Characterising Exoplanet Atmospheres using Traditional Methods and Supervised Machine Learning

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Chloe Fisher

von United Kingdom

Leiter der Arbeit:
Prof. Dr. Kevin Heng
Center for Space and Habitability (CSH)

Characterising Exoplanet Atmospheres using Traditional Methods and Supervised Machine Learning

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Chloe Fisher

von United Kingdom

Leiter der Arbeit:
Prof. Dr. Kevin Heng
Center for Space and Habitability (CSH)

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Bern, 17.09.2021

Der Dekan:
Prof. Dr. Zoltan Balogh

Contents

Abstract

Since the discovery of the first extrasolar planets over 25 years ago, the field of exoplanet research has exploded. Today we have over 4000 confirmed exoplanets, with a wide variety of sizes, orbital separations, and host stars. The characterisation of this diverse population of objects has led to exciting discoveries about the conditions of alien worlds. Future technological advances are expected to provide an abundance of exoplanet spectra with a higher precision and sensitivity than ever before. This calls for a parallel advancement in the accuracy and speed of atmospheric models to interpret this influx of data.

In this thesis, my work on atmospheric retrievals is presented. Starting with traditional techniques, my first thesis paper applies a Bayesian retrieval in combination with an analytical atmospheric model to the Hubble transmission spectra of 38 different exoplanets. My second paper considers the theoretical model of the sodium doublet, and the effect of dropping the assumption of local-thermodynamic equilibrium. From here, I went on to develop a method that uses supervised machine learning to improve the speed and efficiency of the retrieval. This method was explained and tested in a collaborative paper with machine learning experts in Bern. The machine learning retrieval is then applied in several follow-up studies, covering a range of different scenarios. One of these was my final thesis paper, which further extends the new retrieval to high-resolution spectra using the cross-correlation function.

Abstract

In addition to my own papers, I have contributed to a number of studies led by collaborators by running retrievals, assisting other students, and participating in scientific discussions. I have also worked on several observing proposals, both for high-resolution ground-based observatories and for the upcoming James Webb Space Telescope.

I plan to continue my work on exoplanet characterisation and machine learning in the future, using the technique to combine high- and low-resolution spectra to gain further insight into the atmospheres of these distant planets. The speed and efficiency of machine learning will also allow for statistical studies of exoplanets as the quantity of atmospheric spectra from new and upcoming telescopes escalates. Not only will these studies teach us about the conditions and potential habitability of exoplanets, but they will also answer questions about planet formation, diverse chemical processes, and the uniqueness of our solar system.

Introduction

Let's make great discoveries together!

Kevin Heng

1.1 Aim of this Thesis

This thesis covers the work from my PhD over the last four years. The general focus is on atmospheric retrieval of exoplanets. This involved first understanding and running traditional retrievals using Bayesian methods, before moving on to developing a new method of retrievals using machine learning. This led to a series of papers testing the method and applying it to new situations to gain additional insights into different aspects of exoplanets. This thesis covers the work in detail, as well as additional projects I worked on over the last four years.

To put this work into context, Chapter 1 provides an introduction into the field of exoplanets, covering both detection and characterisation. Details of the different components in atmospheric modelling are presented, as well as techniques for

comparing models to data. It also examines the recent development of machine learning methods in exoplanet research, and the wider goal of searching for extraterrestrial life. Finally, new and upcoming telescopes and their impact on the future of exoplanet science are discussed.

In Chapter 2, my first thesis paper is presented and summarised. This paper is a retrieval analysis of a set of 38 exoplanets with Wide-Field Camera 3 transmission spectra from the Hubble Space Telescope. The retrieval results are examined to look for trends between the model parameters, although no evidence for these are found.

Chapter 3 covers my second thesis paper, which introduces a model for the sodium doublet that does not obey local-thermodynamic equilibrium. This non-LTE model is applied to both low- and high-resolution transmission spectra, as well as simulated spectra.

In Chapter 4, the first paper using the Random Forest machine learning method to perform atmospheric retrieval is introduced. Although not a first-author paper, it represents a large part of my thesis work, and I contributed heavily to the study. The Random Forest is explained and applied to the transmission spectrum of the hot Jupiter WASP-12 b.

In Chapter 5, my final thesis paper is presented. This work applies the Random Forest to wideband high-resolution spectra, using the cross-correlation function to overcome the challenges associated with this data. This method is applied to the HARPS-N spectrum of the ultra-hot Jupiter KELT-9 b.

Chapter 6 includes two similar proposals for Cycle 1 of the James Webb Space Telescope, which I either led or contributed to significantly. The two proposals use a similar science case to observe different systems containing multiple mini-Neptunes. Unfortunately, the one I led was not successful, but I intend to resubmit in a future cycle.

In Chapter 7, other studies I supported are presented, with a summary of my contribution.

Chapter 8 includes a concluding summary of my thesis, as well as a discussion of my future projects, which I aim to work on during my SNSF Postdoc.Mobility Fellowship in the coming years.

1.2 Introduction to Exoplanets

The search for life elsewhere in the universe remains one of science's greatest pursuits. The most promising field for this is the detection and characterisation of extrasolar planets. The detailed study of these exoplanets, including those that may harbour life, is a core science goal of the next generation of major observatories. Recent advances in technology have led to major increases in the precision and sensitivity of data, bringing a new era of exoplanet characterisation.

The study of these distant objects has a short but rich history. In 1992, the first ever exoplanets were discovered, orbiting the pulsar PSR B1257+12 (Wolszczan and Frail, 1992). Just three years later, the first exoplanet orbiting a main-sequence star was detected (Mayor and Queloz, 1995). Since then, more than 4000 exoplanets have been confirmed, and the field of exoplanet characterisation has blossomed.

1.3 Exoplanet Detections

In the first few years, exoplanets were typically detected using the radial velocity method, which measures periodic shifts in the stellar spectrum caused by the gravitational effect of the planet on its host star. A landmark discovery for the field of exoplanet characterisation was the first detection of a transiting planet (Charbonneau et al., 2000). By observing small fluctuations in the stellar light caused by the planet transiting between the star and the observer, the presence of an exoplanet, and its radius, can be determined. This new technique would allow astronomers to directly observe the atmosphere of an exoplanet, and the first instance of this came just two years later (Charbonneau et al., 2002).

These two detection methods are still the most fruitful 20 years on. Figure 1.1 shows the distribution of exoplanets across mass and period, with their detection technique highlighted. The majority of the first confirmed exoplanets belonged to a group known as hot Jupiters. These are close-in planets with masses similar or greater than Jupiter. Their large size and short-period orbit make these objects ideal for detecting. However, occurrence studies suggest these are in fact the least common type of exoplanet. The most abundant groups of exoplanets are super-Earths and mini-Neptunes – rocky planets with a size between that of Earth and

Neptune. See Fulton et al. (2017) for an overview of exoplanet occurrence rates from the *Kepler* survey.
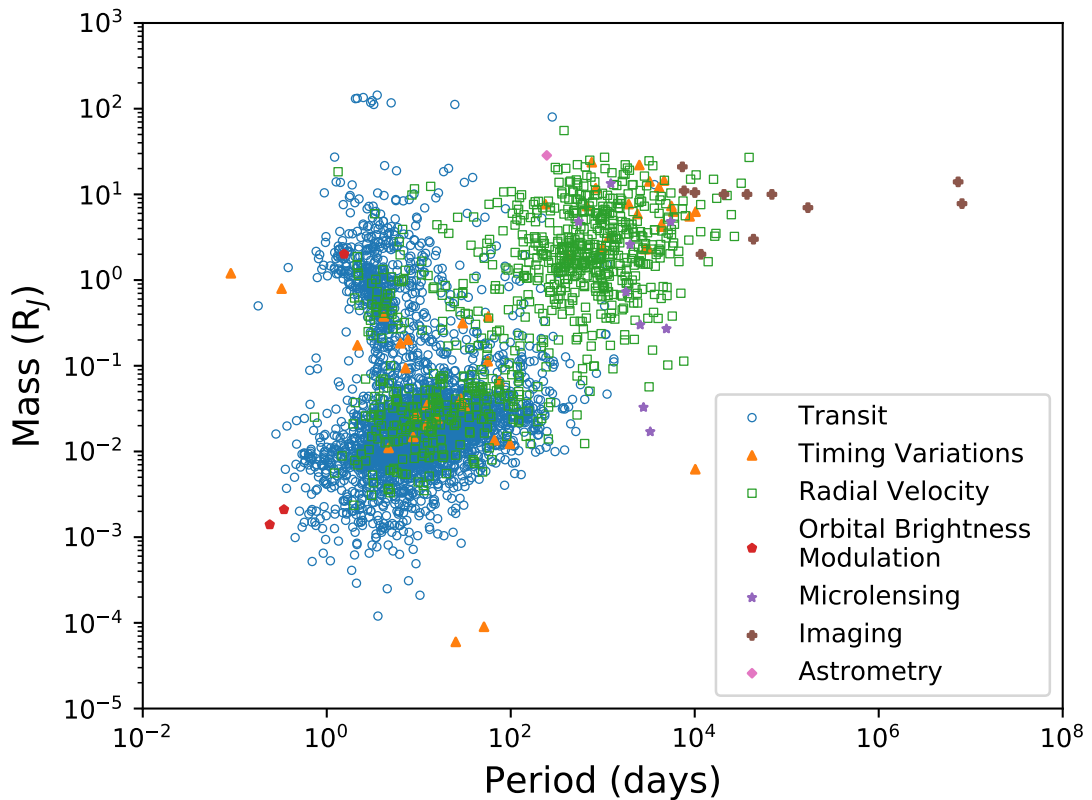
Figure 1.1: Mass versus period for 4261 confirmed exoplanets to-date, and their respective detection technique. Plot made using data from the NASA exoplanet archive.

1.3.1 Radial Velocity

Following its early use in exoplanet detection, the radial velocity method remains one of the most popular techniques. It relies on the Doppler shift of the stellar spectrum, which occurs as it moves along the observer's line of sight (Nelson and Angel, 1998). This movement is caused by the gravitational effect of the planet on the host star. Although one typically imagines the planet orbiting around the star, in reality the two objects orbit their common centre of gravity. This small movement of the star can be detected by high-resolution ground-based spectrographs, down to a precision of 30 cm/s, below which one is limited by the stellar

noise (Petersburg et al., 2020). Figure 1.2 shows a diagram of the radial velocity technique.
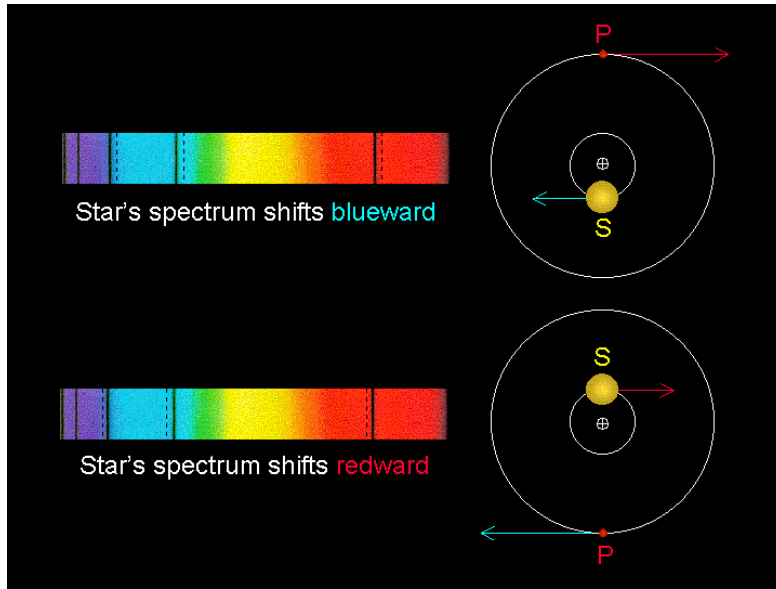
Figure 1.2: The radial velocity technique of detecting exoplanets. Image source: https://faculty.uca.edu/njaustin/PHYS1401/Laboratory/exoplanet.html

One benefit of using the radial velocity technique is the ability to constrain the planet's mass. Using Kepler's third law, one can relate the velocity of the star to the period P, stellar mass m_*, and planetary mass m_{pl}, via the equation

$$K = \left(\frac{2\pi G}{P} \right)^{\frac{1}{3}} \frac{m_{\text{pl}} \sin i}{m_*^{2/3}} \frac{1}{\sqrt{1 - e^2}}, \tag{1.1}$$

where K is the radial velocity semi-amplitude of the star, G is the gravitational constant, i is the inclination of the system with respect to the observer, and e is the eccentricity of the orbit. The caveat of using this equation is that the inclination of the system with respect to the observer is unknown. Therefore, one can only obtain a measurement for $m_{\text{pl}} \sin i$, leading to a *minimum* mass of the planet. However, this still provides useful constraints on the planet's density and potential composition if the radius is known.

1.3.2 Transits

The transit technique is currently responsible for the majority of detected exoplanets. When an exoplanet passes in front of its host star in the observer's line of sight, a small fraction of the stellar light is obscured, causing a dip in its brightness. This is known as a transit (Borucki and Summers, 1984). By measuring the depth of this dip and the periodicity, astronomers can not only detect an exoplanet, but also determine its radius and orbital separation. A diagram of the transit technique is shown in Figure 1.3. Using simple geometry, one obtains the equation for the transit depth as

$$\Delta = \frac{F_{\text{out}} - F_{\text{in}}}{F_{\text{out}}} = 1 - \frac{F_{\text{in}}}{F_{\text{out}}} = 1 - \frac{\left[1 - (R_{\text{pl}}/R_*)^2\right] F_* + F_{\text{pl}}}{F_* + F_{\text{pl}}}, \qquad (1.2)$$

where F_{in} and F_{out} are the fluxes in and out of transit, R_{pl} and R_* are the planetary and stellar radii, and F_{pl} and F_* are the planetary and stellar emitted flux. Assuming F_{pl}, which originates from the nightside of the planet, is negligible, the equation simplifies to

$$\Delta = \left(\frac{R_{\text{pl}}}{R_*}\right)^2. \qquad (1.3)$$

In addition to being an efficient detection method, the transit technique also provides direct observations of the exoplanet's atmosphere. As the planet transits in front of the star, the stellar light filters through its atmosphere, resulting in a transmission spectrum (Seager and Sasselov, 2000) (see Section 1.4.2). Later on in the orbit, the planet's dayside is visible immediately before it passes behind the star. This process, known as the secondary eclipse, allows the observer to measure the planet's emission spectrum (Charbonneau et al., 2005) (see Section 1.4.3). As well as these spectroscopy observations, the fluctuation in the planet's brightness across the entire orbit can be measured, building what's known as a phase curve (Harrington et al., 2006). By analysing this, astronomers can determine the longitudinal temperature structure across the planet's atmosphere.
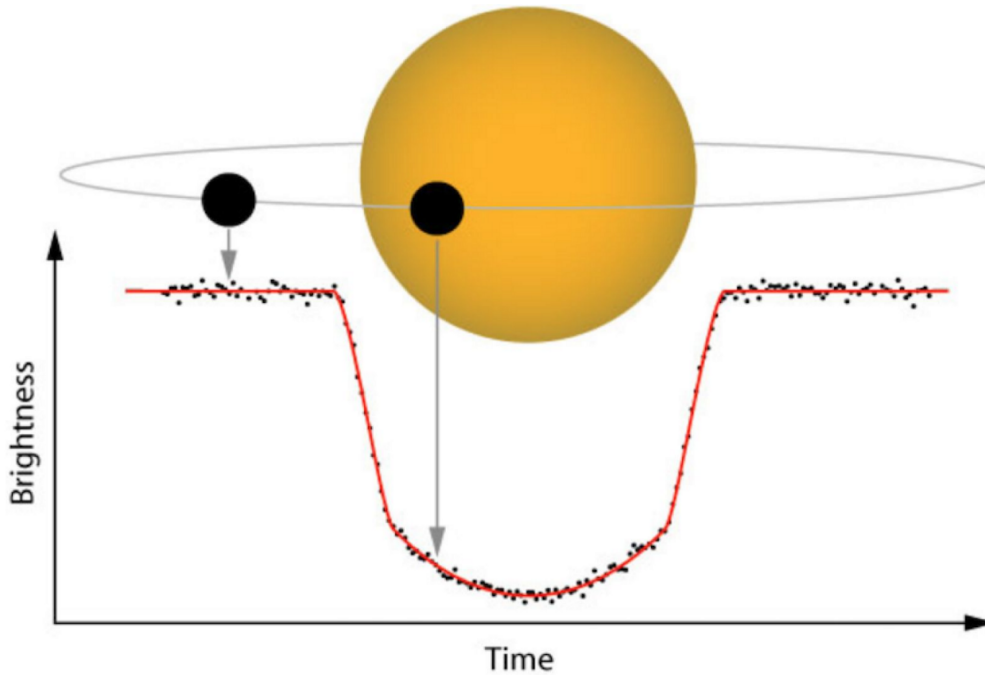
Figure 1.3: A transiting exoplanet and the observed light curve. Image credit: NASA.

1.3.3 Direct Imaging

Although its use is significantly less common that the radial velocity and transit techniques, direct imaging remains one of the most promising exoplanet detection methods. By using a coronagraph to block the stellar light, exoplanets can be resolved from their host star and observed directly (Marois et al., 2010). Figure 1.4 shows an example of directly imaged exoplanets. In order to observe an exoplanet with direct imaging, the planet needs to be on a wide orbit around its host star, with a large size and high temperature to provide enough thermal emission to be detected. This currently limits our ability to directly image most known exoplanets, but future technological advances hope to alleviate this problem.

A key advantage of direct imaging is the ability to directly observe an exoplanet's atmosphere, without the need to separate it from that of the star. This provides a much clearer spectrum for the planet, improving our capability to accurately characterise its atmosphere. It's this advantage that makes direct imaging so promising for future exoplanet analysis, when higher order effects will need to be measured to determine a planet's habitability.
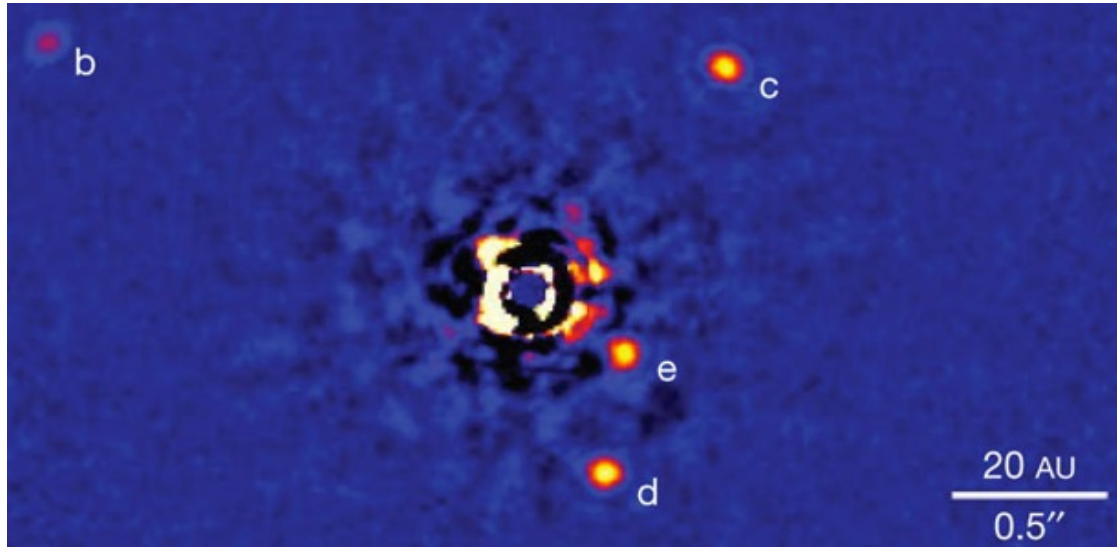
Figure 1.4: The directly imaged exoplanet system orbiting HR 8799 (Marois et al., 2010).

1.3.4 Other Methods

Several other methods of exoplanet detection exist, though their use is less frequent. One of the most common of these is microlensing, where astronomers measure the gravitational effect of the planet on the light of a distant background star (Udalski et al., 2002). This technique is particularly useful for finding low-mass or very distant exoplanets, which are challenging for other detection methods. Another relatively popular technique is known as transit timing variations (TTVs). By observing small variations in the transit time of a known exoplanet, additional planets can be detected in the same system (Ballard et al., 2011). TTVs are an extremely sensitive method of detecting exoplanets, with the capability of discovering planets as small as Earth. They also provide another method for measuring exoplanet masses.

1.4 Exoplanet Characterisation: Observations

Since the first successful detection of an exoplanet atmosphere (Charbonneau et al., 2002), the challenge of characterising these distant objects has become a field of its own (see Seager (2008) for a review). In this section, I present a summary of some observing instruments and techniques used for exoplanet characterisation.

1.4.1 Instruments

Exoplanets are observed with a large number of different instruments, on observatories across the world and in space. Typical observations range from ultraviolet wavelengths, through the optical, and to the far infrared, with each providing distinct and important information about the planet. Additional observations of exoplanets include radio wavelengths, for example using ALMA to observe protoplanetary discs (ALMA Partnership et al., 2015), or SETI's telescopes to search for artificial radio emissions from extraterrestrial life.

Although the first exoplanets were discovered using ground-based spectrographs, the majority to date have been detected from space. The CoRoT telescope (Baglin et al., 2002), operating from 2006 to 2013, discovered about 30 new exoplanets. In 2009, the launch of the Kepler space telescope proved momentous (Borucki et al., 2009). Using the transit technique, Kepler discovered over 2000 confirmed exoplanets, exponentially increasing the field of study. Thanks to the currently operating Transiting Exoplanet Survey Satellite (TESS) (Ricker et al., 2015), we are constantly discovering new exoplanets, and future missions such as PLATO (Rauer et al., 2014) will continue this important work.

When it comes to exoplanet characterisation, both space- and ground-based telescopes have a lot to offer. Space telescopes, such as HST, provide low-resolution spectra with small error bars, allowing for clear molecular features to be studied. On the ground, spectrographs such as HARPS (Mayor et al., 2003), CRIRES (Kaeufl et al., 2004), and ESPRESSO (Pepe et al., 2014) obtain very high-resolution spectra, but with a large level of noise per spectral point. This is useful for detecting certain atoms or molecules in the planet's atmosphere.

We are currently moving into an exciting era for exoplanet observations. The launch of the new James Webb Space Telescope (JWST) later this year will lead to an explosion in precision and sensitivity of our space-based data (Greene et al., 2016). Furthermore, in approximately 6 years first light is expected from the HIRES spectrograph on the Extremely Large Telescope (ELT), currently under development (Zerbi et al., 2014). This will bring a major advancement in the quality of our ground-based spectra. With these next-generation observatories just around the corner, the pressure is on for theorists to advance their analysis techniques and overcome computational limitations.

1.4.2 Transmission Spectroscopy

As previously mentioned, during an exoplanet transit the star light passes through the planet's terminator on its way to the observer. Here, different wavelengths of light are absorbed by different molecules in the atmosphere, obscuring more of the stellar light and causing the planet's radius to appear larger. This is demonstrated in Figure 1.5. The ratio of the planet and stellar radii, or the square of this, is known as the transit depth. By observing this transit depth across multiple wavelengths, we construct what's known as a transmission spectrum. These spectra allow us to characterise the planet's atmosphere, by searching for distinct molecular features and comparing with physical models.
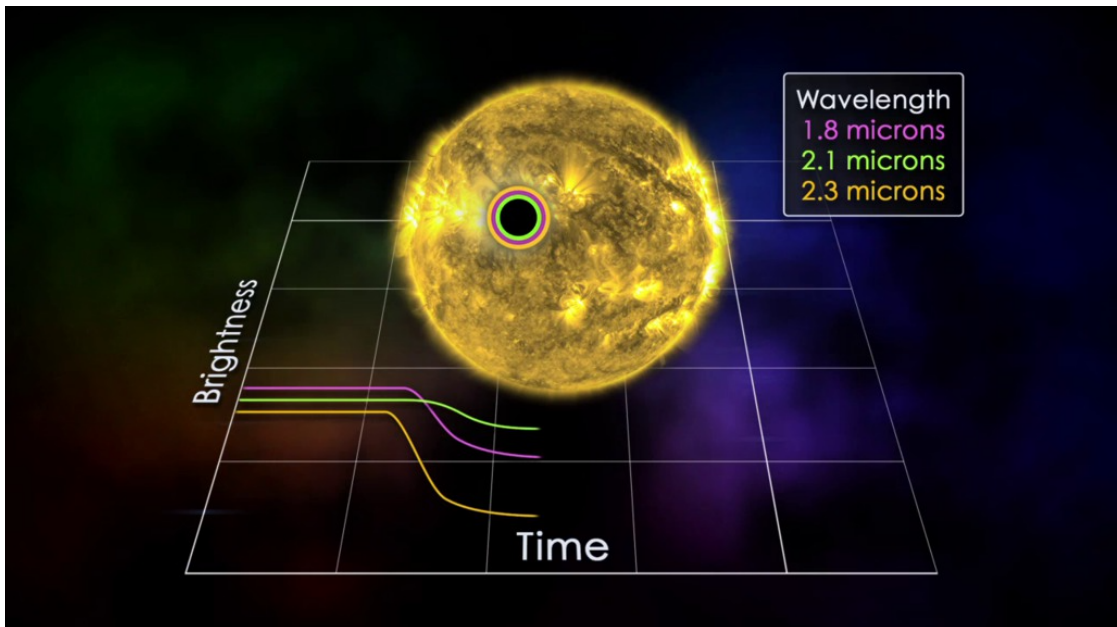
Figure 1.5: Schematic of a transiting planet, with the transit depth shown at different wavelengths. Credit: NASA's Goddard Space Flight Center, additional animations courtesy ESA/Hubble.

The strength of the signal in a planet's transmission spectrum is governed by the size of its atmosphere. We describe this size using a quantity called the scale height, given by

$$H = \frac{k_B T}{mg},$$ (1.4)

where k_B is the Boltzmann constant, T is the temperature of the atmosphere, m is the mean molecular weight, and g is the planet's surface gravity. From this

quantity, we can infer that exoplanets with higher temperature and lower surface gravity will have larger atmospheres, and therefore produce clearer transmission spectra. This indicates that hot Jupiters are the best-suited exoplanet class for transmission spectroscopy, and they still represent the majority of observed transmission spectra we have today. Some prominent examples of well-characterised transmission spectra include the detection of water in WASP-12b (Kreidberg et al., 2015), and sodium in HD 189733 b (Wyttenbach et al., 2015).

1.4.3 Emission Spectroscopy

During a transiting planet's orbit, two eclipses occur. The first is the transit of the planet in front of the star, producing a transmission spectrum as previously discussed. The second is the passing of the planet behind the star, completely obscuring it from the observer's view. This is known as the secondary eclipse. While this event takes place, the flux emitted from the planet is no longer visible. By comparing the flux received from the system immediately before and after the secondary eclipse begins, the wavelength-dependent flux from the planet's dayside can be isolated. This is known as the exoplanet's emission spectrum.

For emission spectra, it is not the scale height that controls the size of the spectral features, but instead the atmospheric temperature profile. If the atmosphere is isothermal then the emission spectrum will be that of a featureless blackbody. These spectra therefore present an opportunity to determine the vertical temperature structure of the atmosphere, as well as constraining molecular abundances. Well-known examples of measured emission spectra include the first observation of a planet's thermal emission for TrES-1 (Charbonneau et al., 2005), and a potential water detection in HD 189733 b (Crouzet et al., 2014).

In addition to observing different regions of an exoplanet, transmission and emission spectra also probe different pressure levels. This makes the study of these two types of spectra highly complementary, and comparing the analysis from both can give us a much greater insight into the planet's atmosphere (e.g. Kreidberg et al., 2014b).

1.4.4 Space-based vs Ground-based Observations

The observation of exoplanets from space, using telescopes such as HST and Spitzer, allows us to measure low-resolution but information-rich atmospheric spectra. These low-resolution spectra can contain clear spectral features for extremely important molecules such as water, methane, and ammonia. Figure 1.6 shows a transmission spectrum of the planet WASP-39 b using a variety of space-based observations (Wakeford et al., 2018). The water feature at 1.4 μm is very clearly resolved by these measurements, allowing the abundance to be precisely constrained.
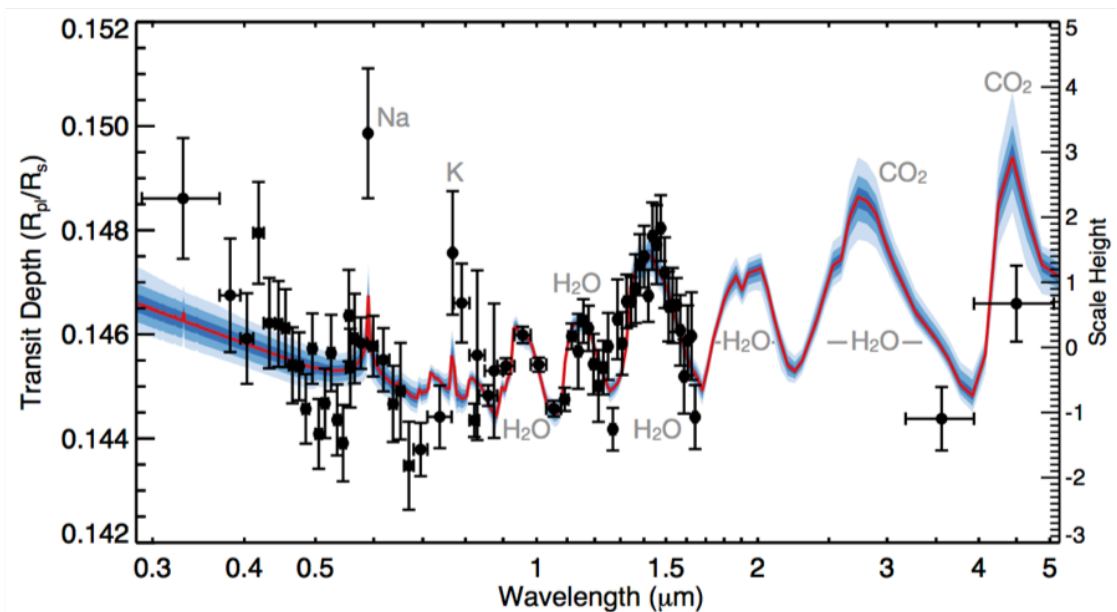
Figure 1.6: Transmission spectrum of WASP-39 b, using data from HST STIS and WFC3, Spitzer IRAC, and VLT FORS2 (Wakeford et al., 2018).

As well as determining molecular abundances, low-resolution spectra offer the opportunity to model the temperature of the atmosphere, and detect the presence of clouds. See Section 1.5 for an explanation of how this is done.

The key disadvantage of space-based observations is that low-resolution spectra can only measure wide spectral features. By using much larger, ground-based spectrographs, we are able to resolve individual spectral lines, allowing us to detect atoms and molecules, such as CO, H_2O, and HCN (e.g. Brogi et al., 2012; Birkby et al., 2013, 2017; de Kok et al., 2013), as well as additional new species like sodium, iron, and other metals (e.g. Wyttenbach et al., 2015, 2017; Hoeij-

makers et al., 2018a, 2019, 2020b; Seidel et al., 2019), which are challenging for space-based analysis. Figure 1.7 shows an example of the resolved sodium lines in the transmission spectrum of WASP-49 b, measured by HARPS (Wyttenbach et al., 2017). Another advantage of resolving individual spectral lines is the opportunity to study the dynamics of an exoplanet's atmosphere. For example, several recent studies have used doppler shifts and broadening of the spectral lines to infer potential wind speeds on the planet (e.g. Wyttenbach et al., 2015; Seidel et al., 2020; Ehrenreich et al., 2020).
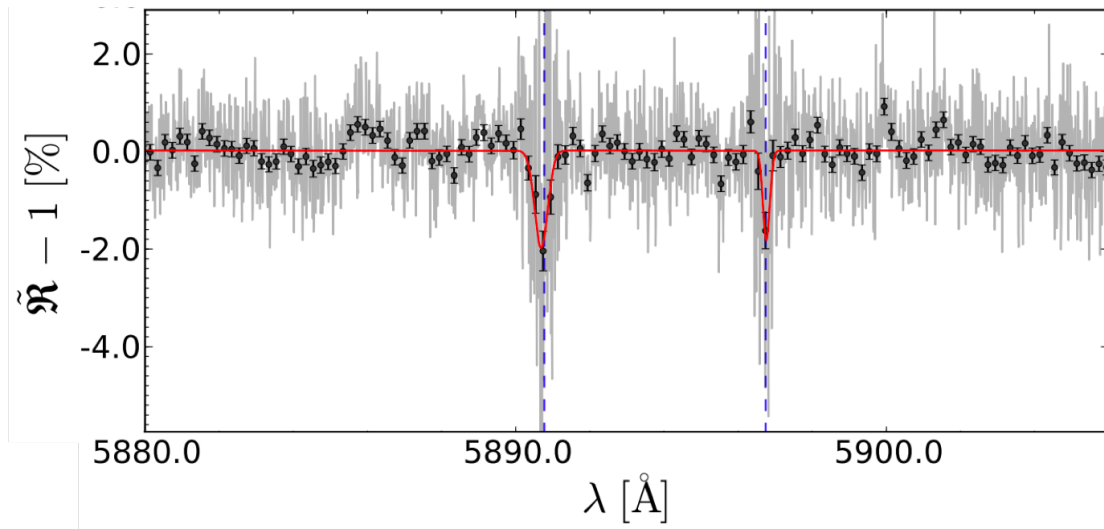
Figure 1.7: Transmission spectrum of WASP-49 b, using data from HARPS (Wyttenbach et al., 2017).

Although ground-based observatories are less restricted than space telescopes in terms of size and flexibility, their measurements are limited by the Earth's atmosphere. This obstacle proves challenging for exoplanet spectroscopy, as not only is it optically thick at multiple wavelengths, but its fluctuations and movement cause additional noise to the desired signal. To navigate this, ground-based spectrographs typically observe shorter wavelengths, such as the optical and UV, and between water absorption bands in the infrared. They also benefit from being built at high altitudes and in dry areas (e.g. Mauna Kea in Hawaii and Cerro Paranal in Chile).

1.5 Exoplanet Characterisation: Theory and Analysis

When creating a physical model of an atmosphere, several components need to be taken into consideration. A summary of those will be presented in this section.

1.5.1 Atmospheric Chemistry

One of the primary aspects of an atmospheric model is the inclusion of molecular and atomic species, and therefore the chemistry governing their abundances. There are multiple options for the chemical model used by atmospheric theorists, which vary in complexity. These include free chemistry, equilibrium chemistry, and more sophisticated disequilibrium effects such as atmospheric mixing and photochemistry.

Free chemistry is essentially the absence of any chemical model at all. In this scenario, chemical abundances are allowed to take any value, giving the model complete freedom to adjust them with little or no constraints. When fitting a model to a spectrum, this allows the model to be entirely driven by the data, without relying on any chemical assumptions. The disadvantage of this is that unphysical situations commonly arise in the models (Heng and Lyons, 2016), and any noise or outliers in the data can cause unrealistic values in the abundances. Free chemistry is most frequently used in a data-driven analysis approach known as atmospheric retrieval (see Section 1.5.6), which benefits from simplified physics due to computational restrictions.

Equilibrium chemistry models determine the molecular abundances for a given temperature and pressure by minimising the Gibbs free energy of the system. Since they do not require a knowledge of the chemical pathways, these models can contain a large number of species with different phases (Van Zeggeren and Storey, 1970). A typical atmospheric equilibrium chemistry model will use the elemental abundances to calculate the chemical composition at each level in the atmosphere (e.g. Stock et al., 2018). Such a model can typically contain hundreds or even thousands of chemical species. Alternatively, for very simple systems approximate analytical equations can be used to give sufficiently accurate estimates of the abundances (Heng and Tsai, 2016). Figure 1.8 shows the molecular abundances for a few species in chemical equilibrium, as a function of the carbon-to-oxygen ratio

(Madhusudhan, 2012). Equilibrium models are a powerful tool as they can teach us about the expected composition of hot exoplanets, and importantly, highlight when surprising abundances are detected. However, they can be highly restrictive, and cooler exoplanets, such as warm Neptunes, are unlikely to be in chemical equilibrium (e.g. Stevenson et al., 2010; Moses et al., 2013).
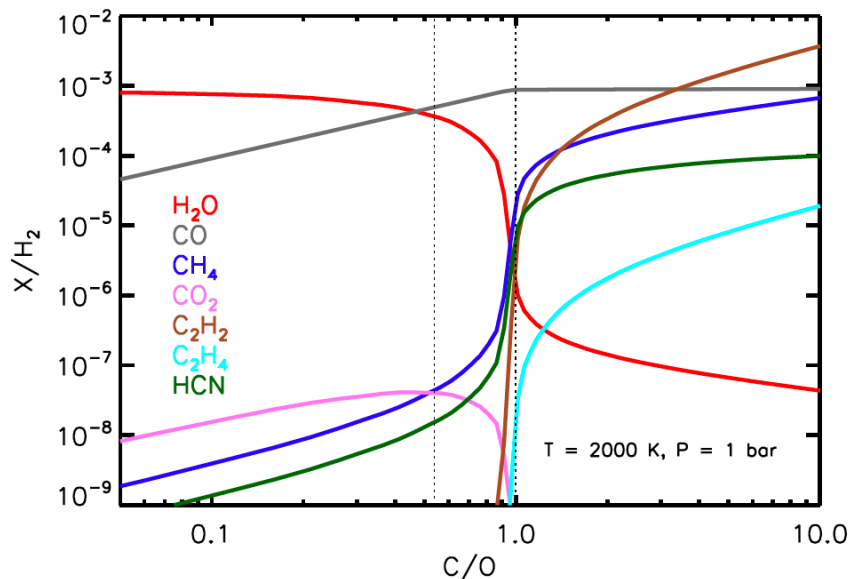
Figure 1.8: Volume mixing ratios of key carbon- and oxygen-carriers in chemical equilibrium, as a function of the C/O ratio. This model is calculated at a temperature of 2000 K and a pressure of 1 bar. Figure taken from Madhusudhan (2012).

Disequilibrium chemistry models can contain a number of different chemical processes which govern the abundances in different regions of the atmosphere. The two most important effects are vertical mixing and photo-dissociation. To account for these, a set of differential equations must be solved to obtain the number densities of each species. Solving these for a full chemical network is a computationally intensive task, so these effects are often approximated (e.g. Tsai et al., 2018). If photochemistry is excluded, timescale arguments can be used to model the effects of vertical mixing. Even with these approximations, disequilibrium chemistry is challenging to model efficiently, and is therefore rarely used inside a full atmospheric model for exoplanets. However, a number of disequilibrium chemistry models exist (e.g. Moses et al., 2011; Venot et al., 2012; Rimmer and Helling, 2016; Tsai et al., 2017), and provide extremely useful information about expected molecular abundances in chemically diverse exoplanet atmospheres.

1.5.2 Radiative Transfer and Temperature Structure

In order to model a spectrum of an exoplanet, one needs to understand how radiation propagates through the atmosphere. This process is known as radiative transfer. Based on the radiation fluxes as a result of solving the radiative transfer equation, we can derive other atmospheric key quantities, such as the temperature-pressure profile, which dictates the temperature at each vertical level of the atmosphere. As with atmospheric chemistry, there are several options for how to model radiative transfer, ranging in complexity. This choice is typically governed by the goal of the model.

If one is aiming to model a theoretical atmosphere of an exoplanet (see Section 1.5.5), with some known or chosen initial conditions, this usually requires a full radiative transfer calculation. In this situation, one models the propagation of radiation in the atmosphere. This involves solving a set of differential equations, with a key variable being τ, the optical depth of the atmosphere. This quantity loosely describes how far a photon will travel before being absorbed or scattered. If a medium is optically thin ($\tau \ll 1$) then interactions with matter are rare, whereas if a medium is optically thick ($\tau \gg 1$) then these interactions are frequent, and a photon is quickly absorbed or scattered. Several approximations can be made to speed up the computation of a radiative transfer model, such as the two-stream approximation (Heng et al., 2014). There are a number of radiative transfer models, or self-consistent atmospheric models containing radiative transfer calculations, in the community (e.g. Irwin et al., 2008; Tremblin et al., 2015; Malik et al., 2017; Piskorz et al., 2018).

If instead one is aiming to determine the temperature structure based on the measure spectrum of an atmosphere, this only requires some parametrisation of the temperature-pressure profile. Similar to the chemistry models, these parametrisations suffer from a trade-off between freedom to fit the data and physical plausibility. One of the first uses of a parametrised T-P profile in exoplanet atmospheres was Madhusudhan and Seager (2009), which uses three layers, each with an exponential profile, shown in Figure 1.9. This results in a 6-parameter model profile. Alternative parametrisations have also been used in atmospheric modelling (e.g. Line et al., 2015; Kitzmann et al., 2020). These parametrised models have the advantage of being data-driven, but can often lead to inconsistent solutions.
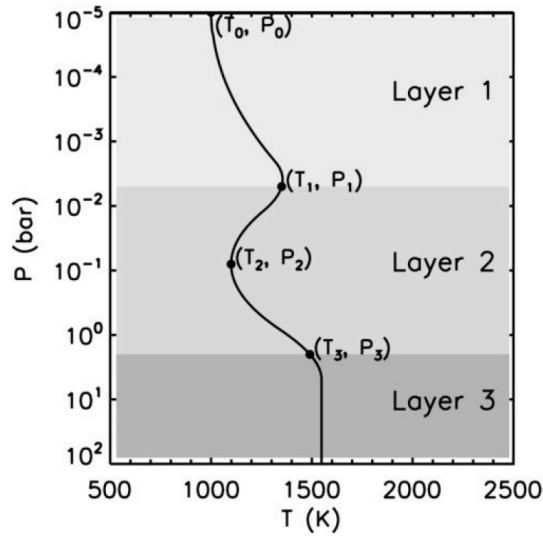
Figure 1.9: Parametrised temperature-pressure profile from Madhusudhan and Seager (2009).

The full radiative transfer calculation and the parametrised T-P profiles represent two extremes in atmospheric modelling. Alternatively, there are semi-analytic formulisms that provide adjustable T-P profiles dictated by the physics (Guillot, 2010; Heng et al., 2012, 2014). Some models take advantage of multiple methods, such as Waldmann et al. (2015b), which uses an analytical T-P profile to set the overall shape, and then adjusts the temperature in each layer to fit the data.

The use of a temperature-pressure profile allows one to model the variable temperature through the height of a planet's atmosphere. One can also assume the temperature remains constant throughout the atmosphere, using an isothermal profile. This is, for example, often done when modelling transmission spectra. However, exoplanets are three-dimensional objects with known temperature variations across all directions. For example, a tidally locked exoplanet, which has one side facing the star at all times, will have a very different temperature profile on its day-side versus its night-side. Recent studies have investigated potential issues associated with only considering one T-P profile in an atmospheric model (e.g. Feng et al., 2016; Taylor et al., 2020). Their results show that multiple profiles will need to be included to fit the precision and sensitivity of the exoplanet spectra predicted from JWST.

1.5.3 Opacities

The unique absorption fingerprint associated with each atom and molecule across different wavelengths is known as its absorption cross-section per unit mass, or opacity. These opacities make up the fundamental building blocks of any spectral model. Figure 1.10 shows the opacities of various molecules and their combined extinction (Malik et al., 2017). The importance of opacities in the accurate characterisation of exoplanet atmospheres is crucial. There are a number of teams calculating the positions and strengths of molecular and atomic transition lines (known as line-lists), and their work supports atmospheric modellers all over the world. The main databases include ExoMol (Tennyson and Yurchenko, 2012; Tennyson et al., 2020), HITRAN (Rothman et al., 2013; Gordon et al., 2017), HITEMP (Rothman et al., 2010), Kurucz (Kurucz, 2017), and NIST (Kramida et al., 2019), among others. The line-lists need to be converted into opacities, which requires a great deal of calculations due to the vast number of transition lines that need to be processed (up to $\sim 10^{10} - 10^{11}$). This can be done inside an atmospheric model, or using open-source opacity calculators and databases (e.g. Grimm and Heng, 2015; Grimm et al., 2021; Chubb et al., 2021).
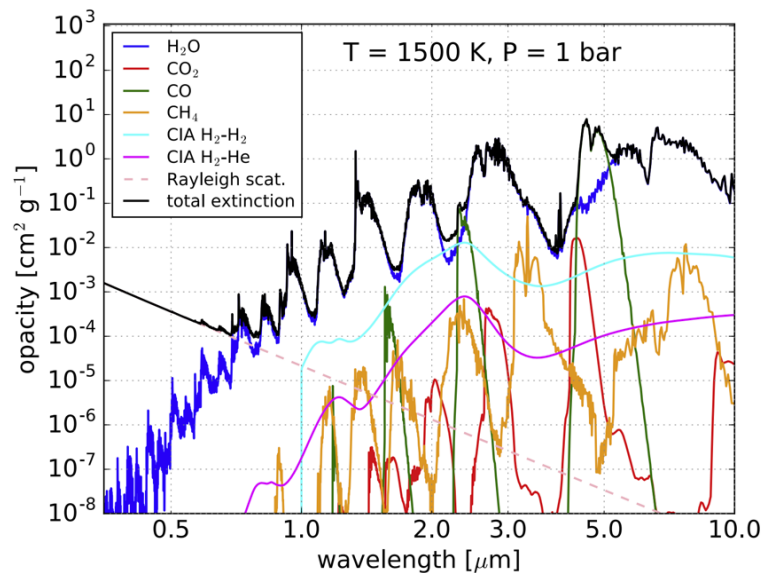
Figure 1.10: Opacities of various molecules at a temperature of 1500 K and a pressure of 1 bar, weighted by some mixing ratios (Malik et al., 2017). The combined extinction is shown in black.

Opacities are calculated at a very high spectral resolution, which can then be binned down for data analysis. In the case of low-resolution spectra, the attention is on the wider, more general shape and height of the spectral features, which can inform us of the molecular abundances and temperature of the planet's atmosphere. This requires the accuracy of the opacities to be focused on the strength of the spectral lines, rather then their specific locations. In contrast, for high-resolution data the individual spectral lines can be resolved, meaning their exact location in wavelength space is the important quantity in the line-lists. Figure 1.11 shows the high- and low-resolution opacities for H_2O, CO_2, and CO (Grimm et al., 2021).
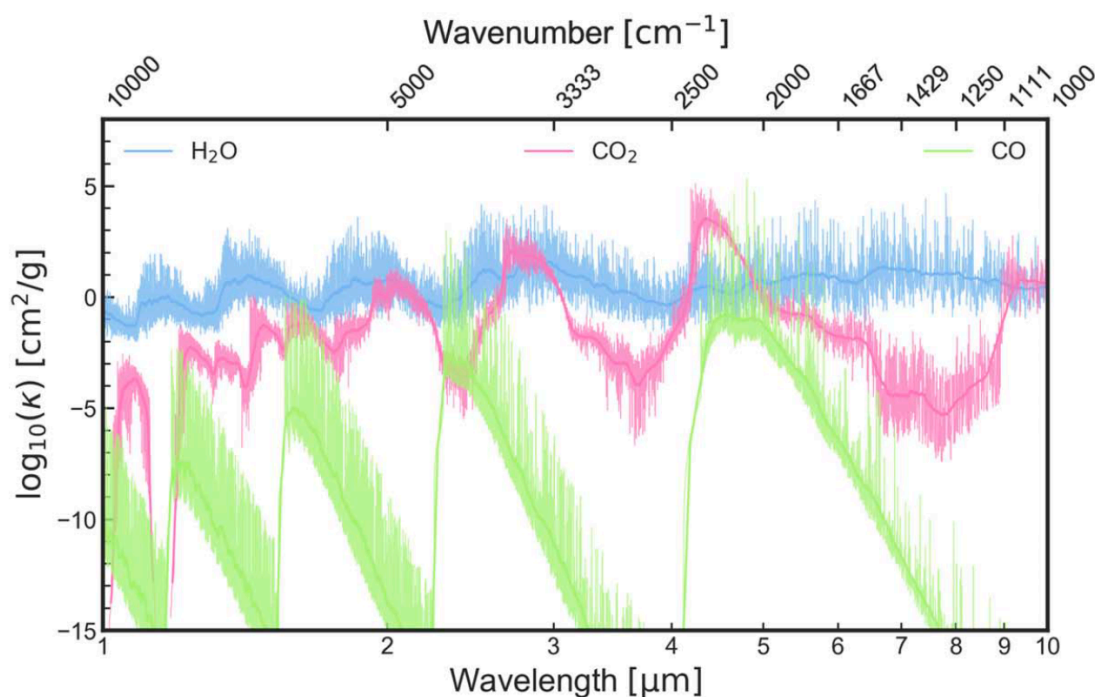
Figure 1.11: Opacities of H_2O, CO_2, and CO, at a temperature of 1500 K and a pressure of 1 mbar (Grimm et al., 2021). The lighter and darker lines show the high- and low-resolution opacities, respectively.

1.5.4 Clouds and Hazes

Ever since the first detection of an exoplanet around a sun-like star, astronomers have predicted the existence of condensates or aerosols on these distant worlds (Guillot et al., 1996; Saumon et al., 1996). The effect of aerosols on an exoplanet's spectrum typically presents as a weakening or complete obstruction of absorption

features. The first measured exoplanet transmission spectrum showed evidence for aerosols based on the obscured wings of the sodium absorption lines (Charbonneau et al., 2002; Fortney et al., 2003). It is now accepted that aerosols are common in the atmospheres of exoplanets, and their inclusion in atmospheric models is essential.

The classification of clouds versus hazes in the exoplanet community is not well-defined, with various distinctions based on their formation processes, particle sizes, or effects on the planet's spectrum. Figure 1.12 shows an example of an optical transmission spectrum with evidence of a Rayleigh-scattering haze, i.e. scattering, sub-micron particles with sizes much smaller than the wavelength (Sing et al., 2011; Fortney et al., 2010; Marley et al., 2013).
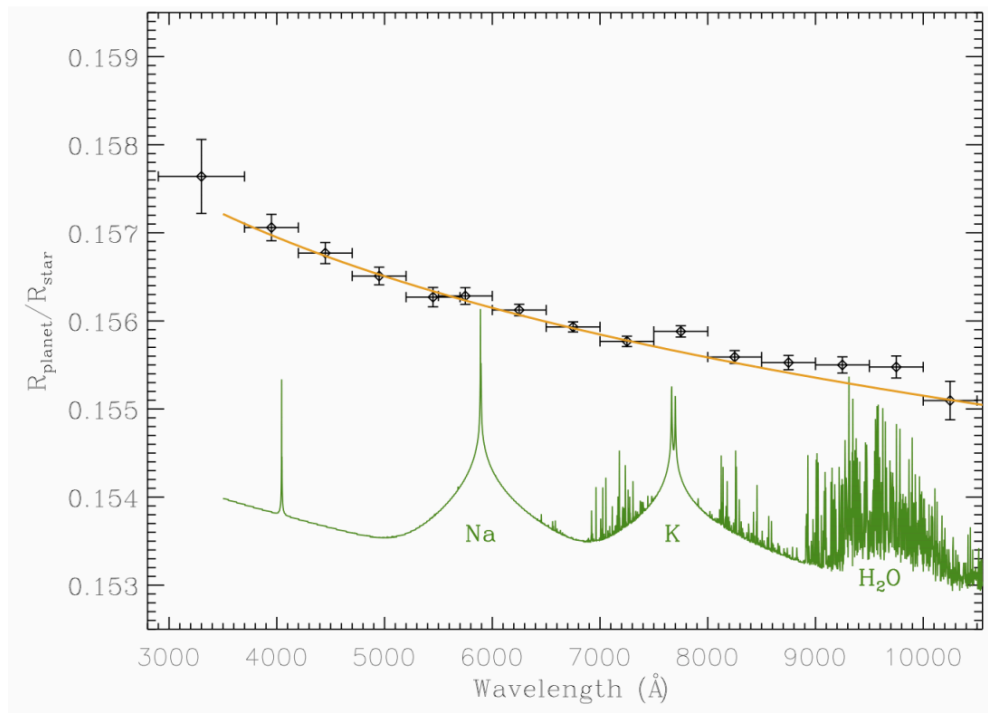
Figure 1.12: Transmission spectrum of HD 189733 b in the optical. The black points show the observed data (from Sing et al. (2011)). The green line shows a model with gaseous absorption only (Fortney et al., 2010). The orange line shows a model with a Rayleigh scattering haze. Figure from Marley et al. (2013).

Again, there are various options for how to model clouds and hazes in exoplanet spectra. In a fully self-consistent model, complex cloud formation processes are included, which can depend on the temperature and chemistry of the atmo-

sphere (e.g. Helling and Woitke, 2006; Helling et al., 2008, 2017). Alternatively, these processes can be approximated in simpler models (e.g. Ackerman and Marley, 2001), though their predictive power is limited. In a data-driven approach, formation modelling is replaced by a parametrisation of the existing cloud. This can be determined by the opacity of the cloud (e.g. Fisher and Heng, 2018), the cloud-top pressure (e.g. Tsiaras et al., 2018), the cloud composition (e.g. Kitzmann et al., 2018), or some combination of these. Comparisons across multiple different parametrisations can even give deeper insights into the cloud structure on a planet (Barstow, 2020). Hazes are often included in atmospheric models as a wavelength-dependent opacity source, leading to slopes in the exoplanet spectra (e.g. Figure 1.12).

1.5.5 Forward Modelling

As previously mentioned, one aspect of analysing exoplanet atmospheres is to use a physics-driven approach to create a fully self-consistent model. These models only depend on the fundamental atmospheric parameters, such as the surface gravity, element abundances, or orbital distance. The atmospheric structure, e.g. the T-P profile or chemical composition, and the spectrum are then an outcome of the model. This approach is known as forward modelling. These models typically contain a radiative transfer model or temperature-pressure profile, a chemistry model containing equilibrium or disequilibrium processes, and a cloud or haze model. An advantage of forward modelling is that one can determine the unique effects of each model parameter on the resulting spectrum. Figure 1.13 shows an example of several transmission spectrum models with different temperature values.

Forward models from different groups can be used to create grids of exoplanet or brown dwarf spectra, which can be used for theoretical analysis or comparison with data (Allard et al., 2001; Marley et al., 2017; Goyal et al., 2018, 2019, 2020). Forward models can also be built into retrieval codes (see Section 1.5.6), although for this the models tend to require a limit on complexity due to computational restrictions.
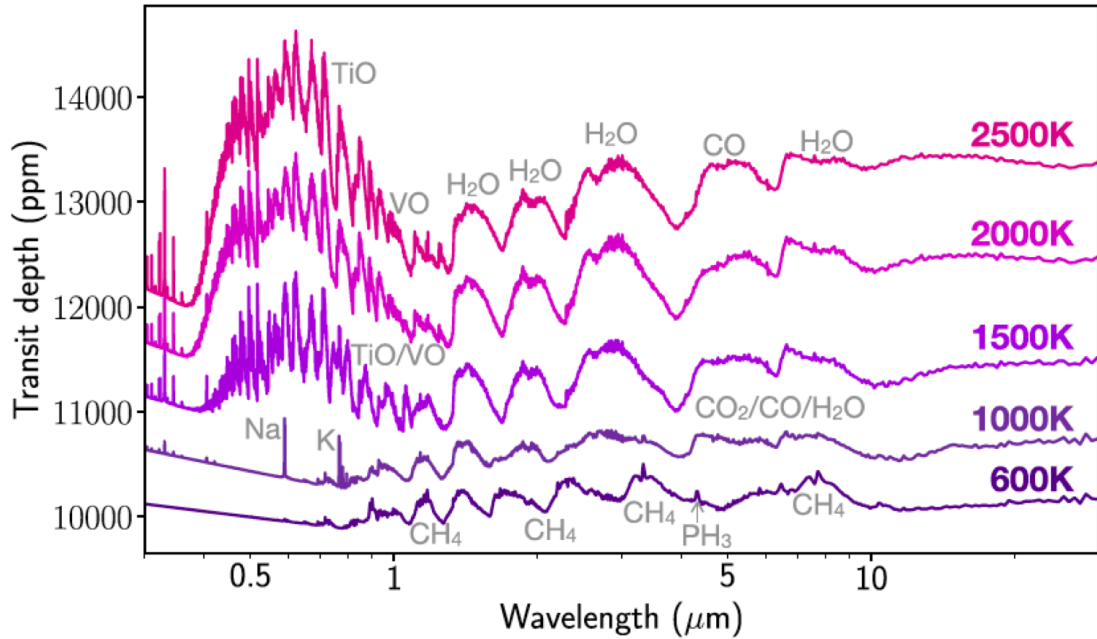
Figure 1.13: Forward models of transmission spectra with varying equilibrium temperature (Goyal et al., 2019).

1.5.6 Atmospheric Retrieval

In contrast to forward modelling, the data-driven approach to analysing exoplanet atmospheres is known as atmospheric retrieval. This involves running an atmospheric model multiple times (usually $\sim 10^4$ times), and using a sampling algorithm to compare the model to the data. The model parameters that produce the spectra which best fit the data are then combined into a posterior distribution. These distributions can also highlight any relationships or degeneracies between the model parameters. The model can either be run in advance, to create a grid of models for the data comparison, or on-the-fly, using more complex sampling algorithms to inform the model on which parameter regions to focus on.

The concept of atmospheric retrieval originates from remote sensing of the Earth's atmosphere (Rodgers, 2000), and has also been used in Solar System studies (Irwin et al., 2008). The first study to use atmospheric retrieval for exoplanets was Madhusudhan and Seager (2009). From here, the field of retrievals blossomed, and is now a major division of exoplanet analysis (see Madhusudhan (2018) and Barstow and Heng (2020) for reviews of the field and its challenges).

As discussed in previous sections, the model used in retrievals is made up of several key components of atmospheric physics. However, the need to run this model $\sim 10^4$ times in a single analysis requires that the model is not too computationally expensive. Therefore, retrieval models typically involve approximations and parametrisations that speed up the computation. One of the challenges of atmospheric retrieval is to determine which approximations can be made without biasing the results. This depends heavily on the resolution and wavelength coverage of the data being analysed. Low-resolution transmission spectra from HST, for example, can be fit with relatively simple models (e.g Heng and Kitzmann, 2017), while the data we predict from JWST is likely to require more physically accurate models (Feng et al., 2016; Taylor et al., 2020). Figure 1.14 shows a comparison of transmission spectra using a full numerical calculation and an analytical approximation, over both JWST-NIRSpec and HST-WFC3 wavelength ranges (Heng and Kitzmann, 2017). This demonstrates a good agreement between the models in the WFC3 range, within the expected error bars for this instrument, but a divergence between the models at longer wavelengths where James Webb will be sensitive.
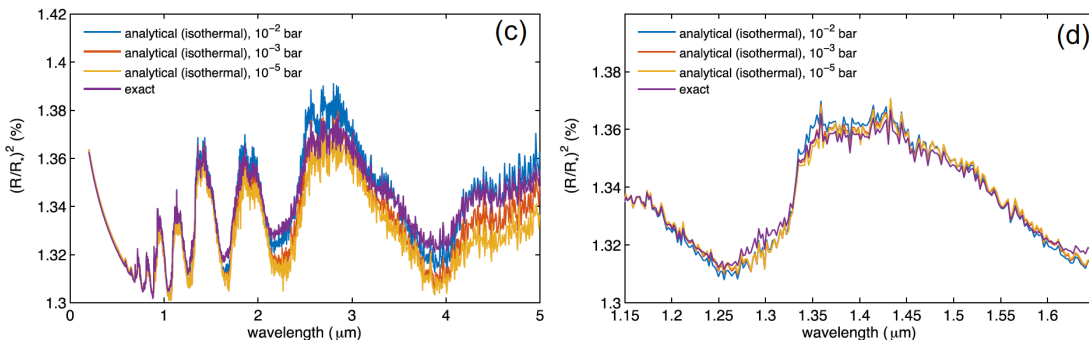
Figure 1.14: Comparison of full numerical calculations of the transmission spectrum of an isothermal atmosphere to those computed using an isothermal, isobaric analytical formula (Heng and Kitzmann, 2017). This calculation is for a temperature of 1000 K. The left- and right-hand panels show JWST-NIRSpec and HST-WFC3 wavelength coverage, respectively.

The other key aspect in a retrieval code is the choice of sampling algorithm. Over the years this has developed from a simple reduced-χ^2 on a grid of models (e.g. Madhusudhan and Seager, 2009), to optimal estimation (e.g. Lee et al., 2012), to more advanced Bayesian inference methods such as Markov Chain Monte Carlo (MCMC) (e.g. Todorov et al., 2016; Wakeford et al., 2017; Evans et al., 2017) and Nested-Sampling (e.g. Benneke and Seager, 2013; Waldmann et al., 2015a; Lavie et al., 2017; MacDonald and Madhusudhan, 2017; Fisher and Heng, 2018). The

latter two use a likelihood function to measure the distance between the model and the data. Open-source algorithms such as `emcee` (Foreman-Mackey et al., 2013) for MCMCs, and MultiNest (Feroz and Hobson, 2008; Feroz et al., 2009) and PyMultiNest (Buchner et al., 2014) for Nested-Sampling, make it relatively straight-forward to build a retrieval code once you have a sufficiently fast atmospheric model.

There are a number of traditional exoplanet retrieval codes in the community, both proprietary and open-source. Examples include NEMESIS (Irwin et al., 2008; Lee et al., 2012; Barstow et al., 2013), CHIMERA (Line et al., 2013, 2014), τ-Rex (Waldmann et al., 2015a,b), ATMO (Evans et al., 2017; Wakeford et al., 2017), Helios-R (Lavie et al., 2017), petitRADTRANS (Mollière et al., 2017), POSEI-DON (MacDonald and Madhusudhan, 2017), Helios-T (Fisher and Heng, 2018), and Helios-r2 (Kitzmann et al., 2020). Each code relies on different assumptions and implementations of the physics, as well as different parameter-sampling techniques. Comparison studies such as Barstow et al. (2020) have investigated these differences, and the effects on the retrieved posterior distributions.

Over the years, several important discoveries have come from atmospheric retrievals. These include the detection of various molecules in hot Jupiters and Neptunes, such as H_2O, CH_4, Na, and K (e.g. Fraine et al., 2014; Sing et al., 2016; Wakeford et al., 2017), the detection of water in the habitable zone sub-Neptune K2-18 b (Benneke et al., 2019b; Tsiaras et al., 2019), and the prevalence of clouds on exoplanets (e.g. Deming et al., 2013; Kreidberg et al., 2014a).

Before using an atmospheric retrieval code on real data, a crucial test should be done to check if known parameter values can be retrieved from a simulated spectrum. This is known as a mock retrieval. Figure 1.15 shows an example of a mock retrieval of a brown dwarf spectrum using Helios-r2 (Kitzmann et al., 2020). By comparing the posterior distributions with the true parameter values, one can assess the retrieval code's capabilities for specific types of data.

Another potential insight to be gained from atmospheric retrievals is an investigation of any parameter trends across multiple exoplanets. In Fisher and Heng (2018) I searched for trends across 38 planets in the water abundances, temperatures, masses, and cloud levels, but found no evidence (see Chapter 2). Other studies have considered a possible relationship between the mass and metallicity
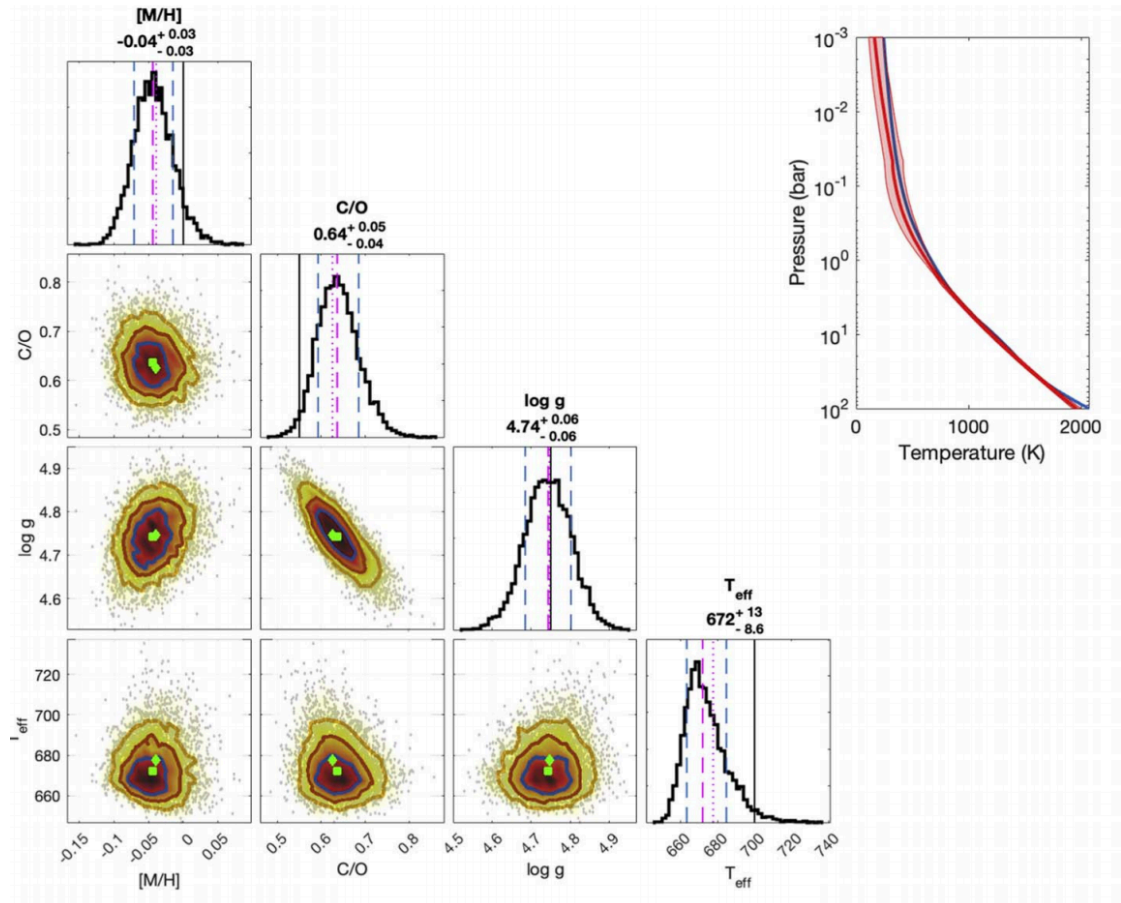
Figure 1.15: Mock retrieval of a simulated brown dwarf spectrum using Helios-r2 (Kitzmann et al., 2020). The true parameter values are shown as solid black lines on the posterior distributions. The true temperature-pressure profile is shown in blue in the top right panel.

of giant exoplanets (e.g. Wakeford et al., 2017; Mansfield et al., 2018; Welbanks et al., 2019; Blain et al., 2021), but this is still uncertain.

The majority of atmospheric retrievals to-date have been performed on low-resolution data, from instruments such as HST's WFC3 and STIS, and Spitzer. However, these low-resolution retrievals are currently only able to constrain molecular abundances to within 1 or more dex, and they often feature major degeneracies between parameters. We predict a great leap in constraining power when James Webb comes online, but in the meantime several studies have looked at performing atmospheric retrieval on high-resolution ground-based data. The biggest challenges in this regard are the large number of spectral points and the high level of noise on each of them. Brogi et al. (2017) and Brogi and Line (2019) were the first to overcome these challenges. Their technique uses a mapping for the likelihood that takes advantage of the cross-correlation function (see Section 1.5.7), which averages information across many spectral points. However, this method is limited to narrowband spectra. Inspired by their work, I developed my own method of high-resolution retrievals (see Chapter 5), which extends to wideband spectra containing $\sim 10^5$ points. Other recent studies have also performed atmospheric retrievals on high-resolution data (e.g. Gibson et al., 2020; Pino et al., 2020).

1.5.7 Cross-correlation Technique

A major breakthrough in the analysis of high-resolution ground-based exoplanet spectra came with the introduction of the cross-correlation technique (Snellen et al., 2008, 2010). This technique uses the cross-correlation function (CCF) to take an average of all the spectral points, weighted by the strengths of the absorption lines of a particular molecular or atomic species. This has the effect of reducing the noise in the spectral lines. The weighting for each species is known as a template, and this template is shifted to different radial velocities at which the CCF is calculated. If a peak in the CCF occurs at the known radial velocity of the exoplanet, this indicates a detection of that species. Slight deviations from the planet's radial velocity can also indicate strong day-to-night winds or heat circulation in the planet's atmosphere (e.g. Snellen et al., 2010; Miller-Ricci Kempton and Rauscher, 2012; Showman et al., 2013; Brogi et al., 2016; Zhang et al., 2017). Figure 1.16 shows an example of the CCFs for CO, H_2O, CH_4, and NH_3, for the emission spectrum of β Pictoris b (Hoeijmakers et al., 2018b). This shows clear

detections of CO and H_2O, and non-detections of CH_4 and NH_3 in this planet's atmosphere.

There are several ways of constructing the cross-correlation function. A commonly used example is a simple template-weighted average,

$$c(v,t) = \sum_{i=0}^{N} x_i(t) T_i(v),$$ (1.5)

where x_i are each of the N points in the spectrum obtained at time t, and $T_i(v)$ are the values of the template that is Doppler shifted to a radial velocity v (Baranne et al., 1996; Pepe et al., 2002; Allart et al., 2017; Hoeijmakers et al., 2019). Using this, one can construct a two-dimensional CCF, where different exposures provide a time-evolution of the spectrum. This is shown in the top row of Figure 1.17, which shows the CCFs for various metals in the atmosphere of the ultra-hot Jupiter KELT-9 b (Hoeijmakers et al., 2018a). The diagonal white lines, clearest for Fe+ and Ti+, track the movement of the planet as it passes in front of the star. The dotted lines therefore indicate the start and end of the transit. These different exposures can be co-added in the rest frame of the planet, giving the plots in the middle row of Figure 1.17. Extracting these at the orbital velocity of the planet (254 km s^{-1}), one can construct the one-dimensional cross-correlation
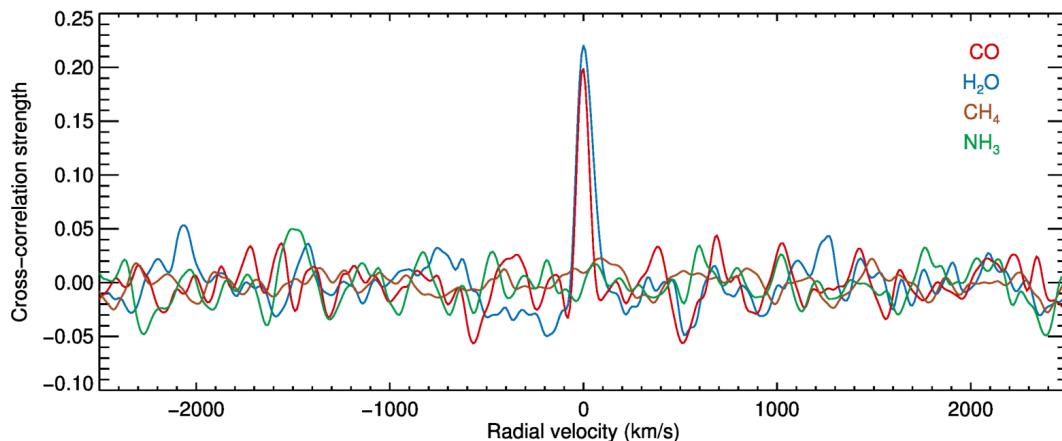
Figure 1.16: One-dimensional cross-correlation function for 4 different species for the emission spectrum of the exoplanet β Pictoris b, observed with the SINFONI integral field spectrograph on the VLT (Hoeijmakers et al., 2018b).

signal-to-noise ratio, shown in the bottom row of Figure 1.17. These plots indicate a detection of Fe, Fe+, and Ti+, but a non-detection of neutral Ti.
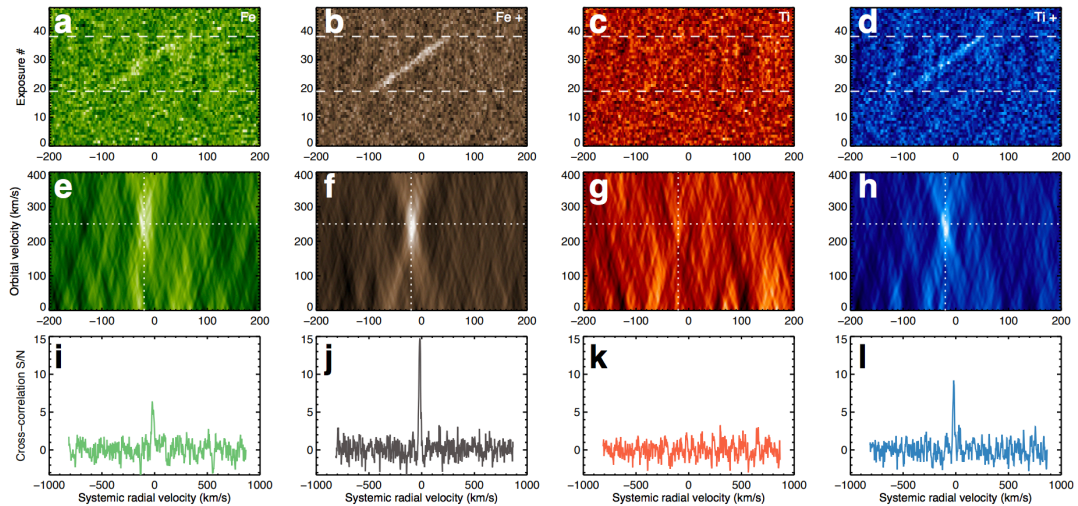
Figure 1.17: Cross-correlations of Fe, Fe+, Ti, and Ti+ with the transmission spectrum of the ultra-hot Jupiter KELT-9 b, observed with HARPS-N (Hoeijmakers et al., 2018a).

Since its first use on exoplanet spectra (Snellen et al., 2010), the cross-correlation technique has been used to detect a vast array of different atoms and molecules. Key examples of molecules include carbon monoxide in τ Boötis b (Brogi et al., 2012) and HD 189733 b (de Kok et al., 2013), and water in HD 189733 b (Birkby et al., 2013) and 51 Peg b (Birkby et al., 2017). More recently, astronomers have taken advantage of the abundance of metal absorption lines in the optical wavelength range to increase our variety of detections. This began with the groundbreaking detection of iron and titanium in KELT-9 b (Hoeijmakers et al., 2018a), and quickly expanded to include magnesium, cromium, scandium, and yttrium in this ultra-hot Jupiter (Hoeijmakers et al., 2019). Additional detections of metals using CCFs have been found in other planets such as MASCARA-2 b (e.g. Casasayas-Barris et al., 2018, 2019; Stangret et al., 2020; Nugroho et al., 2020; Hoeijmakers et al., 2020a), WASP-121 b (Gibson et al., 2020; Hoeijmakers et al., 2020b; Ben-Yami et al., 2020), and WASP-76 b (Ehrenreich et al., 2020; Tabernero et al., 2021).

The ability to detect a much wider array of atoms and molecules make CCFs an extremely valuable tool for analysing exoplanet atmospheres. However, aside from measuring orbital velocities and revealing the presence of extreme winds,

CCFs have been restricted to detecting species in a binary sense. Recent studies have started to push the limits of CCFs, trying to gain additional information by using them in atmospheric retrievals (e.g. Brogi et al., 2017; Brogi and Line, 2019; Fisher et al., 2020). In my work (see Chapter 5), I use machine learning to find the connection between the shape or height of the CCF peak and the properties of the atmosphere, such as the temperature and metallicity.

1.5.8 Machine Learning

In the past few decades, the use of machine learning has exploded in all areas of scientific research. Its ability to vastly improve the speed of data analysis, as well as provide additional insights, makes machine learning ideal for a wide variety of tasks, spanning across all fields. Its use in both exoplanet detection and characterisation has gained popularity in recent years.

For detections, machine learning is an ideal tool for recognising the characteristic shape of a planet transit in a dataset. Thanks to the Kepler mission, the number of planet candidates increased exponentially, motivating a more automated analysis than human judgement. Several studies turned to machine learning algorithms, with varying degrees of complexity. McCauliff et al. (2015) used a Random Forest to classify candidates based on features derived from the Kepler pipeline. Additional studies have also used Random Forests for exoplanet detection (e.g. Mislis et al., 2016; Armstrong et al., 2018; Schanche et al., 2019; Caceres et al., 2019). Thompson et al. (2015) and Armstrong et al. (2017) used unsupervised machine learning to find clusters of candidates with similar light curves, classifying them using known planets or false positives. Machine learning has also been used to identify different signals, such as *non*-transiting planets (Millholland and Laughlin, 2017). The application of neural networks to planet detections has also become increasingly popular. Dittmann et al. (2017) used them to identify real transits in the MEarth dataset, and Pearson et al. (2018) tested them on simulated light curves. Perhaps one of the most successful applications was by Shallue and Vanderburg (2018), in a collaboration with Google. They used a deep convolutional neural network (CNN) to identify a fifth planet around Kepler-80, and an eighth planet around Kepler-90, making it the largest exoplanet system known to-date. CNNs have now been used in a number of exoplanet detection studies (e.g. Ansdell et al., 2018; Dattilo et al., 2019; Chaushev et al., 2019; Yu et al., 2019; Osborn

et al., 2020). Alternative machine learning algorithms, such as Gaussian Processes, have also been used in exoplanet detection (e.g. Armstrong et al., 2021) and for removing stellar activity (e.g. Aigrain et al., 2016).

In contrast, machine learning is still used relatively sparsely in exoplanet characterisation. The first study were Waldmann et al. (2015b) and Waldmann (2016), which used a deep belief neural network to identify the molecules present in an exoplanet emission spectrum. These molecules could then be included in a traditional retrieval to find the abundances, and other atmospheric properties. Following this, I co-developed the use of Random Forests to perform the retrievals themselves, replacing the sampling algorithm (Márquez-Neila et al., 2018) (see Chapter 3). A Random Forest is a supervised machine learning method, meaning the output labels for the training set are provided and the forest learns a mapping between the input and output. This is ideal for exoplanet retrievals, where the goal is to link an observed spectrum to a set of atmospheric parameters. A Random Forest uses a collection of regression trees trained on randomly drawn subsets of the training set. After testing other machine learning techniques, such as neural networks, we settled on the Random Forest due to its transparency, speed, and ability to produce posterior distributions. This technique has since been applied in multiple studies by our group (Oreshenko et al., 2020; Fisher et al., 2020; Guzmán-Mesa et al., 2020) and others (Nixon and Madhusudhan, 2020). Other machine learning techniques have also since been used in retrievals, such as Zingales and Waldmann (2018), who used a Generative Adversarial Network (GAN), and Cobb et al. (2019), who used a Bayesian neural network. In an alternative approach, Hayes et al. (2020) used Principle Component Analysis to inform the priors of a traditional retrieval. Additional studies have also used convolutional networks to predict chemical abundances from reflected light spectra (Soboczenski et al., 2018), and multilayer perceptrons for quick parameter estimations from geometric albedo spectra (Johnsen et al., 2020).

The majority of studies using machine learning are motivated by the need for an increase in the speed and efficiency of data analysis. However, additional insights into the data or model are often provided by these techniques with little to no extra effort. For example, all machine learning algorithms are tested on large datasets, allowing one to analyse the predictability of parameters across a wide range. This is unfeasible with traditional retrievals, which take the order of a few minutes to several hours to run a single time. Additionally, one can run these tests on varied datasets to analyse other aspects of exoplanet spectra. Figure 1.18, from Zingales

and Waldmann (2018), shows how the accuracy of the parameter predictions from their GAN vary with the size of the error bars in the data. Specifically for Random Forests, one obtains a "feature importance", which quantifies the importance for each feature (e.g. properties of the light curve, or individual spectral points) in accurately classifying the data. Figure 1.19 shows an example of the feature importances for identifying real planet transits in WASP data (Schanche et al., 2019).
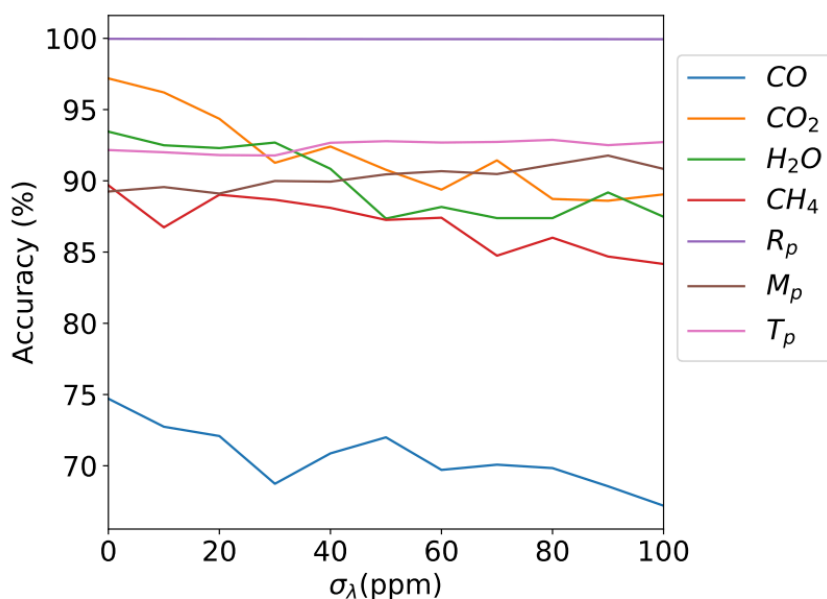
Figure 1.18: Accuracy of the parameter prediction from a GAN, as a function of the spectral error bars (Zingales and Waldmann, 2018).

1.5.9 Life and Habitability

A leading motivation for the detection and characterisation of exoplanets is the search for extra-terrestrial life. By determining the habitability of an exoplanet, we can speculate about its potential to host this alien life. However, that comes with many challenges, and several aspects of habitability are widely disputed among the community.

