



## Reflected Spectroscopy of Small Exoplanets II: Characterization of Terrestrial Exoplanets

Mario Damiano<sup>1</sup> and Renyu Hu<sup>1,2</sup>

<sup>1</sup> Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA; [mario.damiano@jpl.nasa.gov](mailto:mario.damiano@jpl.nasa.gov)

<sup>2</sup> Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA

Received 2022 March 9; accepted 2022 April 25; published 2022 May 31

### Abstract

A space telescope capable of high-contrast imaging has been recognized as the avenue toward finding terrestrial planets around nearby Sun-like stars and characterizing their potential habitability. It is thus essential to quantify the capability of reflected light spectroscopy obtained through direct imaging for terrestrial exoplanets, and existing work focused on planetary analogs of modern Earth. Here we go beyond Earth analogs and use a Bayesian retrieval algorithm, EXOREL<sup>3</sup>, to determine what we could learn about terrestrial exoplanets from their reflected light spectra. Recognizing the potential diversity of terrestrial exoplanets, our focus is to distinguish atmospheric scenarios without any a priori knowledge of the dominant gas. We find that, while a moderate-resolution spectrum in the optical band (0.4–1.0  $\mu\text{m}$ ) may sufficiently characterize a modern Earth analog, it would likely result in incorrect interpretation for planets similar to Archean Earth or having CO<sub>2</sub>-dominated atmospheres. Including observations in the near-infrared bands (1.0–1.8  $\mu\text{m}$ ) can prevent this error, determine the main component (N<sub>2</sub>, O<sub>2</sub>, or CO<sub>2</sub>), and quantify trace gases (H<sub>2</sub>O, O<sub>3</sub>, and CH<sub>4</sub>) of the atmosphere. These results are useful to define the science requirements and design the wavelength bandwidth and observation plans of exoplanet direct imaging missions in the future.

*Unified Astronomy Thesaurus concepts:* [Exoplanet atmospheric composition \(2021\)](#); [Exoplanet atmospheres \(487\)](#); [Bayesian statistics \(1900\)](#); [Bayes' Theorem \(1924\)](#); [Posterior distribution \(1926\)](#); [Prior distribution \(1927\)](#); [Extrasolar rocky planets \(511\)](#)

### 1. Introduction

To characterize the atmosphere, climate, and potential habitability of rocky exoplanets with spectroscopy is a significant endeavor that underpins the exploration of the universe. The rocky exoplanets that have H<sub>2</sub>-dominated atmospheres will be within the reach of the James Webb Space Telescope (JWST) for atmospheric studies through transits (e.g., Deming et al. 2009; Hu et al. 2021a). However, to use the JWST to characterize moderate-temperate and rocky exoplanets that have N<sub>2</sub> or CO<sub>2</sub> atmospheres (i.e., potentially habitable planets that may resemble Earth) will likely require coadding a few tens of transits even for the very best targets due to the small vertical extent of the atmosphere (e.g., Belu et al. 2011; Krissansen-Totton et al. 2018; Wunderlich et al. 2019; Pidhorodetska et al. 2020; Gialluca et al. 2021).

High-contrast imaging of exoplanets is a complementary and recognized avenue to enable the spectroscopic studies of temperate and rocky exoplanets. The level of starlight suppression required to image an Earth-like planet around a Sun-like star has been achieved in laboratories using both the coronagraph (e.g., Trauger & Traub 2007) and star shade (e.g., Harness et al. 2021) technologies. The Nancy Grace Roman Space Telescope (Roman; Spergel et al. 2015; Akeson et al. 2019) will demonstrate the coronagraph technology in space. Studies of future large astrophysics missions commissioned by NASA have indicated that a sizable space telescope coupling with these starlight suppression technologies would be able to find small exoplanets in the habitable zones of nearby stars and study their atmospheres

in the ultraviolet (UV), visible, and near-infrared (NIR) wavelengths (Roberge & Moustakas 2018; Seager et al. 2019; Gaudi et al. 2020; and HabEx Final Report<sup>3</sup> and LUVOIR Final Report<sup>4</sup>). Based on these studies, the Astro2020 decadal survey recommends development toward a potential flagship mission to search for Earth-like habitable exoplanets through direct imaging (National Academies of Sciences et al. 2021).

Earth's spectrum in the UV, visible, and NIR wavelengths contain a wealth of information about the nature of its atmosphere (e.g., Turnbull et al. 2006; Kaltenegger et al. 2007). The spectrum comes from the reflected sunlight and has the absorption features of gases (e.g., H<sub>2</sub>O, O<sub>2</sub>, O<sub>3</sub>) as well as clouds. Feng et al. (2018) applied an inverse retrieval algorithm to the spectrum of modern Earth—as if it is imaged as an exoplanet—to study to what extent one could learn about the atmospheric properties from the spectrum. With an N<sub>2</sub>-dominated atmosphere assumed as the background atmosphere, Feng et al. (2018) demonstrated that the volume mixing ratios of H<sub>2</sub>O, O<sub>2</sub>, O<sub>3</sub> will be constrained from the spectrum. They have also explored a range of signal-to-noise ratios (S/Ns) and spectral resolutions (*R*), and found that to obtain meaningful constraints on the atmosphere, S/N should be at least 15 with a spectral resolution of *R* = 140 and 10 if the spectral resolution is dropped to 70. In the HabEx and LUVOIR final reports, different eons of Earth (e.g. Archean and Proterozoic) have also been discussed, but the evaluation of capabilities of the telescopes was not based on spectral retrievals.

Given the importance of the topic and the scarcity of studies, here we use our Bayesian retrieval method (EXOREL<sup>3</sup>; Damiano & Hu 2020, 2021; Damiano et al. 2020) to explore the possible atmospheric constraints from reflected light spectra of terrestrial



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

<sup>3</sup> <https://www.jpl.nasa.gov/habex/documents/>

<sup>4</sup> <https://asd.gsfc.nasa.gov/luvoir/reports/>

exoplanets. EXOREL<sup>R</sup> preserves the causal relationship between a drop of the mixing ratio of H<sub>2</sub>O and the resulting density of H<sub>2</sub>O clouds, and thus typically yields constraints on the mixing ratio of H<sub>2</sub>O below the clouds (Damiano & Hu 2020, 2021). Also, in our previous work that focuses on the spectral retrieval of sub-Neptunes and water worlds (Damiano & Hu 2021), we made key upgrades (i.e., the centered-log-ratio (CLR) formulation of gas mixing ratios and a new set of prior functions) that allowed us to correctly identify the dominant (i.e., background) gas in the atmosphere. This capability is important because sub-Neptunes may have a diverse atmospheric composition that ranges from H<sub>2</sub> to H<sub>2</sub>O and CO<sub>2</sub> (e.g., Zeng et al. 2019) and is also crucial for the current study.

We use the retrieval algorithm to determine the main component of the atmosphere, rather than assuming the atmosphere to be N<sub>2</sub>-dominated. This is important because we do not know whether terrestrial exoplanets to be found by direct-imaging missions will have N<sub>2</sub>-dominated atmospheres. The rocky planets in the solar system indicate that CO<sub>2</sub>-dominated atmospheres are likely, and atmospheric formation models suggest a wide spectrum of plausible composition (e.g., Elkins-Tanton & Seager 2008; Gaillard et al. 2022). An important objective of this study is to determine whether reflected light spectroscopy can distinguish types of rocky planets (e.g., Earth-like versus Venus-like) and characterize their atmospheric composition. We thus study the ability of spectral retrieval on atmospheric scenarios that correspond to not only modern Earth, but also Archean Earth (e.g., Catling & Zahnle 2020) and planets with CO<sub>2</sub>-dominated atmospheres that resemble Venus and Mars. This is, to our knowledge, the first time a retrieval method is applied to reflected light spectroscopy of small exoplanets beyond modern Earth analogs.

Another important objective of the study is to determine the necessary wavelength coverage to detect and characterize the types of terrestrial exoplanets. We focus on the wavelength coverage, in addition to the S/N and spectral resolution as the observational requirements (e.g., Feng et al. 2018), because it is one of the most important factors that impact the overall architecture of future exoplanet imaging missions. For example, the coronagraph instruments of future missions would probably operate with a wavelength bandpass of  $\sim 20\%$  (e.g., Roberge & Moustakas 2018), and so to obtain a spectrum from 0.4 to, for example,  $1.8 \mu\text{m}$ ,  $\sim 8$  integrations would be necessary—a coverage to only  $1.0 \mu\text{m}$  would halve this requirement. The star shade could result in a wider wavelength bandpass of  $\sim 100\%$  achievable in one shot (Gaudi et al. 2020), but the star shade must be moved to different separations from the telescope to operate between the optical ( $0.4\text{--}1.0 \mu\text{m}$ ) and the NIR band ( $1.0\text{--}1.8 \mu\text{m}$ ), which consumes fuels. Therefore, it is essential to assess whether obtaining the spectrum in the NIR band would be necessary to sufficiently characterize the atmospheres of terrestrial exoplanets.

The paper is organized as follows: Section 2 describes the setup of the retrieval algorithm and the simulation of the atmospheric scenarios. Section 3 shows the results of the retrievals and compares the posterior constraints at varied wavelength coverage, spectral resolution, and S/N for the Earth-like scenario and other atmospheric scenarios. In Section 4, we discuss the results and the implications on the characterization of terrestrial exoplanets through reflected light spectroscopy and the ramification on the design of future missions. Finally, Section 5 summarizes the key findings of this study.

## 2. Methods

### 2.1. Retrieval Setup

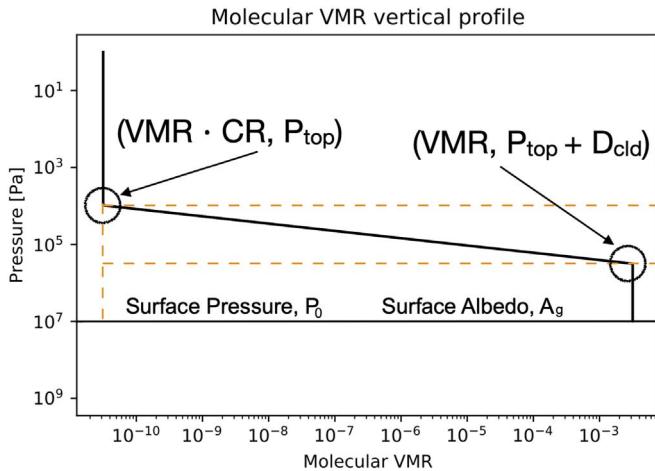
As described in Damiano & Hu (2021), we have updated the radiative transfer model of EXOREL<sup>R</sup> to enable simulations of the reflected light spectra of small exoplanets. We have implemented the CLR (Aitchison 1982) of the mixing ratios of H<sub>2</sub>O, CO<sub>2</sub>, O<sub>2</sub>, O<sub>3</sub>, and CH<sub>4</sub> as free parameters, and N<sub>2</sub> as the filler gas, and applied the set of priors proposed in Damiano & Hu (2021) to ensure that no one gas, including the filler gas, would be preferred as the background gas a priori.

