The theory of transmission spectra revisited: a semi-analytical method for interpreting WFC3 data and an unresolved challenge

Kevin Heng[★] and Daniel Kitzmann

Center for Space and Habitability, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

ABSTRACT

The computation of transmission spectra is a central ingredient in the study of exoplanetary atmospheres. First, we revisit the theory of transmission spectra, unifying ideas from several workers in the literature. Transmission spectra lack an absolute normalization due to the a priori unknown value of a reference transit radius, which is tied to an unknown reference pressure. We show that there is a degeneracy between the uncertainty in the transit radius, the assumed value of the reference pressure (typically set to 10 bar) and the inferred value of the water abundance when interpreting a Wide-Field Camera 3 (WFC3) transmission spectrum. Secondly, we show that the transmission spectra of isothermal atmospheres are nearly isobaric. We validate the isothermal, isobaric analytical formula for the transmission spectrum against full numerical calculations and show that the typical errors are $\sim$0.1 per cent ($\sim$10 ppm) within the WFC3 range of wavelengths for temperatures of 1500 K (or higher). Thirdly, we generalize the previous expression for the transit radius to include a small temperature gradient. Finally, we analyse the measured WFC3 transmission spectrum of WASP-12b and demonstrate that we obtain consistent results with the retrieval approach of Kreidberg et al., if the reference transit radius and reference pressure are fixed to assumed values. The unknown functional relationship between the reference transit radius and reference pressure implies that it is the product of the water abundance and reference pressure that is being retrieved from the data, and not just the water abundance alone. This degeneracy leads to a limitation on how accurately we may extract molecular abundances from transmission spectra using WFC3 data alone. We suggest an approximate expression for this relationship. Finally, we compare our study to that of Griffith and discuss why the degeneracy was missed in previous retrieval studies.

Key words: radiative transfer – planets and satellites: atmospheres.

1 INTRODUCTION

A substantial fraction of the measurements made of exoplanetary atmospheres takes the form of transmission spectra – scrutinizing the small change in the projected size of the exoplanet, across wavelength, as it transits its host star (Seager & Sasselov 2000). To interpret a transmission spectrum using atmospheric retrieval requires that we are able to solve the inverse problem robustly and efficiently: Given the wavelength-dependent transit radius, $R(\lambda)$, we wish to infer the types and abundances of atoms and molecules that contribute to the opacity of the atmosphere. Across a transmission spectrum, the relative size of spectral features is proportional to the pressure scale height, which depends on temperature, surface gravity and mean molecular mass.

Traditionally, one calculates the transmission spectrum by tracing a set of rays through the limb of the atmosphere (Brown 2001; Hubbard et al. 2001; Burrows, Sudarsky & Hubbard 2003). Each ray passes through a transit chord with varying degrees of transparency or opaqueness. By summing up the contributions from all of these chords, one may calculate the effective occulting area of the exoplanet at a given wavelength, πR^2. Several studies describe the theory of transmission spectra (Seager & Sasselov 2000; Brown 2001; Hubbard et al. 2001; Burrows et al. 2003; Fortney 2005; Benneke & Seager 2012; de Wit & Seager 2013; Griffith 2014; Vahidinia et al. 2014; Heng et al. 2015; Bétrémieux & Swain 2017), but a unified analytical approach is missing from the literature. Part of the motivation of this study is to present such a unified approach that thoroughly explores the assumptions, caveats and degeneracies associated with the calculation of transmission spectra. For this study, we ignore the effects of refraction (Seager & Sasselov 2000; Brown 2001; Hubbard et al. 2001; Bétrémieux 2016) and multiple scattering (Brown 2001; Robinson 2017).
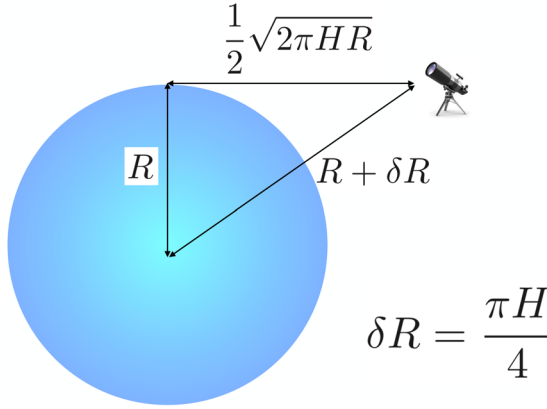
★ E-mail: kevin.heng@csh.unibe.ch

Figure 1. Schematic illustrating the transit geometry for an isothermal atmosphere, which has an effective chord length of $\sqrt{2\pi HR}$, corresponding to the transit radius (Fortney 2005). This works out to $\delta R = \pi H/4$, which is a change in pressure of a factor of $\exp(\pi/4) \approx 2.19$ across the transit chord.

Transmission spectra have previously been shown to be insensitive to temperature variations (Brown 2001; Hubbard et al. 2001). (In this study, we will show that $R \propto \kappa^{\pm 1/b} \ln \kappa$ in the presence of a temperature gradient, where one may reasonably expect $b \gg 1$. The opacity, κ, depends on the temperature.) Retrieval approaches for transmission spectra have assumed isothermal temperature–pressure profiles – either explicitly (Waldmann et al. 2015) or implicitly (Benneke & Seager 2012).[1] For an isothermal atmosphere, it has previously been shown that the transit chord has a length of $\sqrt{2\pi HR}$, with H being the pressure scaleheight (Fortney 2005). The transit geometry is such that $\sqrt{\pi HR/2}$, R and $R + \delta R$ form the three sides of a triangle, from which we may calculate that $\delta R = \pi H/4$ (Fig. 1). This in turn corresponds to a factor of $\exp(\pi/4) \approx 2.19$ for the change in pressure across the transit chord. Thus, isothermal atmospheres are *nearly* isobaric, which leads us to the reasoning that an analytical formula for an isothermal, isobaric atmosphere should adequately serve as an accurate fitting function to data. Furthermore, transmission spectra should be insensitive to the pressure being probed as long as the temperature is high enough and pressure broadening is relatively unimportant at the altitudes being sensed.

The noise floor for the Wide-Field Camera 3 (WFC3) on the *Hubble Space Telescope* (*HST*) is typically 50 ppm and does not reach 20 ppm, and the instruments on the *James Webb Space Telescope* (*JWST*) are expected to have similar noise floors (Greene et al. 2016) due to the shared heritage between the infrared detectors (Beichman et al. 2014). Thus, an isothermal, isobaric formula is deemed to be accurate enough if the average error associated with it is better than 20 ppm. Certainly, the true noise floor for transmission spectra obtained using *JWST* awaits the procurement and analysis of real data.

For completeness, we also generalize the isothermal formula for the transit radius to include a (small) temperature gradient in the region of the atmosphere probed by transmission spectroscopy. Validated analytical formulae are invaluable for studying degeneracies in the model without expending much computational effort.

[1] Benneke & Seager (2012) used the analytical temperature–pressure profile of Guillot (2010), which is perfectly isothermal at high altitudes due to the underlying, simplifying assumptions made, i.e. that the absorption-, flux- and Planck-mean opacities are equal.
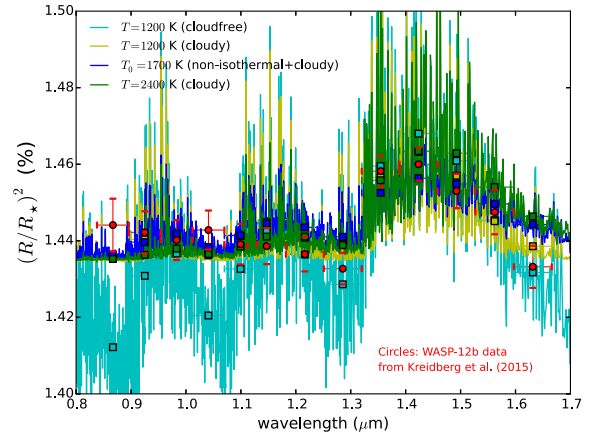

Figure 2. Examples of model transmission spectra with different temperatures, cloud opacities and a finite temperature gradient. The water volume mixing ratio is set to 10^{-3} for illustration. See Table 1 and text for the parameter values assumed for these models.