One of the simplest indicators of a planet's ability to host life is its distance from the star. If the planet orbits too close to its host star, the temperature will
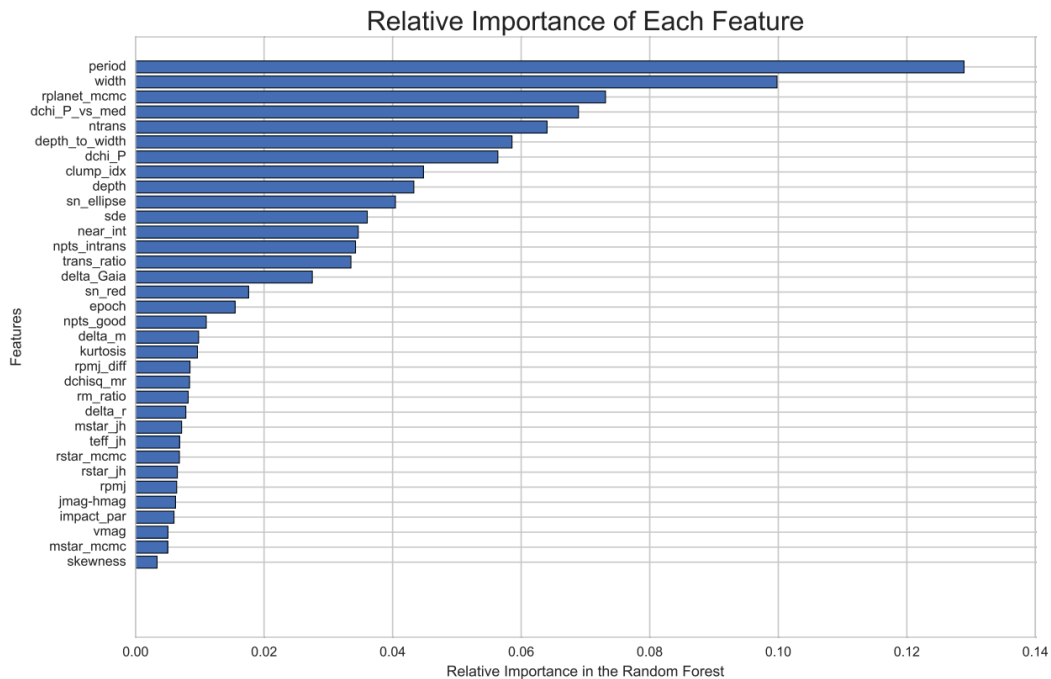
Figure 1.19: Ranked list of the relative feature importance from a Random Forest classifier for identifying real planet transits (Schanche et al., 2019).

increase beyond a level conducive to life. However, a planet orbiting too far out will also be too cold to be habitable. The sweet spot between these points is known as the "habitable zone" (Huang, 1959; Hart, 1978; Kasting et al., 1993). The definition of the habitable zone varies, but is often accepted as the range of orbital distances at which liquid water can exist for an extended period of time on the surface of an Earth-like planet. However, even this definition is unclear, as it contains several Earth-specific assumptions (Ramirez, 2018; Ramirez et al., 2019). The composition and size of the planet's atmosphere will have a big effect on its surface temperature, allowing extensions of the habitable zone in both directions. Figure 1.20, from Seager (2013), shows the traditional habitable zone for Earth-like planets, as well as the extensions for dry terrestrial and hydrogen-rich planets. Other factors can affect the location of the habitable zone, such as the planetary radius and gravity (Yang et al., 2019a), ocean dynamics (Yang et al., 2019b), and rotation rates (Ramirez and Levi, 2018). Despite these caveats, the habitable zone remains a standard factor when considering an exoplanet's potential to host life.
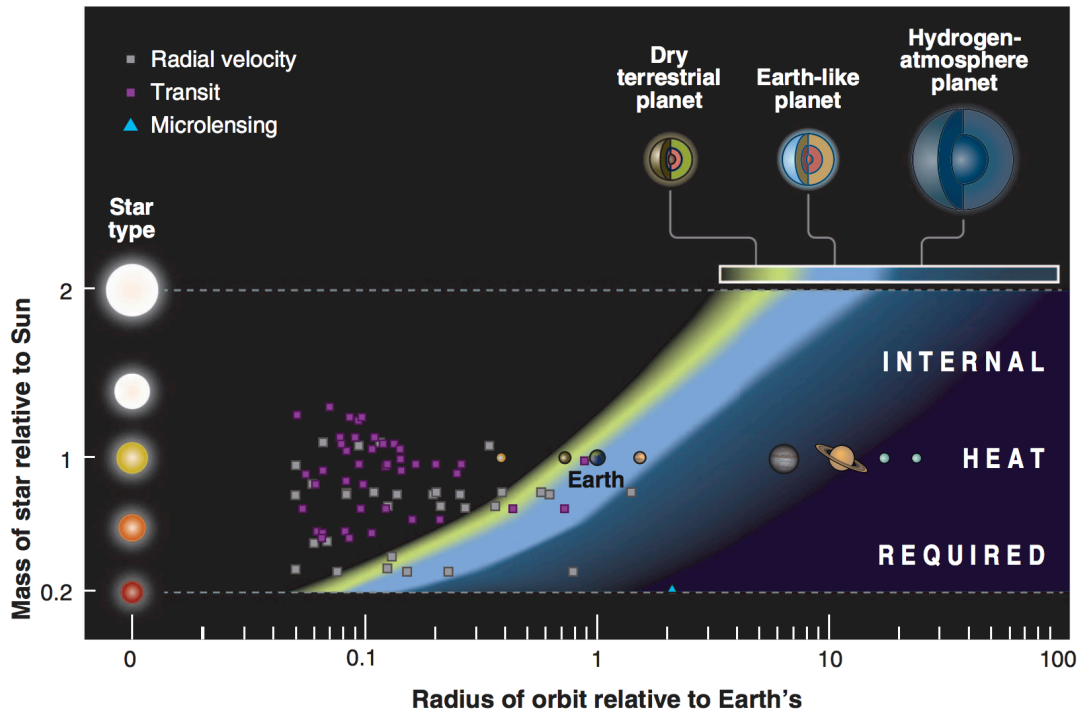
Figure 1.20: The habitable zone, from Seager (2013). The pale blue region shows the traditional habitable zone, for planets with N_2-CO_2-H_2O atmospheres. The yellow region shows the habitable zone extended inwards for dryer planets. The dark blue region shows the outer extension of the habitable zone for planets with hydrogen-rich atmospheres. The solar system planets are shown with images, and some detected exoplanets are shown with symbols.

In contrast to studying a planet's *ability* to host life, astronomers are also searching for indications that a planet *does* host some form of life. The study of possible signals of life on exoplanets is led by the search for "biosignatures". These are gases produced by life's metabolic process that accumulate in the planet's atmosphere at high enough levels to be remotely detected. Biosignatures have been discussed for many years (Lederberg, 1965; Lovelock, 1965), but their detectability remains a challenge for current technology. Several studies have hypothesised about the potential to detect these with upcoming and future observatories (e.g. Schwieterman et al., 2018; Krissansen-Totton et al., 2018b; Lopez-Morales et al., 2019). Again, the definition of a biosignature is somewhat vague as multiple false positives exist, where a gas could also be explained by geological processes, for example. The most favourable biosignatures would be gases produced by life in quantities many orders of magnitude away from chemical equilibrium, with no major abiotic sources. In

addition, we need these biosignatures to have clear, unique spectroscopic features. Seager et al. (2016) set the goal of constructing a list of potential biosignatures for diverse exoplanets, and several studies have looked at individual gases in detail (e.g. Sousa-Silva et al., 2020; Zhan et al., 2021). On Earth, the key current biosignatures would be O_2, O_3, and N_2O. However, it is also important to consider how the Earth has evolved over time, and which biosignatures would have been dominant in different eras (Krissansen-Totton et al., 2018a). Aside from biosignature gases, additional detectable signs of life have been hypothesised, such as biological fluorescence (O'Malley-James and Kaltenegger, 2018, 2019). Studies have also considered the possibility of life existing in the atmospheres of planets without a surface (Seager et al., 2021).

One major issue with the search for life on other planets is that astronomers tend to approach it from a very Earth-centric direction. We search for Earth-sized planets around Sun-like stars, focusing on liquid water and biosignatures we expect from the life we know and understand. However, for many years scientists have speculated about the possibility of completely alien biochemistries, with different building blocks and solvents (e.g. Schulze-Makuch and Irwin, 2006). For example, in contrast to the carbon-based life found on Earth, alternative elements, such as silicon, have been proposed for the basis of extra-terrestrial life (e.g. Petkowski et al., 2020). Despite its plausibility, in the case of completely foreign biochemistry it is unclear which gases would be detectable biosignatures, making this hypothetical life extremely hard to search for. This motivates us to focus on the biology we are familiar with, even though its prevalence in the universe is unknown.

1.6 Looking Forward

The coming years promise to be an exciting time for exoplanet research, due to the next generation of space- and ground-based observatories. Figure 1.21 shows a schematic of the upcoming telescopes relevant for exoplanets.

Firstly, the launch of JWST later this year will be a landmark event. With instruments spanning from 0.6 to 28 μm at varying resolutions, James Webb will provide state-of-the-art observations of many types of exoplanet. Due to several launch delays, the anticipation for JWST has been building for over a decade. An abundance of studies have analysed its predicted capabilities, and the implications
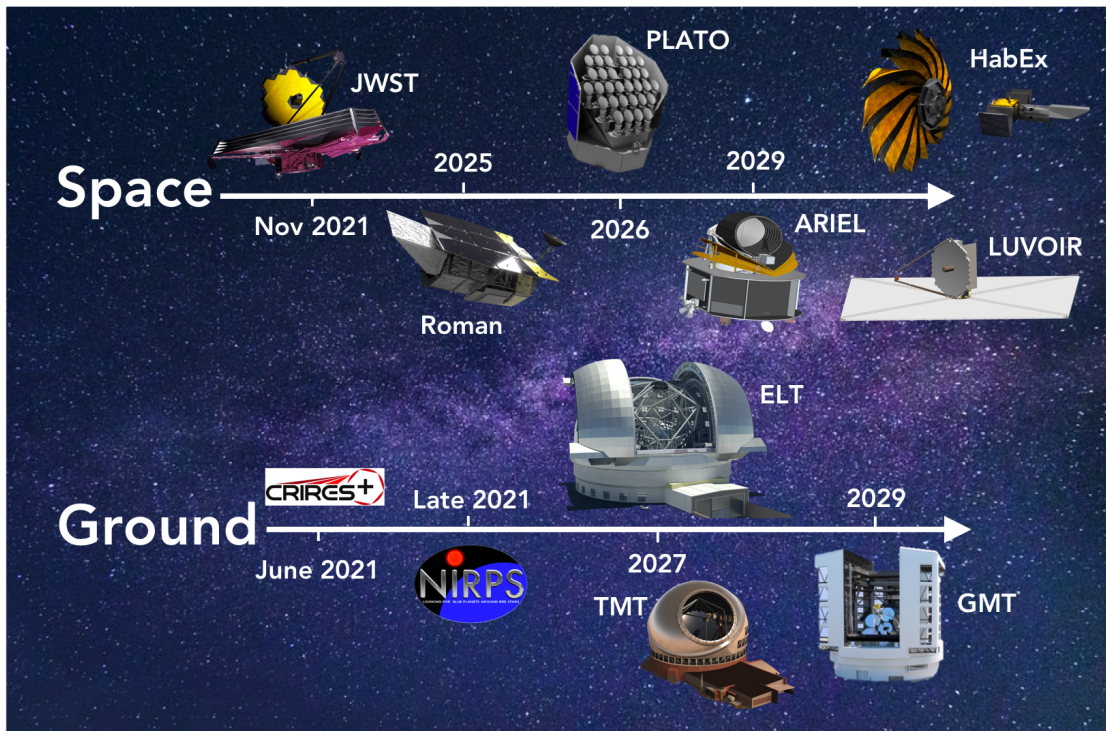
Figure 1.21: Schematic of upcoming and future space- and ground-based telescopes relevant for exoplanet research.

for exoplanet science. For example, Cowan et al. (2015) investigate the possibility of using James Webb to characterise temperature terrestrial planets. Batalha et al. (2015) assess James Webb's potential to characterise the atmospheres of super-Earths, and find that moderate signal-to-noice can be obtained with ~ 25 transits. Greene et al. (2016) explore the constraints of bulk atmospheric properties obtainable from James Webb transmission spectra. Other studies have looked at how to optimise the usage of JWST depending on specific goals in exoplanet characterisation (e.g. Beichman et al., 2014; Batalha et al., 2017; Guzmán-Mesa et al., 2020). However, several papers have discovered issues that could arise when using James Webb, or when analysing the data using the techniques developed for lower resolution spectra. For example, Barstow et al. (2015) use mock retrievals to study the impact of starspots and other systematic errors associated with combining non-simultaneous observations from multiple instruments. Rocchetto et al. (2016) find that the isothermal approximation is not sufficient for analysing JWST transmission spectra, and Feng et al. (2016) and Taylor et al. (2020) show that using a single 1-D thermal profile can bias retrieval results for James Webb emission

spectra. Despite these challenges, there is no doubt that the launch of JWST will bring an abundance of exciting new discoveries in the field of exoplanets.

Following James Webb, a number of space-telescopes are planned that will have important capabilities for exoplanet observations. The Nancy Grace Roman Space Telescope (formally the Wide-Field Infrared Survey Telescope, or WFIRST) will help to detect many more exoplanets using microlensing (Miyazaki et al., 2021). The PLAnetary Transits and Oscillations of stars (PLATO) mission will focus on detecting and characterising rocky, Earth-sized exoplanets in the habitable zone of bright Sun-like stars (Rauer et al., 2014; Grenfell et al., 2020). The Atmospheric Remote-sensing Infrared Exoplanet Large-survey (ARIEL) telescope will observe transits of over 1000 known exoplanets, with the aim of characterising their chemical composition and thermal structure (Tinetti et al., 2018). In addition to these scheduled telescopes, several proposed exoplanet telescopes are in the running for NASA's future Large Strategic Science Missions. These include the Large Ultraviolet Optical Infrared Surveyor (LUVOIR) (The LUVOIR Team, 2019), and the Habitable Exoplanet Imaging Mission (HabEx) (Gaudi et al., 2020), which both have the key science goal of characterising habitable exoplanets.

For ground-based telescopes, a huge leap in the quality of observations is expected with the Extremely Large Telescope (ELT), currently under-development. With an almost 40m aperture, the ELT represents the first in a new era of giant telescopes. It began construction in May 2017, on the Cerro Armazones mountain in the Antofagasta Region in Chile. This location is ideal for astronomy as it is high and dry, with 89% cloudless nights a year. The ELT will have a number of state-of-the-art instruments, including several suited for exoplanet observations. In particular, the HIgh REsolution Spectrograph (HIRES) will have the characterisation of exoplanets as one of its main scientific goals (Marconi et al., 2021). Other instruments with exoplanet capabilities include the High Angular Resolution Monolithic Optical and Near-infrared Integral field spectrograph (HARMONI) (Thatte et al., 2021; Houllé et al., 2021), and the Mid-infrared ELT Imager and Spectrograph (METIS) (Brandl et al., 2021). Closely following the ELT is the construction of the Thirty Meter Telescope (TMT), though its intended construction on Mauna Kea, a dormant volcano regarded as sacred by indigenous Hawaiians, is a point of controversy (e.g. Kahanamoku et al., 2020; Prescod-Weinstein et al., 2020; Witze, 2020). Also under construction is the Giant Magellan Telescope (GMT) in the Atacama region of Chile, another observatory in the generation of extremely large telescopes.

It is worth noting some of the major differences in the development of space- and ground-based observatories. Firstly, there is a large discrepancy between the cost of space-based and ground-based telescopes, due to the expensive launch process and more complex technological challenges. For example, the total cost of JWST has exceeded 10 billion US dollars, whilst the ELT is projected to cost just over 1 billion euros. Secondly, space-telescopes rely on the technology present at the time of construction, and typically cannot be adapted or fixed in situ. This means they must be rigorously tested to check for any issues and ensure they will not be damaged during the launch. This process can take many years, and has contributed to the delay of JWST, now almost 10 years overdue.

The differing budgets required for ground- and space-based observatories has led to a divergence between the two main space agencies, NASA and ESA. The success of pioneering space telescopes, such as HST, have motivated NASA to invest heavily in these, and JWST is predicted to lead the field in space-based observations for the next decade. In contrast, ESA's investment in the ELT will place them at the forefront of ground-based observations. In terms of exoplanet characterisation, this corresponds to a focus on low- and high-resolution spectroscopy for NASA and ESA, respectively. However, studies have shown that the combination of these two data types is essential for building a complete picture of an exoplanet (e.g. Brogi and Line, 2019).

As these next-generation telescopes come online, the abundance of exoplanet spectra ready for characterisation is predicted to increase exponentially. As well as detailed analyses of individual objects, this motivates more statistical studies of the overall population of exoplanets in our galaxy. This is already a popular field of research, with studies using Kepler data to obtain exoplanet occurrence rates (e.g. Petigura et al., 2013; Batalha, 2014; Foreman-Mackey et al., 2014; Sestovic and Demory, 2020), or Hubble transmission spectra to search for trends in atmospheric properties (e.g. Barstow et al., 2017; Fu et al., 2017; Fisher and Heng, 2018; Pinhas et al., 2019). However, the atmospheric studies typically use ~ 10 planets, and larger studies are infeasible with traditional retrievals. In this upcoming era, machine learning retrievals will be imperative for keeping up with the influx of exoplanet spectra, and allowing us to understand the diversity of planets outside our solar system.

Retrieval Analysis of 38 WFC3 Transmission Spectra and Resolution of the Normalization Degeneracy

You should be proud of your achievements and improvement, and that holds regardless of how many marks the examiners choose to give you.

KERRY MAUNDER

2.1 Summary

For my first thesis paper, I performed a retrieval analysis on the Wide-Field Camera 3 (WFC3) transmission spectra of 38 exoplanets, using Bayesian model comparison to determine the conclusions that can be drawn from these low-resolution spectra. 30 of these planets' spectra were taken from Tsiaras et al. (2018), and an additional 4 include Trappist-1 d, e, f, and g from de Wit et al. (2018).

The goal of this paper was to test a hierarchy of models on each planet, and then use a Bayesian comparison to determine the level of model sophistication

that is warranted by the WFC3 data alone. The models range from a flat line, to test a null hypothesis, up to a non-isothermal transmission model including H_2O, HCN, and NH_3, and a non-grey cloud. By using a nested-sampling algorithm to perform the retrievals, I was able to obtain a Bayesian evidence for each model, which could then be compared across the hierarchy to determine the "best" model. The Bayesian evidence penalises models with a higher number of parameters, allowing me to implement a form of Occam's razor. Using the retrieval results from the preferred models, I studied potential trends in planet mass, temperature, molecular abundances, and clouds. In addition, I investigated the so-called "normalization degeneracy", which describes the degenerate relationship between the relative molecular abundances and the absolute normalisation of the transmission spectrum.

From my model comparisons, I found that 8 of these planets' WFC3 spectra are adequately explained by a flat line, and 35 are well-fitted by an isothermal model with only water and a grey cloud. I found little to no evidence for any trends between the retrieved atmospheric parameters of these planets. My results also showed that the normalization degeneracy may be partially broken using the WFC3 data alone, and that the expected resolution from the upcoming James Webb Space Telescope will be sufficient to completely solve the issue. The retrievals of the planets for which a non-grey cloud model is preferred showed that it is not possible to constrain cloud composition using WFC3 data alone.

The intention of this work was to determine what can be learnt from the WFC3 data alone, and therefore the retrieval results may change somewhat significantly when additional datasets are analysed.

2.2 Publication

This work was published in Monthly Notices of the Royal Astronomical Society in 2018.

Retrieval analysis of 38 WFC3 transmission spectra and resolution of the normalisation degeneracy

Chloe Fisher[1,2]* and Kevin Heng[1]†

[1] *University of Bern, Center for Space and Habitability, Gesellschaftsstrasse 6, CH-3012, Bern, Switzerland*
[2] *University of Bern International 2021 Ph.D Fellowship*

ABSTRACT

A comprehensive analysis of 38 previously published Wide Field Camera 3 (WFC3) transmission spectra is performed using a hierarchy of nested-sampling retrievals: with versus without clouds, grey versus non-grey clouds, isothermal versus non-isothermal transit chords and with water, hydrogen cyanide and/or ammonia. We revisit the "normalisation degeneracy": the relative abundances of molecules are degenerate at the order-of-magnitude level with the absolute normalisation of the transmission spectrum. Using a suite of mock retrievals, we demonstrate that the normalisation degeneracy may be partially broken using WFC3 data alone, even in the absence of optical/visible data and without appealing to the presence of patchy clouds, although lower limits to the mixing ratios may be prior-dominated depending on the measurement uncertainties. With James Webb Space Telescope-like spectral resolutions, the normalisation degeneracy may be completely broken from infrared spectra alone. We find no trend in the retrieved water abundances across nearly two orders of magnitude in exoplanet mass and a factor of 5 in retrieved temperature (about 500–2500 K). We further show that there is a general lack of strong Bayesian evidence to support interpretations of non-grey over grey clouds (only for WASP-69b and WASP-76b) and non-isothermal over isothermal atmospheres (no objects). 35 out of 38 WFC3 transmission spectra are well-fitted by an isothermal transit chord with grey clouds and water only, while 8 are adequately explained by flat lines. Generally, the cloud composition is unconstrained.

Key words: planets and satellites: atmospheres

1 INTRODUCTION

At the time of writing, we are in the transitional period between the Hubble and James Webb Space Telescopes (HST and JWST). In the foreseeable future, WFC3 transmission spectra spanning 0.8–1.7 μm will be superceeded by NIR-Spec data ranging from 0.6–5 μm and at enhanced spectral resolution. It is therefore timely to perform a uniform theoretical analysis of a consolidated dataset of WFC3 transmission spectra, which is the over-arching motivation behind the current study.

1.1 Observational motivation: a statistical study of cloudy atmospheres

Following the work of Iyer et al. (2016), Fu et al. (2017) recently conducted a statistical study of the transmission spectra of 34 exoplanets (mostly hot Jupiters) measured using WFC3 onboard HST, which were mostly gathered from

Tsiaras et al. (2018). In order to isolate the spectral feature due to water[1], they quantified the strength of absorption between 1.3–1.65 μm, relative to the continuum, in terms of the number of pressure scale heights, which they represented by A_H. Based on the finding that both A_H and the equilibrium temperature ($T_{\rm eq}$) follow log-normal distributions, Fu et al. (2017) concluded that their sample of A_H is affected by observational bias. Tsiaras et al. (2018) defined an Atmospheric Detectability Index (ADI) to quantify the strength of detection of the water feature, but do not explicitly link the ADI to any trends in cloud properties. They concluded that all of their WFC3 transmission spectra, except for WASP-69b, are consistent with the presence of a grey cloud deck.

Our intention is to build upon the Fu et al. (2017) and Tsiaras et al. (2018) studies by subjecting their WFC3 sample to a detailed atmospheric retrieval study and elucidating the presence of assumptions, limitations and trends. It follows the principle that the same datasets should be analysed by different groups (using different codes and techniques)

---

* Email: chloe.fisher@csh.unibe.ch (CF)
† E-mail: kevin.heng@csh.unibe.ch (KH)

[1] Technically, it is due to a collection of unresolved water lines.
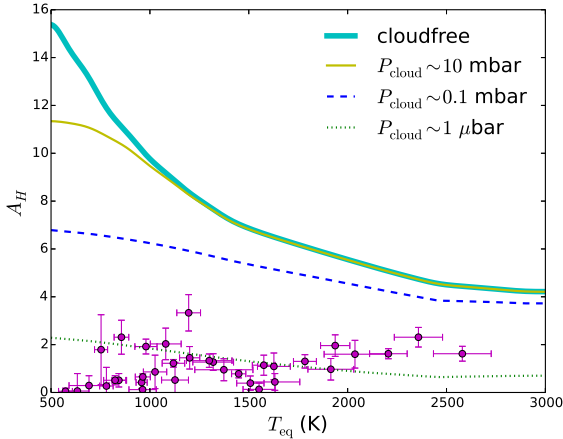
Figure 1. Strength of WFC3 water feature, A_H, in terms of pressure scale heights as a function of the equilibrium temperature. Also shown are the theoretical predictions of A_H for cloud-free and cloudy atmospheres. For the latter, the curves correspond to transit chords probing $P_{\rm cloud} \sim 1$ μbar, ~ 0.1 mbar and ~ 10 mbar if the opacity was solely due to grey clouds and the gravity is $\sim 10^3$ cm s^{-2}. It is apparent that all of the 34 atmospheres are cloudy if only water is assumed to be present.

Figure 2. Opacities of water, hydrogen cyanide and ammonia as functions of wavelength. The `ExoMol` spectroscopic line list was used as input for computing these opacities. For water and hydrogen cyanide, we show a sequence of opacities from 900–2100 K. Darker shades of the same colour correspond to higher temperatures. For ammonia, the temperature sequence is terminated at 1500 K, because `ExoMol` does not provide any data for higher temperatures. The spectral resolution is 5 cm^{-1} and the pressure is fixed at $P = 10$ mbar, because these values are what we use in our retrievals (see text for details).

within the community, so as to check for the consistency and robustness of theoretical interpretations (Fortney et al. 2016).

From a theoretical standpoint, A_H is an elegant quantity to examine, because the difference in transit radii between the peak of the water feature and the continuum is simply
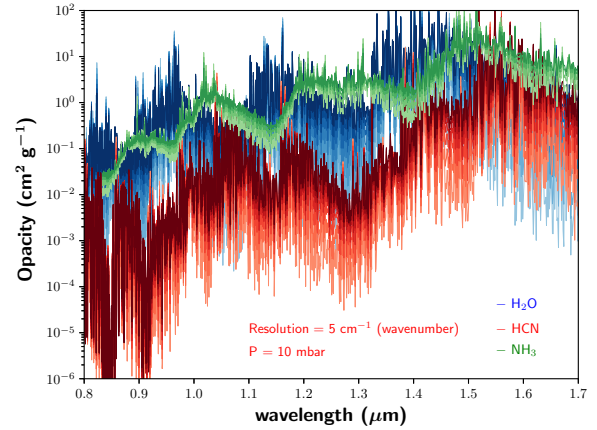
$$A_H = \ln \left(\frac{\kappa_{\max}}{\kappa_{\min}} \right), \tag{1}$$

where $\kappa_{\max}$ and $\kappa_{\min}$ are the maximum and minimum values of the water opacity in the WFC3 range of wavelengths. The preceding equation naturally derives from equation (2), if the volume mixing ratio of water ($X_{\rm H_2O}$) is assumed to be uniform across altitude, and is free of the normalisation degeneracy (see next subsection). Its simplicity allows us to do a first check on if the 34 objects in the sample gathered by Fu et al. (2017) have cloudy atmospheres.

In Figure 1, we show curves of A_H for completely cloud-free atmospheres by assuming that the temperature (sampled by the water opacity) is the equilibrium temperature. Also shown are curves of A_H corresponding to cloudy atmospheres with constant opacities. For example, an opacity of 1 cm^2 g^{-1} corresponds to a transit chord probing a pressure ~ 0.1 mbar if only clouds (and not molecules) are present in the atmosphere. By comparing these theoretical curves to the measured data points of Fu et al. (2017), we tentatively conclude that all of the 34 transiting exoplanets in their sample have cloudy atmospheres. It is one of the goals of the present study to examine if this conclusion is robust. Assuming that the temperature is some fraction of the equilibrium temperature merely translates the theoretical curves along the horizontal axis (not shown).

There is an additional, supporting argument for the atmospheres being cloudy. By visual inspection of measured WFC3 transmission spectra, we noticed that the continuum

blue-wards of the 1.4 μm water feature tends to be somewhat flat, in contrast to the opacity of water which tends to be rather structured at these wavelengths (Figure 2). This suggests that most, if not all, of the WFC3 transmission spectra measured so far are probing cloudy atmospheres—at least at the atmospheric limbs. However, this argument becomes less clear if ammonia and hydrogen cyanide are present, as they may mimic these effects on the spectra.

In the current study, one of our goals is to formalise this finding by performing atmospheric retrieval, within a nested-sampling framework (e.g., Skilling 2006; Feroz & Hobson 2008; Feroz et al. 2009, 2013; Benneke & Seager 2013; Waldmann et al. 2015; Lavie et al. 2017; Tsiaras et al. 2018), on each of the 34 objects in the Fu et al. (2017) sample. We construct a hierarchy of models with increasing levels of sophistication: cloud-free model (2 parameters), cloudy model with constant/grey cloud opacity (3 parameters), cloudy model with non-grey opacity (6 parameters). It is assumed that the main molecular absorber is water. If hydrogen cyanide (HCN) and ammonia (NH$_3$) are added to the analysis (MacDonald & Madhusudhan 2017), then it adds two more free parameters for a maximum of 8 parameters for the isothermal model. Our non-isothermal model adds another parameter. For comparison, MacDonald & Madhusudhan (2017) employ a 16-parameter model based partly on the heritage from Madhusudhan & Seager (2009).

We use the computed Bayesian evidence (Trotta 2008) from the retrievals to select the best model given the quality of the data, and hence determine if the atmospheres are cloudy, if cloud properties may be meaningfully constrained, and if NH$_3$ and/or HCN are detected in a given WFC3 spectrum. Unlike the approach adopted by MacDonald & Madhusudhan (2017), we do not test for patchy clouds. Rather,

we test essentially for whether the cloud particles are small or large (compared to the wavelengths probed).

1.2 Theoretical motivation: the normalisation degeneracy

Atmospheric retrievals of transmission spectrum typically specify a plane-parallel model atmosphere, assume azimuthal symmetry and then trace a transit chord through a set of atmospheric columns (each approximated by a plane-parallel atmosphere) (Madhusudhan & Seager 2009; Benneke & Seager 2012, 2013; Line et al. 2013; Waldmann et al. 2015). This brute-force procedure for calculating the transmission spectrum was previously described by Brown (2001) and Hubbard et al. (2001). In the current study, our intention is to build a nested-sampling retrieval framework around a validated analytical model for computing the transmission spectrum that bypasses the need for a brute-force calculation. Complementary to previous retrieval studies, we make a different set of investments, approximations and simplifications.

Building on the work of Lecavelier des Etangs et al. (2008), de Wit & Seager (2013) and Bétrémieux & Swain (2017), Heng & Kitzmann (2017) demonstrated that an analytical expression for the isothermal transit chord of an atmosphere,

$$R = R_0 + H\left(\gamma + E_1 + \ln\tau\right), \qquad (2)$$

is accurate enough[2] to model WFC3 transmission spectra for atmospheres with temperatures $\sim$ 1000 K or hotter, where we have
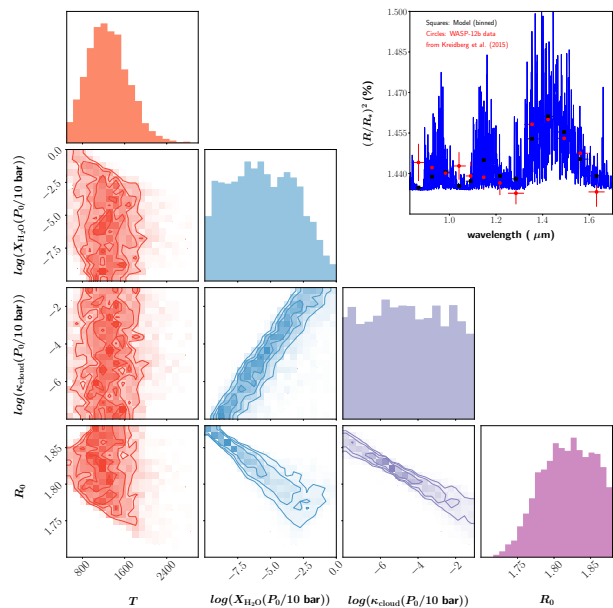
$$\tau \equiv \frac{\kappa P_0}{g}\sqrt{\frac{2\pi R_0}{H}}. \qquad (3)$$

The pressure scale height is given by H, the Euler-Mascheroni constant by $\gamma \approx 0.57721$ and the surface gravity by g. The exponential integral of the first order is given by $E_1(\tau)$ (Abramowitz & Stegun 1970; Arfken & Weber 1995), which has the argument τ. For a WFC3 spectrum dominated by water, the opacity is $\kappa \propto X_{H_2O}$, where X_{H_2O} is the volume mixing ratio of water. Equation (2) assumes that $R_0 < R$; if the layer of the atmosphere located at R_0 is opaque in the WFC3 bandpass ($\tau \gg 1$), then the E_1 term may be dropped.

Equation (2) straightforwardly shows that there exists a three-way degeneracy between the reference transit radius (R_0), reference pressure (P_0) and X_{H_2O}, which was first noticed numerically[3] by Benneke & Seager (2012) and Griffith (2014). The values of R_0 and P_0, as well as the relationship between them, are a priori unknown, because it is akin to having prior knowledge of the structure of the exoplanet. It is apparent that a small change in R_0 causes a large variation in X_{H_2O}. Furthermore, it is $X_{H_2O}P_0$, and not X_{H_2O} alone, that is being retrieved from the data. It is worth

Figure 3. Posterior distributions of the water volume mixing ratio (X_{H_2O}), temperature (T), grey/constant cloud opacity (κ_{cloud}) and reference transit radius (R_0; uniform prior from 1.7–1.88 R_J) from a retrieval analysis of the WASP-12b transmission spectrum. The degeneracies between R_0 and the other quantities is apparent; R_0 is bounded from below because X_{H_2O} is bounded from above by unity. In this test, we have set $P_0 = 10$ bar but in our subsequent retrieval of the WASP-12b WFC3 transmission spectrum we will fit for P_0 (see text and Figure 20). The measured data and best-fit model are shown in the top-right panel. The physical units of T and κ_{cloud} are K and cm^2 g^{-1}, respectively, while R_0 is given in units of Jupiter radii (R_J). This retrieval assumes a constant mean molecular mass and ignores the effect of collisional induced absorption, which we will explore later in the current study.

emphasising that these obstacles do not exist in the forward problem, where one makes a specific set of assumptions (e.g., solar metallicity, chemical equilibrium) and computes the transmission spectrum, but they are front and center in the inverse problem. Heng & Kitzmann (2017) pointed out these issues, but they did not examine them further within a Bayesian retrieval framework, which partially motivates the current study.

Figure 3 shows a retrieval calculation performed using a new code (HELIOS-T) presented as part of the current study, which we constructed specifically to perform fast retrievals on transmission spectra at low spectral resolution.[4] (A detailed description of methodology will come later in §2.) It demonstrates that while the temperature may be robustly retrieved, there are order-of-magnitude degeneracies associated with the water mixing ratio and cloud opacity that arise from small variations of R_0 (in the third significant figure), as previously elucidated by Heng & Kitzmann (2017). In the present study, we wish to examine if R_0 or P_0 may be used as

---

[2] Meaning that the errors incurred are smaller than the noise floor (about 50 parts per million) of HST and the expected noise floor of JWST.

[3] Our stand is that a numerical demonstration of an effect alone does not qualify as attaining full understanding of it, until its theoretical (analytical) formalism has been elucidated.
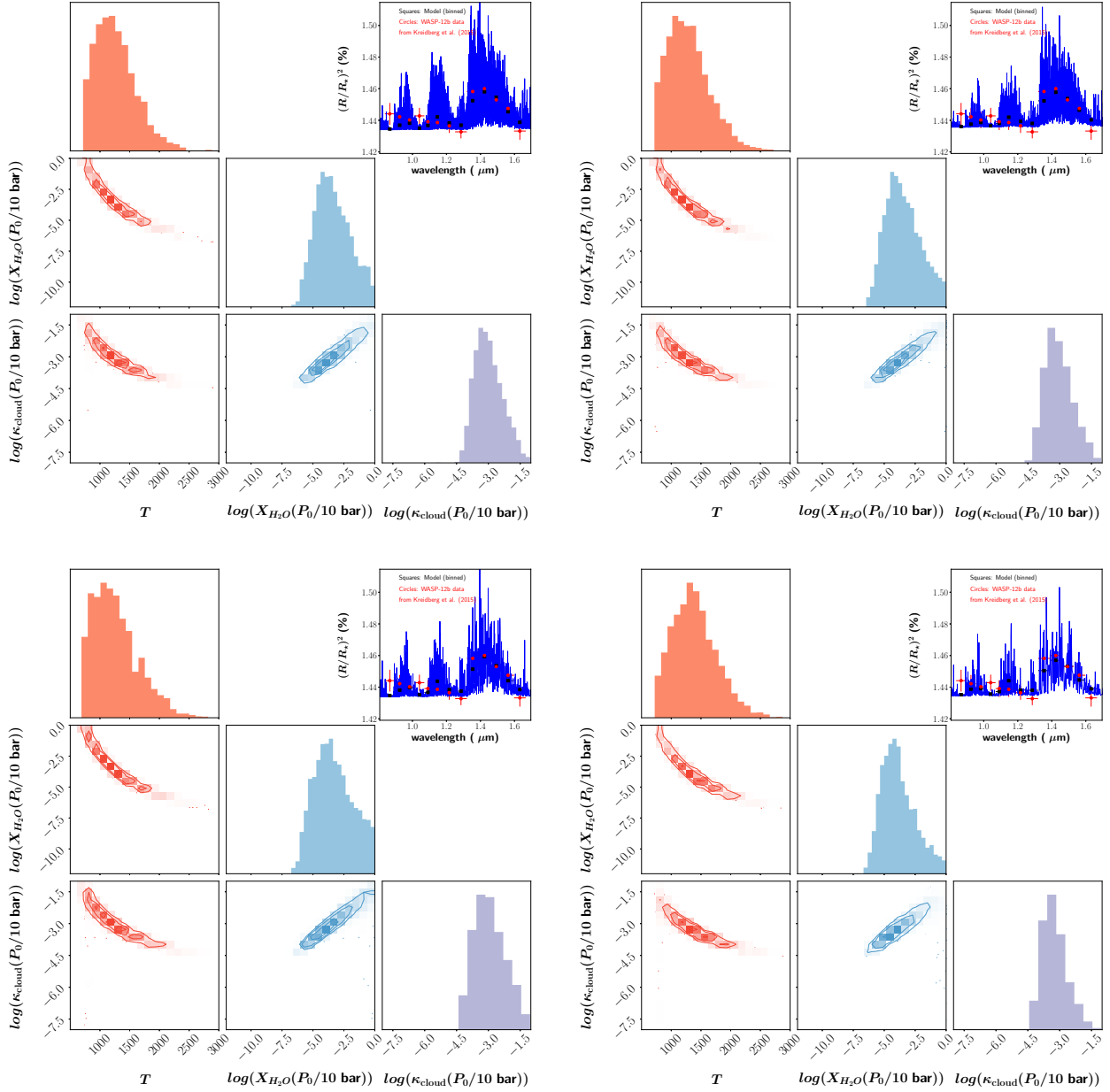
[4] At high spectral resolution, the fully resolved spectral lines may span many orders of magnitude in pressure between the line peaks and wings, thereby violating the isobaric assumption.

Figure 4. Same as Figure 3, but using spectral resolutions of 1 cm^{-1} (top left panel), 2 cm^{-1} (top right panel), 5 cm^{-1} (bottom left panel) and 10 cm^{-1} (bottom right panel) for the `ExoMol` water opacity.

a fitting parameter to break the normalisation degeneracy. We further examine if the normalisation degeneracy may be broken with WFC3 transmission spectra alone, or if JWST-like spectra is needed.

1.3 Layout of current study

In §2, we describe our theoretical methodology, including how we compute transit radii and opacities. In §3, we perform suites of tests, a detailed analysis of the 38 WFC3 transmission spectra in the Tsiaras et al. (2018) and de Wit et al. (2018) samples and elucidate trends among the retrieved quantities. The implications of our results are discussed in §4. Table 1 lists our assumptions for the prior dis-

tributions of parameters. Table 2 summarises our retrieval results. Table 3 summarises some of the input parameters for the retrievals.

Our new nested-sampling retrieval code for transmission spectra, `HELIOS-T`, is part of our open-source suite of tools for analysing exoplanetary atmospheres known as the Exoclimes Simulation Platform (`www.exoclime.org` or `https://github.com/exoclime`).

Table 1. Assumed prior distributions for retrievals and values of physical constants used

Quantity	Symbol	Range	Assumption	Units
Temperature	T	$(100, 2900)$	Uniform	K
Water mixing ratio	X_{H_2O}	$(10^{-13}, 1)$	Log-uniform	–
Hydrogen cyanide mixing ratio	X_{HCN}	$(10^{-13}, 1)$	Log-uniform	–
Ammonia mixing ratio	X_{NH_3}	$(10^{-13}, 1)$	Log-uniform	–
Grey cloud opacity	κ_{cloud}	$(10^{-12}, 10^2)$	Log-uniform	cm^2 g^{-1}
Opacity normalisation for non-grey cloud model	κ_0	$(10^{-10}, 10^{-1})$	Log-uniform	cm^2 g^{-1}
Composition parameter in non-grey cloud model	Q_0	$(1, 100)$	Uniform	–
Index in non-grey cloud model	a	$(3, 6)$	Uniform	–
Monodisperse, spherical cloud particle radius	r_c	$(10^{-7}, 10^{-1})$	Log-uniform	cm
Non-isothermal temperature profile parameter	b	$(-30, -1)$, $(1, 30)$	Uniform	–
Reference transit radius†	R_0	$(1.619, 1.799)$	Uniform	R_J
Reference pressure	P_0	$(10^{-1}, 10^3)$	Log-uniform	bar
Equatorial radius of Jupiter	R_J	7.1492×10^9	–	cm
Mass of hydrogen atom	m_H	m_{amu}	–	cm
Atomic mass unit	m_{amu}	$1.66053904 \times 10^{-24}$	–	g
Boltzmann constant	k_B	$1.38064852 \times 10^{-16}$	–	erg K^{-1}

$\dagger$: Only used in the test retrievals of WASP-17b (§3.2.2).

2 METHODOLOGY

2.1 Transmission spectra

As explained in §1, equation (2) describes our forward model for transforming a given temperature, surface gravity, opacity function, reference transit radius and reference pressure into a transmission spectrum. The accuracy of equation (2) has previously been demonstrated by Heng & Kitzmann (2017) and we will not repeat the analysis and explanation here. To test for non-isothermality, we use another formula derived by Heng & Kitzmann (2017),

$$R = R_0 + H\tau^{1/b} \left(\gamma + E_1 + \ln\tau\right) \tag{4}$$

where the reference optical depth is now given by

$$\tau = \frac{\pi P_0 \kappa}{2g} \sqrt{\frac{2R_0 |b|}{H}}. \tag{5}$$

We again have $E_1 = E_1(\tau)$. The dimensionless index b is the ratio of the non-isothermal to the isothermal scale height. If the values of $|b|$ are much larger than unity, then it means that the behavior is close to being isothermal. Essentially, our simplified temperature-pressure profile is described by 2 parameters.

Our approach is complementary to other approaches in the literature that use more complicated prescriptions for temperature-pressure profiles. For example, Madhusudhan & Seager (2009) and MacDonald & Madhusudhan (2017) use 9- and 7-parameter fitting functions, respectively. Again, we make a different investment: we choose to simplify the temperature profile prescription in order to isolate the effects of the other parameters. It allows us to more cleanly study degeneracies.

2.2 Opacities

3 Our H_2O, HCN and NH_3 opacities are taken from the ExoMol spectroscopic database (Barber et al. 2006; Yurchenko et al. 2011, 2013; Barber et al. 2014; Yurchenko & Tennyson

2014). In a single set of tests (see §3.1.4), we also use the HITRAN (Rothman et al. 1987, 1992, 1998, 2003, 2005, 2009, 2013) and HITEMP (Rothman et al. 2010) databases for water. For a review of the spectroscopic databases, please see Tennyson & Yurchenko (2017). For the procedure on how to use the ExoMol inputs to compute opacities, we refer the reader to Grimm & Heng (2015), Chapter 5 of Heng (2017) and Yurchenko et al. (2018). Examples of opacities for all three molecules are given in Figure 2.

The opacity function used in equation (2) is given by

$$\kappa = \frac{X_{H_2O} m_{H_2O} \kappa_{H_2O}}{m} + \frac{X_{HCN} m_{HCN} \kappa_{HCN}}{m} + \frac{X_{NH_3} m_{NH_3} \kappa_{NH_3}}{m} + \kappa_{cloud}, \tag{6}$$

where m is the mean molecular mass, m_{H_2O} is the mass of the water molecule, κ_{H_2O} is the water opacity, X_{HCN} is the volume mixing ratio of hydrogen cyanide, m_{HCN} is the mass of the hydrogen cyanide molecule, κ_{HCN} is the hydrogen cyanide opacity, X_{NH_3} is the volume mixing ratio of ammonia, m_{NH_3} is the mass of the ammonia molecule, κ_{NH_3} is the ammonia opacity and κ_{cloud} is the cloud opacity.

Denoting the atomic mass unit by m_{amu}, the mean molecular weight ($\mu = m/m_{amu}$) is given by

$$\mu = 2.4X_{H_2} + \frac{X_{H_2O} m_{H_2O}}{m_{amu}} + \frac{X_{HCN} m_{HCN}}{m_{amu}} + \frac{X_{NH_3} m_{NH_3}}{m_{amu}}. \tag{7}$$

The mixing ratio of molecular hydrogen is determined by demanding that all mixing ratios sum to unity,

$$1.1X_{H_2} + X_{H_2O} + X_{HCN} + X_{NH_3} = 1, \tag{8}$$

where we have assumed that the helium mixing ratio follows cosmic abundance ($X_{He} = 0.1X_{H_2}$).

The molecular opacities are sampled at 1 mbar for the first suite of tests (§3.1; to ensure continuity with Heng & Kitzmann 2017) and 10 mbar for our second suite of tests (§3.2.2) and actual results (see §3.3). The cloud mixing ratio is subsumed into κ_{cloud}. The opacity associated with collision-induced absorption (both H_2-H_2 and H_2-He) are taken from Rothman et al. (2013).

An unresolved physics problem inherent in the computation of opacities concerns the effects of pressure broadening. The spectral lines of *isolated* atoms and molecules are described rather well by a Voigt profile. As a population, collisions between them become important at high enough pressures, which modifies the shape of the far line wings of the profile. It remains unknown exactly what "far" means and how to compute these modified profiles. In practice, various workers in the field have resorted to truncating the Voigt profiles at some ad hoc distance from line centre (see Grimm & Heng 2015 and references therein). For this study, we use a line-wing cutoff of 100 cm^{-1}. Fortunately, since transmission spectra probe pressures that are tenuous enough such that pressure broadening has a negligible effect for ~ 1000 K atmospheres, this is not a limiting issue.

Another limitation is that, at the time of writing, the NH$_3$ opacities do not exist for temperatures above 1600 K (Yurchenko et al. 2011). In the absence of these data, we set the opacity for NH$_3$ to be zero for temperatures above 1600 K.

2.3 Cloud models

We consider both grey and non-grey clouds. For our grey cloud model, we assume a constant cloud opacity, which is physically equivalent to assuming that the cloud particles are much larger than the WFC3 wavelengths being probed. Our non-grey cloud model uses the opacity of (Kitzmann & Heng 2018),

$$\kappa_{\text{cloud}} = \frac{\kappa_0}{Q_0 x^{-a} + x^{0.2}}, \tag{9}$$

where $x = 2\pi r_c/\lambda$ is the dimensionless size parameter, r_c is the particle radius and λ is the wavelength. In their study of 32 condensate species, Kitzmann & Heng (2018) showed that $Q_0 \approx 0.1$–65 is a proxy for cloud composition with larger values corresponding to more volatile species. For example, water ice has $Q_0 = 64.98$ and olivine has $Q_0 \approx 10$. The index a ranges from 3 to 7; $a = 4$ corresponds to Rayleigh scattering. Our non-grey cloud model has 4 free parameters: κ_0, Q_0, r_c and a. The immediate implication of the preceding equation is that if the cloud is grey ($a \approx 0$), then the composition cannot be decisively constrained.

Conceptually, the treatment of Lee et al. (2013) and Kitzmann & Heng (2018) are identical in that they both allow smooth transitions between the Rayleigh and large-particle regimes. However, Lee et al. (2013) assumed $a = 4$, whereas Kitzmann & Heng (2018) calibrated Q_0 and a against a larger library of species and composition.

2.4 Data

For 30 out of 38 objects, the WFC3 transmission spectra were obtained from Tsiaras et al. (2018) and provided in electronic form by the first author (A. Tsiaras 2018, private communication). For WASP-17b, WASP-19b, GJ 1214b and HD97658b, the WFC3 transmission spectra were obtained from Mandell et al. (2013), Huitson et al. (2013), Kreidberg et al. (2014a) and Knutson et al. (2014), respectively. The WFC3 transmission spectra of TRAPPIST-1d, e, f and g were taken from de Wit et al. (2018). The stellar radii and surface gravities for each object are listed in Table 3.
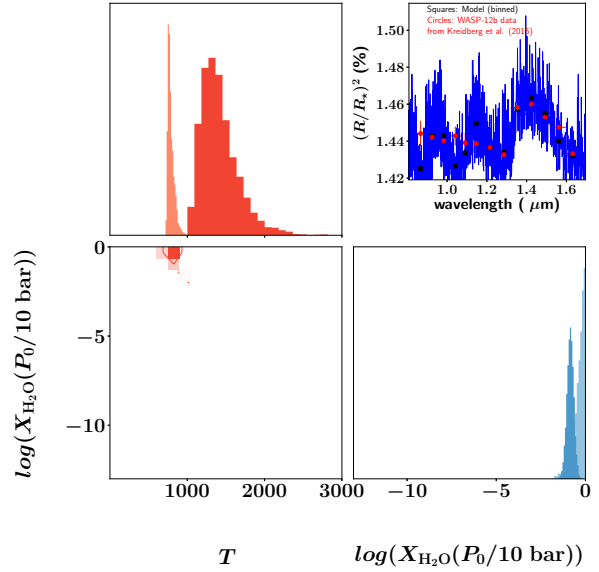
Figure 5. Same as Figure 3, but for cloud-free models in which we fix $m = 2.4m_{\text{H}}$ (log $X_{\text{H}_2\text{O}}$ posterior bumps up against 0) versus a variable m (posteriors distributions are in a darker shade) that takes into account water-rich atmospheres.

Uncertainties in the stellar radii manifest themselves as uncertainties in the normalisation of the transmission spectra.

3 RESULTS

3.1 Suite of tests on WASP-12b transmission spectrum

To provide continuity between Heng & Kitzmann (2017) and the present study, we use the WFC3 transmission spectrum of WASP-12b (13 data points), measured by Kreidberg et al. (2015), as our starting point for tests. To cleanly isolate the effects studied, we begin with using a constant/grey cloud opacity. Note that for these WASP-12b tests only (Figures 4 to 9), the molecular opacities are sampled at 1 mbar, CIA is not included and m is fixed at $2.4m_{\text{H}}$, where m_{H} is the mass of the hydrogen atom (which we take to be one atomic mass unit, m_{amu}), unless otherwise stated. In these tests only, we set $R_0 = 1.79$ R_{J} and $P_0 = 10$ bar. These restrictions are lifted for the rest of the study.

3.1.1 Spectral resolution of opacities

In Figure 4, we perform resolution tests associated with the sampling of the water opacity across wavenumber. We show retrieval outcomes for spectral resolutions of 1, 2, 5 and 10 cm^{-1}. For all of these values, the posterior distributions of T, $X_{\text{H}_2\text{O}}$ and κ_{cloud} are somewhat similar. Specifically, the retrieved temperatures are 1218^{+388}_{-297} K, 1252^{+393}_{-307} K, 1203^{+468}_{-323} K and 1363^{+403}_{-343} K, respectively. The logarithm of the retrieved water volume mixing ratios are $-3.51^{+1.66}_{-1.26}$, $-3.61^{+1.68}_{-1.27}$, $-3.46^{+1.93}_{-1.49}$ and $-4.04^{+1.62}_{-1.14}$, respectively. For the rest of the study, we will adopt a sampling resolution of 5 cm^{-1}.
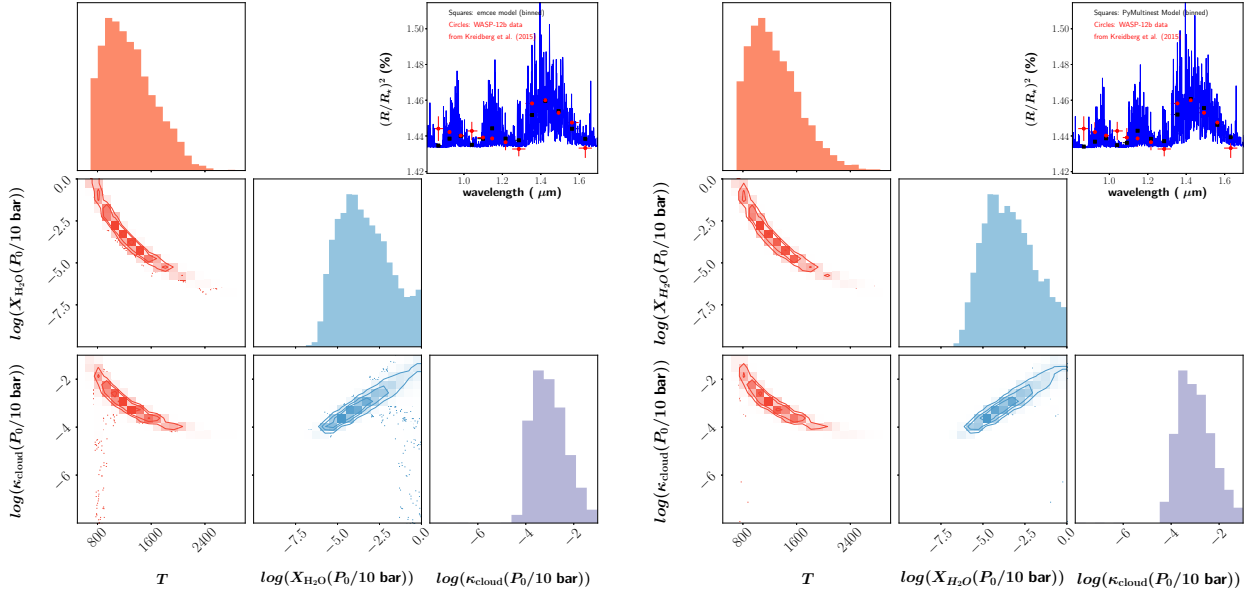
Figure 6. Same as Figure 3, but comparing a MCMC (left panel) versus nested-sampling (right panel) retrieval approach.

3.1.2 Cloudy versus cloud-free

Another necessary check is to determine that cloudy models are necessary in the first place for WASP-12b. In Figure 5, we subject the WASP-12b WFC3 transmission spectrum to two cloud-free retrievals: the first has a fixed $m = 2.4m_H$, while the second has a variable m. For the retrieval with a fixed m, the outcome is implausible as the water volume mixing ratio is $\sim 10\%$–100%. The retrieval with a variable m produces more plausible posteriors, but even by visual inspection it is apparent that the cloud-free model struggles to match the somewhat flat spectral continuum blue-wards of the 1.4 μm water feature. For the rest of the WFC3 transmission spectra, we will not show the posterior distributions associated with the cloud-free retrieval (unless it has the highest Bayesian evidence in the model hierarchy), but we will still include them in the overall analysis.

3.1.3 MCMC versus nested sampling

The next logical test is to compare cloudy retrievals obtained using a Markov Chain Monte Carlo (MCMC) versus nested sampling approach. For the former, we use the open-source `emcee` package (Foreman-Mackey et al. 2013). For the latter, we use the open-source `PyMultiNest` package (Buchner et al. 2014). Figure 6 compares the outcome from this pair of retrievals. It is reassuring that the posterior distributions of T, X_{H_2O} and κ_{cloud} are somewhat similar, although we note that the retrieval performed with MCMC produces higher values of the water volume mixing ratio in the tail of the distribution (towards $X_{H_2O} = 1$). The reason to select the nested-sampling approach over MCMC is because it allows us to straightforwardly compute the Bayesian evidence associated with each model, which in turn allows us to formally apply Occam's Razor.

3.1.4 Choice of spectroscopic databases: `HITRAN` versus `HITEMP` versus `ExoMol`

Perhaps the most surprising outcome of our series of WASP-12b tests is shown in Figure 7, where we examine the retrieval outcomes using the `HITRAN`, `HITEMP` and `ExoMol` spectroscopic databases to construct the water opacity. The main shortcoming with `HITRAN` is that it omits the weak lines of water that contribute prominently to the spectral continuum when $T \sim 1000$ K or hotter. `HITEMP` addresses this issue somewhat, but it is widely accepted by the exoplanet community that `ExoMol` addresses this issue most completely to date (see Tennyson & Yurchenko 2017 for a review). With an equilibrium temperature in excess of 2500 K, WASP-12b is an ideal target for testing if discrepancies from retrievals arise from the use of different line lists. Yet, Figure 7 shows us that the choice of line list for the water opacity is irrelevant at the present spectral resolution and signal-to-noise attainable of the WFC transmission spectrum of WASP-12b. It suggests that the retrievals performed on the other WFC3 transmission spectra are robust to the choice of spectroscopic line list. Despite this finding, we persist in using the `ExoMol` line list in order to dispel any notion that our results lack robustness.

3.1.5 Insensitivity to pressure broadening

Pressure broadening is an ill-defined source of uncertainty, because there is no first-principles theory to describe it (see discussion in §2.2). Nevertheless, we quantify its effect as the final test in this WASP-12b suite. Figure 8 shows two retrievals with $P = 1$ mbar versus 1 bar, which span the conceivable range of pressures probed by the WFC3 transmission spectrum. The effects on the posterior distributions of the temperature, water mixing ratio and grey cloud opac-

ity are minimal, even with a factor of 1000 difference in pressure between the pair of retrievals.

For the rest of the study, we will fix the pressure associated with pressure broadening at 10 mbar. The reasoning is that departures from this value will result in minor errors to the retrieved posterior distributions, which are subsumed as errors in the grey cloud opacity. Given that the exact functional form of pressure broadening cannot be specified from first principles, this is a reasonable approach, because it allows us to include pressure broadening in a more controlled way.

3.1.6   *Comparison of Bayesian evidence*

Following these tests, we analyse the WFC3 transmission spectrum of WASP-12b using a hierarchy of models: with and without clouds, grey versus non-grey clouds, isothermal versus non-isothermal and with various permutations of the three molecules being present. Figure 9 shows the Bayes factor for each model, which is the logarithm of the ratio of the Bayesian evidence of a given model compared to the best model. The value of the Bayes factor may be interpreted as being weak, moderate or strong evidence for the best model in favour of a given model (Trotta 2008). It may also be used to infer that the comparison is inconclusive, i.e., there is no evidence to favour one model over the other, if the Bayes factor between them is less than unity.

A few conclusions may be drawn from Figure 9. First, cloud-free models are disfavoured. Second, there is weak evidence for non-isothermal behaviour, non-grey clouds and the presence of HCN and/or NH_3, but overall the isothermal model with only water present and grey clouds is sufficient to fit the WFC3 transmission spectrum. In other words, there is no evidence for more complicated models to be favoured.

Again, note that the molecular opacities are sampled at 1 mbar, CIA is not included, m is fixed at $2.4m_H$ and we have fixed $R_0 = 1.79\,R_J$ and $P_0 = 10$ bar. These assumptions will be lifted for WASP-12b in Figure 20.

3.2   Breaking the normalisation degeneracy for cloud-free objects

3.2.1   *Deriving $R_0(P_0)$: case study of WASP-17b*

Heng (2016) previously concluded that the atmospheres of WASP-17b and WASP-31b are cloud-free based on optical transmission spectra recorded by STIS (Sing et al. 2016). This conclusion was based on the reasoning that the sodium and potassium lines may serve as diagnostics for cloudiness. The peaks of these resonant lines are hardly affected by clouds, but the line wings are, which makes the distance between the line peak and wing highly sensitive to the degree of cloudiness. If the optical transit chord is cloud-free, then we may associate the measured optical spectral slope with Rayleigh scattering by hydrogen molecules (H_2), which yields a direct measurement of the pressure scale height (Lecavelier des Etangs et al. 2008; Heng 2016),

$$H = -\frac{1}{4}\frac{\partial R}{\partial (\ln \lambda)}, \tag{10}$$

where λ is the wavelength. Such an approach is possible only because we have $\kappa = X_{H_2} m_{H_2} \kappa_{H_2}/m$, $X_{H_2} \approx 1$ and κ_{H_2}

is known from first principles. If the optical transit chord is cloudy, then $\kappa = X_{cloud} m_{cloud} \kappa_{cloud}/m$. The cloud volume mixing ratio (X_{cloud}), composition of the cloud particles (and hence their mass, m_{cloud}) and opacity (κ_{cloud}) are now generally unknown and cannot be uniquely retrieved from either the optical or WFC3 transmission spectra.

We use WASP-17b as a working example, for which Heng (2016) previously estimated $H = 1896$ km using two data points from Sing et al. (2016) and $R_\star = 1.583\,R_\odot$ (Southworth et al. 2012). In the current study, we fit a line to the optical spectral slope (comprising 15 data points) and derive $H = 1950$ km (not shown).

In a hydrogen-dominated atmosphere, the opacity associated with Rayleigh scattering alone is $\kappa = \sigma_{H_2}/m$. The cross section for Rayleigh scattering by hydrogen molecules is (Sneep & Ubachs 2005),

$$\sigma_{H_2} = \frac{24\pi^3}{n_{ref}^2 \lambda^4}\left(\frac{n_r^2 - 1}{n_r^2 + 2}\right)^2, \tag{11}$$

where $n_{ref} = 2.68678 \times 10^{19}$ cm^{-3} and the real part of the index of refraction is (Cox 2000)

$$n_r = 1.358 \times 10^{-4}\left[1 + 7.52 \times 10^{-3}\left(\frac{\lambda}{1\,\mu m}\right)^{-2}\right] + 1. \tag{12}$$

If the optical spectral slope is associated with H_2 Rayleigh scattering alone, then hydrostatic equilibrium allows us to derive a unique solution for P_0,

$$P_0 = \frac{0.56mg}{\sigma_{H_2}}\sqrt{\frac{H}{2\pi R_0}}\,\exp\left(\frac{R - R_0}{H}\right), \tag{13}$$

based on equation (2) and assuming that R_0 is associated with the part of the atmosphere that is opaque to both optical and infrared radiation.

For WASP-17b, we take $R = 1.890\,R_J$ at $\lambda = 0.405$ μm from the measurements of Sing et al. (2016). We then select a reference radius that is three orders of magnitude in pressure greater than that probed by WFC3,

$$R_0 = \bar{R}_{WFC3} - 6.908H, \tag{14}$$

where $\bar{R}_{WFC3}$ is the average value of the transit radius in the measured WFC3 bandpass. The preceding expression assumes hydrostatic equilibrium. For WASP-17b, we have $\bar{R}_{WFC3} = 1.897\,R_J$ and $R_0 = 1.709\,R_J$. Using the measured value of R and equation (13), we estimate that $P_0 = 8$ bar. This means that the pressure probed in the WFC3 bandpass is, on average, about 8 mbar. We note that the pressure probed at $\lambda = 0.405$ μm is about 10 mbar.

We do the same analysis for WASP-31b. We use $R_\star = 1.252\,R_\odot$ (Anderson et al. 2011) and derived $H = 1619$ km. Heng (2016) previously derived $H = 1390$ km based on using $R_\star = 1.12\,R_\odot$ (Anderson et al. 2011). We estimate $R_0 = 1.379\,R_J$ and $P_0 = 26$ bar, based on $R = 1.547\,R_J$ at $\lambda = 0.4032$ μm. This means that the WFC3 bandpass and the optical data point correspond to about 26 mbar and 15 mbar, respectively.

These estimates are broadly consistent with our approach of assuming 10 mbar for the molecular opacities.
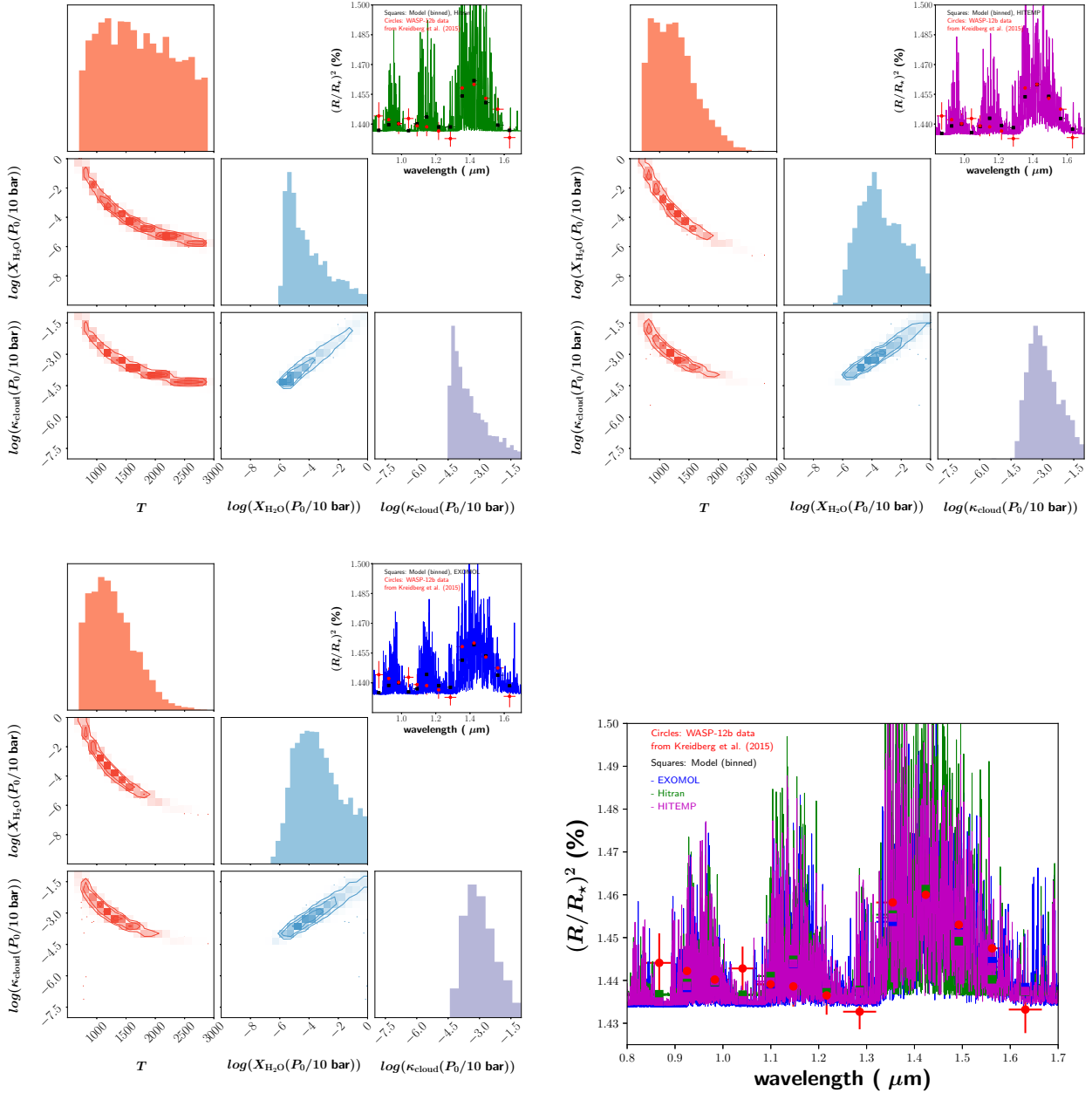
Figure 7. Same as Figure 3, but comparing the use of `HITRAN` (top left panel) versus `HITEMP` (top right panel) versus `ExoMol` (bottom left panel) spectroscopic line lists for water. Additionally, the best-fit spectra are compared in the bottom right panel.

3.2.2 Mock retrievals of WASP-17b: breaking the normalisation degeneracy

Using the derived $R_0 = 1.709$ $R_{\rm J}$ and $P_0 = 8$ bar, we perform suites of mock retrievals to study if the normalisation degeneracy may be broken. A uniform prior distribution of 1.619 to 1.799 $R_{\rm J}$ is set for R_0, while a log-uniform prior distribution of 0.1 to 1000 bar is set for P_0.

First, we create high-resolution mock transmission spectra with 100 data points that are representative of what will be possible with JWST. The uncertainty on each data point is assumed to be 10 parts per million (ppm). We explore pairs of retrievals in which R_0 is held fixed and P_0 is a fitting

parameter, and vice versa. Second, we create a hierarchy of mock spectra to gain understanding into the retrieval outcomes: three molecules with grey clouds, water only with grey clouds and water only (cloud-free). All volume mixing ratios are set to 10^{-3} for illustration, with a grey-cloud opacity of $10^{-2} {\rm cm}^2 {\rm g}^{-1}$.

Figure 10 shows the outcomes of 6 retrievals on high-resolution mock spectra. Unexpectedly, the peaks of the narrow posterior distributions of all 6 parameters, including R_0 or P_0, land exactly on the true values. The pair of cloud-free retrievals with water only also manages to locate the correct solution. In fact, the posterior distribution on the temperature is essentially a narrow spike with no width. This
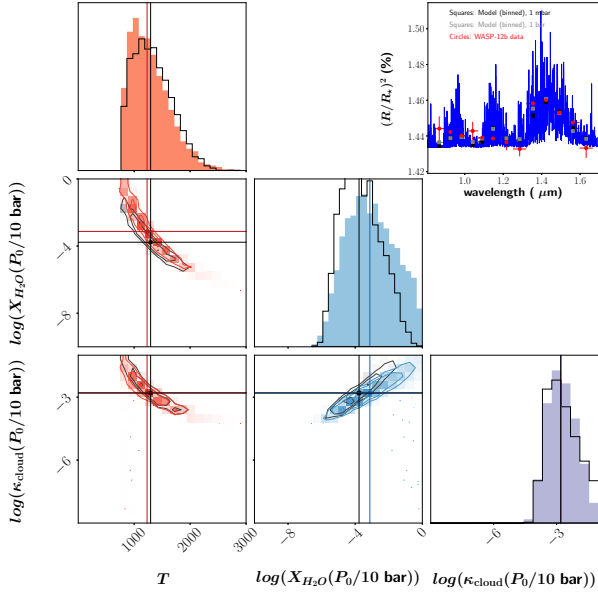
Figure 8. Same as Figure 3, but elucidating the effect of pressure broadening. The posteriors are for $P = 1$ mbar, while the posteriors associated with $P = 1$ bar are overplotted as the solid curves. The vertical and horizontal lines represent the median values.
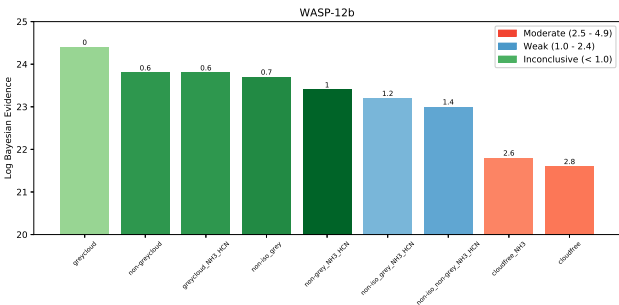
Figure 9. Logarithm of the Bayesian evidence and corresponding Bayes factor between each model compared to the best model (as indicated by the number on top of each bar). The entry marked by "0" is the best model, i.e., the model with the highest Bayesian evidence. The legend lists the correspondence between the Bayes factor and the strength or weakness of the evidence in favour of a given model (compared to the best model). For this set of Bayesian evidences only, we sample the opacities at 1 mbar, ignore CIA and use a fixed $m = 2.4m_{\mathrm{H}}$ (see text).

is straightforward to understand, because the temperature controls the "stretch factor" in the spectrum and a unique solution is obtained by correctly fitting for the difference between the peaks and troughs of the spectrum. By contrast, R_0 or P_0 serves as a "translation factor", which shifts the spectrum up or down in transit radius or depth without altering its shape. Further insight is obtained by examining a pair of retrievals with water only but with grey clouds present. The presence of grey clouds provides an extra degree of freedom in the system in the form of a constant spectral continuum. Grey clouds mute spectral features, which

may be compensated by an increase in the volume mixing ratio of water, which is clearly seen in Figure 10. Note that the normalisation degeneracy is simultaneously present, as increases in $X_{\mathrm{H_2O}}$ and κ_{cloud} are negated by decreases in P_0 or R_0. The lower bound on the water mixing ratio in this pair of retrievals is artificial and is set by the chosen upper limit of our prior on R_0 or P_0. This pair of cloudy retrievals with water only allows us to understand that the degeneracy may be broken, even in the presence of clouds, if multiple molecules are present to provide additional information on the shape of the spectrum.

Figure 11 shows the same suite of retrievals but for a low-resolution, WFC3-like spectrum with 20 data points. The uncertainty on each data point is assumed to be 50 ppm. For each of the 6 retrievals, we perform an additional retrieval in which R_0 or P_0 is held fixed at its true value (1.709 R_{J} or 8 bar). The lessons learnt and insights gleaned from the high-resolution retrievals carry over to the low-resolution ones. Tight constraints are obtained on the temperature. For the volume mixing ratios of the molecules, constraints are obtained at the order-of-magnitude level that encompass the true values, but it is important to note that the lower bounds are artefacts of assuming an upper limit for the prior of R_0 or P_0. Unlike in the high-resolution regime, the low-resolution retrievals do not provide tight constraints on either R_0 or P_0.

The key lesson learnt is that, for meaningful retrieval outcomes to be obtained, we have to assume a reasonable range of prior values for R_0 or P_0. Since we find it easier to have an intuition about P_0, we will set the range of 0.1–1000 bar as the prior on P_0. It then becomes important to set a value of R_0 that corresponds to this range of P_0 values (see §3.2.3).

To illustrate this point, we perform an additional mock retrieval in which the value of R_0 is reduced and the corresponding P_0 value falls outside of the 0.1–1000 bar range. Figure 12 shows that the posterior distribution for P_0 bumps up against the upper boundary of the prior distribution, which results in errors in the retrieved values of temperature and water mixing ratio.

3.2.3   Catalogue of R_0 values for other objects

For the other 36 objects in our sample, we first assume the WFC3 bandpass to probe a pressure of 10 mbar. We then use equation (14) to estimate the value of R_0 that corresponds to 10 bar (Table 3). The pressure scale height is estimated using $H = k_{\mathrm{B}}T_{\mathrm{eq}}/mg$, where k_{B} is the Boltzmann constant and T_{eq} is the equilibrium temperature (as was done by Heng 2016). These R_0 values are then used as input in our retrievals.

We emphasize that while the value of R_0 is fixed to the tabulated value for each object, our retrievals ultimately use P_0 as a fitting parameter as justified by our tests in §3.2.2. The reason to use these values is to have R_0 be in approximately the range of values corresponding to 0.1–1000 bar, such that the retrieval will converge meaningfully.

3.2.4   Collision-induced absorption

As a final test on mock WASP-17b spectra, we consider an isothermal model atmosphere with all three molecules
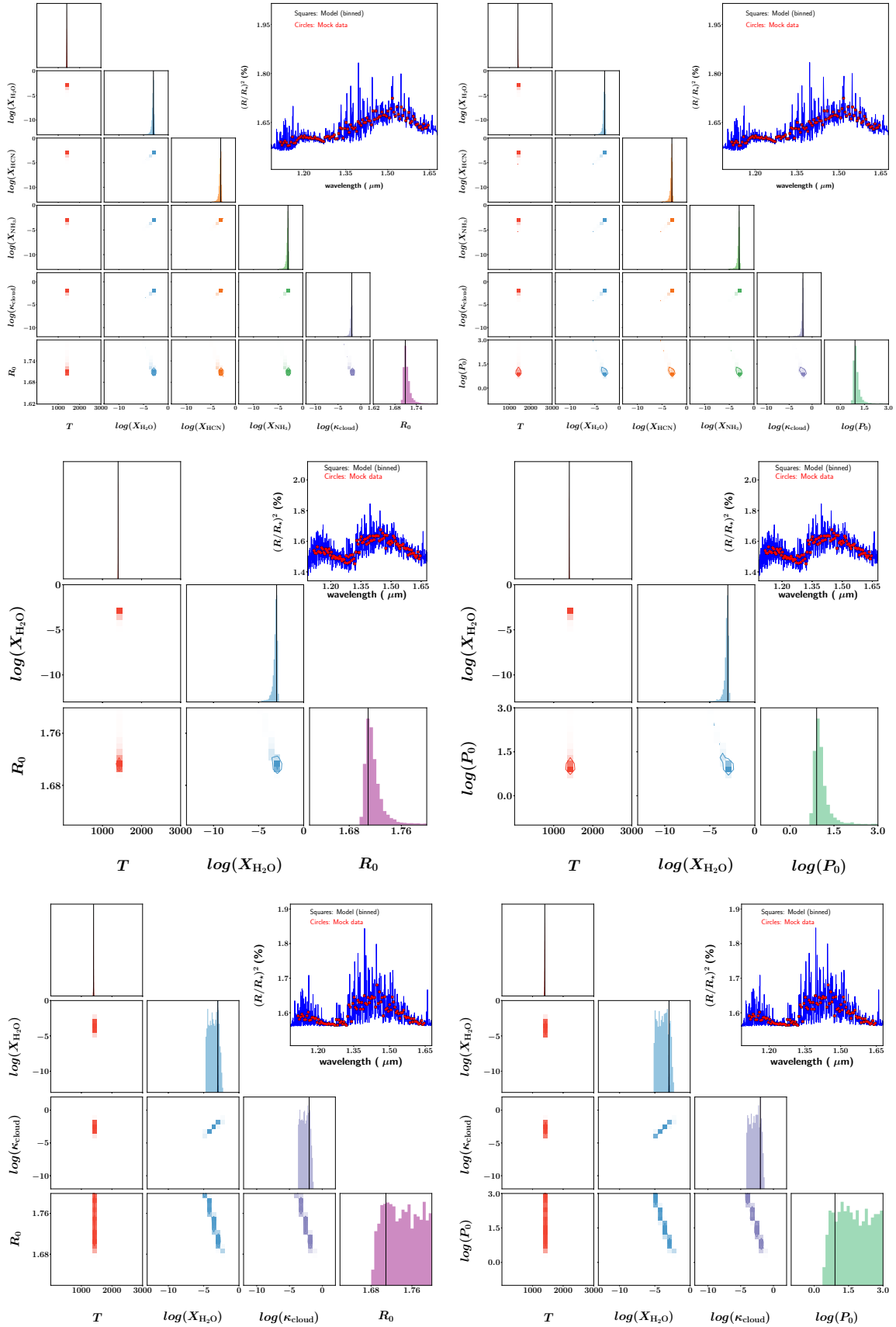
Figure 10. High-resolution (JWST-like) mock retrievals for WASP-17b using $R_0 = 1.709\ R_J$ and $P_0 = 8$ bar. The left column of retrievals hold P_0 fixed at 8 bar and fit for R_0, while the right column holds R_0 fixed at 1.709 R_J and fit for P_0. The top, middle and bottom rows are for three molecules with grey clouds, water only (cloud-free) and water only with grey clouds, respectively. All mock retrievals assume isothermal atmospheres and uncertainties of 10 ppm. Vertical lines indicate the true (input) values of the parameters.
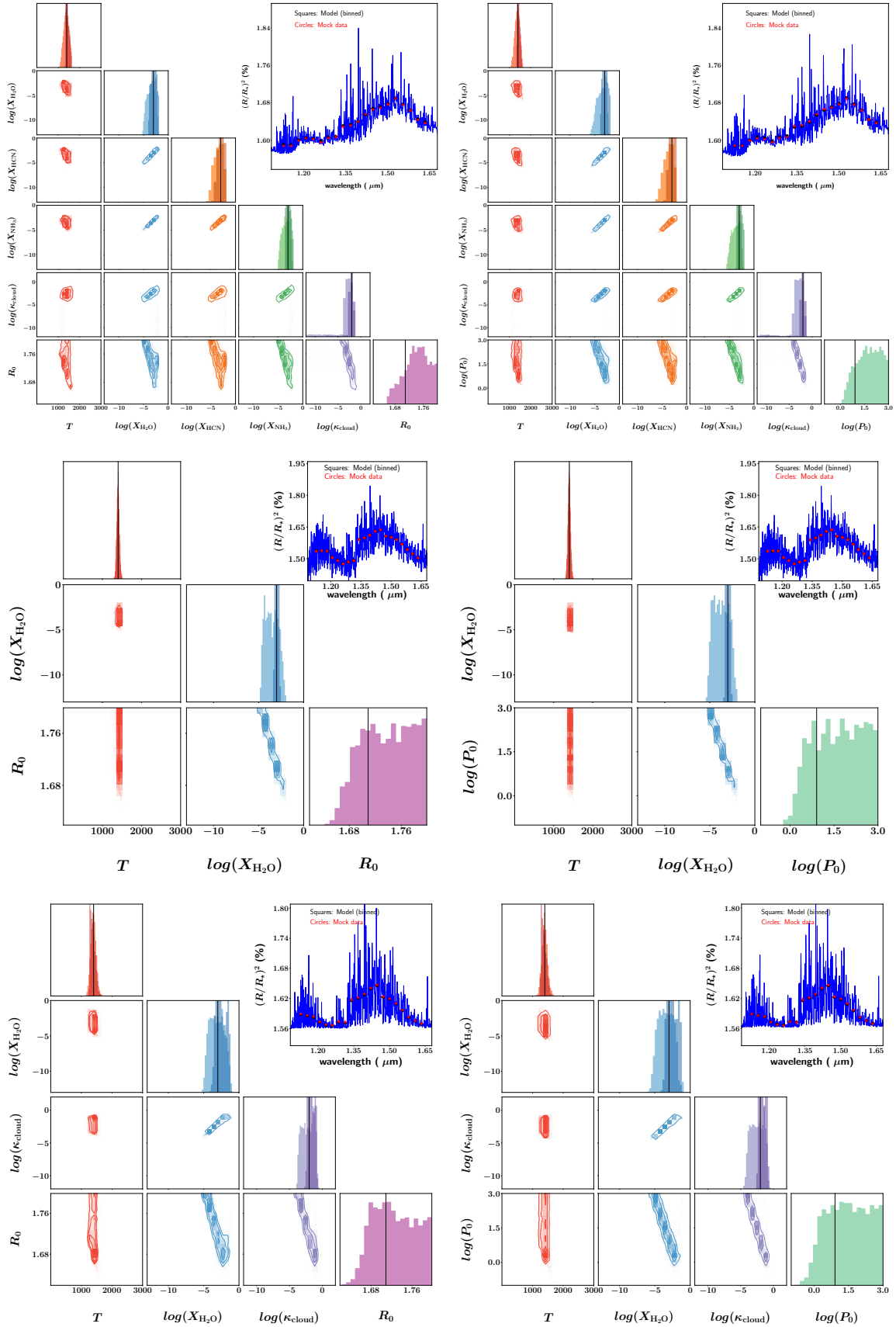
Figure 11. Same as Figure 10, but for low-resolution (WFC3-like) spectra. Additionally, the darker posterior distribution in each panel corresponds to an additional retrieval in which R_0 (left column) or P_0 (right column) is held fixed at its true value (1.709 R_J or 8 bar). The uncertainties on each mock data point is assumed to be 50 ppm.
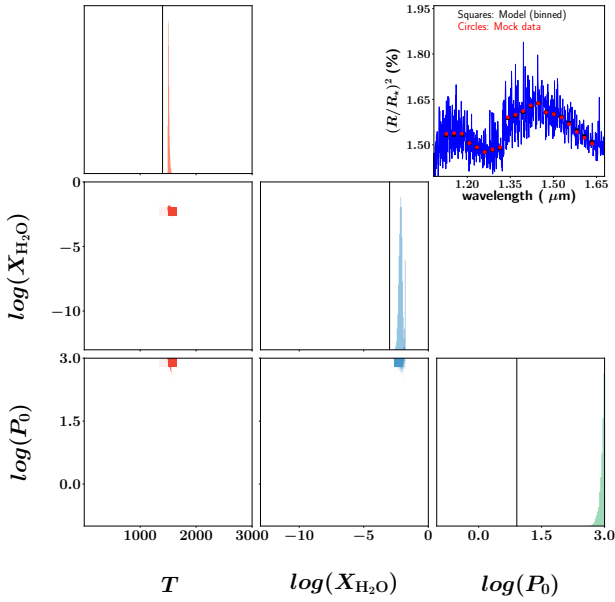
Figure 12. Low-resolution (WFC3-like) mock retrieval of WASP-17b for an isothermal, cloud-free atmosphere with water only. The mock spectrum is created using $R_0 = 1.709\ R_J$ and $P_0 = 8$ bar. The retrieval is performed with a fixed value of R_0 that is reduced by 10% to 1.5381 R_J. The corresponding value of P_0 now lies outside of its assumed prior range (0.1–1000 bar). It is an illustration of how a bad assumption on R_0 can lead to an erroneous retrieval outcome. We emphasise that there is no unique value of R_0 one can assume, but it is related to P_0 via hydrostatic equilibrium. Retrievals with different R_0-P_0 pairs should yield the same outcome as long as the prior range of values of P_0 is set correctly.

Figure 13. Low-resolution (WFC3-like) mock retrieval of WASP-17b for an isothermal atmosphere with grey clouds and all three molecules present. The pressure associated with CIA is a fitting parameter of the retrieval; its true value is $P_{\mathrm{CIA}} = 0.1$ bar. The darker posterior distribution in each panel corresponds to an additional retrieval in which P_{CIA} is held fixed at its true value. Vertical lines indicate the true (input) values of the parameters.

present, grey clouds and CIA. We set the pressure associated with CIA at 0.1 bar, but allow the retrieval to treat this pressure as a fitting parameter (P_{CIA}). Figure 13 shows that the retrieval outcome is insensitive to the retrieved value of P_{CIA}. Similar to our treatment of pressure broadening, we set the pressure associated with CIA to be 0.1 bar for the rest of the study with the reasoning that any deviations from this value may be visualised as errors that are subsumed into the grey cloud opacity. Figure 10 of Tsiaras et al. (2018) shows that CIA contributes a roughly flat continuum to the WFC3 spectrum.

3.2.5 Retrieval analysis of WASP-17b WFC3 transmission spectrum

Following our suite of tests, we now perform a full retrieval analysis on the WFC3 transmission spectrum of WASP-17b using a hierarchy of models. Additionally, we attempt to fit the spectrum with a flat line (one parameter only) and compute its corresponding Bayesian evidence. We see that there is weak evidence against the flat-line fit, but several models are consistent with the data (Figure 14). The isothermal model atmosphere with water only and grey clouds has the highest Bayesian evidence, which motivates us to display the posterior distributions of parameters associated with it in Figure 14. Alongside this retrieval, we perform a second
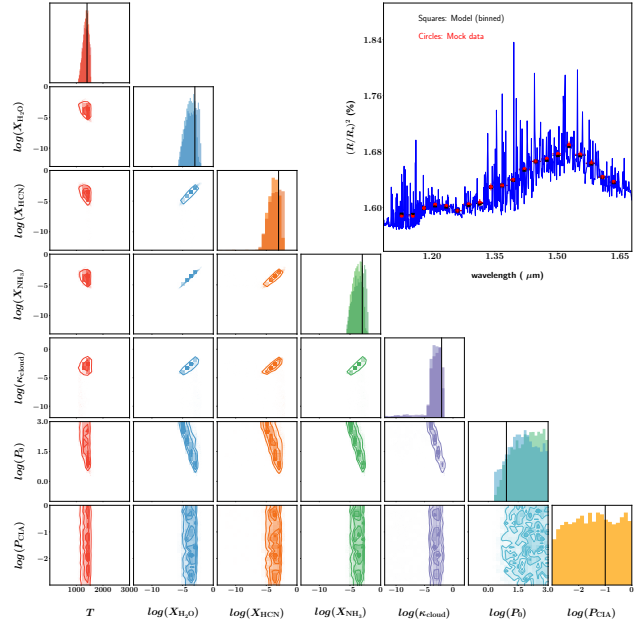
retrieval where $R_0 = 1.709\ R_J$ and $P_0 = 8$ bar as derived using the optical spectral slope. The posterior distribution for P_0 is only loosely constrained. The median value of P_0 is a factor of about 6 larger than its true value (8 bar); its best-fit value hits the upper boundary of the prior distribution at 1000 bar.