A main upgrade of the model is to include the possibility to detect a planetary surface through the characterization of light reflected by a rocky surface. The current model additionally includes the surface pressure and surface albedo as free parameters (see Figure 1). We assume the surface to be isotropically scattering, with the surface albedo as a free parameter. Similar to Feng et al. (2018), we do not include the wavelength dependency of the surface albedo (Hu et al. 2012a). This aspect will be addressed by a future study. Also, instead of calculating the pressure–temperature profile, we assume the atmosphere to be isothermal and assign an estimated temperature given the irradiation level. This greatly simplifies the forward model and is valid because the reflected light spectra are not sensitive to the specifics of the temperature (Feng et al. 2018; also see Figure 2).

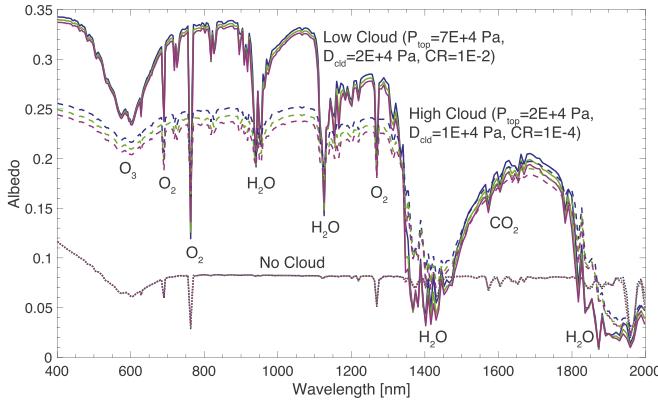
In this study, we include water clouds as the only type of condensates in the atmosphere. This is consistent with the current plans for direct imaging missions. We use the optical properties from Palmer & Williams (1974) to calculate the cross sections and single scattering albedo of water droplets. In this way, our model includes the absorption bands of water clouds at  $1.44$  and  $1.93 \mu\text{m}$ , and gradually more absorption of the clouds outside these bands when the wavelength is longer than  $\sim 1 \mu\text{m}$ . The description of the cloud follows Damiano & Hu (2020), which preserves the causal relationship between the condensation of water and the formation of the cloud (Figure 1). In our model, the cloud is parameterized by the cloud top pressure ( $P_{\text{top}}$ ), the cloud depth (i.e., the difference in pressure between the cloud bottom and the cloud top,  $D_{\text{cld}}$ ), and the condensation ratio (i.e., the ratio between the mixing ratio of water above the cloud and that below the cloud, denoted as CR).

Similar to the test cases presented in Figure 1 of Feng et al. (2018), we have run test cases at the limit of no atmospheric absorption or scattering and confirm that the phase function of the planet to be Lambertian, scaled by the assumed surface albedo.

For a second test, we model the albedo spectra of an Earth-like planet having varied cloud scenarios that correspond to Earth’s cumulus clouds and cirrus clouds (Figure 2). The albedo spectra are highly sensitive to the cloud scenario and are not sensitive to the temperature. The subtle difference in the spectra due to temperature is caused by the cloud particle size’s mild dependency on temperature, which is described in Hu (2019). Absorption features of O<sub>2</sub>, O<sub>3</sub>, H<sub>2</sub>O, and CO<sub>2</sub> are visible, and the descending slope toward longer wavelengths is a combined effect of stronger water absorption and single scattering albedo of cloud particles. The combination of the three components shown in Figure 2 gives a spectrum that resembles Earth’s reflected light spectrum observed in earth-shine (Turnbull et al. 2006).



**Figure 1.** Parameterization of the volume mixing ratio (VMR) vertical profile of the condensing gas (i.e., water vapor) in this study. At the bottom of the atmosphere, we include a reflecting surface characterized by two free parameters: the surface pressure and the surface albedo. The nonconstant VMR allows for the presence of clouds where the gas condenses (i.e., decreases the gas VMR). The four parameters that define the VMR profile and thus the cloud density profile highlighted in the schematics are also free parameters of the model.



**Figure 2.** Modeled spectra of an atmosphere that has a surface pressure of 10<sup>5</sup> Pa (1 bar) and composition similar to modern Earth’s atmosphere. Solid, dashed, and dotted lines correspond to low cloud, high cloud, no cloud scenarios, and the three colors within each group correspond to an assumed temperature of 200, 250, and 300 K. The orbital phase is  $\pi/3$ . The surface albedo is 0.2. The spectra are not sensitive to the assumed temperature.

## 2.2. Simulated Atmospheric Scenarios

Focusing on the atmospheric characterization of terrestrial exoplanets, we simulate a  $1M_{\oplus}$ ,  $1R_{\oplus}$  planet at 1 au from a Sun-like star. We surround the planet with different atmospheric scenarios to explore the capability of reflected light spectroscopy.

We have studied four atmospheric scenarios in this work. (1) An Earth-like atmosphere: Similar to Feng et al. (2018), we also consider a rocky planet having an atmosphere skin to modern Earth’s atmosphere, which is an N<sub>2</sub>-dominated atmosphere ( $\sim 79\%$ ) with the presence of O<sub>2</sub> ( $\sim 20\%$ ) and other minor gases including H<sub>2</sub>O, O<sub>3</sub>, CH<sub>4</sub>, and CO<sub>2</sub> (Tables 1 and 2). (2) An Archean Earth-like atmosphere: Earth’s atmosphere has gone through multiple evolution stages and had a very different composition in the past (e.g., Catling & Zahnle 2020). The paleoatmosphere during the Archean eon is particularly interesting because Archean corresponds to  $>30\%$  of Earth’s history and the atmosphere at the time had a low

mixing ratio of O<sub>2</sub> but much more CO<sub>2</sub> and CH<sub>4</sub> than the present day (Table 3). (3) A CO<sub>2</sub>-dominated atmosphere with O<sub>2</sub> and clouds: In this scenario we start with modern Earth’s atmosphere but swap the mixing ratios between N<sub>2</sub> and CO<sub>2</sub>, effectively creating a hypothetical CO<sub>2</sub>-dominated atmosphere with O<sub>2</sub> (Table 4). The motivation of this scenario is twofold. First, we want to test whether the reflected light spectrum would tell us that the planet should have a CO<sub>2</sub>-dominated atmosphere, simulating the detection of a Venus-like planet. Second, we want to study whether O<sub>2</sub>, a main biosignature gas (e.g., Schwertner et al. 2018), would still be detectable in such an atmosphere. (4) A dry CO<sub>2</sub>-dominated atmosphere: The last scenario we simulate is an arid planet with a CO<sub>2</sub>-dominated atmosphere ( $\sim 99\%$ ). As an end-member, we consider the case without any water vapor in the atmosphere and thus no water clouds either (Table 5). This scenario may correspond to a rocky planet that has lost its ocean (e.g., Wordsworth & Pierrehumbert 2013), or a rocky planet that is too cold for water to exist in liquid or vapor forms with appreciable amounts (such as Mars). We consider minor presence of CH<sub>4</sub> and O<sub>2</sub> to assess the detectability of these gases in such an atmosphere.

In all scenarios, we study the impact of including and excluding the NIR portion of the reflected light spectra in the retrieval. In this work, we adopt the “optical” wavelengths to 0.4–1.0  $\mu\text{m}$  and the “NIR” wavelengths to 1.0–1.8  $\mu\text{m}$ , and adopt a baseline spectral resolution of  $R = 140$  and  $R = 40$  for the optical and NIR bands, respectively. These assumptions are approximately consistent with the large mission studies (Roberge & Moustakas 2018; Gaudi et al. 2020). Additionally for the Earth-like scenario, we explore a S/N of 5, 10, and 20, as well as the spectral resolution in the optical band between 70 and 140. This exploration allows us to compare with previous work (Feng et al. 2018) and better understand the observational capabilities required to extract meaningful information from the spectrum.

Lastly, to simulate the size of the error associated to each of the data points, we considered the maximum value of the spectrum and we divided it by the desired S/N. This effectively assumes that the noise is dominated by the background and not the planet, likely valid for spectroscopy of terrestrial planets (e.g., Hu et al. 2021b). We then added a Gaussian deviation to the data to simulate the random realization of the observation.

## 3. Results

### 3.1. Earth-like Scenario

The retrieval results are reported in Table 1, and the posterior distributions are shown in Appendix A. The simulated data and the best-fit spectra are shown in Figure 3.

The retrieval algorithm is able to successfully determine the basic nature of the planet (i.e., having an N<sub>2</sub>/O<sub>2</sub>-dominated atmosphere with H<sub>2</sub>O and O<sub>3</sub>) with the optical and the NIR bands, except for the case of S/N = 5 and  $R = 70$ . In this minimal-quality case, the noise realization dominates the spectral features and the retrieval algorithm mistakenly returns a scenario closer to a water world (i.e., having an H<sub>2</sub>O-dominated atmosphere). For the other cases, the algorithm returns a fit that corresponds to the input atmospheric composition. The abundance of H<sub>2</sub>O and O<sub>3</sub> is generally well constrained, but that of CH<sub>4</sub> is only moderately constrained in the best-case scenario with S/N = 20 and  $R = 140$ . The

**Table 1**  
Atmospheric Parameters Used to Simulate the Earth-like Scenario and the Retrieval Results

Parameter	Input	S/N = 5, R = 70	S/N = 10, R = 70	S/N = 10, R = 140	S/N = 20, R = 140
Log( $P_0$ ) [Pa]	5.00	$7.25^{+3.39}_{-3.08}$	$7.35^{+3.22}_{-2.65}$	$8.00^{+2.70}_{-3.23}$	$6.84^{+3.47}_{-1.98}$
Log( $P_{\text{top,H}_2\text{O}}$ ) [Pa]	4.85	$1.50^{+2.58}_{-1.39}$	$4.03^{+0.80}_{-2.23}$	$3.96^{+3.08}_{-1.06}$	$4.26^{+0.50}_{-0.41}$
Log( $D_{\text{cl},\text{H}_2\text{O}}$ ) [Pa]	4.30	$4.97^{+1.59}_{-1.73}$	$4.42^{+0.44}_{-0.77}$	$4.34^{+2.56}_{-0.39}$	$4.48^{+0.28}_{-0.26}$
Log(CR <sub>H<sub>2</sub>O</sub> )	-3.00	$-4.66^{+3.67}_{-6.52}$	$-8.45^{+5.25}_{-3.26}$	$-8.30^{+5.02}_{-3.36}$	$-8.21^{+4.75}_{-3.27}$
Log(VMR <sub>H<sub>2</sub>O</sub> )	-2.01	$-0.08^{+0.08}_{-1.22}$	$-1.29^{+1.23}_{-0.62}$	$-1.46^{+0.89}_{-2.47}$	$-1.69^{+0.44}_{-0.47}$
Log(VMR <sub>CH<sub>4</sub></sub> )	-5.96	$-6.41^{+2.70}_{-3.62}$	$-6.97^{+2.38}_{-3.29}$	$-7.32^{+2.60}_{-3.38}$	$-5.22^{+0.73}_{-3.71}$
Log(VMR <sub>CO<sub>2</sub></sub> )	-3.40	$-3.81^{+3.54}_{-5.95}$	$-3.23^{+2.13}_{-5.42}$	$-5.53^{+3.56}_{-4.92}$	$-4.36^{+1.97}_{-4.14}$
Log(VMR <sub>O<sub>2</sub></sub> )	-0.71	$-2.39^{+2.31}_{-6.47}$	$-0.03^{+0.03}_{-1.23}$	$-0.04^{+0.03}_{-4.18}$	$-0.34^{+0.23}_{-0.59}$
Log(VMR <sub>O<sub>3</sub></sub> )	-5.96	$-4.99^{+0.91}_{-0.95}$	$-5.59^{+0.81}_{-0.47}$	$-5.43^{+0.25}_{-2.44}$	$-5.66^{+0.17}_{-0.36}$
Log(VMR <sub>N<sub>2</sub></sub> )	-0.10	$-2.63^{+2.59}_{-5.98}$	$-3.84^{+3.40}_{-5.47}$	$-1.90^{+1.90}_{-4.86}$	$-0.29^{+0.23}_{-0.45}$
$A_g$	0.05	$0.53^{+0.42}_{-0.47}$	$0.53^{+0.42}_{-0.47}$	$0.46^{+0.48}_{-0.42}$	$0.40^{+0.50}_{-0.35}$
Log( $g$ [cgs])	2.99	$3.00^{+0.04}_{-0.04}$	$3.01^{+0.03}_{-0.03}$	$3.00^{+0.03}_{-0.04}$	$2.99^{+0.02}_{-0.01}$
$\mu$	28.70	$20.46^{+13.09}_{-2.44}$	$31.26^{+0.98}_{-11.40}$	$30.94^{+0.79}_{-4.25}$	$29.62^{+1.24}_{-1.31}$