Table 1. Parameter values used for WASP-12b models of transmission spectra in Fig. 2

T (K)	T_0 (K)	T' (K km^{-1})	R_0 (R_J)	$\kappa_{\rm cloud}$ (cm^2 g^{-1})
1200	–	–	1.790	0
1200	–	–	1.786	2×10^{-3}
–	1700	0.1	1.775	2×10^{-3}
2400	–	–	1.711	4×10^{-2}

Note. Following Kreidberg et al. (2015), we assume $R_\star = 1.57 R_\odot$ and $P_0 = 10$ bar. For illustration, we assume a negative temperature gradient in our $T_0 = 1700$ K example. Our water opacities are computed at a pressure of 1 mbar.

An under-emphasized fact is that the normalization of a transmission spectrum depends on specifying a reference transit radius that is associated with a reference pressure. Higher temperatures or a finite temperature gradient may be compensated by lower values of the normalization. The value of the reference transit radius may be matched[2] to the measured white-light radius, but this measurement is associated with an uncertainty. Furthermore, the functional relationship between the reference transit radius and reference pressure is unknown. The reference pressure cannot be extracted from the data and must be assumed. It is typically set to 10 bar (e.g. Line et al. 2013; Kreidberg et al. 2015). A goal of this study is to show that there is a degeneracy between the reference transit radius, reference pressure and water abundance that cannot easily be overcome.

In Fig. 2, we show examples of model transmission spectra specialized to the wavelength range of WFC3 onboard the *HST*, and over-plot the data of WASP-12b measured by Kreidberg et al. (2015). (The parameter values used are listed in Table 1.) It is apparent that the temperature, temperature gradient and degree of cloudiness are somewhat degenerate quantities that produce similar model transmission spectra. The spectral features just bluewards and redwards of the 1.4-µm water feature break these degeneracies partially and allow for a unique fit to be obtained *if the reference transit radius and reference pressure are fixed*. Another goal of this study is to introduce a fast method for performing this fit to data

[2] We emphasize that this is the *choice* of the modeller. Theoretically, any transit radius may be chosen as a reference.

using a validated analytical formula, while pointing out that *our ig-norance of the relationship between the reference transit radius and reference pressure plagues our ability to extract accurate values of the water abundance.* We then suggest an approximate expression to relate the reference transit radius and reference pressure.

In Section 2, we lay down the formalism for computing transmission spectra for both isothermal and non-isothermal atmospheres. In Section 3, we benchmark our isothermal, isobaric analytical formula and report on its accuracy. We apply our formula to the case study of WASP-12b and discuss the degeneracies involved in data interpretation. In Section 4, we discuss the implications of our results.

2 FORMALISM

2.1 Order-of-magnitude expressions

Consider a toy atmosphere with a constant number density n, containing a single molecule with an extinction cross-section σ. A single temperature T describes this atmosphere. The pressure scale-height is $H = kT/mg$ with k being the Boltzmann constant, m being the mass of the molecule and g being the surface gravity. Since we expect $H \ll R$, the characteristic lengthscale in the system is the geometric mean of H and R, which is $\sqrt{HR}$. The chord optical depth associated with the transit radius is

$$\tau \sim n\sigma \sqrt{HR} \sim 1. \tag{1}$$

If we re-arrange this expression and invoke the ideal gas law, we obtain an expression for the transit radius in terms of the opacity (κ; cross-section per unit mass),

$$R \sim H \left(\frac{g}{P\kappa} \right)^2. \tag{2}$$

Consider water to be the only opacity source in the atmosphere (since the WFC3 bandpass probes mostly water). This expression already reveals the presence of a degeneracy between the pressure, temperature and abundance of water. The central limitation is that there is one equation and three unknowns.

Setting $\tau \sim 1$ is reasonable but imprecise because one needs to formally integrate over a range of chords with various values of τ.

2.2 Isothermal atmospheres

We revisit the formalism for isothermal chord optical depths and transit radii, first explored by Fortney (2005) and later expanded upon by de Wit & Seager (2013), Heng et al. (2015) and Bétrémieux & Swain (2017). Our intention is to clarify several aspects of the derivation and elucidate the assumptions made.

Consider that an observer records a transit radius of R'. A radial coordinate r' is defined such that $r' = 0$ is located exactly at R'. It follows that $r' \approx x^2/2R'$, where x is the spatial coordinate along the sightline of the observer (Fortney 2005). For an isothermal atmosphere, we have $n = n' \exp(-r'/H)$, such that n' is the number density probed at R'. By evaluating $\int_{-\infty}^{+\infty} n\sigma \, dx$ and assuming that σ may be taken out of the integral, we obtain $\tau = n'\sigma\sqrt{2\pi HR'}$ (Fortney 2005). An order-of-magnitude estimate of τ misses the $\sqrt{2\pi}$ factor. It is important to note that the characteristic lengthscale, $\sqrt{2\pi HR'}$, only appears when the integration is carried out formally from $-\infty$ to $+\infty$.

Now, we define a radial coordinate r such that $r = 0$ is located at the centre of the exoplanet. We re-scale r' such that $r' = 0$ sits at $r = R_0$. Let the number density associated with the reference transit

radius (R_0) or pressure (P_0) be n_0. It follows that $n' = n_0 \exp[-(r - R_0)/H]$. Since $R' = R_0 + r' \approx R_0$, the chord optical depth is, to a good approximation,

$$\tau = \tau_0 \exp\left(-\frac{r - R_0}{H} \right), \tag{3}$$

which agrees with equation (S.1) of de Wit & Seager (2013) and equation (15) of Bétrémieux & Swain (2017). The reference optical depth is

$$\tau_0 = \frac{P_0 \sigma}{kT} \sqrt{2\pi HR_0}. \tag{4}$$

These considerations yield a relation between τ and r,

$$dr = -\frac{H \, d\tau}{\tau}. \tag{5}$$

Let the effective thickness of the atmosphere, at a given wavelength, be h. The transit radius is then $R = R_0 + h$. The projected size of the exoplanet is πR^2, and it may also be expressed as $\pi R_0^2 + A$, with A being the area of the annulus sitting above the reference radius (Brown 2001; de Wit & Seager 2013),

$$A = 2\pi \int_{R_0}^{+\infty} \left[1 - \exp(-\tau) \right] r \, dr. \tag{6}$$

Note that the integration is formally carried out to $+\infty$, even though the transit radii is effectively located at a finite distance from the centre of the exoplanet. Since we generally expect $A/\pi R_0^2 \ll 1$, the effective thickness of the atmosphere is

$$h = \frac{A}{2\pi R_0}. \tag{7}$$

By performing a change of coordinate from r to τ, we obtain

$$h = H \int_0^{\tau_0} \left(\frac{1 - e^{-\tau}}{\tau} \right) \left[1 + \frac{H}{R_0} \ln\left(\frac{\tau_0}{\tau} \right) \right] d\tau. \tag{8}$$

As previously reasoned by Bétrémieux & Swain (2017), the term associated with H/R_0 is much smaller than the other term and may therefore be neglected, whereas de Wit & Seager (2013) kept both terms and showed that hypergeometric functions obtain from evaluating the smaller term. Following Bétrémieux & Swain (2017), we use the identity in equation (10) of appendix I of Chandrasekhar (1960),

$$E_1 = -\gamma - \ln \tau_0 + \int_0^{\tau_0} \frac{1 - e^{-\tau}}{\tau} \, d\tau, \tag{9}$$

where $\gamma \approx 0.57721$ is the Euler–Mascheroni constant. The quantity E_1 is the exponential integral of the first order with the argument τ_0. As $\tau_0 \to \infty$, we have $E_1 \to 0$. Applying the preceding identity to equation (8) yields

$$h = H (\gamma + \ln \tau_0 + E_1). \tag{10}$$

The extra E_1 term was not explicitly stated in equation (7) of de Wit & Seager (2013), but is implicitly present in their equation (S.4), contrary to the claim of Bétrémieux & Swain (2017).