Yet, despite this inaccuracy in retrieving P_0, the posterior distributions of the pair of retrievals agree well. This is somewhat surprising, because in our mock, low-resolution retrievals of WASP-17b we discovered that the volume mixing ratio of water is prior-dominated on its lower bound (and corresponds to the upper limit set on the prior of P_0). To investigate this issue further, we ran an additional mock retrieval where the uncertainty on each data point is 200 ppm, instead of 50 ppm. Figure 15 shows that the pair of retrievals now have posterior distributions that are more similar to each other, which implies that the retrieval with variable P_0 is no longer as prior-dominated because there is now a larger parameter space of possibilities available to fit the mock spectrum. However, the retrieval outcomes are still better (the posterior distributions are narrower) when the uncertainties are smaller. The lesson learnt is that the lower bounds to volume mixing ratios retrieved from WFC3 transmission spectra may (or may not) be prior-dominated, depending on the measurement uncertainties.

3.2.6 Retrieval analysis of WASP-31b WFC3 transmission spectrum

Since WASP-31b is the other object in our sample where we can robustly derive R_0 and P_0 from the optical spectral
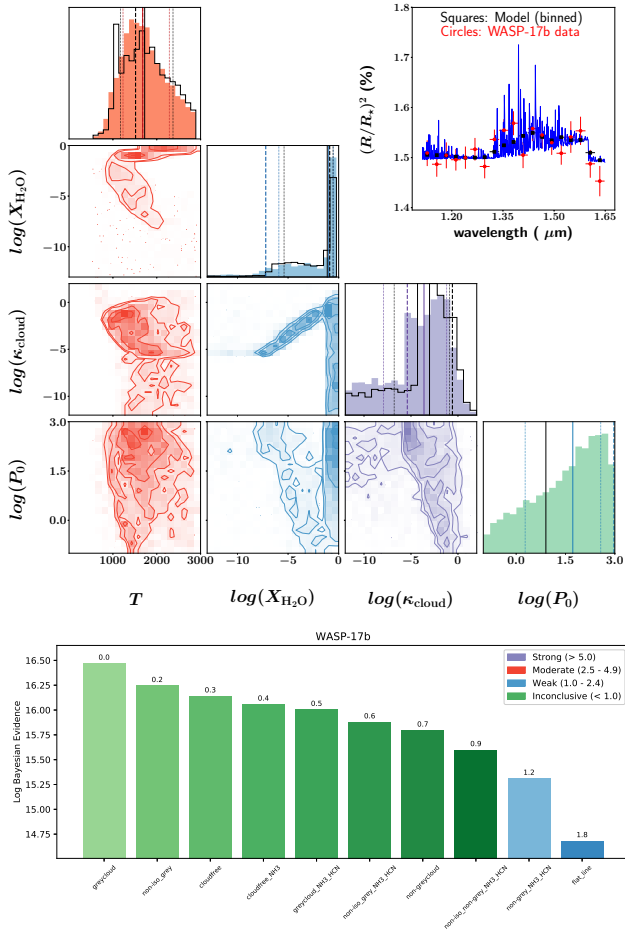
Figure 14. Full atmospheric retrieval analysis of WASP-17b. Top panel: Posterior distributions of parameters for the isothermal model with water only and grey clouds, which has the highest Bayes factor among the model hierarchy. The vertical solid line is the median value of each posterior, while the vertical dotted lines are the 1-σ uncertainties. The vertical dashed line is the best-fit value of each posterior. Also shown is a second retrieval where the R_0-P_0 relationship is determined by the values derived using optical data (see text). Bottom panel: Logarithm of the Bayesian evidence and corresponding Bayes factor between each model compared to the best model.

slope, we subject it to the same retrieval analysis we performed for WASP-17b. In Figure 16, we again subject the WFC3 transmission spectrum to a hierarchy of retrievals. Again, the isothermal model with water only and grey clouds has the highest Bayesian evidence. Two key differences are that the flat-line fit is not ruled out and that there is moderate evidence against cloud-free models. As before, we perform a second retrieval with $R_0 = 1.379\ R_J$ and $P_0 = 26$ bar. The median value of P_0 is about 16 bar and the best-fit value of P_0 almost hits the prior boundary at 594 bar, but despite these outcomes the posterior distributions of parameters from the pair of retrievals agree surprisingly well.

Our general conclusions from studying WASP-17b and WASP-31b are that P_0 can be robustly used as a fitting parameter as long as one's guess for R_0 corresponds to the range of prior values set on P_0. Even if P_0 is not tightly constrained, the posterior distributions of the other param-
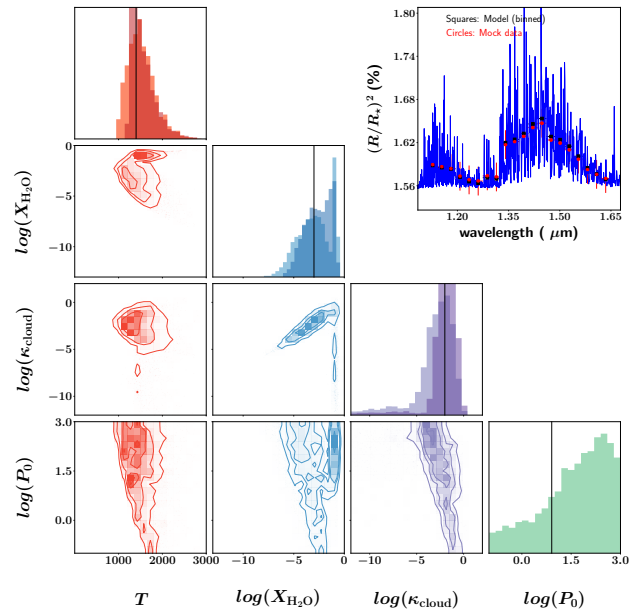
Figure 15. Additional low-resolution (WFC3-like) mock retrieval analysis of WASP-17b, but with larger uncertainties on each data point (200 ppm instead of 50 ppm). The darker posterior distribution in each panel corresponds to a second retrieval in which R_0 is held fixed at its true value (1.709 R_J). This pair of retrievals should be compared to the lower right panel of Figure 11.

eters are, despite the low spectral resolution of the WFC3 transmission spectra.

3.3 Comparison of retrieval models for the other 36 WFC3 transmission spectra

Following our suites of tests in §3.1 and §3.2.2, as well as our retrieval analyses of WASP-17b (§3.2.5) and WASP-31b (§3.2.6), we now apply our retrieval technique to the other 36 WFC3 transmission spectra in our sample. For each object, we use the value of R_0 listed in Table 3 and allow P_0 to be a fitting parameter (with a log-uniform prior between 0.1 and 1000 bar).

The results are shown in Table 2, where the parameter values shown are the median and 1σ uncertainties from the best model (highest Bayesian evidence). Additionally, we ask several questions of the outcome. An atmosphere is deemed to be cloudy if all of the cloud-free models have Bayes factors of unity or more. Cloud-free atmospheres have *only* cloud-free models with Bayes factors of less than unity. If the models with Bayes factors of less than unity are a mixture of cloudy and cloud-free, then we tag the object with "Maybe". For non-grey clouds, our criterion is stricter: it refers only to objects where *only* non-grey cloud models have Bayes factors of less than unity.

If the flat-line fit has a Bayes factor of less than unity, then we deem the retrieval to be inconclusive. In these cases, we do not report any retrieved properties of the WFC3 transit chord.
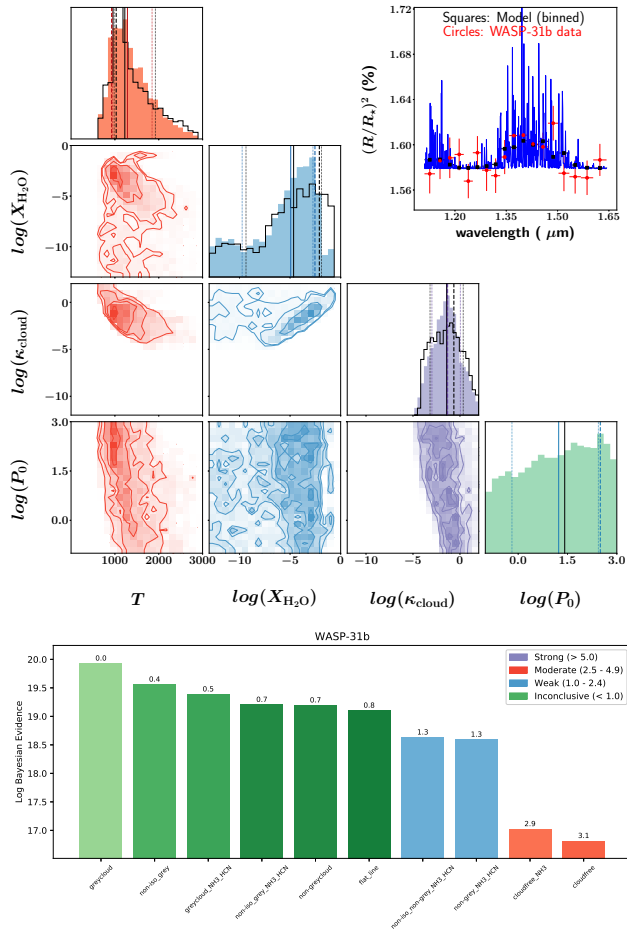
Figure 16. Same as Figure 14, but for WASP-31b.

3.3.1 Ammonia may mimic cloudiness

By visual inspection of the atmospheric opacities (Figure 2), we had suspected that it would be possible for ammonia to mimic the flattening of the spectral continuum blue-wards of the 1.4 μm water feature. Note that the one-parameter flat-line fits are disfavoured. Figure 17 shows 4 examples of objects (HAT-P-1b, HAT-P-3b, HAT-P-41b and XO-1b) where the Bayes factor between the model with grey clouds and water only versus the cloud-free model with water and ammonia is below unity, indicating that there is a lack of Bayesian evidence to favour one model over the other (Trotta 2008). This interpretation holds also for WASP-17b (Figure 14), WASP-19b (Figure 20), HAT-P-38b and HD149026b (Figure 23) and HAT-P-11b (Figure 25).

With WFC3 transmission spectra, the cautionary tale is that cloudiness may be mimicked by the presence of ammonia and this occurs for 9 out of 38 objects in our sample.

3.3.2 Prototypical hot Jupiters

HD 189733b and HD 209458b are among the most studied hot Jupiters so far. In Figure 18, we see that the WFC3 data definitively rules out cloud-free WFC3 transit chords for HD 209458b, and weakly rules them out for HD 189733b. The simplest cloudy model, which is that of an isothermal

atmosphere with grey clouds and water only, explains the WFC3 data well for both prototypical hot Jupiters. For HD 209458b, our retrieved temperature of ≈ 800 K is roughly consistent with MacDonald & Madhusudhan (2017), but our retrieved water abundance of $\log(X_{H_2O}) = -2.65^{+0.81}_{-1.24}$ is more than two orders of magnitude higher than their retrieved value of $-5.24^{+0.36}_{-0.27}$. It is unclear how to compare these values, because it is unclear how MacDonald & Madhusudhan (2017) have broken the normalisation degeneracy. Unlike MacDonald & Madhusudhan (2017), we find a lack of evidence for the detection of either NH$_3$ or HCN. For example, the isothermal model with grey clouds and water only versus that with all three molecules have a Bayes factor of 0.5, indicating that one cannot favour one model over the other (Trotta 2008). For HD 189733b, we compare our results with those of Madhusudhan et al. (2014) in §4.5.

3.3.3 Early Release Science (ERS) objects

WASP-39b and WASP-43b are among the ERS objects proposed for JWST (Batalha et al. 2017). WASP-63b is an ERS object for HST (Kilpatrick et al. 2017). Additionally, WASP-43b is one of the few hot Jupiters to have multi-wavelength phase curves from HST, due to its sub-day orbit that circumvents the thermal breathing obstacle with HST (Stevenson et al. 2014). None of the three objects are cloud-free in the WFC3 bandpass, and the simplest cloudy model fits the WFC3 data well. There is no definitive evidence for the detection of either HCN or NH$_3$. For WASP-63b, this is consistent with the analysis of Kilpatrick et al. (2017). For WASP-43b, our retrieved $\log(X_{H_2O}) = -2.89^{+1.13}_{-3.07}$ is broadly consistent with the $-3.6^{+0.8}_{-0.9}$ value reported by Kreidberg et al. (2014b), although it should be noted that Kreidberg et al. (2014b) included carbon dioxide, carbon monoxide and methane in their analysis, while we excluded these molecules and included ammonia and hydrogen cyanide instead. Interestingly, Kreidberg et al. (2014b) reported a logarithm of the "reference pressure" of $-2.4^{+0.6}_{-0.4}$ (pressure in bar), which is broadly consistent with the pressure of 10 mbar that we assume the WFC3 bandpass to probe. It is unclear how to compare the reference pressures between the two studies.

3.3.4 Very hot Jupiters

In our sample, 4 objects have equilibrium temperatures exceeding 2000 K: WASP-12b, WASP-19b, WASP-76b and WASP-121b. For WASP-12b, the WFC3 transmission spectrum may be explained by models with HCN and NH$_3$ and also models with only water (i.e., these models all fall within Bayes factors of less than unity), which implies that we are unable to offer any estimate on the carbon-to-oxygen ratio, unlike in Kreidberg et al. (2015). Our retrieved $\log(X_{H_2O}) = -3.02^{+1.09}_{-1.36}$ is broadly consistent with the $\sim 10^{-4}$–10^{-2} value reported by Kreidberg et al. (2015). In the case of WASP-19b, a cloud-free model with water only is a viable explanation—a rare occurrence in our sample. WASP-76b is an interesting object in that several scenarios are strongly ruled out: cloud-free with either water only or water and ammonia, the simplest cloudy model, etc. In fact, it seems to show strong evidence for any model featuring a non-grey cloud.

3.3.5   Other hot Jupiters

Figures 21 and 22 show the retrieval outcomes for 7 other hot Jupiters. In all cases, cloud-free models are either unlikely or ruled out. All of these 7 objects have WFC3 transmission spectra that may be explained by model atmospheres with grey clouds, meaning that non-grey clouds are not necessary to explain the data. WASP-101b is the only object where HCN is detected at significant levels, while only upper limits are obtained on the abundances of H_2O and NH_3.

3.3.6   Saturns

Figures 23 and 24 show the retrieval outcomes for 6 Saturn-mass (0.2–$0.4 M_J$) exoplanets. WASP-39b, an ERS object, also belongs to this category. With the exception of WASP-69b, the WFC3 transmission spectra are explained by the simplest cloudy model. WASP-69b requires non-grey clouds along its transit chord to explain the WFC3 data. For HAT-P-18b, HAT-P-38b and HD 149026b, the isothermal cloud-free model with water only provides a viable explanation for the data; several other models also have Bayes factor of less than unity.

3.3.7   Neptunes

There is strong evidence against a cloud-free interpretation of the somewhat flat WFC3 transmission spectra of the exo-Neptunes GJ 436b and GJ 3470b (Figure 25). For GJ 436b, this is consistent with the findings of Knutson et al. (2014b). In fact, the WFC3 transmission spectrum of GJ 436b can simply be fit by a one-parameter flat line, rendering it impossible to report atmospheric properties in a meaningful sense. HAT-P-26b does not have a flat transmission spectrum and cloud-free interpretations are strongly ruled out (Bayes factor exceeding 5.0). Wakeford et al. (2017) previously analysed the transmission spectrum of HAT-P-26b, which includes STIS, WFC3 and Spitzer data spanning 0.5–5 μm, using a suite of models incorporating carbon monoxide, carbon dioxide, methane and water. Using the Bayesian information criterion, they disfavoured cloud-free models. Our WFC3-only analysis is consistent with the conclusion of Wakeford et al. (2017). The best model, in terms of the Bayesian evidence, is the simplest cloudy one: an isothermal atmosphere with grey clouds and water only, but a variety of cloud models have Bayes factors below unity compared to this best model.

3.3.8   Super Earths

Besides being a super Earth, GJ 1214b is the prototypical example of a flat transmission spectrum (Kreidberg et al. 2014a). The retrieval outcome in Figure 26 corroborates this view and it is unsurprisingly that a one-parameter flat-line fit suffices. In our analysis, HD 97658b is inconclusively favoured by a cloud-free model with water and NH_3, though the quantities of ammonia required to match the data may be implausibly high. More data is needed to corroborate or refute this finding.

3.3.9   TRAPPIST-1 exoplanets

de Wit et al. (2018) previously measured somewhat flat WFC3 transmission spectra for TRAPPIST-1d, e, f and g. We note an ongoing debate concerning the robustness of these measured WFC3 transmission spectra, as it has been argued that the shapes of the spectral bandheads may have been contaminated by starspots and faculae from TRAPPIST-1 (Ducrot et al. 2018; Morris et al. 2018; Rackham et al. 2018). Nevertheless, we will analyze these spectra as given. de Wit et al. (2018) ruled out cloud-free, H_2-dominated atmospheres for TRAPPIST-1d, e and f, but not for g. We wish to corroborate or refute this conclusion and also to go slightly further, by considering both Earth-like ($m = 29 \; m_H$) or H_2-dominated (variable m as defined in equation [7]) atmospheres in two separate suites of retrievals.

For Earth-like atmospheres, the WFC3 spectra are explained by the majority of the models in our hierarchy. With the exception of TRAPPIST-1d, there is weak evidence against the WFC3 transmission spectra being explained by a flat line. This is unsurprising (compared to the retrievals with H_2-dominated atmospheres), because for a nitrogen-dominated atmosphere the scale height is an order of magnitude smaller than for the H_2-dominated atmosphere, which implies that even small departures from a flat line require spectral features spanning several scale heights to explain the data. Overall, when Earth-like atmospheres are assumed, the retrieval analyses are inconclusive.

When H_2-dominated atmospheres are assumed, we rule out cloud-free atmospheres with water only for TRAPPIST-1d, e and f. For all four exoplanets, the WFC3 transmission spectrum is adequately explained by a one-parameter flat-line fit, which implies that atmospheric properties cannot be meaningfully retrieved.

We do not consider arguments based on the evolution of the exoplanet or atmospheric escape, as they are out of the scope of the present study. Our inclusion of the TRAPPIST-1 exoplanets is for completeness and they will not be included in our analysis of the trends associated with the water volume mixing ratios in §3.4. However, when compiling population statistics, we will include the outcomes only from the retrievals of the TRAPPIST-1 exoplanets assuming Earth-like atmospheres.

3.4   Trends

All of the techniques developed and tests performed in this study culminate in a singular result: to examine if there are trends in the retrieved atmospheric properties. In particular, we wish to examine if X_{H_2O} correlates with the equilibrium temperature (T_{eq}), retrieved temperature (T) or mass of the exoplanet (M). The equilibrium temperature is a proxy for the strength of insolation or stellar irradiation. Previous studies have plotted the "metallicity" versus the exoplanet mass and claimed a correlation between the two quantities (Kreidberg et al. 2014b; Wakeford et al. 2017, 2018; Arcangeli et al. 2018; Mansfield et al. 2018; Nikolov et al. 2018).

In Figure 29, we find little to no evidence for a correlation between X_{H_2O} and M, T_{eq} or T. If the abundance of water is assumed to be a direct proxy for the elemental abundance of oxygen (see §4.6), then this outcome runs contrary to previous claims. There is a lack of correlation

between κ_{cloud} and T_{eq}, which has two implications. First, it suggests that our inferred $X_{\mathrm{H_2O}}$ values are not biased by the degree of cloudiness (or haziness) in these atmospheres. Second, the majority of atmospheric transit chords probed by WFC3 appear to have $\kappa_{\mathrm{cloud}} \sim 10^{-2}$ cm^2 g^{-1} (corresponding to ~ 10 mbar), regardless of the surface gravity or strength of insolation. The lack of evidence for non-grey clouds implies that the particle radii are $r_c \gtrsim 0.1$ μm. Overall, these outcomes may be interpreted as the transit chords being affected by haze.[5] The ratio of the retrieved to the equilibrium temperatures (T/T_{eq}) appears to have a lower limit of about 0.5.

It is unclear how to relate our results to claimed correlations between the bulk metallicity of exoplanets and their masses based on the analysis of mass-radius relations (Miller & Fortney 2011; Thorngren et al. 2016).

4 DISCUSSION

4.1 Comparison to a previous retrieval study

It is natural to compare our study to Tsiaras et al. (2018), since the WFC3 transmission spectra of 30 objects from our sample are taken from it. Furthermore, some of the modelling choices made by Tsiaras et al. (2018) are the same as ours: isothermal transit chord, nested sampling. Our cloud models differ, because Tsiaras et al. (2018) use the formulation of Lee et al. (2013), which also allows for a smooth transition between the Rayleigh and large-particle regimes but predates Kitzmann & Heng (2018), and also assume a cloud-top boundary (which we do not). Furthermore, Tsiaras et al. (2018) include methane, carbon monoxide, carbon dioxide, titanium oxide (TiO) and vanadium oxide (VO) in their retrievals in addition to water and ammonia; they do not include hydrogen cyanide. By contrast, we only include water, ammonia and hydrogen cyanide in our model hierarchy. Inevitably, these choices lead to differences in some of the retrieval outcomes.

Table 3 of Tsiaras et al. (2018) lists their retrieved water volume mixing ratios. For GJ 436b, HAT-P-12b, WASP-29b, WASP-31b, WASP-67b and WASP-80b, we do not report any retrieved atmospheric properties, unlike for Tsiaras et al. (2018), as the one-parameter flat-line fit is among the models with Bayes factors of less than unity. For GJ 3470b, HAT-P-1b, HAT-P-3b, HAT-P-11b, HAT-P-17b, HAT-P-18b, HAT-P-26b, HAT-P-32b, HAT-P-38b, HD 149026b, HD 189733b, HD 209458b, WASP-12b, HAT-P-41b, WASP-43b, WASP-52b, WASP-63b, WASP-69b, WASP-74b, WASP-101b, WASP-121b and XO-1b, our retrieved water mixing ratios are broadly consistent with those of Tsiaras et al. (2018). For WASP-39b and WASP-76b, our retrieved water mixing ratios differ at the order-of-magnitude level compared to Tsiaras et al. (2018). Interestingly, these two objects also have the highest values of the Atmospheric Detectability Index (ADI) in the Tsiaras et al. (2018) sample of 30 objects.

Of particular interest is WASP-76b, which is one of two objects in our sample that requires a non-grey cloud to fit the data. Tsiaras et al. (2018) reported that their retrieval favours a cloudfree interpretation, because the non-flat spectral continuum blueward of the 1.4-μm water feature may be fitted by the spectral features of TiO and VO. Tsiaras et al. (2018) remark that their retrieved $\log X_{\mathrm{TiO}} \sim -2.5$ is "likely unphysical". Our retrieval yields $\log X_{\mathrm{H_2O}} = -5.3 \pm 0.61$, which is inconsistent with the $\log X_{\mathrm{H_2O}} = -2.7 \pm 1.07$ reported by Tsiaras et al. (2018). The WFC3 transmission spectrum of WASP-76b demonstrates that a wider wavelength range is required to resolve the degeneracy associated with these modelling choices, which will be provided by JWST spectra.

It is unclear why our retrieval outcome for WASP-39b differs from that of Tsiaras et al. (2018), because they did not publish the full set of posterior distributions for this object, unlike for WASP-76b in their Figure 11. For example, it is unclear if the high value of the ADI for WASP-39b translates into a cloud-free interpretation (which is the case for WASP-76b).

4.2 Is there evidence for non-grey clouds? Is cloud composition constrained?

Cloud models of varying sophistication have been employed in retrieval models. Our approach is somewhat different in that we include in our hierarchy of retrievals both grey and non-grey cloud models, as well as a one-parameter flat line. For 8 out of 38 objects, the WFC3 transmission spectrum is explained by a flat line. For 35 out of 38 objects, an isothermal grey cloud model with water only is sufficient to explain the data. Only WASP-69b and WASP-76b have WFC3 transmission spectra that require an explanation by model atmospheres with non-grey clouds along the transit chord. Otherwise, there is generally no evidence for non-grey clouds being present in the sample of 38 objects.

Since the cloud composition may only be inferred for non-grey clouds, this implies that the composition is generally unconstrained, which is consistent with the conclusion drawn by Tsiaras et al. (2018). Even for WASP-69b and WASP-76b, the parameter Q_0 is largely unconstrained because it spans the entire range of values set by the prior.

Given the retrieval outcomes, our approach to not consider patchy clouds (Line & Parmentier 2016) is justified. We have also shown that the normalisation degeneracy may be broken without appealing to the more complicated patchy cloud model, which was invoked by MacDonald & Madhusudhan (2017) to break the degeneracy.

4.3 Is there evidence for non-isothermal transit chords?

For all 38 objects in our sample, we find a lack of strong Bayesian evidence to support non-isothermal transit chords probed by WFC3.

4.4 How prevalent is HCN or NH$_3$?

Based on the best model selected by the Bayesian evidence, we find that only 6 objects have tentative evidence for the detection of ammonia: HAT-P-1b, HAT-P-17b, HAT-P-38b,

---

[5] We adopt the planetary science definition of "cloud" versus "haze": the former is formed thermochemically, while the latter is formed photochemically.

HAT-P-41b, WASP-101b and HD 97658b. However, the retrieved value for HD 97658b is $\log(X_{NH_3}) = -0.48^{+0.19}_{-0.23}$, which may be unphysically high. This is unsurprising as our model contains no chemistry, so there is nothing to prevent unphysical values being retrieved. HAT-P-17b and WASP-101b also have tentative detections of hydrogen cyanide.

4.5   Subsolar water abundances in hot Jupiters?

Madhusudhan et al. (2014) previously analysed the WFC3 transmission spectra of HD 189733b, HD 209458b and WASP-12b using cloud-free retrieval models and found $\log(X_{H_2O}) = -5.20^{+1.68}_{-0.18}$, $-5.27^{+0.65}_{-0.16}$ and $-5.35^{+1.85}_{-1.99}$, respectively. They concluded that the water abundances from these hot Jupiters are subsolar by about 1–2 orders of magnitude. By contrast, our retrievals find $\log(X_{H_2O})$ values that are several orders of magnitude higher: $-2.3^{+0.87}_{-1.26}$, $-2.65^{+0.81}_{-1.24}$ and $-3.02^{+1.09}_{-1.36}$, respectively. We estimate that $\log(X_{H_2O}) \approx -3.2$ assuming chemical equilibrium, solar abundance and a pressure of 10 mbar, which suggests that our retrieved water abundances are super- rather than subsolar as claimed by Madhusudhan et al. (2014). The discrepancy arises from the retrievals of Madhusudhan et al. (2014) being cloud-free, while we have included a cloud model that smoothly transitions between the Rayleigh and large-particle regimes. It is consistent with the fact that cloud opacity diminishes the strength of the water feature, which may be negated by increasing X_{H_2O}.

4.6   What does the "metallicity" mean when interpreting spectra of exoplanetary atmospheres?

Several published studies have plotted the "metallicity" (in "solar" units) versus the mass of the exoplanet with entries from the Solar System gas and ice giants overplotted (Kreidberg et al. 2014b; Wakeford et al. 2017, 2018; Arcangeli et al. 2018; Mansfield et al. 2018; Nikolov et al. 2018). As already elucidated by Heng (2018), there are several caveats to these plots. First, the "metallicity" is predominantly O/H in these studies. Second, the "mixing ratio of water at solar abundance" is a temperature- and pressure-dependent statement. Given a fixed value of O/H, the mixing ratio of water still depends on temperature and pressure. In other words, it is a *function* and not a number. Third, the conversion factor between the water mixing ratio and O/H is not always unity and depends on the elemental abundances (O/H, C/H, etc), carbon-to-oxygen ratio, temperature, pressure, photochemistry, atmospheric mixing, condensation, etc. For all of these reasons, we have chosen to present our retrieved water abundances as they are in Figure 29, rather than convert them to O/H.

REFERENCES

Abramowitz, M., & Stegun, I.A. 1970, Handbook of Mathematical Functions, 9th printing (New York: Dover Publications)

Anderson, D.R., Collier Cameron, A., Hellier, C., et al. 2011, A&A, 531, A60

Anderson, D.R., Collier Cameron, A., Delrez, L., et al. 2014, MNRAS, 445, 1114

Anglada-Escudé, G., Rojas-Ayala, B., Boss, A.P., Weinberger, A.J., & Lloyd, J.P. 2013, A&A, 551, A48

Arcangeli, J., Désert, J.-M., Line, M.R., et al. 2018, ApJL, 855, L30

Arfken, G.B., & Weber, H.J. 1995, Mathematical Methods for Physicists, fourth edition (San Diego: Academic Press)

Barber, R.J., Tennyson, J., Harris, G.J., & Tolchenov, R.N. 2006, MNRAS, 368, 1087

Barber, R.J., Strange, J.K., Hill, C., et al. 2014, MNRAS, 437, 1828

Bakos, G.A., Torres, G., Pál, A., et al. 2010, ApJ, 710, 1724

Batalha, N., Bean, J., Stevenson, K., et al. 2017, JWST Proposal ID 1366, Cycle 0 Early Release Science

Benneke, B., & Seager, S. 2012, ApJ, 753, 100

Benneke, B., & Seager, S. 2013, ApJ, 778, 153

Bétrémieux, Y., & Swain, M.R. 2017, MNRAS, 467, 2834

Biddle, L.I., Pearson, K.A., Crossfield, I.J.M., et al. 2014, MNRAS, 443, 1810

Boyajian, T., von Braun, K., Feiden, G. A., et al. 2015, MNRAS, 447, 846

Brown, T.M. 2001, ApJ, 553, 1006

Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125

Chan, T., Ingemyr, M., Winn, J.N., et al. 2011, AJ, 141, 179

Cox, A.N. 2000, Allen's Astrophysical Quantities, 4th edition (New York: Springer-Verlag)

Delrez, L., Santerne, A., Almenara, J.-M., et al. 2016, MNRAS, 458, 4025

de Wit, J., & Seager, S. 2013, Science, 342, 1473

de Wit, J., Wakeford, H.R., Lewis, N., et al. 2018, Nature Astronomy, 2, 214

Ducrot, E., Sestovic, M., Morris, B.M., et al. 2018, arXiv:1807.01402

Esposito, M., Covino, E., Mancini, L., et al. 2014, A&A, 564, L13

Feroz, F., & Hobson, M.P. 2008, MNRAS, 384, 449

Feroz, F., Hobson, M.P., & Bridges, M. 2009, MNRAS, 398, 1601

Feroz, F., Hobson, M.P., Cameron, E., Pettitt, A.N. 2013, arXiv:1306.2144

Foreman-Mackey, D., Hogg, D.W., Lang, D., & Goodman, J. 2013, PASP, 125, 306

Fortney, J.J., et al. 2016, White Paper for NASA's Nexus for Exoplanet System Science (NExSS) (arXiv:1602.06305)

Freedman, R.S., Lustig-Yaeger, J., Fortney, J.J., et al. 2014, ApJS, 214, 25

Fu, G., Deming, D., Knutson, H., et al. 2017, ApJL, 847, L22

Griffith, C.A. 2014, Philosophical Transactions of the Royal Society A, 372, 86

Grimm, S.L., & Heng, K. 2015, ApJ, 808, 182

Grimm, S.L., Demory, B.-O., Gillon, M., et al. 2018, A&A, 613, A68

Hartman, J. D., Bakos, G. A., Torres, G., et al. 2009, ApJ, 706, 785

Hartman, J.D., Bakos, G.A., Kipping, K.M., et al. 2011a, ApJ, 728, 138

Hartman, J.D., Bakos, G.A., Torres, G., et al. 2011b, ApJ, 742, 59

Hartman, J.D., Bakos, G.A., Béky, B., et al. 2012, AJ, 144, 139

Hebb, L, Collier-Cameron, A., Loeillet, B., et al. 2009, ApJ, 693, 1920
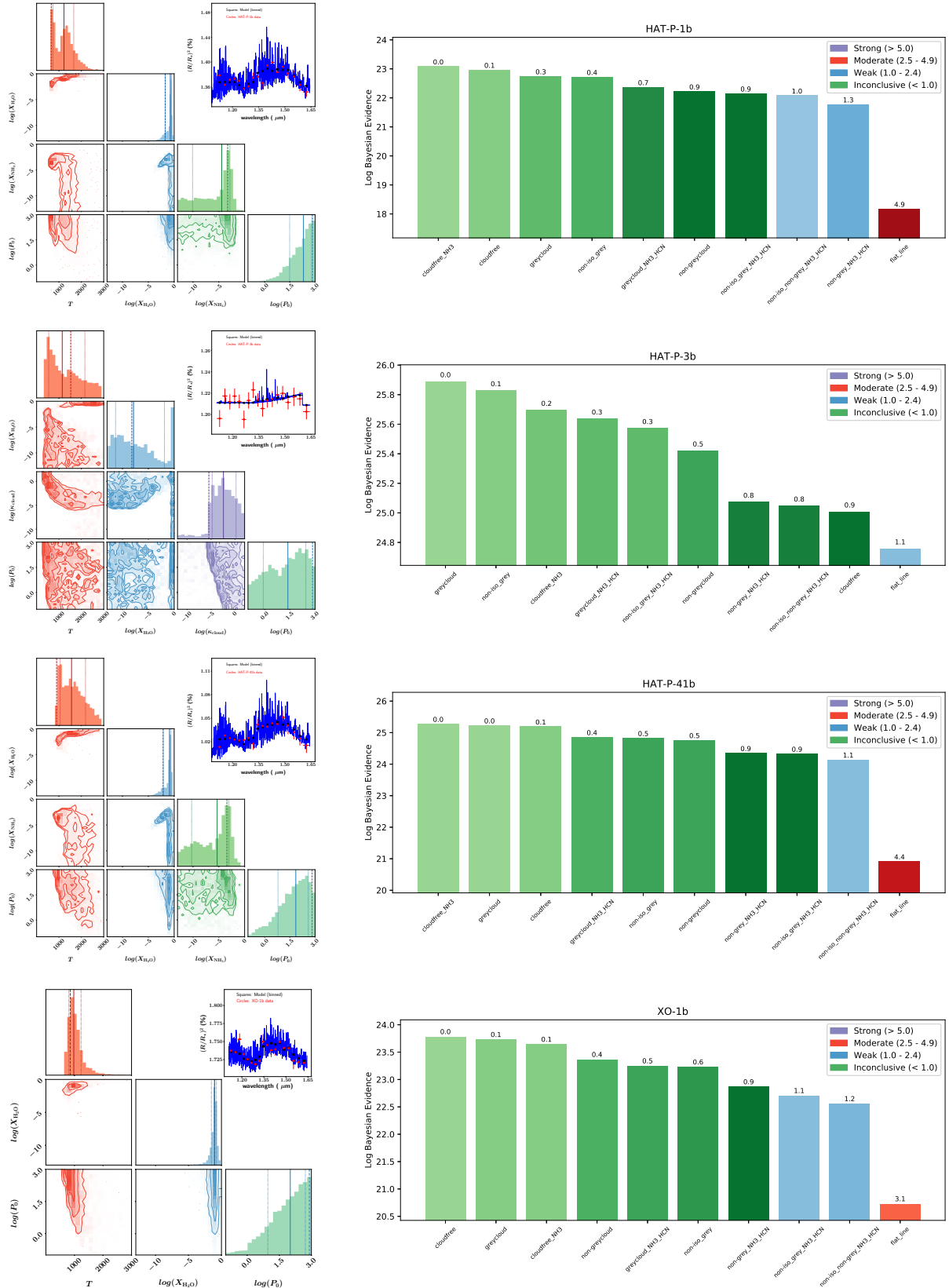
Figure 17. Left column: Posterior distributions and synthetic spectrum for the best model (as selected by the Bayesian evidence). Right column: Comparison of Bayesian evidence for objects for which it is not possible to distinguish between cloudy atmospheres containing water only versus cloud-free atmospheres with both water and ammonia present. The solid, dotted and dashed vertical lines represent the median value, the 1-σ uncertainties associated with the median and the best-fit value of each posterior distribution, respectively. XO-1b is one of two objects with the highest Bayesian evidence for the cloudfree, isothermal model with water only (excluding the TRAPPIST-1 exoplanets).
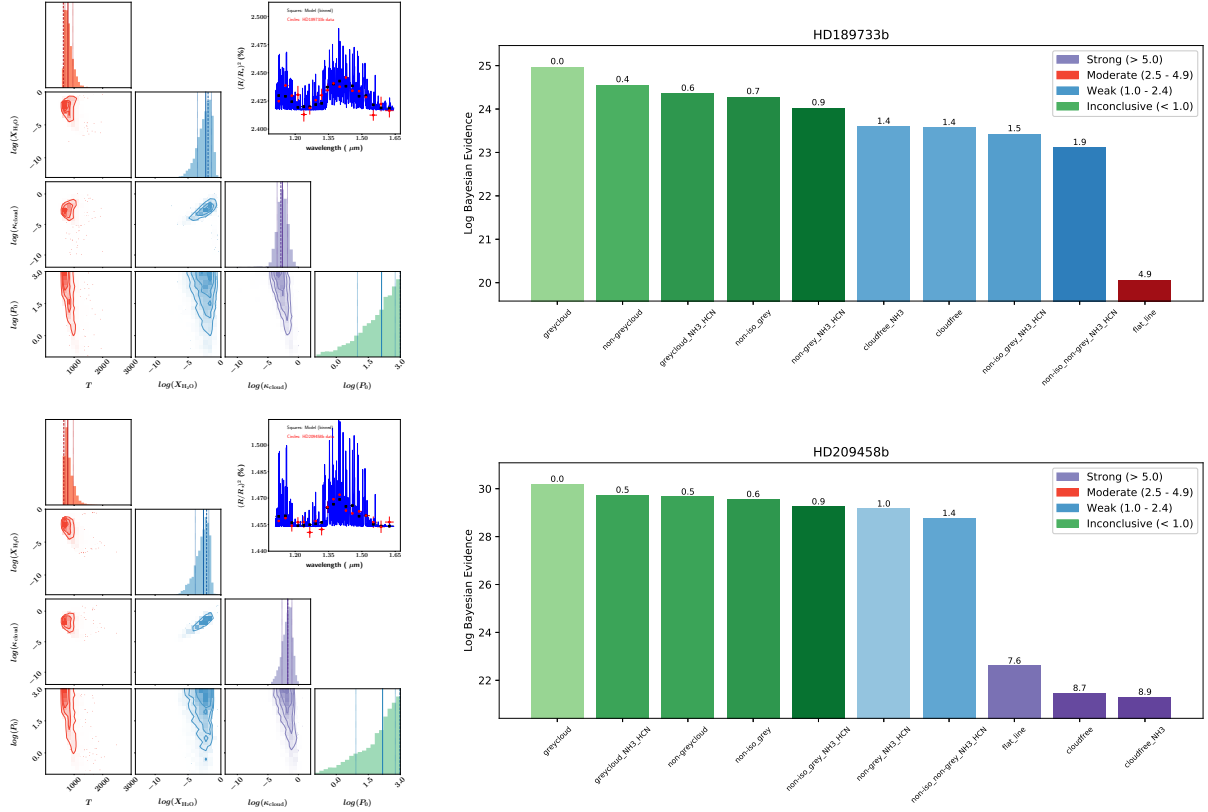
Figure 18. Same as Figure 17, but for the prototypical hot Jupiters HD 189733b and HD 209458b.

Hébrard, G., Collier Cameron, A., Brown, D.J.A., et al. 2013, A&A, 549, A134

Hellier, C., Anderson, D.R., Collier Cameron, A., et al. 2010, ApJL, 723, L60

Hellier, C., Anderson, D.R., Collier Cameron, A., et al. 2011, A&A, 535, L7

Hellier, C., Anderson, D.R., Collier Cameron, A., et al. 2012, MNRAS, 426, 739

Hellier, C., Anderson, D.R., Collier Cameron, A., et al. 2014, MNRAS, 440, 1982

Hellier, C., Anderson, D.R., Collier Cameron, A., et al. 2015, AJ, 150, 18

Heng, K. 2016, ApJL, 826, L16

Heng, K., & Tsai, S.-M. 2016, ApJ, 829, 104

Heng, K., & Kitzmann, D. 2017, MNRAS, 470, 2972

Heng, K. 2017, Exoplanetary Atmospheres: Theoretical Concepts and Foundations (Oxford: Princeton University Press)

Heng, K. 2018, RNAAS, 2, 128

Howard, A.W., Bakos, G.A., Hartman, J., et al. 2012, ApJ, 749, 134

Hubbard, W.B., Fortney, J.J., Lunine, J.I., et al. 2001, ApJ, 560, 413

Huitson, C.M., Sing, D.K., Pont, F., et al. 2013, MNRAS, 434, 3252

Iyer, A.R., Swain, M.R., Zellem, R.T., et al. 2016, ApJ, 823, 109

Johnson, J. A., Winn, J. N., Narita, N., et al. 2008, ApJ, 686, 649

Kilpatrick, B.M., Cubillos, P.E., Stevenson, K.B., et al. 2018, AJ, 156, 103

Kitzmann, D., & Heng, K. 2018, MNRAS, 475, 94

Knutson, H.A., Dragomir, D., Kreidberg, L., et al. 2014, ApJ, 794, 155

Knutson, H.A., Benneke, B., Deming, D., et al. 2014, Nature,

505, 66

Kreidberg, L., Bean, J.L., Désert, J.-M., et al. 2014a, Nature, 505, 69

Kreidberg, L., Bean, J.L., Désert, J.-M., et al. 2014b, ApJL, 793, L27

Kreidberg, L., Line, M.R., Bean, J.L., et al. 2015, ApJ, 814, 66

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., Knutson, H., Deming, D., Wilkins, A., & Desert, J.-M. 2013, ApJ, 778, 183

Line, M.R., & Parmentier, V. 2016, ApJ, 820, 78

MacDonald, R.J., & Madhusudhan, N. 2017, MNRAS, 469, 1979

Maciejewski, G., Dimitrov, D., Mancini, L., et al. 2016, AcA, 66, 55

Madhusudhan, N., & Seager, S. 2009, ApJ, 707, 24

Madhusudhan, N., Crouzet, N., McCullough, P.R., Deming, D., & Hedges, C. 2014, ApJL, 791, L9

Mandell, A.M., Haynes, K., Sinukoff, E., et al. 213, ApJ, 779, 128

Mansfield, M., Bean, J.L., Line, M.R., et al. 2018, AJ, 156, 10

Miller, N., & Fortney, J.J. 2011, ApJL, 736, L29

Morris, B.M., Agol, E., Davenport, J.R.A., & Hawley, S.L. 2018, ApJ, 857, 39

Nikolov, N., Sing, D.K., Fortney, J.J., et al. 2018, Nature, 557, 526

Rackham, B.V., Apai, D., & Giampapa, M.S. 2018, ApJ, 853, 122

Rothman, L.S., Gamache, R.R., Goldman, A., et al. 1987, Applied Optics, 26, 4058

Rothman, L.S., Gamache, R.R., Tipping, R.H., et al. 1992, Jour-
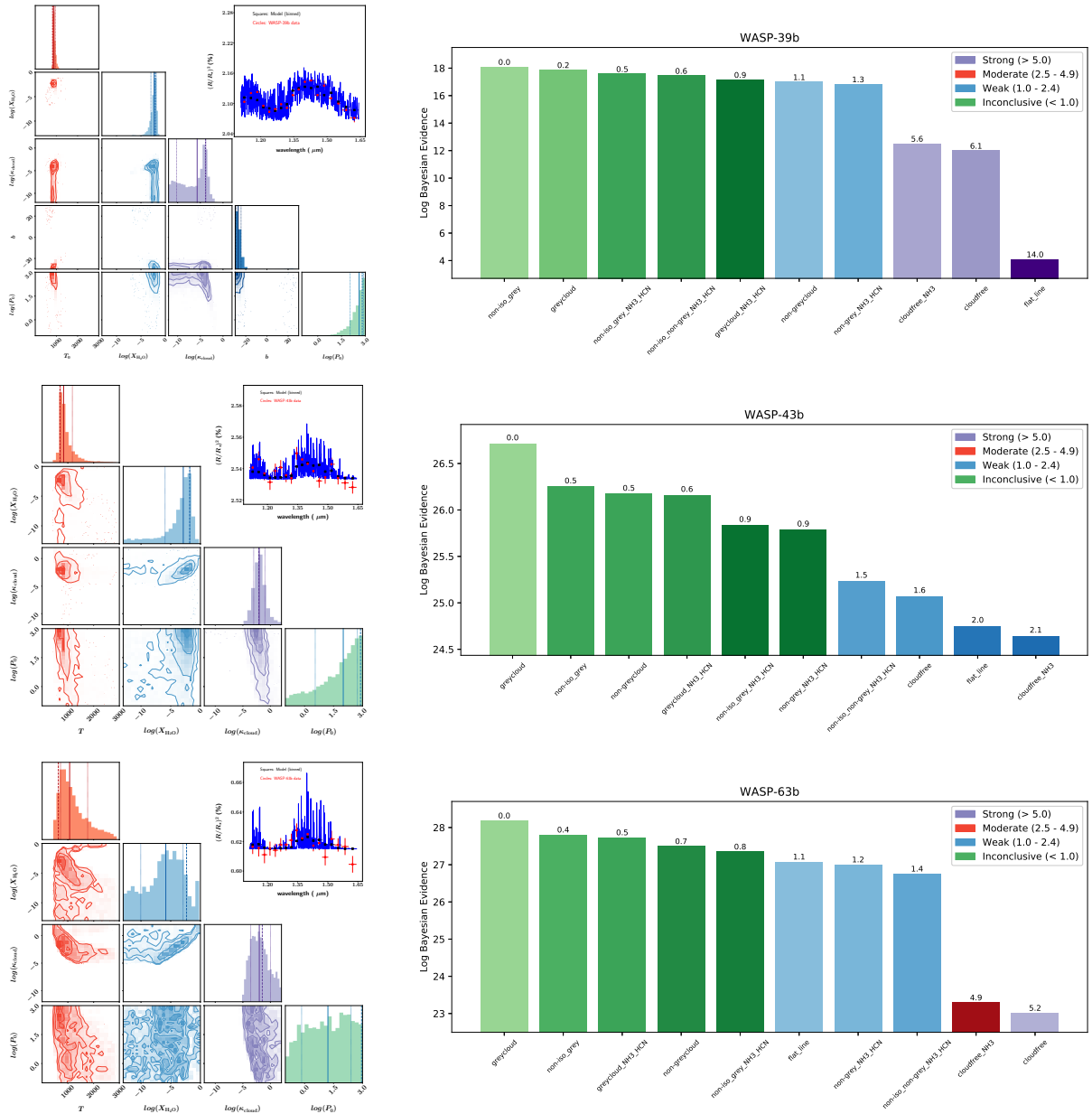
Figure 19. Same as Figure 17, but for Early Release Science (ERS) objects: WASP-39b and WASP-43b (for JWST) and WASP-63b (for HST).

nal of Quantitative Spectroscopy & Radiative Transfer, 48, 469

Rothman, L.S., Rinsland, C.P., Goldman, A., et al. 1996, Journal of Quantitative Spectroscopy & Radiative Transfer, 60, 665

Rothman, L.S., Barbe, A., Benner, D.C., et al. 2003, Journal of Quantitative Spectroscopy & Radiative Transfer, 82, 5

Rothman, L.S., Jacquemar, D., Barbe, A., et al. 2005, Journal of Quantitative Spectroscopy & Radiative Transfer, 96, 139

Rothman, L.S., Gordon, I.E., Barber, R.J., et al. 2010, Journal of Quantitative Spectroscopy & Radiative Transfer, 111, 2139

Rothman, L.S., Gordon, I.E., Barbe, A., et al. 2009, Journal of Quantitative Spectroscopy & Radiative Transfer, 110, 533

Rothman, L.S., Gordon, I.E., Babikov, Y., et al. 2013, Journal of Quantitative Spectroscopy & Radiative Transfer, 130, 4

Sato, B., Hartman, J.D., Bakos, G.A., et al. 2012, PASJ, 64, 97

Sing, D.K., Fortney, J.J., Nikolov, N., et al. 2016, Nature, 529,

59

Skilling, J. 2006, Bayesian Analysis, 1, 833

Sneep, M., & Ubachs, W. 2005, Journal of Quantitative Spectroscopy & Radiative Transfer, 92, 293

Southworth, J., Hinse, T.C., Dominik, M., et al. 2012, MNRAS, 426, 1338

Stevenson, K.B., Désert, J.-M., Line, M.R., et al. 2014, Science, 346, 838

Tennyson, J., & Yurchenko, S.N. 2017, Molecular Astrophysics, 8, 1

Thorngren, D.P., Fortney, J.J., Murray-Clay, R., & Lopez, E.D. 2016, ApJ, 831, 64

Torres, G., Winn, J.N., & Holman, M.J. 2008, ApJ, 677, 1324

Tregloan-Reed, J., Southworth, J., & Tappert, C. 2013, MNRAS, 428, 3671

Triaud, A.H.M.J., Gillon, M., Ehrenreich, D., et al. 2015, MN-
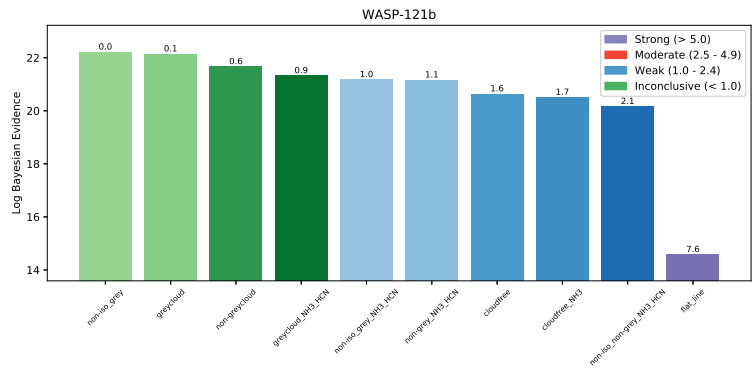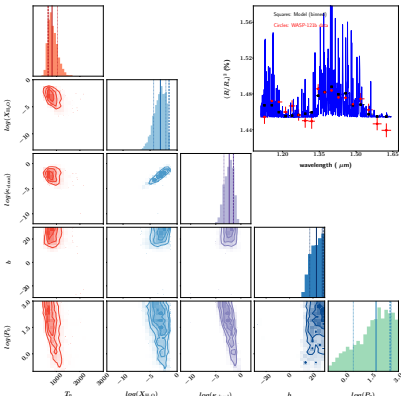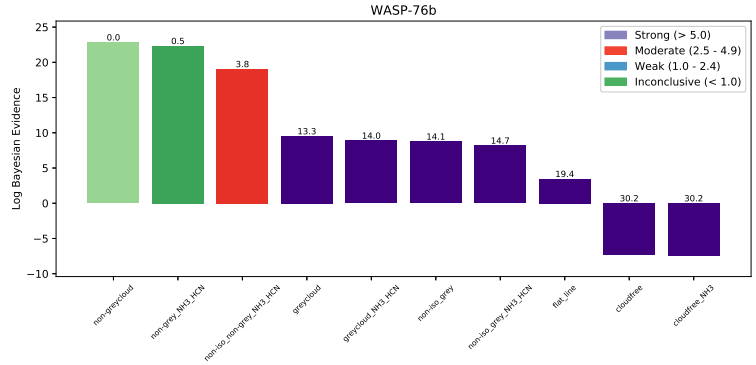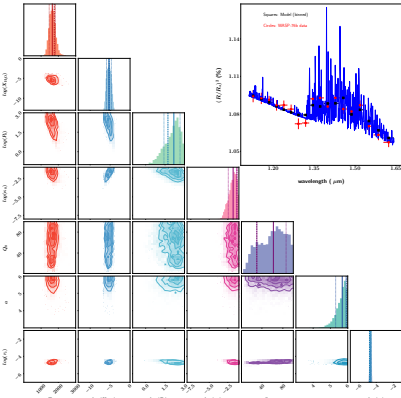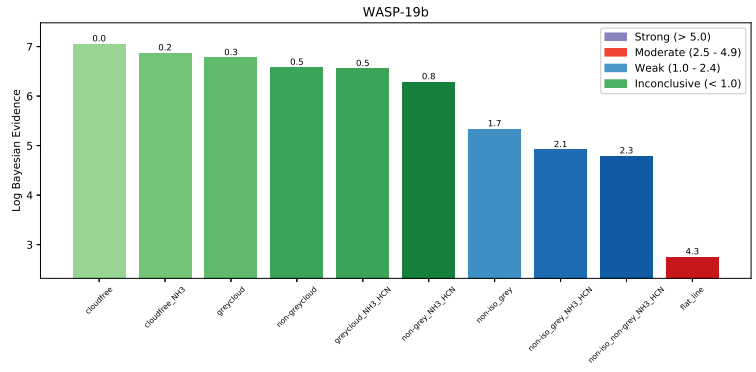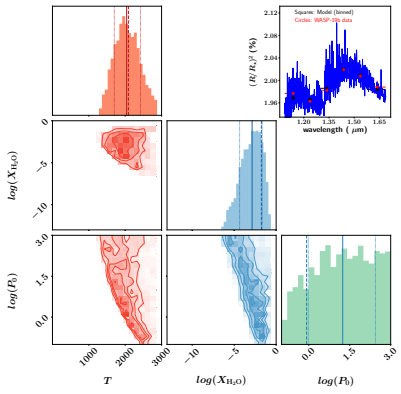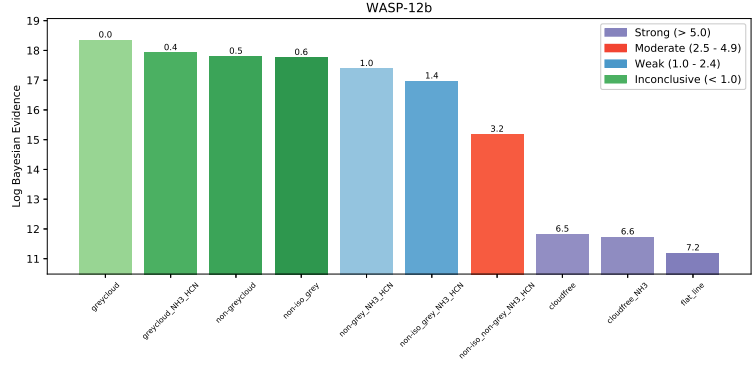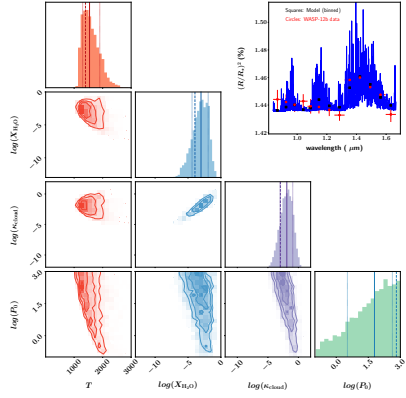
Figure 20. Same as Figure 17, but for very hot Jupiters ($T_{eq} > 2000$ K): WASP-12b, WASP-19b, WASP-76b and WASP-121b. WASP-19b is one of two objects with the highest Bayesian evidence for the cloudfree, isothermal model with water only (excluding the TRAPPIST-1 exoplanets). WASP-76b is one of two objects where non-grey clouds are needed to fit the data.
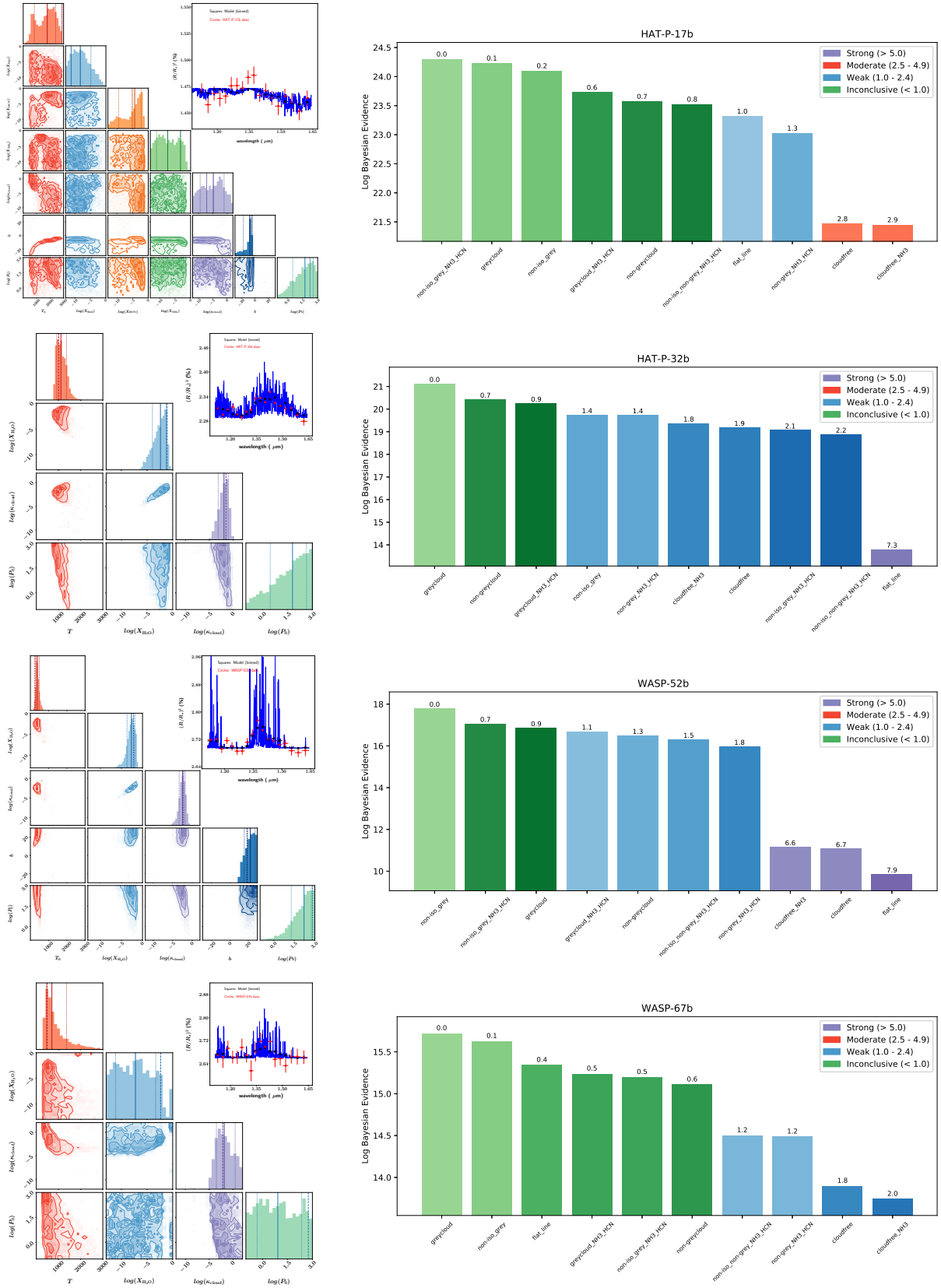
Figure 21. Same as Figure 17, but for the rest of the hot Jupiters: HAT-P-17b, HAT-P-32b, WASP-52b and WASP-67b.
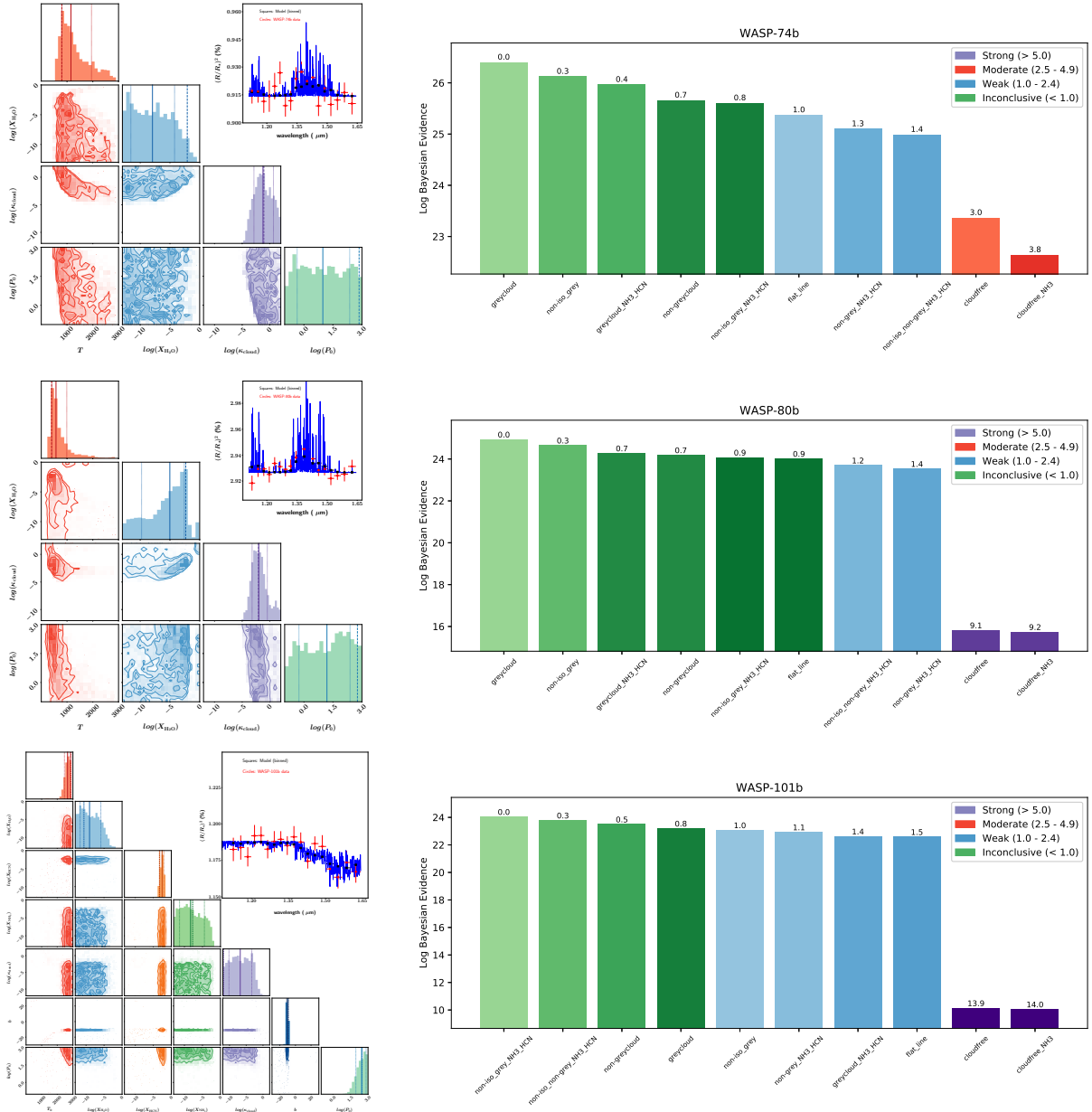
Figure 22. Continuation of Figure 21 for the rest of the hot Jupiters: WASP-74b, WASP-80b and WASP-101b. WASP-101b stands out as the only object for which HCN is significantly detected over water and ammonia.

RAS, 450, 2279

Trotta, R. 2008, Contemporary Physics, 49, 71

Tsiaras, A., Waldmann, I.P., Zingales, T., et al. 2018, AJ, 155, 156

van Grootel, V., Gillon, M., Valencia, D., et al. 2014, ApJ, 786, 2

van Grootel, V., Fernandes, C.S., Gillon, M., et al. 2018, ApJ, 853, 30

von Braun, K., Boyajian, T.S., Kane, S.R., et al. 2012, ApJ, 753, 171

Waldmann, I.P., Tinetti, G., Rocchetto, M., et al. 2015, ApJ, 802, 107

Wakeford, H.R., Sing, D.K., Kataria, T., et al. 2017, Science, 356, 628

Wakeford, H.R., Sing, D.K., Deming, D., et al. 2018, AJ, 155, 29

West, R.G., Hellier, C., Almenara, J.-M., et al. 2016, A&A, 585, A126

Wyttenbach, A., Ehrenreich, D., Lovis, C., Udry, S., & Pepe, F. 2015, A&A, 577, A62

Yurchenko, S.N., Barber, R.J., & Tennyson, J. 2011, MNRAS, 413, 1828

Yurchenko, S.N., Tennyson, J., Barber, R.J., & Thiel, W. 2013, Journal of Molecular Spectroscopy, 291, 69

Yurchenko, S.N., & Tennyson, J. 2014, MNRAS, 440, 1649

Yurchenko, S.N., Al-Refaie, A.F., & Tennyson, J. 2018, A&A, 614, A131

This paper has been typeset from a TeX/LaTeX file prepared by the author.
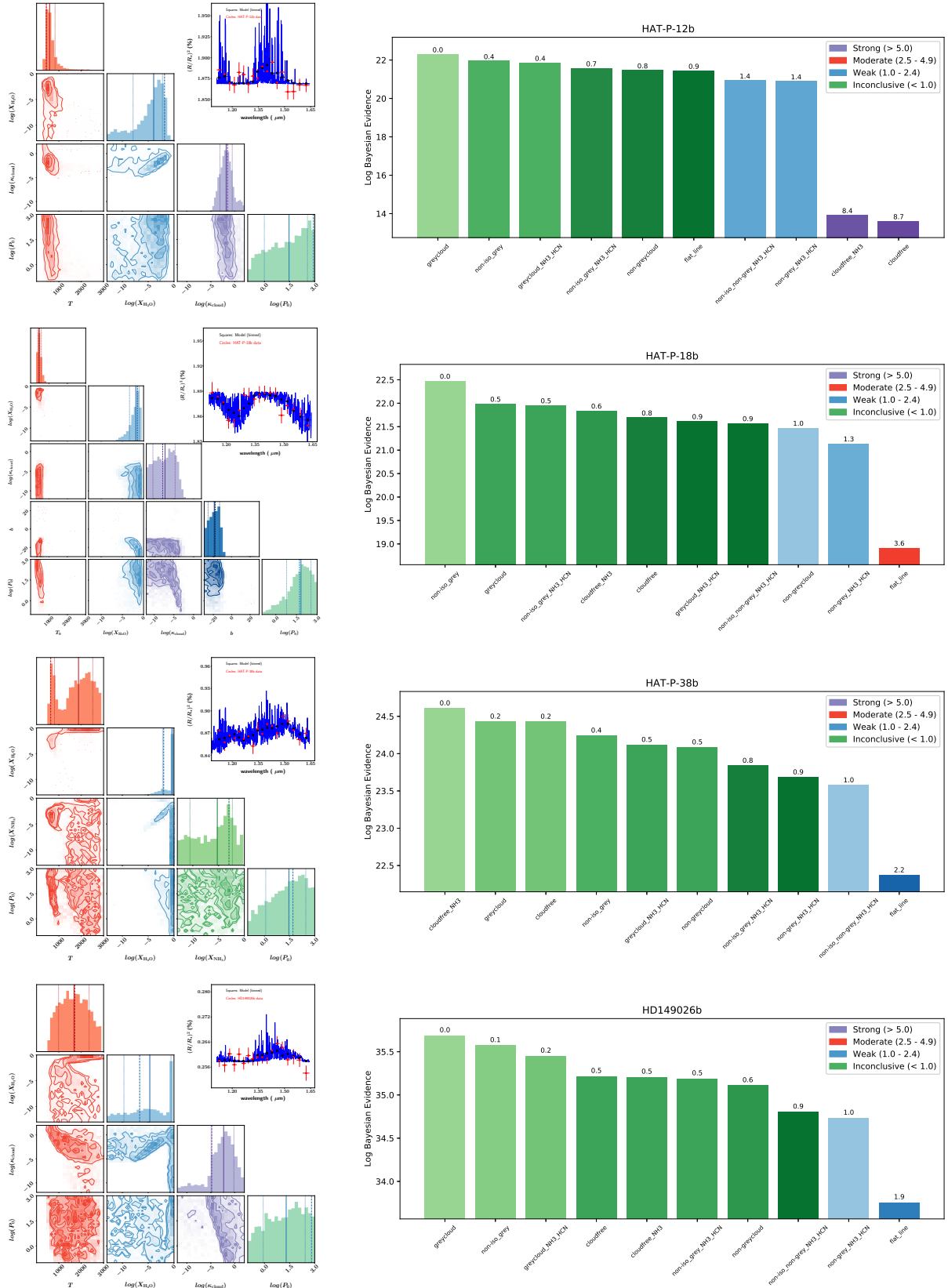
Figure 23. Same as Figure 17, but for exo-Saturns (0.2–$0.4 M_J$): HAT-P-12b, HAT-P-18b, HAT-P-38b and HD 149026b.
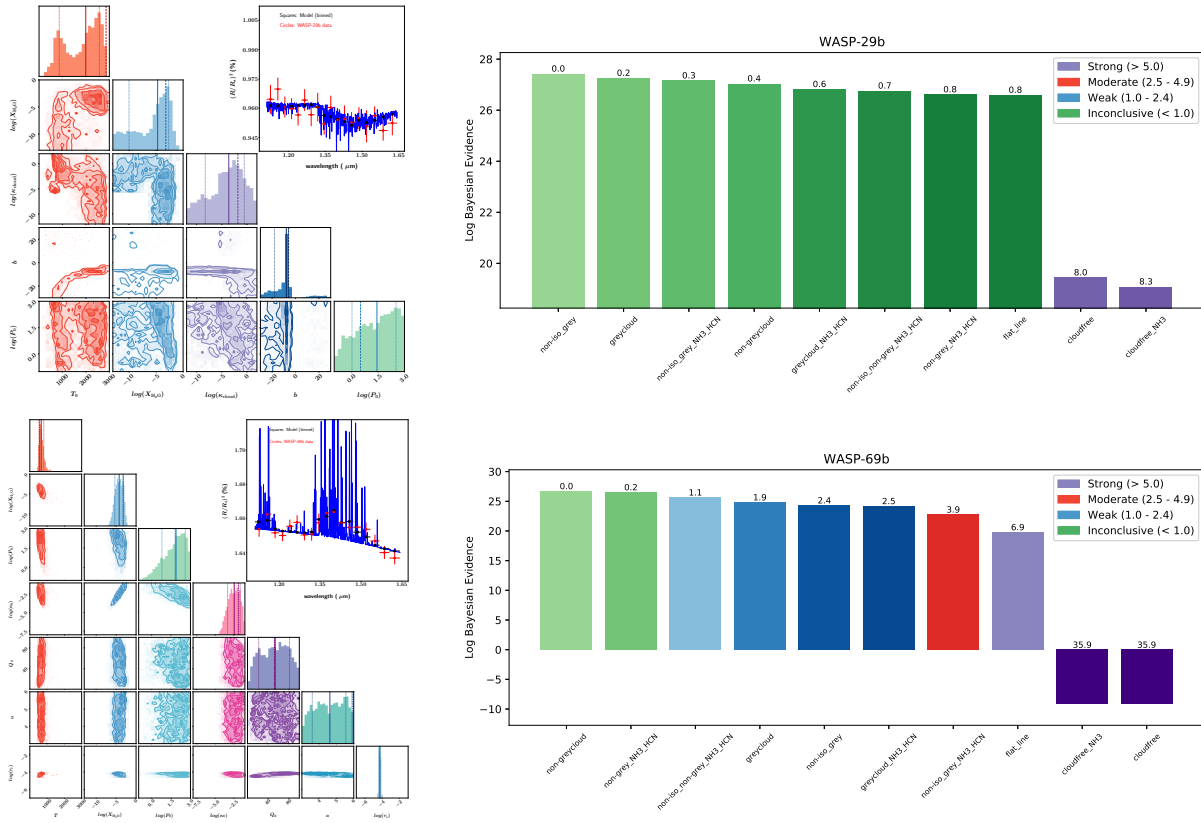
Figure 24. Continuation of Figure 23 for exo-Saturns (0.2–0.4M_J): WASP-29b and WASP-69b. Additionally, WASP-69b is one of two objects where non-grey clouds are needed to fit the data.
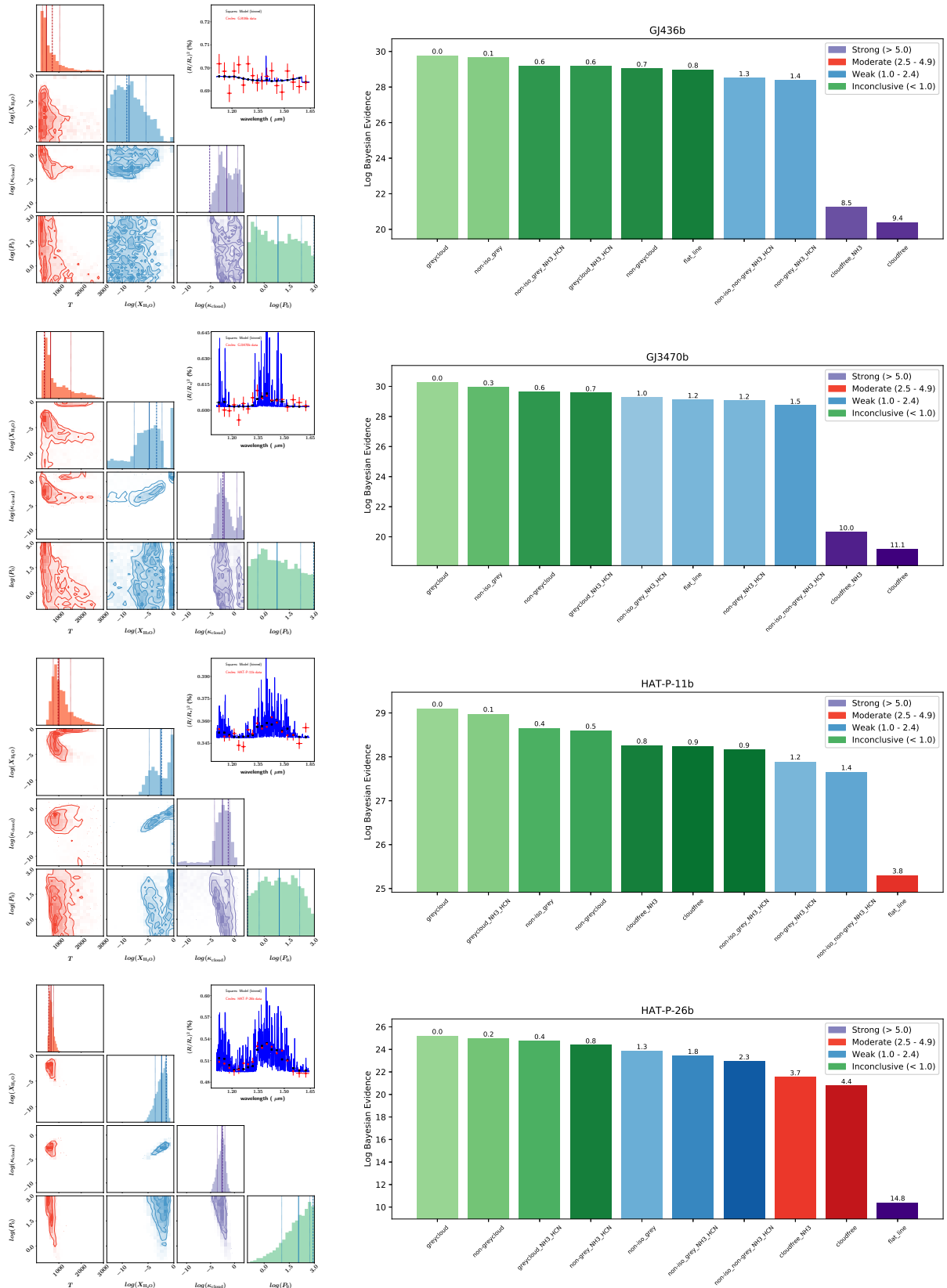
Figure 25. Same as Figure 17, but for exo-Neptunes: GJ 436b, GJ 3470b, HAT-P-11b and HAT-P-26b.
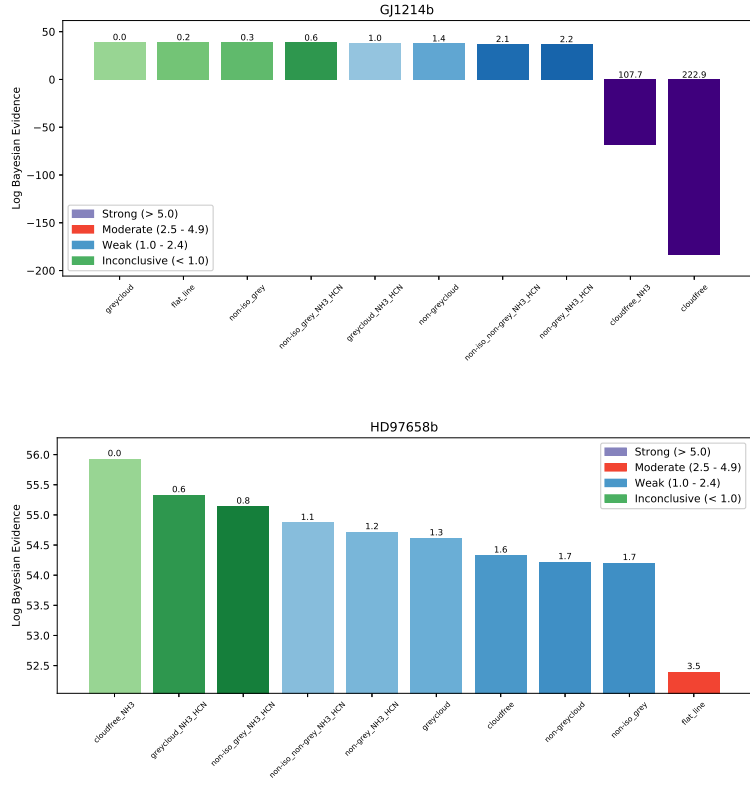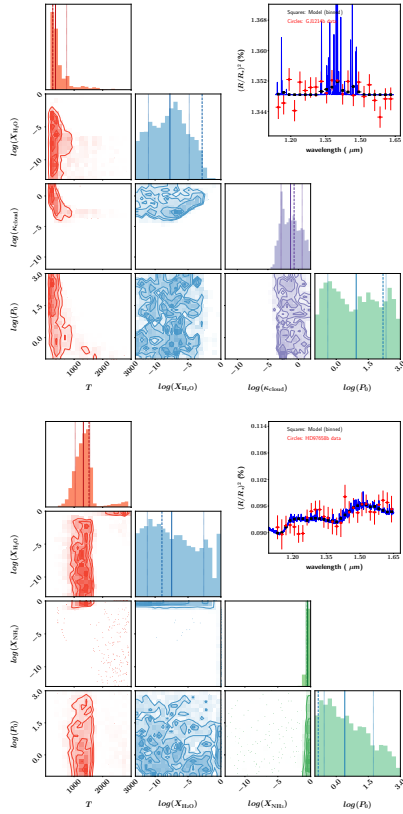
Figure 26. Same as Figure 17, but for super Earths: GJ 1214b and HD 97658b. HD 97658b stands out as an object where ammonia is significantly detected but the abundance of water is essentially unconstrained.
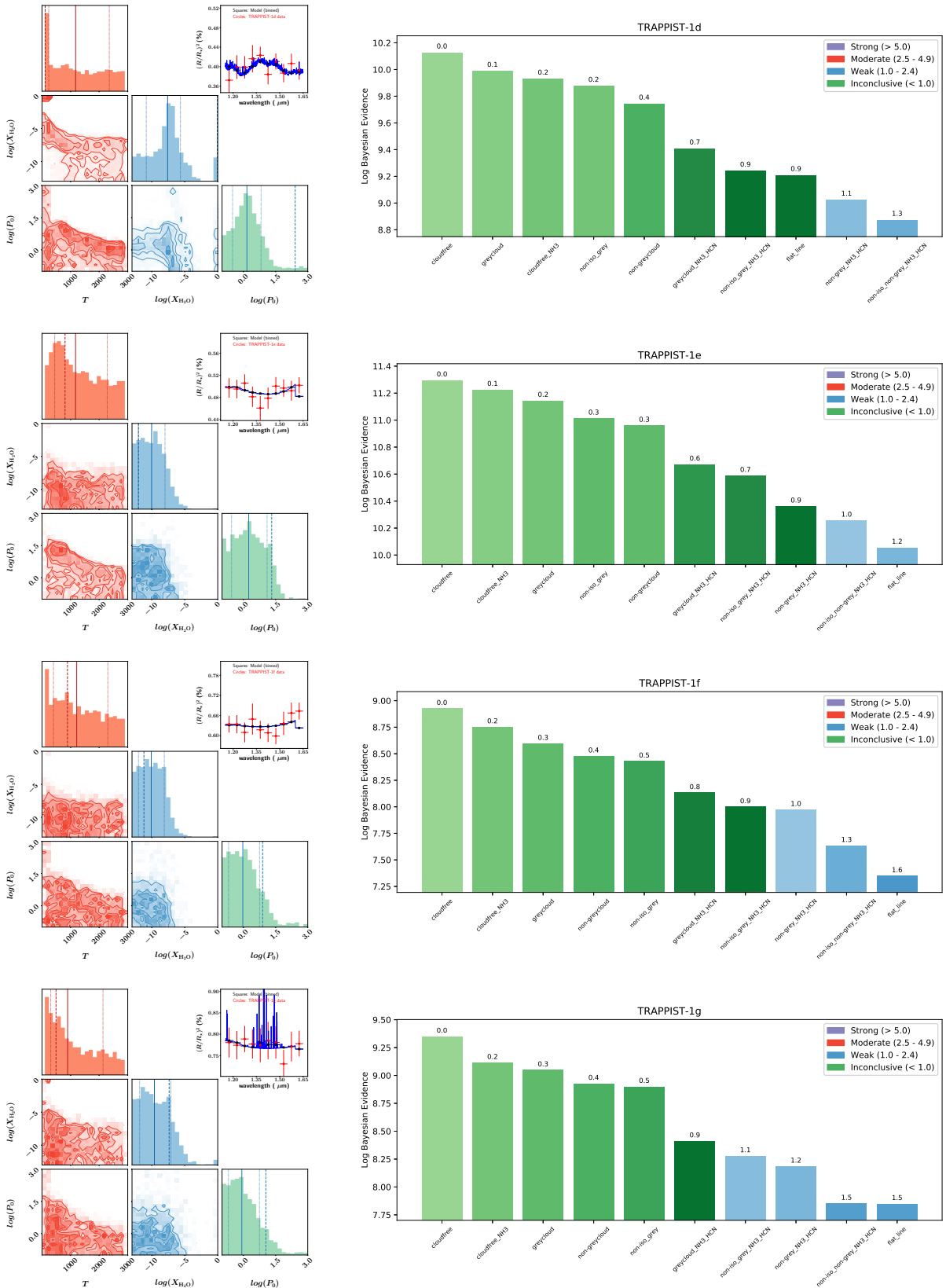
Figure 27. Same as Figure 17, but for the TRAPPIST-1 exoplanets assuming Earth-like atmospheres ($m = 29\ m_{\rm H}$).
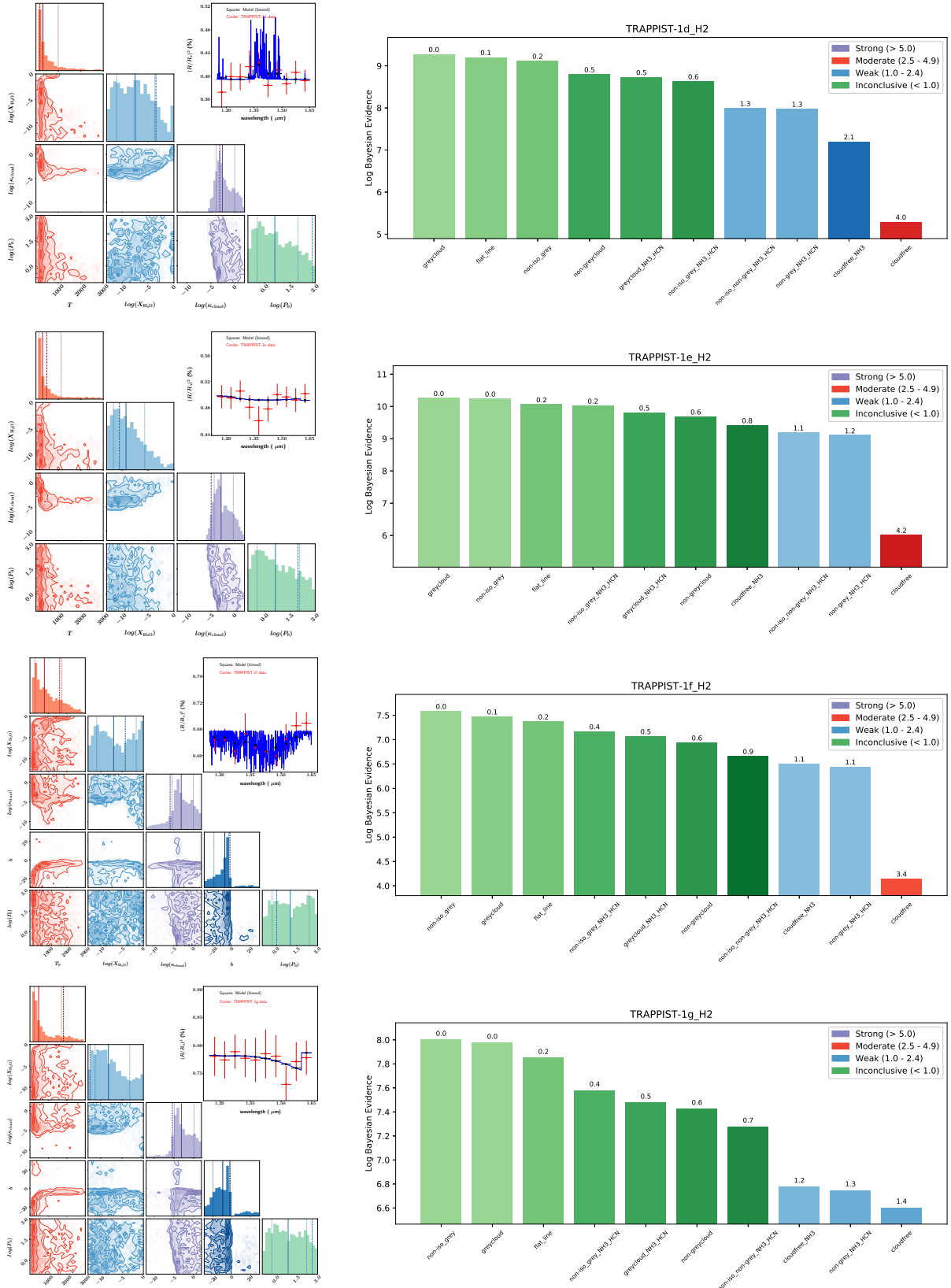
Figure 28. Same as Figure 27, but assuming atmospheres dominated by molecular hydrogen (variable m), where the pressure scale height is larger by about an order of magnitude.
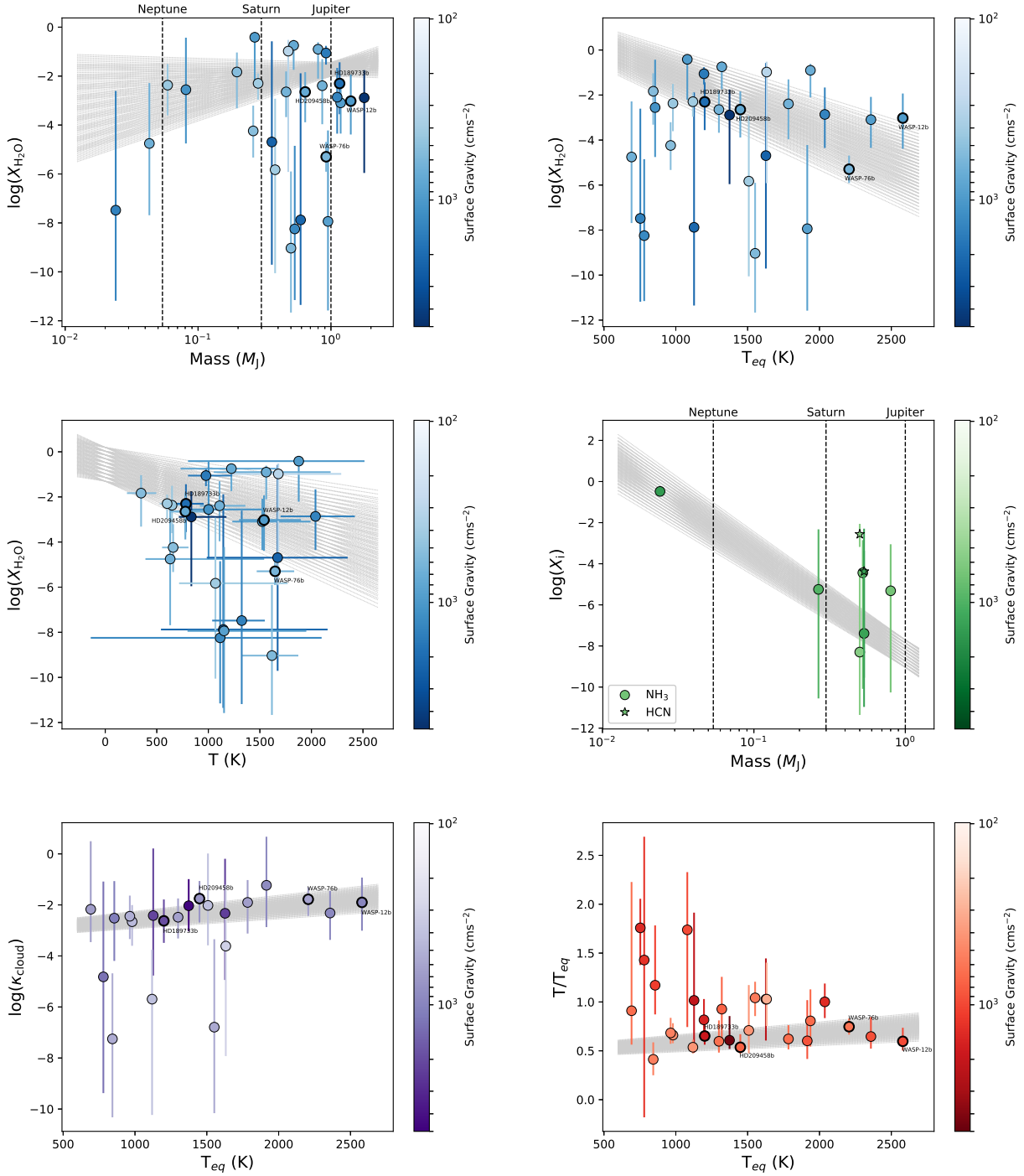
Figure 29. A search for trends between the retrieved atmospheric properties based on the best model (highest Bayesian evidence) of each object. We have excluded 7 objects that can be adequately fitted by a one-parameter flat line. We have also excluded the 4 TRAPPIST-1 exoplanets. The family of lines in each panel shows the Monte Carlo fits of a two-parameter straight line (slope and constant offset). Top row: Water volume mixing ratio versus exoplanet mass (top left panel; slope of 0.94 ± 1.11) and equilibrium temperature (top right panel; slope of -0.00245 ± 0.00054 K^{-1}). Middle left panel: Water volume mixing ratio versus retrieved atmospheric temperature; slope of -0.00134 ± 0.00078 K^{-1}. Middle right panel: Ammonia and hydrogen cyanide volume mixing ratios versus exoplanet mass; slope of -4.81 ± 0.46. Bottom left panel: Grey cloud opacity versus equilibrium temperature; slope of 0.000517 ± 0.000135 K^{-1}. Bottom right panel: Ratio of retrieved to equilibrium temperatures versus equilibrium temperature; slope of 0.000103 ± 0.000038 K^{-1}.

