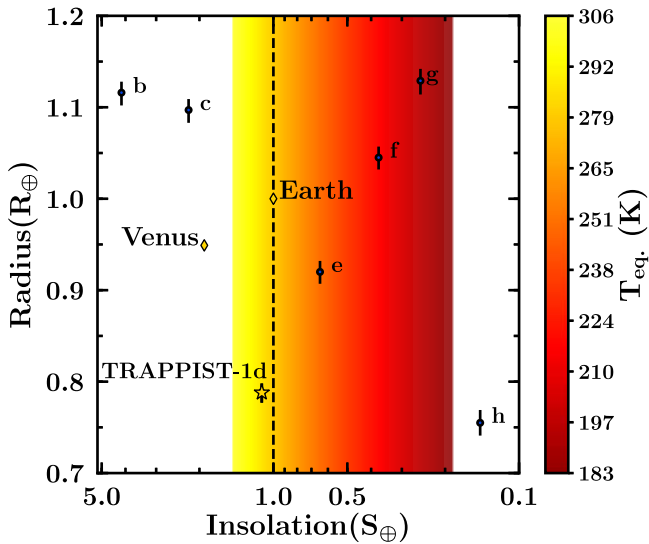
**Figure 11.** Retrieval of synthetic observations of K2-18b. Top: synthetic observations (black error bars with yellow markers) and median retrieved model (dark blue) for STIS G430L and G750L, and NIRSpec G140H, G235H, and G395H. The  $1\sigma$  and  $2\sigma$  confidence intervals (shaded purple areas) are not entirely visible as their span is much smaller than the spread in the data. Horizontal lines at the bottom of the figure show the wavelength coverage of HST-STIS and JWST-NIRSpec. Blue, green, and purple horizontal lines show, respectively, the approximate wavelength regions where  $\text{H}_2\text{O}$ ,  $\text{NH}_3$ , and  $\text{CH}_4$  spectral features are expected. Bottom: retrieved posterior distributions for  $\text{H}_2+\text{He}$ ,  $\text{H}_2\text{O}$ ,  $\text{CH}_4$ , and  $\text{NH}_3$  abundances, and relevant cloud and haze parameters. The green posterior distributions in the foreground correspond to the retrieval using synthetic observations. The vertical red dashed lines indicate the true model input values for this synthetic retrieval. The gray posterior distributions in the background show estimates obtained with existing real observations of K2-18b as discussed in Section 3.3.1, illustrating constraints possible with current data. We note that the red vertical lines are unrelated to the gray posteriors.

planning to observe two transits of the planet using the NIRSpec prism (GTO 1201, PI: David Lafreniere).

Figure 12 shows the planet radius versus stellar insolation for the planets in the TRAPPIST-1 system. When compared to planets in the solar system, TRAPPIST-1 d falls between Venus and Earth in terms of insolation. As such, we consider two possible model configurations for our application of Aurora: a  $\text{CO}_2$ -rich atmosphere (e.g., loosely similar to Venus’ atmosphere) and a  $\text{N}_2$ -rich atmosphere (e.g., loosely similar to Earth’s atmosphere with enhanced  $\text{O}_3$ ). The  $\text{CO}_2$ -rich atmosphere is composed of 98%  $\text{CO}_2$ , 1%  $\text{H}_2\text{O}$ , 11.7 ppm  $\text{H}_2+\text{He}$ , and  $\text{N}_2$  in the remaining percentage ( $\sim 0.99\%$ ). The  $\text{N}_2$ -rich atmosphere is composed of 77.51%  $\text{N}_2$ , 21.38%  $\text{O}_2$ , 1%  $\text{H}_2\text{O}$ ,

0.1%  $\text{CO}_2$ , and 0.01%  $\text{O}_3$ . We note that this  $\text{O}_3$  abundance is  $\sim 10\times$ – $100\times$  higher than present-day Earth’s atmospheric abundance in the stratosphere (e.g., Anderson 1987; Barstow et al. 2016). The atmospheres are modeled to follow an isotherm at 250 K. We considered three cases for each composition: (1) a clear atmosphere, (2) an atmosphere with a gray cloud deck covering the entire planet at a cloud top pressure  $P_{\text{cloud}} = 0.1$  bar, and (3) a gray cloud deck covering the entire planet at a cloud top pressure  $P_{\text{cloud}} = 1$  mbar. The model spectra are generated at a constant resolution of  $R = 5000$  between 0.3 and 5.5  $\mu\text{m}$ .

Besides demonstrating Aurora’s capabilities, we explore the number of JWST transits required for the spectroscopic



**Figure 12.** Planet radius vs. stellar insolation for the TRAPPIST-1 system. The seven TRAPPIST-1 planets (values from Agol et al. 2021) are shown alongside Earth and Venus (yellow diamonds). The shaded region represents the optimistic habitable zone for an M dwarf at the temperature of TRAPPIST-1 (Kopparapu 2013). TRAPPIST-1 d (gold star) is at the inner edge of the optimistic habitable zone of TRAPPIST-1 and is the closest to Earth in terms of insolation. Equilibrium temperature calculated assuming full redistribution and zero albedo.

observations to provide chemical detections and abundance constraints of TRAPPIST-1 d’s atmosphere. We consider observations from 1, 2, 3, 5, and 10 JWST transits. We consider 2 transits motivated by the upcoming GTO 1201 program, and an upper limit at 10 transits based on estimates for characterizing the TRAPPIST-1 system from Batalha et al. (2018; see Section 4.2). We empirically find that 10 transits with JWST-NIRSpec prism can provide chemical constraints of  $\sim 1$  dex or better and robust detections ( $\gtrsim 3\sigma$ ) for multiple chemical species in both the CO<sub>2</sub>-rich and N<sub>2</sub>-rich model atmospheres. The synthetic observations are generated using PANDEXO (Batalha et al. 2017) for the NIRSpec prism using subarray SUB512. The synthetic observations include Gaussian white noise. Additional details are described in Appendix D.2.

The synthetic observations are generated using models that consider CO<sub>2</sub>-CO<sub>2</sub> and N<sub>2</sub>-N<sub>2</sub> CIA, as well as Rayleigh-scattering due to N<sub>2</sub>, CO<sub>2</sub>, and H<sub>2</sub>O. The models in the retrievals do not assume a H-rich atmosphere and consider absorption due to H<sub>2</sub>O, CO<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub>, and O<sub>3</sub>. Including the effects of H<sub>2</sub>-H<sub>2</sub> and H<sub>2</sub>-He CIA, and following the description for non-H-rich atmospheres in Section 2.2, the retrieval considers a total of six chemical components: H<sub>2</sub>+He, H<sub>2</sub>O, CO<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub>, and O<sub>3</sub>. The model also considers one parameter for an isothermal temperature profile and one parameter for  $P_{\text{ref}}$ . We do not consider the full cloud and haze prescriptions presented in this work due to the lack of observations in the optical required to robustly constrain the presence of hazes. However, we consider the presence of inhomogeneous clouds using two parameters: one for the cloud cover ( $\phi_{\text{clouds}}$ ) and one for the cloud top pressure ( $\log_{10}(P_{\text{cloud}})$ ). In summary, the retrieval model has a total of 10 parameters: 6 parameters for the chemical components, 1 parameter for the isothermal temperature, 1 parameter for the reference pressure, and 2 parameters for the presence of inhomogeneous clouds.

We begin by analyzing the results from our exploration of the number of transits required to characterize TRAPPIST-1 d’s atmosphere. Figure 13 shows a summary of the chemical detections for the various numbers of transits considered. We perform this exploration following a conservative approach in which any model preference  $< 2\sigma$  does not constitute a detection (red squares), model preferences  $> 2\sigma$  and  $< 3\sigma$  are suggestive of a chemical detection (yellow squares), and model preferences  $> 3\sigma$  may be considered detections (green squares). Our search suggests that for a CO<sub>2</sub>-rich clear atmosphere, 10 transits with JWST-NIRSpec will be able to provide detections of CO<sub>2</sub> and H<sub>2</sub>O. Likewise, for an N<sub>2</sub>-rich clear atmosphere, 10 transits would provide detections of H<sub>2</sub>O, and possibly O<sub>3</sub> if present at enhanced levels ( $10\times$ – $100\times$  Earth levels) as assumed in the input model described above. For the N<sub>2</sub>-rich atmosphere, although N<sub>2</sub> is found to be the main atmospheric component, its lack of spectral features makes its robust detection difficult.

However, considering clear atmospheres only results in optimistic estimates that may be revised when considering the presence of clouds and hazes. We consider observing the cloudy cases described above using the same number of transits (right columns of Figure 13). As expected, the presence of a cloud deck mutes several of the spectral features resulting in weaker chemical detections or nondetections.

Next, we present the retrieved constraints using Aurora for the cases with the strongest chemical detections, i.e., 10 transits for a cloud-free model, starting with a clear CO<sub>2</sub>-rich atmosphere. The retrieved chemical abundances of interest and retrieved spectrum are shown in Figure 14 along with the synthetic observations. The retrieved abundances for the species included in the true input model are  $\log_{10}(X_{\text{H}_2+\text{He}}) = -5.53_{-3.02}^{+3.01}$ ,  $\log_{10}(X_{\text{H}_2\text{O}}) = -1.64_{-0.92}^{+0.63}$ ,  $\log_{10}(X_{\text{CO}_2}) = -0.07_{-0.78}^{+0.06}$ , and  $\log_{10}(X_{\text{N}_2}) = -4.27_{-3.76}^{+3.35}$ . The retrieved values are consistent within  $\sim 1\sigma$  of the true model input values. Aurora is capable of accurately distinguishing the main constituent of the modeled CO<sub>2</sub>-rich atmosphere, with the posterior distribution of CO<sub>2</sub> corresponding to high abundances. Furthermore, the precisions on the retrieved H<sub>2</sub>O abundance is  $\lesssim 1$  dex, comparable to current chemical constraints for gas giants (e.g., Welbanks et al. 2019).

Although the input model corresponds to a cloud-free atmosphere, we consider in our retrieval model the possibility of inhomogeneous cloud cover. The cloud parameterization retrieves a cloud cover of  $\phi_{\text{clouds}} = 0.37_{-0.25}^{+0.39}$  and  $\log_{10}(P_{\text{cloud}}) = 0.09_{-3.08}^{+1.22}$ , consistent with a clear atmosphere (e.g., relatively small cloud fraction cover) with the gray cloud deck placed below the expected photosphere (e.g., not muting spectral features) in agreement with the input model. Overall, these results indicate that the chemical characterization of a CO<sub>2</sub>-rich, cloud-free atmosphere is possible with 10 JWST-NIRSpec transits. Under these conditions, Aurora is capable of detecting the main component of the atmosphere (CO<sub>2</sub>) at  $4.7\sigma$  and the trace gas (H<sub>2</sub>O) at  $4.3\sigma$ .