Bétrémieux & Swain (2017) have previously interpreted τ_0 to be associated with an optically thick surface such as a cloud deck. In the current derivation, τ_0 is simply associated with a reference pressure corresponding to an atmospheric layer that may – or may not – be chosen to be optically thick. It is a natural outcome of the a priori unknown value of the pressure associated with the reference transit radius R_0. We generally expect $\tau_0 \gg 1$, which means that

the E_1 term vanishes and one ends up with an expression for the effective chord optical depth associated with the transit radius R,

$$\tau_{\text{eff}} = \tau_0 \exp\left(-\frac{h}{H}\right) = \exp(-\gamma) \approx 0.56. \tag{11}$$

As already remarked by de Wit & Seager (2013) and Bétrémieux & Swain (2017), this value of 0.56 is derived from first principles, whereas Lecavelier des Etangs et al. (2008) inferred it by using a graphical solution.

Finally, we state the expression for the transit radius assuming an isothermal atmosphere,

$$R = R_0 + H\left[\gamma + \ln\left(\frac{P_0 \kappa}{g}\sqrt{\frac{2\pi R_0}{H}}\right)\right]. \tag{12}$$

As expected, the transit radius depends linearly on the pressure scaleheight, but is a slowly varying function of the opacity (Brown 2001). The opacity is evaluated at the temperature T and a pressure P that is arbitrarily chosen such that pressure broadening is negligible. However, since the opacity typically varies over many orders of magnitude, its overall effect on the transmission spectrum is comparable to that of H.

The formula in equation (12) teaches us a few lessons about the degeneracies inherent in transmission spectra:

(i) The transit radius may be rewritten as $R = C + H \ln \kappa$, where R_0 and P_0 are absorbed into the constant C. This implies that there is a degeneracy between R_0 and P_0. This degeneracy was noticed numerically by Benneke & Seager (2012). In practice, R_0 may be matched to the measured white-light radius, but the value of P_0 is unknown and cannot be extracted from the data.

(ii) If R_0 is matched to the white-light radius, it means that it does not possess an exact value, but rather a range of values that is associated with the measured uncertainties in the white-light radius. However, since the functional relationship between R_0 and P_0 is unknown, it is not apparent how one should adjust the value of P_0 as R_0 is varied.

(iii) The opacity may generally be written as $\kappa = \chi \kappa_0$, where χ is the mass mixing ratio[3] (and not the volume mixing ratio[4]) and κ_0 is the opacity of a specific molecule (e.g. water). However, χ can be absorbed into the constant C, which informs us that there is a degeneracy between χ, R_0 and P_0. In particular, it is $P_0 \chi$, rather than χ, that one really infers from a fit to data.

(iv) The pressure scaleheight controls the shape of the transmission spectrum, but not its normalization. This implies that the temperature is degenerate with the normalization, which is controlled by R_0, P_0 and χ.

If we specialize to WFC3 transmission spectra, then these degeneracies inform us that the water abundance[5] cannot be uniquely inferred from the data, and depend strongly on the assumed values of the reference transit radius and reference pressure. Since the water mixing ratio appears in the logarithm and the reference transit radius does not, it follows that small variations in R_0 will lead to large variations in χ (Griffith 2014). We will demonstrate this point later during our analysis of WFC3 data of WASP-12b.

[3] Relative abundance by mass.

[4] Relative abundance by number.

[5] From this point on, we will use the terms 'abundance' and 'mixing ratio' interchangeably. The latter may refer to either the volume or mass mixing ratio, depending upon context.

2.3 Non-isothermal atmospheres

Transmission spectra probe the atmosphere at high altitudes (low pressures), where the temperature–pressure profile is approximately linear with distance (Fortney et al. 2008, 2010). Furthermore, since transmission spectra are sensitive to only a limited range in pressure (Fig. 1), one may approximate the temperature–pressure profile as a constant with higher-order corrections. For these reasons, we consider a non-isothermal model atmosphere with a temperature profile given by

$$T = T_0 \pm T'(r - R_0). \tag{13}$$

T' is the radial temperature gradient and is defined to be always a positive number. Positive and negative signs correspond to temperature profiles with positive and negative gradients, respectively. The quantity T_0 is the reference temperature corresponding to R_0 and P_0. We consider this non-isothermal correction to the temperature to be small. For example, we expect $T' \sim 0.1 \text{ K km}^{-1}$ (Huitson et al. 2012; Heng et al. 2015), $T_0 \sim 1000$ K and $H \sim 100$ km for hot Jupiters, which yields $T'H/T_0 \sim 10^{-2}$. We then take the simplification that the opacity or cross-section may be evaluated at $T = T_0$.

There is a need to distinguish between the isothermal pressure scaleheight, $H = kT_0/mg$ and the non-isothermal pressure scaleheight,

$$H' = \frac{T_0}{T'}. \tag{14}$$

The ratio of these two quantities is

$$b \equiv \frac{H'}{H} = \frac{mg}{kT'}. \tag{15}$$

Here, m is the mean molecular mass. For hydrogen-dominated atmospheres with $T' \sim 0.1 \text{ K km}^{-1}$, we have $b \approx 30$. For comparison, the Earth has $T' \sim 1 \text{ K km}^{-1}$.

By integrating the expression for hydrostatic balance, we obtain

$$n = n_0\left(1 \pm \frac{r - R_0}{H'}\right)^{\mp b - 1}, \tag{16}$$

where the positive and negative signs associated with the $(r - R_0)/H'$ term correspond to positive and negative temperature gradients, respectively. The chord optical depth is

$$\tau = \zeta \pi n \sigma \sqrt{\frac{2T_0 R_0}{T'}}. \tag{17}$$

The exact value of the dimensionless coefficient ζ depends on evaluating the integral (Heng et al. 2015),

$$\int_{-\pi/2}^{+\pi/2} (\cos x)^N \, dx = \zeta \pi, \tag{18}$$

where $N = 2b$, if the temperature profile has a positive gradient, and $N = 2b - 1$, if it has a negative gradient. In practice, the value of ζ ranges between 0.1 and 1, and only appears as $\ln \zeta$ in the expression for the transit radius, which suggests that it is sufficient to set it to $\zeta = 0.5$. To do better requires an iteration between the inferred value of b and a re-calculation of ζ until convergence attains.