Table 2. Summary of retrieval outcomes (38 objects, 42 sets of retrievals)

Name	$T_{\rm eq}$ (K)	T (K)	$log(X_{\rm H_2O})$	$log(X_{\rm HCN})$	$log(X_{\rm NH_3})$	Cloudy?	Non-grey clouds?	$log(\kappa_{\rm cloud})$ (cm^2 g^{-1})
GJ436b	633	†	†	†	†	†	†	†
GJ3470b	692	629^{+911}_{-239}	$-4.75^{+2.47}_{-2.93}$	$-$	$-$	Yes	No	$-2.17^{+2.67}_{-1.29}$
HAT-P-1b	1320	1223^{+435}_{-492}	$-0.75^{+0.23}_{-1.0}$	$-$	$-4.44^{+1.59}_{-5.65}$	Maybe	$-$	$-$
HAT-P-3b	1127	1145^{+1011}_{-604}	$-7.88^{+5.99}_{-3.47}$	$-$	$-$	Maybe	$-$	$-2.42^{+2.63}_{-2.35}$
HAT-P-11b	856	1002^{+524}_{-255}	$-2.56^{+2.13}_{-2.2}$	$-$	$-$	Maybe	$-$	$-2.52^{+1.45}_{-1.68}$
HAT-P-12b	958	†	†	†	†	†	†	†
HAT-P-17b	780	1114^{+984}_{-1255}	$-8.25^{+3.4}_{-2.91}$	$-4.37^{+2.08}_{-5.17}$	$-7.39^{+4.02}_{-3.58}$	Yes	No	$-4.82^{+3.74}_{-4.56}$
HAT-P-18b	843	347^{+146}_{-137}	$-1.83^{+0.8}_{-1.48}$	$-$	$-$	Maybe	$-$	$-7.25^{+2.57}_{-3.07}$
HAT-P-26b	980	647^{+118}_{-82}	$-2.37^{+0.86}_{-1.24}$	$-$	$-$	Yes	No	$-2.66^{+0.65}_{-0.94}$
HAT-P-32b	1784	1109^{+251}_{-190}	$-2.39^{+1.09}_{-1.57}$	$-$	$-$	Yes	No	$-1.9^{+0.88}_{-1.22}$
HAT-P-38b	1080	1876^{+637}_{-1074}	$-0.41^{+0.22}_{-1.8}$	$-$	$-5.25^{+2.92}_{-5.31}$	Maybe	$-$	$-$
HAT-P-41b	1937	1561^{+624}_{-507}	$-0.9^{+0.28}_{-1.2}$	$-$	$-5.32^{+2.27}_{-4.94}$	Maybe	$-$	$-$
HD149026b	1627	1672^{+679}_{-687}	$-4.69^{+4.11}_{-5.02}$	$-$	$-$	Maybe	$-$	$-2.33^{+2.14}_{-2.61}$
HD189733b	1201	782^{+172}_{-107}	$-2.3^{+0.87}_{-1.26}$	$-$	$-$	Yes	No	$-2.62^{+0.84}_{-0.86}$
HD209458b	1449	777^{+193}_{-95}	$-2.65^{+0.81}_{-1.24}$	$-$	$-$	Yes	No	$-1.75^{+0.69}_{-0.94}$
WASP-29b	963	†	†	†	†	†	†	†
WASP-31b	1576	†	†	†	†	†	†	†
WASP-39b	1119	600^{+86}_{-72}	$-2.3^{+0.4}_{-0.67}$	$-$	$-$	Yes	No	$-5.69^{+1.94}_{-4.54}$
WASP-43b	1374	835^{+340}_{-121}	$-2.89^{+1.13}_{-3.07}$	$-$	$-$	Yes	No	$-2.03^{+1.04}_{-1.01}$
WASP-52b	1300	776^{+278}_{-149}	$-2.65^{+0.84}_{-1.03}$	$-$	$-$	Yes	No	$-2.48^{+0.74}_{-0.83}$
WASP-63b	1508	1068^{+700}_{-352}	$-5.83^{+2.9}_{-4.23}$	$-$	$-$	Yes	No	$-2.02^{+2.03}_{-1.56}$
WASP-67b	1026	†	†	†	†	†	†	†
WASP-69b	964	658^{+148}_{-107}	$-4.24^{+1.03}_{-1.09}$	$-$	$-$	Yes	Yes	$-2.44^{+0.81}_{-0.89}$
WASP-74b	1915	1152^{+798}_{-354}	$-7.94^{+3.72}_{-3.64}$	$-$	$-$	Yes	No	$-1.23^{+1.9}_{-1.64}$
WASP-76b	2206	1647^{+185}_{-178}	$-5.3^{+0.61}_{-0.61}$	$-$	$-$	Yes	Yes	$-1.78^{+0.47}_{-0.65}$
WASP-80b	824	†	†	†	†	†	†	†
WASP-101b	1552	1616^{+256}_{-288}	$-9.03^{+3.13}_{-2.63}$	$-2.56^{+0.5}_{-0.62}$	$-8.3^{+3.84}_{-3.06}$	Yes	No	$-6.79^{+3.45}_{-3.37}$
WASP-121b	2358	1523^{+468}_{-290}	$-3.09^{+1.01}_{-1.26}$	$-$	$-$	Yes	No	$-2.32^{+0.86}_{-1.05}$
XO-1b	1196	977^{+254}_{-174}	$-1.06^{+0.29}_{-0.47}$	$-$	$-$	Maybe	$-$	$-$
GJ1214b	573	†	†	†	†	†	†	†
HD97658b	753	1323^{+224}_{-286}	$-7.48^{+4.88}_{-3.7}$	$-$	$-0.48^{+0.19}_{-0.23}$	Maybe	$-$	$-$
WASP-17b	1632	1678^{+610}_{-448}	$-0.98^{+0.46}_{-4.94}$	$-$	$-$	Maybe	$-$	$-3.61^{+2.4}_{-4.31}$
WASP-19b	2037	2039^{+381}_{-338}	$-2.86^{+1.2}_{-1.49}$	$-$	$-$	Maybe	$-$	$-$
WASP-12b	2580	1540^{+358}_{-242}	$-3.02^{+1.09}_{-1.36}$	$-$	$-$	Yes	No	$-1.9^{+0.97}_{-1.11}$
TRAPPIST-1d	288	†	†	†	†	†	†	†
		†♣	†♣	†♣	†♣	†♣	†♣	†♣
TRAPPIST-1e	251	†	†	†	†	†	†	†
		1173^{+1108}_{-729}♣	$-10.02^{+2.02}_{-2.02}$♣	$-$	$-$	Maybe♣	$-$	$-$
TRAPPIST-1f	219	†	†	†	†	†	†	†
		1214^{+1089}_{-815}♣	$-10.09^{+1.98}_{-1.94}$♣	$-$	$-$	Maybe♣	$-$	$-$
TRAPPIST-1g	199	†	†	†	†	†	†	†
		896^{+1238}_{-590}♣	$-9.61^{+2.5}_{-2.24}$♣	$-$	$-$	Maybe♣	$-$	$-$

For "Cloudy?": "Yes" refers to cases where all of the cloud-free models have Bayes factors of unity or more. "No" means only cloud-free models have Bayes factor of less than unity. "Maybe" means a mixture of cloud-free and cloudy models have Bayes factor of less than unity. **For "Non-grey clouds?":** "Yes" refers to cases where only non-grey-cloud models have Bayes factors of less than unity. "No" means a mixture of non-grey-cloud and grey-cloud models have Bayes factors of less than unity.
♣: For the TRAPPIST-1 exoplanets, we also examine Earth-like atmospheres ($m = 29m_{\rm H}$).
†: Flat-line fit has Bayes factor of less than unity and no atmospheric properties may be retrieved.

Table 3. Summary of input parameters (38 objects, 42 sets of retrievals)

Name	$R_\star (R_\odot)$	$M\ (M_J)$	$R_0(R_J)$	$g\ (\text{cm s}^{-2})$	References	$\bar{R}_{\text{WFC3}}(R_J)$
GJ436b	0.455	$0.078^{+0.007}_{-0.008}$	0.3532	1318	von Braun et al. (2012)	0.3693
GJ3470b	0.48	0.043 ± 0.005	0.3287	676	Biddle et al. (2014)	0.3630
HAT-P-1b	1.115	0.524 ± 0.031	1.213	750	Johnson et al. (2008); Sing et al. (2016)	1.272
HAT-P-3b	0.799	0.591 ± 0.018	0.8383	2138	Chan et al. (2011)	0.8559
HAT-P-11b	0.75	0.081 ± 0.009	0.4077	1122	Bakos et al. (2010)	0.4332
HAT-P-12b	0.701	0.211 ± 0.012	0.8770	562	Hartman et al. (2009)	0.9341
HAT-P-17b	0.838	0.534 ± 0.018	0.9677	1288	Howard et al. (2012)	0.9880
HAT-P-18b	0.717	0.196 ± 0.008	0.9349	542	Esposito et al. (2014)	0.9552
HAT-P-26b	0.788	0.059 ± 0.007	0.4741	447	Hartman et al. (2011a)	0.5475
HAT-P-32b	1.219	0.860 ± 0.164	1.714	661	Hartman et al. (2011b)	1.804
HAT-P-38b	0.923	0.267 ± 0.020	0.8010	977	Sato et al. (2012)	0.8380
HAT-P-41b	1.683	0.800 ± 0.102	1.568	692	Hartman et al. (2012)	1.662
HD149026b	1.368	$0.359^{+0.022}_{-0.021}$	0.6536	2291	Torres et al. (2008)	0.6774
HD189733b	0.805	1.162 ± 0.058	1.200	1950	Boyajian et al. (2015)	1.221
HD209458b	1.203	0.64 ± 0.09	1.350	759	Boyajian et al. (2015)	1.414
WASP-29b	0.808	0.244 ± 0.020	0.7330	891	Hellier et al. (2010)	0.7692
WASP-31b	1.252	0.478 ± 0.029	1.379	456	Anderson et al. (2011)	1.535
WASP-39b	0.918	0.283 ± 0.041	1.207	414	Maciejewski et al. (2016)	1.297
WASP-43b	0.67	1.78 ± 0.10	1.029	4699	Hellier et al. (2011)	1.039
WASP-52b	0.79	0.46 ± 0.02	1.199	646	Hébrard et al. (2013)	1.266
WASP-63b	1.88	0.38 ± 0.03	1.316	417	Hellier et al. (2012)	1.437
WASP-67b	0.87	0.42 ± 0.04	1.314	501	Hellier et al. (2012)	1.383
WASP-69b	0.813	0.260 ± 0.017	0.9563	532	Anderson et al. (2014)	1.017
WASP-74b	1.64	0.95 ± 0.06	1.456	891	Hellier et al. (2015)	1.528
WASP-76b	1.73	0.92 ± 0.03	1.635	631	West et al. (2016)	1.752
WASP-80b	0.586	$0.538^{+0.035}_{-0.036}$	0.9562	1396	Triaud et al. (2015)	0.9760
WASP-101b	1.29	0.50 ± 0.04	1.274	575	Hellier et al. (2014)	1.364
WASP-121b	1.458	$1.183^{+0.064}_{-0.062}$	1.633	940	Delrez et al. (2016)	1.717
XO-1b	0.934	$0.918^{+0.081}_{-0.078}$	1.172	1626	Torres et al. (2008)	1.197
GJ1214b	0.211	0.019 ± 0.003	0.2135	768	Anglada-Escudé et al. (2013)	0.2385
HD97658b	0.741	$0.024^{+0.003}_{-0.002}$	0.2036	1466	van Grootel et al. (2014)	0.2208
WASP-17b	1.583	0.477 ± 0.033	1.709	316	Southworth et al. (2012)	1.897
WASP-19b	1.004	1.114 ± 0.036	1.311	1419	Tregloan-Reed et al. (2013)	1.378
WASP-12b	1.57	1.41 ± 0.10	1.748	977	Hebb et al. (2009); Kreidberg et al. (2015)	1.836
TRAPPIST-1d	0.121	$9.34^{+1.10}_{-1.23} \times 10^{-4}$	0.05402 / 0.07268♣	474	Grimm et al. (2018); van Grootel et al. (2018)	0.07436
TRAPPIST-1e	0.121	$2.43^{+0.24}_{-0.25} \times 10^{-3}$	0.07329 / 0.08174♣	912	Grimm et al. (2018); van Grootel et al. (2018)	0.08250
TRAPPIST-1f	0.121	$2.94 \pm 0.25 \times 10^{-3}$	0.08490 / 0.09294♣	837	Grimm et al. (2018); van Grootel et al. (2018)	0.09366
TRAPPIST-1g	0.121	$3.61^{+0.30}_{-0.31} \times 10^{-3}$	0.09580 / 0.1030♣	854	Grimm et al. (2018); van Grootel et al. (2018)	0.1036

♣: For the TRAPPIST-1 exoplanets, we also examine Earth-like atmospheres ($m = 29 m_H$).

How Much Information Does the Sodium Doublet Encode?
Retrieval Analysis of Non-LTE Sodium Lines at Low and High
Spectral Resolutions

Pause, take a breath and say, 'Okay, is that the answer then?'

Imre Leader

3.1 Summary

After my work on retrievals, I went on to study the theory of atmospheric spectra in more detail. In my second paper, I investigated the effect of dropping the assumption of local thermodynamic equilibrium (LTE) on sodium lines. In particular, I considered non-Boltzmann distributions for the excited and ground states of the sodium atom, corresponding to the situation where collisions become inefficient higher up in the atmosphere. This is motivated by the extended depth of

the sodium lines, which span about 9 orders of magnitude in pressure between the line wing and peak.

I studied the sodium doublet at high resolution for the exoplanet WASP-49 b, using HARPS data from Wyttenbach et al. (2017). I also studied the sodium doublet at low resolution for 6 exoplanets, using HST STIS data from Sing et al. (2016). For both resolutions I ran nested-sampling retrievals assuming LTE and non-LTE, and then compared the results using the Bayesian evidence.

Using a suite of mock retrievals, I showed that even at high resolution one is unable to distinguish between LTE and non-LTE scenarios using the Bayesian evidence. I also found that the retrievals using LTE tend to obtain a lower temperature value, especially for the high resolution case. The high-resolution retrievals also demonstrated that the sodium lines do not encode enough information to infer the pressures being probed. In addition, my low resolution analysis found no trends between the retrieved parameters for these planets.

3.2 Publication

This work was published in The Astrophysical Journal in 2019.

How Much Information Does the Sodium Doublet Encode? Retrieval Analysis of Non-LTE Sodium Lines at Low and High Spectral Resolutions

Chloe Fisher[1] and Kevin Heng
University of Bern, Center for Space and Habitability, Gesellschaftsstrasse 6, CH-3012, Bern, Switzerland; chloe.fisher@csh.unibe.ch, kevin.heng@csh.unibe.ch

Abstract

Motivated by both ground- and space-based detections of the sodium doublet in the transmission spectra of exoplanetary atmospheres, we revisit the theory and interpretation of sodium lines in non-local thermodynamic equilibrium (NLTE), where collisions are not efficient enough to maintain a Boltzmann distribution for the excited and ground states of the sodium atom. We consider non-Boltzmann distributions that account for the ineffectiveness of collisions. We analyze the sodium doublet in transmission spectra measured at low (HAT-P-1b, HAT-P-12b, HD 189733b, WASP-6b, WASP-17b, and WASP-39b) and high (WASP-49b) spectral resolutions. Nested-sampling retrievals performed on low-resolution optical/visible transmission spectra are unable to break the normalization degeneracy if the spectral continuum is associated with Rayleigh scattering by small cloud particles. Using mock retrievals, we demonstrate that unnormalized ground-based, high-resolution spectra centered on the sodium doublet alone are unable to precisely inform us about the pressure levels probed by the transit chord and hence to identify the region (i.e., thermosphere, exosphere) of the atmosphere being probed. Retrievals performed on the HARPS transmission spectrum of WASP-49b support this conclusion. Generally, we are unable to distinguish between LTE versus NLTE interpretations of the sodium doublet based on the computed Bayesian evidence with the implication that LTE interpretations tend to underestimate the temperature probed by the transit chord. With the current low-resolution data, the sodium line shapes are consistent with Voigt profiles without the need for sub-Lorentzian wings. The retrieved sodium abundances are consistent with being subsolar to solar.

Key words: planets and satellites: atmospheres

1. Introduction

The use of the resonant lines of the sodium atom to probe the atmospheres of exoplanets has a rich history. The detectability of the resonant doublets of sodium and potassium in transmission spectra was first predicted by Seager & Sasselov (2000). In a contemporaneous study, Sudarsky et al. (2000) reached the same conclusion. Later, sodium was detected for the first time in an exoplanet by Charbonneau et al. (2002) for the hot Jupiter HD 209458b. Since the predictions and discovery, the detection of sodium has become routine from space at low spectral resolution (e.g., Sing et al. 2016) and from the ground both at low (e.g., Nikolov et al. 2018) and high spectral resolution (e.g., Redfield et al. 2008; Snellen et al. 2008; Jensen et al. 2012; Wyttenbach et al. 2015, 2017; Khalafinejad et al. 2017; Casasayas-Barris et al. 2018). (But see Gibson et al. 2019 for a retraction of the detection of potassium.) These detections motivate a series of theoretical and phenomenological questions concerning the sodium lines, which we address in the current study.

1.1. Are Sodium Lines in Non-local Thermodynamic Equilibrium?

Local thermodynamic equilibrium (LTE) involves a set of four assumptions: the radiation field is described by a Planck or blackbody function, the velocity field follows a Maxwellian distribution, the distribution of neutrals versus electrons and ions is a solution of the Saha equation, and the energy levels of the ground versus excited states of the atom follow a Boltzmann distribution. One of the few studies to examine

non-LTE (NLTE) effects associated with the sodium lines in exoplanetary atmospheres is Fortney et al. (2003), who considered non-Saha distributions of the electron density caused by photoionization (see also Lavvas et al. 2014).

Here, we take a different approach: we study the NLTE effect associated with non-Boltzmann distributions of the ground and excited states of the sodium atom. Physically, this departure from LTE arises from the diminished role of collisions. To maintain a Boltzmann distribution, collisions between the sodium atoms and the atoms or molecules of the bulk gas they are embedded in (e.g., atomic or molecular hydrogen) need to be efficient. There are two aspects to this NLTE effect. The first is that the sodium doublet, also known as the Fraunhofer D_1 and D_2 lines, is a pair of resonant lines, where the difference in transit radii between line center and wing spans about 20 scale heights (e.g., Heng 2016), corresponding to about 9 orders of magnitude in pressure. It implies that the importance of collisions varies between the peak and wings of each sodium line—they are more important in the line wings, which probe deeper into the atmosphere, and less influential at the line peak.

Figure 1 shows that the ratio of number densities of the first excited state of the sodium atom to its ground state may depart from the LTE value at the order-of-magnitude level, depending on the temperature and pressure. The departure from LTE is nonnegligible for pressures of 1 μbar and 1 mbar. The LTE and NLTE ratios of number densities agree only at a pressure of 1 bar, as shown in the top panel of Figure 1. Since transmission spectra are expected to probe pressures much less than 1 bar, this NLTE effect is worthy of investigation. The departure from LTE produces a correction factor, which reduces to unity in the limit of LTE, to the integrated strength of the sodium lines (as

---

[1] University of Bern International 2021 PhD Fellowship.
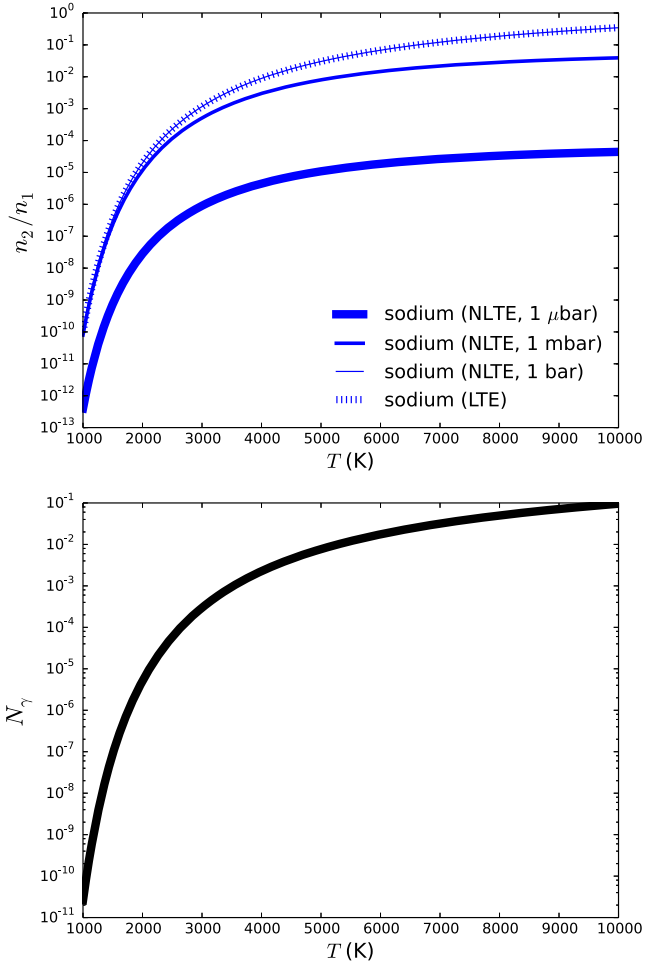
Figure 1. Ratio of number densities of first excited level to ground state (top panel) and photon occupation number (N_γ; bottom panel) associated with the sodium lines. As the Einstein A-coefficients and energy levels of the sodium doublet are very similar (see Table 1), we take their average for illustration. Also for illustration, we set $X_{Na} = 10^{-6}$ (roughly the solar value) and chose three values of the pressure: 1 μbar, 1 mbar, and 1 bar. Only for a pressure of 1 bar is the ratio of number densities described by the LTE limit. Since $N_\gamma \ll 1$, we can ignore both photoabsorption and stimulated emission (see the text).

we will elucidate in Section 2). When the sodium lines are in NLTE, the correction factor exceeds unity and produces weaker opacities. Figure 2 shows that the LTE model tends to overestimate the overall strength of the sodium doublet in the transmission spectrum, regardless of whether clouds are present.

When applied to low-resolution *Hubble Space Telescope* (*HST*) transmission spectra (Sing et al. 2016), we find this NLTE effect produces minor corrections to the retrieved atmospheric properties, except for HD 189733b. However, the interpretation of a high-resolution HARPS transmission spectrum of the hot Jupiter WASP-49b (Lendl et al. 2012, Wyttenbach et al. 2017) is biased toward lower temperatures if LTE is assumed, as we will demonstrate in Section 3.

1.2. Are sub-Lorentzian Line Wings of Sodium Needed to Interpret Spectra?

A classical approach to modeling the spectral lines of atoms or molecules is to assume that the shapes of these lines are

described by the Voigt profile, which is a mathematical convolution of the Lorentz and Doppler profiles (e.g., Chapter 6.5 of Draine 2011, Chapter 5.2 of Heng 2017). For the sodium line observed at $\sim$1000 K, the damping parameter is low enough ($\sim 10^{-3}$) that the Voigt profile consists of a Doppler core and Lorentzian wings, if pressure broadening is ignored (Heng et al. 2015). However, the studies of Burrows et al. (2000) and Allard et al. (2012) suggest that a more realistic description of the sodium line shape consists of sub-Lorentzian wings, even though these studies do not agree on the quantitative details of how to compute these line wings.

From a phenomenological viewpoint, we wish to ask if sub-Lorentzian wings are a necessary ingredient for fitting transmission spectra of sodium lines, both at low and high resolutions. In the current study, we adopt an agnostic and data-driven approach, which is to describe the line wings of sodium by a dimensionless broadening parameter (f_{broad}). A classical Voigt profile with no pressure broadening present has $f_{broad} = 1$. If pressure broadening is present, we have $f_{broad} > 1$. In high-resolution transmission spectra, a resolved sodium line with $f_{broad} > 1$ may also encode information associated with the Rossiter–McLaughlin effect and large-scale atmospheric winds (e.g., Louden & Wheatley 2015). Any value of $f_{broad} < 1$ may be interpreted as the presence of sub-Lorentzian wings. Figure 2 shows that as the broadening parameter increases, the sodium lines become stronger in the wings (while the line-center strengths remain invariant). One expects a degeneracy between the broadening parameter and the mixing ratio of sodium, because the strength of the line wings may be negated by assuming lower abundances of sodium (Figure 2).

From performing retrievals on low-resolution transmission spectra, we find that sub-Lorentzian wings are not necessary to explain the current state of the data (Section 3). This is a phenomenological, rather than theoretical, statement.

1.3. What Are the Limitations of Interpreting Low- versus High-resolution Spectra?

There is a debate in the exoplanet literature concerning how to leverage the value of low- versus high-resolution data off each other to maximize the outcome of retrievals. For example, Brogi et al. (2017) studied a joint emission spectrum of HD 209458b with data from the *HST* Wide Field Camera 3 (WFC3), the *Spitzer Space Telescope* Infrared Array Camera, and the VLT CRIRES spectrograph. These authors demonstrated that the posterior distributions of molecular abundances become narrower at the order-of-magnitude level when the high-resolution data are considered in the retrieval analysis of the low-resolution spectrum. However, as the computations were expensive they were unable to explore the entire parameter space and instead used low-resolution retrievals to guide the exploration. The conclusions of Brogi et al. (2017) on whether the dayside of HD 209458b is isothermal, as well as the retrieved range of metallicity values for the atmosphere, appear to be dependent on their choice of parameterization of the temperature–pressure profile.

As another example, Pino et al. (2018) combined low-resolution data from several *HST* instruments with high-resolution transmission spectra from HARPS to study the hot Jupiter HD 189733b, which includes the detection of the sodium doublet by Wyttenbach et al. (2015). Instead of performing a retrieval, they fixed the temperature–pressure
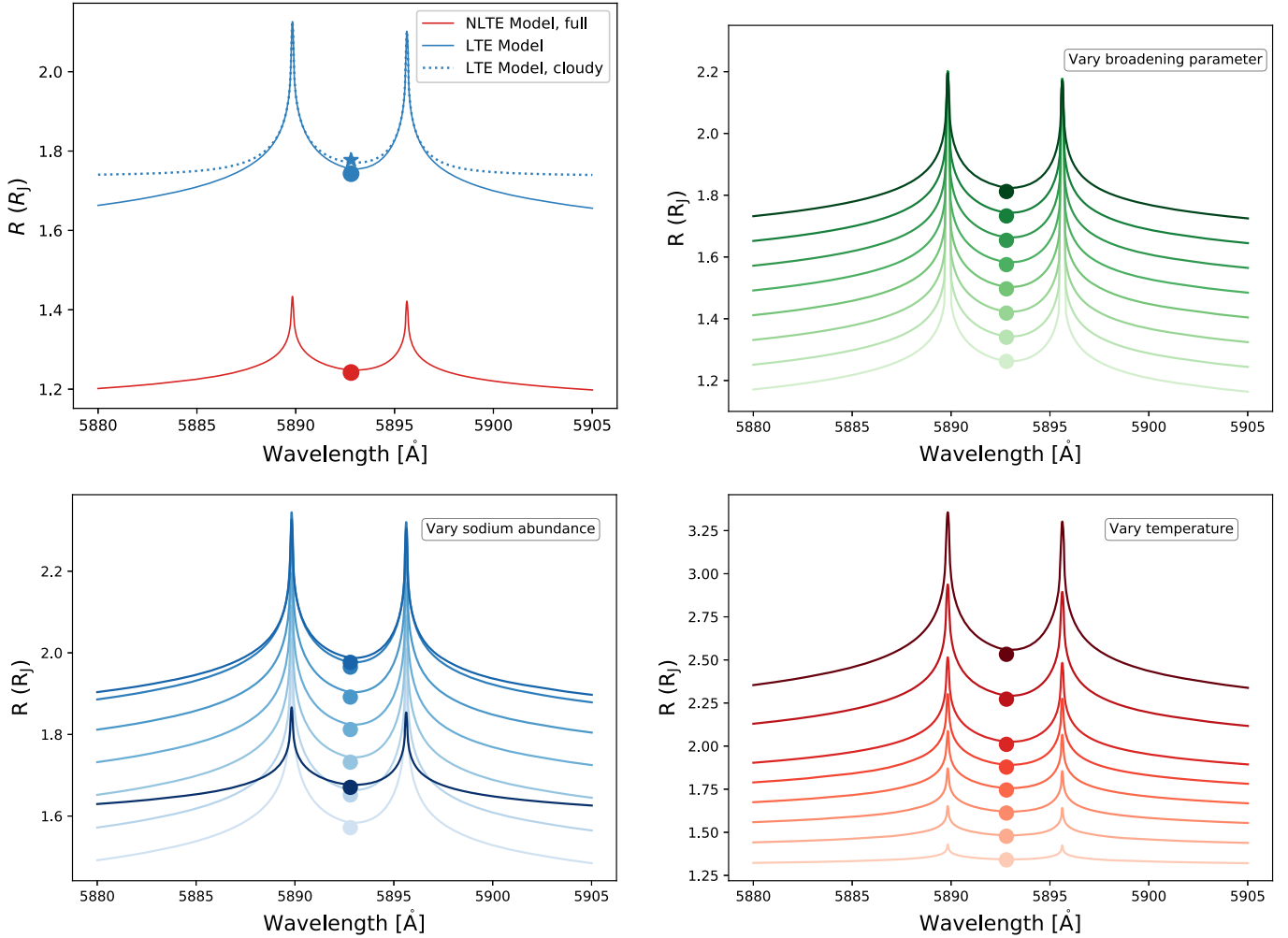
Figure 2. Models of the transmission spectrum of the sodium doublet, assuming WASP-49b-like parameter values ($g = 689$ cm^2 s^{-1}, $T = 4500$ K). For illustration, we assume $R_0 = 1.198R_J$ and $P_0 = 10$ bar for the theoretical normalization. The fiducial model assumes $X_{Na} = 10^{-5}$ and $f_{broad} = 100$. Top left panel: cloudfree NLTE vs. cloudfree and cloudy ($\kappa_{cloud} = 100$ cm^2 g^{-1}) LTE models. Top right panel: varying f_{broad} from 10^{-5} to 10^2 with lighter colors corresponding to lower values of the broadening parameter. Lower left panel: varying X_{Na} from 10^{-8} to 10^{-1} with lighter colors corresponding to lower sodium mixing ratios. When $X_{Na} = 10^{-1}$ (darkest color), the mean molecular mass increases beyond its hydrogen-dominated value ($\approx 2.4 m_{amu}$) and the pressure scale height becomes small, leading to smaller spectral features. Lower right panel: varying temperature from 1000 to 10,000 K, with lighter colors corresponding to lower temperatures. In all of the panels, the filled circle corresponds to the bandpass-averaged transit radius, except for the filled star, which corresponds to that for the cloudy LTE model.

profile to that reported by Wyttenbach et al. (2015), which was itself derived from an isothermal treatment. By paying attention to the line spread function of the different instruments, Pino et al. (2018) were able to fit the combined low- and high-resolution spectrum with a solar-composition model, i.e., the water, sodium, and potassium abundances were held fixed. Pino et al. (2018) explicitly acknowledged that they did not explore the normalization degeneracy (Benneke & Seager 2012; Griffith 2014; Heng & Kitzmann 2017).

In the current study, we ask a different question: what are the limitations encountered when *separately* interpreting low- versus high-resolution transmission spectra of the sodium doublet? This question is nontrivial, because high-resolution, ground-based spectroscopy is a "double differential" technique, where the absolute *empirical* normalization of the transmission spectrum is lost during data calibration and reduction (Wyttenbach et al. 2015, 2017; Pino et al. 2018). In other words, only the relative, and not the absolute, transit depths are measured. Quantitatively, it is unclear how the enhanced spectral resolution (relative to, e.g., *HST* transmission spectra),

diminished wavelength range and lack of an empirical normalization play off or negate one another. Space-based *HST* transmission spectra cover a wider wavelength range and do possess an absolute empirical normalization, but may not encode enough information (due to the low spectral resolution) to completely break the normalization degeneracy (Fisher & Heng 2018), which states that the theoretical normalization of the transmission spectrum is uncertain and leads to order-of-magnitude uncertainties in the chemical abundances (Benneke & Seager 2012; Griffith 2014; Heng & Kitzmann 2017).

To this end, we apply our retrieval technique to the *HST* Space Telescope Imaging Spectrograph (STIS) transmission spectra curated by Sing et al. (2016), which includes HAT-P-1b, HAT-P-12b, HD 189733b, WASP-6b, WASP-17b, and WASP-39b. The sodium doublet is not resolved in these STIS spectra, but is rather detected as a single, unresolved spike of two blended lines in absorption. As a single case study of high-resolution, ground-based spectra, we analyze the HARPS transmission spectrum of the hot Jupiter WASP-49b measured by Wyttenbach et al. (2017). The sodium doublet is resolved in

this HARPS spectrum. Besides the deep transits associated with the sodium doublet (about 2% relative to the continuum), the spectral signature of the Rossiter–McLaughlin effect was not detected (Wyttenbach et al. 2017), which makes WASP-49b an ideal high-resolution case study.

Previously, Fisher & Heng (2018) demonstrated that *HST* WFC3 near-infrared transmission spectra encode enough information to partially break the normalization degeneracy. Specifically, the caveat is that the lower limits to the posterior distributions correspond to the maximum value assumed for the prior distribution of the reference pressure. The *HST* STIS transmission spectra of Sing et al. (2016) cover the optical/visible range of wavelengths, where it has previously been established (e.g., Lecavelier des Etangs et al. 2008; Fisher & Heng 2018) that if the optical/visible spectral slope may be uniquely attributed to molecular hydrogen alone then the normalization degeneracy is completely broken. However, the expectation is that if the optical/visible spectral slope is due to clouds or hazes, then the normalizaton degeneracy cannot be broken (Heng 2016). In the current study, we interpret the STIS transmission spectra by assuming that absorption by sodium atoms, Rayleigh scattering associated with small cloud particles and gray absorption by large cloud particles are present. We find that the normalization degeneracy cannot be broken in the context of such a model interpretation using optical/visible data alone.

For the high-resolution spectra, we find that the theoretical normalization cannot be uniquely determined, but given the lack of an empirical normalization this has little to no effect on the posterior distributions of the retrieved atmospheric properties.

1.4. Is the Degree of Cloudiness of an Exoplanetary Atmosphere Correlated with Any Exoplanet Property?

Previously, Stevenson (2016) and Heng (2016) used WFC3 transmission spectra dominated by water and STIS observations of the sodium doublet, respectively, to study the degree of cloudiness of exoplanetary atmospheres and if it is correlated with any exoplanet property. Collectively, these studies find that hotter objects (Heng 2016; Stevenson 2016) with higher gravity (Stevenson 2016) are more likely to have cloudfree atmospheres.

One of the objectives of the present study is to improve upon and revisit the study of Heng (2016), who designed a diagnostic to interpret the sodium and potassium lines observed in transmission spectra at low resolution, building on the theory developed in Heng et al. (2015). It is based on the reasoning that, since these are strong resonant lines, the line peaks are unaffected by the opacities of clouds or hazes, while the line wings are significantly affected, thus allowing for the peak-to-wing distance to be used as a diagnostic for the degree of cloudiness. The diagnostic was constructed to apply at two points: one at the line peak and one in the line wing. This two-point approach was previously applied to the high-resolution HARPS spectrum of WASP-49b (Wyttenbach et al. 2017).

A more accurate approach is to perform a retrieval that fits for the partially resolved shape of the sodium lines. From the fit, one can compute the distance between the line peak and wing, $\Delta R_{\rm fit}$. One can then remove cloud opacity from the model and compute the corresponding distance in a cloudfree atmosphere, $\Delta R_{\rm fit,CF}$. By taking the ratio of these quantities, we

Table 1
Relevant Physical Properties of Na D Lines

Quantity	Value
$\lambda_{\rm D_1}$ (Å)	5895.92424
$\nu_{\rm D_1}$ (Hz)	5.085×10^{14}
$\lambda_{\rm D_2}$ (Å)	5889.95095
$\nu_{\rm D_2}$ (Hz)	5.090×10^{14}
$A_{\rm D_1}$ (s^{-1})	6.14×10^7
$A_{\rm D_2}$ (s^{-1})	6.16×10^7
$E_{\rm D_1}$ (erg)	3.369×10^{-12}
$E_{\rm D_1}$ (eV)	2.1
$E_{\rm D_2}$ (erg)	3.373×10^{-12}
$E_{\rm D_2}$ (eV)	2.1
$f_{\rm D_1}$	0.320
$f_{\rm D_2}$	0.641

Note. All data extracted from the NIST database except for oscillator strengths ($f_{\rm D_i}$), which are taken from Table 9.3 of Draine (2011). $A_{\rm D_1}$ and $A_{\rm D_2}$ are the A_{21} values for the D_1 and D_2 lines, respectively.

obtain the cloudiness index (Heng 2016),

$$C = \frac{\Delta R_{\rm fit,CF}}{\Delta R_{\rm fit}}, \qquad (1)$$

such that cloudfree and cloudy atmospheres have $C = 1$ and $C > 1$, respectively. The reason to use the best-fit model, rather than the actual data points, in the denominator of C is to avoid situations where an imperfect fit to the measured transmission spectrum produces a best-fit model line peak that sits below the data point, which produces spurious values of $C < 1$. We further investigate if estimates of C are affected by the assumption of LTE.

1.5. Layout of the Paper

In Section 2, we lay out our methodology, including the theory behind the two-level atom treatment, the cross section of the sodium lines, models of the transmission spectrum and treatments of the high-resolution data format. In Section 3, we study a suite of mock retrievals and perform retrieval analyses on the *HST* STIS and HARPS transmission spectra. In Section 4, we discuss the implications of our results, including a scrutiny of the information content of high-resolution spectra with respect to detecting NLTE effects. Table 1 contains the physical constants associated with the sodium doublet. Table 2 states the physical units and prior ranges of the retrieved parameters. Table 3 summarizes the values of the retrieved atmospheric properties for the sample of seven exoplanets considered in the current study.

2. Theory and Methodology

2.1. Two-level Sodium Atom

The two-level atom is not a novel concept and has previously been described in monographs (e.g., Chapter 17.1 of Draine 2011). Here, we recast it in a form that is suitable for computation and application to sodium atoms.

The two-level atom has a ground state and an excited state. The number density of atoms in the excited state (n_2) is described by the following evolution equation (e.g., Equation

Table 2
Retrieved Parameters and Their Prior Ranges

Quantity	Symbol	Units	High-res Range	Low-res Range	Prior Type
Temperature	T	K	0–10,000	0–6000	uniform
Sodium volume mixing ratio	X_{Na}	⋯	10^{-13}–1	10^{-13}–1	log-uniform
Broadening parameter	f_{broad}	⋯	10^{-1}–10^4	10^{-5}–10^{10}	log-uniform
Gray cloud opacity	κ_{cloud}	cm^2 g^{-1}	10^{-4}–10^5	10^{-10}–10^5	log-uniform
Rayleigh scattering scaling parameter (clouds)	f_{cloud}	⋯	⋯	1–10^7	log-uniform
Reference pressure	P_0	bar	1–100	1–100	log-uniform

Note. 1 bar = 10^6 dyn cm^{-2} (cgs units) = 10^5 N m^{-2} (mks units).

Table 3
Summary of Retrieval Outcomes (Both LTE and NLTE Models)

Object	T_{eq} (K)	T (K)	$\log X_{Na}$	$\log f_{broad}$	$\log \kappa_{cloud}$ (cm^2 g^{-1})	$\log f_{cloud}$	C	$\log P_0$ (bar)	R_0 (R_J)
HAT-P-1b[a]	1320	351^{+200}_{-172}	$-8.11^{+4.2}_{-3.15}$	$0.26^{+4.67}_{-3.31}$	$-6.32^{+2.75}_{-2.34}$	$0.95^{+2.67}_{-0.73}$	$1.18^{+1.67}_{-0.16}$	$0.5^{+0.75}_{-0.35}$	1.26
NLTE		339^{+185}_{-152}	$-8.54^{+4.03}_{-2.85}$	$0.2^{+4.89}_{-3.35}$	$-6.4^{+2.96}_{-2.35}$	$1.0^{+2.35}_{-0.76}$	$1.22^{+1.81}_{-0.2}$	$0.5^{+0.75}_{-0.35}$	1.26
HAT-P-12b[a]	960	984^{+308}_{-302}	$-9.61^{+3.59}_{-2.24}$	$-0.56^{+4.63}_{-2.9}$	$-6.57^{+2.1}_{-2.15}$	$0.85^{+1.13}_{-0.55}$	$1.18^{+1.01}_{-0.17}$	$0.57^{+0.64}_{-0.38}$	0.9
NLTE		1018^{+278}_{-283}	$-9.67^{+3.57}_{-2.24}$	$-0.49^{+4.43}_{-2.84}$	$-6.74^{+2.13}_{-2.09}$	$0.72^{+0.86}_{-0.44}$	$1.24^{+0.86}_{-0.21}$	$0.57^{+0.64}_{-0.38}$	0.9
HD 189733b[a]	1200	2797^{+208}_{-211}	$-5.44^{+1.77}_{-1.79}$	$2.24^{+1.94}_{-1.86}$	$-5.37^{+3.65}_{-3.11}$	$4.78^{+0.84}_{-0.75}$	$4.78^{+0.99}_{-0.73}$	$1.04^{+0.65}_{-0.67}$	1.12
NLTE		4782^{+467}_{-472}	$-6.29^{+2.36}_{-2.35}$	$1.2^{+2.51}_{-2.35}$	$-5.97^{+2.9}_{-2.67}$	$2.92^{+0.63}_{-0.58}$	$5.92^{+1.39}_{-1.03}$	$1.04^{+0.65}_{-0.67}$	1.12
WASP-6b[a]	1150	880^{+362}_{-312}	$-9.01^{+3.84}_{-2.53}$	$0.07^{+5.21}_{-3.26}$	$-5.7^{+2.97}_{-2.72}$	$2.55^{+2.22}_{-1.48}$	$1.63^{+3.64}_{-0.6}$	$0.89^{+0.68}_{-0.58}$	1.18
NLTE		1004^{+272}_{-287}	$-8.94^{+3.77}_{-2.48}$	$0.04^{+5.32}_{-3.21}$	$-5.89^{+2.8}_{-2.58}$	$1.95^{+2.31}_{-1.11}$	$1.91^{+4.6}_{-0.82}$	$0.89^{+0.68}_{-0.58}$	1.18
WASP-17b[a]	1740	1489^{+336}_{-399}	$-9.32^{+3.74}_{-3.51}$	$0.38^{+4.45}_{-3.51}$	$-6.91^{+2.39}_{-2.01}$	$0.75^{+0.91}_{-0.47}$	$1.2^{+0.71}_{-0.18}$	$0.58^{+0.54}_{-0.36}$	1.73
NLTE		1568^{+307}_{-329}	$-9.25^{+3.77}_{-2.45}$	$-0.36^{+4.67}_{-3.04}$	$-6.78^{+2.24}_{-2.1}$	$0.58^{+0.57}_{-0.38}$	$1.2^{+0.56}_{-0.18}$	$0.58^{+0.54}_{-0.36}$	1.73
WASP-39b[a]	1120	835^{+131}_{-132}	$-5.56^{+2.3}_{-2.29}$	$2.48^{+2.14}_{-2.4}$	$-6.04^{+2.84}_{-2.71}$	$2.22^{+1.05}_{-0.89}$	$1.39^{+0.1}_{-0.08}$	$1.05^{+0.65}_{-0.68}$	1.18
NLTE		1057^{+205}_{-125}	$-5.52^{+2.96}_{-2.98}$	$1.6^{+2.68}_{-3.06}$	$-5.94^{+2.31}_{-2.7}$	$1.48^{+0.7}_{-0.64}$	$1.42^{+0.12}_{-0.1}$	$1.05^{+0.65}_{-0.68}$	1.18
WASP-49b[b]	1400	7209^{+1763}_{-1892}	$-5.64^{+2.55}_{-2.90}$	$-5.70^{+3.61}_{-2.62}$	$-0.18^{+3.14}_{-2.77}$	⋯	$1.52^{+0.84}_{-0.37}$	$1.02^{+0.63}_{-0.69}$	1.198
NLTE		8415^{+1020}_{-1526}	$-4.44^{+2.02}_{-2.51}$	$-6.56^{+2.91}_{-2.30}$	$-0.90^{+2.82}_{-2.06}$	⋯	$1.36^{+0.42}_{-0.24}$	$1.04^{+0.64}_{-0.68}$	1.198

Notes.
[a] *HST* STIS data from Sing et al. (2016).
[b] HARPS data from Wyttenbach et al. (2017).

(17.5) of Draine 2011),

$$\frac{\partial n_2}{\partial t} = (n_e + n_{total})n_1 C_{12} + \frac{n_1 N_\gamma g_2 A_{21}}{g_1}$$
$$- (n_e + n_{total})n_2 C_{21} - n_2 A_{21}(1 + N_\gamma), \quad (2)$$

where t denotes the time, n_1 is the number density of atoms in the ground state, n_e is the number density of electrons, n_{total} is the total number density (which is the number density of hydrogen molecules in a hydrogen-dominated atmosphere), A_{21} is the Einstein A-coefficient (with units of s^{-1}) and N_γ is the photon occupation number. The statistical weights are $g_j = 2j^2$, where $j = 1, 2$. The rate coefficient for collisional deexcitation is given by Equation (2.27) of Draine (2011),

$$C_{21} = 8.629 \times 10^{-8} \text{ cm}^3 \text{ s}^{-1} T_4^{-1/2} \frac{\Omega_{21}}{g_2}, \quad (3)$$

where $T_4 \equiv T/10^4$ K and T is the temperature. As explained by Draine (2011), the collision strength Ω_{21} is approximately independent of temperature for $T < 10^4$ K and typically has values between 1 and 10. To maximize the NLTE effect, we approximate $\Omega_{21} \approx 1$, which produces the highest possible value of the critical density.

Physically, besides the collisional excitation ($n_{total} n_1 C_{12}$) and deexcitation ($n_{total} n_2 C_{21}$) of the excited level as well as

deexcitation by spontaneous emission ($n_2 A_{21}$), there is also photoabsorption ($n_1 N_\gamma g_2 A_{21}/g_1$) and stimulated emission ($n_2 N_\gamma A_{21}$), as described in Chapter 17.1 of Draine (2011). However, we show in the bottom panel of Figure 1 that the photon occupation number,

$$N_\gamma = (e^{h\nu/k_B T} - 1)^{-1}, \quad (4)$$

with h being Planck's constant, k_B being Boltzmann's constant, and ν being the frequency, is much less than unity, implying that the terms associated with N_γ in Equation (2) may be dropped. Note that the preceding expression for N_γ assumes a Planckian radiation field.

Demanding a steady state ($\partial n_2/\partial t = 0$) yields

$$\frac{n_2}{n_1} = \frac{C_{12}}{C_{21}}(1 + f_e)\left(1 + f_e + \frac{n_{crit}}{n_{total}}\right)^{-1}, \quad (5)$$

where E_{12} is the energy difference between the ground and excited levels and $f_e \equiv n_e/n_{total}$ is the ionization fraction. For $f_e \ll 1$, we obtain the standard NLTE expression for a two-level atom,

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} e^{-E_{12}/k_B T}\left(1 + \frac{n_{crit}}{n_{total}}\right)^{-1}. \quad (6)$$

To arrive at Equation (6), we invoke the principle of detailed balance (e.g., Section 3.5 of Draine 2011),

$$\frac{C_{12}}{C_{21}} = \frac{g_2}{g_1} e^{-E_{12}/k_B T}. \tag{7}$$

The critical density is defined as

$$n_{\text{crit}} \equiv \frac{A_{21}}{C_{21}}. \tag{8}$$

When $n_{\text{total}} \gg n_{\text{crit}}$ (i.e., collisions are important), we recover the Boltzmann distribution and the states obey LTE. At a pressure of 1 bar, n_2/n_1 is well described by its LTE limit for temperatures between 10^3 and 10^4 K (Figure 1). However, for lower pressures, the departures from LTE may become significant at higher temperatures.

For completeness, the Appendix describes an analytical expression for the electron density if all of the electrons are being sourced by collisional ionization of the sodium atom.

2.2. Cross Section of Sodium Lines

The absorption cross section of the sodium atom is the product of the integrated line strength (S) and the line shape (Φ),

$$\sigma_{\text{Na}} = S\Phi. \tag{9}$$

There are three different ways to write S, depending on if the line shape is written in per wavenumber, per frequency, or per wavelength units. We use per frequency units, such that S has units of cm^2 s^{-1} and Φ has units of s.

2.2.1. Integrated Line Strength

The integrated line strength is (Penner 1952)

$$S = \frac{n_2 A_{21} h\nu}{\epsilon\, c n_{\text{total}}}. \tag{10}$$

Following Appendix 2 of Goody & Yung (1989), a factor of $1/n_{\text{total}}$ has been inserted to give S the correct physical units, such that σ_{Na} has units of cm^2. The energy density per unit frequency is given by

$$\epsilon = \frac{4\pi B_\nu}{c}, \tag{11}$$

where c is the speed of light and

$$B_\nu = \frac{2h\nu^3}{c^2}(e^{h\nu/k_B T} - 1)^{-1} \tag{12}$$

is the Planck function in per frequency units. The Einstein A-coefficient is related to the oscillator strength, f_{12}, by (e.g., Equation (6.20) of Draine 2011)

$$A_{21} = \frac{8\pi^2 e^2 \nu^2}{m_e c^3} \frac{g_1}{g_2} f_{12}, \tag{13}$$

where e is the elementary charge and m_e is the mass of the electron.

It follows that the integrated line strength becomes

$$S = \frac{\pi e^2 f_{12}}{m_e c} \frac{n_1}{n_{\text{total}}} (1 - e^{-h\nu/k_B T}) \left(1 + \frac{n_{\text{crit}}}{n_{\text{total}}}\right)^{-1}. \tag{14}$$

There are several noteworthy aspects of the preceding expression. First, since $N_\gamma \ll 1$ and $n_1 \approx n_{\text{total}}$, we obtain

$$S = \frac{\pi e^2 f_{12}}{m_e c} \left(1 + \frac{n_{\text{crit}}}{n_{\text{total}}}\right)^{-1}, \tag{15}$$

which is our final NLTE expression for the two-level atom. Second, in the LTE limit we obtain $S \approx \pi e^2 f_{12}/m_e c$, which is the same as Equation (6.25) of Draine (2011). Third, since Equation (10) is the same starting point for deriving the standard expression for the line strength in HITRAN/HITEMP, it implies that Equation (15) is equivalent to Equation (A5) of Rothman et al. (1998).

2.2.2. Line Shape

The Voigt profile is standard knowledge in monographs (e.g., Chapter 6.5 of Draine 2011, Chapter 5.2 of Heng 2017). However, there is more than one way to define the Lorentz and Doppler widths. Here, we concisely restate the formalism to be explicit about the conventions adopted in the current study.

The Lorentz profile is

$$\Phi_L = \frac{\Gamma_L/\pi}{(\nu - \nu_0)^2 + \Gamma_L^2}, \tag{16}$$

where ν_0 is the frequency at line center and $\Gamma_L = f_{\text{broad}} A_{21}/4\pi$ is the half width at half maximum (HWHM) of the Lorentz profile. This convention differs from, for example, Draine (2011) who defines it as the full width at half maximum (FWHM). We use the dimensionless broadening parameter f_{broad}, which expresses the strength of the Lorentzian wings in terms of the natural width, as an agnostic way of including all sources of broadening. In particular, since the theory of pressure broadening is incomplete, the use of $f_{\text{broad}} > 1$ allows us to report retrieved broadening values in terms of the natural width, which will motivate future studies of the sodium line shape. It also allows us to isolate the pressure dependence of the cross section due to the NLTE term associated with the total density. The use of $f_{\text{broad}} < 1$ allows us to quantify the extent to which the line wings are sub-Lorentzian. We do not truncate the sodium line wings.

The Doppler profile is

$$\Phi_D = \Gamma_D^{-1} \sqrt{\frac{\ln 2}{\pi}} e^{-(\nu-\nu_0)^2/\sigma_{\text{th}}^2}. \tag{17}$$

The thermal speed is $v_{\text{th}} = \sqrt{2k_B T/m}$, while the Doppler shift associated with the thermal speed is $\sigma_{\text{th}} = \nu_0 v_{\text{th}}/c$. $\Gamma_D = \sigma_{\text{th}}\sqrt{\ln 2}$ is the HWHM of the Doppler profile.

By convolving the Lorentz and Doppler profiles, the Voigt profile obtains,

$$\Phi = \sqrt{\frac{\ln 2}{\pi}} \frac{H_V}{\Gamma_D}, \tag{18}$$

where the Voigt H-function is

$$H_V \equiv \frac{a_0}{\pi} \int_{-\infty}^{+\infty} \frac{e^{-y^2}}{(x-y)^2 + a_0^2}\, dy, \tag{19}$$

and $x \equiv (\nu - \nu_0)/\sigma_{\text{th}}$. The damping parameter is defined as $a_0 \equiv \Gamma_L/\sigma_{\text{th}}$.

In theory, the Voigt H-function is evaluated using (Roston & Obaid 2005; Zaghloul 2007)

$$H_V = e^{a_0^2 - x^2} \, \texttt{erfc}\,(a_0) \cos{(2a_0 x)}$$
$$+ \frac{2}{\sqrt{\pi}} \int_0^x e^{y^2 - x^2} \sin{[2a_0(x - y)]} \, dy, \qquad (20)$$

where $\texttt{erfc}(a_0)$ is the complementary error function (e.g., Chapter 10 of Arfken & Weber 1995) with a_0 as its argument.

In practice, the Voigt profile is evaluated using (Schreier 1992)

$$H_V = \Re[w(z)], \qquad (21)$$

where $w(z)$ is the Faddeeva function (Equation $(7.1.3)^2$ of Abramowitz & Stegun 1970) evaluated at (Schreier 1992)

$$z = x + i a_0. \qquad (22)$$

This representation of the Voigt function allows for faster and simpler computation of the line shape.

2.3. Transmission Spectrum

2.3.1. LTE

At a given wavelength or frequency, the expression for the transit radius corresponding to an isothermal transit chord (Lecavelier des Etangs et al. 2008; de Wit & Seager 2013; Bétrémieux & Swain 2017; Heng & Kitzmann 2017; Jordán & Espinoza 2018) is

$$R = R_0 + H\left[\gamma + \ln\left(\frac{P_0 \kappa}{g}\sqrt{\frac{2\pi R_0}{H}}\right)\right], \qquad (23)$$

where $H = k_B T / m g$ is the isothermal pressure scale height, m is the mean molecular mass, g is the gravity, $\gamma = 0.57721$ is the Euler–Mascheroni constant, and κ is the opacity (cross section per unit mass). The reference pressure, P_0, corresponds to a reference transit radius, R_0.

For our model atmospheres, the opacity is

$$\kappa = \frac{X_{Na} \pi e^2 f_{12}}{m m_e c} \Phi + \kappa_{cloud}, \qquad (24)$$

where X_{Na} is the volume mixing ratio of sodium and κ_{cloud} is a gray/constant cloud opacity. The assumption is that the wavelength range probed is narrow enough that it is insensitive to changes in cloud opacity caused by small particles, or that the particles are large (compared to the wavelength).

Denoting the atomic mass unit by m_{amu}, the mean molecular mass is given by

$$m = 2.4 X_{H_2} m_{amu} + X_{Na} m_{Na}, \qquad (25)$$

where $m_{Na} = 23 m_{amu}$ is the mass of the sodium atom. The mixing ratio of molecular hydrogen is determined by demanding that all mixing ratios sum to unity,

$$1.1 X_{H_2} + X_{Na} = 1, \qquad (26)$$

where we have assumed that the helium mixing ratio follows cosmic abundance ($X_{He} = 0.1 X_{H_2}$). When $X_{Na} \ll 1$, we have $X_{H_2} \approx 0.91$. A continuum of possibilities, from hydrogen-dominated to sodium-dominated atmospheres, is allowed.

2.3.2. NLTE

The NLTE expression describing the number densities of the excited and ground states of sodium explicitly contains the pressure P, because $n_{total} = P/k_B T$ if we assume the ideal gas law. This means that we need to explicitly account for the variation of pressure across the sodium line profile. To keep the problem tractable, we assume hydrostatic equilibrium, $P = P_0 \exp[(R_0 - R)/H]$. One of the assumptions involved in deriving Equation (23) is worth emphasizing. Early in the derivation, one assumes that, in writing the optical depth as an integral involving the cross section and number density, one may bring the cross section out of the integral, i.e., assume that it is a constant with respect to the integration variable. Since the integral is carried out over the transit chord, this is equivalent to assuming that the cross section is constant across the chord at a given wavelength. It does not preclude the fact that the pressure probed by the transit chord may be rather different at different wavelengths. In that sense, Equation (23) is only isobaric at a given wavelength and may be used to describe spectral lines where the pressure probed across the line (across different wavelengths) may vary markedly.

The opacity is given by

$$\kappa = \kappa_{LTE}\left(1 + \frac{n_{crit} k_B T}{P}\right)^{-1}, \qquad (27)$$

where we have defined

$$\kappa_{LTE} \equiv \frac{X_{Na} \pi e^2 f_{12}}{m m_e c} \Phi + \kappa_{cloud}. \qquad (28)$$

We assume that the particles associated with the gray cloud opacity track the behavior of the sodium atoms with regards to departures from LTE.

Substituting the expression for the opacity into Equation (23) yields

$$R = R_{LTE} - H \ln\left(1 + \frac{n_{crit} k_B T}{P}\right), \qquad (29)$$

where we have defined

$$R_{LTE} \equiv R_0 + H\left[\gamma + \ln\left(\frac{P_0 \kappa_{LTE}}{g}\sqrt{\frac{2\pi R_0}{H}}\right)\right]. \qquad (30)$$

This way of writing Equation (29) means that the two effects of P_0 have been separated: its regular appearance in R_{LTE} versus its appearance in the NLTE term, which is related to the varying importance of collisions across the sodium line.

By defining

$$x \equiv e^{(R - R_0)/H}, \qquad (31)$$

Equation (29) may be written as a quadratic equation,

$$A x^2 + x - x_0 = 0, \qquad (32)$$

where $A \equiv n_{crit} k_B T / P_0$ and $x_0 \equiv \exp[(R_{LTE} - R_0)/H]$. By solving this quadratic equation, we obtain a closed-form solution for the transit radius,

$$R = R_0 + H \ln\left(\frac{\sqrt{1 + 4A x_0} - 1}{2A}\right). \qquad (33)$$

---

2 The Faddeeva function appears without being specifically attributed to Faddeeva in Abramowitz & Stegun (1970).

It is apparent that the dependence of the transit radius on A, and hence the P_0 term responsible for the varying importance of collisions across the sodium line, is weak.

2.3.3. Values Used for the Reference Transit Radius and Reference Pressure

Following Fisher & Heng (2018), we estimate the reference transit radius using

$$R_0 = \bar{R} - 6.908H. \qquad (34)$$

In Fisher & Heng (2018), $\bar{R}$ is the average transit radius in the WFC3 bandpass. Here, we take $\bar{R}$ to be the *HST* white-light radius as stated in Sing et al. (2016).

Based on WASP-17b, Fisher & Heng (2018) estimated that the WFC3 and STIS bandpasses probe ~10 mbar. This implies that $P_0 \sim 10$ bar. For the low-resolution *HST* STIS retrievals, we therefore set a tight, log-uniform prior of $1 \leqslant P_0 \leqslant 100$ bar.

For the high-resolution spectrum of WASP-49b, we match the spectral continuum to the *HST* white-light radius of $1.198\,R_{\rm J}$ (Wyttenbach et al. 2017), where $R_{\rm J}$ is the radius of Jupiter. This requires setting $R_0 = 0.752R_{\rm J}$ for $P_0 = 10$ bar.

2.4. Nested-sampling Retrievals and Bayesian Evidence

Our exploration of the parameter space of models is performed using a nested-sampling framework (Feroz & Hobson 2008; Feroz et al. 2009, 2013). We use the open-source PyMultiNest package (Buchner et al. 2014), which was previously implemented in Fisher & Heng (2018). Nested-sampling allows us to compute the Bayesian evidence and Bayes factor, which allows us to identify the best model given the quality of the data (Trotta 2008). It also allows us to identify the family of models consistent with the data (e.g., Fisher & Heng 2018).

2.5. Equivalence with Heng (2016)

Let the transit radius at line center and wing be R_c and R_w, respectively. Let also $\Delta R \equiv R_c - R_w$. It follows from Equation (23) that

$$\Delta R = H \ln\left(\frac{\Phi_c}{\Phi_w}\right), \qquad (35)$$

where Φ_c and Φ_w are the Voigt profiles at line center and wing, respectively. If we approximate the Voigt profile as consisting of a Doppler core and Lorentzian wings, then we recover Equation (10) of Heng (2016),

$$\Delta R = H \ln\left[\frac{4\pi^2 c^2 (\lambda - \lambda_0)^2 \lambda_0}{A_{21} \lambda^2 \lambda_0^2 \sqrt{2\pi H g}}\right]. \qquad (36)$$

Thus, the pointwise approach of Heng (2016), as designed for analyzing low-resolution spectra, is a limiting case of the continuous approach given by Equation (23) that we are applying in the current study to high-resolution spectra.

2.6. Data Format for Unnormalized High-resolution Spectra

Let the transit depth be $D = (R/R_\star)^2$, where $R_\star$ is the stellar radius. Ground-based, high-resolution data do not measure the transit depth, but rather the differential transit depth, which is usually denoted by $\tilde{\mathfrak{R}}$ and lacks an absolute empirical normalization. We wish to understand the relationship between D and $\tilde{\mathfrak{R}}$.

Generally, we have $D \ll 1$. Thus, $1 - D$ is a number that is almost unity across wavelength. All of the peaks of D are converted into troughs for $1 - D$. Let the minimum value of D, which corresponds to the continuum, be $D_{\rm min}$. The quantity $\tilde{\mathfrak{R}} = (1 - D)/(1 - D_{\rm min})$ shifts the minimum value of the continuum to unity. The quantity that is typically used for measured spectra is $\tilde{\mathfrak{R}} - 1$ (Wyttenbach et al. 2015, 2017), which shifts the continuum to zero.

3. Results

3.1. Does the Normalization Degeneracy Affect Our Interpretation of the Sodium Doublet?

Before analyzing measured transmission spectra of the sodium doublet, we study a suite of mock retrievals to elucidate the degeneracies between the various model parameters. We also wish to understand the effect of the normalization degeneracy on spectra with a narrow wavelength range centered on the sodium doublet. To generate mock spectra, we assume 100 wavelength bins between 5880 and 5905 Å. Each data point is assumed to have an uncertainty of 0.005 Å on the transit radius. These numbers are guided by the measured HARPS transmission spectrum of WASP-49b (Wyttenbach et al. 2017). We use the measured gravity of WASP-49b of 689 cm^2 s^{-1} (Wyttenbach et al. 2017). For illustration, we assume $T = 4500$ K, $X_{\rm Na} = 10^{-5}$, $f_{\rm broad} = 100$, and $\kappa_{\rm cloud} = 100$ cm^2 g^{-1} in our NLTE model of a mock transmission spectrum.

We study two types of retrievals of mock spectra. Unnormalized spectra are presented in the $\tilde{\mathfrak{R}} - 1$ format previously described in Section 2.6. Normalized spectra are presented in transit radii (R), where we match the continuum of the mock spectrum to the measured white-light radius of $1.198\,R_{\rm J}$ for WASP-49b (Wyttenbach et al. 2017). For the choice of $P_0 = 10$ bar, this requires setting $R_0 = 0.752\ R_{\rm J}$.

The top panel of Figure 3 shows pairs of retrievals where we fix $P_0 = 10$ bar and $R_0 = 0.752\ R_{\rm J}$ versus if we fix only $R_0 = 0.752\ R_{\rm J}$ and allow P_0 to be part of the retrieval. Several striking trends are revealed even by examining the unnormalized spectrum (top left panel). First, the behavior of the temperature and broadening parameter are distinct enough from the other parameters that they may be accurately retrieved, regardless of whether P_0 is known. It demonstrates that our parameterization of line broadening using $f_{\rm broad}$ effectively isolates it from other effects, including temperature variations.

Second, there is value added to the retrieval when the transmission spectrum is normalized (top right panel). The posterior distribution of sodium abundance spans about 5 orders of magnitude (for the FWHM) for the unnormalized spectrum, but shrinks to about 2–3 orders of magnitude simply by matching the spectral continuum to the *HST* white-light radius. Our ignorance of P_0 for the normalized spectrum results in the cloud opacity being loosely unconstrained, but this does not affect our ability to accurately retrieve the temperature and broadening parameter.

Third, for the normalized spectrum (top right panel of Figure 3), both the degeneracy with cloudiness (i.e., any increase in the sodium abundance may be compensated by an increase in the gray cloud opacity) and the normalization degeneracy (i.e., any increase in the sodium abundance may be
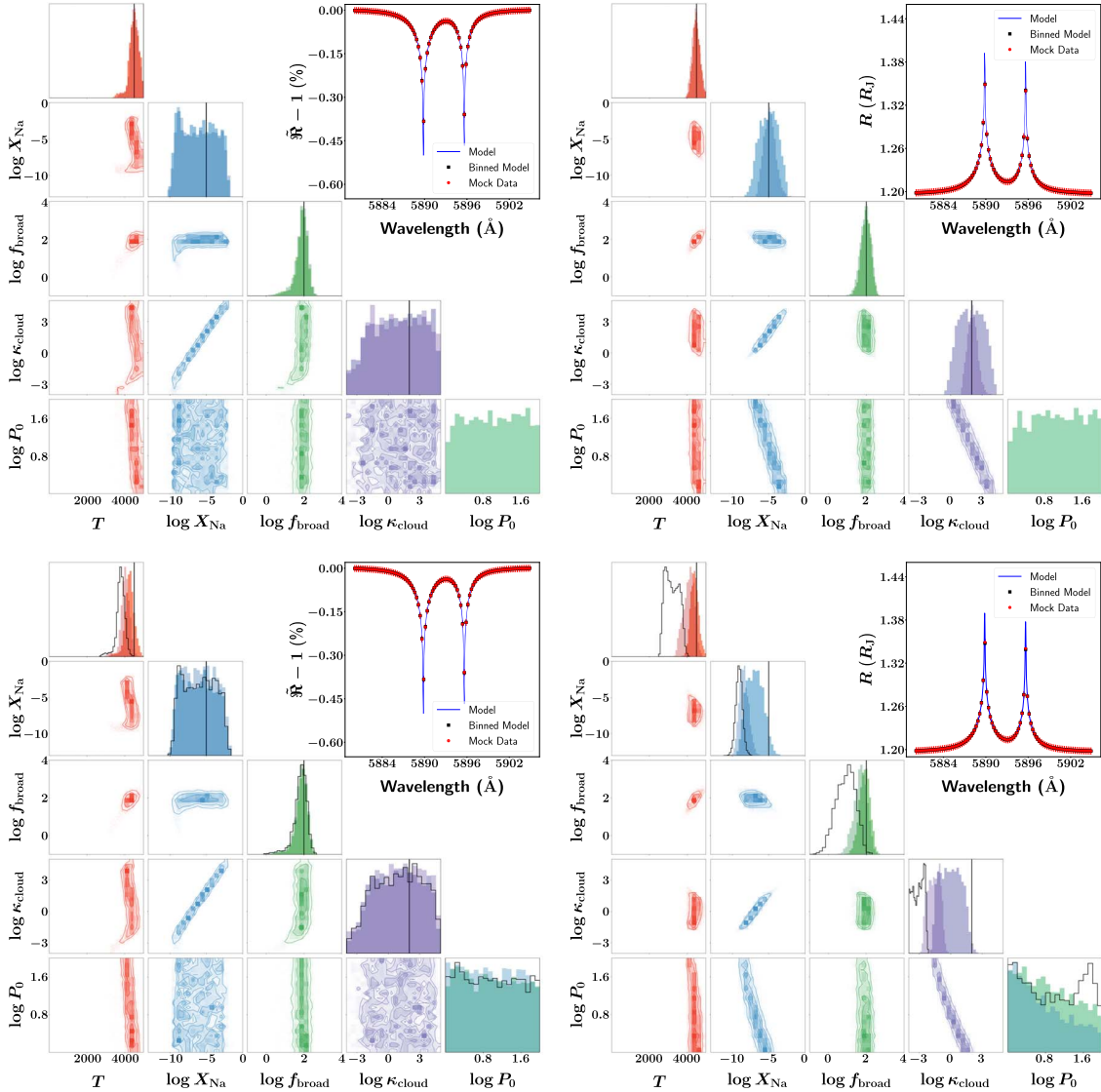
Figure 3. Elucidating the influence of the normalization degeneracy on unnormalized (left column) vs. normalized (right column) spectra, where the spectral continuum of the latter has been matched to the measured white-light *HST* radius of WASP-49b (see the text). This matching produces $R_0 = 0.752\ R_J$, while we have arbitrarily chosen $P_0 = 10$ bar for illustration. We have set $T = 4500$ K, $X_{Na} = 10^{-5}$, $f_{broad} = 100$, and $\kappa_{cloud} = 100\ cm^2\ g^{-1}$, again for illustration. The top row shows mock retrievals where the values of R_0 and P_0 have been fixed to their input values (darker posteriors) vs. those where P_0 is allowed to be a fitting parameter but $R_0 = 0.752\ R_J$ (lighter posteriors). The bottom row shows mock retrievals where the value of R_0 is varied to mimic the real situation where we have no prior knowledge of what its value should be. The light (shaded), dark (shaded), and solid-curve posteriors corresponds to good, not-so-good and bad guesses for R_0, respectively (see the text). Wherever relevant, the vertical lines indicate the true input value of a given parameter.

compensated by a decrease in P_0) are clearly present in the posterior distributions. These degeneracies clearly drive the somewhat large widths of the posterior distributions of both the gray cloud opacity and sodium abundance. When the spectrum is unnormalized (top left panel of Figure 3), the normalization degeneracy vanishes. Essentially, over the width of the posterior distributions of both X_{Na} and κ_{cloud} as determined by the cloudiness degeneracy, the reference pressure is unconstrained. Figure 4 shows the corresponding posterior distributions for the cloudiness index, which is robust to whether P_0 is treated as a fitting parameter and whether the retrieval is performed on unnormalized or normalized spectra.

In a real situation, the value of R_0 is unknown. As demonstrated by Fisher & Heng (2018), it suffices to make a reasonable guess for R_0 and then retrieve for P_0. The bottom

panels of Figure 3 elucidates the effect of good ($R_0 = 0.85R_J$), not-so-good ($R_0 = 0.95R_J$), and bad ($R_0 = 1.05R_J$) guesses. For the normalized spectrum, a bad guess results in a posterior for the cloud opacity that bumps up against its prior value. More insidiously, it also gives a posterior distribution of the sodium abundance that is *narrower* than the one for the good guess. A not-so-good guess also produces misleadingly narrow posterior distributions of the sodium abundance and cloud opacity.

The bottom left panel of Figure 3 shows that these concerns about guessing R_0 are mostly irrelevant for the unnormalized spectrum, because the posterior distributions of parameters are nearly identical for all quantities. The exception is the temperature, which is somewhat different for the bad guess. In other words, the retrieval of unnormalized spectra is unaffected by the
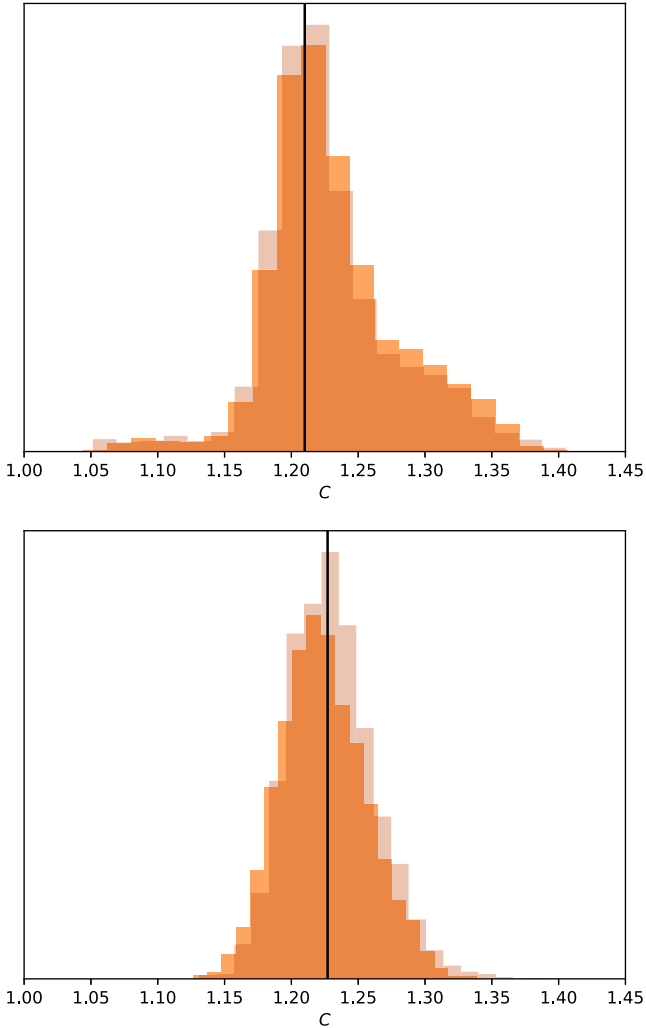
Figure 4. Posterior distribution of the cloudiness index associated with the mock retrieval in Figure 3 for the unnormalized (top panel) and normalized (bottom panel) transmission spectra. The darker posteriors correspond to mock retrievals where P_0 is treated as a fitting parameter.

normalization degeneracy, but at the expense of more uncertain posterior distributions compared to normalized spectra.

3.2. Can We Distinguish between LTE and NLTE Scenarios?

We now use mock retrievals to study a different question, which is whether a suite of nested-sampling retrievals are capable of distinguishing between LTE versus NLTE and cloudfree versus cloudy models? Using the same cloudy NLTE mock spectrum, we perform four retrievals using the cloudfree LTE model, cloudy LTE model, cloudfree NLTE model, and cloudy NLTE model. Figures 5 and 6 show the outcomes of these retrievals. The first major outcome is that the shape of the sodium doublet, in unnormalized transmission spectra, does not encode enough information to distinguish between cloudfree and cloudy models based on the Bayesian evidence. It has the implication that the temperature is underestimated if one uses a cloudfree model to interpret a cloudy transit chord.

A comparison of the Bayesian evidence also does not allow us to distinguish between NLTE and LTE models, which in our

formulation have exactly the same number of parameters. A robust outcome of the retrievals is that the value of the broadening parameter is accurately retrieved regardless of whether cloudy NLTE or LTE models are used. However, when an LTE model is used to perform retrieval on the NLTE mock spectrum, the retrieved temperature is substantially underestimated. In the example shown, we retrieve a temperature of about 2600 K compared to the true input value of 4500 K. This suite of mock retrieval suggests that retrievals performed on unnormalized HARPS transmission spectra of the sodium doublet are unable to discern between NLTE and LTE models.

In both the cloudy LTE and NLTE models, the width of the posterior distribution of the sodium abundance is determined by the degeneracy between the gray cloud opacity and the sodium abundance.

3.3. Retrieval Analysis of HARPS Data of WASP-49b

The mock retrievals of Figure 5 assume idealized conditions, meaning that sources of contamination due to stellar activity, telluric lines, imperfect coaddition of multiple exposures, etc., are not considered. In analyzing the WASP-49b HARPS transmission spectrum of Wyttenbach et al. (2017), we assume that these issues have been addressed by the authors.

Wyttenbach et al. (2017) recorded the peak of each sodium line within a narrow band with a width of 0.4 Å (see their Section 4.2). As discussed in Section 2.6, the reference transit depth, $D_{\min}$, is theoretically taken as the minimum transit depth across the spectrum. However, due to the fluctuations in the continuum points in the data, Wyttenbach et al. (2017) set this reference value to the average measured continuum within two reference bands: blue (5874.94–5886.94 Å) and red (5898.94–5910.94 Å). We apply the same procedure to our models: within the same pair of reference bands, we compute the average value of the continuum. We then take $D_{\min}$ as this value when calculating $\tilde{\mathfrak{R}}$, before shifting it to $\tilde{\mathfrak{R}} = 1$. Our fit to the data using this procedure is shown in Figure 7.