**Note.** The spectral resolution  $R$  is applied to the optical band of  $0.4\text{--}1.0\,\mu\text{m}$ , and the spectrum in the NIR band of  $1.0\text{--}1.8\,\mu\text{m}$  with  $R = 40$  and the same S/N as listed in the table is also included in the retrieval. The error bars of the retrieval results correspond to the 95% confidence interval (i.e.,  $2\sigma$ ). Note—the column of S/N = 10,  $R = 140$  reports the weighted average of the two degenerate solutions found by the retrieval.

abundance of CO<sub>2</sub> is not well constrained in any of the scenarios. The cloud pressures are reasonably well constrained, but the surface pressure (except that it must be greater than the cloud pressure) or the surface albedo cannot be constrained. We see that the constraints of these parameters are the tightest in the case of S/N = 20 and  $R = 140$ , as expected.

Interestingly, the case with S/N = 10 and  $R = 140$  (i.e., high resolution with poor S/N) results in a degenerate solution that involves lower clouds and a lower mixing ratio of O<sub>2</sub> than the truth. While this degenerate solution may be deemed unphysical (e.g., the retrieval still indicates a significant mixing ratio of O<sub>3</sub>), this finding suggests that it may be advisable to test the sensitivity of the retrieval results on multiple choices of spectral binning.

We use the case  $R = 140$  and S/N = 20 to explore the effect of the NIR part of the spectrum ( $1.0\text{--}1.8\,\mu\text{m}$ ) on the atmospheric constraints. The retrieval results with and without the NIR band are compared in Table 2, and the posterior distributions are compared in Appendix A. Without the NIR band but with the “high-quality” option (S/N = 20 and  $R = 140$ ) for the optical band, the retrieval algorithm returns a solution that is moderately different from the truth—an O<sub>2</sub>-dominated atmosphere, with large biases in the VMR of H<sub>2</sub>O and the cloud properties. The constraints on the VMR of O<sub>3</sub> remain solid. The degradation in the retrieval quality ultimately comes from the fact that the NIR band contains information about the pressure of clouds, as the water droplet is not fully reflective in the NIR band and can cause an underlying slope outside of the gas absorption bands (see Figure 2). From a bigger-picture standpoint, we feel that this interpretation can be deemed “somewhat correct” as it still calls for an O<sub>2</sub>/O<sub>3</sub>-rich atmosphere with H<sub>2</sub>O.

### 3.2. Archean Earth-like Scenario

The simulated data and the best-fit models calculated by EXOREL<sup>3</sup> are shown in Figure 4, and the full posterior distributions are reported in Appendix B. The synthesized spectrum shows more deep absorption features in the red end of the optical and the NIR band compared to the Earth-like

scenario, and this is due to the higher mixing ratios of CH<sub>4</sub> and CO<sub>2</sub> in the atmosphere.

The retrieval algorithm correctly characterizes the atmosphere by determining that N<sub>2</sub> is the dominant gas and that the atmosphere has high mixing ratios of CO<sub>2</sub> and CH<sub>4</sub>, with the full spectrum that includes both the optical and NIR bands. The retrieval algorithm also correctly determines that the atmosphere has low amounts of O<sub>2</sub> and O<sub>3</sub>. The pressure of the water cloud, as well as the mixing ratio of water vapor below the cloud, have also been correctly retrieved within  $1\sigma$ .

Without the NIR band, the nature of the atmosphere is not correctly characterized. In this case (see Appendix B), EXOREL<sup>3</sup> reports an H<sub>2</sub>O-dominated atmosphere with a high cloud (at pressures  $<\sim 0.01$  bar). Because of this, the retrieved mixing ratio of CH<sub>4</sub> is biased toward larger values. The retrieval algorithm cannot detect CO<sub>2</sub> but rather returns an erroneous upper limit of  $\sim 1\%$ . Evidently, the retrieval algorithm tends to misinterpret the spectrum in the Archean Earth case without the NIR band. This is because the absorption features of gaseous CO<sub>2</sub> and the clouds of H<sub>2</sub>O are both in the NIR band—without these “anchors” the retrieval algorithm would return a wrong solution.

### 3.3. CO<sub>2</sub>-dominated Atmosphere with O<sub>2</sub> and Clouds

The simulated data and the best-fit models calculated by EXOREL<sup>3</sup> are shown in Figure 5, and the full posterior distribution is reported in Appendix C. The spectrum is similar to the Earth-like scenario in the optical band but has a different shape in the NIR band mainly due to strong CO<sub>2</sub> absorption.

The retrieval algorithm is able to correctly characterize the atmosphere of the simulated planet from the spectrum that includes the optical and the NIR bands. The posteriors clearly show a CO<sub>2</sub>-dominated atmosphere with O<sub>2</sub> as the second most dominant gas. The abundance of H<sub>2</sub>O and O<sub>3</sub> and the cloud pressures are well constrained, and the abundance of CH<sub>4</sub> is poorly (but correctly) constrained. The retrieval algorithm indicates that the data can tolerate an abundance of N<sub>2</sub> (a spectrally inactive gas) that is much higher than the input abundance, and thus the calculated mean molecular mass  $\mu$  is

**Table 2**

Atmospheric Parameters Used to Simulate the Earth-like Scenario and Retrieval Results for the Full Wavelength Band and only the Optical Band

Parameter	Input	Optical + NIR	Optical only
$\text{Log}(P_0)$ [Pa]	5.00	$6.84^{+3.47}_{-1.98}$	$4.39^{+0.13}_{-0.12}$
$\text{Log}(P_{\text{top},\text{H}_2\text{O}})$ [Pa]	4.85	$4.26^{+0.50}_{-0.41}$	$2.07^{+4.60}_{-1.89}$
$\text{Log}(D_{\text{cld},\text{H}_2\text{O}})$ [Pa]	4.30	$4.48^{+0.28}_{-0.26}$	$1.79^{+4.43}_{-1.63}$
$\text{Log}(\text{CR}_{\text{H}_2\text{O}})$	-3.00	$-8.21^{+4.75}_{-3.27}$	$-6.57^{+6.07}_{-4.93}$
$\text{Log}(\text{VMR}_{\text{H}_2\text{O}})$	-2.01	$-1.69^{+0.44}_{-0.47}$	$-3.21^{+1.76}_{-0.29}$
$\text{Log}(\text{VMR}_{\text{CH}_4})$	-5.96	$-5.22^{+0.73}_{-3.71}$	$-6.34^{+1.85}_{-2.83}$
$\text{Log}(\text{VMR}_{\text{CO}_2})$	-3.40	$-4.36^{+1.97}_{-4.14}$	$-3.39^{+2.25}_{-4.00}$
$\text{Log}(\text{VMR}_{\text{O}_2})$	-0.71	$-0.34^{+0.23}_{-0.59}$	$-0.00^{+0.00}_{-0.05}$
$\text{Log}(\text{VMR}_{\text{O}_3})$	-5.96	$-5.66^{+0.17}_{-0.36}$	$-5.52^{+0.14}_{-0.14}$
$\text{Log}(\text{VMR}_{\text{N}_2})$	-0.10	$-0.29^{+0.23}_{-0.45}$	$-3.65^{+2.58}_{-4.64}$
$A_g$	0.05	$0.40^{+0.50}_{-0.35}$	$0.64^{+0.31}_{-0.24}$
$\text{Log}(g)$ [cgs]	2.99	$2.99^{+0.02}_{-0.01}$	$2.89^{+0.17}_{-0.20}$
$\mu$	28.70	$29.62^{+1.24}_{-1.31}$	$31.99^{+0.83}_{-0.63}$

**Note.** The error bars of the retrieval results correspond to the 95% confidence interval (i.e.,  $2\sigma$ ).

**Table 3**

Atmospheric Parameters Used to Simulate the Archean Earth-like Scenario and the Retrieval Results

Parameter	Input	Optical + NIR	Optical only
$\text{Log}(P_0)$ [Pa]	5.00	$7.39^{+3.19}_{-2.21}$	$7.81^{+2.75}_{-3.68}$
$\text{Log}(P_{\text{top},\text{H}_2\text{O}})$ [Pa]	4.85	$4.79^{+0.63}_{-0.33}$	$2.28^{+1.28}_{-1.58}$
$\text{Log}(D_{\text{cld},\text{H}_2\text{O}})$ [Pa]	4.30	$4.58^{+0.52}_{-0.33}$	$3.51^{+0.48}_{-0.27}$
$\text{Log}(\text{CR}_{\text{H}_2\text{O}})$	-4.00	$-9.00^{+4.24}_{-2.73}$	$-9.08^{+4.43}_{-2.64}$
$\text{Log}(\text{VMR}_{\text{H}_2\text{O}})$	-1.00	$-0.72^{+0.43}_{-0.69}$	$-0.12^{+0.09}_{-0.14}$
$\text{Log}(\text{VMR}_{\text{CH}_4})$	-2.00	$-2.03^{+0.30}_{-0.60}$	$-0.67^{+0.26}_{-0.59}$
$\text{Log}(\text{VMR}_{\text{CO}_2})$	-1.00	$-0.71^{+0.40}_{-0.97}$	$-4.94^{+3.47}_{-4.41}$
$\text{Log}(\text{VMR}_{\text{O}_2})$	-7.00	$-5.30^{+2.15}_{-2.81}$	$-4.43^{+2.65}_{-4.42}$
$\text{Log}(\text{VMR}_{\text{O}_3})$	...	$-8.09^{+0.82}_{-1.13}$	$-7.21^{+1.22}_{-1.88}$
$\text{Log}(\text{VMR}_{\text{N}_2})$	-0.10	$-0.25^{+0.20}_{-0.32}$	$-2.62^{+1.90}_{-3.41}$
$A_g$	0.05	$0.64^{+0.33}_{-0.55}$	$0.45^{+0.47}_{-0.40}$
$\text{Log}(g)$ [cgs]	2.99	$2.99^{+0.01}_{-0.01}$	$2.98^{+0.01}_{-0.02}$
$\mu$	28.49	$28.98^{+5.33}_{-4.60}$	$17.75^{+2.30}_{-0.46}$

**Note.** The error bars of the retrieval results correspond to the 95% confidence interval (i.e.,  $2\sigma$ ).

lower than the expected value. However, this does not affect the overall interpretation of the nature of the atmosphere.