We briefly mention the results from considering the more challenging scenario of a cloudy CO<sub>2</sub>-rich atmosphere. As described above, we consider scenarios with 100% cloudy atmospheres with cloud top pressures of 0.1 bar and 1 mbar. While both cases indicate a CO<sub>2</sub>-rich atmosphere, the muted spectral features result in weaker detections of the main atmospheric constituent (e.g.,  $4.3\sigma$  for 0.1 bar and  $\sim 2\sigma$  for 1 mbar; see Figure 13). Although not resulting in strong

CO <sub>2</sub> rich atmosphere		1 Transit	2 Transits	3 Transits	5 Transits	10 Transits	0.1 bar	1 mbar
		CO <sub>2</sub> 98%	<2 $\sigma$	$\sim$ 2 $\sigma$	2.3 $\sigma$	2.5 $\sigma$	4.7 $\sigma$	4.3 $\sigma$
H <sub>2</sub> O 1%	<2 $\sigma$	<2 $\sigma$	$\sim$ 2 $\sigma$	2.3 $\sigma$	4.3 $\sigma$	<2 $\sigma$	<2 $\sigma$	
N <sub>2</sub> rich atmosphere + enhanced O <sub>3</sub>		1 Transit	2 Transits	3 Transits	5 Transits	10 Transits	0.1 bar	1 mbar
		CO <sub>2</sub> 0.1%	<2 $\sigma$	<2 $\sigma$	<2 $\sigma$	<2 $\sigma$	2.4 $\sigma$	$\sim$ 2 $\sigma$
H <sub>2</sub> O 1%	2.5 $\sigma$	4.3 $\sigma$	4.4 $\sigma$	5.2 $\sigma$	6.1 $\sigma$	4.1 $\sigma$	<2 $\sigma$	
N <sub>2</sub> 77.51%	<2 $\sigma$	<2 $\sigma$	<2 $\sigma$	<2 $\sigma$	$\sim$ 2 $\sigma$	<2 $\sigma$	<2 $\sigma$	
O <sub>3</sub> 100 ppm	<2 $\sigma$	<2 $\sigma$	<2 $\sigma$	<2 $\sigma$	$\geq$ 2 $\sigma$	<2 $\sigma$	<2 $\sigma$	

**Figure 13.** Detection significance for different chemical species in the atmosphere of TRAPPIST-1 d as a function of the observed number of transits with JWST-NIRSpec. Top: CO<sub>2</sub>-rich atmosphere. Bottom: N<sub>2</sub>-rich atmosphere, with enhanced O<sub>3</sub>. Left boxes show the results for a cloud-free atmosphere observed with 1, 2, 3, 5, and 10 transits. Right boxes show the results for observing 10 transits of a fully cloudy atmosphere with a cloud-deck top pressure of 0.1 bar and 1 mbar.

detections, Aurora can correctly identify that the observations correspond to a cloudy atmosphere, with the 1 mbar case retrieving  $\phi_{\text{clouds}} = 0.82^{+0.13}_{-0.25}$  and  $\log_{10}(P_{\text{cloud}}) = -2.80^{+1.25}_{-1.24}$  (i.e., relatively high cloud cover fraction and a top cloud-deck pressure above the expected photosphere). In agreement with other studies (see Section 4.2), our results suggest that robustly characterizing the atmosphere of a cloudy rocky exoplanet will need more than 10 JWST-NIRSpec transits.

Next, we present the results from retrieving the clear N<sub>2</sub>-rich atmosphere with 10 JWST-NIRSpec transits. Aurora retrieves abundances of  $\log_{10}(X_{\text{N}_2}) = -0.0037^{+0.0035}_{-2.2195}$ ,  $\log_{10}(X_{\text{H}_2\text{O}}) = -3.06^{+2.27}_{-0.81}$ ,  $\log_{10}(X_{\text{CO}_2}) = -4.18^{+1.97}_{-1.23}$ , and  $\log_{10}(X_{\text{O}_3}) = -4.12^{+1.56}_{-0.68}$ , for the species with detection significances  $\geq 2\sigma$ . The retrieved values are consistent within  $1\sigma$  of the input parameter despite the white noise in the observations. The retrieved cloud parameters are  $\phi_{\text{clouds}} = 0.51^{+0.31}_{-0.34}$  and  $\log_{10}(P_{\text{cloud}}) = -0.34^{+1.12}_{-1.86}$ , consistent with a cloud-free atmosphere due to the retrieved cloud-deck top pressure being below the expected photosphere. Figure 15 presents the synthetic observations as well as the retrieved spectrum and the posterior distributions for the chemical species of interest.

Although the retrieval indicates that N<sub>2</sub> is the main component of the atmosphere, the lack of strong spectral features results in a moderate detection of N<sub>2</sub> with 10 JWST-NIRSpec transits. Besides identifying the main component of the atmosphere, Aurora is able to detect H<sub>2</sub>O (6.1 $\sigma$ ) and O<sub>3</sub> ( $\geq 2\sigma$ ) with 10 transits and provide constraints in their abundances to a precision of  $\sim 1.5$  dex. Although optimistic, these estimates present a tantalizing prospect for the detection of possible biosignatures in habitable zone rocky exoplanets. If

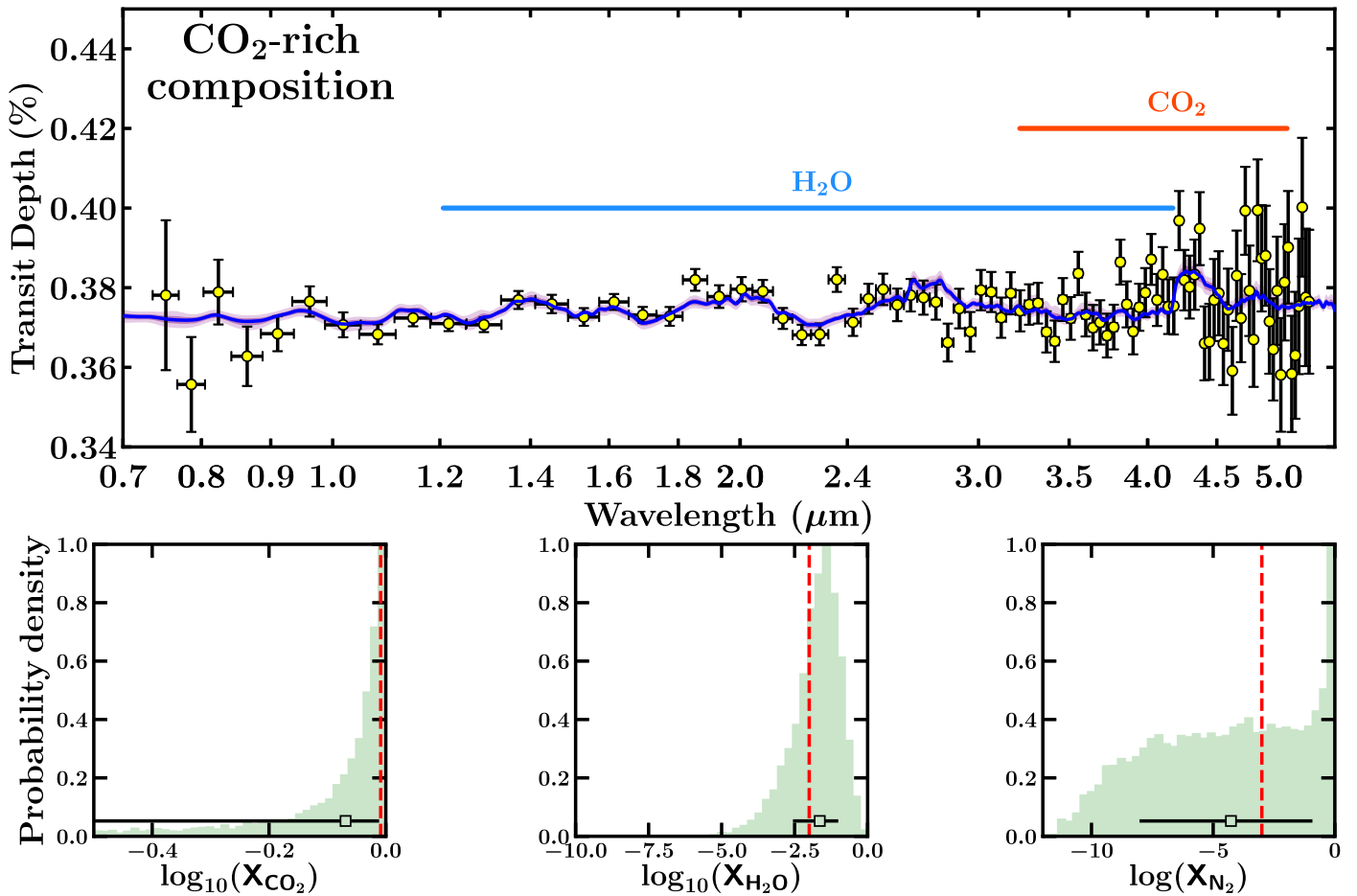
present in  $\sim 10\times$ – $100\times$  higher abundance than present-day Earth stratospheric levels (e.g., Anderson 1987, see Section 4.2), 10 NIRSpec prism transits could provide an initial indication of O<sub>3</sub> in TRAPPIST-1 d.

As performed in the CO<sub>2</sub>-rich case, we investigate the effect of a cloud deck at 0.1 bar and 1 mbar on the estimates above. As shown in Figure 13, the presence of a cloud deck results in weaker or no chemical detections. Nonetheless, the retrieved cloud parameters are mostly consistent with the input cloudy models, with the 1 mbar case retrieving  $\phi_{\text{clouds}} = 0.89^{+0.07}_{-0.12}$  and  $\log_{10}(P_{\text{cloud}}) = -4.62^{+1.07}_{-0.87}$ . In agreement with the CO<sub>2</sub>-rich case, and as expected, this N<sub>2</sub>-rich case suggests that a cloudy atmosphere is more difficult to characterize than a clear atmosphere.

The number of transits with JWST required for the chemical characterization of rocky exoplanets can vary depending on the system parameters, the instrument of choice, and the desired precision on the retrieved atmospheric properties. As such, our result of 10 JWST-NIRSpec transits is specific to the cases considered here. We discuss additional considerations that could revise these results as well as future considerations in Section 4.2.

#### 4. Summary and Discussion

In this work, we introduce Aurora, a next-generation atmospheric retrieval framework for transmission spectra of a wide range of exoplanets: from gas giants with H-rich atmospheres to rocky exoplanets with secondary non-H-rich atmospheres. Aurora retains the capabilities from previous



**Figure 14.** Retrieval of synthetic TRAPPIST-1 d observations for a CO<sub>2</sub>-rich atmosphere. Top: synthetic observations (gold markers, black error bars) and median retrieved model (blue line) for NIRSpec prism. Shaded purple areas indicate 1 $\sigma$  and 2 $\sigma$  confidence intervals. Blue and red horizontal lines show, respectively, the approximate wavelength regions where H<sub>2</sub>O and CO<sub>2</sub> spectral features are expected. Bottom: posterior distributions for CO<sub>2</sub>, H<sub>2</sub>O, and N<sub>2</sub>. Vertical red dashed lines show the true input values.