Again, we need to evaluate the integral,

$$h = \frac{1}{R_0}\int_{R_0}^{+\infty}\left[1 - \exp(-\tau)\right] r \, dr. \tag{19}$$

We find that an analytical solution obtains only if $r\,\mathrm{d}r \propto \mathrm{d}\tau/\tau$, which occurs when we assume $b \gg 1$. This yields

$$r\,\mathrm{d}r \approx -\frac{H'R_0}{b\tau}\tau_0^{\pm 1/b}\,\mathrm{d}\tau. \qquad (20)$$

It follows that the transit radius for a non-isothermal atmosphere is

$$R = R_0 + H\tau_0^{\pm 1/b}\left(\gamma + \ln\tau_0 + E_1\right), \qquad (21)$$

where the reference optical depth is given by

$$\tau_0 = \frac{\zeta\pi P_0 \kappa}{g}\sqrt{\frac{2R_0 b}{H}}. \qquad (22)$$

As before, the E_1 term may be dropped if $\tau_0 \gg 1$ is assumed.

By comparing equation (21) to its isothermal counterpart in equation (12), we see that the formulae for the transit radius are very similar in structure. The reference transit radius and reference pressure are joined by b in the logarithm. The main effect of non-isothermal behaviour is to introduce the $\tau_0^{\pm 1/b}$ correction factor next to the isothermal pressure scaleheight. It accounts for an enhancement to the transit radius when the temperature gradient is positive, and a diminution when it is negative. Since this factor is degenerate with H, it implies that the strength of spectral features in transmission spectra is controlled by a combination of temperature, surface gravity and mean molecular mass (which make up the isothermal pressure scaleheight), as well as a finite temperature gradient. These quantities balance one another out and thus are degenerate with one another. It suggests that the isothermal formula in equation (12) should suffice for fitting data – a statement we will prove in the next section.

3 RESULTS

3.1 Benchmarking and validation

The results in this subsection justify the use of isothermal, isobaric model atmospheres, described by the simple analytical formula in equation (12), to analyse WFC3 transmission spectra for hot exoplanetary atmospheres.

3.1.1 Benchmarking to the isothermal, non-isobaric calculations of Deming, Fortney and Line

We first wish to demonstrate that we are able to perform full calculations of transmission spectra (with no approximations taken) correctly. In their fig. 5, Line et al. (2013) previously published a transmission spectrum of HD 209458b with $R_0 = 1.25\,R_{\mathrm{J}}$ (with R_{J} being the radius of Jupiter), $P_0 = 10$ bar, $R_\star = 1.148\,\mathrm{R}_\odot$ (with $\mathrm{R}_\odot$ being the solar radius) and $g = 10\,\mathrm{m\,s^{-2}}$. The model atmosphere is assumed to be isothermal with $T = 1500$ K and discretized into 90 layers between 0.1 nbar and 10 bar (equally spaced in the logarithm of pressure). The volume mixing ratios of molecular hydrogen, helium and water are assumed to be 0.85, 0.15 and 4.5×10^{-4}, respectively, which translate into $m = 2.3\,m_{\mathrm{amu}}$. The quantity $m_{\mathrm{amu}} = 1.660\,539\,040 \times 10^{-24}$ g is the atomic mass unit. In the same figure, Line et al. (2013) included the calculations of Fortney et al. (2010) and Deming et al. (2013), and demonstrated that all three calculations match well.

In Fig. 3, we repeat this calculation and compare it against the calculations of Fortney et al. (2010), Deming et al. (2013) and Line et al. (2013). Generally, there is excellent agreement and the discrepancies are at the level of ~ 0.1 per cent or less. The discrepancies are due to the different spectral resolutions in
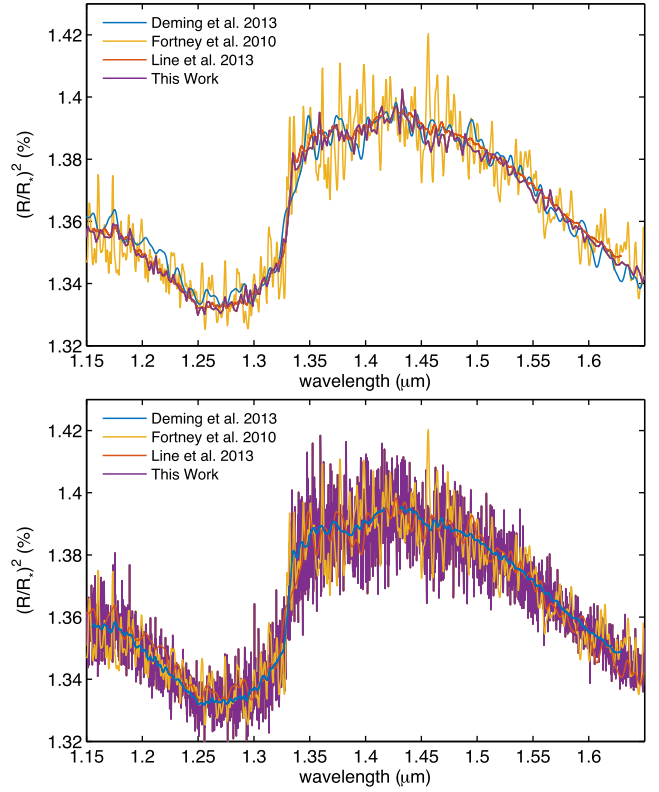

Figure 3. Benchmarking our full numerical calculation of the transmission spectrum of HD 209458b (see text for details of input parameters) to those of Fortney et al., Deming et al. and Line et al. We used a spectral resolution of $\sim 0.1\,\mathrm{cm^{-1}}$ for the opacity function to compute the model transmission spectrum. In the top and bottom panels, we then binned down the model to resolutions of ~ 10 cm and $\sim 1\,\mathrm{cm^{-1}}$, respectively, to illustrate that minor discrepancies result from adopting different spectral resolutions for the water opacity. (Discrepancies may also arise from the use of different spectroscopic line lists to construct the opacity function.)

the opacity function adopted by the different groups. We conclude that our full numerical calculations of transmission spectra are accurate.

3.1.2 Validating isothermal, isobaric formula against full numerical calculations (isothermal, non-isobaric)

Now that we have validated our numerical approach, we compare calculations of the transmission spectrum of the same $T = 1500$ K isothermal atmosphere to those computed using the isothermal formula in equation (12). There is an ambiguity in the isothermal formula in that it is not clear what to assume for the pressure P in the opacity function $\kappa(\lambda, T, P)$. Our intuition is that if the line peaks (rather than the line wings) are dominant as an opacity source, then the computed transmission spectrum should be insensitive to the value of P. In other words, as long as pressure broadening is unimportant, the value of P should be irrelevant. This insight should hold at high temperatures, but break down at low temperatures because of the increasing relative importance of the line wings as an opacity source due to pressure broadening.