The top row of Figure 8 shows our retrieval analysis of the measured high-resolution, unnormalized transmission spectrum of WASP-49b (Wyttenbach et al. 2017). In our initial retrievals (not shown), we struggled to fit the deep line peaks with our model. This motivated us to extend our prior for f_{broad} to lower values, compared with those stated in Table 2. For this data, we use the range $[10^{-10}, 10^4]$. As expected, we are unable to distinguish between the LTE versus NLTE interpretation from the computed Bayesian evidence (448.8 for LTE versus 448.4 for NLTE). On physical grounds alone, we favor the NLTE interpretation. The temperature is 7209^{+1763}_{-1892} K for the LTE interpretation versus 8415^{+1020}_{-1526} K for the NLTE interpretation, consistent with the lesson learned from our mock retrievals that the LTE model tends to predict a lower temperature. These retrieved temperatures are discrepant from the 2950^{+400}_{-500} K value reported by Wyttenbach et al. (2017), but we note that Wyttenbach et al. (2017) fitted Gaussians rather than Voigt profiles to the measured sodium doublets. The use of Gaussians is equivalent to fitting with a Doppler profile in the absence of Lorentzian wings. Our retrieved temperatures remain consistent with the expectation that radiative cooling by collisionally excited atomic hydrogen thermostats the temperature to $\sim 10^4$ K (Murray-Clay et al. 2009). The retrieved sodium volume mixing ratios are loosely constrained and somewhat insensitive to
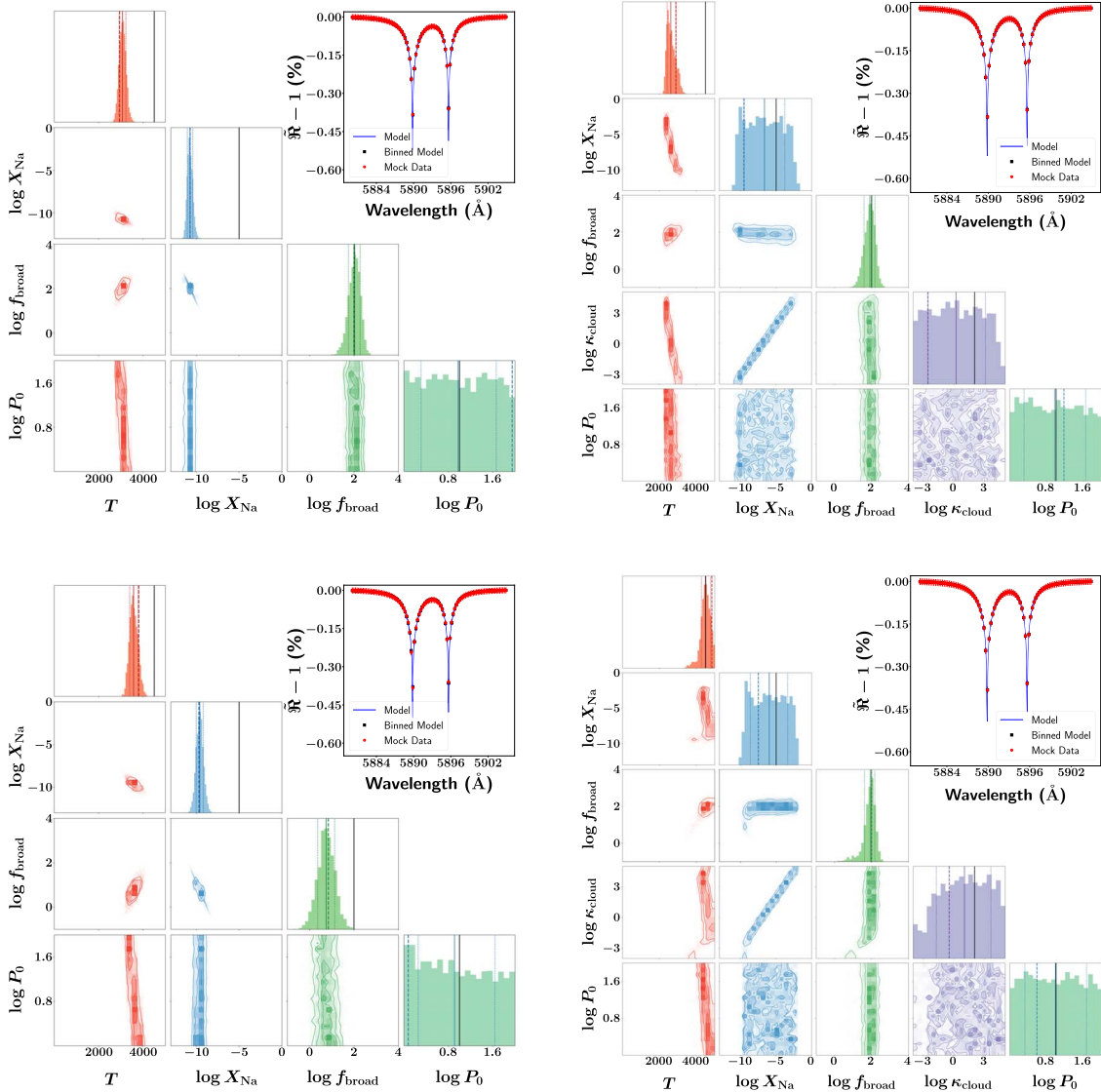
Figure 5. Suite of retrievals on a mock transmission spectrum, with HARPS-like spectral resolution, constructed using a cloudy NLTE model of the sodium doublet. Retrievals are performed assuming a cloudfree LTE model (top left panel), a cloudy LTE model (top right panel), a cloudfree NLTE model (bottom left panel), or a cloudy NLTE model (bottom right panel). The solid and dotted vertical lines show the median and 1σ limits for the posteriors, respectively. The thick dashed lines show the best-fit values. The solid black lines show the truth values for the mock data.

whether the LTE or NLTE interpretation is assumed ($\log X_{\mathrm{Na}} = -5.64^{+2.55}_{-2.90}$ versus $-4.44^{+2.02}_{-2.51}$), as is the retrieved cloud opacity ($\log \kappa_{\mathrm{cloud}} = -0.18^{+3.14}_{-2.77}$ versus $-0.90^{+2.82}_{-2.06}$). The broadening parameter is essentially unconstrained, but takes on $f_{\mathrm{broad}} < 1$ values as the narrow wavelength range precludes full coverage of the sodium line wings.

The continuum of the measured transmission spectrum appears to possess a scatter with a half-width of about 0.005. To test the robustness of our retrieval results, we artificially increased the two line peaks by 0.005 each and reran our retrievals. The retrieval outcomes are shown in the bottom row of Figure 8. For the LTE retrieval, we obtained a reduced temperature of $T = 6756^{+2048}_{-2030}$ K and a similar sodium abundance ($\log X_{\mathrm{Na}} = -6.22^{+3.03}_{-2.86}$). For the NLTE retrieval, the temperature becomes $T = 7925^{+1408}_{-1869}$ K and the sodium abundance becomes $\log X_{\mathrm{Na}} = -4.76^{+2.32}_{-3.01}$. This pair of tests

suggests that the retrieval outcomes are somewhat robust to the uncertainty in the continuum.

3.4. Retrieval Analysis of Low-resolution Transmission Spectra

We examine a sample of six low-resolution transmission spectra: HAT-P-1b, HAT-P-12b, HD 189733b, WASP-6b, WASP-17b, and WASP-39b (Sing et al. 2016). Given the sparseness of the data, we assume a simple model that includes the sodium doublet and a spectral continuum sourced by clouds with both small and large particles of unspecified composition. Large cloud particles are represented by a constant opacity (κ_{cloud}). Following Sing et al. (2016), small cloud particles are assumed to contribute Rayleigh scattering that has a behavior, across wavelength, which is identical to that of molecular hydrogen, but with a magnitude that is offset by some constant, dimensionless factor $f_{\mathrm{cloud}} \geqslant 1$. The cross section for Rayleigh
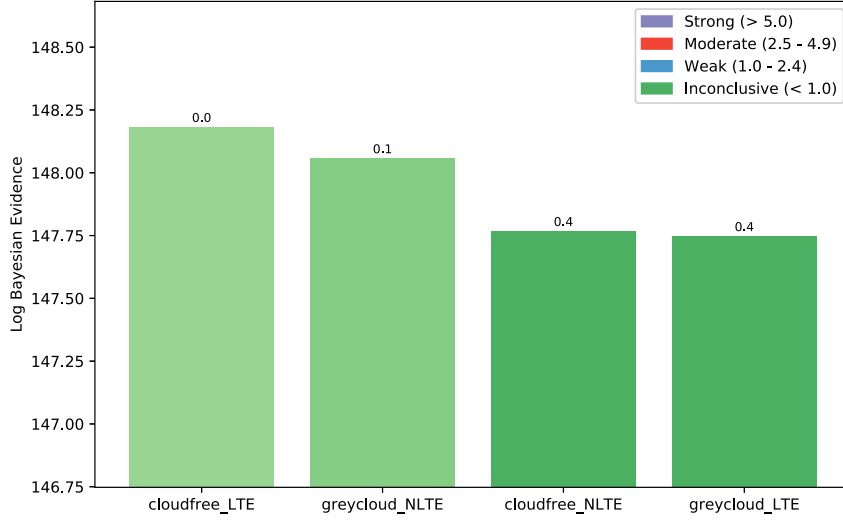
Figure 6. Logarithm of the Bayesian evidence and corresponding Bayes factor between each model compared to the best model (as indicated by the number on top of each bar) for the suite of mock retrievals in Figure 5. The entry marked by "0" is the best model, i.e., the model with the highest Bayesian evidence. The legend lists the correspondence between the Bayes factor and the strength or weakness of the evidence in favor of a given model (compared to the best model).
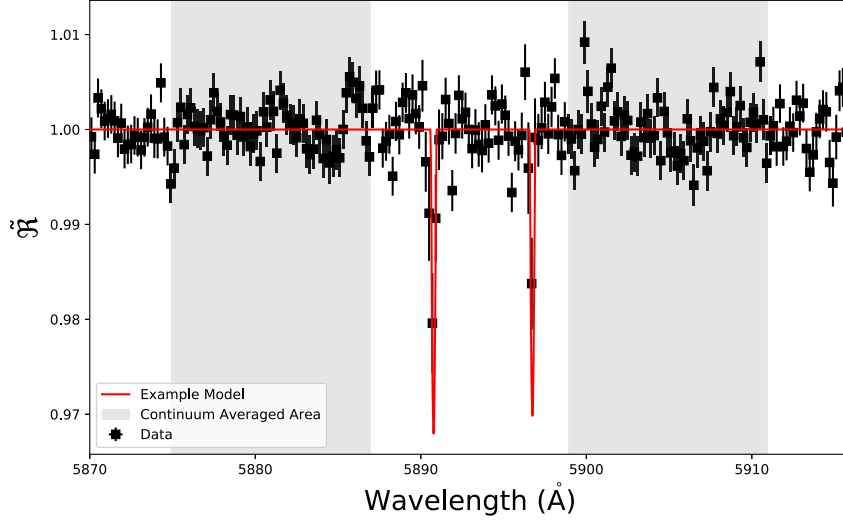
Figure 7. Illustration of how the averaged continuum is measured in blue and red bands (shaded areas) and used to set the reference transit depth, D_{min}. The line peaks are measured in narrow bands with widths of 0.4 Å. As an illustration, we have used a model with $T = 7208$ K, $X_{Na} = 10^{-5.64}$, $\kappa_{cloud} = 10^{-0.18}$ cm^2 g^{-1}, and $f_{broad} = 10^{-5.70}$.

scattering by molecular hydrogen is (Sneep & Ubachs 2005),

$$\sigma_{H_2} = \frac{24\pi^3}{n_{ref}^2 \lambda^4} \left(\frac{n_r^2 - 1}{n_r^2 + 2} \right)^2, \qquad (37)$$

where $n_{ref} = 2.68678 \times 10^{19}$ cm^{-3} and the real part of the index of refraction is (Cox 2000)

$$n_r = 1.358 \times 10^{-4} \left[1 + 7.52 \times 10^{-3} \left(\frac{\lambda}{1\ \mu m} \right)^{-2} \right] + 1. \qquad (38)$$

Collectively, the opacity function is

$$\kappa = \left(\frac{X_{Na} \pi e^2 f_{12}}{m m_e c} \Phi + \frac{X_{H_2} f_{cloud} \sigma_{H_2}}{m} + \kappa_{cloud} \right) \times \left(1 + \frac{n_{crit} k_B T}{P} \right)^{-1}, \qquad (39)$$

where $m_{H_2} = 2 m_{amu}$ is the mass of the hydrogen molecule.

The computed model spectrum is degraded to the measured one by taking the average value of the transit radius in each wavelength bin. Figures 9 and 10 show the outcomes of these 12 retrievals. The reference pressure is either unconstrained or only loosely constrained and its posterior distribution is largely prior-dominated. This is unsurprising, because the pressure probed by the transit chord is

$$P \sim \frac{mg}{f_{cloud} X_{H_2} \sigma_{H_2}} \sqrt{\frac{H}{R}}. \qquad (40)$$

If the atmosphere is cloudfree ($f_{cloud} = 1$) and hydrogen-dominated ($X_{H_2} \approx 0.91$), then the pressure can be uniquely determined. The transit-chord pressure and reference pressure are related by hydrostatic equilibrium. However, if the value of f_{cloud} is a priori unknown, then P and hence P_0 cannot be uniquely determined.
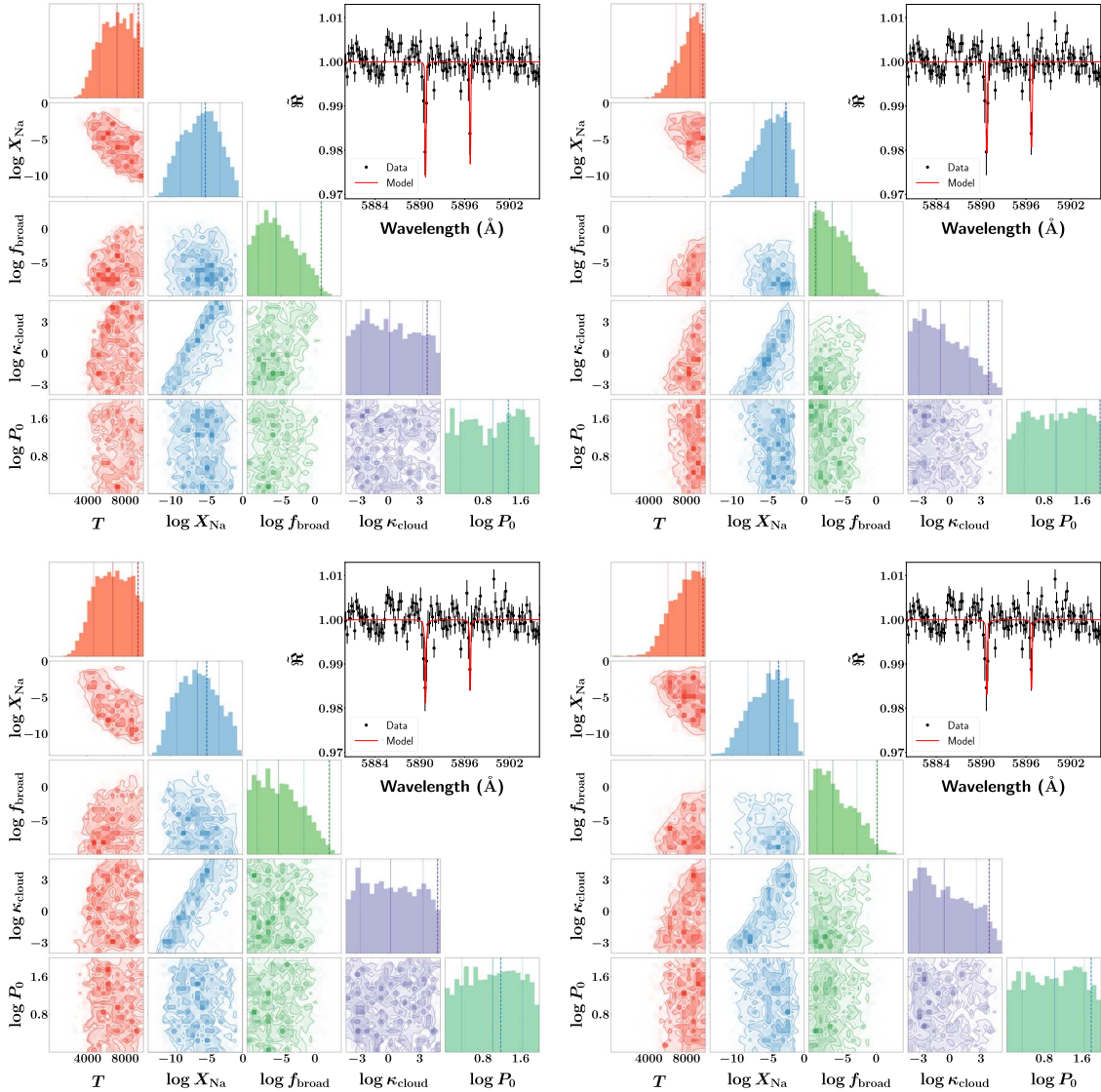
Figure 8. Retrievals on the real HARPS data set for WASP-49b using the LTE (left column) and non-LTE (right column) models. The top row shows the retrievals performed on the original data. For the bottom row, we have artificially increased the two data points corresponding to the line peaks by 0.005 in order to test the sensitivity of the retrieval outcomes to fluctuations in the spectral continuum.

At low spectral resolutions, the trend of decreasing f_{broad} producing strong lines (relative to the continuum) is negated by a higher abundance of sodium, and it is this degeneracy that dominates the width of the posterior distributions of X_{Na}. In all 12 retrievals, the temperature is tightly constrained as it functions like an independent "stretch mode" in the retrieval (Fisher & Heng 2018; see also Figure 2).

HD 189733b is singled out for discussion due to its historical importance in the literature (Pont et al. 2008, 2013; Sing et al. 2011). It is the only object among the sample of six objects that has a clearly cloudy STIS transit chord ($f_{cloud} \gg 1$), but it is important to note that the range of f_{cloud} values retrieved is prior-dominated by our choice of $1 \leqslant P_0 \leqslant 100$ bar. The temperature corresponding to the STIS transit chord is about 2800 K (assuming LTE) or about 4800 K (for NLTE), but these values are not inconsistent with those reported by Wyttenbach et al. (2015) and Heng et al. (2015). In all 12 retrievals, the values of κ_{cloud} retrieved are small, consistent with the continuum being nongray.

Figure 11 shows the empirical trends (or lack thereof) between the various retrieved quantities. For the purpose of computing cloudiness index, the cloudfree scenario is obtained by setting $f_{cloud} = 1$ and $\kappa_{cloud} = 0$. No clear trend between the cloudiness index (C) and the equilibrium temperature, which is a proxy for the insolation or stellar irradiation, is seen, contrary to the tentative trend reported by Heng (2016). The retrieved C values range from being consistent with unity to larger than unity, in broad agreement with the study of Barstow et al. (2017). NLTE interpretations yield a higher degree of cloudiness, because of the diminished strength of the sodium lines relative to LTE.

The uncertainties on the retrieved sodium abundances are large, but the general trend is that X_{Na} is consistent with being solar ($\sim 10^{-6}$) or subsolar. The uncertainties on the retrieved broadening parameters are also large, but the retrieved f_{broad} values are broadly consistent with unity, suggesting that sub-Lorentzian wings are not needed to fit the data.
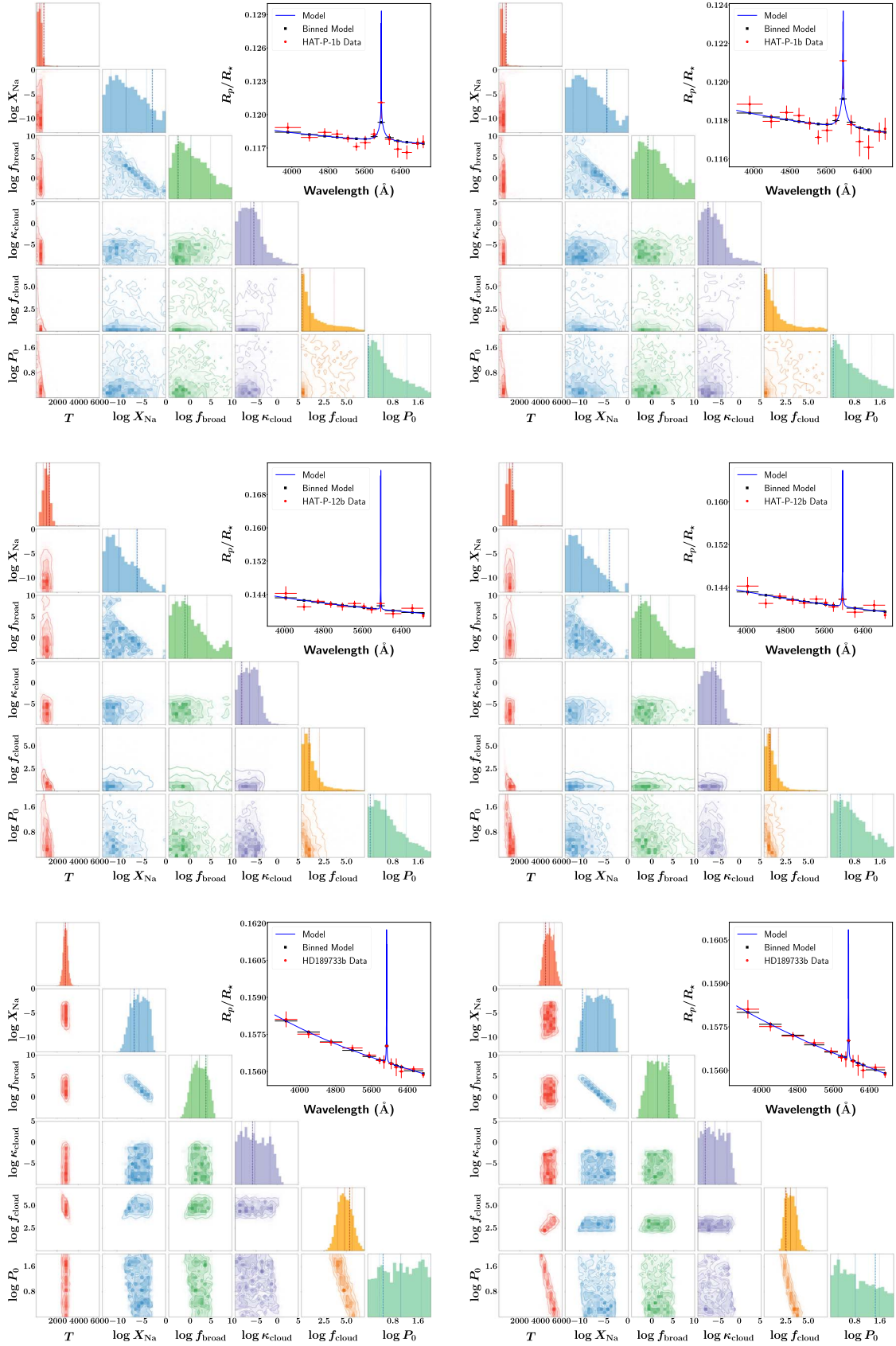
Figure 9. Retrieval analysis of low-resolution transmission spectra. The left and right columns are for the LTE and NLTE interpretations, respectively. The top, middle, and bottom rows are for HAT-P-1b, HAT-P-12b, and HD 189733b, respectively.
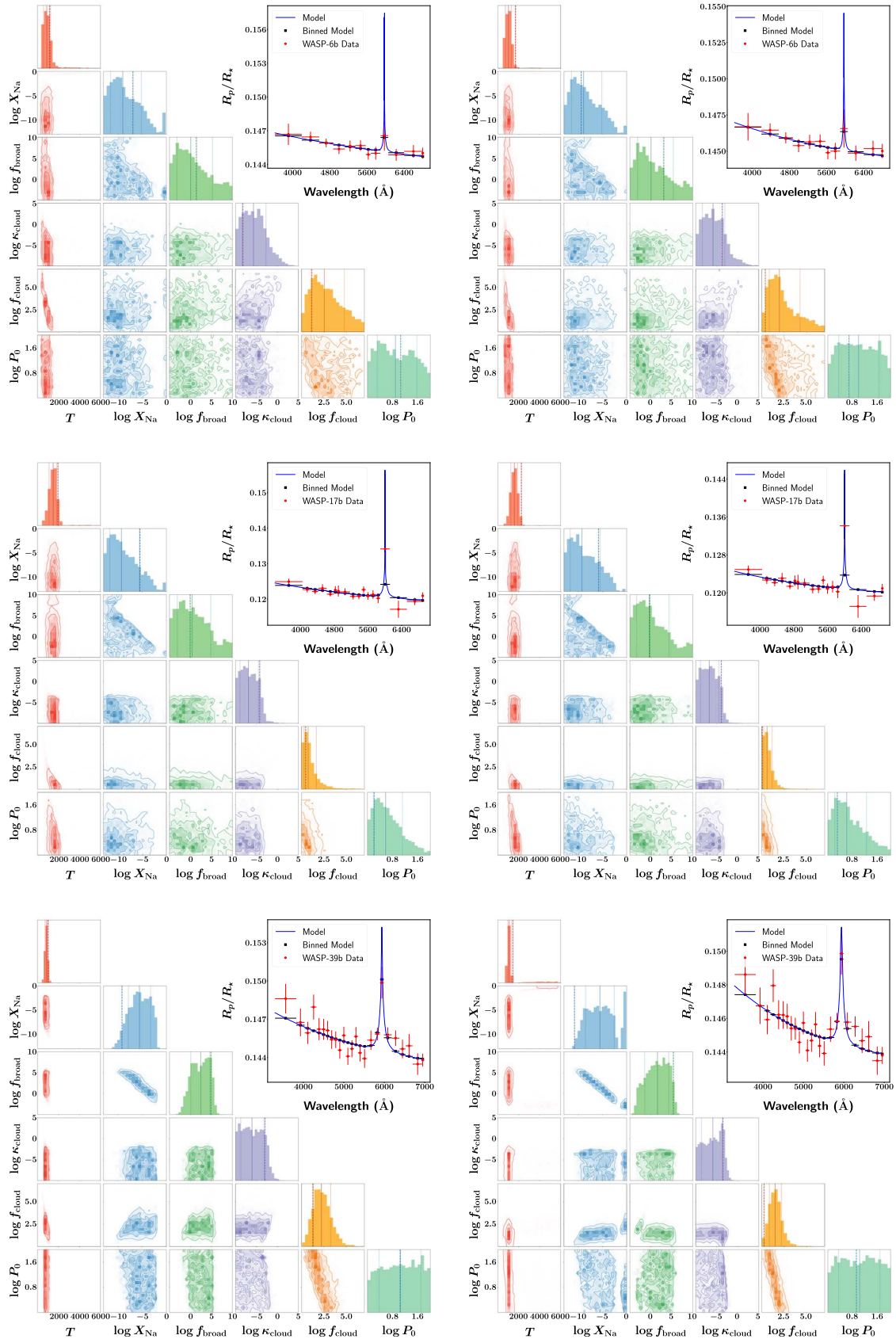
Figure 10. Same as Figure 9, but for WASP-6b (top row), WASP-17b (middle row), and WASP-39b (bottom row).
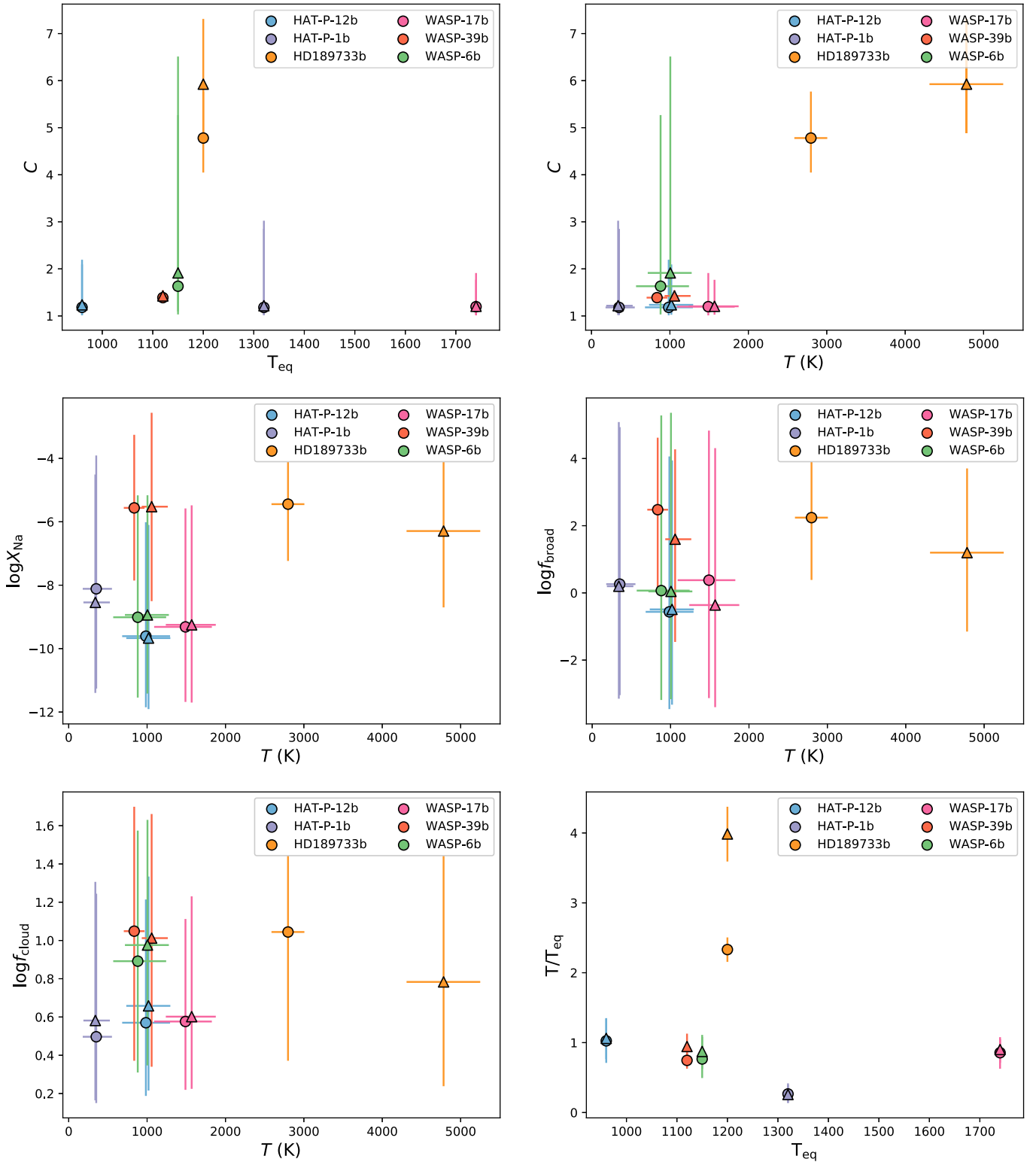
Figure 11. Retrieved properties of the six exoplanets based on the retrieval analysis of their low-resolution transmission spectra. Shown are the cloudiness index vs. equilibrium temperature (top left panel), cloudiness index vs. retrieved temperature (top right panel), sodium abundance vs. temperature (middle left panel), broadening parameter vs. temperature (middle right panel), cloud factor vs. temperature (bottom left panel), and ratio of retrieved to equilibrium temperature vs. equilibrium temperature (bottom right panel). The LTE models are labeled by circles, while the NLTE models are labeled by triangles.

Generally, $T/T_{\rm eq} \sim 1$, where T is the retrieved temperature of the transit chord, with the exception of HD 189733b where $T/T_{\rm eq} > 2$.

4. Discussion

Our retrieval outcomes in Figure 8 clearly demonstrate that ground-based, high-resolution spectra of the sodium doublet alone

do not encode enough information to infer the pressure level being probed by the lines—not even at the order-of-magnitude level. This finding is consistent with the lessons learned from our mock retrievals in Figure 3, where we demonstrated that the normalization degeneracy and the lack of an empirical normalization prevents P_0 from being meaningfully constrained. The finding that ground-based high-resolution spectroscopy is incapable of accurately retrieving the pressures probed appears to contradict the work of Pino et al. (2018) for HD 189733b, but we note that this work did not address the normalization degeneracy and instead assumed a fixed value for R_0 that corresponds to $P_0 = 10$ bar. It is this assumption of fixing the reference radius that allows pressure levels to be inferred. The broader implication of this finding is that one cannot easily infer which part of the atmosphere one is probing (i.e., thermosphere, exosphere) if the sodium lines are analyzed in isolation. Future work should elucidate if ground-based, high-resolution spectra alone (with their lack of an empirical normalization), even across an extended wavelength range, encode enough information (e.g., in the sodium line wings) to provide precise information on the pressures probed by the transit chord.

In the current study, we have chosen to focus only on the optical part of the spectrum that contains the sodium lines in order to elucidate the limitations associated with such a restricted analysis. Future work should combine spectra from the optical and infrared, and elucidate if the near-infrared water lines probed by WFC3 also require an NLTE treatment as has been presented in the current study.

While the retrieved outcomes in the current study are consistent with Voigt profiles for the sodium line shapes, future work should elucidate a theory of $f_{broad} = f_{broad}(P)$ that reconciles and unifies the work of Burrows et al. (2000) and Allard et al. (2012).

Appendix
Electron Density via Solution of the Saha Equation

For completeness, we provide the expression for n_e if the electrons were entirely sourced by the collisional ionization of the sodium atom,

$$\text{Na} \longleftrightarrow \text{Na}^+ + e^-. \quad (41)$$

To render the problem tractable, we retain the assumption that n_e is a solution of the Saha equation, i.e., the ionization states are in LTE, which reads (e.g., Section 3.4 of Draine 2011)

$$\frac{n_e n_{\text{Na}^+}}{n_{\text{Na}}} = \frac{2(2\pi m_e k_B T)^{3/2}}{h^3} \frac{g_{\text{Na}^+}}{g_{\text{Na}}} e^{-E_{\text{ion}}/k_B T}, \quad (42)$$

where E_{ion} is the ionization energy. g_{Na^+} and g_{Na} are the quantum degeneracies associated with Na^+ and Na, respectively. If we assume that the electrons are solely provided by

the ionization of the sodium atom, then we have $n_e = n_{\text{Na}^+}$ and

$$n_e = \frac{(2\pi m_e k_B T)^{3/4}}{h^{3/2}} \sqrt{\frac{g_{\text{Na}^+}}{g_{\text{Na}}}} \sqrt{2X_{\text{Na}} n_{\text{total}}} \; e^{-E_{\text{ion}}/2k_B T}, \quad (43)$$

where $X_{\text{Na}} \equiv n_{\text{Na}}/n_{\text{total}}$ is the volume mixing ratio of sodium. Previously, the preceding expression was used to describe the number density of electrons in protoplanetary disks for the purpose of studying the magnetorotational instability, where it was assumed that all of the electrons were sourced by potassium, which has an ionization energy of $E_{\text{ion}} = 4.3407$ eV (Balbus & Hawley 2000). It was also used to study ohmic dissipation in hot Jupiters (Perna et al. 2010). In the current situation, we assume that the electrons are sourced only by sodium, which has $E_{\text{ion}} = 5.1391$ eV. We have $g_{\text{Na}^+}/g_{\text{Na}} = 1/2$ because the singly charged ion possesses a closed shell of electrons, but the neutral atom has a single valence electron in the $3s$ orbital.

ORCID iDs

Chloe Fisher https://orcid.org/0000-0003-0652-2902
Kevin Heng https://orcid.org/0000-0003-1907-5910

References

Abramowitz, M., & Stegun, I. A. 1970, Handbook of Mathematical Functions (Dover: New York) 9th Printing
Allard, N. F., Kielkopf, J. F., Spiegelman, F., Tinetti, G., & Beaulieu, J. P. 2012, A&A, 543, A159
Arfken, G. B., & Weber, H. J. 1995, Mathematical Methods for Physicists (4th ed.; New York: Academic)
Balbus, S. A., & Hawley, J. F. 2000, SSRv, 92, 39
Barstow, J. K., Aigrain, S., Irwin, P. G. J., & Sing, D. K. 2017, ApJ, 834, 50
Benneke, B., & Seager, S. 2012, ApJ, 753, 100
Bétrémieux, Y., & Swain, M. R. 2017, MNRAS, 467, 2834
Brogi, M., Line, M., Bean, J., Désert, J.-M., & Schwarz, H. 2017, ApJL, 839, L2
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Burrows, A., Marley, M. S., & Sharp, C. M. 2000, ApJ, 531, 438
Casasayas-Barris, N., Pallé, E., Chen, G., et al. 2018, A&A, 616, A151
Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002, ApJ, 568, 377
Cox, A. N. 2000, Allen's Astrophysical Quantities (4th ed.; New York: Springer)
de Wit, J., & Seager, S. 2013, Sci, 342, 1473
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium (Princeton, NJ: Princeton Univ. Press)
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, arXiv:1306.2144
Fisher, C., & Heng, K. 2018, MNRAS, 481, 4698
Fortney, J. J., Sudarsky, D., Hubeny, I., et al. 2003, ApJ, 589, 615
Gibson, N. P., de Mooij, E. J. W., Evans, T. M., et al. 2019, MNRAS, 482, 606
Goody, R. M., & Yung, Y. L. 1989, Atmospheric Radiation: Theoretical Basis (2nd ed.; Oxford: Oxford Univ. Press)
Griffith, C. A. 2014, RSPTA, 372, 86
Heng, K. 2016, ApJL, 826, L16
Heng, K. 2017, Exoplanetary Atmospheres: Theoretical Concepts and Foundations (Princeton, NJ: Princeton Univ. Press)
Heng, K., & Kitzmann, D. 2017, MNRAS, 470, 2972
Heng, K., Wyttenbach, A., Lavie, B., et al. 2015, ApJL, 803, L9
Jensen, A. G., Redfield, S., Endl, M., et al. 2012, ApJ, 743, 203
Jordán, A., & Espinoza, N. 2018, RNAAS, 2, 149
Khalafinejad, S., von Essen, C., Hoeijmakers, H. J., et al. 2017, A&A, 598, A131
Lavvas, P., Koskinen, T., & Yelle, R. V. 2014, ApJ, 796, 15
Lecavelier des Etangs, A., Pont, F., Vidal-Madjar, A., & Sing, D. 2008, A&A, 481, L83
Lendl, M., Anderson, D. R., Collier-Cameron, A., et al. 2012, A&A, 544, A72
Louden, T., & Wheatley, P. J. 2015, ApJL, 814, L24
Murray-Clay, R. A., Chiang, E. I., & Murray, N. 2009, ApJ, 693, 23
Nikolov, N., Sing, D. K., Fortney, J. J., et al. 2018, Natur, 557, 526
Penner, S. S. 1952, JChPh, 20, 507

Perna, R., Menou, K., & Rauscher, E. 2010, ApJ, 724, 313
Pino, L., Ehrenreich, D., Wyttenbach, A., et al. 2018, A&A, 61, A53
Pont, F., Knutson, H., Gilliland, R. L., Moutou, C., & Charbonneau, D. 2008, MNRAS, 385, 109
Pont, F., Sing, D. K., Gibson, N. P., et al. 2013, MNRAS, 432, 2917
Redfield, S., Endl, M., Cochran, W. D., & Koesterke, L. 2008, ApJL, 673, L87
Roston, G. D., & Obaid, F. S. 2005, JQSRT, 94, 255
Rothman, L. S., Rinsland, C. P., Goldman, A., et al. 1998, JQSRT, 60, 665
Schreier, F. 1992, JQSRT, 48, 743
Seager, S., & Sasselov, D. D. 2000, ApJ, 537, 916
Sing, D. K., Fortney, J. J., Nikolov, N., et al. 2016, Natur, 529, 59

Sing, D. K., Pont, F., Aigrain, S., et al. 2011, MNRAS, 416, 1443
Sneep, M., & Ubachs, W. 2005, JQSRT, 92, 293
Snellen, I. A. G., Albrecht, S., de Mooij, E. J. W., & Le Poole, R. S. 2008, A&A, 487, 357
Stevenson, K. B. 2016, ApJL, 817, L16
Sudarsky, D., Burrows, A., & Pinto, P. 2000, ApJ, 538, 885
Trotta, R. 2008, ConPh, 49, 71
Wyttenbach, A., Ehrenreich, D., Lovis, C., Udry, S., & Pepe, F. 2015, A&A, 577, A62
Wyttenbach, A., Lovis, C., Ehrenreich, D., et al. 2017, A&A, 602, A36
Zaghloul, M. R. 2007, MNRAS, 375, 1043

Supervised machine learning for analysing spectra of exoplanetary atmospheres

Being a scientist is like being an explorer. You have this immense curiosity, this stubbornness, this resolute will that you will go forward no matter what other people say.

<div align="right">

SARA SEAGER

</div>

4.1 Summary

In parallel to my work on traditional Bayesian retrievals, I began developing a new method of retrieval using machine learning. This work is part of a close collaboration I have built with two computer scientists, P. Márquez-Neila and R. Sznitman, based at the ARTORG Center for Biomedical Engineering at the University of Bern. Their expertise in machine learning allowed for the fast development of this new retrieval technique, leading to this initial proof-of-concept paper less than a year after our collaboration began.

This work was motivated by several issues that arise from traditional Bayesian retrievals. Firstly, in a single retrieval $\sim$10,000 atmospheric models are computed on-the-fly, with each one compared to the data to check the goodness of fit. These models are typically discarded after the retrieval, thus resulting in repeated computations across multiple retrievals. A single retrieval can range anywhere from a matter of minutes up to several days or even longer, depending on the complexity and efficiency of the model. With the increasing number of exoplanet spectra, it is imperative that a faster and more efficient retrieval method be employed. Secondly, the development of next-generation observatories, such as JWST and the ELT, will lead to an explosion in the sensitivity and precision of exoplanet spectra. The computational restrictions of traditional retrievals limit their use of complex three-dimensional physical models. Several studies have shown that the typical 1-D models used in most retrievals will be insufficient for modelling this new era of data (e.g. Feng et al., 2016; Taylor et al., 2020). The advantage of a machine learning retrieval is that the training set of models are pre-computed, shifting the computational burden offline, and thus allowing for more sophisticated models to be applied.

In this paper, our machine learning retrieval technique is introduced, which uses a Random Forest trained on a large set of simulated spectra. The forest was then applied to the HST WFC3 transmission spectrum of WASP-12 b (Kreidberg et al., 2015) to obtain posteriors for the parameters, and these results were then compared to the traditional nested-sampling retrieval, finding a good agreement. Two additional analyses are also presented that come free with the Random Forest, namely the predicted vs real plots and the information content analysis. The predicted vs real plots show the relationship between the forest's predictions and the true values of the test spectra's parameters. This is the equivalent of performing thousands of mock retrievals, and it allows the user to determine the predictability of each parameter across its range of values. The information content analysis, known as the "feature importance", quantifies the importance of each spectral point in the retrieval of each individual parameter. This information is extremely useful in telescope proposals, and could even be used to inform future instrument design.

Although this paper was led by P. Márquez-Neila, who wrote the Random Forest algorithm, I contributed significantly to the study, working in parallel with Márquez-Neila to test the forest and compare it to the traditional retrieval. I also

computed the training set for the forest. This work now represents one of only a handful of papers using machine learning in exoplanet retrievals.

4.2 Publication

This work was published in Nature Astronomy Letters in 2018.

Supervised machine learning for analysing spectra of exoplanetary atmospheres

Pablo Márquez-Neila[1,2], Chloe Fisher[2], Raphael Sznitman[1] and Kevin Heng [ID][2]*

The use of machine learning is becoming ubiquitous in astronomy[1-3], but remains rare in the study of the atmospheres of exoplanets. Given the spectrum of an exoplanetary atmosphere, a multi-parameter space is swept through in real time to find the best-fit model[4-6]. Known as atmospheric retrieval, this technique originates in the Earth and planetary sciences[7]. Such methods are very time-consuming, and by necessity there is a compromise between physical and chemical realism and computational feasibility. Machine learning has previously been used to determine which molecules to include in the model, but the retrieval itself was still performed using standard methods[8]. Here, we report an adaptation of the 'random forest' method of supervised machine learning[9,10], trained on a precomputed grid of atmospheric models, which retrieves full posterior distributions of the abundances of molecules and the cloud opacity. The use of a precomputed grid allows a large part of the computational burden to be shifted offline. We demonstrate our technique on a transmission spectrum of the hot gas-giant exoplanet WASP-12b using a five-parameter model (temperature, a constant cloud opacity and the volume mixing ratios or relative abundances of molecules of water, ammonia and hydrogen cyanide)[11]. We obtain results consistent with the standard nested-sampling retrieval method. We also estimate the sensitivity of the measured spectrum to the model parameters, and we are able to quantify the information content of the spectrum. Our method can be straightforwardly applied using more sophisticated atmospheric models to interpret an ensemble of spectra without having to retrain the random forest.

We use the previously analysed Hubble Space Telescope Wide Field Camera 3 (WFC3) transmission spectrum of the hot gas-giant WASP-12b, where the volume mixing ratio of water was inferred to be ~10^{-4} to ~10^{-2} and the temperature ~1000 K (ref. [12]). Transmission spectra measure the wavelength-dependent obscuration of starlight by a transiting exoplanet, which encodes signatures of absorption by molecules and clouds in the exoplanetary atmosphere. The choice of this spectrum was to ensure continuity between previous studies[11,12] and because we expect WFC3 to be the workhorse for measuring exo-atmospheric spectra for the immediate future. We implement the random forest method[9,10], which is a supervised form of machine learning. It combines the use of a decision tree[13] and bootstrapping with replacement, and may be used on both discrete and continuous training sets. A decision tree is a way of splitting a training set into subsets on the basis of the common characteristics of its members[14]. The splitting is performed to maximize the gain in information entropy[14]. Since decision trees are sensitive to slight changes in the training set, they are suitable for use with the bootstrapping method, which constructs the decision tree by randomly drawing from the training set[14].

The training set consists of 80,000 synthetic WFC3 transmission spectra, each described by five parameters: temperature (T), the volume mixing ratios (relative abundances of molecules) of water (X_{H_2O}), ammonia (X_{NH_3}) and hydrogen cyanide (X_{HCN}) and a constant cloud opacity (κ_0). Given that these five parameters represent continuous data, we make use of 'regression trees' rather than decision trees (which are used for discrete data) in our random forest[14]. For each spectrum, the values of the five parameters are randomly generated either from a log-uniform (volume mixing ratios and cloud opacities) or uniform (temperature) distribution. In addition to adopting the same wavelength range and 13 bins of the measured WASP-12b WFC spectrum[12], we assume a noise floor of 50 ppm on the transit depth.

In a general machine-learning situation, each member of a training set is associated with a number of characteristics known as features (in the jargon of machine learning), for example, colour, height, type of terrain. For a spectrum, the features are the number of data points it contains. Here, the WFC3 spectrum has 13 features or binned data points. Within the training set, each synthetic spectrum is identified by its values of the five parameters. The training set of 80,000 synthetic spectra resides in a 13-dimensional space, where each dimension corresponds to a wavelength bin. Along the axis of each dimension is a continuous range of values of the transit radii. The goal is to relate an entry (a synthetic spectrum) in this 13-dimensional space with the range of values of each of the five parameters. We accomplish this by subdividing the 13-dimensional space into patches or islands, which is handled using a regression tree. Each patch encompasses some subset of the training set, from which the variance in the transit radii may be computed. The subdivision of the 13-dimensional space is done to minimize the sum of the variances of all of these patches. This is conceptually equivalent to maximizing the gain in information entropy for discrete data[14].

On setting up the regression tree, we use it in tandem with a bootstrapping method. To train each regression tree, we randomly draw from the 80,000 synthetic spectra in the training set. On each draw, the drawn synthetic spectrum is placed back into the training set, allowing it to be drawn more than once. Each regression tree may be visualized as being a predictive 'voter', who returns the ranges of parameter values given the 13 data points of the measured WFC3 transmission spectrum. While a single regression tree produces predictions with large uncertainties, random forests mitigate this pitfall by combining the responses of multiple trees. We performed tests that indicate a convergence of these predictions using 1,000 regression trees (see Methods). Using 1,000 regression trees to form a random forest, we are able to compute the posterior distributions of the parameters[15].

Figure 1 shows the posterior distributions of the temperature, cloud opacity and volume mixing ratios of water, ammonia

[1]ARTORG Center for Biomedical Engineering, University of Bern, Bern, Switzerland. [2]Center for Space and Habitability, University of Bern, Bern, Switzerland.
*e-mail: kevin.heng@csh.unibe.ch
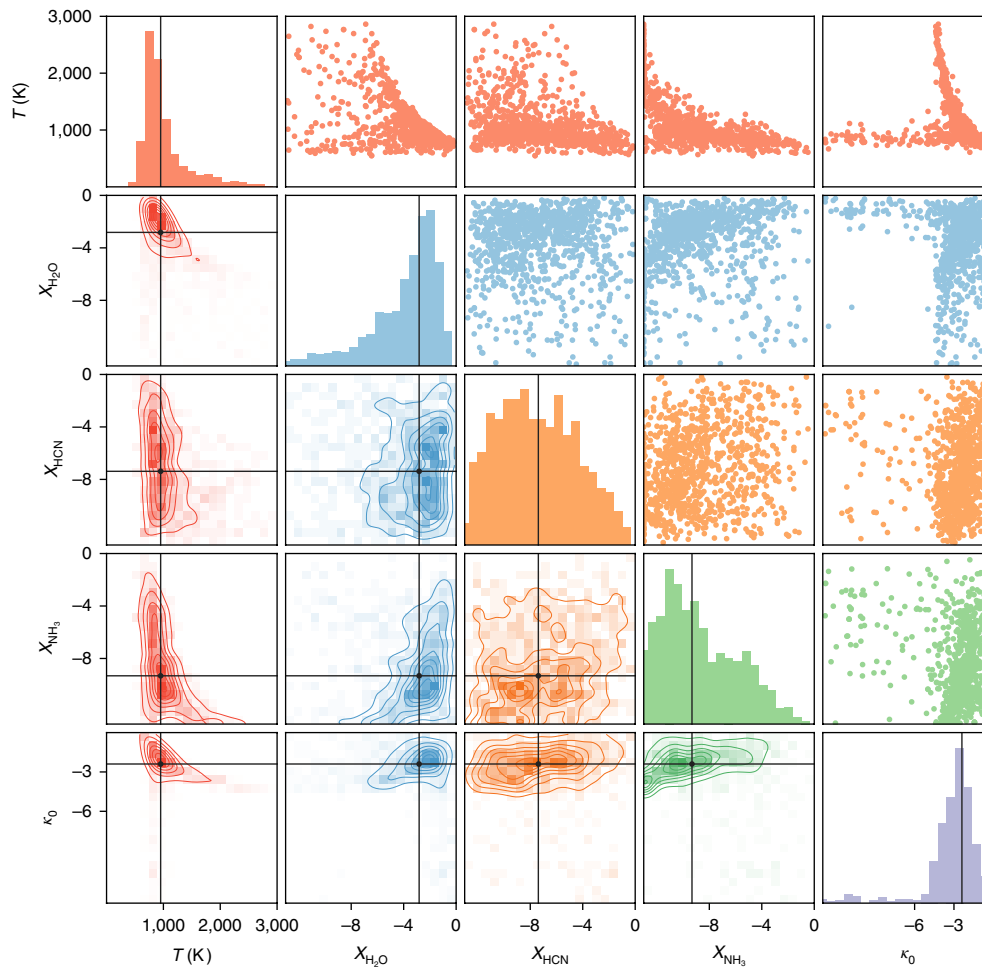
Fig. 1 | Posterior distributions of the volume mixing ratios, temperature and cloud opacity obtained from the machine-learning retrieval analysis of the WFC3 transmission spectrum of WASP-12b. The logarithm (base ten) of the volume mixing ratios (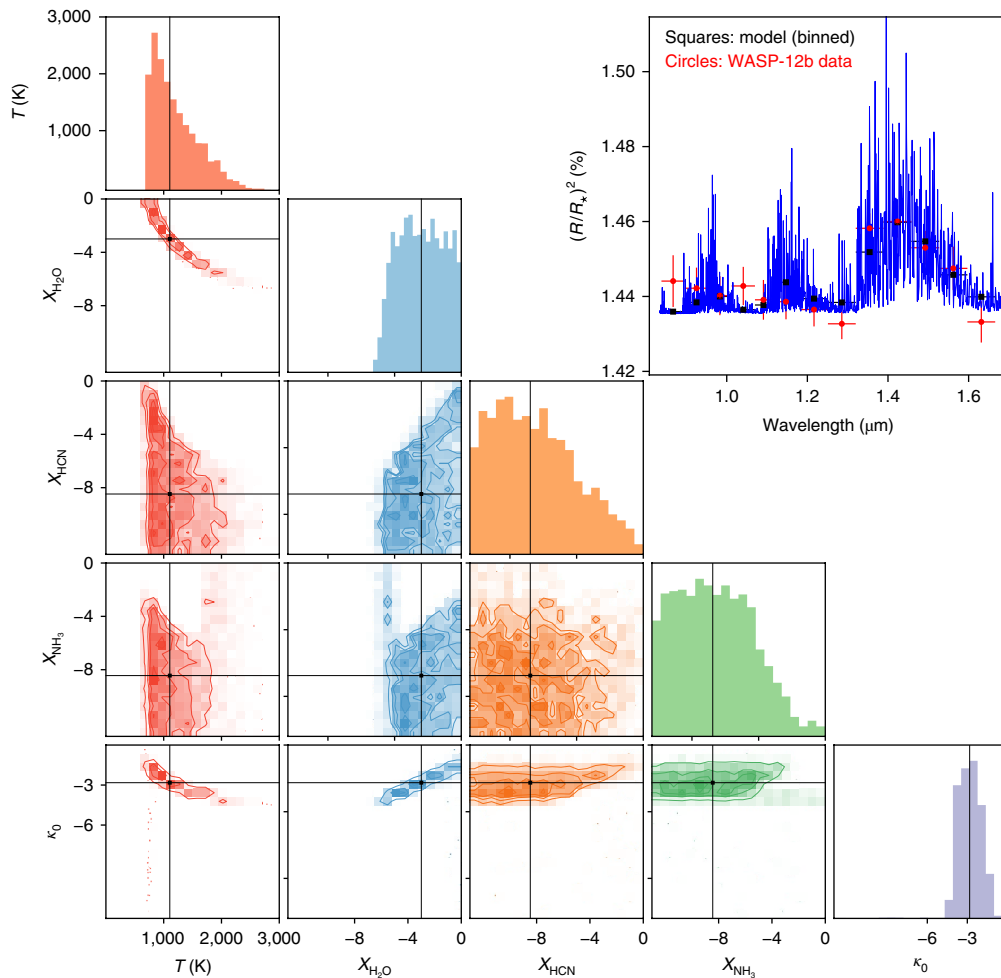relative molecular abundances) of H_2O, HCN and NH_3, and of cloud opacity (κ_0) are shown. Within each scatter plot, each dot is an individual prediction of a single regression tree in the random forest. The straight lines indicate the median values of the parameters. Note that the volume mixing ratios and cloud opacity are associated with a factor ($P_0 = 10$ bar) due to the normalization degeneracy.

and hydrogen cyanide. The retrieved water volume mixing ratio ($\log X_{H_2O} = -2.8^{+1.4}_{-3.6}$) and temperature ($T = 952^{+412}_{-151}$ K) values are broadly consistent with the previous analysis[12]. A non-zero cloud opacity ($\log \kappa_0 = -2.3^{+1.1}_{-1.6}$) is necessary to flatten the spectral continuum blueward of the 1.4 μm water feature. The degeneracies between the temperature, molecular abundances and cloud opacity are consistent with physical expectations. As the temperature increases linearly, the molecular opacities increase exponentially, a property that may be compensated by an order of magnitude decrease in the volume mixing ratio. An increasing temperature also reduces the differences in opacity between the water feature and the spectral continuum blueward of it, a property that may be mimicked to some extent by the cloud opacity. Clouds blunt the strength of molecular features, which may be compensated by order of magnitude increases in the abundances.

The retrieved volume mixing ratios of ammonia and hydrogen cyanide are several orders of magnitude lower than that of water: $\log X_{HCN} = -7.6^{+3.3}_{-3.0}$, $\log X_{NH_3} = -9.2^{+4.2}_{-2.9}$. Running a pair of nested-sampling retrievals shows that the Bayes factor[16] between a model with only water versus one with all three molecules is 0.6 (with the former having the higher Bayesian evidence), implying that there is a lack of evidence for strongly favouring one model over the other. Essentially, there is no evidence for claiming the detection of either hydrogen cyanide or ammonia.

As a consistency check, Fig. 2 shows the posterior distributions of parameters from our nested-sampling retrieval[17,18]. The retrieved parameter values from the nested-sampling retrieval are $T = 1,105^{+545}_{-287}$ K, $\log X_{H_2O} = -3.0^{+2.0}_{-1.9}$, $\log X_{HCN} = -8.5^{+3.8}_{-2.9}$, $\log X_{NH_3} = -8.4^{+3.1}_{-2.9}$, $\log \kappa_0 = -2.8 \pm 0.9$. It is worth noting that the interpretation of transmission spectra suffers from a 'normalization degeneracy'[5,11]. Breaking this normalization degeneracy requires a unique relationship between a reference transit radius (R_0) and reference pressure (P_0) to be specified, which cannot be directly inferred from the WFC3 data alone. In practice, this means that instead of the volume mixing ratio of molecules (X_i), the quantity X_iP_0 is retrieved. In the results shown, we have set $R_0 = 1.79 R_J$ (where R_J is the radius of Jupiter) and $P_0 = 10$ bar to facilitate comparison with a previous study[12].

Having demonstrated that we can use supervised machine learning to perform atmospheric retrieval, we now push beyond the standard analysis. First, we would like to check the values of the five parameters predicted by the random forest method versus 'ground truth' values. For the latter, we generate another 20,000 WFC3 synthetic transmission spectra. We then apply our random forest method, previously trained on the 80,000 synthetic spectra, to predict the parameter values of these 20,000 new synthetic spectra. Figure 3 shows that there is a one-to-one correspondence between

Fig. 2 | Posterior distributions of the volume mixing ratios, temperature and cloud opacity obtained from the nested-sampling retrieval. The logarithm (base ten) of the volume mixing ratios (relative molecular abundances) of H_2O, HCN and NH_3, and of cloud opacity (κ_0) are shown. Within each scatter plot, each dot is an individual prediction of a single regression tree in the random forest. The straight lines indicate the median values of the parameters. Note that the volume mixing ratios and cloud opacity are associated with a factor ($P_0 = 10$ bar) due to the normalization degeneracy. Top right: the measured versus best-fit model transmission spectra, where R is transit radius and $R_\star$ is stellar radius.

the predicted and real values, albeit with some scatter. To verify that the scatter is due to intrinsic model degeneracies (physics) and not due to our implementation of the random forest method itself, we performed other suites of calculations with different numbers of regression trees and noise floors (see Methods).

This comparison between the predicted versus real parameter values provides a rough estimate of the minimum values of the parameters to which the retrieval is sensitive, given the noise model assumed (a constant 50 ppm in our case). For example, the linear trend between the predicted versus real values of the volume mixing ratios of water, hydrogen cyanide and ammonia starts to flatten below $\sim 10^{-6}$, suggesting that volume mixing ratios below 1 ppm are undetectable given the WFC3 transmission spectrum of WASP-12b.

Second, we can use our approach to analyse the information content of the measured WFC3 transmission spectrum. While information content analysis has been previously considered[6,19,20], we offer a complementary analysis and show that this is a natural outcome of the random forest method, called the feature importance analysis. Figure 4 shows the relative weight of each of the 13 data points in the WFC3 transmission spectrum towards determining the value of each parameter. Physical intuition tells us that the data points at around 1.4 μm are the most constraining for the water abundance. The feature importance analysis shows that the two data points near

1.4 μm contain about 30% of the information that goes towards constraining the volume mixing ratio of water. The two bluest data points contain more than 40% of the information needed to constrain the cloud opacity, because they quantify the flatness of the spectral continuum. The two reddest data points are most constraining for hydrogen cyanide.

There are straightforward extensions of random forest retrieval for which no conceptual obstacles exist. We have demonstrated the method on a spectrum with 13 data points, but the random forest method has been shown to work well even for 1,000–10,000 data points[21–25]. This property implies that random forest retrieval is applicable to future James Webb Space Telescope (JWST) spectra spanning a broader range of wavelengths with $\sim$100–1,000 data points[26]. The information content analysis may be used to influence observational campaigns and the design of spectrographs, depending on the intended scientific goal.

Another straightforward extension is to train a random forest once and apply it to an ensemble of spectra. In the current study, we picked a specific object (WASP-12b) to demonstrate our method. There is no conceptual obstacle to making model grids where the surface gravity is allowed to vary. The random forest is trained on this larger grid, but the value of the surface gravity may be fixed to the measured value of a specific object during analysis with no need
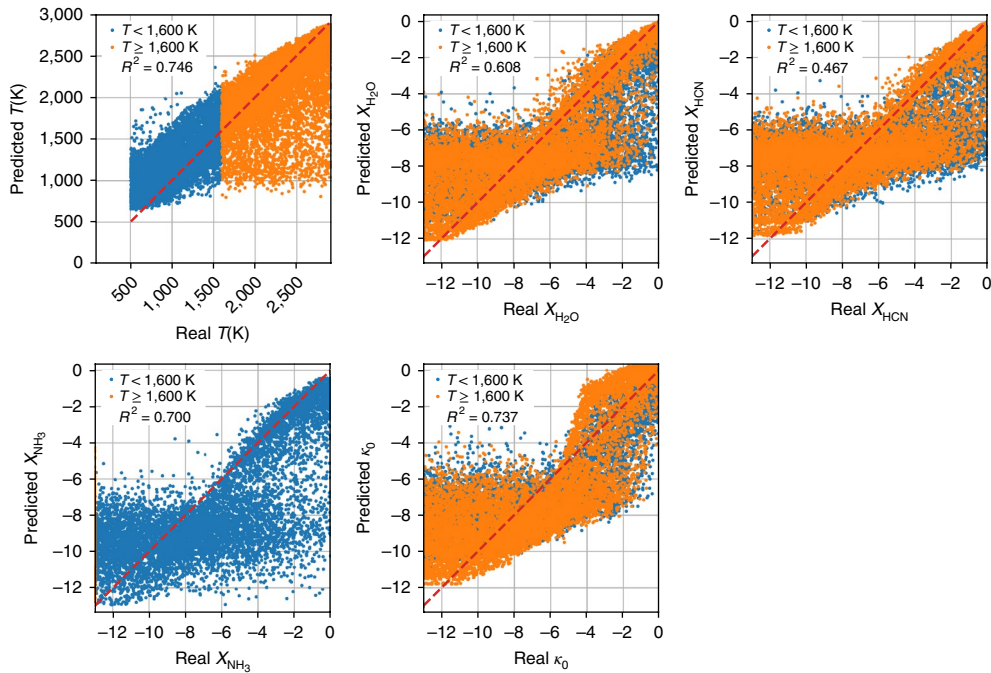
Fig. 3 | True versus random forest predicted values of the five parameters in our transmission spectrum model. The coefficient of determination (R^2) varies from 0 to 1, where values near unity (red dashed line) indicate strong correlations between the predicted and real values of a given parameter, on the basis of the variance of outcomes.
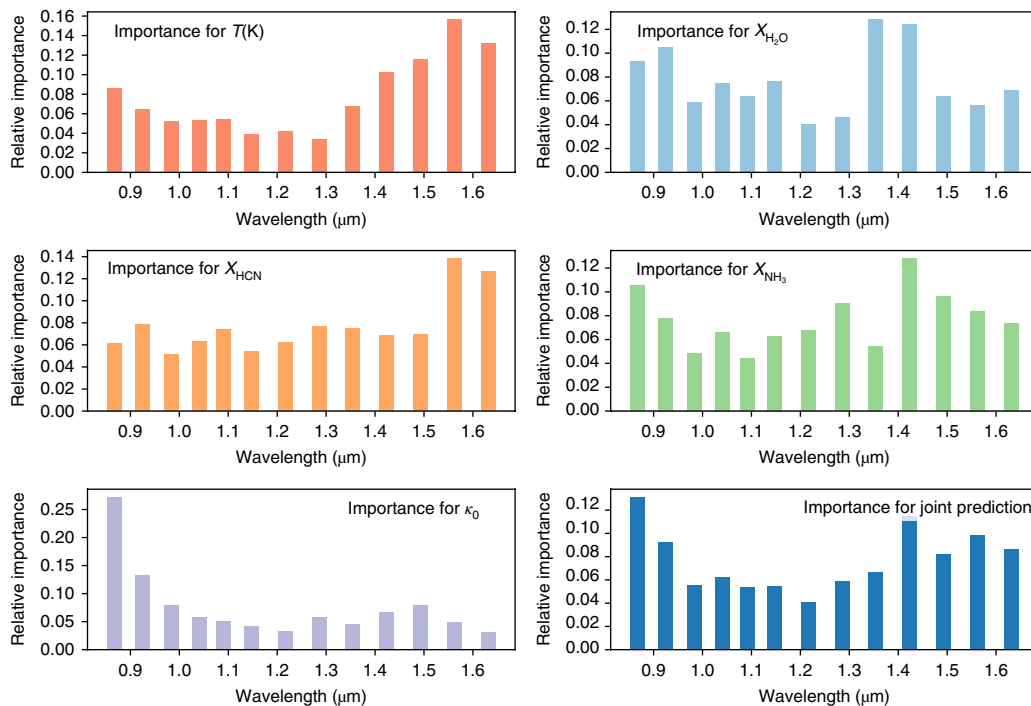
Fig. 4 | Feature importance plots associated with the machine-learning retrieval analysis of the WFC3 transmission spectrum of WASP-12b. Values along the vertical axis indicate the relative importance of a data point for retrieving the value of a given parameter. Within each panel, the bars add up to 1.

for retraining. In the study of stars and brown dwarfs, model grids spanning different ages, luminosities, radii, gravities and cloud configurations have traditionally been used to analyse ensembles of objects[27–29]. It is conceivable that model grids produced by different research groups could be used to perform retrievals even if the computer codes used to generate these grids are proprietary.

For the current study, we have shown that more sophisticated models are not necessary to analyse the WFC3 spectrum of WASP-12b. However, there is nothing to prevent more sophisticated models being considered. For example, using the non-isothermal model of ref. [11] in tandem with the non-grey cloud model of ref. [30] would add four more parameters to the retrieval. A long-standing shortcoming

of atmospheric retrieval, which is the non-self-consistency of the physics and chemistry in the models, may now be obviated using random forest retrieval.

Methods

For the physics input, we choose to use a previously validated analytical formula to convert the temperatures, molecular opacities and relative abundances of molecules into transit radii[11]. The simplicity of this forward model allows us to straightforwardly diagnose coding problems and understand trends in the posterior distributions. We use the simplest incarnation of this formula, which assumes that the atmosphere is isothermal, is isobaric and hosts a grey cloud. Using the nested sampling method[17,18,31,32], we performed regular retrievals, which indicate that non-isothermal behaviour and non-grey clouds are not necessary to explain the data, given their current level of quality and sophistication. We include the opacities of water (H_2O), hydrogen cyanide (HCN) and ammonia (NH_3), computed using the ExoMol spectroscopic line lists[33–37] as input and in the standard way, meaning that the opacities are products of the integrated line strength and line shape, and the line shapes are assumed to be truncated Voigt profiles[38,39].

For each model, we randomly pick values of the parameters over the following ranges: $T = 500$–$2,900$ K, $X_{H_2O} = 10^{-13}$–1, $X_{HCN} = 10^{-13}$–1, $X_{NH_3} = 10^{-13}$–1, $\kappa_0 = 10^{-13}$–10^2 cm^2 g^{-1}. The surface gravity of WASP-12b is taken to be 977 cm s^{-2} (ref. [40]). The spectroscopic database used to construct the NH_3 opacities does not exist for temperatures above 1,600 K (ref. [35]). For computational reasons, we set the NH_3 opacity to be zero and the volume mixing ratio to be small (10^{-13}) if the temperature exceeds this threshold. Fortunately, ammonia is expected to be a minor species at high temperatures, where the dominant nitrogen carrier is instead expected to be molecular nitrogen[41,42].

The features are the 13 values of the transit radius, across wavelength, associated with each transmission spectrum. Thirteen columns may be visualized, each with 80,000 values of the transit radius. A 13-dimensional space can then be visualized, where each dimension is marked by a set of numerical thresholds. Boundaries in this 13-dimensional space are drawn on the basis of splitting the training set to minimize the total variance. Each time a boundary is drawn, the regression tree is split. Once the reduction in the variance of the tree node is negligibly small (0.01 in our case), we stop splitting the training set. Tree pruning methods are not used. Each time the tree is split, only a random subset (4, which is about $\sqrt{13}$) of the 13 spectral bins is used. In other words, both the members of the 80,000 training set as well as the subset of spectral bins associated with each member are randomly drawn to decrease the correlations between predictions from different trees. For a pedagogical summary of the random forest method, see an earlier study[14]. The implementations of the random forest method and R^2 metric (the coefficient of determination) are from the open-source scikit-learn library in the Python programming language.

It has been shown previously that the random forest method is capable of handling systems with 1,000–10,000 features and tree depths of several tens to hundreds[21–25]. Our current problem has 13 features, and the regression trees have on average about 19,000 nodes and depths of 14.

To check the robustness of our results with respect to our implementation of the random forest method, we examine retrieval outcomes with different numbers of regression trees. As before, we train the random forest on 80,000 synthetic spectra and then use it to analyse 20,000 more synthetic spectra. Supplementary Fig. 1 shows that the outcomes of these mock retrievals converge when the number of trees used exceeds about 100. In the same figure, we also checked the retrieval outcomes with different levels of assumed noise floors. For each of the 13 data points in the synthetic WFC3 spectra, we assume a Gaussian uncertainty on the transit depths with full-widths at half-maximum of 10 ppm, 50 ppm and 100 ppm, which represent ideal, typical and easily attainable conditions, respectively. As expected, the variance associated with the true versus predicted values of the five parameters decreases (that is, the R^2 metric increases) when the assumed noise floor is lower. As a further check, we first train a random forest on a model grid with an assumed noise floor of 50 ppm and use it to analyse mock data with assumed noise floors of 10 ppm, 50 ppm and 100 ppm. The resulting R^2 values are 0.676, 0.651 and 0.586, respectively, for the joint predictions.

We also ran the same mock retrievals for a model grid where the atmosphere contains only H_2O versus one that contains HCN and NH_3 (without H_2O), as shown in Supplementary Fig. 2. In the former case, the retrievals return $X_{HCN} \approx 10^{-8}$ and $X_{NH_3} \approx 10^{-10}$ even when neither molecule is present in the mock spectra, which is consistent with our finding in Fig. 2 that volume mixing ratios below ~10^{-6} indicate non-detections of these molecules. In the latter case, we obtain $X_{H_2O} \approx 10^{-8}$, which is consistent with the non-detection of water.

As a final test and precursor for future studies, we generated mock JWST-like data in the Near-Infrared Spectrograph range of wavelengths (0.8 μm to 5.0 μm) at a resolution of 100 (not shown). Despite the increase in the number of features (data points) from 13 to 181, the time needed to train the random forest on 80,000 mock spectra increased by a factor of only 4 (without any attempt to parallelize the computation). The time needed for interpreting the additional 20,000 mock spectra (termed 'testing') is virtually the same in both cases. Furthermore, we note that both the training and testing steps are highly parallelizable.

To determine the spectral resolution used for our opacities, we ran retrievals with resolutions of 1 cm^{-1}, 2 cm^{-1}, 5 cm^{-1} and 10 cm^{-1}, assuming an isothermal atmosphere containing grey clouds and all three molecules. Retrieval practitioners typically use a spectral resolution of 1 cm^{-1} for their opacities[43–45], although it is not uncommon for workers to not state the spectral resolution used. For these four resolutions, the retrievals are shown in Supplementary Fig. 3. The corresponding retrieved parameter values are tabulated. On the basis of this resolution test, we adopt 5 cm^{-1} as our spectral resolution for the opacities.

We assume pressure broadening to be negligible. Since the inferred atmospheric temperature does not fall well below 1,000 K and the volume mixing ratios are typically much smaller than unity, this is not an unreasonable assumption[11]. Operationally, to implement this assumption, we assume a pressure of 1 mbar when computing the opacities. As is accepted practice[46], the lack of data on the physics of pressure broadening forces us to truncate the Voigt profile at some distance from line centre. We have made an ad hoc choice of 100 cm^{-1}, but since pressure broadening is assumed to be negligible, this has little to no effect on the outcome.

To check our assumption of a constant/grey cloud opacity, we ran another retrieval calculation with the non-grey cloud model of ref. [30]. The Bayes factor for the pair of models with grey versus non-grey clouds is 0.6 (with the former having a higher Bayesian evidence), which implies there is no evidence for the data favouring the non-grey over the grey cloud model[16]. In fact, we note that the model with non-grey clouds and only water has the same Bayesian evidence as the one with grey clouds and all three molecules. Similarly, the Bayes factor for a pair of models with isothermal versus non-isothermal atmospheres (both with only water) is 0.7, implying a lack of evidence for non-isothermal behaviour. The latter has lower Bayesian evidence and was computed using the non-isothermal analytical formula derived by ref. [11].

References

1. Banerji, M. et al. Galaxy Zoo: reproducing galaxy morphologies via machine learning. *Mon. Not. R. Astron. Soc.* **406**, 342–353 (2010).
2. Graff, P., Feroz, F., Hobson, M. P. & Lasenby, A. SKYNET: an efficient and robust neural network training tool for machine learning in astronomy. *Mon. Not. R. Astron. Soc.* **441**, 1741–1759 (2014).
3. Pearson, K. A., Palafox, L. & Griffith, C. A. Searching for exoplanets using artificial intelligence. *Mon. Not. R. Astron. Soc.* **474**, 478–491 (2018).
4. Madhusudhan, N. & Seager, S. A temperature and abundance retrieval method for exoplanet atmospheres. *Astrophys. J.* **707**, 24–39 (2009).
5. Benneke, B. & Seager, S. Atmospheric retrieval for super-Earths: uniquely constraining the atmospheric composition with transmission spectroscopy. *Astrophys. J.* **753**, 100 (2012).
6. Line, M. R. et al. Information content of exoplanetary transit spectra: an initial look. *Astrophys. J.* **749**, 93 (2012).
7. Rodgers, C. D. *Inverse Methods for Atmospheric Sounding: Theory and Practice* (World Scientific, Singapore, 2000).
8. Waldmann, I. P. Dreaming of atmospheres. *Astrophys. J.* **820**, 107 (2016).
9. Ho, T. K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 832–844 (1998).
10. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
11. Heng, K. & Kitzmann, D. The theory of transmission spectra revisited: a semi-analytical method for interpreting WFC3 data and an unresolved challenge. *Mon. Not. R. Astron. Soc.* **470**, 2972–2981 (2017).
12. Kreidberg, L. et al. A detection of water in the transmission spectrum of the hot Jupiter WASP-12b and implications for its atmospheric composition. *Astrophys. J.* **814**, 66 (2015).
13. Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. *Classification and Regression Trees* (Chapman & Hall/CRC, Boca Raton, 1984).
14. Kelleher, J. D., Mac Namee, B. & D'Arcy, A. *Fundamentals of Machine Learning for Predictive Data Analytics* (MIT Press, Cambridge, MA, 2015).
15. Criminisi, A., Shotton, J. & Konukoglu, E. *Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning* Technical Report TR-2011-114 (Microsoft Research, 2011).
16. Trotta, R. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.* **49**, 71–104 (2008).
17. Skilling, J. et al. Nested sampling for general Bayesian computation. *Bayesian Anal.* **1**, 833–859 (2006).
18. Feroz, F. & Hobson, M. P. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *Mon. Not. R. Astron. Soc.* **384**, 449–463 (2008).

19. Batalha, N. E. & Line, M. R. Information content analysis for selection of optimal JWST observing modes for transiting exoplanet atmospheres. *Astron. J.* **153**, 151 (2017).

20. Howe, A. R., Burrows, A. & Deming, D. An information-theoretic approach to optimize JWST observations and retrievals of transiting exoplanet atmospheres. *Astrophys. J.* **835**, 96 (2017).

21. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer, New York, 2001).

22. Sznitman, R., Becker, C., Fleuret, F. & Fua, P. Fast object detection with entropy-driven evaluation. In *Proc. 2013 IEEE Conference on Computer Vision and Pattern Recognition* 3270–3277 (IEEE, 2013).

23. Zikic, D., Glocker, B. & Criminisi, A. Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. *Med. Image Anal.* **18**, 1262–1273 (2014).

24. Rieke, N. et al. Surgical tool tracking and pose estimation in retinal microsurgery. In *Proc. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention* (eds Navab, N. et al.) 266–273 (Lecture Notes in Computer Science 9349, Springer, 2015).

25. Zhang, L., Varadarajan, J., Suganthan, P. N., Ahuja, N. & Moulin, P. Robust visual tracking using oblique random forests. *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 5825–5834 (IEEE, 2017).

26. Greene, T. P. et al. Characterizing transiting exoplanet atmospheres with JWST. *Astrophys. J.* **817**, 17 (2016).

27. Marley, M. S. et al. Atmospheric, evolutionary, and spectral models of the brown dwarf Gliese 229 B. *Science* **272**, 1919–1921 (1996).

28. Burrows, A. et al. A non-gray theory of extrasolar giant planets and brown dwarfs. *Astrophys. J.* **491**, 856–875 (1997).

29. Baraffe, I., Chabrier, G., Allard, F. & Hauschildt, P. H. Evolutionary models for low-mass stars and brown dwarfs: uncertainties and limits at very young ages. *Astron. Astrophys.* **382**, 563–572 (2002).

30. Kitzmann, D. & Heng, K. Optical properties of potential condensates in exoplanetary atmospheres. *Mon. Not. R. Astron. Soc.* **475**, 94–107 (2018).

31. Feroz, F., Hobson, M. P. & Bridges, M. MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Mon. Not. R. Astron. Soc.* **398**, 1601–1614 (2009).

32. Buchner, J. et al. X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *Astron. Astrophys.* **564**, A125 (2014).

33. Barber, R. J., Tennyson, J., Harris, G. J. & Tolchenov, R. N. A high-accuracy computed water line list. *Mon. Not. R. Astron. Soc.* **368**, 1087–1094 (2006).

34. Barber, R. J. et al. ExoMol line lists—III. An improved hot rotation-vibration line list for HCN and HNC. *Mon. Not. R. Astron. Soc.* **437**, 1828–1835 (2014).

35. Yurchenko, S. N., Barber, R. J. & Tennyson, J. A variationally computed line list for hot NH_3. *Mon. Not. R. Astron. Soc.* **413**, 1828–1834 (2011).

36. Yurchenko, S. N., Tennyson, J., Barber, R. J. & Thiel, W. Vibrational transition moments of CH_4 from first principles. *J. Mol. Spectrosc.* **291**, 69–76 (2013).

37. Yurchenko, S. N. & Tennyson, J. ExoMol line lists—IV. The rotation-vibration spectrum of methane up to 1500 K. *Mon. Not. R. Astron. Soc.* **440**, 1649–1661 (2014).

38. Rothman, L. S. et al. The HITRAN molecular spectroscopic database and HAWKS (HITRAN atmospheric workstation): 1996 edition. *J. Quant. Spectrosc. Radiat. Transf.* **60**, 665–710 (1998).

39. Grimm, S. L. & Heng, K. HELIOS-K: an ultrafast, open-source opacity calculator for radiative transfer. *Astrophys. J.* **808**, 182 (2015).

40. Hebb, L. et al. WASP-12b: the hottest transiting extrasolar planet yet discovered. *Astrophys. J.* **693**, 1920–1928 (2009).

41. Burrows, A. & Sharp, C. M. Chemical equilibrium abundances in brown dwarf and extrasolar giant planet atmospheres. *Astrophys. J.* **512**, 843–863 (1999).

42. Heng, K. & Tsai, S.-M. Analytical models of exoplanetary atmospheres. III. Gaseous C-H-O-N chemistry with nine molecules. *Astrophys. J.* **829**, 104 (2016).

43. Line, M. R. et al. A systematic retrieval analysis of secondary eclipse spectra. I. A comparison of atmospheric retrieval techniques. *Astrophys. J.* **775**, 137 (2013).

44. Waldmann, I. P. et al. Tau-REx I: a next generation retrieval code for exoplanetary atmospheres. *Astrophys. J.* **802**, 107 (2015).

45. Lavie, B. et al. HELIOS–RETRIEVAL: an open-source, nested sampling atmospheric retrieval code; application to the HR 8799 exoplanets and inferred constraints for planet formation. *Astron. J.* **154**, 91 (2017).

46. Sharp, C. M. & Burrows, A. Atomic and molecular opacities for brown dwarf and giant planet atmospheres. *Astrophys. J.* **168**, 140–166 (2007).

Acknowledgements

Author contributions

P.M.-N. led the development of computer codes used for this study, performed the machine-learning-related calculations, participated in the experimental design and made the majority of the figures. C.F. computed the grid of atmospheric models used as the training set, participated in the experimental design and performed the nested-sampling retrievals. R.S. co-led the scientific vision and experimental design and co-wrote the manuscript. K.H. co-led the scientific vision and experimental design and led the writing and typesetting of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41550-018-0504-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

CHAPTER 5

Interpreting High-resolution Spectroscopy of Exoplanets using Cross-correlations and Supervised Machine Learning

My methods are really methods of working and thinking; this is why they have crept in everywhere anonymously.

EMMY NOETHER

5.1 Summary

Following on from the initial work on Random Forest retrievals, I decided to turn the focus to high-resolution spectra. Ground-based observatories are able to provide exoplanet spectra with resolutions of $\sim$10,000-100,000, allowing one to resolve individual spectral lines. However, the noise level and quantity of points in the data pose a challenge for traditional retrievals.

In this paper, I developed a method of retrieval on high-resolution HARPS-N data using the Random Forest. I initially tested direct retrievals on the spec-

tra with the forest, but this proved unsuccessful due to the large error bars on each spectral point. Inspired by the work of Brogi and Line (2019), I created a mapping of the spectra to a lower dimensional space using the well-established cross-correlation function (CCF). These CCFs are able to beat down the noise in the spectra by averaging across a large number of points. I applied the Random Forest to the CCFs and was able to retrieve our test parameters successfully. I then used the forest to retrieve on the HARPS-N spectrum of KELT-9 b. My results showed that the abundance of ionised iron in this planet is far above the quantities that can be achieved in the atmospheric model, indicative of missing physics.

In addition, I discussed the degeneracy in metallicity at high values, arising from the increase in mean molecular weight of the atmosphere, which reduces the spectral features. I also showed that the likelihood-free aspect of the Random Forest can be an advantage, when compared to traditional retrievals, if the assumed error distribution is incorrect. Furthermore, I tested an alternative machine learning technique, known as a Bayesian neural network, and found it was unable to reproduce complex posteriors. Finally, I addressed the claim in Cobb et al. (2019) that the Random Forest can be overconfident but incorrect. This turned out to be an artifact of the model used in the training set, which leads to the same results in a nested-sampling retrieval.

This work is the first study able to perform atmospheric retrieval on wideband high-resolution exoplanet spectra. With the development of larger, more advanced ground-based observatories, and an increasing abundance of exoplanet spectra, this method will be essential for optimising our analysis of these objects.

5.2 Publication

This work was published in The Astronomical Journal in 2020.

Interpreting High-resolution Spectroscopy of Exoplanets using Cross-correlations and Supervised Machine Learning

Chloe Fisher[1] , H. Jens Hoeijmakers[1,2], Daniel Kitzmann[1] , Pablo Márquez-Neila[3], Simon L. Grimm[1] ,
Raphael Sznitman[3], and Kevin Heng[1]

[1] University of Bern, Center for Space and Habitability, Gesellschaftsstrasse 6, CH-3012, Bern, Switzerland; chloe.fisher@csh.unibe.ch,
jens.hoeijmakers@space.unibe.ch, kevin.heng@csh.unibe.ch

[2] Observatoire astronomique de l'Université de Genève, 51 Chemin des Maillettes, 1290 Versoix, Switzerland

[3] ARTORG Center for Biomedical Engineering, University of Bern, Bern, Switzerland

Abstract

We present a new method for performing atmospheric retrieval on ground-based, high-resolution data of exoplanets. Our method combines cross-correlation functions with a random forest, a supervised machine-learning technique, to overcome challenges associated with high-resolution data. A series of cross-correlation functions are concatenated to give a "CCF-sequence" for each model atmosphere, which reduces the dimensionality by a factor of ~100. The random forest, trained on our grid of $\sim65,000$ models, provides a likelihood-free method of retrieval. The precomputed grid spans 31 values of both temperature and metallicity, and incorporates a realistic noise model. We apply our method to HARPS-N observations of the ultra-hot Jupiter KELT-9b and obtain a metallicity consistent with solar ($\log M = -0.2 \pm 0.2$). Our retrieved transit chord temperature ($T = 6000^{+0}_{-200}$K) is unreliable as strong ion lines lie outside of the extent of the training set, which we interpret as being indicative of missing physics in our atmospheric model. We compare our method to traditional nested sampling, as well as other machine-learning techniques, such as Bayesian neural networks. We demonstrate that the likelihood-free aspect of the random forest makes it more robust than nested sampling to different error distributions, and that the Bayesian neural network we tested is unable to reproduce complex posteriors. We also address the claim in Cobb et al. 2019 that our random forest retrieval technique can be overconfident but incorrect. We show that this is an artifact of the training set, rather than of the machine-learning method, and that the posteriors agree with those obtained using nested sampling.

Unified Astronomy Thesaurus concepts: Exoplanet atmospheres (487)

1. Introduction

1.1. Observational Motivation I: The Rise of Ground-based High-resolution Spectra

The observational characterization of exoplanetary atmospheres via the measurement of transmission and emission spectra is occurring on two fronts: low-resolution, space-based spectroscopy (mainly with the *Hubble Space Telescope* and the *Spitzer Space Telescope*), and high-resolution spectroscopy using a wide variety of ground-based spectrographs (Table 1). Spectra measured from space have the advantage that the spectral continuum, which encodes information on chemistry and clouds/hazes, may be measured in an absolute sense. Ground-based spectra lose the spectral continuum—and effectively measure *relative* transit depths or fluxes—due to having to correct for the presence of the Earth's atmosphere, but offer the key advantage that individual spectral lines may be resolved with spectral resolution $\sim10^5$. A plausible approach is to combine the advantages each has to offer and jointly analyze space- and ground-based spectra (e.g., Brogi et al. 2017).

Following the pioneering work of Snellen et al. (2008, 2010); (see also Wiedemann et al. 2001; Brown et al. 2002; Deming et al. 2005), the use of high-resolution, ground-based spectroscopy to identify the presence of atoms and molecules has become routine (Redfield et al. 2008; Brogi et al. 2012; Birkby et al.

