The Astronomical Journal, 160:15 (20pp), 2020 July https://doi.org/10.3847/1538-3881/ab9176

Information Content of JWST NIRSpec Transmission Spectra of Warm Neptunes

Andrea Guzmán-Mesa[1], Daniel Kitzmann[1], Chloe Fisher[1,7], Adam J. Burgasser[2,8], H. Jens Hoeijmakers[1,3], Pablo Márquez-Neila[1,4], Simon L. Grimm[1], Avi M. Mandell[5], Raphael Sznitman[4], and Kevin Heng[1,6]
[1] University of Bern, Center for Space and Habitability, Gesellschaftsstrasse 6, CH-3012, Bern, Switzerland; andrea.guzmanmesa@space.unibe.ch, kevin.heng@csh.unibe.ch
[2] Center for Astrophysics and Space Science, University of California San Diego, La Jolla, CA 92093, USA
[3] Observatoire astronomique de l'Université de Genève, 51 chemin des Maillettes, 1290 Versoix, Switzerland
[4] University of Bern, ARTORG Center for Biomedical Engineering, Murtenstrasse 50, CH-3008, Bern, Switzerland
[5] Solar System Exploration Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[6] University of Warwick, Department of Physics, Astronomy & Astrophysics Group, Coventry CV4 7AL, UK
Received 2019 November 8; revised 2020 May 5; accepted 2020 May 6; published 2020 June 12

Abstract

Warm Neptunes offer a rich opportunity for understanding exo-atmospheric chemistry. With the upcoming James Webb Space Telescope (JWST), there is a need to elucidate the balance between investments in telescope time versus scientific yield. We use the supervised machine-learning method of the random forest to perform an information content (IC) analysis on a 11-parameter model of transmission spectra from the various NIRSpec modes. The three bluest medium-resolution NIRSpec modes (0.7–1.27 μm, 0.97–1.84 μm, 1.66–3.07 μm) are insensitive to the presence of CO. The reddest medium-resolution mode (2.87–5.10 μm) is sensitive to all of the molecules assumed in our model: CO, CO_2, CH_4, C_2H_2, H_2O, HCN, and NH_3. It competes effectively with the three bluest modes on the information encoded on cloud abundance and particle size. It is also competitive with the low-resolution prism mode (0.6–5.3 μm) on the inference of every parameter except for the temperature and ammonia abundance. We recommend astronomers to use the reddest medium-resolution NIRSpec mode for studying the atmospheric chemistry of 800–1200 K warm Neptunes; its corresponding high-resolution counterpart offers diminishing returns. We compare our findings to previous JWST IC analyses that favor the blue orders and suggest that the reliance on chemical equilibrium could lead to biased outcomes if this assumption does not apply. A simple, pressure-independent diagnostic for identifying chemical disequilibrium is proposed based on measuring the abundances of H_2O, CO, and CO_2.

Unified Astronomy Thesaurus concepts: Exoplanet atmospheres (486); Exoplanets (498)

1. Introduction

With the much anticipated launch of the James Webb Space Telescope (JWST) in 2021 (and Cycle 1 proposals due in 2020), the exoplanet community is studying the balance between investments of telescope time and scientific yield (Beichman et al. 2014; Barstow et al. 2015; Greene et al. 2016; Batalha & Line 2017; Howe et al. 2017). Both the Guaranteed Time Observations and the Early Release Science programs are designed to gain an understanding of systematics and data reduction strategies (Stevenson et al. 2016; Bean et al. 2018; Kilpatrick et al. 2018) and will provide the first opportunities to obtain JWST transit spectroscopy data over a wide range of infrared wavelengths for many of the best-known transiting exoplanets.

1.1. Motivation I: Anticipated Chemical Diversity of Warm Neptunes

One of the unexpected outcomes of the Kepler mission is that ~1000 K sub-Neptune- to Neptune-sized exoplanets on short-period orbits are common (e.g., Petigura et al. 2013; Crossfield et al. 2016), which we will collectively term "warm Neptunes" in the current study. Their bulk densities indicate the presence of a hydrogen- and/or helium-dominated atmosphere. The Transiting Exoplanet Survey Satellite (TESS) is discovering warm Neptunes orbiting bright stars (e.g., Dragomir et al. 2019; Esposito et al. 2019; Quinn et al. 2019; Trifonov et al. 2019). With no example in our solar system, a deeper understanding of the properties of warm Neptunes is expected to shed light on exoplanet formation processes. The complete chemical inventory of their atmospheres is currently unknown, and it is expected that JWST spectra will allow the exoplanet community to make significant progress on this question.

Across a temperature range of 800–1200 K, warm Neptunes are theoretically predicted to exhibit remarkable chemical diversity with water (H_2O), methane (CH_4), carbon dioxide (CO_2), and carbon monoxide (CO) having a wide range of volume mixing ratios as the elemental abundance of carbon (C/H) and the carbon-to-oxygen ratio (C/O) vary (Moses et al. 2013). At ~1000 K, equilibrium chemistry predicts a transition from CH_4- to CO-dominated atmospheres toward higher temperatures (e.g., Moses et al. 2011; Madhusudhan 2012; Venot et al. 2014; Heng & Tsai 2016). However, 800–1200 K is also the temperature range where the assumption of chemical equilibrium breaks down, because the chemical and dynamical timescales become comparable and photochemistry may not be negated by high temperatures. For example, Madhusudhan & Seager (2011) find tentative evidence for the overabundance of CO (compared to expectations from chemical equilibrium) in the warm Neptune GJ 436b, which has an equilibrium

[7] University of Bern International 2021 Ph.D Fellowship.
[8] On sabbatical at the Center for Space and Habitability in Fall 2019.

temperature of about 649 ± 60 K (Torres et al. 2008); see also Morley et al. (2017). In our own Jupiter, the overabundance of CO was interpreted as a sign of disequilibrium chemistry due to atmospheric mixing (Prinn & Barshay 1977). Similarly, Oppenheimer et al. (1998) detected an excess of CO in the brown dwarf Gliese 229B.

For all of these reasons, warm Neptunes with atmospheric temperatures in the range of 800–1200 K are the next frontier in understanding atmospheric chemistry from transmission spectroscopy.

1.2. Motivation II: Accuracy of Constraining Elemental Abundances and C/O

The key controlling parameters of atmospheric chemistry are the set of elemental abundances (mainly C/H, O/H, N/H) and C/O (e.g., Burrows & Sharp 1999; Madhusudhan 2012; Heng & Tsai 2016). Atmospheric mixing and photolysis act to complicate the translation between the elemental and molecular abundances (e.g., Moses et al. 2011; Tsai et al. 2017). As already noted by Line et al. (2013) and Greene et al. (2016), the best approach for inferring the elemental abundances and C/O from spectra is to directly retrieve the abundances of the major carbon, oxygen, and nitrogen molecular carriers,

$$\begin{aligned}
\mathrm{C/H} &= \frac{X_{\mathrm{CO}} + X_{\mathrm{CO_2}} + X_{\mathrm{CH_4}} + X_{\mathrm{HCN}} + 2X_{\mathrm{C_2H_2}}}{X_{\mathrm{H}}},\\
\mathrm{O/H} &= \frac{X_{\mathrm{CO}} + 2X_{\mathrm{CO_2}} + X_{\mathrm{H_2O}}}{X_{\mathrm{H}}},\\
\mathrm{N/H} &= \frac{X_{\mathrm{NH_3}} + X_{\mathrm{HCN}}}{X_{\mathrm{H}}},\\
\mathrm{C/O} &= \frac{X_{\mathrm{CO}} + X_{\mathrm{CO_2}} + X_{\mathrm{CH_4}} + X_{\mathrm{HCN}} + 2X_{\mathrm{C_2H_2}}}{X_{\mathrm{CO}} + 2X_{\mathrm{CO_2}} + X_{\mathrm{H_2O}}}, \qquad (1)
\end{aligned}$$

where X_i are the volume mixing ratios of molecules and $X_{\mathrm{H}} = 2X_{\mathrm{H_2}} + 4X_{\mathrm{CH_4}} + X_{\mathrm{HCN}} + 2X_{\mathrm{C_2H_2}} + 2X_{\mathrm{H_2O}} + 3X_{\mathrm{NH_3}}$. In H_2-dominated atmospheres (as studied here), $X_{\mathrm{H}} \approx 2X_{\mathrm{H_2}}$. It is important to note that these are inferred quantities in the gas phase, which may differ from their bulk values owing to condensation, e.g., sequestration of oxygen into olivine. Only in extremely hot conditions, such as for ultrahot Jupiters and main-sequence stars, may we reasonably assume that the photospheric and bulk elemental abundances are similar (e.g., Kitzmann et al. 2018). Line et al. (2013) caution that retrieving directly for the molecular abundances results in a nonuniform prior for C/O (see their Section 3.3).

In the current study, we consider seven molecules. CO and CH_4 are the major carbon carriers (Burrows & Sharp 1999), with CO_2 being a minor carbon carrier unless the metallicity is highly enriched (e.g., Moses et al. 2013; Heng & Lyons 2016). H_2O and CO are major oxygen carriers (Burrows & Sharp 1999). Acetylene (C_2H_2) becomes nonnegligible as C/O approaches unity (e.g., Moses et al. 2011; Madhusudhan 2012; Heng & Tsai 2016). NH_3 competes with molecular nitrogen (N_2) as the major nitrogen carrier (Burrows & Sharp 1999), while hydrogen cyanide (HCN) is an important link between the carbon and nitrogen reservoirs (e.g., Moses et al. 2011). The accuracy of retrieving for the elemental abundances hinges on a spectrum having sufficient spectral resolution, signal-to-noise ratio, and wavelength coverage to accurately account for the molecules that are present in sufficient amounts. If not all of the molecules are properly accounted for, it will lead to erroneous inferences about C/O.

A second approach is to assume chemical equilibrium and parameterize all of the molecular abundances by two numbers: C/O and the metallicity. Chemical equilibrium is a *local* approximation in the sense that each patch of atmosphere has no memory of its past and all of the molecular abundances may be completely determined once one has knowledge of the local temperature and pressure. Metallicity has three definitions in the astronomical literature: stellar astrophysicists refer to the relative abundance of all elements heavier than helium *by mass* (Section 3.12 of Asplund et al. 2009), observational spectroscopists refer to the elemental abundance of iron *by number* (Section 4.2 of Asplund et al. 2009), and atmospheric chemists typically refer to the elemental abundance of a volatile element (e.g., carbon) by number (Moses et al. 2013). In the third definition, it is usually assumed that the ratios of the elemental abundances are kept fixed to their solar values with the exception of C/H or O/H, which are allowed to be free parameters in order to allow for a variable C/O (e.g., Moses et al. 2013; Heng 2018; Drummond et al. 2019). In chemical equilibrium, knowledge of the abundance of a single carbon or oxygen carrier is sufficient to constrain C/H or O/H, respectively. However, if chemical equilibrium is a poor assumption, then misleading conclusions will follow. None of the atmospheres of solar system bodies are well described by chemical equilibrium.

One of the goals of the current study is to examine the relationship between the accuracy of retrieving for the elemental abundances and hence C/O.

1.3. Motivation III: Novel Information Content Analysis Approach, Feasible for Complex Models

Classical information content (IC) analysis is based on computing Jacobians, which are the derivatives of the model output (e.g., transit depth) with respect to the parameters. See Section 2 of Batalha & Line (2017) for a recent review.[9] Classical IC analysis is a time-consuming process. For example, Section 3.1 of Batalha & Line (2017) states, "For each of these 84 combinations of planet types, we compute a *separate* Jacobian" (these authors' emphasis). Howe et al. (2017) introduce the use of "mutual information" (see their Section 2) but remark how "the difficulty with the use of mutual information is that it is computationally intensive, especially for the dense data sets produced by JWST." For reasons of computational feasibility, Howe et al. (2017) adopted a simple three-parameter model that assumed an isothermal transit chord, gray clouds, and a metallicity.[10] Batalha & Line (2017) assumed chemical equilibrium models described by the metallicity and C/O and a nongray treatment of clouds.

In the current study, we adopt a qualitatively different approach to IC analysis. Recently, Márquez-Neila et al. (2018) demonstrated that the classical machine-learning method of the "random forest" (Ho 1998; Breiman 2001; Criminisi et al. 2011) may be adapted to perform atmospheric retrieval, as a complement to

[9] Note that the treatment in Batalha & Line (2017) requires the assumption of Gaussian probability distributions.

[10] Presumably, this requires the assumption of chemical equilibrium, but Howe et al. (2017) do not explicitly state this beyond the following sentence: "Most notably, the alkali metal lines and the CO bands grow much stronger with increasing temperature as the concentrations of these species in chemical equilibrium increase."