In Fig. 4, we validate this hypothesis by demonstrating that the computed transmission spectra for $T = 1500$ K are almost identical for $P = 0.01$ mbar versus $P = 10$ mbar over the wavelength range probed by WFC3. (For the rest of the study, we
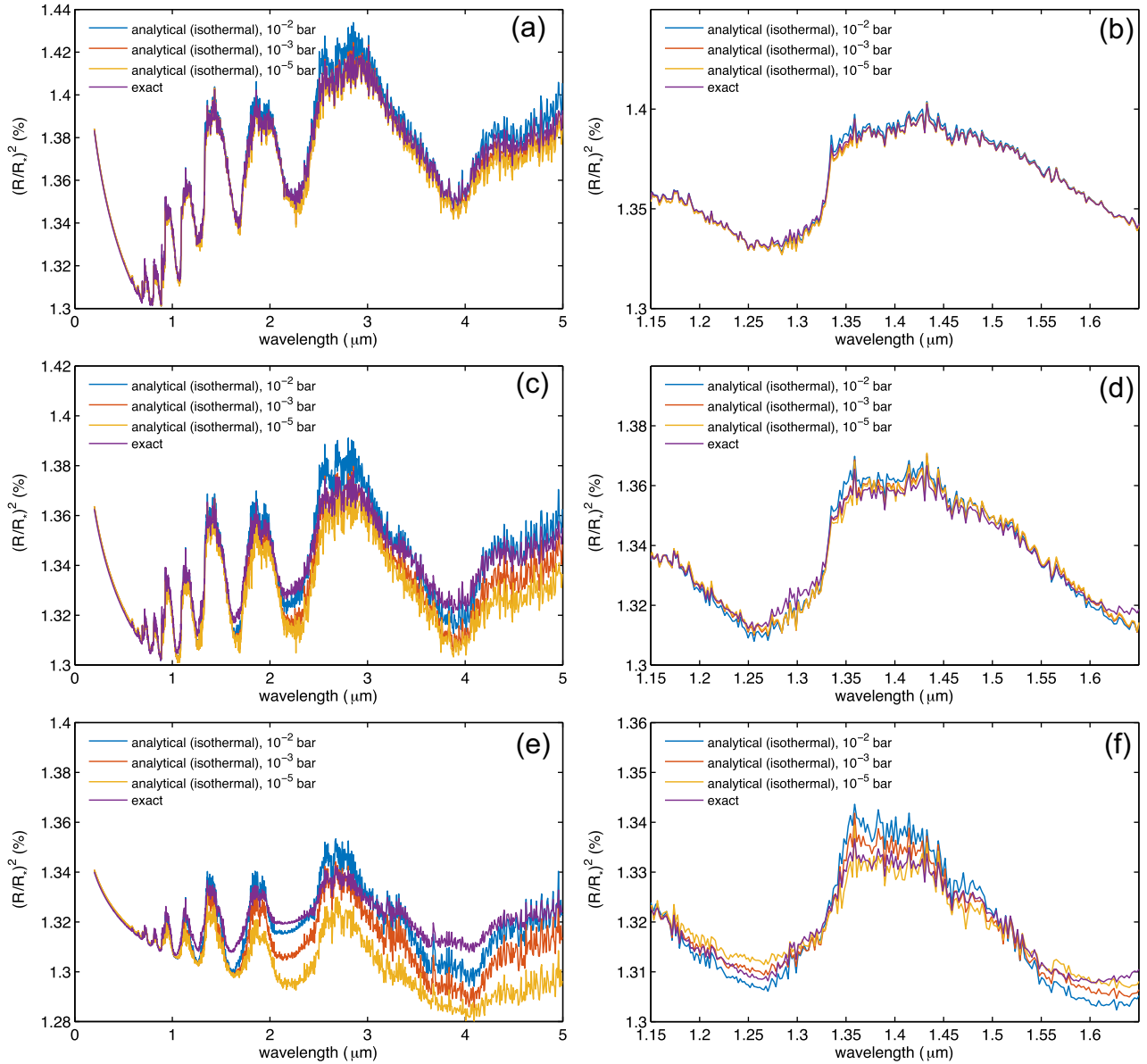
Figure 4. Comparing our full numerical calculations of the transmission spectrum of an isothermal atmosphere to those computed using our isothermal, isobaric analytical formula. The first, second and third rows are for $T = 1500$, 1000 and 500 K, respectively. The left- and right-hand columns are for *JWST*-NIRSpec and *HST*-WFC3 wavelength coverage, respectively. We used a spectral resolution ∼0.1 cm^{-1} for the opacity function and then binned the model down to a resolution ∼10 cm^{-1}. These calculations demonstrate that WFC3 transmission spectra may be approximated as being isobaric (constant pressure) in some instances.

assume $P = 1$ mbar.) Even over the wavelength range probed by the Near Infrared Spectrograph (NIRSpec) on the *JWST*, the agreement is rather remarkable. Table 2 lists the minimum, maximum and mean errors associated with comparing our analytical formula to the full numerical calculations. Generally, the error worsens as the temperature decreases and the wavelength range increases, as seen for the 1000 and 500 K case studies in Fig. 4. It is worth emphasizing that these discrepancies are *not* due to the isothermal assumption (as the numerical model itself is isothermal). Rather, the influence of the line wings, which depends on pressure, becomes stronger as the temperature becomes lower. For example, we see in Fig. 4 that the use of the analytical formula becomes sensitive to the choice of pressure value for the 500 K case study.

3.1.3 Validating isothermal, isobaric formula against full numerical calculations (non-isothermal, non-isobaric)

As a final validation step, we wish to demonstrate that our isothermal, isobaric analytical formula does a decent job of matching full, non-isothermal, non-isobaric numerical calculations. When using the analytical formula, we fix the pressure associated with the water opacities at 1 mbar. We assume a typical temperature gradient of $T' = 0.1$ K km^{-1}, which at the order-of-magnitude level is the value inferred from measurements of HD 189733b (Huitson et al. 2012; Heng et al. 2015; Wyttenbach et al. 2015). We consider two representative cases: $T_0 = 1000$ K with a positive temperature gradient and $T_0 = 2000$ K with a negative temperature gradient. In Fig. 5, we compare the pair of non-isothermal calculations against a set of isothermal calculations done using the analytical formula. We see

Table 2. Errors associated with comparing full numerical calculations to analytical formula in equation (12).

Figure	T probed	Minimum relative (per cent)	Minimum absolute (ppm)	Maximum relative (per cent)	Maximum absolute (ppm)	Mean relative (per cent)	Mean absolute (ppm)
3a	1500 K	1.53×10^{-4}	1.89×10^{-2}	0.61	72.98	0.11	13.19
3b	1500 K	4.75×10^{-5}	5.69×10^{-3}	0.56	67.72	0.11	13.85
3c	1000 K	4.20×10^{-4}	5.25×10^{-2}	1.38	167.00	0.29	35.34
3d	1000 K	1.31×10^{-4}	1.58×10^{-2}	1.10	132.91	0.31	37.96
3e	500 K	2.96×10^{-4}	3.61×10^{-2}	1.79	213.67	0.41	49.09
3f	500 K	3.51×10^{-5}	4.24×10^{-3}	1.24	148.14	0.47	56.53
4a	1250 K	1.84×10^{-4}	2.19×10^{-2}	0.83	100.80	0.23	28.02
4b	1600 K	7.09×10^{-5}	8.79×10^{-3}	0.88	114.40	0.18	22.80

Note. when using the analytical formula, a pressure of 1 mbar was assumed for the water opacity.
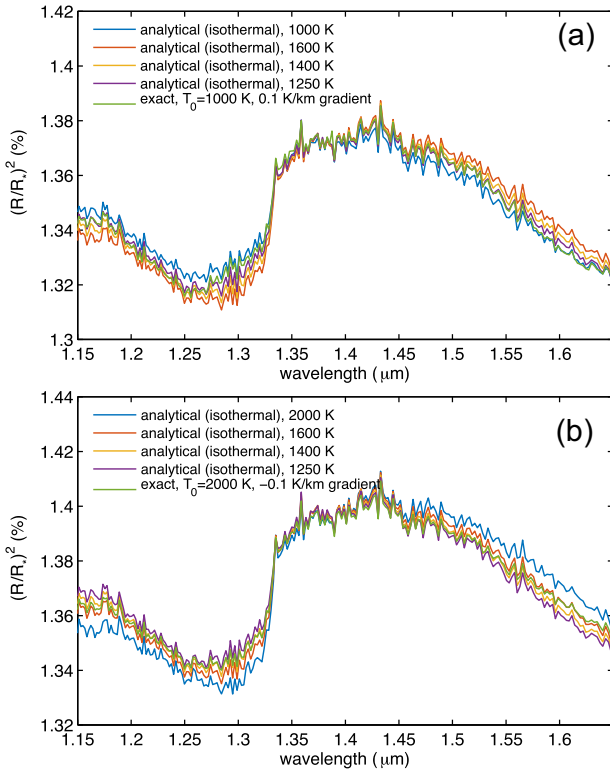

Figure 5. Comparing full, non-isothermal numerical calculations to our isothermal, isobaric analytical formula. The top and bottom panels show model atmospheres with positive and negative temperature gradients, respectively. We used a spectral resolution ~0.1 cm^{-1} for the opacity function and then binned the model down to a resolution ~10 cm^{-1}. These calculations demonstrate that WFC3 transmission spectra may be accurately analysed using an isothermal, isobaric model, which assumes a pressure of 1 mbar for the water opacities.

that the isothermal models with temperatures of 1250 and 1600 K provide satisfactory matches to the numerical calculations for the positive- and negative-gradient case studies, respectively. If a higher level of accuracy is desired, then one may use the non-isothermal formula.