2013, 2017; Brogi et al. 2013, 2014, 2018; de Kok et al. 2013; Lockwood et al. 2014; Wyttenbach et al. 2015, 2017; Piskorz et al. 2016, 2017, 2018; Khalafinejad et al. 2017, 2018; Nugroho et al. 2017; Hoeijmakers et al. 2018, 2019; Cauley et al. 2019; Guilluy et al. 2019; Seidel et al. 2019). These identifications are essentially model independent, relying only on knowledge of the cross sections or opacities of these atoms and molecules as determined by quantum physics (e.g., Rothman et al. 1998; Heng 2017). Line transition databases contain the positions and relative strengths of individual lines, either from experimental measurement or derived from first principles, which are then cross-correlated against the lines detected in the high-resolution spectrum. By matching dozens to hundreds of lines using cross-correlation, robust identifications of atoms and molecules may be obtained (but see Hoeijmakers et al. 2015; Brogi & Line 2019 for examples of detections being dependent on the accuracy of the line database used to compute these opacities). In contrast, the claimed detections of molecules other than water in the Wide Field Camera 3 (WFC3) spectra of exoplanetary atmospheres remains model dependent and an active topic of debate (e.g., Fisher & Heng 2018), because at these resolutions (~10) only the shapes of the overall opacities, consisting of a large collection of lines averaged together, are measured.

Interpreting ground-based, high-resolution spectra using the cross-correlation technique has one major shortcoming: cross-correlation is mainly capable of answering the binary question of whether an atom or molecule is absent or present, either in emission or absorption. It does not yield the abundance of that atom or molecule, nor the atmospheric temperature and pressure of the environment in which it lies. It similarly does

1

Table 1
High-resolution Cross-dispersed Echelle (grating) Spectrographs with Wide Instantaneous Wavelength Coverage

Name	Telescope	Resolving power	Wavelength Range (nm)	Status	Reference(s)
HARPS	ESO 3.6 m	120,000	378–691	Active	Mayor et al. (2003)
HARPS-N	TNG	120,000	378–691	Active	Cosentino et al. (2012)
ESPRESSO	VLT	70,000–190,000	378–691	Active	Pepe et al. (2014)
CARMENES	CAHA 3.5	80,000–100,000	520–1710	Active	Quirrenbach et al. (2010)
GIANO	TNG	50,000	950–2450	Active	Origlia et al. (2014)
CRIRES+	VLT	50,000–100,000	Y, J, H, K, L, M bands	Under development	Follert et al. (2014)
UVES	VLT	40,000–110,000	300–1100	Active	Dekker et al. (2000)
NIRSPEC	Keck	25,000	960–5500	Active	McLean et al. (1998)
PEPSI	LBT	43,000–270,000	383–912	Active	Strassmeier et al. (2015)
HDS	Subaru	90,000–165,000	298–1016	Active	Noguchi et al. (2002)
EXPRES	DCT	150,000	380–844	Active	Fischer et al. (in prep)
HIRES	ELT	100,000	397–2500	Under development	Zerbi et al. (2014)
NIRPS	ESO 3.6 m	80,000	974–1809	Under development	Wildi et al. (2017)
SPIRou	CFHT	70,000	980–2440	Active	Donati et al. (2018)
iShell	IRTF	75,000	J, H, K, L, M bands	Active	Rayner et al. (2016)
IGRINS	HJS	40,000	1450–2450	Active	Park et al. (2014)

not yield cloud or haze properties of the atmosphere. The first study to decisively address this shortcoming was Brogi & Line (2019), who re-analyzed CRIRES observations and derived an analytical expression that maps the cross-correlation function to the likelihood function. The ability to compute the likelihood function implies that Bayes's Theorem may subsequently be invoked to compute posterior distributions of chemical abundances, temperature, etc.

CRIRES was an infrared echelle spectrograph mounted on UT1 of ESO's VLT (Kaeufl et al. 2004). Although the spectrograph achieved high spectral resolution of ~100,000, the instantaneous wavelength coverage was small because the spectrograph was not cross-dispersed. Consequently, the spectra analyzed by Brogi & Line (2019) contain only 4096 data points (1.9626–2.0045 μm, 2.2875–2.3454 μm in two different modes). As every model being computed in the atmospheric retrieval needs to be cross-correlated against the spectrum, it becomes computationally prohibitive to scale this method up to spectra of cross-dispersed echelle spectrographs that contain ~10^5–10^6 data points, because this increases the computational time by a factor ~10^2–10^3. However, elucidating such a scalable method is crucial in the era of high-resolution spectrographs with wide *instantaneous* wavelength coverage, an overview of which we list in Table 1. Gibson et al. (2020) also recently performed retrieval on high-resolution data from the blue arm of UVES with ~10^3 data points using a Markov Chain Monte Carlo (MCMC) method.

A novel method to analyze ground-based, high-resolution spectra with ~10^5–10^6 data points is therefore needed that will allow the computational effort to be reduced at the order-of-magnitude level *and* allow for the computation of posterior distributions of parameters.

1.2. Observational Motivation II: Failure of Direct Retrievals on Noisy Spectra

Another major limitation of ground-based, high-resolution spectra is the observational uncertainty. The level of noise on each individual spectral data point is typically much greater than the signal itself, which causes the direct retrieval to fail (see Section 3.1). While each individual spectral point contains little information, the entire spectrum does encode valuable information on the atmospheric abundances and properties. Any successful interpretation method needs to leverage the

information content of the entire spectrum against the high level of noise present.

This is the rationale behind the cross-correlation technique, which has been adopted by many workers (e.g., Snellen et al. 2010; Brogi et al. 2012; Birkby et al. 2013; de Kok et al. 2013; Lockwood et al. 2014; Wyttenbach et al. 2015; Piskorz et al. 2016; Nugroho et al. 2017; Hoeijmakers et al. 2018; Guilluy et al. 2019; Seidel et al. 2019), including Brogi & Line (2019).

In the current study, we will incorporate the cross-correlation technique into a novel method for performing retrievals on noisy, high-resolution spectra, but in a way that is distinct from Brogi & Line (2019).

1.3. Theoretical Motivation I: Likelihood-free Inference Methods using Machine Learning

In the published exoplanet literature, atmospheric retrievals typically assume the likelihood function to be a Gaussian when implementing the Markov Chain Monte Carlo (MCMC) or nested-sampling routines (e.g., Benneke & Seager 2012; Line et al. 2013; Waldmann et al. 2015; Lavie et al. 2017; MacDonald & Madhusudhan 2017; Fisher & Heng 2018; Brogi & Line 2019),

$$\ln \mathcal{L} = -\frac{1}{2}\sum_{i}^{n}\left(\frac{R_i - R_{i,\mathrm{obs}}}{\sigma_i}\right)^2 - \frac{\ln(2\pi\sigma_i^2)}{2}, \quad (1)$$

where the transmission spectrum has n measurements of transit radii ($R_{i,\mathrm{obs}}$) that are compared to the theoretical values of the transit radii (R_i) computed using a model. The standard deviation of the uncertainty on each data point, assumed to follow a Gaussian distribution, is σ_i. It is further assumed that the uncertainties are uncorrelated with one another.

One of the motivations of the current study is to provide an alternative inference approach that is likelihood free, meaning that one does not have to explicitly assume the functional form of the likelihood function. In practice, these likelihood-free inference approaches belong to the class of Approximate Bayesian Computation (ABC) methods (Sisson et al. 2019). Specifically, we use the supervised machine-learning method of the random forest (Ho 1998; Breiman 2001), which was previously adapted by Márquez-Neila et al. (2018) to interpret low-resolution *Hubble*-WFC3 transmission spectra. The
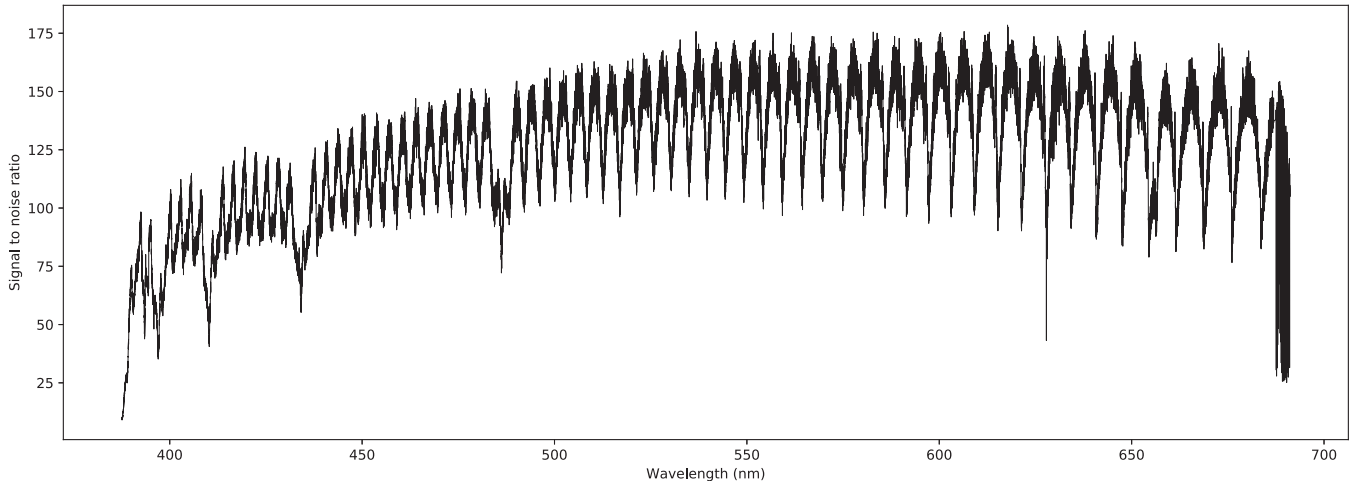
Figure 1. The signal-to-noise level of the spectrum of the host star KELT-9 achieved in a 600 s exposure obtained with the HARPS-N instrument. The signal-to-noise ratio is dominated by the photon (shot) noise, which decreases toward shorter wavelengths due to a reduced efficiency of the instrument, transmission of the Earth's atmosphere, and lower intrinsic luminosity of the star. The significant narrowband variation is due to the efficiency of the spectrograph falling off at the edges of spectral orders, as well as absorption lines in the star and the Earth's atmosphere.

method relies on using a grid of precomputed atmospheric models combined with an arbitrary noise model as a training set for the random forest. The uncertainties on each data point in the measured spectrum are incorporated into the noise-free model grid to generate a training set of noisy models. This approach is not unlike that of standard retrieval techniques, which typically compute a grid of atmospheric models on the fly.

The random forest consists of a collection of regression trees. Each regression tree is trained on a subset of the grid of atmospheric models. By identifying regions of the multi-dimensional parameter space that predict similar transmission spectra, each regression tree quantifies the "distance" between the model and measured transmission spectra. This plays the role of the Euclidean distance $(R_i - R_{i,\mathrm{obs}})$ in the Gaussian likelihood function, except that the likelihood is implicitly learned from the training set of noisy models. (See Section 2.6.1 for more information about the random forest.)

Other advantages offered by the random forest retrieval method include the ability to run large suites of mock retrievals to both validate the model grid used and quantify its sensitivity to the parameters, as well as information content analysis to quantify the relative importance of each data point in the spectrum toward determining the value of each parameter (Márquez-Neila et al. 2018).

1.4. Theoretical Motivation II: Feature Engineering

Feature engineering is the process by which the training set used in a machine-learning method is optimized, e.g., a reduction in the dimensionality of the problem. Deep learning methods perform feature engineering in an automated way, but they are significantly more expensive to implement than the random forest. One of the novel aspects of the current study is the use of feature engineering to efficiently interpret noisy, high-resolution spectra. Instead of using the spectra themselves as the training set, we demonstrate that it is sufficient to use a set of cross-correlation functions (CCFs) that sparsely sample the parameter space. The resulting "cross-correlation sequence" serves as the training set for the random forest, resulting in a reduction in the size of the training set by a factor of ∼100.

This feature engineering step allows the random forest retrieval method to be scaled up to interpret high-resolution spectra with ∼10^5–10^6 data points in a computationally feasible way.

1.5. Layout of Study

In Section 2, we describe our methodology, including the computation of the model grid of transmission spectra (radiative transfer, opacities, and chemistry), the implementation of the random forest method, etc. In Section 3, we show our results from testing the method, and also the retrieval on HARPS-N observations of KELT-9b. In Section 4, we discuss the results and compare our method to nested sampling and other machine-learning techniques. In Section 5, we summarize our conclusions.

2. Methods

2.1. KELT-9b

As a proof of concept and in order to test the method, we have focused the retrieval on the ultra-hot Jupiter, KELT-9b. The brightness of the star combined with the extremely high temperatures allow for a higher signal-to-noise ratio (SNR) than for other exoplanets (see Figure 1), making it a good test subject for a retrieval on ground-based data. Furthermore, this object has been previously studied with high-resolution data in Hoeijmakers et al. (2018, 2019). Kitzmann et al. (2018) demonstrated that chemical equilibrium is a reasonable assumption, significantly reducing the number of parameters required in the atmospheric model, and that it is cloud-free with a continuum dominated by H^- (Arcangeli et al. 2018). However, Hoeijmakers et al. (2019) suggested that there is most likely missing physics in this model, due to the discrepancy between the expected cross-correlation function for Fe^+ and the one obtained from the data. We will discuss this further in Section 3.4.

2.2. Model Grid

To construct the grid of models of KELT-9b, we adopt the system parameters reported by Gaudi et al. (2017) and Hoeijmakers et al. (2019). We generate the models using an

observation simulator, Helios-o (Bower et al. 2019), which follows the method described in Gaidos et al. (2017). This algorithm has been validated in Heng & Kitzmann (2017), where it was compared against the models from Fortney et al. (2010), Deming et al. (2013), and Line et al. (2013).

The model atmosphere is one-dimensional, plane-parallel, isothermal, in hydrostatic equilibrium, and in chemical equilibrium. It has 199 layers with 200 pressure levels ranging from 10^{-15}–2 bar. Each one-dimensional model atmosphere may be visualized as an atmospheric column. Ray tracing is performed through a collection of these atmospheric columns to construct the transit chord at each wavelength, taking into account the variation of gravity as different pressure levels are probed. The variation of the effective transit radius with wavelength due to the chemical composition of the atmosphere is the transmission spectrum (Brown 2001).

The volume mixing ratios (relative abundances by number) of atoms, ions, and molecules are computed using the FastChem chemical-equilibrium code, which considers gas-phase chemistry for more than 550 molecular species with elements more abundant than germanium (Stock et al. 2018). Additionally, we add most of the first and doubly ionized ions as well as anions for atoms lighter than neptunium (Hoeij-makers et al. 2019). Our volume mixing ratios computed using FastChem are pressure dependent, because of our nonisobaric treatment of the transit chord. The opacities are computed using the open-source HELIOS-K opacity calculator (Grimm & Heng 2015). The inputs for the Fe, Fe^+, Ti, and Ti^+ opacities are sourced from the Kurucz database[4] (Kurucz 2017). The hydrogen anion (H^-) cross section is taken from John (1988). For completeness, collision-induced absorption associated with H–He, H_2–H_2, and H_2–He collisions are included (Richard et al. 2012). Pressure broadening is neglected as the spectral continuum in ultra-hot Jupiters is dominated by absorption associated with the hydrogen anion (H^-), which masks the line wings. The line shape is assumed to be a Voigt profile. The natural line width and thermal broadening are included (Kurucz 2017). Opacities are sampled uniformly across wavenumber with a spectral resolution of 0.01 cm^{-1}, and the transmission spectra are calculated at a resolution of 0.03 cm^{-1}.

The assumption of chemical equilibrium allows us to greatly simplify the theoretical analysis because the abundances of atoms and ions are completely specified by the temperature, pressure, and elemental abundances. By assuming the ratios of elemental abundances follow those of the Sun, we reduce the chemical parameters down to a single number known as the metallicity. Therefore, we have just two parameters in our model—temperature and metallicity. The temperature range of the grid spans from 3000 to 6000 K, in steps of 100 K, and the metallicity ranges from 0.1 to 100 times solar (-1 to 2 for the logarithm of the metallicity, $\log M$, in steps of 0.1). This results in 31 values for each parameter, and thus 961 models in the grid in total.

2.3. Modeling HARPS-N Observations

We use existing observations of KELT-9b produced by the HARPS-N spectrograph (Hoeijmakers et al. 2018) to convert the resulting model grid to models of the observed transmission spectrum. First, the transmission spectrum is convolved with a Gaussian with a full-width-at-half-maximum of 2.7 km s^{-1}

---

[4] http://kurucz.harvard.edu/

(equivalent to the resolving power of the HARPS-N spectrograph), as well as a rotation-broadening profile that matches the rotation period of KELT-9b. It is subsequently interpolated onto the wavelength grid of the stitched, resampled pipeline-reduced (s1d) observations from HARPS-N. The continuum of the transmission spectrum is removed using a high-pass filter, in the same way as the observations with the HARPS-N spectrograph are filtered to remove broadband spectral variations that are due to the instrument and variable observing conditions (Hoeijmakers et al. 2018).

It would be possible to use this retrieval method for other instruments, such as those listed in Table 1, however these would require different training sets to account for other observational effects. The noise model (see Section 2.5) would also need to be adjusted for different instruments.

2.4. CCF-sequences

We use the cross-correlation operator defined as

$$C(v) = \frac{\sum_i F_i \mathcal{T}_i(v)}{\sum_i \mathcal{T}_i(v)}, \tag{2}$$

where F is the transmission spectrum, $\mathcal{T}$ is the cross-correlation template interpolated onto the same wavelength grid as the spectrum, v is the velocity, and the summation takes place over the spectral data points. The denominator is a normalization factor, and thus the fluxes of the templates do not need to be rescaled when performing the cross-correlation.

Four subsets of cross-correlation templates, consisting of the spectral lines of neutral iron (Fe), singly ionized iron (Fe^+), neutral titanium (Ti), and singly ionized titanium (Ti^+), are created. Within each subset, there are 16 templates consisting of 4 values of temperature (3000, 4000, 5000, and 6000 K) and 4 values of metallicity (0.1, 1, 10, and 100×solar). In total, there are 64 cross-correlation templates. These templates are generated in the same way as the models (Section 2.2) with all but the relevant species' opacities removed from the final model, leaving only the required species' spectral lines. Broadening is not included as we are not aiming to retrieve dynamic properties. (See Section 4.2 for tests involving velocity parameters.)

Each synthetic transmission spectrum in the model grid is cross-correlated with each of the 64 templates to create a set of 64 CCFs. Additionally, each template is shifted in velocity space from -20 to 20 km s^{-1} in steps of 1 km s^{-1}, resulting in 40 CCF values per template. These 64 CCFs are concatenated together to give a single sequence containing 2560 points, which we term a "CCF-sequence" (Figure 2). Each of the 64 templates probes different components of the information contained in the spectral lines. In this way, the resulting CCF-sequence encodes the physical properties of the atmosphere over multiple axes. This feature engineering step has essentially reduced the dimensions of the model spectra by a factor of ~100.

2.5. Noise Model

Because KELT-9 is a bright star, the noise is dominated by photon noise, and the SNR mainly varies due to the wavelength-dependent efficiency of the instrument, the stellar spectrum, and Earth's atmospheric transmission function (see Figure 1). The noise per spectral pixel is empirically measured from the time series of observations used by Hoeijmakers et al. (2018).
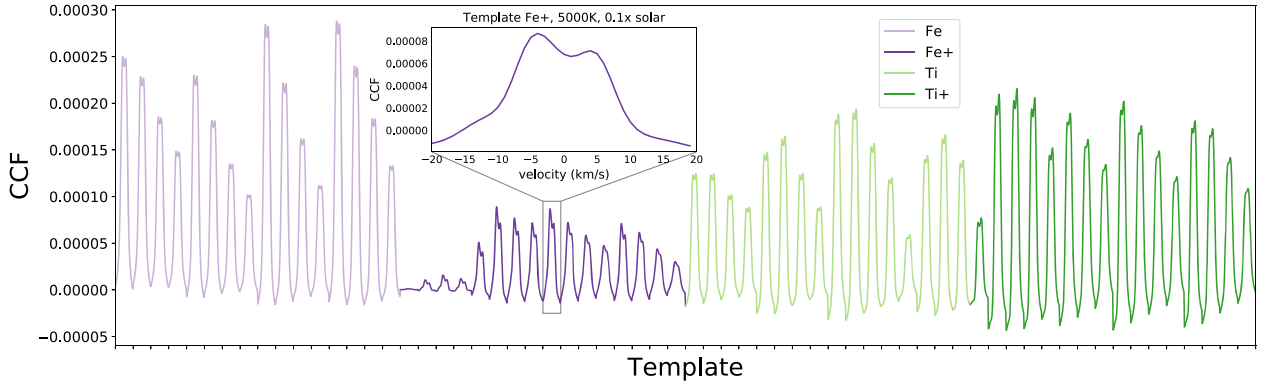
Figure 2. Example of a CCF-sequence constructed by cross-correlating 64 templates with a model transmission spectrum with $T = 3500$ K and $\log M = 0.8$. Each CCF has 40 points across velocity for a total of 2560 points for the entire CCF-sequence. The insert magnifies one of the CCFs (Fe^+, $T = 5000$ K, and $\log M = 0.1$) for illustration.

For each spectral pixel a value may be drawn randomly from an assumed Gaussian distribution, creating a model of the noise of the entire spectrum that can be propagated through the cross-correlation function.

We assume each point in the spectrum F has a Gaussian error bar with standard deviation σ_{F_i}. The noise model for the CCF then becomes a linear combination of Gaussians, therefore also a Gaussian, with a variance of

$$\sigma_C^2 = \frac{\sum_i \sigma_{F_i}^2 \mathcal{T}_i(v)^2}{\sum_i \mathcal{T}_i(v)^2}. \tag{3}$$

We can then add the noise to the model grid of CCF-sequences. Since we require many instances of noise for the random forest, and the cross-correlation is computationally quite expensive, this provides a great advantage over applying the CCF to the noisy spectra.

2.6. Random Forest

2.6.1. Theory

The random forest consists of a collection of regression trees—decision trees for interpreting continuous data. Each regression tree is trained on a subset of the grid of atmospheric models. During training, a tree is constructed by locating divisions in each wavelength dimension that sort the training spectra into groups with similar parameter values, known as leaves. Each leaf then has an assigned set of parameter values given by the training spectra in its group. When predicting on a real data set, the spectrum is passed down each tree until it lands in a leaf, and the predicted parameter values are given by the corresponding set. The sets for every tree in the forest are then combined to give a distribution for each parameter.

The random forest falls into a class of inference methods known as "ABC" (Sisson et al. 2019). ABC methods were invented to treat problems where it was either infeasible or impossible to explicitly specify the functional form of the likelihood (e.g., in the study of human populations). Instead of seeking the maximum likelihood in a multidimensional parameter space, ABC methods seek to minimize some abstract distance (with the Euclidean distance being one specific example) between a set of simulated models and data to below some stated tolerance (Chapter 1.3 of Sisson et al. 2019). If the tolerance is formally zero, then ABC methods become exact Bayesian methods, which have been shown to produce accurate posteriors (Chapter 1.6 of Sisson et al. 2019). In practice,
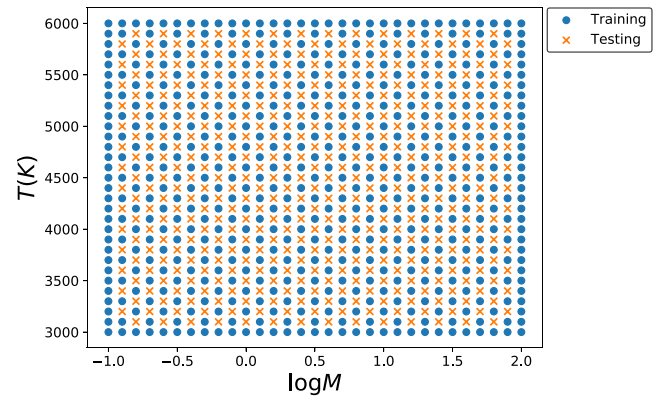
Figure 3. Separation of the 961 members of the model grid into training and testing sets for the random forest. The edges of this parameter space are intentionally included in the training set as the forest is unable to extrapolate.

nonzero tolerances generally imply that the computed posterior distributions are approximate (hence the "A" in "ABC"), where the degree of accuracy depends on the tolerance specified (Chapter 1.5 of Sisson et al. 2019). ABC methods often employ "summary statistics" as a dimensionality reduction step (Chapter 1.7 of Sisson et al. 2019). In the current study, the use of the CCF-sequence qualifies as a use of summary statistics.

2.6.2. Setup

Starting from our grid of CCF-sequences, we divide the parameter space into training and testing sets, as shown in Figure 3. This is to ensure the two sets are sufficiently distinct such that we can accurately test the performance of the forest. Next, we sample each point of the CCF-sequence within its respective uncertainty to generate 120 noisy instances of each CCF-sequence. We do this by drawing from Gaussian distributions with variance defined by Equation (3). The entire set therefore amounts to 115,320 noisy CCF-sequences, with 64,920 in training and 50,400 in testing.

Our random forests consists of 1000 trees. Tree splitting is performed using a threshold variance of 0.01. Each time a tree is split, a random subset of 50 (approximately the square root) of the 2560 sequence points is used. Tree pruning methods are not used (see Breiman et al. 1984; Hastie et al. 2001 for clarification of the terminology). For the predictions, the data is passed down through each tree until it reaches an end point,
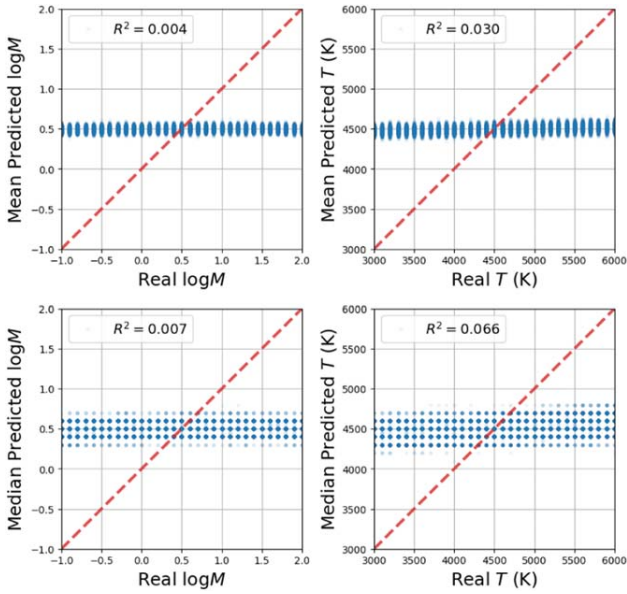
Figure 4. Predicted vs. real values of the logarithm of metallicity ($\log M$) and temperature (T) for the random forest trained using a section of the high-resolution spectrum containing 10^4 points from 400 to 410 nm. The top and bottom sets of plots correspond to the mean and median predictions, respectively. The coefficient of determination (R^2) varies from -1 to 1, where values near unity indicate strong anticorrelations or correlations between the real and predicted values of a given parameter, based on the variance of outcomes. See Figure 5 for a mock retrieval.

Figure 5. A mock retrieval using a section of the high-resolution spectrum containing 10^4 points from 400 to 410 nm, from the test set shown in Figure 4. The mock spectrum has solar metallicity and a temperature of 4100 K. In the top left and bottom right panels, the solid posteriors show the results of the retrieval using the random forest (RF), and the empty line posteriors show the results from nested sampling (NS). The purple, dashed lines show the true values. The top right panel shows the data points (lilac) with the error region (gray), along with the model (dark purple) corresponding to the medians from the $\log M$ and T posteriors.

known as a leaf. The set of all training parameters that lie in this leaf are then given as the prediction for that tree. We call this the "full-leaf" prediction. These training parameters come from the bootstrapped training data set—built using random sampling with replacement from the original training data set—that was used to train each tree. The final posterior is constructed by combining these predictions for all of the 1000 trees. This full-leaf prediction is an improvement on the previous method in Márquez-Neila et al. (2018), i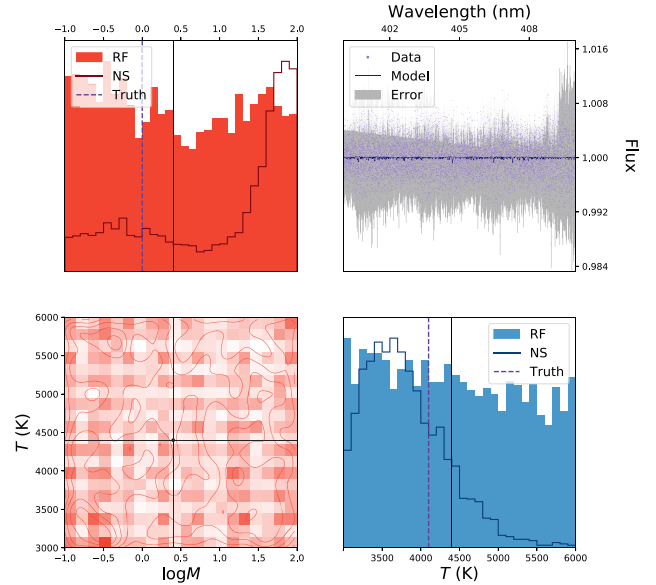n which only the mean parameter values corresponding to the predicted leaf were used, as it gives a more accurate approximation of the posterior. The implementation of the random forest method and R^2 metric are adopted from the open-source `scikit.learn` library (Pedregosa et al. 2011) in the `Python` programming language.

3. Results

3.1. Failure of Direct Retrieval

Initially we attempted to perform the random forest retrieval directly on the transmission spectra, set up in the same way as described in Section 2.6.2 but with the model spectra instead of the CCF-sequences. Since the random forest method has been demonstrated to work for a dimensionality of at most $\sim 10^4$ (Hastie et al. 2001; Sznitman et al. 2013; Zikic et al. 2014; Rieke et al. 2015; Zhang et al. 2017), we consider only a section of 10^4 wavelength points from 400 to 410 nm in each synthetic spectrum. Other sampling strategies (e.g., selecting line peaks only) produce similar outcomes[5] (not shown). Figure 4 shows the results of testing this forest, using both the mean (top panels) and median (bottom panels) predictions. The

coefficient of determination, R^2, which measures the degree of agreement between the real versus predicted parameter values, is essentially zero for temperature and metallicity for both mean and median predictions, implying that the random forest has no predictive power when applied to the synthetic spectra themselves. Figure 5 includes an example of the posterior distributions of temperature and metallicity for a mock retrieval, which are unconstrained and consistent with their prior distributions. In addition, we tested a traditional retrieval algorithm using nested sampling (Skilling 2006; Feroz & Hobson 2008; Feroz et al. 2009, 2019) with the open-source `PyMultinest` package (Buchner et al. 2014). Due to the high number of spectral points and complex forward model, we are unable to compute models on the fly as in a regular nested-sampling retrieval (see Section 4.1). Instead, we take the same grid of models as the forest, but without the added noise, and interpolate on it to produce forward models. Figure 5 also shows the results from the nested-sampling mock retrieval. These posteriors span essentially the whole prior, with peaks offset from the correct values.

In summary, ground-based high-resolution spectra of exoplanets reside in a qualitatively different regime than the same measurements of stars or space-based low- to medium-resolution spectra of exoplanets. Individual data points hold little information as they are overwhelmed by noise, but the entire spectrum does encode useful information. This motivates our use of the CCFs, which effectively select the most informative lines in the spectrum.

3.2. Random Forest Mock Retrievals

Figure 6 shows the results of testing the random forest trained on the CCF-sequences. The predictive power of the

---

[5] While selecting line peaks is conceptually similar to a cross-correlation, by not averaging the spectral lines the noise remains high and hence the retrieval still fails.
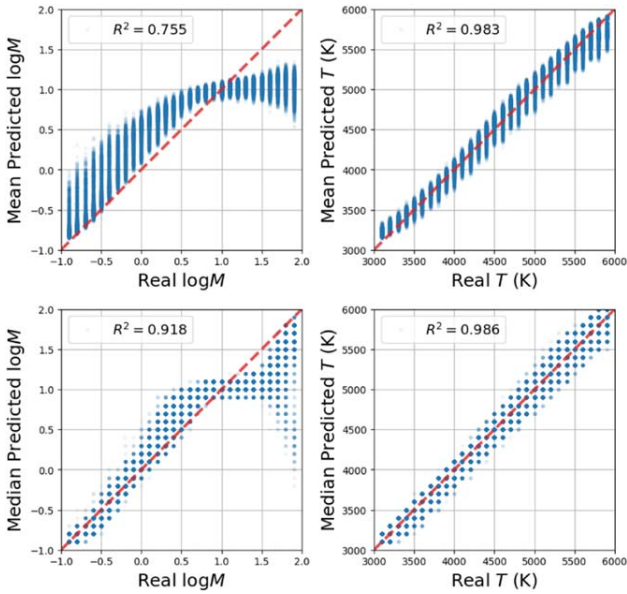
Figure 6. Predicted vs. real values of the logarithm of metallicity ($\log M$) and temperature (T) for the random forest trained on the CCF-sequences. The top and bottom panels show the results using the mean and median predictions, respectively. The coefficient of determination (R^2) varies from -1 to 1, where values near unity indicate strong anticorrelations or correlations between the real and predicted values of a given parameter, based on the variance of the outcomes. See Figure 7 for a mock retrieval.

Figure 7. A mock retrieval performed on a model with solar metallicity and $T = 4100$ K, using the random forest trained on the CCF-sequences (see Figure 6). The black lines show the median values. The purple, dashed lines show the true values. The top right panel shows the data points (lilac) with the error region (gray), along with the model (dark purple) corresponding to the medians from the $\log M$ and T posteriors.

random forest has increased significantly. The difference in the predictability of the two parameters, metallicity and temperature, follows our intuition. The strength of spectral features are proportional to the logarithm of the opacity multiplied by the abundance of an atom. Because opacities have an exponential dependence on temperature (Rothman et al. 1998; Heng 2017), the line strengths are highly sensitive to temperature, and the ability of the random forest to predict temperature is strong. The ability to predict metallicity is somewhat weaker, because the metallicity linearly controls the atomic abundances, the logarithm of which determines the line depths (e.g., Heng & Kitzmann 2017). At high metallicities, the predictive power of the random forest tapers off, because the pressure scale height of the atmosphere decreases and the size of spectral features starts to decrease (see Section 3.3). The top and bottom panels of Figure 6 correspond to the mean and median predictions of the trees, respectively. Traditionally, random forests produce mean predictions, but given the focus of atmospheric retrieval on posteriors and confidence intervals, we are more interested in the medians, which are more robust against asymmetric posteriors. The increase in R^2 scores when using the median comes particularly from these more complex posteriors. Figure 7 also shows an example of the posterior distributions obtained from the hybrid CCF retrieval, which recovers the injected values of temperature and metallicity accurately.

A useful, natural outcome of the random forest is the information content analysis known as the "feature importance." This determines which data points hold the most importance for retrieving each parameter. Figure 8 shows the feature importance when predicting metallicity and temperature. As suggested by the bottom panel of Figure 8, the ion species control the temperature prediction. Rising temperatures cause the neutral species to collisionally ionize, initially increasing the abundances of Fe^+ and Ti^+ by orders of
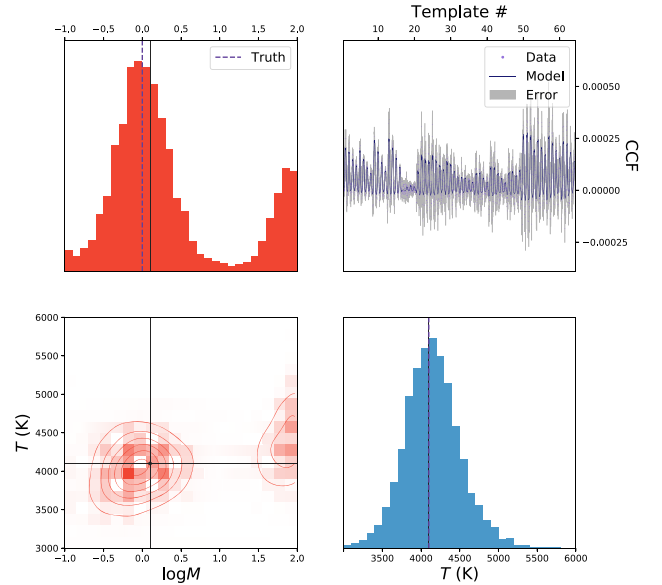
magnitude while the corresponding decrease in neutral abundance is relatively small.

As the metallicity increases, the depths of all metal absorption lines will tend to increase. However, in Figure 8 there appears to be a greater feature importance for the neutrals when predicting metallicity. A possible explanation for this is that as metallicity increases, the atmosphere will be more laden by free electrons from easily ionized species. Following the Saha equation (Saha 1920), this will lead to a decrease in the ionization fraction, partially negating the enhancement to the ion mixing ratios that stems from the higher metal abundance. Therefore, the neutral species are expected to be more sensitive to metallicity.

3.3. Metallicity Degeneracy

From our tests on the random forest in Figure 6, we can see that some of the high metallicity spectra yield much lower metallicity predictions. This is demonstrated further in Figure 9, which shows a retrieval on one of these high metallicity spectra. The double-peaked posterior leads to a mean prediction that is heavily offset from the true value. This multimodal structure is due to a degeneracy between line depth and metal abundance for high metallicity values. As discussed in Section 3.2, as the metallicity increases to very high levels, the atmosphere is no longer hydrogen dominated, causing the mean molecular weight to increase significantly. This in turn decreases the scale height and absorption line depths, reminiscent of lower metallicity values. We tested all the spectra with the highest metallicity value in the testing set ($\log M = 1.9$), and plotted the median predictions in Figure 10. This plot shows that the degeneracy is stronger at lower temperatures. This follows our physical intuition because at lower temperatures the pressure scale height is smaller, thus compressing the features and reducing the spectrum's

Figure 8. Feature importance plots describing the relative importance of each CCF in the sequence toward constraining metallicity and temperature.
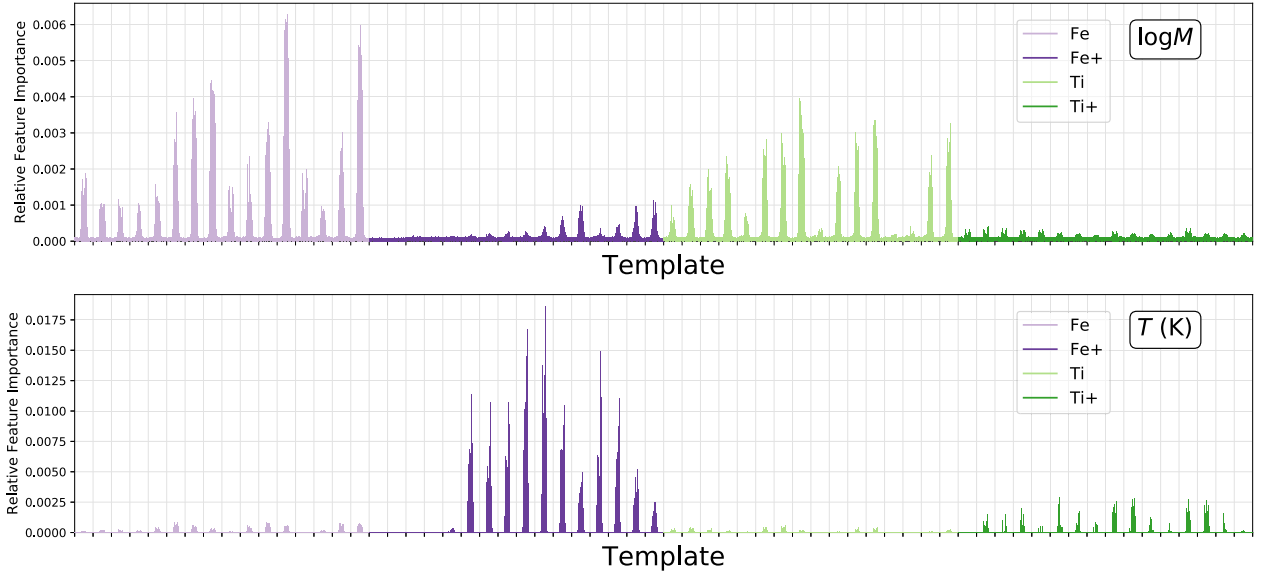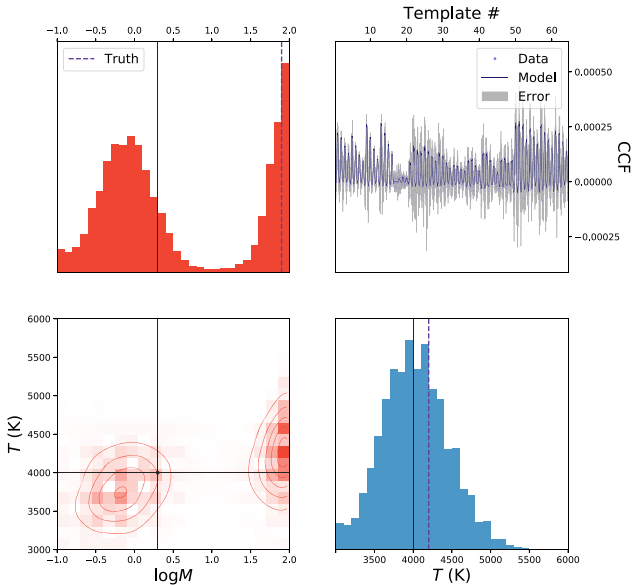

Figure 9. A mock retrieval performed on a model with $\log M = 1.9$ and $T = 4200$ K, using the random forest trained on the CCF-sequences (see Figure 6). The black lines show the median values. The purple, dashed lines show the true values. The top right panel shows the data points (lilac) with the error region (gray), along with the model (dark purple) corresponding to the medians from the $\log M$ and T posteriors.

sensitivity to metallicity. This makes these spectra harder to distinguish from one another for a given SNR.

This degeneracy is also visible in Figure 11, which shows noise-free spectra with $T = 3000$ K and varying metallicities, and a cross-correlation with those spectra. As the metallicity increases, the troughs in the left-hand plot deepen up to a point, after which they become shallower again. Similarly, the height of the CCFs in the right-hand plot increase with metallicity until $\log M \gtrsim 1.0$, after which the peaks decrease again. While the shape of the high and low metallicity noise-free spectra do differ slightly from each other, these variations are within the error bars of the data, making the noisy spectra indistinguishable.
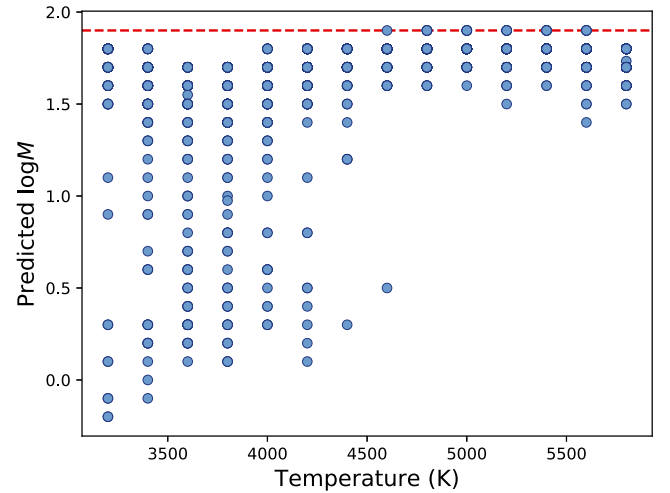
Figure 10. Median predictions for metallicity vs. the true temperature value for the test spectra with $\log M = 1.9$, from Figure 6. The red, dashed line shows the true metallicity value, 1.9.

3.4. KELT-9b Retrieval

Finally, we performed the hybrid CCF retrieval on the real HARPS-N data set for the ultra-hot Jupiter KELT-9b. Figure 12 shows our results for several different retrievals. As described in Hoeijmakers et al. (2019), the ionized iron lines in the spectrum of KELT-9b appear to be much larger than predicted, possibly resulting from an outflowing envelope not present in the model. This leads to a CCF-sequence for the real KELT-9b data that features significantly higher peaks in the Fe^+ CCFs when compared to the training set, as shown in Figure 13. With the intent of comparing the effects of the different species, we performed three independent retrievals on the KELT-9b data set—one containing the full CCF-sequence, as described in Section 2.4, a second containing only the neutral elements, and a third containing only the ions. Each retrieval uses a separate random forest trained on the corresponding sections of the model CCF-sequences. The three retrievals are compared in Figure 12, where the empty
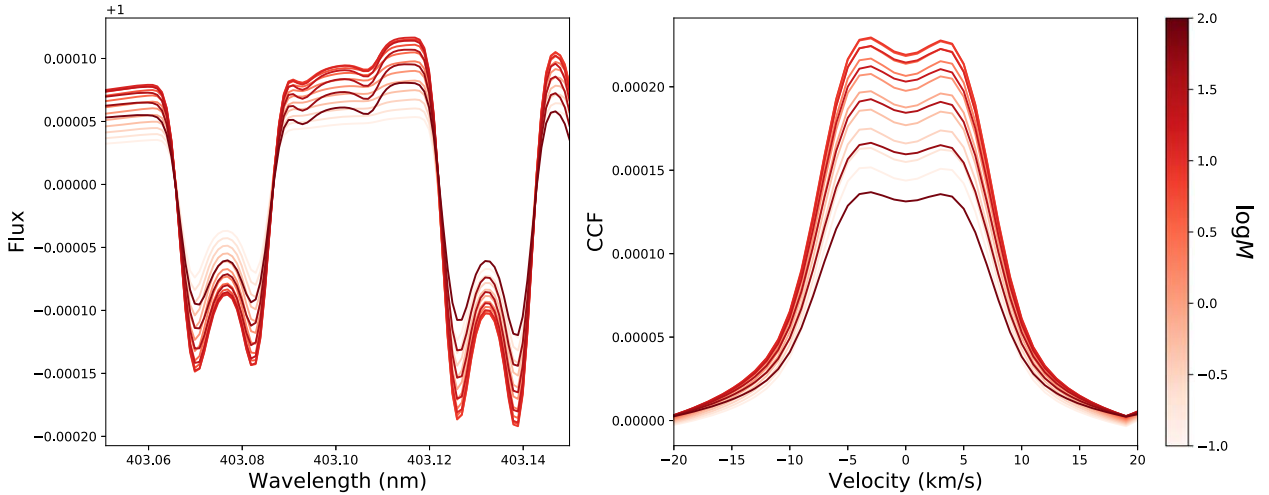
Figure 11. Noise-free synthetic spectra with $T = 3100$ K and varying metallicity values. The left-hand plot shows a zoomed-in section of the transmission spectra themselves, while the right-hand plot shows a single cross-correlation with each spectrum and the template for Fe at $T = 3000$ K and $\log M = -1.0$. The darker color corresponds to higher metallicity values.
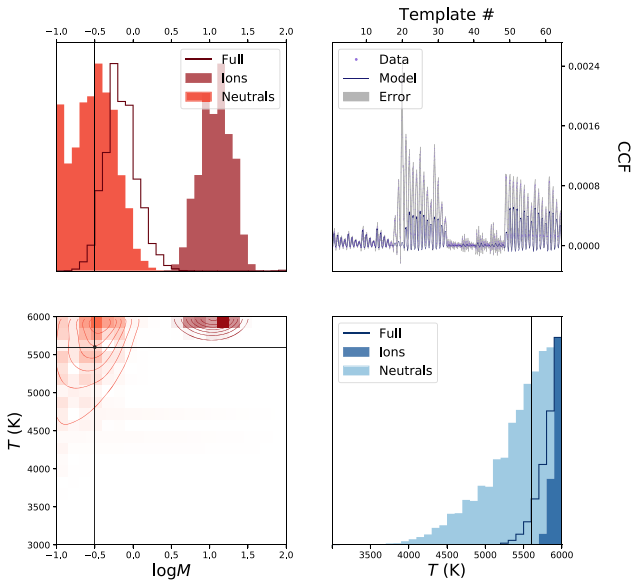
Figure 12. Retrieval performed on the CCF-sequence of the transmission spectrum of KELT-9b measured by the HARPS-N spectrograph. The retrieval is performed in three different ways: using only neutrals (Fe, Ti); (see Figure 14), using only ions (Fe$^+$, Ti$^+$) or using all four species ("Full"). The vertical and horizontal lines indicate the median values of the posterior distributions corresponding to the neutrals-only retrieval. The top right panel shows the data points (lilac) with the error region (gray) for the CCF-sequence produced by the KELT-9b HARPS-N data, along with the model (dark purple) corresponding to the medians from the $\log M$ and T posteriors.

lined, darker colored, and lighter colored posteriors show the results from the full, ionized, and neutral retrievals, respectively.

The metallicity prediction greatly varies between the different retrievals, which is not unexpected here. The extremely high temperatures cause most of the neutral species to be ionized, leading to low abundances for Fe and Ti. Thus, in the neutral retrieval we predict a low $\log M$ value of $-0.5^{+0.2}_{-0.4}$, while the ion retrieval predicts 1.0 ± 0.2. The full retrieval lies further toward the neutral prediction, with $\log M = -0.2 \pm 0.2$, which is unsurprising due to the stronger feature importance in the neutral CCFs for metallicity.

When the Fe$^+$ CCFs are included, i.e., in the full and ion retrievals, the temperature prediction is forced to its upper limit in an attempt to match the strong Fe$^+$ lines ($T = 6000^{+0}_{-200}$K and $T = 6000^{+0}_{-100}$K for the full and ion retrievals, respectively). However, in the neutral retrieval we still obtain a very high temperature value of 5600^{+400}_{-600}K, suggesting it is not only the excess Fe$^+$ that escalates the temperature prediction. Figure 14 shows the "predicted versus real" graphs for the forest trained only on the neutrals. As the temperature increases, this forest's predictive ability decreases, as expected due to ionization. This suggests that the neutral posterior for temperature in Figure 12 may not be reliable. A positive conclusion is that this method is able to identify when a model is flawed.

Using *TESS* photometry, Wong et al. (2019) constrain the dayside and nightside temperatures of KELT-9b to be 4570 ± 90 K and 3020 ± 90 K, respectively. However, this is not inconsistent with a higher retrieved temperature from transmission spectroscopy. The dayside spectrum traces higher pressures than the transmission spectrum, which probes tenuous layers of the upper atmosphere. The present retrieval would be consistent with the scenario of an inversion layer, as is predicted in highly irradiated exoplanets (Hubeny et al. 2003; Fortney et al. 2008).

4. Discussion

4.1. Comparison to Nested Sampling

One of the most common techniques for performing atmospheric retrieval is nested sampling (Skilling 2006; Feroz & Hobson 2008; Feroz et al. 2009, 2019). In a traditional retrieval, a relatively computationally inexpensive forward model is used to generate spectra on the fly, while the sampling method searches the parameter space for the optimal solution. Brogi & Line (2019) demonstrate a method for performing retrieval on high-resolution data with nested sampling, but are restricted to ~4000 spectral data points of the CRIRES instrument. As the number of spectral points increases, so does the time required to compute the models, making this method infeasible for a full HARPS-N spectrum with ~300,000 points and multiple free parameters.
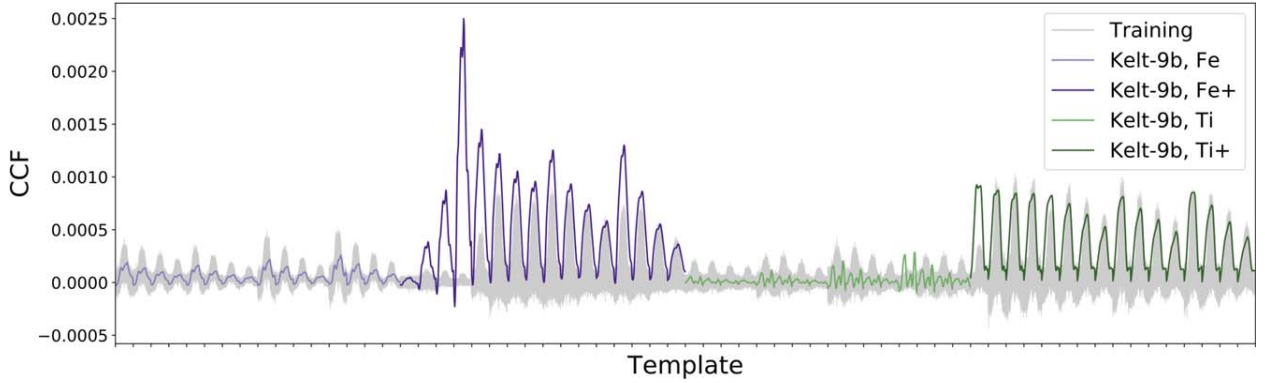
Figure 13. Training vs. measured KELT-9b CCF-sequences. The measured CCF-sequence for Fe$^+$ lies outside of the range of the model CCF-sequences, thus flagging missing physics in the model grid.
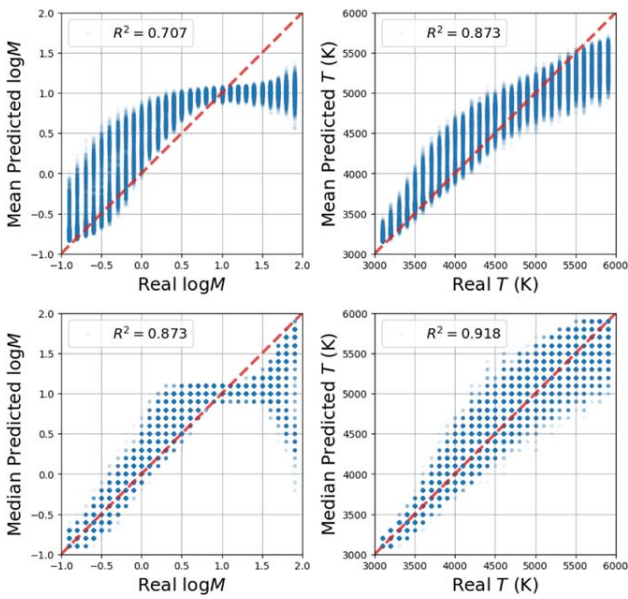
Figure 14. Predicted vs. real values for the forest trained on the CCF-sequences with only neutral species, Fe and Ti. The top and bottom panels show the predictions using the means and medians, respectively. The coefficient of determination (R^2) varies from -1 to 1, where values near unity indicate strong anticorrelations or correlations between the real and predicted values of a given parameter, based on the variance of the outcomes. The retrieval using this forest on the KELT-9b data is shown as the lighter colored posteriors in Figure 12.

Our method of constructing CCF-sequences allows us to reduce the dimensionality down from $\sim$300,000 to $\sim$2500. However, now the computational time for each model is much greater as it involves first generating the spectrum and then cross-correlating 64 times with the different templates. Therefore, it remains infeasible to use a standard nested-sampling retrieval for this technique. The random forest requires a grid of precomputed models to train on, allowing the computational burden to be shifted offline. An alternative method using nested sampling could be employed by interpolating on the same grid of models, but without the added noise. There are a few disadvantages involved with this when compared with the forest.

First, the prediction time on a single spectrum is still orders of magnitudes slower than the pretrained forest ($\sim$20 s versus $\sim$0.05 s). This increased computational speed allows the forest to produce "predicted versus real" graphs, as shown in Figure 6 for $\sim$50,000 models. These graphs give crucial information

about the ability to predict each parameter and the performance of one's retrieval over a vast range of models. We also gain additional information from the random forest, such as the feature importance plots shown in Figure 8. This quantifies the information content in each spectral point with respect to each parameter being retrieved and can be used to infer which areas of the spectrum are most affected by each parameter. It gives us a deeper insight into how the retrieval works, and even indicates which spectral regions might be most informative when considering future observations.

Second, the use of the likelihood function in nested sampling assumes that the error bars on each spectral point are independent. While this is usually a good assumption, in the process of generating the CCF-sequences we repeatedly cross-correlate a single spectrum with multiple templates and then concatenate these into a sequence. This implies that the noise samples corresponding to each individual cross-correlation cannot be independent as they propagate from the same spectrum. With this assumption broken, it becomes unclear how to proceed with a nested-sampling retrieval on the CCF-sequences.

Third, as discussed in Section 1.3, another assumption one needs to make with nested sampling is a form for the likelihood function, and thus the error bars. For example, it is commonly assumed that the error bars are Gaussians, leading to a likelihood function as shown in Equation (1). The forest also requires an assumption of a random distribution when adding noise to the training set, however it does not depend on a likelihood function. As a test, we generated a model CCF-sequence for a mock retrieval, but this time we added noise by drawing from a Cauchy distribution as opposed to a Gaussian. The motivation behind using a Cauchy distribution is that it does not obey the central limit theorem, and thus the likelihood across many points in a spectrum does not behave as a Gaussian. This provides a challenging test for retrieval methods that assume normally distributed error bars. We performed these retrievals using a forest trained on models with Gaussian errors and a nested-sampling algorithm assuming a Gaussian likelihood function, as shown in Equation (1). Note that this likelihood does not use the cross-correlation function, unlike in Brogi & Line (2019). The results are shown in Figure 15. We can see that while the posteriors are wide for the forest, they still encapsulate the true values, whereas the nested-sampling retrieval produces tightly constricted, incorrect posteriors. This suggests that the forest is more robust to differences in error distributions.
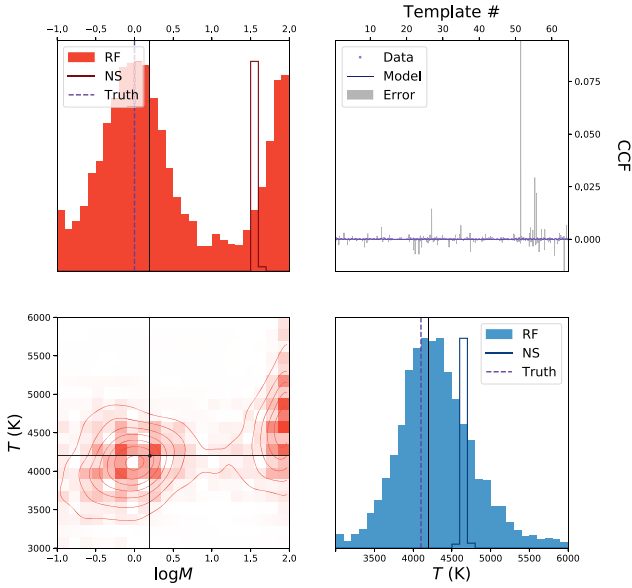
Figure 15. A mock retrieval performed on a model with $\log M = 0$ and $T = 4100$ K, using the CCF-sequence where the noise has been drawn from a Cauchy distribution. The solid posteriors show the random forest (RF) retrieval results, trained on the CCF-sequences with Gaussian noise models (see Figure 6). The empty line posteriors show the nested-sampling (NS) retrieval results using a model that interpolates on the grid of noise-free CCF-sequences and has a Gaussian likelihood. The black lines show the median values of the random forest. The purple, dashed lines show the true values. The top right panel shows the data points (lilac) with the error region (gray), along with the model (dark purple) corresponding to the medians from the $\log M$ and T posteriors.

4.2. Velocity–Velocity Space Performance

So far we have only explored the effects of temperature and metallicity and assumed the velocity parameters, $V_{\rm sys}$ and K_p, are fixed to previously determined values. It is possible that neglecting velocities could lead to severe biases in retrievals. To investigate this, we added the systemic velocity ($V_{\rm sys}$) and the error in the semiamplitude of the planet radial velocity (ΔK_p) to the method. We took $V_{\rm sys}$ from $-10\,\mathrm{km\,s^{-1}}$ to $+10\,\mathrm{km\,s^{-1}}$ in steps of $2\,\mathrm{km\,s^{-1}}$, and ΔK_p from $0\,\mathrm{km\,s^{-1}}$ to $60\,\mathrm{km\,s^{-1}}$ in steps of $6\,\mathrm{km\,s^{-1}}$. Figure A1 shows the results of testing the random forest trained on the CCF-sequences, including the velocity parameters.

This test shows that the addition of the velocity parameters does somewhat reduce the predictive ability of the other parameters, however this reduction is extremely minor for the temperature and not too problematic for the metallicity. The method is able to perfectly retrieve the systemic velocity ($V_{\rm sys}$), but struggles with the error in semiamplitude of the planet radial velocity (ΔK_p). An error in the assumed value of K_p leads to a misalignment of the planet absorption line when summing in the planet rest frame, effectively resulting in a broadening of the CCF. This makes it more challenging to distinguish between sequences of different metallicities. This explains the greater uncertainty in metallicity that we see in Figure A1 when compared with the results from the method without the velocity parameters (Figure 6).

4.3. Comparison to Other Machine-learning Techniques

There are several other machine-learning methods that can be used to perform atmospheric retrieval (Waldmann 2016; Zingales & Waldmann 2018; Cobb et al. 2019), each with their

own advantages. We tested the same CCF-sequence retrieval as before, but now using a standard neural network and a standard Bayesian neural network (BNN); (Gal 2016). In both cases we used a standard multilayer perceptron architecture with three layers. Each layer consists of a linear transformation with bias followed by a ReLU activation, except the last layer, which does not apply an activation function. The first layer transforms spectra from the input space $\mathbb{R}^{2560}$ to an intermediate representation $\mathbb{R}^{512}$. Similarly, layer 2 maps elements to $\mathbb{R}^{32}$, and layer 3 maps elements to the space of parameters $\mathbb{R}^2$. The BNN also applies dropout (Srivastava et al. 2014) with probability 0.15 on the output of layers 1 and 2, as explained in Gal (2016). We implemented both networks using the `PyTorch` library for automatic differentiation (Paszke et al. 2017) and used `Adam` (Kingma & Ba 2014) as the optimization method.

The results of the test predictions are shown in Figure A2. Compared to the random forest, they both perform with slightly improved R^2 scores. However, this is only a measure of the average prediction. In atmospheric retrieval, we are predominantly interested in the range of possible parameter values given by a retrieval, and therefore the posteriors of each parameter. A traditional neural network does not produce posteriors, so it cannot be meaningfully applied to this retrieval problem. The BNN does provide posteriors, so we are able to compare these to the forest. Figure A3 shows the comparison for two mock retrievals, one with $\log M = 1.0$ and $T = 5100$ K (top panel), and one with $\log M = 1.9$ and $T = 4200$ K (bottom panel). For the first retrieval, the forest and the BNN produce very similar results, with the BNN posteriors slightly tighter and more centered on the true values. However, in the second retrieval the BNN does not perform well for the metallicity prediction. This mock spectrum was selected as one of the retrievals with a strong metallicity degeneracy, as discussed in Section 3.3, in order to test how the two methods deal with these issues. The results for the metallicity prediction are $\log M = 0.3^{+1.7}_{-0.7}$ for the forest, and $\log M = 0.7^{+0.2}_{-0.2}$ for the BNN. Both the average predictions are heavily offset from the correct value, however the posterior from the forest captures the degenerate behavior in metallicity, and therefore encompasses the correct value inside the 1σ interval. In contrast, the BNN posterior sits in the middle of the degenerate peaks and remains tightly constrained around the offset value. It is worth noting that this implementation of the BNN is not equivalent to the one used in Cobb et al. (2019), as they use a different form of the likelihood which has not been tested on such high-resolution data.

4.4. Clarification with Respect to Cobb et al. (2019)

In Cobb et al. (2019), it was suggested that the random forest in Márquez-Neila et al. (2018) has the potential to produce overconfident, incorrect posteriors based on a mock retrieval from a test data set. This forest was trained on WFC3 spectra with 13 data points and predicted 5 parameters—temperature, free chemical abundances of H_2O, HCN, and NH_3, and a gray cloud opacity, κ_0. The opacities were calculated with `HELIOS-K` (Grimm & Heng 2015), using the `ExoMol`[6] (Tennyson et al. 2016) spectroscopic linelists for H_2O (Polyansky et al. 2018), HCN (Barber et al. 2014), and NH_3 (Yurchenko et al. 2011).

---

[6] http://exomol.com

The mock spectrum tested on by Cobb et al. (2019) has $T = 1479.6$ K, $\log X_{H_2O} = -9.79$, $\log X_{HCN} = -9.04$, $\log X_{NH_3} = -5.91$, and $\log \kappa_0 = 1.87$. The retrieved posterior for NH$_3$ was tightly constrained and offset from the correct value, which was used to infer that the forest could produce spurious results. However, we ran the same retrieval with nested sampling, using the same model with the same assumptions. Figure A4 shows the results from the random forest retrieval (left panel) and the nested-sampling retrieval (right panel). The posteriors appear very comparable, with the same behavior in the ammonia abundance.

At the time of publishing Márquez-Neila et al. (2018), there were no opacity linelists available for NH$_3$ for temperatures above 1500 K. To deal with this, as stated in Márquez-Neila et al. (2018), the opacity for NH$_3$ was artificially set to zero, and the abundance to the minimum in the range, 10^{-13}. Notable in this particular mock spectrum is the high cloud opacity, equivalent to a cloud top pressure of ~ 1 μbar. This results in an essentially flat spectrum. When retrieving on a flat line, the only two parameters in this model having an effect are the temperature and the cloud opacity, which are perfectly degenerate with each other (i.e., an increase in either results in an upwards shift of the line, so by decreasing the other, one obtains the same spectrum). This degeneracy means one can only obtain lower bounds for the temperature and cloud opacity, corresponding to the upper bound of the other parameter's prior. A consequence of this is a collection of posterior samples in the region $T > 1500$ K, which, as forced by the model, have $\log X_{NH_3} = -13$, resulting in the peaked posterior for NH$_3$. Therefore, this offset posterior is actually an artifact of the training set rather than of the random forest. This is shown conclusively in Figure A4, as the forest's posteriors agree with the true Bayesian posteriors from nested sampling.

5. Conclusion

This paper presents a new method for performing atmospheric retrieval on ground-based, high-resolution spectroscopic observations of exoplanets. By using a combination of CCFs we are able to reduce the dimensionality of the problem and decrease the high levels of uncertainty on each spectral data point. Using our previously demonstrated random forest retrieval technique (Márquez-Neila et al. 2018), we can execute the retrieval quickly and run a multitude of tests of the method. These show that the method performs well on mock observations, with a high predictive power for metallicity and temperature ($R^2 = 0.918$ and 0.986, respectively). The random forest also provides feature importance plots, which show that the neutral cross-correlations are most important for determining the metallicity, while the temperature prediction relies predominantly on the ions. Our method also highlights the metallicity degeneracy in the model, which accounts for the reduced predictability at high metallicity values.

We performed the retrieval on HARPS-N observations for the ultra-hot Jupiter KELT-9b. The metallicity appears to be consistent with solar, with the retrieval seemingly driven by the neutral species. The prediction for temperature is forced up to exceptionally high values, due to excess Fe$^+$ absorption that appears in the high-resolution transmission spectrum, suggesting the need for more complex physics in the model. This can be seen when comparing the data to the training set, which also implies that this method is able to recognize when the model is incomplete.

We also compared the use of our random forest to other approaches, such as the traditional nested-sampling technique and other machine-learning methods. We showed that the forest is more robust to the use of different error distributions than nested sampling, due to it being likelihood free. When compared with a BNN, although the BNN obtains marginally improved R^2 scores, only the forest was able to produce complex posteriors, e.g., in the case of degenerate metallicity values. We also demonstrated that the claim in Cobb et al. (2019), that the forest can be overconfident but incorrect, is actually an outcome of the atmospheric model itself and that the forest's posteriors agree with the results from nested sampling.

Appendix
Additional Figures

Figure A1 shows the results for the forest including the velocity parameters, as discussed in Section 4.2. Figures A2 and A3 show the results for the neural network and Bayesian neural network trained on the CCF-sequences, as discussed in Section 4.3. Figure A4 shows the comparison between a random forest and a nested sampling retrieval for the model and spectrum considered in Cobb et al. (2019), as discussed in Section 4.4.
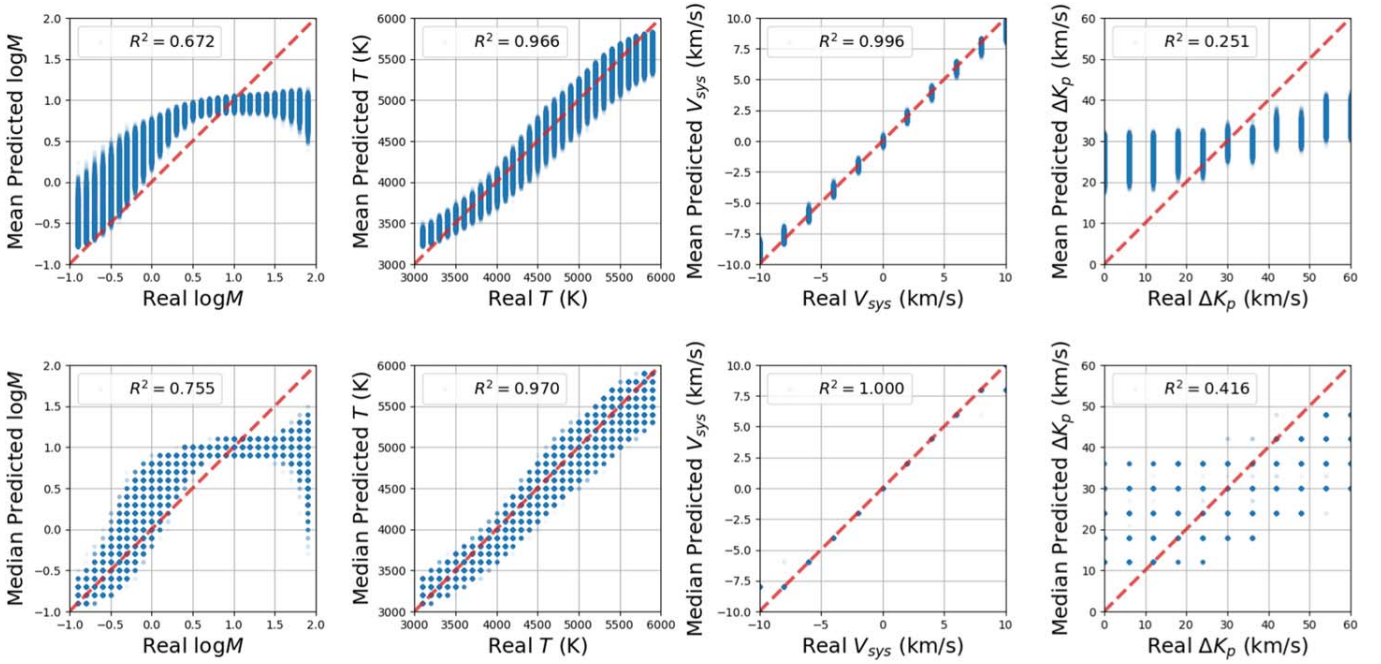
Figure A1. Predicted vs. real values of the logarithm of metallicity (log$M$), temperature (T), systemic velocity (V_{sys}), and error in semiamplitude of the planet radial velocity (ΔK_p) for the random forest trained on the CCF-sequences. The top and bottom rows show the predictions using the means and medians, respectively. The coefficient of determination (R^2) varies from -1 to 1, where values near unity indicate strong anticorrelations or correlations between the real and predicted values of a given parameter, based on the variance of the outcomes.
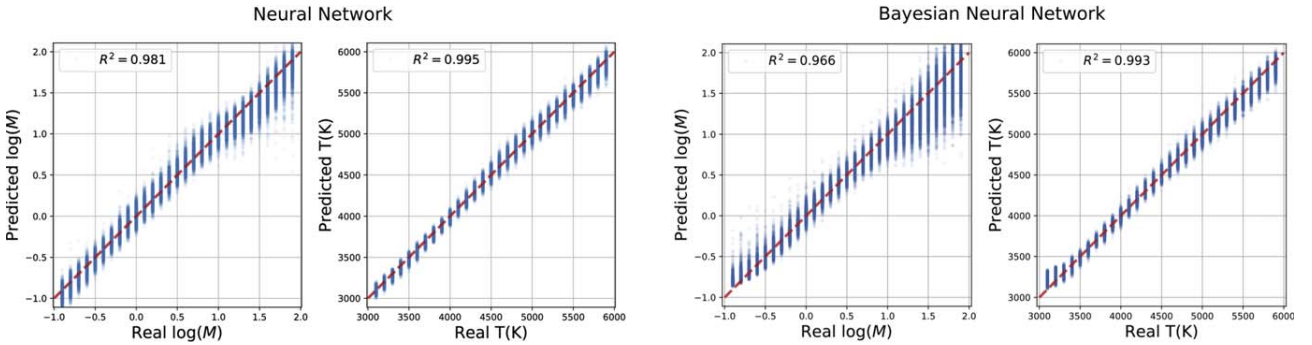
Figure A2. Predicted vs. real values for neural networks trained on the CCF-sequences. The left and right pairs show the results for a standard neural network and a Bayesian neural network (BNN), respectively. The coefficient of determination (R^2) varies from -1 to 1, where values near unity indicate strong anticorrelations or correlations between the real and predicted values of a given parameter, based on the variance of the outcomes. Mock retrievals for the BNN are shown as the empty line posteriors in Figure A3.
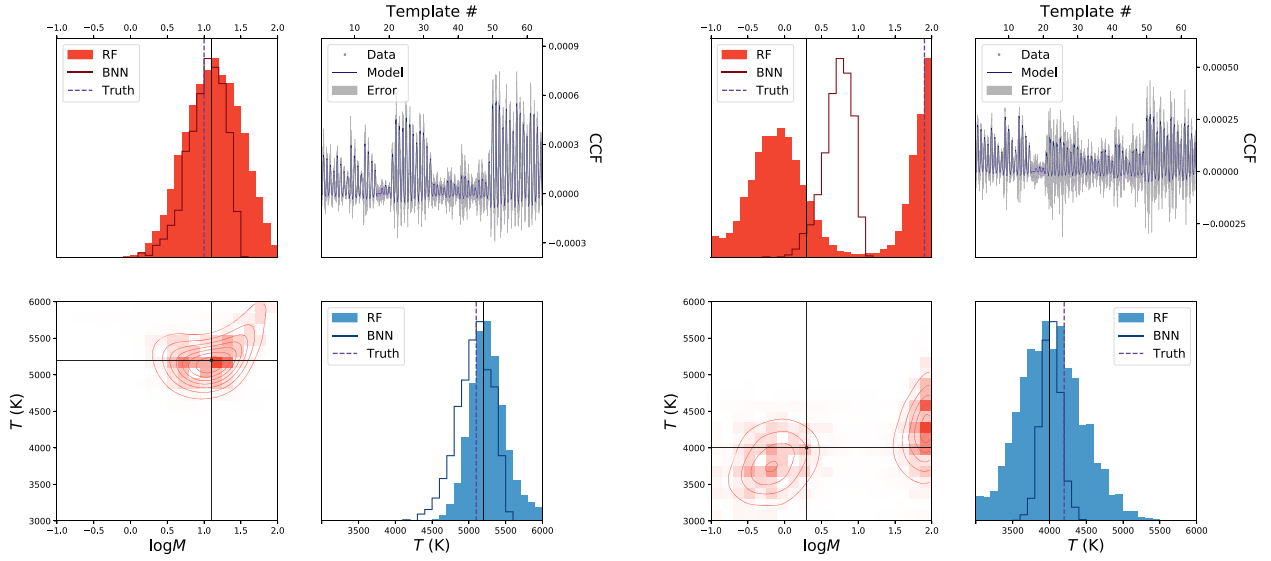
Figure A3. Two mock retrievals performed using the CCF-sequences. The solid posteriors show the random forest retrieval results (see Figure 6). The empty line posteriors show the Bayesian neural network retrieval results (see bottom panels of Figure A2). The black lines show the median values of the random forest, and the purple, dashed lines show the true values. The left figure corresponds to a retrieval on a model with $\log M = 1.0$ and $T = 5100$ K. The right figure corresponds to a retrieval on a model with $\log M = 1.9$ and $T = 4200$ K.
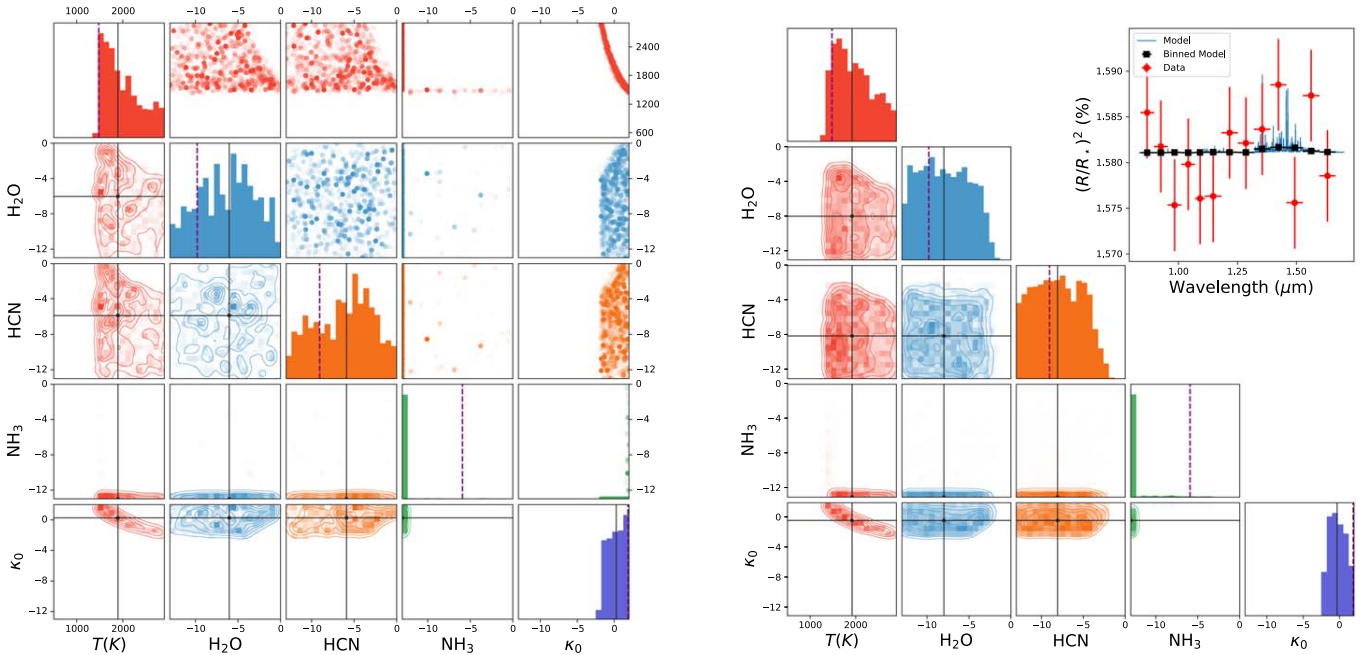
Figure A4. Retrieval results for a mock spectrum with $T = 1479.6$ K, $\log X_{H_2O} = -9.79$, $\log X_{HCN} = -9.04$, $\log X_{NH_3} = -5.91$, and $\log \kappa_0 = 1.87$. The left- and right-hand plots show the results using a random forest and nested sampling, respectively. The black lines show the median predictions. The purple, dashed lines show the true values.

ORCID iDs

Chloe Fisher ⓘ https://orcid.org/0000-0003-0652-2902
Daniel Kitzmann ⓘ https://orcid.org/0000-0003-4269-3311
Simon L. Grimm ⓘ https://orcid.org/0000-0002-0632-4407
Kevin Heng ⓘ https://orcid.org/0000-0003-1907-5910

References

Arcangeli, J., Désert, J.-M., Line, M. R., et al. 2018, ApJL, 855, L30
Barber, R. J., Strange, J. K., Hill, C., et al. 2014, MNRAS, 437, 1828
Benneke, B., & Seager, S. 2012, ApJ, 753, 100
Birkby, J. L., de Kok, R. J., Brogi, M., et al. 2013, MNRASL, 436, L35
Birkby, J. L., de Kok, R. J., Brogi, M., et al. 2017, ApJ, 153, 138
Bower, D. J., Kitzmann, D., Wolf, A. S., et al. 2019, A&A, 631, A103
Breiman, L. 2001, Machine Learning, 45, 5
Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. 1984, Classification and Regression Trees (Boca Raton, FL: Chapman & Hall/CRC Press)
Brogi, M., de Kok, R. J., Birkby, J. L., et al. 2014, A&A, 565, A124
Brogi, M., Giacobbe, P., Guilluy, G., et al. 2018, A&A, 615, A16
Brogi, M., Line, M., Bean, J., et al. 2017, ApJL, 839, L2
Brogi, M., & Line, M. R. 2019, AJ, 157, 114
Brogi, M., Snellen, I. A. G., de Kok, R. J., et al. 2012, Natur, 486, 502
Brogi, M., Snellen, I. A. G., de Kok, R. J., et al. 2013, ApJ, 767, 27
Brown, T. M. 2001, ApJ, 553, 1006
Brown, T. M., Libbrecht, K. G., & Charbonneau, D. 2002, PASP, 114, 826
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125

Cauley, W., Shkolnik, E. L., Ilyin, I., et al. 2019, AJ, 157, 69
Cobb, A. D., Himes, M. D., Soboczenski, F., et al. 2019, AJ, 158, 33
Cosentino, R., Lovis, C., Pepe, F., et al. 2012, Proc. SPIE, 8446, 84461V
Criminisi, A., Shotton, J., & Konukoglu, E. 2011, Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning, Microsoft Research Technical Report, TR-2011-114
de Kok, R. J., Brogi, M., Snellen, I. A. G., et al. 2013, A&A, 554, A82
Dekker, H., D'Odorico, S., Kaufer, A., et al. 2000, Proc. SPIE, 4008, 534
Deming, D., Brown, T. M., Charbonneau, D., et al. 2005, ApJ, 622, 1149
Deming, D., Wilkins, A., McCullough, P., et al. 2013, ApJ, 774, 95
Donati, J.-F., Kouach, D., Lacombe, M., et al. 2018, in Handbook of Exoplanets, ed. H. Deeg & J. Belmonte (Cham: Springer), 107
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium (Princeton, NJ: Princeton Univ. Press)
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, OJAp, 2, 19
Fisher, C., & Heng, K. 2018, MNRAS, 481, 4698
Follert, R., Dorn, R. J., Oliva, E., et al. 2014, Proc. SPIE, 9147, 19
Fortney, J., Lodders, K., Marley, M., & Freedman, R. 2008, ApJ, 678, 1419
Fortney, J., Shabram, M., Showman, A. P., et al. 2010, ApJ, 709, 1396
Gaidos, E., Kitzmann, D., & Heng, K. 2017, MNRAS, 468, 3418
Gal, Y. 2016, PhD thesis, Univ. Cambridge
Gaudi, B. S., Stassun, K. G., Collins, K. A., et al. 2017, Natur, 546, 514
Gibson, N. P., Merritt, S., Nugroho, S. K., et al. 2020, MNRAS, 493, 2215
Grimm, S. L., & Heng, K. 2015, ApJ, 808, 182
Guilluy, G., Sozzetti, A., Brogi, M., et al. 2019, A&A, 625, A107
Hastie, T., Tibshirani, R., & Friedman, J. 2001, The Elements of Statistical Learning (New York: Springer)
Heng, K. 2017, Exoplanetary Atmospheres: Theoretical Concepts and Foundations (Princeton, NJ: Princeton Univ. Press)
Heng, K., & Kitzmann, D. 2017, MNRAS, 470, 2972
Ho, T. K. 1998, ITPAM, 20, 832
Hoeijmakers, H. J., de Kok, R. J., Snellen, I. A. G., et al. 2015, A&A, 575, A20
Hoeijmakers, H. J., Ehrenreich, D., Heng, K., et al. 2018, Natur, 560, 453
Hoeijmakers, H. J., Ehrenreich, D., Kitzmann, D., et al. 2019, A&A, 627, A165
Hubeny, I., Burrows, A., & Sudarsky, D. 2003, ApJ, 594, 1011
Hunter, J. D. 2007, CSE, 9, 90
John, T. L. 1988, A&A, 193, 189
Kaeufl, H.-U., Ballester, P., Biereichel, P., et al. 2004, SPIE, 5492, 1218
Khalafinejad, S., Salz, M., & Cubillos, P. E. 2018, A&A, 618, A98
Khalafinejad, S., von Essen, C., & Hoeijmakers, H. J. 2017, A&A, 598, A131
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Kitzmann, D., Heng, K., Rimmer, P. B., et al. 2018, ApJ, 863, 183
Kurucz, R. L. 2017, CaJPh, 95, 825
Lavie, B., Mendonça, J. M., Mordasini, C., et al. 2017, AJ, 154, 91
Line, M. R., Knutson, H., Deming, D., et al. 2013, ApJ, 778, 183
Line, M. R., Wolf, A. S., Zhang, X., et al. 2013, ApJ, 775, 137
Lockwood, A. C., Johnson, J. A., Bender, C. F., et al. 2014, ApJL, 783, L29
MacDonald, R. J., & Madhusudhan, N. 2017, MNRAS, 469, 1979
Márquez-Neila, P., Fisher, C., Sznitman, R., & Heng, K. 2018, NatAs, 2, 719
Mayor, M., Pepe, F., Queloz, D., et al. 2003, Msngr, 114, 20
McLean, I. S., Becklin, E. E., Bendiksen, O., et al. 1998, Proc. SPIE, 3354, 566
Noguchi, K., Aoki, W., Kawanomoto, S., et al. 2002, PASJ, 54, 855

Nugroho, S. K., Kawahara, H., Masuda, K., et al. 2017, AJ, 154, 221
Origlia, L., Oliva, E., Baffa, C., et al. 2014, Proc. SPIE, 9147, 91471E
Park, C., Jaffe, D. T., Yuk, I.-S., et al. 2014, Proc. SPIE, 9147, 91471D
Paszke, A., Gross, S., Chintala, S., et al. 2017, Automatic differentiation in PyTorch, https://openreview.net/forum?id=BJJsrmfCZ
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Pepe, F., Molaro, P., Cristiani, S., et al. 2014, AN, 335, 8
Piskorz, D., Benneke, B., Crockett, N. R., et al. 2016, ApJ, 832, 131
Piskorz, D., Benneke, B., Crockett, N. R., et al. 2017, AJ, 154, 78
Piskorz, D., Buzard, C., Line, M. R., et al. 2018, AJ, 156, 133
Polyansky, O. L., Kyuberis, A. A., Zobov, N. F., et al. 2018, MNRAS, 480, 2597
Quirrenbach, A., Amado, P. J., Mandel, H., et al. 2010, Proc. SPIE, 7735, 773513
Rayner, J., Tokunaga, A., Jaffe, D., et al. 2016, Proc. SPIE, 9908, 990884
Redfield, S., Endl, M., Cochran, W. D., & Koesterke, L. 2008, ApJL, 673, L87
Richard, C., Gordon, I. E., Rothman, L. S., et al. 2012, JQSRT, 113, 1276
Rieke, N., Tan, D. J., Alsheakhali, M., et al. 2015, in Medical Image Computing and Computer-assisted Intervention—MICCAI 2015, ed. N. Navab et al. (Cham: Springer), 266
Rothman, L. S., Rinsland, C. P., Goldman, A., et al. 1998, JQSRT, 60, 665
Saha, M. N. 1920, PMag, 40, 472
Seidel, J. V., Ehrenreich, D., Wyttenbach, A., et al. 2019, A&A, 623, A166
Sisson, S. A., Fan, Y., & Beaumont, M. A. 2019, Handbook of Approximate Bayesian Computation (Boca Raton, FL: CRC Press)
Skilling, J. 2006, BayAn, 1, 833
Snellen, I. A. G., Albrecht, S., de Mooij, E. J. W., & Le Poole, R. S. 2008, A&A, 487, 357
Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., & Albrecht, S. 2010, Natur, 465, 1049
Srivastava, N., Hinton, G., Krizhevsky, A., et al. 2014, JMLR, 15, 1929
Stock, J. W., Kitzmann, D., Patzer, A. B. C., & Sedlmayr, E. 2018, MNRAS, 479, 865
Strassmeier, K. G., Ilyin, I., Järvinen, A., et al. 2015, AN, 336, 324
Sznitman, R., Becker, C., Fleuret, F., & Fua, P. 2013, in 2013 IEEE Conf. on Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE), 3270
Tennyson, J., Yurchenko, S. N., Al-Refaie, A. F., et al. 2016, JMoSp, 327, 73
The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Virtanen, P., Gommers, R., & Oliphant, T. E. 2020, Nature Methods, 17, 261
Waldmann, I. P. 2016, ApJ, 820, 107
Waldmann, I. P., Tinetti, G., Rocchetto, M., et al. 2015, ApJ, 802, 107
Wiedemann, G., Deming, D., & Bjoraker, G. 2001, ApJ, 546, 1068
Wildi, F., Blind, N., Reshetov, V., et al. 2017, Proc SPIE, 10400, 1040018
Wong, I., Shporer, A., Morris, B. M., et al. 2019, arXiv:1910.01607
Wyttenbach, A., Ehrenreich, D., Lovis, C., et al. 2015, A&A, 577, A62
Wyttenbach, A., Lovis, C., Ehrenreich, D., et al. 2017, A&A, 602, A36
Yurchenko, S. N., Barber, R. J., & Tennyson, J. 2011, MNRAS, 413, 1828
Zerbi, F. M., Bouchy, F., Fynbo, J., et al. 2014, Proc. SPIE, 9147, 914723
Zhang, L., Varadarajan, J., Suganthan, P., et al. 2017, in 2017 IEEE Conf. on Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE), 5825
Zikic, D., Glocker, B., & Criminisi, A. 2014, Medical Image Analysis, 18, 1262
Zingales, T., & Waldmann, I. P. 2018, AJ, 156, 268

JWST Cycle 1 Proposals

We are now post lagoon.