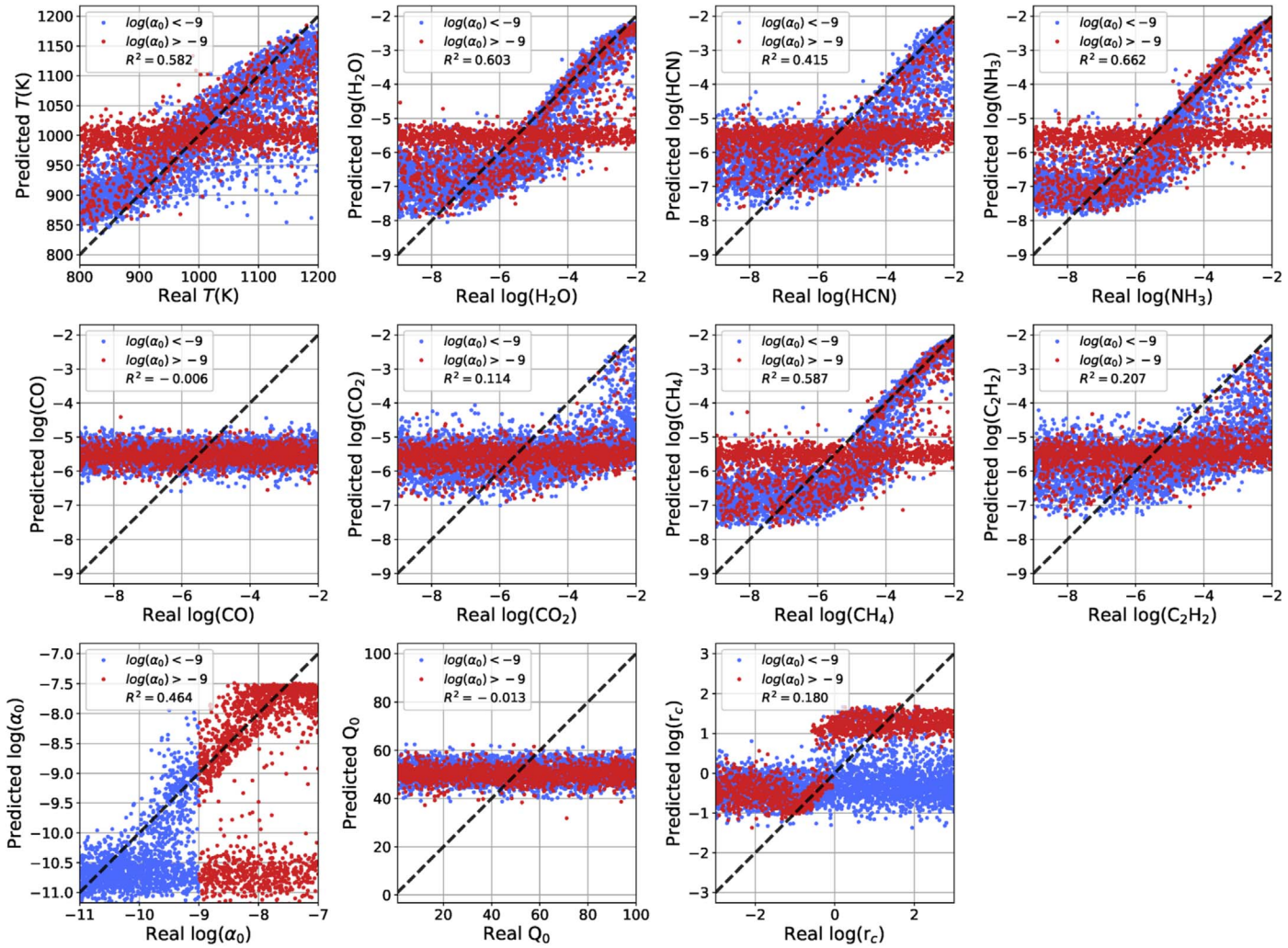

Figure 1. RvP values of the various parameters from a suite of 20,000 mock retrievals on HST-WFC3 transmission spectra using the random forest method. For clarity (and with no loss of generality), we only show 5000 of these mock retrievals. The stellar and exoplanetary parameters of GJ 436 and the warm Neptune GJ 436b, respectively, are assumed (see text). The synthetic spectra are composed of 13 wavelength bins from 0.8 to 1.7 μm following Kreidberg et al. (2015) and to provide continuity with Márquez-Neila et al. (2018). Each synthetic data point assumes an optimistic photon-limited uncertainty of 20 parts per million (ppm). The blue and red points correspond to cloud-free ($\alpha_0 < 10^{-9}$ cm^{-1}; see text for definition) and cloudy ($\alpha_0 > 10^{-9}$ cm^{-1}) models, respectively. Negative and positive values of the coefficient of determination ($\mathcal{R}$) correspond to negative and positive correlations, respectively.

standard methods such as nested sampling (Skilling 2006; Feroz & Hobson 2008; Feroz et al. 2009, 2019). Fisher et al. (2020) compare random forest retrieval to other methods (nested sampling, Bayesian neural networks). There are three distinct advantages of random forest retrieval in terms of practical implementation. First, it performs "feature importance," which in the context of spectra means that it is able to compute the relative importance of each data point for constraining each parameter of a chosen model used to interpret the spectra. Second, it is able to easily perform large suites of mock retrievals in the form of "real versus predicted" (RvP) plots (Márquez-Neila et al. 2018; Fisher et al. 2020; Oreshenko et al. 2020). Third, since the random forest may be trained on a pre-computed model grid of arbitrary sophistication, the obstacles of computational feasibility encountered by Howe et al. (2017) and Batalha & Line (2017) may be overcome. Instead of assuming chemical equilibrium, we allow each of our seven molecules to take on a broad range of abundances and infer the elemental abundances and C/O from the retrieved abundances.

Examples of RvP plots are shown in Figure 1, where we perform a suite of 20,000 mock retrievals for Hubble Space Telescope (HST) Wide Field Camera 3 (WFC3) transmission spectra of the warm Neptune GJ 436b. These RvP plots may be used to quantify the ability of a retrieval to accurately recover each parameter value of the model. The random forest reports the *mean* predicted values of the parameters in the RvP plots. The figure of merit used is the "coefficient of determination" ($\mathcal{R}^2$), where $\mathcal{R}^2 = 0$ means zero predictability (zero correlation between the RvP values of a parameter) and $\mathcal{R}^2 = 1$ means perfect predictability. Model degeneracies will generally lower the value of $\mathcal{R}^2$ (Márquez-Neila et al. 2018; Fisher et al. 2020; Oreshenko et al. 2020). The RvP plots reproduce widely accepted knowledge in the exoplanet retrieval literature: WFC3 transmission spectra probe mainly H_2O, CH_4, and NH_3, with some sensitivity to HCN, but are insensitive to CO and CO_2. Furthermore, while cloud particle radius and abundance may be retrieved, one is blind to the retrieval of cloud composition. If CO and H_2O are present in comparable abundances, then the retrieval will only accurately infer the H_2O abundance, leading to an inaccurate estimate of C/O.

Figure 2 shows the accompanying feature importance plots. Each feature importance plot quantifies the relative importance

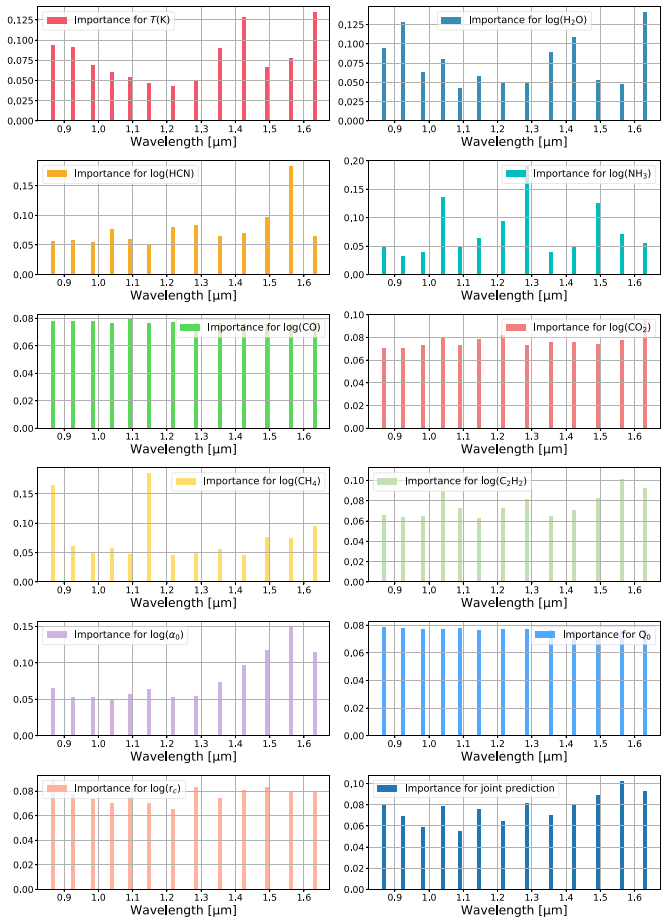

Figure 2. Companion figure montage to Figure 1, which shows the "feature importance" plots from the random forest retrieval analysis. Each feature importance plot quantifies the relative importance of each data point in an HST-WFC3 transmission spectrum for determining the value of a given parameter. The entries in each panel add up to unity.

of each of the 13 data points in the WFC3 transmission spectrum for determining the value of a parameter. For example, the two data points near 1.4 μm constrain the water abundance, which matches our intuition of a water feature being present at these wavelengths. The bluest data points constrain the cloud abundance and particle radius. When the feature importance is about equal for all 13 data points (e.g., for CO), it often indicates a lack of sensitivity to a given parameter, which can only be confirmed by cross-matching with the $\mathcal{R}^2 \approx 0$ value from the RvP plot (Figure 1).

In the current study, we will demonstrate the usefulness of both feature importance and RvP analysis for understanding the IC of JWST NIRSpec transmission spectra of warm Neptunes. The random forest technique has also been applied to ground-based spectra of brown dwarfs at medium spectral resolution (Oreshenko et al. 2020) and ultrahot Jupiters at high spectral resolution (Fisher et al. 2020).

1.4. Motivation IV: Planning NIRSpec Observations on JWST

The current study is restricted to transmission spectroscopy at optical to near-infrared wavelengths. Specifically, we consider the NIRSpec instrument on JWST.[11] In the low-resolution (∼100) prism mode, NIRSpec has a simultaneous wavelength coverage of 0.6–5.3 μm. It is suitable for stars fainter than $J \approx 10$, corresponding to Kepler and fainter TESS targets. At medium resolution (∼1000), NIRSpec has four modes: 0.7–1.27 μm (G140M/F070LP), 0.97–1.84 μm (G140M/F100LP), 1.66–3.07 μm (G235M/F170LP), and 2.87–5.10 μm (G395M/F290LP). These modes are suitable for stars fainter than $J \approx 6$–8. Four high-resolution (∼2700) modes exist as well, but as we will show, these do not add much interpretational value, in terms of retrieving elemental and molecular abundances, to what the medium-resolution modes already offer. Table 1 provides a summary of the JWST NIRSpec modes we will consider in the current study.

Table 1
JWST NIRSpec Modes Considered

Shorthand	Wavelengths (μm)	Configuration	Resolution
L	0.6–5.3	PRISM/CLEAR	100
M1	0.7–1.27	G140M/F070LP	600
M2	0.97–1.84	G140M/F100LP	1000
M3	1.66–3.07	G235M/F170LP	1000
M4	2.87–5.10	G395M/F290LP	1000
H4	2.87–5.10	G395H/F290LP	2700

1.5. Layout of Study

In Section 2, we describe our methods of computation. In Section 3, we present the results from our IC analyses and also an improved diagnostic for chemical disequilibrium. In Section 4, we compare our results to those of previous studies and discuss their implications for planning JWST observations.

2. Methodology

2.1. Opacities

2.1.1. Molecules

The molecular opacities of H_2O, HCN, NH_3, CO, CO_2, CH_4, and C_2H_2 are taken from the `ExoMol` (Barber et al. 2006, 2014; Yurchenko et al. 2011, 2013; Yurchenko & Tennyson 2014), `HITRAN` (Rothman et al. 1987, 1992, 1996, 2003, 2005, 2010, 2013), and `HITEMP` (Rothman et al. 2009) spectroscopic databases; the pressure broadening parameters for H_2–He mixtures are taken from the `ExoMol` database. A review of the spectroscopic databases may be found in Tennyson & Yurchenko (2017). For a review of how to compute opacities given inputs from the spectroscopic databases, we refer the reader to, for example, the appendix of Rothman et al. (1996), Grimm & Heng (2015), Chapter 5 of Heng (2017), and Yurchenko et al. (2018). All opacities are calculated with the HELIOS-K opacity calculator (Grimm & Heng 2015) from 10^{-8} to 10^{3} bar. Table 2 states the details of the opacities, including the range of wavenumbers/wavelengths over which spectroscopic data needed as input exist. When computing transmission spectra, we use opacity sampling at the resolutions stated in Table 1.

2.1.2. Clouds

In the current study, we will use the terms "cloud," "haze," and "aerosol" interchangeably, based on the reasoning that while these terms may reflect different formation pathways, the effects on a spectrum follow a common phenomenological treatment. There is

[11] https://jwst-docs.stsci.edu/near-infrared-spectrograph/nirspec-observing-modes/nirspec-bright-object-time-series-spectroscopy

Table 2
Spectroscopic Line Lists Used to Generate Opacities

Molecule	Line List	Shortest Wavelength (μm)	Wavenumber Range (cm^{-1})	References
CO	12C–16O__Li2015	0.45	0–22000	Li et al. (2015)
CO_2	HITEMP 2010	1.04	258–9648	Rothman et al. (2009)
CH_4	12C–1H4__YT10to10	0.83	0–12100	Yurchenko et al. (2013), Yurchenko & Tennyson (2014)
C_2H_2	HITRAN 2016	1.01	0–9889	Gordon et al. (2017)
H_2O	1H2–16O__POKAZATEL	0.24	0–41200	Polyansky et al. (2018)
HCN	1H–12C–14N__Harris	0.57	0–17585	Harris et al. (2006), Barber et al. (2014)
NH_3	14N–1H3__BYTe	0.83	0–12000	Yurchenko et al. (2011)

no consensus on the use of these terms: Earth scientists use "haze" versus "cloud" as a measure of particle size, while planetary scientists use these terms to refer to photochemical and thermochemical formation origins, respectively.

If a cloud consists of spherical particles of a single radius (i.e., a monodisperse cloud), then its cross section is

$$\sigma_c = Q\pi r_c^2, \tag{2}$$

where Q is the extinction efficiency. It may be computed using Mie theory (Mie 1908). Kitzmann & Heng (2018) use the open-source `LX_MIE` Mie code to calibrate a convenient fitting function for Q,

$$Q = \frac{Q_1}{Q_0 x^{-a} + x^{0.2}}, \tag{3}$$

where $Q_1 \approx 4$ (Kitzmann & Heng 2018), the dimensionless size parameter is $x = 2\pi r_c/\lambda$, and λ is the wavelength. This fitting function smoothly transitions between the regimes of small ($x \ll 1$; Rayleigh and nongray continuum) and large ($x \gg 1$; gray continuum) particles. For simplicity, we assume $a = 4$; see Table 2 of Kitzmann & Heng (2018) for the values of a as a function of the composition. Refractory and volatile condensates correspond to $Q_0 \sim 10$ and $\sim$100, respectively (see Table 2 of Kitzmann & Heng 2018). The cloud extinction coefficient is assumed to be uniform along the transit chord.