3.2 Analyzing the WFC3 transmission spectrum of WASP-12b

3.2.1 Comparison to specific models to study trends (no data fitting performed)

Now that we have demonstrated the accuracy of the isothermal, isobaric analytical formula for computing WFC3 transmission spectra,

we return to the models presented in Fig. 2. Kreidberg et al. (2015) have previously measured and interpreted the WFC3 transmission spectrum of the hot Jupiter WASP-12b. They assumed the stellar radius to be $R_\star = 1.57 \, R_\odot$. The reference transit radius was fixed at $R_0 = 1.79 \, R_J$, while the reference pressure was assumed to be $P_0 = 10$ bar. As a proof of concept, we adopt the Kreidberg et al. (2015) values for $R_\star$ and P_0, but explore four models where we varied the value of R_0. The values of R_0 adopted are listed in Table 1. Note that these values of R_0 are well within the uncertainties associated with the white-light radius of $1.79 \pm 0.09 \, R_J$ measured by Hebb et al. (2009). Kreidberg et al. (2015) do not report the value of the surface gravity used, but we follow Hebb et al. (2009) and use log g = 2.99 (cgs units) or $g = 977$ cm s^{-2}.

To compute model transmission spectra, we use the isothermal formula in equation (12) and also the non-isothermal formula in equation (21). We add a constant opacity associated with clouds or aerosols to the total opacity, which is described by

$$\kappa = \frac{m_{H_2O}}{m} X_{H_2O} \kappa_{H_2O} + \kappa_{cloud}, \qquad (23)$$

where $m_{H_2O} = 18 \, m_{amu}$ is the mass of the water molecule, $X_{H_2O} = 10^{-3}$ is the volume mixing ratio of water and κ_{H_2O} is the water opacity. In other words, $X_{H_2O} m_{H_2O}/m$ is the mass mixing ratio. We set $m = 2.4 \, m_{amu}$. The values of κ_{cloud} and T' adopted are listed in Table 1. Our constant cloud opacity assumes that the (spherical) cloud particles have circumferences that are larger than the longest wavelength probed, as described by Mie theory (e.g. Pierrehumbert 2010). Generally, higher values of the cloud opacity are needed as the temperature increases (which increases the water opacity).

As already discussed in the beginning of the study, Fig. 2 shows us that the temperature, temperature gradient and degree of cloudiness are not easily teased apart when interpreting the data. This proof-of-concept comparison with the Kreidberg et al. (2015) data informs us that cloudfree models are ruled out, because it is difficult to match the relatively flat continuum bluewards of the 1.4-µm water feature with cloudfree models, but it is harder to pin down the values of temperature and cloud opacity, and also if a temperature gradient is present. Higher temperatures may be compensated by lower values of R_0. The values of R_0 and κ_{cloud} may in turn be adjusted to compensate for each other. The most discerning data points appear to be at 1.3 and 1.6 µm, just bluewards and redwards of the water absorption feature, respectively. The sensitivity of these data points to temperature and cloudiness allows us to perform a fit to the data, provided we fix the values of R_0 and P_0.
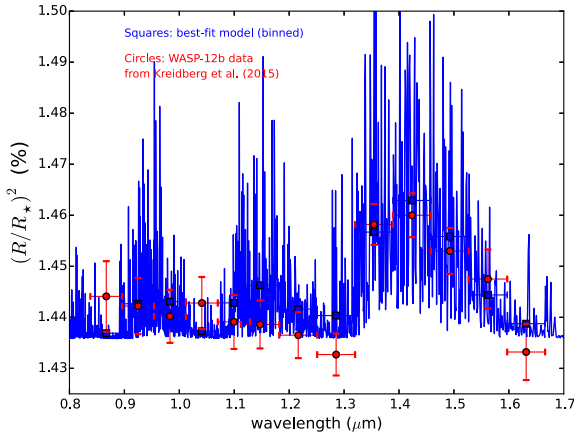
Figure 6. Best-fitting model using our analytical isothermal, isobaric formula to the measured WFC3 transmission spectrum of WASP-12b from Kreidberg et al. (2015).

3.2.2 Fits to data: the X_{H_2O}-R_0-P_0 degeneracy

Fig. 6 shows our formal, three-parameter fit to the WFC transmission spectrum of WASP-12b measured by Kreidberg et al. (2015), using the isothermal, isobaric formula in equation (12). The details of this procedure are described in the Appendix. We first wish to demonstrate that our simpler approach yields the same answer as the full retrieval method of Kreidberg et al. (2015). By setting $R_0 = 1.79\,R_J$ and $P_0 = 10$ bar, we obtain $T = 1020$ K and $X_{H_2O} = 3.9 \times 10^{-3}$, which is consistent with the range of values reported in table 4 of Kreidberg et al. (2015). The cloud opacity is $\kappa_{cloud} = 3.5 \times 10^{-3}$ cm^2 g^{-1}. Kreidberg et al. (2015) mention the use of a planetary radius scaling factor 'to account for uncertainty in the pressure level in the atmosphere at a given radius'. Their fig. 11 shows that this scale factor is $0.99^{+0.01}_{-0.02}$, essentially very close to unity. Recall that variations in the reference pressure appear as $H \ln P_0$ in the formula for the transit radius, which implies that they are muted by the logarithm even if P_0 is varied by several orders of magnitude. By contrast, variations in R_0 affect the transit radius linearly and produce large changes in the water mixing ratio. It is unclear if this scaling factor accounts for the uncertainties in the white-light radius of $1.79 \pm 0.09\,R_J$ measured by Hebb et al. (2009).

Next, we set $R_0 = 1.85\,R_J$, which is a 3.4 per cent increase from its previous value. Remarkably, such a small increase in the reference transit radius leads to more than 5 orders of magnitude of decrease in the water mixing ratio obtained from the fit: $X_{H_2O} = 2.8 \times 10^{-8}$. The cloud opacity becomes $\kappa_{cloud} = 2.5 \times 10^{-8}$ cm^2 g^{-1}. Essentially, cloudiness and water abundance may be cancelled out by a larger normalization. Furthermore, it is really $P_0 X_{H_2O}$ that is the fitting parameter and not X_{H_2O} alone. Any linear change in the reference pressure is compensated by the same factor in the water mixing ratio. In Fig. 7, we show the values of $X_{H_2O}(P_0/10\,\text{bar})^{-1}$ obtained from fitting to the data as a function of the value of R_0 assumed.

A more detailed exploration of these degeneracies and the posterior distributions of parameters is beyond the scope of this study and deferred to future work.