Without the NIR band, however, the retrieval algorithm significantly misinterprets the nature of the atmosphere. In this case, the retrieval algorithm mistakenly reports a 100 bar  $\text{N}_2$ -dominated atmosphere, without clouds, and with mixing ratios of  $\text{H}_2\text{O}$ ,  $\text{O}_2$ , and  $\text{O}_3$  smaller than the true values by several orders of magnitudes. The retrieval algorithm also mistakenly reports an upper limit of  $\text{CO}_2$  of approximately 1% and suggests the planet would have a highly reflective surface. Evidently the spectral interpretation without the NIR band would be wrong, and this is again because the absorption features of gaseous  $\text{CO}_2$  and the clouds of  $\text{H}_2\text{O}$  are both in the NIR band.

### 3.4. Dry $\text{CO}_2$ -dominated Atmosphere

The simulated data and the best-fit models calculated by EXOREL<sup>R</sup> are shown in Figure 6. The spectrum on the blue end

**Table 4**

Atmospheric Parameters Used to Simulate the  $\text{CO}_2$ -dominated Atmosphere with  $\text{O}_2$  and Clouds Scenario and the Retrieval Results

Parameter	Input	Optical + NIR	Optical only
$\text{Log}(P_0)$ [Pa]	5.00	$5.80^{+2.77}_{-0.60}$	$6.98^{+0.23}_{-0.24}$
$\text{Log}(P_{\text{top},\text{H}_2\text{O}})$ [Pa]	4.85	$4.88^{+0.25}_{-0.26}$	$3.05^{+2.98}_{-2.83}$
$\text{Log}(D_{\text{cld},\text{H}_2\text{O}})$ [Pa]	4.30	$4.84^{+0.19}_{-0.27}$	$3.11^{+2.97}_{-2.87}$
$\text{Log}(\text{CR}_{\text{H}_2\text{O}})$	-3.00	$-8.33^{+4.37}_{-3.29}$	$-5.46^{+5.06}_{-6.03}$
$\text{Log}(\text{VMR}_{\text{H}_2\text{O}})$	-2.01	$-2.14^{+0.65}_{-0.50}$	$-6.39^{+0.27}_{-0.28}$
$\text{Log}(\text{VMR}_{\text{CH}_4})$	-5.96	$-7.07^{+1.40}_{-1.52}$	$-8.71^{+1.75}_{-2.18}$
$\text{Log}(\text{VMR}_{\text{CO}_2})$	-0.10	$-0.32^{+0.15}_{-0.34}$	$-7.12^{+4.64}_{-3.60}$
$\text{Log}(\text{VMR}_{\text{O}_2})$	-0.71	$-1.16^{+0.25}_{-0.46}$	$-3.83^{+0.23}_{-0.25}$
$\text{Log}(\text{VMR}_{\text{O}_3})$	-5.96	$-6.21^{+0.12}_{-0.22}$	$-8.01^{+0.20}_{-0.19}$
$\text{Log}(\text{VMR}_{\text{N}_2})$	-3.40	$-0.35^{+0.23}_{-0.30}$	$-0.00^{+0.00}_{-0.00}$
$A_g$	0.05	$0.36^{+0.52}_{-0.29}$	$0.96^{+0.03}_{-0.04}$
$\text{Log}(g)$ [cgs]	2.99	$2.99^{+0.02}_{-0.03}$	$3.04^{+0.01}_{-0.02}$
$\mu$	43.54	$35.78^{+3.12}_{-4.26}$	$28.01^{+0.05}_{-0.00}$

**Note.** The error bars of the retrieval results correspond to the 95% confidence interval (i.e.,  $2\sigma$ ).

**Table 5**

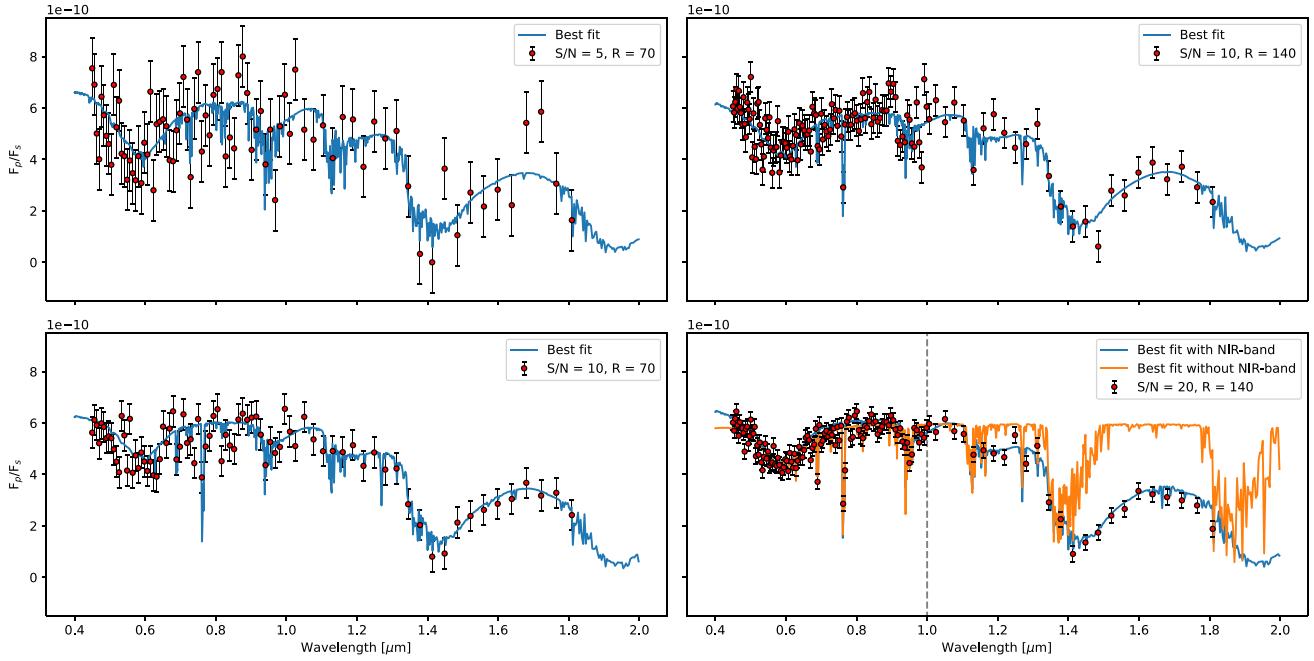
Atmospheric Parameters Used to Simulate the Dry  $\text{CO}_2$ -dominated Atmosphere Scenario and the Retrieval Results

Parameter	Input	Optical + NIR	Optical only
$\text{Log}(P_0)$ [Pa]	5.00	$5.71^{+0.26}_{-0.18}$	$5.51^{+1.77}_{-0.78}$
$\text{Log}(P_{\text{top},\text{H}_2\text{O}})$ [Pa]	...	$6.47^{+1.44}_{-2.49}$	$4.10^{+3.57}_{-3.73}$
$\text{Log}(D_{\text{cld},\text{H}_2\text{O}})$ [Pa]	...	$2.98^{+1.89}_{-2.47}$	$4.95^{+3.29}_{-4.47}$
$\text{Log}(\text{CR}_{\text{H}_2\text{O}})$	...	$-8.49^{+3.85}_{-3.09}$	$-5.72^{+5.25}_{-5.77}$
$\text{Log}(\text{VMR}_{\text{H}_2\text{O}})$	...	$-7.25^{+3.93}_{-3.42}$	$-13.94^{+7.63}_{-9.90}$
$\text{Log}(\text{VMR}_{\text{CH}_4})$	-5.96	$-7.54^{+1.73}_{-2.88}$	$-13.68^{+7.08}_{-10.18}$
$\text{Log}(\text{VMR}_{\text{CO}_2})$	-0.01	$-0.93^{+0.24}_{-0.23}$	$-13.61^{+9.38}_{-9.74}$
$\text{Log}(\text{VMR}_{\text{O}_2})$	-2.79	$-6.67^{+2.89}_{-3.70}$	$-14.28^{+8.59}_{-9.48}$
$\text{Log}(\text{VMR}_{\text{O}_3})$	...	$-9.34^{+1.22}_{-1.28}$	$-14.36^{+5.95}_{-9.60}$
$\text{Log}(\text{VMR}_{\text{N}_2})$	-1.55	$-0.05^{+0.02}_{-0.05}$	$-0.01^{+0.01}_{-0.01}$
$A_g$	0.2	$0.34^{+0.05}_{-0.04}$	$0.30^{+0.48}_{-0.12}$
$\text{Log}(g)$ [cgs]	2.99	$3.22^{+0.06}_{-0.05}$	$3.17^{+0.37}_{-0.23}$
$\mu$	43.55	$29.89^{+1.41}_{-0.77}$	$28.01^{+0.01}_{-0.01}$

**Note.** The error bars of the retrieval results correspond to the 95% confidence interval (i.e.,  $2\sigma$ ).

is dominated by Rayleigh scattering as there are no clouds, and the flux ratio is about 4 to 5 times smaller compared with the previous case and the overall flux ratio is on the order of  $10^{-10}$ . This may thus be a limiting scenario that corresponds to the nominal precision achievable with exoplanet direct-imaging missions (e.g., Roberge & Moustakas 2018; Seager et al. 2019). With the assumed S/N of 20, the full posterior distributions are reported in Appendix D.

The posterior distributions show that no clouds are detected and the abundance of gaseous  $\text{H}_2\text{O}$  should be low and negligible, as expected. The surface pressure and the surface albedo are therefore the most important parameters responsible for the baseline of the spectrum. The two parameters are therefore correlated with each other. The posterior distribution of the surface albedo, as well as the surface pressure, suggests a small bias from the true value. This bias could be linked with the moderate overestimation of the surface gravity (which is correlated with both the surface albedo and pressure) and



**Figure 3.** Simulated data and the best-fit spectra of the Earth-like scenario. Combinations of different S/N and optical-band spectral resolutions have been considered to generate the four different cases. The blue lines depict the best-fit spectra from the retrieval, and the orange line depicts the best-fit spectrum without fitting the NIR band.

therefore the radius of the planet. On the atmospheric composition, the retrieval algorithm is able to tell that the atmosphere has a very high abundance of CO<sub>2</sub> and does not have a significant presence of other absorbing gases, but also permits the interpretation of an N<sub>2</sub>-dominated atmosphere. We consider this interpretation “somewhat correct,” as N<sub>2</sub> is a largely inactive gas.