JENS HOEIJMAKERS

6.1 Introduction

In November 2020, the cycle 1 deadline for the upcoming James Webb Space Telescope took place. In an effort to pool resources and foster collaborations, we set up a proposal team involving researchers from Bern, Lund, Copenhagen, Dublin, and the Space Telescope Science Institute in Maryland. In total we submitted 12 proposals, of which 4 were approved. I led one of the unsuccessful proposals, and supported heavily one of the successful ones. These two proposals were based on very similar science cases, though the intended objects were different. In this chapter I will briefly describe these proposals and their objectives.

6.2 Science Case

6.2.1 Atmospheric Chemistry and Planet Formation

The distance from the star at which a specific volatile compound condenses, such as H_2O, CO or CO_2, is known as the snow/ice line. Protoplanetary disk models are able to predict the evolution of the location of these lines over time, during which planets can form in the disk. By comparing elemental abundances in an exoplanetary atmosphere with that of the host star, one can draw conclusions about the formation and evolution of the system. In particular, the C/O ratio is considered to be one of the key indicators for this (Öberg et al., 2011; Madhusudhan et al., 2014; Öberg and Bergin, 2016). Assuming no strong mixing between the gas and solid components of a planet, a C/O ratio close to that of the host star suggests in situ formation of the planet, inside the water ice line. In contrast, a vastly differing C/O ratio implies the planet has likely migrated across the H_2O, CO and CO_2 ice lines. The study of these ratios will teach us valuable lessons about the formation of exoplanetary systems and our own solar system, as well as providing a powerful empirical test of generalized planet formation/migration theories.

In practice, there are a number of caveats that make it difficult to definitively determine the formation-migration history of a planet from its atmospheric abundances alone. Specifically, there are the effects of photoevaporation, late-stage planetesimal accretion (e.g., from comets), and core-envelope mixing (Madhusudhan, 2019). Also, the specific properties of the initial protoplanetary disk can vary. Multi-planet systems are extremely valuable case studies, since the redundancies of having several planets, as well as the dynamical constraints on possible migration scenarios, improve one's ability to develop a self-consistent formation-migration narrative.

An interesting class of exoplanets for this study are warm Neptunes. Unlike larger exoplanets, these Neptune-sized objects are predicted to have extremely diverse atmospheric compositions, varying significantly with temperature, pressure, and metallicity (Moses et al., 2013). However, until now the limitations of current instruments have prohibited the study of atmospheric features in these smaller planets, with the exception of only a handful – e.g. GJ 3470 b (Benneke et al., 2019a); K2-18 b (Benneke et al., 2019b; Tsiaras et al., 2019); HAT-P-11 b (Fraine

et al., 2014). The James Webb Space Telescope will open up this field by providing the resolution and wavelength coverage required to characterize these planets.

6.2.2 Transmission Spectroscopy with HST and Spitzer

The field of exoplanet atmospheres has flourished in the era of the Hubble Space Telescope (HST) (Deming and Seager, 2017). By observing the starlight that passes through a planet's atmosphere during a transit, observers can obtain a transmission spectrum. Chemical species in the atmosphere will absorb this light at different wavelengths, each imprinting its unique set of absorption features in the spectrum. Current state-of-the-art observing techniques and analysis methodologies using HST have yielded detections of a wide variety of molecules, including H_2O, CH_4, Na and K (e.g. Sing et al. 2016). However, the spectral coverage and resolution provided by HST is limited, and therefore so are the analyses of these spectra.

The highly anticipated James Webb Space Telescope will greatly advance the field of transmission spectroscopy of exoplanets. The array of different instruments and observing modes will provide spectra across a wide range of wavelengths at significantly higher spectral resolutions than HST (e.g. Batalha et al. 2017). In particular, the ability to measure the $> 2\mu$m range at medium resolutions will allow us to constrain the carbon-bearing molecules better than ever before. This in turn will provide a vast improvement on our constraints of the C/O ratios of exoplanet atmospheres.

6.3 TOI-125

The proposal I led aimed to observe one transit of two planets, b and c, in the TOI-125 system. In this section I will introduce this system and discuss the motivation for our choice of objects.

6.3.1 The TOI-125 System

The TOI-125 system is a relatively recent discovery from the TESS mission. It contains three confirmed sub-Neptune planets on relatively close-in orbits (parameters shown in Table 6.1). Due to intensive HARPS radial velocity measurements, the masses for these planets have been precisely constrained, unlike most planets of this size. In addition, the majority of warm Neptunes favourable for transmission spectroscopy are not known to reside in multi-planet systems. These three planets also have comparable radii and masses, differing only in orbital separation. TOI-125 therefore provides a rare opportunity to precisely characterize and compare similar planets in a single system. This system also has the benefit of being visible by JWST from early May to December every year, providing many observing opportunities.
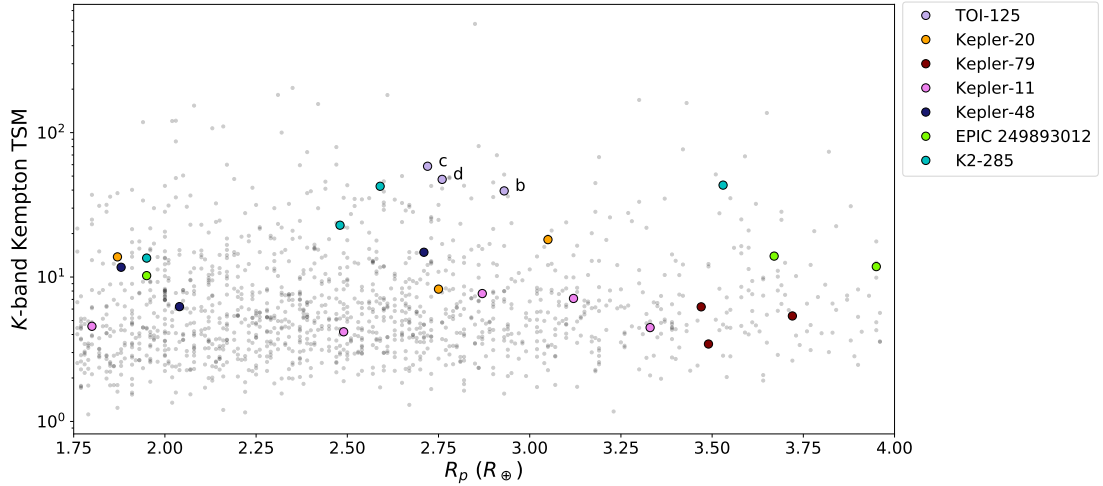
Figure 6.1: Kempton Transmission Spectroscopy Metric (TSM) vs planet radius, for sub-Neptunes ($1.75R_\oplus < R < 4.0R_\oplus$). Planets in a multi-planet system and with measured masses are highlighted.

Planet	Radius ($R_\oplus$)	Period (days)	$\log g$	T_{eq} (K)	H_{eq} (km)
b	2.726 ± 0.075	$4.654^{+0.00033}_{-0.00031}$	3.097 ± 0.047	1037 ± 11	~ 287
c	2.759 ± 0.1	$9.151^{+0.00070}_{-0.00082}$	$2.931^{+0.068}_{-0.076}$	828 ± 9	~ 336
d	2.93 ± 0.17	$19.98^{+0.0050}_{-0.0056}$	3.192 ± 0.064	638 ± 7	~ 142

Table 6.1: Planetary parameters for the TOI-125 system (Nielsen et al., 2020).

The TOI-125 system has several attributes that make it attractive for JWST transmission spectroscopy. Three sub-Neptune exoplanets with radii of $R = 2.7$–$2.9R_\oplus$, gravities of $\log g \approx 2.9$–3.2 and orbital periods of 4.7–20 days orbit a K0 star with $K_s = 8.995$ (Nielsen et al., 2020). With these radii, it is likely that these exoplanets have hydrogen-dominated atmospheres (Fulton and Petigura, 2018), which implies relatively large pressure scale heights (e.g. for planets b and c: $H \sim 300$ km, $H/R \sim 0.01$) that are favorable for transmission spectroscopy. Figure 6.1 shows a plot of a modified Kempton Transmission Spectroscopy Metric (Kempton et al., 2018) for sub-Neptunes. The highlighted planets are those residing in multi-planet systems with measured masses. This clearly shows how the TOI-125 system is one of the best for classification with transmission spectroscopy. Furthermore, the radii being smaller than $3R_\oplus$ suggest that magma-atmosphere interactions that occurred during their formation history may be imprinted in the transmission spectra—in the form of altered water abundances (Kite et al., 2020).

Formation models predict the evolving locations of the snow/ice lines of H_2O, CO, carbon dioxide (CO_2) and ammonia (NH_3) (Burn et al., 2019) within the protoplanetary disk of TOI-125 (Figure 6.2). Since all the planets in the TOI-125 system reside at semi-major axes < 0.2 AU from their star, they are all currently within the water snow/ice line. If these exoplanets formed in-situ, then the formation models predict that their atmospheres should, under some assumptions, have carbon-to-oxygen ratios that match that of the star (Öberg et al., 2011). Hence, a falsifiable hypothesis presents itself, and any outcome that yields C/O measurements for any of these planets will be interesting for constraining formation models.

6.3.2 Mock Retrievals

A distinct challenge of these proposed observations and subsequent interpretations is to accurately retrieve molecular abundances in the face of varying temperature, chemical disequilibrium and clouds. Figures 6.3 and 6.4 show the results from mock retrievals of one transit of TOI-125 b and c with realistic noise for the NIR-Spec BOTS mode with the G235M disperser. For this retrieval we assumed an isothermal, isobaric atmosphere with free chemistry and a non-grey cloud model. The abundances of the molecules in the synthetic spectrum are given by the blue lines in Figure 6.3, and the mean molecular weight of the atmosphere was adjusted
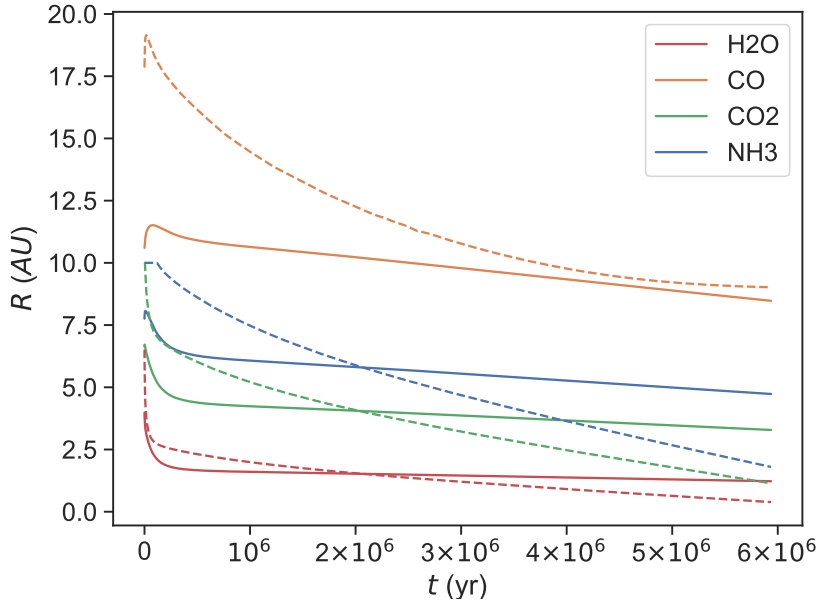
Figure 6.2: Evolution of different ice lines for a disk model of the TOI-125 system (1000g cm^{-2} at 5 AU) for two typical values of the Shakura-Sunyaev α-disk viscosity (dashed: $\alpha = 10^{-2}$; solid: $\alpha = 2 \times 10^{-3}$).

accordingly. The low temperatures of these planets suggest that chemical equilibrium is unlikely (Moses et al., 2013), and free chemistry allows us to explore the parameter space with only the data points to inform us. Including a realistic chemical model would only tighten our constraints further, so therefore our free chemistry retrieval provides an upper limit on the ability to infer the chemical abundances in these atmospheres.

The retrieved posteriors for TOI-125 b in Figure 6.3 shows that key carbon and oxygen carriers are well-constrained (e.g. H_2O, CO_2 and CH_4), and we can place useful upper limits on CO. Due to the larger scale height for planet c (see Table 6.1), the constraints on its atmospheric parameters are even tighter. The cloud parameters correspond to an abundance (κ_0), a cloud particle size (r_c), and a composition (Q_0), which denotes the value of the size parameter $x = 2\pi r_c/\lambda$ at which the extinction efficiency peaks (Kitzmann et al., 2018). Our synthetic data contains a cloud with an opacity of 10^{-3} cm^2 g^{-1}, corresponding to a cloud-top pressure of ~ 10 mbar. Since transmission spectra are thought to probe roughly these pressures, this gives a reasonable cloud coverage. Although we cannot con-

strain the cloud parameters, which is unsurprising due to the wavelength range considered, this demonstrates that we are able to obtain meaningful results on the other atmospheric parameters, even in the presence of a realistic cloud.

Figure 6.3: Mock retrieval of simulated spectra of one transit of TOI-125 b in the G235M mode, with error bars provided by PandExo. The blue lines in the posteriors show the true values for the synthetic spectra. The inset shows the data points and the best-fit model (dark blue) and 2σ uncertainty (light blue).

From these results we could build a posterior for the retrieved C/O ratio, which is shown in Figure 6.4. This shows an accurate retrieval of our true C/O ratio from the simulated spectra of TOI-125 b and c, confirming that our observations would be able to put powerful constraints on planet formation models. We could also put some constraints on a diagnostic of chemical equilibrium. By comparing the
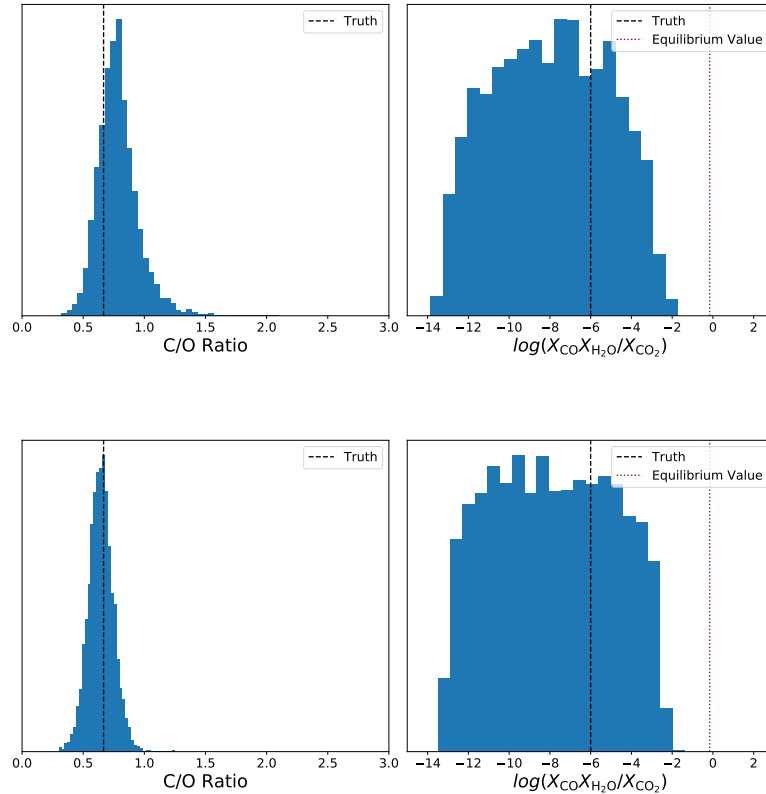
Figure 6.4: Retrieved C/O ratio and chemical equilibrium diagnostic posteriors for TOI-125 b (top) and c (bottom).

inferred value of $X_{CO}X_{H_2O}/X_{CO_2}$ (where X_i is the relative abundance of molecule i) with its equilibrium value, one may empirically determine if the atmosphere is out of chemical equilibrium (Heng and Tsai, 2016), as demonstrated in Figure 6.4.

6.3.3 Observational Details

We requested to observe one transit each of TOI-125 b and c, using the NIRSpec BOTS mode. We aimed to use the G235M disperser, which would provide a wavelength coverage of 1.66-3.07 μm at a resolution of 1000. Although the results from Guzmán-Mesa et al. (2020) concluded that G395M was optimal for studying warm Neptunes, the SNR of the TOI-125 planets is significantly lower than GJ436 b, the planet from this study. Our mock retrievals showed that G235M was a better choice for our system.

We had originally planned to observe all three TOI-125 planets to enable a complete comparison of the system. However, our mock retrievals of TOI-125 d showed significantly less constrained posteriors. This is likely due to both the planet's colder temperature and higher surface gravity, leading to a lower scale height and therefore smaller spectral features.

6.4 TOI-178

This proposal was led by Dr. Matthew Hooten, and involved a collaboration with the CHEOPS (CHaracterising ExOPlanets Satellite) science team. It aimed to observe one transit of TOI-178 b, d, and g. I supported this proposal from the theoretical side, providing mock retrievals and working on the science case, as it relied on the same reasoning as the TOI-125 proposal.

6.4.1 The TOI-178 System

The TESS mission originally found 3 sub-Neptune planets orbiting a nearby late K-type star, TOI-178 (Leleu et al., 2019). However, using data from CHEOPS, transit observations revealed a compact system of at least 6 transiting planets (Leleu et al., 2020)[1]. In addition to the transit observations, the ESPRESSO spectrograph was able to constrain the planetary masses. Table 6.2 shows the parameters for these planets.

From the predicted densities of these planets, we can deduce that the system contains two inner rocky planets (b and c), as well as two potential water worlds (e and f). Planets d and g appear to have maintained significant gaseous envelopes. The planet density versus equilibrium temperature is shown in Figure 6.5, which compares the density trends for multi-planet systems. This variation in density across the TOI-178 system contradicts the primary theories of planet formation (e.g. Mordasini et al., 2012; Alibert et al., 2013; Emsenhuber et al., 2020), which predict a decrease in density with orbital separation due to decreased evaporation of the primordial envelope.

---

[1]This paper was not public at the time of the cycle 1 deadline, but the proposal involved the entire CHEOPS science team (including M. Hooten).

Planet	Radius ($R_\oplus$)	Period (days)	Density [$\rho_\oplus$]	T_{eq} (K)	H_{eq} (km)
b	1.177 ± 0.074	$1.914557^{+0.000016}_{-0.000018}$	$0.93^{+0.29}_{-0.33}$	952 ± 19	~ 308
c	$1.710^{+0.094}_{-0.082}$	$3.238458^{+0.000020}_{-0.000018}$	$0.95^{+0.18}_{-0.21}$	798 ± 16	~ 173
d	2.640 ± 0.069	6.557694 ± 0.000013	$0.16^{+0.04}_{-0.05}$	631 ± 13	~ 536
e	2.169 ± 0.079	$9.961872^{+0.000034}_{-0.000038}$	$0.39^{+0.12}_{-0.10}$	549 ± 11	~ 226
f	2.379 ± 0.086	$15.231930^{+0.00011}_{-0.000085}$	$0.58^{+0.13}_{-0.12}$	475.8 ± 9.6	~ 120
g	2.91 ± 0.11	$20.709460^{+0.000077}_{-0.000069}$	$0.19^{+0.05}_{-0.09}$	429.4 ± 8.5	~ 310

Table 6.2: Planetary parameters for the TOI-178 system (Leleu et al., 2020).
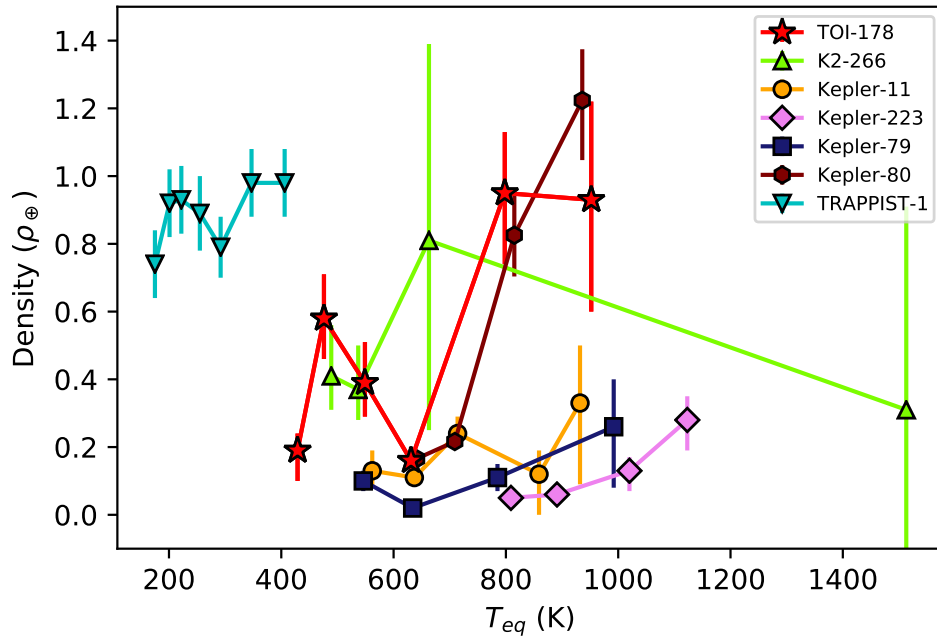
Figure 6.5: Planet density vs equilibrium temperature for a selection of systems with 4 or more transiting planets.

TOI-178 has all the advantages of the TOI-125 system, but with the added benefit of three additional planets. This provided the opportunity to select the best planets for atmospheric characterisation, whilst still enabling a comparison within a single system.

6.4.2 Mock Retrievals

As with TOI-125, we performed mock retrievals on the planets in this system to determine our ability to characterise their atmospheres. Figures 6.6 and 6.7 show the results from the retrievals of one transit of TOI-178 b, d, and g, with realistic noise for the NIRSpec BOTS mode with the G395M disperser. We assumed the same model as for TOI-125. Figure 6.6 shows the best fit model and 3σ uncertainty, along with the simulated data points. These fits clearly show the retrieval's constraining power, and the significant visible spectral features, especially for planets d and g. The posteriors from the retrieval of planet b, the least amenable of the three to transmission spectroscopy, are shown in Figure 6.7. This demonstrates the ability to accurately measure the chemical abundances in this planet's atmosphere, leading to a reasonable constraint on the planet's C/O ratio.

6.4.3 Observational Details

We requested to observe one transit each of TOI-178 b, d, and g, using the NIRSpec BOTS mode with the G395M disperser, which provides a wavelength coverage of 2.87-5.10 μm at a resolution of 1000. Guzmán-Mesa et al. (2020) found that G395M provided the best constraints on the chemical abundances for warm Neptunes, and our retrieval results agreed in this case.
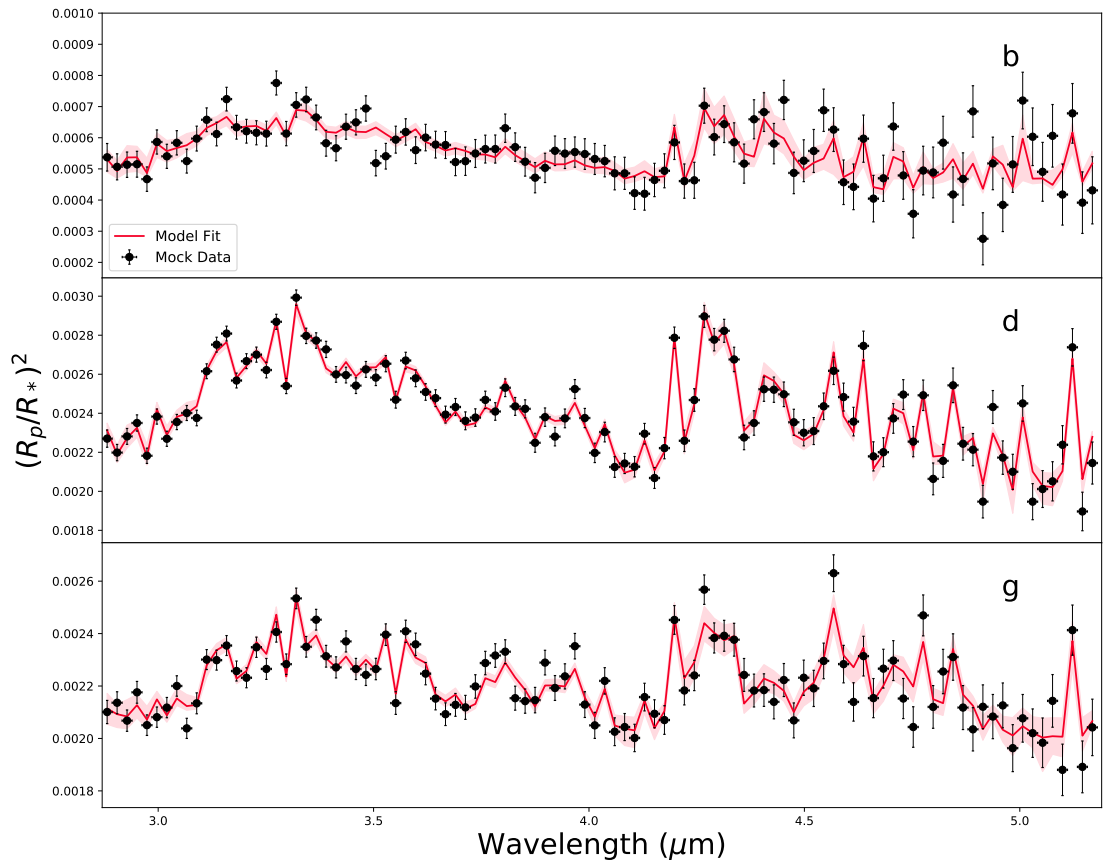
Figure 6.6: Model fits from our mock retrievals of TOI-178 b, d, and g. The black points show the simulated transmission spectra for the G395M mode. The solid red line shows the best-fit model and the pale red area shows the 3σ uncertainty.
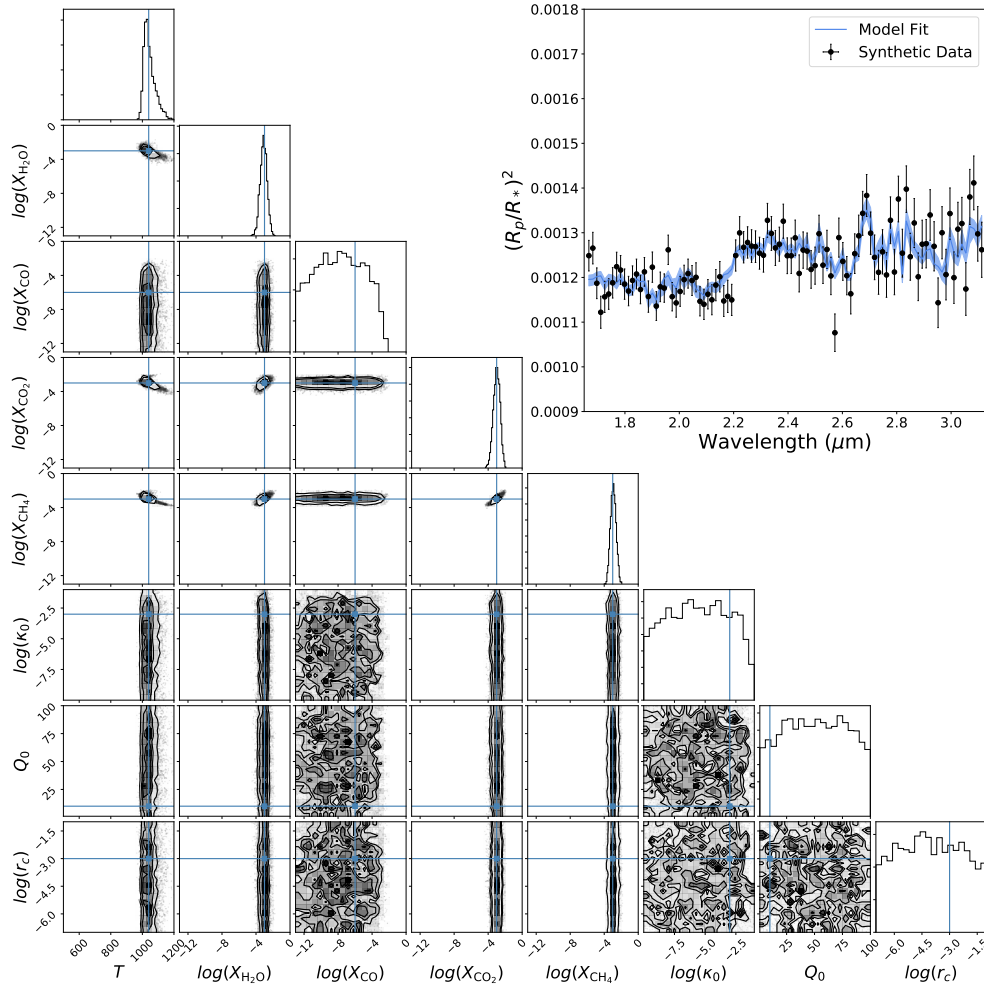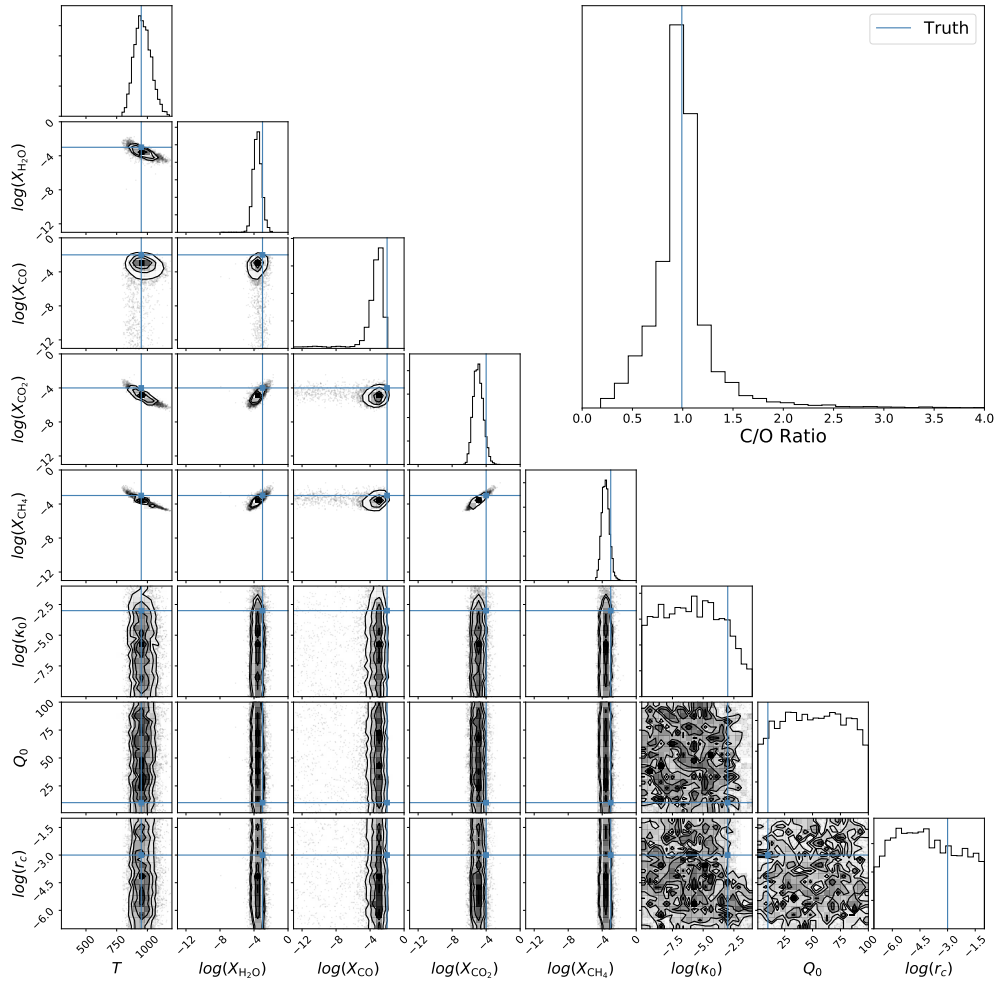
Figure 6.7: Mock retrieval of simulated spectra of one transit of TOI-178 b in the G395M mode, with error bars provided by PandExo. The blue lines in the posteriors show the true values for the synthetic spectra. The inset shows the posterior of the derived C/O ratio.

6.5 Proposal Assessments

Due to the similarity of these two proposals, it is unsurprising that only one was selected. The privileged access our proposal team had to the TOI-178 data gave us the upper hand, and the more optimal targets for transmission spectroscopy made this proposal the better choice. Although the TOI-125 proposal, led by myself, was not selected in this cycle, it was graded in the second quintile. I aim to improve and resubmit this proposal in a future James Webb cycle. I also hope to be heavily involved in the TOI-178 data analysis, the experience of which will be invaluable in improving the TOI-125 proposal.

CHAPTER 7

Supported Research

It has been said that astronomy is a humbling and character-building experience. There is perhaps no better demonstration of the folly of human conceits than this distant image of our tiny world. To me, it underscores our responsibility to deal more kindly with one another, and to preserve and cherish the pale blue dot, the only home we've ever known.

CARL SAGAN

As well as working on my own publications, I have supported several other studies, in collaborations with researchers in Bern and elsewhere. My contributions include performing retrieval analyses, training and assisting other students in running code, and general discussions and text editing. Each study, and my contribution to it, is summarised below.

7.1 Seidel et al. (2019)

Hot Exoplanet Atmospheres Resolved with Transit Spectroscopy (HEARTS) II. A broadened sodium feature on the ultra-hot giant WASP-76 b

In this study, the ultra-hot Jupiter WASP-76 b was observed using the HARPS spectrograph at the European Southern Observatory (ESO) 3.6 m telescope. These observations showed a clear detection of neutral sodium in the planet's atmosphere. An interesting finding in these observations was the significant broadening of the sodium lines, indicative of super-rotation in the upper atmosphere of this planet. Figure 7.1 shows the comparison between the Gaussian fit for the co-added sodium lines and the HARPS instrumental line spread function. It also shows that the FWHM of the Gaussian fit is well below the escape velocity of this planet, suggesting that atmospheric escape of sodium is unlikely.

For this work, I performed a retrieval analysis on the sodium lines, following the methods from my previous work (Fisher and Heng, 2019). However, the extensive broadening of the sodium lines caused too much discrepancy between our model and the data, leading to unreliable posteriors. This retrieval was not included in the final version of the publication.
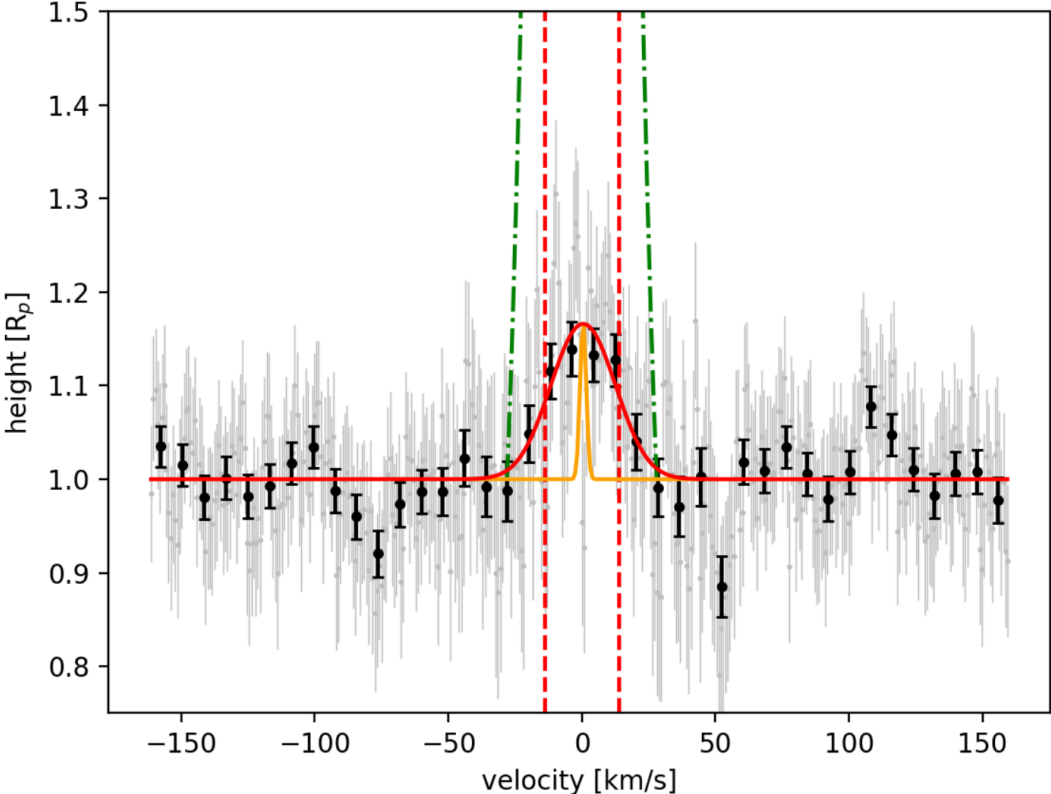
Figure 7.1: Co-added lines of the HARPS transmission spectrum sodium doublet for WASP-76 b, as a function of velocity. The data are shown in grey, and the binned points in black. The red line shows the Gaussian fit to the data, with the FWHM shown by the dashed red lines. The yellow line shows the HARPS instrumental line spread function, and the escape velocity is shown by the green dotted-dashed lines. Figure from Seidel et al. (2019).

7.2 Hoeijmakers et al. (2019)

A spectral survey of an ultra-hot Jupiter
Detection of metals in the transmission spectrum of KELT-9 b

Following the ground-breaking discovery of iron and titanium in the ultra-hot Jupiter KELT-9 b by Hoeijmakers et al. (2018a), this work performed a detection analysis for 39 different species in this planet. The results showed confident detections for 8 species, including Na I, Mg I, Cr II, Sc II, and Y II, as well as the previously detected Fe I, Fe II, and Ti II. Evidence was also found for Ca I, Cr I, Co I, and Sr II.

During this study, I was collaborating with J. Hoeijmakers on our high-resolution retrieval paper (Fisher et al., 2020). This involved working closely with Hoeijmakers and learning a great deal from him on the topic of high-resolution spectroscopy and cross-correlation functions. We also discussed this spectral survey and any implications it might have for our retrieval work.

7.3 Oreshenko et al. (2020)

Supervised Machine Learning for Intercomparison of Model Grids of Brown Dwarfs:
Application to GJ 570D and the Epsilon Indi B Binary System

After our initial proof-of-concept paper on using the Random Forest for atmospheric retrieval (Márquez-Neila et al., 2018), I supported several projects using this method, led by other PhD students in our group. The first was this study, which used the forest to compare brown dwarf model grids from several different groups across the world.

In the field of brown dwarfs, a number of teams create grids of model spectra using their own codes, and then share these grids publicly. However, the proprietary nature of the models themselves prohibit their use in atmospheric retrievals by other groups. An advantage of the Random Forest is that the model itself is not required in this case, as one can build a training set from the model grid directly.

In this work, three model grids were used to create three different training sets. These include the `Sonora` grid (Marley et al., 2017, 2021), the `AMES-cond` grid (Allard et al., 2001), and a new grid built using the `Helios` code (Malik et al., 2017, 2019). A forest was trained on each of these grids, and then tests were run by predicting on the other grids, and on real spectra of the brown dwarfs GJ 570D and Epsilon Indi Ba and Bb.

The results showed that the different implementations of alkali line modelling leads to discrepancies between the grids, which has a drastic effect on the retrieval of the surface gravity. If the spectrum below 1.2 μm (where the alkali lines occur) is removed, the agreement in gravity is significantly improved. The retrieval of temperature is robust across all three models. Figure 7.2 demonstrates these findings.

For this study, I assisted M. Oreshenko with her implementation of the Random Forest. We worked closely with our collaborators at ARTORG on this project, meeting them on a frequent basis to discuss the results and improve the forest.

(a) Tested on `AMES-cond`, full spectral range   (b) Tested on `AMES-cond`, spectrum cut below 1.2 μm

(c) Tested on `Sonora`, full spectral range   (d) Tested on `Sonora`, spectrum cut below 1.2 μm
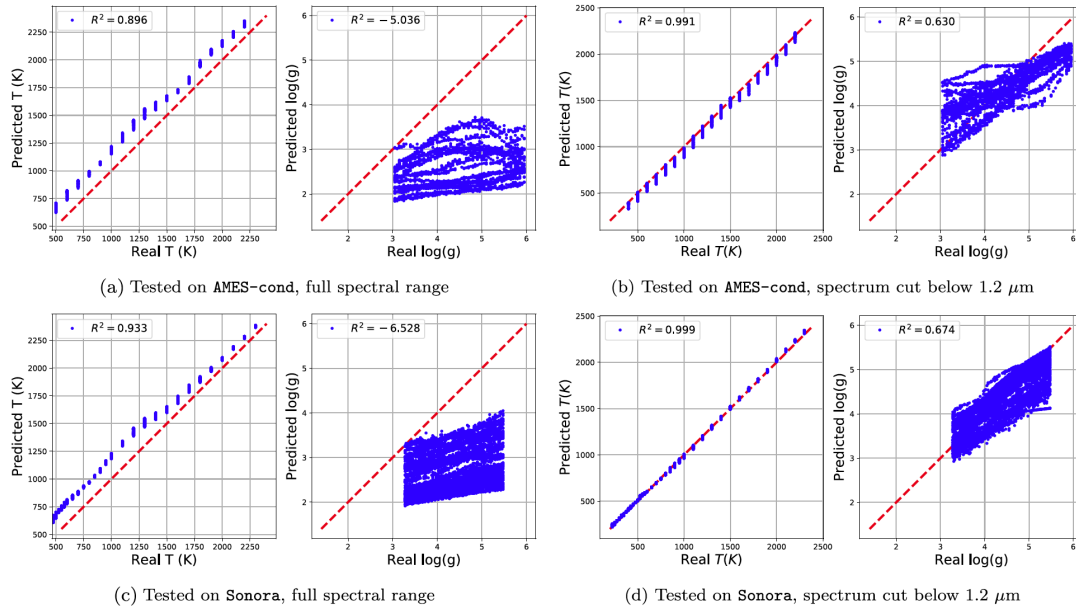
Figure 7.2: Testing the retrieval of temperature and gravity for the Random Forest trained on the `Helios` grid. The top and bottom rows show the forest tested on the `AMES-cond` and `Sonora` grids, respectively. The left- and right-hand panel pairs show the retrievals with and without the wavelength range below 1.2 μm, respectively.

7.4 Guzmán-Mesa et al. (2020)

Information Content of JWST NIRSpec Transmission Spectra of Warm Neptunes

The second study I supported using the Random Forest was this paper, which investigated the optimal JWST NIRSpec mode for studying warm Neptunes. Published 4 months before the James Webb cycle 1 deadline, this paper provided useful support for many observing proposals.

Due to the lower temperatures of these objects, warm Neptunes provide an exciting opportunity to study exoplanet atmospheres that may be in chemical disequilibrium. By training a forest on a set of atmospheric models using free chemistry, this study used the predictability and feature importance to determine which wavelength ranges provide the most information on the chemistry of these planets. The model included a temperature parameter, the abundances of 7 species (H_2O, HCN, NH_3, CO, CO_2, CH_4, and C_2H_2), and a non-grey cloud with 3 parameters.

The results showed that the G395M mode, covering 2.87-5.10 μm at low resolution, offers the best solution for studying warm Neptunes with JWST. This mode was able to retrieve the CO abundance, in particular, significantly better than the other modes. There exists a high-resolution version of this mode, although the results showed this offered diminishing returns. Figure 7.3 shows the R^2 score of each mode, for each parameter in the model.

During this project, I supported A. Guzmán Mesa in her use of the Random Forest. I taught her initially how to implement it, and then assisted her when problems arose along the way.
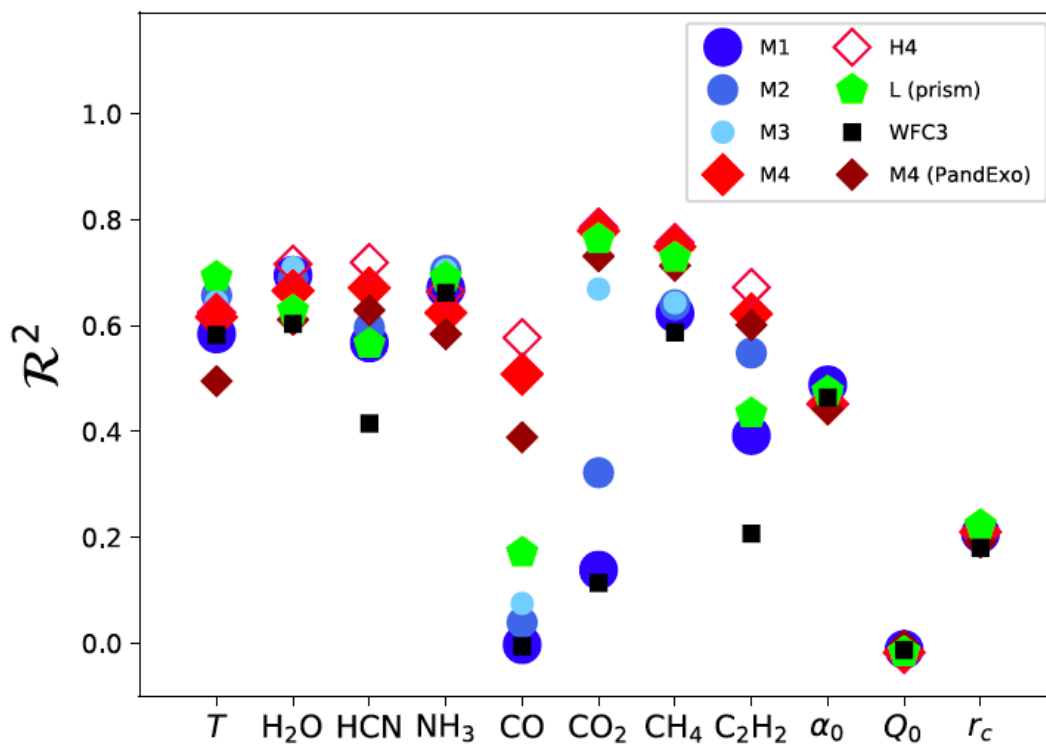
Figure 7.3: R^2 score of each model parameter, for each of the modes tested. Here "M4" refers to the G395M mode. H4 refers to the high-resolution version of M4. The performance of HST's WFC3 instrument is also included, for comparison.

7.5 Grimm et al. (2021)

HELIOS-K 2.0 Opacity Calculator and Open-source Opacity Database for Exoplanetary Atmospheres

This work presented an update to the opacity calculator HELIOS-K, developed by S. Grimm. As well as reporting on improvements to the code's efficiency and implementation, this study introduced a new opacity database (`https://dace.unige.ch/opacityDatabase/?`), where users can download and visualise opacities for a wide variety of atomic and molecular species.

As a frequent user of these opacities, I worked with S. Grimm over several years, testing and implementing the opacity files. I also contributed directly to editing the text.

7.6 Co-I on proposals

As well as supporting a number of scientific publications, I have also participated in several successful observing proposals, shown in Table 7.1. More details on the TOI-178 JWST proposal, led by M. Hooten, can be found in Chapter 6.

Year	Facility	PI	Program/objects
2019	HARPS	H. J. Hoeijmakers	Iron and titanium in WASP-189 b
2021	ESPRESSO	H. J. Hoeijmakers	Metals on the day-side of WASP-121 b
2021	JWST	M. Hooten	TOI-178: the best laboratory for testing planetary formation theories
2021	JWST	A. Rathcke	Probing the Terrestrial Planet TRAPPIST-1 c for the Presence of an Atmosphere
2021	JWST	N. Espinoza	The first near-infrared spectroscopic phase-curve of a super-Earth: K2-141 b
2021	JWST	N. Espinoza	Exploring the morning and evening limbs of a transiting exoplanet: WASP-63 b

Table 7.1: Table of successful observing proposals on which I am a co-investigator.

Conclusions and Future Outlook

We are just an advanced breed of monkeys on a minor planet of a very average star. But we can understand the Universe. That makes us something very special.

<div align="right">STEPHEN HAWKING</div>

8.1 Conclusions

I have spent four years working on my PhD in the Center for Space and Habitability at the University of Bern. In this time I have trained myself in the field of exoplanet characterisation and atmospheric retrieval, developed a new method of performing retrievals using machine learning, and established myself as a member of the worldwide exoplanet community through international collaborations.

In my first thesis paper (see Chapter 2), I wrote and applied my own atmospheric retrieval code to a set of 38 different exoplanets with spectra from Hubble's WFC3 instrument. By using a simplified analytical model for transmission spectra, shown

to be sufficient for this resolution and wavelength range (Heng and Kitzmann, 2017), I was able to run a suite of retrievals with varied parameters, using nested-sampling to perform a model comparison. The results showed that the preferred model for the majority of planets contained only water and a grey cloud, indicating that a wider wavelength coverage or higher resolutions are required to gain more information about these atmospheres. The retrieved parameters for the planets showed no clear evidence for any trends in the values. Since publishing, I have made the retrieval code open-source[1]. This paper has now received over 50 citations, and I have presented the work at a number of international conferences.

In 2019 I published my second thesis paper (see Chapter 3), in which I applied a non-LTE model of the sodium doublet to several datasets, and tested whether this effect could be detected through a series of mock retrievals. My results showed that the Bayesian evidence from the nested-sampling retrievals was unable to distinguish between LTE and non-LTE scenarios, even at high resolution. They also showed a bias towards lower temperatures when the LTE model was used. This work represents one of only a few papers considering non-LTE effects in exoplanet spectra.

In parallel to this work, I began collaborating with two machine learning experts at the ARTORG Center for Biomedical Engineering in Bern. Working closely with them, I helped to develop a new method using Random Forests to perform atmospheric retrieval on transmission spectra. This led to our publication in Nature Astronomy in 2018 (Márquez-Neila et al., 2018) (see Chapter 4). For this paper I generated the large training set of model spectra, on which the forest was trained and tested. I also performed traditional nested-sampling retrievals using the same atmospheric model for comparison. This paper has received over 30 citations, and has been influential in the field as one of only a handful of studies using machine learning for exoplanet analysis. These new techniques are paving the way for the future of exoplanet characterisation, when we predict to have an abundance of spectra from next-generation telescopes.

After the success of our Random Forest retrievals on low-resolution spectra, I decided to test it on the more challenging high-resolution spectra from ground-based spectrographs. This led to my last thesis paper (see Chapter 5). The initial tests applying the forest directly to the data proved unsuccessful, and the recent work on high-resolution retrievals using the cross-correlation function by Brogi

---

[1] https://github.com/exoclime/HELIOS-T

and Line (2019) inspired me to apply the forest to the CCFs instead. This enabled me to successfully retrieve the temperature and metallicity of wideband HARPS-N spectra. I then applied this method to the real HARPS-N data for KELT-9 b. The results showed an overabundance of Fe+ in the planet's atmosphere that could not be explained by the model, suggesting more complex physics needs to be included. I also performed various comparisons between the forest and traditional retrievals, as well as more advanced machine learning techniques such as neural networks. With approximately one citation per month since publishing, this paper remains one of the only studies to perform retrievals on high-resolution data. However, this field is gaining popularity, with more papers beginning to use these high-resolution retrievals. With the upcoming opening of the ELT and other next-generation ground-based observatories, retrieval of high-resolution data will play a key role in exoplanet characterisation.

In addition to my own papers, I also supported several studies throughout my PhD led by colleagues at Bern and around the world. In Seidel et al. (2019) I ran a retrieval on the sodium doublet in the HARPS spectrum of WASP-76 b. In Hoeijmakers et al. (2019) I supported the work on a spectral survey of KELT-9 b with discussions with the lead author on high-resolution retrievals and the implications of this work. In Oreshenko et al. (2020) I assisted the lead author with her implementation of the forest on several brown dwarf model grids, showing her how to use it and helping when problems arose. Similarly for Guzmán-Mesa et al. (2020), I taught the lead author how to run the forest and apply it to her dataset for an information content analysis of the different modes of JWST NIRSpec. In Grimm et al. (2021) I tested and implemented the opacities calculated by the HELIOS-K code, and helped with editing the text of the paper.

As well as working on scientific publications, I led and contributed to several observing proposals. I was a co-investigator on two successful ground-based proposals. The first used HARPS to observe the planet WASP-189 b. The second used ESPRESSO to look for metals on the dayside of WASP-121 b. I was also a co-investigator on four successful James Webb Cycle 1 proposals. For one of these, which will use JWST to observe several planets in the TOI-178 system, I contributed heavily, providing simulated spectra and mock retrievals of all the planets. In parallel, I worked on a similar proposal for which I was the principle investigator. This intended to observe two planets in the TOI-125 system with the NIRSpec instrument. Although this proposal was unsuccessful, I intend to develop it further and resubmit in a future cycle.

In conclusion, over four years of my PhD I have worked on a variety of different projects that have led to several publications, observing proposals, and open-source codes. I have attended a number of international conferences and presented my work to the community with a positive response. I have received interest in my work from international researchers, and secured an independent fellowship for the coming years. But most importantly, I have greatly enjoyed my years as a PhD student, and I look forward to what's in store for my time as a postdoc.

8.2 Future Outlook

In February 2021 I successfully applied for the SNSF Postdoc.Mobility Fellowship to spend two years undertaking independent research, hosted at the University of Oxford, UK. The research I plan to carry out is a continuation of my PhD work on using machine learning to analyse exoplanet atmospheres. However, I also intend to incorporate more advanced modelling techniques such as General Circulation Models, as well as developing the Random Forest for simultaneous analysis of high- and low-resolution spectra. A detailed research plan is explained in Section 8.2.1.

After receiving the fellowship acceptance, I postponed the start date by 9 months, to June 2022. This allows me to work on an exciting short-term project as a postdoc at the University of Bern. The project intends to study cold Haber worlds, which are hypothetical exoplanets where the Haber process for reacting hydrogen and nitrogen produces ammonia as a biosignature.

8.2.1 Fellowship Research Plan

The detailed characterisation of exoplanets, including those that may harbour life, is a core science goal of the next generation of major observatories, such as the James Webb Space Telescope (JWST) and the Extremely Large Telescope (ELT). These will lead to an explosion in the sensitivity and precision of exoplanet spectra, allowing us to study exoplanets as the three-dimensional, dynamic systems that they are. Current techniques comparing observations to models (known as atmospheric retrievals) rely on simplified 1-D models that will soon become insufficient. However, due to computational restrictions, traditional retrievals will be unable to cope with more complex physical models. During my PhD, I developed

a brand-new machine learning retrieval method, which uses a Random Forest to drastically speed up the process. I have proven this method to be effective on both low-resolution, spaced-based data (Márquez-Neila et al., 2018) and high-resolution, ground-based data (Fisher et al., 2020).

For the SNSF Postdoc.Mobility fellowship, I plan to use the Random Forest to perform a combined analysis on both high- and low-resolution data simultaneously, offering new insights that cannot be obtained from either alone. By undertaking this fellowship at Oxford, I will be able to work with experts on 3-D models of exoplanets, necessary for correctly interpreting the next generation of datasets. I will use my machine learning method to incorporate these into retrievals for the first time ever.

In the short term, my proposed projects will contribute towards improving our understanding of phenomena such as condensate formation and transport, molecular dissociation, and chemical disequilibrium in a wide range of exoplanets, from hot Jupiters to mini-Neptunes and super-Earths. In the longer term, this fellowship will help me to establish and disseminate the use of machine learning methods to interpret exoplanet spectra, and hone them to the point where they – and I – can play a key role in the search for signs of life beyond the solar system.

Project 1: Combining Ground- and Space-based Data with the Random Forest

The combination of ground- and space-based data in retrievals has recently been shown to give tighter constraints on the atmospheric parameters (Brogi et al., 2017). These precise constraints are important for accurately characterising atmospheres, and will be crucial in the future for determining habitability. However, due to computation time, current retrievals on high-resolution spectra are limited to narrow-band data (Brogi and Line, 2019). In my recent study (Fisher et al., 2020), I developed and tested a method that uses the Random Forest to perform atmospheric retrieval on wide-band data. **I propose to combine high- and low-resolution retrievals, and quantify the information from each dataset.**

I will use pre-computed, sophisticated models to form a training set. The forest enables us to also test multiple models from different research groups, and quanti-

tatively compare the results. The speed and efficiency of the Random Forest will allow us to easily combine data from various telescopes. By modelling a planet's properties, such as the temperature profile and atmospheric composition, we hope to learn not only the conditions of the planet, but also its potential formation location, and the diversity across the range of planets in our universe. High-resolution data provides us with individually resolved spectral lines, which are unobtainable from low-resolution spectra. This information helps us to constrain atmospheric chemistry, learn more about atmospheric dynamics, and even determine the vertical structure of the atmosphere (see Figure 8.1 for an outline of the project). In addition, I will perform an information content analysis with the Random Forest. This provides a unique opportunity to analyse which dataset is driving the retrieval for each parameter, which will be a valuable insight for the community. The speed of individual retrievals will also allow for statistical studies on certain types of exoplanets.

This project is incredibly timely with the launch of JWST in the coming year. I am currently involved in four successful Cycle 1 proposals, and intend to work on several proposals in Cycle 2, thus I hope to have access to some non-public data. Alternatively, the uniqueness of my method will enable me to extract information from public datasets, such as the ERS program, that other techniques cannot. My
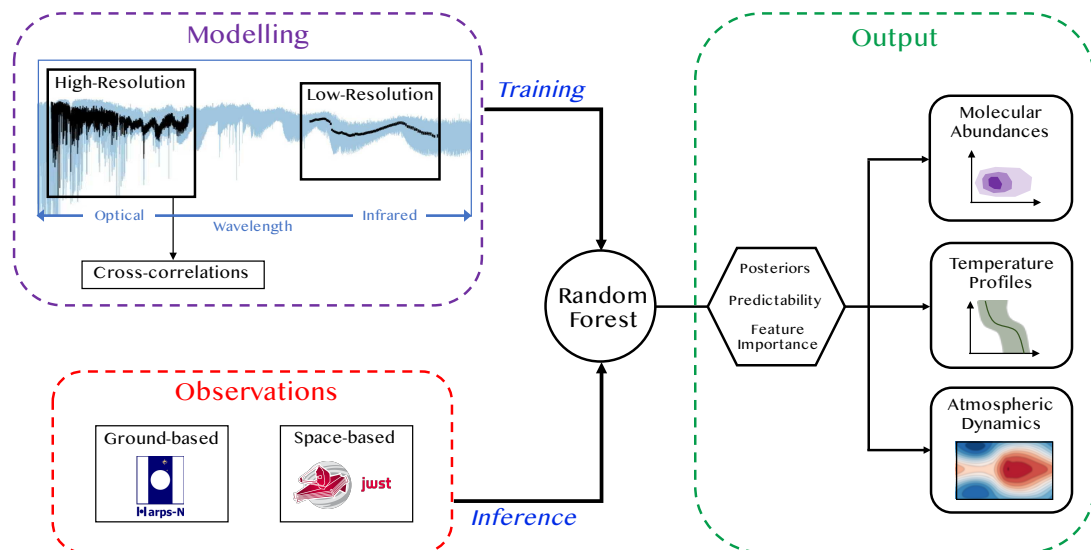
Figure 8.1: Schematic outlining the process of the retrieval on combined datasets using the Random Forest.

combined retrieval would allow for the most detailed analysis of the state-of-the-art data, and clearly evaluate the importance and limits of JWST. My method could also support proposals for telescope time, and even inform future telescope development. I will focus primarily on studying gas giants and mini-Neptunes with this method, as these provide excellent high-resolution data, and will be well-observed by JWST. These planets teach us about the diversity of planetary systems unlike our own, and can give us crucial information about planet formation.

I anticipate that this project will lead to a series of papers, first to demonstrate the method and subsequently to analyse data for individual objects, both now with HST data and later with JWST. These papers will focus on advising the community about how to optimise their observations, which will be imperative given the high demand for current and future telescope time. I also anticipate writing papers that perform statistical analyses on large groups of exoplanets, which will improve our knowledge of the bigger picture of exoplanets, and help put our solar system into the context of the rest of the universe.

With only a handful of people performing retrievals on high-resolution data, this work has the potential for a great number of collaborations across the world. The uniqueness of my technique and its many advantages will help promote me as an international exoplanet researcher. I am already involved in a collaboration with researchers at Lund and Geneva, analysing our new high-resolution spectra of WASP-189 b. Extending beyond the fellowship, this project will clearly position me as one of the key people in the field for jointly analysing ground- and space-based telescope data.

Project 2: Using GCM Outputs for Atmospheric Retrieval

With the improving quality of data, we require more detailed physics in the models used for retrieval. Previous studies have shown that using 1-D models will no longer be sufficient for analysing JWST data (e.g. Feng et al., 2016; Taylor et al., 2020), and high-resolution data already suggests the need for higher-dimensional effects (e.g. Wyttenbach et al., 2015). 3-D effects can be studied using General Circulation Models (GCMs), which solve for the complex geometry of tidally locked, irradiated exoplanets. However, their long computation time currently prohibits their direct use in atmospheric retrieval. **I propose to directly use GCM *outputs* in atmospheric retrievals for the first time ever, via the Random Forest.**

Since the forest does not require models to be calculated on-the-fly, we are able to compute these offline, making the use of GCMs feasible. I plan to identify the key parameters that significantly affect the global temperature structure of the atmosphere, then vary these parameters to construct a core set of ∼10-100 GCMs. I will then post-process these GCMs to produce ∼1000 members of the training set (transmission/emission spectra, phase curves, etc), by varying parameters such as the chemical abundances.

The use of GCMs in retrieval would allow us to analyse transmission and emission spectra, and phase curves consistently. These three datasets measure the terminator, dayside, and longitudinal variations of an exoplanet atmosphere. Since 1- and 2-D models cannot naturally relate these aspects of the atmosphere, additional parameters must be introduced into the models. GCMs allow us to automatically account for the complex geometry associated with these different datasets, resulting in fewer free parameters overall. By using GCMs to build our training models, we will no longer suffer from the biases associated with analysing a planet with a 1-D model. These 3-D retrievals will also give us an entirely new window into atmospheric circulation, which will complement previous analyses of phase curves. Again, the feature importance from the Random Forest will be able to highlight which datasets are most sensitive to the parameters in the model. This will indicate to the community which data to focus on for different cases. Ground-based, high-resolution data has already been used to detect dynamics on exoplanets (e.g. Wyttenbach et al., 2015; Seidel et al., 2019; Ehrenreich et al., 2020). Using GCMs in our retrievals would allow for an exhaustive analysis of these effects. The development of these next-generation methods is also an investment for the future, when data of potentially Earth-like exoplanets become available.

This project has huge potential for collaborations, as many groups across the world have their own GCMs with different assumptions and approximations. The Random Forest provides a great way to compare these models in the context of retrievals. I also anticipate a number of future studies combining well-measured phase curves with spectra and analysing them consistently, bringing in expertise from other aspects of exoplanet atmospheric research. In the future, this will lead to more cooperative projects for studying atmospheres, involving efforts from all corners of the field.

Schedule and Milestones

See Figure 8.2 for a detailed timeline of the two projects.
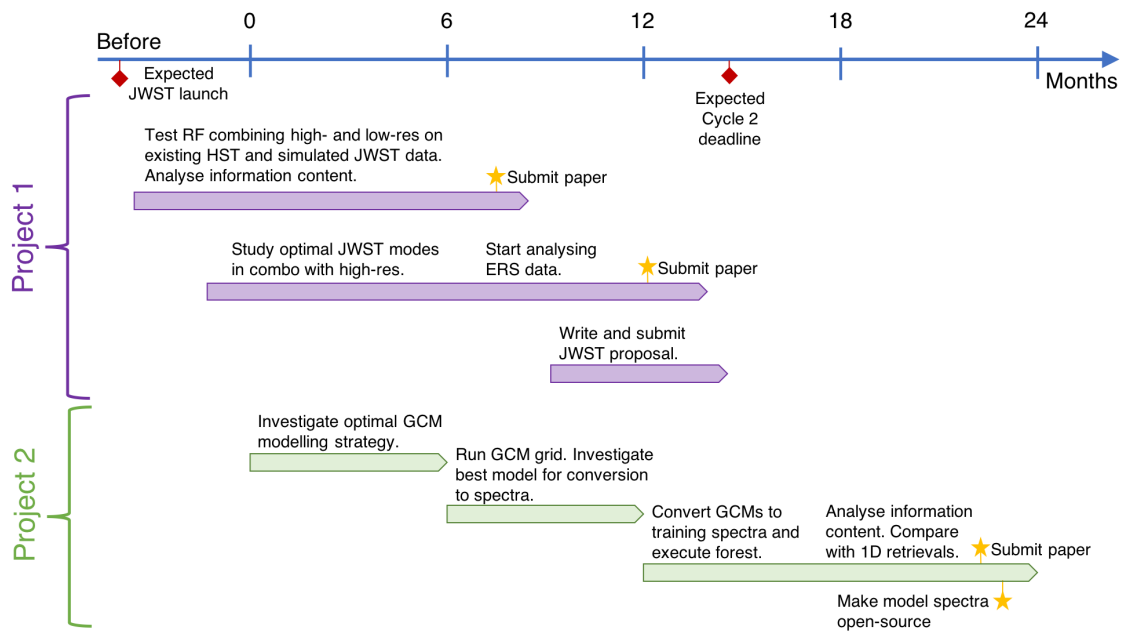
Figure 8.2: Timeline for SNF Mobility Fellowship projects.

Output **Project 1: 2 papers; Project 2: 1 paper and an open-source set of model spectra.**

Relevance and Impact

I predict that the impact of this work will be extremely significant in the field of exoplanet science. By allowing us to optimise our observations and extract the maximum information from our data, these methods will advance our understanding of exoplanetary atmospheres. With the launch of JWST this year, and the completion of the ELT in 2027, it will become increasingly imperative that this valuable observing time is used wisely. The studies from my first project will provide the community with the advice they require, and the necessary methods to analyse the state-of-the-art data. This work will also allow for the fast analysis of multiple planets, providing statistical comparisons and leading to the study of trends in these planets' properties.

The 3-D modelling project will provide the community with a more detailed insight into the structure and dynamics of an exoplanet. By making these models open-source, this will allow for researchers to improve their analysis of 3-D effects, essential for the new generation of telescopes. The conditions of these exoplanets will teach us about the bigger picture of the universe, by answering questions about planet formation, diverse chemical processes, and the uniqueness of our solar system (see Figure 8.3). Furthermore, by establishing the use of machine learning methods in exoplanet analysis, the work from my fellowship will open up the field to a whole new set of cutting-edge techniques. I will publish results from both my projects as articles in scientific journals (see schedule for details).
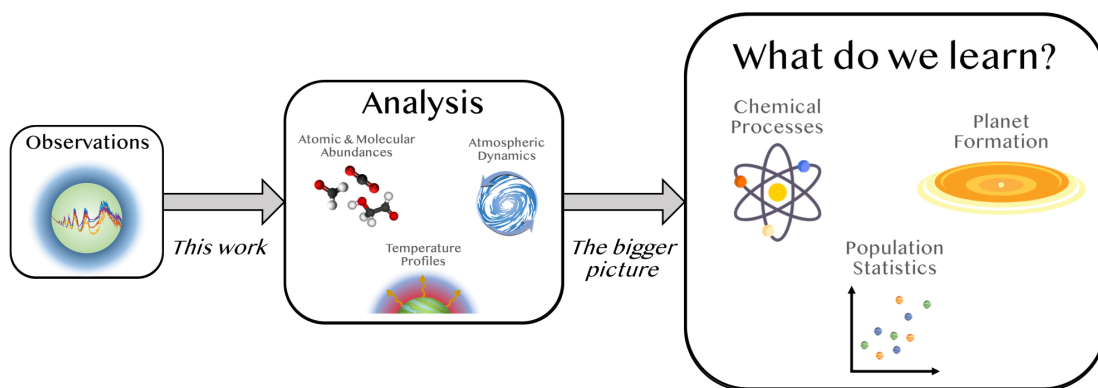
Figure 8.3: Schematic showing the placement of this work in the context of the bigger picture, and what we hope to learn from our results.

8.3 Epilogue

After an enjoyable and productive four years as a PhD student, I can only hope my postdoc years will be just as fruitful. I am excited for what the future holds for exoplanets, with the next-generation of telescopes just around the corner. The exoplanet community has been welcoming to me, and although the academic system has many flaws, I have felt very grateful to be able to work on a topic I have found so fascinating. Few people have the privilege to study the Universe, and I am lucky to be one of them.

CHAPTER 9

Acknowledgements

If you put your left-over pizza slices back in the box, you will forget about them. Then when you find them later, you have surprise pizza.

<div align="right">

Marko Sestovic

</div>

Firstly, I would like to thank Kevin Heng for supervising me the past four years. Your motivation, guidance, and support has been invaluable, and your confidence in me is greatly appreciated. I highly doubt I would've achieved what I have if it wasn't for you.

Secondly, I would like to thank Pablo Márquez-Neila and Raphael Sznitman, for collaborating with us. This work would not have been possible without your expertise in machine learning. You showed extreme patience with us, especially when explaining topics that must be so trivial to you. I have learnt so much.

Finally, I would like to thank my colleagues and friends at Bern, and back in England, who made these four years so much fun.

The Exoclimes Group, past and present

Jens Hoeijmakers, for providing advice, both scientific and otherwise. For teaching me about the wonders of observations. And for making me laugh. I hope you return to the blue lagoon one day.

Daniel Kitzmann, for always knowing the answer to my questions, and being the expert on everything theoretical. And for the best homemade cakes in Bern.

Andrea Guzmán Mesa, for being so entertaining, and generally a great friend. The office is never dull when you're around.

Kathryn Jones, for being so funny and reminding me of home. See you on the sloops.

Sinan Li, for being so kind and making me smile.

Anna Lüber, for bringing enthusiasm back to the office.

Brett Morris (and Brienne), for providing advice and time with the best doggo in town.

Russell Deitrick, for the best bad jokes.

Simon Grimm, for a wealth of wisdom, and great stories about life growing up in Graubunden.

Kaustubh Hakim, for funny conversation, and tips on the best indian food around.

Elsie Lee, for having the brightest office on our corridor, and next-level coffee.

Shami Tsai, for entertaining boardgame nights and PhD advice.

Matej Malik, for never taking things too seriously.

Maria Oreshenko, for great ice cream, and all the fresh air in the office.

And others

Brice-Olivier Demory, for giving great advice, and for the unforgettable flight around the Matterhorn.

Dan Bower, for proposal advice and mountain tips.

Tobias Meier, for the weekend hikes and movie nights. Team hearts.

Apurva Oza, suh dude, for Bhangra lessons and wild nights out.

Marko Sestovic... There are no words.

And special thanks to Triple-T for the hiking weekends, and for the great company who joined.

Anika, for the great day trips, fun nights out, and for being the best flatmate and friend I could've asked for.

All my family, especially my mum, for supporting and encouraging me throughout the last four years. I am extremely lucky to have you all.

And Rob, for always being there for me, and making the distance bearable. For coming to visit so many times, and putting up with me. For all your support. I don't know if I could've have done it without you, but I do know it would've been a lot less fun.

~

❊ DEDICATION ❊

I would like to dedicate this thesis to my dad, who passed away in May 2021. I wouldn't be who I am today if it wasn't for him. He was an incredible father, supporter, and friend. He is greatly missed.

There's a starman waiting in the sky
He'd like to come and meet us
But he thinks he'd blow our minds
There's a starman waiting in the sky
He's told us not to blow it
'Cause he knows it's all worthwhile
He told me
Let the children lose it
Let the children use it
Let all the children boogie

DAVID BOWIE

Bibliography

Ackerman, A. S. and Marley, M. S. (2001). Precipitating Condensation Clouds in Substellar Atmospheres. *ApJ*, 556(2):872–884.

Aigrain, S., Parviainen, H., and Pope, B. J. S. (2016). K2SC: flexible systematics correction and detrending of K2 light curves using Gaussian process regression. *MNRAS*, 459(3):2408–2419.

Alibert, Y., Carron, F., Fortier, A., Pfyffer, S., Benz, W., Mordasini, C., and Swoboda, D. (2013). Theoretical models of planetary system formation: mass vs. semi-major axis. *A&A*, 558:A109.

Allard, F., Hauschildt, P. H., Alexander, D. R., Tamanai, A., and Schweitzer, A. (2001). The Limiting Effects of Dust in Brown Dwarf Model Atmospheres. *ApJ*, 556(1):357–372.

Allart, R., Lovis, C., Pino, L., Wyttenbach, A., Ehrenreich, D., and Pepe, F. (2017). Search for water vapor in the high-resolution transmission spectrum of HD 189733b in the visible. *A&A*, 606:A144.

ALMA Partnership, Brogan, C. L., Pérez, L. M., Hunter, T. R., Dent, W. R. F., Hales, A. S., Hills, R. E., Corder, S., Fomalont, E. B., Vlahakis, C., Asaki, Y., Barkats, D., Hirota, A., Hodge, J. A., Impellizzeri, C. M. V., Kneissl, R., Liuzzo, E., Lucas, R., Marcelino, N., Matsushita, S., Nakanishi, K., Phillips, N., Richards, A. M. S., Toledo, I., Aladro, R., Broguiere, D., Cortes, J. R., Cortes, P. C., Espada, D., Galarza, F., Garcia-Appadoo, D., Guzman-Ramirez, L., Humphreys, E. M., Jung, T., Kameno, S., Laing, R. A., Leon, S., Marconi, G., Mignano, A., Nikolic, B., Nyman, L. A., Radiszcz, M., Remijan, A., Rodón, J. A., Sawada, T., Takahashi, S., Tilanus, R. P. J., Vila Vilaro, B., Watson, L. C., Wiklind, T., Akiyama, E., Chapillon, E., de Gregorio-Monsalvo, I., Di Francesco, J., Gueth, F., Kawamura, A., Lee,

C. F., Nguyen Luong, Q., Mangum, J., Pietu, V., Sanhueza, P., Saigo, K., Takakuwa, S., Ubach, C., van Kempen, T., Wootten, A., Castro-Carrizo, A., Francke, H., Gallardo, J., Garcia, J., Gonzalez, S., Hill, T., Kaminski, T., Kurono, Y., Liu, H. Y., Lopez, C., Morales, F., Plarre, K., Schieven, G., Testi, L., Videla, L., Villard, E., Andreani, P., Hibbard, J. E., and Tatematsu, K. (2015). The 2014 ALMA Long Baseline Campaign: First Results from High Angular Resolution Observations toward the HL Tau Region. *ApJL*, 808(1):L3.

Ansdell, M., Ioannou, Y., Osborn, H. P., Sasdelli, M., 2018 NASA Frontier Development Lab Exoplanet Team, Smith, J. C., Caldwell, D., Jenkins, J. M., Räissi, C., Angerhausen, D., and NASA Frontier Development Lab Exoplanet Mentors, . (2018). Scientific Domain Knowledge Improves Exoplanet Transit Classification with Deep Learning. *ApJL*, 869(1):L7.

Armstrong, D. J., Gamper, J., and Damoulas, T. (2021). Exoplanet validation with machine learning: 50 new validated Kepler planets. *MNRAS*, 504(4):5327–5344.

Armstrong, D. J., Günther, M. N., McCormac, J., Smith, A. M. S., Bayliss, D., Bouchy, F., Burleigh, M. R., Casewell, S., Eigmüller, P., Gillen, E., Goad, M. R., Hodgkin, S. T., Jenkins, J. S., Louden, T., Metrailler, L., Pollacco, D., Poppenhaeger, K., Queloz, D., Raynard, L., Rauer, H., Udry, S., Walker, S. R., Watson, C. A., West, R. G., and Wheatley, P. J. (2018). Automatic vetting of planet candidates from ground-based surveys: machine learning with NGTS. *MNRAS*, 478(3):4225–4237.

Armstrong, D. J., Pollacco, D., and Santerne, A. (2017). Transit shapes and self-organizing maps as a tool for ranking planetary candidates: application to Kepler and K2. *MNRAS*, 465(3):2634–2642.

Baglin, A., Auvergne, M., Barge, P., Buey, J. T., Catala, C., Michel, E., Weiss, W., and COROT Team (2002). COROT: asteroseismology and planet finding. In Battrick, B., Favata, F., Roxburgh, I. W., and Galadi, D., editors, *Stellar Structure and Habitable Planet Finding*, volume 485 of *ESA Special Publication*, pages 17–24.

Ballard, S., Fabrycky, D., Fressin, F., Charbonneau, D., Desert, J.-M., Torres, G., Marcy, G., Burke, C. J., Isaacson, H., Henze, C., Steffen, J. H., Ciardi, D. R., Howell, S. B., Cochran, W. D., Endl, M., Bryson, S. T., Rowe, J. F., Holman, M. J., Lissauer, J. J., Jenkins, J. M., Still, M., Ford, E. B., Christiansen, J. L., Middour, C. K., Haas, M. R., Li, J., Hall, J. R., McCauliff, S., Batalha, N. M., Koch, D. G., and Borucki, W. J. (2011). The Kepler-19 System: A Transiting 2.2 R$_{\oplus}$ Planet and a Second Planet Detected via Transit Timing Variations. *ApJ*, 743(2):200.

Baranne, A., Queloz, D., Mayor, M., Adrianzyk, G., Knispel, G., Kohler, D., Lacroix, D., Meunier, J. P., Rimbaud, G., and Vin, A. (1996). ELODIE: A spectrograph for accurate radial velocity measurements. *A&AS*, 119:373–390.

Barstow, J. K. (2020). Unveiling cloudy exoplanets: the influence of cloud model choices on retrieval solutions. *MNRAS*, 497(4):4183–4195.

Barstow, J. K., Aigrain, S., Irwin, P. G. J., Fletcher, L. N., and Lee, J. M. (2013). Constraining

the atmosphere of GJ 1214b using an optimal estimation technique. *MNRAS*, 434(3):2616–2628.

Barstow, J. K., Aigrain, S., Irwin, P. G. J., Kendrew, S., and Fletcher, L. N. (2015). Transit spectroscopy with James Webb Space Telescope: systematics, starspots and stitching. *MNRAS*, 448(3):2546–2561.

Barstow, J. K., Aigrain, S., Irwin, P. G. J., and Sing, D. K. (2017). A Consistent Retrieval Analysis of 10 Hot Jupiters Observed in Transmission. *ApJ*, 834(1):50.

Barstow, J. K., Changeat, Q., Garland, R., Line, M. R., Rocchetto, M., and Waldmann, I. P. (2020). A comparison of exoplanet spectroscopic retrieval tools. *MNRAS*, 493(4):4884–4909.

Barstow, J. K. and Heng, K. (2020). Outstanding Challenges of Exoplanet Atmospheric Retrievals. *SSRv*, 216(5):82.

Batalha, N., Kalirai, J., Lunine, J., Clampin, M., and Lindler, D. (2015). Transiting Exoplanet Simulations with the James Webb Space Telescope. *arXiv e-prints*, page arXiv:1507.02655.

Batalha, N. E., Mandell, A., Pontoppidan, K., Stevenson, K. B., Lewis, N. K., Kalirai, J., Earl, N., Greene, T., Albert, L., and Nielsen, L. D. (2017). PandExo: A Community Tool for Transiting Exoplanet Science with JWST & HST. *PASP*, 129(976):064501.

Batalha, N. M. (2014). Exploring exoplanet populations with NASA's Kepler Mission. *Proceedings of the National Academy of Science*, 111(35):12647–12654.

Beichman, C., Benneke, B., Knutson, H., Smith, R., Lagage, P.-O., Dressing, C., Latham, D., Lunine, J., Birkmann, S., Ferruit, P., Giardino, G., Kempton, E., Carey, S., Krick, J., Deroo, P. D., Mandell, A., Ressler, M. E., Shporer, A., Swain, M., Vasisht, G., Ricker, G., Bouwman, J., Crossfield, I., Greene, T., Howell, S., Christiansen, J., Ciardi, D., Clampin, M., Greenhouse, M., Sozzetti, A., Goudfrooij, P., Hines, D., Keyes, T., Lee, J., McCullough, P., Robberto, M., Stansberry, J., Valenti, J., Rieke, M., Rieke, G., Fortney, J., Bean, J., Kreidberg, L., Ehrenreich, D., Deming, D., Albert, L., Doyon, R., and Sing, D. (2014). Observations of Transiting Exoplanets with the James Webb Space Telescope (JWST). *PASP*, 126(946):1134.

Ben-Yami, M., Madhusudhan, N., Cabot, S. H. C., Constantinou, S., Piette, A., Gandhi, S., and Welbanks, L. (2020). Neutral Cr and V in the Atmosphere of Ultra-hot Jupiter WASP-121 b. *ApJL*, 897(1):L5.

