



# Impact of Planetary Mass Uncertainties on Exoplanet Atmospheric Retrievals

Q. Changeat , L. Keyte , I. P. Waldmann , and G. Tinetti

Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK; [quentin.changeat.18@ucl.ac.uk](mailto:quentin.changeat.18@ucl.ac.uk)*Received 2019 August 17; revised 2020 April 11; accepted 2020 May 1; published 2020 June 17*

## Abstract

In current models used to interpret exoplanet atmospheric observations, the planetary mass is treated as a prior and is measured/estimated independently with external methods, such as radial velocity or transit timing variation techniques. This approach is necessary as available spectroscopic data do not have sufficient wavelength coverage and/or signal-to-noise to infer the planetary mass. We examine here whether the planetary mass can be directly retrieved from transit spectra as observed by future space observatories, which will provide higher quality spectra. More in general, we quantify the impact of mass uncertainties on spectral retrieval analyses for a host of atmospheric scenarios. Our approach is both analytical and numerical: we first use simple approximations to extract analytically the influence of each atmospheric/planetary parameter to the wavelength-dependent transit depth. We then adopt a fully Bayesian retrieval model to quantify the propagation of the mass uncertainty onto other atmospheric parameters. We found that for clear-sky, gaseous atmospheres the posterior distributions are the same when the mass is known or retrieved. The retrieved mass is very accurate, with a precision of more than 10%, provided the wavelength coverage and signal-to-noise ratio are adequate. When opaque clouds are included in the simulations, the uncertainties in the retrieved mass increase, especially for high altitude clouds. However, atmospheric parameters such as the temperature and trace-gas abundances are unaffected by the knowledge of the mass. Secondary atmospheres, expected to be present in many super-Earths, are more challenging due to the higher degree of freedom for the atmospheric main component, which is unknown. For broad wavelength range and adequate signal-to-noise observations, the mass can still be retrieved accurately and precisely if clouds are not present, and so are all the other atmospheric/planetary parameters. When clouds are added, we find that the mass uncertainties may impact substantially the retrieval of the mean molecular weight: an independent characterization of the mass would therefore be helpful to capture/confirm the main atmospheric constituent.

*Unified Astronomy Thesaurus concepts:* [Exoplanet atmospheres \(487\)](#)

## 1. Introduction

In recent years, the study of exoplanetary atmospheres has shifted from the investigation of individual planets to the characterization of populations (e.g., Barstow et al. 2017; Tsiaras et al. 2018; Pinhas et al. 2019). In parallel, detection missions, such as Kepler and TESS, and ground-based observatories are enabling the identification of an increasing number of interesting targets suitable for atmospheric studies. New space observatories and dedicated missions, such as the NASA James Webb Space Telescope (JWST; Bean et al. 2018) and the ESA ARIEL mission (Tinetti et al. 2018), are expected to revolutionize our understanding of the physical and chemical properties of a large and diverse sample of extrasolar worlds. To prepare for these missions, significant resources will be allocated to acquire/refine basic planetary, orbital, and stellar parameters. To maximize the efficiency of the community effort, it is important to prioritize follow-up activities where these are particularly needed. For example, the recent ARIEL ExoClock project<sup>1</sup> provides priorities to guide amateur astronomers in the selection of targets for ephemeris refinement. While for most planets considered for transit spectroscopy, the mass is already constrained from radial velocity observations, it is important to evaluate whether these measurements are precise enough for atmospheric characterization. In the case of low-gravity exoplanets, current masses may have large uncertainties and it is therefore foreseen that refinements from radial velocity (López-Morales et al. 2016) or

transit timing variation techniques (Borsato et al. 2019) will have to be made in preparation for JWST and ARIEL.

Current transit spectroscopic data do not have sufficient wavelength coverage and/or signal-to-noise to infer the planetary mass (Line et al. 2012); therefore, spectral retrieval models include this key parameter as a prior estimated through external methods, such as radial velocity or transit timing variation techniques. This limitation will no longer apply to future space missions and observatories (JWST, ARIEL, Twinkle; see Edwards et al. 2018) designed to provide spectroscopic observations over a broader wavelength range, higher spectral resolution, and signal-to-noise. de Wit & Seager (2013) showed that for atmospheres dominated by a single species, the mass could be retrieved from transit spectra only. de Wit & Seager (2013) also stressed the importance of Rayleigh scattering and collision induced absorption, which are particularly valuable to constraint masses from retrievals. However, in their examples, they only considered atmospheres dominated by a single species. Batalha et al. (2017) highlighted the degeneracy between mean molecular weight and main atmospheric components for planets with a secondary atmosphere—i.e., an atmosphere that has evolved from a pure H/He composition. However, they restricted their analysis to the comparison of forward models for the specific case of a H<sub>2</sub>/H<sub>2</sub>O atmosphere.

Here we aim to determine whether the planetary mass can be directly retrieved from transit spectra observed by future space observatories and, more generally, to quantify the impact of mass uncertainties on spectral retrieval analyses. Compared to previous studies in the literature, we investigate a more

<sup>1</sup> <https://www.exoclock.space>

comprehensive suite of atmospheric scenarios and we adopt a fully Bayesian retrieval to quantify the propagation of the mass uncertainty onto other atmospheric parameters. The paper is comprised of two main sections. The first section explores analytically the role of the mass in transit spectroscopy and the contribution of the different parameters to the optical depth. In the second section, we use atmospheric retrieval techniques to illustrate/confirm the predictions made in Section 1 and estimate the mass uncertainties in various key examples.

## 2. Analytical Study

### 2.1. Derivation

We investigate here the impact of the planetary mass to the wavelength-dependent transit depth,  $C_{\text{atm}}$ . Here, the goal is to use simple approximations to extract analytically the influence of each atmospheric/planetary parameter to the transit depth. We present in this section the key steps but the detailed derivation can be found in Appendix B. We follow the approach taken by Brown (2001), Fortney (2005), Lecavelier des Etangs et al. (2008), de Wit & Seager (2013), and Heng et al. (2015). For a clear-sky atmosphere we have:

$$C_{\text{atm}}(\lambda) = 2\pi \int_0^{z_{\text{max}}} (R_0 + z)(1 - \exp[\tau(z, \lambda)]) dz, \quad (1)$$

where  $R_0$  is the radius at which the atmosphere becomes opaque at all wavelengths,  $z$  is the altitude from  $R_0$ , and  $\lambda$  is the wavelength.  $\tau$  is the optical depth, i.e.,

$$\tau(z, \lambda) = 2 \int_0^{x_f} \sum_i n_{0i} e^{-\frac{z}{H}} e^{-\frac{x^2}{2(R_0+z)H}} \sigma_i(p, T, \lambda) dx, \quad (2)$$

where  $x$  is the distance from the planet normal,  $n_{0i}$  is number density of species  $i$  at  $z = 0$ ,  $H$  is the scale height,  $\sigma_i$  is the cross section of the species  $i$ ,  $p$  is the pressure, and  $T$  is the temperature. We can estimate the temperature and pressure dependence of the cross sections  $\sigma$  by assuming a linear interpolation from tables of known values of  $\sigma$ , as currently done in most retrieval models (Hill et al. 2013 and Barton et al. 2017).

$$\sigma_i(T) = \sigma_i(T_1) + \frac{\sigma_i(T_2) - \sigma_i(T_1)}{T_2 - T_1}(T - T_1), \quad (3)$$

$$\sigma_i(p) = \sigma_i(p_j) + \frac{\sigma_i(p_{j+1}) - \sigma_i(p_j)}{p_{j+1} - p_j}(p - p_j), \quad (4)$$

where  $T_1$ ,  $T_2$ ,  $p_j$  and  $p_{j+1}$  are temperatures and pressures known from cross-section tables. Since the pressure differences across the  $x$ -axis are large (larger than the interpolation intervals), we sum over intervals  $(x_j, x_{j+1})$  of known pressures  $(p_j, p_{j+1})$ . This approximation allows us to derive analytically the path integral along the line of sight.

$$\begin{aligned} \tau(z, \lambda) = & \sum_j \sum_i n_{0i} e^{-\frac{z}{H}} \sqrt{\pi(R_0 + z)H} (\sqrt{2} I_j \sigma_i(p_j, T_1) \\ & + K_{ij}^T \sqrt{2} I_j (T - T_1) + (K_{ij}^p + K_{ij}^X (T - T_1))(p_0 I_j' e^{-\frac{z}{H}} \\ & - \sqrt{2} I_j p_j). \end{aligned} \quad (5)$$

The coefficients  $K_{ij}^{T,p,X}$  are the derivatives of the cross section with respect to either  $T$ ,  $p$  or both (their expression is given in Appendix B).  $I_j$  represents the integration of the opacity along the  $x$ -axis. Finally, the scale height  $H$  is

defined by:

$$H = \frac{k_b T (R_0 + z)^2}{\mu M_p G}, \quad (6)$$

where  $\mu$  is the mean molecular mass of the atmosphere,  $G$  is the gravitational constant,  $k_b$  is the Boltzmann constant, and  $M_p$  is the planetary mass.

