



Reflected Spectroscopy of Small Exoplanets I: Determining the Atmospheric Composition of Sub-Neptunes Planets

Mario Damiano¹  and Renyu Hu^{1,2} 

¹ Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA; mario.damiano@jpl.nasa.gov

² Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA

Received 2021 June 18; revised 2021 August 20; accepted 2021 August 27; published 2021 October 18

Abstract

Direct imaging of widely separated exoplanets from space will obtain their reflected light spectra and measure their atmospheric properties, and small and temperate planets will be the focus for the next generation of telescopes. In this work, we used our Bayesian retrieval algorithm EXOREL³ to determine the constraints on the atmospheric properties of sub-Neptune planets from observations taken with a HabEx-like telescope. Small and temperate planets may have a non-H₂-dominated atmosphere, and therefore we introduced the compositional analysis technique in our framework to explore the bulk atmospheric chemistry composition without any prior knowledge about it. We have developed a novel set of prior functions for the compositional analysis free parameters. We compared the performances of the framework with the flat prior and the novel prior and we reported a better performance when using the novel priors set. We found that the retrieval algorithm cannot only identify the dominant gas of the atmosphere but also to constrain other less abundant gases with high statistical confidence without any prior information on the composition. The results presented here demonstrates that reflected light spectroscopy can characterize small exoplanets with diverse atmospheric composition. The Bayesian framework should be applied to design the instrument and the observation plan of exoplanet direct-imaging experiments 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 \(1927\)](#); [Ocean planets \(1151\)](#)

1. Introduction

Up to date, the atmospheric exoplanet characterization of small planets with a H₂-dominated atmosphere have been performed through transmission spectroscopy (e.g., Knutson et al. 2014; Tsiaras et al. 2016, 2019; Wakeford et al. 2017; Benneke et al. 2019a, 2019b). However, if the atmosphere is dominated by heavier gases (e.g., H₂O or CO₂), the atmosphere is less extended and it would be more difficult to characterize it through transmission spectroscopy. Since reflected light observations are not primarily sensitive to the atmospheric scale height, the direct imaging with reflection spectroscopy is well suited to studying a wide range of atmospheric types including a non-H₂-dominated atmosphere. For this reason, direct imaging is gaining traction in the interest of the scientific community.

A new generation of telescopes with direct-imaging capabilities are being developed to allow atmospheric characterization in the reflected light. The Nancy Grace Roman Space Telescope (Roman; Spergel et al. 2015; Akeson et al. 2019) will be capable of collecting starlight reflected by giant exoplanets through high-contrast imaging. The *Starshade Rendezvous Mission* (Seager et al. 2019) could further enhance the Roman's capability to image smaller planets. The Habitable Exoplanet Observatory (HabEx; Gaudi et al. 2020), a concept of a 4m space telescope, and its main objective is to image small planets and study their atmospheres. The Large Ultraviolet/Optical/InfraRed Surveyor (LUVOIR; Roberge & Moustakas 2018) is a concept mission that would find and characterize even more Earth exoplanets similar to Earth by direct imaging.

Most of the Bayesian frameworks, used to interpret atmospheric spectra, assume that the exoplanetary atmosphere is

H₂-dominated. The assumption simplifies the problem as only the mixing ratio of minor gases are free parameters of the model. When studying the atmospheric chemistry composition, the constraint that the sum of the mixing ratio of the gases equals unity must be respected. The assumed knowledge of the dominant gas implicitly respect the above condition. However, small planets might not have a H₂-dominated atmosphere and the dominant gas is often unknown. Under these condition, compositional analysis (Aitchison 1982) must be introduced in the Bayesian framework. In Benneke & Seager (2012), the authors introduced compositional analysis to apply Bayesian statistics to transmission spectra of small planets. They used the centered-log ratio (CLR) of the gases as free parameters opposed to the log-volume-mixing ratio (Log(VMR)) as commonly used in other Bayesian framework. The authors also chose to adopt *ignorance prior* (constant prior) functions for the CLR. In this work, we also adopt the CLR of gases as free parameters; however, we introduced a new set of priors. The volume-mixing ratio (VMR) is a direct measurement of the concentration of the gas in the atmosphere, and in our framework, we would like to explore different values of VMR and assigning them the same probability. Since the relation between the CLR and the VMR is nonlinear, reparameterization of a constant probability distribution is not another constant distribution. We noted that with the new set of priors, the Bayesian framework does not overestimate or underestimate the concentration of gases, which occurred when we used a constant prior function for CLR.

Here we use a robust Bayesian inverse retrieval method (EXOREL³; Damiano & Hu 2020; Damiano et al. 2020) to interpret reflected light spectra of small and temperate sub-Neptune planets. This is, to our knowledge, the first time a

retrieval method is applied to reflected light spectroscopy of small exoplanets beyond Earth-like planets (which was discussed in Feng et al. 2018). Our model has the following specific advancements. (1) We can use retrieval to determine the main component of the atmosphere, rather than assuming the atmosphere to be, for example, N_2 -dominated or H_2 -dominated. (2) We can explore atmosphere in which multiple gases have equal or similarly dominant abundances. (3) We can retrieve the mixing ratio of H_2O below a water cloud, leading to potential inference of an ocean.

The paper is organized as follows: Section 2 describes the setup of the retrieval, Section 3 discusses the scientific scenarios we explored in this work, Section 4 shows the results of the inverse retrieval performed on the scientific scenarios, and Section 5 explores the implications of the results obtained and the retrieval setup used.

2. Methods

2.1. Retrieval Setup

We update the atmospheric scenario and radiative transfer model in Hu (2019), Damiano & Hu (2020), and Damiano et al. (2020) to enable simulations of super-Earths' reflected light spectra. Compared with giant planet models, a key difference for super-Earths' atmospheres is that they can have a wider range of bulk atmospheric compositions (e.g., Hu & Seager 2014). In order to consistently calculate the VMR of the constituents of an exoplanetary atmosphere, models must respect the following relations:

$$0 < \text{VMR}_i < 1 \quad (1)$$

and

$$\sum_{i=1}^n \text{VMR}_i = 1 \quad (2)$$

where n represents the total number of gases considered in the model. The equations above underline that the values of the gases are not independent as one value can be calculated by knowing the other $n - 1$ values. Therefore, in the gaseous giant planets scenario, where H_2 is always the dominant gas, the Bayesian statistical process uses the $\text{Log}(\text{VMR})$ of the other gases as free parameters. This scheme always ensure that Equations (1) and (2) are respected. In the case of smaller planets, it is unknown which gas is the dominant one a priori. For this reason, it is challenging using the $\text{Log}(\text{VMR})$ as the free parameter for each of the considered gases in the Bayesian process while respecting Equations (1) and (2).

To overcome this difficulty, we implemented the CLR (Aitchison 1982) of the mixing ratios of H_2O , CO_2 , CH_4 , N_2 , O_2 , and O_3 as free parameters. The CLR is defined for a composition X of n elements as follows:

$$\text{clr}(X) = \ln \left[\frac{X_1}{g_m(X)}, \dots, \frac{X_n}{g_m(X)} \right] \quad (3)$$

where

$$g_m(X) = \left(\prod_{i=1}^n X_i \right)^{1/n} \quad (4)$$

is the geometric mean of the composition X .

Let us define the VMR as a composition of n gases. By using the CLR, the restrictions the model must respect can be simplified. Indeed, Equations (1) and (2), in the CLR space, simply translate into:

$$\sum_{i=1}^n \text{clr}(\text{VMR})_i = 0. \quad (5)$$

By definition, the CLR can assume positive and negative values and in particular the CLR could assume any values inside the range $(-\infty, +\infty)$. Lower values of CLR correspond to the low value of VMR and vice versa. One can thus specify $n - 1$ CLR values and derive the last CLR value without favoring or disfavoring any species. This approach for compositional retrieval has been previously applied to transmission spectroscopy (Benneke & Seager 2012).

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 μm , and gradually more absorption of the clouds outside the 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. In our model, the cloud is parametrized by the cloud top pressure (P_{top}), the cloud depth (i.e., the difference in pressure between the cloud bottom and the cloud top, D_{clid}), 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). Finally, 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; Robinson 2018).

In the vision of performing the retrieval of super-Earth reflected light spectroscopy, we introduce the surface pressure and albedo as fixed or free parameters of the model. Even though they are not leveraged by the scenarios we present here (in this work they are fixed to a specific value; see Section 3), we will describe and explore the effect of the surface (pressure position and ground albedo) in a subsequent paper on small planets (M. Damiano & R. Hu 2021, in preparation).