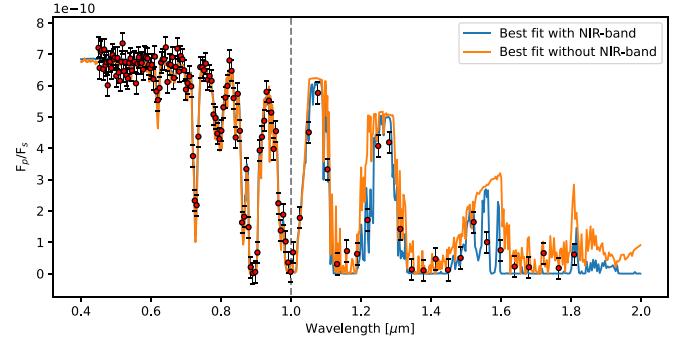
The abundance of O<sub>2</sub> in this atmosphere is moderately constrained, with a tail toward the low abundance. This is because the assumed abundance of O<sub>2</sub> is small, and thus the O<sub>2</sub> features in the spectrum are not as prominent as in the previous cases.

Without the NIR band (see Appendix D), however, the very high abundance of CO<sub>2</sub> cannot be retrieved. In this case, the retrieval algorithm mistakenly reports a 100% N<sub>2</sub> atmosphere, without clouds or any other gases. Also, degenerate solutions are found regarding the surface pressure, the surface albedo, and the gravity of the planet, as these parameters are correlated with each other. We consider the spectral interpretation without the NIR band to be incorrect as it misses the main feature of the atmosphere, and this error is again due to the absorption features of gaseous CO<sub>2</sub> in the NIR band.

## 4. Discussion

### 4.1. Comparison with Previous Studies

The results on the spectral characterization of a modern Earth analog presented in Section 3.1 can be compared with the results presented in Feng et al. (2018). That paper also considered an Earth-like atmosphere and explored the effects on the atmospheric constraints with different S/N and spectral resolutions of a synthesized reflected spectrum in the wavelength range between 0.4 and 1.0 μm. That paper suggested that if a spectral resolution of 70 is considered, the minimum S/N required is 10 in order to characterize clouds



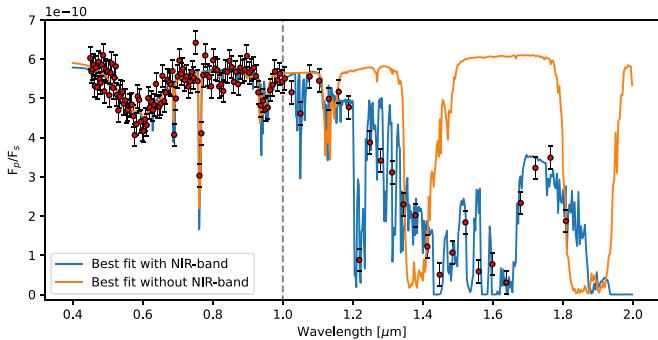
**Figure 4.** Simulated data of the Archean Earth-like scenario (red points) and the best-fit models with (blue line) and without the NIR band (orange line).

and VMR of absorbing gases in the atmosphere. The required S/N bumps to 15 if a spectral resolution of 140 is instead assumed.

Our findings are in general agreement with Feng et al. (2018). EXOREL<sup>33</sup> presents the feature to not assume the dominant gas but let the spectrum speak for itself. With the inclusion of the NIR band (1.0–1.8 μm), we find that an N<sub>2</sub>–O<sub>2</sub> atmosphere is retrieved at the S/N and spectral resolution consistent with Feng et al. (2018). Without the NIR band, the retrieval can still find the atmosphere to be O<sub>2</sub>-dominated but will miss the N<sub>2</sub>. These results show that a modern Earth analog is really an ideal case for spectral characterization in the reflected light, thanks to the richness of the spectral features (Figure 3). We suggest that S/N = 5 is not enough to characterize the atmospheric composition and clouds in this case, and an S/N > 10 and ideally 20 would be required.

### 4.2. The Importance of the NIR Band

A modern Earth analog is one of the many possibilities we might encounter when exploring terrestrial planets through



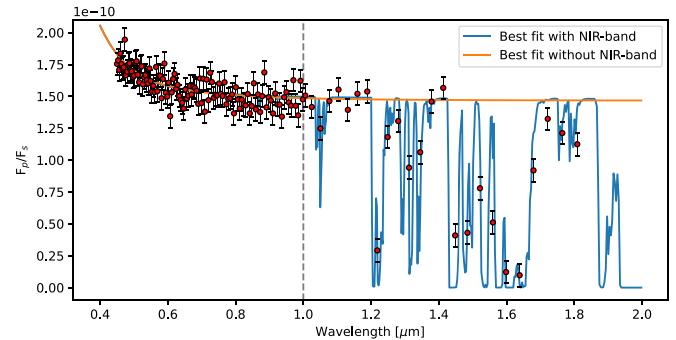
**Figure 5.** Simulated data of the CO<sub>2</sub>-dominated atmosphere with O<sub>2</sub> and clouds scenario (red points) and best-fit models with (blue line) and without the NIR band (orange line).

direct imaging. Earth itself had an atmosphere very different from the present day for most of its past (e.g., Catling & Zahnle 2020), and the other rocky planets in the solar system have CO<sub>2</sub>-dominated atmospheres, indicating the outcomes of rocky planet evolution are diverse. Therefore, the ability to measure the bulk atmospheric composition appears to be an inalienable part of the ability to characterize a terrestrial exoplanet. EXOREL<sup>R</sup> was designed to quantify the requirement to measure the bulk atmospheric composition (Damiano & Hu 2021). Here, based on the limited exploration of the scenarios that resemble Archean Earth and planets having CO<sub>2</sub>-dominated atmospheres, we suggest that the optical band alone (0.4–1.0  $\mu\text{m}$ ) is not sufficient to characterize terrestrial exoplanets found by direct imaging, and including the NIR band (1.0–1.8  $\mu\text{m}$ ) can provide an adequate solution.

Table 6 summarizes the main findings from the retrieval studies in Section 3 and compares what we would learn about the terrestrial planets with the optical band alone to that from both the optical and the NIR bands. Recall that our optical only cases assume a spectral resolution of 140 and an S/N of 20 in 0.4–1.0  $\mu\text{m}$ , which is consistent with the best capability envisioned by existing mission studies for the optical band (e.g., Roberge & Moustakas 2018; Seager et al. 2019; Gaudi et al. 2020). Even so, the retrievals misinterpret the optical-only spectra of all the scenarios that go beyond the modern Earth analog. Without the NIR band, the retrievals converge to solutions that are spectroscopically degenerate (Figures 3–6) and have fundamentally different atmospheric composition from the truths (Table 6). It is striking that the modern Earth analog is the only scenario, among the four explored, for which the retrieval results from the optical-only spectrum can be deemed acceptable. In the other cases where CO<sub>2</sub> plays a more prominent role in shaping the spectra, the retrievals from the optical-only spectra return erroneous solutions.

The specifics of the degenerate solutions may depend on the build-in assumptions of EXOREL<sup>R</sup>, for example, the parameterization of the cloud density (Figure 1). However, our results indicate that at least one degenerate solution exists for each of the scenarios. Allowing more freedom in the cloud parameterization (e.g., Feng et al. 2018; Carrión-González et al. 2020) will result in potentially more degenerate solutions.

The degenerate solutions found appear to be physical and thus unlikely to be ruled out by planetary models. The degenerate solution in the Archean Earth case resembles an H<sub>2</sub>O steam atmosphere with high mixing ratios of CH<sub>4</sub>, which by itself may be plausible from an atmospheric evolution standpoint (e.g., Zahnle et al. 1988). CH<sub>4</sub> should react with



**Figure 6.** Simulated data of the dry CO<sub>2</sub>-dominated atmosphere scenario (red points) and the best-fit models with (blue line) and without the NIR band (orange line).

H<sub>2</sub>O to produce CO<sub>2</sub>, but the optical-only retrieval does not provide any constrain on the mixing ratio of CO<sub>2</sub>, and the lifetime of CH<sub>4</sub> can be long if the atmosphere is massive and has some H<sub>2</sub> in the mixture (e.g., Hu & Seager 2014). To maintain a steam atmosphere would require the planet to be sufficiently irradiated, but for a planet close to the inner edge of the habitable zone, it would be difficult to eliminate this degeneracy solely based on theoretical calculations of the planet's temperature, as those calculations have uncertainties and are sensitive to minor gases in the atmosphere (e.g., Way & Del Genio 2020; Turbet et al. 2021). Similarly, the degenerate solution found in the CO<sub>2</sub>-dominated atmosphere with clouds case involves a dry and massive atmosphere made of N<sub>2</sub>, which would be uninhabitable and cannot be easily discounted from theoretical models (e.g., planetary and atmospheric evolution, thermal structure, and atmospheric chemistry).

What if we let go of the objective of determining the bulk atmospheric composition and only focus on detecting the biosignature gases in these scenarios? Both O<sub>2</sub> and O<sub>3</sub> are detectable with the optical-only spectra, but their mixing ratios in a CO<sub>2</sub>-dominated atmosphere would not be correctly retrieved unless the NIR band is included (Table 6). Several false-positive scenarios have been proposed for O<sub>2</sub> and O<sub>3</sub> as biosignature gases (e.g., Hu et al. 2012b; Domagal-Goldman et al. 2014; Wordsworth & Pierrehumbert 2014; Meadows et al. 2018), and it would be important to measure their mixing ratios correctly and interpret the mixing ratios in the context of planetary environments. For a planet like Archean Earth, the combination of CO<sub>2</sub> and CH<sub>4</sub> can be a biosignature (e.g., Schwamb et al. 2018). However, this biosignature would not be detected from the optical-only spectrum. From the above, we see that an optical-only spectrum is probably also insufficient for detecting biosignature gases in terrestrial exoplanets.