### 2.2. Interpretation

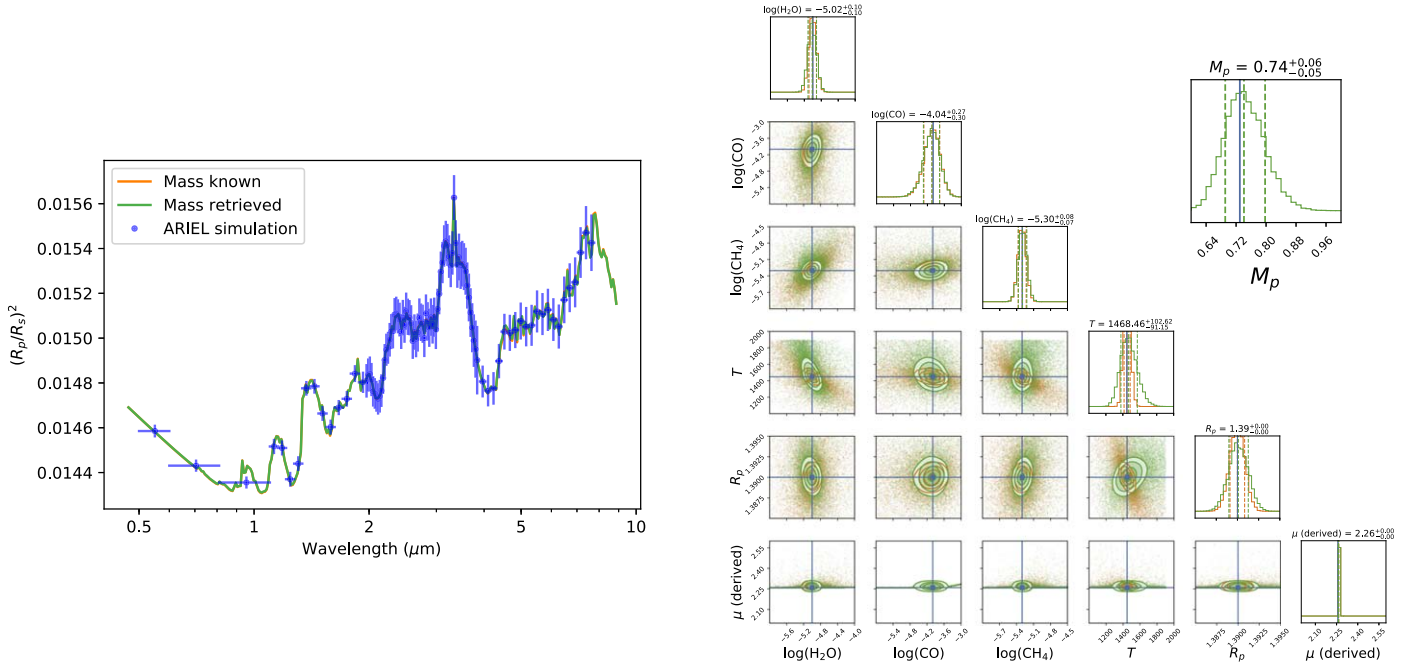
The equations derived in the previous section can be used to predict the degeneracies we expect in retrieval simulations. Similar equations and degeneracies have been studied in previous works (Brown 2001; Fortney 2005; Lecavelier des Etangs et al. 2008; de Wit & Seager 2013; Griffith 2014; Line & Parmentier 2016; Rocchetto et al. 2016; Batalha et al. 2017; Heng & Kitzmann 2017; Fisher & Heng 2018; Tinetti et al. 2018; Welbanks & Madhusudhan 2019). We summarize here the key findings that are relevant for our discussion on planetary mass. In general, the mass is expected to be well retrieved as its contribution to the transit depth calculation is uniquely constrained by the atmospheric scale height. The mass appears only in the scale height definition, while the other parameters are constrained from other individual contributions to the opacity.

1.  $R_0$  is the radius at which a clear-sky atmosphere becomes opaque at all wavelengths. In the case of a cloudy atmosphere (gray clouds), degeneracies may exist as  $R_0$  cannot be detected accurately below the cloud deck.
2. For gaseous planets,  $\mu$  is usually equal to roughly  $\sim 2.3$ , defined by the ratio  $\text{H}_2/\text{He}$  only. In secondary atmospheres, a wider range of main atmospheric components may exist and therefore  $\mu$  is degenerate with  $M_p$  in Equation (5).
3. The temperature has a similar role to the mass in the definition of the scale height (i.e., when an increase of the mass translates into a contraction of the atmosphere, a decrease of the temperature essentially plays the same role). However, the temperature is expected to change with altitude. Also, the temperature dependence of the cross sections could allow the temperature contribution to be distinguishable from the mass contribution if the observations are good enough and depending on the considered species and the atmospheric conditions.
4. The trace gases' number densities,  $n_{0i}$  may change with altitude but otherwise are independent from the other parameters, including the mass.

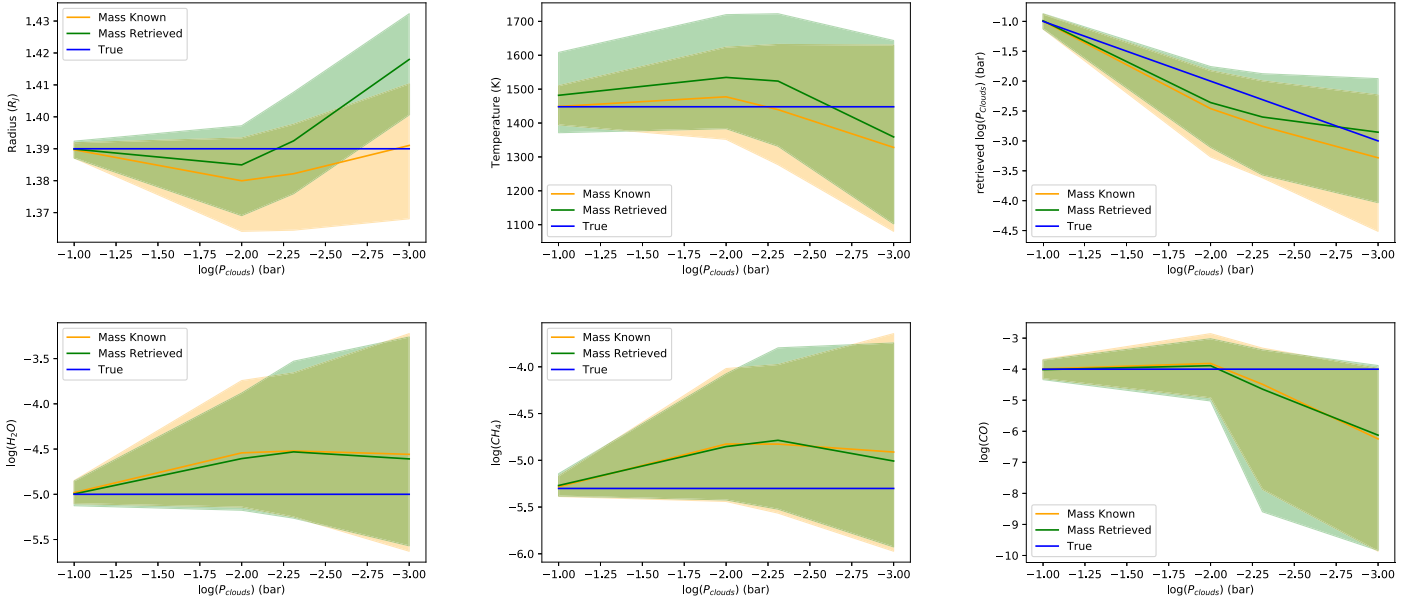
## 3. Retrieval Analysis

### 3.1. Methodology

In this section, we complement the analytical derivation in Section 2 with a number of relevant examples from retrieval simulations. We consider both primary and secondary atmospheres relevant to gaseous planets and super-Earths. We make extensive use of the open-source TauREx model (Waldmann et al. 2015a, 2015b) to simulate different atmospheric scenarios and perform retrievals. TauREx is a fully Bayesian radiative transfer and retrieval framework that encompasses molecular line-lists from the ExoMol project (Tennyson et al. 2016), HITEMP (Rothman & Gordon 2014), and HITRAN (Gordon et al. 2016).



**Figure 1.** Spectra (left) and posteriors distribution (right) for a hot-Jupiter with a clear-sky atmosphere. Orange plots: the mass is known. Green plots: the mass is retrieved. The blue crosses indicate either the simulated ARIEL observations (left plot) or the ground truth values (right plot).



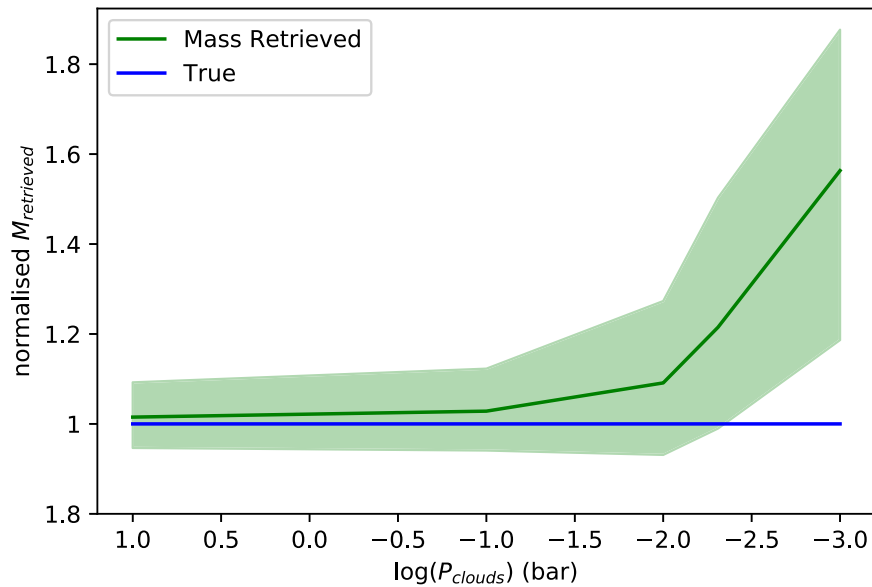
**Figure 2.** Comparison between the known/retrieved mass cases as a function of cloud pressure. The clear-sky case is rendered by placing the cloud deck at 10 bar.

For each case we begin by using TauREx in forward mode to generate a high-resolution theoretical spectrum. For the purpose of this investigation, we focus only on transit spectra and assume isothermal profiles. We will cover eclipse spectra and more complex temperature–pressure profiles in a future work. Our model allows us to specify the main constituents of the atmosphere using their relative abundances (ratios of two molecules).