2.2. A New Set of Priors

In EXOREL²¹ (Hu 2019; Damiano & Hu 2020; Damiano et al. 2020), we defined the $\text{Log}(\text{VMR})$ of H_2O , NH_3 , and CH_4 as free parameters of our Bayesian model to retrieve insights into the atmosphere of gaseous giant planets. We used uniform priors for the $\text{Log}(\text{VMR})$ to assign the same probability to the values in the range $(-12.0, 0.0)$. To characterize small planets, we pointed to the CLR of the gases to be the free parameters of the Bayesian process. However, the CLR is a nonlinear transformation of the VMR space. For this reason, we cannot use a uniform set of priors for the CLR, otherwise we would assign different probabilities to different values of VMR, favoring some VMR values rather than equally sampling the

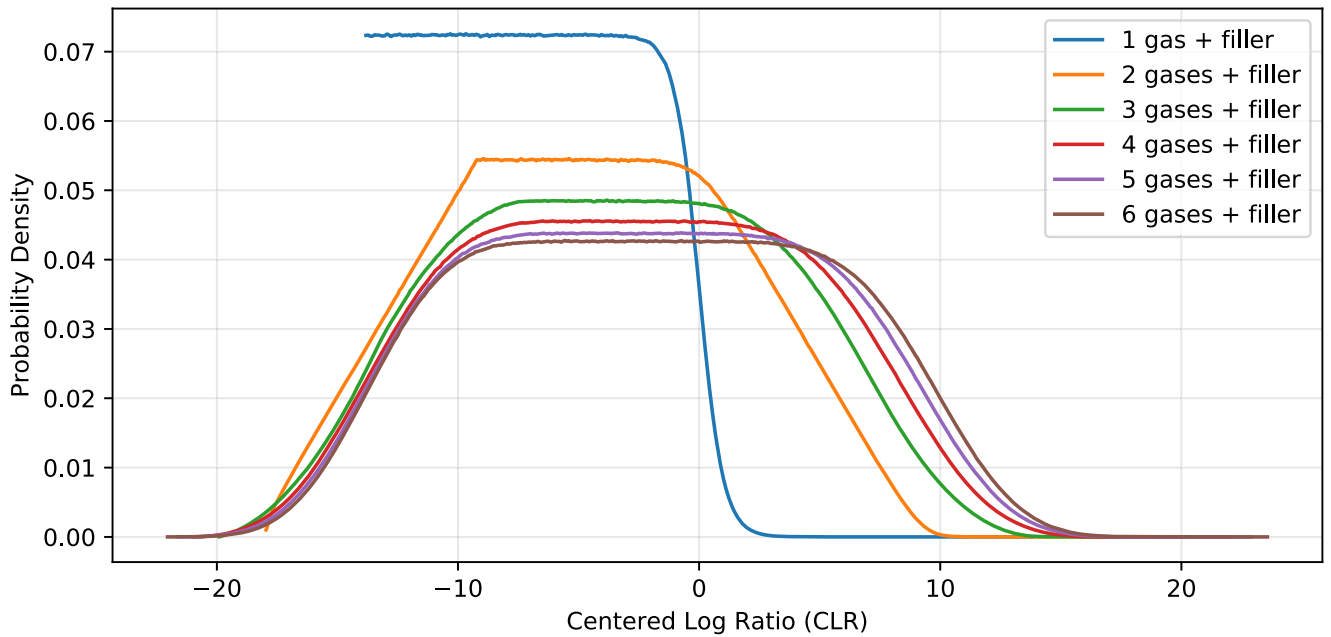


Figure 1. Prior functions in the CLR space for different number of gases considered in the retrieval. The functions are normalized by their integral.

whole range. This is particularly true for values of VMR close to unity, this leads the gas with the strongest absorption features to be misinterpreted as the dominant gas in the atmosphere.

For this reason, in this work, we introduce a new set of priors for the CLR. First of all, we underline that the number of free parameters related to gases is equal to $n - 1$ with n being the total number of gases. This is the result of Equation (5). Therefore, we define a “filler” gas, and in our model H_2 is always considered as the “filler” given its nature of inactive gas (i.e., it does not show spectral features in the moderate resolution and signal to noise ratio (S/N) relevant to exoplanet observations). The choice of the filler gas does not favor or disfavor that gas by itself.

Let us consider a VMR range defined within $(10^{-12}, 1)$. For each $n - 1$ VMR_{*i*}, we randomly draw a value from the range according to an *uniform distribution*. We keep only those sets of VMR where the sum is less than 1, therefore we can include the dependent VMR of the filler gas to sum to unity. We use Equation (3) to transform the sets of VMR into sets of CLR. By doing so we transform an *uniform distribution* defined in the VMR space into a distribution of the CLR space. Since the CLR is not a linear transformation, the distribution defined in the CLR space is not uniform. Moreover, we observed that the distribution in the CLR space depends on the value of n , meaning that the shape of the CLR prior is different for different number of gases to be fit in the Bayesian process. In Figure 1, we reported the resulting probability-density function for the CLR for different number of gases considered in the Bayesian process to be fit. Note that once the number n of gases has been chosen, the same prior function is used for all the n gases.

Finally, the prior functions associated to all the other free parameters, i.e., surface pressure P_0 , cloud top pressure P_{top} , the cloud depth D_{cld} , condensation ratio CR, the surface albedo A_g , and the Log of the gravity $\text{Log}(g)$, have uniform distributions as introduced in (Damiano & Hu 2020).

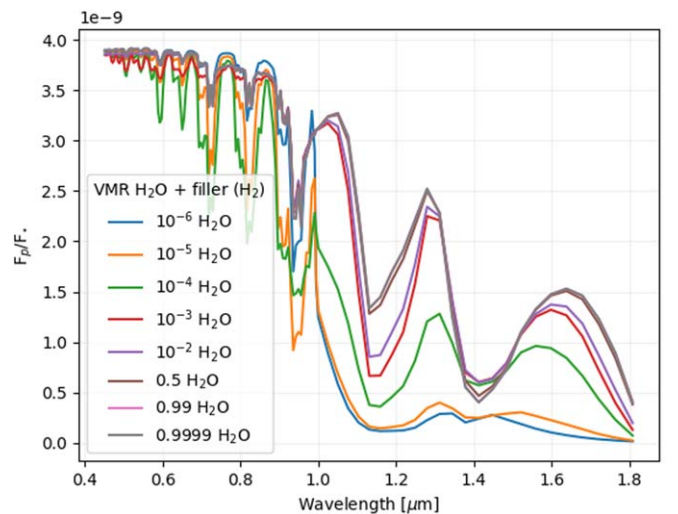


Figure 2. Reflected light spectra of a hypothetical H_2 - H_2O atmosphere with different amount of water. When water is below 10^{-2} in VMR the spectra are distinguishable, while above 10^{-2} the spectra are degenerate and not sensitive anymore to the VMR of the two gases.

3. Planetary Scenarios

3.1. Idealized Scenarios

To test the performance of the new EXOREL^{3A}’s setup, we include here the analysis of the simplest scenario possible, i.e., an atmosphere made solely of water and hydrogen. We simulated eight spectra to cover the VMR of water and hydrogen between 10^{-6} and 0.999999. Figure 2 shows the simulated flux ratio of a $10 M_{\oplus}$ and $2.4 R_{\oplus}$ planet with a gravitational acceleration of $g = 17 \text{ m s}^{-2}$ around a Sun-like star at 1 au with different atmospheric compositions. It is worth noting before running the statistical retrieval process that when the amount of water is significant ($\geq 10^{-2}$) the spectra are degenerate and cannot be discerned from each other. When the water amount is low, the water cloud layer is thin and the starlight passes through. On the contrary, when the water