It is worth noting that this simplified treatment of the cloud cross section does not capture composition-specific spectral features (e.g., Cushing et al. 2009; Lee et al. 2014).

As it is calibrated on first-principles calculations, our treatment of clouds is an improvement over the gray cloud assumption of Howe et al. (2017) and the approach of Greene et al. (2016) and Batalha & Line (2017), who used a combination of a "cloud top pressure" (for gray clouds) and a power-law parameterization (for nongray clouds).

2.1.3. Total Extinction Coefficient

The total extinction coefficient is

$$\alpha = \alpha_c + \sum_i \frac{m_i}{m} X_i \kappa_i \rho, \tag{4}$$

where m_i is the mass, X_i is the volume mixing ratio, and κ_i is the opacity of each molecule. The sum is performed over all of the molecules in the system. The mass density and mean molecular mass of the atmosphere are given by ρ and m, respectively. The extinction coefficient associated with clouds is written as

$$\alpha_c = \frac{\alpha_0}{Q_0 x^{-4} + x^{0.2}}. \tag{5}$$

The mean molecular mass, cloud volume mixing ratio (X_c), and Q_1 are subsumed into a single fitting parameter,

$$\alpha_0 \propto \frac{Q_1 \pi r_c^2 X_c}{m}. \tag{6}$$

2.2. Transmission Spectra

Consistent with Greene et al. (2016) and Howe et al. (2017), we assume isothermal, nonisobaric transit chords.[12] We use the `HELIOS-O` code to compute transmission spectra (Gaidos et al. 2017; Bower et al. 2019). Each model atmosphere is divided into 150 annuli in pressure (P) from 10^{-8} to 10 bar. The limit of 10 bar is chosen to ensure that the atmosphere is fully opaque at the lower boundary and has no bearing on the final outcome of the calculation.

At each wavelength, the slant optical depth is computed using (Brown 2001)

$$\tau = \int_{-\infty}^{\infty} \alpha \, dx, \tag{7}$$

where x is the spatial coordinate along the line of sight. The transmission function along each line of sight is

$$\mathcal{T} = e^{-\tau}. \tag{8}$$

Integrating along the radial coordinate yields the transit depth (Brown 2001),

$$\left(\frac{R}{R_\star}\right)^2 = \frac{1}{R_\star^2} \int_0^{\infty} 2r(1 - \mathcal{T}) \, dr. \tag{9}$$

JWST spectra are expected to encode enough information (Fisher & Heng 2018) to break the normalization degeneracy (Benneke & Seager 2012; Griffith 2014; Barstow et al. 2015; Heng & Kitzmann 2017; Heng 2019). Nevertheless, we account for this degeneracy by matching the computed white-light radius of each model to the measured one ($R = 0.3767\ R_J$ from 0.5 to 1.0 μm; Torres et al. 2008).

Across the wavenumber range of $1/\lambda = 1800$–$17{,}000\ cm^{-1}$ (wavelength range of 0.6–5.5 μm), we assume a uniform spacing in $\log(1/\lambda)$ corresponding to 6700 points, such that the spectral resolution is approximately constant with an average

[12] It is not equivalent to assuming an isothermal atmosphere; rather, it is the assumption that the region of the atmosphere probed by transmission spectroscopy is isothermal over the wavelength (and hence pressure) range considered.

value of 3000. The spectra are then restricted in wavelength and binned down to a spectral resolution of 100, 600, or 1000, depending on which modes of NIRSpec one is studying (see Table 1). An optimistic, photon-limited uncertainty of 20 ppm per data point is assumed, consistent with Greene et al. (2016). The intention is to identify the possible weaknesses of each NIRSpec mode even under idealized conditions.

On a desktop computer (Intel Core i9-7960X CPU), it takes `HELIOS-O`, which is written in the C++ programming language, about 1 s to compute each model. For the entire grid of 100,000 models, this amounts to about 30 hr of computing time.

2.3. Random Forest Retrieval

The "random forest" is a classical, supervised method of machine learning (Ho 1998; Breiman 2001). It belongs to a class of methods known as Approximate Bayesian Computation (ABC). Within the ABC framework, it has been demonstrated that one may compute approximate posterior distributions and perform model comparison via computation of the Bayesian evidence (Sisson et al. 2019).

As is appropriate for continuous quantities such as transit depths or radii, a regression tree (rather than a decision tree) is used to classify transmission spectra with different sets of parameter values (treated as "labels"; Márquez-Neila et al. 2018). A bootstrapping method is used to generate an uncorrelated forest of regression trees, and the combined output of the random forest yields the posterior distributions of parameters (Criminisi et al. 2011). Following Fisher et al. (2020), we take as output all of the entries in a leaf, rather than the average of the leaf, as the sampled posterior distribution of a parameter.

As demonstrated by Márquez-Neila et al. (2018), the random forest produces two additional diagnostics: feature importance plots, which quantify the relative importance of each data point in the transmission spectrum for constraining each parameter; and RvP plots, which quantify the degree to which each parameter may be predicted in mock retrievals given the noise model. The RvP analysis is essentially an efficient way to generate large suites ($\sim 10^4$) of mock retrievals, which is computationally challenging to accomplish using standard retrieval methods (e.g., Barstow et al. 2015).

The range of values of the model parameters, as well as the assumptions on their prior distributions, are stated in Table 3. Each parameter is randomly drawn from its prior and a noise-free transmission spectrum is generated, as explained in Section 2.2. In order to add noise, each point in the synthetic spectrum is assumed to follow a Gaussian distribution with a standard deviation of 20 ppm. The points are then randomly sampled from these distributions, centered on their noise-free values.

This is repeated to build a grid of 100,000 models for the forest, split into 80,000 for training and 20,000 for testing. The random forest consists of 1000 trees. Tree splitting is performed using the following steps: the range of values of each parameter is normalized such that its maximum value is 100; tree splitting ceases when the change in total variance of the parameter values (as a node is split into two branches) is less than a stated tolerance, which is set to 0.01. Each time a tree is split, a random subset of $\sim\sqrt{N}$ points is used, where N is the total number of spectral points, to reduce biases. Tree pruning methods are not used. The implementations of the random forest method and R^2 metric are from the open-source `scikit.learn` library (Pedregosa et al. 2011) in the `Python` programming language.

On a desktop computer (Intel Core i7 CPU), it takes `HELA`, which is written in the `Python` programming language, about 10 minutes to train the random forest.

Table 3
Retrieved Parameters and Their Prior Distributions

Quantity	Symbol	Units	Range	Prior Type
Temperature	T	K	800–1200	uniform
Volume mixing ratios	X_i	⋯	10^{-9}–10^{-2}	log-uniform
Cloud extinction coefficient normalization	α_0	cm^{-1}	10^{-11}–10^{-7}	log-uniform
Proxy for cloud composition	Q_0	⋯	1–100	uniform
Cloud particle radius	r_c	μm	10^{-3}–10^{3}	log-uniform

3. Results

As an illustration, we will use the example of GJ 436b for our calculations: the GJ 436 star has a stellar radius of $R_\star = 0.455\ R_\odot$, and GJ 436b has a surface of $g = 1318\ \mathrm{cm\ s}^{-2}$ (von Braun et al. 2012). The qualitative conclusions of our study do not depend on the choice of these parameter values.

3.1. RvP Analysis of Different JWST NIRSpec Modes

Table 1 lists the four medium-resolution modes of JWST NIRSpec. The expectation is that the M1 (0.7–1.27 μm) and M2 (0.97–1.84 μm) modes, which probe a collective wavelength range similar to the WFC3 instrument of HST, encode the most information on cloud properties (e.g., Lecavelier des Etangs et al. 2008) but may be insensitive to important carbon-bearing molecules such as CO and CO_2. Therefore, we begin the discussion by comparing the M1 and M4 (2.87–5.10 μm) modes.

Figure 3 shows the outcomes of performing 20,000 mock retrievals for each of the modes in turn. For clarity of presentation (and with no loss of generality), we display only 5000 out of the 20,000 mock retrievals. It is emphasized again that the random forest reports the *mean* predicted value of each parameter.[13] Based on the similar $\mathcal{R}^2$ values obtained, the M1 and M4 modes do comparably well at constraining the H_2O and NH_3 abundances, as well as the temperature. The M4 mode outperforms the M1 mode by more than 0.1 in $\mathcal{R}^2$ value for constraining the abundances of HCN, CH_4, and C_2H_2. As demonstrated by the low $\mathcal{R}^2$ values, the M1 mode is insensitive to CO_2 ($\mathcal{R}^2 = 0.138$) and essentially blind to CO ($\mathcal{R}^2 = -0.003$). The M4 mode offers drastic improvements on constraining CO ($\mathcal{R}^2 = 0.508$) and CO_2 ($\mathcal{R}^2 = 0.779$) owing to their spectral features across 4–5 μm (Figure 4).

When a mock retrieval fails to predict the value of a given parameter, the RvP analysis returns values that are the mean of the range considered. In the case of CO, since the range of volume mixing ratios considered is 10^{-9} to 10^{-2} (in log-uniform spacing), the random forest returns $X_{\mathrm{CO}} = 10^{-6}$ to 10^{-5} for the M1 mode. In other RvP plots where the predicted values of the parameters level off at a value that is below the mean of the range considered, these indicate the minimum or

[13] The median value may also be reported, which is the approach followed by Fisher et al. (2020).

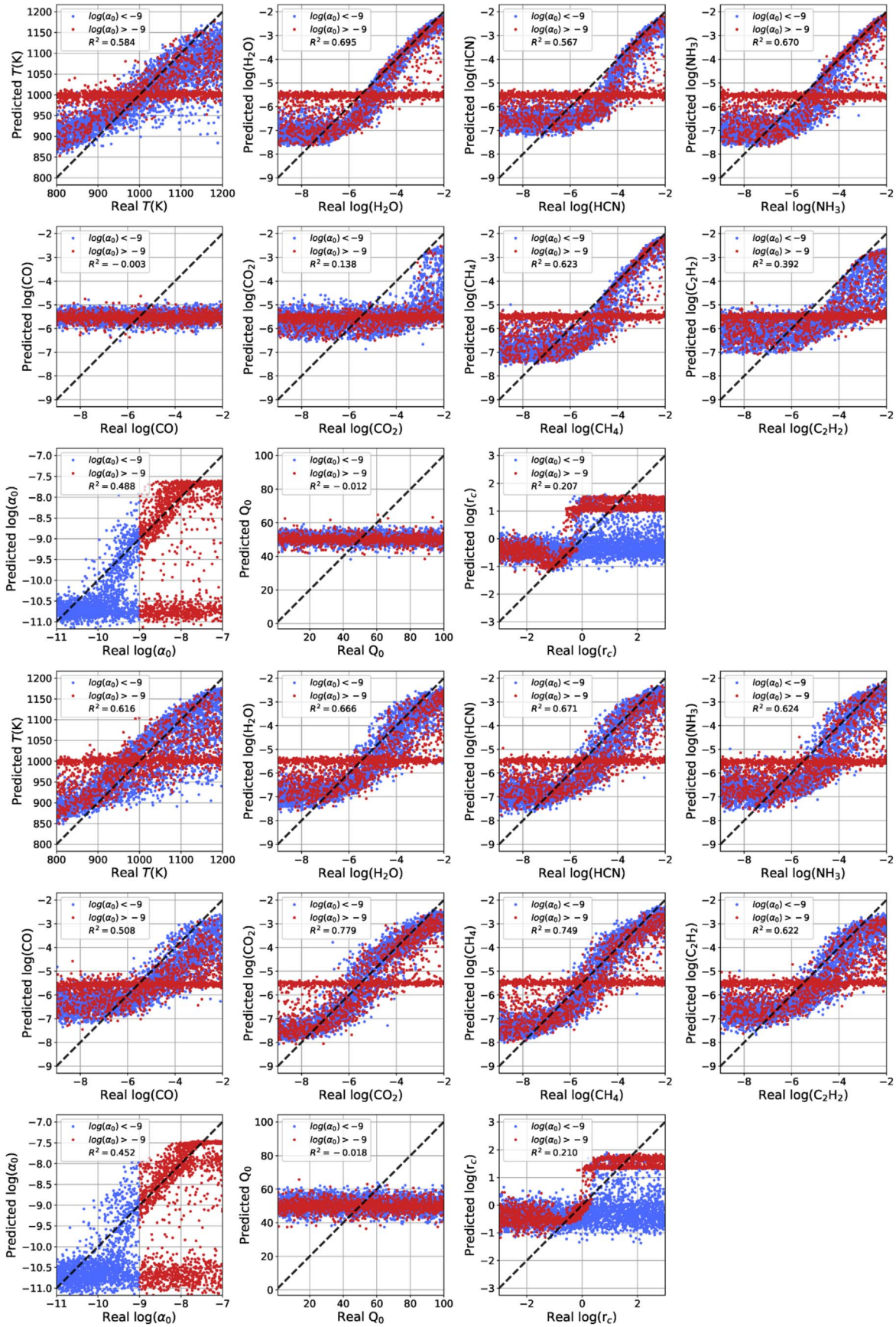

Figure 3. RvP analysis of the medium-resolution M1 (top montage of 11 panels) vs. M4 (bottom montage of 11 panels) modes of JWST NIRSpec. As in Figure 1, the blue and red points correspond to cloud-free ($\alpha_0 < 10^{-9}$ cm g^{-1}) and cloudy ($\alpha_0 > 10^{-9}$ cm g^{-1}) transmission spectra, respectively. For clarity of presentation, we show only 5000 out of the actual 20,000 mock retrievals performed.