The high-resolution spectrum is convolved through the instrument model of L. Mugnai et al. (2019, in preparation) to simulate a spectrum as observed by ARIEL. Said synthetic spectrum acts as the input to the retrieval. In this study, we focus on observations obtainable with a single transit, except in Section 3.4 where we investigate the benefits on an increased

signal-to-noise obtained by coadding multiple transits. For each case considered we perform two retrievals: in the first case, the planetary mass is assumed to be known; in the second, it is retrieved as a free parameter. The latter allows us to investigate whether the mass can be reliably estimated from transit spectra and assess the impact of mass uncertainties onto the retrieval of other atmospheric properties, such as the concentration of the trace gases, the temperature, and the cloud pressure.

In Section 3.2, we investigate the case of a hypothetical hot-Jupiter, with parameters based on HD 209458b (see Table A1 in Appendix A). We first present the case of a clear atmosphere and then extend the study to consider the impact of clouds at different cloud pressures. In Section 3.3, we investigate the



**Figure 3.** Normalized retrieved mass in the case of a gaseous planet as a function of cloud pressure. The green curve is the retrieved mass with its  $1\sigma$  uncertainty. The blue line is the real value. The clear case is represented by a cloud deck at 10 bar. The retrieved mass is not affected by low altitude clouds ( $P_{\text{clouds}} \approx 0.1$  bar), while for high altitude completely opaque clouds, the retrieved mass starts to diverge from its true value (60% for  $P_{\text{clouds}} = 10^{-3}$  bar).

case of a hypothetical super-Earth with a heavy atmosphere containing a significant fraction of  $\text{N}_2$  or any other inert gas that cannot be detected through the identification of spectroscopic features. Section 3.4 is dedicated to the impact of the signal-to-noise on the retrieved mass, while in Section 3.5 we compare the results of mass retrievals on HST data. Finally, we consider key examples of secondary atmospheres with clouds.

### 3.2. Retrievals of Gaseous Planets

The first set of retrievals focuses on primary atmospheres, i.e., composed mainly of  $\text{H}_2$ , He. The simulated hot-Jupiter is based on HD 209458b and its parent star: the stellar and planetary parameters have been taken from Stassun et al. (2017). For trace gases, we have included  $\text{H}_2\text{O}$ ,  $\text{CH}_4$ , and  $\text{CO}$ , with mixing ratios  $10^{-5}$ ,  $5 \times 10^{-6}$  and  $10^{-4}$ , respectively (e.g., Tsiraras et al. 2018). We first simulate a clear atmosphere case, and then investigate the behavior of the retrievals when clouds are present by varying the pressure of the cloud deck.

*Clear atmosphere.* The fitted spectra and posteriors for both retrievals (“mass known” and “mass retrieved”) in the case of a clear atmosphere are shown in Figure 1. Here the predictions of our analytic derivation still hold: molecular abundances and other parameters exhibit the same posterior distributions, showing that in this case the knowledge of the mass does not impact the results. The  $1\sigma$  mass uncertainty corresponds to about 7% of its value. This uncertainty is propagated to the temperature posteriors, which are slightly larger when the mass is retrieved. However, the temperature is still very well constrained.

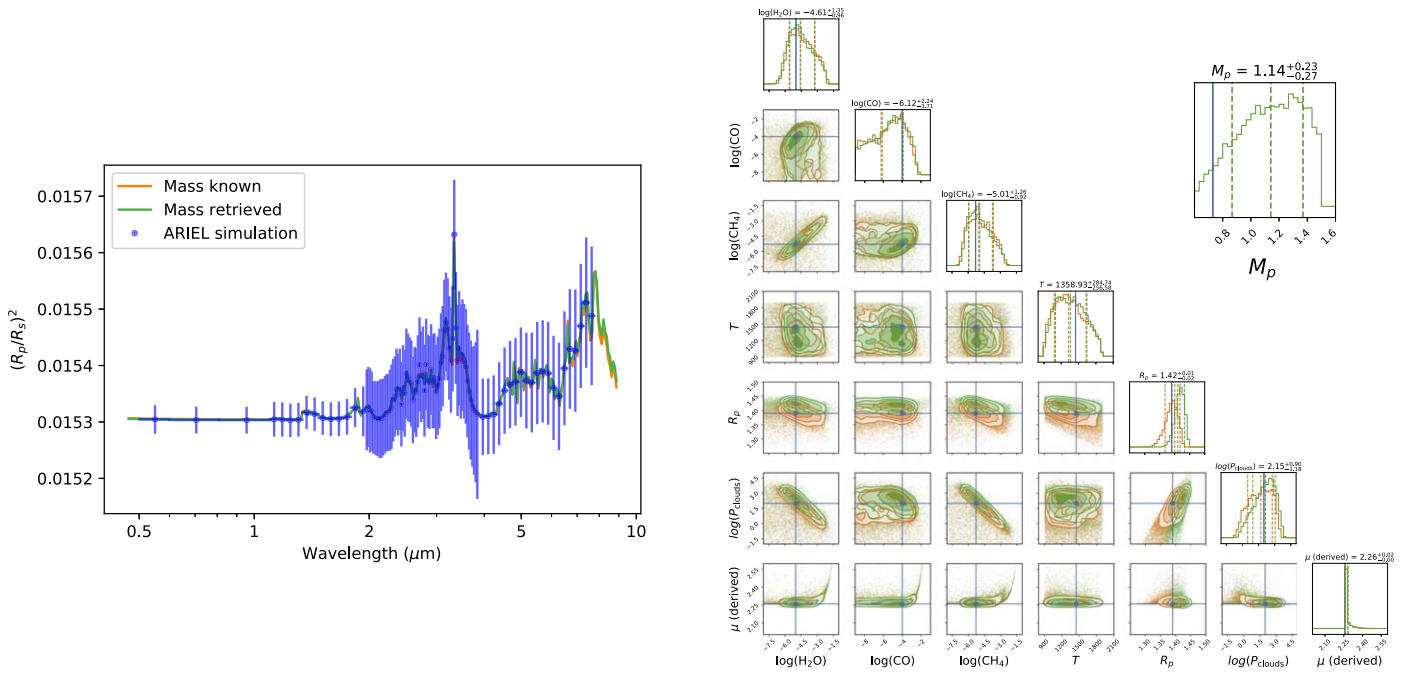
*Overcast atmosphere.* Clouds are modeled by including a completely opaque cloud deck, where the cloud is optically thick below the cloud-top pressure. As mentioned previously, this choice represents the worst case scenario, due to the maximum degeneracy with  $R_0$ , see Equation (B28) in Appendix B. In addition to this issue, we note that with the observing time fixed and optimized for the clear sky case, for the high altitude clouds the signal-to-noise ratio decreases noticeably. Five cases are considered in our analysis:

1. Clear sky case, see Figure 1.
2. Opaque cloud case at  $10^{-1}$  bar.
3. Opaque cloud case at  $10^{-2}$  bar.
4. Opaque cloud case at  $5 \times 10^{-2}$  bar.
5. Opaque cloud case at  $10^{-3}$  bar.

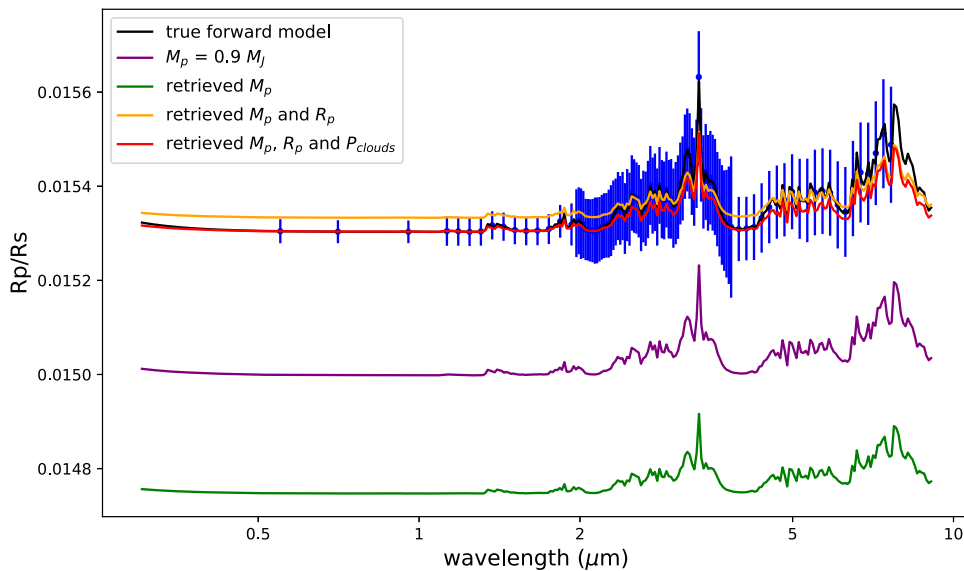
In Figure 2 we plot the comparison between the known/retrieved mass cases as a function of cloud pressure. Some discrepancies appear only in the retrieval of the radius when the cloud pressure gets closer to  $10^{-3}$  bar. For all the other atmospheric parameters, the knowledge of the mass does not impact the retrieved values nor the uncertainties. While the uncertainty of the retrieved values increases when the cloud pressure decreases, as expected, we do not observe a difference between the known and retrieved mass cases. The retrieved trace-gas abundances and temperature are within  $1\sigma$  of the true value.

Focusing on the retrieval of the mass, the results of the normalized retrieved mass for each of the five cases are shown in Figure 3. We appreciate that the mass is well retrieved for all cases with clouds at low altitudes, while the retrieved mass becomes less accurate when the cloud pressure is lower than  $10^{-2}$  bar. At the same time, the  $1\sigma$  spread around the retrieved value also increases with the cloud altitude. The inaccuracy of the retrieved mass for high altitude, opaque clouds appears to be correlated with the inaccuracy of the retrieved radius, as shown in Figure 2. To investigate this important point further, we discuss in detail a specific example of a gas-giant planet with high altitude, opaque clouds.