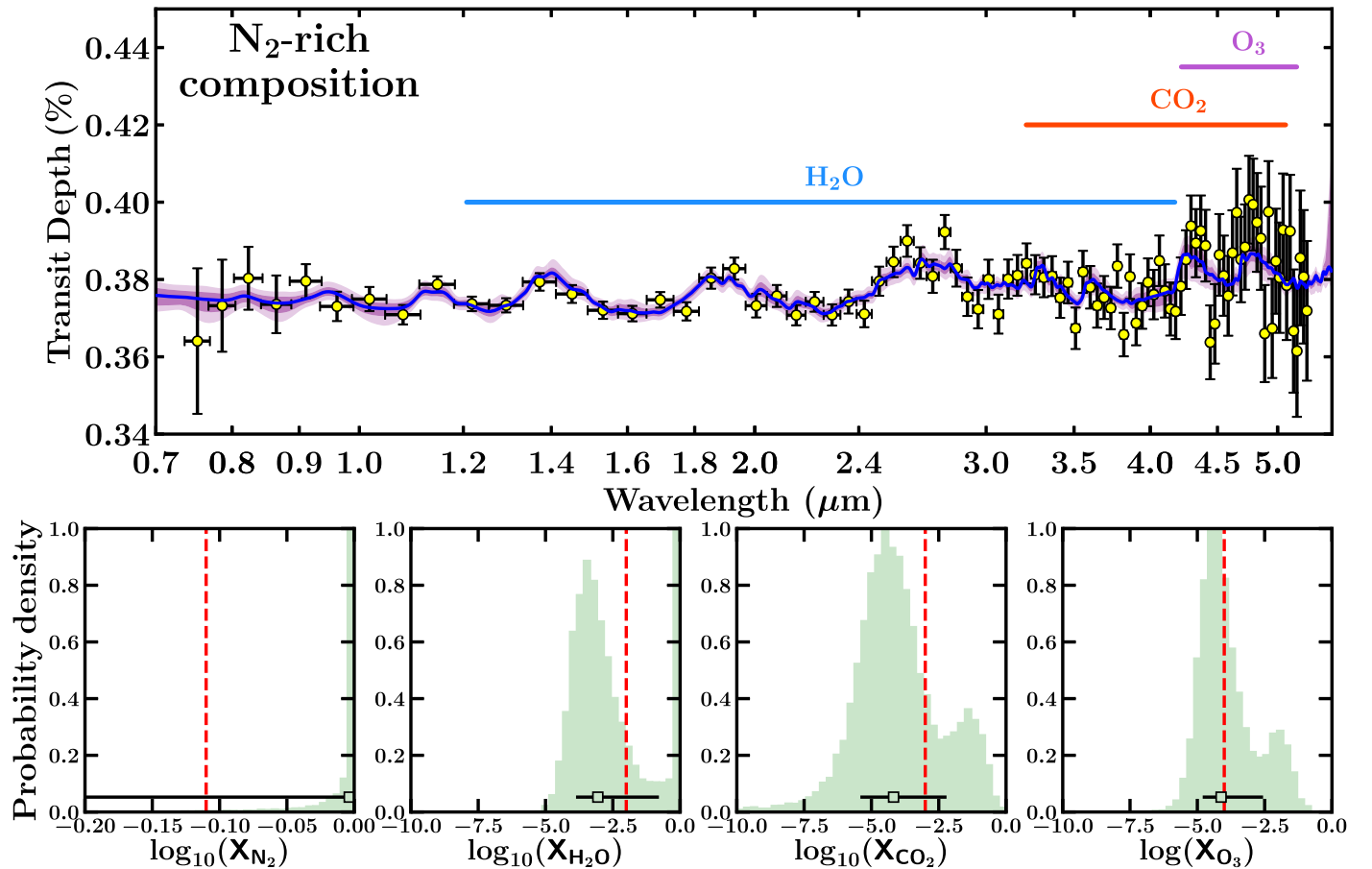
retrieval codes and presents advancements for the analysis of future observations. These key advancements are as follows:

1. Aurora can retrieve the bulk composition of any exoplanet atmosphere without the assumptions of a H-rich atmosphere. The retrieved parameter estimates are robust for H-rich and H-poor atmospheres. We demonstrate this on current and synthetic observations of hot Jupiters, mini Neptunes, and rocky exoplanets.
2. We introduce a new generalized treatment for inhomogeneous clouds and hazes. The new cloud and haze prescription explores a larger parameter space compared to previous treatments. Our new approach mitigates some biases and limitations in previous prescriptions and is robust when assuming a H-rich atmosphere or not.
3. Aurora incorporates in its retrieval framework the next generation of nested sampling algorithms. These are highly adaptable and designed for handling highly degenerate problems and problems of higher dimensionality. This advancement is key in the development of multidimensional retrieval techniques and alleviates some of their computational needs.
4. Aurora has a modular structure designed to evolve with the needs of future spectroscopic observations. The new modular capabilities include:

- (a) Noise modeling capabilities beyond the traditional assumed independently distributed Gaussian errors. These include the ability to retrieve inflation for the standard deviation of observations and consider correlated noise using Gaussian processes.
- (b) Forward models considering the effects of light refraction, forward scattering, and Mie scattering due to condensates.

In this work, we have validated Aurora’s Bayesian retrieval framework using up-to-date existing observations of the hot Jupiter HD 209458 b and the mini Neptune K2-18b. We further validate Aurora’s retrieval framework using HST-STIS and JWST-NIRSpec synthetic observations for K2-18b, and JWST-NIRSpec synthetic observations for the rocky exoplanet TRAPPIST-1 d. Our results highlight four findings:

1. For hot Jupiters, the retrieved parameter estimates are robust against assumptions of a H-rich atmosphere or not. The cloud and haze prescription introduced in this work results in a higher model evidence than previous inhomogeneous cloud and haze prescriptions when applied to current observations of the well-studied hot Jupiter HD 209458 b.



**Figure 15.** Retrieval of synthetic TRAPPIST-1 d observations for a  $\text{N}_2$ -rich atmosphere with enhanced  $\text{O}_3$ . Top: synthetic observations (gold markers, black error bars) and median retrieved model (blue line) for NIRSpect prism. Shaded purple areas indicate  $1\sigma$  and  $2\sigma$  confidence intervals. Blue, red, and purple horizontal lines show, respectively, the approximate wavelength regions where  $\text{H}_2\text{O}$ ,  $\text{CO}_2$ , and  $\text{O}_3$  spectral features are expected. Bottom: posterior distributions for  $\text{N}_2$ ,  $\text{H}_2\text{O}$ ,  $\text{CO}_2$ , and  $\text{O}_3$ . Vertical dashed lines show the true input values.

2. For current observations of mini Neptunes, we have demonstrated that the atmosphere of K2-18b is H rich. Furthermore, the nested sampling algorithms included in Aurora retrieve almost identical parameter estimates. The retrieved properties of K2-18b are consistent with previous results.
3. For future observations of mini Neptunes with JWST, abundance estimates could result in precisions of  $\sim 0.5$  dex or better. Abundance estimates obtained with JWST observations can be robust even in the absence of observations in the optical wavelengths, for relatively low cloud/haze covers as for the case of K2-18b.
4. For future observations of rocky exoplanets, Aurora can robustly identify their dominant atmospheric composition as well as reliably detect and constrain the abundance of trace gases. For example, 10 JWST transits of TRAPPIST-1 d could enable clear detections and abundance constraints of  $\text{H}_2\text{O}$  in a cloud-free,  $\text{N}_2$ -rich or  $\text{CO}_2$ -rich atmosphere. Furthermore, 10 JWST transits could enable initial indications of  $\text{O}_3$  if present at enhanced levels ( $\sim 10\times$ – $100\times$  present-day Earth’s stratospheric abundances) in a cloud-free  $\text{N}_2$ -rich atmosphere.

We discuss the implications of our results for the analysis of current and future observations of hot Jupiters, mini Neptunes, and rocky exoplanets.

#### 4.1. Constraining the Composition of Mini Neptunes

Recent spectroscopic observations (e.g., Benneke et al. 2019a, 2019b; Tsiaras et al. 2019) have demonstrated that mini Neptunes are advantageous targets in the search of  $\text{H}_2\text{O}$  vapor and other possible molecular features in low-mass exoplanets. The lack of an analog for this type of planet in our solar system represents a unique opportunity to learn about the diversity of planet configurations and compositions. Straddling the gap between terrestrial planets and ice giants, it is not always clear whether the atmospheres of some of these planets are H rich or not. Similar to our approach in Section 3.3.1, we suggest that the interpretation of future observations of mini Neptunes and possible super-Earths should begin by not assuming a bulk atmospheric composition. One should perform an agnostic retrieval first, and retrieve the main atmospheric constituent. Then, and if the data suggests the planet’s atmosphere is H-rich, a second retrieval assuming an H-rich atmospheric composition could be informative and should be performed. In this context, the two retrieval approaches should be seen as complementary and informative.

Considerations about the presence of clouds and hazes in these mini Neptunes remain to be explored. The possible absence of clouds in the observable atmosphere of temperate planets like K2-18b, as presented here and in agreement with previous studies (Welbanks et al. 2019; Madhusudhan et al. 2020), represents a

surprise when compared to hotter cloudy planets like GJ 1214b (Kreidberg et al. 2014b). On the other hand, the possibility of Mie-scattering clouds as recently argued in the atmosphere of GJ 3470b (Benneke et al. 2019a) could indicate a complex diversity in the presence of condensates in this type of planets.

Future studies could investigate the need for more complex cloud and haze models in retrievals when interpreting observations of mini Neptunes. For instance, while Benneke et al. (2019a) develop and implement a new Mie-scattering cloud parameterization for atmospheric retrievals in order to explain the observations of GJ 3470b, Welbanks et al. (2019) explain the same observations without invoking Mie scattering and implementing previous prescriptions for inhomogeneous clouds and hazes. Investigating the performance of those cloud prescriptions and the one introduced here could elucidate whether the apparent drop in transit depth in the spectrophotometric observations of GJ 3470b indeed requires invoking Mie-scattering particles. Furthermore, incorporating the Mie-scattering module available in Aurora into its retrieval framework could provide further insights into the atmospheric nature of this and other planets.

Likewise, future studies may investigate how the limitations of previous inhomogeneous cloud and haze prescriptions unveiled in this work affect recent studies that investigate the influence of cloud model choices on retrieval solutions (e.g., Barstow 2020). A full discussion of the different degeneracies and biases in different prescriptions for the clouds and hazes is beyond the scope of this work and reserved for future investigations.