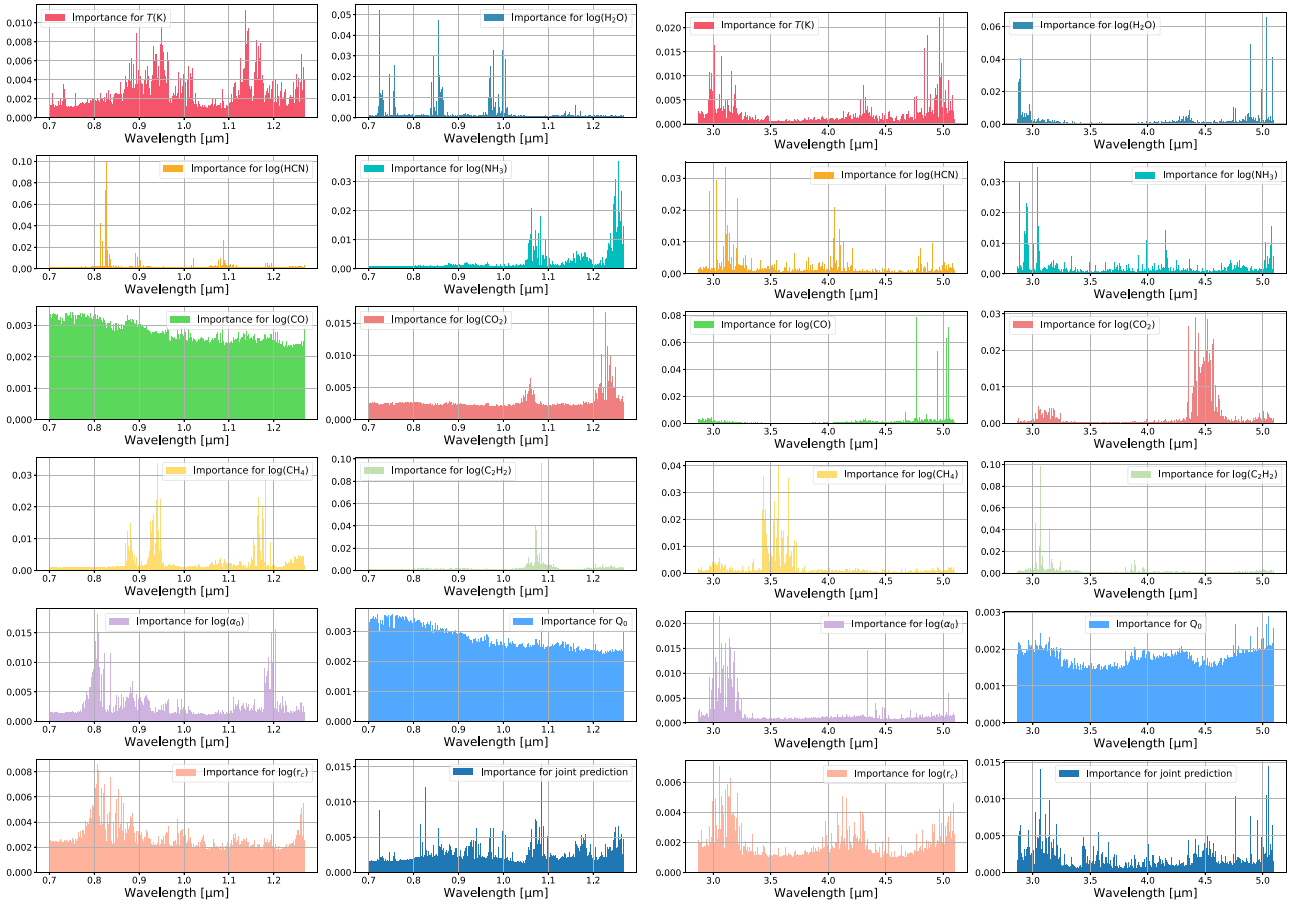

Figure 4. Feature importance analysis of the medium-resolution M1 (left montage of 12 panels) vs. M4 (right montage of 12 panels) modes of JWST NIRSpec.

threshold value of a parameter that can be constrained given the noise model. For example, $X_{H_2O} \gtrsim 10^{-7}$ for both the M1 and M4 modes. Generally, volume mixing ratios as low as $\sim 10^{-8}$ may be constrained given the assumed 20 ppm noise floor.

In Figure 3, the points have been color-coded blue ($\alpha_0 < 10^{-9}$ cm^{-1}) or red ($\alpha_0 > 10^{-9}$ cm^{-1}) to correspond to cloud-free or cloudy atmospheres, respectively. This threshold value of α_0 was obtained by trial and error and is guided mainly by inspecting the RvP behavior of both α_0 and r_c. The bimodal behavior of α_0 above this threshold is an indication of the degeneracy between the degree of cloudiness and the molecular abundances. The trend of r_c leveling off at $\gtrsim 1$ μm is the outcome of the cloud opacity becoming gray/constant as the cloud particles become large compared to the wavelengths probed. This trend is consistent with the basic principles of Mie theory. In all of the RvP plots of the molecular abundances and temperature, a subpopulation of the red (cloudy) points cluster in the middle of the range of values considered, indicating that the random forest does not predict a value for the given parameter.

The M1 and M4 modes constrain α_0 ($\mathcal{R}^2 = 0.488$ vs. 0.452) and r_c ($\mathcal{R}^2 = 0.207$ vs. 0.210) almost equally well. Both the M1 and M4 modes have no sensitivity to the cloud composition (via Q_0; $\mathcal{R}^2 \approx 0$), which implies that it is challenging to identify cloud composition by constraining changes in the gradient of the spectral continuum alone. It does not rule out the possibility that higher-order spectral features that are composition specific may retain constraining power (e.g., Cushing et al. 2009; Lee et al. 2014).

For completeness, the RvP plots of the M2, M3, and L modes are included in the Appendix as Figures A2, A4, and A6, respectively. The M2 mode exhibits similar behavior to the M1 mode in that it is somewhat insensitive to CO_2 ($\mathcal{R}^2 = 0.322$) and nearly blind to CO ($\mathcal{R}^2 = 0.039$). The M3 mode (1.66–3.07 μm) is blind to CO ($\mathcal{R}^2 = 0.075$) but sensitive to CO_2 ($\mathcal{R}^2 = 0.669$). The L mode has good sensitivity to CO_2 ($\mathcal{R}^2 = 0.763$) but is largely insensitive to CO ($\mathcal{R}^2 = 0.171$). Section 4.2 and Figure 9 perform a detailed comparison of the $\mathcal{R}^2$ values of every parameter for all of the modes considered in the present study.

3.2. Feature Importance Analysis of JWST NIRSpec Modes

Figure 4 shows the feature importance analysis of the M1 versus M4 modes. Each panel shows the fractional importance of each data point for constraining a given parameter. It cannot be overemphasized that the feature importance values cannot be compared between panels, because the entries are normalized such that they add up to unity within the same panel.

The feature importance analysis of the M4 mode reproduces our intuition about the warm Spitzer Space Telescope channels. Channel 1 of the IRAC instrument, which ranges from about 3.1 to 3.9 μm and is often quoted as the "3.6 μm channel," probes several spectral features of methane (e.g., Sudarsky et al. 2003; Fortney et al. 2005, 2006, 2010). Channel 2 of IRAC, which ranges from about 3.9 to 5.1 μm and is often quoted as the "4.5 μm channel," probes carbon monoxide (e.g., Sudarsky et al. 2003; Fortney et al. 2005, 2006, 2010; Charbonneau et al. 2008). It is consistent with the narrative that the flux ratios of these

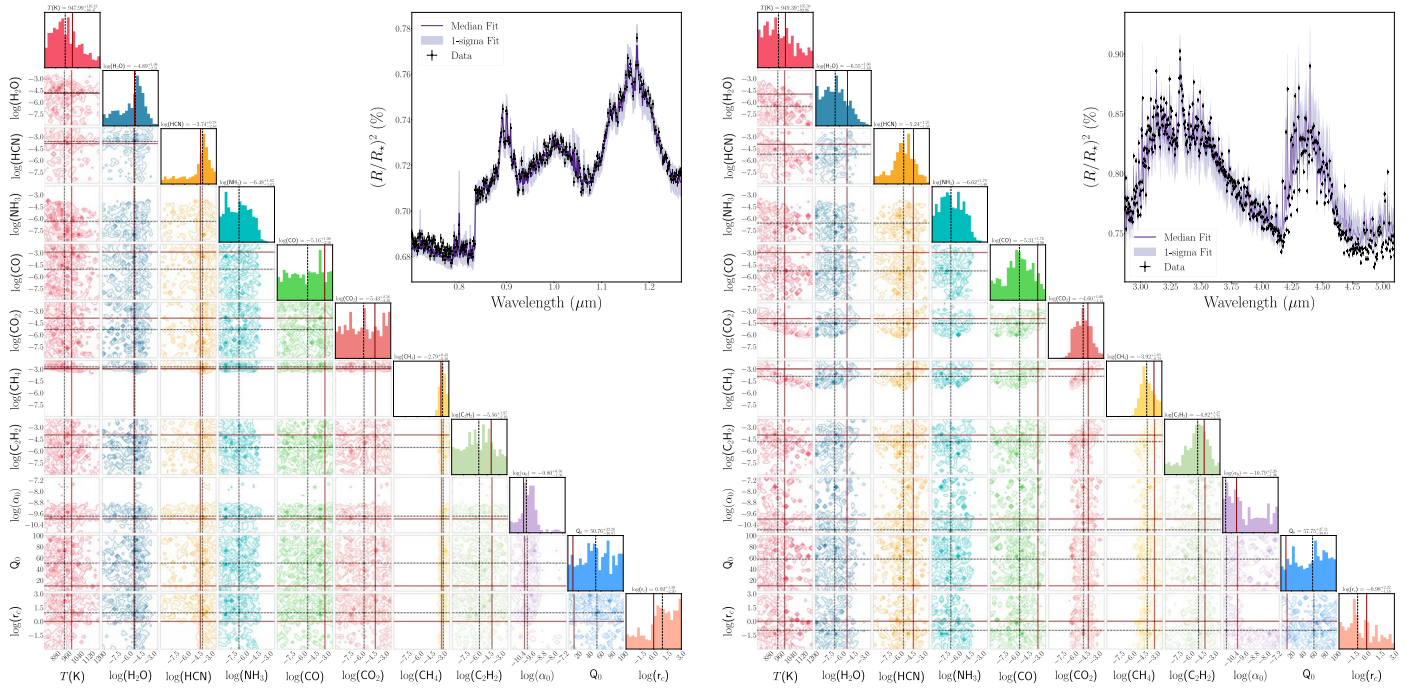

Figure 5. Posterior distributions from mock retrievals of a carbon-rich (water-poor) atmosphere with $T = 1000$ K, $X_{CO} = X_{CH_4} = 10^{-3}$, $X_{HCN} = X_{C_2H_2} = X_{CO_2} = 10^{-4}$, $X_{H_2O} = 10^{-5}$, $\alpha_0 = 10^{-10}$ cm^{-1}, $Q_0 = 10$, and $r_c = 1$ μm. The left and right montages are for the M1 and M4 modes, respectively. The kink in the synthetic spectrum associated with the M1 mode is due to the nonexistence of CH_4 line list data at bluer wavelengths. The vertical black dotted lines show the median value of each posterior distribution. Wherever applicable, the vertical red solid lines show the truth values of a parameter.

channels probe the relative abundances of CH_4 to CO, and is thus a measure of disequilibrium chemistry (e.g., Madhusudhan & Seager 2011; Moses et al. 2011).

Other properties are less apparent without detailed scrutiny of the feature importance plots. Generally, molecules such as H_2O, HCN, NH_3, CH_4, and C_2H_2 have multiple spectral lines distributed across the wavelength ranges of both the M1 and M4 modes. For the M4 mode, there are strong CO_2 features between 4 and 5 μm. It also encompasses a CO feature at 4.7 μm, which explains the ability of the 4.5 μm channel of IRAC to constrain carbon monoxide. The M1 and M4 modes are equally good at constraining α_0 and the cloud particle radius (based on comparing the $\mathcal{R}^2$ values as discussed earlier), but these constraints come from different wavelength regions.

Parameters associated with $\mathcal{R}^2 \approx 0$ typically have almost equal feature importance distributed across wavelength, as is the case for Q_0 (both M1 and M4) and CO (only M1). For completeness, we include in the Appendix the feature importance plots of the M2, M3, and L modes in Figures A3, A5, and A7, respectively.

3.3. Posterior Distributions from Mock Retrievals

As an illustration, we consider a case study that is motivated by qualitative trends in gaseous, equilibrium chemistry at ~1000 K (e.g., Moses et al. 2011, 2013; Madhusudhan 2012; Heng & Tsai 2016): a carbon-rich (water-poor) atmosphere consisting of $X_{CO} = X_{CH_4} = 10^{-3}$, $X_{HCN} = X_{C_2H_2} = X_{CO_2} = 10^{-4}$, and $X_{H_2O} = 10^{-5}$, which corresponds to $C/O \approx 1.98$ or $\log C/O \approx 0.30$. For illustration, we assume $T = 1000$ K, $\alpha_0 = 10^{-10}$ cm^{-1}, $Q_0 = 10$, and $r_c = 1$ μm.

Consistent with the insensitivity of the M1 mode to CO, CO_2, and C_2H_2, the posterior distributions of these molecules are unconstrained (Figure 5). The M4 mode does surprisingly poorly on CO, but this is because its spectral lines are being masked by those of CO_2 and CH_4 (see Appendix). Both modes obtain only an upper limit for NH_3, which is absent from this model atmosphere. Overall, the M4 mode does somewhat better at retrieving the C/O ratio compared to the M1 mode (Figure 6).

Identifying the minimum set of molecules needed to explain a spectrum may be achieved using Bayesian model comparison (e.g., Benneke & Seager 2012; Waldmann et al. 2015; Fisher & Heng 2018) or deep learning methods (e.g., Waldmann 2016), which are beyond the scope of the present study.