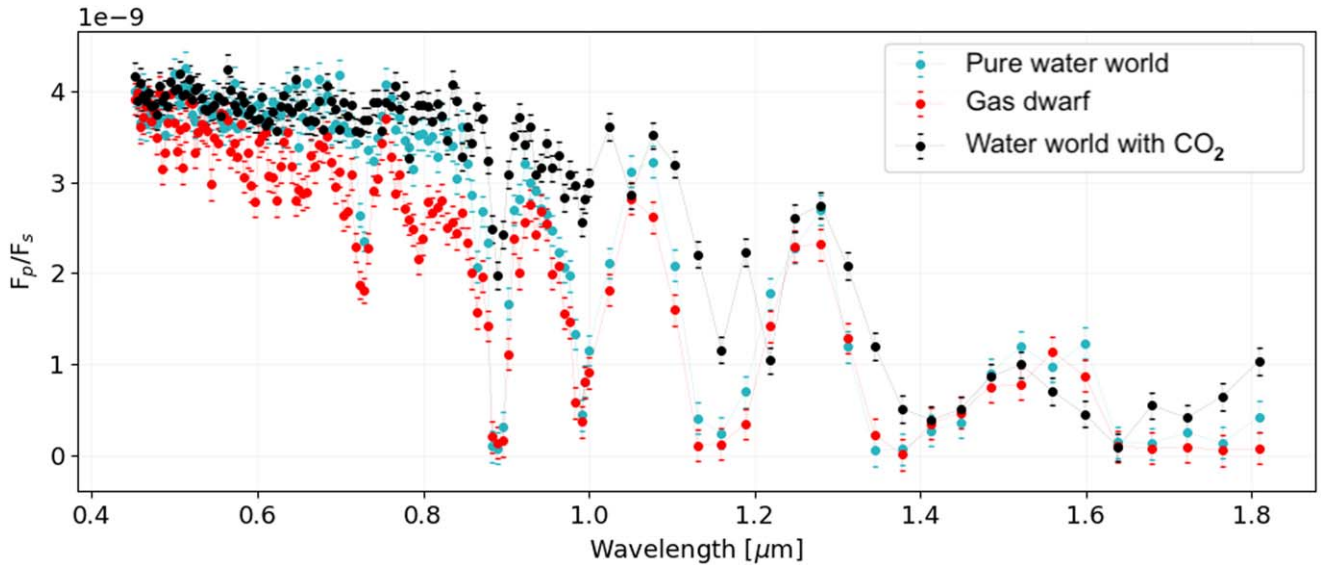


Figure 3. Simulated reflected spectra of the three scenarios of sub-Neptune exoplanets.

concentration is higher, the clouds become optically thick and the modulation observed in the flux ratio is mostly due to the water cloud single-scattering albedo. Note that in all the simulations the cloud position has been kept the same around 10^5 Pa.

Although the spectra are degenerate when the water mixing ratio is $>10^{-2}$, the mean molecular mass and therefore the scale height of the atmosphere is different between, for example, 1% water and 50% water. This difference can be useful to break the degeneracy when other absorbing gases are included (e.g., CO_2 and CH_4), because the atmospheric scale height will affect the depth of the absorption features of the other absorbing gases (as shown in the realistic scenarios).

3.2. Realistic Scenarios

For more complex scenarios, we simulate the same type of planet as in the idealized scenario, i.e., $10 M_{\oplus}$, $2.4 R_{\oplus}$, $g = 17.0 \text{ m s}^{-2}$, around a Sun-like star at 1 au, but with different atmospheric compositions. This type of planet is close to the center of the high-radius population of the $1.7\text{--}3.5 R_{\oplus}$ planet population found by the Kepler survey (Fulton & Petigura 2018). Such a planet can be a rocky core with an extended H_2 -dominated atmosphere (i.e., a gas dwarf), or a planet with 50% of mass being water (i.e., a water world; Zeng et al. 2019). If it is a water world, the extended H_2 -dominated atmosphere is not required to explain the radius, and it can have a thin H_2 -dominated atmosphere, or an N_2 - or CO_2 -dominated atmosphere. This reduces to determining the mean molecular weight of the atmosphere.

We include three different scenarios to cover these possibilities. (1) Pure water world (Figure 3; cyan model): in this scenario we have a high concentration of water below the water cloud layer. On top of the cloud, instead, hydrogen is the dominant gas as the water condensed in the cloud form. We also included a small concentration of CO_2 , CH_4 and N_2 . (2) Gas dwarf (Figure 3; red model): this scenario represent the sub-Neptune class of planets with hydrogen dominated atmosphere, a water cloud layer and absorbing gas (e.g., CO_2 , CH_4 , and N_2). (3) Water world with CO_2 (Figure 3; black model): in this scenario we will test the Bayesian retrieval in an

Table 1
Atmospheric Parameters Used to Simulate the Realistic Scenarios

Parameter	Pure Water World	Gas Dwarf	Water World With CO_2
$\text{Log}(P_{\text{top,H}_2\text{O}})$ [Pa]	4.85	4.85	4.85
$\text{Log}(D_{\text{clid,H}_2\text{O}})$ [Pa]	4.30	4.30	4.30
$\text{Log}(\text{CR}_{\text{H}_2\text{O}})$	-5.00	-5.00	-5.00
$\text{Log}(\text{VMR}_{\text{H}_2\text{O}})$	-0.30	-3.00	-0.35
$\text{Log}(\text{VMR}_{\text{CO}_2})$	-3.40	-3.40	-0.35
$\text{Log}(\text{VMR}_{\text{CH}_4})$	-3.40	-3.40	-3.40
$\text{Log}(\text{VMR}_{\text{N}_2})$	-3.40	-3.40	-3.40
g [m/s^2]	17.0	17.0	17.0
μ	10.05	2.06	28.13

environment in which there are two dominant gases (i.e., H_2O and CO_2) of similar abundance, a water cloud layer and absorbing gas (i.e., CH_4) and spectrally inactive gases (i.e., N_2 and H_2).

As we can note from Figure 3, the three models have the same baseline in terms of flux ratio because the physical parameters of the planet and the simulated system have not been changed. The differences in the spectra are only due to different chemical composition of the simulated atmosphere. The three spectra show some similarities but the main constituent is different in each of them. The atmospheric values used to simulate the three spectra are reported in Table 1.

For all the scenarios presented here, we considered an S/N of 20 at $0.7 \mu\text{m}$ at a spectral resolution of 140 in the optical ($\lambda \leq 1.0 \mu\text{m}$) and 40 in the infrared ($\lambda \geq 1.0 \mu\text{m}$) without making any specific assumption on the integration time which is not the focus of this work. This is consistent with the science plan of HabEx (Gaudi et al. 2020).

4. Results

4.1. Idealized Scenario

As we described in Section 3, we initially tested our updated model with the simplest idealized case: an atmosphere with

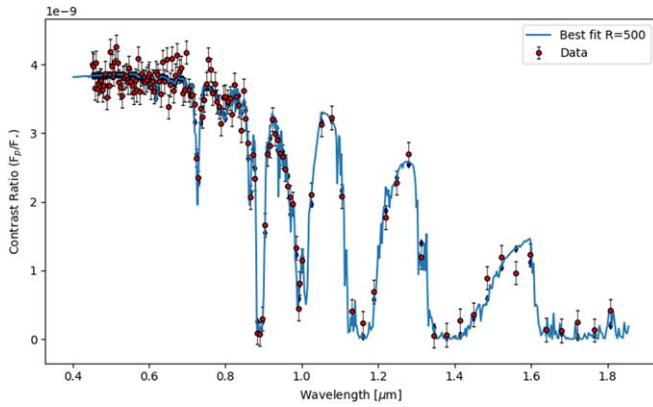


Figure 4. Simulated data and retrieved model in the water world scenario.

solely water and the filling gas, i.e., H_2 . We run eight retrievals as the simulated cases in which the amount of water and hydrogen in the atmosphere was changed. When the mixing ratio of water (below the cloud) is in the range of $10^{-5} \sim 10^{-3}$, the cloud position and the water abundance below the cloud can be retrieved (Figure 7). As shown in the spectra (Figure 2), when the VMR of water is larger than 10^{-2} the spectra are degenerate and the model is not able to correctly identify the truths (i.e., producing a broad posterior distribution of water mixing ratio between 10^{-2} and 1; Figure 7). The retrieval model always identifies H_2 as the dominant gas even though the role of the two gases swaps in the eight simulations (Figure 7). The major driver of these degeneracies is the water cloud in the atmosphere. When the water amount is low, the clouds are not completely opaque in all layers, and therefore some of the light can reach deep into the cloud layers of the atmosphere. On the contrary, when the VMR of water is higher than 10^{-2} the cloud becomes optically thick and behaves more like a reflecting deck, causing the spectrum to be dominated solely by the cloud single-scattering albedo.

Using the idealized scenarios we have also tested the performance of the new prior functions developed in Section 2.2 and the flat prior in the CLR space. Figure 8 shows the posterior distribution calculated by EXOREL^{3R} when the two set of priors are used to retrieve one of the idealized cases, i.e., $10^{-3} H_2O$ scenario. In the case of the flat prior, the framework found two solutions. One of the results is consistent with the true value, the other result shows different atmospheric characteristics. In particular, the retrieved amount of water is close to unity, and the cloud layers extend deeper in the atmosphere. If the modified priors are used (blue model in Figure 8), only one solution is found and it is the one consistent with the truths.

4.2. Water World

The first of the three realistic scenarios we considered in this work is a water world. In this scenario, the amount of water below the clouds is considered to be high, then it drops with altitude as it condenses into cloud form.