Nonetheless, in order to robustly constrain the presence and properties of clouds and hazes, spectroscopic observations in optical wavelengths are essential (e.g., Line & Parmentier 2016; Welbanks & Madhusudhan 2019). While the results in Section 3.3.2 suggest that state-of-the-art spectroscopic observations in the optical with HST-STIS may not be precise enough to robustly constrain the properties of cloud and hazes for the marginal cloud/haze cover in our K2-18b test case, achievable constraints with HST-STIS for instances with enhanced cloud and haze cover and for other mini Neptunes remain to be explored. Future studies could also explore constraints on the properties of clouds and hazes using observations from multiple instruments on JWST, HST, and ground-based facilities.

#### 4.2. Constraining the Composition of Rocky Exoplanets

Constraining the chemical composition of rocky exoplanets with heavy mean molecular weight atmospheres needs dedicated observational efforts. Exploratory studies using synthetic observations, as performed here and other studies discussed below, can inform the requirements for future observational campaigns. On the number of transits required for the characterization of a TRAPPIST-1 d-like planet, our results are broadly consistent with previous studies of the TRAPPIST-1 planets (e.g., Morley et al. 2017; Batalha et al. 2018; Krissansen-Totton et al. 2018; Lustig-Yaeger et al. 2019; Wunderlich et al. 2019).

For instance, Batalha et al. (2018), using models for TRAPPIST-1 f that consider the presence of a gray cloud deck and an information-content-based approach (Batalha & Line 2017), find that the NIRSpec prism could detect and constrain the dominant atmospheric absorber in H<sub>2</sub>O-rich and CO<sub>2</sub>-rich atmospheres of rocky exoplanets by the 10th transit to

uncertainties smaller than 0.5 dex. Batalha et al. (2018) argue that if the dominant absorber has not been observed by the 10th transit it is unlikely that more transits could provide more information. Naturally, our suggestion of characterizing a rocky exoplanet using 10 JWST transits is valid only if a featureless spectrum has been ruled out using the first few transits.

If not pursuing a robust chemical characterization, a fewer number of transits could help identify spectroscopic features and reject a featureless spectrum. Our results broadly agree with the study of Lustig-Yaeger et al. (2019), who find that two NIRSpec prism transits are enough to rule out a featureless spectrum for TRAPPIST-1 d, although they use a signal-to-noise metric and we use a Bayesian detection significance metric. Additionally, Lustig-Yaeger et al. (2019) employ atmospheric models from Lincowski et al. (2018) considering self-consistent atmospheric compositions and find that CO<sub>2</sub> could be weakly detected in TRAPPIST-1 d using one transit of the JWST-NIRSpec prism in a variety of O<sub>2</sub>- and CO<sub>2</sub>-rich atmospheres. While based on simpler atmospheric models and limited to the CO<sub>2</sub>-rich composition, our results suggest that for a cloud-free atmosphere, two transits of JWST-NIRSpec prism may suffice to provide initial detections ( $\sim 2\sigma$ ) of the main atmospheric component in TRAPPIST-1 d.

Nevertheless, our results are limited to the specific model considerations we have investigated. Modifying our model assumptions of a cloud-free atmosphere, a limited number of absorbers, lack of stellar contamination, isothermal temperature profile, among others, could result in a larger number of transits required for the desired atmospheric constraints. For instance, using more complex general circulation models, Komacek et al. (2020) demonstrate that  $\sim 10$  NIRSpec prism transits would be required to detect H<sub>2</sub>O vapor in the atmosphere of a terrestrial-size exoplanet orbiting a late-type M dwarf when ignoring the effect of clouds and using a similar signal-to-noise metric to Lustig-Yaeger et al. (2019). This result is broadly consistent with our estimate and that of other studies (e.g., Batalha et al. 2018). However, when the effect of clouds is considered, Komacek et al. (2020) find that 63 or more transits are required to detect water. Their results, and our exploration of cloudy models in CO<sub>2</sub>-rich and N<sub>2</sub>-rich atmospheres, suggest that the presence of clouds may significantly increase the number of transits required to detect water features.

Similarly, considering the effect of stellar contamination could significantly affect our interpretations. Recently, Rackham et al. (2018) argue that the stellar contamination impact in the transmission spectra of the TRAPPIST-1 planets can be comparable to or larger than the signal produced by an atmospheric feature. In that case, not accounting for stellar contamination could result in a false positive and be a limiting factor in obtaining reliable abundance constraints. Future studies could investigate the effect of stellar contamination in retrievals (e.g., as in Pinhas et al. 2018) for super-Earths/mini Neptunes using the stellar heterogeneity module included in Aurora and revisit our reported estimates.

Furthermore, considering more complex atmospheric compositions with multiple absorbers as investigated by Morley et al. (2017) could better inform our estimates. In their study, Morley et al. (2017) use radiative-convective models of the TRAPPIST-1 planets assuming Earth-like, Venus-like, and Titan-like atmospheres to determine the number of NIRSpec/NIRISS transit observations required to rule out a flat spectrum

at  $\sim 5\sigma$  confidence. Their results suggest that as few as 13 transits could rule out a flat spectrum for a Venus-like atmospheric composition on TRAPPIST-1 d. While our results seem to be broadly consistent with those of Morley et al. (2017), albeit more optimistic, the impact of non-isothermal profiles and other chemical compositions on our results remains to be investigated.

Lastly, the number of transits required for characterizing H<sub>2</sub>O and CO<sub>2</sub> may not be representative of the requirements for detecting and characterizing other chemical species, including possible biosignatures. For instance, Barstow & Irwin (2016) investigate the number of transits required to detect O<sub>3</sub> in the atmosphere of TRAPPIST-1 d. Their study assumes atmospheric chemistry identical to Earth's present-day atmosphere and employs an optimal estimation retrieval algorithm with isothermal models with clouds deep in the atmosphere where they do not have a significant effect on the spectrum. Their results suggest that present-day Earth levels of O<sub>3</sub> would be detectable with 30 transits of NIRSpec prism and MIRI Low-Resolution Spectrometer. Our results suggest that 10 transits of TRAPPIST-1 d with NIRSpec prism could provide initial indications of O<sub>3</sub> in a N<sub>2</sub>-rich cloud-free atmosphere if present at enhanced abundances ( $\sim 10\times$ – $100\times$  present-day stratospheric Earth levels, e.g., Anderson 1987; Barstow & Irwin 2016; Barstow et al. 2016). Future studies using Aurora could further investigate the requirements for the detection and robust characterization of O<sub>3</sub> and other possible biosignatures (e.g., Krissansen-Totton et al. 2018; Wunderlich et al. 2019).

The characterization of rocky exoplanets with JWST remains an attractive avenue in the search for atmospheric features in habitable zone planets and the search for possible biosignatures. Although our results indicate that precise abundance constraints will be possible with the upcoming generation of telescopes, several outstanding considerations mentioned above need to be explored. Particularly, the presence of clouds, hazes, and stellar contamination may present a significant hindrance in the characterization of rocky exoplanets. If true, temperate cloud-free sub-Neptunes like K2-18b may be the best targets for atmospheric characterization of low-mass exoplanets.

#### 4.3. On Multidimensional Effects

Modeling the presence of inhomogeneities in the atmospheric properties of a planet requires models beyond one-dimensional considerations. Depending on their degree of inhomogeneity, these irregularities can potentially affect the retrieved atmospheric properties. For instance, in this work, we have explored how inhomogeneities in cloud and haze cover affect the retrieved chemical abundances when assuming or not a H-rich atmosphere. Aurora currently employs a combination of one-dimensional models to capture the multidimensional effect of inhomogeneous cloud and haze cover. Nonetheless, other heterogeneities and their multidimensional nature can also affect the retrieved atmospheric properties. Recent studies have explored possible limitations of one-dimensional retrievals in the context of transmission spectroscopy (e.g., Changeat et al. 2019; MacDonald et al. 2020; Pluriel et al. 2020).

For instance, compositional differences in the atmospheric chemistry of exoplanets are an effect largely expected for ultrahot Jupiters (UHJs) with day-side temperatures  $\gtrsim 2200$  K (e.g., Arcangeli et al. 2018; Parmentier et al. 2018). These highly irradiated, tidally locked planets, can exhibit large

contrasts between the atmospheric temperature of their day side and their night side, which in turn can result in strong variations in their atmospheric composition. Retrievals of UHJs should not assume a homogeneous chemical composition or temperature structure. Aurora's retrieval framework is currently designed for planets without strong temperature inhomogeneities across the terminator (e.g., low or moderately irradiated planets). However, we have designed Aurora with the implementation of multidimensional effects in mind.

Our current retrieval framework can be readily generalized to incorporate multidimensional P-T profiles and nonuniform mixing ratios, just as we have done for clouds and hazes. The inclusion of new nested sampling algorithms, optimized for the treatment of high-dimensional parameter spaces and highly degenerate solutions, aids our retrieval framework in these future developments. Future works can expand the current retrieval framework to consider the effects necessary for the appropriate study of UHJs. Nevertheless, these considerations are not imperative for the planets considered in this study where large compositional/temperature gradients between the day side and night side of a planet are not expected.

#### 4.4. Concluding Remarks

Currently, over 50 transiting exoplanets have been observed with transmission spectroscopy and nearly 20 chemical species have been detected in exoplanetary atmospheres (e.g., Madhusudhan 2019). While observations of hot gas giants with H-rich atmospheres have been the most abundant, advancements in observing facilities (e.g., Gillon et al. 2011, 2017), large observing campaigns (e.g., Kreidberg et al. 2014b; Benneke et al. 2019b), as well as the so-called M-dwarf opportunity (e.g., Charbonneau & Deming 2007; Scalo et al. 2007) have allowed for tantalizing transmission spectra of mini Neptunes and super-Earths. The imminent launch of JWST and the continuous observing efforts with HST and ground facilities promise to reveal many more spectra of transiting exoplanets, including the prospect of spectral features in non-H-rich atmospheres. From hot Jupiters to temperate low-mass exoplanets, their spectra could provide further insights into their formation paths, possible trends in compositions, and maybe even their prospects for habitability. It is in this context that retrieval capabilities like Aurora could play an important role in the accurate interpretation of spectroscopic observations.

L.W. thanks the Gates Cambridge Trust for support toward his doctoral studies. This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation OPP1144. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the published version that might arise from this submission. This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service ([www.csd3.cam.ac.uk](http://www.csd3.cam.ac.uk)), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council ([www.dirac.ac.uk](http://www.dirac.ac.uk)). This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National



Aeronautics and Space Administration under the Exoplanet Exploration Program. L.W. thanks Siddharth Gandhi and Arazi Pinhas for their help with the AURA retrieval code which is the progenitor of Aurora and for various helpful discussions. L.W. thanks Peter McGill for helpful discussions, in particular on Gaussian processes. L.W. thanks Anjali Piette for helpful discussions, in particular on sources of opacity. L.W. thanks Matthew Nixon for helpful discussions and his input generating Figure 12. We thank the anonymous referee for their helpful comments.

*Software:* Astropy (Astropy Collaboration et al. 2018), Celerite (Foreman-Mackey et al. 2017), Dynesty (Speagle 2020), George (Ambikasaran et al. 2015), Matplotlib (Hunter 2007), MultiNest

(Feroz et al. 2009, 2013), Numpy (Harris et al. 2020), Pandexo (Batalha et al. 2017), PyMultiNest (Buchner et al. 2014), PyPolyChord (Handley et al. 2015a, 2015b), SciPy (Virtanen et al. 2020).