3.4. An Alternative Diagnostic for Detecting Chemical Disequilibrium

Line & Yung (2013) previously proposed a simple diagnostic for identifying chemical disequilibrium in an atmosphere, based on measuring the volume mixing ratios associated with the following chemical reaction (e.g., Moses et al. 2011):

$$CH_4 + H_2O \leftrightarrows CO + 3H_2. \tag{10}$$

When rewritten in the formalism of Heng & Tsai (2016), Equation (2) of Line & Yung (2013) is the reciprocal of

$$\frac{X_{CO}}{X_{CH_4} X_{H_2O}} \left(\frac{P}{P_0}\right)^2, \tag{11}$$

where $P_0 = 1$ bar is an arbitrary reference pressure. If the transit chord probed is in chemical equilibrium, then the preceding expression *is* the equilibrium constant,

$$K_{eq} = \exp\left(-\frac{\Delta \tilde{G}_{0,1}}{\mathcal{R}_{univ} T}\right), \tag{12}$$

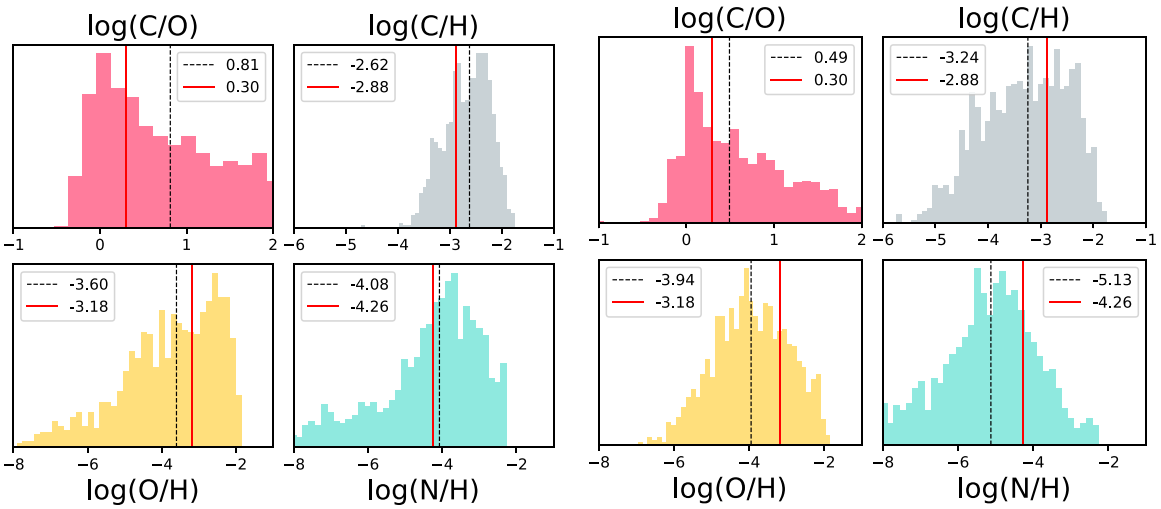

Figure 6. Posterior distributions of C/O, C/H, O/H, and N/H, computed by post-processing the primary posteriors obtained in Figure 5. The left and right montages are for the M1 and M4 modes, respectively. The vertical black dotted lines show the median value of each posterior distribution, which are also indicated numerically in each panel (as the logarithm of each quantity). The vertical red solid lines correspond to the truth values.

where $\mathcal{R}_{\rm univ} = 8.3144621$ J K^{-1} mol^{-1} is the universal gas constant and $\Delta\tilde{G}_{0,1}$ is the molar Gibbs free energy of the reaction (at the reference pressure) tabulated in the JANAF database[14] and listed in, for example, Table 2 of Heng & Lyons (2016).

The key idea proposed by Line & Yung (2013) is to retrieve for the volume mixing ratios ($X_{\rm CO}$, $X_{\rm CH_4}$, $X_{\rm H_2O}$) and obtain an estimate for Equation (11). If this estimate disagrees with $K_{\rm eq}$ (which requires the retrieved temperature as an input), then the region of the atmosphere probed by transmission spectroscopy is in chemical disequilibrium. The major uncertainty with this approach is that the pressure probed in transmission (P) needs to be accurately and precisely known, especially since it appears as the square of itself in Equation (11). See Section 6.3 of Greene et al. (2016) for a critique of Line & Yung (2013).

Using the same concept, we propose to focus on another chemical reaction (e.g., Moses et al. 2011),

$$\rm CO_2 + H_2 \leftrightarrows CO + H_2O, \tag{13}$$

where the corresponding combination of volume mixing ratios has no dependence on pressure (e.g., Heng & Tsai 2016),

$$\frac{X_{\rm CO}X_{\rm H_2O}}{X_{\rm CO_2}}, \tag{14}$$

because the number of molecules associated with the reactants and products is the same. As we will see in Section 4.2, only the M4 mode of JWST NIRSpec is highly sensitive to the presence of CO, CO_2, and H_2O. By retrieving for their mixing ratios and obtaining an estimate for the preceding expression, one may then compare it to the corresponding equilibrium constant,

$$K_{\rm eq,2} = \exp\left(-\frac{\Delta\tilde{G}_{0,2}}{\mathcal{R}_{\rm univ}T}\right), \tag{15}$$

where $\Delta\tilde{G}_{0,2}$ is again listed in Table 2 of Heng & Lyons (2016). In chemical equilibrium, Equations (14) and (15) are equal. Figure 7 shows that $K_{\rm eq,2}$ varies by a factor of about 7 from 800 to 1200 K.

[14] https://janaf.nist.gov

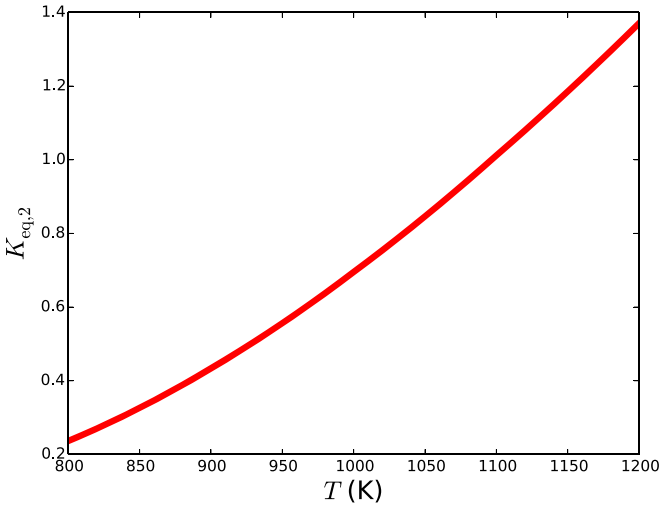

Figure 7. Equilibrium constant associated with the chemical reaction $\rm CO_2 + H_2 \leftrightarrows CO + H_2O$. There is no dependence on pressure.

To accurately employ this diagnostic, the spectra measured using JWST NIRSpec would have to be of a good enough quality to demonstrate that $X_{\rm CO}X_{\rm H_2O}/X_{\rm CO_2}$ is sufficiently different from $K_{\rm eq,2}$. In Figure 8, we show as an illustration the pair of posterior distributions of $X_{\rm CO}X_{\rm H_2O}/X_{\rm CO_2}$ from retrievals on a mock spectrum corresponding to the M1 and M4 modes for the carbon-rich case study considered in Figure 5. The posterior corresponding to the M4 mode firmly excludes the equilibrium value of $K_{\rm eq,2} \approx 0.7$, indicating that the carbon-rich model atmosphere considered is out of chemical equilibrium. The posterior corresponding to the M1 mode is only marginally consistent with the equilibrium value.

4. Discussion

4.1. Comparison to Previous Work

4.1.1. Greene et al. (2016)

Greene et al. (2016) did not perform an IC analysis, but they did study mock retrievals across several exoplanet types

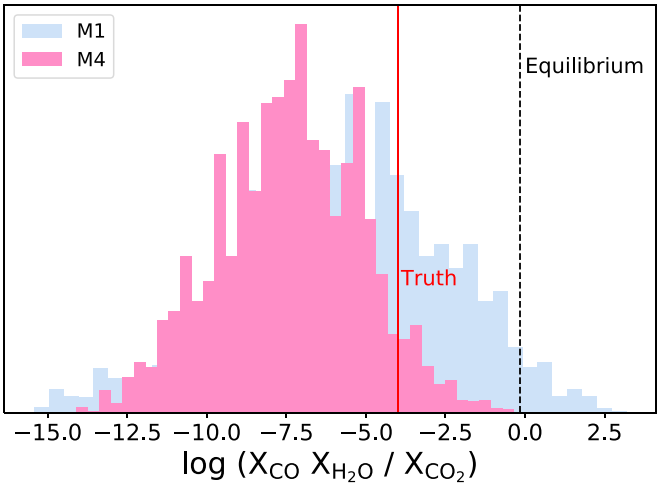

Figure 8. Posterior distributions of the chemical disequilibrium diagnostic corresponding to the carbon-rich case study presented in Figure 5. The solid vertical line is the truth value, while the dashed vertical line is the value in chemical equilibrium (0.695).

(see their Tables 1–3) and JWST modes (see their Table 4), in both emission and transmission. Six molecules were explicitly considered in the mock retrievals: CO, CO_2, H_2O, CH_4, NH_3, and N_2. For transmission spectra, the transit chord was assumed to be isothermal. The cloud model consists of a cloud top pressure (for gray clouds) and a power-law prescription (for nongray clouds consisting of small particles). A key finding of Greene et al. (2016) is the following: "$\lambda = 1$–$2.5\ \mu$m transmission spectra will often constrain the major molecular constituents of clear solar-composition atmospheres well."

The fourth rows of Figures 7 and 8 of Greene et al. (2016) show mock retrievals for a warm Neptune (700 K) and warm sub-Neptune (600 K), respectively, with clouds and solar metallicity. It is worth noting that Greene et al. (2016) have fixed $X_{CO_2} = 3.16 \times 10^{-11}$ and $X_{CO} = 10^{-9}$ for both cases (see their Table 3). While the comparison is imperfect, the M2 mode (0.97–1.84 μm) may be compared to the NIRISS mode (1–2.5 μm) considered by Greene et al. (2016). Our RvP analysis in Figure A2 (Appendix) suggests that CO (with $\mathcal{R}^2 = 0.039$ for the M2 mode) is undetectable across the wavelength range of NIRISS, which is consistent with the unconstrained posterior distributions of X_{CO} obtained by Greene et al. (2016) in the fourth rows of their Figures 7 and 8. Since the contributions of CO and CO_2 to C/O are negligible in both cases, we have

$$\mathrm{C/O} = \frac{X_{CO} + X_{CO_2} + X_{CH_4}}{X_{CO} + 2X_{CO_2} + X_{H_2O}} \approx \frac{X_{CH_4}}{X_{H_2O}}. \quad (16)$$

This explains why the posterior distributions of C/O associated with the 1–2.5 μm versus 1–5 μm retrievals are similar in the fourth rows of Figures 7 and 8 of Greene et al. (2016).

As a further check, the third row of Figure 6 of Greene et al. (2016), which describes a mock retrieval for a hot Jupiter ($X_{CO} \sim 10^{-4}$), shows an unconstrained posterior distribution of X_{CO} associated with 1–2.5 μm. However, the posterior distribution of X_{CO} associated with 1–5 μm is bounded on both sides, consistent with the findings of the current study.

4.1.2. Howe et al. (2017)

Howe et al. (2017) traded model sophistication for a broad exploration of the JWST modes of the NIRcam, NIRISS, NIRSpec, and MIRI instruments (see their Table 1), including the proposal of a set of observing programs for hot Jupiters (see their Table 2). Mock atmospheric retrievals are performed using a Markov Chain Monte Carlo code. Their Table 3 lists the 11 hot Jupiters considered in their study. Figure 7 of Howe et al. (2017) shows examples of calculations of Jacobians with respect to the metallicity, temperature, and pressure. Even though Howe et al. (2017) suggest the use of Jacobians to diagnose cloud properties, they ultimately do not explore this option in their study. For reasons of computational feasibility, Howe et al. (2017) opted for a three-parameter model that explores the temperature (of the isothermal transit chord), metallicity, and cloud top pressure (or, equivalently, a constant cloud opacity).

Howe et al. (2017) remarked, "For our simple forward model, the instrument that consistently gives the most information is the NIRISS G700XD mode," which corresponds to a wavelength range of 0.6–2.8 μm. At face value, the statement about the NIRISS G700XD mode appears to be at odds with the conclusions of the current study that the blue modes of NIRSpec are suboptimal for constraining the elemental abundances and C/O (see Section 4.2). The solution to this conundrum lies in the assumption of chemical equilibrium made by Howe et al. (2017). In chemical equilibrium, knowledge of the elemental abundances, temperature, and pressure allows one to fully specify all of the molecular abundances. Equivalently, one can back out the elemental abundances if the temperature, pressure, and only a subset of the molecular abundances are known.

For example, at a given temperature and pressure one can infer O/H given only X_{H_2O} if chemical equilibrium is assumed (e.g., Heng 2018). It bypasses the need to detect CO or CO_2, which are generally needed, in a chemical disequilibrium situation, for accurately inferring O/H. If the ratios of O/H to the other elemental abundances are further assumed to take on their solar values, then the metallicity may be inferred as a single number (e.g., Heng 2018). Otherwise, the metallicity is generally a set of numbers given by the different elemental abundances. The inferred IC of the 0.6–2.8 μm mode hinges on accepting these assumptions.

4.1.3. Batalha & Line (2017)

Batalha & Line (2017) used an approach to IC analysis that is similar to that of Howe et al. (2017), which is based on computing Jacobians. Their model explorations are based on a WASP-62b-like gas giant, where the main parameters are the temperature, C/O, and metallicity. It is unclear whether the C/H or O/H has a fixed (solar) ratio to the other elemental abundances. The cloud model follows that of Greene et al. (2016). Key conclusions from Batalha & Line (2017) include the following: "A single observation with NIRISS always yields the highest IC content spectra with the tightest constraints, regardless of temperature, C/O, [M/H], cloud effects or precision." As elucidated in Section 4.1.2, this conclusion hinges on the assumption of chemical equilibrium. The temperature range considered by Batalha & Line (2017;

600–1800 K) crosses the transition where chemical equilibrium starts to break down at low temperatures.

4.1.4. Nixon & Madhusudhan (2020)

In a recent study, Nixon & Madhusudhan (2020) assessed the random forest technique for atmospheric retrieval. They compared several retrievals using both random forests and the traditional nested-sampling method. They also added the extension of a likelihood function to the forest to produce posteriors that match the nested-sampling retrievals. The close agreement between their extended random forest and the nested-sampling posteriors is unsurprising, as the same likelihood function is used in both. The agreement implies consistency and not necessarily veracity.