The simulated data and the best fit calculated by EXOREL^{3R} are shown in Figure 4. Our model was able to fit the data, and in particular it was able to reconstruct all the spectroscopic features due to the absorption of CO_2 and CH_4 . In Appendix C we reported the 1D probability distribution functions of the free and derived parameters along the 2D correlation maps of the parameters. In the posterior distribution (Figure 9) the 2σ

Table 2
Atmospheric Retrieved Parameters for the Water World Scenario

Parameter	3σ	2σ	1σ
$\text{Log}(P_{\text{top},H_2O})$	✓	✓	
$\text{Log}(D_{\text{cld},H_2O})$	✓	✓	
$\text{Log}(CR_{H_2O})$	✓		
$\text{Log}(VMR_{H_2O})$			
$\text{Log}(VMR_{CO_2})$	✓	✓	✓
$\text{Log}(VMR_{CH_4})$	✓	✓	✓
$\text{Log}(VMR_{N_2})$	✓	✓	✓
$\text{Log}(VMR_{H_2})^*$			
g [m/s^2]			
μ^*			

Note. The parameters that are usefully constrained are checked.

confidence level is reported. In Table 2, we indicated whether or not the true value of the free parameters falls inside 1, 2, or 3σ of the posterior distribution in Figure 4. The result calculated by EXOREL^{3R} (see Figure 9) gives a clear overview of the performances of the model in handling cases with completely unknown atmospheric composition. First of all, the model was able to correctly retrieve the position of the water cloud in the atmosphere (P_{top,H_2O} and D_{H_2O}) within 2σ . Second, the overall composition of the atmosphere has been retrieved with CH_4 , CO_2 , and N_2 retrieved within 1σ . The amount of water, and therefore the concentration of hydrogen (derived), was not retrieved within 3σ . However, the model was able to identify water as the dominant gas below the cloud layer, and hydrogen dominated above it. This can be easily visualized on the retrieved chemistry profile of the atmosphere reported in Appendix C. Because of the overestimation of the water content in the atmosphere, the fitted gravity of the planet and the mean molecular mass (μ , derived) are also slightly off.

Finally, it is worth noting that the model was able to identify the atmosphere to be a water-dominated one. In comparison, the model applied to the same atmosphere without CH_4 or CO_2 (i.e., the 50% water case in the idealized scenarios; Section 4.1) would return a water mixing ratio between 10^{-2} and unity without preference, and thus cannot identify the atmosphere being hydrogen or water-dominated. Including other absorbing gases breaks the degeneracy. The absorption features of CH_4 and CO_2 vary in depth accordingly to the thickness of the atmosphere. If the atmosphere was lighter, i.e., an H_2 -dominated atmosphere, the absorption features would have been deeper.

4.3. Gas Dwarf

In this scenario, we simulated a small planet with an H_2 -dominated atmosphere with a water cloud layer and other absorbing gases.

The simulated data and the best fit calculated by EXOREL^{3R} are shown in Figure 5. The model was able to fit the data, and to recover all the spectroscopic features due to the absorption of CO_2 and CH_4 . In the Appendix D we reported the full posterior distribution calculated by EXOREL^{3R} (see Figure 10). In Table 3, we indicated whether or not the true value of the free parameters falls inside 1, 2, or 3σ of the posterior distribution in Figure 5. The result (see Figure 10) gives a clear overview of the performances of the model in handling cases in which the dominant gas is a spectrally inactive gas. Also in this case, the model was able to correctly retrieve the position of the water

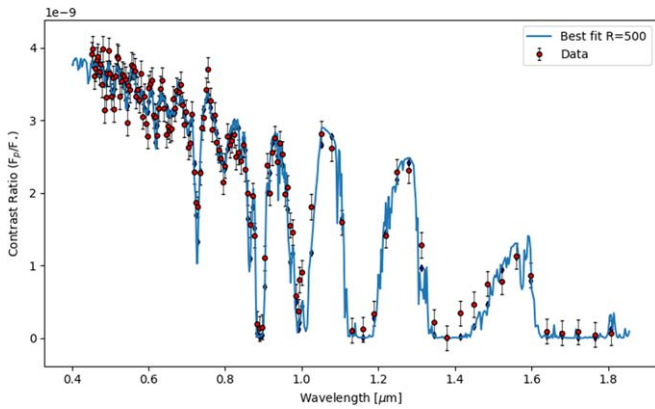


Figure 5. Simulated data and retrieved model in the gas dwarf scenario.

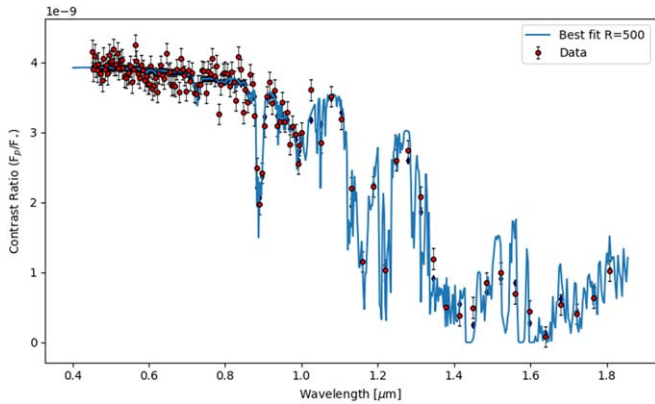


Figure 6. Simulated data and retrieved model in the water world with CO₂ scenario.

cloud in the atmosphere ($P_{\text{top,H}_2\text{O}}$ and $D_{\text{H}_2\text{O}}$) within 1σ . The overall composition of the atmosphere has been retrieved within 1σ including water and hydrogen. This can be easily visualized on the retrieved chemistry profile of the atmosphere reported in Appendix D. In this case, the gravity of the planet and the derived mean molecular mass match the simulated values within 1σ .

4.4. Water World with CO₂

In this last case, we wanted to simulate an heavy atmosphere without having a single dominant gas. Below the cloud the concentrations of H₂O and CO₂ are ~ 0.5 of the total atmosphere. Above the cloud, the water abundance dropped as it condensed, while the VMR of CO₂ was kept constant.

This is the case where the compositional analysis shows its strength. The simulated data and the best fit calculated by EXOREL³ are shown in Figure 6. Also in this case, the model calculated by EXOREL³ fits the simulated data. Moreover, also the free parameters have been retrieved with high confidence. In Table 4, we indicated whether or not the true value of the free parameters falls inside 1, 2, or 3σ of the posterior distribution in Figure 6. We reported the full posterior distribution in the Appendix E. The cloud has been positioned with high confidence at 10^5 Pa, and the gravity of the planet was correctly estimated.

The important result here is that the model was able to identify both CO₂ and H₂O as dominant gases without showing

Table 3
Atmospheric Retrieved Parameters for the Sub-Neptune Scenario

Parameter	3σ	2σ	1σ
$\text{Log}(P_{\text{top,H}_2\text{O}})$	✓	✓	✓
$\text{Log}(D_{\text{cld,H}_2\text{O}})$	✓	✓	✓
$\text{Log}(\text{CR}_{\text{H}_2\text{O}})$	✓	✓	
$\text{Log}(\text{VMR}_{\text{H}_2\text{O}})$	✓	✓	✓
$\text{Log}(\text{VMR}_{\text{CO}_2})$	✓	✓	✓
$\text{Log}(\text{VMR}_{\text{CH}_4})$	✓	✓	
$\text{Log}(\text{VMR}_{\text{N}_2})$	✓	✓	✓
$\text{Log}(\text{VMR}_{\text{H}_2})^*$	✓	✓	✓
g [m/s^2]	✓	✓	✓
μ^*	✓	✓	✓

Note. The parameters that are usefully constrained are checked.

Table 4
Atmospheric Retrieved Parameters for the H₂O–CO₂ Atmosphere Scenario

Parameter	3σ	2σ	1σ
$\text{Log}(P_{\text{top,H}_2\text{O}})$	✓	✓	
$\text{Log}(D_{\text{cld,H}_2\text{O}})$	✓	✓	✓
$\text{Log}(\text{CR}_{\text{H}_2\text{O}})$	✓	✓	
$\text{Log}(\text{VMR}_{\text{H}_2\text{O}})$	✓	✓	✓
$\text{Log}(\text{VMR}_{\text{CO}_2})$	✓	✓	✓
$\text{Log}(\text{VMR}_{\text{CH}_4})$	✓	✓	
$\text{Log}(\text{VMR}_{\text{N}_2})$	✓	✓	✓
$\text{Log}(\text{VMR}_{\text{H}_2})^*$	✓	✓	✓
g [m/s^2]	✓	✓	✓
μ^*	✓	✓	✓

degenerate solutions. The overall composition has been retrieved correctly within $1\text{--}2\sigma$. It is worth nothing that the concentration of N₂ has a broad probability distribution, suggesting that there might no be enough information to exactly fit it. Also, the active gases (i.e., H₂O, CO₂, and CH₄) show correlations with each other within a small range of values therefore not impacting the overall solution. As the other cases, the retrieved chemical profile is provided (see Appendix E). From here, it is possible to easily visualize that EXOREL³ was able to correctly interpret the chemical structure of the atmosphere.