## Appendix A Priors Used in This Work

As explained in Section 2.3, the priors for the parameters in this work are mostly standard prescriptions and adopted from previous studies (e.g., Pinhas et al. 2019; Welbanks et al. 2019). Table A1 presents the parameters and priors used in this work.

**Table A1**  
Parameters and Priors Used in This Work

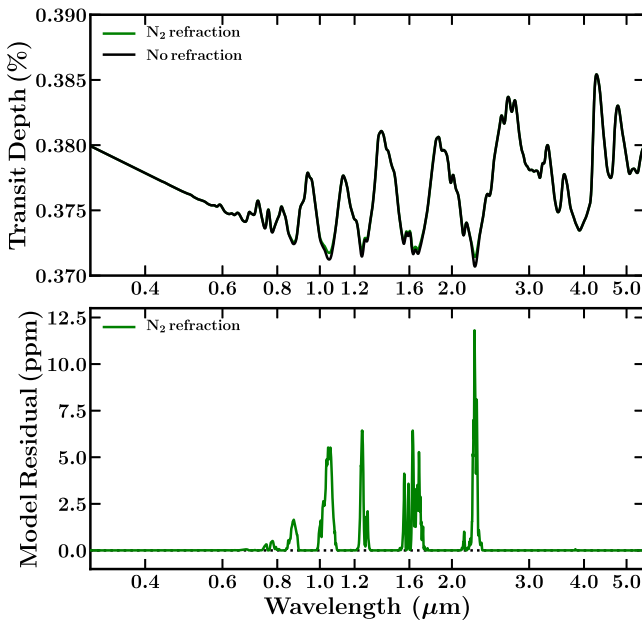
Parameter	Prior Distribution	Prior Range
$X_i$	Log-uniform	$10^{-12}$ – $10^0$ for H-rich retrievals of HD 209485b (Section 3.1) $10^{-12}$ – $10^{-0.3}$ for H-rich retrievals of K2-18b (Section 3.2 and 3.3.1)
$z_i$	Uniform	–4.61–23.03 for non-H-rich retrievals with 6 chemical components (Section 3.4) –3.45–24.18 for non-H-rich retrievals with 8 chemical components (Section 3.3.1) –3.07–24.56 for non-H-rich retrievals with 9 chemical components (Section 3.1) –2.76–24.87 for non-H-rich retrievals with 10 chemical components (Section 3.3.1)
$T_0$	Uniform	0–300 K for retrievals of K2-18b 0–400 K for retrievals of TRAPPIST-1 d 800–1550 K for retrievals of HD 209458 b
$\alpha_{1,2}$	Uniform	$0.02$ – $2.00$ $\text{K}^{-1/2}$
$P_{1,2,\text{cloud,ref}}$	Log-uniform	$10^{-6}$ – $10^2$ bar
$P_3$	Log-uniform	$10^{-2}$ – $10^2$ bar
$a$	Log-uniform	$10^{-4}$ – $10^{10}$
$\gamma$	Uniform	–20–2
$\phi_{\text{clouds}}$		
$\phi_{\text{hazes}}$	Uniform	0–1
$\phi_{\text{clouds+hazes}}$		

**Note.** The priors for the compositional parameters ( $z_i$ ) are reported here to two decimal places and calculated for the purposes of this table. The compositional parameter priors are automatically calculated by Aurora given the condition that they must span the range  $10^{-12}$ – $10^0$  in the  $X_i$  space and that they must satisfy the conditions explained in Section 2.2.

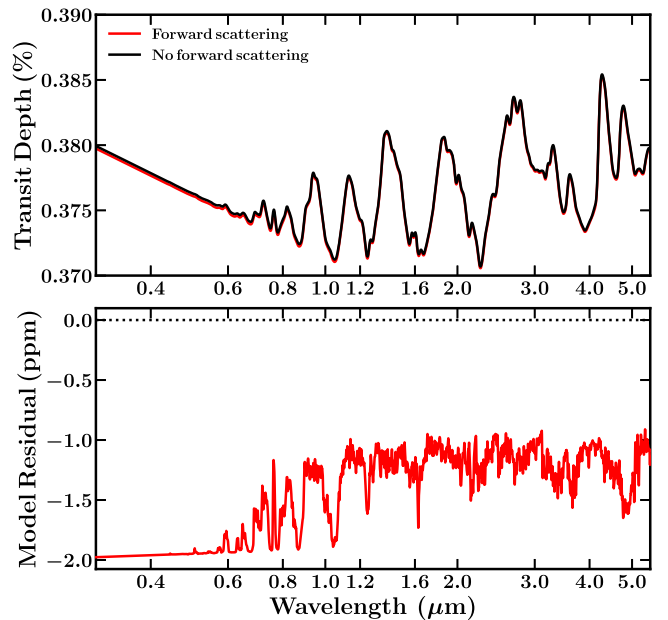
## Appendix B Additional Forward-scattering and Refraction Models for a Rocky Exoplanet

For completion, we consider the effects of wavelength-dependent refraction and forward scattering presented in Section 2.4.3 on the rocky exoplanet TRAPPIST-1 d. We consider the same  $N_2$ -rich atmospheric model used in Section 3.4. Namely, we consider an atmosphere composed of 77.51%  $N_2$ , 21.38%  $O_2$ , 1%  $H_2O$ , 0.1%  $CO_2$ , and 0.01%  $O_3$ . The model spectra are generated at a constant resolution of  $R=5000$  between 0.3 and  $5.5 \mu\text{m}$ , and smoothed for the purposes of Figure B1.

When considering the effects of wavelength-dependent refraction, we specifically include  $N_2$  refraction as our atmospheric model is  $N_2$  rich. The forward model considering the effects of forward scattering uses the same assumptions as in Section 2.4.3 (i.e.,  $g = 0.95$  and  $\tilde{\omega}_0 = 1$ ) and the planet–star orbital separation reported by Agol et al. (2021). Figure B1 shows these forward models and their respective model residuals.



(a) Wavelength-dependent refraction



(b) Forward scattering

**Figure B1.** (a) Effects of wavelength-dependent refraction and (b) forward scattering for a TRAPPIST-1 d-like rocky exoplanet. The top panels show the forward model without these effects considered (black), the model with  $N_2$  refraction (green), and the model considering forward scattering (red). The bottom panels show the model residuals.

These model considerations have a stronger effect in this rocky exoplanet test case than in the mini-Neptune test case presented in Section 2.4.3. The impact of  $N_2$  refraction results in a model difference of  $\sim 12$  ppm. On the other hand, forward scattering results in a model difference  $\lesssim 2$  ppm. The impact of these considerations is  $\sim 3\times$  stronger in this TRAPPIST-1 d case than in the K2-18b case presented in the main text. Nonetheless, these results are dependent on the assumed synthetic model considerations and should not be generalized to other test cases beyond the ones considered here.

## Appendix C Validation of Aurora on HD 209458 b

In Section 3.1, we validate Aurora’s retrieval capabilities on the transmission spectrum of HD 209458 b from Sing et al. (2016). Table C1 presents the retrieved parameters for this case under the different model considerations explained in Section 3.1.

**Table C1**  
Retrieved Parameters for the Spectrum of HD 209458 b Using Different Cloud and Haze Models, as well as Assuming a H-rich Atmosphere or Not as Explained in Section 3.1