In their comparisons, Nixon & Madhusudhan (2020) show some discrepancies between the standard random forest and the nested-sampling retrievals. In an improvement on the implementation of Márquez-Neila et al. (2018), we have upgraded the trees in the forest to predict the entire set of parameters in the given leaf, as opposed to taking the average value of each leaf (as described in Section 2.3). This gives a more accurate sampling of the posterior. This upgrade is not included in the standard random forest used in Nixon & Madhusudhan (2020) and could account for the discrepancies in their Figure 13.

Nixon & Madhusudhan (2020) also discuss the issue in Cobb et al. (2019), who showed an example where the forest predicts an overconfident, incorrect value of ammonia at the prior minimum in a mock retrieval. As discussed in Section 4.4 and Figure A4 of Fisher et al. (2020), this effect arises from a limitation of the training set used and not because of the random forest. Specifically, because spectroscopic line list data needed to compute the ammonia opacity did not exist above 1500 K, the ammonia mixing ratio was artificially set to 10^{-13} when the temperature crossed this threshold. Fisher et al. (2020) showed that this artifact was also detected using the nested-sampling method. In other words, Cobb et al. (2019) succeeded in identifying the limitation of the training set but drew the wrong conclusion from their findings.

In Section 4 of Nixon & Madhusudhan (2020), it is suggested that the forest cannot be used for a retrieval with many parameters, claiming that "a Random Forest retrieval with n free parameters appears to require $\gtrsim 10^n$ models for an adequate training set." There is in fact no explicit rule for the size of the training set, which will likely depend on many variables such as the relationships between the parameters, the prior ranges, the resolution of the spectra, etc. We found no issues in the current study when using our 11-parameter model on both the WFC3- and JWST-like spectra. One can see from the predicted versus real plots that the forest's performance is quite reasonable given the degeneracies one expects from multiple parameters.

4.2. Recommendations for JWST Observing Proposals

In Figure 9, we consolidate all of our findings into a summary plot that quantifies the predictive power of every JWST NIRSpec mode considered in the current study. Several key points arise from inspecting Figure 9:

1. The three bluest medium-resolution modes (M1, M2, and M3) are essentially blind to CO ($\mathcal{R}^2 \approx 0$), implying that the derived elemental abundances of carbon and oxygen may be inaccurate if CO is a major constituent, data are only available in these modes ($\lesssim$1.8 μm), and the atmospheric abundances are out of chemical equilibrium.
2. All of the modes do equally well at constraining α_0 (which subsumes the cloud abundance) and the cloud particle size (which is constrained by the slope of the spectral continuum) but do equally poorly at identifying cloud composition via constraining the change in slope of the spectral continuum.
3. Perhaps the most surprising finding is that the M4 mode (2.87–5.10 μm) outperforms the low-resolution ($\sim$100) prism mode (0.6–5.3 μm) on the ability to constrain every parameter except for the temperature and ammonia abundance. Both modes constrain the cloud properties equally well (or poorly). In the trade-off between spectral resolution (by a factor of $\sim$10) and wavelength coverage, the former triumphs.
4. While increasing the resolution from $\sim$100 to $\sim$1000 enhances the constraining power substantially, a further increase of resolution to $\sim$2700, corresponding to the high-resolution modes of JWST NIRSpec, adds diminishing value. We demonstrate this by performing an RvP analysis of the M4 mode with a resolution of $\sim$2700, which we label as "H4" in Figure 9. On average, the $\mathcal{R}^2$ value increases by 0.026 or about 5.3% across the 11 parameters. The biggest improvement in $\mathcal{R}^2$ is associated with CO: from 0.508 to 0.577 (increase of 13.6%).

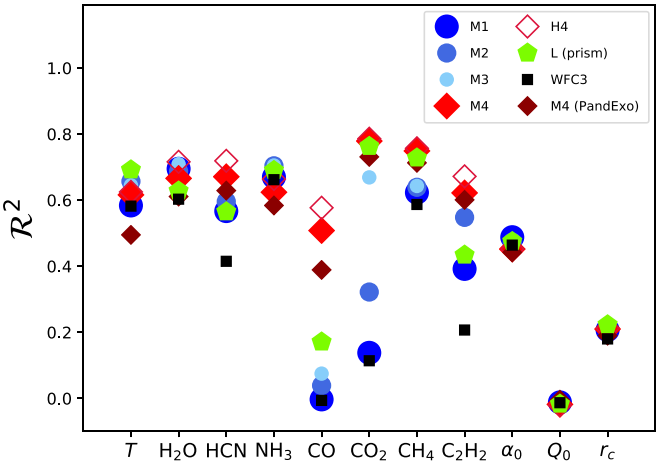

Figure 9. Constraining power of various JWST NIRSpec observing modes as quantified by the coefficient of determination ($\mathcal{R}^2$). See Table 1 for an explanation of the modes and wavelength coverage. Zero predictability and perfect predictability correspond to $\mathcal{R}^2 = 0$ and $\mathcal{R}^2 = 1$, respectively. For comparison, the WFC3 channel (0.8–1.7 μm) of HST is included. The H4 mode covers the same wavelength range as the M4 mode, but at a higher resolution of $\sim$2700.

In this study, we have adopted a fiducial noise model in which every spectral value is sampled with an uncertainty of 20 ppm, which is the optimistic theoretical noise floor of JWST (Beichman et al. 2014). As a sensitivity test, we perform another set of calculations using `PandExo` (Batalha et al. 2017) to simulate a Bright Object Time-Series observation of GJ 436b and to obtain a more realistic noise model for application to the M4 mode. The noise model is simulated by assuming a single transit time series of GJ 436b with the G395M grating and the sub2048 subarray read-out mode. The standard deviation as predicted by `PandExo` varies between 200 and 550 ppm over the M4 wavelength range. Our model spectra that serve as training data are subsequently interpolated

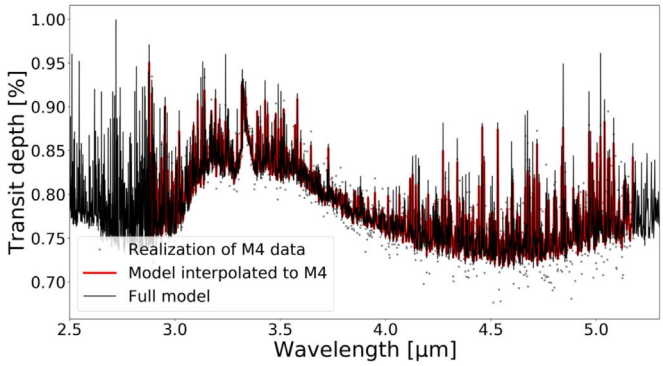

Figure 10. Sample model spectrum from the training set, computed at a resolution of 3000 over the full wavelength range (black) and interpolated onto the wavelength grid of the M4 mode (NIRSpec G395M; red). The gray points denote a single realization of the model with errors sampled from the noise model as computed by `PandExo`.

onto the wavelength grid simulated by `PandExo`. An example model spectrum after interpolation and addition of noise is shown in Figure 10.

Figure 9 shows the constraining power for various model parameters obtained using the different modes, including M4 with the realistic noise model obtained with `PandExo`. Despite the fact that the noise of the realistic model is ∼10 to 20× higher than initially assumed, the qualitative conclusions remain unchanged: the M4 mode's ability (or inability) to constrain the 11 parameters of the model are similar to when 20 ppm uncertainties are assumed. The exception is CO, where the $\mathcal{R}^2$ value drops from 0.508 to 0.389. However, the $\mathcal{R}^2$ value associated with CO_2 remains high: 0.731 versus 0.779.

Overall, we recommend that the medium-resolution M4 mode be used as it offers the most balanced portfolio of constraining power across temperature, molecular abundances, and cloud properties. If the goal is to constrain these parameters accurately in order to infer the elemental abundances and C/O without assuming chemical equilibrium, the medium-resolution M4 mode is sufficient; the corresponding high-resolution mode is unnecessary.

We acknowledge financial support from the Swiss National Science Foundation, the European Research Council (via a Consolidator grant to K.H.; grant No. 771620), the PlanetS National Center of Competence in Research (NCCR), the Center for Space and Habitability (CSH), and the Swiss-based MERAC Foundation. We are grateful to Brice-Olivier Demory for constructive discussions and advice on the manuscript.

Appendix A
Additional Figures

Figure A1 shows various transmission spectra associated with the carbon-rich case study of Section 3.3. It is apparent that the transmission spectra with $X_{CO} = 0$, 10^{-3} and 10^{-2} are very similar. The similarity of these spectra is due to the spectral lines of CO being masked by those of CH_4 and CO_2 at the chosen abundances ($X_{CO} = X_{CH_4} = 10^{-3}$, $X_{CO_2} = 10^{-4}$). The transmission spectrum with $X_{CO} = 0.1$ is markedly different only because CO is so abundant that it changes the mean molecular mass—and hence the pressure scale height—significantly.

For completeness, we include in Figures A2–A7 the RvP and feature importance plots of the M2, M3, and L modes.

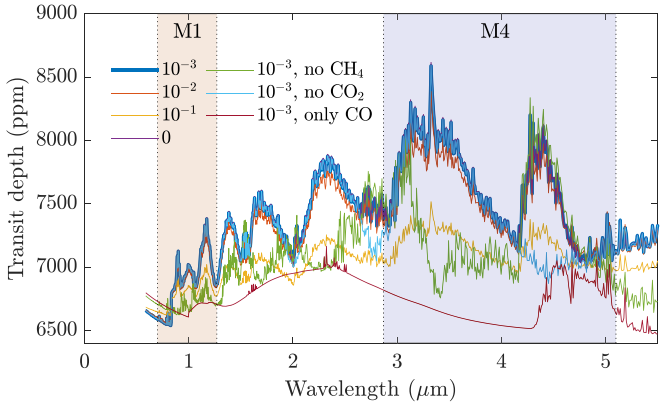

Figure A1. Transmission spectra corresponding to the carbon-rich case study of Figure 5, but with the CO abundance removed (labeled "0") or varied from 10^{-3} (its default value) to 10^{-1}. Three additional curves with CO only, CH_4 removed, and CO_2 removed are included.

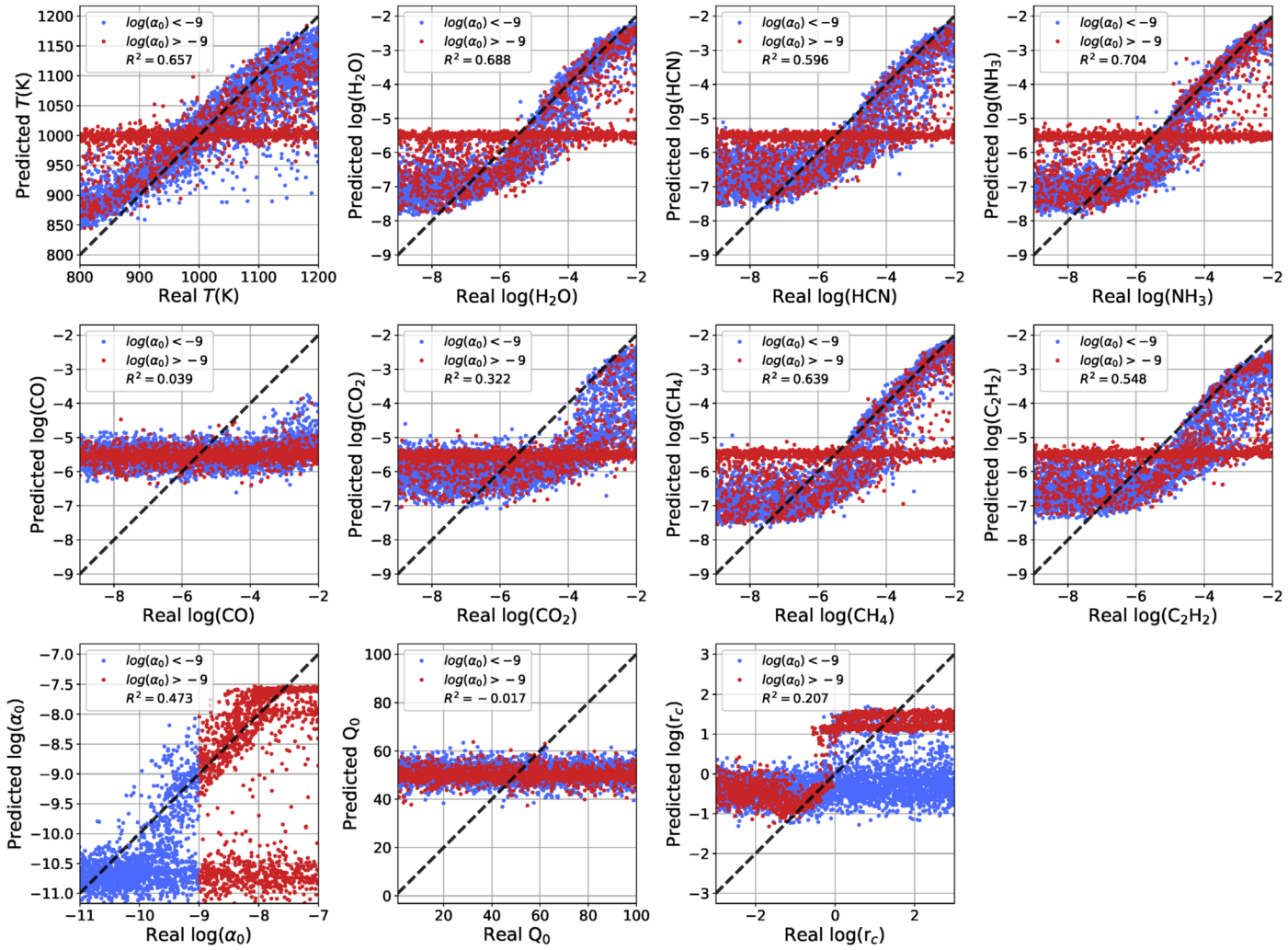

Figure A2. Same as Figure 3, but for the M2 mode.