Figure 4 illustrates an example where the cloud deck is located at  $10^{-3}$  bar: the “known” and “retrieved” mass scenarios are compared. As expected from Figure 2, in this example the retrieved radius  $R_0$  is no longer accurate and the mass is no longer centered around its true value and has large uncertainties. More specifically, the retrieval shows a bias in selecting a larger radius  $R_0$  to fit the spectrum. To compensate this bias, the mass retrieved is centered around a larger value, i.e.,  $1.14 M_J$ , compared to the true value, which is  $0.88 M_J$ . We illustrate the retrieval degeneracy between planetary mass, planetary radius, and cloud top pressure



**Figure 4.** Spectra (left) and posteriors distribution (right) for a hot-Jupiter with a cloudy atmosphere (opaque cloud deck at  $10^{-3}$  bar). Orange plots: the mass is known. Green plots: the mass is retrieved. The blue crosses indicate either the simulated ARIEL observations (left plot) or the ground truth values (right plot).

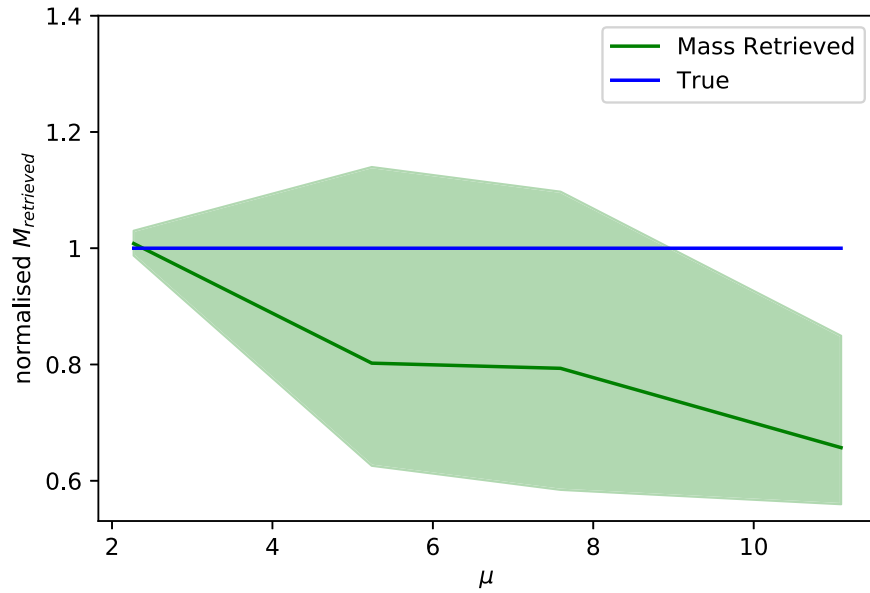


**Figure 5.** Comparison of different forward models based on the cloudy case with cloud top pressure at  $10^{-3}$  bar. Black: true model. Purple: true model where only the mass is changed to  $M_p = 0.9M_J$ . Green: true model where the mass is changed to the retrieved mean value. Orange: true model where the mass and the radius are changed to the retrieved value. Red: true model where the mass, the radius, and the cloud pressure are changed to the retrieved value.

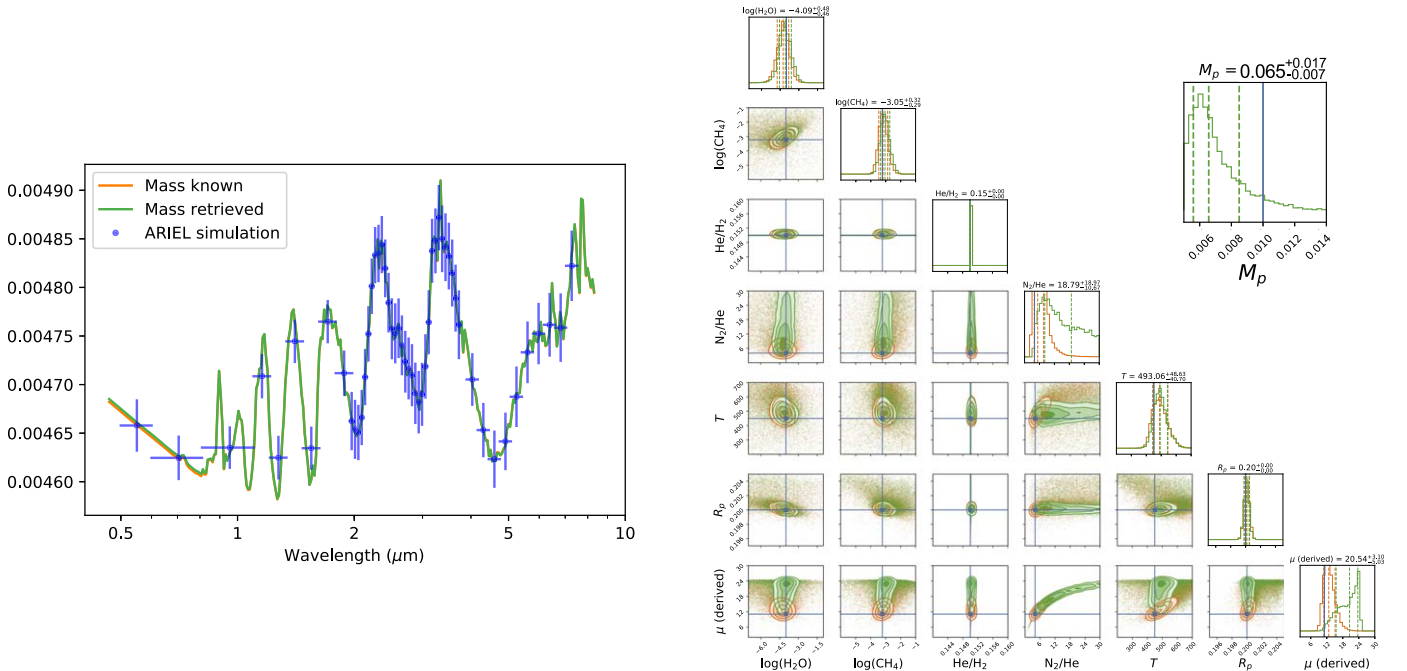
by showing forward models for different cases in Figure 5. We note that, only small variations in the radius—less than 3%—are necessary to compensate for large mass offsets of  $\sim 60\%$ . To mitigate this issue, the target could be observed for a longer time to increase the signal-to-noise ratio, therefore reducing the level of degeneracies among these three parameters (see the discussion in Section 3.4). Again as expected from Figure 2, we do not see significant differences in the other retrieved atmospheric parameters when the mass is known/unknown, which is reassuring for a mission or observing campaign dedicated to probe the atmospheric composition/thermal structure.

### 3.3. Retrieval on Secondary Atmosphere Planets

In this section, we consider secondary atmospheres consisting of elements heavier than H/He. The super-Earth simulated here is taken from the ARIEL Target list (Edwards et al. 2019). The parameters used in our model are reported in Appendix A. We use the inactive gas  $N_2$  to increase the mean molecular weight  $\mu$  of the atmosphere and simulate a host of heavy atmospheres around a rocky planet. In our example, the atmosphere contains  $H_2O$  and  $CH_4$  as trace gases: their absolute abundances are fixed at, respectively,  $10^{-4}$  and  $6 \times 10^{-4}$ . The rest of the atmosphere is filled with a combination of  $H_2$ , He, and  $N_2$ . By varying the



**Figure 6.** Normalized retrieved mass ( $M_{\text{retrieved}}$  in green) for planets with a secondary atmosphere as a function of the mean molecular weight. The blue line represents the real value.



**Figure 7.** ARIEL simulated spectra (left) and posteriors distribution (right) for a cloud-free atmosphere with  $\mu = 11.1$  (i.e.,  $\text{N}_2/\text{He} = 4$ ). Orange plots: the mass is known. Green plots: the mass is retrieved. Blue crosses: simulated ARIEL observations obtained in one transit (left plot) and true values in the posterior distributions (right plots).

$\text{N}_2/\text{He}$  ratio, we essentially control the value of the mean molecular weight.

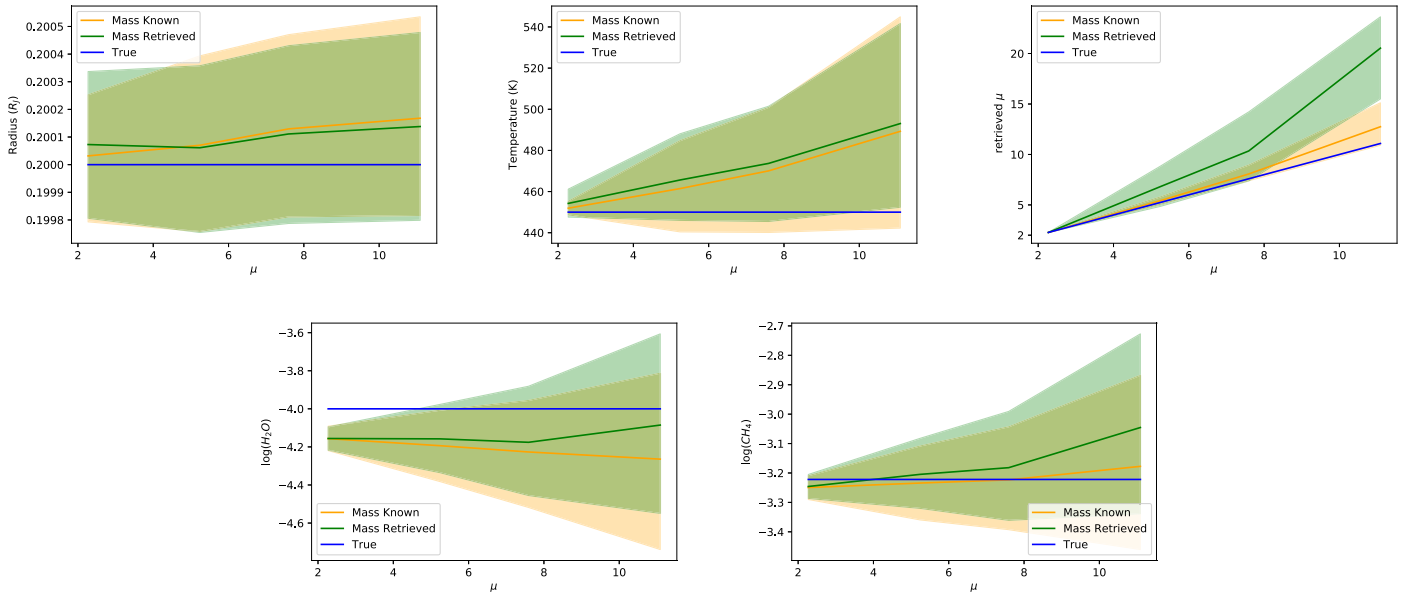
We have deliberately selected  $\text{H}_2$ ,  $\text{He}$ , and  $\text{N}_2$  in our simulations, so that the retrievals will not be guided by any spectral features of these molecules. This choice represents the worst case scenario to assess the degeneracy between the mass and the mean molecular weight. Atmospheres dominated by species such as  $\text{H}_2\text{O}/\text{CO}_2$ /etc., would have traceable molecular features and would therefore represent a more favorable scenario for the inverse models. In this section, we consider the four following cases:

1.  $\mu = 2.3$  ( $\text{N}_2/\text{He} = 0$ ).
2.  $\mu = 5.2$  ( $\text{N}_2/\text{He} = 1$ ).

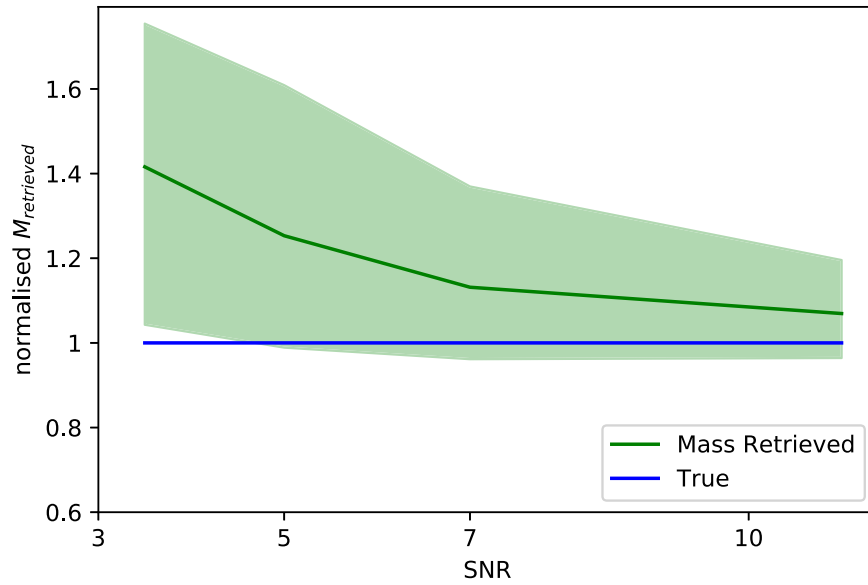
3.  $\mu = 7.6$  ( $\text{N}_2/\text{He} = 2$ ).
4.  $\mu = 11.1$  ( $\text{N}_2/\text{He} = 4$ ).

We show in Figure 6 the normalized mass retrieved as a function of the mean molecular weight  $\mu$ .

At small  $\mu$ , the atmosphere is dominated by a single gas species:  $\text{H}_2$ . This case has already been considered in Section 3.2. The degeneracy mass/mean molecular weight becomes more important for increasing  $\mu$  (see Figure 6). For  $\mu \geq 9$ , the mass is not correctly retrieved due to the degeneracy predicted in Section 2. We show the case  $\mu = 11.1$ , i.e.,  $\text{N}_2/\text{He} = 4$ , in Figure 7, which clearly illustrates the discrepancy between the ground truth and the retrieved  $\mu$  and planetary mass. We find that the space of possible solutions for



**Figure 8.** Impact of the mass on the retrieval of the radius, temperature, mean molecular weight, and trace-gas abundances for different scenarios of heavy atmospheres represented by increasing values of  $\mu$ . The simulated ARIEL observations are obtained in one transit



**Figure 9.** Normalized retrieved mass ( $M_{\text{retrieved}}$  in green) for a  $\text{N}_2$ -rich heavy atmosphere case ( $\mu = 27.8$ ) as a function of S/N. Blue line: real value.

the retrieved mass in the case of secondary atmospheres is not centered around the true value. These results match the conclusions reached by Batalha et al. (2017): by analyzing different cases of heavy atmospheres, they found they could reproduce the same spectra with different sets of parameters. In our simulations, however, we show also that the trace gases, the temperature, and the planetary radius are accurately retrieved with the same posteriors for both the known and retrieved mass cases.

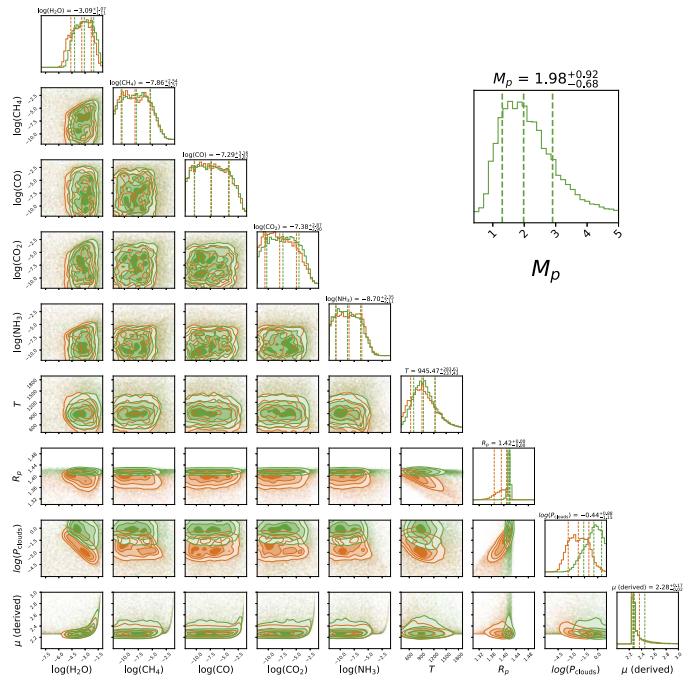
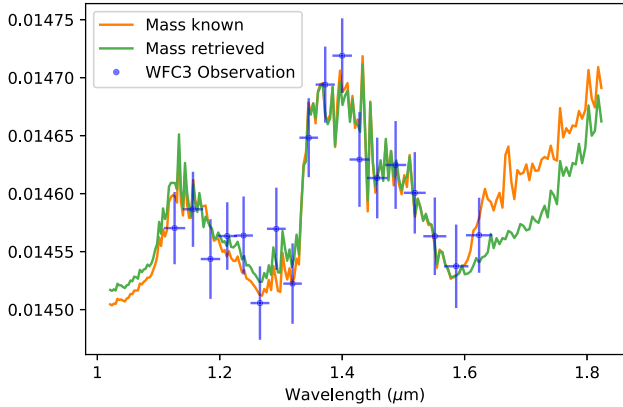
We plot in Figure 8 a comparison of the retrieved parameters as a function of the mean molecular weight. This shows that the temperature, the trace gases, and the radius have similar uncertainties when the mass is known and retrieved for different values of  $\mu$ . The retrieved  $\mu$  is degenerate with the mass and tends to be larger than the true value, hence the complementary smaller retrieved mass in Figure 6: the mean

molecular weight and the mass are inversely correlated in these retrievals. Additionally, for all cases, the retrieved  $\mu$  presents larger uncertainties when the mass is retrieved at the same time.

We conclude that for planets with a secondary atmosphere, an independent determination of the mass can help to break the degeneracy with the mean molecular weight.

### 3.4. Importance of the Signal-to-noise and Wavelength Coverage of the Transit Spectrum to Retrieve the Mass

The larger  $\mu$  is, the smaller the spectral signal is, and therefore it is important to guarantee an adequate signal-to-noise ratio (S/N) when we observe heavy atmospheres. We show in Figure 9 an example of secondary atmosphere with large amounts of  $\text{N}_2$  ( $\mu = 27.8$ ): we plot the normalized retrieved mass error as a function of the S/N.



**Figure 10.** Hubble transit spectra (left) and posteriors distribution (right) for HD 209458b (Tsiaras et al. 2018). Orange plots: the mass is known. Green plots: the mass is retrieved. Blue crosses: Hubble observations.

An adequate wavelength coverage is also very important to reliably retrieve the mass. To illustrate this point, we compare the results of Hubble observations for HD 209458 b (Tsiaras et al. 2018) when the mass is known and retrieved with uniform priors. The best fitted spectra and posterior distributions for both cases are presented in Figure 10.

Similarly to the high-altitude cloud case presented in Section 3.2, the retrieved trace gas abundances and temperature, while not being very precise, are not affected by the mass uncertainties. The main differences appear in the retrieved radius and cloud top pressure. The retrieved mass is not accurate:  $1.98 M_J$  instead of  $0.73 M_J$ . The difference is significant (170%) and demonstrates that a broad wavelength coverage and an adequate S/N is necessary to estimate correctly the mass through transit spectroscopy.

### 3.5. Cloudy Secondary Atmospheres

Finally, we investigate the case of cloudy secondary atmospheres. In this case, the mass is expected to degenerate with both the mean molecular weight (see Figure 7) and the cloud top pressure (see Figures 4 and 5). In Figure 11, we show the simulated spectra and posteriors for two different mean molecular weights:  $\mu = 11.1$  and  $\mu = 7.6$  (corresponding  $N_2/He$  ratios of 4 and 2); opaque clouds are added at  $10^{-2}$  bar.