Parameter	Model 0 H-rich	Model 0 No H-rich	Model 1 H-rich	Model 1 No H-rich	Model 2 H-rich	Model 2 No H-rich	Model 3 H-rich	Model 3 No H-rich
$\log_{10}(X_{\text{H}_2+\text{He}})$	N/A	$-6.03^{+1.38}_{-2.18} \times 10^{-06}$	N/A	$-1.99^{+0.97}_{-2.84} \times 10^{-05}$	N/A	$-9.52^{+7.49}_{-120.49} \times 10^{-05}$	N/A	$-1.94^{+1.68}_{-34.22} \times 10^{-04}$
$\log_{10}(X_{\text{H}_2\text{O}})$	$-4.98^{+0.13}_{-0.14}$	$-4.99^{+0.13}_{-0.14}$	$-4.53^{+0.34}_{-0.28}$	$-4.48^{+0.29}_{-0.27}$	$-4.06^{+0.73}_{-0.51}$	$-4.00^{+0.68}_{-0.51}$	$-3.89^{+0.78}_{-0.62}$	$-3.81^{+0.73}_{-0.63}$
$\log_{10}(X_{\text{Na}})$	$-6.75^{+0.20}_{-0.21}$	$-6.74^{+0.20}_{-0.22}$	$-5.43^{+0.62}_{-0.49}$	$-5.35^{+0.54}_{-0.43}$	$-4.34^{+1.55}_{-0.99}$	$-4.23^{+1.45}_{-0.99}$	$-3.95^{+1.71}_{-1.27}$	$-3.80^{+1.57}_{-1.27}$
$\log_{10}(X_{\text{K}})$	$-7.94^{+0.24}_{-0.25}$	$-7.92^{+0.23}_{-0.26}$	$-6.96^{+0.63}_{-0.51}$	$-6.90^{+0.54}_{-0.45}$	$-6.04^{+1.44}_{-0.96}$	$-5.96^{+1.40}_{-0.93}$	$-5.73^{+1.65}_{-1.25}$	$-5.58^{+1.51}_{-1.18}$
$\log_{10}(X_{\text{CH}_4})$	$-6.25^{+0.58}_{-3.44}$	$-6.13^{+0.47}_{-2.45}$	$-8.62^{+2.11}_{-2.23}$	$-8.09^{+1.62}_{-1.84}$	$-8.28^{+2.28}_{-2.44}$	$-7.71^{+1.83}_{-2.14}$	$-8.19^{+2.33}_{-2.52}$	$-7.56^{+1.88}_{-2.23}$
$\log_{10}(X_{\text{NH}_3})$	$-5.93^{+0.11}_{-0.13}$	$-5.92^{+0.11}_{-0.13}$	$-6.57^{+0.80}_{-3.48}$	$-6.42^{+0.67}_{-2.80}$	$-6.40^{+1.10}_{-3.59}$	$-6.18^{+0.96}_{-2.93}$	$-6.77^{+1.50}_{-3.48}$	$-6.33^{+1.17}_{-2.95}$
$\log_{10}(X_{\text{HCN}})$	$-9.83^{+1.97}_{-1.47}$	$-9.02^{+1.64}_{-1.46}$	$-9.44^{+2.26}_{-1.74}$	$-8.76^{+1.71}_{-1.50}$	$-9.37^{+2.37}_{-1.76}$	$-8.50^{+2.00}_{-1.75}$	$-9.25^{+2.44}_{-1.87}$	$-8.39^{+2.12}_{-1.79}$
$\log_{10}(X_{\text{CO}})$	$-7.08^{+1.43}_{-1.65}$	$-6.76^{+1.26}_{-1.43}$	$-6.33^{+1.62}_{-1.85}$	$-6.16^{+1.42}_{-1.49}$	$-5.93^{+1.79}_{-1.98}$	$-5.62^{+1.66}_{-1.76}$	$-5.76^{+1.90}_{-2.02}$	$-5.38^{+1.70}_{-1.87}$
$\log_{10}(X_{\text{CO}_2})$	$-10.19^{+1.47}_{-1.22}$	$-9.42^{+1.17}_{-1.21}$	$-9.94^{+1.64}_{-1.40}$	$-9.22^{+1.23}_{-1.22}$	$-9.69^{+1.80}_{-1.55}$	$-9.02^{+1.51}_{-1.40}$	$-9.65^{+1.85}_{-1.61}$	$-8.89^{+1.60}_{-1.49}$
$T_0$ (K)	$810.55^{+14.87}_{-7.71}$	$810.36^{+14.43}_{-7.51}$	$972.42^{+265.03}_{-126.67}$	$984.32^{+225.59}_{-125.41}$	$1064.53^{+283.25}_{-195.12}$	$1062.14^{+282.43}_{-192.28}$	$1081.01^{+277.22}_{-202.50}$	$1076.33^{+267.44}_{-197.43}$
$\alpha_1$	$1.74^{+0.19}_{-0.32}$	$1.74^{+0.18}_{-0.32}$	$1.17^{+0.54}_{-0.52}$	$1.17^{+0.49}_{-0.46}$	$1.10^{+0.59}_{-0.53}$	$1.11^{+0.58}_{-0.53}$	$1.04^{+0.62}_{-0.54}$	$1.06^{+0.61}_{-0.54}$
$\alpha_2$	$1.48^{+0.37}_{-0.65}$	$1.47^{+0.37}_{-0.63}$	$0.97^{+0.68}_{-0.58}$	$1.03^{+0.57}_{-0.54}$	$1.00^{+0.65}_{-0.59}$	$1.01^{+0.64}_{-0.59}$	$0.99^{+0.67}_{-0.59}$	$0.99^{+0.65}_{-0.59}$
$\log_{10}(P_1)$ (bar)	$-1.78^{+1.93}_{-1.94}$	$-1.78^{+1.96}_{-1.88}$	$-1.61^{+1.68}_{-1.72}$	$-1.51^{+1.46}_{-1.51}$	$-1.56^{+1.66}_{-1.76}$	$-1.57^{+1.66}_{-1.74}$	$-1.61^{+1.67}_{-1.77}$	$-1.62^{+1.68}_{-1.73}$
$\log_{10}(P_2)$ (bar)	$-3.89^{+1.83}_{-1.40}$	$-3.90^{+1.85}_{-1.35}$	$-4.13^{+1.92}_{-1.29}$	$-4.01^{+1.60}_{-1.26}$	$-4.07^{+1.92}_{-1.34}$	$-4.11^{+1.91}_{-1.28}$	$-4.14^{+1.92}_{-1.30}$	$-4.09^{+1.84}_{-1.31}$
$\log_{10}(P_3)$ (bar)	$0.52^{+1.06}_{-1.67}$	$0.46^{+1.09}_{-1.66}$	$0.62^{+0.96}_{-1.32}$	$0.67^{+0.84}_{-1.11}$	$0.64^{+0.95}_{-1.37}$	$0.62^{+0.94}_{-1.33}$	$0.59^{+0.97}_{-1.34}$	$0.60^{+0.95}_{-1.34}$
$\log_{10}(P_{\text{ref}})$ (bar)	$-0.71^{+0.06}_{-0.06}$	$-0.72^{+0.06}_{-0.05}$	$-2.64^{+0.66}_{-0.63}$	$-2.70^{+0.55}_{-0.56}$	$-3.15^{+0.79}_{-0.73}$	$-3.18^{+0.76}_{-0.71}$	$-3.33^{+0.79}_{-0.64}$	$-3.37^{+0.80}_{-0.63}$
$\log_{10}(a)$	N/A	N/A	$4.35^{+0.71}_{-1.01}$	$4.49^{+0.56}_{-0.65}$	$2.88^{+0.91}_{-0.85}$	$2.96^{+0.85}_{-0.85}$	$3.28^{+1.01}_{-1.13}$	$3.34^{+0.92}_{-1.07}$
$\gamma$	N/A	N/A	$-14.04^{+4.53}_{-3.94}$	$-14.54^{+3.85}_{-3.31}$	$-16.57^{+3.15}_{-2.37}$	$-16.60^{+2.93}_{-2.29}$	$-16.15^{+3.36}_{-2.60}$	$-16.27^{+3.25}_{-2.50}$
$\log_{10}(P_{\text{cloud}})$ (bar)	N/A	N/A	$-4.41^{+0.80}_{-0.57}$	$-4.47^{+0.56}_{-0.50}$	$-4.60^{+0.99}_{-0.91}$	$-4.65^{+0.97}_{-0.86}$	$-4.72^{+0.99}_{-0.84}$	$-4.74^{+0.99}_{-0.82}$
$\phi_{\text{clouds}}$	N/A	N/A	N/A	N/A	$0.60^{+0.07}_{-0.11}$	$0.59^{+0.07}_{-0.11}$	$0.34^{+0.18}_{-0.20}$	$0.35^{+0.17}_{-0.20}$
$\phi_{\text{hazes}}$	N/A	N/A	N/A	N/A	$0.30^{+0.09}_{-0.07}$	$0.30^{+0.08}_{-0.07}$	$0.27^{+0.09}_{-0.10}$	$0.27^{+0.09}_{-0.09}$
$\phi_{\text{cloud+hazes}}$	N/A	N/A	$0.54^{+0.09}_{-0.12}$	$0.53^{+0.08}_{-0.10}$	N/A	N/A	$0.24^{+0.19}_{-0.16}$	$0.24^{+0.19}_{-0.15}$
$\log(\mathcal{Z})$	949.74	946.68	957.70	951.35	958.14	955.38	958.40	955.74

**Note.** N/A means that the parameter not considered in the model by construction.

## Appendix D Procedure for Simulating JWST Observations

### D.1. PANDEXO Input for K2-18b

The assumed stellar spectrum for K2-18 is interpolated from the Phoenix Stellar Atlas (Husser et al. 2013) by PANDEXO assuming  $T_{\text{eff}} = 3457$  K (Benneke et al. 2019b),  $[\text{Fe}/\text{H}] = 0.12$  (Sarkis et al. 2018),  $\log_{10}(g) = 4.858$  cgs (Crossfield et al. 2016),  $R_{\text{star}} = 0.4445 R_{\odot}$  (Benneke et al. 2019b), and a  $K$ -band magnitude of 8.899. We assume a transit duration of 2.663 hr (Benneke et al. 2017). The in-transit and out-of-transit fluxes are computed using this stellar spectrum and a constant transit-depth model assuming a planet radius of  $2.61 R_{\oplus}$  (Benneke et al. 2019b).

Our simulated observations assume a saturation limit of 80% full well and an extremely optimistic noise floor of 5 ppm. Our assumed noise floor is much smaller than expectations from Greene et al. (2016) and Beichman & Greene (2018). The upcoming JWST Cycle 1 will better inform the existence of a possible systematic noise floor. Nonetheless, given that the noise in our synthetic observations at the native resolution of the gratings is much higher than noise floor expectations, our choice of 5 ppm has little impact on our results. We generate these synthetic observations for one transit and the number of groups per integration suggested by PANDEXO following the *optimize* option. The number of groups per integration is 13, 14, and 25 for G140H, G235H, and G395H respectively. Once the observations are simulated using PANDEXO, we proceed to bin every two pixels to obtain individual spectral resolution elements. Subsequently, we remove any resolution elements that fall within the gaps in the spectral configurations chosen; namely 1.31–1.35  $\mu\text{m}$  for G140H/F100LP, 2.20–2.27  $\mu\text{m}$  for G235H/F170LP and 3.72–3.82  $\mu\text{m}$  for G395H/F290LP. After this, we remove any outlier elements with noise levels above

500 ppm for G140H and G235H, and 400 ppm for G395H. We proceed to bin the observations every 40 resolution elements, remove any binned elements that fall within the gaps in the spectral configurations chosen, and obtain an average resolution of  $R \sim 70$ . The synthetic observations have a mean precision of  $\sim 26$  ppm,  $\sim 30$  ppm, and  $\sim 39$  ppm, for G140H, G235H, and G395H, respectively.