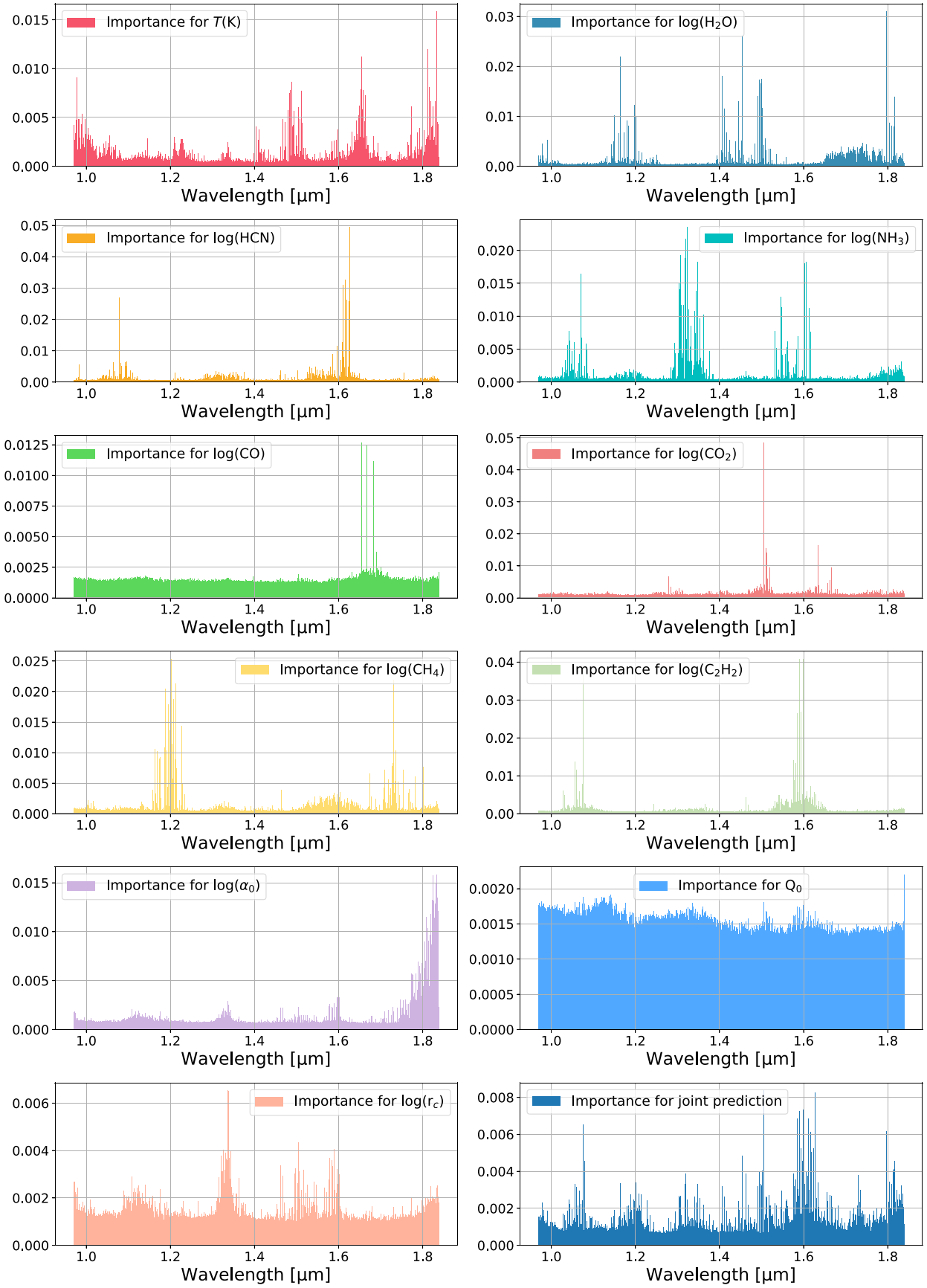

Figure A3. Same as Figure 4, but for the M2 mode.

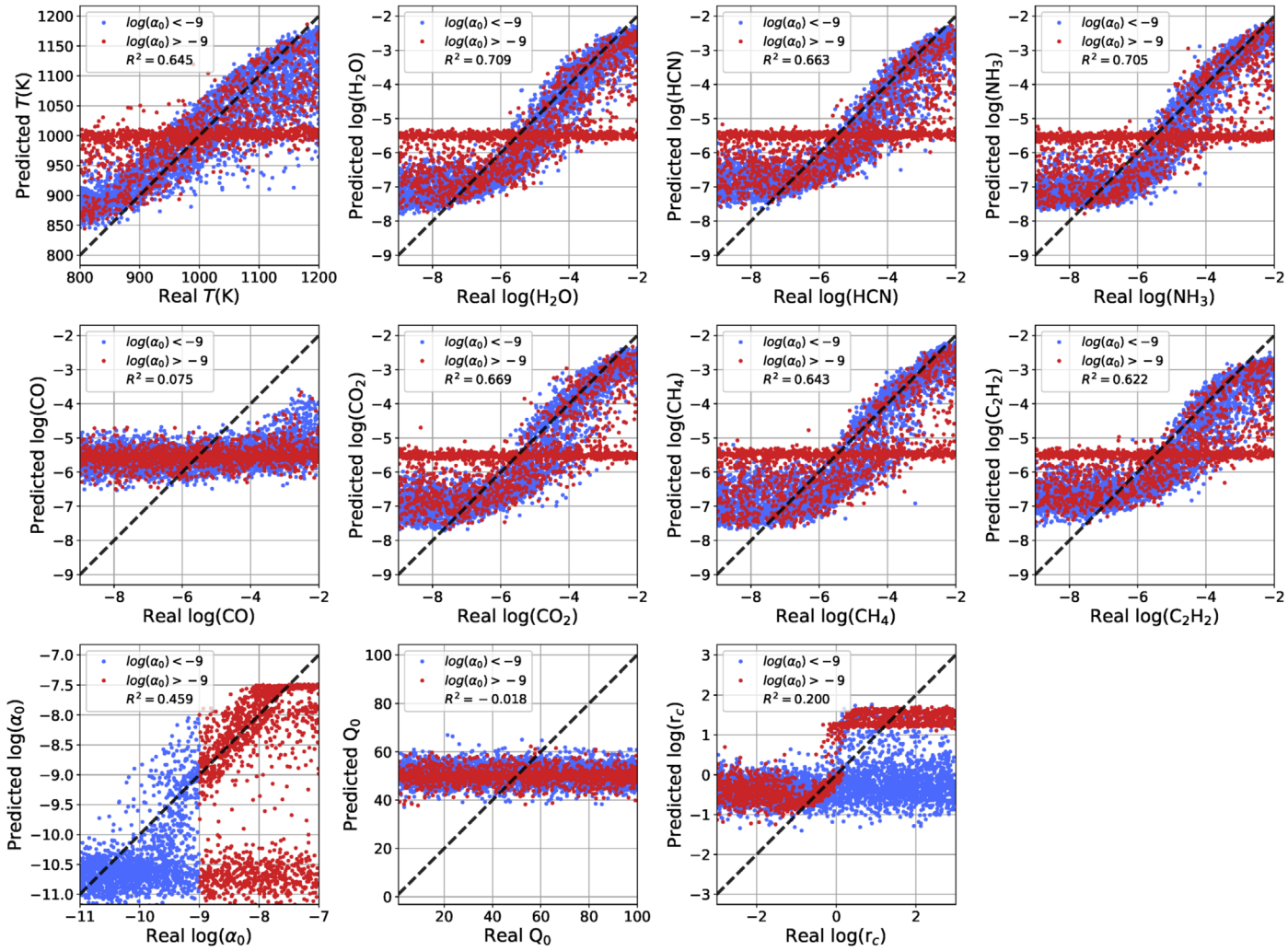

Figure A4. Same as Figure 3, but for the M3 mode.

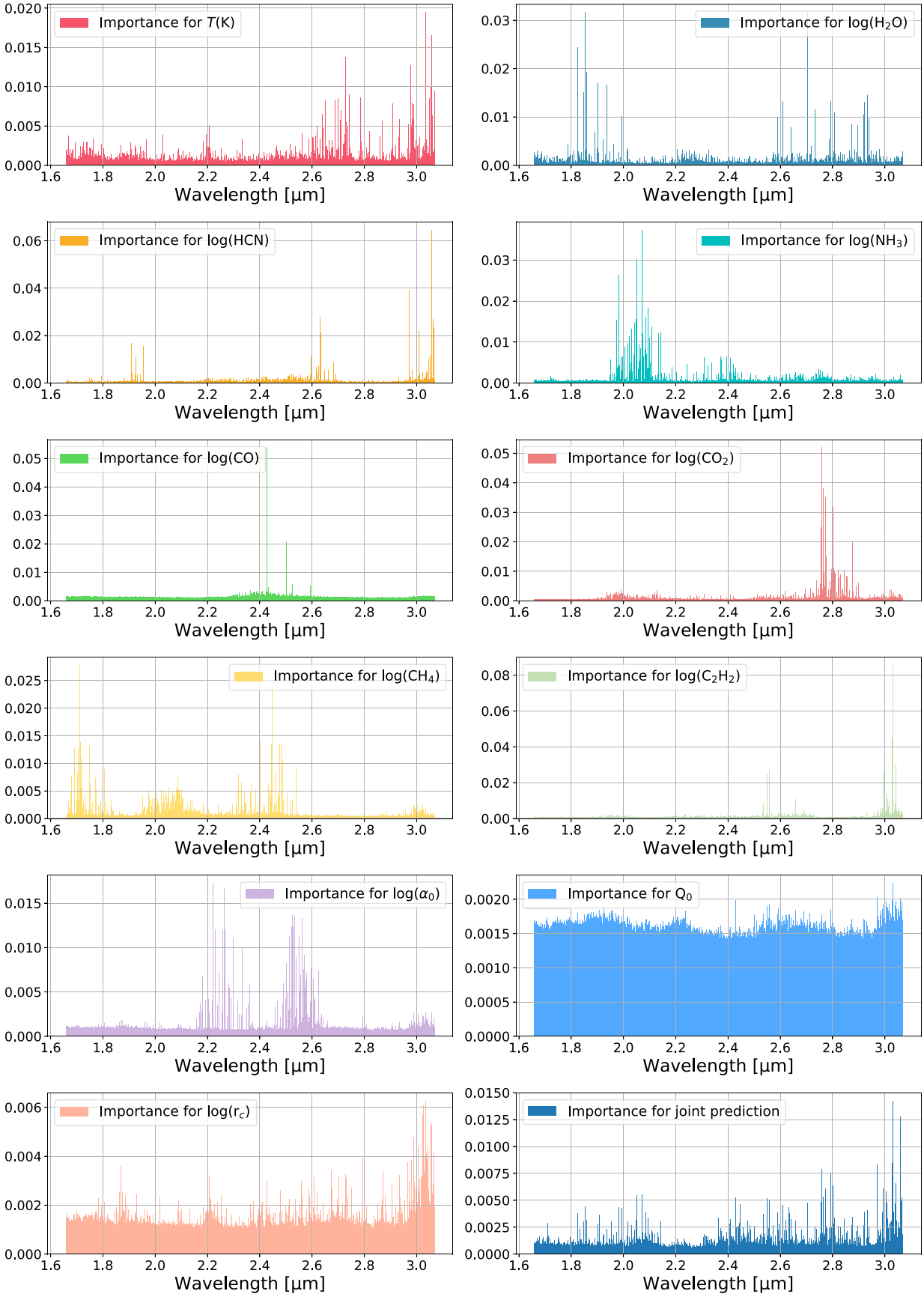

Figure A5. Same as Figure 4, but for the M3 mode.

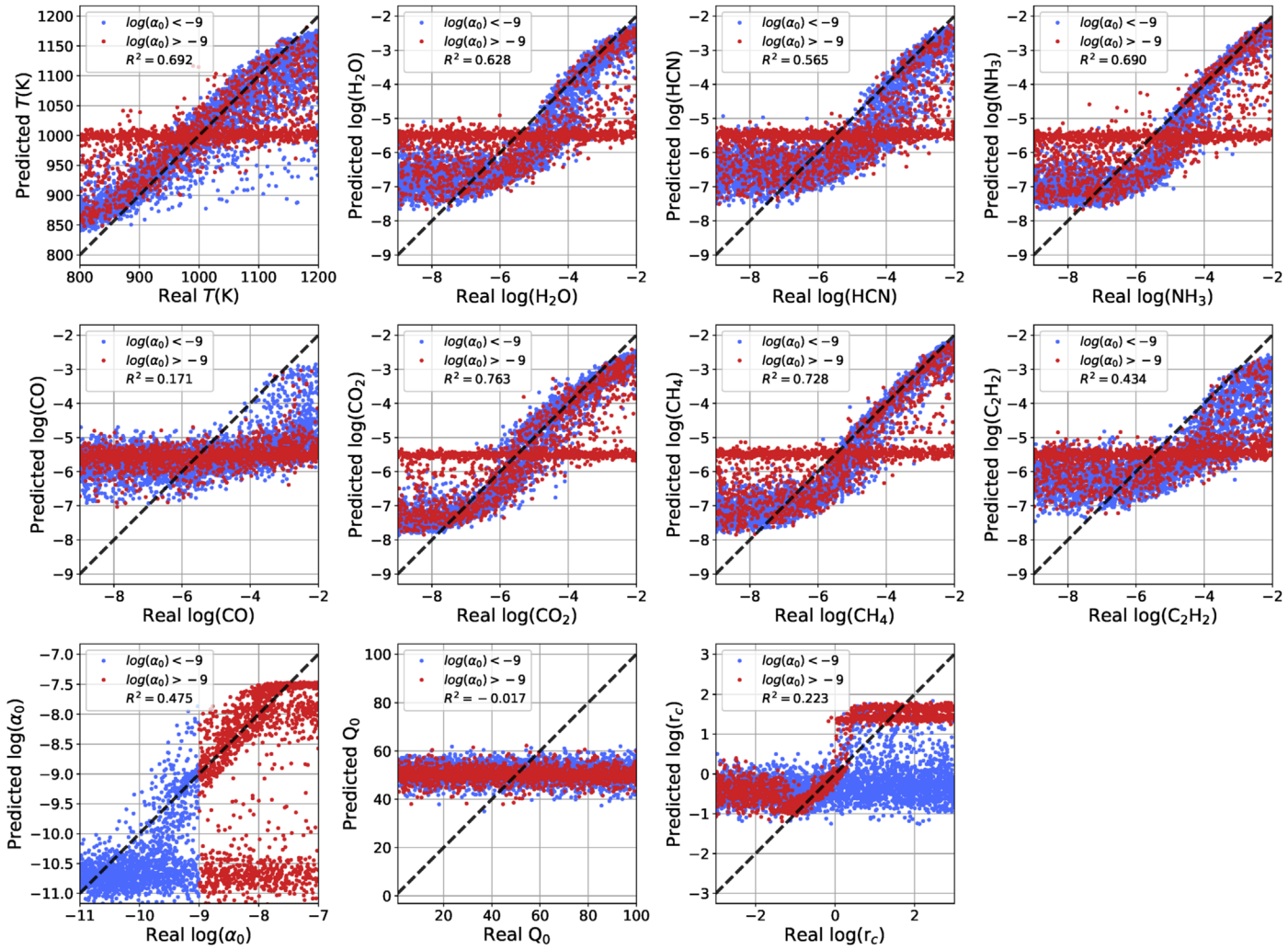

Figure A6. Same as Figure 3, but for the L mode.