In Figure 11 we show the cases  $\mu = 11.1$  (top) and  $\mu = 7.6$  (bottom). In the case  $\mu = 7.6$ , the atmosphere is lighter and presents a better signal. Here prior knowledge of the mass allows us to break the degeneracy and retrieve the appropriate  $N_2/He$ , as well as the cloud pressure. Without prior information about the mass, the model retrieves the trace gas abundances and the temperature with equal accuracy/precision but it is not able to constrain the mean molecular weight. In the case of  $\mu = 11.1$ , the retrieval does not properly constrain the  $N_2/He$  ratio and provides a wrong lower limit on its value,

leading to a biased estimate of the mean molecular weight  $\mu$ . Here, additional observations are needed to increase the S/N and to constrain the mean molecular weight.

From these examples, we deduce that the degeneracy with the mean molecular weight is more serious than the degeneracy with clouds, especially as the gray cloud assumption adopted here is pessimistic. Other more realistic cloud models (Lee et al. 2013 or Madhusudhan & Seager 2009) would be more transparent at least in some spectral windows, so that information from the deeper atmosphere could be captured.

## 4. Discussion and Conclusions

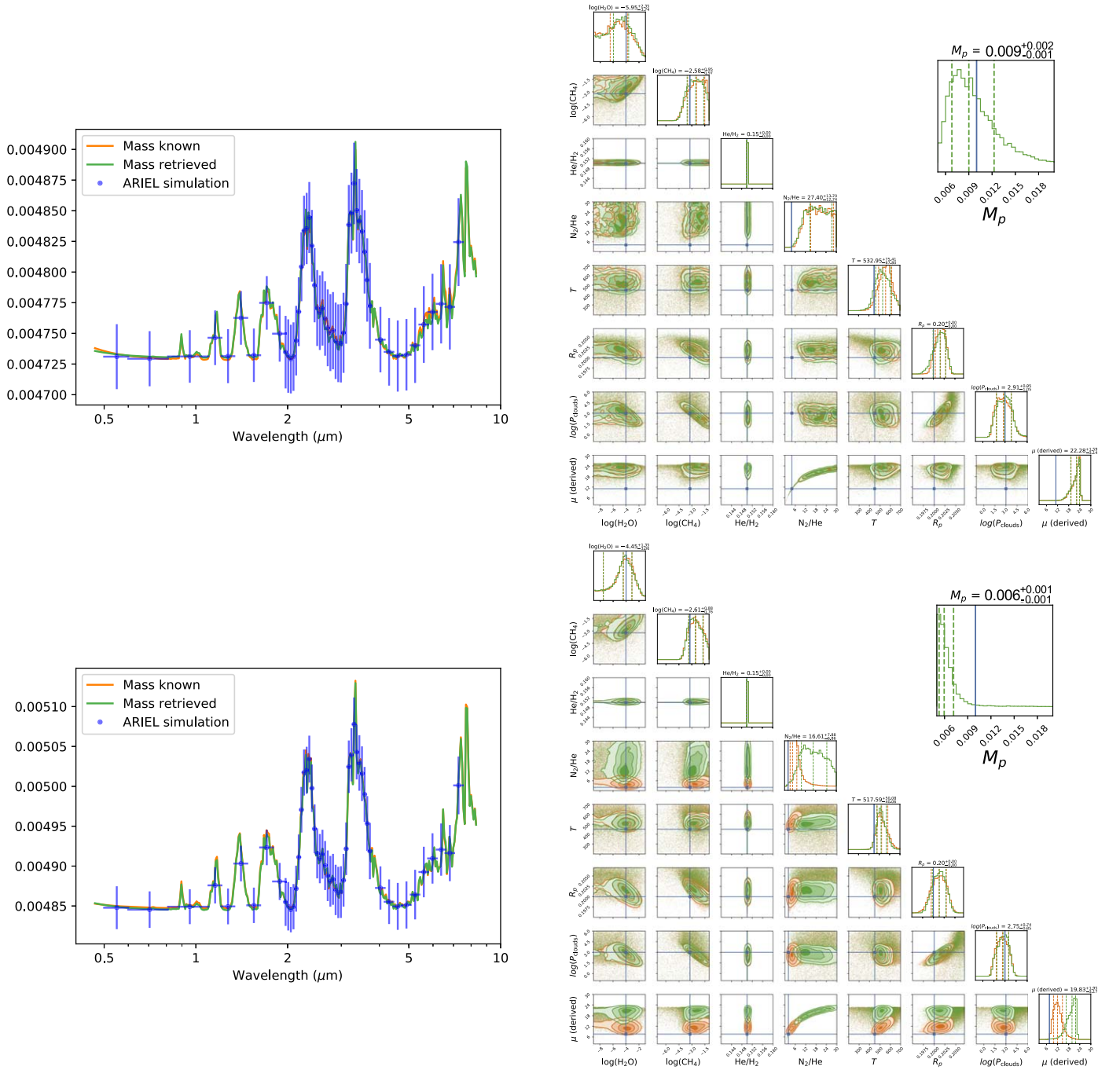
Table 1 summarizes the cases investigated in our study, showing the mass uncertainties in percent and the parameters affected if we ignore the planetary mass.

For clear-sky, gaseous atmospheres we find the same posterior distributions when the mass is known or retrieved. The retrieved mass is very accurate, with a precision of more than 10%, provided the wavelength coverage and S/N are adequate.

When opaque clouds are included in the simulations, the uncertainties in the retrieved radius and mass increase, especially for high altitude clouds. The error in the retrieved mass is up to 60% for our worst case scenario, i.e., a cloud pressure at  $10^{-3}$  bar. Additionally, we find that the posterior distributions of the retrieved radius and mass are no longer centered around their true values, indicating that solutions with different masses, radii and cloud parameters present similar likelihood. By contrast, atmospheric parameters such as the temperature and trace gas abundances appear to be unaffected by the knowledge of—or lack of—the mass.

Secondary atmospheres are more challenging due to the higher degree of freedom for the atmospheric main component. For broad wavelength ranges and adequate S/N observations,





**Figure 11.** ARIEL simulated spectra (left) and posteriors distribution (right) for a planet with a cloudy secondary atmosphere. The top cloud pressure is  $10^{-2}$  bar. Top:  $\mu = 11.1$ , bottom:  $\mu = 7.6$ . Orange plots: the mass is known. Green plots: the mass is retrieved. Blue crosses: simulated ARIEL observations obtained in one transit (left) and ground truth values (right).

**Table 1**  
Summary of the Mass Uncertainties and the Degeneracies from Our Retrieval Simulations

Type of Planet	HJ or SN, Clear-sky	HJ, High Opaque Clouds	SE, Clear-sky	SE, High Opaque Clouds	HJ HST
Mass uncertainty	<10%	10%–60%	<10% <sup>a</sup>	degenerate	170%
Temperature	no	no	no	no	no
Trace composition	no	no	no	no	no
Main composition	no	no	yes	yes	no
Radius	no	yes	no	yes	yes
Clouds	NA	yes	NA	yes	yes

**Notes.** HJ—hot Jupiter; SN—sub-Neptune with H/He-rich atmosphere; SE—super-Earth with secondary atmosphere; yes—affected by the knowledge of the mass; no—unaffected by the knowledge of the mass; NA—not applicable.

<sup>a</sup> With adequate S/N and wavelength coverage.

the mass can still be retrieved accurately and precisely if clouds are not present, and so can all the other parameters. We confirm the results in Batalha et al. (2017) concerning secondary atmospheres dominated by multiple species, for which a degeneracy may exist. Here, prior information about the mass may help to extract the main constituent ratios. However, we also show that it is possible to retrieve the mass of full H<sub>2</sub> and full N<sub>2</sub> planets down to an accuracy of 10% when the S/N is sufficient. This confirms the results from de Wit & Seager (2013).

When clouds are added, we find that the mass uncertainties may impact substantially the retrieval of the mean molecular weight: an independent characterization of the mass would therefore be helpful to capture/confirm the main constituents.

In the context of large scale surveys (ARIEL) and dedicated studies (JWST) of exoplanetary atmospheres, our results indicate that constraining the planetary mass for secondary atmospheres is important to ensure that we fully exploit the information content of the spectra. Current mass estimates found in exoplanet databases, which are mainly coming from radial velocity follow-up confirmations, have typical error bars of the order of 10%. Such small uncertainties guarantee an excellent prior knowledge for the mass in retrieval simulations, even for overcast planets.

Planets smaller than Neptune have larger mass errors, often larger than 50%. This uncertainty may contribute to the degeneracy in retrieving the mean molecular weight of the atmosphere, especially when clouds are present. Radial velocity campaigns should therefore prioritize the mass characterization of low-gravity planets as in the other cases, transit spectroscopy retrievals appear to be sufficiently robust to mass uncertainties.

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 758892, ExoAI; No. 776403, Exoplanets A) and under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ ERC grant agreement numbers 617119 (Exo-Lights). Furthermore, we acknowledge funding by the Science and Technology Funding Council (STFC) grants: ST/K502406/1, ST/P000282/1, ST/P002153/1, and ST/S002634/1. We also thank the referee for relevant comments that have greatly improved the quality of this manuscript.

## Appendix A

### Star–Planet Parameters Used for This Study

The planet and star parameters used to compute the models in Section 3 are shown in Table A1.

**Table A1**  
Planet and Star Parameters Used in Our Retrieval Simulations

Parameters	Hot Jupiter	Super-Earth
$R_s (R_{\text{sun}})$	1.19	0.3
$T_s (K)$	6091	3671
$M_s (M_{\text{sun}})$	1.23	0.4
Distance (pc)	48	2.6
$R_p (R_{\text{Jupiter}})$	1.39	0.2
$M_p (M_{\text{Jupiter}})$	0.73	0.01
$T_p (K)$	1450	450

## Appendix B

### Derivation of Transit Equation

The transit geometry and relevant variables are illustrated in Figure B1.