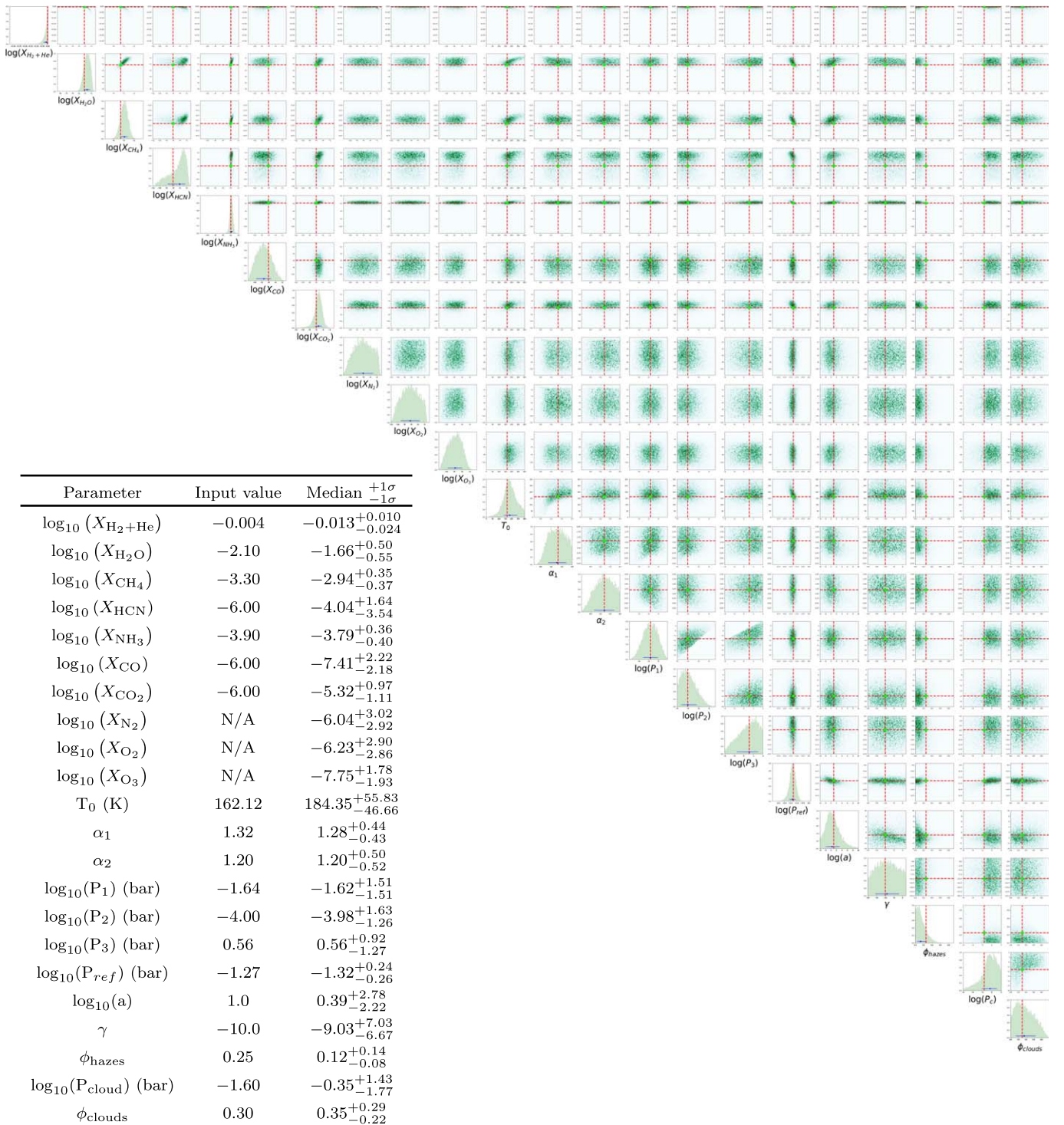
### D.2. PANDEXO Input for TRAPPIST-1 d

The assumed stellar spectrum is interpolated from the Phoenix Stellar Atlas assuming  $T_{\text{eff}} = 2566$  K (Agol et al. 2021),  $[\text{Fe}/\text{H}] = 0.04$ ,  $R_{\text{star}} = 0.1170 R_{\odot}$  (Gillon et al. 2017),  $\log_{10}(g) = 5.2396$  cgs (Agol et al. 2021), and a  $K$ -band magnitude of 10.296. We assume a transit duration of 48.87 minutes (Agol et al. 2021). The in-transit and out-of-transit fluxes are computed using this stellar spectrum and a constant transit-depth model assuming a planet radius of  $0.788 R_{\oplus}$  (Agol et al. 2021).

We maintain the assumption of a saturation limit of 80% full well and an optimistic noise floor of 5 ppm. We generate the observations for 10 NIRSpec prism transits and 2 groups per integration. The instrument configuration for PANDEXO uses aperture S1600A1 and subarray SUB512. We bin every two pixels to obtain individual resolution elements. We remove any elements with precisions larger than 300 ppm, leaving elements  $\gtrsim 0.7$  and  $\lesssim 5.3 \mu\text{m}$ .

## Appendix E Posterior Distributions for the Retrieval of K2-18b Using Synthetic Observations

We present in Figure E1 the posterior distribution from the retrieval on synthetic observations of K2-18b as performed in Section 3.3.2.



**Figure E1.** Posterior distribution for the retrieval of synthetic HST-STIS and JWST-NIRSpec observations of K2-18b (see Section 3.3.2). The retrieval does not assume a H-rich atmosphere. The input parameters for the true model are shown using vertical dashed red lines. Additional examples of posterior distributions are available online at [osf.io/3m94k](https://osf.io/3m94k).

### ORCID iDs

Luis Welbanks  <https://orcid.org/0000-0003-0156-4564>  
 Nikku Madhusudhan  <https://orcid.org/0000-0002-4869-000X>

### References

Adams, D., Gao, P., de Pater, I., & Morley, C. V. 2019, *ApJ*, **874**, 61  
 Agol, E., Dorn, C., Grimm, S. L., et al. 2021, *PSJ*, **2**, 1

Aitchison, J. 1986, *The Statistical Analysis of Compositional Data* (Dordrecht: Springer)  
 Aitchison, J. J., & Egozcue, J. 2005, *MatG*, **37**, 829  
 Allard, N. F., Spiegelman, F., & Kielkopf, J. F. 2016, *A&A*, **589**, A21  
 Allard, N. F., Spiegelman, F., Leininger, T., & Molliere, P. 2019, *A&A*, **628**, A120  
 Al-Refai, A. F., Changeat, Q., Waldmann, I. P., & Tinetti, G. 2019, *arXiv:1912.07759*  
 Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O’Neil, M. 2015, *ITPAM*, **38**, 252

- Anderson, J. G. 1987, *ARPC*, **38**, 489
- Arcangeli, J., Désert, J.-M., Line, M. R., et al. 2018, *ApJL*, **855**, L30
- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, **47**, 481
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Barbary, K. 2105, Nestle Sampling Library, <https://github.com/kbarbary/nestle>, GitHub
- Barber, R. J., Strange, J. K., Hill, C., et al. 2014, *MNRAS*, **437**, 1828
- Barstow, J. K. 2020, *MNRAS*, **497**, 4183
- Barstow, J. K., Aigrain, S., Irwin, P. G. J., et al. 2014, *ApJ*, **786**, 154
- Barstow, J. K., Aigrain, S., Irwin, P. G. J., Kendrew, S., & Fletcher, L. N. 2016, *MNRAS*, **458**, 2657
- Barstow, J. K., Aigrain, S., Irwin, P. G. J., & Sing, D. K. 2017, *ApJ*, **834**, 50
- Barstow, J. K., & Irwin, P. G. J. 2016, *MNRAS*, **461**, L92
- Batalha, N. E., Lewis, N. K., Line, M. R., Valenti, J., & Stevenson, K. 2018, *ApJL*, **856**, L34
- Batalha, N. E., Lewis, T., Fortney, J. J., et al. 2019, *ApJL*, **885**, L25
- Batalha, N. E., & Line, M. R. 2017, *AJ*, **153**, 151
- Batalha, N. E., Mandell, A., Pontoppidan, K., et al. 2017, *PASP*, **129**, 064501
- Beichman, C. A., & Greene, T. P. 2018, in *Observing Exoplanets with the James Webb Space Telescope*, ed. H. J. Deeg & J. A. Belmonte (Berlin: Springer), 85
- Benneke, B., Knutson, H. A., Lothringer, J., et al. 2019a, *NatAs*, **3**, 813
- Benneke, B., & Seager, S. 2012, *ApJ*, **753**, 100
- Benneke, B., & Seager, S. 2013, *ApJ*, **778**, 153
- Benneke, B., Werner, M., Petigura, E., et al. 2017, *ApJ*, **834**, 187
- Benneke, B., Wong, I., Piaulet, C., et al. 2019b, *ApJL*, **887**, L14
- Blecic, J. 2016, arXiv:1604.02692
- Blecic, J., Dobbs-Dixon, I., & Greene, T. 2017, *ApJ*, **848**, 127
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. 1984, *Classification and Regression Trees* (London: Taylor and Francis)
- Broggi, M., & Line, M. R. 2019, *AJ*, **157**, 114
- Bruno, G., Lewis, N. K., Alam, M. K., et al. 2020, *MNRAS*, **491**, 5361
- Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, *A&A*, **564**, A125
- Burningham, B., Marley, M. S., Line, M. R., et al. 2017, *MNRAS*, **470**, 1177
- Caldas, A., Leconte, J., Selsis, F., et al. 2019, *A&A*, **623**, A161
- Changeat, Q., & Al-Refaie, A. F. 2020, *ApJ*, **898**, 155
- Changeat, Q., Edwards, B., Waldmann, I. P., & Tinetti, G. 2019, *ApJ*, **886**, 39
- Changeat, Q., Keyte, L., Waldmann, I. P., & Tinetti, G. 2020, *ApJ*, **896**, 107
- Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, *ApJL*, **529**, L45
- Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002, *ApJ*, **568**, 377
- Charbonneau, D., & Deming, D. 2007, arXiv:0706.1047
- Chayes, F. 1960, *JGR*, **65**, 4185
- Chen, G., Pallé, E., Welbanks, L., et al. 2018, *A&A*, **616**, A145
- Cobb, A. D., Himes, M. D., Soboczenski, F., et al. 2019, *AJ*, **158**, 33
- Colón, K. D., Kreidberg, L., Welbanks, L., et al. 2020, *AJ*, **160**, 280
- Crossfield, I. J. M., Ciardi, D. R., Petigura, E. A., et al. 2016, *ApJS*, **226**, 7
- Cubillos, P. E. 2016, arXiv:1604.01320
- Dalgarno, A., & Williams, D. A. 1962, *ApJ*, **136**, 690
- Damiano, M., & Hu, R. 2020, *AJ*, **159**, 175
- de Wit, J., & Seager, S. 2013, *Sci*, **342**, 1473
- Deming, D., Wilkins, A., McCullough, P., et al. 2013, *ApJ*, **774**, 95
- Evans, T. M., Sing, D. K., Kataria, T., et al. 2017, *Natur*, **548**, 58
- Feng, Y. K., Line, M. R., & Fortney, J. J. 2020, *AJ*, **160**, 137
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, **398**, 1601
- Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, arXiv:1306.2144
- Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, *OJAS*, **2**, 10
- Fisher, C., & Heng, K. 2018, *MNRAS*, **481**, 4698
- Fisher, C., & Heng, K. 2019, *ApJ*, **881**, 25
- Fisher, C., Hoeijmakers, H. J., Kitzmann, D., et al. 2020, *AJ*, **159**, 192
- Foreman-Mackey, D., Agol, E., Angus, R., & Ambikasaran, S. 2017, *AJ*, **154**, 220
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Foreman-Mackey, D., Montet, B. T., Hogg, D. W., et al. 2015, *ApJ*, **806**, 215
- Gandhi, S., Brogi, M., Yurchenko, S. N., et al. 2020, *MNRAS*, **495**, 224
- Gandhi, S., & Madhusudhan, N. 2017, *MNRAS*, **472**, 2334
- Gandhi, S., & Madhusudhan, N. 2018, *MNRAS*, **474**, 271
- Gandhi, S., Madhusudhan, N., Hawker, G., & Piette, A. 2019, *AJ*, **158**, 228
- Gillon, M., Jehin, E., Magain, P., et al. 2011, *EPJWC*, **11**, 06002
- Gillon, M., Triaud, A. H. M. J., Demory, B.-O., et al. 2017, *Natur*, **542**, 456
- Greene, T. P., Line, M. R., Montero, C., et al. 2016, *ApJ*, **817**, 17
- Guzmán-Mesa, A., Kitzmann, D., Fisher, C., et al. 2020, *AJ*, **160**, 15
- Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015a, *MNRAS*, **450**, L61
- Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015b, *MNRAS*, **453**, 4384
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hayes, J. J. C., Kerins, E., Awiphan, S., et al. 2020, *MNRAS*, **494**, 4492
- Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, *ApJL*, **529**, L41
- Heney, L. G., & Greenstein, J. L. 1941, *ApJ*, **93**, 70
- Higson, E., Handley, W., Hobson, M., & Lasenby, A. 2019, *Stat. Comput.*, **29**, 891
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Husser, T. O., Wende-von Berg, S., Dreizler, S., et al. 2013, *A&A*, **553**, A6
- Irwin, P. G. J., Barstow, J. K., Bowles, N. E., et al. 2014, *Icar*, **242**, 172
- Irwin, P. G. J., & Dyudina, U. 2002, *Icar*, **156**, 52
- Irwin, P. G. J., Parmentier, V., Taylor, J., et al. 2020, *MNRAS*, **493**, 106
- Irwin, P. G. J., Teanby, N. A., de Kok, R., et al. 2008, *JQSRT*, **109**, 1136
- Irwin, P. G. J., Weir, A. L., Taylor, F. W., Calcutt, S. B., & Carlson, R. W. 2001, *Icar*, **149**, 397
- Iyer, A. R., & Line, M. R. 2020, *ApJ*, **889**, 78
- Karman, T., Gordon, I. E., van der Avoird, A., et al. 2019, *Icar*, **328**, 160
- Kass, R. E., & Raftery, A. E. 1995, *J. Am. Stat. Assoc.*, **90**, 773
- Kitzmann, D., Heng, K., Oreshenko, M., et al. 2020, *ApJ*, **890**, 174
- Komacek, T. D., Fauchez, T. J., Wolf, E. T., & Abbot, D. S. 2020, *ApJL*, **888**, L20
- Kopparapu, R. K. 2013, *ApJL*, **767**, L8
- Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014a, *ApJL*, **793**, L27
- Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014b, *Natur*, **505**, 69
- Krissansen-Totton, J., Garland, R., Irwin, P., & Catling, D. C. 2018, *AJ*, **156**, 114
- Lacy, B. I., & Burrows, A. S. 2020, *ApJ*, **905**, 131
- Lavie, B., Mendonça, J. M., Mordasini, C., et al. 2017, *AJ*, **154**, 91
- Lavvas, P., Koskinen, T., Steinrueck, M. E., García Muñoz, A., & Showman, A. P. 2019, *ApJ*, **878**, 118
- Lecavelier Des Etangs, A., Pont, F., Vidal-Madjar, A., & Sing, D. 2008, *A&A*, **481**, L83
- Lee, J. M., Fletcher, L. N., & Irwin, P. G. J. 2012, *MNRAS*, **420**, 170
- Lee, J.-M., Heng, K., & Irwin, P. G. J. 2013, *ApJ*, **778**, 97
- Lincowski, A. P., Meadows, V. S., Crisp, D., et al. 2018, *ApJ*, **867**, 76
- Line, M. R., Fortney, J. J., Marley, M. S., & Sarahana, S. 2014a, *ApJ*, **793**, 33
- Line, M. R., Knutson, H., Wolf, A. S., & Yung, Y. L. 2014b, *ApJ*, **783**, 70
- Line, M. R., & Parmentier, V. 2016, *ApJ*, **820**, 78
- 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. E., Marley, M. S., Lewis, N., et al. 2016, *AJ*, **152**, 217
- Lustig-Yaeger, J., Meadows, V. S., & Lincowski, A. P. 2019, *AJ*, **158**, 27
- MacDonald, R. J., Goyal, J. M., & Lewis, N. K. 2020, *ApJL*, **893**, L43
- MacDonald, R. J., & Madhusudhan, N. 2017, *MNRAS*, **469**, 1979
- Madhusudhan, N. 2012, *ApJ*, **758**, 36
- Madhusudhan, N. 2018, *Atmospheric Retrieval of Exoplanets* (Berlin: Springer), 104
- Madhusudhan, N. 2019, *ARA&A*, **57**, 617
- Madhusudhan, N., Crouzet, N., McCullough, P. R., Deming, D., & Hedges, C. 2014, *ApJL*, **791**, L9
- Madhusudhan, N., Nixon, M. C., Welbanks, L., Piette, A. A. A., & Booth, R. A. 2020, *ApJL*, **891**, L7
- Madhusudhan, N., & Seager, S. 2009, *ApJ*, **707**, 24
- Madhusudhan, N., & Seager, S. 2011, *ApJ*, **729**, 41
- Mai, C., & Line, M. R. 2019, *ApJ*, **883**, 144
- Márquez-Neila, P., Fisher, C., Sznitman, R., & Heng, K. 2018, *NatAs*, **2**, 719
- Min, M., Ormel, C. W., Chubb, K., Helling, C., & Kawashima, Y. 2020, *A&A*, **642**, A28
- Mollière, P., Wardenier, J. P., van Boekel, R., et al. 2019, *A&A*, **627**, A67
- Morley, C. V., Kreidberg, L., Rustamkulov, Z., Robinson, T., & Fortney, J. J. 2017, *ApJ*, **850**, 121
- Moses, J. I., Line, M. R., Visscher, C., et al. 2013, *ApJ*, **777**, 34
- Nayak, M., Lupu, R., Marley, M. S., et al. 2017, *PASP*, **129**, 034401
- Nikolov, N., Sing, D. K., Fortney, J. J., et al. 2018, *Natur*, **557**, 526
- Nixon, M. C., & Madhusudhan, N. 2020, *MNRAS*, **496**, 269
- Ohno, K., & Kawashima, Y. 2020, *ApJL*, **895**, L47
- Parmentier, V., Line, M. R., Bean, J. L., et al. 2018, *A&A*, **617**, A110
- Pawlowsky-Glahn, V., & Buccianti, A. 2011, *Compositional Data Analysis* (New York: Wiley)
- Pearson, K. 1897, *RSPS*, **60**, 489
- Piette, A. A. A., & Madhusudhan, N. 2020, *MNRAS*, **497**, 5136
- Pinhas, A., & Madhusudhan, N. 2017, *MNRAS*, **471**, 4355
- Pinhas, A., Madhusudhan, N., Gandhi, S., & MacDonald, R. 2019, *MNRAS*, **482**, 1485