5. Discussion

In the work presented here, we updated our model EXOREL³ to interpret exoplanetary atmospheres through compositional analysis without providing any prior information about the dominant gas. This is an important step to achieve as small planets (e.g., Sub-Neptunes, Super-Earths, Earth-like planets) might not have H₂ as dominant constituent of their atmosphere.

5.1. The Set of Priors

We introduced in our model the CLR of the gases as free parameter. The CLR is a more robust tool for compositional analysis compared to the VMR, which is usually used for Bayesian inverse retrieval processes, because it does not favor or disfavor the “filling gas.” The CLR and the VMR are related to each other in a nonlinear relation (Equation (3)). However, we are interested in exploring the VMR as it is used within the

radiative transfer model and it is a direct way to assess the composition of the atmosphere. Therefore, we developed a novel set of prior functions for the CLR within the Bayesian statistical process. This was necessary since we wanted to explore the VMR in a range within $[10^{-12}, 1]$ assigning the same probability to the values within the range. Since the relation between CLR and VMR is nonlinear the resulted prior functions for the CLR are nonconstant across the range (Figure 1).

The resulted shape of the CLR prior functions prevents the abundance for a gas to be extremely high or extremely low unless the data suggests otherwise. However, this does not have an impact in the ability of the model to identify the dominant gas in the atmosphere as suggested by the studied scenarios presented in this work. In Figure 8, we showed that in the case where the framework with the flat prior, a solution with high concentration of water was found. This solution was not found when the modified prior functions were used within the statistical calculation.

5.2. The H_2O – H_2 Degeneracy

When we applied our updated model to a simple idealized case, we found degenerate solutions. In particular, when only water and hydrogen were included in the atmosphere, the model found degenerate solutions when the VMR of water was larger than 10^{-2} . As we explained in the Section 4, since hydrogen does not have strong spectral features in the wavelength range probed in this work (0.4 – $1.8 \mu\text{m}$), when the amount of water is high, the water cloud that forms is dense enough to be optically thick, and the reflected spectrum is dominated by the single-scattering albedo of the cloud. Otherwise, if the water content in the atmosphere is between $10^{-5} \sim 10^{-3}$ the Bayesian process is able to correctly retrieve the VMR of water and hydrogen.

As we have shown with the water world scenarios with and without CO_2 , the degeneracies might be broken when spectrally active gases are included in the model. This is because the spectral signatures are affected by the scale height of the atmosphere, and changing the H_2O/H_2 ratio will affect the mean molecular mass (μ), which will ultimately change the atmospheric scale height.

5.3. Atmospheric Scenarios

We applied our model to different atmospheric scenarios of cold sub-Neptunes planets. The model demonstrated that it will be possible to identify the dominant gas of the atmosphere (even if it is not hydrogen) and the overall composition with high confidence without any prior knowledge. In the case of the gas dwarf scenario, the framework was able to determine that the atmosphere is H_2 -dominated, to retrieve with high confidence the composition of the atmosphere, and to place correctly the water cloud in the atmosphere.

The framework also demonstrated that it would be able to interpret an atmosphere in which there are multiple spectrally active dominant gases. In the case of the water world with CO_2 case, the retrieved values of the VMR of the gases are within 1 – 2σ . The water cloud position has also been retrieved within 1σ . We also reported that the framework did not perform optimally in case of two dominant gases in which one of the two is a spectrally inactive gas, e.g., pure water world scenario. In that case even though the two dominant gases have been identified, their concentrations were not constrained within 3σ .

In Section 4.1, we reported that the model always identifies H_2 as the dominant gas when the water concentration is greater than

10^{-2} . In the water world scenario, the framework suggests the water as the dominant gas. In that scenario, we included similar concentration for water and hydrogen (~ 0.5). The major difference between the water world scenario and the simple idealized scenario is the inclusion of more spectrally active gases. CH_4 and CO_2 add important details to the reflected spectrum that are useful to the Bayesian framework. The depth of the absorption features is indeed affected by the scale height of the atmosphere. If the atmosphere is H_2 or H_2O -dominated atmosphere, the mean molecular mass changes and therefore the atmospheric scale height. The inclusion of spectrally active gases helped to better converge toward the true values.

Interestingly, the retrievals show broad constraints on the mixing ratio of N_2 , a radiatively inactive gas (Figures 9, 10, and 11). This is because the main role of N_2 is mainly to change the mean molecular mass. High value of N_2 would result in a heavy atmosphere and spectral features would be suppressed. This mechanism will define a possible upper limit. The low value of N_2 will instead create a lighter atmosphere where spectral absorption features are enhanced. This mechanism will help to define a lower limit. Finally, N_2 , being a free parameter, is described by the modified priors introduced in this work. Extremely high values or very low values are unlikely, unless required by the data.

6. Conclusion

In this work, we presented the updated version of EXOREL^{3A} for the direct imaging spectroscopy of sub-Neptune exoplanets. The update now allows the Bayesian framework to interpret reflected light spectra of temperate atmospheres with diverse bulk composition without any prior information about the composition. Small planets like those targeted in this study will not necessarily have an H_2 -dominated atmosphere, and therefore the dominant gas is unknown. We introduced the centered-log ratio (CLR) as the free parameter in our framework to allow compositional analysis. However, our attention is always on the more commonly used volume-mixing ratio (VMR). To allow an unbiased exploration of VMR values we developed a new set of prior functions for the CLR. The shape of the distribution of the prior changes according to the number of gases to be fitted in the atmosphere. The more gases are introduced the flatter the prior distribution gets.

We tested the retrieval framework with idealized and realistic atmospheric scenarios. Initially, we simulated the simplest case: an atmosphere with only water and hydrogen. We simulated multiple instances of this scenario in which the water and the hydrogen concentration was changed. We reported that if the water VMR is comprised between 10 – 5 and 10^{-2} , the water cloud present in the atmosphere is not opaque and photons are able to reach lower part of the atmosphere. As a result, the cloud pressure and the water VMR can be retrieved simultaneously from the reflected light spectra. On the contrary, if the water VMR is greater than 10^{-2} , the water cloud is optically thick and more photons are reflected. Increasing the VMR of water above 10^{-2} does not change the modulation of the reflected spectrum resulting in spectral degeneracy. We showed that adding spectral active gases (e.g., CH_4 and CO_2) helps break the degeneracy since now the dominant gas has spectral features and can thus be constrained from data. In fact, the depth of the absorption features depend on the scale height of the atmosphere and the dominant gas plays a crucial role in determining the mean molecular mass value.

With future space missions like HabEx and LUVOIR, temperate sub-Neptune planets will be at the forefront of the exoplanet exploration. Exploring models and implementing new statistic tools is therefore an essential step. In a future work, we will push forward the capabilities of our model to also implement the interpretation of super-Earths and Earth-like planets with a rocky surface that contributes to the overall light reflected by the planet (M. Damiano & R. Hu 2021, in preparation).

This work was supported in part by the NASA WFIRST Science Investigation Teams grant #NNN16D016T. 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 Idealized Scenario

The idealized scenario is the simplest atmospheric scenario we used to test the capabilities of EXOREL³¹. We test the model varying the abundance of water and hydrogen used for the scenarios. As described in Section 4.1, the albedo spectra become degenerate when the water abundance is greater than 10^{-2} (Figures 2 and 7). Moreover, the cloud position has been retrieved for all the cases but the case with water abundance of 10^{-6} . In this case, the forming cloud is too thin to be statistically constrained.