The normalized differential flux  $\Delta$  between in-transit  $F_{\text{in}}$  and out-transit  $F_{\text{out}}$  can be calculated as:

$$\Delta = \frac{F_{\text{out}} - F_{\text{in}}}{F_{\text{out}}} = \frac{R_p(\lambda)^2}{R_s^2}, \quad (\text{B1})$$

where  $R_p(\lambda)$  is the wavelength-dependent radius, which includes the atmospheric contribution, and  $R_s$  is the stellar radius.

The wavelength-dependent contribution of the atmosphere starts at  $R_0$ , we have:

$$\begin{aligned} \pi R_p(\lambda)^2 &= C_{\text{surf}} + C_{\text{atm}}(\lambda) = 2\pi \int_0^{R_0} r dr \\ &+ 2\pi \int_{R_0}^{\infty} r(1 - e^{-\tau(r,\lambda)}) dr, \end{aligned} \quad (\text{B2})$$

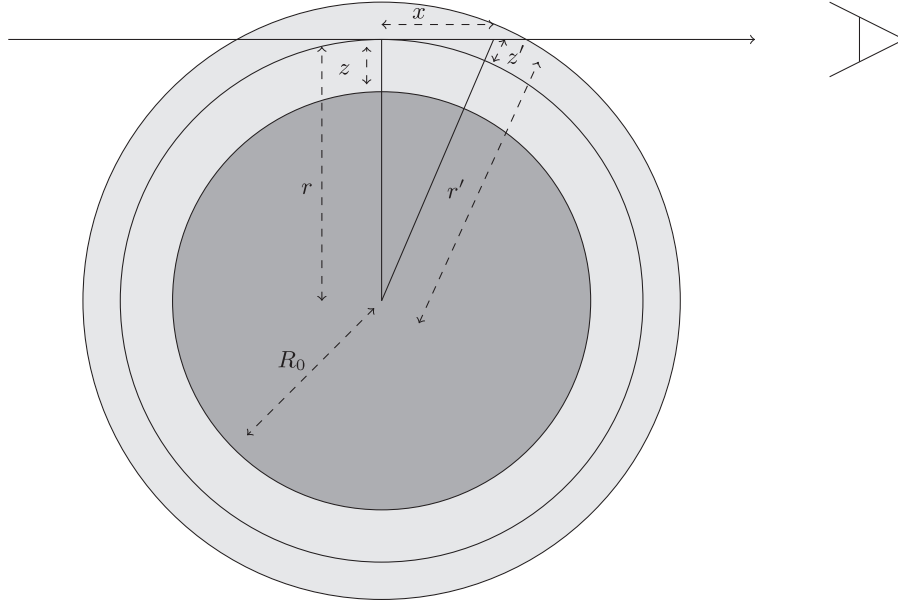
where we have introduced the optical depth  $\tau(r, \lambda)$ .

The optical depth represents the atmospheric absorption at a given altitude integrated along the line of sight:

$$\tau(r, \lambda) = 2 \int_0^{x_f} \sum_i n_i(r') \sigma_i(r', \lambda) dx, \quad (\text{B3})$$

where  $n_i$  is the number density of the  $i$ th species and  $\sigma_i$  is the cross section of the  $i$ th species.  $x_f$  is the maximum distance in the atmospheric layer along the line of sight.

For the rest of this derivation, we consider standard assumptions in current retrieval models published in the literature (Irwin et al. 2008; Madhusudhan & Seager 2009; Line et al. 2013; Benneke 2015; Waldmann et al. 2015a, 2015b; Harrington 2016; Lavie et al. 2017; MacDonald & Madhusudhan 2017; Cubillos 2018; Gandhi & Madhusudhan 2018; Al-Refaie et al. 2019; Mollière et al. 2019; Ormel & Min 2019; Zhang et al. 2019; Kitzmann et al. 2020). We assume



**Figure B1.** Illustration of the transmission of the stellar radiation through an exoplanet atmosphere during a transit event.  $R_0$  is the radius at which the planet becomes fully opaque in absence of clouds. For a given point in the atmosphere,  $z$  is the altitude normal to the Sun-observer connecting line and  $x$  is the projected distance from that normal to the point.  $r'$  is the distance from the point to the planetary center. In addition, we define  $r = R_0 + z$  and  $z' = r' - r$ .

the atmosphere is isothermal and in hydrostatic equilibrium. The scale height  $H$  is therefore defined as:

$$H = \frac{k_b T (R_0 + z)^2}{\mu M_p G}, \quad (\text{B4})$$

where  $k_b$  is the Boltzmann constant,  $T$  is the temperature,  $\mu$  is the mean molecular weight, and  $G$  is the gravitational constant. One can write the number density as:

$$n_i(r') = n_i(R_0) e^{-\frac{r'-R_0}{H}} = n_{0i} e^{-\frac{z+z'}{H}}, \quad (\text{B5})$$

For simplicity we write  $n_i(R_0) = n_{0i}$ .

Using Pythagoras' theorem and neglecting second-order terms of  $z'$ :

$$(R_0 + z)^2 + x^2 = (R_0 + z + z')^2, \quad (\text{B6})$$

$$z' = \frac{x^2}{2(R_0 + z)}, \quad (\text{B7})$$

we can now rewrite  $\tau$  as:

$$\tau(z, \lambda) = 2 \int_0^{x_f} \sum_i n_{0i} e^{-\frac{z}{H}} e^{-\frac{x^2}{2(R_0+z)H}} \sigma_i(p, T, \lambda) dx. \quad (\text{B8})$$

Finally, the contribution of the entire atmosphere can be estimated as:

$$C_{\text{atm}}(\lambda) = 2\pi \int_{R_0}^{\infty} (R_0 + z) \left( 1 - \exp \left[ -2 \int_0^{x_f} \sum_i n_{0i} e^{-\frac{z}{H}} e^{-\frac{x^2}{2(R_0+z)H}} \sigma_i(p, T, \lambda) dx \right] \right) dz. \quad (\text{B9})$$

The planetary mass appears only in the exponent, influencing solely the optical depth  $\tau$ . We use here the same assumptions made in the retrieval analysis to estimate the cross sections  $\sigma_i$ . We note that  $\sigma_i$  depend only on temperature and pressure:

$\sigma_i(r', \lambda) = \sigma_i(p(r'), T(r'), \lambda)$ . Considering  $\sigma_i(p, T, \lambda)$ , we approximate their values by using a linear interpolation with respect to  $p$  and  $T$ . We assume the temperature of the planet is isothermal and will not change dramatically with variation of the mass. Therefore, the atmospheric temperature can be interpolated between two known values from cross-section tables (for example: ExoMol, ExoTransmit Kempton et al. 2017, HITEMP or HITRAN). The pressure, however, needs to be integrated along the line of sight and we expect large variations: we therefore use a list of values labeled  $p_j$  and interpolate  $\sigma_i$  between two of these values. We have:

$$\sigma_i(T) = \sigma_i(T_1) + \frac{\sigma_i(T_2) - \sigma_i(T_1)}{T_2 - T_1} (T - T_1), \quad (\text{B10})$$

$$\sigma_i(p) = \sigma_i(p_j) + \frac{\sigma_i(p_{j+1}) - \sigma_i(p_j)}{p_{j+1} - p_j} (p - p_j), \quad (\text{B11})$$

where  $T_1, T_2, p_j$ , and  $p_{j+1}$  are fixed temperatures and pressures. Since the pressure differences across the  $x$ -axis are large—larger than the interpolation intervals—we sum over intervals  $(x_j, x_{j+1})$  of known pressures  $(p_j, p_{j+1})$ . We define:

$$K_{ij}(T) = \frac{\sigma_i(p_{j+1}, T, \lambda) - \sigma_i(p_j, T, \lambda)}{p_{j+1} - p_j}, \quad (\text{B12})$$

and estimate  $\sigma_i|_{x_j}^{x_{j+1}}$  in the interval  $(x_j, x_{j+1})$ :

$$\sigma_i(p, T, \lambda)|_{x_j}^{x_{j+1}} = \sigma_i(p_j, \lambda) + K_{ij}(T, \lambda)(p - p_j). \quad (\text{B13})$$

We include Equation (B13) in the expression of  $\tau$ , we get:

$$\tau(z, \lambda) = \sum_j \int_{x_j}^{x_{j+1}} \sum_i n_{0i} e^{-\frac{z}{H}} e^{-\frac{x^2}{2(R_0+z)H}} (\sigma_i(p_j, T, \lambda) - K_{ij}(T, \lambda)p_j + K_{ij}(T, \lambda)p) dx. \quad (\text{B14})$$

Knowing that the atmospheric pressure as a function of  $x$ ,  $z$  is  $p = p_0 e^{-\frac{x}{H}} e^{-\frac{z^2}{2(R_0+z)H}}$ , we can now calculate the integral:

$$\int_{x_j}^{x_{j+1}} e^{-\frac{x^2}{L}} = \frac{1}{2} \sqrt{\pi L} \left[ \operatorname{erf}\left(\frac{x}{\sqrt{L}}\right) \right]_{x_j}^{x_{j+1}}, \quad (\text{B15})$$

where  $L$  is a normalization constant. Concerning the integration boundaries, we have  $x_j = x(p_j)$ , which translates into:

$$x_j = \sqrt{-2(R_0 + z)H \left[ \ln\left(\frac{p_j}{p_0}\right) + \frac{z}{H} \right]}. \quad (\text{B16})$$