- Pinhas, A., Rackham, B. V., Madhusudhan, N., & Apai, D. 2018, *MNRAS*, **480**, 5314
- Pluriel, W., Zingales, T., Leconte, J., & Parmentier, V. 2020, *A&A*, **636**, A66
- Pont, F., Knutson, H., Gilliland, R. L., Moutou, C., & Charbonneau, D. 2008, *MNRAS*, **385**, 109
- Rackham, B., Espinoza, N., Apai, D., et al. 2017, *ApJ*, **834**, 151
- Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, *ApJ*, **853**, 122
- Rasmussen, C. E., & Williams, C. K. I. 2006, *Gaussian Processes for Machine Learning* (Cambridge, MA: MIT Press)
- Richard, C., Gordon, I. E., Rothman, L. S., et al. 2012, *JQSRT*, **113**, 1276
- Robinson, T. D., Fortney, J. J., & Hubbard, W. B. 2017, *ApJ*, **850**, 128
- Rodgers, C. D. 2000, *Inverse Methods for Atmospheric Sounding: Theory and Practice* (Singapore: World Scientific)
- Rothman, L. S., Gordon, I. E., Barber, R. J., et al. 2010, *JQSRT*, **111**, 2139
- Sarkis, P., Henning, T., Kürster, M., et al. 2018, *AJ*, **155**, 257
- Scalo, J., Kaltenegger, L., Segura, A. G., et al. 2007, *AsBio*, **7**, 85
- Seager, S., & Deming, D. 2010, *ARA&A*, **48**, 631
- Seager, S., & Sasselov, D. D. 2000, *ApJ*, **537**, 916
- Sedaghati, E., Boffin, H. M. J., MacDonald, R. J., et al. 2017, *Natur*, **549**, 238
- Seidel, J. V., Ehrenreich, D., Pino, L., et al. 2020, *A&A*, **633**, A86
- Shardanand, & Prasad Rao, A. D. 1977, NASA Technical Note **TN D-8842**
- Sing, D. K., Fortney, J. J., Nikolov, N., et al. 2016, *Natur*, **529**, 59
- Sivia, D., & Skilling, J. 2006, *Data Analysis: A Bayesian Tutorial* (Oxford: Oxford Univ. Press)
- Skilling, J. 2004, in *AIP Conf. Ser. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, ed. R. Fischer, R. Preuss, & U. V. Toussaint (Melville, NY: AIP), 395
- Skilling, J. 2006, *BayAn*, **1**, 833
- Sneep, M., & Ubachs, W. 2005, *JQSRT*, **92**, 293
- Soboczenski, F., Himes, M. D., O'Beirne, M. D., et al. 2018, arXiv:1811.03390
- Speagle, J. S. 2020, *MNRAS*, **493**, 3132
- Tanner, J. M. 1949, *JAPh*, **2**, 1
- Tegmark, M., Strauss, M. A., Blanton, M. R., et al. 2004, *PhRvD*, **69**, 103501
- Thalman, R., Zarzana, K. J., Tolbert, M. A., & Volkamer, R. 2014, *JQSRT*, **147**, 171
- Trotta, R. 2008, *ConPh*, **49**, 71
- Trotta, R. 2017, arXiv:1701.01467
- Tsiaras, A., Waldmann, I. P., Tinetti, G., Tennyson, J., & Yurchenko, S. N. 2019, *NatAs*, **3**, 1086
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- von Essen, C., Mallonn, M., Welbanks, L., et al. 2019, *A&A*, **622**, A71
- Wakeford, H. R., Sing, D. K., Deming, D., et al. 2018, *AJ*, **155**, 29
- Wakeford, H. R., Sing, D. K., Kataria, T., et al. 2017, *Sci*, **356**, 628
- Waldmann, I. P. 2016, *ApJ*, **820**, 107
- Waldmann, I. P., Rocchetto, M., Tinetti, G., et al. 2015a, *ApJ*, **813**, 13
- Waldmann, I. P., Tinetti, G., Rocchetto, M., et al. 2015b, *ApJ*, **802**, 107
- Welbanks, L., & Madhusudhan, N. 2019, *AJ*, **157**, 206
- Welbanks, L., Madhusudhan, N., Allard, N. F., et al. 2019, *ApJL*, **887**, L20
- Woitke, P., Helling, C., Hunter, G. H., et al. 2018, *A&A*, **614**, A1
- Wunderlich, F., Godolt, M., Grenfell, J. L., et al. 2019, *A&A*, **624**, A49
- Wytenbach, A., Ehrenreich, D., Lovis, C., Udry, S., & Pepe, F. 2015, *A&A*, **577**, A62
- Yurchenko, S. N., Barber, R. J., & Tennyson, J. 2011, *MNRAS*, **413**, 1828
- Yurchenko, S. N., & Tennyson, J. 2014, *MNRAS*, **440**, 1649
- Zalesky, J. A., Line, M. R., Schneider, A. C., & Patience, J. 2019, *ApJ*, **877**, 24
- Zhang, M., Chachan, Y., Kempton, E. M. R., et al. 2020, *ApJ*, **899**, 27
- Zhang, M., Chachan, Y., Kempton, E. M. R., & Knutson, H. A. 2019, *PASP*, **131**, 034501
- Zhang, X. 2020, *RAA*, **20**, 099
- Zingales, T., & Waldmann, I. P. 2018, *AJ*, **156**, 268