Taken together, the results shown in this paper indicate that spectroscopy in the optical band (0.4–1.0  $\mu\text{m}$ ) is not sufficient to characterize the terrestrial exoplanets found by direct imaging, in either determining the state of the atmosphere or detecting the biosignature gases. As an initial step, our retrieval exploration suggests that including the NIR band (1.0–1.8  $\mu\text{m}$ ) would eliminate the degenerate solutions and enable the proper measurements of the state of the atmosphere and the biosignature gases. Further extending the wavelength coverage to  $\sim 2$ –5  $\mu\text{m}$  could result in better characterization and sensitivity to more biosignature gases (S. R. Martin et al. 2022, private communication). Extending the wavelength coverage into the UV may be an alternative solution

**Table 6**  
Capability to Characterize Terrestrial Planets with the Reflected Light Spectroscopy in the Optical and NIR Bands

Scenario	Optical + NIR	Optical only
Modern Earth	<i>Correct:</i> N <sub>2</sub> /O <sub>2</sub> atmosphere with H <sub>2</sub> O, O <sub>3</sub> and CH <sub>4</sub>	<i>Somewhat Correct:</i> O <sub>2</sub> atmosphere with H <sub>2</sub> O and O <sub>3</sub>
Archean Earth	<i>Correct:</i> N <sub>2</sub> atmosphere with H <sub>2</sub> O, CO <sub>2</sub> , CH <sub>4</sub> , and little O <sub>2</sub>	<i>Incorrect:</i> H <sub>2</sub> O atmosphere with high clouds, CH <sub>4</sub> more than truth by >1 order of magnitude, and little CO <sub>2</sub>
Habitable Venus (CO <sub>2</sub> atmosphere with O <sub>2</sub> and clouds)	<i>Correct:</i> CO <sub>2</sub> atmosphere with O <sub>2</sub> , H <sub>2</sub> O, O <sub>3</sub> , and possibly N <sub>2</sub>	<i>Incorrect:</i> Massive N <sub>2</sub> atmosphere without clouds and with O <sub>2</sub> , H <sub>2</sub> O and O <sub>3</sub> less than truth by >2 orders of magnitude
Dry CO <sub>2</sub> atmosphere	<i>Somewhat Correct:</i> N <sub>2</sub> /CO <sub>2</sub> atmosphere without clouds or H <sub>2</sub> O	<i>Incorrect:</i> N <sub>2</sub> atmosphere without clouds, H <sub>2</sub> O, or CO <sub>2</sub>

**Note.** The high-level results are summarized from the retrievals on the four planetary scenarios with moderate spectral resolution and S/N.

(Lisman & Schwieterman 2019). We will assess these options in future studies.

#### 4.3. Vertical Profile of H<sub>2</sub>O

It is also notable that the mixing ratio of H<sub>2</sub>O below the cloud is retrieved correctly and precisely (typical  $2\sigma$  uncertainty  $\sim 0.5$  dec) for all the scenarios with H<sub>2</sub>O based on the optical and NIR spectra (Tables 1–4). This ability to constrain the mixing ratio of H<sub>2</sub>O below the cloud comes from the EXOREL<sup>R</sup> feature that connects the cloud density to the drop off of the mixing ratio of H<sub>2</sub>O, as shown in Figure 1. We see that the cloud position is also retrieved correctly in these scenarios, and this is because the water cloud is not fully reflective but has its spectral features in the NIR band.

The characterization of the cloud position and the mixing ratio of H<sub>2</sub>O below the cloud will result in the constraints of the vertical profile of water vapor on a terrestrial exoplanet. This vertical profile can then be compared with climate model predictions to provide insight into the planet's hydrologic cycles and the existence of a large water reservoir on the surface.

### 5. Conclusions

We use a Bayesian retrieval model EXOREL<sup>R</sup> to study the spectral characterization capabilities of terrestrial exoplanets found by direct imaging. We consider plausible scenarios that range from planets similar to modern Earth and Archean Earth to planets having CO<sub>2</sub>-dominated atmospheres, and perform spectral retrievals without any prior knowledge of the composition of the atmosphere. We find that a spectrum in 0.4–1.8  $\mu$ m with moderate resolution ( $R = 140$  in the optical and 40 in the NIR band) and an S/N = 20 can distinguish each of these scenarios, determine the main atmospheric component (N<sub>2</sub>, O<sub>2</sub>, or CO<sub>2</sub>), measure the mixing ratios of trace gases including H<sub>2</sub>O, O<sub>3</sub>, and CH<sub>4</sub>, and constrain the pressure level of a cloud layer. These constraints amount to a comprehensive understanding of the state of the atmosphere, detection of O<sub>2</sub>/O<sub>3</sub> or CO<sub>2</sub>–CH<sub>4</sub> biosignature gases, and potential inference of liquid-water oceans on the surface.

To achieve this vision of terrestrial exoplanet characterization will require spectroscopy that goes beyond the optical band of 0.4–1.0  $\mu$ m. Consistent with Feng et al. (2018), we show that the optical-band spectroscopy at the resolution of at least 70–140 and the S/N of at least 10–20 can provide reasonable constraints on an Earth-like atmosphere. However,

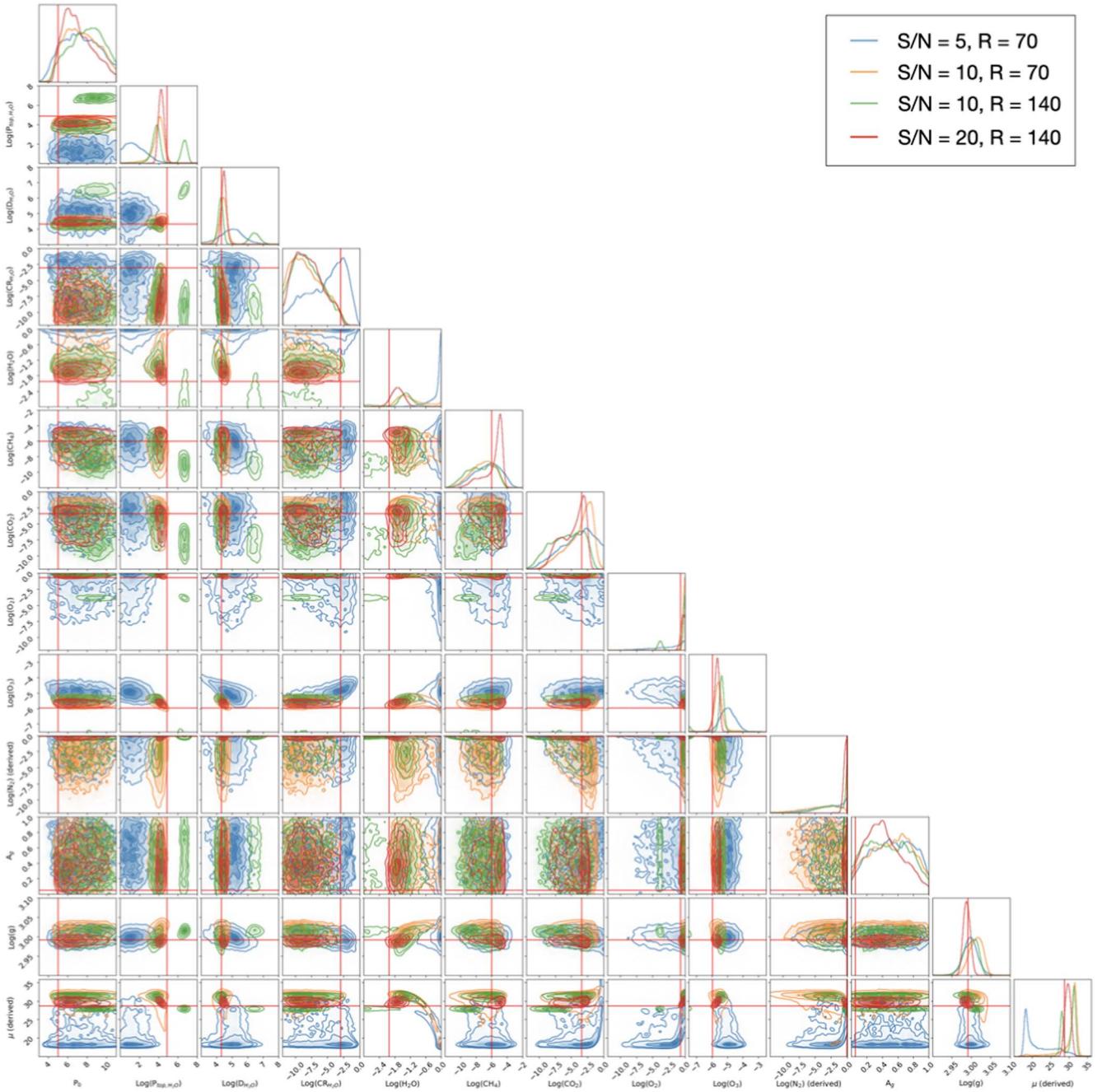
for planets like Archean Earth or having CO<sub>2</sub>-dominated atmospheres, the optical-band spectroscopy will be insufficient and suffer from physically plausible degenerate solutions. Including the NIR band will eliminate the degeneracies and prevent incorrect interpretation.

It is thus crucial to consider a broad wavelength coverage that goes beyond the optical band when designing the future space telescope capable of terrestrial exoplanet imaging. This is, in a general sense, consistent with the proposed mission concepts HabEx (Gaudi et al. 2020) and LUVOIR (Roberge & Moustakas 2018) that would provide the spectroscopic capability in the NIR band. Regardless of the technologies used for starlight suppression (i.e., coronagraph and star shade), it should be considered a high priority to obtain the broad-wavelength spectroscopy for as many targets as possible when evaluating mission architectures. The minimum requirement for spectral characterization of potentially habitable exoplanets should include, at least a fraction of, the NIR band. Future studies will further pinpoint the short and long wavelength cutoffs for an adequate spectrum of directly imaged terrestrial exoplanets.

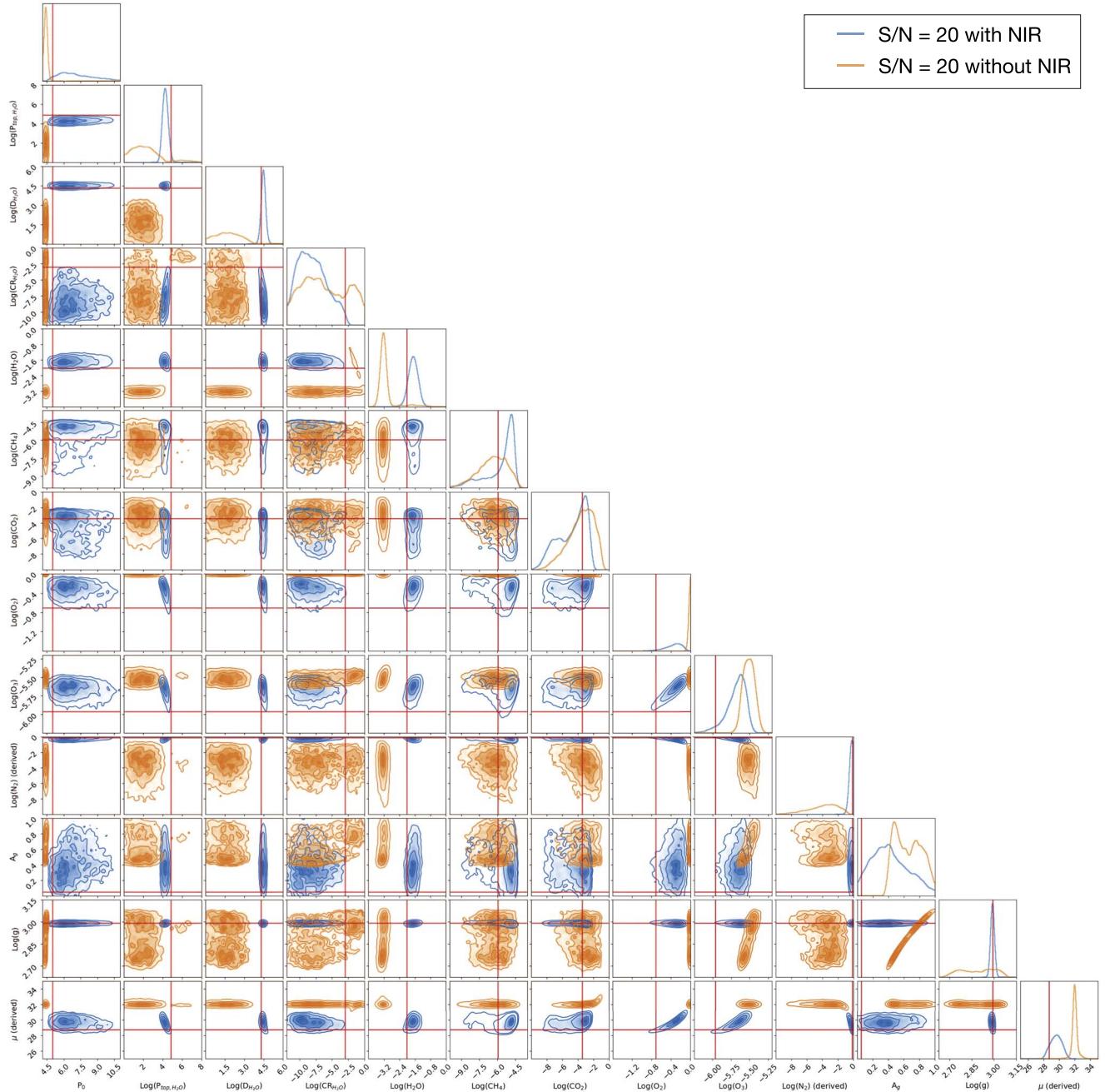
We thank Charles Lawrence, Bertrand Mennesson, Keith Warfield, and Rhonda Morgan for helpful discussion. This work was supported in part by the NASA WFIRST Science Investigation Teams grant #NNN16D016T and Exoplanets Research Program grant #80NM0018F0612. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