3.3 An unresolved challenge: How do we relate R_0 and P_0 when analysing data?

An outstanding issue with setting R_0 to be the white-light radius is that it is inconsistent with the assumption of $\tau_0 \gg 1$. At all wavelengths, a transmission spectrum is probing transit chords with
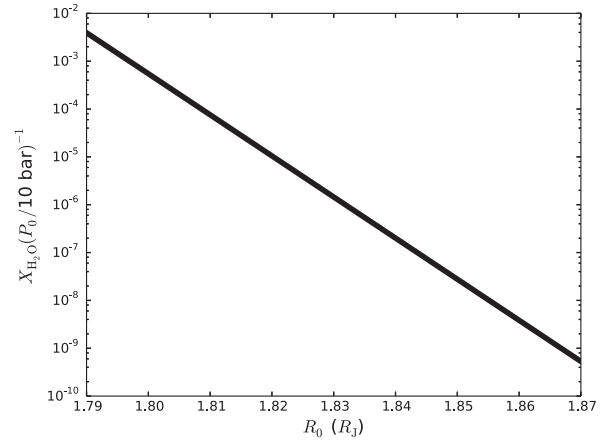

Figure 7. Inferred water abundances, obtained from best fits to the measured WFC3 transmission spectrum of WASP-12b from Kreidberg et al. (2015), versus the assumed value of the reference transit radius.

optical depths ~ 1. A white-light transit radius would correspond to $\tau_0 \sim 1$, not $\tau_0 \gg 1$. This inconsistency may be alleviated by restoring the $E_1 = E_1(\tau_0)$ term in equation (12), yielding

$$R = R_0 + H \left[\gamma + E_1 + \ln \left(\frac{P_0 \kappa}{g} \sqrt{\frac{2\pi R_0}{H}} \right) \right]. \quad (24)$$

The preceding expression is now valid for all values of τ_0, but the issue remains that the functional form of $R_0(P_0)$ is unknown.

Phenomenologically, one may fix P_0 and fit for the value of R_0. The fit to the data will produce a solution for R_0, but it is unclear if this value of the reference transit radius corresponds correctly to the chosen value of P_0. To give a concrete example, we return to the case of WASP-12b. We use the white-light radius to set $R_0 = 1.79 \pm 0.09\,R_J$ (Hebb et al. 2009). But what is the value of P_0? The measured transmission spectrum has nothing to say on this issue, as one can only fit for the value of $P_0 X_{H_2O}$ and not X_{H_2O} alone.

Ideally, we would like to fix R_0 to the white-light radius and use theory to inform us what P_0 is. For isothermal, isobaric atmospheres, we have (Heng 2016)

$$P_0 \sim 0.56 \frac{g}{\bar{\kappa}} \sqrt{\frac{H}{2\pi R_0}}, \quad (25)$$

where we interpret $\bar{\kappa}$ as the geometric mean opacity within the white-light bandpass. The preceding expression is correct only at the order-of-magnitude level because of the use of the geometric mean opacity. But it would break the degeneracy associated with the reference transit radius and reference pressure. Having determined the values of both R_0 and P_0, we may then perform a fit to the data to infer the water abundance.

Using $g = 977$ cm s^{-2}, $T = 1020$ K, $m = 2.4\,m_{amu}$ and $R_0 = 1.79\,R_J$, we obtain $H = 362$ km and $P_0 \sim 1$ mbar $(\bar{\kappa}/0.01$ cm^2 g$^{-1})^{-1}$.

4 DISCUSSION

4.1 Implications and possible solutions

There are several major implications of our findings.

(i) Beyond identifying the presence of water in the WFC3 data, our ability to accurately infer the water abundance is limited, due to transmission spectra lacking an absolute normalization. Our

ignorance of this absolute normalization is due to the unknown relationship between the reference transit radius and the reference pressure.

(ii) It has previously been proposed that the *J* band (1.22–1.30 μm) and water spectral feature (1.36–1.44 μm), located within WFC transmission spectra, serve as diagnostics for the degree of cloudiness in an exoplanetary atmosphere (Stevenson 2016). The lack of an absolute normalization and the degeneracies associated with temperature, a finite temperature gradient and the degree of cloudiness suggest that the true picture is more complicated and the interpretation is multi-dimensional.

(iii) We suggest an approximate relationship between the reference transit radius and reference pressure. However, since it is correct only at the order-of-magnitude level, it implies that the inferred abundance of water is only accurate at the order-of-magnitude level.

Other solutions to break the degeneracy have been suggested. If pressure broadening of the line wings is a major source of opacity, then it will break the insensitivity of the transmission spectrum to the pressure being sensed, as previously suggested by Griffith (2014), although this is only useful at low temperatures. This behaviour is clearly seen for the 500 K case study in Fig. 4. The caveat is that the theory of pressure broadening is incomplete, and the choice of line-wing cut-off[6] may impact the inference made (Sharp & Burrows 2007; Grimm & Heng 2015). A cleaner idea is to use the measured transit radius associated with the optical/visible Rayleigh slope to pin down the pressure being sensed (Griffith 2014). This approach is unambiguous only if the atmosphere is cloud-free, such that Rayleigh scattering may be robustly associated with the bulk gas in the atmosphere (Heng 2016). In this case, the opacity due to Rayleigh scattering and the atmospheric pressure (dominated by the bulk gas) are straightforwardly related.

Yet another solution is to perform a joint analysis of emission and transmission spectra, as the former is not subjected to the degeneracies discussed (Griffith 2014). However, it remains unclear if the atmospheric conditions probed in emission and transmission may be described by a single, one-dimensional (1D) model, as hot exoplanets are expected to be 3D entities (Burrows et al. 2010; Heng, Menou & Phillipps 2011; Heng & Showman 2015).

Generally, breaking this degeneracy in the face of additional degeneracies introduced by the presence of clouds and hazes is an outstanding problem. For all of these reasons, we consider this challenge to be unresolved.

4.2 Comparison to Griffith (2014)

Taken at face value, some of the conclusions made by Griffith (2014) are qualitatively similar to those of this study. As such, we will now perform a detailed comparison to, and analysis of, the study of Griffith (2014).

No formal derivation of the transit radius was performed in Griffith (2014). In fact, Griffith (2014) does not show an analytical expression for *R* at all, which is a key difference between that and this study. Rather, the study of Griffith (2014) relies on analytical scalings, rather than analytical derivations – the latter refers to starting from first principles, whereas the former uses heuristic arguments. One consequence of this approach is that the origin of the relationship between the reference radius (R_0) and reference

pressure (P_0) was not elucidated. Griffith (2014) asserts that R_0 is the transit radius at a pressure where the atmosphere is opaque, but our derivation shows that this is not a necessary condition, and that $R_0(P_0)$ is simply a mathematical constant of integration. By retaining the extra term involving the exponential integral, it is possible to associate R_0 with any optical depth.

Griffith (2014) states that a degeneracy arises from 'uncertainties in the assignment of the pressure level associated with the retrieved radius at unity optical depth'. This is consistent with our finding that transmission spectra are insensitive to the pressure probed. However, this is a separate and distinct degeneracy from the unknown relationship between R_0 and P_0, since the reference radius is generally not located at an optical depth of unity. In fact, the lack of a first-principles analytical formula for *R* is probably the reason why Griffith (2014) did not discover the degeneracy between R_0 and $\ln(P_0 X)$, where *X* is the volume mixing ratio. Griffith (2014) discusses two models of an isothermal atmosphere with slightly different values of the reference radius: R_0 and R_0'. Using our analytical formula, we realize that equation (3.6) of Griffith (2014) is actually a simplified version of a more general, first-principles expression,

$$\frac{P_0 X}{P_0' X'} = \exp\left(-\frac{\Delta R_0}{H}\right) \sqrt{\frac{R_0'}{R_0}}, \tag{26}$$

where $\Delta R_0 \equiv R_0 - R_0'$. While $\sqrt{R_0'/R_0} \approx 1$ is probably a reasonable assumption (and it is made implicitly by Griffith 2014), the real degeneracy is not just between the two mixing ratios (*X* and *X'*), but rather between $P_0 X$ and $P_0' X'$.