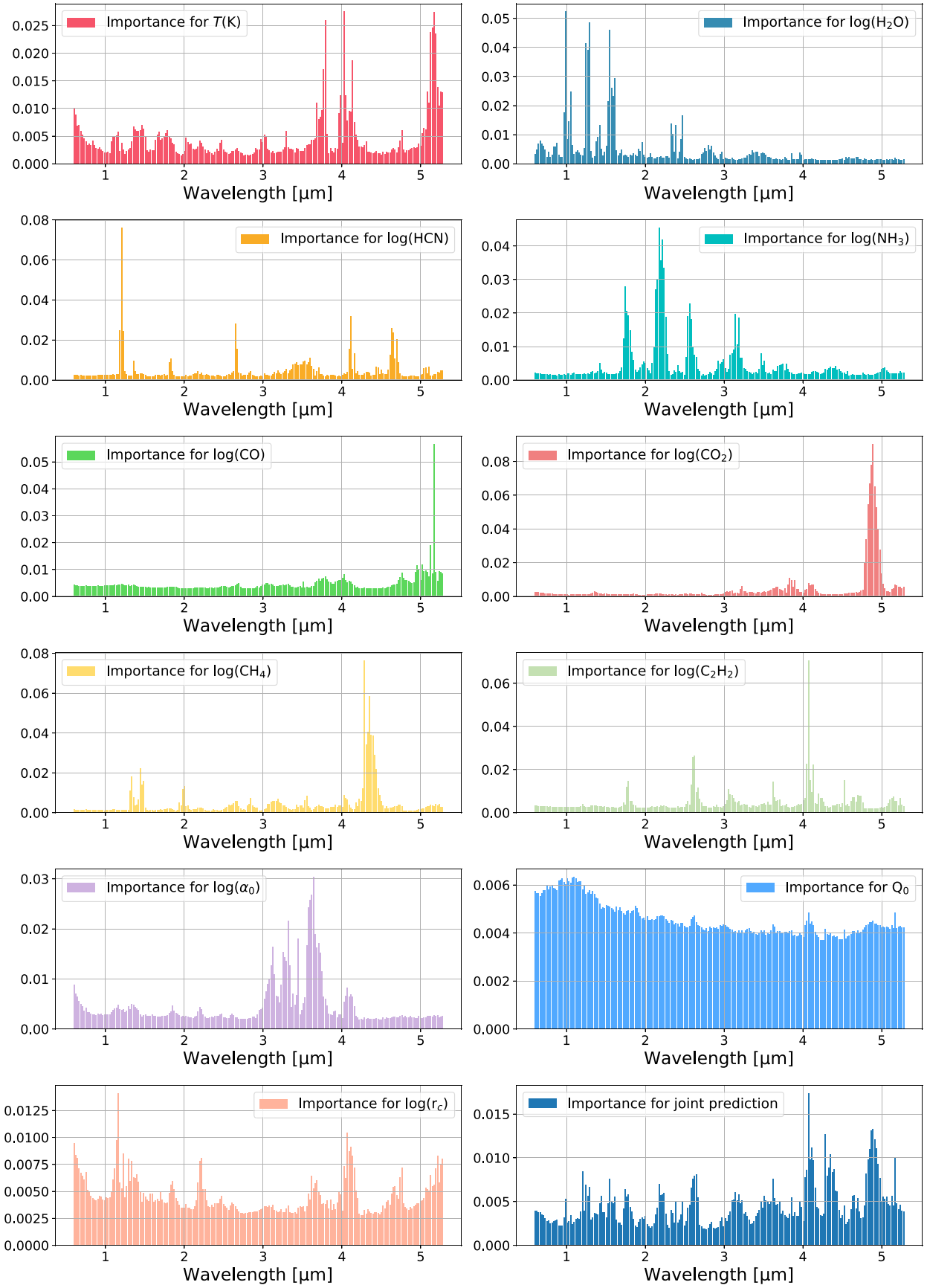

Figure A7. Same as Figure 4, but for the L mode.

ORCID iDs

Andrea Guzmán-Mesa https://orcid.org/0000-0001-5762-0276
Daniel Kitzmann https://orcid.org/0000-0003-4269-3311
Chloe Fisher https://orcid.org/0000-0003-0652-2902
Adam J. Burgasser https://orcid.org/0000-0002-6523-9536
Simon L. Grimm https://orcid.org/0000-0002-0632-4407
Avi M. Mandell https://orcid.org/0000-0002-8119-3355
Kevin Heng https://orcid.org/0000-0003-1907-5910

References

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Barber, R. J., Strange, J. K., Hill, C., et al. 2014, MNRAS, 437, 1828
Barber, R. J., Tennyson, J., Harris, G. J., & Tolchenov, R. N. 2006, MNRAS, 368, 1087
Barstow, J. K., Aigrain, S., Irwin, P. G. J., Kendrew, S., & Fletcher, L. N. 2015, MNRAS, 448, 2546
Batalha, N. E., & Line, M. R. 2017, AJ, 153, 151
Batalha, N. E., Mandell, A., Pontoppidan, K., et al. 2017, PASP, 129, 064501
Bean, J. L., Stevenson, K. B., Batalha, N. M., et al. 2018, PASP, 130, 114402
Beichman, C., Benneke, B., Knutson, H., et al. 2014, PASP, 126, 1134
Benneke, B., & Seager, S. 2012, ApJ, 753, 100
Bower, D. J., Kitzmann, D., Wolf, A. S., et al. 2019, A&A, 631, A103
Breiman, L. 2001, Machine Learning, 45, 5
Brown, T. M. 2001, ApJ, 553, 1006
Burrows, A., & Sharp, C. M. 1999, ApJ, 512, 843
Charbonneau, D., Knutson, H. A., Barman, T., et al. 2008, ApJ, 686, 1341
Cobb, A. D., Himes, M. D., Soboczenski, F., et al. 2019, AJ, 158, 33
Criminisi, A., Konukoglu, E., & Shotton, J. 2011, Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning, Microsoft Research Tech. Rep. TR-2011-114
Crossfield, I. J. M., Ciardi, D. R., Petigura, E. A., et al. 2016, ApJS, 226, 7
Cushing, M. C., Looper, D., Burgasser, A. J., et al. 2009, ApJ, 696, 986
Dragomir, D., Teske, J., Günther, M. N., et al. 2019, ApJL, 875, L7
Drummond, B., Carter, A. L., Hébrard, E., et al. 2019, MNRAS, 486, 1123
Esposito, M., Armstrong, D. J., Gandolfi, D., et al. 2019, A&A, 62, A165
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, OJAp, 2, 10
Fisher, C., & Heng, K. 2018, MNRAS, 481, 4698
Fisher, C., Hoeijmakers, H. J., Kitzmann, D., et al. 2020, AJ, 159, 192
Fortney, J. J., Cooper, C. S., Showman, A. P., Marley, M. S., & Freedman, R. S. 2006, ApJ, 652, 746
Fortney, J. J., Marley, M. S., Lodders, K., Saumon, D., & Freedman, R. 2005, ApJL, 627, L69
Fortney, J. J., Shabram, M., Ahowman, A. P., et al. 2010, ApJ, 709, 1396
Gaidos, E., Kitzmann, D., & Heng, K. 2017, MNRAS, 468, 3418
Gordon, I. E., Rothman, L. S., Hill, C, et al. 2017, JQSRT, 203, 3
Greene, T. P., Line, M. R., Montero, C., et al. 2016, ApJ, 817, 17
Griffith, C. A. 2014, RSPTA, 372, 86
Grimm, S. L., & Heng, K. 2015, ApJ, 808, 182
Harris, G. J., Tennyson, J., Kaminsky, B. M., Pavlenko, Ya. V., & Jones, H. R. A. 2006, MNRAS, 367, 400
Heng, K. 2017, Exoplanetary Atmospheres: Theoretical Concepts and Foundations (Princeton, NJ: Princeton Univ. Press)
Heng, K. 2018, RNAAS, 2, 128
Heng, K. 2019, MNRAS, 490, 3378
Heng, K., & Kitzmann, D. 2017, MNRAS, 470, 2972
Heng, K., & Lyons, J. R. 2016, ApJ, 817, 149
Heng, K., & Tsai, S.-M. 2016, ApJ, 829, 104
Ho, T. K. 1998, ITPAM, 20, 832
Howe, A. R., Burrows, A., & Deming, D. 2017, ApJ, 835, 96
Kilpatrick, B. M., Cubillos, P. E., Stevenson, K. B., et al. 2018, AJ, 156, 103
Kitzmann, D., & Heng, K. 2018, MNRAS, 475, 94
Kitzmann, D., Heng, K., Rimmer, P. B., et al. 2018, ApJ, 863, 183
Kreidberg, L., Line, M. R., Bean, J. L., et al. 2015, ApJ, 814, 66
Lecavelier des Etangs, A., Pont, F., Vidal-Madjar, A., & Sing, D. 2008, A&A, 481, L83
Lee, J.-M., Irwin, P. G. J., Fletcher, L. N., Heng, K., & Barstow, J. K. 2014, ApJ, 789, 14
Li, G., Gordon, I. E., Rothman, L. S., et al. 2015, ApJS, 216, 15
Line, M. R., Wolf, A. S., Zhang, X., et al. 2013, ApJ, 775, 137
Line, M. R., & Yung, Y. 2013, ApJ, 779, 3
Madhusudhan, N. 2012, ApJ, 758, 36
Madhusudhan, N., & Seager, S. 2011, ApJ, 729, 41
Márquez-Neila, P., Fisher, C., Sznitman, R., & Heng, K. 2018, NatAs, 2, 719
Mie, G. 1908, AnP, 330, 377
Morley, C. V., Knutson, H., Line, M., et al. 2017, AJ, 153, 86
Moses, J. I., Line, M. R., Visscher, C., et al. 2013, ApJ, 777, 34
Moses, J. I., Visscher, C., Fortney, J. J., et al. 2011, ApJ, 737, 15
Nixon, M. C., & Madhusudhan, N. 2020, MNRAS, in press (arXiv:2004.10755)
Oppenheimer, B. R., Kulkarni, S. R., Matthews, K., & van Kerkwijk, M. H. 1998, ApJ, 502, 932
Oreshenko, M., Kitzmann, D., Márquez-Neila, P., et al. 2020, AJ, 159, 6
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Petigura, E. A., Marcy, G. W., & Howard, A. W. 2013, ApJ, 770, 69
Polyansky, O. L., Kyuberis, A. A., Zobov, N. F., et al. 2018, MNRAS, 480, 2597
Prinn, R. G., & Barshay, S. S. 1977, Sci, 198, 1031
Quinn, S. N., Becker, J. C., Rodriguez, J. E., et al. 2019, AJ, 158, 177
Rothman, L. S., Barbe, A., Benner, D. C., et al. 2003, JQSRT, 82, 5
Rothman, L. S., Gamache, R. R., Goldman, A., et al. 1987, ApOpt, 26, 4058
Rothman, L. S., Gamache, R. R., Tipping, R. H., et al. 1992, JQSRT, 48, 469
Rothman, L. S., Gordon, I. E., Babikov, Y., et al. 2013, JQSRT, 130, 4
Rothman, L. S., Gordon, I. E., Barbe, A., et al. 2009, JQSRT, 110, 533
Rothman, L. S., Gordon, I. E., Barber, R. J., et al. 2010, JQSRT, 111, 2139
Rothman, L. S., Jacquemar, D., Barbe, A., et al. 2005, JQSRT, 96, 139
Rothman, L. S., Rinsland, C. P., Goldman, A., et al. 1996, JQSRT, 60, 665
Sisson, S. A., Fan, Y., & Beaumont, M. A. 2019, Handbook of Approximate Bayesian Computation (Boca Raton, FL: CRC Press)
Skilling, J. 2006, BayAn, 1, 833
Stevenson, K. B., Lewis, N. K., Bean, J. L., et al. 2016, PASP, 128, 094401
Sudarsky, D., Burrows, A., & Hubeny, I. 2003, ApJ, 588, 1121
Tennyson, J., & Yurchenko, S. N. 2017, MolAs, 8, 1
Torres, G., Winn, J. N., & Holman, M. J. 2008, ApJ, 677, 1324
Trifonov, T., Rybizki, J., & Kürster, M. 2019, A&A, 622, L7
Tsai, S.-M., Lyons, J. R., Grosheintz, L., et al. 2017, ApJS, 228, 20
Venot, O., Agúndez, M., Selsis, F., Tessenyi, M., & Iro, N. 2014, A&A, 562, A51
von Braun, K., Boyajian, T. S., Kane, S. R., et al. 2012, ApJ, 753, 171
Waldmann, I. P. 2016, ApJ, 820, 107
Waldmann, I. P., Tinetti, G., Rocchetto, M., et al. 2015, ApJ, 802, 107
Yurchenko, S. N., Al-Refaie, A. F., & Tennyson, J. 2018, A&A, 614, A131
Yurchenko, S. N., Barber, R. J., & Tennyson, J. 2011, MNRAS, 413, 1828
Yurchenko, S. N., & Tennyson, J. 2014, MNRAS, 440, 1649
Yurchenko, S. N., Tennyson, J., Barber, R. J., & Thiel, W. 2013, JMoSp, 291, 69