Benneke, B., Knutson, H. A., Lothringer, J., Crossfield, I. J. M., Moses, J. I., Morley, C., Kreidberg, L., Fulton, B. J., Dragomir, D., Howard, A. W., Wong, I., Désert, J.-M., McCullough, P. R., Kempton, E. M. R., Fortney, J., Gilliland , R., Deming, D., and Kammer, J. (2019a). A sub-Neptune exoplanet with a low-metallicity methane-depleted atmosphere and Mie-scattering clouds. *Nature Astronomy*, 3:813–821.

Benneke, B. and Seager, S. (2013). How to Distinguish between Cloudy Mini-Neptunes and Water/Volatile-dominated Super-Earths. *ApJ*, 778(2):153.

Benneke, B., Wong, I., Piaulet, C., Knutson, H. A., Lothringer, J., Morley, C. V., Crossfield, I. J. M., Gao, P., Greene, T. P., Dressing, C., Dragomir, D., Howard, A. W., McCullough, P. R., Kempton, E. M. R., Fortney, J. J., and Fraine, J. (2019b). Water Vapor and Clouds on the Habitable-zone Sub-Neptune Exoplanet K2-18b. *ApJL*, 887(1):L14.

Birkby, J. L., de Kok, R. J., Brogi, M., de Mooij, E. J. W., Schwarz, H., Albrecht, S., and Snellen, I. A. G. (2013). Detection of water absorption in the day side atmosphere of HD 189733 b using ground-based high-resolution spectroscopy at 3.2μm. *MNRAS*, 436:L35–L39.

Birkby, J. L., de Kok, R. J., Brogi, M., Schwarz, H., and Snellen, I. A. G. (2017). Discovery of Water at High Spectral Resolution in the Atmosphere of 51 Peg b. *AJ*, 153(3):138.

Blain, D., Charnay, B., and Bézard, B. (2021). 1D atmospheric study of the temperate sub-Neptune K2-18b. *A&A*, 646:A15.

Borucki, W., Koch, D., Batalha, N., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., Dunham, E., Gautier, T. N., Geary, J., Gilliland, R., Jenkins, J., Kjeldsen, H., Lissauer, J. J., and Rowe, J. (2009). KEPLER: Search for Earth-Size Planets in the Habitable Zone. In Pont, F., Sasselov, D., and Holman, M. J., editors, *Transiting Planets*, volume 253, pages 289–299.

Borucki, W. J. and Summers, A. L. (1984). The photometric method of detecting other planetary systems. *Icarus*, 58(1):121–134.

Brandl, B., Bettonvil, F., van Boekel, R., Glauser, A., Quanz, S., Absil, O., Amorim, A., Feldt, M., Glasse, A., Güdel, M., Ho, P., Labadie, L., Meyer, M., Pantin, E., van Winckel, H., and METIS Consortium (2021). METIS: The Mid-infrared ELT Imager and Spectrograph. *The Messenger*, 182:22–26.

Brogi, M., de Kok, R. J., Albrecht, S., Snellen, I. A. G., Birkby, J. L., and Schwarz, H. (2016). Rotation and Winds of Exoplanet HD 189733 b Measured with High-dispersion Transmission Spectroscopy. *ApJ*, 817(2):106.

Brogi, M., Line, M., Bean, J., Désert, J. M., and Schwarz, H. (2017). A Framework to Combine Low- and High-resolution Spectroscopy for the Atmospheres of Transiting Exoplanets. *ApJL*, 839(1):L2.

Brogi, M. and Line, M. R. (2019). Retrieving Temperatures and Abundances of Exoplanet Atmospheres with High-resolution Cross-correlation Spectroscopy. *AJ*, 157(3):114.

Brogi, M., Snellen, I. A. G., de Kok, R. J., Albrecht, S., Birkby, J., and de Mooij, E. J. W. (2012). The signature of orbital motion from the dayside of the planet τ Boötis b. *Nature*, 486(7404):502–504.

Buchner, J., Georgakakis, A., Nandra, K., Hsu, L., Rangel, C., Brightman, M., Merloni, A., Salvato, M., Donley, J., and Kocevski, D. (2014). X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *A&A*, 564:A125.

Burn, R., Marboeuf, U., Alibert, Y., and Benz, W. (2019). Radial drift and concurrent ablation of boulder-sized objects. *A&A*, 629:A64.

Caceres, G. A., Feigelson, E. D., Jogesh Babu, G., Bahamonde, N., Christen, A., Bertin, K., Meza, C., and Curé, M. (2019). Autoregressive Planet Search: Application to the Kepler Mission. *AJ*, 158(2):58.

Casasayas-Barris, N., Pallé, E., Yan, F., Chen, G., Albrecht, S., Nortmann, L., Van Eylen, V., Snellen, I., Talens, G. J. J., González Hernández, J. I., Rebolo, R., and Otten, G. P. P. L. (2018). Na I and Hα absorption features in the atmosphere of MASCARA-2b/KELT-20b. *A&A*, 616:A151.

Casasayas-Barris, N., Pallé, E., Yan, F., Chen, G., Kohl, S., Stangret, M., Parviainen, H., Helling, C., Watanabe, N., Czesla, S., Fukui, A., Montañés-Rodríguez, P., Nagel, E., Narita, N., Nortmann, L., Nowak, G., Schmitt, J. H. M. M., and Zapatero Osorio, M. R. (2019). Atmospheric characterization of the ultra-hot Jupiter MASCARA-2b/KELT-20b. Detection of CaII, FeII, NaI, and the Balmer series of H (Hα, Hβ, and Hγ) with high-dispersion transit spectroscopy. *A&A*, 628:A9.

Charbonneau, D., Allen, L. E., Megeath, S. T., Torres, G., Alonso, R., Brown, T. M., Gilliland, R. L., Latham, D. W., Mandushev, G., O'Donovan, F. T., and Sozzetti, A. (2005). Detection of Thermal Emission from an Extrasolar Planet. *ApJ*, 626(1):523–529.

Charbonneau, D., Brown, T. M., Latham, D. W., and Mayor, M. (2000). Detection of Planetary Transits Across a Sun-like Star. *ApJL*, 529(1):L45–L48.

Charbonneau, D., Brown, T. M., Noyes, R. W., and Gilliland, R. L. (2002). Detection of an Extrasolar Planet Atmosphere. *ApJ*, 568(1):377–384.

Chaushev, A., Raynard, L., Goad, M. R., Eigmüller, P., Armstrong, D. J., Briegal, J. T., Burleigh, M. R., Casewell, S. L., Gill, S., Jenkins, J. S., Nielsen, L. D., Watson, C. A., West, R. G., Wheatley, P. J., Udry, S., and Vines, J. I. (2019). Classifying exoplanet candidates with convolutional neural networks: application to the Next Generation Transit Survey. *MNRAS*, 488(4):5232–5250.

Chubb, K. L., Rocchetto, M., Yurchenko, S. N., Min, M., Waldmann, I., Barstow, J. K., Mollière, P., Al-Refaie, A. F., Phillips, M. W., and Tennyson, J. (2021). The ExoMolOP database: Cross sections and k-tables for molecules of interest in high-temperature exoplanet atmospheres. *A&A*, 646:A21.

Cobb, A. D., Himes, M. D., Soboczenski, F., Zorzan, S., O'Beirne, M. D., Güneş Baydin, A., Gal, Y., Domagal-Goldman, S. D., Arney, G. N., Angerhausen, D., and 2018 NASA FDL Astrobiology Team, I. (2019). An Ensemble of Bayesian Neural Networks for Exoplanetary Atmospheric Retrieval. *AJ*, 158(1):33.

Cowan, N. B., Greene, T., Angerhausen, D., Batalha, N. E., Clampin, M., Colón, K., Crossfield, I. J. M., Fortney, J. J., Gaudi, B. S., Harrington, J., Iro, N., Lillie, C. F., Linsky, J. L.,

Lopez-Morales, M., Mandell, A. M., and Stevenson, K. B. (2015). Characterizing Transiting Planet Atmospheres through 2025. *PASP*, 127(949):311.

Crouzet, N., McCullough, P. R., Deming, D., and Madhusudhan, N. (2014). Water Vapor in the Spectrum of the Extrasolar Planet HD 189733b. II. The Eclipse. *ApJ*, 795(2):166.

Dattilo, A., Vanderburg, A., Shallue, C. J., Mayo, A. W., Berlind, P., Bieryla, A., Calkins, M. L., Esquerdo, G. A., Everett, M. E., Howell, S. B., Latham, D. W., Scott, N. J., and Yu, L. (2019). Identifying Exoplanets with Deep Learning. II. Two New Super-Earths Uncovered by a Neural Network in K2 Data. *AJ*, 157(5):169.

de Kok, R. J., Brogi, M., Snellen, I. A. G., Birkby, J., Albrecht, S., and de Mooij, E. J. W. (2013). Detection of carbon monoxide in the high-resolution day-side spectrum of the exoplanet HD 189733b. *A&A*, 554:A82.

de Wit, J., Wakeford, H. R., Lewis, N. K., Delrez, L., Gillon, M., Selsis, F., Leconte, J., Demory, B.-O., Bolmont, E., Bourrier, V., Burgasser, A. J., Grimm, S., Jehin, E., Lederer, S. M., Owen, J. E., Stamenković, V., and Triaud, A. H. M. J. (2018). Atmospheric reconnaissance of the habitable-zone Earth-sized planets orbiting TRAPPIST-1. *Nature Astronomy*, 2:214–219.

Deming, D., Wilkins, A., McCullough, P., Burrows, A., Fortney, J. J., Agol, E., Dobbs-Dixon, I., Madhusudhan, N., Crouzet, N., Desert, J.-M., Gilliland, R. L., Haynes, K., Knutson, H. A., Line, M., Magic, Z., Mandell, A. M., Ranjan, S., Charbonneau, D., Clampin, M., Seager, S., and Showman, A. P. (2013). Infrared Transmission Spectroscopy of the Exoplanets HD 209458b and XO-1b Using the Wide Field Camera-3 on the Hubble Space Telescope. *ApJ*, 774(2):95.

Deming, L. D. and Seager, S. (2017). Illusion and reality in the atmospheres of exoplanets. *Journal of Geophysical Research (Planets)*, 122(1):53–75.

Dittmann, J. A., Irwin, J. M., Charbonneau, D., Bonfils, X., Astudillo-Defru, N., Haywood, R. D., Berta-Thompson, Z. K., Newton, E. R., Rodriguez, J. E., Winters, J. G., Tan, T.-G., Almenara, J.-M., Bouchy, F., Delfosse, X., Forveille, T., Lovis, C., Murgas, F., Pepe, F., Santos, N. C., Udry, S., Wünsche, A., Esquerdo, G. A., Latham, D. W., and Dressing, C. D. (2017). A temperate rocky super-Earth transiting a nearby cool star. *Nature*, 544(7650):333–336.

Ehrenreich, D., Lovis, C., Allart, R., Zapatero Osorio, M. R., Pepe, F., Cristiani, S., Rebolo, R., Santos, N. C., Borsa, F., Demangeon, O., Dumusque, X., González Hernández, J. I., Casasayas-Barris, N., Ségransan, D., Sousa, S., Abreu, M., Adibekyan, V., Affolter, M., Allende Prieto, C., Alibert, Y., Aliverti, M., Alves, D., Amate, M., Avila, G., Baldini, V., Bandy, T., Benz, W., Bianco, A., Bolmont, É., Bouchy, F., Bourrier, V., Broeg, C., Cabral, A., Calderone, G., Pallé, E., Cegla, H. M., Cirami, R., Coelho, J. M. P., Conconi, P., Coretti, I., Cumani, C., Cupani, G., Dekker, H., Delabre, B., Deiries, S., D'Odorico, V., Di Marcantonio, P., Figueira, P., Fragoso, A., Genolet, L., Genoni, M., Génova Santos, R., Hara, N., Hughes, I., Iwert, O., Kerber, F., Knudstrup, J., Landoni, M., Lavie, B., Lizon, J.-L., Lendl,

M., Lo Curto, G., Maire, C., Manescau, A., Martins, C. J. A. P., Mégevand, D., Mehner, A., Micela, G., Modigliani, A., Molaro, P., Monteiro, M., Monteiro, M., Moschetti, M., Müller, E., Nunes, N., Oggioni, L., Oliveira, A., Pariani, G., Pasquini, L., Poretti, E., Rasilla, J. L., Redaelli, E., Riva, M., Santana Tschudi, S., Santin, P., Santos, P., Segovia Milla, A., Seidel, J. V., Sosnowska, D., Sozzetti, A., Spanò, P., Suárez Mascareño, A., Tabernero, H., Tenegi, F., Udry, S., Zanutta, A., and Zerbi, F. (2020). Nightside condensation of iron in an ultrahot giant exoplanet. *Nature*, 580(7805):597–601.

Emsenhuber, A., Mordasini, C., Burn, R., Alibert, Y., Benz, W., and Asphaug, E. (2020). The New Generation Planetary Population Synthesis (NGPPS). I. Bern global model of planet formation and evolution, model tests, and emerging planetary systems. *arXiv e-prints*, page arXiv:2007.05561.

Evans, T. M., Sing, D. K., Kataria, T., Goyal, J., Nikolov, N., Wakeford, H. R., Deming, D., Marley, M. S., Amundsen, D. S., Ballester, G. E., Barstow, J. K., Ben-Jaffel, L., Bourrier, V., Buchhave, L. A., Cohen, O., Ehrenreich, D., García Muñoz, A., Henry, G. W., Knutson, H., Lavvas, P., Lecavelier Des Etangs, A., Lewis, N. K., López-Morales, M., Mandell, A. M., Sanz-Forcada, J., Tremblin, P., and Lupu, R. (2017). An ultrahot gas-giant exoplanet with a stratosphere. *Nature*, 548(7665):58–61.

Feng, Y. K., Line, M. R., Fortney, J. J., Stevenson, K. B., Bean, J., Kreidberg, L., and Parmentier, V. (2016). The Impact of Non-uniform Thermal Structure on the Interpretation of Exoplanet Emission Spectra. *ApJ*, 829(1):52.

Feroz, F. and Hobson, M. P. (2008). Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *MNRAS*, 384(2):449–463.

Feroz, F., Hobson, M. P., and Bridges, M. (2009). MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS*, 398(4):1601–1614.

Fisher, C. and Heng, K. (2018). Retrieval analysis of 38 WFC3 transmission spectra and resolution of the normalization degeneracy. *MNRAS*, 481(4):4698–4727.

Fisher, C. and Heng, K. (2019). How Much Information Does the Sodium Doublet Encode? Retrieval Analysis of Non-LTE Sodium Lines at Low and High Spectral Resolutions. *ApJ*, 881(1):25.

Fisher, C., Hoeijmakers, H. J., Kitzmann, D., Márquez-Neila, P., Grimm, S. L., Sznitman, R., and Heng, K. (2020). Interpreting High-resolution Spectroscopy of Exoplanets using Cross-correlations and Supervised Machine Learning. *AJ*, 159(5):192.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: The mcmc hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306–312.

Foreman-Mackey, D., Hogg, D. W., and Morton, T. D. (2014). Exoplanet Population Inference and the Abundance of Earth Analogs from Noisy, Incomplete Catalogs. *ApJ*, 795(1):64.

Fortney, J. J., Shabram, M., Showman, A. P., Lian, Y., Freedman, R. S., Marley, M. S., and Lewis, N. K. (2010). Transmission Spectra of Three-Dimensional Hot Jupiter Model Atmospheres. *ApJ*, 709(2):1396–1406.

Fortney, J. J., Sudarsky, D., Hubeny, I., Cooper, C. S., Hubbard, W. B., Burrows, A., and Lunine, J. I. (2003). On the Indirect Detection of Sodium in the Atmosphere of the Planetary Companion to HD 209458. *ApJ*, 589(1):615–622.

Fraine, J., Deming, D., Benneke, B., Knutson, H., Jordán, A., Espinoza, N., Madhusudhan, N., Wilkins, A., and Todorov, K. (2014). Water vapour absorption in the clear atmosphere of a Neptune-sized exoplanet. *Nature*, 513(7519):526–529.

Fu, G., Deming, D., Knutson, H., Madhusudhan, N., Mandell, A., and Fraine, J. (2017). Statistical Analysis of Hubble/WFC3 Transit Spectroscopy of Extrasolar Planets. *ApJL*, 847(2):L22.

Fulton, B. J. and Petigura, E. A. (2018). The California-Kepler Survey. VII. Precise Planet Radii Leveraging Gaia DR2 Reveal the Stellar Mass Dependence of the Planet Radius Gap. *AJ*, 156(6):264.

Fulton, B. J., Petigura, E. A., Howard, A. W., Isaacson, H., Marcy, G. W., Cargile, P. A., Hebb, L., Weiss, L. M., Johnson, J. A., Morton, T. D., Sinukoff, E., Crossfield, I. J. M., and Hirsch, L. A. (2017). The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets. *AJ*, 154(3):109.

Gaudi, B. S., Seager, S., Mennesson, B., Kiessling, A., Warfield, K., Cahoy, K., Clarke, J. T., Domagal-Goldman, S., Feinberg, L., Guyon, O., Kasdin, J., Mawet, D., Plavchan, P., Robinson, T., Rogers, L., Scowen, P., Somerville, R., Stapelfeldt, K., Stark, C., Stern, D., Turnbull, M., Amini, R., Kuan, G., Martin, S., Morgan, R., Redding, D., Stahl, H. P., Webb, R., Alvarez-Salazar, O., Arnold, W. L., Arya, M., Balasubramanian, B., Baysinger, M., Bell, R., Below, C., Benson, J., Blais, L., Booth, J., Bourgeois, R., Bradford, C., Brewer, A., Brooks, T., Cady, E., Caldwell, M., Calvet, R., Carr, S., Chan, D., Cormarkovic, V., Coste, K., Cox, C., Danner, R., Davis, J., Dewell, L., Dorsett, L., Dunn, D., East, M., Effinger, M., Eng, R., Freebury, G., Garcia, J., Gaskin, J., Greene, S., Hennessy, J., Hilgemann, E., Hood, B., Holota, W., Howe, S., Huang, P., Hull, T., Hunt, R., Hurd, K., Johnson, S., Kissil, A., Knight, B., Kolenz, D., Kraus, O., Krist, J., Li, M., Lisman, D., Mandic, M., Mann, J., Marchen, L., Marrese-Reading, C., McCready, J., McGown, J., Missun, J., Miyaguchi, A., Moore, B., Nemati, B., Nikzad, S., Nissen, J., Novicki, M., Perrine, T., Pineda, C., Polanco, O., Putnam, D., Qureshi, A., Richards, M., Eldorado Riggs, A. J., Rodgers, M., Rud, M., Saini, N., Scalisi, D., Scharf, D., Schulz, K., Serabyn, G., Sigrist, N., Sikkia, G., Singleton, A., Shaklan, S., Smith, S., Southerd, B., Stahl, M., Steeves, J., Sturges, B., Sullivan, C., Tang, H., Taras, N., Tesch, J., Therrell, M., Tseng, H., Valente, M., Van Buren, D., Villalvazo, J., Warwick, S., Webb, D., Westerhoff, T., Wofford, R., Wu, G., Woo, J., Wood, M., Ziemer, J., Arney, G., Anderson, J., Maíz-Apellániz, J., Bartlett, J., Belikov, R., Bendek, E., Cenko, B., Douglas, E., Dulz, S., Evans, C., Faramaz, V., Feng, Y. K., Ferguson, H., Follette, K., Ford, S., García, M., Geha, M., Gelino, D., Götberg, Y., Hildebrandt, S., Hu, R., Jahnke, K., Kennedy, G., Kreidberg, L., Isella, A., Lopez, E., Marchis, F., Macri, L., Marley, M., Matzko, W., Mazoyer,

J., McCandliss, S., Meshkat, T., Mordasini, C., Morris, P., Nielsen, E., Newman, P., Petigura, E., Postman, M., Reines, A., Roberge, A., Roederer, I., Ruane, G., Schwieterman, E., Sirbu, D., Spalding, C., Teplitz, H., Tumlinson, J., Turner, N., Werk, J., Wofford, A., Wyatt, M., Young, A., and Zellem, R. (2020). The Habitable Exoplanet Observatory (HabEx) Mission Concept Study Final Report. *arXiv e-prints*, page arXiv:2001.06683.

Gibson, N. P., Merritt, S., Nugroho, S. K., Cubillos, P. E., de Mooij, E. J. W., Mikal-Evans, T., Fossati, L., Lothringer, J., Nikolov, N., Sing, D. K., Spake, J. J., Watson, C. A., and Wilson, J. (2020). Detection of Fe I in the atmosphere of the ultra-hot Jupiter WASP-121b, and a new likelihood-based approach for Doppler-resolved spectroscopy. *MNRAS*, 493(2):2215–2228.

Gordon, I. E., Rothman, L. S., Hill, C., Kochanov, R. V., Tan, Y., Bernath, P. F., Birk, M., Boudon, V., Campargue, A., Chance, K. V., Drouin, B. J., Flaud, J. M., Gamache, R. R., Hodges, J. T., Jacquemart, D., Perevalov, V. I., Perrin, A., Shine, K. P., Smith, M. A. H., Tennyson, J., Toon, G. C., Tran, H., Tyuterev, V. G., Barbe, A., Császár, A. G., Devi, V. M., Furtenbacher, T., Harrison, J. J., Hartmann, J. M., Jolly, A., Johnson, T. J., Karman, T., Kleiner, I., Kyuberis, A. A., Loos, J., Lyulin, O. M., Massie, S. T., Mikhailenko, S. N., Moazzen-Ahmadi, N., Müller, H. S. P., Naumenko, O. V., Nikitin, A. V., Polyansky, O. L., Rey, M., Rotger, M., Sharpe, S. W., Sung, K., Starikova, E., Tashkun, S. A., Auwera, J. V., Wagner, G., Wilzewski, J., Wcisło, P., Yu, S., and Zak, E. J. (2017). The HITRAN2016 molecular spectroscopic database. *JQSRT*, 203:3–69.

Goyal, J. M., Mayne, N., Drummond, B., Sing, D. K., Hébrard, E., Lewis, N., Tremblin, P., Phillips, M. W., Mikal-Evans, T., and Wakeford, H. R. (2020). A library of self-consistent simulated exoplanet atmospheres. *MNRAS*, 498(4):4680–4704.

Goyal, J. M., Mayne, N., Sing, D. K., Drummond, B., Tremblin, P., Amundsen, D. S., Evans, T., Carter, A. L., Spake, J., Baraffe, I., Nikolov, N., Manners, J., Chabrier, G., and Hebrard, E. (2018). A library of ATMO forward model transmission spectra for hot Jupiter exoplanets. *MNRAS*, 474(4):5158–5185.

Goyal, J. M., Wakeford, H. R., Mayne, N. J., Lewis, N. K., Drummond, B., and Sing, D. K. (2019). Fully scalable forward model grid of exoplanet transmission spectra. *MNRAS*, 482(4):4503–4513.

Greene, T. P., Line, M. R., Montero, C., Fortney, J. J., Lustig-Yaeger, J., and Luther, K. (2016). Characterizing Transiting Exoplanet Atmospheres with JWST. *ApJ*, 817(1):17.

Grenfell, J. L., Godolt, M., Cabrera, J., Carone, L., Muñoz, A. G., Kitzmann, D., Smith, A. M. S., and Rauer, H. (2020). Atmospheric Characterization via Broadband Color Filters on the PLAnetary Transits and Oscillations of stars (PLATO) Mission. *Experimental Astronomy*, 50(1):1–49.

Grimm, S. L. and Heng, K. (2015). HELIOS-K: An Ultrafast, Open-source Opacity Calculator for Radiative Transfer. *ApJ*, 808(2):182.

Grimm, S. L., Malik, M., Kitzmann, D., Guzmán-Mesa, A., Hoeijmakers, H. J., Fisher, C., Mendonça, J. M., Yurchenko, S. N., Tennyson, J., Alesina, F., Buchschacher, N., Burnier, J., Segransan, D., Kurucz, R. L., and Heng, K. (2021). HELIOS-K 2.0 Opacity Calculator and Open-source Opacity Database for Exoplanetary Atmospheres. *ApJS*, 253(1):30.

Guillot, T. (2010). On the radiative equilibrium of irradiated planetary atmospheres. *A&A*, 520:A27.

Guillot, T., Burrows, A., Hubbard, W. B., Lunine, J. I., and Saumon, D. (1996). Giant Planets at Small Orbital Distances. *ApJL*, 459:L35.

Guzmán-Mesa, A., Kitzmann, D., Fisher, C., Burgasser, A. J., Hoeijmakers, H. J., Márquez-Neila, P., Grimm, S. L., Mandell, A. M., Sznitman, R., and Heng, K. (2020). Information Content of JWST NIRSpec Transmission Spectra of Warm Neptunes. *AJ*, 160(1):15.

Harrington, J., Hansen, B. M., Luszcz, S. H., Seager, S., Deming, D., Menou, K., Cho, J. Y. K., and Richardson, L. J. (2006). The Phase-Dependent Infrared Brightness of the Extrasolar Planet υ Andromedae b. *Science*, 314(5799):623–626.

Hart, M. H. (1978). The evolution of the atmosphere of the earth. *Icarus*, 33(1):23–39.

Hayes, J. J. C., Kerins, E., Awiphan, S., McDonald, I., Morgan, J. S., Chuanraksasat, P., Komonjinda, S., Sanguansak, N., Kittara, P., and SPEARNet Collaboration (2020). Optimizing exoplanet atmosphere retrieval using unsupervised machine-learning classification. *MNRAS*, 494(3):4492–4508.

Helling, C., Tootill, D., Woitke, P., and Lee, E. (2017). Dust in brown dwarfs and extrasolar planets. V. Cloud formation in carbon- and oxygen-rich environments. *A&A*, 603:A123.

Helling, C. and Woitke, P. (2006). Dust in brown dwarfs. V. Growth and evaporation of dirty dust grains. *A&A*, 455(1):325–338.

Helling, C., Woitke, P., and Thi, W. F. (2008). Dust in brown dwarfs and extra-solar planets. I. Chemical composition and spectral appearance of quasi-static cloud layers. *A&A*, 485(2):547–560.

Heng, K., Hayek, W., Pont, F., and Sing, D. K. (2012). On the effects of clouds and hazes in the atmospheres of hot Jupiters: semi-analytical temperature-pressure profiles. *MNRAS*, 420(1):20–36.

Heng, K. and Kitzmann, D. (2017). The theory of transmission spectra revisited: a semi-analytical method for interpreting WFC3 data and an unresolved challenge. *MNRAS*, 470(3):2972–2981.

Heng, K. and Lyons, J. R. (2016). Carbon Dioxide in Exoplanetary Atmospheres: Rarely Dominant Compared to Carbon Monoxide and Water in Hot, Hydrogen-dominated Atmospheres. *ApJ*, 817(2):149.

Heng, K., Mendonça, J. M., and Lee, J.-M. (2014). Analytical Models of Exoplanetary Atmospheres. II. Radiative Transfer via the Two-stream Approximation. *ApJS*, 215(1):4.

Heng, K. and Tsai, S.-M. (2016). Analytical Models of Exoplanetary Atmospheres. III. Gaseous C-H-O-N Chemistry with Nine Molecules. *ApJ*, 829(2):104.

Hoeijmakers, H. J., Cabot, S. H. C., Zhao, L., Buchhave, L. A., Tronsgaard, R., Davis, A. B., Kitzmann, D., Grimm, S. L., Cegla, H. M., Bourrier, V., Ehrenreich, D., Heng, K., Lovis, C., and Fischer, D. A. (2020a). High-resolution transmission spectroscopy of MASCARA-2 b with EXPRES. *A&A*, 641:A120.

Hoeijmakers, H. J., Ehrenreich, D., Heng, K., Kitzmann, D., Grimm, S. L., Allart, R., Deitrick, R., Wyttenbach, A., Oreshenko, M., Pino, L., Rimmer, P. B., Molinari, E., and Di Fabrizio, L. (2018a). Atomic iron and titanium in the atmosphere of the exoplanet KELT-9b. *Nature*, 560(7719):453–455.

Hoeijmakers, H. J., Ehrenreich, D., Kitzmann, D., Allart, R., Grimm, S. L., Seidel, J. V., Wyttenbach, A., Pino, L., Nielsen, L. D., Fisher, C., Rimmer, P. B., Bourrier, V., Cegla, H. M., Lavie, B., Lovis, C., Patzer, A. B. C., Stock, J. W., Pepe, F. A., and Heng, K. (2019). A spectral survey of an ultra-hot Jupiter. Detection of metals in the transmission spectrum of KELT-9 b. *A&A*, 627:A165.

Hoeijmakers, H. J., Schwarz, H., Snellen, I. A. G., de Kok, R. J., Bonnefoy, M., Chauvin, G., Lagrange, A. M., and Girard, J. H. (2018b). Medium-resolution integral-field spectroscopy for high-contrast exoplanet imaging. Molecule maps of the β Pictoris system with SINFONI. *A&A*, 617:A144.

Hoeijmakers, H. J., Seidel, J. V., Pino, L., Kitzmann, D., Sindel, J. P., Ehrenreich, D., Oza, A. V., Bourrier, V., Allart, R., Gebek, A., Lovis, C., Yurchenko, S. N., Astudillo-Defru, N., Bayliss, D., Cegla, H., Lavie, B., Lendl, M., Melo, C., Murgas, F., Nascimbeni, V., Pepe, F., Ségransan, D., Udry, S., Wyttenbach, A., and Heng, K. (2020b). Hot Exoplanet Atmospheres Resolved with Transit Spectroscopy (HEARTS). IV. A spectral inventory of atoms and molecules in the high-resolution transmission spectrum of WASP-121 b. *A&A*, 641:A123.

Houllé, M., Vigan, A., Carlotti, A., Choquet, É., Cantalloube, F., Phillips, M. W., Sauvage, J.-F., Schwartz, N., Otten, G. P. P. L., Baraffe, I., Emsenhuber, A., and Mordasini, C. (2021). Direct imaging and spectroscopy of exoplanets with the ELT/HARMONI high-contrast module. *arXiv e-prints*, page arXiv:2104.11251.

Huang, S.-S. (1959). Occurrence of Life in the Universe. *American Scientist*, 47(3):397–402.

Irwin, P. G. J., Teanby, N. A., de Kok, R., Fletcher, L. N., Howett, C. J. A., Tsang, C. C. C., Wilson, C. F., Calcutt, S. B., Nixon, C. A., and Parrish, P. D. (2008). The NEMESIS planetary atmosphere radiative transfer and retrieval tool. *JQSRT*, 109:1136–1150.

Johnsen, T. K., Marley, M. S., and Gulick, V. C. (2020). A Multilayer Perceptron for Obtaining Quick Parameter Estimations of Cool Exoplanets from Geometric Albedo Spectra. *PASP*,

132(1010):044502.

Kaeufl, H.-U., Ballester, P., Biereichel, P., Delabre, B., Donaldson, R., Dorn, R., Fedrigo, E., Finger, G., Fischer, G., Franza, F., Gojak, D., Huster, G., Jung, Y., Lizon, J.-L., Mehrgan, L., Meyer, M., Moorwood, A., Pirard, J.-F., Paufique, J., Pozna, E., Siebenmorgen, R., Silber, A., Stegmeier, J., and Wegerer, S. (2004). CRIRES: a high-resolution infrared spectrograph for ESO's VLT. In Moorwood, A. F. M. and Iye, M., editors, *Ground-based Instrumentation for Astronomy*, volume 5492 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 1218–1227.

Kahanamoku, S., 'Anolani Alegado, R., Kagawa-Viviani, A., Leimomi Kamelamela, K., Kamai, B., Walkowicz, L. M., Prescod-Weinstein, C., Alexa de los Reyes, M., and Neilson, H. (2020). A Native Hawaiian-led summary of the current impact of constructing the Thirty Meter Telescope on Maunakea. *arXiv e-prints*, page arXiv:2001.00970.

Kasting, J. F., Whitmire, D. P., and Reynolds, R. T. (1993). Habitable Zones around Main Sequence Stars. *Icarus*, 101(1):108–128.

Kempton, E. M. R., Bean, J. L., Louie, D. R., Deming, D., Koll, D. D. B., Mansfield, M., Christiansen, J. L., López-Morales, M., Swain, M. R., Zellem, R. T., Ballard, S., Barclay, T., Barstow, J. K., Batalha, N. E., Beatty, T. G., Berta-Thompson, Z., Birkby, J., Buchhave, L. A., Charbonneau, D., Cowan, N. B., Crossfield, I., de Val-Borro, M., Doyon, R., Dragomir, D., Gaidos, E., Heng, K., Hu, R., Kane, S. R., Kreidberg, L., Mallonn, M., Morley, C. V., Narita, N., Nascimbeni, V., Pallé, E., Quintana, E. V., Rauscher, E., Seager, S., Shkolnik, E. L., Sing, D. K., Sozzetti, A., Stassun, K. G., Valenti, J. A., and von Essen, C. (2018). A Framework for Prioritizing the TESS Planetary Candidates Most Amenable to Atmospheric Characterization. *PASP*, 130(993):114401.

Kite, E. S., Fegley, Bruce, J., Schaefer, L., and Ford, E. B. (2020). Atmosphere Origins for Exoplanet Sub-Neptunes. *ApJ*, 891(2):111.

Kitzmann, D., Heng, K., Oreshenko, M., Grimm, S. L., Apai, D., Bowler, B. P., Burgasser, A. J., and Marley, M. S. (2020). Helios-r2: A New Bayesian, Open-source Retrieval Model for Brown Dwarfs and Exoplanet Atmospheres. *ApJ*, 890(2):174.

Kitzmann, D., Heng, K., Rimmer, P. B., Hoeijmakers, H. J., Tsai, S.-M., Malik, M., Lendl, M., Deitrick, R., and Demory, B.-O. (2018). The Peculiar Atmospheric Chemistry of KELT-9b. *ApJ*, 863(2):183.

Kramida, A., Yu. Ralchenko, Reader, J., and and NIST ASD Team (2019). NIST Atomic Spectra Database (ver. 5.7.1), [Online]. Available: `https://physics.nist.gov/asd` [2017, April 9]. National Institute of Standards and Technology, Gaithersburg, MD.

Kreidberg, L., Bean, J. L., Désert, J.-M., Benneke, B., Deming, D., Stevenson, K. B., Seager, S., Berta-Thompson, Z., Seifahrt, A., and Homeier, D. (2014a). Clouds in the atmosphere of the super-Earth exoplanet GJ1214b. *Nature*, 505(7481):69–72.

Kreidberg, L., Bean, J. L., Désert, J.-M., Line, M. R., Fortney, J. J., Madhusudhan, N., Stevenson, K. B., Showman, A. P., Charbonneau, D., McCullough, P. R., Seager, S., Burrows, A., Henry, G. W., Williamson, M., Kataria, T., and Homeier, D. (2014b). A Precise Water Abundance Measurement for the Hot Jupiter WASP-43b. *ApJL*, 793(2):L27.

Kreidberg, L., Line, M. R., Bean, J. L., Stevenson, K. B., Désert, J.-M., Madhusudhan, N., Fortney, J. J., Barstow, J. K., Henry, G. W., Williamson, M. H., and Showman, A. P. (2015). A Detection of Water in the Transmission Spectrum of the Hot Jupiter WASP-12b and Implications for Its Atmospheric Composition. *ApJ*, 814(1):66.

Krissansen-Totton, J., Olson, S., and Catling, D. C. (2018a). Disequilibrium biosignatures over Earth history and implications for detecting exoplanet life. *Science Advances*, 4(1):eaao5747.

Krissansen-Totton, J., Garland, R., Irwin, P., and Catling, D. C. (2018b). Detectability of Biosignatures in Anoxic Atmospheres with the James Webb Space Telescope: A TRAPPIST-1e Case Study. *AJ*, 156(3):114.

Kurucz, R. L. (2017). Including all the lines: data releases for spectra and opacities. *Canadian Journal of Physics*, 95(9):825–827.

Lavie, B., Mendonça, J. M., Mordasini, C., Malik, M., Bonnefoy, M., Demory, B.-O., Oreshenko, M., Grimm, S. L., Ehrenreich, D., and Heng, K. (2017). HELIOS-RETRIEVAL: An Opensource, Nested Sampling Atmospheric Retrieval Code; Application to the HR 8799 Exoplanets and Inferred Constraints for Planet Formation. *AJ*, 154(3):91.

Lederberg, J. (1965). Signs of Life: Criterion-System of Exobiology. *Nature*, 207(4992):9–13.

Lee, J. M., Fletcher, L. N., and Irwin, P. G. J. (2012). Optimal estimation retrievals of the atmospheric structure and composition of HD 189733b from secondary eclipse spectroscopy. *MNRAS*, 420(1):170–182.

Leleu, A., Alibert, Y., Hara, N., et al. (2020). *A&A (submitted):* *https://tinyurl.com/y4kktcdm*.

Leleu, A., Lillo-Box, J., Sestovic, M., Robutel, P., Correia, A. C. M., Hara, N., Angerhausen, D., Grimm, S. L., and Schneider, J. (2019). Co-orbital exoplanets from close-period candidates: the TOI-178 case. *A&A*, 624:A46.

Line, M. R., Knutson, H., Wolf, A. S., and Yung, Y. L. (2014). A Systematic Retrieval Analysis of Secondary Eclipse Spectra. II. A Uniform Analysis of Nine Planets and their C to O Ratios. *ApJ*, 783(2):70.

Line, M. R., Teske, J., Burningham, B., Fortney, J. J., and Marley, M. S. (2015). Uniform Atmospheric Retrieval Analysis of Ultracool Dwarfs. I. Characterizing Benchmarks, Gl 570D and HD 3651B. *ApJ*, 807(2):183.

Line, M. R., Wolf, A. S., Zhang, X., Knutson, H., Kammer, J. A., Ellison, E., Deroo, P., Crisp, D., and Yung, Y. L. (2013). A Systematic Retrieval Analysis of Secondary Eclipse Spectra. I. A Comparison of Atmospheric Retrieval Techniques. *ApJ*, 775(2):137.

Lopez-Morales, M., Currie, T., Teske, J., Gaidos, E., Kempton, E., Males, J., Lewis, N., Rackham, B. V., Ben-Ami, S., Birkby, J., Charbonneau, D., Close, L., Crane, J., Dressing, C., Froning, C., Hasegawa, Y., Konopacky, Q., Kopparapu, R. K., Mawet, D., Mennesson, B., Ramirez, R., Stelter, D., Szentgyorgyi, A., Wang, J., Alam, M., Collins, K., Dupree, A., Karovska, M., Kirk, J., Levi, A., McGruder, C., Packman, C., Rugheimer, S., and Rukdee, S. (2019). Detecting Earth-like Biosignatures on Rocky Exoplanets around Nearby Stars with Ground-based Extremely Large Telescopes. *BAAS*, 51(3):162.

Lovelock, J. E. (1965). A Physical Basis for Life Detection Experiments. *Nature*, 207(4997):568–570.

MacDonald, R. J. and Madhusudhan, N. (2017). HD 209458b in new light: evidence of nitrogen chemistry, patchy clouds and sub-solar water. *MNRAS*, 469(2):1979–1996.

Madhusudhan, N. (2012). C/O Ratio as a Dimension for Characterizing Exoplanetary Atmospheres. *ApJ*, 758(1):36.

Madhusudhan, N. (2018). *Atmospheric Retrieval of Exoplanets*, page 104.

Madhusudhan, N. (2019). Exoplanetary Atmospheres: Key Insights, Challenges, and Prospects. *ARA&A*, 57:617–663.

Madhusudhan, N., Amin, M. A., and Kennedy, G. M. (2014). Toward Chemical Constraints on Hot Jupiter Migration. *ApJL*, 794(1):L12.

Madhusudhan, N. and Seager, S. (2009). A Temperature and Abundance Retrieval Method for Exoplanet Atmospheres. *ApJ*, 707(1):24–39.

Malik, M., Grosheintz, L., Mendonça, J. M., Grimm, S. L., Lavie, B., Kitzmann, D., Tsai, S.-M., Burrows, A., Kreidberg, L., Bedell, M., Bean, J. L., Stevenson, K. B., and Heng, K. (2017). HELIOS: An Open-source, GPU-accelerated Radiative Transfer Code for Self-consistent Exoplanetary Atmospheres. *AJ*, 153(2):56.

Malik, M., Kitzmann, D., Mendonça, J. M., Grimm, S. L., Marleau, G.-D., Linder, E. F., Tsai, S.-M., and Heng, K. (2019). Self-luminous and Irradiated Exoplanetary Atmospheres Explored with HELIOS. *AJ*, 157(5):170.

Mansfield, M., Bean, J. L., Line, M. R., Parmentier, V., Kreidberg, L., Désert, J.-M., Fortney, J. J., Stevenson, K. B., Arcangeli, J., and Dragomir, D. (2018). An HST/WFC3 Thermal Emission Spectrum of the Hot Jupiter HAT-P-7b. *AJ*, 156(1):10.

Marconi, A., Abreu, M., Adibekyan, V., Aliverti, M., Allende Prieto, C., Amado, P., Amate, M., Artigau, E., Augusto, S., Barros, S., Becerril, S., Benneke, B., Bergin, E., Berio, P., Bezawada,

N., Boisse, I., Bonfils, X., Bouchy, F., Broeg, C., Cabral, A., Calvo-Ortega, R., Canto Martins, B. L., Chazelas, B., Chiavassa, A., Christensen, L., Cirami, R., Coretti, I., Covino, S., Cresci, G., Cristiani, S., Cunha Parro, V., Cupani, G., de Castro Leão, I., Renan de Medeiros, J., Furlande Souza, M. A., Di Marcantonio, P., Di Varano, I., D'Odorico, V., Doyon, R., Drass, H., Figueira, P., Belen Fragoso, A., Uldall Fynbo, J. P., Gallo, E., Genoni, M., González Hernández, J., Haehnelt, M., Hlavacek-Larrondo, J., Hughes, I., Huke, P., Humphrey, A., Kjeldsen, H., Korn, A., Kouach, D., Landoni, M., Liske, J., Lovis, C., Lunney, D., Maiolino, R., Malo, L., Marquart, T., Martins, C., Mason, E., Molaro, P., Monnier, J., Monteiro, M., Mordasini, C., Morris, T., Mucciarelli, A., Murray, G., Niedzielski, A., Nunes, N., Oliva, E., Origlia, L., Pallé, E., Pariani, G., Parr-Burman, P., Peñate, J., Pepe, F., Pinna, E., Piskunov, N., Rasilla Piñeiro, J. L., Rebolo, R., Rees, P., Reiners, A., Riva, M., Romano, D., Rousseau, S., Sanna, N., Santos, N., Sarajlic, M., Shen, T. C., Sortino, F., Sosnowska, D., Sousa, S., Stempels, E., Strassmeier, K., Tenegi, F., Tozzi, A., Udry, S., Valenziano, L., Vanzi, L., Weber, M., Woche, M., Xompero, M., Zackrisson, E., and Zapatero Osorio, M. R. (2021). HIRES, the High-resolution Spectrograph for the ELT. *The Messenger*, 182:27–32.

Marley, M. S., Ackerman, A. S., Cuzzi, J. N., and Kitzmann, D. (2013). *Clouds and Hazes in Exoplanet Atmospheres*, page 367.

Marley, M. S., Saumon, D., Fortney, J. J., Morley, C., Lupu, R. E., Freedman, R., and Visscher, C. (2017). Sonora: A New Generation Model Atmosphere Grid for Brown Dwarfs and Young Extrasolar Giant Planets. In *American Astronomical Society Meeting Abstracts #230*, volume 230 of *American Astronomical Society Meeting Abstracts*, page 315.07.

Marley, M. S., Saumon, D., Visscher, C., Lupu, R., Freedman, R., Morley, C., Fortney, J. J., Seay, C., Smith, A. J. R. W., Teal, D. J., and Wang, R. (2021). The Sonora Brown Dwarf Atmosphere and Evolution Models I. Model Description and Application to Cloudless Atmospheres in Rainout Chemical Equilibrium. *arXiv e-prints*, page arXiv:2107.07434.

Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., and Barman, T. (2010). Images of a fourth planet orbiting HR 8799. *Nature*, 468(7327):1080–1083.

Márquez-Neila, P., Fisher, C., Sznitman, R., and Heng, K. (2018). Supervised machine learning for analysing spectra of exoplanetary atmospheres. *Nature Astronomy*, 2:719–724.

Mayor, M., Pepe, F., Queloz, D., Bouchy, F., Rupprecht, G., Lo Curto, G., Avila, G., Benz, W., Bertaux, J. L., Bonfils, X., Dall, T., Dekker, H., Delabre, B., Eckert, W., Fleury, M., Gilliotte, A., Gojak, D., Guzman, J. C., Kohler, D., Lizon, J. L., Longinotti, A., Lovis, C., Megevand, D., Pasquini, L., Reyes, J., Sivan, J. P., Sosnowska, D., Soto, R., Udry, S., van Kesteren, A., Weber, L., and Weilenmann, U. (2003). Setting New Standards with HARPS. *The Messenger*, 114:20–24.

Mayor, M. and Queloz, D. (1995). A Jupiter-mass companion to a solar-type star. *Nature*, 378(6555):355–359.

McCauliff, S. D., Jenkins, J. M., Catanzarite, J., Burke, C. J., Coughlin, J. L., Twicken, J. D., Tenenbaum, P., Seader, S., Li, J., and Cote, M. (2015). Automatic Classification of Kepler Planetary Transit Candidates. *ApJ*, 806(1):6.

Miller-Ricci Kempton, E. and Rauscher, E. (2012). Constraining High-speed Winds in Exoplanet Atmospheres through Observations of Anomalous Doppler Shifts during Transit. *ApJ*, 751(2):117.

Millholland, S. and Laughlin, G. (2017). Supervised Learning Detection of Sixty Non-transiting Hot Jupiter Candidates. *AJ*, 154(3):83.

Mislis, D., Bachelet, E., Alsubai, K. A., Bramich, D. M., and Parley, N. (2016). SIDRA: a blind algorithm for signal detection in photometric surveys. *MNRAS*, 455(1):626–633.

Miyazaki, S., Johnson, S. A., Sumi, T., Penny, M. T., Koshimoto, N., and Yamawaki, T. (2021). Revealing Short-period Exoplanets and Brown Dwarfs in the Galactic Bulge Using the Microlensing Xallarap Effect with the Nancy Grace Roman Space Telescope. *AJ*, 161(2):84.

Mollière, P., van Boekel, R., Bouwman, J., Henning, T., Lagage, P. O., and Min, M. (2017). Observing transiting planets with JWST. Prime targets and their synthetic spectral observations. *A&A*, 600:A10.

Mordasini, C., Alibert, Y., Benz, W., Klahr, H., and Henning, T. (2012). Extrasolar planet population synthesis . IV. Correlations with disk metallicity, mass, and lifetime. *A&A*, 541:A97.

Moses, J. I., Line, M. R., Visscher, C., Richardson, M. R., Nettelmann, N., Fortney, J. J., Barman, T. S., Stevenson, K. B., and Madhusudhan, N. (2013). Compositional Diversity in the Atmospheres of Hot Neptunes, with Application to GJ 436b. *ApJ*, 777(1):34.

Moses, J. I., Visscher, C., Fortney, J. J., Showman, A. P., Lewis, N. K., Griffith, C. A., Klippenstein, S. J., Shabram, M., Friedson, A. J., Marley, M. S., and Freedman, R. S. (2011). Disequilibrium Carbon, Oxygen, and Nitrogen Chemistry in the Atmospheres of HD 189733b and HD 209458b. *ApJ*, 737(1):15.

Nelson, A. F. and Angel, J. R. P. (1998). The Range of Masses and Periods Explored by Radial Velocity Searches for Planetary Companions. *ApJ*, 500(2):940–957.

Nielsen, L. D., Gandolfi, D., Armstrong, D. J., Jenkins, J. S., Fridlund, M., Santos, N. C., Dai, F., Adibekyan, V., Luque, R., Steffen, J. H., Esposito, M., Meru, F., Sabotta, S., Bolmont, E., Kossakowski, D., Otegi, J. F., Murgas, F., Stalport, M., Rodler, F., Díaz, M. R., Kurtovic, N. T., Ricker, G., Vanderspek, R., Latham, D. W., Seager, S., Winn, J. N., Jenkins, J. M., Allart, R., Almenara, J. M., Barrado, D., Barros, S. C. C., Bayliss, D., Berdiñas, Z. M., Boisse, I., Bouchy, F., Boyd, P., Brown, D. J. A., Bryant, E. M., Burke, C., Cochran, W. D., Cooke, B. F., Demangeon, O. D. S., Díaz, R. F., Dittman, J., Dorn, C., Dumusque, X., García, R. A., González-Cuesta, L., Grziwa, S., Georgieva, I., Guerrero, N., Hatzes, A. P., Helled, R., Henze, C. E., Hojjatpanah, S., Korth, J., Lam, K. W. F., Lillo-Box, J., Lopez, T. A., Livingston, J., Mathur, S., Mousis, O., Narita, N., Osborn, H. P., Palle, E., Rojas, P. A. P., Persson,

C. M., Quinn, S. N., Rauer, H., Redfield, S., Santerne, A., dos Santos, L. A., Seidel, J. V., Sousa, S. G., Ting, E. B., Turbet, M., Udry, S., Vand erburg, A., Van Eylen, V., Vines, J. I., Wheatley, P. J., and Wilson, P. A. (2020). Mass determinations of the three mini-Neptunes transiting TOI-125. *MNRAS*, 492(4):5399–5412.

Nixon, M. C. and Madhusudhan, N. (2020). Assessment of supervised machine learning for atmospheric retrieval of exoplanets. *MNRAS*, 496(1):269–281.

Nugroho, S. K., Gibson, N. P., de Mooij, E. J. W., Watson, C. A., Kawahara, H., and Merritt, S. (2020). Searching for thermal inversion agents in the transmission spectrum of KELT-20b/MASCARA-2b: detection of neutral iron and ionised calcium H&K lines. *MNRAS*, 496(1):504–522.

Öberg, K. I. and Bergin, E. A. (2016). Excess C/O and C/H in Outer Protoplanetary Disk Gas. *ApJL*, 831(2):L19.

Öberg, K. I., Murray-Clay, R., and Bergin, E. A. (2011). The Effects of Snowlines on C/O in Planetary Atmospheres. *ApJL*, 743(1):L16.

O'Malley-James, J. T. and Kaltenegger, L. (2018). Biofluorescent worlds: global biological fluorescence as a biosignature. *MNRAS*, 481(2):2487–2496.

O'Malley-James, J. T. and Kaltenegger, L. (2019). Biofluorescent Worlds - II. Biological fluorescence induced by stellar UV flares, a new temporal biosignature. *MNRAS*, 488(4):4530–4545.

Oreshenko, M., Kitzmann, D., Márquez-Neila, P., Malik, M., Bowler, B. P., Burgasser, A. J., Sznitman, R., Fisher, C. E., and Heng, K. (2020). Supervised Machine Learning for Intercomparison of Model Grids of Brown Dwarfs: Application to GJ 570D and the Epsilon Indi B Binary System. *AJ*, 159(1):6.

Osborn, H. P., Ansdell, M., Ioannou, Y., Sasdelli, M., Angerhausen, D., Caldwell, D., Jenkins, J. M., Räissi, C., and Smith, J. C. (2020). Rapid classification of TESS planet candidates with convolutional neural networks. *A&A*, 633:A53.

Pearson, K. A., Palafox, L., and Griffith, C. A. (2018). Searching for exoplanets using artificial intelligence. *MNRAS*, 474(1):478–491.

Pepe, F., Mayor, M., Galland, F., Naef, D., Queloz, D., Santos, N. C., Udry, S., and Burnet, M. (2002). The CORALIE survey for southern extra-solar planets VII. Two short-period Saturnian companions to <ASTROBJ>HD 108147</ASTROBJ> and <ASTROBJ>HD 168746</ASTROBJ>. *A&A*, 388:632–638.

Pepe, F., Molaro, P., Cristiani, S., Rebolo, R., Santos, N. C., Dekker, H., Mégevand, D., Zerbi, F. M., Cabral, A., Di Marcantonio, P., Abreu, M., Affolter, M., Aliverti, M., Allende Prieto, C., Amate, M., Avila, G., Baldini, V., Bristow, P., Broeg, C., Cirami, R., Coelho, J., Conconi, P., Coretti, I., Cupani, G., D'Odorico, V., De Caprio, V., Delabre, B., Dorn, R., Figueira, P., Fragoso, A., Galeotta, S., Genolet, L., Gomes, R., González Hernández, J. I., Hughes,

I., Iwert, O., Kerber, F., Landoni, M., Lizon, J. L., Lovis, C., Maire, C., Mannetta, M., Martins, C., Monteiro, M., Oliveira, A., Poretti, E., Rasilla, J. L., Riva, M., Santana Tschudi, S., Santos, P., Sosnowska, D., Sousa, S., Spanó, P., Tenegi, F., Toso, G., Vanzella, E., Viel, M., and Zapatero Osorio, M. R. (2014). ESPRESSO: The next European exoplanet hunter. *Astronomische Nachrichten*, 335(1):8.

Petersburg, R. R., Ong, J. M. J., Zhao, L. L., Blackman, R. T., Brewer, J. M., Buchhave, L. A., Cabot, S. H. C., Davis, A. B., Jurgenson, C. A., Leet, C., McCracken, T. M., Sawyer, D., Sharov, M., Tronsgaard, R., Szymkowiak, A. E., and Fischer, D. A. (2020). An Extreme-precision Radial-velocity Pipeline: First Radial Velocities from EXPRES. *AJ*, 159(5):187.

Petigura, E. A., Howard, A. W., and Marcy, G. W. (2013). Prevalence of Earth-size planets orbiting Sun-like stars. *Proceedings of the National Academy of Science*, 110(48):19273–19278.

Petkowski, J. J., Bains, W., and Seager, S. (2020). On the potential of silicon as a building block for life. *Life*, 10(6).

Pinhas, A., Madhusudhan, N., Gandhi, S., and MacDonald, R. (2019). H_2O abundances and cloud properties in ten hot giant exoplanets. *MNRAS*, 482(2):1485–1498.

Pino, L., Désert, J.-M., Brogi, M., Malavolta, L., Wyttenbach, A., Line, M., Hoeijmakers, J., Fossati, L., Bonomo, A. S., Nascimbeni, V., Panwar, V., Affer, L., Benatti, S., Biazzo, K., Bignamini, A., Borsa, F., Carleo, I., Claudi, R., Cosentino, R., Covino, E., Damasso, M., Desidera, S., Giacobbe, P., Harutyunyan, A., Lanza, A. F., Leto, G., Maggio, A., Maldonado, J., Mancini, L., Micela, G., Molinari, E., Pagano, I., Piotto, G., Poretti, E., Rainer, M., Scandariato, G., Sozzetti, A., Allart, R., Borsato, L., Bruno, G., Di Fabrizio, L., Ehrenreich, D., Fiorenzano, A., Frustagli, G., Lavie, B., Lovis, C., Magazzù, A., Nardiello, D., Pedani, M., and Smareglia, R. (2020). Neutral Iron Emission Lines from the Dayside of KELT-9b: The GAPS Program with HARPS-N at TNG XX. *ApJL*, 894(2):L27.

Piskorz, D., Buzard, C., Line, M. R., Knutson, H. A., Benneke, B., Crockett, N. R., Lockwood, A. C., Blake, G. A., Barman, T. S., Bender, C. F., Deming, D., and Johnson, J. A. (2018). Ground- and Space-based Detection of the Thermal Emission Spectrum of the Transiting Hot Jupiter KELT-2Ab. *AJ*, 156(3):133.

Prescod-Weinstein, C., Walkowicz, L. M., Tuttle, S., Nord, B., and Neilson, H. R. (2020). Re-framing astronomical research through an anticolonial lens – for TMT and beyond. *arXiv e-prints*, page arXiv:2001.00674.

Ramirez, R., Abbot, D. S., Fujii, Y., Hamano, K., Kite, E., Levi, A., Lingam, M., Lueftinger, T., Robinson, T. D., Rushby, A., Schaefer, L., Tasker, E., Vladilo, G., and Wordsworth, R. D. (2019). Habitable zone predictions and how to test them. *BAAS*, 51(3):31.

Ramirez, R. M. (2018). A More Comprehensive Habitable Zone for Finding Life on Other Planets. *Geosciences*, 8(8):280.

Ramirez, R. M. and Levi, A. (2018). The ice cap zone: a unique habitable zone for ocean worlds. *MNRAS*, 477(4):4627–4640.

Rauer, H., Catala, C., Aerts, C., Appourchaux, T., Benz, W., Brandeker, A., Christensen-Dalsgaard, J., Deleuil, M., Gizon, L., Goupil, M. J., Güdel, M., Janot-Pacheco, E., Mas-Hesse, M., Pagano, I., Piotto, G., Pollacco, D., Santos, Ċ., Smith, A., Suárez, J. C., Szabó, R., Udry, S., Adibekyan, V., Alibert, Y., Almenara, J. M., Amaro-Seoane, P., Eiff, M. A.-v., Asplund, M., Antonello, E., Barnes, S., Baudin, F., Belkacem, K., Bergemann, M., Bihain, G., Birch, A. C., Bonfils, X., Boisse, I., Bonomo, A. S., Borsa, F., Brandão, I. M., Brocato, E., Brun, S., Burleigh, M., Burston, R., Cabrera, J., Cassisi, S., Chaplin, W., Charpinet, S., Chiappini, C., Church, R. P., Csizmadia, S., Cunha, M., Damasso, M., Davies, M. B., Deeg, H. J., Díaz, R. F., Dreizler, S., Dreyer, C., Eggenberger, P., Ehrenreich, D., Eigmüller, P., Erikson, A., Farmer, R., Feltzing, S., de Oliveira Fialho, F., Figueira, P., Forveille, T., Fridlund, M., García, R. A., Giommi, P., Giuffrida, G., Godolt, M., Gomes da Silva, J., Granzer, T., Grenfell, J. L., Grotsch-Noels, A., Günther, E., Haswell, C. A., Hatzes, A. P., Hébrard, G., Hekker, S., Helled, R., Heng, K., Jenkins, J. M., Johansen, A., Khodachenko, M. L., Kislyakova, K. G., Kley, W., Kolb, U., Krivova, N., Kupka, F., Lammer, H., Lanza, A. F., Lebreton, Y., Magrin, D., Marcos-Arenal, P., Marrese, P. M., Marques, J. P., Martins, J., Mathis, S., Mathur, S., Messina, S., Miglio, A., Montalban, J., Montalto, M., Monteiro, M. J. P. F. G., Moradi, H., Moravveji, E., Mordasini, C., Morel, T., Mortier, A., Nascimbeni, V., Nelson, R. P., Nielsen, M. B., Noack, L., Norton, A. J., Ofir, A., Oshagh, M., Ouazzani, R. M., Pápics, P., Parro, V. C., Petit, P., Plez, B., Poretti, E., Quirrenbach, A., Ragazzoni, R., Raimondo, G., Rainer, M., Reese, D. R., Redmer, R., Reffert, S., Rojas-Ayala, B., Roxburgh, I. W., Salmon, S., Santerne, A., Schneider, J., Schou, J., Schuh, S., Schunker, H., Silva-Valio, A., Silvotti, R., Skillen, I., Snellen, I., Sohl, F., Sousa, S. G., Sozzetti, A., Stello, D., Strassmeier, K. G., Švanda, M., Szabó, G. M., Tkachenko, A., Valencia, D., Van Grootel, V., Vauclair, S. D., Ventura, P., Wagner, F. W., Walton, N. A., Weingrill, J., Werner, S. C., Wheatley, P. J., and Zwintz, K. (2014). The PLATO 2.0 mission. *Experimental Astronomy*, 38(1-2):249–330.

Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., Dunham, E. W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M. J., Howard, A. W., Ida, S., Jenkins, J. M., Jernigan, G., Johnson, J. A., Kaltenegger, L., Kawai, N., Kjeldsen, H., Laughlin, G., Levine, A. M., Lin, D., Lissauer, J. J., MacQueen, P., Marcy, G., McCullough, P. R., Morton, T. D., Narita, N., Paegert, M., Palle, E., Pepe, F., Pepper, J., Quirrenbach, A., Rinehart, S. A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K. G., Sullivan, P., Szentgyorgyi, A., Torres, G., Udry, S., and Villasenor, J. (2015). Transiting Exoplanet Survey Satellite (TESS). *Journal of Astronomical Telescopes, Instruments, and Systems*, 1:014003.

Rimmer, P. B. and Helling, C. (2016). A Chemical Kinetics Network for Lightning and Life in Planetary Atmospheres. *ApJS*, 224(1):9.

Rocchetto, M., Waldmann, I. P., Venot, O., Lagage, P. O., and Tinetti, G. (2016). Exploring Biases of Atmospheric Retrievals in Simulated JWST Transmission Spectra of Hot Jupiters. *ApJ*, 833(1):120.

Rodgers, C. D. (2000). Inverse Methods for Atmospheric Sounding - Theory and Practice. Inverse Methods for Atmospheric Sounding - Theory and Practice. Series: Series on Atmospheric Oceanic and Planetary Physics.

Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Chris Benner, D., Bernath, P. F., Birk, M., Bizzocchi, L., Boudon, V., Brown, L. R., Campargue, A., Chance, K., Cohen, E. A., Coudert, L. H., Devi, V. M., Drouin, B. J., Fayt, A., Flaud, J. M., Gamache, R. R., Harrison, J. J., Hartmann, J. M., Hill, C., Hodges, J. T., Jacquemart, D., Jolly, A., Lamouroux, J., Le Roy, R. J., Li, G., Long, D. A., Lyulin, O. M., Mackie, C. J., Massie, S. T., Mikhailenko, S., Müller, H. S. P., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E. R., Richard, C., Smith, M. A. H., Starikova, E., Sung, K., Tashkun, S., Tennyson, J., Toon, G. C., Tyuterev, V. G., and Wagner, G. (2013). The HITRAN2012 molecular spectroscopic database. *JQSRT*, 130:4–50.

Rothman, L. S., Gordon, I. E., Barber, R. J., Dothe, H., Gamache, R. R., Goldman, A., Perevalov, V. I., Tashkun, S. A., and Tennyson, J. (2010). HITEMP, the high-temperature molecular spectroscopic database. *JQSRT*, 111:2139–2150.

Saumon, D., Hubbard, W. B., Burrows, A., Guillot, T., Lunine, J. I., and Chabrier, G. (1996). A Theory of Extrasolar Giant Planets. *ApJ*, 460:993.

Schanche, N., Collier Cameron, A., Hébrard, G., Nielsen, L., Triaud, A. H. M. J., Almenara, J. M., Alsubai, K. A., Anderson, D. R., Armstrong, D. J., Barros, S. C. C., Bouchy, F., Boumis, P., Brown, D. J. A., Faedi, F., Hay, K., Hebb, L., Kiefer, F., Mancini, L., Maxted, P. F. L., Palle, E., Pollacco, D. L., Queloz, D., Smalley, B., Udry, S., West, R., and Wheatley, P. J. (2019). Machine-learning approaches to exoplanet transit detection and candidate validation in wide-field ground-based surveys. *MNRAS*, 483(4):5534–5547.

Schulze-Makuch, D. and Irwin, L. (2006). The prospect of alien life in exotic forms on other worlds. *Die Naturwissenschaften*, 93:155–72.

Schwieterman, E. W., Kiang, N. Y., Parenteau, M. N., Harman, C. E., DasSarma, S., Fisher, T. M., Arney, G. N., Hartnett, H. E., Reinhard, C. T., Olson, S. L., Meadows, V. S., Cockell, C. S., Walker, S. I., Grenfell, J. L., Hegde, S., Rugheimer, S., Hu, R., and Lyons, T. W. (2018). Exoplanet Biosignatures: A Review of Remotely Detectable Signs of Life. *Astrobiology*, 18(6):663–708.

Seager, S. (2008). Exoplanet Transit Spectroscopy and Photometry. *SSRv*, 135(1-4):345–354.

Seager, S. (2013). Exoplanet Habitability. *Science*, 340(6132):577–581.

Seager, S., Bains, W., and Petkowski, J. J. (2016). Toward a List of Molecules as Potential Biosignature Gases for the Search for Life on Exoplanets and Applications to Terrestrial

Biochemistry. *Astrobiology*, 16(6):465–485.

Seager, S., Petkowski, J. J., Günther, M. N., Bains, W., Mikal-Evans, T., and Deming, D. (2021). Possibilities for an Aerial Biosphere in Temperate Sub Neptune-Sized Exoplanet Atmospheres. *Universe*, 7(6):172.

Seager, S. and Sasselov, D. D. (2000). Theoretical Transmission Spectra during Extrasolar Giant Planet Transits. *ApJ*, 537(2):916–921.

Seidel, J. V., Ehrenreich, D., Pino, L., Bourrier, V., Lavie, B., Allart, R., Wyttenbach, A., and Lovis, C. (2020). Wind of change: retrieving exoplanet atmospheric winds from high-resolution spectroscopy. *A&A*, 633:A86.

Seidel, J. V., Ehrenreich, D., Wyttenbach, A., Allart, R., Lendl, M., Pino, L., Bourrier, V., Cegla, H. M., Lovis, C., Barrado, D., Bayliss, D., Astudillo-Defru, N., Deline, A., Fisher, C., Heng, K., Joseph, R., Lavie, B., Melo, C., Pepe, F., Ségransan, D., and Udry, S. (2019). Hot Exoplanet Atmospheres Resolved with Transit Spectroscopy (HEARTS). II. A broadened sodium feature on the ultra-hot giant WASP-76b. *A&A*, 623:A166.

Sestovic, M. and Demory, B.-O. (2020). Occurrence rate of exoplanets orbiting ultracool dwarfs as probed by K2. *A&A*, 641:A170.

Shallue, C. J. and Vanderburg, A. (2018). Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90. *AJ*, 155(2):94.

Showman, A. P., Fortney, J. J., Lewis, N. K., and Shabram, M. (2013). Doppler Signatures of the Atmospheric Circulation on Hot Jupiters. *ApJ*, 762(1):24.

Sing, D. K., Fortney, J. J., Nikolov, N., Wakeford, H. R., Kataria, T., Evans, T. M., Aigrain, S., Ballester, G. E., Burrows, A. S., Deming, D., Désert, J.-M., Gibson, N. P., Henry, G. W., Huitson, C. M., Knutson, H. A., Lecavelier Des Etangs, A., Pont, F., Showman, A. P., Vidal-Madjar, A., Williamson, M. H., and Wilson, P. A. (2016). A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. *Nature*, 529(7584):59–62.

Sing, D. K., Pont, F., Aigrain, S., Charbonneau, D., Désert, J. M., Gibson, N., Gilliland, R., Hayek, W., Henry, G., Knutson, H., Lecavelier Des Etangs, A., Mazeh, T., and Shporer, A. (2011). Hubble Space Telescope transmission spectroscopy of the exoplanet HD 189733b: high-altitude atmospheric haze in the optical and near-ultraviolet with STIS. *MNRAS*, 416(2):1443–1455.

Snellen, I. A. G., Albrecht, S., de Mooij, E. J. W., and Le Poole, R. S. (2008). Ground-based detection of sodium in the transmission spectrum of exoplanet HD 209458b. *A&A*, 487(1):357–362.

Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., and Albrecht, S. (2010). The orbital motion, absolute mass and high-altitude winds of exoplanet HD209458b. *Nature*, 465(7301):1049–1051.

Soboczenski, F., Himes, M. D., O'Beirne, M. D., Zorzan, S., Gunes Baydin, A., Cobb, A. D., Gal, Y., Angerhausen, D., Mascaro, M., Arney, G. N., and Domagal-Goldman, S. D. (2018). Bayesian Deep Learning for Exoplanet Atmospheric Retrieval. *arXiv e-prints*, page arXiv:1811.03390.

Sousa-Silva, C., Seager, S., Ranjan, S., Petkowski, J. J., Zhan, Z., Hu, R., and Bains, W. (2020). Phosphine as a Biosignature Gas in Exoplanet Atmospheres. *Astrobiology*, 20(2):235–268.

Stangret, M., Casasayas-Barris, N., Pallé, E., Yan, F., Sánchez-López, A., and López-Puertas, M. (2020). Detection of Fe I and Fe II in the atmosphere of MASCARA-2b using a cross-correlation method. *A&A*, 638:A26.

Stevenson, K. B., Harrington, J., Nymeyer, S., Madhusudhan, N., Seager, S., Bowman, W. C., Hardy, R. A., Deming, D., Rauscher, E., and Lust, N. B. (2010). Possible thermochemical disequilibrium in the atmosphere of the exoplanet GJ 436b. *Nature*, 464(7292):1161–1164.

Stock, J. W., Kitzmann, D., Patzer, A. B. C., and Sedlmayr, E. (2018). FastChem: A computer program for efficient complex chemical equilibrium calculations in the neutral/ionized gas phase with applications to stellar and planetary atmospheres. *MNRAS*, 479(1):865–874.

Tabernero, H. M., Zapatero Osorio, M. R., Allart, R., Borsa, F., Casasayas-Barris, N., Demangeon, O., Ehrenreich, D., Lillo-Box, J., Lovis, C., Pallé, E., Sousa, S. G., Rebolo, R., Santos, N. C., Pepe, F., Cristiani, S., Adibekyan, V., Allende Prieto, C., Alibert, Y., Barros, S. C. C., Bouchy, F., Bourrier, V., D'Odorico, V., Dumusque, X., Faria, J. P., Figueira, P., Génova Santos, R., González Hernández, J. I., Hojjatpanah, S., Lo Curto, G., Lavie, B., Martins, C. J. A. P., Martins, J. H. C., Mehner, A., Micela, G., Molaro, P., Nunes, N. J., Poretti, E., Seidel, J. V., Sozzetti, A., Suárez Mascareño, A., Udry, S., Aliverti, M., Affolter, M., Alves, D., Amate, M., Avila, G., Bandy, T., Benz, W., Bianco, A., Broeg, C., Cabral, A., Conconi, P., Coelho, J., Cumani, C., Deiries, S., Dekker, H., Delabre, B., Fragoso, A., Genoni, M., Genolet, L., Hughes, I., Knudstrup, J., Kerber, F., Landoni, M., Lizon, J. L., Maire, C., Manescau, A., Di Marcantonio, P., Mégevand, D., Monteiro, M., Monteiro, M., Moschetti, M., Mueller, E., Modigliani, A., Oggioni, L., Oliveira, A., Pariani, G., Pasquini, L., Rasilla, J. L., Redaelli, E., Riva, M., Santana-Tschudi, S., Santin, P., Santos, P., Segovia, A., Sosnowska, D., Spanò, P., Tenegi, F., Iwert, O., Zanutta, A., and Zerbi, F. (2021). ESPRESSO high-resolution transmission spectroscopy of WASP-76 b. *A&A*, 646:A158.

Taylor, J., Parmentier, V., Irwin, P. G. J., Aigrain, S., Lee, G. K. H., and Krissansen-Totton, J. (2020). Understanding and mitigating biases when studying inhomogeneous emission spectra with JWST. *MNRAS*, 493(3):4342–4354.

Tennyson, J. and Yurchenko, S. N. (2012). ExoMol: molecular line lists for exoplanet and other atmospheres. *MNRAS*, 425(1):21–33.

Tennyson, J., Yurchenko, S. N., Al-Refaie, A. F., Clark, V. H. J., Chubb, K. L., Conway, E. K., Dewan, A., Gorman, M. N., Hill, C., Lynas-Gray, A. E., Mellor, T., McKemmish, L. K., Owens, A., Polyansky, O. L., Semenov, M., Somogyi, W., Tinetti, G., Upadhyay, A.,

Waldmann, I., Wang, Y., Wright, S., and Yurchenko, O. P. (2020). The 2020 release of the ExoMol database: Molecular line lists for exoplanet and other hot atmospheres. *JQSRT*, 255:107228.

Thatte, N., Tecza, M., Schnetler, H., Neichel, B., Melotte, D., Fusco, T., Ferraro-Wood, V., Clarke, F., Bryson, I., O'Brien, K., Mateo, M., Garcia Lorenzo, B., Evans, C., Bouché, N., Arribas, S., and HARMONI Consortium (2021). HARMONI: the ELT's First-Light Near-infrared and Visible Integral Field Spectrograph. *The Messenger*, 182:7–12.

The LUVOIR Team (2019). The LUVOIR Mission Concept Study Final Report. *arXiv e-prints*, page arXiv:1912.06219.

Thompson, S. E., Mullally, F., Coughlin, J., Christiansen, J. L., Henze, C. E., Haas, M. R., and Burke, C. J. (2015). A Machine Learning Technique to Identify Transit Shaped Signals. *ApJ*, 812(1):46.

Tinetti, G., Drossart, P., Eccleston, P., Hartogh, P., Heske, A., Leconte, J., Micela, G., Ollivier, M., Pilbratt, G., Puig, L., Turrini, D., Vandenbussche, B., Wolkenberg, P., Beaulieu, J.-P., Buchave, L. A., Ferus, M., Griffin, M., Guedel, M., Justtanont, K., Lagage, P.-O., Machado, P., Malaguti, G., Min, M., Nørgaard-Nielsen, H. U., Rataj, M., Ray, T., Ribas, I., Swain, M., Szabo, R., Werner, S., Barstow, J., Burleigh, M., Cho, J., du Foresto, V. C., Coustenis, A., Decin, L., Encrenaz, T., Galand, M., Gillon, M., Helled, R., Morales, J. C., Muñoz, A. G., Moneti, A., Pagano, I., Pascale, E., Piccioni, G., Pinfield, D., Sarkar, S., Selsis, F., Tennyson, J., Triaud, A., Venot, O., Waldmann, I., Waltham, D., Wright, G., Amiaux, J., Auguères, J.-L., Berthé, M., Bezawada, N., Bishop, G., Bowles, N., Coffey, D., Colomé, J., Crook, M., Crouzet, P.-E., Da Peppo, V., Sanz, I. E., Focardi, M., Frericks, M., Hunt, T., Kohley, R., Middleton, K., Morgante, G., Ottensamer, R., Pace, E., Pearson, C., Stamper, R., Symonds, K., Rengel, M., Renotte, E., Ade, P., Affer, L., Alard, C., Allard, N., Altieri, F., André, Y., Arena, C., Argyriou, I., Aylward, A., Baccani, C., Bakos, G., Banaszkiewicz, M., Barlow, M., Batista, V., Bellucci, G., Benatti, S., Bernardi, P., Bézard, B., Blecka, M., Bolmont, E., Bonfond, B., Bonito, R., Bonomo, A. S., Brucato, J. R., Brun, A. S., Bryson, I., Bujwan, W., Casewell, S., Charnay, B., Pestellini, C. C., Chen, G., Ciaravella, A., Claudi, R., Clédassou, R., Damasso, M., Damiano, M., Danielski, C., Deroo, P., Di Giorgio, A. M., Dominik, C., Doublier, V., Doyle, S., Doyon, R., Drummond, B., Duong, B., Eales, S., Edwards, B., Farina, M., Flaccomio, E., Fletcher, L., Forget, F., Fossey, S., Fränz, M., Fujii, Y., García-Piquer, Á., Gear, W., Geoffray, H., Gérard, J. C., Gesa, L., Gomez, H., Graczyk, R., Griffith, C., Grodent, D., Guarcello, M. G., Gustin, J., Hamano, K., Hargrave, P., Hello, Y., Heng, K., Herrero, E., Hornstrup, A., Hubert, B., Ida, S., Ikoma, M., Iro, N., Irwin, P., Jarchow, C., Jaubert, J., Jones, H., Julien, Q., Kameda, S., Kerschbaum, F., Kervella, P., Koskinen, T., Krijger, M., Krupp, N., Lafarga, M., Landini, F., Lellouch, E., Leto, G., Luntzer, A., Rank-Lüftinger, T., Maggio, A., Maldonado, J., Maillard, J.-P., Mall, U., Marquette, J.-B., Mathis, S., Maxted, P., Matsuo, T., Medvedev, A., Miguel, Y., Minier, V., Morello, G., Mura, A., Narita, N., Nascimbeni, V., Nguyen Tong, N., Noce, V., Oliva, F., Palle, E., Palmer, P., Pancrazzi, M., Papageorgiou, A., Parmentier, V., Perger, M., Petralia, A., Pezzuto, S., Pierrehumbert, R., Pillitteri, I., Piotto, G., Pisano, G., Prisinzano, L., Radioti, A., Réess,

J.-M., Rezac, L., Rocchetto, M., Rosich, A., Sanna, N., Santerne, A., Savini, G., Scandariato, G., Sicardy, B., Sierra, C., Sindoni, G., Skup, K., Snellen, I., Sobiecki, M., Soret, L., Sozzetti, A., Stiepen, A., Strugarek, A., Taylor, J., Taylor, W., Terenzi, L., Tessenyi, M., Tsiaras, A., Tucker, C., Valencia, D., Vasisht, G., Vazan, A., Vilardell, F., Vinatier, S., Viti, S., Waters, R., Wawer, P., Wawrzaszek, A., Whitworth, A., Yung, Y. L., Yurchenko, S. N., Osorio, M. R. Z., Zellem, R., Zingales, T., and Zwart, F. (2018). A chemical survey of exoplanets with ARIEL. *Experimental Astronomy*, 46(1):135–209.

Todorov, K. O., Line, M. R., Pineda, J. E., Meyer, M. R., Quanz, S. P., Hinkley, S., and Fortney, J. J. (2016). The Water Abundance of the Directly Imaged Substellar Companion κ And b Retrieved from a Near Infrared Spectrum. *ApJ*, 823(1):14.

Tremblin, P., Amundsen, D. S., Mourier, P., Baraffe, I., Chabrier, G., Drummond, B., Homeier, D., and Venot, O. (2015). Fingering Convection and Cloudless Models for Cool Brown Dwarf Atmospheres. *ApJL*, 804(1):L17.

Tsai, S.-M., Kitzmann, D., Lyons, J. R., Mendonça, J., Grimm, S. L., and Heng, K. (2018). Toward Consistent Modeling of Atmospheric Chemistry and Dynamics in Exoplanets: Validation and Generalization of the Chemical Relaxation Method. *ApJ*, 862(1):31.

Tsai, S.-M., Lyons, J. R., Grosheintz, L., Rimmer, P. B., Kitzmann, D., and Heng, K. (2017). VULCAN: An Open-source, Validated Chemical Kinetics Python Code for Exoplanetary Atmospheres. *ApJS*, 228(2):20.

Tsiaras, A., Waldmann, I. P., Tinetti, G., Tennyson, J., and Yurchenko, S. N. (2019). Water vapour in the atmosphere of the habitable-zone eight-Earth-mass planet K2-18 b. *Nature Astronomy*, 3:1086–1091.

Tsiaras, A., Waldmann, I. P., Zingales, T., Rocchetto, M., Morello, G., Damiano, M., Karpouzas, K., Tinetti, G., McKemmish, L. K., Tennyson, J., and Yurchenko, S. N. (2018). A Population Study of Gaseous Exoplanets. *AJ*, 155(4):156.

Udalski, A., Paczynski, B., Zebrun, K., Szymanski, M., Kubiak, M., Soszynski, I., Szewczyk, O., Wyrzykowski, L., and Pietrzynski, G. (2002). The Optical Gravitational Lensing Experiment. Search for Planetary and Low-Luminosity Object Transits in the Galactic Disk. Results of 2001 Campaign. *AcA*, 52:1–37.

Van Zeggeren, F. and Storey, S. H, j. a. (1970). *The computation of chemical equilibria.* [London] : Cambridge University Press. Bibliography: p. [163]-171.

Venot, O., Hébrard, E., Agúndez, M., Dobrijevic, M., Selsis, F., Hersant, F., Iro, N., and Bounaceur, R. (2012). A chemical model for the atmosphere of hot Jupiters. *A&A*, 546:A43.

Wakeford, H. R., Sing, D. K., Deming, D., Lewis, N. K., Goyal, J., Wilson, T. J., Barstow, J., Kataria, T., Drummond, B., Evans, T. M., Carter, A. L., Nikolov, N., Knutson, H. A., Ballester, G. E., and Mandell, A. M. (2018). The Complete Transmission Spectrum of WASP-39b with a Precise Water Constraint. *AJ*, 155(1):29.

Wakeford, H. R., Sing, D. K., Kataria, T., Deming, D., Nikolov, N., Lopez, E. D., Tremblin, P., Amundsen, D. S., Lewis, N. K., Mandell, A. M., Fortney, J. J., Knutson, H., Benneke, B., and Evans, T. M. (2017). HAT-P-26b: A Neptune-mass exoplanet with a well-constrained heavy element abundance. *Science*, 356(6338):628–631.

Waldmann, I. P. (2016). Dreaming of Atmospheres. *ApJ*, 820(2):107.

Waldmann, I. P., Tinetti, G., Rocchetto, M., Barton, E. J., Yurchenko, S. N., and Tennyson, J. (2015a). Tau-REx I: A Next Generation Retrieval Code for Exoplanetary Atmospheres. *ApJ*, 802(2):107.

Waldmann, I. P., Rocchetto, M., Tinetti, G., Barton, E. J., Yurchenko, S. N., and Tennyson, J. (2015b). Tau-REx II: Retrieval of Emission Spectra. *ApJ*, 813(1):13.

Welbanks, L., Madhusudhan, N., Allard, N. F., Hubeny, I., Spiegelman, F., and Leininger, T. (2019). Mass-Metallicity Trends in Transiting Exoplanets from Atmospheric Abundances of H_2O, Na, and K. *ApJL*, 887(1):L20.

Witze, A. (2020). How the fight over a Hawaii mega-telescope could change astronomy. *Nature*, 577(7791):457–458.

Wolszczan, A. and Frail, D. A. (1992). A planetary system around the millisecond pulsar PSR1257 + 12. *Nature*, 355(6356):145–147.

Wyttenbach, A., Ehrenreich, D., Lovis, C., Udry, S., and Pepe, F. (2015). Spectrally resolved detection of sodium in the atmosphere of HD 189733b with the HARPS spectrograph. *A&A*, 577:A62.

Wyttenbach, A., Lovis, C., Ehrenreich, D., Bourrier, V., Pino, L., Allart, R., Astudillo-Defru, N., Cegla, H. M., Heng, K., Lavie, B., Melo, C., Murgas, F., Santerne, A., Ségransan, D., Udry, S., and Pepe, F. (2017). Hot Exoplanet Atmospheres Resolved with Transit Spectroscopy (HEARTS). I. Detection of hot neutral sodium at high altitudes on WASP-49b. *A&A*, 602:A36.

Yang, H., Komacek, T. D., and Abbot, D. S. (2019a). Effects of Radius and Gravity on the Inner Edge of the Habitable Zone. *ApJL*, 876(2):L27.

Yang, J., Abbot, D. S., Koll, D. D. B., Hu, Y., and Showman, A. P. (2019b). Ocean Dynamics and the Inner Edge of the Habitable Zone for Tidally Locked Terrestrial Planets. *ApJ*, 871(1):29.

Yu, L., Vanderburg, A., Huang, C., Shallue, C. J., Crossfield, I. J. M., Gaudi, B. S., Daylan, T., Dattilo, A., Armstrong, D. J., Ricker, G. R., Vanderspek, R. K., Latham, D. W., Seager, S., Dittmann, J., Doty, J. P., Glidden, A., and Quinn, S. N. (2019). Identifying Exoplanets with Deep Learning. III. Automated Triage and Vetting of TESS Candidates. *AJ*, 158(1):25.

Zerbi, F. M., Bouchy, F., Fynbo, J., Maiolino, R., Piskunov, N., Rebolo Lopez, R., Santos, N., Strassmeier, K., Udry, S., Vanzi, L., Riva, M., Basden, A., Boisse, I., Bonfils, X., Buscher, D., Cabral, A., Dimarcantonio, P., Di Varano, I., Henry, D., Monteiro, M., Morris, T., Murray,

G., Oliva, E., Parry, I., Pepe, F., Quirrenbach, A., Rasilla, J. L., Rees, P., Stempels, E., Valenziano, L., Wells, M., Wildi, F., Origlia, L., Allende Prieto, C., Chiavassa, A., Cristiani, S., Figueira, P., Gustafsson, B., Hatzes, A., Haehnelt, M., Heng, K., Israelian, G., Kochukhov, O., Lovis, C., Marconi, A., Martins, C. J. A. P., Noterdaeme, P., Petitjean, P., Puzia, T., Queloz, D., Reiners, A., and Zoccali, M. (2014). HIRES: the high resolution spectrograph for the E-ELT. In Ramsay, S. K., McLean, I. S., and Takami, H., editors, *Ground-based and Airborne Instrumentation for Astronomy V*, volume 9147 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 914723.

Zhan, Z., Seager, S., Jurand Petkowski, J., Sousa-Silva, C., Ranjan, S., Huang, J., and Bains, W. (2021). Assessment of Isoprene as a Possible Biosignature Gas in Exoplanets with Anoxic Atmospheres. *arXiv e-prints*, page arXiv:2103.14228.

Zhang, J., Kempton, E. M. R., and Rauscher, E. (2017). Constraining Hot Jupiter Atmospheric Structure and Dynamics through Doppler-shifted Emission Spectra. *ApJ*, 851(2):84.

Zingales, T. and Waldmann, I. P. (2018). ExoGAN: Retrieving Exoplanetary Atmospheres Using Deep Convolutional Generative Adversarial Networks. *AJ*, 156(6):268.