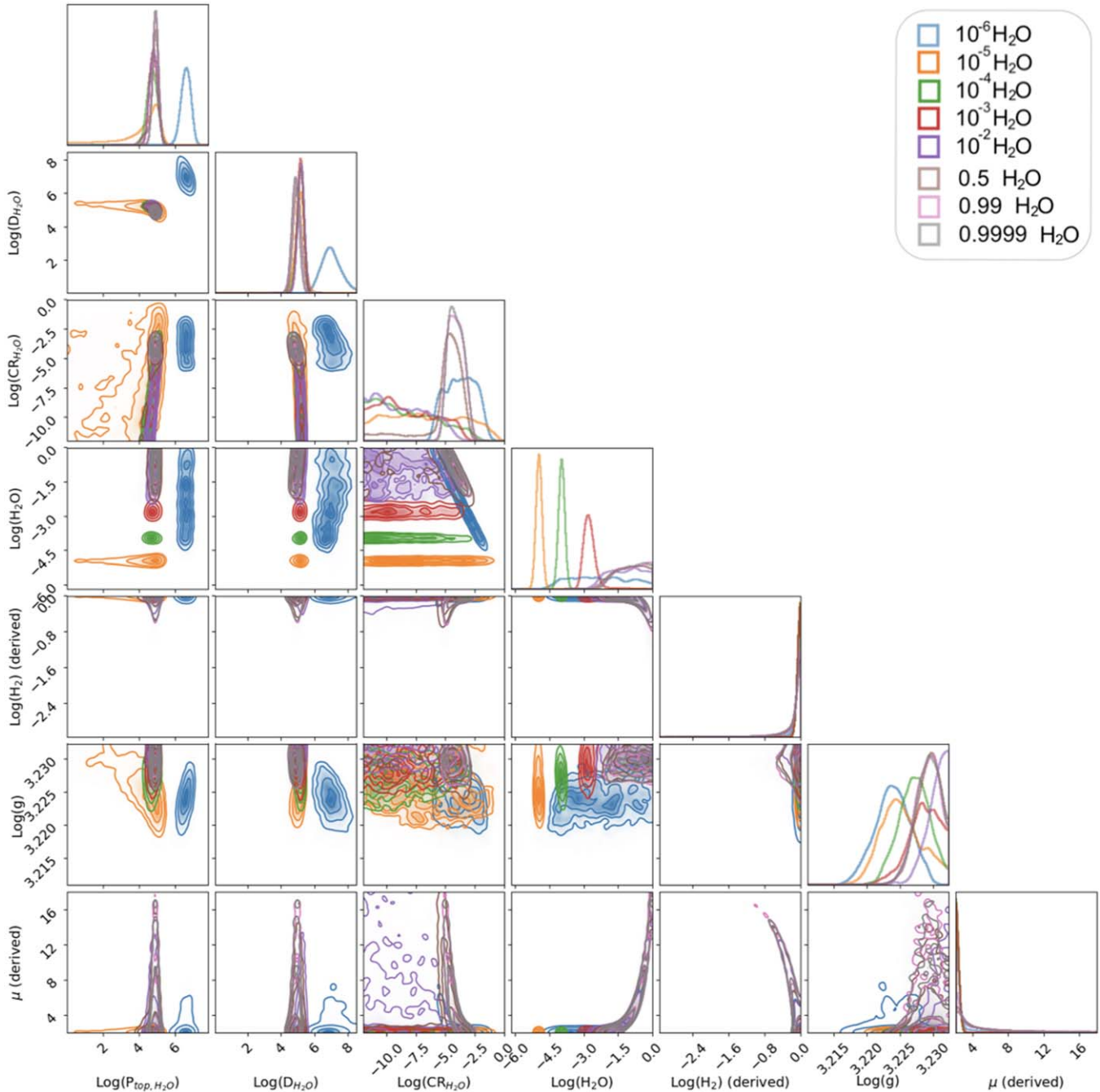


Figure 7. Full posterior distribution of the idealized cases considered in this work. The distribution shows that if the water abundance is $\geq 10^{-2}$ the framework is not able to constrain the atmospheric composition. Moreover, if the mixing ratio of water is $\leq 10^{-6}$, the forming water cloud is too thin to be statistically constrained.

Appendix B Modified Prior Functions Versus Constant Prior in the CLR Space

In Section 2.2 we introduced a new set of prior functions. The choice was justified by the nonlinear transformation from the VMR space to the CLR space. In Figure 8, we compared the result of EXOREL^{3t} when the constant or modified prior functions are used for one of the idealized scenario (i.e., the 10^{-3} H₂O case). As

discussed in Section 2.2, when the constant priors are used the framework tends to statistically overevaluate high-abundance values of the gases (e.g., H₂O abundance in this case). Moreover, the retrieval using the constant prior results in a degenerate solution with a thicker cloud and a higher water mixing ratio below the cloud than the truth. The modified priors introduced in this work prevent the degenerate solution and allow the framework to constrain all the free parameters without degeneracies.

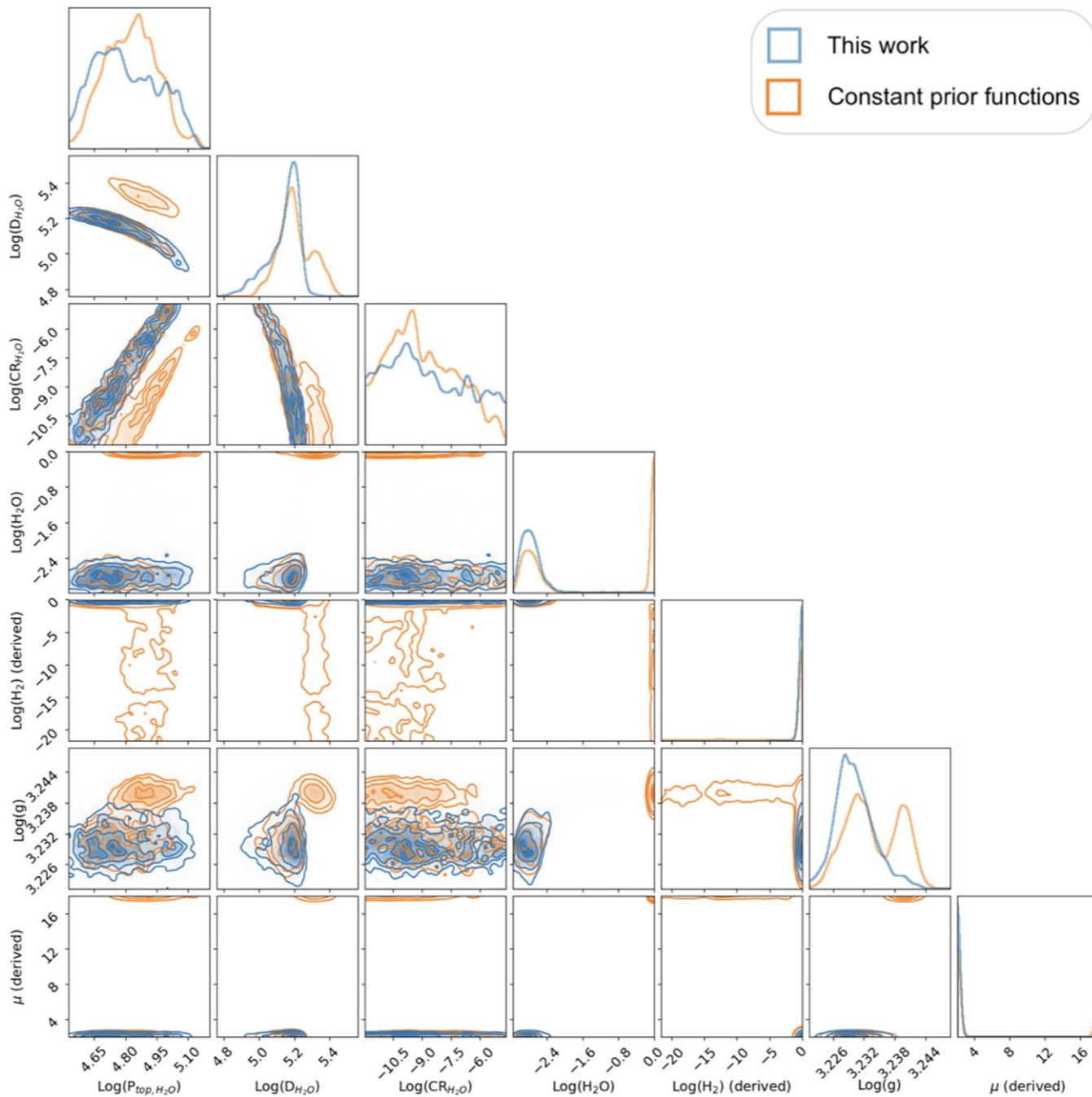


Figure 8. The full posterior distribution of the idealized case with water abundance of 10^{-3} in the case of constant or modified prior functions in the CLR space. The distribution shows that if the constant priors are used, a degeneracies are found in the water concentration and cloud properties.

Appendix C Scenario 1: Pure Water World

When we refer to water world, we intend to describe a planet with high amount of water in the lower part of the atmosphere.

We simulated a H₂O-dominated atmosphere with other absorbing gases. In Figure 9 we show the full posterior distribution and the retrieved chemical profile calculated by EXOREL[®].