Defining for convenience

$$I_j(z) = \operatorname{erf}\left(\sqrt{-\left(\ln\left(\frac{p_{j+1}}{p_0}\right) + \frac{z}{H}\right)}\right) - \operatorname{erf}\left(\sqrt{-\left(\ln\left(\frac{p_j}{p_0}\right) + \frac{z}{H}\right)}\right), \quad (\text{B17})$$

and

$$I_j'(z) = \operatorname{erf}\left(\sqrt{-2\left(\ln\left(\frac{p_{j+1}}{p_0}\right) + \frac{z}{H}\right)}\right) - \operatorname{erf}\left(\sqrt{-2\left(\ln\left(\frac{p_j}{p_0}\right) + \frac{z}{H}\right)}\right), \quad (\text{B18})$$

we get

$$\int_{x_j}^{x_{j+1}} e^{-\frac{x^2}{2(R_0+z)H}} = \frac{1}{2} \sqrt{2\pi(R_0+z)H} I_j(z), \quad (\text{B19})$$

and

$$\int_{x_j}^{x_{j+1}} e^{-\frac{x^2}{(R_0+z)H}} = \frac{1}{2} \sqrt{\pi(R_0+z)H} I_j'(z). \quad (\text{B20})$$

This result leads to:

$$\tau(z, \lambda) = \sum_j \sum_i n_{0i} e^{-\frac{z}{H}} \sqrt{\pi(R_0+z)H} (\sigma_i(p_j, T) - K_{ij}^p) \sqrt{2} I_j + K_{ij}^x e^{-\frac{z}{H}} p_0 I_j'. \quad (\text{B21})$$

Now the temperature dependence of  $\sigma$  can be added in the same way. Here we assume the atmosphere to be isothermal, so we do not have to consider the temperature outside our reference  $\sigma(T_2)$  and  $\sigma(T_1)$ . This leads to:

$$\begin{aligned} \tau &= \sum_j \sum_i n_{0i} e^{-\frac{z}{H}} \sqrt{\pi(R_0+z)H} (\sqrt{2} I_j \sigma_i(p_j, T) \\ &+ K_{ij}^T \sqrt{2} I_j (T - T_1) + (K_{ij}^p + K_{ij}^x (T - T_1)) (p_0 I_j' e^{-\frac{z}{H}} \\ &- \sqrt{2} I_j p_j), \end{aligned} \quad (\text{B22})$$

where the coefficients  $K_{ij}^p$ ,  $K_{ij}^T$ , and  $K_{ij}^x$  are only wavelength dependent and can be calculated from tables:

$$K_{ij}^p = \frac{\sigma_i(p_{j+1}, T_1) - \sigma_i(p_j, T_1)}{p_{j+1} - p_j}, \quad (\text{B23})$$

$$K_{ij}^T = \frac{\sigma_i(p_j, T_2) - \sigma_i(p_j, T_1)}{T_2 - T_1}, \quad (\text{B24})$$

$$\begin{aligned} K_{ij}^x &= \frac{1}{p_{j+1} - p_j} \frac{1}{T_2 - T_1} (\sigma_i(p_{j+1}, T_2) - \sigma_i(p_{j+1}, T_1) \\ &- \sigma_i(p_j, T_2) + \sigma_i(p_j, T_1)). \end{aligned} \quad (\text{B25})$$

We replace:

$$H = \frac{k_b T (R_0 + z)^2}{\mu M_p G}. \quad (\text{B26})$$

By considering the cross sections constant with pressure and temperature,  $p_{j+1} = 0$  and  $p_j = p_0 e^{-\frac{z}{H}}$  the equations for  $I_j = 1$  and  $\tau$  are simplified and we finally get:

$$\tau = \sum_i n_{0i} \sigma_i(p_0, T_0) e^{-\frac{z}{H}} \sqrt{2\pi(R_0+z)H}. \quad (\text{B27})$$

We can investigate the contribution of gray clouds by separating the atmospheric terms below and above clouds in Equation (B2), we obtain:

$$\begin{aligned} \pi R_p(\lambda)^2 &= C_{\text{surf}} + C_{\text{clouds}} + C_{\text{atm}}(\lambda), = 2\pi \int_0^{R_0} r dr \\ &+ 2\pi \int_0^{z_c} (R_0 + z) dz + 2\pi \int_{z_c}^{\infty} (R_0 + z) \\ &\times (1 - e^{-\tau(z, \lambda)}) dz, \end{aligned} \quad (\text{B28})$$

where  $\tau(z, \lambda)$  is given in Equation (B22) and:

$$z_c = -H \ln\left(\frac{P_c}{P_s}\right). \quad (\text{B29})$$

$P_c$  is the cloud-top pressure and  $P_s$  is the pressure at the reference radius  $R_0$  (10 bar in this paper).

## ORCID iDs

Q. Changeat  <https://orcid.org/0000-0001-6516-4493>  
L. Keyte  <https://orcid.org/0000-0001-5849-577X>  
I. P. Waldmann  <https://orcid.org/0000-0002-4205-5267>  
G. Tinetti  <https://orcid.org/0000-0001-6058-6654>

## References

- Al-Refaie, A. F., Changeat, Q., Waldmann, I. P., & Tinetti, G. 2019, arXiv:1912.07759
- Barstow, J. K., Aigrain, S., Irwin, P. G. J., & Sing, D. K. 2017, *ApJ*, 834, 50
- Barton, E. J., Hill, C., Yurchenko, S. N., et al. 2017, *JQSRT*, 187, 453
- Batalha, N. E., Kempton, E. M. R., & Mbarek, R. 2017, *ApJL*, 836, L5
- Bean, J. L., Stevenson, K. B., Batalha, N. M., et al. 2018, *PASP*, 130, 114402
- Benneke, B. 2015, arXiv:1504.07655
- Borsato, L., Malavolta, L., Piotto, G., et al. 2019, *MNRAS*, 484, 3233
- Brown, T. M. 2001, *ApJ*, 553, 1006
- Cubillos 2018, PyratBay retrieval code, <https://pcubillos.github.io/pyratbay/index.html>
- de Wit, J., & Seager, S. 2013, *Sci*, 342, 1473
- Edwards, B., Mugnai, L., Tinetti, G., Pascale, E., & Sarkar, S. 2019, *AJ*, 157, 242
- Edwards, B., Rice, M., Zingales, T., et al. 2018, *ExA*, 47, 29
- Fisher, C., & Heng, K. 2018, *MNRAS*, 481, 4698
- Fortney, J. J. 2005, *MNRAS*, 364, 649
- Gandhi, S., & Madhusudhan, N. 2018, *MNRAS*, 474, 271
- Gordon, I., Rothman, L. S., Wilzewski, J. S., et al. 2016, AAS/DPS Meeting, 48, 421.13
- Griffith, C. A. 2014, *RSPTA*, 372, 20130086
- Harrington, J. 2016, Atmospheric Retrievals from Exoplanet Observations and Simulations with BART, NASA Proposal, 16-XPR16-10
- Heng, K., & Kitzmann, D. 2017, *MNRAS*, 470, 2972

- Heng, K., Wyttenbach, A., Lavie, B., et al. 2015, *ApJL*, 803, L9
- Hill, C., Yurchenko, S. N., & Tennyson, J. 2013, *Icar*, 226, 1673
- Irwin, P. G. J., Teanby, N. A., de Kok, R., et al. 2008, *JQSRT*, 109, 1136
- Kempton, E. M.-R., Lupu, R., Owusu-Asare, A., Slough, P., & Cale, B. 2017, *PASP*, 129, 044402
- Kitzmann, D., Heng, K., Oreshenko, M., et al. 2020, *ApJ*, 890, 174
- Lavie, B., Mendonça, J. M., Mordasini, C., et al. 2017, *AJ*, 154, 91
- Lecavelier des Etangs, A., Pont, F., Vidal-Madjar, A., & Sing, D. 2008, *A&A*, 481, L83
- Lee, J.-M., Heng, K., & Irwin, P. G. J. 2013, *ApJ*, 778, 97
- Line, M. R., & Parmentier, V. 2016, *ApJ*, 820, 78
- Line, M. R., Wolf, A. S., Zhang, X., et al. 2013, *ApJ*, 775, 137
- Line, M. R., Zhang, X., Vasisht, G., et al. 2012, *ApJ*, 749, 93
- López-Morales, M., Haywood, R. D., Coughlin, J. L., et al. 2016, *AJ*, 152, 204
- MacDonald, R. J., & Madhusudhan, N. 2017, *MNRAS*, 469, 1979
- Madhusudhan, N., & Seager, S. 2009, *ApJ*, 707, 24
- Mollière, P., Wardenier, J. P., van Boekel, R., et al. 2019, *A&A*, 627, A67
- Ormel, C. W., & Min, M. 2019, *A&A*, 622, A121
- Pinhas, A., Madhusudhan, N., Gandhi, S., & MacDonald, R. 2019, *MNRAS*, 482, 1485
- Rocchetto, M., Waldmann, I. P., Venot, O., Lagage, P. O., & Tinetti, G. 2016, *ApJ*, 833, 120
- Rothman, L. S., & Gordon, I. E. 2014, in 13th Biennial HITRAN Conference (Cambridge, MA: Harvard-Smithsonian Center for Astrophysics), 49
- Stassun, K. G., Collins, K. A., & Gaudi, B. S. 2017, *AJ*, 153, 136
- Tennyson, J., Yurchenko, S. N., Al-Refaie, A. F., et al. 2016, *JMoSp*, 327, 73
- Tinetti, G., Drossart, P., Eccleston, P., et al. 2018, *ExA*, 46, 135
- Tsiaras, A., Waldmann, I. P., Zingales, T., et al. 2018, *AJ*, 155, 156
- Waldmann, I. P., Rocchetto, M., Tinetti, G., et al. 2015a, *ApJ*, 813, 13
- Waldmann, I. P., Tinetti, G., Rocchetto, M., et al. 2015b, *ApJ*, 802, 107
- Welbanks, L., & Madhusudhan, N. 2019, *AJ*, 157, 206
- Zhang, M., Chachan, Y., Kempton, E. M. R., & Knutson, H. A. 2019, *PASP*, 131, 034501