There is a discrepancy between our derivation that the variation in radial distance, across the transit chord, is $\pi H/4$ and the claim of Griffith (2014) that it is about 4*H*. Both claims are made for isothermal atmospheres. Griffith (2014) arrived at this estimate by using an approximate scaling law for δR, bounded by a pair of transmission[7] values. By choosing these transmission values to be 0.05 and 0.95 (corresponding to optical depths of about 3 and 0.05, respectively), Griffith (2014) estimated that $\delta R \approx 4H$. These are rather ad hoc choices. For example, choosing 0.01 and 0.99 instead yields $\delta R \approx 6H$. fig. 6 of Griffith (2014) does not actually validate these choices because it simply plots the value of the transmission as a function of depth/pressure. In our approach, we have used the result that there is an effective transit chord with a chord length of $\sqrt{2\pi H R}$ (Fortney 2005; Lecavelier des Etangs et al. 2008; de Wit & Seager 2013; Heng et al. 2015), and used geometry to arrive at $\delta R = \pi H/4$ (Fig. 1).

Pedantically (and for completeness), we note that Griffith (2014) (re)discovered the $\sqrt{2\pi H R}$ expression for the transit chord without crediting Fortney (2005). Furthermore, there is no discussion or derivation of the result that the effective chord optical depth is 0.56.

4.3 Why were these degeneracies not uncovered in retrieval studies?

At first sight, these degeneracies do not seem to exist in the traditional forward problem (e.g. Hubbard et al. 2001; Fortney et al. 2008, 2010). Given a set of assumptions (e.g. identities and abundances of molecules, chemistry, irradiation conditions), one may construct a model atmosphere and trace lines of sight (chords) through it to obtain the transmission spectrum. For a fixed set of

---

[6] In this study, we adopt a cut-off of 25 cm^{-1} for the water lines.

[7] Defined by Griffith (2014) as being $\exp(-\tau)$, where τ is the optical depth.

parameter values, one may uniquely calculate the relationship between each transit chord, the transit radius and pressure, and also infer the reference pressure (given a choice of the reference transit radius).

A retrieval modeller might plausibly object by stating that he or she has set up a grid in pressure or distance and explored a suite of model atmospheres with different chemistries, temperature–pressure profiles, etc. For each model in the suite, the pressure corresponding to each transit radius, as well as the reference transit radius and reference pressure, may be straightforwardly calculated. However, recall that because transmission spectra are insensitive to the pressure probed (unless pressure broadening comes into play), there exist multiple families of model atmospheres where the grid in pressure may be shifted up or down with little to no effect on one's ability to fit the observed spectrum. For each choice of grid, there is a different set of $R_0(P_0)$ relationships. Model suites with a fixed choice of pressure grid would not fully explore the degeneracies unearthed in this study.

ACKNOWLEDGEMENTS

REFERENCES

Beichman C. et al., 2014, PASP, 126, 1134
Benneke B., Seager S., 2012, ApJ, 753, 100
Bétrémieux Y., 2016, MNRAS, 456, 4051
Bétrémieux Y., Swain M. R., 2017, MNRAS, 467, 2834
Brown T. M., 2001, ApJ, 553, 1006
Burrows A., Sudarsky D., Hubbard W. B., 2003, ApJ, 594, 545
Burrows A., Rauscher E., Spiegel D. S., Menou K., 2010, ApJ, 719, 341
Chandrasekhar S., 1960, Radiative Transfer. Dover, New York
de Wit J., Seager S., 2013, Science, 342, 1473
Deming D. et al., 2013, ApJ, 774, 95
Fortney J. J., 2005, MNRAS, 364, 649
Fortney J. J., Lodders K., Marley M. S., Freedman R. S., 2008, ApJ, 678, 1419
Fortney J. J., Shabram M., Showman A. P., Lian Y., Freedman R. S., Marley M. S., Lewis N. K., 2010, ApJ, 709, 1396
Greene T. P., Line M. R., Montero C., Fortney J. J., Lustig-Yaeger J., Luther K., 2016, ApJ, 817, 17
Griffith C. A., 2014, Phil. Trans. R. Soc. A, 372, 86
Grimm S. L., Heng K., 2015, ApJ, 808, 182
Guillot T., 2010, A&A, 520, A27
Hebb L et al., 2009, ApJ, 693, 1920
Heng K., 2016, ApJ, 826, L16
Heng K., Showman A. P., 2015, Annu. Rev. Earth Planet. Sci., 43, 509
Heng K., Menou K., Phillipps P. J., 2011, MNRAS, 413, 2380
Heng K., Wyttenbach A., Lavie B., Sing D. K., Ehrenreich D., Lovis C., 2015, ApJL, 803, L9
Hubbard W. B., Fortney J. J., Lunine J. I., Burrows A., Sudarsky D., Pinto P., 2001, ApJ, 560, 413
Huitson C. M., Sing D. K., Vidal-Madjar A., Ballester G. E., Lecavelier des Etangs A., Désert J.-M., Pont F., 2012, MNRAS, 422, 2477
Kreidberg L. et al., 2015, ApJ, 814, 66
Lecavelier des Etangs A., Pont F., Vidal-Madjar A., Sing D., 2008, A&A, 481, L83
Line M. R., Knutson H., Deming D., Wilkins A., Desert J.-M., 2013, ApJ, 778, 183
Pierrehumbert R. T., 2010, Principles of Planetary Climate. Cambridge Univ. Press, Cambridge
Robinson T. D., 2017, 836, 236
Seager S., Sasselov D. D., 2000, ApJ, 537, 916
Sharp C. M., Burrows A., 2007, ApJS, 168, 140
Stevenson K. B., 2016, ApJL, 817, L16
Vahidinia S., Cuzzi J. N., Marley M., Fortney J., 2014, ApJL, 789, L11
Waldmann I., Tinetti G., Rocchetto M., Barton E. J., Yurchenko S. N., Tennyson J., 2015, ApJ, 802, 107
Wyttenbach A., Ehrenreich D., Lovis C., Udry S., Pepe F., 2015, A&A, 577, A62

APPENDIX: DETAILS OF MODEL FITTING

We perform a fit of equation (12) to the binned data points listed in table 3 of Kreidberg et al. (2015) by implementing the `curve_fit` routine in PYTHON. We take the following approximation when computing the model transmission spectrum: We compute the water opacities (with a pressure fixed at 1 mbar) in each of the wavebands listed in table 3 of Kreidberg et al. (2015). Within each wandband, we compute the *geometric mean* of the opacities. We then use these mean opacities to compute the model transmission spectrum, which we fit to the binned data. We check that our approach is sound by using the values of the fitting parameters to compute the transmission spectrum at full spectral resolution (~ 0.1 cm^{-1}; shown as the curve in Fig. 6), which we then bin down (shown as the squares in Fig. 6) and compare to the data points (shown as the circles in Fig. 6). When this approach is repeated with the arithmetic mean of the opacities, it is apparent that the binned model does not match the data (not shown). Physically, we expect the geometric mean to be a fair averaging of the opacities, since one is dealing with values that span many orders of magnitude.

This paper has been typeset from a TEX/LATEX file prepared by the author.