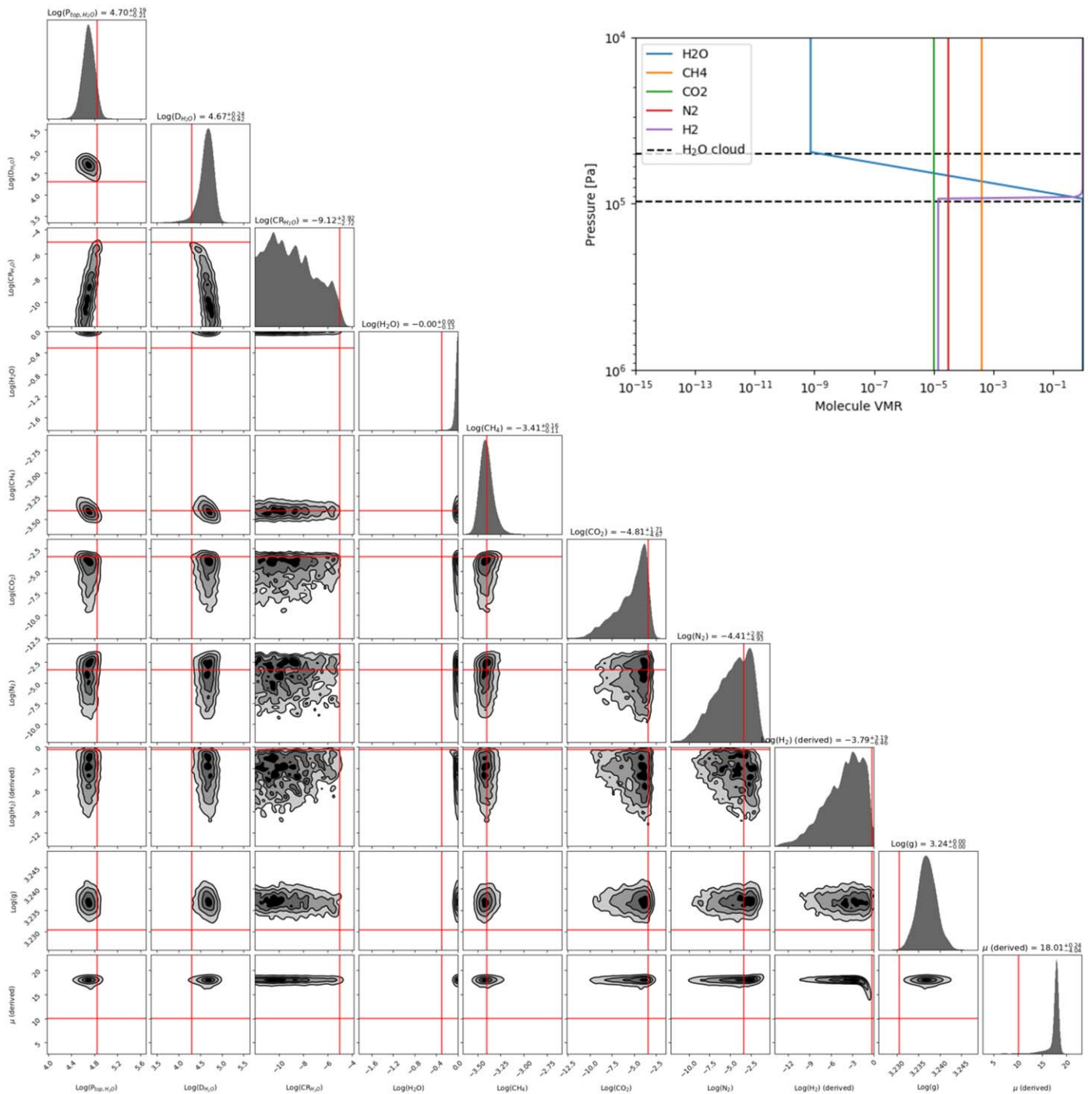


Figure 9. The full posterior distribution (corner plot) and the retrieved chemical profile of the scenario are shown. The red lines in the corner plot refer to the true value used to simulate the data.

Appendix D Scenario 2: Gas Dwarf

The sub-Neptunes are a class of planets which have the dimension of a super-Earth but with a gaseous envelope.

We simulated a H₂-dominated atmosphere with other absorbing gases. In Figure 10 we show the full posterior distribution and the retrieved chemical profile calculated by EXOREL³⁸.

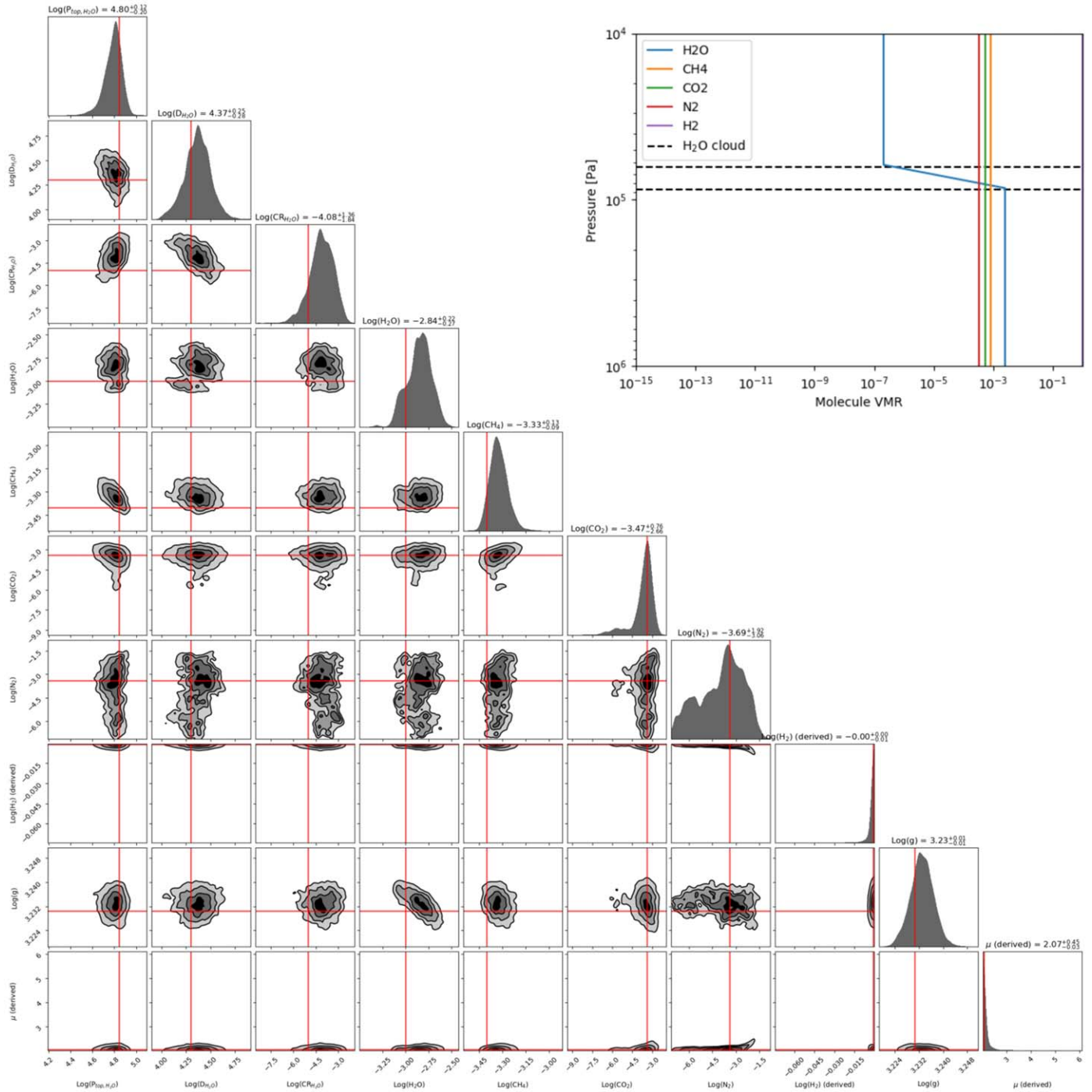


Figure 10. The full posterior distribution (corner plot) and the retrieved chemical profile of the scenario are shown. The red lines in the corner plot refer to the true value used to simulate the data.

Appendix E Scenario 3: Water World with CO₂

We simulated a fictitious atmosphere without a clear single dominant gas. We rather chose to include H₂O and CO₂ with

same VMR (~0.5), and with other absorbing gases. In Figure 11 we show the full posterior distribution and the retrieved chemical profile calculated by EXOREL⁹³.