### Appendix A Scenario 1: An Earth-like Atmosphere

We simulate the reflected spectrum of an N<sub>2</sub>-dominated atmosphere with O<sub>2</sub> being the second most abundant gas (Figure 3). We also include H<sub>2</sub>O, CH<sub>4</sub>, CO<sub>2</sub>, and O<sub>3</sub> as minor absorbing gases. We also include a water cloud layer and a surface with albedo of 0.05. We use the synthesized data as input for EXOREL<sup>R</sup>, and Figure 7 comprises the resulted posterior distributions of all the S/N and spectral resolution combinations discussed in Section 2.2. Also, we run EXOREL<sup>R</sup> on only the optical part of the spectrum to explore the benefit of having the NIR wavelength band in addition to the optical band. The result is shown in Figure 8.



**Figure 7.** Full posterior distributions (corner plots) of the Earth-like scenario. The different cases explore a combination of different optical-band spectral resolution and S/N. The red lines in the corner plots refer to the true value used to simulate the data.



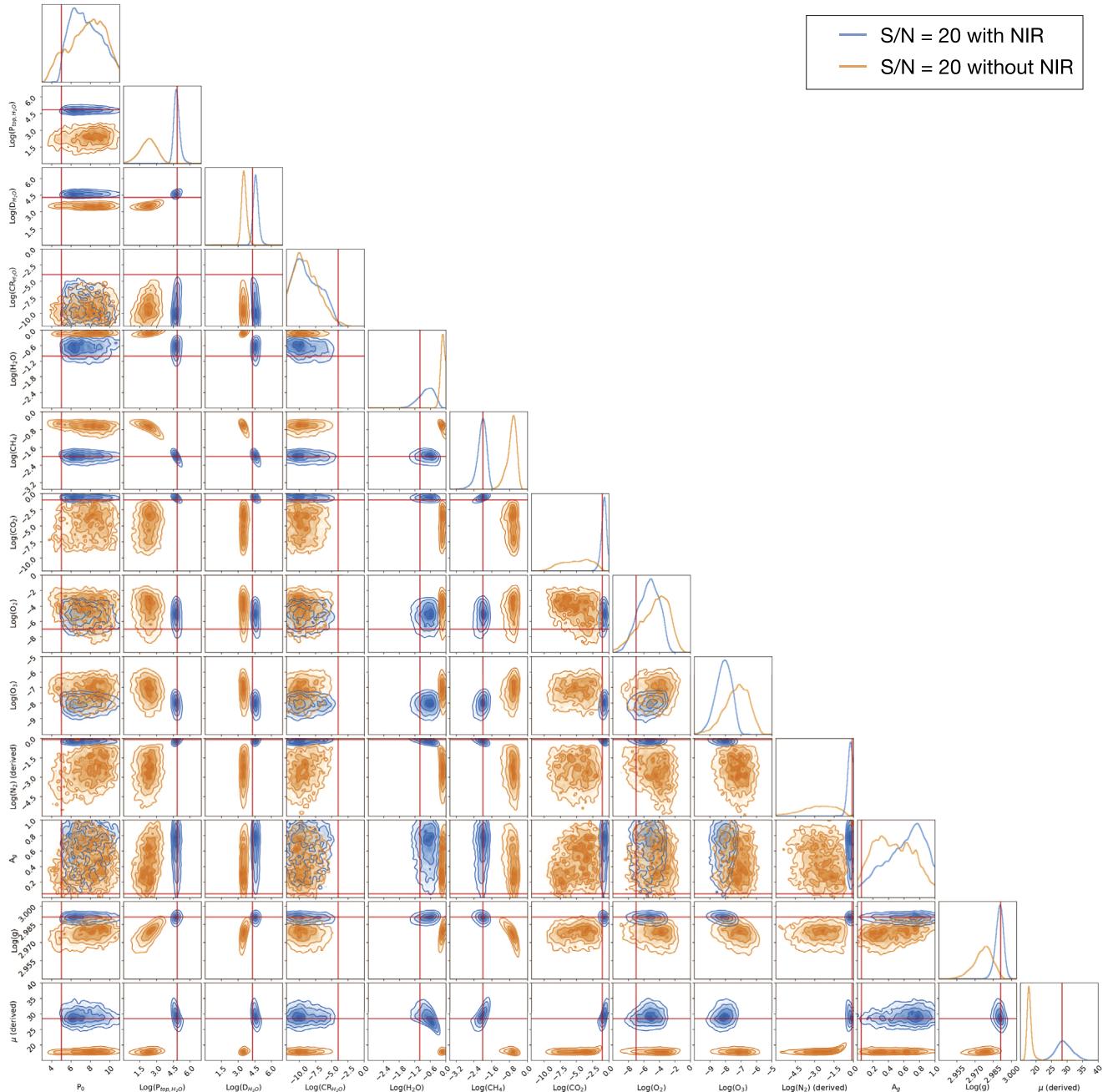
**Figure 8.** Full posterior distributions of the Earth-like scenario with (blue lines) and without (orange lines) the NIR wavelength band in 1.0–1.8  $\mu\text{m}$ . The red lines in the corner plots refer to the true value used to simulate the data.

## Appendix B

### Scenario 2: An Archean Earth-like Atmosphere

Earth's atmosphere has experienced multiple evolutionary stages before reaching the current state. One of the main stages early in the history is called Archean. During this stage, the atmosphere had minimal O<sub>2</sub> and a higher amount of CO<sub>2</sub> and CH<sub>4</sub> with respect to the contemporary atmosphere. Therefore, we also simulate the reflected spectrum of a planet like the

Archean Earth, assuming an N<sub>2</sub>-dominated atmosphere with H<sub>2</sub>O, CH<sub>4</sub>, CO<sub>2</sub>, and O<sub>2</sub> as minor absorbing gases (Figure 4). We still include a water cloud layer and a surface with albedo of 0.05. The synthesized data are used as input for EXOREL<sup>R</sup>. We perform the retrieval with and without the IR part of the spectrum to explore the benefit of having the NIR wavelength band in addition to the optical. The result is shown in Figure 9.



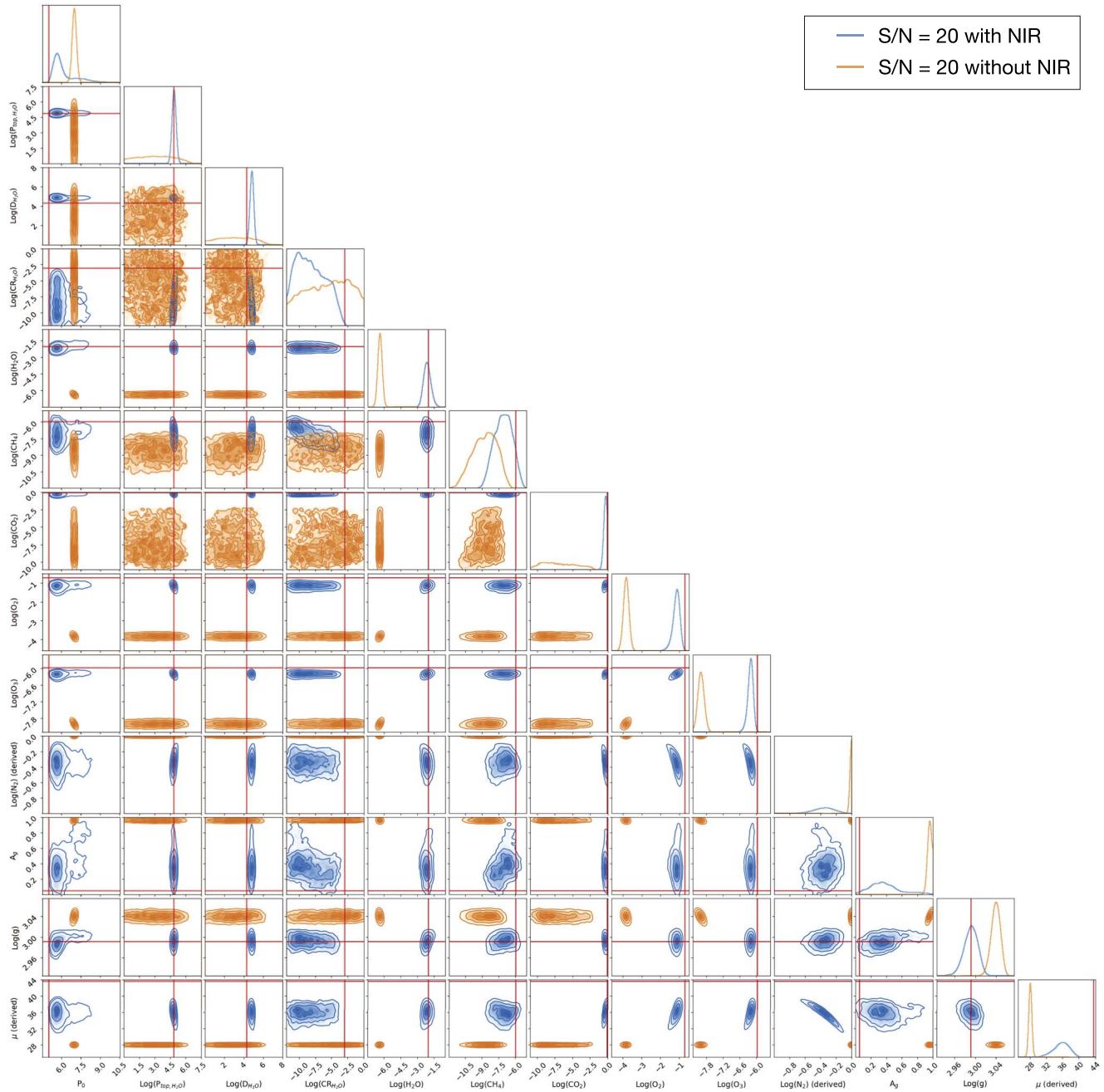
**Figure 9.** Full posterior distributions of the Archean Earth-like scenario with (blue lines) and without (orange lines) the NIR wavelength band. The red lines in the corner plots refer to the true value used to simulate the data.