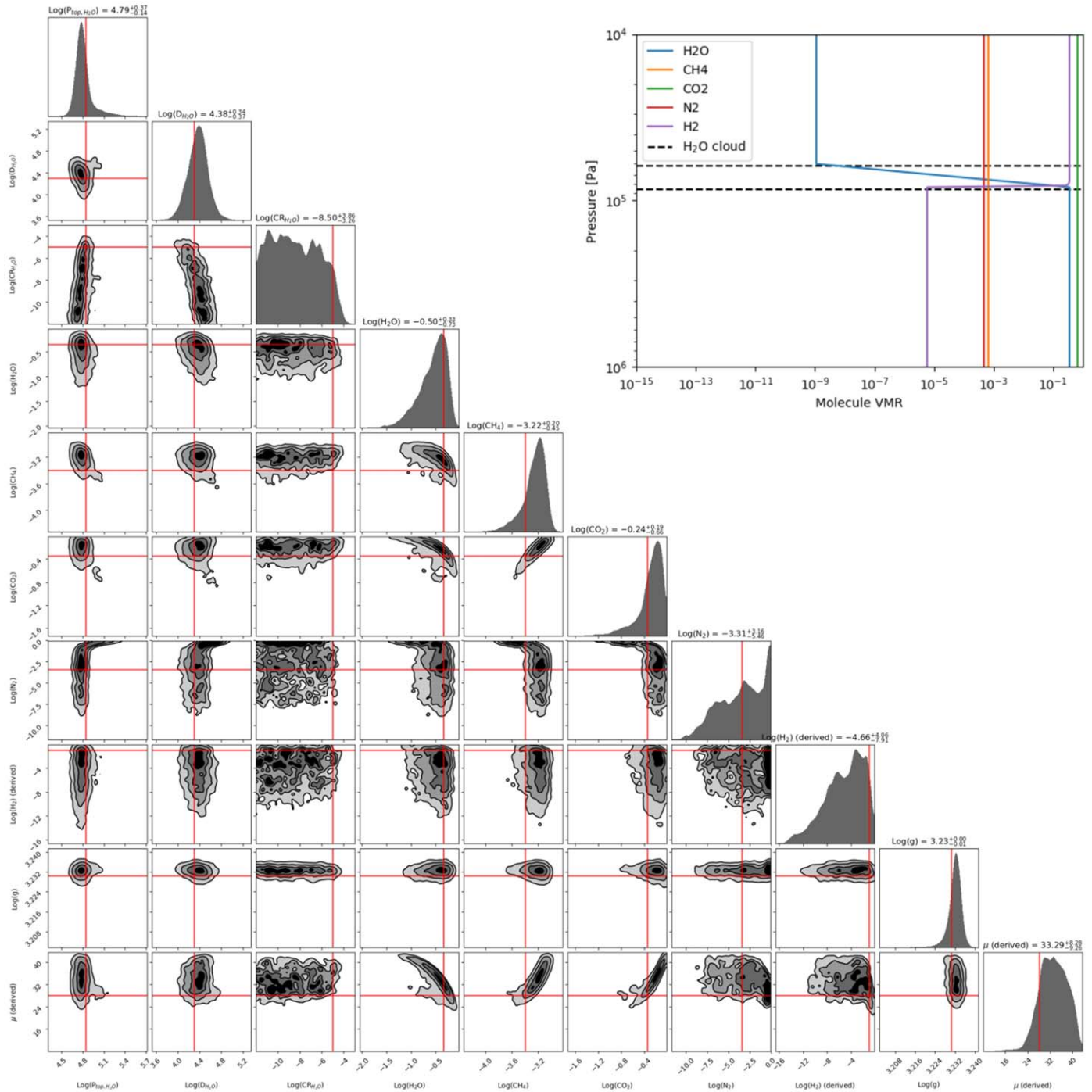


Figure 11. The full posterior distribution (corner plot) and the retrieved chemical profile of the scenario are shown. The red lines in the corner plot refer to the true value used to simulate the data.

Appendix F MULTINEST Implementation

As we describe in Damiano & Hu (2020), we use MULTINEST (Buchner et al. 2014) as the Bayesian engine. The prior functions in MULTINEST are defined by transforming the uniform cube $[0, 1]$ into the range of the physical quantities of interest. In the case of the work presented here, we need to

calculate the inverse of the *cumulative density function* of the probability-density function of the CLR, obtained as described in Section 2.2. Figure 12 shows the relation between the uniform cube and the CLR of the gases for different number of gases to be fit in the model. Note that once the number n of gases has been chosen, the same prior function is used for all the n gases.

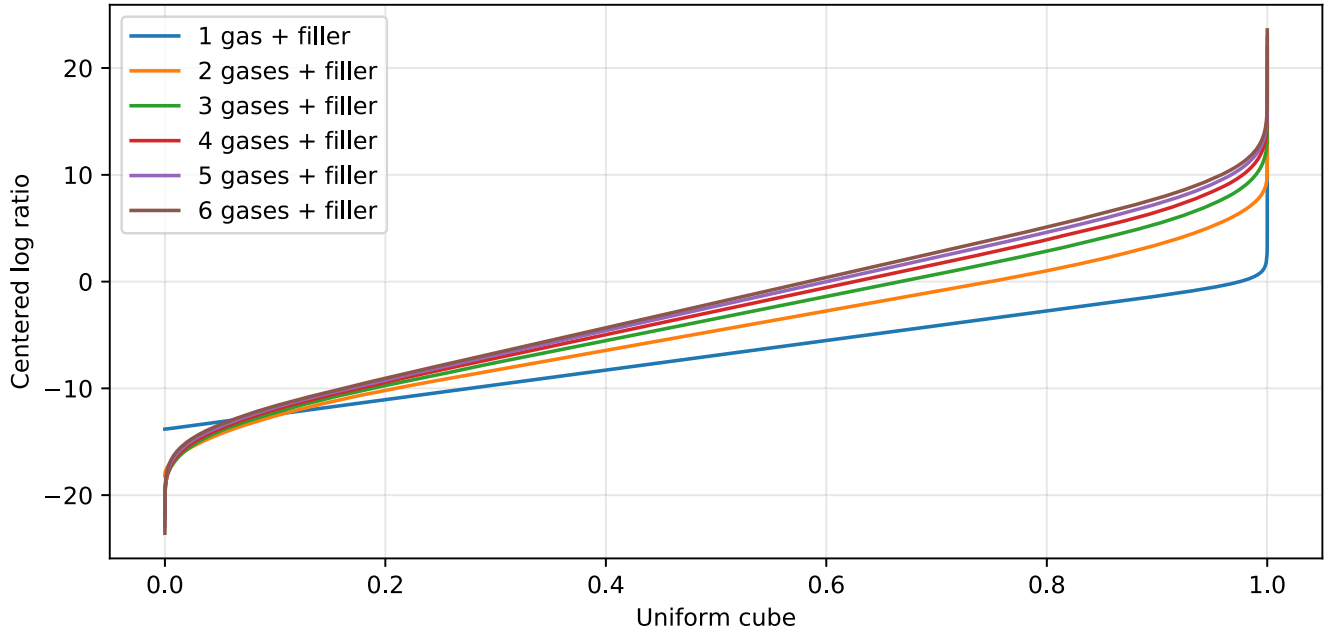


Figure 12. Relation between the uniform cube defined inside MULTINEST and the CLR defined in Section 2.2 for different number of gases retrieved at the same time.

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. Series B Stat. Methodol.*, 44, 139
 Akeson, R., Armus, L., Bachelet, E., et al. 2019, arXiv:1902.05569
 Benneke, B., Knutson, H. A., Lothringer, J., et al. 2019a, *NatAs*, 3, 813
 Benneke, B., & Seager, S. 2012, *ApJ*, 753, 100
 Benneke, B., Wong, I., Piaulet, C., et al. 2019b, *ApJL*, 887, L14
 Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, *A&A*, 564, A125
 Damiano, M., & Hu, R. 2020, *AJ*, 159, 175
 Damiano, M., Hu, R., & Hildebrandt, S. R. 2020, *AJ*, 160, 206
 Feng, Y. K., Robinson, T. D., Fortney, J. J., et al. 2018, *AJ*, 155, 200
 Fulton, B. J., & Petigura, E. A. 2018, *AJ*, 156, 264
 Gaudi, B. S., Seager, S., Mennesson, B., et al. 2020, arXiv:2001.06683
 Hu, R. 2019, *ApJ*, 887, 166
 Hu, R., & Seager, S. 2014, *ApJ*, 784, 63
 Knutson, H. A., Benneke, B., Deming, D., & Homeier, D. 2014, *Natur*, 505, 66
 Palmer, K. F., & Williams, D. 1974, *JOSA*, 64, 1107
 Roberge, A., & Moustakas, L. A. 2018, *NatAs*, 2, 605
 Robinson, T. D. 2018, in *Handbook of Exoplanets*, ed. H. J. Deeg & J. A. Belmonte (Berlin: Springer), 67
 Seager, S., Kasdin, N. J., Booth, J., et al. 2019, *BAAS*, 51, 106, <https://baas.aas.org/pub/2020n7i106/release/1>
 Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
 Tsiaras, A., Rocchetto, M., Waldmann, I. P., et al. 2016, *ApJ*, 820, 99
 Tsiaras, A., Waldmann, I. P., Tinetti, G., Tennyson, J., & Yurchenko, S. N. 2019, *NatAs*, 3, 1086
 Wakeford, H. R., Sing, D. K., Kataria, T., et al. 2017, *Sci*, 356, 628
 Zeng, L., Jacobsen, S. B., Sasselov, D. D., et al. 2019, *PNAS*, 116, 9723