### Appendix C

#### Scenario 3: A CO<sub>2</sub>-dominated Atmosphere with O<sub>2</sub> and Clouds

We simulate the reflected spectrum of a CO<sub>2</sub>-dominated atmosphere with O<sub>2</sub> and a cloud layer (Figure 5). We include

H<sub>2</sub>O, CH<sub>4</sub>, and O<sub>3</sub> as minor absorbing gases. We also include a surface with albedo of 0.05. We run EXOREL<sup>R</sup> on only the optical part of the spectrum to explore the benefit of having the NIR wavelength band in addition to the optical. The result is shown in Figure 10.



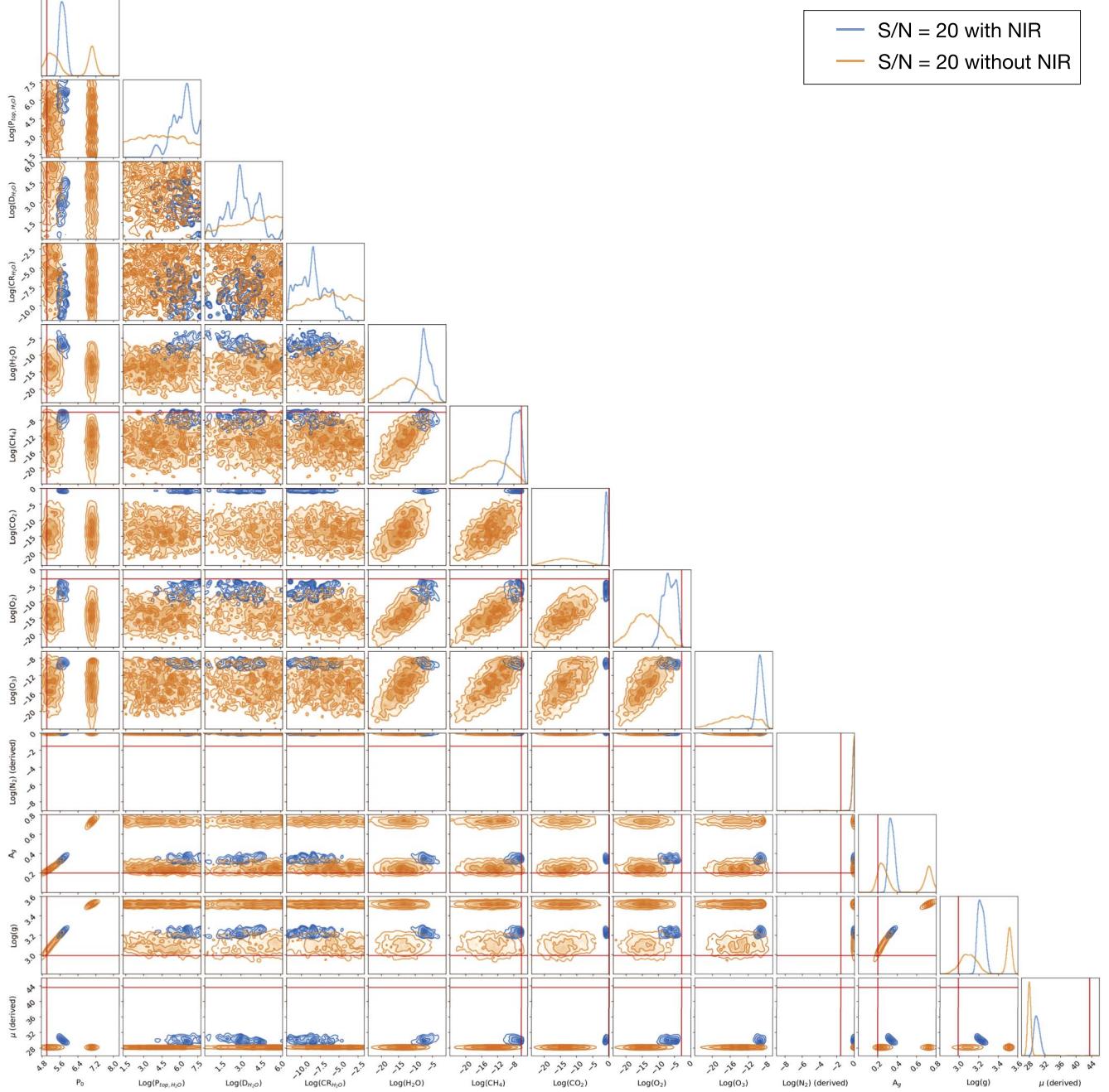
**Figure 10.** Full posterior distribution of the CO<sub>2</sub>-dominated atmosphere with O<sub>2</sub> and clouds scenario with (blue lines) and without (orange lines) the NIR wavelength band. The red lines in the corner plots refer to the true value used to simulate the data.

## Appendix D

### Scenario 4: A Dry CO<sub>2</sub>-dominated Atmosphere

We simulate the reflected spectrum of a dry CO<sub>2</sub>-dominated atmosphere without clouds (Figure 6). We use a CO<sub>2</sub>-dominated atmosphere with CH<sub>4</sub> and O<sub>2</sub> as minor absorbing gases. We

include a surface with albedo of 0.2. We run EXOREL<sup>R</sup> on only the optical part of the spectrum to explore the benefit of having the NIR wavelength band in addition to the optical. We use the synthesized data as input for EXOREL<sup>R</sup>, and the result is shown in Figure 11.



**Figure 11.** Full posterior distribution of the dry CO<sub>2</sub>-dominated atmosphere without clouds scenario with (blue lines) and without (orange lines) the NIR wavelength band. The red lines in the corner plots refer to the true value used to simulate the data.

## ORCID iDs

Mario Damiano <https://orcid.org/0000-0002-1830-8260>  
 Renyu Hu <https://orcid.org/0000-0003-2215-8485>

## References

- Aitchison, J. 1982, *J. R. Stat. Soc. Ser. B Methodol.*, 44, 139
- Akeson, R., Armus, L., Bachelet, E., et al. 2019, arXiv:[1902.05569](https://arxiv.org/abs/1902.05569)
- Astro 2020 2021, Pathways to Discovery in Astronomy and Astrophysics for the 2020s (Washington DC: National Academies of Sciences, Engineering, and Medicine)
- Belu, A., Selsis, F., Morales, J.-C., et al. 2011, *A&A*, **525**, A83
- Carrión-González, Ó., Muñoz, A. G., Cabrera, J., et al. 2020, *A&A*, **640**, A136
- Catling, D. C., & Zahnle, K. J. 2020, *SciA*, **6**, eaax1420
- Damiano, M., & Hu, R. 2020, *AJ*, **159**, 175
- Damiano, M., & Hu, R. 2021, *AJ*, **162**, 200
- Damiano, M., Hu, R., & Hildebrandt, S. R. 2020, *AJ*, **160**, 206
- Deming, D., Seager, S., Winn, J., et al. 2009, *PASP*, **121**, 952
- Domagal-Goldman, S. D., Segura, A., Claire, M. W., Robinson, T. D., & Meadows, V. S. 2014, *ApJ*, **792**, 90
- Elkins-Tanton, L. T., & Seager, S. 2008, *ApJ*, **685**, 1237
- Feng, Y. K., Robinson, T. D., Fortney, J. J., et al. 2018, *AJ*, **155**, 200
- Gaillard, F., Bernadou, F., Roskosz, M., et al. 2022, *E&PSL*, **577**, 117255
- Gaudi, B. S., Seager, S., Mennesson, B., et al. 2020, arXiv:[2001.06683](https://arxiv.org/abs/2001.06683)
- Gialluca, M. T., Robinson, T. D., Rugheimer, S., & Wunderlich, F. 2021, *PASP*, **133**, 054401
- Harness, A., Shaklan, S., Willems, P., et al. 2021, *JATIS*, **7**, 021207
- Hu, R. 2019, *ApJ*, **887**, 166
- Hu, R., Damiano, M., Scheucher, M., et al. 2021a, *ApJL*, **921**, L8
- Hu, R., Ehlmann, B. L., & Seager, S. 2012a, *ApJ*, **752**, 7
- Hu, R., Lisman, D., Shaklan, S., et al. 2021b, *JATIS*, **7**, 021205
- Hu, R., & Seager, S. 2014, *ApJ*, **784**, 63
- Hu, R., Seager, S., & Bains, W. 2012b, *ApJ*, **761**, 166
- Kaltenegger, L., Traub, W. A., & Jucks, K. W. 2007, *ApJ*, **658**, 598
- Krissansen-Totton, J., Garland, R., Irwin, P., & Catling, D. C. 2018, *AJ*, **156**, 114
- Lisman, D., & Schwieiterman, E. W. 2019, *BAAS*, **51**, 217
- Meadows, V. S., Reinhard, C. T., Arney, G. N., et al. 2018, *AsBio*, **18**, 630
- Palmer, K. F., & Williams, D. 1974, *JOSA*, **64**, 1107
- Pidhorodetska, D., Fauchez, T. J., Villanueva, G. L., Domagal-Goldman, S. D., & Kopparapu, R. K. 2020, *ApJL*, **898**, L33
- Roberge, A., & Moustakas, L. A. 2018, *NatAs*, **2**, 605
- Schwieiterman, E. W., Kiang, N. Y., Parenteau, M. N., et al. 2018, *AsBio*, **18**, 663
- Seager, S., Kasdin, N. J., Booth, J., et al. 2019, *BAAS*, **51**, 106
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:[1503.03757](https://arxiv.org/abs/1503.03757)
- Trauger, J. T., & Traub, W. A. 2007, *Natur*, **446**, 771
- Turbet, M., Bolmont, E., Chaverot, G., et al. 2021, *Natur*, **598**, 276
- Turnbull, M. C., Traub, W. A., Jucks, K. W., et al. 2006, *ApJ*, **644**, 551
- Way, M. J., & Del Genio, A. D. 2020, *JGRE*, **125**, e2019JE006276
- Wordsworth, R., & Pierrehumbert, R. 2013, *ApJ*, **778**, 154
- Wordsworth, R., & Pierrehumbert, R. 2014, *ApJL*, **785**, L20
- Wunderlich, F., Godolt, M., Grenfell, J. L., et al. 2019, *A&A*, **624**, A49
- Zahnle, K. J., Kasting, J. F., & Pollack, J. B. 1988, *Icar*, **74**, 62
- Zeng, L., Jacobsen, S. B., Sasselov, D. D., et al. 2019, *PNAS*, **116**, 9723